MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Forecasting hierarchical and grouped time series through trace minimization

Shanika L Wickramasuriya, George Athanasopoulos and Rob J Hyndman

November 2015

Working Paper 15/15

Forecasting hierarchical and grouped time series through trace minimization

Shanika L Wickramasuriya
Department of Econometrics and Business Statistics,
Monash University,
VIC 3800, Australia.
Email: shanika.wickramasuriya@monash.edu

George Athanasopoulos
Department of Econometrics and Business Statistics,
Monash University,
VIC 3800, Australia.
Email: george.athanasopoulos@monash.edu

Rob J Hyndman
Department of Econometrics and Business Statistics,
Monash University,
VIC 3800, Australia.
Email: rob.hyndman@monash.edu

26 November 2015

JEL classification: C32, C53

Forecasting hierarchical and grouped time series through trace minimization

Abstract

Large collections of time series often have aggregation constraints due to product or geographical hierarchies. The forecasts for the disaggregated series are usually required to add up exactly to the forecasts of the aggregated series, a constraint known as "aggregate consistency". The combination forecasts proposed by Hyndman et al. (2011) are based on a Generalized Least Squares (GLS) estimator and require an estimate of the covariance matrix of the reconciliation errors (i.e., the errors that arise due to aggregate inconsistency). We show that this is impossible to estimate in practice due to identifiability conditions.

We propose a new combination forecasting approach that incorporates the information from a full covariance matrix of forecast errors in obtaining a set of aggregate consistent forecasts. Our approach minimizes the mean squared error of the aggregate consistent forecasts across the entire collection of time series under the assumption of unbiasedness. The minimization problem has a closed form solution. We make this solution scalable by providing a computationally less demanding alternative representation.

We evaluate the performance of the proposed method compared to alternative methods using a series of simulation designs which take into account various features of the collected time series. This is followed by an empirical application using Australian domestic tourism data. The results indicate that the proposed method works well with artificial and real data.

Keywords: Hierarchical time series, forecasting, reconciliation, contemporaneous error correlation, trace minimization

1 Introduction

Many applications require forecasts of large collections of related time series which can be organized in a hierarchical structure. For example, sales of a multinational company can be disaggregated in a geographical hierarchy into countries, regions, cities and stores. The company will usually require forecasts of total sales, national sales, regional sales, down to sales for an individual store, and these forecasts should add up appropriately across the hierarchy.

The company may also produce many products that form a product hierarchy, divided into groups and sub-groups of products. Then forecasts of total sales, and sales within each product grouping are also required. The cross-product of these two hierarchies often results in a very large collection (often consisting of millions) of individual time series of sales for each product type in each store.

A large collection of time series with aggregation constraints is called a "grouped time series" (Hyndman, Lee, and Wang, 2016). When the grouping structure is a single hierarchy, we call the collection a "hierarchical time series". When forecasting these structures, it is essential to ensure that the forecasts reconcile across the aggregation constraints (i.e., the forecasts of aggregates should be equal to the sum of the corresponding disaggregated forecasts). Specifically, the forecasts should mimic the properties of the real data.

A simple approach is to forecast all of the most disaggregated series, and simply add the forecasts to form forecasts of the various aggregated series. This is known as "bottom-up" forecasting. However, this ignores relationships between series, and performs particularly poorly on highly disaggregated data which tend to have a low signal-to-noise ratio.

If we ignore the aggregation constraints, we could simply forecast all the series in a collection independently. However, it is very unlikely (unless extremely simple forecasting methods are used) that the resulting set of forecasts will be aggregate consistent. Further, this approach also ignores the relationships between series.

Hyndman et al. (2011) proposed a solution to this problem using least squares reconciliation. Their method involves forecasting all series at all levels of aggregation independently (we refer to these as *base* forecasts), and then uses a regression model to optimally combine and reconcile these forecasts (we refer to these as *reconciled* forecasts). In the regression, the independent base forecasts are modeled as the sum of the expected values of the future series and an error term. If the covariance matrix of the error is known, then the Generalized Least Squares (GLS) gives the minimum variance unbiased estimate of the expected values of the disaggregated time series. However, we shall show that the covariance matrix is non-identifiable and therefore impossible to estimate.

Hyndman et al. (2011) and Athanasopoulos, Ahmed, and Hyndman (2009) reverted to using Ordinary Least Squares (OLS) to compute reconciled forecasts and showed that their method works well compared to the most commonly implemented approaches. Hyndman, Lee, and Wang (2016) suggested using Weighted Least Squares (WLS), taking account of the variances on the diagonal of the variance-covariance matrix but ignoring the off-diagonal covariance elements. Furthermore, estimates of the variances are not readily available in practice, and so Hyndman, Lee, and Wang (2016) proposed that the variances of the base forecast errors be used instead. However they provided no theoretical justification for this proxy; we provide the missing justification in this paper. Moreover, they introduced several algorithms to speed up the computations involved so that these methods (OLS and WLS) could handle very large collections of time series efficiently.

Another theoretical contribution to the field of hierarchical forecasting was proposed by van Erven and Cugliari (2014) using a Game-Theoretically OPtimal (GTOP) reconciliation method. They select the set of reconciled predictions such that the total weighted quadratic loss of the reconciled predictions will never be greater than the total weighted quadratic loss of the base predictions. While this approach has some advantages in that it requires fewer assumptions about the forecasts and forecast errors, it does not have a closed form solution, and does not scale well for very large collections of time series.

More recently, Park and Nassar (2014) introduced a disaggregation ("top-down") approach to forecasting hierarchical time series using a Bayesian framework. They proposed a probabilistic model with dynamically evolving latent variables to detect changes of proportions in time series at each level of disaggregation. However, Hyndman et al. (2011) showed theoretically that any top-down method introduces bias into the reconciled forecasts at each disaggregation level even if the base forecasts are unbiased. Additionally, the approach of Park and Nassar (2014) cannot be easily generalized to handle grouped but non-hierarchical collections of time series.

Our approach is to extend the work of Hyndman, Lee, and Wang (2016) and Hyndman et al. (2011) and frame the problem in terms of finding a set of minimum variance unbiased estimates of future values of all time series across the entire collection. That is, we minimize the sum of variances of the reconciled forecast errors under the property of unbiasedness.

An interesting feature of the proposed method is that it results in a unique analytical solution which incorporates information about the correlation structure of the collection. Furthermore, the proposed solution has an equivalent representation which allows for greater computational efficiency in obtaining a set of reconciled forecasts for very large collections of time series.

2 Forecast reconciliation for hierarchical and grouped time series

2.1 Notation

Following the notation in Hyndman, Lee, and Wang (2016), we let $\boldsymbol{y}_t$ be an m-vector containing all observations at time t, and $\boldsymbol{b}_t$ be an n-vector containing the observations at the most disaggregated level only. This leads to the convenient general matrix representation

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

where $\boldsymbol{S}$ is a "summing matrix" of order $m \times n$ which aggregates the bottom level series to the series at aggregation levels above.

To take a specific example, consider the small hierarchical time series depicted in the tree diagram in Figure 1. Each parent comprises the sum of its children. Let y_t denote the observation at time t at the most aggregate level 0; $y_{A,t}$ and $y_{B,t}$ the observations at aggregation level 1; and $y_{AA,t}, y_{AB,t}, \dots, y_{BB,t}$ the observations at the lowest disaggregate level. In this example, $n = 5$, $m = 8$, $\boldsymbol{y}_t = [y_t, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{AC,t}, y_{BA,t}, y_{BB,t}]'$, $\boldsymbol{b}_t = [y_{AA,t}, y_{AB,t}, y_{AC,t}, y_{BA,t}, y_{BB,t}]'$ and the summing matrix is given by

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ & & \boldsymbol{I}_n & & \end{bmatrix},$$

where $\boldsymbol{I}_n$ is an identity matrix of dimension $n = 5$.

The same general notation can apply to any large collection of time series subject to aggregation constraints. Each aggregation constraint is represented by a row in the summing matrix $\boldsymbol{S}$.

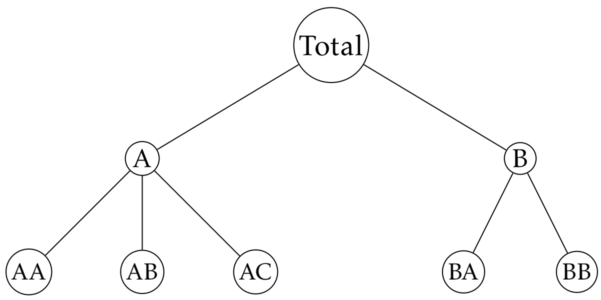

Figure 1: *A two level hierarchical tree diagram.*

Let $\hat{\boldsymbol{y}}_T(h)$ be a vector of h-step-ahead base forecasts for each time series in the collection, made at time T, and stacked in the same order as $\boldsymbol{y}_t$. Then all reconciliation methods can be written as

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_T(h), \tag{1}$$

for some appropriately selected "reconciliation matrix" $\boldsymbol{P}$ of order $n \times m$, where $\tilde{\boldsymbol{y}}_T(h)$ is a set of reconciled forecasts which are now by construction aggregate consistent. Hence the basic idea of forecast reconciliation methods is to linearly map a given set of base forecasts to a set of reconciled forecasts. The role of $\boldsymbol{P}$ is to extract or combine the base forecasts into bottom level disaggregated forecasts which are then summed by $\boldsymbol{S}$. For unbiased forecasts, we require $\boldsymbol{SPS} = \boldsymbol{S}$ (Hyndman et al., 2011). See Athanasopoulos, Ahmed, and Hyndman (2009) for further discussion of $\boldsymbol{P}$ matrices for various hierarchical forecasting methods.

2.2 GLS reconciliation

Hyndman et al. (2011) proposed the "optimal combination" approach, based on the regression model

$$\hat{\boldsymbol{y}}_T(h) = \boldsymbol{S}\boldsymbol{\beta}_T(h) + \boldsymbol{\varepsilon}_h, \tag{2}$$

where $\boldsymbol{\beta}_T(h) = \mathrm{E}[\boldsymbol{b}_{T+h}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T]$ is the unknown mean of the most disaggregate series at the bottom level and $\boldsymbol{\varepsilon}_h$ is the reconciliation error with mean zero and covariance matrix $\mathrm{Var}(\boldsymbol{\varepsilon}_h) = \Sigma_h$. If Σ_h was known, the GLS estimator of $\boldsymbol{\beta}_T(h)$ would be the minimum variance unbiased estimator, which would lead to reconciled forecasts given by

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_T(h) \tag{3}$$

where $\Sigma_h^{\dagger}$ is a generalized inverse of Σ_h which is often (near) singular due to the aggregation involved in $\boldsymbol{y}_t$. Hence for the optimal combination approach $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$.

In general Σ_h is not known and is not identifiable. To see this, note that the residuals from the regression model (2) are given by

$$\tilde{\boldsymbol{\varepsilon}}_h = \hat{\boldsymbol{y}}_T(h) - \tilde{\boldsymbol{y}}_T(h) = (\boldsymbol{I}_m - \boldsymbol{SP})\hat{\boldsymbol{y}}_T(h), \tag{4}$$

so that

$$\mathrm{Var}(\tilde{\varepsilon}_h) = (\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P})\Sigma_h(\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P})'.$$

In order for Σ_h to be identified, $(\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P})$ must be invertible and therefore of full rank. However, $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$ implies that $\boldsymbol{S}'\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}'\boldsymbol{S}$, and because $\boldsymbol{S}'\boldsymbol{S}$ is positive definite, we have $\boldsymbol{P}\boldsymbol{S} = \boldsymbol{I}_n$, and $\boldsymbol{S}\boldsymbol{P} = \boldsymbol{I}_m$. Therefore,

$$\mathrm{Rank}(\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P}) = \mathrm{tr}(\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P}) = \mathrm{tr}(\boldsymbol{I}_m) - \mathrm{tr}(\boldsymbol{I}_n) = m - n,$$

and $(\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P})$ is rank deficient, and consequently Σ_h cannot be identified.

Hyndman et al. (2011) avoided estimating Σ_h by using OLS, replacing Σ_h by $k_h\boldsymbol{I}_m$ where k_h is a constant.

2.3 Minimum trace (MinT) reconciliation

Let the h-step-ahead base and reconciled forecast errors be defined as,

$$\hat{\boldsymbol{e}}_t(h) = \boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_t(h) \tag{5}$$

$$\tilde{\boldsymbol{e}}_t(h) = \boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_t(h) \tag{6}$$

for $t = 1, 2, \ldots,$ where $\hat{\boldsymbol{y}}_t(h)$ and $\tilde{\boldsymbol{y}}_t(h)$ are the h-step-ahead base and reconciled forecasts using information up to and including time t, and $\boldsymbol{y}_{t+h}$ are the observed values of all series at time $t+h$. It should be noted that the reconciled forecast errors, $\tilde{\boldsymbol{e}}_t$ given by (6), and the reconciliation errors, $\tilde{\varepsilon}_t$ given by (4), are conceptually different. The former is the forecast error of reconciled forecasts while the latter is the error due to the aggregation inconsistency of the base forecasts. Hyndman et al. (2011) confused these two errors.

Lemma 1. *For any $\boldsymbol{P}$ such that $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$, the covariance matrix of the h-step-ahead reconciled forecast errors is given by,*

$$\mathrm{Var}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_t(h)] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}'$$

where $\tilde{\boldsymbol{y}}_t(h)$ is given by (1) and $\boldsymbol{W}_h = \mathrm{E}[\hat{\boldsymbol{e}}_t(h)\hat{\boldsymbol{e}}_t'(h)]$ is the variance-covariance matrix of the h-step-ahead base forecast errors.

Proof. See Appendix A.1. □

Lemma 1 can be used to construct prediction intervals for hierarchical and grouped forecasts.

We wish to find the value of $\boldsymbol{P}$ that minimizes the trace of $\text{Var}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_t(h)]$ subject to it satisfying $\boldsymbol{SPS} = \boldsymbol{S}$. This would give the best (minimum variance) linear unbiased reconciled forecasts. We refer to this as the MinT approach.

Theorem 1. *Let $\boldsymbol{W}_h$ be the positive definite covariance matrix of the h-step-ahead base forecast errors. Then the optimal reconciliation matrix, which minimizes* $\text{tr}[\boldsymbol{SPW}_h\boldsymbol{P}'\boldsymbol{S}']$ *such that $\boldsymbol{SPS} = \boldsymbol{S}$, is given by*

$$\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}. \tag{7}$$

or equivalently by

$$\boldsymbol{P} = \boldsymbol{J} - \boldsymbol{J}\boldsymbol{W}_h\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{W}_h\boldsymbol{U})^{-1}\boldsymbol{U}', \tag{8}$$

where the $m \times n$ summing matrix is partitioned as $\boldsymbol{S}' = \left[\boldsymbol{C}' \;\vdots\; \boldsymbol{I}_n\right]$, $\boldsymbol{J} = \left[\boldsymbol{0}_{n\times m^*} \;\vdots\; \boldsymbol{I}_n\right]$, $\boldsymbol{U}' = \left[\boldsymbol{I}_{m^*} \;\vdots\; -\boldsymbol{C}\right]$, *and $m^* = m - n$.*

Proof. See Appendix A.2. □

Computing $\boldsymbol{P}$ using (8) requires the inversion of an $m^* \times m^*$ matrix (where $m^* < n$), compared to the inversion of two matrices of orders $n \times n$ and $m \times m$ (where $n < m$) required for the formulation in Theorem 1. Hence, the alternative representation is computationally significantly less demanding.

The reconciled forecasts from the MinT approach are computed by,

$$\tilde{\boldsymbol{y}}_T(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\hat{\boldsymbol{y}}_T(h) = \boldsymbol{S}[\boldsymbol{J} - \boldsymbol{J}\boldsymbol{W}_h\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{W}_h\boldsymbol{U})^{-1}\boldsymbol{U}']\hat{\boldsymbol{y}}_T(h). \tag{9}$$

Dagum and Cholette (2006) also describe this alternative solution, although they derive it under the assumption that $\text{Cov}(\hat{\boldsymbol{e}}_{1,h}, \hat{\boldsymbol{e}}_{2,h}) = \boldsymbol{0}$ where $\hat{\boldsymbol{e}}_{2,h}$ is the vector of base forecast errors from the most disaggregated series, and $\hat{\boldsymbol{e}}_{1,h}$ is the vector of base forecast errors from all other series. Moreover, this result coincides with the work of Stone (1976) and Byron (1978) in the area of balancing national income accounts.

Because Hyndman et al. (2011) confused the forecast errors with the reconciliation errors, they expressed the GLS solution (3) using the matrix $\boldsymbol{W}_h$ rather than Σ_h, thus giving equation (7). They justified the use of OLS by assuming that the base forecast errors of the bottom level series are additive like the data (the "error-additivity assumption"), and showed that using the Moore-Penrose inverse would then convert the GLS solution to an OLS solution.

Theorem 1 provides further insight into this result. Let $\hat{\boldsymbol{a}}_t(h)$ be the bottom-level base forecast errors and let $\boldsymbol{V}_h = \text{Var}(\hat{\boldsymbol{a}}_t(h))$. Then the error-additivity assumption is that $\hat{\boldsymbol{e}}_t(h) = \boldsymbol{S}\hat{\boldsymbol{a}}_t(h)$, and so $\boldsymbol{W}_h = \boldsymbol{S}\boldsymbol{V}_h\boldsymbol{S}'$ is singular. Then the variance covariance matrix of the reconciled forecast errors is $\text{Var}[\boldsymbol{y}_{n+h} - \tilde{\boldsymbol{y}}_n(h)] = \boldsymbol{SPSV}_h(\boldsymbol{SPS})' = \boldsymbol{SV}_h\boldsymbol{S}'$ which is independent of $\boldsymbol{P}$. Consequently, under the error-additivity assumption, *any* matrix that satisfies $\boldsymbol{SPS} = \boldsymbol{S}$ will be a minimium variance MinT solution. Applying the Moore-Penrose inverse of $\boldsymbol{W}_h$ when computing (7) will lead to the OLS solution $\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$ (using Fact 6.4.8 of Bernstein, 2005). On the other hand, applying the Moore-Penrose inverse when computing (9) yields

$$\boldsymbol{P} = \boldsymbol{J} - \boldsymbol{JSV}_h\boldsymbol{S}'\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{SV}_h\boldsymbol{S}'\boldsymbol{U})^{-1}\boldsymbol{U}' = \boldsymbol{J},$$

as $\boldsymbol{S}'\boldsymbol{U} = \boldsymbol{0}$. This is the bottom-up approach.

2.4 Alternative estimators for $\boldsymbol{W}_h$

Note that the MinT and GLS solutions differ only in the covariance matrix that enters the estimators. Σ_h in the GLS estimator (3) is the covariance matrix of the reconciliation errors, which we have shown cannot be identified, whereas $\boldsymbol{W}_h$ in the MinT estimator (9) is the covariance matrix of the base forecast errors. Although $\boldsymbol{W}_h$ does not suffer from a lack of identification, it is nevertheless challenging to estimate, especially for $h > 1$. In this section we provide some alternatives which we evaluate later in some Monte Carlo experiments and an empirical application.

1. Set $\boldsymbol{W}_h = k_h\boldsymbol{I}$ $\forall h$ where $k_h > 0$. This is the most simplifying assumption to make, and the MinT estimator collapses to the OLS estimator of Hyndman et al. (2011).

2. Set $\boldsymbol{W}_h = k_h\text{diag}(\hat{\boldsymbol{W}}_1)$ $\forall h$ where $k_h > 0$ and

$$\hat{\boldsymbol{W}}_1 = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{e}}_t(1)\hat{\boldsymbol{e}}_t(1)'$$

   is the unbiased sample estimator of the in-sample one-step-ahead base forecast errors as defined in (5). In this case MinT can be described as a WLS estimator. A similar estimator is used in Hyndman, Lee, and Wang (2016), however no theoretical justification is provided there. Athanasopoulos et al. (2015) also implement this estimator when applying temporal hierarchies. In what follows we denote this as WLS_v (i.e., WLS applying variance scaling).

3. Set $\boldsymbol{W}_h = k_h\boldsymbol{\Lambda}$, $\forall h$ where $k_h > 0$ and $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{S1})$ where $\boldsymbol{1}$ is a unit column vector of dimension n. This specification assumes that the bottom level base forecast errors each have variance k_h and are uncorrelated between nodes. Hence each element of the diagonal $\boldsymbol{\Lambda}$ matrix contains the number of forecast error variances contributing to that aggregation level. This estimator only depends on the grouping structure of the collection and therefore we refer to it as an estimator that applies structural scaling and we denote it as WLS_s. This estimator was proposed by Athanasopoulos et al. (2015) for temporal hierarchies. Its advantage over OLS is that it assumes equivariant forecast errors only within levels of the hierarchy and not across levels which is unrealistically assumed by OLS. It is particularly useful in cases where forecast errors are not available. For example, in cases where the base forecasts are generated by judgemental forecasting.

4. Set $\boldsymbol{W}_h = k_h\hat{\boldsymbol{W}}_1$ $\forall h$ where $k_h > 0$, the unrestricted sample covariance estimator for $h = 1$. Even though this is relatively simple to obtain, it may not be a good estimate when $m > T$ or m is of the same order as T. In the results that follow we denote this as MinT(Sample).

5. Set $\boldsymbol{W}_h = k_h\hat{\boldsymbol{W}}^*_{1,D}$ $\forall h$ where $k_h > 0$, $\hat{\boldsymbol{W}}^*_{1,D} = \lambda_D\hat{\boldsymbol{W}}_{1,D} + (1-\lambda_D)\hat{\boldsymbol{W}}_1$ is a shrinkage estimator with diagonal target, $\hat{\boldsymbol{W}}_{1,D}$ is a diagonal matrix consisting of the diagonal entries of $\hat{\boldsymbol{W}}_1$, and λ_D is the shrinkage intensity parameter. Thus, off-diagonal elements of $\hat{\boldsymbol{W}}_1$ are shrunk towards zero and diagonal elements (variances) remain unchanged. Schäfer and Strimmer (2005) proposed the shrinkage parameter

$$\hat{\lambda}_D = \frac{\sum_{i\neq j}\hat{\text{Var}}(\hat{r}_{ij})}{\sum_{i\neq j}\hat{r}_{ij}^2},$$

   where $\hat{r}_{ij}$ is the ij-th element of $\hat{\boldsymbol{R}}_1$, the 1-step-ahead sample correlation matrix. We refer to this as MinT(Shrink) in the results that follow.

3 Monte Carlo experiments

In order to evaluate the performance of the MinT approach we carried-out simulations for three different designs. Sections 3.1 and 3.2 explore the impact of the correlation structure of the series on the reconciled forecasts. Section 3.3 aims to capture an important observed feature of grouped time series, namely that more aggregated series are smoother than less aggregated series.

3.1 Exploring the effect of correlation

We first consider the simplest possible hierarchy for which two bottom level series are generated and aggregated to the total y_t. The assumed data generating processes for the bottom level series is a bivariate VAR(1),

$$\boldsymbol{B}\boldsymbol{z}_t = \boldsymbol{\Phi}\boldsymbol{z}_{t-1} + \boldsymbol{\eta}_t,$$

where

$$\boldsymbol{B} = \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Phi} = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}, \quad \boldsymbol{z}_t = \begin{bmatrix} y_{A,t} \\ y_{B,t} \end{bmatrix}, \quad \boldsymbol{\eta}_t = \begin{bmatrix} \eta_{A,t} \\ \eta_{B,t} \end{bmatrix},$$

$\eta_{A,t} \sim \mathcal{N}(0, \sigma_A^2)$ and $\eta_{B,t} \sim \mathcal{N}(0, \sigma_B^2)$. This structure implies that $y_{A,t}$ does not have a contemporaneous effect on $y_{B,t}$. The contemporaneous error correlation matrix is given by

$$\mathrm{Cor}(\boldsymbol{B}^{-1}\boldsymbol{\eta}_t) = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \qquad \text{where} \quad \rho = \frac{-\gamma\sigma_B}{\sqrt{\sigma_A^2 + \sigma_B^2\gamma^2}}.$$

Therefore we control the correlation by setting

$$\gamma = \frac{-\rho}{\sqrt{1-\rho^2}} \frac{\sigma_A}{\sigma_B}.$$

Simulation setup

Using two sets of parameters, $\{\alpha = 0.8, \beta = 0.4\}$ and $\{\alpha = 0.4, \beta = 0.8\}$, we generated 192 observations for each of the two bottom level series which were then aggregated to give the third series. In both cases $\sigma_A^2 = 2$ and $\sigma_B^2 = 4$ were used. The data were divided into training and test sets comprising the first 180 and the last 12 observations respectively. For each series, 1 to 12-step-ahead base forecasts were generated from an ARIMA model fitted to the training set using the default settings in the automated algorithm of Hyndman and Khandakar (2008) and implemented in the *forecast* package for R (see Hyndman, 2015). The base forecasts were then reconciled using the various approaches discussed in Section 2. Also shown are results for bottom-up forecasts (designated as BU).

We have considered alternative parameter values and sample sizes for this simulation setting, but to save space we do not present all results. The omitted results are qualitatively similar and are available upon request.

The top panels in Figures 2 and 3 show the variation in RMSE (Root Mean Squared Error) for the 1-step-ahead base and reconciled forecasts. The left panels show the results for the top level, and the right panels show the results for the bottom level. The top left panels for both sets of parameters show that the stronger the negative correlation between the bottom level series, the lower the RMSE for the one-step-ahead top level forecasts. This demonstrates that the negative correlation between the two bottom level series has a smoothing effect making the aggregate series easier to forecast. As the correlation increases from −0.8 to 0.8, the RMSE increases almost monotonically. The top right panels of both figures show U-shaped curves for the RMSE for the 1-step-ahead bottom level forecasts. As the error correlation between the two bottom level series increases in magnitude (either positively or negatively), the forecast accuracy deteriorates. This is most likely due to the inability of the univariate models to fully capture the dynamics across the bottom-level series generated from the multivariate process which increase in strength as the correlation increases in absolute value.

The panels below show the relative ratio in average RMSE between the reconciled forecasts and the base forecasts over the horizons $h = 1$, 1–4 and 1–12. A ratio below 1 shows that the average RMSE of the reconciled forecasts is lower than that of the base forecasts. For example, for $h = 1$ the average RMSE for MinT(Shrink) is approximately 30% lower than the RMSE of the base forecasts indicating a 30% improvement in forecast accuracy. For both sets of parameters and for almost every setting across the entire correlation range, the MinT approach forecasts are amongst the most accurate and they almost always improve on the average RMSE of the base forecasts for both top and bottom level series (with a few rare exceptions).

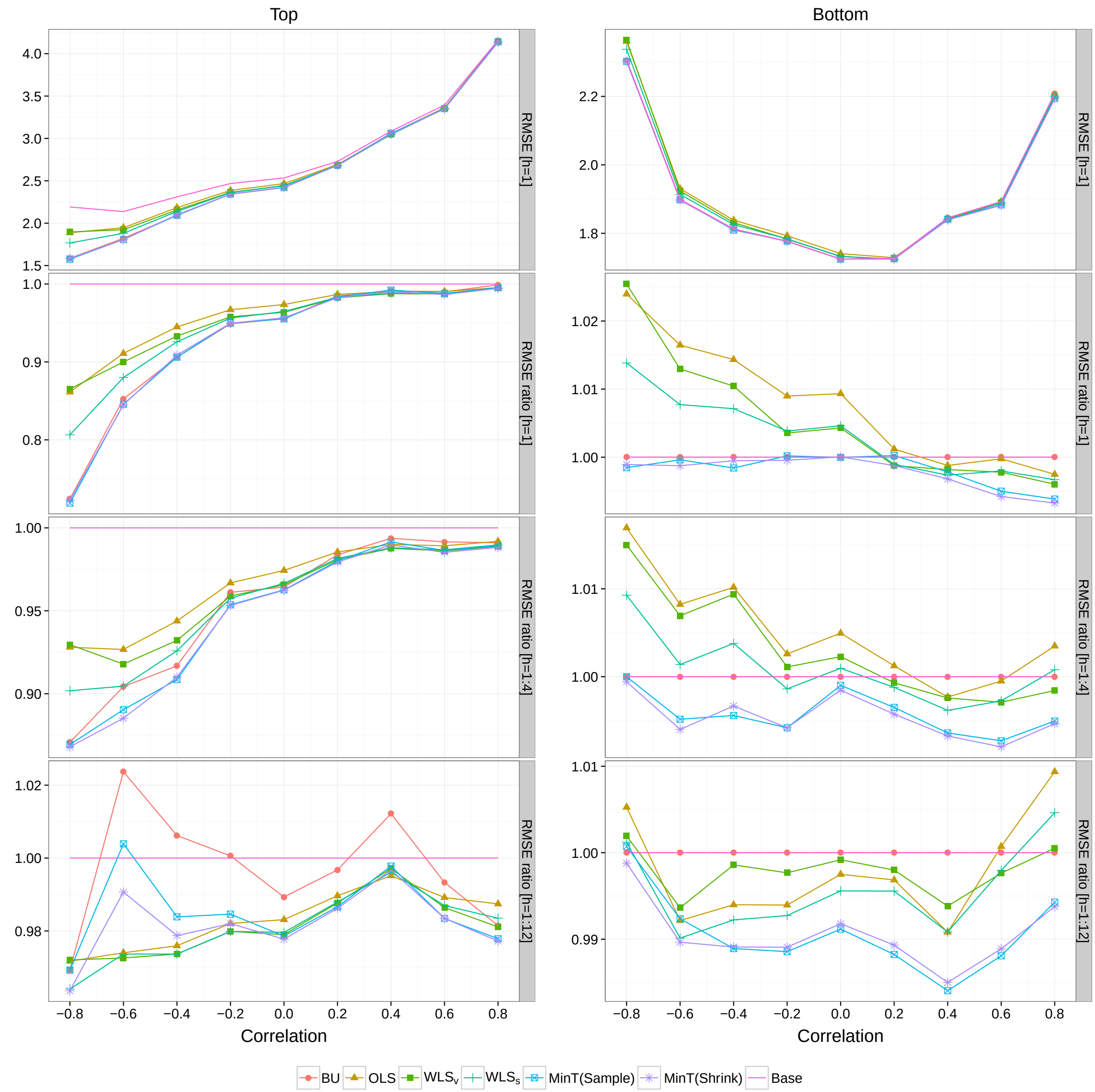

Figure 2: *Forecasting performance of base and reconciled forecasts for the top level series (shown in the left panels) and the bottom level series (shown in the right panels). The top panels show the average RMSE for one-step-ahead forecasts. The panels below show the relative ratio in average RMSE between the reconciled forecasts and the base forecasts for $h = 1$, 1–4 and 1–12 steps ahead respectively. The sample size is $T = 192$ and the parameters used are $\alpha = 0.8$, $\beta = 0.4$, $\sigma_A^2 = 2$ and $\sigma_B^2 = 4$.*

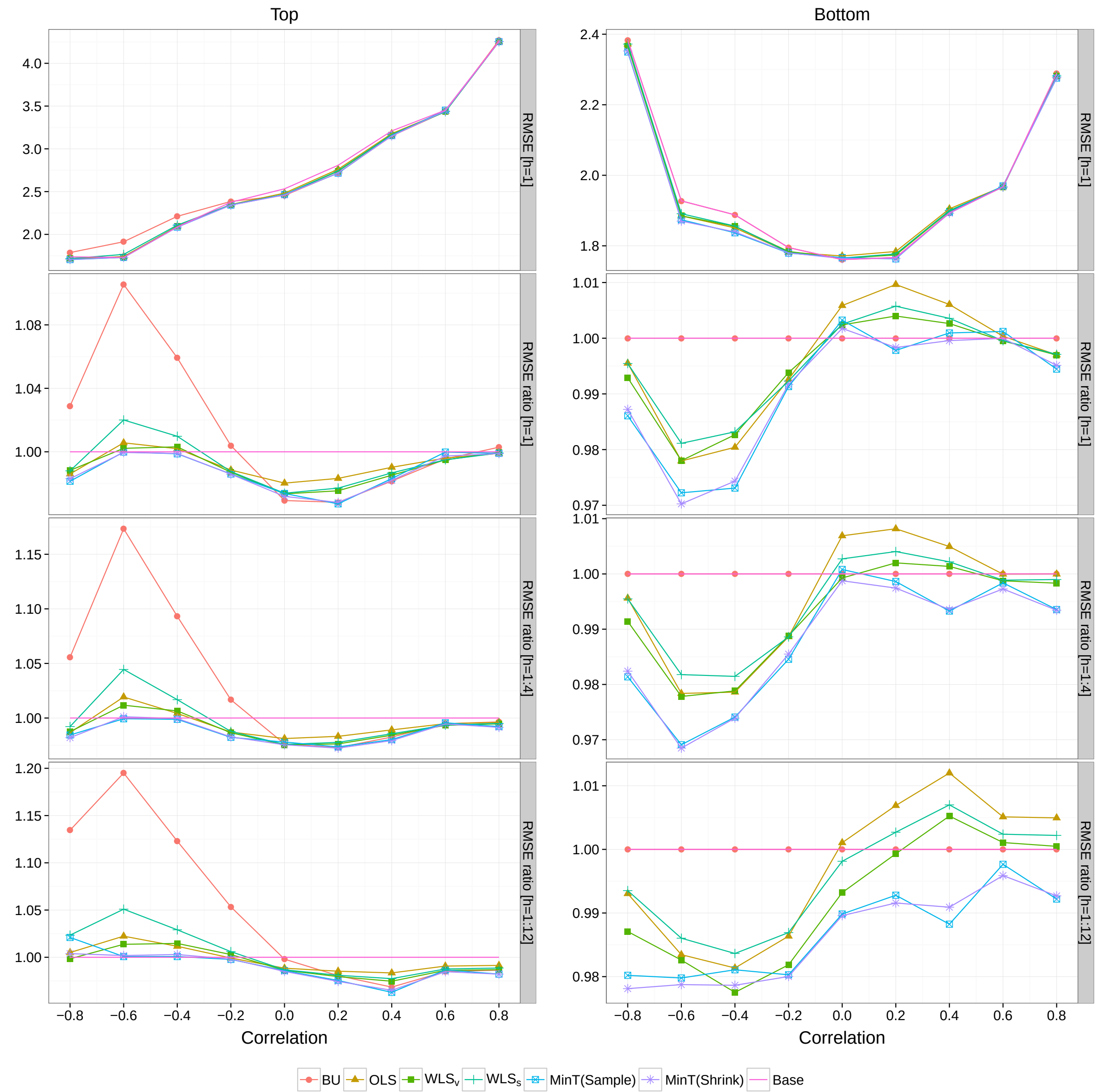

Figure 3: *Forecasting performance of base and reconciled forecasts for the top level series (shown in the left panels) and the bottom level series (shown in the right panels). The top panels show the average RMSE for one-step-ahead forecasts. The panels below show the relative ratio in average RMSE between the reconciled forecasts and the base forecasts for $h = 1$, 1–4 and 1–12 steps ahead respectively. The sample size is $T = 192$ and the parameters used are $\alpha = 0.4$, $\beta = 0.8$, $\sigma_A^2 = 2$ and $\sigma_B^2 = 4$.*

3.2 Exploring the effect of correlation on a larger hierarchy

We now consider a slightly larger hierarchy with 2 levels of aggregation and $m = 7$ series in total. The structure of the hierarchy is given in Figure 4.

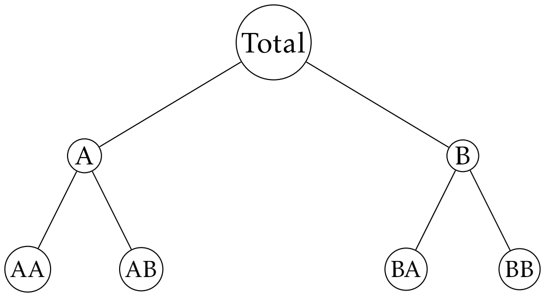

Figure 4: *Hierarchical structure for simulation design 2.*

Once again the bottom level series were first generated and then summed appropriately to obtain the series for the levels above. Each series in the bottom level was generated from an ARIMA(p,d,q) process with p and q taking values $0,1$ and 2 with equal probability and d taking values 0 and 1 with equal probability. For each constructed series, the parameters were chosen randomly from a uniform distribution satisfying stationarity and invertibility conditions. Table 1 shows the parameter space for each component of the ARIMA generating processes. Note that the purpose of using 3.4 in the parameter space of θ_1 is to obtain the same upper and lower bounds as for the parameter space of ϕ_1.

Table 1: *Parameter space of AR and MA components.*

	AR component	
p	Coefficient	Parameter space
1	ϕ_1	$[0.5, 0.7]$
2	ϕ_1	$[\phi_2 - 1, 1 - \phi_2]$
	ϕ_2	$[0.5, 0.7]$
	MA component	
q	Coefficient	Parameter space
1	θ_1	$[0.5, 0.7]$
2	θ_1	$[-(1+\theta_2)/3.4, (1+\theta_2)/3.4]$
	θ_2	$[0.5, 0.7]$

The bottom level ARIMA data generating processes have the following contemporaneous error covariances in order to obtain a hierarchical structure with correlated series,

$$\begin{bmatrix} 5 & 3 & 2 & 1 \\ 3 & 4 & 2 & 1 \\ 2 & 2 & 5 & 3 \\ 1 & 1 & 3 & 4 \end{bmatrix}.$$

This structure allows a strongly positive error correlation among the series with the same parent, while a moderately positive correlation among the series belonging to different parents.

Simulation setup

For each series $T = 60$, 180 and 300 observations were generated and the final $h = 8$, 16 and 24 observations respectively were withheld as test sets. Using the remaining observations as training data, base forecasts were generated from ExponenTial Smoothing (ETS) models using the default settings as implemented in the *forecast* package for R (see Hyndman, 2015) and described in Hyndman and Khandakar (2008). The purpose of fitting ETS models to series generated from ARIMA processes is to emulate what happens in practice where the true generating processes is not known and mis-specified. This process was repeated 1000 times.

The results are presented in the left panel of Table 4 under the heading 'Simulation design 2'. Each entry in the table shows the percentage difference in the average RMSE between the reconciled forecasts and the base forecasts. A negative (positive) entry shows a percentage decrease (increase) in average RMSE relative to the independent unreconciled base forecasts. The bold entries identify the best performing method. The results show that the MinT(Shrink) approach generates the most accurate forecasts overall. MinT(Shrink) is the only reconcililation method that always improves on the base forecasts for all levels, all forecast horizons and all sample sizes. It also always ranks in the top two most accurate methods across forecast horizons or aggregation levels for all sample sizes. It is interesting to note that MinT(Sample) also performs well and approaches MinT(Shrink) in forecast accuracy as the sample size increases.

From the approaches that use a strictly diagonal covariance matrix, WLS_v performs best and would overall be ranked as the second best approach especially for $h > 1$. Recall that for the WLS and the MinT approaches, one-step-ahead in-sample base forecast errors are used and we assume that $\boldsymbol{W}_h \propto \hat{\boldsymbol{W}}_1$ for all $h > 1$. Controlling for the possibly adverse effect of this assumption and therefore evaluating only the one-step-ahead forecast performance between WLS_v and

MinT(Shrink), it seems that specifying a non-diagonal covariance matrix and using shrinkage to estimate it, leads to a substantial improvement in forecast accuracy.

The bottom-up (BU) approach performs especially well for small sample sizes. This is not surprising as covariance estimation is challenging and unreliable in this situation. Obviously the BU approach cannot improve the bottom level forecasts, so that all percentage differences are zero at the bottom level.

3.3 Exploring the smoothing effect of aggregation

A common characteristic of aggregated time series is that aggregation will smooth-out random fluctuations and thereby more aggregated series are less noisy compared to their disaggregate components. In order to capture this feature within a simulation setting, we implemented a simulation design based on van Erven and Cugliari (2014) for the hierarchical structure of Figure 1. The bottom level series were generated from ARIMA(p,d,q) processes as specified in Section 3.2. A noise component was then added to make the bottom series noisier than the aggregated series. Specifically, the bottom level series were generated using

$$\begin{aligned}
y_{AA,t} &= z_{AA,t} - \nu_t - 0.5\omega_t,\\
y_{AB,t} &= z_{AB,t} + \nu_t - 0.5\omega_t,\\
y_{BA,t} &= z_{BA,t} - \nu_t + 0.5\omega_t,\\
\text{and} \qquad y_{BB,t} &= z_{BB,t} + \nu_t + 0.5\omega_t,
\end{aligned}$$

where $z_{AA,t}$, $z_{AB,t}$, $z_{BA,t}$ and $z_{BB,t}$ are ARIMA(p,d,q) processes with error terms $\tau_{AA,t}$, $\tau_{AB,t}$, $\tau_{BA,t}$, $\tau_{BB,t} \sim \mathcal{N}(0,\sigma_0^2)$, while $\nu_t \sim \mathcal{N}(0,\sigma_1^2)$ and $\omega_t \sim \mathcal{N}(0,\sigma_2^2)$ are white noise processes independent of each other. The data for the aggregated levels were obtained by summing the bottom level series. Hence,

$$\begin{aligned}
y_{A,t} &= z_{AA,t} + z_{AB,t} - \omega_t,\\
y_{B,t} &= z_{BA,t} + z_{BB,t} + \omega_t,\\
\text{and} \qquad y_t &= z_{AA,t} + z_{AB,t} + z_{BA,t} + z_{BB,t}.
\end{aligned}$$

This allows the bottom-level series to be correlated with other series in the hierarchy. We set

$$\mathrm{Var}(\tau_{AA,t} + \tau_{AB,t} + \tau_{BA,t} + \tau_{BB,t}) \le \mathrm{Var}(\tau_{AA,t} + \tau_{AB,t} - \omega_t) \le \mathrm{Var}(\tau_{AA,t} - \nu_t - 0.5\omega_t),$$

which simplifies to

$$2\sigma_0^2 \leq \sigma_2^2 \leq \frac{4}{3}(\sigma_1^2 - \sigma_0^2),$$

ensuring that the aggregate series are less noisy than their disaggregate components. We set $\sigma_0^2 = 1, \sigma_1^2 = 10$ and $\sigma_2^2 = 6$. Using these data generating processes, we implement the same simulation setting as in Section 3.2 in order to evaluate the forecasting performance of the alternative forecast reconciliation approaches.

The results are presented in the right panel of Table 4 under the heading 'Simulation design 3'. As expected, there is a tremendous loss in accuracy from forecasting such a structure using the bottom-up approach. As in the previous simulation setting, the best performing methods seem to be the MinT(Shrink) and the WLS$_v$ approaches. Also we again observe that as the sample size increases, MinT(Sample) approaches MinT(Shrink) in forecast accuracy, and occasionally does better.

4 Forecasting Australian Domestic Tourism

Domestic tourism flows within any country form a natural geographical hierarchy. Australia comprises seven states and territories which can be further divided into 27 zones and 76 regions (see Appendix B for further details). Furthermore, stakeholders in tourism are interested in forecasts of tourist flows by purpose of travel as this affects tourist behaviour in many ways. We consider four purposes of travel: holiday, visiting friends and relatives, business and other. In total we have 555 time series resulting from grouping by geography and purpose of travel. Table 2 presents further details.

Table 2: *Grouped time series for Australian tourism flows.*

Geographical grouping	Number of series per geographical grouping	Number of series per purpose of travel	Total
Australia	1	4	5
States	7	28	35
Zones	27	108	135
Regions	76	304	380
Total	111	444	555

For more details on the geographical hierarchy refer to Appendix B.

We use 'visitor nights', the total number of nights spent by Australians away from home, as a measure of tourism flows. The data come from the National Visitor Survey which is managed by Tourism Research Australia and are collected throughout the year using computer assisted telephone interviews from nearly 120,000 Australian residents aged 15 years and over (Tourism Research Australia, 2015). They are monthly time series and span the period January 1998 to December 2013. A time series plot of the most aggregate level is presented in Figure 5. As expected tourism flows are highly seasonal. The plot also shows that overall domestic tourism flows in Australia have been stable over time, with a slight cyclical variation around a relatively flat trend.

The first 96 observations were used as the training set and the remaining observations were used as test data. Using the training data, 1 to 12-step-ahead base forecasts were generated across the entire hierarchy from both ARIMA and ETS models using the default settings in the automated algorithms of Hyndman and Khandakar (2008) and implemented in the *forecast* for R (see Hyndman, 2015). The base forecasts were then reconciled using the alternative approaches we have discussed so far. We should note that the length of each series is less than the total number of series in the hierarchy. Hence, the sample variance covariance is not a positive definite matrix and therefore we did not consider the MinT(Sample) approach. The out-of-sample forecast evaluation was implemented in an expanding window fashion. That is, while increasing the training sample by one observation each time, the above procedure was repeated until November 2013. This gives a total of 96 1-step-ahead forecasts, 95 2-step-ahead forecasts, and up to 85 12-step-ahead forecasts per series.

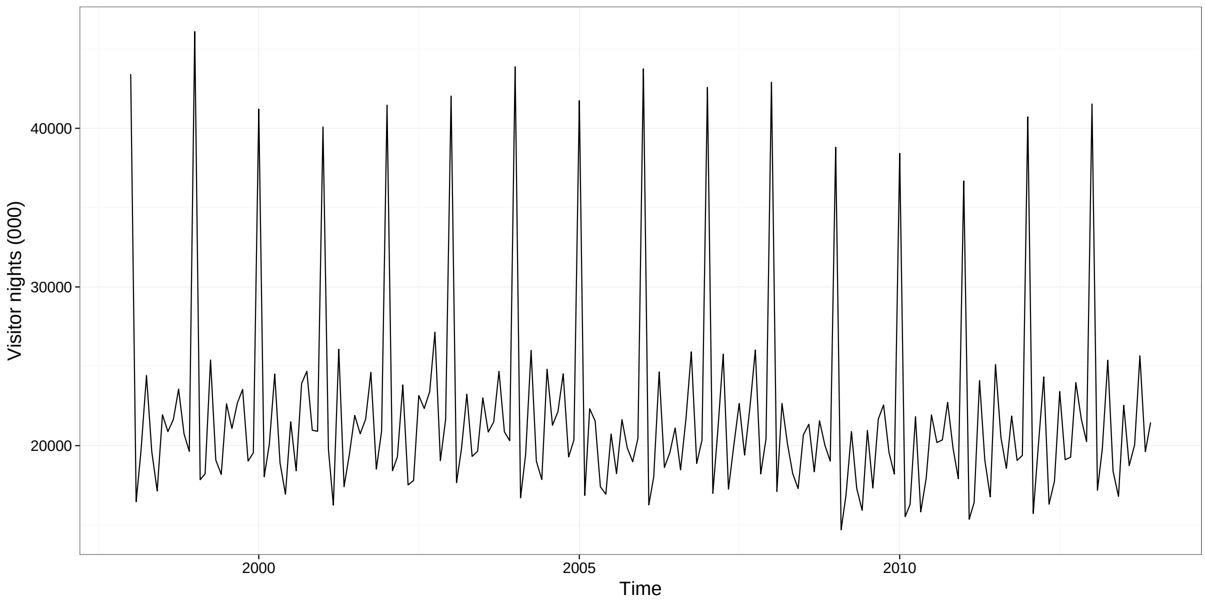

Figure 5: *Visitor nights for Australia.*

The results are presented in Table 5. Each positive (negative) entry shows a percentage increase (decrease) in average RMSE relative to the base forecasts. The top panel shows the results using the ARIMA base forecasts while the bottom panel shows the results using the ETS base forecasts. It is immediately clear from both panels that MinT(Shrink) is the best performing forecast reconciliation approach. In almost all cases the MinT(Shrink) reconciled forecasts show an improvement over the base forecasts. This is observed for both ARIMA and ETS base forecasts (with a few rare exceptions).

All other MinT based approaches also perform well showing mostly improvements over the base forecasts. The most poorly performing approach is clearly the bottom-up method showing some large losses for both sets of base forecasts but in particular for the ARIMA base forecasts. The biggest disadvantage of the bottom-up approach is that the series are modelled and forecasts are generated only at the most disaggregated bottom level. The series at the bottom level are highly disaggregated and therefore very noisy making it challenging to identify any seasonal and possibly trending components. These components become prominent at the higher levels of aggregation and are clearly visible at the top-most aggregated level in Figure 5. The MinT based approaches are able to bring this information to the levels below, improving the forecast accuracy of the base forecasts.

5 Conclusions and discussion

In the existing literature on forecasting hierarchical and grouped time series, the GLS estimator introduced by Hyndman et al. (2011) is the only available method which attempts to utilise information from the covariance structure between the series. However, in this paper we have shown that it is impossible to compute such a solution in practice due to identifiability conditions.

While overcoming this challenge, we have proposed a new combination approach for forecasting hierarchical and grouped time series. A remarkable feature of this approach is that the minimizer has an analytical solution which is identical to a GLS estimator. Fortunately, the required estimate of the covariance structure is estimable in practice, thus leading to a feasible GLS solution. By exploring the features of the matrices involved in the GLS estimator, we have been able to derive an alternative representation which involves inverting only a single matrix of lower dimension compared to the inversions involved in the original GLS solution. This makes our estimator scalable for handling large collections of time series.

A natural candidate estimator for the covariance matrix in the GLS solution is the sample covariance matrix, although it is not always positive-definite or well estimated with large collections of time series. Consequently, we have discussed several alternative estimates including structured and unstructured diagonal matrices and a shrinkage type estimator. Using a series of simulation designs and a sizeable empirical application we have illustrated that the proposed method clearly outperforms existing methods.

The covariance estimator is not only needed for computing the reconciled point forecasts, but for computing prediction intervals as shown in Lemma 1. We leave to a later paper a discussion of whether our shrinkage estimator is appropriate for this purpose, and whether the specific covariance structure of hierarchical and grouped time series can be exploited to obtain better a better covariance estimator than those considered here.

A Proofs

A.1 Lemma 1

Proof. Consider the following two regression models for data and forecasts respectively:

$$\begin{aligned} \boldsymbol{y}_{t+h} &= \boldsymbol{S}\boldsymbol{\beta}_t(h) + \boldsymbol{\eta}_h, \quad \boldsymbol{\eta}_h \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_h), \\ \hat{\boldsymbol{y}}_t(h) &= \boldsymbol{S}\boldsymbol{\beta}_t(h) + \boldsymbol{\varepsilon}_h, \quad \boldsymbol{\varepsilon}_h \sim \mathcal{N}(\boldsymbol{0}, \Sigma_h), \end{aligned}$$

where $t = 1, 2, \dots$, $\hat{\boldsymbol{y}}_t(h)$ are the h-step-ahead base forecasts using information up to and including time t, and $\boldsymbol{y}_{t+h}$ are the observed values of the series at time $t+h$. Note that $\boldsymbol{\eta}_h$ is aggregate consistent because $\boldsymbol{\beta}_t(h)$ and $\boldsymbol{y}_{t+h}$ are aggregate consistent. However, $\boldsymbol{\varepsilon}_h$ is not aggregate consistent since $\hat{\boldsymbol{y}}_t(h)$ is not in general.

The covariance matrix of the h-step-ahead base forecast errors is

$$\boldsymbol{W}_h = \mathrm{Var}[\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_t(h)] = \mathrm{Var}[\boldsymbol{\eta}_h - \boldsymbol{\varepsilon}_h] = \boldsymbol{\Omega}_h + \Sigma_h - \boldsymbol{\Gamma}_h - \boldsymbol{\Gamma}_h', \tag{10}$$

where $\boldsymbol{\Gamma}_h = \mathrm{E}[\boldsymbol{\eta}_h \boldsymbol{\varepsilon}_h']$. Recall from (1) that

$$\tilde{\boldsymbol{y}}_t(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_t(h) = \boldsymbol{S}\boldsymbol{P}[\boldsymbol{S}\boldsymbol{\beta}_t(h) + \boldsymbol{\varepsilon}_h] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{S}\boldsymbol{\beta}_t(h) + \boldsymbol{S}\boldsymbol{P}\boldsymbol{\varepsilon}_h = \boldsymbol{S}\boldsymbol{\beta}_t(h) + \boldsymbol{S}\boldsymbol{P}\boldsymbol{\varepsilon}_h$$

since $\boldsymbol{SPS} = \boldsymbol{S}$ for unbiased reconciled forecasts. Then the covariance matrix of the h-step-ahead reconciled forecast errors is given by,

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_t(h)] &= \mathrm{Var}[\boldsymbol{\eta}_h - \boldsymbol{SP}\boldsymbol{\varepsilon}_h] \\
&= \boldsymbol{\Omega}_h + \boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}' - \mathrm{E}(\boldsymbol{\eta}_h\boldsymbol{\varepsilon}_h'\boldsymbol{P}'\boldsymbol{S}') - [\mathrm{E}(\boldsymbol{\eta}_h\boldsymbol{\varepsilon}_h'\boldsymbol{P}'\boldsymbol{S}')]' \\
&= \boldsymbol{\Omega}_h + [\boldsymbol{SP}\Sigma_h - \mathrm{E}(\boldsymbol{\eta}_h\boldsymbol{\varepsilon}_h')]\boldsymbol{P}'\boldsymbol{S}' - [\mathrm{E}(\boldsymbol{\eta}_h\boldsymbol{\varepsilon}_h'\boldsymbol{P}'\boldsymbol{S}')]'.
\end{aligned}$$

Using equation (10),

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_t(h)] &= \boldsymbol{\Omega}_h + (\boldsymbol{SP}\Sigma_h + \boldsymbol{W}_h - \boldsymbol{\Omega}_h - \Sigma_h + \boldsymbol{\Gamma}_h')\boldsymbol{P}'\boldsymbol{S}' - \boldsymbol{SP}\boldsymbol{\Gamma}_h' \\
&= \boldsymbol{\Omega}_h + [-(\boldsymbol{I} - \boldsymbol{SP})\Sigma_h + \boldsymbol{W}_h - \boldsymbol{\Omega}_h + \boldsymbol{\Gamma}']\boldsymbol{P}'\boldsymbol{S}' - \boldsymbol{SP}\boldsymbol{\Gamma}_h' \\
&= -(\boldsymbol{I} - \boldsymbol{SP})\Sigma_h\boldsymbol{P}'\boldsymbol{S}' + \boldsymbol{\Omega}_h(\boldsymbol{I} - \boldsymbol{P}'\boldsymbol{S}') + \boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}' - \boldsymbol{SP}\boldsymbol{\Gamma}_h' + \boldsymbol{\Gamma}_h'\boldsymbol{P}'\boldsymbol{S}'. \qquad (11)
\end{aligned}$$

Let $\mathcal{I}_t = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$ denote the data available at time t. Then

$$\boldsymbol{\Omega}_h = \mathrm{Var}(\boldsymbol{y}_{t+h} \mid \mathcal{I}_t) = \mathrm{Var}(\boldsymbol{\eta}_h) = \boldsymbol{S}\mathrm{E}\{\boldsymbol{b}_{t+h} - \mathrm{E}[\boldsymbol{b}_{t+h} \mid \mathcal{I}_t]\}\{\boldsymbol{b}_{t+h} - \mathrm{E}[\boldsymbol{b}_{t+h} \mid \mathcal{I}_t]\}'\boldsymbol{S}',$$

yields $\boldsymbol{\Omega}_h(\boldsymbol{I} - \boldsymbol{P}'\boldsymbol{S}') = \boldsymbol{0}$ and $(\boldsymbol{I} - \boldsymbol{SP})\boldsymbol{\Omega}_h = \boldsymbol{0}$. Additionally (10) gives

$$(\boldsymbol{I} - \boldsymbol{SP})\boldsymbol{W}_h = (\boldsymbol{I} - \boldsymbol{SP})\Sigma_h - (\boldsymbol{I} - \boldsymbol{SP})[\boldsymbol{\Gamma}_h + \boldsymbol{\Gamma}_h']$$

and

$$\begin{aligned}
\boldsymbol{\Gamma}_h &= \mathrm{E}[\boldsymbol{\eta}_h\boldsymbol{\varepsilon}_h'] \\
&= \mathrm{E}\{\boldsymbol{y}_{t+h} - \mathrm{E}[\boldsymbol{y}_{t+h} \mid \mathcal{I}_t]\}\{\hat{\boldsymbol{y}}_t(h) - \mathrm{E}[\hat{\boldsymbol{y}}_t(h)]\}' \\
&= \boldsymbol{S}\mathrm{E}\{\boldsymbol{b}_{t+h} - \mathrm{E}[\boldsymbol{b}_{t+h} \mid \mathcal{I}_t]\}\{\hat{\boldsymbol{y}}_t(h) - \mathrm{E}[\hat{\boldsymbol{y}}_t(h)]\}',
\end{aligned}$$

gives $(\boldsymbol{I} - \boldsymbol{SP})\boldsymbol{\Gamma}_h = \boldsymbol{0}$.

Applying the above derived results on (11) leads to (7):

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_t(h)] &= -\Big\{(\boldsymbol{I} - \boldsymbol{SP})\boldsymbol{W}_h + (\boldsymbol{I} - \boldsymbol{SP})\boldsymbol{\Gamma}_h'\Big\}\boldsymbol{P}'\boldsymbol{S}' + \boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}' - \boldsymbol{SP}\boldsymbol{\Gamma}_h' + \boldsymbol{\Gamma}_h'\boldsymbol{P}'\boldsymbol{S}' \\
&= \boldsymbol{SPW}_h\boldsymbol{P}'\boldsymbol{S}' - (\boldsymbol{I} - \boldsymbol{SP})\mathrm{E}\Big\{\hat{\boldsymbol{y}}_t(h) - \mathrm{E}[\hat{\boldsymbol{y}}_t(h)]\Big\}\Big\{\boldsymbol{S}\boldsymbol{b}_{t+h} - \mathrm{E}[\boldsymbol{S}\boldsymbol{b}_{t+h} \mid \mathcal{I}_t]\Big\}'\boldsymbol{P}'\boldsymbol{S}' - \boldsymbol{SP}\boldsymbol{\Gamma}_h' + \boldsymbol{\Gamma}_h'\boldsymbol{P}'\boldsymbol{S}' \\
&= \boldsymbol{SPW}_h\boldsymbol{P}'\boldsymbol{S}' - (\boldsymbol{I} - \boldsymbol{SP})\mathrm{E}\Big\{\hat{\boldsymbol{y}}_t(h) - \mathrm{E}[\hat{\boldsymbol{y}}_t(h)]\Big\}\Big\{\boldsymbol{S}\boldsymbol{b}_{t+h} - \mathrm{E}[\boldsymbol{S}\boldsymbol{b}_{t+h} \mid \mathcal{I}_t]\Big\}' - \boldsymbol{SP}\boldsymbol{\Gamma}_h' + \boldsymbol{\Gamma}_h'\boldsymbol{P}'\boldsymbol{S}' \\
&= \boldsymbol{SPW}_h\boldsymbol{P}'\boldsymbol{S}' - (\boldsymbol{I} - \boldsymbol{SP})\mathrm{E}\Big\{\hat{\boldsymbol{y}}_t(h) - \mathrm{E}[\hat{\boldsymbol{y}}_t(h)]\Big\}\Big\{\boldsymbol{y}_{t+h} - \mathrm{E}[\boldsymbol{y}_{t+h} \mid \mathcal{I}_t]\Big\}' - \boldsymbol{SP}\boldsymbol{\Gamma}_h' + \boldsymbol{\Gamma}_h'\boldsymbol{P}'\boldsymbol{S}'
\end{aligned}$$

$$
\begin{aligned}
&= \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}' - (\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P})\boldsymbol{\Gamma}'_h - \boldsymbol{S}\boldsymbol{P}\boldsymbol{\Gamma}'_h + \boldsymbol{\Gamma}'_h\boldsymbol{P}'\boldsymbol{S}' \\
&= \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}' - \boldsymbol{\Gamma}'_h + \boldsymbol{\Gamma}'_h\boldsymbol{P}'\boldsymbol{S}' \\
&= \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}' - \boldsymbol{\Gamma}'_h(\boldsymbol{I} - \boldsymbol{P}'\boldsymbol{S}') \\
&= \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}'.
\end{aligned}
$$

□

A.2 Theorem 1

Proof. Reformulate the objective function such that,

$$\mathrm{tr}[\boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}'] = \mathrm{tr}[\boldsymbol{S}'\boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'].$$

As $\boldsymbol{S}'\boldsymbol{S}$ and $\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'$ are both symmetric, and positive definite and positive semidefinite respectively, using Lemma 1 of Wang, Kuo, and Hsu (1986) we have

$$\mathrm{tr}[\boldsymbol{S}'\boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'] \geq \lambda_{\min}(\boldsymbol{S}'\boldsymbol{S})\mathrm{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'] \geq \lambda_{\min}(\boldsymbol{C}'\boldsymbol{C} + \boldsymbol{I})\mathrm{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'],$$

where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of matrix $\boldsymbol{A}$ and $\boldsymbol{S} = \begin{bmatrix}\boldsymbol{C} \\ \boldsymbol{I}_n\end{bmatrix}$. Since $\boldsymbol{C}'\boldsymbol{C}$ and $\boldsymbol{I}$ are both symmetric, and positive semidefinite and positive definite respectively, applying Weyl's inequalities (Horn and Johnson, 1990),

$$\lambda_{\min}(\boldsymbol{C}'\boldsymbol{C} + \boldsymbol{I})\mathrm{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'] \geq [\lambda_{\min}(\boldsymbol{C}'\boldsymbol{C}) + \lambda_{\min}(\boldsymbol{I})]\mathrm{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'] \geq \mathrm{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'],$$

using the fact that all the eigenvalues of an identity matrix are unity.

Now $\boldsymbol{P}\boldsymbol{S} = \boldsymbol{I}_n$, so the minimization problem can be restated as

$$\min_{\boldsymbol{P}} \quad \mathrm{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'] \qquad \text{such that} \quad \boldsymbol{P}\boldsymbol{S} = \boldsymbol{I}.$$

If $\boldsymbol{W}_h$ is positive definite and we let $\boldsymbol{L} = \boldsymbol{W}_h^{-\frac{1}{2}}\boldsymbol{S}$ and $\boldsymbol{H} = \boldsymbol{P}\boldsymbol{W}_h^{\frac{1}{2}}$, the minimization problem becomes

$$\min_{\boldsymbol{H}} \quad \mathrm{tr}[\boldsymbol{H}\boldsymbol{H}'] \qquad \text{such that} \quad \boldsymbol{H}\boldsymbol{L} = \boldsymbol{I}.$$

The unique solution of this well-known minimization problem is given by the Moore-Penrose generalized inverse of $\boldsymbol{L}$ (Penrose, 1956). As we know that $\boldsymbol{L}$ is full rank, $\boldsymbol{H} = (\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'$ which gives

$$P = (S'W_h^{-1}S)^{-1}S'W_h^{-1}.$$

Now let W_h be partitioned as

$$W_h = \left[\begin{array}{c:c} W_{11} & W_{12} \\ \hdashline W_{12}' & W_{22} \end{array}\right],$$

where W_{11}, W_{12} and W_{22} are matrices of dimension $m^* \times m^*$, $m^* \times n$ and $n \times n$. Assuming that W_{22} and $W_{11} - W_{12}W_{22}^{-1}W_{12}'$ are non-singular, we can write the inverse of W_h (using Fact 2.15.3 of Bernstein, 2005) as

$$\begin{aligned}
& W_h^{-1} = \left[\begin{array}{c:c} W_{11}^* & W_{12}^* \\ \hdashline W_{12}^{*'} & W_{22}^* \end{array}\right], \\
\text{where} \quad & W_{11}^* = \left(W_{11} - W_{12}W_{22}^{-1}W_{12}'\right)^{-1} \\
& W_{12}^* = -\left(W_{11} - W_{12}W_{22}^{-1}W_{12}'\right)^{-1} W_{12}W_{22}^{-1} \\
\text{and} \quad & W_{22}^* = W_{22}^{-1}W_{12}'\left(W_{11} - W_{12}W_{22}^{-1}W_{12}'\right)^{-1} W_{12}W_{22}^{-1} + W_{22}^{-1}.
\end{aligned}$$

Then if we partition $S' = \left[\begin{array}{c:c} C' & I_n \end{array}\right]$, it is easy to show that matrix P in Theorem 1 can be written as

$$\begin{aligned}
& P = (S'W_h^{-1}S)^{-1}S'W_h^{-1} = \left[\begin{array}{c:c} P_1^* & P_2^* \end{array}\right], \\
\text{where} \quad & P_1^* = A(C'W_{11}^* + W_{12}^{*'}), \\
& P_2^* = A(C'W_{12}^* + W_{22}^*), \\
\text{and} \quad & A = (C'W_{11}^*C + W_{12}^{*'}C + C'W_{12}^* + W_{22}^*)^{-1}.
\end{aligned}$$

Let's consider the two terms separately.

$$\begin{aligned}
P_1^* &= (C'W_{11}^*C + W_{12}^{*'}C + C'W_{12}^* + W_{22}^*)^{-1}(C'W_{11}^* + W_{12}^{*'}) \\
&= \left[(C' - W_{22}^{-1}W_{12}')(W_{11} - W_{12}W_{22}^{-1}W_{12}')^{-1}(C - W_{12}W_{22}^{-1}) + W_{22}^{-1}\right]^{-1} \\
&\qquad\qquad \times (C' - W_{22}^{-1}W_{12}')(W_{11} - W_{12}W_{22}^{-1}W_{12}')^{-1}.
\end{aligned}$$

Assuming that $I + \left(C - W_{12}W_{22}^{-1}\right)W_{22}\left(C' - W_{22}^{-1}W_{12}'\right)\left(W_{11} - W_{12}W_{22}^{-1}W_{12}'\right)^{-1}$ is non-singular, Fact 2.14.19 of Bernstein (2005) gives,

$$\begin{aligned}
P_1^* = {} & W_{22}(C' - W_{22}^{-1}W_{12}')(W_{11} - W_{12}W_{22}^{-1}W_{12}')^{-1} \\
& \times \left[I + (C - W_{12}W_{22}^{-1})W_{22}(C' - W_{22}^{-1}W_{12}')(W_{11} - W_{12}W_{22}^{-1}W_{12}')^{-1}\right]^{-1}
\end{aligned}$$

$$
\begin{aligned}
&= (\boldsymbol{W}_{22}\boldsymbol{C}' - \boldsymbol{W}_{12}')\Big[(\boldsymbol{W}_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}') \\
&\qquad + (\boldsymbol{C}\boldsymbol{W}_{22} - \boldsymbol{W}_{12})(\boldsymbol{C}' - \boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')(\boldsymbol{W}_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')^{-1}(\boldsymbol{W}_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')\Big]^{-1} \\
&= -(\boldsymbol{W}_{12}' - \boldsymbol{W}_{22}\boldsymbol{C}')\Big(\boldsymbol{W}_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}' + \boldsymbol{C}\boldsymbol{W}_{22}\boldsymbol{C}' - \boldsymbol{C}\boldsymbol{W}_{12}' - \boldsymbol{W}_{12}\boldsymbol{C}' + \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}'\Big)^{-1} \\
&= -(\boldsymbol{W}_{12}' - \boldsymbol{W}_{22}\boldsymbol{C}')(\boldsymbol{W}_{11} - \boldsymbol{C}\boldsymbol{W}_{12}' - \boldsymbol{W}_{12}\boldsymbol{C}' + \boldsymbol{C}\boldsymbol{W}_{22}\boldsymbol{C}')^{-1}.
\end{aligned}
$$

Similarly, simplifying the matrix $\boldsymbol{P}_2^*$ gives

$$
\boldsymbol{P}_2^* = \boldsymbol{X}^{-1}\Big[-(\boldsymbol{C}' - \boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')(\boldsymbol{W}_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')^{-1}\boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1} + \boldsymbol{W}_{22}^{-1}\Big]
$$

where $\boldsymbol{X} = \Big[(\boldsymbol{C}' - \boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')(\boldsymbol{W}_{22} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')^{-1}(\boldsymbol{C} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}) + \boldsymbol{W}_{22}^{-1}\Big]$.

Now re-write the above expression as

$$
\begin{aligned}
\boldsymbol{P}_2^* &= \boldsymbol{X}^{-1}\Big[(\boldsymbol{C}' - \boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')(\boldsymbol{W}_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')^{-1}(\boldsymbol{C} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}) + \boldsymbol{W}_{22}^{-1} \\
&\qquad\qquad - (\boldsymbol{C}' - \boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')(\boldsymbol{W}_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')^{-1}\boldsymbol{C}\Big] \\
&= \boldsymbol{X}^{-1}\Big[\boldsymbol{X} - (\boldsymbol{C}' - \boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')(\boldsymbol{W}_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')^{-1}\boldsymbol{C}\Big] \\
&= \boldsymbol{I} - \boldsymbol{X}^{-1}(\boldsymbol{C}' - \boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')(\boldsymbol{W}_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{12}')^{-1}\boldsymbol{C} \\
&= \boldsymbol{I} - \boldsymbol{P}_1^* \\
&= \boldsymbol{I} + (\boldsymbol{W}_{12}' - \boldsymbol{W}_{22}\boldsymbol{C}')(\boldsymbol{W}_{11} - \boldsymbol{C}\boldsymbol{W}_{12}' - \boldsymbol{W}_{12}\boldsymbol{C}' + \boldsymbol{C}\boldsymbol{W}_{22}\boldsymbol{C}')^{-1}.
\end{aligned}
$$

Combining the simplified results of $\boldsymbol{P}_1^*$ and $\boldsymbol{P}_2^*$, the matrix $\boldsymbol{P}$ can be written as,

$$
\begin{aligned}
\boldsymbol{P} &= \begin{pmatrix} \boldsymbol{0}_{n\times m^*} & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{0}_{n\times m^*} & \boldsymbol{I} \end{pmatrix}\begin{pmatrix} \boldsymbol{W}_{11} & \boldsymbol{W}_{12} \\ \boldsymbol{W}_{12}' & \boldsymbol{W}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{I} \\ -\boldsymbol{C}' \end{pmatrix}\left[\begin{pmatrix} \boldsymbol{I} & -\boldsymbol{C} \end{pmatrix}\begin{pmatrix} \boldsymbol{W}_{11} & \boldsymbol{W}_{12} \\ \boldsymbol{W}_{12}' & \boldsymbol{W}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{I} \\ -\boldsymbol{C}' \end{pmatrix}\right]^{-1}\begin{pmatrix} \boldsymbol{I} & -\boldsymbol{C} \end{pmatrix} \\
&= \boldsymbol{J} - \boldsymbol{J}\boldsymbol{W}_h\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{W}_h\boldsymbol{U})^{-1}\boldsymbol{U}',
\end{aligned}
$$

where $\boldsymbol{J} = \begin{bmatrix} \boldsymbol{0}_{n\times m^*} & \boldsymbol{I}_n \end{bmatrix}$, $\boldsymbol{U}' = \begin{bmatrix} \boldsymbol{I}_{m^*} & -\boldsymbol{C} \end{bmatrix}$, and $m^* = m - n$. □

B Tourism hierarchy

Table 3: *Geographical division of Australia.*

Series	Name	Label	Series	Name	Label
Total			*Regions continued*		
1	Australia	Total	55	Lakes	BCA
States			56	Gippsland	BCB
2	NSW	A	57	Phillip Island	BCC
3	VIC	B	58	Central Murray	BDA
4	QLD	C	59	Goulburn	BDB
5	SA	D	60	High Country	BDC
6	WA	E	61	Melbourne	BDD
7	TAS	F	62	Upper Yarra	BDE
8	NT	G	63	Murray East	BDF
Zones			64	Wimmera+Mallee	BEA
9	Metro NSW	AA	65	Western Grampians	BEB
10	Nth Coast NSW	AB	66	Bendigo Loddon	BEC
11	Sth Coast NSW	AC	67	Macedon	BED
12	Sth NSW	AD	68	Spa Country	BEE
13	Nth NSW	AE	69	Ballarat	BEF
14	ACT	AF	70	Central Highlands	BFG
15	Metro VIC	BA	71	Gold Coast	CAA
16	West Coast VIC	BB	72	Brisbane	CAB
17	East Coast VIC	BC	73	Sunshine Coast	CAC
18	Nth East VIC	BD	74	Central Queensland	CBA
19	Nth West VIC	BE	75	Bundaberg	CBB
20	Metro QLD	CA	76	Fraser Coast	CBC
21	Central Coast QLD	CB	77	Mackay	CBD
22	Nth Coast QLD	CC	78	Whitsundays	CCA
23	Inland QLD	CD	79	Northern	CCB
24	Metro SA	DA	80	Tropical North Queensland	CCC
25	Sth Coast SA	DB	81	Darling Downs	CDA
26	Inland SA	DC	82	Outback	CDB
27	West Coast SA	DD	83	Adelaide	DAA
28	West Coast WE	EA	84	Barossa	DAB
29	Nth WA	EB	85	Adelaide Hills	DAC
30	Sth WA	EC	86	Limestone Coast	DBA
31	Sth TAS	FA	87	Fleurieu Peninsula	DBB
32	Nth East TAS	FB	88	Kangaroo Island	DBC
33	Nth West TAS	FC	89	Murraylands	DCA
34	Nth Coast NT	GA	90	Riverland	DCB
35	Central NT	GB	91	Clare Valley	DCC
Regions			92	Flinders Range and Outback	DCD
36	Sydney	AAA	93	Eyre Peninsula	DDA
37	Central Coast	AAB	94	Yorke Peninsula	DDB
38	Hunter	ABA	95	Australia's Coral Coast	EAA
39	North Coast NSW	ABB	96	Experience Perth	EAB
40	Northern Rivers Tropical NSW	ABC	97	Australia's South West	EAC
41	South Coast	ACA	98	Australia's North West	EBA
42	Snowy Mountains	ADA	99	Australia's Golden Outback	ECA
43	Capital Country	ADB	100	Hobart and the South	FAA
44	The Murray	ADC	101	East Coast	FBA
45	Riverina	ADD	102	Launceston, Tamar and the North	FBB
46	Central NSW	AEA	103	North West	FCA
47	New England North West	AEB	104	Wilderness West	FCB
48	Outback NSW	AEC	105	Darwin	GAA
49	Blue Mountains	AED	106	Kakadu Arnhem	GAB
50	Canberra	AFA	107	Katherine Daly	GAC
51	Melbourne	BAA	108	Barkly	GBA
52	Peninsula	BAB	109	Lasseter	GBB
53	Geelong	BAC	110	Alice Springs	GBC
54	Western	BBA	111	MacDonnell	GBD

C Tables

Table 4: *Out-of-sample forecast performance.*

	Simulation design 2									Simulation design 3								
	$T = 60$			$T = 180$			$T = 300$			$T = 60$			$T = 180$			$T = 300$		
	$h = 1$	*1–4*	*1–8*	$h = 1$	*1–8*	*1–16*	$h = 1$	*1–12*	*1–24*	$h = 1$	*1–4*	*1–8*	$h = 1$	*1–8*	*1–16*	$h = 1$	*1–12*	*1–24*
	Top level									Top level								
BU	**-10.8**	-6.6	-4.1	-7.3	-3.8	-4.0	-7.4	-2.6	-3.2	50.2	26.6	12.7	33.2	14.9	6.0	43.4	7.2	2.7
OLS	-5.5	-5.3	-5.2	-3.9	-3.2	-3.6	-3.6	-2.2	-2.5	1.1	-1.5	**-3.5**	-1.0	-1.7	**-2.5**	0.5	**-3.3**	**-3.3**
WLS_v	-9.3	**-7.7**	**-6.7**	-6.6	-4.6	-5.2	-6.4	**-3.3**	**-3.9**	-1.1	**-2.0**	-3.1	-3.9	-2.0	-2.3	-1.3	-2.8	-2.7
WLS_s	-8.1	-7.1	-6.5	-5.7	-4.3	-4.9	-5.4	-3.1	-3.5	9.1	2.6	-1.9	4.0	0.8	-1.7	7.7	-2.4	-3.2
MinT(Sample)	-9.4	-5.1	-3.9	-8.0	-3.8	-4.2	-8.3	-2.6	-2.9	0.9	-0.7	-1.6	**-6.1**	-1.6	-1.8	**-2.8**	-2.2	-2.3
MinT(Shrink)	-10.3	**-7.7**	-6.6	**-8.2**	**-4.7**	**-5.3**	**-8.6**	-3.1	-3.6	**-1.3**	-1.9	-2.9	-5.8	**-2.1**	-2.2	-2.5	-2.4	-2.5
	Level 1									Level 1								
BU	**-5.5**	-2.2	-0.1	-2.9	0.0	0.6	-4.9	-0.6	-0.7	10.5	6.1	2.2	6.6	3.2	1.3	7.4	2.8	2.0
OLS	-1.0	-1.2	-1.2	0.1	0.3	0.5	-1.3	-0.5	-0.4	-4.0	-4.9	-5.4	-3.4	-4.0	-3.4	-3.6	-2.6	-1.7
WLS_v	-4.2	-3.3	**-2.6**	-2.2	**-0.9**	**-0.7**	-3.7	**-1.4**	**-1.4**	**-4.6**	**-5.3**	**-5.5**	-4.0	-4.5	-3.6	-4.2	-2.7	-1.6
WLS_s	-3.2	-2.8	-2.4	-1.5	-0.6	-0.5	-3.0	-1.1	-1.1	-2.2	-3.8	-5.1	-2.4	-3.3	-3.1	-2.2	-2.4	-1.7
MinT(Sample)	-4.6	-0.8	0.7	-3.8	0.2	0.5	-6.0	-0.8	-0.6	-1.3	-3.2	-3.2	**-5.2**	-4.7	-3.1	-4.0	-2.9	-1.7
MinT(Shrink)	-5.4	**-3.3**	-2.3	**-4.0**	-0.7	-0.5	**-6.2**	-1.3	-1.2	-4.1	-5.2	-5.2	-5.1	**-5.2**	**-3.8**	**-4.2**	**-3.0**	**-1.8**
	Bottom level									Bottom level								
BU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OLS	3.6	0.9	-0.7	2.3	0.2	-0.1	2.8	0.1	0.2	-2.8	-2.9	-2.4	-2.0	-2.3	-1.8	-2.2	-1.9	-1.5
WLS_v	1.5	-0.7	-1.7	0.7	-0.6	-0.9	1.4	-0.5	**-0.5**	-3.3	-3.6	-3.1	-2.5	-3.1	-2.4	-2.8	-2.5	**-1.9**
WLS_s	1.8	-0.3	-1.5	1.0	-0.4	-0.7	1.5	-0.3	-0.3	-2.4	-2.6	-2.4	-1.8	-2.1	-1.7	-2.0	-1.9	-1.6
MinT(Sample)	0.1	1.0	0.1	-1.5	0.0	-0.6	**-1.7**	-0.2	-0.1	-2.1	-2.2	-0.9	**-3.9**	-3.2	-1.2	**-4.5**	-2.5	-0.8
MinT(Shrink)	**-0.4**	**-1.3**	**-2.2**	**-1.5**	**-0.8**	**-1.3**	**-1.7**	**-0.6**	**-0.5**	**-3.8**	**-4.1**	**-3.4**	**-3.9**	**-4.1**	**-2.6**	-4.4	**-3.1**	-1.7
	Average									Average								
BU	**-5.0**	-2.6	-1.2	-3.0	-1.1	-0.9	-3.8	-0.9	-1.1	8.2	5.6	2.9	5.7	3.5	1.6	6.6	2.2	1.2
OLS	-0.6	-1.6	-2.1	-0.2	-0.7	-0.9	-0.4	-0.7	-0.7	-2.7	-3.3	-3.5	-2.3	-2.7	-2.4	-2.3	-2.4	-1.9
WLS_v	-3.5	-3.5	**-3.4**	-2.4	-1.8	-2.0	-2.5	**-1.5**	**-1.7**	-3.4	-3.9	-3.9	-3.1	-3.3	-2.7	-3.0	-2.6	**-2.0**
WLS_s	-2.7	-3.1	-3.2	-1.8	-1.6	-1.8	-1.9	-1.3	-1.4	-1.1	-2.2	-3.1	-1.3	-2.0	-2.1	-1.0	-2.1	-2.0
MinT(Sample)	-4.2	-1.3	-0.8	-4.1	-1.0	-1.2	-5.0	-1.0	-1.0	-1.6	-2.3	-1.7	**-4.5**	-3.4	-1.9	**-4.2**	-2.6	-1.4
MinT(Shrink)	**-5.0**	**-3.8**	**-3.4**	**-4.2**	**-1.8**	**-2.1**	**-5.2**	**-1.5**	-1.6	**-3.6**	**-4.1**	**-3.9**	**-4.5**	**-4.1**	**-2.9**	**-4.2**	**-3.0**	-1.9

Note: Each entry shows the percentage difference in the average RMSE between reconciled forecasts and base forecasts. A negative (positive) entry shows a percentage decrease (increase) in average RMSE relative to the base forecasts. The bold entries identify the best performing method.

Table 5: *Out-of-sample forecast performance for Australian domestic tourism flows.*

	h = 1	2	3	6	12	1–6	1–12	*h* = 1	2	3	6	12	1–6	1–12
	ARIMA													
	Australia							Australia by purpose of travel						
BU	33.2	25.9	23.0	24.3	20.3	26.2	25.1	41.7	34.3	33.3	32.6	24.7	34.6	31.8
OLS	**-2.5**	-2.7	-3.0	-2.0	-1.5	-2.4	-2.0	**0.6**	**0.3**	**0.7**	**-0.3**	-0.8	**0.1**	**-0.2**
WLS_v	1.1	-1.4	-2.7	-0.0	-2.1	-0.7	-0.4	9.7	7.3	7.0	6.6	3.4	7.2	6.2
WLS_s	-0.5	-2.9	-3.9	-1.3	-2.6	-2.1	-1.6	8.5	5.9	5.8	5.2	2.6	6.0	5.1
MinT(Shrink)	-2.3	**-4.7**	**-5.5**	**-2.3**	**-3.1**	**-3.6**	**-2.8**	2.9	0.9	0.8	0.6	**-1.9**	0.9	0.3
	States							States by purpose of travel						
BU	15.0	14.4	12.1	12.4	7.6	13.4	12.1	10.5	9.3	8.7	10.2	7.8	9.6	9.3
OLS	-2.2	-1.2	-1.1	-1.9	-2.3	-1.5	-1.8	-0.4	-0.0	0.2	0.2	0.5	0.1	0.0
WLS_v	-4.1	-3.2	-4.2	-3.7	-6.1	-3.7	-4.0	-3.9	-3.6	-3.7	-2.9	-3.1	-3.4	-3.4
WLS_s	-4.5	-3.7	-4.4	-4.1	-6.0	-4.0	-4.3	-3.3	-3.2	-3.2	-2.5	-2.5	-3.0	-2.9
MinT(Shrink)	**-5.5**	**-4.6**	**-5.4**	**-4.7**	**-6.5**	**-5.0**	**-5.1**	**-5.7**	**-5.4**	**-5.3**	**-4.7**	**-4.8**	**-5.2**	**-5.1**
	Zones							Zones by purpose of travel						
BU	5.9	6.1	6.5	6.2	5.3	6.2	6.0	3.5	3.1	3.0	3.0	2.4	3.1	3.1
OLS	-2.4	-1.9	-0.7	-1.2	-0.3	-1.5	-1.2	0.9	0.7	0.9	0.4	0.5	0.7	0.6
WLS_v	-4.0	-2.8	-2.0	-2.1	-2.4	-2.6	-2.5	-3.4	-3.0	-2.9	-3.0	-3.0	-3.1	-3.0
WLS_s	-4.3	-3.6	-2.7	-2.8	-2.9	-3.2	-3.1	-2.3	-2.4	-2.2	-2.5	-2.4	-2.4	-2.3
MinT(Shrink)	**-5.4**	**-4.2**	**-3.3**	**-3.4**	**-3.5**	**-3.9**	**-3.7**	**-4.5**	**-4.0**	**-3.9**	**-4.0**	**-4.1**	**-4.1**	**-4.0**
	Regions							Regions by purpose of travel						
BU	2.7	1.7	1.8	2.0	2.0	1.9	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OLS	-1.5	-1.7	-1.5	-1.5	-0.6	-1.6	-1.3	1.4	1.4	1.4	1.1	1.3	1.3	1.2
WLS_v	-3.5	-3.5	-3.4	-3.0	-2.7	-3.3	-3.1	-3.5	-3.1	-3.0	-3.0	-2.8	-3.1	-3.1
WLS_s	-3.2	-3.4	-3.2	-2.9	-2.5	-3.2	-2.9	-1.6	-1.5	-1.4	-1.6	-1.4	-1.5	-1.5
MinT(Shrink)	**-4.2**	**-4.3**	**-4.1**	**-3.6**	**-3.2**	**-4.0**	**-3.7**	**-4.1**	**-3.6**	**-3.5**	**-3.5**	**-3.3**	**-3.6**	**-3.5**
	ETS													
	Australia							Australia by purpose of travel						
BU	7.0	4.0	4.3	2.5	1.8	3.9	4.1	7.0	6.9	7.8	7.0	4.4	7.0	6.5
OLS	**-0.2**	-1.6	-1.5	-1.2	-1.1	-1.2	-1.1	**-1.5**	0.3	0.3	-0.3	-0.9	-0.2	-0.4
WLS_v	1.5	-2.8	-2.4	-2.4	-3.1	-1.9	-1.7	-0.3	-0.3	0.2	-0.2	-3.0	-0.2	-0.6
WLS_s	0.9	**-3.1**	-2.7	-2.5	-3.0	**-2.2**	-1.9	-0.6	-0.2	0.2	-0.2	-2.7	-0.2	-0.6
MinT(Shrink)	1.1	**-3.1**	**-2.9**	**-2.6**	**-3.2**	**-2.2**	**-2.0**	-1.3	**-1.1**	**-1.0**	**-1.0**	**-3.4**	**-1.1**	**-1.4**
	States							States by purpose of travel						
BU	6.1	6.5	4.9	5.1	3.5	5.3	5.2	2.4	3.0	2.0	2.5	3.3	2.4	2.5
OLS	**0.0**	0.6	-0.1	0.8	-0.1	0.4	0.4	-1.1	-0.2	-0.6	-0.7	0.0	-0.7	-0.7
WLS_v	0.7	0.4	-0.3	0.5	-1.0	0.3	0.5	-1.4	-1.0	-1.5	-1.5	-1.3	-1.4	-1.5
WLS_s	0.6	0.3	-0.4	0.4	-0.9	0.2	0.3	-1.3	-0.7	-1.3	-1.2	-1.0	-1.2	-1.3
MinT(Shrink)	0.6	**0.2**	**-0.6**	**0.3**	**-1.3**	**0.1**	**0.2**	**-2.0**	**-1.4**	**-2.1**	**-2.0**	**-1.8**	**-1.9**	**-2.0**
	Zones							Zones by purpose of travel						
BU	2.7	2.9	2.6	2.4	4.1	2.6	3.0	0.8	0.9	0.4	1.2	1.7	0.8	1.0
OLS	-0.9	-0.6	-0.6	-0.7	0.1	-0.7	-0.6	-0.4	-0.5	-0.6	-0.4	-0.1	-0.5	-0.4
WLS_v	-0.7	-0.8	-0.7	-1.1	-0.1	-0.8	-0.6	-1.4	-1.4	-1.7	-1.4	-1.3	-1.5	-1.5
WLS_s	-0.9	-1.0	-0.9	-1.3	-0.2	-1.0	-0.8	-1.1	-1.1	-1.3	-1.1	-0.8	-1.2	-1.2
MinT(Shrink)	**-1.0**	**-1.1**	**-1.1**	**-1.4**	**-0.6**	**-1.1**	**-1.0**	**-1.9**	**-1.8**	**-2.1**	**-1.9**	**-1.9**	**-1.9**	**-2.0**
	Regions							Regions by purpose of travel						
BU	-0.1	0.4	0.6	1.4	2.2	0.6	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OLS	-1.2	-1.0	-0.9	-0.6	-0.6	-0.9	-0.9	0.9	0.8	0.9	0.7	0.7	0.8	0.7
WLS_v	-2.1	-1.9	-1.6	-1.4	-1.0	-1.7	-1.5	-1.1	-1.2	-1.1	-1.4	-1.5	-1.2	-1.3
WLS_s	-1.9	-1.7	-1.5	-1.2	-0.9	-1.5	-1.4	-0.3	-0.4	-0.3	-0.5	-0.5	-0.4	-0.4
MinT(Shrink)	**-2.4**	**-2.2**	**-2.0**	**-1.7**	**-1.5**	**-2.0**	**-1.9**	**-1.6**	**-1.6**	**-1.6**	**-1.9**	**-2.0**	**-1.7**	**-1.8**

Note: Each entry shows the percentage difference in average RMSE between reconciled forecasts and base forecasts. A negative (positive) entry shows a percentage decrease (increase) in average RMSE relative to the base forecasts. The bold entries identify the best performing method.

References

Athanasopoulos, G, Ahmed, RA, and Hyndman, RJ (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting* **25**, 146–166.

Athanasopoulos, G, Hyndman, RJ, Kourentzes, N, and Petropoulos, F (2015). *Forecasting with Temporal Hierarchies*. Working paper 16/15. Department of Econometrics & Business Statistics, Monash University.

Bernstein, DS (2005). *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press.

Byron, RP (1978). The Estimation of Large Social Account Matrices. *Journal of Royal Statistical Society, Series A* **141**(3), 359–367.

Dagum, EB and Cholette, PA (2006). *Benchmarking, Temporal, Distribution, and Reconciliation Methods for Time Series*. Springer-Verlag.

Horn, RA and Johnson, CR (1990). *Matrix Analysis*. Cambridge University Press.

Hyndman, RJ (2015). *forecast: Forecasting functions for time series and linear models*. R package version 6.2. http://github.com/robjhyndman/forecast.

Hyndman, RJ and Khandakar, Y (2008). Automatic Time Series Forecasting : The forecast Package for R. *Journal of Statistical Software* **27**(3).

Hyndman, RJ, Lee, AJ, and Wang, E (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics and Data Analysis*. to appear.

Hyndman, RJ, Ahmed, RA, Athanasopoulos, G, and Shang, HL (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis* **55**, 2579–2589.

Park, M and Nassar, M (2014). “Variational Bayesian inference for forecasting hierarchical time series”. www.gatsby.ucl.ac.uk/~mijung/ICMLworkshop_PARK_NASSAR.pdf.

Penrose, R (1956). On best approximate solutions of linear matrix equations. *Mathematical Proceedings of the Cambridge Philosophical Society* **52**, 17–19.

Schäfer, J and Strimmer, K (2005). A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology* **4**(1), 32.

Stone, R (1976). “The development of economic data systems”. In: *Social Accounting for Development Planning with Special Reference to Sri Lanka*. By Graham Pyatt and Alan Roe. Cambridge University Press, pp.xvi–xvii.

Tourism Research Australia (2015). *Tourism forecasts*. Tech. rep. Canberra. arXiv: arXiv:1011.1669v3.

van Erven, T and Cugliari, J (2014). “Game-theorically Optimal Reconciliation of Contemporaneous Hierarchical Time Series Forecasts”. In: *Modeling and Stochastic Learning for Forecasting in High Dimension*. Ed. by A Antoniadis, X Brossat, and JM Poggi. Springer Lecture Notes in Statistics. Paris: Springer, pp. 297–317.

Wang, SD, Kuo, TS, and Hsu, CF (1986). Trace Bounds on the Solution of the Algebraic Matrix Riccati and Lyapunov Equation. *IEEE Transactions on Autimatic Control* **AC-31**, 1986.