

AVALIAÇÃO DA ESTACIONARIEDADE E TESTE DE COINTEGRAÇÃO EM SÉRIES TEMPORAIS: O CASO DA DEMANDA DE ENERGIA ELÉTRICA RESIDENCIAL NO BRASIL

Anderson Garcia Silveira¹; Viviane Leite Dias de Mattos²; Andrea Cristina Konrath³

Resumo

Uma das maneiras mais recomendadas para modelar séries econômicas não estacionárias, com a finalidade de evitar o fenômeno da regressão espúria, é através do conceito de cointegração. Para que seja possível aplicar esse conceito, é necessário que as séries estudadas possuam a mesma ordem de integração. Os testes de raiz unitária são desenvolvidos com o intuito de verificar a presença de tendência estocástica em séries temporais. Os testes de ADF e KPSS são aplicados para encontrar a ordem de integração de duas séries temporais relacionadas ao consumo de energia elétrica no Brasil: consumo e tarifa. A aplicação dos testes indica que ambas as séries possuem a mesma ordem de integração: $I(1)$. Com a informação da ordem de integração das séries, os testes de Johansen são utilizados para verificar se as séries possuem relação de cointegração. A aplicação dos testes indica que existe relação de longo prazo entre as variáveis.

Palavras chave: Séries Temporais, Cointegração, Testes de Raiz Unitária.

Abstract

One of the most recommended ways to model non-stationary economics time series, in order to avoid spurious regression, is by the concept of cointegration. To be possible applying this concept, it is necessary that the studied series have the same integration order. The unit root tests have been developed on the way to determinate the presence of stochastic trend on the time series. The ADF and KPSS tests are applied to find out the integration order of two time-series related to electrical energy in Brazil: consumption and price. The application of the tests resulted in the indication that the two series have the same integration order: $I(1)$. With the information about the integration order, the Johansen's tests are applied to verify if the series have cointegration relationship. The tests application indicate that exists long run relationship between the variables.

Keywords: Time series, cointegration, Unit roots tests.

Introdução

A escolha correta de um modelo que descreva o comportamento de uma série temporal depende, além de outros fatores, da análise da sua estacionariedade. Segundo Gujarati e Porter (2011), uma série é estacionária se sua média e sua variância são invariantes ao longo do tempo e o valor da covariância entre dois períodos depende somente do grau de defasagens entre as

¹ Mestrando em Modelagem Computacional pela Universidade Federal do Rio Grande; e-mail: a.garcia.ee@gmail.com.

² Professora Doutora do Programa de Pós Graduação em Modelagem Computacional na Universidade Federal do Rio Grande; e-mail: viviane.leite.mattos@gmail.com.

³ Professora Doutora do Departamento de Informática e Estatística na Universidade Federal de Santa Catarina; e-mail: andreack@gmail.com.

observações e não do período efetivo de tempo em que a covariância é calculada. A presença de tendência estocástica ou raiz unitária em séries temporais, caracterizando a não estacionariedade, faz com que as previsões se tornem mais imprecisas conforme o distanciamento do último ponto da amostra aumente.

Além disso, séries com tendência estocástica levam ao problema da regressão espúria. Durante muito tempo utilizou-se procedimentos de diferenciação para evitar o esse tipo de problema. Porém, segundo Engle e Granger (1987), estes procedimentos removem toda a lógica de equilíbrio da teoria econômica, razão pela qual foi introduzido o conceito de cointegração nessa teoria.

A base do raciocínio da cointegração é que mesmo que as séries sejam não estacionárias, pode existir uma relação de equilíbrio a longo prazo entre elas, de modo que se diz que essas séries cointegram. Para que seja possível utilizar o conceito de cointegração, deve-se garantir que as séries analisadas possuam a mesma ordem de integração, que representa o número de vezes que é necessário diferenciar a série para que ela se torne estacionária. Quando é necessário diferenciar d vezes uma série para que ela se torne estacionária, essa série é dita integrada de ordem d , sendo denotada como $I(d)$.

Deste modo, o primeiro passo para a aplicação da cointegração consiste na verificação da existência de raiz unitária nas séries estudadas, o que pode ser feito por meio dos testes Dickey-Fulley (DF) e Dickey-Fulley Aumentado (ADF), os quais consideram como hipótese nula a presença de raiz unitária nos dados, indicando não estacionariedade. Outro teste utilizado com essa finalidade é o teste de Kwiatkowski-Phillips-Schmidt-Shin (KPSS), que considera como hipótese nula a ausência de raiz unitária, indicando que a série é estacionária. Com a informação sobre a ordem de integração de cada série envolvida na análise, aplica-se o procedimento de Johansen para testar a existência de relação de longo prazo entre as variáveis.

Na modelagem do consumo de energia elétrica, com o intuito de avaliar a relação de variáveis a longo prazo, essa técnica tem sido bastante utilizada, como por exemplo em Andrade e Lobão (1997), Schmidt e Lima (2004) e Mattos e Lima (2005), sendo fundamental para sua aplicação verificar se as séries estudadas possuem a mesma ordem de integração. Nesse sentido, o presente estudo é desenvolvido com o objetivo de avaliar a relação entre consumo e tarifa de energia elétrica residencial, atualizando estudos já desenvolvidos por considerar os últimos 30 anos, além de analisar o desempenho das técnicas utilizadas.

1 Metodologia

Os dados sobre o consumo e a tarifa de energia elétrica do setor residencial no Brasil entre os anos de 1985 e 2014 têm como fonte Eletrobrás (2016) e são disponibilizados pelo Instituto de Pesquisas Econômicas Aplicadas (IPEA) por meio de seu banco de dados, o IPEADATA. Inicialmente é feita uma transformação logarítmica nas séries, seguida de análise gráfica de ambas em nível e em primeira diferença. Após, são realizados os testes para verificar a estacionariedade das séries e determinar sua ordem de integração: testes DF e ADF, mostrados na seção 1.1, e KPSS, explanado na seção 1.2. Com a informação sobre a ordem de integração de cada série, aplica-se o teste de Johansen, abordado na seção 1.3, para verificar a existência de cointegração entre as variáveis. Todos esses procedimentos são realizados com o auxílio do software livre *Gnu Regression, Econometrics and Times-Séries Library (Gretl)*.

1.1 Testes de Dickey-Fulley

O teste de Dickey-Fulley considera o modelo:

$$Y_t = \rho Y_{t-1} + u_t \quad -1 \leq \rho \leq 1 \quad t = 1, 2, 3, \dots, N, \quad (1)$$

em que Y_t é a série a ser modelada, ρ o coeficiente de Y_{t-1} e u_t é um termo de ruído branco.

Para a realização do teste, a Eq. (1) é trabalhada na forma de diferenças, conforme Eq. (2):

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + u_t = (\rho - 1)Y_{t-1} + u_t \quad (2)$$

que pode ser escrita conforme Eq. (3):

$$\Delta Y_t = \delta Y_{t-1} + u_t \quad (3)$$

onde $\delta = \rho - 1$ e Δ é o primeiro operador de diferença.

Em Dickey e Fuller (1981) é demonstrado que, sob a hipótese nula de que $\delta = 0$, o valor estimado do coeficiente de Y_{t-1} na Eq. (3), segue a estatística τ (tau). Mediante a utilização da simulação de Monte Carlo, os autores computaram os valores fundamentais dessa estatística.

Na prática, testa-se a hipótese de que $\delta = 0$, o que implica em $\rho = 1$, e, conseqüentemente, na presença de uma raiz unitária, ou seja:

$H0$: $\delta = 0$ (o modelo possui raiz unitária, a série não é estacionária);

$H1$: $\delta < 0$ (o modelo não possui raiz unitária, a série é estacionária).

Nos casos de u_t correlacionados, pode-se aplicar o teste de Dickey-Fuller Aumentado, que parte do mesmo princípio mostrado acima, porém considerando valores defasados da variável

dependente ΔY_t . Neste teste define-se o número de defasagens das diferenças empiricamente, primeiro escolhendo um valor elevado e verificando a significância da última defasagem. Diminui-se então o número de defasagens até que a última seja estatisticamente significativa. Após encontrar o número ótimo de defasagens, parte-se para a inclusão de constante e tendência, onde são testados os seguintes modelos, baseados em Dickey e Fuller (1981): sem constante ou tendência - Eq. (4), com a inclusão de constante (β_1) - Eq. (5), ou ainda com constante e tendência ($\beta_1 + \beta_2 t$) - Eq. (6).

$$\Delta Y_t = \delta Y_{t-1} + u_t \quad (4)$$

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t \quad (5)$$

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t \quad (6)$$

Compara-se o valor da estatística τ com valores tabelados para servir de suporte à tomada de decisão. Nesse estudo adotou-se o nível de significância de 5%.

1.2 Teste de KPSS

O teste KPSS, proposto em Kwiatkowski et al. (1992), avalia a estacionariedade considerando um modelo com tendência, passeio aleatório e erro, conforme Eq. (7):

$$Y_t = \xi t - r_t + \varepsilon_t \quad (7)$$

em que ξ é uma constante que representa o coeficiente de t , ε_t é um erro estacionário e um passeio aleatório $r_t = r_{t-1} + u_t$, onde u_t é iid. com distribuição normal, média zero e variância σ^2 ; e Y_t a variável estudada.

Considerando que ε_t representa os resíduos de uma regressão em Y_t , explicado pelas componentes de tendência, passeio aleatório e intercepto, um estimador para a variância dos erros dessa regressão, de acordo com Kwiatkowski et al (1992), é dado pela Eq. (8).

$$\hat{\sigma}_\varepsilon^2 = \frac{SQE}{N} \quad (8)$$

onde SQE é a soma dos quadrados dos erros e N é o número de observações.

Define-se a soma parcial dos resíduos, S_t , conforme Eq. (9),

$$S_t = \sum_{i=1}^t \varepsilon_i, \quad t = 1, 2, 3, \dots, N \quad (9)$$

de modo que a estatística de teste LM é dada pela Eq. (10):

$$LM = \sum_{t=1}^N \frac{S_t^2}{N^2 \hat{\sigma}_\varepsilon^2} \quad (10)$$

onde $\hat{\sigma}_\varepsilon^2$ é a estimativa da variância de longo prazo de Y_t . Essa estatística é utilizada para testar as hipóteses:

$H0: \sigma_u^2 = 0$ (o modelo não possui raiz unitária, a série é estacionária);

$H1: \sigma_u^2 > 0$ (o modelo possui raiz unitária, a série não é estacionária).

Os valores da estatística de teste LM são comparados com valores tabelados para dar suporte à tomada de decisão. Nesse estudo foi adotado o valor de 5% como nível de significância.

1.3 Ordem do modelo VAR

Antes de realizar o teste de cointegração é preciso definir a ordem do modelo VAR (Vetores Autorregressivos) a ser utilizado. Para isso, são consideradas várias defasagens, sendo escolhida a que minimiza os resultados dos testes de critério de informação de Akaike (CIA) (AKAIKE, 1973), Schwarz (CIS) (SCHWARZ et al., 1978) e Hannan-Quinn (CIHQ) (HANNAN; QUINN, 1979), conforme as Eqs. 11, 12 e 13:

$$\text{LnCIA} = \frac{2k}{N} + \text{Ln}\left(\frac{SQR}{N}\right) \quad (11)$$

$$\text{LnCIS} = \frac{k}{N} \text{Ln}(N) + \text{Ln}\left(\frac{SQR}{N}\right) \quad (12)$$

$$\text{LnCIHQ} = \frac{k}{N} 2\text{Ln}(\text{Ln}(N)) + \text{Ln}\left(\frac{SQR}{N}\right) \quad (13)$$

onde:

N é o número de observações;

k é o número de regressores; e

SQR é a soma dos quadrados dos resíduos.

Caso sejam encontradas respostas diferentes entre os três critérios, é preciso escolher um deles, o que é feito levando-se em conta a possível perda de graus de liberdade.

1.4 Teste de cointegração

De acordo com Enders (2008), o procedimento de Johansen está baseado na relação entre o posto de uma matriz e suas raízes características e pode ser visto como um teste de raiz unitária de Dickey-Fuller para n variáveis. O procedimento consiste em estimar os vetores de cointegração através de um vetor autorregressivo, por máxima verossimilhança.

Para que seja possível demonstrar o procedimento de Johansen em um teste de cointegração de n variáveis, consideremos o vetor x com d defasagens, conforme a Eq. 14:

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + \dots + A_d x_{t-d} + \zeta_t, \quad (14)$$

onde:

x_t = vetor de cointegração de variáveis ($n \times 1$);

A_i = matriz dos coeficientes ($n \times n$); e

ζ_t = vetor de erros ($n \times 1$).

Subtraindo x_{t-1} de ambos os lados da Eq. (14), tem-se:

$$\Delta x_t = (A_1 - I)x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + \dots + A_d x_{t-d} + \zeta_t.$$

Agora, somando e subtraindo $(A_1 - I)x_{t-2}$, tem-se:

$$\Delta x_t = (A_1 - I)\Delta x_{t-1} + (A_2 + A_1 - I)x_{t-2} + A_3 x_{t-3} + \dots + A_d x_{t-d} + \zeta_t.$$

Continuando, somando e subtraindo $(A_2 + A_1 - I)x_{t-3}$, tem-se:

$$\Delta x_t = (A_1 - I)\Delta x_{t-1} + (A_2 + A_1 - I)\Delta x_{t-2} + (A_3 + A_2 + A_1 - I)x_{t-3} + \dots + A_d x_{t-d} + \zeta_t.$$

Seguindo, obtém-se a Eq. (15):

$$\Delta x_t = \sum_{i=1}^{d-1} \pi_i \Delta x_{t-i} + \pi x_{t-d} + \zeta_t \quad (15)$$

onde:

$$\pi = -\left(I - \sum_{i=1}^d A_i\right) \quad \text{e} \quad \pi = -\left(I - \sum_{j=1}^i A_j\right)$$

A Eq. 15 representa o posto da matriz π , que é igual ao número de vetores independentes de cointegração. Caso o posto seja zero, a matriz é nula, não existe cointegração e as variáveis são estacionárias. Caso o posto seja um, existe apenas um vetor de cointegração. Nos demais casos ($1 < \text{posto}(\pi) < n$) existem múltiplos vetores de cointegração. Como o posto de uma matriz é igual ao número de suas raízes características diferentes de zero, a análise dos vetores de cointegração pode ser feita através da análise de significância das raízes estimadas da matriz π . Dessa maneira, o teste de cointegração proposto por Johansen é composto pela Eq. (16) e pela Eq. (17):

$$\lambda_{\text{traces}}(r) = -N \sum_{i=r+1}^n \text{Ln}(1 - \hat{\lambda}_i) \quad (16)$$

$$\lambda_{\text{max}}(r, r+1) = -N \text{Ln}(1 - \hat{\lambda}_{r+1}) \quad (17)$$

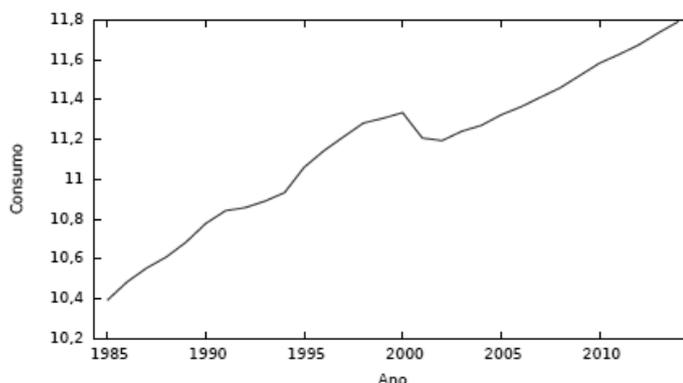
onde $\hat{\lambda}_i$ são os valores estimados das raízes características (autovalores), obtidos da matriz π estimada; e N é o número de observações.

Na Eq. (16) assume-se como hipótese nula a existência de r^* vetores de cointegração e como hipótese alternativa $r > r^*$ vetores de cointegração. A Eq. (17) é usada para testar a hipótese existência de r vetores de cointegração contra a hipótese alternativa de $r+1$ vetores.

2 Resultados e discussão

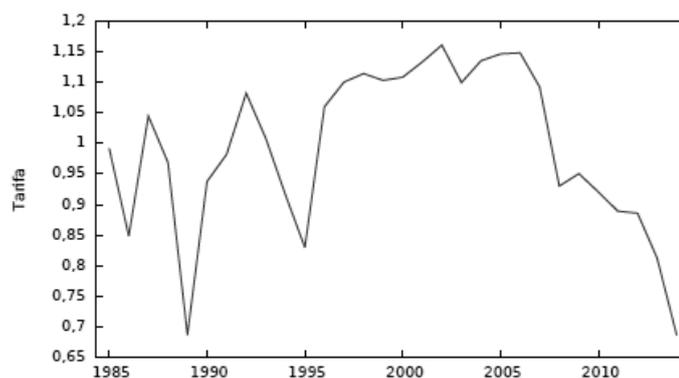
A partir da representação de uma série temporal por um gráfico em linhas, é possível fazer uma avaliação inicial sobre a necessidade de inclusão ou não de tendência e constante. O gráfico da série temporal dos dados originais para o consumo pode ser visto na Fig. (1) e para a tarifa, na Fig. (2).

Figura 1 - Consumo de energia ao longo do tempo



Fonte: Os autores.

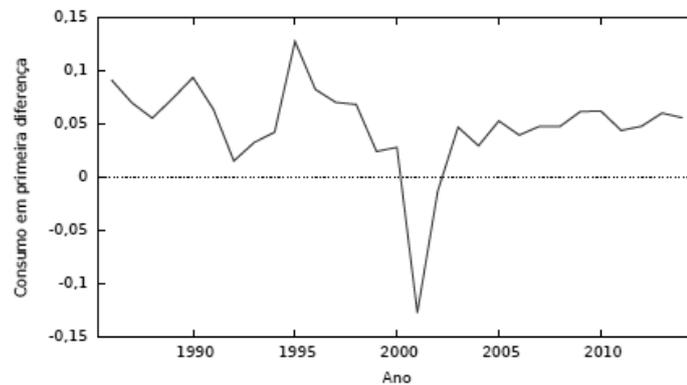
Figura 2 - Tarifa da energia ao longo do tempo



Fonte: Os autores.

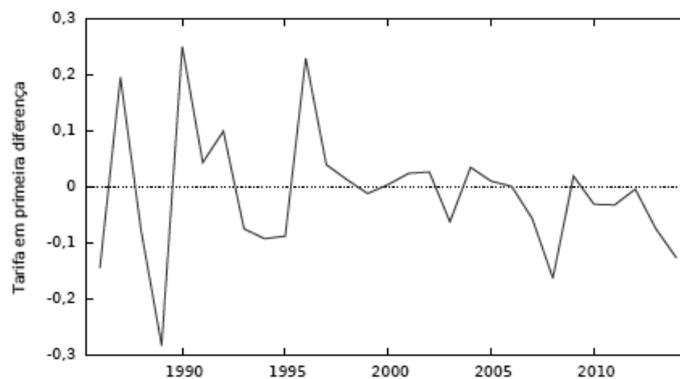
As séries em primeira diferença para o consumo e para a tarifa podem ser observadas nos gráficos em linha na Fig. (3) e na Fig. (4), respectivamente.

Figura 3 - Variação do consumo de energia ao longo do tempo.



Fonte: Os autores.

Figura 4 - Variação da tarifa de energia ao longo do tempo.



Fonte: Os autores.

Este número foi encontrado conforme descrito na seção 1.1, resultando na inclusão de uma defasagem para a variável consumo e zero defasagens para a variável tarifa. Em seguida, foram estimadas as equações Eq. (4), Eq. (5) e Eq. (6), verificando se a constante e a tendência eram significativas. Para a variável consumo, a Eq. (4) apresentou o coeficiente de Y_{t-1} positivo, devendo ser rejeitada, pois implicaria que a série seria explosiva. Foi verificado que tanto a constante quanto a tendência são significativas. Para a variável tarifa, apenas a constante foi significativa. Os p-valores encontrados pelos testes ADF e KPSS são apresentados na Tab. (1)

O teste de KPSS apresentou constante e tendência significativas para ambas as variáveis e os p-valores também são apresentados na Tab. (1).

Tabela 1 - Testes de estacionariedade para as séries em nível

Teste	p-valor (consumo)	p-valor (tarifa)
ADF	> 0.05	> 0.05
KPSS	< 0.05	< 0.05

Fonte: Elaborada pelos autores.

Em seguida, foram realizados os mesmos testes para as variáveis em primeira diferença. Tanto para o teste ADF, quanto para o KPSS apenas a constante se apresentou significativa. Os resultados dos testes para as variáveis em primeira diferença são apresentados na Tab. (2).

Tabela 2 - Testes de estacionariedade para as séries em primeira diferença

Teste	p-valor (consumo)	p-valor (tarifa)
ADF	< 0.05	< 0.05
KPSS	> 0.05	> 0.05

Fonte: Elaborada pelos autores.

De acordo com as informações apresentadas na Tab. (1) e na Tab. (2), pode-se notar que ambos os testes apresentam indícios de que ambas as séries se apresentam como não estacionárias para os dados originais e estacionárias em primeira diferença. Considerando nível de significância 5%, no primeiro caso, com os dados originais, segundo o teste ADF (p-valor > 0,05), não são encontradas evidências de que as séries são estacionárias. Já, segundo o teste KPSS (p-valor < 0,05), são encontradas evidências de que as séries não são estacionárias. Os testes com os dados em primeira diferença permitem chegar às seguintes conclusões: segundo o teste ADF (p-valor < 0,05), são encontradas evidências de que as séries são estacionárias; já para o teste KPSS (p-valor > 0,05), não são encontradas evidências de que as séries não são estacionárias.

Sendo ambas as séries I(1), são aplicados os testes das para a definição do número de defasagens a serem incluídas no modelo VAR, cujos resultados podem ser observados na Tab. (3)

Tabela 3 - Aplicação dos testes dos critérios de informação

Defasagem	CIA	CIS	CIHQ
1	-5,14	-4,85	-5,05
2	-4,92	-4,44	-4,78

Fonte: Elaborada pelos autores.

De acordo com a Tab. (3), os três critérios de seleção apontam que o modelo a ser escolhido deve conter uma defasagem.

Com o modelo definido, considerando uma defasagem, são aplicados os testes de Johansen para verificar a existência de cointegração entre as variáveis. Os resultados dos testes de λ_{trace} e λ_{max} são apresentados na Tab. (4).

Tabela 4 - Testes de Johansen

Número de Vetores	Autovalor	λ_{trace} (p-valor)	λ_{max} (p-valor)
0	0,70	< 0,05	< 0,05
1	0,12	> 0,05	> 0,05

Elaborada pelos autores.

Através dos resultados apresentados na Tab. (4), é possível concluir que tanto o teste λ_{trace} , quanto o λ_{max} apontam que existe um vetor de cointegração entre as variáveis Consumo e Tarifa. O vetor de cointegração entre as variáveis pode ser visualizado na Tab. (5):

Tabela 5 - Vetor de cointegração normalizado

Consumo	Tarifa	Constante
1	6,68	-19,54

Fonte: Elaborada pelos autores.

O vetor de cointegração indica que o aumento da tarifa impacta de maneira a diminuir o consumo de energia elétrica.

3 Conclusões

Os testes utilizados para avaliar a estacionariedade trabalham com hipóteses opostas e podem ser considerados complementares. Para tomar a decisão com mais segurança, é indicado que se realize ambos os testes, de modo que quando seja rejeitada hipótese nula do ADF e aceita a do KPSS, há fortes indícios de que a série seja estacionária. Já no caso em que a hipótese nula do teste ADF é aceita e a do KPSS é rejeitada, há indícios de que a série seja não-estacionária.

No presente estudo, os resultados obtidos pelos dois testes permitem chegar à conclusão de que existem evidências de que as séries estudadas são integradas de ordem um, portanto as séries estão aptas para aplicação de testes de cointegração.

Ao aplicar os testes de cointegração propostos por Johansen (1988), conclui-se que ambos indicam que existe um vetor de cointegração entre as séries, apontando a existência de relação de longo prazo entre as variáveis. O vetor normalizado, resultante do teste, mostra que há indícios que a tarifa impacta negativamente no consumo de energia elétrica.

Como sugestão para trabalhos futuros, indica-se a inclusão de novas variáveis no modelo, como renda e preços de bens eletrodomésticos, por exemplo. Além disso, recomenda-se a aplicação de um modelo de correção de erros para ajustar a relação de curto prazo entre as

variáveis e possibilitar a realização de previsões, uma vez que o vetor de cointegração está relacionado apenas à relação de longo prazo.

Referências

- AKAIKE, H. Maximum likelihood identification of Gaussian autoregressive moving average models. **Biometrika**, v. 60, n. 2, p. 255-265, 1973.
- ANDRADE, T. A.; LOBÃO, W.J.A. **Elasticidade renda e preço da demanda residencial de energia elétrica no Brasil**. 1997.
- DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. **Econometrica: Journal of the Econometric Society**, p. 1057-1072, 1981.
- ELETROBRÁS, 2016. “**Boletim siese: consumo de energia elétrica no brasil**”. Disponível em: <http://www.ipeadata.gov.br>. Acesso em 20-Jul.-2016.
- ENDERS, W., 2008. **Applied econometric time series**. John Wiley & Sons.
- GUJARATI, D.N. and PORTER, D.C., 2011. **Econometria Básica-5**. AMGH Editora.
- HANNAN, E.J. and QUINN, B.G., 1979. “The determination of the order of an autoregression”. **Journal of the Royal Statistical Society. Series B (Methodological)**, pp. 190–195.
- JOHANSEN, S., 1988. “Statistical analysis of cointegration vectors”. **Journal of Economic Dynamics and Control**, VOL. 12, NO. 2, PP. 231–254.
- KWIATKOWSKI, D., PHILLIPS, P.C., SCHMIDT, P. and SHIN, Y., 1992. “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?” **Journal of econometrics**, Vol. 54, No. 1-3, pp. 159–178.
- MATTOS, L.B.d. and LIMA, J.E.d., 2005. “Demanda residencial de energia elétrica em minas gerais: 1970-2002”. **Nova Economia**, Vol. 15, No. 3, pp. 31–52.
- SCHMIDT, C.A.J. and LIMA, M.A., 2004. “A demanda por energia elétrica no brasil”. **Revista brasileira de economia**, Vol. 58, No. 1, pp. 68–98.
- SCHWARZ, G. et al., 1978. “Estimating the dimension of a model”. *The annals of statistics*, Vol. 6, No. 2, pp. 461–464. Engle, R.F. and Granger, C.W., 1987. “Co-integration and error correction: representation, estimation, and testing”. **Econometrica: journal of the Econometric Society**, pp. 251–276.