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

May 2, 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 Coherent forecasts

2.1 Notation and preliminaries

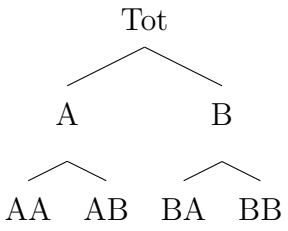


Figure 1: An example of a two level hierarchical structure.

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

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

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 point 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.

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.

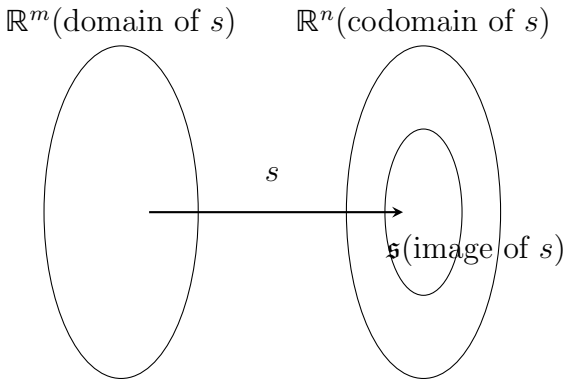


Figure 2: The domain, codomain and image of the mapping s.

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 $\breve{\boldsymbol{y}}_{t+h|t} \in \mathfrak{s}$.

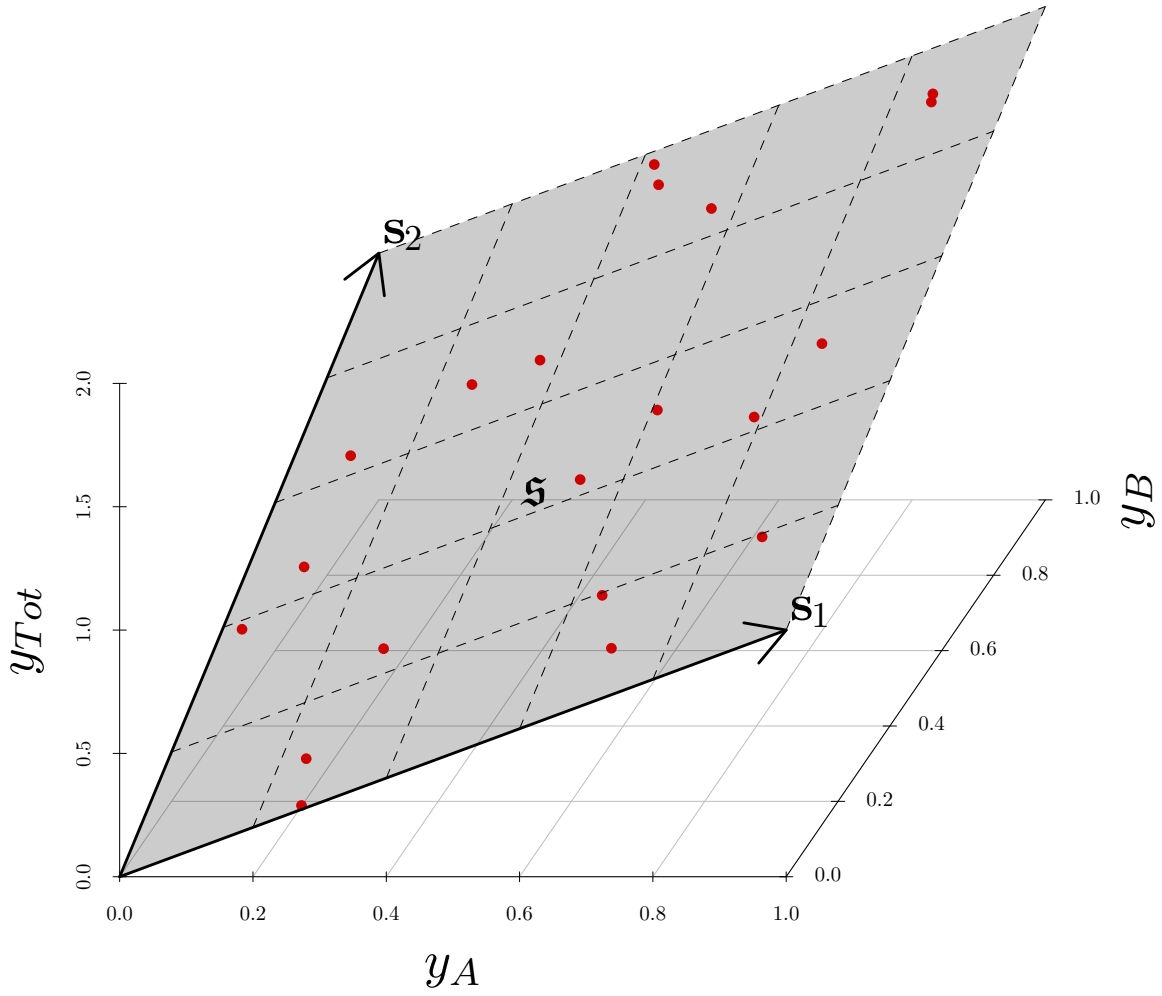


Figure 3: Depiction of a three dimensional hierarchy with $y_{\text{Tot}} = y_{\text{A}} + y_{\text{B}}$. The gray colour two dimensional plane reflects the coherent subspace $\mathfrak{s}$ where $\vec{s}_1 = (1, 1, 0)'$ and $\vec{s}_2 = (1, 0, 1)'$ are basis vectors that spans $\mathfrak{s}$. The points in $\mathfrak{s}$ represents realisations or coherent forecasts

These definitions of the coherent space $\mathfrak{s}$ and coherent point forecasts in terms of the mapping $s(.)$, may give the impression that the bottom-level series play an important role. However, alternative definitions could be formed using any set of basis vectors that span $\mathfrak{s}$. For example, consider the most simple three variable hierarchy where $y_{Tot,t} = y_{A,t} + y_{B,t}$. Figure 3 represent this in a 3-D diagram. In this case the matrix $\boldsymbol{S}$ has columns $\vec{s}_1 = (1, 1, 0)'$ and $\vec{s}_2 = (1, 0, 1)'$ spanning $\mathfrak{s}$ which is depicted in gray colour two dimensional plane. Pre-multiplying by $\boldsymbol{S}$ transforms arbitrary values of $y_{A,t}$ and $y_{B,t}$ into a coherent vector for the full hierarchy. These coherent points always lies in the $\mathfrak{s}$ plane as the red points depicted in the figure.

However the columns $(1, 0, 1)'$ and $(0, 1, -1)'$ also span $\mathfrak{s}$ and define a mapping that transforms arbitrary values of $y_{Tot,t}$ and $y_{A,t}$ into a coherent vector for the full hierarchy. The definitions above could be made in terms of any series and not just the bottom-level series. In general, we call the series (or linear combinations thereof) used in the definitions of coherence *basis series*. Unless stated otherwise, we will always assume that the basis series are the bottom-level series as in Definition 2.2, since this facilitates comparison with existing approaches in the literature.

3 Forecasts reconciliation

3.1 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 3.1. 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 $\boldsymbol{d} = \boldsymbol{0}$, where, $\boldsymbol{R}_\perp$ 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 $\boldsymbol{s}$. 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 $\boldsymbol{y}$. In OLS reconciliation, $\boldsymbol{R}'_\perp = \boldsymbol{S}'$ whereas in MinT or WLS reconciliation $\boldsymbol{R}'_\perp$ takes the form $\boldsymbol{S}'\boldsymbol{W}^{-1}$. We will be discussing these projections distinctly in the next subsection.

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

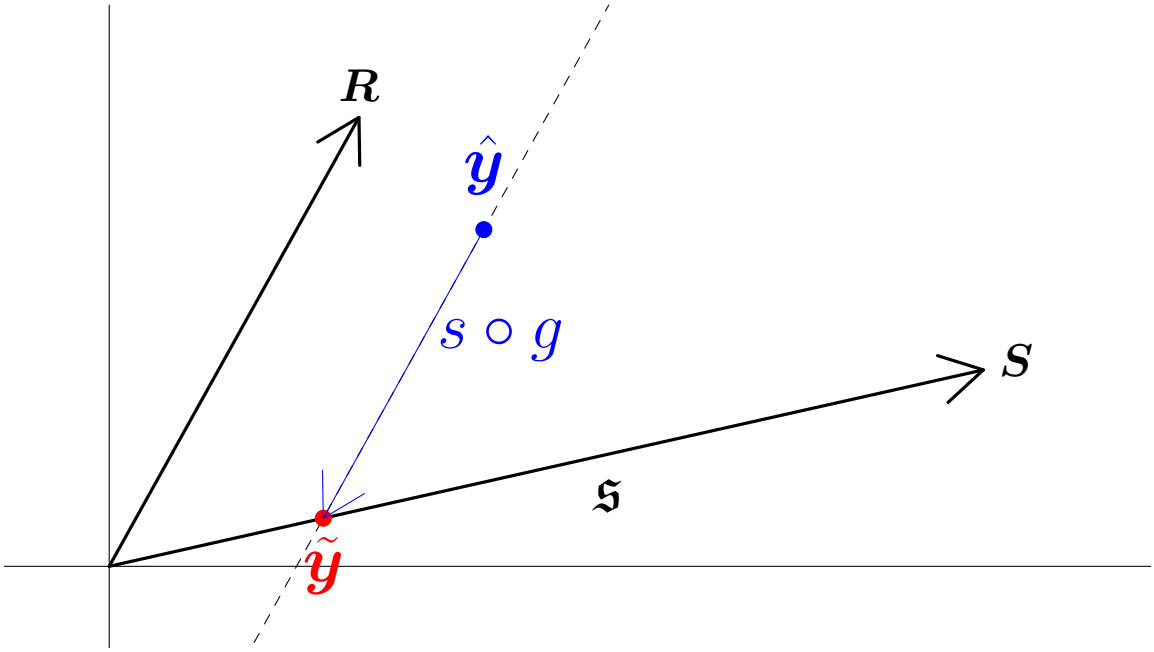


Figure 4: Summary of probabilistic point reconciliation. The mapping $s \circ g$ projects the unreconciled forecast $\hat{\boldsymbol{y}}_{t+h|h}$ onto $\mathfrak{s}$, yielding the reconciled forecast $\tilde{\boldsymbol{y}}_{t+h|h}$ with subscripts dropped in the figure for ease of presentation. Since the smallest hierarchy involves three dimensions, this figure is only a schematic.

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.

3.2 Motivation of using projections

We have seen from the above discussion that projection is playing an important role in point forecast reconciliation. Now we turn our attention to the following two theorems that explains the motivation of using projection in this context.

First, let $\boldsymbol{\mu}_{t+h|t} := \mathrm{E}(\boldsymbol{y}_{t+h} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$ and assume $\hat{\boldsymbol{y}}_{t+h|t}$ is an unbiased prediction; that is $\mathrm{E}_{1:t}(\hat{\boldsymbol{y}}_{t+h|t}) = \boldsymbol{\mu}_{t+h|t}$, where the subscript $1:t$ denotes an expectation taken over the training sample.

Theorem 3.1 (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 onto $\mathfrak{s}$.*

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

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

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

We note the above result holds when the projection $s \circ g$ is only onto the coherent subspace $\mathfrak{s}$. That is the result does not hold for any general g even when the range of $s \circ g$ is $\mathfrak{s}$. To describe this more explicitly suppose $s \circ g$ is a projection to any linear subspace $\mathfrak{L}$ of $\mathfrak{s}$. Then $\boldsymbol{SG}\boldsymbol{\mu}_{t+h|t} \neq \boldsymbol{\mu}_{t+h|t}$ as the projection will move $\boldsymbol{\mu}_{t+h|t}$ to a point $\bar{\boldsymbol{\mu}}$ in $\mathfrak{L}$ as depicted in the Figure 5. Thus $\mathrm{E}_{1:t}(\tilde{\boldsymbol{y}}_{t+h|t}) \neq \boldsymbol{\mu}_{t+h|t}$ which breaks the unbiasedness. Recall the top-down method (Gross & Sohl 1990) with

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{p} & \mathbf{0}_{(m\times n-1)} \end{pmatrix} \tag{4}$$

where $\boldsymbol{p} = (p_1, \ldots, p_m)'$ is an m-dimensional vector consisting a set of proportions which is use to disaggregate the top-level forecasts along the hierarchy. Hyndman et al. (2011) claimed that this method is not producing unbiased coherent forecasts even if the base forecasts are unbiased since $\boldsymbol{SGS} \neq \boldsymbol{S}$ for $\boldsymbol{G}$ in (4). However the more rational explanation is that, in top-down approach the projection $s \circ g$ is not onto $\mathfrak{s}$, but to a linear subspace of $\mathfrak{s}$ spanned by $\boldsymbol{p}$. Thus from above explanation it follows that $\boldsymbol{SG}\boldsymbol{\mu}_{t+h|t} \neq \boldsymbol{\mu}_{t+h|t}$ and hence not producing unbiased forecasts.

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.

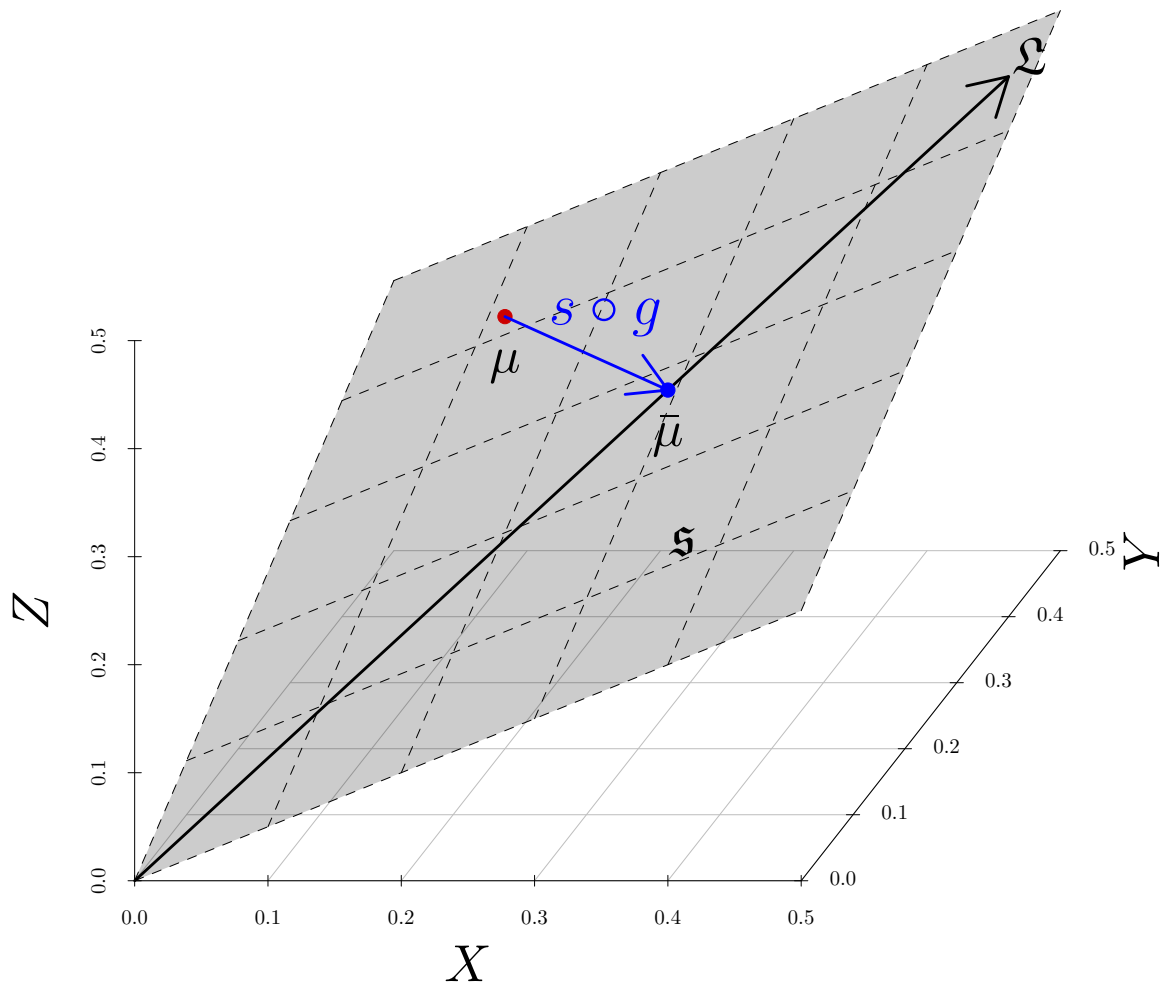


Figure 5: $\mathfrak{L}$ is a linear subspace of the coherent subspace $\mathfrak{s}$. If $s \circ g$ is a projection not onto $\mathfrak{s}$ but onto $\mathfrak{L}$, then $\boldsymbol{\mu} \in \mathfrak{s}$ will be moved to $\bar{\boldsymbol{\mu}} \in \mathfrak{L}$.

Theorem 3.2 (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:*

$$\|(\tilde{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_2^2 \leq \|(\hat{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_2^2. \tag{5}$$

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 onto $\mathfrak{s}$. Therefore,

$$\|(\tilde{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_2 = \|(\boldsymbol{SG}\hat{\boldsymbol{y}}_{t+h|t} - \boldsymbol{SG}\boldsymbol{y}_{t+h})\|_2 \tag{6}$$

$$= \|\boldsymbol{SG}(\hat{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_2. \tag{7}$$

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 $\hat{\boldsymbol{y}}_{t+h|t} \notin \mathfrak{s}$.

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 $\boldsymbol{R}$. Following subsections will explicitly discuss the different projection based reconciliation methods and their optimality based on distinct distance measures.

3.2.1 OLS reconciliation

Recall that in OLS reconciliation, $\boldsymbol{R}_\perp = \boldsymbol{S}$ and thus it orthogonally projects $\hat{\boldsymbol{y}}$ to the coherent subspace. Further, it minimises the Euclidean distance between $\hat{\boldsymbol{y}}_{t+h|t}$ and $\tilde{\boldsymbol{y}}_{t+h|t}$. In addition to that Figure 6 also shows that $\tilde{\boldsymbol{y}}$ is always closer to $\boldsymbol{y}$ than $\hat{\boldsymbol{y}}$ in terms of

the Euclidean distance which is directly followed from the Pythagorean theorem. It also implies that the sum of squared error for OLS reconciled forecasts are always less than that for base forecasts.

3.2.2 MinT reconciliation

In MinT reconciliation, $\boldsymbol{R}'_{\perp}$ is taking the form $\boldsymbol{S}'\boldsymbol{W}^{-1}$, where it can be thought of as orthogonal projections after pre-multiplying by $\boldsymbol{W}^{-1/2}$. That is, the coordinates of incoherent space will be scaled by $\boldsymbol{W}^{-1/2}$ which is then followed by the orthogonal projection. Alternatively this can be interpreted as an oblique projections in Euclidean space where the columns of $\boldsymbol{R}$ is the 'direction' along which incoherent point forecasts are projected onto the coherent space $\mathfrak{s}$ as depicted in Figure 4. In terms of distances, MinT minimises the Euclidean distance between $\hat{\boldsymbol{y}}_{t+h|t}$ and $\tilde{\boldsymbol{y}}_{t+h|t}$ in the transformed space which is same as the scaled Euclidean distance in the original space. Latter is also referred to as the Mahalonobis distance. We also note that the WLS is a special case of MinT where $\boldsymbol{W}^{-1}$ is a diagonal matrix.

Wickramasuriya et al. (2018) showed that the MinT is optimal with respect to the mean squared forecast errors. We can provide a more general geometrical explanation to this optimality using the schematic in Figure 7. Consider the h-step ahead reconciled forecast errors. These can be always approximated by the insample h-step ahead forecast errors. Since these errors are coherent, they lies in a direction that is closer to the coherent subspace $\mathfrak{s}$. Therefore if you project $\hat{\boldsymbol{y}}$ along the direction of these in-sample forecast errors, then you can get closer to the true value $\boldsymbol{y}$ as depicted in the schematic. Further, unlike OLS, the squared error for MinT reconciled forecasts is not always less than that of base forecasts in every single replication although it outperforms on average.

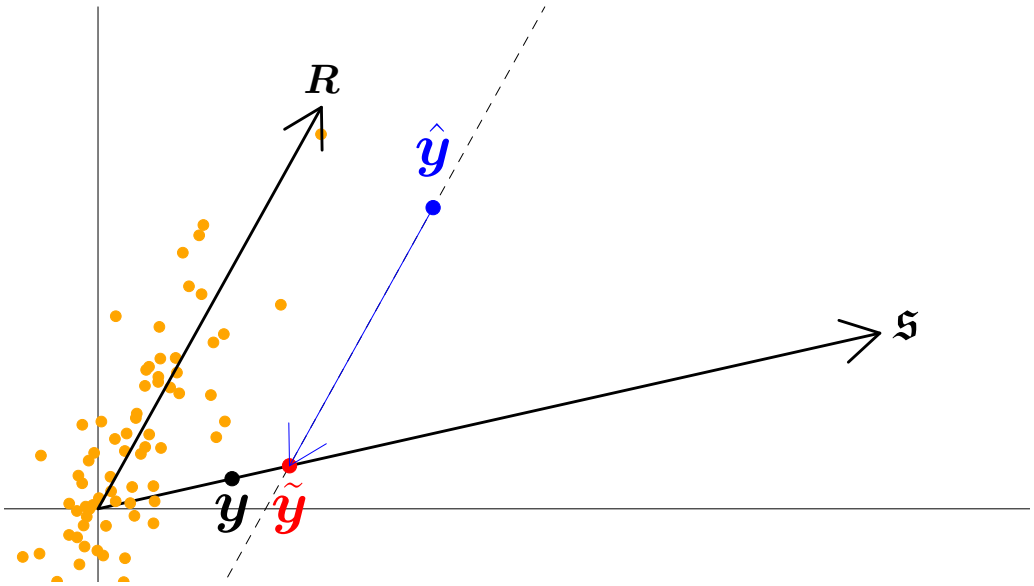


Figure 7: A schematic to represent MinT reconciliation. Points in orange colour represent the insample errors. $\boldsymbol{R}$ shows the direction of the insample errors. $\hat{\boldsymbol{y}}$ is projected onto $\mathfrak{s}$ along the the direction of $\boldsymbol{R}$.

3.2.3 Bottom-up method

Bottom-up method is one of the traditional and simplest ways of producing coherent forecasts. Under this approach, the incoherent forecasts are projected to the coherent subspace along the direction which is perpendicular to the bottom level series. In terms of distances, this method minimises the distance between reconciled and unreconciled forecasts only

along the dimension 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.

4 Bias correction

5 Application

6 Conclusions

References

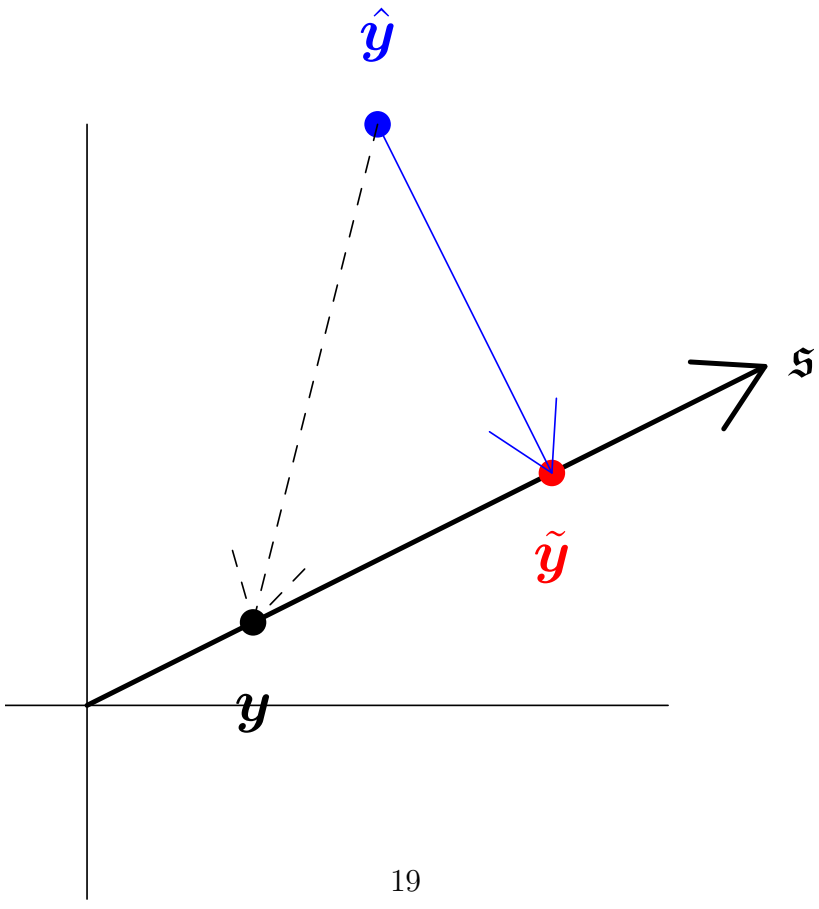
Gross, C. W. & Sohl, J. E. (1990), 'Disaggregation methods to expedite product line forecasting', *Journal of Forecasting* **9**(3), 233–254.

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.

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.


$\hat{\boldsymbol{y}}$
$\mathfrak{S}$
$\tilde{\boldsymbol{y}}$
$\boldsymbol{y}$