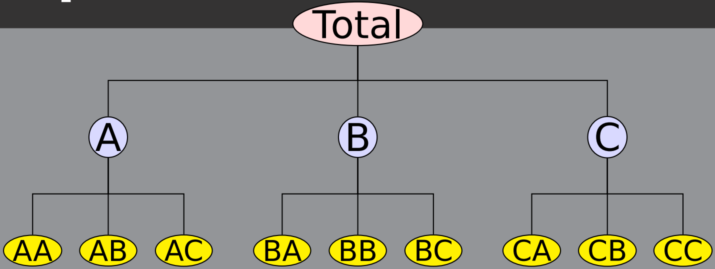
George Athanasopoulos
(with Rob J. Hyndman, Nikos Kourentzes and Fotis Petropoulos)

# Forecasting hierarchical (and grouped) time series



# Outline

# Summary

3 key developments:

1. We generalise the forecasting process in "properly" accounting for grouped data. (Empirical Application 1).
2. We advance the “optimal combination” approach by proposing two new estimators based on WLS.
   Both now implemented in the hts package.
3. We introduce temporal hierarchies. (Empirical Application 2).

# Summary

3 key developments:

1. We generalise the forecasting process in "properly" accounting for grouped data. (Empirical Application 1).
2. We advance the “optimal combination” approach by proposing two new estimators based on WLS.

   Both now implemented in the hts package.
3. We introduce temporal hierarchies. (Empirical Application 2).

# Summary

3 key developments:

1. We generalise the forecasting process in "properly" accounting for grouped data. (Empirical Application 1).
2. We advance the “optimal combination” approach by proposing two new estimators based on WLS.
   Both now implemented in the hts package.
3. We introduce temporal hierarchies. (Empirical Application 2).

# Summary

3 key developments:

1. We generalise the forecasting process in "properly" accounting for grouped data. (Empirical Application 1).
2. We advance the “optimal combination” approach by proposing two new estimators based on WLS.
   Both now implemented in the hts package.
3. We introduce temporal hierarchies. (Empirical Application 2).

# Hierarchical versus grouped

**Table:** Geographical Hierarchy

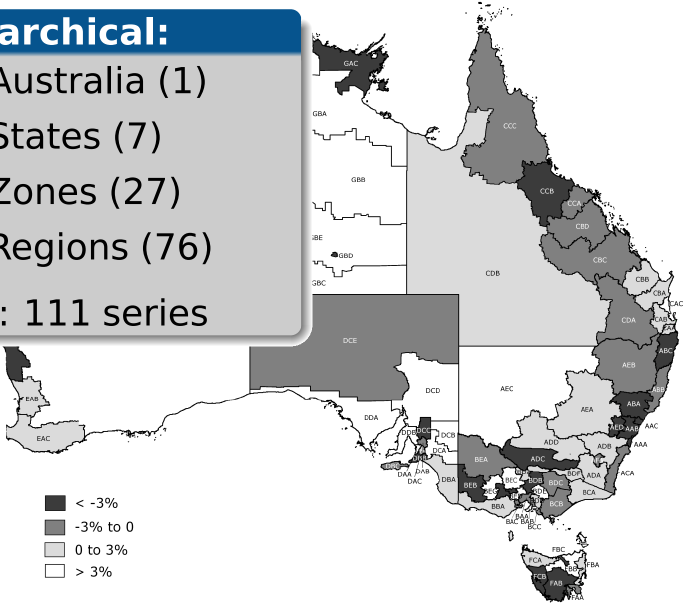
| Level | Total series per level |
| --- | --- |
| Australia | 1 |
| States and Territories | 7 |
| VIC, NSW, QLD, SA, WA, NT, TAS; | |
| Zones | 27 |
| VIC (5): Metro, West Coast, East Coast, Nth East, Nth West; | |
| NSW (6): Metro, Nth Coast, Sth Coast, Sth, Nth, ACT; | |
| QLD (4): Metro, Central Coast, Nth Coast, Inland; | |
| Regions | 76 |
| Metro VIC: Melbourne, Peninsula, Geelong; | |
| Metro NSW: Sydney, Illawarra, Central Coast; | |
| Metro QLD: Brisbane, Gold Coast, Sunshine Coast; | |

# Australian domestic tourism

## Hierarchical:

- Australia (1)
- States (7)
- Zones (27)
- Regions (76)

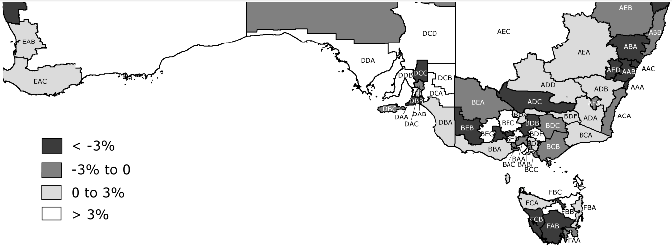
Total: 111 series



# Australian domestic tourism


## Purpose of travel (PoT):

- Holiday
- Visiting Friends and Relatives
- Business
- Other



# Australian domestic tourism

## Hierarchical:

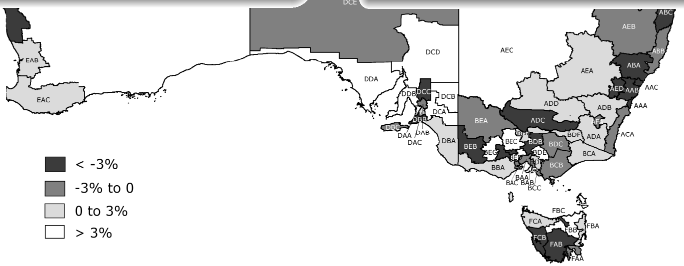
- Australia (1)
- States (7)
- Zones (27)
- Regions (76)

Total: 111 series

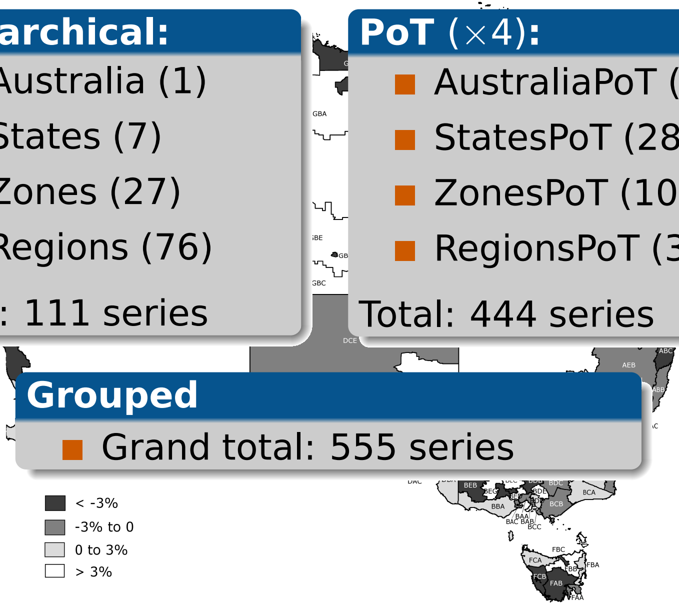
## PoT (×4):

- AustraliaPoT (4)
- StatesPoT (28)
- ZonesPoT (108)
- RegionsPoT (304)

Total: 444 series



# Australian domestic tourism




## Hierarchical:

- Australia (1)
- States (7)
- Zones (27)
- Regions (76)

Total: 111 series

## PoT ($\times 4$):

- AustraliaPoT (4)
- StatesPoT (28)
- ZonesPoT (108)
- RegionsPoT (304)

Total: 444 series

## Grouped

- Grand total: 555 series

# Forecasting such structures

Existing methods:

- Bottom-up
- Top-down
- Middle-out

**Key idea: forecast reconciliation**

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

# Forecasting such structures

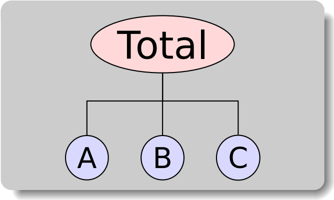
Existing methods:

- Bottom-up
- Top-down
- Middle-out

**Key idea: forecast reconciliation**

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

# Hierarchical data

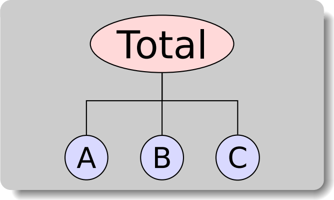



$Y_t$ : observed aggregate of all series at time $t$.

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

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

# Hierarchical data

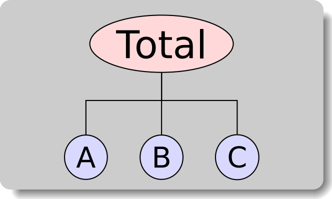



$Y_t$ : observed aggregate of all series at time $t$.

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

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

# Hierarchical data



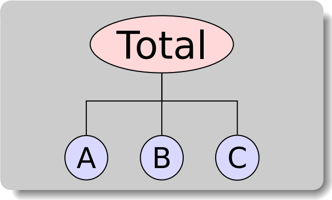

$Y_t$ : observed aggregate of all series at time $t$.

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

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

$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix}$$

# Hierarchical data



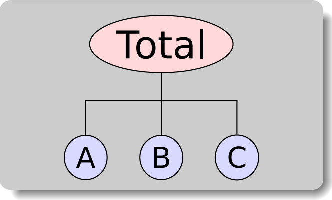

$Y_t$ : observed aggregate of all series at time $t$.

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

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

$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \begin{pmatrix} Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix}$$

# Hierarchical data



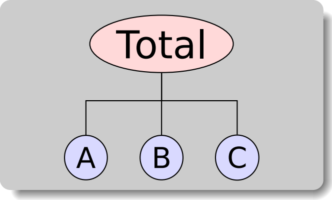

$Y_t$ : observed aggregate of all series at time $t$.

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

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

$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

# Hierarchical data




$Y_t$ : observed aggregate of all series at time $t$.

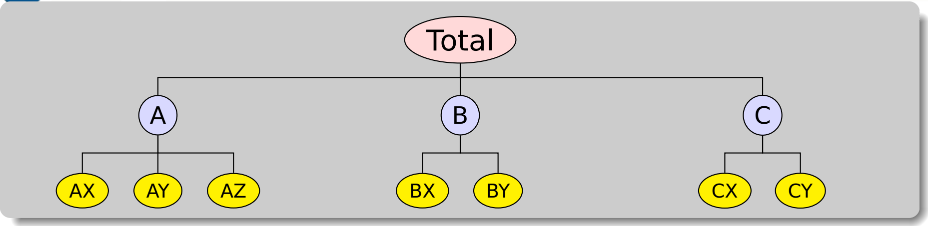
$Y_{X,t}$ : observation on series $X$ at time $t$.

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

$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \end{pmatrix}}_{\boldsymbol{B}_t}$$

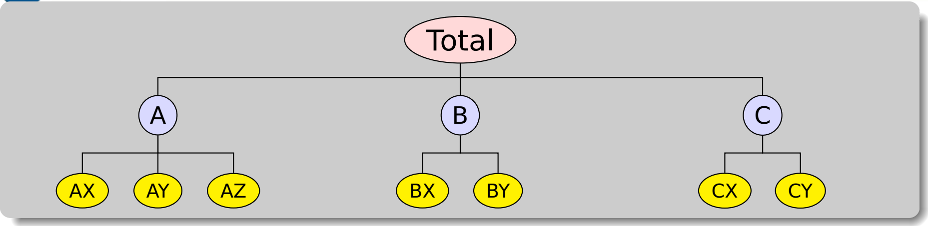
$$\boxed{\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{B}_t}$$

# Hierarchical data



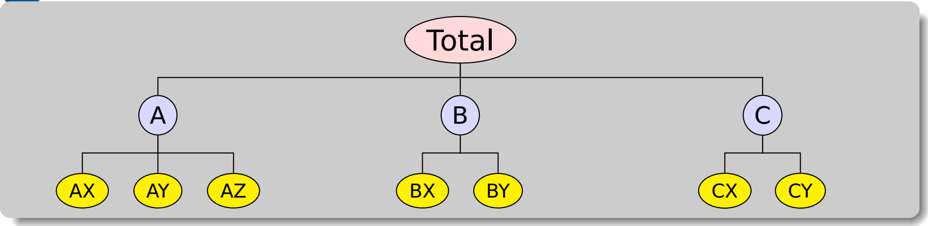

$$\boldsymbol{Y}_t = \begin{pmatrix} Y_t \\ Y_{A,t} \\ Y_{B,t} \\ Y_{C,t} \\ Y_{AX,t} \\ Y_{AY,t} \\ Y_{AZ,t} \\ Y_{BX,t} \\ Y_{BY,t} \\ Y_{CX,t} \\ Y_{CY,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 &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

# Hierarchical data




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

# Hierarchical data




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

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

# 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


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

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

# Outline

# Optimal forecasts

**Key idea: forecast reconciliation**

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

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

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

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

# Optimal forecasts

**Key idea: forecast reconciliation**

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

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

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

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

# Optimal forecasts

**Key idea: forecast reconciliation**

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

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

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

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

# Optimal forecasts

**Key idea: forecast reconciliation**

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

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

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

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

# Optimal forecasts

**Key idea: forecast reconciliation**

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

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

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

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

# Optimal forecasts

**Key idea: forecast reconciliation**

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

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

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

- $\beta_n(h) = \mathrm{E}[\boldsymbol{B}_{n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.
- $\varepsilon_h$ has zero mean and covariance $\Sigma_h$.
- 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 $\boldsymbol{P} = (\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma_h^{\dagger}$.
- $\Sigma_h^{\dagger}$ is generalized inverse of $\Sigma_h$.
- Revised forecasts *unbiased*: $\boldsymbol{SPS} = \boldsymbol{S}$.
- Revised forecasts *minimum variance*:

$$\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'$$

- **Problem:** $\Sigma_h$ hard to estimate.

# Optimal combination forecasts

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

Initial forecasts

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

$$\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'.$$

- **Problem:** $\Sigma_h$ 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

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

$$\mathrm{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'.$$

- **Problem:** $\Sigma_h$ 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

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

$$\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'.$$

- **Problem:** $\Sigma_h$ 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

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

$$\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'.$$

- **Problem:** $\Sigma_h$ 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

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

$$\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}'.$$

- **Problem:** $\Sigma_h$ 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

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

$$\text{Var}[\tilde{\boldsymbol{Y}}_n(h)|\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n] = \boldsymbol{S}(\boldsymbol{S}'\Sigma_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'.$$

- **Problem:** $\Sigma_h$ hard to estimate.

# Outline

# Approx. optimal forecasts

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

## Solution 1: OLS

- Assume $\varepsilon_h \approx \boldsymbol{S}\varepsilon_{B,h}$ where $\varepsilon_{B,h}$ is the forecast error at bottom level.
- If Moore-Penrose generalized inverse used, then $(\boldsymbol{S}'\Sigma^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\Sigma^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$.

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

# Approx. optimal forecasts

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

## Solution 2: Rescaling

- Suppose we approximate $\Sigma_h$ by its diagonal.
- Let $\Lambda = \left[\text{diagonal}\left(\hat{\Sigma}_1\right)\right]^{-1}$ contain inverse one-step ahead in-sample forecast error variances.

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

# Approx. optimal forecasts

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

## Solution 3: Averaging

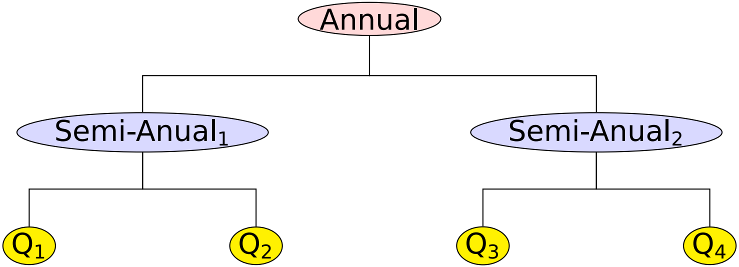
- If the bottom level error series are approximately uncorrelated and have similar variances, then $\Lambda$ is inversely proportional to the number of series contributing to each node.
- So set $\Lambda$ to be the inverse row sums of $\mathbf{S}$:

$$\Lambda = diag(\mathbf{S} \times \mathbf{1})^{-1}$$

where $\mathbf{1} = (1, 1, \ldots, 1)'$.

# Outline

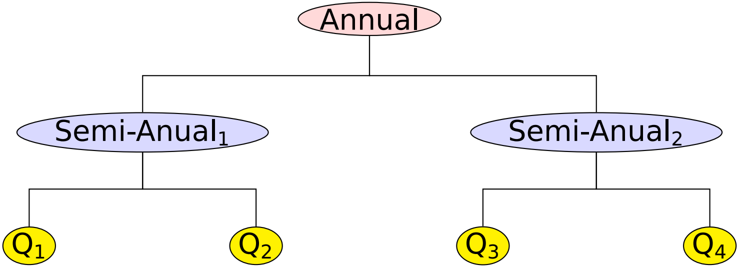
# Temporal hierarchies: quarterly




**Basic idea:**

- Forecast series at each available frequency.
- Optimally combine forecasts within the same year.

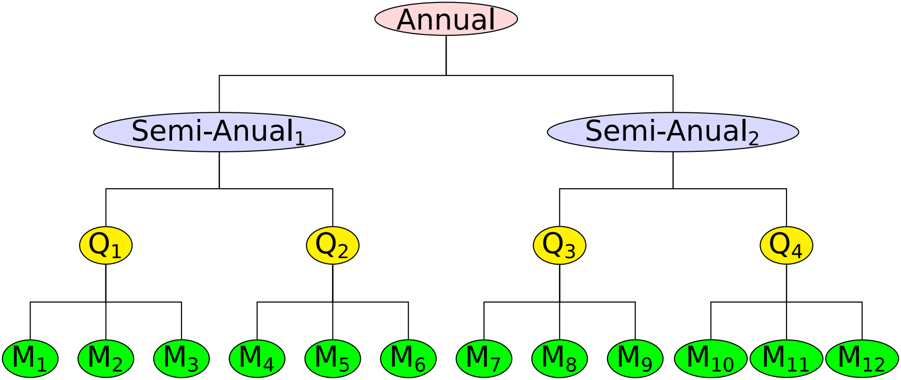
# Temporal hierarchies: quarterly




**Basic idea:**

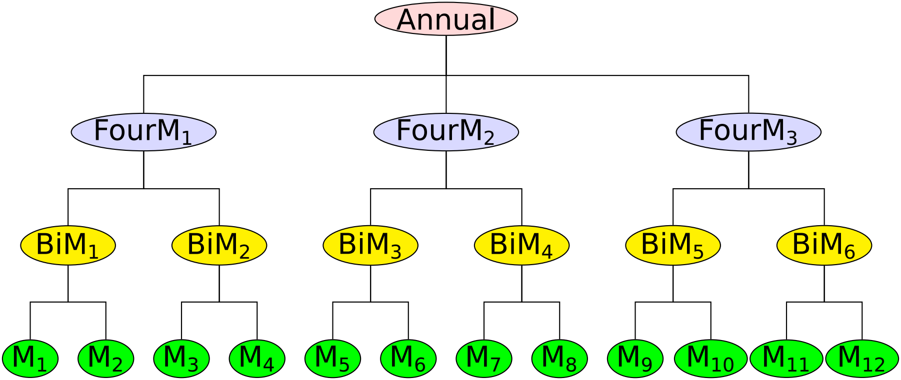
- Forecast series at each available frequency.
- Optimally combine forecasts within the same year.

# Temporal hierarchies: monthly



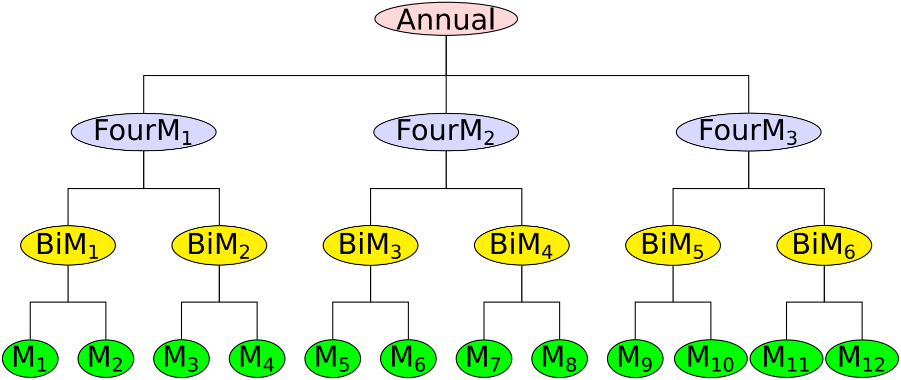

- Aggregate: 3, 6, 12
- Alternatively: 2, 4, 12.
- How about: 2, 3, 4, 6, 12?

# Temporal hierarchies: monthly




- Aggregate: $3, 6, 12$
- Alternatively: $2, 4, 12$.
- How about: $2, 3, 4, 6, 12$?

# Temporal hierarchies: monthly




- Aggregate: $3, 6, 12$
- Alternatively: $2, 4, 12$.
- How about: $2, 3, 4, 6, 12$?

# Monthly data

$$\underbrace{\begin{pmatrix} A \\ SemiA_1 \\ SemiA_2 \\ FourM_1 \\ FourM_2 \\ FourM_3 \\ Q_1 \\ \vdots \\ Q_4 \\ BiM_1 \\ \vdots \\ BiM_6 \\ M_1 \\ \vdots \\ M_{12} \end{pmatrix}}_{(28\times1)} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ & & & & & & & & & & & \\ & & & & & & \boldsymbol{I}_{12} & & & & & \\ & & & & & & & & & & & \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \\ M_5 \\ M_6 \\ M_7 \\ M_8 \\ M_9 \\ M_{10} \\ M_{11} \\ M_{12} \end{pmatrix}}_{\boldsymbol{B}_t}$$

# Experimental setup:

- M3 forecasting competition (Makridakis and Hibon, 2000, *IJF*). In total 3003 series.
- 1,428 monthly series with a test sample of 12 observations each.
- 756 quarterly series with a test sample of 8 observations each.
- Forecast each series with ETS (ARIMA) models. Methods performed well in the actual competition.

# Experimental setup:

- M3 forecasting competition (Makridakis and Hibon, 2000, *IJF*). In total 3003 series.
- 1,428 monthly series with a test sample of 12 observations each.
- 756 quarterly series with a test sample of 8 observations each.
- Forecast each series with ETS (ARIMA) models. Methods performed well in the actual competition.

# Experimental setup:

- M3 forecasting competition (Makridakis and Hibon, 2000, *IJF*). In total 3003 series.
- 1,428 monthly series with a test sample of 12 observations each.
- 756 quarterly series with a test sample of 8 observations each.
- Forecast each series with ETS (ARIMA) models. Methods performed well in the actual competition.

# Experimental setup:

- M3 forecasting competition (Makridakis and Hibon, 2000, *IJF*). In total 3003 series.
- 1,428 monthly series with a test sample of 12 observations each.
- 756 quarterly series with a test sample of 8 observations each.
- Forecast each series with ETS (ARIMA) models. Methods performed well in the actual competition.

# Results: Monthly

| sMAPE *(obs)* | *Forecast Horizon (h)* Annual (1) | SemiA (2) | FourM (3) | Q (4) | BiM (6) | M (12) | Average |
|---|---|---|---|---|---|---|---|
| | | | *ETS* | | | | |
| Initial | 9.66 | 9.18 | 9.76 | 10.14 | 10.82 | 12.85 | 10.40 |
| Bottom-up | 8.38 | 9.14 | 9.78 | 10.06 | 11.04 | 12.85 | 10.21 |
| OLS | 7.80 | 8.64 | 9.39 | 9.72 | 10.68 | 12.68 | 9.82 |
| Scaling | 7.64 | 8.44 | 9.15 | 9.49 | 10.45 | 12.40 | 9.60 |
| Averaging | **7.51** | **8.31** | **9.05** | **9.38** | **10.34** | **12.30** | **9.48** |

# Results: Quarterly

| sMAPE *(obs)* | *Forecast Horizon (h)* Annual *(2)* | Semi-Ann *(4)* | Quart *(8)* | Average |
|---|---|---|---|---|
| | | *ETS* | | |
| Initial | 10.50 | 9.97 | 9.84 | 10.10 |
| Bottom-up | 8.87 | 9.35 | 9.84 | 9.35 |
| OLS | 9.31 | 9.78 | 10.28 | 9.79 |
| Scaling | **8.75** | **9.19** | **9.70** | **9.21** |
| Averaging | 8.81 | 9.26 | 9.78 | 9.28 |

# Outline

# 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

# 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.

G Athanasopoulos, RA Ahmed, RJ Hyndman,(2009). “Hierarchical forecasts for Australian domestic tourism”. *International Journal of Forecasting* 25, 146-166.

RJ Hyndman, et al., (2014). hts: Hierarchical time series. cran.r-project.org/package=hts.

RJ Hyndman and G Athanasopoulos (2014). *Forecasting: principles and practice.* OTexts. `www.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.

- Email: **george.athanasopoulos@monash.edu**

- Acknowledgments: Rob J Hyndman, Roman Ahmed, Han Shang, Shanika Wickramasuriya, Nikolaos Kourentzes, Fotios Petropoulos.
- Acknowledgments: Excellent research assistance by Earo Wang.