Bayesian Time Series modelling and Analysis

6.12.2024

$\{y_t\}$ $\quad \{y_t, t = 1, 2, \dots\}$ $\quad y_{1:T}$

strong stationarity

$\{y_t\}$ any $n > 0$, any sequence $t_1, \dots, t_n$
and any $h > 0$

$(y_{t_1}, \dots, y_{t_n})'$ $\quad (y_{t+h}, \dots, y_{t_n+h})$

weak stationarity (second order stationarity)

$\{y_t\}$ first and second moments of the two sequences exist and are identical.

$E[y_t] = \mu$, $\quad var(y_t) = v$

$cov(y_t, y_s)$ depends on $|t-s|$

Strong stationarity implies weak stationarity

autocorrelation function (ACF)

$\{Y_t\}$

<u>autocovariance</u>

$$\gamma(t,s) = cov(Y_t, Y_s) = E[(Y_t - \mu_t)(Y_s - \mu_s)]$$

$\mu_t = E[Y_t]$, $\mu_s = E[Y_s]$

stationarity $\Rightarrow$ $E[Y_t] = \mu$ for all t

and $\gamma(t,s) = \gamma(|t-s|)$

if $h > 0$ $\quad \gamma(h) = cov(Y_t, Y_{t-h})$

<u>autocorrelation</u>

$$\rho(t,s) = \frac{\gamma(t,s)}{\sqrt{\gamma(t,t)}\sqrt{\gamma(s,s)}}$$

stationarity $\Rightarrow$ $\rho(h) = \gamma(h)/\gamma(0)$

$\gamma(0) = var(Y_t)$

How to get estimates of γ and ρ ?

$y_{1:T}$

sample ACF: assuming stationarity

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=1}^{T-h} (y_{t+h} - \bar{y})(y_t - \bar{y})$$

with $\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$,

$$\hat{\rho}(h) = \hat{\gamma}(h) / \hat{\gamma}(0)$$

The AR(1)

$$y_t = \phi y_{t-1} + \varepsilon_t \quad , \quad \varepsilon_t \overset{iid}{\sim} N(0, v)$$

$$\begin{aligned}
y_t &= \phi(\phi y_{t-2} + \varepsilon_t) + \varepsilon_t \\
&= \phi^2 y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\
&= \dots \\
&= \phi^k y_{t-k} + \sum_{j=0}^{k-1} \phi^j \varepsilon_{t-j}
\end{aligned}$$

if $\phi \in [-1, 1] \Rightarrow$

$$y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \qquad (a)$$

$$E[Y_t] = 0 \quad , \quad var(Y_t) = var(a)$$

$$= \sum_{j=0}^{\infty} \phi^{2j} v$$

$$= \frac{v}{1-\phi^2}$$ does not dependent (stationary)

$$\gamma(h) = E[Y_t Y_{t-h}]$$

$$= E\left[\left(\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\right)\left(\sum_{k=0}^{\infty} \phi^k \varepsilon_{t-h-k}\right)\right]$$

$$= E\left[(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots + \phi^h\varepsilon_{t-h}) (\varepsilon_{t-h} + \phi\varepsilon_{t-h-1} + \phi^2\varepsilon_{t-h-2} + \dots +)\right]$$

if time is not the same in the cross-products the expectation is 0 because cov() and are iid.

$$= E\left[\phi^h \varepsilon_{t-h}^2 + \phi^{h+1}\phi\varepsilon_{t-h-1}^2 + \phi^{h+2}\phi^2\varepsilon_{t-h-2} + \dots\right]$$

$$= v \sum_{j=0}^{\infty} \phi^{h+j}\phi^j$$

$$= v\phi^h \sum_{j=0}^{\infty} \phi^{2j}$$

$$= \frac{v\phi^h}{1-\phi^2}$$

$$\rho(h) = \gamma(h) / \gamma(0) = \phi^h$$

In general, for any integer h, we have

$$\rho(h) = \phi^{|h|}$$

$$\gamma(h) = \frac{v \phi^{|h|}}{1 - \phi^2}$$

The larger the value of ϕ, the decay will be slower.

Maximum likelihood Estimation for AR(1)

$$y_t = \phi y_{t-1} + \varepsilon_t \quad , \quad \varepsilon_t \overset{iid}{\sim} N(0, v)$$

(we work will the conditional likelihood)

$$\phi \in [-1, 1] \Rightarrow \{y_t\} \text{ is stationary}$$

$$y_1 \sim N(0, v/1-\phi^2)$$

$$y_t | y_{t-1} \sim N(\phi y_{t-1}, v)$$

$$y_{1:T}$$

$$p(Y_{1:T} \mid \phi, v) = p(Y_1 \mid \phi, v) \prod_{t=2}^{T} p(Y_t \mid Y_{t-1}, \phi, v)$$

$$= \frac{1 \cdot (1-\phi^2)^{1/2}}{(2\pi v)^{1/2}} \exp\left\{ - \frac{Y_1^2 (1-\phi^2)}{2v} \right\} \cdot \frac{1}{(2\pi v)^{(T-1)1/2}} \exp\left\{ - \frac{\sum_{t=2}^{T} (Y_t - \phi Y_{t-1})^2}{2v} \right\}$$

$$\underbrace{}_{Q^*(\phi)}$$

$$Q^*(\phi) = Y_1^2 (1-\phi^2) + \underbrace{\sum_{t=2}^{T} (Y_t - \phi Y_{t-1})^2}_{Q(\phi)}$$

$$= \frac{(1-\phi)^{1/2}}{(2\pi v)^{T/2}} \exp\left\{ - \frac{Q^*(\phi)}{2v} \right\}$$

This is the full likelihood.

<u>Conditional likelihood</u> (conditioned on the first obs)

$$p(Y_{2:T} \mid Y_1, \phi, v) = \prod_{t=2}^{T} \frac{1}{(2\pi v)^{1/2}} \exp\left\{ - \frac{(Y_t - \phi Y_{t-1})^2}{2v} \right\}$$

$$= \frac{1}{(2\pi v)^{(T-1)/2}} \exp\left\{ - \frac{Q(\phi)}{2v} \right\}$$

$$\begin{pmatrix} Y_2 \\ \vdots \\ Y_T \end{pmatrix} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_{T-1} \end{pmatrix} \phi + \begin{pmatrix} \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{pmatrix}$$

$\Rightarrow$

$$\underset{\sim}{\tilde{y}} = \underset{\sim}{X}\phi + \underset{\sim}{\varepsilon}, \quad \underset{\sim}{\varepsilon} \sim N(0, v\underset{\sim}{I})$$

if $\underset{\sim}{X}$ is full rank, then

$$\hat{\phi} = (\underset{\sim}{X}^T X)^{-1} \underset{\sim}{X}^T \underset{\sim}{\tilde{y}}$$

$$\hat{v} = s^2 = \frac{(\underset{\sim}{\tilde{y}} - \underset{\sim}{X}\hat{\phi})^T(\underset{\sim}{\tilde{y}} - \underset{\sim}{X}\hat{\phi})}{|\dim(\underset{\sim}{\tilde{y}}) - \dim(\hat{\phi})|}$$

$$\boxed{\begin{cases} \hat{\phi}_{mle} = \left(\sum_{t=2}^{T} y_t y_{t-1}\right) \Big/ \left(\sum_{t=2}^{T} y_{t-1}^2\right) \\ s^2 = \sum_{t=2}^{T} (y_t - \hat{\phi}_{mle} y_{t-1})^2 / (T-2) \end{cases}}$$

unbiased

We have to resort to numerical maximization for the AR coefficient if we work with the full likelihood.

$$Y_t = \phi Y_{t-1} + \varepsilon_t \quad , \quad \varepsilon_t \overset{iid}{\sim} N(0,1)$$

$$Y_1 \mid \phi \sim N\left(0, \frac{1}{(1-\phi^2)}\right)$$

$$Y_t \mid Y_{t-1}, \phi \sim N(\phi Y_{t-1}, 1)$$

<u>full likelihood</u>

$$p(Y_{1:T} \mid \phi) = p(Y_1 \mid \phi) \cdot \prod_{t=2}^{T} p(Y_t \mid Y_{t-1}\, \phi) =$$

$$= \frac{(1-\phi^2)^{1/2}}{(2\pi)^{1/2}} \exp\left\{-\frac{Y_t (1-\phi^2)}{2}\right\} \cdot$$

$$\frac{1}{(2\pi)^{(T-1)/2}} \exp\left\{-\sum_{t=2}^{T} \frac{(Y_t - \phi Y_{t-1})^2}{2}\right\}$$

$$= \frac{(1-\phi^2)^{1/2}}{(2\pi)^{T/2}} \exp\left\{-\frac{1}{2} Q^*(\phi)\right\}$$

$$\Rightarrow \log p(Y_{1:T} \mid \phi) = \frac{1}{2} \log(1-\phi^2) - \frac{1}{2} Q^*(\phi) + k$$

$\Rightarrow$ numerical optimization

e.g. Newton-Raphson.

Bayesian Inference in the AR(1)

model based on the conditional likelihood and a reference prior.

$$y_t = \phi y_{t-1} + \varepsilon_t \,, \quad \varepsilon_t \overset{iid}{\sim} N(0, v)$$

we can write as:

$$\underset{\sim}{\tilde{y}} = \underset{\sim}{X} \phi + \underset{\sim}{\varepsilon}$$

ϕ is a scalar or β

$\underset{\sim}{\varepsilon} \sim N(0, v\underline{I})$

simplify posterior

$$\begin{pmatrix} y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_{T-1} \end{pmatrix} \begin{matrix} \phi \\ (\beta) \end{matrix} \begin{pmatrix} \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

conditional likelihood:

$$p(y_{2:T} \mid \phi, v, y_1) = \frac{1}{(2\pi v)^{(T-1)/2}} \exp\left\{ -\frac{(\underset{\sim}{\tilde{y}} - \underset{\sim}{X}\phi)^T (\underset{\sim}{\tilde{y}} - \underset{\sim}{X}\phi)}{2v} \right\}$$

reference prior:

$$p(\phi, v) \propto \frac{1}{v}$$

posterior

$$p(\phi, v \mid y_{1:T}) \propto p(\phi, v)\, p(y_{2:T} \mid y_1, \phi, v)$$

$$p(\phi, v) \propto \frac{1}{v} \cdot (\beta \mid v, \underline{X}, \underline{\tilde{y}}) \sim N\left(\hat{\underline{\beta}}_{mle}, v(\underline{X}^T X)^{-1}\right)$$

$$\cdot (v \mid \underline{X}, \underline{\tilde{y}}) \sim IG\left(\frac{(T-2)}{2}, \frac{Q(\hat{\underline{\beta}}_{mle})}{2}\right)$$

$$\hat{\underline{\beta}}_{mle} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\tilde{y}}$$

$$\hat{\underline{\beta}}_{mle} = \hat{\underline{\phi}}_{mle} = \left(\sum_{t=2}^{T} y_t y_{t-1}\right) \Big/ \sum_{t=2}^{T} y_{t-1}^2$$

$$Q(\hat{\underline{\phi}}_{mle}) = \sum_{t=2}^{T} (y_t - \hat{\underline{\phi}}_{mle} y_{t-1})^2$$

We will obtain samples for $\hat{\underline{\beta}}_{mle}$ and $\hat{\underline{\phi}}_{mle}$.

Definition of the state-space representation

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t,$$

$\varepsilon_t \overset{iid}{\sim} N(0, v)$ general AR(p) setting

The AR characteristic polynomial is

$$\Phi(u) = 1 - \phi_1 u - \ldots - \phi_p u^p$$

u is any complex value.

<u>stable</u> : if $\Phi(u) = 0$ only if $|u| > 1$
(outside of the unit circle)

stable $\Rightarrow$ stationarity.

Then it has a moving average representation.

$$Y_t = \Psi(B)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$ with $\psi = 1$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$

$B^j Y_t = Y_{t-j}$: B : Backshift operator

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots$$

we can also write:

$$\tilde{\Phi}(u) = \prod_{j=1}^{p} (1 - \alpha_j u) \quad (\alpha_1, \alpha_2, \dots, \alpha_p) \text{ are characteristic roots}$$

State-space (or dynamic linear model) representation:

$$Y_t = F^T \underline{X}_t$$

$$\underline{X}_t = G \underline{X}_{t-1} + \underline{w}_t$$

For AR(p), we have

$$\underline{X}_t = (Y_t, Y_{t-1}, \dots, Y_{t-p+1})^T$$

$$F = (1, 0, \dots, 0)^T$$

$$\underline{w}_t = (\varepsilon_t, 0, 0, \dots, 0)$$

$$G = \begin{pmatrix} \phi_1 & \phi_2, & \dots, & \phi_{p-1}, & \phi_p \\ 1 & \ddots & & 0 & 0 \\ 0 & & \ddots & \vdots & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & 1 & 0 \end{pmatrix}$$

if we do the matrix computation, we get back the general formula for AR(p).

$$f_t(h) = E[y_{t+h} \mid y_{1:t}]$$

for forecasting h steps ahead

$$= F^T E[\underset{\sim}{x}_{t+h} \mid y_{1:t}]$$

$$= F^T G \, E[\underset{\sim}{x}_{t+h-1} \mid y_{1:t}]$$

$$= F^T G^h E[\underset{\sim}{x}_t \mid y_{1:t}]$$

$$= F^T G^h \underline{x}_t$$

Eigenstructure of G? The eigenvalues corresponds to the characteristic roots of the process.

Assume we have p distinct reciprocal roots

$$G = E \Lambda E^{-1}$$

E : matrix of eigenvectors

$$\Lambda = \text{diagonal}(\alpha_1, \dots, \alpha_p)$$

$$G^h = E \Lambda^h E^{-1}$$

the power h (number of steps ahead) are the α at the power of h.

$$f_t(h) = \sum_{j=1}^{p} (c+j) \alpha_j^h$$

If the process is stable, the reciprocal roots are gonna decay exponentially. It provides a way to interpret the AR process.

This variance can change for each y_i, so assumption 3 is not met.

On residuals vs fitted values we can see if assumptions are met. Here, not the case.

GLM (Generalized Linear Models)

if there is a correlation structure, mixed models can be used.

ACF of the AR(p)

$$\rho(h) - \phi \rho(h-1) - \dots - \phi_p \rho(h-p) = 0$$

$\alpha_1, \dots, \alpha_r$: reciprocal roots

$\downarrow$

$m_1, \dots, m_r$: multiplicity $\qquad \sum_{i=1}^{r} m_i = p$

in this case

$$\rho(h) = \alpha_1^h p_1(h) + \dots + \alpha_r^h p_r(h)$$

$p_j(h)$ polynomial of order m

- AR(1)

  $\rho(h) = \phi^h$

- AR(2)

  $\alpha_1 = r \exp(i\omega)$

  $\alpha_2 = \exp(i\omega)$

$$\rho(h) = a\alpha_1^h + b\alpha_2^h = c r^h \cos(wh + d)$$
$$|r| < 1$$

The property of the ACF is related to the reciprocal roots

Bayesian Inference in the AR(p)

Reference prior and conditional likelihood

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t \ , \ \varepsilon_t \overset{iid}{\sim} N(0, v)$$

parameters: $\phi_1, \dots, \phi_p, v$

$$(y_t \mid y_{t-1}, \dots, y_{t-p}, \phi_1, \dots, \phi_p, v) \sim \text{Normal}$$
$$\sim N\left(\sum_{j=1}^{p} \phi_j y_{t-j} \ , \ v\right)$$

$$p(y_{(p+1):T} \mid y_{1:p}, \phi_1, \dots, \phi_p, v) = \prod_{t=p+1}^{T} p(y_t \mid y_{t-1}, \dots, y_{t-p}, \phi, v)$$

$$\Rightarrow \quad \underset{\sim}{\tilde{y}} = \underset{\sim}{X}\beta + \underset{\sim}{\varepsilon} \quad , \ \underset{\sim}{\varepsilon} \sim N(0, v\underline{I})$$

$$\Downarrow$$

$$\underset{\sim}{y} = \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \vdots \\ y_T \end{pmatrix} \quad \underset{\sim}{\beta} = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} \quad \underline{X} = \begin{pmatrix} y_p & y_{p-1} & \dots & y_1 \\ \vdots & & & \\ y_{T-1} & y_{T-2} & \dots & y_{T-p} \end{pmatrix}$$

$$\hat{\underset{\sim}{\beta}}_{mle} = (\underset{\sim}{X}^T \underset{\sim}{X})^{-1} \underset{\sim}{X}^T \underset{\sim}{y}$$ given that $\underline{X}$ is full rank

$$p(\underset{\sim}{\beta}, v) \propto \frac{1}{v}$$

$$(\underset{\sim}{\beta} \mid y_{1:T}, v) \sim N(\hat{\underset{\sim}{\beta}}_{mle}, v(\underset{\sim}{X}^T X)^{-1})$$

$$(v \mid y_{1:T}) \sim IG\left(\frac{T-2p}{2}, \frac{S^2}{2}\right)$$

Model order selection

How to chose p? sample ACF and PACF

We can also consider criteria, or assume that p is unknown and chose a prior on p.

p^* : maximum so we will consider

$p = 1 : p^*$

$S_p^2 \qquad y_{(p^*+1):T}$

$$AIC_p = (T - p^*) \log(s_p^2) + \underbrace{2p}_{\text{overfitting penalty}}$$

The optimal p is the one that minimizes the AIC/BIC.

$$BIC_p = (T - p^*) \log(s_p^2) + \underbrace{p \log(T - p^*)}_{\text{overfitting penalty}}$$

we can also look at the posterior predictive distribution.

Spectral Represesentation of AR(p)

(Frequency domain also called spectral domain)

$\phi_1, \dots, \phi_p, v$

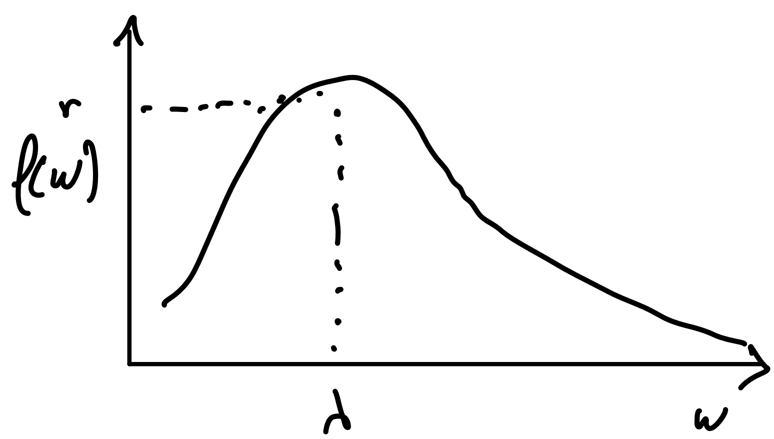
$$f(w) = \frac{v}{|(1 - \phi_1 e^{-iw} - \dots - \phi_p e^{-iwp})|^2 \, 2\pi}$$

frequency between 0 and π

$\Rightarrow$ related to the characteristic roots of the polynomial

eg. α_1 $r = 0.7$ (modulus)

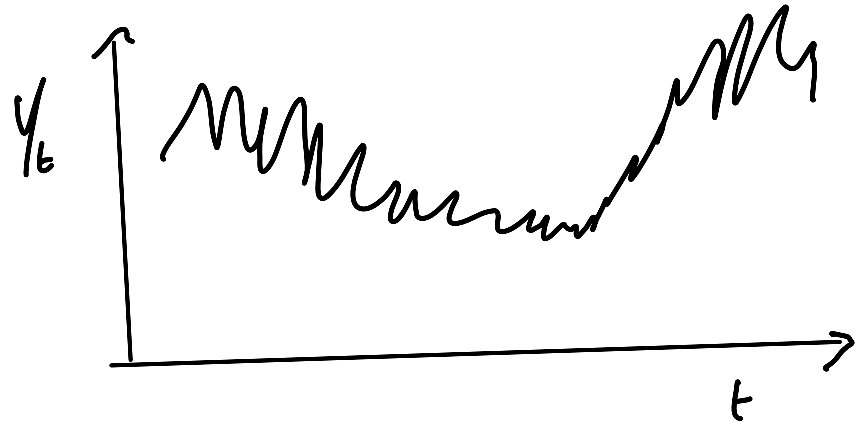
α_2 $\lambda = 12$ (periodicity)



The Normal Dynamic Linear Model (NDLM)

For modelling and analysing nonstationary time series.

$$Y_t = \mu + v_t \quad , \quad v_t \overset{iid}{\sim} N(0, v)$$



$$E[\mu_t] = \mu$$

$$v(Y_t) = v$$

$$Y_t = \mu_t + v_t \quad , \quad v_t \overset{iid}{\sim} N(0, v)$$

$$\mu_t = \mu_{t-1} + w_t \quad , \quad w_t \overset{iid}{\sim} N(0, w)$$

(the mean level is changing over time)

The general class of DLM can be written as follows

$$Y_t = F_t^T \underset{\sim}{\theta}_t + v_t \quad , \quad v_t \overset{iid}{\sim} N(0, V_t) \qquad \text{observation equation}$$

$$\underset{\sim}{\theta}_t = G_t \theta_{t-1} + w_t \quad , \quad w_t \overset{iid}{\sim} N(0, W_t) \qquad \text{system equation}$$

F_t : vector of dimension k

$\underset{\sim}{\theta}_t$: vector of dimension K of parameters

G_t : $K \times K$ matrix

V_t : variance matrix obs. level

W_t : variance matrix system level

$$\underset{\sim}{\theta}_0 \sim N(m_0, C_0) \qquad \text{prior}$$

given

for inference, we are interesting in

1. $p(\underset{\sim}{\theta}_t \mid D_t)$ filtering distribution

$$D_t = \{D_0, Y_{1:T}\}$$

2. $p(Y_{t+h} \mid D_t)$ forecasting distribution

3. $p(\theta_t \mid D_T)$ smoothing distribution

$t < T$

$$f_t(h) = E[Y_{t+h} \mid D_t] = F_{t+h}^T G_{t+h} \cdots G_{t+1} E[\underset{\sim}{\theta}_t \mid D_t]$$

$$F_t = F \qquad G_t = G \qquad \Rightarrow f_t(h) = F^T G^h E[\underset{\sim}{\theta}_t \mid D_t]$$

$$\{F_t, G_t, V_t, W_t\}$$

Polynomial Trend Models

- first order polynomial model

$$Y_t = \theta_t + v_t \quad , \quad v_t \overset{iid}{\sim} N(0, V_t)$$

$$\theta_t = \theta_{t-1} + w_t \quad , \quad w_t \overset{iid}{\sim} N(0, W_t)$$

$$\{1, 1, V_t, W_t\}$$

forecast function

$$f_t(h) = E[\theta_t \mid D_t]$$

- second order polynomial model

$$Y_t = \theta_{t,1} + v_t \quad , \quad v_t \overset{iid}{\sim} N(0, V_t)$$

$$\theta_{t,1} = \theta_{t-1,1} + \theta_{t-1,2} + w_{t,1}$$

$$\theta_{t,2} = \theta_{t-1,2} + w_{t,2}$$

$$\underset{\sim}{w}_t = \begin{pmatrix} w_{t,1} \\ w_{t,2} \end{pmatrix} \qquad \underset{\sim}{w}_t \sim N(0, W_t)$$

Bivariate Normal

$$\underset{\sim}{\theta}_t = \begin{pmatrix} \theta_{t,1} \\ \theta_{t,2} \end{pmatrix} \qquad F^T = (1, 0) \qquad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

past values of the second component

$$f_t(h) = F^T G^h E[\underset{\sim}{\theta}_t \mid D_t] = (1, h) E[\underset{\sim}{\theta}_t \mid D_t]$$

$$G^h = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix}$$

$$= (K_{t,1} + K_{t,2}\, h)$$

linear function on the number of steps ahead h

- general polynomial model
  p^{th} order model

$$\underset{\sim}{\theta}_t = \begin{pmatrix} \theta_{t,1} \\ \vdots \\ \theta_{t,p} \end{pmatrix} \qquad F^T = (1, 0, \dots, 0)$$

$$G = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & \\ & & & \ddots & 1 \\ & & & & 1 \\ 0 & & & & \end{pmatrix} = J_p(I)$$

$$f_t(h) = (K_{t,0} + K_{t,1}\, h + \dots + K_{t,p-1}\, h^{p-1})$$

Other parametrizations exist.

Depending on the "shape" of the time series we chose the degree of the polynomial model accordingly.

The superposition principle

$$f_t(h) = \left(K_{t,0} + K_{t,1} h\right) + \left(K_{t,2} X_{t+h}\right)$$

general structure that we want for the forecast function

$$f_{1t}(h) = \left(K_{t,0} + K_{t,1} h\right)$$

$$f_{2t}(h) = K_{t,2} X_{t+h}$$

$$\left.\begin{array}{l} \{F_1, G_1, \cdot, \cdot\} \\ F_1 = \epsilon_1 = (1,0)^T \qquad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \end{array}\right\} \text{for } f_{1t}(h)$$

$$\left.\begin{array}{l} \{F_2, G_2, \cdot, \cdot\} \\ F_{2,t} = x_t \qquad G_2 = 1 \end{array}\right\} \text{for } f_{2t}(h)$$

$$\{F, G, \cdot, \cdot\}$$

$$F = (1, 0, x_t) \qquad G = \text{block-diagonal structure } (G_1, G_2)$$

$$= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Filtering

Bayesian approach to filtering, smoothing and forecasting equations

$$Y_t = F_t^T \theta_t + \nu_t, \quad \nu_t \overset{iid}{\sim} N(0, v_t)$$

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \overset{iid}{\sim} N(0, W_t)$$

$$(\theta_0 \mid D_0) \sim N(m_0, C_0)$$

prior distribution : Normal-Normal conjugate

assumption : $(\theta_{t-1} \mid D_{t-1}) \sim N(m_{t-1}, C_{t-1})$

- prior at time t

$$(\theta_t \mid D_{t-1}) \sim N(\quad , \quad)$$

$$E[\theta_t \mid D_{t-1}] = G_t E[\theta_{t-1} \mid D_{t-1}]$$
$$= G_t m_{t-1} = a_t$$

$$var(\theta_t \mid D_{t-1}) = G_t var(\theta_{t-1} \mid D_{t-1}) G_t^T + W_t$$

$$= G_t C_{t-1} G_t^T w_t = R_t$$

- $(Y_t \mid D_{t-1})$ one step ahead forecast

$$(Y_t \mid D_{t-1}) \sim N(\quad , \quad)$$

$$E[Y_t \mid D_{t-1}] = F_t^T E[\theta_t \mid D_{t-1}] = F_t^T a_t = f_t$$

$$\begin{aligned} var(Y_t \mid D_{t-1}) &= F_t^T var(\theta_t \mid D_{t-1}) F_t + V_t \\ &= F_t^T R_t F^T + V_t = q_t \end{aligned}$$

posterior

- $(\theta_t \mid D_t)$ $\qquad D_t = \{D_{t-1}, Y_t\}$

$$\left(\begin{pmatrix} \theta_t \\ Y_t \end{pmatrix} \middle| \underset{\text{prior}}{D_{t-1}} \right) \sim N\left(\begin{pmatrix} a_t \\ f_t \end{pmatrix}, \begin{pmatrix} R_t & R_t F_t \\ F_t^T R_t & q_t \end{pmatrix} \right)$$

one step ahead forecast

? $cov(\theta_t, Y_t \mid D_{t-1})$

$$\mathrm{cov}\left(\theta_t, Y_t \mid D_{t-1}\right) = \mathrm{cov}\left(\theta_t, F_t^T\theta_t + v_t \mid D_{t-1}\right)$$
$$= \mathrm{var}\left(\theta_t \mid D_{t-1}\right) F_t$$
$$= \underline{R_t F_t}$$

- $(\theta_t \mid Y_t, D_{t-1})$

  same as

  $(\theta_t \mid D_t)$

from Normal theory, we know that if we have:

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

$$(X_1 \mid X_2) \sim N\left(\alpha^*, \Sigma^*\right)$$
$$\alpha^* = \alpha_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(X_2 - \alpha_2\right)$$
$$\Sigma^* = \Sigma_{11} - \Sigma_{12}\Sigma_{22}\Sigma_{21}$$

So

$$(\theta_t \mid Y_t, D_{t-1}) \sim N(m_t, C_t)$$
$$(\theta_t \mid D_t)$$

$$m_t = a_t + R_t F_t \; q_t^{-1} \, (Y_t - f_t)$$

and $$C_t = R_t - R_t F_t \, q_t^{-1} \, F_t^T R_t$$

set $$e_t = Y_t - f_t$$
$$A_t = R_t F_t \, q_t^{-1}$$

$$\left.\begin{aligned} m_t &= a_t + A_t e_t \\ C_t &= R_t - A_t q_t A_t^T \end{aligned}\right\}$$ posterior mean and var. of the filtering equations

Smoothing and forecasting

$$Y_t = F_t^T \theta_t + V_t \quad , \quad V_t \sim N(0, V)$$ obs. equation

$$\theta_t = G_t \theta_{t-1} + w_t \quad , \quad w_t \sim N(0, W)$$ system equation

$$\{ F_t , G_t , V_t , W_t \}$$

illustrate the process:

		Y_1	Y_2	Y_3	Y_4
	θ_0	θ_1	θ_2	θ_3	θ_4
conj prior	$N(m_0, C_0) \to$	$a_1, R_1 \to$	a_2, R_2	$\to \dots$	$\to \dots$
	forecast	f_1, q_1	f_2, q_2		

using filtering

$$m_3, C_3$$
$$m_1, C_1 \qquad m_2, C_2 \qquad a_3(0), R_3(0)$$
$$a_3(-2), R_3(-2) \quad a_3(-1), R_3(-1)$$

smoothing $(\theta_t \mid D_T) \sim N\left(a_T(t-T), R_T(t-T)\right)$

$$t \leq T$$

recursions

$$a_T(t-T) = m_t + B_t\left[a_T(t-T+1) - a_{t+1}\right]$$

$$R_T(t-T) = C_t + B_t\left[R_T(t-T+1) - R_{t+1}\right]B_t^T$$

$$\begin{cases} a_T(0) = m_T \quad , \quad R_T(0) = C_t \\ B_t = C_t G_{t+1}^T C_t \end{cases}$$

<u>forecasting distributions</u>:

$$(\theta_{t+h} \mid D_t) \sim N\left(a_t(h), R_t(h)\right) \qquad h \geq 0$$

$$\begin{cases} a_t(h) = G_{t+h}\, a_t(h-1) \quad , \quad a_t(0) = m_t \\ R_t(h) = G_{t+h} R_t(h-1) G_{t+h}^T + W_{t+h} \quad R_t(0) = C_t \end{cases}$$

$$(Y_{t+h} \mid D_t) \sim N(f_t(h), q_t(h))$$

$$f_t(h) = F_{t+h}^T a_t(h)$$

$$q_t(h) = F_{t+h}^T R_t(h) F_{t+h} + V_{t+h}$$

Fourier Representation (seasonal representation)

- $\omega \in [0, \pi]$ $\quad$ ω: frequency

$\{E_2, J_2(1, \omega, \cdot, \cdot\}$

$E_2 = (1, 0)^T \quad J_2(1, \omega) = \begin{pmatrix} \cos(\omega) & \sin(\omega \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}$

$$f_t(h) = E_2^T [J_2(1, \omega)]^h \begin{pmatrix} a_t \\ b_t \end{pmatrix}$$

$$= (1, 0) \begin{pmatrix} \cos(\omega h) & \sin(\omega h) \\ -\sin(\omega h) & \cos(\omega h) \end{pmatrix} \begin{pmatrix} a_t \\ b_t \end{pmatrix}$$

$$= a_t \cos(\omega h) + b_t \sin(\omega h)$$

$$= \underset{\substack{\uparrow \\ \text{amplitude}}}{A_t} \cos(\omega h + \underset{\substack{\uparrow \\ \text{phase}}}{B_t})$$

• if $w = \pi$

$\{1, -1, \cdot, \cdot\}$

$\rho_t(h) = (-1)^h a_t$ oscilatory behaviour

p : period

• when p is an odd number

$p = 2m - 1$

$$w_j = \frac{2\pi j}{p} \qquad j = 1, \ldots, 2m-1$$

Now, using the superposition principle

$\{F, G, \cdot, \cdot\}$ $\qquad \theta_t$: $2(m-1)$-vector

$$F = (E_2^T, E_2^T, \ldots, E_2^T)^T$$

$$G = \text{blockdiagonal}(J_2(1, w_1), \ldots, J_2(1, w_{m-1}))$$

$$f_t(h) = \sum_{j=1}^{m-1} A_{t,j} \cos(\omega_j h + B_{t,j})$$ forecast function

- p even number $\qquad p = 2m$

$$\omega_j = \frac{2\pi j}{p} \qquad j = 1, \dots, m-1$$

$$\omega_m = \pi$$

$$\{F, G, \cdot, \cdot\}$$

$$F = (E_2^T, E_2^T, \dots, E_2^T, 1)^T$$

$$G = \text{blockdiagonal}(J_2(1, \omega_1), \dots, J_2(1, \omega_{m-1}), -1)$$

$$f_t(h) = \sum_{j=1}^{m-1} A_{t,j} \cos(\omega_j h + B_{t,j}) + A_{t,m}(-1)^h$$

There are other representations, e.g. seasonal factor representations.

Building NDLM's with multiple components

- Linear trend + seasonal component $p = 4$

(a) linear trend

$$\{F_1, G_1, \cdot, \cdot\}$$

$$F_1 = (1, 0)^T \qquad G_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$f_{1,t}(h) = K_{t,0} + K_{t,1}h$$

(b) seasonal component

$$p = 4 = 2 \cdot 2m$$

$$w_1 = \frac{2\pi}{4} = \frac{\pi}{2} \qquad \{F_2, G_2, \cdot, \cdot\}$$

$$w_2 = \pi \qquad F_2 = (E_2^T, 1) = (1, 0, 1)^T$$

$$G_2 = \text{blockdiagonal}(J_2(1, w_1), -1)$$

$$G_2 = \begin{pmatrix} \cos(\pi/2) & \sin(\pi/2) & 0 \\ -\sin(\pi/2) & \cos(\pi/2) & 0 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Linear + seasonal

$$\{F, G, \cdot, \cdot\}$$

$$F = (F_1^T, F_2^T)^T$$

$$G = \text{blockdiagonal}(G_1, G_2)$$

$$f_{2t}(h) = A_{t,1} \cos\left(\frac{\pi}{2} h + B_{t,1}\right) + A_{t,2}(-1)^h$$

$$f_t(h) = f_{1t}(h) + f_{2t}(h)$$

Filtering, Smoothing and Forecasting:

Unknown observational variance

Filtering $\qquad V_t = v$, v unknown

we will place a prior distribution on the variance

$$Y_t = F_t^T \theta_t + \nu_t \quad , \nu_t \sim N(0, v)$$

$$\theta_t = G_t \theta_{t-1} + w_t \, , \; w_t \sim N(0, v W_t^*)$$

$(\theta_0 | D_0, v) \sim N(m_0, vC_0)$

$(v | D_0) \sim IG(n_0/2, d_0/2)$, $d_0 = n_0 S_0$

Filtering and smoothing equations in closed-form.

<u>conditional filtering equations:</u>

$(\theta_{t-1} | D_{t-1}, v) \sim N(m_{t-1}, vC^*_{t-1})$

$(v | D_{t-1}) \sim IG(n_{t-1}/2, d_{t-1}/2)$ $\quad d_{t-1} = n_{t-1} S_{t-1}$

- $(\theta_t | D_{t-1}, v) \sim N(a_t, vR^*_t)$

  $a_t = G_t m_{t-1}$, $R^*_t = G_t C^*_{t-1} G_t^T + W^*_t$

- $(Y_t | D_{t-1}, v) \sim N(f_t, vq^*_t)$

  $f_t = F_t^T a_t$, $q^*_t = F_t^T R^*_t F_t + 1$

posterior distribution for the variance

- $(v | D_t) \sim IG(n_t/2, d_t/2)$

  $n_t = n_{t-1} + 1$, $d_t = n_t S_t$

$$s_t = s_{t-1} + \frac{s_{t-1}}{n_t}\left(\frac{e_t^2}{q_t} - 1\right), \quad q_t = s_{t-1} q_t^*$$

$$e_t = y_t - f_t$$

- $(\theta_t \mid D_t, v) \sim N(m_t, v C_t^*)$

  $m_t = a_t + A_t e_t \,, \quad C_t^* = R_t^* - A_t A_t^T q_t^*$

  $A_t = R_t^* F_t / q_t^*$

we also have:

- $(\theta_t \mid D_{t-1}) \sim T_{n_{t-1}}(a_t, R_t)$

  $R_t = s_{t-1} R_t^*$

- $(y_t \mid D_{t-1}) \sim T_{n_{t-1}}(f_t, q_t) \,, \quad q_t = s_{t-1} q_t^*$

- $(\theta_t \mid D_t) \sim T_{n_t}(m_t, C_t) \,, \quad C_t = s_t C_t^*$

Smoothing equations:

$$(\theta_t \mid D_T) \sim T_{n_T}\left(a_T(t-T),\, R_T(t-T)\, S_T/S_t\right)$$

Forecasting

- $(\theta_{t+h} \mid D_t) \sim T_{n_t}\left(a_t(h),\, R_t(h)\right)$

$$a_t(h) = G_{t+h}\, a_t(h-1)$$
$$R_t(h) = G_{t+h} R_t(h-1) G_{t+h}^T + W_{t+h}$$

- $(Y_{t+h} \mid D_t) \sim T_{n_t}\left(f_t(h),\, q_t(h)\right)$

$$f_t(h) = F_{t+h}^T a_t(h) \quad , \quad q_t(h) = F_{t+h}^T R_t(h) F_{t+h} + S_t$$

Specifying the system Covariance matrix via discount factors

$$\{F_t, G_t, V_t, W_t\}$$

- V_t, W_t known for all t
- $V_t = V$, with V unknown and W_t known

for all t

- V_t is known for all t

$$\mathrm{var}(\theta_t \mid D_{t-1}) = R_t = \underbrace{G_t C_{t-1} G_t^T}_{P_t} + W_t$$

$P_t = G_t C_{t-1} G_t^T$ corresponds to the prior variance in a model of the form

$$\{F_t, G_t, V_t, 0\}$$

$R_t = \frac{P_t}{\delta}$ with $\delta \in [0,1]$ discount factor

δ close to $0 \Rightarrow$ more uncertainty

$$R_t = P_t + W_t = \frac{P_t}{\delta} \Rightarrow W_t = \left(\frac{1-\delta}{\delta}\right) P_t$$

e.g.

$$P_1 = G_1 C_0 G_1^T$$

- $V_t = V$, V unknown

$$W_t^* = \left(\frac{1-\delta}{\delta}\right) P_t^*$$

with $P_t^* = G_t C_{t-1}^* G_t^T$

How do we chose δ in practice?

$\delta \in (0,1]$ usually we consider $\delta \geq 0.8$

we can use a one-step ahead predictive distributions

$$\underbrace{\log\left(p\left(y_{1:T} \mid D_0, \delta\right)\right)}_{\ell(\delta)} = \sum_{t=1}^{T} \log\left(p\left(y_t \mid D_{t-1}, \delta\right)\right)$$

$\ell(\delta)$ can be used as a likelihood function for δ.

"optimal" δ by maximizing $\ell(\delta)$

we can also consider

$$MSE(\delta) = \sum_{t=1}^{T} e_t^2(\delta) / T$$ to be minimized

with $e_t(\delta) = y_t - f_t(\delta)$

"optimal" δ by minimizing $MSE(\delta)$

Code notes

NDLM, unknown observational variance example

for filtering, smoothing and forecasting

unknown v

set_up_dlm_matrices_unknown_v : pass arguments Ft, Gt, Wt_star to the function

set_up_initial_states_unknown_v : pass priors parameters. So is an initial estimate for the variance. Inverse Gamma prior.

forward_filter_unknown_v : if we pass delta, the function assume we use a discount factor.

nt : degrees of freedom parameters

backward_smoothing_unknown_v : delta for providing a discount factor. No delta: Wt provided by the user.

forecast_function_unknown_v : forecast function

dataset : Nile : 100 annual measurements to the Nile river

we use a first order polynomial level

break 95 to use the model, 5 to estimate

set up the matrices.

m_0 : prior parameter, a priori the mean is 800.

C_0 : 10 (variance)

1. Compute the moments of the filtering distribution and credible interval.
2. Smoothing function and distribution of the state parameter vector.

3. $K=5$ steps ahead forecast function and credible interval.

4. Plot the results filtering : red
smoothing : blue

Case study : EEG data (Electroancephalogram)

Has been analysed in multiple papers.

We will compute the optimal discoant factor

Time varying autoregressive model of order 12

analysis with 3200 observations first

data : higher frequency for lower quasi periodicity at the begining.
lower amplitude over time towards the end

G_t is constant over time.

F_t varies over time.

Co_star : prior variance of 10

n_0 : 1 degree of freedom (prior)

W_t defined using a discount factor.

use MSE to pick the optimal discount factor. 0.994 (fairly large: variation smoothly over time)

Blue : mean of the smoothing distribution (capturing the patterns of the data)

reciprocal roots for each time t.
and also obtain a spectral representation.

moduli
periods

model order of 12.

Plot : roots that have the largest modulus over time.

Trajectory based on the smoothing distribution.
starts high then oscilates but fairly persistant

Frequency (in herz) : declining.

Frequency starting a little bit higher than 5 herz
then progressively declining.
Dominant component mostly above the delta band.

library astsa

arma.spec

results corresponds to one spectra density.
picking different times.
· location of the dominant peak is decreasing over
time

Google trends

library (gtrends R)

1. download
2. Plot (trends? Periodicity?
3. Chose model (linear trend for forecast function)
4. Fourier representation with a period of 12

5. Define the structure of the DLM
6. dlm package to find parameters / matrices
   m_0 = mean level
7. names (model) check we have F and G.
   n_0 = degrees of freedom
8. Set up the matrices
9. Use optimal discount factor
10. Retrieving results for filtering, smoothing and forecasting distributions and credible intervals
11. Plots / smoothing distribution is capturing the trend.
12. Analyse (variations in amplitude?)

Own Project