MONASH University

MONASH BUSINESS SCHOOL

Forecasting: principles and practice

Rob J Hyndman

3.3 Hierarchical forecasting

Outline

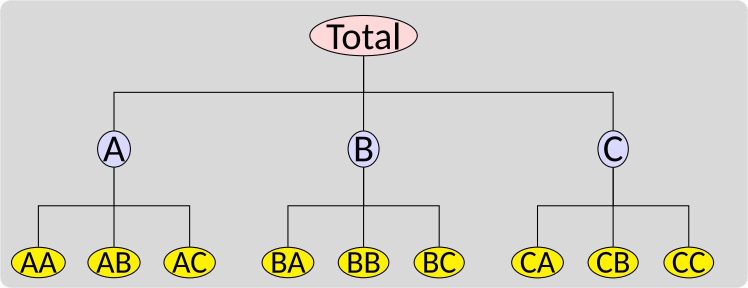
1 Hierarchical and grouped time series

2 Forecasting framework

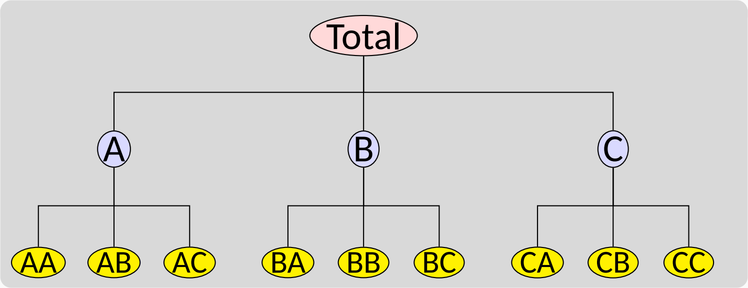
3 Application: Australian tourism

4 hts package for R

Hierarchical and grouped time series



Hierarchical and grouped time series



Examples

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

Forecasting the PBS

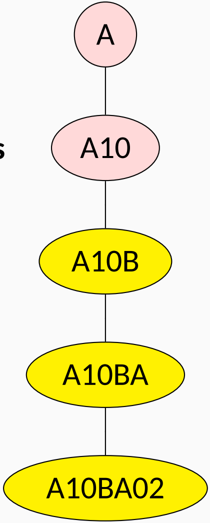
ATC drug classification

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

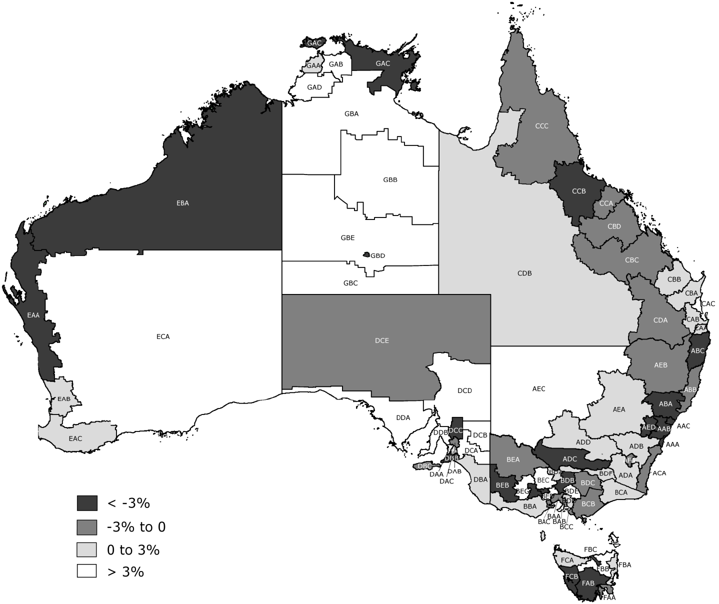
Drugs used in diabetes

Blood glucose lowering drugs

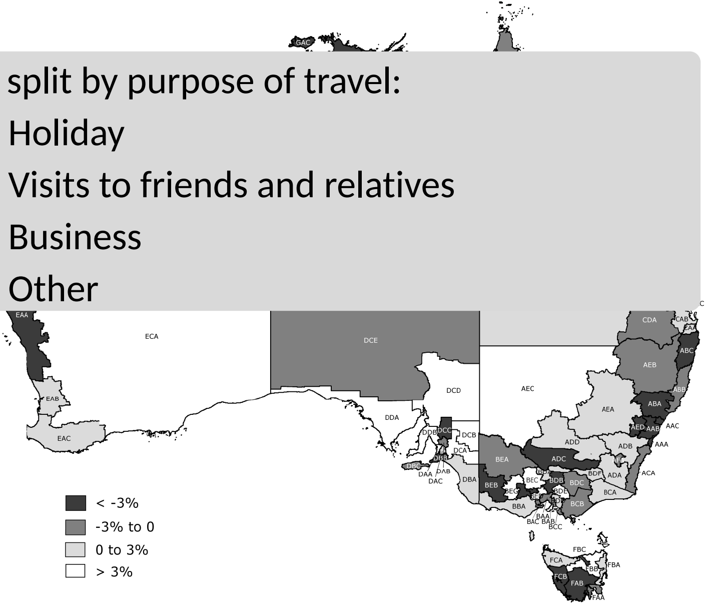
Biguanides

Metformin

Australian tourism



Australian tourism



Also split by purpose of travel:

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

Hierarchical/grouped time series

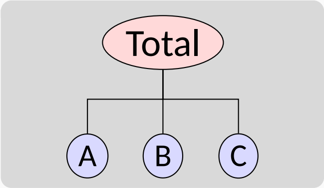
- A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.

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

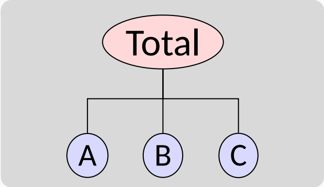
- A **grouped time series** is a collection of time series that are aggregated in a number of non-hierarchical ways.

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

Hierarchical data



Hierarchical data

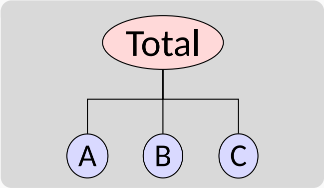


Y_t : observed aggregate of all series at time t.

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

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

Hierarchical data



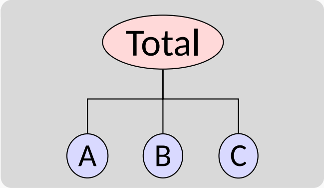
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: 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.

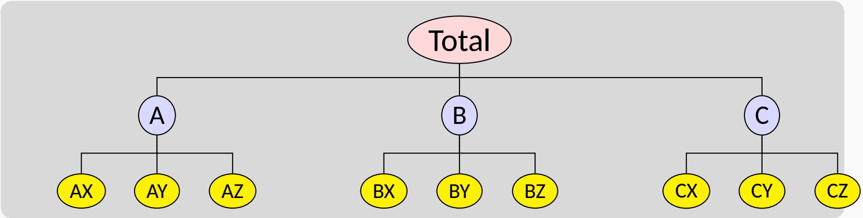
$Y_{X,t}$: observation on series X at time t.

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

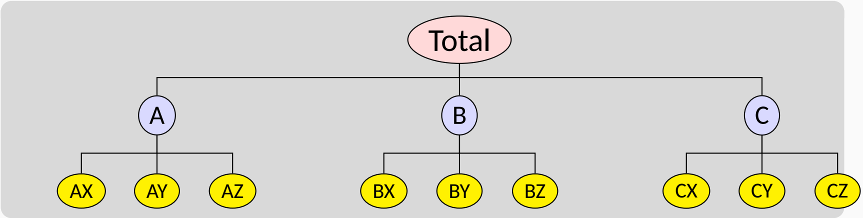
$$\boldsymbol{y}_t = [Y_t, Y_{A,t}, Y_{B,t}, Y_{C,t}]' = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

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

Hierarchical data

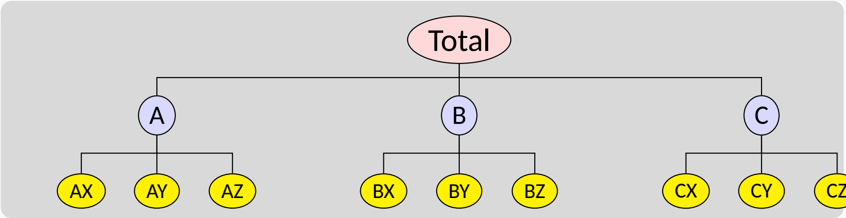


Hierarchical data



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

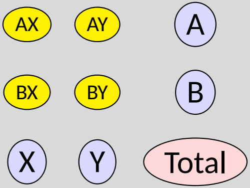
Hierarchical data



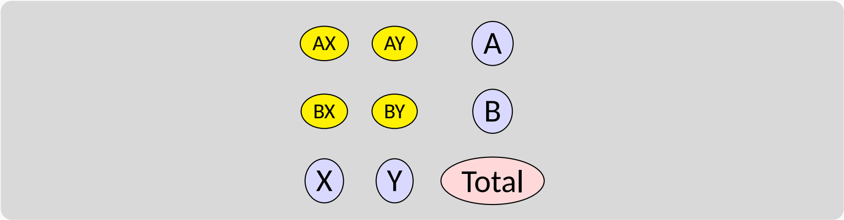
$$\boldsymbol{y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0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

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

Grouped data

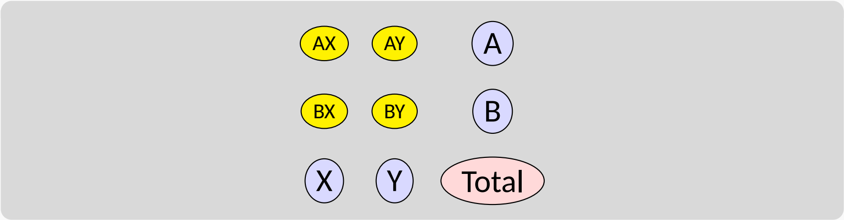


Grouped data



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

Grouped data



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

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

Outline

1 Hierarchical and grouped time series

2 Forecasting framework

3 Application: Australian tourism

4 hts package for R

Forecasting notation

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

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.)

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 $\boldsymbol{P}$.

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

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

Bottom-up forecasts are obtained using

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

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

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

Top-down forecasts are obtained using

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

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 and variance

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

General properties: bias and variance

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

Bias

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

General properties: bias and variance

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

Bias

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

Variance

Let error variance of h-step base forecasts $\hat{\boldsymbol{y}}_n(h)$ be

$$\Sigma_h = \text{Var}[\boldsymbol{y}_{n+h} - \hat{\boldsymbol{y}}_n(h) \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_n]$$

Then error variance of the reconciled forecasts is

$$\text{Var}[\boldsymbol{y}_{n+h} - \tilde{\boldsymbol{y}}_n(h) \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_n] = \boldsymbol{S}\boldsymbol{P}\Sigma_h\boldsymbol{P}'\boldsymbol{S}'$$

Optimal forecast reconciliation

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

Optimal forecast reconciliation

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

Theorem: MinT Reconciliation

If $\boldsymbol{P}$ satisfies $\boldsymbol{SPS} = \boldsymbol{S}$, then

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

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

Optimal forecast reconciliation

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

Theorem: MinT Reconciliation

If $\boldsymbol{P}$ satisfies $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$, then

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

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

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

Optimal forecast reconciliation

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

Assume that $\Sigma_h = k_h\Sigma_1$ to simplify computations.

Optimal forecast reconciliation

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

Assume that $\Sigma_h = k_h\Sigma_1$ to simplify computations.

WLS solution

- Approximate Σ_1 by its diagonal.

Optimal forecast reconciliation

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

Assume that $\Sigma_h = k_h\Sigma_1$ to simplify computations.

WLS solution

- Approximate Σ_1 by its diagonal.

GLS solution

- Estimate Σ_1 using shrinkage to the diagonal.

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.
- Conceptually easy to implement: regression of base forecasts on structure matrix.
- Weights are independent of the data and of the covariance structure of the hierarchy.

Outline

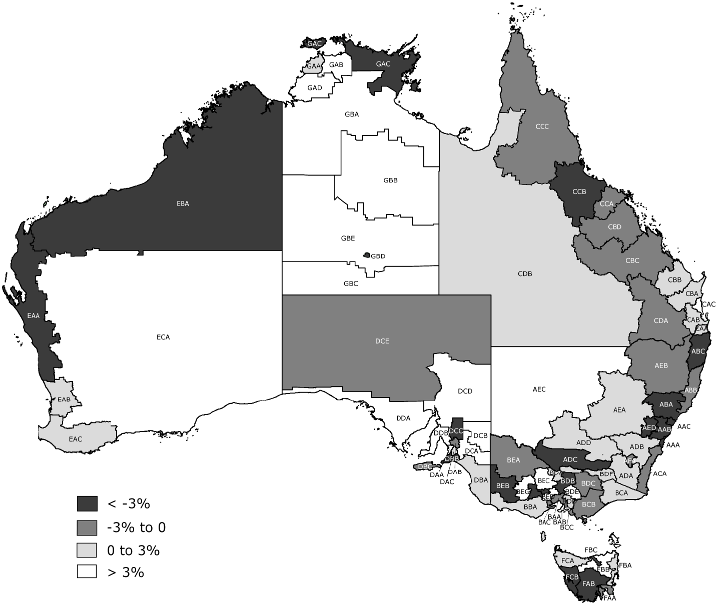
1 Hierarchical and grouped time series

2 Forecasting framework

3 Application: Australian tourism

4 hts package for R

Australian tourism

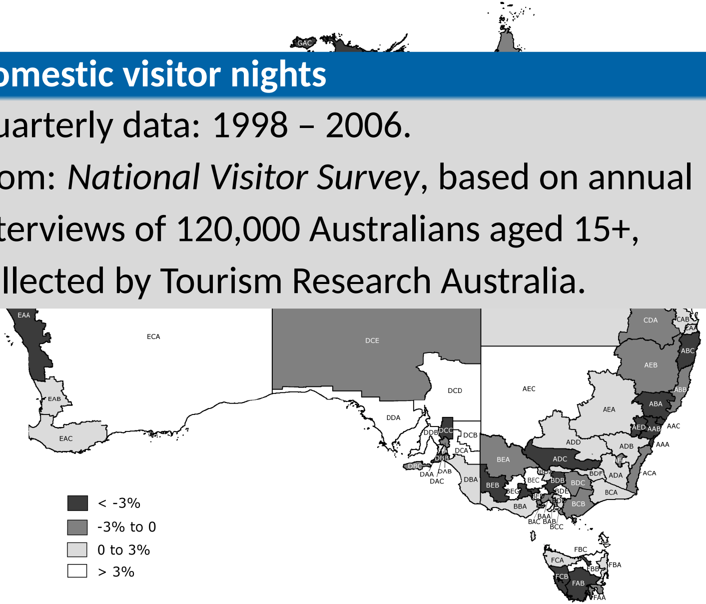


Australian tourism

Domestic visitor nights

Quarterly data: 1998 – 2006.

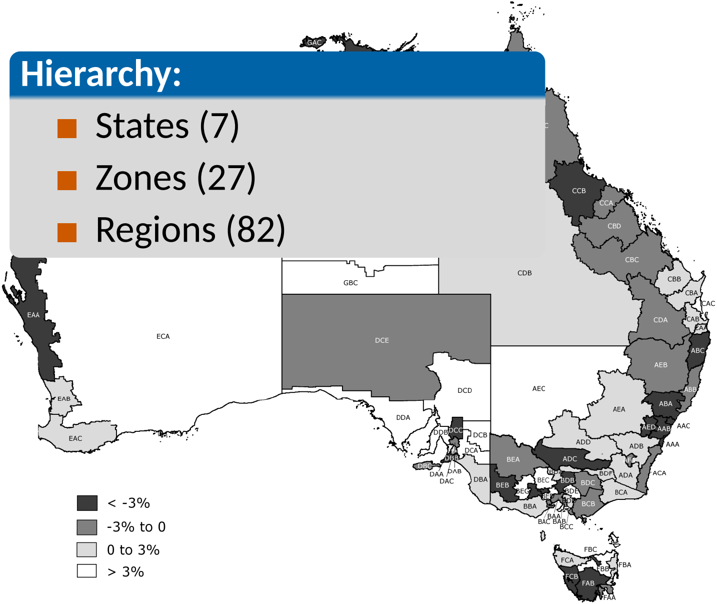
From: *National Visitor Survey*, based on annual interviews of 120,000 Australians aged 15+, collected by Tourism Research Australia.



Australian tourism

Hierarchy:

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



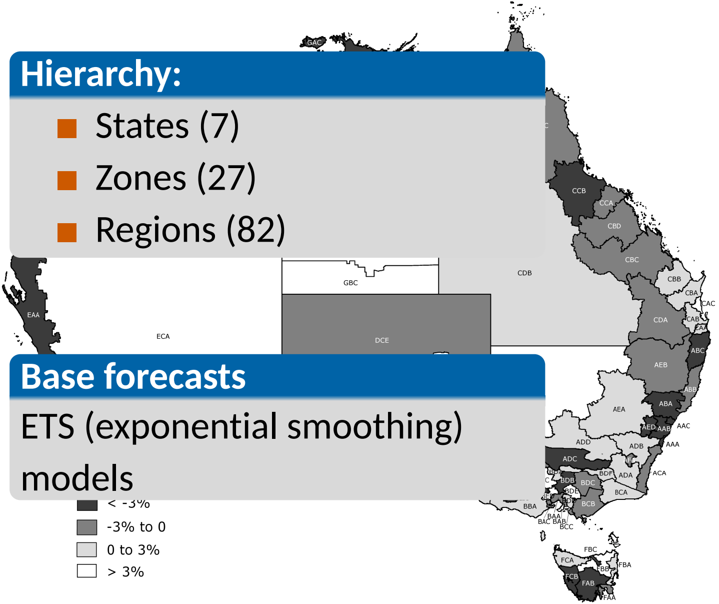
Australian tourism

Hierarchy:

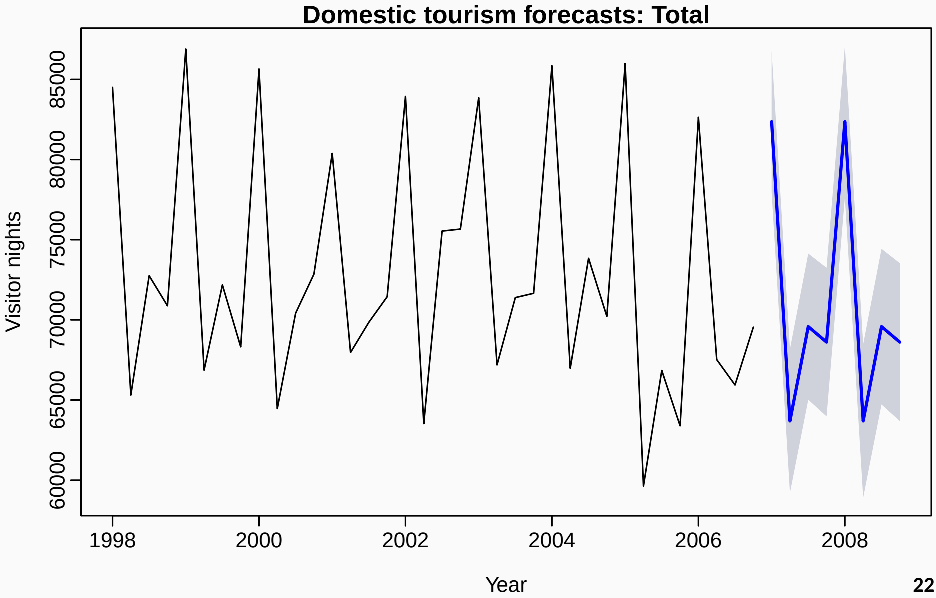
- States (7)
- Zones (27)
- Regions (82)

Base forecasts

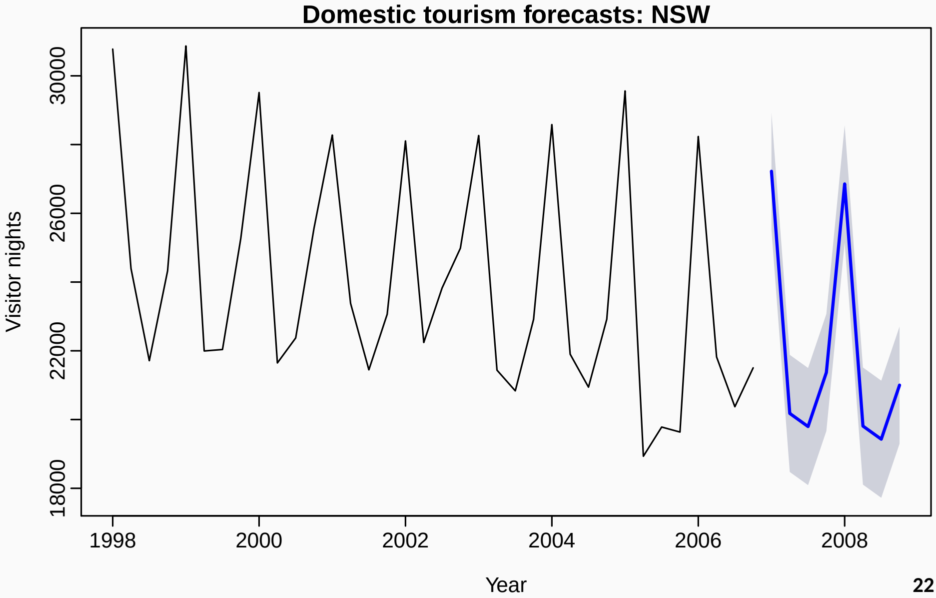
ETS (exponential smoothing) models



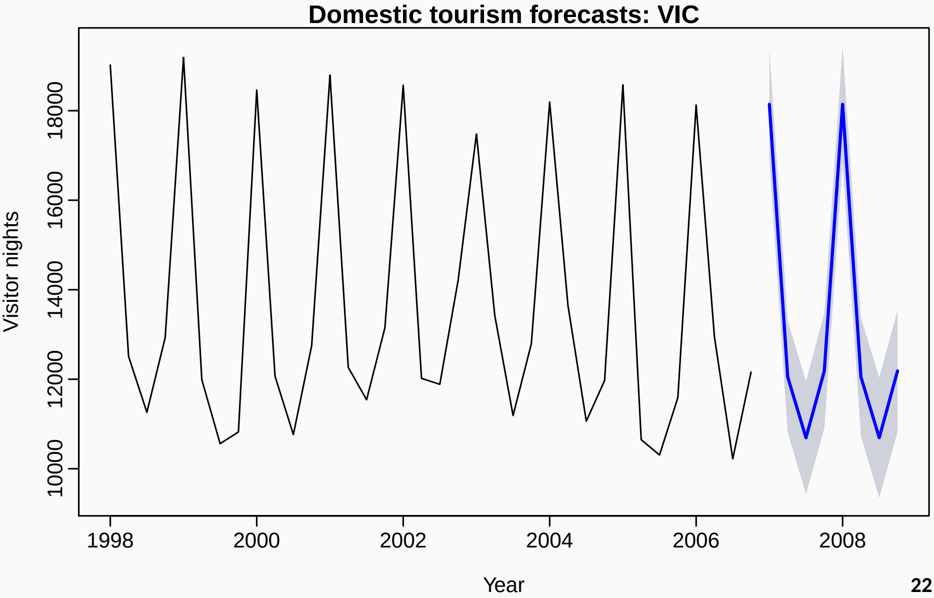
Base forecasts



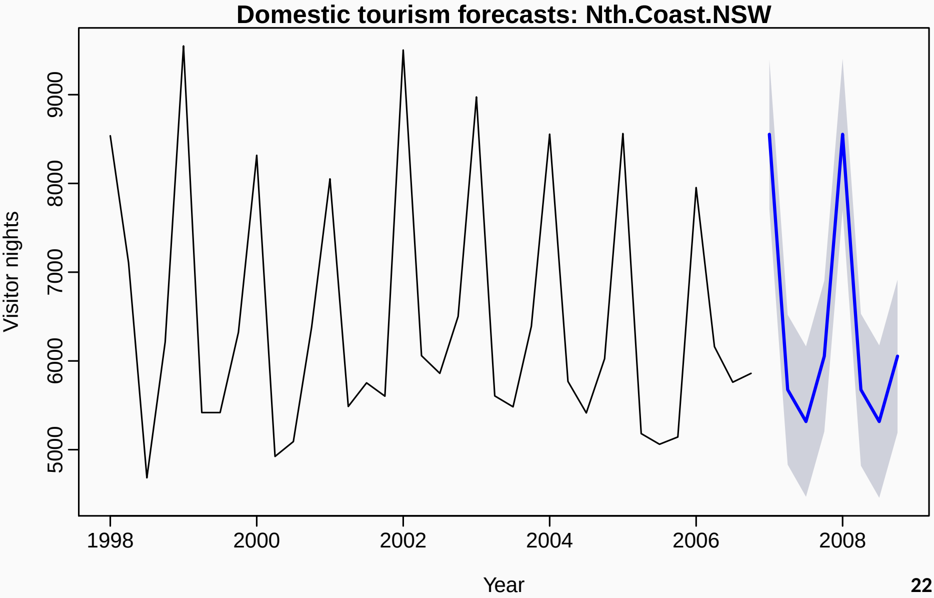
Base forecasts



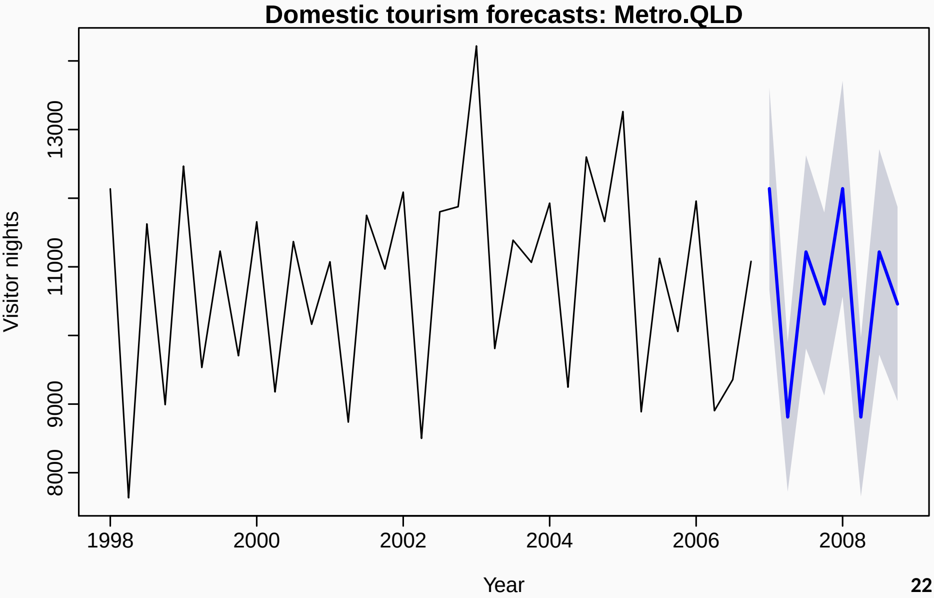
Base forecasts



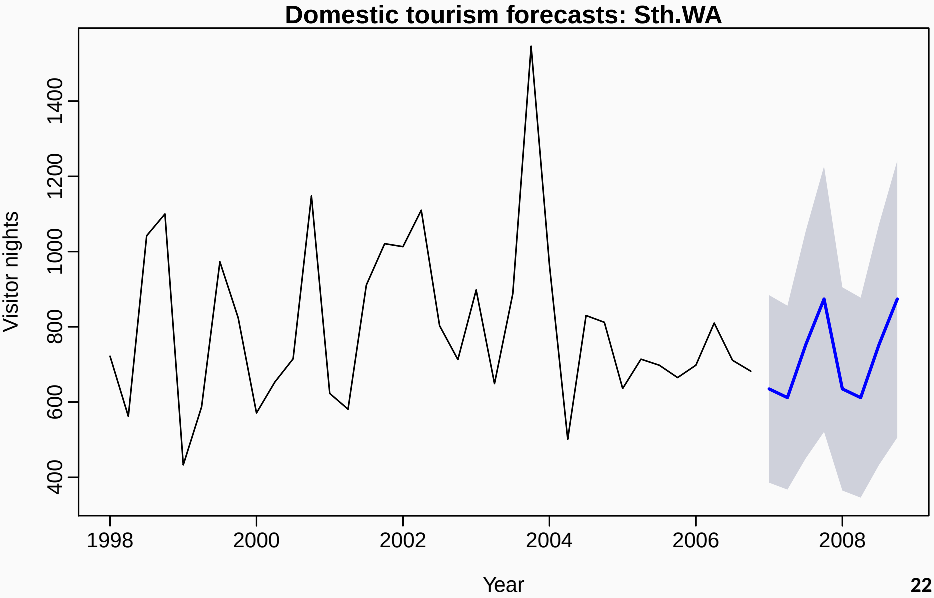
Base forecasts



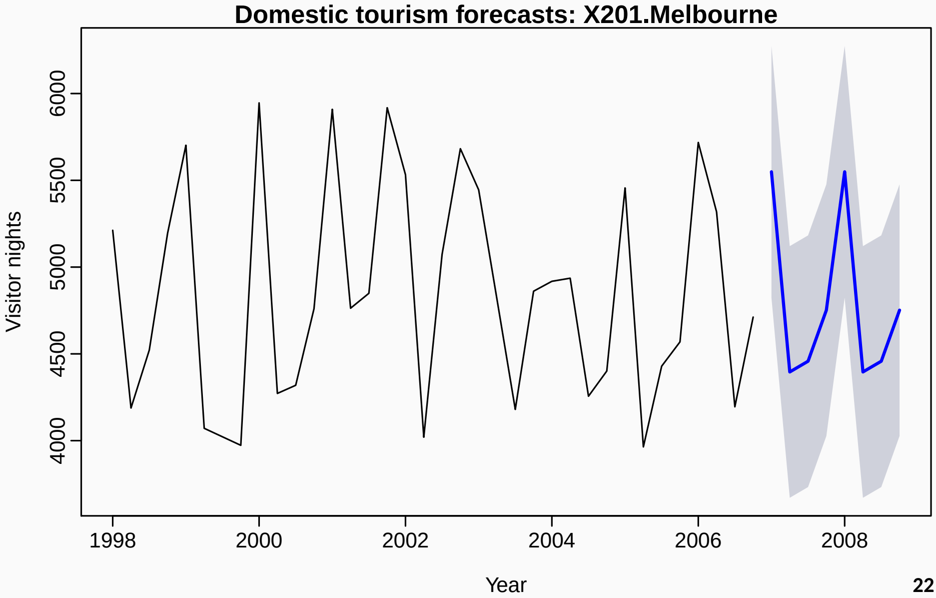
Base forecasts



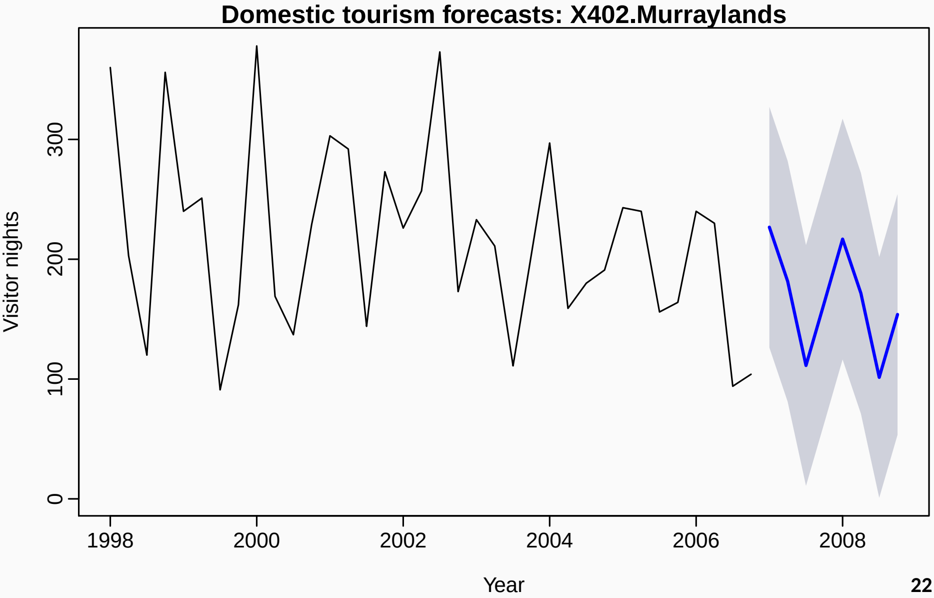
Base forecasts



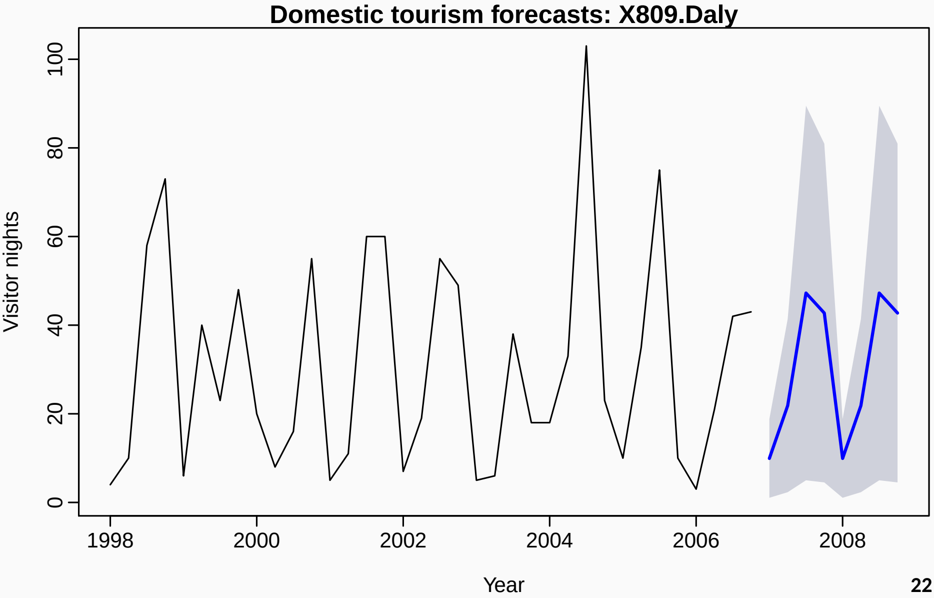
Base forecasts



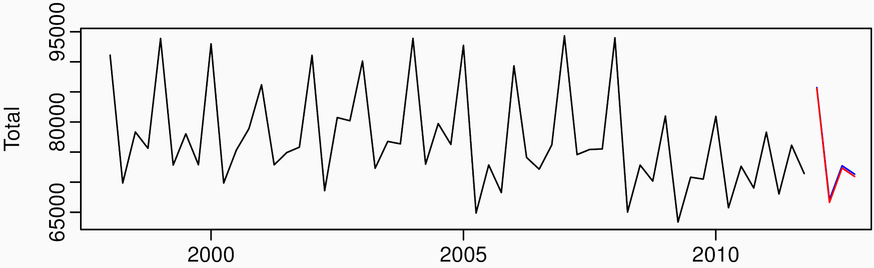
Base forecasts



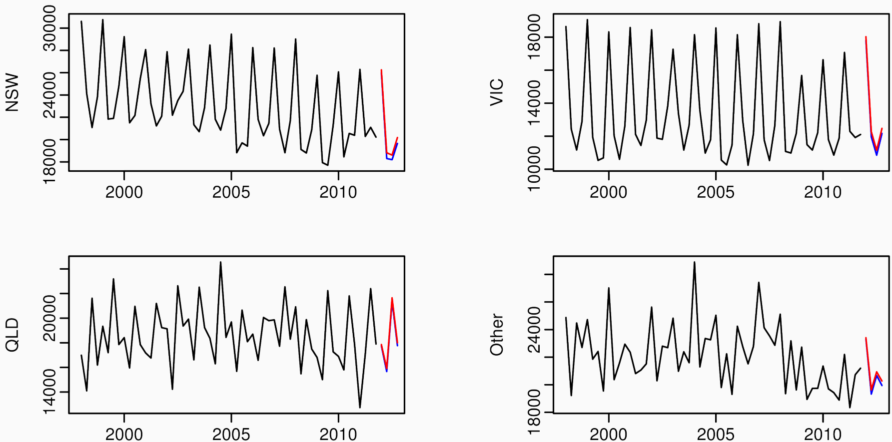
Base forecasts



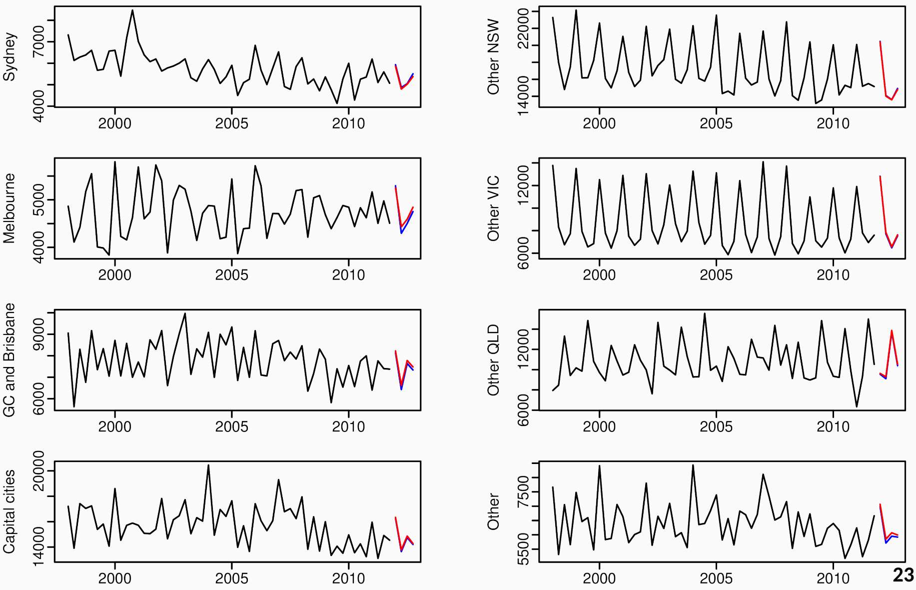
Reconciled forecasts



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.

Hierarchy: states, zones, regions

RMSE	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	**Ave**
				Australia			
Base	1762.04	1770.29	1766.02	1818.82	1705.35	1721.17	**1757.28**
Bottom	1736.92	1742.69	1722.79	1752.74	1666.73	1687.43	**1718.22**
WLS	1705.21	1715.87	1703.75	1729.56	1627.79	1661.24	**1690.57**
GLS	1704.64	1715.60	1705.31	1729.04	1626.36	1661.64	**1690.43**
				States			
Base	399.77	404.16	401.92	407.26	395.38	401.17	**401.61**
Bottom	404.29	406.95	404.96	409.02	399.80	401.55	**404.43**
WLS	398.84	402.12	400.71	405.03	394.76	398.23	**399.95**
GLS	398.84	402.16	400.86	405.03	394.59	398.22	**399.95**
				Regions			
Base	93.15	93.38	93.45	93.79	93.50	93.56	**93.47**
Bottom	93.15	93.38	93.45	93.79	93.50	93.56	**93.47**
WLS	93.02	93.32	93.38	93.72	93.39	93.53	**93.39**
GLS	92.98	93.27	93.34	93.66	93.34	93.46	**93.34**

Forecast horizon

Outline

1 Hierarchical and grouped time series

2 Forecasting framework

3 Application: Australian tourism

4 hts package for R

hts package for R

hts: Hierarchical and grouped time series

Methods for analysing and forecasting hierarchical and grouped time series

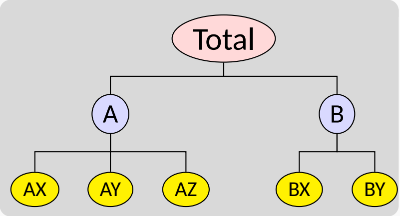
Version:	5.1.5
Depends:	R ($\geq$ 3.2.0), forecast ($\geq$ 8.1)
Published:	2018-03-26
Author:	Rob J Hyndman, Earo Wang, Shanika Wickramasuriya, Alan Lee
Maintainer:	Earo Wang $<$Earo.Wang at monash.edu$>$
BugReports:	https://github.com/earowang/hts/issues
License:	GPL ($\geq$ 2)
URL:	http://pkg.earo.me/hts

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)))
```

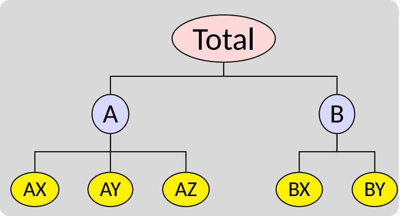
Example using R

```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)))
```



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)))
```




```r
# 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("wls", "ols", "minmt", "nseries"),
  covariance = c("shr","sam"),
  positive = TRUE,
  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.
`parallel`	If TRUE, allow parallel processing
`num.cores`	If parallel = TRUE, specify how many cores are going to be used