

# 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  $\gamma$  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  $\phi$ , 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,

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

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

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

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  $\hat{\phi}_{mle}$ .

## Definition of the state-space representation

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

$\varepsilon_t \stackrel{iid}{\sim} N(0, V)$  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  $h$  (<sup>number of</sup> steps ahead) are the  $\alpha$  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:  $\phi_1, \dots, \phi_p, v$

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

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

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

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



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

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

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

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

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

## 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  $\pi$

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

e.g.  $\alpha_1$   $r = 0.7$

$\alpha_2$   $\lambda = 12$

periodicity

