MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://monash.edu/business/ebs/research/publications

Optimal forecast reconciliation with time series selection

Xiaoqian Wang, Rob J Hyndman, Shanika L Wickramasuriya

September 2023

Working Paper no/yr

AACSB ACCREDITED EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

Optimal forecast reconciliation with time series selection

Xiaoqian Wang
Monash University, VIC 3800, Australia
Email: xiaoqian.wang@monash.edu
Corresponding author

Rob J Hyndman
Monash University, VIC 3800, Australia
Email: rob.hyndman@monash.edu

Shanika L Wickramasuriya
Monash University, VIC 3145, Australia

21 September 2023

Optimal forecast reconciliation with time series selection

Abstract

Abstract

Keywords: Keyword 1, Keyword 2

1 Introduction

Hierarchical time series and forecast reconciliation.

Single-level approaches, least squares-based reconciliation approaches, geometric intuition, other extensions with constraints.

However… Two issues. The choice of W can have significant effect on the quality of the reconciled forecasts. Some time series perform poorly.

In this paper, our focus will be on…

The remainder of the paper is structured as follows.

2 Preliminaries

2.1 Notation

We denote the set $\{1, \ldots, k\}$ by $[k]$ for any non-negative integer k. A *hierarchical time series* can be considered as an n-dimensional multivariate time series, $\{\boldsymbol{y}_t, t \in [T]\}$, that adheres to known linear constraints. Let $\boldsymbol{y}_t \in \mathbb{R}^n$ be a vector comprising observations of all time series in the hierarchy at time t, and $\boldsymbol{b}_t \in \mathbb{R}^{n_b}$ be a vector comprising observations of all bottom-level time series at time t. The full hierarchy at time t can be written as

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t,$$

where $\boldsymbol{S}$ is an $n \times n_b$ *summing matrix* that shows aggregation constraints present in the structure. We can write the summing matrix as $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{I}_{n_b} \end{bmatrix}$, where $\boldsymbol{A}$ is an $n_a \times n_b$ *aggregation matrix* with $n = n_a + n_b$, and $\boldsymbol{I}_{n_b}$ is an n_b-dimensional identity matrix.

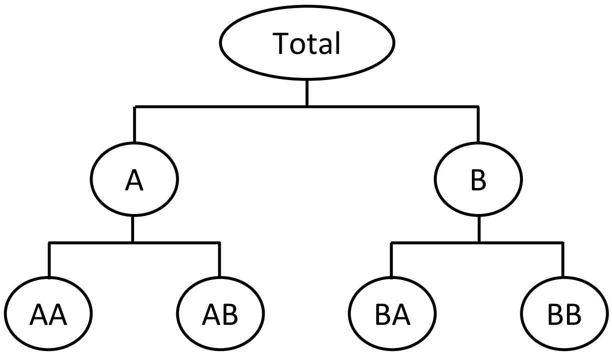

Figure 1: *An example of a two-level hierarchical time series.*

To clarify these notations, consider the example of the hierarchy in Figure 1. For this two-level hierarchy, $n = 7$, $n_b = 4$, $n_a = 3$, $\boldsymbol{y}_t = [y_{\text{Total},t}, y_{\text{A},t}, y_{\text{B},t}, y_{\text{AA},t}, y_{\text{AB},t}, y_{\text{BA},t}, y_{\text{BB},t}]'$, $\boldsymbol{b}_t = [y_{\text{AA},t}, y_{\text{AB},t}, y_{\text{BA},t}, y_{\text{BB},t}]'$, and

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ & \boldsymbol{I}_4 & & \end{bmatrix}.$$

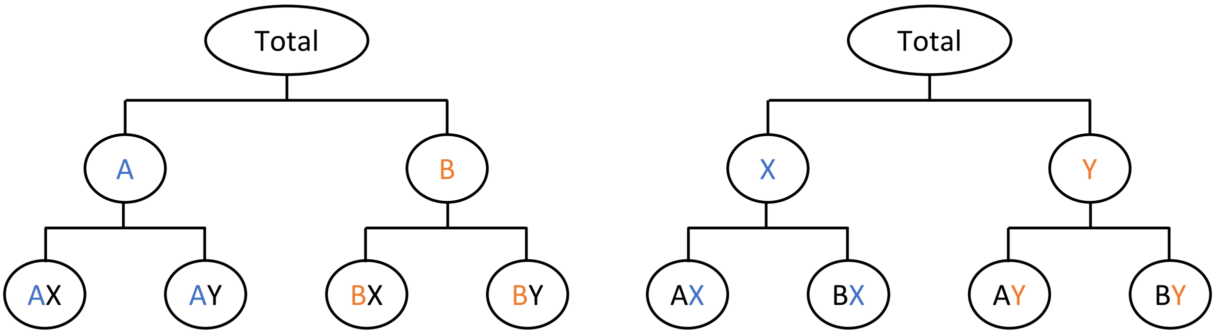

Figure 2: *An example of a two level grouped time series.*

When data structure does not naturally disaggregate in a unique hierarchical manner, we can combine these hierarchical structures to form a *grouped time series*. Thus, grouped time series can also be considered as hierarchical time series with more than one grouping structure. Figure 2 shows an example of a two level grouped time series with two alternative aggregation structures. For this example, $n = 9$, $n_b = 4$, $n_a = 5$, $\boldsymbol{y}_t = [y_{\text{Total},t}, y_{\text{A},t}, y_{\text{B},t}, y_{\text{X},t}, y_{\text{Y},t}, y_{\text{AX},t}, y_{\text{AY},t}, y_{\text{BX},t}, y_{\text{BY},t}]'$, $\boldsymbol{b}_t = [y_{\text{AX},t}, y_{\text{AY},t}, y_{\text{BX},t}, y_{\text{BY},t}]'$, and

$$
\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ & \boldsymbol{I}_4 & & \end{bmatrix}.
$$

2.2 Linear forecast reconciliation

Let $\hat{\boldsymbol{y}}_{T+h|T} \in \mathbb{R}^n$ be a vector of h-step-ahead *base forecasts* for all time series in the hierarchy, given observations up to time T, and stacked in the same order as $\boldsymbol{y}_t$. We can use any method to generate these forecasts, but In general they will not add up especially when we forecast each series independently.

When forecasting hierarchical time series, we expect the forecasts to be *coherent* (i.e., aggregation constraints are satisfied). Let $\tilde{\boldsymbol{y}}_{T+h|T} \in \mathbb{R}^n$ denote a vector of h-step-ahead *reconciled forecasts* which are coherent by construction, ψ a mapping that reconciles base forecasts, $\hat{\boldsymbol{y}}_{T+h|T}$. Then we have *forecast reconciliation* $\tilde{\boldsymbol{y}}_{T+h|T} = \psi(\hat{\boldsymbol{y}}_{T+h|T})$, which is essentially a post-processing method. In this paper, we focus on linear forecast reconciliation given by

$$
\tilde{\boldsymbol{y}}_{T+h|T} = \boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{y}}_{T+h|T},
$$

where

- $\boldsymbol{G}_h$ is an $n_b \times n$ weighting matrix that maps the base forecasts into the bottom level. In other words, it combines all base forecasts to form reconciled forecasts for bottom-level series.
- $\boldsymbol{S}$ is an $n \times n_b$ summing matrix that sums up bottom-level reconciled forecasts to produce coherent forecasts of all levels. It identifies the linear constraints involved in the hierarchy.

2.2.1 Minimum trace reconciliation

Let the h-step-ahead *base forecast errors* be defined as $\hat{\boldsymbol{e}}_{T+h|T} = \boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h|T}$, and the h-step-ahead *reconciled forecast errors* be defined as $\tilde{\boldsymbol{e}}_{T+h|T} = \boldsymbol{y}_{T+h} - \tilde{\boldsymbol{y}}_{T+h|T}$. Wickramasuriya, Athanasopoulos & Hyndman (2019) formulated a linear reconciliation problem as minimizing the trace (MinT) of the h-step-ahead covariance matrix of the reconciled forecast errors, $\text{Var}(\tilde{\boldsymbol{e}}_{T+h|T})$.

Under the assumption of unbiasedness, the unique solution of the minimization problem is given by

$$G_h = \left(S' W_h^{-1} S\right)^{-1} S' W_h^{-1}, \tag{1}$$

where W_h is the positive definite covariance matrix of the h-step-ahead base forecast errors, $\mathrm{Var}(\hat{e}_{T+h|T})$.

The trace minimization problem can be reformulated as a least squares problem with linear constraints given by

$$\begin{aligned} \min_{\tilde{y}_{T+h|T}} \quad & \frac{1}{2}(\hat{y}_{T+h|T} - \tilde{y}_{T+h|T})' W_h^{-1} (\hat{y}_{T+h|T} - \tilde{y}_{T+h|T}) \\ \text{s.t.} \quad & \tilde{y}_{T+h|T} = S\tilde{b}_{T+h|T}, \end{aligned}$$

where $\tilde{b}_{T+h|T} \in \mathbb{R}^{n_b}$ is the vector comprising h-step-ahead bottom-level reconciled forecasts, made at time T. Focusing on W_h, the intuitive behind the MinT reconciliation is that the larger the estimated variance of the base forecast errors, the larger the range of adjustments permitted for forecast reconciliation.

It's challenging to estimate W_h, especially for $h > 1$. Assuming that $W_h = k_h W_1$, $\forall h$, where $k_h > 0$, the MinT solution of G does not change with the forecast horizon. Hence, we will drop the subscript h for the ease of exposition. The most popularly used candidate estimators for W in the forecast reconciliation literature are listed as follows.

1. $W_{\mathrm{OLS}} = I$ is the *OLS estimator* proposed by Hyndman et al. (2011), assuming that the base forecast errors are uncorrelated and equivariant.
2. $W_{\mathrm{WLSs}} = \mathrm{diag}(S\mathbf{1})$ is the *WLS estimator applying structural scaling* proposed by Athanasopoulos et al. (2017). This estimator depends only on the aggregation structure of the hierarchy. It assumes that the variance of each bottom-level base forecast error is equivalent and uncorrelated between nodes.
3. $W_{\mathrm{WLSv}} = \mathrm{diag}(\hat{W}_1)$ is the *WLS estimator applying variance scaling* proposed by Hyndman, Lee & Wang (2016), where $\hat{W}_1$ denotes the unbiased covariance estimator based on the in-sample one-step-ahead base forecast errors (i.e., residuals).
4. $W_{\mathrm{MinT}} = \hat{W}_1$ is referred to as the *MinT estimator* based on the sample covariance matrix proposed by Wickramasuriya, Athanasopoulos & Hyndman (2019).

5. $\boldsymbol{W}_{\text{MinTs}} = \lambda \operatorname{diag}(\hat{\boldsymbol{W}}_1) + (1-\lambda)\hat{\boldsymbol{W}}_1$ is the *MinT shrinkage estimator* suggested by Wickramasuriya, Athanasopoulos & Hyndman (2019), in which off-diagonal elements of $\hat{\boldsymbol{W}}_1$ are shrunk toward zero.

It's hard to say which estimator for $\boldsymbol{W}$ works better. Pritularga, Svetunkov & Kourentzes (2021) demonstrated that the performance of forecast reconciliation is affected by two sources of uncertainties, i.e., the base forecast uncertainty and the reconciliation weight uncertainty. Recall that the uncertainty in the MinT solution in Equation 1 is introduced by the uncertainty in the reconciliation weighting matrix as the summing matrix is fixed for a certain hierarchy. This indicates that OLS and WLSs estimators for $\boldsymbol{W}$ may lead to less volatile reconcliation performance compared to WLSv, MinT, and MinTs estimators. Panagiotelis et al. (2021) provided a geometric intuition for reconciliation and showed that, when considering the Euclidean distance loss function, OLS reconciliation yields results that are at least as favorable as the base forecasts, whereas MinT reconciliation performs poorly relative to the base forecasts. However, when considering the mean squared reconciled forecast error, Wickramasuriya (2021) indicated that MinT reconciliation is better than OLS reconciliation. Therefore, which estimator for $\boldsymbol{W}$ to use hinges on the specific hierarchical time series of interest, the targeted level or series, and the selected loss function.

2.2.2 Reconciliation with regularization

Ben Taieb & Koo (2019) proposed a reconciliation method without posing unbiasedness assumption on the base or reconciled forecasts. They formulated the reconciliation problem as a regularized empirical risk minimization problem given by

$$\min_{\boldsymbol{G}} \frac{1}{(T-T_1-h+1)n} \left\| \boldsymbol{Y} - \hat{\boldsymbol{Y}}\boldsymbol{G}'\boldsymbol{S}' \right\|_F^2 + \lambda \| \operatorname{vec}(\boldsymbol{G}) \|_1,$$

where T_1 denotes the minimum number of observations used for model training, $\|\cdot\|_F$ is the Frobenius norm, $\boldsymbol{Y} = \left[\boldsymbol{y}_{T_1+h}, \ldots, \boldsymbol{y}_T\right]' \in \mathbb{R}^{(T-T_1-h+1)\times n}$, $\hat{\boldsymbol{Y}} = \left[\hat{\boldsymbol{y}}_{T_1+h|T_1}, \ldots, \hat{\boldsymbol{y}}_{T|T-h}\right]' \in \mathbb{R}^{(T-T_1-h+1)\times n}$, and $\lambda \geq 0$ is a regularization parameter. Imposing such a L_1 penalty on $\boldsymbol{G}$ will introduce sparsity and reduce estimation variance, albeit at the cost of introducing some bias. In addition, they also proposed another strategy that penalizes the matrix $\boldsymbol{G}$ towards the solution obtained by bottom-up method, i.e., $\boldsymbol{G}_{\text{BU}} = [\boldsymbol{0}_{n_b \times n_a} \mid \boldsymbol{I}_{n_b}]$.

In practice, a prevalent challenge in forecast reconciliation arises when the base forecasts of some time series within the hierarchical structure may perform poorly, especially for large hierarchies.

This can be attributed to either the inherent complexity of forecasting these series or potential model misspecification. In such cases, the effectiveness of forecast reconciliation may diminish, as the role of the weighting matrix $\boldsymbol{G}$ is to assimilate *all* information of the base forecasts and map them into bottom-level disaggregated forecasts which are subsequently summed by $\boldsymbol{S}$. While the regularized empirical reconciliation method proposed by Ben Taieb & Koo (2019) introduces sparsity by shrinking some elements of $\boldsymbol{G}$ towards zero, it remains incapable of mitigating the adverse impact of underperforming base forecasts on the quality of the reconciled forecasts. Moreover, the method is time-consuming because it uses expanding windows to recursively generate out-of-sample base forecasts, which are then used in the minimization problem.

We therefore propose two branches of innovative methods, constrained (out-of-sample-based) and unconstrained (in-sample-based) reconciliation with selection. These methods aim to identify and address the negative effect of some base forecasts of poor performance in a hierarchy on the overall performance of the reconciled forecasts. Additionally, through the incorporation of regularization in our objective function, our method can enhances reconciliation outcomes produced by using a "bad" choice of $\boldsymbol{W}$, thus reducing the risk of choosing estimator for $\boldsymbol{W}$. Moreover, our method generalizes to grouped hierarchies.

3 Forecast reconciliation with time series selection

For the ease of exposition, we will use the subscript h rather than $T + h \mid T$.

3.1 Constrained reconciliation with selection

Unbiasedness constraint

3.2 Unconstrained reconciliation with selection

4 Monte Carlo simulations

4.1 Model misspecification in a hierarchy

4.2 Exploring the effect of correlation

5 Applications

5.1 Forecasting Australian domestic tourism

5.2 Forecasting Australian prison population

6 Conclusion

Acknowledgement

References

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

Ben Taieb, S & B Koo (2019). Regularized Regression for Hierarchical Forecasting Without Unbiasedness Conditions. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '19. Anchorage, AK, USA: Association for Computing Machinery, pp.1337–1347.

Hyndman, RJ, RA Ahmed, G Athanasopoulos & HL Shang (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis* **55**(9), 2579–2589.

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

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

Pritularga, KF, I Svetunkov & N Kourentzes (2021). Stochastic coherency in forecast reconciliation. *International Journal of Production Economics* **240**, 108221.

Wickramasuriya, SL (2021). Properties of point forecast reconciliation approaches. arXiv: 2103.11129 `[stat.ME]`.

Wickramasuriya, SL, G Athanasopoulos & RJ Hyndman (2019). Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. *Journal of the American Statistical Association* **114**(526), 804–819.