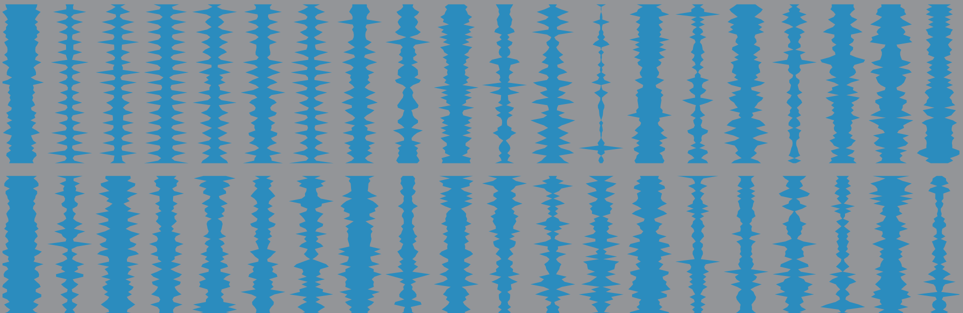
MONASH BUSINESS SCHOOL

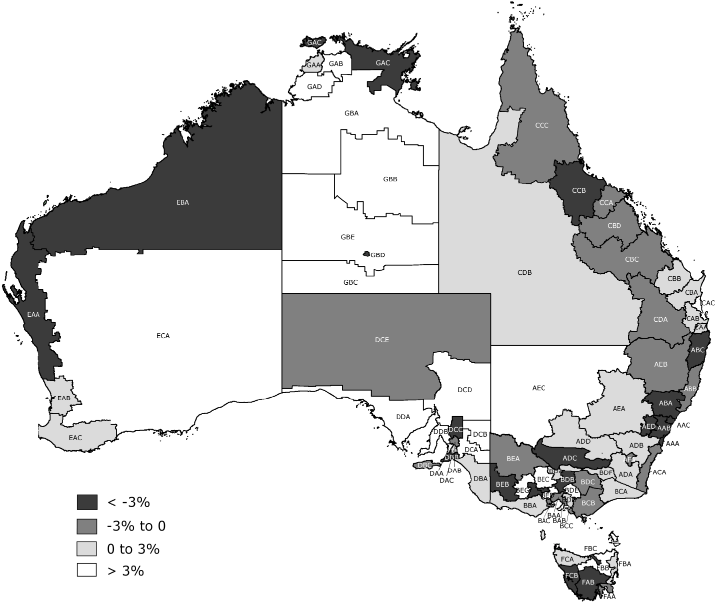
Rob J Hyndman

Visualizing and forecasting big time series data



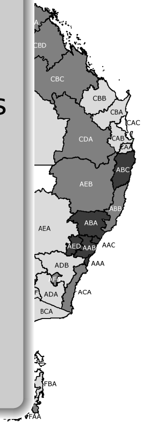
Outline

1. Australian tourism demand



1. Australian tourism demand

- Quarterly data on visitor night from 1998:Q1 – 2013:Q4
- From: *National Visitor Survey*, based on annual interviews of 120,000 Australians aged 15+, collected by Tourism Research Australia.
- Split by 7 states, 27 zones and 76 regions (a geographical hierarchy)
- Also split by purpose of travel
  - Holiday
  - Visiting friends and relatives (VFR)
  - Business
  - Other
- 304 bottom-level series



2. Labour market participation

Australia and New Zealand Standard Classification of Occupations

- 8 major groups
  - 43 sub-major groups
    - 97 minor groups
      - – 359 unit groups
        - * 1023 occupations

Example: statistician

2 Professionals

22 Business, Human Resource and Marketing Professionals

224 Information and Organisation Professionals

2241 Actuaries, Mathematicians and Statisticians

224113 Statistician

2. Labour market participation

Australia and New Zealand Standard Classification of Occupations

- 8 major groups
  - 43 sub-major groups
    - 97 minor groups
      - – 359 unit groups
        - * 1023 occupations

Example: statistician

2 Professionals

22 Business, Human Resource and Marketing Professionals

224 Information and Organisation Professionals

2241 Actuaries, Mathematicians and Statisticians

224113 Statistician

3. Spectacle sales

- Monthly UK sales data from 2000 – 2014
- Provided by a large spectacle manufacturer
- Split by brand (26), gender (3), price range (6), materials (4), and stores (600)
- About 1 million bottom-level series

3. Spectacle sales

- Monthly UK sales data from 2000 – 2014
- Provided by a large spectacle manufacturer
- Split by brand (26), gender (3), price range (6), materials (4), and stores (600)
- About 1 million bottom-level series

3. Spectacle sales

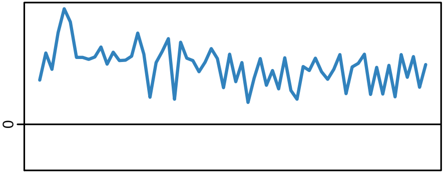
- Monthly UK sales data from 2000 – 2014
- Provided by a large spectacle manufacturer
- Split by brand (26), gender (3), price range (6), materials (4), and stores (600)
- About 1 million bottom-level series

3. Spectacle sales

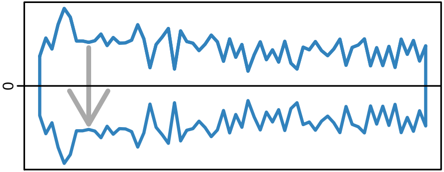
- Monthly UK sales data from 2000 – 2014
- Provided by a large spectacle manufacturer
- Split by brand (26), gender (3), price range (6), materials (4), and stores (600)
- About 1 million bottom-level series

Outline

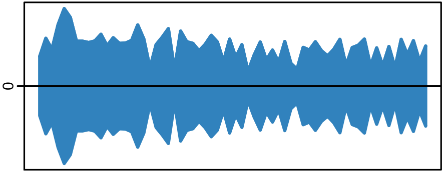
Kite diagrams



Line graph profile

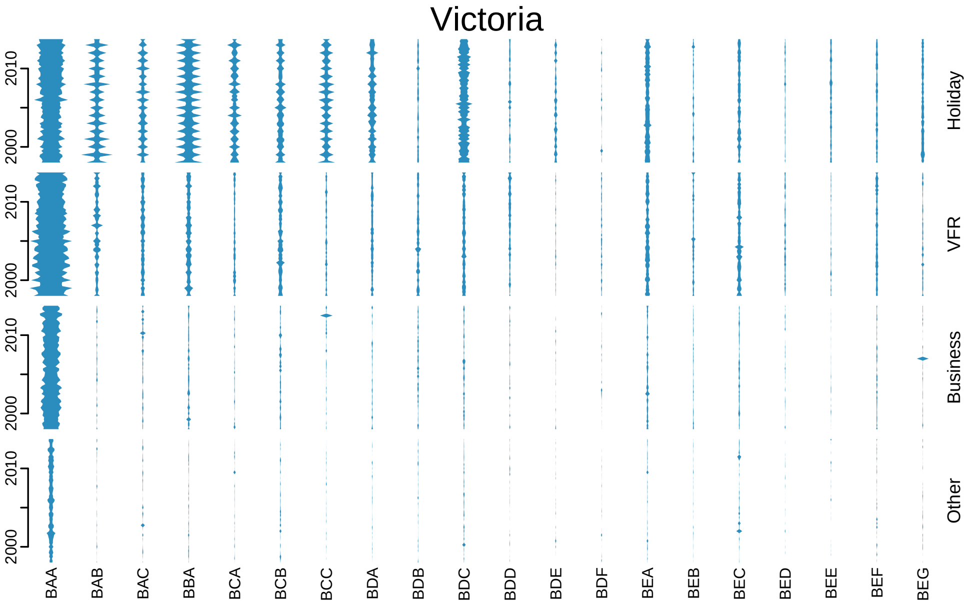


Duplicate & flip around the horizontal axis

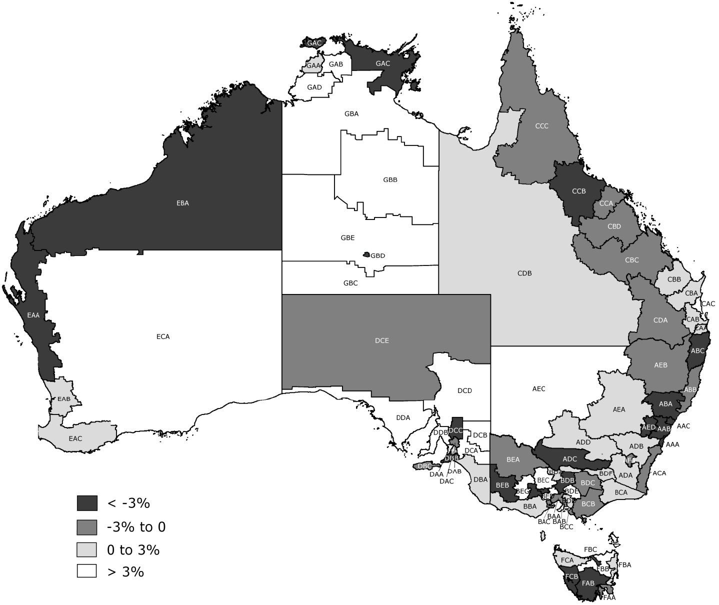


Fill the colour

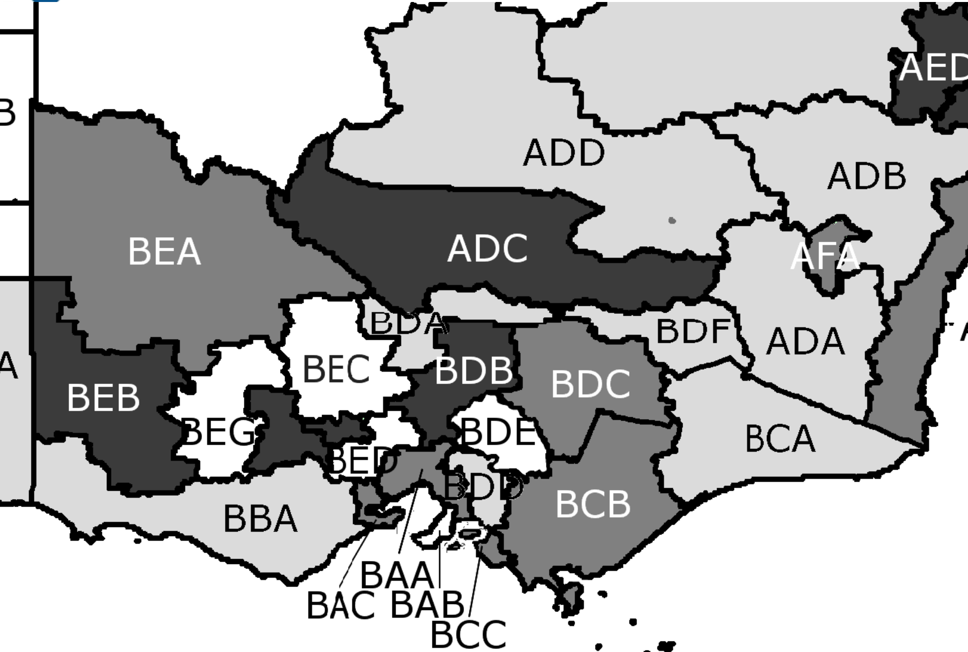
Kite diagrams: Victorian tourism



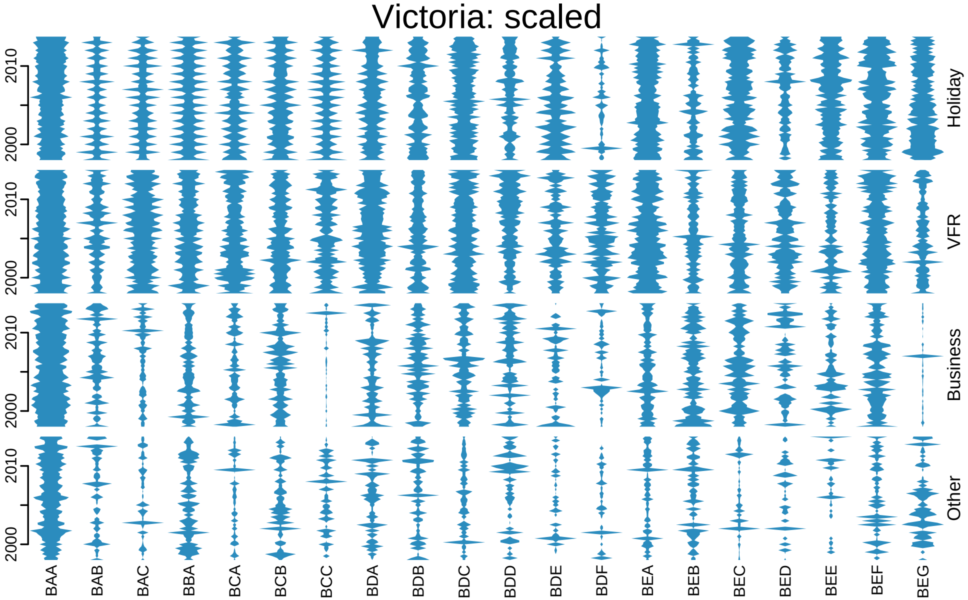
Kite diagrams: Victorian tourism



Kite diagrams: Victorian tourism

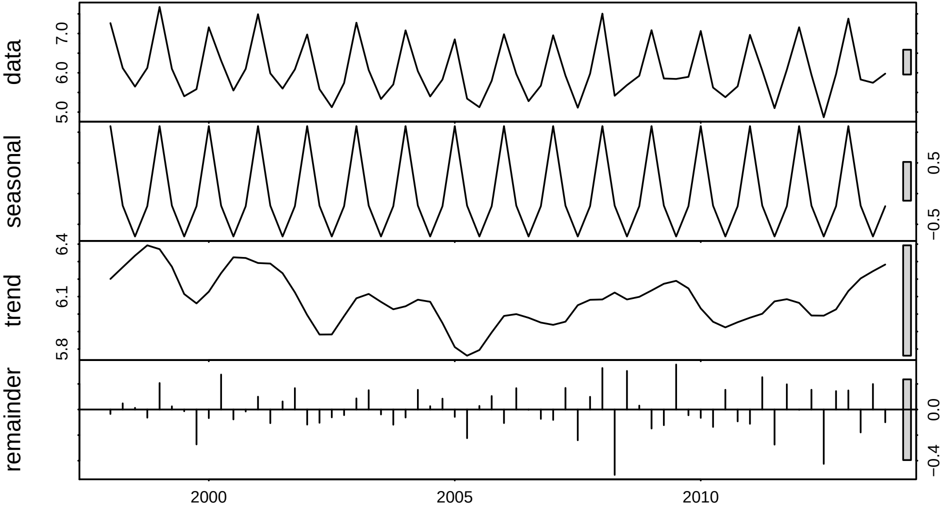


Kite diagrams: Victorian tourism



An STL decomposition

STL decomposition of tourism demand for holidays in Peninsula

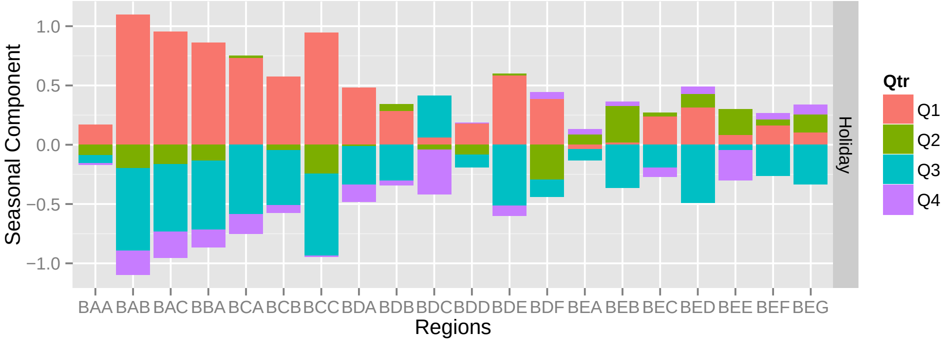


Seasonal stacked bar chart

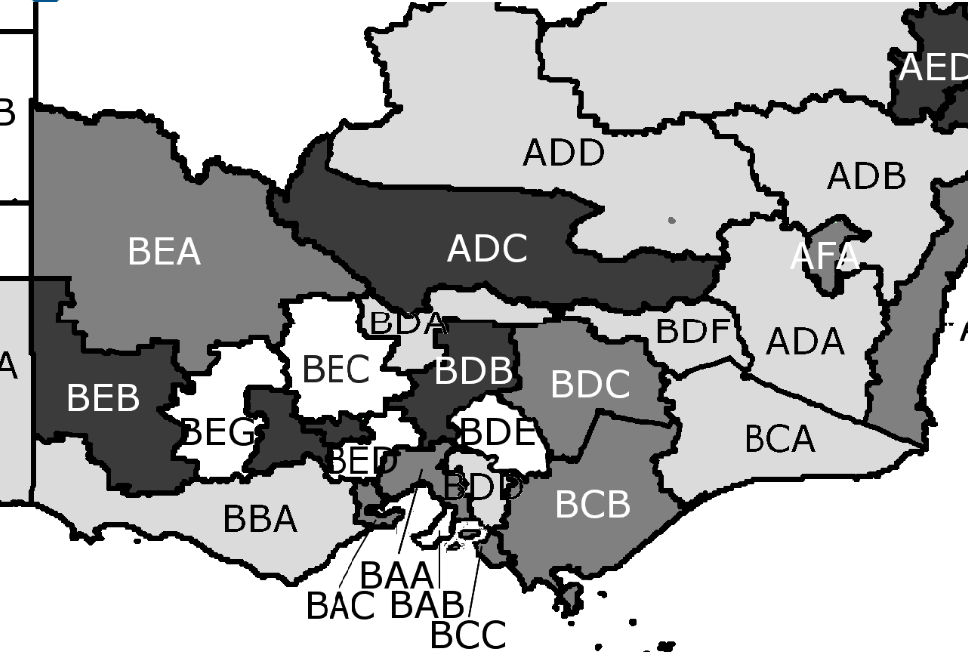
- Place positive values above the origin while negative values below the origin
- Map the bar length to the magnitude
- Encode quarters by colours

Seasonal stacked bar chart

- Place positive values above the origin while negative values below the origin
- Map the bar length to the magnitude
- Encode quarters by colours



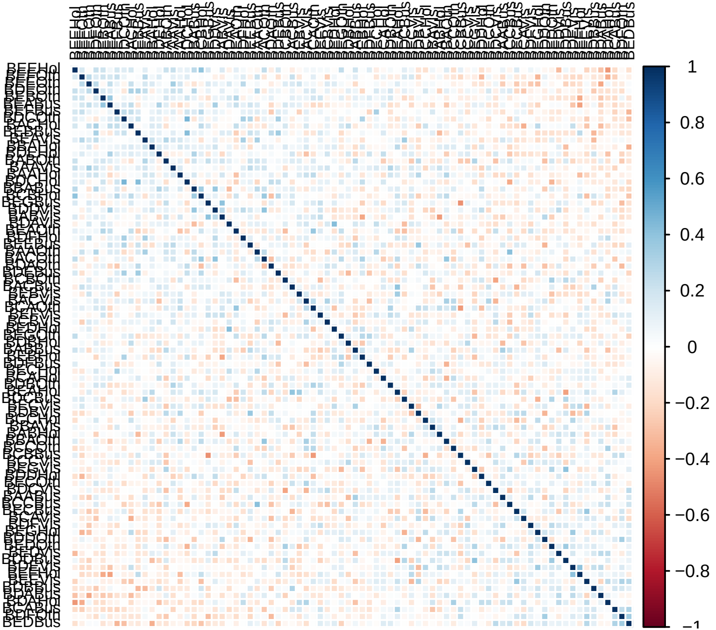
Seasonal stacked bar chart: VIC



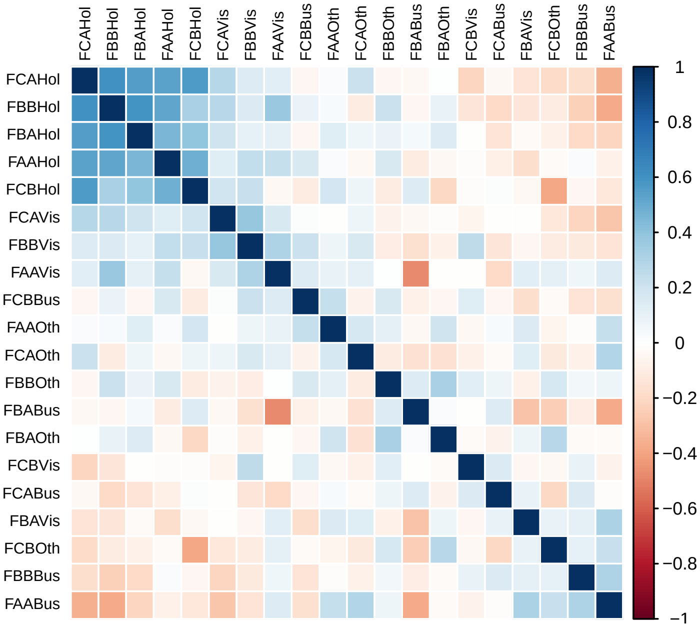
Corrgram of remainder

- Compute the correlations among the remainder components
- Render both the sign and magnitude using a colour mapping of two hues
- Order variables according to the first principal component of the correlations.

Corrgram of remainder: VIC



Corrgram of remainder: TAS

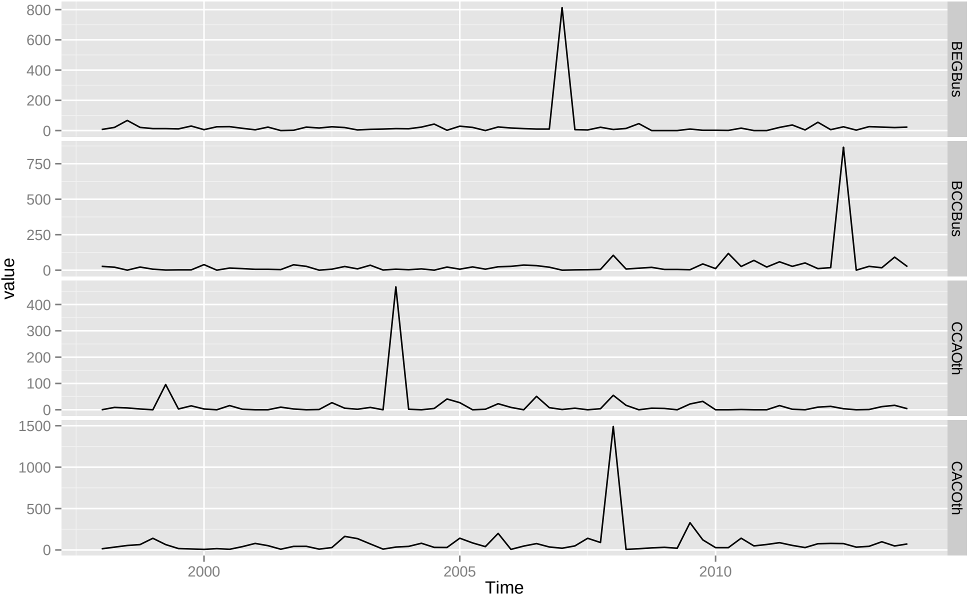


Feature analysis

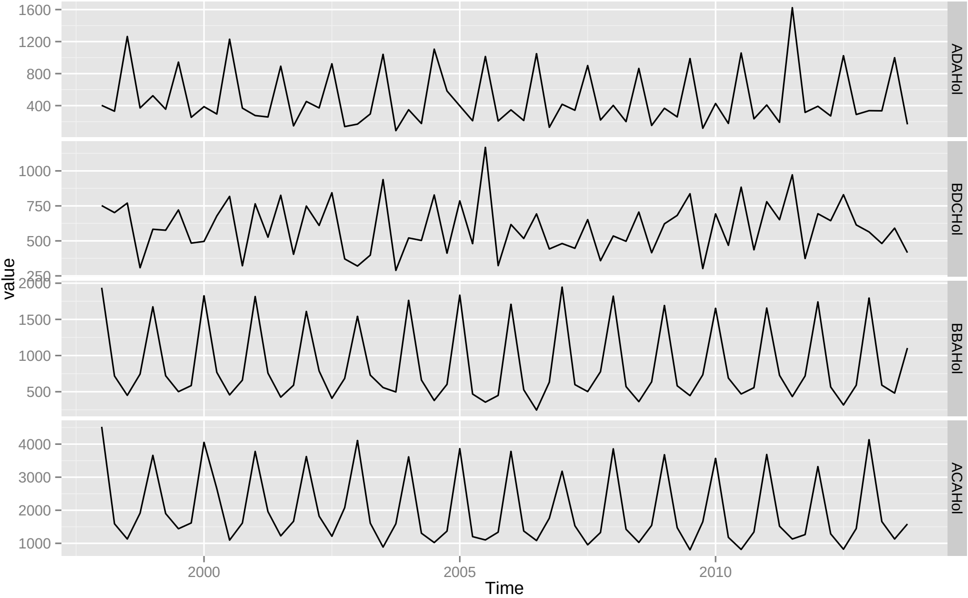
- Summarize each time series with a feature vector:
  - strength of trend
  - lumpiness (variance of annual variances of remainder)
  - strength of seasonality
  - size of seasonal peak
  - size of seasonal trough
  - ACF1
  - linearity of trend
  - curvature of trend
  - spectral entropy
- Do PCA on feature matrix

Feature analysis

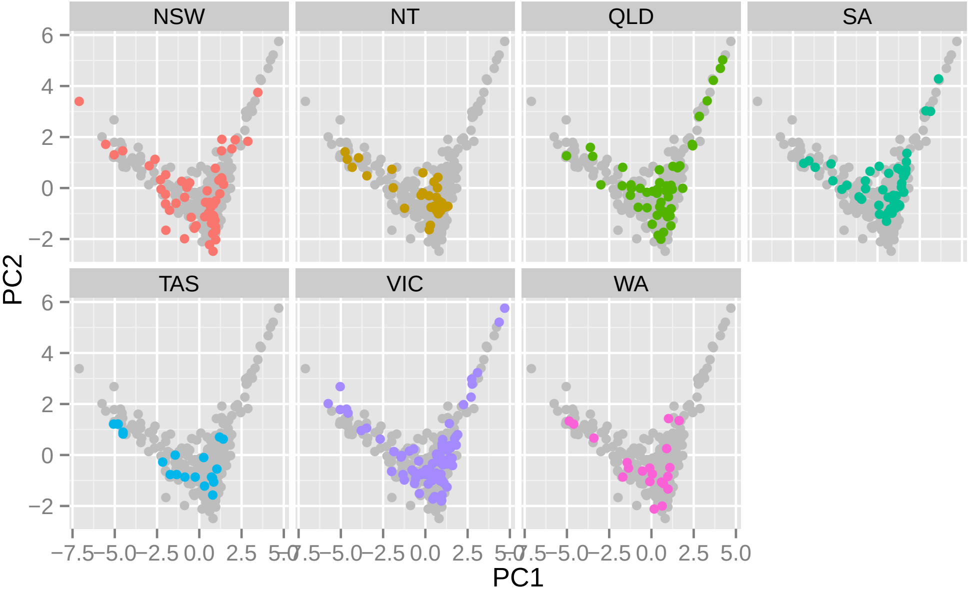
Feature analysis



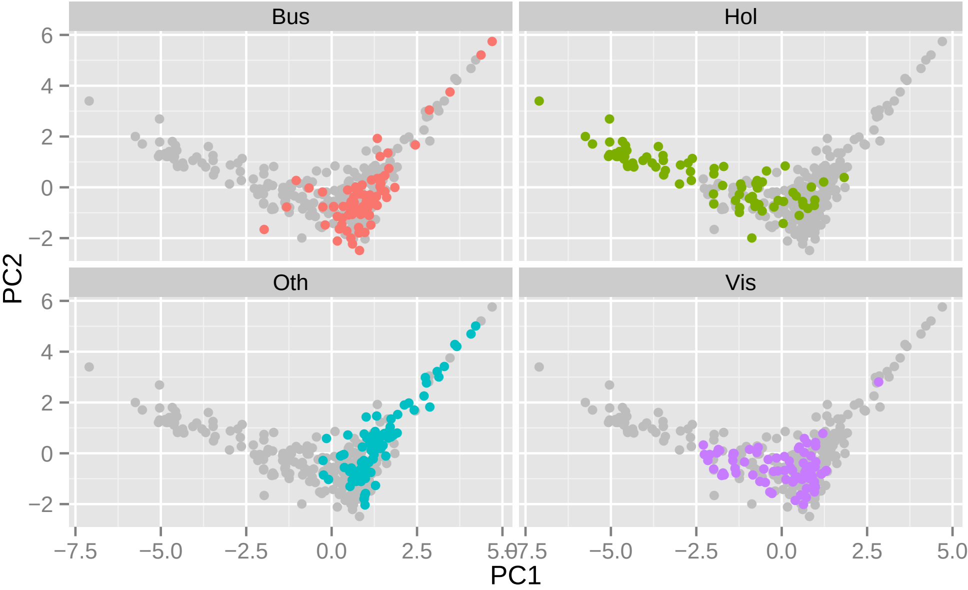
Feature analysis



Feature analysis



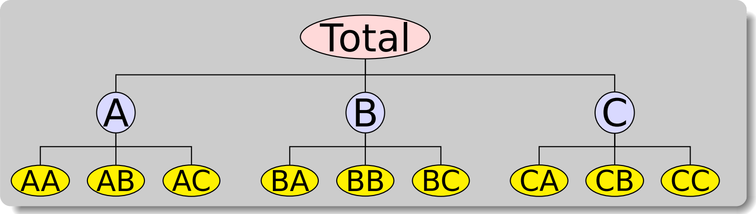
Feature analysis



Outline

Hierarchical time series

A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.

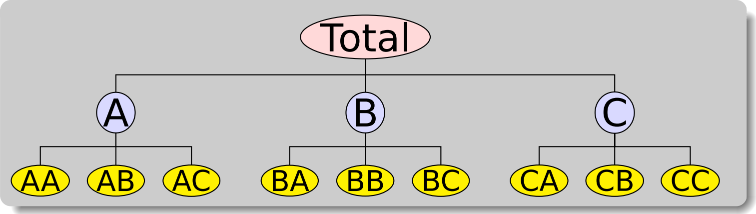


Examples

- Net labour turnover
- Tourism by state and region

Hierarchical time series

A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.

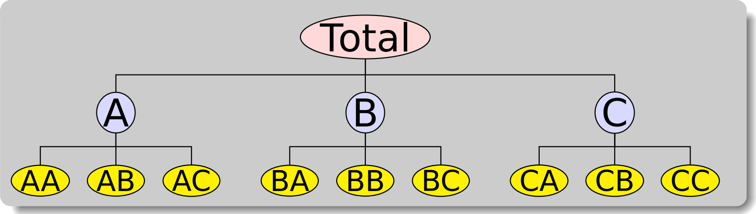


Examples

- Net labour turnover
- Tourism by state and region

Hierarchical time series

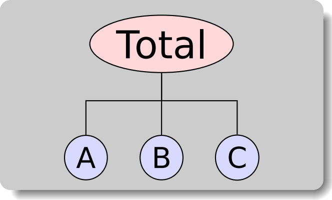
A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.



Examples

- Net labour turnover
- Tourism by state and region

Hierarchical time series

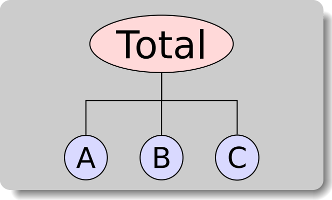


Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: observation on series X at time t.

$\boldsymbol{B}_t$: vector of all series at bottom level in time t.

Hierarchical time series

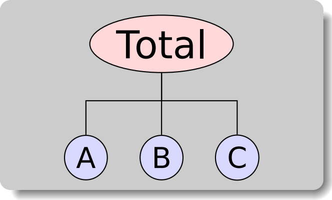


Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: observation on series X at time t.

$\boldsymbol{B}_t$: vector of all series at bottom level in time t.

Hierarchical time series



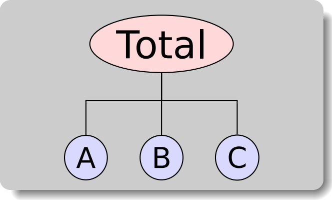
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: observation on series X at time t.

$\boldsymbol{B}_t$: vector of all series at bottom level in time t.

$$\boldsymbol{y}_t = [Y_t, Y_{A,t}, Y_{B,t}, Y_{C,t}]' = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix}$$

Hierarchical time series



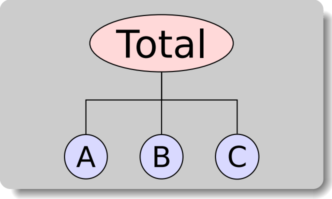
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: observation on series X at time t.

$\boldsymbol{B}_t$: vector of all series at bottom level in time t.

$$\boldsymbol{y}_t = [Y_t, Y_{A,t}, Y_{B,t}, Y_{C,t}]' = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \begin{pmatrix} Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix}$$

Hierarchical time series



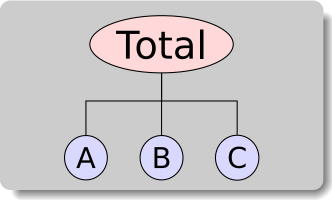
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: observation on series X at time t.

$\boldsymbol{B}_t$: vector of all series at bottom level in time t.

$$\boldsymbol{y}_t = [Y_t, Y_{A,t}, Y_{B,t}, Y_{C,t}]' = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

Hierarchical time series



Y_t : observed aggregate of all series at time t.

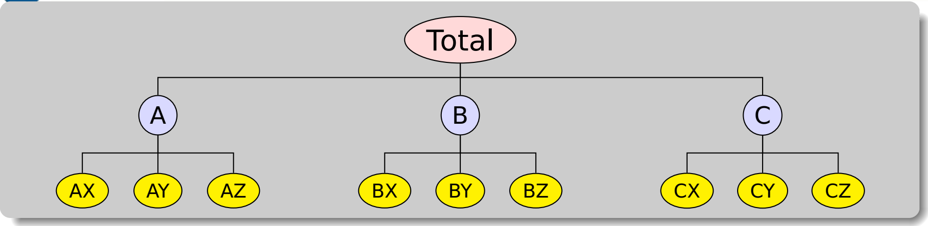
$Y_{X,t}$: observation on series X at time t.

$\boldsymbol{B}_t$: vector of all series at bottom level in time t.

$$\boldsymbol{y}_t = [Y_t, Y_{A,t}, Y_{B,t}, Y_{C,t}]' = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

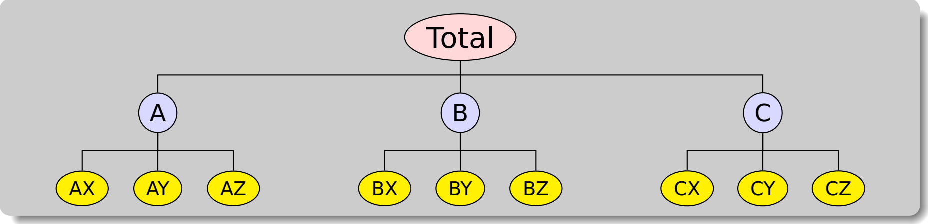
$$\boxed{\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{B}_t}$$

Hierarchical time series

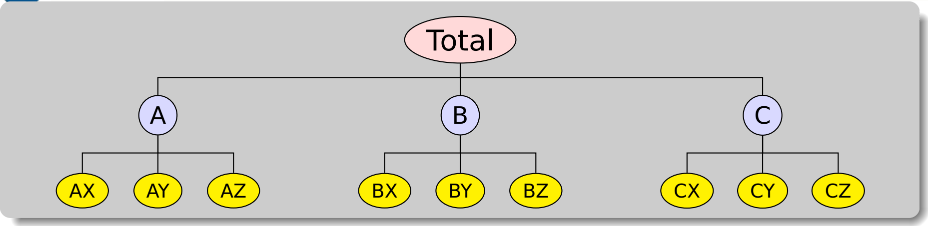


$$\boldsymbol{y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

Hierarchical time series



$$\boldsymbol{y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$



$$\boldsymbol{y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{B}_t$$

Forecasting notation

Let $\hat{\boldsymbol{y}}_n(h)$ be vector of initial h-step forecasts, made at time n, stacked in same order as $\boldsymbol{y}_t$.

(They may not add up.)

Reconciled forecasts are of the form:

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_n(h)$ to get bottom-level forecasts.

Forecasting notation

Let $\hat{\boldsymbol{y}}_n(h)$ be vector of initial h-step forecasts, made at time n, stacked in same order as $\boldsymbol{y}_t$. (They may not add up.)

Reconciled forecasts are of the form:

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_n(h)$ to get bottom-level forecasts.
- $\boldsymbol{S}$ adds them up

Forecasting notation

Let $\hat{\boldsymbol{y}}_n(h)$ be vector of initial h-step forecasts, made at time n, stacked in same order as $\boldsymbol{y}_t$. (They may not add up.)

Reconciled forecasts are of the form:

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_n(h)$ to get bottom-level forecasts.
- $\boldsymbol{S}$ adds them up

Forecasting notation

Let $\hat{\boldsymbol{y}}_n(h)$ be vector of initial h-step forecasts, made at time n, stacked in same order as $\boldsymbol{y}_t$. (They may not add up.)

Reconciled forecasts are of the form:

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_n(h)$ to get bottom-level forecasts.
- $\boldsymbol{S}$ adds them up

Forecasting notation

Let $\hat{\boldsymbol{y}}_n(h)$ be vector of initial h-step forecasts, made at time n, stacked in same order as $\boldsymbol{y}_t$. (They may not add up.)

Reconciled forecasts are of the form:

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ extracts and combines base forecasts $\hat{\boldsymbol{y}}_n(h)$ to get bottom-level forecasts.
- $\boldsymbol{S}$ adds them up

General properties: bias

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Assume: base forecasts $\hat{\boldsymbol{y}}_n(h)$ are unbiased:

$$\mathrm{E}[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n] = \mathrm{E}[\boldsymbol{y}_{n+h}|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$$

- Let $\hat{\boldsymbol{B}}_n(h)$ be bottom level base forecasts with $\beta_n(h) = \mathrm{E}[\hat{\boldsymbol{B}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$.
- Then $\mathrm{E}[\hat{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\beta_n(h)$.
- We want the revised forecasts to be unbiased: $\mathrm{E}[\tilde{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{S}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.

Revised forecasts are unbiased iff $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$.

General properties: bias

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{SP}\hat{\boldsymbol{y}}_n(h)$$

Assume: base forecasts $\hat{\boldsymbol{y}}_n(h)$ are unbiased:

$$\mathrm{E}[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n] = \mathrm{E}[\boldsymbol{y}_{n+h}|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$$

- Let $\hat{\boldsymbol{B}}_n(h)$ be bottom level base forecasts with $\beta_n(h) = \mathrm{E}[\hat{\boldsymbol{B}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$.
- Then $\mathrm{E}[\hat{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\beta_n(h)$.
- We want the revised forecasts to be unbiased: $\mathrm{E}[\tilde{\boldsymbol{y}}_n(h)] = \boldsymbol{SPS}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.

Revised forecasts are unbiased iff $\boldsymbol{SPS} = \boldsymbol{S}$.

General properties: bias

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Assume: base forecasts $\hat{\boldsymbol{y}}_n(h)$ are unbiased:

$$\mathrm{E}[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n] = \mathrm{E}[\boldsymbol{y}_{n+h}|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$$

- Let $\hat{\boldsymbol{B}}_n(h)$ be bottom level base forecasts with $\boldsymbol{\beta}_n(h) = \mathrm{E}[\hat{\boldsymbol{B}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$.
- Then $\mathrm{E}[\hat{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\boldsymbol{\beta}_n(h)$.
- We want the revised forecasts to be unbiased: $\mathrm{E}[\tilde{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{S}\boldsymbol{\beta}_n(h) = \boldsymbol{S}\boldsymbol{\beta}_n(h)$.

Revised forecasts are unbiased iff $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$.

General properties: bias

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Assume: base forecasts $\hat{\boldsymbol{y}}_n(h)$ are unbiased:

$$\mathrm{E}[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n] = \mathrm{E}[\boldsymbol{y}_{n+h}|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$$

- Let $\hat{\boldsymbol{B}}_n(h)$ be bottom level base forecasts with $\beta_n(h) = \mathrm{E}[\hat{\boldsymbol{B}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$.
- Then $\mathrm{E}[\hat{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\beta_n(h)$.
- We want the revised forecasts to be unbiased: $\mathrm{E}[\tilde{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{S}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.

Revised forecasts are unbiased iff $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$.

General properties: bias

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{SP}\hat{\boldsymbol{y}}_n(h)$$

Assume: base forecasts $\hat{\boldsymbol{y}}_n(h)$ are unbiased:

$$\mathrm{E}[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n] = \mathrm{E}[\boldsymbol{y}_{n+h}|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$$

- Let $\hat{\boldsymbol{B}}_n(h)$ be bottom level base forecasts with $\boldsymbol{\beta}_n(h) = \mathrm{E}[\hat{\boldsymbol{B}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$.
- Then $\mathrm{E}[\hat{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\boldsymbol{\beta}_n(h)$.
- We want the revised forecasts to be unbiased: $\mathrm{E}[\tilde{\boldsymbol{y}}_n(h)] = \boldsymbol{SPS}\boldsymbol{\beta}_n(h) = \boldsymbol{S}\boldsymbol{\beta}_n(h)$.

Revised forecasts are unbiased iff $\boldsymbol{SPS} = \boldsymbol{S}$.

General properties: bias

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{SP}\hat{\boldsymbol{y}}_n(h)$$

Assume: base forecasts $\hat{\boldsymbol{y}}_n(h)$ are unbiased:

$$\mathrm{E}[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n] = \mathrm{E}[\boldsymbol{y}_{n+h}|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$$

- Let $\hat{\boldsymbol{B}}_n(h)$ be bottom level base forecasts with $\beta_n(h) = \mathrm{E}[\hat{\boldsymbol{B}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$.
- Then $\mathrm{E}[\hat{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\beta_n(h)$.
- We want the revised forecasts to be unbiased: $\mathrm{E}[\tilde{\boldsymbol{y}}_n(h)] = \boldsymbol{SPS}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.

Revised forecasts are unbiased iff $\boldsymbol{SPS} = \boldsymbol{S}$.

General properties: bias

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Assume: base forecasts $\hat{\boldsymbol{y}}_n(h)$ are unbiased:

$$\mathrm{E}[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n] = \mathrm{E}[\boldsymbol{y}_{n+h}|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$$

- Let $\hat{\boldsymbol{B}}_n(h)$ be bottom level base forecasts with $\beta_n(h) = \mathrm{E}[\hat{\boldsymbol{B}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n]$.
- Then $\mathrm{E}[\hat{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\beta_n(h)$.
- We want the revised forecasts to be unbiased: $\mathrm{E}[\tilde{\boldsymbol{y}}_n(h)] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{S}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.

Revised forecasts are unbiased iff $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$.

General properties: variance

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Let variance of base forecasts $\hat{\boldsymbol{y}}_n(h)$ be given by

$$\Sigma_h = \text{Var}[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1, \dots, \boldsymbol{y}_n]$$

Then the variance of the revised forecasts is given by

$$\text{Var}[\tilde{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1, \dots, \boldsymbol{y}_n] = \boldsymbol{S}\boldsymbol{P}\Sigma_h\boldsymbol{P}'\boldsymbol{S}'.$$

General properties: variance

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{SP}\hat{\boldsymbol{y}}_n(h)$$

Let variance of base forecasts $\hat{\boldsymbol{y}}_n(h)$ be given by

$$\Sigma_h = \text{Var}[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1, \dots, \boldsymbol{y}_n]$$

Then the variance of the revised forecasts is given by

$$\text{Var}[\tilde{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1, \dots, \boldsymbol{y}_n] = \boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}'.$$

General properties: variance

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Let variance of base forecasts $\hat{\boldsymbol{y}}_n(h)$ be given by

$$\Sigma_h = \text{Var}[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1, \dots, \boldsymbol{y}_n]$$

Then the variance of the revised forecasts is given by

$$\text{Var}[\tilde{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1, \dots, \boldsymbol{y}_n] = \boldsymbol{S}\boldsymbol{P}\Sigma_h\boldsymbol{P}'\boldsymbol{S}'.$$

BLUF via trace minimization

Theorem

For any $\boldsymbol{P}$ satisfying $\boldsymbol{SPS} = \boldsymbol{S}$, then
$$\min_{\boldsymbol{P}} = \text{trace}[\boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}']$$
has solution $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$.

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts | Base forecasts

BLUF via trace minimization

Theorem

For any $\boldsymbol{P}$ satisfying $\boldsymbol{SPS} = \boldsymbol{S}$, then

$$\min_{\boldsymbol{P}} = \text{trace}[\boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}']$$

has solution $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$.

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts

Base forecasts

- Equivalent to GLS estimate of regression $\hat{\boldsymbol{y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$ where $\varepsilon \sim N(0, \Sigma_h)$.

BLUF via trace minimization

Theorem

For any $\boldsymbol{P}$ satisfying $\boldsymbol{SPS} = \boldsymbol{S}$, then

$$\min_{\boldsymbol{P}} = \text{trace}[\boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}']$$

has solution $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$.

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts | Base forecasts

- Equivalent to GLS estimate of regression $\hat{\boldsymbol{y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$ where $\varepsilon \sim N(\boldsymbol{0}, \Sigma_h)$.
- Problem: Σ_h hard to estimate.

BLUF via trace minimization

Theorem

For any $\boldsymbol{P}$ satisfying $\boldsymbol{SPS} = \boldsymbol{S}$, then

$$\min_{\boldsymbol{P}} = \text{trace}[\boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}']$$

has solution $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$.

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts | Base forecasts

- Equivalent to GLS estimate of regression $\hat{\boldsymbol{y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$ where $\varepsilon \sim \text{N}(\boldsymbol{0}, \Sigma_h)$.
- **Problem:** Σ_h hard to estimate.

BLUF via trace minimization

Theorem

For any $\boldsymbol{P}$ satisfying $\boldsymbol{SPS} = \boldsymbol{S}$, then

$$\min_{\boldsymbol{P}} = \text{trace}[\boldsymbol{SP}\Sigma_h \boldsymbol{P}'\boldsymbol{S}']$$

has solution $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$.

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts — Base forecasts

- Equivalent to GLS estimate of regression $\hat{\boldsymbol{y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$ where $\varepsilon \sim \text{N}(\boldsymbol{0}, \Sigma_h)$.
- **Problem:** Σ_h hard to estimate.

BLUF via trace minimization

Theorem

For any $\boldsymbol{P}$ satisfying $\boldsymbol{SPS} = \boldsymbol{S}$, then

$$\min_{\boldsymbol{P}} = \text{trace}[\boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}']$$

has solution $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$.

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts   Base forecasts

- Equivalent to GLS estimate of regression $\hat{\boldsymbol{y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$ where $\varepsilon \sim \text{N}(\boldsymbol{0}, \Sigma_h)$.
- **Problem:** Σ_h hard to estimate.

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^\dagger\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts

Base forecasts

Solution 1: OLS

- Assume $\varepsilon_h \approx \boldsymbol{S}\varepsilon_{B,h}$ where $\varepsilon_{B,h}$ is the forecast error at bottom level.
- Then $\Sigma_h \approx \boldsymbol{S}\Omega_h\boldsymbol{S}'$ where $\Omega_h = \text{Var}(\varepsilon_{B,h})$.
- If Moore-Penrose generalized inverse used, then $(\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^\dagger = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts Base forecasts

Solution 1: OLS

- Assume $\varepsilon_h \approx \boldsymbol{S}\varepsilon_{B,h}$ where $\varepsilon_{B,h}$ is the forecast error at bottom level.
- Then $\Sigma_h \approx \boldsymbol{S}\Omega_h\boldsymbol{S}'$ where $\Omega_h = \text{Var}(\varepsilon_{B,h})$.
- If Moore-Penrose generalized inverse used, then $(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts

Base forecasts

Solution 1: OLS

- Assume $\varepsilon_h \approx \boldsymbol{S}\varepsilon_{B,h}$ where $\varepsilon_{B,h}$ is the forecast error at bottom level.
- Then $\Sigma_h \approx \boldsymbol{S}\Omega_h\boldsymbol{S}'$ where $\Omega_h = \mathrm{Var}(\varepsilon_{B,h})$.
- If Moore-Penrose generalized inverse used, then $(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts

Base forecasts

Solution 1: OLS

- Assume $\varepsilon_h \approx \boldsymbol{S}\varepsilon_{B,h}$ where $\varepsilon_{B,h}$ is the forecast error at bottom level.
- Then $\Sigma_h \approx \boldsymbol{S}\Omega_h\boldsymbol{S}'$ where $\Omega_h = \text{Var}(\varepsilon_{B,h})$.
- If Moore-Penrose generalized inverse used, then $(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts Base forecasts

Solution 1: OLS

- Assume $\varepsilon_h \approx \boldsymbol{S}\varepsilon_{B,h}$ where $\varepsilon_{B,h}$ is the forecast error at bottom level.
- Then $\Sigma_h \approx \boldsymbol{S}\Omega_h\boldsymbol{S}'$ where $\Omega_h = \text{Var}(\varepsilon_{B,h})$.
- If Moore-Penrose generalized inverse used, then $(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts Base forecasts

Solution 1: OLS

- Assume $\varepsilon_h \approx \boldsymbol{S}\varepsilon_{B,h}$ where $\varepsilon_{B,h}$ is the forecast error at bottom level.
- Then $\Sigma_h \approx \boldsymbol{S}\Omega_h\boldsymbol{S}'$ where $\Omega_h = \text{Var}(\varepsilon_{B,h})$.
- If Moore-Penrose generalized inverse used, then $(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\color{red}{\tilde{\boldsymbol{y}}_n(h)} = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\color{blue}{\hat{\boldsymbol{y}}_n(h)}$$

Revised forecasts

Base forecasts

Solution 2: WLS

- Suppose we approximate Σ_1 by its diagonal.
- Easy to estimate, and places weight where we have best forecasts.
- Empirically, it gives better forecasts than other available methods.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts — Base forecasts

Solution 2: WLS

- Suppose we approximate Σ_1 by its diagonal.
- Easy to estimate, and places weight where we have best forecasts.
- Empirically, it gives better forecasts than other available methods.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts Base forecasts

Solution 2: WLS

- Suppose we approximate Σ_1 by its diagonal.
- Easy to estimate, and places weight where we have best forecasts.
- Empirically, it gives better forecasts than other available methods.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts   Base forecasts

Solution 2: WLS

- Suppose we approximate Σ_1 by its diagonal.
- Easy to estimate, and places weight where we have best forecasts.
- Empirically, it gives better forecasts than other available methods.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts

Base forecasts

Solution 2: WLS

- Suppose we approximate Σ_1 by its diagonal.
- Easy to estimate, and places weight where we have best forecasts.
- Empirically, it gives better forecasts than other available methods.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda\hat{\boldsymbol{y}}_n(h)$$

Optimal combination forecasts

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}\hat{\boldsymbol{y}}_n(h)$$

Revised forecasts   Base forecasts

Solution 2: WLS

- Suppose we approximate Σ_1 by its diagonal.
- Easy to estimate, and places weight where we have best forecasts.
- Empirically, it gives better forecasts than other available methods.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda\hat{\boldsymbol{y}}_n(h)$$

Challenges

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda\hat{\boldsymbol{y}}_n(h)$$

- Computational difficulties in big hierarchies due to size of the $\boldsymbol{S}$ matrix and singular behavior of $(\boldsymbol{S}'\Lambda\boldsymbol{S})$.
- Loss of information in ignoring covariance matrix in computing point forecasts.
- Still need to estimate covariance matrix to produce prediction intervals.

Challenges

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda\hat{\boldsymbol{y}}_n(h)$$

- Computational difficulties in big hierarchies due to size of the $\boldsymbol{S}$ matrix and singular behavior of $(\boldsymbol{S}'\Lambda\boldsymbol{S})$.
- Loss of information in ignoring covariance matrix in computing point forecasts.
- Still need to estimate covariance matrix to produce prediction intervals.

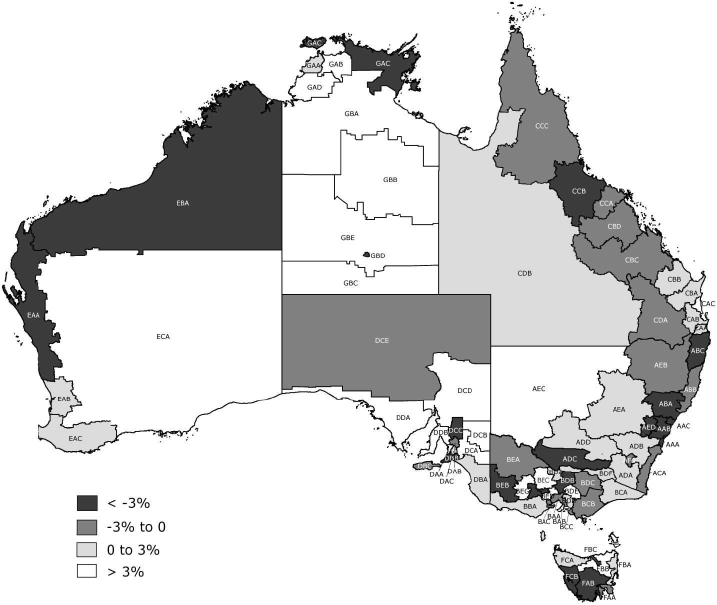
Challenges

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda\hat{\boldsymbol{y}}_n(h)$$

- Computational difficulties in big hierarchies due to size of the $\boldsymbol{S}$ matrix and singular behavior of $(\boldsymbol{S}'\Lambda\boldsymbol{S})$.
- Loss of information in ignoring covariance matrix in computing point forecasts.
- Still need to estimate covariance matrix to produce prediction intervals.

Outline

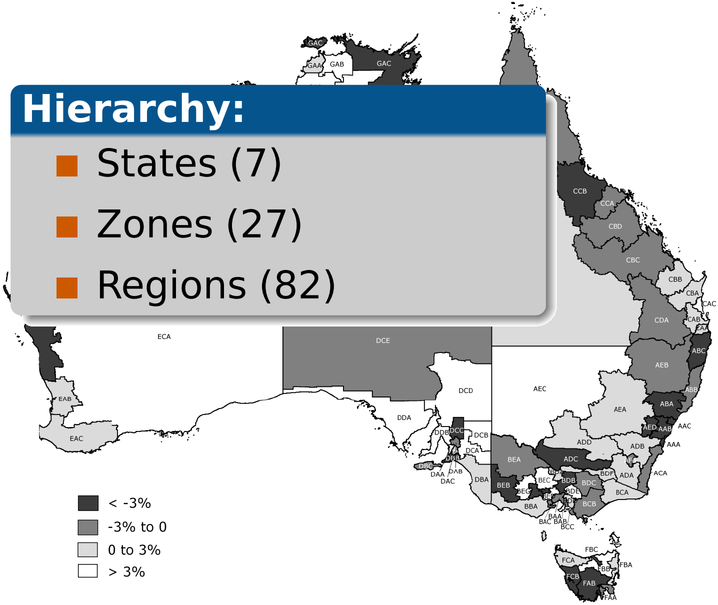
Australian tourism



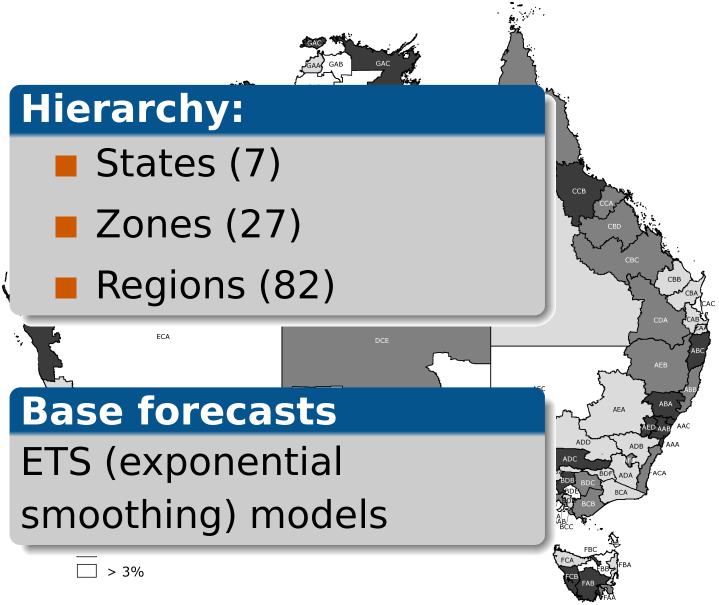
Australian tourism

Hierarchy:

- States (7)
- Zones (27)
- Regions (82)



Australian tourism



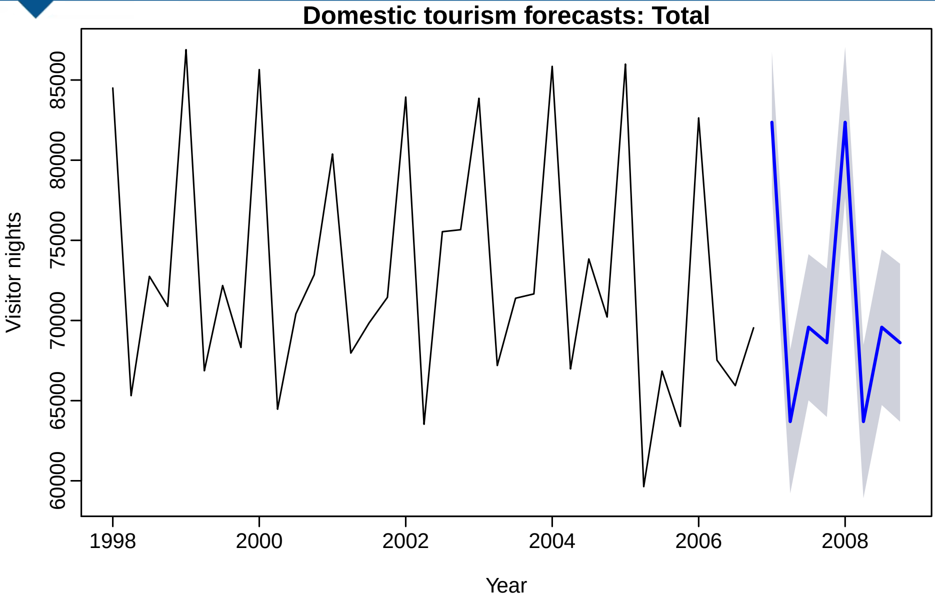
Hierarchy:

- States (7)
- Zones (27)
- Regions (82)

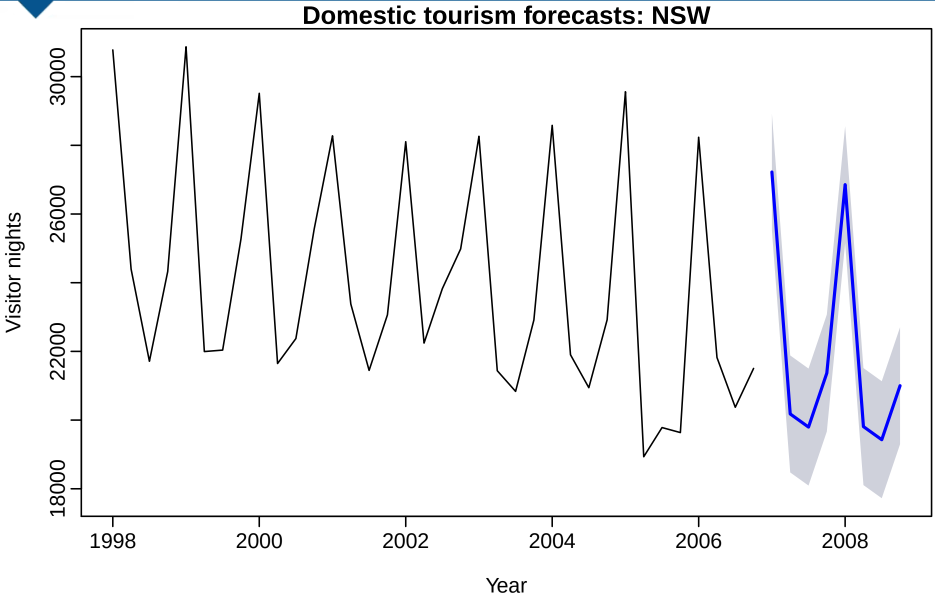
Base forecasts

ETS (exponential smoothing) models

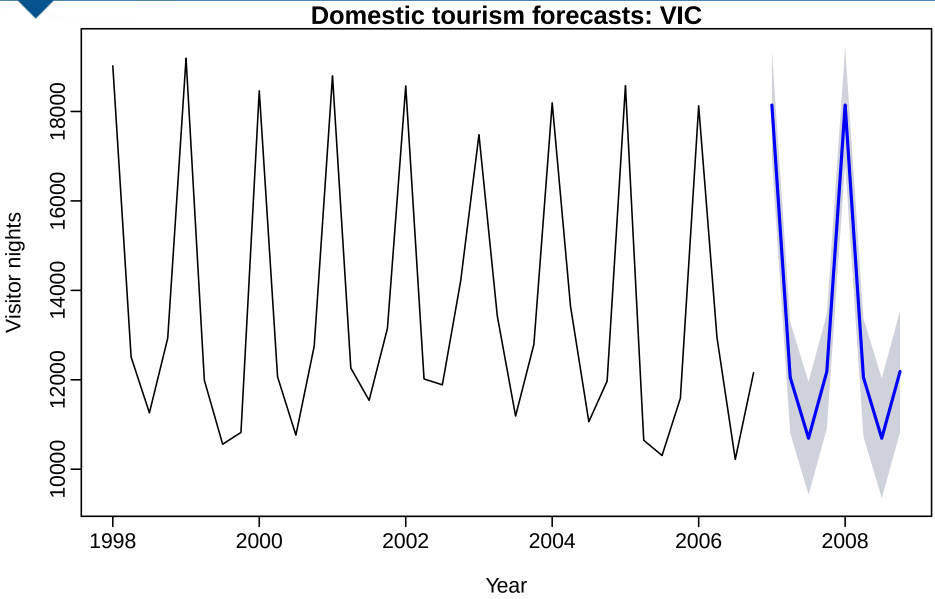
Base forecasts



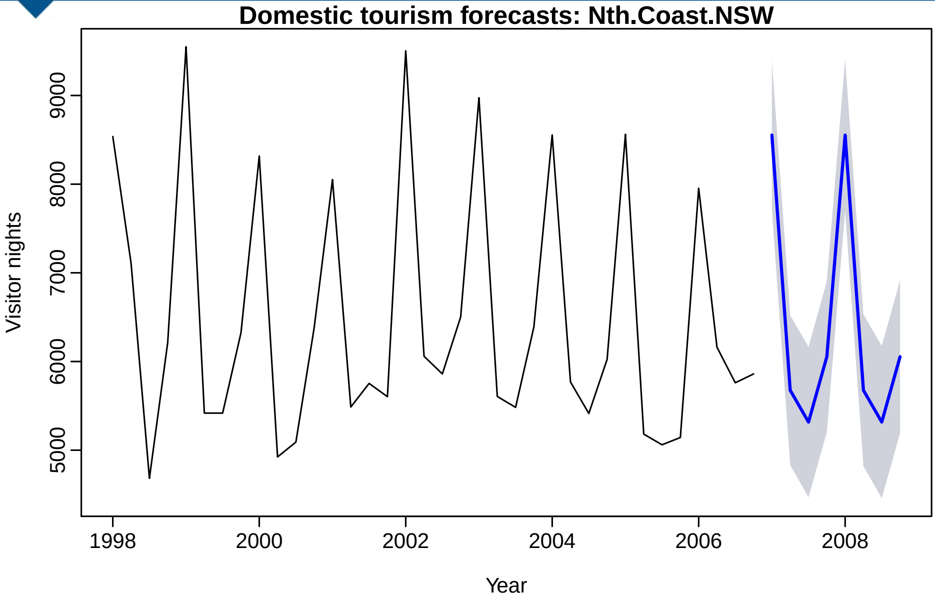
Base forecasts



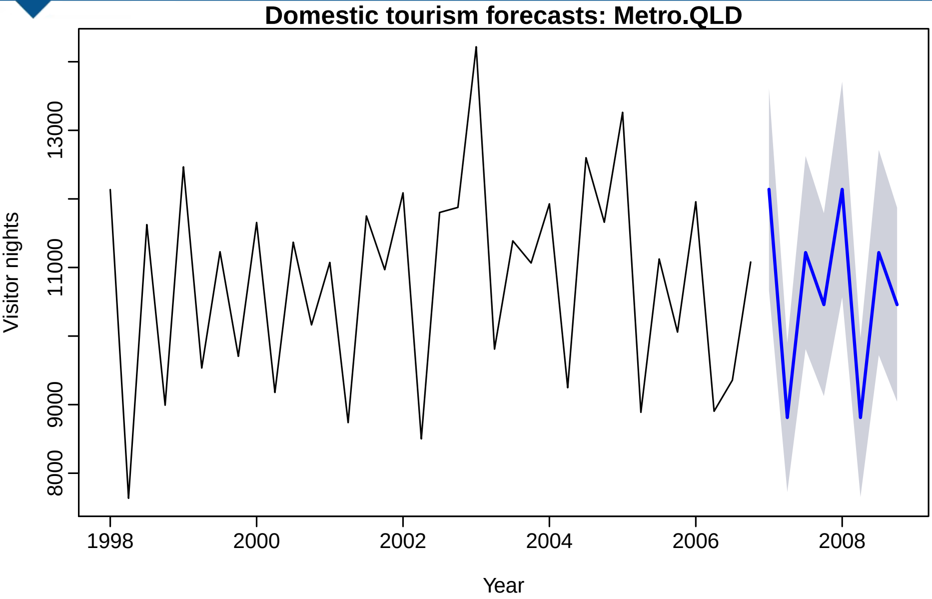
Base forecasts



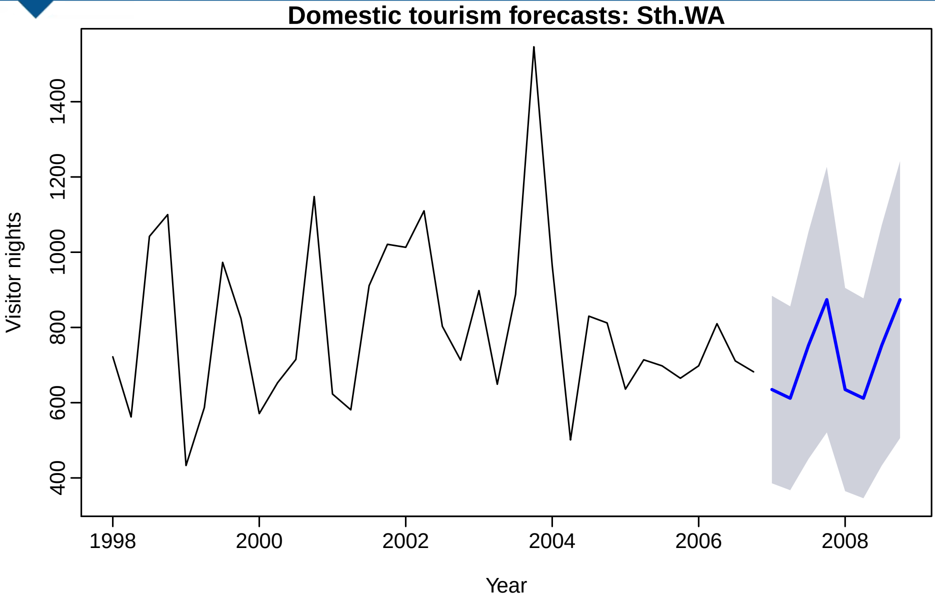
Base forecasts



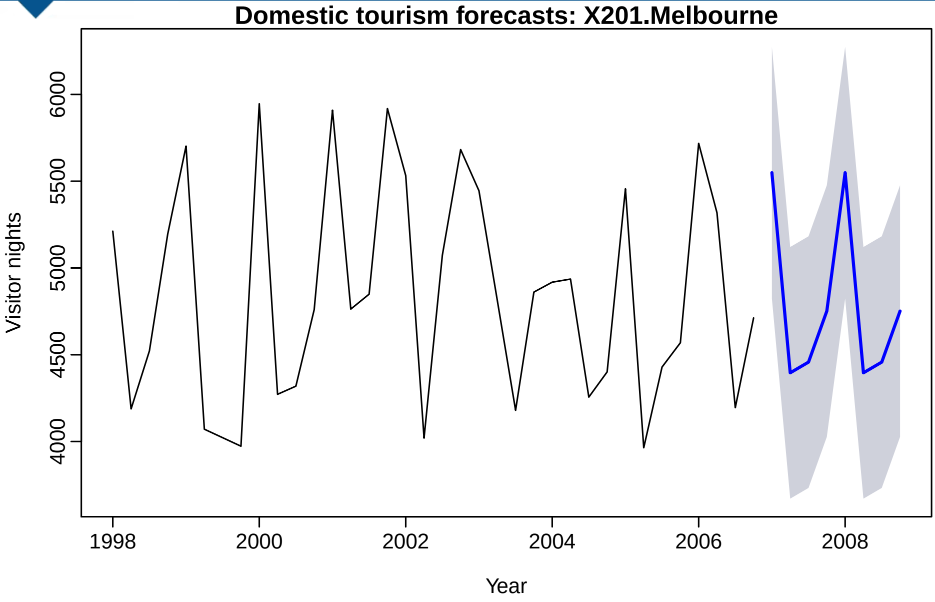
Base forecasts



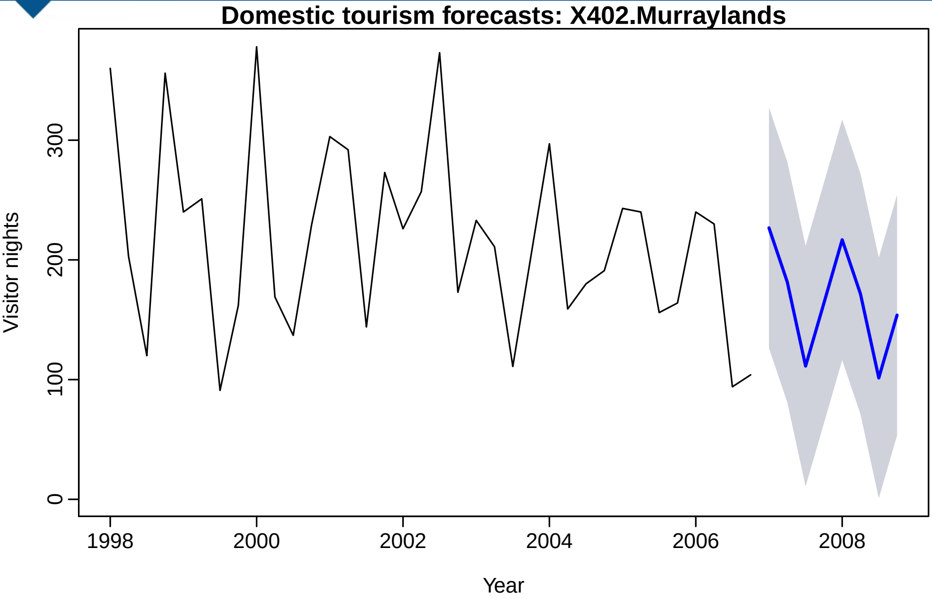
Base forecasts



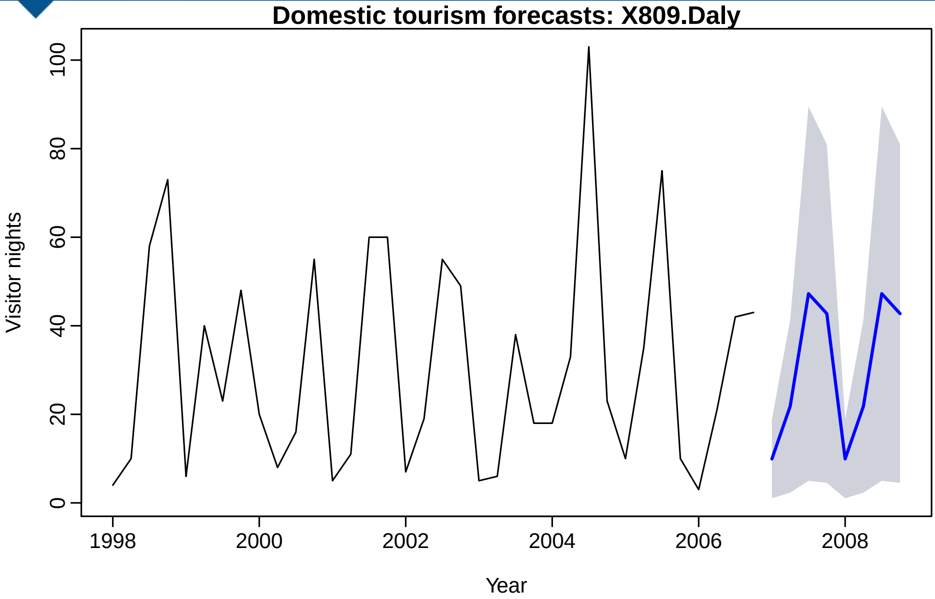
Base forecasts



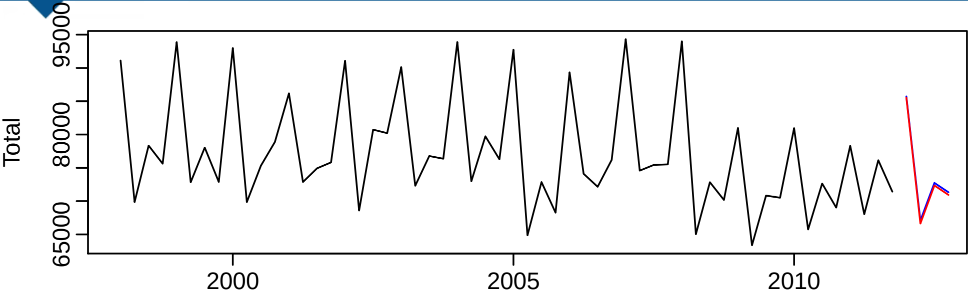
Base forecasts



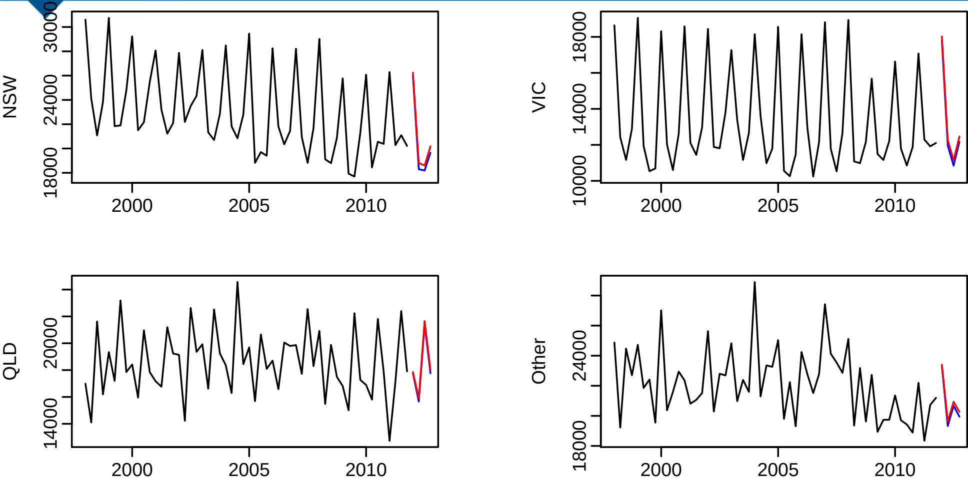
Base forecasts



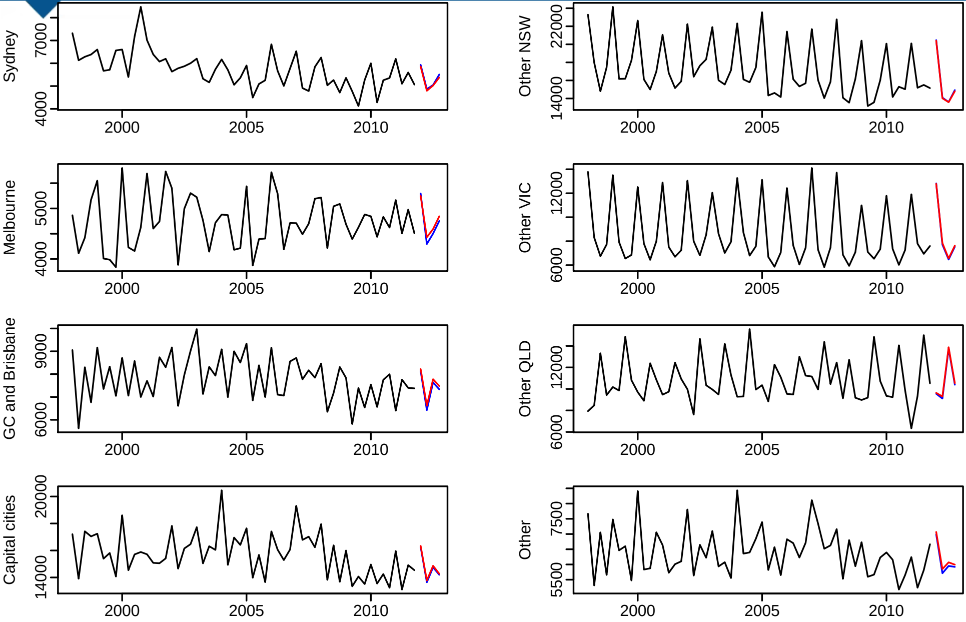
Reconciled forecasts



Reconciled forecasts



Reconciled forecasts



Forecast evaluation

- Select models using all observations;
- Re-estimate models using first 12 observations and generate 1- to 8-step-ahead forecasts;
- Increase sample size one observation at a time, re-estimate models, generate forecasts until the end of the sample;
- In total 24 1-step-ahead, 23 2-steps-ahead, up to 17 8-steps-ahead for forecast evaluation.

Forecast evaluation

- Select models using all observations;
- Re-estimate models using first 12 observations and generate 1- to 8-step-ahead forecasts;
- Increase sample size one observation at a time, re-estimate models, generate forecasts until the end of the sample;
- In total 24 1-step-ahead, 23 2-steps-ahead, up to 17 8-steps-ahead for forecast evaluation.

Forecast evaluation

- Select models using all observations;
- Re-estimate models using first 12 observations and generate 1- to 8-step-ahead forecasts;
- Increase sample size one observation at a time, re-estimate models, generate forecasts until the end of the sample;
- In total 24 1-step-ahead, 23 2-steps-ahead, up to 17 8-steps-ahead for forecast evaluation.

Forecast evaluation

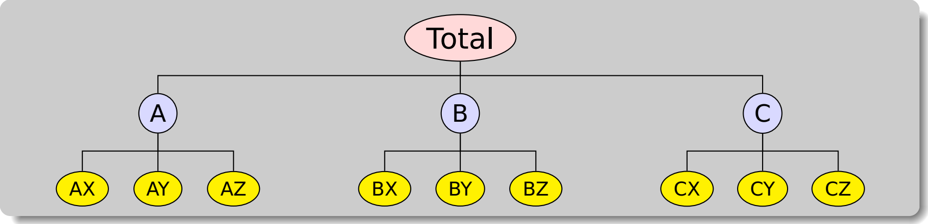
- Select models using all observations;
- Re-estimate models using first 12 observations and generate 1- to 8-step-ahead forecasts;
- Increase sample size one observation at a time, re-estimate models, generate forecasts until the end of the sample;
- In total 24 1-step-ahead, 23 2-steps-ahead, up to 17 8-steps-ahead for forecast evaluation.

Hierarchy: states, zones, regions

MAPE	$h = 1$	$h = 2$	$h = 4$	$h = 6$	$h = 8$	Average
Top Level: Australia						
Bottom-up	3.79	3.58	4.01	4.55	4.24	4.06
OLS	3.83	3.66	**3.88**	**4.19**	4.25	**3.94**
WLS	**3.68**	**3.56**	3.97	4.57	4.25	4.04
Level: States						
Bottom-up	10.70	10.52	10.85	11.46	11.27	11.03
OLS	11.07	10.58	11.13	11.62	12.21	11.35
WLS	**10.44**	**10.17**	**10.47**	**10.97**	**10.98**	**10.67**
Level: Zones						
Bottom-up	14.99	14.97	14.98	15.69	15.65	15.32
OLS	15.16	15.06	15.27	15.74	16.15	15.48
WLS	**14.63**	**14.62**	**14.68**	**15.17**	**15.25**	**14.94**
Bottom Level: Regions						
Bottom-up	33.12	32.54	32.26	33.74	33.96	33.18
OLS	35.89	33.86	34.26	36.06	37.49	35.43
WLS	**31.68**	**31.22**	**31.08**	**32.41**	**32.77**	**31.89**

Outline

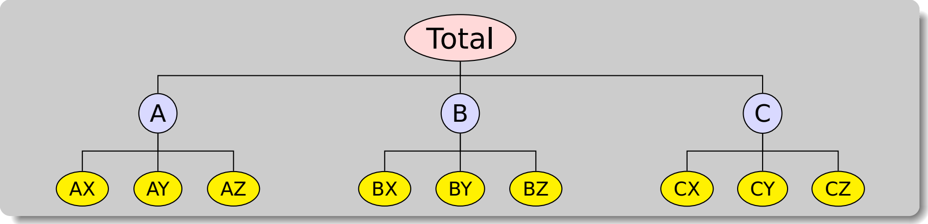
Fast computation: hierarchical data



$$\boldsymbol{y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{B}_t$$

Fast computation: hierarchical data

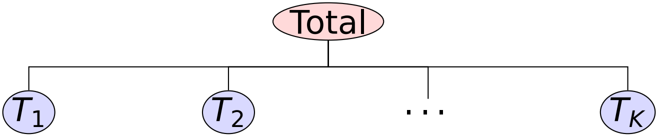


$$\boldsymbol{y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{B,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{C,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{B}_t$$

Fast computation: hierarchies

Think of the hierarchy as a tree of trees:



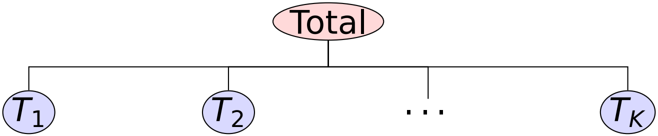
Then the summing matrix contains k smaller summing matrices:

$$\boldsymbol{S} = \begin{bmatrix} \mathbf{1}'_{n_1} & \mathbf{1}'_{n_2} & \cdots & \mathbf{1}'_{n_K} \\ \boldsymbol{S}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{S}_K \end{bmatrix}$$

where $\mathbf{1}_n$ is an n-vector of ones and tree T_i has n_i terminal nodes.

Fast computation: hierarchies

Think of the hierarchy as a tree of trees:



Then the summing matrix contains k smaller summing matrices:

$$\boldsymbol{S} = \begin{bmatrix} \mathbf{1}'_{n_1} & \mathbf{1}'_{n_2} & \cdots & \mathbf{1}'_{n_K} \\ \boldsymbol{S}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{S}_K \end{bmatrix}$$

where $\mathbf{1}_n$ is an n-vector of ones and tree T_i has n_i terminal nodes.

Fast computation: hierarchies

$$\boldsymbol{S}'\Lambda\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_1'\Lambda_1\boldsymbol{S}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_2'\Lambda_2\boldsymbol{S}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{S}_K'\Lambda_K\boldsymbol{S}_K \end{bmatrix} + \lambda_0\boldsymbol{J}_n$$

- λ_0 is the top left element of Λ;
- Λ_k is a block of Λ, corresponding to tree T_k;
- $\boldsymbol{J}_n$ is a matrix of ones;
- $n = \sum_k n_k$.

Now apply the Sherman-Morrison formula . . .

Fast computation: hierarchies

$$\boldsymbol{S}'\Lambda\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_1'\Lambda_1\boldsymbol{S}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_2'\Lambda_2\boldsymbol{S}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{S}_K'\Lambda_K\boldsymbol{S}_K \end{bmatrix} + \lambda_0\boldsymbol{J}_n$$

- λ_0 is the top left element of Λ;
- Λ_k is a block of Λ, corresponding to tree T_k;
- $\boldsymbol{J}_n$ is a matrix of ones;
- $n = \sum_k n_k$.

Now apply the Sherman-Morrison formula . . .

Fast computation: hierarchies

$$(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1} = \begin{bmatrix} (\boldsymbol{S}_1'\Lambda_1\boldsymbol{S}_1)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{S}_2'\Lambda_2\boldsymbol{S}_2)^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\boldsymbol{S}_K'\Lambda_K\boldsymbol{S}_K)^{-1} \end{bmatrix} - c\boldsymbol{S}_0$$

- $\boldsymbol{S}_0$ can be partitioned into K^2 blocks, with the (k,ℓ) block (of dimension $n_k \times n_\ell$) being

$$(\boldsymbol{S}_k'\Lambda_k\boldsymbol{S}_k)^{-1}\boldsymbol{J}_{n_k,n_\ell}(\boldsymbol{S}_\ell'\Lambda_\ell\boldsymbol{S}_\ell)^{-1}$$

- $\boldsymbol{J}_{n_k,n_\ell}$ is a $n_k \times n_\ell$ matrix of ones.
- $c^{-1} = \lambda_0^{-1} + \sum_k \mathbf{1}_{n_k}'(\boldsymbol{S}_k'\Lambda_k\boldsymbol{S}_k)^{-1}\mathbf{1}_{n_k}$.
- Each $\boldsymbol{S}_k'\Lambda_k\boldsymbol{S}_k$ can be inverted similarly.
- $\boldsymbol{S}'\Lambda\boldsymbol{y}$ can also be computed recursively.

Fast computation: hierarchies

$$(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1} = \begin{bmatrix} (\boldsymbol{S}_1'\Lambda_1\boldsymbol{S}_1)^{-1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{S}_2'\Lambda_2\boldsymbol{S}_2)^{-1} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & (\boldsymbol{S}_K'\Lambda_K\boldsymbol{S}_K)^{-1} \end{bmatrix} - c\boldsymbol{S}_0$$

- $\boldsymbol{S}_0$ can be partitioned into K^2 blocks, with the (k,ℓ) block (of dimension $n_k \times n_\ell$) being
- $\boldsymbol{J}_{n_k,n_\ell}$
- c^{-1}
- Each $\boldsymbol{S}_k'\Lambda_k\boldsymbol{S}_k$ can be inverted similarly.
- $\boldsymbol{S}'\Lambda\boldsymbol{y}$ can also be computed recursively.

The recursive calculations can be done in such a way that we never store any of the large matrices involved.

Fast computation

A similar algorithm has been developed for grouped time series with two groups.

When the time series are not strictly hierarchical and have more than two grouping variables:

- Use sparse matrix storage and arithmetic.
- Use iterative approximation for inverting large sparse matrices.

Paige & Saunders (1982)
ACM Trans. Math. Software

Fast computation

A similar algorithm has been developed for grouped time series with two groups.

When the time series are not strictly hierarchical and have more than two grouping variables:

- Use sparse matrix storage and arithmetic.
- Use iterative approximation for inverting large sparse matrices.

Paige & Saunders (1982)
ACM Trans. Math. Software

Fast computation

A similar algorithm has been developed for grouped time series with two groups.

When the time series are not strictly hierarchical and have more than two grouping variables:

- Use sparse matrix storage and arithmetic.
- Use iterative approximation for inverting large sparse matrices.

Paige & Saunders (1982)
ACM Trans. Math. Software

Outline

hts package for R

hts: Hierarchical and grouped time series

Methods for analysing and forecasting hierarchical and grouped time series

Version:	4.5
Depends:	forecast ($\geq$ 5.0), SparseM
Imports:	parallel, utils
Published:	2014-12-09
Author:	Rob J Hyndman, Earo Wang and Alan Lee
Maintainer:	Rob J Hyndman $<$Rob.Hyndman at monash.edu$>$
BugReports:	`https://github.com/robjhyndman/hts/issues`
License:	GPL ($\geq$ 2)

Example using R

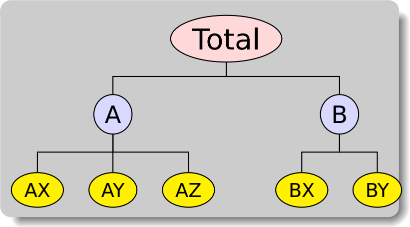
```
library(hts)

# bts is a matrix containing the bottom level time series
# nodes describes the hierarchical structure
y <- hts(bts, nodes=list(2, c(3,2)))
```

Example using R

```
library(hts)

# bts is a matrix containing the bottom level time series
# nodes describes the hierarchical structure
y <- hts(bts, nodes=list(2, c(3,2)))
```



Example using R

```
library(hts)

# bts is a matrix containing the bottom level time series
# nodes describes the hierarchical structure
y <- hts(bts, nodes=list(2, c(3,2)))

# Forecast 10-step-ahead using WLS combination method
# ETS used for each series by default
fc <- forecast(y, h=10)
```

forecast.gts function

Usage

```
forecast(object, h,
  method = c("comb", "bu", "mo", "tdgsf", "tdgsa", "tdfp"),
  fmethod = c("ets", "rw", "arima"),
  weights = c("sd", "none", "nseries"),
  positive = FALSE,
  parallel = FALSE, num.cores = 2, ...)
```

Arguments

`object`	Hierarchical time series object of class `gts`.
`h`	Forecast horizon
`method`	Method for distributing forecasts within the hierarchy.
`fmethod`	Forecasting method to use
`positive`	If TRUE, forecasts are forced to be strictly positive
`weights`	Weights used for "optimal combination" method. When weights = "sd", it takes account of the standard deviation of forecasts.
`parallel`	If TRUE, allow parallel processing
`num.cores`	If parallel = TRUE, specify how many cores are going to be used

Outline

1. Examples of biggish time series
2. Time series visualisation
3. BLUF: Best Linear Unbiased Forecasts
4. Application: Australian tourism
5. Fast computation tricks
6. hts package for R
7. **References**

References

RJ Hyndman, RA Ahmed, G Athanasopoulos, and HL Shang (2011). "Optimal combination forecasts for hierarchical time series". *Computational statistics & data analysis* **55**(9), 2579–2589.

RJ Hyndman, AJ Lee, and E Wang (2014). *Fast computation of reconciled forecasts for hierarchical and grouped time series*. Working paper 17/14. Department of Econometrics & Business Statistics, Monash University

RJ Hyndman, AJ Lee, and E Wang (2014). *hts: Hierarchical and grouped time series*. `cran.r-project.org/package=hts`.

RJ Hyndman and G Athanasopoulos (2014). *Forecasting: principles and practice*. OTexts. `OTexts.org/fpp/`.

References

RJ Hyndman, RA Ahmed, G Athanasopoulos, and HL Shang (2011). "Optimal combination forecasts for hierarchical time series". *Computational statistics & data analysis* **55**(9), 2579–2589.

RJ Hyndman, AJ Lee, and E Wang (2014). *Fast computation of reconciled forecasts for hierarchical and grouped time series*. Working paper 17/14. Department of Econometrics & Business Statistics, Monash University

RJ Hyndman, AJ Lee, and E Wang (2014). *hts:*

- Papers and R code:
  robjhyndman.com
- Email: **Rob.Hyndman@monash.edu**