

Optimal forecast reconciliation for 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

15 June 2017

Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization

Abstract

Large collections of time series often have aggregation constraints due to product or geographical groupings. The forecasts for the most disaggregated series are usually required to add-up exactly to the forecasts of the aggregated series, a constraint we refer to as “coherent forecasts”. The combination forecasts proposed by Hyndman, Ahmed, 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 incoherent forecasts). We show that this is impossible to estimate in practice due to identifiability conditions.

We propose a new reconciliation forecasting approach that incorporates the information from a full covariance matrix of forecast errors in obtaining a set of coherent forecasts. Our approach minimizes the mean squared error of the coherent 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 that 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 comprising 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 desired 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 add the results to form forecasts of the various aggregated series. This is known as “bottom-up” forecasting (see Dunn, Williams, and Dechaine, 1976; Orcutt, Watt, and Edwards, 1968; Shlifer and Wolff, 1979, among others). 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 coherent. Further, this approach also ignores the relationships between series.

Hyndman, Ahmed, 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 base forecasts are unbiased and 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, Ahmed, 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, Ahmed, 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 easily be generalized to handle grouped but non-hierarchical collections of time series.

Our approach is to extend the work of Hyndman, Ahmed, et al. (2011) and Hyndman, Lee, and Wang (2016) 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. We refer to the approach as MinT (minimum trace) reconciliation in what follows.

An interesting feature of the MinT approach 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.

A related strand of literature considers incoherencies in the temporal dimension. Specifically, this literature deals with forecasts of a time series observed across different sampling frequencies. Two of the latest contributions are Athanasopoulos, Hyndman, et al. (2017) and Sayal et al. (2016). Athanasopoulos, Hyndman, et al. (2017) build what they refer to as temporal hierarchies using aggregation of non-overlapping observations to construct aggregate series of different frequencies up to the annual level. They then generate forecasts for each level independently and combine these using OLS and WLS (among other methods) to generate a set of optimally reconciled forecasts across all levels of temporal aggregation. They show that forecasting with temporal hierarchies increases forecast accuracy particularly under increased model uncertainty. Sayal et al. (2016) introduce a new statistical procedure, namely wavelet benchmarking, for reconciling inconsistencies across different frequencies in data sets such as those often observed from national accounts. They show that their method substantially outperforms regularly used approaches.

The rest of the paper is structured as follows. Section 2 introduces the necessary notation, revises the GLS reconciliation of Hyndman, Ahmed, et al. (2011) and introduces the MinT approach. Section 3 presents a variety of Monte-Carlo experiments followed by a detailed empirical example for forecasting tourism flows in Australia in Section 4. Section 5 concludes with a discussion of possible future research.

2 Forecast reconciliation for hierarchical and grouped time series

2.1 Notation

Following the notation in Hyndman, Lee, and Wang (2016), we let \mathbf{y}_t be an m -vector containing all observations at time t , and \mathbf{b}_t be an n -vector containing the observations at the most

disaggregated level only. This leads to the convenient general matrix representation

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t, \quad (1)$$

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

Insert Figure 1 about here.

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, $n = 5$, $m = 8$, $\mathbf{y}_t = [y_t, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{AC,t}, y_{BA,t}, y_{BB,t}]'$, $\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{AC,t}, y_{BA,t}, y_{BB,t}]'$ and the summing matrix is given by

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

where \mathbf{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 \mathbf{S} .

Let $\hat{\mathbf{y}}_T(h)$ be a vector of h -step-ahead base forecasts for each time series in the collection, made using data to time T , and stacked in the same order as \mathbf{y}_t . (Any method may have been used to produce these forecasts.) Then all linear reconciliation methods can be written as

$$\tilde{\mathbf{y}}_T(h) = \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_T(h), \quad (2)$$

for some appropriately selected matrix \mathbf{P} of order $n \times m$, where $\tilde{\mathbf{y}}_T(h)$ is a set of reconciled forecasts which are now coherent by construction. 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 \mathbf{P} is to project the base forecasts into bottom level disaggregated forecasts which are then summed by \mathbf{S} .

For example, setting $\mathbf{P} = [\mathbf{0}_{n \times (m-n)} \mid \mathbf{I}_n]$ where $\mathbf{0}_{i \times j}$ is the $i \times j$ null matrix, bottom-level base forecasts are extracted from $\hat{\mathbf{y}}_T(h)$ and then summed by \mathbf{S} to return bottom-up forecasts. On the other hand, setting $\mathbf{P} = [\mathbf{p} \mid \mathbf{0}_{n \times (m-1)}]$, the set of proportions $\mathbf{p} = [p_1, p_2, \dots, p_n]$ distribute the top-level base forecasts to the bottom level which are then summed by \mathbf{S} to return top-down forecasts. For further discussion of \mathbf{P} matrices for various hierarchical forecasting methods see Athanasopoulos, Ahmed, and Hyndman (2009).

Assume that the base forecasts are unbiased, i.e., $E[\hat{\mathbf{y}}_T(h) | \mathcal{I}_T] = E[\mathbf{y}_{T+h} | \mathcal{I}_T]$, where $\mathcal{I}_T = \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$ denote data observed up to time T . Let $\hat{\mathbf{b}}_T(h)$ be the bottom level base forecasts with $E[\hat{\mathbf{b}}_T(h) | \mathcal{I}_T] = \boldsymbol{\beta}_T(h)$, hence $E[\hat{\mathbf{y}}_T(h) | \mathcal{I}_T] = \mathbf{S}\boldsymbol{\beta}_T(h)$. Then a set of reconciled forecasts will also be unbiased iff $\mathbf{S}\mathbf{P}\mathbf{S} = \mathbf{S}$ or equivalently $\mathbf{P}\mathbf{S} = \mathbf{I}_n$, as then from Eq. (2) we have $E[\tilde{\mathbf{y}}_T(h) | \mathcal{I}_T] = \mathbf{S}\boldsymbol{\beta}_T(h)$.

2.2 GLS reconciliation

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

$$\hat{\mathbf{y}}_T(h) = \mathbf{S}\boldsymbol{\beta}_T(h) + \boldsymbol{\varepsilon}_h, \quad (3)$$

where $\boldsymbol{\beta}_T(h) = E[\mathbf{b}_{T+h} | \mathcal{I}_T]$ is the unknown mean of the most disaggregate series at the bottom level and $\boldsymbol{\varepsilon}_h$ is the reconciliation error which is independent of observations $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$ with mean zero and covariance matrix $\text{Var}(\boldsymbol{\varepsilon}_h) = \boldsymbol{\Sigma}_h$. If $\boldsymbol{\Sigma}_h$ 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

$$\tilde{\mathbf{y}}_T(h) = \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}_h^+\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}_h^+\hat{\mathbf{y}}_T(h), \quad (4)$$

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

In general $\boldsymbol{\Sigma}_h$ is not known and is not identifiable. To see this, note that the residuals from the regression model in Eq. (3) are given by

$$\boldsymbol{\varepsilon}_h = \hat{\mathbf{y}}_T(h) - \tilde{\mathbf{y}}_T(h) = (\mathbf{I}_m - \mathbf{S}\mathbf{P})\hat{\mathbf{y}}_T(h), \quad (5)$$

so that

$$\text{Var}(\tilde{\boldsymbol{\varepsilon}}_h | \mathcal{I}_t) = (\mathbf{I}_m - \mathbf{SP})\boldsymbol{\Sigma}_h(\mathbf{I}_m - \mathbf{SP})'. \quad (6)$$

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

$$\text{Rank}(\mathbf{I}_m - \mathbf{SP}) = \text{tr}(\mathbf{I}_m - \mathbf{SP}) = \text{tr}(\mathbf{I}_m) - \text{tr}(\mathbf{I}_n) = m - n,$$

and $(\mathbf{I}_m - \mathbf{SP})$ is rank deficient, and consequently $\boldsymbol{\Sigma}_h$ cannot be identified.

Hyndman, Ahmed, et al. (2011) avoided estimating $\boldsymbol{\Sigma}_h$ by using OLS, replacing $\boldsymbol{\Sigma}_h$ by $k_h \mathbf{I}_m$ where $k_h > 0$ is a constant.

2.3 Minimum trace (MinT) reconciliation

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

$$\hat{\boldsymbol{\varepsilon}}_t(h) = \mathbf{y}_{t+h} - \hat{\boldsymbol{y}}_t(h) \quad (7)$$

$$\tilde{\boldsymbol{\varepsilon}}_t(h) = \mathbf{y}_{t+h} - \tilde{\boldsymbol{y}}_t(h) \quad (8)$$

for $t = 1, 2, \dots$, 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 \mathbf{y}_{t+h} are the observed values of all series at time $t + h$. It should be noted that the reconciled forecast errors, $\tilde{\boldsymbol{\varepsilon}}_t$ given by Eq. (8), and the estimated reconciliation errors, $\hat{\boldsymbol{\varepsilon}}_t$ given by Eq. (5), are conceptually different. The former is the forecast error of reconciled forecasts while the latter is the error due to the incoherency of the base forecasts. Hyndman, Ahmed, et al. (2011) confused these two errors.

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

$$\text{Var}[\mathbf{y}_{t+h} - \tilde{\boldsymbol{y}}_t(h) | \mathcal{I}_t] = \mathbf{SPW}_h \mathbf{P}' \mathbf{S}'$$

where $\tilde{\boldsymbol{y}}_t(h)$ is given by Eq. (2) and $\mathbf{W}_h = \text{E}[\hat{\boldsymbol{\varepsilon}}_t(h)\hat{\boldsymbol{\varepsilon}}_t'(h) | \mathcal{I}_t]$ is the variance-covariance matrix of the h -step-ahead base forecast errors.

Proof. See Appendix A.1. □

An important feature of Lemma 1 is that it can be used to construct prediction intervals for any hierarchical or grouped forecasting approaches that generates unbiased reconciled forecasts, conditional on the base forecasts being unbiased.

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

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

$$\mathbf{P} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}, \quad (9)$$

or equivalently by

$$\mathbf{P} = \mathbf{J} - \mathbf{J}\mathbf{W}_h\mathbf{U}(\mathbf{U}'\mathbf{W}_h\mathbf{U})^{-1}\mathbf{U}', \quad (10)$$

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

Proof. See Appendix A.2. □

Computing \mathbf{P} using Eq. (10) requires the inversion of an $m^* \times m^*$ matrix (where $m^* < n$ in most applications), 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 significantly less demanding in terms of computation.

The reconciled forecasts from the MinT approach are computed by

$$\hat{\mathbf{y}}_T(h) = \mathbf{S}(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}\hat{\mathbf{y}}_T(h) = \mathbf{S}[\mathbf{J} - \mathbf{J}\mathbf{W}_h\mathbf{U}(\mathbf{U}'\mathbf{W}_h\mathbf{U})^{-1}\mathbf{U}']\hat{\mathbf{y}}_T(h). \quad (11)$$

Dagum and Cholette (2006) also describe this alternative solution, although they derive it under the assumption that $\text{Cov}(\hat{\mathbf{e}}_{1,h}, \hat{\mathbf{e}}_{2,h}) = \mathbf{0}$ where $\hat{\mathbf{e}}_{2,h}$ is the vector of base forecast errors from the most disaggregated series, and $\hat{\mathbf{e}}_{1,h}$ 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, Ahmed, et al. (2011) confused the forecast errors with the reconciliation errors, they expressed the GLS solution in Eq. (4) using the matrix \mathbf{W}_h rather than $\mathbf{\Sigma}_h$, thus giving

Eq. (9). 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 generalized inverse would then reduce the GLS solution to an OLS solution.

Theorem 1 provides further insight into this result. Let $\hat{\mathbf{a}}_t(h)$ be the bottom-level base forecast errors and let $\mathbf{V}_h = \text{Var}[\hat{\mathbf{a}}_t(h)|\mathcal{I}_t]$. Then the error-additivity assumption is that $\hat{\mathbf{e}}_t(h) = \mathbf{S}\hat{\mathbf{a}}_t(h)$, and so $\mathbf{W}_h = \mathbf{S}\mathbf{V}_h\mathbf{S}'$ is singular. The variance covariance matrix of the reconciled forecast errors is then given by $\text{Var}[\mathbf{y}_{n+h} - \hat{\mathbf{y}}_n(h)|\mathcal{I}_t] = \mathbf{S}\mathbf{P}\mathbf{S}\mathbf{V}_h(\mathbf{S}\mathbf{P}\mathbf{S})' = \mathbf{S}\mathbf{V}_h\mathbf{S}'$ which is independent of \mathbf{P} . Consequently, under the error-additivity assumption, *any* matrix that satisfies $\mathbf{S}\mathbf{P}\mathbf{S} = \mathbf{S}$ will be a minimum variance MinT solution. Applying the Moore-Penrose inverse of \mathbf{W}_h when computing Eq. (9) will lead to the OLS solution $\mathbf{P} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$ (using Fact 6.4.8 of Bernstein, 2005). On the other hand, applying the Moore-Penrose inverse when computing Eq. (10) yields

$$\mathbf{P} = \mathbf{J} - \mathbf{J}\mathbf{S}\mathbf{V}_h\mathbf{S}'\mathbf{U}(\mathbf{U}'\mathbf{S}\mathbf{V}_h\mathbf{S}'\mathbf{U})^{-1}\mathbf{U}' = \mathbf{J},$$

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

2.4 Alternative estimators for \mathbf{W}_h

Note that the MinT and GLS solutions differ only in the covariance matrix that enters the estimators. In the GLS estimator in Eq. (4), $\boldsymbol{\Sigma}_h$ is the covariance matrix of the reconciliation errors, which we have shown cannot be identified, whereas in the MinT estimator in Eq. (11), \mathbf{W}_h is the covariance matrix of the base forecast errors. Although \mathbf{W}_h does not suffer from a lack of identification, it is nevertheless challenging to estimate, especially for $h > 1$. This section discusses several alternatives, which the next section then evaluates using a series of Monte Carlo experiments and empirical application.

1. Set $\mathbf{W}_h = k_h\mathbf{I}$, $\forall h$, where $k_h > 0$. This is the most simplifying assumption to make, and collapses the MinT estimator to the OLS estimator of Hyndman, Ahmed, et al. (2011). This is optimal only under some special conditions, such as when the base forecast errors are uncorrelated and equivariant. However, these conditions are impossible to satisfy in applications of hierarchical and grouped time series.
2. Set $\mathbf{W}_h = k_h\text{diag}(\hat{\mathbf{W}}_1)$, $\forall h$, where $k_h > 0$ and

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

is the unbiased sample estimator of the in-sample one-step-ahead base forecast errors as defined in Eq. (7). In this case MinT can be described as a WLS estimator. A similar estimator is used in Hyndman, Lee, and Wang (2016), but they provide no theoretical justification. Athanasopoulos, Hyndman, et al. (2017) also implemented this estimator when applying temporal hierarchies. In what follows, we denote this as WLS_v (i.e., WLS applying variance scaling).

3. Set $W_h = k_h \Lambda$, $\forall h$, where $k_h > 0$ and $\Lambda = \text{diag}(\mathbf{S1})$ with $\mathbf{1}$ being a unit column vector of dimension n . This specification is proposed by Athanasopoulos, Hyndman, et al. (2017) for temporal hierarchies and assumes that each of the bottom-level base forecast errors has a variance k_h and are uncorrelated between nodes. Hence, each element of the diagonal Λ matrix contains the number of forecast error variances contributing to that aggregation level. This estimator depends only on the grouping structure of the collection, and therefore we refer to it as an estimator that applies structural scaling and denote it by WLS_s . Its advantage over OLS is that it assumes equivariant forecast errors only within levels of the structure 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 $W_h = k_h \hat{W}_1$, $\forall h$, 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 when m is of the same order as T . In the results that follow, we denote this as MinT(Sample).
5. Set $W_h = k_h \hat{W}_{1,D}^*$, $\forall h$, where $k_h > 0$, $\hat{W}_{1,D}^* = \lambda_D \hat{W}_{1,D} + (1 - \lambda_D) \hat{W}_1$ is a shrinkage estimator with diagonal target, $\hat{W}_{1,D}$ is a diagonal matrix comprising the diagonal entries of \hat{W}_1 , and λ_D is the shrinkage intensity parameter. Thus, off-diagonal elements of \hat{W}_1 are shrunk towards zero and diagonal elements (variances) remain unchanged. Schäfer and Strimmer (2005) proposed a scale and location invariant shrinkage estimator by parameterizing the shrinkage in terms of variances and correlations rather than variances and covariances. Assuming that the variances are constant, they proposed the shrinkage intensity parameter

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

where \hat{r}_{ij} is the ij th element of $\hat{\mathbf{R}}_1$, the 1-step-ahead sample correlation matrix to shrink it towards an identity matrix. We denote 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 five different designs. Sections 3.1 and 3.2 explore the impact of the correlation structure between series on the reconciled forecasts. Section 3.3 aims to capture an important observed feature of hierarchical and grouped time series, namely that more aggregated series are smoother than less aggregated series. Section 3.4 considers reconciliation where the data contain a seasonal component. Finally, Section 3.5 considers a genuinely ‘large’ hierarchy.

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)

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

where

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

and $\eta_{A,t} \sim \mathcal{N}(0, \sigma_A^2)$ and $\eta_{B,t} \sim \mathcal{N}(0, \sigma_B^2)$ are independent Gaussian white noise error processes. 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

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

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 (Hyndman, 2017). 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 upper panels of 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 upper-left panels for both sets of parameters show that the stronger the negative error correlation between the bottom level series, the lower the RMSE for the 1-step-ahead top level forecasts. This demonstrates that the negative error correlation between the two bottom level series has a smoothing effect, making the aggregate series easier to forecast. As the error 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 RMSEs of 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.

Insert Figure 2 about here.

Insert Figure 3 about here.

The lower panels show the relative ratio in average RMSEs between the reconciled and the base forecasts over the horizons $h = 1, 1-4$ and $1-12$. A ratio less than one 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) at the top level is approximately 30% lower than the RMSE of the base forecasts, indicating a 30% improvement in forecast accuracy when the error correlation between the bottom-level series is -0.8 . 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 the top and bottom level series (with a few rare exceptions).

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.

Insert Figure 4 about here.

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 of 0, 1 and 2, with equal probability and d taking values of 0 and 1 with equal probability. For each series constructed, the parameters were chosen randomly from a uniform distribution over a space well within the stationary and invertible parameter space. Table 1 shows the parameter space for each component of the ARIMA generating processes.

Insert Table 1 about here.

The bottom level ARIMA data generating processes have a contemporaneous error covariance matrix given by

$$\begin{bmatrix} 5 & 3 & 2 & 1 \\ 3 & 4 & 2 & 1 \\ 2 & 2 & 5 & 3 \\ 1 & 1 & 3 & 4 \end{bmatrix} \quad (12)$$

in order to obtain a hierarchical structure with correlated series. This structure allows a strongly positive error correlation among the series with the same parent, but a moderately positive correlation among the series belonging to different parents. The errors were generated from a multivariate normal distribution with zero mean and the aforementioned covariance structure.

Simulation setup

For each series, $T = 60, 180$ or 300 observations were generated; the final $h = 8, 16$ or 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 (Hyndman, 2017) 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 unknown and misspecified. This process was repeated 1000 times.

The results are presented in the left panel of Table 2 under the heading ‘Section 3.2: effect of correlation’. Each entry in the table shows the percentage difference in the average RMSE between the reconciled and the base forecasts. A negative (positive) entry shows a percentage decrease (increase) in average RMSE relative to the unreconciled base forecasts. The bold entries identify the best performing approaches. The results show that the MinT(Shrink) approach generates the most accurate forecasts overall. MinT(Sample) also performs well, at least for $h = 1$, and its forecast accuracy approaches that of MinT(Shrink) as the sample size increases.

Insert Table 2 about here.

Of the approaches that use a strictly diagonal covariance matrix, WLS_v performs best at $h = 1$ and shows a performance similar to that of MinT(Shrink) when $h > 1$. Recall that 1-step-ahead in-sample base forecast errors are used for the WLS and the MinT approaches, and it is assumed that $W_h \propto \hat{W}_1$ for all $h > 1$. Controlling for the possibly adverse effects of this assumption, and therefore evaluating only the 1-step-ahead forecast performance between WLS_v and MinT(Shrink), it seems that specifying a non-diagonal covariance matrix and estimating it using shrinkage leads to a substantial improvement in forecast accuracy.

The bottom-up (BU) approach performs especially well for small sample sizes. This is not surprising, given that 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. We captured this feature within a simulation setting by implementing a simulation design based on the work of 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

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

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

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

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

$$\text{Var}(\tau_{AA,t} + \tau_{AB,t} + \tau_{BA,t} + \tau_{BB,t}) \leq \text{Var}(\tau_{AA,t} + \tau_{AB,t} - \omega_t) \leq \text{Var}(\tau_{AA,t} - v_t - 0.5\omega_t),$$

which simplifies to

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

thus 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 setup as in Section 3.2 for evaluating the forecasting performances of the alternative forecast reconciliation approaches.

The results are presented in the right panel of Table 2 under the heading ‘Section 3.3: effect of smoothing’. As expected, there is a tremendous loss in accuracy from using the bottom-up approach to forecast such a structure. MinT(Shrink) outperforms all of the forecast reconciliation

approaches, with WLS_v being second best. Again, we observe that $\text{MinT}(\text{Sample})$ approaches $\text{MinT}(\text{Shrink})$ in forecast accuracy as the sample size increases.

3.4 Reconciliation with seasonal data

Using the same hierarchical structure as in Section 3.2 (shown in Figure 4), we now consider a setting where the simulated series include a seasonal component. The bottom level series were generated using the basic structural time series model

$$\mathbf{b}_t = \boldsymbol{\mu}_t + \boldsymbol{\gamma}_t + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\mu}_t$, $\boldsymbol{\gamma}_t$ and $\boldsymbol{\eta}_t$ are the trend, seasonal and error components respectively,

$$\begin{aligned} \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \mathbf{v}_t + \boldsymbol{\varrho}_t, & \boldsymbol{\varrho}_t &\sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{\varrho}}^2 \mathbf{I}_4), \\ \mathbf{v}_t &= \mathbf{v}_{t-1} + \boldsymbol{\zeta}_t, & \boldsymbol{\zeta}_t &\sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{\zeta}}^2 \mathbf{I}_4), \\ \boldsymbol{\gamma}_t &= -\sum_{i=1}^{s-1} \boldsymbol{\gamma}_{t-i} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{\omega}}^2 \mathbf{I}_4), \end{aligned}$$

and $\boldsymbol{\varrho}_t$, $\boldsymbol{\zeta}_t$ and $\boldsymbol{\omega}_t$ are errors independent of each other and over time. The error variances were set to $\sigma_{\boldsymbol{\varrho}}^2 = 2$, $\sigma_{\boldsymbol{\zeta}}^2 = 0.007$ and $\sigma_{\boldsymbol{\omega}}^2 = 7$. The number of seasons in the year was set to $s = 4$ for quarterly data. The initial values for $\boldsymbol{\mu}_0$, \mathbf{v}_0 , $\boldsymbol{\gamma}_0$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$ were generated independently from a multivariate normal distribution with mean zero and covariance matrix, $\boldsymbol{\Sigma}_0 = \mathbf{I}_4$. Each component of $\boldsymbol{\eta}_t$ was generated from an $\text{ARIMA}(p, 0, q)$ process with p and q taking values of 0 and 1 with equal probability. The parameter space of the coefficients and the error covariance structure used are the same as those in Section 3.2, shown in Table 1 and Eq. (12) respectively. The bottom level series were then appropriately summed to obtain the data for higher levels.

The simulation setup implemented is identical to that outlined in Section 3.2. Base forecasts were generated from both ETS and ARIMA models using the default settings as implemented in the forecast package for R (Hyndman, 2017) and described in Hyndman and Khandakar (2008). This process was repeated 500 times.

The results are presented in Table 3. The left panel shows the results using the ARIMA base forecasts while the right panel shows the results using the ETS base forecasts. All of the reconciliation approaches improve the forecast accuracy of both sets of base forecasts. $\text{MinT}(\text{Shrink})$ consistently shows the largest improvements. $\text{MinT}(\text{Sample})$ also performs well for the larger

sample sizes, i.e., $T = 300$, for which a more accurate estimate of the covariance structure can be obtained.

Although the aim of these simulations is not to compare the forecast accuracy between the base forecasts per se, it is worth noting that the ETS base forecasts tend to be more accurate than the ARIMA base forecasts. This is particularly noticeable for the bottom-level base forecasts and also for $T = 300$. An important result here is that the severe model mis-specification occurring with the ARIMA base forecasts is compensated for by the reconciliation approaches, and especially MinT(Shrink) and MinT(Sample), leading to substantial improvements over the base forecasts. Hence, the covariance information used in the MinT approach is especially beneficial when there is severe model mis-specification. We should also note that for these cases, the bottom-up approach does not improve the forecast accuracy over the base forecasts.

3.5 Forecasting a 'large' hierarchy

In our final design we consider a much larger hierarchy. The structure comprises 5-levels and 2,047 series in total. More specifically, 1,536 series at the bottom level were aggregated in groups of four for the next four levels resulting in 384, 96, 24 and 6 series respectively for levels 4 to 1; which were then aggregated to give the top-level most aggregate series.

The challenge here is to generate hierarchical time series at this very large scale that reflect features inherent in real data. To achieve this we generalise, expand and combine the designs of Sections 3.2 and 3.3. In particular, the data generating mechanism was designed to reflect (i) correlations across series; and (ii) a higher signal-to-noise ratio for aggregate series compared to their disaggregate counterparts. The details of the data generating mechanism are presented in Appendix B.

For each series, $T = 180$ observations were generated with the final $h = 18$ withheld as the test set. Using the remaining observations as training data, base forecasts were generated from ARIMA and ETS models using the default settings as implemented in the forecast package for R (Hyndman, 2017) and described in Hyndman and Khandakar (2008). This process was repeated 200 times.

The results are presented in Table 5. We should note here that the length of each series is less than the total number of series in the structure. Hence, the sample variance covariance is not a positive definite matrix, and therefore we did not consider the MinT(Sample) approach. The reconciliation approaches overwhelmingly improve the forecast accuracy of both sets of base

forecasts. MinT(Shrink) consistently shows the largest improvements for the shorter forecast horizons $h = 1$ and 4 and WLS_v for the longer forecast horizons $h = 8$ and 16. The bottom-up approach shows substantial losses relative to the base forecasts for $h = 1$ to 4.

4 Forecasting Australian Domestic Tourism

Domestic tourism flows within any country form a natural geographical hierarchy. Australia comprises seven states and territories which are further divided into 27 zones and 76 regions (Appendix C shows further details on the geographical hierarchy). Furthermore, grouping tourism flows by purpose of travel is also of interest as travel patterns and spending behaviour vary across these groups. We consider four purposes of travel: holiday, visiting friends and relatives (VFR), business and other. Grouping Australian domestic tourism flows by geographical divisions and purposes of travel results in a total of 555 time series (both aggregated and disaggregated). Table 7 presents further details.

Insert Table 7 about here.

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). The data are monthly time series and span the period January 1998 to December 2016.

The left panel of Figure 6 plots the total number of visitor nights for 2016. The map shows the diversity in the volume of tourism flows across the geographical divisions with the highest number of visitor nights concentrating around the state capitals. Figure 7 plots visitor nights grouped by the four purposes of travel together with the aggregate. The left panel shows homogeneous seasonal variation across the aggregate, holiday and VFR series in contrast to the business and other series plotted in the right panel. The series also show diverse trends. Generating accurate forecasts for each of the 555 time series within the Australian tourism grouping structure is imperative for the planning purposes of National, State and Local tourism authorities. Ensuring that these forecasts are reconciled across all geographical disaggregation levels and purposes of travel leads to coherent decision making across the various levels of management. Moreover, in the empirical results that follow we find overwhelming evidence

that reconciling independent base forecasts using the MinT approach, substantially improves forecast accuracy across all levels of disaggregation and groupings.

One advantage of using a reconciliation approach to forecasting is that it implicitly models spatial autocorrelations in the data. With such a large collection of time series, it would be challenging to model the spatial autocorrelations directly, but through reconciliation we can implicitly account for spatial autocorrelations, especially when the Mint method is used.

Insert Figure 6 about here.

Insert Figure 7 about here.

We evaluate the forecast accuracy of MinT using a rolling window. We set the training window to 96 observations and generate 1- to 12-steps-ahead base forecasts for each of the 555 series. The base forecasts are generated from ARIMA and ETS models fitted using the default settings in the automated algorithms of Hyndman and Khandakar (2008) and implemented in the `forecast` for R (Hyndman, 2017). The base forecasts are then combined or reconciled using the alternative approaches. We roll the training window forward by one observation and repeat the process until November 2016. This gives a total of 132 1-step-ahead, 131 2-step-ahead, down to 121 12-step-ahead forecasts for each of the 555 series. We should note here that the number of observations per series is less than the total number of series in the structure. Hence, the sample variance covariance is not a positive definite matrix, and therefore the MinT(Sample) approach will not be considered.

The results are presented in Table 8. The rows labelled '*Base*' show the average RMSE ($\times 10^{-3}$) for the base forecasts. A positive (negative) entry above this row shows a percentage increase (decrease) in average RMSE relative to the base forecasts. The top half of the table shows the results using ARIMA base forecasts, while the bottom half shows the results using ETS base forecasts. It is immediately clear that for all levels of disaggregation, MinT(Shrink) forecasts improve (and in many cases substantially improve) on the accuracy of the base forecasts. The improvements are more pronounced when the base forecasts are less accurate, which is the case with the ARIMA base forecasts compared to the ETS base forecasts.

Furthermore, the MinT(Shrink) approach returns the most accurate reconciled forecasts for all levels. The only exception being the most aggregate level using ARIMA base forecasts, for which OLS returns the most accurate forecasts. The two variations of WLS and OLS also perform well, generally improving forecast accuracy over the base forecasts.

Insert Table 8 about here.

Clearly the least accurate approach is the bottom-up showing increases in the average RMSE relative to the base forecasts for almost all levels of aggregation and all forecast horizons. These losses in forecast accuracy are substantial at the higher levels of aggregation and especially when using the least accurate ARIMA base forecasts. The biggest disadvantage of the bottom-up approach is that the series are modelled and forecasts are generated only at the bottom level. The series at the bottom level are highly disaggregated and therefore very noisy, making it challenging to identify any seasonal or possibly trending components. Table 9 shows the composition of ARIMA and ETS models identified at each level over the final estimation window. Fewer than 50% of the models identified at the bottom level, for both ARIMA and ETS models, include a seasonal component, and only 5% of them include both a seasonal and a trending component. These components become more prominent at the higher levels of aggregation and therefore they are more readily identified by the automated forecasting algorithms. These empirical results show that when forecasting such hierarchical or grouped structures, it is imperative that the approach used can bring informative signals from the higher levels of aggregation to the lower levels and vice versa. MinT, the two variations of WLS, and OLS are all able to do this, and hence improvements in forecast accuracy over the base forecasts are attained using these approaches even at the extremely noisy bottom level.

Finally, the right panel of Figure 6 plots the projected annual growth rates for visitor nights for 2017. These are generated from MinT(Shrink) using the most accurate ETS base forecasts. The map shows the diversity in the projected growth rates across the geographical regions of Australia. Combining these with the left panel brings a plethora of invaluable information to the tourism authorities. These help identify key areas of growth and/or decline of Australian domestic tourism.

5 Conclusions and discussion

In the existing literature on forecasting hierarchical and grouped time series, the GLS estimator introduced by Hyndman, Ahmed, et al. (2011) is the only previous approach which attempts to utilize 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 forecast reconciliation 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 derived 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 a large collection 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 using shrinkage clearly outperforms the 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 a better covariance estimator than those considered here.

A Proofs

A.1 Lemma 1

Proof. Let the h -step-ahead base and reconciled forecast errors be given as

$$\begin{aligned}\hat{\boldsymbol{e}}_t(h) &= \boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_t(h) \\ \text{and } \tilde{\boldsymbol{e}}_t(h) &= \boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_t(h),\end{aligned}$$

for $t = 1, 2, \dots$, 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 $t + h$.

We have that

$$\begin{aligned}\tilde{\boldsymbol{e}}_t(h) &= \hat{\boldsymbol{e}}_t(h) + (\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P}) \hat{\boldsymbol{y}}_t(h) \\ &= \hat{\boldsymbol{e}}_t(h) + (\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P}) [\boldsymbol{y}_{t+h} - \hat{\boldsymbol{e}}_t(h)] \\ &= \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{e}}_t(h) + (\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P})\boldsymbol{y}_{t+h}.\end{aligned}$$

Since $\boldsymbol{y}_{t+h} = \boldsymbol{S}\boldsymbol{b}_{t+h}$, where \boldsymbol{b}_{t+h} is the vector of observations at the bottom level, $(\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P})\boldsymbol{S} = \mathbf{0}$. Hence $\tilde{\boldsymbol{e}}_t(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{e}}_t(h)$ giving

$$\text{Var}[\tilde{\boldsymbol{e}}_t(h)] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}',$$

where \boldsymbol{W}_h is the variance covariance matrix of the h -step-ahead base forecast errors. □

A.2 Theorem 1

Proof. Reformulate the objective function such that,

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

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

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

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

$$\lambda_{\min}(C'C + I)\text{tr}[PW_hP'] \geq [\lambda_{\min}(C'C) + \lambda_{\min}(I)]\text{tr}[PW_hP'] \geq \text{tr}[PW_hP'],$$

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

Now $PS = I_n$, so the minimization problem can be restated as

$$\min_P \text{tr}[PW_hP'] \quad \text{such that } PS = I.$$

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

$$\min_H \text{tr}[HH'] \quad \text{such that } HL = I.$$

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

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

As the above unique solution of MinT has a similar representation to a GLS estimator of a least squares problem, we can reformulate the trace minimization problem in terms of a linear equality constrained least squares problem as follows:

$$\begin{aligned} \min_{\check{y}_T(h)} \frac{1}{2} [\hat{y}_T(h) - \check{y}_T(h)]' W_h^{-1} [\hat{y}_T(h) - \check{y}_T(h)] \\ \text{s.t. } \check{y}_T(h) = S\check{b}_T(h), \end{aligned}$$

or equivalently,

$$\begin{aligned} \min_{\check{y}_T(h)} \frac{1}{2} [\hat{y}_T(h) - \check{y}_T(h)]' W_h^{-1} [\hat{y}_T(h) - \check{y}_T(h)] \\ \text{s.t. } \mathbf{U}'\check{y}_T(h) = \mathbf{0}, \end{aligned}$$

where $\check{b}_T(h)$ is the vector of last m_K elements of $\check{y}_T(h)$, $\mathbf{U}' = \begin{bmatrix} I_{m^*} \\ -C \end{bmatrix}$ and $m^* = m - n$.

The Lagrangian function for the minimization problem is

$$\mathcal{L}[\tilde{\mathbf{y}}_T(h), \boldsymbol{\lambda}] = \frac{1}{2}[\hat{\mathbf{y}}_T(h) - \tilde{\mathbf{y}}_T(h)]' \mathbf{W}_h^{-1} [\hat{\mathbf{y}}_T(h) - \tilde{\mathbf{y}}_T(h)] - \boldsymbol{\lambda}' \mathbf{U}' \tilde{\mathbf{y}}_T(h),$$

where $\boldsymbol{\lambda}$ is a Lagrange multiplier vector.

The first-order necessary conditions for $\tilde{\mathbf{y}}_T(h)$ to be a solution of the minimization problem require that there is a vector $\tilde{\boldsymbol{\lambda}}$ for which the following system of equations is satisfied:

$$\begin{bmatrix} \mathbf{W}_h^{-1} & \mathbf{U} \\ \mathbf{U}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{y}}_T(h) \\ -\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{W}_h^{-1} \hat{\mathbf{y}}_T(h) \\ \mathbf{0} \end{bmatrix}.$$

Assuming that \mathbf{W}_h^{-1} is positive definite yields the solution of the above system of equations as

$$\begin{aligned} \tilde{\mathbf{y}}_T(h) &= [\mathbf{I} - \mathbf{W}_h \mathbf{U} (\mathbf{U}' \mathbf{W}_h \mathbf{U})^{-1} \mathbf{U}'] \hat{\mathbf{y}}_T(h) \\ \text{and } \tilde{\boldsymbol{\lambda}} &= -(\mathbf{U}' \mathbf{W}_h \mathbf{U})^{-1} \mathbf{U}' \hat{\mathbf{y}}_T(h). \end{aligned}$$

The h -step-ahead reconciled forecasts at the bottom level are given by

$$\tilde{\mathbf{b}}_T(h) = \left[[\mathbf{0}_{n \times m^*} \mid \mathbf{I}_n] - [\mathbf{0}_{n \times m^*} \mid \mathbf{I}_n] \mathbf{W}_h \mathbf{U} (\mathbf{U}' \mathbf{W}_h \mathbf{U})^{-1} \mathbf{U}' \right] \hat{\mathbf{y}}_T(h) := \mathbf{P} \hat{\mathbf{y}}_T(h).$$

Hence we can define an alternative representation for the \mathbf{P} matrix by

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

where $\mathbf{J} = [\mathbf{0}_{n \times m^*} \mid \mathbf{I}_n]$. □

B Data generating mechanism for Section 3.5

In order to present the details of the simulation setup, we consider a simplified 2-level hierarchy as shown in Figure 5. The structure has two parents, denoted by nodes A and B, and each one of them has four children denoted respectively by AA, AB, AC, AD and BA, BB, BC, BD. The bottom level series were generated by summing the two components shown in Table 4, reflecting features commonly observed in hierarchical time series.

Insert Figure 5 about here.

Insert Table 4 about here.

C Tourism hierarchy

Insert Table 6 about here.

Figures

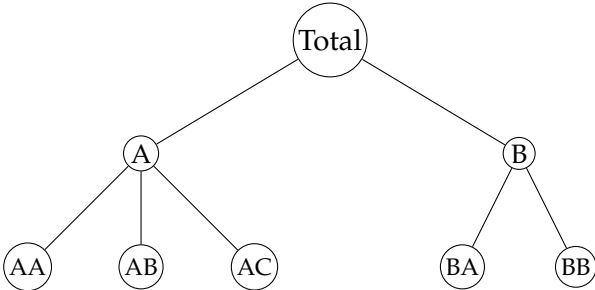


Figure 1: A two-level hierarchical tree diagram.

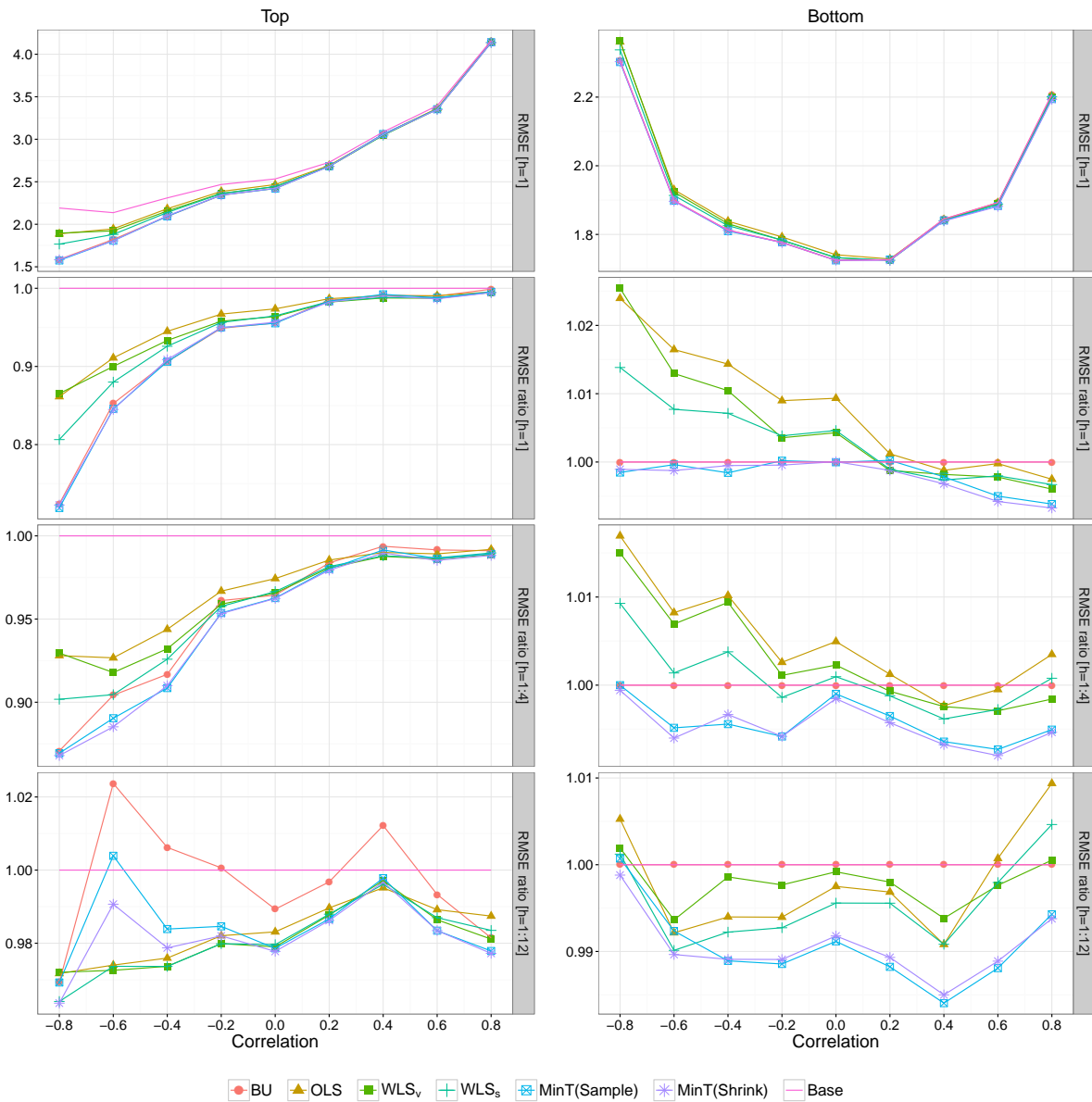


Figure 2: Forecasting performances of the base and reconciled forecasts for the top level series (shown in the left panels) and the bottom level series (shown in the right panels). The upper panels show the average RMSEs for 1-step-ahead forecasts. The lower panels show the relative ratio of average RMSEs 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 $\sigma_B^2 = 4$.

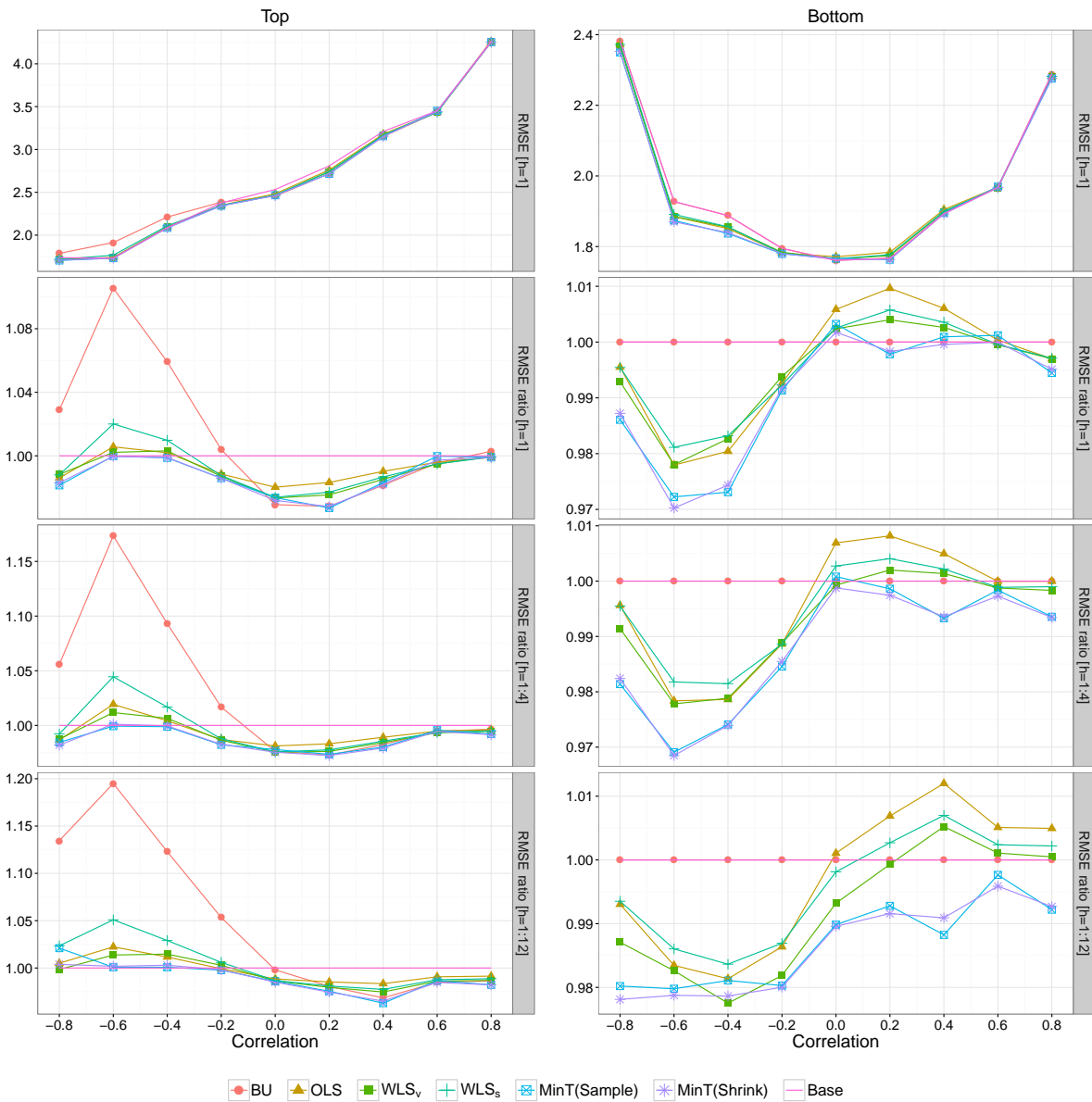


Figure 3: Forecasting performances of the base and reconciled forecasts for the top level series (shown in the left panels) and the bottom level series (shown in the right panels). The upper panels show the average RMSEs for 1-step-ahead forecasts. The lower panels show the relative ratio of average RMSEs between the reconciled and 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 $\sigma_B^2 = 4$.

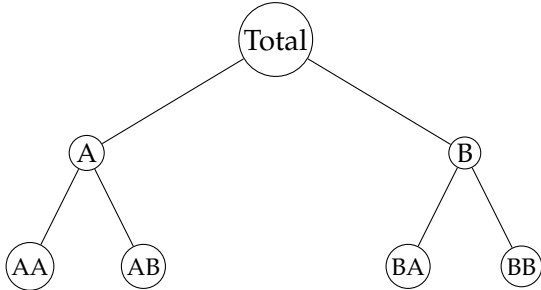


Figure 4: Hierarchical structure for simulations of Section 3.2.

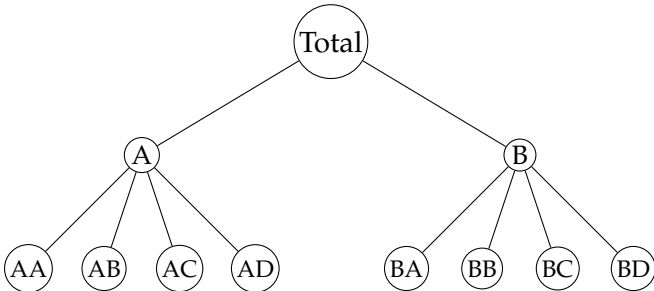


Figure 5: A simplified two-level hierarchical tree diagram.

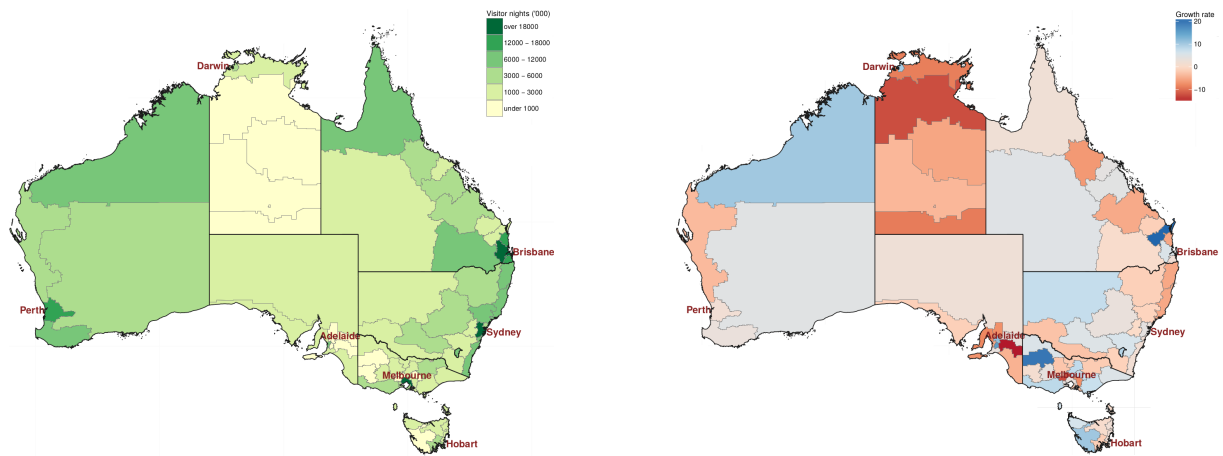


Figure 6: *Left: Total visitor nights in 2016. Right: Projected growth rates for 2017 using ETS base forecasts and reconciling using MinT(Shrink).*

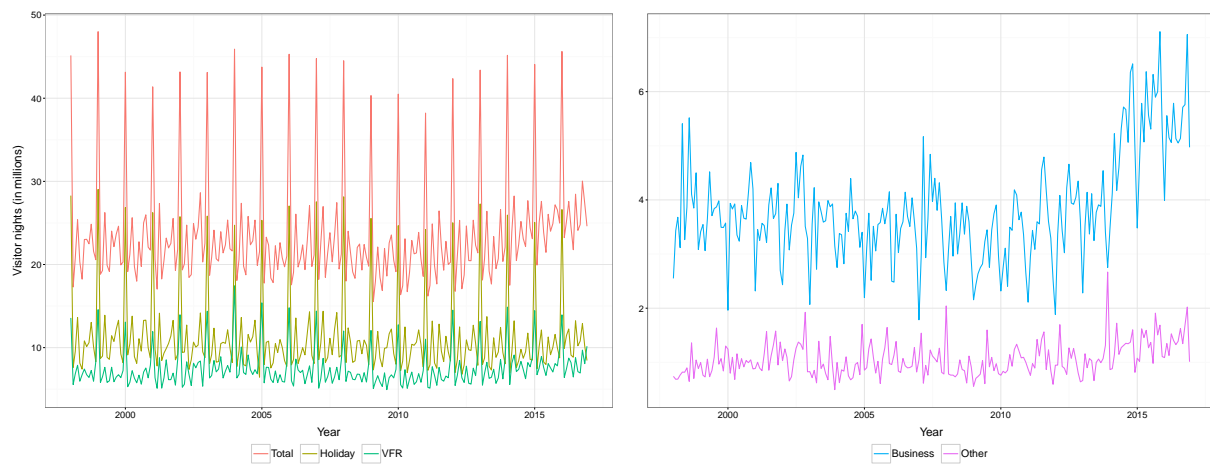


Figure 7: *Visitor nights for Australia by purpose of travel.*

Tables

Table 1: *Parameter space for the AR and MA components.*

AR component		
p	Coefficient	Parameter space
1	ϕ_1	$[0.5, 0.7]$
2	ϕ_1	$[\phi_2 - 0.9, 0.9 - \phi_2]$
	ϕ_2	$[0.5, 0.7]$
MA component		
q	Coefficient	Parameter space
1	θ_1	$[0.5, 0.7]$
2	θ_1	$[-(0.9 + \theta_2)/3.2, (0.9 + \theta_2)/3.2]$
	θ_2	$[0.5, 0.7]$

The purpose of including 3.2 in the parameter space of θ_1 is to obtain the same lower and upper bounds as in the parameter space of ϕ_1 .

Table 2: Out-of-sample forecast results for simulations exploring the effect of correlation and smoothing

	Section 3.2: effect of correlation									Section 3.3: effect of smoothing								
	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	-8.7	-5.2	-2.5	-5.5	-0.4	2.0	-5.5	-1.2	-0.6	48.9	22.6	12.8	42.8	12.2	5.7	51.4	6.2	3.7
OLS	-5.1	-4.7	-4.1	-2.9	-2.0	-1.4	-3.2	-1.7	-1.6	1.4	-1.7	-3.1	-0.5	-2.6	-2.5	1.9	-3.0	-1.9
WLS _v	-8.2	-6.5	-4.8	-5.0	-2.3	-1.0	-5.3	-2.3	-2.0	-1.0	-2.4	-3.2	-2.0	-2.5	-2.1	0.2	-2.2	-1.4
WLS _s	-7.3	-6.2	-4.9	-4.3	-2.4	-1.4	-4.6	-2.2	-2.0	9.3	1.8	-1.4	6.3	-1.0	-2.0	10.4	-2.6	-1.6
MinT(Sample)	-5.8	-3.1	-1.2	-6.3	0.5	3.4	-6.2	-1.8	-1.4	-0.6	-3.3	-3.9	-2.3	-2.2	-1.9	-0.3	-2.1	-1.3
MinT(Shrink)	-8.0	-5.9	-4.1	-6.3	-0.7	1.5	-6.5	-2.3	-1.9	-1.8	-3.2	-3.7	-2.4	-2.5	-2.1	0.0	-2.1	-1.3
	Level 1									Level 1								
BU	-4.8	-2.6	-1.4	-3.5	1.3	3.3	-2.3	0.4	0.9	8.2	4.9	3.4	6.5	2.1	0.6	6.7	0.8	0.8
OLS	-1.3	-2.2	-2.8	-1.1	-0.4	-0.3	-0.6	-0.3	-0.3	-4.6	-4.9	-5.0	-4.3	-4.2	-3.4	-4.2	-3.1	-2.1
WLS _v	-4.2	-3.8	-3.5	-3.0	-0.6	0.4	-2.2	-0.7	-0.6	-5.2	-5.3	-5.2	-4.7	-4.2	-3.3	-4.6	-2.7	-1.8
WLS _s	-3.4	-3.5	-3.6	-2.4	-0.7	-0.1	-1.8	-0.6	-0.6	-3.0	-3.9	-4.4	-3.2	-3.9	-3.5	-2.9	-3.2	-2.1
MinT(Sample)	-3.1	-0.7	-0.3	-4.7	2.1	4.7	-3.2	-0.2	0.0	-3.6	-4.3	-4.0	-5.0	-4.1	-3.0	-4.8	-3.1	-1.9
MinT(Shrink)	-4.8	-3.5	-3.1	-4.6	0.8	2.7	-3.4	-0.7	-0.6	-5.5	-5.5	-5.3	-5.1	-4.5	-3.4	-5.0	-3.1	-2.0
	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	2.7	0.3	-1.0	1.8	-1.1	-2.2	1.3	-0.4	-0.8	-2.5	-2.6	-2.7	-2.1	-2.0	-1.5	-2.2	-1.3	-1.1
WLS _v	0.8	-0.8	-1.4	0.4	-1.2	-1.8	0.1	-0.8	-1.0	-3.1	-3.2	-3.4	-2.5	-2.5	-1.9	-2.7	-1.6	-1.4
WLS _s	1.1	-0.7	-1.6	0.8	-1.3	-2.1	0.4	-0.7	-0.9	-2.2	-2.4	-2.6	-1.9	-1.9	-1.5	-1.9	-1.3	-1.1
MinT(Sample)	0.9	1.1	0.2	-1.7	0.3	0.4	-1.5	-0.5	-0.9	-3.4	-1.7	-0.2	-3.3	-1.7	-0.6	-3.9	-2.1	-1.6
MinT(Shrink)	-0.6	-1.3	-2.0	-1.6	-0.7	-1.0	-1.4	-0.9	-1.3	-4.1	-3.8	-3.4	-3.4	-2.8	-2.1	-4.0	-2.5	-1.9
	Average									Average								
BU	-4.1	-2.3	-1.1	-2.7	0.3	1.6	-2.3	-0.2	0.1	7.5	4.8	3.3	6.4	2.7	1.3	7.1	1.4	1.0
OLS	-0.9	-1.9	-2.4	-0.5	-1.1	-1.4	-0.6	-0.7	-0.8	-2.7	-3.1	-3.5	-2.5	-2.7	-2.3	-2.3	-2.1	-1.6
WLS _v	-3.5	-3.4	-3.0	-2.3	-1.3	-0.9	-2.2	-1.1	-1.1	-3.4	-3.7	-3.9	-3.0	-3.0	-2.4	-3.0	-2.0	-1.5
WLS _s	-2.8	-3.1	-3.1	-1.7	-1.4	-1.3	-1.7	-1.1	-1.1	-1.2	-2.2	-2.9	-1.4	-2.3	-2.2	-1.0	-2.1	-1.5
MinT(Sample)	-2.4	-0.7	-0.3	-4.0	1.0	2.6	-3.4	-0.7	-0.8	-3.1	-2.7	-2.0	-3.7	-2.5	-1.6	-3.8	-2.4	-1.6
MinT(Shrink)	-4.2	-3.3	-2.9	-3.9	-0.2	0.9	-3.5	-1.2	-1.2	-4.3	-4.2	-4.0	-3.8	-3.3	-2.5	-3.9	-2.6	-1.8

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

Table 3: Out-of-sample forecast results from simulating seasonal data.

	ARIMA									ETS								
	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	9.3	11.2	10.3	3.0	2.6	0.9	12.5	-8.5	-30.9	-5.3	-6.4	-7.9	1.0	-2.0	-4.5	1.2	-4.6	-8.8
OLS	-1.9	-1.2	-1.4	-1.8	-2.7	-4.3	-2.7	-22.2	-33.3	-4.5	-5.0	-5.7	-0.8	-2.3	-4.3	-1.2	-4.5	-7.7
WLS _v	-2.0	-1.4	-1.7	-1.5	-3.2	-5.4	-3.5	-37.7	-62.0	-6.4	-6.9	-8.4	-0.6	-3.1	-6.0	-1.0	-6.1	-10.8
WLS _s	-1.2	-0.1	-0.4	-1.6	-2.8	-4.9	-1.4	-28.3	-46.1	-6.1	-6.7	-7.8	-0.7	-2.9	-5.7	-1.2	-5.9	-10.2
MinT(Sample)	-0.4	0.6	0.5	-1.2	-3.0	-4.4	-3.3	-37.5	-62.5	-3.6	-3.7	-5.6	-1.1	-2.4	-4.5	-1.1	-5.1	-8.7
MinT(Shrink)	-2.4	-1.4	-1.7	-1.8	-3.4	-5.2	-3.7	-37.9	-62.8	-4.9	-5.3	-7.2	-1.1	-2.7	-5.1	-1.1	-5.6	-9.6
Base	17.0	17.3	20.0	17.0	20.4	25.0	15.7	33.2	74.7	17.9	18.4	21.8	17.1	20.6	25.6	15.4	22.2	30.5
	Level 1									Level 1								
BU	8.3	10.2	8.6	1.1	2.1	1.6	13.1	24.4	16.8	-2.9	-5.0	-5.4	-0.3	-1.3	-2.4	0.6	-1.4	-2.5
OLS	-3.4	-3.4	-3.7	-2.0	-2.4	-2.7	-1.4	-0.9	0.2	-2.1	-3.4	-3.4	-1.1	-1.3	-2.0	-0.5	-1.0	-1.5
WLS _v	-5.0	-5.6	-5.9	-2.7	-3.7	-4.5	-3.9	-19.9	-38.6	-4.0	-5.5	-6.0	-1.2	-2.2	-3.7	-0.7	-2.6	-4.3
WLS _s	-2.8	-2.2	-2.7	-2.1	-2.6	-3.4	-0.3	-4.3	-11.7	-3.7	-5.2	-5.5	-1.3	-2.0	-3.4	-0.7	-2.3	-3.7
MinT(Sample)	-4.2	-4.3	-4.6	-2.3	-3.7	-4.0	-4.4	-21.0	-41.8	-1.8	-3.0	-4.1	-1.9	-2.1	-3.4	-1.1	-2.2	-3.5
MinT(Shrink)	-5.6	-5.9	-6.2	-2.9	-4.1	-4.8	-4.7	-21.6	-42.3	-3.2	-4.5	-5.5	-1.8	-2.3	-3.7	-1.0	-2.7	-4.3
Base	10.6	10.7	12.6	10.3	12.3	15.5	9.6	16.2	30.6	10.6	10.8	12.8	10.2	12.2	15.7	9.2	13.2	18.4
	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	-6.2	-7.1	-6.4	-2.0	-2.7	-2.4	-6.9	-8.6	-5.5	0.5	1.0	1.3	-0.6	0.0	0.2	-0.7	0.2	0.6
WLS _v	-14.1	-16.2	-14.6	-3.6	-5.2	-5.3	-15.6	-35.1	-45.3	-0.5	-0.2	-0.4	-0.6	-0.6	-0.8	-0.8	-0.7	-1.0
WLS _s	-5.9	-6.5	-5.9	-2.1	-2.9	-2.8	-6.4	-10.2	-10.5	-0.5	-0.2	-0.1	-0.7	-0.5	-0.6	-0.8	-0.5	-0.6
MinT(Sample)	-13.7	-15.7	-14.4	-3.8	-5.9	-5.8	-16.3	-38.3	-51.8	0.9	0.8	0.2	-1.0	-0.8	-1.3	-1.0	-0.9	-1.4
MinT(Shrink)	-14.9	-16.9	-15.5	-4.2	-6.2	-6.4	-16.6	-38.6	-52.1	-0.2	-0.2	-0.7	-1.1	-1.0	-1.5	-1.1	-1.2	-1.9
Base	7.4	7.6	8.9	6.4	7.9	10.2	7.0	13.6	24.6	6.5	6.5	7.9	6.2	7.5	9.9	5.8	8.6	12.2
	Average									Average								
BU	4.9	6.0	5.2	1.2	1.4	0.7	7.1	4.2	-5.5	-2.4	-3.4	-3.9	0.2	-1.0	-2.0	0.5	-1.7	-3.1
OLS	-4.2	-4.5	-4.3	-1.9	-2.6	-3.0	-4.2	-10.3	-12.9	-1.7	-2.1	-2.2	-0.8	-1.0	-1.7	-0.8	-1.4	-2.3
WLS _v	-8.2	-9.2	-8.7	-2.7	-4.2	-5.1	-9.0	-31.7	-48.9	-3.3	-3.8	-4.4	-0.8	-1.8	-3.1	-0.8	-2.8	-4.6
WLS _s	-3.8	-3.6	-3.6	-2.0	-2.8	-3.5	-3.3	-13.6	-22.1	-3.1	-3.6	-4.0	-0.9	-1.7	-2.8	-0.9	-2.5	-4.1
MinT(Sample)	-7.4	-8.1	-7.7	-2.6	-4.4	-4.9	-9.4	-33.4	-52.6	-1.2	-1.7	-2.8	-1.3	-1.7	-2.8	-1.0	-2.5	-4.0
MinT(Shrink)	-8.9	-9.6	-9.2	-3.1	-4.8	-5.6	-9.7	-33.8	-53.0	-2.5	-3.1	-4.0	-1.4	-1.9	-3.2	-1.1	-2.9	-4.7
Base	9.7	9.9	11.5	9.0	10.9	13.9	9.0	17.2	33.4	9.3	9.4	11.3	8.9	10.7	13.8	8.2	11.8	16.6

Base shows the average RMSE of the base forecasts. A negative (positive) entry above this row shows the percentage decrease (increase) in average RMSE of the reconciled forecasts relative to the base forecasts. Bold entries identify the best performing approaches.

Table 4: Data generating mechanism for simulating a large hierarchy.

Component	DGP
Correlation	<p>This component ensures that children belonging to the same parent are correlated, while children belonging to different parents are independent. For the children series this component was generated from $ARIMA(p, d, q)$ processes with variances drawn from a uniform distribution on the interval $(0.05, 0.1)$. A compound symmetric block diagonal correlation structure with blocks of size 4 was used to generate contemporaneous error correlations. The correlations for each block were drawn from a uniform distribution on the interval $(0.3, 0.8)$ allowing for moderate to high correlations among the children.</p>
Signal-to-noise ratio	<p>In order to ensure that aggregate series have a higher signal-to-noise ratio compared to their disaggregate counterparts, this component was generated from processes with off-setting errors. Specifically</p> $ \begin{aligned} x_{AA,t} &= z_t + 0.25\omega_t + \eta_t, & x_{BA,t} &= z_t - 0.25\omega_t + \nu_t, \\ x_{AB,t} &= z_t + 0.25\omega_t - \eta_t, & x_{BB,t} &= z_t - 0.25\omega_t - \nu_t, \\ x_{AC,t} &= +0.25\omega_t, & x_{BC,t} &= -0.25\omega_t, \\ x_{AD,t} &= +0.25\omega_t, & x_{BD,t} &= -0.25\omega_t, \end{aligned} $ <p>where z_t is an $ARIMA(p, 1, q)$ process with normally distributed errors with mean zero and variance 0.005. ω_t, η_t and ν_t were independently generated from $ARIMA(p, 0, q)$ processes with normally distributed errors with mean zero and variance 0.4.</p>

For each $ARIMA(p, d, q)$ process, $p, q \in \{0, 1, 2\}$ and $d \in \{0, 1\}$ with equal probability. The parameter spaces for the coefficients are given in Table 1.

Note: For hierarchies involving more than 2-levels, the weights of exactly off-setting errors from each level to the bottom level nodes (0.25 in this case) need to be calculated appropriately. The weights for a given node are simply the reciprocal of the number of descendant children at the bottom level.

Table 5: Out-of-sample forecast performances from 5-level hierarchy.

	ARIMA								ETS							
	<i>h</i> = 1	1-4	1-8	1-16	<i>h</i> = 1	1-4	1-8	1-16	<i>h</i> = 1	1-4	1-8	1-16	<i>h</i> = 1	1-4	1-8	1-16
	Top level				Level 1				Top level				Level 1			
BU	63.3	21.0	-1.9	-18.9	36.0	13.9	-2.0	-16.3	57.3	18.9	-1.8	-16.0	34.5	13.6	-1.7	-11.8
OLS	-0.7	-1.6	-2.0	-2.2	-2.8	-0.1	1.4	2.1	-0.5	-1.6	-2.2	-3.3	-1.7	0.2	1.1	3.5
WLS _v	14.2	-4.7	-14.1	-20.8	1.7	-6.8	-12.4	-18.0	13.6	-3.5	-12.3	-19.8	3.5	-4.7	-10.7	-15.2
WLS _s	7.9	-6.9	-13.8	-18.8	-2.1	-8.1	-11.9	-16.0	8.1	-5.5	-12.0	-18.0	0.1	-6.1	-10.3	-13.4
MinT(Shrink)	3.1	-7.1	-11.6	-14.5	-5.3	-8.5	-10.4	-12.4	1.8	-6.5	-10.0	-14.9	-4.2	-6.9	-8.8	-10.8
Base	64.2	142.5	268.1	543.5	14.2	27.7	48.9	94.8	69.9	145.9	263.3	506.1	15.0	27.9	47.9	86.6
	Level 2				Level 3				Level 2				Level 3			
BU	16.5	6.6	-2.2	-10.9	4.6	1.2	-2.2	-5.6	15.0	6.0	-1.8	-6.8	4.5	1.4	-1.2	-2.3
OLS	-2.7	0.2	2.6	5.5	-2.7	-0.9	1.1	4.3	-2.5	-0.2	2.3	6.9	-2.1	-0.8	1.4	5.2
WLS _v	-4.0	-6.8	-9.5	-12.4	-5.3	-5.6	-6.2	-6.8	-2.7	-5.6	-8.0	-9.3	-3.8	-4.5	-4.6	-3.9
WLS _s	-5.7	-7.5	-9.2	-10.9	-5.8	-5.9	-6.1	-6.2	-4.3	-6.4	-7.8	-8.0	-4.3	-4.8	-4.5	-3.4
MinT(Shrink)	-7.2	-7.7	-8.2	-8.3	-6.3	-6.0	-5.7	-4.9	-6.6	-6.9	-6.8	-6.2	-5.3	-5.1	-4.1	-2.5
Base	5.3	9.4	15.3	27.4	2.5	4.0	6.1	9.9	5.6	9.4	15.0	25.1	2.5	4.0	5.9	9.2
	Level 4				Bottom level				Level 4				Bottom level			
BU	0.8	-0.1	-1.3	-2.8	0.0	0.0	0.0	0.0	1.1	0.6	-0.1	-0.5	0.0	0.0	0.0	0.0
OLS	-1.5	-0.9	-0.1	1.2	0.7	0.7	1.0	1.8	-1.1	-0.6	0.3	1.7	0.5	0.5	0.7	1.1
WLS _v	-2.7	-2.9	-3.2	-3.6	-1.6	-1.3	-1.0	-0.3	-2.0	-2.1	-1.9	-1.6	-1.2	-1.1	-0.9	-0.6
WLS _s	-2.8	-3.0	-3.1	-3.4	-0.3	-0.3	-0.3	0.0	-2.0	-2.1	-1.9	-1.5	-0.2	-0.3	-0.2	-0.2
MinT(Shrink)	-2.9	-3.0	-3.0	-2.9	-1.6	-1.3	-0.7	0.4	-2.3	-2.2	-1.7	-1.0	-1.2	-1.0	-0.5	0.1
Base	1.4	2.0	2.9	4.5	0.6	0.9	1.2	1.8	1.4	2.0	2.8	4.2	0.7	0.9	1.2	1.8
	Average				Average				Average				Average			
BU	5.4	2.3	-1.0	-5.1					5.1	2.4	-0.5	-3.1				
OLS	-0.7	-0.1	0.7	2.1					-0.5	-0.1	0.7	2.2				
WLS _v	-1.8	-3.1	-4.4	-6.0					-1.1	-2.4	-3.5	-4.6				
WLS _s	-1.8	-3.0	-4.0	-5.3					-1.1	-2.3	-3.2	-3.9				
MinT(Shrink)	-2.9	-3.4	-3.8	-4.0					-2.3	-2.8	-2.8	-3.0				
Base	1.0	1.5	2.2	3.5					1.0	1.5	2.2	3.3				

Base shows the average RMSE of the base forecasts. A negative (positive) entry above this row shows the percentage decrease (increase) in average RMSE of the reconciled forecasts relative to the base forecasts. Bold entries identify the best performing approaches.

Table 6: Geographical divisions of Australia.

Series	Name	Label	Series	Name	Label
<i>Total</i>			<i>Regions continued</i>		
1	Australia	Total	55	Lakes	BCA
<i>States</i>			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 East	BDD
7	TAS	F	62	Upper Yarra	BDE
8	NT	G	63	Murray East	BDF
<i>Zones</i>			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	BEG
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 WA	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
<i>Regions</i>			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

Table 7: *Grouped time series for Australian tourism flows.*

Geographical division	Number of series per geographical division	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 C](#).

Table 8: Out-of-sample forecast evaluation 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	26.2	22.2	20.0	22.9	17.1	22.2	22.2	12.9	8.1	9.6	6.9	6.5	8.6	7.2
OLS	-3.1	-2.6	-3.2	-0.9	-1.0	-2.3	-1.3	-3.6	-3.7	-2.7	-4.6	-4.0	-3.7	-4.2
WLS _v	2.3	2.3	1.1	3.0	1.3	2.2	3.1	-6.0	-7.7	-6.1	-8.6	-7.9	-7.4	-8.0
WLS _s	1.2	0.9	-0.3	2.4	0.6	0.9	2.0	-5.6	-7.3	-5.8	-7.6	-7.6	-6.8	-7.5
MinT(Shrink)	-1.3	-1.0	-1.0	0.4	-0.3	-0.5	0.5	-11.9	-12.9	-11.1	-13.0	-11.3	-12.2	-12.4
Base	173.7	176.3	180.6	182.1	198.9	179.4	183.9	77.2	79.2	78.6	82.2	84.5	79.8	82.1
States								States by purpose of travel						
BU	3.5	7.0	8.9	7.9	7.1	7.0	7.0	9.4	7.0	6.9	6.5	0.8	7.1	5.9
OLS	-6.8	-4.7	-3.1	-3.9	-3.1	-4.2	-4.1	3.0	1.5	1.3	1.5	-0.6	1.8	1.4
WLS _v	-9.3	-5.2	-2.9	-4.3	-3.3	-4.9	-4.6	-1.3	-2.6	-2.4	-3.1	-7.2	-2.4	-3.2
WLS _s	-9.0	-5.6	-3.4	-4.4	-3.6	-5.2	-4.8	-0.4	-2.0	-1.7	-1.8	-6.2	-1.6	-2.4
MinT(Shrink)	-10.8	-7.1	-4.6	-6.1	-4.6	-6.7	-6.4	-3.1	-4.4	-4.3	-5.0	-8.5	-4.2	-4.9
Base	46.1	44.5	43.8	45.2	46.3	44.8	45.5	18.6	18.9	19.0	19.3	20.5	19.0	19.4
Zones								Zones by purpose of travel						
BU	-0.2	1.9	1.1	2.2	2.8	1.4	1.8	1.7	2.1	2.3	2.0	2.7	2.1	2.6
OLS	-3.5	-3.2	-3.7	-3.6	-3.0	-3.5	-3.5	0.9	0.9	0.7	1.0	2.1	0.9	1.3
WLS _v	-7.8	-5.2	-5.7	-5.7	-4.3	-5.8	-5.6	-4.0	-3.0	-2.8	-4.0	-2.5	-3.3	-2.9
WLS _s	-6.4	-4.9	-5.2	-4.8	-4.1	-5.2	-5.1	-4.0	-3.0	-2.8	-4.0	-2.5	-3.3	-2.9
MinT(Shrink)	-8.8	-6.4	-7.0	-7.2	-5.4	-7.1	-7.0	-5.0	-4.0	-4.0	-5.1	-3.4	-4.3	-3.9
Base	19.9	19.3	19.5	19.6	19.6	19.6	19.7	8.5	8.4	8.4	8.5	8.5	8.4	8.5
Regions								Regions by purpose of travel						
BU	-1.3	0.1	-0.8	0.4	1.7	-0.3	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OLS	-2.5	-2.1	-2.5	-2.4	-1.7	-2.4	-2.1	1.9	1.6	1.5	1.7	1.6	1.6	1.5
WLS _v	-6.4	-4.5	-5.2	-5.2	-3.5	-5.2	-4.7	-3.1	-2.8	-2.8	-3.4	-3.0	-3.0	-3.1
WLS _s	-4.9	-3.8	-4.3	-3.8	-2.7	-4.1	-3.6	-0.9	-0.9	-0.9	-0.8	-1.0	-0.9	-1.0
MinT(Shrink)	-6.9	-5.3	-6.2	-6.1	-4.3	-5.9	-5.6	-3.7	-3.4	-3.6	-4.0	-3.6	-3.5	-3.7
Base	10.1	9.9	10.0	10.1	10.0	10.0	10.0	4.4	4.3	4.3	4.4	4.4	4.3	4.4
ETS														
Australia								Australia by purpose of travel						
BU	12.1	5.5	4.3	3.7	5.6	5.4	6.0	8.1	4.5	4.8	4.8	5.1	5.2	5.1
OLS	-0.5	-1.1	-1.8	-1.3	0.0	-1.2	-0.7	-0.6	-0.1	0.6	0.2	0.0	0.1	0.0
WLS _v	3.3	0.4	-1.6	-1.2	1.8	-0.1	0.8	1.6	0.5	0.6	0.3	1.2	0.6	0.9
WLS _s	2.4	0.0	-1.9	-1.4	1.4	-0.6	0.4	0.9	0.2	0.3	0.0	1.1	0.3	0.7
MinT(Shrink)	1.4	-0.8	-3.2	-3.3	-0.1	-1.7	-1.0	0.1	-0.4	-0.5	-0.9	0.0	-0.4	-0.2
Base	162.9	164.1	170.1	177.7	194.3	170.3	177.1	63.7	63.7	64.5	67.0	72.8	65.0	67.6
States								States by purpose of travel						
BU	1.5	2.5	2.3	3.3	3.5	2.2	2.9	1.4	1.0	0.4	1.1	0.3	0.9	0.8
OLS	-2.6	-1.0	-0.6	-0.6	-1.0	-1.2	-1.2	-0.6	-0.8	-0.7	-0.6	-1.8	-0.7	-1.0
WLS _v	-2.8	-0.9	-0.8	-0.8	0.0	-1.3	-0.9	-1.3	-1.3	-1.4	-1.5	-2.5	-1.5	-1.7
WLS _s	-3.0	-1.0	-0.9	-0.9	-0.2	-1.5	-1.0	-1.3	-1.4	-1.4	-1.4	-2.3	-1.4	-1.6
MinT(Shrink)	-3.3	-1.2	-1.4	-1.6	-1.0	-1.8	-1.6	-1.7	-1.6	-1.8	-1.9	-3.0	-1.9	-2.1
Base	41.8	40.8	41.0	41.6	43.9	41.4	42.1	17.9	17.9	18.0	18.2	19.1	18.0	18.4
Zones								Zones by purpose of travel						
BU	0.9	0.1	0.4	1.0	2.0	0.5	0.9	0.2	-0.2	-0.7	-0.3	0.7	-0.3	0.1
OLS	-1.3	-1.4	-1.1	-1.1	-1.1	-1.2	-1.3	-0.4	-0.4	-0.4	-0.5	-0.5	-0.4	-0.4
WLS _v	-1.7	-2.1	-1.7	-1.5	-0.9	-1.7	-1.6	-1.4	-1.6	-1.9	-1.8	-1.5	-1.7	-1.6
WLS _s	-1.8	-2.0	-1.7	-1.6	-0.9	-1.7	-1.6	-1.4	-1.6	-1.9	-1.8	-1.5	-1.7	-1.6
MinT(Shrink)	-2.2	-2.4	-2.1	-2.1	-1.5	-2.2	-2.2	-1.8	-2.0	-2.2	-2.1	-1.9	-2.0	-2.0
Base	18.0	18.1	18.1	18.1	18.6	18.1	18.3	8.0	8.0	8.0	8.1	8.2	8.0	8.1
Regions								Regions by purpose of travel						
BU	-0.7	-0.8	-1.0	-0.3	0.4	-0.7	-0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OLS	-1.4	-1.4	-1.6	-1.4	-1.6	-1.5	-1.6	0.8	0.9	1.1	0.9	0.7	0.9	0.8
WLS _v	-2.4	-2.4	-2.5	-2.3	-2.1	-2.4	-2.3	-0.7	-0.8	-0.6	-0.8	-1.1	-0.7	-0.8
WLS _s	-2.4	-2.3	-2.5	-2.2	-2.0	-2.4	-2.3	-0.3	-0.2	-0.1	-0.2	-0.3	-0.2	-0.2
MinT(Shrink)	-2.8	-2.7	-2.9	-2.7	-2.5	-2.8	-2.7	-1.1	-1.1	-1.0	-1.2	-1.5	-1.1	-1.2
Base	9.4	9.4	9.4	9.5	9.6	9.4	9.5	4.1	4.1	4.1	4.2	4.2	4.1	4.2

Base shows the average RMSE ($\times 10^{-3}$) of the base forecasts. A negative (positive) entry above this row shows the percentage decrease (increase) in average RMSE of the reconciled forecasts relative to the base forecasts. Bold entries identify the best performing approaches.

Table 9: ARIMA and ETS models identified for the final rolling window ending November 2016.

Geographical division	Australia	States	Zones	Regions
Total number of series grouped by purpose of travel including the aggregates	5	35	135	380
Proportion of ARIMA models with				
Drift	0.80	0.31	0.19	0.07
Seasonal	0.80	0.91	0.73	0.47
Both	0.80	0.29	0.16	0.05
Proportion of ETS models with				
Trend	0.80	0.32	0.17	0.11
Seasonal	1.00	1.00	0.84	0.48
Both	0.80	0.32	0.14	0.05

References

- Athanasopoulos, G, Hyndman, RJ, Kourentzes, N, and Petropoulos, F (2017). Forecasting with Temporal Hierarchies. *European Journal of Operational Research* **262**(1), 60–74.
- Athanasopoulos, G, Ahmed, RA, and Hyndman, RJ (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting* **25**, 146–166.
- 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.
- Dunn, DM, Williams, WH, and Dechaine, TL (1976). Aggregate Versus Subaggregate Models in Local Area Forecasting. *Journal of American Statistical Association* **71**(353), 68–71.
- Horn, RA and Johnson, CR (1990). *Matrix Analysis*. Cambridge University Press.
- Hyndman, RJ (2017). *forecast: Forecasting Functions for Time Series and Linear Models*. R package version 8.0. <http://github.com/robjhyndman/forecast>.
- 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.
- 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* **97**, 16–32.
- Orcutt, GH, Watt, WH, and Edwards, JB (1968). Data Aggregation and Information Loss. *The American Economic Review* **58**(4), 773–787.
- 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.
- Sayal, H, Aston, JAD, Elliott, D, and Ombao, H (2016). An introduction to applications of wavelet benchmarking with seasonal adjustment. *Journal of the Royal Statistical Society. Series A: Statistics in Society* **180** (3), 863–889.
- 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.
- Shlifer, E and Wolff, RW (1979). Aggregation and Proration in Forecasting. *Management Science* **25**(6), 594–603.
- 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.
- van Erven, T and Cugliari, J (2014). “Game-theoretically 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 Automatic Control* **AC-31**, 1986.