MONASH University

Rob J Hyndman

Functional time series

with applications in demography

2. Automatic time series forecasting

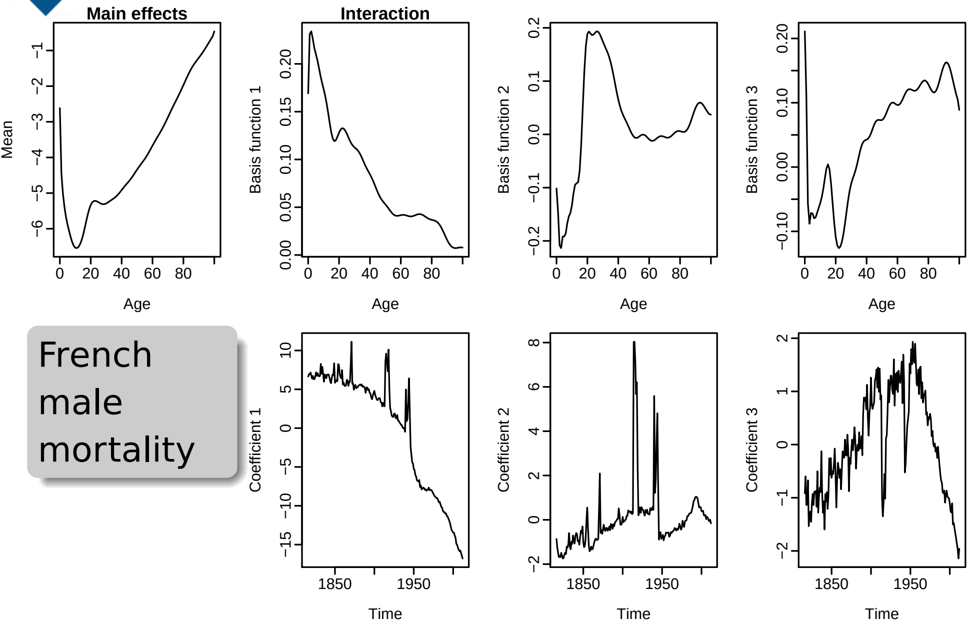
Outline

Functional principal components

$$y_t(x_i) = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i},$$
$$s_t(x) = \mu(x) + \sum_{k=1}^{T-1} \beta_{t,k}\,\phi_k(x)$$

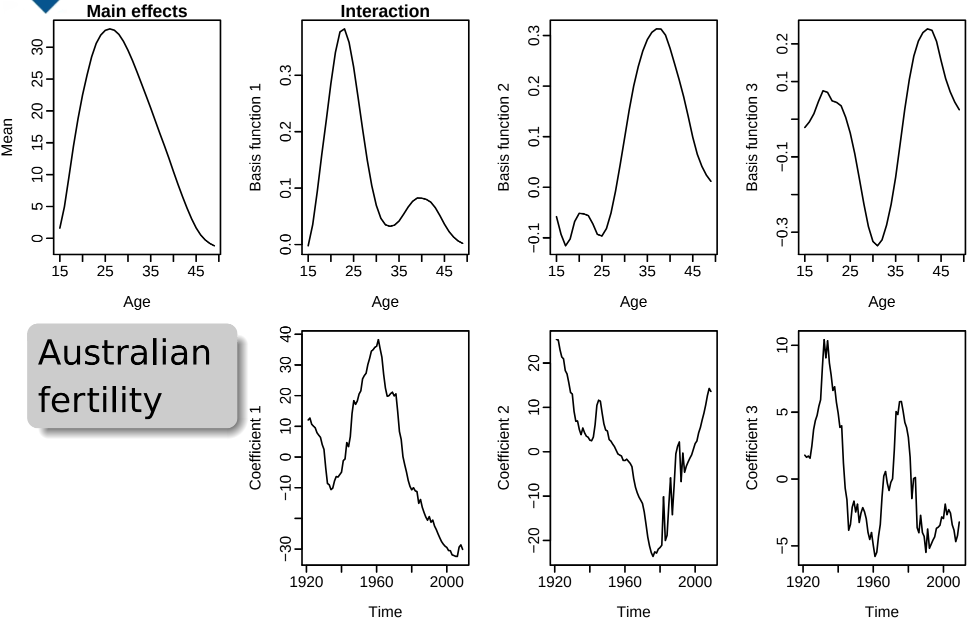
1. Estimate smooth functions $s_t(x)$ using weighted penalized regression splines.
2. Compute $\mu(x)$ as $\bar{s}(x)$ across years.
3. Compute $\beta_{t,k}$ and $\phi_k(x)$ using functional principal components.
4. To forecast $y_t(x_i)$, we need forecasts of $\{\beta_{t,k}\}$.

Functional principal components



French male mortality

Functional principal components



Australian fertility

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

		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

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

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

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

		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

- 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

		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

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

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

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

Trend **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

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

ETS(A,N,N)

SES with additive errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t$
Observation equation	$y_t = \ell_{t-1} + \varepsilon_t$
State equation	$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1}$
- “innovations” or “single source of error” because same error process, ε_t.
- Observation equation: relationship between observations and states.
- State equation: evolution of the state through time.

ETS(A,N,N)

SES with additive errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t$
Observation equation	$y_t = \ell_{t-1} + \varepsilon_t$
State equation	$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1}$
- "innovations" or "single source of error" because same error process, ε_t.
- Observation equation: relationship between observations and states.
- State equation: evolution of the state through time.

ETS(A,N,N)

SES with additive errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t$
Observation equation	$y_t = \ell_{t-1} + \varepsilon_t$
State equation	$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1}$
- "innovations" or "single source of error" because same error process, ε_t.
- Observation equation: relationship between observations and states.
- State equation: evolution of the state through time.

ETS(A,N,N)

SES with additive errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t$
Observation equation	$y_t = \ell_{t-1} + \varepsilon_t$
State equation	$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1}$
- "innovations" or "single source of error" because same error process, ε_t.
- Observation equation: relationship between observations and states.
- State equation: evolution of the state through time.

ETS(M,N,N)

SES with multiplicative errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t$
Observation equation	$y_t = \ell_{t-1}(1 + \varepsilon_t)$
State equation	$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

- Relative forecast errors: $\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}}$
- Models with additive and multiplicative errors with the same parameters generate the same point forecasts but different prediction intervals.

ETS(M,N,N)

SES with multiplicative errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t$
Observation equation	$y_t = \ell_{t-1}(1 + \varepsilon_t)$
State equation	$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

- Relative forecast errors: $\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}}$
- Models with additive and multiplicative errors with the same parameters generate the same point forecasts but different prediction intervals.

ETS(M,N,N)

SES with multiplicative errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t$
Observation equation	$y_t = \ell_{t-1}(1 + \varepsilon_t)$
State equation	$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

- Relative forecast errors: $\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}}$
- Models with additive and multiplicative errors with the same parameters generate the same point forecasts but different prediction intervals.

ETS(A,A,N)

Holt's linear method with additive errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t + hb_t$
Observation equation	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$
State equations	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$
	$b_t = b_{t-1} + \beta\varepsilon_t$

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1}$

ETS(A,A,A)

Holt-Winters additive method with additive errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$
Observation equation	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$
State equations	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$
	$b_t = b_{t-1} + \beta\varepsilon_t$
	$s_t = s_{t-m} + \gamma\varepsilon_t$

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1}$
- $h_m^+ = \lfloor (h-1) \mod m \rfloor + 1$.

Additive error models

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
$\mathbf{A_d}$	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
$\mathbf{M_d}$	$y_t = \ell_{t-1}b_{t-1}^{\phi} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1}^{\phi} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^{\phi}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^{\phi})$

Multiplicative error models

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}(1+\alpha\varepsilon_t)$	$y_t = (\ell_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}+\alpha(\ell_{t-1}+s_{t-m})\varepsilon_t$ $s_t = s_{t-m}+\gamma(\ell_{t-1}+s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}(1+\alpha\varepsilon_t)$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1}+b_{t-1})(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1}+b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}+\beta(\ell_{t-1}+b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1}+b_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}+b_{t-1}+\alpha(\ell_{t-1}+b_{t-1}+s_{t-m})\varepsilon_t$ $b_t = b_{t-1}+\beta(\ell_{t-1}+b_{t-1}+s_{t-m})\varepsilon_t$ $s_t = s_{t-m}+\gamma(\ell_{t-1}+b_{t-1}+s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1}+b_{t-1})s_{t-m}(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1}+b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}+\beta(\ell_{t-1}+b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
A$_\mathbf{d}$	$y_t = (\ell_{t-1}+\phi b_{t-1})(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1}+\phi b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1}+\phi b_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}+\phi b_{t-1}+\alpha(\ell_{t-1}+\phi b_{t-1}+s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1}+s_{t-m})\varepsilon_t$ $s_t = s_{t-m}+\gamma(\ell_{t-1}+\phi b_{t-1}+s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1}+\phi b_{t-1})s_{t-m}(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1}+\phi b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
M	$y_t = \ell_{t-1}b_{t-1}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}(1+\beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}+\alpha(\ell_{t-1}b_{t-1}+s_{t-m})\varepsilon_t$ $b_t = b_{t-1}+\beta(\ell_{t-1}b_{t-1}+s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m}+\gamma(\ell_{t-1}b_{t-1}+s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}(1+\beta\varepsilon_t)$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
M$_\mathbf{d}$	$y_t = \ell_{t-1}b_{t-1}^{\phi}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1+\beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}+\alpha(\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^{\phi}+\beta(\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m}+\gamma(\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^{\phi}s_{t-m}(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1+\beta\varepsilon_t)$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$

Innovations state space models

Let $\boldsymbol{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})$ and $\varepsilon_t \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$.

$$y_t = \underbrace{h(\boldsymbol{x}_{t-1})}_{\hat{y}_{t|t-1}} + \underbrace{k(\boldsymbol{x}_{t-1})\varepsilon_t}_{e_t}$$ Observation equation

$$\boldsymbol{x}_t = f(\boldsymbol{x}_{t-1}) + g(\boldsymbol{x}_{t-1})\varepsilon_t$$ State equation

Additive errors:

$$k(\boldsymbol{x}_{t-1}) = 1. \qquad y_t = \hat{y}_{t|t-1} + \varepsilon_t.$$

Multiplicative errors:

$$k(\boldsymbol{x}_{t-1}) = \hat{y}_{t|t-1}. \qquad y_t = \hat{y}_{t|t-1}(1 + \varepsilon_t).$$

$\varepsilon_t = (y_t - \hat{y}_{t|t-1})/\hat{y}_{t|t-1}$ is relative error.

Some unstable models

- Some of the combinations of (Error, Trend, Seasonal) can lead to numerical difficulties; see equations with division by a state.
- These are: ETS(M,M,A), ETS(M,M_d,A), ETS(A,N,M), ETS(A,A,M), ETS(A,A_d,M), ETS(A,M,N), ETS(A,M,A), ETS(A,M,M), ETS(A,M_d,N), ETS(A,M_d,A), and ETS(A,M_d,M).

Some unstable models

- Some of the combinations of (Error, Trend, Seasonal) can lead to numerical difficulties; see equations with division by a state.
- These are: ETS(M,M,A), ETS(M,M_d,A), ETS(A,N,M), ETS(A,A,M), ETS(A,A_d,M), ETS(A,M,N), ETS(A,M,A), ETS(A,M,M), ETS(A,M_d,N), ETS(A,M_d,A), and ETS(A,M_d,M).

Exponential smoothing models

Additive Error

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

Multiplicative Error

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

Estimation and model selection

Estimation

$$L^*(\theta, \boldsymbol{x}_0) = T \log\left(\sum_{t=1}^{T} \varepsilon_t^2 / k^2(\boldsymbol{x}_{t-1})\right) + 2\sum_{t=1}^{T} \log|k(\boldsymbol{x}_{t-1})|$$
$$= -2\log(\text{Likelihood}) + \text{constant}$$

- Minimize wrt $\theta = (\alpha, \beta, \gamma, \phi)$ and initial states $\boldsymbol{x}_0 = (\ell_0, b_0, s_0, s_{-1}, \dots, s_{-m+1})$.

Model selection

- Select amongst all models using AIC (or similar).

Estimation and model selection

Estimation

$$L^*(\boldsymbol{\theta},\boldsymbol{x}_0) = T\log\left(\sum_{t=1}^{T}\varepsilon_t^2/k^2(\boldsymbol{x}_{t-1})\right) + 2\sum_{t=1}^{T}\log|k(\boldsymbol{x}_{t-1})|$$
$$= -2\log(\text{Likelihood}) + \text{constant}$$

- Minimize wrt $\boldsymbol{\theta} = (\alpha,\beta,\gamma,\phi)$ and initial states $\boldsymbol{x}_0 = (\ell_0, b_0, s_0, s_{-1}, \dots, s_{-m+1})$.

Model selection

- Select amongst all models using AIC (or similar).

Estimation and model selection

Estimation

$$L^*(\boldsymbol{\theta}, \boldsymbol{x}_0) = T \log\left(\sum_{t=1}^{T} \varepsilon_t^2 / k^2(\boldsymbol{x}_{t-1})\right) + 2\sum_{t=1}^{T} \log|k(\boldsymbol{x}_{t-1})|$$
$$= -2\log(\text{Likelihood}) + \text{constant}$$

- Minimize wrt $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \phi)$ and initial states $\boldsymbol{x}_0 = (\ell_0, b_0, s_0, s_{-1}, \dots, s_{-m+1})$.

Model selection

- Select amongst all models using AIC (or similar).

Automatic forecasting

From Hyndman et al. (IJF, 2002):

- Apply each model that is 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 model that is 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 model that is 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 model that is 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 model that is 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.

Forecasting with ETS models

- Point forecasts obtained by iterating equations for $t = T+1, \dots, T+h$, setting $\varepsilon_t = 0$ for $t > T$.
- Not the same as $\text{E}(y_{t+h} | y_1, \dots, y_t)$ unless trend and seasonality are both additive.
- Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.
- Prediction intervals will differ between models with additive and multiplicative methods.
- Exact PI available for many models.
- Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.

Forecasting with ETS models

- Point forecasts obtained by iterating equations for $t = T + 1, \ldots, T + h$, setting $\varepsilon_t = 0$ for $t > T$.
- Not the same as $\mathrm{E}(y_{t+h}|y_1, \ldots, y_t)$ unless trend and seasonality are both additive.
- Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.
- Prediction intervals will differ between models with additive and multiplicative methods.
- Exact PI available for many models.
- Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.

Forecasting with ETS models

- Point forecasts obtained by iterating equations for $t = T+1, \dots, T+h$, setting $\varepsilon_t = 0$ for $t > T$.
- Not the same as $\mathrm{E}(y_{t+h}|y_1, \dots, y_t)$ unless trend and seasonality are both additive.
- Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.
- Prediction intervals will differ between models with additive and multiplicative methods.
- Exact PI available for many models.
- Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.

Forecasting with ETS models

- Point forecasts obtained by iterating equations for $t = T+1, \dots, T+h$, setting $\varepsilon_t = 0$ for $t > T$.
- Not the same as $\mathrm{E}(y_{t+h}|y_1, \dots, y_t)$ unless trend and seasonality are both additive.
- Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.
- Prediction intervals will differ between models with additive and multiplicative methods.
- Exact PI available for many models.
- Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.

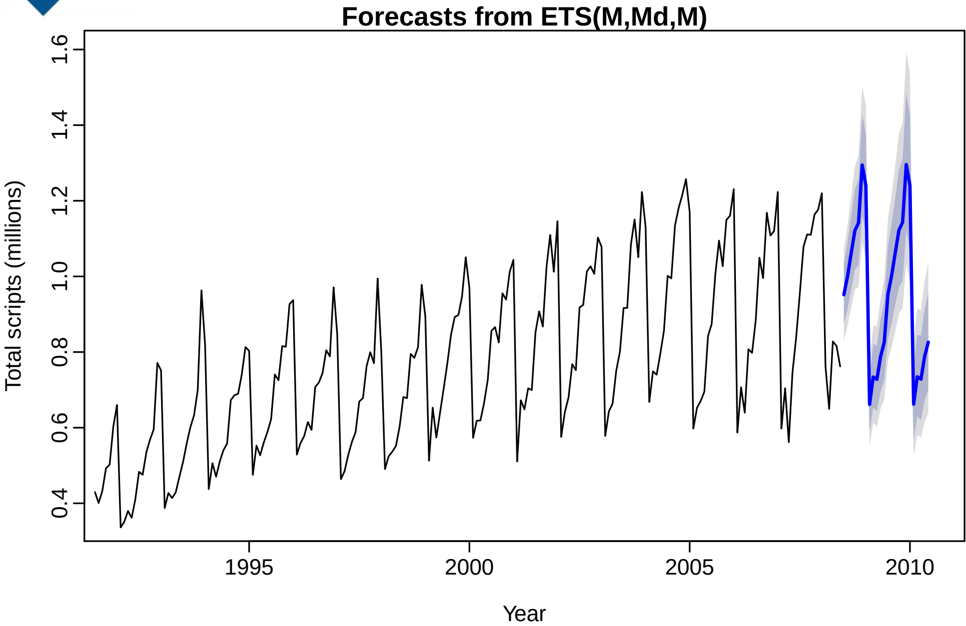
Forecasting with ETS models

- Point forecasts obtained by iterating equations for $t = T + 1, \dots, T + h$, setting $\varepsilon_t = 0$ for $t > T$.
- Not the same as $\mathrm{E}(y_{t+h}|y_1, \dots, y_t)$ unless trend and seasonality are both additive.
- Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.
- Prediction intervals will differ between models with additive and multiplicative methods.
- Exact PI available for many models.
- Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.

Forecasting with ETS models

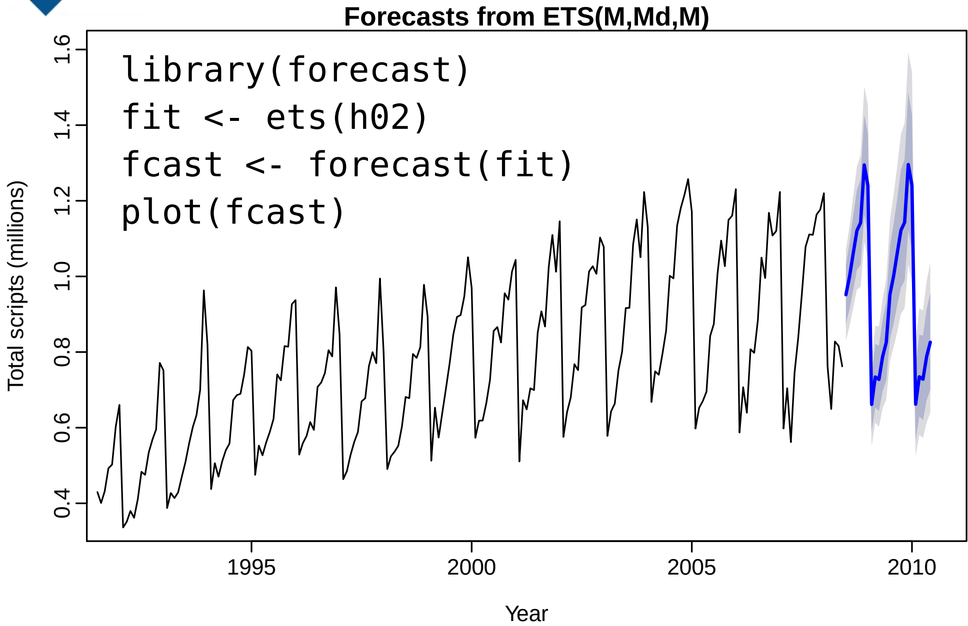
- Point forecasts obtained by iterating equations for $t = T + 1, \dots, T + h$, setting $\varepsilon_t = 0$ for $t > T$.
- Not the same as $\mathrm{E}(y_{t+h}|y_1, \dots, y_t)$ unless trend and seasonality are both additive.
- Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.
- Prediction intervals will differ between models with additive and multiplicative methods.
- Exact PI available for many models.
- Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.

Exponential smoothing



Exponential smoothing

```
library(forecast)
fit <- ets(h02)
fcast <- forecast(fit)
plot(fcast)
```



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
```

Outline

ARIMA modelling

Conventional ARIMA forecasting

- calculate forecasts from the best fitting ARIMA model
- Not necessarily the best forecasting ARIMA model.
- Model identification subjective and complex.

ARIMA modelling

Conventional ARIMA forecasting

- calculate forecasts from the best fitting ARIMA model
- Not necessarily the best forecasting ARIMA model.
- Model identification subjective and complex.

ARIMA modelling

Conventional ARIMA forecasting

- calculate forecasts from the best fitting ARIMA model
- Not necessarily the best forecasting ARIMA model.
- Model identification subjective and complex.

Non-seasonal ARIMA model

$$y_t \sim \text{ARIMA}(p, d, q)$$
$$y'_t = (1 - B)^d y_t$$
$$y'_t = c + \sum_{j=1}^{p} \phi_j y'_{t-j} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

p AR parameters: $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$
q MA parameters: $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$
d is the differencing order

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

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.

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?

$$\text{AICc} = -2\log(L) + 2(p+q+k+1)\left(\frac{T}{T-p-q-k-2}\right).$$

where L is the maximised likelihood fitted to the *differenced* data, $k = 1$ if $c \neq 0$ and $k = 0$ otherwise.

Step 1: Select current model (with smallest AIC) from:
ARIMA$(2, d, 2)$
ARIMA$(0, d, 0)$
ARIMA$(1, d, 0)$
ARIMA$(0, d, 1)$

Step 2: Consider variations of current model:
- vary one of $p, q,$ from current model by ± 1
- p, q both vary from current model by ± 1
- Include/exclude c from current model

Model with lowest AIC becomes current model.

Repeat Step 2 until no lower AIC can be found.

How does auto.arima() work?

$$\text{AICc} = -2\log(L) + 2(p+q+k+1)\left(\frac{T}{T-p-q-k-2}\right).$$

where L is the maximised likelihood fitted to the *differenced* data, $k = 1$ if $c \neq 0$ and $k = 0$ otherwise.

Step 1: Select current model (with smallest AIC) from:
ARIMA$(2, d, 2)$
ARIMA$(0, d, 0)$
ARIMA$(1, d, 0)$
ARIMA$(0, d, 1)$

Step 2: Consider variations of current model:
- vary one of p, q, from current model by ± 1
- p, q both vary from current model by ± 1
- Include/exclude c from current model

Model with lowest AIC becomes current model.

Repeat Step 2 until no lower AIC can be found.

How does auto.arima() work?

$$\text{AICc} = -2\log(L) + 2(p+q+k+1)\left(\frac{T}{T-p-q-k-2}\right).$$

where L is the maximised likelihood fitted to the *differenced* data, $k=1$ if $c \neq 0$ and $k=0$ otherwise.

Step 1: Select current model (with smallest AIC) from:
ARIMA$(2,d,2)$
ARIMA$(0,d,0)$
ARIMA$(1,d,0)$
ARIMA$(0,d,1)$

Step 2: Consider variations of current model:
- vary one of $p, q,$ from current model by ± 1
- p, q both vary from current model by ± 1
- Include/exclude c from current model

Model with lowest AIC becomes current model.

Repeat Step 2 until no lower AIC can be found.

How does auto.arima() work?

$$\text{AIC}c = -2\log(L) + 2(p+q+k+1)\left(\frac{T}{T-p-q-k-2}\right).$$

where L is the maximised likelihood fitted to the *differenced* data, $k = 1$ if $c \neq 0$ and $k = 0$ otherwise.

Step 1: Select current model (with smallest AIC) from:
ARIMA$(2, d, 2)$
ARIMA$(0, d, 0)$
ARIMA$(1, d, 0)$
ARIMA$(0, d, 1)$

Step 2: Consider variations of current model:
- vary one of p, q, from current model by ± 1
- p, q both vary from current model by ± 1
- Include/exclude c from current model

Model with lowest AIC becomes current model.

Repeat Step 2 until no lower AIC can be found.

How does auto.arima() work?

A seasonal ARIMA process

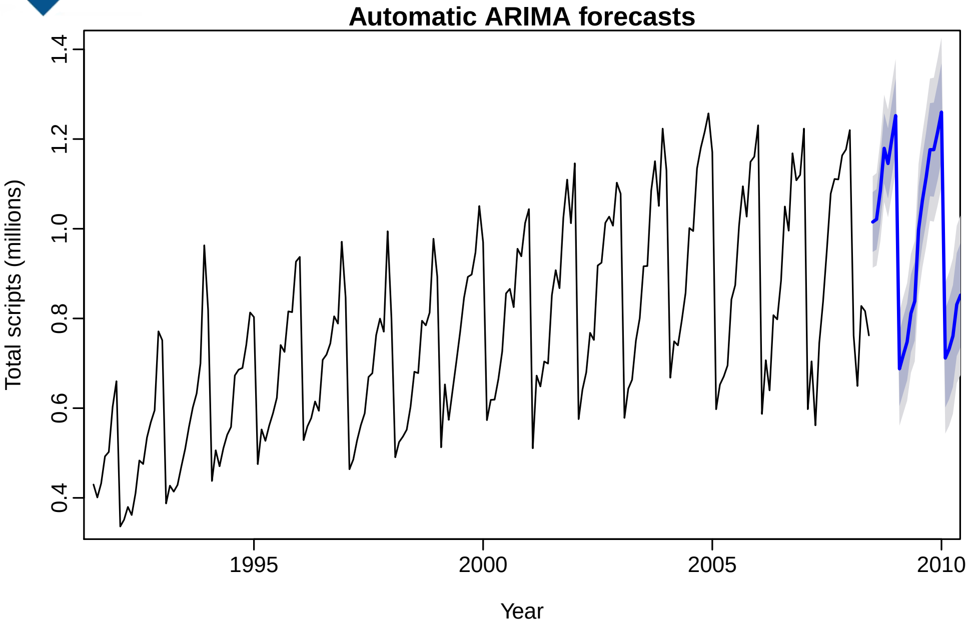
$$\Phi(B^m)\phi(B)(1-B)^d(1-B^m)^D y_t = c + \Theta(B^m)\theta(B)\varepsilon_t$$

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

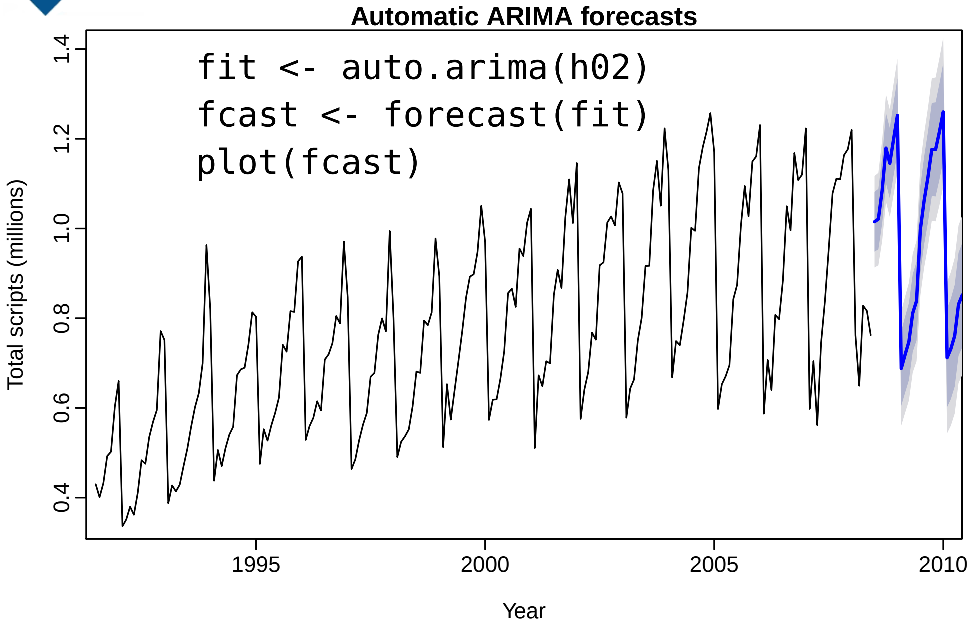
Hyndman & Khandakar (JSS, 2008) algorithm:

- 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.

Automatic seasonal ARIMA



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
1995
2000
2005
2010
Year

Auto ARIMA

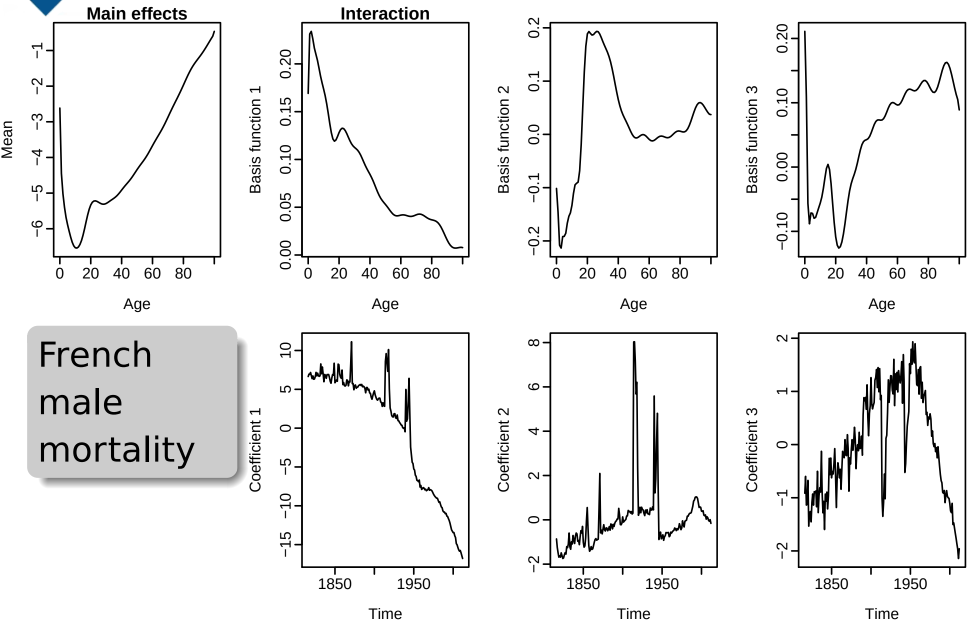
```
> fit
Series: h02
ARIMA(3,1,3)(0,1,1)[12]

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
```

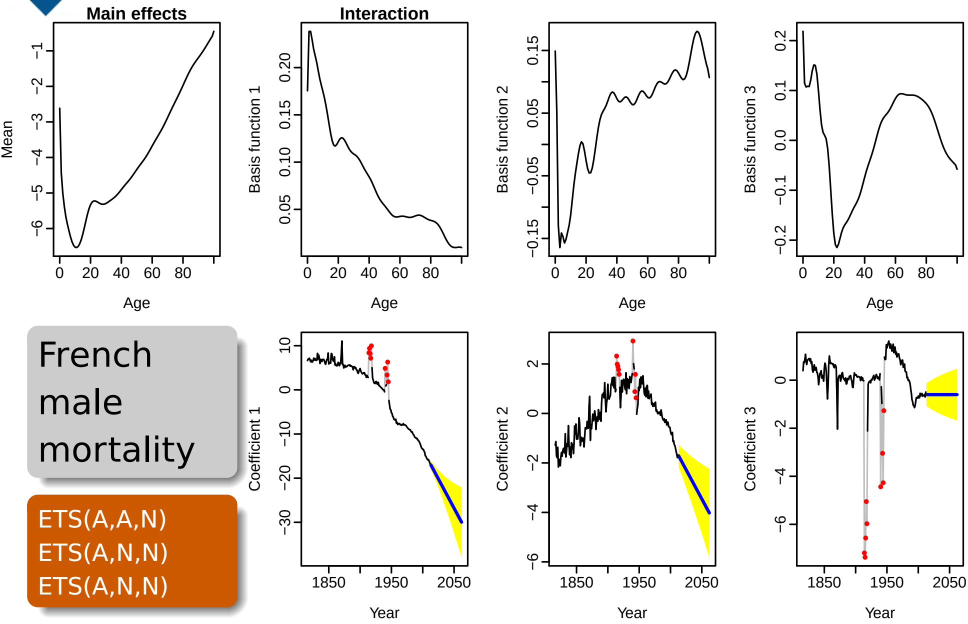
Outline

Functional principal components



French male mortality

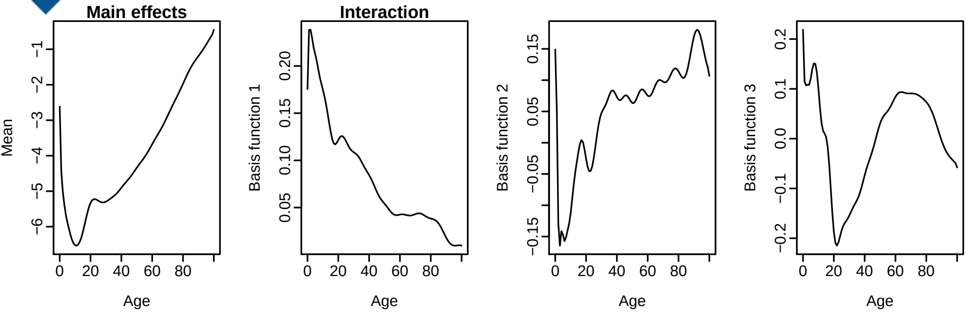
Functional principal components



French male mortality

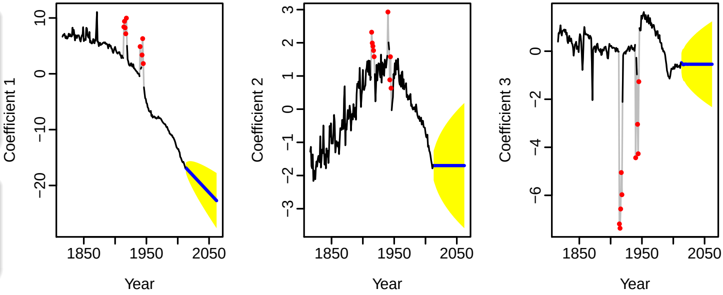
ETS(A,A,N)
ETS(A,N,N)
ETS(A,N,N)

Functional principal components

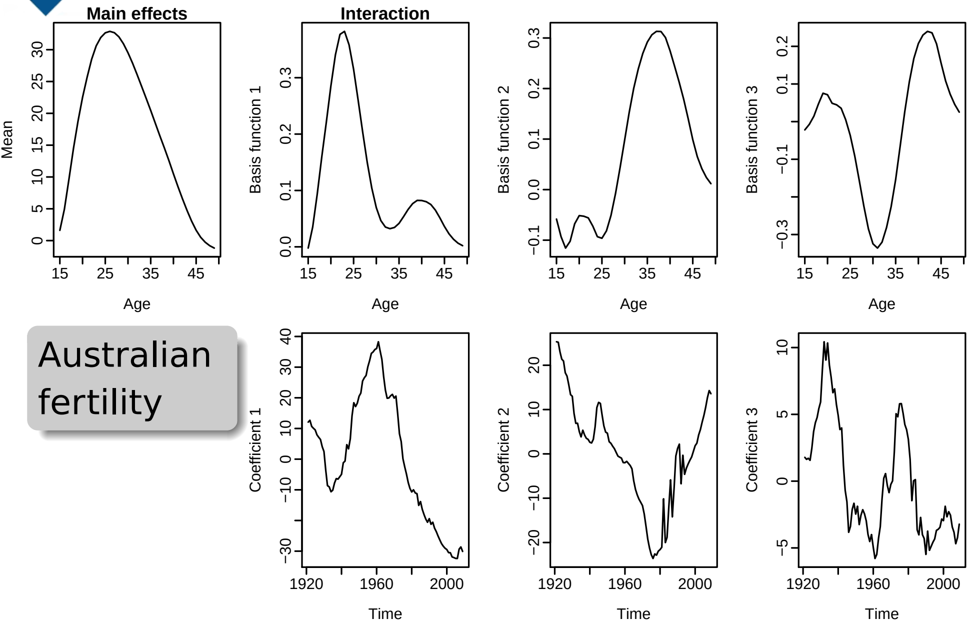


French male mortality

ARIMA(1,1,1)+c
ARIMA(1,1,1)
ARIMA(0,1,1)

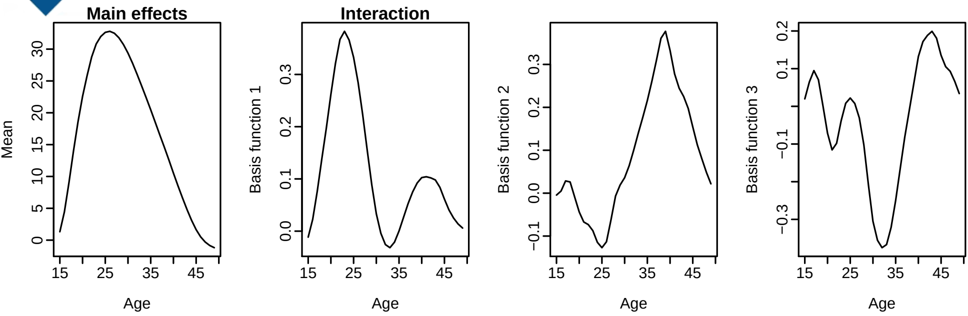


Functional principal components



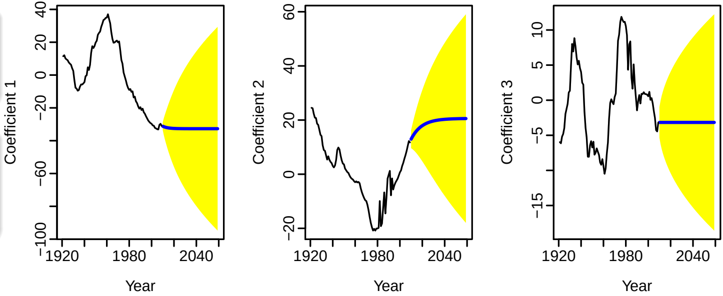
Australian fertility

Functional principal components

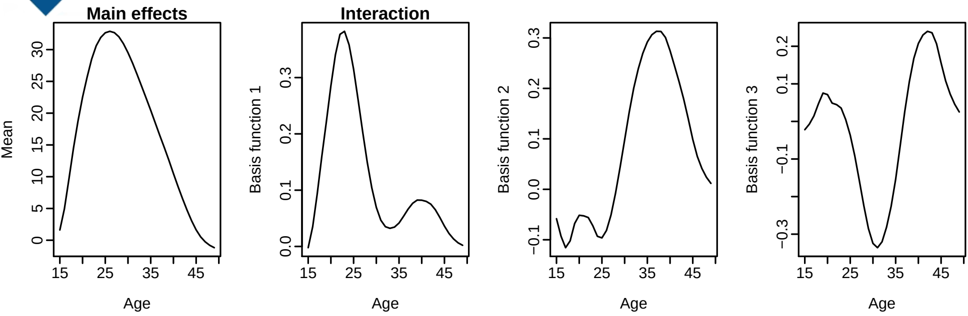


Australian fertility

ETS(A,A_d,N)
ETS(A,A_d,N)
ETS(A,A_d,N)

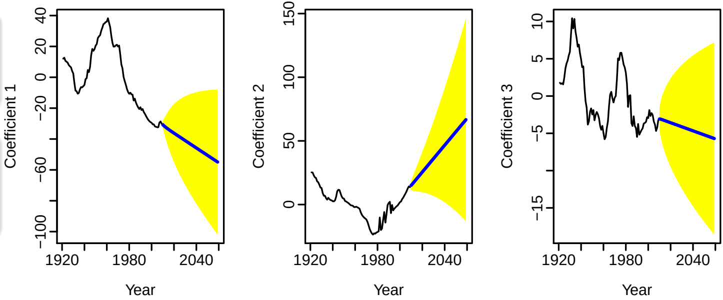


Functional principal components



Australian fertility

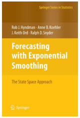
ARIMA(1,1,1)+c
ARIMA(5,2,2)
ARIMA(2,1,0)



Outline

Selected 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, Khandakar (2008). “Automatic time series forecasting : the forecast package for R”. . *Journal of Statistical Software* **26**(3), 1–22.

Hyndman (2014). *forecast: Forecasting functions for time series and linear models*. `cran.r-project.org/package=forecast`