

# **Modelos Vetoriais Autoregressivos**

Alessandro Martim Marques

**Finanças Quantitativas - MPFE - EESP - FGV**

22 de novembro de 2012

# Sumário

Objetivos

Modelos VAR Estáveis

Representação MA de um processo VAR( $p$ )

Processos Estacionários

Forecasts em Modelos VAR

Estimação dos Modelos VAR

CrITÉrios para Seleção de Modelos

# Modelos Vetoriais Autoregressivos

## Objetivos

Modelos VAR Estáveis

Representação MA de um processo VAR( $p$ )

Processos Estacionários

Forecasts em Modelos VAR

Estimação dos Modelos VAR

CrITÉrios para Seleção de Modelos

# Objetivos

Se **possuímos observações passadas** de uma série temporal e se os **dados passados contém informação** sobre o comportamento futuro de alguma variável, é plausível utilizar como *forecast* alguma função dos dados coletados.

Formalmente: sendo  $y_t$  o valor da variável de interesse no período  $t$ , um *forecast* para o período  $T + h$ , feito no final do período  $T$ , pode ter a forma

$$\hat{y}_{T+h} = f(y_T, y_{T-1}, \dots) \quad (1)$$

O maior objetivo da **análise univariada** de séries temporais é especificar as formas sensatas para  $f(\bullet)$ .

Em muitas aplicações, **funções lineares** têm sido utilizada, de modo que, por exemplo, pode-se fazer a seguinte especificação:

$$\hat{y}_{T+h} = v + \alpha_1 y_T + \alpha_2 y_{T-1} + \dots$$

Ao lidar-se com **variáveis econômicas** freqüentemente os valores de uma variável dependem não apenas de seus próprios valores passados mas também dos valores passados de outras variáveis relacionadas à primeira.

Denotando as variáveis relacionadas por  $y_{1t}, y_{2t}, \dots, y_{Kt}$  temos formalmente:

$$\hat{y}_{1,T+h} = f_1(y_{1,T}, y_{2,T}, \dots, y_{K,T}, y_{1,T-1}, y_{2,T-1}, \dots, y_{K,T-1}, y_{1,T-2}, \dots)$$

Esta última especificação pode valer para todo um conjunto de variáveis inter-relacionadas de modo que, de uma forma geral, o *forecast* para a  $k$ -ésima variável pode ser expresso como:

$$\hat{y}_{k,T+h} = f_k(y_{1,T}, \dots, y_{K,T}, y_{1,T-1}, \dots, y_{K,T-1}, \dots) \quad (2)$$

Um conjunto de séries temporais

$$y_{kt} \quad k = 1, \dots, K \quad t = 1, \dots, T$$

é chamado uma **série temporal múltipla** e a equação (2) expressa o *forecast*  $\hat{y}_{k,T+h}$  como uma função dessa série temporal múltipla.

Em analogia com o caso univariado, o objetivo principal da análise de séries temporais múltiplas é especificar formas adequadas para as funções  $f_1, \dots, f_K$  as quais possam ser utilizadas para bons *forecasts*.

# Processos autoregressivos vetoriais

Por razões de simplicidade comecemos com *forecasts* lineares. Consideremos uma série univariada e um *forecast* de  $h = 1$  período no futuro. Se  $f(\bullet)$  na equação (1) é linear, teremos

$$\hat{y}_{T+1} = \nu + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots .$$

Assumindo que apenas um número finito  $p$  das observações passadas de  $y$  são utilizados na fórmula de predição obtemos

$$\hat{y}_{T+1} = \nu + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots + \alpha_p y_{T-p+1} . \quad (3)$$

O valor verdadeiro  $y_{T+1}$  não será, em geral, igual ao *forecast*  $\hat{y}_{T+1}$ . Haverá um erro no *forecast*, o qual denotaremos por

$$u_{T+1} := y_{T+1} - \hat{y}_{T+1}$$

de modo que

$$\begin{aligned} y_{T+1} &= \hat{y}_{T+1} + u_{T+1} \\ &= \nu + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots + \alpha_p y_{T-p+1} + u_{T+1} \end{aligned} \quad (4)$$



Assumindo agora que os valores da série temporal são **realizações de uma variável aleatória** e que o mesmo **processo gerador de dados** prevalece em cada período, a equação (4) tem a forma de um **processo autoregressivo**,

$$y_t = \nu + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + u_t, \quad (5)$$

onde as quantidades  $y_t, y_{t-1}, \dots, y_{t-p}$  e  $u_t$  são agora **variáveis aleatórias**.

Assumindo agora que os valores da série temporal são **realizações de uma variável aleatória** e que o mesmo **processo gerador de dados** prevalece em cada período, a equação (4) tem a forma de um **processo autoregressivo**,

$$y_t = \nu + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + u_t, \quad (5)$$

onde as quantidades  $y_t, y_{t-1}, \dots, y_{t-p}$  e  $u_t$  são agora **variáveis aleatórias**.

### Importante

Para verdadeiramente obtermos um processo AR assumimos que os erros  $u_s$  e  $u_t$  são não correlacionados para  $s \neq t$ . Isto equivale a dizer que toda a **informação útil** contida no passado de  $y_t$  foi utilizada nos *forecasts* de modo que **não existem erros sistemáticos** nos *forecasts*.

Se considerarmos agora séries temporais múltiplas a extensão natural de

$$\hat{y}_{T+1} = \nu + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots + \alpha_p y_{T-p+1}$$

seria

$$\begin{aligned} \hat{y}_{k,T+1} &= \nu + \alpha_{k1,1} y_{1,T} + \alpha_{k2,1} y_{2,T} + \cdots + \alpha_{kK,1} y_{K,T} \\ &\quad + \cdots + \alpha_{k1,p} y_{1,T-p+1} + \cdots + \alpha_{kK,p} y_{K,T-p+1} \end{aligned} \quad (6)$$

$$k = 1, \dots, K.$$

Para simplificar a notação, vamos definir

$$y_t := (y_{1t}, \dots, y_{Kt})^T,$$

$$\hat{y}_t := (\hat{y}_{1t}, \dots, \hat{y}_{Kt})^T,$$

$$v_t := (v_{1t}, \dots, v_{Kt})^T$$

e

$$A_i := \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} & \cdots & \alpha_{1K,i} \\ \alpha_{21,i} & \alpha_{22,i} & \cdots & \alpha_{2K,i} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{K1,i} & \alpha_{K2,i} & \cdots & \alpha_{KK,i} \end{bmatrix}.$$

Utilizando esta notação

$$\begin{aligned}\hat{y}_{k,T+1} = & \nu + \alpha_{k1,1} y_{1,T} + \alpha_{k2,1} y_{2,T} + \cdots + \alpha_{kK,1} y_{K,T} \\ & + \cdots + \alpha_{k1,p} y_{1,T-p+1} + \cdots + \alpha_{kK,p} y_{K,T-p+1}\end{aligned}$$

pode ser escrita de forma compacta como

$$\hat{y}_{T+1} = \nu + A_1 y_T + \cdots + A_p y_{T-p+1} . \quad (7)$$

Se os  $y_t$ 's são considerados vetores aleatórios, este preditor é o *forecast* ótimo obtido de um **modelo autoregressivo vetorial** da forma

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad (8)$$

na qual os erros

$$u_t = (u_{1t}, \dots, u_{Kt})^T$$

formam uma sequência de  $K$ -vetores i.i.d, com média zero.

Se os  $y_t$ 's são considerados vetores aleatórios, este preditor é o *forecast* ótimo obtido de um **modelo autoregressivo vetorial** da forma

$$y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \quad (8)$$

na qual os erros

$$u_t = (u_{1t}, \dots, u_{Kt})^T$$

formam uma seqüência de  $K$ -vetores i.i.d, com média zero.

É sempre bom lembrar ...

O modelo descrito pela equação (8) é uma tremenda simplificação se comparado com a forma geral dada pela equação

$$\hat{y}_{k,T+h} = f_k(y_{1,T}, \dots, y_{K,T}, y_{1,T-1}, \dots, y_{K,T-1}, \dots).$$

Por causa desta estrutura simples tal modelo desfruta de grande popularidade.

# Modelos Vetoriais Autoregressivos

Objetivos

**Modelos VAR Estáveis**

Representação MA de um processo VAR( $p$ )

Processos Estacionários

Forecasts em Modelos VAR

Estimação dos Modelos VAR

CrITÉrios para Seleção de Modelos



## VAR( $p$ ) Estável

Seja o seguinte modelo VAR( $p$ ) (VAR de ordem  $p$ ):

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \quad t = 0, \pm 1, \pm 2, \dots \quad (9)$$

no qual

$y_t = (y_{1t}, \dots, y_{Kt})^T$  : vetor aleatório ( $K \times 1$ )

$A_i$  : matrizes ( $K \times K$ ) de coeficientes constantes

$v_t = (v_{1t}, \dots, v_{Kt})^T$  : vetor constante ( $K \times 1$ ) de interceptos

e  $u_t = (u_{1t}, \dots, u_{Kt})^T$  é um **ruído branco**  $K$ -dimensional ou **processo de inovação**, i.e.,

$$\mathbf{E}[u_t] = 0 \quad \mathbf{E}[u_t u_t^T] = \Sigma_u \quad \mathbf{E}[u_t u_s^T] = 0 \text{ para } s \neq t$$

A **matriz de covariância**  $\Sigma_u$  é considerada **não singular**.

Para melhor compreender o processo descrito pela equação

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$$

consideremos o modelo VAR(1):

$$y_t = v + A_1 y_{t-1} + u_t \quad (10)$$

Se o processo inicia em um tempo  $t = 1$  temos

$$\begin{aligned} y_1 &= v + A_1 y_0 + u_1 \\ y_2 &= v + A_1 y_1 + u_2 = v + A_1 (v + A_1 y_0 + u_1) + u_2 \\ &= (\mathbb{I}_k + A_1) v + A_1^2 y_0 + A_1 u_1 + u_2 \\ &\vdots \\ y_t &= (\mathbb{I}_k + A_1 + \cdots + A_1^{t-1}) v + A_1^t y_0 + \sum_{i=0}^{t-1} A_1^i u_{t-i} \\ &\vdots \end{aligned} \quad (11)$$

Desta última expressão

$$y_t = (\mathbb{1}_K + A_1 + \cdots + A_1^{t-1}) v + A_1^t y_0 + \sum_{i=0}^{t-1} A_1^i u_{t-i}$$

podemos notar que:

- ▶ os vetores  $y_1, \dots, y_t$  são determinados por  $y_0, u_1, \dots, u_t$
- ▶ a **distribuição conjunta**  $y_1, \dots, y_t$  é determinada pela **distribuição conjunta** de  $y_0, u_1, \dots, u_t$ .

Apesar de, em geral, assumirmos que o processo iniciou-se num período específico freqüentemente é conveniente assumir-se que ele iniciou-se num **passado infinito**.

Esse pressuposto já estava presente na definição do processo VAR( $p$ ):

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \quad t = 0, \pm 1, \pm 2, \dots ,$$

Apesar de, em geral, assumirmos que o processo iniciou-se num período específico freqüentemente é conveniente assumir-se que ele iniciou-se num **passado infinito**.

Esse pressuposto já estava presente na definição do processo VAR( $p$ ):

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \quad t = 0, \pm 1, \pm 2, \dots ,$$

**Pergunta: que tipo de processo é consistente com o mecanismo definido na equação acima?**

Consideremos novamente o processo VAR(1): fazendo  $j = t - 1$  na equação (11) obtemos

$$y_t = v + A_1 y_{t-1} + u_t$$

$$y_t = (\mathbb{I}_K + A_1 + \cdots + A_1^j) v + A_1^{j+1} y_{t-j-1} + \sum_{i=0}^j A_1^i u_{t-i}.$$

Se todos os autovalores de  $A_1$  possuem módulo menor do que 1, a seqüência  $A_1^i$ ,  $i = 0, 1, \dots$  é **absolutamente somável**.

Como consequência desta "somabilidade" temos que:

- ▶ a soma infinita

$$\sum_{i=1}^{\infty} A_1^i u_{t-i}$$

existe em média quadrática.

- ▶

$$(\mathbb{I}_k + A_1 + \cdots + A_1^j) v \xrightarrow{j \rightarrow \infty} (\mathbb{I}_k - A_1)^{-1} v$$

- ▶  $A_1^{j+1}$  converge a zero rapidamente com  $j \rightarrow \infty$  o que nos leva a ignorar o termo  $A_1^{j+1} y_{t-j-1}$  neste limite.

Nestas condições dizer que  $y_t$  é o processo VAR(1) da equação

$$y_t = v + A_1 y_{t-1} + u_t$$

equivale a dizer que  $y_t$  é um processo estocástico bem definido o qual pode ser escrito como

$$y_t = \mu + \sum_{i=0}^{\infty} A_1^i u_{t-1}, \quad t = 0, \pm 1, \pm 2, \dots \quad (12)$$

no qual

$$\mu := (\mathbb{I}_k - A_1)^{-1} v.$$

O primeiros momentos do processo  $y_t$  são dados por:

$$\mathbf{E}[y_t] = \mu \quad \forall t. \quad (13)$$



Já os segundos momentos do processo  $y_t$  são dados por:

$$\begin{aligned}
 \Gamma_y(h) &= \mathbf{E}[(y_t - \mu)(y_{t-h} - \mu)^T] \\
 &= \lim_{n \rightarrow \infty} \sum_{i=0}^n \sum_{j=0}^n A_1^i \mathbf{E}[u_{t-i} u_{t-h-j}^T] A_1^{Tj} \\
 &= \lim_{n \rightarrow \infty} \sum_{i=0}^n A_1^{h+i} \Sigma_u A_1^{Tj} \\
 &= \sum_{i=0}^{\infty} A_1^{h+i} \Sigma_u A_1^{Tj}
 \end{aligned} \tag{14}$$

pois

$$\mathbf{E}[u_t u_s^T] = 0 \text{ para } s \neq t$$

e

$$\mathbf{E}[u_t u_t^T] = \Sigma_u \quad \forall t.$$

## Estabilidade:

Dizemos que um processo VAR(1) é **estável** se todos os autovalores de  $A_1$  possuem módulo menor do que 1. Esta condição é equivalente a

$$\det(\mathbb{I}_k - A_1 z) \neq 0 \text{ para } |z| \leq 1. \quad (15)$$

## Estabilidade:

Dizemos que um processo VAR(1) é **estável** se todos os autovalores de  $A_1$  possuem módulo menor do que 1. Esta condição é equivalente a

$$\det(\mathbb{I}_k - A_1 z) \neq 0 \text{ para } |z| \leq 1. \quad (15)$$

## Importante:

O processo  $y_t$  para  $t = 0, \pm 1, \pm 2, \dots$  pode ainda ser definido mesmo quando a condição (15) não é satisfeita. Daqui por diante sempre assumiremos a estabilidade dos processos definidos para todo  $t \in \mathbb{Z}$ .

## Estabilidade:

Dizemos que um processo VAR(1) é **estável** se todos os autovalores de  $A_1$  possuem módulo menor do que 1. Esta condição é equivalente a

$$\det(\mathbb{I}_k - A_1 z) \neq 0 \text{ para } |z| \leq 1. \quad (15)$$

## Importante:

O processo  $y_t$  para  $t = 0, \pm 1, \pm 2, \dots$  pode ainda ser definido mesmo quando a condição (15) não é satisfeita. Daqui por diante sempre assumiremos a estabilidade dos processos definidos para todo  $t \in \mathbb{Z}$ .

A discussão anterior pode ser estendida para processos VAR( $p$ ) com  $p > 1$  pois todo processo VAR( $p$ ) pode ser escrito na forma de um processo VAR(1).

Se  $y_t$  é um processo VAR( $p$ ), o processo VAR(1)  $Kp$ -dimensional correspondente é definido como

$$Y_t = \mathbf{v} + \mathbf{A}Y_{t-1} + U_t \quad (16)$$

no qual

$$Y_t := \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}_{(Kp \times 1)} \quad \mathbf{v} := \begin{bmatrix} v \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(Kp \times 1)}$$

$$\mathbf{A} := \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ \mathbb{1}_K & 0 & \cdots & 0 & 0 \\ 0 & \mathbb{1}_K & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbb{1}_K & 0 \end{bmatrix}_{(Kp \times Kp)} \quad U_t := \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(Kp \times 1)}$$

Agora,  $Y_t$  é **estável** se

$$\det(\mathbb{I}_{\text{kp}} - \mathbf{A}z) \neq 0 \text{ para } |z| \leq 1 \quad (17)$$

Já o vetor média de  $Y_t$  é

$$\boldsymbol{\mu} := \mathbf{E}[Y_t] = (\mathbb{I}_{\text{kp}} - \mathbf{A})^{-1} \boldsymbol{\nu}$$

e suas autocovariâncias são

$$\Gamma_Y(h) = \sum_{i=0}^{\infty} \mathbf{A}^{h+i} \Sigma_U \mathbf{A}^{\text{T}i} \quad (18)$$

na qual

$$\Sigma_U := \mathbf{E}[U_t U_t^{\text{T}}] .$$

Definindo a seguinte matriz  $K \times Kp$ :

$$J := \begin{bmatrix} \mathbb{1}_K : 0 : \cdots : 0 \end{bmatrix} \quad (19)$$

o processo  $y_t$  pode ser recuperado a partir de  $Y_t$  como

$$y_t = J Y_t.$$

A média de  $y_t$  é

$$\mathbf{E}[y_t] = J \boldsymbol{\mu} \quad (\text{constante } \forall t)$$

e suas autocovariâncias

$$\Gamma_y(h) = J \Gamma_Y(h) J^T \quad (\text{também invariantes})$$

Não é muito difícil ver que

$$\det(\mathbb{1}_{K^p} - \mathbf{A}z) = \det(\mathbb{1}_K - A_1 z - \cdots - A_p z^p)$$

Este é o chamado **polinômio característico reverso** do processo  $\text{VAR}(p)$ .



Não é muito difícil ver que

$$\det(\mathbb{1}_K - \mathbf{A}z) = \det(\mathbb{1}_K - A_1 z - \cdots - A_p z^p)$$

Este é o chamado **polinômio característico reverso** do processo  $\text{VAR}(p)$ .

### Condição de estabilidade:

O processo

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \quad t = 0, \pm 1, \pm 2, \dots$$

é **estável** se seu polinômio característico reverso não possui raízes **sobre ou dentro** do círculo unitário complexo. Formalmente,  $y_t$  é estável se

$$\det(\mathbb{1}_K - A_1 z - \cdots - A_p z^p) \neq 0 \text{ para } |z| \leq 1. \quad (20)$$

**Em resumo:** dizemos que  $y_t$  é um processo VAR( $p$ ) estável se vale a condição (20) e

$$y_t = J Y_t = J \boldsymbol{\mu} + J \sum_{i=0}^{\infty} \mathbf{A}^i U_{t-i}. \quad (21)$$

Devido ao fato de

$$U_t := (u_t^T, 0, \dots, 0)^T$$

envolver o processo de ruído branco  $u_t$ , o processo  $y_t$  é dito ser determinado por seu processo de inovação e freqüentemente são tomados pressupostos específicos relativos a  $u_t$ .

Um exemplo importante é o pressuposto de que  $u_t$  é um **ruído branco gaussiano**, *i.e.*:

$$u_t \sim \mathcal{N}(0, \Sigma_u) \quad \forall t$$

e  $u_t$  e  $u_s$  independentes para  $t \neq s$ .

Neste caso pode-se mostrar que  $y_t$  é um **processo gaussiano**, *i.e.*, subconjuntos  $y_t, \dots, y_{t+h}$  possuem distribuições **normais multivariadas** para todo  $t$  e  $h$ .

A condição de estabilidade da equação (20), *i.e.*,

$$\det(\mathbb{I}_k - A_1 z - \cdots - A_p z^p) \neq 0 \text{ para } |z| \leq 1$$

nos provê uma ferramenta útil para a verificação da estabilidade de um processo VAR.

## Exemplo:

Processo VAR(1) de dimensão 3:

$$y_t = v + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} y_{t-1} + u_t$$

Polinômio característico reverso:

$$\det \left( \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} z \right)$$

## Exemplo: (cont.)

Polinômio característico reverso:

## Exemplo: (cont.)

Polinômio característico reverso:

$$\det \left( \begin{bmatrix} 1 - 0.5z & 0 & 0 \\ -0.1z & 1 - 0.1z & -0.3z \\ 0 & -0.2z & 1 - 0.3z \end{bmatrix} \right) =$$
$$= (1 - 0,5z)(1 - 0,4z - 0,03z^2)$$

## Exemplo: (cont.)

Polinômio característico reverso:

$$\det \left( \begin{bmatrix} 1 - 0.5z & 0 & 0 \\ -0.1z & 1 - 0.1z & -0.3z \\ 0 & -0.2z & 1 - 0.3z \end{bmatrix} \right) =$$
$$= (1 - 0,5z)(1 - 0,4z - 0,03z^2)$$

Raízes do polinômio:

$$z_1 = 2 \quad z_2 = 2,1525 \quad z_3 = -15,4858$$

$|z_1|, |z_2|, |z_3| > 1 \implies$  o processo é **estável**.



## Outro exemplo:

Processo VAR(2) de dimensão 1:

$$y_t = v + \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} y_{t-2} + u_t$$

Polinômio característico reverso:

## Outro exemplo:

Processo VAR(2) de dimensão 1:

$$y_t = v + \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} y_{t-2} + u_t$$

Polinômio característico reverso:

$$\det \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} z \begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} z^2 \right) =$$
$$= 1 - z + 0,21 z^2 - 0,025 z^3$$

## Outro exemplo:

Processo VAR(2) de dimensão 1:

$$y_t = v + \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} y_{t-2} + u_t$$

Polinômio característico reverso:

$$\det \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} z \begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} z^2 \right) =$$
$$= 1 - z + 0,21 z^2 - 0,025 z^3$$

Raízes do polinômio:

$$z_1 = 1,3 \quad z_2 = 3,55 + 4,26 i \quad z_3 = 3,55 - 4,26 i$$

## Outro exemplo:

Processo VAR(2) de dimensão 1:

$$y_t = v + \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} y_{t-2} + u_t$$

Polinômio característico reverso:

$$\det \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} z \begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} z^2 \right) =$$
$$= 1 - z + 0,21 z^2 - 0,025 z^3$$

Raízes do polinômio:

$$z_1 = 1,3 \quad z_2 = 3,55 + 4,26 i \quad z_3 = 3,55 - 4,26 i$$

**Estável?**

## Outro exemplo:

Processo VAR(2) de dimensão 1:

$$y_t = v + \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} y_{t-2} + u_t$$

Polinômio característico reverso:

$$\det \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} z \begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} z^2 \right) =$$
$$= 1 - z + 0,21 z^2 - 0,025 z^3$$

Raízes do polinômio:

$$z_1 = 1,3 \quad z_2 = 3,55 + 4,26 i \quad z_3 = 3,55 - 4,26 i$$

**Estável?**    **Sim!**     $|z_2| = |z_3| = \sqrt{3,55^2 + 4,26^2} = 5,545 > 1.$

# Modelos Vetoriais Autoregressivos

Objetivos

Modelos VAR Estáveis

**Representação MA de um processo VAR( $p$ )**

Processos Estacionários

Forecasts em Modelos VAR

Estimação dos Modelos VAR

CrITÉrios para Seleção de Modelos

# Representação de Média Móvel

**Relembrando:** o processo VAR( $p$ )

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \quad t = 0, \pm 1, \pm 2, \dots$$

tem a seguinte representação VAR(1):

$$Y_t = \mathbf{v} + \mathbf{A}Y_{t-1} + U_t$$

a qual, assumindo estabilidade pode ser escrita como

$$Y_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}^i U_{t-i}. \quad (22)$$

Esta é a chamada **representação de média móvel** do processo, ou **representação MA**.

A representação MA de  $y_t$  pode ser encontrada se fizermos uso novamente da matriz

$$J := \left[ \mathbb{1}_K : 0 : \cdots : 0 \right].$$

Se multiplicarmos

$$Y_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}^i U_{t-i}$$

por  $J$  obtemos:

$$\begin{aligned} y_t &= JY_t = J\boldsymbol{\mu} + \sum_{i=0}^{\infty} J\mathbf{A}^i J^T U_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \end{aligned} \tag{23}$$

na qual

$$\mu := J\boldsymbol{\mu},$$

$$\Phi_i := J\mathbf{A}^i J^T.$$



Agora, dada a estrutura do ruído branco  $U_t$ , temos

$$U_t = J^T J U_t \text{ e } J U_t = u_t.$$

Dado que  $A^i$  é absolutamente somável, o mesmo vale para as matrizes  $\Phi_i$ .

A média de  $y_t$  é dada por

$$\mathbf{E}[y_t] = \mu$$

e suas autocovariâncias

$$\begin{aligned}\Gamma_y(h) &= \mathbf{E}[(y_t - \mu)(y_{t-h} - \mu)^T] \\ &= \mathbf{E}\left[\left(\sum_{i=0}^{h-1} \Phi_i u_{t-i} + \sum_{i=0}^{\infty} \Phi_{h+i} u_{t-h-i}\right)\left(\sum_{i=0}^{\infty} \Phi_i u_{t-h-i}\right)^T\right] \\ &= \sum_{i=0}^{\infty} \Phi_{h+i} \Sigma_u \Phi_i^T.\end{aligned}\tag{24}$$

Existe uma alternativa mais direta para o cálculo das matrizes  $\Phi_i$  através da utilização do **operador de lag**  $L$ . Este operador é definido da seguinte forma:

$$Ly_t = y_{t-1},$$

*i.e.*, ele desloca o índice um período para trás (**operador de backshift**).

Utilizando este operador a formulação original do processo VAR( $p$ ),

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \quad t = 0, \pm 1, \pm 2, \dots$$

pode ser reescrita como

$$y_t = v + (A_1 L + \cdots + A_p L^p) y_t + u_t$$

ou

$$A(L) y_t = v + u_t, \tag{25}$$

na qual

$$A(L) := \mathbb{1}_K - A_1 L - \cdots - A_p L^p.$$

Seja agora

$$\Phi(L) := \sum_{i=0}^{\infty} \Phi_i L^i$$

um operador tal que

$$\Phi(L) A(L) = \mathbb{1}_K. \quad (26)$$

Multiplicando

$$A(L) y_t = v + u_t$$

por  $\Phi(L)$  nos dá

$$\begin{aligned} y_t &= \Phi(L) v + \Phi(L) u_t \\ &= \left( \sum_{i=0}^{\infty} \Phi_i \right) v + \sum_{i=0}^{\infty} \Phi_i u_{t-i}. \end{aligned} \quad (27)$$

O operador  $\Phi(L)$  é o **inverso** de  $A(L)$  e muitas vezes denotado  $A(L)^{-1}$ .

Geralmente o operador  $A(L)$  é dito **inversível** se

$$|A(z)| \neq 0 \quad \text{para} \quad |z| \leq 1 .$$

Se esta condição é satisfeita as matrizes

$$\Phi(L) = A(L)^{-1}$$

são absolutamente somáveis e daí o processo

$$\Phi(L) u_t = A(L)^{-1} u_t$$

é bem definido.

As matrizes  $\Phi_i$  pode ser obtidas de

$$\Phi(L) A(L) = \mathbb{I}_K,$$

utilizando-se

$$\begin{aligned}\mathbb{I}_K &= \Phi(L) A(L) \\ &= (\Phi_0 + \Phi_1 L + \Phi_2 L^2 + \dots) (\mathbb{I}_K - A_1 L - \dots - A_p L^p) \\ &= \Phi_0 + (\Phi_1 - \Phi_0 A_1) L + (\Phi_2 - \Phi_1 A_1 - \Phi_0 A_2) L^2 + \dots \\ &\quad + \left( \Phi_i - \sum_{j=1}^i \Phi_{i-j} A_j \right) L^i + \dots\end{aligned}$$

Esta última equação nos dá:

$$\begin{aligned}
 \mathbb{1}_K &= \Phi_0 \\
 0 &= \Phi_1 - \Phi_0 A_1 \\
 0 &= \Phi_2 - \Phi_1 A_2 - \Phi_0 A_2 \\
 &\vdots \\
 0 &= \Phi_i - \sum_{j=1}^i \Phi_{i-j} A_j \\
 &\vdots
 \end{aligned}
 \tag{28}$$

nas quais  $A_j = 0$  para  $j > p$ .



Assim podemos calcular  $\Phi_i$  recursivamente usando

$$\begin{aligned}\Phi_0 &= \mathbb{I}_K \\ \Phi_i &= \sum_{j=1}^i \Phi_{i-j} A_j, \quad i = 1, 2, \dots\end{aligned}\tag{29}$$

A média  $\mu$  de  $y_t$  é dada por:

$$\mu = \Phi(1) \nu = A(1)^{-1} \nu = (\mathbb{I}_K - A_1 - \dots - A_p)^{-1} \nu\tag{30}$$

Para um processo VAR(1), as recursões da equação (29) implicam em

$$\Phi_0 = \mathbb{I}_K, \quad \Phi_1 = A_1, \dots, \quad \Phi_i = A_1^i, \dots$$

## Exemplo:

Retornando ao VAR(1) de dimensão 3:

$$y_t = v + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} y_{t-1} + u_t$$

Temos  $\Phi_0 = \mathbb{I}_3$ ,

$$\Phi_1 = \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0 \\ 0 & 0.2 & 0.3 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0.25 & 0 & 0 \\ 0.06 & 0.07 & 0.12 \\ 0.02 & 0.08 & 0.15 \end{bmatrix},$$

$$\Phi_3 = \begin{bmatrix} 0.125 & 0 & 0 \\ 0.037 & 0.031 & 0.057 \\ 0.018 & 0.038 & 0.069 \end{bmatrix}, \quad \text{etc.}$$

Já para um processo VAR(2) a recursão (29) resulta em

$$\Phi_1 = A_1$$

$$\Phi_2 = \Phi_1 A_1 + A_2 = A_1^2 + A_2$$

$$\Phi_3 = \Phi_2 A_1 + \Phi_1 A_2 = A_1^3 + A_2 A_1 + A_1 A_2$$

$$\vdots \quad \quad \vdots$$

$$\Phi_i = \Phi_{i-1} A_1 + \Phi_{i-2} A_2$$

$$\vdots \quad \quad \vdots$$

## Outro exemplo:

Retornando ao VAR(2) de dimensão 2:

$$y_t = v + \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} y_{t-2} + u_t$$

Temos as seguintes matrizes de coeficientes MA:  $\Phi_0 = \mathbb{I}_2$ ,

$$\Phi_1 = \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0.29 & 0.1 \\ 0.65 & 0.29 \end{bmatrix}, \quad \Phi_3 = \begin{bmatrix} 0.21 & 0.079 \\ 0.566 & 0.21 \end{bmatrix}, \quad \text{etc.}$$

É importante notar que, para ambos os processos usados como exemplos,

$$\lim_{i \rightarrow \infty} \Phi_i = 0.$$

Esta é uma **consequência da estabilidade** de ambos os processos.

Faz-se importante é o fato de que a representação MA de um processo VAR( $p$ ) estável **não é necessariamente de ordem infinita**.

Ou seja,  $\Phi_i$  pode ser **zero** para  $i$  maior que algum inteiro finito  $q$ .

## Exemplo: VAR(1) bivariado

$$y_t = v + \begin{bmatrix} 0 & \alpha \\ 0 & 0 \end{bmatrix} y_{t-1} + u_t$$

Representação MA:

$$y_t = \mu + u_t + \begin{bmatrix} 0 & \alpha \\ 0 & 0 \end{bmatrix} u_{t-1}$$

pois

$$\begin{bmatrix} 0 & \alpha \\ 0 & 0 \end{bmatrix}^i = 0 \text{ para } i > 1.$$

# Modelos Vetoriais Autoregressivos

Objetivos

Modelos VAR Estáveis

Representação MA de um processo  $\text{VAR}(p)$

**Processos Estacionários**

Forecasts em Modelos VAR

Estimação dos Modelos VAR

Critérios para Seleção de Modelos

## ]Processos Estacionários

### Relembrando ...

Um processo estocástico é dito estacionário se seu primeiro e segundo momentos são invariantes com o tempo.

Formalmente, um processo estocástico  $y_t$  é estacionário se:

$$\mathbf{E}[y_t] = \mu < \infty \quad \forall t \quad (31)$$

(**todo  $y_t$  tem o mesmo vetor de médias finito  $\mu$** ) e

$$\mathbf{E}[(y_t - \mu)(y_{t-h} - \mu)^T] = \Gamma_y(h) = \Gamma_y^T(-h) < \infty \quad \forall t, h = 0, 1, 2, \dots \quad (32)$$

(**as autocovariâncias não dependem de  $t$  mas apenas do intervalo  $h$  que separa  $y_t$  e  $y_{t-h}$** )



O ruído branco que utilizamos na definição do processo VAR( $p$ )

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \quad t = 0, \pm 1, \pm 2, \dots$$

é um exemplo de processo estacionário.

Já vimos também que, para um processo VAR( $p$ ) estável

$$\mathbf{E}[y_t] = \mu \quad \text{e} \quad \Gamma_y(h) = \sum_{i=0}^{\infty} \Phi_{h+i} \Sigma_u \Phi_i^T$$

ou seja:

**Proposição:** (condição de estacionariedade)

Um processo VAR( $p$ ) estável  $y_t$ ,  $t = 0, \pm 1, \pm 2, \dots$  é estacionário.

### Importante:

1. Estabilidade implica estacionariedade e por isso a condição de estabilidade

$$\det(\mathbb{I}_k - A_1 z - \cdots - A_p z^p) \neq 0 \text{ para } |z| \leq 1$$

é muitas vezes chamada de **condição de estacionariedade**.

2. **O reverso da proposição acima não é verdade**. Ou seja, um processo não estável não é necessariamente não estacionário.

# Cálculo de Autocovariâncias e Autocorrelações

É possível obter-se expressões analíticas para as matrizes de autocovariâncias e autocorrelações para processos  $\text{VAR}(p)$  estáveis.

Os resultados são obtidos a partir das chamadas **equações de Yule-Walker** e os detalhes podem ser encontrados em [Lütkepohl 2005, capítulo 2, seção 2.1.4]

# Modelos Vetoriais Autoregressivos

Objetivos

Modelos VAR Estáveis

Representação MA de um processo VAR( $p$ )

Processos Estacionários

**Forecasts em Modelos VAR**

Estimação dos Modelos VAR

Critérios para Seleção de Modelos

## Forecasts

Situação:

- ▶ Precisamos fazer afirmações sobre valores futuros  $y_1, \dots, y_K$
- ▶ Temos em mãos:
  1. um modelo para o processo gerador de dados e
  2. um conjunto informacional  $\Omega_t$  contendo a informação disponível no período  $t$ .

# Forecasts

Situação:

- ▶ Precisamos fazer afirmações sobre valores futuros  $y_1, \dots, y_K$
- ▶ Temos em mãos:
  1. um modelo para o processo gerador de dados e
  2. um conjunto informacional  $\Omega_t$  contendo a informação disponível no período  $t$ .

Exemplo:

- ▶ PGD: processo VAR( $p$ )
- ▶  $\Omega_t = \{y_s | s \leq t\}$
- ▶  $y_s = y_{1s}, \dots, y_{Ks}^T$
- ▶  $t$  é a **origem do forecast**
- ▶  $h$  é o **horizonte do forecast**
- ▶ um preditor,  $h$  períodos à frente, é um **preditor de  $h$ -passos**.

**Estratégia comum:** uma **função custo** específica é associada aos erros de *forecast* e um *forecast* é considerado **ótimo** se ele minimiza essa função custo.

**Estratégia comum:** uma **função custo** específica é associada aos erros de *forecast* e um *forecast* é considerado **ótimo** se ele minimiza essa função custo.

No contexto de modelos VAR, preditores que minimizam o erro quadrático médio (MSE) são os mais utilizados.



**Estratégia comum:** uma **função custo** específica é associada aos erros de *forecast* e um *forecast* é considerado **ótimo** se ele minimiza essa função custo.

No contexto de modelos VAR, preditores que minimizam o erro quadrático médio (MSE) são os mais utilizados.

Seja  $y_t = y_{1t}, \dots, y_{Kt}^\top$  um processo VAR( $p$ )  $K$ -dimensional estável. O preditor de mínimo MSE para um *forecast* com origem em  $t$  e horizonte  $h$  é esperança condicional

$$\mathbf{E}_t[y_{t+h}] := \mathbf{E}[y_{t+h} | \Omega_t] = \mathbf{E}[y_{t+h} | \{y_s | s \leq t\}] . \quad (33)$$

**Estratégia comum:** uma **função custo** específica é associada aos erros de *forecast* e um *forecast* é considerado **ótimo** se ele minimiza essa função custo.

No contexto de modelos VAR, preditores que minimizam o erro quadrático médio (MSE) são os mais utilizados.

Seja  $y_t = y_{1t}, \dots, y_{Kt}^T$  um processo VAR( $p$ )  $K$ -dimensional estável. O preditor de mínimo MSE para um *forecast* com origem em  $t$  e horizonte  $h$  é esperança condicional

$$\mathbf{E}_t[y_{t+h}] := \mathbf{E}[y_{t+h} | \Omega_t] = \mathbf{E}[y_{t+h} | \{y_s | s \leq t\}]. \quad (33)$$

Este preditor minimiza o MSE para cada componente de  $y_t$ .

Formalmente, se  $\bar{y}_t(h)$  é qualquer preditor de  $h$ -passos na origem  $t$ ,

$$\begin{aligned}\mathbf{MSE}[\bar{y}_t(h)] &= \mathbf{E}[(y_{t+h} - \bar{y}_t(h))(y_{t+h} - \bar{y}_t(h))^T] \\ &\geq \mathbf{MSE}[\mathbf{E}_t[y_{t+h}]] \\ &= \mathbf{E}[(y_{t+h} - \mathbf{E}_t[y_{t+h}])(y_{t+h} - \mathbf{E}_t[y_{t+h}])^T], \quad (34)\end{aligned}$$

na qual o sinal de desigualdade entre duas matrizes significa que a diferença é uma matriz não negativa definida (também denominada positiva semidefinida).

De forma equivalente, para qualquer vetor  $c$  ( $K \times 1$ ),

$$\mathbf{MSE}[c^T \bar{y}_t(h)] \geq \mathbf{MSE}[c^T \mathbf{E}_t[y_{t+h}]]$$

Uma forma de verificar-se o carácter ótimo da esperança condicional é notarmos que:

$$\begin{aligned}\mathbf{MSE}[\bar{y}_t(h)] &= \mathbf{E}[(y_{t+h} - \mathbf{E}_t[y_{t+h}] - \bar{y}_t(h)) \times \\ &\quad \times (y_{t+h} - \mathbf{E}_t[y_{t+h}] - \bar{y}_t(h))^T] \\ &= \mathbf{MSE}[\mathbf{E}_t[y_{t+h}]] + \\ &\quad + \mathbf{E}[(\mathbf{E}_t[y_{t+h}] - \bar{y}_t(h)) (\mathbf{E}_t[y_{t+h}] - \bar{y}_t(h))^T]\end{aligned}$$

na qual foi usado que

$$\mathbf{E}[(y_{t+h} - \mathbf{E}_t[y_{t+h}]) (y_{t+h} - \mathbf{E}_t[y_{t+h}])^T] = 0$$

A última igualdade

$$\mathbf{E}[(y_{t+h} - \mathbf{E}_t[y_{t+h}]) (y_{t+h} - \mathbf{E}_t[y_{t+h}])^T] = 0$$

é válida pois

$$(y_{t+h} - \mathbf{E}_t[y_{t+h}])$$

é uma função das inovações posteriores ao período  $t$  e, portanto, não correlacionadas com os termos contidos em

$$(\mathbf{E}_t[y_{t+h}] - \bar{y}_t(h))$$

os quais são funções de  $y_s$ ,  $s \leq t$ .

O caráter ótimo da esperança condicional implica que o preditor de  $h$ -passos ótimo para o processo VAR( $p$ ) é dado por

$$\mathbf{E}_t[y_{t+h}] = \nu + A_1 \mathbf{E}_t[y_{t+h-1}] + \cdots + A_p \mathbf{E}_t[y_{t+h-p}] . \quad (35)$$

O caráter ótimo da esperança condicional implica que o preditor de  $h$ -passos ótimo para o processo VAR( $p$ ) é dado por

$$\mathbf{E}_t[y_{t+h}] = \nu + A_1 \mathbf{E}_t[y_{t+h-1}] + \cdots + A_p \mathbf{E}_t[y_{t+h-p}]. \quad (35)$$

Esta última equação é válida apenas se o processo de inovação é um ruído branco independente tal que  $u_t$  e  $u_s$  são independentes para  $s \neq t$  e, portanto,

$$\mathbf{E}_t[u_{t+h}] = 0 \quad \text{para } h > 0.$$

A equação (35) pode ser usada para calcular-se recursivamente o preditor de  $h$ -passos, começando com  $h = 1$ :

$$\begin{aligned}\mathbf{E}_t[y_{t+1}] &= \nu + A_1 y_t + \cdots + A_p y_{t-p+1} \\ \mathbf{E}_t[y_{t+2}] &= \nu + A_1 \mathbf{E}_t[y_{t+1}] + A_2 y_t + \cdots + A_p y_{t-p+2} \\ &\vdots\end{aligned}$$



A equação (35) pode ser usada para calcular-se recursivamente o preditor de  $h$ -passos, começando com  $h = 1$ :

$$\begin{aligned}\mathbf{E}_t[y_{t+1}] &= \boldsymbol{\nu} + A_1 y_t + \cdots + A_p y_{t-p+1} \\ \mathbf{E}_t[y_{t+2}] &= \boldsymbol{\nu} + A_1 \mathbf{E}_t[y_{t+1}] + A_2 y_t + \cdots + A_p y_{t-p+2} \\ &\vdots\end{aligned}$$

Utilizando essas recursões obtemos, para o processo VAR(1):

$$\mathbf{E}_t[y_{t+h}] = \left( \mathbb{1}_K + A_1 + \cdots + A_1^{h-1} \right) \boldsymbol{\nu} + A_1^h y_t.$$

## Exemplo:

Retornando ao VAR(1) de dimensão 3:

$$y_t = v + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} y_{t-1} + u_t$$

Assumindo  $y_t = (-6, 3, 5)^T$  e  $v = (0, 2, 1)^T$  teremos:

$$\mathbf{E}_t[y_{t+1}] =$$

## Exemplo:

Retornando ao VAR(1) de dimensão 3:

$$y_t = v + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} y_{t-1} + u_t$$

Assumindo  $y_t = (-6, 3, 5)^T$  e  $v = (0, 2, 1)^T$  teremos:

$$\mathbf{E}_t[y_{t+1}] = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} -6 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} -3.0 \\ 3.2 \\ 3.1 \end{bmatrix}$$

## Exemplo:

Retornando ao VAR(1) de dimensão 3:

$$y_t = v + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} y_{t-1} + u_t$$

Assumindo  $y_t = (-6, 3, 5)^T$  e  $v = (0, 2, 1)^T$  teremos:

$$\mathbf{E}_t[y_{t+1}] = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} -6 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} -3.0 \\ 3.2 \\ 3.1 \end{bmatrix}$$

$$\mathbf{E}_t[y_{t+2}] =$$

## Exemplo:

Retornando ao VAR(1) de dimensão 3:

$$y_t = v + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} y_{t-1} + u_t$$

Assumindo  $y_t = (-6, 3, 5)^T$  e  $v = (0, 2, 1)^T$  teremos:

$$\mathbf{E}_t[y_{t+1}] = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} -6 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} -3.0 \\ 3.2 \\ 3.1 \end{bmatrix}$$

$$\mathbf{E}_t[y_{t+2}] = (\mathbb{I}_3 + A_1) v + A_1^2 y_t = \begin{bmatrix} -1.50 \\ 2.95 \\ 2.57 \end{bmatrix}, \text{ etc.}$$

## Outro exemplo:

Retornando ao VAR(2) de dimensão 2:

$$y_t = v + \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} y_{t-2} + u_t$$

Com  $y_t = (0.06, 0.03)^T$ ,  $y_{t-1} = (0.055, 0.03)^T$  e  $v = (0.02, 0.03)^T$  teremos:

## Outro exemplo:

Retornando ao VAR(2) de dimensão 2:

$$y_t = v + \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} y_{t-2} + u_t$$

Com  $y_t = (0.06, 0.03)^T$ ,  $y_{t-1} = (0.055, 0.03)^T$  e  $v = (0.02, 0.03)^T$  teremos:

$$\begin{aligned} \mathbf{E}_t[y_{t+1}] &= \begin{bmatrix} 0.02 \\ 0.03 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} \begin{bmatrix} 0.06 \\ 0.03 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} \begin{bmatrix} 0.055 \\ 0.03 \end{bmatrix} \\ &= \begin{bmatrix} 0.053 \\ 0.08275 \end{bmatrix} \end{aligned}$$

Outro exemplo: (cont.)



## Outro exemplo: (cont.)

$$\begin{aligned}\mathbf{E}_t[y_{t+2}] &= \begin{bmatrix} 0.02 \\ 0.03 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} \begin{bmatrix} 0.053 \\ 0.08275 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} \begin{bmatrix} 0.06 \\ 0.03 \end{bmatrix} \\ &= \begin{bmatrix} 0.0548 \\ 0.1076 \end{bmatrix}\end{aligned}$$

A esperança condicional tem as seguintes propriedades:

1. É um preditor não viesado (*unbiased*):

$$\mathbf{E}[y_{t+h} - \mathbf{E}_t[y_{t+h}]] = 0.$$

2. Se  $u_t$  é independente,

$$\mathbf{MSE}[\mathbf{E}_t[y_{t+h}]] = \mathbf{MSE}[\mathbf{E}_t[y_{t+h} | y_t, y_{t-1}, \dots]],$$

ou seja, o MSE do preditor é igual ao MSE condicional dados  $y_t, y_{t-1}, \dots$

## Intervalos de Confiança para os *Forecasts*

Para realizar-se o cálculo dos intervalos ou regiões de confiança faz-se normalmente o pressuposto de que estamos lidando com **processos gaussianos** nos quais  $y_t, y_{t+1}, \dots, y_{t+h}$  possuem uma distribuição multivariada normal para quaisquer  $t$  e  $h$ .

De forma equivalente pode-se supor que  $u_t$  é multivariado normal, *i.e.*,

$$u_t \sim \mathcal{N}(0, \Sigma_u)$$

com  $u_t$  e  $u_s$  independentes para  $t \neq s$ .

Sob estas condições **os erros de *forecast* também são normalmente distribuídos** pois são transformações lineares de vetores gaussianos:

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} \sim \mathcal{N}(0, \Sigma_y(h)). \quad (36)$$

Sob estas condições **os erros de *forecast* também são normalmente distribuídos** pois são transformações lineares de vetores gaussianos:

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} \sim \mathcal{N}(0, \Sigma_y(h)). \quad (36)$$

Este resultado implica em que os erros de *forecast* de cada componente individual do processo possuem distribuição gaussiana de modo que:

$$\frac{y_{k,t+h} - y_{k,t}(h)}{\sigma_k(h)} \sim \mathcal{N}(0, 1), \quad (37)$$

na qual

- ▶  $y_{k,t}(h)$  é a  $k$ -ésima componente de  $y_t(h)$
- ▶  $\sigma_k(h)$  é a raiz quadrada do  $k$ -ésimo elemento da diagonal de  $\Sigma_y(h)$

Denotando por  $z_{(\alpha)}$  o  $\alpha \times 100$  ponto percentual superior da distribuução normal obtemos

$$\begin{aligned}
 1 - \alpha &= \Pr \left[ -z_{(\alpha/2)} \leq \frac{y_{k,t+h} - y_{k,t}(h)}{\sigma_k(h)} \leq z_{(\alpha/2)} \right] \\
 &= \Pr \left[ y_{k,t}(h) - z_{(\alpha/2)} \sigma_k(h) \leq y_{k,t+h} \leq y_{k,t}(h) + z_{(\alpha/2)} \sigma_k(h) \right].
 \end{aligned}$$

Portanto, um intervalo de confiança  $(1 - \alpha)100\%$ , do *forecast*  $h$  períodos à frente, para o  $k$ -ésimo componente de  $y_t$  é dado por

$$y_{k,t}(h) \pm z_{(\alpha/2)} \sigma_k(h) \quad (38)$$

ou

$$\left[ y_{k,t}(h) - z_{(\alpha/2)} \sigma_k(h) , y_{k,t}(h) + z_{(\alpha/2)} \sigma_k(h) \right]. \quad (39)$$

O resultado (36), *i.e.*,

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} \sim \mathcal{N}(0, \Sigma_y(h))$$

também pode ser utilizado para estabelecer intervalos de confiança conjuntos para duas ou mais variáveis.

Os detalhes, que fazem uso do chamado **método de Bonferroni**, podem ser encontrados em [Lütkepohl 2005, capítulo 2, seção 2.2.3].

# Análise Estrutural com Modelos VAR

Análise de certos aspectos das relações entre variáveis de interesse presentes num modelo do tipo VAR.

Estes aspectos incluem:

- ▶ Causalidade de Granger
- ▶ Causalidade Instantânea
- ▶ Causalidade *Multi-Step*
- ▶ Análise de Impulso-Resposta

Discussão completa e detalhada em [Lütkepohl 2005, capítulo 2, seção 2.3].



# Modelos Vetoriais Autoregressivos

Objetivos

Modelos VAR Estáveis

Representação MA de um processo VAR( $p$ )

Processos Estacionários

Forecasts em Modelos VAR

**Estimação dos Modelos VAR**

Critérios para Seleção de Modelos

## Estimação do VAR

Vamos assumir possuímos os dados de uma série temporal múltipla,  $K$ -dimensional,

$$y_1, \dots, y_T \quad \text{com} \quad y_t = (y_{1t}, \dots, y_{Kt})^T$$

e que sabemos que esta série é gerada por um processo VAR( $p$ ) estável, *i.e.*:

$$y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \quad (40)$$

com

$$\mathbf{E}[y_t] = \mu \quad \forall t$$

e

$$\mathbf{E}[(y_t - \mu)(y_{t-h} - \mu)^T] = \Gamma_y(h) = \Gamma_y^T(-h) \quad \forall t, h = 0, 1, 2, \dots$$

Assumimos agora que as matrizes de coeficientes  $v, A_1, \dots, A_p$ , bem como  $\Sigma_u$  são desconhecidos.

Nossa tarefa agora é, a partir dos dados da série temporal, estimar os coeficientes.

Esta tarefa será realizada utilizando-se uma **estimação de mínimos quadrados multivariada**.

Vamos assumir que temos uma **amostra de comprimento  $T$**  para cada uma das  $K$  variáveis, todas as amostras correspondendo ao mesmo período temporal.

Adicionalmente temos à nossa disposição  $p$  valores pré-amostrais de cada variável:

$$y_{-p+1}, y_{-p+2}, \dots, y_{-1}, y_0.$$

Este particionamento em amostra e pré-amostra é conveniente para simplificação notacional

Em seguida definimos:

$$\begin{aligned}
 Y &:= (y_1, \dots, y_T) & (K \times T) \\
 B &:= (v, A_1, \dots, A_p) & (K \times (Kp + 1)) \\
 Z_t &:= \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} & ((Kp + 1) \times 1) \\
 Z &:= (Z_0, \dots, Z_{T-1}) & ((Kp + 1) \times t) \\
 U &:= (u_1, \dots, u_T) & (K \times T) \\
 \mathbf{y} &:= \text{vec}(Y) & (KT \times 1) \\
 \boldsymbol{\beta} &:= \text{vec}(B) & ((K^2 p + K) \times 1) \\
 \mathbf{b} &:= \text{vec}(B^T) & ((K^2 p + K) \times 1) \\
 \mathbf{u} &:= \text{vec}(U) & (KT \times 1)
 \end{aligned} \tag{41}$$

Fazendo uso destas últimas definições, para  $t = 1, \dots, T$ , o modelos VAR( $p$ )

$$y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t$$

pode ser reescritos das seguintes três formas compactas:

$$Y = BZ + U \quad (42)$$

ou

$$\text{vec}(Y) = \text{vec}(BZ) + \text{vec}(U) = (Z^T \otimes \mathbb{I}_K) \text{vec}(B) + \text{vec}(U)$$

ou

$$\mathbf{y} = (Z^T \otimes \mathbb{I}_K) \boldsymbol{\beta} + \mathbf{u}. \quad (43)$$

Cabe notar que a matriz de covariância de  $\mathbf{u}$  é

$$\Sigma_{\mathbf{u}} = \mathbb{I}_T \otimes \Sigma_u. \quad (44)$$

Sendo assim, a estimação multivariada de mínimos quadrados para  $\boldsymbol{\beta}$  equivale escolher um estimador que minimize

$$\begin{aligned}
 S(\boldsymbol{\beta}) &= \mathbf{u}^T (\mathbb{1}_T \otimes \Sigma_u)^{-1} \mathbf{u} = \mathbf{u}^T (\mathbb{1}_T \otimes \Sigma_u^{-1})^{-1} u \\
 &= [\mathbf{y} - (Z^T \otimes \mathbb{1}_K) \boldsymbol{\beta}]^T (\mathbb{1}_T \otimes \Sigma_u^{-1})^{-1} [\mathbf{y} - (Z^T \otimes \mathbb{1}_K) \boldsymbol{\beta}] \\
 &= \text{vec}(Y - BZ)^T (\mathbb{1}_T \otimes \Sigma_u^{-1})^{-1} \text{vec}(Y - BZ) \\
 &= \text{Tr}[(Y - BZ)^T \Sigma_u^{-1} (Y - BZ)] .
 \end{aligned} \tag{45}$$

Para encontrar o mínimo desta função notemos que

$$\begin{aligned}
 S(\boldsymbol{\beta}) &= \mathbf{y}^T (\mathbb{1}_T \otimes \Sigma_u^{-1}) \mathbf{y} \\
 &\quad + \boldsymbol{\beta}^T (Z \otimes \mathbb{1}_K) (\mathbb{1}_T \otimes \Sigma_u^{-1}) (Z^T \otimes \mathbb{1}_K) \boldsymbol{\beta} \\
 &\quad - 2 \boldsymbol{\beta}^T (Z \otimes \mathbb{1}_K) (\mathbb{1}_T \otimes \Sigma_u^{-1}) \mathbf{y} \\
 &= \mathbf{y}^T (\mathbb{1}_T \otimes \Sigma_u^{-1}) \mathbf{y} \\
 &\quad + \boldsymbol{\beta}^T (Z Z^T \otimes \Sigma_u^{-1}) \boldsymbol{\beta} \\
 &\quad - 2 \boldsymbol{\beta}^T (Z \otimes \Sigma_u^{-1}) \mathbf{y}.
 \end{aligned}$$

Isto nos permite escrever a seguinte derivada:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 2 (Z Z^T \otimes \Sigma_u^{-1}) \boldsymbol{\beta} - 2 (Z \otimes \Sigma_u^{-1}) \mathbf{y}.$$

Igualando a derivada a zero obtemos as chamadas **equações normais**:

$$(Z Z^T \otimes \Sigma_u^{-1}) \hat{\boldsymbol{\beta}} = (Z \otimes \Sigma_u^{-1}) \mathbf{y} \quad (46)$$

e, conseqüentemente, o estimador de mínimos quadrados será:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= ((Z Z^T)^{-1} \otimes \Sigma_u^{-1}) (Z \otimes \Sigma_u^{-1}) \mathbf{y} \\ &= ((Z Z^T)^{-1} Z \otimes \mathbb{1}_k) \mathbf{y}. \end{aligned} \quad (47)$$

A matriz hessiana de  $S(\boldsymbol{\beta})$

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = 2 (Z Z^T \otimes \Sigma_u^{-1}),$$

é positiva definida, confirmando que  $\hat{\boldsymbol{\beta}}$  é o vetor que minimiza  $S(\boldsymbol{\beta})$ .



O estimador de mínimos quadrados pode ser escrito de diferentes formas:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= ((ZZ^T)^{-1}Z \otimes \mathbb{I}_k) [(Z^T \otimes \mathbb{I}_k)\boldsymbol{\beta} + \mathbf{u}] \\ &= \boldsymbol{\beta} + ((ZZ^T)^{-1}Z \otimes \mathbb{I}_k) \mathbf{u}\end{aligned}\tag{48}$$

ou

$$\begin{aligned}\text{vec}(\hat{B}) = \hat{\boldsymbol{\beta}} &= ((ZZ^T)^{-1}Z \otimes \mathbb{I}_k) \text{vec}(Y) \\ &= \text{vec}(YZ^T(ZZ^T)^{-1}).\end{aligned}$$

Portanto,

$$\begin{aligned}\hat{B} &= YZ^T(ZZ^T)^{-1} \\ &= (BZ + U)Z^T(ZZ^T)^{-1} \\ &= B + UZ^T(ZZ^T)^{-1}\end{aligned}\tag{49}$$

Outra possibilidade é multiplicar

$$y_t = BZ_{t-1} + u_t$$

por  $Z_{t-1}^T$  e tomar o valor esperado:

$$\mathbf{E}[y_t Z_{t-1}^T] = B\mathbf{E}[Z_{t-1} Z_{t-1}^T] . \quad (50)$$

Estimando  $\mathbf{E}[y_t Z_{t-1}^T]$  por

$$\frac{1}{T} \sum_{t=1}^T y_t \mathbf{Z}_{t-1}^T = \frac{1}{T} \mathbf{Y} \mathbf{Z}^T ,$$

e  $\mathbf{E}[Z_{t-1} Z_{t-1}^T]$  por

$$\frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{t-1} \mathbf{Z}_{t-1}^T = \frac{1}{T} \mathbf{Z} \mathbf{Z}^T$$

Obtemos as equações normais

$$\frac{1}{T} \mathbf{Y} \mathbf{Z}^T = \hat{\mathbf{B}} \frac{1}{T} \mathbf{Z} \mathbf{Z}^T$$

e, desta última,

$$\hat{\mathbf{B}} = \mathbf{Y} \mathbf{Z}^T (\mathbf{Z} \mathbf{Z}^T)^{-1}.$$

Uma terceira possibilidade é escrever o estimador de mínimos quadrados como

$$\hat{\mathbf{b}} = \text{vec}(\hat{\mathbf{B}}^T) = (\mathbb{I}_K \otimes (\mathbf{Z}\mathbf{Z}^T)^{-1}\mathbf{Z}) \text{vec}(\mathbf{Y}^T) .$$

Escrito desta forma é fácil ver que a estimação de mínimos quadrados multivariada é equivalente à estimação de mínimos quadrados ordinária, separadamente, para cada uma das  $K$  equações em

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t .$$

Mostremos este fato ...

Seja  $b^T_k$  a  $k$ -ésima linha de  $B$ , *i.e.*,  $b_k$  contém todos os parâmetros da  $k$ -ésima equação.

Obviamente  $\mathbf{b}^T = (b^T_1, \dots, b^T_k)$ .

Além disso, seja

$$y(k) = (y_{k1}, \dots, y_{kT})^T$$

a série temporal disponível para a  $k$ -ésima variável, tal que

$$\text{vec}(\mathbf{Y}^T) = \begin{bmatrix} y_{(1)} \\ \vdots \\ y_{(K)} \end{bmatrix}.$$

Com esta notação

$$\hat{b}_k = (Z Z^T)^{-1} Z y_{(k)}$$

é o estimador de mínimos quadrados ordinário do modelo

$$y_{(k)} = Z^T b_k + u_{(k)},$$

no qual

$$u_{(k)} = (u_{k1}, \dots, u_{kT})^T$$

e

$$\mathbf{b}^T = (\hat{b}_1^T, \dots, \hat{b}_K^T).$$

Das propriedades assintóticas do estimador de mínimos quadrados e assumindo-se normalidade do ruído branco obtém-se o seguinte estimador para matriz de covariâncias do ruído:

$$\hat{\Sigma}_u = \frac{T}{T - KP - 1} \tilde{\Sigma}_u, \quad (51)$$

na qual

$$\tilde{\Sigma}_u = \frac{1}{T} T (\mathbb{I}_T - Z^T (Z Z^T)^{-1} Z) Z^T. \quad (52)$$

Já para a matriz de covariâncias do estimador  $\hat{\beta}$  pode-se obter:

$$\Sigma_{\beta} = \Gamma^{-1} \otimes \Sigma_u,$$

na qual

$$\Gamma^{-1} = \frac{ZZ^T}{T}.$$

Maiores detalhes sobre a derivação destas matrizes de covariâncias podem ser encontrados em [Lütkepohl 2005, capítulo 3, seção 3.2.2].



# Modelos Vetoriais Autoregressivos

Objetivos

Modelos VAR Estáveis

Representação MA de um processo VAR( $p$ )

Processos Estacionários

Forecasts em Modelos VAR

Estimação dos Modelos VAR

**CrITÉrios para Seleção de Modelos**

# Critérios para Seleção de Modelos I

Quando o principal objetivo de um modelo VAR é prover um bom *forecast* é importante selecionar a ordem correta do modelo utilizado com base neste objetivo.

Tendo este objetivo em mente faz sentido escolher a ordem do modelo tal que nossos *forecasts* tenham a melhor precisão.

Sendo assim o **erro quadrático médio** (MSE) do *forecast* é a medida que buscamos e um primeiro critério baseado nesta medida foi construído por Akaike.

Seja  $\tilde{\Sigma}_u(m)$  o estimador de  $\Sigma_u$  (matriz de covariância dos resíduos) resultante do ajuste de um modelo VAR( $m$ )  $K$ -dimensional.

O critério proposto por Akaike é definido como:

### Critério de Erro de Predição Final

$$\begin{aligned} \text{FPE}(m) &= \det \left[ \frac{T + Km + 1}{T} \frac{T}{T - Km - 1} \tilde{\Sigma}_u(m) \right] \\ &= \left[ \frac{T + Km + 1}{T - Km - 1} \right]^K \det \tilde{\Sigma}_u(m) \end{aligned} \quad (53)$$

Baseado no critério FPE, o estimador  $\hat{p}(\text{FPE})$  para  $p$  é escolhido de forma que

$$\text{FPE}(\hat{p}(\text{FPE})) = \min \{ \text{FPE}(m) | m = 0, 1, \dots, M \} .$$

Ou seja,

1. estimam-se modelos  $\text{VAR}(m)$  de ordens  $m = 0, 1, \dots, M$  e os correspondentes  $\text{FPE}(m)$  são calculados,
2. a ordem  $m$  que minimiza o FPE é a ordem escolhida como estimador de  $p$ .

Akaike (novamente!), baseado num raciocínio completamente diferente derivou um outro critério, bastante similar ao FPE, definido da seguinte forma:

### Critério de Informação de Akaike

$$\begin{aligned} \text{AIC}(m) &= \ln |\tilde{\Sigma}_u(m)| + \frac{2}{T} (\# \text{ de parâmetros}) \\ &= \ln |\tilde{\Sigma}_u(m)| + \frac{2mK^2}{T} \end{aligned} \quad (54)$$

Novamente o estimador  $\hat{p}(\text{AIC})$  para  $p$  é escolhido de forma a minimizar o valor de AIC.

A similaridade entre os critérios FPE e AIC vem do fato de que, para uma constante  $N$ ,

$$\frac{T+N}{T-N} = 1 + \frac{2N}{T} + O(T^{-2})$$

a qual permite que escrevamos

$$\ln \text{FPE}(m) = \text{AIC}(m) + \frac{2K}{T} + O(T^{-2})$$

Conseqüentemente, AIC e  $\ln \text{FPE}$  diferem essencialmente por um termo de ordem  $O(T^{-2})$  e, portanto, **os dois critérios são aproximadamente equivalentes para grandes valores de  $T$ .**

É sempre desejável que um estimador tenha certas **propriedades amostrais**.

Uma propriedade assintótica desejável para um estimador é a chamada **consistência**.

Um estimador  $\hat{p}$  da ordem  $p$  de um processo VAR é dito consistente se

$$\lim_{T \rightarrow \infty} \Pr[\hat{p} = p] = 1 . \quad (55)$$

O estimador  $\hat{p}$  é dito **fortemente consistente** se

$$\Pr \left[ \lim_{T \rightarrow \infty} \hat{p} = p \right] = 1 . \quad (56)$$

Desta forma a seleção da ordem de um modelo VAR é dita **consistente** ou **fortemente consistente** se os estimadores resultantes possuem estas propriedades. **Mas ....**

Pode-se mostrar que: **se  $M > p$ ,  $\hat{p}(\text{FPE})$  e  $\hat{p}(\text{AIC})$  são não consistentes.**

Para sanar este problema foram introduzidos dois outros critérios ...



O primeiro desses critérios é devido a Hannan e Quinn:

O primeiro desses critérios é devido a Hannan e Quinn:

### Critério de Hannan-Quinn

$$\begin{aligned}\text{HQ}(m) &= \ln |\tilde{\Sigma}_u(m)| + \frac{2 \ln \ln T}{T} (\# \text{ de parâmetros}) \\ &= \ln |\tilde{\Sigma}_u(m)| + \frac{2 \ln \ln T}{T} m K^2\end{aligned}\tag{57}$$

O primeiro desses critérios é devido a Hannan e Quinn:

### Critério de Hannan-Quinn

$$\begin{aligned}\text{HQ}(m) &= \ln |\tilde{\Sigma}_u(m)| + \frac{2 \ln \ln T}{T} (\# \text{ de parâmetros}) \\ &= \ln |\tilde{\Sigma}_u(m)| + \frac{2 \ln \ln T}{T} m K^2\end{aligned}\tag{57}$$

O estimador  $\hat{p}(\text{HQ})$  é a ordem que minimiza  $\text{HQ}(m)$  para  $m = 0, 1, \dots, M$ .

O primeiro desses critérios é devido a Hannan e Quinn:

### Critério de Hannan-Quinn

$$\begin{aligned}\text{HQ}(m) &= \ln |\tilde{\Sigma}_u(m)| + \frac{2 \ln \ln T}{T} (\# \text{ de parâmetros}) \\ &= \ln |\tilde{\Sigma}_u(m)| + \frac{2 \ln \ln T}{T} m K^2\end{aligned}\tag{57}$$

O estimador  $\hat{p}(\text{HQ})$  é a ordem que minimiza  $\text{HQ}(m)$  para  $m = 0, 1, \dots, M$ .

**Pode-se mostrar que HQ é fortemente consistente para  $K > 1$ .**

Já o segundo critério foi derivado por Schwarz e é baseado em critério bayesianos:

Já o segundo critério foi derivado por Schwarz e é baseado em critério bayesianos:

### Critério de Schwarz

$$\begin{aligned}\text{HQ}(m) &= \ln |\tilde{\Sigma}_u(m)| + \frac{\ln T}{T} (\# \text{ de parâmetros}) \\ &= \ln |\tilde{\Sigma}_u(m)| + \frac{\ln T}{T} m K^2\end{aligned}\tag{58}$$

Já o segundo critério foi derivado por Schwarz e é baseado em critério bayesianos:

### Critério de Schwarz

$$\begin{aligned}\text{HQ}(m) &= \ln |\tilde{\Sigma}_u(m)| + \frac{\ln T}{T} (\# \text{ de parâmetros}) \\ &= \ln |\tilde{\Sigma}_u(m)| + \frac{\ln T}{T} m K^2\end{aligned}\tag{58}$$

Novamente o estimador  $\hat{p}(\text{SC})$  é a ordem que minimiza  $\text{SC}(m)$  para  $m = 0, 1, \dots, M$ .

Já o segundo critério foi derivado por Schwarz e é baseado em critério bayesianos:

### Critério de Schwarz

$$\begin{aligned}\text{HQ}(m) &= \ln |\tilde{\Sigma}_u(m)| + \frac{\ln T}{T} (\# \text{ de parâmetros}) \\ &= \ln |\tilde{\Sigma}_u(m)| + \frac{\ln T}{T} m K^2\end{aligned}\tag{58}$$






Novamente o estimador  $\hat{p}(\text{SC})$  é a ordem que minimiza  $\text{SC}(m)$  para  $m = 0, 1, \dots, M$ .

**Pode-se mostrar que SC é fortemente consistente para qualquer  $K$ .**



Uma discussão mais detalhada, com todos os detalhes técnicos, sobre os critérios de seleção aqui apresentados pode ser encontrada em [Lütkepohl 2005, capítulo 4, seção 4.3].

# Referências

-  Lütkepohl, H., *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin, 2005.
-  Hamilton, J., *Time Series Analysis*, Cambridge University Press, Cambridge, 1994.
-  Moretin, P.A., *Econometria Financeira*, Edgard Blucher, São Paulo, 2011.
-  Hayashi, F., *Econometrics*, Princeton University Press, Princeton, 2000.
-  Greene, W.H., *Econometric Analysis*, Prentice-Hall, Up Saddle River, 2004.

# Homework 1 (Quer dizer que tem mais!?)

```
function [v, A, B, Yhat, critFPE, ...  
         critAIC, critHQ, critSC] = estimVAR(Y, p, h)  
% Inputs:  
%   Y           Series das variaveis  
%   p           Ordem do VAR  
%   h           Horizonte de forecast  
%  
% Outputs:  
%   v           Estimadores dos interceptos do VAR(p)  
%   A           A(i).  
%   B           [v, A(1),...,A(p)]  
%   Yhat        Forecast do processo para T+1,...,T+h  
%   critFPE     Forecast precision error  
%   critAIC     Akaike's Information Criterion (AIC)  
%   critHQ      Hannan-Quinn criterion (HQ)  
%   critSC      Schwarz's criterion (SC)
```

Seu codigo aqui ....

end

## Homework 2 (Sim, tem mais!)

### Causalidade de Granger

1. Estude o assunto.
2. Entenda o teste.
3. Implemente.