Hierarchical Forecasts Reconciliation

Puwasala Gamakumara*
Department of Econometrics and Business Statistics,
Monash University,
VIC 3800, Australia.
Email: Puwasala.Gamakumara@monash.edu
and
Anastasios Panagiotelis
Department of Econometrics and Business Statistics,
Monash University,
VIC 3800, Australia.
Email: Anastasios.Panagiotelis@monash.edu
and
George Athanasopoulos
Department of Econometrics and Business Statistics,
Monash University,
VIC 3800, Australia.
Email: george.athanasopoulos@monash.edu
and
Rob J Hyndman
Department of Econometrics and Business Statistics,
Monash University,
VIC 3800, Australia.
Email: rob.hyndman@monash.edu

April 8, 2019

Abstract

TBC

*The authors gratefully acknowledge the support of Australian Research Council Grant DP140103220. We also thank Professor Mervyn Silvapulle for valuable comments.

1 Introduction

2 Forecasts reconciliation

2.1 Notation and preliminaries

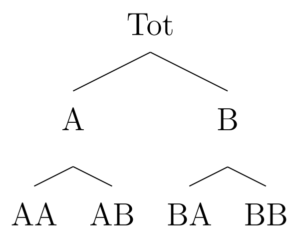

Figure 1: An example of a two level hierarchical structure.

A *hierarchical time series* is a collection of n variables indexed by time, where some variables are aggregates of other variables. We let $\boldsymbol{y}_t \in \mathbb{R}^n$ be a vector comprising observations of all variables in the hierarchy at time t. The *bottom-level series* are defined as those m variables that cannot be formed as aggregates of other variables; we let $\boldsymbol{b}_t \in \mathbb{R}^m$ be a vector comprised of observations of all bottom-level series at time t. The hierarchical structure of the data implies that

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t, \tag{1}$$

where $\boldsymbol{S}$ is an $n \times m$ constant matrix that encodes the aggregation constraints, holds for all t.

To clarify these concepts consider the example of the hierarchy in Figure 1. For this hierarchy, $n = 7$, $\boldsymbol{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{C,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]'$, $m = 4$, $\boldsymbol{b}_t =$

$[y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]'$ and

$$\boldsymbol{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ & \multicolumn{2}{c}{\boldsymbol{I}_4} & \end{pmatrix},$$

where $\boldsymbol{I}_4$ is the 4×4 identity matrix.

While most applications of hierarchical time series to date have involved data that respect an aggregation structure, in principle the matrix $\boldsymbol{S}$ can encode any linear constraints including weighted sums or even cases where some variables in the hierarchy are formed by taking the difference of two other variables.

2.2 Coherent point forecasts

It is desirable that forecasts, whether point forecasts or probabilistic forecasts, should in some sense respect inherent aggregation constraints. We follow other authors (Wickramasuriya et al. 2018, Hyndman & Athanasopoulos 2018) in using the nomenclature *coherence* to describe this property. We now provide new definitions for coherent forecasts in terms of vector spaces that give a geometric understanding of the problem, thus facilitating the development of probabilistic forecast reconciliation in Section 2.3.

Definition 2.1 (Coherent subspace)**.** The m-dimensional linear subspace $\mathfrak{s} \subset \mathbb{R}^n$ that is spanned by the columns of $\boldsymbol{S}$, i.e. $\mathfrak{s} = \text{span}(\boldsymbol{S})$, is defined as the *coherent space*.

It will sometimes be useful to think of pre-multiplication by $\boldsymbol{S}$ as a mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$, in which case we use the notation $s(.)$. Although the codomain of $s(.)$ is $\mathbb{R}^n$, its image is the coherent space $\mathfrak{s}$ as depicted in Figure 2.

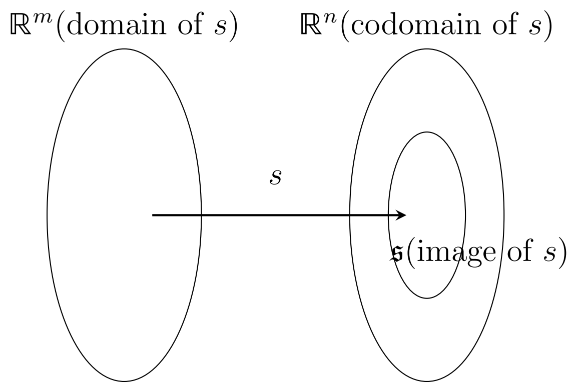

Figure 2: The domain, codomain and image of the mapping s.

Definition 2.2 (Coherent Point Forecasts)**.** Let $\breve{\boldsymbol{y}}_{t+h|t} \in \mathbb{R}^n$ be a point forecast of the values of all series in the hierarchy at time $t+h$, made using information up to and including time t. Then $\breve{\boldsymbol{y}}_{t+h|t}$ is *coherent* if $\breve{\boldsymbol{y}}_{t+h|t} \in \mathfrak{s}$.

2.3 Point forecast reconciliation

As discussed, reconciliation is distinct from coherence, since the former refers to a process whereby incoherent forecasts are made coherent. Although reconciliation methods for point forecasts are extant in the literature they are rarely defined explicitly. We do so here in slightly more general terms than usual.

Let $\hat{\boldsymbol{y}}_{t+h|t} \in \mathbb{R}^n$ be any set of incoherent point forecasts at time $t+h$ conditional on information up to and including time t. We now introduce a linear function that converts unreconciled forecasts into new bottom level forecasts. Let $\boldsymbol{G}$ and $\boldsymbol{d}$ be an $m \times n$ matrix and $m \times 1$ vector respectively, and let $g : \mathbb{R}^n \to \mathbb{R}^m$ be the mapping $g(\boldsymbol{y}) = \boldsymbol{G}\boldsymbol{y} + \boldsymbol{d}$. A composition of g and $s(.)$ gives the following definition for point forecast reconciliation.

Definition 2.3. The point forecast $\tilde{\boldsymbol{y}}_{t+h|t}$ "reconciles" $\hat{\boldsymbol{y}}_{t+h|t}$ with respect to the mapping $g(.)$ iff

$$\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}(\boldsymbol{G}\hat{\boldsymbol{y}}_{t+h|t} + \boldsymbol{d})\,. \tag{2}$$

Several choices of $g(.)$ currently extant in the literature, including the OLS, WLS and MinT methods, are special cases where $s \circ g$ is a projection. These can be defined so that $\boldsymbol{G} = (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}$ and $\boldsymbol{d} = \boldsymbol{0}$, were, $\boldsymbol{R}_{\perp}$ is a $n \times m$ orthogonal complement to an $n \times (n-m)$ matrix $\boldsymbol{R}$, where the columns of the latter span the null space of $\mathfrak{s}$. For example, a straightforward choice of $\boldsymbol{R}$ for the most simple three variable hierarchy where $y_{1,t} = y_{2,t} + y_{3,t}$, is the vector $(1,-1,-1)$ which is orthogonal (in the Euclidean sense) to the columns of $\boldsymbol{S}$. In this case, the matrix $\boldsymbol{R}$ can be interpreted as a 'restrictions' matrix since it has the property that $\boldsymbol{R}'\boldsymbol{y} = \boldsymbol{0}$ for coherent $\boldsymbol{y}$. For this three variable hierarchy, $\boldsymbol{R}'_{\perp} = \boldsymbol{S}$ and reconciliation corresponds to the OLS method. For the case where $\boldsymbol{R}'_{\perp} \neq \boldsymbol{S}$, for example WLS and MinT, there are two possible interpretations. One is that these are oblique projections in Euclidean space where the columns of $\boldsymbol{R}$ are 'directions' along which incoherent point forecasts are projected onto the coherent space $\mathfrak{s}$. Alternatively, since $\boldsymbol{R}'_{\perp}$ is usually written in the form $\boldsymbol{S}'\boldsymbol{W}^{-1}$, these projections can be thought of as orthogonal projections after pre-multiplying by $\boldsymbol{W}^{-1/2}$. A schematic providing a geometric interpretation of point reconciliation is given in Figure 3, while Table 1 summarises existing reconciliation methods.

The columns of $\boldsymbol{S}$ and $\boldsymbol{R}$ provide a basis for $\mathbb{R}^n$. Therefore any incoherent set of point forecasts $\hat{\boldsymbol{y}}_{t+h|t} \in \mathbb{R}^n$ can be expressed in terms of coordinates in the basis defined by $\boldsymbol{S}$ and $\boldsymbol{R}$. Let $\tilde{\boldsymbol{b}}_{t+h|t}$ and $\tilde{\boldsymbol{a}}_{t+h|t}$ be the coordinates corresponding to $\boldsymbol{S}$ and $\boldsymbol{R}$ respectively, after a change of basis. The process of reconciliation involves setting the values of the reconciled bottom-level series to be $\tilde{\boldsymbol{b}}_{t+h|t}$, and ignoring $\tilde{\boldsymbol{a}}_{t+h|t}$ to ensure coherence. From properties

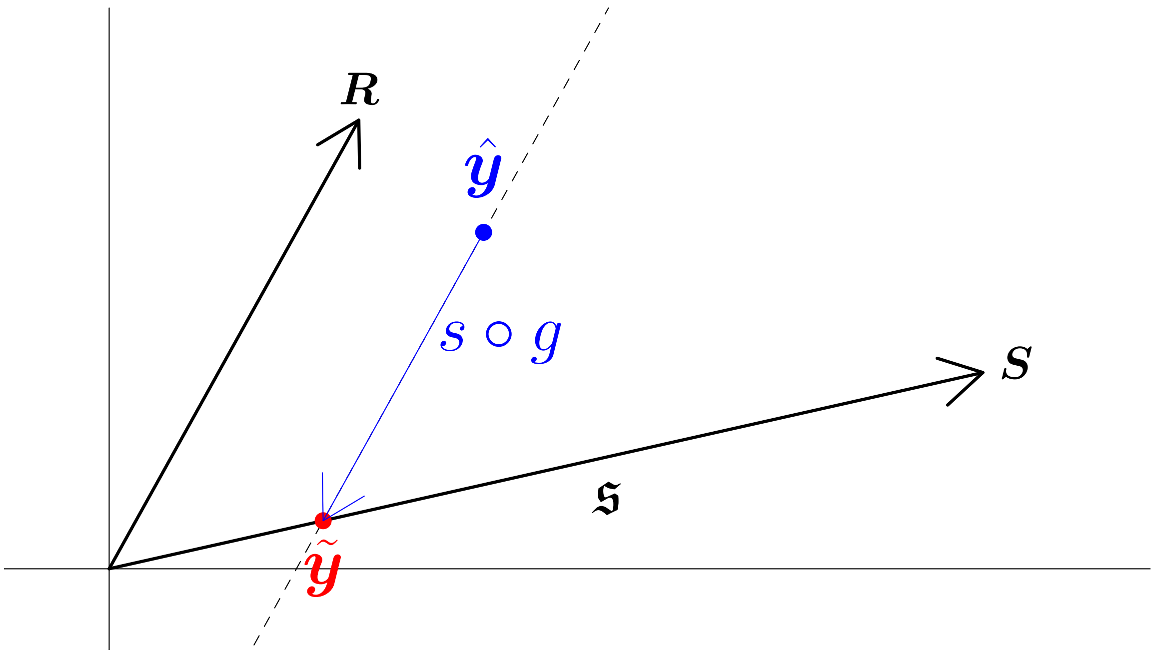

Figure 3: Summary of probabilistic point reconciliation. The mapping $s \circ g$ projects the unreconciled forecast $\hat{\boldsymbol{y}}_{t+h|h}$ onto $\mathfrak{s}$, yielding the reconciled forecast $\tilde{\boldsymbol{y}}_{t+h|h}$ with subscripts dropped in the figure for ease of presentation. Since the smallest hierarchy involves three dimensions, this figure is only a schematic.

Table 1: Summary of reconciliation methods that are projections. Here, $\hat{\boldsymbol{W}}^{sam}$ is the variance covariance matrix of one-step ahead forecast errors, $\hat{\boldsymbol{W}}^{shr}$ is a shrinkage estimator more suited to large dimensions proposed by Schäfer & Strimmer (2005), $\hat{\boldsymbol{W}}^{wls}$ is the diagonal matrix with diagonal elements w_{ii}, and $\tau = \frac{\sum_{i\neq j} \hat{\mathrm{Var}}(\hat{w}_{ij})}{\sum_{i\neq j} \hat{w}_{ij}^2}$, where w_{ij} denotes the (i,j)th element of $\hat{\boldsymbol{W}}^{sam}$.

Method	$\boldsymbol{W}$	$\boldsymbol{R}'_{\perp}$
OLS	$\boldsymbol{I}$	$\boldsymbol{S}'$
MinT(Sample)	$\hat{\boldsymbol{W}}^{sam}$	$\boldsymbol{S}'(\hat{\boldsymbol{W}}^{sam})^{-1}$
MinT(Shrink)	$\tau \mathrm{Diag}(\hat{\boldsymbol{W}}^{sam}) + (1-\tau)\hat{\boldsymbol{W}}^{sam}$	$\boldsymbol{S}'(\hat{\boldsymbol{W}}^{shr})^{-1}$
WLS	$\mathrm{Diag}(\hat{\boldsymbol{W}}^{shr})$	$\boldsymbol{S}'(\hat{\boldsymbol{W}}^{wls})^{-1}$

of linear algebra it follows that

$$\hat{\boldsymbol{y}}_{t+h|t} = (\boldsymbol{S}\ \boldsymbol{R}) \begin{pmatrix} \tilde{\boldsymbol{b}}_{t+h|t} \\ \tilde{\boldsymbol{a}}_{t+h|t} \end{pmatrix} = \boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h|t} + \boldsymbol{R}\tilde{\boldsymbol{a}}_{t+h|t},$$

while the reconciled point forecast is

$$\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h|t}.$$

In order to find $\tilde{\boldsymbol{b}}_{t+h|t}$ we require the inverse $(\boldsymbol{S}\ \boldsymbol{R})^{-1}$ which is given by

$$(\boldsymbol{S}\ \boldsymbol{R})^{-1} = \begin{pmatrix} (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp} \\ (\boldsymbol{S}'_{\perp}\boldsymbol{R})^{-1}\boldsymbol{S}'_{\perp} \end{pmatrix}, \tag{3}$$

where $\boldsymbol{S}_{\perp}$ is the orthogonal complements of $\boldsymbol{S}$. Thus it follows that $\tilde{\boldsymbol{b}}_{t+h|t} = (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}\hat{\boldsymbol{y}}_{t+h|t}$ and $\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}(\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}\hat{\boldsymbol{y}}_{t+h|t}$. Here $(\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}$ corresponds to $\boldsymbol{G}$ as defined previously.

Point reconciliation methods based on projections will always minimise the distance between unreconciled and reconciled forecasts, however the specific distance will depend on the choice of $\boldsymbol{R}$. For example OLS minimises the Euclidean distance between $\hat{\boldsymbol{y}}_{t+h|t}$ and $\tilde{\boldsymbol{y}}_{t+h|t}$, while Wickramasuriya et al. (2018) show that MinT minimises the Mahalonobis distance between $\hat{\boldsymbol{y}}_{t+h|t}$ and $\tilde{\boldsymbol{y}}_{t+h|t}$. Explain how MinT works well. Bottom-up methods minimise distance between reconciled and unreconciled forecasts only along dimensions corresponding to the bottom-level series. Therefore bottom-up methods should be thought of as a boundary case of reconciliation methods, since they ultimately do not use information at all levels of the hierarchy.

We now state two theorems that motivate the use of projections for point forecast reconciliation. First, let $\boldsymbol{\mu}_{t+h|t} := \mathrm{E}(\boldsymbol{y}_{t+h} \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_t)$ and assume $\hat{\boldsymbol{y}}_{t+h|t}$ is an unbiased prediction; that is $\mathrm{E}_{1:t}(\hat{\boldsymbol{y}}_{t+h|t}) = \mu_{t+h|t}$, where the subscript $1:t$ denotes an expectation taken over the training sample.

Theorem 2.1 (Unbiasedness preserving property)**.** *For unbiased $\hat{\boldsymbol{y}}_{t+h|t}$, the reconciled point forecast is also an unbiased prediction as long as $s \circ g$ is a projection.*

Proof. The expected value of the reconciled forecast is given by

$$\mathrm{E}_{1:t}(\tilde{\boldsymbol{y}}_{t+h|t}) = \mathrm{E}_{1:t}(\boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{t+h|t}) = \boldsymbol{S}\boldsymbol{G}\mathrm{E}_{1:t}(\hat{\boldsymbol{y}}_{t+h|t}) = \boldsymbol{S}\boldsymbol{G}\boldsymbol{\mu}_{t+h|t}.$$

Since the aggregation constraints hold for the true data generating process, $\boldsymbol{\mu}_{t+h|t}$ must lie in $\mathfrak{s}$. If $\boldsymbol{S}\boldsymbol{G}$ is a projection, then it is equivalent to the identity map for all vectors that lie in its range. Therefore $\boldsymbol{S}\boldsymbol{G}\boldsymbol{\mu}_{t+h|t} = \boldsymbol{\mu}_{t+h|t}$ when $\boldsymbol{S}\boldsymbol{G}$ is a projection matrix. □

We note the same result does not hold for general g even when the range of $s \circ g$ is $\mathfrak{s}$. Now let $\boldsymbol{y}_{t+h}$ be the realisation of the data generating process at time $t+h$, and let $\|\boldsymbol{v}\|_2$ be the L_2 norm of vector $\boldsymbol{v}$. The following theorem shows that reconciliation never increases, and in most cases reduces, the sum of squared errors of point forecasts.

Theorem 2.2 (Distance reducing property)**.** *If* $\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{SG}\hat{\boldsymbol{y}}_{t+h|t}$*, where* $\boldsymbol{G}$ *is such that* $\boldsymbol{SG}$ *is an orthogonal projection onto* $\mathfrak{s}$*, then the following inequality holds:*

$$\|(\tilde{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_2^2 \leq \|(\hat{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_2^2. \tag{4}$$

Proof. Since the aggregation constraints must hold for all realisations, $\boldsymbol{y}_{t+h} \in \mathfrak{s}$ and $\boldsymbol{y}_{t+h} = \boldsymbol{SG}\boldsymbol{y}_{t+h}$ whenever $\boldsymbol{SG}$ is a projection. Therefore

$$\begin{aligned}
\|(\tilde{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_2 &= \|(\boldsymbol{SG}\hat{\boldsymbol{y}}_{t+h|t} - \boldsymbol{SG}\boldsymbol{y}_{t+h})\|_2 && (5)\\
&= \|\boldsymbol{SG}(\hat{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_2 && (6)\\
& && (7)
\end{aligned}$$

The Cauchy-Schwarz inequality can be used to show that orthogonal projections are bounded operators (Hunter & Nachtergaele 2001), therefore

$$\|\boldsymbol{SG}(\hat{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_2 \leq \|(\hat{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_2.$$

□

The inequality is strict whenever $\hat{\boldsymbol{y}}_{t+h|t} \notin \mathfrak{s}$.

3 Bias correction

4 Application

5 Conclusions

References

Athanasopoulos, G., Ahmed, R. A. & Hyndman, R. J. (2009), 'Hierarchical forecasts for australian domestic tourism', *International Journal of Forecasting* **25**(1), 146 – 166.

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

Dunn, D. M., Williams, W. H. & Dechaine, T. L. (1976), 'Aggregate Versus Subaggregate Models in Local Area Forecasting', *Journal of American Statistical Association* **71**(353), 68–71.

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

Hunter, J. K. & Nachtergaele, B. (2001), *Applied analysis*, World Scientific Publishing Company.

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.

Hyndman, R. J. & Athanasopoulos, G. (2018), *Forecasting: principles and practice, 2nd Edition*, OTexts.

Schäfer, J. & Strimmer, K. (2005), 'A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics', *Statistical Applications in Genetics and Molecular Biology* **4**(1).

van Erven, T. & Cugliari, J. (2014), *Game-Theoretically Optimal reconciliation of contemporaneous hierarchical time series forecasts.*

Wickramasuriya, S. L., Athanasopoulos, G. & Hyndman, R. J. (2018), 'Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization', *J American Statistical Association* . to appear.