Forecasts Reconciliation: A geometric view with new insights on bias correction.

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

August 22, 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

The past decade has seen rapid development in methodologies for forecasting time series that follow a hierarchical aggregation structure. Of particular prominence have been *forecast reconciliation* methods involving two steps; first separate forecasts are produced for all series, then these are adjusted ex post to ensure coherence with aggregation constraints. Forecast reconciliation has mostly been formulated using a regression model, see Hyndman et al. (2011) and Wickramasuriya et al. (2018) for examples. This setup can be counter-intuitive since a vector comprised of forecasts from different time series models is also assumed to be the dependent variable in a regression model. In this paper, we eschew a regression interpretation in favour of a novel, geometric understanding of forecast reconciliation. This allows us to develop novel proofs and a clearer understanding of the interplay between forecast bias and reconciliation methods.

Multivariate time series following an aggregation structure arise in many disciplines such as manufacturing, engineering, marketing and medicine. Forecasts of these series should adhere to aggregation constraints to ensure aligned decision making. Earlier studies achieved this by only forecasting a single level of the hierarchy and then either aggregating in a bottom up fashion (Dunn et al. 1976) or disaggregating in a top-down fashion (Gross & Sohl 1990, Athanasopoulos et al. 2009). For reviews of these approaches including a discussion of their advantages and disadvantages see Schwarzkopf et al. (1988), Kahn (1998), Lapide (1998), Fliedner (2001).

include references

In contrast to these methods, Hyndman et al. (2011) proposed forecasting all series in the hierarchy, referring to these as *base* forecasts. Since base forecasts were produced independently they were not guaranteed to adhere to aggregation constraints and could thus be improved via further adjustment. A framework was proposed whereby the base forecasts were assumed to follow a regression model. The predicted values from this model were guaranteed to adhere to the linear constraints by construction and could thus be used as a new set of forecasts. This approach and later modifications have subsequently been shown to outperform bottom up and top down approaches in a variety of empirical settings. references

Some theoretical insight into the performance of forecast reconciliation methods has been provided by Van Erven & Cugliari (2014) and Wickramasuriya et al. (2018). Both papers provide a proof that reconciliation is guaranteed to improve base forecasts. The latter paper also proposes a particular version of reconciliation known as the Minimum Trace (MinT) method. This is optimal in the sense of minimising the trace of reconciled forecast error covariance matrix under the assumption that base forecasts are unbiased.

Our main contribution is to propose a geometric interpretation of the entire hierarchical forecasting problem. In this setting, we show that reconciled forecasts will have a number of attractive properties when they are obtained via projections. We believe that this is clearer and more intuitive than explanations based on regression modelling, notwithstanding the fact that regression based-methods themselves are indeed projections. As such, this paper is in part a review of existing results cast in a new light, but one that we believe to be warranted as forecast reconciliation methodologies have become more popular. In addition, we also propose three major and novel results.

First, our approach makes it clear that the defining characteristic of so-called *hierarchical time series* is not aggregation but linear constraints. As a result forecast reconciliation can be applied in contexts where there are no clear candidates of *bottom level* series, an insight that is not apparent when the problem is viewed through the lens of regression modelling. Second, we provide a new proof that reconciled forecasts dominate unreconciled forecasts which makes explicit the link between a reconciliation method and a loss function. We believe that this link is lacking in previous work that attempts to establish similar results, in particular Van Erven & Cugliari (2014) and Wickramasuriya et al. (2018). Futhermore, unlike Van Erven & Cugliari (2014) and Wickramasuriya et al. (2018) our proof does not require an assumption about convexity. Third, we revisit the issue of bias. We prove that reconciliation using certain projection matrices guarantees unbiased reconciled forecasts as long as base forecasts are also unbiased. A natural question that arises is what to do in the case of biased reconciled forecasts. Rather than addressing this issue by considering matrices that are not projections, we propose to bias-correct before reconciliation. This is evaluated in an extensive empirical study where we find that even when bias correction fails, the extent of the problem is mitigated by reconciling forecasts.

The remainder of this paper is structured as follows. Section 2 deals with the concept of coherence and defines so called hierarchical time series in a way that does not depend on any notion of bottom level series. Section 3 defines forecast reconciliation in terms of projections and includes a proof that reconciled forecasts dominate base forecasts with respect to a specific loss function. In Section 4 we prove the unbiasedness preserving property of reconciliation via certain projection matrices and propose methods for bias correction. In Section 5 we evaluate these methods for in an extensive emprirical application to domestic tourism flow in Australia. Section 6 concludes with some discussion and thoughts on the future of research in forecast reconciliation.

2 Coherent forecasts

2.1 Notation and preliminaries

We briefly define the concept of a *hierarchical time series* in a fashion similar to Wickramasuriya et al. (2018), Hyndman & Athanasopoulos (2018) and others, before 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 $\boldsymbol{y}_t \in \mathbb{R}^n$ 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 the following holds for all t

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

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

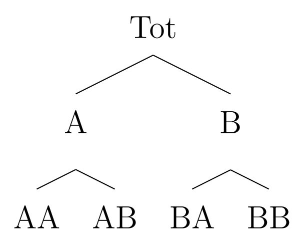

Figure 1: An example of a two level hierarchical structure.

To clarify these concepts consider the example of the hierarchy in Figure 1. For this hierarchy, $n = 7$, $\boldsymbol{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{C,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]'$, $m = 4$, $\boldsymbol{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]'$ 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 such a definition is completely serviceable, it obscures the full generality of the literature on so-called hierarchical time series. In fact, concepts such as coherence and reconciliation, defined in full below, only require the data to have two important characteristics; the first is that they are multivariate, the second is that they adhere to linear constraints.

2.2 Coherence

The property that data adhere to some linear constraints is referred to as *coherence*. We now provide definitions aimed at providing geometric intuition of hierarchical time series.

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 $\boldsymbol{y} \in \mathfrak{s}$ 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 $\boldsymbol{S}$, i.e. $\mathfrak{s} = \text{span}(\boldsymbol{S})$. 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 '$\boldsymbol{S}$ matrix' would have columns $(1, 0, 1)'$ and $(0, 1, -1)'$. However, while there are multiple ways to define an $\boldsymbol{S}$ matrix, in all cases the columns will span the same coherent subspace, which is unique.

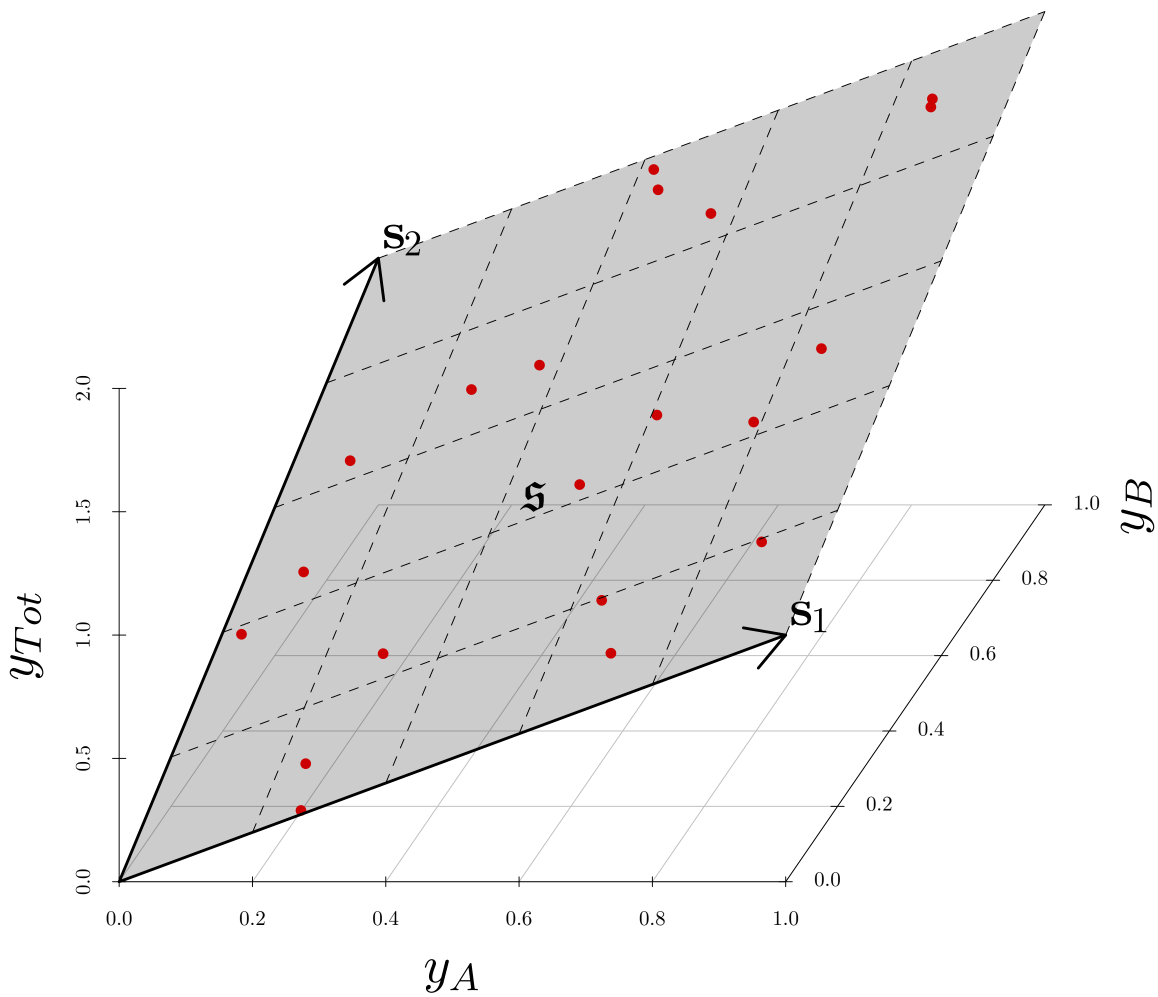

Figure 2: Depiction of a three dimensional hierarchy with $y_{\text{Tot}} = y_{\text{A}} + y_{\text{B}}$. The gray coloured two dimensional plane depicts the coherent subspace $\mathfrak{s}$ where $\vec{s}_1 = (1, 1, 0)'$ and $\vec{s}_2 = (1, 0, 1)'$ are basis vectors that spans $\mathfrak{s}$. The red points in $\mathfrak{s}$ represent 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 $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$ and all future values $\boldsymbol{y}_{T+1}, \boldsymbol{y}_{T+2}, \ldots$ lie in the coherent subspace, i.e. $\boldsymbol{y}_t \in \mathfrak{s} \quad \forall t$.

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. Also notable by its absence in the above definition is any reference to *aggregation*. In some ways, terms such as *hierarchical* and *aggregation* can be misleading since the literature has covered instances that cannot be depicted in a similar fashion to Figure1 and/or do not involve aggregation. **Include brief summary of all non-traditional hierarchies - e.g. grouped hierarchies, temporal hierarchies with weird overlapping, problems where we look at differences between variables etc.** Finally, although Definition 2.2 makes 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.

Definition 2.3 (Coherent Point Forecasts)**.** Let $\breve{\boldsymbol{y}}_{t+h|t} \in \mathbb{R}^n$ be a vector of point forecasts of all series in the hierarchy where the subscript $t+h|h$ implies that the forecast is made as time t for a period h steps into the future. Then $\breve{\boldsymbol{y}}_{t+h|t}$ is *coherent* if $\breve{\boldsymbol{y}}_{t+h|t} \in \mathfrak{s}$.

Without any loss of generality, the above definition could also be applied to prediction for multivariate data in general, rather than just forecasting of time series. Much of the early literature that dealt with the problem of forecasting hierarchical time series (see Gross & Sohl 1990, and references therein) produced forecasts at a single level of the hierarchy in the first stage. Subsequently forecasts for all series were recovered through aggregation, disaggregation according to historical or forecast proportions or some combination of both. As such incoherent forecasts were not a problem in these earlier papers.

Forecasting a single level of the hierarchy did not, however echo common practice within many industries. In many organisations different departments or 'silos' each produced their own forecasts, often with their own information sets and judgemental adjustments. This approach does have several advantages over only forecasting a single level. First, there is no loss of information since all bottom levels are modelled. Second, modelling top level series often identifies features such as trend and seasonality that cannot be detected in noisy disaggregate data. Unfortunately, however, when forecasts are produced independently at all levels, forecasts are likely to be incoherent. This problem of incoherent forecasts cannot in general be solved by multivariate modelling either. Instead, the solution to incoherent forecasts is to make an ex post adjustment that ensures coherence, a process known as *forecast reconciliation*

might need refs

3 Forecast reconciliation

The problem of forecast reconciliation is predicated on there being an n-vector of forecasts that are incoherent. We will call these *base forecasts* and denote them as $\hat{\boldsymbol{y}}_{t+h|h}$. In the sequel, this subscript will be dropped at times for ease of exposition. In the most general terms, reconciliation can be defined as follows

Definition 3.1 (Reconciled forecasts)**.** Let ψ be a mapping, $\psi : \mathbb{R}^n \to \mathfrak{s}$. The point forecast $\tilde{\boldsymbol{y}}_{t+h|t} = \psi\left(\hat{\boldsymbol{y}}_{t+h|t}\right)$ is said to "reconcile" a base forecast $\hat{\boldsymbol{y}}_{t+h|t}$ with respect to the mapping $\psi(.)$

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 $\boldsymbol{S}\boldsymbol{G}$, where $\boldsymbol{G}$ is an $(n-m) \times n$ matrix (some authors use $\boldsymbol{P}$ used in place of $\boldsymbol{G}$). This facilitates an interpretation of reconciliation as a two-step process, in the first step, base forecasts $\hat{\boldsymbol{y}}_{t+h|t}$ 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 $\boldsymbol{S}$.

Although pre-multiplying base forecasts by $\boldsymbol{SG}$ will result in coherent forecasts, a number of desirable properties arise when $\boldsymbol{SG}$ has the specific structure of a *projection* matrix onto $\mathfrak{s}$. In general a projection matrix is defined via the idemoptence property, i.e. $\boldsymbol{SG}^2 = \boldsymbol{SG}$. However a much more important property of projection matrices, used in multiple instances below, is that any vector lying in the image of the projection will be mapped to itself by that projection (see Lemma 2.4 in Rao 1974, for a proof). In our context this implies that for any $\boldsymbol{v} \in \mathfrak{s}$, $\boldsymbol{SGv} = \boldsymbol{v}$.

We begin by considering the special case of an orthogonal projection whereby $\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$. This is equivalent to so called OLS reconciliation as introduced by Hyndman et al. (2011). We refrain from any discussion of regression models focusing instead on geometric interpretations. However the connection between OLS and orthogonal projection should be clear, in the context of regression modelling predicted values from OLS are obtained via an orthogonal projection of the response onto the span of the regressors.

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 realised value targeted by the forecast.

To address the first of these properties we make the concept of closeness more concrete, by considering the Euclidean distance between the base forecast $\hat{\boldsymbol{y}}$ and the reconciled forecast $\tilde{\boldsymbol{y}}$. A property of an orthogonal projection is that the distance between $\hat{\boldsymbol{y}}$ and $\tilde{\boldsymbol{y}}$ is minimal for over any possible $\tilde{\boldsymbol{y}} \in \mathfrak{s}$. In this sense reconciliation via orthogonal projection leads to the smallest possible adjustments of the base forecasts.

The property that reconciliation should improve forecasts was touched upon in Section 2.3 of Wickramasuriya et al. (2018). The discussion in that paper focuses on the case of MinT. Here we provide a new explicit proof of that result. We do so first in the case of an orthogonal projection where the geometric intuition of the proof is clear and then generalise the result to reconciliation using any projection matrix in Section 3.2.

Consider the Euclidean distance between a forecast and the target. This is equivalent to the root of the sum of squared errors over the entire hierarchy. Let $\boldsymbol{y}_{t+h}$ be the realisation of the data generating process at time $t+h$. 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 $\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{SG}\hat{\boldsymbol{y}}_{t+h|t}$, where $\boldsymbol{G}$ is such that $\boldsymbol{SG}$ is an orthogonal (in the Euclidean sense) projection onto $\mathfrak{s}$ and let $\|\boldsymbol{v}\|$ be the L_2 norm (in the Euclidean sense) of vector $\boldsymbol{v}$ then:*

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

Proof. Since, $\boldsymbol{y}_{t+h}, \tilde{\boldsymbol{y}}_{t+h} \in \mathfrak{s}$ and since the projection is orthogonal, by Pythagoras' theorem

$$\|(\hat{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|^2 = \|(\tilde{\boldsymbol{y}}_{t+h|t} - \hat{\boldsymbol{y}}_{t+h})\|^2 + \|(\tilde{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|^2. \tag{3}$$

Since $\|(\tilde{\boldsymbol{y}}_{t+h|t} - \hat{\boldsymbol{y}}_{t+h})\|^2 \geq 0$ this implies,

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

with equality only holding when $\tilde{\boldsymbol{y}}_{t+h|t} = \hat{\boldsymbol{y}}_{t+h}$. Taking the square root of both sides proves the desired result. $\square$

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 $\hat{\boldsymbol{y}}$ is shown as a blue dot. Since $\hat{\boldsymbol{y}}$ is incoherent, $\hat{\boldsymbol{y}}_{t+h|t} \notin \mathfrak{s}$ and in this case the inequality is strict. Reconciliation is an orthogonal projection from $\hat{\boldsymbol{y}}$ to the coherent subspace yielding the reconciled forecast $\tilde{\boldsymbol{y}}$ shown in red. Finally, the target of the forecast $\boldsymbol{y}$ 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.

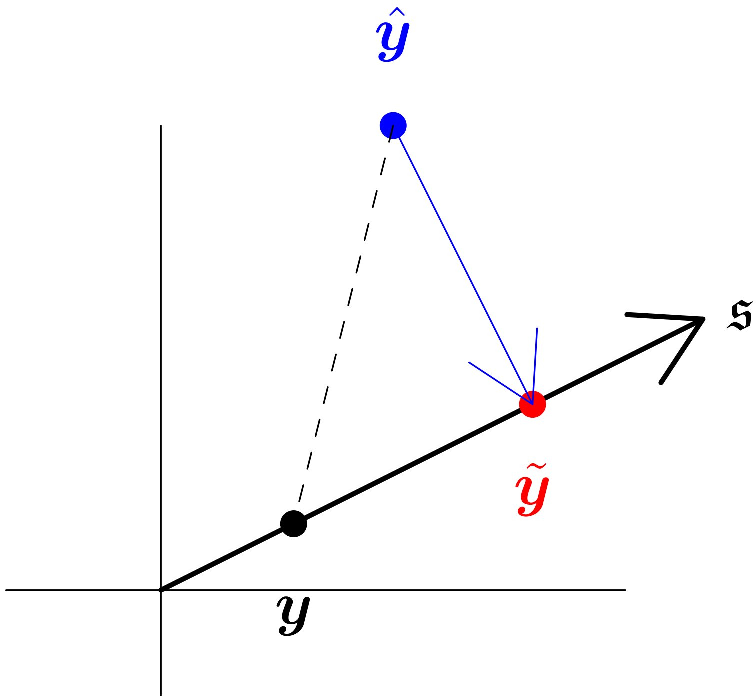

Figure 3: Orthogonal projection of $\hat{\boldsymbol{y}}$ onto $\mathfrak{s}$ yielding the reconciled forecast $\tilde{\boldsymbol{y}}$

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

Theorem 3.1 is in some ways more powerful than perhaps previously understood. Crucially, the result is not a result that requires taking expectations. This distance reducing property will hold for any realisation and any forecast and not just on average. Nothing needs to be assumed about the statistical properties of the data generating process or the process by which forecasts are made.

However, in other ways, Theorem 3.1 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, 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

One justification for using an orthogonal projection is that it leads to improved forecast accuracy in terms of the root of the sum of squared errors of *all* variables in the hierarchy. A clear shortcoming of this measure of forecast accuracy is that forecasts errors in all series should not necessarily be treated equally. For example, in hierarchies, top-level series tend to have a much larger scale than bottom level series. Even when two series are on a similar scale, series that are more predictable or less variable will tend to be downweighted by simply aggregating square errors. An even more sophisticated understanding may take the correlation between series into account. All of these considerations lead towards reconciliation of the form $\tilde{\boldsymbol{y}} = \boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}\hat{\boldsymbol{y}}$, where $\boldsymbol{W}$ is a symmetric matrix. Generally, it is assumed that $\boldsymbol{W}$ is invertible, otherwise a pseudo inverse can be used.

It should be noted that $\boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}$ is an oblique, rather than an orthogonal projection matrix in the usual Euclidean geometry. However this matrix can be considered to be an orthogonal projection for a different geometry defined by the norm $||\boldsymbol{v}||_{\boldsymbol{W}^{-1}} = \boldsymbol{v}'\boldsymbol{W}^{-1}\boldsymbol{v}$, which we will call the generalised Euclidean geometry with respect to $\boldsymbol{W}^{-1}$. One way to understand this geometry is that it is the same as Euclidean geometry when all vectors are first transformed by pre-multiplying by $\boldsymbol{W}^{-1/2}$. This leads to a transformed $\boldsymbol{S}$ matrix $\boldsymbol{S}^* = \boldsymbol{W}^{-1/2}\boldsymbol{S}$ and transformed $\hat{\boldsymbol{y}}$ and $\tilde{\boldsymbol{y}}$ vectors $\hat{\boldsymbol{y}}^* = \boldsymbol{W}^{-1/2}\hat{\boldsymbol{y}}$ and $\tilde{\boldsymbol{y}}^* = \boldsymbol{W}^{-1/2}\tilde{\boldsymbol{y}}$. The transformed reconciled forecast results from an orthogonal projection in the transformed space since

$$\begin{aligned}
\tilde{\boldsymbol{y}}^* &= \boldsymbol{W}^{-1/2}\tilde{\boldsymbol{y}} && (5)\\
&= \boldsymbol{W}^{-1/2}\boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}\hat{\boldsymbol{y}} && (6)\\
&= \boldsymbol{S}^*\left(\boldsymbol{S}^{*\prime}\boldsymbol{S}^*\right)^{-1}\boldsymbol{S}^{*\prime}\hat{\boldsymbol{y}}^* && (7)
\end{aligned}$$

Thinking of the problem in terms of a geometry defined by the norm $\boldsymbol{v}'\boldsymbol{W}^{-1}\boldsymbol{v}$ is also quite instructive when it comes to thinking about the connection between distances and loss functions. In the generalised Euclidean geometry, the distance between the reconciled forecast and the realisation is given by $(\hat{\boldsymbol{y}} - \boldsymbol{y})'\boldsymbol{W}^{-1}(\hat{\boldsymbol{y}} - \boldsymbol{y})$. For diagonal $\boldsymbol{W}^{-1}$, this is equivalent to a weighted sum of squared error loss function and when $\boldsymbol{W}$ is a covariance matrix, this is equivalent to a Mahalanobis distance. As such Theorem 3.1 can easily be generalised as follows:

Theorem 3.2 (General distance reducing property)**.** *If* $\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{t+h|t}$*, where* $\boldsymbol{G}$ *is such that* $\boldsymbol{SG}$ *is an orthogonal (in the generalised Euclidean sense) projection onto* $\mathfrak{s}$ *then:*

$$\|(\tilde{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_{\boldsymbol{W}^{-1}} \leq \|(\hat{\boldsymbol{y}}_{t+h|t} - \boldsymbol{y}_{t+h})\|_{\boldsymbol{W}^{-1}}. \tag{8}$$

Proof. The proof is identical to the proof for Theorem 3.1 but relies on the generalised Pythagorean Theorem (applicable to Generalised Euclidean space) rather than the Pythagorean Theorem. □

The implication of Theorem 3.2 is that if the objective function is some weighted sum of squared errors, or a Mahalanobis distance, then the projection matrix $\boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}$ is guaranteed to improve forecast accuracy over base forecasts, for an appropriately selected $\boldsymbol{W}$.

Note here that we rely here on the generalised Pythagorean Theorem (which involves an equality). In contrast, Wickramasuriya et al. (2018) follow Van Erven & Cugliari (2014) in stating their result in terms of the Generalised Pythagorean Inequality. The proof of Wickramasuriya et al. (2018) requires an assumptions about convexity so that the angle between the base forecast and coherent subspace must be greater than 90 degrees. The proof we have provided here requires no such assumption, since this may not hold for an arbitrary $\boldsymbol{W}$. As such the statement from Wickramasuriya et al. (2018) that *"MinT reconciled forecasts are at least as good as the incoherent forecasts"* should be qualified; this is only true in with respect to a loss function that depends on $\boldsymbol{W}$. If Euclidean distance (or mean squared error) is used, there will be cases where the MinT estimator does not improve upon base forecasts.

Discuss results in Figure 4 here

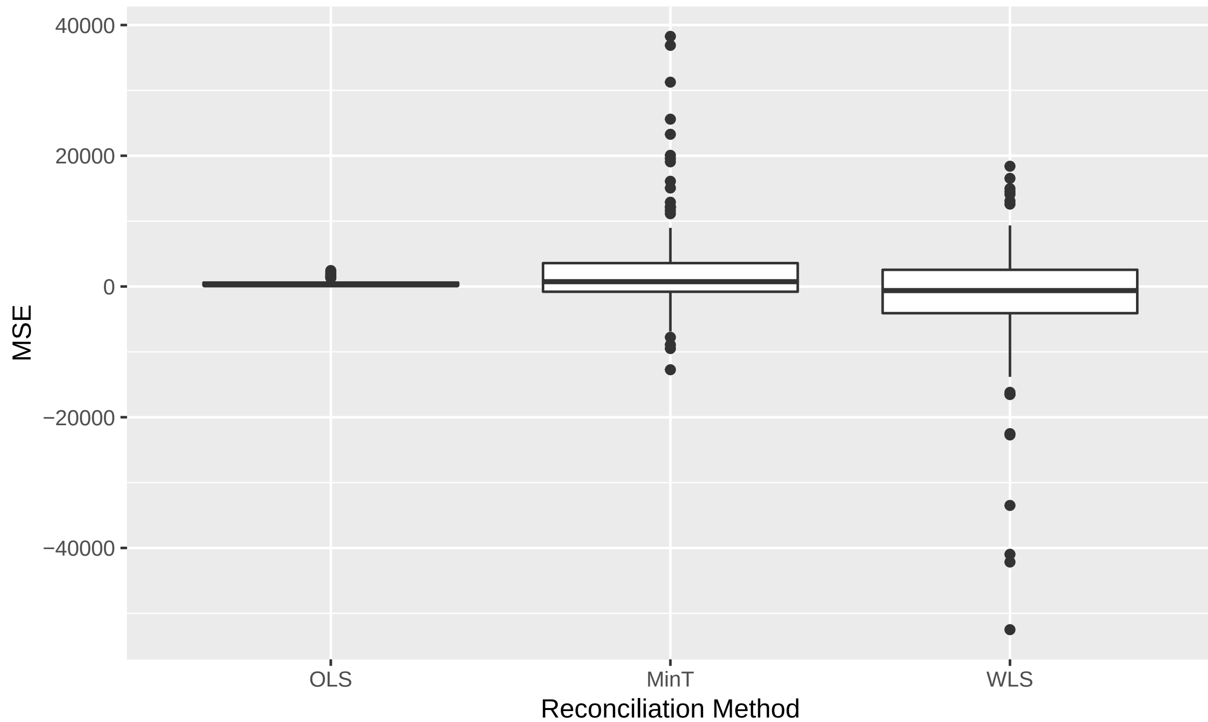

Figure 4: Difference of MSE between Base forecasts and reconciled forecasts over the replications is presented. Positive values of the difference implies reconciliation improves the forecast accuracy than base forecasts.

3.3 MinT

While the properties discussed so far hold for any projection matrix, the MinT method of Wickramasuriya et al. (2018) has an additional optimality property. Wickramasuriya et al. (2018) show that for unbiased base forecasts, the trace of the forecast error covariance matrix of reconciled forecasts is minimised by an oblique projection with a particular choice of $\boldsymbol{W}$. This choice is that $\boldsymbol{W}$ should be the forecast error covariance matrix where errors come from using the base forecasts. Although the base forecast error covariance matrix is unknown, it can be estimated using in-sample errors.

Figure **??** provides geometrical intuition into the MinT method. Suppose the in-sample errors are given by the orange points. They provide information on the most likely direction of large deviations from the coherent subspace. This direction is denoted by $\boldsymbol{R}$. Figure **??** then shows a target value of $\boldsymbol{y}$, while the grey points indicate possible values for the base forecasts (the base forecasts are of course stochastic). One possible value of the forecast is depicted in blue as $\hat{\boldsymbol{y}}$. An oblique projection of the blue point back along the direction of $\boldsymbol{R}$ yields a reconciled forecast closer to the target, especially compared to an orthogonal projection showed as in Figure **??**. Figure **??** depicts a similar oblique projection along $\boldsymbol{R}$ for all the gray points yield reconciled forecasts tightly packed near the target $\boldsymbol{y}$. In this sense, the oblique MinT projection minimises the forecast error variance of reconciled forecasts. In contrast to the result in Theorem 3.2, this property is a statistical property in the sense that MinT is optimal in expectation.

4 Bias in forecast reconciliation

Before turning our attention to the issue of bias itself it is important to state a sensible property that any reconciliation method should have. That is if base forecasts are already coherent then reconciliation should not change the forecast. As stated in Section 3, this property holds only when $\boldsymbol{SG}$ is a projection matrix. As a corollary, reconciling using an arbitrary $\boldsymbol{G}$, may in fact change an already coherent forecast.

The property that projections map all vectors in the coherent subspace onto themselves is also useful in proving the unbiasedness preserving property of reconciliation Wickramasuriya et al. (2018). Before restating this proof using a clear geometric interpretation we

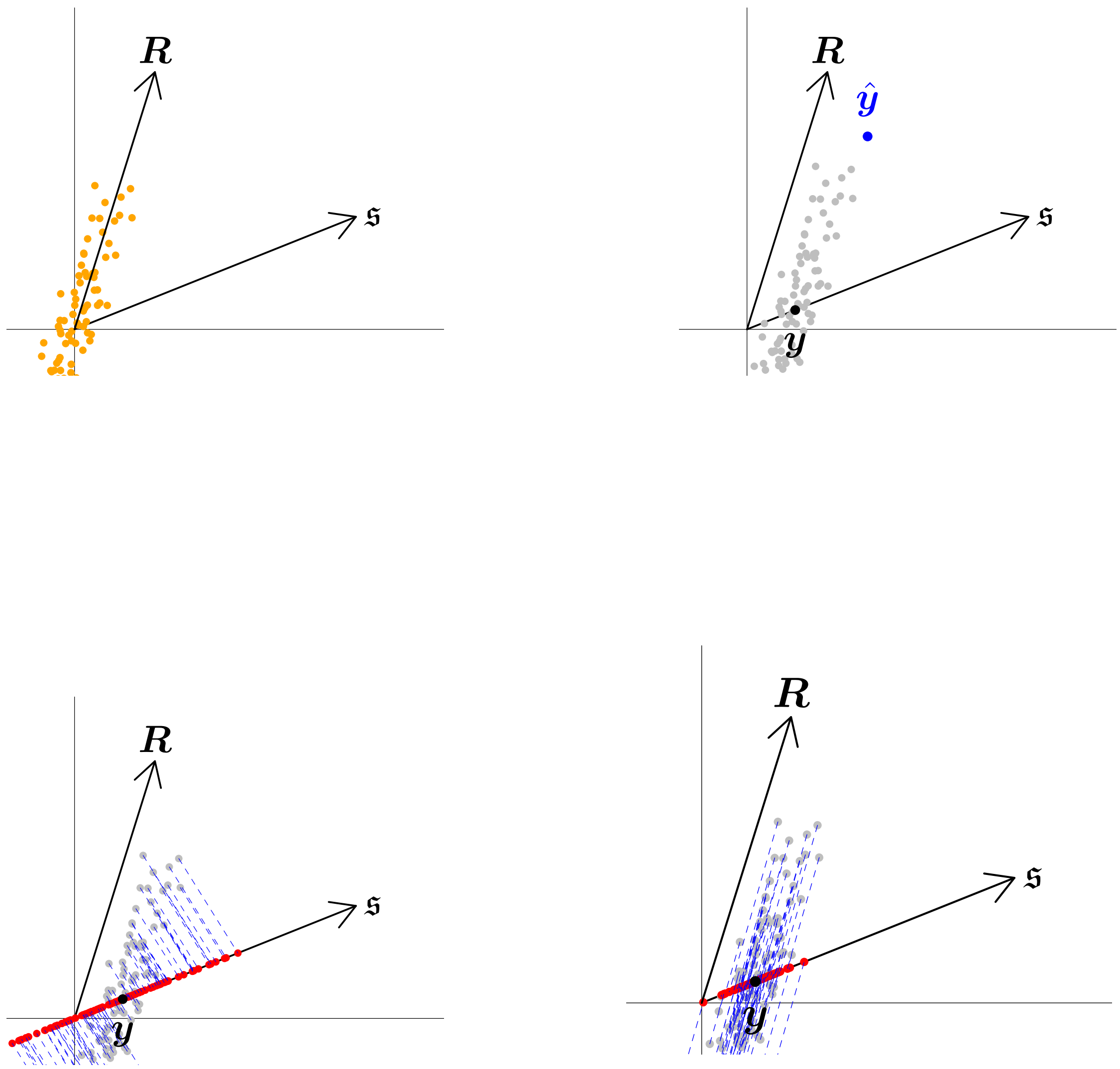

Figure 5: A schematic to represent orthogonal and oblique reconciliation. Points in orange colour in top left figure represent the insample errors. $\boldsymbol{R}$ shows the most likely direction of deviations from the coherent subspace $\mathfrak{s}$. Grey points in top right figure indicate potential realisations of the base forecast while the blue dot $\hat{\boldsymbol{y}}$ indicates one such realisation. The black dot $\boldsymbol{y}$ denotes the (unknown) target of the forecast. Bottom left figure shows the orthogonal projection of all potential realisations onto the coherent subspace while bottom right figure shows the oblique projection.

discuss in a precise fashion what is meant by unbiasedness.

Suppose that the target of a point forecast is $\boldsymbol{\mu}_{t+h|t} := \mathrm{E}(\boldsymbol{y}_{t+h} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$ where the expectation is taken over the predictive density. Our point forecast can be thought of as an estimate of this quantity. The forecast is random due to uncertainty in the training sample and it is with respect to this uncertainty that unbiasedness refers. More concretely, the point forecast will be unbiased if $\mathrm{E}_{1:t}(\hat{\boldsymbol{y}}_{t+h|t}) = \boldsymbol{\mu}_{t+h|t}$, where the subscript $1:t$ denotes an expectation taken over the training sample.

or do we want to define unbiasedness of a forecast as the expected value of the forecast equals realisation?

Theorem 4.1 (Unbiasedness preserving property)**.** *For unbiased $\hat{\boldsymbol{y}}_{t+h|t}$, the reconciled point forecast is also an unbiased prediction as long as $\boldsymbol{SG}$ 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 $\boldsymbol{\mu}_{t+h|t}$ is an expectation taken with respect to the degenerate predictive density it must lie in $\mathfrak{s}$. We have already established that when $\boldsymbol{SG}$ is a projection onto $\mathfrak{s}$ then it maps all vectors in $\mathfrak{s}$ onto themselves. As such $\boldsymbol{SG}\boldsymbol{\mu}_{t+h|t} = \boldsymbol{\mu}_{t+h|t}$ when $\boldsymbol{SG}$ is a projection matrix. □

We note that the above result holds when the projection $\boldsymbol{SG}$ has the coherent subspace $\mathfrak{s}$ as its image and not for all projection matrices in general. To describe this more explicitly suppose $\boldsymbol{SG}$ 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 $\left\{\boldsymbol{\mu}_{t+h|t} : \boldsymbol{\mu}_{t+h|t} \in \mathfrak{s}, \boldsymbol{\mu}_{t+h|t} \notin \mathfrak{L}\right\}$, $\boldsymbol{SG}\boldsymbol{\mu}_{t+h|t} \neq \boldsymbol{\mu}_{t+h|t}$. This is depicted in Figure 6 where $\boldsymbol{\mu}_{t+h|t}$ is projected to a point $\bar{\boldsymbol{\mu}}$ in $\mathfrak{L}$. In this case, the expectation of reconciled forecast will be $\bar{\boldsymbol{\mu}}$ rather than $\boldsymbol{\mu}_{t+h|t}$ and hence biased.

This result has implications in practice. The top-down method (Gross & Sohl 1990) has

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{p} & \boldsymbol{0}_{(m\times n-1)} \end{pmatrix} \tag{9}$$

where $\boldsymbol{p} = (p_1, \dots, p_m)'$ is an m-dimensional vector consisting a set of proportions used to disaggregate the top-level forecast. In this case it can be verified that $\boldsymbol{SG}$ is idempotent, i,e. $\boldsymbol{SGSG} = \boldsymbol{SG}$ and therefore $\boldsymbol{SG}$ 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 produces biased forecasts even when the base forecasts are unbiased.

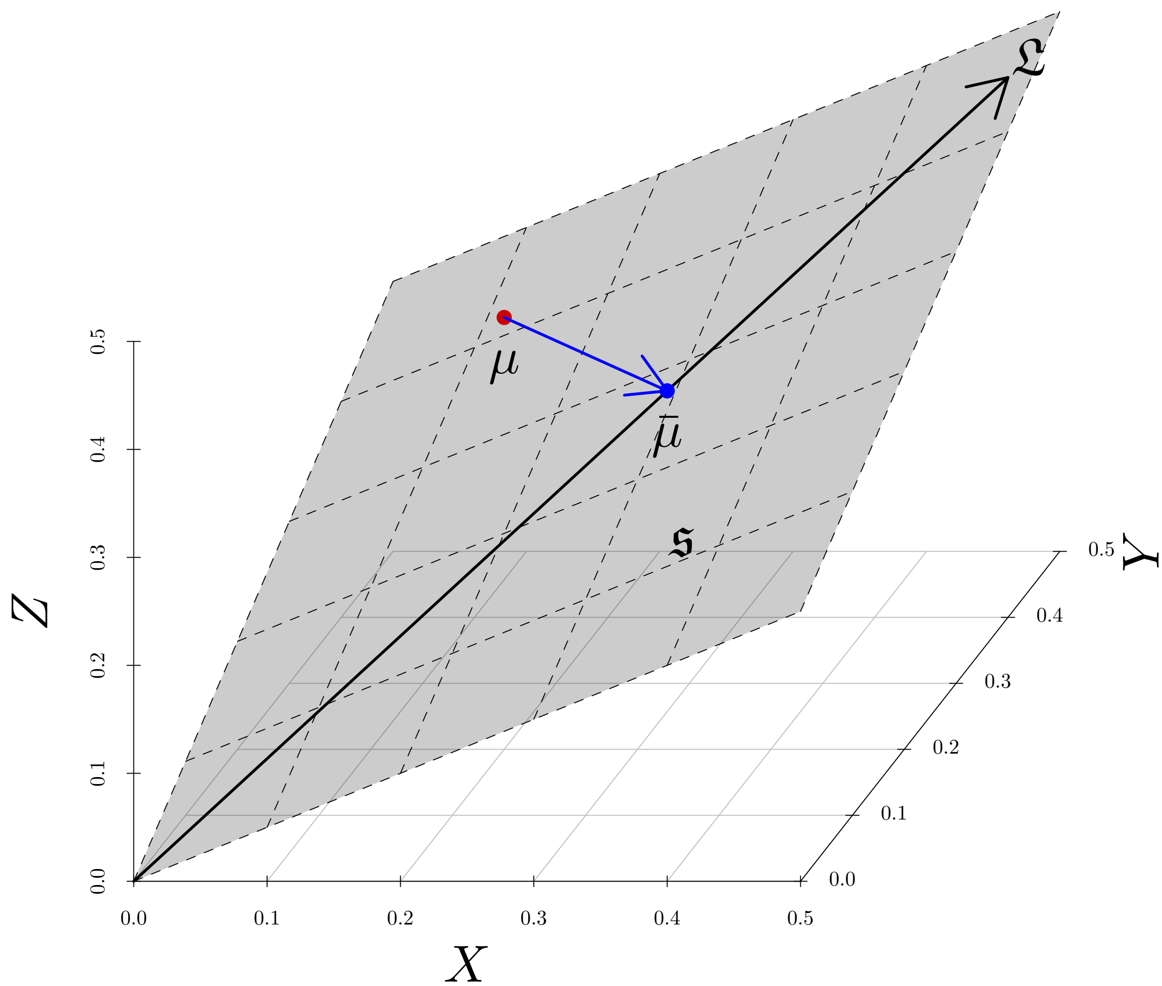

Figure 6: $\mathfrak{L}$ is a linear subspace of the coherent subspace $\mathfrak{s}$. If $s \circ g$ is a projection not onto $\mathfrak{s}$ but onto $\mathfrak{L}$, then $\boldsymbol{\mu} \in \mathfrak{s}$ will be moved to $\bar{\boldsymbol{\mu}} \in \mathfrak{L}$.

Finally, it is often stated that an assumption required to prove the unbiasedness preserving property is that $\boldsymbol{SGS} = \boldsymbol{S}$ or alternatively that $\boldsymbol{GS} = \boldsymbol{I}$. Both of these conditions are equivalent to assuming that $\boldsymbol{SG}$ is a projection matrix. However, problems arise when viewing the preservation of unbiasedness through the prism of imposing a constraint $\boldsymbol{GS} = \boldsymbol{I}$. This thinking suggests that a way to deal with biased forecasts is to select $\boldsymbol{G}$ in an unconstrained manner. However, equipped with a geometric understanding of the problem, we would advise against this approach. The constraint $\boldsymbol{GS} = \boldsymbol{I}$ is not just about bias and dropping the constraint compromises all of the attractive properties of projections. Opening the door to reconciliation methods that change already coherent base forecast would seem to suggest an increase in the variability of the forecast. This seems particularly perverse when the motivation for using a biased method in the first place is to reduce variance.

perhaps elaborate in a proof in appendix

Our own solution to dealing with biased forecasts is to bias correct *before* reconciliation. In many cases the method for bias correction will be context specific. For instance, in our empirical study in Section 5 we consider a scenario where bias is induced via taking a Box-Cox transformation before modelling. In this well-known case a number of bias correction methods exist. Our particular choice of bias correction will be the Guererro method (Guerrero 1993).

Alternatively, a more general purpose approach to bias correction is to simply estimate the bias by taking the sample mean of $\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h|t}$ for all $t+h$ in the training sample. This can be then subtracted from future forecasts. As stated in the discussion of MinT, in-sample errors are already used to estimate the optimal direction of projection. As such we see no problems with using the same errors to bias correct. Geometrically, the intuition is simple. In top left panel in Figure 5, the orange points are centered around the origin as would be expected from an unbiased forecast. If forecasts are biased, then errors should simply be translated until they are centered at the origin.

5 Empirical study

Using an empirical application to forecast Australian domestic tourism flow, we illustrate the usefulness of projection based reconciliation in practice. Previous studies have pronounced that the reconciliation improves point forecast accuracy in domestic tourism flow in Australia (Athanasopoulos et al. (2009), Hyndman et al. (2011), Wickramasuriya et al. (2018)). However, our motivation in this study is to demonstrate how the bias correction methods discussed in previous section along with the projection-based reconciliation help to improve the forecast accuracy.

5.1 Data

We use "overnight trips" as a measure of domestic tourism flow in Australia. Total "overnight trips" in Australia can be disaggregated into 7 states, 27 zones and 75 regions forming a geographical hierarchy with 110 total number of series and 75 bottom level series. More information about the series of this hierarchy is given in Table 2 in the appendix. Data were obtained from the National Visitor Survey (NVS) which were collected through telephone interviews from an annual sample of $120,000$ Australian residents aged 15 years or more. Data form a monthly series starting from January 1998 to December 2017 which gives a total of 240 observations per series.

As we can see from the time plots in Figure 7, the total overnight trips has a strong seasonal pattern. Most overnight trips seems to be happening in January which is associated with the summer season in southern hemisphere. On the other hand the least trips are happening in February perhaps due to the end of summer holidays. We could also observe a slight downward trend in total overnight trips from 2005 to 2010 and then started increasing since 2010.

Figure 8 shows time plots for some selected series from states, zones and regions of the hierarchy. While observing the series in state level, we could see that Northern Territory(NT) has a different seasonal patterns compared to the other states. Highest overnight trips to NT happens during June - August where July being the peak. Lowest trips happens during December - February. This is perhaps Australians prefer to visit North of Australia during its dry season rather than in wet season. This pattern also reflect in zones and regions associated to NT as well as regions in north Queensland. All other states and their associated zones and regions show peak overnight trips in January.

Further we noticed that the signal-to-noise ratio decreases with the level of disaggregation of the hierarchy. We can see much noisier series in regional and zone levels whereas the noise is comparatively stabilised in the upper levels. We have also observed that the overnight trips to 'Adelaide Hills' has an anomalous observation on December-2002. We replace this observation with the average overnight trips on December-2001 and December-2003 for the same destination.

Using this data we produce $h = 1$ to $h = 6$ month ahead forecasts based on a rolling window approach. First a training window of 100 observations is considered from Jan-1998 to Apr-2006 and forecasts produce for May-2006 to Oct-2006. Then we roll the training window one month at a time where the final forecast is produced for Dec-2017. This leads to 140 1-step-ahead, 139 2-step-ahead through 135 6-step-ahead forecasts left for evaluation.

5.2 Base forecasts and reconciliation:

First we transform each series in the hierarchy to stabilise any variations in the data. We use two types of transformation, namely, log transformation and more general Box-Cox transformation. Equation (10) gives the mathematical formula for these transformations.

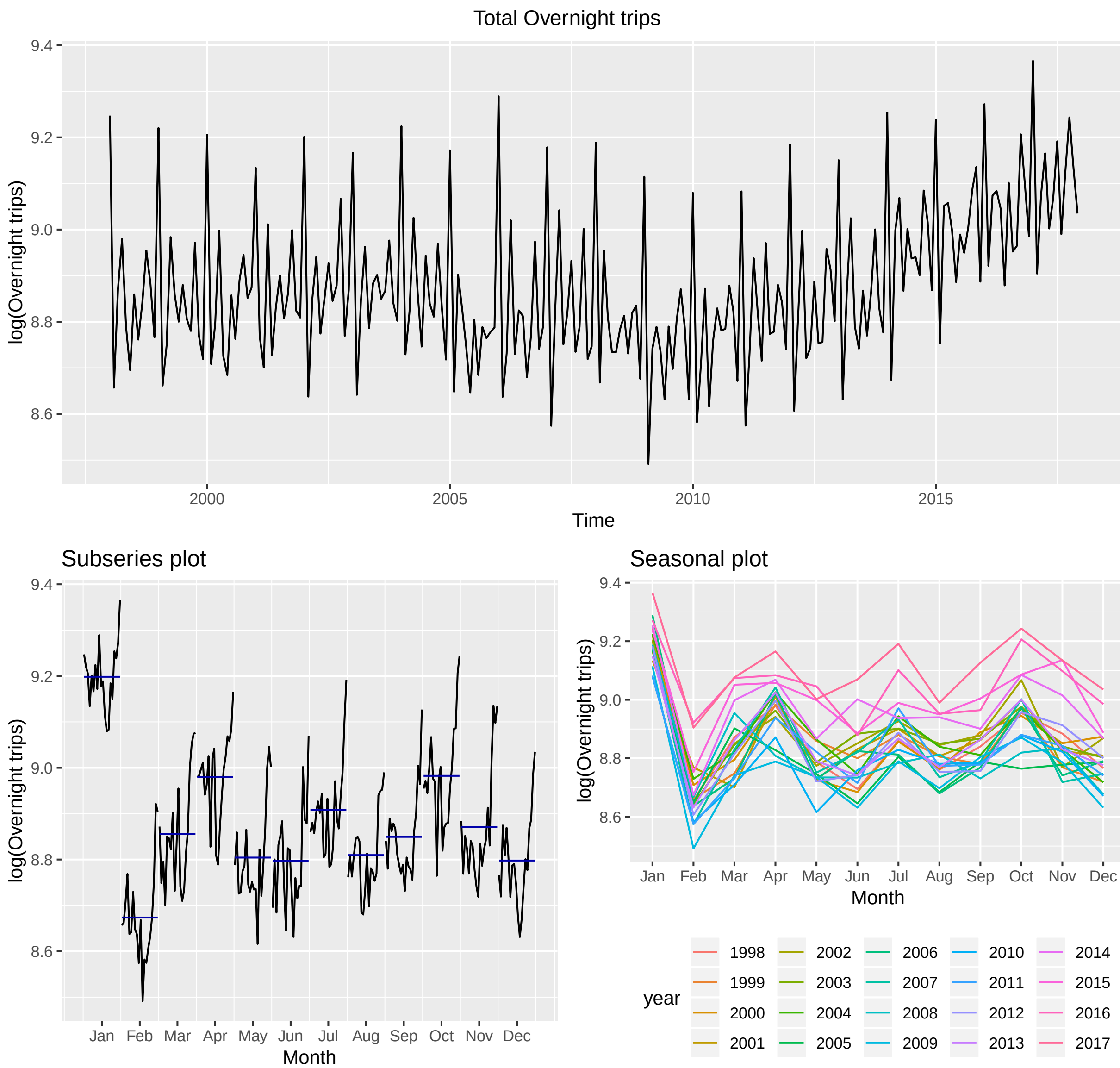

Figure 7: Total overnight trips are presented in log scale. Top panel shows the time plot. Bottom left panel shows the sub-series plot for each month whereas bottom right panel shows the seasonal plot coloured by year

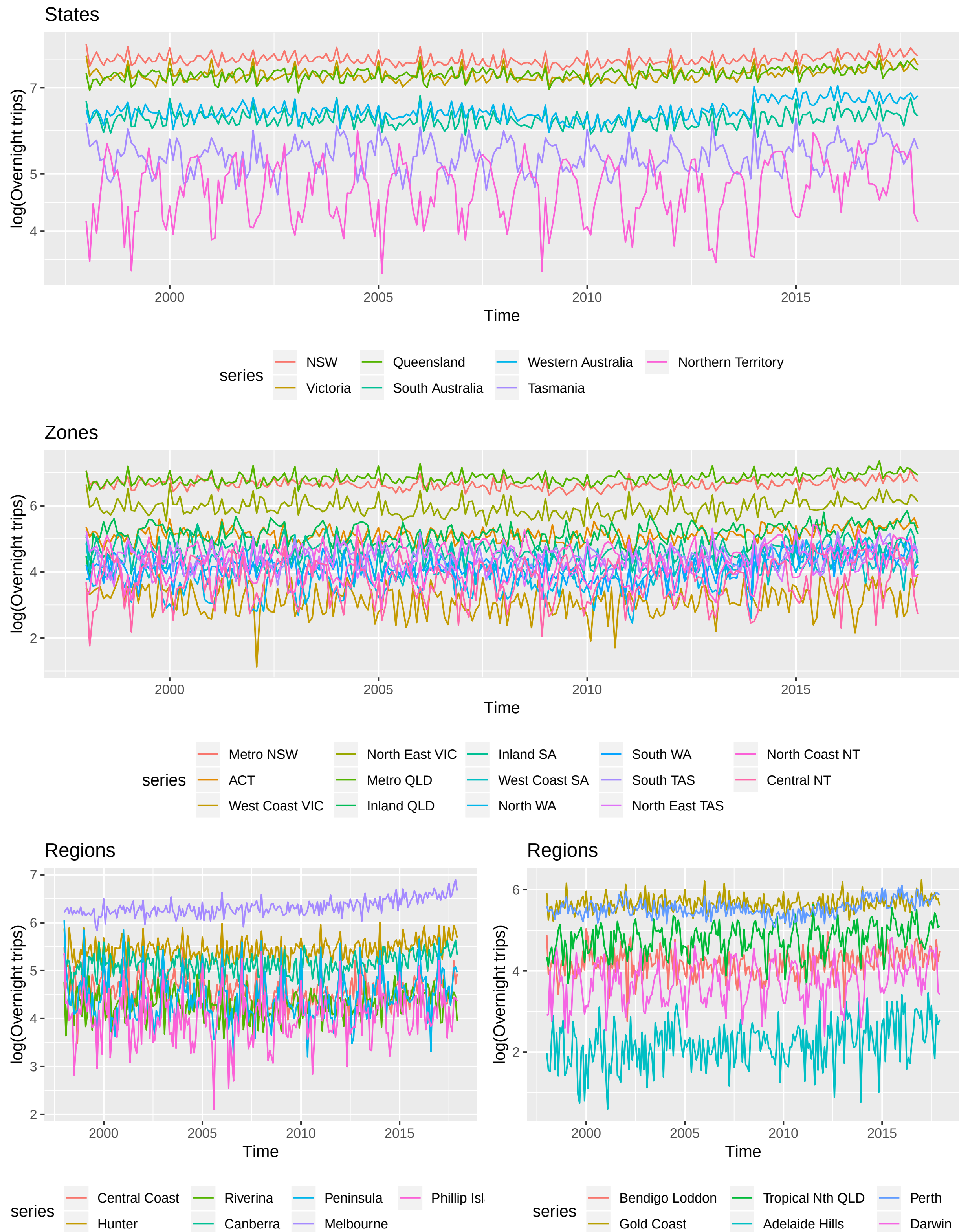

Figure 8: Time plot of overnight trips for some selected series from different disaggregate levels of the hierarchy. All values are presented in log scale. To avoid impact from the zero values we added a constant 1 to all observations

A constant is added to each series to avoid the impact of zero observations when taking log transformations. We note here that the two transformations were applied separately to the series in the hierarchy and treated distinctly.

$$x_t = \begin{cases} \log(y_t) & \text{if log transformation, } \lambda = 0 \\ \frac{y_t^\lambda - 1}{\lambda} & \text{if Box-Cox transformation, } \lambda \neq 0 \end{cases} \tag{10}$$

The "Guerrero" method (Guerrero 1993) implemented in `BoxCox.lambda()` function in `forecast` package in R software (Hyndman et al. 2019) is use to select optimal λ for Box-Cox transformation. For log transformation we simply choose $\lambda = 0$. Then we fit univariate ARIMA models for the transformed series. `auto.arima()` function in `forecast` package is use to choose the best model that minimises the AIC. Using the fitted models, forecasts are produced for 6 forecast horizons for each series in the hierarchy. These forecasts are then back-transformed using equation (11) to scale them back into the original space. However, the back transformed forecasts will be biased (Guerrero 1993).

$$y_{t+h} = \begin{cases} \exp(x_{t+h}) & \text{if log transformation, } \lambda = 0 \\ (\lambda x_{t+h} + 1)^{1/\lambda} & \text{if Box-Cox transformation, } \lambda \neq 0 \end{cases} \tag{11}$$

Hence the reconciled forecasts that follow from these forecasts will also be biased. This is the exact scenario that we want to demonstrate in this study. As proposed before, we first adjust for the bias in base forecasts prior to the reconciliation. We use the de-biasing factor proposed by Taylor (1986) to calculate the back-transformed bias-adjusted forecasts. These are given in equation (12) .

$$y_{t+h} = \begin{cases} \exp(x_{t+h})[1 + \frac{\sigma_h^2}{2}] & \text{if log transformation, } \lambda = 0 \\ (\lambda x_{t+h} + 1)^{1/\lambda}[1 + \frac{\sigma_h^2(1-\lambda)}{2(\lambda x_{t+h}+1)^2}] & \text{if Box-Cox transformation, } \lambda \neq 0 \end{cases} \tag{12}$$

where, x_{t+h} is the h-step-ahead forecasts from the log/Box-Cox transformed series and σ_h^2 is the variance of x_{t+h}. Unbiased forecasts from this method is referred to as Unbiased (Method-1). We also applied a more general bias adjustment method by estimating the bias using in-sample residual means as explained in Section 4. Unbiased forecasts from this general method is referred to as Unbiased (Method-2).

Should we give more concise words for these two methods??

Once we get the base forecasts for all series in the hierarchy, we can do the projection-based reconciliation to get coherent forecasts. We reconcile both biased and bias-corrected (Unbiased) base forecasts to demonstrate how the reconciliation follows from unbiased base forecasts improves the forecast accuracy. MinT(Shrink), WLS, OLS and bottom-up methods are used for reconciliation.

5.3 Results and discussion

Mean Squared Error (MSE) is used to measure the forecast accuracy. Average MSE over 140 replications for 1-step-ahead forecasts are presented for log transformation and Box-Cox transformation separately in Table 1.

Let us first consider the results from log transformation. We see that the unbiased forecasts from Method-1 has less MSE compared to the biased forecasts in base level. This holds for all reconciled forecasts as well. It implies that the bias correction has improved the forecast accuracy. Further, unbiased reconciled forecasts from OLS and MinT perform better than the base forecasts. Moreover, Unbiased MinT reconciled forecasts are outperforming. We also note that the bias correction via Method-2 has not necessarily

Table 1: Average MSE($\times 10^3$) of base and reconciled 1-step-ahead point forecasts are presented for log transformation and Box-Cox transformation. Unbiased(Method-1) follows from the bias adjustment via Taylor's de-biasing factor whereas Unbiased(Method-2) follows from residual mean adjustment.

	Log Transformation			Box-Cox Transformation		
R.method	Biased	Unbiased (Method-1)	Unbiased (Method-2)	Biased	Unbiased (Method-1)	Unbiased (Method-2)
Base	12.06	11.87	12.52	12.27	139.94	13.45
Bottom-up	17.37	15.47	21.58	18.06	15.76	23.33
OLS	11.59	11.41	11.98	11.84	116.91	12.96
WLS	15.00	13.83	17.55	15.83	14.20	19.51
MinT(Shrink)	9.36	**9.25**	10.35	10.27	**10.11**	12.43

work well in this empirical example.

Now turning to the results from Box-Cox transformation, we see that the bias correction via Method-1 has worsen the base forecasts. The same result holds for OLS reconciled forecasts as well. We also observed that this has happened because the bias correction is doing worst for some replications in "Total" series as depicted in Figure 9. This because for these replications the selected model has a drift term with a very large standard error. This will blow up the bias-adjusted forecasts.

However reconciled forecast from, Bottom-up, WLS or MinT has mitigate the effect from this unusual result in unbiased base forecasts. Most importantly, Unbiased MinT is outperforming all forecasts.

We have presented results only for one forecast horizon. However, results for other forecast horizons follows the same conclusions.

6 Conclusions

Defining concepts such as coherence and reconciliation in geometric terms provides new insights into forecast reconciliation methods. We have also provided evidence that reconciliation, particularly using the MinT method, can mitigate the effect of poor bias correction. Our intention in proposing a geometric interpretation is also to provoke research into new areas. We now discuss three such possibilites.

First, it should be possible to extend to the concept of coherence to non-linear constraints. In these cases the coherent space may need to be defined by a manifold. Although much more challenging, it is still possible define reconciled forecasts in terms of projections onto a manifold. Second, since we have established that the concept of bottom-level series is not crucial in forecast reconciliation an open question is whether is may be better

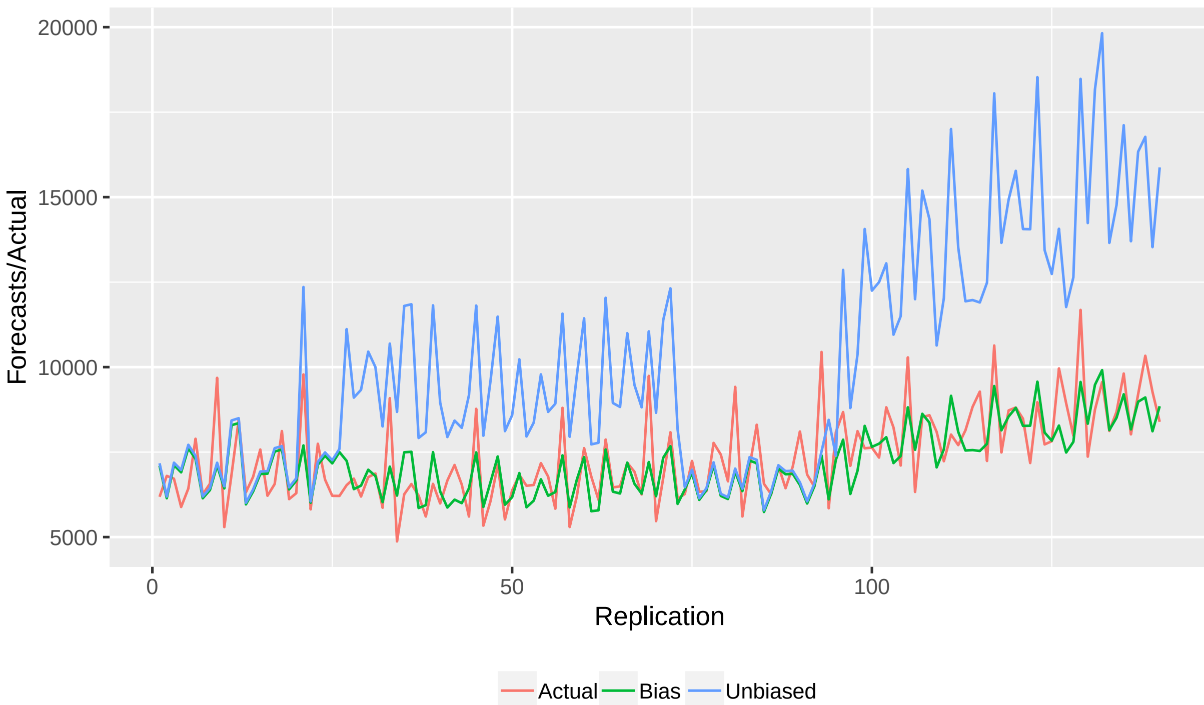

Figure 9: Actual values vs the base forecasts for Total overnight trips follows from Box-Cox transformation

to construct base forecasts of linear combinations of the time series rather than the time series themselves. Finally, the geometric interpretations of hierarchical forecast reconciliation facilitates an extension into a probabilistic framework, an issue that we investigate in a separate paper.

7 Appendix

Table 2: Geographical hierarchy of Australian tourism flow

Level 0 - Total			*Zones cont.*			*Regions cont.*		
1	Tot	Australia	48	AED	Blue Mountains	99	FAA	Hobart and South
Level 1 - States			49	AFA	Canberra	100	FBA	East Coast
2	A	NSW	50	BAA	Melbourne	101	FBB	Launceston, Tamar & North
3	B	Victoria	51	BAB	Peninsula	102	FCA	North West
4	C	Queensland	52	BAC	Geelong	103	FCB	Wilderness West
5	D	South Australia	53	BBA	Western	104	GAA	Darwin
6	E	Western Australia	54	BCA	Lakes	105	GAB	Kakadu Arnhem
7	F	Tasmania	55	BCB	Grippsland	106	GAC	Katherine Daly
8	G	Northern Territory	56	BCD	Phillip Island	107	GBA	Barkly
Level 2 - Zones			57	BDA	Central Murray	108	GBB	Lasseter
9	AA	Metro NSW	58	BDB	Goulburn	109	GBC	Alice Springs
10	AB	North Coast NSW	59	BDC	High Country	110	GBD	MacDonnell
11	AC	South Coast NSW	60	BDD	Melbourne East			
12	AD	South NSW	61	BDE	Upper Yarra			
13	AE	North NSW	62	BDF	Murray East			
14	AC	ACT	63	BEA	Wimmera+Mallee			
15	BA	Metro VIC	64	BEB	Western Grampians			
16	BB	West Coast VIC	65	BEC	Bendigo Loddon			
17	BC	East Coast VIC	66	BED	Macedon			
18	BC	North East VIC	67	BEE	Spa Country			
19	BD	North West VIC	68	BEF	Ballarat			
20	CA	Metro QLD	69	BEG	Central Highlands			
21	CB	Central Coast QLD	70	CAA	Gold Coast			
22	CC	North Coast QLD	71	CAB	Brisbane			
23	CD	Inland QLD	72	CAC	Sunshine Coast			
24	DA	Metro SA	73	CBA	Central Queensland			
25	DB	South Coast SA	74	CBB	Bundaberg			
26	DC	Inland SA	75	CBC	Fraser Coast			
27	DD	West Coast SA	76	CBD	Mackay			
28	EA	West Coast WA	77	CCA	Whitsundays			
29	EB	North WA	78	CCB	Northern			
30	EC	South WA	79	CCC	Tropical North Queensland			
31	FA	South TAS	80	CDA	Darling Downs			
32	FB	North East TAS	81	CDB	Outback			
33	FC	North West TAS	82	DAA	Adelaide			
34	GA	North Coast NT	83	DAB	Barossa			
35	GB	Central NT	84	DAC	Adelaide Hills			
Level 2 - Regions			85	DBA	Limestone Coast			
36	AAA	Sydney	86	DBB	Fleurieu Peninsula			
37	AAB	Central Coast	87	DBC	Kangaroo Island			
38	ABA	Hunter	88	DCA	Murraylands			
39	ABB	North Coast NSW	89	DCB	Riverland			
40	ABC	Hunter	90	DCC	Clare Valley			
41	ACA	South Coast	91	DCD	Flinders Range and Outback			
42	ADA	Snowy Mountains	92	DDA	Eyre Peninsula			
43	ADB	Capital Country	93	DDB	Yorke Peninsula			
44	ADC	The Murray	94	EAA	Australia's Coral Coast			
45	ADD	Riverina	95	EAB	Experience Perth			
45	AEA	Central NSW	96	EAC	Australia's South West			
46	AEB	New England North West	97	EBA	Australia's North West			
47	AEC	Outback NSW	98	ECA	Australia's Golden Outback			

References

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

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.

Fliedner, G. (2001), 'Hierarchical forecasting: issues and use guidelines', *Industrial Management & Data Systems* **101**(1), 5–12.

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

Guerrero, V. M. (1993), 'Time-series analysis supported by power transformations', *Journal of Forecasting* **12**(1), 37–48.

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*, OTexts.

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F., R Core Team, Ihaka, R., Reid, D., Shaub, D., Tang, Y. & Zhou, Z. (2019), *forecast: Forecasting Functions for Time Series and Linear Models*. Version 8.5.
URL: *https://CRAN.R-project.org/package=forecast*

Kahn, K. B. (1998), 'Revisiting top-down versus bottom-up forecasting'.
URL: *http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=985713&lang=pt-br&site=ehost-live*

Lapide, L. (1998), 'A simple view of top-down vs bottom-up forecasting', *Journal of Business Forecasting Methods & Systems* **17**, 28–31.

Rao, C. R. (1974), 'Projectors, generalized inverses and the blue's', *Journal of the Royal Statistical Society: Series B (Methodological)* **36**(3), 442–448.

Schwarzkopf, A. B., Tersine, R. J. & Morris, J. S. (1988), 'Top-down versus bottom-up forecasting strategies.', *nternational Journal of Production Research* **26**(11), 1833.

Taylor, J. M. (1986), 'The retransformed mean after a fitted power transformation', *Journal of the American Statistical Association* **81**(393), 114–118.

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', *Journal of the American Statistical Association* **1459**, 1–45.