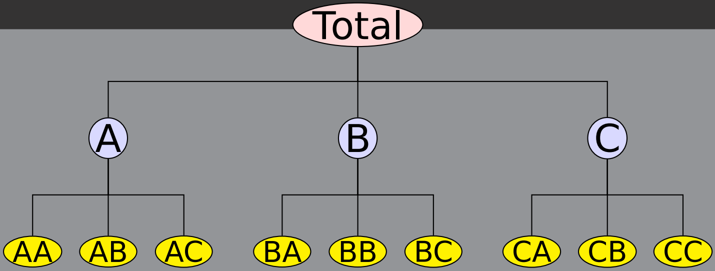
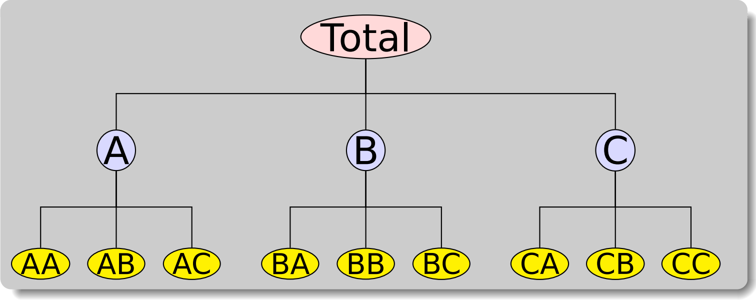
MONASH University

Rob J Hyndman

tools for hierarchical time series



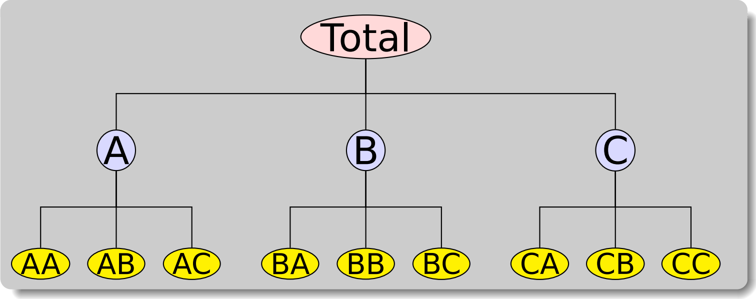
Introduction



Examples

- Manufacturing product hierarchies
- Pharmaceutical sales
- Net labour turnover

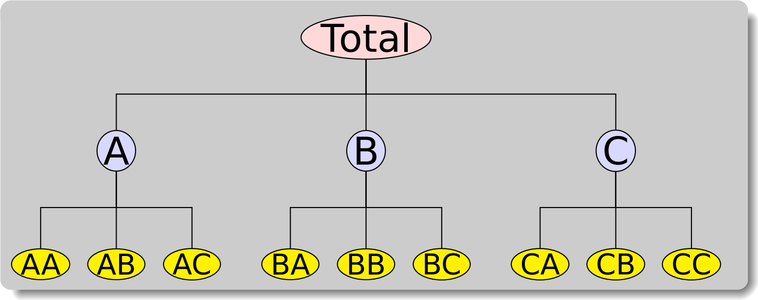
Introduction



Examples

- Manufacturing product hierarchies
- Pharmaceutical sales
- Net labour turnover

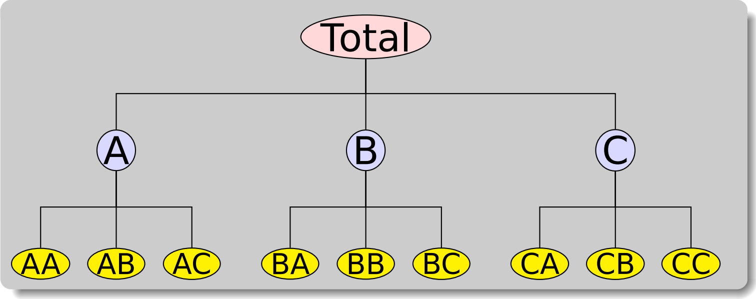
Introduction



Examples

- Manufacturing product hierarchies
- Pharmaceutical sales
- Net labour turnover

Introduction



Examples

- Manufacturing product hierarchies
- Pharmaceutical sales
- Net labour turnover

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: daily numbers of calls to HP call centres are grouped by product type and location of call centre.

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: daily numbers of calls to HP call centres are grouped by product type and location of call centre.

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: daily numbers of calls to HP call centres are grouped by product type and location of call centre.

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: daily numbers of calls to HP call centres are grouped by product type and location of call centre.

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.

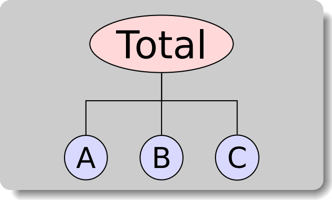
Hierarchical/grouped time series

- 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.

Hierarchical data

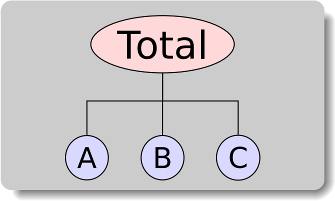


Y_t : observed aggregate of all series at time t.

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

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

Hierarchical data

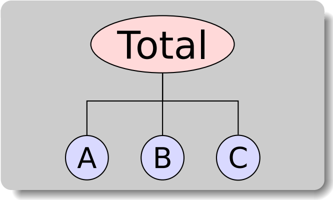


Y_t : observed aggregate of all series at time t.

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

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

Hierarchical data



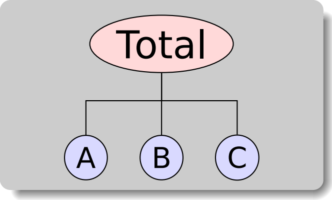
Y_t : observed aggregate of all series at time t.

$Y_{X,t}$: 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



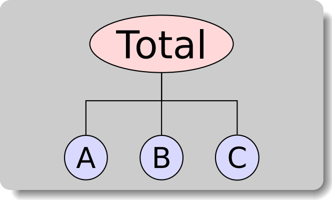
Y_t : observed aggregate of all series at time t.

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

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

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

Hierarchical data



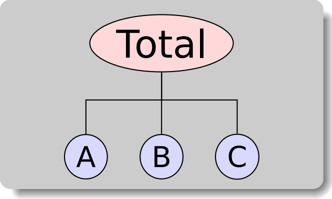
Y_t : observed aggregate of all series at time t.

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

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

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

Hierarchical data



Y_t : observed aggregate of all series at time t.

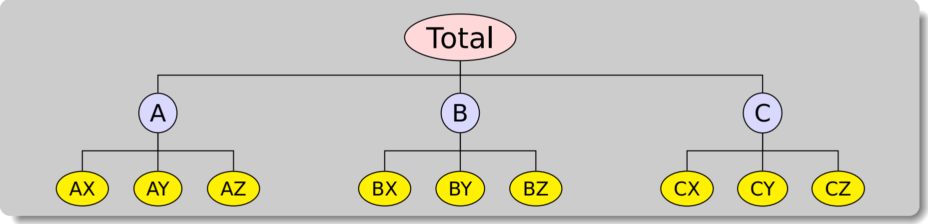
$Y_{X,t}$: observation on series X at time t.

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

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

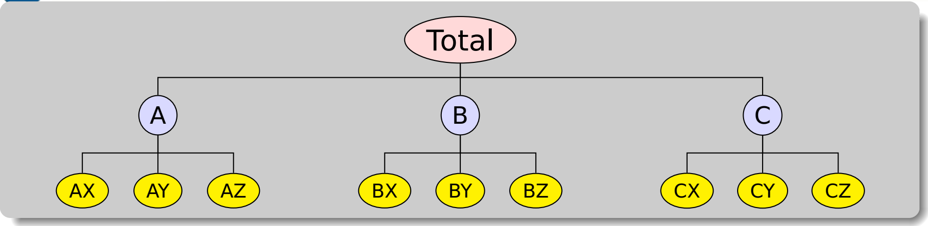
$$\boxed{\boldsymbol{Y}_t = S\boldsymbol{B}_t}$$

Hierarchical data



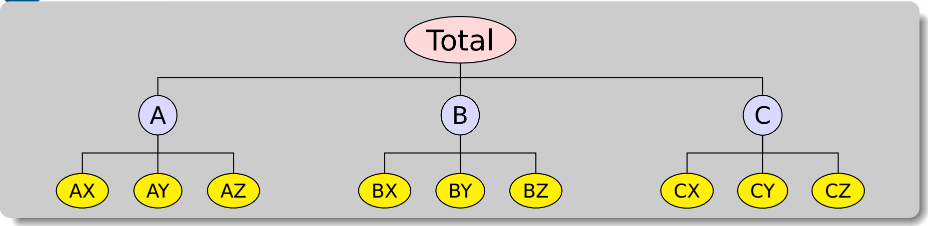
$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{BZ,t} \\ Y_{CX,t} \\ Y_{CY,t} \\ Y_{CZ,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0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

Hierarchical data



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

Hierarchical data



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

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

Grouped data

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

Grouped data

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

Grouped data

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

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

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 $\boldsymbol{Y}_t$.

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

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

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 $\boldsymbol{Y}_t$.

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

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 $\boldsymbol{Y}_t$.

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

Optimal combination forecasts

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

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

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

Optimal combination forecasts

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

Initial forecasts

Optimal combination forecasts

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

Revised forecasts

Initial forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- **Problem:** Σ_h hard to estimate.
- **Solution:** Assume $\epsilon_h \approx S\epsilon_{K,h}$ where $\epsilon_{K,h}$ is the forecast error at bottom level.
  Then $\Sigma_h \approx S\Omega_h S'$ where $\Omega_h = \text{Var}(\epsilon_{K,h})$.
  If Moore-Penrose generalized inverse used, then $(S'\Sigma^{\dagger}S)^{-1}S'\Sigma^{\dagger} = (S'S)^{-1}S'$.

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

Optimal combination forecasts

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

Revised forecasts

Initial forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- **Problem:** Σ_h hard to estimate.
- **Solution:** Assume $\varepsilon_h \approx S\varepsilon_{K,h}$ where $\varepsilon_{K,h}$ is the forecast error at bottom level.

Then $\Sigma_h \approx S\Omega_h S'$ where $\Omega_h = \text{Var}(\varepsilon_{K,h})$.

If Moore-Penrose generalized inverse used, then $(S'\Sigma^{\dagger}S)^{-1}S'\Sigma^{\dagger} = (S'S)^{-1}S'$.

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

Optimal combination forecasts

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

Revised forecasts

Initial forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- **Problem:** Σ_h hard to estimate.
- **Solution:** Assume $\varepsilon_h \approx S\varepsilon_{K,h}$ where $\varepsilon_{K,h}$ is the forecast error at bottom level.

  Then $\Sigma_h \approx S\Omega_h S'$ where $\Omega_h = \mathrm{Var}(\varepsilon_{K,h})$.

  If Moore-Penrose generalized inverse used, then $(S'\Sigma_h^{\dagger}S)^{-1}S'\Sigma_h^{\dagger} = (S'S)^{-1}S'$.

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

Optimal combination forecasts

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

Revised forecasts   Initial forecasts

- $\Sigma_h^\dagger$ is generalized inverse of Σ_h.
- **Problem:** Σ_h hard to estimate.
- **Solution:** Assume $\varepsilon_h \approx S\varepsilon_{K,h}$ where $\varepsilon_{K,h}$ is the forecast error at bottom level.

  Then $\Sigma_h \approx S\Omega_h S'$ where $\Omega_h = \mathrm{Var}(\varepsilon_{K,h})$.

  If Moore-Penrose generalized inverse used, then $(S'\Sigma^\dagger S)^{-1}S'\Sigma^\dagger = (S'S)^{-1}S'$.

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

Optimal combination forecasts

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

Revised forecasts — Initial forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- **Problem:** Σ_h hard to estimate.
- **Solution:** Assume $\varepsilon_h \approx S\varepsilon_{K,h}$ where $\varepsilon_{K,h}$ is the forecast error at bottom level.
  Then $\Sigma_h \approx S\Omega_h S'$ where $\Omega_h = \text{Var}(\varepsilon_{K,h})$.
  If Moore-Penrose generalized inverse used, then $(S'\Sigma^{\dagger}S)^{-1}S'\Sigma^{\dagger} = (S'S)^{-1}S'$.

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

Optimal combination forecasts

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

Revised forecasts — Initial forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- **Problem:** Σ_h hard to estimate.
- **Solution:** Assume $\varepsilon_h \approx S\varepsilon_{K,h}$ where $\varepsilon_{K,h}$ is the forecast error at bottom level.
  Then $\Sigma_h \approx S\Omega_h S'$ where $\Omega_h = \text{Var}(\varepsilon_{K,h})$.
  If Moore-Penrose generalized inverse used, then $(S'\Sigma^{\dagger}S)^{-1}S'\Sigma^{\dagger} = (S'S)^{-1}S'$.

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

Optimal combination forecasts

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

Revised forecasts

Initial forecasts

- $\Sigma_h^\dagger$ is generalized inverse of Σ_h.
- **Problem:** Σ_h hard to estimate.
- **Solution:** Assume $\varepsilon_h \approx S\varepsilon_{K,h}$ where $\varepsilon_{K,h}$ is the forecast error at bottom level.
  Then $\Sigma_h \approx S\Omega_h S'$ where $\Omega_h = \mathrm{Var}(\varepsilon_{K,h})$.
  If Moore-Penrose generalized inverse used, then $(S'\Sigma^\dagger S)^{-1}S'\Sigma^\dagger = (S'S)^{-1}S'$.

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

Optimal combination forecasts

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

Revised forecasts Initial forecasts

- $\Sigma_h^\dagger$ is generalized inverse of Σ_h.
- **Problem:** Σ_h hard to estimate.
- **Solution:** Assume $\varepsilon_h \approx S\varepsilon_{K,h}$ where $\varepsilon_{K,h}$ is the forecast error at bottom level.
  Then $\Sigma_h \approx S\Omega_h S'$ where $\Omega_h = \text{Var}(\varepsilon_{K,h})$.
  If Moore-Penrose generalized inverse used, then $(S'\Sigma^\dagger S)^{-1}S'\Sigma^\dagger = (S'S)^{-1}S'$.

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

Optimal combination forecasts

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

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

Optimal combination forecasts

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

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

Optimal combination forecasts

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

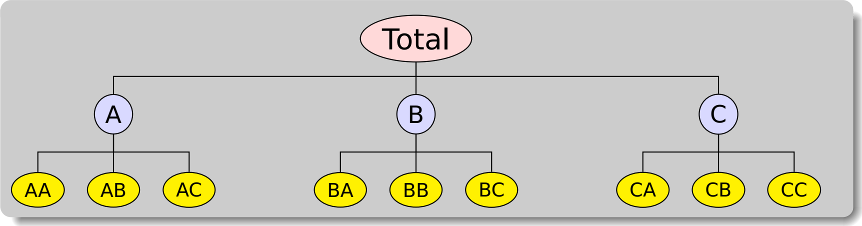
- GLS = OLS.
- Optimal weighted average of initial forecasts.
- Optimal reconciliation weights are $S(S'S)^{-1}S'$.
- Weights are independent of the data and of the covariance structure of the hierarchy!

Optimal combination forecasts

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

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

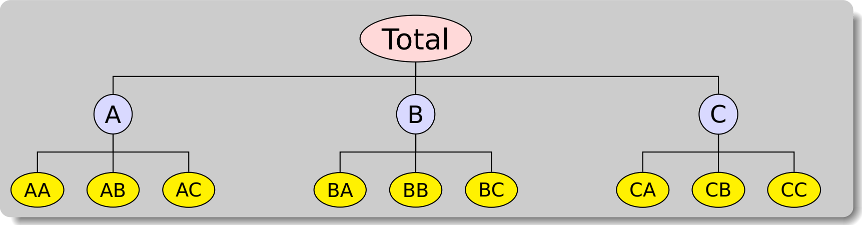
Optimal combination forecasts



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

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

Optimal combination forecasts



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

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

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

- Computational difficulties in big hierarchies due to size of the S matrix and non-singular behavior of $(S'S)$.
- Need to estimate covariance matrix to produce prediction intervals.

Challenges

- Computational difficulties in big hierarchies due to size of the S matrix and non-singular behavior of $(S'S)$.
- Need to estimate covariance matrix to produce prediction intervals.

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 | GPL-3 [expanded from: GPL ($\geq$ 2)]

Example using R

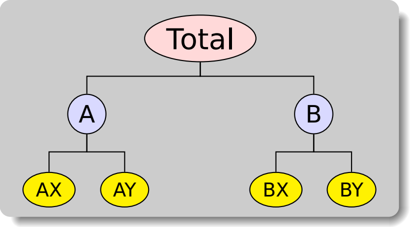
```
library(hts)

# 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

```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 optimal 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 currently available in the **hts** package, for forecasting hierarchical time series. The methods included are: top-down, buttom-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 related time series can be organized in a hierarchical structure based on dimensions such as gender, geography or product type. This has led to the problem of hierarchical time series

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.com/fpp/`.