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 18, 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 methodology for forecasting time series that adhere to linear aggregation constraints. 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 the linear constraints. Forecast reconciliation has mostly been formulated as a regression model, see Hyndman et al. (2011) and Wickramasuriya et al. (2018) for examples. This setup can be counter-intuitive since a vector of forecasts coming 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. In doing so, we review and provide clearer insight into existing approaches, and achieve a clearer understanding of the interplay between forecasts bias and reconciliation methods.

Multivariate time series that follow 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 (Gross & Sohl 1990, Athanasopoulos et al. 2009) fashion. For reviews of these approaches see Schwarzkopf et al. (1988), Kahn (1998), Lapide (1998), Fliedner (2001).

include references

In contrast to these approaches 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 be improved via further adjustment. A framework was proposed whereby the 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 good 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.

While reviewing the existing studies in this literature, our attempt is to give a new understanding to the whole hierarchical forecasting problem with a geometric intuition. We first provide a geometric interpretation of coherency and reconciliation in terms of projections. Then we rehash all existing reconciliation methods into a single framework based on projections. In addition to being highly intuitive, this allows us to establish new theoretical results. We prove two new theorems about point forecast reconciliation, the first showing that reconciled forecasts dominate unreconciled forecasts via the distance reducing property of projections while the second shows that reconciliation via projections preserves the unbiasedness of base forecasts.

Unbiasedness of base forecasts is a strong assumption in many of the existing methods which will be limited only to a selective number of forecasting methods. We propose a simple solution to this barrier. That is to correct the bias in base forecasts before doing the projection-based reconciliation. This allows using any advanced statistical method despite the unbiasedness to produce base forecasts without worrying about the unbiasedness. Using an empirical study, we further show that the reconciled forecasts follow from biased corrected base forecasts will improve the forecast accuracy compared to that of the biased forecasts.

The remainder of this paper is structured as follows. In Section 2 coherent point forecasting is defined geometrically while section 3 contains definitions of projection-based reconciliation as well as our main theoretical results. Section 4 discuss the unbiased preserving property of projection-based reconciliation and Section **??** contains our proposed method to correct the bias in base forecasts. In Section 5 we apply these methods to get coherent forecasts for domestic tourism flow in Australia and Section 6 concludes with some discussion and thoughts on future research.

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.

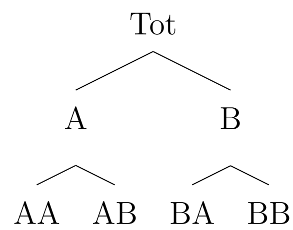

Figure 1: An example of a two level hierarchical structure.

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

$$\boldsymbol{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ \multicolumn{4}{c}{\boldsymbol{I}_4} \end{pmatrix},$$

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.

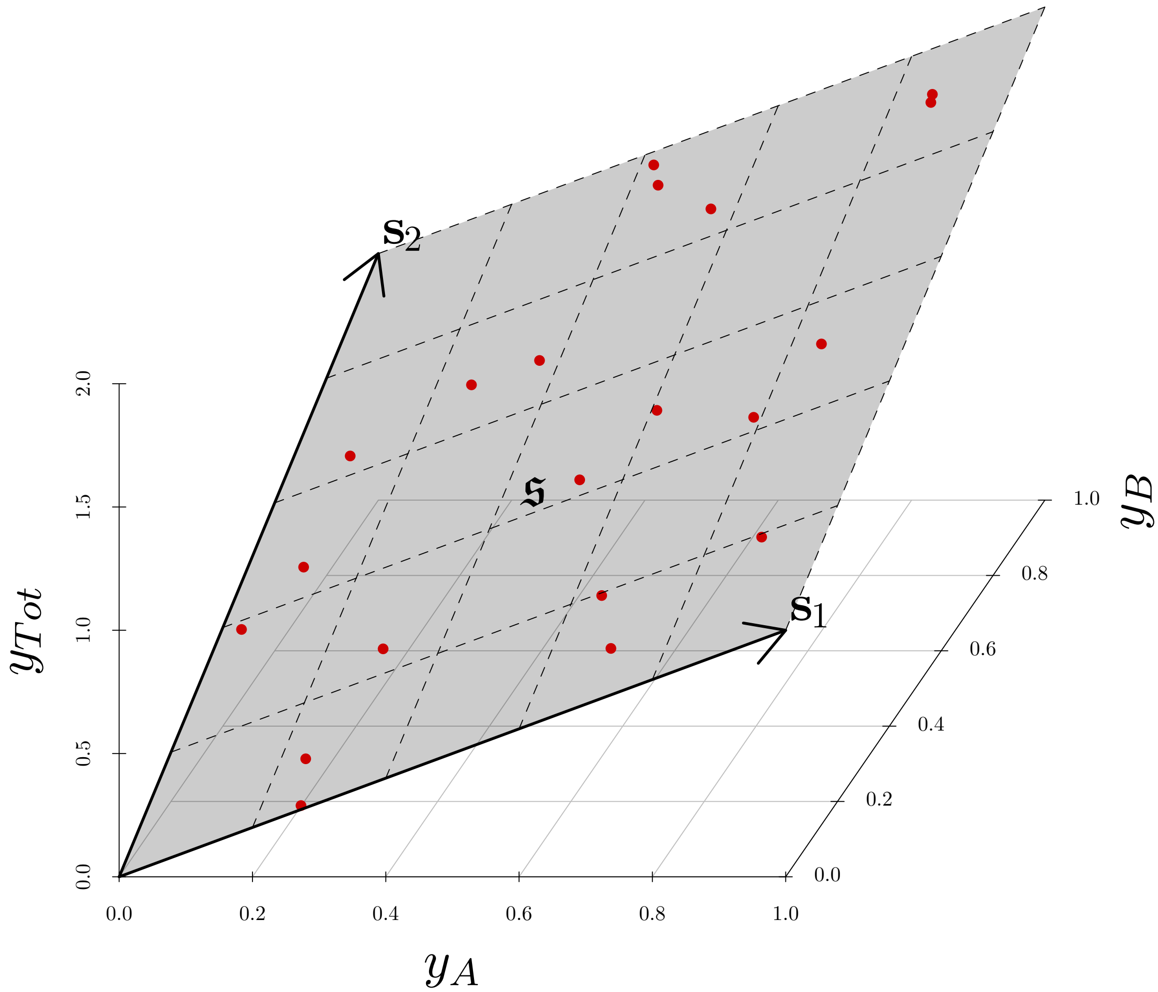

Figure 2: Depiction of a three dimensional hierarchy with $y_{\text{Tot}} = y_{\text{A}} + y_{\text{B}}$. The gray coloured two dimensional plane depicts the coherent subspace $\mathfrak{s}$ where $\vec{s}_1 = (1, 1, 0)'$ and $\vec{s}_2 = (1, 0, 1)'$ are basis vectors that spans $\mathfrak{s}$. The red 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 Figure 1 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, $\psi : \mathbb{R}^n \rightarrow \mathfrak{s}$. 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 $\boldsymbol{SG}$, 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 $\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{t+h|t}$, where $\boldsymbol{G}$ is such that $\boldsymbol{S}\boldsymbol{G}$ 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.

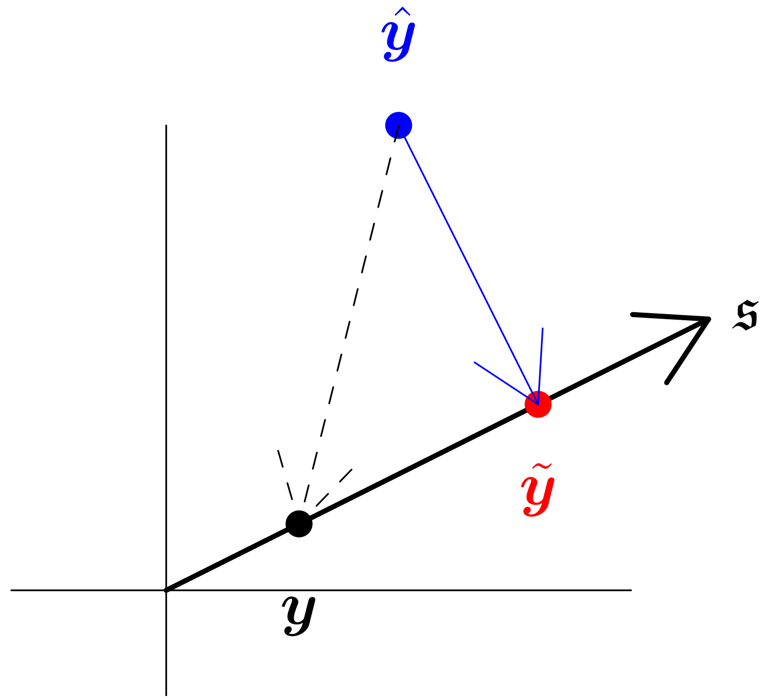

Figure 3: Orthogonal projection of $\hat{\boldsymbol{y}}$ onto $\mathfrak{s}$ yielding the reconciled forecast $\tilde{\boldsymbol{y}}$

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

$$\tilde{\boldsymbol{y}}^* = \boldsymbol{W}^{-1/2}\tilde{\boldsymbol{y}} \tag{5}$$

$$= \boldsymbol{W}^{-1/2}\boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}\hat{\boldsymbol{y}} \tag{6}$$

$$= \boldsymbol{S}^*\left(\boldsymbol{S}^{*'}\boldsymbol{S}^*\right)^{-1}\boldsymbol{S}^{*'}\hat{\boldsymbol{y}}^* \tag{7}$$

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 $\boldsymbol{S}\boldsymbol{G}$ 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 11 here

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 4 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 5 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 6. Figure 7 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.

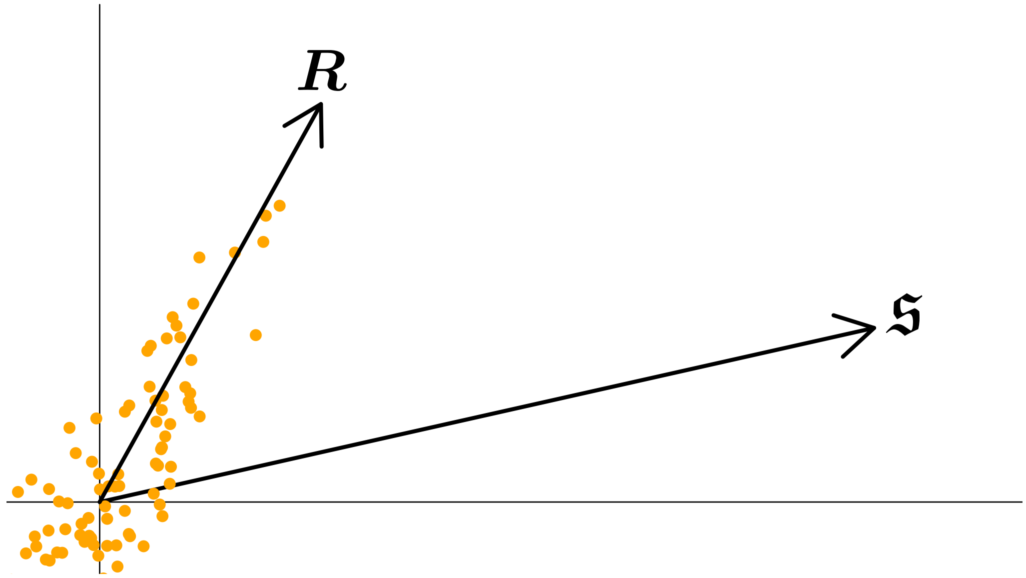

Figure 4: A schematic to represent MinT reconciliation. Points in orange colour represent the insample errors. $\boldsymbol{R}$ shows the most likely direction of deviations from the coherent subspace. $\hat{\boldsymbol{y}}$ is projected onto $\mathfrak{s}$ along the the direction of $\boldsymbol{R}$.

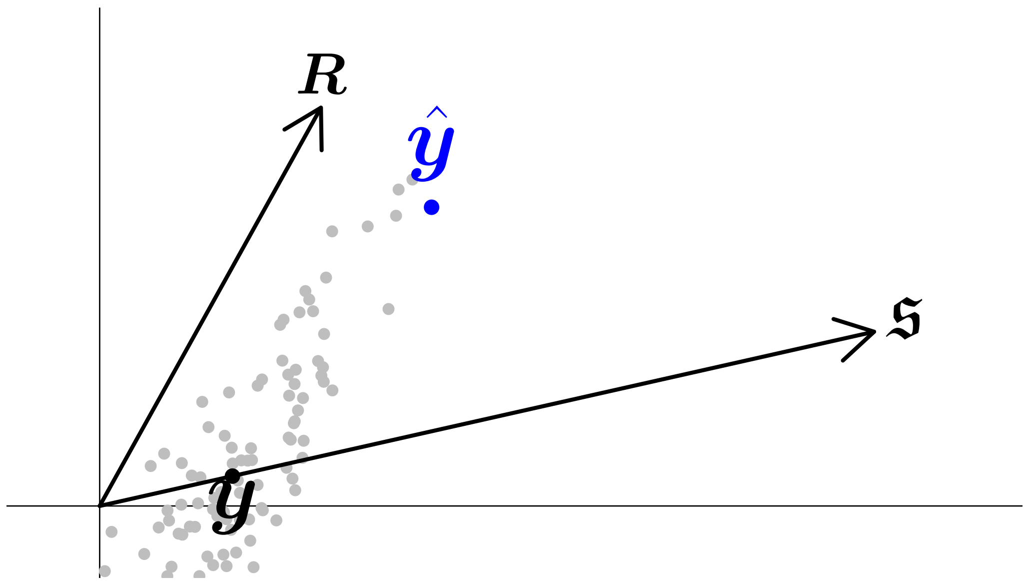

Figure 5: A schematic to represent MinT reconciliation. Grey points indicate potential realisations of the base forecast while the blue dot indicates one such realisation. The black dot $\boldsymbol{y}$ denotes the (unknown) target of the forecast.

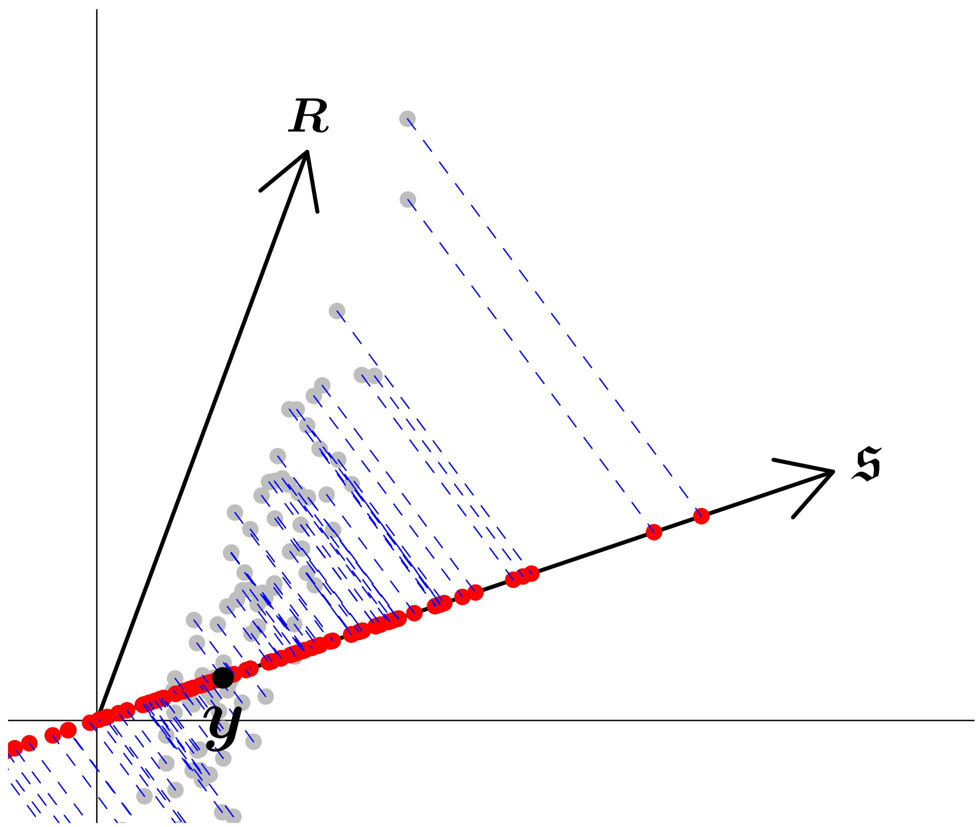

Figure 6: All gray points are orthogonally projected onto the coherent subspace $\mathfrak{s}$. Reconciled forecasts marked in red dots are widely spread about the target $\boldsymbol{y}$.

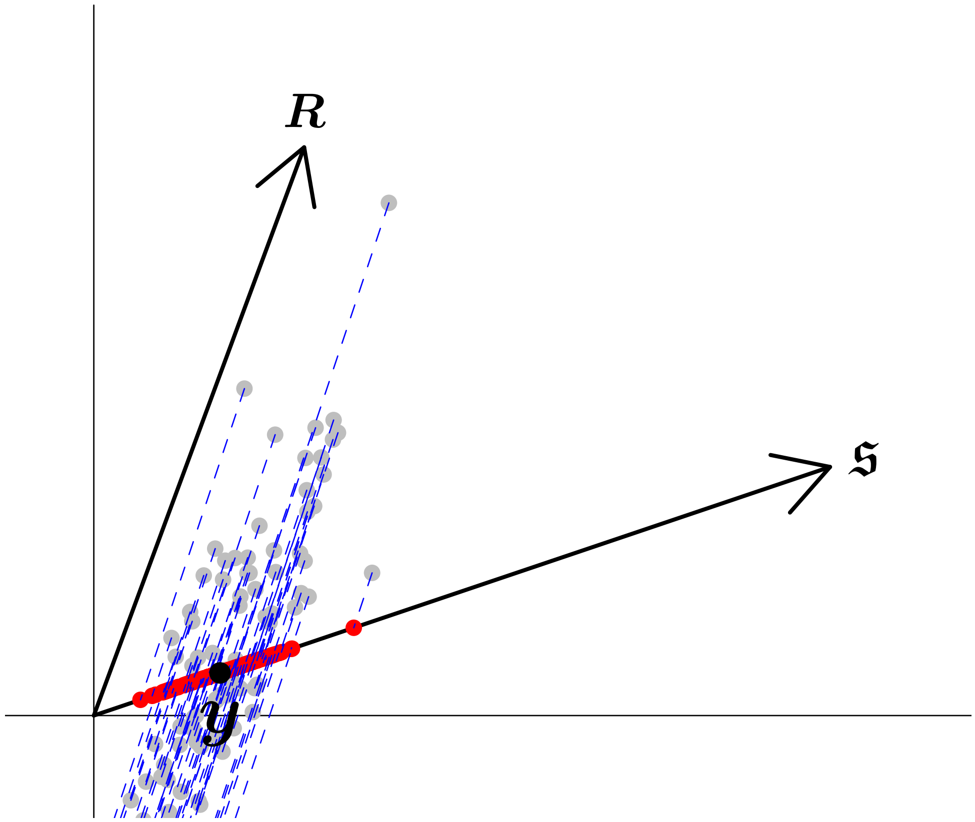

Figure 7: All gray points are obliquely projected along the direction of $\boldsymbol{R}$. Reconciled points denoted in red dots are concentrated about the target $\boldsymbol{y}$.

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 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. $\square$

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 8 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.

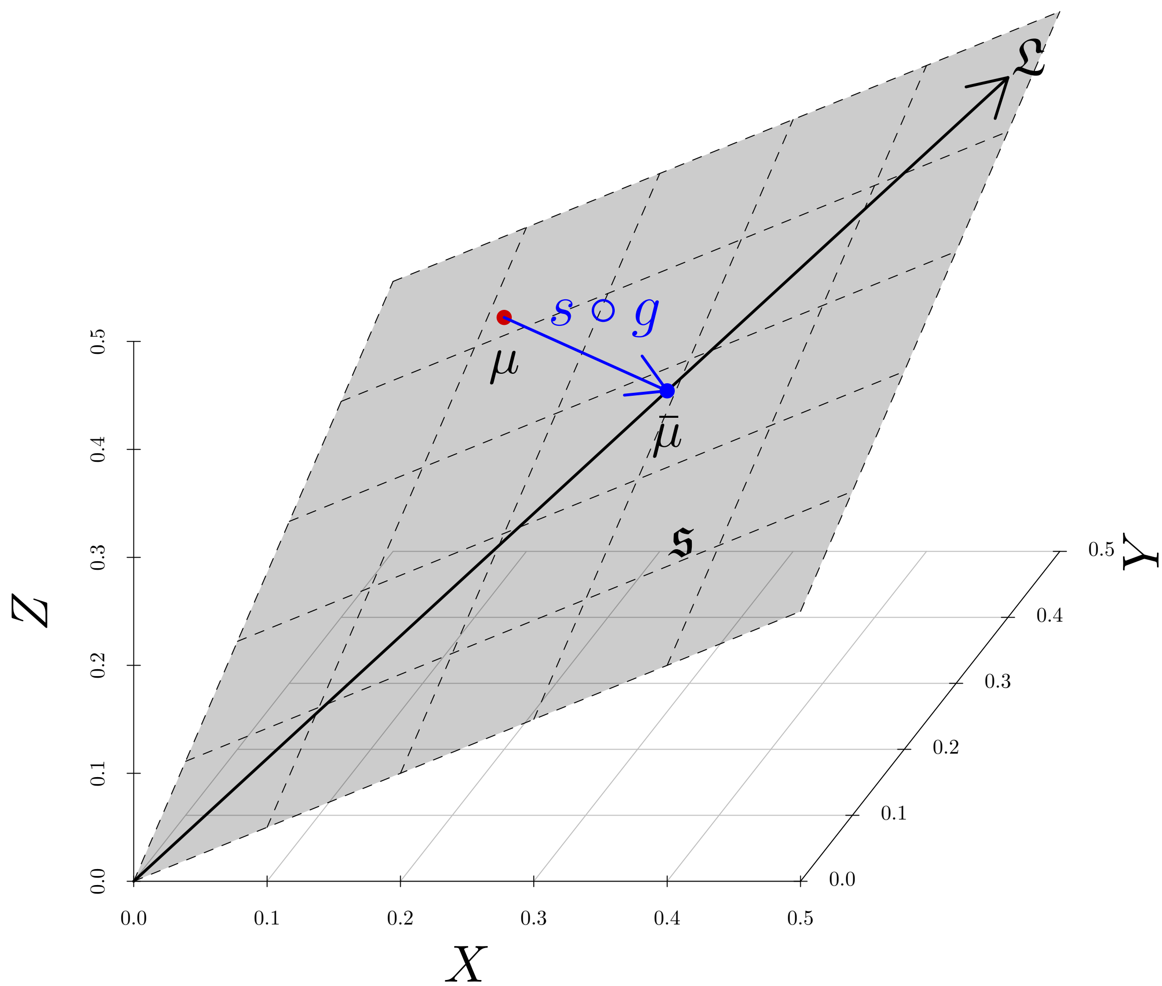

Figure 8: $\mathfrak{L}$ is a linear subspace of the coherent subspace $\mathfrak{s}$. If $s \circ g$ is a projection not onto $\mathfrak{s}$ but onto $\mathfrak{L}$, then $\boldsymbol{\mu} \in \mathfrak{s}$ will be moved to $\bar{\boldsymbol{\mu}} \in \mathfrak{L}$.

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 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 Figure 4, 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

Total tourism flow of any country can naturally disaggregate along a geographical hierarchy. Producing accurate forecasts for these hierarchical time series while preserving the coherency is important for making align business decisions in the tourism industry. Many applications related to Australian tourism flow have pronounced that reconciliation improves the point forecast accuracy (Athanasopoulos et al. (2009), Hyndman et al. (2011), Wickramasuriya et al. (2018)).

Using a similar empirical application to tourism forecasting in Australia, we show how to use projection-based reconciliation when we have a set of biased incoherent forecasts. As we mentioned in the previous section our focus is to bias adjust the incoherent forecasts first and then do the reconciliation.

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.

We produce $h = 1$ to $h = 6$ months ahead forecasts using a rolling window approach. First training window of 100 observations is considered from Jan-1998 to Apr-2006 and produce forecasts 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 directs to 140 1-step-ahead, 139 2-step-ahead through 135 6-step-ahead forecasts left for evaluation.

Preliminary analysis:

While observing the patterns of individual series, we noticed that the signal-to-noise ratio decreases with the disaggregate level of the hierarchy. From Figure 9 and 10 this is depicted well as we see much noisier series in bottom levels whereas the noise level is comparatively stabilised in the upper levels. We have also observed that the overnight trips for 'Adelaide Hills' has an anomalous observation on December-2002. We replaced this observation with the average overnight trips on December-2001 and December-2003 for the same destination.

Base forecasts and reconciliation:

We first transform each series in the hierarchy using the Box-Cox transformation as given in equation (10) to stabilise any variations in the data.

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

We use "guerrero" method (Guerrero 1993) implemented in `BoxCox.lambda()` function in `forecast` package R software (Hyndman et al. 2019) to select optimal λ. Then we fit univariate ARIMA models for the transformed series. We use `auto.arima()` function in `forecast` package to choose the best model that minimises the AIC. Using the fitted models we produce $h = 1, ..., 6$ months ahead forecasts 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 (Hyndman & Athanasopoulos 2018).

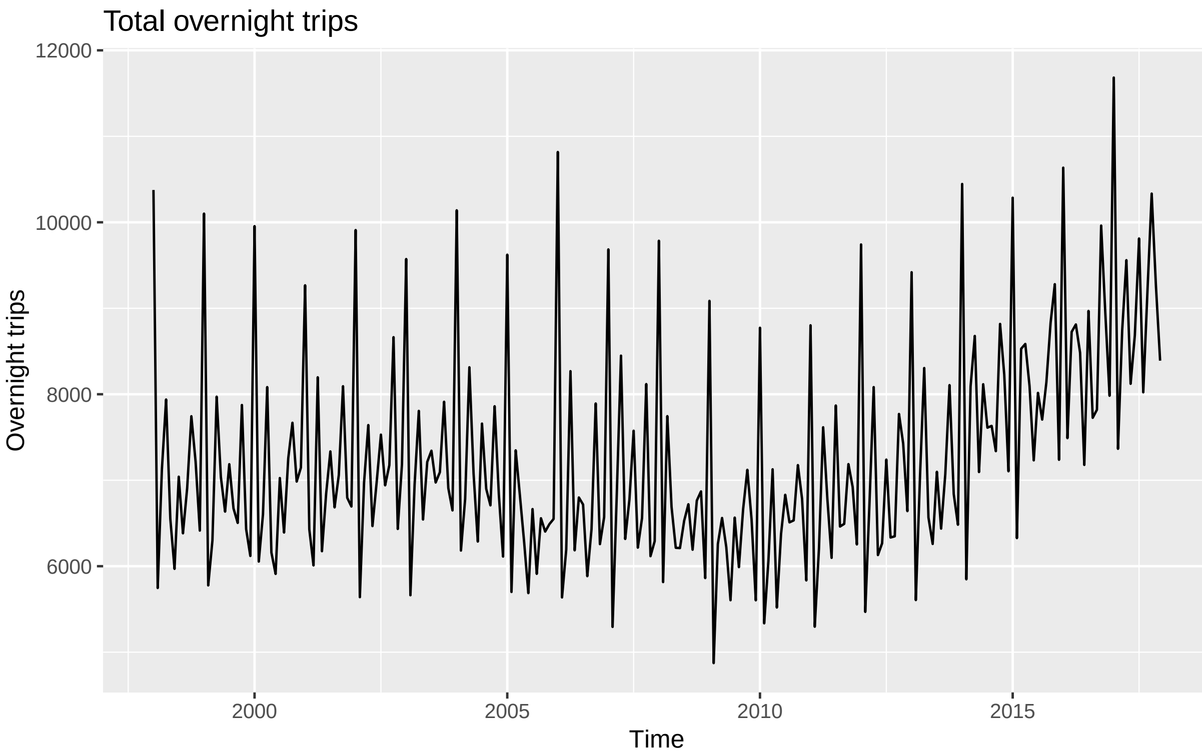

Figure 9: Time plot for total overnight trips.

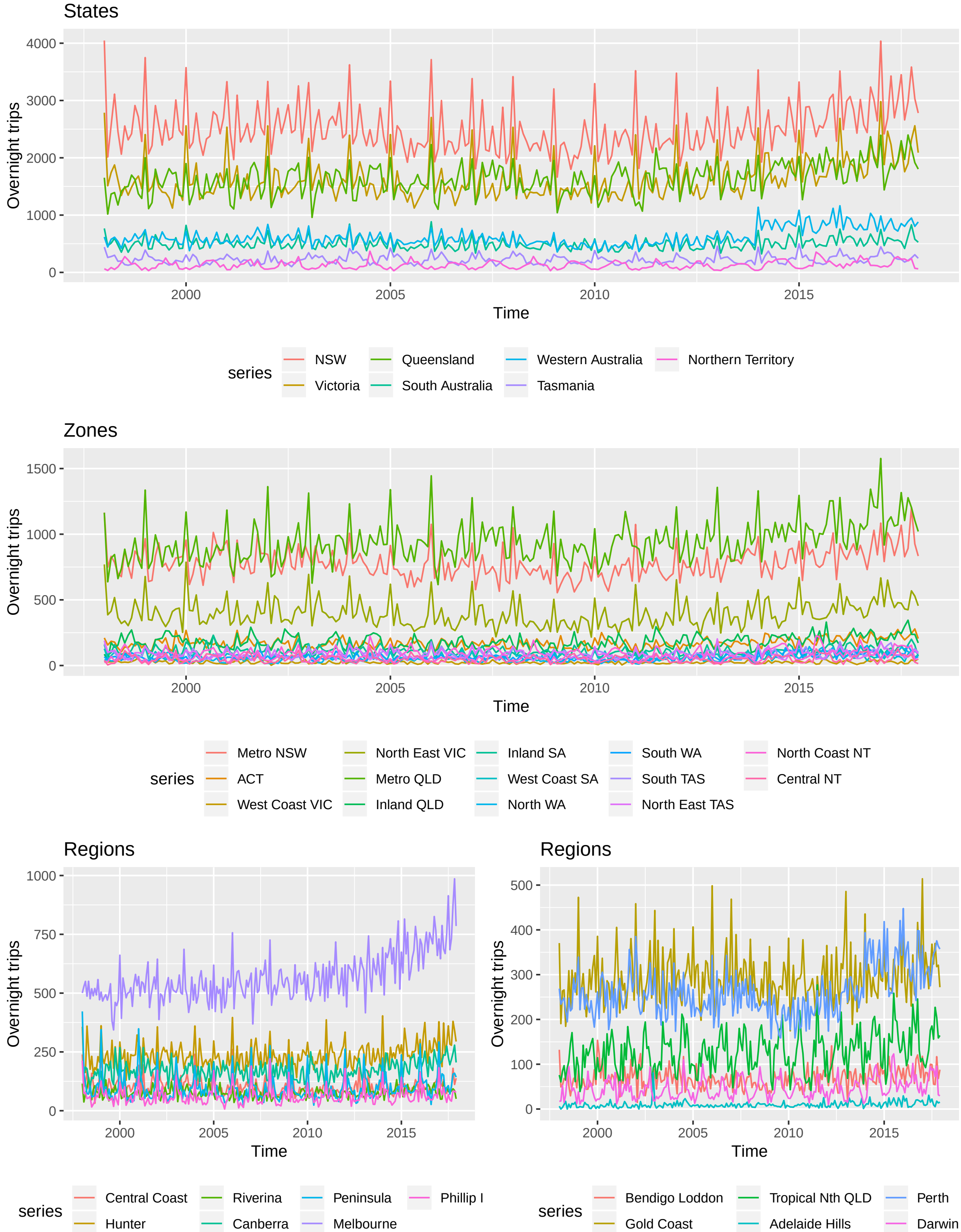

Figure 10: Time plot for some selected series from different disaggregate levels of the hierarchy.

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

Hence the reconciled forecasts follow from these will also biased. This is the exact scenario that we want to demonstrate in this study. As we proposed before, we first bias correct the base forecasts prior to the reconciliation. We use the mathematical formula in equation (12) to calculate the back-transformed and bias-adjusted forecasts.

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

where, x_{t+h} is the h-step-ahead forecasts from the Box-Cox transformed series and σ_h^2 is the variance of x_{t+h}.

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.

Results and discussion:

Mean Squared Error (MSE) is used to measure the forecast accuracy and results are presented in Table 1. Recall that projections preserve the unbiasedness in reconciled forecasts only if the base forecasts are unbiased. As a consequence, the reconciled forecasts follow from the biased base forecasts will also be biased. We can see from the results that unbiased-reconciled forecasts from MinT(Shrink) and WLS are better than that of biased-reconciled forecasts although the unbiased-base forecasts are not necessarily better than the

Table 1: Average MSE($\times 10^3$) of base and reconciled point forecasts at forecasts horizons $h = 1, ..., 6$ are presented. Bias and unbiased columns represent the results for biased-base/biased-reconciled and unbiased-base/unbiased-reconciled forecasts respectively. Comparisons can be made across biased vs unbiased forecasts as well as base vs reconciled forecasts.

	h=1		h=2		h=3		h=4		h=5		h=6	
Method	Biased	Unbiased	Biased	Unbiased	Biased	Unbiased	Biased	Unbiased	Biased	Unbiased	Biased	Unbiased
Base	12.18	161.63	13.14	243.13	14.08	269.20	15.88	325.41	15.49	332.11	15.79	344.40
MinT(Shrink)	10.14	**9.96**	11.47	**10.78**	12.55	**11.94**	15.53	**14.03**	14.63	**13.46**	15.28	**13.78**
WLS	15.62	13.97	15.93	13.71	16.65	14.50	18.91	16.26	18.76	16.20	18.85	15.83
OLS	11.73	135.17	12.80	204.42	13.73	225.59	15.56	272.41	15.17	278.08	15.49	288.21
Bottom-up	17.86	15.55	17.86	14.90	18.47	15.61	20.50	17.21	20.34	17.04	20.39	16.69

biased-base forecasts. Furthermore, unbiased MinT reconciled forecasts are outperforming all biased and unbiased forecasts. These results are true for all forecast horizons.

Following the distance reducing property in Theorem 3.1 and the discussion therein, we have noted that the orthogonal projections will always improve the base forecasts with respect to the mean squared errors whereas, the oblique projections will not always, but improves only on average. Figure 11 shows the MSE difference between base forecasts and reconciled forecasts - MinT, OLS and WLS for all replications in different forecast horizons. Recall that OLS reconciliation is doing orthogonal projection whereas MinT and WLS are doing oblique projections onto the coherent subspace. It is apparent from the plot that the MSE difference for OLS is always positive, implying that OLS improves the base in terms of MSE in all replications. On the other hand, there are some instances that this difference for MinT and WLS is negative which is reflected from the short negative tails of box-plots. This implies that oblique projections will not improve the base forecasts in every case. However, these projections will outperform the base with respect to MSE on average which is again reflected from the highly positive skewed MSE differences.

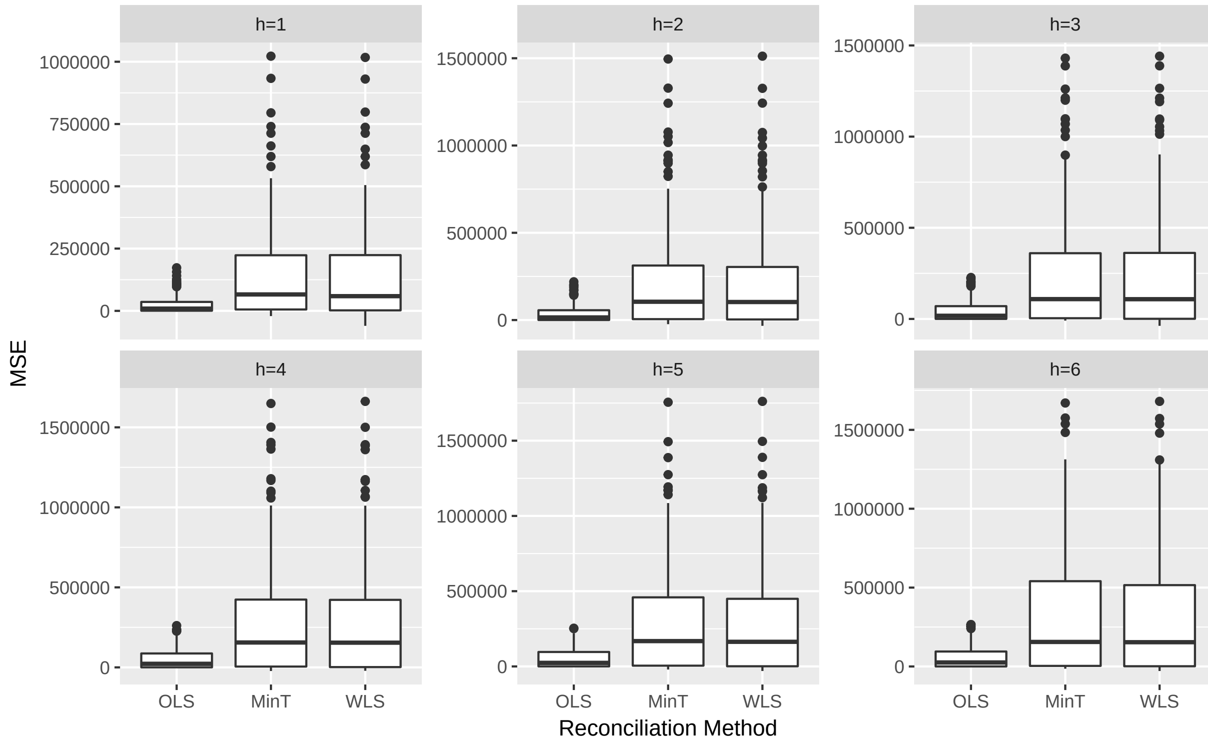

Figure 11: Difference of MSE between Base forecasts and reconciled forecasts over the replications in distinct forecast horizons. Positive values of the difference implies reconciliation improves the forecast accuracy than base forecasts.

6 Conclusions

By redefining coherent point forecast and point forecast reconciliation, we rehash all existing reconciliation methods into a single projection based geometric framework. We have also established new theoretical results that support the use of projections for point forecast reconciliation. We show that projection of unbiased base forecasts onto the coherent subspace will always produce unbiased reconciled forecasts. Yet the projection-based reconciliation can be used to reconciled bias base forecasts after bias adjustments. Empirical results from the application of these methods to forecasting Australian domestic tourism flow show that reconciled forecast follows from bias-adjusted base improves the forecast accuracy and unbiased MinT reconciliation is outperforming.

These new geometric interpretations of hierarchical forecast reconciliation facilitate extensions of the problem to a probabilistic framework. We leave this discussion into a different paper.

7 Appendix

Table 2: Geographical hierarchy of Australian tourism flow

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

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.pdf', *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.

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.