Bayesian Time Series modelling and Analysis

6.12.2024

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

strong stationarity

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

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

weak stationarity (second order stationarity)

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

$E[y_t] = \mu$, $var(y_t) = v$
$cov(y_t, y_s)$ depends on $|t - s|$

Strong stationarity implies weak stationarity

autocorrelation function (ACF)

$\{Y_t\}$

autocovariance

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

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

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

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

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

autocorrelation

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

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

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

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)

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

$$y_t = \phi(\phi y_{t-2} + \varepsilon_t) + \varepsilon_t$$

$$= \phi^2 y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t$$

$$= \dots$$

$$= \phi^k y_{t-k} + \sum_{j=0}^{k-1} \phi^j \varepsilon_{t-j}$$

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 dependont

(stationary)

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

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)

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

(we work will the conditional likelihood)

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

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

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

$$Y_{1:T}$$

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

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

$$Q^*(\phi)$$

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

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

This is the full likelihood.

Conditional likelihood (conditioned on the first obs)

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

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

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

$\Rightarrow$

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

if $\underline{X}$ is full rank, then

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

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

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

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

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

$$\varepsilon_t \overset{iid}{\sim} N(0, v)$$

general AR(p) setting

The AR characteristic polynomial is

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

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

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

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

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

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

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

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

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

$$= F^T E[\underline{X}_{t+h} \mid Y_{1:t}]$$

$$= F^T G\, E[\underline{X}_{t+h-1} \mid Y_{1:t}]$$

$$= F^T G^h E[\underline{X}_t \mid Y_{1:t}]$$

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

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

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

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

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

$$(y_t \mid y_{t-1}, \dots, y_{t-p}, \phi_1, \dots, \phi_p, v) \sim \text{Normal}$$

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

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

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

$$\Downarrow$$

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

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

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

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

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

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)

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

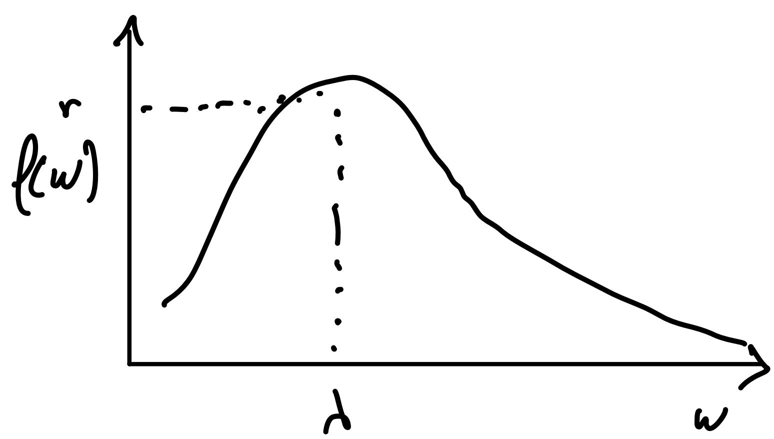
$$f(w) = \frac{v}{|(1-\phi_1 e^{-iw} - \dots - \phi_p e^{-iwp})|^2 \, 2\pi}$$

$f(w)$ ↓ frequency between 0 and π

⇒ related to the characteristic roots of the polynomial

eg. α_1 $r = 0.7$ ← modulus

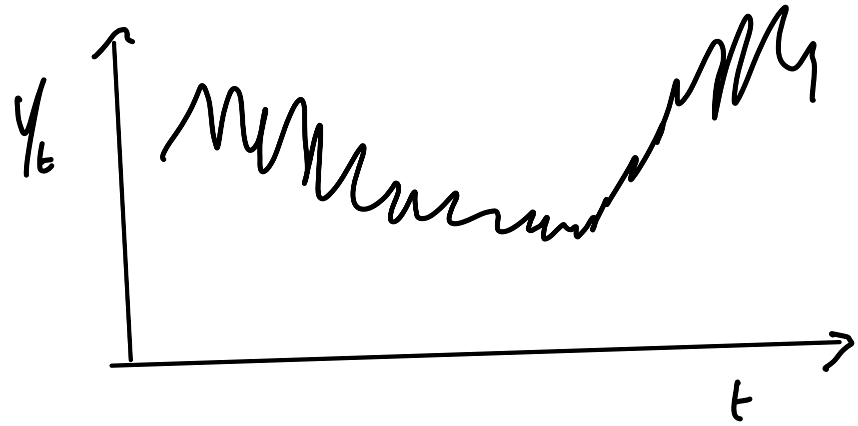
α_2 $\lambda = 12$ ↓ periodicity



The Normal Dynamic Linear Model (NDLM)

For modelling and analysing nonstationary time series.

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



$$E[Y_t] = \mu$$

$$V(Y_t) = V$$

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

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

(the mean level is changing over time)

The general class of DLM can be written as follows

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

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

$t < T$

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

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

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

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

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

$$G^h = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \qquad = (K_{t,1} + K_{t,2} h)$$

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

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

general structure that we want for the forecast function

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

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

$$\left.\begin{array}{l} \{F_1, G_1, \cdot, \cdot\} \\ F_1 = e_1 = (1,0)^T \qquad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \end{array}\right\} \text{ for } f_{1t}(h)$$

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

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

$F = (1, 0, X_t)$ $\qquad$ G = 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

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

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

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

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

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

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

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

$$var(Y_t \mid D_{t-1}) = F_t^T var(\theta_t \mid D_{t-1}) F_t + V_t$$

$$= F_t^T R_t F^T + V_t = q_t$$

posterior

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

prior

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

one step ahead forecast

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

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

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

  same as

  $(\theta_t \mid D_t)$

from Normal theory, we know that if we have:

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

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

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:

$$\begin{array}{ccccc} & Y_1 & Y_2 & Y_3 & Y_4 \\ \theta_0 & \theta_1 & \theta_2 & \theta_3 & \theta_4 \end{array}$$

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

forecast $f_1, q_1 \quad f_2, q_2$

using filtering

m_1, C_1 m_2, C_2 m_3, C_3

$a_3(-2), R_3(-2)$ $a_3(-1), R_3(-1)$ $a_3(0), R_3(0)$

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

$$t \leq T$$

recursions

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

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

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

<u>forecasting distributions</u> :

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

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

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

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

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

Fourier Representation (seasonal representation)

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

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

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

$$\begin{aligned} f_t(h) &= E_2^T [J_2(1,\omega)]^h \begin{pmatrix} a_t \\ b_t \end{pmatrix} \\ &= (1,0) \begin{pmatrix} \cos(\omega h) & \sin(\omega h) \\ -\sin(\omega h) & \cos(\omega h) \end{pmatrix} \begin{pmatrix} a_t \\ b_t \end{pmatrix} \\ &= a_t \cos(\omega h) + b_t \sin(\omega h) \\ &= \underset{\text{amplitude}}{A_t} \cos(\omega h + \underset{\text{phase}}{B_t}) \end{aligned}$$

- if $w = \pi$

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

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

p : period

- when p is an odd number

$$p = 2m - 1$$

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

Now, using the superposition principle

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

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

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

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

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

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

$$w_m = \pi$$

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

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

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

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

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

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

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

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

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

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