

# EAE1223: ECONOMETRIA III

## AULA 7 - COINTEGRAÇÃO

Luis A. F. Alvarez

16 de junho de 2025

# NATUREZA DA INFERÊNCIA ESPÚRIA

- Em algumas aulas atrás, vimos que a inferência baseada nos estimadores de MQO de um modelo linear de um processo  $I(1)$  em outro processo  $I(1)$ , **completamente independentes**, gerava conclusões espúrias.
  - Relação verdadeira é 0, mas testes de hipótese proviam forte evidência contra hipótese de associação nula (p-valores baixíssimos).
- Uma pergunta que cabe é: se há dependência entre os processos  $I(1)$ , há condições sob as quais estimadores de MQO produzem inferência **não espúria**?

## UM EXEMPLO

- Considere o seguinte processo bivariado :

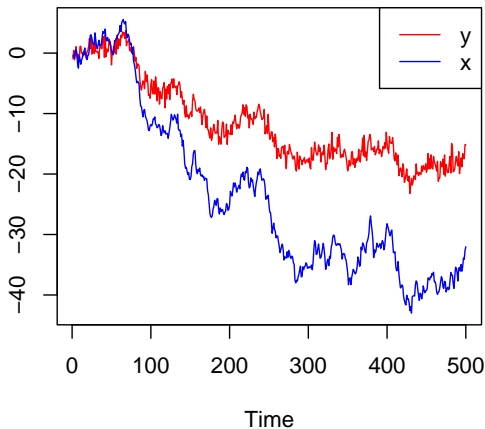
$$x_t = x_{t-1} + u_t ,$$

$$y_t = \gamma x_t + v_t ,$$

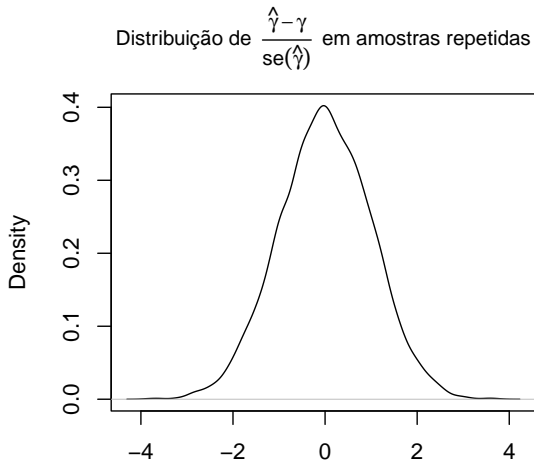
onde  $\{u_t\}_t$  e  $\{v_t\}_t$  são ruídos brancos independentes uns dos outros.

- Observe que ambos os processos são  $I(1)$
- Vamos considerar as propriedades do estimador de MQO  $\hat{\gamma}$  de  $\gamma$ , sobre realizações repetidas da incerteza econômica:

# UMA REALIZAÇÃO DO PROCESSO POR $T = 500$ PERÍODOS ( $\gamma = 0.5$ )



# DISTRIBUIÇÃO DE $\hat{t} = \frac{\hat{\gamma} - \gamma}{\text{se}(\hat{\gamma})}$ EM REALIZAÇÕES REPETIDAS DA INCERTEZA ( $T = 500$ E $\gamma = 0.5$ )



# TENDÊNCIA ESTOCÁSTICA COMUM

- Note que a estatística  $t$ , sob a nula, tem distribuição convencional (normal) sobre realizações repetidas da incerteza.
- Isso ocorre mesmo com ambas as séries sendo  $I(1)$ .
- Por que isso ocorre?
  - Processos possuem tendência estocástica **comum**.
  - A tendência comum “amarra” o comportamento explosivo dos processos, recuperando a validade da aproximação normal.
  - A processos  $I(1)$  com tendência estocástica comum, damos o nome de **processos cointegrados**.

# COINTEGRAÇÃO

- Considere um processo vetorial  $\{\mathbf{Z}_t\}_{t \in \mathbb{Z}}$  com  $n$  variáveis.
- Dizemos que o vetor é cointegrado (de ordem  $(1,1)$ ), ou  $CI(1,1)$ , se:
  1. Cada uma das séries  $\{Z_{jt}\}_{t \in \mathbb{Z}}$ ,  $j = 1, \dots, n$ , é  $I(1)$ .
  2. Existe um vetor  $\gamma \in \mathbb{R}^n$ ,  $\gamma \neq \mathbf{0}_{n \times 1}$ , tal que  $\gamma' \mathbf{Z}_t$ , é  $I(0)$ .
- Se processo é  $CI(1,1)$ , existe combinação linear não trivial tal que processo resultante é estacionário.
  - Existe uma **relação de longo prazo**, fruto de uma tendência comum, amarrando os processos, de modo que desvios dessa relação são  $I(0)$ .
  - Note que a relação não é única, visto que sempre podemos multiplicá-la por uma constante não nula.
- De modo mais geral, um processo vetorial é dito cointegrado de ordem  $(d,b)$ , ou  $CI(d,b)$ ,  $0 < b \leq d$ , se:
  - Cada uma das séries  $\{Z_{jt}\}_{t \in \mathbb{Z}}$ ,  $j = 1, \dots, n$ , é  $I(d)$ .
  - Existe um vetor  $\gamma \in \mathbb{R}^n$ ,  $\gamma \neq 0$ , tal que  $\gamma' \mathbf{Z}_t$ , é  $I(d-b)$ .
- Focaremos no caso  $CI(1,1)$ , que é o mais economicamente interessante.

# COINTEGRAÇÃO E TEORIA ECONÔMICA

- O conceito de cointegração é de natureza estatística.
  - Não obstante, a evidência ou não de uma relação de longo prazo pode ter uma interpretação econômica.
- Considere uma medida dos preços praticados para uma cesta de bens no Brasil, em reais; uma outra medida, para cesta similar, dos preços praticados em dólares nos Estados Unidos, e a taxa de câmbio real/Estados Unidos.
  - Pela teoria da paridade do poder de compra, embora processos individualmente apresentem comportamento não estacionário (estocástico), deveria haver uma relação de cointegração entre estas variáveis.
- Se há neutralidade da moeda no longo prazo, deveria haver relação de cointegração entre base monetária **nominal** e nível de preços que **excluisse** produto real.
  - Pela equação quantitativa da moeda, deveria haver relação de longo prazo entre produto real e base monetária **real**.



## TESTANDO PELA PRESENÇA DE COINTEGRAÇÃO

- Engle e Granger (1987) popularizaram um procedimento bastante intuitivo para se testar a nula de **não cointegração**.
- Considere um processo vetorial  $\mathbf{Z}_t = (y_t, \mathbf{x}_t)$  em que cada entrada é  $I(1)$ , e para o qual acreditamos que possa haver alguma relação de cointegração envolvendo  $y_t$ .
  - Para sistemas com duas variáveis ( $\mathbf{x}_t$  escalar), crença é sem perda de generalidade, visto que cointegração necessariamente envolve  $y_t$ .
  - Para sistemas com mais de duas variáveis, efetivamente nos restringimos a alternativas em que  $y_t$  participa da relação.
- Ideia de Engle e Granger (1987): estimar por MQO uma regressão de  $y_t$  em  $\mathbf{x}_t$  (e um intercepto):

$$y_t = \hat{\alpha} + \hat{\gamma}' \mathbf{x}_t + \hat{u}_t, \quad t = 1, \dots, T$$

e fazer um teste da nula de raiz unitária nos resíduos  $\{\hat{u}_t\}_{t=1}^T$ .

- Se não há cointegração, resíduos devem apresentar tendência estocástica.
- Se há cointegração envolvendo  $y_t$ , resíduos deveriam ser estacionários.

# PROCEDIMENTO DE PHILLIPS E OULIARIS (1990)

- Teste não pode ser feito usando valores críticos convencionais, pois estimação preliminar afeta valores críticos.
  - Phillips e Ouliaris (1990) tabularam os valores críticos para as estatísticas do teste ADF e de Phillips e Perron (1988).
- Tabulação depende de existência de tendência linear no nível da série.
  - Se há somente tendência estocástica em nível, usar [Caso 2](#).
  - Se séries apresentam tendência estocástica, e ao menos uma das séries em  $\mathbf{x}_t$  apresentar tendência determinística linear em nível, usar [Caso 3](#).
  - Se séries apresentam tendência estocástica, nenhuma das séries em  $\mathbf{x}_t$  apresentar tendência determinística em nível, mas  $y_t$  apresentar tendência linear, há que se modificar o conceito de cointegração. Nesse caso, é impossível gerar combinação linear estacionária, mas ainda é possível gerar combinação linear *trend-stationary* (cointegração em torno de tendência determinística). Se quisermos testar isso, rodamos

$$y_t = \hat{\alpha} + \hat{\delta}t + \hat{\gamma}'\mathbf{x}_t + \hat{u}_t, \quad t = 1, \dots, T,$$

e usamos os valores críticos do [Caso 3](#), supondo que há  $n + 1$  variáveis.

- Se sabemos que relação de cointegração tem média zero, podemos omitir o intercepto da regressão e considerando o [Caso 1](#).

# VALORES CRÍTICOS

**TABLE B.9**

**Critical Values for the Phillips  $Z_t$  Statistic or the Dickey-Fuller  $t$  Statistic When Applied to Residuals from Spurious Cointegrating Regression**

Number of right-hand variables in regression, excluding trend or constant ( $n - 1$ )	Sample size ( $T$ )	Probability that $(\hat{\rho} - 1)/\hat{\sigma}_{\hat{\rho}}$ is less than entry						
		0.010	0.025	0.050	0.075	0.100	0.125	0.150
Case 1								
1	500	-3.39	-3.05	-2.76	-2.58	-2.45	-2.35	-2.26
2	500	-3.84	-3.55	-3.27	-3.11	-2.99	-2.88	-2.79
3	500	-4.30	-3.99	-3.74	-3.57	-3.44	-3.35	-3.26
4	500	-4.67	-4.38	-4.13	-3.95	-3.81	-3.71	-3.61
5	500	-4.99	-4.67	-4.40	-4.25	-4.14	-4.04	-3.94
Case 2								
1	500	-3.96	-3.64	-3.37	-3.20	-3.07	-2.96	-2.86
2	500	-4.31	-4.02	-3.77	-3.58	-3.45	-3.35	-3.26
3	500	-4.73	-4.37	-4.11	-3.96	-3.83	-3.73	-3.65
4	500	-5.07	-4.71	-4.45	-4.29	-4.16	-4.05	-3.96
5	500	-5.28	-4.98	-4.71	-4.56	-4.43	-4.33	-4.24
Case 3								
1	500	-3.98	-3.68	-3.42	—	-3.13	—	—
2	500	-4.36	-4.07	-3.80	-3.65	-3.52	-3.42	-3.33
3	500	-4.65	-4.39	-4.16	-3.98	-3.84	-3.74	-3.66
4	500	-5.04	-4.77	-4.49	-4.32	-4.20	-4.08	-4.00
5	500	-5.36	-5.02	-4.74	-4.58	-4.46	-4.36	-4.28

The probability shown at the head of the column is the area in the left-hand tail.

Source: P. C. B. Phillips and S. Ouliaris, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica* 58 (1990), p. 190. Also Wayne A. Fuller, *Introduction to Statistical Time Series*, Wiley, New York, 1976, p. 373.

## O QUE O MQO ESTIMA SOB COINTEGRAÇÃO?

- Para um processo vetorial  $\mathbf{Z}_t = (y_t, \mathbf{x}_t)$   $n$ -variado cointegrado, há no mínimo uma, e no máximo  $n - 1$  relações de cointegração linearmente independentes disponíveis.
  - Se houvesse  $n$  relações de cointegração, é possível mostrar que o processo na verdade seria estacionário.
- Sejam  $\left\{ (1, \mathbf{b}'_1)', (1, \mathbf{b}'_2)', \dots, (1, \mathbf{b}'_q)' \right\}$  as  $q$  relações linearmente independentes que envolvem  $y_t$ .
- Se  $q > 0$ , possível mostrar que estimador de MQO de  $y_t$  em  $\mathbf{x}_t$  e em um intercepto é consistente para o parâmetro  $\gamma^* \in \text{span}\{\mathbf{b}_1, \dots, \mathbf{b}_q\}$  do **melhor preditor linear de  $y_t$  em  $\mathbf{x}_t$**  definido de tal forma que:

$$y_t = \alpha^* + \mathbf{x}'_t \gamma^* + u_t$$

com  $u_t$  estacionário,  $\mathbb{E}[u_t] = 0$  e  $\text{cov}(\mathbf{x}_t, u_t) = \mathbf{0}$ .

- Se  $\text{cov}(\Delta \mathbf{x}_t, u_s) = \mathbf{0}$  para todo  $t$  e  $s$ , podemos realizar inferência sobre  $\gamma^*$  com base no estimador de MQO, estatísticas  $t$  com erros padrão HAC e valores críticos normais. Este é o caso da simulação anterior.
- Se  $q = 0$ , inferência com base no MQO é espúria.

# LIMITAÇÕES DA METODOLOGIA DE ENGLE-GRANGER

- Não obstante sua praticidade, a metodologia de Engle-Granger possui algumas limitações:
  - Ao efetivamente escolher uma variável do lado esquerdo, nos restringimos a relações de cointegração que envolvem  $y_t$ .
  - Se  $n > 2$ , não conseguimos separar as relações de cointegração distintas (linearmente independentes) potencialmente existentes, cada uma correspondente a uma relação de longo prazo diferente.
  - Além disso, para conduzir inferência sobre o vetor de cointegração estimado, necessitamos de hipóteses mais fortes sobre os erros.
- Como veremos a seguir, o **teorema da representação de Granger** nos diz que, para um sistema de variáveis cointegrado, há restrições no processo gerador que são úteis para separar as relações de cointegração existentes.
  - Essas relações podem ser usadas para construir modelos preditivos superiores (a um VAR em primeiras diferenças).
  - Além disso, a metodologia nos permite testar hipóteses sobre as diferentes relações diretamente.

# TEOREMA DE REPRESENTAÇÃO DE GRANGER

- Seja  $\{\mathbf{Z}_t\}_{t \in \mathbb{Z}}$  um processo vetorial em  $\mathbb{R}^n$  em que cada uma das variáveis é  $I(1)$ , e o processo é descrito por um VAR(p):

$$\mathbf{Z}_t = \mathbf{c} + \sum_{j=1}^p A_j \mathbf{Z}_{t-j} + \mathbf{v}_t. \quad (1)$$

- Teorema de representação de Granger:  $\mathbf{Z}_t$  é  $CI(1,1)$  se, e somente se, admite a representação de correção de erros:

$$\Delta \mathbf{Z}_t = \mathbf{c} + \alpha' \beta \mathbf{Z}_{t-1} + \sum_{j=1}^{p-1} \mathbf{B}_j \Delta \mathbf{Z}_{t-j} + \mathbf{v}_t, \quad (2)$$

onde  $\alpha$  e  $\beta$  são matrizes  $r \times n$  com posto  $r$ , e cada uma das linhas de  $\beta$  corresponde a uma das  $r \in \{1, \dots, n-1\}$  relações de cointegração linearmente independentes do sistema.

# COINTEGRAÇÃO E CORREÇÃO DE ERROS

- Teorema de representação de Granger nos diz que, se sistema é cointegrado, variações de curto prazo  $\Delta \mathbf{Z}_t$  respondem a desvios nas  $r$  relações de longo prazos, de modo a garantir a tendência comum de longo prazo do sistema.
- Do ponto de vista preditivo, o teorema nos diz que, se as variáveis são cointegradas, a relação de cointegração deve ser usada na predição (relativamente a trabalhar com um VAR em primeiras diferenças).
  - Por outro lado, embora, em um sistema cointegrado, um estimador de VAR irrestrito em nível é consistente para os parâmetros populacionais (a relação não é espúria), o teorema de representação de Granger nos produz restrições que permitem realizar inferência sobre os parâmetros.
- Veremos a seguir como o modelo vetorial de correção de erros (VECM) pode ser estimado, e suas estimativas podem ser usadas para se testar o número de relações de cointegração existentes no sistema.

# DETECTANDO O NÚMERO DE RELAÇÕES DE COINTEGRAÇÃO

- Considere o estimador de máxima verossimilhança condicional de (2), onde a verossimilhança é derivada sob a hipótese auxiliar de que  $\mathbf{v}_t \stackrel{iid}{\sim} N(\mathbf{0}_{n \times 1}, \Sigma)$ , e a estimação impõe que  $r = q$ .
  - Podemos escolher  $p$  preliminarmente ajustando um VAR **em nível** e fazendo os procedimentos vistos na aula anterior.
- Defina por  $\hat{L}_q$  a log-verossimilhança, no máximo atingido, calculada sob a hipótese de que  $r = q$ .
- Johansen (1991) propõe testar:

$$H_0 : r = q, \quad H_1 : r = q + 1,$$

usando a estatística de razão de verossimilhança

$$LR(q, q + 1) = 2 \cdot (\max_{j \in \{q, q+1\}} \hat{L}_j - \hat{L}_q) = 2 \cdot (\hat{L}_{q+1} - \hat{L}_q).$$

- Distribuição não padrão sob a nula, e tabulada pelo autor.
- Estatística de teste conhecida como do **máximo autovalor**, visto que pode ser obtida através de um autovalor de uma matriz auxiliar calculada a partir dos dados.



# PROCEDIMENTO SEQUENCIAL E ESTATÍSTICA DO TRAÇO

- Com base no teste do máximo autovalor, podemos detectar  $r$  procedendo sequencialmente, começando os testes com  $q = 0$  e parando no momento em que não rejeitamos mais a nula.
  - Valor de  $r$  é dado pelo menor  $q$  tal que não rejeitamos a nula do teste.
- Um outro teste possível é:

$$H_0 : r = q, \quad H_1 : r \geq q + 1,$$

que pode ser conduzido a partir da estatística de traço

$LR(q, n) = 2 \cdot (\max_{n \geq j \geq q} \hat{L}_j - \hat{L}_q) = 2 \cdot (\hat{L}_n - \hat{L}_q)$ , e os valores críticos tabulados por Johansen (1991) para esse caso.

- Também procedemos sequencialmente, parando no  $r$  em que não rejeitamos a nula.

## COMPONENTES DETERMINÍSTICOS

- Em nossa discussão, consideramos o modelo (2), em que, a princípio,  $\mathbf{c} \in \mathbb{R}^n$  toma qualquer valor:
  - Como há no máximo  $r = n - 1$  relações de cointegração, possível mostrar que a constante irrestrita permite  $n - r > 0$  tendências lineares linearmente independentes afetando o nível das séries; além de  $r$  médias distintas em cada uma das relações de cointegração.
  - Essa flexibilidade vem ao custo de poder reduzido.
- Se sabemos que não há tendências lineares no nível das séries, mas ainda assim queremos permitir que as relações de cointegração tenham média diferente de zero, podemos considerar o modelo:

$$\Delta \mathbf{Z}_t = \alpha' [\mathbf{a} + \beta \mathbf{Z}_{t-1}] + \sum_{j=1}^{p-1} \mathbf{B}_j \Delta \mathbf{Z}_{t-j} + \mathbf{v}_t,$$

com  $\mathbf{a} \in \mathbb{R}^r$ . Ao reduzir o número de parâmetros ganhamos poder (valores críticos para os testes mudam nesse caso).

## COMPONENTES DETERMINÍSTICOS (CONT.)

- Por outro lado, se queremos considerar que a cointegração somente remove a tendência estocástica, devemos considerar o modelo:

$$\Delta \mathbf{Z}_t = \mathbf{a} + \alpha'[\mathbf{b}t + \beta \mathbf{Z}_{t-1}] + \sum_{j=1}^{p-1} \mathbf{B}_j \Delta \mathbf{Z}_{t-j} + \mathbf{v}_t,$$

- Por fim, também é possível considerar testes nos diferentes modelos sob a presença adicional de *dummies* sazonais.
  - Valores críticos tabulados em Johansen (1995).

# INFERÊNCIA SOBRE PARÂMETROS DO VECM

- Estimado um VECM, podemos fazer inferência individual (conjunta) sobre os  $\beta$  usando estatísticas  $t$  ( $F$ ) e valores críticos normais ( $F$ ).
  - Um ponto importante é que existem diferentes pares  $(\alpha, \beta)$  compatíveis com  $r$  relações de cointegração (linearmente independentes)  $\implies$  precisamos normalizar  $\beta$  para separar  $\alpha$  de  $\beta$ .
- Como alternativa, é possível testar algumas nulas conjuntas sobre os  $\beta$  usando a estatística LR, sem necessidade de normalização.
  - Por exemplo, podemos testar que somente um subconjunto de  $s \geq r$  das variáveis participa das relações.
- Podemos fazer inferência individual ou conjunta sobre os  $B_j$  e  $\alpha$  usando estatística  $t$  ou testes  $F$  convencionais, respectivamente.
  - Para testar  $\alpha$ , precisamos de normalizações em  $\beta$ , embora não seja o caso para testes que só envolvem os  $B_j$ .
- No entanto, **não** é possível, no geral, fazer testes de causalidade de Granger usando as distribuições convencionais de referência.
  - Isso se deve à nula envolver restrições sobre coeficientes associados a variáveis  $I(0)$  e  $I(1)$   $\implies$  falha de aproximações convencionais.

# METODOLOGIA VAR COM SÉRIES $I(1)$

- A discussão anterior sugere o seguinte procedimento para se ajustar um VAR com séries  $I(1)$ .
  1. Checar pela presença de cointegração.
  2. Se não há cointegração ( $r = 0$ ), estimar um VAR( $p-1$ ) em primeiras diferenças.
    - Nesse caso, inferência tradicional é válida para todos os parâmetros estimados.
  3. Se há cointegração ( $r > 0$ ), estimar o modelo de correção de erros, impondo o número de relações de cointegração detectadas.
    - Nesse caso, é possível fazer inferência como discutido no slide anterior.

## VAR(p) EM NÍVEL COM SÉRIES I(1)

- Embora seja possível mostrar que, com séries I(1), o estimador do VAR(p) em nível seja consistente para os parâmetros tanto com  $r = 0$  como com  $r > 0$ , há dois motivos para não fazê-lo.
- Estimar um VAR(p-1) em primeira diferença quando  $r = 0$  é preferível do ponto de vista de eficiência estatística, visto que efetivamente restringimos os coeficientes da representação VECM equivalente.
- Além disso, estimar o VAR(p-1) em primeira diferença quando  $r = 0$  ou a representação VECM quando  $r > 0$  é preferível à estimação direta VAR(p) em nível, visto que temos garantias de inferência válida para diferentes subconjuntos dos parâmetros.
  - De modo geral, inferência no VAR(p) em nível não é padrão.

# METODOLOGIA VAR COM SÉRIES $I(0)$ E $I(1)$

- E se há as variáveis de interesse envolvem processos  $I(1)$  e  $I(0)$ ?
  - Nesse caso, comumente se estima um VAR com as séries  $I(1)$  em primeiras diferenças, e as séries  $I(0)$  em nível.
    - No entanto, isso ignora a possível cointegração entre as séries  $I(1)$ .
  - Uma alternativa é detectar as  $r$  relações de cointegração entre as séries  $I(1)$ , e, se  $r > 0$ , estimar um VECM com  $r + n_0$  relações de cointegração, onde  $n_0$  é o número de variáveis  $I(0)$ .
    - Recorde-se que, numa representação VECM com  $s$  variáveis, se há  $s$  relações de cointegração, os processos são efetivamente estacionários.
    - Se  $r = 0$ , estimamos o VAR misto, com as séries  $I(1)$  em diferença e as  $I(0)$  em nível.

# REFERÊNCIAS I



Engle, Robert F. e C. W. J. Granger (1987). “Co-Integration and Error Correction: Representation, Estimation, and Testing”. Em: *Econometrica* 55.2, pp. 251–276. ISSN: 00129682, 14680262. URL: <http://www.jstor.org/stable/1913236> (acesso em 02/05/2024).



Johansen, Søren (1991). “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models”. Em: *Econometrica* 59.6, pp. 1551–1580. ISSN: 00129682, 14680262. URL: <http://www.jstor.org/stable/2938278> (acesso em 06/05/2024).



— (dez. de 1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press. ISBN: 9780198774501. DOI: 10.1093/0198774508.001.0001. URL: <https://doi.org/10.1093/0198774508.001.0001>.



Phillips, P. C. B. e S. Ouliaris (1990). “Asymptotic Properties of Residual Based Tests for Cointegration”. Em: *Econometrica* 58.1, pp. 165–193. ISSN: 00129682, 14680262. URL: <http://www.jstor.org/stable/2938339> (acesso em 02/05/2024).



## REFERÊNCIAS II



Phillips, Peter C. B. e Pierre Perron (1988). “Testing for a Unit Root in Time Series Regression”. Em: *Biometrika* 75.2, pp. 335–346. ISSN: 00063444. URL: <http://www.jstor.org/stable/2336182> (acesso em 13/03/2024).