Forecasting combination of hierarchical time series: a novel method with an application to CoVid-19

Abstract:

Multiple, hierarchically organized time series are routinely submitted to the forecaster upon request to provide estimates of their future values, regardless the level occupied in the hierarchy. In this paper, a novel method for the prediction of hierarchically structured time series will be presented. The idea is to enhance the quality of the predictions obtained using a technique of the type forecast reconciliation, by applying this procedure to a set of optimally combined predictions, generated by different statistical models. The goodness of the proposed method will be evaluated using the official time series related to the number of people tested positive to the SARS-CoV-2 in each of the Italian regions, between February 24th 2020 and August 31th 2020.

1. Introduction

In many applications, it is often the case that accurate forecasts are needed for time series showing an inherent hierarchical structure. For example, in economics the forecaster is routinely asked to provide separate forecasts for the industrial production index at the most aggregated level as well as for specific (sub–) classes of economic activities. The estimation of the future demand of domestic tourism usually follows a geographical proximity criterion, based on which the related time series are organized (and predicted) according to homogeneous groups. Sometimes, emergency situations require close monitoring of the spread of a disease not only at a national but also at a regional level, e.g. in order to set up more appropriate countermeasures for elderly and chronically ill people. These are all cases where a single line of hierarchy generates the overall structure of the data which therefore is referred to as "hierarchical time series". The present paper is concerned with the forecast of such data structures. While on one hand it is always possible to disregard the underlying hierarchical arrangement and thus carry out the prediction exercise considering each time series singularly, on the other hand, by doing so, it is very unlikely for the resulting higher level forecasts to be equal to the sum of the lower level ones. It goes by itself that such a situation is not acceptable in many instances, e.g. in the field of official statistics where aggregation consistency is generally a *conditio sine qua non* and even a slight misalignment needs to be dealt with. Therefore, many combination techniques have been designed to preserve the

needed adding-up conditions, by accounting for the position occupied in the hierarchy by each and every time series, regardless their level of (dis–)aggregation. However, this type of approach – usually referred to as "forecast reconciliation" – is in general dependent on the statistical model *a priori* chosen to carry out the forecasting exercise. Undoubtedly, this choice might negatively impact the quality of the generated forecasts, e.g. by conveying not negligible amount of uncertainty into the analysis. This is especially true in the case of real-life data – where problems, such as small sample size, noise and systematic and/or non systematic errors – might hinder the choice of the "right" statistical model.

Motivated by this, in the present paper a method built upon the forecast reconciliation procedure devised by Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. (2011) will be presented. In other words, a joint hierarchical forecasting system will be formulated, where an additional optimality condition, derived in a multi-model setup of the type forecast combination, drives the choice of the "best" statistical model generating the predicted values. The final part of the procedure is designed to lower the bias of the selected forecasts. The main novelty of the method is that the forecast combination is applied directly on forecasts which have already been reconciled. In essence, it is an optimization procedure articulated in four steps: one performed at a cross-section level (reconciliation), two at a cross-model level (forecast combination) and the final one on the chosen prediction vector, for bias adjustment purposes. Surprisingly, to the best of the author's knowledge, this is the first attempt of this sort in the case of cross-sectional hierarchical time series.

The rest of the paper is structured as follows: Section 2 is devoted to the literature review concerning the two statistical methods the proposed procedure is based on, which will be detailed in the following Section 3. The proposed method, as well as its justification, will be respectively illustrated in Section 4 and 5. The following Section 6 will be devoted to an extensive empirical application, carried out using the official Italian data related to the SARS-CoV-2 positive cases, which will demonstrate the validity of the proposed approach. Section 7, containing the conclusions and the future directions of this work, will end the paper.

2. Literature review

As already mentioned, the proposed procedure is based on two classes of methods, usually referred to as "forecast reconciliation" and "forecast combination". The former serves the purpose of achieving aggregation consistency of individual, aggregation inconsistent, forecasts whereas the latter will be employed to combine different reconciled forecasts, each of them generated according to different statistical models.

A rigorous and theoretically sound investigation on forecasts combination dates back to the late 60s – with the famous seminal paper by Bates, J. M.,

and Granger, C. W. (1969). Here, the Authors showed that the combination of forecasts often leads to a better forecast accuracy and, by doing so, provided an alternative way to the notion that a "best" method exists and can be identified. Ever since this paper, the integration of a number of forecasts, independently estimated on a single time series, has attracted a great deal of research interest and, as a result, a vast literature is today available. Much of it is aimed at presenting empirical applications documenting the appealing features of this approach, which in many cases can improve even upon the best individual forecast, in terms of forecast risk, forecast error variance and consistency between in-sample and out-of-sample error distributions, as pointed out by Barrow, D. K., and Kourentzes, N. (2016). This can happen for a variety of reasons, many of them related to the fact that the choice of the "right" model, in general, implies the injection of not negligible amounts of uncertainty into the analysis (Chambers, J. C., Mullick, S. K., and Smith, D. D. (1971) and Chatfield, C. (1996). In the same line of thinking, many Authors, see for example Makridakis, S. (1989), Stock, J. H., and Watson, M. W (2001) and Stock, J. H., and Watson, M. W. (2004), emphasize the dangers related to misspecification errors, which, on the other hand, can be mitigated by combining the forecasts yielded by a number of models. In support of this argument, there are a number of studies which show that it is very unlikely, using a well calibrated portfolio of models, that one of them consistently dominates the others across the whole prediction window. Such an argument is consistent with the view that the "true" underlying data generating process is, saved for trivial or lab controlled cases, way too complicated to be adequately captured by a single model. This is the position, for example, of Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997), according to whom the data can never support, and we can never identify, the "true" model. Therefore, the selection of a statistical model is more realistically the process of identifying the best approximating one. Once defined models as approximations, the concept of the identification of the "true" one ceases to be decisive in favor of approaches pursuing, in the first place, the goal of achieving good forecasting performances.

Many practical uses of forecast combination are discussed in the excellent work of Clemen, R. T. (1989), where the author covers a wide spectrum of applications, ranging from economics, demography and politics to meteorology and outcomes of football games. In the same spirit is the more recent paper by Mancuso, A. C. B., and Werner, L. (2013), which presents a classification of 174 articles focusing on forecast combination. In particular, new applications are reported from different sectors, such as commercial, tourism, urban traffic, betting market and propagation of successful innovations. The analysis of the outcomes of the M3 forecast competition, discussed in Makridakis, S., and Hibon, M. (2000), goes in favor of the forecast combination approach, which on average proved to outperform the methods individually applied. Same conclusion applies to the more recent M4 forecast competition, discussed in Mancuso, A. C. B., and Werner, L. (2020), where 12 out of a group of 17 most accurate methods are based on the combination of forecasts. Other considerations

in favor of this approach are more closely related to the features of the time series under investigation. For instance, in many real-life cases they are affected by structural breaks, induced by a variety of factors whose real-time detection is generally difficult to achieve. However, in a multi–model setup it is not unreasonable to have models showing different degrees of ability in handling such events. Such a situation can translate into gains in terms of forecast accuracy, as it has been argued not only since the very beginning, in the above mentioned paper by Bates and Granger, but also in more recent times, by, among others, Clements, M. P., and Hendry, D. F. (2002), Sessions, D. N., and Chatterjee, S. (1989) and Makridakis, S. (1989). All these Authors concur on the premise that, on average, combining the forecasts yielded by models with different reaction times to a given intervention – and thus requiring stretches of post-break data of different length – can do a better job than individual models. For what said, it comes at no surprise that such appealing results might lead to a change in the perspective many researchers and practitioners look at the forecasting methods, i.e. from model selection – based on the assumption of the existence of one, "true" data generating process – to model averaging.

When the data set under investigation show a hierarchical structure, forecast combination techniques (but this holds true for any univariate forecasting procedure) are insensitive to the level of aggregation at which they are applied. Consistently, in such cases, the independent forecasting of the component time series is always possible, even if not advisable, due to the very likely lack of consistency occurring between the sum of the predictions generated at one level with those available at the level above. But this is not the whole story: by applying forecasting procedures at the components level, rather than limit them to the most aggregate one, it is possible to adequately capture the data covariance structure and thus achieve not negligible gains in terms of quality of the predictions. This fact has been pointed out, *inter alia*, by Fair, R. C., and Shiller, R. J. (1990), Marcellino, M., Stock, J. H., and Watson, M. W. (2003) and Hubrich, K. (2005). Their conclusions are related to two of the most traditional approaches, usually referred to as Top-Down and Bottom-Up (see, for example, Schwarzkopf, A. B., Tersine, R. J., and Morris, J. S. (1988), Lapide, L. (2006) and Athanasopoulos, G., Ahmed, R. A., and Hyndman, R. J. (2009)). While the former envisions a two-step procedure – i.e. the forecasting is first performed at the top level and then, by disaggregating these data based on the historical percentage of each data point, within the whole group – in the latter each and every time series is first individually predicted and then all the forecasts are summed up. There is another approach which has gained widespread acceptance over the years, known as "middle-out". It can be considered an extension of the top-down approach, since the forecast is first generated at two separate levels (upper and lower) and then combined in a proper manner to form a composite forecast. It is worth outlining how all of these methods can be considered sub-optimal insofar they neglect the correlation structure existing among the series belonging to the same level. Finally, a more recent approach known as optimal combination, envisions a two-step procedure

where first the sequential and exhaustive forecast of each and every time series is independently performed and then – by optimally combining the predicted values obtained – are aggregated to achieve consistency across the hierarchical levels (reconciliation). Theoretically, cross-level coherency can be attained by means of Generalized Least Square (GLS) whose employment, however, turns out to be unfeasible due to the unidentifiability of the covariance matrix of the reconciliation errors (Wickramasuriya, S. L., Athanasopoulos, G., and Hyndman, R. J. (2019)). However, other methods, e.g. of the type OLS (Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. (2011)) or WLS (Hyndman, R. J., Lee, A. J., and Wang, E. (2016)) can be used to circumvent this hurdle.

The method proposed in the present paper is of the type mixed, in the sense that exploits the approaches related to both forecasts reconciliation and forecasts combination. It is noted how mixed methods of this sort are not often encountered in literature. Such a situation might be due to the relatively recent introduction of reconciliation methods capable, unlike more traditional procedures, of accounting for the correlation structures among the series within a given hierarchical level, and thus able to deliver better performances. On the other hand, the need of a unified framework combining these two approaches has been recently brought up by Di Fonzo, T., and Girolimetto, D. (2021), which discussed an *ad hoc*, bi-dimensional (cross-sectional and temporal) procedure built upon a recent proposal by Wickramasuriya, S. L., Athanasopoulos, G., and Hyndman, R. J. (2019). Their method employs all the summation constraints arising in the cross-temporal hierarchical structure to reconciliate the base forecasts, using simple projections in a suitable linear space. On the other hand, the recent proposal by Spiliotis, E., Petropoulos, F., and Assimakopoulos, V. (2019) envisions a common framework where both forecast reconciliation and combination of forecasts generated by multiple models work together. Their method arises from the consideration that base forecasts should not be derived from a single method but a combination of methods. In the same direction goes the early work by Van Erven, T., and Cugliari, J. (2015), where both combination and reconciliation of the forecasts are applied in a two-step procedure, i.e. "first one comes up with the best possible forecasts for the time series without worrying about aggregation consistency and then a reconciliation procedure is used to make the forecasts aggregate consistent". As it will be seen, the procedure illustrated in the present paper significantly differs from the above mentioned ones, in that different base forecasts are generated according to an arbitrary, pre-specified portfolio of statistical models, so that the combination exercise is performed on a set of already reconciled forecasts. Finally, while these authors focus on temporal aggregation this paper considers cross-sectional aggregations.

3. The framework

In this Section, an explanation of the framework within which the proposed method operates is given. In particular, the hierarchy structure of reference

5

along with the employed reconciliation method are illustrated. The forecast combination part will be explained using two real-life examples of portfolios – which will be both used in the empirical Section – related to the statistical prediction models and the forecast combination techniques entertained.

3.1. Hierarchical cross-sectional reconciliation: the chosen method

This paper focuses on structures of the type summation constrained, in the sense that the underlying hierarchical structure of a given $m-$dimensional time series $\boldsymbol{x}_t$, arises by summing up the bottom-level series into the higher ones. Figure 1 is an example of such a structure, under the condition that the constraints $x_t = x_{a,t} + x_{b,t}$, $x_{a,t} = x_{aa,t} + x_{ab,t} + x_{ac,t}$ and $x_{b,t} = x_{ba,t} + x_{bb,t}$ are all satisfied.
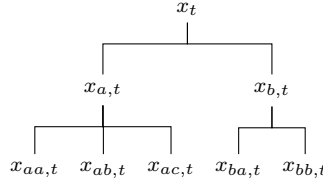

Fig. 1: A two-level hierarchical structure

Formally, we have that the observed data $\boldsymbol{x}_t$ – as well as their estimated future values, defined as $\boldsymbol{x}_h$; $h = 1, 2, \ldots, H$, with H the prediction horizon – lie in the summation-coherent subspace $\{\mathcal{U}\}$; $\forall t = 1, 2, \ldots, T$ and $\forall h = 1, 2, \ldots, H$. The prediction step subscript h has been omitted in Figure 1, for the sake of a better readability. In total, this hierarchy contains $m = 8$ time series, $n = 5$ of which are the lowest level time series, which therefore constitute the highest level of disaggregation. The observed series $\boldsymbol{x}_t \in \mathbb{R}^m$ can be broken down as follows: $\boldsymbol{x}_t = [\boldsymbol{u}_t', \boldsymbol{b}_t']'$, where $\boldsymbol{b}_t' \in \mathbb{R}^n$ and $\boldsymbol{u}_t' \in \mathbb{R}^{m-n}$ respectively contain the data pertaining to the bottom and upper series. Therefore, according this representation, the structure of Figure 1 (omitting the subscript t) can be broken down as follows: $[\boldsymbol{u}_t', \boldsymbol{b}_t']' \equiv [x, x_a, x_b, x_{aa}, x_{ab}, x_{ac}, x_{ba}, x_{bb}]'$, $\boldsymbol{u}_t' \equiv [x_a, x_b]'$ and $\boldsymbol{b}_t' \equiv [x_{aa}, x_{ab}, x_{ac}, x_{ba}, x_{bb}]'$. The hierarchical structure – satisfying $\boldsymbol{x} \subset \{\mathcal{U}\}$ – is induced by the summing matrix $\boldsymbol{S}$ of dimension $m \times n$ such that $\boldsymbol{x}_t = \boldsymbol{S}\boldsymbol{b}_t$. Formally: $\boldsymbol{x} \subset \{\mathcal{U}\} \iff \boldsymbol{x}_t = \boldsymbol{S}\boldsymbol{b}_t$ (the symbol $\iff$ replacing the locution "if

and only if"). The $\boldsymbol{S}$ matrix for the hierarchy in Figure 1 is as follows:

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Using the symbols ˜ and ˆ respectively to refer to the case of coherent and base (generally non coherent) forecasts, the reconciled forecast $h-$ step ahead can be expressed as proposed by Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. (2011), i.e.

$$\tilde{\boldsymbol{x}}(h) = \boldsymbol{SP}\hat{\boldsymbol{x}}(h), \tag{1}$$

for some appropriately chosen matrix $\boldsymbol{P} \in \mathbb{R}^{m \times n}$. Assuming unbiased base forecasts, the best linear unbiased revised forecasts (minimizer of the sum of the variances of the lower hierarchical level) are given by Equation 1 with

$$\boldsymbol{P} = (\boldsymbol{S'W^{-1}S})^{-1}\boldsymbol{S'W^{-1}} \tag{2}$$

and thus (see Taieb, S. B., Taylor, J. W., andHyndman, R. J. (2017), Theorem 1)

$$\tilde{\boldsymbol{x}}(h) = \boldsymbol{S}(\boldsymbol{S'W^{-1}S})^{-1}\boldsymbol{S'W^{-1}}\hat{\boldsymbol{x}}(h), \tag{3}$$

where $\boldsymbol{S}$ is as above defined and $\hat{\boldsymbol{x}}(h)$ and $\tilde{\boldsymbol{x}}(h)$; $h = 1, 2, \ldots, H$ represent respectively the set of H predictions independently generated and the ones made coherent. Finally, $\boldsymbol{W}$ is the positive definite covariance matrix of the base forecast errors, i.e. $\hat{\boldsymbol{e}}_t(h) = \hat{\boldsymbol{x}}_t(h) - \boldsymbol{x}_t(h)$, so that $\boldsymbol{W}(h) = \mathbb{E}[\hat{\boldsymbol{e}}_t(h) - \hat{\boldsymbol{e}}'_t(h)]$. As shown by Wickramasuriya, S. L., Athanasopoulos, G., and Hyndman, R. J. (2019), matrix $\boldsymbol{W}(h)$ appears in the equation for the estimation of the error variance of the reconciled forecasts, i.e.

$$\boldsymbol{V}(h) = Var[\boldsymbol{x}(T + h) - \tilde{\boldsymbol{x}}(h)] = \boldsymbol{SPW}(h)\boldsymbol{P'S'}, \tag{4}$$

whose diagonal elements are the variances of the forecast errors. Their minimization can thus be performed in terms of the trace of $\boldsymbol{V}(h)$ and given by Equation 2 (therefore, this method is called Minimum Trace Estimator). Unfortunately, as proved by the same Authors, $\boldsymbol{W}$ is not identifiable, therefore, in the empirical section, the work around proposed by them will be adopted. In essence, it is assumed $\boldsymbol{W}_h = k(h)\text{diag}\hat{\boldsymbol{W}}_1$; $\forall h$ and assuming $k(h) > 0$ and denoting with $\boldsymbol{W}_1$ the forecast errors covariance matrix estimated at horizon $h = 1$ – i.e. $\hat{\boldsymbol{W}}_1 = \frac{1}{T}\sum_1^T \hat{\boldsymbol{e}}_t\hat{\boldsymbol{e}}'_t$ – and with K is an unknown constant depending on the time horizon h.

3.2. The Forecast combination methods adopted

As already mentioned, the proposed method uses a set of combination methods, out of which the winner is selected according to a suitable loss function. In many empirical studies, it is shown how forecast combinations on average delivers better performances than methods based on a single forecasting statistical models. The theoretical validity of this approach is rooted in the assumption that the dimension of the sample sizes available in real-life applications are usually finite and, as a result, the correct specification of the "true" underlying data generation process is not attainable.

In what follows it is assumed $\boldsymbol{x}_t$ to be the variable of interest and that $\boldsymbol{f}_t = (f_{1t}, f_{2t}, \ldots, f_{Nt})'$ are the N, not perfectly collinear, available forecasts, whose combination is expressed as $f = \sum_{i=1}^{N} w_i f_{it}$, or, equivalently, $f = \boldsymbol{f}_t' \boldsymbol{w}$, being w's the combination weights.

The first method considered in this paper is of the type simple average. Despite its inherent simplicity (it ignores the correlation structure of the forecast errors) this method has been adopted given its ability, proved true in many cases, to "dominate more refined combination schemes aimed at estimating the theoretically optimal combination weights" (Atiya, A. F. (2020)). The simple average assigns equal weights to all predictors, i.e. $\boldsymbol{w}^{sa} = \frac{1}{N}$ and thus the combined forecast is

$$f = \boldsymbol{f}_t' \boldsymbol{w}^{sa}.$$

In the second method chosen, the forecast combination weights:

$$\boldsymbol{w}^{ols} = (w_1, w_2, \ldots, w_N), \tag{5}$$

along with the intercept b, are computed using ordinary least squares (OLS) regression (Granger, C. W., and Ramanathan, R. (1984)), i.e.

$$f = b + \boldsymbol{f}_t' \boldsymbol{w}^{ols}. \tag{6}$$

The third method applied – of the type Least Absolute Deviation (LAD) – is a modification of the OLS method, and it is expressed as in Equation 6, replacing the superscript *ols* with *lad*. Since the method of least squares assigns heavy weights on the error terms, the more robust estimator LAD – of the type Gauss–Laplace (see, e.g. Dielman, Terry E. (2005)) – minimizes the absolute values and not the squared values of the error term. This features is particularly useful when the error term is generated by distributions having a infinite variance (fat tails) caused by outliers in the disturbance term.

Finally, a modification of the method proposed by Newbold, P., and Granger, C. W. (1974), built upon an earlier methodology of Bates, J. M., and Granger, C. W. (1969), is our fourth approach. Let $\sum$ be the positive definite matrix of the mean squared prediction errors (MSPE) of $\boldsymbol{f}_t$ and $\boldsymbol{g}$ is an $N \times 1$ vector

8

of $(1, 1, \ldots, 1)'$ their method relies on a constrained minimization of the MSPE under the normalizing condition $\boldsymbol{g'w} = 1$. The resulting combination of weights is

$$\boldsymbol{w}^{ng} = \frac{\sum^{-1} \boldsymbol{g}}{\boldsymbol{g'} \sum^{-1} \boldsymbol{g}},$$

so that the combined forecast is

$$f = \boldsymbol{f}_t' \boldsymbol{w}^{ng}. \tag{7}$$

However, unlike the original method, the variant employed here follows the proposal by Hsiao, C., and Wan, S. K. (2014), which does not impose the prior restriction that the matrix $\sum$ is diagonal.

3.3. The entertained statistical models

Before delving into the proposed method, a quick presentation of the forecasting methods employed in the empirical section is in order. The first two statistical models considered are of the type ARIMA (Auto Regressive Fractional Moving Average) (Box, G. E., Jenkins, G. M., and Reinsel, G. (1970)) and *ARFIMA* (Auto Regressive Fractional Moving Average) (Granger, C. W., and Ramanathan, R. (1984) and Granger, C. W., and Joyeux, R. (1980)). Being the latter a generalization of the former, the two models will be presented conjointly.

ARFIMA (Auto Regressive Integrated Fractional Moving Average) models are useful in circumstances where the underlying stochastic process exhibits hyperbolic decay patterns in their estimated autocorrelation function. ARFIMA-type processes are usually expressed as follows:

$$\boldsymbol{\Phi}(B)(1 - B)^d x_t = \boldsymbol{\Theta}(B)\varepsilon_t; \quad \varepsilon_t \sim \text{i.i.d.}(0, \sigma^2),$$

where d is a parameter – assumed to take non-integer values in the difference operator $(1 - B)^d$, with B identifying the backward operator, that is $B^k x_t = x_{t-k}$. The fractional differencing operator is defined by the binomial expansion $(1 - B)^d = \sum_0^\infty \binom{d}{i} (-B)^i$. The process is stationary and invertible if the roots of the autoregressive polynomial of order p, $\boldsymbol{\Phi}(B) = 1 - \phi_1 B - \phi_2 B^2, \ldots, -\phi_p B^p$, and the order q moving-average part, $\boldsymbol{\Theta}(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$, lie outside the unit circle with $|d| < 0.5$.

ARFIMA models generalize the ARIMA(p,d,q) representation where the parameter d is constrained to integer values. This type of model has been designed to capture approximately parabolic decay patterns of the empirical autocorrelation function. As such, they are suitable to model persistence structures embedded in the underlying stochastic process of the type short-memory.

Theta method – the third forecasting model considered – is a powerful class of models which have been proposed by Assimakopoulos, V., and Nikolopoulos, K. (2000). Define with the symbol ∇ the difference operator – i.e. $\nabla x_t = x_t - x_{t-1}$, x_t being the original time series – this method is the solution of the equation

$$\nabla^2 z_t(\theta) = \theta \nabla^2 x_t; \quad t = 3, 4, \ldots, n, \tag{8}$$

with $z_t(\theta)$'s analytical solution reading as following: $z_t(\theta)\theta x_t + (1 - \theta)(A_n + B_n t); \quad t = 1, 2, \ldots, n$, where A_n and B_n are the minimum square coefficient of a linear regression equation of the series x_t against $\mathbf{1}_n$, i.e. the vector of ones of length n. These terms are given by $A_n = \frac{1}{N} \sum_{t=1}^{n} x_t - \frac{n+1}{2} B_n$ and $B_n = \frac{6}{n^2 - 1} \left[\frac{2}{N} \sum_{t=1}^{N} x_t t - \frac{1+n}{n} \sum_{t=1}^{N} x_t \right]$. Finally, the initial values z_1 and z_2 in Equation 8 are estimated by minimization of $\sum_{t=1}^{n} (|x_t - z_t(\theta)|^2)$.

The fourth and last model employed is of the type exponential smoothing, proposed in 1944 by Robert G. Brown, a US Navy operations research analyst (Gardner Jr, E. S. (2006)). Specifically, two schemes have been employed here: Additive Holt Error Model (AEM) and multiplicative Holt Error Model (MEM) (as it will be seen, the procedure automatically will select the "best" one). As for the AEM, let $\mu_t = \hat{x}_t = l_{t-1} + b_{t-1}$ be the one-step ahead forecast of the observed time series x_t generated by the forecasting equation $x_t = l_{t-1} + b_{t-1} + \varepsilon_t$, being l_t a measure of the level of the series, b_t an estimate of the slope (or growth) at time t and $\varepsilon_t = x_t - \mu_t$ the one-step-ahead forecast error, referred to the time t. The level and slope equations for AEM are respectively represented as

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t$$
$$b_t = b_{t-1} + \beta(l_t + l_{t-1} - b_{t-1}) = b_{t-1} + \alpha \beta \varepsilon_t.$$

By re-expressing the error term ε_t as $\varepsilon_t = \frac{(x_t - \mu_t)}{\mu_t}$ (relative errors), the forecast, level and slope equations for the MAM model are as follows:

$$l_t = (l_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t)$$
$$b_t = b_{t-1} + \beta(lt - 1 + b_{t-1})\varepsilon_t.$$

In the above two sets of equations α and β are the model parameters to be estimated.

4. The proposed method

Let us indicate with the symbols $\mathcal{R}$ and $|\cdot|$ respectively a suitable reconciliation method and the cardinality function (assuming the number of elements in a given set to be finite, $|\cdot|$ simply returns the number of the elements belonging to that set). Let the symbol **ncol** identify the function which, applied to a given matrix, returns its number of columns and $\mathcal{M} \equiv \{\mu_1, \mu_2, \ldots, \mu_M\}$ and

10

$\mathcal{D} \equiv \{\delta_1, \delta_2, \ldots, \delta_D\}$ respectively the set of $|\mathcal{M}| = M$ prediction models and the set $|\mathcal{D}| = D$ of forecast combination methods entertained, both arbitrarily chosen. Once applied to the time series of interest x_t; $t = 1, 2, \ldots, T$, each model $\{\mu_j \in \mathcal{M}; \ j = 1, 2, \ldots M\}$ generates a set, called $\mathcal{F}^H$, made up with M H–step ahead predictions, i.e.: $\{\mathcal{F}^H(\mu_j); \ j = 1, 2, \ldots M\}$. Each of the elements of this set is a base forecasts, in the sense that it is generated by individually applying a given statistical model μ_j to the observed time series without any attempt of reconciliation.
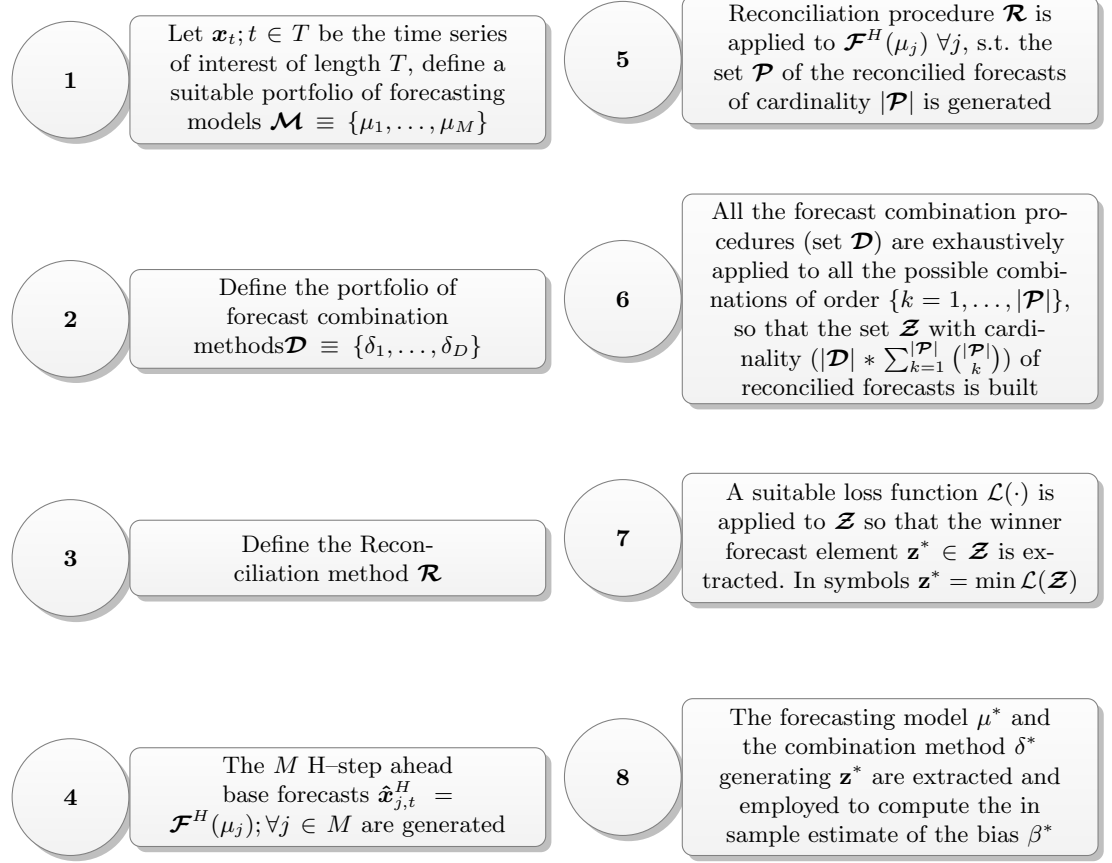
Each of these M elements in $\mathcal{F}$ (the M forecast vectors) is individually reconciled through the reconciliation procedure $\mathcal{R}$, i.e. $\{\mathcal{R}(\mathcal{F}(\mu_j); \ j = 1, 2, \ldots M)\}$ (the superscript h is omitted for brevity). At this point, the resulting set $\{\mathcal{P}(\mu_j); \ j = 1, 2, \ldots M\}$ of M model–dependent reconciled forecasts (first optimization) is optimally combined by applying each method in the set $\mathcal{D}$ to any possible combination (without repetition) of order $\{k = 1, 2, \ldots, M\}$ to the set $\mathcal{P}$ (second optimization). The resulting set $\mathcal{Z}$ – with cardinality ($|\mathcal{D}| *$ $\sum_{k=1}^{|\mathcal{P}|} \binom{M}{k}$)) – contains all the possible combinations – $\forall k$-order – of the model-dependent reconciled forecasts. The third optimization step is carried out by applying to $\mathcal{Z}$ a suitable loss function, here denoted with the symbol $\mathcal{L}(\cdot)$. The optimal vector of forecasts is thus the element $\mathbf{z}^* \in \mathcal{Z}$ minimizing this function, i.e. $\mathbf{z}^* = \min \mathcal{L}(\mathcal{Z})$. This optimality condition is expressed as

$$\mathbf{z}^* = f(\mu^*, \delta^*), \tag{9}$$

being the arguments of f respectively the "best" forecasting model and forecast combination technique. This last step, by ruling out the less performing combination method(s), has been introduced in order to reduce the overall uncertainty level of the analysis. In fact, suppose that the original set $\mathcal{D}$ reduces to $\mathcal{D}'$ – being clearly $|\mathcal{D}'| < |\mathcal{D}|$ – the additional amounts of undesired fluctuations and noise – which one can reasonably expect as a consequence of the employment of one (more) under-performing combination method(s) – are avoided. Finally, the model bias β^* is empirically estimated using the in–sample residuals generated by employing the winners techniques μ^* and δ^*, according to an optimal choice made on a set of suitable central tendency functions (fourth optimization).
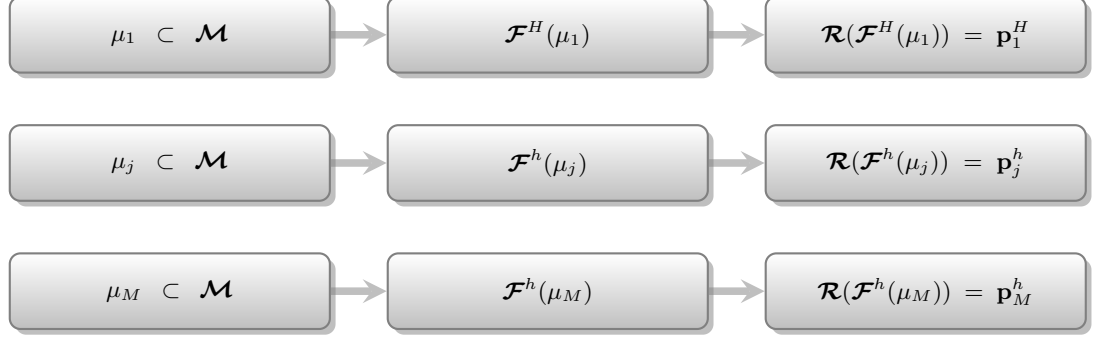
For the sake of clarity, a more schematic description of the method is given below, in the form of algorithm presented in a step-by-step fashion.

Figure 1 Algorithm of the proposed method

1 — Let $\boldsymbol{x}_t; t \in T$ be the time series of interest of length T, define a suitable portfolio of forecasting models $\boldsymbol{\mathcal{M}} \equiv \{\mu_1, \ldots, \mu_M\}$

5 — Reconciliation procedure $\boldsymbol{\mathcal{R}}$ is applied to $\boldsymbol{\mathcal{F}}^H(\mu_j) \,\forall j$, s.t. the set $\boldsymbol{\mathcal{P}}$ of the reconciled forecasts of cardinality $|\boldsymbol{\mathcal{P}}|$ is generated

2 — Define the portfolio of forecast combination methods $\boldsymbol{\mathcal{D}} \equiv \{\delta_1, \ldots, \delta_D\}$

6 — All the forecast combination procedures (set $\boldsymbol{\mathcal{D}}$) are exhaustively applied to all the possible combinations of order $\{k = 1, \ldots, |\boldsymbol{\mathcal{P}}|\}$, so that the set $\boldsymbol{\mathcal{Z}}$ with cardinality $(|\boldsymbol{\mathcal{D}}| * \sum_{k=1}^{|\boldsymbol{\mathcal{P}}|} \binom{|\boldsymbol{\mathcal{P}}|}{k})$ of reconciled forecasts is built

3 — Define the Reconciliation method $\boldsymbol{\mathcal{R}}$

7 — A suitable loss function $\mathcal{L}(\cdot)$ is applied to $\boldsymbol{\mathcal{Z}}$ so that the winner forecast element $\mathbf{z}^* \in \boldsymbol{\mathcal{Z}}$ is extracted. In symbols $\mathbf{z}^* = \min \mathcal{L}(\boldsymbol{\mathcal{Z}})$

4 — The M H–step ahead base forecasts $\hat{\boldsymbol{x}}_{j,t}^H = \boldsymbol{\mathcal{F}}^H(\mu_j); \forall j \in M$ are generated

8 — The forecasting model μ^* and the combination method δ^* generating $\mathbf{z}^*$ are extracted and employed to compute the in sample estimate of the bias β^*

The novelty of the proposed method is basically captured by steps 5–7 and 8. For the sake of a more operational comprehension, step 5–7 are now discussed using the matrix notation whereas step 8 will be detailed in Subsection 4.1. Step 5 indicates that once a number M of different forecasts, generated by M models, become available for each level of the hierarchy, they are reconciled one at a time through $\boldsymbol{\mathcal{R}}$. In practice, the reconciliation function $\boldsymbol{\mathcal{R}}$, applied to the given hierarchical structure, generates a sequence of $H \times 1$ vectors of reconciled predictions $\mathbf{p}_j^H$, as below schematized.

Figure 2   Illustration of the procedure to obtain the vectors of reconciled forecasts

$$\mu_1 \; \subset \; \mathcal{M} \qquad \longrightarrow \qquad \mathcal{F}^H(\mu_1) \qquad \longrightarrow \qquad \mathcal{R}(\mathcal{F}^H(\mu_1)) \; = \; \mathbf{p}_1^H$$

$$\mu_j \; \subset \; \mathcal{M} \qquad \longrightarrow \qquad \mathcal{F}^h(\mu_j) \qquad \longrightarrow \qquad \mathcal{R}(\mathcal{F}^h(\mu_j)) \; = \; \mathbf{p}_j^h$$

$$\mu_M \; \subset \; \mathcal{M} \qquad \longrightarrow \qquad \mathcal{F}^h(\mu_M) \qquad \longrightarrow \qquad \mathcal{R}(\mathcal{F}^h(\mu_M)) \; = \; \mathbf{p}_M^h$$

Each vector $\mathbf{p}_{H,j}$ can be seen as the column of a matrix, say $\mathbf{P}_{H,M}$, containing all the M model-dependent reconcilied forecasts. In the following step 6, these M (column)-vectors of predictions are combined according to a number D of different combination methods $(\delta_1, \ldots, \delta_D)$. In essence, they are sequencially and exhaustively applied to each of the possible combinations of order $\{k = 1, \ldots, \mathrm{ncol}(\boldsymbol{P})\}$ of the column vectors of $\mathbf{P}_{H,M}$. Defining the combination (without repetition) function with C_k and setting, for instance, the combination order $k = k_0 < M$, the submatrix $\mathbf{P}_{H,M}^{k_0} = C_{k_0}(\mathbf{P}_{H,M})$ stores all the $\binom{M}{k_0}$ combinations of the forecasts $\mathbf{p}_j^H$; $j = 1, \ldots, \mathrm{ncol}(\mathbf{P}_{H,M}^{k_0})$, called $\boldsymbol{z}_j^H$, as illustrated in the Figure below.

Figure 3   Illustration of the procedure to obtain the vectors of model dependent reconcilied forecasts

$$\delta_1 \mathbf{P}_{H,M}^k \;\; = \;\; \boldsymbol{z}_1^H$$

$$\delta_j \mathbf{P}_{H,M}^k \;\; = \;\; \boldsymbol{z}_j^H$$

$$\delta_D \mathbf{P}_{H,M}^k \;\; = \;\; \boldsymbol{z}_{\mathrm{ncol}(\boldsymbol{P}_{H,M}^k)}^H$$
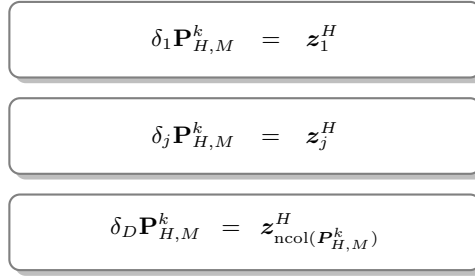
By looping over all of the combination orders $k = 1, \ldots, \mathrm{ncol}(\mathbf{P}_{H,M})$, the matrix $\boldsymbol{Z}$ containing all the possible combination of the M model dependent reconcilied forecasts is obtained. This matrix is called $\boldsymbol{Z}$ and has dimensions

$$H \times D * \mathrm{ncol}(\mathbf{P}_{H,M}). \tag{10}$$

Step 7 translate into simply applying a suitable loss function to $\boldsymbol{Z}$ (columnwise), until the final vector of predictions $\boldsymbol{z}^*$, verifying the minimum condition $\min \mathcal{L}(\mathbf{Z})$, is extracted.

4.1. The bias correction procedure (step 8)

It is well known that a perfectly unbiased forecast is a condition not frequently met in many real-life applications. Unfortunately, the unbiasedness of the forecasts reconciliation method chosen (Equation 2), depends on the unbiasedness of each and every base forecasts, as proved by Wickramasuriya, S. L., Athanasopoulos, G., and Hyndman, R. J. (2019). The proposed method can alleviate this problem as it is designed to generate a "big" competition set ($\mathcal{Z}$), made up with more "balanced" forecasts (thanks to the forecast combinations techniques applied) and thus more likely to perform better than methods generating fewer or just one forecast vector.

The bias correction of the forecast values stored in the vector $\mathbf{z}^*$ – obtained in step 7 of Figure 1 – is performed using an improved version of the simple, yet effective, procedure discussed in Spiliotis, E., Petropoulos, F., and Assimakopoulos, V. (2019). In more details, the adopted method translates into a six–step iterative procedure, designed to empirically estimate a set of in-sample tentative biases $\{\boldsymbol{\beta} \equiv \beta_1, \beta_2, \ldots, \beta_B\}$, each of them obtained according to a predefined, arbitrary set of suitable central tendency functions $\{a_1, a_2, \ldots, a_A \subset \mathcal{A}\}$, being $|\boldsymbol{\beta}| = |\mathcal{A}|$ (or, equivalently, $B = A$).

Recalling that $\boldsymbol{x}_t$ is the observed time series, let us denote with $\hat{\boldsymbol{x}}_t^* = f(\mu^*, \delta^*)$ its one-step-ahead predictions – obtained according to the optimal forecasting model and prediction combination method (see Equation 9) – and with $\boldsymbol{\varepsilon_t}|\beta_j$ the vector of residuals between these two series conditional to a given central tendency function, i.e. $\boldsymbol{\varepsilon_t}|\beta_j = \boldsymbol{x}_t - \hat{\boldsymbol{x}}_t^*|\beta_j(\alpha_j)$. In what follows, the term α_j is omitted as it is understood the dependency relationship between bias and central tendency function, i.e. bias $= \beta_j(\alpha_j)$. The set $\boldsymbol{\beta}$ is thus generated by iteratively and exhaustively applying each function in $\mathcal{A}$ to the bias adjusting equations which, once expressed in terms of residuals, read as follows:

$$^{\beta_j}\boldsymbol{\varepsilon_t} = \boldsymbol{x}_t - \hat{\boldsymbol{x}}_t^*|\beta_j; \quad j = 1, 2, \ldots, B, \tag{11}$$

$$^{\beta_j}\boldsymbol{\eta_t} = \frac{\boldsymbol{x}_t}{\hat{\boldsymbol{x}}_t^*|\beta_j}; \quad j = 1, 2, \ldots, B. \tag{12}$$

Equations 11 – 12 differ only for that the one-step ahead predictions are respectively adjusted additively and multiplicatively. Finally, by applying a suitable loss function, $\mathcal{E}(\cdot)$ to each of the vectors $^{\beta_j}\boldsymbol{\varepsilon_t}$ and $^{\beta_j}\boldsymbol{\eta_t}$, the optimal bias estimation is its minimzer, i.e.

$$\beta^* = \min_{\mathcal{E}}(^{\beta_j}\boldsymbol{\varepsilon_t}; ^{\beta_j}\boldsymbol{\eta}); \quad j = 1, 2, \ldots, B. \tag{13}$$

Once the final bias is computed, it can be readily applied in a forward looking fashion, i.e.

$$^a\boldsymbol{y_{t,H}} = \mathbf{z}_{,H}^* + \beta^*(\boldsymbol{x}_t - \mathbf{z}_t^*) \tag{14}$$

14

or

$$^{m}\boldsymbol{y_{t,H}} = \mathbf{z}^{*}_{,H} * \beta^{*} \frac{\boldsymbol{x}_t}{\mathbf{z}^{*}_t},$$ (15)

according to whether the winner central tendency function is applied in an additive (Equation 14) or multiplicative fashion (Equation 15). The generalized notations for the first term in Equations 4.1 and 4.1 is

$$^{u}\boldsymbol{y_{t,H}},$$ (16)

which represent the final predictor. In such an approach, the future and the past are assumed to be affected by the same amount of bias. Such an assumption, under stationaity of the observed time series and a "sufficient" sample size, might not be considered unreasonable.

In the case of $\{\boldsymbol{\beta} \equiv \beta_1\}$ with β_1 the mean function, Equations 14–15 respectively are as follows:

$$^{a}\boldsymbol{y_{t,H}} = \mathbf{z}^{*}_{t,H} + \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_t - \mathbf{z}^{*}_t)$$

or

$$^{m}\boldsymbol{y_{t,H}} = \mathbf{z}^{*}_{t,H} * \frac{1}{T}\sum_{t=1}^{T}\frac{\boldsymbol{x}_t}{\mathbf{z}^{*}_t},$$

and thus equivalent (saved for the notation) to the procedure discussed in Spiliotis, E., Petropoulos, F., and Assimakopoulos, V. (2019) (Equations 3–4), page 21.

In the empirical application of Section 6, the central tendency functions applied are: mean, median and root mean square, respectively denoted by the lowercase Greek letters π, τ and ξ. Their mathematical representations are as follows:

$$\pi(\cdot) = \frac{1}{T}\sum_{t=1}^{T}x_t, \qquad \tau(\cdot) = \frac{1}{2}x_{(\lfloor\ T+1)/2\rfloor} + x_{\lceil(T+1)/2\rceil}, \qquad \xi(\cdot) = \frac{1}{T}\sqrt{\sum_{t=1}^{T}x_t^2}.$$ (17)

In the case of the median ($\tau(\cdot)$) x is an ordered list of T values, and the symbols $\lfloor\cdot\rfloor$ $\lceil\cdot\rceil$ denote the floor and ceiling functions, respectively.

5. Justification of the method

The effectiveness of the proposed method is, in general, conditioned to the choices of the statistical prediction models and the forecast combination techniques included in the sets $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{D}}$, as well as to the selection of the most suitable central tendency functions (the set $\boldsymbol{\mathcal{A}}$). A careful building of those sets

15

(our multidimensional search space), is a prerequisite for the proposed method to properly perform. Its final dimensions are as in Expression 10 plus twice the number of central density functions considered, used in both additive and multiplicative fashion, i.e.

$$H \times [2 * |\mathcal{A}| + D * \text{ncol}(\mathbf{P}_{H,M})] \, .$$

The point of strength of the method is thus related to the availability of a potentially large number of multiple choices, all of them derived in a multiple combination set-up (in terms of statistical prediction model, forecast combination and bias correction), so that the selected forecast vector are the minimizer of a bi-dimensional loss function $(\mathcal{L}, \mathcal{E})$. In addition, the method is very flexible, as it can work with all the methods deemed suitable for the problem at hand, being the only limit the computational time. This is certainly an issue, which, however, can be easily circumvented thanks to the structure of the method itself, which is naturally prone to be parallelized.

According to the bias–variance decomposition approach, the mean square error (MSE) can be decomposed into a bias β and a variance (V) terms, i.e.

$$MSE = \beta^2 + V \tag{18}$$

In what follows, the advantages related to the proposed method will be illustrated in terms of Equation 18. Firstly, it is noted how, in general, simpler models tend to produce large biases and small variances whereas complex models behave in the opposite way. The proposed method is designed to overcome such an issue not only because it can employ a combination of several models with different levels of complexity but also because it selects the "best" combination technique (stored in the set $\mathcal{D}$) according to the data set under investigation. This last feature is clearly a plus, since there is not such combination techniques able to perform optimally in any circumstances. For example, Palm, F. C., and Zellner, A. (1992) found that there are cases where a simple average combination may be more robust than weighted average combinations. Therefore, by iteratively testing many different techniques, one is more likely to find the most suitable (if not the "optimal") one.

Bias-wise, the advantages of this method are related to its self-balancing and self-adjusting features, the former being induced by the bias compensation phenomenon, more likely to occur in a multi-model set up, whereas the latter relies on the bias correction procedure, given in Section 4.1. In particular, its effectiveness in bias reduction is related to the fact that the self-adjustment procedure uses a vector of forecast which, by design, has already been controlled for bias. To see this, let us express the generic forecast combination $\tilde{\boldsymbol{x}}_t^H$ as

$$\tilde{\boldsymbol{x}}_t^H = \sum_{i=1}^{M} w_i \boldsymbol{\mathcal{F}}^H(\mu_i),$$

16

with $w_i = f(\tilde{\delta})$ the combination weights generated by the combination method $\tilde{\delta}$ and $\mathcal{F}^H(\mu_j); j = 1, \ldots, M$ are base forecasts generated by the M statistical models entertained (see Figure 1 step 1 and 4). Assuming $0 \leq w_j \leq 1$, $\sum_{i=1}^{M} w_i = 1$ and the vector of "future" observations of length H to be known, the total amount of bias of the combined forecast is given by

$$
\begin{aligned}
\beta &= \mathbf{E}(\tilde{\boldsymbol{x}}_t^H - \boldsymbol{x}_t^H) \\
&= \sum_{i=1}^{M} w_i [\mathbf{E}\tilde{\boldsymbol{x}}_{t,i}^H - \mathbf{E}\boldsymbol{x}_t^H] \\
&= \sum_{i=1}^{M} w_i \beta_i,
\end{aligned}
\tag{19}
$$

where the subscript i is used to refer to a specific model, in terms of generated bias (β_i) and forecast ($\tilde{\boldsymbol{x}}_{t,i}^H$). The right term of Equation 19 shows that the bias of the forecast combination is the weighted average of the biases of the base forecasts and thus, provided that their magnitude is comparable, one can reasonably expect an overall bias reduction due to cancellation effects. Such a phenomenon is not rare, since – in general – it is not common for all the biases to show the same sign. Since the bias-correcting method – discussed in Section 4.1 – is applied on an already optimally combined vector of forecasts $\mathbf{z}_t^* = \boldsymbol{x}_t(\delta^*, \mu^*)$, a less pronounced bias can be expected in the final predictions, given by

$$
{}^u\boldsymbol{y_{t,H}} = \boldsymbol{z}_{t,H}^* + \boldsymbol{\beta}^*(\boldsymbol{x}_t - \boldsymbol{z}_t^*),
\tag{20}
$$

where β^* is as in Equation 13.

The proposed method can also help keep low the variance of the forecasts in an amount inversely proportional to the correlation coefficients computed between the competing forecasts and proportional to the reduction in the standard errors induced by the reconciliation procedure adopted $\mathcal{R}$. To see this, denoting with σ_i^2 the variance of the individual forecast i and with $\gamma_{i,j}$ the correlation coefficient computed on a generic pair (i, j) of forecasts, we use the following inequality (derived in Atiya, A. F. (2020)), i.e.

$$
V \leq \sum_{i=1}^{N} w_i \sigma_i^2 - 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} w_i w_j (1 - \gamma_{i,j}) \sigma_i \sigma_j,
\tag{21}
$$

where V is the variance of the combination of forecasts. Inequality 21 states that V tends to be considerably less than the average of the individual forecasts in an amount depending on $\frac{1}{\gamma_{i,j}}$, meaning that less correlated forecasts are beneficial in terms of variance reduction. Such a situation is more likely to occur in procedures which, as the one proposed, grant a "sufficient" number of predictions. However, there is another point in favor of the proposed method on this matter. In fact, recalling that the different combinations of the forecasts are performed on already reconciled vectors of predictions $\mathbf{z}_t^H$ – according to

17

the adopted procedure $\mathcal{R}$ – by virtue of Equation 4 their variances obey to the following inequality:

$$Var(\mathbf{z}_t^H) \ < Var(\boldsymbol{x}_t^H); \forall \mathbf{z} \in \boldsymbol{\mathcal{Z}}. \tag{22}$$

Therefore, the overall level of variance in Equation 21 decreases of an amount inversely proportional to $Var(\mathbf{z}_t^H)$.

6. Empirical study

In this section the goodness of the proposed method will be evaluated using the official time series related to the number of people tested positive to the SARS-CoV-2 in each of the Italian regions, between February 24th 2020 and October 7th 2020. The whole data set – issued by the Italian National Institute of Health – are publicly and freely available at the web address *https://github.com/pcm-dpc/COVID-19/tree/master/dati-regioni*. The data, sampled at a daily frequency, are stored in a matrix called $\mathcal{O}$ (see Table 1) of dimension 227×21, where 21 are the Italian regions. From a strictly administrative point of view, the number of the Italian regions amounts to 20, however, for one of them, called Trentino Alto Adige, the data are split according to its two main provinces: Trento and Bolzano. As reported in the same Table 1, the proposed procedure is trained on a portion of the data matrix, called $\mathcal{O}^{train}$, of dimensions 197×21 and time span February 24th – September 7th, whereas the test part is carried out on a set, called $\mathcal{O}^{test}$, whose dimensions are $H = 30 \times 21$ (the time span is from September 8th to October 7th). Finally, 30 days ahead "real-life" estimates – in the sense that they are related to future values which are unknown at the time of their computations – for the time window October 8th – November 7th will be stored in the matrix $\mathcal{O}^{fore}$. Since the proposed procedure combines a num-

TABLE 1
The employed data set and its portions defined according to the different purposes served

Symbol	Start date	End Date	Sample size
$\mathcal{O}$	February 24th	October 7th	227
$\mathcal{O}^{train}$	February 24th	September 7th	197
$\mathcal{O}^{test}$	September 8th	October 7th	30
$\mathcal{O}^{fore}$	October 8th	November 7th	30

ber of models (μ's), combination methods (δ's) and central tendency functions ($\mathcal{E}'s$), for each of those, conventional symbols are respectively given in Tables 2, 3 and 4. Consistently with the convention introduced in Equation 16, in Table 4 the superscript u is used to indicate the type of bias considered, i.e. additive ($u = a$) or multiplicative ($u = m$). Finally, to efficiently keep track of the outcomes of the method, in Table 5 the whole set of model combinations employed – respectively of class $k = 4, 3, 2$ – are provided. Since we have four different combination methods, each of the 11 model combination (reported in Table 5) are performed four times, which yields a total of 44 method-dependent combinations. One of them, for example, is the forecasts combination generated by

18

combining an ARFIMA and an ETS models using the method Ordinary Least Squares. This information is conveniently conveyed by the symbol *OLS–AE*.

TABLE 2
Symbols employed to identify the statistical models

Model (μ)	Symbol
ARFIMA	A
ETS	E
THETA	T
ARIMA	B

TABLE 3
Symbols employed to identify the central tendency functions

Method (δ)	Symbol
Ordinary Least Square	OLS
Least Absolute Deviation	LAD
Newbold and Granger	NG
Simple Average	SA

TABLE 4
Symbols employed to identify the statistical models

Central tendency function	Symbol
Mean	$^{u}\pi$
Median	$^{u}\tau$
RMS	$^{u}\xi$

TABLE 5
Combinations of models of class $K = 4, 3, 2$ attempted for each of the four model-dependent reconcilied forecasts

Number	K	Model combination
1	4	A–E–T–B
2		A–B–E
3	3	A–B–T
4		A–E–T
5		B–E–T
6		A–B
7		A–E
8	2	A–T
9		B–E
10		B–T
11		E–T

Recalling that with $^{u}\boldsymbol{y_{t,H}}$ the final predictions yielded by the proposed method are denoted (see Equation 16), the loss function employed ($\mathcal{L}$) is the Root Mean Square Error (RMSE), given by $\sqrt{\frac{1}{T}\sum_{h=1}^{30}(\boldsymbol{x_{t,H}} - \, ^{u}\boldsymbol{y_{t,H}})^2}$. The same function is adopted in–sample to select the best 3–tuple ($\mu^*, \delta^*, \mathcal{E}^*$), i.e. $\sqrt{\frac{1}{T}\sum_{t=1}^{227}(\boldsymbol{x_t} - \, ^{u}\boldsymbol{y_t})^2}$ and to evaluate the method's performance in the test set $\mathcal{O}^{test}$. Finally, the out of sample estimate of the bias, in the sequel denoted by the symbol β^{out}, has been computed on the set $\mathcal{O}^{test}$, using the formula

19

$$\frac{1}{T} \sum_{t=1}^{T} \left(\boldsymbol{x}_t \quad - \quad {}^u\boldsymbol{y}_{t,\boldsymbol{H}} \right).$$

6.1. Performances of the method

The performance of the method are summarized in Appendix A and in Table 6. In particular, in Appendix A the observations belonging to the test set $\mathcal{O}^{test}$ (black line) and the related estimates ${}^u\boldsymbol{y}_{t,\boldsymbol{H}}$ (red line) are depicted. The first three columns of Table 6 respectively indicate the name of the Italian regions, the winner combination (μ^*, δ^*) and the related RMSE values $(\mathcal{L}^*)$. The best central tendency function $(\mathcal{E}^*)$, the selected bias corrector (β^*) and the estimated out-of-sample bias (β^{out}) are given in columns four, five, six. In the last two columns the $RMSE$ values $(\mathcal{L})$ relative to each of the forecasting models (reported in Table 2) taken separately, are recorded. The accuracy of the proposed method seems to be very good as, in almost all the cases, the winner combinations deliver better predictions than all the statistical models singularly considered and, in many cases, outperform them. This is the case, for example, of the Campania and Calabria regions. Here (see Table 6), the recorded $RMSE$ is respectively equal to 34.7 and 63.5, far below the values obtained using the best statistical models, i.e. $Theta$ and ETS, which respectively scored $\mathcal{L} = 143.6$ and 265.1. In addition, the proposed procedure shows always from very low to negligible in-sample amounts of bias. The most selected central tendency functions is ${}^a\tau$ and ${}^a\pi$, whereas it is noted that the RMS function $({}^u\xi)$ has never been chosen. Regarding the out-of-sample bias, if on one hand, as expected, it is always $\beta^{out} > \beta^*$, on the other hand it can be said that the magnitude of the values assumed by β^{out} can be deemed acceptable. This is especially true if one considers the length of the prediction window ($H = 30$ days) compared with the available data set. In particular, six regions (Valle d'Aosta, Trento, Molise, Lazio, Abruzzo, Calabria) show interesting values for the out of sample bias, being $\beta^{out} < 10$. The worse performances of the method refers to the region of Sardegna, where the consistent underestimations of the true values lead to a recorded bias of around 504. This fact can be explained by looking at the irregular, bumpy shape of this time series, reported in Appendix 2 (Sardegna), which might have introduced distortions in the model estimators. In Appendix B, the graphical results of a pure out-of-sample application of each of the winner combinations are reported. In more details, the region-specific winner combinations (μ^*, δ^*) are applied to the whole set $\mathcal{O}$, so that the $H = 30$ days–ahead resulting forecasts – stored in the set $\mathcal{O}^{fore}$ – are the pure forecasts for the period October 8th – November 7th. In Appendix B, the regional time series in $\mathcal{O}$ (true observations) are plotted in black whereas the predictions are in red color. The analysis of these Figures suggest a slower acceleration in the growth of positive cases in some of the north regions (e.g. Lombardia, Trento, Liguria) whereas the center and south regions might face a strong increase of positives. This seems likely to happen in the Campania, Basilicata and Molise regions. The number of positive for Italy, predicted for the end of the pure forecast period (November 7th), is about of 115,854.

Table 6: *Performances of the method for each of the Italian regions. Outcomes of the winner models and of the single statistical models. See text for details*

Region	Winner combination	$\mathcal{L}^*$	$\mathcal{E}^*$	β^*	β^{out}	Single models	$\mathcal{L}$
Piemonte	$LAD - BET$	135.6	$^a\tau$	≈ 0	54.09	$ARFIMA$	1050.8
						ETS	1065.8
						θ	1271.8
						$ARIMA$	683.6
Val d'Aosta	$SA - BT$	8.6	$^a\tau$	≈ 0	-2.25	$ARFIMA$	60.7
						ETS	177.5
						θ	30.5
						$ARIMA$	29.5
Lombardia	$SA - ET$	172.6	$^m\tau$	1.0	70.83	$ARFIMA$	3184.0
						ETS	911.7
						θ	966.1
						$ARIMA$	1713.8
Bolzano	$OLS - ET$	223.2	$^a\pi$	≈ 0	190.9	$ARFIMA$	468.9
						ETS	227.6
						θ	271.2
						$ARIMA$	248.8
Trento	$NG - BE$	29.4	$^a\tau$	-.13	8.95	$ARFIMA$	113.7
						ETS	30.0
						θ	238.2
						$ARIMA$	211.5
Veneto	$NG - BE$	58.5	$^a\pi$	0.55	49.01	$ARFIMA$	620.1
						ETS	95.5
						θ	259.9
						$ARIMA$	49.6
Friuli Venezia Giulia	$SA - BE$	65.9	$^a\pi$	0.70	18.08	$ARFIMA$	475.4
						ETS	140.6
						θ	763.5
						$ARIMA$	155.6
Liguria	$LAD - ET$	124.2	$^a\tau$	≈ 0	-28.13	$ARFIMA$	4123.8
						ETS	126.8
						θ	1057.7
						$ARIMA$	241.9
Emilia Romagna	$LAD - BT$	296.7	$^a\tau$	≈ 0	-160.01	$ARFIMA$	2203.4
						ETS	343.5
						θ	724.5

Continued on next page

21

Region	model	$\mathcal{L}^*$	$\mathcal{E}^*$	β^*	α^*	Single models	$\mathcal{L}$
						$ARIMA$	274.9
Toscana	$LAD-AB$	259.6	$^a\tau$	≈ 0	149.8	$ARFIMA$	3417.7
						ETS	1202.4
						θ	2485.9
						$ARIMA$	310.0
Umbria	$OLS-AB$	133.3	$^a\pi$	≈ 0	104.54	$ARFIMA$	701.9
						ETS	254.5
						θ	344.4
						$ARIMA$	168.1
Marche	$SA-BET$	211.0	$^m\tau$	1.01	136.65	$ARFIMA$	2191.6
						ETS	320
						θ	1185.7
						$ARIMA$	355.5
Lazio	$SA-ET$	45.9	$^a\tau$	0.13	7.43	$ARFIMA$	522.2
						ETS	153.5
						θ	180.7
						$ARIMA$	64.7
Abruzzo	$LAD-BET$	40.0	$^a\tau$	≈ 0	6.97	$ARFIMA$	740.2
						ETS	70.3
						θ	294.6
						$ARIMA$	40.7
Molise	$SA-AB$	39.2	$^a\pi$	$-.05$	-5.8	$ARFIMA$	250.9
						ETS	1229.0
						θ	147.0
						$ARIMA$	265.1
Campania	$SA-BT$	34.7	$^a\tau$	.05	-30.52	$ARFIMA$	480.7
						ETS	359.8
						θ	143.6
						$ARIMA$	201.3
Puglia	$LAD-BET$	419.4	$^a\tau$	≈ 0	-249.58	$ARFIMA$	1991.5
						ETS	680.6
						θ	2507.5
						$ARIMA$	674.1
Basilicata	$LAD-AB$	75.2	$^a\tau$	≈ 0	-54.68	$ARFIMA$	117.5
						ETS	1453.5
						θ	37.0
						$ARIMA$	417.7

Region	model	$\mathcal{L}^*$	$\mathcal{E}^*$	β^*	α^*	Single models	$\mathcal{L}$
Calabria	$OLS - AETB$	63.5	$^a\pi$	≈ 0	4.13	$ARFIMA$	2124.9
						ETS	265.1
						θ	1109.9
						$ARIMA$	302.1
Sicilia	$SA - BET$	70.7	$^m\pi$	0.39	57.68	$ARFIMA$	1436.8
						ETS	333.0
						θ	715.7
						$ARIMA$	195.2
Sardegna	$NG - AB$	581.9	$^m\pi$	-0.05	504.75	$ARFIMA$	983.7
						ETS	998.9
						θ	1204.6
						$ARIMA$	605.4

7. Conclusions and the future directions

The present paper provides sufficient evidences that reconciliation serves the double purpose of generating coherent forecast with improved accuracy, under a multidimensional optimization constraint. The proposed method is designed to handle the increased amount of uncertainty surrounding the forecasting, as one carries out the prediction exercise at a progressively more disaggregated levels. The novelty of this procedure is the application of the forecast combination techniques to already optimally combined forecasts, generated by different prediction models. In this regard, the proposed procedure can be improved by increasing the portfolio of the statistical models entertained under a suitable program architecture. The second line of future research refers to the bias estimation, whose uncertainty might estimated using a suitable resampling scheme.

8. Data availability

The data that support the findings of this study are openly available in the section "COVID-19/dati-regioni/" at https://github.com/pcm-dpc/COVID-19/tree/master/dati-regioni.

9. Disclaimer

The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics.

References

Assimakopoulos, V., and Nikolopoulos, K. (2000). *The theta model: a decomposition approach to forecasting. International journal of forecasting International journal of forecasting* **16**(4), 521-530.

Athanasopoulos, G., Ahmed, R. A., and Hyndman, R. J. (2011). Optimal combination forecasts for hierarchical time series *Computational statistics and data analysis* **55**(9), 2579-2589.

Atiya, A. F. (2020). *Why does forecast combination work so well?. International Journal of Forecasting* **36**(1), 197-200

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

Barrow, D. K., and Kourentzes, N. (2016). Distributions of forecasting errors of forecast combinations: implications for inventory management. *International Journal of Production Economics* **177**, 24-33.

Bates, J. M., and Granger, C. W. (1969). The combination of forecasts *Journal of the Operational Research Society* **20**(4), 451-468.

Box, G. E., Jenkins, G. M., and Reinsel, G. (1970). *Time series analysis: forecasting and control Holden-day San Francisco* 1970

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: an integral part of inference *Biometrics* 603-618.

Chambers, J. C., Mullick, S. K., and Smith, D. D. (1971). How to choose the right forecasting technique *Harvard Business Rev.* **49**, 45-74.

Chatfield, C. (1996). Model uncertainty and forecast accuracy *Journal of Forecasting* **15**(7), 495-508.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting* **5**(4), 559-583.

Clements, M. P., and Hendry, D. F. (2002). Modelling methodology and forecast failure *The Econometrics Journal* **5**(2), 319-344.

Di Fonzo, T., and Girolimetto, D. (2021). *Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives. International Journal of Forecasting*

Fair, R. C., and Shiller, R. J. (1990). Comparing information in forecasts from econometric models. *The American Economic Review* **3**75-389.

Gardner Jr, E. S. (2006). *Exponential smoothing: The state of the art—Part II International journal of forecasting* **22**(4), 637-666.

Dielman, Terry E (2005). *Least absolute value regression: recent contributions Journal of statistical computation and simulation* **75**(4), 263–286.

Granger, C. W., and Joyeux, R. (1980). *An introduction to long-memory time series models and fractional differencing. Journal of time series analysis* **1**(1), 15-29.

Granger, C. W. (1980). *Long memory relationships and the aggregation of dynamic models Journal of econometrics* **14**(2), 227-238.

Granger, C. W., and Ramanathan, R. (1984). *Improved methods of combining forecasts. Journal of forecasting Journal of forecasting* **3**(2) 197-204.

Hsiao, C., and Wan, S. K. (2014). *Is there an optimal forecast combination?. Journal of Econometrics* **178**, 294-309.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series *Computational statistics and data analysis* **55**(9), 2579-2589.

Hyndman, R. J., Lee, A. J., and Wang, E. (2016). *Fast computation of reconciled forecasts for hierarchical and grouped time series Computational statistics and data analysis* **97** 16-32.

Hubrich, K. (2005). Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy?. *International Journal of Forecasting* **21**(1), 119-136.

Schwarzkopf, A. B., Tersine, R. J., and Morris, J. S. (2006). Top-down and bottom-up forecasting in S&OP. *International Journal of Forecasting The Journal of Business Forecasting* **25**(2), 14-16.

Makridakis, S. (1989). Why combining works?. International Jour-

NAL OF FORECASTING *International Journal of Forecasting* **5**(4), 601-603.

MANCUSO, A. C. B., AND WERNER, L. (2013). REVIEW OF COMBINING FORECASTS APPROACHES. *Independent journal of management and production* **4**(1), 248-277.

MAKRIDAKIS, S., SPILIOTIS, E., AND ASSIMAKOPOULOS, V. (2020). THE M4 COMPETITION: 100,000 TIME SERIES AND 61 FORECASTING METHODS. *International Journal of Forecasting* **36**(1), 54-74.

MAKRIDAKIS, S., AND HIBON, M. (2000). THE M3-COMPETITION: RESULTS, CONCLUSIONS AND IMPLICATIONS. *International Journal of Forecasting* **16**(4), 451-476.

MARCELLINO, M., STOCK, J. H., AND WATSON, M. W. (2003). MACROECONOMIC FORECASTING IN THE EURO AREA: COUNTRY SPECIFIC VERSUS AREA-WIDE INFORMATION. *European Economic Review* **47**(1), 1-18.

NEWBOLD, P., AND GRANGER, C. W. (1974). *Experience with forecasting univariate time series and the combination of forecasts. Journal of the Royal Statistical Society: Series A (General)* **137**(2), 131-146.

PALM, F. C., AND ZELLNER, A. (2016). *To combine or not to combine? Issues of combining forecasts. Journal of Forecasting* **11**(8) 687-701.

SESSIONS, D. N., AND CHATTERJEE, S. (1989). THE COMBINING OF FORECASTS USING RECURSIVE TECHNIQUES WITH NON-STATIONARY WEIGHTS *Journal of Forecasting* **8**(3), 239-251.

SCHWARZKOPF, A. B., TERSINE, R. J., AND MORRIS, J. S. (1988). TOP-DOWN VERSUS BOTTOM-UP FORECASTING STRATEGIES. *International Journal of Forecasting The International Journal Of Production Research* **26**(11), 1833-1843.

SPILIOTIS, E., PETROPOULOS, F., AND ASSIMAKOPOULOS, V. (2019). IMPROVING THE FORECASTING PERFORMANCE OF TEMPORAL HIERARCHIES *Plos one* **14**(10) .

STOCK, J. H., AND WATSON, M. W (2001). A COMPARISON OF LINEAR AND NONLINEAR UNIVARIATE MODELS FOR FORECASTING MACROECONOMIC TIME SERIES in *Hengle, R.F. and White, H. (eds.)* Festschrift in honor of Clive Granger, 1-44.

STOCK, J. H., AND WATSON, M. W. (2004). COMBINATION FORECASTS OF OUTPUT GROWTH IN A SEVEN-COUNTRY DATA SET. *International Journal of Forecasting* **23**(6), 405-430.

TAIEB, S. B., TAYLOR, J. W., ANDHYNDMAN, R. J. (2017). *Coherent probabilistic forecasts for hierarchical time series.* In International Conference on Machine Learning. 3348–3357

VAN ERVEN, T., AND CUGLIARI, J. (2015). *Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts*

WICKRAMASURIYA, S. L., ATHANASOPOULOS, G., AND HYNDMAN, R. J. (2019). OPTIMAL FORECAST RECONCILIATION FOR HIERARCHICAL AND GROUPED TIME SERIES THROUGH TRACE MINIMIZATION *Journal of the American Statistical Association* **114**(526), 804-819.

WICKRAMASURIYA, S. L., ATHANASOPOULOS, G., AND HYNDMAN, R. J. (2019). *Optimal forecast reconciliation for hierarchical and grouped time*
26

series through trace minimization. *Journal of the American Statistical Association* **114(526)**, 804-819.