Rob J Hyndman

Forecasting:

Principles and Practice

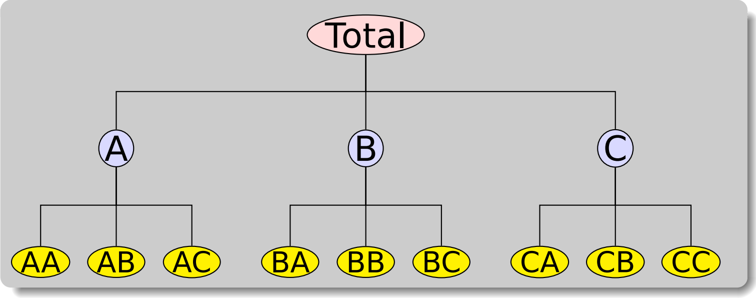
11. Hierarchical forecasting

OTexts.com/fpp/9/4/

Outline

1. **Hierarchical and grouped time series**
2. Forecasting framework
3. Optimal forecasts
4. Approximately optimal forecasts
5. Application: Australian tourism
6. Application: Australian labour market
7. hts package for R

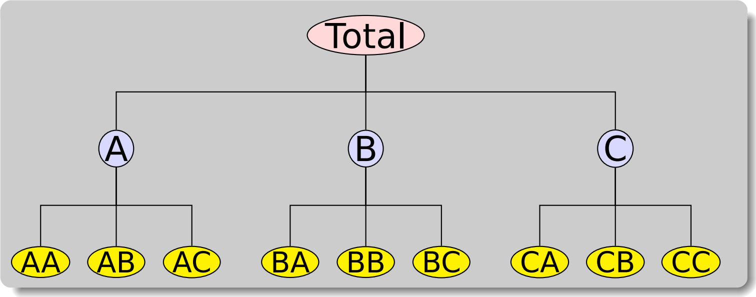
Introduction



Examples

- Manufacturing product hierarchies
- Net labour turnover
- Pharmaceutical sales
- Tourism demand by region and purpose

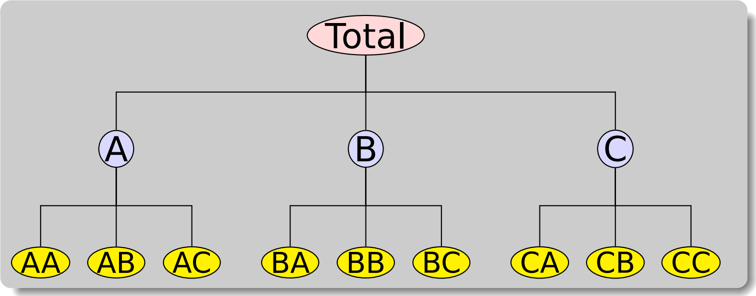
Introduction



Examples

- Manufacturing product hierarchies
- Net labour turnover
- Pharmaceutical sales
- Tourism demand by region and purpose

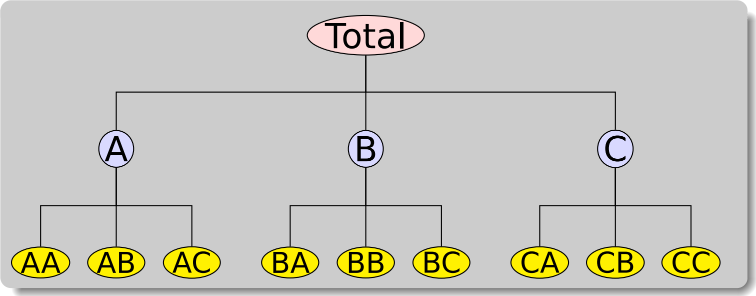
Introduction



Examples

- Manufacturing product hierarchies
- Net labour turnover
- Pharmaceutical sales
- Tourism demand by region and purpose

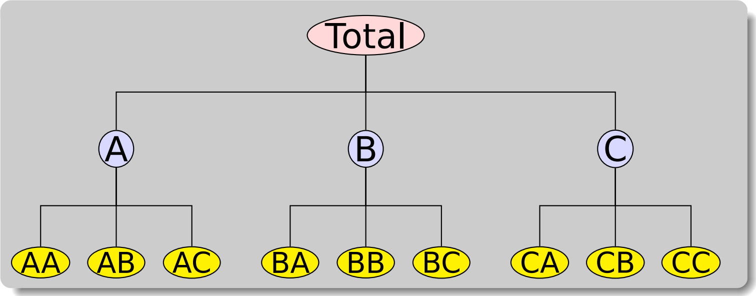
Introduction



Examples

- Manufacturing product hierarchies
- Net labour turnover
- Pharmaceutical sales
- Tourism demand by region and purpose

Introduction



Examples

- Manufacturing product hierarchies
- Net labour turnover
- Pharmaceutical sales
- Tourism demand by region and purpose

Forecasting the PBS

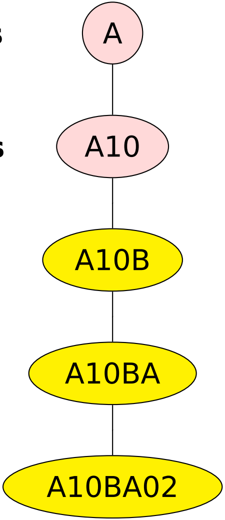
ATC drug classification

A	Alimentary tract and metabolism
B	Blood and blood forming organs
C	Cardiovascular system
D	Dermatologicals
G	Genito-urinary system and sex hormones
H	Systemic hormonal preparations, excluding sex hormones and insulins
J	Anti-infectives for systemic use
L	Antineoplastic and immunomodulating agents
M	Musculo-skeletal system
N	Nervous system
P	Antiparasitic products, insecticides and repellents
R	Respiratory system
S	Sensory organs
V	Various

ATC drug classification

14 classes — Alimentary tract and metabolism

84 classes — Drugs used in diabetes

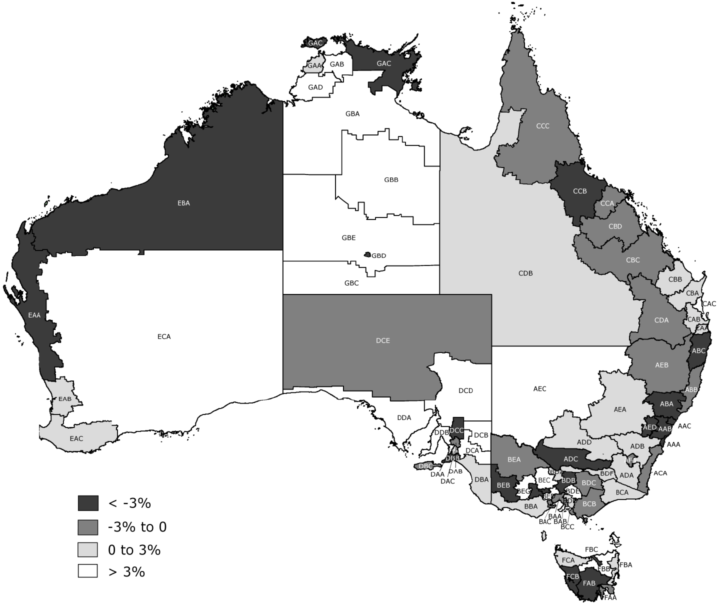


Blood glucose lowering drugs

Biguanides

Metformin

Australian tourism



Australian tourism

Also split by purpose of travel:

- Holiday
- Visits to friends and relatives
- Business
- Other

Hierarchical/grouped time series

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

Example: Pharmaceutical products are organized in a hierarchy under the Anatomical Therapeutic Chemical (ATC) Classification System.

- A **grouped time series** is a collection of time series that are aggregated in a number of non-hierarchical ways.

Example: Australian tourism demand is grouped by region and purpose of travel.

Hierarchical/grouped time series

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

Example: Pharmaceutical products are organized in a hierarchy under the Anatomical Therapeutic Chemical (ATC) Classification System.

- A **grouped time series** is a collection of time series that are aggregated in a number of non-hierarchical ways.

Example: Australian tourism demand is grouped by region and purpose of travel.

Hierarchical/grouped time series

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

Example: Pharmaceutical products are organized in a hierarchy under the Anatomical Therapeutic Chemical (ATC) Classification System.

- A **grouped time series** is a collection of time series that are aggregated in a number of non-hierarchical ways.

Example: Australian tourism demand is grouped by region and purpose of travel.

Hierarchical/grouped time series

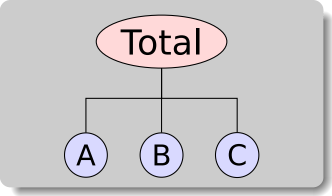
- A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.

Example: Pharmaceutical products are organized in a hierarchy under the Anatomical Therapeutic Chemical (ATC) Classification System.

- A **grouped time series** is a collection of time series that are aggregated in a number of non-hierarchical ways.

Example: Australian tourism demand is grouped by region and purpose of travel.

Hierarchical data

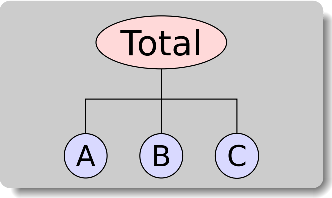


Y_t : observed aggregate of all series at time t.

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

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

Hierarchical data

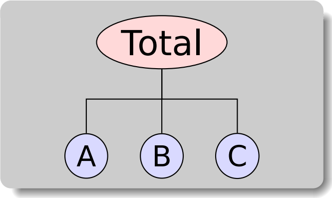


Y_t : observed aggregate of all series at time t.

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

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

Hierarchical data



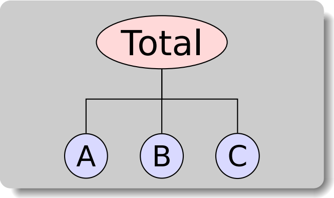
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: 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 data



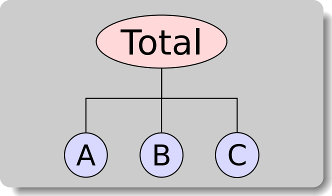
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: 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 data



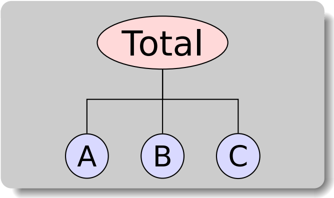
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: 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 data



Y_t : observed aggregate of all series at time t.

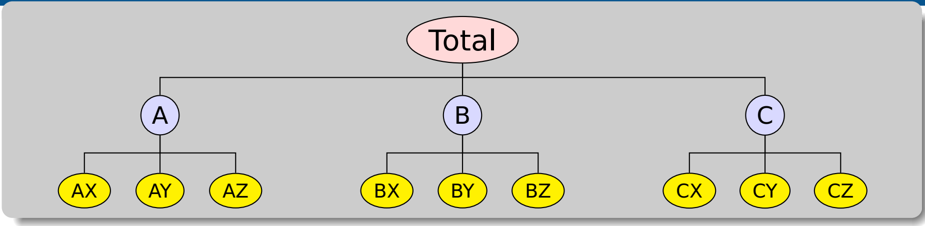
$Y_{X,t}$: observation on series X at time t.

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

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

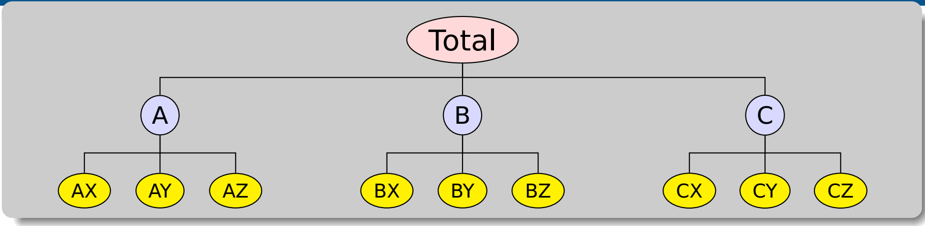
$$\boxed{\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t}$$

Hierarchical data

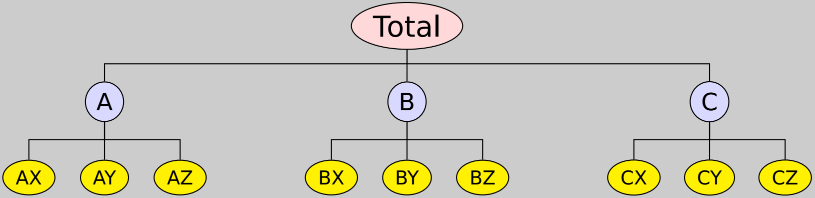


$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0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

Hierarchical data



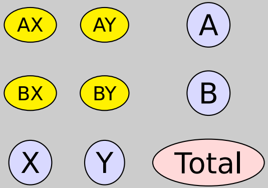
$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0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



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

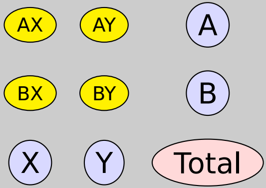
$$\boxed{\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t}$$

Grouped data



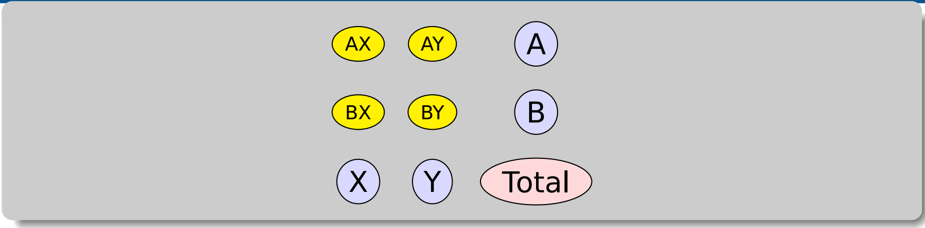
$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{X,t} \\ Y_{Y,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{BX,t} \\ Y_{BY,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{AX,t} \\ Y_{AY,t} \\ Y_{BX,t} \\ Y_{BY,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

Grouped data



$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{X,t} \\ Y_{Y,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{BX,t} \\ Y_{BY,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{AX,t} \\ Y_{AY,t} \\ Y_{BX,t} \\ Y_{BY,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

Grouped data



$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{X,t} \\ Y_{Y,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{BX,t} \\ Y_{BY,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{AX,t} \\ Y_{AY,t} \\ Y_{BX,t} \\ Y_{BY,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

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

Outline

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.)

Hierarchical forecasting methods 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.)

Hierarchical forecasting methods 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
- Revised reconciled forecasts: $\tilde{\boldsymbol{Y}}_n(h)$.

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.)

Hierarchical forecasting methods 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
- Revised reconciled forecasts: $\tilde{\boldsymbol{Y}}_n(h)$.

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.)

Hierarchical forecasting methods 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
- Revised reconciled forecasts: $\tilde{\boldsymbol{Y}}_n(h)$.

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.)

Hierarchical forecasting methods 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
- Revised reconciled forecasts: $\tilde{\boldsymbol{Y}}_n(h)$.

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.)

Hierarchical forecasting methods of the form:

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

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
- Revised reconciled forecasts: $\tilde{\boldsymbol{Y}}_n(h)$.

Bottom-up forecasts

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

Bottom-up forecasts are obtained using

$$\boldsymbol{P} = [\mathbf{0} \mid \mathbf{I}],$$

where $\mathbf{0}$ is null matrix and $\mathbf{I}$ is identity matrix.

- $\boldsymbol{P}$ matrix extracts only bottom-level forecasts from $\hat{\boldsymbol{Y}}_n(h)$
- $\boldsymbol{S}$ adds them up to give the bottom-up forecasts.

Bottom-up forecasts

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

Bottom-up forecasts are obtained using

$$\boldsymbol{P} = [\mathbf{0} \mid \mathbf{I}] \, ,$$

where $\mathbf{0}$ is null matrix and $\mathbf{I}$ is identity matrix.

- $\boldsymbol{P}$ matrix extracts only bottom-level forecasts from $\hat{\boldsymbol{Y}}_n(h)$
- $\boldsymbol{S}$ adds them up to give the bottom-up forecasts.

Bottom-up forecasts

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

Bottom-up forecasts are obtained using

$$\boldsymbol{P} = [\mathbf{0} \mid \mathbf{I}]\,,$$

where $\mathbf{0}$ is null matrix and $\mathbf{I}$ is identity matrix.

- $\boldsymbol{P}$ matrix extracts only bottom-level forecasts from $\hat{\boldsymbol{Y}}_n(h)$
- $\boldsymbol{S}$ adds them up to give the bottom-up forecasts.

Top-down forecasts

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

Top-down forecasts are obtained using

$$\boldsymbol{P} = [\boldsymbol{p} \mid \boldsymbol{0}]$$

where $\boldsymbol{p} = [p_1, p_2, \dots, p_{m_K}]'$ is a vector of proportions that sum to one.

- $\boldsymbol{P}$ distributes forecasts of the aggregate to the lowest level series.
- Different methods of top-down forecasting lead to different proportionality vectors $\boldsymbol{p}$.

Top-down forecasts

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

Top-down forecasts are obtained using

$$\boldsymbol{P} = [\boldsymbol{p} \mid \boldsymbol{0}]$$

where $\boldsymbol{p} = [p_1, p_2, \dots, p_{m_K}]'$ is a vector of proportions that sum to one.

- $\boldsymbol{P}$ distributes forecasts of the aggregate to the lowest level series.
- Different methods of top-down forecasting lead to different proportionality vectors $\boldsymbol{p}$.

Top-down forecasts

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

Top-down forecasts are obtained using

$$\boldsymbol{P} = [\boldsymbol{p} \mid \boldsymbol{0}]$$

where $\boldsymbol{p} = [p_1, p_2, \dots, p_{m_K}]'$ is a vector of proportions that sum to one.

- $\boldsymbol{P}$ distributes forecasts of the aggregate to the lowest level series.
- Different methods of top-down forecasting lead to different proportionality vectors $\boldsymbol{p}$.

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: $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{SPS}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.
- Result will hold provided $\boldsymbol{SPS} = \boldsymbol{S}$.
- True for bottom-up, but not for any top-down method or middle-out method

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: $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{SPS}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.
- Result will hold provided $\boldsymbol{SPS} = \boldsymbol{S}$.
- True for bottom-up, but not for any top-down method or middle-out method

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: $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{SPS}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.
- Result will hold provided $\boldsymbol{SPS} = \boldsymbol{S}$.
- True for bottom-up, but not for any top-down method or middle-out method

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: $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{S}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.
- Result will hold provided $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$.
- True for bottom-up, but not for any top-down method or middle-out method.

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: $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{SPS}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.
- Result will hold provided $\boldsymbol{SPS} = \boldsymbol{S}$.
- True for bottom-up, but not for any top-down method or middle-out method

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: $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{SPS}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.
- Result will hold provided $\boldsymbol{SPS} = \boldsymbol{S}$.
- True for bottom-up, but not for *any* top-down method or middle-out method.

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: $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{SPS}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.
- Result will hold provided $\boldsymbol{SPS} = \boldsymbol{S}$.
- True for bottom-up, but not for *any* top-down method or middle-out method.

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: $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{SPS}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.
- Result will hold provided $\boldsymbol{SPS} = \boldsymbol{S}$.
- True for bottom-up, but not for *any* top-down method or middle-out method.

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

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

Then the variance of the revised forecasts is given by

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

This is a general result for all existing methods.

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

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

This is a general result for all existing methods.

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

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

This is a general result for all existing methods.

Outline

Forecasts

Key idea: forecast reconciliation

- Ignore structural constraints and forecast every series of interest independently.
- Adjust forecasts to impose constraints.

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

$\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t$. So $\hat{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$.

Forecasts

Key idea: forecast reconciliation

- Ignore structural constraints and forecast every series of interest independently.
- Adjust forecasts to impose constraints.

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

$\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t$. So $\hat{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$.

- $\beta_n(h) = \mathrm{E}[\boldsymbol{B}_{n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.

Forecasts

Key idea: forecast reconciliation

- Ignore structural constraints and forecast every series of interest independently.
- Adjust forecasts to impose constraints.

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

$\boxed{\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t}$. So $\hat{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$.

- $\beta_n(h) = \mathrm{E}[\boldsymbol{B}_{n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.

Forecasts

Key idea: forecast reconciliation

- ➥ Ignore structural constraints and forecast every series of interest independently.
- ➥ Adjust forecasts to impose constraints.

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

$\boxed{\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t}$. So $\boxed{\hat{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h}$.

- $\beta_n(h) = \mathrm{E}[\boldsymbol{B}_{n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.
- ε_h has zero mean and covariance Σ_h.
- Estimate $\beta_n(h)$ using GLS?

Forecasts

Key idea: forecast reconciliation

- ➥ Ignore structural constraints and forecast every series of interest independently.
- ➥ Adjust forecasts to impose constraints.

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

$\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t$. So $\hat{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$.

- $\beta_n(h) = \mathrm{E}[\boldsymbol{B}_{n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.
- ε_h has zero mean and covariance Σ_h.
- Estimate $\beta_n(h)$ using GLS?

Forecasts

Key idea: forecast reconciliation

- ➥ Ignore structural constraints and forecast every series of interest independently.
- ➥ Adjust forecasts to impose constraints.

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

$\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t$. So $\hat{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$.

- $\beta_n(h) = \mathrm{E}[\boldsymbol{B}_{n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.
- ε_h has zero mean and covariance Σ_h.
- Estimate $\beta_n(h)$ using GLS?

Forecasts

Key idea: forecast reconciliation

- Ignore structural constraints and forecast every series of interest independently.
- Adjust forecasts to impose constraints.

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

$\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t$. So $\hat{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\beta_n(h) + \varepsilon_h$.

- $\beta_n(h) = \mathrm{E}[\boldsymbol{B}_{n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.
- ε_h has zero mean and covariance Σ_h.
- Estimate $\beta_n(h)$ using GLS?

Optimal combination forecasts

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

- $\Sigma_h^\dagger$ is generalized inverse of Σ_h.
- $\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'$
- **Problem:** Σ_h hard to estimate.

Optimal combination forecasts

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

Initial forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- $\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'$
- **Problem:** Σ_h hard to estimate.

Optimal combination forecasts

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

Revised forecasts

Initial forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- $\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'$
- **Problem:** Σ_h hard to estimate.

Optimal combination forecasts

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

Revised forecasts Initial forecasts

- $\Sigma_h^\dagger$ is generalized inverse of Σ_h.
- $\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'$
- **Problem:** Σ_h hard to estimate.

Optimal combination forecasts

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

Revised forecasts | Initial forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- $\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'$
- **Problem:** Σ_h hard to estimate.

Optimal combination forecasts

$$\tilde{\mathbf{Y}}_n(h) = \mathbf{S}\hat{\beta}_n(h) = \mathbf{S}(\mathbf{S}'\Sigma_h^{\dagger}\mathbf{S})^{-1}\mathbf{S}'\Sigma_h^{\dagger}\hat{\mathbf{Y}}_n(h)$$

Revised forecasts

Initial forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- $\text{Var}[\tilde{\mathbf{Y}}_n(h)|\mathbf{Y}_1,\dots,\mathbf{Y}_n] = \mathbf{S}(\mathbf{S}'\Sigma_h^{\dagger}\mathbf{S})^{-1}\mathbf{S}'$
- **Problem:** Σ_h hard to estimate.

Optimal combination forecasts

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

Revised forecasts

Initial forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- $\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'$
- **Problem:** Σ_h hard to estimate.

Outline

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

- Approximate $\Sigma_1^{\dagger}$ by $c\mathbf{I}$.
- Or 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

- Approximate $\Sigma_1^\dagger$ by $c\mathbf{I}$.
- Or 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

- Approximate $\Sigma_1^{\dagger}$ by $c\mathbf{I}$.
- Or 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

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

Revised forecasts   Base forecasts

Solution 1: OLS

- Approximate $\Sigma_1^{\dagger}$ by $c\mathbf{I}$.
- Or assume $\varepsilon_h \approx \mathbf{S}\varepsilon_{B,h}$ where $\varepsilon_{B,h}$ is the forecast error at bottom level.
- Then $\Sigma_h \approx \mathbf{S}\Omega_h\mathbf{S}'$ where $\Omega_h = \text{Var}(\varepsilon_{B,h})$.
- If Moore-Penrose generalized inverse used, then $(\mathbf{S}'\Sigma_h^{\dagger}\mathbf{S})^{-1}\mathbf{S}'\Sigma_h^{\dagger} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$.

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

Optimal combination forecasts

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

Revised forecasts

Base forecasts

Solution 1: OLS

- Approximate $\Sigma_1^{\dagger}$ by $c\mathbf{I}$.
- Or 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 $(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.

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

Optimal combination forecasts

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

Revised forecasts

Base forecasts

Solution 1: OLS

- Approximate $\Sigma_1^{\dagger}$ by $c\mathbf{I}$.
- Or 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 $(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.

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

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

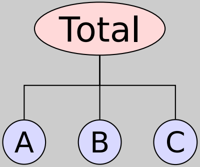
Base forecasts

Solution 1: OLS

- Approximate $\Sigma_1^{\dagger}$ by $c\mathbf{I}$.
- Or 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 $(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.

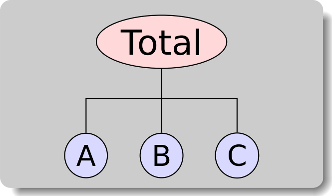
$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{Y}}_n(h)$$

Optimal combination forecasts



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

Optimal combination forecasts

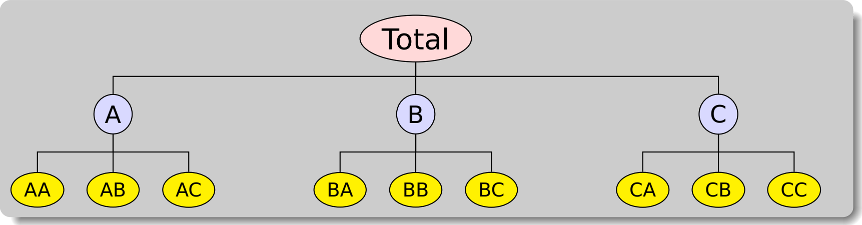


$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{Y}}_n(h)$$

Weights:

$$\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}' = \begin{bmatrix} 0.75 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.75 & -0.25 & -0.25 \\ 0.25 & -0.25 & 0.75 & -0.25 \\ 0.25 & -0.25 & -0.25 & 0.75 \end{bmatrix}$$

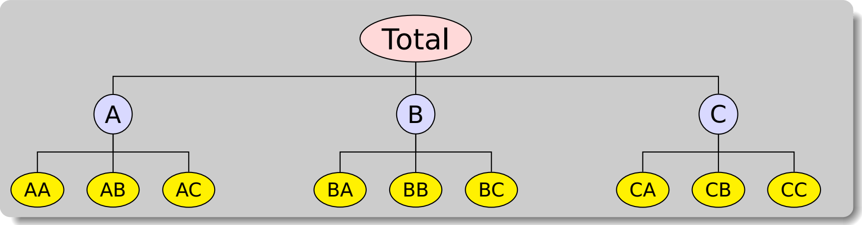
Optimal combination forecasts



Weights: $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}' =$

$$\begin{bmatrix}
0.69 & 0.23 & 0.23 & 0.23 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 \\
0.23 & 0.58 & -0.17 & -0.17 & 0.19 & 0.19 & 0.19 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 \\
0.23 & -0.17 & 0.58 & -0.17 & -0.06 & -0.06 & -0.06 & 0.19 & 0.19 & 0.19 & -0.06 & -0.06 & -0.06 \\
0.23 & -0.17 & -0.17 & 0.58 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 & 0.19 & 0.19 & 0.19 \\
0.08 & 0.19 & -0.06 & -0.06 & 0.73 & -0.27 & -0.27 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 \\
0.08 & 0.19 & -0.06 & -0.06 & -0.27 & 0.73 & -0.27 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 \\
0.08 & 0.19 & -0.06 & -0.06 & -0.27 & -0.27 & 0.73 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & 0.19 & -0.06 & -0.02 & -0.02 & -0.02 & 0.73 & -0.27 & -0.27 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & 0.19 & -0.06 & -0.02 & -0.02 & -0.02 & -0.27 & 0.73 & -0.27 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & 0.19 & -0.06 & -0.02 & -0.02 & -0.02 & -0.27 & -0.27 & 0.73 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & -0.06 & 0.19 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & 0.73 & -0.27 & -0.27 \\
0.08 & -0.06 & -0.06 & 0.19 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.27 & 0.73 & -0.27 \\
0.08 & -0.06 & -0.06 & 0.19 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.27 & -0.27 & 0.73
\end{bmatrix}$$

Optimal combination forecasts



Weights: $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}' =$

$$
\begin{bmatrix}
0.69 & 0.23 & 0.23 & 0.23 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 & 0.08 \\
0.23 & 0.58 & -0.17 & -0.17 & 0.19 & 0.19 & 0.19 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 \\
0.23 & -0.17 & 0.58 & -0.17 & -0.06 & -0.06 & -0.06 & 0.19 & 0.19 & 0.19 & -0.06 & -0.06 & -0.06 \\
0.23 & -0.17 & -0.17 & 0.58 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 & -0.06 & 0.19 & 0.19 & 0.19 \\
0.08 & 0.19 & -0.06 & -0.06 & 0.73 & -0.27 & -0.27 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 \\
0.08 & 0.19 & -0.06 & -0.06 & -0.27 & 0.73 & -0.27 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 \\
0.08 & 0.19 & -0.06 & -0.06 & -0.27 & -0.27 & 0.73 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & 0.19 & -0.06 & -0.02 & -0.02 & -0.02 & 0.73 & -0.27 & -0.27 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & 0.19 & -0.06 & -0.02 & -0.02 & -0.02 & -0.27 & 0.73 & -0.27 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & 0.19 & -0.06 & -0.02 & -0.02 & -0.02 & -0.27 & -0.27 & 0.73 & -0.02 & -0.02 & -0.02 \\
0.08 & -0.06 & -0.06 & 0.19 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & 0.73 & -0.27 & -0.27 \\
0.08 & -0.06 & -0.06 & 0.19 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.27 & 0.73 & -0.27 \\
0.08 & -0.06 & -0.06 & 0.19 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.02 & -0.27 & -0.27 & 0.73
\end{bmatrix}
$$

Features

- Covariates can be included in initial forecasts.
- Adjustments can be made to initial forecasts at any level.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.
- $\boldsymbol{SPS} = \boldsymbol{S}$ so reconciled forcasts are unbiased.
- Conceptually easy to implement: OLS on base forecasts.
- Weights are independent of the data and of the covariance structure of the hierarchy.

Features

- Covariates can be included in initial forecasts.
- Adjustments can be made to initial forecasts at any level.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.
- $\boldsymbol{SPS} = \boldsymbol{S}$ so reconciled forcasts are unbiased.
- Conceptually easy to implement: OLS on base forecasts.
- Weights are independent of the data and of the covariance structure of the hierarchy.

Features

- Covariates can be included in initial forecasts.
- Adjustments can be made to initial forecasts at any level.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.
- $\boldsymbol{SPS} = \boldsymbol{S}$ so reconciled forcasts are unbiased.
- Conceptually easy to implement: OLS on base forecasts.
- Weights are independent of the data and of the covariance structure of the hierarchy.

Features

- Covariates can be included in initial forecasts.
- Adjustments can be made to initial forecasts at any level.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.
- $\boldsymbol{SPS} = \boldsymbol{S}$ so reconciled forcasts are unbiased.
- Conceptually easy to implement: OLS on base forecasts.
- Weights are independent of the data and of the covariance structure of the hierarchy.

Features

- Covariates can be included in initial forecasts.
- Adjustments can be made to initial forecasts at any level.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.
- $\boldsymbol{SPS} = \boldsymbol{S}$ so reconciled forcasts are unbiased.
- Conceptually easy to implement: OLS on base forecasts.
- Weights are independent of the data and of the covariance structure of the hierarchy.

Features

- Covariates can be included in initial forecasts.
- Adjustments can be made to initial forecasts at any level.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.
- $\boldsymbol{SPS} = \boldsymbol{S}$ so reconciled forcasts are unbiased.
- Conceptually easy to implement: OLS on base forecasts.
- Weights are independent of the data and of the covariance structure of the hierarchy.

Challenges

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

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

Challenges

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

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

Challenges

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

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

Optimal combination forecasts

$$\tilde{\mathbf{Y}}_n(h) = \mathbf{S}(\mathbf{S}'\Sigma_1^{\dagger}\mathbf{S})^{-1}\mathbf{S}'\Sigma_1^{\dagger}\hat{\mathbf{Y}}_n(h)$$

Solution 1: OLS

- Approximate $\Sigma_1^{\dagger}$ by $c\mathbf{I}$.

Solution 2: Rescaling

- Suppose we approximate Σ_1 by its diagonal.
- Let $\Lambda = \left[\text{diagonal}\left(\Sigma_1\right)\right]^{-1}$ contain inverse one-step forecast variances.

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

Optimal combination forecasts

$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_1^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_1^{\dagger}\hat{\boldsymbol{Y}}_n(h)$$

Solution 1: OLS

- Approximate $\Sigma_1^{\dagger}$ by $c\mathbf{I}$.

Solution 2: Rescaling

- Suppose we approximate Σ_1 by its diagonal.
- Let $\Lambda = \left[\text{diagonal}\left(\Sigma_1\right)\right]^{-1}$ contain inverse one-step forecast variances.

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

Optimal combination forecasts

$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_1^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_1^{\dagger}\hat{\boldsymbol{Y}}_n(h)$$

Solution 1: OLS

- Approximate $\Sigma_1^{\dagger}$ by $c\mathbf{I}$.

Solution 2: Rescaling

- Suppose we approximate Σ_1 by its diagonal.
- Let $\Lambda = \left[\text{diagonal}\left(\Sigma_1\right)\right]^{-1}$ contain inverse one-step forecast variances.

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

Optimal combination forecasts

$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_1^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_1^{\dagger}\hat{\boldsymbol{Y}}_n(h)$$

Solution 1: OLS

- Approximate $\Sigma_1^{\dagger}$ by $c\mathbf{I}$.

Solution 2: Rescaling

- Suppose we approximate Σ_1 by its diagonal.
- Let $\Lambda = \left[\text{diagonal}\left(\Sigma_1\right)\right]^{-1}$ contain inverse one-step forecast variances.

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

Optimal combination forecasts

$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Sigma_1^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_1^\dagger\hat{\boldsymbol{Y}}_n(h)$$

Solution 1: OLS

- Approximate $\Sigma_1^\dagger$ by $c\mathbf{I}$.

Solution 2: Rescaling

- Suppose we approximate Σ_1 by its diagonal.
- Let $\Lambda = \left[\text{diagonal}\left(\Sigma_1\right)\right]^{-1}$ contain inverse one-step forecast variances.

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

Optimal reconciled forecasts

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

- Easy to estimate, and places weight where we have best forecasts.
- Ignores covariances.
- For large numbers of time series, we need to do calculation without explicitly forming S or $(S'\Lambda S)^{-1}$ or $S'\Lambda$.

Optimal reconciled forecasts

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

Initial forecasts

- Easy to estimate, and places weight where we have best forecasts.
- Ignores covariances.
- For large numbers of time series, we need to do calculation without explicitly forming $\boldsymbol{S}$ or $(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}$ or $\boldsymbol{S}'\Lambda$.

Optimal reconciled forecasts

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

Revised forecasts — Initial forecasts

- Easy to estimate, and places weight where we have best forecasts.
- Ignores covariances.
- For large numbers of time series, we need to do calculation without explicitly forming $\boldsymbol{S}$ or $(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}$ or $\boldsymbol{S}'\Lambda$.

Optimal reconciled forecasts

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

Revised forecasts

Initial forecasts

- Easy to estimate, and places weight where we have best forecasts.
- Ignores covariances.
- For large numbers of time series, we need to do calculation without explicitly forming $\boldsymbol{S}$ or $(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}$ or $\boldsymbol{S}'\Lambda$.

Optimal reconciled forecasts

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

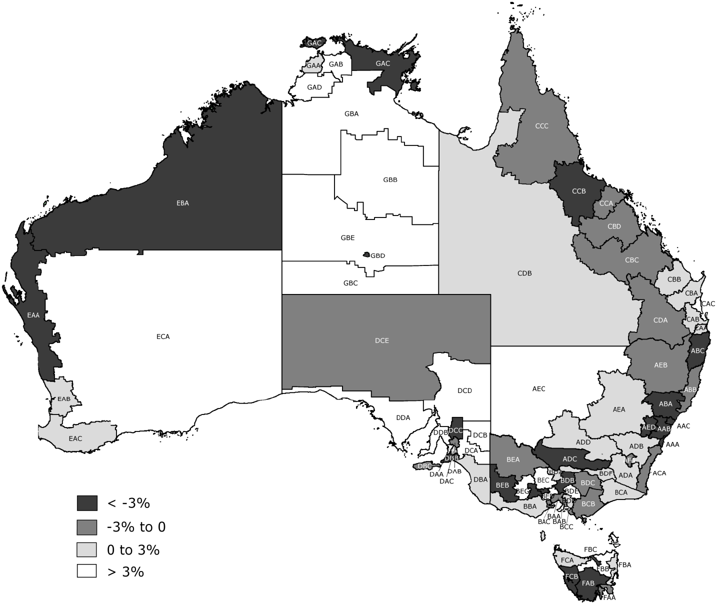
Revised forecasts

Initial forecasts

- Easy to estimate, and places weight where we have best forecasts.
- Ignores covariances.
- For large numbers of time series, we need to do calculation without explicitly forming $\boldsymbol{S}$ or $(\boldsymbol{S}'\Lambda\boldsymbol{S})^{-1}$ or $\boldsymbol{S}'\Lambda$.

Outline

Australian tourism



Australian tourism

Domestic visitor nights

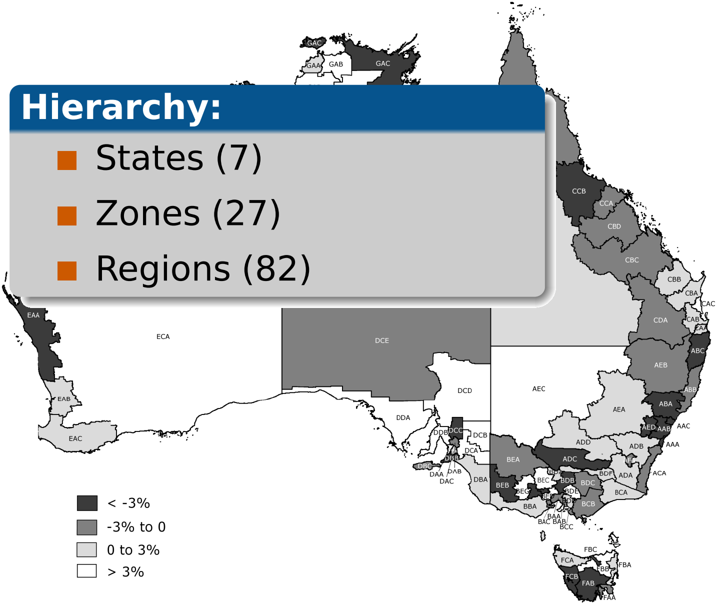
Quarterly data: 1998 – 2006.

From: *National Visitor Survey*, based on annual interviews of 120,000 Australians aged 15+, collected by Tourism Research Australia.

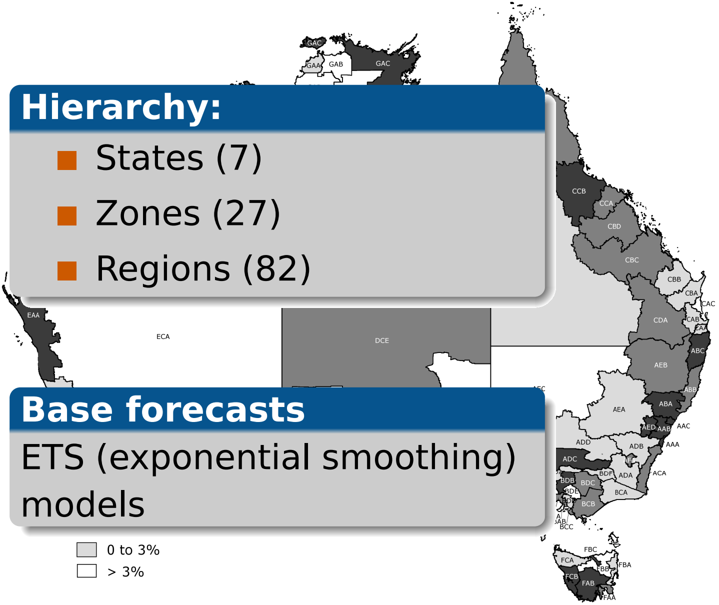
Australian tourism

Hierarchy:

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



Australian tourism



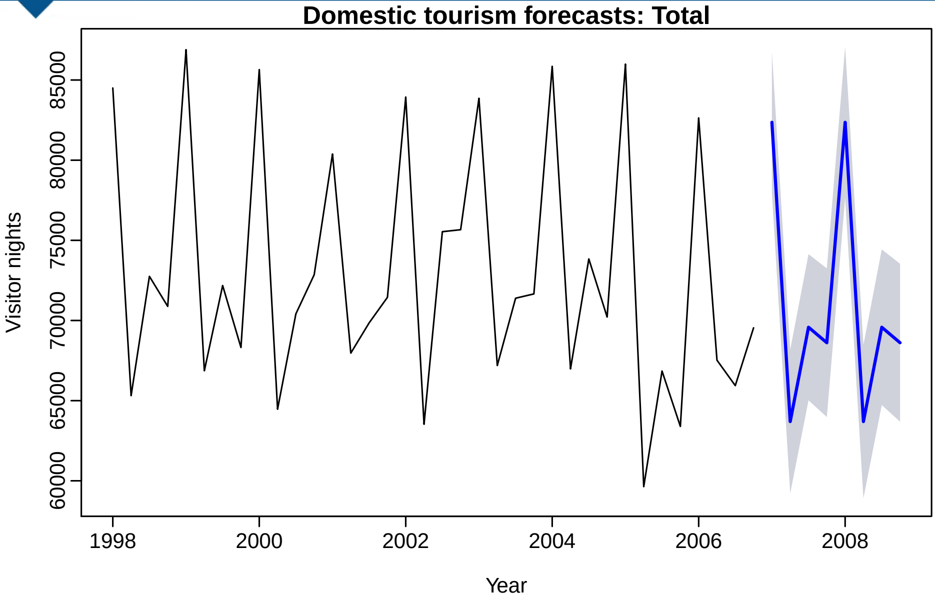
Hierarchy:

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

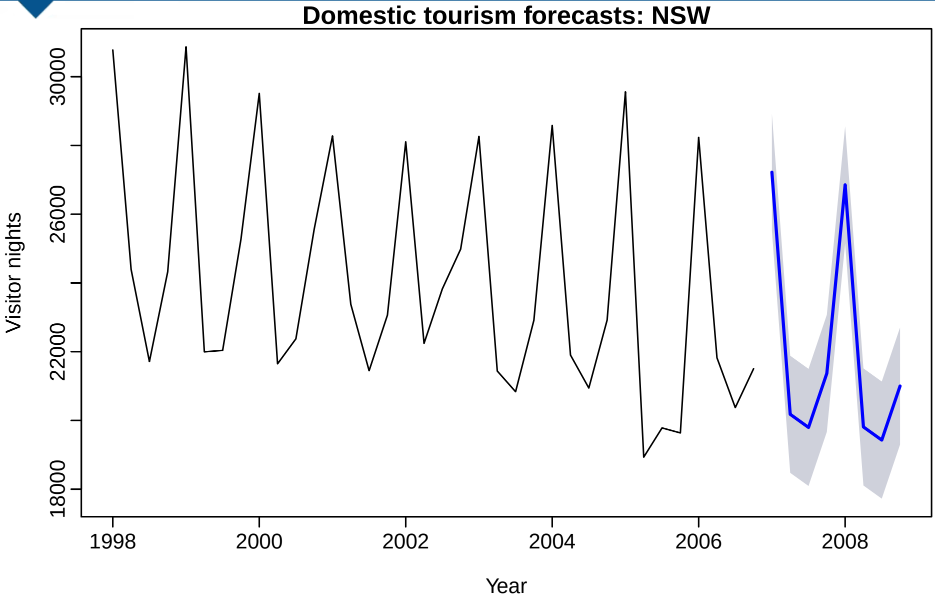
Base forecasts

ETS (exponential smoothing) models

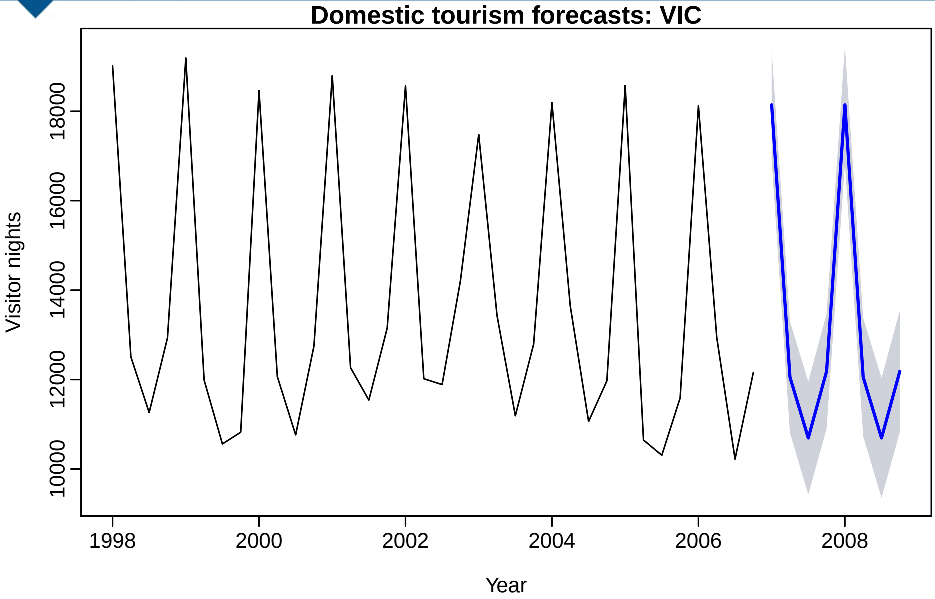
Base forecasts



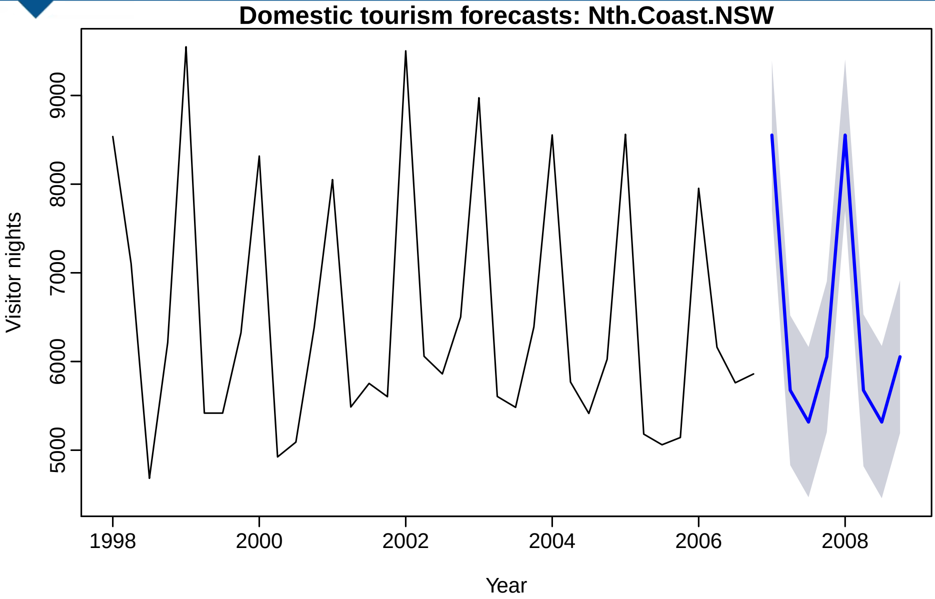
Base forecasts



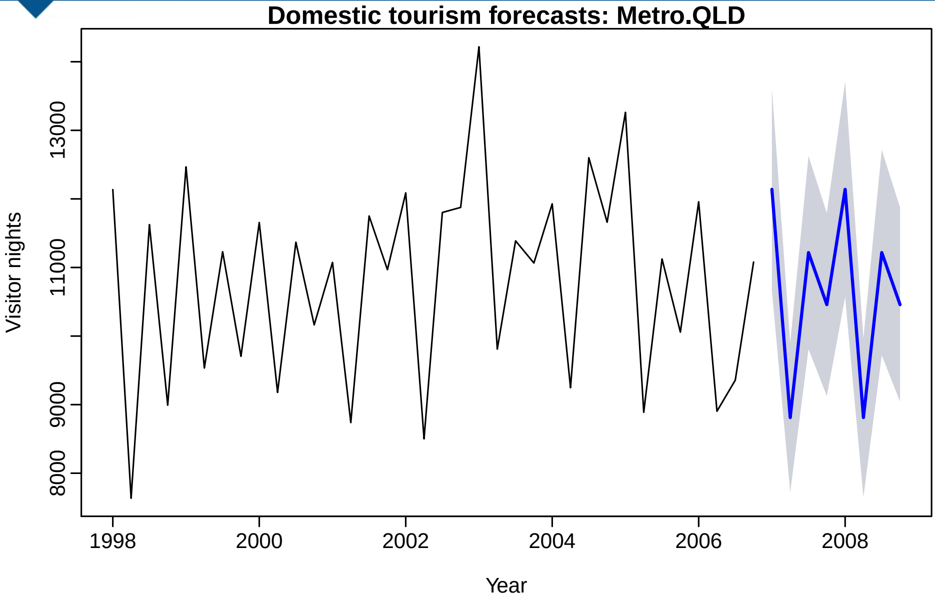
Base forecasts



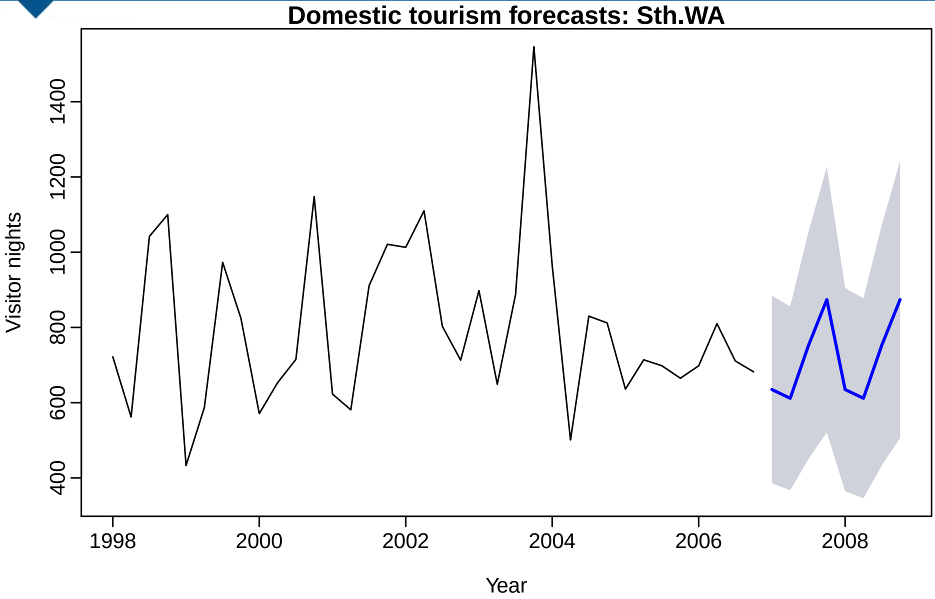
Base forecasts



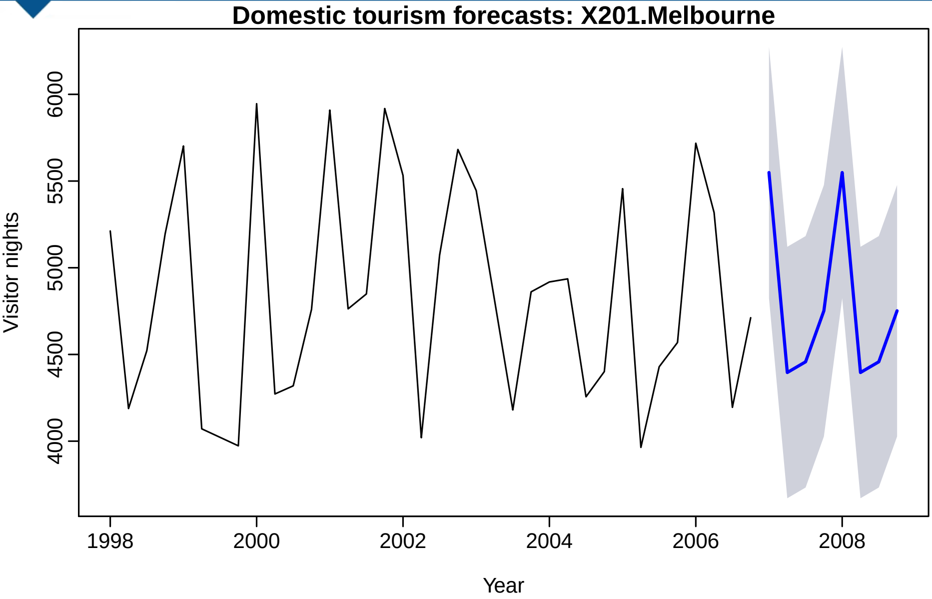
Base forecasts



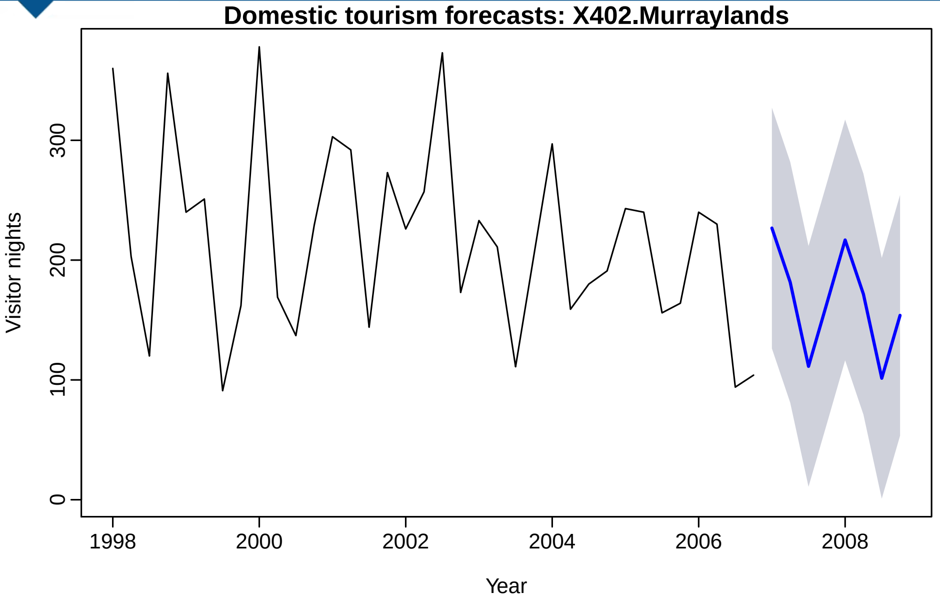
Base forecasts



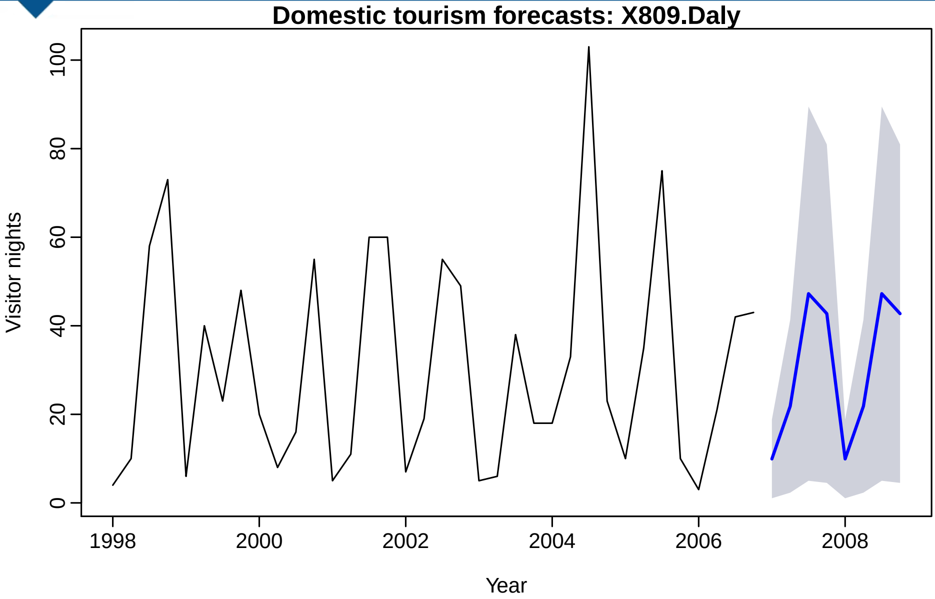
Base forecasts



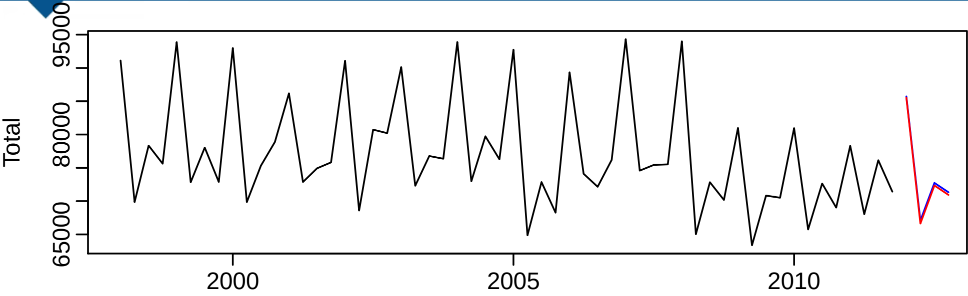
Base forecasts



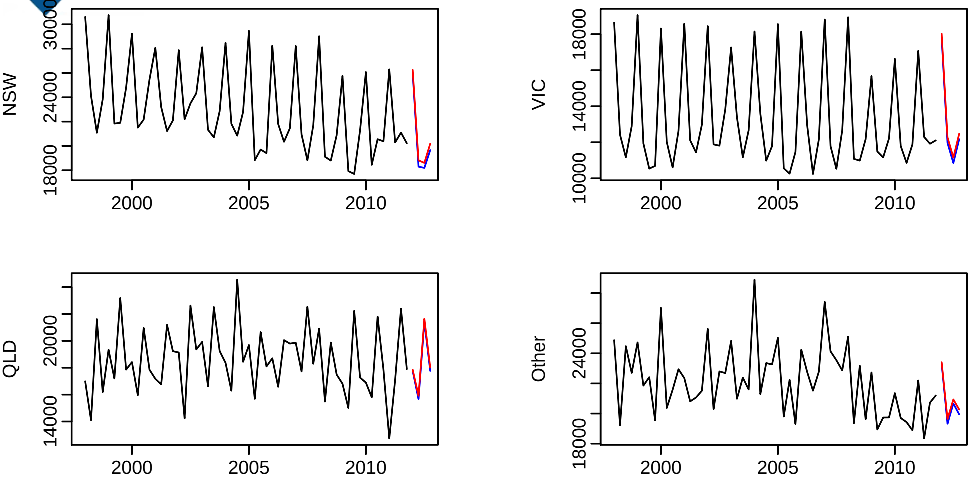
Base forecasts



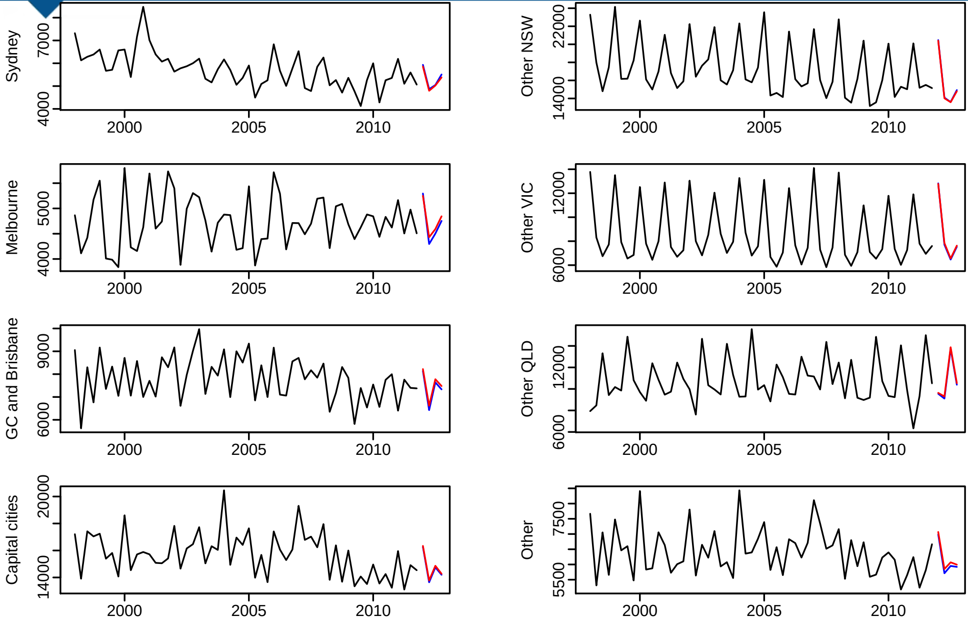
Reconciled forecasts



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**
Scaling (st. dev.)	**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
Scaling (st. dev.)	**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
Scaling (st. dev.)	**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
Scaling (st. dev.)	**31.68**	**31.22**	**31.08**	**32.41**	**32.77**	**31.89**

Outline

ANZSCO

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

ANZSCO

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

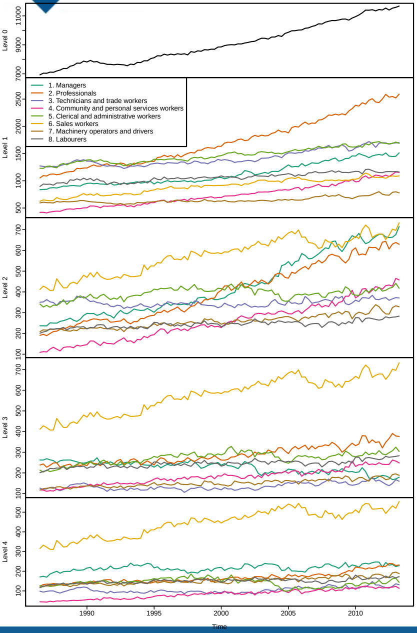
22 Business, Human Resource and Marketing Professionals

224 Information and Organisation Professionals

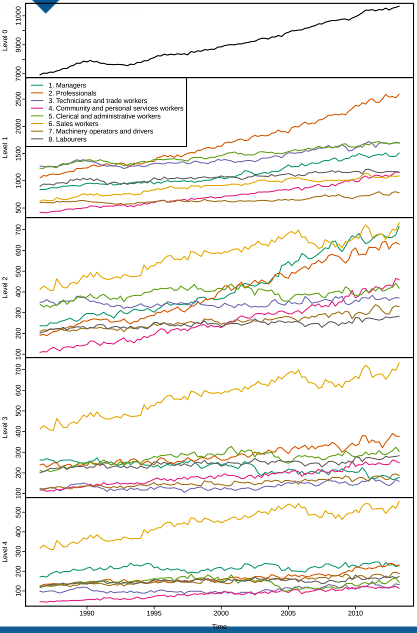
2241 Actuaries, Mathematicians and Statisticians

224113 Statistician

Australian Labour Market data

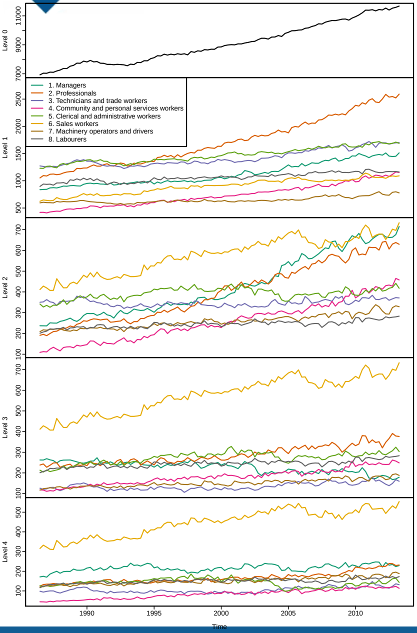


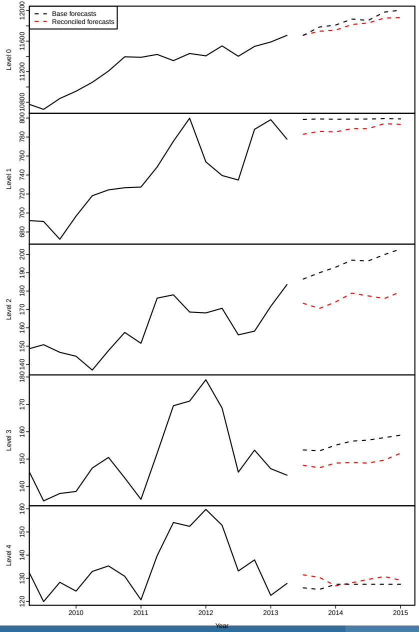
Australian Labour Market data



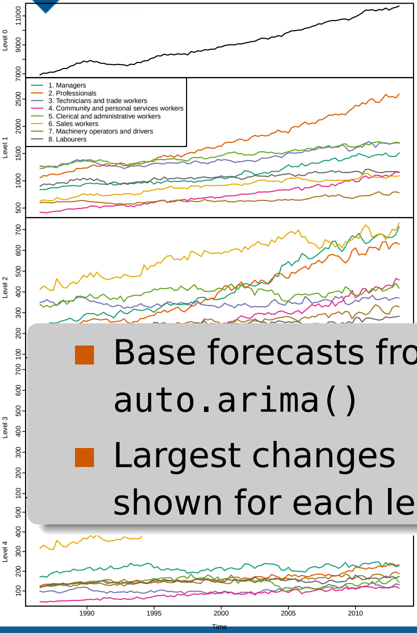
- Lower three panels show largest sub-groups at each level.

Australian Labour Market data

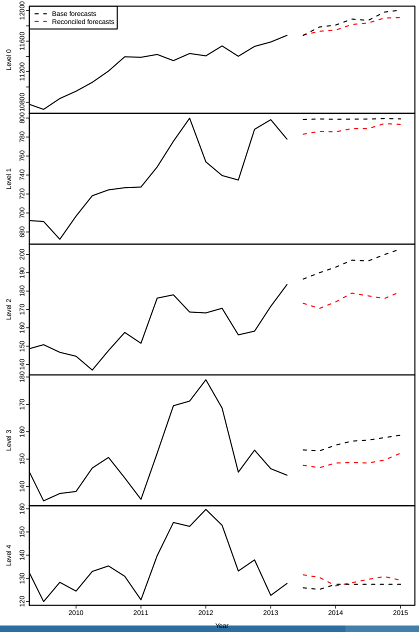




Australian Labour Market data



- Base forecasts from `auto.arima()`
- Largest changes shown for each level



Forecast evaluation (rolling origin)

RMSE	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	Average
Top level									
Bottom-up	74.71	102.02	121.70	131.17	147.08	157.12	169.60	178.93	135.29
OLS	**52.20**	**77.77**	**101.50**	**119.03**	138.27	150.75	160.04	166.38	**120.74**
WLS	61.77	86.32	107.26	119.33	**137.01**	**146.88**	**156.71**	**162.38**	122.21
Level 1									
Bottom-up	21.59	27.33	30.81	32.94	35.45	37.10	39.00	40.51	33.09
OLS	21.89	28.55	32.74	35.58	38.82	41.24	43.34	45.49	35.96
WLS	**20.58**	**26.19**	**29.71**	**31.84**	**34.36**	**35.89**	**37.53**	**38.86**	**31.87**
Level 2									
Bottom-up	8.78	10.72	11.79	12.42	13.13	13.61	14.14	14.65	12.40
OLS	9.02	11.19	12.34	13.04	13.92	14.56	15.17	15.77	13.13
WLS	**8.58**	**10.48**	**11.54**	**12.15**	**12.88**	**13.36**	**13.87**	**14.36**	**12.15**
Level 3									
Bottom-up	5.44	6.57	7.17	7.53	7.94	8.27	8.60	8.89	7.55
OLS	5.55	6.78	7.42	7.81	8.29	8.68	9.04	9.37	7.87
WLS	**5.35**	**6.46**	**7.06**	**7.42**	**7.84**	**8.17**	**8.48**	**8.76**	**7.44**
Bottom Level									
Bottom-up	2.35	2.79	3.02	3.15	3.29	3.42	3.54	3.65	3.15
OLS	2.40	2.86	3.10	3.24	3.41	3.55	3.68	3.80	3.25
WLS	**2.34**	**2.77**	**2.99**	**3.12**	**3.27**	**3.40**	**3.52**	**3.63**	**3.13**

Outline

hts package for R

hts: Hierarchical and grouped time series

Methods for analysing and forecasting hierarchical and grouped time series

Version:	4.3
Depends:	forecast ($\geq$ 5.0)
Imports:	SparseM, parallel, utils
Published:	2014-06-10
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

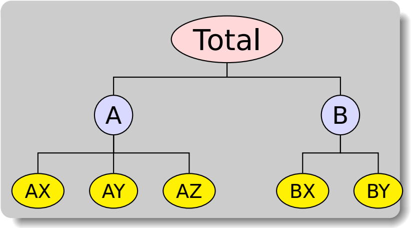
```
library(hts)

# 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

```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