

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 $h > 0$

$$(y_{t_1}, \dots, y_{t_n})' \quad (y_{t_1+h}, \dots, y_{t_n+h})$$

weak stationarity (second order stationarity)

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

$$E[y_t] = \mu, \quad \text{var}(y_t) = \sigma^2$$

$\text{cov}(y_t, y_s)$ depends on $|t-s|$

Strong stationarity implies weak stationarity

autocorrelation function (ACF)

$\{Y_t\}$

autocovariance

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

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

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

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

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

autocorrelation

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

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

$$\gamma(0) = \text{var}(Y_t)$$

How to get estimates of γ and P ?

$$Y_{1:T}$$

sample ACF : assuming stationarity

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

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

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

The AR(1)

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$$

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

$$\text{if } \phi \in [-1, 1] \Rightarrow$$

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

$$E[Y_t] = 0 \quad , \quad \text{var}(Y_t) = \text{var}(a) \\ = \sum_{j=0}^{\infty} \phi^{2j} V \\ = \frac{V}{1-\phi^2} \quad \begin{array}{l} \text{does not} \\ \text{depend on } t \\ (\text{stationary}) \end{array}$$

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

$$= E\left[\left(\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\right)\left(\sum_{k=0}^{\infty} \phi^k \varepsilon_{t-h-k}\right)\right] \\ = E\left[\left(\varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots + \phi^h \varepsilon_{t-h}\right) \right. \\ \left. \left(\varepsilon_{t-h} + \phi \varepsilon_{t-h-1} + \phi^2 \varepsilon_{t-h-2} + \dots + \right)\right]$$

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

$$= E\left[\phi^h \varepsilon_{t-h}^2 + \phi^{h+1} \phi \varepsilon_{t-h-1}^2 + \phi^{h+2} \phi^2 \varepsilon_{t-h-2}^2 \right. \\ \left. + \dots \right] \\ = V \sum_{j=0}^{\infty} \phi^{h+j} \phi^j \\ = V \phi^h \sum_{j=0}^{\infty} \phi^{2j} \\ = \frac{V \phi^h}{1-\phi^2}$$

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

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)

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

(we work with the conditional likelihood)

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

$$Y_1 \sim N(0, v/(1-\phi^2))$$

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

$$Y_{1:T}$$

$$\begin{aligned}
P(Y_{1:T} | \phi, v) &= P(Y_1 | \phi, v) \prod_{t=2}^T P(Y_t | Y_{t-1}, \phi, v) \\
&= \frac{1 \cdot (1-\phi^2)^{1/2}}{(2\pi v)^{1/2}} \exp \left\{ -\frac{y_1^2(1-\phi^2)}{2v} \right\} \\
&\quad \frac{1}{(2\pi v)^{(T-1)/2}} \exp \left\{ -\frac{\sum_{t=2}^T (y_t - \phi y_{t-1})^2}{2v} \right\} \\
&\quad \frac{Q^*(\phi)}{Q^*(\phi)}
\end{aligned}$$

$$\begin{aligned}
Q^*(\phi) &= y_1^2(1-\phi^2) + \underbrace{\sum_{t=2}^T (y_t - \phi y_{t-1})^2}_{Q(\phi)} \\
&= \frac{(1-\phi)^{1/2}}{(2\pi v)^{T/2}} \exp \left\{ -\frac{Q^*(\phi)}{2v} \right\}
\end{aligned}$$

This is the full likelihood.

Conditional Likelihood (conditioned on the first obs)

$$\begin{aligned}
P(Y_{2:T} | Y_1, \phi, v) &= \prod_{t=2}^T \frac{1}{(2\pi v)^{1/2}} \exp \left\{ -\frac{(y_t - \phi y_{t-1})^2}{2v} \right\} \\
&= \frac{1}{(2\pi v)^{(T-1)/2}} \exp \left\{ -\frac{Q(\phi)}{2v} \right\}
\end{aligned}$$

$$\begin{pmatrix} y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_{T-1} \end{pmatrix} \phi + \begin{pmatrix} \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}$$

\Rightarrow

$$\tilde{Y} = \underline{X}\phi + \underline{\epsilon}, \underline{\epsilon} \sim N(0, \sqrt{T})$$

if \underline{X} is full rank, then

$$\hat{\phi} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \tilde{Y}$$

$$\hat{\sigma}^2 = S^2 = \frac{(\tilde{Y} - \underline{X}\hat{\phi})^T (\tilde{Y} - \underline{X}\hat{\phi})}{|\dim(\tilde{Y}) - \dim(\phi)|}$$

$$\left\{ \begin{array}{l} \hat{\phi}_{mle} = \left(\sum_{t=2}^T Y_t Y_{t-1} \right) / \left(\sum_{t=2}^T Y_{t-1}^2 \right) \\ S^2 = \sum_{t=2}^T (Y_t - \hat{\phi}_{mle} Y_{t-1})^2 / (T-2) \end{array} \right.$$

unbiased

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

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

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

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

full likelihood

$$\begin{aligned} P(Y_{1:T} | \phi) &= P(Y_1 | \phi) \cdot \prod_{t=2}^T P(Y_t | Y_{t-1}, \phi) = \\ &= \frac{(1-\phi^2)^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{Y_t (1-\phi^2)}{2} \right\} \\ &\quad \frac{1}{(2\pi)^{(T-1)/2}} \exp \left\{ -\sum_{t=2}^T \frac{(Y_t - \phi Y_{t-1})^2}{2} \right\} \\ &= \frac{(1-\phi^2)^{1/2}}{(2\pi)^{T/2}} \exp \left\{ -\frac{1}{2} Q^*(\phi) \right\} \end{aligned}$$

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

\Rightarrow numerical optimization
e.g. Newton-Raphson.

Bayesian Inference in the AR(1)

model based on the conditional likelihood and a reference prior.

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

we can write as,

$$\tilde{Y} = X\phi + \underline{\varepsilon}$$

ϕ is a scalar or β
 $\underline{\varepsilon} \sim N(0, vI)$

$$\begin{pmatrix} Y_2 \\ \vdots \\ Y_T \end{pmatrix} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_{T-1} \end{pmatrix} \begin{pmatrix} \phi \\ (\beta) \end{pmatrix} \begin{pmatrix} \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

Simplify posteriors

conditional likelihood:

$$P(Y_{2:T} | \phi, v, Y_1) = \frac{1}{(2\pi v)^{(T-1)/2}} \exp \left\{ -\frac{(\tilde{Y} - X\phi)^T (\tilde{Y} - X\phi)}{2v} \right\}$$

reference prior:

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

posterior

$$p(\phi, v | Y_{1:T}) \propto p(\phi, v) P(Y_{2:T} | Y_1, \phi, v)$$

$$P(\phi, \nu) \propto \frac{1}{\nu} \cdot (\beta | v, \underline{x}, \tilde{\underline{y}}) \sim N(\hat{\beta}_{mle}, \nu(\underline{X}^T \underline{X})^{-1})$$

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

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

$$\hat{\sigma}_{mle}^2 = \hat{Q}_{mle} = \left(\sum_{t=2}^T y_t y_{t-1} \right) / \sum_{t=2}^T y_{t-1}^2$$

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

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

Definition of the state-space representation

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

$\varepsilon_t \stackrel{\text{iid}}{\sim} N(0, V)$ general AR(p) setting

The AR characteristic polynomial is

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

u is any complex value.

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

stable \Rightarrow stationarity.

Then it has a moving average representation.

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

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

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

$$\Psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

we can also write :

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

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

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

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

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

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

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

$$\begin{aligned}
 f_t(h) &= E[Y_t + h | Y_{1:t}] \\
 &= F^T E[\underline{X}_{t+h} | Y_{1:t}] \\
 &= F^T G E[\underline{X}_{t+h-1} | Y_{1:t}] \\
 &= F^T G^h E[\underline{X}_t | Y_{1:t}] \\
 &= F^T G^h \underline{X}_t
 \end{aligned}$$

for forecasting h
steps ahead

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

$$\sum_{i=1}^r m_i = p$$

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)

$$\begin{aligned}\alpha_1 &= r \exp(i\omega) \\ \alpha_2 &= \exp(i\omega)\end{aligned}$$

$$p(h) = a\alpha_1^h + b\alpha_2^h = cr^h \cos(wh+d)$$

$$|r| < 1$$

The property of the ACF is related to the reciprocal roots

Bayesian Inference in the AR(p)

Reference prior and conditional likelihood

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} N(0, v)$$

parameters: ϕ_1, \dots, ϕ_p, v

$$(Y_t | Y_{t-1}, \dots, Y_{t-p}, \phi_1, \dots, \phi_p, v) \sim \text{Normal}$$

$$\sim N\left(\sum_{j=1}^p \phi_j Y_{t-j}, v\right)$$

$$P(Y_{(p+1):T} | Y_{1:p}, \phi_1, \dots, \phi_p, v) = \prod_{t=p+1}^T P(Y_t | Y_{t-1}, \dots, Y_{t-p}, \phi, v)$$

$$\Rightarrow \tilde{Y} = X\beta + \varepsilon, \quad \varepsilon \sim N(0, vI)$$



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

$$\hat{\underline{\beta}}_{\text{MLE}} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \tilde{\underline{y}} \quad \text{given that } \underline{X} \text{ is full rank}$$

$$P(\underline{\beta}, v) \propto \frac{1}{v}$$

$$(\underline{\beta} | Y_{1:T}, v) \sim N(\hat{\underline{\beta}}_{\text{MLE}}, v(\underline{X}^T \underline{X})^{-1})$$

$$(v | Y_{1:T}) \sim IG\left(\frac{T-2p}{2}, \frac{s^2}{2}\right)$$

Model order selection

How to choose 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

$$p = 1: p^*$$

$$S_p^2 \quad Y_{(p^*+1):T}$$

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

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 Representation of AR(p)

(Frequency domain also called spectral domain)

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

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

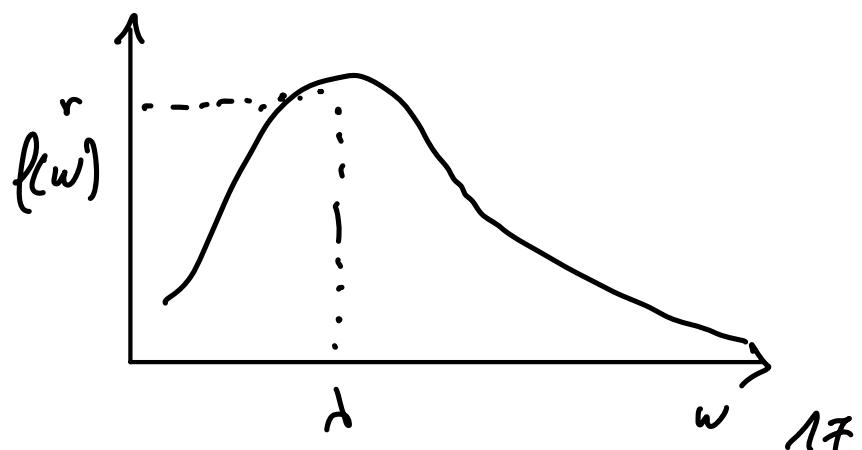
frequency between
0 and π

\Rightarrow related to the characteristic
roots of the polynomial

e.g. α_1 $r = 0.7$

α_2 $\lambda = 12$

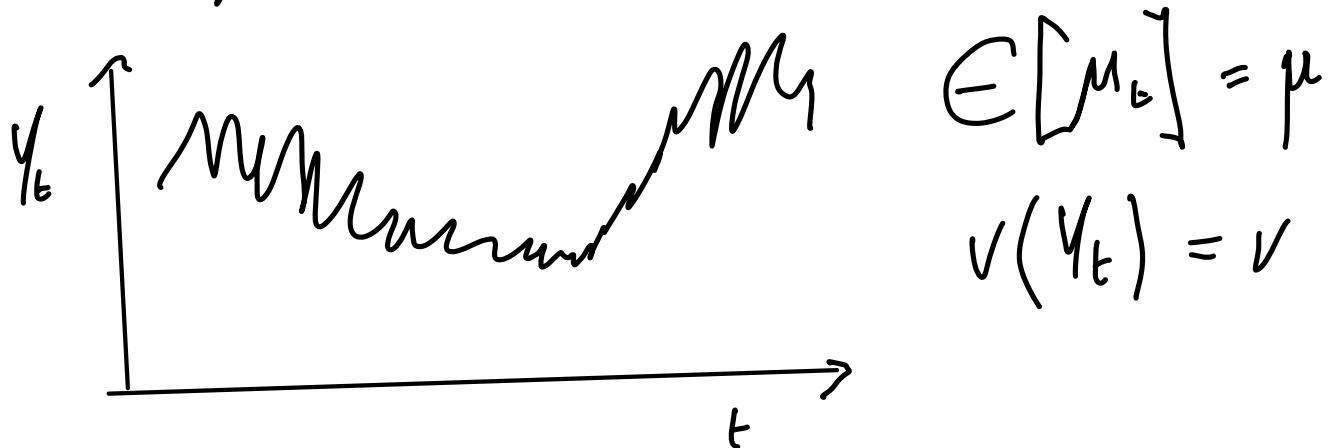
periodicity



The Normal Dynamic Linear Model (NDLM)

For modelling and analysing nonstationary time series.

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



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

$$\mu_t = \mu_{t-1} + w_t, \quad w_t \stackrel{iid}{\sim} N(0, w) \quad \left. \begin{array}{l} \text{the mean level} \\ \text{is changing over} \\ \text{time} \end{array} \right\}$$

The general class of DLM can be written as follows

$$Y_t = F_t^T \underline{\theta}_t + V_t, \quad V_t \stackrel{iid}{\sim} N(0, V_t) \quad \begin{array}{l} \text{observation} \\ \text{equation} \end{array}$$

$$\underline{\theta}_t = G_t \underline{\theta}_{t-1} + w_t, \quad w_t \stackrel{iid}{\sim} N(0, W_t) \quad \begin{array}{l} \text{system} \\ \text{equation} \end{array}$$

F_t : vector of dimension K

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

G_t : $K \times K$ matrix

V_b : variance matrix obs. level
 W_t : variance matrix system level

$$\underline{\theta}_0 \sim N(m_0, C_0) \quad \text{prior}$$

↓
given

For inference, we are interesting in

$$1. \ p(\underline{\theta}_t | D_t) \quad \text{filtering distribution}$$

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

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

$$3. \ p(\underline{\theta}_t | D_T) \quad \text{smoothing distribution}$$

$$t < T$$

$$f_t(h) = E[Y_{t+h} | D_t] = F_{t+h}^T G_{t+h} \dots \\ G_{t+1} E[\underline{\theta}_t | D_t]$$

$$F_t = F \Rightarrow f_t(h) = F^T G^h E[\underline{\theta}_t | D_t]$$

$$G_t = G$$

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

Polynomial Trend Models

- first order polynomial model

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

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

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

forecast function

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

- second order polynomial model

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

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

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

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

Bivariate Normal

$$\underline{\theta}_G = \begin{pmatrix} \theta_{t,1} \\ \theta_{t,2} \end{pmatrix} \quad F^T = (1, 0) \quad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Past values of the
second component

$$f_t(h) = F^T G^h \in [\underline{\theta}_t | \bar{\theta}_t] = (1, h) \in [\underline{\theta}_t | \bar{\theta}_t]$$

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

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

linear function on the number of steps ahead h

- general polynomial model
 p^{th} order model

$$\underline{\theta}_t = \begin{pmatrix} \underline{\theta}_{t,1} \\ \vdots \\ \underline{\theta}_{t,p} \end{pmatrix} \quad F^T = (1, 0, \dots, 0)$$

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

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

Other parametrizations exist.

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

The superposition principle

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

general structure that we want for the forecast function

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

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

$$\left\{ F_1, G_1, \dots, \right\} \quad F_1 = E_1 = (1, 0)^T \quad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \left. \right\} \text{ for } f_{1t}(h)$$

$$\left\{ F_2, G_2, \dots, \right\} \quad F_2 = X_t \quad G_2 = 1 \quad \left. \right\} \text{ for } f_{2t}(h)$$

$$\left\{ F, G, \dots, \right\} \quad F = (1, 0, X_t) \quad G = \text{block-diagonal structure} \\ (G_1, G_2)$$

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

Filtering

Bayesian approach to filtering, smoothing and forecasting equations

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

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

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

prior distribution : Normal-Normal conjugate

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

- prior at time t

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

$$E[\theta_t | D_{t-1}] = G_t E[\theta_{t-1} | D_{t-1}]$$

$$= G_t m_{t-1} = a_t$$

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

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

• $(Y_t | D_{t-1})$ one step ahead forecast

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

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

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

posterior

$$\bullet (\theta_t | D_E) \quad D_E = \{D_{t-1}, Y_t\}$$

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

one step ahead forecast ? $\text{cov}(\theta_t, Y_t | D_{t-1})$

$$\begin{aligned}\text{cov}(\theta_t, y_t | D_{t-1}) &= \text{cov}(\theta_t, F_t^T \theta_t + v_t | D_{t-1}) \\ &= \text{var}(\theta_t | D_{t-1}) F_t \\ &= \underline{R_t F_t}\end{aligned}$$

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

Same as

$$(\theta_t | D_t)$$

from Normal theory, we know that if we have:

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

$$(X_1, X_2) \sim N(\alpha^*, \Sigma^*)$$

$$\alpha^* = \alpha_1 + \Sigma_{12} \Sigma_{22}^{-1} (X_2 - \alpha_2)$$

$$\Sigma^* = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

So

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

$$(\theta_t | D_t)$$

$$m_t = a_t + R_t F_t q_t^{-1} (y_t - p_t)$$

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

set $e_t = y_t - p_t$

$$A_t = R_t F_t q_t^{-1}$$

$$m_t = a_t + A_t e_t$$

$$C_t = R_t - A_t q_t A_t^T$$

} posterior mean of
the filtering equations
and var.

Smoothing and Forecasting

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

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

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

illustrate the process:

$$\theta_0 \quad \theta_1$$

$$y_1 \quad y_2$$

$$\theta_2 \quad \theta_3$$

$$y_3$$

$$y_4$$

conj prior $N(m_0, C_0) \rightarrow a_1, R_1 \rightarrow a_2, R_2 \rightarrow \dots \rightarrow \dots$

forecast $p_1, q_1 \rightarrow p_2, q_2 \rightarrow \dots \rightarrow \dots$

using filtering

$$\begin{matrix} m_1, c_1 \\ a_3(-2), R_3(-2) \end{matrix} \quad \begin{matrix} m_2, c_2 \\ a_3(-1), R_3(-1) \end{matrix} \quad \begin{matrix} m_3, c_3 \\ a_3(0), R_3(0) \end{matrix}$$

smoothing $(\theta_{t|T} | D_T) \sim N(a_{t|T}, R_{t|T})$

$$t \leq T$$

recursions

$$a_{t|T} = m_t + B_t [a_{t|T+1} - a_{t+1}]$$

$$R_{t|T} = C_t + B_t [R_{t|T+1} - R_{t+1}] B_t^T$$

$$\begin{cases} a_{t|T} = m_t & , R_{t|T} = C_t \\ B_t = C_t G_{t+1}^T C_t \end{cases}$$

forecasting distributions :

$$(\theta_{t+h} | D_t) \sim N(a_{t+h}, R_{t+h}) \quad h \geq 0$$

$$\begin{cases} a_{t+h} = G_{t+h} a_{t+h-1} & , a_{t+h}(0) = m_t \\ R_{t+h} = G_{t+h} R_{t+h-1} G_{t+h}^T + W_{t+h} & R_{t+h}(0) = C_t \end{cases}$$

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

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

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

Fourier Representation (Seasonal representation)

- $\omega \in [0, \pi]$ ω : Frequency

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

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

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

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

$$= a_t \cos(wh) + b_t \sin(wh)$$

$$= A_t \cos(wh + \beta_t)$$

↑ ↑
amplitude phase

- if $\omega = \pi$

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

$$P_t(h) = (-1)^h a_e \quad \text{oscillatory behaviour}$$

P : period

- when P is an odd number

$$P = 2m - 1$$

$$\omega_j = \frac{2\pi j}{P} \quad j = 1, \dots, 2m - 1$$

Now, using the superposition principle

$$\{F, G, \cdot, \cdot\} \quad \theta_E : 2(m-1) \text{-vector}$$

$$F = (\epsilon_1^T, \epsilon_2^T, \dots, \epsilon_n^T)^T$$

$$G = \text{block-diagonal} \left(J_2(1, \omega_1), \dots, J_2(1, \omega_{m-1}) \right)$$

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

- P even number $P = 2m$

$$\omega_j = \frac{2\pi j}{P} \quad j = 1, \dots, m-1$$

$$\omega_m = \pi$$

$$\{F, G, \dots\}$$

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

$$G = \text{block-diagonal} \left(J_2(1, \omega_1), \dots, J_2(1, \omega_{m-1}), -1 \right)$$

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

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

Building NDLM's with multiple components

- Linear trend + seasonal component $P=4$

(a) linear trend

$$\{F_1, G_1, \dots\}$$

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

$$f_{1,6}(h) = K_{t,0} + K_{\epsilon,1h}$$

(b) seasonal component

$$P = 4 = 2 \cdot 2m$$

$$w_1 = \frac{2\pi}{4} = \frac{\pi}{2} \quad \{F_2, G_2, \dots\}$$

$$w_2 = \pi \quad F_2 = (E_2^\top, 1) = (1, 0, 1)^\top$$

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

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

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 $V_t = V$, V unknown

we will place a prior distribution on the variance

$$Y_t = F_t^T \theta_t + v_t, v_t \sim N(0, V)$$

$$\theta_t = G_t \theta_{t-1} + w_t, w_t \sim N(0, V_w)$$

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

$$(v | D_0) \sim IG(n_0/2, d_0/2), d_0 = n_0 s_0$$

Filtering and smoothing equations in closed-form.

Conditional Filtering equations:

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

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

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

$$a_t = G_t m_{t-1}, R_t^* = G_t C_{t-1}^* G_t^T + W_t^*$$

- $(y_t | D_{t-1}, v) \sim N(p_t, vq_t^*)$

$$p_t = F_t^T a_t, q_t^* = F_t^T R_t^* F_t + 1$$

posterior distribution for the variance

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

$$n_t = n_{t-1} + 1, \quad d_t = n_t s_t$$

$$S_t = S_{t-1} + \frac{S_{t-1}}{n_t} \left(\frac{\rho_t^2}{q_t} - 1 \right), \quad q_t = S_{t-1} q_t^*$$

$$e_t = y_t - p_t$$

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

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

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

we also have :

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

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

- $(y_t | D_{t-1}) \sim T_{n_{t-1}}(p_t, q_t), \quad q_t = S_{t-1} q_t^*$

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

Smoothing equations:

$$(\theta_t | D_T) \sim T_{n_T} (\alpha_T(t-T), R_T(t-T) s_T / s_t)$$

Forecasting

$$\bullet (\theta_{t+h} | D_t) \sim T_{n_t} (\alpha_t(h), R_t(h))$$

$$\alpha_t(h) = G_{t+h} \alpha_t(h-1)$$

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

$$\bullet (Y_{t+h} | D_t) \sim T_{n_t} (f_t(h), q_t(h))$$

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

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 known and W_t known

for all t

- V_t is known for all t

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

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

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

$$R_t = \frac{P_t}{\delta} \quad \text{with} \quad \delta \in [0, 1] \quad \begin{matrix} \text{discount} \\ \text{factor} \end{matrix}$$

δ close to 0 \Rightarrow more uncertainty

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

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 choose δ in practice?

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

We can use a one-step ahead predictive distributions

$$\underbrace{\log(P(Y_{1:T} | D_0, \delta))}_{l(\delta)} = \sum_{t=1}^T \log(P(Y_t | D_{t-1}, \delta))$$

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

"optimal" δ by maximizing $l(\delta)$

We can also consider

$$T^2 \text{SE}(\delta) = \sum_{t=1}^T e_t^2(\delta) / T \quad \begin{matrix} \text{to be} \\ \text{minimized} \end{matrix}$$

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

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

Code notes

NDLFT, unknown observational variance example

for filtering, smoothing and forecasting

Unknown V

set-up-dlm-matrices-unknown-v : pass arguments
 F_t, G_t, W_t -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.

n_t : degrees of freedom parameters

backward_smoothing_unknown_v: delta for providing a discount factor. No delta: W_t 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 : ECG data (Electroencephalogram)

Has been analysed in multiple papers.

We will compute the optimal discount factor

Time varying autoregressive model of order 12

analysis with 3200 observations first

data : higher frequency for lower quasi periodicity
at the beginning.

lower amplitude over time towards the end

G_t is constant over time.

F_t varies over time.

Co-star : prior variance of 10

No : 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)

β_{true} : 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 oscillates 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