QBUS 6840: Lecture 12

Hierarchical and Group Time Series

Professor Junbin Gao

The University of Sydney Business School

Outlines

- Hierarchical time series
- Grouped time series
- Modeling approaches
- Mapping matrices

Readings:
Online textbook Chapter 10
`https://otexts.com/fpp2/hierarchical.html`; and Slides

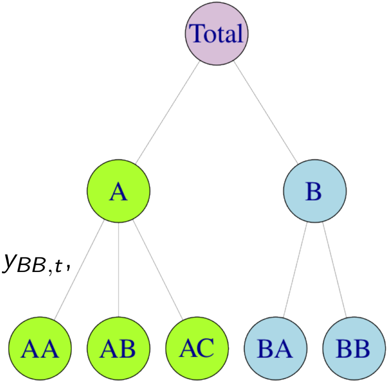
Overview

- Time series can often be naturally disaggregated by various attributes of interest.
- These categories are nested within the larger group categories, and so the collection of time series follow a hierarchical aggregation structure. We refer to these as "hierarchical time series"
- Sometimes we have a more complicated aggregation structure where the product hierarchy and the geographic hierarchy can both be used together. We usually refer to these as "grouped time series"

Hierarchical Time Series: Structure

2-level hierarchical structure

- Total series is denoted by y_t for $t = 1, \ldots, T$. The Total is disaggregated into two series at level 1, which in turn are divided into three and two series respectively at the bottom-level of the hierarchy.
- These disaggregated time series are denoted by $y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{AC,t}, y_{BA,t}, y_{BB,t}$, or you number them from level 1 to level 2, from left to right as $y_{1,t}, y_{2,t}, y_{3,t}, y_{4,t}, y_{5,t}, y_{6,t}, y_{7,t}$. The total can be named as $y_t = y_{0,t}$.



Relations

- In total there are eight time series
- We have

$$y_t = y_{A,t} + y_{B,t}$$

and

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t}, \quad y_{B,t} = y_{BA,t} + y_{BB,t}$$

Hence

$$y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BB,t}$$

- These equations can be thought of as aggregation constraints or summing equalities
- At each level, the sum gives the total

Matrix Representation

- We will use the time series at bottom level as building blocks, then the relations can be summarized in the following matrix representation

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix}$$

- In compact notation

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$$

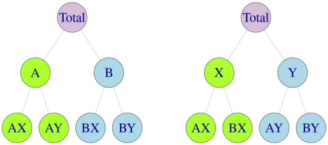
Grouped Time Series

- Grouped time series involve more general aggregation structures than hierarchical time series
- With grouped time series, the structure does not naturally disaggregate in a unique hierarchical manner
- Often the disaggregating factors are both nested and crossed
- For example, we could further disaggregate all geographic levels of the Australian tourism data by purpose of travel (such as holidays, business, etc.)
- So we could consider visitors nights split by purpose of travel for the whole of Australia, and for each state, and for each zone.
- Then we describe the structure as involving the purpose of travel "crossed" with the geographic hierarchy

Grouped Time Series: Structure

2-level grouped structure

- Total, the most aggregate level of the data, again represented by y_t.
- The Total can be disaggregated by attributes (A, B) forming series $y_{A,t}$ and $y_{B,t}$, or by attributes (X, Y) forming series $y_{X,t}$ and $y_{Y,t}$.
- At the bottom level, the data are disaggregated by both attributes.



Relations

- The previous example shows that there are alternative aggregation paths for grouped structures
  - $y_t = y_{AX,t} + y_{AY,t} + y_{BX,t} + y_{BY,t}$.
  - At level 1

$$y_{A,t} = y_{AX,t} + y_{AY,t}, \quad y_{B,t} = y_{BX,t} + y_{BY,t}$$

  - Or at second level 1

$$y_{X,t} = y_{AX,t} + y_{BX,t}, \quad y_{Y,t} = y_{AY,t} + y_{BY,t}$$

- There are in total 9 different time series. At both bottom levels, all the times series are the same.

Matrix Representation

- We can use the time series at bottom level as building blocks, then the relations can be summarized in the following matrix representation

$$
\begin{bmatrix} y_t \\ y_A t \\ y_B t \\ y_X t \\ y_Y t \\ y_{AX} t \\ y_{AY} t \\ y_{BX} t \\ y_{BY} t \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AX} t \\ y_{AY} t \\ y_{BX} t \\ y_{BY} t \end{bmatrix}
$$

- Or

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

The bottom-up approach

- A simple method for generating coherent forecasts is the bottom-up approach
  - first generating forecasts for each series at the bottom-level, and
  - then summing these to produce forecasts for all the series in the structure.
- For example, for the hierarchy we discussed, we first do h-step-ahead forecasts for each of the bottom-level series:

$$\widehat{y}_{AA,h}, \quad \widehat{y}_{AB,h}, \quad \widehat{y}_{AC,h}, \quad \widehat{y}_{BA,h} \quad \text{and} \quad \widehat{y}_{BB,h}.$$

- Summing these, we get h-step-ahead coherent forecasts for the rest of the series:

$$\begin{aligned} \tilde{y}_h &= \widehat{y}_{AA,h} + \widehat{y}_{AB,h} + \widehat{y}_{AC,h} + \widehat{y}_{BA,h} + \widehat{y}_{BB,h}, \\ \tilde{y}_{A,h} &= \widehat{y}_{AA,h} + \widehat{y}_{AB,h} + \widehat{y}_{AC,h}, \\ \text{and} \quad \tilde{y}_{B,h} &= \widehat{y}_{BA,h} + \widehat{y}_{BB,h}. \end{aligned}$$

where the "tilde" notation indicates coherent forecasts

The bottom-up approach

- Or the structure matrix equation can be written for forecasts

$$\begin{bmatrix} \tilde{y}_h \\ \tilde{y}_{A,h} \\ \tilde{y}_{B,h} \\ \tilde{y}_{AA,h} \\ \tilde{y}_{AB,h} \\ \tilde{y}_{AC,h} \\ \tilde{y}_{BA,h} \\ \tilde{y}_{BB,h} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \widehat{y}_{AA,h} \\ \widehat{y}_{AB,h} \\ \widehat{y}_{AC,h} \\ \widehat{y}_{BA,h} \\ \widehat{y}_{BB,h} \end{bmatrix}$$

- Can you write out the forecast for the grouped time series?

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.
- Assume there are m disaggregated time series at the bottom level. We use $\{p_1, \ldots, p_m\}$ to denote the disaggregation proportions dictating how the forecasts of the Total series are to be distributed to obtain forecasts for each series at the bottom-level of the structure.

Top-down approaches: Example

- For example, in the previous two level hierarchical aggregation structures, we have

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

$$\text{and} \qquad \tilde{y}_{BB,t} = p_5 \hat{y}_t.$$

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

$$\tilde{\boldsymbol{b}}_t = \boldsymbol{p} \hat{y}_t$$

- 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

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

Top-down approaches: How to get $\boldsymbol{p}$

- Average historical proportions

$$p_j = \frac{1}{T}\sum_{t=1}^{T}\frac{y_{j,t}}{y_t}, \quad \text{for } j = 1, \ldots, m$$

  where $y_{j,t}$ is the historical values at the bottom level time series

- Proportions of the historical averages

$$p_j = \sum_{t=1}^{T}\frac{y_{j,t}}{T}\Big/\sum_{t=1}^{T}\frac{y_t}{T}, \quad \text{for } j = 1, \ldots, m$$

- Other strategies, see the online text.

Mapping matrices

- Suppose we forecast all series independently, ignoring the aggregation constraints.
- We call these the base forecasts and denote them by $\widehat{\boldsymbol{y}}_h$ where h is the forecast horizon. They are stacked in the same order as the data $\boldsymbol{y}_t$.
- Then all forecasting approaches for either hierarchical or grouped structures can be represented as

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_h$$

where $\boldsymbol{G}$ is a matrix that maps the base forecasts into the bottom-level, and the summing matrix $\boldsymbol{S}$ sums these up using the aggregation structure to produce a set of coherent forecasts $\tilde{\boldsymbol{y}}_h$.

Designing $\boldsymbol{G}$

- Bottom-up Approach: As each bottom series comes from itself, so for previous 2-level hierarchical structure

$$\boldsymbol{G} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- The top-down approaches: the bottom series is portion of the total series forecast, so

$$\boldsymbol{G} = \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Forecast reconciliation

- The key requirement for the hierarchical or grouped forecasts is to satisfy the coherent constraint conditions
- We write $\tilde{\boldsymbol{y}}_h = \boldsymbol{SG}\hat{\boldsymbol{y}}_h$ as

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{P}\hat{\boldsymbol{y}}_h$$

  where $\boldsymbol{P} = \boldsymbol{SG}$ is a "projection" or a "reconciliation matrix".
- It takes the incoherent base forecasts $hat\boldsymbol{y}_h$, and reconciles them to produce coherent forecasts $\tilde{\boldsymbol{y}}_h$.

The Best $\boldsymbol{G}$

- We shall find the optimal G matrix to give the most accurate reconciled forecasts
- Optimal forecast reconciliation will occur if we can find the G matrix which minimises the forecast error of the set of coherent forecasts.
- Suppose we generate coherent forecasts using

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_h$$

- Theory has proved that, to have unbiased forecasts coherent forecasts, we shall have $\boldsymbol{SGS} = \boldsymbol{S}$.
- Interestingly, no top-down method satisfies this constraint, so all top-down methods are biased.

The Best $\boldsymbol{G}$

- Wickramasuriya et al. show that the variance-covariance matrix of the h-step-ahead coherent forecast errors is given by

$$\boldsymbol{V}_h = \mathrm{Var}[\boldsymbol{y}_{T+h} - \tilde{\boldsymbol{y}}_h] = \boldsymbol{S}\boldsymbol{G}\boldsymbol{W}_h\boldsymbol{G}^T\boldsymbol{S}^T$$

  where $\boldsymbol{W}_h = \mathrm{Var}[(\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_h)]$ is the variance-covariance matrix of the corresponding base forecast errors.
- The objective is to find a matrix $\boldsymbol{G}$ that minimises the error variances of the coherent forecasts.
- Under certain conditions, we can find

$$\boldsymbol{G} = (\boldsymbol{S}^T\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{W}_h^{-1}.$$

Practical Estimate of $\boldsymbol{W}_h$

There are couple of suggested estimate for $\boldsymbol{W}_h$

- $\boldsymbol{W}_h = k_h \boldsymbol{I}$ for all h, where $k_h > 0$ is a constant.
- $\boldsymbol{W}_h = k_h \text{diag}(\hat{\boldsymbol{W}}_1)$, for all h, where $k_h > 0$ is a constant

  $$\hat{\boldsymbol{W}}_1 = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{e}_t \boldsymbol{e}_t^T,$$

  and $\boldsymbol{e}_t$ is an n-dimensional vector of residuals of the models that generated the base forecasts stacked in the same order as the data.
- $\boldsymbol{W}_h = k_h \boldsymbol{\Lambda}$ for all h, where $k_h > 0$ is a constant.
  $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{S}\boldsymbol{1})$ and $\boldsymbol{1}$ is a unit vector of dimension n.