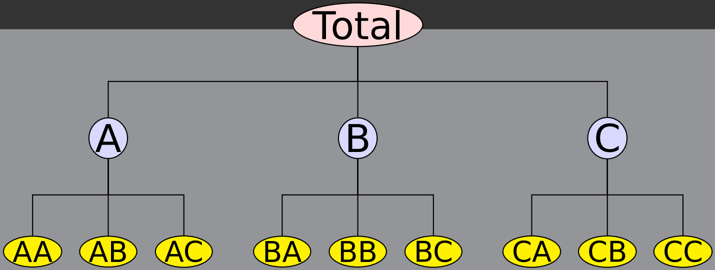
MONASH University

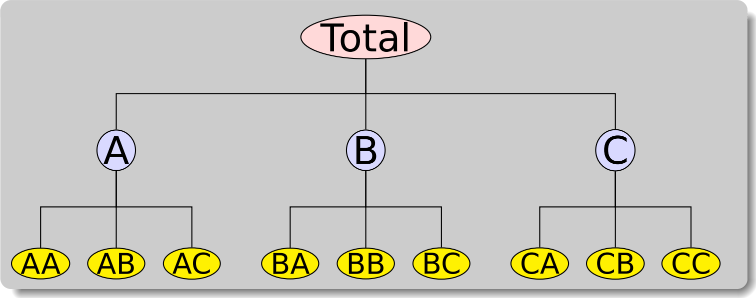
Rob J Hyndman

Forecasting hierarchical time series



Outline

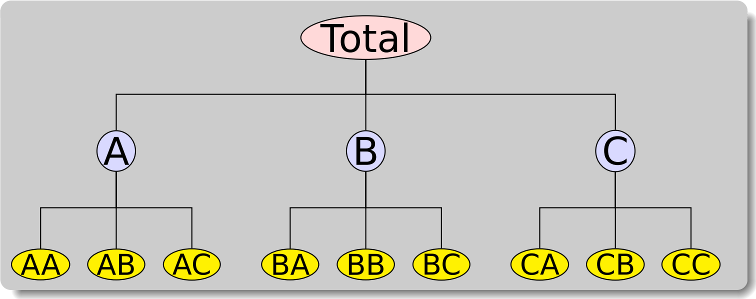
Introduction



Examples

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

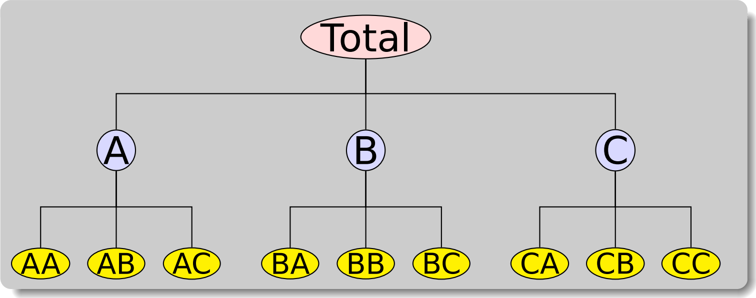
Introduction



Examples

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

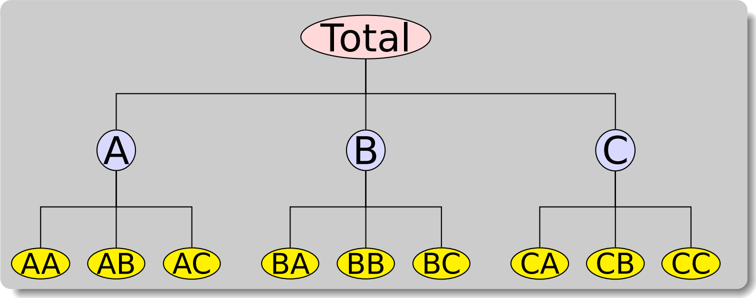
Introduction



Examples

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

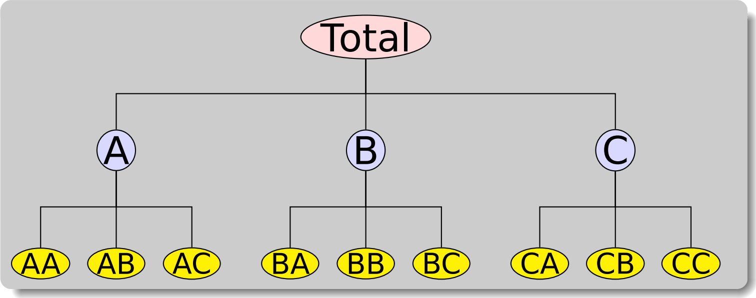
Introduction



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

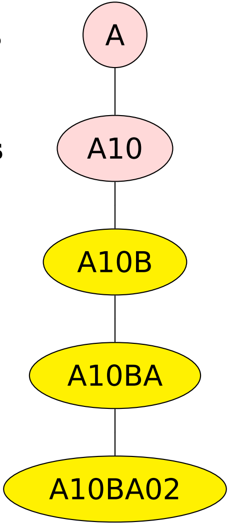
ATC drug classification

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

84 classes



Alimentary tract and metabolism

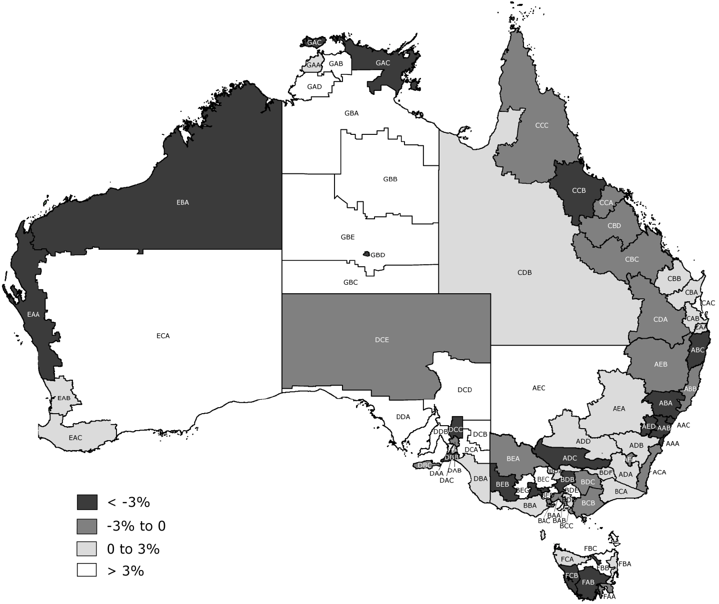
Drugs used in diabetes

Blood glucose lowering drugs

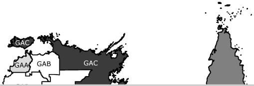
Biguanides

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/grouped time series

- Forecasts should be "aggregate consistent", unbiased, minimum variance.
- Existing methods:
  - Bottom-up
  - Top-down
  - Middle-out
- How to compute forecast intervals?
- Most research is concerned about relative performance of existing methods.
- There is **no** research on how to deal with forecasting grouped time series.

Hierarchical/grouped time series

- Forecasts should be "aggregate consistent", unbiased, minimum variance.
- Existing methods:
  - Bottom-up
  - Top-down
  - Middle-out
- How to compute forecast intervals?
- Most research is concerned about relative performance of existing methods.
- There is **no** research on how to deal with forecasting grouped time series.

Hierarchical/grouped time series

- Forecasts should be “aggregate consistent”, unbiased, minimum variance.
- Existing methods:
  - Bottom-up
  - Top-down
  - Middle-out
- How to compute forecast intervals?
- Most research is concerned about relative performance of existing methods.
- There is **no** research on how to deal with forecasting grouped time series.

Hierarchical/grouped time series

- Forecasts should be "aggregate consistent", unbiased, minimum variance.
- Existing methods:
  - Bottom-up
  - Top-down
  - Middle-out
- How to compute forecast intervals?
- Most research is concerned about relative performance of existing methods.
- There is **no** research on how to deal with forecasting grouped time series.

Hierarchical/grouped time series

- Forecasts should be "aggregate consistent", unbiased, minimum variance.
- Existing methods:
  - Bottom-up
  - Top-down
  - Middle-out
- How to compute forecast intervals?
- Most research is concerned about relative performance of existing methods.
- There is **no** research on how to deal with forecasting grouped time series.

Hierarchical/grouped time series

- Forecasts should be “aggregate consistent”, unbiased, minimum variance.
- Existing methods:
  - Bottom-up
  - Top-down
  - Middle-out
- How to compute forecast intervals?
- Most research is concerned about relative performance of existing methods.
- There is **no** research on how to deal with forecasting grouped time series.

Hierarchical/grouped time series

- Forecasts should be “aggregate consistent”, unbiased, minimum variance.
- Existing methods:
  - Bottom-up
  - Top-down
  - Middle-out
- How to compute forecast intervals?
- Most research is concerned about relative performance of existing methods.
- There is **no** research on how to deal with forecasting grouped time series.

Hierarchical/grouped time series

- Forecasts should be “aggregate consistent”, unbiased, minimum variance.
- Existing methods:
  - Bottom-up
  - Top-down
  - Middle-out
- How to compute forecast intervals?
- Most research is concerned about relative performance of existing methods.
- There is **no** research on how to deal with forecasting grouped time series.

Top-down method

Advantages

- Works well in presence of low counts.
- Single forecasting model easy to build
- Provides reliable forecasts for aggregate levels.

Disadvantages

- Loss of information, especially individual series dynamics.
- Distribution of forecasts to lower levels can be difficult
- No prediction intervals

Top-down method

Advantages

- Works well in presence of low counts.
- Single forecasting model easy to build
- Provides reliable forecasts for aggregate levels.

Disadvantages

- Loss of information, especially individual series dynamics.
- Distribution of forecasts to lower levels can be difficult
- No prediction intervals

Top-down method

Advantages

- Works well in presence of low counts.
- Single forecasting model easy to build
- Provides reliable forecasts for aggregate levels.

Disadvantages

- Loss of information, especially individual series dynamics.
- Distribution of forecasts to lower levels can be difficult
- No prediction intervals

Top-down method

Advantages

- Works well in presence of low counts.
- Single forecasting model easy to build
- Provides reliable forecasts for aggregate levels.

Disadvantages

- Loss of information, especially individual series dynamics.
- Distribution of forecasts to lower levels can be difficult
- No prediction intervals

Top-down method

Advantages

- Works well in presence of low counts.
- Single forecasting model easy to build
- Provides reliable forecasts for aggregate levels.

Disadvantages

- Loss of information, especially individual series dynamics.
- Distribution of forecasts to lower levels can be difficult
- No prediction intervals

Top-down method

Advantages

- Works well in presence of low counts.
- Single forecasting model easy to build
- Provides reliable forecasts for aggregate levels.

Disadvantages

- Loss of information, especially individual series dynamics.
- Distribution of forecasts to lower levels can be difficult
- No prediction intervals

Bottom-up method

Advantages

- No loss of information.
- Better captures dynamics of individual series.

Disadvantages

- Large number of series to be forecast.
- Constructing forecasting models is harder because of noisy data at bottom level.
- No prediction intervals

Bottom-up method

Advantages

- No loss of information.
- Better captures dynamics of individual series.

Disadvantages

- Large number of series to be forecast.
- Constructing forecasting models is harder because of noisy data at bottom level.
- No prediction intervals

Bottom-up method

Advantages

- No loss of information.
- Better captures dynamics of individual series.

Disadvantages

- Large number of series to be forecast.
- Constructing forecasting models is harder because of noisy data at bottom level.
- No prediction intervals

Bottom-up method

Advantages

- No loss of information.
- Better captures dynamics of individual series.

Disadvantages

- Large number of series to be forecast.
- Constructing forecasting models is harder because of noisy data at bottom level.
- No prediction intervals

Bottom-up method

Advantages

- No loss of information.
- Better captures dynamics of individual series.

Disadvantages

- Large number of series to be forecast.
- Constructing forecasting models is harder because of noisy data at bottom level.
- No prediction intervals

A new approach

We propose a new statistical framework for forecasting hierarchical time series which:

1. provides point forecasts that are consistent across the hierarchy;
2. allows for correlations and interaction between series at each level;
3. provides estimates of forecast uncertainty which are consistent across the hierarchy;
4. allows for ad hoc adjustments and inclusion of covariates at any level.

A new approach

We propose a new statistical framework for forecasting hierarchical time series which:

1. provides point forecasts that are consistent across the hierarchy;
2. allows for correlations and interaction between series at each level;
3. provides estimates of forecast uncertainty which are consistent across the hierarchy;
4. allows for ad hoc adjustments and inclusion of covariates at any level.

A new approach

We propose a new statistical framework for forecasting hierarchical time series which:

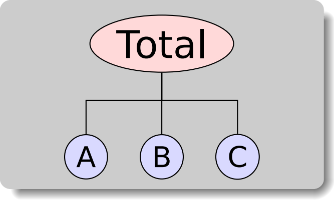
1. provides point forecasts that are consistent across the hierarchy;
2. allows for correlations and interaction between series at each level;
3. provides estimates of forecast uncertainty which are consistent across the hierarchy;
4. allows for ad hoc adjustments and inclusion of covariates at any level.

A new approach

We propose a new statistical framework for forecasting hierarchical time series which:

1. provides point forecasts that are consistent across the hierarchy;
2. allows for correlations and interaction between series at each level;
3. provides estimates of forecast uncertainty which are consistent across the hierarchy;
4. allows for ad hoc adjustments and inclusion of covariates at any level.

Hierarchical data

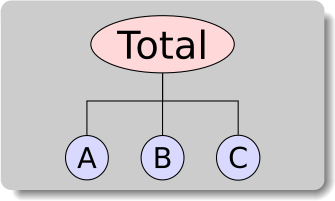


Y_t : observed aggregate of all series at time t.

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

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

Hierarchical data

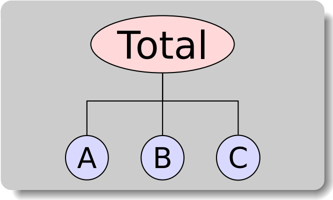


Y_t : observed aggregate of all series at time t.

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

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

Hierarchical data



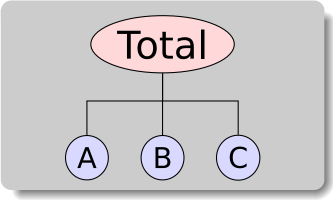
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: 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



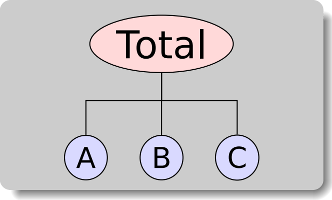
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: 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



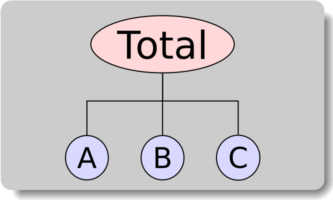
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: 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.

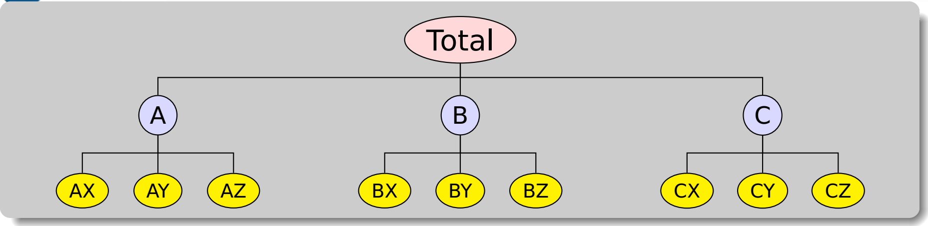
$Y_{X,t}$: observation on series X at time t.

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

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

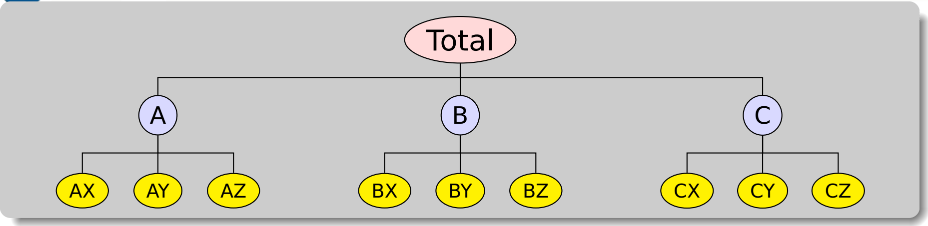
$$\boxed{\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t}$$

Hierarchical data



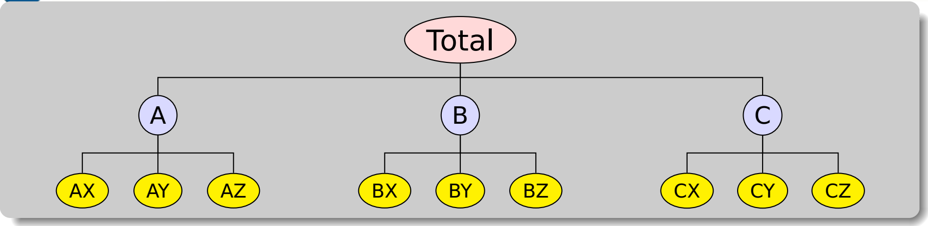
$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0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

Hierarchical data



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

Hierarchical data



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

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

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:

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

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: $\tilde{\boldsymbol{Y}}_n(h)$.

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: $\tilde{\boldsymbol{Y}}_n(h)$.

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: $\tilde{\boldsymbol{Y}}_n(h)$.

Bottom-up forecasts

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

Bottom-up forecasts are obtained using

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

where $\boldsymbol{0}$ is null matrix and $\boldsymbol{I}$ 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

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

where $\boldsymbol{0}$ is null matrix and $\boldsymbol{I}$ 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

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

where $\boldsymbol{0}$ is null matrix and $\boldsymbol{I}$ 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 $\boldsymbol{p} = [p_1, p_2, \ldots, p_{m_K}]'$ 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

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

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 $E[\hat{\boldsymbol{Y}}_n(h)] = \boldsymbol{S}\beta_n(h)$.
- 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:

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

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

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

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: $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{S}\beta_n(h) = \boldsymbol{S}\beta_n(h)$.
- 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: 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

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

Then the variance of the revised forecasts is given by

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

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

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

Then the variance of the revised forecasts is given by

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

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

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

Then the variance of the revised forecasts is given by

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

This is a general result for all existing methods.

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 $\boldsymbol{Y}_t$.

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

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

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 $\boldsymbol{Y}_t$.

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

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

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 $\boldsymbol{Y}_t$.

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

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

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

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

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

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

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

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

Optimal combination forecasts

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

Optimal combination forecasts

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

Initial forecasts

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.
- Optimal $\mathbf{P} = (\mathbf{S}'\Sigma_h^\dagger\mathbf{S})^{-1}\mathbf{S}'\Sigma_h^\dagger$
- Revised forecasts *unbiased*: $\mathbf{SPS} = \mathbf{S}$.
- Revised forecasts *minimum variance*:

$$\begin{aligned} \mathrm{V}[\tilde{\mathbf{Y}}_n(h)|\mathbf{Y}_1,\dots,\mathbf{Y}_n] &= \mathbf{SP}\Sigma_h\mathbf{P}'\mathbf{S}' \\ &= \mathbf{S}(\mathbf{S}'\Sigma_h^\dagger\mathbf{S})^{-1}\mathbf{S}' \end{aligned}$$

- **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.
- Optimal $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$
- Revised forecasts *unbiased*: $\boldsymbol{SPS} = \boldsymbol{S}$.
- Revised forecasts *minimum variance*:

$$\begin{aligned} \mathrm{V}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] &= \boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}' \\ &= \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}' \end{aligned}$$

- **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.
- Optimal $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^\dagger$
- Revised forecasts *unbiased*: $\boldsymbol{SPS} = \boldsymbol{S}$.
- Revised forecasts *minimum variance*:

$$\begin{aligned}\mathrm{V}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] &= \boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}' \\ &= \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'\end{aligned}$$

- **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.
- Optimal $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$
- Revised forecasts *unbiased*: $\boldsymbol{SPS} = \boldsymbol{S}$.
- Revised forecasts *minimum variance*:

$$\begin{aligned} \mathrm{V}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] &= \boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}' \\ &= \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}' \end{aligned}$$

- **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.
- Optimal $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$
- Revised forecasts *unbiased*: $\boldsymbol{SPS} = \boldsymbol{S}$.
- Revised forecasts *minimum variance*:

$$\begin{aligned} \mathrm{V}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] &= \boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}' \\ &= \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}' \end{aligned}$$

- **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.
- Optimal $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^\dagger$
- Revised forecasts *unbiased*: $\boldsymbol{SPS} = \boldsymbol{S}$.
- Revised forecasts *minimum variance*:

$$\begin{aligned} \mathrm{V}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] &= \boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}' \\ &= \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}' \end{aligned}$$

- **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.
- Optimal $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^\dagger$
- Revised forecasts *unbiased*: $\boldsymbol{SPS} = \boldsymbol{S}$.
- Revised forecasts *minimum variance*:

$$\begin{aligned} \mathrm{V}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] &= \boldsymbol{SP}\Sigma_h\boldsymbol{P}'\boldsymbol{S}' \\ &= \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}' \end{aligned}$$

- **Problem:** Σ_h hard to estimate.

Outline

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

- 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 $\Omega_h = \mathrm{V}(\varepsilon_{B,h})$.
- 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

- 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 $\Omega_h = \mathrm{V}(\varepsilon_{B,h})$.
- 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

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

Revised forecasts

Base forecasts

Solution 1: OLS

- 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 $\Omega_h = \mathrm{V}(\varepsilon_{B,h})$.
- 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

- 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 $\Omega_h = \mathrm{V}(\varepsilon_{B,h})$.
- 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

- 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 $\Omega_h = \mathrm{V}(\varepsilon_{B,h})$.
- 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

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

Revised forecasts   Base forecasts

Solution 1: OLS

- 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 $\Omega_h = \mathrm{V}(\varepsilon_{B,h})$.
- 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

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

- GLS = OLS.
- Optimal weighted average of initial forecasts.
- Optimal reconciliation weights are $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.
- Weights are independent of the data and of the covariance structure of the hierarchy!

Optimal combination forecasts

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

- GLS = OLS.
- Optimal weighted average of initial forecasts.
- Optimal reconciliation weights are $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.
- Weights are independent of the data and of the covariance structure of the hierarchy!

Optimal combination forecasts

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

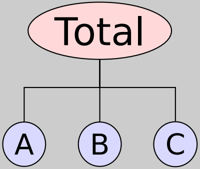
- GLS = OLS.
- Optimal weighted average of initial forecasts.
- Optimal reconciliation weights are $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.
- Weights are independent of the data and of the covariance structure of the hierarchy!

Optimal combination forecasts

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

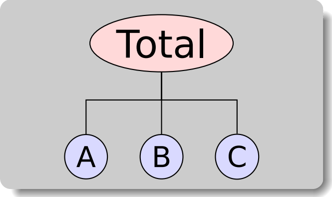
- GLS = OLS.
- Optimal weighted average of initial forecasts.
- Optimal reconciliation weights are $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.
- Weights are independent of the data and of the covariance structure of the hierarchy!

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

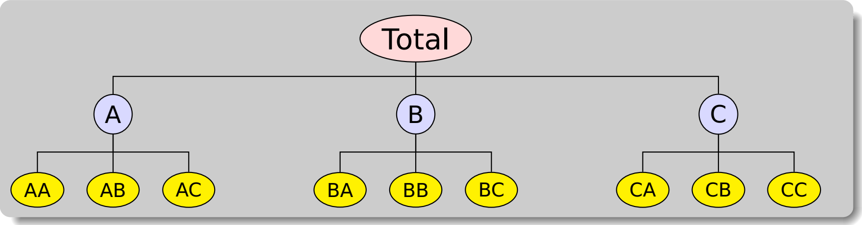


$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{Y}}_n(h)$$

Weights:

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

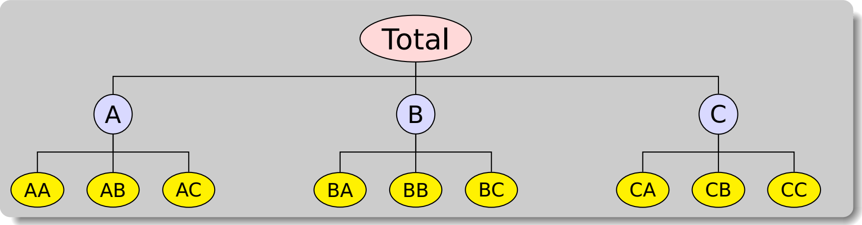
Optimal combination forecasts



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

- Forget “bottom up” or “top down”. This approach combines all forecasts optimally.
- Method outperforms bottom-up and top-down, especially for middle levels.
- 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.
- Conceptually easy to implement: OLS on base forecasts.

Features

- Forget "bottom up" or "top down". This approach combines all forecasts optimally.
- Method outperforms bottom-up and top-down, especially for middle levels.
- 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.
- Conceptually easy to implement: OLS on base forecasts.

Features

- Forget "bottom up" or "top down". This approach combines all forecasts optimally.
- Method outperforms bottom-up and top-down, especially for middle levels.
- 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.
- Conceptually easy to implement: OLS on base forecasts.

Features

- Forget "bottom up" or "top down". This approach combines all forecasts optimally.
- Method outperforms bottom-up and top-down, especially for middle levels.
- 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.
- Conceptually easy to implement: OLS on base forecasts.

Features

- Forget “bottom up” or “top down”. This approach combines all forecasts optimally.
- Method outperforms bottom-up and top-down, especially for middle levels.
- 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.
- Conceptually easy to implement: OLS on base forecasts.

Features

- Forget "bottom up" or "top down". This approach combines all forecasts optimally.
- Method outperforms bottom-up and top-down, especially for middle levels.
- 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.
- Conceptually easy to implement: OLS on base 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 non-singular behavior of $(\boldsymbol{S}'\boldsymbol{S})$.
- Need to estimate covariance matrix to produce prediction intervals.
- Assumption might be unrealistic.
- 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 non-singular behavior of $(\boldsymbol{S}'\boldsymbol{S})$.
- Need to estimate covariance matrix to produce prediction intervals.
- Assumption might be unrealistic.
- 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 non-singular behavior of $(\boldsymbol{S}'\boldsymbol{S})$.
- Need to estimate covariance matrix to produce prediction intervals.
- Assumption might be unrealistic.
- 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 non-singular behavior of $(\boldsymbol{S}'\boldsymbol{S})$.
- Need to estimate covariance matrix to produce prediction intervals.
- Assumption might be unrealistic.
- Ignores covariance matrix in computing point forecasts.

Optimal combination forecasts

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

Solution 2: Rescaling

- Suppose we rescale the original forecasts by Λ, reconcile using OLS, and backscale:

$$\tilde{\boldsymbol{Y}}^*_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda^2\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda^2\hat{\boldsymbol{Y}}_n(h).$$

- If $\Lambda = \left(\Sigma_h^\dagger\right)^{1/2}$, we get the GLS solution.
- Approximately optimal solution: $\Lambda = \text{diagonal}\left(\Sigma_1^\dagger\right)^{1/2}$
- That is, Λ contains inverse one-step forecast standard deviations.

Optimal combination forecasts

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

Solution 2: Rescaling

- Suppose we rescale the original forecasts by Λ, reconcile using OLS, and backscale:

$$\tilde{\boldsymbol{Y}}^*_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda^2\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda^2\hat{\boldsymbol{Y}}_n(h).$$

- If $\Lambda = \left(\Sigma_h^\dagger\right)^{1/2}$, we get the GLS solution.
- Approximately optimal solution: $\Lambda = \text{diagonal}\left(\Sigma_1^\dagger\right)^{1/2}$
- That is, Λ contains inverse one-step forecast standard deviations.

Optimal combination forecasts

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

Solution 2: Rescaling

- Suppose we rescale the original forecasts by Λ, reconcile using OLS, and backscale:

$$\tilde{\boldsymbol{Y}}_n^*(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda^2\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda^2\hat{\boldsymbol{Y}}_n(h).$$

- If $\Lambda = \left(\Sigma_h^\dagger\right)^{1/2}$, we get the GLS solution.
- Approximately optimal solution: $\Lambda = \text{diagonal}\left(\Sigma_1^\dagger\right)^{1/2}$
- That is, Λ contains inverse one-step forecast standard deviations.

Optimal combination forecasts

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

Solution 2: Rescaling

- Suppose we rescale the original forecasts by Λ, reconcile using OLS, and backscale:

$$\tilde{\boldsymbol{Y}}_n^*(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda^2\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda^2\hat{\boldsymbol{Y}}_n(h).$$

- If $\Lambda = \left(\Sigma_h^\dagger\right)^{1/2}$, we get the GLS solution.
- Approximately optimal solution:
  $\Lambda = \text{diagonal}\left(\Sigma_1^\dagger\right)^{1/2}$
- That is, Λ contains inverse one-step forecast standard deviations.

Optimal combination forecasts

$$\tilde{\boldsymbol{Y}}^*_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda^2\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda^2\hat{\boldsymbol{Y}}_n(h)$$

Solution 3: Averaging

- If the bottom level error series are approximately uncorrelated and have similar variances, then Λ is inversely proportional to the number of series making up each element of $\boldsymbol{Y}$.
- So set Λ to be the inverse row sums of $\boldsymbol{S}$.
- Then $\Lambda\hat{\boldsymbol{Y}}_n(h)$ is the *average* at each node rather than the *sum* at each node.

Optimal combination forecasts

$$\tilde{\boldsymbol{Y}}^*_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda^2\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda^2\hat{\boldsymbol{Y}}_n(h)$$

Solution 3: Averaging

- If the bottom level error series are approximately uncorrelated and have similar variances, then Λ is inversely proportional to the number of series making up each element of $\boldsymbol{Y}$.
- So set Λ to be the inverse row sums of $\boldsymbol{S}$.
- Then $\Lambda\hat{\boldsymbol{Y}}_n(h)$ is the *average* at each node rather than the *sum* at each node.

Optimal combination forecasts

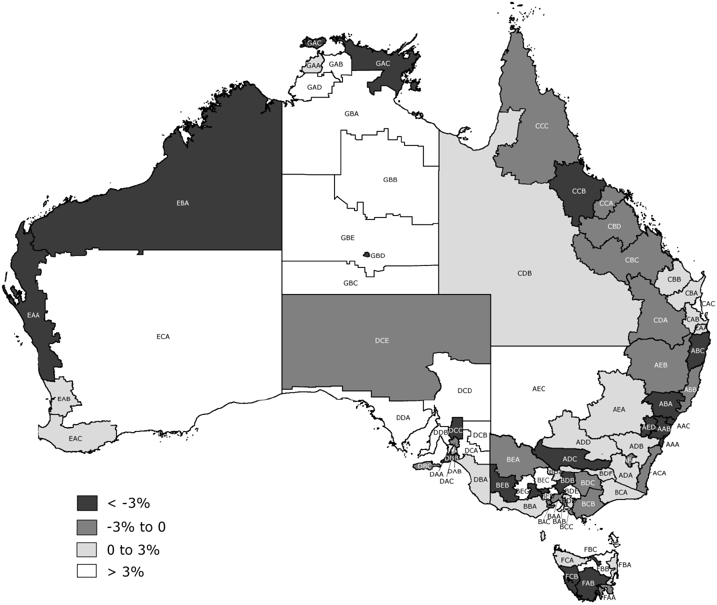
$$\tilde{\boldsymbol{Y}}^*_n(h) = \boldsymbol{S}(\boldsymbol{S}'\Lambda^2\boldsymbol{S})^{-1}\boldsymbol{S}'\Lambda^2\hat{\boldsymbol{Y}}_n(h)$$

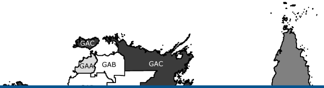
Solution 3: Averaging

- If the bottom level error series are approximately uncorrelated and have similar variances, then Λ is inversely proportional to the number of series making up each element of $\boldsymbol{Y}$.
- So set Λ to be the inverse row sums of $\boldsymbol{S}$.
- Then $\Lambda\hat{\boldsymbol{Y}}_n(h)$ is the *average* at each node rather than the *sum* at each node.

Outline

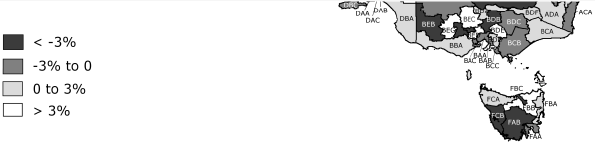
Application to Australian tourism





Quarterly data on visitor nights

Domestic visitor nights
from 1998 – 2006
Data from: *National Visitor Survey*, based on annual interviews of 120,000 Australians aged 15+, collected by Tourism Research Australia.



Application to Australian tourism

Also split by purpose of travel:

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

Exponential smoothing methods

		Seasonal Component		
Trend Component		N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	Seasonal Component A (Additive)	Seasonal Component M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

N,N: Simple exponential smoothing

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

N,N: Simple exponential smoothing
A,N: Holt's linear method

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

N,N: Simple exponential smoothing
A,N: Holt's linear method
A_d,N: Additive damped trend method

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

N,N: Simple exponential smoothing
A,N: Holt's linear method
A_d,N: Additive damped trend method
M,N: Exponential trend method

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

N,N: Simple exponential smoothing
A,N: Holt's linear method
A_d,N: Additive damped trend method
M,N: Exponential trend method
M_d,N: Multiplicative damped trend method

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

N,N: Simple exponential smoothing
A,N: Holt's linear method
A_d,N: Additive damped trend method
M,N: Exponential trend method
M_d,N: Multiplicative damped trend method
A,A: Additive Holt-Winters' method

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

N,N: Simple exponential smoothing
A,N: Holt's linear method
A_d,N: Additive damped trend method
M,N: Exponential trend method
M_d,N: Multiplicative damped trend method
A,A: Additive Holt-Winters' method
A,M: Multiplicative Holt-Winters' method

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

- There are 15 separate exponential smoothing methods.

Exponential smoothing methods

	Trend Component	Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

- There are 15 separate exponential smoothing methods.
- Each can have an additive or multiplicative error, giving 30 separate models.

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation E T S : **E**xponen**T**ial **S**mooth

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation E T S : **E**xponen**T**ial **S**mooth

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	Seasonal Component A (Additive)	Seasonal Component M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation E T S : **E**xponen**T**ial **S**mooth

Trend

Examples:

A,N,N: Simple exponential smoothing with additive errors

A,A,N: Holt's linear method with additive errors

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation E T S : **E**xponen**T**ial **S**mooth

↑ ↖

Trend **S**easonal

Examples:

A,N,N: Simple exponential smoothing with additive errors

A,A,N: Holt's linear method with additive errors

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation E T S : **E**xponen**T**ial **S**mooth

Error **T**rend **S**easonal

Examples:

A,N,N: Simple exponential smoothing with additive errors

A,A,N: Holt's linear method with additive errors

Exponential smoothing methods

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation E T S : **E**xponen**T**ial **S**mooth

Error **T**rend **S**easonal

Examples:

A,N,N: Simple exponential smoothing with additive errors

A,A,N: Holt's linear method with additive errors

Exponential smoothing methods

Innovations state space models

- All ETS models can be written in innovations state space form (IJF, 2002).
- Additive and multiplicative versions give the same point forecasts but different prediction intervals.

Error **T**rend **S**easonal

Examples:

A,N,N: Simple exponential smoothing with additive errors

A,A,N: Holt's linear method with additive errors

Automatic forecasting

From Hyndman et al. (IJF, 2002):

- Apply each of 30 models that are appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AIC:
$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$
where $p = \#$ parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Automatic forecasting

From Hyndman et al. (IJF, 2002):

- Apply each of 30 models that are appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AIC:
  $$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$
  where $p = \#$ parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Automatic forecasting

From Hyndman et al. (IJF, 2002):

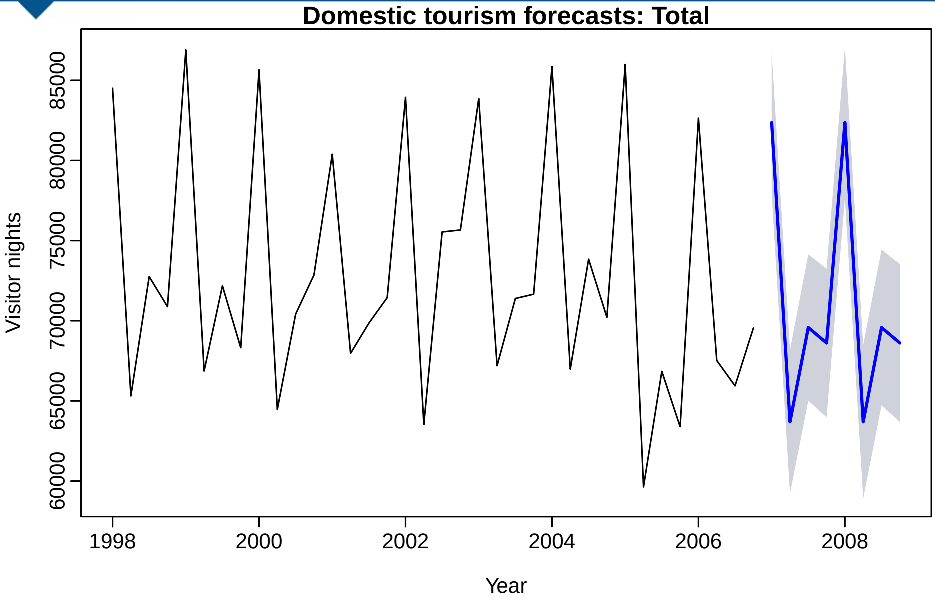
- Apply each of 30 models that are appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AIC:
  $$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$
  where $p = \#$ parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Automatic forecasting

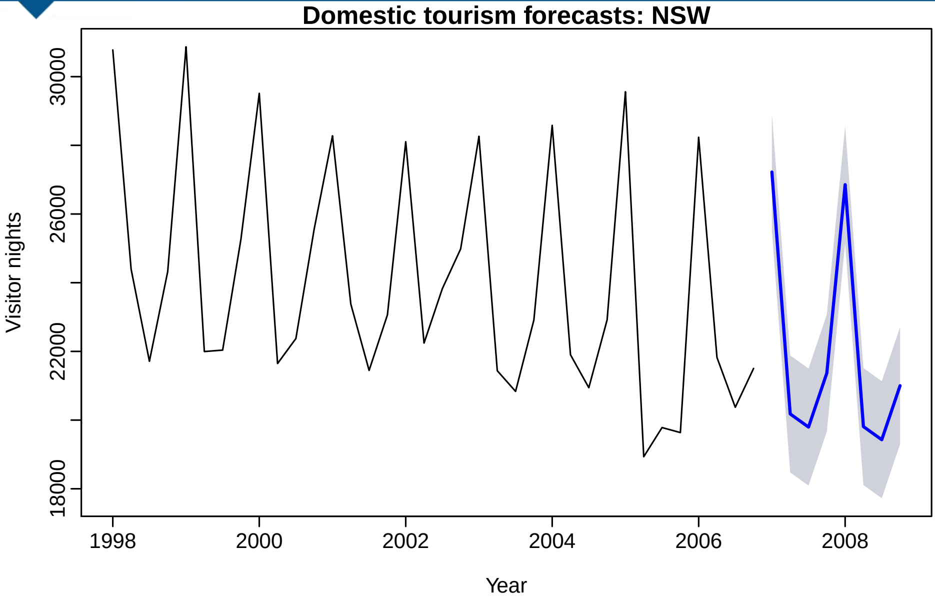
From Hyndman et al. (IJF, 2002):

- Apply each of 30 models that are appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AIC:
$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$
where $p = \#$ parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

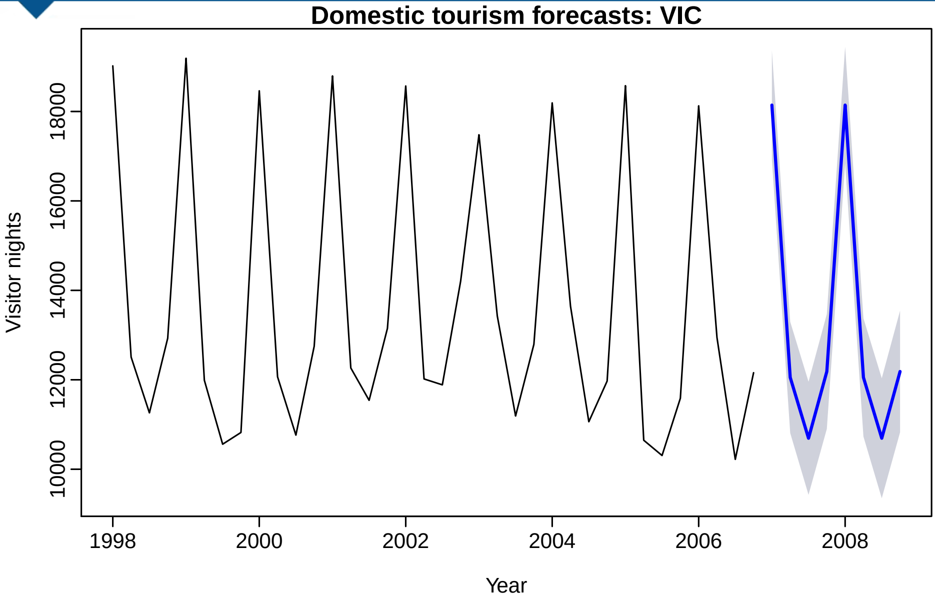
Base forecasts



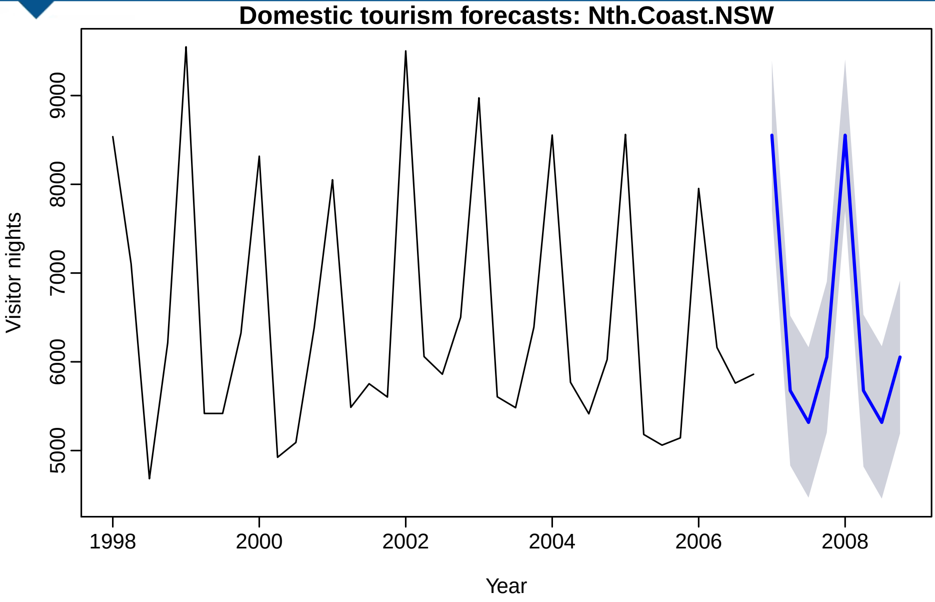
Base forecasts



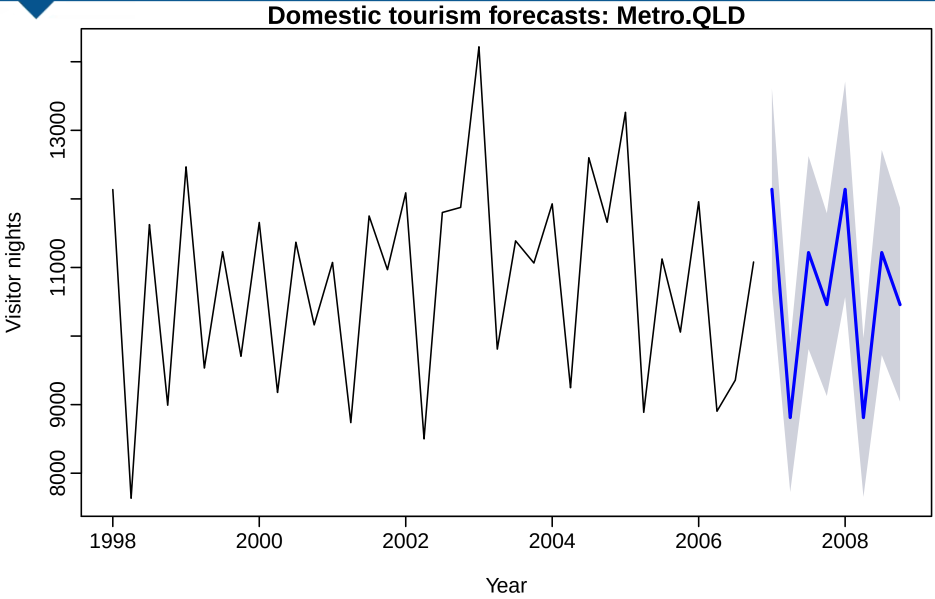
Base forecasts



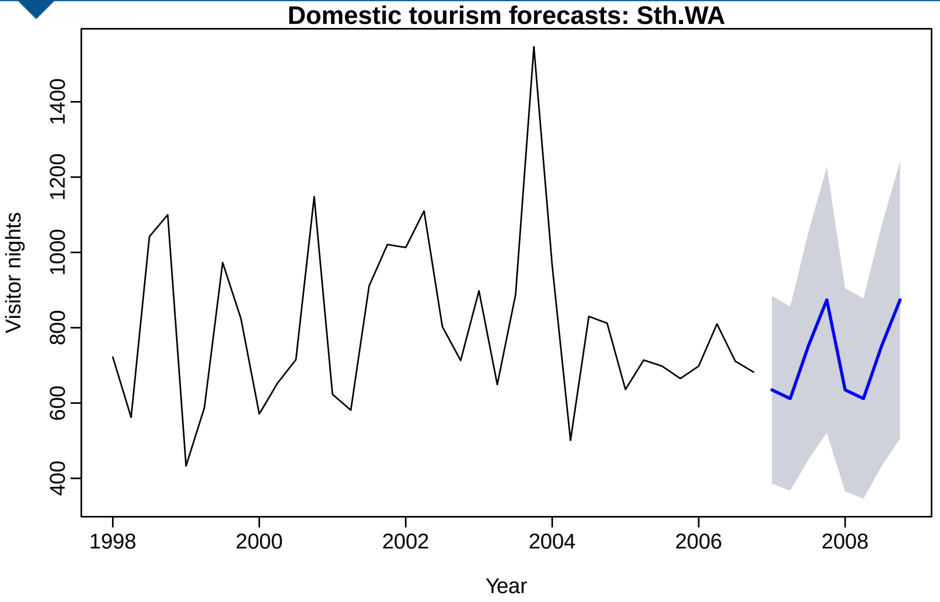
Base forecasts



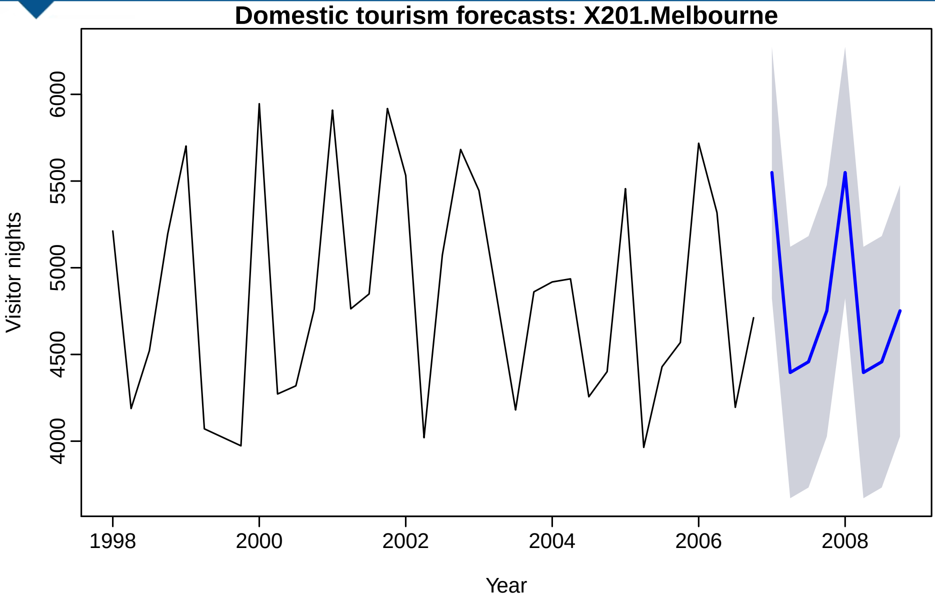
Base forecasts



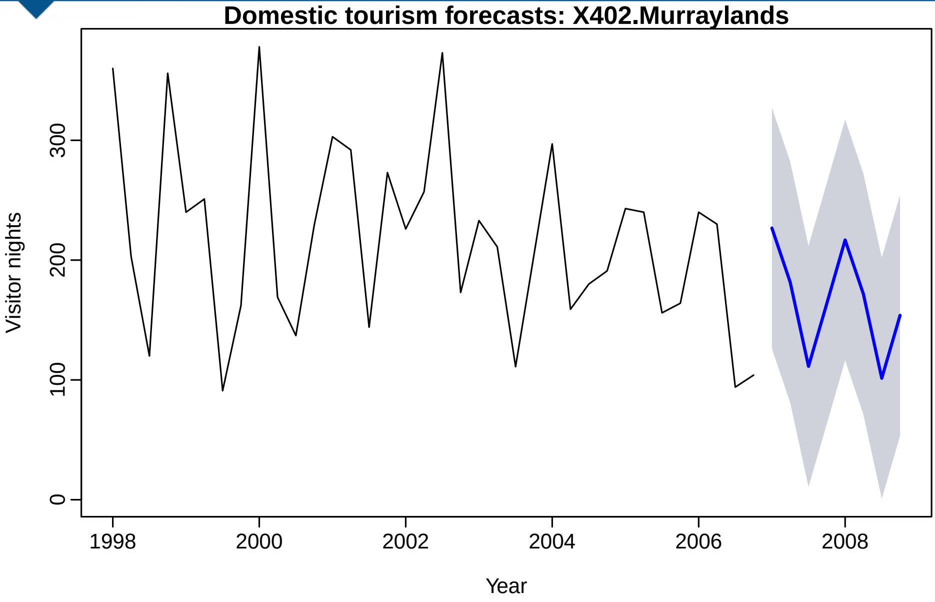
Base forecasts



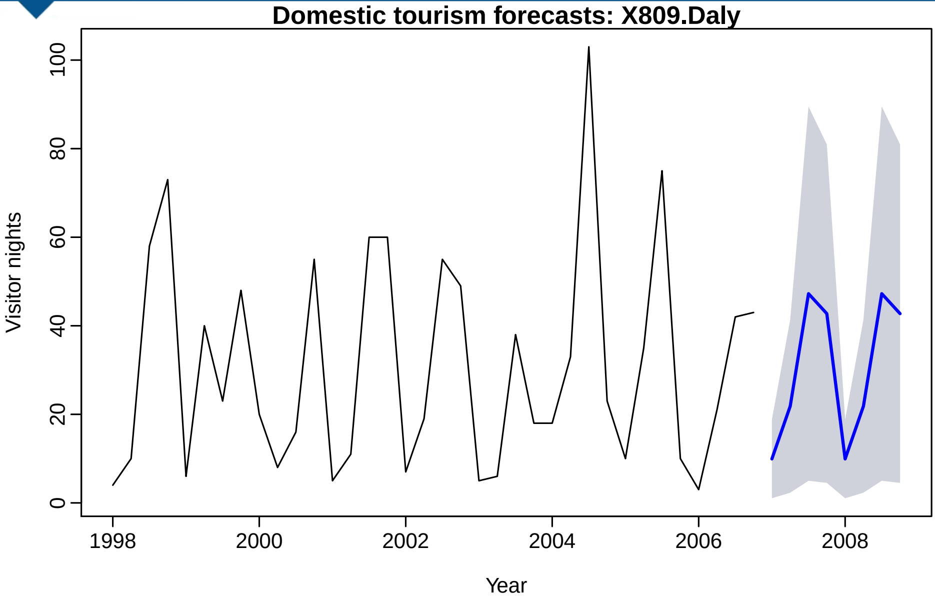
Base forecasts



Base forecasts



Base forecasts



Hierarchy: states, zones, regions

MAPE	*Forecast Horizon (h)* 1	2	4	6	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	**3.68**	**3.56**	3.97	4.57	4.25	4.04
Averaging	3.76	3.60	4.01	4.58	**4.22**	4.06
			Level 1: 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	**10.44**	**10.17**	**10.47**	**10.97**	**10.98**	**10.67**
Averaging	10.59	10.36	10.69	11.27	11.21	10.89

Based on a rolling forecast origin with at least 12 observations in the training set.

Hierarchy: states, zones, regions

MAPE	Forecast Horizon (h) 1	2	4	6	8	Average
			Level 2: 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	**14.63**	**14.62**	**14.68**	**15.17**	**15.25**	**14.94**
Averaging	14.79	14.79	14.85	15.46	15.49	15.14
			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	**31.68**	**31.22**	**31.08**	**32.41**	**32.77**	**31.89**
Averaging	32.84	32.20	32.06	33.44	34.04	32.96

Based on a rolling forecast origin with at least 12 observations in the training set.

Groups: Purpose, states, capital

MAPE	*Forecast Horizon (h)* 1	2	4	6	8	Average
			Top Level: Australia			
Bottom-up	**3.48**	**3.30**	4.04	4.56	4.58	4.03
OLS	3.80	3.64	**3.94**	**4.22**	**4.35**	**3.95**
Scaling	3.65	3.45	4.00	4.52	4.57	4.04
Averaging	3.59	3.33	3.99	4.56	4.58	4.04
			Level 1: Purpose of travel			
Bottom-up	8.14	8.37	9.02	9.39	9.52	8.95
OLS	**7.94**	**7.91**	8.66	**8.66**	**9.29**	**8.54**
Scaling	7.99	8.10	**8.59**	9.09	9.43	8.71
Averaging	8.04	8.21	8.79	9.25	9.44	8.82

Based on a rolling forecast origin with at least 12 observations in the training set.

Groups: Purpose, states, capital

MAPE	1	2	4	6	8	Average
	Forecast Horizon (h)					
Level 2: States						
Bottom-up	**21.34**	21.75	22.39	23.26	23.31	22.58
OLS	22.17	21.80	23.53	23.15	23.90	22.99
Scaling	21.49	21.62	**22.20**	**23.13**	23.25	22.51
Averaging	21.38	**21.61**	22.30	23.17	**23.24**	**22.51**
Bottom Level: Capital city versus other						
Bottom-up	31.97	31.65	32.19	33.70	33.47	32.62
OLS	32.31	**30.92**	32.41	**33.35**	34.13	32.55
Scaling	32.12	31.36	32.18	33.36	33.43	32.52
Averaging	**31.92**	31.39	**32.04**	33.51	**33.39**	**32.49**

Based on a rolling forecast origin with at least 12 observations in the training set.

Outline

hts package for R

hts: Hierarchical and grouped time series

Methods for analysing and forecasting hierarchical and grouped time series

Version:	3.01
Depends:	forecast
Imports:	SparseM
Published:	2013-05-07
Author:	Rob J Hyndman, Roman A Ahmed, and Han Lin Shang
Maintainer:	Rob J Hyndman <Rob.Hyndman at monash.edu>
License:	GPL-2 $\vert$ GPL-3 [expanded from: GPL ($\geq$ 2)]

Example using R

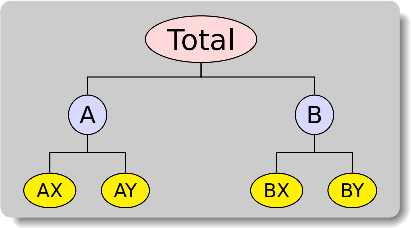
```
library(hts)

# bts is a matrix containing the bottom level time series
# g describes the grouping/hierarchical structure
y <- hts(bts, g=c(1,1,2,2))
```

Example using R

```
library(hts)

# bts is a matrix containing the bottom level time series
# g describes the grouping/hierarchical structure
y <- hts(bts, g=c(1,1,2,2))
```



Example using R

```
library(hts)

# bts is a matrix containing the bottom level time series
# g describes the grouping/hierarchical structure
y <- hts(bts, g=c(1,1,2,2))

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

Example using R

```
library(hts)

# bts is a matrix containing the bottom level time series
# g describes the grouping/hierarchical structure
y <- hts(bts, g=c(1,1,2,2))

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

# Select your own methods
ally <- allts(y)
allf <- matrix(, nrow=10, ncol=ncol(ally))
for(i in 1:ncol(ally))
  allf[,i] <- mymethod(ally[,i], h=10)
allf <- ts(allf, start=2004)
# Reconcile forecasts so they add up
fc2 <- combinef(allf, Smatrix(y))
```

hts function

Usage

```
hts(y, g)
gts(y, g, hierarchical=FALSE)
```

Arguments

`y`	Multivariate time series containing the bottom level series
`g`	Group matrix indicating the group structure, with one column for each series when completely disaggregated, and one row for each grouping of the time series.
`hierarchical`	Indicates if the grouping matrix should be treated as hierarchical.

Details

`hts` is simply a wrapper for `gts(y,g,TRUE)`. Both return an object of class `gts`.

forecast.gts function

Usage

```
forecast(object, h,
  method = c("comb", "bu", "mo", "tdgsf", "tdgsa", "tdfp", "all"),
  fmethod = c("ets", "rw", "arima"), level, positive = FALSE,
  xreg = NULL, newxreg = NULL, ...)
```

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
`level`	Level used for "middle-out" method (when `method="mo"`)
`positive`	If TRUE, forecasts are forced to be strictly positive
`xreg`	When `fmethod = "arima"`, a vector or matrix of external regressors, which must have the same number of rows as the original univariate time series
`newxreg`	When `fmethod = "arima"`, a vector or matrix of external regressors, which must have the same number of rows as the original univariate time series
`...`	Other arguments passing to `ets` or `auto.arima`

Utility functions

`allts(y)`	Returns all series in the hierarchy
`Smatrix(y)`	Returns the summing matrix
`combinef(f)`	Combines initial forecasts optimally.

More information

hts: An R Package for Forecasting Hierarchical or Grouped Time Series

Rob J Hyndman, George Athanasopoulos, Han Lin Shang

Vignette on CRAN

Abstract

This paper describes several methods that are curr
for forecasting hierarchical time series. The methods included are: top-down, bottom-up, middle-out and optimal combination. The implementation of these methods is illustrated by using regional infant mortality counts in Australia.

Keywords: top-down, bottom-up, middle-out, optimal combination .

Introduction

Advances in data collection and storage have resulted in large numbers of time series that are hierarchical in structure, and clusters of which may be correlated. In many applications the

Outline

References

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

RJ Hyndman, RA Ahmed, and HL Shang (2013). *hts: Hierarchical time series*. `cran.r-project.org/package=hts`.

RJ Hyndman and G Athanasopoulos (2013). *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 and Data Analysis* **55**(9), 2579–2589

RJ Hyndman, RA Ahmed, and HL Shang (2013). *hts: Hierarchical time series*.

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