NOTAS DE AULA - VIII

- Exemplo: considere o ME de nível local

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

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

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 $k \to \infty$.

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

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

Cálculo da verossimilhança :

$$\psi = (\sigma_\eta^2, \sigma_\varepsilon^2) = (q\sigma_\varepsilon^2, \sigma_\varepsilon^2)$$

$$\psi = (q, \sigma_\varepsilon^2), \quad por\tan to \ \psi^* = q.$$

$$\log L_c(q) = -\frac{(T-1)}{2}\log(2\pi+1) - \frac{(T-1)}{2}\log\sigma_\varepsilon^2(q) - \frac{1}{2}\sum_{i=2}^{T}\log f_t(q)$$

onde :

$$f_t = p_{t|t-1}(q) + 1$$

$$p_{t+1|t} = \frac{p_{t|t-1}}{1 + p_{t|t-1}} + q.$$

- Um panorama s/ otimização não-linear

$$f(\psi) \rightarrow \text{função objetivo}$$

$$f(\psi) = \log L(\psi)$$

$$g(\psi) = \frac{\partial f}{\partial \psi} = \left(\frac{\partial f}{\partial \psi_1}, \frac{\partial f}{\partial \psi_2}, \ldots, \frac{\partial f}{\partial \psi_p} \right)' \rightarrow \text{px1(gradiente)}$$

$$H(\Psi) = \frac{\partial^2 f}{\partial \psi \partial \psi'} = \begin{pmatrix} \frac{\partial^2 f}{\partial \psi_1^2} & \frac{\partial^2 f}{\partial \psi_1 \psi_2} & & \frac{\partial^{f\,2}}{\partial \psi_1 \partial \psi_p} \\ \frac{\partial^2 f}{\partial \psi_1 \partial \psi_2} & \frac{\partial^2 f}{\partial \psi_2^2} & & \frac{\partial^2 f}{\partial \psi_1 \partial \psi_p} \\ \vdots & & \ddots & \\ \frac{\partial^2 f}{\partial \psi_1 \partial \psi_p} & \cdots & \cdots & \frac{\partial^2 f}{\partial \psi_p^2} \end{pmatrix} \rightarrow \text{Hessiano}$$

- Os algorítmos de otimização não linear baseada no método de quasi-Newton são baseados na expansão quadrática da função custo.

$$f(\psi) = f(\tilde{\psi}) + (\psi - \tilde{\psi})' g(\tilde{\psi}) + \frac{1}{2}(\psi - \tilde{\psi})' H(\tilde{\psi})(\psi - \tilde{\psi})$$

$$\frac{\partial f}{\partial \psi} = g(\psi) = g(\tilde{\psi}) + 1/2(H + H')(\psi - \tilde{\psi})$$

$$g(\psi) = g(\tilde{\psi}) + H(\psi - \tilde{\psi}) = 0, \; resolvendo \;\; p/\psi:$$

$$\psi = \tilde{\psi} - H^{-1}(\tilde{\psi}) g(\tilde{\psi})$$

$$\psi^{(i+1)} = \psi^{i} - H^{-1}(\psi^{i}) g(\psi^{i}), \;\; i = 1, 2, \ldots, niter$$

$$Obs: \partial(x' A x)/\partial x = (A + A')x.$$

- Problema: mesmo se a direção for de gradiente crescente, o passo tomado pode ser exagerado.

$$\psi^{(i+1)} = \psi^{(i)} - \lambda_i H^{-1}(\psi^{(i)}) g(\psi^{(i)}), \;\; i = 1,2,\ldots, \text{niter}$$

onde λ_i controla a extensão do passo, garantindo que $f(\psi^{(i+1)}) \geq f(\psi^{(i)})$.

- Quando a função custo é o log da verossimilhança, o método de scoring, o qual substitui o Hessiano pelo seu valor esperado torna-se:

$$\psi^{(i+1)} = \psi^{(i)} + \lambda_i I^{-1}(\psi^{(i)}) g(\psi^{(i)}), \ i = 1,2,\ldots, \text{niter}$$

$$E(H(\psi^{(i)})) = -E\left(\frac{\partial^2 \log L}{\partial\psi \partial\psi'}\right) = I(\psi^{(i)}) \ , \ \text{onde } I(\psi^{(i)})^{-1} \text{ é o}$$

limite inferior de Cramér – Rao para a matriz variância – covariância dos hiperparâmetros.

- Critérios de convergência: o processo de otimização é finalizado quando os três critérios abaixo são satisfeitos:

i. veross. $\quad \text{crit}_1 = |\ell(\psi_i) - \ell(\psi_{i+1})| \, / \, |\ell(\psi_i)| < \varepsilon$

ii. gradiente $\quad \text{crit}_2 = \frac{1}{p} \sum_{j=1}^{p} |g_j(\psi_i)| < 10\varepsilon$

iii. parâmetro $\quad \text{crit}_3 = \frac{1}{p} \sum_{j=1}^{t} \frac{|\psi_{i+1,j} - \psi_{i,j}|}{|\psi_{i,j}|} < 100\varepsilon$

- A mensagem final do processo de otimização pode tomar as seguintes formas:

	crit1	crit2	crit3
- very strong	< e	< e	< e
- strong	< e	< e	< 10e
- weak	< e	< 10e	< 10e
- very weak	< 10e	< 10e	< 10e

Exemplo da saída de STAMP:

```
Lairline = log(airline);
Method of estimation is Maximum likelihood
The present sample is: 49 (1) to 60 (12)

 MaxLik initialising...
it  1  f=   2.75996 e0=   0.30426 step=   1.00000
it  2  f=   2.76849 e0=   0.64547 step=   1.00000
it  3  f=   2.84211 e0=   0.26834 step=   1.00000
it  4  f=   2.87831 e0=   0.11764 step=   0.50000
it  5  f=   2.88073 e0=   0.04613 step=   0.50000

 MaxLik iterating...
it  4  f=   2.88365 df=   0.00005 e1=   0.00068 e2=   0.06698 step=   1.00000
it  8  f=   2.88366 df=   0.00000 e1=   0.00000 e2=   0.00000 step=   0.00105

Equation  1.

Lairline = Trend + Dummy seasonal + Irregular

Estimation report
Model with  4 parameters ( 2 restrictions).
Parameter estimation sample is   49. 1 -   60.12. (T =  144).
Log-likelihood kernel is 2.883664.
Very strong convergence in   8 iterations.
( likelihood cvg 2.772031e-015
  gradient cvg   6.377121e-008
  parameter cvg  2.360301e-009 )
```

Eq 1 : Estimated variances of disturbances.

Component	Lairline	(q-ratio)
Irr	0.00012951	(0.1852)
Lvl	0.00069945	(1.0000)
Slp	0.00000	(0.0000)
Sea	6.4129e-005	(0.0917)

Eq 1 : Estimated coefficients of final state vector.

Variable	Coefficient	R.m.s.e.	t-value	
Lvl	6.1809	0.016985	363.91	[0.0000]
Slp	0.0093707	0.0022176	4.2256	[0.0000]
Sea_ 1	-0.11016	0.015203	-7.2465	[0.0000]
Sea_ 2	-0.21568	0.013792	-15.638	[0.0000]
Sea_ 3	-0.069632	0.013744	-5.0664	[0.0000]
Sea_ 4	0.040004	0.013746	2.9102	[0.0042]
Sea_ 5	0.21936	0.013746	15.958	[0.0000]
Sea_ 6	0.23184	0.013744	16.868	[0.0000]
Sea_ 7	0.10554	0.013746	7.6779	[0.0000]
Sea_ 8	-0.0029544	0.013753	-0.21482	[0.8302]
Sea_ 9	-0.0024482	0.013766	-0.17784	[0.8591]
Sea_10	-0.014385	0.013781	-1.0438	[0.2985]
Sea_11	-0.11648	0.013786	-8.4487	[0.0000]

Eq 1 : Seasonal analysis (at end of period).

Seasonal Chi^2(11) test is 848.84 [0.0000].

	Seas 1	Seas 2	Seas 3	Seas 4	Seas 5
Value	-0.065006	-0.11648	-0.014385	-0.0024482	-0.0029544

	Seas 6	Seas 7	Seas 8	Seas 9	Seas 10
Value	0.10554	0.23184	0.21936	0.040004	-0.069632

	Seas 11	Seas 12
Value	-0.21568	-0.11016

- A vantagem deste método é que a saída da optimização não só oferece os valores ótimos, mas também a matriz variância, cujo elementos da diagonal são utilizados na construção de testes de hipótese, pois:

$$\sqrt{T}(\psi - \tilde{\psi}) \sim N(0, \mathrm{Asy\,var}(\tilde{\psi})),$$
$$\text{onde } \mathrm{Asy\,var}(\tilde{\psi}) = I^{-1}(\tilde{\psi}).$$

- Deve-se observar que este resultado só é válido se:

i. Ψ é um ponto interior do espaço paramétrico;

ii. as terceiras derivadas de logL existem e são contínuas na vizinhança do valor verdadeiro do parâmetro;

iii. Ψ é identificável.

- Na prática o gradiente e o Hessiano, podem ser obtidos numericamente ou analiticamente.

⇒Tratamento de observações faltantes

- Suponha que existe uma observação faltante em t=m. O que ocorrerá com o FK ?

eq. de previsão:

$$a_{m|m-1} = Ta_{m-1} + c_m$$

$$P_{m|m-1} = TP_{m-1}T' + RQR'$$

eq. de atualização: usaremos a seguinte substituição

$$a_m = a_{m|m-1}$$

$P_m = P_{m|m-1}$. Portanto em $t = m+1$, teremos:

$$a_{m+1|m-1} = Ta_m + c_{m+1} = Ta_{m|m-1} + c_{m+1} = T^2 a_{m-1} + Tc_{m+1} + c_m$$

$$\begin{aligned} P_{m+1|t} &= TP_m T' + RQR' \\ &= TP_{m|m-1}T' + RQR' \\ &= T(TP_{m-1}T' + RQR')T' + RQR' \\ &= T^2 P_{m-1} T'^2 + TRQR'T' + RQR'. \end{aligned}$$

Para estimar y em $t = m$, usaremos o algoritmo de suavização:

$$\hat{y}_m = E(y_{m|T}) = Ta_{m|T}.$$

Como a veross. é afetada?

- Não existe $p(y_m|Y_{m-1})$, pulamos e usamos

$$p(y_{m+1}|Y_m) = p(y_{m+1}|Y_{m-1})$$