MODELOS MULTIVARIADOS

1. Introdução

- Interesse: modelos para uma série temporal vetorial $\mathbf{X}_t$, com n componentes $X_{1t}, X_{2t}, \ldots, X_{nt}$, observadas em $t \in Z$.

  Estudo de relações dinâmicas entre as séries componentes.

  Notação: $\mathbf{X}_t = (X_{1t}, X_{2t}, \ldots, X_{nt})'$, $t \in Z$ e X_{it} ou $X_{i,t}$, indistintamente, para a i-ésima componente, $i = 1, \ldots, p$.

- **Exemplo 1.** Podemos pensar o vetor $\mathbf{X}_t$ como constituído pelos retornos de n ativos financeiros de um fundo de investimentos no instante t e o objetivo é analisar o desempenho do fundo ao longo do tempo. Numa outra situação, um investidor pode ter uma carteira com ações da Petrobrás, Banco do Brasil, Banespa e Banco Itaú e neste caso $n = 4$.

- vetor de médias (depende de t):

$$\boldsymbol{\mu}_t = E(\mathbf{X}_t) = (\mu_{1t}, \mu_{2t}, \ldots, \mu_{nt})' \quad (1)$$

- matriz de covariâncias ($n \times n$, depende de t):

$$\begin{aligned} \boldsymbol{\Gamma}(t+\tau, t) &= \\ &= E\{(\mathbf{X}_{t+\tau} - \boldsymbol{\mu}_{t+\tau})(\mathbf{X}_t - \boldsymbol{\mu}_t)'\} \end{aligned} \quad (2)$$

- $\gamma_{ij}(t+\tau, t),\ i, j = 1, \ldots, n$: componentes de $\boldsymbol{\Gamma}(t+\tau, t)$:

$$\gamma_{ij}(t+\tau, t) = \text{Cov}\{X_{i,t+\tau}, X_{j,t}\}, \quad (3)$$

$i, j = 1, \ldots, n$ é a covariância entre as séries $X_{i,t+\tau}$ e $X_{j,t}$.

- **Exemplo 1.**(cont.) $\mathbf{X}_t = (X_{1t}, \ldots, X_{4t})'$, $\boldsymbol{\mu}_t = (\mu_{1t}, \ldots, \mu_{4t})'$ é o vetor de médias e a matriz (2) ficará

$$\boldsymbol{\Gamma}(t+\tau, t) =$$

$$\begin{bmatrix} \gamma_{11}(t+\tau,t) & \gamma_{12}(t+\tau,t) & \cdots & \gamma_{14}(t+\tau,t) \\ \gamma_{21}(t+\tau,t) & \gamma_{22}(t+\tau,t) & \cdots & \gamma_{24}(t+\tau,t) \\ \cdots & \cdots & & \cdots \\ \gamma_{41}(t+\tau,t) & \gamma_{42}(t+\tau,t) & \cdots & \gamma_{44}(t+\tau,t) \end{bmatrix}$$

2. **Séries Estacionárias**

- vetor de médias: não depende de t:

$$\boldsymbol{\mu} = E(\mathbf{X}_t) = (\mu_1, \ldots, \mu_n)', \qquad (4)$$

e, para $\tau \in Z$,

$$\boldsymbol{\Gamma}(\tau) = E\{(\mathbf{X}_{t+\tau} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'\} = [\gamma_{ij}(\tau)]_{i,j=1}^{n}. \qquad (5)$$

Neste caso, $\gamma_{ii}(\tau)$ será a f.a.cov. da série estacionária X_{it} e $\gamma_{ij}(\tau)$ será a função de covariância cruzada (f.c.c.) de X_{it} e X_{jt}. Notemos que, em geral, $\gamma_{ij}(\tau) \neq \gamma_{ji}(\tau)$.

- No caso particular de $\tau = 0$ em (5) obtemos

$$\mathbf{\Gamma}(0) = E\{(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'\}, \qquad (6)$$

que é a *matriz de covariâncias contemporâneas*. Em particular,

$\gamma_{ii}(0) = \text{Var}(X_{it})$,

$\gamma_{ij}(0) = \text{Cov}\{X_{it}, X_{jt}\}$.

- O *coeficiente de correlação contemporâneo* entre X_{it} e X_{jt} é então dado por

$$\rho_{ij}(0) = \frac{\gamma_{ij}(0)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}. \qquad (7)$$

$\rho_{ij}(0) = \rho_{ji}(0),\ \rho_{ii}(0) = 1$

$-1 \leq \rho_{ij}(0) \leq 1$, para todo $i, j = 1, \ldots, n$, logo $\boldsymbol{\rho}(0) = [\rho_{ij}(0)]_{i,j=1}^{n}$ é uma matriz simétrica, com elementos na diagonal principal todos iguais a um.

- A *matriz de correlações* de lag τ é definida por

$$\boldsymbol{\rho}(\tau) = \mathbf{D}^{-1}\boldsymbol{\Gamma}(\tau)\mathbf{D}^{-1}, \qquad (8)$$

sendo $\mathbf{D} = \text{diag}\{\sqrt{\gamma_{11}(0)}, \ldots, \sqrt{\gamma_{nn}(0)}\}$.

- Ou seja, denotando $\boldsymbol{\rho}(\tau) = [\rho_{ij}(\tau)]_{i,j=1}^{n}$, temos

$$\rho_{ij}(\tau) = \frac{\gamma_{ij}(\tau)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}, \qquad (9)$$

que é o coeficiente de correlação entre $X_{i,t+\tau}$ e $X_{j,t}$.

- $\tau > 0$: coeficiente mede a dependência linear de X_{it} sobre X_{jt}, que ocorreu antes do instante $t + \tau$.

  $\rho_{ij}(\tau) \neq 0$, $\tau > 0$, dizemos que X_{jt} é *antecedente a* X_{it} ou que X_{jt} *lidera* X_{it} no lag τ.

  De modo análogo, $\rho_{ji}(\tau)$ mede a dependência linear de X_{jt} sobre $X_{it}, \tau > 0$.

- **Proposição 1.** As seguintes propriedades são válidas:

(i) $\mathbf{\Gamma}(\tau) = \mathbf{\Gamma}'(-\tau)$.

(ii) $|\gamma_{ij}(\tau)| \leq [\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}$, $i, j = 1, \ldots, n$.

(iii) $\gamma_{ii}(\tau)$ é uma função de auto-covariância, para todo i.

(iv) $\sum_{j,k=1}^{m} \mathbf{a}_j^{'} \mathbf{\Gamma}(j-k)\mathbf{a}_k \geq 0$, para quaisquer m e $\mathbf{a}_1, \ldots, \mathbf{a}_m$ vetores de $\mathbb{R}^n$.

3. Estimação de Médias e Covariâncias

- $\{\mathbf{X}_t, t = 1, \ldots, T\}$: observações do processo estacionário $\{\mathbf{X}_t, t \in Z\}$;

  média $\boldsymbol{\mu}$ pode ser estimada pelo vetor de médias amostrais

$$\overline{\boldsymbol{X}} = \frac{\sum_{t=1}^{T} \mathbf{X}_t}{T}. \qquad (10)$$

  Segue-se que a média μ_j de X_{jt} é estimada por $\sum_{t=1}^{T} X_{jt}/T$.

- Para estimar $\mathbf{\Gamma}(\tau)$ usamos

$$\hat{\mathbf{\Gamma}}(\tau) = \frac{1}{T}\sum_{t=1}^{T-\tau}(\mathbf{X}_{t+\tau} - \overline{\boldsymbol{X}})(\mathbf{X}_t - \overline{\boldsymbol{X}})^{'}, \; 0 \leq \tau \leq T-1. \qquad (11)$$

- A matriz de correlações pode ser estimada por

$$\widehat{\boldsymbol{\rho}}(\tau) = \widehat{\mathbf{D}}'\widehat{\mathbf{\Gamma}}\widehat{\mathbf{D}}^{-1}, \tag{12}$$

onde $\hat{\mathbf{D}}$ é a matriz diagonal $n \times n$ dos desvios padrões amostrais das séries individuais.

- **Exemplo 2.** X_{1t} : retornos diários da Petrobrás ;

  X_{2t}: retornos diários do Ibovespa, de 3/1/95 a 27/12/2000, $T = 1498$;

  $\mathbf{X}_t = (X_{1t}, X_{2t})'$.

  SCA: usa a notação

$$+ : \rho_{ij}(\tau) \geq 2/\sqrt{T},$$
$$- : \rho_{ij}(\tau) \leq -2/\sqrt{T},$$
$$\bullet : -2/\sqrt{T} < \rho_{ij}(\tau) < 2/\sqrt{T}.$$

$$\widehat{\boldsymbol{\rho}}(0) = \begin{bmatrix} 1,00 & 0,83 \\ 0,83 & 1,00 \end{bmatrix},$$

enquanto

$$\widehat{\boldsymbol{\rho}}(1) = \begin{bmatrix} 0,05 & 0,02 \\ 0,10 & 0,11 \end{bmatrix}.$$

Como $1/\sqrt{1498} = 0,02584$, o elemento $\rho_{12}(1)$ pode ser considerado estatisticamente nulo, de modo que a representação pictórica dessa matriz de correlações amostrais é

$$\begin{bmatrix} + & \cdot \\ + & + \end{bmatrix}.$$

Note que a correlação contemporânea entre as duas séries é 0,83.

- **Exemplo 3.** Consideremos, agora, a série bivariada, consistindo dos retornos mensais do Ibovespa e da taxa de juros dos títulos C-Bond da dívida brasileira, ambas de julho de 1994 a agosto de 2001, $T = 86$.

Vemos que $\boldsymbol{\rho}(\tau)$, $\tau = 1, 2, \ldots, 4$, podem ser consideradas nulas, o que sugere que estamos na presença de um ruído branco bivariado. É fácil verificar que a correlação contemporânea entre as duas séries é negativa (-$0,77$).

4. **Modelos VAR**

- Dizemos que o processo $\mathbf{X}_t$, de ordem $n \times 1$, segue um modelo VAR(p) se

$$\mathbf{X}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \ldots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \mathbf{a}_t, \tag{13}$$

onde $\mathbf{a}_t \sim \text{RB}(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Phi}_0 = (\phi_{10}, \ldots, \phi_{n0})'$ é um vetor $n \times 1$ de constantes e $\mathbf{\Phi}_k$ são matrizes $n \times n$ constantes, com elementos $\phi_{ij}^{(k)}$, $i, j = 1, \ldots, n$, $k = 1, \ldots, p$.

- Se $\mathbf{I}_n$ é a matriz identidade de ordem n, o modelo (13) pode ser escrito na forma

$$\mathbf{\Phi}(B)\mathbf{X}_t = \mathbf{\Phi}_0 + \mathbf{a}_t, \tag{14}$$

onde $\mathbf{\Phi}(B) = \mathbf{I}_n - \mathbf{\Phi}_1 B - \ldots - \mathbf{\Phi}_p B^p$ é o operador auto-regressivo vetorial de ordem p, ou ainda, um polinômio matricial $n \times n$ em B. O elemento genérico de $\mathbf{\Phi}(B)$ é $[\delta_{ij} - \phi_{ij}^{(1)} B - \ldots - \phi_{ij}^{(p)} B^p]$, para $i, j = 1, \ldots, n$ e $\delta_{ij} = 1$, se $i \neq j$ e igual a zero, caso contrário.

- VAR(1):

$$\mathbf{X}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{a}_t. \tag{15}$$

caso especial: $n = 2$ e (15) reduz-se a

$$\begin{aligned} X_{1t} &= \phi_{10} + \phi_{11} X_{1,t-1} + \phi_{12} X_{2,t-1} + a_{1t}, \\ X_{2t} &= \phi_{20} + \phi_{21} X_{1,t-1} + \phi_{22} X_{2,t-1} + a_{2t}. \end{aligned} \tag{16}$$

- $\phi_{12} = 0 \quad \rightarrow \quad X_{1t}$ não dependerá de $X_{2,t-1}$;

  $\phi_{21} = 0 \quad \rightarrow \quad X_{2,t}$ não dependerá de $X_{1,t-1}$.

  $\phi_{12} = 0$ e $\phi_{21} \neq 0$: existe uma relação linear unidirecional de X_{1t} para X_{2t}.

  $\phi_{12} = \phi_{21} = 0$ dizemos que não existe relação linear entre as séries, ou que elas são *não-acopladas*.

  $\phi_{12} \neq 0$, $\phi_{21} \neq 0$, dizemos que existe uma relação de *feedback* entre as duas séries.

  Note também que se $\sigma_{12} = 0$ em $\boldsymbol{\Sigma}$, não existe relação linear contemporânea entre X_{1t} e X_{2t}.

- O processo $\boldsymbol{X}_t$ em (15) será estacionário se a média for constante e $E(\mathbf{X}_{t+\tau}\mathbf{X}_t^{\prime})$

independente de t. Neste caso, se $\boldsymbol{\mu} = E(\mathbf{X}_t)$, teremos

$$\boldsymbol{\mu} = (\mathbf{I}_n - \boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}_0.$$

Segue-se que o modelo poderá ser escrito na forma

$$\mathbf{X}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}(\mathbf{X}_{t-1} - \boldsymbol{\mu}) + \mathbf{a}_t,$$

ou ainda, se $\tilde{\mathbf{X}}_t = \mathbf{X}_t - \boldsymbol{\mu}$,

$$\tilde{\mathbf{X}}_t = \boldsymbol{\Phi}\tilde{\mathbf{X}}_{t-1} + \mathbf{a}_t. \qquad (17)$$

- Assim como no caso de um AR(1) univariado, obtemos de (17) que

$$\tilde{\mathbf{X}}_t = \mathbf{a}_t + \boldsymbol{\Phi}\mathbf{a}_{t-1} + \boldsymbol{\Phi}^2\mathbf{a}_{t-2} + \ldots, \qquad (18)$$

ou seja, temos a representação MA(∞) do modelo. Também, é fácil ver que Cov($\mathbf{a}_t, \mathbf{X}_{t-1}$) $= \mathbf{0}$ e Cov($\mathbf{a}_t, \mathbf{X}_t$) $= \mathbf{\Sigma}$.

- **Proposição 2.** O processo $\mathbf{X}_t$ seguindo um modelo VAR(1) será estacionário se todas as soluções de

$$|\mathbf{I}_n - \mathbf{\Phi} z| = 0 \qquad (19)$$

estiverem fora do círculo unitário.

Como as soluções de (19) são inversas dos autovalores de $\mathbf{\Phi}$, uma condição equivalente é que todos os auto-valores de $\mathbf{\Phi}$ sejam menores do que um, em módulo. Ou ainda, $|\mathbf{I}_n - \mathbf{\Phi} z| \neq 0$, $|z| \leq 1$.

- **Exemplo 4.** No caso de um VAR(1) bivariado, temos que (19) fica

$$\begin{bmatrix} 1 - \phi_{11} z & -\phi_{22} z \\ -\phi_{21} z & 1 - \phi_{22} z \end{bmatrix}$$

$$= (1 - \phi_{11}z)(1 - \phi_{22}z) - \phi_{12}\phi_{21}z^2 = 0,$$

ou seja, obtemos a equação

$$1 - \text{tr}(\mathbf{\Phi})z - |\mathbf{\Phi}|z^2 = 0,$$

onde $\text{tr}(\mathbf{\Phi}) = \phi_{11} + \phi_{22}$ indica o traço de $\mathbf{\Phi}$. Logo as duas séries são (conjuntamente) estacionárias se as soluções desta equação de segundo grau estiverem fora do círculo unitário. Por exemplo, se

$$\mathbf{\Phi} = \begin{bmatrix} 0,5 & 0,3 \\ -0,6 & -0,1 \end{bmatrix},$$

então $\text{tr}(\mathbf{\Phi}) = 0,4$, $|\mathbf{\Phi}| = 0,13$ e as raízes da equação terão módulos maiores do que um.

- Matriz de Covariâncias:

$$\mathbf{\Gamma}(0) = \mathbf{\Sigma} + \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}' + \mathbf{\Phi}^2\mathbf{\Sigma}(\mathbf{\Phi}^2)' + \ldots$$

$$= \sum_{j=0}^{\infty} \mathbf{\Phi}^j \mathbf{\Sigma} (\mathbf{\Phi}^j)^{'}, \; \mathbf{\Phi}_0 = \mathbf{I}_n.$$

Se pós-multiplicarmos (17) por $\tilde{\mathbf{X}}_{t-\tau}^{'}$ teremos

$$E(\tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_{t-\tau}^{'}) = \mathbf{\Phi} E(\tilde{\mathbf{X}}_{t-1} \tilde{\mathbf{X}}_{t-\tau}^{'}) + E(\mathbf{a}_t \tilde{\mathbf{X}}_{t-\tau}^{'}).$$

Fazendo $\tau = 0$ obtemos

$$\mathbf{\Gamma}(0) = \mathbf{\Phi}\mathbf{\Gamma}(-1) + \mathbf{\Sigma} = \mathbf{\Phi}\mathbf{\Gamma}(1)^{'} + \mathbf{\Sigma}.$$

- **Proposição 3.** Para o modelo VAR(p) dado em (14) temos os seguintes resultados:

(i) O processo $\mathbf{X}_t$ será estacionário se as soluções de

$$|\mathbf{I}_n - \mathbf{\Phi}_1 B - \ldots - \mathbf{\Phi}_p B^p| = 0$$

estiverem fora do círculo unitário.

(ii) Se $\mathbf{X}_t$ for estacionário,

$$\boldsymbol{\mu} = E(\mathbf{X}_t) = (\mathbf{I}_n - \boldsymbol{\Phi}_1 B - \ldots - \boldsymbol{\Phi}_p B^p)^{-1} \boldsymbol{\Phi}_0.$$

(iii) Escrevendo (14) na forma

$$\tilde{\mathbf{X}}_t = \boldsymbol{\Phi}_1 \tilde{\mathbf{X}}_{t-1} + \ldots + \boldsymbol{\Phi}_p \tilde{\mathbf{X}}_{t-p} + \mathbf{a}_t,$$

com $\tilde{\mathbf{X}}_t = \mathbf{X}_t - \boldsymbol{\mu}$ e multiplicando esta equação por $\tilde{\mathbf{X}}'_{t-\tau}$ obtemos

$$\boldsymbol{\Gamma}(\tau) = \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}(\tau - 1) + \ldots + \boldsymbol{\Phi}_p \boldsymbol{\Gamma}(\tau - p), \quad \tau > 0,$$

que são as equações de Yule-Walker no caso de um modelo VAR(p).

5. **Construção de Modelos VAR**

- Uma maneira de identificar a ordem p de um modelo VAR(p) consiste em ajustar seqüencialmente modelos auto-regressivos vetoriais de ordens $1, 2, \ldots, k$ e testar a significância dos coeficientes (matrizes). Considere, pois, os modelos

$$
\begin{aligned}
\mathbf{X}_t &= \mathbf{\Phi}_0{}^{(1)} + \mathbf{\Phi}_1{}^{(1)}\mathbf{X}_{t-1} + \mathbf{a}_t^{(1)}, \\
\mathbf{X}_t &= \mathbf{\Phi}_0{}^{(2)} + \mathbf{\Phi}_1^{(2)}\mathbf{X}_{t-1} + \mathbf{\Phi}_2^{(2)}\mathbf{X}_{t-2} + \mathbf{a}_t^{(2)}, \\
\cdots & \quad \cdots \\
\mathbf{X}_t &= \mathbf{\Phi}_0^{(k)} + \mathbf{\Phi}_1^{(k)}\mathbf{X}_{t-1} + \ldots + \mathbf{\Phi}_k^{(k)}\mathbf{X}_{t-k} + \mathbf{a}_t^{(k)}.
\end{aligned}
\tag{20}
$$

Os parâmetros podem ser estimados por MQ ordinários, que fornecem estimadores consistentes e eficientes. Testamos, então,

$$
\begin{aligned}
H_0 &: \mathbf{\Phi}_k^{(k)} = \mathbf{0}, \\
H_1 &: \mathbf{\Phi}_k^{(k)} \neq 0, k = 1, 2, \ldots .
\end{aligned}
\tag{21}
$$

O teste da razão de verossimilhanças é baseado nas estimativas das matrizes de covariâncias dos resíduos dos modelos ajustados. Para a k-ésima equação, considere

$$\hat{\mathbf{a}}_t^{(k)} = \mathbf{X}_t - \hat{\mathbf{\Phi}}_0^{(k)} - \hat{\mathbf{\Phi}}_1^{(k)} \mathbf{X}_{t-1} - \ldots - \hat{\mathbf{\Phi}}_k^{(k)} \mathbf{X}_{t-k}.$$

A matriz de covariância dos resíduos, que estima $\mathbf{\Sigma}$, é dada então por

$$\hat{\mathbf{\Sigma}}_k = \frac{1}{T-k} \sum_{t=k+1}^{T} \hat{\mathbf{a}}_t^{(k)} (\hat{\mathbf{a}}_t^{(k)})', \ k \geq 0, \tag{22}$$

onde para $k = 0$, $\hat{\mathbf{a}}_t^{(0)} = \mathbf{X}_t - \overline{\mathbf{X}}$. A estatística da RV para o teste (21) é dada por

$$\mathrm{RV}(k) = (T-k) \ln \frac{|\hat{\mathbf{\Sigma}}_{k-1}|}{|\hat{\mathbf{\Sigma}}_k|}, \tag{23}$$

que tem distribuição qui-quadrado com n^2 graus de liberdade, $\chi^2(n^2)$.

- Outra maneira de identificar a ordem de um VAR é usar algum critério de informação, tais como:

$$
\begin{aligned}
\text{AIC}(k) &= \ln(|\hat{\Sigma}_k|) + 2kn^2/T \text{ (Akaike)}, \\
\text{BIC}(k) &= \ln(|\hat{\Sigma}_k|) + kn^2 \ln(T)/T \text{ (Schwarz)}, \\
\text{HQC}(k) &= \ln(|\hat{\Sigma}_k|) + kn^2 \ln(\ln(T))/T \text{ (Hannan-Quinn)}
\end{aligned}
$$

- Estimação

  Identificado o valor de p e supondo $\mathbf{a}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, podemos estimar os coeficientes por maxima verossimilhança. Neste caso, os estimadores de MQ são equivalentes a estimadores de MV condicionais.
  No caso de um VAR(1), os EMV condicionais são obtidos maximizando-se

$$
\begin{aligned}
\ell &= -\frac{n(T+1)}{2}\ln(2\pi) + \frac{(T-1)}{2}\ln|\mathbf{\Sigma}^{-1}| \\
&\quad -\frac{1}{2}\sum_{t=2}^{T}(\mathbf{X}_t - \mathbf{\Phi}\mathbf{X}_{t-1})^{'}\mathbf{\Sigma}^{-1}(\mathbf{X}_t - \mathbf{\Phi}\mathbf{X}_{t-1}),
\end{aligned}
$$

obtendo-se

$$
\begin{aligned}
\hat{\mathbf{\Phi}} &= [\sum_{t=2}^{T}\mathbf{X}_t\mathbf{X}_{t-1}^{'}][\sum_{t=2}^{T}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^{'}]^{-1}, \\
\hat{\mathbf{\Sigma}} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\mathbf{a}}_t(\hat{\mathbf{a}}_t)^{'}, \\
\hat{\mathbf{a}}_t &= \mathbf{X}_t - \hat{\mathbf{\Phi}}\mathbf{X}_{t-1}.
\end{aligned}
$$

No caso geral de um VAR(p), os EMV condicionais são obtidos por métodos de maximização numérica.

- Diagnóstico

  Para testar se o modelo é adequado, usamos os resíduos para construir a versão multivariada da estatística de Box-Ljung-Pierce, dada por

$$Q(m) = T^2 \sum_{\tau=1}^{m} \frac{1}{T-\tau} \text{tr}(\hat{\mathbf{\Gamma}}(\tau)^{'}\hat{\mathbf{\Gamma}}(0)^{-1}\hat{\mathbf{\Gamma}}(\tau)\hat{\mathbf{\Gamma}}(0)^{-1}), \tag{24}$$

que sob H_0 : a série $\mathbf{a}_t$ é ruído branco, tem distribuição $\chi^2(n^2(m-p))$. Para que o número de graus de liberdade seja positivo, m deve ser maior do que p.

- Previsão

Considere o VAR(1) dado em (17) e suponha que o parâmetro $\mathbf{\Phi}$ seja conhecido. A previsão de origem T e horizonte h é dada por

$$\hat{\mathbf{X}}_T(h) = \mathbf{\Phi}\hat{\mathbf{X}}_T(h-1),$$

da qual segue

$$\hat{\mathbf{X}}_T(h) = \mathbf{\Phi}^h \mathbf{X}_T, \quad h = 1, 2, \ldots. \quad (25)$$

Como

$$\mathbf{X}_{t+h} = \mathbf{\Phi}\mathbf{X}_{T+h-1} + \mathbf{a}_{T+h},$$

temos que o erro de previsão h passos a frente é dado por

$$\mathbf{e}_T(h) = \mathbf{X}_{t+h} - \hat{\mathbf{X}}_T(h) = \sum_{j=0}^{h-1} \mathbf{\Phi}^j \mathbf{a}_{T+h-j}, \quad (26)$$

de modo que o erro quadrático médio do previsor (25) fica

$$\mathbf{\Sigma}(h) = \mathsf{EQMP}(h) = \sum_{j=0}^{h-1} \mathbf{\Phi}^j \mathbf{\Sigma}(\mathbf{\Phi}^j)'. \quad (27)$$

Considerando, agora, o modelo VAR(p), com parâmetros supostos conhecidos, $\mathbf{a}_t$ uma seqüência i.i.d. e $\mathcal{F}_t = \{\mathbf{X}_s : s \leq t\}$, obtemos

$$E(\mathbf{X}_{t+h}|\mathcal{F}_t) = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 E(\mathbf{X}_{t+h-1}|\mathcal{F}_t) + \ldots +$$

$$+\mathbf{\Phi} E(\mathbf{X}_{t+h-p}|\mathcal{F}_t),$$

pois $E(\mathbf{a}_{t+h}|\mathcal{F}_t) = 0$, para todo $h > 0$.

Para $h = 1$ obtemos

$$\hat{\mathbf{X}}_t(1) = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{X}_t + \ldots + \mathbf{\Phi}_p \mathbf{X}_{t-p+1},$$

e para $h = 2$ temos

$$\hat{\mathbf{X}}_t(2) = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \hat{\mathbf{X}}_t(1) + \mathbf{\Phi}_2 \mathbf{X}_t + \ldots + \mathbf{\Phi}_p \mathbf{X}_{t-p+2},$$

de modo que as previsões podem ser obtidas recursivamente.

Neste caso, o erro de previsão de horizonte h é dado por

$$\mathbf{e}_T(h) = \sum_{j=0}^{h-1} \mathbf{\Psi}_j \mathbf{a}_{T+h-j}, \tag{28}$$

onde as matrizes $\mathbf{\Psi}_j$ são obtidas recursivamente por

$$\mathbf{\Psi}_j = \sum_{k=1}^{p-1} \mathbf{\Psi}_{j-k} \mathbf{\Phi}_k, \tag{29}$$

com $\mathbf{\Psi}_0 = \mathbf{I}_n$ e $\mathbf{\Phi}_j = 0, j > p$. Segue-se que a matriz de EQM de previsão fica

$$\mathbf{\Sigma}(h) = \sum_{j=0}^{p-1} \mathbf{\Psi}_j \mathbf{\Sigma} \mathbf{\Psi}_j'. \tag{30}$$

- Quando os parâmetros do modelo VAR(p) são estimados, o melhor preditor de X_{T+h} é agora dado por

$$\tilde{\mathbf{X}}_T(h) = \hat{\mathbf{\Phi}}_0 + \hat{\mathbf{\Phi}}_1 \tilde{\mathbf{X}}_T(h-1) + \ldots + \hat{\mathbf{\Phi}}_p \tilde{\mathbf{X}}_T(h-p), \quad h > 1. \tag{31}$$

Neste caso, a matriz de EQM de previsão torna-se

$$\hat{\mathbf{\Sigma}}(h) = \mathbf{\Sigma}(h) + EQM(\mathbf{X}_{T+h} - \tilde{\mathbf{X}}_T(h)). \tag{32}$$

Na prática, o segundo termo em (32) é ignorado e $\hat{\mathbf{\Sigma}}(h)$ é calculada por

$$\hat{\mathbf{\Sigma}}(h) = \sum_{j=0}^{h-1} \hat{\mathbf{\Psi}}_j \hat{\mathbf{\Sigma}} \hat{\mathbf{\Psi}}_j', \tag{33}$$

com $\hat{\mathbf{\Psi}}_j = \sum_{k=1}^{p-1} \hat{\mathbf{\Psi}}_{j-k} \hat{\mathbf{\Phi}}_k$. Lütkepohl (1991) dá uma aproximação para o segundo membro de (32).

- **Exemplo 5.** Retornemos ao exemplo 4 e ajustemos um modelo VAR(p) à série $\mathbf{X}_t$, onde:

  X_{1t}: retornos diários do Ibovespa,

  X_{2t}: retornos diários da Petrobrás, $T = 1498$.

  Na Tabela 7.3 temos os resultados de ajustes de modelos auto-regressivos até ordem 8. Usando os valores de (23) ou dos AIC correspondentes, selecionamos a ordem $p = 6$. A tabela foi obtida usando-se o SCA.

  Tabela 7.3 - Estatísticas resultantes de ajustes de modelos VAR(p), $p = 1, \ldots, 8$, para os retornos diários do Ibovespa e Petrobrás.

- EVIEWS: fornece o modelo VAR(6):

$$X_{1t} = 0,1018 X_{1,t-1} - 0,1113 X_{1,t-6} +$$

$$+0,0790 X_{2,t-3} + a_{1t},$$

$$X_{2t} = -0,1348X_{1,t-6} + 0,1111X_{2,t-1} -$$

$$-0,0973X_{2,t-3} + a_{2,t}.$$

retornos diários da Petrobrás: influenciados por valores passados dos retornos diários do Ibovespa, ou seja, dependem de preços do mercado financeiro;

retornos do Ibovespa: influenciados por valores defasados dos retornos da Petrobrás, o que é razoável, dado que as ações da Petrobrás fazem parte do índice.

Seque-se que há uma relação de "feedback' entre as duas séries.

O coeficiente de correlação entre as duas séries é $0,83$, o que mostra uma relação contemporânea forte.

6. **Causalidade de Granger**

- Para sistemas temporais, Granger (1969) define causalidade em termos de *previsibilidade*: a variável X causa a variável Y, com respeito a um dado universo de informação (que inclui X e Y), se o presente de Y pode ser previsto mais eficientemente usando valores passados de X, do que não usando esse passado, toda e qualquer outra informação disponível (incluindo valores passados de Y) sendo usada em ambos os casos. A definição não requer que o sistema seja linear; se o for, as previsões serão lineares.

- Seja $\{A_t, t = 0, \pm 1, \pm 2, \ldots\}$ o conjunto de informação relevante até (e incluindo) o instante t, contendo pelo menos X_t, Y_t. Defina $\overline{A}_t = \{A_s : s < t\}$, $\overline{\overline{A}}_t = \{A_s : s \leq t\}$, e definições análogas para $\overline{X}_t, \overline{Y}_t$, etc. Seja $P_t(Y|B)$ o preditor de EQM mínimo de Y_t, usando o conjunto de informação B e $\sigma^2(Y|B)$ o correspondente EQM do preditor.

- **Definição 1.** Dizemos que:

(a) $X_t \rightarrow Y_t$: X_t *causa* Y_t no sentido de Granger se

$$\sigma^2(Y_t|\overline{A}_t) < \sigma^2(Y_t|\overline{A}_t - \overline{X}_t).$$

Ou seja, Y_t pode ser melhor prevista usando toda a informação disponível, incluindo o passado de Y_t e X_t.

Dizemos também que X_t é *exógena* ou *antecedente* a Y_t.

(b) $X_t \Rightarrow Y_t$: X_t *causa instantaneamente* Y_t no sentido de Granger se:

$$\sigma^2(Y_t|A_t, \overline{\overline{X}}_t) < \sigma^2(Y_t|\overline{A}_t)$$

Ou seja, o valor presente de Y_t é melhor previsto se o valor presente de X_t for incluído.

(c) Há *feedback*, e escrevemos $X_t \leftrightarrow Y_t$, se X_t causa Y_t e Y_t causa X_t.

(d) Há *causalidade unidirecional* de X_t para Y_t se $X_t \to Y_t$ e *não há feedback*.

- É fácil ver que se $X_t \Rightarrow Y_t$, então $Y_t \Rightarrow X_t$. Portanto usualmente dizemos que há causalidade instantânea entre X_t e Y_t.

- Há várias propostas para operacionalizar as definições anteriores:

  Pierce e Haugh (1977) propõem ajustar modelos ARIMA a transformações adequadas de ambas as séries e depois estabelecer padrões de causalidade entre os resíduos por meio de correlações cruzadas. Veja também Layton (1984).

Hsiao (1979) sugere ajustar modelos auto-regressivos via AIC.

No caso de mais de duas séries, Boudjellaba et al. (1992) sugerem ajustar modelos VARMA às séries.

Uma resenha desses procedimentos é feita por Cunha (1997).

- Trataremos do assunto por meio da representação VAR da série multivariada $\mathbf{X}_t$, de ordem $n \times 1$. A representação MA do processo é dada por

$$\mathbf{X}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(B)\mathbf{a}_t, \quad \boldsymbol{\Psi}_0 = \mathbf{I}_n. \tag{34}$$

Suponha que

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Z}_t \end{bmatrix},$$

onde $\mathbf{Y}_t$ é um vetor $r \times 1$ e $\mathbf{Z}_t$ é um vetor $s \times 1$, $r + s = n$. Então podemos escrever

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Z}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Psi}_{11}(B) & \boldsymbol{\Psi}_{12}(B) \\ \boldsymbol{\Psi}_{21}(B) & \boldsymbol{\Psi}_{22}(B) \end{bmatrix} \begin{bmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{bmatrix}, \tag{35}$$

particionando $\boldsymbol{\mu}, \boldsymbol{\Psi}(B)$ e $\mathbf{a}_t$ de acordo com a partição de $\mathbf{X}_t$. Se houver causalidade unidirecional de $\mathbf{Y}_t$ para $\mathbf{Z}_t$, isto é, se $\mathbf{Z}_t$ for melhor prevista pelo presente e passado de $\mathbf{Y}_t$, mas não o contrário, deveremos ter $\boldsymbol{\Psi}_{12}(B) = 0$ e obteremos

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{11}(B)\mathbf{a}_{1t}, \\ \mathbf{Z}_t &= \boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{21}(B)\mathbf{a}_{1t} + \boldsymbol{\Psi}_{22}(B)\mathbf{a}_{2t}. \end{aligned}$$

- Na realidade, é possível demonstrar o seguinte resultado, que é uma caracterização de não-causalidade de Granger.

Note que $\mathbf{\Psi}(B) = \mathbf{I}_n + \mathbf{\Psi}_1 b + \mathbf{\Psi}_2 B^2 + \ldots$ e

$$\mathbf{\Psi}_i = \begin{bmatrix} \mathbf{\Psi}_{11,i} & \mathbf{\Psi}_{12,i} \\ \mathbf{\Psi}_{21,i} & \mathbf{\Psi}_{22,1} \end{bmatrix}, \quad i = 1, 2, \ldots.$$

- **Proposição 4.** O previsor ótimo de $\mathbf{Y}_t$ baseado em $\overline{\overline{\overline{\mathbf{X}}}}_t$ é igual ao previsor ótimo de $\mathbf{Y}_t$ baseado em $\overline{\overline{\overline{\mathbf{Y}}}}_t$ se e somente se $\mathbf{\Psi}_{12,i} = \mathbf{0}$, $i = 1, 2, \ldots$.

- Do ponto de vista prático, convêm considerar o modelo VAR de ordem finita, ou seja,

$$\begin{aligned} \mathbf{X}_t \;\; = \;\; & \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Z}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{\Phi}_{11,1} & \mathbf{\Phi}_{12,1} \\ \mathbf{\Phi}_{21,1} & \mathbf{\Phi}_{22,1} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-1} \\ \mathbf{Z}_{t-1} \end{bmatrix} + \quad (36) \\ & + \begin{bmatrix} \mathbf{\Phi}_{11,p} & \mathbf{\Phi}_{12,p} \\ \mathbf{\Phi}_{21,p} & \mathbf{\Phi}_{22,p} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-p} \\ \mathbf{Z}_{t-p} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{bmatrix}, \end{aligned}$$

e a condição da Proposição 4 estará satisfeita se e somente se $\mathbf{\Phi}_{12,i} = \mathbf{0}$, $i = 1, 2, \ldots, p$. Ou seja, se $\mathbf{X}_t$ seguir um modelo VAR(p), com matriz de covariâncias não-singular, então $\boldsymbol{Z}_t$ não causa $\boldsymbol{Y}_t$ se e somente se $\boldsymbol{\Phi}_{12,i} = \mathbf{0}$, $i = 1, 2, \ldots, p$.

- Uma caracterização de não-existência de causalidade instantânea é dada pela proposição seguinte. A prova é dada em Lütkepohl (1991).

  Proposição 5. Se $\mathbf{X}_t$ for como em (36), com matriz de covariâncias não singular, então *não existe causalidade instantânea* entre $\mathbf{Y}_t$ e $\mathbf{Z}_t$ se e somente se $E(\mathbf{a}_{1t}\mathbf{a}_{2t}^{'}) = 0$.

- A equação (36) pode ser escrita como:

$$
\begin{aligned}
\mathbf{Y}_t &= \boldsymbol{\mu}_1 + \sum_{i=1}^{p} \boldsymbol{\Phi}_{11,i}\mathbf{Y}_{t-i} + \sum_{i=1}^{p} \boldsymbol{\Phi}_{12,i}\mathbf{Z}_{t-i} + \mathbf{a}_{1t}, \quad (37) \\
\mathbf{Z}_t &= \boldsymbol{\mu}_2 + \sum_{i=1}^{p} \boldsymbol{\Phi}_{21,i}\mathbf{Y}_{t-i} + \sum_{i=i}^{p} \boldsymbol{\Phi}_{22,i}\mathbf{Z}_{t-i} + \mathbf{a}_{2t}. \quad (38)
\end{aligned}
$$

Suponha, também, a matriz $\mathbf{\Sigma}$ particionada como

$$\mathbf{\Sigma} = \left[\begin{array}{cc} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{array} \right],$$

sendo que $\mathbf{\Sigma}_{ij} = E(\mathbf{a}_{it}\mathbf{a}_{jt}^{'})$, $i, j = 1, 2$. Então:

(i) $\mathbf{Z}_t$ não causa $\mathbf{Y}_t \leftrightarrow \mathbf{\Phi}_{12,i} = \mathbf{0}$, para todo i;

(ii) $\mathbf{Y}_t$ não causa $\mathbf{Z}_t \leftrightarrow \mathbf{\Phi}_{21,i} = \mathbf{0}$, para todo i.

- **Proposição 6** (i) $\mathbf{Z}_t$ não causa $\mathbf{Y}_t \leftrightarrow |\mathbf{\Sigma}_{11}| = |\mathbf{\Sigma}_1|$, onde $\mathbf{\Sigma}_1 = E(\mathbf{c}_{1t}\mathbf{c}_{1t}^{'})$ é obtida da regressão restrita

$$\mathbf{Y}_t = \boldsymbol{\nu}_1 + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{Y}_{t-i} + \mathbf{c}_{1t}. \tag{39}$$

(ii) $\mathbf{Y}_t$ não causa $\mathbf{Z}_t \leftrightarrow |\mathbf{\Sigma}_{22}| = |\mathbf{\Sigma}_2|$, onde $\mathbf{\Sigma}_2 = E(\mathbf{c}_{2t}\mathbf{c}_{2t}^{'})$ é obtida da regressão restrita

$$\mathbf{Z}_t = \boldsymbol{\nu}_2 + \sum_{i=1}^{p} \mathbf{C}_i \mathbf{Z}_{t-i} + \mathbf{c}_{2t}. \qquad (40)$$

As regressões (37)-(40) podem ser estimadas por MQO e a partir dos resíduos de MQ as matrizes de covariâncias envolvidas são estimadas por:

$$\hat{\boldsymbol{\Sigma}}_i = (T-p)^{-1} \sum_{t=p+1}^{T} \hat{\boldsymbol{c}}_{it} \hat{\boldsymbol{c}}_{it}^{'},$$

$$\hat{\boldsymbol{\Sigma}}_{ii} = (T-p)^{-1} \sum_{t=p+1}^{T} \hat{\boldsymbol{a}}_{it} \hat{\boldsymbol{a}}_{it}^{'}, \quad i = 1, 2.$$

- Os testes e respectivas estatísticas da razão de verossimilhanças são dados por:

(i) $H_{01}: \boldsymbol{\Phi}_{12,i} = \mathbf{0}$, para todo i ($\mathbf{Z}_t$ não causa $\mathbf{Y}_t$),

$$RV_1 = (T-p)[\log|\hat{\boldsymbol{\Sigma}}_1| - \log|\hat{\boldsymbol{\Sigma}}_{11}|] \sim \chi^2(prs).$$

(ii) $H_{02}: \ \mathbf{\Phi}_{21,i} = \mathbf{0}$, para todo i ($\mathbf{Y}_t$ não causa $\mathbf{Z}_t$),

$$RV_2 = (T-p)[\log|\hat{\mathbf{\Sigma}}_2| - \log|\hat{\mathbf{\Sigma}}_{22}|] \sim \chi^2(prs).$$

- **Exemplo 6.** Para o exemplo 5, vemos que $X_{1t} \to X_{2t}$, ou seja, retornos diários do Ibovespa causam, no sentido de Granger, retornos diários da Petrobrás. Também, $X_{2t} \to X_{1t}$. Logo, há "feedback' entre as duas séries de retornos.

Exemplo 7. Um modelo VAR(1) para as séries de retornos diários do Banespa (X_{1t}) e da Petrobrás (X_{2t}) é dado por

$$\begin{aligned}
X_{1t} &= 0,122 X_{1,t-1} + a_{1t}, \\
X_{2t} &= 0,333 X_{1,t-1} + 0,081 X_{2,t-1} + a_{2t},
\end{aligned}$$

sendo $\begin{bmatrix} 1 & 0,007 \\ 0,007 & 1 \end{bmatrix}$ a matriz de correlação estimada. Vemos que Banespa causa Petrobrás, mas não o contrário.

Escrevendo o modelo na forma estrutural, obtemos

$$X_{2t} = 0,005X_{1t} + 0,332X_{1,t-1} + 0,081X_{2,t-1} + b_{2t}, \quad (41)$$

o que mostra que provavelmente não há causalidade instantânea entre as séries, dada a magnitude do coeficiente de X_{1t}. Mas um teste formal teria que ser feito.