3.3 Hierarchical forecasting
1 Hierarchical and grouped time series

2 Lab session 15

3 Temporal hierarchies

4 Lab session 16
Australian tourism demand

Forecasting: principles and practice

Hierarchical and grouped time series
Australian tourism demand

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

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

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

Provided by a large spectacle manufacturer

Split by brand (26), gender (3), price range (6), materials (4), and stores (600)

About 1 million bottom-level series
A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.

**Examples**
- Tourism by state and region
Hierarchical time series

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

**Examples**

- Tourism by state and region
A **grouped time series** is a collection of time series that can be grouped together in a number of non-hierarchical ways.

Examples

- Labour turnover by occupation and state
- Spectacle sales by brand, gender, stores, etc.
A **grouped time series** is a collection of time series that can be grouped together in a number of non-hierarchical ways.

Examples
- Labour turnover by occupation and state
- Spectacle sales by brand, gender, stores, etc.
Grouped time series

A grouped time series is a collection of time series that can be grouped together in a number of non-hierarchical ways.

Examples

- Labour turnover by occupation and state
- Spectacle sales by brand, gender, stores, etc.
The problem

1. How to forecast time series at all nodes such that the forecasts add up in the same way as the original data?

2. Can we exploit relationships between the series to improve the forecasts?
The problem

1. How to forecast time series at all nodes such that the forecasts add up in the same way as the original data?

2. Can we exploit relationships between the series to improve the forecasts?
Forecast all series at all levels of aggregation using an automatic forecasting algorithm (e.g., ets, auto.arima, ...)

Reconcile the resulting forecasts so they add up correctly using least squares optimization (i.e., find closest reconciled forecasts to the original forecasts).

This is all available in the hts package in R.
The solution

1. Forecast all series at all levels of aggregation using an automatic forecasting algorithm (e.g., ets, auto.arima, ...)

2. Reconcile the resulting forecasts so they add up correctly using least squares optimization (i.e., find closest reconciled forecasts to the original forecasts).

3. This is all available in the hts package in R.
The solution

1. Forecast all series at all levels of aggregation using an automatic forecasting algorithm (e.g., ets, auto.arima, ...)

2. Reconcile the resulting forecasts so they add up correctly using least squares optimization (i.e., find closest reconciled forecasts to the original forecasts).

3. This is all available in the hts package in R.
hts: Hierarchical and Grouped Time Series

Methods for analysing and forecasting hierarchical and grouped time series

Version: 5.0
Depends: R (≥ 3.0.2), forecast (≥ 5.0), SparseM, Matrix, matrixcalc
Imports: parallel, utils, methods, graphics, grDevices, stats
LinkingTo: Rcpp (≥ 0.11.0), RcppEigen
Suggests: testthat
Published: 2016-04-06
Author: Rob J Hyndman, Earo Wang, Alan Lee, Shanika Wickramasuriya
Maintainer: Rob J Hyndman <Rob.Hyndman at monash.edu>
BugReports: https://github.com/robjhyndman/hts/issues
License: GPL (≥ 2)
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)))
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)))
library(hts)

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

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

Usage

\texttt{gts(y, characters)}

Arguments

\texttt{y} \hspace{1cm} \text{Multivariate time series containing the bottom level series}

\texttt{characters} \hspace{1cm} \text{Vector of integers, or list of vectors, showing how column names indicate group structure.}

Example

\texttt{bnames <- c("VIC1F","VIC1M","VIC2F","VIC2M","VIC3F","VIC3M", "NSW1F","NSW1M","NSW2F","NSW2M","NSW3F","NSW3M")}
\texttt{bts <- matrix(ts(rnorm(120)), ncol = 12)}
\texttt{colnames(bts) <- bnames}
\texttt{x <- gts(bts, characters = c(3, 1, 1))}
Example 2

bnames <-
  c("VICMelbAA","VICMelbAB",
     "VICGeelAA","VICGeelAB",
     "VICMelbBA","VICMelbBB",
     "VICGeelBA","VICGeelBB",
     "NSWSyndAA","NSWSyndAB",
     "NSWWollAA","NSWWollAB",
     "NSWSyndBA","NSWSyndBB",
     "NSWWollBA","NSWWollBB")
bts <- matrix(ts(rnorm(160)), ncol = 16)
colnames(bts) <- bnames
x <- gts(bts, characters = list(c(3, 4), c(1, 1))))
forecast.gts function

Usage
forecast(object, h, 
   method = c("comb", "bu", "mo","tdgsa", "tdgsf", "tdfp"), 
   weights = c("wls", "ols", "mint", "nseries"), 
   fmethod = c("ets", "arima", "rw"), 
   algorithms = c("lu", "cg", "chol", "recursive", "slm"), 
   covariance = c("shr", "sam"), 
   positive = FALSE, 
   parallel = FALSE, num.cores = 2, ...)

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

Hierarchy:

- States (7)
- Zones (27)
- Regions (82)
Example: Australian tourism

**Hierarchy:**
- States (7)
- Zones (27)
- Regions (82)

**Base forecasts**
ETS (exponential smoothing) models

Forecasting: principles and practice  Hierarchical and grouped time series  15
Domestic tourism forecasts: Total

<table>
<thead>
<tr>
<th>Year</th>
<th>Visitor nights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>60000</td>
</tr>
<tr>
<td>2000</td>
<td>65000</td>
</tr>
<tr>
<td>2002</td>
<td>70000</td>
</tr>
<tr>
<td>2004</td>
<td>75000</td>
</tr>
<tr>
<td>2006</td>
<td>80000</td>
</tr>
<tr>
<td>2008</td>
<td>85000</td>
</tr>
</tbody>
</table>
Domestic tourism forecasts: VIC

Visitor nights over time for VIC from 1998 to 2008.
Domestic tourism forecasts: Nth.Cost.NSW
Domestic tourism forecasts: Metro.QLD

<table>
<thead>
<tr>
<th>Year</th>
<th>Visitor nights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>8000</td>
</tr>
<tr>
<td>2000</td>
<td>9000</td>
</tr>
<tr>
<td>2002</td>
<td>11000</td>
</tr>
<tr>
<td>2004</td>
<td>13000</td>
</tr>
<tr>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
</tr>
</tbody>
</table>
Base forecasts

Domestic tourism forecasts: Sth.WA

Visitor nights vs. Year:
- 1998: 400
- 2000: 600
- 2002: 800
- 2004: 1200
- 2006: 1400

Year range: 1998 to 2008
Domestic tourism forecasts: X402.Murraylands

Year
Visitor nights
0 100 200 300
Domestic tourism forecasts: X809.Daly

<table>
<thead>
<tr>
<th>Year</th>
<th>Visitor nights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
</tr>
</tbody>
</table>
Forecast evaluation

Training sets

Test sets $h = 1$

→ time
Forecast evaluation

Training sets

Test sets $h = 1$
Forecast evaluation

Training sets

Test sets $h = 1$

Time
Forecast evaluation

Training sets

Test sets $h = 1$
Forecast evaluation

Training sets

Test sets $h = 1$
Forecast evaluation

Training sets

Test sets $h = 1$

Forecasting: principles and practice
Hierarchical and grouped time series
17
Forecast evaluation

Training sets

Test sets $h = 1$
Forecast evaluation

**Training sets**

**Test sets** $h = 1$

![Diagram showing training and test sets](image-url)
Forecast evaluation

Training sets

Test sets $h = 1$
Forecast evaluation

**Training sets**

**Test sets** \( h = 1 \)
Forecast evaluation

Training sets

Test sets $h = 1$

Forecasting: principles and practice
Hierarchical and grouped time series
Forecast evaluation

Training sets

Test sets \( h = 1 \)
Forecast evaluation

Training sets

Test sets $h = 1$

Forecasting: principles and practice  Hierarchical and grouped time series  17
Forecast evaluation

Training sets

Test sets $h = 1$
Forecast evaluation

Training sets

Test sets $h = 1$
Forecast evaluation

Training sets

Test sets $h = 1$

$\rightarrow$ time
Forecast evaluation

Training sets

Test sets $h = 1$

$\Rightarrow$ time

Forecasting: principles and practice
Hierarchical and grouped time series
17
Forecast evaluation

Training sets

Test sets $h = 1$

→ time

Forecasting: principles and practice
Hierarchical and grouped time series
17
Forecast evaluation

Training sets

Test sets $h = 1$
Forecast evaluation

Training sets  Test sets $h = 1$

Forecasting: principles and practice  Hierarchical and grouped time series
Forecast evaluation

Training sets

Test sets \( h = 2 \)
Forecast evaluation

Training sets

Test sets $h = 3$

Forecasting: principles and practice
Hierarchical and grouped time series
Forecast evaluation

Training sets

Test sets \( h = 4 \)
Forecast evaluation

Training sets

Test sets $h = 5$

Forecasting: principles and practice

Hierarchical and grouped time series
Forecast evaluation

Training sets

Test sets $h = 6$
## Hierarchy: states, zones, regions

### Forecast horizon

<table>
<thead>
<tr>
<th>RMSE</th>
<th>$h = 1$</th>
<th>$h = 2$</th>
<th>$h = 3$</th>
<th>$h = 4$</th>
<th>$h = 5$</th>
<th>$h = 6$</th>
<th>Ave</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>1762.04</td>
<td>1770.29</td>
<td>1766.02</td>
<td>1818.82</td>
<td>1705.35</td>
<td>1721.17</td>
<td>1757.28</td>
</tr>
<tr>
<td>Bottom</td>
<td>1736.92</td>
<td>1742.69</td>
<td>1722.79</td>
<td>1752.74</td>
<td>1666.73</td>
<td>1687.43</td>
<td>1718.22</td>
</tr>
<tr>
<td>OLS</td>
<td>1747.60</td>
<td>1757.68</td>
<td>1751.77</td>
<td>1800.67</td>
<td>1686.00</td>
<td>1706.45</td>
<td>1741.69</td>
</tr>
<tr>
<td>WLS</td>
<td>1705.21</td>
<td>1715.87</td>
<td><strong>1703.75</strong></td>
<td>1729.56</td>
<td>1627.79</td>
<td><strong>1661.24</strong></td>
<td><strong>1690.57</strong></td>
</tr>
<tr>
<td>GLS</td>
<td><strong>1704.64</strong></td>
<td><strong>1715.60</strong></td>
<td>1705.31</td>
<td><strong>1729.04</strong></td>
<td>1626.36</td>
<td>1661.64</td>
<td><strong>1690.43</strong></td>
</tr>
<tr>
<td><strong>States</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>399.77</td>
<td>404.16</td>
<td>401.92</td>
<td>407.26</td>
<td>395.38</td>
<td>401.17</td>
<td><strong>401.61</strong></td>
</tr>
<tr>
<td>Bottom</td>
<td>404.29</td>
<td>406.95</td>
<td>404.96</td>
<td>409.02</td>
<td>399.80</td>
<td>401.55</td>
<td><strong>404.43</strong></td>
</tr>
<tr>
<td>OLS</td>
<td>404.47</td>
<td>407.62</td>
<td>405.43</td>
<td>413.79</td>
<td>401.10</td>
<td>404.90</td>
<td><strong>406.22</strong></td>
</tr>
<tr>
<td>WLS</td>
<td><strong>398.84</strong></td>
<td><strong>402.12</strong></td>
<td><strong>400.71</strong></td>
<td><strong>405.03</strong></td>
<td>394.76</td>
<td>398.23</td>
<td><strong>399.95</strong></td>
</tr>
<tr>
<td>GLS</td>
<td><strong>398.84</strong></td>
<td><strong>402.16</strong></td>
<td><strong>400.86</strong></td>
<td><strong>405.03</strong></td>
<td><strong>394.59</strong></td>
<td><strong>398.22</strong></td>
<td><strong>399.95</strong></td>
</tr>
<tr>
<td><strong>Regions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>93.15</td>
<td>93.38</td>
<td>93.45</td>
<td>93.79</td>
<td>93.50</td>
<td>93.56</td>
<td><strong>93.47</strong></td>
</tr>
<tr>
<td>Bottom</td>
<td>93.15</td>
<td>93.38</td>
<td>93.45</td>
<td>93.79</td>
<td>93.50</td>
<td>93.56</td>
<td><strong>93.47</strong></td>
</tr>
<tr>
<td>OLS</td>
<td>93.28</td>
<td>93.53</td>
<td>93.64</td>
<td>94.17</td>
<td>93.78</td>
<td>93.88</td>
<td><strong>93.71</strong></td>
</tr>
<tr>
<td>WLS</td>
<td>93.02</td>
<td>93.32</td>
<td>93.38</td>
<td>93.72</td>
<td>93.39</td>
<td>93.53</td>
<td><strong>93.39</strong></td>
</tr>
<tr>
<td>GLS</td>
<td><strong>92.98</strong></td>
<td><strong>93.27</strong></td>
<td><strong>93.34</strong></td>
<td><strong>93.66</strong></td>
<td><strong>93.34</strong></td>
<td><strong>93.46</strong></td>
<td><strong>93.34</strong></td>
</tr>
</tbody>
</table>
Hierarchical time series

- $y_t$: observed aggregate of all series at time $t$.
- $y_{X,t}$: observation on series $X$ at time $t$.
- $b_t$: vector of all series at bottom level in time $t$. 

```
+------------------+
|  Total           |
+------------------+
|     A             |
|                B  |
|                C  |
```
Hierarchical time series

\[ y_t : \text{observed aggregate of all series at time } t. \]

\[ y_{X, t} : \text{observation on series } X \text{ at time } t. \]

\[ b_t : \text{vector of all series at bottom level in time } t. \]
Hierarchical time series

\[ y_t = [y_t, y_{A,t}, y_{B,t}, y_{C,t}]' = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix} \]

\( y_t \) : observed aggregate of all series at time \( t \).
\( y_{X,t} \) : observation on series \( X \) at time \( t \).
\( b_t \) : vector of all series at bottom level in time \( t \).
Hierarchical time series

\[ y_t = [y_t, y_{A,t}, y_{B,t}, y_{C,t}]' = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix} \]

- \( y_t \): observed aggregate of all series at time \( t \).
- \( y_{X,t} \): observation on series \( X \) at time \( t \).
- \( b_t \): vector of all series at bottom level in time \( t \).
Hierarchical time series

\[ y_t = [y_t, y_{A,t}, y_{B,t}, y_{C,t}]' = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix} \]

- \( y_t \): observed aggregate of all series at time \( t \).
- \( y_{X,t} \): observation on series \( X \) at time \( t \).
- \( b_t \): vector of all series at bottom level in time \( t \).
Hierarchical time series

\[ y_t := \text{observed aggregate of all series at time } t. \]
\[ y_{X,t} := \text{observation on series } X \text{ at time } t. \]
\[ b_t := \text{vector of all series at bottom level in time } t. \]

\[
y_t = [y_t, y_{A,t}, y_{B,t}, y_{C,t}]' = \begin{pmatrix}
1 & 1 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
y_{A,t} \\
y_{B,t} \\
y_{C,t}
\end{pmatrix} = S b_t
\]
Hierarchical time series

\[ y_t = \begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{pmatrix} + b_t \]
Hierarchical time series

\[ y_t = \begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} b_t \\ s \end{pmatrix} \]
Hierarchical time series

\[ y_t = S b_t \]
Grouped data

\[
y_t = \begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}
\]

Forecasting: principles and practice
Hierarchical and grouped time series

21
Grouped data

\[
\begin{pmatrix}
y_t \\
y_{A,t} \\
y_{B,t} \\
y_{X,t} \\
y_{Y,t} \\
y_{AX,t} \\
y_{AY,t} \\
y_{BX,t} \\
y_{BY,t}
\end{pmatrix} =
\begin{pmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

\[
\begin{pmatrix}
y_{AX,t} \\
y_{AY,t} \\
y_{BX,t} \\
y_{BY,t}
\end{pmatrix} = \mathbf{S} \mathbf{b}_t
\]
Grouped data

\[ y_t = S b_t \]

where

\[ y_t = \begin{pmatrix} y_{t} \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} \]

and

\[ b_t = \begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} \]

\[ S = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \]
Hierarchical and grouped time series

Every collection of time series with aggregation constraints can be written as

\[ y_t = Sb_t \]

where

- \( y_t \) is a vector of all series at time \( t \)
- \( b_t \) is a vector of the most disaggregated series at time \( t \)
- \( S \) is a “summing matrix” containing the aggregation constraints.
Let $\hat{y}_n(h)$ be vector of initial $h$-step forecasts, made at time $n$, stacked in same order as $y_t$. 
(In general, they will not “add up”.)

Reconciled forecasts must be of the form:

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

for some matrix $P$. $P$ extracts and combines base forecasts $\hat{y}_n(h)$ to get bottom-level forecasts.
Forecasting notation

Let \( \hat{y}_n(h) \) be vector of initial \( h \)-step forecasts, made at time \( n \), stacked in same order as \( y_t \). (In general, they will not “add up”.)

Reconciled forecasts must be of the form:

\[
\tilde{y}_n(h) = SP\hat{y}_n(h)
\]

for some matrix \( P \).

- \( P \) extracts and combines base forecasts \( \hat{y}_n(h) \) to get bottom-level forecasts.
- \( S \) adds them up.
Let $\hat{y}_n(h)$ be vector of initial $h$-step forecasts, made at time $n$, stacked in same order as $y_t$. (In general, they will not “add up”.)

Reconciled forecasts must be of the form:

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

for some matrix $P$.
Let \( \hat{y}_n(h) \) be vector of initial \( h \)-step forecasts, made at time \( n \), stacked in same order as \( y_t \).
(In general, they will not “add up”.)

Reconciled forecasts must be of the form:
\[
\tilde{y}_n(h) = SP\hat{y}_n(h)
\]
for some matrix \( P \).

- \( P \) extracts and combines base forecasts \( \hat{y}_n(h) \) to get bottom-level forecasts.
- \( S \) adds them up.
Let $\hat{y}_n(h)$ be vector of initial $h$-step forecasts, made at time $n$, stacked in same order as $y_t$. (In general, they will not “add up”.)

Reconciled forecasts must be of the form:

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

for some matrix $P$.

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

Main result
The best (minimum sum of variances) unbiased forecasts are obtained when $P = (S'\Sigma_h^{-1}S)^{-1}S'\Sigma_h^{-1}$, where $\Sigma_h$ is the $h$-step base forecast error covariance matrix.

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

Reconciled forecasts

Problem: $\Sigma_h$ hard to estimate, especially for $h > 1$.
Solutions:
- Ignore $\Sigma_h$ (OLS)
- Assume $\Sigma_h$ diagonal (WLS) [Default in hts]
- Try to estimate $\Sigma_h$ (GLS)
Main result
The best (minimum sum of variances) unbiased forecasts are obtained when
$$P = (S'\Sigma_h^{-1}S)^{-1}S'\Sigma_h^{-1}$$,
where $\Sigma_h$ is the $h$-step base forecast error covariance matrix.

Reconciled forecasts

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

Base forecasts

Problem: $\Sigma_h$ hard to estimate, especially for $h > 1$.
Solutions:
- Ignore $\Sigma_h$ (OLS)
- Assume $\Sigma_h$ diagonal (WLS) [Default in hts]
- Try to estimate $\Sigma_h$ (GLS)
Optimal combination forecasts

Main result
The best (minimum sum of variances) unbiased forecasts are obtained when $P = (S'^{-1}S)^{-1}S'^{-1}$, where $\Sigma_h$ is the $h$-step base forecast error covariance matrix.

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

Reconciled forecasts
Base forecasts

Problem: $\Sigma_h$ hard to estimate, especially for $h > 1$.

Solutions:
- Ignore $\Sigma_h$ (OLS)
- Assume $\Sigma_h$ diagonal (WLS) [Default in hts]
- Try to estimate $\Sigma_h$ (GLS)
1 Hierarchical and grouped time series

2 Lab session 15

3 Temporal hierarchies

4 Lab session 16
1 Hierarchical and grouped time series
2 Lab session 15
3 Temporal hierarchies
4 Lab session 16
Temporal hierarchies

Basic idea:

- Forecast series at each available frequency.
- Optimally reconcile forecasts within the same year.
Basic idea:

- Forecast series at each available frequency.
- Optimally reconcile forecasts within the same year.
**Monthly series**

![Diagram of temporal hierarchies]

- $k = 2, 4, 12$ nodes
- $k = 3, 6, 12$ nodes
- Why not $k = 2, 3, 4, 6, 12$ nodes?
Monthly series

- $k = 2, 4, 12$ nodes
- $k = 3, 6, 12$ nodes
- Why not $k = 2, 3, 4, 6, 12$ nodes?
Monthly series

Why not $k = 2, 3, 4, 6, 12$ nodes?
Monthly data

\[
\begin{pmatrix}
A \\
SemiA_1 \\
SemiA_2 \\
FourM_1 \\
FourM_2 \\
FourM_3 \\
Q_1 \\
\vdots \\
Q_4 \\
BiM_1 \\
\vdots \\
BiM_6 \\
M_1 \\
\vdots \\
M_{12}
\end{pmatrix}
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
M_1 \\
M_2 \\
M_3 \\
M_4 \\
M_5 \\
M_6 \\
M_7 \\
M_8 \\
M_9 \\
M_{10} \\
M_{11} \\
M_{12} \\
B_t
\end{pmatrix}
\]

\[S = I_{12}\]
In general

For a time series \( y_1, \ldots, y_T \), observed at frequency \( m \), we generate aggregate series

\[
y_j^{[k]} = \sum_{t=1+(j-1)k}^{jk} y_t, \quad \text{for } j = 1, \ldots, \left\lfloor \frac{T}{k} \right\rfloor
\]

- \( k \in F(m) = \{ \text{factors of } m \} \).
- A single unique hierarchy is only possible when there are no coprime pairs in \( F(m) \).
- \( M_k = m/k \) is seasonal period of aggregated series.
In general

For a time series $y_1, \ldots, y_T$, observed at frequency $m$, we generate aggregate series

$$y_j^{[k]} = \sum_{t=1+(j-1)k}^{jk} y_t, \quad \text{for } j = 1, \ldots, \lfloor T/k \rfloor$$

- $k \in F(m) = \{\text{factors of } m\}$.
- A single unique hierarchy is only possible when there are no coprime pairs in $F(m)$.
- $M_k = m/k$ is seasonal period of aggregated series.

Forecasting: principles and practice  

Temporal hierarchies  

31
In general

For a time series \( y_1, \ldots, y_T \), observed at frequency \( m \), we generate aggregate series

\[
\begin{align*}
y_j^{[k]} &= \sum_{t=1+(j-1)k}^{jk} y_t, & \text{for } j = 1, \ldots, \lfloor T/k \rfloor
\end{align*}
\]

- \( k \in F(m) = \{ \text{factors of } m \} \).
- A single unique hierarchy is only possible when there are no coprime pairs in \( F(m) \).
- \( M_k = m/k \) is seasonal period of aggregated series.
UK Accidents and Emergency Demand

Temporal hierarchies

Annual (k=52)

Semi-annual (k=26)

Quarterly (k=13)

Monthly (k=4)

Bi-weekly (k=2)

Weekly (k=1)

Forecast

Forecast

Forecast

Forecast

Forecast

Forecast

- - - - base

- - - - reconciled
UK Accidents and Emergency Demand

1. Type 1 Departments — Major A&E
2. Type 2 Departments — Single Specialty
3. Type 3 Departments — Other A&E/Minor Injury
4. Total Attendances
5. Type 1 Departments — Major A&E > 4 hrs
6. Type 2 Departments — Single Specialty > 4 hrs
7. Type 3 Departments — Other A&E/Minor Injury > 4 hrs
8. Total Attendances > 4 hrs
9. Emergency Admissions via Type 1 A&E
10. Total Emergency Admissions via A&E
11. Other Emergency Admissions (i.e., not via A&E)
12. Total Emergency Admissions
13. Number of patients spending > 4 hrs from decision to admission
UK Accidents and Emergency Demand

- **Minimum training set**: all data except the last year
- Base forecasts using `auto.arima()`.
- Mean Absolute Scaled Errors for 1, 4 and 13 weeks ahead using a rolling origin.

<table>
<thead>
<tr>
<th>Aggr. Level</th>
<th>$h$</th>
<th>Base</th>
<th>Reconciled</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>$-17.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>4</td>
<td>1.9</td>
<td>1.5</td>
<td>$-18.6%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>13</td>
<td>2.3</td>
<td>1.9</td>
<td>$-16.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>1–52</td>
<td>2.0</td>
<td>1.9</td>
<td>$-5.0%$</td>
</tr>
<tr>
<td>Annual</td>
<td>1</td>
<td>3.4</td>
<td>1.9</td>
<td>$-42.9%$</td>
</tr>
</tbody>
</table>
**UK Accidents and Emergency Demand**

- **Minimum training set**: all data except the last year
- **Base forecasts using auto.arima()**.
- Mean Absolute Scaled Errors for 1, 4 and 13 weeks ahead using a rolling origin.

<table>
<thead>
<tr>
<th>Aggr. Level</th>
<th>$h$</th>
<th>Base</th>
<th>Reconciled</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>−17.2%</td>
</tr>
<tr>
<td>Weekly</td>
<td>4</td>
<td>1.9</td>
<td>1.5</td>
<td>−18.6%</td>
</tr>
<tr>
<td>Weekly</td>
<td>13</td>
<td>2.3</td>
<td>1.9</td>
<td>−16.2%</td>
</tr>
<tr>
<td>Weekly</td>
<td>1–52</td>
<td>2.0</td>
<td>1.9</td>
<td>−5.0%</td>
</tr>
<tr>
<td>Annual</td>
<td>1</td>
<td>3.4</td>
<td>1.9</td>
<td>−42.9%</td>
</tr>
</tbody>
</table>
**UK Accidents and Emergency Demand**

- **Minimum training set:** all data except the last year
- **Base forecasts using** `auto.arima()`.
- **Mean Absolute Scaled Errors for 1, 4 and 13 weeks ahead using a rolling origin.**

<table>
<thead>
<tr>
<th>Aggr. Level</th>
<th>$h$</th>
<th>Base</th>
<th>Reconciled</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>$-17.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>4</td>
<td>1.9</td>
<td>1.5</td>
<td>$-18.6%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>13</td>
<td>2.3</td>
<td>1.9</td>
<td>$-16.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>1–52</td>
<td>2.0</td>
<td>1.9</td>
<td>$-5.0%$</td>
</tr>
<tr>
<td>Annual</td>
<td>1</td>
<td>3.4</td>
<td>1.9</td>
<td>$-42.9%$</td>
</tr>
</tbody>
</table>
UK Accidents and Emergency Demand

- **Minimum training set**: all data except the last year
- Base forecasts using `auto.arima()`.
- Mean Absolute Scaled Errors for 1, 4 and 13 weeks ahead using a rolling origin.

<table>
<thead>
<tr>
<th>Aggr. Level</th>
<th>$h$</th>
<th>Base</th>
<th>Reconciled</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>$-17.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>4</td>
<td>1.9</td>
<td>1.5</td>
<td>$-18.6%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>13</td>
<td>2.3</td>
<td>1.9</td>
<td>$-16.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>1–52</td>
<td>2.0</td>
<td>1.9</td>
<td>$-5.0%$</td>
</tr>
<tr>
<td>Annual</td>
<td>1</td>
<td>3.4</td>
<td>1.9</td>
<td>$-42.9%$</td>
</tr>
</tbody>
</table>
UK Accidents and Emergency Demand

- **Minimum training set:** all data except the last year
- Base forecasts using `auto.arima()`.
- Mean Absolute Scaled Errors for 1, 4 and 13 weeks ahead using a rolling origin.

<table>
<thead>
<tr>
<th>Aggr. Level</th>
<th>$h$</th>
<th>Base</th>
<th>Reconciled</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>$-17.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>4</td>
<td>1.9</td>
<td>1.5</td>
<td>$-18.6%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>13</td>
<td>2.3</td>
<td>1.9</td>
<td>$-16.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>1–52</td>
<td>2.0</td>
<td>1.9</td>
<td>$-5.0%$</td>
</tr>
<tr>
<td>Annual</td>
<td>1</td>
<td>3.4</td>
<td>1.9</td>
<td>$-42.9%$</td>
</tr>
</tbody>
</table>
thief: Temporal Hierarchical Forecasting

Install from CRAN
install.packages("thief")

Usage
library(thief)
thief(y)
thief: Temporal HIErarchical Forecasting

Install from CRAN
install.packages("thief")

Usage
library(thief)
thief(y)
1. Hierarchical and grouped time series
2. Lab session 15
3. Temporal hierarchies
4. Lab session 16
Lab Session 16