

Modelos VAR e VEC*

Pedro Valls[†]

EESP-FGV

22 de julho de 2024

Sumário

1	Introdução	4
2	Modelos MA Vetorial - VMA	8
2.1	Especificação do VMA	8
2.2	Função de Autocovariância para VMA(1)	10
3	Modelos AR Vetorial - VAR	10
3.1	Especificação do VAR	10
3.2	Função de Autocovariância para VAR(1)	12
3.3	Representação de Wold para Model VAR Estacionário	14
15	section.4	
4.1	Primeiro Caso: autovalores distintos e menor do que um	16
4.2	Segundo Caso: autovalores distintos sendo um menor do que um e o outro igual a um	16
4.3	Terceiro Caso: autovalores ambos iguais a um	18
4.4	Exercícios	21
5	Condições para que um VAR(1) Trivariado se reduza a AR univariados	22
5.1	Primeiro Caso: um autovalor igual a 1 e os outros dois distintos e menores do que um	22
5.2	Segundo Caso: dois autovalores iguais a 1 e o outro menor do que um . .	24
5.3	Exercícios	25
6	Especificação do Modelo VAR Geral	26

*© 2024 - Pedro Valls

[†]CEQEF e EESP-FGV, Rua Dr. Plínio Barreto 365 sala 1319, 01313-020, São Paulo, S.P. Brasil.
E-mail:pedro.valls@fgv.br.

7	Estimação de Modelos VAR	27
7.1	Estimação do $VAR(1)$	28
7.2	Distribuição dos Estimadores e Teste de Hipóteses	28
7.3	Estimação por SURE	29
7.4	Determinando a ordem do $VAR(p)$	30
7.5	Causalidade de Granger	31
7.6	Exemplo com os dados gerados segundo um $VAR(1)$ estacionário	31
8	Previsão para um $VAR(1)$	37
8.1	Exemplo com horizonte de previsão igual a 10	38
9	Função de Resposta ao Impulso	39
10	Decomposição da variância do erro de previsão	43
11	VAR Estrutural com restrições de longo prazo	45
12	Modelo de Correção de Erros Vetorial (VEC)	45
12.1	Introdução	45
12.2	Regressão Espúria	47
12.3	Implicações Estatísticas da Regressão Espúria	48
12.4	Cointegração no contexto univariado	48
12.4.1	Normalização	49
12.4.2	Exemplos de Cointegração em Economia e Finanças num contexto univariado	49
12.4.3	Cointegração e Tendências Comuns	50
12.4.4	Alguns Sistemas Cointegrados Simulados - sistema bivariado	51
12.4.5	Alguns Sistemas Cointegrados Simulados - sistema trivariado	52
12.4.6	Cointegração e Modelo de Correção do Equilíbrio - EqCM	54
12.4.7	Testes de Cointegração no contexto univariado	54
12.4.8	Testes de Cointegração baseado nos resíduos	54
12.4.9	Estimando o Modelo de Correção do Equilíbrio usando M.Q.O.	57
12.5	Os modelos VAR e Cointegração	58
12.5.1	Introdução	58
12.5.2	VAR e cointegração	58
12.5.3	Não unicidade da fatorização da matriz de longo-prazo, Π	60
12.5.4	Alguns exemplos	60
12.6	Teste de Johansen	60
12.7	Um Exemplo de Determinação da Dimensão do Espaço de Cointegração - Dados Gerados pelo programa abaixo	62
12.8	Exercício	66
12.9	Um Exemplo de Determinação da Dimensão do Espaço de Cointegração - Relação entre mercado spot e futuro para o S&P500 dados intradiária de Maio de 1996.	67
12.10	Restrições na matriz de carga e nos vetores de cointegração	70
12.11	Teste de Exogeneidade Fraca	72

13 Apêndice - Operador Vec e Produto de Kronecker	74
14 Apêndice - Demonstração de (44) e (45)	75

Lista de Figuras

1	Vetor MA(1) com 200 observações	9
2	MA(1) Univariados com 200 observações	9
3	Vetor AR(1) com 200 observações	11
4	Correlações Cruzadas para Vetor AR(1) com 200 observações	14
5	Simulação do Vetor AR(1) com 200 observações $\lambda_1 = 1.0$ e $\lambda_2 = 0.3$	17
6	Relação de cointegração $z_{2,t} = y_{1,t} - 0.5y_{2,t}$	18
7	VAR com séries $I(2)$	20
8	Relação de cointegração entre y_1 e y_2	20
9	$VAR(1)$ trivariado com $\lambda_1 = 1.0$, $\lambda_2 = 0.5$ e $\lambda_3 = 0.3$	23
10	Duas relações de Cointegração	23
11	$VAR(1)$ trivariado com $\lambda_1 = \lambda_2 = 1$ e $\lambda_3 = 0.5$	25
12	Resíduos do $VAR(1)$ sem constante e tendência	34
13	Autovalores da Matriz de Longo Prazo	36
14	Previsão 10 passos à frente para o $VAR(1)$ estacionário	38
15	IRF ortogonal para y_1	41
16	IRF ortogonal para y_2	41
17	IRF cumulativa para y_1	42
18	IRF cumulativa para y_2	42
19	IRF não ortogonal e cumulativa para y_1	43
20	IRF não ortogonal e cumulativa para y_2	43
21	Log do SP500 Spot e Fut	46
22	Diagrama de dispersão - Log do SP500 Spot e Fut	46
23	Diagrama de dispersão de $y_{1,t}$ e $y_{2,t}$	47
24	Diagrama de dispersão de $\Delta(y_{1,t})$ e $\Delta(y_{2,t})$	47
25	EqCM para Spot e Futuro de SP500	49
26	As duas séries do sistema (84 - 88)	52
27	EqCM para sistema (84 - 88)	52
28	As três séries do sistema (89-92)	53
29	EqCM para sistema (89-92)	53
30	EqCM com $\beta = (1, -1)$	55
31	Acf e Pacf para EqCM com $\beta = (1, -1)$	55
32	EqCM com β estimado por M.Q.O.	56
33	ACF e PACF para EqCM com β estimado por M.Q.O.	57
34	Relação de Cointegração $\beta'y_{t-1}$	66
35	Log do Spot e do Futuro do S&P500 intradiário Maio de 1996	67
36	Relação de Cointegração $\beta'y_{t-1}$ para LSPOT e LFUT	70

Lista de Tabelas

1	Critérios de Informação para VAR	32
2	Estimação do $VAR(1)$	33

3	Estimação do $VAR(1)$ sem constante e tendência	34
4	Portmanteau Test for Autocorrelation	35
5	Normality test for Residuals	35
6	ARCH test for Residuals	35
7	$H_0 : y_1$ não causa no sentido de Granger y_2	36
8	$H_0 : y_1$ não causa no sentido de Granger y_2	36
9	Decomposição da variância para y_2	44
10	Decomposição da variância para y_1	44
11	ADF teste para EqCM com $\beta = (1, -1)$	56
12	EqCM como resíduo da regressão estática	56
13	ADF teste para EqCM com β estimado por M.Q.O.	57
14	EqCM para (93)	58
15	EqCM para (94)	58
16	Seleção da defasagem ótima	63
17	Dimensão do espaço de cointegração para os Casos	64
18	Critérios de Informação para todos os Casos	65
19	VEC para o melhor modelo	65
20	Bondade de Ajuste e Testes de Especificação para Tabela (19)	66
21	Dimensão do espaço de cointegração para os Casos	68
22	Critérios de Informação para todos os Casos	68
23	VEC para o melhor modelo LSPOT e LFUT	69
24	Bondade de Ajuste e Testes de Especificação para Tabela (23)	70
25	Restrictions on β :	71
26	Restrictions on α :	71

1 Introdução

Neste capítulo iremos generalizar os modelos univariados de séries temporais para o caso multivariado.

Vamos considerar n Séries Temporais, $\{y_{1,t}\}, \dots, \{y_{n,t}\}$, que serão agrupadas no vetor, $n \times 1$, de séries temporais, \mathbf{y}_t , onde o $i^{ésimo}$ componente é $\{y_{i,t}\}$, isto é para o instante t temos, $\mathbf{y}_t = (y_{1,t}, \dots, y_{n,t})$.

A análise de séries temporais multivariadas é usada quando desejamos modelar ou explicar a interação e movimentos conjuntos entre um grupo de variáveis de séries temporais:

- (i) Consumo e Renda;
- (ii) Preço de ações e dividendos;
- (iii) Preços a termo e à vista de taxas de câmbio; e
- (iv) Taxas de juros, crescimento monetário, renda, inflação.

Stock and Watson (2001) afirmam que macroeconometricistas realizam quatro atividades com séries temporais multivariadas:

1. Descrever e resumir dados macroeconômicos;

2. Fazer previsões macroeconômicas;
3. Quantificar o que sabemos, ou não, sobre a verdadeira estrutura da macroeconomia;
e
4. Aconselhar os formuladores de políticas macroeconômicas.

A série temporal multivariada \mathbf{y}_t é covariância estacionária e ergódica¹, se todos os componentes são estacionários e ergódicos.

Temos então:

(i)

$$\begin{aligned} E[\mathbf{y}_t] &= \boldsymbol{\mu} = (\mu_1, \dots, \mu_n)' \\ Var[\mathbf{y}_t] &= \boldsymbol{\Gamma}_0 = E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})'] = \end{aligned}$$

(ii)

$$\begin{pmatrix} var(y_{1,t}) & cov(y_{1,t}, y_{2,t}) & \cdots & cov(y_{1,t}, y_{n,t}) \\ cov(y_{2,t}, y_{1,t}) & var(y_{2,t}) & \cdots & cov(y_{2,t}, y_{n,t}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(y_{n,t}, y_{1,t}) & cov(y_{n,t}, y_{2,t}) & \cdots & var(y_{n,t}) \end{pmatrix}$$

A matrix de correlação de \mathbf{y}_t é uma matriz $(n \times n)$, tal que:

$$Corr[\mathbf{y}_t] = \mathbf{R}_0 = \mathbf{D}^{-1} \boldsymbol{\Gamma}_0 \mathbf{D}^{-1}$$

onde \mathbf{D} é uma matriz $(n \times n)$ diagonal com j^{esimo} elemento na diagonal dado por $(\gamma_{jj}^0)^{1/2} = var(Y_{jt})^{1/2}$

Os parâmetros $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}_0$ and \mathbf{R}_0 são estimados através dos dados $(\mathbf{y}_1, \dots, \mathbf{y}_T)$ usando os momentos amostrais, isto é:

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \bar{\mathbf{y}} = \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t \xrightarrow{p} E[\mathbf{y}_t] = \boldsymbol{\mu} \\ \hat{\boldsymbol{\Gamma}}_0 &= \frac{1}{T} \sum_{t=1}^T (\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})' \xrightarrow{p} Var(\mathbf{y}_t) = \boldsymbol{\Gamma}_0 \\ \hat{\mathbf{R}}_0 &= \hat{\mathbf{D}}^{-1} \hat{\boldsymbol{\Gamma}}_0 \hat{\mathbf{D}}^{-1} \xrightarrow{p} Corr(\mathbf{Y}_t) = \mathbf{R}_0 \end{aligned}$$

onde $\hat{\mathbf{D}}$ é uma matriz $(n \times n)$ diagonal com os desvios padrões amostrais de y_{jt} ao longo da diagonal.

O Teorema Ergódico justifica a convergência dos momentos da amostra para seus correspondentes populacionais.

Numa série temporal multivariada \mathbf{y}_t , cada componente possui autocovariâncias e autocorrelações, mas também há covariância e correlações entre defasagens cruzadas entre todos os pares possíveis de componentes. As autocovariâncias e autocorrelações de $y_{j,t}$ para $j = 1, \dots, n$ são definidas como

¹Ergodicidade garante que autocovariância converge para zero quando as defasagens crescem. Também garante que a média e autocovariâncias amostrais convergem para os respectivos populacionais.

$$\begin{aligned}\gamma_{jj}^k &= \text{cov}(y_{j,t}, y_{j,t-k}) \\ \rho_{jj}^k &= \text{corr}(y_{j,t}, y_{j,t-k}) = \frac{\gamma_{jj}^k}{\gamma_{jj}^0}\end{aligned}$$

e são simétricos em relação a k : $\gamma_{jj}^k = \gamma_{jj}^{-k}$, $\rho_{jj}^k = \rho_{jj}^{-k}$

As covariâncias defasadas cruzadas e as correlações defasadas cruzadas entre $y_{i,t}$ e $y_{j,t}$ são definidas como

$$\begin{aligned}\gamma_{ij}^k &= \text{cov}(Y_{it}, Y_{jt-k}) \\ \rho_{ij}^k &= \text{corr}(Y_{it}, Y_{jt-k}) = \frac{\gamma_{ij}^k}{\sqrt{\gamma_{ii}^0 \gamma_{jj}^0}}\end{aligned}$$

e não são necessariamente simétricas em k .

Em geral

$$\gamma_{ij}^k = \text{cov}(Y_{it}, Y_{jt-k}) \neq \text{cov}(Y_{it}, Y_{jt+k}) = \gamma_{ij}^{-k}$$

Temos então as seguintes definições:

Definição 1 Se $\gamma_{ij}^k \neq 0$ para algum $k > 0$ então $y_{j,t}$ é dita ser *antecedente a* $y_{i,t}$.

Definição 2 Se $\gamma_{ij}^{-k} \neq 0$ para algum $k > 0$ então $y_{i,t}$ é dita ser *antecedente a* $y_{j,t}$.

Definição 3 É possível que $y_{i,t}$ anteceda $y_{j,t}$ e vice-versa. Neste caso, dizemos que existe *feedback* entre as duas séries.

Todas as covariâncias e correlações cruzadas de defasagem k são resumidas nas matrizes de covariância cruzada de defasagem k ($n \times n$) e nas matrizes de correlação cruzada de defasagem k .

$$\begin{aligned}\Gamma_k &= E[(\mathbf{y}_t - \mu)(\mathbf{y}_{t-k} - \mu)'] = \\ &\begin{pmatrix} \text{cov}(y_{1,t}, y_{1,t-k}) & \text{cov}(y_{1,t}, y_{2,t-k}) & \cdots & \text{cov}(y_{1,t}, y_{n,t-k}) \\ \text{cov}(y_{2,t}, y_{1,t-k}) & \text{cov}(y_{2,t}, y_{2,t-k}) & \cdots & \text{cov}(y_{2,t}, y_{n,t-k}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(y_{n,t}, y_{1,t-k}) & \text{cov}(y_{n,t}, y_{2,t-k}) & \cdots & \text{cov}(y_{n,t}, y_{n,t-k}) \end{pmatrix} \\ \mathbf{R}_k &= \mathbf{D}^{-1} \Gamma_k \mathbf{D}^{-1}\end{aligned}$$

As matrizes Γ_k e \mathbf{R}_k não são simétricas em k mas é fácil mostrar que $\Gamma_{-k} = \Gamma_k'$ e $\mathbf{R}_{-k} = \mathbf{R}_k'$

As matrizes Γ_k e \mathbf{R}_k são estimadas através dos dados $(\mathbf{y}_1, \dots, \mathbf{y}_T)$ usando:

$$\begin{aligned}\hat{\Gamma}_k &= \frac{1}{T} \sum_{t=1}^T (\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_{t-k} - \bar{\mathbf{y}})' \\ \hat{\mathbf{R}}_k &= \hat{\mathbf{D}}^{-1} \hat{\Gamma}_k \hat{\mathbf{D}}^{-1}\end{aligned}$$

Qualquer série temporal multivariada ($n \times 1$) covariância estacionária \mathbf{y}_t tem uma **Representação de Wold** ou **representação de processo linear** da seguinte forma:

$$\begin{aligned}\mathbf{y}_t &= \mu + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \cdots \\ &= \mu + \sum_{k=0}^{\infty} \Psi_k \varepsilon_{t-k} \\ \Psi_0 &= \mathbf{I}_n \\ \varepsilon_t &\sim NI(\mathbf{0}, \Sigma)\end{aligned}$$

onde Ψ_k é a matriz ($n \times n$) cujo elemento (ij) é ψ_{ij}^k .

Na notação do operador defasagem, a representação de Wold pode ser escrita da seguinte forma:

$$\begin{aligned}\mathbf{y}_t &= \mu + \Psi(L) \varepsilon_t \\ \Psi(L) &= \sum_{k=0}^{\infty} \Psi_k L^k \\ \sum_{k=0}^{\infty} |\Psi_k| &< \infty\end{aligned}$$

onde os primeiros dois momentos de \mathbf{y}_t são dados por:

$$\begin{aligned}E[\mathbf{y}_t] &= \mu \\ Var[\mathbf{y}_t] &= \sum_{k=0}^{\infty} \Psi_k \Sigma \Psi_k'\end{aligned}$$

Seja \mathbf{y}_t uma série temporal multivariada ($n \times 1$) estacionária e ergódica com $E[\mathbf{y}_t] = \mu$.

Então o Teorema de Limite Central para processos estacionários e ergódicos implica em:

$$\sqrt{T}(\bar{\mathbf{y}} - \mu) \xrightarrow{d} N(\mathbf{0}, \sum_{j=-\infty}^{\infty} \Gamma_j)$$

ou

$$\bar{\mathbf{y}} \overset{A}{\sim} N(\mu, \frac{1}{T} \sum_{j=-\infty}^{\infty} \Gamma_j)$$

Portanto, a **variância de longo-prazo** (*long-run variance*) de \mathbf{y}_t é T vezes a variância assintótica de $\bar{\mathbf{y}}$, isto é:

$$LRV(\mathbf{y}_t) = T \cdot Avar(\bar{\mathbf{y}}) = \sum_{j=-\infty}^{\infty} \Gamma_j$$

Como $\Gamma_{-j} = \Gamma_j'$, $LRV(\mathbf{y}_t) = \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j + \Gamma_j')$

E usando a representação de Wold de \mathbf{y}_t podemos mostrar que

$$LRV(\mathbf{y}_t) = \Psi(1)\Sigma\Psi(1)'$$

onde

$$\Psi(1) = \sum_{k=0}^{\infty} \Psi_k$$

Um estimador consistente de $LRV(\mathbf{y}_t)$ pode ser calculado usando métodos não paramétricos. Um estimador popular é o estimador de autocovariância ponderada de Newey-West.(see Newey and West (1987)), isto é:

$$\widehat{LRV}_{NW}(\mathbf{y}_t) = \widehat{\Gamma}_0 + \sum_{j=1}^{M_T} \omega_{j,T} \cdot (\widehat{\Gamma}_j + \widehat{\Gamma}_j')$$

onde $\omega_{j,T}$ são pesos que somam um, M_T é um parâmetro de truncamento da defasagem que satisfaz $M_T = O(T^{1/3})$.

Usualmente os pesos de Bartlett são usados, isto é:

$$\omega_{j,T}^{Bartlett} = 1 - \frac{j}{M_T + 1}$$

Inicialmente iremos apresentar os modelo de séries temporais multivariados estacionários, a saber Média Móveis Vetoriais e Autorregressivos Vetoriais e a seguir vamos permitir não estacionaridades nestes modelos. Neste caso é necessário introduzir o conceito de co-integração, uma vez que podemos ter séries que são não estacionárias mas pode existir uma combinação delas que as torna estacionárias. Neste caso, os modelos Autorregressivos Vetoriais se transformam em Modelo de Correção de Erros Vetoriais (VEC) que contemplam variáveis em nível e em primeira diferença. Nesta classe de modelos é possível definir um teste para exogeneidade fraca impondo restrições na relação de longo prazo entre as variáveis.

2 Modelos MA Vetorial - VMA

2.1 Especificação do VMA

Nos modelos de séries temporais multivariadas podemos ter representações semelhantes ao caso univariado, a saber, representações AR, MA e ARMA.

Primeiramente vamos apresentar a representação MA Vetorial que é apresentada abaixo

$$\mathbf{y}_t = \sum_{j=0}^q \Theta_j \varepsilon_{t-j} \quad (1)$$

onde \mathbf{y}_t é um vetor $n \times 1$, Θ_j são matrizes $n \times n$ com $\Theta_0 = I$ e $\varepsilon_t \sim NI_n[0, \Omega]$. Observe que estamos assumindo que o vetor de observações tem média zero, mas poderíamos ter um vetor de médias que denotaremos pelo vetor \mathbf{m} . Observe também que os choques são ruídos brancos mas é permitido que tenham correlação contemporânea.

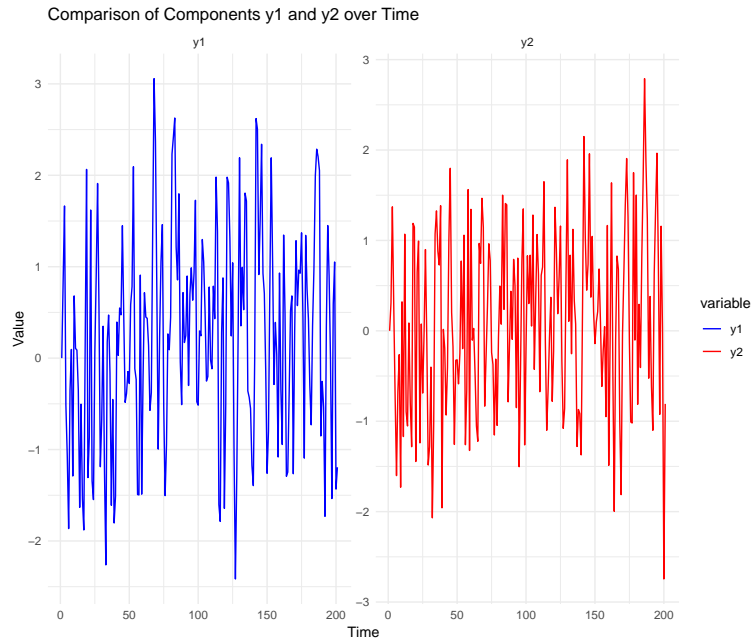
O programa

Simula um VMA(1) no R

gera uma vetor MA(1) onde $\Theta_1 = \begin{bmatrix} 0.55 & -0.1 \\ 0.1 & 0.3 \end{bmatrix}$, $\Omega = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.2 \end{bmatrix}$ e $\mathbf{m} = \mathbf{0}$

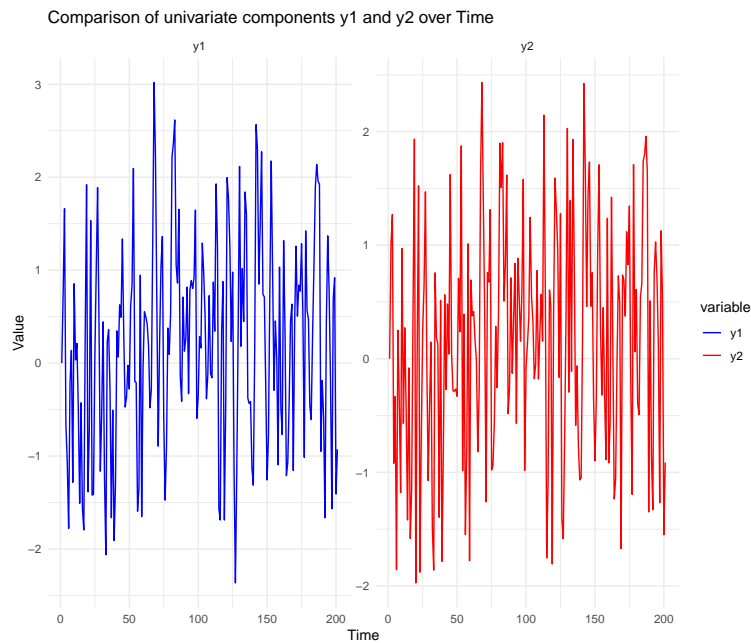
Os gráficos abaixo apresenta as duas séries geradas segundo este Processo Gerador de Dados (P.G.D.).

Figura 1: Vetor MA(1) com 200 observações



Modelos MA univariados foram gerados usando os mesmos erros do modelo multivariado. As seguintes séries foram obtida:

Figura 2: MA(1) Univariados com 200 observações



Observe que as séries univariadas geradas tem um padrão parecido com as multivariadas. Somente pelos gráficos fica difícil identificar a relação de interligação entre as duas séries.

2.2 Função de Autocovariância para VMA(1)

Considere o VMA(1) dado por:

$$\mathbf{y}_t = \varepsilon_t + \Theta_1 \varepsilon_{t-1} \quad (2)$$

podemos então definir a variância de \mathbf{y}_t , que será denotada por $\Gamma(0)$, da seguinte forma:

$$\begin{aligned} \Gamma(0) &= E(\mathbf{y}_t \mathbf{y}_t') = E(\varepsilon_t + \Theta_1 \varepsilon_{t-1})(\varepsilon_t + \Theta_1 \varepsilon_{t-1})' \\ \Gamma(0) &= \Omega + \Theta_1 \Omega \Theta_1' \end{aligned} \quad (3)$$

e a função de autocovariância será definida por:

$$\Gamma(\tau) = E(\mathbf{y}_t \mathbf{y}_{t-\tau}') = \begin{cases} \Theta_1 \Omega & \text{para } \tau = 1 \\ \Omega \Theta_1' & \text{para } \tau = -1 \\ 0 & \text{para } |\tau| \geq 2 \end{cases} \quad (4)$$

3 Modelos AR Vetorial - VAR

3.1 Especificação do VAR

O **Modelo Vetor Autorregressivo (VAR)** é um dos modelos mais bem-sucedidos, flexíveis e fáceis de usar para análise de séries temporais multivariadas.

Este modelo foi introduzido por Cris Sims no artigo "Macroeconomics and Reality" (veja Sims (1980)).

É uma extensão natural do modelo autorregressivo univariado para séries temporais multivariadas dinâmicas.

Tem se mostrado especialmente útil para descrever o comportamento dinâmico de séries temporais econômicas e financeiras e para previsão.

Frequentemente, fornece previsões superiores às dos modelos de séries temporais univariadas e a modelos de equações simultâneas elaborados baseados em teoria.

Usado para inferência estrutural e análise de políticas. Na análise estrutural, certas suposições sobre a estrutura causal dos dados em investigação são impostas, e os impactos causais resultantes de choques inesperados em variáveis especificadas sobre as variáveis no modelo são geralmente resumidos. Esses impactos causais são normalmente resumidos com **funções de resposta ao impulso** e **variância do erro de previsão**.

Uma outra representação possível para o Vetor Autorregressivo, cujo processo gerador dos dados pode ser definido por:

$$\Phi_0 \mathbf{y}_t = \mathbf{m} + \sum_{j=0}^p \Phi_j \mathbf{y}_{t-j} + \varepsilon_t \quad (5)$$

onde \mathbf{y}_t é um vetor $n \times 1$, Φ_j são matrizes $n \times n$ com $\Phi_0 = I$ e $\varepsilon_t \sim NI_N[0, \Omega]$. O caso em que Φ_0 não é identidade corresponde ao que é chamado na literatura de VAR Estrutural que não será considerado aqui.

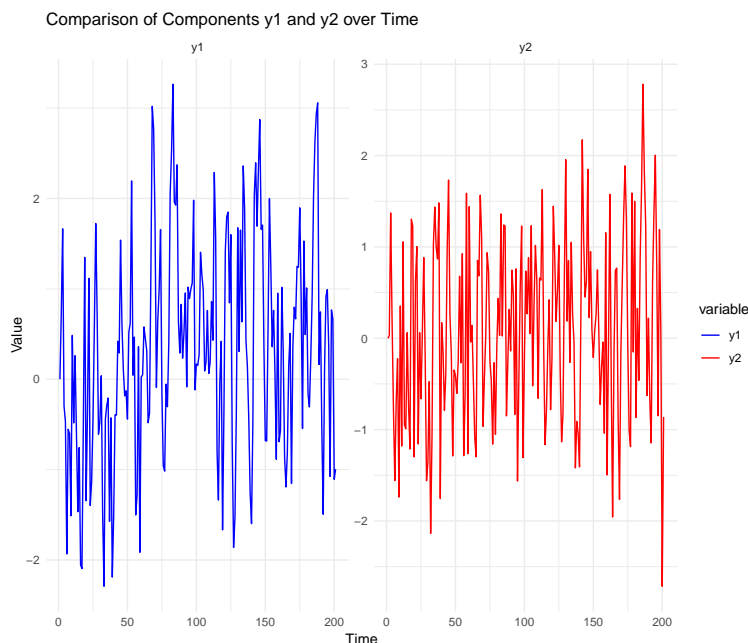
O programa abaixo

Simula um VAR(1) no R

simula um vetor AR(1) onde $\Phi_1 = \begin{bmatrix} 0.55 & -0.1 \\ 0.1 & 0.3 \end{bmatrix}$, $\Omega = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.2 \end{bmatrix}$ e $\mathbf{m} = \mathbf{0}$

Os gráficos abaixo apresentam as séries geradas segundo este P.G.D.

Figura 3: Vetor AR(1) com 200 observações



Temos o seguinte exemplo para um VAR(1) bivariado

Exemplo 4 *VAR(1) bivariado*

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

ou

$$\begin{aligned} y_{1,t} &= m_1 + \phi_{11}y_{1,t-1} + \phi_{12}y_{2,t-1} + \varepsilon_{1,t} \\ y_{2,t} &= m_2 + \phi_{21}y_{1,t-1} + \phi_{22}y_{2,t-1} + \varepsilon_{2,t} \end{aligned}$$

onde $cov(\varepsilon_{1,t}, \varepsilon_{2,t}) = \sigma_{12}$ para $t = s$ e igual a 0 caso contrário.

Cada equação possui os mesmos regressores - constante e valores defasados de $y_{1,t}$ e $y_{2,t}$.

A endogeneidade é evitada usando valores defasados de $y_{1,t}$ e $y_{2,t}$.

O modelo $VAR(p)$ é apenas um modelo de regressão aparentemente não relacionada (SURE, do inglês "seemingly unrelated regression") com variáveis defasadas e termos determinísticos como regressores comuns.

Na notação do operador de defasagem, o $VAR(p)$ é escrito como:

$$\begin{aligned}\Phi(L)\mathbf{y}_t &= \mathbf{m} + \varepsilon_t \\ \Phi(L) &= \mathbf{I}_n - \Phi_1 L - \dots - \Phi_p L^p\end{aligned}$$

O $VAR(p)$ é estável (estacionário) se as raízes de:

$$\det(\mathbf{I}_n - \Phi_1 z - \dots - \Phi_p z^p) = 0$$

estiverem fora do círculo unitário (tiverem módulo maior que um), ou, equivalentemente, se os autovalores da matriz companheira (em inglês *companion matrix*)

$$\mathbf{F} = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ & \mathbf{I}_{np-1} & & & \mathbf{0} \end{bmatrix}$$

tem módulos menores do que um.

Um processo $VAR(p)$ estável é estacionário e ergódico, com médias, variâncias e autocovariâncias invariantes no tempo.

No exemplo (4) a condição de estacionaridade é $\det(\mathbf{I}_n - \Phi_1 z) = 0$ torna-se

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

A condição de estabilidade envolve termos cruzados ϕ_{12} e ϕ_{21} .

Se $\phi_{12} = \phi_{21} = 0$ (VAR diagonal), então a condição de estabilidade bivariada se reduz às condições de estabilidade univariadas para cada equação.

3.2 Função de Autocovariância para VAR(1)

Considere um VAR(1) dado por

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \varepsilon_t$$

podemos definir a função de autocovariância por:

$$E(\mathbf{y}_t \mathbf{y}'_{t-\tau}) = \Phi_1 E(\mathbf{y}_{t-1} \mathbf{y}'_{t-\tau}) + E(\varepsilon_t \mathbf{y}'_{t-\tau}) \quad (6)$$

Agora observe que para $\tau \geq 1$ temos que $E(\varepsilon_t \mathbf{y}'_{t-\tau}) = 0$ e para $\tau = 0$ temos:

$$\begin{aligned}E(\varepsilon_t \mathbf{y}'_t) &= E[\varepsilon_t (\Phi_1 \mathbf{y}_{t-1} + \varepsilon_t)'] \\ &= E[\varepsilon_t \varepsilon'_t] \\ &= \Omega\end{aligned}$$

então para $\tau \geq 1$, (6) implica em:

$$\Gamma(\tau) = \Phi_1 \Gamma(\tau - 1) \quad (7)$$

e para $\tau = 0$, (6) implica em

$$\Gamma(0) = \Phi_1 \Gamma(-1) + \Omega \quad (8)$$

Observe que

$$\Gamma(-1) = \Gamma'(1) \quad (9)$$

e fazendo $\tau = 1$, (7) temos

$$\Gamma(1) = \Phi_1 \Gamma(0) \quad (10)$$

e substituindo (9) e (10) em (8) temos:

$$\Gamma(0) = \Phi_1 \Gamma(0) \Phi_1' + \Omega \quad (11)$$

cujas soluções são dadas por²:

$$vec(\Gamma(0)) = (\mathbf{I} - \Phi_1 \otimes \Phi_1')^{-1} vec(\Omega) \quad (12)$$

Observe que a matriz $\Gamma(\tau)$ representa a covariância cruzada entre os componentes de \mathbf{y}_t . Por exemplo se $\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}$ temos que

$$\Gamma(0) = \begin{bmatrix} Var(y_1) & Cov(y_1 : y_2) \\ Cov(y_1 : y_2) & Var(y_2) \end{bmatrix} \quad (13)$$

que é a matriz de variância e covariância de \mathbf{y}_t e quando $\tau = 1$ temos

$$\Gamma(1) = \begin{bmatrix} E(y_{1,t}y_{1,t-1}) & E(y_{1,t}y_{2,t-1}) \\ E(y_{2,t}y_{1,t-1}) & E(y_{2,t}y_{2,t-1}) \end{bmatrix} \quad (14)$$

onde na diagonal temos a primeira autocovariância para cada um dos componentes de \mathbf{y}_t e fora da diagonal principal a primeira covariância cruzada entre as componentes de \mathbf{y}_t .

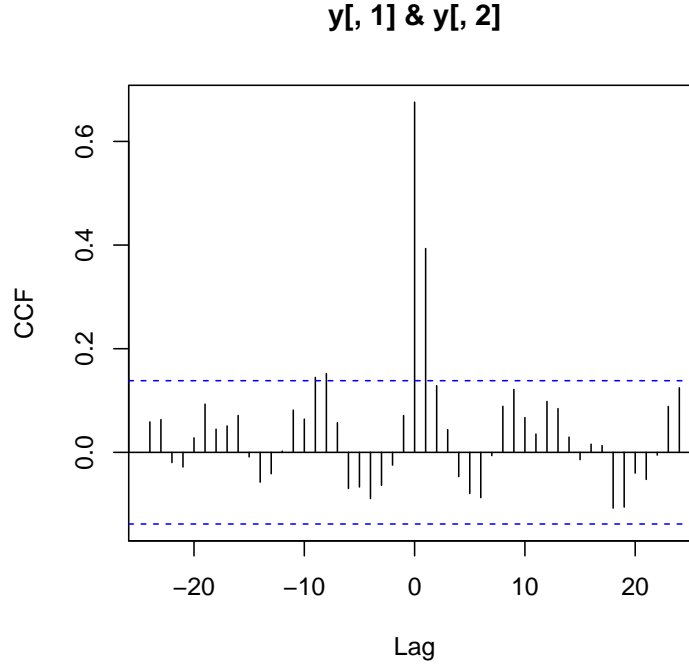
Para definir a correlação cruzada basta dividir $\Gamma(\tau)$ pelas respectivas variâncias, por exemplo em (14) teremos

$$R(1) = \begin{bmatrix} \frac{E(y_{1,t}y_{1,t-1})}{Var(y_{1,t})} & \frac{E(y_{1,t}y_{2,t-1})}{\sqrt{Var(y_{1,t})}\sqrt{Var(y_{2,t})}} \\ \frac{E(y_{2,t}y_{1,t-1})}{\sqrt{Var(y_{1,t})}\sqrt{Var(y_{2,t})}} & \frac{E(y_{2,t}y_{2,t-1})}{Var(y_{2,t})} \end{bmatrix}$$

Voltando aos dados gerados e calculando estas correlações para $\tau = 1, \dots, 24$, temos

²Veja apêndice (13) para definição e propriedades do operador \otimes e vec

Figura 4: Correlações Cruzadas para Vetor AR(1) com 200 observações



que tem um decaimento exponencial tanto para defasagens quanto para antecedenças, caracterizando um processo autorregressivo.

3.3 Representação de Wold para Model VAR Estacionário

Voltando para o $VAR(p)$, se \mathbf{Y}_t é estacionário em covariância, então a média incondicional é dada por:

$$\mu = (\mathbf{I}_n - \Phi_1 - \dots - \Phi_p)^{-1} \mathbf{m}$$

A forma ajustada para a média do $VAR(p)$ é então

$$(\mathbf{y}_t - \mu) = \Phi_1(\mathbf{y}_{t-1} - \mu) + \Phi_2(\mathbf{y}_{t-2} - \mu) + \dots + \Phi_p(\mathbf{y}_{t-p} - \mu) + \varepsilon_t$$

O modelo básico $VAR(p)$ pode ser muito restritivo para representar adequadamente as principais características dos dados. A forma geral do modelo $VAR(p)$ com termos determinísticos e variáveis exógenas é dada por:

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \dots + \Phi_p \mathbf{y}_{t-p} + \Psi \mathbf{D}_t + \mathbf{G} \mathbf{X}_t + \varepsilon_t$$

$$\mathbf{D}_t = \text{termos determinísticos}$$

$$\mathbf{X}_t = \text{variáveis exógenas} (E[\mathbf{X}_t' \varepsilon_t] = 0)$$

Considere o modelo $VAR(p)$ estacionário

$$\Phi(L) \mathbf{y}_t = \mathbf{m} + \varepsilon_t$$

$$\Phi(L) = \mathbf{I}_n - \Phi_1 L - \dots - \Phi_p L^p$$

Como \mathbf{y}_t é estacionário, $\Phi(L)^{-1}$ existe de modo que:

$$\begin{aligned}\mathbf{y}_t &= \Phi(L)^{-1}\mathbf{m} + \Phi(L)^{-1}\varepsilon_t \\ &= \mu + \sum_{k=0}^{\infty} \Psi_k \varepsilon_{t-k} \\ \Psi_0 &= \mathbf{I}_n \\ \lim_{k \rightarrow \infty} \Psi_k &= 0\end{aligned}$$

Note que

$$\Theta(L) = \sum_{k=0}^{\infty} \Psi_k L^k$$

que é uma representação de Média Móvel Vetorial de ordem infinita ($VMA(\infty)$).

Os coeficientes da representação de Wold, Ψ_k , podem ser determinados a partir dos coeficientes do VAR , Π_k , resolvendo:

$$\Psi(L)\Phi(L) = \mathbf{I}_n$$

que implica em:

$$\begin{aligned}\Psi_1 &= \Phi_1 \\ \Psi_2 &= \Phi_1 \Phi_1 + \Phi_2 = \Phi_1 \Psi_1 + \Phi_2 \\ &\vdots \\ \Psi_s &= \Phi_1 \Psi_{s-1} + \dots + \Phi_p \Psi_{s-p}\end{aligned}$$

Como $\Phi(L)^{-1} = \Psi(L)$, a variância de longo prazo para \mathbf{y}_t tem a seguinte forma:

$$\begin{aligned}LRV(\mathbf{y}_t) &= \Psi(1)\Sigma\Psi(1)' \\ &= \Phi(1)^{-1}\Sigma\Phi(1)^{-1'} \\ &= (\mathbf{I}_n - \Phi_1 - \dots - \Phi_p)^{-1}\Sigma(\mathbf{I}_n - \Phi_1 - \dots - \Phi_p)^{-1'}\end{aligned}$$

4 Condições para que um VAR(1) bivariado se reduza a AR univariados³

Considere uma VAR(1) para o caso de duas variáveis, isto é:

$$\mathbf{y}_t = \mathbf{m} + \Phi\mathbf{y}_{t-1} + \varepsilon_t \quad (15)$$

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \quad (16)$$

³Esta sessão é baseada em Johnston and DiNardo (1996).

com $\varepsilon_t \sim NI(0, \Omega)$.

Se a matriz Φ puder ser decomposta em $P\Lambda P^{-1}$ onde P é a matriz de autovetores de Λ e esta é uma matriz diagonal dos autovalores então podemos definir $\mathbf{z}_t = P^{-1}\mathbf{y}_t$ e (15) poderá ser escrito da seguinte forma

$$\mathbf{z}_t = \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} m_1^* \\ m_2^* \end{bmatrix} + \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t}^* \\ \varepsilon_{2,t}^* \end{bmatrix} \quad (17)$$

onde esta representação equivale a AR univariados para as novas variáveis \mathbf{z}_t .

Observe que esta representação só é possível se a matriz Φ for diagonalizável. Podemos ter três casos dependendo se os autovalores são distintos e menores do que um, distintos sendo um menor do que um e o outro igual a um e o terceiro caso em que ambos os autovalores são iguais a um. Vamos apresentar estes três casos com as suas implicações em termos de estacionaridade para as variáveis \mathbf{y}_t .

4.1 Primeiro Caso: autovalores distintos e menor do que um

Suponha que $|\lambda_1| < 1$ e $|\lambda_2| < 1$ então por (17) temos que tanto $z_{1,t}$ quanto $z_{2,t}$ são AR(1) estacionários e portanto com $z_{1,t}$ e $z_{2,t}$ são combinações lineares de $y_{1,t}$ e $y_{2,t}$ temos que estes últimos são estacionários.

Neste caso inferência usual é válida e podemos definir o equilíbrio estático que será dado por:

$$\begin{aligned} (I - \Phi) \bar{\mathbf{y}} &= \mathbf{m} \\ \Pi \bar{\mathbf{y}} &= \mathbf{m} \\ \bar{\mathbf{y}} &= \Pi^{-1} \mathbf{m} \end{aligned} \quad (18)$$

Observe que $\Pi = (I - \Phi)$ tem autovalores iguais a $1 - \lambda_i$ para $i = 1, 2$ que são diferentes de zero uma vez que os autovalores $|\lambda_i| < 1$. Observe que (18) representa o equilíbrio estático e os desvios deste equilíbrio são transitórios e tendem a zero.

4.2 Segundo Caso: autovalores distintos sendo um menor do que um e o outro igual a um

Neste caso temos que $\lambda_1 = 1$ e $|\lambda_2| < 1$ e (17) será dada por:

$$\mathbf{z}_t = \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} m_1^* \\ m_2^* \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t}^* \\ \varepsilon_{2,t}^* \end{bmatrix}$$

então, $z_{1,t}$ será um passeio aleatório com *drift*, uma vez que tem uma constante diferente de zero e $z_{2,t}$ será um AR(1) estacionário. Deste modo temos que $y_{1,t}$ e $y_{2,t}$ são $I(1)$ mas existe uma combinação linear deles que gera um processo estacionário, isto é dizemos que $y_{1,t}$ e $y_{2,t}$ co-integram.

A representação VAR dada por (16) pode ser escrita da seguinte forma:

$$\Delta \mathbf{y}_t = \mathbf{m} - \Pi \mathbf{y}_{t-1} + \varepsilon_t \quad (19)$$

onde $\Pi = (I - \Phi)$ tem autovalores iguais a 0 e $1 - \lambda_2$. Como um dos autovalores de Π é zero esta matriz tem posto reduzido e pode ser escrita como o produto de duas matrizes

2×1 , i.e. $\Pi = \alpha\beta'$ onde o vetor β é chamado de vetor de co-integração e α é chamada de matriz de cargas e dá a importância da relação de co-integração em cada equação.

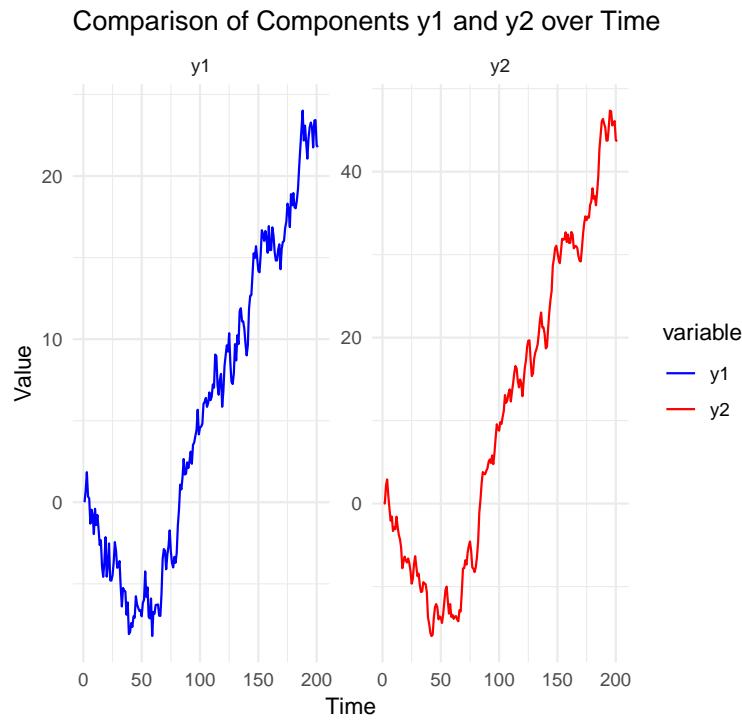
Como a matriz Π pode ser escrita como $P \text{diag}\{0, 1 - \lambda_2\} P^{-1}$ temos que (19) pode ser re-escrito da seguinte forma:

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} - \begin{bmatrix} p_{12}(1 - \lambda_2) \\ p_{22}(1 - \lambda_2) \end{bmatrix} z_{2,t-1} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \quad (20)$$

onde $z_{2,t-1} = p^{21}y_{1,t-1} + p^{22}y_{2,t-1}$ e mede os desvios do equilíbrio pois representa a relação de cointegração entre $y_{1,t}$ e $y_{2,t}$.

Foi usando o programa para gerar VAR(1) com matriz $\Phi = \begin{bmatrix} 0.8 & 0.1 \\ 1.0 & 0.5 \end{bmatrix}$ que tem autovalores iguais a 1 e 0.3. Abaixo são apresentados os gráficos de y_1 e y_2 .

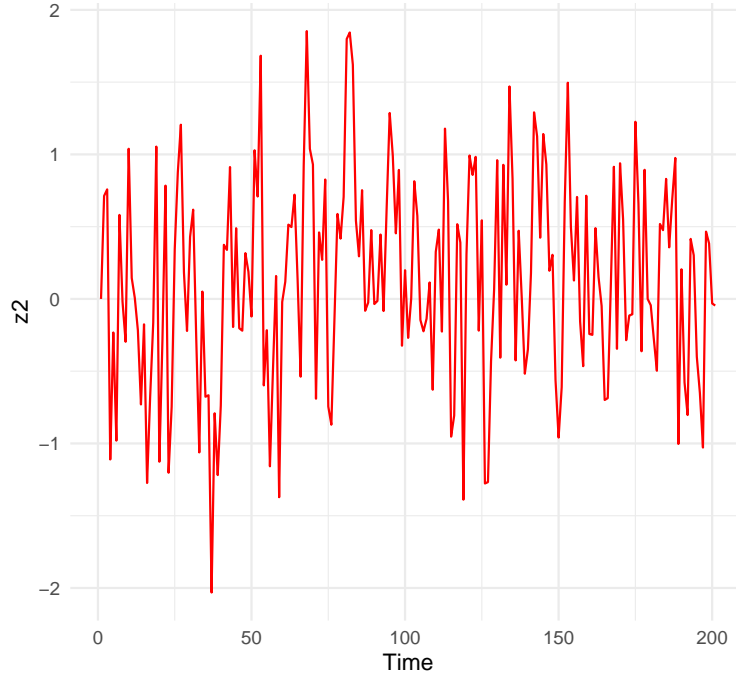
Figura 5: Simulação do Vetor AR(1) com 200 observações $\lambda_1 = 1.0$ e $\lambda_2 = 0.3$



Observe que ambas as variáveis tem tendência estocástica, isto é são $I(1)$, mas se usarmos $\beta = (1 : -0.5)^4$ teremos o vetor z_{2t} cujo gráfico é apresentado abaixo e não tem tendência estocástica

⁴A segunda linha da matriz inversa dos autovetores é dada por $(1.4569, -0.72843)$ e normalizando pela primeira componente temos $(1.0, -0.5)$.

Figura 6: Relação de cointegração $z_{2,t} = y_{1,t} - 0.5y_{2,t}$



4.3 Terceiro Caso: autovalores ambos iguais a um

Neste caso temos que $\lambda_1 = 1$ e $\lambda_2 = 1$ e os autovetores associados a estes autovalores não são linearmente independentes e portanto a matriz P é singular e Φ não pode ser diagonalizada da forma usual.

Este é um caso particular daquele em que os autovalores são iguais e a decomposição possível é a de Jordan que obtém uma matriz C tal que $C^{-1}\Phi C = J$ onde J é dada por:

$$J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \quad (21)$$

que é conhecida por matriz de Jordan para autovalores com multiplicidade dois.

Como $C^{-1}\Phi C = J$ temos então que $\Phi = CJC^{-1} = J$ e (15) pode ser escrito da seguinte forma:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{m} + CJC^{-1}\mathbf{y}_{t-1} + \varepsilon_t \\ C^{-1}\mathbf{y}_t &= C^{-1}\mathbf{m} + JC^{-1}\mathbf{y}_{t-1} + C^{-1}\varepsilon_t \\ \mathbf{z}_t &= \mathbf{m}^* + J\mathbf{z}_{t-1} + \varepsilon_t^* \end{aligned} \quad (22)$$

onde $\mathbf{z}_t = C^{-1}\mathbf{y}_t$, $\mathbf{m}^* = C^{-1}\mathbf{m}$ e $\varepsilon_t^* = C^{-1}\varepsilon_t$.

Observe que (22) é dado por:

$$\begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} m_1^* \\ m_2^* \end{bmatrix} + \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t}^* \\ \varepsilon_{2,t}^* \end{bmatrix} \quad (23)$$

e no caso em que os autovalores são iguais a um (23) pode ser re-escrito da seguinte forma:

$$z_{1,t} = m_1^* + z_{1,t-1} + z_{2,t-1} + \varepsilon_{1,t}^* \quad (24)$$

$$z_{2,t} = m_2^* + z_{2,t-1} + \varepsilon_{2,t}^* \quad (25)$$

Observe que (25) pode ser escrito da seguinte forma:

$$\begin{aligned} (1-L)z_{2,t} &= m_2^* + \varepsilon_{2,t}^* \\ z_{2,t} &= \frac{m_2^*}{(1-L)} + \frac{\varepsilon_{2,t}^*}{(1-L)} \end{aligned} \quad (26)$$

e (24) pode ser escrito da seguinte forma:

$$(1-L)z_{1,t} = m_1^* + z_{2,t-1} + \varepsilon_{1,t}^* \quad (27)$$

e substituindo (26) em (27) temos:

$$\begin{aligned} (1-L)z_{1,t} &= m_1^* + \frac{m_2^*}{(1-L)} + \frac{\varepsilon_{2,t-1}^*}{(1-L)} + \varepsilon_{1,t}^* \\ (1-L)^2 z_{1,t} &= m_1^*(1-L) + m_2^* + \varepsilon_{2,t-1}^* + (1-L)\varepsilon_{1,t}^* \\ (1-L)^2 z_{1,t} &= m_2^* + \varepsilon_{2,t-1}^* + \varepsilon_{1,t}^* - \varepsilon_{1,t-1}^* \end{aligned} \quad (28)$$

e por (25) temos que $z_{2,t}$ é $I(1)$ e por (28) que $z_{1,t}$ é $I(2)$ e portanto \mathbf{y}_t será $I(2)$. Observe que como $z_{2,t}$ é $I(1)$ temos que existe uma combinação linear de $y_{1,t}$ e $y_{2,t}$ que reduz a ordem de integração de dois para um, logo estas duas variáveis co-integração.

Podemos deste modo definir co-integração do seguinte modo.

Definição 5 *Os componentes do vetor \mathbf{y}_t são ditos co-integrados de ordem d , b e denotados por $\mathbf{y}_t \sim CI(d, b)$, se*

- (i) *todos os componentes de \mathbf{y}_t são $I(d)$; e*
- (ii) *existe um vetor não nulo β tal que $\beta' \mathbf{y}_t \sim I(d-b)$, $d \geq b > 0$ e o vetor β é chamado de vetor de co-integração.*

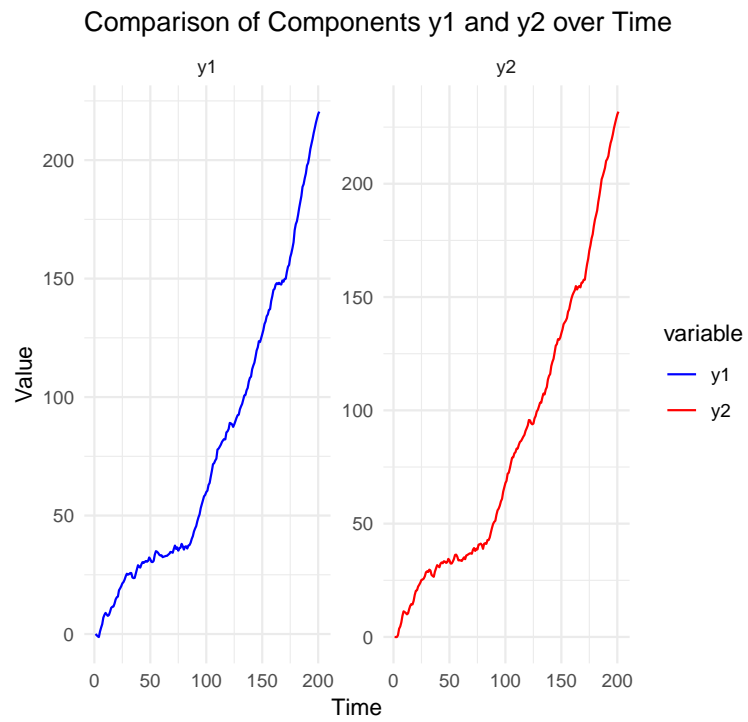
Observe que um caso particular em que $d = b = 1$ corresponde ao segundo caso apresentado acima e no terceiro caso temos $d = 2$ e $b = 1$.

O programa abaixo

Simulando VAR(1) com $\lambda_1 = \lambda_2 = 1$

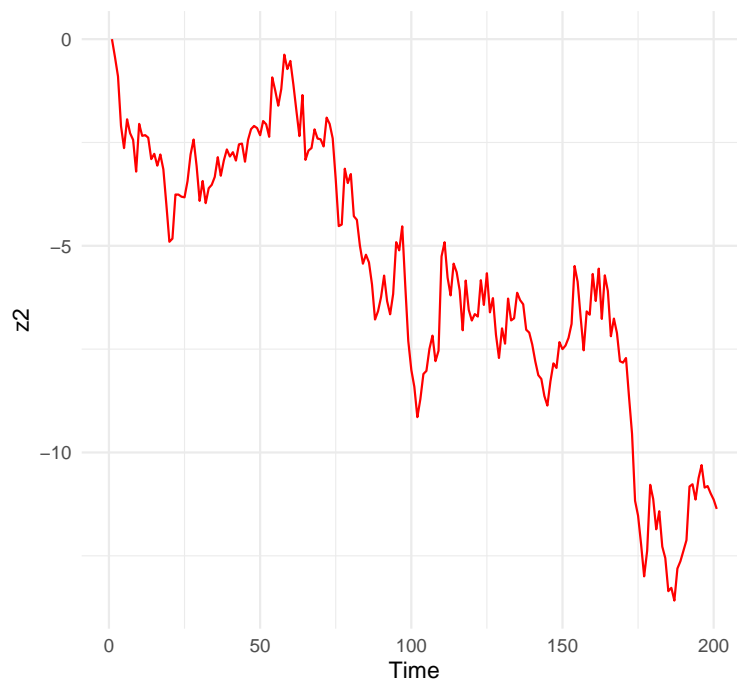
simula um VAR(1) com ambos os autovalores iguais a um.

Figura 7: VAR com séries $I(2)$



Observe que as séries tem tendência estocástica e testando raiz unitária não se rejeita que estas séries são $I(2)$. Mas se considerarmos a seguinte combinação linear $y_{1t} - y_{2t}$ obtem-se a seguinte série:

Figura 8: Relação de cointegração entre y_1 e y_2



que claramente tem tendência estocástica e é $I(1)$.

4.4 Exercícios

1. Considere o seguinte VAR(1)

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} .8 & .6 \\ -.6 & .8 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

onde

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \sim NI \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} : \begin{pmatrix} .3 & .2 \\ .2 & .2 \end{pmatrix} \right]$$

- (a) Escreva um programa para gerar uma realização de VAR.
- (b) Faça o gráfico contra o tempo de y_{1t} e de y_{2t} , o que voce pode dizer destas variáveis? Porque?
- (c) Obtenha os autovalores da matriz de coeficientes Φ , o que voce pode dizer a respeito destes autovalores?

2. Considere o seguinte VAR(1)

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} .8 & .8 \\ .6 & .6 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

onde

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \sim NI \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} : \begin{pmatrix} .3 & .2 \\ .2 & .2 \end{pmatrix} \right]$$

- (a) Escreva um programa para gerar uma realização de VAR.
- (b) Faça o gráfico contra o tempo de y_{1t} e de y_{2t} , o que voce pode dizer destas variáveis? Porque?
- (c) Obtenha os autovalores da matriz de coeficientes Φ , o que voce pode dizer a respeito destes autovalores?

3. Considere o seguinte VAR(1)

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} .8 & .2 \\ -.2 & 1.2 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

onde

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \sim NI \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} : \begin{pmatrix} .3 & .2 \\ .2 & .2 \end{pmatrix} \right]$$

- (a) Escreva um programa para gerar uma realização de VAR.
- (b) Faça o gráfico contra o tempo de y_{1t} e de y_{2t} , o que voce pode dizer destas variáveis? Porque?
- (c) Obtenha os autovalores da matriz de coeficientes Φ , o que voce pode dizer a respeito destes autovalores?
- (d) Os componentes de \mathbf{y}_t cointegram? Porque?
- (e) Caso a resposta no ítem anterior tenha sido afirmativa, qual é o vetor de cointegração. Esta combinação linear de y_{1t} e y_{2t} é estacionária?

5 Condições para que um VAR(1) Trivariado se reduza a AR univariados

Vamos considerar o caso de um VAR(1) com três variáveis e iremos considerar dois casos. No primeiro um dos autovalores é igual a um e os outros dois são distintos e menores do que um. No segundo caso dois dos autovalores serão iguais a um e o terceiro será menor do que um.

5.1 Primeiro Caso: um autovalor igual a 1 e os outros dois distintos e menores do que um

Neste caso a matriz Φ tem autovalores $\lambda_1 = 1$, $|\lambda_2| < 1$ e $|\lambda_3| < 1$ com $\lambda_2 \neq \lambda_3$. Neste caso temos o seguinte sistema para as variáveis \mathbf{z}_t

$$\begin{bmatrix} z_{1t} \\ z_{2t} \\ z_{3t} \end{bmatrix} = \begin{bmatrix} m_1^* \\ m_2^* \\ m_3^* \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} z_{1t-1} \\ z_{2t-1} \\ z_{3t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t}^* \\ \varepsilon_{2t}^* \\ \varepsilon_{3t}^* \end{bmatrix} \quad (29)$$

que implica que z_{1t} é $I(1)$ e z_{2t} e z_{3t} são $I(0)$. Deste modo temos que os componentes de \mathbf{y}_t são todos $I(1)$.

Observe que o vetor \mathbf{y}_t pode ser escrito como combinação linear de z_{1t} , z_{2t} e z_{3t} , isto é,

$$\mathbf{y}_t = p_1 z_{1t} + p_2 z_{2t} + p_3 z_{3t} \quad (30)$$

onde p_i é a i -ésima coluna da matriz P de autovetores associados aos autovalores λ_i 's.

Para que (30) resulte numa série estacionária, precisamos eliminar z_{1t} desta expressão. Isto pode ser conseguido multiplicando-se por $p^{(2)}$ ou $p^{(3)}$ segunda e terceira coluna da matriz P^{-1} uma vez que $p_1' p^{(2)} = p_1' p^{(3)} = 0$. Temos então duas relações de co-integração dadas por:

$$\begin{aligned} z_{2t} &= p^{(2)} \mathbf{y}_t \\ z_{3t} &= p^{(3)} \mathbf{y}_t \end{aligned}$$

Observe que diferente do caso bivariado em que a relação de co-integração era única agora temos uma **infinitude de relações de co-integração**. Isto porque como temos duas relações de co-integração estas relações vão gerar um plano onde qualquer vetor pertencente a este plano é uma combinação linear de variáveis estacionárias e portanto, estacionária, gerando uma nova relação de cointegração válida.

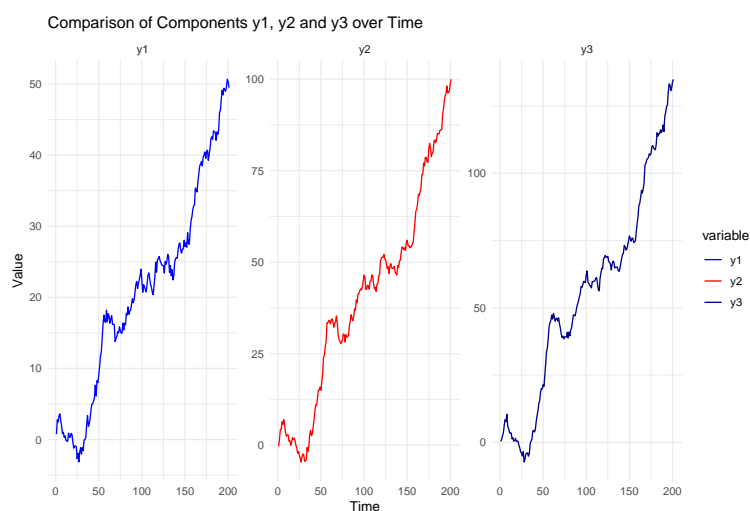
No programa abaixo

VAR(1) trivariado com $\lambda_1 = 1.0$, $\lambda_2 = 0.5$ e $\lambda_3 = 0.3$

gera um VAR(1) trivariado com $\lambda_1 = 1.0$, $\lambda_2 = 0.5$ e $\lambda_3 = 0.3$.

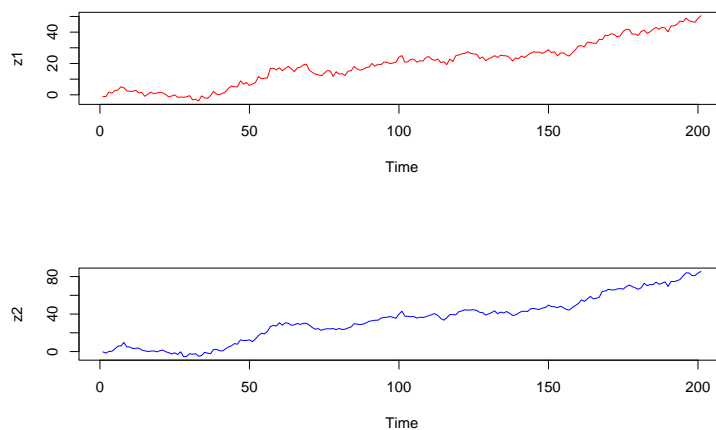
Abaixo são apresentadas as séries geradas.

Figura 9: VAR(1) trivariado com $\lambda_1 = 1.0$, $\lambda_2 = 0.5$ e $\lambda_3 = 0.3$



Observe que as três séries tem tendência estocástica e portanto são $I(1)$. Mas as seguintes combinações lineares resultam em séries estacionárias - $y_{2t} - y_{1t}$ e $y_{3t} - y_{1t}$ - cujos gráficos estão abaixo.

Figura 10: Duas relações de Cointegração



Como a matriz Φ pode ser escrita como $P \text{diag}\{0, 1 - \lambda_2, 1 - \lambda_3\} P^{-1}$, ela pode ser re-escrita como o produto de duas matrizes 3×2 ambas com posto dois, isto é $\Phi = \alpha\beta'$. O VAR pode então ser escrito da seguinte forma:

$$\begin{aligned}\Delta y_{1t} &= m_1 - p_{12}(1 - \lambda_2)z_{2t-1} - p_{13}(1 - \lambda_3)z_{3t-1} + \varepsilon_{1t} \\ \Delta y_{2t} &= m_2 - p_{22}(1 - \lambda_2)z_{2t-1} - p_{23}(1 - \lambda_3)z_{3t-1} + \varepsilon_{2t} \\ \Delta y_{3t} &= m_3 - p_{32}(1 - \lambda_2)z_{2t-1} - p_{33}(1 - \lambda_3)z_{3t-1} + \varepsilon_{3t}\end{aligned}$$

que em forma matricial é dado por:

$$\begin{aligned}\Delta \mathbf{y}_t &= \mathbf{m} - \Pi \mathbf{y}_{t-1} + \varepsilon_t \\ \Delta \mathbf{y}_t &= \mathbf{m} - \alpha\beta' \mathbf{y}_{t-1} + \varepsilon_t \\ \Delta \mathbf{y}_t &= \mathbf{m} - \alpha \mathbf{z}_{t-1} + \varepsilon_t\end{aligned}$$

com $\mathbf{z}_{t-1} = \beta' \mathbf{y}_{t-1}$ contendo as duas relações de co-integração.

5.2 Segundo Caso: dois autovalores iguais a 1 e o outro menor do que um

Neste caso a matriz Φ tem autovalores $\lambda_1 = 1$, $\lambda_2 = 1$ e $|\lambda_3| < 1$. Neste caso temos que usar a decomposição de Jordan, isto é, existe uma matriz C tal que $C\Phi C^{-1} = J$ onde J é dada por:

$$J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

Definindo $\mathbf{z}_t = C^{-1} \mathbf{y}_t$ temos que o VAR pode ser escrito da seguinte forma:

$$z_{1t} = m_1^* + z_{1t-1} + z_{2t-1} + \varepsilon_{1t}^* \quad (31)$$

$$z_{2t} = m_2^* + z_{2t-1} + \varepsilon_{2t}^* \quad (32)$$

$$z_{3t} = m_3^* + \lambda_3 z_{3t-1} + \varepsilon_{3t}^* \quad (33)$$

e portanto por (33) z_{3t} é $I(0)$, por (32) z_{2t} é $I(1)$ e por (31) z_{1t} é $I(2)$. Isto implica que os componentes de \mathbf{y}_t são no máximo $I(2)$ sendo que existem duas relações de cointegração sendo que uma delas gera uma série $I(1)$ e a outra $I(0)$.

Da mesma forma que foi feito no caso bivariado podemos escrever $\mathbf{y}_t = C \mathbf{z}_t$, isto é

$$\mathbf{y}_t = c_1 z_{1t} + c_2 z_{2t} + c_3 z_{3t} \quad (34)$$

onde em (34) c_i representa a i -ésima coluna da matriz C .

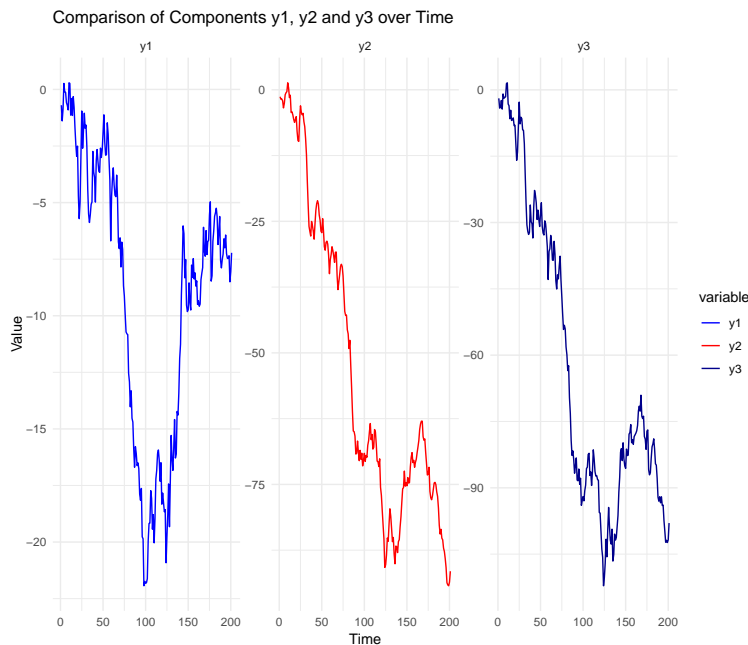
Para eliminarmos z_{1t} e z_{3t} de (34) devemos multiplicar esta expressão por $c^{(2)}$ obtendo-se então $z_{2t} = c^{(2)} \mathbf{y}_t$ que é $I(1)$. Por outro lado para eliminarmos z_{1t} e z_{2t} de (34) devemos multiplicar esta expressão por $c^{(3)}$ obtendo-se então $z_{3t} = c^{(3)} \mathbf{y}_t$ que é $I(0)$.

O programa

VAR(1) Trivariado com $\lambda_1 = \lambda_2 = 1$ e $\lambda_3 = 0.5$

gera um $VAR(1)$ trivariado com $\lambda_1 = \lambda_2 = 1$ e $\lambda_3 = 0.5$. Temos as seguintes séries:

Figura 11: $VAR(1)$ trivariado com $\lambda_1 = \lambda_2 = 1$ e $\lambda_3 = 0.5$



Observe que as séries tem tendência estocástica e são $I(2)$. Usando a combinação linear $y_{2t} - y_{1t}$ obtemos uma série $I(1)$ e usando a combinação $y_{3t} - y_{2t}$ obtemos uma série $I(0)$. Os gráficos abaixo ilustram estas combinações lineares.

5.3 Exercícios

1. Considere o seguinte $VAR(1)$

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 1.0 & 1.0 & 0.9 \\ 0.0 & 1.0 & 0.5 \\ 0.0 & 0.0 & 0.5 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

onde

$$\begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} \sim NI \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} : \begin{pmatrix} 1.0 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 1.0 \end{pmatrix} \right]$$

- (a) Escreva um programa para gerar uma realização de VAR.
- (b) Faça o gráfico contra o tempo de y_{1t} , y_{2t} e y_{3t} , o que voce pode dizer destas variáveis? Porque?

- (c) Obtenha os autovalores da matriz de coeficientes Φ , o que voce pode dizer a respeito destes autovalores?

2. Considere o seguinte VAR(1)

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 1.0 & 1.0 & 0.9 \\ 0.0 & 0.5 & 0.5 \\ 0.0 & 0.0 & 0.3 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

onde

$$\begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} \sim NI \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} : \begin{pmatrix} 1.0 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 1.0 \end{pmatrix} \right]$$

- (a) Escreva um programa para gerar uma realização de VAR.
- (b) Faça o gráfico contra o tempo de y_{1t} , y_{2t} e y_{3t} , o que voce pode dizer destas variáveis? Porque?
- (c) Obtenha os autovalores da matriz de coeficientes Φ , o que voce pode dizer a respeito destes autovalores?

6 Especificação do Modelo VAR Geral

Considere o seguinte VAR

$$\mathbf{y}_t = \mathbf{m} + \Phi_1 \mathbf{y}_{t-1} + \dots + \Phi_p \mathbf{y}_{t-p} + \varepsilon_t \quad (35)$$

onde \mathbf{y}_t é um vetor $k \times 1$ e $\varepsilon_t \sim NI(0, \Omega)$.

Usando as identidades

$$\mathbf{y}_t = \Delta \mathbf{y}_t + \mathbf{y}_{t-1} \quad (36)$$

e

$$\mathbf{y}_{t-j} = \mathbf{y}_{t-1} - \sum_{k=1}^{j-1} \Delta \mathbf{y}_{t-k} \quad j = 2, \dots, p \quad (37)$$

podemos escrever (35) da seguinte forma

$$\Delta \mathbf{y}_t = \mathbf{m} + \mathbf{B}_1 \Delta \mathbf{y}_{t-1} + \dots + \mathbf{B}_{p-1} \Delta \mathbf{y}_{t-p+1} - \Pi \mathbf{y}_{t-1} + \varepsilon_t \quad (38)$$

onde $\mathbf{B}_i = - \sum_{k=i+1}^p \Phi_k$ para $i = 1, \dots, p-1$ e $\Pi = I - \Phi_1 - \dots - \Phi_p$.

Como vimos nos caso bi e tri-variado o comportamento de \mathbf{y}_t vai depender dos autovalores da matriz de longo-prazo Π , que podem ser iguais a zero - implicando em raíz unitária - ou menores do que zero - implicando em componentes estacionários.

Podemos então ter os seguintes casos:

1. $Posto(\Pi) = k$ - todos os autovalores de Π são diferentes de zero e portanto esta matriz tem posto completo e implicando que os autovalores de $\Phi(1) = \Phi_1 + \dots + \Phi_p$ são todos menores do que um, portanto todos os componentes de \mathbf{y}_t são estacionários e a representação válida é o $VAR(p)$ em nível dado por (35)
2. $Posto(\Pi) = 0$ - todos os autovalores de Π são zero e portanto esta matriz é indistinguível da matriz nula. Implica também que $\Phi(1)$ tem todos os autovalores iguais a um e portanto os componentes de \mathbf{y}_t são no mínimo $I(1)$ e a representação válida é um $VAR(p-1)$ em primeira diferença, isto é (78) sem o termo em nível.
3. $0 < Posto(\Pi) = r < k$ - neste caso temos $k - r$ autovalores iguais a zero e r autovalores diferentes de zero. Os componentes de \mathbf{y}_t são no mínimo $I(1)$ e a representação válida é (78) com $\Pi = \alpha\beta'$ onde α e β são matrizes $k \times r$ de posto r . Esta representação é chamada Vetor de Correção do Equilíbrio, $VCEq$, e nela estão presentes r relações de co-integração.

7 Estimação de Modelos VAR

Sem perda de generalidade vamos considerar um $VAR(1)$ uma vez que num $VAR(p)$ se definirmos o vetor $\xi_t = [\mathbf{y}'_t : \mathbf{y}'_{t-1} : \dots : \mathbf{y}'_{t-p+1}]$ de dimensão $kp \times 1$, (35) pode ser re-escrito da seguinte forma:

$$\xi_t = \mathbf{F}\xi_{t-1} + v_t \quad (39)$$

onde a matriz \mathbf{F} é dada por:

$$\mathbf{F} = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ & \mathbf{I}_{kp-1} & & & \mathbf{0} \end{bmatrix} \quad (40)$$

e o vetor v_t é dado por:

$$v_t = \begin{bmatrix} \varepsilon_t \\ \mathbf{0}_{kp-1} \end{bmatrix} \quad (41)$$

e satisfaz as seguintes propriedades:

1. $E(v_t) = 0$
2. $E(v_t v'_s) = \begin{cases} \mathbf{Q} & \text{para } t = s \\ \mathbf{0} & \text{para } t \neq s \end{cases}$
3. $\mathbf{Q} = \begin{bmatrix} \Omega & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$

Se todos os componentes de \mathbf{y}_t forem estacionários o modelo pode ser estimado por Máxima Verossimilhança - exata ou aproximada.

Se os componentes de \mathbf{y}_t forem não estacionários temos que determinar a dimensão do espaço de cointegração para depois estimarmos o VAR .

Em ambos os casos é necessário determinar a ordem do VAR .

Primeiro vamos apresentar a estimação de um $VAR(1)$ por Máxima Verossimilhança para depois obter as propriedades dos estimadores assim como testes de hipóteses a respeito dos parâmetros que podem ser usados para determinar a ordem do VAR . A seguir

a estimação por Regressões Aparentemente não Relacionadas (SURE) será apresentada. E por último a determinação da ordem do VAR será obtida por critérios de informação.

7.1 Estimação do $VAR(1)$

Vamos assumir que a ordem é conhecida e igual a um.

Neste caso a função de verossimilhança para um $VAR(1)$ é dada por:

$$\begin{aligned} \ln \ell(\Psi) = & -\frac{k(T-1)}{2} \ln(2\pi) + \frac{(T-1)}{2} \ln |\Omega^{-1}| \\ & -\frac{1}{2} \sum_{t=2}^T (\mathbf{y}_t - \Phi_1 \mathbf{y}_{t-1})' \Omega^{-1} (\mathbf{y}_t - \Phi_1 \mathbf{y}_{t-1}) \\ & -\frac{N}{2} \ln(2\pi) - \frac{(T-1)}{2} \ln |\Gamma(0)| - \frac{1}{2} \mathbf{y}_1' \Gamma^{-1}(0) \mathbf{y}_1 \end{aligned} \quad (42)$$

e a maximização desta função em relação a Φ_1 , Ω e $\Gamma(0)$ é feita de forma numérica.

Ao se desconsiderar \mathbf{y}_1 a maximização de $\ln \ell(\Psi)$ reduz-se a minimização de:

$$\begin{aligned} S(\Psi) = & \frac{k(T-1)}{2} \ln(2\pi) - \frac{(T-1)}{2} \ln |\Omega^{-1}| \\ & + \frac{1}{2} \sum_{t=2}^T (\mathbf{y}_t - \Phi_1 \mathbf{y}_{t-1})' \Omega^{-1} (\mathbf{y}_t - \Phi_1 \mathbf{y}_{t-1}) \end{aligned} \quad (43)$$

e os estimadores de máxima verossimilhança de Φ_1 e Ω serão dados por⁵:

$$\hat{\Phi}_1 = \left[\sum_{t=2}^T \mathbf{y}_t \mathbf{y}_{t-1}' \right] \left[\sum_{t=2}^T \mathbf{y}_{t-1} \mathbf{y}_{t-1}' \right]^{-1} \quad (44)$$

e

$$\hat{\Omega} = \frac{1}{T} \left[\sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t' \right] \quad (45)$$

onde os resíduos são dados por:

$$\hat{\varepsilon}_t = \mathbf{y}_t - \hat{\Phi}_1 \mathbf{y}_{t-1}$$

7.2 Distribuição dos Estimadores e Teste de Hipóteses

Como Φ_1 é uma matriz é mais fácil obter a distribuição para o $vec(\Phi_1)$ que será denotado por ϕ . Temos então

$$\hat{\phi} \sim N(\phi, \mathbf{AVAR}(\hat{\phi})) \quad (46)$$

onde

⁵Veja apêndice (14) para a demonstração

$$\mathbf{AVAR}(\widehat{\phi}) = T^{-1}(\Omega \otimes Q^{-1}) \quad (47)$$

e

$$Q = T^{-1} \sum_{t=2}^T \mathbf{y}_t \mathbf{y}_t' \quad (48)$$

Dada a distribuição do E.M.V para $\phi = \text{vec}(\Phi_1)$ podemos construir testes de hipóteses para ϕ usando-se a estatística de Wald ou da razão de verossimilhança. Por exemplo para se testar $H_0 : \phi = \mathbf{0}$ a estatística de Wald é dada por:

$$\mathbf{W} = \widehat{\phi}' [\mathbf{AVAR}(\widehat{\phi})]^{-1} \widehat{\phi} \sim \chi_{n^2}^2$$

Para o teste da razão de verossimilhança necessitamos obter o máximo de $\ln \ell(\Psi)$ tanto para o modelo restrito quanto para o modelo irrestrito.

Observe que

$$\begin{aligned} \ln \ell(\widehat{\Psi}_R) &\propto -\frac{T}{2} \ln |\widehat{\Omega}_{ols,R}| \\ \ln \ell(\widehat{\Psi}_I) &\propto -\frac{T}{2} \ln |\widehat{\Omega}_{ols,I}| \end{aligned}$$

e a estatística de teste será dada por:

$$\begin{aligned} LR &= -2(\ln \ell(\widehat{\Psi}_R) - \ln \ell(\widehat{\Psi}_I)) \\ &= T\{\ln |\widehat{\Omega}_{ols,R}| - \ln |\widehat{\Omega}_{ols,I}|\} \end{aligned} \quad (49)$$

Observe que (49) pode ser usada para determinar a ordem do VAR uma vez que o modelo restrito é um $VAR(p-1)$ e o irrestrito é um $VAR(p)$.

7.3 Estimação por SURE

Considere o $VAR(1)$ dado por

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \varepsilon_t \quad \text{para } t = 1, \dots, T \quad (50)$$

onde \mathbf{y}_t é um vetor $n \times 1$.

Agora defina $\mathbf{y}_v = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]'$ onde cada \mathbf{y}_i para $i = 1, \dots, n$ é um vetor $T \times 1$ e portanto \mathbf{y}_v será um vetor $Tn \times 1$.

Seja $\mathbf{X}_v = \text{diag}\{\mathbf{X}, \mathbf{X}, \dots, \mathbf{X}\} = \mathbf{X} \otimes \mathbf{I}_n$ matriz $Tn \times n^2$ onde cada matriz \mathbf{X} é $T \times n$ dada por $\mathbf{X} = [\mathbf{y}_1^{(d)}, \mathbf{y}_2^{(d)}, \dots, \mathbf{y}_n^{(d)}]$ e cada $\mathbf{y}_i^{(d)}$ representa a defasagem de um período de \mathbf{y}_i .

Seja também $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_n]'$ um vetor $n^2 \times 1$ em que cada $\boldsymbol{\beta}_i$ corresponde a i -ésima linha de Φ_1 .

E seja também $\boldsymbol{\varepsilon}_v = [\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n]'$ um vetor $Tn \times 1$ com distribuição normal com média zero e matriz de covariância dada por $\Omega \otimes \mathbf{I}_T$.

Então (50) pode ser rescrito da seguinte forma:

$$\mathbf{y}_v = \mathbf{X}_v \boldsymbol{\beta} + \boldsymbol{\varepsilon}_v \quad (51)$$

Primeiro vamos mostrar que o estimador de MQO de $\boldsymbol{\beta}$ em (51) é equivalente os estimador de MQO equação por equação.

O estimador de MQO de $\boldsymbol{\beta}$ em (51) é dado por:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{MQO} &= (\mathbf{X}'_v \mathbf{X}_v)^{-1} \mathbf{X}'_v \mathbf{y}_v \\ &= [(\mathbf{X}' \otimes \mathbf{I}_k)(\mathbf{X} \otimes \mathbf{I}_k)]^{-1} (\mathbf{X}' \otimes \mathbf{I}_k) \mathbf{y}_v \\ &= [\mathbf{X}' \mathbf{X} \otimes \mathbf{I}_k]^{-1} (\mathbf{X}' \otimes \mathbf{I}_k) \mathbf{y}_v \\ &= \text{diag}\{(\mathbf{X}' \mathbf{X})^{-1}, (\mathbf{X}' \mathbf{X})^{-1}, \dots, (\mathbf{X}' \mathbf{X})^{-1}\} \text{diag}\{\mathbf{X}' \mathbf{y}_1, \mathbf{X}' \mathbf{y}_2, \dots, \mathbf{X}' \mathbf{y}_k\} \\ &= \text{diag}\{(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_1, (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_2, \dots, (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_k\} \\ &= \text{diag}\{\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \dots, \hat{\boldsymbol{\beta}}_k\} \end{aligned} \quad (52)$$

A seguir vamos mostrar que o estimador de MQG de $\boldsymbol{\beta}$ em (51) é equivalente os estimador de MQO.

O estimador de MQG de $\boldsymbol{\beta}$ em (51) é dado por:

$$\hat{\boldsymbol{\beta}}_{MQG} = (\mathbf{X}'_v (\Omega \otimes \mathbf{I}_T)^{-1} \mathbf{X}_v)^{-1} \mathbf{X}'_v (\Omega \otimes \mathbf{I}_T)^{-1} \mathbf{y}_v \quad (53)$$

Agora observe que

$$\begin{aligned} \mathbf{X}'_v (\Omega \otimes \mathbf{I}_T)^{-1} &= (\mathbf{I}_k \otimes \mathbf{X}') (\Omega \otimes \mathbf{I}_T)^{-1} \\ &= (\mathbf{I}_k \otimes \mathbf{X}') (\Omega^{-1} \otimes \mathbf{I}_T) \\ &= (\Omega^{-1} \otimes \mathbf{X}') \\ &= (\Omega^{-1} \otimes \mathbf{I}_k) (\mathbf{I}_k \otimes \mathbf{X}') \\ &= (\Omega^{-1} \otimes \mathbf{I}_k) \mathbf{X}'_v \end{aligned} \quad (54)$$

e substituindo (54) em (53) temos:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{MQG} &= [(\Omega^{-1} \otimes \mathbf{I}_k) \mathbf{X}'_v \mathbf{X}_v]^{-1} (\Omega^{-1} \otimes \mathbf{I}_k) \mathbf{X}'_v \mathbf{y}_v \\ &= (\mathbf{X}'_v \mathbf{X}_v)^{-1} (\Omega^{-1} \otimes \mathbf{I}_k)^{-1} (\Omega^{-1} \otimes \mathbf{I}_k) \mathbf{X}'_v \mathbf{y}_v \\ &= (\mathbf{X}'_v \mathbf{X}_v)^{-1} \mathbf{X}'_v \mathbf{y}_v \\ &= \hat{\boldsymbol{\beta}}_{MQOs} \end{aligned}$$

7.4 Determinando a ordem do $VAR(p)$

A ordem to $VAR(p)$ pode ser determinada usando o teste da razão de verossimilhança apresentado na seção (7.2).

O procedimento geral consistente em estimar modelos $VAR(p)$ com ordens $p = p_{max}, \dots, 0$ e escolher o valor de p que minimiza algum critério de seleção de modelos, com a preocupação de manter o tamanho da amostra fixada em $p_{max} + 1, \dots, T$.

A expressão geral dos critérios de seleção de modelos é dado por:

$$\begin{aligned}
MSC(p) &= \ln |\hat{\Omega}(p)| + c_T \cdot \varphi(n, p), \\
\hat{\Omega}(p) &= \frac{1}{T} \left[\sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t' \right], \\
c_T &= \text{função do tamanho da amostra}, \\
\varphi(n, p) &= \text{função de penalização}.
\end{aligned}$$

Os três critérios de informação mais usados são Akaike(AIC) (veja Akaike (1973)), Schwarz ou Bayesian (BIC) (veja Schwarz (1978)) e Hannan-Quinn (HQ) (veja Hannan and Quinn (1979)), que são dados por:

$$\begin{aligned}
AIC(p) &= \ln |\hat{\Omega}(p)| + \frac{2}{T}pn^2, \\
BIC(p) &= \ln |\hat{\Omega}(p)| + \frac{\ln(T)}{T}pn^2, \\
HQ(p) &= \ln |\hat{\Omega}(p)| + \frac{2 \ln(\ln(T))}{T}pn^2.
\end{aligned}$$

E o objetivo é minimizar estes critérios de informação.

O critério de informação AIC superestima, assintoticamente, a order com probabilidade positiva.

Os critérios BIC e HQ estimam a ordem de forma consistente sob condições bastante gerais se a ordem verdadeira p for menor ou igual a p_{max} .

7.5 Causalidade de Granger

Um dos principais usos dos modelos VAR é a previsão. A seguinte noção intuitiva da capacidade de previsão de uma variável é devida a Granger (veja Granger (1969)):

1. Se uma variável, ou grupo de variáveis, y_1 for útil para prever outra variável, ou grupo de variáveis y_2 , então diz-se que y_1 causa no sentido de Granger y_2 ; caso contrário, diz-se que não causa Granger a y_2 ;
2. Formalmente, y_1 não causa no sentido de Granger a y_2 se, para todos os $s > 0$, o Erro Médio Quadrático (EMQ) ou em inglês *Mean Squared Error (MSE)* de uma previsão de $y_{2,t+s}$ baseada em $(y_{2,t}, y_{2,t-1}, \dots)$ tem o mesmo *MSE* de a previsão de $y_{2,t+s}$ baseada em $(y_{2,t}, y_{2,t-1}, \dots)$ e $(y_{1,t}, y_{1,t-1}, \dots)$;
3. Observe que implicitamente estamos assumindo que os componentes de \mathbf{y}_t são estacionários.
4. A noção de causalidade de Granger não implica em relação causa entre as variáveis. Ela apenas implica capacidade de previsão.

7.6 Exemplo com os dados gerados segundo um $VAR(1)$ estacionário

Usando o programa

$VAR(1)$ estacionário.

Primeiro supondo que o número de defasagens é desconhecido, temos que determinar a ordem do VAR . Para se obter tal ordem, usamos o comando `VARselect()` usando como defasagem máxima 10. Temos os seguintes resultados:

Tabela 1: Critérios de Informação para VAR

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-0.9903	-0.9751	-0.9564	-0.9348	-0.9169	-0.8985	-0.8883	-0.8523	-0.8199	-0.8084
HQ(n)	-0.9330	-0.8891	-0.8418	-0.7915	-0.7449	-0.697	-0.6591	-0.5944	-0.5333	-0.4932
SIC(n)	-0.8490	-0.7630	-0.6737	-0.5813	-0.4927	-0.4037	-0.3229	-0.2161	-0.1130	-0.0309
FPE(n)	0.3714	0.3772	0.3843	0.3928	0.3999	0.4074	0.4117	0.4270	0.4413	0.4466

segundo todos os critérios de informação a ordem ótima é a primeira.

Observe que nas variáveis endógenas listou-se as duas variáveis, a saber, y_1 e y_2 . Como variáveis exógenas a constante e a tendência linear determinística e o VAR estimado foi irrestrito.

Obtém-se as seguintes estimativas:

Tabela 2: Estimação do $VAR(1)$

	$y_{1,t}$	$y_{2,t}$
$y_{1,t-1}$	0.5019	0.0476
	(0.0817)	(0.0952)
	[6.1410]	[0.5000]
	{0.0000}	{0.6174}
$y_{2,t-1}$	-0.1379	0.2948
	(0.0802)	(0.0935)
	[-1.7180]	[3.1754]
	{0.0874}	{0.0019}
$Const$	-0.0703	-0.1504
	(0.1416)	(0.1391)
	[-0.4970]	[-1.0820]
	{0.6201}	{0.0454}
$trend$	0.0022	0.0025
	(0.0012)	(0.0012)
	[1.7810]	[2.0130]
	{0.0765}	{0.4544}
R-squared	0.3198	0.0703
Adj. R-squared	0.3094	0.0560
Sum sq. resids	193.0992	189.6300
S.E. equation	0.98520	0.9675
F-statistic	30.726	4.936
Log likelihood	-278.7774	-275.1656
Akaike AIC	2.8277	2.7917
Schwarz SC	2.8937	2.8576
Mean dependent	0.3247	0.0792
S.D. dependent	1.1854	0.9958
Determinant Residual Covariance	0.3637	
Log Likelihood (d.f. adjusted)	-462.4007	
Akaike Information Criteria	4.7040	
Schwarz Criteria	4.8359	
Nota: amostra de 2 a 201, desvios padrões em () e estatística t em [] e p-valor em { }.		

Observe que na segunda e terceira coluna temos não só as estimativas dos coeficientes, assim como os desvios padrões, as estatísticas t — *Student* e os p-valores. Além disto temos as medidas de qualidade de ajuste para cada uma das equações em separado assim como nas quatro últimas linhas as medidas de qualidade de ajuste para as duas equações ao mesmo tempo.

Observe que a constante e a tendência não são significativas em ambas as equações e portanto podemos retirá-la. Obtém-se então os seguintes resultados:

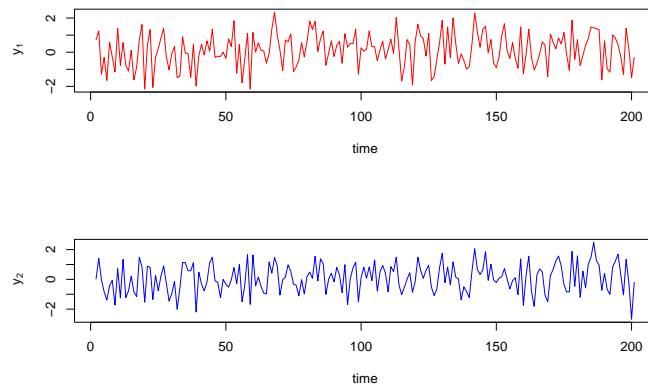
Tabela 3: Estimação do $VAR(1)$ sem constante e tendência

	$y_{1,t}$	$y_{2,t}$
$y_{1,t-1}$	0.5777 (0.0776) [7.4454] {0.0000}	-0.0765 (0.0759) [-1.0080] {0.8470}
$y_{2,t-1}$	0.0185 (0.0955) [0.1934] {0.3147}	0.2751 (0.0934) [2.9443] {0.0036}
R-squared	0.294963	0.043624
Adj. R-squared	0.291403	0.038794
Sum sq. resids	197.1732	188.7319
S.E. equation	0.997910	0.976315
F-statistic	82.83648	9.031626
Log likelihood	-282.3642	-277.9887
Akaike AIC	2.843642	2.799887
Schwarz SC	2.876625	2.832870
Mean dependent	0.324772	0.079258
S.D. dependent	1.185473	0.995822
Determinant resid covariance (dof adj.)	0.3717	
Determinant resid covariance	0.3643	
Log Likelihood	-466.6099	
Akaike information criterion	4.70609	
Schwarz criterion	4.772065	
Number of coefficients	4	

Esta simplificação é válida como pode ser visto pela redução do Critério de Informação de Schwarz.

A seguir apresentamos os gráficos dos resíduos nas duas equações

Figura 12: Resíduos do $VAR(1)$ sem constante e tendência



Os resíduos parecem ruídos brancos e se testarmos autocorrelação nos resíduos não se rejeita que os mesmos são não correlacionados conforme o teste de Portmanteau para autocorrelação apresentado abaixo:

Tabela 4: Portmanteau Test for Autocorrelation

Portmanteau Test (asymptotic)		
$\chi^2 = 53.852$	$df = 60$	$p - value = 0.6985$

Testando-se normalidade obtém-se:

Tabela 5: Normality test for Residuals

Component	Skewness	χ^2	df	Prob*
1	-0.1815	1.0985	1	0.2946
2	-0.0415	0.0574	1	0.8106
Joint		1.1560	2	0.5610
Component	Kurtosis	χ^2	df	Prob*
1	2.4552	2.4733	1	0.1158
2	2.9166	0.0580	1	0.8097
Join		2.5313	2	0.2821
Component	Jarque-Bera	df	Prob*	
1	3.5719	2	0.1676	
2	3.1527	2	0.2067	
Joint	3.6873	4	0.4500	
*Approximate p-values do not				
account for coefficient estimation				

Testando-se heteroscedasticidade, pelo teste de ARCH, obtém-se:

Tabela 6: ARCH test for Residuals

Component	χ^2	df	Prob*
y_1	11.719	16	0.2946
y_2	19.072	16	0.269
Joint	38.794	45	0.731
*Approximate p-values do not			
account for coefficient estimation			

Portanto não se pode rejeitar que os resíduos são ruídos branco.

Desejamos também verificar se o VAR estimado tem componentes estacionários. Uma forma de verificar é através dos autovalores da matriz de longo-prazo. O gráfico abaixo apresenta tais autovalores:

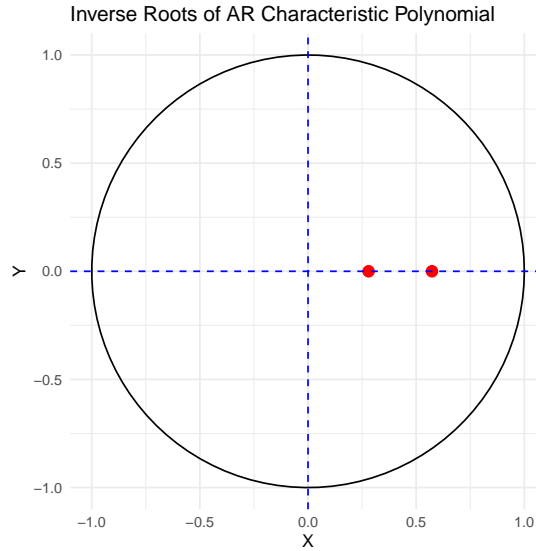


Figura 13: Autovalores da Matriz de Longo Prazo

Pelo gráfico fica claro que ambos os autovalores são menores do que um e portanto ambas as componentes são estacionárias.

A seguir vamos testar não causalidade de Granger. O programa a seguir faz este teste

Teste de causalidade de Granger

primeiro vamos testar se $H_0 : y_1$ não causa no sentido de Granger y_2 , e, temos os seguintes resultados:

Tabela 7: $H_0 : y_1$ não causa no sentido de Granger y_2

Granger causality test	
$F(1, 396) = 1.016$	$p - value = 0.3141$

que não se rejeita a nula neste caso.

Segundo vamos testar se $H_0 : y_2$ não causa no sentido de Granger y_2 , e, temos os seguintes resultados:

Tabela 8: $H_0 : y_1$ não causa no sentido de Granger y_2

Granger causality test	
$F(1, 396) = 0.0374$	$p - value = 0.8468$

que não se rejeita a nula neste caso.

8 Previsão para um $VAR(1)$

Como vimos acima que todo $VAR(p)$ pode ser reescrito como um $VAR(1)$ basta apresentar o previsor ótimo assim como o erro médio quadrático de previsão para o $VAR(1)$.

Considere o $VAR(1)$ dado em (50) temos que o previsor l passos a frente é dado por:

$$\hat{\mathbf{y}}_{T+l|T} = \Phi_1 \hat{\mathbf{y}}_{T+l-1|T} \quad (55)$$

o que implica em

$$\hat{\mathbf{y}}_{T+l|T} = \Phi_1^l \mathbf{y}_T \quad (56)$$

que vale para $l = 1, 2, \dots$

Este método é usualmente definido como [previsão iterada](#).

Podemos também fazer a previsão usando o [modelo direto](#) isto é o $VAR(1)$ é agora escrito da seguinte forma:

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-h} + \varepsilon_t \quad (57)$$

cujas previsões l passos a frente é dado por:

$$\tilde{\mathbf{y}}_{T+l|T} = \Phi_1 \mathbf{y}_T \quad (58)$$

Sob correta especificação, $\hat{\mathbf{y}}_{T+l}$ é mais eficiente do que $\tilde{\mathbf{y}}_{T+l}$. No entretanto, na presença de especificação incorreta do modelo $\hat{\mathbf{y}}_{T+l}$ pode ser mais robusta do que $\tilde{\mathbf{y}}_{T+l}$.

Por outro lado podemos obter as previsões ótimas usando a representação $MA(\infty)$ dada por:

$$\hat{\mathbf{y}}_{T+l|T} = \sum_{j=0}^{\infty} \Psi_{j+l} \varepsilon_{T-l} \quad (59)$$

onde o associado erro de previsão é dado por:

$$\mathbf{v}_{T+l} = \mathbf{y}_{T+l} - \hat{\mathbf{y}}_{T+l|T} = \sum_{j=0}^{l-1} \Psi_1^j \varepsilon_{T+l-j} \quad (60)$$

Portanto a matriz de variância-covariância do erro de previsão é dada por:

$$V(\mathbf{v}_{T+l}) = \Omega + \Psi_1 \Omega \Psi_1' + \dots + \Psi_{l-1} \Omega \Psi_{l-1}' \quad (61)$$

onde os elementos da diagonal de $V(\mathbf{v}_{T+l})$ podem ser usados para construir intervalos de confiança para as previsões das variáveis $\hat{y}_{j,T+l}$, onde $j = 1, \dots, n$.

A expressão (59) que dá a previsão ótima pode ser reescrita da seguinte forma:

$$\begin{aligned} \hat{\mathbf{y}}_{T+l|T} &= \Psi_l \varepsilon_T + \sum_{j=0}^{\infty} \Psi_{j+l+1} \varepsilon_{T-1-l} \\ &= \hat{\mathbf{y}}_{T+l|T-1} + \Psi_l \varepsilon_T \end{aligned} \quad (62)$$

e usando (61) temos:

$$\hat{\mathbf{y}}_{T+h} = \hat{\mathbf{y}}_{T+h|T-1} + \mathbf{C}_h \left(\underbrace{\mathbf{y}_T - \hat{\mathbf{y}}_{T|T-1}}_{\text{forecast error}} \right) \quad (63)$$

O lado esquerdo da equação (63) é a **previsão ótima para o período $T + h$ feita no período T**

e o lado direito é a **previsão ótima para o período $T + h$ feita no período $T - 1$ e erro de previsão um passo à frente cometido na previsão de y_T no período $T - 1$** .

Presumiu-se que os parâmetros eram conhecidos.

Quando os parâmetros são estimados, as expressões para a variância do erro de previsão devem ser modificadas para levar em conta a incerteza na estimativa dos parâmetros.

Veja Lütkepohl (2007) para uma explicação detalhada dessas expressões.

8.1 Exemplo com horizonte de previsão igual a 10

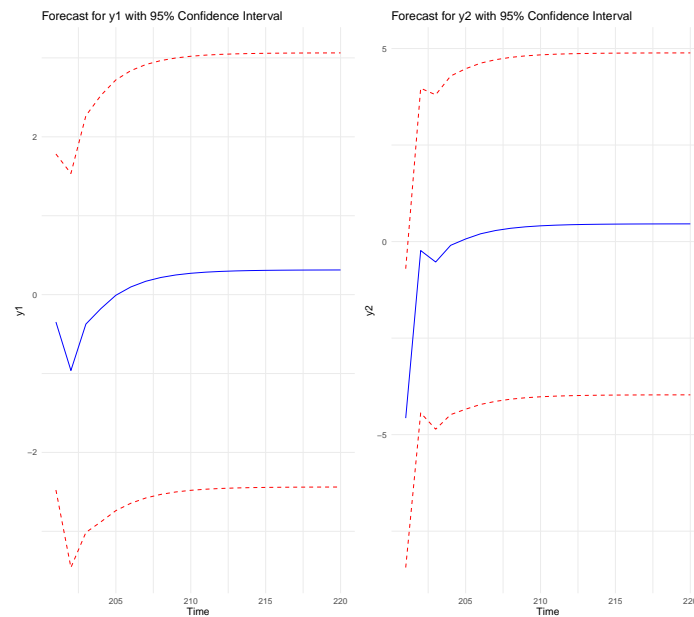
No programa

Previsão 10 passos à frente para o $VAR(1)$ estacionário.

gera um $VAR(1)$ e faz previsões dez passos a frente.

O resultado deste programa é a previsão pontual para cada uma das variáveis, que neste caso são duas, assim como os intervalos de previsão usando 2 desvios padrões como limites dos intervalos. Temos então o seguinte gráfico:

Figura 14: Previsão 10 passos à frente para o $VAR(1)$ estacionário



9 Função de Resposta ao Impulso

Um VAR estacionário pode ser escrito como um $MA(\infty)$, isto é:

$$\begin{aligned}\mathbf{y}_t &= \Phi^{-1}(L)\varepsilon_t \\ &= \Psi(L)\varepsilon_t\end{aligned}\tag{64}$$

com $\varepsilon_t \sim NI(0, \Omega)$.

Como Ω é positiva definida, existe uma matriz não singular P tal que:

$$P\Omega P' = \mathbf{I}\tag{65}$$

Então (64) pode ser reescrito da seguinte forma:

$$\begin{aligned}\mathbf{y}_t &= \Psi(L)P^{-1}P\varepsilon_t \\ &= \Theta(L)v_t\end{aligned}\tag{66}$$

com $v_t = P\varepsilon_t$ e portanto temos que:

1. $E(v_t) = 0$;
2. $E(v_t v_t') = P\Omega P'$.

Então a expressão (66) é a representação $MA(\infty)$ do seguinte modelo:

$$P\Phi(L)\mathbf{y}_t = v_t\tag{67}$$

A expressão acima é um VAR Estrutural ($SVAR$) pois existem relações contemporâneas entre as variáveis por causa da matriz P .

Os erros ortogonais v_t são os choques estruturais (econômicos).

Como a variável em análise reagirá aos choques estruturais escrevendo o polinômio $\Psi(L)$ como

$$\begin{aligned}\Theta(L) &= P^{-1} - \Psi_1 P^{-1}L - \Psi_2 P^{-1}L^2 - \dots \\ &= \Theta_0 - \Theta_1 L - \Theta_2 L^2 - \dots\end{aligned}\tag{68}$$

Um choque é um vetor com um elemento igual a um e todos os outros iguais a zero, por exemplo:

$$v_{1,t+1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad v_{2,t+1} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad v_{n,t+1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}\tag{69}$$

Os elementos diferentes de zero também podem ser definidos como qualquer outra constante, como o desvio padrão do elemento correspondente de ε_t .

Como o sistema é linear, o tamanho do choque é irrelevante porque a resposta é apenas proporcional ao tamanho do choque.

A resposta em período $t + i$ de y a um choque no período $t + 1$ será:

$$\begin{aligned}\frac{\partial \mathbf{y}_{t+1}}{\partial v_{t+1}} &= P^{-1} = \Theta_0 \\ \frac{\partial \mathbf{y}_{t+2}}{\partial v_{t+1}} &= -\Theta_1 \\ \frac{\partial \mathbf{y}_{t+3}}{\partial v_{t+1}} &= -\Theta_2 \\ &\vdots\end{aligned}\tag{70}$$

onde, por exemplo:

$$\Theta_0 = \begin{bmatrix} \frac{\partial y_{1,t+1}}{\partial v_{1,t+1}} & \cdots & \frac{\partial y_{1,t+1}}{\partial v_{n,t+1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_{n,t+1}}{\partial v_{1,t+1}} & & \frac{\partial y_{n,t+1}}{\partial v_{n,t+1}} \end{bmatrix}\tag{71}$$

e os termos $\frac{\partial y_{k,t+i}}{\partial v_{j,t+1}}$ são conhecidos como **resposta ao impulso** no período $t + i$ à variável k ao choque j com $k, j = 1, \dots, n$ e $i = 1, 2, \dots$.

A coleção de todas as respostas ao impulso é conhecida como a **função de resposta ao impulso (IRF)**.

A matriz P não é única e, como consequência, a função de resposta ao impulso (IRF) também não é única.

Como Ω tem $\frac{n(n+1)}{2}$ elementos distintos, este é o número máximo de elementos irrestitos em P . Assim, tipicamente, P é escolhido como uma matriz triangular, ou seja,

$$P = \begin{bmatrix} p_{11} & 0 & \cdots & 0 \\ p_{21} & p_{22} & \cdots & 0 \\ \vdots & & \ddots & \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

onde os elementos p_{ij} são tais que $P\Omega P' = \mathbf{I}$.

O programa

IRF.

obtem as IRF para o $VAR(1)$ estacionário.

No primeiro caso as IRF são obtidas com a opção orthogonal. Temos os seguintes gráficos:

Figura 15: IRF ortogonal para y_1

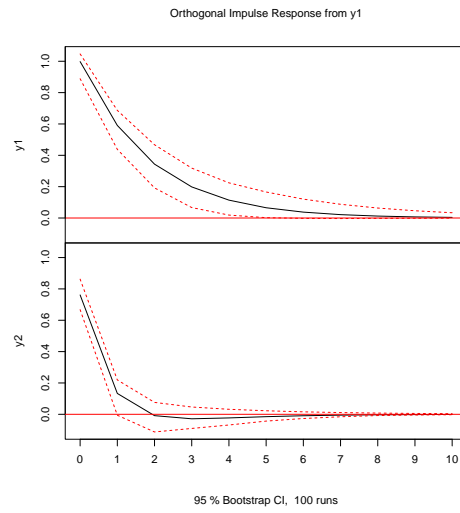
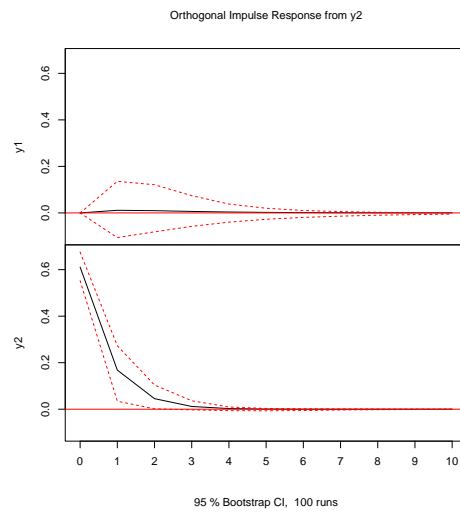


Figura 16: IRF ortogonal para y_2



No segundo caso as IRF são obtidas com a opção cumulativa. Temos os seguintes gráficos:

Figura 17: IRF cumulativa para y_1

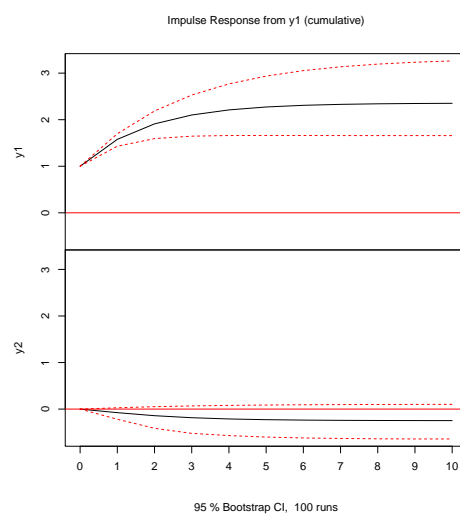
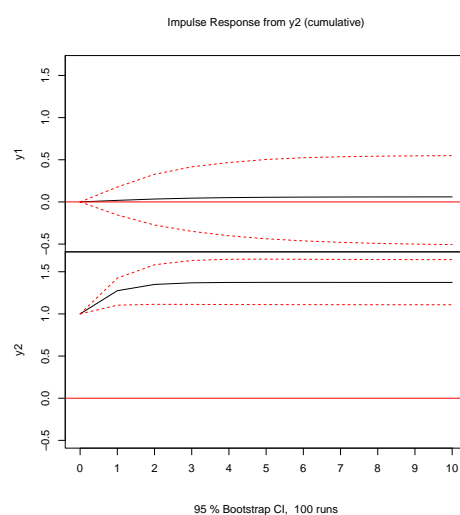


Figura 18: IRF cumulativa para y_2



No terceiro caso as IRF são obtidas com as opções ortogonal e cumulativa desligadas. Temos os seguintes gráficos:

Figura 19: IRF não ortogonal e cumulativa para y_1

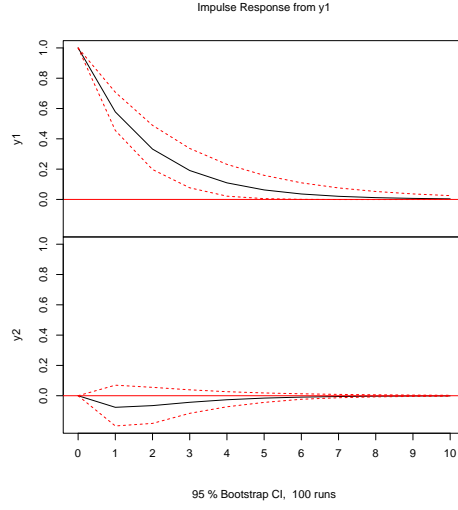
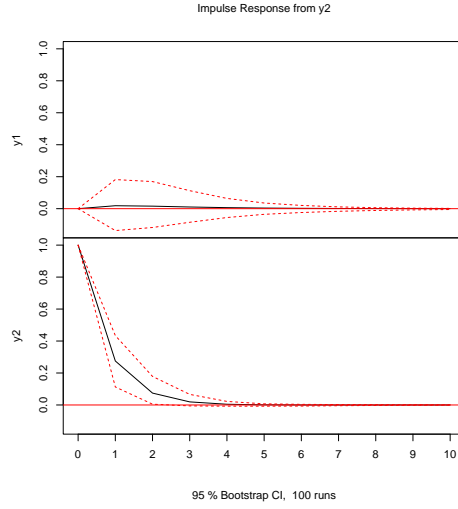


Figura 20: IRF não ortogonal e cumulativa para y_2



10 Decomposição da variância do erro de previsão

O erro de previsão l -passos à frente pode ser escrito como:

$$\mathbf{v}_{T+l} = \mathbf{y}_{T+l} - \hat{\mathbf{y}}_{T+l | T} = \sum_{j=0}^{l-1} \Psi_1^j \varepsilon_{T+l-j} \quad (72)$$

Portanto a matriz de variância-covariância do erro de previsão é dada por:

$$V(\mathbf{v}_{T+l}) = \Omega + \Psi_1 \Omega \Psi_1' + \dots + \Psi_{l-1} \Omega \Psi_{l-1}' \quad (73)$$

que pode ser reescrita da seguinte forma:

$$\begin{aligned}
V(\mathbf{v}_{T+l}) &= P^{-1}P\Omega P'(P^{-1})' + \Psi_1 P^{-1}P\Omega P'(P^{-1})'\Psi_1' + \\
&+ \dots + \Psi_{l-1} P^{-1}P\Omega P'(P^{-1})'\Psi_{l-1}' \\
&= \Theta_0\Theta_0' + \Theta_1\Theta_1' + \dots + \Theta_{l-1}\Theta_{l-1}'
\end{aligned} \tag{74}$$

Segue-se que

$$\Theta_{ij,0}^2 + \Theta_{ij,1}^2 + \dots + \Theta_{ij,l-1}^2 \tag{75}$$

representa a contribuição das inovações na j -ésima variável na explicação da variância do erro de previsão l -passos à frente para y_i com $i, j = 1, 2, \dots, n$.

Isto é chamado de **decomposição da variância do erro de previsão** (em inglês *forecast error variance decomposition - FEVD*).

O programa

Decomposição da variância - FEVD

obtem a decomposição da variância do erro de previsão para o $VAR(1)$ estacionário.

O FEVD para y_1 é apresentado na tabela abaixo:

Tabela 9: Decomposição da variância para y_2

passos à frente	y_1	y_2
1	1.0000000	0.0000000
2	0.9999053	0.0000947
3	0.9998497	0.0001503
4	0.9998264	0.0001736
5	0.9998177	0.0001823
6	0.9998147	0.0001853
7	0.9998137	0.0001863
8	0.9998134	0.0001866
9	0.9998133	0.0001867
10	0.9998132	0.0001868

O FEVD para y_2 é apresentado na tabela abaixo:

Tabela 10: Decomposição da variância para y_1

passos à frente	y_1	y_2
1	0.6083670	0.3916330
2	0.5981126	0.4018874
3	0.5969120	0.4030880
4	0.5971598	0.4028402
5	0.5973689	0.4026311
6	0.5974600	0.4025400
7	0.5974936	0.4025064
8	0.5975053	0.4024947
9	0.5975092	0.4024908
10	0.5975105	0.4024895

11 VAR Estrutural com restrições de longo prazo

Blanchard and Quah (1989) explicam como identificar $VAR's$ impondo restrições de longo prazo.

Considere o $VAR(1)$ bivariado para o crescimento do PIB e o desemprego definido por:

$$\mathbf{y}_t = A\mathbf{y}_{t-1} + Bv_t \quad (76)$$

onde v_t são os choques estruturais e $B = P^{-1}$ como antes.

A representação MA é dada por:

$$\mathbf{y}_t = (I - AL)^{-1}Bv_t \quad (77)$$

a resposta cumulativa de longo prazo de \mathbf{y} a choques em v é:

$$(I - AL)^{-1}B = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \quad (78)$$

O elemento (i, j) da matriz acima captura os efeitos de longo prazo do choque j na variável i .

A partir da matriz de variância-covariância Ω , obtemos 3 parâmetros, então, para identificar B precisamos impor uma restrição a priori em seus elementos.

Uma possibilidade é impor $b_{21} = 0$ como na decomposição de Cholesky.

Outra possibilidade é, se $\mathbf{y}_t = [\text{crescimento do PIB}_t : \text{desemprego}_t]'$, então o choque de demanda v_{1t} não tem efeito de longo prazo sobre o crescimento do PIB se

$$\pi_{11}b_{11} + \pi_{12}b_{21} = 0 \quad (79)$$

De fato essa restrição implica

$$(I - AL)^{-1}B = \begin{bmatrix} 0 & * \\ * & * \end{bmatrix} \quad (80)$$

em vez de impor $b_{21} = 0$, a restrição (79) implica que b_{11} e b_{21} estão ligados por essa relação linear.

12 Modelo de Correção de Erros Vetorial (VEC)

12.1 Introdução

Os modelos VAR discutidos até agora são apropriados para modelar dados $I(0)$, como por exemplo, retornos de ativos ou taxas de crescimento de séries temporais macroeconômicas.

A teoria econômica, no entanto, muitas vezes implica relações de equilíbrio entre os níveis das variáveis de séries temporais que são melhor descritas como sendo $I(1)$.

De forma similar, argumentos de arbitragem implicam que o logaritmo dos preços de ativos são $I(1)$ de certas séries temporais financeiras estão relacionados.

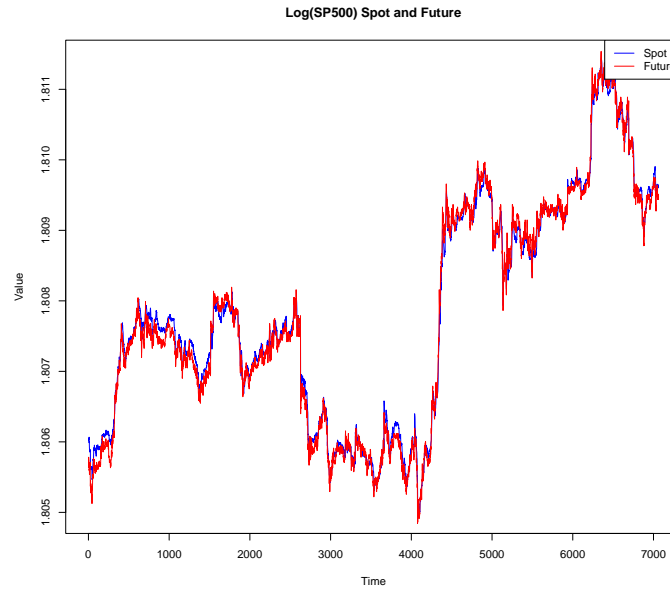
O programa abaixo

S&P500 Spot e Future

apresenta os dados à vista e futuro para o índice S&P500.

Por exemplo, o logaritmo do preço de SP500 à vista ($\log(\text{SP500_SPOT})$) e futuro ($\log(\text{SP500_FUT})$) são apresentados no gráfico abaixo

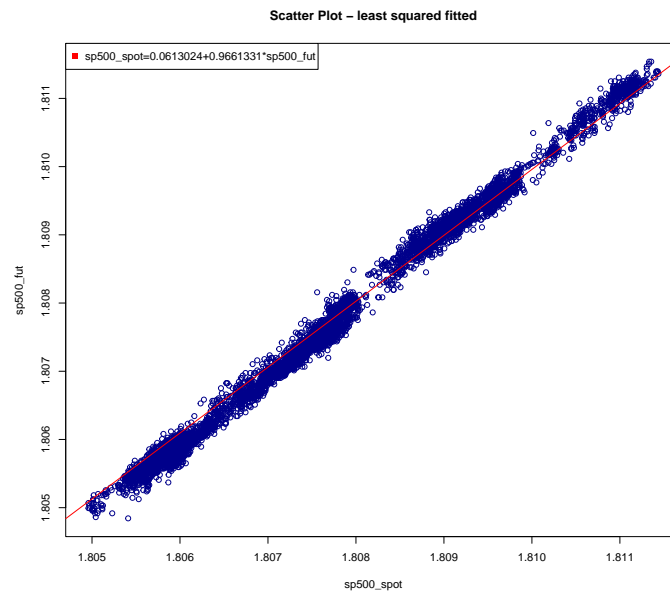
Figura 21: Log do SP500 Spot e Fut



observe que ambas as séries são $I(1)$, mas elas andam juntas.

O diagrama de dispersão abaixo mostra que a reta de regressão, linha vermelha, é um atrator dos pontos do diagrama

Figura 22: Diagrama de dispersão - Log do SP500 Spot e Fut



O conceito estatístico de cointegração é necessário para entender os modelos de regressão e os modelos VAR com dados $I(1)$. Uma vez que a regressão com variáveis $I(1)$

pode ser **espúria** .

12.2 Regressão Espúria

Considere dois processos estocástico $I(1)$ independente e que não cointegram $I(1)$. Denote este dois processos por y_{1t} e y_{2t} , que são gerados da seguinte forma:

$$\begin{aligned} y_{it} &= y_{it-1} + \varepsilon_{it} \\ \varepsilon_{it} &\sim GWN(0, 1) \quad i = 1, 2 \end{aligned} \quad (81)$$

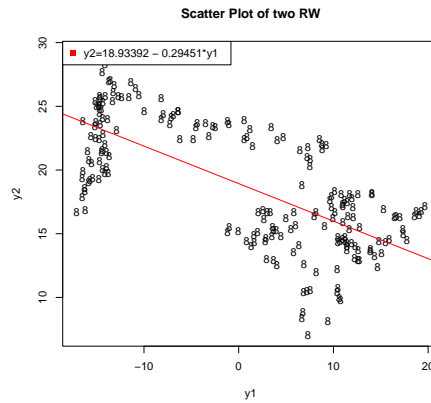
Vamos fazer a regressão de $y_{2,t}$ em constante e $y_{1,t}$ tanto em nível, que é um exemplo de regressão espúria (veja Granger and Newbold (1974)), quanto em primeira diferença, que mostrará que as duas variáveis não são relacionadas.

Temos o seguinte programa para gerar as variáveis e estimar as duas regressões

Regressão Espúria

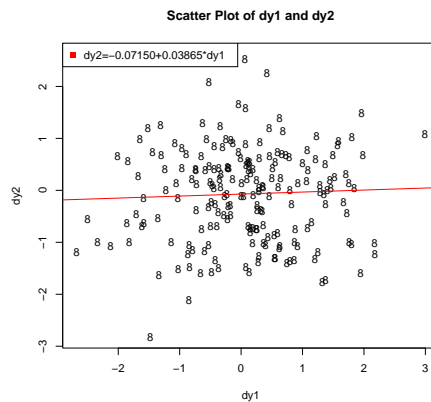
O gráfico abaixo apresenta o diagrama de dispersão de $y_{1,t}$ e $y_{2,t}$.

Figura 23: Diagrama de dispersão de $y_{1,t}$ e $y_{2,t}$



Observe que a reta de regressão não é um atrator dos pontos do diagrama. Agora o gráfico abaixo apresenta o diagrama de dispersão de $\Delta(y_{1,t})$ e $\Delta(y_{2,t})$.

Figura 24: Diagrama de dispersão de $\Delta(y_{1,t})$ e $\Delta(y_{2,t})$



Observe que a reta de regressão indica que as duas séries não estão relacionadas.

12.3 Implicações Estatísticas da Regressão Espúria

Seja $\mathbf{y}_t = (y_{1,t}, \dots, y_{n,t})'$, um vetor $(n \times 1)$ de séries temporais $I(1)$ que não são cointegradas. Escreva

$$\begin{aligned}\mathbf{y}_t &= (y_{1,t}, \mathbf{y}_{2,t})' \\ \mathbf{y}_{2,t} &= (y_{2,t}, \dots, y_{n,t})'\end{aligned}$$

e considere a regressão de $y_{1,t}$ em $\mathbf{y}_{2,t}$ dada por:

$$y_{1,t} = \hat{\beta}_2' \mathbf{y}_{2,t} + \hat{\varepsilon}_t \quad (82)$$

Como $y_{1,t}$ e $\mathbf{y}_{2,t}$ não cointegram temos que:

- (i) o valor verdadeiro de β_2 é zero;
- (ii) a regressão (82) é uma regressão espúria;
- (iii) $\hat{\varepsilon}_t$ é $I(1)$;
- (iv) $\hat{\beta}_2$ não converge em probabilidade para zero, mas converge em distribuição para uma variável aleatória não normal, não necessariamente centrada em zero;
- (v) as estatísticas t -Student usuais de MQO para testar se os elementos de β_2 são zero divergem para $\pm\infty$ à medida que $T \rightarrow \infty$. Portanto, com uma amostra grande, parecerá que $y_{1,t}$ e $\mathbf{y}_{2,t}$ cointegram quando na realidade não copintegram;
- (vi) O R^2 usual da regressão converge para a unidade à medida que $T \rightarrow \infty$, de modo que o modelo parecerá se ajustar bem, mesmo estando mal especificado.
- (vii) A regressão com dados $I(1)$ só faz sentido quando os dados são cointegrados.

12.4 Cointegração no contexto univariado

Seja $\mathbf{y}_t = (y_{1,t}, \dots, y_{n,t})'$, um vetor $(n \times 1)$ de séries temporais $I(1)$. Dizemos que os componentes de \mathbf{y}_t **cointegram** se existe um vetor, $(n \times 1)$, $\beta = (\beta_1, \dots, \beta_n)$ tal que:

$$\beta' \mathbf{y}_t = \beta_1 y_{1t} + \dots + \beta_n y_{nt} \sim I(0) \quad (83)$$

Em palavras, as séries temporais não estacionárias \mathbf{Y}_t são cointegradas se houver uma combinação linear delas que seja estacionária ou $I(0)$.

- (i) A combinação linear $\beta' \mathbf{y}_t$ é motivada pela teoria econômica e referida como relação de **equilíbrio de longo prazo**;
- (ii) séries temporais $I(1)$ com uma relação de equilíbrio de longo prazo não podem se desviar muito do equilíbrio, pois forças econômicas atuarão para restaurar a relação de equilíbrio.

O gráfico (22) mostra que $\text{Log}(SP500_Fut)$ e $\text{Log}(SP500_Spot)$ cointegram e a reta de regressão dá a relação de equilíbrio de longo prazo entre estas duas variáveis.

12.4.1 Normalização

O vetor de cointegração β não é único, pois para qualquer escalar c temos:

$$c \cdot \beta' y_t = \beta^{*'} y_t \sim I(0)$$

Alguma suposição de normalização é necessária para identificar β de forma única. Uma normalização típica é:

$$\beta = (1, -\beta_2, \dots, -\beta_n)'$$

de modo que

$$\beta' y_t = y_{1t} - \beta_2 y_{2t} - \dots - \beta_n y_{nt} \sim I(0)$$

ou

$$\begin{aligned} y_{1t} &= \beta_2 y_{2t} + \dots + \beta_n y_{nt} + \varepsilon_t \\ \varepsilon_t &\sim I(0) = \text{resíduo de cointegração} \end{aligned}$$

No equilíbrio de longo prazo $\varepsilon_t = 0$ e a relação de equilíbrio de longo prazo é:

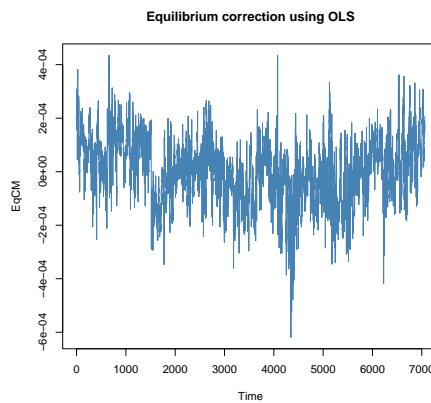
$$y_1 = \beta_2 y_2 + \dots + \beta_n y_n$$

e a correção do equilíbrio de longo prazo (em inglês *equilibrium correction model (EqCM)*), será dada por:

$$y_1 - \beta_2 y_2 - \dots - \beta_n y_n$$

No exemplo de Spot e Futuro para SP500 o EqCM é apresentado no gráfico abaixo:

Figura 25: EqCM para Spot e Futuro de SP500



12.4.2 Exemplos de Cointegração em Economia e Finanças num contexto univariado

O modelo de renda permanente implica cointegração entre consumo e renda.

Modelos de demanda de moeda implicam cointegração entre moeda, renda, preços e taxa de juros.

Modelos de teoria do crescimento implicam cointegração entre renda, consumo e investimento.

A paridade do poder de compra implica cointegração entre a taxa de câmbio nominal e os preços domésticos e estrangeiros.

A paridade coberta de taxa de juros implica cointegração entre as taxas de câmbio a termo e à vista.

A equação de Fisher implica cointegração entre a taxa de juros nominal e a inflação.

A hipótese de expectativas da estrutura a termo implica cointegração entre taxas de juros nominais de diferentes maturidades.

O modelo de valor presente dos preços das ações afirma que o preço de uma ação é o valor presente descontado esperado de seus dividendos ou lucros futuros esperados.

As relações de equilíbrio implicadas por essas teorias econômicas são chamadas de relações de equilíbrio de longo prazo, porque as forças econômicas que atuam em resposta aos desvios do equilíbrio podem levar muito tempo para restaurar o equilíbrio.

Como resultado, a cointegração é modelada usando longos períodos de dados de séries temporais de baixa frequência, medidos mensalmente, trimestralmente ou anualmente.

Em finanças, a cointegração pode ser uma relação de alta frequência ou de baixa frequência.

A cointegração em alta frequência é motivada por argumentos de arbitragem.

1. A Lei do Preço Único implica que ativos idênticos devem ser vendidos pelo mesmo preço para evitar oportunidades de arbitragem. Isso implica cointegração entre os preços do mesmo ativo negociado em diferentes mercados.
2. Argumentos de arbitragem semelhantes implicam cointegração entre preços à vista e futuros, preços à vista e a termo, e preços de compra e venda.

Aqui, a terminologia relação de equilíbrio de longo prazo é um tanto enganosa, porque as forças econômicas que atuam para eliminar as oportunidades de arbitragem trabalham muito rapidamente.

A cointegração é adequadamente modelada usando curtos períodos de dados de alta frequência em segundos, minutos, horas ou dias.

12.4.3 Cointegração e Tendências Comuns

Se a série temporal vetorial, $(n \times 1)$, \mathbf{y}_t for cointegrada com $0 < r < n$ vetores de cointegração, então há $n - r$ **tendências estocásticas comuns** que são $I(1)$.

Por exemplo seja

$$\begin{aligned}\mathbf{y}_t &= (y_{1,t}, y_{2,t})' \sim I(1) \\ \varepsilon_t &= (\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})' \sim I(0)\end{aligned}$$

e suponha que \mathbf{y}_t cointegra com vetor de cointegração $\boldsymbol{\beta} = (1, -\beta_2)'$.

Essa relação de cointegração pode ser representada por:

$$\begin{aligned}y_{1,t} &= \beta_2 \sum_{s=1}^t \varepsilon_{1,s} + \varepsilon_{3,t} \\ y_{2,t} &= \sum_{s=1}^t \varepsilon_{1,s} + \varepsilon_{2,t}\end{aligned}$$

E a tendência estocástica comum é dada por:

$$\sum_{s=1}^t \varepsilon_{1,s}$$

Observe que a relação de cointegração anula a tendência estocástica comum:

$$\begin{aligned} \beta' \mathbf{y}_t &= y_{1,t} - \beta_2 y_{2,t} \\ &= \beta_2 \sum_{s=1}^t \varepsilon_{1,s} + \varepsilon_{3,t} - \beta_2 \left(\sum_{s=1}^t \varepsilon_{1,s} + \varepsilon_{2,t} \right) \\ &= \varepsilon_{3,t} - \beta_2 \varepsilon_{2,t} \sim I(0) \end{aligned}$$

12.4.4 Alguns Sistemas Cointegrados Simulados - sistema bivariado

Considere um sistema bivariado cointegrado para $\mathbf{y}_t = (y_{1,t}, y_{2,t})'$ com vetor de cointegração $\beta = (1, -\beta_2)'$.

Por exemplo a **representação triangular** tem a seguinte forma:

$$\begin{aligned} y_{1,t} &= \beta_2 y_{2,t} + u_t, \text{ onde } u_t \sim I(0) \\ y_{2,t} &= y_{2,t-1} + \nu_t, \text{ onde } \nu_t \sim I(0) \end{aligned}$$

a primeira equação descreve a relação de equilíbrio de longo prazo com um erro de desequilíbrio u_t que é $I(0)$.

A segunda equação especifica $y_{2,t}$ como a tendência estocástica comum com inovação ν_t :

$$y_{2,t} = y_{2,0} + \sum_{j=1}^t \nu_j$$

Em geral, as inovações u_t e ν_t podem ser contemporaneamente e serialmente correlacionadas. A estrutura temporal dessas inovações caracteriza a dinâmica de curto prazo do sistema de cointegração.

O sistema com $\beta_2 = 1$, por exemplo, pode ser usado para modelar o comportamento do logaritmo de preços à vista e a termo, preços à vista e futuros, preços de ações e dividendos, ou consumo e renda.

Vamos gerar um sistema bivariado com vetor de cointegração $\beta = (1, -1)'$, com o seguinte processo gerador dos dados:

$$y_{1,t} = y_{2,t} + u_t \tag{84}$$

$$y_{2,t} = y_{2,t-1} + \nu_t \tag{85}$$

$$u_t = 0.75u_{t-1} + \varepsilon_t \tag{86}$$

$$\varepsilon_t \sim NI(0, (0.5)^2) \tag{87}$$

$$\nu_t \sim NI(0, (0.5)^2) \tag{88}$$

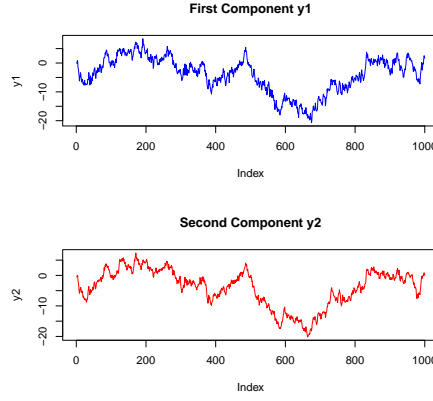
Observe que $y_{2,t}$ define a tendência comum.

Temos o seguinte programa no R para simular o sistema (84 - 88):

Simula Sistema (84 - 88)

O gráfico abaixo apresenta as duas séries

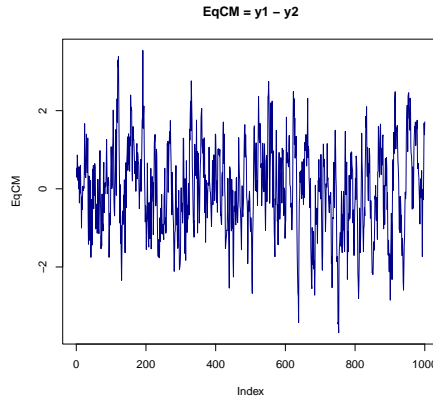
Figura 26: As duas séries do sistema (84 - 88)



observe que ambas tem tendência e parecem andar juntas.

O gráfico abaixo apresenta o $EqCM = y_{1,t} - y_{2,t}$:

Figura 27: EqCM para sistema (84 - 88)



observe que a série parece estacionária.

12.4.5 Alguns Sistemas Cointegrados Simulados - sistema trivariado

Considere um sistema trivariado cointegrado para $\mathbf{y}_t = (y_{1,t}, y_{2,t}, y_{3,t})'$ com vetor de cointegração $\boldsymbol{\beta} = (1, -0.5, -0.5)'$.

Por exemplo a **representação triangular** tem a seguinte forma:

$$y_{1,t} = 0.5 * y_{2,t} + 0.5 * y_{2,t} + 0.5 * y_{3,t} + u_t \quad (89)$$

$$u_t = 0.75 * u_{t-1} + \varepsilon_t \text{ onde } \varepsilon_t \sim NI(0, (0.5)^2) \quad (90)$$

$$y_{2,t} = y_{2,t-1} + \nu_t, \text{ onde } \nu_t \sim NI(0, (0.5)^2) \quad (91)$$

$$y_{3,t} = y_{3,t-1} + \omega_t, \text{ onde } \omega_t \sim NI(0, (0.5)^2) \quad (92)$$

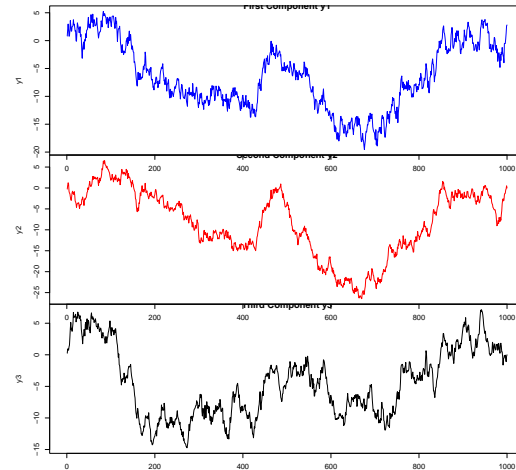
Observe que $y_{2,t}$ e $y_{3,t}$ são as tendências comuns.

Temos o seguinte programa no R para simular o sistema (89-92):

Simula Sistema (89-92)

O gráfico abaixo apresenta as três séries

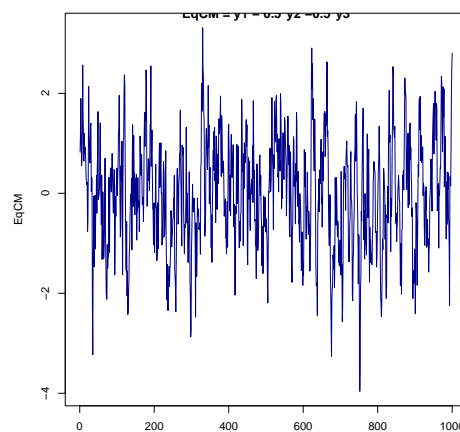
Figura 28: As três séries do sistema (89-92)



observe que as três séries tem tendência.

O gráfico abaixo apresenta o $EqCM = y_{1,t} - 0.5 * y_{2,t} - 0.5 * y_{3,t}$:

Figura 29: EqCM para sistema (89-92)



observe que a série parece estacionária.

12.4.6 Cointegração e Modelo de Correção do Equilíbrio - EqCM

Considere um vetor bivariado $I(1) \mathbf{y}_t = (y_{1,t}, y_{2,t})'$ e suponha que \mathbf{y}_t seja cointegrado com o vetor de cointegração $\boldsymbol{\beta} = (1, -\beta_2)'$, de forma que $\boldsymbol{\beta}'\mathbf{y}_t = y_{1,t} - \beta_2 y_{2,t}$ seja $I(0)$. Engle and Granger (1987) mostraram que a cointegração implica a existência de um **Modelo de Correção do Equilíbrio, em inglês *Equilibrium correction model (EqCM)*** da seguinte forma:

$$\begin{aligned}\Delta y_{1,t} &= c_1 + \alpha_1(y_{1,t-1} - \beta_2 y_{2,t-1}) + \sum_j \psi_{11}^j \Delta y_{1,t-j} + \sum_j \psi_{12}^j \Delta y_{2,t-j} + \varepsilon_{1,t} \\ \Delta y_{2,t} &= c_2 + \alpha_2(y_{1,t-1} - \beta_2 y_{2,t-1}) + \sum_j \psi_{21}^j \Delta y_{1,t-j} + \sum_j \psi_{22}^j \Delta y_{2,t-j} + \varepsilon_{2,t}\end{aligned}$$

O modelo de correção do equilíbrio(EqCM) vincula a relação de equilíbrio de longo prazo, implícita pela cointegração, ao mecanismo de ajuste dinâmico de curto prazo que descreve como as variáveis reagem quando se afastam do equilíbrio de longo prazo.

12.4.7 Testes de Cointegração no contexto univariado

Os testes de cointegração cobrem duas situações:

Há no máximo um vetor de cointegração

- (i) Considerado por Engle and Granger (1987). Eles desenvolveram um procedimento simples de teste em duas etapas baseado em resíduos usando técnicas de regressão;
- (ii) Possivelmente há $0 \leq r < n$ vetores de cointegração Considerado por Johansen (1988). Ele desenvolveu um procedimento sequencial para determinar a existência de cointegração e para determinar o numero de relações de cointegração com base em técnicas de máxima verossimilhança.

Primeiramente iremos considerar só o item (i) e posteriormente na seção (12.5) iremos apresentar o item (ii).

12.4.8 Testes de Cointegração baseado nos resíduos

O procedimento em duas etapas de Engle and Granger (1987) para determinar se o vetor $\boldsymbol{\beta}$, $(n \times 1)$, é um vetor de cointegração é o seguinte:

1. Fazer a regressão estática de $y_{1,t}$ em constante e $y_{2,t}, \dots, y_{n,t}$. Formar o resíduo de cointegração, isto é, $\hat{u}_t = \hat{\boldsymbol{\beta}}'\mathbf{y}_t = y_{1,t} - \hat{\beta}_1 - \hat{\beta}_2 y_{2,t} - \dots - \hat{\beta}_n y_{n,t}$;
2. Realizar um teste de raiz unitária em \hat{u}_t para determinar se ele é $I(0)$, onde os valores críticos são dados por MacKinnon (1991);
3. A hipótese nula no procedimento de Engle-Granger é de não-cointegração e a alternativa é de cointegração.

Existem dois casos a considerar:

- (a) O vetor de cointegração proposto, β , é pré-especificado (não estimado). Por exemplo, a teoria econômica pode implicar valores específicos para os elementos em β , como $\beta = (1, -1)'$. O resíduo de cointegração é então prontamente construído usando o vetor de cointegração pré-especificado;
- (b) O vetor de cointegração proposto é estimado a partir dos dados e uma estimativa do resíduo de cointegração $\hat{u}_t = \hat{\beta}'\mathbf{y}_t$ é formada.

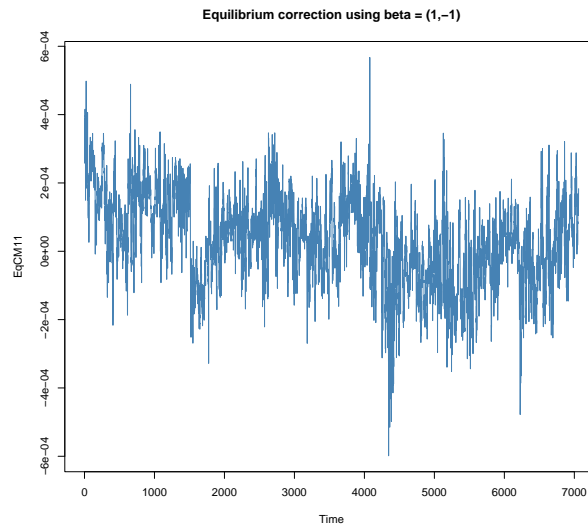
O programa abaixo

EqCM com $\beta = (1, -1)$

gera um EqCM com $\beta = (1, -1)$.

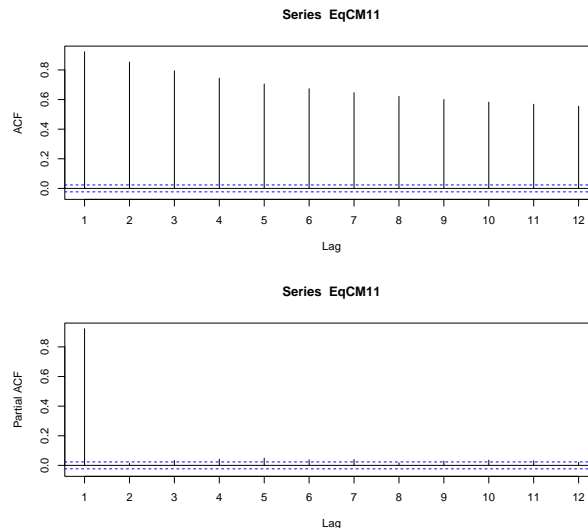
O gráfico abaixo apresenta este EqCM com $\beta = (1, -1)$:

Figura 30: EqCM com $\beta = (1, -1)$



A seguir são apresentados o ACF e PACF até defasagens de ordem 12. Temos os seguinte gráficos:

Figura 31: Acf e Pacf para EqCM com $\beta = (1, -1)$



Observe que embora a primeira autocorrelação seja alta (0.9231) o decaindo indica uma série estacionária.

Podemos comprovar através do teste de raiz unitária de Dickey & Fuller. Temos os seguintes resultados:

Tabela 11: ADF teste para EqCM com $\beta = (1, -1)$

ADF Teste		
$ADF = -9.1017$	Lag order = 19	$p - value = 0.01$

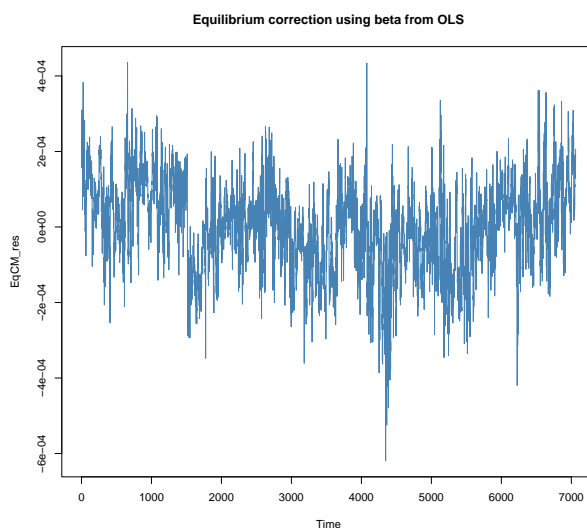
Agora vamos estimar o EqCM através do resíduo da regressão estática de sp500_spot em constante e sp500_fut. Temos os seguintes resultados:

Tabela 12: EqCM como resíduo da regressão estática

Coefficiente	Estimate	Std Error	t-value	p-value
constante	0.0613	0.0015	39.86	0.0000
sp500_fut	0.9661	0.0009	1135.54	0.0000
$\hat{\sigma} = 0.0001$				
$R^2 = 0.9946$		$\overline{R}^2 = 0.9946$		
$F(1, 6059) = 1.289 * 10^6$		$p - value = 0.000$		

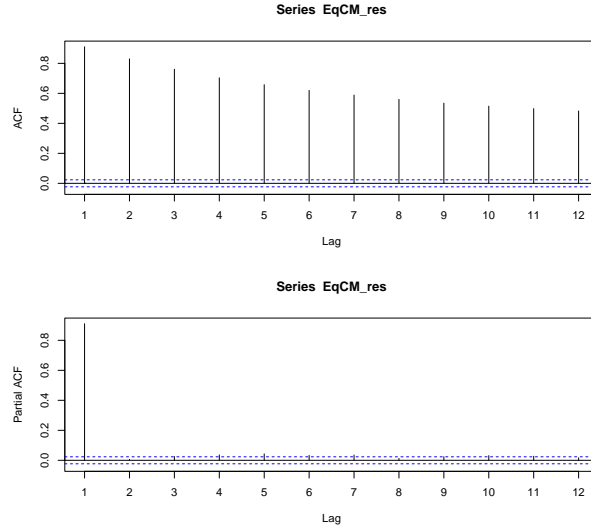
O gráfico abaixo apresenrta este EqCM com β estimado por M.Q.O.

Figura 32: EqCM com β estimado por M.Q.O.



A seguir são apresentados o ACF e PACF até defasagens de ordem 12. Temos os seguinte gráficos:

Figura 33: ACF e PACF para EqCM com β estimado por M.Q.O.



Podemos comprovar através do teste de raiz unitária de Dickey & Fuller. Temos os seguintes resultados:

Tabela 13: ADF teste para EqCM com β estimado por M.Q.O.

ADF Teste		
$ADF = -9.0205$	Lag order = 19	$p - value = 0.01$

12.4.9 Estimando o Modelo de Correção do Equilíbrio usando M.Q.O.

Considere o vetor bivariado $\mathbf{y}_t = (y_{1,t}, y_{2,t})'$, cujo componentes são $I(1)$, e suponha que \mathbf{y}_t seja cointegrado com o vetor de cointegração $\boldsymbol{\beta} = (1, -\beta_2)'$, de forma que $\boldsymbol{\beta}'\mathbf{y}_t = y_{1,t} - \beta_2 y_{2,t}$ seja $I(0)$.

Suponha que se tenha uma estimativa $\hat{\beta}_2$ (por Mínimos Quadrados Ordinários) do coeficiente de cointegração e que se esteja interessado em estimar o modelo de correção de equilíbrio correspondente para $\Delta y_{1,t}$ e $\Delta y_{2,t}$ usando:

$$\Delta y_{1,t} = c_1 + \alpha_1(y_{1,t-1} - \hat{\beta}_2 y_{2,t-1}) + \sum_j \psi_{11}^j \Delta y_{1,t-j} + \sum_j \psi_{12}^j \Delta y_{2,t-j} + \varepsilon_{1,t} \quad (93)$$

$$\Delta y_{2,t} = c_2 + \alpha_2(y_{1,t-1} - \hat{\beta}_2 y_{2,t-1}) + \sum_j \psi_{21}^j \Delta y_{1,t-j} + \sum_j \psi_{22}^j \Delta y_{2,t-j} + \varepsilon_{2,t} \quad (94)$$

Porque $\hat{\beta}_2$ é super consistente, ele pode ser tratado como conhecido no EqCM, de forma que o erro de desequilíbrio estimado $y_{1,t} - \hat{\beta}_2 y_{2,t}$ possa ser tratado como o desequilíbrio conhecido $y_{1,t} - \beta_2 y_{2,t}$.

Como todas as variáveis no modelo EqCM são $I(0)$, as duas equações de regressão podem ser consistentemente estimadas usando Mínimos Quadrados Ordinários (OLS).

Alternativamente, o sistema EqCM pode ser estimado por regressões aparentemente não relacionadas (SUR) para aumentar a eficiência, caso o número de defasagens nas duas equações seja diferente.

Voltando para o exemplo do Log do S&P500 Spot e Fut o programa abaixo estima o sistema (93-94) por M.Q.O. equação por equação:

S&P500 Spot e Future

temos os seguintes resultados assumindo que o número de lags é um:

Tabela 14: EqCM para (93)

$\Delta \log(sp500_spot)_t = c_1 + \alpha_1 EqCM_res_{t-1} + \psi_{11} \Delta \log(sp500_spot)_{t-1} + \psi_{12} \Delta \log(sp500_fut)_{t-1} + \epsilon_{1t}$				
Coefficients	Estimate	Std Error	t-value	p-value
c_1	$4.652e-07$	$3.699e-07$	1.258	0.209
α_1	$-8.581e-02$	$3.418e-03$	-25.109	0.0000
ψ_{11}	$8.463e-02$	$1.227e-02$	6.896	0.0000
ψ_{12}	$-9.596e-03$	$7.968e-03$	-1.204	0.228
$\hat{\sigma} = 3.107e-05$				
$R^2 = 0.1039$		$\bar{R}^2 = 0.1035$		
$F(3, 7055) = 272.6$		$p - value = 0.0000$		

Tabela 15: EqCM para (94)

$\Delta \log(sp500_fut)_t = c_2 + \alpha_2 EqCM_res_{t-1} + \psi_{21} \Delta \log(sp500_spot)_{t-1} + \psi_{22} \Delta \log(sp500_fut)_{t-1} + \epsilon_{2t}$				
Coefficients	Estimate	Std Error	t-value	p-value
c_2	$5.508e-07$	$6.244e-07$	0.882	0.37770
α_2	$-5.460e-03$	$5.769e-03$	-0.946	0.34394
ψ_{21}	$17.905e-03$	$2.071e-02$	0.382	0.70275
ψ_{22}	$-3.889e-02$	$1.345e-02$	-2.891	0.00385
$\hat{\sigma} = 5.245e-05$				
$R^2 = 0.001247$		$\bar{R}^2 = 0.0008226$		
$F(3, 7055) = 2.937$		$p - value = 2.937$		

12.5 Os modelos VAR e Cointegração

12.5.1 Introdução

O teorema de representação de Granger (veja Granger (1983) liga a cointegração aos modelos de correção de erros (MCE), embora os modelos MCE apareceram pela primeira vez no artigo Sargan (1964).

Em uma série de artigos importantes e em um excelente livro didático (veja Johansen (1995)), Soren Johansen firmemente fundamenta a cointegração e os modelos de correção de erros em uma estrutura de vetor autorregressivo.

Esta seção descreve a abordagem de Johansen para a modelagem de cointegração.

12.5.2 VAR e cointegração

Considere um $VAR(p)$ em nível para um vetor \mathbf{y}_t ($n \times 1$) dado por:

$$\mathbf{y}_t = \mathbf{m} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \dots + \Phi_p \mathbf{y}_{t-p} + \epsilon_t \quad (95)$$

onde \mathbf{m} são termos determinísticos, que podem ser só a constante ou tendência linear.

Temos que o $VAR(p)$ é estável se

$$\det(\mathbf{I}_n - \Phi_1 z - \dots - \Phi_p z^p) = 0 \quad (96)$$

se todas as raízes estão fora do círculo unitário.

Se houver raízes no círculo unitário, então algumas ou todas as variáveis em \mathbf{y}_t são $I(1)$ e elas também podem ser cointegradas.

Se \mathbf{y}_t for cointegrado, então a representação VAR não é a mais adequada para análise, pois as relações de cointegração não estão explicitadas.

As relações de cointegração tornam-se explicitadas se a VAR em níveis for transformada no modelo de correção de erros vetorial (em inglês *Vector Error Correction Model (VECM)*) ou modelo de correção do equilíbrio vetorial (em inglês *Vector Equilibrium Correction Model (VEqCM)*).

Usando as identidades (36) e (37) podemos escrever (95) da seguinte forma:

$$\begin{aligned} \Delta \mathbf{y}_t &= \mathbf{m} + \Pi \mathbf{y}_{t-1} + \Gamma_1 \Delta \mathbf{y}_{t-1} \\ &+ \dots + \Gamma_{p-1} \Delta \mathbf{y}_{t-p+1} + \varepsilon_t \\ \Pi &= \Phi_1 + \dots + \Phi_p - \mathbf{I}_n \\ \Gamma_k &= - \sum_{j=k+1}^p \Phi_j, k = 1, \dots, p-1 \end{aligned} \quad (97)$$

Na representação VEqCM, $\Delta \mathbf{y}_t$ e todas as suas defasagens são $I(0)$.

O termo $\Pi \mathbf{y}_{t-1}$ é o único que inclui variáveis potencialmente $I(1)$. Mas para que este seja um regressor válido, isto porque todas as outras variáveis são $I(0)$, devemos ter que $\Pi \mathbf{y}_{t-1}$ tem que ser também $I(0)$. Portanto, $\Pi \mathbf{y}_{t-1}$ deve conter as relações de cointegração, se existirem.

Se o $VAR(p)$ tem uma raiz unitária ($z = 1$), então:

$$\begin{aligned} \det(\mathbf{I}_n - \Phi_1 z - \dots - \Phi_p z^p) &= 0 \\ \Rightarrow \det(\Pi) &= 0 \\ \Rightarrow \Pi &\text{ é singular} \end{aligned} \quad (98)$$

Se Π é singular então $\Pi(\mathbf{1}) = \mathbf{0}$ e portanto a matriz $\Pi(\mathbf{1})$ terá posto reduzido, e $\text{posto}(\Pi) = r < n$.

Existem três casos a considerar:

1. Se $\text{posto}(\Pi) = 0$, então:

$$\Pi = \mathbf{0},$$

o que implica que todos os componentes de \mathbf{y}_t são $I(1)$ e não cointegram. O $VEqCM$ se reduz a um $VAR(p-1)$ em primeira diferença, isto é:

$$\Delta \mathbf{y}_t = \mathbf{m} + \Gamma_1 \Delta \mathbf{y}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{y}_{t-p+1} + \varepsilon_t.$$

2. Se $\text{posto}(\Pi) = n$, então, tem posto completo, o que implica que todos os componentes de \mathbf{y}_t são pelo menos $I(0)$. O $VEqCM$ se reduz a um $VAR(p)$ em nível, isto é, equação (95).

3. Se $0 < \text{posto}(\mathbf{\Pi}) = r < n$, então, tem posto reduzido, o que implica que os componentes de todos os componentes de \mathbf{y}_t são pelo menos $I(1)$, com r vetores de cointegração linearmente independentes e $n - r$ tendência estocásticas comuns (raízes unitárias). Como $\mathbf{\Pi}$ tem posto igual a r , como vimos na seção 6 existem matrizes α e β de ordem $n \times r$ tais que:

$$\mathbf{\Pi}_{(n \times n)} = \alpha_{(n \times r)} \beta'_{(n \times r)}$$

onde α e β são matrizes $(n \times r)$ ambas com posto r . As linhas de β' formam uma base para os r vetores de cointegração e os elementos de α forma a matriz de cargas ou matriz de coeficientes de ajustamento, que dão a importância das relações de cointegração em cada equação. Neste caso o *VEqCM* pode ser escrito da seguinte forma:

$$\begin{aligned} \Delta \mathbf{y}_t &= \mathbf{m} + \alpha \beta' \mathbf{y}_{t-1} + \mathbf{\Gamma}_1 \Delta \mathbf{y}_{t-1} \\ &+ \dots + \mathbf{\Gamma}_{p-1} \Delta \mathbf{y}_{t-p+1} + \varepsilon_t \end{aligned}$$

onde $\beta' \mathbf{y}_{t-1} \sim I(0)$ dado que β' é a matriz dos vetores de cointegração.

12.5.3 Não unicidade da fatorização da matriz de longo-prazo, $\mathbf{\Pi}$

É importante reconhecer que a fatoração $\mathbf{\Pi} = \alpha \beta'$ não é única, pois para qualquer matriz H não singular de dimensão $r \times r$, temos:

$$\begin{aligned} \alpha \beta' &= \alpha H H^{-1} \beta' = (\alpha H)(\beta H^{-1})' = \alpha^* \beta'^* \\ \alpha^* &= \alpha H, \beta^* = \beta H^{-1} \end{aligned}$$

Portanto, a fatoração $\mathbf{\Pi} = \alpha \beta'$ apenas identifica o espaço gerado pelas relações de cointegração.

Para obter valores únicos de α e β' são necessárias restrições adicionais no modelo (veja Johansen (1995)).

12.5.4 Alguns exemplos

12.6 Teste de Johansen

Vamos supor que o posto de $\mathbf{\Pi}(1) = \mathbf{r} < \mathbf{n}$. Como vimos na seção 6 existem matrizes α e β de ordem $n \times r$ tais que:

$$\mathbf{\Pi}(1) = \alpha \beta' \tag{99}$$

então (97) pode ser reescrito da seguinte forma:

$$\Delta \mathbf{y}_t = \mathbf{m} - \mathbf{\Phi}_2 \Delta \mathbf{y}_{t-1} - \alpha \beta' \mathbf{y}_{t-1} + \varepsilon_t \tag{100}$$

Como estamos assumindo que os componentes de \mathbf{y}_t são $I(1)$ e em geral combinação linear de variáveis que tem tendência estocástica também terão tendência estocástica, $\beta' \mathbf{y}_{t-1}$ em princípio seria também $I(1)$.

Observe que neste caso (100) não seria uma representação válida uma vez que a variável dependente é $I(0)$ assim como $\Delta \mathbf{y}_{t-1}$.

A única possibilidade para que (100) estivesse balanceada seria $\alpha = \mathbf{0}$, mas este caso não é possível uma vez que o posto de $\Pi(\mathbf{1}) > \mathbf{0}$.

Deste modo devemos ter $\beta' \mathbf{y}_{t-1} \sim I(0)$, isto é os componentes de \mathbf{y}_t co-integram.

Este termo é chamado de termo de correção do equilíbrio e a matriz β é composta pelos vetores de co-integração.

Por outro lado a matriz α , é chamada de matriz de cargas, e, dá a importância dos vetores de co-integração em cada uma das equações.

Observe que testar cointegração entre os componentes de \mathbf{y}_t , corresponde a testar se o posto da matriz $\Pi(\mathbf{1})$ é reduzido e igual a r o número de vetores de co-integração.

Este teste é conhecido na literatura como teste de Johansen. Ele sugere isolar $\Pi(\mathbf{1})$ em (97) através de duas regressões auxiliares, a saber:

1. faça a regressão de $\Delta \mathbf{y}_t$ em \mathbf{m} e $\Delta \mathbf{y}_{t-1}(\Delta \mathbf{y}_{t-2}, \dots, \Delta \mathbf{y}_{t-k})$ e obtenha o resíduo denotado por \mathbf{R}_{0t}
2. faça a regressão de \mathbf{y}_{t-1} em \mathbf{m} e $\Delta \mathbf{y}_{t-1}(\Delta \mathbf{y}_{t-2}, \dots, \Delta \mathbf{y}_{t-k})$ e obtenha o resíduo denotado por \mathbf{R}_{1t}

Agora (97) é equivalente a:

$$\mathbf{R}_{0t} = -\Pi(\mathbf{1})\mathbf{R}_{1t} + \varepsilon_t \quad (101)$$

ou se usarmos (100) teremos então:

$$\mathbf{R}_{0t} = -\alpha \beta' \mathbf{R}_{1t} + \varepsilon_t \quad (102)$$

Maximizando-se a verossimilhança e assumindo-se que β é conhecido temos que o estimador de α será dado por:

$$\hat{\alpha}(\beta) = -\mathbf{S}_{01}\beta(\beta'\mathbf{S}_{11}\beta)^{-1} \quad (103)$$

e o estimador de Ω será dado por:

$$\hat{\Omega}(\beta) = \mathbf{S}_{00} - \mathbf{S}_{01}\beta(\beta'\mathbf{S}_{11}\beta)^{-1}\beta'\mathbf{S}_{10} \quad (104)$$

onde

$$\mathbf{S}_{ij} = \frac{1}{T} \sum_{t=1}^T \mathbf{R}_{it}\mathbf{R}_{jt}' \quad \text{para } i = j = 0, 1 \quad (105)$$

Observe que como β em (102) é assumido conhecido, o modelo é linear e o estimador de α é obtido pela regressão de \mathbf{R}_{0t} em $\beta'\mathbf{R}_{1t}$, cujo estimador é dado pela covariância entre \mathbf{R}_{0t} e $\beta'\mathbf{R}_{1t}$, vezes a inversa da variância de $\beta'\mathbf{R}_{1t}$ que é exatamente a expressão (103).

Por outro lado o estimador da matriz de variância residual é dado por:

$$\begin{aligned}
\widehat{\Omega}(\beta) &= (1/T) \sum_{t=1}^T \widehat{\varepsilon}_t \widehat{\varepsilon}_t' \\
&= (1/T) \sum_{t=1}^T (\mathbf{R}_{0t} - \mathbf{S}_{01}\beta(\beta'\mathbf{S}_{11}\beta)^{-1} \beta'\mathbf{R}_{1t})(\mathbf{R}_{0t} - \mathbf{S}_{01}\beta(\beta'\mathbf{S}_{11}\beta)^{-1} \beta'\mathbf{R}_{1t})' \\
&= (1/T) \sum_{t=1}^T (\mathbf{R}_{0t}\mathbf{R}_{0t}' - \mathbf{R}_{0t}\mathbf{R}_{1t}'\beta(\beta'\mathbf{S}_{11}\beta)^{-1} \beta'\mathbf{S}_{10}) \\
&= \mathbf{S}_{00} - \mathbf{S}_{01}\beta(\beta'\mathbf{S}_{11}\beta)^{-1} \beta'\mathbf{S}_{10}
\end{aligned}$$

Johansen mostrou que maximizar a verossimilhança em relação a β é equivalente a minimização de:

$$\frac{|\beta'\mathbf{S}_{11}\beta - \beta'\mathbf{S}_{10}\mathbf{S}_{00}\mathbf{S}_{01}\beta|}{|\beta'\mathbf{S}_{11}\beta|} \quad (106)$$

cuja solução é dada pela obtenção dos autovalores de :

$$(\lambda\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}\mathbf{S}_{01}\beta) = 0 \quad (107)$$

Temos então se $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$ são os r maiores autovalores então $\beta = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r)$ com \mathbf{v}_i autovetor associado ao autovalor λ_i .

Neste caso a verossimilhança é proporcional a

$$\sum_{i=1}^r \ln(1 - \lambda_i) \quad (108)$$

Johansen propõe duas estatísticas de teste:

1. Maior autovalor λ_{\max}
2. Traço $\lambda_{traco} = \sum_{i=1}^r \ln(1 - \lambda_i)$

Pode-se mostrar, por exemplo que a estatística λ_{traco} tem distribuição $\chi^2(2m^2)$ onde $m = T - r$.

Observe que foi assumido que o vetor de constantes \mathbf{m} não entrava na relação de cointegração. Caso seja necessário que a constante faça parte do vetor de cointegração o procedimento pode ser modificado.

12.7 Um Exemplo de Determinação da Dimensão do Espaço de Cointegração - Dados Gerados pelo programa abaixo

Usando os dados gerados pelo programa

Segunda e Terceira componente do VAR(1) Trivariado

vamos considerar o vetor $\mathbf{y}_t = (y_{2t}, y_{3t})$.

Vamos determinar a defasagem ótima. Temos então o seguinte resultado:

Tabela 16: Seleção da defasagem ótima

	1	2	3	4	5	6	7	8	9	10
AIC(n)	0.9350	0.8712	0.9015	0.9055	0.9121	0.9200	0.9416	0.9568	0.9879	1.0142
HQ(n)	0.9766	0.9404	0.9985	1.0301	1.0644	1.1000	1.1493	1.1922	1.2509	1.3049
SC(n)	1.0376	1.0421	1.1408	1.2131	1.2881	1.3643	1.4543	1.5379	1.6373	1.7320
FPE(n)	2.5473	2.3898	2.4636	2.4735	2.4902	2.5103	2.5658	2.6059	2.6891	2.7621

então temos que AIC, HQ e FPE escolhem ordem 2 e SC ordem 1.

Temos que usar 2 ou 1 defasagens no teste de Johansen.

O pacote que faz o teste de Johansen no R, **urca**, não permite que o número de defasagens no VEC seja menor do que 2.

Vamos usar o CATS no OxMetrics que permite os seguintes casos:

1. Dados não tem tendência determinística - temos as duas primeiras possibilidades
 - tanto a constante quanto uma tendência determinística estão ausentes na relação de cointegração e no VAR - Caso 1.
 - a constante, mas não a tendência determinística, está presente na relação de cointegração mas não no VAR - Caso 2.
2. Dados tem tendência determinística - temos o terceiro e quarto caso
 - a constante mas não a tendência determinística na relação de cointegração e sem constante e tendência determinística no VAR^6 - Caso 3.
 - sem constante mas com tendência determinística na relação de cointegração e constante no VAR - Caso 4.

O caso que não é contemplado no CATS é aquele em que os dados tem tendência quadrática

O seguinte programa no CATS determina a dimensão do espaço de cointegração para os quatro casos acima usando primeiro duas defasagens e em seguida uma defasagem para o VAR .

Dimensão do espaço de cointegração para todos os Casos

Temos os seguintes resultados

⁶Observe que num passeio aleatório com drift (constante) quando se integra o processo o drift passa a ser uma tendência determinística. Isto é um caso especial de: se y_t é um polinômio de ordem p , então $\Delta^d y_t$ é um polinômio de ordem $p - d$.

Tabela 17: Dimensão do espaço de cointegração para os Casos

Model		I(1) Analysis				
	Lags	r	Eigenvalues	Trace	Crit 5%	p-value
Case 1	2	0	0.3503	85.84	12.28	0.0000
		1	0.0001	0.0200	4.07	0.9260
Case 1	1	0	0.5722	196.06	12.28	0.0000
		1	0.0005	0.1000	4.07	0.8140
Case 2	2	0	0.3588	89.66	20.16	0.0000
		1	0.0062	1.24	9.14	0.9040
Case 2	1	0	0.5755	171.91	20.16	0.0000
		1	0.0070	1.4100	9.14	0.8780
Case 3	2	0	0.3578	88.13	15.41	0.0000
		1	0.0001	0.010	3.84	0.9180
Case 3	1	0	0.5706	168.21	15.41	0.0000
		1	0.0000	0.0000	3.84	0.9580
Case 4	2	0	0.3655	101.54	25.73	0.0000
		1	0.0539	11.02	12.45	0.0880
Case 4	1	0	0.5846	186.90	25.73	0.0000
		1	0.0588	12.06	12.45	0.0580

Para todos os casos a dimensão do espaço de integração é um.

Impondo que a dimensão do espaço de cointegração vamos escolher o melhor modelo usando o seguinte batch

Critérios de Informação para todos os Casos

Temos então a seguinte tabela:

Tabela 18: Critérios de Informação para todos os Casos

Model	T	p	Lags	Method	log-Lik	BIC	HQ	AIC
Case 1	199	7	2	RRR	-643.1676	6.6502	6.5812	6.5343
Case 1	199	3	1	RRR	-653.6498	6.6491	6.6196	6.5995
Case 2	199	8	2	RRR	-641.8645	6.6637	6.5849	6.5313
Case 2	199	4	1	RRR	-652.8715	6.6679	6.6285	6.6017
Case 3	199	9	2	RRR	-641.2512	6.6841	6.5955	6.5352
Case 3	199	5	1	RRR	-652.1700002	6.6875	6.6382	6.6047
Case 4	199	10	2	RRR	-640.0526	6.6987	6.6002	6.5332
Case 4	199	6	1	RRR	-648.8499	6.6807	6.6216	6.5814

Pelos critérios de BIC e HQ impondo uma relação de cointegração, o melhor modelo é o modelo 4 com uma defasagem no VAR para o critério BIC e com duas defasagens para o critério de HQ. Pelo critério de AIC também impondo uma relação de cointegração, o melhor modelo é o modelo 2 mas com duas defasagens no VAR .

Estimando o VEC para o melhor modelo pelo critério BIC, mas se este modelo for estimado o teste de autocorelação apresenta problemas. Deste modo optou-se por estimar o melhor modelo pelo critério HQ que não apresenta problemas de especificação. Obtemos os seguintes resultados:⁷

Tabela 19: VEC para o melhor modelo

t-statistics in in ()	p-values in []	
Cointegrating Eq	CointEq1	
$y_3(-1)$	1.000000	
$y_2(-1)$	-1.3600 (-265.4) [0.9123]	
Error Correction:	$D(y_2)$	$D(y_3)$
CointEq1	0.1200 (1.3) [0.4523]	-0.5870 (-6.90) [0.7211]

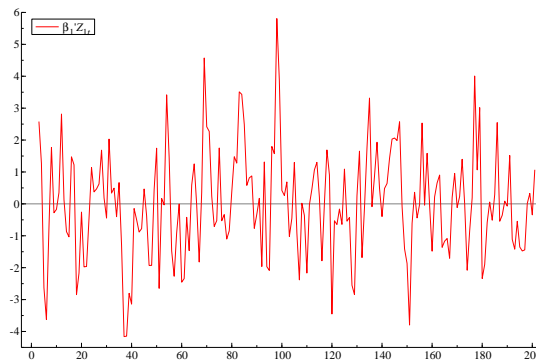
⁷Embora a ordenação seja primeiro y_2 e depois y_3 , portanto a normalização no vetor de cointegração seria 1 para y_2 e irrestrito para y_3 , o CATS força que α seja negativo. Por isto a normalização neste caso fica 1 para y_3 e irrestrito para y_2 .

Tabela 20: Bondade de Ajuste e Testes de Especificação para Tabela (19)

Determinant Residual Covariance (d.f. adjusted)	0.7882
Log Likelihood (d.f. adjusted)	-643.1676
Akaike Information Criteria	6.5343
Schwarz Criteria	6.6501
Test for autocorrelation LM(1):	$\chi^2(4) = 3.0258[0.5535]$
Test for autocorrelation LM(2):	$\chi^2(4) = 6.4943[0.1651]$
Test for normality:	$\chi^2(4) = 5.0845[0.2787]$
Test for ARCH LM(1):	$\chi^2(9) = 0.7305[0.9998]$
Test for ARCH LM(2):	$\chi^2(18) = 16.2055[0.5782]$

O gráfico abaixo apresenta a relação de co-integração entre as duas variáveis

Figura 34: Relação de Cointegração $\beta' y_{t-1}$



12.8 Exercício

1. Considere as séries y_1 e y_3 gerados pelo programa

Segunda e Terceira componente do VAR(1) Trivariado

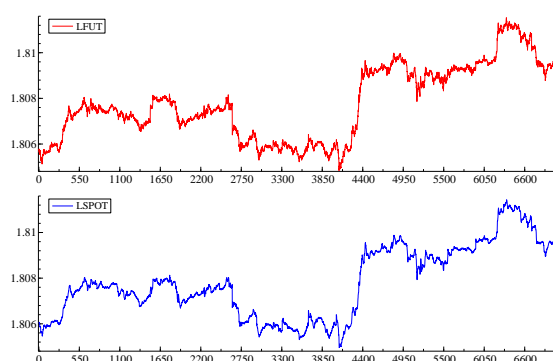
- (a) Obtenha a defasagem ótima para o VAR
- (b) Verifique se estas séries co-integram
- (c) Estime o melhor VEC para estes dados.
- (d) Você acha que o modelo é adequado? Porque?

2. No arquivo PETR4_0024.XLS temos dados de Petrobrás PN para o período de 3/01/2000 até 02/01/2024. Usando os dados de Máximo e Mínimo.
- (a) Como os dados tem observações faltando use um modelo de Nível Local para estimar estas observações.
 - (b) Obtenha a defasagem ótima para o VAR
 - (c) Verifique se estas séries co-integram
 - (d) Estime o melhor VEC para estes dados.
 - (e) Voce acha que o modelo é adequado? Porque?

12.9 Um Exemplo de Determinação da Dimensão do Espaço de Cointegração - Relação entre mercado spot e futuro para o S&P500 dados intradiária de Maio de 1996.

No arquivo S&P500 temos os dados intradiários para o valor spot e futuro do S&P500 em Maio de 1996. Os gráficos abaixo apresentam estas séries após serem transformadas usando logaritmo.

Figura 35: Log do Spot e do Futuro do S&P500 intradiário Maio de 1996



Observe que ambas as séries são não estacionária e parecem andar juntas.

Estimamos os casos 1 - 4, com 20 defasagens que foi a escolha pelo critério AIC uma vez que pelos critérios BIC (6 defasagens) e HQ (8 defasagens) o modelo apresenta problemas de especificação, isto é autocorrelação dos resíduos. Estimando com 20 defasagens usando o programa

Dimensão do Espaço de Cointegração

temos que em todos os casos o espaço de cointegração é um como pode ser visto pela tabela abaixo:

Tabela 21: Dimensão do espaço de cointegração para os Casos

Model		I(1) Analysis				
	Lags	r	Eigenvalues	Trace	Crit 5%	p-value
Case 1	20	0	0.0073	54.41	12.28	0.0000
		1	0.0001	0.82	4.07	0.4270
Case 2	20	0	0.0095	70.59	20.16	0.0000
		1	0.0004	3.13	9.14	0.567
Case 3	20	0	0.0095	69.75	15.41	0.0000
		1	0.0003	2.34	3.84	0.126
Case 4	20	0	0.0104	76.64	25.73	0.0000
		1	0.0005	3.32	12.45	0.8280

Usando que a dimensão do espaço de cointegração é um, os casos 1- 4 foram estimados usando o seguinte programa:

Estimação de todos os casos com dimensão do espaço de cointegração igual a um temos os seguintes resultados:

Tabela 22: Critérios de Informação para todos os Casos

Model	T	p	Lags	Method	log-Lik	BIC	HQ	AIC
Case 1	7041	79	20	RRR	123423.67	-34.9590	-35.0100	-3.0360
Case 2	7041	80	20	RRR	123431.60	-34.9600	-35.0110	-35.0380
Case 3	7041	81	20	RRR	123432.00	-34.9590	-35.0110	-36.0380
Case 4	7041	82	20	RRR	123432.00	-34.9590	-35.0110	-35.0385

Pelos critérios de BIC e HQ o melhor modelo é o caso 2, mas pelo critério AIC o melhor modelo é o caso 4.

O modelo escolhido pelo AIC pode ser justificado pelo Custo de Carregamento (*Cost of Carry*). Em mercados de commodities, o preço futuro reflete o preço spot ajustado pelos custos de carregamento, que incluem custos de armazenamento, seguro e taxas de juros. A relação pode ser expressa como:

$$F = S + (C \times t)$$

onde F é o preço futuro, S é o preço spot e C é Custo de Carregamento (*Cost of Carry*) por unidade de tempo por isso multiplicamos por t para que o custo de carregamento seja medido na mesma unidade que spot e futuro.

Como o modelo do caso 2 tem constante na relação de cointegração mas não no VAR , as séries não tem tendência determinística no nível.

Essa constante é conhecida como *basis* ou base e representa a diferença entre o preço spot e o preço futuro de um ativo. A base pode ser positiva ou negativa, dependendo das condições de mercado.

Fatores que influenciam a relação spot-futuro:

1. Custo de transporte e armazenamento: Em mercados de commodities, os custos associados ao transporte e armazenamento do produto até a data de vencimento do contrato futuro afetam a base.
2. Taxa de juros: Em mercados financeiros, a diferença entre os preços spot e futuros pode ser influenciada pelas taxas de juros, devido ao custo de oportunidade do dinheiro investido.
3. Expectativas de mercado: Se o mercado espera que o preço de um ativo suba ou desça no futuro, isso se refletirá na diferença entre os preços spot e futuros.
4. *Convenience Yield*: Em mercados de commodities, o convenience yield representa o benefício não financeiro de possuir a *commodity* física, o que também pode afetar a base.

Estimando este modelo que corresponde a segundo caso temos:

Tabela 23: VEC para o melhor modelo LSPOT e LFUT

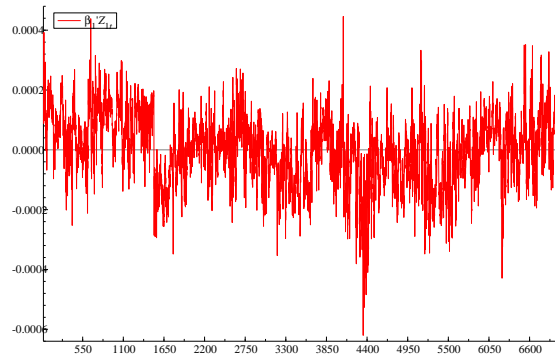
t-statistics in in ()	p-values in []	
Cointegrating Eq	CointEq1	
$LSPOT_{-1}$	1.000000	
$LFUT_{-1}$	-0.97 (-153.8) [0.8340]	
<i>Constant</i>	-0.0544 (-4.8) [0.5217]	
Error Correction:	$D(LSPOT)$	$D(LFUT)$
CointEq1	-0.0342 (-7.9) [0.5136]	-0.0103 (-1.40) [0.7211]

Tabela 24: Bondade de Ajuste e Testes de Especificação para Tabela (23)

Determinant Residual Covariance (d.f. adjusted)	-40.7365
Log Likelihood (d.f. adjusted)	123431.605
Akaike Information Criteria	-35.0381
Schwarz Criteria	-34.9602
Test for autocorrelation LM(1):	$\chi^2(4) = 1.2909[0.8629]$
Test for autocorrelation LM(2):	$\chi^2(4) = 2.6654[0.6153]$
Test for normality:	$\chi^2(4) = 1.5156e + 05[0.0000]$
Test for ARCH LM(1):	$\chi^2(9) = 217.99[0.0000]$
Test for ARCH LM(2):	$\chi^2(18) = 247.53[0.0000]$

O gráfico da relação de cointegração é apresentado abaixo:

Figura 36: Relação de Cointegração $\beta' y_{t-1}$ para LSPOT e LFUT



Para se verificar a validade desta especificação devemos testar não correlação serial, normalidade e homoscedasticidade dos resíduos. Não se rejeita a não correlação dos resíduos mas são rejeitadas as hipóteses de normalidade e homoscedasticidade e deste modo os testes de cointegração não são válidos. Uma explicação para este fato é a própria natureza dos dados, uma vez que dados financeiros tem distribuições com caudas pesadas e tem estrutura na variância condicional.

Um outro problema diz respeito a potência do teste de Johansen. Este teste depende da história, *span*, dos dados e não da frequência. No caso acima o *span* é de um mês, pouca história, embora a frequência dos dados seja muito alta. Deste modo os resultados devem ser interpretados com certo cuidado.

12.10 Restrições na matriz de carga e nos vetores de cointegração

Em algumas situações, como por exemplo no caso da estimação da relação entre spot e futuro, existem restrições que podem ser testadas tanto na relação de longo prazo quanto na matriz de cargas.

A restrição de normalização para LSPOT e que o coeficiente de LFUT é -1 e o da constante é irrestrito, será dado por:

Tabela 25: Restrictions on β :

$H_1 :$	$\beta = (-1, 1, *)$
---------	----------------------

O programa para fazer este teste é dado por:

Teste H_1

e o resultado do teste é:

$$\text{Test of restrictions on } \beta: \chi^2(1) = 15.851[0.0001]$$

como o p-valor é 0,0001 a hipótese é rejeitada.

Em outras situações podemos estar interessados em restrições na matriz de cargas, por exemplo a relação de cointegração não entra do LFUT, mas entra na equação do LSPOT. Esta hipótese corresponde a testar:

Tabela 26: Restrictions on α :

$H_2 :$	$\alpha = (0, *)$
---------	-------------------

O programa para fazer este teste é dado por:

Teste H_2

e o resultado do teste é:

$$\text{Test of restrictions on } \alpha: \chi^2(1) = 1.7913[0.1808]$$

como o p-valor é 0.1808 a hipótese não é rejeitada.

Podemos também testar as duas hipóteses conjuntamente ($H_1 \cap H_2$)

O programa para fazer este teste é dado por:

Teste $H_1 \cap H_2$

e o resultado é apresentado abaixo:

$$\text{Test of restrictions on } \alpha \text{ and } \beta: \chi^2(2) = 16.046[0.0003]$$

e como o p-valor é 0.0003 rejeitamos estas hipóteses.

12.11 Teste de Exogeneidade Fraca

Para se testar exogeneidade fraca devemos testar se os parâmetros do modelo condicional e do(s) modelo(s) marginal(is) são variação livre. Uma forma de fazer isto é através de teste de restrições lineares na matriz de carga.

Suponha que $\mathbf{z}_t = (y_t : x_t : w_t)$ e que existem dois vetores de cointegração, isto é, β' é uma matriz 2×3 , por exemplo:

$$\beta' = \begin{bmatrix} 1 & -\beta_1 & 0 \\ 0 & 1 & -\beta_2 \end{bmatrix} \quad (109)$$

Observe que os coeficientes unitários não implicam em perda de generalidade pois a normalização pode ser feita por qualquer componente do vetor. Por outro lado os coeficientes nulos implicam em perda de generalidade e devem ser testados usando testes do tipo H_1 .

A representação VEC neste caso é dada por:

$$\begin{aligned} \Delta y_t &= \alpha_{11}(y_{t-1} - \beta_1 x_{t-1}) + \alpha_{12}(x_{t-1} - \beta_2 w_{t-1}) + \varepsilon_{1t} \\ \Delta x_t &= \alpha_{21}(y_{t-1} - \beta_1 x_{t-1}) + \alpha_{22}(x_{t-1} - \beta_2 w_{t-1}) + \varepsilon_{2t} \\ \Delta w_t &= \alpha_{31}(y_{t-1} - \beta_1 x_{t-1}) + \alpha_{32}(x_{t-1} - \beta_2 w_{t-1}) + \varepsilon_{3t} \end{aligned} \quad (110)$$

e a representação do modelo condicional e marginal é dada por:

$$\begin{aligned} \Delta y_t &= \delta_1 \Delta x_t + \delta_2 \Delta w_t + \delta_3(y_{t-1} - \beta_1 x_{t-1}) + \delta_4(x_{t-1} - \beta_2 w_{t-1}) + v_t \\ \Delta x_t &= \alpha_{21}(y_{t-1} - \beta_1 x_{t-1}) + \alpha_{22}(x_{t-1} - \beta_2 w_{t-1}) + \varepsilon_{2t} \\ \Delta w_t &= \alpha_{31}(y_{t-1} - \beta_1 x_{t-1}) + \alpha_{32}(x_{t-1} - \beta_2 w_{t-1}) + \varepsilon_{3t} \end{aligned} \quad (111)$$

Os parâmetros do modelo condicional são:

$$\lambda_1 = (\delta_1 : \delta_2 : \delta_3 : \delta_4 : \beta_1 : \beta_2 : \sigma^2) \quad (112)$$

e dos modelos marginais

$$\lambda_2 = (\alpha_{21} : \alpha_{22} : \alpha_{31} : \alpha_{32} : \beta_1 : \beta_2 : \sigma_{22} : \sigma_{23} : \sigma_{33}) \quad (113)$$

Para que λ_1 e λ_2 sejam variação livre devemos ter:

$$\alpha_{21} = \alpha_{22} = \alpha_{31} = \alpha_{32} = 0 \quad (114)$$

isto é, os vetores de cointegração só entram na primeira equação implicando que z_t e w_t são exógenos fracos para λ_1 .

Logo para se testar exogeneidade fraca devemos testar se "as relações de cointegração" não entram nos modelos marginais, isto é, um teste do tipo H_2 .

Voltando ao exemplo da taxa spot e futura o teste H_2 que foi feito acima é o teste de exogeneidade fraca da taxa fut para o modelo condicional da taxa spot. Como não se rejeitou a hipótese nula temos que a taxa fut é exogena fraca para os parâmetros do modelo condicional.

13 Apêndice - Operador Vec e Produto de Kronecker

Seja A uma matriz $n \times m$ podemos definir o operador vec por

$$vec(A) = \begin{pmatrix} a_{11} \\ \cdot \\ \cdot \\ \cdot \\ a_{m1} \\ \cdot \\ \cdot \\ \cdot \\ a_{1n} \\ \cdot \\ \cdot \\ \cdot \\ a_{mn} \end{pmatrix}$$

que é um vetor $mn \times 1$ que consiste em empilhar as colunas da matriz A .

Uma outra operação importante é o produto de Kronecker. Considere as matrizes A e B de dimensões $m \times n$ e $p \times q$. Podemos então definir

$$A \otimes B = (a_{ij}B)_{mp,nq}$$

por exemplo se $\Omega = (\omega_{ij})_{2,2}$ e $S = (s_{ij})_{2,2}$ temos

$$\Omega \otimes S = \begin{pmatrix} \omega_{11}s_{11} & \omega_{11}s_{12} & \omega_{12}s_{11} & \omega_{12}s_{12} \\ \omega_{11}s_{21} & \omega_{11}s_{22} & \omega_{12}s_{21} & \omega_{12}s_{22} \\ \omega_{21}s_{11} & \omega_{21}s_{12} & \omega_{22}s_{11} & \omega_{22}s_{12} \\ \omega_{21}s_{21} & \omega_{21}s_{22} & \omega_{22}s_{21} & \omega_{22}s_{22} \end{pmatrix}$$

Algumas propriedades do produto de Kronecker e do vec

1. $(A \otimes B)' = A' \otimes B'$
2. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ se A e B são não singulares.
3. $|A \otimes B| = |A|^m |B|^n$ se A e B são matrizes $n \times n$ e $m \times m$ respectivamente.
4. $(A \otimes B)(C \otimes D) = AC \otimes BD$
5. $c \otimes A = cA = Ac = A \otimes c$ se c for constante.
6. $tr(A \otimes B) = tr(A)tr(B)$
7. $(vec A)'(vec B) = tr(A'B)$ se A e B são ambas matrizes $m \times n$
8. $vec(a') = vec(a) = a$ se a é $n \times 1$
9. $vec(ABC) = (C' \otimes A)vec(B)$ se o produto ABC estiver definido

10. $tr(ABCD) = vec(D')'(A)vec(B) = (vec(D))'(A \otimes C')(vec(B'))$ porque $vec(B') = (vec(B))'$.

14 Apêndice - Demonstração de (44) e (45)

O último termo de (43) pode ser escrito da seguinte forma

$$\begin{aligned}
& \sum_{t=2}^T (\mathbf{y}_t - \Phi_1 \mathbf{y}_{t-1})' \Omega^{-1} (\mathbf{y}_t - \Phi_1 \mathbf{y}_{t-1}) = \\
&= \sum_{t=2}^T (\mathbf{y}_t - \hat{\Phi}_1 \mathbf{y}_{t-1} + \hat{\Phi}_1 \mathbf{y}_{t-1} - \Phi_1 \mathbf{y}_{t-1})' \Omega^{-1} (\mathbf{y}_t - \hat{\Phi}_1 \mathbf{y}_{t-1} + \hat{\Phi}_1 \mathbf{y}_{t-1} - \Phi_1 \mathbf{y}_{t-1}) \\
&= \sum_{t=2}^T (\hat{\varepsilon}_t + (\hat{\Phi}_1 - \Phi_1) \mathbf{y}_{t-1})' \Omega^{-1} (\hat{\varepsilon}_t + (\hat{\Phi}_1 - \Phi_1) \mathbf{y}_{t-1}) \\
&= \sum_{t=2}^T \hat{\varepsilon}_t' \Omega^{-1} \hat{\varepsilon}_t + 2 \sum_{t=2}^T \hat{\varepsilon}_t' \Omega^{-1} (\hat{\Phi}_1 - \Phi_1) \mathbf{y}_{t-1} + \sum_{t=2}^T \mathbf{y}_{t-1}' (\hat{\Phi}_1 - \Phi_1)' \Omega^{-1} (\hat{\Phi}_1 - \Phi_1) \mathbf{y}_{t-1}
\end{aligned} \tag{115}$$

Agora considere o termo do medio de (115) como este termo é um escalar ele permanece inalterado se aplicarmos o operador traço, isto é:

$$\begin{aligned}
\sum_{t=2}^T \hat{\varepsilon}_t' \Omega^{-1} (\hat{\Phi}_1 - \Phi_1) \mathbf{y}_{t-1} &= trace \left[\sum_{t=2}^T \hat{\varepsilon}_t' \Omega^{-1} (\hat{\Phi}_1 - \Phi_1) \mathbf{y}_{t-1} \right] \\
&= trace \left[\sum_{t=2}^T \Omega^{-1} (\hat{\Phi}_1 - \Phi_1) \mathbf{y}_{t-1} \hat{\varepsilon}_t' \right] \\
&= trace \left[\Omega^{-1} (\hat{\Phi}_1 - \Phi_1) \sum_{t=2}^T \mathbf{y}_{t-1} \hat{\varepsilon}_t' \right]
\end{aligned} \tag{116}$$

mas os resíduos são ortogonais aos regressores e portanto $\sum_{t=2}^T \mathbf{y}_{t-1} \hat{\varepsilon}_t' = 0$ e (115) reduz-se a:

$$\sum_{t=2}^T \hat{\varepsilon}_t' \Omega^{-1} \hat{\varepsilon}_t + \sum_{t=2}^T \mathbf{y}_{t-1}' (\hat{\Phi}_1 - \Phi_1)' \Omega^{-1} (\hat{\Phi}_1 - \Phi_1) \mathbf{y}_{t-1} \tag{117}$$

Observe que em (117) só o segundo termo depende de Φ_1 e como Ω é positiva definida Ω^{-1} também será e se definirmos $\mathbf{y}_{t-1}^* = (\hat{\Phi}_1 - \Phi_1) \mathbf{y}_{t-1}$ temos que este segundo termo pode ser escrito da seguinte forma:

$$\sum_{t=2}^T \mathbf{y}_{t-1}'^* \Omega^{-1} \mathbf{y}_{t-1}^* \tag{118}$$

que é sempre positivo uma vez que é uma forma quadrática e só será nulo quando $\mathbf{y}_{t-1}^* = 0$ que será o menor valor para (116). Observe que $\mathbf{y}_{t-1}^* = 0 \Rightarrow \Phi_1 = \hat{\Phi}_1$.

Para obter o estimador de máxima verossimilhança de Ω vamos precisar das seguintes propriedades de derivada de forma quadrática e de determinante:

1. Seja \mathbf{A} uma matriz $n \times n$ com determinante positivo então $\frac{\partial \log |\mathbf{A}|}{\partial \mathbf{A}} = (\mathbf{A}')^{-1}$
2. Seja \mathbf{A} uma matriz $n \times n$ e x um vetor $n \times 1$ então $\frac{\partial x' \mathbf{A} x}{\partial \mathbf{A}} = x x'$

Observe que (43) pode ser escrito da seguinte forma:

$$S(\Omega : \hat{\Phi}_1) = \frac{kT}{2} \ln(2\pi) - \frac{T}{2} \ln |\Omega^{-1}| + \frac{1}{2} \sum_{t=1}^T \hat{\varepsilon}_t' \Omega^{-1} \hat{\varepsilon}_t \quad (119)$$

e derivando (119) em relação a Ω^{-1} temos:

$$\begin{aligned} \frac{\partial S(\Omega : \hat{\Phi}_1)}{\partial \Omega^{-1}} &= -\frac{T}{2} \frac{\partial \ln |\Omega^{-1}|}{\partial \Omega^{-1}} + \frac{1}{2} \sum_{t=1}^T \frac{\partial \hat{\varepsilon}_t' \Omega^{-1} \hat{\varepsilon}_t}{\partial \Omega^{-1}} \\ &= -\frac{T}{2} \Omega' + \frac{1}{2} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t' \\ \hat{\Omega} &= (1/T) \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t' \end{aligned}$$

Referências

- Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 60(2):255–265.
- Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, 79:655–673.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.
- Granger, C. W. J. (1983). Co-integrated variables and error-correcting models. Technical Report 93-13, UCSD.
- Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2):111–120.
- Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):190–195.

- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254.
- Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
- Johnston, J. and DiNardo, J. (1996). *Econometric Methods*. McGraw-Hill/Irwin.
- Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer, 2nd edition edition.
- MacKinnon, J. G. (1991). Critical values for cointegration tests. In Engle, R. F. and Granger, C. W. J., editors, *Long- Run Economic Relationships*, pages 267–276. Oxford University Press.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.
- Sargan, J. D. (1964). Wages and prices in the united kingdom: A study in econometric methodology. In Hart, P.E., M. G. and Whitaker, J. K., editors, *Econometric Analysis for National Economic Planning (with discussions)*, pages 25–63. Butterworths.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.
- Stock, J. H. and Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, 15(4):101–115.