

Rob J Hyndman

Forecasting: Principles and Practice



9. State space models

Outline

- 1 Recall ETS models
- 2 Simple structural models
- 3 Linear Gaussian state space models
- 4 Kalman filter
- **5** ARIMA models in state space form
- **6** Kalman smoothing
- 7 Time varying parameter models

		Seasonal Component		
	Trend	N	Α	M
	Component	(None)	(Additive)	(Multiplicative)
N	(None)	N,N	N,A	N,M
Α	(Additive)	A,N	A,A	A,M
A_{d}	(Additive damped)	A_d , N	A_d , A	A_d , M
М	(Multiplicative)	M,N	M,A	M,M
M_{d}	(Multiplicative damped)	M_d,N	M_d ,A	M _d ,M

General notation ETS: ExponenTial Smoothing

		Seasonal Component		
Trend		N	Α	M
	Component	(None)	(Additive)	(Multiplicative)
N	(None)	N,N	N,A	N,M
Α	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A _d ,N	A_d , A	A_d , M
М	(Multiplicative)	M,N	M,A	M,M
M_{d}	(Multiplicative damped)	M _d ,N	M_d ,A	M_d , M

General notation ETS: ExponenTial Smoothing

		Seasonal Component		
	Trend		Α	М
	Component	(None)	(Additive)	(Multiplicative)
N	(None)	N,N	N,A	N,M
Α	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A _d ,N	A_d , A	A_d , M
М	(Multiplicative)	M,N	M,A	M,M
\mathbf{M}_{d}	(Multiplicative damped)	M _d ,N	M_d ,A	M_d , M

General notation E T S : Exponen Tial Smoothing

Trend

Examples:

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

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

		Seasonal Component		
Trend		N	Α	М
	Component	(None)	(Additive)	(Multiplicative)
N	(None)	N,N	N,A	N,M
Α	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d , N	A_d , A	A _d ,M
М	(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

		Seasonal Component		
Trend		N	Α	M
	Component	(None)	(Additive)	(Multiplicative)
N	(None)	N,N	N,A	N,M
Α	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A _d ,N	A_d , A	A_d , M
М	(Multiplicative)	M,N	M,A	M,M
M_{d}	(Multiplicative damped)	M _d ,N	M_d ,A	M_d , M

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

Error Trend Seasonal

Examples:

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

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

		Seasonal Component		
Trend		N	Α	М
	Component	(None)	(Additive)	(Multiplicative)
N	(None)	N,N	N,A	N,M
Α	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A _d ,N	A_d , A	A_d , M
М	(Multiplicative)	M,N	M,A	M,M
\mathbf{M}_{d}	(Multiplicative damped)	M _d ,N	M_d ,A	M_d , M

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

Error Trend Seasonal

Examples:

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

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

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 Trend Seasonal

Examples:

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

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

Innovations state space models

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

$$y_{t} = \underbrace{h(\mathbf{x}_{t-1})}_{\mu_{t}} + \underbrace{k(\mathbf{x}_{t-1})\varepsilon_{t}}_{e_{t}}$$
$$\mathbf{x}_{t} = f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\varepsilon_{t}$$

Additive errors:

$$k(x) = 1.$$
 $y_t = \mu_t + \varepsilon_t.$

Multiplicative errors:

$$k(\mathbf{x}_{t-1}) = \mu_t.$$
 $y_t = \mu_t(1 + \varepsilon_t).$ $\varepsilon_t = (y_t - \mu_t)/\mu_t$ is relative error.

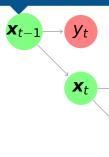
Outline

- 1 Recall ETS models
- 2 Simple structural models
- 3 Linear Gaussian state space models
- 4 Kalman filter
- **5** ARIMA models in state space form
- **6** Kalman smoothing
- 7 Time varying parameter models

State space models

 y_{t+1}

 \mathbf{x}_{t+1}



ETS state vector

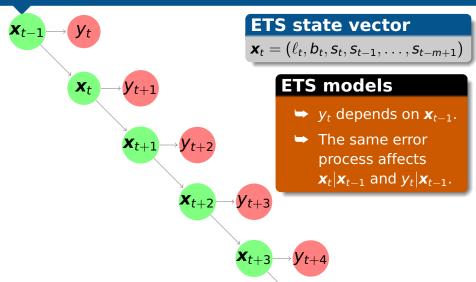
$$\mathbf{x}_{t} = (\ell_{t}, b_{t}, s_{t}, s_{t-1}, \dots, s_{t-m+1})$$

$$\mathbf{x}_{t+2} \longrightarrow \mathbf{y}_{t+3}$$

 y_{t+2}

$$x_{t+3} \longrightarrow y_{t+4}$$

State space models



State space models



$$\mathbf{x}_{t+1} \longrightarrow \mathbf{y}_{t+1}$$

$$\mathbf{x}_{t+2} \longrightarrow \mathbf{y}_{t+2}$$

ETS state vector

$$oldsymbol{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})$$

Structural models

- \rightarrow y_t depends on \mathbf{x}_t .
- A different error process affects $\mathbf{x}_t | \mathbf{x}_{t-1}$ and $\mathbf{y}_t | \mathbf{x}_t$.

 $y_{t+3} \longrightarrow y_{t+3}$

 $\mathbf{x}_{t+4} \longrightarrow \mathbf{y}_{t+4}$

 $X_{t+5} \longrightarrow V_{t+5}$

$$y_t = \ell_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$

- ullet ε_t and ξ_t are independent Gaussian white noise processes.
- Compare ETS(A,N,N) where $\xi_t = \alpha \varepsilon_{t-1}$.
- Parameters to estimate: σ_{ε}^2 and σ_{ξ}^2 .
- If $\sigma_{\varepsilon}^2 = 0$, $y_t \sim \text{NID}(\ell_0, \sigma_{\varepsilon}^2)$.

$$y_t = \ell_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$

- $\mathbf{\epsilon}_t$ and ξ_t are independent Gaussian white noise processes.
- Compare ETS(A,N,N) where $\xi_t = \alpha \varepsilon_{t-1}$.
- Parameters to estimate: σ_{ε}^2 and σ_{ξ}^2 .
- If $\sigma_{\varepsilon}^2 = 0$, $y_t \sim \text{NID}(\ell_0, \sigma_{\varepsilon}^2)$.

$$y_t = \ell_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$

- $\mathbf{\epsilon}_t$ and ξ_t are independent Gaussian white noise processes.
- Compare ETS(A,N,N) where $\xi_t = \alpha \varepsilon_{t-1}$.
- Parameters to estimate: σ_{ε}^2 and σ_{ξ}^2 .
- If $\sigma_{\varepsilon}^2 = 0$, $y_t \sim \text{NID}(\ell_0, \sigma_{\varepsilon}^2)$.

$$y_t = \ell_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$

- $\mathbf{\epsilon}_t$ and ξ_t are independent Gaussian white noise processes.
- Compare ETS(A,N,N) where $\xi_t = \alpha \varepsilon_{t-1}$.
- Parameters to estimate: σ_{ε}^2 and σ_{ξ}^2 .
- If $\sigma_{\varepsilon}^2 = 0$, $y_t \sim \mathsf{NID}(\ell_0, \sigma_{\varepsilon}^2)$.

$$y_t = \ell_t + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \xi_t$$

$$b_t = b_{t-1} + \zeta_t$$

- \bullet ε_t , ξ_t and ζ_t are independent Gaussian white noise processes.
- Compare ETS(A,A,N) where $\xi_t = (\alpha + \beta)\varepsilon_{t-1}$ and $\zeta_t = \beta\varepsilon_{t-1}$
- Parameters to estimate: σ_{ε}^2 , σ_{ε}^2 , and σ_{ζ}^2 .
- If $\sigma_{\zeta}^2 = \sigma_{\xi}^2 = 0$, $y_t = \ell_0 + tb_0 + \varepsilon_t$.
- Model is a time-varying linear regression

$$y_t = \ell_t + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \xi_t$$

$$b_t = b_{t-1} + \zeta_t$$

- $\mathbf{\epsilon}_t$, ξ_t and ζ_t are independent Gaussian white noise processes.
- Compare ETS(A,A,N) where $\xi_t = (\alpha + \beta)\varepsilon_{t-1}$ and $\zeta_t = \beta\varepsilon_{t-1}$
- Parameters to estimate: σ_{ε}^2 , σ_{ε}^2 , and σ_{ζ}^2 .
- If $\sigma_{\zeta}^2 = \sigma_{\xi}^2 = 0$, $y_t = \ell_0 + tb_0 + \varepsilon_t$.
- Model is a time-varying linear regression.

$$y_t = \ell_t + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \xi_t$$

$$b_t = b_{t-1} + \zeta_t$$

- $\mathbf{\epsilon}_t$, ξ_t and ζ_t are independent Gaussian white noise processes.
- Compare ETS(A,A,N) where $\xi_t = (\alpha + \beta)\varepsilon_{t-1}$ and $\zeta_t = \beta\varepsilon_{t-1}$
- Parameters to estimate: σ_{ε}^2 , σ_{ε}^2 , and σ_{ζ}^2 .
- If $\sigma_{\zeta}^2 = \sigma_{\xi}^2 = 0$, $y_t = \ell_0 + tb_0 + \varepsilon_t$.
- Model is a time-varying linear regression

$$y_t = \ell_t + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \xi_t$$

$$b_t = b_{t-1} + \zeta_t$$

- $\mathbf{\epsilon}_t$, ξ_t and ζ_t are independent Gaussian white noise processes.
- Compare ETS(A,A,N) where $\xi_t = (\alpha + \beta)\varepsilon_{t-1}$ and $\zeta_t = \beta\varepsilon_{t-1}$
- Parameters to estimate: σ_{ε}^2 , σ_{ε}^2 , and σ_{ζ}^2 .
- If $\sigma_{\zeta}^2 = \sigma_{\xi}^2 = 0$, $y_t = \ell_0 + tb_0 + \varepsilon_t$.
- Model is a time-varying linear regression.

$$y_t = \ell_t + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \xi_t$$

$$b_t = b_{t-1} + \zeta_t$$

- $\mathbf{\epsilon}_t$, ξ_t and ζ_t are independent Gaussian white noise processes.
- Compare ETS(A,A,N) where $\xi_t = (\alpha + \beta)\varepsilon_{t-1}$ and $\zeta_t = \beta\varepsilon_{t-1}$
- Parameters to estimate: σ_{ε}^2 , σ_{ε}^2 , and σ_{ζ}^2 .
- If $\sigma_{\zeta}^2 = \sigma_{\xi}^2 = 0$, $y_t = \ell_0 + tb_0 + \varepsilon_t$.
- Model is a time-varying linear regression.

$$y_t = \ell_t + s_{1,t} + \varepsilon_t$$
 $\ell_t = \ell_{t-1} + b_{t-1} + \xi_t$
 $b_t = b_{t-1} + \zeta_t$
 $s_{1,t} = -\sum_{j=1}^{m-1} s_{j,t-1} + \eta_t$
 $s_{j,t} = s_{j-1,t-1}, \qquad j = 2, \dots, m-1$

- \blacksquare ε_t , ξ_t , ζ_t and η_t are independent Gaussian white noise processes.
- Compare ETS(A,A,A).
- Parameters to estimate: σ_{ε}^2 , σ_{ξ}^2 , σ_{ζ}^2 and σ_{η}^2
- Deterministic seasonality if $\sigma_n^2 = 0$

$$egin{aligned} y_t &= \ell_t + s_{1,t} + arepsilon_t \ \ell_t &= \ell_{t-1} + b_{t-1} + \xi_t \ b_t &= b_{t-1} + \zeta_t \ s_{1,t} &= -\sum_{j=1}^{m-1} s_{j,t-1} + \eta_t \ s_{j,t} &= s_{j-1,t-1}, \qquad j = 2, \dots, m-1 \end{aligned}$$

- \mathbf{z}_t , ξ_t , ζ_t and η_t are independent Gaussian white noise processes.
- Compare ETS(A,A,A).
- Parameters to estimate: σ_{ε}^2 , σ_{ξ}^2 , σ_{ζ}^2 and σ_{η}^2
- Deterministic seasonality if $\sigma_n^2 = 0$.

$$egin{aligned} y_t &= \ell_t + s_{1,t} + arepsilon_t \ \ell_t &= \ell_{t-1} + b_{t-1} + \xi_t \ b_t &= b_{t-1} + \zeta_t \ s_{1,t} &= -\sum_{j=1}^{m-1} s_{j,t-1} + \eta_t \ s_{j,t} &= s_{j-1,t-1}, \qquad j = 2, \dots, m-1 \end{aligned}$$

- \mathbf{z}_t , ξ_t , ζ_t and η_t are independent Gaussian white noise processes.
- Compare ETS(A,A,A).
- Parameters to estimate: σ_{ε}^2 , σ_{ξ}^2 , σ_{ζ}^2 and σ_{η}^2
- Deterministic seasonality if $\sigma_n^2 = 0$

$$egin{aligned} y_t &= \ell_t + s_{1,t} + arepsilon_t \ \ell_t &= \ell_{t-1} + b_{t-1} + \xi_t \ b_t &= b_{t-1} + \zeta_t \ s_{1,t} &= -\sum_{j=1}^{m-1} s_{j,t-1} + \eta_t \ s_{j,t} &= s_{j-1,t-1}, \qquad j = 2, \dots, m-1 \end{aligned}$$

- \mathbf{z}_t , ξ_t , ζ_t and η_t are independent Gaussian white noise processes.
- Compare ETS(A,A,A).
- Parameters to estimate: σ_{ε}^2 , σ_{ξ}^2 , σ_{ζ}^2 and σ_{η}^2
- Deterministic seasonality if $\sigma_n^2 = 0$.

$$y_t = \ell_t + \sum_{j=1}^{J} s_{j,t} + \varepsilon_t$$
 $\ell_t = \ell_{t-1} + b_{t-1} + \xi_t$
 $b_t = b_{t-1} + \zeta_t$
 $s_{j,t} = \cos \lambda_j s_{j,t-1} + \sin \lambda_j s_{j,t-1}^* + \omega_{j,t}$
 $s_{j,t}^* = -\sin \lambda_j s_{j,t-1} + \cos \lambda_j s_{j,t-1}^* + \omega_{j,t}^*$

- $\lambda_i = 2\pi j/m$
- $\mathbf{\varepsilon}_t$, ξ_t , ζ_t , $\omega_{j,t}$, $\omega_{j,t}^*$ are independent Gaussian white noise processes
- lacksquare $\omega_{j,t}$ and $\omega_{j,t}^*$ have same variance $\sigma_{\omega_j}^2$
- **Equivalent to BSM when** $\sigma_{\omega,j}^2 = \sigma_{\omega}^2$ and J = m/2
- Choose J < m/2 for fewer degrees of freedom

$$egin{aligned} y_t &= \ell_t + \sum_{j=1}^J s_{j,t} + arepsilon_t \ \ell_t &= \ell_{t-1} + b_{t-1} + \xi_t \ b_t &= b_{t-1} + \zeta_t \ s_{j,t} &= \cos \lambda_j s_{j,t-1} + \sin \lambda_j s_{j,t-1}^* + \omega_{j,t} \ s_{j,t}^* &= -\sin \lambda_j s_{j,t-1} + \cos \lambda_j s_{j,t-1}^* + \omega_{j,t}^* \end{aligned}$$

- $\lambda_i = 2\pi j/m$
- lacksquare $\omega_{j,t}$ and $\omega_{j,t}^*$ have same variance $\sigma_{\omega,j}^2$
- **Equivalent to BSM when** $\sigma_{\omega,j}^2 = \sigma_{\omega}^2$ and J = m/2
- lacksquare Choose J < m/2 for fewer degrees of freedom

$$egin{aligned} y_t &= \ell_t + \sum_{j=1}^J s_{j,t} + arepsilon_t \ \ell_t &= \ell_{t-1} + b_{t-1} + \xi_t \ b_t &= b_{t-1} + \zeta_t \ s_{j,t} &= \cos \lambda_j s_{j,t-1} + \sin \lambda_j s_{j,t-1}^* + \omega_{j,t} \ s_{j,t}^* &= -\sin \lambda_j s_{j,t-1} + \cos \lambda_j s_{j,t-1}^* + \omega_{j,t}^* \end{aligned}$$

- $\lambda_i = 2\pi j/m$
- $\mathbf{\varepsilon}_t$, ξ_t , ζ_t , $\omega_{j,t}$, $\omega_{j,t}^*$ are independent Gaussian white noise processes
- lacksquare $\omega_{j,t}$ and $\omega_{j,t}^*$ have same variance $\sigma_{\omega,j}^2$
- lacksquare Equivalent to BSM when $\sigma_{\omega,j}^2=\sigma_\omega^2$ and J=m/2
- Choose J < m/2 for fewer degrees of freedom

$$y_t = \ell_t + \sum_{j=1}^{J} s_{j,t} + \varepsilon_t$$
 $\ell_t = \ell_{t-1} + b_{t-1} + \xi_t$
 $b_t = b_{t-1} + \zeta_t$
 $s_{j,t} = \cos \lambda_j s_{j,t-1} + \sin \lambda_j s_{j,t-1}^* + \omega_{j,t}$
 $s_{j,t}^* = -\sin \lambda_j s_{j,t-1} + \cos \lambda_j s_{j,t-1}^* + \omega_{j,t}^*$

- $\lambda_i = 2\pi j/m$
- $\mathbf{\varepsilon}_t$, ξ_t , ζ_t , $\omega_{j,t}$, $\omega_{j,t}^*$ are independent Gaussian white noise processes
- lacksquare $\omega_{j,t}$ and $\omega_{j,t}^*$ have same variance $\sigma_{\omega,j}^2$
- Equivalent to BSM when $\sigma_{\omega,j}^2 = \sigma_{\omega}^2$ and J = m/2
- Choose J < m/2 for fewer degrees of freedom

$$y_t = \ell_t + \sum_{j=1}^{J} s_{j,t} + \varepsilon_t$$
 $\ell_t = \ell_{t-1} + b_{t-1} + \xi_t$
 $b_t = b_{t-1} + \zeta_t$
 $s_{j,t} = \cos \lambda_j s_{j,t-1} + \sin \lambda_j s_{j,t-1}^* + \omega_{j,t}$
 $s_{j,t}^* = -\sin \lambda_j s_{j,t-1} + \cos \lambda_j s_{j,t-1}^* + \omega_{j,t}^*$

- $\lambda_i = 2\pi j/m$
- $\mathbf{\varepsilon}_t$, ξ_t , ζ_t , $\omega_{j,t}$, $\omega_{j,t}^*$ are independent Gaussian white noise processes
- lacksquare $\omega_{j,t}$ and $\omega_{j,t}^*$ have same variance $\sigma_{\omega,j}^2$
- Equivalent to BSM when $\sigma_{\omega,j}^2 = \sigma_{\omega}^2$ and J = m/2
- Choose J < m/2 for fewer degrees of freedom

- ETS models are much more general as they allow non-linear (multiplicative components).
- ETS allows automatic forecasting due to its larger model space.
- Additive ETS models are almost equivalent to the corresponding structural models.
- ETS models have a larger parameter space. Structural models parameters are always non-negative (variances).
- Structural models are much easier to generalize (e.g., add covariates).
- It is easier to handle missing values with structural models.

- ETS models are much more general as they allow non-linear (multiplicative components).
- ETS allows automatic forecasting due to its larger model space.
- Additive ETS models are almost equivalent to the corresponding structural models.
- ETS models have a larger parameter space. Structural models parameters are always non-negative (variances).
- Structural models are much easier to generalize (e.g., add covariates).
- It is easier to handle missing values with structural models.

- ETS models are much more general as they allow non-linear (multiplicative components).
- ETS allows automatic forecasting due to its larger model space.
- Additive ETS models are almost equivalent to the corresponding structural models.
- ETS models have a larger parameter space. Structural models parameters are always non-negative (variances).
- Structural models are much easier to generalize (e.g., add covariates).
- It is easier to handle missing values with structural models.

- ETS models are much more general as they allow non-linear (multiplicative components).
- ETS allows automatic forecasting due to its larger model space.
- Additive ETS models are almost equivalent to the corresponding structural models.
- ETS models have a larger parameter space. Structural models parameters are always non-negative (variances).
- Structural models are much easier to generalize (e.g., add covariates).
- It is easier to handle missing values with structural models.

ETS vs Structural models

- ETS models are much more general as they allow non-linear (multiplicative components).
- ETS allows automatic forecasting due to its larger model space.
- Additive ETS models are almost equivalent to the corresponding structural models.
- ETS models have a larger parameter space.
 Structural models parameters are always non-negative (variances).
- Structural models are much easier to generalize (e.g., add covariates).
- It is easier to handle missing values with structural models.

ETS vs Structural models

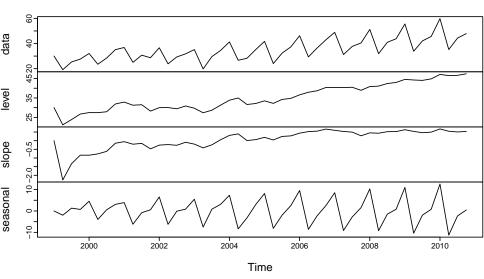
- ETS models are much more general as they allow non-linear (multiplicative components).
- ETS allows automatic forecasting due to its larger model space.
- Additive ETS models are almost equivalent to the corresponding structural models.
- ETS models have a larger parameter space.
 Structural models parameters are always non-negative (variances).
- Structural models are much easier to generalize (e.g., add covariates).
- It is easier to handle missing values with structural models.

Structural models in R

```
StructTS(oil, type="level")
StructTS(ausair, type="trend")
StructTS(austourists, type="BSM")
fit <- StructTS(austourists, type = "BSM")</pre>
decomp <- cbind(austourists, fitted(fit))</pre>
colnames(decomp) <- c("data","level","slope",</pre>
   "seasonal")
plot(decomp, main="Decomposition of
  International visitor nights")
```

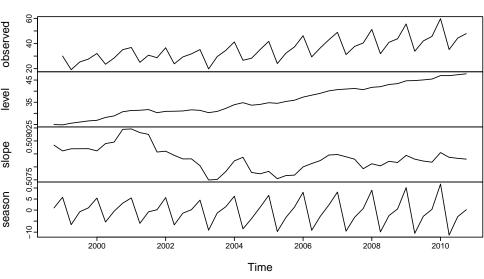
Structural models in R

Decomposition of International visitor nights



ETS decomposition

Decomposition by ETS(A,A,A) method



Outline

- 1 Recall ETS models
- 2 Simple structural models
- 3 Linear Gaussian state space models
- 4 Kalman filter
- **5** ARIMA models in state space form
- **6** Kalman smoothing
- 7 Time varying parameter models

Observation equation
State equation

$$y_t = \mathbf{f}' \mathbf{x}_t + \varepsilon_t$$

$$oldsymbol{x}_t = oldsymbol{G} oldsymbol{x}_{t-1} + oldsymbol{w}_t$$

Observation equation
$$egin{aligned} y_t &= oldsymbol{f}' oldsymbol{x}_t + arepsilon_t \end{aligned}$$
 State equation $oldsymbol{x}_t &= oldsymbol{G} oldsymbol{x}_{t-1} + oldsymbol{w}_t \end{aligned}$

- State vector \mathbf{x}_t of length p
- **G** a $p \times p$ matrix, **f** a vector of length p
- lacksquare $\varepsilon_t \sim \mathsf{NID}(0, \sigma^2)$, $m{w}_t \sim \mathsf{NID}(m{0}, m{W})$.

Local level model:

$$extbf{\emph{f}} = extbf{\emph{G}} = 1, \quad extbf{\emph{x}}_t = \ell_t.$$

$$oldsymbol{f}' = [1 \ 0], \ oldsymbol{\mathsf{x}}_t = egin{bmatrix} \ell_t \ \ell_t \end{bmatrix}$$

$$G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \sigma_{\xi}^2 & 0 \\ 0 & \sigma_{\zeta}^2 \end{bmatrix}$$

Observation equation
$$egin{aligned} y_t &= oldsymbol{f}' oldsymbol{x}_t + arepsilon_t \end{aligned}$$
 State equation $oldsymbol{x}_t &= oldsymbol{G} oldsymbol{x}_{t-1} + oldsymbol{w}_t \end{aligned}$

- State vector \mathbf{x}_t of length p
- **G** a $p \times p$ matrix, **f** a vector of length p
- lacksquare $\varepsilon_t \sim \mathsf{NID}(0, \sigma^2), \quad \mathbf{w}_t \sim \mathsf{NID}(\mathbf{0}, \mathbf{W}).$

Local level model:

$$\mathbf{f} = \mathbf{G} = 1, \quad \mathbf{x}_t = \ell_t.$$

$$extbf{f}' = [1 \ 0], \ egin{bmatrix} \ell_t \end{bmatrix}$$

$$G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \sigma_{\xi}^2 & 0 \\ 0 & \sigma_{\xi}^2 \end{bmatrix}$$

Observation equation
$$egin{aligned} y_t &= oldsymbol{f}' oldsymbol{x}_t + arepsilon_t \end{aligned}$$
 State equation $oldsymbol{x}_t &= oldsymbol{G} oldsymbol{x}_{t-1} + oldsymbol{w}_t \end{aligned}$

- State vector \mathbf{x}_t of length p
- **G** a $p \times p$ matrix, **f** a vector of length p
- \bullet $\varepsilon_t \sim \mathsf{NID}(0, \sigma^2)$, $w_t \sim \mathsf{NID}(\mathbf{0}, \mathbf{W})$.

Local level model:

$$m{f} = m{G} = m{1}, \quad \ \ m{x}_t = \ell_t.$$

$$m{f}' = egin{bmatrix} 1 & 0 \end{bmatrix}, \ m{x}_t = egin{bmatrix} \ell_t \ b_t \end{bmatrix} \quad m{G} = egin{bmatrix} 1 & 1 \ 0 & 1 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \sigma_{\xi}^2 & 0 \\ 0 & \sigma_{\zeta}^2 \end{bmatrix}$$

Observation equation
$$egin{aligned} y_t &= extbf{\textit{f}}' extbf{\textit{x}}_t + arepsilon_t \end{aligned}$$
 State equation $egin{aligned} extbf{\textit{x}}_t &= extbf{\textit{G}} extbf{\textit{x}}_{t-1} + extbf{\textit{w}}_t \end{aligned}$

- State vector \mathbf{x}_t of length p
- **G** a $p \times p$ matrix, **f** a vector of length p
- $\mathbf{E}_t \sim \mathsf{NID}(0, \sigma^2)$, $\mathbf{W}_t \sim \mathsf{NID}(\mathbf{0}, \mathbf{W})$.

Local level model:

$$m{f} = m{G} = m{1}, \quad \ \ m{x}_t = \ell_t.$$

$$m{f}' = [m{1} \ m{0}], \ m{x}_t = egin{bmatrix} \ell_t \\ b_t \end{bmatrix} \quad m{G} = egin{bmatrix} m{1} & m{1} \\ m{0} & m{1} \end{bmatrix} \quad m{W} = egin{bmatrix} \sigma_{\xi}^2 & 0 \\ m{0} & \sigma_{\zeta}^2 \end{bmatrix}$$

Observation equation
$$egin{aligned} y_t &= extbf{\textit{f}}' extbf{\textit{x}}_t + arepsilon_t \end{aligned}$$
 State equation $egin{aligned} extbf{\textit{x}}_t &= extbf{\textit{G}} extbf{\textit{x}}_{t-1} + extbf{\textit{w}}_t \end{aligned}$

- State vector \mathbf{x}_t of length p
- **G** a $p \times p$ matrix, **f** a vector of length p
- $\mathbf{E}_t \sim \mathsf{NID}(0, \sigma^2)$, $\mathbf{W}_t \sim \mathsf{NID}(\mathbf{0}, \mathbf{W})$.

Local level model:

$$m{f} = m{G} = m{1}, \quad m{x}_t = \ell_t.$$

$$m{f}' = [m{1} \ m{0}], \ m{x}_t = egin{bmatrix} \ell_t \ b_t \end{bmatrix} \qquad m{G} = egin{bmatrix} m{1} & m{1} \ m{0} & m{1} \end{bmatrix} \qquad m{W} = egin{bmatrix} \sigma_{\xi}^2 & m{0} \ m{0} & \sigma_{\zeta}^2 \end{bmatrix}$$

Observation equation
$$egin{aligned} y_t &= oldsymbol{f}' oldsymbol{x}_t + arepsilon_t \end{aligned}$$
 State equation $oldsymbol{x}_t &= oldsymbol{G} oldsymbol{x}_{t-1} + oldsymbol{w}_t \end{aligned}$

- State vector x_t of length p
- **G** a $p \times p$ matrix, **f** a vector of length p
- ullet $\varepsilon_t \sim \mathsf{NID}(0, \sigma^2), \quad \mathbf{w}_t \sim \mathsf{NID}(\mathbf{0}, \mathbf{W}).$

Local level model:

$$m{f} = m{G} = m{1}, \quad m{x}_t = \ell_t.$$

$$\mathbf{f}' = \begin{bmatrix} 1 & 0 \end{bmatrix}$$
,

$$\mathbf{x}_t = \begin{bmatrix} \ell_t \\ b_t \end{bmatrix}$$

$$m{G} = egin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$m{x}_t = egin{bmatrix} \ell_t \ b_t \end{bmatrix} \qquad m{G} = egin{bmatrix} 1 & 1 \ 0 & 1 \end{bmatrix} \qquad m{W} = egin{bmatrix} \sigma_{\xi}^2 & 0 \ 0 & \sigma_{\zeta}^2 \end{bmatrix}$$

Basic structural model

Linear Gaussian state space model

$$egin{aligned} \mathbf{y}_t &= \mathbf{f}' \mathbf{x}_t + arepsilon_t, & arepsilon_t &\sim \mathsf{N}(\mathsf{0}, \sigma^2) \ \mathbf{x}_t &= \mathbf{G} \mathbf{x}_{t-1} + \mathbf{w}_t & \mathbf{w}_t &\sim \mathsf{N}(\mathbf{0}, \mathbf{W}) \end{aligned}$$

Basic structural model

Linear Gaussian state space model

$$egin{align} egin{align} eg$$

$$extbf{\emph{f}}' = [extbf{1} extbf{0} extbf{1} extbf{0} extbf{1} \cdots extbf{0}]$$
 , $extbf{\emph{W}} = ext{diagonal}(\sigma_{\xi}^2, \sigma_{\zeta}^2, \sigma_{\eta}^2, 0, \dots, 0)$

$$\mathbf{x}_t = \begin{bmatrix} \ell_t \\ b_t \\ s_{1,t} \\ s_{2,t} \\ s_{3,t} \\ \vdots \\ s_{m-1,t} \end{bmatrix} \mathbf{G} = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & -1 & \dots & -1 & -1 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix}$$

Outline

- 1 Recall ETS models
- 2 Simple structural models
- 3 Linear Gaussian state space models
- 4 Kalman filter
- **5** ARIMA models in state space form
- **6** Kalman smoothing
- 7 Time varying parameter models

Notation:

$$\hat{\mathbf{x}}_{t|t} = \mathsf{E}[\mathbf{x}_t|y_1, \dots, y_t] \ \hat{\mathbf{x}}_{t|t-1} = \mathsf{E}[\mathbf{x}_t|y_1, \dots, y_{t-1}] \ \hat{\mathbf{y}}_{t|t-1} = \mathsf{E}[y_t|y_1, \dots, y_{t-1}]$$

$$\hat{m{P}}_{t|t} = extsf{Var}[m{x}_t|y_1,\ldots,y_t] \ \hat{m{P}}_{t|t-1} = extsf{Var}[m{x}_t|y_1,\ldots,y_{t-1}] \ \hat{m{v}}_{t|t-1} = extsf{Var}[y_t|y_1,\ldots,y_{t-1}]$$

Forecasting:

$$\begin{split} \hat{\mathbf{y}}_{t|t-1} &= \mathbf{f}' \hat{\mathbf{x}}_{t|t-1} \\ \hat{\mathbf{v}}_{t|t-1} &= \mathbf{f}' \hat{\mathbf{P}}_{t|t-1} \mathbf{f} + \sigma^2 \end{split}$$

Updating or State Filtering:

$$\hat{m{x}}_{t|t} = \hat{m{x}}_{t|t-1} + \hat{m{P}}_{t|t-1} m{f} \hat{m{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1}) \\ \hat{m{P}}_{t|t} = \hat{m{P}}_{t|t-1} - \hat{m{P}}_{t|t-1} m{f} \hat{m{v}}_{t|t-1}^{-1} m{f}' \hat{m{P}}_{t|t-1}$$

$$\hat{m{\mathcal{X}}}_{t+1|t} = m{G}\hat{m{\mathcal{X}}}_{t|t} \ \hat{m{P}}_{t+1|t} = m{G}\hat{m{P}}_{t|t}m{G}' + m{W}$$

Notation:

$$\hat{\mathbf{x}}_{t|t} = \mathsf{E}[\mathbf{x}_t|y_1, \dots, y_t] \\ \hat{\mathbf{x}}_{t|t-1} = \mathsf{E}[\mathbf{x}_t|y_1, \dots, y_{t-1}] \\ \hat{y}_{t|t-1} = \mathsf{E}[y_t|y_1, \dots, y_{t-1}]$$

$$\hat{m{P}}_{t|t} = extsf{Var}[m{x}_t|y_1,\ldots,y_t] \ \hat{m{P}}_{t|t-1} = extsf{Var}[m{x}_t|y_1,\ldots,y_{t-1}] \ \hat{m{v}}_{t|t-1} = extsf{Var}[y_t|y_1,\ldots,y_{t-1}]$$

Forecasting:

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{x}}_{t|t-1} \ \hat{\mathbf{v}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{P}}_{t|t-1}\mathbf{f} + \sigma^2$$

Updating or State Filtering:

$$\begin{split} \hat{\pmb{x}}_{t|t} &= \hat{\pmb{x}}_{t|t-1} + \hat{\pmb{\rho}}_{t|t-1} \mathbf{f} \hat{\pmb{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1}) \\ \hat{\pmb{\rho}}_{t|t} &= \hat{\pmb{\rho}}_{t|t-1} - \hat{\pmb{\rho}}_{t|t-1} \mathbf{f} \hat{\pmb{v}}_{t|t-1}^{-1} \mathbf{f}' \hat{\pmb{\rho}}_{t|t-1} \end{split}$$

$$\hat{oldsymbol{x}}_{t+1|t} = oldsymbol{G}\hat{oldsymbol{x}}_{t|t} \ \hat{oldsymbol{P}}_{t+1|t} = oldsymbol{G}\hat{oldsymbol{P}}_{t|t}oldsymbol{G}' + oldsymbol{W}$$

Notation:

$$\begin{split} \hat{\boldsymbol{x}}_{t|t} &= \mathsf{E}[\boldsymbol{x}_t|y_1,\ldots,y_t] \\ \hat{\boldsymbol{x}}_{t|t-1} &= \mathsf{E}[\boldsymbol{x}_t|y_1,\ldots,y_{t-1}] \\ \hat{\boldsymbol{y}}_{t|t-1} &= \mathsf{E}[\boldsymbol{y}_t|y_1,\ldots,y_{t-1}] \\ \end{split} \qquad \begin{aligned} \hat{\boldsymbol{P}}_{t|t} &= \mathsf{Var}[\boldsymbol{x}_t|y_1,\ldots,y_t] \\ \hat{\boldsymbol{P}}_{t|t-1} &= \mathsf{Var}[\boldsymbol{x}_t|y_1,\ldots,y_{t-1}] \\ \hat{\boldsymbol{v}}_{t|t-1} &= \mathsf{Var}[\boldsymbol{y}_t|y_1,\ldots,y_{t-1}] \end{aligned}$$

Forecasting:

$$egin{aligned} \hat{y}_{t|t-1} &= extbf{f}' \hat{ extbf{x}}_{t|t-1} \ \hat{v}_{t|t-1} &= extbf{f}' \hat{ extbf{P}}_{t|t-1} extbf{f} + \sigma^2 \end{aligned}$$

Updating or State Filtering:

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \hat{\mathbf{P}}_{t|t-1} \mathbf{f} \hat{\mathbf{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1})$$
 $\hat{\mathbf{P}}_{t|t} = \hat{\mathbf{P}}_{t|t-1} - \hat{\mathbf{P}}_{t|t-1} \mathbf{f} \hat{\mathbf{v}}_{t|t-1}^{-1} \mathbf{f}' \hat{\mathbf{P}}_{t|t-1}$

$$\hat{oldsymbol{x}}_{t+1|t} = oldsymbol{G}\hat{oldsymbol{x}}_{t|t} \ \hat{oldsymbol{P}}_{t+1|t} = oldsymbol{G}\hat{oldsymbol{P}}_{t|t}oldsymbol{G}' + oldsymbol{W}$$

Notation:

Forecasting:

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{f}' \hat{\mathbf{x}}_{t|t-1}$$

$$\hat{\mathbf{v}}_{t|t-1} = \mathbf{f}' \hat{\mathbf{P}}_{t|t-1} \mathbf{f} + \sigma^2$$

Updating or State Filtering:

$$\begin{split} \hat{\pmb{x}}_{t|t} &= \hat{\pmb{x}}_{t|t-1} + \hat{\pmb{P}}_{t|t-1} \pmb{f} \hat{\pmb{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1}) \\ \hat{\pmb{P}}_{t|t} &= \hat{\pmb{P}}_{t|t-1} - \hat{\pmb{P}}_{t|t-1} \pmb{f} \hat{\pmb{v}}_{t|t-1}^{-1} \pmb{f}' \hat{\pmb{P}}_{t|t-1} \end{split}$$

$$\hat{m{x}}_{t+1|t} = m{G}\hat{m{x}}_{t|t} \ \hat{m{P}}_{t+1|t} = m{G}\hat{m{P}}_{t|t}m{G}' + m{W}$$

Notation:

$$\hat{\mathbf{x}}_{t|t} = \mathsf{E}[\mathbf{x}_t|y_1, \dots, y_t] \\ \hat{\mathbf{x}}_{t|t-1} = \mathsf{E}[\mathbf{x}_t|y_1, \dots, y_{t-1}] \\ \hat{y}_{t|t-1} = \mathsf{E}[y_t|y_1, \dots, y_{t-1}]$$

$$\hat{m{P}}_{t|t} = extsf{Var}[m{x}_t|y_1,\dots,y_t]$$

$$\hat{\boldsymbol{P}}_{t|t-1} = \mathsf{Var}[\boldsymbol{x}_t|y_1,\ldots,y_{t-1}]$$

 $\hat{v}_{t|t-1} = \mathsf{Var}[y_t|y_1,\ldots,y_{t-1}]$

Forecasting:

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{x}}_{t|t-1}$$

$$\hat{\mathbf{v}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{P}}_{t|t-1}\mathbf{f} + \sigma^2$$

Iterate for $t = 1, \dots, T$

Updating or State Filtering:

$$\hat{\pmb{x}}_{t|t} = \hat{\pmb{x}}_{t|t-1} + \hat{\pmb{P}}_{t|t-1} \pmb{f} \hat{\pmb{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1})$$
 $\hat{\pmb{P}}_{t|t} = \hat{\pmb{P}}_{t|t-1} - \hat{\pmb{P}}_{t|t-1} \pmb{f} \hat{\pmb{v}}_{t|t-1}^{-1} \pmb{f}' \hat{\pmb{P}}_{t|t-1}$

$$\hat{m{x}}_{t+1|t} = m{G}\hat{m{x}}_{t|t} \ \hat{m{P}}_{t+1|t} = m{G}\hat{m{P}}_{t|t}m{G}' + m{W}$$

Notation:

$$\hat{\mathbf{x}}_{t|t} = \mathsf{E}[\mathbf{x}_t|y_1, \dots, y_t] \\ \hat{\mathbf{x}}_{t|t-1} = \mathsf{E}[\mathbf{x}_t|y_1, \dots, y_{t-1}] \\ \hat{y}_{t|t-1} = \mathsf{E}[y_t|y_1, \dots, y_{t-1}]$$

$$\hat{m{P}}_{t|t} = extsf{Var}[m{x}_t|y_1,\ldots,y_t] \ \hat{m{P}}_{t|t-1} = extsf{Var}[m{x}_t|y_1,\ldots,y_{t-1}] \ \hat{m{v}}_{t|t-1} = extsf{Var}[y_t|y_1,\ldots,y_{t-1}]$$

Forecasting:

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{x}}_{t|t-1}$$

$$\hat{\mathbf{v}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{P}}_{t|t-1}\mathbf{f} + \sigma^2$$

Iterate for $t = 1, \dots, T$

Assume we know $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$.

Updating or State Filtering:

$$\hat{\pmb{x}}_{t|t} = \hat{\pmb{x}}_{t|t-1} + \hat{\pmb{P}}_{t|t-1} \pmb{f} \hat{\pmb{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1})$$
 $\hat{\pmb{P}}_{t|t} = \hat{\pmb{P}}_{t|t-1} - \hat{\pmb{P}}_{t|t-1} \pmb{f} \hat{\pmb{v}}_{t|t-1}^{-1} \pmb{f}' \hat{\pmb{P}}_{t|t-1}$

$$\hat{m{x}}_{t+1|t} = m{G}\hat{m{x}}_{t|t} \ \hat{m{P}}_{t+1|t} = m{G}\hat{m{P}}_{t|t}m{G}' + m{W}$$

Notation:

$$\hat{\mathbf{x}}_{t|t} = \mathsf{E}[\mathbf{x}_t|y_1, \dots, y_t] \\ \hat{\mathbf{x}}_{t|t-1} = \mathsf{E}[\mathbf{x}_t|y_1, \dots, y_{t-1}] \\ \hat{y}_{t|t-1} = \mathsf{E}[y_t|y_1, \dots, y_{t-1}]$$

$$\hat{m{P}}_{t|t} = \mathsf{Var}[m{x}_t|y_1,\ldots,y_t]$$
 $\hat{m{P}}_{t|t-1} = \mathsf{Var}[m{x}_t|y_1,\ldots,y_{t-1}]$
 $\hat{m{v}}_{t|t-1} = \mathsf{Var}[y_t|y_1,\ldots,y_{t-1}]$

Forecasting:

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{f}' \hat{\mathbf{x}}_{t|t-1}$$

$$\hat{\mathbf{v}}_{t|t-1} = \mathbf{f}' \hat{\mathbf{P}}_{t|t-1} \mathbf{f} + \sigma^2$$

Iterate for $t = 1, \dots, T$

Assume we know $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$.

Updating or State Filtering:

$$\hat{m{x}}_{t|t} = \hat{m{x}}_{t|t-1} + \hat{m{P}}_{t|t-1} m{f} \hat{m{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1})$$

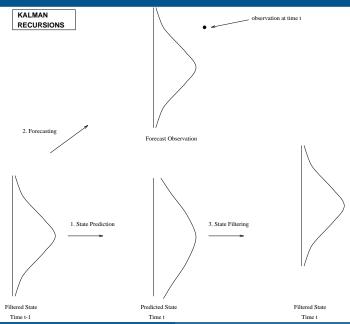
$$\hat{\boldsymbol{\rho}}_{t|t} = \hat{\boldsymbol{\rho}}_{t|t-1} - \hat{\boldsymbol{\rho}}_{t|t-1} \boldsymbol{f} \hat{\boldsymbol{v}}_{t|t-1}^{-1} \boldsymbol{f}' \hat{\boldsymbol{\rho}}_{t|t-1}$$

State Prediction

$$\hat{m{x}}_{t+1|t} = m{G}\hat{m{x}}_{t|t} \ \hat{m{P}}_{t+1|t} = m{G}\hat{m{P}}_{t|t}m{G}' + m{W}$$

Just conditional expectations. So this gives minimum MSE estimates.

Kalman recursions



- Need $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$ to get started.
- Common approach for structural models: set $\mathbf{x}_{1|0} = 0$ and $\mathbf{P}_{1|0} = k\mathbf{I}$ for a very large k.
- Lots of research papers on optimal initialization choices for Kalman recursions.
- ETS approach was to estimate $\mathbf{x}_{1|0}$ and avoid $\mathbf{P}_{1|0}$ by assuming error processes identical.
- A random $\mathbf{x}_{1|0}$ could be used with ETS models, and then a form of Kalman filter would be required for estimation and forecasting.
- This gives more realistic prediction intervals.

- Need $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$ to get started.
- Common approach for structural models: set $\mathbf{x}_{1|0} = 0$ and $\mathbf{P}_{1|0} = k\mathbf{I}$ for a very large k.
- Lots of research papers on optimal initialization choices for Kalman recursions.
- ETS approach was to estimate $\mathbf{x}_{1|0}$ and avoid $\mathbf{P}_{1|0}$ by assuming error processes identical.
- A random $\mathbf{x}_{1|0}$ could be used with ETS models and then a form of Kalman filter would be required for estimation and forecasting.
- This gives more realistic prediction intervals.

- Need $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$ to get started.
- Common approach for structural models: set $\mathbf{x}_{1|0} = 0$ and $\mathbf{P}_{1|0} = k\mathbf{I}$ for a very large k.
- Lots of research papers on optimal initialization choices for Kalman recursions.
- ETS approach was to estimate $\mathbf{x}_{1|0}$ and avoid $\mathbf{P}_{1|0}$ by assuming error processes identical.
- A random $\mathbf{x}_{1|0}$ could be used with ETS models, and then a form of Kalman filter would be required for estimation and forecasting.
- This gives more realistic prediction intervals.

- Need $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$ to get started.
- Common approach for structural models: set $\mathbf{x}_{1|0} = 0$ and $\mathbf{P}_{1|0} = k\mathbf{I}$ for a very large k.
- Lots of research papers on optimal initialization choices for Kalman recursions.
- ETS approach was to estimate $\mathbf{x}_{1|0}$ and avoid $\mathbf{P}_{1|0}$ by assuming error processes identical.
- A random $\mathbf{x}_{1|0}$ could be used with ETS models, and then a form of Kalman filter would be required for estimation and forecasting.
- This gives more realistic prediction intervals.

- Need $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$ to get started.
- Common approach for structural models: set $\mathbf{x}_{1|0} = 0$ and $\mathbf{P}_{1|0} = k\mathbf{I}$ for a very large k.
- Lots of research papers on optimal initialization choices for Kalman recursions.
- ETS approach was to estimate $\mathbf{x}_{1|0}$ and avoid $\mathbf{P}_{1|0}$ by assuming error processes identical.
- A random $\mathbf{x}_{1|0}$ could be used with ETS models, and then a form of Kalman filter would be required for estimation and forecasting.
- This gives more realistic prediction intervals.

- Need $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$ to get started.
- Common approach for structural models: set $\mathbf{x}_{1|0} = 0$ and $\mathbf{P}_{1|0} = k\mathbf{I}$ for a very large k.
- Lots of research papers on optimal initialization choices for Kalman recursions.
- ETS approach was to estimate $\mathbf{x}_{1|0}$ and avoid $\mathbf{P}_{1|0}$ by assuming error processes identical.
- A random $\mathbf{x}_{1|0}$ could be used with ETS models, and then a form of Kalman filter would be required for estimation and forecasting.
- This gives more realistic prediction intervals.

Local level model

$$egin{aligned} y_t &= \ell_t + arepsilon_t \ \ell_t &= \ell_{t-1} + u_t \end{aligned} \qquad egin{aligned} arepsilon_t &\sim \mathsf{NID}(0, \sigma^2) \ u_t &\sim \mathsf{NID}(0, q^2) \end{aligned}$$

Kalman recursions:

$$\begin{split} \hat{y}_{t|t-1} &= \hat{\ell}_{t-1|t-1} \\ \hat{v}_{t|t-1} &= \hat{p}_{t|t-1} + \sigma^2 \\ \hat{\ell}_{t|t} &= \hat{\ell}_{t-1|t-1} + \hat{p}_{t|t-1} \hat{v}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1}) \\ \hat{p}_{t+1|t} &= \hat{p}_{t|t-1} (1 - \hat{v}_{t|t-1}^{-1} \hat{p}_{t|t-1}) + q^2 \end{split}$$

Local level model

$$egin{aligned} y_t &= \ell_t + arepsilon_t \ \ell_t &= \ell_{t-1} + u_t \end{aligned} \qquad egin{aligned} arepsilon_t &\sim \mathsf{NID}(0, \sigma^2) \ u_t &\sim \mathsf{NID}(0, q^2) \end{aligned}$$

Kalman recursions:

$$\begin{split} \hat{y}_{t|t-1} &= \hat{\ell}_{t-1|t-1} \\ \hat{v}_{t|t-1} &= \hat{p}_{t|t-1} + \sigma^2 \\ \hat{\ell}_{t|t} &= \hat{\ell}_{t-1|t-1} + \hat{p}_{t|t-1} \hat{v}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1}) \\ \hat{p}_{t+1|t} &= \hat{p}_{t|t-1} (1 - \hat{v}_{t|t-1}^{-1} \hat{p}_{t|t-1}) + q^2 \end{split}$$

Handling missing values

Forecasting:

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{x}}_{t|t-1}$$

$$\hat{\mathbf{v}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{P}}_{t|t-1}\mathbf{f} + \sigma^2$$

Iterate for t = 1, ..., Tstarting with $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$.

Updating or State Filtering:

$$\hat{m{x}}_{t|t} = \hat{m{x}}_{t|t-1} + \hat{m{P}}_{t|t-1} m{f} \hat{m{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1})$$

$$\hat{\boldsymbol{P}}_{t|t} = \hat{\boldsymbol{P}}_{t|t-1} - \hat{\boldsymbol{P}}_{t|t-1} \boldsymbol{f} \hat{\boldsymbol{v}}_{t|t-1}^{-1} \boldsymbol{f}' \hat{\boldsymbol{P}}_{t|t-1}$$

$$\hat{oldsymbol{x}}_{t|t-1} = oldsymbol{G}\hat{oldsymbol{x}}_{t-1|t-1}$$

$$\hat{oldsymbol{P}}_{t|t-1} = oldsymbol{G}\hat{oldsymbol{P}}_{t-1|t-1}oldsymbol{G}' + oldsymbol{W}'$$

Handling missing values

Forecasting:

$$\begin{split} \hat{y}_{t|t-1} &= \mathbf{f}' \hat{\mathbf{x}}_{t|t-1} \\ \hat{v}_{t|t-1} &= \mathbf{f}' \hat{\mathbf{P}}_{t|t-1} \mathbf{f} + \sigma^2 \end{split}$$

Iterate for t = 1, ..., Tstarting with $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$.

Updating or State Filtering:

$$\begin{split} \hat{\pmb{x}}_{t|t} &= \hat{\pmb{x}}_{t|t-1} + \hat{\pmb{P}}_{t|t-1} \pmb{f} \hat{\pmb{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1}) \\ \hat{\pmb{P}}_{t|t} &= \hat{\pmb{P}}_{t|t-1} - \hat{\pmb{P}}_{t|t-1} \pmb{f} \hat{\pmb{v}}_{t|t-1}^{-1} \pmb{f}' \hat{\pmb{P}}_{t|t-1} \end{split}$$

State Prediction

$$egin{aligned} \hat{oldsymbol{x}}_{t|t-1} &= oldsymbol{G}\hat{oldsymbol{x}}_{t-1|t-1} \ \hat{oldsymbol{P}}_{t|t-1} &= oldsymbol{G}\hat{oldsymbol{P}}_{t-1|t-1}oldsymbol{G}' + oldsymbol{W} \end{aligned}$$

Ignored greyed out section if y_t missing.

Handling missing values

Forecasting:

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{x}}_{t|t-1}$$

$$\hat{\mathbf{v}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{P}}_{t|t-1}\mathbf{f} + \sigma^2$$

Iterate for t = 1, ..., Tstarting with $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$.

Updating or State Filtering:

$$\begin{split} \hat{\boldsymbol{x}}_{t|t} &= \hat{\boldsymbol{x}}_{t|t-1} + \hat{\boldsymbol{\rho}}_{t|t-1} \boldsymbol{f} \hat{\boldsymbol{v}}_{t|t-1}^{-1} (\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1}) \\ \hat{\boldsymbol{\rho}}_{t|t} &= \hat{\boldsymbol{\rho}}_{t|t-1} - \hat{\boldsymbol{\rho}}_{t|t-1} \boldsymbol{f} \hat{\boldsymbol{v}}_{t|t-1}^{-1} \boldsymbol{f}' \hat{\boldsymbol{\rho}}_{t|t-1} \end{split}$$

State Prediction

$$\hat{oldsymbol{x}}_{t|t-1} = oldsymbol{G}\hat{oldsymbol{x}}_{t-1|t-1} \ \hat{oldsymbol{P}}_{t|t-1} = oldsymbol{G}\hat{oldsymbol{P}}_{t-1|t-1}oldsymbol{G}' + oldsymbol{W}$$

Ignored greyed out section if y_t missing.

Multi-step forecasting

Forecasting:

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{x}}_{t|t-1}$$

$$\hat{\mathbf{v}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{P}}_{t|t-1}\mathbf{f} + \sigma^2$$

Iterate for $t = T + 1, \dots, T + h$ starting with $\mathbf{x}_{T|T}$ and $\mathbf{P}_{T|T}$.

Updating or State Filtering:

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \hat{\mathbf{p}}_{t|t-1} \mathbf{f} \hat{\mathbf{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1})$$

$$\hat{\mathbf{p}}_{t|t} = \hat{\mathbf{p}}_{t|t-1} - \hat{\mathbf{p}}_{t|t-1} \mathbf{f} \hat{\mathbf{v}}_{t|t-1}^{-1} \mathbf{f}' \hat{\mathbf{p}}_{t|t-1}$$

$$egin{aligned} \hat{m{x}}_{t|t-1} &= m{G}\hat{m{x}}_{t-1|t-1} \ \hat{m{P}}_{t|t-1} &= m{G}\hat{m{P}}_{t-1|t-1}m{G}' + m{W} \end{aligned}$$

Multi-step forecasting

Forecasting:

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{x}}_{t|t-1} \ \hat{\mathbf{v}}_{t|t-1} = \mathbf{f}'\hat{\mathbf{P}}_{t|t-1}\mathbf{f} + \sigma^2$$

Iterate for $t = T + 1, \dots, T + h$ starting with $\mathbf{x}_{T|T}$ and $\mathbf{P}_{T|T}$.

Updating or State Filtering:

$$\begin{split} \hat{\pmb{x}}_{t|t} &= \hat{\pmb{x}}_{t|t-1} + \hat{\pmb{\rho}}_{t|t-1} \mathbf{f} \hat{\pmb{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1}) \\ \hat{\pmb{\rho}}_{t|t} &= \hat{\pmb{\rho}}_{t|t-1} - \hat{\pmb{\rho}}_{t|t-1} \mathbf{f} \hat{\pmb{v}}_{t|t-1}^{-1} \mathbf{f}' \hat{\pmb{\rho}}_{t|t-1} \end{split}$$

State Prediction

$$\hat{oldsymbol{x}}_{t|t-1} = oldsymbol{G}\hat{oldsymbol{x}}_{t-1|t-1} \ \hat{oldsymbol{P}}_{t|t-1} = oldsymbol{G}\hat{oldsymbol{P}}_{t-1|t-1}oldsymbol{G}' + oldsymbol{W}$$

Treat future values as missing.

- Very general equations for any model in state space format.
- Any model in state space format can easily be generalized.
- Optimal MSE forecasts
- Easy to handle missing values.
- Easy to compute likelihood.

- Very general equations for any model in state space format.
- Any model in state space format can easily be generalized.
- Optimal MSE forecasts
- Easy to handle missing values.
- Easy to compute likelihood.

- Very general equations for any model in state space format.
- Any model in state space format can easily be generalized.
- Optimal MSE forecasts
- Easy to handle missing values.
- Easy to compute likelihood.

- Very general equations for any model in state space format.
- Any model in state space format can easily be generalized.
- Optimal MSE forecasts
- Easy to handle missing values.
- Easy to compute likelihood.

- Very general equations for any model in state space format.
- Any model in state space format can easily be generalized.
- Optimal MSE forecasts
- Easy to handle missing values.
- Easy to compute likelihood.

Likelihood calculation

 $\dot{\theta}=$ all unknown parameters $f_{\theta}(y_t|y_1,y_2,\ldots,y_{t-1})=$ one-step forecast density.

Likelihood

$$L(y_1,\ldots,y_T;\boldsymbol{\theta}) = \prod_{t=1}^T f_{\boldsymbol{\theta}}(y_t|y_1,\ldots,y_{t-1})$$

Gaussian log likelihood

$$\log L = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log\hat{v}_{t|t-1} - \frac{1}{2}\sum_{t=1}^{T}e_{t}^{2}/\hat{v}_{t|t-1}$$

where $e_t = y_t - \hat{y}_{t|t-1}$.

All terms obtained from Kalman filter equations

Likelihood calculation

 $\dot{ heta}=$ all unknown parameters $f_{ heta}(y_t|y_1,y_2,\ldots,y_{t-1})=$ one-step forecast density.

Likelihood

$$L(y_1,\ldots,y_T;\boldsymbol{\theta}) = \prod_{t=1}^T f_{\boldsymbol{\theta}}(y_t|y_1,\ldots,y_{t-1})$$

Gaussian log likelihood

$$\log L = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log\hat{v}_{t|t-1} - \frac{1}{2}\sum_{t=1}^{T}e_{t}^{2}/\hat{v}_{t|t-1}$$

where $e_t = y_t - \hat{y}_{t|t-1}$.

All terms obtained from Kalman filter equations.

Likelihood calculation

 $\dot{\theta}$ = all unknown parameters $f_{\theta}(y_t|y_1, y_2, \dots, y_{t-1})$ = one-step forecast density.

Likelihood

$$L(y_1,\ldots,y_T;\boldsymbol{\theta})=\prod_{t=1}^T f_{\boldsymbol{\theta}}(y_t|y_1,\ldots,y_{t-1})$$

Gaussian log likelihood

$$\log L = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log\hat{v}_{t|t-1} - \frac{1}{2}\sum_{t=1}^{T}e_{t}^{2}/\hat{v}_{t|t-1}$$

where $e_t = y_t - \hat{y}_{t|t-1}$.

All terms obtained from Kalman filter equations.

Outline

- 1 Recall ETS models
- 2 Simple structural models
- 3 Linear Gaussian state space models
- 4 Kalman filter
- **5** ARIMA models in state space form
- **6** Kalman smoothing
- 7 Time varying parameter models

AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t, \qquad e_t \sim \mathsf{NID}(0, \sigma^2)$$

Let
$$\mathbf{x}_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ 0 \end{bmatrix}$.

$$\mathbf{x}_t = \begin{bmatrix} \mathbf{1} & \mathbf{0} \end{bmatrix} \mathbf{x}_t$$
 $\mathbf{x}_t = \begin{bmatrix} \phi_1 & \phi_2 \\ \mathbf{1} & \mathbf{0} \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{w}_t$

AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t, \qquad e_t \sim \text{NID}(0, \sigma^2)$$

Let
$$\mathbf{x}_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ 0 \end{bmatrix}$.

Then

$$egin{aligned} oldsymbol{y}_t &= egin{bmatrix} oldsymbol{0} oldsymbol{x}_t \ oldsymbol{x}_t &= egin{bmatrix} \phi_1 & \phi_2 \ oldsymbol{1} & oldsymbol{0} \end{bmatrix} oldsymbol{x}_{t-1} + oldsymbol{w}_t \end{aligned}$$

Now in state space form

AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t, \qquad e_t \sim \mathsf{NID}(0, \sigma^2)$$

Let
$$\mathbf{x}_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ 0 \end{bmatrix}$.

$$egin{aligned} m{y}_t &= egin{bmatrix} m{0} m{x}_t \ m{x}_t &= egin{bmatrix} \phi_1 & \phi_2 \ m{1} & m{0} \end{bmatrix} m{x}_{t-1} + m{w}_t \end{aligned}$$

- Now in state space form
- We can use Kalman filter to compute likelihood and forecasts.

AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$
, $e_t \sim \text{NID}(0, \sigma^2)$

Let
$$\mathbf{x}_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ 0 \end{bmatrix}$.

$$egin{aligned} m{y}_t &= egin{bmatrix} 1 & 0 m{y}_t \ m{x}_t &= egin{bmatrix} \phi_1 & \phi_2 \ 1 & 0 \end{bmatrix} m{x}_{t-1} + m{w}_t \end{aligned}$$

- Now in state space form
- We can use Kalman filter to compute likelihood and forecasts.

AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$
, $e_t \sim \text{NID}(0, \sigma^2)$

Let
$$\mathbf{x}_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ 0 \end{bmatrix}$.

$$egin{aligned} m{y}_t &= egin{bmatrix} 1 & 0 m{y}_t \ m{x}_t &= egin{bmatrix} \phi_1 & \phi_2 \ 1 & 0 \end{bmatrix} m{x}_{t-1} + m{w}_t \end{aligned}$$

- Now in state space form
- We can use Kalman filter to compute likelihood and forecasts.

AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t, \qquad e_t \sim \mathsf{NID}(0, \sigma^2)$$

Let
$$\mathbf{x}_t = \begin{bmatrix} y_t \\ \phi_2 y_{t-1} \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ 0 \end{bmatrix}$.
$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_t$$
$$\mathbf{x}_t = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{w}_t$$

AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t, \qquad e_t \sim \mathsf{NID}(0, \sigma^2)$$

Alternative formulation

Let
$$m{x}_t = egin{bmatrix} y_t \\ \phi_2 y_{t-1} \end{bmatrix}$$
 and $m{w}_t = egin{bmatrix} e_t \\ 0 \end{bmatrix}$.

$$\mathbf{x}_t = \begin{bmatrix} \phi_1 & \mathbf{1} \\ \phi_2 & 0 \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{w}_t$$

Alternative state space form

We can use Kalman filter to compute likelihoor

AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t, \qquad e_t \sim \mathsf{NID}(0, \sigma^2)$$

Let
$$\mathbf{x}_t = \begin{bmatrix} y_t \\ \phi_2 y_{t-1} \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ 0 \end{bmatrix}$.
$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_t$$

$$\mathbf{x}_t = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{w}_t$$

- Alternative state space form
- We can use Kalman filter to compute likelihood and forecasts

AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t, \qquad e_t \sim \mathsf{NID}(0, \sigma^2)$$

Let
$$m{x}_t = egin{bmatrix} y_t \ \phi_2 y_{t-1} \end{bmatrix}$$
 and $m{w}_t = egin{bmatrix} e_t \ 0 \end{bmatrix}$. $y_t = egin{bmatrix} 1 & 0 \end{bmatrix} m{x}_t \ m{x}_t = egin{bmatrix} \phi_1 & 1 \ \phi_2 & 0 \end{bmatrix} m{x}_{t-1} + m{w}_t$

- Alternative state space form
- We can use Kalman filter to compute likelihood and forecasts.

AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$
, $e_t \sim \text{NID}(0, \sigma^2)$

Let
$$\mathbf{x}_t = \begin{bmatrix} y_t \\ \phi_2 y_{t-1} \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ 0 \end{bmatrix}$.
$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_t$$

$$\mathbf{x}_t = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{w}_t$$

- Alternative state space form
- We can use Kalman filter to compute likelihood and forecasts.

$\mathsf{AR}(p)$ model

$$\mathbf{y}_t = \phi_1 \mathbf{y}_{t-1} + \dots + \phi_p \mathbf{y}_{t-p} + \mathbf{e}_t, \qquad \mathbf{e}_t \sim \mathsf{NID}(\mathbf{0}, \sigma^2)$$

Let
$$m{x}_t = egin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}$$
 and $m{w}_t = egin{bmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

$$m{x}_t = egin{bmatrix} 1 & 0 & 0 & \dots & 0 \end{bmatrix} m{x}_t \ m{x}_t = egin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \ 1 & 0 & \dots & 0 & 0 \ dots & \ddots & dots & dots \ 0 & \dots & 0 & 1 & 0 \end{bmatrix} m{x}_{t-1} + m{w}_t$$

$\mathsf{AR}(p)$ model

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + e_t, \qquad e_t \sim \mathsf{NID}(0, \sigma^2)$$

Let
$$\mathbf{x}_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.
$$\mathbf{y}_t = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \end{bmatrix} \mathbf{x}_t$$

$$\mathbf{x}_t = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{w}_t$$

ARMA(1,1) model

$$\mathbf{y}_t = \phi \mathbf{y}_{t-1} + \theta \mathbf{e}_{t-1} + \mathbf{e}_t, \qquad \mathbf{e}_t \sim \mathsf{NID}(\mathbf{0}, \sigma^2)$$

Let
$$\mathbf{x}_t = \begin{bmatrix} \mathbf{y}_t \\ \theta \mathbf{e}_t \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ \theta \mathbf{e}_t \end{bmatrix}$.

$$egin{aligned} oldsymbol{y}_t &= egin{bmatrix} 1 & 0 \end{bmatrix} oldsymbol{x}_t \ oldsymbol{x}_t &= egin{bmatrix} \phi & 1 \ 0 & 0 \end{bmatrix} oldsymbol{x}_{t-1} + oldsymbol{w}_t \end{aligned}$$

ARMA(1,1) model

$$y_t = \phi y_{t-1} + \theta e_{t-1} + e_t, \qquad e_t \sim \mathsf{NID}(0, \sigma^2)$$

Let
$$\mathbf{x}_t = \begin{bmatrix} \mathbf{y}_t \\ \theta \mathbf{e}_t \end{bmatrix}$$
 and $\mathbf{w}_t = \begin{bmatrix} e_t \\ \theta \mathbf{e}_t \end{bmatrix}$.

$$egin{aligned} m{y}_t &= egin{bmatrix} 1 & 0 \end{bmatrix} m{x}_t \ m{x}_t &= egin{bmatrix} \phi & 1 \ 0 & 0 \end{bmatrix} m{x}_{t-1} + m{w}_t \end{aligned}$$

ARMA(p,q) model

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t$$

Let
$$r = \max(p, q + 1)$$
, $\theta_i = 0$, $q < i \le r$, $\phi_i = 0$, $p < j \le r$.

$$egin{aligned} m{y}_t &= egin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} m{x}_t \ m{x}_t &= egin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \ \phi_2 & 0 & 1 & \ddots & dots \ dots & dots & \ddots & \ddots & 0 \ \phi_{r-1} & 0 & \dots & 0 & 1 \ \phi_r & 0 & 0 & \dots & 0 \end{bmatrix} m{x}_{t-1} + egin{bmatrix} 1 \ heta_1 \ dots \ heta_{r-1} \end{bmatrix} e_t \end{aligned}$$

The arima function in R is implemented using this formulatior

ARMA(p,q) model

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t$$

Let $r = \max(p, q + 1)$, $\theta_i = 0$, $q < i \le r$, $\phi_j = 0$, $p < j \le r$.

$$\mathbf{x}_{t} = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \mathbf{x}_{t}$$

$$\mathbf{x}_{t} = \begin{bmatrix} \phi_{1} & 1 & 0 & \dots & 0 \\ \phi_{2} & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \phi_{r-1} & 0 & \dots & 0 & 1 \\ \phi_{r} & 0 & 0 & \dots & 0 \end{bmatrix} \mathbf{x}_{t-1} + \begin{bmatrix} 1 \\ \theta_{1} \\ \vdots \\ \theta_{r-1} \end{bmatrix} e_{t}$$

The arima function in R is implemented using this formulation

ARMA(p,q) model

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t$$

Let
$$r = \max(p, q + 1)$$
, $\theta_i = 0$, $q < i \le r$, $\phi_j = 0$, $p < j \le r$.

$$egin{aligned} m{y}_t &= egin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} m{x}_t \ m{x}_t &= egin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \ \phi_2 & 0 & 1 & \ddots & dots \ dots & dots & \ddots & \ddots & 0 \ \phi_{r-1} & 0 & \dots & 0 & 1 \ \phi_r & 0 & 0 & \dots & 0 \end{bmatrix} m{x}_{t-1} + egin{bmatrix} 1 \ heta_1 \ dots \ heta_{r-1} \end{bmatrix} m{e}_t \end{aligned}$$

The arima function in R is implemented using this formulation

ARMA(p,q) model

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t$$

Let
$$r = \max(p, q + 1)$$
, $\theta_i = 0$, $q < i \le r$, $\phi_j = 0$, $p < j \le r$.

$$m{x}_t = egin{bmatrix} m{0} & 0 & 0 & 0 & m{x}_t \ m{x}_t = egin{bmatrix} \phi_1 & 1 & 0 & 0 & 0 & 0 \ \phi_2 & 0 & 1 & \ddots & \vdots \ \vdots & \vdots & \ddots & \ddots & 0 \ \phi_{r-1} & 0 & \dots & 0 & 1 \ \phi_r & 0 & 0 & \dots & 0 \end{bmatrix} m{x}_{t-1} + egin{bmatrix} 1 \ heta_1 \ \vdots \ heta_{r-1} \end{bmatrix} m{e}_t$$

The arima function in R is implemented using this formulation.

Outline

- 1 Recall ETS models
- 2 Simple structural models
- 3 Linear Gaussian state space models
- 4 Kalman filter
- **5** ARIMA models in state space form
- **6** Kalman smoothing
- 7 Time varying parameter models

Kalman smoothing

Want estimate of $\mathbf{x}_t | y_1, \dots, y_T$ where t < T. That is, $\hat{\mathbf{x}}_{t|T}$.

$$\hat{m{x}}_{t|T} = \hat{m{x}}_{t|t} + m{A}_t \left(\hat{m{x}}_{t+1|T} - \hat{m{x}}_{t+1|t} \right)$$
 $\hat{P}_{t|T} = \hat{P}_{t|t} + m{A}_t \left(\hat{P}_{t+1|T} - \hat{P}_{t+1|t} \right) m{A}_t'$
where $m{A}_t = \hat{P}_{t|t} m{G}' \left(\hat{P}_{t+1|t} \right)^{-1}$.

- Uses all data, not just previous data.
- Useful for estimating missing values: $\hat{y}_{t|T} = \mathbf{f}'\hat{\mathbf{x}}_{t|T}$.
- Useful for seasonal adjustment when one of the states is a seasonal component.

Kalman smoothing

Want estimate of $\mathbf{x}_t | y_1, \dots, y_T$ where t < T. That is, $\hat{\mathbf{x}}_{t|T}$.

$$\begin{array}{rcl} \hat{\boldsymbol{x}}_{t|T} & = & \hat{\boldsymbol{x}}_{t|t} + \boldsymbol{A}_t \left(\hat{\boldsymbol{x}}_{t+1|T} - \hat{\boldsymbol{x}}_{t+1|t} \right) \\ \hat{P}_{t|T} & = & \hat{P}_{t|t} + \boldsymbol{A}_t \left(\hat{P}_{t+1|T} - \hat{P}_{t+1|t} \right) \boldsymbol{A}_t' \end{array}$$
 where $A_t = \hat{P}_{t|t} \boldsymbol{G}' \left(\hat{P}_{t+1|t} \right)^{-1}$.

- Uses all data, not just previous data.
- Useful for estimating missing values: $\hat{y}_{t|T} = \mathbf{f}'\hat{\mathbf{x}}_{t|T}$.
- Useful for seasonal adjustment when one of the states is a seasonal component.

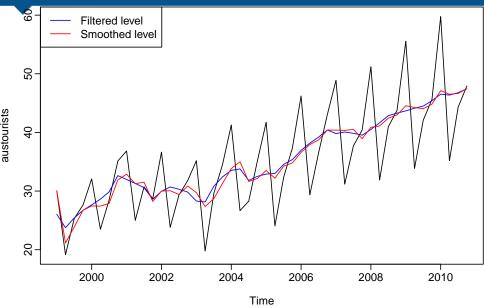
Kalman smoothing

Want estimate of $\mathbf{x}_t | y_1, \dots, y_T$ where t < T. That is, $\hat{\mathbf{x}}_{t|T}$.

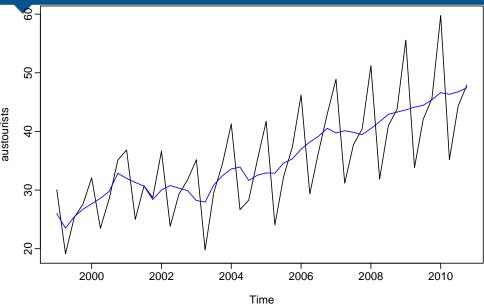
$$\begin{array}{rcl} \hat{\boldsymbol{x}}_{t|T} & = & \hat{\boldsymbol{x}}_{t|t} + \boldsymbol{A}_t \left(\hat{\boldsymbol{x}}_{t+1|T} - \hat{\boldsymbol{x}}_{t+1|t} \right) \\ \hat{P}_{t|T} & = & \hat{P}_{t|t} + \boldsymbol{A}_t \left(\hat{P}_{t+1|T} - \hat{P}_{t+1|t} \right) \boldsymbol{A}_t' \end{array}$$
 where $A_t = \hat{P}_{t|t} \boldsymbol{G}' \left(\hat{P}_{t+1|t} \right)^{-1}$.

- Uses all data, not just previous data.
- Useful for estimating missing values: $\hat{y}_{t|T} = \mathbf{f}'\hat{\mathbf{x}}_{t|T}$.
- Useful for seasonal adjustment when one of the states is a seasonal component.

```
fit <- StructTS(austourists, type = "BSM")
sm <- tsSmooth(fit)</pre>
plot(austourists)
lines(sm[,1],col='blue')
lines(fitted(fit)[,1],col='red')
legend("topleft",col=c('blue','red'),lty=1,
  legend=c("Filtered level","Smoothed level")
```



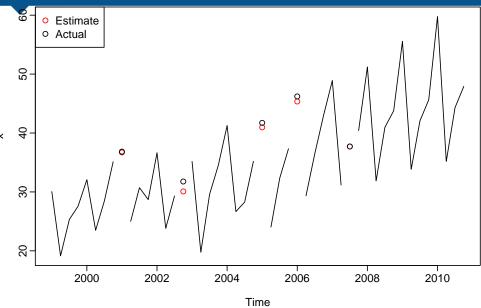
```
fit <- StructTS(austourists, type = "BSM")
sm <- tsSmooth(fit)</pre>
plot(austourists)
# Seasonally adjusted data
aus.sa <- austourists - sm[,3]
lines(aus.sa,col='blue')
```



Kalman smoothing in R

```
x <- austourists
miss <- sample(1:length(x), 5)
x[miss] <- NA
fit <- StructTS(x, type = "BSM")
sm <- tsSmooth(fit)</pre>
estim <- sm[,1]+sm[,3]
plot(x, ylim=range(austourists))
points(time(x)[miss], estim[miss],
  col='red', pch=1)
points(time(x)[miss], austourists[miss],
  col='black', pch=1)
legend("topleft", pch=1, col=c(2,1),
  legend=c("Estimate","Actual"))
```

Kalman smoothing in R



Outline

- 1 Recall ETS models
- 2 Simple structural models
- 3 Linear Gaussian state space models
- 4 Kalman filter
- **5** ARIMA models in state space form
- 6 Kalman smoothing
- 7 Time varying parameter models

Time varying parameter models

Linear Gaussian state space model

$$egin{aligned} \mathbf{y}_t &= \mathbf{f}_t' \mathbf{x}_t + arepsilon_t, & arepsilon_t &\sim \mathsf{N}(\mathbf{0}, \sigma_t^2) \ \mathbf{x}_t &= \mathbf{G}_t \mathbf{x}_{t-1} + \mathbf{w}_t, & \mathbf{w}_t &\sim \mathsf{N}(\mathbf{0}, \mathbf{W}_t) \end{aligned}$$

Kalman recursions:

$$\begin{split} \hat{y}_{t|t-1} &= \mathbf{f}_t' \hat{\mathbf{x}}_{t|t-1} \\ \hat{v}_{t|t-1} &= \mathbf{f}_t' \hat{\mathbf{P}}_{t|t-1} \mathbf{f}_t + \sigma_t^2 \\ \hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \hat{\mathbf{P}}_{t|t-1} \mathbf{f}_t \hat{v}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1}) \\ \hat{\mathbf{P}}_{t|t} &= \hat{\mathbf{P}}_{t|t-1} - \hat{\mathbf{P}}_{t|t-1} \mathbf{f}_t \hat{v}_{t|t-1}^{-1} \mathbf{f}_t' \hat{\mathbf{P}}_{t|t-1} \\ \hat{\mathbf{x}}_{t|t-1} &= \mathbf{G}_t \hat{\mathbf{x}}_{t-1|t-1} \\ \hat{\mathbf{P}}_{t|t-1} &= \mathbf{G}_t \hat{\mathbf{P}}_{t-1|t-1} \mathbf{G}_t' + \mathbf{W}_t \end{split}$$

Time varying parameter models

Linear Gaussian state space model

$$egin{aligned} oldsymbol{y}_t &= oldsymbol{f}_t' oldsymbol{x}_t + arepsilon_t, & arepsilon_t \sim \mathsf{N}(\mathbf{0}, \sigma_t^2) \ oldsymbol{x}_t &= oldsymbol{G}_t oldsymbol{x}_{t-1} + oldsymbol{w}_t & oldsymbol{w}_t \sim \mathsf{N}(oldsymbol{0}, oldsymbol{W}_t) \end{aligned}$$

Kalman recursions:

$$\begin{split} \hat{y}_{t|t-1} &= \mathbf{f}_t' \hat{\mathbf{x}}_{t|t-1} \\ \hat{v}_{t|t-1} &= \mathbf{f}_t' \hat{\mathbf{P}}_{t|t-1} \mathbf{f}_t + \sigma_t^2 \\ \hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \hat{\mathbf{P}}_{t|t-1} \mathbf{f}_t \hat{\mathbf{v}}_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1}) \\ \hat{\mathbf{P}}_{t|t} &= \hat{\mathbf{P}}_{t|t-1} - \hat{\mathbf{P}}_{t|t-1} \mathbf{f}_t \hat{\mathbf{v}}_{t|t-1}^{-1} \mathbf{f}_t' \hat{\mathbf{P}}_{t|t-1} \\ \hat{\mathbf{x}}_{t|t-1} &= \mathbf{G}_t \hat{\mathbf{x}}_{t-1|t-1} \\ \hat{\mathbf{P}}_{t|t-1} &= \mathbf{G}_t \hat{\mathbf{P}}_{t-1|t-1} \mathbf{G}_t' + \mathbf{W}_t \end{split}$$

Local level with covariate

$$y_t = \ell_t + \beta z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$

$$m{f}_t' = [m{1} \ m{z}_t] \quad m{x}_t = egin{bmatrix} \ell_t \\ m{eta} \end{bmatrix} \quad m{G} = egin{bmatrix} m{1} & 0 \\ 0 & 1 \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & 0 \\ 0 & 0 \end{bmatrix}$$

Assumes z_t is fixed and known (as in

B Estimate of β is given by \hat{x}_{TT}

Equivalent to simple linear regression with tilt varying intercept.

Easy to extend to multiple regression with

$$y_t = \ell_t + \beta z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$

$$m{f}_t' = [1 \ z_t] \quad m{x}_t = egin{bmatrix} \ell_t \\ eta \end{bmatrix} \quad m{G} = egin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$y_t = \ell_t + \beta z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$

$$m{f}_t' = egin{bmatrix} 1 \ z_t \end{bmatrix} \quad m{x}_t = egin{bmatrix} \ell_t \ eta \end{bmatrix} \quad m{G} = egin{bmatrix} 1 & 0 \ 0 & 1 \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & 0 \ 0 & 0 \end{bmatrix}$$

- Assumes z_t is fixed and known (as in regression)
- **E**stimate of β is given by $\hat{x}_{T|T}$.
- Equivalent to simple linear regression with time varying intercept.
- Easy to extend to multiple regression with additional terms.

$$y_t = \ell_t + \beta z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$

$$m{f}_t' = egin{bmatrix} 1 \ z_t \end{bmatrix} \quad m{x}_t = egin{bmatrix} \ell_t \ eta \end{bmatrix} \quad m{G} = egin{bmatrix} 1 & 0 \ 0 & 1 \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & 0 \ 0 & 0 \end{bmatrix}$$

- Assumes z_t is fixed and known (as in regression)
- Estimate of β is given by $\hat{x}_{T|T}$.
- Equivalent to simple linear regression with time varying intercept.
- Easy to extend to multiple regression with additional terms.

$$y_t = \ell_t + \beta z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$

$$m{f}_t' = egin{bmatrix} 1 \ z_t \end{bmatrix} \quad m{x}_t = egin{bmatrix} \ell_t \ eta \end{bmatrix} \quad m{G} = egin{bmatrix} 1 & 0 \ 0 & 1 \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & 0 \ 0 & 0 \end{bmatrix}$$

- Assumes z_t is fixed and known (as in regression)
- **E**stimate of β is given by $\hat{x}_{T|T}$.
- Equivalent to simple linear regression with time varying intercept.
- Easy to extend to multiple regression with additional terms.

$$y_t = \ell_t + \beta z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$

$$m{f}_t' = [1\ z_t] \quad m{x}_t = egin{bmatrix} \ell_t \ eta \end{bmatrix} \quad m{G} = egin{bmatrix} 1 & 0 \ 0 & 1 \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & 0 \ 0 & 0 \end{bmatrix}$$

- Assumes z_t is fixed and known (as in regression)
- **E**stimate of β is given by $\hat{x}_{T|T}$.
- Equivalent to simple linear regression with time varying intercept.
- Easy to extend to multiple regression with additional terms.

Simple linear regression with time varying parameters

$$y_{t} = \ell_{t} + \beta_{t} z_{t} + \varepsilon_{t}$$
$$\ell_{t} = \ell_{t-1} + \xi_{t}$$
$$\beta_{t} = \beta_{t-1} + \zeta_{t}$$

$$m{f}_t' = [m{1} \ z_t] \quad m{x}_t = egin{bmatrix} \ell_t \\ eta_t \end{bmatrix} \quad m{G} = egin{bmatrix} m{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & \mathbf{0} \\ \mathbf{0} & \sigma_{\zeta}^2 \end{bmatrix}$$

Allows for a linear regression with parameters that change slowly over time.

Forecasting: Principles and Practice _______Time vary

Simple linear regression with time varying parameters

$$y_t = \ell_t + \beta_t z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$
$$\beta_t = \beta_{t-1} + \zeta_t$$

$$m{f}_t' = [\mathbf{1} \ z_t] \quad m{x}_t = egin{bmatrix} \ell_t \ eta_t \end{bmatrix} \quad m{G} = egin{bmatrix} \mathbf{1} & \mathbf{0} \ \mathbf{0} & \mathbf{1} \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & \mathbf{0} \ \mathbf{0} & \sigma_{\zeta}^2 \end{bmatrix}$$

that change slowly over time.

Parameters follow independent random walks

Simple linear regression with time varying parameters

$$y_t = \ell_t + \beta_t z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$
$$\beta_t = \beta_{t-1} + \zeta_t$$

$$m{f}_t' = [\mathbf{1} \ z_t] \quad m{x}_t = egin{bmatrix} \ell_t \ eta_t \end{bmatrix} \quad m{G} = egin{bmatrix} \mathbf{1} & \mathbf{0} \ \mathbf{0} & \mathbf{1} \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & \mathbf{0} \ \mathbf{0} & \sigma_{\zeta}^2 \end{bmatrix}$$

- Allows for a linear regression with parameters that change slowly over time.
- Parameters follow independent random walks.
- Estimates of parameters given by $\hat{x}_{t|t}$ or $\hat{x}_{t|T}$.

Simple linear regression with time varying parameters

$$y_{t} = \ell_{t} + \beta_{t}z_{t} + \varepsilon_{t}$$
$$\ell_{t} = \ell_{t-1} + \xi_{t}$$
$$\beta_{t} = \beta_{t-1} + \zeta_{t}$$

$$m{f}_t' = [\mathbf{1} \ z_t] \quad m{x}_t = egin{bmatrix} \ell_t \ eta_t \end{bmatrix} \quad m{G} = egin{bmatrix} \mathbf{1} & \mathbf{0} \ \mathbf{0} & \mathbf{1} \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & \mathbf{0} \ \mathbf{0} & \sigma_{\zeta}^2 \end{bmatrix}$$

- Allows for a linear regression with parameters that change slowly over time.
- Parameters follow independent random walks.
- Estimates of parameters given by $\hat{x}_{t|t}$ or $\hat{x}_{t|T}$.

Simple linear regression with time varying parameters

$$y_{t} = \ell_{t} + \beta_{t} z_{t} + \varepsilon_{t}$$
$$\ell_{t} = \ell_{t-1} + \xi_{t}$$
$$\beta_{t} = \beta_{t-1} + \zeta_{t}$$

$$m{f}_t' = [\mathbf{1} \ z_t] \quad m{x}_t = egin{bmatrix} \ell_t \ eta_t \end{bmatrix} \quad m{G} = egin{bmatrix} \mathbf{1} & \mathbf{0} \ \mathbf{0} & \mathbf{1} \end{bmatrix} \quad m{W}_t = egin{bmatrix} \sigma_{\xi}^2 & \mathbf{0} \ \mathbf{0} & \sigma_{\zeta}^2 \end{bmatrix}$$

- Allows for a linear regression with parameters that change slowly over time.
- Parameters follow independent random walks.
- Estimates of parameters given by $\hat{x}_{t|t}$ or $\hat{x}_{t|T}$.

Same idea can be used to estimate a regression iteratively as new data arrives.

$$y_t = \ell_t + \beta_t z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$
$$\beta_t = \beta_{t-1} + \zeta_t$$

$$m{f}_t' = [1 \ z_t] \quad m{x}_t = egin{bmatrix} \ell_t \\ eta_t \end{bmatrix} \quad m{G} = egin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad m{W}_t = egin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Same idea can be used to estimate a regression iteratively as new data arrives.

Simple linear regression with updating parameters

$$y_t = \ell_t + \beta_t z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$
$$\beta_t = \beta_{t-1} + \zeta_t$$

$$m{f}_t' = [m{1} \ z_t] \quad m{x}_t = egin{bmatrix} \ell_t \ eta_t \end{bmatrix} \quad m{G} = egin{bmatrix} m{1} & 0 \ 0 & 1 \end{bmatrix} \quad m{W}_t = egin{bmatrix} 0 & 0 \ 0 & 0 \end{bmatrix}$$

Updated parameter estimates given by $\hat{x}_{t|t}$

Same idea can be used to estimate a regression iteratively as new data arrives.

$$y_{t} = \ell_{t} + \beta_{t} z_{t} + \varepsilon_{t}$$
$$\ell_{t} = \ell_{t-1} + \xi_{t}$$
$$\beta_{t} = \beta_{t-1} + \zeta_{t}$$

$$m{f}_t' = [1 \ z_t] \quad m{x}_t = egin{bmatrix} \ell_t \ eta_t \end{bmatrix} \quad m{G} = egin{bmatrix} 1 & 0 \ 0 & 1 \end{bmatrix} \quad m{W}_t = egin{bmatrix} 0 & 0 \ 0 & 0 \end{bmatrix}$$

Same idea can be used to estimate a regression iteratively as new data arrives.

$$y_{t} = \ell_{t} + \beta_{t} z_{t} + \varepsilon_{t}$$
$$\ell_{t} = \ell_{t-1} + \xi_{t}$$
$$\beta_{t} = \beta_{t-1} + \zeta_{t}$$

$$m{f}_t' = [m{1} \ m{z}_t] \quad m{x}_t = egin{bmatrix} \ell_t \ m{eta}_t \end{bmatrix} \quad m{G} = egin{bmatrix} m{1} & m{0} \ m{0} & m{1} \end{bmatrix} \quad m{W}_t = egin{bmatrix} m{0} & m{0} \ m{0} & m{0} \end{bmatrix}$$

- Updated parameter estimates given by $\hat{x}_{t|t}$.
- Recursive residuals given by $y_t \hat{y}_{t|t-1}$.

Same idea can be used to estimate a regression iteratively as new data arrives.

$$y_t = \ell_t + \beta_t z_t + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \xi_t$$
$$\beta_t = \beta_{t-1} + \zeta_t$$

$$m{f}_t' = egin{bmatrix} 1 \ z_t \end{bmatrix} \quad m{x}_t = egin{bmatrix} \ell_t \ eta_t \end{bmatrix} \quad m{G} = egin{bmatrix} 1 & 0 \ 0 & 1 \end{bmatrix} \quad m{W}_t = egin{bmatrix} 0 & 0 \ 0 & 0 \end{bmatrix}$$

- Updated parameter estimates given by $\hat{x}_{t|t}$.
- Recursive residuals given by $y_t \hat{y}_{t|t-1}$.