

## 10.4 Top-down approaches

---

Top-down approaches only work with strictly hierarchical aggregation structures, and not with grouped structures. They involve first generating forecasts for the Total series  $y_t$ , and then disaggregating these down the hierarchy.

We let  $p_1, \dots, p_m$  be a set of disaggregation proportions which dictate how the forecasts of the Total series are to be distributed to obtain forecasts for each series at the bottom-level of the structure. For example, for the hierarchy of Figure 10.1 using proportions  $p_1, \dots, p_5$  we get,

$$\tilde{y}_{AA,t} = p_1 \hat{y}_t, \quad \tilde{y}_{AB,t} = p_2 \hat{y}_t, \quad \tilde{y}_{AC,t} = p_3 \hat{y}_t, \quad \tilde{y}_{BA,t} = p_4 \hat{y}_t \quad \text{and} \quad \tilde{y}_{BB,t} = p_5 \hat{y}_t.$$

Using matrix notation we can stack the set of proportions in a  $m$ -dimensional vector  $\mathbf{p} = (p_1, \dots, p_m)'$  and write

$$\tilde{\mathbf{b}}_t = \mathbf{p} \hat{y}_t.$$

Once the bottom-level  $h$ -step-ahead forecasts have been generated, these are aggregated to generate coherent forecasts for the rest of the series. In general, for a specified set of proportions, top-down approaches can be represented as

$$\tilde{\mathbf{y}}_h = \mathbf{S} \mathbf{p} \hat{y}_t.$$

The two most common top-down approaches specify disaggregation proportions based on the historical proportions of the data. These performed well in the study of Gross & Sohl (1990).

### Average historical proportions

$$p_j = \frac{1}{T} \sum_{t=1}^T \frac{y_{j,t}}{y_t}$$

for  $j = 1, \dots, m$ . Each proportion  $p_j$  reflects the average of the historical proportions of the bottom-level series  $y_{j,t}$  over the period  $t = 1, \dots, T$  relative to the total aggregate  $y_t$ .

This approach is implemented in the `forecast()` function by setting `method="tdgsa"`, where `tdgsa` stands for “top-down Gross-Sohl method A.”

## Proportions of the historical averages

$$p_j = \sum_{t=1}^T \frac{y_{j,t}}{T} / \sum_{t=1}^T \frac{y_t}{T}$$

for  $j = 1, \dots, m$ . Each proportion  $p_j$  captures the average historical value of the bottom-level series  $y_{j,t}$  relative to the average value of the total aggregate  $y_t$ .

This approach is implemented in the `forecast()` function by setting `method="tdgsf"`, where `tdgsf` stands for “top-down Gross-Sohl method F.”

A convenient attribute of such top-down approaches is their simplicity. One only needs to model and generate forecasts for the most aggregated top-level series. In general, these approaches seem to produce quite reliable forecasts for the aggregate levels and they are useful with low count data. On the other hand, one disadvantage is the loss of information due to aggregation. Using such top-down approaches, we are unable to capture and take advantage of individual series characteristics such as time dynamics, special events, etc.

## Forecast proportions

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 ([Athanasopoulos, Ahmed, & Hyndman, 2009](#)).

Consider a one level hierarchy. We first generate  $h$ -step-ahead forecasts for all of the series. We don't use these forecasts directly, and they are not coherent (they don't add up correctly). Let's call these “initial” forecasts. We calculate the proportion of each  $h$ -step-ahead initial forecast at the bottom level, to the aggregate of all the  $h$ -step-ahead initial forecasts at this level. We refer to these as the forecast proportions, and we use them to disaggregate the top-level  $h$ -step-ahead initial forecast in order to generate coherent forecasts for the whole of the hierarchy.

For a  $K$ -level hierarchy, this process is repeated for each node, going from the top to the bottom level. Applying this process leads to the following general rule for obtaining the forecast proportions:

$$p_j = \prod_{\ell=0}^{K-1} \frac{\hat{y}_{j,h}^{(\ell)}}{\hat{S}_{j,h}^{(\ell+1)}}$$

where  $j = 1, 2, \dots, m$ ,  $\hat{y}_{j,h}^{(\ell)}$  is the  $h$ -step-ahead initial forecast of the series that corresponds to the node which is  $\ell$  levels above  $j$ , and  $\hat{S}_{j,h}^{(\ell)}$  is the sum of the  $h$ -step-ahead initial forecasts below the node that is  $\ell$  levels above node  $j$  and are directly connected to that node. These forecast proportions disaggregate the  $h$ -step-ahead initial forecast of the Total series to get  $h$ -step-ahead coherent forecasts of the bottom-level series.

We will use the hierarchy of Figure 10.1 to explain this notation and to demonstrate how this general rule is reached. Assume we have generated initial forecasts for each series in the hierarchy. Recall that for the top-level “Total” series,  $\tilde{y}_h = \hat{y}_h$ , for any top-down approach. Here are some examples using the above notation:

- $\hat{y}_{A,h}^{(1)} = \hat{y}_{B,h}^{(1)} = \hat{y}_h = \tilde{y}_h$ ;
- $\hat{y}_{AA,h}^{(1)} = \hat{y}_{AB,h}^{(1)} = \hat{y}_{AC,h}^{(1)} = \hat{y}_{A,h}^{(1)}$ ;
- $\hat{y}_{AA,h}^{(2)} = \hat{y}_{AB,h}^{(2)} = \hat{y}_{AC,h}^{(2)} = \hat{y}_{BA,h}^{(2)} = \hat{y}_{BB,h}^{(2)} = \hat{y}_h = \tilde{y}_h$ ;
- $\hat{S}_{AA,h}^{(1)} = \hat{S}_{AB,h}^{(1)} = \hat{S}_{AC,h}^{(1)} = \hat{y}_{AA,h}^{(1)} + \hat{y}_{AB,h}^{(1)} + \hat{y}_{AC,h}^{(1)}$ ;
- $\hat{S}_{AA,h}^{(2)} = \hat{S}_{AB,h}^{(2)} = \hat{S}_{AC,h}^{(2)} = \hat{S}_{A,h}^{(1)} = \hat{S}_{B,h}^{(1)} = \hat{S}_h = \hat{y}_{A,h}^{(1)} + \hat{y}_{B,h}^{(1)}$ .

Moving down the farthest left branch of the hierarchy, coherent forecasts are given by

$$\tilde{y}_{A,h} = \left( \frac{\hat{y}_{A,h}^{(1)}}{\hat{S}_{A,h}^{(1)}} \right) \tilde{y}_h = \left( \frac{\hat{y}_{AA,h}^{(1)}}{\hat{S}_{AA,h}^{(2)}} \right) \tilde{y}_h$$

and

$$\tilde{y}_{AA,h} = \left( \frac{\hat{y}_{AA,h}^{(1)}}{\hat{S}_{AA,h}^{(1)}} \right) \tilde{y}_{A,h} = \left( \frac{\hat{y}_{AA,h}^{(1)}}{\hat{S}_{AA,h}^{(1)}} \right) \left( \frac{\hat{y}_{AA,h}^{(1)}}{\hat{S}_{AA,h}^{(2)}} \right) \tilde{y}_h.$$

Consequently,

$$p_1 = \left( \frac{\hat{y}_{AA,h}}{\hat{S}_{AA,h}^{(1)}} \right) \left( \frac{\hat{y}_{AA,h}^{(1)}}{\hat{S}_{AA,h}^{(2)}} \right).$$

The other proportions can be obtained similarly.

One disadvantage of all top-down approaches, including this one, is that it does not produce unbiased coherent forecasts ([Hyndman, Ahmed, Athanasopoulos, & Shang, 2011](#)).

This approach is implemented in the `forecast()` function by setting `method="tdfp"`, where `tdfp` stands for “top-down forecast proportions.”