NOTAS DE AULA - VI

⇒ Modelos em Espaço de Estados

- O tratamento estatístico dos ME está baseado na forma de espaço de estados (EE).

- Uma vez que um modelo estatístico é colocado na forma de EE, então o FK fornece as estimativas do estado e dos hiperparâmetros do modelo.

- A forma em espaço de estados (EE) é uma maneira específica de de escrever modelo lineares a tempo discreto, sendo definida através de duas equações estocásticas de diferenças. Considere a sua forma univariada:

$-eq\ das\ observações\ ou\ das\ medidas:$

$$y_t = Z_t^{'}\alpha_t + d_t + \varepsilon_t, \qquad t = 1, 2, ..., T$$

$-eq.\ do\ estado,\ do\ sistema\ ou\ de\ transição:$

$$\alpha_{t+1} = T_t\alpha_t + c_t + R_t\eta_t$$

$onde:$

$-\ \alpha_t\ (m\,x1),\ é\ o\ vetor\ de\ estado,\ não\ observável$

$-\varepsilon_t\ \sim NID(0, H_t)$

$-\eta_t\ (\mathrm{g\,x1}) \sim NID(0, Q_t)$

$-\alpha_0 \sim N(a_0, P_0)$

$-E[\varepsilon_t\eta_t'] = 0, \quad \forall s, t$

$-E(\varepsilon_t\alpha_0') = E(\eta_t\alpha_0') = 0, \quad \forall t$

$-\ as\ matrizes\ do\ sistema: Z_t\ (mx1);\ d_t (1x1); T_t\ \ (m\,x\,m); c_t\ (m\,x1),$
$R_t\ \ (m\,x\,g), h_t\ (1x1)\ \ e\ Q_t\ \ (g\,x\,g)\ \ são\, não\ estocásticas,\ podendo$
$conter\ elementos\ desconhecidos.$

- Muitos modelos estatísticos podem ser colocados na forma de EE: regressão linear, ARMA, ME, etc.

- É importante observar que a forma de EE não é única para um dado modelo estatístico.

- Entretanto se o estado entre duas formas de EE diferentes seguem uma transformação linear e inversível, i.e.,

$$\alpha_t^* = B\alpha_t, \; B(m \times m), \text{tal que } \det(B) \neq 0$$

então, pode-se demonstrar que a verossimilhança e a função de previsão permanecerão invariantes.

- Exemplos:

1. $y_t \sim AR(2)$: $y_t = \phi_1 y_{t-1} + \phi y_{t-2} + \varepsilon_t$

i) $y_t = (1 \quad 0)\begin{pmatrix} y_t \\ \phi_2 y_{t-1} \end{pmatrix}$

$$\begin{pmatrix} y_t \\ \phi_2 y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix}\begin{pmatrix} y_{t-1} \\ \phi_2 y_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\varepsilon_t.$$

ii) $y_t = (1 \quad 0)\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}$

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\varepsilon_t$$

- Observe que nestas duas formas de EE, a relação entre os vetores de estado é dada por:

$$\alpha_t^* = \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1/\phi_2 \end{pmatrix} \begin{pmatrix} y_t \\ \phi_2 y_{t-1} \end{pmatrix}, \text{ e que}$$

$$\det B = 1/\phi_2 \neq 0.$$

2. $y_t \sim AR(p)$

$$y_t = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix}$$

$$\begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ 0 & 0 & & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \varepsilon_t$$

3. $y_t \sim MA(1)$: $y_t = \varepsilon_t + \theta \varepsilon_{t-1}$

i. $y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ \theta \varepsilon_t \end{pmatrix}$

$$\begin{pmatrix} y_t \\ \theta \varepsilon_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \theta \varepsilon_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ \theta \end{pmatrix} \varepsilon_t$$

ii. $y_t = \begin{pmatrix} 1 & \theta \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix}$

$$\begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \varepsilon_t$$

- Finalmente, pode-se demonstrar que um modelo ARMA(p,q) pode ser colocado sob as seguintes forma de EE:

\- <u>forma 1</u>: primeiro mostre que uma ARMA(p,q) pode ser escrito como ARMA(m,m-1), onde m= máx(p,q+1), com θ_j=0, i >q e ϕ_i =0, i>p.

$$y_t = \phi_1 y_t + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

$$y_t = \begin{pmatrix} 1 & \theta_1 & \theta_2 & \ldots & \theta_{m-1} \end{pmatrix} \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \\ \vdots \\ \alpha_{m,t} \end{pmatrix}$$

$$\begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \\ \vdots \\ \alpha_{m,t} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_{m-1} & \phi_m \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \\ \alpha_{3,t-1} \\ \vdots \\ \alpha_{m,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- Por exemplo, y ~ ARMA(2,1)

$$y_t = \phi_1 y_t + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \qquad (I)$$

$$m = \text{máx}(2,2) = 2$$

$$y_t = (1 \;\; \theta_1)\begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{pmatrix}$$

$$\begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$$

i. $y_t = \alpha_{1,t} + \theta_1 \alpha_{2,t} = (1 + \theta_1 L)\alpha_{1,t}$, usando ii.

ii. $\alpha_{2,t} = \alpha_{1,t-1}$

iii. $\alpha_{1,t} = \phi_1 \alpha_{1,t-1} + \phi_2 \alpha_{2,t-1} + \varepsilon_t \therefore (1 + \theta_1 L + \theta_2 L^2)\alpha_{1,t} = \varepsilon_t$,

usando i.

ou

$$\alpha_{1,t} = \varepsilon_t \,/\, (1 + \theta_1 L + \theta_2 L^2).$$

Subst. em i, chegamos a (I).

- <u>forma 2:</u> primeiro mostre que uma ARMA(p,q) pode ser escrito como ARMA(m,m), onde m= max(p,q+1), com θ_j=0, i >q e ϕ_i =0, i >p.

$$y_t = \phi_1 y_t + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

$$y_t = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \end{pmatrix} \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \vdots \\ \alpha_{m,t} \end{bmatrix}_{mx1}$$

$$\begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \vdots \\ \alpha_{m,t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 & 0 & 0 & \vdots & 0 \\ \phi_2 & 0 & 1 & 0 & \vdots & 0 \\ \ldots & \ldots & \ldots & \ldots & \vdots & 1 \\ \phi_m & 0 & 0 & 0 & \vdots & 0 \end{bmatrix} \begin{bmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \\ \vdots \\ \alpha_{m,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{m-1} \end{bmatrix}_{mx1} \varepsilon_t$$

Por exemplo, y ~ ARMA(2,1):

$$y_t = \phi_1 y_t + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \qquad \text{(I)}$$

$$m = \text{mín}(2,2) = 2$$

$$y_t = (1 \quad 0)\begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{pmatrix}$$

$$\begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{pmatrix} = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix}\begin{pmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ \theta_1 \end{pmatrix}\varepsilon_t$$

i. $y_t = \alpha_{1,t}$

ii. $\alpha_{1,t} = \phi_1 \alpha_{1,t-1} + \alpha_{2,t-1} + \varepsilon_t$. Usando i, segue que

$y_t = \phi_1 y_{t-1} + \alpha_{2,t-1} + \varepsilon_t$

iii. $\alpha_{2,t} = \phi_2 \alpha_{1,t-1} + \theta_1 \varepsilon_t = \phi_2 y_{t-1} + \theta_1 \varepsilon$, usando i.

Susbtituindo iii em ii, obtemos (I).

⇒ **Propriedades de Sistemas Lineares Dinâmicos**

- Nem todo modelo que é colocado na forma EE possui boas propriedades estatísticas.

- Por exemplo, pode ser que um determinado elemento do vetor de estado não seja estimável (identificável) a partir dos dados.

- As propriedades que apresentaremos a seguir são estabelecidas, verificando-se algumas características das matrizes do sistema.

- Para definir estas propriedades é necessário assumir que o sistema é invariante no tempo, isto é,que todas as matrizes do sistema não dependam explicitamente do tempo.

1. <u>Estacionariedade</u> = a estacionariedade de y está intimamente relacionada à estacionariedade do processo do vetor de estado.

- Em seguida temos que estabelecer sob que condições a eq. do sistema, que se trata de um processo autoregressivo vetorial de ordem 1, é estacionária.

- Calculando a média e variância para o vetor de estado:

$$\alpha_t = T^t\alpha_0 + \sum_{j=0}^{t-1} T^j \upsilon_{t-j}, \qquad \upsilon_t = R\eta_t$$

$$E(\alpha_t) = T^t\ a_0$$

$$Var(\alpha_t) = E[(\alpha_t - E(\alpha_t))(\alpha_t - E(\alpha_t))']$$

$$= T^t P_o T^t{}' + \sum_{j=0}^{t-1} T^j W T^j{}', \quad W = RQR'$$

- Fica óbvio que ambas são funções explícitas de t. Entretanto se t$\rightarrow\infty$, está dependência pode desaparecer, dependendo dos elementos da matriz T (estacionariedade assintótica).

- Para investigar este comportamento iremos utilizar a decomposição de T:

$$T = F\Lambda F^{-1} \qquad \text{(i)}$$

, onde:

– F é a matriz de autovetores (F_i);

– $\Lambda = \text{diag}(\lambda_i),\ i = 1,\dots,m$, é a matriz de autovalores;

os quais satisfazem $TF_i = \lambda_i F_i$.

- Iremos agora utilizar o resultado em (i) para obter uma expressão para T^t em termos dos seus autovalores e autovetores:

$$T = F\Lambda F^{-1}$$
$$T^j = (F\Lambda F^{-1})^j = F\,\Lambda^j\,F^{-1}, \text{ onde } \Lambda^j = \operatorname{diag}(\lambda_i^j),\ i = 1,\ldots,m$$

- Para simplificar as demonstrações vamos admitir que m=2:

$$T^j = F\,\Lambda^j\,F^{-1}$$
$$= \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} \begin{pmatrix} \lambda_1^j & 0 \\ 0 & \lambda_2^j \end{pmatrix} \begin{pmatrix} f^{11} & f^{12} \\ f^{21} & f^{22} \end{pmatrix}$$
$$= \begin{pmatrix} f_{11}f^{11}\lambda_1^j + f_{12}f^{21}\lambda_2^j & f_{11}f^{12}\lambda_1^j + f_{12}f^{22}\lambda_2^j \\ f_{21}f^{11}\lambda_1^j + f_{22}f^{21}\lambda_2^j & f_{21}f^{12}\lambda_1^j + f_{22}f^{22}\lambda_2^j \end{pmatrix}$$

- Usando esta expressão podemos calcular, por exemplo:

$$E(\alpha_t) = T^t\,a_0 = \begin{pmatrix} f_{11}f^{11}\lambda_1^t + f_{12}f^{21}\lambda_2^t & f_{11}f^{12}\lambda_1^t + f_{12}f^{22}\lambda_2^t \\ f_{21}f^{11}\lambda_1^t + f_{22}f^{21}\lambda_2^t & f_{21}f^{12}\lambda_1^t + f_{22}f^{22}\lambda_2^t \end{pmatrix} \begin{pmatrix} a_{0,1} \\ a_{0,2} \end{pmatrix}.$$

• Isolando o primeiro elemento deste valor esperado:

$$E(\alpha_{1t}) = (f_{11} f^{11} \lambda_1^t + f_{12} f^{21} \lambda_2^t) a_{0,1} + (f_{11} f^{12} \lambda_1^t + f_{12} f^{22} \lambda_2^t) a_{0,2}$$

• Se $|\lambda_i(T)| < 1$, i=1,2, então qdo $t \to \infty$, a média incondicional torna-se zero, portanto, independente de t. O mesmo comportamento pode ser observado para a matriz variância covariância e FAC.

• Portanto a condição suficiente para estacionariedade do processo do vetor de estado é que ao autovalores da matriz T sejam todos menores do que um, em módulo.

• Se isto ocorrer, necessariamente, o processo para y também será estacionário.

• O ME de TLL é não estacionário, enquanto o de ciclo é estacionário.

$$\begin{pmatrix} \vdots \\ Z'T^{m-1} \end{pmatrix}$$

2. Observabilidade=

- Esta propriedade investiga se todos os elementos do vetor de estado podem ser estimados em um tempo t arbitrário a partir do conhecimento de y_t:

$$E(\alpha_t) = TE(\alpha_{t-1}) \therefore E(\alpha_t) = T^t\, a_0$$

$$Mas\ E(y_t) = Z'E(\alpha_t) = Z'T^t\, a_0$$

$$\therefore\ \mathrm{E}(y_0) = Z'a_0$$

$$\mathrm{E}(y_1) = Z'Ta_0$$

$$\mathrm{E}(y_2) = Z'T^2a_0$$

$$E\,(y_{m-1}) = Z'T^{m-1}a_0.\ \ Agrupando\ vetorialmente:$$

$$E\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{m-1} \end{pmatrix} = \begin{pmatrix} Z' \\ Z'T \\ Z'T^2 \\ \vdots \\ Z'T^{m-1} \end{pmatrix} a_0 = M\begin{pmatrix} a_{0,1} \\ a_{0,2} \\ \vdots \\ a_{0,m} \end{pmatrix}$$

$$E(Y_m) = Ma_0 \therefore a_0 = M^{-1}E(Y_m).$$

$$\acute{E}\ necess\acute{a}rio\ e\ suficiente,\ pois,\ que\ a\ matriz\ M,\ seja\ invert\acute{\imath}vel \therefore$$
$$posto(M) = m \text{ ou } \det(M) \neq 0.$$

- Exemplo1 : considere o modelo LLT.

$$Z' = (1,0), \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$M = \begin{pmatrix} Z' \\ Z'T \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad \det M = 1 \Rightarrow \text{sistema é observável.}$$

- Exemplo 2:

$$y_t = \mu_t + \beta_t + \varepsilon_t \qquad y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \varepsilon_t$$

$$\mu_t = \mu_{t-1} + \eta_t \qquad \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix}$$

$$\beta_t = \beta_{t-1} + \xi_t$$

$$Z' = (1,1), \quad T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$M = \begin{pmatrix} Z' \\ Z'T \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \det M = 0 \Rightarrow \text{sistema é não observável.}$$

Ou seja um dos elementos do vetor de estado é não identificável a partir dos dados.

- Reparametrizando o sistema, a não identificabilidade ficará mais clara:

Sejam:

$$\theta_t \triangleq \mu_t + \beta_t \qquad \delta_{1t} = \eta_t + \xi_t$$

$\psi_t \triangleq \mu_t - \beta_t \qquad \delta_{2t} = \eta_t - \xi_t$, ou seja a reparametrização é do tipo:

$$\alpha_t^* = B\alpha_t,\ B(m \, x \, m),\ \text{onde}\ \det(B) \neq 0.$$

$$\begin{pmatrix} \theta_t \\ \psi_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}$$

$\therefore \quad y_t = \theta_t + \varepsilon_t$, mas

$\mu_t + \beta_t = (\mu_{t-1} + \beta_{t-1}) + (\eta_t + \xi_t)$, i.e.

$\theta_t = \theta_{t-1} + \delta_{1t}$,

Portanto o sistema pode ser re – escrito como:

$$\therefore \quad y_t = \theta_t + \varepsilon_t$$
$$\theta_t = \theta_{t-1} + \delta_{1t}$$
$$\psi_t = \psi_{t-1} + \delta_{2t}$$

Nesta parametrização fica claro que um dos elementos do vetor de estado é redundante.

$\Rightarrow$ **O filtro de Kalman**

- Dado o sistema linear dinâmico, ou a representação em EE, a questão é saber como estimar o vetor de estado dada a ST de observações:

$$y_t = Z'_t \alpha_t + d_t + \varepsilon_t$$
$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t$$

$$\textit{com as seguintes condições}:$$
$$\alpha_0 \sim N(a_0, P_0);$$
$$\varepsilon_t \sim N(0, h_t) \quad \eta_t \sim N(0, Q_t)$$
$$E(\varepsilon_t \eta'_s) = 0, \forall\ t, s$$
$$E(\xi_t \alpha'_0) = E(\eta'_t \alpha_0) = 0$$

- A resposta é o FK, que se trata de um algoritmo que fornece, recursivamente, a média e variância do vetor de estado, condicional nas observações, para todo t.

- Como o sistema é Gaussiano, estas informações serão suficientes para caracterizar toda a distribuição condicional do estado, dado as observações.

- É importante distinguir sob que conjunto de informação de Y estamos condicionando a distribuição do vetor de estado:

$$\mathrm{FK}\begin{cases} p(\alpha_t|Y_{t-1}) \rightarrow \text{previsão ou filtragem: distr. a priori.} \\ p(\alpha_t|Y_t) \rightarrow \text{atualização: distr. a posteriori} \end{cases}$$

$$p(\alpha_t|Y_T) \rightarrow \text{suavização (smoothing).}$$

- Seja a seguinte notação:

$$a_t = E(\alpha_t|Y_t)$$

$$P_t = E[(\alpha_t - a_t)(\alpha_t - a_t)'|Y_t] \rightarrow \text{matriz de variância ou MSE}$$

$$a_{t|t-1} = E(\alpha_t|Y_{t-1})$$

$$P_{t|t-1} = E[(\alpha_t - a_{t|t-1})(\alpha_t - a_{t|t-1})']|Y_{t-1}]$$

- O passo inicial do FK é calcular a média e variância no tempo t, condicional às observações até o instante t-1.

$$\alpha_t = T_t\alpha_{t-1} + C_t + R_t\eta_t$$

$$E(\alpha_t|Y_{t-1}) = T_t\, E(\alpha_{t-1}|Y_{t-1}) + c_t + R_t\, E(\eta_t|Y_{t-1})$$

$$a_{t|t-1} = T_t a_{t-1} + c_t$$

$$P_{t|t-1} = E[(\alpha_t - a_{t|t-1})(\alpha_t - a_{t|t-1})'|Y_{t-1}] = T_t P_{t-1} T_t' + R_t Q_t R_t'$$

- Ou seja: $(\alpha_t|Y_{t-1}) \sim N(a_{t|t-1}, P_{t|t-1})$.

- Este resultado pode ser obtido de forma genérica através da solução da seguinte integral:

$$
\begin{aligned}
p(\alpha_t|Y_{t-1}) &= \int p(\alpha_t,\alpha_{t-1}|Y_{t-1})d\alpha_{t-1} \\
&= \int p(\alpha_t|\alpha_{t-1})p(\alpha_{t-1}|Y_{t-1})d\alpha_{t-1}
\end{aligned}
$$

- No caso do ambiente Gaussiano a solução desta integral coincide com o resultado anterior.

- Para outros tipos de distribuição dos choques, geralmente não existirá solução analítica para esta integral, exigindo a utilização de métodos numéricos ou aproximações analíticas.

- A partir dos anos 90 ficou bastante em voga a solução destas integrais por métodos de simulação Monte Carlo, como, por exemplo, MCMC, muito utilizado pelos Bayesianos.

- Seja a definição de inovação em modelos de EE:

$$
\begin{aligned}
\upsilon_t &= y_t - \hat{y}_{t|t-1} = y_t - E(y_t|Y_{t-1}) \\
&= y_t - E(Z_t\alpha_t + d_r + \varepsilon_t|Y_{t-1}) \\
&= y_t - (Z'_t a_{t|t-1} + d_t) \\
&= (Z'_t \alpha_t + d_t + \varepsilon_t) - (Z'_t a_{t|t-1} + d_t) \\
\upsilon_t &= Z'_t(\alpha_t - a_{t|t-1}) + \varepsilon_t
\end{aligned}
$$

- Pode-se provar que:

$$\begin{aligned}
&E(\upsilon_t) = 0\\
&E(\upsilon_t \upsilon_t') = f_t = Z'_t P_{t|t-1} Z_t + h_t\\
&E(\upsilon_t \upsilon_s') = 0, \quad t \neq s\\
&\upsilon_t \sim \text{Normal} \quad \therefore \quad \upsilon_t \sim NID(0, F_t).
\end{aligned}$$

- Ao tornar-se disponível no tempo t, a observação y_t é incorporada na estimação do vetor de estado, resultando nas equações de atualização:

$$\begin{aligned}
a_t &= a_{t|t-1} + P_{t|t-1} Z_t f_t^{-1} \upsilon_t\\
P_t &= P_{t|t-1} - P_{t|t-1} Z_t f_t^{-1} Z_t' P_{t|t-1},\\
f_t &= Z_t' P_{t|t-1} Z_t + h_t
\end{aligned}$$

$$ou \ seja \ (\alpha_t \mid Y_t) \sim N(a_t, P_t).$$

- Prova por indução:

$$t = 1$$

$$eq.\ estado: \ \alpha_1 = T_1 \alpha_0 + c_1 + R_1 \eta_1, \ donde \ segue \ que:$$

$$\begin{aligned}
a_{1|0} &= T_1 a_0 + c_1\\
P_{1|0} &= T_1 P_0 T_0' + R_1 Q_1 R_1'.
\end{aligned}$$

$$Portanto: \quad \alpha_1 = a_{1|0} + (\alpha_1 - a_{1|0}) \sim Normal \ \text{mult.}$$

- Trabalhando agora com a eq. das observações:

$$eq.\ obs:\ y_1 = Z'_1\alpha_1 + d_1 + \varepsilon_1 = Z_1^{'}a_{1|0} + d_1 + Z_1(\alpha_1 - a_{1|0}) + \varepsilon_1$$

$$\hat{y}_{1|0} = Z'_1 a_{1|0} + d_1$$

$$\begin{aligned} MSE(y_1 - \hat{y}_{1|0} \mid I_o) &= E[(y_1 - Z'_1 a_{1|0} - d_1)(y_1 - Z'_1 a_{1|0} - d_1)'] \\ &= E[(Z'_1(\alpha_1 - a_{1|0}) + \varepsilon_1)(Z'_1(\alpha_1 - a_{1|0}) + \varepsilon_1)'] \\ &= Z'_1 P_{1|0} Z_1 + h_1 \\ &= f_1 \end{aligned}$$

$$\begin{aligned} Cov[(\alpha_1, y_1) \mid I_0] &= E[(\alpha_1 - a_{1|0})(y_1 - Z'_1 a_{1|0} - d_1)'] = \\ &= E[(\alpha_1 - a_{1|0})(Z'_1(\alpha_1 - a_{1|0}) + \varepsilon_1)'] \\ &= E[(\alpha_1 - a_{1|0})(\alpha_1 - a_{1|0})' Z_1] + E[(\alpha_1 - a_{1|0})\varepsilon'_1] \\ &= P_{1|0} Z_1 \end{aligned}$$

$$Portanto: \begin{pmatrix} \alpha_1 \\ y_1 \end{pmatrix} \mid I_o \sim N\left[\begin{pmatrix} a_{1|0} \\ Z'_{1|0} a_{1|0} + d_1 \end{pmatrix}, \begin{pmatrix} P_0 & P_{1|0} Z_1 \\ Z'_1 P_{1|0} & Z'_1 P_{1|0} Z_1 + h_1 \end{pmatrix}\right]$$

- Em seguida utilizamos um resultado da distribuição normal multivariada:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \ \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}\right),\ então,\ P(X \mid Y) \sim N(\mu_{X|Y}, \Sigma_{X/Y})$$

$$\mu_{X|Y} = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y)$$

$$\Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$$

- Estabelecendo as seguintes correspondências:

$$X = \alpha_1$$
$$Y = y_1$$
$$\Sigma_{XX} = P_{1|0} \qquad \mu_X = a_{1|0}$$
$$\Sigma_{YY} = Z'_1 P_{1|0} Z_1 + h_1 = F_1$$
$$\mu_Y = Z'_1 a_{1|0} + d_1$$
$$\Sigma_{XY} = P_{1/0} Z_1$$

- Segue que:

$$(\alpha_1 | Y_1) \sim N(a_1, P_1),\ onde$$

$$a_1 = a_{1|0} + P_{1|0}\, Z_1\, f_1^{-1}(y_1 - (Z'_1\, a_{1|0} + d_1))$$
$$P_1 = P_{1|0} - P_{1|0}\, Z_1\, f_1^{-1}\, Z'_1\, P_{1|0}$$

- Repetindo os passos acima p/ t= 2,3, ..., verificamos as fórmulas de atualização do FK.

- 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:

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

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

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

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

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

$p(y_t|Y_{t-1})$ 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:

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

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

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, o FK fornece o melhor estimador do vetor de estado, minimizando o MSE.

ii. Fora do ambiente Gaussiano, 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.

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

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

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

- 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:

$$\Rightarrow a_{t+1|t} = \bar{T}a_{t|t-1} + \bar{K}\upsilon_t,$$
$$\bar{T} = T - \bar{K}Z \text{ e } \bar{K} = T\bar{P}Z\bar{F}^{-1} = T\bar{P}Z(Z'\bar{P}Z + h)^{-1}$$

$$\Rightarrow P_{t+1|t} = \bar{P}$$

- Sob determinadas condições, entretanto, é possível estabelecer condições p/ solução da ERA:

i. Se o sistema é estacionário, $|\lambda_i(T)| < 1, \forall = i, i = 1,...,m,$ 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: $\|P_{t+1|t} - P_{t|t-1}\| \leq \beta\alpha^t$; $\beta \geq 1$ 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/modelos estacionários

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:

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

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:

$$A = \begin{pmatrix} a_{17} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \rightarrow Vec(A) = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}$$

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

$$A \otimes B = \begin{pmatrix} a_4 B & a_{12} B & a_{nm} B \\ \vdots & \vdots & \vdots \\ a_{n1} B & a_{n2} B & a_{nm} B \end{pmatrix}_{mpxnq} \quad)$$

- 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 compon./modelos não estacionários

- 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".

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

- 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.

- Exemplo: considere o ME de nível local

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim N(0,\sigma^2)$$
$$\mu_t = \mu_{t-1} + \eta_t \qquad \eta_t \sim N(0,\sigma^2_\eta)$$

$$a_{t+1|t} = a_{t|t-1} + k_t \upsilon_t$$
$$= a_{t|t-1} + k_t (y_t - a_{t|t-1})$$
$$= (1 - k_t)\, a_{t|t-1} + k_t y_t, \quad \text{(EWMA !)}$$

onde $k_t = p_{t|t-1} / f_t = p_{t|t-1} / (p_{t|t-1} + h_t)$

$$p_{t+1|t} = p_{t|t-1} - p^2_{t|t-1} / f_t + \sigma^2_\eta$$

Usando que $h_t = \sigma^2, \sigma^2_\eta = q\sigma^2, p_{t|t-1} = \sigma^2\, p_{t|t-1}$ segue que:

$$k_t = p_{t|t-1} / (p_{t|t-1} + 1)$$
$$p_{t+1|t} = p_{t|t-1} - [p^2_{t|t-1} / (1 + p_{t|t-1})] + q$$

prior difusa: $\alpha_{1|0} \sim N(a_{1|0} = 0, p_{1|0} = k,\ k \to \infty)$

$\to t = 1$, fazendo $k \to \infty$.

$$a_{2|1} = (1 - k_1)\, a_{1|0} + k_1 y_1 = [1 - p_{1|0} / (p_{1|0} + 1)] a_{1|0} + [p_{1|0} / (p_{1|0} + 1)] y_1$$
$$= [1 - k / (k+1)] 0 + [k / (k+1)] y_1 = y_1,$$

$$p_{2|1} = p_{1|0} - [p^2_{1|0} / (1 + p_{1|0})] + q = k - [k^2 / (1+k)] + q = 1 + q.$$

- 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

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

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

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

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

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

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

- Métodos de estimação dos hiperparâmetros:

  - máxima veross (MV)
  - algorítmo EM

- Cálculo da MV:

função de densidade conjunta:

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

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

Aplicando sucessivamente, temos que:

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

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:

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

Para o ambiente Gaussiano a densidade preditiva é fácilmente calculável, sendo dada por:

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

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

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

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

Portanto:

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

$$= (2\pi)^{-T/2} \prod_{t=1}^{t} f_t^{-1/2} \exp\left\{-\frac{1}{2}\upsilon_t f_t^{-1} \upsilon^1_t\right\}, \text{ tirando o log}$$

$$\log L(\psi|Y_t) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log f_t - \frac{1}{2}\sum_{t=1}^{t} \upsilon_t f_t^{-1} \upsilon'_t,$$

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 na solução de:

$$\text{máx} \log L(\Psi)$$

- Através de uma artifício de transformação, STAMP transforma o problema em otimização sem restrição, adotando as seguintes transformações:

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

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

$$-\text{frequência:}\; \lambda = 2\pi / (2 + \exp(\theta)),\;\; 0 < \lambda < \pi$$

$\Rightarrow$ Concentração da Verossimilhança

- A dimensão da busca na optimização 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.

- 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:

$$-Q \to Q\sigma^2 \qquad - p_0 \to p_0\sigma^2$$
$$-f_t \to f_t\sigma^2 \qquad - p_{t|t-1} \to p_{t|t-1}\sigma^2$$

Portanto as eqs do FK, serão:

$$f_t = Z'_t P_{t|t-1} Z_t + \sigma^2$$
$$\sigma_\varepsilon^2 f_t = Z'_t \sigma_\varepsilon^2 P_{t|t-1} Z_t + \sigma^2$$
$$f_t = Z_t' P_{t|t-1} Z_t + 1 \to \text{ não envolve } \sigma^2\text{, apenas } \Psi^*.$$

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

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

$$K_t = T_{t+1} P_{t|t-1} Z_t' f_t^{-1} = T_{t+1} P_{t|t-1} \sigma^2 Z' f_t^{-1} \sigma^{-2} = T_{t+1} P_{t|t-1} Z_t' f_t^{-1}.$$

Ou seja, υ_t não envolverá σ^2.

- Em seguida obtemos a verossimilhança na nova parametrização:

$$\log L(\psi^*,\sigma^2) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum \log f_t - \frac{1}{2}\sum \upsilon_t^2 / f_t$$

$$f_t \rightarrow f_t\sigma^2$$

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

Calculando $\dfrac{\partial L}{\partial\sigma^2} = -\dfrac{T}{2\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\sum \upsilon_t^2 / f_t = 0$

Chegamos à $\sigma^2(\psi^*) = \dfrac{1}{T-d}\displaystyle\sum_{t=d+1}^{T} \upsilon_t^2 / f_t.$

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

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