

Optimally Reconciling Forecasts in a Hierarchy

ROB J. HYNDMAN AND GEORGE ATHANASOPOULOS

PREVIEW We know that when forecasting the product hierarchy, reconciliation is invariably needed to make the sum of lower-level forecasts equate to the upper-level forecasts. The authors argue that the traditional Bottom-Up, Top-Down, and Middle-Out procedures for reconciliation all fail to make best use of the available data. They show we can do better by taking weighted averages of the forecasts from different levels, an approach they call optimal reconciliation.

HIERARCHICAL STRUCTURES

Time series can often be naturally disaggregated using attributes such as product type. The total number of bicycles sold by a cycling warehouse, for example, can be disaggregated into a hierarchy of bicycle types. Such a warehouse will sell road bikes, mountain bikes, children's bikes, and hybrids. Each of these can be disaggregated further into finer categories. Children's bikes can be divided into balance bikes for children under 4 years old, single-geared bikes for children between 4 and 6, and multi-geared bikes for children over the age of 6. Hybrid bikes can be divided into city, commuting, comfort, and trekking bikes; and so on. Such disaggregation imposes a *hierarchical structure* and we refer to these as *hierarchical time series*.

To take another example, a manufacturing company may track sales by country, and within a country by region, and within each region by sales outlet. Again, this is a hierarchical structure, and the complete collection of sales data can be thought of as hierarchical time series.

We get a similar structure when the two attributes above are combined; for example, a bicycle manufacturer may disaggregate sales both by product and by geographical location. In this case, we still have a hierarchical structure; however, the structure does not disaggregate in a unique way. We can disaggregate by product type and then by geographical location, but also vice versa. We usually refer to these structures as “grouped time series.”

The *optimal reconciliation method* we discuss here can handle hierarchical and grouped time series. In contrast, traditional forecasting approaches completely ignore the structures of grouped time series.

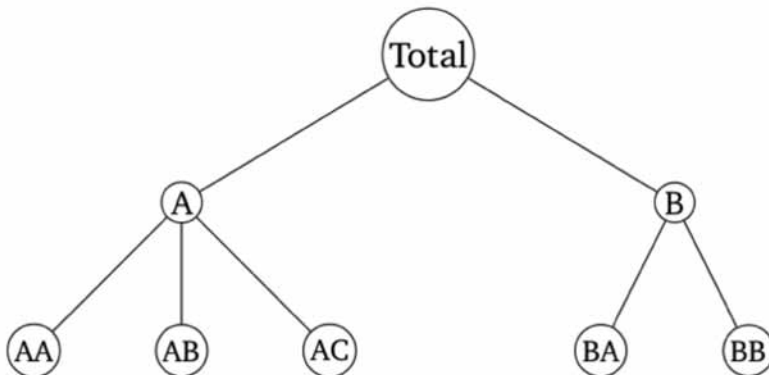
Figure 1 shows a very simple hierarchy. At the top of the hierarchy is the “Total,” the most aggregate level of the data. This total is split into two categories, A and B, which are in turn split into three and two subcategories respectively.

If the letters now stand for the values in each node, then

$$AA + AB + AC = A, \quad BA + BB = B \quad \text{and} \quad A + B = \text{Total}.$$

While the data within the hierarchical structure will add up appropriately, the forecasts generally will not. Consider the first level of the hierarchy in Figure 1. If we forecast the time series for each node independently, the forecasts for A and B will not necessarily add up to the forecasts for Total. This can cause confusion when making decisions based on forecasts (e.g. how much stock of different types to ship, or how much money to allocate to different divisions of a company). Consequently, it is standard to require hierarchical forecasts to add up

Figure 1. A Two-Level Hierarchical Tree



appropriately (a feature we call “aggregate consistency”): the sum of the forecasts must equal the forecast of the sum.

TRADITIONAL FORECASTING APPROACHES

One common method for hierarchical forecasting is *bottom-up*. In this approach, we first generate independent forecasts for each series at the bottom level of the hierarchy and then aggregate these to produce forecasts for the upper levels of the hierarchy.

For example, for the hierarchy of Figure 1, we first generate forecasts for the bottom-level series:

$$\widehat{AA}, \widehat{AB}, \widehat{AC}, \widehat{BA} \text{ and } \widehat{BB}.$$

Aggregating these up the hierarchy, we get forecasts for the rest of the series. We add

$$\widehat{AA} + \widehat{AB} + \widehat{AC}$$

to obtain \widehat{A} ,

$$\widehat{BA} + \widehat{BB}$$

to obtain \widehat{B} , and

$$\widehat{A} + \widehat{B}$$

to obtain \widehat{Total} .

An alternative is to start with forecasting the most aggregated series (the Total), and then disaggregate the forecasts down the hierarchy. This is known as a *top-down* approach. There are several possible ways to disaggregate; the most common is to disaggregate by the average proportions computed on the historical data. For example, if A has averaged 65% of the Total, we would multiply the forecast for the Total by 65% to obtain \widehat{A} .

Because historical proportions used for disaggregation do not take account of how those proportions may change over time, top-down approaches based on historical proportions tend to produce less accurate forecasts at lower levels of the hierarchy than bottom-up approaches. To address this issue, proportions based on forecasts rather than historical data can be used. As described by Athanasopoulos and colleagues (2009), we would disaggregate the relative to the sum of the forecasts \widehat{A} and \widehat{B}

A third approach that is widely used in practice is *middle-out* forecasting. A middle level

Key Points

- In a product or geographic hierarchy, the sum of forecasts at a granular level cannot be expected to equal the forecast at the group level, hence reconciliation is necessary. Traditional reconciliation methods include bottom-up, top-down, and middle-out.
- Major problems exist with these traditional methods, such as failure to utilize all available information and the lack of any straightforward means to determine an optimal reconciliation approach.
- We propose an *optimal combination* of the forecasts made at all levels of the hierarchy that gives the best possible reconciled forecasts. This approach uses more of the available information than the traditional approaches and, in empirical comparisons, tends to give more accurate forecasts than any of the bottom-up, top-down, or middle-out methods.
- We present an illustrative example of a hierarchy of tourism in different regions and subregions, and show in an appendix the requisite calculations using the R statistical programming language.

of the hierarchy is selected, and forecasts are computed for each node at that level. Forecasts for levels above are computed by aggregation, and forecasts for levels below by disaggregation. For example, in the hierarchy of Figure 1, forecasts may be calculated for A and B (the middle level). Then the Total forecast is obtained by adding these two forecasts, and the forecasts for the bottom level would be obtained by disaggregating the A and B forecasts, as in a top-down approach.

Until recently, these three approaches (bottom-up, top-down, and middle-out) were the only possible ways of ensuring that forecasts across the hierarchy added up appropriately.

OPTIMAL RECONCILIATION

There is now another possible approach to handling hierarchical time series. As suggested by Hyndman and colleagues (2011), we can generate independent forecasts for all nodes, initially ignoring the fact that they will not add up. Then an *optimal reconciliation* step is used to adjust the independent forecasts, giving revised forecasts that do add up in a way consistent with the hierarchical structure. The revised forecast at each node will be a weighted average of the forecasts from all nodes.

Advantages

Optimal reconciliation is based on a combination of the independent forecasts that gives the best possible reconciled forecasts, taking account of all forecasts at all levels. It uses more of the available information within a hierarchy than any other existing method, allows for correlations and interactions between series at each level of the hierarchy, and accounts for ad hoc adjustments of forecasts at any level. Provided that the base forecasts are unbiased, optimal reconciliation produces unbiased revised forecasts. In our empirical comparisons, the optimal-reconciliation approach tends to give more accurate forecasts than any of the bottom-up, top-down, or middle-out methods.

Difficulties

The major difficulty with the optimal reconciliation approach lies in the computation of the *reconciliation weights* used to form a weighted average of forecasts at an individual node. The weights are obtained using linear regression, where all the independent forecasts (from all nodes) are regressed

against a set of dummy variables indicating which of the bottom-level series contribute to each node. They are “optimal” in the sense that the mean squared reconciliation error, computed using the differences between the reconciled and independent forecasts, is as small as possible. Thus, the weights enable us to revise the independent forecasts by the smallest possible amount until they add up.

Importantly, the weights depend *only* on the hierarchical structure and not on the actual observed data. That way, the weights can be computed once for the hierarchy, and reapplied whenever forecasts need to be reconciled.

We write the initial forecast for the time series at node X as \hat{X} (X-hat) and the revised forecast \tilde{X} (X-tilde). Then, for the hierarchy in Figure 1, the weights are given in **Table 1**.

Note again that the weights depend only on the structure of the hierarchy, regardless of the data at each node.

The revised forecast for node A, for example, is equal to a combination of all the original forecasts with weights taken from the second row of this table:

$$A = 0.310\widehat{Total} + 0.517A - 0.207B + 0.172\widehat{AA} + 0.172\widehat{AB} + 0.172\widehat{AC} - 0.103\widehat{BA} - 0.103\widehat{BB}$$

Notice that the weights can be negative. This is to effectively remove unwanted components. For example, in the above equation, the value of \widehat{Total} contains information from node B as well as node A. Consequently, terms associated with node B have negative coefficients to offset the inclusion of the Total.

Table 1. Weights for the Figure 1 Hierarchy

	Independent forecasts							
	\widehat{Total}	A	B	\widehat{AA}	\widehat{AB}	\widehat{AC}	\widehat{BA}	\widehat{BB}
Revised forecasts								
\widehat{Total}	0.586	0.310	0.276	0.103	0.103	0.103	0.138	0.138
A	0.310	0.517	-0.207	0.172	0.172	0.172	-0.103	-0.103
B	0.276	-0.207	0.483	-0.069	-0.069	-0.069	0.241	0.241
\widehat{AA}	0.103	0.172	-0.069	0.724	-0.276	-0.276	-0.034	-0.034
\widehat{AB}	0.103	0.172	-0.069	-0.276	0.724	-0.276	-0.034	-0.034
\widehat{AC}	0.103	0.172	-0.069	-0.276	-0.276	0.724	-0.034	-0.034
\widehat{BA}	0.138	-0.103	0.241	-0.034	-0.034	-0.034	0.621	-0.379
\widehat{BB}	0.138	-0.103	0.241	-0.034	-0.034	-0.034	-0.379	0.621

Figure 2. Australian Tourism Hierarchy

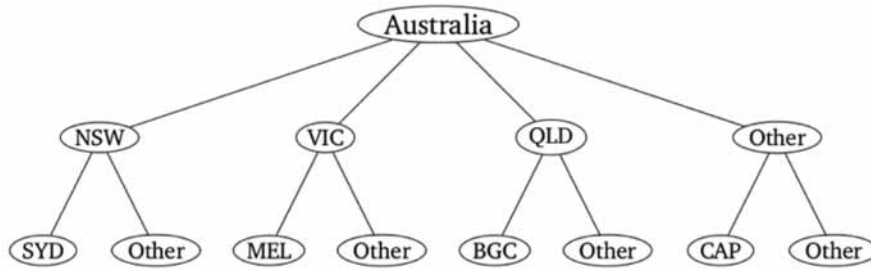
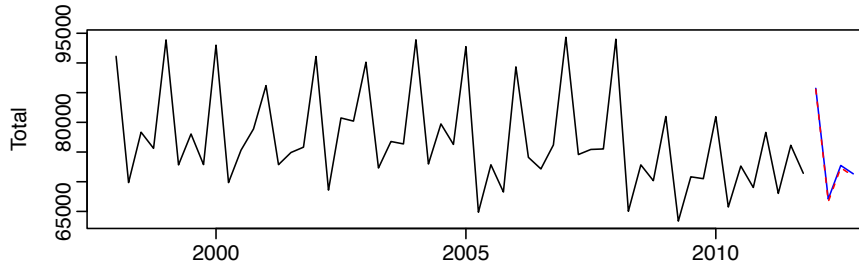


Figure 3. Total Australian Visitor Nights over the Period 1998–2012



The optimal reconciliation weights are designed so that the revised forecasts will always add up appropriately:

$$A = \widehat{AA} + \widehat{AB} + \widehat{AC} \quad B = \widehat{BA} + \widehat{BB} \quad \text{and} \quad \widehat{Total} = A + B.$$

This can easily be checked in Table 1.

The traditional approaches and the optimal-reconciliation approach are all implemented in the “hts” package for R (Hyndman and colleagues, 2014). The package also allows for automatic forecasting of time series, using either exponential smoothing or ARIMA modeling (Kolassa and Hyndman, 2010). The appendix shows the implementation instructions.

EXAMPLE: AN AUSTRALIAN TOURISM HIERARCHY

We demonstrate the optimal reconciliation of independent forecasts with an example based on forecasts of Australian domestic tourism demand, measured by the visitor nights spent away from home by Australians.

The hierarchy is shown in **Figure 2**. The first level of the hierarchy contains visitor nights for each of the three largest states of Australia: New South Wales, Victoria,

and Queensland. The remaining states are classified as “Other.” At the second level, we have the capital cities of each state: Sydney, Melbourne, Brisbane, and Gold Coast (a coastal city and a major tourism attraction near Brisbane), and the other five capital cities of the remaining states and territories of Australia. The historical data are quarterly and cover the time period 1998 to 2012.

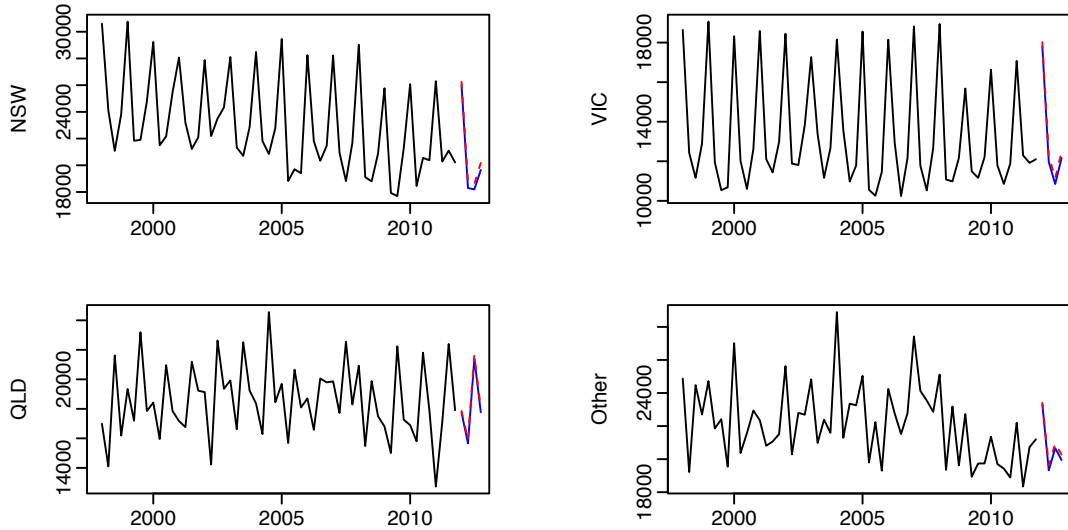
Step 1. Using the data from 1998 through 2011 — that is, holding out the quarters of 2012 to test forecast accuracy — we generate forecasts for each of the four quarters of 2012. The forecasts were generated from an automatic exponential-smoothing algorithm developed by Hyndman and colleagues (2002) and implemented in Hyndman (2014). **Figure 3** shows the historical time series and 2012 forecasts for the Total.

Step 2. We calculate the optimal reconciliation weights.

Step 3. We reconcile the forecasts using the optimal reconciliation weights.

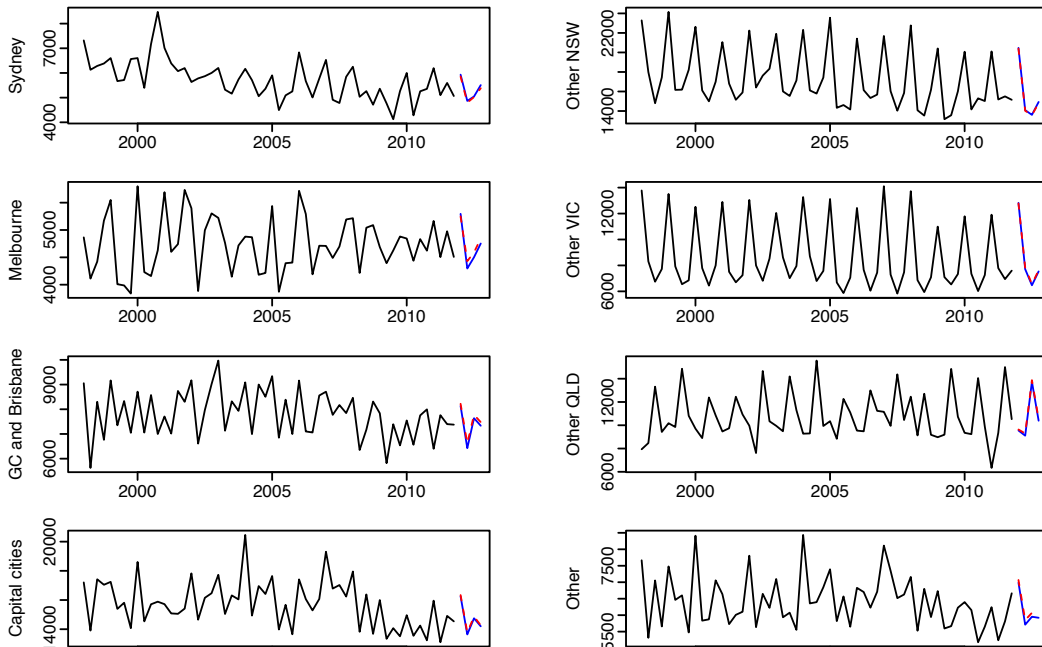
The plots in **Figures 3-5** demonstrate the difference between the initial independent forecasts and the reconciled forecasts. In **Figures 3-5**, the blue line shows the independent forecasts for the four quarters of 2012,

Figure 4: Visitor Nights by State over the Period 1998–2012.



In Figures 3-5, the blue line shows the independent forecasts for the four quarters of 2012, and the red line shows the corresponding reconciled forecasts.

Figure 5. Visitor Nights by Capital Cities over the Period 1998–2012



and the red line shows the corresponding reconciled forecasts.

The MAPE for the optimal reconciliation forecasts (over all the series and over the four quarters of 2012) is 5.35% compared to 5.58% for the bottom-up forecasts and 5.70% for the independent unreconciled forecasts.

The example demonstrates the basic idea of the optimal reconciliation approach, which is to forecast each series in the hierarchy as accurately as possible, taking advantage of the characteristics of each, and then revise these forecasts by the **smallest possible amount** to make them consistent in the aggregate. The key phrase here is *smallest possible amount*.

PROS AND CONS AND FUTURE DIRECTIONS

As we noted earlier, the most commonly applied method of hierarchical forecasting is the bottom-up approach, due to its sheer simplicity. Its greatest advantage is that we are modeling and forecasting the data at the most disaggregate level, and therefore we do not lose any information due to aggregation. However, as discussed in the accompanying article by Zotteri and colleagues (2014), the more we thin our data by disaggregation, the noisier it becomes. Furthermore, at very disaggregate levels, individual new products enter the market while others are discontinued, making it even more challenging to forecast at these levels.

The relative simplicity of the top-down approach using historical proportions is its greatest attribute. Here we only need to build one model for the most aggregate level, and we then distribute the forecasts generated from this model to lower levels. In general, top-down approaches seem to produce quite reliable forecasts for the aggregate levels, and they are very useful when the most disaggregated data consist of low counts (e.g., when demand for a product contains a lot of zeros and ones at the most disaggregated levels).

A disadvantage of this approach is the loss of information due to aggregation. Using historical proportions, we are unable to capture and take advantage of individual series characteristics such as time dynamics, special events, etc. For example, the tourism data above are highly seasonal. The seasonal pattern of visitor nights may vary across series depending on the tourism destination, e.g., a ski resort versus a beach resort. Using forecasted proportions instead of historical proportions will improve this aspect when using a top-down approach.

Hyndman and colleagues (2011) derived an important theoretical result concerning top-down forecasts. They show that no matter what proportions are used, disaggregation of forecasts inevitably introduces bias. Therefore, even if our independent forecasts are unbiased, disaggregating them to get forecasts at lower levels (as is required with

any top-down or middle-out approach) will generate biased forecasts for the lower levels. We consider this to be the greatest disadvantage of all top-down and middle-out approaches.

In contrast to all previously existing methods for forecasting hierarchical time series, optimal reconciliation allows us to take advantage of all the available data and model the dynamics of each individual series. It also can allow for correlations and interactions between time series in the hierarchy (although we did not do this in the illustrative example). We are currently developing better procedures for determining these weights, so that the scale of the original data and the correlations between series can be also accounted for.

One feature that should be highlighted here is that with grouped data, where alternative disaggregation paths exist, the optimal reconciliation approach allows us to forecast every time series from every disaggregation path. Reconciled forecasts are then calculated for the whole hierarchy. Unlike all top-down and middle-out approaches, the optimal reconciliation weights do not introduce any bias to the reconciled forecasts.

We are also currently looking at some possible approaches for computing prediction intervals for the reconciled forecasts. It is currently not possible to produce proper prediction intervals for bottom-up, top-down, or middle-out forecasts either, and we hope to develop methods to allow the calculation of prediction intervals for these approaches as well.

REFERENCES

- Athanasopoulos, G., Ahmed, R.A. & Hyndman, R.J. (2009). Hierarchical Forecasts for Australian Domestic Tourism, *International Journal of Forecasting*, 25(1), 146–166.
- Hyndman, R.J. (2014). *Forecast: Forecasting Functions for Time Series and Linear Models*, URL: cran.r-project.org/package=forecast
- Hyndman, R.J., Ahmed, R.A., Athanasopoulos, G. & Shang, H.L. (2011). Optimal Combination Forecasts for Hierarchical Time Series, *Computational Statistics & Data Analysis*, 55(9), 2579–2589.



Rob Hyndman is Professor of Statistics in the Department of Econometrics and Business Statistics at Monash University in Australia and Director of the Monash University Business and Economic Forecasting Unit. He is also Editor-in-Chief of the *International Journal of Forecasting* and a Director of the International Institute of Forecasters. Rob is the author of over 100 research papers in statistical science. In 2007, he received the Moran Medal from the Australian Academy of Science for his contributions to statistical research, especially in the area of statistical forecasting. For 30 years, Rob has maintained an active consulting practice, assisting hundreds of companies and organizations.

as Associate Editor for the *International Journal of Forecasting* and is on the Editorial Board of the *Journal of Travel Research*. His research interests include multivariate time-series analysis and forecasting, and wealth and tourism economics.

Rob.Hyndman@monash.edu



George Athanasopoulos

is an Associate Professor and Deputy Head (Caulfield campus) in the Department of Econometrics and Business Statistics at Monash University. This year he was elected a Director of the International Institute of Forecasters. He has recently been appointed

as Associate Editor for the *International Journal of Forecasting* and is on the Editorial Board of the *Journal of Travel Research*. His research interests include multivariate time-series analysis and forecasting, and wealth and tourism economics.

george.athanasopoulos@monash.edu

Hyndman, R.J., Koehler, A.B., Snyder, R.D. & Grose, S. (2002). A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods, *International Journal of Forecasting*, 18(3), 439–454.

Hyndman, R.J., Lee, A.J. & Wang, E. (2014). *hts: Hierarchical and Grouped Time Series*, URL: cran.r-project.org/package=hts

Kolassa, S. & Hyndman, R.J. (2010). Free Open-Source Forecasting Using R, *Foresight*, Issue 17 (Spring 2010), 19–24.

Zotteri, G., Kalchschmidt, M. & Sacconi, N. (2014). Forecasting by Cross-Sectional Aggregation, *Foresight*, Issue 35 (Fall 2014), 35–41.

APPENDIX: IMPLEMENTATION IN R

To implement the method, it is necessary to be able to automatically generate the independent forecasts (using any time series forecasting algorithm), and to be able

to generate the optimal weights by solving the large linear regression problem. Many software packages allow for automatic forecasting. Generating the weights is tricky, as standard regression software will not handle the size of the problem when there are many thousands of time series involved. We have developed some fast and efficient procedures for handling this in R.

The traditional approaches and the optimal reconciliation approach are all implemented in the *hts* package for R. The package also allows for automatic forecasting of time series (using either exponential smoothing or ARIMA modeling).

Suppose we have a collection of time series with the hierarchical structure given in Figure 1. Then the following R code will compute bottom-up forecasts for all series in the hierarchy, using exponential smoothing for the bottom-level series, and aggregation for all other series:

R code

```
# bts is a time series matrix containing the
# bottom level series
# The first three series belong to one group,
# and the last two series
# belong to a different group
y <- hts(bts, nodes=list(2, c(3,2)))
fc <- forecast(y, method="bu", fmethod="ets",
h=12)
```

The *nodes* argument specifies the hierarchical structure, the *method* argument specifies we want bottom-up forecasting, the *fmethod* argument shows we want to use exponential smoothing, and *h=12* indicates we want to forecast 12 steps ahead.

Alternatively, if we wanted to use the optimal reconciliation approach with ARIMA models, we would replace the last line with the following command:

R code

```
fc <- forecast(y, method="comb",
fmethod="arima", h=12)
```

To see all the available options, use `help(forecast.gts)`.

That's it! Everything else is handled automatically. The reconciliation is very fast, even for tens of thousands of time series. However, the forecasts for each node can be time consuming, as each individual series must be modeled and forecast independently.