NOTAS DE AULA - VII

- Da mesma forma que no passo preditivo, o passo de atualização corresponde ao cálculo de uma integral, uma vez que corresponde ao teorema de Bayes:

$$\text{posterior} \propto \text{veross prior}$$

$$p(\alpha_t|Y_t) \propto p(y_t|\alpha_t)p(\alpha_t|Y_{t-1})$$

$$p(\alpha_t|Y_t) = k\,p(y_t|\alpha_t)p(\alpha_t|Y_{t-1}), \text{ onde k é tal que}$$

$$\int p(\alpha_t|Y_t)\,d\alpha_t = k\int p(y_t|\alpha_t)p(\alpha_t|Y_{t-1})d\alpha_t = 1, \text{ ou}$$

$$k^{-1} = \int p(y_t|\alpha_t)p(\alpha_t|Y_{t-1})d\alpha_t = p(y_t|Y_{t-1})$$

$p(y_t|Y_{t-1})$ é a densidade preditiva.

- No caso do ambiente Gaussiano esta densidade de atualização tem a mesma forma da densidade de previsão do estado, sendo ambas Gaussianas.

- Daí a necessidade de atualizar apenas a média e variância condicionais.

- Uma outra forma de apresentar as eqs do FK é fazer t=t+1 nas eqs. de previsão, e então substituir as eqs. de atualização, eliminando-as do FK:

$$a_{t+1|t} = (T_{t+1} - K_t Z_t) a_{t|t-1} + K_t y_t + (c_{t+1} - K_t d_t)$$
$$= T_{t+1} a_{t|t-1} + K_t \upsilon_t + c_{t+1}$$

$$P_{t+1|t} = T_{t+1}(P_{t|t-1} - P_{t|t-1} Z_t f_t^{-1} Z_t^{'} P_{t|t-1}) T'_{t+1} + R_{t+1} Q_{t+1} R'_{t+1}$$
$$= T_{t+1} P_{t+1} T'_{t+1} - K_t f_t K'_t + R_{t+1} Q_{t+1} R'_{t+1} \quad (eq.\ de\ Riccati)$$

$$onde\ \ K_t = T_{t+1} P_{t|t-1} Z_t f_t^{-1}\ \ é\ o\ ganho\ de\ Kalman.$$

- **Propriedades ótimas do FK:**

i. No ambiente Gaussiano e linear, o FK fornece a média condicional do vetor de estado, que é o estimador ótimo do estado no sentido de minimizar o MSE.

ii. Fora do ambiente Gaussiano e linear, o FK fornece o melhor estimador linear do vetor de estado, minimizando o MSE.

iii. A estimativa da matriz P_t, $P_{t+1|t,}$ não depende das observações, portanto será também a variância incondiconal do estado, podendo assim ser estimada *off-line.*

$\Rightarrow$ **Solução *steady state* da eq. de Riccati:**

- Para um sistema invariante no tempo, é possível obter uma solução para a equação de Ricatti.

$$P_{t+1|t} = T(P_{t|t-1} - P_{t|t-1} Z F_t^{-1} Z' P_{t|t-1})T' + RQR'$$

$$P_{t+1|t} = T(P_{t|t-1} - P_{t|t-1} Z (Z P_{t|t-1} Z' + h)^{-1} Z' P_{t|t-1})T' + RQR'$$

$Fazendo \quad P_{t+1|t} = P_{t|t-1} = \overline{P}$, *obtemos a eq. de Riccati algébrica* (ERA):

$$\overline{P} - T\overline{P}T' + T\overline{P}Z(Z'\overline{P}Z + h)^{-1} Z'\overline{P}T' - RQR' = 0$$

- A questão agora é saber sob que condições a ERA apresenta uma solução.

- Geralmente é difícil obter um solução explícita da ERA, e neste caso, saber se esta solução é única e positiva definida.

- Se existir uma solução da ERA, então isto pode ser utilizado para aumentar a eficiência computacional do FK:

$$\Rightarrow \mathrm{a}_{t+1|t} = \overline{\mathrm{T}}\mathrm{a}_{t|t-1} + \overline{\mathrm{K}}\upsilon_t ,$$

$$\overline{\mathrm{T}} = \mathrm{T} - \overline{\mathrm{K}}\mathrm{Z} \ \text{ e } \ \overline{\mathrm{K}} = \mathrm{T}\overline{\mathrm{P}}\mathrm{Z}\overline{\mathrm{F}}^{-1} = \mathrm{T}\overline{\mathrm{P}}\mathrm{Z}(\mathrm{Z}'\overline{\mathrm{P}}\mathrm{Z} + \mathrm{h})^{-1}$$

$$\Rightarrow \mathrm{P}_{t+1|t} = \overline{\mathrm{P}}$$

- Sob determinadas condições, pode-se mostrar que é possível estabelecer condições p/ solução da ERA:

i. Se o sistema é estacionário, $|\lambda_i(T)| < 1$, i= 1,...,m, e $P_{1|0}$ é psd, então: $\lim_{t\to\infty} P_{t+1|t} = \bar{P}$. Se $\bar{P}$ é única, a converg. é exponencial: $||P_{t+1|t} - P_{t|t-1}|| < \beta\alpha^{t}$; $|\beta|<1$ e $\alpha \in [0,1]$.

ii. Se o sistema é observável e $P_{1|0} - \bar{P}$ é pd, ou $P_{1|0} = \bar{P}$, então $\lim_{t\to\infty} P_{t+1|t} = \bar{P}$.

- Observe que para os ME, a segunda condição é que nos interessa, pois o sistema não é estacionário.

$\Rightarrow$ **Condições iniciais do FK:**

- Dependendo de forma do FK utilizado (1+1) ou (2 em 1) a condição inicial dirá respeito a:

$$\alpha_0 \sim N(a_0, P_0) \text{ ou}$$
$$\alpha_{1|0} \sim N(a_{1|0}, P_{1|0}).$$

- É também importante discriminar as componentes estacionárias das não estacionárias.

1. Condição inicial para componentes estacionárias

Inicialmente, considere o seguinte modelo:

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2).$$
$$\mu_t = \phi\mu_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2). \ |\phi| < 1.$$

A distr. incondicional de μ_t será normal, com:

$$E(\mu_t) = 0$$
$$Var(\mu_t) = \phi^2 Var(\mu_t) + \sigma_\eta^2$$
$$Var(\mu_t) = \sigma_\eta^2 / (1 - \phi^2).$$

Portanto é natural que inicializemos o FK com

$$a_0 = 0, \ p_0 = \frac{\sigma_\eta^2}{1 - \phi^2}.$$

- Numa situação mais geral, teremos:

$$\alpha_t = T\alpha_{t-1} + c + R\eta_t, \ \ assu\min do \ que$$
$$|\lambda_i \text{(T)}|\text{<1},\forall\text{=i, i=1,...,m.}$$

Calculando a média e variância incondicional:

$$E(\alpha_t) = T\,E(\alpha_{t-1}) + c$$
$$E(\alpha_t)(I-T) = c \ \ \therefore \ E(\alpha_t) = (I-T)^{-1}c.$$

$$Var(\alpha_t) = T \ Var(\alpha_{t-1})T' + RQR'$$
$$Vec(Var(\alpha_t)) = Vec(T \ Var(\alpha_{t-1})T' + Vec(RQR^*)$$
$$= (T \otimes T)Vec[Var \ \alpha_t] + R \otimes R \ Vec \ Q.$$

$(usando \ que:$

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \rightarrow Vec(A) = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}$$

$$Vec(ABC) = (C \otimes A') \ Vec \ B)$$

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & a_{1m}B \\ \vdots & \vdots & \vdots \\ a_{n1}B & a_{n2}B & a_{nm}B \end{pmatrix}_{mpxnq}$$

- Portanto, segue que:

$$\alpha_0 \sim N(a_0, P_0) \text{ ou}$$
$$a_0 = (I - T)^{-1} c \quad \text{e}$$
$$\text{Vect } P_0 = (I - T \otimes T)^{-1} (R \otimes R') \text{ Vec } Q.$$

2. Condição inicial para componentes não estacionárias

- Esta é a situação mais relevante p/ ME, onde a maioria das componentes são não estacionárias.

- Existem vários algoritmos disponíveis na literatura, mas iremos apenas considerar a solução da priori não informativa ou difusa: representa a situação de total ignorância sobre a distribuição dos valores possíveis de α; todos os valores são "equiprováveis".

$$\alpha_0 \sim N(a_0, P_0), \text{ com}$$
$$a_0 = 0 \text{ e}$$
$$P_0 = kI, \text{ sendo } k \text{ "muito grande".}$$

- Por conveniência computacional, geralmente o FK é parametrizado em termos da variância da eq. das observações, σ^2_ε ou simplesmente σ^2. A explicação virá qdo abordarmos a estimação dos hiperparâmetros via MV.

- Portanto, inicializando o FK com uma prior difusa, para o modelo de nível local, faz com que no segundo passo preditivo a distribuição já seja própria.

- Pode-se demonstrar, que em geral, se o vetor de estado possui “d” componentes não estacionárias, então em t=d+1, a distribuição a priori será própria.

- Portanto no MEB, apenas a partir de t=14 as inovações e distribuições serão bem definidas.

$\Rightarrow$ **Máxima Verossimilhança e decomposição pelo erro de previsão**

- Tipicamente alguns ou todos os elementos das matrizes do sistema são desconhecidos. Estes são aglutinados num vetor ψ, denominado do vetor de hiperparâmetros.

- Ex: considere o modelo de TTL com ciclo estocástico

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

$$T = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \rho\cos\lambda_c & \rho\sin\lambda_c \\ 0 & 0 & -\rho\sin\lambda_c & \rho\cos\lambda_c \end{pmatrix}$$

$$R = I_{(4x4)}$$

$$h = \sigma_\varepsilon^2 \qquad \sigma_k^2 = (1-\rho^2)\sigma_\psi^2$$

$$Q = \begin{pmatrix} \sigma_\eta^2 & & \underset{\sim}{0} & \\ & \sigma_\zeta^2 & & \\ \underset{\sim}{0} & & \sigma_k^2 & \\ & & & \sigma_k^2 \end{pmatrix}$$

$$\Psi = (\rho, \lambda_c, \sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\psi^2)$$

- Métodos de estimação dos hiperparâmetros:

  - máxima veross (MV)
  - algorítmo EM

- Cálculo da MV:

função de densidade conjunta:

$$f(y_t, y_{t-1}, ..., y_2, y_1 | \psi) = f(y_t | y_{t-1} y_2, .., y_1) f(y_{t-1}, y_{t-2}, ..., y_1),$$

usando que $P(A, B) = P(A|B)P(B)$.

Aplicando sucessivamente, temos que:

$$f(y_t, y_{t-1}, ..., y_2, y_1 | \psi) = \prod_{t=1}^{T} f(y_t | Y_{t-1}; \psi),$$

onde $f(y_t | Y_{t-1}; \psi)$ é a densidade preditiva, calculada pela solução da integral:

$$\begin{aligned} f(y_t | Y_{t-1}; \psi) &= \int f(y_t, \alpha_t | Y_{t-1}; \psi)\, d\alpha_t \\ &= \int f(y_t | \alpha_t, Y_{t-1}; \psi) f(\alpha_t | Y_{t-1}; \psi)\, d\alpha_t \\ &= \int f(y_t | \alpha_t) f(\alpha_t | Y_{t-1})\, d\alpha_t \end{aligned}$$

onde abandonamos a dependência em Ψ, por simplicidade de notação.

- A função verossimilhança é obtida diretamente a partir da função densidade conjunta, invertendo o seu argumento:

- função verossimilhança (sem concentrar):

$$L(\psi)=f(y_t,y_{t-1},...,y_2,y_1|\psi)=\prod_{t=d+1}^{T} f(y_t|\mathbf{Y}_{t-1};\psi),$$

onde d= n^o de componentes não estacionárias do vetor de estado.

>> Para o ambiente Gaussiano a densidade preditiva é dada por:

$f(y_t|Y_{t-1};\psi)\sim N(\hat{y}_{t|t-1}, F_t)$, onde

$\hat{y}_{t|t-1} = Z'_t a_{t|t-1} + d_t$

$F_t = Z'_t P_{t|t-1} Z_t + \sigma^2$ e o MEE é dado por:

(na notação do Koopman: $a_{t|t-1} = a_t$, $P_{t|t-1} = P_t$)

$$\begin{cases} y_t = Z_t \alpha_{t-1} + d_t + \varepsilon_t \\ \alpha_t = T_t \alpha_{t-1} + c_t + R_r \eta_t. \end{cases}$$

Portanto:

$$L(\psi)=\prod_{t=d+1}^{T} (2\pi)^{-1/2} F_t^{-1/2} \exp\left\{-\frac{1}{2}\upsilon_t F_t^{-1} \upsilon'_t\right\} = \prod_{t=d+1}^{T} (2\pi)^{-1/2} F_t^{-1/2} \exp\left\{-\frac{1}{2}\upsilon_t^2/F_t\right\}$$

$$= (2\pi)^{-(T-d)/2} \prod_{t=d+1}^{t} F_t^{-1/2} \exp\left\{-\frac{1}{2}\upsilon_t^2/F_t\right\},\ \text{tirando o log (base e):}$$

$$l(\psi) = \log\ L(\psi) = -\frac{(T-d)}{2}\log(2\pi) - \frac{1}{2}\sum_{t=d+1}^{T}\log F_t - \frac{1}{2}\sum_{t=d+1}^{T}\upsilon_t^2/F_t,$$

observando que $F_t = F_t(\psi)$ e $\upsilon_t = \upsilon_t(\psi)$.

- O problema de otimização (com restrições e não -linear) a ser resolvido consiste em determinar:

$$\hat{\Psi} = \text{máx } \log L(\Psi)$$

- Através de um artifício, pode-se utilizar um algoritmo de para otimização sem restrição, para resolver a otimização com restrição adotando transformações. Por exemplo;

- variâncias: $\sigma^2 = \exp(2\theta), \quad 0 < \sigma^2 < \infty$

$$\text{- ctes. de amortecimento: } \rho = \begin{cases} |\theta|(1+\theta^2)^{-1/2} & \text{p/ciclo}, 0 < \rho < 1 \\ \theta(1+\theta^2)^{-1/2} & \text{p/ AR(1)}, |\rho| < 1 \end{cases}$$

- frequência: $\lambda = 2\pi / (2 + \exp(\theta)), \quad 0 < \lambda < \pi$

Em todos estes casos $-\infty < \theta < \infty$, i.e., a otimização é efetuada sem restrição

$\Rightarrow$ Concentração da Verossimilhança

- A dimensão da busca na otimização pode ser diminuída em uma dimensão se reparametrizarmos o vetor de hiperparâmetros Ψ.

- Em particular utilizaremos que $\Psi=(\Psi^*,\sigma^2)$, considerando o modelo univariado. Geralmente $\sigma^2 = \text{Var}(\varepsilon_t)$.

- Como resultado, as equações do FK não envolverão σ^2, e assim a verossimilhança também não envolverá σ^2.

- Formalmente efetuando a nova parametrização nas expressões das variâncias do modelo:

Nova parametrização:

- $Q_t^* = Q\sigma_\varepsilon^2$ - $P_0^* = P_0\sigma_\varepsilon^2$
- $F_t^* = F_t\sigma_\varepsilon^2$ - $P_{t|t-1}^*(P_t^*) = P_{t|t-1}(P_t)\sigma_\varepsilon^2$

(entre parênteses a notação do Koopman)

Portanto as eqs do FK na nova parametrização, serão:

$F_t^* = Z'_t P_{t|t-1}^*(P_t^*)Z_t + \sigma_\varepsilon^2$

$\sigma_\varepsilon^2 F_t = Z'_t \sigma_\varepsilon^2 P_{t|t-1}(P_t)Z_t + \sigma_\varepsilon^2$

$F_t = Z_t' P_{t|t-1}(P_t)Z_t + 1$, não envolve σ_ε^2, apenas Ψ^*.

$\upsilon_t = y_t - Z'_t a_{t|t-1}(a_t)$, mas

$a_{t+1|t}(a_{t+1}) = T_{t+1} a_{t|t-1}(a_t) + K_t \upsilon_t$ e por sua vez usando a nova parametrização, chegamos à seguinte expressão p/ K_t:

$K_t = T_{t+1} P_{t|t-1}^*(P_t^*)Z_t F_t^{*-1} = T_{t+1} P_{t|t-1}(P_t)\sigma_\varepsilon^2 Z_t F_t^{-1} \sigma_\varepsilon^{-2} = T_{t+1} P_{t|t-1}(P_t)Z_t F_t^{-1}$.

Ou seja, ambos υ_t e F_t, na nova parametrização, não envolverão σ_ε^2. Consequência: ao maximizarmos a veross. na nova parametrização os estimadores de Ψ^* não envolverão σ_ε^2.

- Em seguida obtemos a verossimilhança na nova parametrização:

$$l(\psi^*,\sigma_\varepsilon^2)= \log L(\psi^*,\sigma_\varepsilon^2) = -\frac{(T-d)}{2}\log 2\pi - \frac{1}{2}\sum_{t=d+1}^{T}\log F_t^* - \frac{1}{2}\sum_{t=d+1}^{T}\upsilon_t^2/F_t^*$$

$$F_t^* = F_t\sigma_\varepsilon^2$$

$$l(\psi^*,\sigma_\varepsilon^2)= -\frac{(T-d)}{2}\log 2\pi - \frac{(T-d)}{2}\log\sigma_\varepsilon^2 - \frac{1}{2}\sum_{t=d+1}^{T}\log F_t - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=d+1}^{T}\upsilon_t^2/F_t$$

Calculando $\dfrac{\partial L}{\partial\sigma_\varepsilon^2} = -\dfrac{(T-d)}{2\sigma_\varepsilon^2} + \dfrac{1}{2(\sigma_\varepsilon^2)^2}\displaystyle\sum_{t=d+1}^{T}\upsilon_t^2/F_t = 0$

chegamos ao estimador de MV de σ_ε^2 sem necessidade de, diretamente, realizar umabusca não linear, pois:

$$\hat{\sigma}_\varepsilon^2(\hat{\psi}^*)=\frac{1}{T-d}\sum_{t=d+1}^{T}\upsilon_t^2(\hat{\psi}^*)/F_t\ (\hat{\psi}^*).$$

- Finalmente, substituindo esta expressão na eq. da veross., chegamos à veross. concentrada:

$$l_c(\psi^*)= -\frac{(T-d)}{2}(\log 2\pi+1)- \frac{(T-d)}{2}\log\hat{\sigma}_\varepsilon^2(\psi^*) - \frac{1}{2}\sum_{t=d+1}^{T}\log F_t\ (\psi^*)$$