

Department of Econometrics and Business Statistics

<http://monash.edu/business/ebs/research/publications>

Fast forecast reconciliation using linear models

Mahsa Ashouri, Rob J Hyndman, Galit Shmueli

October 2020

Working Paper 29/19

Fast forecast reconciliation using linear models

Mahsa Ashouri

Institute of Service Science, National Tsing Hua University, Taiwan

Email: mahsa.ashouri@iss.nthu.edu.tw

Corresponding author

Rob J Hyndman

Monash University, Clayton VIC 3800, Australia

Email: rob.hyndman@monash.edu

Galit Shmueli

Institute of Service Science, National Tsing Hua University, Taiwan

Email: galit.shmueli@iss.nthu.edu.tw

27 October 2020

JEL classification: C10,C14,C22

Fast forecast reconciliation using linear models

Abstract

Forecasting hierarchical or grouped time series using a reconciliation approach involves two steps: computing base forecasts and reconciling the forecasts. Base forecasts can be computed by popular time series forecasting methods such as Exponential Smoothing (ETS) and Autoregressive Integrated Moving Average (ARIMA) models. The reconciliation step is a linear process that adjusts the base forecasts to ensure they are coherent. However using ETS or ARIMA for base forecasts can be computationally challenging when there are a large number of series to forecast, as each model must be numerically optimized for each series. We propose a linear model that avoids this computational problem and handles the forecasting and reconciliation in a single step. The proposed method is very flexible in incorporating external data, handling missing values and model selection. We illustrate our approach using a dataset on monthly Australian domestic tourism, as well as a simulated dataset. We compare our approach to reconciliation using ETS and ARIMA, and show that our approach is much faster while providing similar levels of forecast accuracy.

Keywords: hierarchical forecasting, grouped forecasting, reconciling forecast, linear regression

1 Introduction

Modern data collection tools have dramatically increased the amount of available time series data (Januschowski et al. 2013). For example, the internet of things and point-of-sale scanning produce huge volumes of time series in a short period of time. Naturally, there is an interest in forecasting these time series, yet forecasting large collections of time series is computationally challenging.

1.1 Hierarchical and grouped time series

In many cases, these time series can be structured and disaggregated based on hierarchies or groups such as geographic location, product type, gender, etc. An example of hierarchical time series is the monthly number of Australian domestic tourists, which can be disaggregated into different states and then into different zones. Figure 1 shows a schematic of such a hierarchical

time series structure with three levels. The top level is the total series, formed by aggregating all the bottom level series. In the middle level, series are aggregations of their own child series; for instance, series A is the aggregation of AW and AX. Finally, the bottom level series, includes the most disaggregated series. In our example, A might represent the Northern Territory (NT) state, which can be disaggregated into northern coast NT and central NT.

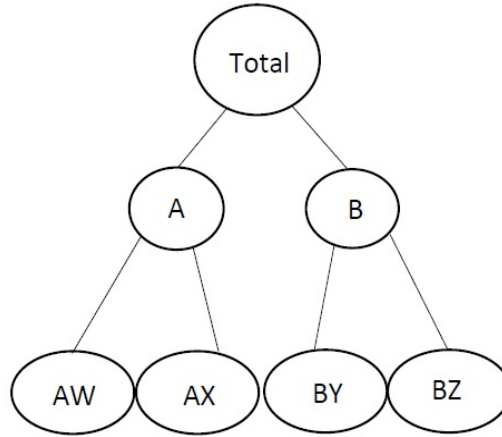


Figure 1: An example of a two level hierarchical structure.

Grouped time series involve more complicated aggregation structures compared to strictly hierarchical time series. To continue our Australian tourism example: suppose we have two grouping factors for each tourist that are not nested: purpose of travel (Business/Holiday) and sex (Male/Female). The disaggregated series for each combination of purpose of travel and sex can be combined to form purpose of travel subtotals, or sex subtotals. These subtotals can be combined to give the overall total. Both subtotals are of interest.

We can think of such structures as hierarchical time series without a unique hierarchy. A schematic of this grouped time series structure is shown in Figure 2 with two grouping factors, each of two levels (A/B and C/D). The series in this structure can be split first into groups A and B and then subdivided further into C and D (left side), or split first into C and D and then subdivided into A and B (right side). The final disaggregation is identical in both cases, but the middle level aggregates are different.

We use the same notation (following Hyndman & Athanasopoulos 2018) for both hierarchical and grouped time series. We denote the total series at time t by y_t , and the series at node Z (subaggregation level Z) and time t by $y_{Z,t}$. For describing the relationships between series, we use an $N \times M$ matrix, called the “summing matrix”, denoted by S , in which N is the overall number of nodes and M is the number of bottom level nodes. For example in Figure 1, $N = 7$ and $M = 4$, while in Figure 2, $N = 9$ and $M = 4$. Then we can write $y_t = S b_t$, where y_t is a

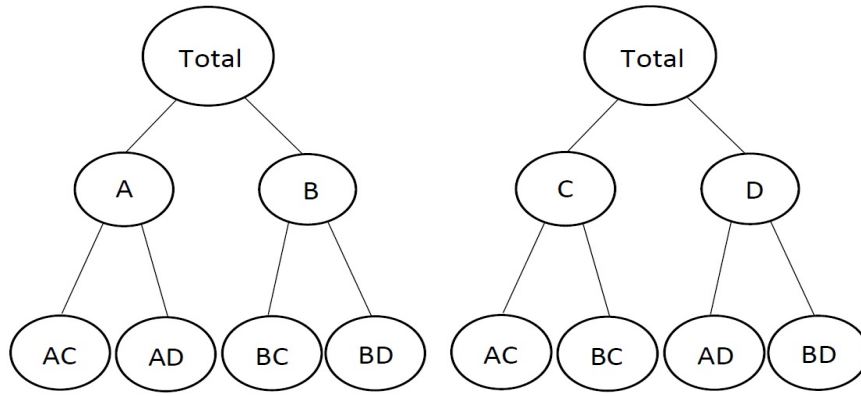


Figure 2: An example of a two level grouped structure.

vector of all the level nodes at time t and \mathbf{b}_t is the vector of all the bottom level nodes at time t . For the example shown in Figure 2, the equation can be written as follows:

$$\begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{D,t} \\ y_{AC,t} \\ y_{AD,t} \\ y_{BC,t} \\ y_{BD,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AC,t} \\ y_{AD,t} \\ y_{BC,t} \\ y_{BD,t} \end{pmatrix}.$$

1.2 Forecasting hierarchical time series

If we just forecast each series individually, we are ignoring the hierarchical or grouping structure, and the forecasts will not be “coherent”. That is, forecasts for lower level series will not necessarily add up to forecasts for higher level series. This means forecasts will not add up in a way that is consistent with the aggregation structure of the time series collection (Hyndman & Athanasopoulos 2018).

There are several available methods that consider the hierarchical structure information when forecasting time series. These include the top-down (Gross & Sohl 1990; Fliedner 2001), bottom-up (Kahn 1998), middle-out and optimal combination (Hyndman et al. 2011) approaches. In the top-down approach, we first forecast the total series and then disaggregate the forecast to form

lower level series forecasts based on a set of historical and forecasted proportions (for details see Athanasopoulos, Ahmed & Hyndman 2009). In the bottom-up approach, the forecasts in each level of the hierarchy can be computed by aggregating the bottom level series forecasts. However, we may not get good upper-level forecasts because the most disaggregated series can be noisy and so their forecasts are often inaccurate. In the middle-out approach, the process can be started from one of the middle levels and other forecasts can be computed using aggregation for upper levels and disaggregation for lower levels. Finally, optimal combination uses all the N forecasts for all of the series in the entire structure, and then uses an optimization process to reconcile the resulting forecasts. The advantage of the optimal combination method, compared with the other methods, is that it considers all information in the hierarchy, including any correlations among the series.

In the optimal combination method, reconciled forecasts can be computed using the following equation known as weighted least squares (WLS) (Wickramasuriya, Athanasopoulos & Hyndman 2019)

$$\tilde{\mathbf{y}}_{t+h} = \mathbf{S}(\mathbf{S}'\mathbf{W}_{t+h}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_{t+h}^{-1}\hat{\mathbf{y}}_{t+h}, \quad (1)$$

where $\hat{\mathbf{y}}_{t+h}$ represents a vector of h -step-ahead base forecasts for all levels of the hierarchy, and \mathbf{W}_{t+h} is the covariance matrix of forecast errors for the h -step-ahead base forecasts.

Several possible simple methods for estimating \mathbf{W}_{t+h} are available. Wickramasuriya, Athanasopoulos & Hyndman (2019) discuss a simple approximation whereby $\mathbf{W}_{t+h} = k_{t+h}\mathbf{\Lambda}$ with k_{t+h} being a positive constant, $\mathbf{\Lambda} = \text{diag}(\mathbf{S}\mathbf{1})$, and $\mathbf{1}$ being a unit vector of dimension M (the number of bottom level series). Note that $\mathbf{\Lambda}$ simply contains the row sums of the summing matrix \mathbf{S} , and that k_{t+h} will cancel out in (1).¹ Thus

$$\tilde{\mathbf{y}}_{t+h} = \mathbf{S}(\mathbf{S}'\mathbf{\Lambda}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{\Lambda}^{-1}\hat{\mathbf{y}}_{t+h}. \quad (2)$$

The most computationally challenging part of the optimal combination method is to produce all the base forecasts that make up $\hat{\mathbf{y}}_{t+h}$. In many applications, there may be thousands or even millions of individual series, and each of them must be forecast independently. The most popular time series forecasting methods such as ETS and ARIMA models (Hyndman & Athanasopoulos 2018) involve non-linear optimization routines to estimate the parameters via maximum likelihood estimation. Usually, multiple models are fitted for each series, and the best is selected

¹For simplicity of exposition, we used the structural scaling (wls_struct) summing matrix for reconciliation in all the results. In Appendix B, we provide two alternatives: Tables 15 and 16, Appendix B display the results for two other summing matrices using a shrinkage estimator (mint_shrink) and variance scaling (wls_var) for the Australian tourism example.

by minimizing Akaike's Information Criterion (Akaike 1998). This computational challenges increases with the number of lower level series as well as in the number of aggregations of interest.

We therefore propose a new approach to compute the base forecasts that is both computationally fast while maintaining an acceptable forecasting accuracy level. Our proposed approach avoids the computational challenge of using ETS or ARIMA that require numerical optimization for each series. It is very flexible in terms of incorporating external data, handling missing values, and model selection. And finally, our approach handles the forecasting and reconciliation in a single step.

2 Proposed approach: Linear model

Our proposed approach is based on using linear regression models for computing base forecasts. Suppose we have a linear model that we use for forecasting, and we wish to apply it to N different series which have some aggregation constraints. We have observations $y_{t,i}$ from times $t = 1, \dots, T$ and series $i = 1, \dots, N$. Then

$$y_{t,i} = \beta_i' x_{t,i} + \varepsilon_{t,i}$$

where $x_{t,i} = (1, x_{t,1,i}, \dots, x_{t,p,i})$ is a $(p + 1)$ -vector of regression variables, $\beta_i = (\beta_{0,i}, \beta_{1,i}, \beta_{2,i}, \dots, \beta_{p,i})$ is a $(p + 1)$ -vector of coefficients and $\varepsilon_{t,i}$ is the error². This equation for all the observations in matrix form can be written as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} X_1 & 0 & 0 & \dots & 0 \\ 0 & X_2 & 0 & \dots & 0 \\ 0 & 0 & X_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & X_N \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix} \quad (3)$$

$$Y = BX + E,$$

²If different predictors are chosen for different series, we can still choose a constant p that includes all the predictors, and then set the relevant coefficients in the X matrix to 0.

where $\mathbf{y}_i = (y_{1,i}, y_{2,i}, \dots, y_{T,i})$ is a T -vector, $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, \beta_{2,i}, \dots, \beta_{p,i})$ is a $(p+1)$ -vector, $\boldsymbol{\varepsilon}_i = (\varepsilon_{1,i}, \varepsilon_{2,i}, \dots, \varepsilon_{T,i})$ is a T -vector and \mathbf{X}_i is the $T \times (p+1)$ -matrix

$$\mathbf{X}_i = \begin{pmatrix} 1 & x_{1,i,1} & x_{1,i,2} & \dots & x_{1,i,p} \\ 1 & x_{2,i,1} & x_{2,i,2} & \dots & x_{2,i,p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{T,i,1} & x_{T,i,2} & \dots & x_{T,i,p} \end{pmatrix}.$$

Equation (3) can be written as $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$, with parameter estimates given by $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Then the base forecasts are obtained using

$$\hat{\mathbf{y}}_{t+h} = \mathbf{X}_{t+h}^* \hat{\mathbf{B}}, \quad (4)$$

where $\hat{\mathbf{y}}_{t+h}$ is an N -vector of forecasts, $\hat{\mathbf{B}}$ comprises N stacked $(p+1)$ -vectors of estimated coefficients, and \mathbf{X}_{t+h}^* is the $N \times N(p+1)$ matrix

$$\mathbf{X}_{t+h}^* = \begin{pmatrix} \mathbf{x}'_{t+h,1} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{x}'_{t+h,2} & 0 & \dots & 0 \\ 0 & 0 & \mathbf{x}'_{t+h,3} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & \mathbf{x}'_{t+h,N} \end{pmatrix}.$$

Note that we use \mathbf{X}_t^* to distinguish this matrix, which combines $\mathbf{x}_{t,i}$ across all series for one time from \mathbf{X}_i which combines $\mathbf{x}_{t,i}$ across all time for one series.

Finally, we can combine the two linear equations for computing base forecasts and reconciled forecasts (Equations (2) and (4)) to obtain the reconciled forecasts with a single equation:

$$\tilde{\mathbf{y}}_{t+h} = \mathbf{S}(\mathbf{S}'\boldsymbol{\Lambda}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Lambda}(\mathbf{X}_{t+h}^* \hat{\mathbf{B}}) = \mathbf{S}(\mathbf{S}'\boldsymbol{\Lambda}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Lambda}\mathbf{X}_{t+h}^* (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \quad (5)$$

2.1 Simplified formulation for a fixed set of predictors (X)

If we have the same set of predictor variables, \mathbf{X} , for all the series, we can write Equations (3) to (5) more easily using multivariate regression equations, and we can obtain all the reconciled

forecasts for all the series in one equation. In that case, Equation (3) can be rearranged as follows:

$$\begin{pmatrix} y_{11} & \dots & y_{1N} \\ y_{21} & \dots & y_{2N} \\ \vdots & & \vdots \\ y_{T1} & \dots & y_{TN} \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{T1} & \dots & X_{Tp} \end{pmatrix} \begin{pmatrix} \beta_{01} & \dots & \beta_{0N} \\ \beta_{11} & \dots & \beta_{1N} \\ \vdots & & \vdots \\ \beta_{p1} & \dots & \beta_{pN} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} & \dots & \varepsilon_{1N} \\ \varepsilon_{21} & \dots & \varepsilon_{2N} \\ \vdots & & \vdots \\ \varepsilon_{T1} & \dots & \varepsilon_{TN} \end{pmatrix}, \quad (6)$$

where \mathbf{Y} , \mathbf{X} , \mathbf{B} and \mathbf{E} are now matrices of size $T \times N$, $T \times (p+1)$, $(p+1) \times N$ and $T \times N$, respectively. Equations (4) to (5) can be written accordingly using Equation (6) and here $\mathbf{X}_{t+h,i}^* = \mathbf{X}_{t+h}^*$, where \mathbf{X}_{t+h}^* is an $h \times (p+1)$ matrix. This simpler formulation translates into a computational advantage, as the \mathbf{X} matrix is smaller, thereby easing matrix multiplication operations.

2.2 OLS predictors

As an example of the \mathbf{X}_t matrix in Equation (3), we can refer to the set of predictors proposed in Ashouri, Shmueli & Sin (2018) for modeling trend, seasonality and autocorrelation by using lagged values (y_{t-1}, y_{t-2}, \dots), trend variables and seasonal dummy variables:

$$y_t = \alpha_0 + \alpha_1 t + \beta_1 s_{1,t} + \dots + \beta_{m-1} s_{m-1,t} + \gamma_1 y_{t-1} + \dots + \gamma_p y_{t-p} + \delta z_t + \varepsilon_t.$$

Here, $s_{j,t}$ is a dummy variable taking value 1 if time t is in season j ($j = 1, 2, \dots, m$), y_{t-k} is the k th lagged value for y_t and z_t is some external information at time t . The seasonal period m depends on the problem; for instance, if we have daily data with day-of-week seasonality, then $m = 7$.

When a single OLS model is fitted to a collection of time series (e.g. several bottom-level series), then trend and seasonality predictors are the same for all series and we can use the simpler multivariate regression models in Equation (6). However, this formulation is inappropriate when including lags, which differ for each series, and/or series-specific external series, in which case we use the formulation in Equation (3).

When there are many options for choosing predictors, such as many seasonal dummy variables, lags, or high order trend terms, we can consider applying a model selection approach such as Akaike's Information Criterion or leave-one-out cross-validation (LOOCV) to select the best set of predictors in terms of prediction. In practice, LOOCV can be computationally heavy except in the special case of linear models (Christensen 2020) and therefore using linear models provide a viable solution. Also, when the number of seasons m is large (e.g. in hourly data), Fourier terms

can result in fewer predictors than dummy variables. The number of Fourier terms can also be determined using the same AIC or LOOCV approach (Hyndman & Athanasopoulos 2018).

2.3 Computational considerations

There are two ways for computing the above forecasts. First, we could create the matrices Y , X and E , and then directly use the above equations (taking advantage of sparse matrix routines) to obtain the forecasts. Alternatively, we could use separate regression models to compute the coefficients for each linear model individually. Although the matrix, $X'X$, which we need to invert is sparse and block diagonal, it is still faster to use the second approach involving separate regression models.

2.4 Prediction intervals

For obtaining prediction intervals, we need to compute the variance of reconciled forecasts as follows (Wickramasuriya, Athanasopoulos & Hyndman 2019):

$$\text{Var}(\tilde{y}_{t+h}) = SP\Sigma_{t+h}P'S', \quad (7)$$

where $P = (S'\Lambda S)^{-1}S'\Lambda$ and Σ_{t+h} denotes the variance of the base forecasts given by the usual linear model formula (Hyndman & Athanasopoulos 2018)

$$\Sigma_{t+h} = \sigma^2 \left[1 + X_{t+h}^* (X'X)^{-1} (X_{t+h}^*)' \right].$$

where σ^2 is the variance of the base model residuals. Assuming normally distributed errors, we can easily obtain any required prediction intervals corresponding to elements of \tilde{y}_{t+h} using the diagonals of (7).

3 Applications

In this section we illustrate our approach using a real data set and a simulated dataset³. The real data study includes forecasting monthly Australian domestic tourism. This dataset contains 304 series with both hierarchical and grouped structure with strong seasonality. In the simulation studies, we simulate series based on the monthly Australian domestic tourism data and systematically modify the forecasting horizon, noise level, hierarchy levels, and number of series. In [Appendix B](#), we use another real dataset on daily Wikipedia pageviews. In all these cases,

³All methods were run on a Linux server with Intel Xeon Silver 4108 (1.80GHz / 8-Cores / 11MB Cache)*2 and 8GB DDR4 2666 DIMM ECC Registered Memory. R version 3.6.1. All the displayed computation times are only for the reconciled point forecasts.

we compare the forecasting accuracy of ETS, ARIMA⁴ and the proposed linear OLS forecasting model, with and without the reconciliation step. In these applications, we used the weighted reconciliation approach from Equation (2). For comparing these methods, we use the average of Root Mean Square Errors (RMSEs) across all series and also display box plots for forecast errors along with the raw forecast errors. To aid visibility, we suppress plotting the outliers.

We apply two methods for generating forecasts that align with two different practical forecasting scenarios. The first approach is *rolling origin* forecasting, where we generate one-step-ahead forecasts (\tilde{y}_{t+1} where t changes). This mimics the scenario where data are refreshed every time period. In the second *fixed origin* method, forecasts are generated at a fixed time t for h steps ahead: $\tilde{y}_{t+1}, \tilde{y}_{t+2}, \dots, \tilde{y}_{t+h}$ (we replace lagged values of y by their forecasts if they occur at periods after the forecast origin; See algorithms ?? and ?? for more details in rolling and fixed origin approaches).

Algorithm 1 Hierarchical and grouped time series rolling origin OLS forecast reconciliation

```

1:  $data \leftarrow matrix(y_{1:T,1:N}, x_{1:T,1:N,1:p})$ 
2: for  $i \in \{1, \dots, N\}$  do
3:   for  $k \in \{1, \dots, h\}$  do
4:      $data \leftarrow data_{(1:(T-h)+(k-1)),i}$ 
5:      $newdata \leftarrow data_{((T-h)+k),i}$ 
6:      $fit \leftarrow lm(y_{t,i} \sim x_{t,i,1} + x_{t,i,2} + \dots + x_{t,i,p}, data = data)$ 
7:      $\hat{y}_{t+k,i} \leftarrow predict(fit, newdata = newdata)$ 
8:   end for
9: end for
10:  $S \leftarrow summing\ matrix(groups_{data})$ 
11:  $\Lambda \leftarrow diag(S1)$ 
12:  $adjust \leftarrow S(S'\Lambda S)^{-1}S\Lambda$ 
13: for  $i \in 1, \dots, h$  do
14:    $\tilde{y}_{(T-h)+i,1:N} \leftarrow adjust \times \hat{y}_{(T-h)+i,1:N}$ 
15: end for
16: return  $\tilde{y}_{((T-h):T),1:N}$ 

```

3.1 Australian domestic tourism

This dataset has 19 years of monthly visitor nights in Australia by Australian tourists, a measure used as an indicator of tourism activity (Wickramasuriya, Athanasopoulos & Hyndman 2019). The data were collected by computer-assisted telephone interviews with 120,000 Australians aged 15 and over (Tourism Research Australia 2005). The dataset includes 304 time series each of length 228 observations. The hierarchy and grouping structure for this dataset is made using geographic and purpose of travel information.

⁴For running ETS and ARIMA, we applied ETS' and ARIMA' functions from the 'fable' package (O Hara-Wild, Hyndman & Wang 2019). The two sets of functions were run independently and not immediately one after the other.

Algorithm 2 Hierarchical and grouped time series fixed origin OLS forecast reconciliation

```

1:  $data \leftarrow matrix(y_{1:T,1:N}, x_{1:T,1:N,1:p})$ 
2: for  $i \in \{1, \dots, N\}$  do
3:    $data \leftarrow data_{(1:(T-h),i)}$ 
4:    $fit \leftarrow lm(y_{t,i} \sim x_{t,i,1} + x_{t,i,2} + \dots + x_{t,i,p}, data = data)$ 
5:    $newdata \leftarrow data_{((T-h)+1),i}$ 
6:   for  $k \in \{1, \dots, h\}$  do
7:      $\hat{y}_{t+k,i} \leftarrow predict(fit, newdata = newdata)$ 
8:      $newdata \leftarrow \hat{data}_{((T-h)+k),i}$ 
9:   end for
10: end for
11:  $S \leftarrow summing\ matrix(groups_{data})$ 
12:  $\Lambda \leftarrow diag(S1)$ 
13:  $adjust \leftarrow S(S'\Lambda S)^{-1}S\Lambda$ 
14: for  $i \in 1, \dots, h$  do
15:    $\tilde{y}_{(T-h)+i,1:N} \leftarrow adjust \times \hat{y}_{(T-h)+i,1:N}$ 
16: end for
17: return  $\tilde{y}_{((T-h):T),1:N}$ 

```

Table 1: Australia geographic hierarchical structure.

| Series | Name | Label | Series | Name | Label |
|--------|----------------|-------|--------|--------------------|-------|
| Total | | | Region | | |
| 1 | Australia | Total | 55 | Lakes | BCA |
| State | | | 56 | Gippsland | BCB |
| 2 | NSW | A | 57 | Phillip Island | BCC |
| 3 | VIC | B | 58 | General 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 |
| Zone | | | 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 |
| Region | | | 92 | Flinders Range and Outback | DCD |
| | 36 | Sydney | AAA | 93 | Eyre Peninsula |
| 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 SouthWest | 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 |

In this dataset we have three levels of geographic divisions in Australia. In the first level, Australia is divided into seven “States” including New South Wales (NSW), Victoria (VIC), Queensland (QLD), South Australia (SA), Western Australia (WA), Tasmania (TAS) and Northern Territory (NT). In the second and third levels, it is divided into 27 “Zones” and 76 “Regions” (for

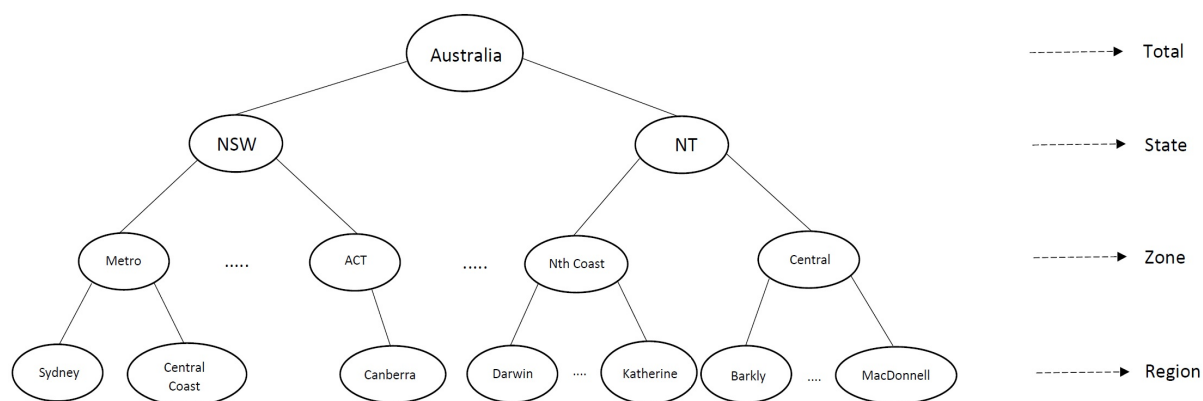


Figure 3: *Australian geographic hierarchical structure.*

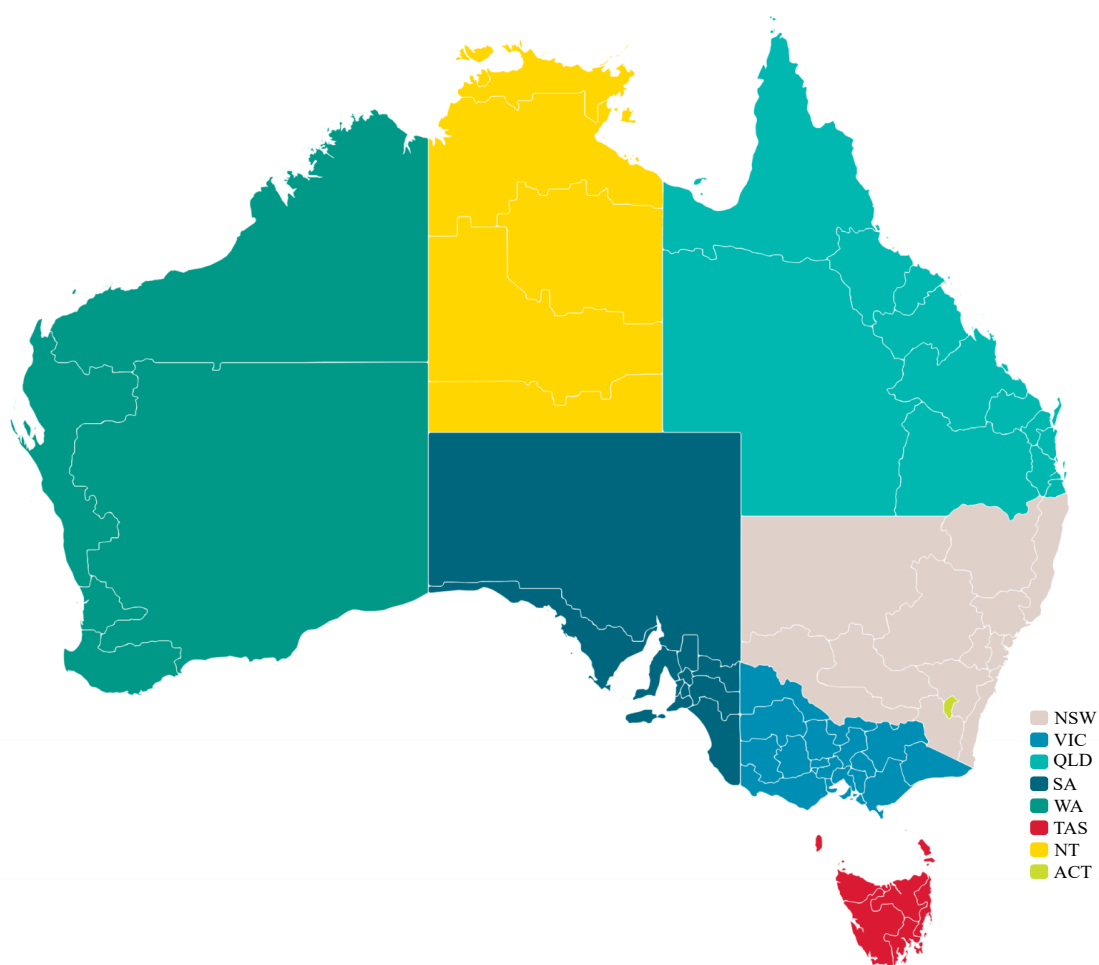


Figure 4: *Australia tourism region map - colors represent states.*

details about Australia geographic divisions see Figure 3 and Table 1 and also Figure 4 which shows Australia map divided by tourism region and colored by states⁵).

We have four purposes of travel: Holiday (Hol), Visiting friends and relatives (Vis), Business (Bus) and Other (Oth). So there are $76 \times 4 = 304$ series at the most disaggregate level. Based on the geographic hierarchy and purpose grouping, we end up with 8 aggregation levels with 555 series in total as shown in Table 2.

Table 2: Number of Australian domestic tourism series at each aggregation level.

| Division | Series |
|------------------|--------|
| Australia | 1 |
| State | 7 |
| Zone | 27 |
| Region | 76 |
| Purpose | 4 |
| State x Purpose | 28 |
| Zone x Purpose | 108 |
| Region x Purpose | 304 |
| Total | 555 |

We report the forecast results for all these aggregation levels, as well as the average RMSE across all the levels of the hierarchy. We used the same predictors in the OLS predictor matrix for the rolling and fixed origin approaches. For the rolling and fixed origin model, we include a quadratic trend, 11 dummy variables, and lags 1 and 12. This is intended to capture the monthly seasonality. In addition, before running the model, we partition the data into training and test sets, with the last 24 months (2 years) as our test set, and the rest as our training set.

Table 3: Mean(RMSE) on 24 months test set for ETS, ARIMA and OLS with and without reconciliation - Rolling origin.

| Level | Unreconciled | | | Reconciled | | |
|------------------|--------------|-------|------|------------|-------|------|
| | ETS | ARIMA | OLS | ETS | ARIMA | OLS |
| Total | 1516 | 1504 | 1634 | 1733 | 1840 | 1864 |
| State | 511 | 501 | 498 | 497 | 482 | 509 |
| Zone | 215 | 217 | 213 | 211 | 210 | 213 |
| Region | 123 | 125 | 117 | 118 | 120 | 117 |
| Purpose | 676 | 674 | 682 | 673 | 713 | 713 |
| State x Purpose | 213 | 217 | 213 | 208 | 209 | 213 |
| Zone x Purpose | 98 | 103 | 98 | 96 | 99 | 97 |
| Region x Purpose | 56 | 58 | 56 | 56 | 57 | 56 |

⁵www.tra.gov.au/tra/2016/Tourism_Region_Profiles/Region_profiles/index.html

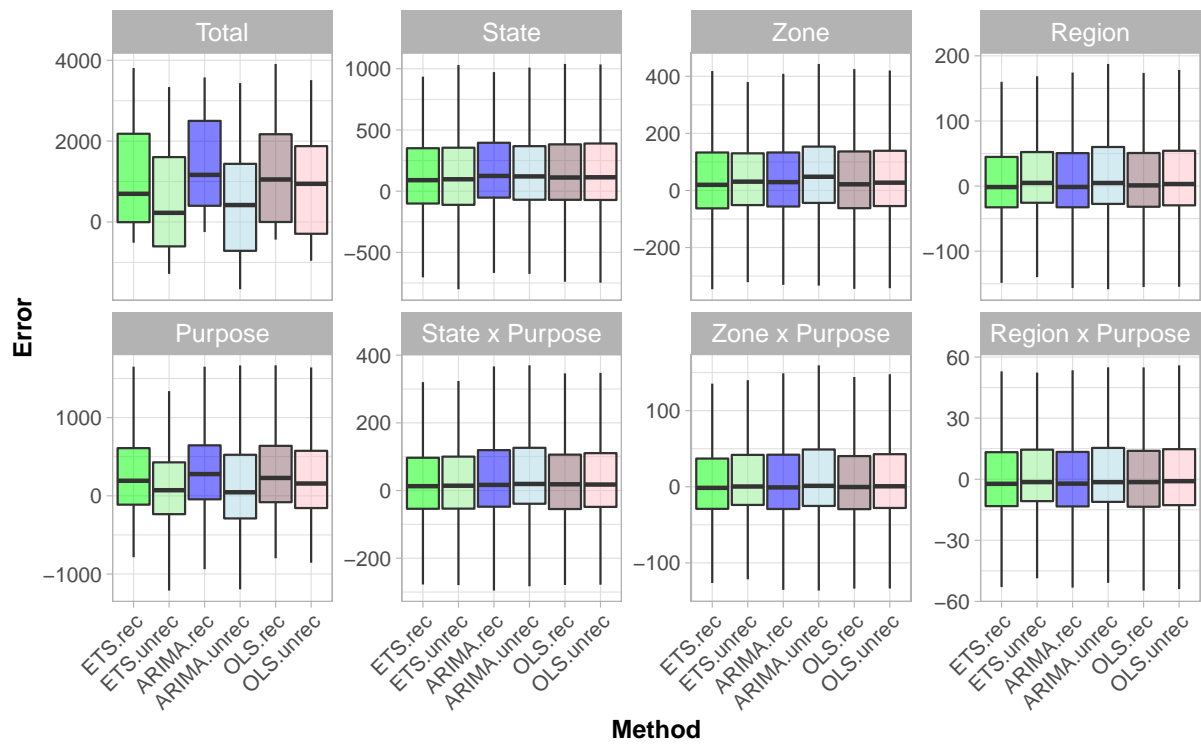


Figure 5: Box plots of rolling origin forecast errors from reconciled and unreconciled ETS, ARIMA and OLS methods at each hierarchical level.

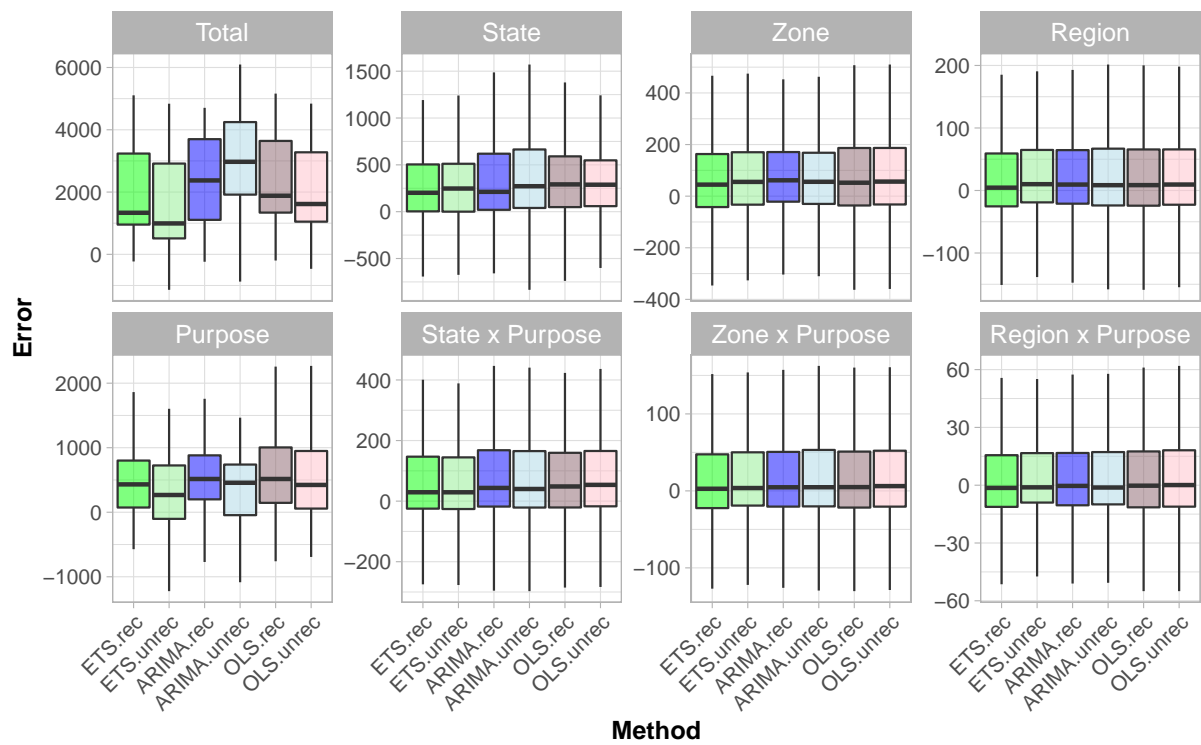


Figure 6: Box plots of fixed origin forecast errors for reconciled and unreconciled ETS, ARIMA and OLS methods at each hierarchical level.

Table 4: *Mean(RMSE) on 24 months test set for ETS, ARIMA and OLS with and without reconciliation - Fixed origin.*

| Level | Unreconciled | | | Reconciled | | |
|------------------|--------------|-------|------|------------|-------|------|
| | ETS | ARIMA | OLS | ETS | ARIMA | OLS |
| Total | 2239 | 3433 | 2529 | 2492 | 2744 | 2819 |
| State | 594 | 610 | 597 | 573 | 583 | 612 |
| Zone | 240 | 230 | 243 | 237 | 234 | 243 |
| Region | 133 | 132 | 127 | 127 | 127 | 126 |
| Purpose | 767 | 829 | 876 | 822 | 889 | 921 |
| State x Purpose | 227 | 233 | 237 | 222 | 226 | 236 |
| Zone x Purpose | 103 | 106 | 105 | 102 | 102 | 104 |
| Region x Purpose | 59 | 59 | 59 | 58 | 58 | 58 |

In Figures 5 and 6 we display the error box plots for both reconciled and unreconciled forecasts using all three methods, for the rolling origin and fixed origin forecasts. In these figures we see the error distributions across all the models.

Together with Tables 3 and 4, results show that our proposed OLS forecasting model produces forecast accuracy similar to ETS and ARIMA, which are computationally heavy for many time series (see Table 5). We also see the usefulness of the reconciliation in decreasing the average RMSE in all three methods. Except for the total series, reconciliation improves forecasts in all the hierarchy levels. Also, because the higher level series have higher counts, the errors are larger in magnitude (Appendix A shows the box plots with scaled errors⁶, to better compare errors across all the hierarchy levels). In addition, we see that (as expected) by applying rolling origin 1-step-ahead forecasts, the error densities are closer and more tightly distributed around zero than the fixed origin multi-step-ahead forecasts.

Figures 7 and 8 show the rolling and fixed origin forecast results for the total series and one of the bottom level series, AAVis (Sydney - Business). In these plots we have both reconciled (dashed lines) and unreconciled (dotted lines) forecasts and we see that the reconciliation step improves the forecasts in this series. We also see that the OLS model forecast accuracy is similar to the other two methods.

Figures 9 and 10 display the prediction interval for the OLS approach, with and without reconciliation forecasts for the total series and one of the bottom level series, AAVis (Sydney - Visiting).

Table 5 compares the computation time of the three methods for rolling and fixed origin forecasting. We see that the OLS forecasting model is much faster compared to the other methods.

⁶Scaled errors are computed by subtracting the mean and dividing by the standard deviation.

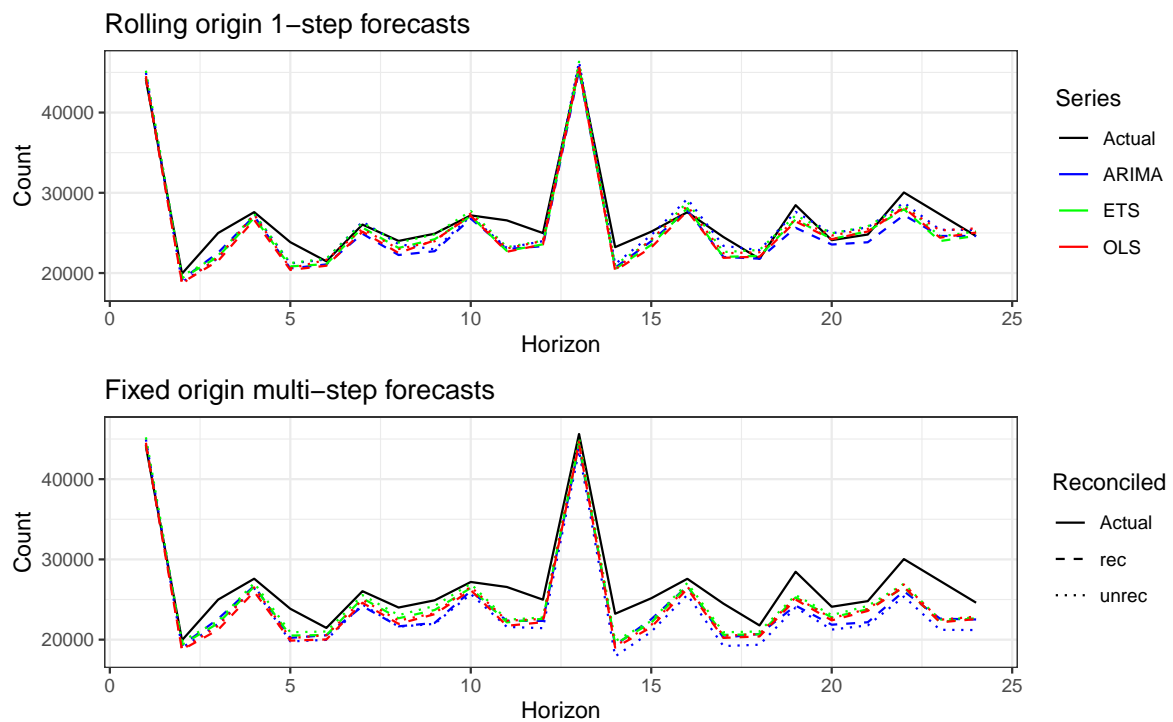


Figure 7: Comparing ETS, ARIMA and OLS forecasts (reconciled and unreconciled) for 'Total' series. (Top: rolling origin, Bottom: fixed origin).

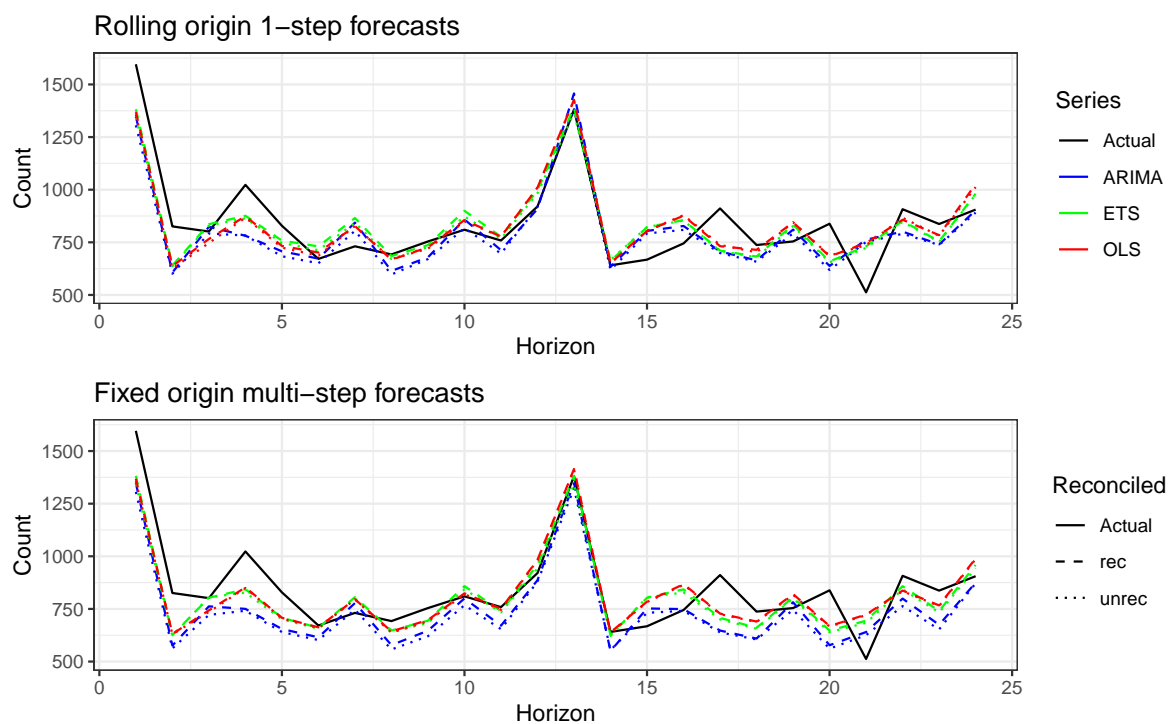


Figure 8: Comparing ETS, ARIMA and OLS forecasts (reconciled and unreconciled) for 'AAAVis' bottom-level series. (Top: rolling origin, Bottom: fixed origin).

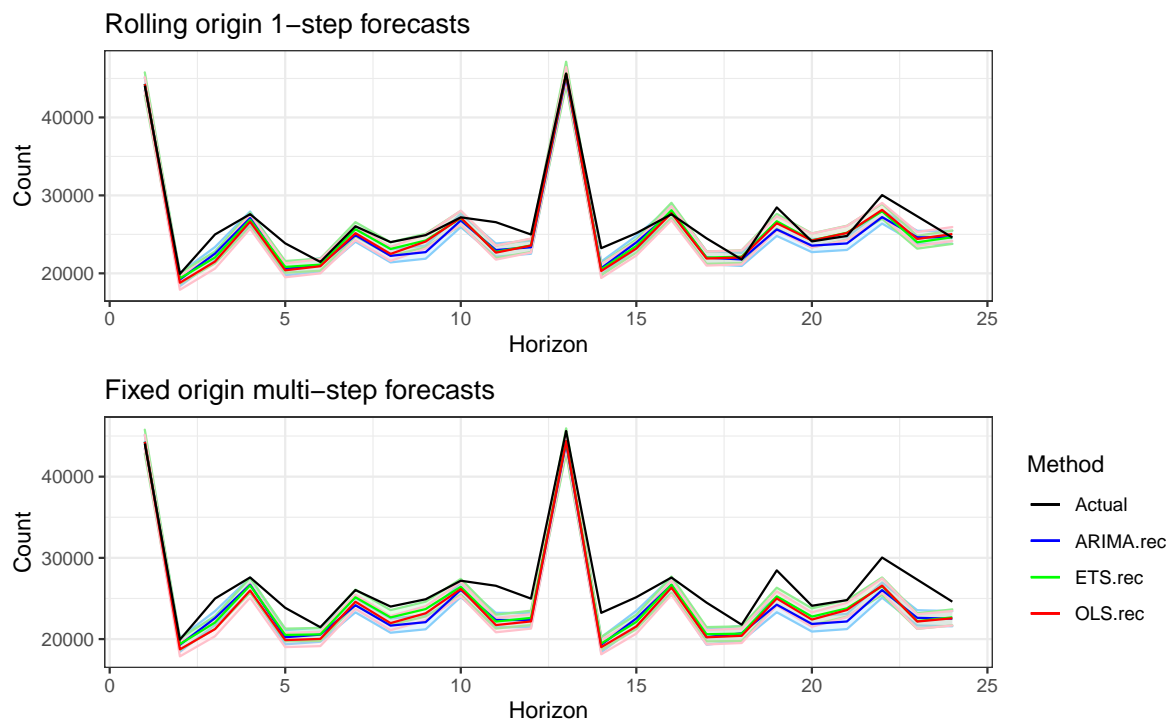


Figure 9: Comparing ETS, ARIMA and OLS reconciled forecasts and prediction intervals for 'Total' series. (Top: rolling origin, Bottom: fixed origin).

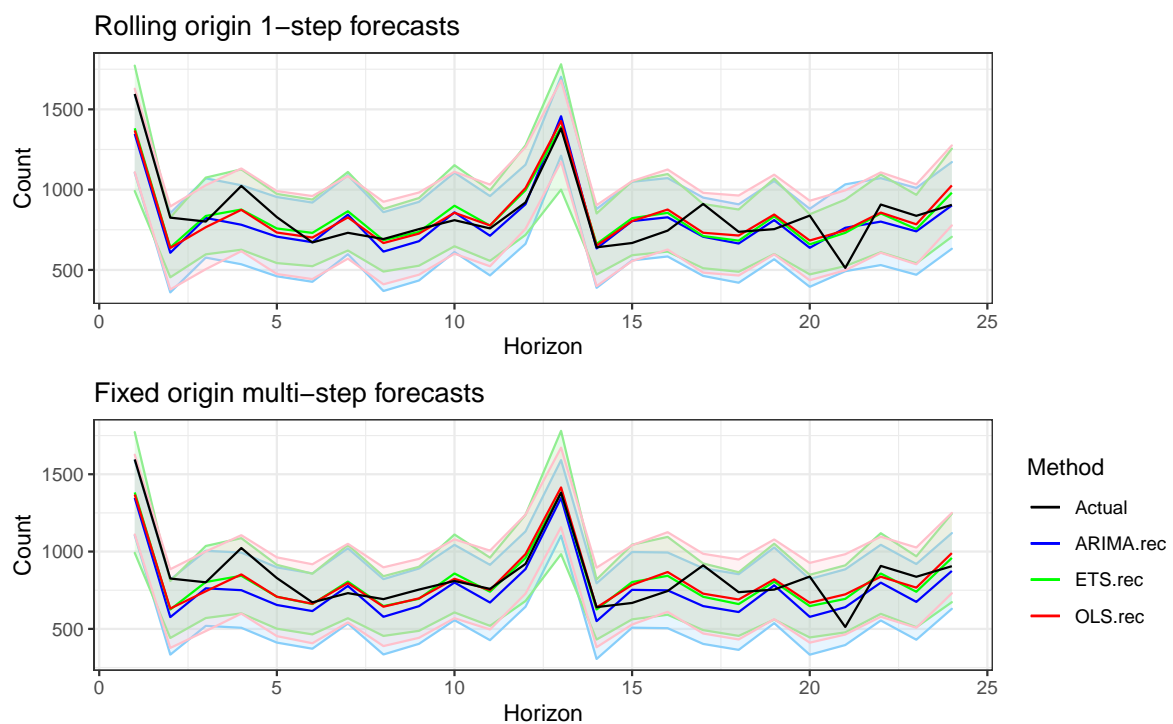


Figure 10: Comparing ETS, ARIMA and OLS reconciled forecasts and prediction intervals for 'AAAVis' bottom-level series. (Top: rolling origin, Bottom: fixed origin).

Also, since reconciliation is a linear process, in all methods it is very fast and does not affect computation time significantly.

Since we are using a linear model, we can easily include exogenous variables which can often be helpful in improving forecast accuracy. In this application, we tried including an “Easter” dummy variable indicating the timing of Easter, but its affect on forecast accuracy was minimal, so it was omitted in the model reported here.

Finally, Table 6 shows that, as mentioned in Section 2.3, computation is faster using separate regression models compared to the matrix approach (even using sparse matrix algebra).

Table 5: *Computation time (seconds) for ETS, ARIMA and OLS with reconciliation, for Rolling and fixed origin forecasts, on a 24 months test set.*

| | Rolling origin | Fixed origin |
|-------|----------------|--------------|
| ETS | 14648 | 618 |
| ARIMA | 30346 | 1085 |
| OLS | 48 | 18 |

Table 6: *Computation time (seconds) for OLS using the matrix approach and separate regression models, with reconciliation, for rolling and fixed origin, on a 12 months test set.*

| | Rolling origin | Fixed origin |
|-----------------|----------------|--------------|
| Matrix approach | 210 | 106 |
| Separate models | 48 | 18 |

3.2 Australian domestic tourism simulation study

We provide results from two simulation studies based on the Australian domestic tourism dataset, to evaluate the sensitivity of our results to several factors. In the first study, we simulate bottom-level series similar to the real bottom-level series of the tourism data, with the same number of series and the same length. We then generate forecasts for four forecast horizons (12, 24, 36 and 48 months) with four different noise levels (standard deviation=0.01, 0.1, 0.5 and 1)⁷.

Tables 7 and 8 display the average of the RMSEs for 12 to 48 month-ahead forecasts with different noise levels. Results are shown for the base and the reconciled forecasts for both rolling and fixed origin approaches. The results show that, as expected, by increasing the forecast horizon and/or noise level, the average RMSE increases in all the three methods. Also, the proposed OLS approach shows similar results compared with ETS and ARIMA. It should be noted that for both rolling and fixed origin forecasts in the OLS approach we use the same set of predictors as the Australian domestic tourism example.

⁷Since the level of the series are different, we first scale the simulated series (subtracting by mean and dividing by standard deviation), add the white noise series and then we rescale the series.

Table 7: Mean RMSE of rolling origin forecasts for simulated data (304 bottom-level series and 8 levels of hierarchy), for different error levels, methods (ETS, ARIMA, OLS), forecast horizons, with/without reconciliation.

| Reconciliation | Error | Forecast horizon | ETS | ARIMA | OLS |
|----------------|-------|------------------|-------|-------|-------|
| rec | 0.01 | 12 | 123.8 | 119.8 | 130.4 |
| rec | 0.01 | 24 | 122.6 | 119.4 | 128.8 |
| rec | 0.01 | 36 | 124.6 | 122.3 | 131.8 |
| rec | 0.01 | 48 | 122.1 | 120.4 | 129.0 |
| rec | 0.10 | 12 | 123.7 | 120.1 | 129.7 |
| rec | 0.10 | 24 | 122.9 | 120.3 | 129.2 |
| rec | 0.10 | 36 | 125.8 | 123.7 | 133.1 |
| rec | 0.10 | 48 | 123.5 | 122.1 | 130.5 |
| rec | 0.50 | 12 | 143.9 | 143.2 | 146.9 |
| rec | 0.50 | 24 | 149.9 | 146.3 | 153.2 |
| rec | 0.50 | 36 | 155.8 | 151.5 | 160.0 |
| rec | 0.50 | 48 | 154.5 | 150.8 | 159.7 |
| rec | 1.00 | 12 | 192.9 | 198.1 | 193.9 |
| rec | 1.00 | 24 | 207.1 | 209.0 | 209.3 |
| rec | 1.00 | 36 | 215.7 | 215.4 | 218.0 |
| rec | 1.00 | 48 | 218.4 | 217.5 | 220.9 |
| unrec | 0.01 | 12 | 132.2 | 125.8 | 132.9 |
| unrec | 0.01 | 24 | 128.5 | 125.4 | 131.0 |
| unrec | 0.01 | 36 | 130.2 | 128.5 | 134.0 |
| unrec | 0.01 | 48 | 127.8 | 126.6 | 131.7 |
| unrec | 0.10 | 12 | 132.0 | 126.2 | 132.2 |
| unrec | 0.10 | 24 | 129.5 | 126.0 | 131.4 |
| unrec | 0.10 | 36 | 131.9 | 129.9 | 135.1 |
| unrec | 0.10 | 48 | 129.5 | 128.3 | 133.0 |
| unrec | 0.50 | 12 | 149.5 | 149.7 | 148.2 |
| unrec | 0.50 | 24 | 156.5 | 153.7 | 154.5 |
| unrec | 0.50 | 36 | 163.3 | 158.1 | 160.7 |
| unrec | 0.50 | 48 | 161.9 | 157.7 | 160.7 |
| unrec | 1.00 | 12 | 200.9 | 205.4 | 194.5 |
| unrec | 1.00 | 24 | 214.9 | 216.3 | 210.1 |
| unrec | 1.00 | 36 | 222.0 | 221.9 | 218.1 |
| unrec | 1.00 | 48 | 223.6 | 223.5 | 221.2 |

Figures 11 and 12 display one of the bottom level series with its ARIMA, ETS, and OLS 12 to 48 months ahead reconciled forecasts and prediction intervals, while changing the noise levels. From these two figures we see the OLS approach prediction intervals are almost identical to those from ARIMA and ETS.

Figures 13 and 14 show the computation time (seconds) for ETS, ARIMA and OLS methods on rolling and fixed origin forecasts. From these figures we see that increasing the forecast horizon from one to 48 months increases computation time almost linearly, while the noise level does not change the computation time. Also, the computation time for ARIMA and ETS is much

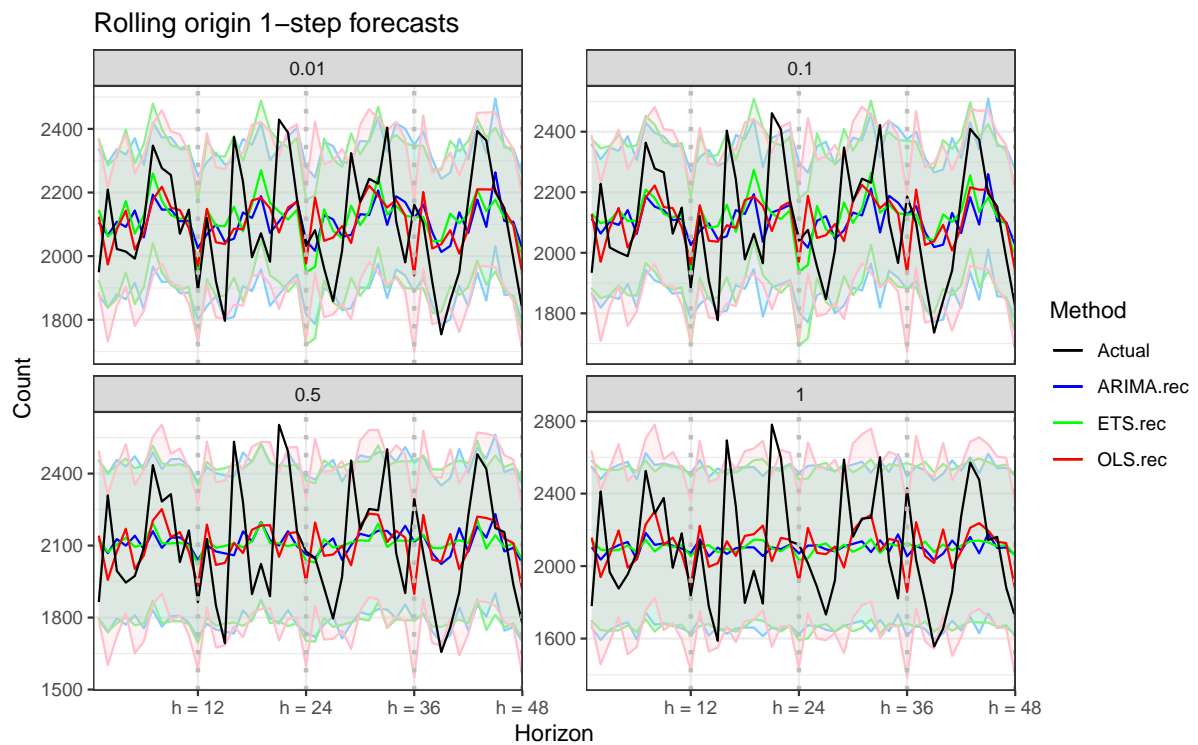


Figure 11: Comparing reconciled 'rolling origin' forecasts and prediction intervals for a sample bottom-level series across different error levels (different panels).

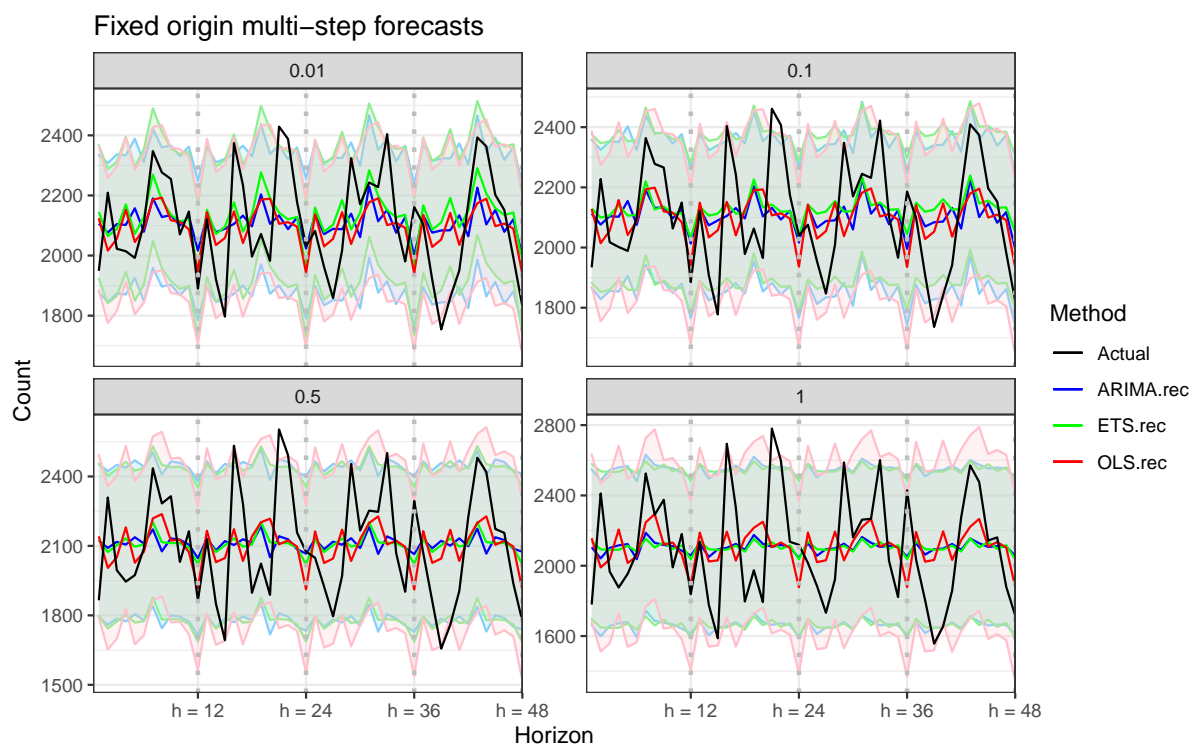


Figure 12: Comparing reconciled 'fixed origin' forecasts and prediction intervals for a sample bottom-level series across different error levels (different panels).

Table 8: Mean RMSE of fixed origin forecasts for simulated data (304 bottom-level series and 8 levels of hierarchy), for different error levels, methods (ETS, ARIMA, OLS), forecast horizons, with/without reconciliation.

| Reconciliation | Error | Forecast horizon | ETS | ARIMA | OLS |
|----------------|-------|------------------|-------|-------|-------|
| rec | 0.01 | 12 | 123.6 | 119.8 | 131.4 |
| rec | 0.01 | 24 | 126.5 | 124.2 | 137.3 |
| rec | 0.01 | 36 | 133.4 | 131.9 | 148.6 |
| rec | 0.01 | 48 | 133.5 | 133.9 | 152.3 |
| rec | 0.10 | 12 | 124.8 | 120.3 | 130.7 |
| rec | 0.10 | 24 | 128.0 | 124.5 | 137.2 |
| rec | 0.10 | 36 | 135.3 | 132.2 | 148.9 |
| rec | 0.10 | 48 | 135.5 | 134.3 | 152.6 |
| rec | 0.50 | 12 | 144.6 | 143.4 | 147.8 |
| rec | 0.50 | 24 | 154.5 | 151.6 | 158.2 |
| rec | 0.50 | 36 | 162.8 | 160.7 | 170.0 |
| rec | 0.50 | 48 | 164.1 | 163.5 | 174.3 |
| rec | 1.00 | 12 | 194.1 | 197.7 | 194.7 |
| rec | 1.00 | 24 | 212.4 | 213.3 | 213.0 |
| rec | 1.00 | 36 | 221.9 | 222.1 | 224.8 |
| rec | 1.00 | 48 | 225.8 | 227.1 | 231.0 |
| unrec | 0.01 | 12 | 133.1 | 125.3 | 133.3 |
| unrec | 0.01 | 24 | 136.3 | 129.7 | 139.1 |
| unrec | 0.01 | 36 | 143.4 | 138.9 | 150.2 |
| unrec | 0.01 | 48 | 143.8 | 141.0 | 153.8 |
| unrec | 0.10 | 12 | 133.4 | 126.2 | 132.5 |
| unrec | 0.10 | 24 | 136.9 | 130.7 | 138.9 |
| unrec | 0.10 | 36 | 144.4 | 140.0 | 150.4 |
| unrec | 0.10 | 48 | 145.0 | 142.4 | 154.0 |
| unrec | 0.50 | 12 | 150.3 | 147.8 | 148.7 |
| unrec | 0.50 | 24 | 161.0 | 156.6 | 159.2 |
| unrec | 0.50 | 36 | 170.4 | 167.5 | 170.8 |
| unrec | 0.50 | 48 | 172.2 | 171.0 | 175.0 |
| unrec | 1.00 | 12 | 202.2 | 204.2 | 195.0 |
| unrec | 1.00 | 24 | 220.9 | 220.4 | 213.5 |
| unrec | 1.00 | 36 | 230.1 | 229.1 | 225.1 |
| unrec | 1.00 | 48 | 233.8 | 233.3 | 231.3 |

longer than OLS. Note that the computation time for the reconciliation step is less than a second and therefore that would be similar for base and reconciled forecasts.

In the second simulation study, we fix the forecast horizon at $h = 24$ and the noise at 0.5, and then create 4 different hierarchy levels (8, 10, 12 and 18) — obtained using the hierarchy structures in Table 2, 9, 10 and 11 (8 = same as Australian domestic tourism data; 9 = adding one hierarchy factor, resulting in 10 levels; 10 = adding two hierarchy factors, resulting in 12 levels; and 11 = adding two hierarchy factors and one grouping factor, resulting in 18 levels)⁸).

⁸For simplicity we just include 2-way aggregation combinations.

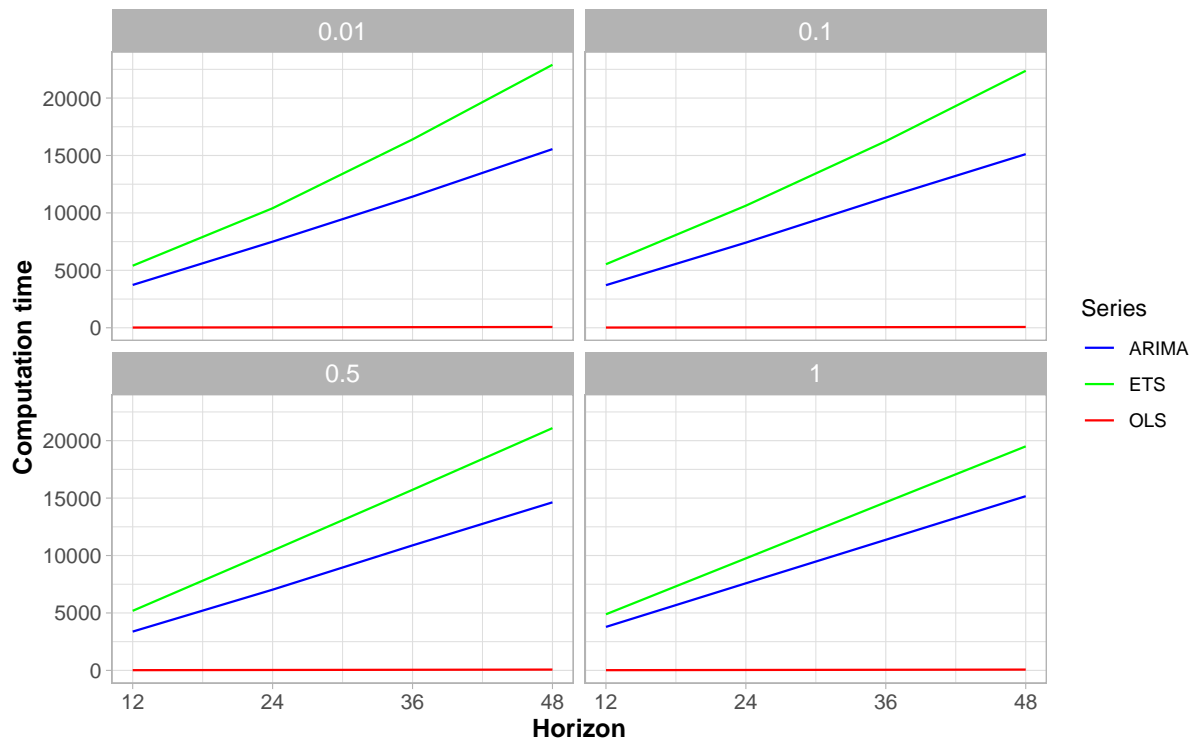


Figure 13: Computation time (seconds) for rolling origin reconciled forecasts using ETS, ARIMA and OLS, by horizon and by different error values (panels), for 304 bottom-level series and 8 levels of hierarchy.

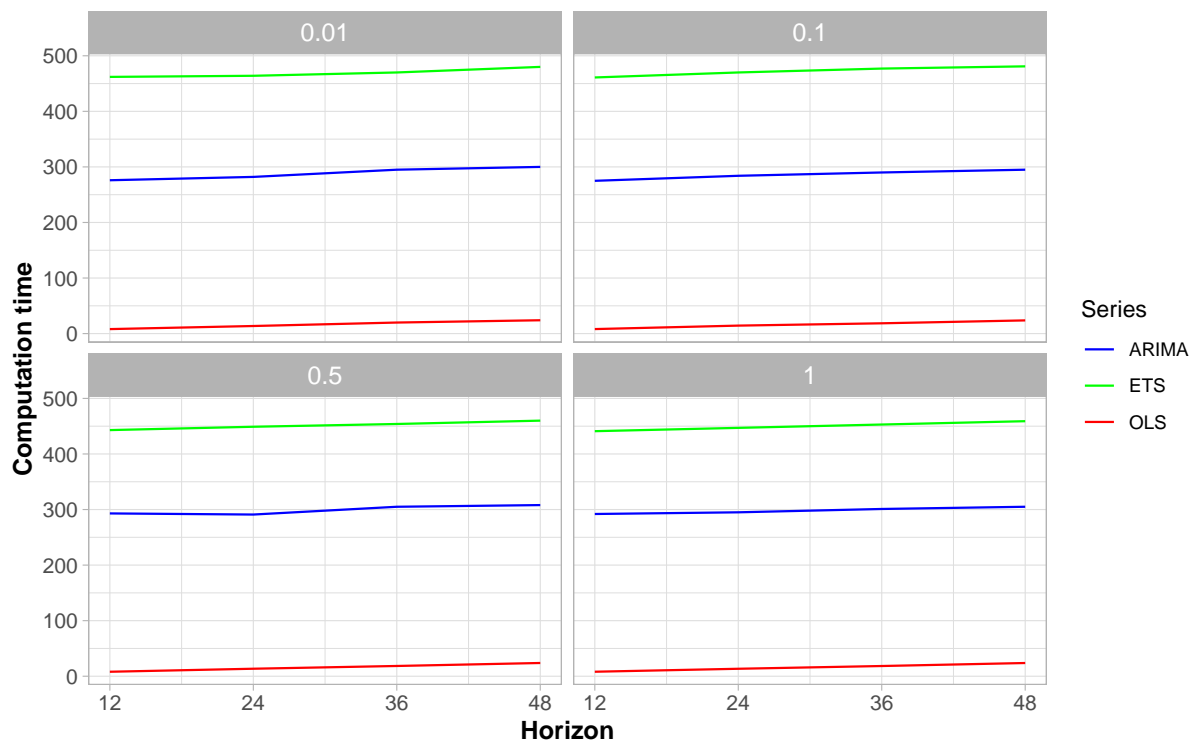


Figure 14: Computation time (seconds) for fixed origin reconciled forecasts using ETS, ARIMA and OLS, by horizon and by different error values (panels), for 304 bottom-level series and 8 levels of hierarchy.

Table 9: *Number of simulated Australian domestic tourism series at each aggregation level - adding one hierarchy variable (Level 1)*

| Division | Series |
|-------------------|--------|
| Australia | 1 |
| Level 1 | 3 |
| State | 7 |
| Zone | 27 |
| Region | 76 |
| Purpose | 4 |
| Level 1 x Purpose | 12 |
| State x Purpose | 28 |
| Zone x Purpose | 108 |
| Region x Purpose | 304 |
| Total | 570 |

Table 10: *Number of simulated Australian domestic tourism series at each aggregation level - adding two hierarchy variables (Level 1 and Level 2)*

| Division | Series |
|-------------------|--------|
| Australia | 1 |
| Level 1 | 3 |
| Level 2 | 5 |
| State | 7 |
| Zone | 27 |
| Region | 76 |
| Purpose | 4 |
| Level 1 x Purpose | 12 |
| Level 2 x Purpose | 20 |
| State x Purpose | 28 |
| Zone x Purpose | 108 |
| Region x Purpose | 304 |
| Total | 595 |

We also simulated four sizes of bottom-level series (304, 608, 1520 and 3040). In order to add series, we change the number of ‘Purpose’ categories (grouping factor) in the Australian domestic tourism example. Table 12 displays the total number of series based on 304, 608, 1520 and 3040 bottom levels series with 8, 10, 12 and 18 hierarchy levels.

Tables 13 and 14 display the average RMSEs by number of bottom level series (304, 608, 1520 and 3040), number of hierarchy levels (8, 10, 12 and 18), for ARIMA, ETS and OLS, with rolling and fixed origin forecasts. These results are both for reconciled and unreconciled forecasts. We see from these tables that reconciling the forecasts decreases the mean RMSE, while increasing the number of series increases the mean RMSE.

Table 11: *Number of simulated Australian domestic tourism series at each aggregation level - adding two hierarchy and one grouping variables (Level 1, Level 2 and Group 1)*

| Division | Series |
|-------------------|--------|
| Australia | 1 |
| Level 1 | 3 |
| Level 2 | 5 |
| State | 7 |
| Zone | 27 |
| Region | 76 |
| Purpose | 4 |
| Group 1 | 5 |
| Level 1 x Purpose | 12 |
| Level 2 x Purpose | 20 |
| State x Purpose | 28 |
| Zone x Purpose | 108 |
| Level 1 x Group 1 | 5 |
| Level 2 x Group 1 | 6 |
| State x Group 1 | 7 |
| Zone x Group 1 | 27 |
| Purpose x Group 1 | 20 |
| Bottom level | 304 |
| Total | 665 |

Table 12: *Total number of the series in the hierarchy structure based on the different number of series with 8, 10, 12 and 18 levels of the hierarchy.*

| Bottom level series | Total series | | | |
|---------------------|--------------|------|------|------|
| | 8 | 10 | 12 | 18 |
| 304 | 555 | 570 | 595 | 665 |
| 608 | 999 | 1026 | 1071 | 1161 |
| 1520 | 2331 | 2394 | 2499 | 2649 |
| 3040 | 4551 | 4674 | 2643 | 5129 |

In Figures 15 and 16, we compare the reconciled ETS, ARIMA, and OLS forecasts and prediction intervals for one of the bottom level series. From these figures we see that all three approaches have similar prediction intervals.

Finally, in Figures 17 and 18, we display the computation time for rolling and fixed origin approaches with different levels and number of bottom level series. We see that increasing the number of series increases computation time linearly. Increasing the number of levels slightly increases the computation time. In both figures we see a substantial difference in computation time between ARIMA and ETS compared with our OLS approach.⁹

⁹The reconciliation step computation time varies from less than one second to around three seconds, as a function of the number of series. All figures display only the base forecasts computation times.

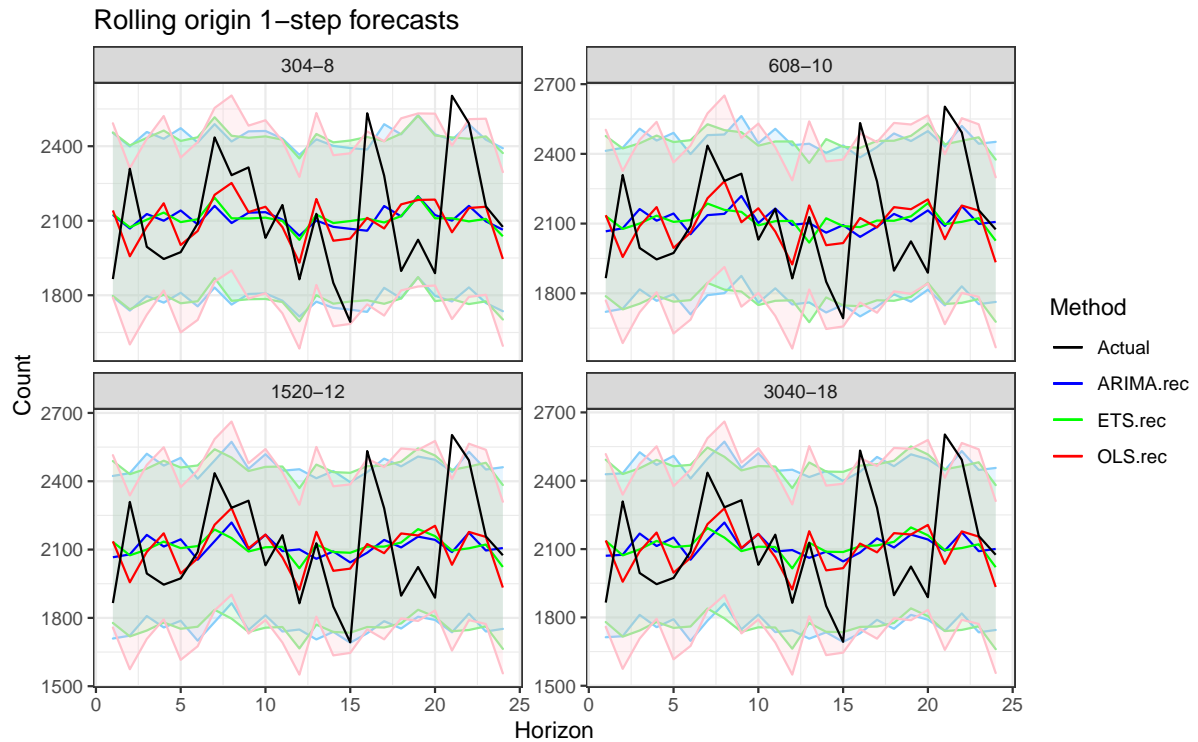


Figure 15: Comparing reconciled 'rolling origin' forecasts and prediction intervals for a sample bottom-level series, for different number of bottom-level series and hierarchy levels (different panels). Simulated series has error value 0.5 and 24 months test set.

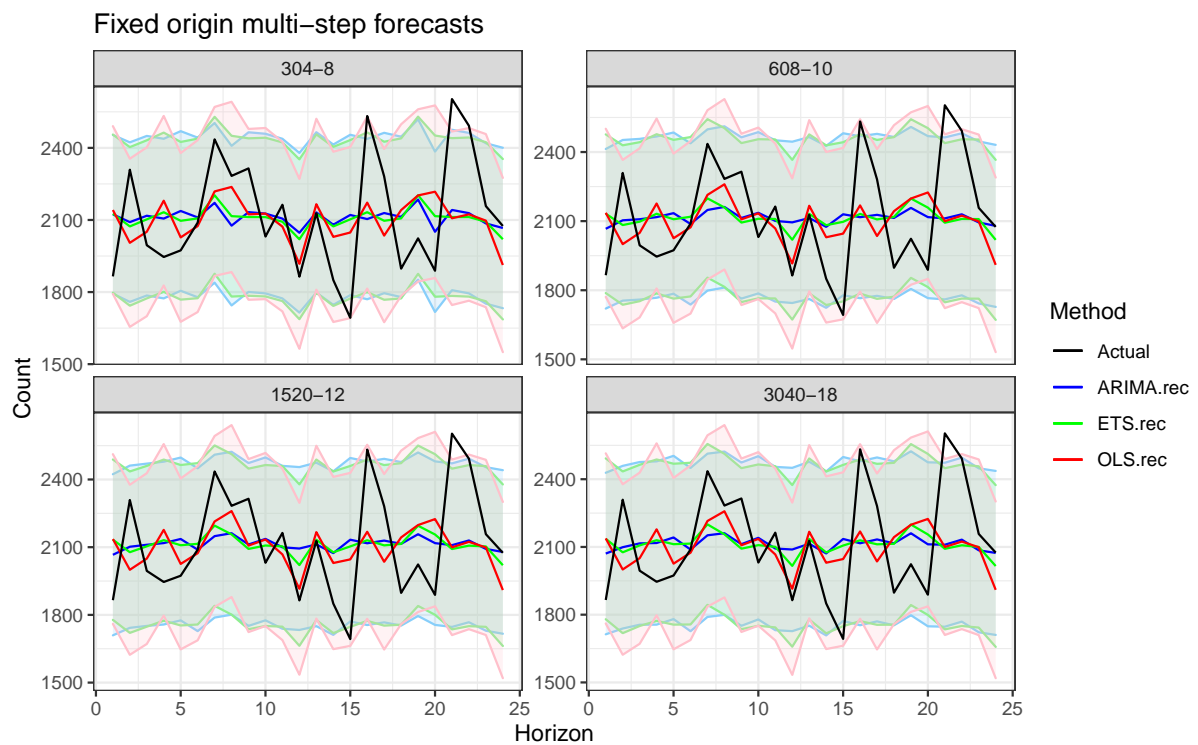


Figure 16: Comparing reconciled fixed origin forecasts and prediction intervals for a sample bottom-level series, for different number of bottom-level series and hierarchy levels (different panels). Simulated series has error value 0.5 and 24 months test set.

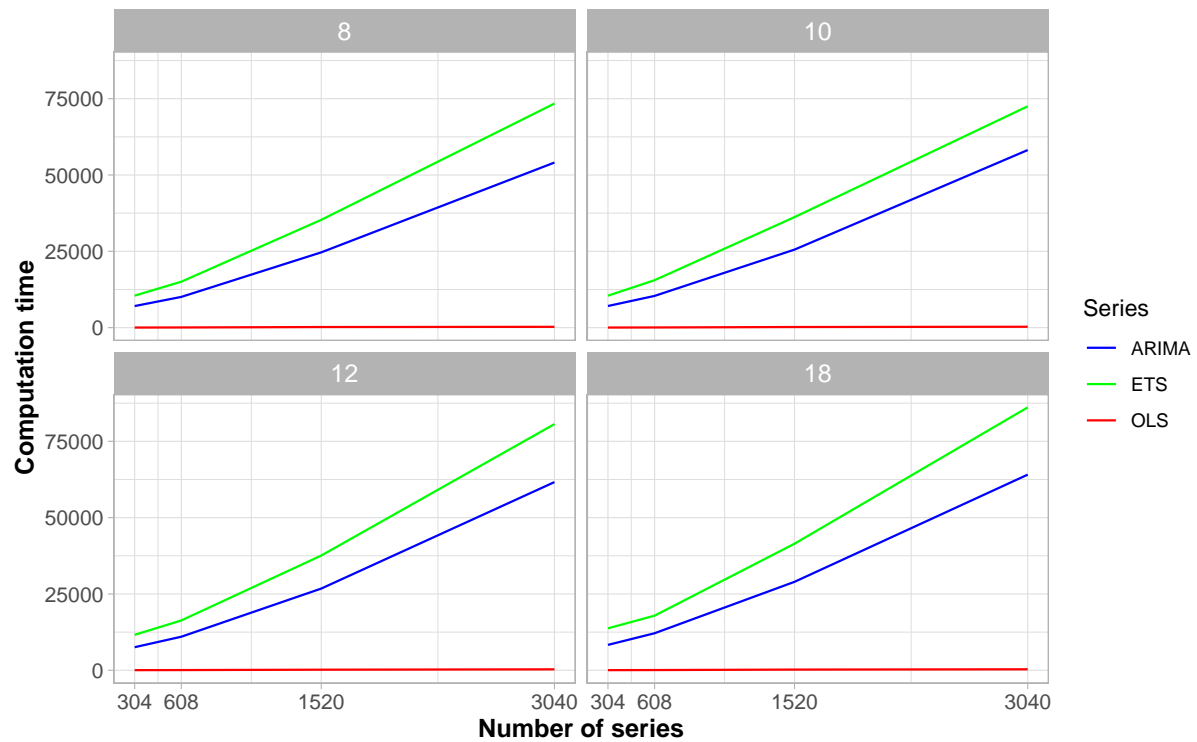


Figure 17: Computation time (seconds) for rolling origin reconciled forecasts using ETS, ARIMA and OLS, by number of bottom-level series (x-axis), and by levels of hierarchy (panels). Simulated series has error value 0.5 and 24 months test set.

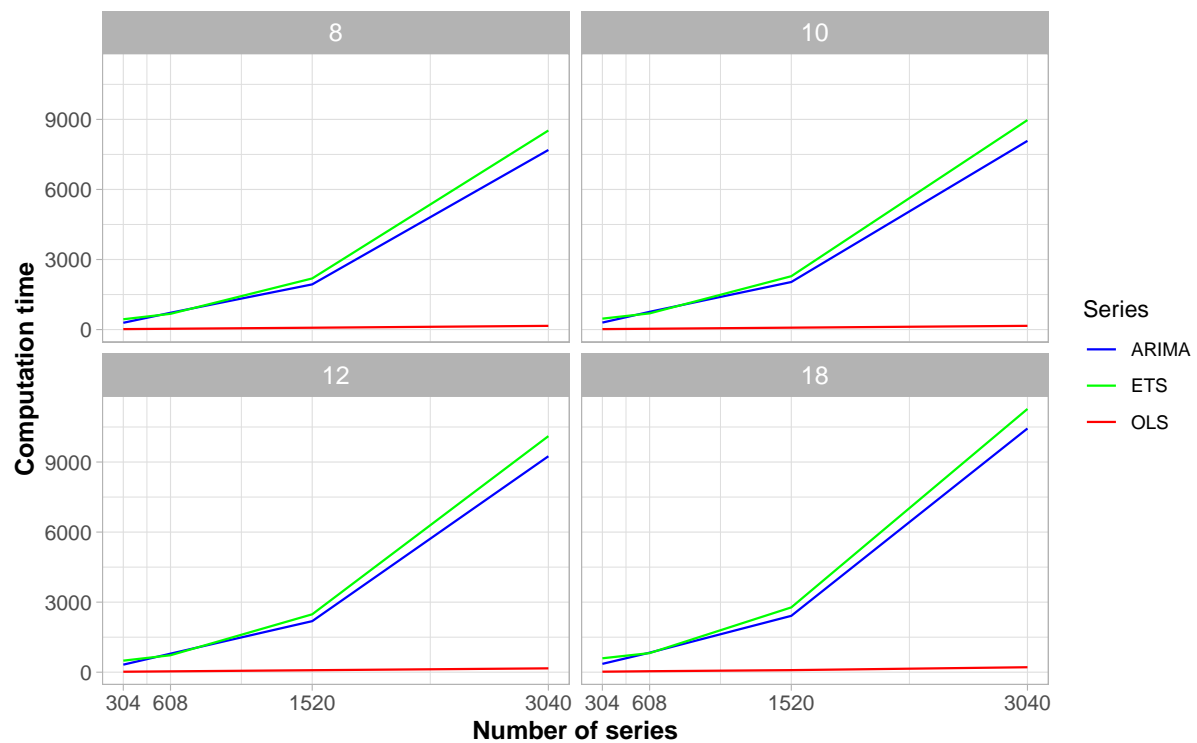


Figure 18: Computation time (seconds) for fixed origin reconciled forecasts using ETS, ARIMA and OLS, by number of bottom-level series (x-axis), and by levels of hierarchy (panels). Simulated series has error value 0.5 and 24 months test set.

Table 13: Mean RMSE by number of hierarchy levels, number of bottom-level series, method, with/without reconciliation. Based on rolling origin forecasts for a 24-month horizon, with error value 0.5.

| Reconciliation | Series | Levels | ETS | ARIMA | OLS |
|----------------|--------|--------|-------|-------|-------|
| rec | 304 | 8 | 149.9 | 146.3 | 153.2 |
| rec | 304 | 10 | 164.6 | 160.5 | 168.1 |
| rec | 304 | 12 | 176.6 | 172.3 | 180.2 |
| rec | 304 | 18 | 204.6 | 208.4 | 207.1 |
| rec | 608 | 8 | 196.8 | 194.6 | 205.5 |
| rec | 608 | 10 | 262.5 | 262.5 | 225.3 |
| rec | 608 | 12 | 231.0 | 229.6 | 241.2 |
| rec | 608 | 18 | 283.3 | 280.1 | 294.5 |
| rec | 1520 | 8 | 288.0 | 284.9 | 301.0 |
| rec | 1520 | 10 | 314.2 | 312.1 | 329.3 |
| rec | 1520 | 12 | 336.4 | 334.5 | 352.4 |
| rec | 1520 | 18 | 432.0 | 427.1 | 450.5 |
| rec | 3040 | 8 | 394.2 | 390.6 | 412.5 |
| rec | 3040 | 10 | 429.5 | 427.2 | 451.0 |
| rec | 3040 | 12 | 459.7 | 457.6 | 482.6 |
| rec | 3040 | 18 | 600.9 | 595.6 | 628.7 |
| unrec | 304 | 8 | 156.5 | 153.7 | 154.5 |
| unrec | 304 | 10 | 172.3 | 168.9 | 169.7 |
| unrec | 304 | 12 | 185.1 | 180.9 | 181.8 |
| unrec | 304 | 18 | 212.9 | 199.6 | 208.9 |
| unrec | 608 | 8 | 203.0 | 202.3 | 205.0 |
| unrec | 608 | 10 | 265.4 | 263.5 | 224.8 |
| unrec | 608 | 12 | 238.4 | 237.6 | 240.5 |
| unrec | 608 | 18 | 291.4 | 286.8 | 293.8 |
| unrec | 1520 | 8 | 298.1 | 296.4 | 300.1 |
| unrec | 1520 | 10 | 325.2 | 324.2 | 328.4 |
| unrec | 1520 | 12 | 347.7 | 346.2 | 351.2 |
| unrec | 1520 | 18 | 445.0 | 436.5 | 449.0 |
| unrec | 3040 | 8 | 407.9 | 406.4 | 411.2 |
| unrec | 3040 | 10 | 444.9 | 443.7 | 449.5 |
| unrec | 3040 | 12 | 475.9 | 473.4 | 480.7 |
| unrec | 3040 | 18 | 618.4 | 608.5 | 626.4 |

Table 14: Mean RMSE by number of hierarchy levels, number of bottom-level series, method, with/without reconciliation. Simulated series has error value 0.5. Forecasting uses fixed origin for a 24-month horizon.

| Reconciliation | Series | Levels | ETS | ARIMA | OLS |
|----------------|--------|--------|-------|-------|-------|
| rec | 304 | 8 | 154.5 | 151.7 | 158.2 |
| rec | 304 | 10 | 169.6 | 166.4 | 174.1 |
| rec | 304 | 12 | 181.8 | 178.2 | 186.4 |
| rec | 304 | 18 | 209.8 | 204.9 | 213.4 |
| rec | 608 | 8 | 199.1 | 199.3 | 213.4 |
| rec | 608 | 10 | 217.8 | 219.4 | 235.4 |
| rec | 608 | 12 | 233.1 | 235.9 | 252.2 |
| rec | 608 | 18 | 286.1 | 288.5 | 312.6 |
| rec | 1520 | 8 | 291.6 | 291.8 | 312.9 |
| rec | 1520 | 10 | 317.8 | 320.1 | 345.1 |
| rec | 1520 | 12 | 340.2 | 343.4 | 369.7 |
| rec | 1520 | 18 | 435.4 | 440.1 | 479.9 |
| rec | 3040 | 8 | 399.1 | 400.0 | 429.1 |
| rec | 3040 | 10 | 434.6 | 438.0 | 473.2 |
| rec | 3040 | 12 | 465.1 | 469.5 | 507.0 |
| rec | 3040 | 18 | 607.0 | 613.6 | 670.9 |
| unrec | 304 | 8 | 161.0 | 156.6 | 159.2 |
| unrec | 304 | 10 | 177.1 | 172.9 | 175.4 |
| unrec | 304 | 12 | 189.9 | 185.4 | 187.7 |
| unrec | 304 | 18 | 217.9 | 212.5 | 214.7 |
| unrec | 608 | 8 | 207.0 | 209.5 | 213.8 |
| unrec | 608 | 10 | 227.0 | 230.6 | 235.8 |
| unrec | 608 | 12 | 243.7 | 247.1 | 252.6 |
| unrec | 608 | 18 | 299.3 | 299.0 | 312.7 |
| unrec | 1520 | 8 | 302.2 | 307.5 | 313.5 |
| unrec | 1520 | 10 | 330.7 | 336.9 | 345.6 |
| unrec | 1520 | 12 | 353.4 | 360.1 | 370.2 |
| unrec | 1520 | 18 | 449.8 | 456.0 | 480.0 |
| unrec | 3040 | 8 | 414.6 | 421.8 | 429.9 |
| unrec | 3040 | 10 | 453.7 | 461.3 | 473.9 |
| unrec | 3040 | 12 | 484.9 | 492.6 | 507.7 |
| unrec | 3040 | 18 | 628.3 | 635.7 | 670.9 |

4 Conclusion

We have proposed a linear model approach to fast forecasting of hierarchical or grouped time series, with accuracy that nearly matches that of forecast methods such as ETS and ARIMA. This is especially useful in large collections of time series, as is typical in hierarchical and grouped structures. Although ETS and ARIMA are advantageous in terms of forecasting power and accuracy, they can be computationally heavy when facing large collections of time series in the hierarchy.

An important feature of our proposed OLS model is its ability to easily include external information such as holiday dummies or other external series. We also note that OLS has the additional practical advantage of handling missing data while ETS and ARIMA require imputation.

Pennings & Dalen (2017) proposed another approach for forecasting hierarchical time series using state space models. Although their approach is flexible in handling outliers, missing data and external features, it is less flexible across different kinds of datasets and is computationally much more demanding.

Another advantage of our approach is that it can be computed in a single matrix equation (5). This makes it extremely fast and easy to implement, and enables standard results to be derived with minimal effort (e.g., prediction intervals).

In this paper, we computed prediction intervals based on a normality assumption for all the real and simulated examples. One future direction for potentially improving prediction intervals is applying bootstrap-based methods which only assume that the forecast errors are uncorrelated.

Acknowledgements

Ashouri and Shmueli were partially funded by Ministry of Science and Technology (MOST), Taiwan [Grant 106-2420-H-007-019]. Hyndman's research is supported by the Australian Center of Excellence in Mathematical and Statistical Frontiers.

A Appendix A

Figures 19 and 20 display boxplots of the scaled forecast errors for the tourism example. These plots are displayed for both rolling forward and multiple-step-ahead forecasts, respectively. We see that scaled error magnitudes are similar at different hierarchy / grouping levels.

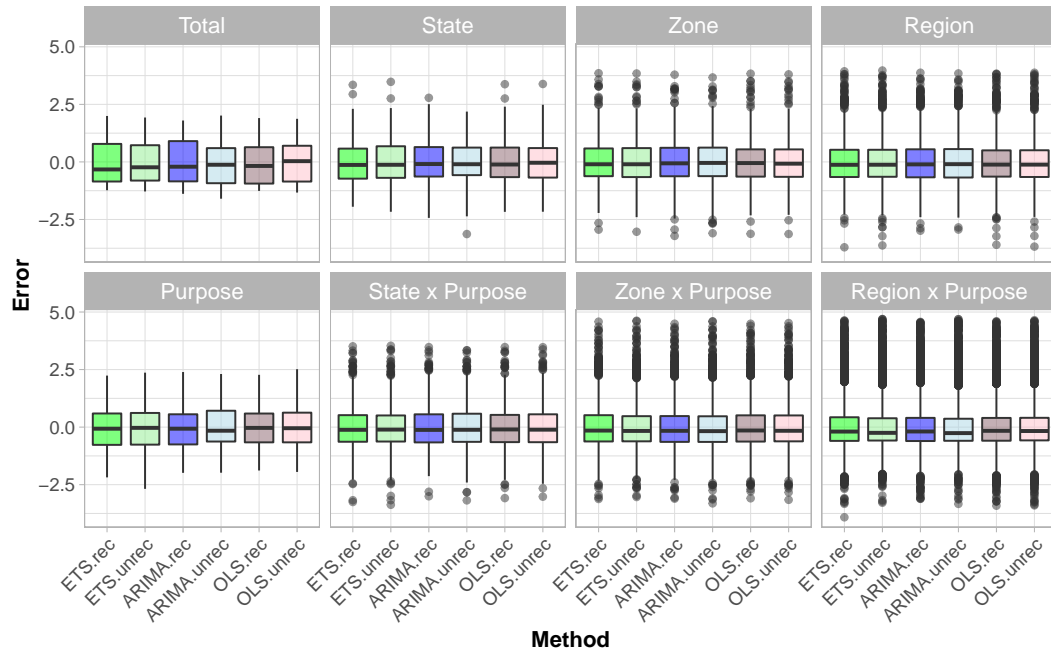


Figure 19: Box plots of scaled forecast errors from reconciled and unreconciled ETS, ARIMA, and OLS rolling origin forecasts for tourism demand data. Panels are different hierarchy levels.

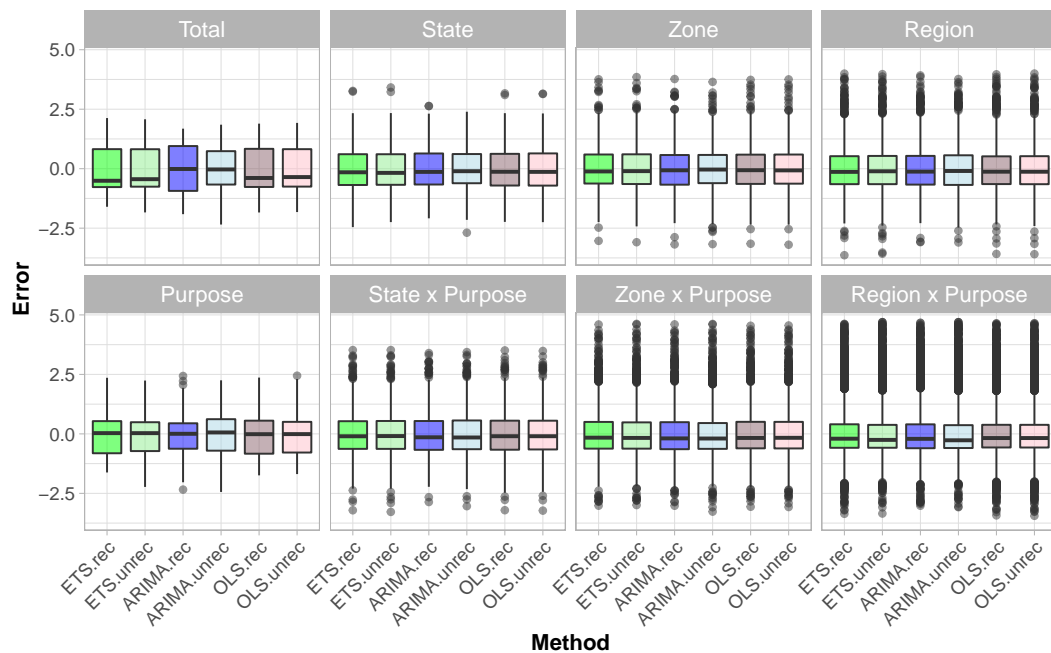


Figure 20: Box plots of scaled forecast errors from reconciled and unreconciled ETS, ARIMA, and OLS fixed origin forecasts for tourism demand data. Panels are different hierarchy levels.

Tables 15 and 16 display the reconciliation results using two other reconciliation matrices introduced in (Wickramasuriya, Athanasopoulos & Hyndman 2019): shrinkage estimator (mint_shrink) and variance scaling (wls_var). These results are presented for both rolling origin (Table 15) and fixed origin forecasts (Table 16). We see that in this case the shrinkage estimator provides the best reconciled forecasts.

Table 15: Comparing Mean(RMSE) of three different reconciliation matrices for rolling origin forecasts on a 24 months test set.

| Level | mint_shrink | | | wls_var | | | wls_struct | | |
|------------------|-------------|-------|------|---------|-------|------|------------|-------|------|
| | ETS | ARIMA | OLS | ETS | ARIMA | OLS | ETS | ARIMA | OLS |
| Total | 1666 | 1637 | 1834 | 1787 | 1938 | 1910 | 1733 | 1840 | 1864 |
| State | 492 | 460 | 500 | 499 | 491 | 511 | 497 | 482 | 509 |
| Zone | 208 | 202 | 209 | 210 | 210 | 213 | 211 | 210 | 213 |
| Region | 117 | 116 | 115 | 117 | 119 | 116 | 118 | 120 | 117 |
| Purpose | 666 | 655 | 703 | 683 | 734 | 720 | 673 | 713 | 713 |
| State x Purpose | 207 | 204 | 209 | 208 | 209 | 213 | 208 | 209 | 213 |
| Zone x Purpose | 96 | 96 | 96 | 96 | 98 | 97 | 96 | 99 | 97 |
| Region x Purpose | 56 | 56 | 55 | 56 | 57 | 55 | 56 | 57 | 56 |

Table 16: Comparing Mean(RMSE) of three different reconciliation matrices for fixed origin forecasts on a 24 months test set.

| Level | mint_shrink | | | wls_var | | | wls_struct | | |
|------------------|-------------|-------|------|---------|-------|------|------------|-------|------|
| | ETS | ARIMA | OLS | ETS | ARIMA | OLS | ETS | ARIMA | OLS |
| Total | 2444 | 2721 | 2506 | 2541 | 2738 | 2880 | 2492 | 2744 | 2819 |
| State | 567 | 580 | 569 | 577 | 585 | 618 | 573 | 583 | 612 |
| Zone | 232 | 233 | 228 | 235 | 236 | 244 | 237 | 234 | 243 |
| Region | 125 | 126 | 120 | 125 | 127 | 126 | 127 | 127 | 126 |
| Purpose | 819 | 868 | 849 | 836 | 898 | 936 | 822 | 889 | 921 |
| State x Purpose | 221 | 224 | 224 | 223 | 226 | 238 | 222 | 226 | 236 |
| Zone x Purpose | 101 | 102 | 100 | 101 | 103 | 104 | 102 | 102 | 104 |
| Region x Purpose | 58 | 58 | 56 | 58 | 58 | 58 | 58 | 58 | 58 |

In Figure 21, we compare the prediction intervals for reconciled and unreconciled forecasts for rolling and fixed origin forecasts. We can see the effect of the reconciliation on improving the forecasts' standard deviations (i.e. narrower prediction intervals).

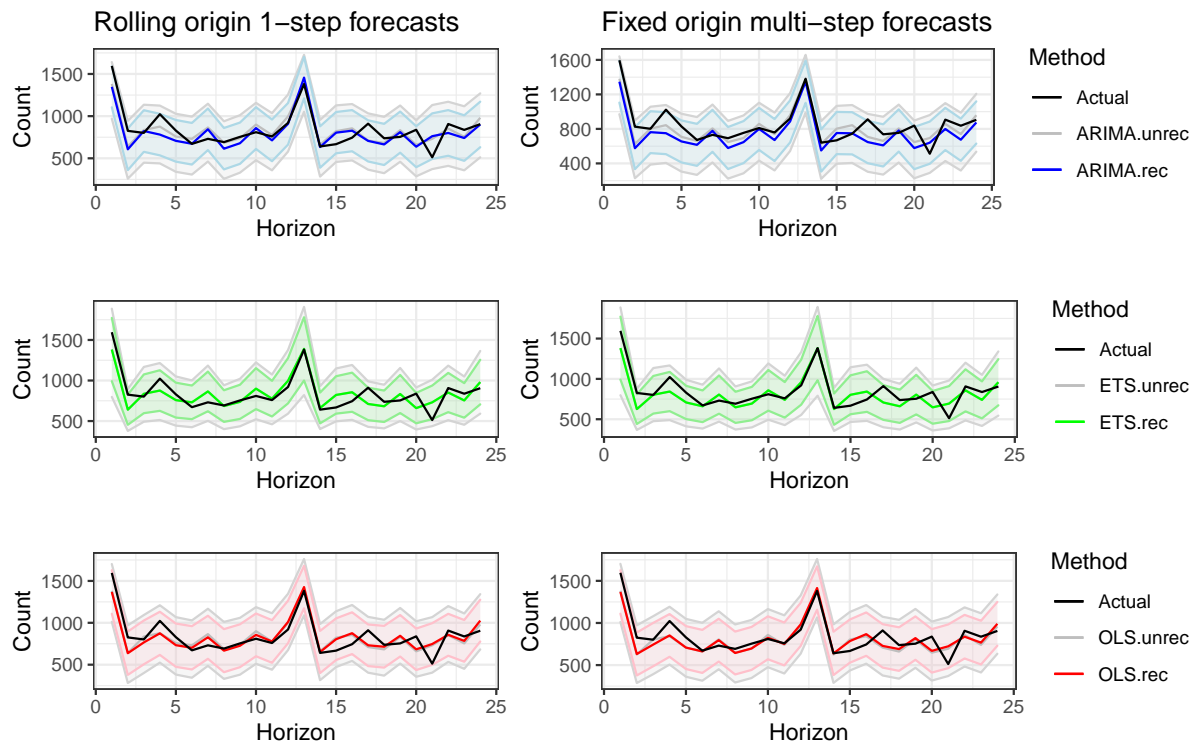


Figure 21: Comparing forecasts and prediction intervals of 'AAAVis' bottom-level series across methods. Left: rolling origin. Right: fixed origin.

B Appendix B: Wikipedia pageviews: Grouped structure

We illustrate our method on another real dataset. The Wikipedia dataset includes 12 months of daily data (2016-06-01 to 2017-06-29) on Wikipedia pageviews for the most popular social networks articles (Ashouri, Shmueli & Sin 2018). This dataset is noisier than the Australian monthly tourism data, making forecasting more challenging. The data has a grouped structure with the following attributes (see Table 17):

- *Agent*: Spider, User;
- *Access*: Desktop, Mobile app, Mobile web;
- *Language*: en (English), de (German), es (Spanish), zh (Chinese); and
- *Purpose*: Blogging related, Business, Gaming, General purpose, Life style, Photo sharing, Reunion, Travel, Video.

Figure 22 shows one possible hierarchy for this dataset, but the order of the hierarchy can be switched.

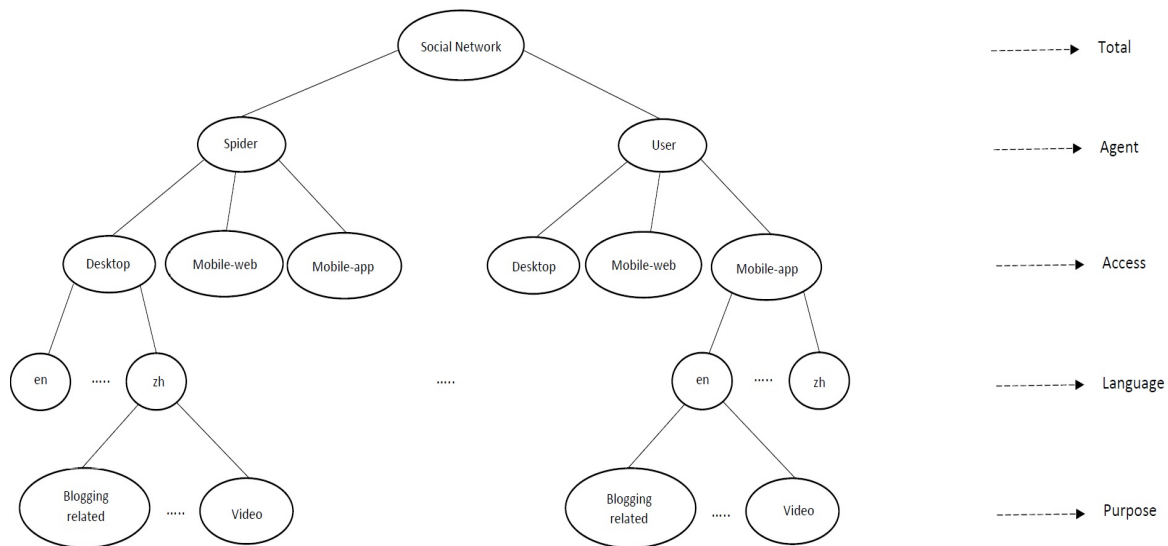


Figure 22: One possible hierarchical structure for the Wikipedia pageviews dataset.

Table 17: Social networking Wikipedia article grouping structure

| Grouping | Series | Grouping | Series |
|----------|-------------------|----------|----------------------|
| Total | | Language | |
| | 1. Social Network | | 10. zh (Chinese) |
| Access | | Purpose | |
| | 2. Desktop | | 11. Blogging related |
| | 3. Mobile app | | 12. Business |
| Agent | | | 13. Gaming |
| | 4. Mobile web | | 14. General purpose |
| | 5. Spider | | 15. Life style |
| | 6. User | | 16. Photo sharing |
| Language | | | 17. Reunion |
| | 7. en (English) | | 18. Travel |
| | 8. de (German) | | 19. Video |
| | 9. es (Spanish) | | |

We consider the main aggregation factors and their two-way combinations.¹⁰ The final dataset includes 913 time series, each with length 394. Table 18 shows the group structure's different levels and the number of series in each level.

For this daily dataset, in the OLS forecasting model we include in the predictor matrix the following terms: a quadratic trend, 6 seasonal dummies, and lags 1 and 7 for rolling and fixed

¹⁰First, while we present results for single groups, we applied all the two-way combinations in the reconciliation step. Second, there are four more 3-way aggregation combinations that we do not include: Agent \times Access \times Language, Agent \times Access \times Purpose, Agent \times Language \times Purpose, and Access \times Language \times Purpose. Including these four additional aggregations might slightly improve the results but for simplicity, we excluded them.

Table 18: *Number of Wikipedia pageviews series at each aggregation level.*

| Division | Series |
|--------------------|--------|
| Total pageviews | 1 |
| Access | 3 |
| Agent | 2 |
| Language | 4 |
| Purpose | 9 |
| Access x Agent | 5 |
| Access x Language | 12 |
| Access x Purpose | 27 |
| Agent x Language | 8 |
| Agent x Purpose | 18 |
| Language x Purpose | 33 |
| Bottom level | 913 |
| Total | 1035 |

origin models. We partitioned the data into training and test sets. We used the last 28 days for our test set and the rest for the training set.

Tables 19 and 20 show the RMSE results. Although these time series are noisier, we still get acceptable results for the OLS forecasting model compared with ETS and ARIMA. In this case, we get similar results with and without the reconciliation step.

Table 19: *Mean RMSE for ETS, ARIMA and OLS with and without reconciliation - Rolling origin - Wikipedia dataset*

| Level | Unreconciled | | | Reconciled | | |
|--------------|--------------|---------|---------|------------|---------|---------|
| | ETS | ARIMA | OLS | ETS | ARIMA | OLS |
| Total | 10773.7 | 15019.6 | 12968.1 | 12763.6 | 13341.0 | 11959.0 |
| Access | 6524.7 | 6666.3 | 6021.1 | 6794.9 | 6827.8 | 6099.5 |
| Agent | 8272.9 | 10282.7 | 9372.2 | 8578.0 | 8989.7 | 8503.6 |
| Language | 4870.1 | 6279.3 | 5688.2 | 6132.9 | 5901.8 | 5447.5 |
| Purpose | 5233.5 | 4672.5 | 4106.9 | 4541.4 | 4149.7 | 3805.2 |
| Bottom level | 358.1 | 239.7 | 261.5 | 363.8 | 241.6 | 262.8 |

Figures 23 and 24 display the forecast error box plot. These plots are for rolling and fixed origin forecasts over 28 days in each level of grouping. We see that the error distribution is similar in all levels across the different methods. The only exception is the Total series, where ETS performs significantly better than ARIMA and OLS. We also note that the reconciliation is less effective. As in the tourism example, in higher levels series have higher counts and therefore their error magnitudes are larger.

In Figures 25 and 26, we display results for the total and one of the bottom level series, “desktopuserPho04” (desktop-user-english-photo sharing). The plot shows rolling and fixed origin

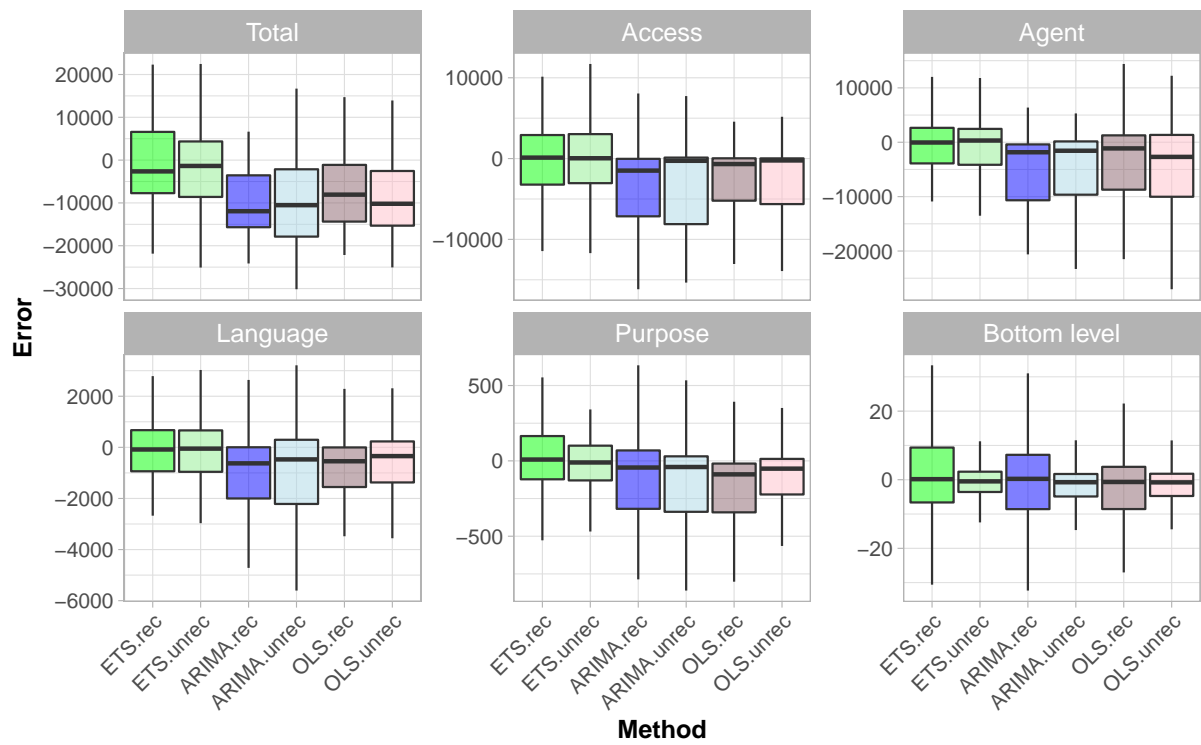


Figure 23: Box plots of forecast errors for reconciled and unreconciled ETS, ARIMA and OLS methods at each hierarchy level for rolling origin forecasts of Wikipedia pageviews.

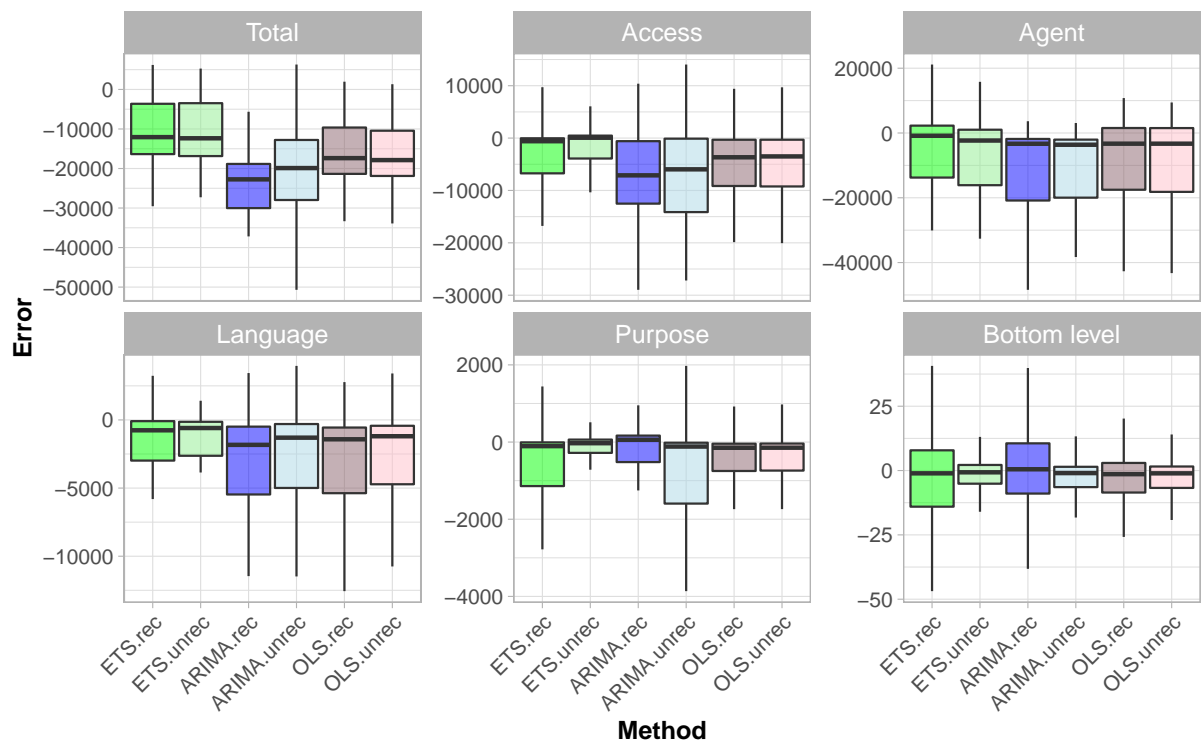


Figure 24: Box plots of forecast errors for reconciled and unreconciled ETS, ARIMA and OLS methods at each hierarchy level for fixed origin forecasts of Wikipedia pageviews.

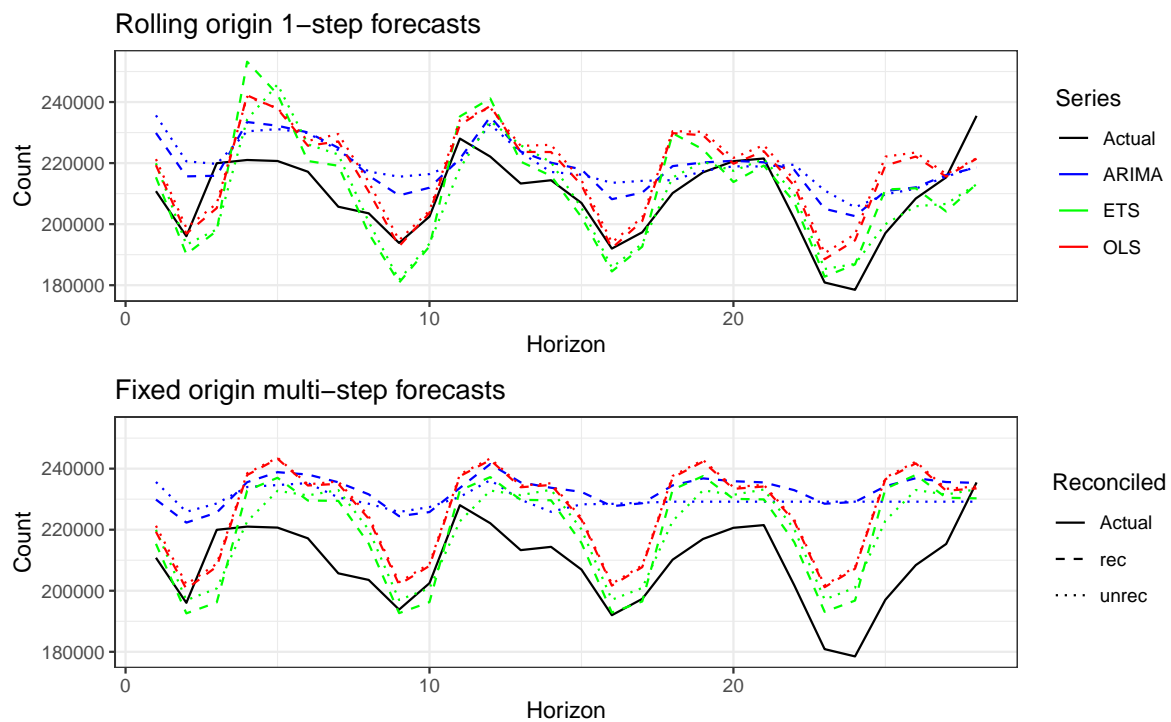


Figure 25: Comparing reconciled and unreconciled ETS, ARIMA and OLS rolling and fixed origin forecasts for Wikipedia pageviews 'Total' Series.

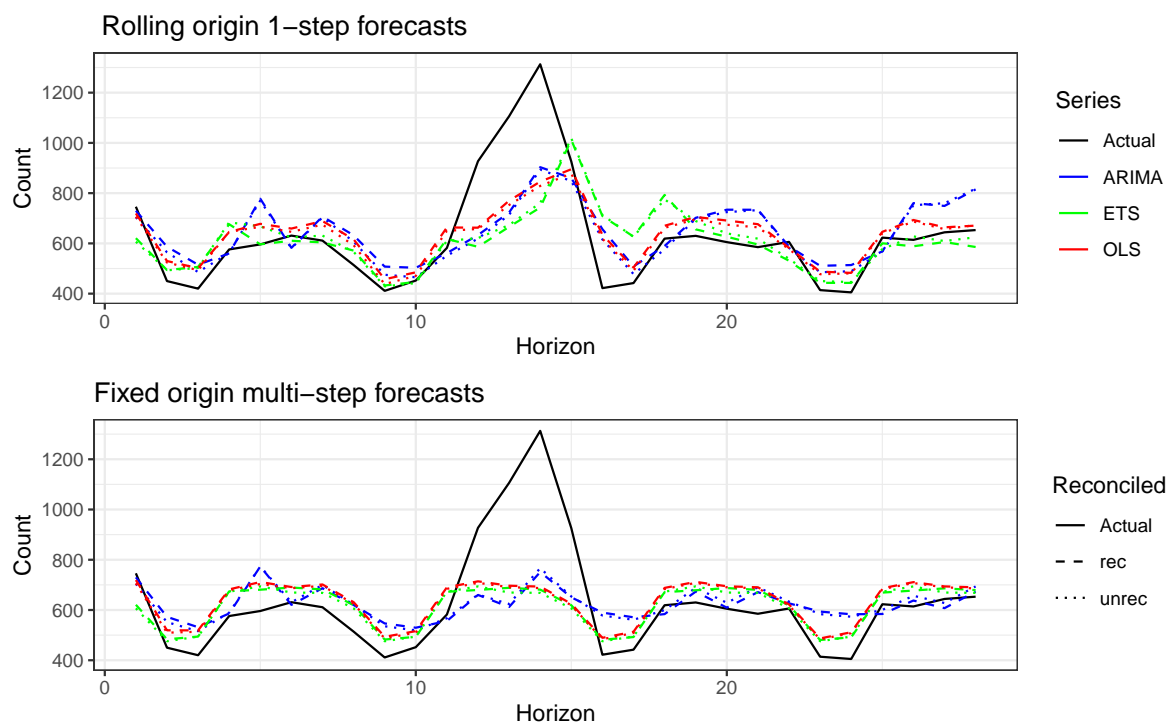


Figure 26: Comparing reconciled and unreconciled ETS, ARIMA and OLS rolling and fixed origin forecasts for Wikipedia pageviews 'desktopusenPho04' bottom-level series

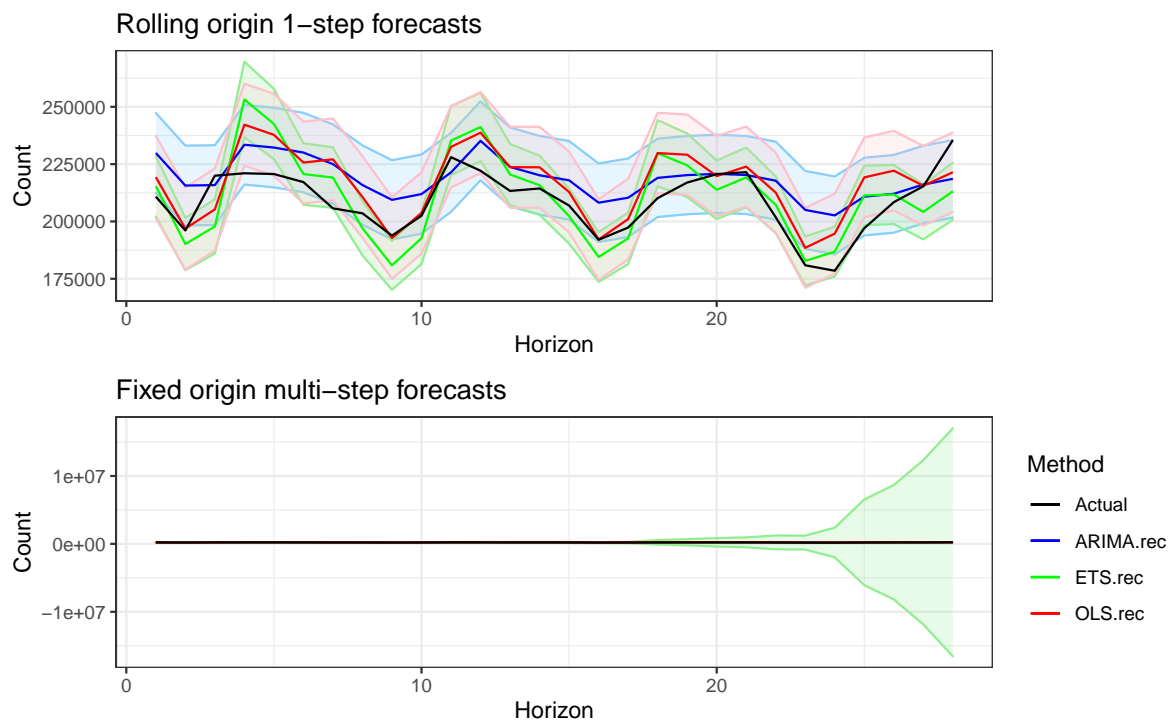


Figure 27: The actual test set for the 'Total series' compared to the forecasts from reconciled and unreconciled OLS methods with prediction interval for rolling and fixed origin Wikipedia pageviews.

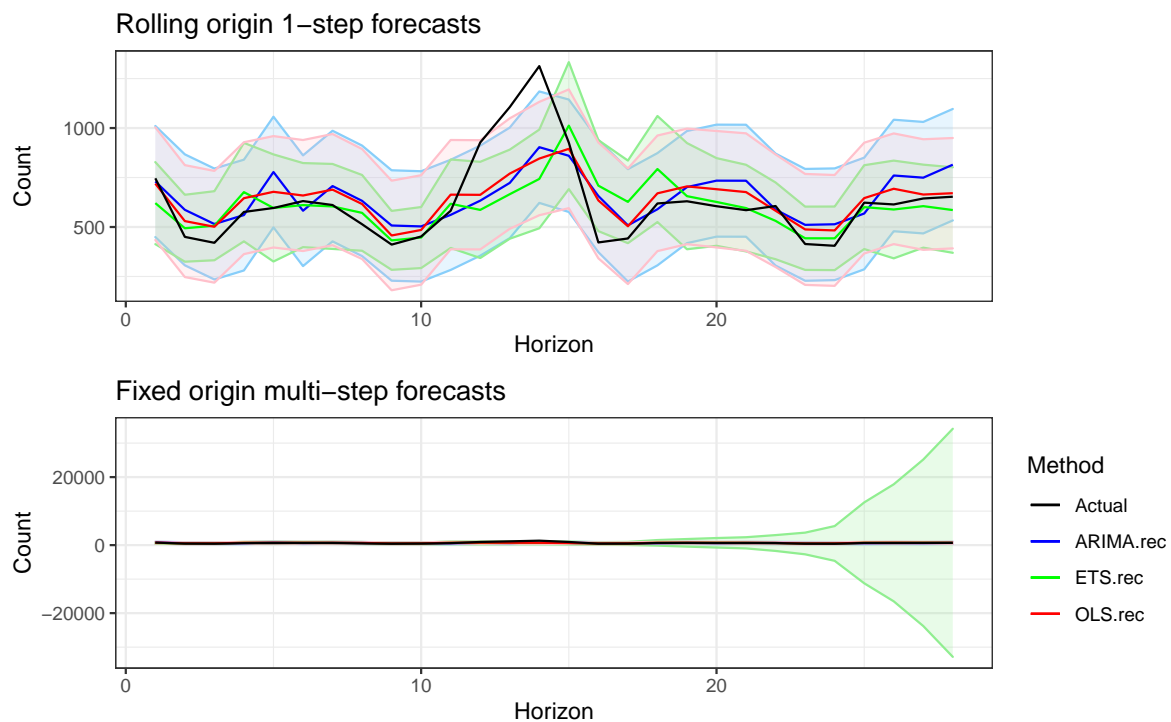


Figure 28: The actual test set for the 'desktopusenPho04' bottom level series compared to the forecasts from reconciled and unreconciled OLS methods with prediction interval for rolling and fixed origin Wikipedia pageviews.

Table 20: Mean RMSE for ETS, ARIMA and OLS with and without reconciliation - Fixed origin - Wikipedia dataset

| Level | Unreconciled | | | Reconciled | | |
|--------------|--------------|---------|---------|------------|---------|---------|
| | ETS | ARIMA | OLS | ETS | ARIMA | OLS |
| Total | 14846.9 | 24298.8 | 20203.7 | 15787.6 | 26193.9 | 19691.1 |
| Access | 7117.4 | 10722.8 | 8866.4 | 8520.2 | 11532.4 | 8853.3 |
| Agent | 13608.7 | 18168.8 | 14985.7 | 12130.3 | 17639.5 | 14580.2 |
| Language | 6475.9 | 9527.0 | 7913.7 | 6792.9 | 10783.0 | 8022.2 |
| Purpose | 5302.7 | 8638.5 | 5694.1 | 5141.4 | 7536.7 | 5541.8 |
| Bottom level | 436.9 | 388.0 | 366.2 | 440.8 | 389.6 | 365.9 |

forecast results over the 28 day test set for ETS, ARIMA and OLS, with (dashed lines) and without (dotted lines) applying reconciliation. We see that the OLS forecasting model performs close to the other two methods, and reconciliation improves the forecasts.

Table 21 presents the computation times for all three methods. ETS and ARIMA are clearly much more computationally heavy compared with OLS. As in the Australian tourism dataset, running reconciliation does not have much effect on computation time.

Table 21: Computation time (seconds) for ETS, ARIMA and OLS with reconciliation - Rolling and fixed origin forecasts - Wikipedia dataset

| | Rolling origin | Fixed origin |
|-------|----------------|--------------|
| ETS | 22592 | 882 |
| ARIMA | 20016 | 1682 |
| OLS | 116 | 61 |

References

- Akaike, H (1998). “Information theory and an extension of the maximum likelihood principle”. In: *Selected Papers of Hirotugu Akaike*. Springer Series in Statistics (Perspectives in Statistics). Springer, pp.199–213.
- Ashouri, M, G Shmueli & CY Sin (2018). Clustering time series by domain-relevant features using model-based trees. *Proceedings of the 2018 Data Science, Statistics & Visualization (DSSV)*.
- Athanasopoulos, G, RA Ahmed & RJ Hyndman (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting* **25**(1), 146–166.
- Christensen, R (2020). *Plane answers to complex questions: the theory of linear models*. 5th. Springer.
- Fliedner, G (2001). Hierarchical forecasting: issues and use guidelines. *Industrial Management & Data Systems* **101**(1), 5–12.
- Gross, CW & JE Sohl (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting* **9**(3), 233–254.
- Hyndman, RJ, RA Ahmed, G Athanasopoulos & HL Shang (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis* **55**(9), 2579–2589.
- Hyndman, RJ & G Athanasopoulos (2018). *Forecasting: principles and practice*. Melbourne, Australia: OTexts. <https://OTexts.org/fpp2>.
- Januschowski, T, S Kolassa, M Lorenz, C Schwarz, et al. (2013). Forecasting with in-memory technology. *Foresight: The International Journal of Applied Forecasting* **31**, 14–20.
- Kahn, KB (1998). Revisiting top-down versus bottom-up forecasting. *The Journal of Business Forecasting* **17**(2), 14.
- O Hara-Wild, M, R Hyndman & E Wang (2019). *fable: Forecasting Models for Tidy Time Series*. R package version 0.1. 0.
- Pennings, CL & J van Dalen (2017). Integrated hierarchical forecasting. *European Journal of Operational Research* **263**(2), 412–418.
- Tourism Research Australia (2005). Travel by Australians, September Quarter 2005. *Tourism Australia*.
- Wickramasuriya, SL, G Athanasopoulos & RJ Hyndman (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association* **114**(526), 804–819.