



Contents lists available at ScienceDirect

## International Journal of Forecasting

journal homepage: [www.elsevier.com/locate/ijforecast](http://www.elsevier.com/locate/ijforecast)

# Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives

Tommaso Di Fonzo\*, Daniele Girolimetto

Department of Statistical Sciences, University of Padua, Via C. Battisti 241, 35121 Padova, Italy



## ARTICLE INFO

## Keywords:

Linearly constrained multiple time series  
Combining forecasts  
Heuristic techniques  
Evaluating forecasts  
GDP from Income and Expenditure side

## ABSTRACT

Forecast reconciliation is a post-forecasting process aimed to improve the quality of the base forecasts for a system of hierarchical/grouped time series. Cross-sectional and temporal hierarchies have been considered in the literature, but generally, these two features have not been fully considered together. The paper presents two new results by adopting a notation that simultaneously deals with both forecast reconciliation dimensions. (i) The closed-form expression of the optimal (in the least squares sense) point forecasts fulfilling both contemporaneous and temporal constraints. (ii) An iterative procedure that produces cross-temporally reconciled forecasts by alternating forecast reconciliation along one single dimension (either cross-sectional or temporal) at each iteration step. The feasibility of the proposed procedures, along with first evaluations of their performance as compared to the most performing 'single dimension' (either cross-sectional or temporal) forecast reconciliation procedures, is studied through a forecasting experiment on the 95 quarterly time series of the Australian Gross Domestic Product from Income and Expenditure sides. For this dataset, the new procedures, in addition to providing fully coherent forecasts in both cross-sectional and temporal dimensions, improve the forecast accuracy of the state-of-the-art point forecast reconciliation techniques.

© 2021 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

## 1. Introduction

In several operational fields, decisions to be successful require the support of accurate, detailed, and coherent forecasts. Forecasts are coherent when the predicted values at the disaggregate and aggregate scales are equal when brought to the same level. For example, temporally coherent monthly predictions sum up to annual ones, and similarly, geographically coherent regional predictions add up to country-level ones. Such coherence is an important qualifier for forecasts, to support aligned decision making across different planning units and horizons, while avoiding that different decision making units plan on different views of the future. For example, Kourntzes

and Athanasopoulos (2019) generate forecasts for Australian domestic tourism that are coherent across multiple geographical divisions (regions, zones, states, and whole country), but are also coherent across time (at monthly, bi-monthly, quarterly, 4-monthly, 6-monthly, and annual levels), i.e., for different planning horizons.

As a matter of fact, in the growing literature on hierarchical forecast reconciliation, the cross-sectional (contemporaneous) and temporal coherency dimensions are mostly considered in separate ways: either 'time-by-time' cross-sectional forecast reconciliation for a  $n$ -dimensional time series (Hyndman et al., 2011, 2020), or temporal coherency for the forecasts of a single variable for different time frequencies (Athanasopoulos et al., 2017; Hyndman & Kourntzes, 2018). The issue of getting coherent forecasts along both cross-sectional and temporal dimensions (i.e., cross-temporal coherency) has been dealt with by Yagli et al. (2019) and Spiliotis et al. (2020b). However,

\* Corresponding author.

E-mail address: [difonzo@stat.unipd.it](mailto:difonzo@stat.unipd.it) (T. Di Fonzo).

**Symbols and notation used in the paper**

$n_b, n_a, n$	Number of bottom, upper, and total time series ( $n = n_a + n_b$ )
$m$	Highest available sampling frequency per seasonal cycle (max. order of temporal aggregation)
$h \geq 1$	Forecast horizon for the lowest frequency time series
$T$	Number of high-frequency observations used in the forecasting models
$N$	Number of observations at the lowest frequency: $N = \frac{T}{m}$
$\mathcal{K}$	Set of factors of $m$ in descending order: $\mathcal{K} = \{k_p, k_{p-1}, \dots, k_2, k_1\}$ , $k_p = m$ , $k_1 = 1$
$k^*$	$\sum_{j=2}^p \frac{m}{k_j}$
$M_k$	$\frac{m}{k}$ , $k \in \mathcal{K}$
$\mathbf{b}_t \in \mathbb{R}^{n_b}$	vector containing the bottom time series (bts) at time $t$
$\mathbf{a}_t \in \mathbb{R}^{n_a}$	vector containing the upper time series (uts) at time $t$
$\mathbf{y}_t \in \mathbb{R}^n$	vector containing the time series $\mathbf{y}_t = [\mathbf{a}'_t \ \mathbf{b}'_t]'$ at time $t$
$\mathbf{Y}, \hat{\mathbf{Y}}, \tilde{\mathbf{Y}} \in \mathbb{R}^{n \times h(k^*+m)}$	Matrices of target, base and cross-temporally reconciled forecasts
$\mathbf{Y}^{[k]}, k \in \mathcal{K}$	$(n \times hM_k)$ matrix containing the target forecasts of the level $k$ temporally aggregated series. Component of matrix $\mathbf{Y} = [\mathbf{Y}^{[m]} \ \mathbf{Y}^{[k_{p-1}]} \ \dots \ \mathbf{Y}^{[k_2]} \ \mathbf{Y}^{[1]}]$
$\hat{\mathbf{Y}}^{[k]}, k \in \mathcal{K}$	$(n \times hM_k)$ matrix containing the base forecasts of the level $k$ temporally aggregated series. Component of matrix $\hat{\mathbf{Y}} = [\hat{\mathbf{Y}}^{[m]} \ \hat{\mathbf{Y}}^{[k_{p-1}]} \ \dots \ \hat{\mathbf{Y}}^{[k_2]} \ \hat{\mathbf{Y}}^{[1]}]$
$\tilde{\mathbf{Y}}^{[k]}, k \in \mathcal{K}$	$(n \times hM_k)$ matrix containing the cross-temporally reconciled forecasts of the level $k$ temporally aggregated series. Component of matrix $\tilde{\mathbf{Y}} = [\tilde{\mathbf{Y}}^{[m]} \ \tilde{\mathbf{Y}}^{[k_{p-1}]} \ \dots \ \tilde{\mathbf{Y}}^{[k_2]} \ \tilde{\mathbf{Y}}^{[1]}]$
$\mathbf{A}^{[k]}, \mathbf{B}^{[k]}, k \in \mathcal{K}$	$(n_a \times hM_k)$ and $(n_b \times hM_k)$ components of matrix $\mathbf{Y}^{[k]} = \begin{bmatrix} \mathbf{A}^{[k]} \\ \mathbf{B}^{[k]} \end{bmatrix}$

$\mathbf{y}, \hat{\mathbf{y}}, \tilde{\mathbf{y}} \in \mathbb{R}^{hn(k^*+m)}$	$\mathbf{y} = \text{vec}(\mathbf{Y})$ , $\hat{\mathbf{y}} = \text{vec}(\hat{\mathbf{Y}})$ , $\tilde{\mathbf{y}} = \text{vec}(\tilde{\mathbf{Y}})$
$\mathbf{P} \in \mathbb{R}^{hn(k^*+m) \times hn(k^*+m)}$	Commutation matrix, such that $\mathbf{P}[\text{vec}(\mathbf{Y})] = \text{vec}(\mathbf{Y}')$
$\mathbf{C} \in \mathbb{R}^{n_a \times n_b}$	Cross-sectional (contemporaneous) aggregation matrix
$\mathbf{S} \in \mathbb{R}^{n \times n_b}$	Cross-sectional (contemporaneous) summing matrix
$\mathbf{U}' \in \mathbb{R}^{n_a \times n}$	Zero constraints cross-sectional kernel matrix: $\mathbf{U}'\mathbf{Y} = \mathbf{0}_{[n_a \times (k^*+m)]}$
$\mathbf{K}_h \in \mathbb{R}^{hk^* \times hm}$	Temporal aggregation matrix
$\mathbf{R}_h \in \mathbb{R}^{h(k^*+m) \times hm}$	Temporal summing matrix
$\mathbf{Z}'_h \in \mathbb{R}^{hk^* \times h(k^*+m)}$	Zero constraints temporal kernel matrix: $\mathbf{Z}'_h \mathbf{Y}' = \mathbf{0}_{[hk^* \times n]}$
$\mathbf{H}' \in \mathbb{R}^{hn_a^* \times nh(k^*+m)}$	Zero constraints full row-rank cross-temporal kernel matrix: $\mathbf{H}'\mathbf{y} = \mathbf{0}$

as far as we know, the procedure proposed by [Kourentzes and Athanasopoulos \(2019\)](#) is the only one to simultaneously fulfill both cross-sectional and temporal coherency in the final reconciled point forecasts at any considered aggregation dimension.

Whereas the most recent forecast reconciliation procedures for each single coherence dimension are based on some optimality criterion ([van Erven & Cugliari, 2015](#); [Wickramasuriya et al., 2019](#)), the cross-temporal reconciliation procedure by [Kourentzes and Athanasopoulos \(2019\)](#) can be considered as a heuristic with a simple and effective structure, i.e., an approach that employs a practical method that is not guaranteed to be optimal, but which is nevertheless sufficient for reaching an immediate goal, which in our case is the coherency along all the considered dimensions of the reconciled forecasts. This fact is probably because of the consideration that in a cross-temporal forecast reconciliation framework, the complexity and dimensions of the problem grow very quickly along with the requested computational time and memory ([Kourentzes & Athanasopoulos, 2019](#); [Nystrup et al., 2020](#)). This is certainly true, but appropriate, thorough use of some well known linear algebra tools, and the expanding computation facilities, in terms of both calculation power and memory, makes it feasible to look for the optimal solution (in the least squares sense), expanding the field of application of the ‘forecast reconciliation methodology’ to simultaneously encompass both contemporaneous (cross-sectional) and temporal coherency dimensions.

Bottom-up and top-down approaches to forecast reconciliation are well known to both practitioners and researchers. In short, according to the bottom-up approach

(Dunn et al., 1976), the forecasts at the most disaggregated level are summed up to obtain the aggregates of interest. On the contrary, in the top-down approach (Gross & Sohl, 1990), the most aggregated series is forecasted first and then is disaggregated to provide lower-level predictions (Fliedner, 2001; Athanasopoulos et al., 2009, and the references therein). A combination of these two approaches, known as middle-out (Athanasopoulos et al., 2009), selects an intermediate level of (dis)aggregation and moves downward in a top-down fashion, and onwards according to bottom-up. However, in the last decade, there have been several contributions exploiting a regression-based optimal combination approach (Hyndman et al., 2011). These have proven convincing from a mathematical-statistical point of view, flexible enough to be adapted to both cross-sectional and temporal frameworks (Wickramasuriya et al., 2019, and Athanasopoulos et al., 2017, respectively), and very effective in improving the base forecasts from many different application fields (economics, demography, energy, tourism, etc.).

This paper considers an optimal combination approach, which takes the base (incoherent and however obtained) forecasts of all nodes in the hierarchy and reconciles them. We show that all the linear constraints induced by the cross-temporal hierarchy underlying the time series may be used to reconcile the base forecasts through a simple projection in a well-chosen linear space. At this end, we extend the optimal (in the least squares sense) solutions separately proposed for each coherency dimension, developing the optimal point forecasting procedure, which - while exploiting both cross-sectional and temporal hierarchies - simultaneously fulfills both contemporaneous and temporal constraints. We refer to this as optimal combination cross-temporal forecast reconciliation. In addition, grounding on the existing literature on this topic (Wickramasuriya et al., 2019; Athanasopoulos et al., 2017; Nystrup et al., 2020), we discuss some simple approximations of the covariance matrix to be used in the statistical point forecast reconciliation, focusing on those making use of the in-sample residuals (when available) of the models used to get the base forecasts. The strictly, and very important related issue of probabilistic forecast reconciliation (Ben Taieb et al., 2021; Panagiotelis et al., 2020) is not considered in this paper and will be dealt with in the next future.

The paper is organized as follows. We start by presenting the general framework of point forecast reconciliation according to Byron (1978) (see also van Erven & Cugliari, 2015, and Panagiotelis et al., 2021), avoiding reference to cross-sectional or time indices (Section 2). Hierarchical and grouped systems of multivariate time series are introduced in Section 3. The temporal hierarchies which characterize a single time series are discussed in Section 4. The cross-sectional and temporal coherency dimensions are dealt with simultaneously in Section 5, and the optimal (in the least squares sense) solution to the cross-temporal forecast reconciliation problem is then developed in Section 6. An iterative cross-temporal forecast reconciliation procedure that extends the heuristics proposed by Kourentzes and Athanasopoulos (2019) is described in Section 7. The feasibility of all the proposed

new procedures, along with the evaluation of their performance when compared to the most performing ‘single dimension’ (either cross-sectional or temporal) forecast reconciliation procedures, is studied in Section 8. This is completed through a forecasting experiment on the 95 quarterly time series of the Australian GDP from Income and Expenditure sides considered by Athanasopoulos et al. (2019) and Bisaglia et al. (2020). Section 9 presents conclusions and lists some topics in this field worth to be considered for future research. An online appendix contains complementary materials on methodological and practical issues, and supplementary tables and graphs related to the empirical application.

## 2. Optimal point forecast reconciliation

Forecast reconciliation is a post-forecasting process aimed to improve the quality of the base forecasts for a system of hierarchical/grouped, and more generally linearly constrained, time series (Hyndman et al., 2011; Panagiotelis et al., 2021) by exploiting the constraints that the series in the system must fulfill, whereas in general, the base forecasts do not. Let  $\mathbf{y}$  be a  $(s \times 1)$  vector of target point forecasts that is needed to satisfy the system of linearly independent constraints

$$\mathbf{H}'\mathbf{y} = \mathbf{0}_{(r \times 1)}, \tag{1}$$

where  $\mathbf{H}'$  is a  $(r \times s)$  matrix, with  $\text{rank}(\mathbf{H}') = r < s$ , and  $\mathbf{0}_{(r \times 1)}$  is a  $(r \times 1)$  null vector. Let  $\hat{\mathbf{y}}$  be a  $(s \times 1)$  vector of unbiased point forecasts, not fulfilling the linear constraints (1) (i.e.,  $\mathbf{H}'\hat{\mathbf{y}} \neq \mathbf{0}$ ).

Drawing upon Stone et al. (1942) and Byron (1978), we consider a regression-based reconciliation method assuming that  $\hat{\mathbf{y}}$  is related to  $\mathbf{y}$  by

$$\hat{\mathbf{y}} = \mathbf{y} + \boldsymbol{\varepsilon}, \tag{2}$$

where  $\boldsymbol{\varepsilon}$  is a  $(s \times 1)$  vector of zero mean disturbances, with known p.d. covariance matrix  $\mathbf{W}$ . The reconciled forecasts  $\tilde{\mathbf{y}}$  are found by minimizing the generalized least squares (GLS) objective function  $(\hat{\mathbf{y}} - \mathbf{y})' \mathbf{W}^{-1} (\hat{\mathbf{y}} - \mathbf{y})$  constrained by (1):

$$\tilde{\mathbf{y}} = \underset{\mathbf{y}}{\text{argmin}} (\mathbf{y} - \hat{\mathbf{y}})' \mathbf{W}^{-1} (\mathbf{y} - \hat{\mathbf{y}}), \quad \text{s.t. } \mathbf{H}'\mathbf{y} = \mathbf{0}.$$

The solution is given by (Byron, 1978, p. 360):

$$\tilde{\mathbf{y}} = \hat{\mathbf{y}} - \mathbf{W}\mathbf{H}(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1}\mathbf{H}'\hat{\mathbf{y}} = \mathbf{M}\hat{\mathbf{y}}, \tag{3}$$

where  $\mathbf{M} = \mathbf{I}_s - \mathbf{W}\mathbf{H}(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1}\mathbf{H}'$  is a  $(s \times s)$  projection matrix.<sup>1</sup> Denoting with  $\mathbf{d}_{\hat{\mathbf{y}}} = \mathbf{0} - \mathbf{H}'\hat{\mathbf{y}}$  the  $(r \times 1)$  vector containing the base forecasts’ ‘coherency’ errors, we can re-state expression (3) as  $\tilde{\mathbf{y}} = \hat{\mathbf{y}} + \mathbf{W}\mathbf{H}(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1}\mathbf{d}_{\hat{\mathbf{y}}}$ , which makes it clear that the reconciliation formula (3) simply ‘adjusts’ the original forecasts vector  $\hat{\mathbf{y}}$  with a linear combination - according to the smoothing matrix  $\mathbf{W}\mathbf{H}(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1}$  - of the coherency errors in the base forecasts. In addition, if the error term of the model (2) is

<sup>1</sup> A geometric interpretation of the entire hierarchical forecasting problem is given by Panagiotelis et al. (2021).

Gaussian, the reconciliation error  $\tilde{\varepsilon} = \tilde{\mathbf{y}} - \mathbf{y}$  is a zero-mean Gaussian vector with covariance matrix

$$E(\tilde{\mathbf{y}} - \mathbf{y})(\tilde{\mathbf{y}} - \mathbf{y})' = \mathbf{W} - \mathbf{W}\mathbf{H}(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1}\mathbf{H}' = \mathbf{M}\mathbf{W}.$$

Hyndman et al. (2011) propose an alternative formulation as for the reconciled estimates, equivalent to the expression (3) and obtained by GLS estimation of the model  $\hat{\mathbf{y}} = \mathbf{S}\boldsymbol{\beta} + \varepsilon$ , where  $\mathbf{S}$  is a ‘structural summation matrix’ describing the aggregation relationships operating on  $\mathbf{y}$ , and  $\boldsymbol{\beta}$  is a subset of  $\mathbf{y}$  containing the target forecasts of the bottom level series, such that  $\mathbf{y} = \mathbf{S}\boldsymbol{\beta}$  (see Section 3). Since the hypotheses on  $\varepsilon$  remain unchanged, it can be shown (Wickramasuriya et al., 2019) that  $\tilde{\boldsymbol{\beta}} = (\mathbf{S}'\mathbf{W}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}^{-1}\hat{\mathbf{y}}$  is the best linear unbiased estimate of  $\boldsymbol{\beta}$ , and that the whole reconciled forecasts vector is given by  $\tilde{\mathbf{y}} = \mathbf{S}\tilde{\boldsymbol{\beta}} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}$ , where  $\mathbf{G} = (\mathbf{S}'\mathbf{W}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}^{-1}$ .

As witnessed by the extensive literature on adjusting preliminary data (as the base forecasts can be considered) in order to fulfill some externally imposed constraints (see, *inter alia*, Stone et al., 1942 and Byron, 1978), the distinctive feature of the generalized least-squares reconciliation approach is that it can take into account the ‘quality’, however measured, of the preliminary estimates, through an appropriate choice of the covariance matrix  $\mathbf{W}$ . However, for a long time, these procedures have depended on the assumption that this matrix (or any other indicators of the estimates’ accuracy) of the figures to be reconciled was known. In many practical situations,  $\mathbf{W}$  is assumed to be diagonal, and the data are adjusted in the light of their relative variances to satisfy the linear restrictions. But another - perhaps more delicate - challenge arises when either any reliability measure is available, or it can hardly be deduced by the data. The solutions proposed in literature for this case are basically of two types, both consistent with the least-squares approach shown so far:

1. mathematical/mechanical solutions: the base forecasts are balanced by minimizing a penalty criterion which ‘induces’ a covariance matrix (which is simply a statistical artifact);
2. data-based solutions: the variability of the base forecasts to be reconciled is estimated through the models and the data used to produce the forecasts.

As for point forecast reconciliation, in the following, we will consider both approaches, with an explicit preference towards approximations of  $\mathbf{W}$  based on the in-sample residuals (when available), which appear more convincing from a statistical point of view, and generally well-performing in practical applications. However, this topic deserves further attention (Jeon et al., 2019, p. 368), and will be considered for future research.

### 3. Hierarchical and grouped time series

Following Panagiotelis et al. (2021), we define a linearly constrained time series  $\mathbf{y}_t$  as a  $n$ -dimensional time series such that all observed values  $\mathbf{y}_1 \dots \mathbf{y}_T$  and all future values  $\mathbf{y}_{T+1}, \mathbf{y}_{T+2} \dots$  lie in the coherent linear subspace  $S \subset \mathbb{R}^n$ , that is:  $\mathbf{y}_t \in S, \forall t$ . In many situations, the time

series are linked through summation constraints, inducing a hierarchy. Fig. 1 (left) gives an example of a hierarchical system with eight variables and three levels: the top-variable at level 0, two variables (A and B) at level 1, and five variables at level 2 (AA, AB, BA, BB, BC). The temporal observations of these variables form a hierarchical time series, consisting of five bottom time series (bts) and three aggregated upper time series (uts).

Assuming that the relationship mapping the lower-level series in the hierarchy of Fig. 1 (left) into the higher ones always be a simple summation, the bottom-level series can be thought of as building blocks that cannot be obtained as a sum of other series in the hierarchy, while all the series at upper levels can be expressed by appropriately summing part or all of them. For all time periods  $t = 1, \dots, T$ , the link between the top level series  $y_t$  and the bottom level series is given by:

$$y_t = y_{AA,t} + y_{AB,t} + y_{BA,t} + y_{BB,t} + y_{BC,t}. \tag{4}$$

At the same time, the nodes at the intermediate level of the hierarchy satisfy the aggregation constraints:

$$\begin{aligned} y_{A,t} &= y_{AA,t} + y_{AB,t} \\ y_{B,t} &= y_{BA,t} + y_{BB,t} + y_{BC,t} \end{aligned} \tag{5}$$

Consider now the matrices  $\mathbf{C}$ ,  $\mathbf{S}$ , and  $\mathbf{U}'$ , of dimension  $(3 \times 5)$ ,  $(8 \times 5)$ , and  $(3 \times 8)$ , respectively:

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

$$\mathbf{S} = \begin{bmatrix} \mathbf{C} \\ \mathbf{I}_5 \end{bmatrix}, \quad \mathbf{U}' = [\mathbf{I}_3 \quad -\mathbf{C}],$$

where matrix  $\mathbf{U}'$  encodes each summation relationship in a row, with 1 at the associated node, and -1 at its leaves. Expressions (4) and (5) can be written in a more compact way if we define the vectors of *bottom level* ( $\mathbf{b}_t$ ) and *upper level* ( $\mathbf{a}_t$ ) time series at time  $t$  as, respectively,

$$\mathbf{b}_t = \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{BC,t} \end{bmatrix}, \quad \mathbf{a}_t = \begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \end{bmatrix}. \text{ Denoting by } \mathbf{y}_t \text{ the } (8 \times 1)$$

vector  $\mathbf{y}_t = [\mathbf{a}_t' \quad \mathbf{b}_t']'$ , the relationships linking bottom and upper time series can be equivalently expressed as:

$$\mathbf{a}_t = \mathbf{C}\mathbf{b}_t, \quad \mathbf{y}_t = \mathbf{S}\mathbf{b}_t, \quad \mathbf{U}'\mathbf{y}_t = \mathbf{0}_{(3 \times 1)}, \quad t = 1, \dots, T. \tag{6}$$

Thus, for any time index  $t$ ,  $\mathbf{y}_t$  is in the kernel of  $\mathbf{U}'$ , also known as null-space of the linear transformation induced by matrix  $\mathbf{U}'$ , which is given by the set of vectors  $\mathbf{v} \in \mathbb{R}^8$ , such that  $\mathbf{U}'\mathbf{v} = \mathbf{0}_{(3 \times 1)}$  (Harville, 2008, p. 591). We call *structural representation* of series  $\mathbf{y}_t$  the formulation  $\mathbf{y}_t = \mathbf{S}\mathbf{b}_t, t = 1, \dots, T$ , and *zero constraints kernel representation* of series  $\mathbf{y}_t$  the equivalent expression  $\mathbf{U}'\mathbf{y}_t = \mathbf{0}, t = 1, \dots, T$ .

A linearly constrained time series formed by two or more hierarchical time series sharing the same top level series, and the same bottom level series, is called grouped

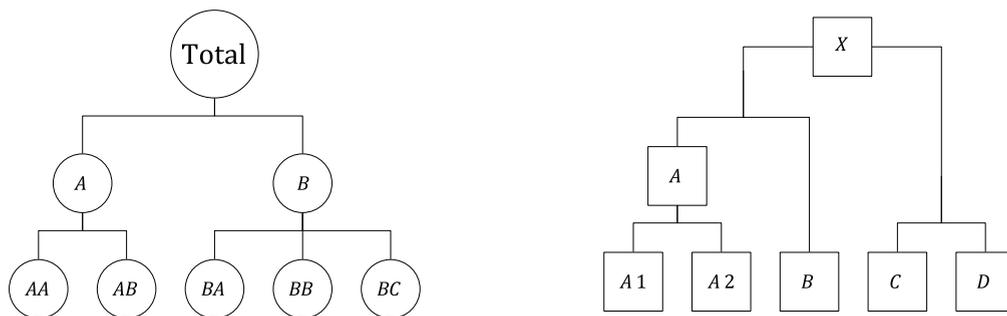


Fig. 1. Two examples of linearly constrained time series. Left: a simple three-level hierarchical structure. Right: two hierarchies sharing the same top-level series.

time series (Hyndman et al., 2011; Hyndman & Athanasopoulos, 2018). Provided matrix  $C$  is appropriately designed, the definitions of matrices  $S$  and  $U'$ , depending solely on matrix  $C$ , remain unchanged.

It should be noted that we can face linearly constrained time series for which the structural representation  $y_t = Sb_t$  does not give a straightforward view of the links between the bottom and upper level time series. Fig. 1 (right) shows two very simple hierarchies, where the variables of each hierarchy contribute (from different sides) to the same top level variable  $X$ , and the bottom level series of the hierarchy on the left side ( $A1, A2, B$ ) are independent of those on the right side ( $C, D$ ). The aggregation relationships between the upper variables  $X$  and  $A$ , and the bottom ones  $A1, A2, B, C$ , and  $D$  are given by:

$$\begin{aligned} X &= A1 + A2 + B \\ X &= C + D \\ A &= A1 + A2 \end{aligned} \quad (7)$$

Expression (7) cannot be represented as a mapping from the bottom variables into (themselves and) the upper variables. Nevertheless, it is possible to set up the constraints valid for all the component series in  $y = [X \ A \ A1 \ A2 \ B \ C \ D]'$  through the matrix

$$U' = \begin{bmatrix} 1 & 0 & -1 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & -1 & -1 & 0 & 0 & 0 \end{bmatrix},$$

such that  $U'y = 0_{(3 \times 1)}$ . After simple operations on expression (7), it is found:

$$\begin{aligned} X &= C + D \\ A &= -B + C + D \\ A1 &= -A2 - B + C + D \end{aligned} \quad (8)$$

so we can write  $U'y = 0_{(3 \times 1)}$ , with

$$U' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 1 & 1 & -1 & -1 \end{bmatrix} = [I_3 \quad -C].$$

While there is no practical problem for working with such constraints, it is clear that they do not conform to the visual pattern of the linearly constrained time series in Fig. 1 (right), where  $A1$  appears as a 'bottom variable'. However, in Eq. (8) it is expressed as a linear combination of series  $A2, B, C$ , and  $D$ .

### 3.1. Alternative approximations of the covariance matrix for cross-sectional point forecast reconciliation

Suppose we have the  $(n \times 1)$  vector  $\hat{y}_h$  of unbiased base forecasts for the  $n$  variables of the linearly constrained series  $y_t$  for the forecast horizon  $h$ . If the base forecasts have been independently computed, generally they do not fulfill the cross-sectional aggregation constraints, i.e.  $U'\hat{y}_h \neq 0_{(n \times 1)}$ . By adapting the general point forecast reconciliation formula (3), the vector of reconciled forecasts is given by:

$$\tilde{y}_h = \hat{y}_h - W_{cs}U(U'W_{cs}U)^{-1}U'\hat{y}_h, \quad (9)$$

where  $W_{cs}$  is a  $(n \times n)$  p.d. matrix, assumed known, and suffix 'cs' stands for 'cross-sectional'. Alternative choices for  $W_{cs}$  proposed in literature are the following:

- identity (cs-ols):  $W_{cs} = I_n$  (Hyndman et al., 2011),
- structural (cs-struct):  $W_{cs} = \text{diag}(S1_{n_b})$  (Athanasopoulos et al., 2017),
- series variance (cs-wls):  $W_{cs} = \hat{W}_{cs-var} = I_n \odot \hat{W}_1$  (Hyndman et al., 2016),
- MinT-shr (cs-shr):  $W_{cs} = \hat{W}_{cs-shr} = \hat{\lambda}\hat{W}_{cs-var} + (1 - \hat{\lambda})\hat{W}_1$  (Wickramasuriya et al., 2019),
- MinT-sam (cs-sam):  $W_{cs} = \hat{W}_1$  (Wickramasuriya et al., 2019),

where the symbol  $\odot$  denotes the Hadamard product,  $\hat{\lambda}$  is an estimated shrinkage coefficient (Ledoit & Wolf, 2004),  $\hat{W}_1$  is the covariance matrix of the one-step ahead in-sample forecast errors  $\hat{e}_t = y_t - \hat{y}_t, t = 1, \dots, T$ :

$$\hat{W}_1 = \frac{1}{T} \sum_{t=1}^T \hat{e}_t \hat{e}_t'. \quad (10)$$

The first three matrices are diagonal, and in the first case, the projection is orthogonal. In contrast, the latter two ones (cs-shr and cs-sam) have been proposed within the minimum-trace point forecast reconciliation approach by Wickramasuriya et al. (2019). It should be noted that the quality of the estimate  $\hat{W}_1$  crucially depends on the dimension of  $T$ . In particular, when  $T < n$ , matrix  $\hat{W}_1$  is singular, which prevents the matrix inversion in expression (9). The shrunk version  $\hat{W}_{cs-shr}$  is a feasible alternative, well-performing in many practical situations (Wickramasuriya et al., 2019).

### 3.2. Matrix representation of the cross-sectional constraints

Let us denote with

$$\mathbf{b}_t = [b_{1t} \dots b_{it} \dots b_{n_b t}]', \quad t = 1, \dots, T, \quad (11)$$

the  $T$  vectors, each of dimension  $(n_b \times 1)$ , containing the *high-frequency bottom-time series* (hf-bts), that is the bottom series of the hierarchy/group observed at the highest available temporal frequency. As we shall see in Section 5, in cross-temporal hierarchies of time series, the hf-bts should be considered as the ‘very’ bottom series of the system, since they cannot be formed as either contemporaneous or temporal sum of other variables. Likewise, let us denote with

$$\mathbf{a}_t = [a_{1t} \dots a_{jt} \dots a_{n_a t}]', \quad t = 1, \dots, T, \quad (12)$$

the  $T$  vectors, each of dimension  $(n_a \times 1)$ , containing the *high-frequency upper-time series* (hf-uts), which are the cross-sectionally aggregated series of the hierarchy/group, observed at the highest temporal frequency. At each time  $t = 1, \dots, T$ , the cross-sectional (contemporaneous) aggregation constraints that map the hf-bts into the hf-uts can be written as:

$$\mathbf{a}_t = \mathbf{C}\mathbf{b}_t, \quad t = 1, \dots, T, \quad (13)$$

where  $\mathbf{C}$  is a  $(n_a \times n_b)$  *contemporaneous aggregation matrix*. The structural representation of the linearly constrained time series  $\mathbf{y}_t$  is in turn given by (Hyndman et al., 2011)  $\begin{bmatrix} \mathbf{a}_t \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{C} \\ \mathbf{I}_{n_b} \end{bmatrix} \mathbf{b}_t$ , that is,  $\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$ ,  $t = 1, \dots, T$ , where  $\mathbf{S} = \begin{bmatrix} \mathbf{C} \\ \mathbf{I}_{n_b} \end{bmatrix}$  is a  $(n \times n_b)$  *contemporaneous summing matrix*, with  $n = n_a + n_b$ . The constraints valid for  $\mathbf{y}_t$  can be expressed in kernel form through the  $(n_a \times n)$  *zero constraints matrix*  $\mathbf{U}' = [\mathbf{I}_{n_a} - \mathbf{C}]$ , that is:  $\mathbf{U}'\mathbf{y}_t = \mathbf{0}_{(n_a \times 1)}$ ,  $t = 1, \dots, T$ . Now, denote  $\mathbf{B}$  the  $(n_b \times T)$  matrix containing the  $T$  observations of the  $n_b$  hf-bts of the system:

$$\mathbf{B} = \begin{bmatrix} b_{11} & \dots & b_{1t} & \dots & b_{1T} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{i1} & \dots & b_{it} & \dots & b_{iT} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{n_b 1} & \dots & b_{n_b t} & \dots & b_{n_b T} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^* \\ \vdots \\ \mathbf{b}_i^* \\ \vdots \\ \mathbf{b}_{n_b}^* \end{bmatrix},$$

where  $\mathbf{b}_t$  has been defined by (11), and  $\mathbf{b}_i^* = [b_{i1} \dots b_{it} \dots b_{iT}]'$ ,  $i = 1, \dots, n_b$ , is the  $(T \times 1)$  vector containing all the observations of the  $i$ th univariate hf-bts, where the asterisk in  $\mathbf{b}_i^*$  is used to distinguish this vector, which combines  $b_{it}$  across all times for one series, from

$\mathbf{b}_t$ , which combines  $b_{it}$  across all series for one time. We consider the  $(n_a \times T)$  matrix  $\mathbf{A}$  for the hf-uts as well:

$$\mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1t} & \dots & a_{1T} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{j1} & \dots & a_{jt} & \dots & a_{jT} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n_a 1} & \dots & a_{n_a t} & \dots & a_{n_a T} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_j^* \\ \vdots \\ \mathbf{a}_{n_a}^* \end{bmatrix},$$

where  $\mathbf{a}_t$  was defined by (12), and  $\mathbf{a}_j^* = [a_{j1} \dots a_{jt} \dots a_{jT}]'$ ,  $j = 1, \dots, n_a$ , is the  $(T \times 1)$  vector containing all the observations of the  $j$ th univariate component hf-uts. The cross-sectional (contemporaneous) aggregation relationships (13) linking bottom and upper time series of  $\mathbf{y}_t$  can thus be expressed in compact form by simultaneously encompassing all  $T$  time periods, as

$$\mathbf{A} = \mathbf{C}\mathbf{B}, \quad (14)$$

which is equivalent to

$$\mathbf{U}'\mathbf{Y} = \mathbf{0}_{(n_a \times T)}, \quad (15)$$

where  $\mathbf{Y} = \begin{bmatrix} \mathbf{A}' & \mathbf{B}' \end{bmatrix}'$  is the  $(n \times T)$  matrix containing the observations of all  $n$  series. It is worth noting that the cross-sectional constraints (14) and (15) hold at any time observation index of any temporal frequency. This has to be considered when dealing with cross-temporal hierarchies (see Section 5).

Now, let us consider the vectorized forms of matrices  $\mathbf{B}'$  and  $\mathbf{A}'$ , according to which the data is organized ‘by-variable-first-and-then-by-time’:  $\mathbf{b} = \text{vec}(\mathbf{B}')$ ,  $\mathbf{a} = \text{vec}(\mathbf{A}')$ , where  $\mathbf{b} = [\mathbf{b}_1^* \dots \mathbf{b}_i^* \dots \mathbf{b}_{n_b}^*]'$ , and  $\mathbf{a} = [\mathbf{a}_1^* \dots \mathbf{a}_j^* \dots \mathbf{a}_{n_a}^*]'$  have dimensions  $(Tn_b \times 1)$ , and  $(Tn_a \times 1)$ , respectively. The cross-sectional constraints (14) can thus be expressed as (Harville, 2008, p. 345):

$$\mathbf{a} = (\mathbf{C} \otimes \mathbf{I}_T) \mathbf{b}, \quad (16)$$

where the symbol  $\otimes$  denotes the Kronecker product. Since  $\mathbf{y} = \text{vec}(\mathbf{Y}') = \begin{bmatrix} \text{vec}(\mathbf{A}') \\ \text{vec}(\mathbf{B}') \end{bmatrix} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$ , expression (16) can be also re-stated as:

$$(\mathbf{U}' \otimes \mathbf{I}_T) \mathbf{y} = \mathbf{0}_{(Tn_a \times 1)}. \quad (17)$$

### 4. Temporal hierarchies

Following Athanasopoulos et al. (2017), we consider a time series  $\{x_t\}_{t=1}^T$  observed at the highest available sampling frequency per seasonal cycle, say  $m$  (e.g., month per year,  $m = 12$ , quarter per year,  $m = 4$ , hour per day,  $m =$

24). Given a factor  $k$  of  $m$ ,<sup>2</sup> we can construct a temporally aggregated version of the time series  $x_t$ , through the non-overlapping sums of its  $k$  successive values, which has a seasonal period equal to  $M_k = m/k$ . To avoid ragged-edge data, we assume that the total number of observations of  $x_t$  involved in the non-overlapping aggregation is a multiple of  $m$ , i.e.,  $T = N \cdot m$ , where  $N$  is the length of the most temporally aggregated version of the series, i.e., the series observed at the lowest available frequency. We denote with  $\mathcal{K} = \{k_p, k_{p-1}, \dots, k_2, k_1\}$  the set of the  $p$  factors of  $m$ , in descending order, where  $k_p = m$  and  $k_1 = 1$ . The temporally aggregated series of order  $k$  can be written as

$$x_l^{[k]} = \sum_{t=(l-1)k+1}^{lk} x_t, \quad l = 1, \dots, \frac{T}{k}, \quad k \in \mathcal{K}. \quad (18)$$

Expression (18) accounts also for the temporal aggregation transforming  $x_t$  in itself (i.e.,  $x_t \equiv x_t^{[1]}$ ,  $l = t$ ). Since the observation index  $l$  in (18) varies with each aggregation level  $k$ , to express a common index for all levels, we define  $\tau$  as the observation index of the most aggregated series, such that  $l = \tau$  at that level, i.e.  $x_\tau^{[m]}$ ,  $\tau = 1, \dots, N$ . As for the other temporally aggregated series defined in expression (18), we stack the observations for each aggregation level below  $m$  in the  $(M_k \times 1)$  column vectors

$$\mathbf{x}_\tau^{[k]} = \left[ x_{M_k(\tau-1)+1}^{[k]} \ x_{M_k(\tau-1)+2}^{[k]} \ \dots \ x_{M_k\tau}^{[k]} \right]', \quad (19)$$

$$\tau = 1, \dots, N, \quad k \in \{k_{p-1}, \dots, k_2, 1\}.$$

We may collect  $x_\tau^{[m]}$  and the  $p - 1$  vectors defined by expression (19) in a single column vector by keeping distinct the temporally aggregated data from the high-frequency one:

$$\mathbf{x}_\tau = \left[ x_\tau^{[m]} \ \mathbf{x}_\tau^{[k_{p-1}]} \ \dots \ \mathbf{x}_\tau^{[k_2]} \ \mathbf{x}_\tau^{[1]} \right]'$$

$$= \left[ \mathbf{t}_{x_\tau} \ \mathbf{x}_\tau^{[1]} \right]', \quad \tau = 1, \dots, N,$$

where  $\mathbf{t}_{x_\tau} = \left[ x_\tau^{[m]} \ \mathbf{x}_\tau^{[k_{p-1}]} \ \dots \ \mathbf{x}_\tau^{[k_2]} \right]'$  is a  $(k^* \times 1)$  vector,

with  $k^* = \sum_{j=2}^p \frac{m}{k_j}$ , containing all the temporally aggregated series at the observation index  $\tau$ ,  $\mathbf{x}_\tau^{[1]}$  is the  $(m \times 1)$  vector of observations of the time series at the highest available frequency within the complete  $\tau$ -th cycle, and thus each  $\mathbf{x}_\tau$  has dimension  $[(k^* + m) \times 1]$ . The relationships linking the nodes in the hierarchy can be expressed as we did in (6) for the cross-sectional (contemporaneous) hierarchy case:

$$\mathbf{t}_{x_\tau} = \mathbf{K}_1 \mathbf{x}_\tau^{[1]}, \quad \mathbf{x}_\tau = \mathbf{R}_1 \mathbf{x}_\tau^{[1]}, \quad (20)$$

$$\mathbf{Z}'_1 \mathbf{x}_\tau = \mathbf{0}_{(k^* \times 1)}, \quad \tau = 1, \dots, N,$$

<sup>2</sup> If  $k$  is not a factor of  $m$ , then the seasonality of the aggregate series is a non-integer, and so forecasts of the aggregate are more difficult to compute.

where  $\mathbf{K}_1 = \begin{bmatrix} \mathbf{1}'_m \\ \mathbf{I}_{\frac{m}{k_{p-1}}} \otimes \mathbf{1}'_{k_{p-1}} \\ \vdots \\ \mathbf{I}_{\frac{m}{k_2}} \otimes \mathbf{1}'_{k_2} \end{bmatrix}$  is the  $(k^* \times m)$  temporal aggregation matrix converting the high-frequency observations into lower-frequency (temporally aggregated) ones,

$\mathbf{R}_1 = \begin{bmatrix} \mathbf{K}_1 \\ \mathbf{I}_m \end{bmatrix}$  is the  $[(k^* + m) \times m]$  temporal summing matrix, and  $\mathbf{Z}'_1 = [\mathbf{I}_{k^*} \ - \mathbf{K}_1]$  is the zero constraints kernel matrix valid for  $\mathbf{x}_\tau$ . For example, with quarterly data it is  $m = 4$ ,  $k^* = 3$ , and  $\mathbf{K}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ ,  $\mathbf{R}_1 = \begin{bmatrix} \mathbf{K}_1 \\ \mathbf{I}_3 \end{bmatrix}$ ,

and  $\mathbf{Z}'_1 = [\mathbf{I}_3 \ - \mathbf{K}_1]$ .

The temporal aggregation relationships can be extended to the whole time span covered by series  $x_t$ . Denoting by  $\mathbf{x} = (\mathbf{x}'_1 \ \dots \ \mathbf{x}'_N)$  the  $[N(k^* + m) \times 1]$  vector containing all the data of series  $X$  at any observed temporal frequency, the complete set of temporal aggregation constraints is given by

$$\mathbf{Z}'_N \mathbf{x} = \mathbf{0}_{(Nk^* \times 1)}, \quad (21)$$

where  $\mathbf{Z}'_N = [\mathbf{I}_{Nk^*} \ - \mathbf{K}_N]$ , and  $\mathbf{K}_N = \begin{bmatrix} \mathbf{I}_N \otimes \mathbf{1}'_m \\ \mathbf{I}_{N\frac{m}{k_{p-1}}} \otimes \mathbf{1}'_{k_{p-1}} \\ \vdots \\ \mathbf{I}_{N\frac{m}{k_2}} \otimes \mathbf{1}'_{k_2} \end{bmatrix}$ .

#### 4.1. Alternative approximations of the covariance matrix for point forecast reconciliation through temporal hierarchies

Suppose we have the  $[(k^* + m) \times 1]$  vector  $\hat{\mathbf{x}}_h$  of unbiased base forecasts for the  $p$  temporal aggregates of a single time series  $X$  within a complete time cycle, i.e. at the forecast horizon  $h = 1$  for the lowest (most aggregated) time frequency. If the base forecasts have been independently computed, generally they do not fulfill the temporal aggregation constraints, i.e.  $\mathbf{Z}'_1 \hat{\mathbf{x}}_h \neq \mathbf{0}_{(k^* \times 1)}$ . By adapting the general point forecast reconciliation formula (3), and not considering suffix  $h$  to simplify the notation, the vector of temporally reconciled forecasts is given by:

$$\tilde{\mathbf{x}} = \hat{\mathbf{x}} - \Omega \mathbf{Z}_1 (\mathbf{Z}'_1 \Omega \mathbf{Z}_1)^{-1} \mathbf{Z}'_1 \hat{\mathbf{x}}, \quad (22)$$

where  $\Omega$  is a  $[(k^* + m) \times (k^* + m)]$  p.d. matrix, assumed known.

To consider possible residual-based estimates of matrix  $\Omega$ , denote

$$\hat{\mathbf{e}}_\tau^{[k]} = \mathbf{x}_\tau^{[k]} - \hat{\mathbf{x}}_\tau^{[k]}, \quad \tau = 1, \dots, N, \quad k \in \mathcal{K}, \quad (23)$$

the  $(M_k \times 1)$  vectors of the in-sample residuals at time index  $\tau$  for the models used to generate the base forecasts of the temporally aggregated series of order  $k$ . These vectors can be organized as

$$\widehat{\mathbf{E}}_x^{[k]} = [\hat{\mathbf{e}}_1^{[k]} \ \dots \ \hat{\mathbf{e}}_\tau^{[k]} \ \dots \ \hat{\mathbf{e}}_N^{[k]}]', \quad k \in \mathcal{K}, \quad (24)$$

where each matrix  $\widehat{\mathbf{E}}_x^{[k]}$  has dimension  $(N \times M_k)$ , and then grouped in the  $[N \times (k^* + m)]$  matrix of in-sample

residuals  $\widehat{\mathbf{E}}_x = \begin{bmatrix} \widehat{\mathbf{E}}_x^{[m]} & \widehat{\mathbf{E}}_x^{[k_{p-1}]} & \dots & \widehat{\mathbf{E}}_x^{[k_2]} & \widehat{\mathbf{E}}_x^{[1]} \end{bmatrix}$ . Each column of this matrix contains the in-sample residuals pertaining to a specific node of the temporal hierarchy, thus the sample cross-covariance matrix of the  $k^* + m$  nodes of the temporal hierarchy is given by:

$$\widehat{\Omega} = \frac{1}{N} (\widehat{\mathbf{E}}_x)' \widehat{\mathbf{E}}_x. \tag{25}$$

This matrix is well defined if  $N > (k^* + m)$ , otherwise there might be singularity issues which would prevent its use in expression (22) in place of matrix  $\Omega$ . Athanasopoulos et al. (2017) and Hyndman and Kourntzes (2018) consider the following alternative choices for  $\Omega$  (the suffix ‘t’ stands for ‘temporal’, to keep the ‘t’-procedures distinct from the ‘cs’-ones shown in Section 3.1):

- identity (t-ols):  $\Omega = \mathbf{I}_{k^*+m}$ ,
- structural (t-struct):  $\Omega = \widehat{\Omega}_{t\text{-struct}} = \text{diag}(\mathbf{R}_1 \mathbf{1}_m)$
- hierarchy variance scaling (t-wlsh):  $\Omega = \widehat{\Omega}_{t\text{-wlsh}} = \mathbf{I}_{k^*+m} \odot \widehat{\Omega}$
- series variance scaling (t-wlsv):  $\Omega = \widehat{\Omega}_{t\text{-wlsv}}$
- MinT-shr (t-shr):  $\Omega = \widehat{\Omega}_{t\text{-shr}} = \hat{\lambda} \widehat{\Omega}_{t\text{-wlsh}} + (1 - \hat{\lambda}) \widehat{\Omega}$
- MinT-sam (t-sam):  $\Omega = \widehat{\Omega}$

The series variance scaling matrix  $\widehat{\Omega}_{t\text{-wlsv}}$  is a diagonal matrix “which contains estimates of the in-sample one-step-ahead error variances across each level” (Athanasopoulos et al., 2017, p. 64), that requires a reduced number ( $p$  instead of  $k^* + m$ ) of variances to be estimated as compared to the hierarchy variance scaling matrix  $\widehat{\Omega}_{t\text{-wlsh}}$ . To include potential information in the residuals’ autocorrelation, we also consider two matrices recently proposed by Nystrup et al. (2020):

- auto-covariance scaling (t-acov):  $\Omega = \widehat{\Omega}_{t\text{-acov}}$
- series Markov (t-sar1):  $\Omega = \widehat{\Omega}_{t\text{-sar1}}$

The auto-covariance scaling makes use of the estimates of the full autocovariance matrices within each aggregation level, while ignoring correlations between aggregation

levels:  $\widehat{\Omega}_{t\text{-acov}} = \begin{bmatrix} \widehat{\Omega}^{[m]} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \widehat{\Omega}^{[1]} \end{bmatrix}$ , where the  $(M_k \times$

$M_k)$  matrices  $\widehat{\Omega}^{[k]}$  are given by:

$$\widehat{\Omega}^{[k]} = \frac{1}{N} \sum_{\tau=1}^N \widehat{\mathbf{e}}_{\tau}^{[k]} (\widehat{\mathbf{e}}_{\tau}^{[k]})' = \frac{1}{N} (\widehat{\mathbf{E}}_x^{[k]})' \widehat{\mathbf{E}}_x^{[k]}, \quad k \in \mathcal{K}, \tag{26}$$

with vector  $\widehat{\mathbf{e}}_{\tau}^{[k]}$  and matrix  $\widehat{\mathbf{E}}_x^{[k]}$  given by (23) and (24), respectively. Because it is sometimes difficult to estimate the covariance matrix within each aggregation level without assuming that it has some special form, Nystrup et al. (2020) consider the Toeplitz matrix for the estimated first-order autocorrelation coefficients of the in-sample residuals for the  $p - 1$  levels  $k = k_1, \dots, k_{p-1}$ , of the series’ temporal hierarchy. Denoting these autocorrelation coefficients with  $\rho_{[k]}$ , it is  $\mathbf{r}^{[k]} = \mathbf{1}$ ,  $\mathbf{r}^{[k]} =$

$$\begin{bmatrix} 1 & \dots & \rho_{[k]}^{M_k-1} \\ \vdots & \ddots & \vdots \\ \rho_{[k]}^{M_k-1} & \dots & 1 \end{bmatrix}, \quad k = k_1, \dots, k_{p-1}, \text{ where each}$$

matrix  $\mathbf{r}^{[k]}$ ,  $k \in \mathcal{K}$ , has dimension  $(M_k \times M_k)$ . The  $p$  matrices are used to build the  $[(k^* + m) \times (k^* + m)]$  matrix

$$\mathbf{r} = \begin{bmatrix} 1 & \mathbf{0}' & \dots & \mathbf{0}' \\ \mathbf{0} & \mathbf{r}^{[k_{p-1}]} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{r}^{[1]} \end{bmatrix}, \text{ which can be used to}$$

estimate matrix  $\Omega$  as  $\widehat{\Omega}_{t\text{-sar1}} = \widehat{\Omega}_{t\text{-wlsv}}^{\frac{1}{2}} \mathbf{r} \widehat{\Omega}_{t\text{-wlsv}}^{\frac{1}{2}}$ .

### 5. The cross-temporal forecast reconciliation framework

#### 5.1. Cross-temporal aggregation constraints

The cross-sectional aggregation relationships (17), linking  $n$  series observed at the same time frequency over a  $T$ -length time span, and the equivalent temporal aggregation relationships (20) and (21), valid for an individual variable expressed at different time frequencies, can be simultaneously considered, by extending (i) the cross-sectional constraints to all observation frequencies, and (ii) the temporal aggregation relationships to all variables.

Considering contemporaneous and temporal dimensions in the same framework requires extending and adapting the notations used so far. To this end, define the  $p$  matrices  $\mathbf{Y}^{[k]}$ , each of dimension  $(n \times NM_k)$ , as

$$\mathbf{Y}^{[k]} = \begin{bmatrix} \mathbf{A}^{[k]} \\ \mathbf{B}^{[k]} \end{bmatrix}, \quad k \in \mathcal{K},$$

where

$$\mathbf{B}^{[k]} = \begin{bmatrix} \mathbf{b}_1^{[k]} & \dots & \mathbf{b}_i^{[k]} & \dots & \mathbf{b}_{n_b}^{[k]} \end{bmatrix},$$

$$\mathbf{A}^{[k]} = \begin{bmatrix} \mathbf{a}_1^{[k]} & \dots & \mathbf{a}_j^{[k]} & \dots & \mathbf{a}_{n_a}^{[k]} \end{bmatrix}, \quad k \in \mathcal{K},$$

are the matrices containing the  $k$ -order temporal aggregates of the bts ( $\mathbf{B}^{[k]}$ ) and uts ( $\mathbf{A}^{[k]}$ ), of dimension  $(n_b \times NM_k)$  and  $(n_a \times NM_k)$ , respectively. To be consistent with the notation so far,  $\mathbf{Y}^{[1]}$ ,  $\mathbf{B}^{[1]}$ , and  $\mathbf{A}^{[1]}$  denote the matrices containing data at the highest available sampling frequency, while  $\mathbf{Y}$ ,  $\mathbf{B}$ , and  $\mathbf{A}$  are used now to denote the matrices containing the data at any considered temporal frequency:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{[m]} & \mathbf{A}^{[k_{p-1}]} & \dots & \mathbf{A}^{[k_2]} & \mathbf{A}^{[1]} \\ \mathbf{B}^{[m]} & \mathbf{B}^{[k_{p-1}]} & \dots & \mathbf{B}^{[k_2]} & \mathbf{B}^{[1]} \end{bmatrix},$$

where  $\mathbf{Y}$ ,  $\mathbf{A}$ , and  $\mathbf{B}$  have  $n$ ,  $n_a$  and  $n_b$  rows, respectively, and the same number of columns,  $[N(k^* + m)]$ .

#### Cross-sectional aggregation constraints

By exploiting the results shown in Section 3.2, the cross-sectional aggregation relationships operating along all the time observation indices can be expressed in compact form as  $\mathbf{U}'\mathbf{Y} = \mathbf{0}_{[n_a \times N(k^*+m)]}$ , or equivalently in vectorized form as

$$(\mathbf{U}' \otimes \mathbf{I}_{N(k^*+m)}) \mathbf{y} = \mathbf{0}_{[n_a N(k^*+m) \times 1]}. \tag{27}$$

#### Temporal aggregation constraints

The temporal aggregation relationships (20), valid for a single series, can be extended to each component of the

time series  $\mathbf{y}_t$  as follows:

$$\begin{bmatrix} \mathbf{A}^{[m]'} & \mathbf{B}^{[m]'} \\ \vdots & \vdots \\ \mathbf{A}^{[k_2]'} & \mathbf{B}^{[k_2]'} \end{bmatrix} = \mathbf{K}_N \begin{bmatrix} \mathbf{A}^{[1]'} & \mathbf{B}^{[1]'} \end{bmatrix}, \quad (28)$$

which can be equivalently written as  $[\mathbf{I}_{Nk^*} \quad -\mathbf{K}_N] \mathbf{Y}' = \mathbf{0}_{(Nk^* \times n)}$ , that is  $\mathbf{Z}'_N \mathbf{Y}' = \mathbf{0}_{(Nk^* \times n)}$ . This last expression can be re-stated in vectorized form as:

$$(\mathbf{I}_n \otimes \mathbf{Z}'_N) \mathbf{y} = \mathbf{0}_{(nNk^* \times 1)}. \quad (29)$$

In summary, by considering expressions (27) and (29) together, the cross-temporal constraints working on the complete set of observations can be expressed as  $\check{\mathbf{H}} \mathbf{y} = \mathbf{0}_{(n^* \times 1)}$ , where  $n^* = n_a N(k^* + m) + nNk^*$ , and  $\check{\mathbf{H}} = \begin{bmatrix} \mathbf{U}' \otimes \mathbf{I}_{N(k^*+m)} \\ \mathbf{I}_n \otimes \mathbf{Z}'_N \end{bmatrix}$  is a  $[n^* \times nN(k^* + m)]$  cross-temporal zero-constraints kernel matrix. Due to the simultaneous consideration of temporal and cross-sectional relationships linking the various time series of the system, some rows of  $\check{\mathbf{H}}$  are redundant, and can be eliminated to get a full row-rank zero-constraints kernel matrix. In detail, matrix  $\check{\mathbf{H}}$  consists in:

- $Nn_a k^*$  rows defining the cross-sectional aggregation constraints operating on the lf-uts;
- $Nn_a m$  rows defining the cross-sectional aggregation constraints operating on the hf-bts;
- $Nn_a (k^* + m)$  rows defining the temporal aggregation constraints operating on both hf- and lf-uts;
- $Nn_b (k^* + m)$  rows defining the temporal aggregation constraints operating on both hf- and lf-bts.

Since the first set of  $Nn_a k^*$  constraints is linearly dependent from the other rows of matrix  $\check{\mathbf{H}}$ , a full row-rank cross-temporal zero constraints kernel matrix  $\mathbf{H}'$  can be obtained by:

1. considering the  $[Nn(k^* + m) \times Nn(k^* + m)]$  commutation matrix (Magnus & Neudecker, 2019, p. 54)  $\mathbf{P}$ , such that  $\mathbf{P}[\text{vec}(\mathbf{Y})] = \text{vec}(\mathbf{Y}')$ ;
2. defining  $\mathbf{U}^* = [\mathbf{0}_{(Nn_a m \times Nn k^*)} \quad \mathbf{I}_{Nm} \otimes \mathbf{U}'] \mathbf{P}'$ ;
3. considering the  $[N(n_a m + nk^*) \times Nn(k^* + m)]$  matrix:

$$\mathbf{H}' = \begin{bmatrix} \mathbf{U}^* \\ \mathbf{I}_n \otimes \mathbf{Z}'_N \end{bmatrix}, \quad (30)$$

which has full row-rank equal to  $N(n_a m + nk^*) = n^* - Nn_a k^*$ , and allows to re-state the complete cross-temporal constraints as:

$$\mathbf{H}' \mathbf{y} = \mathbf{0}. \quad (31)$$

### 5.2. Cross-temporal point forecast reconciliation: introduction

Let us assume to have unbiased base forecasts for all the individual time series of a linearly constrained multiple time series, and for all levels of the temporal hierarchies built from the highest available sampling frequency. In addition, assume that the forecast horizon for

the most temporally aggregated time series be  $h = 1$ ,<sup>3</sup> and that the forecast horizons for the other temporally aggregated series cover the entire time cycle. This means that (i) the forecast horizon for the highest frequency time series is equal to  $m$ , and (ii) in general, the forecast horizon for a temporally aggregated time series of order  $k$  spans from 1 to  $M_k$ .

The base forecasts for each bottom time series of the system form the vectors  $\hat{\mathbf{b}}_i^{[k]}$ ,  $i = 1, \dots, n_b$ ,  $k \in \mathcal{K}$ , where  $\hat{\mathbf{b}}_i^{[1]} = \{\hat{b}_{il}^{[1]}\}_{l=1}^m$  is the  $(m \times 1)$  vector containing the base forecasts for the  $i$ th high-frequency bottom time series (hf-bts), which are the ‘very’ bottom time series in the cross-temporal framework, while the remaining  $\hat{\mathbf{b}}_i^{[k]}$ ’s (for  $k \neq 1$ ) contain the  $M_k$  forecasts for the lower-frequency ones (lf-bts). The base forecasts for the upper time series can be similarly defined as  $\hat{\mathbf{a}}_j^{[k]}$ ,  $j = 1, \dots, n_a$ ,  $k \in \mathcal{K}$ , where  $\hat{\mathbf{a}}_j^{[1]} = \{\hat{a}_{jl}^{[1]}\}_{l=1}^m$  is the  $(m \times 1)$  vector containing the base forecasts for the high-frequency  $j$ th upper time series (hf-uts), and the  $\hat{\mathbf{a}}_j^{[k]}$ ’s (for  $k \neq 1$ ) are  $(M_k \times 1)$  vectors of low-frequency upper time series (lf-uts) forecasts. Let us collect these base forecasts in the  $(n_b \times M_k)$  and  $(n_a \times M_k)$ , respectively, matrices

$$\begin{aligned} \hat{\mathbf{B}}^{[k]} &= \begin{bmatrix} \hat{\mathbf{b}}_1^{[k]} & \dots & \hat{\mathbf{b}}_{n_b}^{[k]} \end{bmatrix}, \\ \hat{\mathbf{A}}^{[k]} &= \begin{bmatrix} \hat{\mathbf{a}}_1^{[k]} & \dots & \hat{\mathbf{a}}_{n_a}^{[k]} \end{bmatrix}, \quad k \in \mathcal{K}. \end{aligned} \quad (32)$$

The matrix containing the base bts forecasts is given by  $\hat{\mathbf{B}} = \begin{bmatrix} \hat{\mathbf{B}}^{[m]} & \hat{\mathbf{B}}^{[k_{p-1}]} & \dots & \hat{\mathbf{B}}^{[k_2]} & \hat{\mathbf{B}}^{[1]} \end{bmatrix}$ , where  $\hat{\mathbf{B}}$  has dimension  $[n_b \times (k^* + m)]$ . The base uts forecasts can be similarly arranged in the  $[n_a \times (k^* + m)]$  matrix  $\hat{\mathbf{A}} = \begin{bmatrix} \hat{\mathbf{A}}^{[m]} & \hat{\mathbf{A}}^{[k_{p-1}]} & \dots & \hat{\mathbf{A}}^{[k_2]} & \hat{\mathbf{A}}^{[1]} \end{bmatrix}$ . From expression (32) we can define the  $p$  matrices  $\hat{\mathbf{Y}}^{[k]}$ , each of dimension  $(n \times M_k)$ , containing the base forecasts for the temporal aggregation

level  $k$  of both uts and bts:  $\hat{\mathbf{Y}}^{[k]} = \begin{bmatrix} \hat{\mathbf{A}}^{[k]} \\ \hat{\mathbf{B}}^{[k]} \end{bmatrix}$ ,  $k \in \mathcal{K}$ . Finally,

denoting with  $\hat{\mathbf{Y}}$  the  $[n \times (k^* + m)]$  matrix containing the base forecasts of all series and for all temporal aggregation levels, it is  $\hat{\mathbf{Y}} = \begin{bmatrix} \hat{\mathbf{Y}}^{[m]} & \hat{\mathbf{Y}}^{[k_{p-1}]} & \dots & \hat{\mathbf{Y}}^{[k_2]} & \hat{\mathbf{Y}}^{[1]} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{A}} \\ \hat{\mathbf{B}} \end{bmatrix}$ . In

general, the base forecasts fulfill neither cross-sectional (contemporaneous) nor temporal aggregation constraints (i.e., respectively,  $\mathbf{U}' \hat{\mathbf{Y}} \neq \mathbf{0}_{[n_a \times (k^* + m)]}$ , and  $\mathbf{Z}'_1 \hat{\mathbf{Y}} \neq \mathbf{0}_{(k^* \times n)}$ ). The cross-temporal point forecast reconciliation problem can thus be stated as follows: we are looking for a reconciled point forecast matrix, say  $\tilde{\mathbf{Y}}$ , which is ‘as-close-as-possible’ (according to a pre-specified metric) to the base forecast matrix  $\hat{\mathbf{Y}}$ , and simultaneously in line with the cross-sectional and temporal aggregation constraints:

$$\mathbf{U}' \tilde{\mathbf{Y}} = \mathbf{0}_{n_a \times (k^* + m)}, \quad \mathbf{Z}'_1 \tilde{\mathbf{Y}} = \mathbf{0}_{(k^* \times n)}. \quad (33)$$

As we have previously shown, the relationships (33) can be expressed in vectorized form as  $\mathbf{H}' \tilde{\mathbf{y}} = \mathbf{0}$ , where  $\tilde{\mathbf{y}} =$

<sup>3</sup> The general case  $h \geq 1$  can be dealt with straightforwardly.

$\text{vec}(\tilde{\mathbf{Y}})$  and, since  $h = 1$ , the full row-rank matrix  $\mathbf{H}'$  in (30) becomes

$$\mathbf{H}' = \begin{bmatrix} \mathbf{U}^* \\ \mathbf{I}_n \otimes \mathbf{Z}'_1 \end{bmatrix}, \quad (34)$$

with  $\mathbf{U}^* = [\mathbf{0}_{(n_a m \times n k^*)} \quad \mathbf{I}_m \otimes \mathbf{U}'] \mathbf{P}'$ ,  $\mathbf{P}$  being the  $[n(k^* + m) \times n(k^* + m)]$  commutation matrix defined above.

### 6. Cross-temporal optimal forecast combination

Let us consider the multivariate regression model

$$\hat{\mathbf{Y}} = \mathbf{Y} + \mathbf{E}, \quad (35)$$

where the involved matrices have each dimension  $[n \times (k^* + m)]$  and contain, respectively, the base ( $\tilde{\mathbf{Y}}$ ) and the target forecasts ( $\mathbf{Y}$ ), and the coherency errors ( $\mathbf{E}$ ) for the  $n$  component variables of the linearly constrained time series of interest. For each variable,  $k^* + m$  base forecasts are available, pertaining to all aggregation levels of the temporal hierarchy for a complete cycle of high-frequency observation,  $m$ . Consider now two vectorized versions of model (35), by transforming the matrices either in original form:

$$\text{vec}(\hat{\mathbf{Y}}) = \text{vec}(\mathbf{Y}) + \text{vec}(\mathbf{E}) \Leftrightarrow \hat{\mathbf{y}} = \mathbf{y} + \boldsymbol{\varepsilon}, \quad (36)$$

or in transposed form:

$$\text{vec}(\hat{\mathbf{Y}}') = \text{vec}(\mathbf{Y}') + \text{vec}(\mathbf{E}') \Leftrightarrow \hat{\mathbf{y}}' = \mathbf{y}' + \boldsymbol{\eta}. \quad (37)$$

The target forecasts must fulfill both the cross-sectional (contemporaneous) constraints  $\mathbf{U}'\mathbf{Y} = \mathbf{0}_{[n_a \times (k^* + m)]}$ , and the temporal aggregation constraints  $\mathbf{Z}'_1 \mathbf{Y}' = \mathbf{0}_{(k^* \times n)}$ . That is, in vectorized form:

$$\begin{aligned} (\mathbf{I}_{k^* + m} \otimes \mathbf{U}') \mathbf{y} &= \mathbf{0}_{[n_a(k^* + m) \times 1]} \\ \Leftrightarrow (\mathbf{U}' \otimes \mathbf{I}_{k^* + m}) \mathbf{y} &= \mathbf{0}_{[n_a(k^* + m) \times 1]} \end{aligned} \quad (38)$$

$$(\mathbf{Z}'_1 \otimes \mathbf{I}_n) \mathbf{y} = \mathbf{0}_{(k^* \times n \times 1)} \Leftrightarrow (\mathbf{I}_n \otimes \mathbf{Z}'_1) \mathbf{y} = \mathbf{0}_{(k^* \times n \times 1)} \quad (39)$$

Denote with  $\mathbf{P}$  the  $[n(k^* + m) \times n(k^* + m)]$  commutation matrix such that  $\mathbf{P}\text{vec}(\mathbf{Y}) = \text{vec}(\mathbf{Y}')$  (i.e.  $\mathbf{P}\mathbf{y} = \mathbf{y}'$ ). As a consequence, using the full row-rank matrix  $\mathbf{H}'$  defined by expression (34), the constraints (38) and (39) can be re-stated as either  $\mathbf{H}'\mathbf{y} = \mathbf{0}$ , or  $\mathbf{H}'\mathbf{P}\mathbf{y} = \mathbf{0}$ . Now, denote  $\mathbf{W} = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']$ , and  $\boldsymbol{\Omega} = E[\boldsymbol{\eta}\boldsymbol{\eta}']$ .  $\mathbf{W}$  and  $\boldsymbol{\Omega}$  are different parameterizations of the same statistical object, the covariance matrix of the random disturbances in the multivariate regression model (35), for which the following relationships hold:  $\boldsymbol{\Omega} = \mathbf{P}\mathbf{W}\mathbf{P}'$ , and  $\mathbf{W} = \mathbf{P}'\boldsymbol{\Omega}\mathbf{P}$ . Thus, to apply the general point forecast reconciliation formula (3) to a cross-temporal forecast reconciliation problem, we may consider either the expression  $\tilde{\mathbf{y}} = \hat{\mathbf{y}} - \boldsymbol{\Omega}\mathbf{H}(\mathbf{H}'\boldsymbol{\Omega}\mathbf{H})^{-1}\mathbf{H}'\hat{\mathbf{y}}$ , or equivalently re-state it as  $\tilde{\mathbf{y}} = \hat{\mathbf{y}} - \mathbf{W}\mathbf{P}'\mathbf{H}(\mathbf{H}'\mathbf{P}\mathbf{W}\mathbf{P}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{P}\hat{\mathbf{y}}$ .

#### 6.1. Simple approximations of the covariance matrix for cross-temporal point forecast reconciliation

Consider the column vectorized form of the multivariate regression (36), whose random disturbances can be

written as  $\boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1^{[m]} & \boldsymbol{\varepsilon}_1^{[k_{p-1}]} & \dots & \boldsymbol{\varepsilon}_m^{[k_{p-1}]} & \dots & \boldsymbol{\varepsilon}_1^{[1]} & \dots & \boldsymbol{\varepsilon}_m^{[1]} \end{bmatrix}'$ ,

where each  $(n \times 1)$  vector  $\boldsymbol{\varepsilon}_l^{[k]}$ ,  $k \in \mathcal{K}$ ,  $l = 1, \dots, M_k$ , contains contemporaneous random disturbances, i.e. at the same observation index for a given temporal aggregation order. A simple generalization to the cross-temporal framework of the cross-sectional approach (see Section 3.1) consists of assuming that only the disturbances at the same time index of the same temporal aggregation level are correlated. No temporal dependence (either within the same series at different times or between the  $n$  series) is admitted:

$$\begin{aligned} E \left[ \boldsymbol{\varepsilon}_r^{[k_i]} \left( \boldsymbol{\varepsilon}_s^{[k_j]} \right)' \right] &= \begin{cases} \mathbf{W}_i^{[k]} & \text{if } k_i = k_j = k, \quad r = s = l \quad k \in \mathcal{K}, \\ \mathbf{0} & \text{otherwise} \end{cases}, \quad l = 1, \dots, M_k. \end{aligned}$$

Furthermore, if it is assumed that within each temporal aggregation level the random disturbances follow a multivariate white noise, which means that the contemporaneous covariance matrices are constant in time (i.e.,  $\mathbf{W}_i^{[k]} = \mathbf{W}^{[k]}$ ,  $k \in \mathcal{K}$ ,  $l = 1, \dots, M_k$ ), the matrix  $\mathbf{W}$  has the following block-diagonal pattern:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}^{[m]} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \left( \mathbf{I}_{\frac{m}{k_{p-1}}} \otimes \mathbf{W}^{[k_{p-1}]} \right) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \left( \mathbf{I}_m \otimes \mathbf{W}^{[1]} \right) \end{bmatrix}. \quad (40)$$

From a practical point of view, each  $(n \times n)$  matrix  $\mathbf{W}^{[k]}$ ,  $k \in \mathcal{K}$ , may be approximated like in the cross-sectional forecast reconciliation case, possibly using the in-sample residuals (see Section 3.1). Expression (40) can thus be seen as a simple extension to the cross-temporal case of the approach developed in the cross-sectional framework, where no temporal dependence is accounted for both within and between the  $n$  series. We may similarly propose a simplified pattern of the disturbances covariance matrix of the multivariate regression model (35) by considering the row vectorization form (37). In this case, the random disturbances vector  $\boldsymbol{\eta}$  can be written as  $\boldsymbol{\eta} = [\boldsymbol{\eta}'_1 \dots \boldsymbol{\eta}'_i \dots \boldsymbol{\eta}'_n]'$ , where each  $[(k^* + m) \times 1]$  vector  $\boldsymbol{\eta}_i$ ,  $i = 1, \dots, n$ , contains the random disturbances at different observation indices of the various temporal aggregation levels for the same series  $i$ . If we assume that the  $n$  series are uncorrelated at any observation index for any temporal aggregation level (i.e., neither contemporaneous nor temporal correlation is admitted between the series) denoting with  $\boldsymbol{\Omega}_{ii} = E[\boldsymbol{\eta}_i \boldsymbol{\eta}'_i]$ ,  $i = 1, \dots, n$ , the  $[(k^* + m) \times (k^* + m)]$  covariance matrix of the coherency errors of the temporal hierarchies of series  $i$ , the complete matrix  $\boldsymbol{\Omega}$  can be written as follows:

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \boldsymbol{\Omega}_{nn} \end{bmatrix}, \quad (41)$$

where each matrix  $\Omega_{ii}$ ,  $i = 1, \dots, n$ , may be approximated as in the temporal forecast reconciliation case, possibly using the in-sample residuals (see Section 4.1). Thus expression (41) can be seen as a simple extension to the cross-temporal case of the approach developed in the temporal hierarchies framework, where no correlation is admitted between the random errors of the  $n$  series. The covariance patterns (40) and (41) (i) are placed at opposite ends of possible ways of dealing with cross-temporal variables, and (ii) should be considered as the first practical devices to make the optimal combination forecast approach feasible for the cross-temporal framework as well.

Residual-based estimates of the covariance matrix  $\mathbf{W}$  (and its re-parameterized counterpart  $\Omega$ ) use the in-sample residuals of the models used to forecast the  $n$  time series considered at any temporal aggregation level. Denote by  $\hat{\mathbf{E}}_l^{[k]}$ ,  $k \in \mathcal{K}$ ,  $l = 1, \dots, M_k$ , the  $(n \times N)$  matrix containing the in-sample residuals for a single node of the cross-temporal hierarchy (i.e., the  $i$ th row contains the residuals for the  $N$  sub-periods  $l$  of the model used to forecast the temporal aggregate of order  $k$  of series  $i$ ). For each temporal aggregation level  $k \in \mathcal{K}$ , the  $M_k$  matrices  $\hat{\mathbf{E}}_l^{[k]}$  can be grouped into the  $(n \times NM_k)$  matrix  $\hat{\mathbf{E}}^{[k]} = [\hat{\mathbf{E}}_1^{[k]} \dots \hat{\mathbf{E}}_l^{[k]} \dots \hat{\mathbf{E}}_{M_k}^{[k]}]$ ,  $k \in \mathcal{K}$ . The  $(n(k^* + m) \times N)$  matrix containing all the residuals at any time observation index and any temporal aggregation level can, in turn, be written as:

$$\hat{\mathbf{E}} = \begin{bmatrix} \hat{\mathbf{E}}_1^{[m]} \\ \hat{\mathbf{E}}_1^{[k_{p-1}]} \\ \vdots \\ \hat{\mathbf{E}}_1^{[k_{p-1}]} \\ \vdots \\ \hat{\mathbf{E}}_1^{[1]} \\ \vdots \\ \hat{\mathbf{E}}_m^{[1]} \end{bmatrix} = [\hat{\mathbf{e}}_1 \dots \hat{\mathbf{e}}_\tau \dots \hat{\mathbf{e}}_N],$$

where each  $[n(k^* + m) \times 1]$  vector  $\hat{\mathbf{e}}_\tau$ ,  $\tau = 1, \dots, N$ ,

$$\text{is given by } \hat{\mathbf{e}}_\tau = \begin{bmatrix} \hat{e}_{1,\tau}^{[m]} \dots (\hat{e}_{1,\tau}^{[1]})' \dots \hat{e}_{n,\tau}^{[m]} \dots (\hat{e}_{n,\tau}^{[1]})' \end{bmatrix}.$$

The sample residual covariance matrix can be calculated according to both parameterization as either  $\hat{\Omega}_{\text{sam}} = \frac{1}{N} \sum_{\tau=1}^N \hat{\mathbf{e}}_\tau (\hat{\mathbf{e}}_\tau)' = \frac{1}{N} \hat{\mathbf{E}} \hat{\mathbf{E}}'$ , or  $\hat{\mathbf{W}}_{\text{sam}} = \mathbf{P}' \hat{\Omega}_{\text{sam}} \mathbf{P}$ . However, in many practical situations matrix  $\hat{\mathbf{E}}$  has a number of rows - which is equal to the number of nodes in the cross-temporal hierarchy - much larger than the number of columns, which is equal to  $N = \frac{T}{m}$ . Thus matrices  $\hat{\Omega}_{\text{sam}}$  and  $\hat{\mathbf{W}}_{\text{sam}}$  might not have good properties (in particular, they are not p.d. if  $N \leq n(k^* + m)$ ), and simplified

approximations must be found. Two feasible alternatives are given by either the diagonalization or the shrinkage of matrix  $\hat{\mathbf{W}}_{\text{sam}}$ , that is, respectively,  $\hat{\mathbf{W}}_{\text{wlsh}} = \mathbf{I}_{n(k^*+m)} \odot \hat{\mathbf{W}}_{\text{sam}}$ , and  $\hat{\mathbf{W}}_{\text{shr}} = \hat{\lambda} \hat{\mathbf{W}}_{\text{wlsh}} + (1 - \hat{\lambda}) \hat{\mathbf{W}}_{\text{sam}}$ , where  $\hat{\mathbf{W}}_{\text{wlsh}}$  is a diagonal matrix containing the estimates of the ‘hierarchy variances’ for each node of the cross-temporal hierarchy,  $\hat{\mathbf{W}}_{\text{shr}}$  is the matrix obtained by shrinkage of  $\hat{\mathbf{W}}_{\text{sam}}$  with target  $\hat{\mathbf{W}}_{\text{wlsh}}$ , and  $\hat{\lambda}$  is an estimate of the coefficient of shrinkage intensity  $\lambda$ ,  $0 \leq \lambda \leq 1$ . Both  $\hat{\mathbf{W}}_{\text{sam}}$  and  $\hat{\mathbf{W}}_{\text{shr}}$  refer to all the  $n(k^* + m)$  hierarchy nodes simultaneously taken, but unlike the former matrix, the latter should not suffer from possible singularity problems. An alternative choice is the block-diagonal matrix (40). Following Wickramasuriya et al. (2019), full and shrunk estimates of matrices  $\mathbf{W}^{[k]}$ ,  $k \in \mathcal{K}$ , forming the blocks on the diagonal of this matrix, may be computed as:

$$\hat{\mathbf{W}}_{\text{sam}}^{[k]} = \frac{1}{NM_k} \hat{\mathbf{E}}^{[k]} (\hat{\mathbf{E}}^{[k]})', \quad k \in \mathcal{K}, \tag{42}$$

$$\hat{\mathbf{W}}_{\text{shr}}^{[k]} = \hat{\lambda}_k (\mathbf{I}_n \odot \hat{\mathbf{W}}_{\text{sam}}^{[k]}) + (1 - \hat{\lambda}_k) \hat{\mathbf{W}}_{\text{sam}}^{[k]}, \quad k \in \mathcal{K}, \tag{43}$$

and used to approximate  $\mathbf{W}$  as follows (the superscript ‘BD’ stands for ‘Block-Diagonal’):

$$\hat{\mathbf{W}}_{\text{sam}}^{BD} = \begin{bmatrix} \hat{\mathbf{W}}_{\text{sam}}^{[m]} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{I}_m \otimes \hat{\mathbf{W}}_{\text{sam}}^{[1]} \end{bmatrix},$$

$$\hat{\mathbf{W}}_{\text{shr}}^{BD} = \begin{bmatrix} \hat{\mathbf{W}}_{\text{shr}}^{[m]} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{I}_m \otimes \hat{\mathbf{W}}_{\text{shr}}^{[1]} \end{bmatrix}.$$

Most of the choices for  $\mathbf{W}$  (or  $\Omega$ ) shown so far are simple extensions to the cross-temporal framework of the approximations for  $\mathbf{W}$  (or  $\Omega$ ) considered either in cross-sectional or in temporal forecast reconciliation. We consider the following approximations (‘oct’ stands for ‘optimal cross-temporal’):

- identity (oct-ols):  $\mathbf{W} = \Omega = \mathbf{I}_{n(k^*+m)}$
- hierarchy variance scaling (oct-wlsh):  $\mathbf{W} = \hat{\mathbf{W}}_{\text{wlsh}}$
- series variance scaling (oct-wlsv):  $\mathbf{W} = \hat{\mathbf{W}}_{\text{wlsv}} = \mathbf{P}' \hat{\Omega}_{\text{wlsv}} \mathbf{P}$ , where  $\hat{\Omega}_{\text{wlsv}}$  is a straightforward extension of  $\hat{\Omega}_{\text{t-wlsv}}$  (see Section 4.1)
- block-diagonal shrunk cross-covariance scaling (oct-bdshr):  $\mathbf{W} = \hat{\mathbf{W}}_{\text{shr}}^{BD}$
- block-diagonal cross-covariance scaling (oct-bdsam):  $\mathbf{W} = \hat{\mathbf{W}}_{\text{sam}}^{BD}$
- auto-covariance scaling (acov):  $\mathbf{W} = \hat{\mathbf{W}}_{\text{acov}} = \mathbf{P}' \hat{\Omega}_{\text{acov}} \mathbf{P}$ , where  $\hat{\Omega}_{\text{acov}}$  is a straightforward extension of  $\hat{\Omega}_{\text{t-acov}}$  (see Section 4.1)
- MinT-shr (oct-shr):  $\mathbf{W} = \hat{\mathbf{W}}_{\text{shr}}$
- MinT-sam (oct-sam):  $\mathbf{W} = \hat{\mathbf{W}}_{\text{sam}}$

## 7. Heuristic cross-temporal reconciliation procedures

### 7.1. The KA procedure

Kourentzes and Athanasopoulos (2019), henceforth KA, have proposed a cross-temporal reconciliation method that can be viewed as an ensemble forecasting procedure that exploits the simple averaging of different forecasts. The procedure consists of the following steps (it is assumed  $h = 1$ ):

#### Step 1

For each individual variable, compute the temporally reconciled forecasts and collect them in the  $[n \times (k^* + m)]$  matrix  $\check{Y}$ . This result can be obtained by applying the point forecast reconciliation formula (22) to each column of matrix  $\hat{Y} = \begin{bmatrix} \hat{a}_1 & \dots & \hat{a}_{n_a} & \hat{b}_1 & \dots & \hat{b}_{n_b} \\ \hat{a}_1^{[1]} & \dots & \hat{a}_{n_a}^{[1]} & \hat{b}_1^{[1]} & \dots & \hat{b}_{n_b}^{[1]} \end{bmatrix}$ . These reconciled forecasts are in line with the temporal aggregation constraints, i.e.  $Z_1' \check{Y} = \mathbf{0}_{(k^* \times n)}$ , but in general they are not in line with the cross-sectional constraints, that is:  $U' \check{Y} \neq \mathbf{0}_{[n_a \times (k^* + m)]}$ .

#### Step 2

Transform  $\check{Y}$  by computing time-by-time cross-sectional reconciled forecasts for all the temporal aggregation levels, and collect them in the  $[n \times (k^* + m)]$  matrix  $\check{Y} = [\check{Y}^{[m]} \check{Y}^{[k_{p-1}]} \dots \check{Y}^{[k_2]} \check{Y}^{[1]}]$ , where  $\check{Y}^{[k]}$ ,  $k \in \mathcal{K}$ , has dimension  $(n \times M_k)$ . Thus, the cross-sectionally reconciled forecasts can be computed by transforming each  $\check{Y}^{[k]}$  as  $\check{Y}^{[k]} = \mathbf{M}^{[k]} \check{Y}^{[k]}$ ,  $k \in \mathcal{K}$ , where  $\mathbf{M}^{[k]} = \mathbf{I}_n - \mathbf{W}^{[k]} \mathbf{U} (\mathbf{U}' \mathbf{W}^{[k]} \mathbf{U})^{-1} \mathbf{U}'$  denotes the  $(n \times n)$  projection matrix used to reconcile forecasts of  $k$ -level temporally aggregated time series, and  $\mathbf{W}^{[k]}$  is a  $(n \times n)$  known p.d. matrix. Since it is  $\mathbf{U}' \mathbf{M}^{[k]} = \mathbf{0}_{(n_a \times n)}$ ,  $k \in \mathcal{K}$ , the reconciled forecasts are cross-sectionally coherent, i.e.  $\mathbf{U}' \check{Y} = \mathbf{0}_{[n_a \times (k^* + m)]}$ , but not temporally:  $Z_1' \check{Y} \neq \mathbf{0}_{(k^* \times n)}$ .

#### Step 3

Transform again the step 1 forecasts  $\check{Y}$ , by computing time-by-time cross-sectional reconciled forecasts for all the temporal aggregation levels using the  $(n \times n)$  matrix  $\bar{\mathbf{M}}$ , given by the average of the matrices  $\mathbf{M}^{[k]}$  obtained at step 2. Matrix  $\bar{\mathbf{M}}$  can be expressed as  $\bar{\mathbf{M}} = \frac{1}{p} \sum_{k \in \mathcal{K}} \mathbf{M}^{[k]}$ , and the final cross-temporal reconciled forecasts are given by:

$$\check{Y}^{KA} = \bar{\mathbf{M}} \check{Y}. \tag{44}$$

Since  $\mathbf{U}' \bar{\mathbf{M}} = \frac{1}{p} \sum_{k \in \mathcal{K}} \mathbf{U}' \mathbf{M}^{[k]} = \mathbf{0}_{(n_a \times n)}$ , and  $Z_1' \check{Y} = \mathbf{0}_{(k^* \times n)}$ , the reconciled forecasts (44) fulfill both cross-sectional and temporal aggregation constraints:

$$\mathbf{U}' \check{Y}^{KA} = \mathbf{U}' \bar{\mathbf{M}} \check{Y} = \mathbf{0}_{[n_a \times (k^* + m)]},$$

$$Z_1' (\check{Y}^{KA})' = Z_1' \check{Y}' \bar{\mathbf{M}}' = \mathbf{0}_{(k^* \times n)}.$$

### 7.2. Some remarks

To perform step 1, KA consider either t-struc or t-wlsv (see Section 4.1) as forecast reconciliation procedures through temporal hierarchies, while in step 2 either cs-wls or cs-shr are used (see Section 3.1).

**Remark 1.** These two steps can be seen as the successive applications of two distinct multivariate reconciliation procedures: in the first step it is solved a linearly constrained quadratic minimization problem, where only temporal aggregation constraints are considered:

$$\check{y} = \underset{y}{\operatorname{argmin}} (y - \hat{y})' \Omega^{-1} (y - \hat{y}), \quad \text{s.t. } (\mathbf{I}_n \otimes Z_1') y = \mathbf{0},$$

where  $\Omega$  is the block-diagonal matrix in (41). The second step consists of another quadratic minimization problem, where only cross-sectional (contemporaneous) constraints are considered:

$$\check{y} = \underset{y}{\operatorname{argmin}} (y - \check{y})' \mathbf{W}^{-1} (y - \check{y}), \quad \text{s.t. } (\mathbf{I}_{k^* + m} \otimes \mathbf{U}') y = \mathbf{0},$$

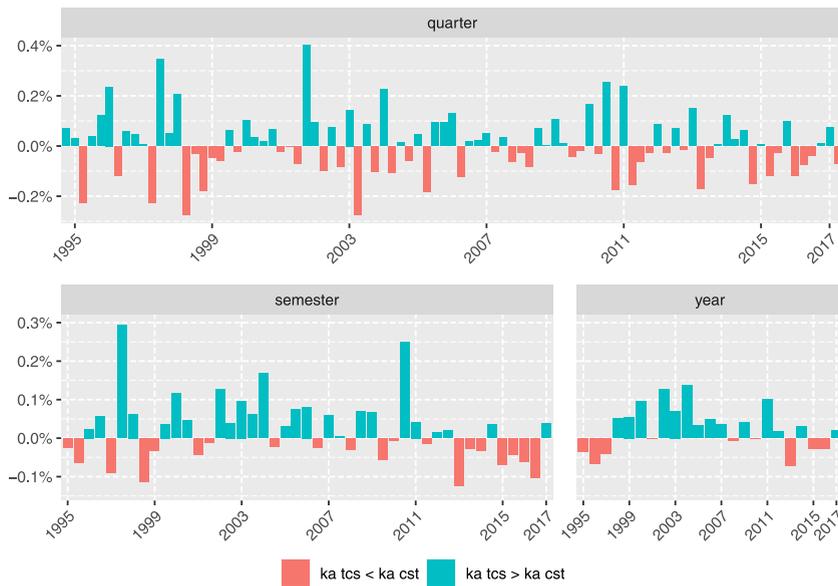
where  $\mathbf{W}$  is the block-diagonal matrix in (40). It is worth noting that one may switch between the two data representations using the permutation matrix  $\mathbf{P}$ , that is  $\check{y} = \mathbf{P}' \check{y}$ .

**Remark 2.** In general, the final result of the reconciliation procedure would change if the user inverts the order of application of the two reconciliation steps. Since the differences between the reconciled point forecasts according to these two approaches could be not negligible (see Section 7.3), this seems to be a weakness of the procedure, and calls for a decision rule about the final reconciled forecasts to retain. A practical way of doing this could be choosing the reconciled forecasts, which are the ‘closest’ (according to a given metric) to the base forecasts.

**Remark 3.** The calculation of the average matrix  $\bar{\mathbf{M}}$  in the final step of the procedure, needed to recover the cross-temporal coherency across the point forecasts, requires the availability of the projection matrices used in the second step. This poses no problem when closed form reconciliation formulae can be used. Unfortunately, this is not the case when non-negativity of the final reconciled estimates is wished, calling for appropriate numerical procedures (Wickramasuriya et al., 2020; Kourentzes & Athanasopoulos, 2021).

### 7.3. An iterative heuristic cross-temporal reconciliation procedure

Taking inspiration from the KA reconciliation procedure, we consider an iterative procedure that produces cross-temporal reconciled forecasts by alternating forecast reconciliation along one dimension (cross-sectional or temporal). Each iteration consists of the first two steps of the KA procedure, so the forecasts are reconciled by alternating reconciliation through temporal hierarchies and cross-sectional reconciliation in a cyclic fashion. Starting from the base forecasts  $\hat{Y}$ , denote with  $d_{cs}$  and  $d_{te}$ , respectively, the cross-sectional and the temporal gross



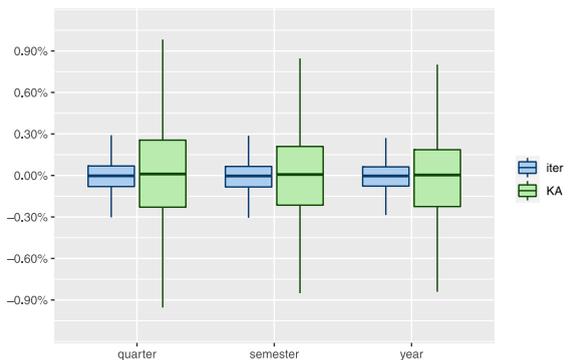
**Fig. 2.** Quarterly, semi-annual and annual Australian GDP one-step-ahead reconciled forecasts according to the Kourentzes and Athanasopoulos (2019) cross-temporal reconciliation approach (t-wlsv for the temporal step, cs-shr for the cross-sectional step) by alternating the constraint dimensions to be fulfilled: percentage differences between the reconciled forecasts obtained through (i) temporal-then-cross-sectional reconciliation, and (ii) cross-sectional-then-temporal reconciliation. The differences between the two reconciled forecasts are divided by their arithmetic mean.

discrepancies, given by  $d_{cs} = \|\mathbf{U}\hat{\mathbf{Y}}\|_1$ , and  $d_{te} = \|\mathbf{Z}'_1\tilde{\mathbf{Y}}\|_1$ , where  $\|\mathbf{X}\|_1 = \sum_{i,j} |x_{i,j}|$ . Since the base forecasts are not in line with either type of constraints, in general both  $d_{cs}$  and  $d_{te}$  are greater than zero. The iterative procedure can be described as follows:

1. Start the iterations by calculating the temporally reconciled forecasts  $\tilde{\mathbf{Y}}^{(1)}$ , such that  $\mathbf{Z}'_1(\tilde{\mathbf{Y}}^{(1)})' = \mathbf{0}$ , and  $d_{cs}^{(1)} = \|\mathbf{U}\tilde{\mathbf{Y}}^{(1)}\|_1 \geq 0$ .
2. The point forecasts in matrix  $\tilde{\mathbf{Y}}^{(1)}$  are then cross-sectionally reconciled, obtaining  $\tilde{\mathbf{Y}}^{(2)}$ , which is such that  $\mathbf{U}\tilde{\mathbf{Y}}^{(2)} = \mathbf{0}$ , and  $d_{te}^{(1)} = \|\mathbf{Z}'_1(\tilde{\mathbf{Y}}^{(2)})'\|_1 \geq 0$ .
3. The updates in steps 1. and 2. are performed at each iteration  $j$ ,  $j = 1, 2, \dots$ , until a convergence criterion is met, that is  $d_{te}^{(j)} < \delta$ , where  $\delta$  is a positive tolerance value (e.g.,  $\delta = 10^{-6}$ ), and matrix  $\tilde{\mathbf{Y}}^{(2j)}$  contains the final cross-temporal reconciled forecasts.

The above procedure can be seen as an extension of the well known iterative proportional fitting procedure (Deming & Stephan, 1940), also known as the RAS method (Miller & Blair, 2009), to adjust the internal cell values of a two-dimensional matrix iteratively until they sum to some predetermined row and column totals. In that case, the adjustment follows a proportional adjustment scheme. In contrast, each adjustment step is made in the cross-temporal reconciliation framework according to the penalty function associated with the single-dimension reconciliation procedure adopted.

Indeed, the choice of dimension along which the first reconciliation step in each iteration is performed is up



**Fig. 3.** Quarterly, semi-annual and annual one-step-ahead reconciled forecasts of 79 out of 95 time series of the Australian GDP from Income and Expenditure sides using both the original KA cross-temporal reconciliation procedure (t-wlsv for the temporal step, and cs-shr for the cross-sectional one), and its iterative variant: boxplots of the percentage differences between the reconciled forecasts obtained through (i) temporal-then-cross-sectional reconciliation, and (ii) cross-sectional-then-temporal reconciliation. The differences between each pair of reconciled forecasts are divided by their arithmetic mean.

to the user. There is no particular reason to perform the temporal reconciliation first and the cross-sectional reconciliation second. Fig. 2 shows the percentage discrepancies in the Australian GDP at current prices one-step-ahead forecasts for any temporal aggregation level (quarterly, semi-annual, annual, see Section 8), when the cross-temporal reconciliation is performed according to either the KA approach, or to the analogous procedure where the cross-sectional constraints are considered first, and then the temporal dimension is accounted for (see

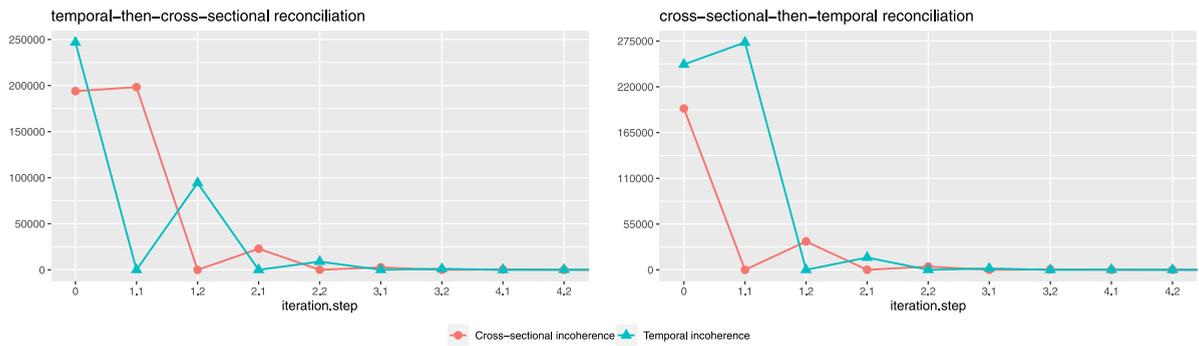


Fig. 4. Incoherence at each iteration step of the iterative cross-temporal forecast reconciliation procedure (t-wlsv + cs-shr) for the Australian GDP time series, at the first forecast origin 1994:Q3.

section A.7 in the online appendix). Percentage differences in the reconciled forecasts for this single, important variable, are visually evident, though bounded within (−0.3% – +0.4%).

Fig. 3 completes the results shown so far by considering the forecasts of the strictly positive 79 (out of 95) variables from both Income and Expenditure sides, cross-temporally reconciled according to the KA procedure and its iterative variant. The boxplots show the distributions of the percentage discrepancies between the reconciled forecasts obtained using temporal reconciliation first, and cross-sectional reconciliation then, vis-à-vis the results obtained by inverting the order of application of the two reconciliation procedures. It appears that the iterative variant of the original KA proposal produces less pronounced discrepancies.

It must be said that the convergence speed of the iterative procedure does not seem to be affected by the choice of the first dimension to be fulfilled when the iteration starts. Fig. 4 shows an example of the convergence speed of the iterative procedure, either starting with the cross-sectional (right panel) or temporal (left panel) reconciliation procedure for the Australian GDP forecasts. In both cases, the convergence is achieved very quickly: fixing  $\delta = 10^{-6}$ , 15 (14) iterates are needed when starting from the temporal (cross-sectional) dimension. Furthermore, from the fourth iteration onwards, the constraints are practically fulfilled in both cases. Nevertheless, since the final reconciled values depend on this choice, it would be useful having an ex-ante ‘choice rule’ between the two alternatives. For ease of presentation, in the following, we maintain the original choice made by KA, performing temporal forecast reconciliation first and cross-sectional reconciliation second.

### 8. Cross-temporal reconciliation of the Australian GDP forecasts from Income and Expenditure sides

In a recent paper, Athanasopoulos et al. (2019) propose “the application of state-of-the-art forecast reconciliation methods to macroeconomic forecasting” to perform aligned decision making and to improve forecast accuracy. In their empirical study, they consider the cross-sectional forecast reconciliation for 95 Australian Quarterly National Accounts time series, describing the Gross Domestic

Product (GDP) at current prices from Income and Expenditure sides, interpreted as two distinct hierarchical structures. In the former case (Income), GDP is on the top of 15 lower level aggregates, while in the latter (Expenditure), GDP is the top level aggregate of a hierarchy of 79 time series (for details, refer to Athanasopoulos et al., 2019, pp. 702–705, and Figures 21.4–21.7 therein). By managing the complete set of 95 time series following the approach described in Section 3, Bisaglia et al. (2020) have extended the results of Athanasopoulos et al. (2019), showing that fully cross-sectional reconciled quarterly GDP forecasts, coherent with all the reconciled forecasts from both Expenditure and Income sides can be obtained through the general reconciliation approach described in Section 2. According to the notation adopted so far, the  $(33 \times 95)$  kernel matrix accounting for the cross-sectional zero constraints is given by:

$$U' = \begin{bmatrix} 1 & 0'_{(5 \times 1)} & -1'_{(10 \times 1)} & 0'_{(26 \times 1)} & 0'_{(53 \times 1)} \\ 1 & 0'_{(5 \times 1)} & 0'_{(10 \times 1)} & 0'_{(26 \times 1)} & -1'_{(53 \times 1)} \\ 0_{(5 \times 1)} & I_5 & -C^I & 0_{(5 \times 26)} & 0_{(5 \times 53)} \\ 0_{(26 \times 1)} & 0_{(26 \times 5)} & 0_{(26 \times 10)} & I_{26} & -C^E \end{bmatrix},$$

where  $C^I$  and  $C^E$  are the contemporaneous aggregation matrix mapping the bts into the uts for the Income and the Expenditure sides, respectively (Bisaglia et al., 2020). By exploiting this result, cross-temporal forecast reconciliation is now applied within the same forecasting experiment designed by Athanasopoulos et al. (2019)<sup>4</sup> extended to consider semi-annual and annual forecasts as well. For the available time series span (1984:Q4 – 2018:Q1), quarterly base forecasts from 1 up to 4 quarters ahead have been obtained for the  $n = 95$  separate time series through simple univariate ARIMA models selected using the auto.arima function of the R-package forecast (Hyndman et al., 2020). The forecasting experiment uses a recursive training sample with expanding window length. The first training sample is set from 1984:Q4 to 1994:Q3, and the last ends on 2017:Q1, for a total of 91 forecast origins. Similarly, in the same automatic fashion, we have computed (i) one- and two-step-ahead forecasts

<sup>4</sup> We did not change this first, crucial step in the forecast reconciliation workflow, since the focus is on the potential of cross-temporal forecast reconciliation.

for the time series obtained by temporal aggregation of two successive quarters, and (ii) one-step-ahead forecasts for the time series obtained by temporal aggregation of four successive quarters. All the base forecasts have been reconciled using the R-package FoReco (Girolimetto & Di Fonzo, 2021).

### 8.1. Performance measures for multiple comparisons

We evaluate the performance of multiple (say,  $J > 1$ ) forecast reconciliation procedures through forecast accuracy indices calculated on the forecast error

$$\hat{e}_{i,j,t}^{[k],h} = y_{i,t+h}^{[k]} - \hat{y}_{i,j,t}^{[k],h},$$

$$i = 1, \dots, 95, \quad t = 1, \dots, 91, \quad k \in \mathcal{K},$$

$$j = 0, \dots, J, \quad h = 1, \dots, h_k,$$

where  $y$  and  $\hat{y}$  are the observed and forecasted values, respectively,  $i$  denotes the series ( $i = 1, \dots, 32$ , for the uts,  $i = 33, \dots, 95$ , for the bts),  $j = 0$  denotes the base forecasts,  $t$  is the forecast origin ( $t = 1$  corresponds to 1994:Q3),  $\mathcal{K} = \{4, 2, 1\}$ , and  $h_4 = 1$ ,  $h_2 = 2$ ,  $h_1 = 4$ , are the forecast horizons for annual, semi-annual, and quarterly time series, respectively. The accuracy is evaluated using the Average Relative Mean Square Error (AvgRelMSE, Davydenko & Fildes, 2013; Kourentzes & Athanasopoulos, 2019, 2020), obtained by transforming the MSE index, given by the average across all 91 forecast origins of the squared forecast errors:

$$\text{MSE}_{i,j}^{[k],h} = \frac{1}{91} \sum_{t=1}^{91} \left( \hat{e}_{i,j,t}^{[k],h} \right)^2, \tag{45}$$

$$i = 1, \dots, 95, \quad k \in \mathcal{K},$$

$$j = 0, \dots, J, \quad h = 1, \dots, h_k.$$

The AvgRelMSE is the geometric mean across all 95 series of the MSE ratio<sup>5</sup> of a forecast over a benchmark given by the base, incoherent ARIMA forecasts, across all evaluation samples, for a given horizon  $h$ :

$$\text{AvgRelMSE}_j^{[k],h} = \left( \prod_{i=1}^{95} \text{rMSE}_{i,j}^{[k],h} \right)^{\frac{1}{95}}, \tag{46}$$

$$j = 0, \dots, J, \quad k \in \mathcal{K},$$

$$h = 1, \dots, h_k,$$

where  $\text{rMSE}_{i,j}^{[k],h} = \frac{\text{MSE}_{i,j}^{[k],h}}{\text{MSE}_{i,0}^{[k],h}}$  is the relative MSE. If a forecast outperforms the base forecasts, then the AvgRelMSE becomes smaller than one and vice-versa, and the percentage improvement in accuracy over the benchmark can be calculated as  $\left( 1 - \text{AvgRelMSE}_j^{[k],h} \right) \times 100$ . Expression (46), which refers to all 95 time series, can be re-stated for (i) groups of variables (e.g., bts and uts), (ii) multiple forecast horizons (e.g.,  $h = 1 - 4$  for quarterly

forecasts,  $k = 1$ ;  $h = 1 - 2$  for semi-annual forecasts,  $k = 2$ ), (iii) different temporal aggregation levels over the whole forecast horizon (e.g., accuracy indices for the whole temporal hierarchy of each series). To give a complete picture of the evaluation results, in the following subsection, we show and discuss the MSE-based accuracy indices, at multiple timescales and forecast horizons, for a set of selected forecast reconciliation procedures. Furthermore, we use the non-parametric Friedman and the post hoc ‘Multiple Comparison with the Best’ (MCB) Nemenyi tests (Koning et al., 2005; Kourentzes & Athanasopoulos, 2019; Makridakis et al., 2020) to establish if the forecasting performances of the considered techniques are significantly different.

### 8.2. The considered forecast reconciliation procedures

The empirical application mainly aims to evaluate the performance of the most convincing new cross-temporal reconciliation procedures, which are those using residual-based approximations of the covariance matrix, compared to the state-of-the-art point forecast reconciliation procedures. At this end, we consider five selected procedures recently proposed in the hierarchical forecasting literature, five (two-step and iterative) variants of the KA approach, and three optimal combination forecast procedures:

- cs-shr (Wickramasuriya et al., 2019),
- t-wlsv (Athanasopoulos et al., 2017),
- t-acov (Nystrup et al., 2020),
- t-sar1 (Nystrup et al., 2020),
- kah-wlsv-shr (Kourentzes & Athanasopoulos, 2019),
- tcs-acov-shr, i.e. two-step t-acov + cs-shr,
- tcs-sar1-shr, i.e. two-step t-sar1 + cs-shr,
- ite-wlsv-shr, i.e. iterative t-wlsv + cs-shr,
- ite-acov-shr, i.e. iterative t-acov + cs-shr,
- ite-sar1-shr, i.e. iterative t-sar1 + cs-shr,
- oct-wlsv, i.e.  $\mathbf{W} = \widehat{\mathbf{W}}_{BD}^{wlsv}$ ,
- oct-bdshr, i.e.  $\mathbf{W} = \widehat{\mathbf{W}}_{shr}$ ,
- oct-acov, i.e.  $\mathbf{W} = \widehat{\mathbf{W}}_{acov}$ .

The first five procedures have proven well-performing in several empirical applications (Athanasopoulos et al., 2019, 2017; Wickramasuriya et al., 2019; Bisaglia et al., 2020; Nystrup et al., 2020, among others). The one-dimension reconciliation procedures (cs-shr, t-wlsv, t-acov, and t-sar1) do not give fully coherent forecasts. Rather, as far as they are expected to improve the base forecasts, the best-practice one-dimension procedures should be viewed as stricter benchmarks for the cross-temporal forecast reconciliation techniques. These are requested to give accurate one-number forecasts as well. The forecasting experiment was designed to evaluate the capability of the cross-temporal forecast reconciliation procedures to improve the forecast accuracy as compared to (i) the base forecasts and (ii) the most performing one-dimension forecast reconciliation methods. In addition, the experiment should help in assessing (iii) the performance of both KA-variants (two-step and iterative procedures) and optimal combination forecasts as compared to the original proposal by KA, and (iv) the feasibility and the accuracy of the optimal combination

<sup>5</sup> Davydenko and Fildes (2013) develop the Average Relative MAE, based on the Mean Absolute Error, and suggest that this formulation ‘can also be extended to other measures of dispersion or loss functions’, as the AvgRelMSE in (46).

**Table 1**  
AvgRelMSE at any temporal aggregation level and any forecast horizon.

Reconciliation procedure	Quarterly					Semi-annual			Annual	All
	1	2	3	4	1–4	1	2	1–2	1	
<i>all 95 series</i>										
base	1	1	1	1	1	1	1	1	1	1
cs-shr	0.958	0.970	0.976	0.982	0.972	0.953	0.978	0.965	0.966	0.969
t-wlsv	<b>1.002</b>	0.999	0.988	0.985	0.993	0.844	0.932	0.887	0.773	0.928
t-acov	0.978	0.991	0.999	0.989	0.989	0.825	0.935	0.879	0.769	0.923
t-sar1	<b>1.002</b>	0.999	0.988	0.985	0.994	0.845	0.932	0.887	0.773	0.928
kah-wlsv-shr	0.968	0.970	0.960	<b>0.960</b>	0.965	0.818	<b>0.909</b>	0.862	0.752	0.901
tcs-acov-shr	0.945	<b>0.958</b>	0.971	0.963	0.959	0.798	0.912	0.853	0.748	0.895
tcs-sar1-shr	0.968	0.970	0.960	0.960	0.965	0.818	0.909	0.862	0.752	0.901
ite-wlsv-shr	0.961	0.968	<b>0.959</b>	0.960	0.962	0.815	0.909	0.861	0.751	0.900
ite-acov-shr	<b>0.940</b>	0.958	0.971	0.965	<b>0.959</b>	<b>0.796</b>	0.913	<b>0.852</b>	<b>0.748</b>	<b>0.895</b>
ite-sar1-shr	0.961	0.968	0.959	0.961	0.962	0.815	0.909	0.861	0.751	0.900
oct-wlsv	0.969	0.972	0.962	0.963	0.967	0.820	0.913	0.865	0.756	0.904
oct-bdshr	0.984	0.980	0.962	0.967	0.973	0.830	0.914	0.871	0.757	0.910
oct-acov	0.955	0.965	0.977	0.971	0.967	0.801	0.919	0.858	0.753	0.902
<i>32 upper series</i>										
base	1	1	1	1	1	1	1	1	1	1
cs-shr	<b>0.916</b>	<b>0.927</b>	0.930	0.932	<b>0.926</b>	0.917	0.939	0.928	0.923	0.926
t-wlsv	<b>1.006</b>	<b>1.009</b>	0.991	0.992	1.000	0.856	0.939	0.896	0.768	0.933
t-acov	<b>1.002</b>	<b>1.015</b>	0.992	0.993	<b>1.000</b>	0.854	0.938	0.895	0.768	0.9332
t-sar1	<b>1.007</b>	<b>1.009</b>	0.991	0.992	1.000	0.856	0.939	0.896	0.768	0.933
kah-wlsv-shr	0.940	0.947	0.928	0.930	0.936	0.800	0.877	0.837	0.715	0.873
tcs-acov-shr	0.941	0.944	0.931	0.933	0.937	0.796	0.878	0.836	0.715	0.873
tcs-sar1-shr	0.940	0.946	0.928	0.930	0.936	0.800	0.877	0.837	0.715	0.873
ite-wlsv-shr	0.925	0.942	0.922	0.927	0.929	0.793	0.874	0.833	0.711	<b>0.867</b>
ite-acov-shr	0.928	0.940	0.926	0.931	0.931	<b>0.789</b>	0.875	<b>0.831</b>	<b>0.711</b>	0.868
ite-sar1-shr	0.926	0.942	<b>0.922</b>	<b>0.927</b>	0.929	0.794	<b>0.874</b>	0.833	0.711	0.867
oct-wlsv	0.941	0.951	0.932	0.933	0.939	0.803	0.881	0.841	0.720	0.876
oct-bdshr	0.945	0.956	0.925	0.934	0.940	0.809	0.879	0.843	0.717	0.877
oct-acov	0.939	0.950	0.935	0.937	0.940	0.798	0.884	0.840	0.719	0.876
<i>63 bottom series</i>										
base	1	1	1	1	1	1	1	1	1	1
cs-shr	0.981	0.993	1.000	<b>1.009</b>	0.996	0.971	0.999	0.985	0.988	0.991
t-wlsv	0.999	0.995	0.986	0.982	0.990	0.839	0.928	0.882	0.775	0.925
t-acov	0.966	0.980	<b>1.002</b>	0.986	0.983	0.811	0.934	0.870	0.770	0.917
t-sar1	0.999	0.994	0.986	0.982	0.990	0.839	0.928	0.882	0.775	0.925
kah-wlsv-shr	0.983	0.982	<b>0.976</b>	<b>0.976</b>	0.979	0.827	<b>0.925</b>	0.875	0.771	0.916
tcs-acov-shr	0.947	<b>0.966</b>	0.992	0.978	<b>0.971</b>	<b>0.799</b>	0.929	<b>0.862</b>	<b>0.766</b>	<b>0.907</b>
tcs-sar1-shr	0.983	0.982	0.976	0.976	0.979	0.827	0.925	0.875	0.771	0.916
ite-wlsv-shr	0.980	0.981	0.978	0.978	0.979	0.826	0.928	0.875	0.772	0.917
ite-acov-shr	<b>0.946</b>	0.968	0.995	0.983	0.973	0.799	0.932	0.863	0.767	0.909
ite-sar1-shr	0.980	0.982	0.978	0.978	0.979	0.826	0.928	0.876	0.773	0.917
oct-wlsv	0.984	0.983	0.978	0.979	0.981	0.829	0.929	0.878	0.775	0.919
oct-bdshr	<b>1.004</b>	0.992	0.981	0.984	0.990	0.840	0.933	0.885	0.778	0.927
oct-acov	0.964	0.973	0.998	0.988	0.981	0.803	0.936	0.867	0.771	0.915

cross-temporal reconciliation procedures, which for the time being - even when they are computed using the in-sample residuals - are based on simple/unrealistic approximations of the covariance matrix (see Section 6.1). As for this last point, we investigate any significant difference between the reconciled forecasts produced by the most performing heuristic and optimal combination procedures.

### 8.3. Main results

Table 1 presents the AvgRelMSE's obtained for the forecasting techniques (base + 13 reconciliation procedures) listed in the previous sub-section. We provide results for all 95 component time series, and for the 32 upper-level and the 63 bottom-level time series separately. The results are shown by the level of temporal

aggregation and forecast horizon. The lowest error is highlighted in red boldface at each column, while values greater than one, which means that the reconciled forecasts are worse than the base ones, are highlighted in black boldface.

Most of the data in the table are represented in Fig. 5, containing the graphs of the AvgRelMSE's for the considered procedures, across all forecast horizons, by temporal aggregation level of the forecasted series. The procedures have been put in the order given by the overall AvgRelMSE. This seems like a good compromise to represent such a multiple comparison. Fig. 6 shows the Multiple Comparison with the Best (MCB) Nemenyi test. The Friedman test has shown that the forecasts given by the considered procedures are different only when all temporal aggregation levels and forecast horizons (left panel) and one-step-ahead quarterly forecasts

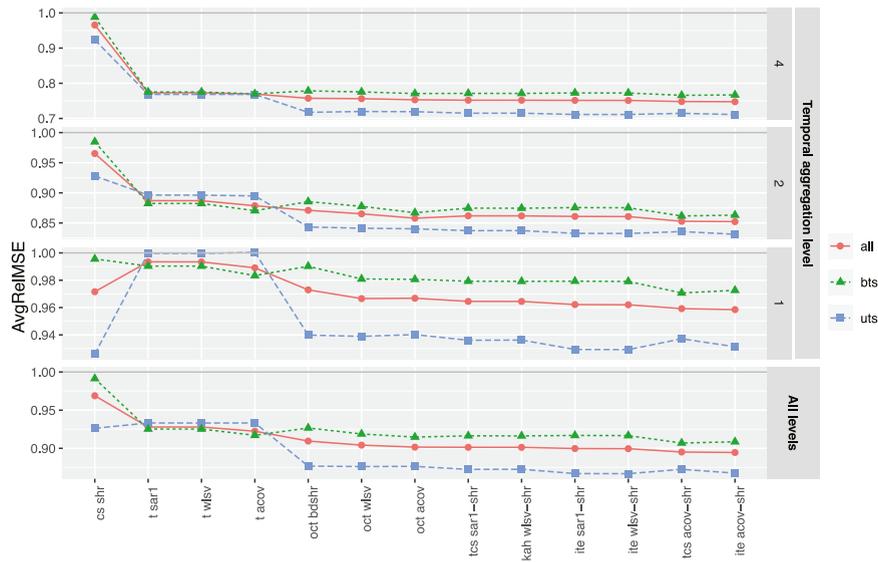


Fig. 5. Average Relative MSE across all series and forecast horizons, by temporal aggregation level.

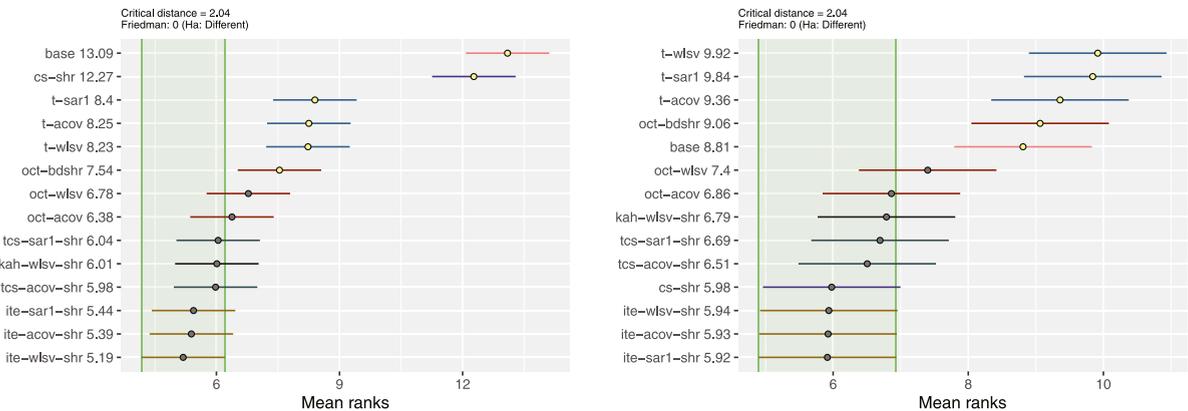


Fig. 6. MCB Nemenyi test results: average ranks and 95% confidence intervals. The reconciliation procedures are sorted vertically according to the MSE mean rank (i) across all time frequencies and forecast horizons (left), and (ii) for one-step-ahead quarterly forecasts (right). The mean rank of each method is displayed to the right of their names. If the intervals of two forecast reconciliation procedures do not overlap, this indicates a statistically different performance. Thus, methods that do not overlap with the green interval are considered significantly worse than the best and vice-versa.

(right panel) are considered. The main results found on this dataset can be summarized as follows:

- the cross-temporal reconciliation reduces the AvgRelMSE for the uts (the most important variables for the decision-maker, e.g., GDP) at any temporal aggregation level and any forecast horizon;
- this accuracy improvement is less marked, though yet visually evident, for the bottom level series, as compared to the reconciled forecasts through temporal hierarchies alone, which however are cross-sectionally incoherent;
- each iterative procedure performs better than its two-step counterpart;
- within the cross-temporal procedures, the heuristic procedures provide better results than the optimal combination ones.

Looking at the performances of each procedure, it is worth noting that cs-shr scores first as for the quarterly

forecasts of the uts, and almost always improves on the base forecasts' accuracy, regardless of series' group, temporal aggregation level, and forecast horizon. The only exception is an AvgRelMSE greater than 1 (1.009) for the bts quarterly forecasts at horizon 4. In addition, from the right panel of Fig. 6 we observe that, when considered only on quarterly basis, the one-step-ahead forecasts for all series provided by cs-shr are (temporally incoherent and) not significantly different from those provided by the best procedure (which in this case is ite-acov-shr). However, since this reconciliation procedure does not account for the temporal dimension, its relative performance worsens (i.e., the cross-temporal procedures improve on the base forecasts more than cs-shr, as shown in the top panels in Fig. 5) as the temporal aggregation level increases. Overall, ite-acov-shr always scores best for all series and all forecast horizons, and second-best for the bts series and all forecast horizons, while

ts-acov-shr scores second and first, in turn. However, ite-acov-shr shows good results for the uts forecasts as well. In this case, the best performances are given by ite-sar1-shr and ite-wlsv-shr. The differences in the forecasts produced by all the considered heuristic procedures are not statistically significant at any temporal aggregation level and forecast horizon.<sup>6</sup> Furthermore, two optimal combination procedures (oct-acov and oct-wlsv) produce reconciled forecasts not significantly different from the best procedure according to the Nemenyi test, while oct-bdshr is significantly (worse and) different from the best forecast reconciliation procedure.

To conclude, the general improvement registered on average (last column of Table 1) by the cross-temporal reconciliation procedures may be considered a positive outcome. They combine an acceptable forecasting performance at the quarterly level with a good performance at the semi-annual and annual levels, with the additional feature that the complete system of forecasts is internally and temporally coherent.

## 9. Conclusions

The hierarchical framework is currently considered an effective way to improve the accuracy of forecasts in many different fields of application. In this paper, we give contributions and extensions to a topic that has been widely studied in the last decade by connecting it to the widespread literature on least-squares adjustment of preliminary data (Stone et al., 1942; Byron, 1978), with a focus on a projection approach for linearly constrained multiple time series. This *de facto* encompasses and extends the modeling framework by Hyndman et al. (2011) (see also Panagiotelis et al., 2021, and Wickramasuriya et al., 2019). However, we do agree with Jeon et al. (2019), p. 368, that a “shortcoming of many of the approaches above (...) is that the weights (...) are a function of in-sample errors and are not directly determined with reference to an objective function ultimately used to assess forecast quality”. This problem, yet present for cross-temporal hierarchies, is added to the dimensionality issues that characterize these structures. The number of nodes is considerably larger than the relevant single-dimension hierarchies and calls for alternative estimation strategies, based for example, on cross-validation, as proposed by Jeon et al. (2019), or - when enough data is available - on Machine Learning techniques (Mancuso et al., 2021; Spiliotis et al., 2020a). Nevertheless, cross-temporal point forecast reconciliation seems to be a promising theme worth considering for future research. Some topics in this field are (i) simulation experiments to better understand behavior, potentiality, and possible shortcomings of the proposed procedures, (ii) more realistic (and hopefully effective) approximations of the covariance matrices for cross-temporal reconciliation; (iii) extending the cross-temporal framework to the reconciliation of probabilistic forecasts (Ben Taieb et al., 2021; Panagiotelis et al.,

2020; Jeon et al., 2019), and for Bayesian (Eckert et al., 2021) and fast (Ashouri et al., 2021) forecast reconciliation procedures; (iv) extending the cross-temporal optimal combination approach to the case of intermittent demand forecasts (Petropoulos & Kourentzes, 2015), with the related non-negativity issues (Wickramasuriya et al., 2020; Kourentzes & Athanasopoulos, 2021), and possible consideration of ‘soft’ constraints (Danilov & Magnus, 2008).

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We would like to thank the Associate Editor and an anonymous referee for their useful feedback on an earlier version of the manuscript.

## Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.ijforecast.2021.08.004>.

## References

- Ashouri, M., Hyndman, R. J., & Shmueli, G. (2021). Fast forecast reconciliation using linear models. *Journal of Computational and Graphical Statistics*, <http://dx.doi.org/10.1080/10618600.2021.1939038>.
- Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, 25(1), 146–166. <http://dx.doi.org/10.1016/j.ijforecast.2008.07.004>.
- Athanasopoulos, G., Gamakumara, P., Panagiotelis, A., Hyndman, R. J., & Affan, M. (2019). Hierarchical forecasting. In P. Fuleky (Ed.), *Macroeconomic forecasting in the era of big data* (pp. 689–719). Cham: Springer, [http://dx.doi.org/10.1007/978-3-030-31150-6\\_21](http://dx.doi.org/10.1007/978-3-030-31150-6_21).
- Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1), 60–74. <http://dx.doi.org/10.1016/j.ejor.2017.02.046>.
- 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. <http://dx.doi.org/10.1080/01621459.2020.1736081>.
- Bisaglia, L., Di Fonzo, T., & 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* (pp. 951–956). Pearson, URL: <https://it.pearson.com/content/dam/region-core/italy/pearson-italy/pdf/Docenti/Universit%C3%A0/Pearson-SIS-2020-atti-convegno.pdf>.
- Byron, R. P. (1978). The estimation of large social account matrices. *Journal of the Royal Statistical Society. Series A*, 141(3), 359–367. <http://dx.doi.org/10.2307/2344807>.
- Danilov, D., & Magnus, J. R. (2008). On the estimation of a large sparse Bayesian system: The Snaer program. *Computational Statistics and Data Analysis*, 52(9), 4203–4224. <http://dx.doi.org/10.1016/j.csda.2008.02.019>.
- Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3), 510–522. <http://dx.doi.org/10.1016/j.ijforecast.2012.09.002>.

<sup>6</sup> The Nemenyi tests for each temporal aggregation level and each forecast horizon are available on request from the corresponding author.

- Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11(4), 427–444. <http://dx.doi.org/10.1214/aoms/1177731829>.
- 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. <http://dx.doi.org/10.1080/01621459.1976.10481478>.
- Eckert, F., Hyndman, R. J., & Panagiotelis, A. (2021). Forecasting Swiss exports using Bayesian forecast reconciliation. *European Journal of Operational Research*, 291(2), 693–710. <http://dx.doi.org/10.1016/j.ejor.2020.09.046>.
- Fliedner, G. (2001). Hierarchical forecasting: Issues and use guidelines. *Industrial Management & Data Systems*, 101(1), 5–12. <http://dx.doi.org/10.1108/02635570110365952>.
- Girolimetto, D., & Di Fonzo, T. (2021). Package FoReco: Point forecast reconciliation. URL: <https://cran.r-project.org/package=FoReco>.
- Gross, C. W., & Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, 9(3), 233–254. <http://dx.doi.org/10.1002/for.3980090304>.
- Harville, D. A. (2008). *Matrix algebra from a statistician's perspective*. New York: Springer-Verlag.
- Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, 55(9), 2579–2589. <http://dx.doi.org/10.1016/j.csda.2011.03.006>.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). Melbourne, Australia: OTexts, URL: [OTexts.com/fpp2](http://OTexts.com/fpp2).
- Hyndman, R. J., & Kourentzes, N. (2018). Package thief: Temporal HIERarchical Forecasting. URL: <https://cran.r-project.org/package=thief>.
- Hyndman, R. J., Lee, A. J., & Wang, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics and Data Analysis*, 97, 16–32. <http://dx.doi.org/10.1016/j.csda.2015.11.007>.
- Hyndman, R. J., Lee, A. J., Wang, E., & Wickramasuriya, S. L. (2020). hts: Hierarchical and grouped time series. URL: <https://cran.r-project.org/package=hts>.
- 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. <http://dx.doi.org/10.1016/j.ejor.2019.05.020>.
- Koning, A. J., Franses, P. H., Hibon, M., & Stekler, H. O. (2005). The M3 competition: Statistical tests of the results. *International Journal of Forecasting*, 21(3), 397–409. <http://dx.doi.org/10.1016/j.ijforecast.2004.10.003>.
- Kourentzes, N., & Athanasopoulos, G. (2019). Cross-temporal coherent forecasts for Australian tourism. *Annals of Tourism Research*, 75, 393–409. <http://dx.doi.org/10.1016/j.annals.2019.02.001>.
- Kourentzes, N., & Athanasopoulos, G. (2020). On the evaluation of hierarchical forecasts. URL: <https://www.monash.edu/business/ebs/research/publications/ebs/wp02-2020.pdf>.
- Kourentzes, N., & Athanasopoulos, G. (2021). Elucidate structure in intermittent demand series. *European Journal of Operational Research*, 288(1), 141–152. <http://dx.doi.org/10.1016/j.ejor.2020.05.046>.
- Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365–411. [http://dx.doi.org/10.1016/S0047-259X\(03\)00096-4](http://dx.doi.org/10.1016/S0047-259X(03)00096-4).
- Magnus, J. R., & Neudecker, H. (2019). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M5 accuracy competition: Results, findings and conclusions. URL: [https://www.researchgate.net/publication/344487258\\_The\\_M5\\_Accuracy\\_competition\\_Results\\_findings\\_and\\_conclusions](https://www.researchgate.net/publication/344487258_The_M5_Accuracy_competition_Results_findings_and_conclusions).
- Mancuso, P., Piccialli, V., & Sudoso, A. M. (2021). A machine learning approach for forecasting hierarchical time series. *Expert Systems with Applications*, <http://dx.doi.org/10.1016/j.eswa.2021.115102>.
- Miller, R. E., & Blair, P. D. (2009). *Input-output analysis: foundations and extensions*. New York: Cambridge University Press.
- 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. <http://dx.doi.org/10.1016/j.ejor.2019.07.061>.
- 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), 343–359. <http://dx.doi.org/10.1016/j.ijforecast.2020.06.004>.
- Panagiotelis, A., Gamakumara, P., Athanasopoulos, G., & Hyndman, R. J. (2020). Probabilistic forecast reconciliation: Properties, evaluation and score optimisation. (pp. 1–43). URL: <https://www.monash.edu/business/ebs/research/publications/ebs/wp26-2020.pdf>.
- Petropoulos, F., & Kourentzes, N. (2015). Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, 66(6), 914–924. <http://dx.doi.org/10.1057/jors.2014.62>.
- Spiliotis, E., Abolghasemi, M., Hyndman, R. J., Petropoulos, F., & Assimakopoulos, V. (2020). Hierarchical forecast reconciliation with machine learning. [arXiv:2006.02043](https://arxiv.org/abs/2006.02043).
- Spiliotis, E., Petropoulos, F., Kourentzes, N., & Assimakopoulos, V. (2020). Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption. *Applied Energy*, 261(November 2019), Article 114339. <http://dx.doi.org/10.1016/j.apenergy.2019.114339>.
- Stone, R., Champernowne, D. G., & Meade, J. E. (1942). The precision of national income estimates. *Review of Economic Studies*, 9(2), 111–125. <http://dx.doi.org/10.2307/2967664>.
- van Erven, T., & Cugliari, J. (2015). Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. In A. Antoniadis, J.-M. Poggi, & X. Brossat (Eds.), *Modeling and stochastic learning for forecasting in high dimensions (vol. 217)* (pp. 297–317). Springer Science and Business Media, LLC. [http://dx.doi.org/10.1007/978-3-319-18732-7\\_15](http://dx.doi.org/10.1007/978-3-319-18732-7_15).
- 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. <http://dx.doi.org/10.1080/01621459.2018.1448825>.
- Wickramasuriya, S. L., Turlach, B. A., & Hyndman, R. J. (2020). Optimal non-negative forecast reconciliation. *Statistics and Computing*, 30(5), 1167–1182. <http://dx.doi.org/10.1007/s11222-020-09930-0>.
- Yagli, G. M., Yang, D., & Srinivasan, D. (2019). Reconciling solar forecasts: Sequential reconciliation. *Solar Energy*, 179(December 2018), 391–397. <http://dx.doi.org/10.1016/j.solener.2018.12.075>.