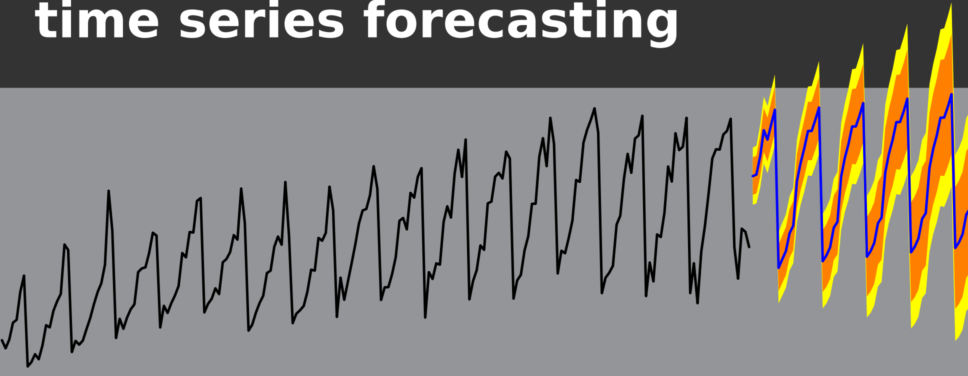
MONASH University

Rob J Hyndman

Advances in automatic time series forecasting



Outline

Motivation

Motivation

Australian Government

Department of Health and Ageing

Motivation

FOXTEL digital

Australian Government

Department of Health and Ageing

Motivation

FOXTEL digital

Incitec Pivot

Australian Government

Department of Health and Ageing

Motivation

FOXTEL digital

hp

Incitec Pivot

AUSTRALIA

Australian Government

Department of Health and Ageing

Motivation

Motivation

1. Common in business to have over 1000 products that need forecasting at least monthly.
2. Forecasts are often required by people who are untrained in time series analysis.
3. Some types of data can be decomposed into a large number of univariate time series that need to be forecast.

Specifications

Automatic forecasting algorithms must:

- determine an appropriate time series model;
- estimate the parameters;
- compute the forecasts with prediction intervals.

Motivation

1. Common in business to have over 1000 products that need forecasting at least monthly.
2. Forecasts are often required by people who are untrained in time series analysis.
3. Some types of data can be decomposed into a large number of univariate time series that need to be forecast.

Specifications

Automatic forecasting algorithms must:

- determine an appropriate time series model;
- estimate the parameters;
- compute the forecasts with prediction intervals.

Motivation

1. Common in business to have over 1000 products that need forecasting at least monthly.
2. Forecasts are often required by people who are untrained in time series analysis.
3. Some types of data can be decomposed into a large number of univariate time series that need to be forecast.

Specifications

Automatic forecasting algorithms must:

- determine an appropriate time series model;
- estimate the parameters;
- compute the forecasts with prediction intervals.

Motivation

1. Common in business to have over 1000 products that need forecasting at least monthly.
2. Forecasts are often required by people who are untrained in time series analysis.
3. Some types of data can be decomposed into a large number of univariate time series that need to be forecast.

Specifications

Automatic forecasting algorithms must:

- determine an appropriate time series model;
- estimate the parameters;
- compute the forecasts with prediction intervals.

Motivation

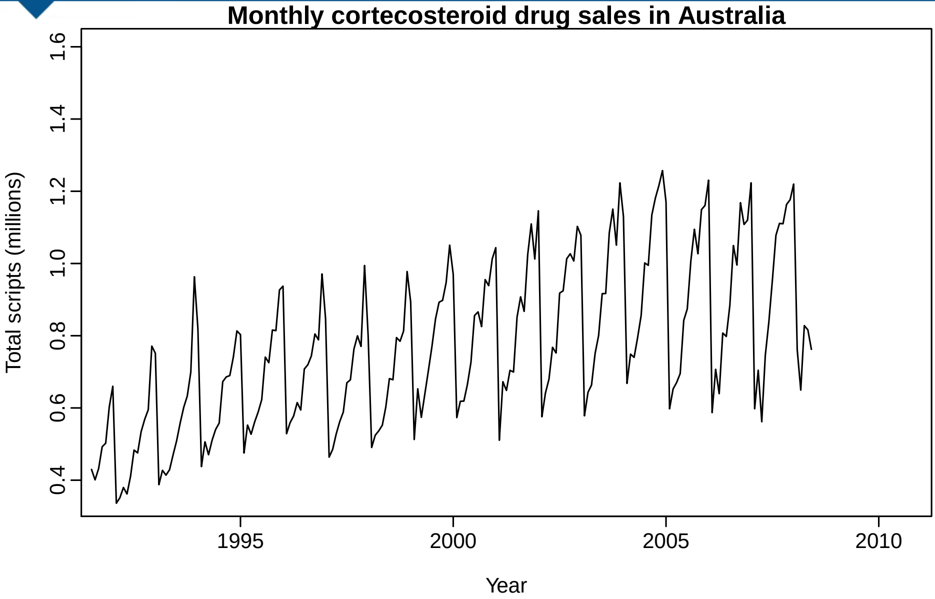
1. Common in business to have over 1000 products that need forecasting at least monthly.
2. Forecasts are often required by people who are untrained in time series analysis.
3. Some types of data can be decomposed into a large number of univariate time series that need to be forecast.

Specifications

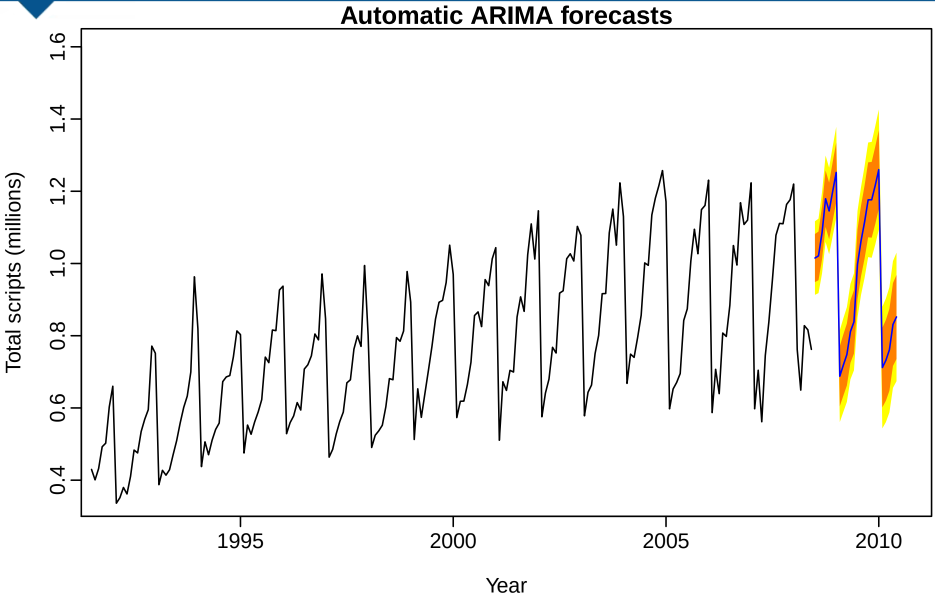
Automatic forecasting algorithms must:

- determine an appropriate time series model;
- estimate the parameters;
- compute the forecasts with prediction intervals.

Example: Cortecosteroid sales



Example: Cortecosteroid sales



Outline

Exponential smoothing methods

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

Exponential smoothing methods

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

N,N: Simple exponential smoothing

Exponential smoothing methods

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

N,N: Simple exponential smoothing

A,N: Holt's linear method

Exponential smoothing methods

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

N,N: Simple exponential smoothing

A,N: Holt's linear method

A_d,N: Additive damped trend method

Exponential smoothing methods

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

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

Exponential smoothing methods

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

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

Exponential smoothing methods

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

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

Exponential smoothing methods

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

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

Exponential smoothing methods

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

- There are 15 separate exponential smoothing methods.

Exponential smoothing methods

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

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

Exponential smoothing methods

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

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

Exponential smoothing methods

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

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

Examples:

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

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

M,A,M: Multiplicative Holt-Winters' method with multiplicative errors

Exponential smoothing methods

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

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

$\uparrow$

Trend

Examples:

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

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

M,A,M: Multiplicative Holt-Winters' method with multiplicative errors

Exponential smoothing methods

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

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

↑ Trend ↖ Seasonal

Examples:

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

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

M,A,M: Multiplicative Holt-Winters' method with multiplicative errors

Exponential smoothing methods

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

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

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

Examples:

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

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

M,A,M: Multiplicative Holt-Winters' method with multiplicative errors

Exponential smoothing methods

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

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

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

Examples:

A,N,N: Simple exponential smoothing with additive errors
A,A,N: Holt's linear method with additive errors
M,A,M: Multiplicative Holt-Winters' method with multiplicative errors

Exponential smoothing methods

Innovations state space models

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

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

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

Examples:

A,N,N:	Simple exponential smoothing with additive errors
A,A,N:	Holt's linear method with additive errors
M,A,M:	Multiplicative Holt-Winters' method with multiplicative errors

Automatic forecasting

From Hyndman et al. (IJF, 2002):

- Apply each of 30 models that are appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AIC:

$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$

  where $p = \#$ parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.

Automatic forecasting

From Hyndman et al. (IJF, 2002):

- Apply each of 30 models that are appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AIC:

$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$

where $p = \#$ parameters.

- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.

Automatic forecasting

From Hyndman et al. (IJF, 2002):

- Apply each of 30 models that are appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AIC:

$$\mathrm{AIC} = -2\log(\mathrm{Likelihood}) + 2p$$

  where $p = \#$ parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.

Automatic forecasting

From Hyndman et al. (IJF, 2002):

- Apply each of 30 models that are appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AIC:

$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$

  where $p = \#$ parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.

Automatic forecasting

From Hyndman et al. (IJF, 2002):

- Apply each of 30 models that are appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AIC:

$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$

  where $p = \#$ parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.

Automatic forecasting

From Hyndman et al. (IJF, 2002):

- Apply each of 30 models that are appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best method using AIC:

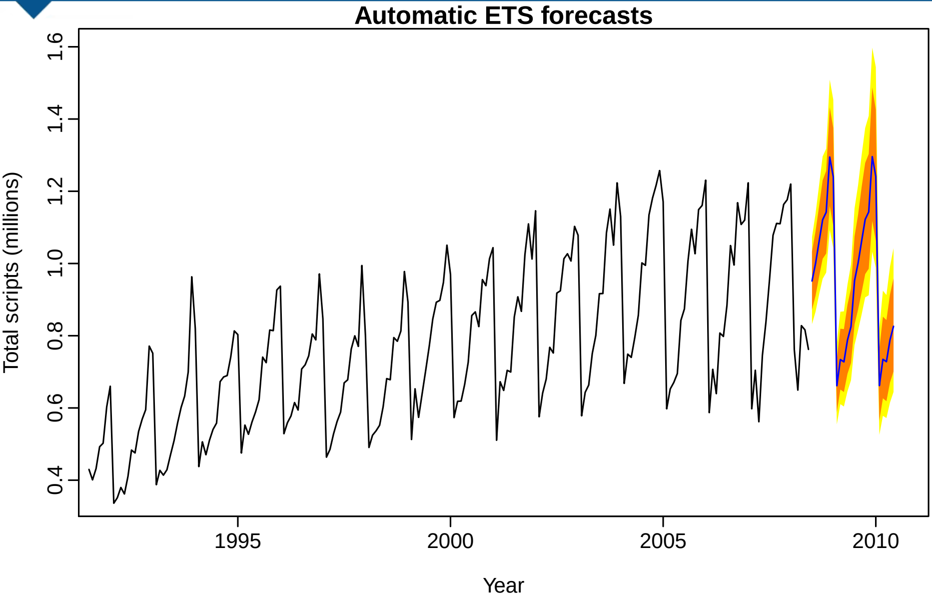
$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$

where $p = \#$ parameters.

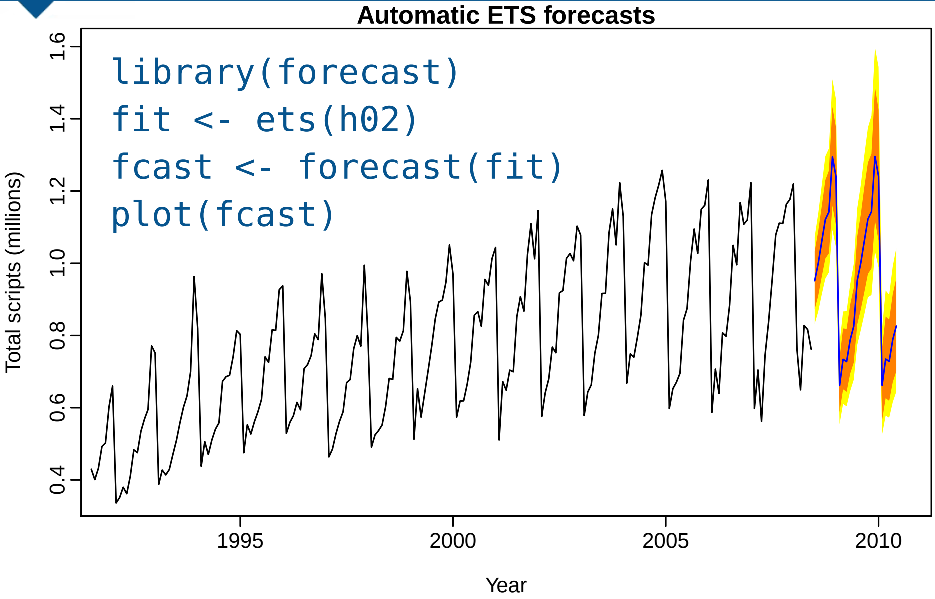
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.

Exponential smoothing



Exponential smoothing

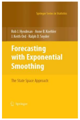


Exponential smoothing

```
> fit
ETS(M,Md,M)

  Smoothing parameters:
    alpha = 0.3318
    beta  = 4e-04
    gamma = 1e-04
    phi   = 0.9695
  Initial states:
    l = 0.4003
    b = 1.0233
    s = 0.8575 0.8183 0.7559 0.7627 0.6873 1.2884
        1.3456 1.1867 1.1653 1.1033 1.0398 0.9893

  sigma:  0.0651

     AIC       AICc        BIC
-121.97999 -118.68967  -65.57195
```

References

Hyndman, Koehler, Snyder, Grose (2002). “A state space framework for automatic forecasting using exponential smoothing methods”. *International Journal of Forecasting* **18**(3), 439–454



Hyndman, Koehler, Ord, Snyder (2008). *Forecasting with exponential smoothing: the state space approach*. Berlin: Springer-Verlag. `www.exponentialsmoothing.net`

Hyndman (2012). *forecast: Forecasting functions for time series*. `cran.r-project.org/package=forecast`

Outline

How does auto.arima() work?

A non-seasonal ARIMA process

$$\phi(B)(1-B)^d y_t = c + \theta(B)\varepsilon_t$$

Need to select appropriate orders p, q, d, and whether to include c.

How does auto.arima() work?

A non-seasonal ARIMA process

$$\phi(B)(1-B)^d y_t = c + \theta(B)\varepsilon_t$$

Need to select appropriate orders p, q, d, and whether to include c.

Hyndman & Khandakar (JSS, 2008) algorithm:

- Select no. differences d via KPSS unit root test.
- Select p, q, c by minimising AIC.
- Use stepwise search to traverse model space, starting with a simple model and considering nearby variants.

How does auto.arima() work?

A seasonal ARIMA process

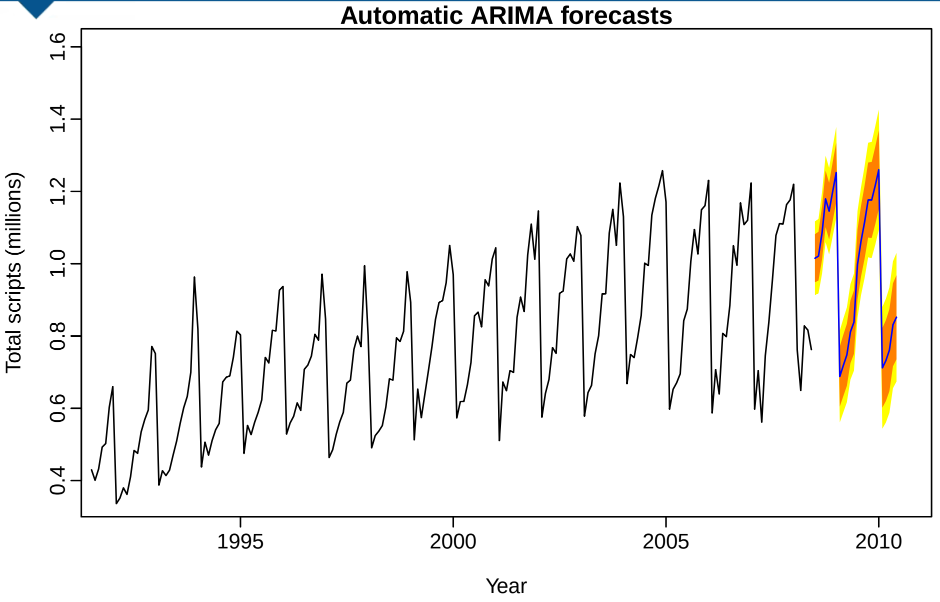
$$\Phi(B^m)\phi(B)(1-B)^d(1-B^m)^D y_t = c + \Theta(B^m)\theta(B)\varepsilon_t$$

Need to select appropriate orders p, q, d, P, Q, D, and whether to include c.

Hyndman & Khandakar (JSS, 2008) algorithm:

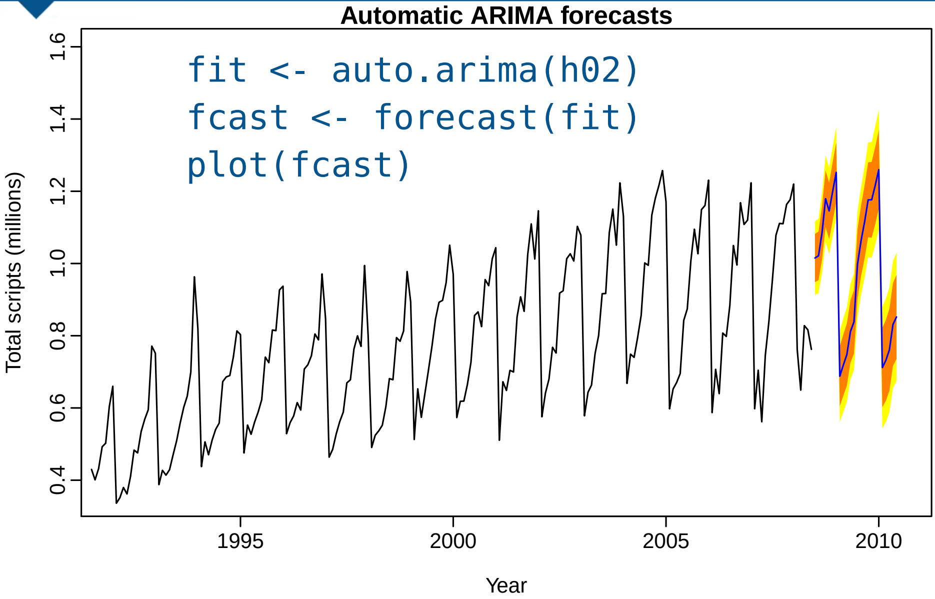
- Select no. differences d via KPSS unit root test.
- Select D using OCSB unit root test.
- Select p, q, P, Q, c by minimising AIC.
- Use stepwise search to traverse model space, starting with a simple model and considering nearby variants.

Auto ARIMA



Automatic ARIMA forecasts
Total scripts (millions)
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1995
2000
2005
2010
Year

Auto ARIMA



Automatic ARIMA forecasts
fit <- auto.arima(h02)
fcast <- forecast(fit)
plot(fcast)
Total scripts (millions)
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1995
2000
2005
2010
Year

Auto ARIMA

```
> fit
Series: h02
ARIMA(3,1,3)(0,1,1)[12]

Coefficients:
          ar1      ar2     ar3      ma1     ma2     ma3
      -0.3648  -0.0636  0.3568  -0.4850  0.0479  -0.353
s.e.   0.2198   0.3293  0.1268   0.2227  0.2755   0.212
         sma1
      -0.5931
s.e.   0.0651

sigma^2 estimated as 0.002706:  log likelihood=290.25
AIC=-564.5   AICc=-563.71   BIC=-538.48
```

References

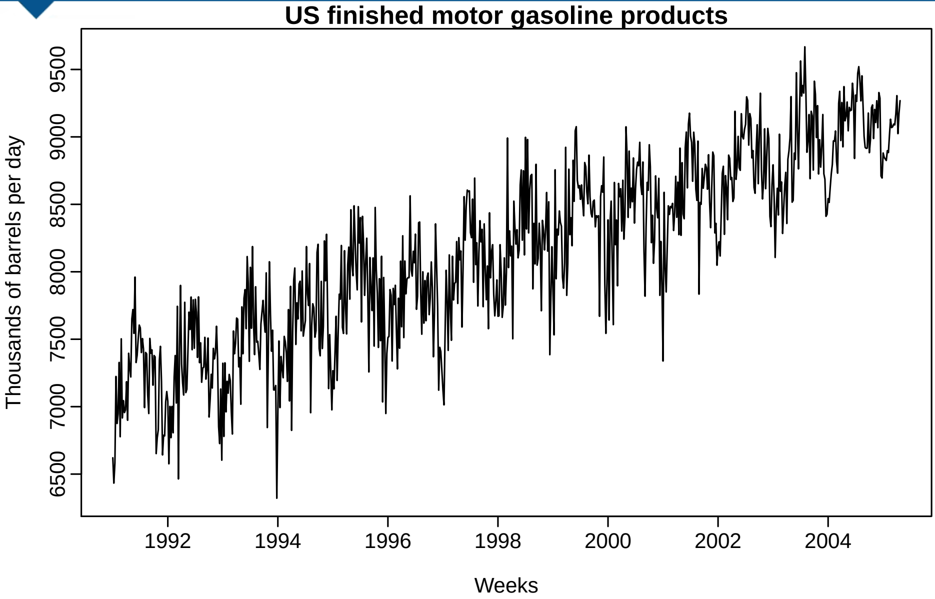
Hyndman, Khandakar (2008). "Automatic time series forecasting : the forecast package for R". *Journal of Statistical Software* **26**(3)

Hyndman (2011). "Major changes to the forecast package". `robjhyndman.com/researchtips/forecast3/`.

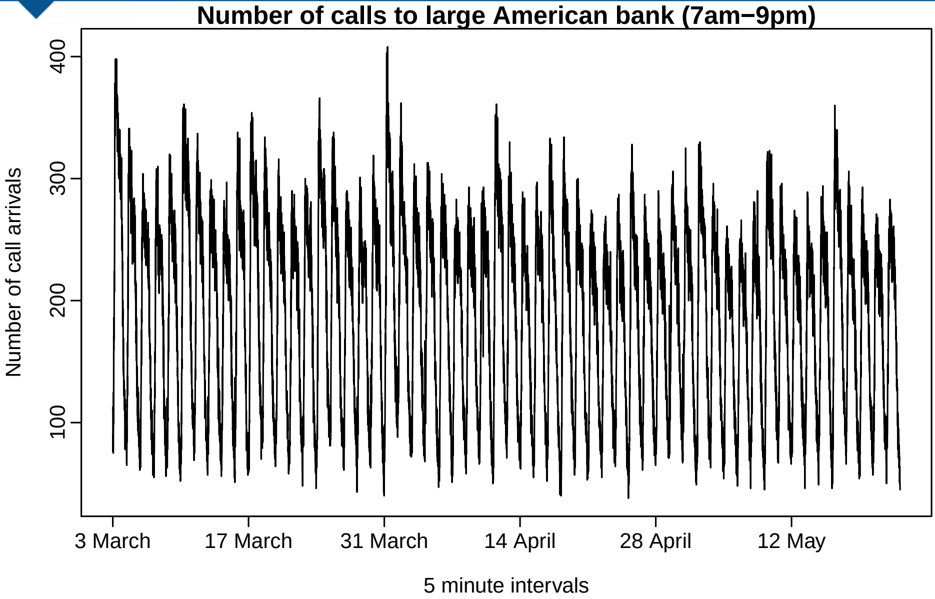
Hyndman (2012). *forecast: Forecasting functions for time series*. `cran.r-project.org/package=forecast`

Outline

Examples

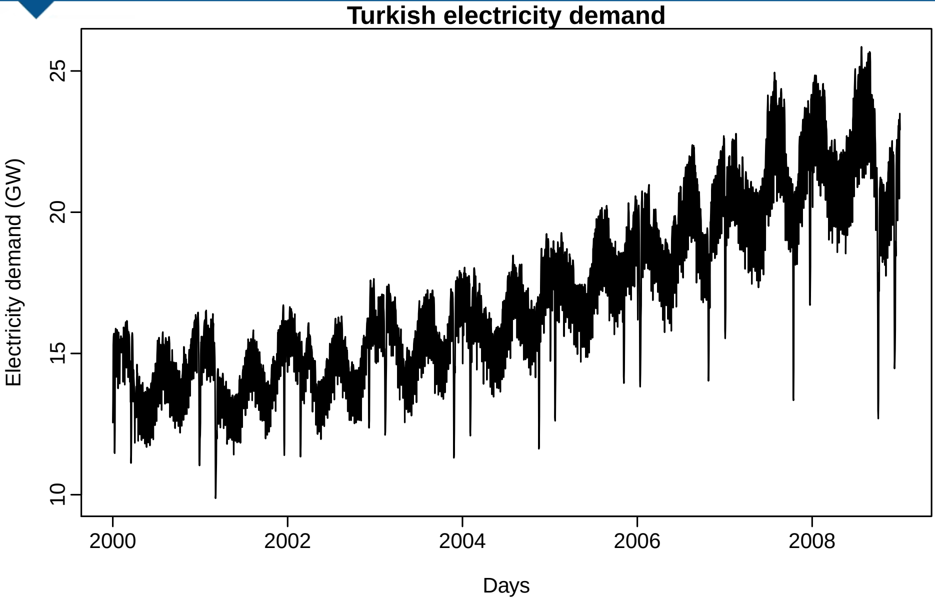

US finished motor gasoline products
Thousands of barrels per day
6500
7000
7500
8000
8500
9000
9500
1992
1994
1996
1998
2000
2002
2004
Weeks

Examples



Number of calls to large American bank (7am−9pm)
Number of call arrivals
100
200
300
400
3 March
17 March
31 March
14 April
28 April
12 May
5 minute intervals

Examples


Turkish electricity demand
Electricity demand (GW)
10
15
20
25
2000
2002
2004
2006
2008
Days

TBATS model

$$y_t = \text{observation at time } t$$

$$y_t^{(\omega)} = \begin{cases} (y_t^\omega - 1)/\omega & \text{if } \omega \neq 0; \\ \log y_t & \text{if } \omega = 0. \end{cases}$$

$$y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + d_t$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^{p} \phi_i d_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$\begin{aligned} s_{j,t}^{(i)} &= s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t \\ s_{j,t}^{(i)} &= -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t \end{aligned}$$

TBATS model

$$y_t = \text{observation at time } t$$

$$y_t^{(\omega)} = \begin{cases} (y_t^\omega - 1)/\omega & \text{if } \omega \neq 0; \\ \log y_t & \text{if } \omega = 0. \end{cases}$$

Box-Cox transformation

$$y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + d_t$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^{p} \phi_i d_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$\begin{aligned} s_{j,t}^{(i)} &= s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t \\ s_{j,t}^{(i)} &= -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t \end{aligned}$$

TBATS model

$$y_t = \text{observation at time } t$$

$$y_t^{(\omega)} = \begin{cases} (y_t^\omega - 1)/\omega & \text{if } \omega \neq 0; \\ \log y_t & \text{if } \omega = 0. \end{cases}$$

Box-Cox transformation

$$y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + d_t$$

M seasonal periods

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^{p} \phi_i d_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$\begin{aligned} s_{j,t}^{(i)} &= s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t \\ s_{j,t}^{(i)} &= -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t \end{aligned}$$

TBATS model

$$y_t = \text{observation at time } t$$

$$y_t^{(\omega)} = \begin{cases} (y_t^\omega - 1)/\omega & \text{if } \omega \neq 0; \\ \log y_t & \text{if } \omega = 0. \end{cases}$$

Box-Cox transformation

$$y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + d_t$$

M seasonal periods

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t$$

global and local trend

$$d_t = \sum_{i=1}^{p} \phi_i d_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$\begin{aligned} s_{j,t}^{(i)} &= s_{j,t-1}^{(i)} \cos\lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin\lambda_j^{(i)} + \gamma_1^{(i)} d_t \\ s_{j,t}^{(i)} &= -s_{j,t-1}^{(i)} \sin\lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos\lambda_j^{(i)} + \gamma_2^{(i)} d_t \end{aligned}$$

TBATS model

$$y_t = \text{observation at time } t$$

$$y_t^{(\omega)} = \begin{cases} (y_t^\omega - 1)/\omega & \text{if } \omega \neq 0; \\ \log y_t & \text{if } \omega = 0. \end{cases}$$

Box-Cox transformation

$$y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + d_t$$

M seasonal periods

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t$$

global and local trend

$$d_t = \sum_{i=1}^{p} \phi_i d_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

ARMA error

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$\begin{aligned} s_{j,t}^{(i)} &= \quad s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t \\ s_{j,t}^{(i)} &= -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t \end{aligned}$$

TBATS model

$$y_t = \text{observation at time } t$$

$$y_t^{(\omega)} = \begin{cases} (y_t^\omega - 1)/\omega & \text{if } \omega \neq 0; \\ \log y_t & \text{if } \omega = 0. \end{cases}$$

Box-Cox transformation

$$y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + d_t$$

M seasonal periods

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t$$

global and local trend

$$d_t = \sum_{i=1}^{p} \phi_i d_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

ARMA error

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$s_{j,t}^{(i)} = \quad s_{j,t-1}^{(i)} \text{c}$$

$$s_{j,t}^{(i)} = -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t$$

Fourier-like seasonal terms

TBATS model

$y_t =$ observation at time t

$y_t^{(\omega)} = \begin{cases} (y_t^{\omega} - 1)/\omega & \text{if } \omega \neq 0; \\ \text{lo} \end{cases}$

$y_t^{(\omega)} = \ell_{t-1}$

$\ell_t = \ell_{t-1}$

$b_t = (1 -$

$d_t = \sum_{i=1}^{p}$

$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$

$s_{j,t}^{(i)} = s_{j,t-1}^{(i)}$ c

$s_{j,t}^{(i)} = -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t$

TBATS

Trigonometric

Box-Cox

ARMA

Trend

Seasonal

Box-Cox transformation

M seasonal periods

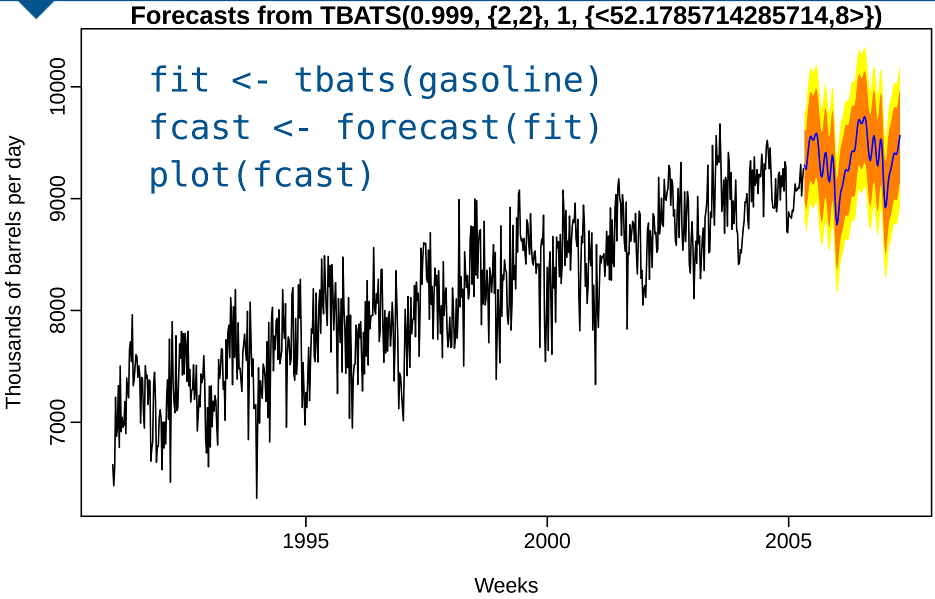
global and local trend

ARMA error

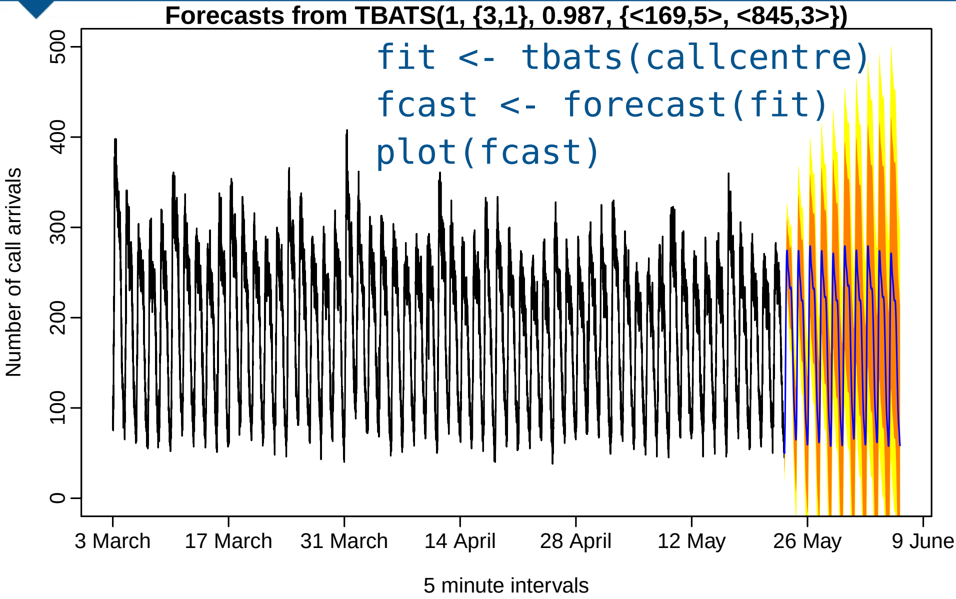
Fourier-like seasonal terms

Examples

```
fit <- tbats(gasoline)
fcast <- forecast(fit)
plot(fcast)
```



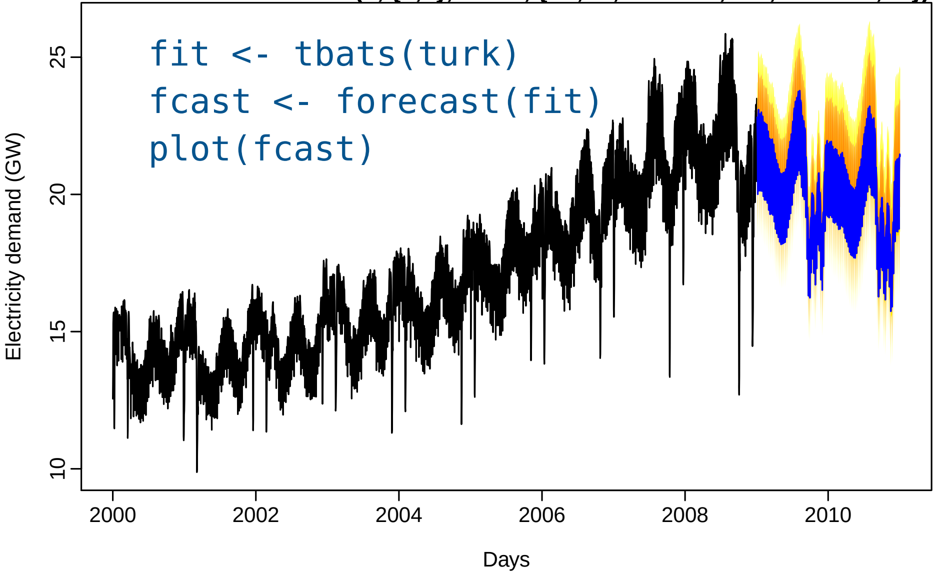
Examples


Forecasts from TBATS(1, {3,1}, 0.987, {<169,5>, <845,3>})
fit <- tbats(callcentre)
fcast <- forecast(fit)
plot(fcast)
Number of call arrivals
0
100
200
300
400
500
3 March
17 March
31 March
14 April
28 April
12 May
26 May
9 June
5 minute intervals

Examples

```r
fit <- tbats(turk)
fcast <- forecast(fit)
plot(fcast)
```

Forecasts from TBATS(0, {5,3}, 0.997, {<7,3>, <354.37,12>, <365.25,4>})



References

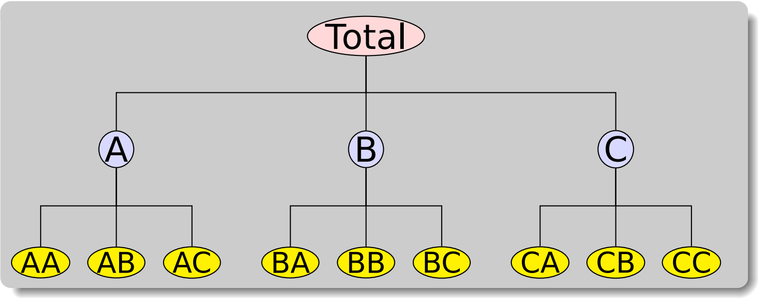
Automatic algorithm described in
De Livera, Hyndman, Snyder (2011). "Forecasting time series with complex seasonal patterns using exponential smoothing". *Journal of the American Statistical Association* **106**(496), 1513–1527.

Slightly improved algorithm implemented in
Hyndman (2012). *forecast: Forecasting functions for time series*. `cran.r-project.org/package=forecast`.

More work required!

Outline

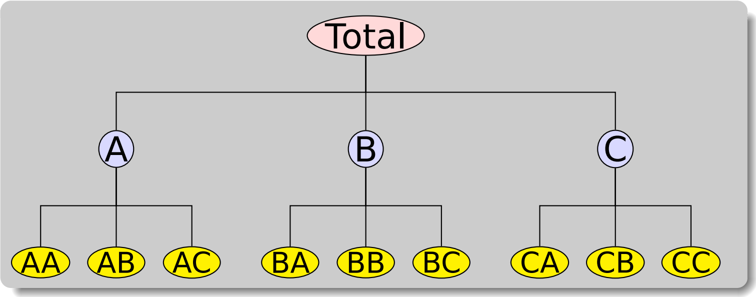
Introduction



Examples

- Manufacturing product hierarchies
- Pharmaceutical sales
- Net labour turnover

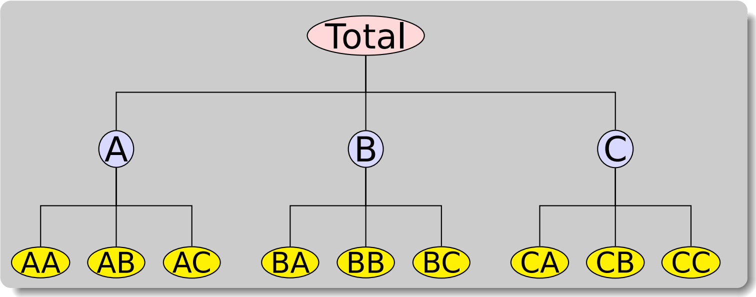
Introduction



Examples

- Manufacturing product hierarchies
- Pharmaceutical sales
- Net labour turnover

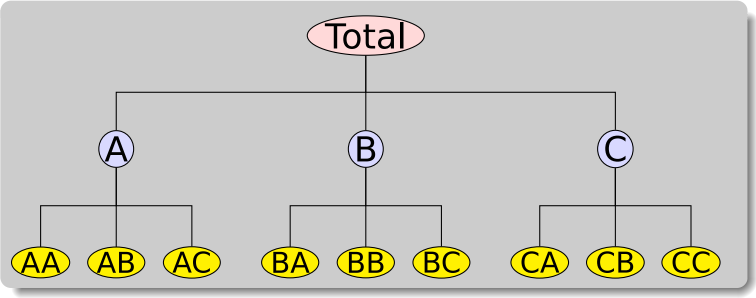
Introduction



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.
- 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.
- Forecasts should be “aggregate consistent”, unbiased, minimum variance.
- How to compute forecast intervals?

Hierarchical/grouped time series

- A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.
- 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.
- Forecasts should be “aggregate consistent”, unbiased, minimum variance.
- How to compute forecast intervals?

Hierarchical/grouped time series

- A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.
- 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.
- Forecasts should be “aggregate consistent”, unbiased, minimum variance.
- How to compute forecast intervals?

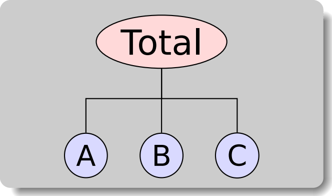
Hierarchical/grouped time series

- A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.
- 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.
- Forecasts should be “aggregate consistent”, unbiased, minimum variance.
- How to compute forecast intervals?

Hierarchical/grouped time series

- A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.
- 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.
- Forecasts should be "aggregate consistent", unbiased, minimum variance.
- How to compute forecast intervals?

Notation



K : number of levels in hierarchy (excl. Total).

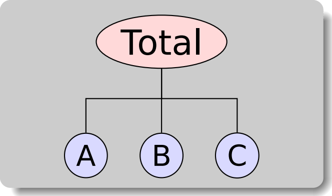
Y_t : observed aggregate of all series at time t.

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

$\boldsymbol{Y}_{i,t}$: vector of all series at level i in time t.

$\boldsymbol{Y}_t = [Y_t, \boldsymbol{Y}_{1,t}, \dots, \boldsymbol{Y}_{K,t}]'$

Notation



K : number of levels in hierarchy (excl. Total).

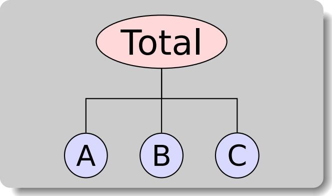
Y_t : observed aggregate of all series at time t.

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

$\boldsymbol{Y}_{i,t}$: vector of all series at level i in time t.

$$\boldsymbol{Y}_t = [Y_t, \boldsymbol{Y}_{1,t}, \dots, \boldsymbol{Y}_{K,t}]'$$

Notation



K : number of levels in hierarchy (excl. Total).

Y_t : observed aggregate of all series at time t.

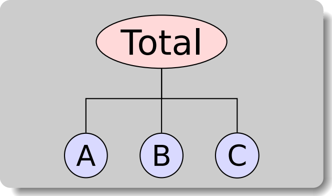
$Y_{X,t}$: observation on series X at time t.

$\boldsymbol{Y}_{i,t}$: vector of all series at level i in time t.

$\boldsymbol{Y}_t = [Y_t, \boldsymbol{Y}_{1,t}, \dots, \boldsymbol{Y}_{K,t}]'$

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

Notation



K : number of levels in hierarchy (excl. Total).

Y_t : observed aggregate of all series at time t.

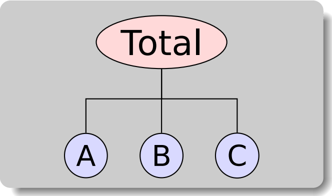
$Y_{X,t}$: observation on series X at time t.

$\boldsymbol{Y}_{i,t}$: vector of all series at level i in time t.

$\boldsymbol{Y}_t = [Y_t, \boldsymbol{Y}_{1,t}, \dots, \boldsymbol{Y}_{K,t}]'$

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

Notation



K : number of levels in hierarchy (excl. Total).

Y_t : observed aggregate of all series at time t.

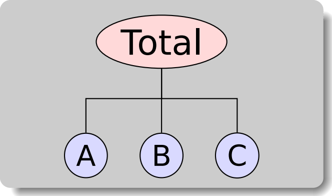
$Y_{X,t}$: observation on series X at time t.

$\boldsymbol{Y}_{i,t}$: vector of all series at level i in time t.

$\boldsymbol{Y}_t = [Y_t, \boldsymbol{Y}_{1,t}, \ldots, \boldsymbol{Y}_{K,t}]'$

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

Notation



K : number of levels in hierarchy (excl. Total).

Y_t : observed aggregate of all series at time t.

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

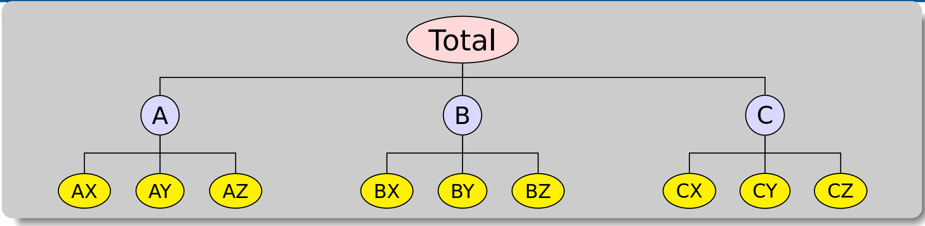
$\boldsymbol{Y}_{i,t}$: vector of all series at level i in time t.

$\boldsymbol{Y}_t = [Y_t, \boldsymbol{Y}_{1,t}, \dots, \boldsymbol{Y}_{K,t}]'$

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

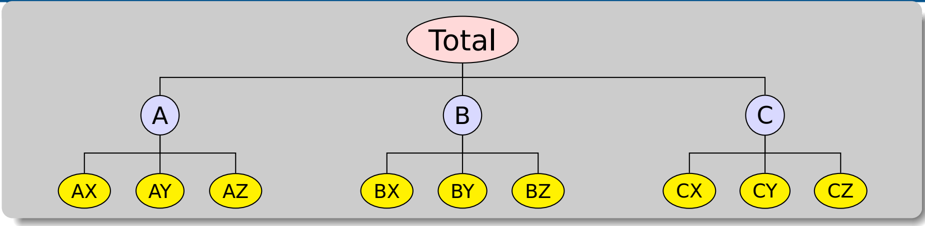
$$\boxed{\boldsymbol{Y}_t = S\boldsymbol{Y}_{K,t}}$$

Hierarchical data



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

Hierarchical data



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

Grouped data

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

Grouped data

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

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 forecasts for horizon h, made at time n, stacked in same order as $\boldsymbol{Y}_t$.

$\boldsymbol{Y}_t = S\boldsymbol{Y}_{K,t}$. 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 forecasts for horizon h, made at time n, stacked in same order as $\boldsymbol{Y}_t$.

$\boldsymbol{Y}_t = S\boldsymbol{Y}_{K,t}$. 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 forecasts for horizon h, made at time n, stacked in same order as $\boldsymbol{Y}_t$.

$\boldsymbol{Y}_t = S\boldsymbol{Y}_{K,t}$. So $\hat{\boldsymbol{Y}}_n(h) = S\beta_n(h) + \varepsilon_h$.

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

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 forecasts for horizon h, made at time n, stacked in same order as $\boldsymbol{Y}_t$.

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

- $\beta_n(h) = E[\boldsymbol{Y}_{K,n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.
- ε_h has zero mean and covariance matrix Σ_h.
- Estimate $\beta_n(h)$ using GLS?
- Revised forecasts: $\tilde{\boldsymbol{Y}}_n(h) = S\hat{\beta}_n(h)$

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 forecasts for horizon h, made at time n, stacked in same order as $\boldsymbol{Y}_t$.

$\boldsymbol{Y}_t = S\boldsymbol{Y}_{K,t}$. So $\hat{\boldsymbol{Y}}_n(h) = S\beta_n(h) + \varepsilon_h$.

- $\beta_n(h) = E[\boldsymbol{Y}_{K,n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.
- ε_h has zero mean and covariance matrix Σ_h.
- Estimate $\beta_n(h)$ using GLS?
- Revised forecasts: $\tilde{\boldsymbol{Y}}_n(h) = S\hat{\beta}_n(h)$

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 forecasts for horizon h, made at time n, stacked in same order as $\boldsymbol{Y}_t$.

$\boldsymbol{Y}_t = S\boldsymbol{Y}_{K,t}$. So $\hat{\boldsymbol{Y}}_n(h) = S\beta_n(h) + \varepsilon_h$.

- $\beta_n(h) = E[\boldsymbol{Y}_{K,n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.
- ε_h has zero mean and covariance matrix Σ_h.
- Estimate $\beta_n(h)$ using GLS?
- Revised forecasts: $\tilde{\boldsymbol{Y}}_n(h) = S\hat{\beta}_n(h)$

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 forecasts for horizon h, made at time n, stacked in same order as $\boldsymbol{Y}_t$.

$\boldsymbol{Y}_t = S\boldsymbol{Y}_{K,t}$. So $\hat{\boldsymbol{Y}}_n(h) = S\beta_n(h) + \varepsilon_h$.

- $\beta_n(h) = E[\boldsymbol{Y}_{K,n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.
- ε_h has zero mean and covariance matrix Σ_h.
- Estimate $\beta_n(h)$ using GLS?
- Revised forecasts: $\tilde{\boldsymbol{Y}}_n(h) = S\hat{\beta}_n(h)$

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 forecasts for horizon h, made at time n, stacked in same order as $\boldsymbol{Y}_t$.

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

- $\beta_n(h) = E[\boldsymbol{Y}_{K,n+h} \mid \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n]$.
- ε_h has zero mean and covariance matrix Σ_h.
- Estimate $\beta_n(h)$ using GLS?
- Revised forecasts: $\boxed{\tilde{\boldsymbol{Y}}_n(h) = S\hat{\beta}_n(h)}$

Optimal combination forecasts

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

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

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

Optimal combination forecasts

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

Base forecasts

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

Base forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- **Problem:** Don't know Σ_h and 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

$$(S'\Sigma_h^{\dagger}S)^{-1}S'\Sigma_h^{\dagger} = (S'S)^{-1}S'$$

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

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 Base forecasts

- $\Sigma_h^\dagger$ is generalized inverse of Σ_h.
- **Problem:** Don't know Σ_h and 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 | Base forecasts

- $\Sigma_h^\dagger$ is generalized inverse of Σ_h.
- **Problem:** Don't know Σ_h and 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   Base forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- **Problem:** Don't know Σ_h and 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 — Base forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- **Problem:** Don't know Σ_h and 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 — Base forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- **Problem:** Don't know Σ_h and 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

Base forecasts

- $\Sigma_h^{\dagger}$ is generalized inverse of Σ_h.
- **Problem:** Don't know Σ_h and 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   Base forecasts

- $\Sigma_h^\dagger$ is generalized inverse of Σ_h.
- **Problem:** Don't know Σ_h and 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 base forecasts.
- Computational difficulties in big hierarchies due to size of S matrix.
- Optimal weights are $S(S'S)^{-1}S'$
- Weights are independent of the data!

Optimal combination forecasts

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

- GLS = OLS.
- Optimal weighted average of base forecasts.
- Computational difficulties in big hierarchies due to size of S matrix.
- Optimal weights are $S(S'S)^{-1}S'$
- Weights are independent of the data!

Optimal combination forecasts

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

- GLS =OLS.
- Optimal weighted average of base forecasts.
- Computational difficulties in big hierarchies due to size of S matrix.
- Optimal weights are $S(S'S)^{-1}S'$
- Weights are independent of the data!

Optimal combination forecasts

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

- GLS =OLS.
- Optimal weighted average of base forecasts.
- Computational difficulties in big hierarchies due to size of S matrix.
- Optimal weights are $S(S'S)^{-1}S'$
- Weights are independent of the data!

Optimal combination forecasts

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

- GLS =OLS.
- Optimal weighted average of base forecasts.
- Computational difficulties in big hierarchies due to size of S matrix.
- Optimal weights are $S(S'S)^{-1}S'$
- Weights are independent of the data!

Optimal combination forecasts

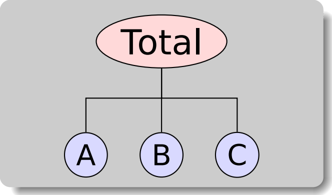
$$\tilde{\mathbf{Y}}_n(h) = S(S'S)^{-1}S'\hat{\mathbf{Y}}_n(h)$$

- GLS =OLS.
- Optimal weighted average of base forecasts.
- Computational difficulties in big hierarchies due to size of S matrix.
- Optimal weights are $S(S'S)^{-1}S'$
- Weights are independent of the data!

Optimal combination forecasts

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

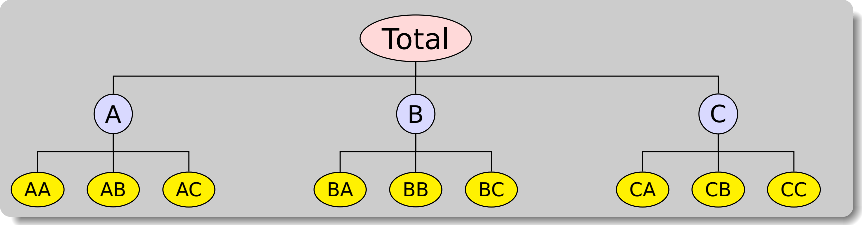
- GLS =OLS.
- Optimal weighted average of base forecasts.
- Computational difficulties in big hierarchies due to size of S matrix.
- Optimal weights are $S(S'S)^{-1}S'$
- Weights are independent of the data!



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

$$\begin{bmatrix} 0.75 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.75 & -0.25 & -0.25 \\ 0.25 & -0.25 & 0.75 & -0.25 \\ 0.25 & -0.25 & -0.25 & 0.75 \end{bmatrix}$$

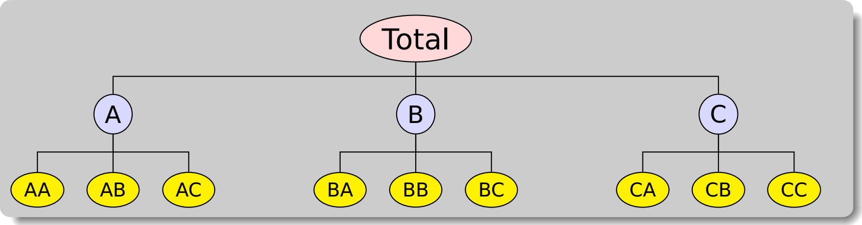
Optimal combination forecasts



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

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

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 and problems

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

$$\text{Var}[\tilde{\boldsymbol{Y}}_n(h)] = S\Omega_h S' = \Sigma_h.$$

- Covariates can be included in base forecasts.
- Point forecasts are always consistent.
- Need to estimate Ω_h to produce prediction intervals.
- Very simple and flexible method. Can work with any hierarchical or grouped time series.

Features and problems

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

$$\mathrm{Var}[\tilde{\boldsymbol{Y}}_n(h)] = S\Omega_h S' = \Sigma_h.$$

- Covariates can be included in base forecasts.
- Point forecasts are always consistent.
- Need to estimate Ω_h to produce prediction intervals.
- Very simple and flexible method. Can work with any hierarchical or grouped time series.

Features and problems

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

$$\text{Var}[\tilde{\boldsymbol{Y}}_n(h)] = S\Omega_h S' = \Sigma_h.$$

- Covariates can be included in base forecasts.
- Point forecasts are always consistent.
- Need to estimate Ω_h to produce prediction intervals.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.

Features and problems

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

$$\text{Var}[\tilde{\mathbf{Y}}_n(h)] = S\Omega_h S' = \Sigma_h.$$

- Covariates can be included in base forecasts.
- Point forecasts are always consistent.
- Need to estimate Ω_h to produce prediction intervals.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.

Features and problems

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

$$\text{Var}[\tilde{\mathbf{Y}}_n(h)] = S\Omega_h S' = \Sigma_h.$$

- Covariates can be included in base forecasts.
- Point forecasts are always consistent.
- Need to estimate Ω_h to produce prediction intervals.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.

Features and problems

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

$$\text{Var}[\tilde{\mathbf{Y}}_n(h)] = S\Omega_h S' = \Sigma_h.$$

- Covariates can be included in base forecasts.
- Point forecasts are always consistent.
- Need to estimate Ω_h to produce prediction intervals.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.

Features and problems

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

$$\text{Var}[\tilde{\mathbf{Y}}_n(h)] = S\Omega_h S' = \Sigma_h.$$

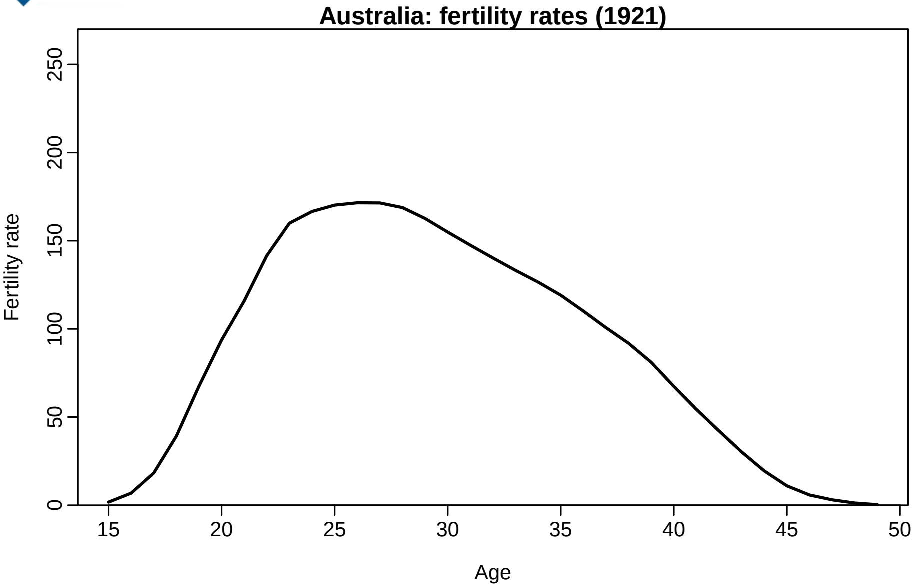
- Covariates can be included in base forecasts.
- Point forecasts are always consistent.
- Need to estimate Ω_h to produce prediction intervals.
- Very simple and flexible method. Can work with *any* hierarchical or grouped time series.

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

Hyndman, Ahmed, Shang (2011). *hts: Hierarchical time series*. `cran.r-project.org/package=hts`

Outline

Fertility rates



Some notation

Let $y_{t,x}$ be the observed data in period t at age x, $t = 1, \dots, n$.

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- Estimate $f_t(x)$ using penalized regression splines.
- Estimate $\mu(x)$ as mean $f_t(x)$ across years.
- Estimate $\beta_{t,k}$ and $\phi_k(x)$ using functional (weighted) principal components.

Some notation

Let $y_{t,x}$ be the observed data in period t at age x, $t = 1, \dots, n$.

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- Estimate $f_t(x)$ using penalized regression splines.
- Estimate $\mu(x)$ as mean $f_t(x)$ across years.
- Estimate $\beta_{t,k}$ and $\phi_k(x)$ using functional (weighted) principal components.
- $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \mathrm{N}(0, v(x))$.

Some notation

Let $y_{t,x}$ be the observed data in period t at age x, $t = 1, \dots, n$.

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- Estimate $f_t(x)$ using penalized regression splines.
- Estimate $\mu(x)$ as mean $f_t(x)$ across years.
- Estimate $\beta_{t,k}$ and $\phi_k(x)$ using functional (weighted) principal components.
- $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \mathrm{N}(0, v(x))$.

Some notation

Let $y_{t,x}$ be the observed data in period t at age x, $t = 1, \dots, n$.

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- Estimate $f_t(x)$ using penalized regression splines.
- Estimate $\mu(x)$ as mean $f_t(x)$ across years.
- Estimate $\beta_{t,k}$ and $\phi_k(x)$ using functional (weighted) principal components.
- $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \text{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \text{N}(0, v(x))$.

Some notation

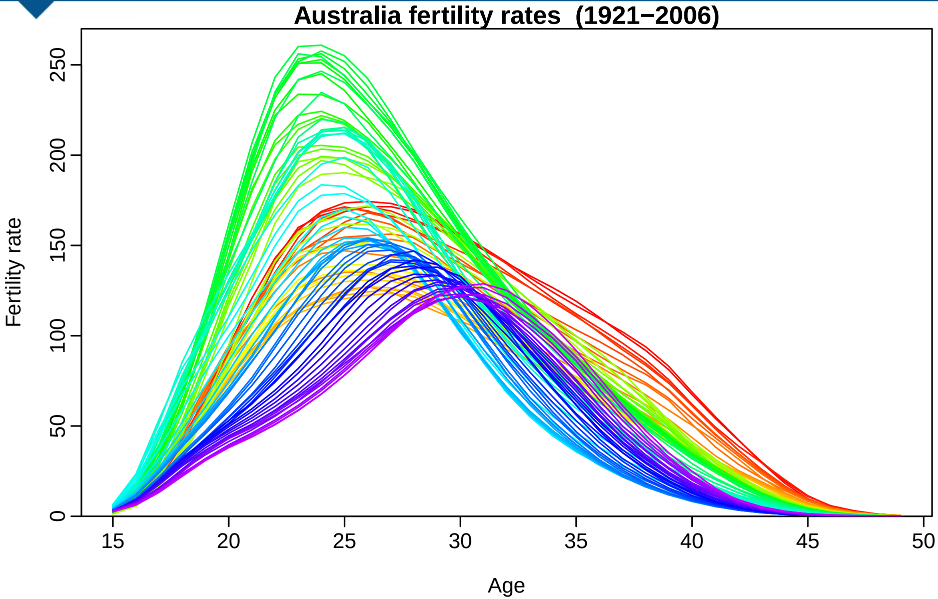
Let $y_{t,x}$ be the observed data in period t at age x, $t = 1, \dots, n$.

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

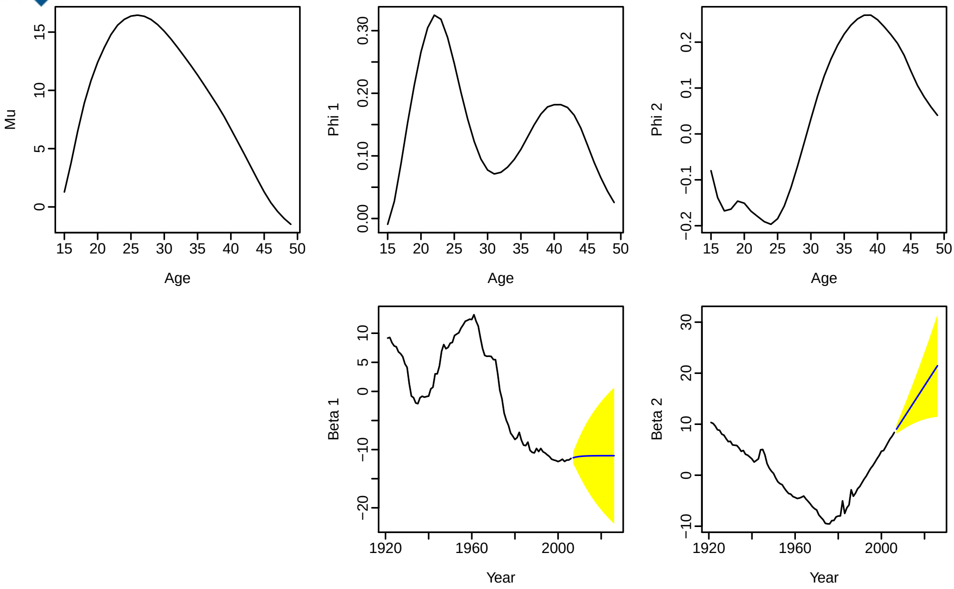
$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- Estimate $f_t(x)$ using penalized regression splines.
- Estimate $\mu(x)$ as mean $f_t(x)$ across years.
- Estimate $\beta_{t,k}$ and $\phi_k(x)$ using functional (weighted) principal components.
- $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \text{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \text{N}(0, v(x))$.

Fertility application



Fertility model



Functional time series model

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- The eigenfunctions $\phi_k(x)$ show the main regions of variation.
- The scores $\{\beta_{t,k}\}$ are uncorrelated by construction. So we can forecast each $\beta_{t,k}$ using a univariate time series model.
- Univariate ARIMA models used for automatic forecasting of scores.

Functional time series model

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- The eigenfunctions $\phi_k(x)$ show the main regions of variation.
- The scores $\{\beta_{t,k}\}$ are uncorrelated by construction. So we can forecast each $\beta_{t,k}$ using a univariate time series model.
- Univariate ARIMA models used for automatic forecasting of scores.

Functional time series model

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- The eigenfunctions $\phi_k(x)$ show the main regions of variation.
- The scores $\{\beta_{t,k}\}$ are uncorrelated by construction. So we can forecast each $\beta_{t,k}$ using a univariate time series model.
- Univariate ARIMA models used for automatic forecasting of scores.

Functional time series model

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- The eigenfunctions $\phi_k(x)$ show the main regions of variation.
- The scores $\{\beta_{t,k}\}$ are uncorrelated by construction. So we can forecast each $\beta_{t,k}$ using a univariate time series model.
- Univariate ARIMA models used for automatic forecasting of scores.

Forecasts

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

Forecasts

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

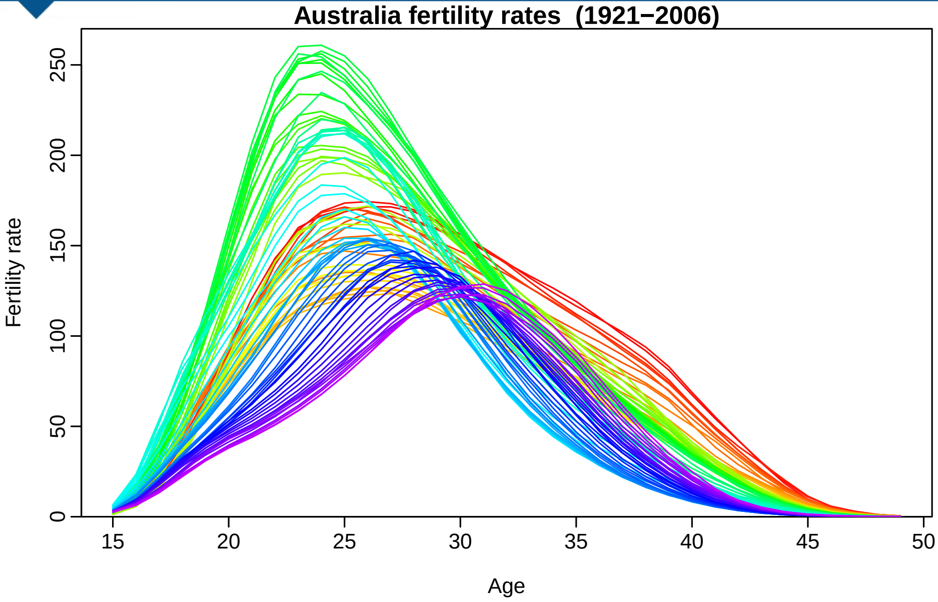
$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

$$\text{E}[y_{n+h,x} \mid \boldsymbol{y}] = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{n+h,k}\,\hat{\phi}_k(x)$$

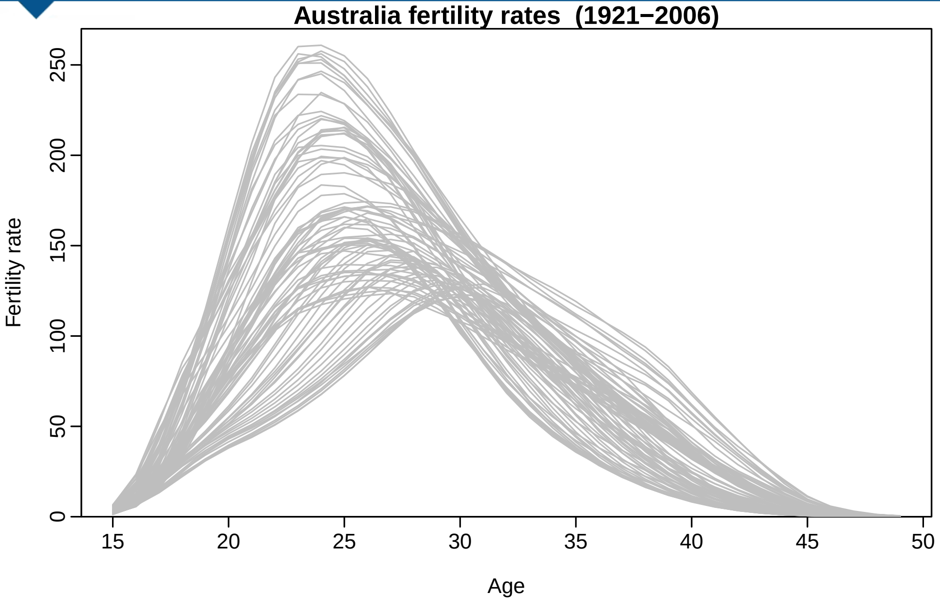
$$\text{Var}[y_{n+h,x} \mid \boldsymbol{y}] = \hat{\sigma}_\mu^2(x) + \sum_{k=1}^{K} v_{n+h,k}\,\hat{\phi}_k^2(x) + \sigma_t^2(x) + v(x)$$

where $v_{n+h,k} = \text{Var}(\beta_{n+h,k} \mid \beta_{1,k}, \dots, \beta_{n,k})$
and $\boldsymbol{y} = [y_{1,x}, \dots, y_{n,x}]$.

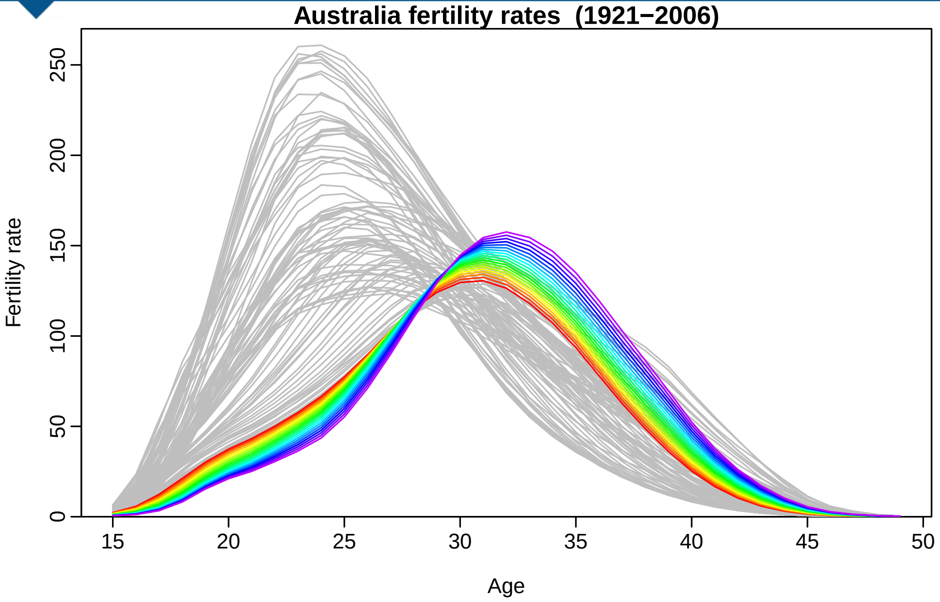
Forecasts of $f_t(x)$



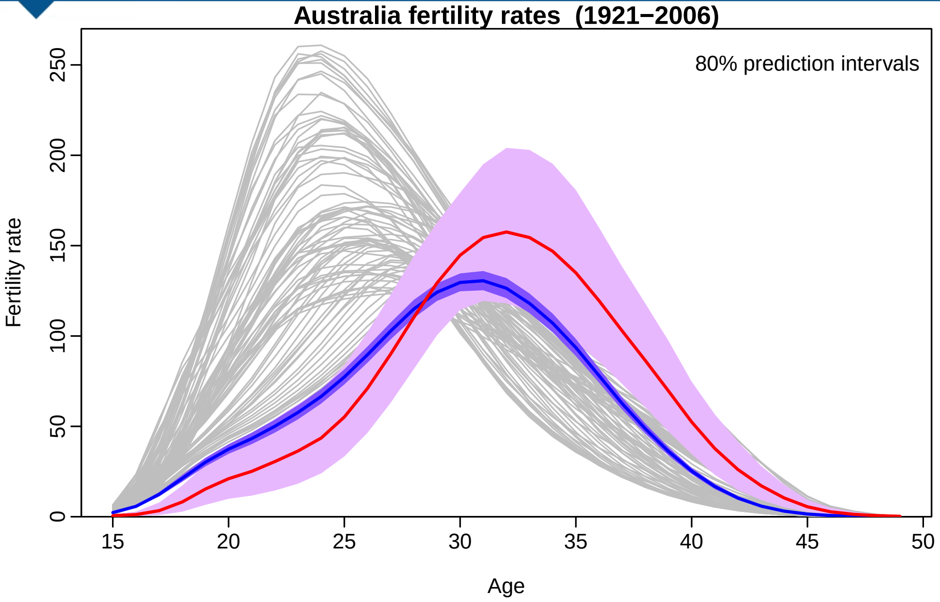
Forecasts of $f_t(x)$



Forecasts of $f_t(x)$



Forecasts of $f_t(x)$



References

Hyndman, Ullah (2007). “Robust forecasting of mortality and fertility rates: A functional data approach”. *Computational Statistics and Data Analysis* **51**(10), 4942–4956

Hyndman, Shang (2009). “Forecasting functional time series (with discussion)”. *Journal of the Korean Statistical Society* **38**(3), 199–221

Hyndman (2012). *demography: Forecasting mortality, fertility, migration and population data*. `cran.r-project.org/package=demography`

Outline

Forecasting groups

Let $f_{t,j}(x)$ be the smoothed mortality rate for age x in group j in year t.

Forecasting groups

Let $f_{t,j}(x)$ be the smoothed mortality rate for age x in group j in year t.

- Groups may be males and females.
- Groups may be states within a country.
- Expected that groups will behave similarly.
- **Coherent** forecasts do not diverge over time.
- Existing functional models do not impose coherence.

Forecasting groups

Let $f_{t,j}(x)$ be the smoothed mortality rate for age x in group j in year t.

- Groups may be males and females.
- Groups may be states within a country.
- Expected that groups will behave similarly.
- **Coherent** forecasts do not diverge over time.
- Existing functional models do not impose coherence.

Forecasting groups

Let $f_{t,j}(x)$ be the smoothed mortality rate for age x in group j in year t.

- Groups may be males and females.
- Groups may be states within a country.
- Expected that groups will behave similarly.
- **Coherent** forecasts do not diverge over time.
- Existing functional models do not impose coherence.

Forecasting groups

Let $f_{t,j}(x)$ be the smoothed mortality rate for age x in group j in year t.

- Groups may be males and females.
- Groups may be states within a country.
- Expected that groups will behave similarly.
- **Coherent** forecasts do not diverge over time.
- Existing functional models do not impose coherence.

Forecasting groups

Let $f_{t,j}(x)$ be the smoothed mortality rate for age x in group j in year t.

- Groups may be males and females.
- Groups may be states within a country.
- Expected that groups will behave similarly.
- **Coherent** forecasts do not diverge over time.
- Existing functional models do not impose coherence.

Forecasting the coefficients

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\, \phi_k(x) + e_t(x)$$

- We use ARIMA models for each coefficient $\{\beta_{1,j,k}, \dots, \beta_{n,j,k}\}$.
- The ARIMA models are non-stationary for the first few coefficients ($k = 1, 2$)
- Non-stationary ARIMA forecasts will diverge. Hence the mortality forecasts are not coherent.

Forecasting the coefficients

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- We use ARIMA models for each coefficient $\{\beta_{1,j,k}, \dots, \beta_{n,j,k}\}$.
- The ARIMA models are non-stationary for the first few coefficients ($k = 1, 2$)
- Non-stationary ARIMA forecasts will diverge. Hence the mortality forecasts are not coherent.

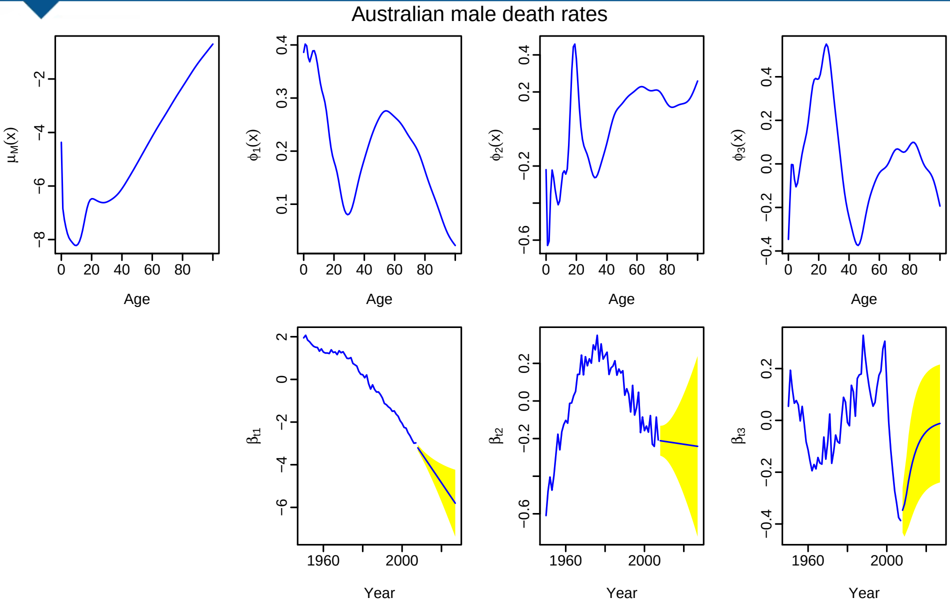
Forecasting the coefficients

$$y_{t,x} = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

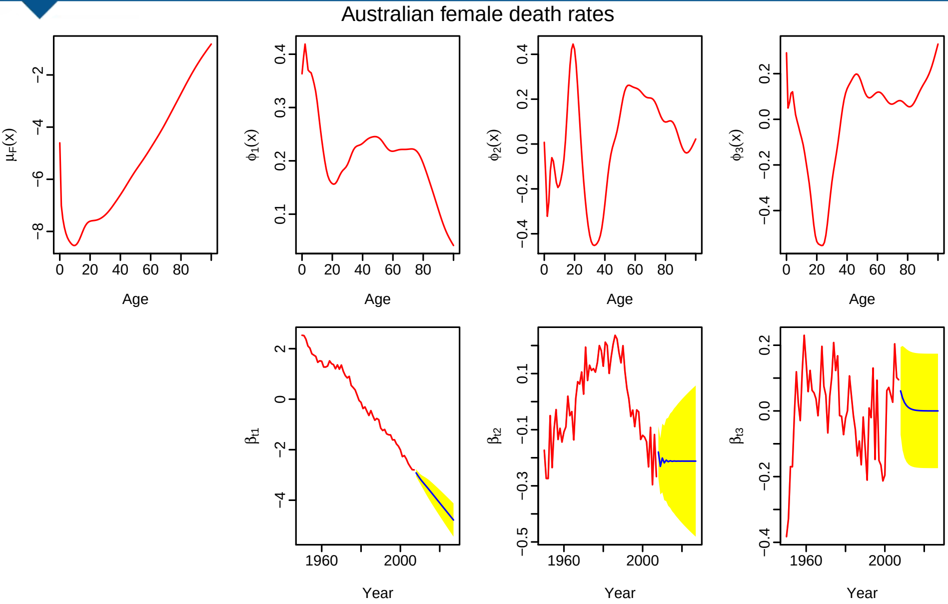
$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

- We use ARIMA models for each coefficient $\{\beta_{1,j,k}, \dots, \beta_{n,j,k}\}$.
- The ARIMA models are non-stationary for the first few coefficients ($k = 1, 2$)
- Non-stationary ARIMA forecasts will diverge. Hence the mortality forecasts are not coherent.

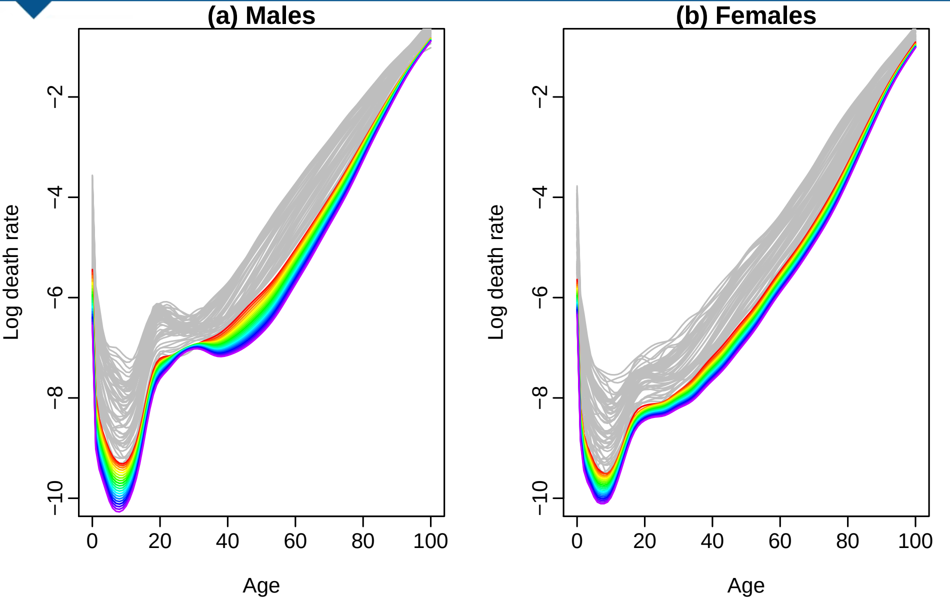
Male fts model



Female fts model



Australian mortality forecasts



Mortality product and ratios

Key idea

Model the geometric mean and the mortality ratio instead of the individual rates for each sex separately.

$$p_t(x) = \sqrt{f_{t,M}(x) f_{t,F}(x)} \quad \text{and} \quad r_t(x) = \sqrt{f_{t,M}(x) \big/ f_{t,F}(x)}.$$

- Product and ratio are approximately independent
- Ratio should be stationary (for coherence) but product can be non-stationary.

Mortality product and ratios

Key idea

Model the geometric mean and the mortality ratio instead of the individual rates for each sex separately.

$$p_t(x) = \sqrt{f_{t,M}(x) f_{t,F}(x)} \quad \text{and} \quad r_t(x) = \sqrt{f_{t,M}(x) \big/ f_{t,F}(x)}.$$

- Product and ratio are approximately independent
- Ratio should be stationary (for coherence) but product can be non-stationary.

Mortality product and ratios

Key idea

Model the geometric mean and the mortality ratio instead of the individual rates for each sex separately.

$$p_t(x) = \sqrt{f_{t,M}(x) f_{t,F}(x)} \quad \text{and} \quad r_t(x) = \sqrt{f_{t,M}(x) \big/ f_{t,F}(x)}.$$

- Product and ratio are approximately independent
- Ratio should be stationary (for coherence) but product can be non-stationary.

Mortality product and ratios

Key idea

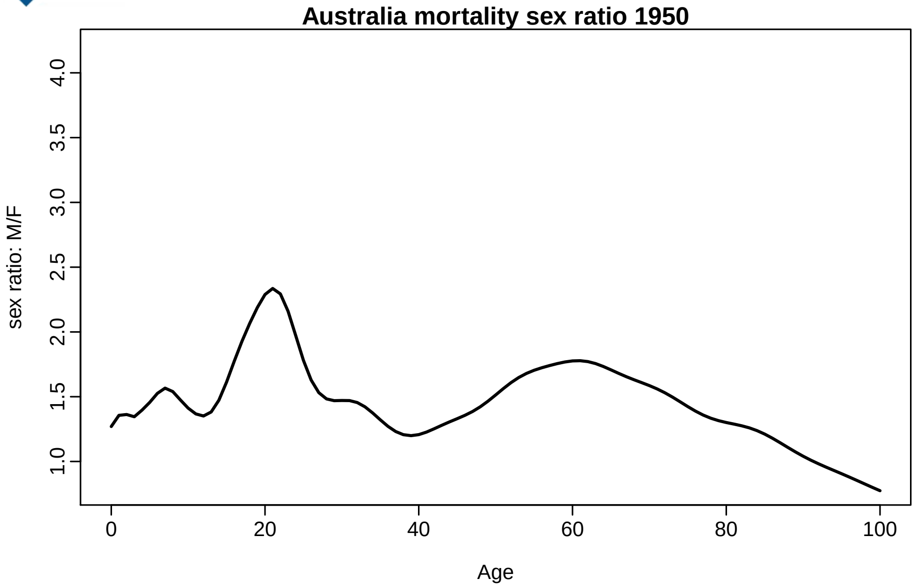
Model the geometric mean and the mortality ratio instead of the individual rates for each sex separately.

$$p_t(x) = \sqrt{f_{t,M}(x) f_{t,F}(x)} \quad \text{and} \quad r_t(x) = \sqrt{f_{t,M}(x) \big/ f_{t,F}(x)}.$$

- Product and ratio are approximately independent
- Ratio should be stationary (for coherence) but product can be non-stationary.

Mortality rates

Mortality rates



Model product and ratios

$$p_t(x) = \sqrt{f_{t,M}(x) f_{t,F}(x)} \quad \text{and} \quad r_t(x) = \sqrt{f_{t,M}(x) / f_{t,F}(x)}.$$

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x)$$

$$\log[r_t(x)] = \mu_r(x) + \sum_{\ell=1}^{L} \gamma_{t,\ell}\psi_\ell(x) + w_t(x).$$

- $\{\gamma_{t,\ell}\}$ restricted to be stationary processes: either ARMA(p,q) or ARFIMA(p,d,q).
- No restrictions for $\beta_{t,1},\dots,\beta_{t,K}$.
- **Forecasts:** $f_{n+h|n,M}(x) = p_{n+h|n}(x) r_{n+h|n}(x)$
  $f_{n+h|n,F}(x) = p_{n+h|n}(x) / r_{n+h|n}(x).$

Model product and ratios

$$p_t(x) = \sqrt{f_{t,M}(x) f_{t,F}(x)} \quad \text{and} \quad r_t(x) = \sqrt{f_{t,M}(x)/f_{t,F}(x)}.$$

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x)$$

$$\log[r_t(x)] = \mu_r(x) + \sum_{\ell=1}^{L} \gamma_{t,\ell}\psi_\ell(x) + w_t(x).$$

- $\{\gamma_{t,\ell}\}$ restricted to be stationary processes: either ARMA(p, q) or ARFIMA(p, d, q).
- No restrictions for $\beta_{t,1}, \dots, \beta_{t,K}$.
- **Forecasts:** $f_{n+h|n,M}(x) = p_{n+h|n}(x) r_{n+h|n}(x)$

  $f_{n+h|n,F}(x) = p_{n+h|n}(x) / r_{n+h|n}(x).$

Model product and ratios

$$p_t(x) = \sqrt{f_{t,M}(x) f_{t,F}(x)} \quad \text{and} \quad r_t(x) = \sqrt{f_{t,M}(x)/f_{t,F}(x)}.$$

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x)$$

$$\log[r_t(x)] = \mu_r(x) + \sum_{\ell=1}^{L} \gamma_{t,\ell}\psi_\ell(x) + w_t(x).$$

- $\{\gamma_{t,\ell}\}$ restricted to be stationary processes: either ARMA(p, q) or ARFIMA(p, d, q).
- No restrictions for $\beta_{t,1}, \dots, \beta_{t,K}$.
- **Forecasts:** $f_{n+h|n,M}(x) = p_{n+h|n}(x) r_{n+h|n}(x)$
  $f_{n+h|n,F}(x) = p_{n+h|n}(x) / r_{n+h|n}(x).$

Model product and ratios

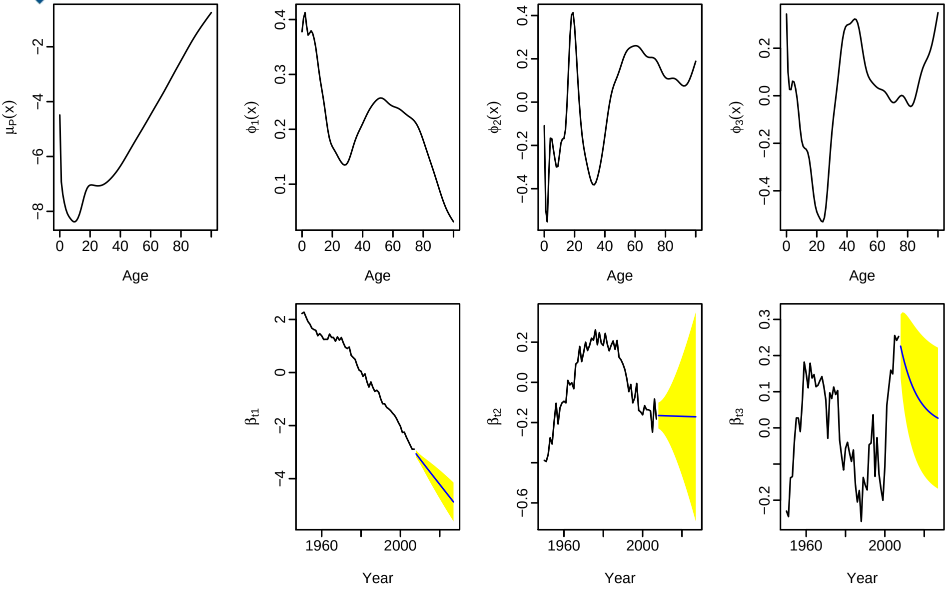
$$p_t(x) = \sqrt{f_{t,M}(x) f_{t,F}(x)} \quad \text{and} \quad r_t(x) = \sqrt{f_{t,M}(x) / f_{t,F}(x)}.$$

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x)$$

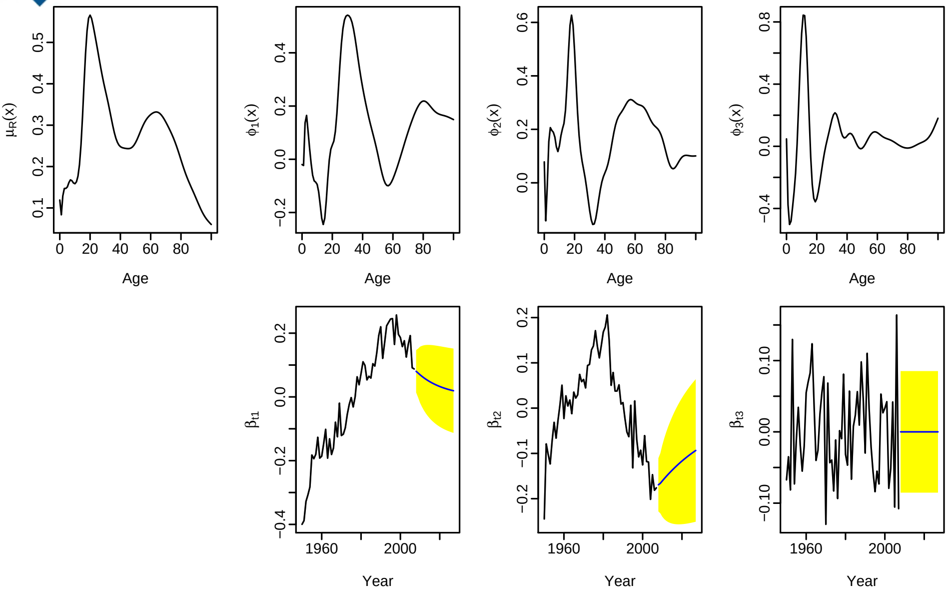
$$\log[r_t(x)] = \mu_r(x) + \sum_{\ell=1}^{L} \gamma_{t,\ell}\psi_\ell(x) + w_t(x).$$

- $\{\gamma_{t,\ell}\}$ restricted to be stationary processes: either ARMA(p, q) or ARFIMA(p, d, q).
- No restrictions for $\beta_{t,1}, \dots, \beta_{t,K}$.
- **Forecasts:** $f_{n+h|n,M}(x) = p_{n+h|n}(x) r_{n+h|n}(x)$
  $f_{n+h|n,F}(x) = p_{n+h|n}(x) / r_{n+h|n}(x).$

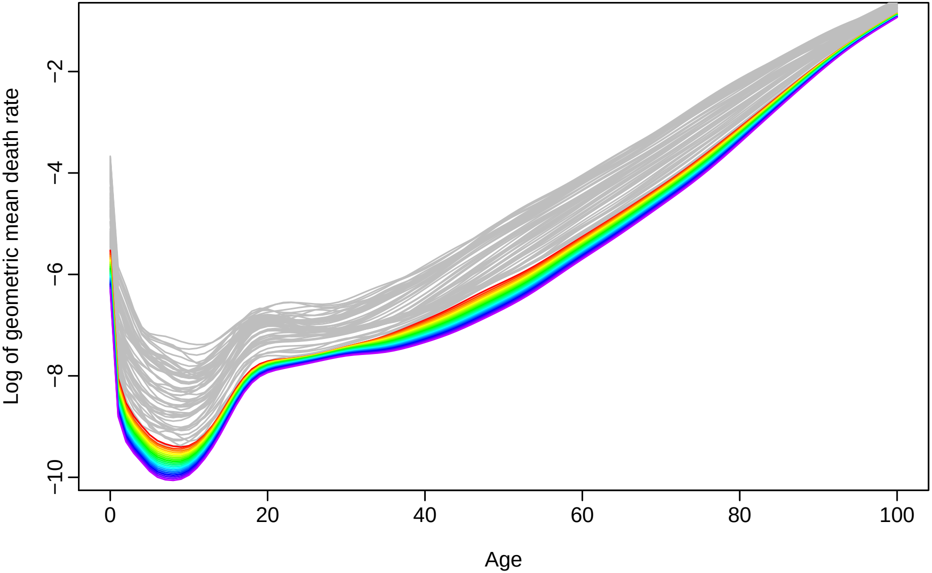
Product model



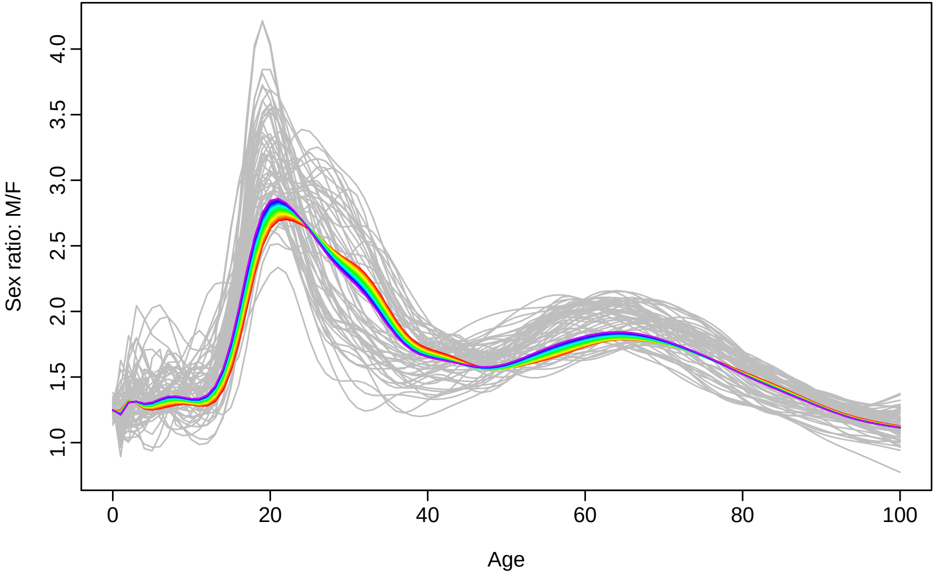
Ratio model



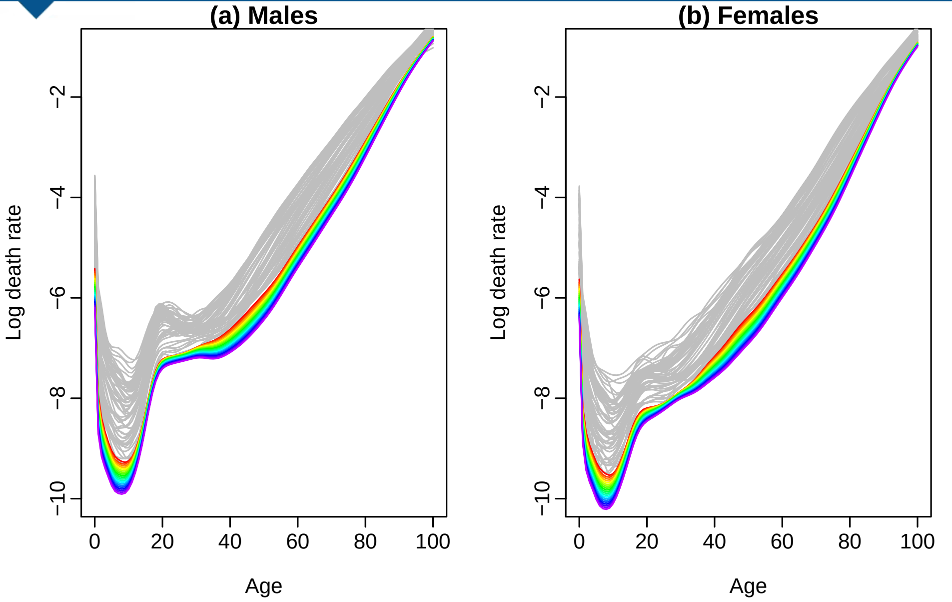
Product forecasts



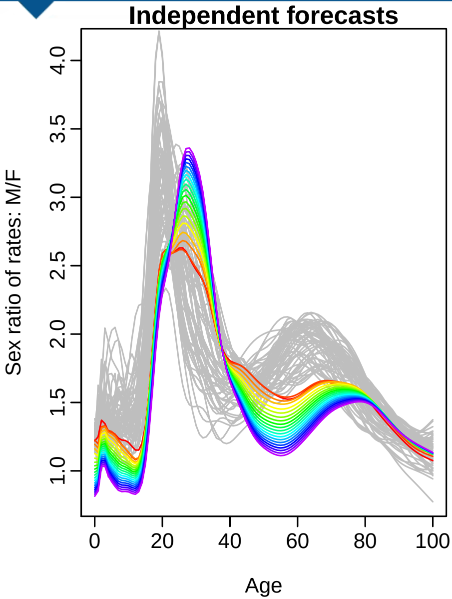
Ratio forecasts

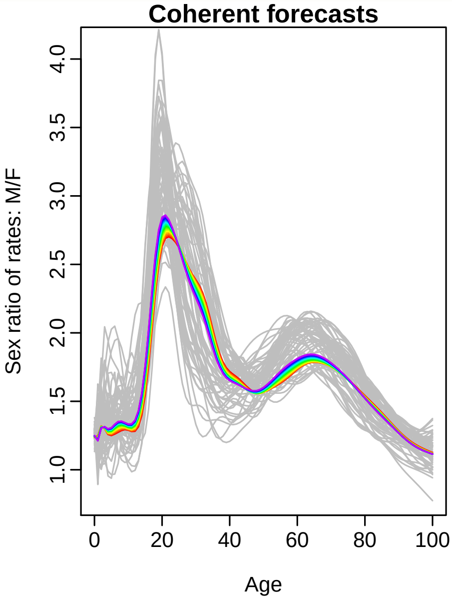


Coherent forecasts

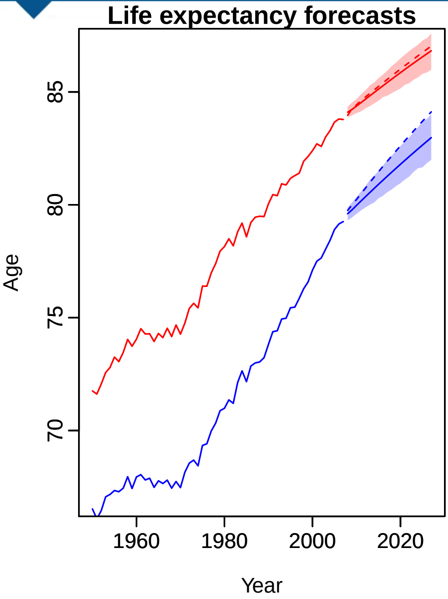


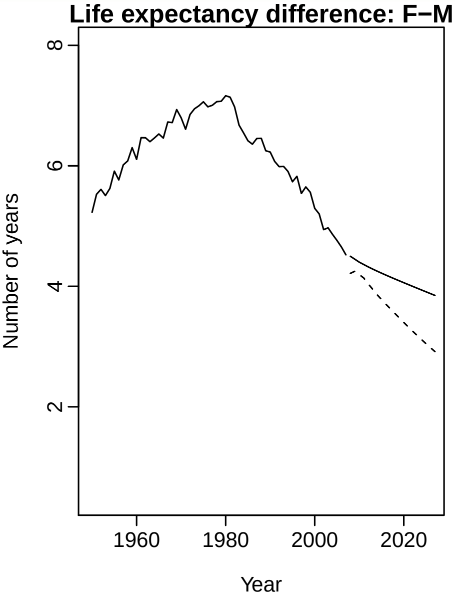
Ratio forecasts





Life expectancy forecasts





Coherent forecasts for J groups

$$p_t(x) = [f_{t,1}(x) f_{t,2}(x) \cdots f_{t,J}(x)]^{1/J}$$

$$\text{and} \qquad r_{t,j}(x) = f_{t,j}(x) \big/ p_t(x),$$

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^{K} \beta_{t,k} \phi_k(x) + e_t(x)$$

$$\log[r_{t,j}(x)] = \mu_{r,j}(x) + \sum_{l=1}^{L} \gamma_{t,l,j} \psi_{l,j}(x) + w_{t,j}(x).$$

- $p_t(x)$ and all $r_{t,j}(x)$ are approximately independent.
- Ratios satisfy constraint $r_{t,1}(x) r_{t,2}(x) \cdots r_{t,J}(x) = 1$.
- $\log[f_{t,j}(x)] = \log[p_t(x) r_{t,j}(x)]$

Coherent forecasts for J groups

$$p_t(x) = [f_{t,1}(x) f_{t,2}(x) \cdots f_{t,J}(x)]^{1/J}$$

and $$r_{t,j}(x) = f_{t,j}(x) \big/ p_t(x),$$

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x)$$

$$\log[r_{t,j}(x)] = \mu_{r,j}(x) + \sum_{l=1}^{L} \gamma_{t,l,j}\psi_{l,j}(x) + w_{t,j}(x).$$

- $p_t(x)$ and all $r_{t,j}(x)$ are approximately independent.
- Ratios satisfy constraint $r_{t,1}(x) r_{t,2}(x) \cdots r_{t,J}(x) = 1$.
- $\log[f_{t,j}(x)] = \log[p_t(x) r_{t,j}(x)]$

Coherent forecasts for J groups

$$p_t(x) = [f_{t,1}(x) f_{t,2}(x) \cdots f_{t,J}(x)]^{1/J}$$

and $$r_{t,j}(x) = f_{t,j}(x) / p_t(x),$$

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^{K} \beta_{t,k} \phi_k(x) + e_t(x)$$

$$\log[r_{t,j}(x)] = \mu_{r,j}(x) + \sum_{l=1}^{L} \gamma_{t,l,j} \psi_{l,j}(x) + w_{t,j}(x).$$

- $p_t(x)$ and all $r_{t,j}(x)$ are approximately independent.
- Ratios satisfy constraint $r_{t,1}(x) r_{t,2}(x) \cdots r_{t,J}(x) = 1$.
- $\log[f_{t,j}(x)] = \log[p_t(x) r_{t,j}(x)]$

Coherent forecasts for J groups

$$p_t(x) = [f_{t,1}(x) f_{t,2}(x) \cdots f_{t,J}(x)]^{1/J}$$

and $$r_{t,j}(x) = f_{t,j}(x) \big/ p_t(x),$$

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^{K} \beta_{t,k} \phi_k(x) + e_t(x)$$

$$\log[r_{t,j}(x)] = \mu_{r,j}(x) + \sum_{l=1}^{L} \gamma_{t,l,j} \psi_{l,j}(x) + w_{t,j}(x).$$

- $p_t(x)$ and all $r_{t,j}(x)$ are approximately independent.
- Ratios satisfy constraint $r_{t,1}(x) r_{t,2}(x) \cdots r_{t,J}(x) = 1$.
- $\log[f_{t,j}(x)] = \log[p_t(x) r_{t,j}(x)]$

Coherent forecasts for J groups

$$p_t(x) = [f_{t,1}(x) f_{t,2}(x) \cdots f_{t,J}(x)]^{1/J}$$

and $$r_{t,j}(x) = f_{t,j}(x) / p_t(x),$$

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x)$$

$$\log[r_{t,j}(x)] = \mu_{r,j}(x) + \sum_{l=1}^{L} \gamma_{t,l,j}\psi_{l,j}(x) + w_{t,j}(x).$$

- $p_t(x)$ and all $r_{t,j}(x)$ are approximately independent.
- Ratios satisfy constraint $r_{t,1}(x) r_{t,2}(x) \cdots r_{t,J}(x) = 1$.
- $\log[f_{t,j}(x)] = \log[p_t(x) r_{t,j}(x)]$

Coherent forecasts for J groups

$$
\begin{aligned}
\log[f_{t,j}(x)] &= \log[p_t(x) r_{t,j}(x)] = \log[p_t(x)] + \log[r_{t,j}] \\
&= \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,\ell,j}\psi_{\ell,j}(x) + z_{t,j}(x)
\end{aligned}
$$

- $\mu_j(x) = \mu_p(x) + \mu_{r,j}(x)$ is group mean
- $z_{t,j}(x) = e_t(x) + w_{t,j}(x)$ is error term.
- $\{\gamma_{t,\ell}\}$ restricted to be stationary processes: either ARMA(p,q) or ARFIMA(p,d,q).
- No restrictions for $\beta_{t,1},\dots,\beta_{t,K}$.

Coherent forecasts for J groups

$$\log[f_{t,j}(x)] = \log[p_t(x) r_{t,j}(x)] = \log[p_t(x)] + \log[r_{t,j}]$$
$$= \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,\ell,j}\psi_{\ell,j}(x) + z_{t,j}(x)$$

- $\mu_j(x) = \mu_p(x) + \mu_{r,j}(x)$ is group mean
- $z_{t,j}(x) = e_t(x) + w_{t,j}(x)$ is error term.
- $\{\gamma_{t,\ell}\}$ restricted to be stationary processes: either ARMA(p,q) or ARFIMA(p,d,q).
- No restrictions for $\beta_{t,1},\dots,\beta_{t,K}$.

Coherent forecasts for J groups

$$\log[f_{t,j}(x)] = \log[p_t(x) r_{t,j}(x)] = \log[p_t(x)] + \log[r_{t,j}]$$
$$= \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,\ell,j}\psi_{\ell,j}(x) + z_{t,j}(x)$$

- $\mu_j(x) = \mu_p(x) + \mu_{r,j}(x)$ is group mean
- $z_{t,j}(x) = e_t(x) + w_{t,j}(x)$ is error term.
- $\{\gamma_{t,\ell}\}$ restricted to be stationary processes: either ARMA(p, q) or ARFIMA(p, d, q).
- No restrictions for $\beta_{t,1}, \dots, \beta_{t,K}$.

Coherent forecasts for J groups

$$
\begin{aligned}
\log[f_{t,j}(x)] = \log[p_t(x) r_{t,j}(x)] &= \log[p_t(x)] + \log[r_{t,j}] \\
&= \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,\ell,j}\psi_{\ell,j}(x) + z_{t,j}(x)
\end{aligned}
$$

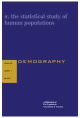
- $\mu_j(x) = \mu_p(x) + \mu_{r,j}(x)$ is group mean
- $z_{t,j}(x) = e_t(x) + w_{t,j}(x)$ is error term.
- $\{\gamma_{t,\ell}\}$ restricted to be stationary processes: either ARMA(p, q) or ARFIMA(p, d, q).
- No restrictions for $\beta_{t,1}, \ldots, \beta_{t,K}$.

Coherent forecasts for J groups

$$\log[f_{t,j}(x)] = \log[p_t(x) r_{t,j}(x)] = \log[p_t(x)] + \log[r_{t,j}]$$
$$= \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,\ell,j}\psi_{\ell,j}(x) + z_{t,j}(x)$$

- $\mu_j(x) = \mu_p(x) + \mu_{r,j}(x)$ is group mean
- $z_{t,j}(x) = e_t(x) + w_{t,j}(x)$ is error term.
- $\{\gamma_{t,\ell}\}$ restricted to be stationary processes: either ARMA(p, q) or ARFIMA(p, d, q).
- No restrictions for $\beta_{t,1}, \dots, \beta_{t,K}$.

References



Hyndman, Booth, Yasmeen (2012). “Coherent mortality forecasting: the product-ratio method with functional time series models”. *Demography*, to appear

Hyndman (2012). *demography: Forecasting mortality, fertility, migration and population data*. `cran.r-project.org/package=demography`

For further information

robjhyndman.com

- Slides and references for this talk.
- Links to all papers and books.
- Links to R packages.
- A blog about forecasting research.