Processos autoregressivos (parte 3)

Prof. Caio Azevedo

Introdução

- Vimos como estimar, supondo apenas a estacionariedade fraca, de um dado processo, sua média, função de auto-covariância (FACV) e função de auto-correlação (FAC) (aqui).
- Usualmente, quando outras características desse processo são levadas em consideração, por exemplo, no processo de estimação, podemos obter melhores resultados (menor(es), viés, variância, eqm).
- Vamos considerar, neste momento, processos AR(p).

Introdução

- Um processo estocástico $\{Y_t\}$ é dito ser um processo AR(p) (essencialmente) se:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t. \quad (1)$$

- Algumas propriedades foram discutidas (ao menos para casos particulares) aqui, aqui e aqui.
- Vamos s focar no caso em que o processo (1) é **estacionário**, **causal** e que $\epsilon_t \sim RB(0, \sigma^2)$.

Introdução

- Não é difícil verificar que $\mathcal{E}(Y_t) = 0, \forall t$.
- A versão do processo (1) de média μ ($\mathcal{E}(Y_t) = \mu, \forall t$) é dada por:

$$Y_t = \mu + \phi_1 (Y_{t-1} - \mu) + \cdots + \phi_p (Y_{t-p} - \mu) + \epsilon_t, \ \epsilon_t \sim RB(0, \sigma^2).$$

- Considere, inicialmente, o problema de estimar a FACV e FAC de um processo AR(p), com $\mu = 0$.

FAC

- Note que (multiplicando-se ambos os lados da primeira Equação (1) por Y_t):

$$
\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t \\
Y_t^2 &= \phi_1 Y_t Y_{t-1} + \cdots + \phi_p Y_t Y_{t-p} + Y_t \epsilon_t \\
E(Y_t^2) &= \phi_1 E(Y_t Y_{t-1}) + \cdots + \phi_p E(Y_t Y_{t-p}) + E(Y_t \epsilon_t) \\
\gamma(0) &= \phi_1 \gamma(1) + \cdots + \phi_p \gamma(p) + E(Y_t \epsilon_t). \qquad (2)
\end{aligned}
$$

- Lembrando que $\gamma(h)$ é a função de auto-covariância do processo.

FAC

- Além disso, como o processo é causal então $E(\epsilon_t Y_{t-k}) = 0$ para $k \geq 1$, logo (multiplicando-se ambos os lados da primeira Equação (1) por ϵ_t):

$$
\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t \\
\epsilon_t Y_t &= \phi_1 \epsilon_t Y_{t-1} + \cdots + \phi_p \epsilon_t Y_{t-p} + \epsilon_t^2 \\
E(\epsilon_t Y_t) &= \phi_1 E(\epsilon_t Y_{t-1}) + \cdots + E(\phi_p \epsilon_t Y_{t-p}) + E(\epsilon_t^2) \\
E(\epsilon_t Y_t) &= E(\epsilon_t^2) = \sigma^2. \qquad (3)
\end{aligned}
$$

FAC

- Portanto, aplicando (3) em (2), temos que:

$$
\begin{aligned}
\gamma(0) &= \phi_1\gamma(1) + \cdots + \phi_p\gamma(p) + \sigma^2 \\
\rightarrow 1 &= \phi_1\rho(1) + \cdots + \phi_p\rho(p) + \frac{\sigma^2}{\gamma(0)} \\
\rightarrow \gamma(0) &= \frac{\sigma^2}{1 - \phi_1\rho(1) - \cdots - \phi_p\rho(p)}. \qquad (4)
\end{aligned}
$$

FAC

- Por outro lado, temos que (multiplicando-se ambos os lados da Equação (5) por Y_{t-k} e dividindo-se ambos os lados da Equação (6) por $\gamma(0)$):

$$
\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t \\
Y_{t-k} Y_t &= \phi_1 Y_{t-k} Y_{t-1} + \cdots + \phi_p Y_{t-k} Y_{t-p} + Y_{t-k}\epsilon_t \qquad (5) \\
E(Y_{t-k} Y_t) &= \phi_1 E(Y_{t-k} Y_{t-1}) + \cdots + \phi_p E(Y_{t-k} Y_{t-p}) + E(Y_{t-k}\epsilon_t) \\
\gamma(k) &= \phi_1 \gamma(k-1) + \cdots + \phi_p \gamma(k-p) \qquad (6) \\
\rho(k) &= \phi_1 \rho(k-1) + \cdots + \phi_p \rho(k-p) \qquad (7)
\end{aligned}
$$

FAC

- Lembrando que, em um processo estacionário, $\rho(h) = \dfrac{\gamma(h)}{\gamma(0)}$.
- Em termos matriciais, a Equação (7), para $k = 1, \ldots, p$ temos que:

$$\begin{bmatrix} \rho(0) & \rho(1) & \ldots & \rho(p-1) \\ \rho(1) & \rho(0) & \ldots & \rho(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \ldots & \rho(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix} \quad (8)$$

FAC

- Esse sistema de equações (Equação (8)) é conhecido na literatura como equações de Yule-Walker.
- De posse de $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ e de $\boldsymbol{\rho} = (\rho(1), \ldots, \rho(p))'$ podemos obter σ^2 através de (4), ou seja:

$$\widehat{\sigma}^2 = \widehat{\phi}(0)\left(1 - \widehat{\phi}_1\widehat{\rho}(1) - ... - \widehat{\phi}_p\widehat{\rho}(p)\right). \tag{9}$$

- Como estimar $\boldsymbol{\rho}$ e $\boldsymbol{\phi}$?

Exemplo

- Considere o seguinte processo:

$$
\begin{aligned}
Y_t &= 0,33Y_{t-1} + 0,5Y_{t-2} + \epsilon_t,\ \epsilon_t \sim RB(0,\sigma^2) \\
&\rightarrow (1 - 0,33B - 0,5B^2)Y_t = \epsilon_t,
\end{aligned}
$$

- Fazendo $\phi(z) = 1 - 0,33z - 0,5z^2 = 0$ temos que $z = -1,78221$ e $z = 1,12221$.
- Logo o processo é estacionário e causal e, de (8), temos que:

$$\rho(k) = 0,33\rho(k-1) + 0,5\rho(k-2)$$

Exemplo

- Portanto, temos que as equações de Yule-Walker são dadas por:

$$\begin{aligned} \rho(1) &= 0,33 + 0,5\rho(1) \\ \rho(2) &= 0,33\rho(1) + 0,5. \end{aligned}$$

- Resolvendo o sistema de equações acima, temos que $\rho(1) = 0,66$ e $\rho(2) = 0,7178$ e portanto $\gamma(0) = 2,362391\sigma^2$ (da Equação (9)).

Exemplo

- Para resolver as equações em diferença

$$\rho(k) = 0,33\rho(k-1) + 0,5\rho(k-2),$$

precisamos calcular as raízes de $\phi(z) = 1 - 0,33z - 0,5z^2 = 0$, que são $m_1 = -1,78221$ e $m_2 = 1,12221$. Adicionalmente, como elas são reais e diferentes, temos que:

$$\rho(k) = c_1\left(\frac{1}{m_1}\right)^k + c_2\left(\frac{1}{m_2}\right)^k.$$

- Para obter c_1 e c_2 utilizamos as primeiras autocorrelações estimadas, ou seja $\rho(1) = 0,66$ e $\rho(2) = 0,7178$.

Exemplo

- Dito isso, temos que:

$$
\begin{aligned}
0,66 &= c_1\left(\frac{1}{m_1}\right)+c_2\left(\frac{1}{m_2}\right)\\
0,7178 &= c_1\left(\frac{1}{m_1}\right)^2+c_2\left(\frac{1}{m_2}\right)^2.
\end{aligned}
$$

- Resolvendo a equação acima temos que $c_1 = 0,159143$ e $c_2 = 0,840867$, portanto:

$$
\rho(k) = 0,159143\left(-\frac{1}{1,78221}\right)^k + 0,840867\left(\frac{1}{1,12221}\right)^k. \quad (10)
$$

- Através a Equação (10) é possível calcular as demais auto-correlações $(k \geq 2)$.

Estimação (usando o modelo) - Introdução

- Considere $\{Y_t\}$ um processo AR(p) com média μ, estacionário e causal:

  $$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t, \ \epsilon_t \sim RB(0, \sigma^2).$$

- Dada uma amostra $(y_1, \ldots, y_n)'$ (ST observada), como podemos estimar os parâmetros $\boldsymbol{\theta} = (\mu, \phi_1, \ldots, \phi_p, \sigma^2)' = \left(\mu, \boldsymbol{\phi}', \sigma^2\right)'$, em que $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$?

- Há algumas opções como : métodos dos momentos, mínimos quadrados condicionais, máxima verossimilhança e métodos bayesianos ref1, ref2. Discutiremos os três primeiros.

Estimação - Método dos momentos (MM)

- Consiste em igualar os momentos populacionais aos amostrais, resolvendo o respectivo sistema de equações resultante (aqui).
- Para μ, como este é a média do processo, o estimador pelo métodos dos momentos (EMM) é $\widehat{\mu} = \overline{Y}$.

Estimação - Método dos momentos (MM)

- Para $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ podemos usar o sistema de Yule-Walker, substituindo $\rho(1), \ldots, \rho(p)$ pelas autocorrelações amostrais $\widehat{\rho}(1), \ldots, \widehat{\rho}(p)$ de sorte que:

$$\begin{bmatrix} \widehat{\rho}(0) & \widehat{\rho}(1) & \ldots & \widehat{\rho}(p-1) \\ \widehat{\rho}(1) & \widehat{\rho}(0) & \ldots & \widehat{\rho}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\rho}(p-1) & \widehat{\rho}(p-2) & \ldots & \widehat{\rho}(0) \end{bmatrix} \begin{bmatrix} \widehat{\phi}_1 \\ \widehat{\phi}_2 \\ \vdots \\ \widehat{\phi}_p \end{bmatrix} = \begin{bmatrix} \widehat{\rho}(1) \\ \widehat{\rho}(2) \\ \vdots \\ \widehat{\rho}(p) \end{bmatrix},$$

Estimação - Método dos momentos

- Portanto, temos que o EMM para ϕ é dado por $\widehat{\boldsymbol{\phi}} = \boldsymbol{R}^{-1}\boldsymbol{r}$, em que $\widehat{\boldsymbol{\phi}} = (\widehat{\phi}_1, \ldots, \widehat{\phi}_p)^\top$, $\boldsymbol{r} = (\widehat{\rho}(1), \ldots, \widehat{\rho}(p))'$ e

$$\boldsymbol{R} = \begin{bmatrix} 1 & \widehat{\rho}(1) & \ldots & \widehat{\rho}(p-1) \\ \widehat{\rho}(1) & 1 & \ldots & \widehat{\rho}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\rho}(p-1) & \widehat{\rho}(p-2) & \ldots & 1 \end{bmatrix}.$$

Estimação - Método dos momentos - Exemplos

- **AR(1):** para o modelo AR(1) temos que $\widehat{\phi} = (1)^{-1}\widehat{\rho}(1) = \widehat{\rho}(1)$.
- **AR(2):** para o modelo AR(2) temos que:

$$
\begin{aligned}
\begin{bmatrix} \widehat{\phi}_1 \\ \widehat{\phi}_2 \end{bmatrix} &= \begin{bmatrix} 1 & \widehat{\rho}(1) \\ \widehat{\rho}(1) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\rho}(1) \\ \widehat{\rho}(2) \end{bmatrix} \\
&= \frac{1}{1-\widehat{\rho}(1)^2} \begin{bmatrix} 1 & -\widehat{\rho}(1) \\ -\widehat{\rho}(1) & 1 \end{bmatrix} \begin{bmatrix} \widehat{\rho}(1) \\ \widehat{\rho}(2) \end{bmatrix} \\
&= \frac{1}{1-\widehat{\rho}(1)^2} \begin{bmatrix} \widehat{\rho}(1) - \widehat{\rho}(1)\widehat{\rho}(2) \\ \widehat{\rho}(2) - \widehat{\rho}(1)^2 \end{bmatrix}.
\end{aligned}
$$

Estimação - Método dos momentos

- Para σ^2, uma vez que:

$$\gamma(0) = \frac{\sigma^2}{1 - \phi_1\rho(1) - \cdots - \phi_p\rho(p)},$$

então um estimador para σ^2 é:

$$\widehat{\sigma}^2 = S^2(1 - \widehat{\phi}_1\widehat{\rho}(1) - \cdots - \widehat{\phi}_p\widehat{\rho}(p)), \tag{11}$$

em que $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \widehat{\mu})^2$ é a variância amostral dos valores observados do processo.

Estimação - Método dos momentos

- Para obter outros resultados inferenciais (EP, IC, TH), o seguinte resultado assintótico é útil: sob certas condições de regularidade (aqui):

$$\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow[n\to\infty]{D} N(\mathbf{0}, \sigma^2 \boldsymbol{R}^{-1}).$$

- Note que, uma vez que estamos estimando parâmetros associados à um modelo de regressão, utilizando determinados métodos de estimação, os estimadores da média (μ) e da função de auto-covariância/ auto-correlação ($\gamma(h), \phi(h)$), podem ser diferentes daqueles propostos aqui.

Estimação - Mínimos quadrados condicionais (MQC)

- Assemelha-se ao método de mínimos quadrados (aqui), no sentido de minimizar uma soma de quadrados.
- Para um modelo AR(p) podemos escrever os erros como:

$$\epsilon_t = Y_t - \mu - \phi_1(Y_{t-1} - \mu) - \cdots - \phi_p(Y_{t-p} - \mu).$$

- Contudo, no método de MQC, considera-se apenas as os erros cujas esperanças condicionais de $Y_t|(Y_1, ..., Y_p)$ são observáveis ($t \geq p$).
- Por exemplo, se p $= 1$ (AR(1)), não se observa (Y_0). Logo, $Y_1 = \mu + \phi Y_0 + \epsilon_1$ e, assim $\mathcal{E}(Y_1|y_0) = \mu + \phi y_0$ não é observável.

Estimação - Mínimos quadrados condicionais (MQC)

- Logo, os estimadores de mínimos quadrados condicionais de $\boldsymbol{\theta} = (\mu, \phi_1, \ldots, \phi_p)' = (\mu, \boldsymbol{\phi}')'$ são obtidos minimizando:

$$S(\mu, \boldsymbol{\phi}) = \sum_{t=p+1}^{n} \epsilon_t^2,$$

em relação à $(\mu, \boldsymbol{\phi}')'$.

- Note que se utiliza $n - p$ observações no processo de estimação.

Estimação - Mínimos quadrados condicionais (MQC)

1. Se μ for conhecido, os estimadores de MQC de ϕ, possuem forma fechada.

2. Se μ for desconhecido, os estimadores de MQC de ϕ não possuem forma explícita. Assim, ou métodos numéricos (resolução de sistemas de equações não lineares, como Newton-Raphson, Escore de Fisher, BFGS etc), ou aproximações analíticas/assintóticas como expansão em série de Taylor, convergência estocástica, devem ser utilizados.

3. Uma alternativa é utilizar o EMM de μ ($\widehat{\mu} = \overline{Y}$), na expressão de $\widehat{\phi}$, mencionada no item 1).

4. Para mais detalhes veja aqui e aqui.

Estimação - Mínimos quadrados condicionais

- Essencialmente, temos que resolver o seguinte sistema de equações não-lineares:

$$\begin{cases} S(\mu) = \dfrac{\partial}{\partial \mu} S(\mu,\phi) \mid_{\mu=\widehat{\mu},\phi=\widehat{\phi}} = 2 \displaystyle\sum_{t=p+1}^{n} \epsilon_t \dfrac{\partial}{\partial \mu} \epsilon_t \mid_{\mu=\widehat{\mu},\phi=\widehat{\phi}} = 0 \\ S(\phi_1) = \dfrac{\partial}{\partial \phi_1} S(\mu,\phi) \mid_{\mu=\widehat{\mu},\phi=\widehat{\phi}} = 2 \displaystyle\sum_{t=p+1}^{n} \epsilon_t \dfrac{\partial}{\partial \phi_1} \epsilon_t \mid_{\mu=\widehat{\mu},\phi=\widehat{\phi}} = 0 \\ \vdots \\ S(\phi_p) = \dfrac{\partial}{\partial \phi_p} S(\mu,\phi) \mid_{\mu=\widehat{\mu},\phi=\widehat{\phi}} = 2 \displaystyle\sum_{t=p+1}^{n} \epsilon_t \dfrac{\partial}{\partial \phi_p} \epsilon_t \mid_{\mu=\widehat{\mu},\phi=\widehat{\phi}} = 0 \end{cases}$$

Estimação - Mínimos quadrados condicionais

- Como exemplo, considere um modelo AR(1). Os estimadores de mínimos quadrados condicionais para (μ, ϕ) são obtidos minimizando:

$$S(\mu, \phi) = \sum_{t=2}^{n} \left[y_t - \mu - \phi(y_{t-1} - \mu)\right]^2 , \qquad (12)$$

ou seja, resolvendo o seguinte sistema de equações:

$$\begin{cases} S(\mu) = \dfrac{\partial}{\partial \mu} S(\mu, \phi) \mid_{\mu=\widehat{\mu}, \phi=\widehat{\phi}} = 0 \\ S(\phi) = \dfrac{\partial}{\partial \phi} S(\mu, \phi) \mid_{\mu=\widehat{\mu}, \boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}} = 0 \end{cases}$$

Estimação - Mínimos quadrados condicionais

- Derivando a Equação em (12) com relação a μ temos que:

$$
\begin{aligned}
\frac{\partial}{\partial \mu} S(\mu,\phi) &= 2\sum_{t=2}^{n}(\phi-1)\left[y_t-\mu-\phi(y_{t-1}-\mu)\right] \\
&= 2(\phi-1)\sum_{t=2}^{n} y_t - 2\mu(\phi-1)(n-1) \\
&- 2\phi(\phi-1)\sum_{t=2}^{n}(y_{t-1}-\mu) \\
&= 2(\phi-1)(n-1)\left[\overline{y}_1-\mu-\phi\overline{y}_2+2\phi\mu\right],
\end{aligned}
$$

em que $\overline{y}_1 = \sum_{t=2}^{n} y_t/(n-1)$ e $\overline{y}_2 = \sum_{t=2}^{n} y_{t-1}/(n-1)$.

Estimação - Mínimos quadrados condicionais

- Agora, derivando a Equação em (12) com relação a ϕ temos que:

$$
\begin{aligned}
\frac{\partial}{\partial \phi} S(\mu, \phi) &= -2 \sum_{t=2}^{n} (y_{t-1} - \mu) \left[y_t - \mu - \phi (y_{t-1} - \mu) \right] \\
&= -2 \left[\sum_{t=2}^{n} (y_{t-1} - \mu)(y_t - \mu) - \phi \sum_{t=2}^{n} (y_{t-1} - \mu)^2 \right].
\end{aligned}
$$

Estimação - Mínimos quadrados condicionais

- Por fim fazendo $\dfrac{\partial}{\partial \mu} S(\mu, \phi) = 0$ e $\dfrac{\partial}{\partial \phi} S(\mu, \phi) = 0$, temos que:

$$
\begin{aligned}
\widehat{\mu} &= \frac{\overline{Y}_1 - \widehat{\phi}\overline{Y}_2}{1 - \widehat{\phi}}, \\
\widehat{\phi} &= \frac{\sum_{t=2}^{n}(Y_{t-1} - \widehat{\mu})(Y_t - \widehat{\mu})}{\sum_{t=2}^{n}(Y_{t-1} - \widehat{\mu})^2}.
\end{aligned} \tag{13}
$$

- Ou seja, não é possível obter uma solução explícita.
- Para estimar σ^2 podemos utilizar (11), por exemplo.
- **Proposição**: sob algumas condições, assintoticamente, os estimadores de mínimos quadrados condicionais equivalem aos estimadores pelo MM.

Estimação - Mínimos quadrados condicionais

- **Prova**: Note que se $n \to \infty$ então $\overline{Y}_1 = \overline{Y}_2 \approx \overline{Y}$ e se $\widehat{\phi} \not\to 1$, então:

$$\widehat{\mu} \approx \frac{\overline{Y} - \widehat{\phi}\overline{Y}}{1 - \widehat{\phi}} = \overline{Y}. \tag{14}$$

- Assim, utilizando a Equação (14) na Equação (13), temos (se $n \to \infty$), que:

$$\widehat{\phi} \approx \frac{\sum_{t=2}^{n}(Y_{t-1} - \overline{Y})(Y_t - \overline{Y})}{\sum_{t=2}^{n}(Y_{t-1} - \overline{Y})^2} = \widehat{\rho}(1).$$

Estimação - máxima verossimilhança (MV)

- O método de MV consiste em, essencialmente, maximizar a verossimilhança associado ao modelo estatístico de interesse (aqui).
- Adicionalmente às suposições usuais, assumiremos que $\epsilon_t \sim NID(0, \sigma^2)$.
- Serão apresentados detalhes apenas para o modelo AR(1).
- Para o modelo geral AR(p) o raciocínio é análogo (veja aqui).

Estimação - máxima verossimilhança (MV)

- Seja, assim, o modelo AR(1) gaussiano:

$$Y_t = \mu + \phi(Y_{t-1} - \mu) + \epsilon_t,\ \epsilon_t \sim N(0, \sigma^2),\ |\phi| < 1. \qquad (15)$$

- Da Equação (15), temos que $Y_t | y_{t-1} \sim N(\mu + \phi(Y_{t-1} - \mu), \sigma^2)$.
- Logo, considerando $y_1, \ldots, y_n$ os valores observados do processo, temos que a verossimilhança (condicional) é dada por:

$$L(\mu, \phi, \sigma^2) = f(y_1, \ldots, y_n; \mu, \phi, \sigma^2) = f(y_1|y_0)f(y_2|y_1)\ldots f(y_n|y_{n-1}).$$

Estimação - máxima verossimilhança (MV)

- Note que, para o modelo AR(1), teremos e lidar com a “ausência” de Y_0, pois $Y_1 = \mu + \phi(Y_0 - \mu) + \epsilon_1$.
- Logo, em $f(y_1|y_0)$, y_0 não é observável.
- Analogamente, no modelo $AR(p)$, teremos que lidar com a ausência de $(Y_0, Y_{-1}, ... Y_{p-1})'$.

Estimação - máxima verossimilhança (MV)

- Algumas soluções possíveis (veja aqui):
  - Usar a verossimilhança exata ou não condicional (sempre que possível, utilizaremos este método), que leva à não dependência de tais valores (podemos ter verossimilhanças exata e não condicional, diferentes entre si).
  - Usar a verossimilhança condicional e
    - Substituir os valores "ausentes" por valores considerados razoáveis (por especialista(s) na área em questão) ou por estimativas obtidas por outro método.
    - Usar métodos de "previsão para o passado" ("backforecasting").

Estimação - máxima verossimilhança

- Voltando, temos que $Y_t|y_{t-1} \sim N(\mu+\phi(y_{t-1}-\mu), \sigma^2)$ com densidade:

$$f(y_t|y_{t-1}) = f_Z((y_t - \mu) - \phi(y_{t-1} - \mu)),$$

em que $f_Z(.)$ é a densidade de $Z \sim N(0, \sigma^2)$.

- Logo, temos que:

$$L(\mu, \phi, \sigma^2) = f(y_1) \prod_{t=2}^{n} f_Z((y_t - \mu) - \phi(y_{t-1} - \mu)).$$

- Sabendo que $\{y_t\}$ é causal, então $y_1 = \mu + \sum_{j=0}^{\infty} \phi^j \epsilon_{1-j}$ e portanto $Y_1 \sim N(\mu, \sigma^2/(1 - \phi^2))$.

Estimação - máxima verossimilhança

- Finalmente, temos que:

$$L(\mu, \phi, \sigma^2) = \frac{\sqrt{1-\phi^2}}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} S_1(\mu, \phi)\right\},$$

em que

$$S_1(\mu, \phi) = (1-\phi^2)(y_1-\mu)^2 + \sum_{t=2}^{n} [(y_t-\mu) - \phi(y_{t-1}-\mu)]^2,$$

é a soma de quadrados não condicional.

Estimação - máxima verossimilhança

- Logo, temos que maximizar

$$l \equiv l(\mu, \phi, \sigma^2) = \frac{1}{2}\ln(1-\phi^2) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}S_1(\mu,\phi) + \text{const.}$$

- Além disso:

$$\begin{aligned} S_1(\mu,\phi,\sigma^2) &= \frac{\partial l}{\partial \mu} = -\frac{1}{\sigma^2}(1-\phi^2)(y_1-\mu) \\ &+ \frac{1}{\sigma^2}\sum_{t=2}^{n}[(y_t-\mu)-\phi(y_{t-1}-\mu)](1-\phi) \quad (16) \end{aligned}$$

Estimação - máxima verossimilhança

- Adicionalmente:

$$
\begin{aligned}
S_2(\mu,\phi,\sigma^2) &= \frac{\partial l}{\partial \phi} = -\frac{\phi}{(1-\phi^2)} \\
&+ \frac{\phi(y_1-\mu)^2}{\sigma^2} + \frac{1}{\sigma^2}\sum_{t=2}^{n}[(y_t-\mu)-\phi(y_{t-1}-\mu)](y_t-\mu) \qquad (17)
\end{aligned}
$$

- Finalmente:

$$
S_3(\mu,\phi,\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} S_1(\mu,\phi) .
$$

Estimação - máxima verossimilhança

- Note que $S_1(\widehat{\mu}, \widehat{\phi}, \widehat{\sigma}^2) = 0$ e $S_2(\widehat{\mu}, \widehat{\phi}, \widehat{\sigma}^2) = 0$ não tem solução analítica.
- Contudo,

$$S_3(\widehat{\mu}, \widehat{\phi}, \widehat{\sigma}^2) = 0 \rightarrow \widehat{\sigma}^2 = \frac{1}{n-1} S_1\left(\widehat{\mu}, \widehat{\phi}\right). \tag{18}$$

- Logo, utilizamos (18) em (16) e (17) e resolvemos, numericamente (Newton-Raphson, Escore de Fisher, L-BFGS-B), o seguinte sistema de equações:

$$\begin{cases} S_1(\widehat{\mu}, \widehat{\phi}, \widehat{\sigma}^2) = 0 \\ S_2(\widehat{\mu}, \widehat{\phi}, \widehat{\sigma}^2) = 0 \end{cases}$$

Estimação - máxima verossimilhança

- Para o modelo AR(p), dada uma amostra de tamanho n, a ideia é decompor a verossimilhança como:

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= f_{Y_1,Y_2,\ldots,Y_{t-1},Y_t}(y_1,y_2,\ldots,y_{t-1},y_t;\boldsymbol{\theta}) \\
&= f_{Y_{p+1},\ldots,Y_t|Y_1,\ldots Y_p}(y_{p+1},\ldots,y_t|y_1,\ldots y_p;\boldsymbol{\theta}) \\
&\times f_{Y_1,Y_2,\ldots,Y_p}(y_1,y_2,\ldots,y_p;\boldsymbol{\theta}) \\
&= \left\{\prod_{t=p+1}^{T} f_{Y_t|Y_{t-1},Y_{t-2},\ldots,Y_{t-p}}(y_t|y_{1-1},y_{t-2},\ldots,y_{t-p};\boldsymbol{\theta})\right\} \\
&\times f_{Y_1,Y_2,\ldots,Y_p}(y_1,y_2,\ldots,y_p;\boldsymbol{\theta}).
\end{aligned}
$$

Comentários

- Assintoticamente, os EMM e os EMQC são assintoticamente equivalentes ($n \to \infty$).
- Um procedimento usual que tende a levar aos melhores resultados (entre os que foram apresentados) consiste em:
  - Obter as estimativas de MQC (de modo iterativo) a partir de estimativas iniciais pelo MM ou pelo MQC (de forma não iterativa).
  - Usar as estimativas obtidas anteriormente como valores iniciais no processo iterativo para a obtenção das estimativas de MV.

Comentários

- Sob certas condições de regularidade (aqui e aqui), temos, para n suficientemente grande, que:

$$\widehat{\boldsymbol{\theta}}_{MV} \approx N_{p+2}\left(\boldsymbol{\theta}, \boldsymbol{I}\left(\boldsymbol{\theta}\right)^{-1}\right),$$

em que $\boldsymbol{I}(\boldsymbol{\theta})$ é a respectiva informação de Fisher.