

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

July 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

We briefly define the concept of a *hierarchical time series* in a fashion similar to [, before](#) citations elaborating on some of the limitations of this understanding. A *hierarchical time series* is a collection of n variables indexed by time, where some variables are aggregates of other variables. We let $\mathbf{y}_t \in \mathbb{R}^n$ 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 $\mathbf{b}_t \in \mathbb{R}^m$ be a vector comprised of observations of all bottom-level series at time t . The hierarchical structure of the data implies that

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

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

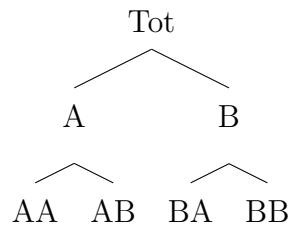


Figure 1: An example of a two level hierarchical structure.

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

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

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

While such a definition is completely serviceable, it obscures the full generality of many methodologies developed in the literature on so-called hierarchical time series. In fact, to apply concepts such as coherence and reconciliation, the data only require two important characteristics; the first is that they are multivariate, the second is that they adhere to linear constraints. Below we redefine the concepts of coherence and hierarchical time series using a geometric interpretation, before turning our attention to forecast reconciliation in Section 3.

2.2 Coherence

Definition 2.1 (Coherent subspace). The m -dimensional linear subspace $\mathfrak{s} \subset \mathbb{R}^n$ for which a set of linear constraints holds for all $\mathbf{y} \in \mathfrak{s}$ is defined as the *coherent subspace*.

To further illustrate, Figure 2 depicts the most simple three variable hierarchy where $y_{Tot,t} = y_{A,t} + y_{B,t}$. The coherent subspace is depicted as a grey 2-dimensional plane within 3-dimensional space, i.e. $m = 2$ and $n = 3$. It is worth noting that the coherent subspace is spanned by the columns of \mathbf{S} , i.e. $\mathfrak{s} = \text{span}(\mathbf{S})$. In Figure 2, these columns are $\vec{s}_1 = (1, 1, 0)'$ and $\vec{s}_2 = (1, 0, 1)'$. However, it is equally important to recognise that the hierarchy could

also have been defined in terms of $y_{Tot,t}$ and $y_{A,t}$ rather than the bottom level series, $y_{A,t}$ and $y_{B,t}$. In this case the corresponding ‘ \mathbf{S} matrix’ would have columns $(1, 0, 1)'$ and $(0, 1, -1)'$. However, while there are multiple ways to define an \mathbf{S} matrix, in all cases the columns will span the same coherent subspace, which is unique.

Also notable by its absence in the above definition is any reference to *aggregation*. As the literature has shown, the linear constraints need not be aggregation constraints at all. For example consider weighted sums, while consider an example where one variable is the difference of two other variables.

find
refer-
ence

include
Li and
Tang
refer-
ence

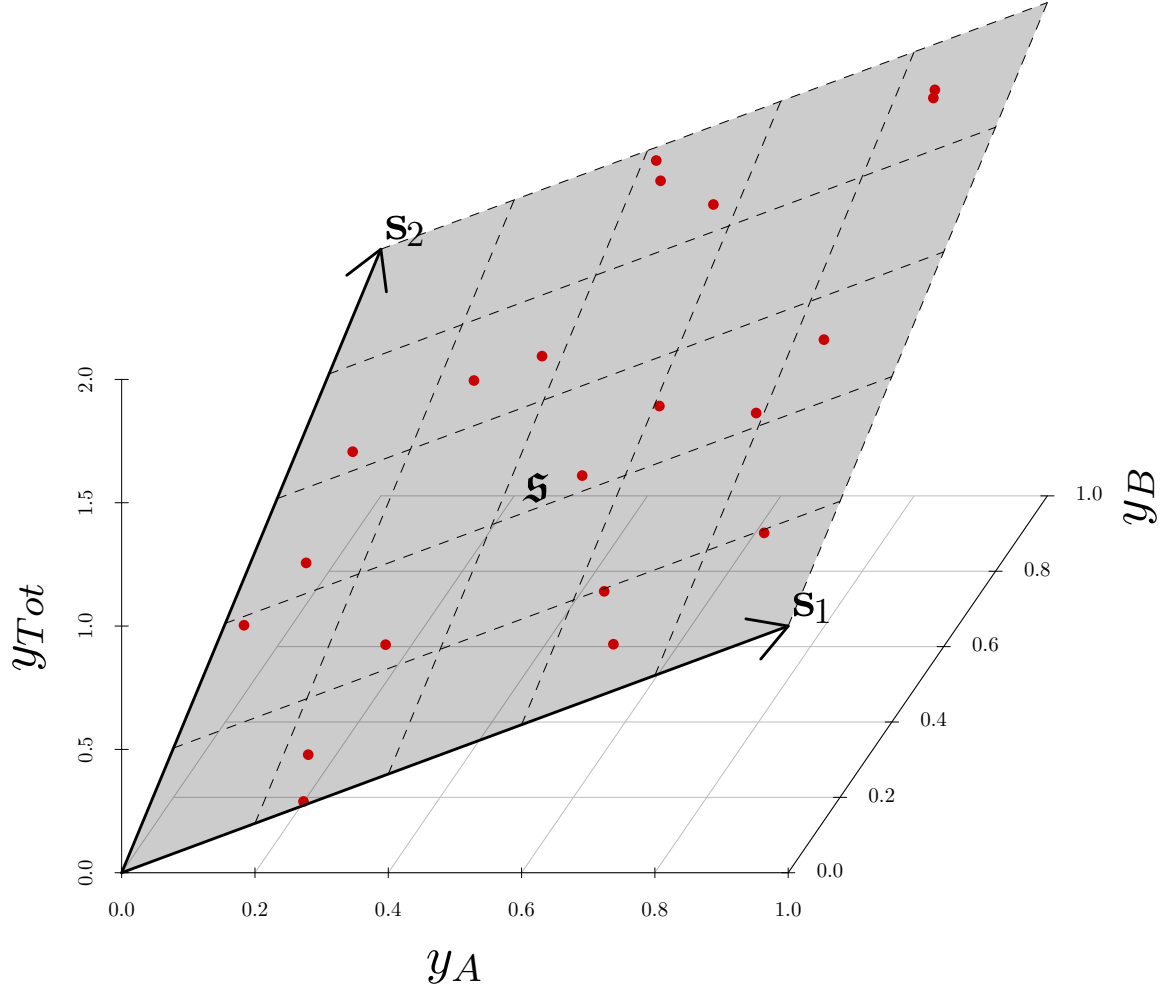


Figure 2: Depiction of a three dimensional hierarchy with $y_{Tot} = y_A + y_B$. The gray colour two dimensional plane reflects the coherent subspace \mathfrak{s} where $\vec{s}_1 = (1, 1, 0)'$ and $\vec{s}_2 = (1, 0, 1)'$ are basis vectors that spans \mathfrak{s} . The points in \mathfrak{s} represents realisations or coherent forecasts

Definition 2.2 (Hierarchical Time Series). A hierarchical time series is an n -dimensional multivariate time series such that all observed values $\mathbf{y}_1, \dots, \mathbf{y}_T$ and all future values $\mathbf{y}_{T+1}, \mathbf{y}_{T+2}, \dots$ lie in the coherent subspace, i.e. $\mathbf{y}_t \in \mathfrak{s} \quad \forall t$.

Despite the common use of the term *hierarchical time series*, it should be clear from the definition that the data need not necessarily follow a hierarchy. In fact, the term *hierarchical* is misleading since the literature has covered instances that cannot easily be depicted as hierarchies as in Figure 1. These include . Furthermore, although the definition makes clear reference to time series, this definition can be easily generalised to any vector-valued data for which some linear constraints are known to hold for all realisations.

include
refer-
ences

Definition 2.3 (Coherent Point Forecasts). Let $\check{\mathbf{y}}_{t+h|t} \in \mathbb{R}^n$ 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 $\check{\mathbf{y}}_{t+h|t}$ is *coherent* if $\check{\mathbf{y}}_{t+h|t} \in \mathfrak{s}$.

Without any loss of generality, that above definition could also be applied to prediction for multivariate data in general, rather than just forecasting of time series. While the observed data will be coherent by definition, it is important to note that there are a number of reasons why forecasts or predictions may be incoherent.

First, since applications of hierarchical forecasting tend to be very high dimensional a common strategy in practice is to produce forecasts for each time series independently using univariate models. Second, even where a multivariate model is used for the full vector of observations, it may be difficult to capture the linear constraints inherent in the data particularly for complicated non-linear models. Third, in some cases judgemental adjustments may be made inducing incoherent forecasts.

some
discus-
sion
about
why
recon-
cilia-
tion v
single
level

3 Forecast reconciliation

As discussed in the previous section, for a number of reasons, coherence is not guaranteed when forecasts are produced for all series. To ensure aligned decision making, it is desirable to adjust forecasts ex post to ensure coherence. This process is referred to as *reconciliation*. In the most general terms, reconciliation can be defined as follows

Definition 3.1 (Reconciled forecasts). Let ψ be a mapping, $\psi : \mathbb{R}^n \rightarrow \mathfrak{s}$. The point forecast $\tilde{\mathbf{y}}_{t+h|t}$ “reconciles” $\hat{\mathbf{y}}_{t+h|t}$ with respect to the mapping $\psi(\cdot)$ iff

$$\tilde{\mathbf{y}}_{t+h|t} = \psi(\hat{\mathbf{y}}_{t+h|t}) . \quad (2)$$

All reconciliation methods that we are aware of consider a linear mapping for ψ , which involves pre-multiplying base forecasts by an $n \times n$ matrix that has \mathfrak{s} as its image. One way to achieve this is with a matrix \mathbf{SG} (with \mathbf{P} used in place of \mathbf{G} in some cases), where \mathbf{G} is an $(n - m) \times n$ matrix. This facilitates an interpretation of reconciliation as a two-step process, in the first step, base forecasts $\hat{\mathbf{y}}_{t+h|t}$ are combined to form a new set of bottom level forecasts, in the second step, these mapped to a full vector of coherent forecasts via pre-multiplication by \mathbf{S} .

Although pre-multiplying base forecasts by \mathbf{SG} will result in coherent forecasts, a number of desirable properties arise when \mathbf{SG} has the specific structure of a *projection* matrix. We begin by considering the special case of an orthogonal projection whereby $\mathbf{G} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$. This is equivalent to so called OLS reconciliation as introduced by Hyndman et al. (2011) where the connection between OLS and orthogonal projection should be clear.

3.1 Orthogonal projection

In this section we discuss two sensible properties that can be achieved by reconciliation via orthogonal projection. The first is that reconciliation should adjust the base forecasts as little as possible, i.e. the base and reconciled forecast should be ‘close’. The second is that reconciliation in some sense should improve forecast accuracy, or more loosely, that the reconciled forecast should be ‘closer’ to the truth.

To address the first of these properties we make the concept of closeness more concrete, by considering the Euclidean distance between the base forecast $\hat{\mathbf{y}}$ and the reconciled forecast $\tilde{\mathbf{y}}$. A property of an orthogonal projection is that the distance between $\hat{\mathbf{y}}$ and $\tilde{\mathbf{y}}$ will be as small as possible given that $\tilde{\mathbf{y}} \in \mathfrak{s}$. In this way reconciliation via orthogonal projection does not lead to large adjustments of the base forecasts.

The second property introduced above has been the focus of theoretical results in the forecast reconciliation literature. Here we provide a more streamlined version of proofs by van Erven & Cugliari (2014) and Wickramasuriya et al. (2018) before providing geometrical intuition aimed at simplifying the reader’s understanding of these proofs.

Consider the Euclidean distance between a forecast and the target, which is equivalent to the root total squared error. Let \mathbf{y}_{t+h} be the realisation of the data generating process at time $t + h$, and let $\|\mathbf{v}\|_2$ be the L_2 norm of vector \mathbf{v} . The following theorem shows that reconciliation never increases, and in most cases reduces, the sum of squared errors of point forecasts.

Theorem 3.1 (Distance reducing property). *If $\tilde{\mathbf{y}}_{t+h|t} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{t+h|t}$, where \mathbf{G} is such that $\mathbf{S}\mathbf{G}$ is an orthogonal projection onto \mathfrak{s} , then the following inequality holds:*

$$\|(\tilde{\mathbf{y}}_{t+h|t} - \mathbf{y}_{t+h})\|_2^2 \leq \|(\hat{\mathbf{y}}_{t+h|t} - \mathbf{y}_{t+h})\|_2^2. \quad (3)$$

Proof. Since the aggregation constraints must hold for all realisations, $\mathbf{y}_{t+h} \in \mathfrak{s}$ and $\mathbf{y}_{t+h} =$

$SG\mathbf{y}_{t+h}$ whenever SG is a projection onto \mathfrak{s} . Therefore,

$$\|(\tilde{\mathbf{y}}_{t+h|t} - \mathbf{y}_{t+h})\|_2 = \|(SG\hat{\mathbf{y}}_{t+h|t} - SG\mathbf{y}_{t+h})\|_2 \quad (4)$$

$$= \|SG(\hat{\mathbf{y}}_{t+h|t} - \mathbf{y}_{t+h})\|_2. \quad (5)$$

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

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

□

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

The simple geometric intuition behind the proof is demonstrated in Figure ?? . In this schematic, the coherent subspace is depicted as a black arrow. The base forecast $\hat{\mathbf{y}}$ is shown as a blue dot. Since it is incoherent it does not lie of \mathfrak{s} . Reconciliation is an orthogonal projection from $\hat{\mathbf{y}}$ to the coherent subspace yielding the reconciled forecast $\tilde{\mathbf{y}}$ shown in red. Finally, the target of the forecast \mathbf{y} is displayed as a black point, and although its exact location is unknown to the forecaster, it is known that it will lie somewhere along the coherent subspace.

compare
to van
Erven
and
Wick-
rema-
suriya
more
ex-
plicitly

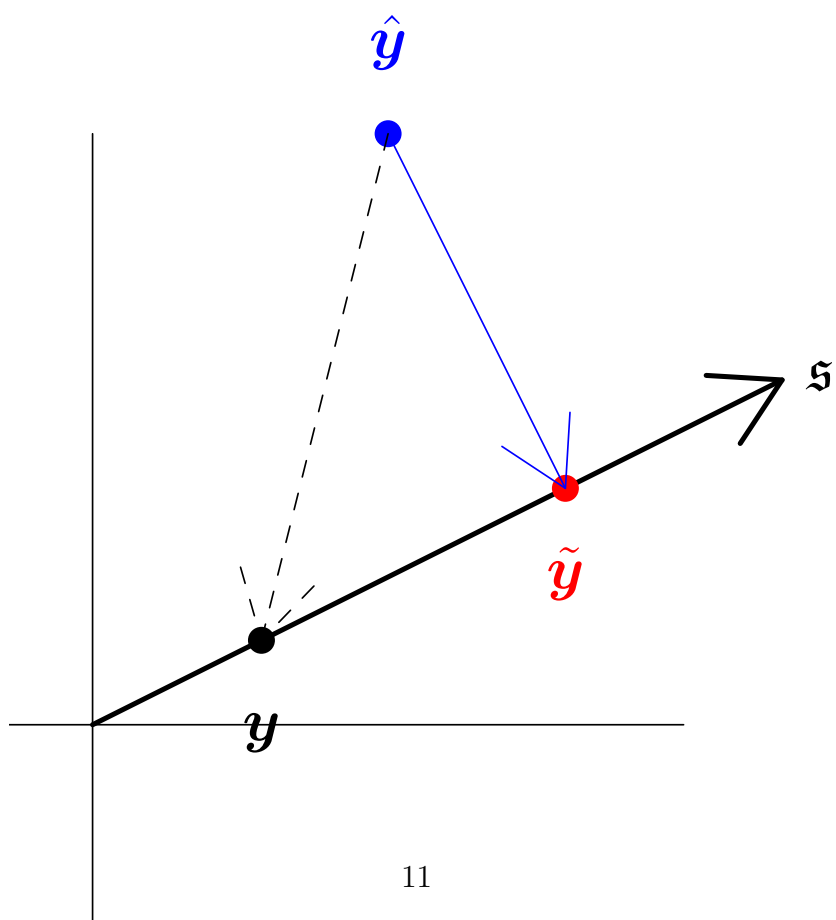


Figure ?? clearly shows that $\hat{\mathbf{y}}$, $\tilde{\mathbf{y}}$ and \mathbf{y} form a right angled triangle. In this triangle the distance between \mathbf{y} and $\hat{\mathbf{y}}$ is the hypotenuse and therefore must be longer than the distance between \mathbf{y} and $\tilde{\mathbf{y}}$. As such reconciliation is guaranteed to reduce the squared error of the forecast.

The theorem is in some ways perhaps more powerful than perhaps previously understood. The theorem is often stated in terms of expectations. However, the distance reducing property result is stronger than that since it will hold for any realisation and any forecast. Nothing needs to be assumed about the statistical properties of the data generating process or the process by which forecasts are made.

check
other
proofs

However, in other ways, the theorem is weaker than perhaps often understood. First, when improvements in forecast accuracy are discussed in the context of the theorem, this refers to a very specific measure of forecast accuracy. In particular, this measure is the root of the sum of squared errors of *all* variables in the hierarchy. As such, while forecast improvement is guaranteed for the hierarchy overall, reconciliation can lead to worse forecasts for individual series. Second, although orthogonal projections are guaranteed to improve on base forecasts both for all realisations and in expectation, they are not necessarily the projection that leads to the greatest improvement in forecast accuracy. As such referring to reconciliation via orthogonal projections as ‘optimal’ is somewhat misleading since it does not have the optimality properties of some oblique projections, in particular MinT. It is to oblique projections that we now turn our attention.

3.2 Oblique Projections

First, the linear subspace onto which all points are projected, or the image of the projection, must be defined. In our context this can be defined by the m columns of the matrix \mathbf{S} . Second, the direction along which points are projected must be defined. This will be

achieved by defining a matrix \mathbf{R} with $n - m$ columns then span the direction of projection. A schematic of this is presented . A projection matrix can then be constructed as $\mathbf{S}(\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}$ where, \mathbf{R}_{\perp} is an $n \times m$ orthogonal complement to \mathbf{R} such that $\mathbf{R}'_{\perp}\mathbf{R} = \mathbf{0}$. It is simple to verify that this construction satisfies the properties of a projection matrix, namely symmetry and idempotence.

A straightforward choice of \mathbf{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 \mathbf{S} . In this case, the matrix \mathbf{R} can be interpreted as a ‘restrictions’ matrix since it has the property that $\mathbf{R}'\mathbf{y} = \mathbf{0}$ for coherent \mathbf{y} . In OLS reconciliation, $\mathbf{R}'_{\perp} = \mathbf{S}'$ whereas in MinT or WLS reconciliation \mathbf{R}'_{\perp} takes the form $\mathbf{S}'\mathbf{W}^{-1}$. We will be discussing these projections distinctly in the next subsection.

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

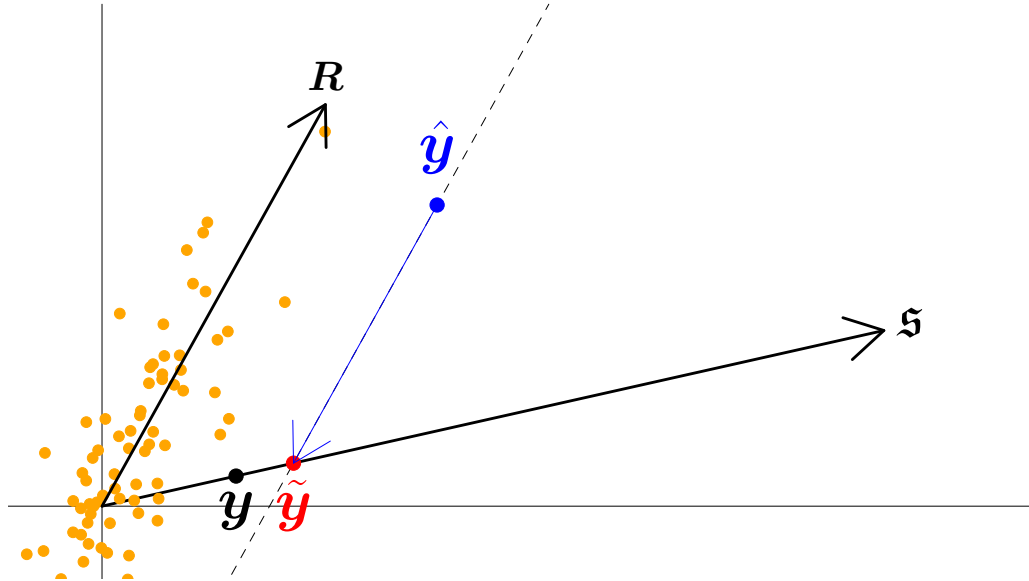


Figure 4: A schematic to represent MinT reconciliation. Points in orange colour represent the insample errors. \mathbf{R} shows the direction of the insample errors. $\hat{\mathbf{y}}$ is projected onto \mathfrak{s} along the the direction of \mathbf{R} .

4 The issue of bias

A sensible property is that reconciliation should not change base forecasts that are already coherent. That is

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{t+h|t} = \hat{\mathbf{y}}_{t+h|t} \text{ iff } \hat{\mathbf{y}}_{t+h|t} \in \mathfrak{s}. \quad (6)$$

Using geometric intuition, it should be clear that the property in Equation 6 will hold when $\mathbf{S}\mathbf{G}$ is a projection matrix. This also implies that the property in Equation 6 does not hold for arbitrary \mathbf{G} . For this reason it is often assumed that $\mathbf{S}\mathbf{G}\mathbf{S} = \mathbf{S}$ or alternatively that $\mathbf{G}\mathbf{S} = \mathbf{I}$. Enforcing these assumptions are equivalent to assuming that $\mathbf{S}\mathbf{G}$ is a projection matrix.

The property that projections map all vectors in the coherent subspace onto themselves is useful in proving the unbiasedness preserving property of reconciliation, previously proven by. Suppose that the target of a point forecast is $\boldsymbol{\mu}_{t+h|t} := \mathbb{E}(\mathbf{y}_{t+h} \mid \mathbf{y}_1, \dots, \mathbf{y}_t)$ where the expectation is taken over the predictive density. Our point forecast is an estimate of this quantity that is random due to sampling variation. The point forecast will be unbiased if $\mathbb{E}_{1:t}(\hat{\mathbf{y}}_{t+h|t}) = \boldsymbol{\mu}_{t+h|t}$, where the subscript $1:t$ denotes an expectation taken over the training sample.

Theorem 4.1 (Unbiasedness preserving property). *For unbiased $\hat{\mathbf{y}}_{t+h|t}$, the reconciled point forecast is also an unbiased prediction as long as $\mathbf{S}\mathbf{G}$ is a projection onto \mathfrak{s} .*

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

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

Since $\boldsymbol{\mu}_{t+h|t}$ is an expectation taken with respect to the degenerate predictive density it must lie in \mathfrak{s} . We have already established that when $\mathbf{S}\mathbf{G}$ is a projection onto \mathfrak{s} then it

perhaps
find a
reference -
it is so
funda-
men-
tal
that is
is on
wikipedia
and
wol-
fram
math-
world

perhaps
elab-
orate
in a
proof
in ap-
pendix

reference
Shanika
and
maybe
van

maps all vectors in \mathfrak{s} onto themselves. As such $\mathbf{SG}\boldsymbol{\mu}_{t+h|t} = \boldsymbol{\mu}_{t+h|t}$ when \mathbf{SG} is a projection matrix. \square

We note that the above result holds when the projection \mathbf{SG} is only onto the coherent subspace \mathfrak{s} and not for all projection matrices in general. To describe this more explicitly suppose \mathbf{SG} has as its image \mathfrak{L} which is itself a lower dimensional linear subspace of \mathfrak{s} , i.e. $\mathfrak{L} \subset \mathfrak{s}$. Then for $\{\boldsymbol{\mu}_{t+h|t} : \boldsymbol{\mu}_{t+h|t} \in \mathfrak{s}, \boldsymbol{\mu}_{t+h|t} \notin \mathfrak{L}\}$, $\mathbf{SG}\boldsymbol{\mu}_{t+h|t} \neq \boldsymbol{\mu}_{t+h|t}$. This is depicted in Figure 5 where $\boldsymbol{\mu}_{t+h|t}$ is projected to a point $\bar{\boldsymbol{\mu}}$ in \mathfrak{L} . Therefore in this case, the reconciled forecast will have as its expectation $\bar{\boldsymbol{\mu}}$ rather than $\boldsymbol{\mu}_{t+h|t}$ and be biased. This result has implications in practice, in particular, the top-down method (Gross & Sohl 1990) has

$$\mathbf{G} = \begin{pmatrix} \mathbf{p} & \mathbf{0}_{(m \times n-1)} \end{pmatrix} \quad (7)$$

where $\mathbf{p} = (p_1, \dots, p_m)'$ is an m -dimensional vector consisting a set of proportions which is use to disaggregate the top-level forecasts along the hierarchy. In this case it can be verified that \mathbf{SG} is idempotent, i.e. $\mathbf{SGSG} = \mathbf{SG}$ and therefore \mathbf{SG} is a projection matrix. However the image of this projection is not an m -dimensional subspace but a 1-dimensional subspace. As such, top-down reconciliation will bias base forecasts when those base forecasts are unbiased.

5 Bias correction

6 Application

7 Conclusions

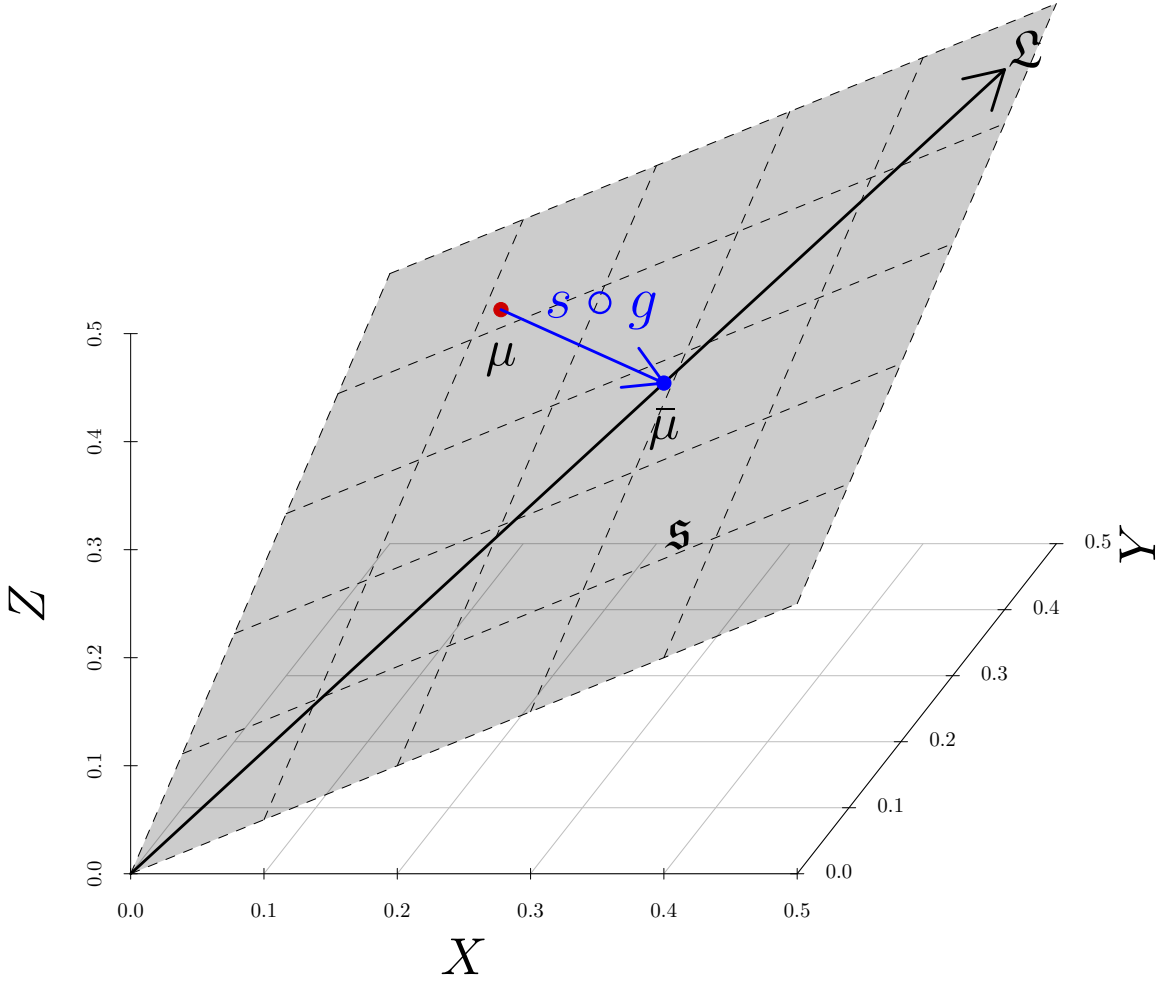


Figure 5: \mathcal{L} is a linear subspace of the coherent subspace \mathfrak{s} . If $s \circ g$ is a projection not onto \mathfrak{s} but onto \mathcal{L} , then $\mu \in \mathfrak{s}$ will be moved to $\bar{\mu} \in \mathcal{L}$.

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.
- 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.