



MODELO ARIMA PARA COMPREENSÃO DAS VARIAÇÕES DE OFERTA E DEMANDA DO GÁS NATURAL LIQUEFEITO NO BRASIL

Bruno Dantas Michelena¹

Karla Patrícia Santos Oliveira Rodríguez Esquerre²

Tarssio Brito Barreto³

Salvador Ávila Filho⁴

Resumo

O Gás Natural Liquefeito (GNL) se tornou um produto com grande importância no que diz respeito à introdução de novas fontes de energia para a geração de energia elétrica, muito em função da grande flexibilidade que existe nesse mercado. O Brasil iniciou as atividades de regaseificação de GNL em 2009. Desde seu início, houve muita oscilação na quantidade oferta de gás natural liquefeito, bem como a demanda de energia elétrica. Este trabalho teve por objetivo descrever as séries temporais mensais de oferta de GNL e demanda de energia elétrica a partir do uso de gás natural, de modo a compreender as características existentes nessas duas séries. Para realizar essa descrição, foi desenvolvido um modelo ARIMA com a utilização do Software R, onde foi encontrado como melhor modelo para descrever os dados o modelo ARIMA (0,1,0), modelo este conhecido como random walk, com características de grande aleatoriedade, essa informação possibilita a tentativa de compreensão das variações de oferta e demanda do GNL.

Palavras-chave: Gás Natural Liquefeito, Séries Temporais, ARIMA, Oferta, Demanda.

Abstract

Liquefied Natural Gas (LNG) has become a product of great importance with regard to the introduction of new sources of energy for the generation of electric energy, much due to the great flexibility that exists in this market. Brazil started its LNG regasification activities in 2009. Since its inception, there has been a lot of fluctuation in the quantity of liquefied natural gas supplied, as well as the demand for electric energy. The objective of this work was to describe the monthly time series of LNG supply and demand for electricity from the use of natural gas in order to understand the characteristics of these two series. To perform this description, an ARIMA model was developed with the use of Software R, where it was found as the best model to describe the data ARIMA model (0,1,0), this model known as random walk, with characteristics of great randomness, this information enables an attempt to understand LNG supply and demand variations.

Keywords: Liquefied Natural Gas, Time Series, ARIMA, Supply, Demand.

¹ Programa de Pós-Graduação em Engenharia Industrial da Universidade Federal da Bahia (PEI-UFBA), brunomichelena@gmail.com.

² Universidade Federal da Bahia (UFBA), karlaesquerre@ufba.br.

³ Universidade Federal da Bahia (UFBA), tarssio.disap@hotmail.com.

⁴ Universidade Federal da Bahia (UFBA), avilasal@ufba.br.



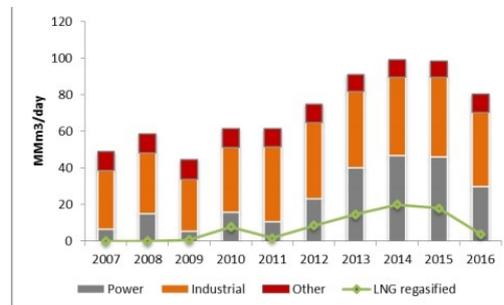
Introdução

A cadeia produtiva do gás natural (GN) brasileiro tem participado de uma evolução constante, recebendo investimentos, melhorias operacionais e logísticas ao longo dos últimos anos, porém, também, vivenciou momentos de crise, principalmente no que diz respeito ao fornecimento de gás proveniente da Bolívia. Nesse sentido, houve um grande avanço na diversificação das fontes de gás natural para a produção nacional e o Gás Natural Liquefeito (GNL) foi inserido no contexto brasileiro em 2009 para ampliar a oferta de gás natural do país, bem como reduzir a dependência em relação ao fornecimento boliviano (ANP, 2010). A demanda de gás natural é representada em boa parte pelo consumo da indústria e a geração de energia elétrica, de acordo com o Boletim Mensal de Gás Natural do Ministério de Minas e Energia (MME), 87% (média dos boletins lançados em 2017) do GN consumido é proveniente dessas duas fontes. A oferta e demanda do GNL apresentou uma variação ao longo dos últimos anos como pode ser observado na Figura 1. Segundo Prade (2015), o Brasil realiza a importação de GNL basicamente para suprimento da demanda de energia elétrica.

O aumento na demanda trouxe alguns aspectos interessantes para a produção e distribuição do gás natural no Brasil. Primeiro, houve uma forte regulamentação de suas características para manter a qualidade exigida pelos mercados, atividade essa desenvolvida principalmente pela Agência Nacional de Petróleo, Gás Natural e Biocombustíveis – ANP. Em seguida houve um impulso governamental na criação dos terminais de regaseificação de Gás Natural Liquefeito (GNL), de modo a ampliar a produção de gás natural e atender as demandas nacionais constantemente crescentes pela redução nos outros modos de geração de energia.

A regaseificação de GNL é uma atividade que ocorre em alguns países há um tempo e no Brasil se iniciou por volta de 2009 com as criações dos Terminais de Regaseificação de Pecém com capacidade de regaseificação de 7 MMm³/dia (milhões de metros cúbicos por dia) e da Bahia de Guanabara com capacidade inicialmente de 14 MMm³/dia, porém com uma ampliação para 20 MMm³/dia. Mais tarde, em 2014, houve a construção do Terminal de Regaseificação da Bahia com capacidade de 14 MMm³/dia e a capacidade total de regaseificação brasileira atingiu 41 MMm³/dia superando os 32 MMm³/dia do suprimento boliviano (PETROBRAS, 2014).

Figura 1 – Demanda de Gás Natural e GNL regaseificado



Fonte: Autores

Uma abordagem analítica quanto a essas variações se faz necessária, pois essa cadeia passou por diversas mudanças e influências nos últimos anos, enquanto em 2014 houve um recorde de consumo de gás natural, onde foi atingido o número de 99,3 MMm³/dia, com a indústria consumindo 42,9 MMm³/ dia e a geração de energia elétrica utilizando 46,8 MMm³/dia desse montante. A crise econômica brasileira trouxe alguns impactos e o consumo foi reduzido para 80,3 MMm³/dia, com a indústria consumindo 40,8 MMm³/dia e a geração de energia elétrica 29,6 MMm³/dia no final de Abril de 2016 (GOMES, 2017).

A avaliação de características temporais tem sido pouco utilizada em pesquisas envolvendo o Gás Natural, tornando esta pesquisa um desafio. Os modelos de séries temporais baseiam-se nas variações ocorridas ao longo do tempo nas variáveis econômicas. Dentre os objetivos das análises de séries temporais temos: a investigação do mecanismo gerador da série; realização de previsões de valores futuros de curto e longo prazo; descrição do comportamento da série, verificação gráfica de existências de tendências, ciclo e variações sazonais, busca de periodicidade nos dados (MORETTIN; TOLOI, 2004).

Neste contexto, o ambiente R tem se demonstrado promissor para a modelagem e análise de fenômenos temporais que possuem características atreladas a questões econômicas e políticas.

Objetivo

O objetivo desse trabalho é descrever o comportamento da série temporal de oferta e demanda de GNL buscando analisar suas características frente ao contexto econômico do Brasil.

Material e Método

Neste trabalho foi analisada a variação da oferta e demanda do GNL, mensalmente, durante os períodos de janeiro de 2009 até junho de 2017, totalizando 102 observações de cada uma das variáveis (oferta e demanda). Os dados referentes à demanda do GNL são os dados de demanda para geração de energia elétrica, visto que o GNL é regaseificado basicamente para esta finalidade. Todos os dados foram retirados do Boletim Mensal do Gás Natural, informações disponíveis no site do Ministério de Minas e Energia do Brasil (<http://www.mme.gov.br>).

Em um primeiro momento foi realizada uma análise descritiva dos dados observando estatísticas básicas dos mesmos.

Em seguida foi utilizado o modelo ARIMA (AutoRegressive Integrated Moving Average) ou modelo de Box & Jenkins que representa um conjunto de processos estocásticos, onde em cada instante de tempo t existe uma série de valores que a série pode assumir, de acordo com possibilidades de ocorrência. Esse modelo foi utilizado para auxiliar a compreensão das características temporais observadas na oferta e demanda do GNL. Todos os testes envolvidos no trabalho foram realizados com o auxílio do Software R i386 3.4.2.

ESTACIONARIDADE E TESTE DE RAÍZ UNITÁRIA

Segundo Box e Jenkins (1976), para a construção do modelo ARIMA é necessário a realização de três etapas: 1) Identificação; 2) Estimação; e 3) Verificação. Este modelo é adequado para a previsão de séries temporais cujo processo estocástico não é estacionário.

Os modelos AR, MA e ARMA, são apropriados para descrever séries temporais estacionárias, isto é, séries que se desenvolvem no tempo ao redor de uma média constante. Em muitos casos, as séries encontradas não são estacionárias, mas quando tomamos a série diferenciada esta se torna estacionária. Os modelos autorregressivos, integrados e de médias móveis (ARIMA) representam a série diferenciada por um modelo ARMA.

Seja Z_t uma série temporal não estacionária. Tomamos $W_t = \Delta Z_t = Z_t - Z_{t-1}$ sendo a série diferenciada uma vez de Z_t , denotamos por $W_t = \Delta^{d*} Z_t$ a série temporal diferenciada d vezes de Z_t .

Podemos representar W_t por um modelo ARMA, como W_t é uma diferença de Z_t então, Z_t é uma integral de W_t , assim dizemos que Z_t segue um modelo autorregressivo, integrado

e de médias móveis, isto é, um modelo ARIMA de ordem (p,d,q) e escrevemos $ARIMA(p,d,q)$ onde p é a ordem da componente autorregressiva, d é o número de diferenças tomadas na série e q é a ordem da componente de médias móveis.

Portanto, podemos descrever todos os modelos vistos anteriormente utilizando a nomenclatura ARIMA, isto é:

i) $ARIMA(p,0,0) = AR(p)$;

ii) $ARIMA(0,0,q) = MA(q)$;

iii) $ARIMA(p,0,q) = ARMA(p,q)$.

No caso de uma série não estacionária, utiliza-se o modelo completo, $ARIMA(p,d,q)$ com d diferenças na série original.

Logo, a série original passará por algumas diferenciações a fim de torná-la estacionária, em muitos casos é comum se chegar à estacionaridade após a primeira diferença (BOX & JENKINS, 1976).

Considerando um modelo autoregressivo de primeira ordem, descrito por:

$$X_t = \phi X_{t-1} + \varepsilon_t \quad (1)$$

A equação (1) mostra que o valor da série X no instante t é igual ao seu valor no tempo $t-1$ multiplicado pela constante ϕ e somado a um choque aleatório, o que caracteriza um modelo autoregressivo de primeira ordem, ou $AR(1)$. O termo ε_t é erro de ruído branco, com média zero e variância σ^2 .

Se $\delta=1$, o processo da equação (1) se torna modelo de passeio aleatório, um exemplo clássico de processo não estacionário. Nesse caso, o valor da série X no instante t é tão somente seu valor no tempo $t-1$ somado a um choque aleatório.

A alternativa foi elaborada por Dickey e Fuller (1979), que mostraram que, sob a hipótese nula de $\delta=0$, o valor t estimado do coeficiente de X_{t-1} segue a estatística τ . Os valores críticos dessa estatística foram calculados por simulações de Monte Carlo, e na literatura o teste tau é conhecido como teste de Dickey-Fuller (DF).

Uma variação desse teste é o teste de Dickey-Fuller Aumentado, para o caso em que o termo de erro ε_s é correlacionado. Como demonstrado na equação (2), este é conduzido com a adição de valores defasados da variável ΔY_t :

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^m a_i \Delta y_{t-i} + \varepsilon_t \quad (2)$$

Onde ε_t é um termo de erro puro de ruído branco, β_1 e β_2 são constantes. O número de lags da variável ΔY_t incluídos é determinado empiricamente, de forma que o termo de erro não possua correlação serial. O termo $\beta_2 t$ inclui uma tendência determinística, enquanto o termo β_1 se aplica para o caso em que há suspeita de que o processo estocástico seja um passeio aleatório com *drift*, em vez de passeio aleatório puro.

Definiu-se o modelo ARIMA como sendo de máxima ordem três, para isto, buscou-se testar as possíveis combinações entre (p,d,q) que foram avaliadas pelo método de Akaike (AIC), visto na equação 3, no qual o menor valor definia a modelagem a ser assumida. Para determinar a ordem de um modelo ARIMA não sazonal, um critério útil é o critério de informação de Akaike. É escrito como:

$$AIC = -2\log(L) + 2(p + q + k + 1) \quad (3)$$

Em que L é verossimilhança dos dados, p é a ordem da parte autoregressiva e q é a ordem da parte de média móvel. O parâmetro k neste critério é definido como o número de parâmetros no modelo sendo ajustado aos dados. Para o AIC, se k=1, então c é diferente de 0 e se k=0, então c é igual 0. O AIC corrigido para modelos ARIMA pode ser escrito como visto na equação 4:

$$AIC_c = AIC + \frac{(2(p + q + k + 1)(p + q + k + 2))}{(T - p - q - k - 2)} \quad (4)$$

Teste de autocorrelação dos resíduos

Foi observado na literatura que o teste de Ljung - Box apresenta melhores resultados do que o teste de Box - Pierce. As hipóteses do teste são as mesmas, na hipótese nula os resíduos são independentes e identicamente distribuídos (i.i.d.), já na hipótese alternativa os resíduos não são i.i.d.

Calculamos as autocorrelações por:

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^n \hat{a}_t^2} \quad (5)$$

É possível mostrar que:

$$\hat{r}_k \sim N\left(0, \frac{1}{n}\right) \quad (6)$$

Se o modelo for apropriado, a estatística do teste

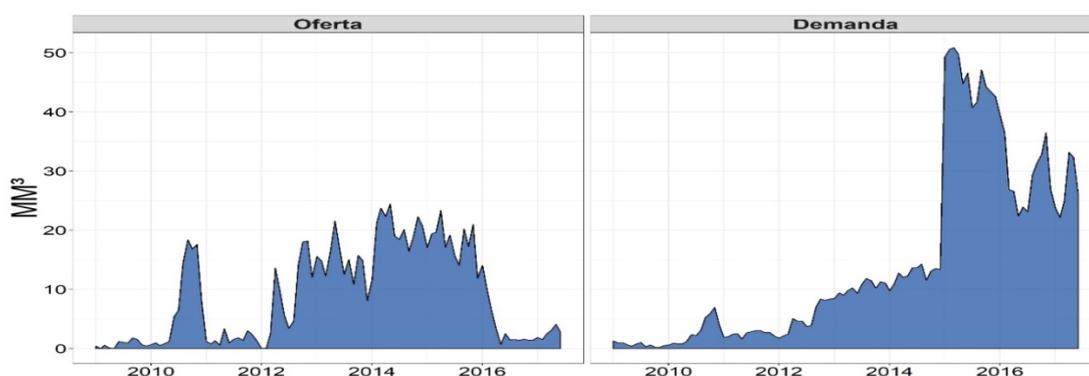
$$Q(k) = n(n-2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n-j)} \quad (7)$$

Terá aproximadamente uma distribuição χ^2 com $(K - p - q)$ graus de liberdade, onde K é o número de defasagens tomada na função de autocorrelação, p e q são as ordens do modelo ajustado. Portanto, rejeitamos a hipótese nula se $Q > \chi_{1-\alpha, k-p-q}^2$ com um nível de significância α .

Resultados e Discussão

Como pode ser observado no Gráfico 1, existe uma grande variação tanto na oferta, como na demanda do gás natural liquefeito, após o ano de 2010, fica evidenciado um aumento da oferta de GNL e isso se dá em função da criação dos terminais de regaseificação da Baía de Guanabara no Rio de Janeiro e de Pecém no Ceará. Em 2014 há um novo aumento na oferta, que se justifica pela criação do terceiro terminal localizado na Ilha dos Frades na Bahia. A partir de 2016 a oferta é reduzida em função de uma combinação de fatores, como a crise econômica, que levou a demanda de energia elétrica a diminuir, bem como o aumento da oferta de energia elétrica através de outras fontes de energia como eólica e biomassa (GOMES, 2017).

Gráfico 1 – Séries de Oferta e Demanda



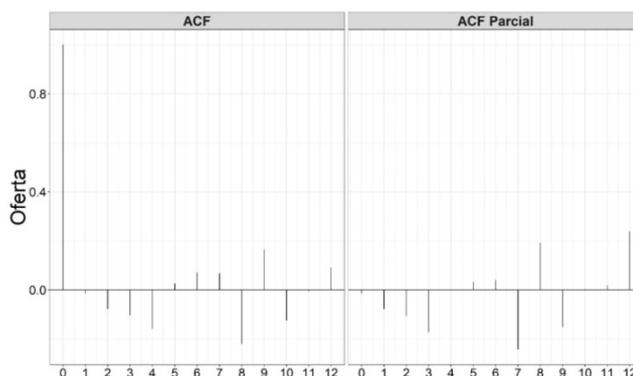
Fonte: Autores.

Quanto à demanda de energia elétrica a partir de gás natural, seu momento de crescimento mais importante se dá entre 2014 e 2015 e sua justificativa é a crise hídrica que o país sofreu nesses anos, ativando as termelétricas. Também é possível avaliar com esse gráfico as características de não-estacionariedade, pois os dados variam na média.

Autocorrelação Total (ACF) e Autocorrelação Parcial (PACF)

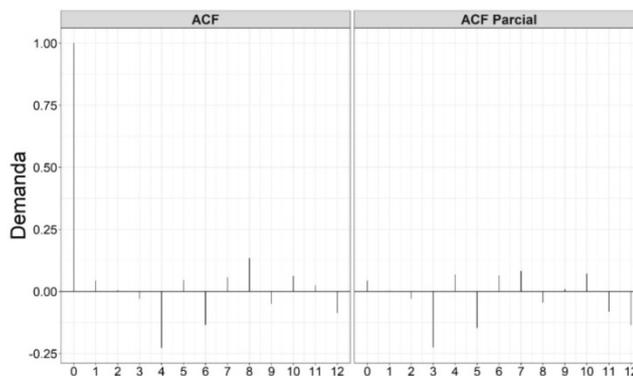
Os testes de autocorrelação total e parcial aplicados nas séries diferenciadas foram realizados para verificar a estacionariedade das séries como pode ser observado nos gráficos 2 e 3. O valor de ACF diminui rapidamente quando os lags aumentam, isso caracteriza a estacionariedade. Já o teste de PACF demonstra a estacionariedade, pois corta o 0 após k defasagens. Em ambos os testes os dados não são autocorrelacionados.

Gráfico 2 – Teste de autocorrelação total e parcial dos dados da oferta



Fonte: Autores.

Gráfico 3 – Teste de autocorrelação total e parcial dos dados da demanda



Fonte: Autores.

Teste de Dickey-Fuller Aumentado (ADF)

Também para realizar a verificação de estacionariedade, foi utilizado o Teste de Dickey-Fuller Aumentado (ADF) que permitiu a compreensão de que a hipótese nula, onde as variáveis possuem uma raiz unitária, ou ordem de integração igual a 1, não é rejeitada. Para os dados referentes à oferta o ADF foi igual a -10,047 e o p-valor igual a 0,01 e já para os dados da demanda -9,3037 com o p-valor 0,01. Quando foi aplicado o teste de hipótese nas séries diferenciadas, há um indício estatístico de se rejeitar a hipótese nula, pois as séries se demonstram estacionárias.

Verificação

Na etapa de verificação, o modelo que melhor representa os dados de oferta foi o ARIMA (0,1,0) com o AICc no valor de 542,5913 e σ^2 igual a 12,36 como pode ser visto na Tabela 1.

O modelo ARIMA (0,1,0) é conhecido como random walk, ou caminho aleatório, que pode ser considerado um caso limite de um modelo AR (1). Esse modelo específico tem algumas características como as mudanças na série são completamente aleatórias, os impactos dos choques aleatórios não decaem ao longo do tempo, pois o modelo conta com uma forte memória, esses choques tem um efeito permanente na série.

Tabela 1 – Identificação do modelo para dados da oferta

Oferta	
Modelos	AICc
ARIMA(2,1,2)	
ARIMA(0,1,0)	544,6688
ARIMA(1,1,0)	546,7718
ARIMA(0,1,1)	546,7677
ARIMA(0,1,0)	542,5913
ARIMA(1,1,1)	546,5553

Fonte: Autores.

A aleatoriedade discutida na verificação do modelo, para a oferta de GNL, pode ser causada porque essa oferta é disponibilizada de acordo com decisões políticas, além de depender de relações comerciais com outros países, seja com os fornecedores de GNL

propriamente ditos, como também a quantidade de gás natural que pode ser ofertado principalmente na relação com a Bolívia.

Na verificação do modelo para a demanda também se identificou como melhor modelo o ARIMA (0,1,0), com AICc no valor de 583,9612 e σ^2 igual a 18,61.

Para a demanda, a aleatoriedade pode ser explicada principalmente por conta da inserção de novas fontes geradoras de energia elétrica nos últimos anos, pois como foi discutido por Gomes (2017), houve um grande avanço em desenvolvimento de fontes alternativas de energia elétrica como a criação de parques eólicos e solares, utilização de biomassa, além das usinas hidrelétricas que representa a fonte de maior geração de energia elétrica no país.

Tabela 2 – Identificação do modelo para os dados da demanda

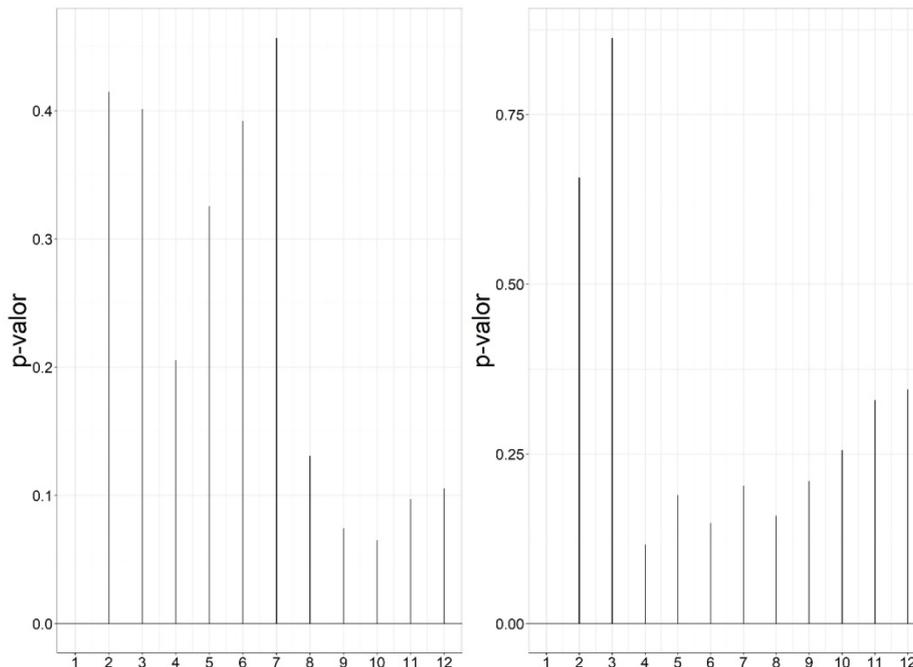
Demanda	
Modelos	AICc
ARIMA(2,1,2)	
ARIMA(0,1,0)	585,7228
ARIMA(1,1,0)	587,6579
ARIMA(0,1,1)	587,6596
ARIMA(0,1,0)	583,9612
ARIMA(1,1,1)	589,8270

Fonte: Autores.

Teste dos resíduos

Utilizando o teste de Ljung-Box (Gráfico 4) para os modelos de oferta e demanda do GNL, fica evidenciado que os lags no período de um ano tem de maneira geral o p-valor menor que 0,05, o que caracteriza a não autocorrelação dos resíduos, ou também comportamento de ruído branco, o que corrobora a ideia de que os modelos de séries temporais em questão representam bem os dados trabalhados.

Gráfico 4 – Teste de Ljung-Box para os resíduos



Fonte: Autores

Conclusão

O presente trabalho descreveu as séries temporais referentes à oferta e demanda do GNL no Brasil de 2009 até 2017, para essa finalidade foi identificado o modelo ARIMA (0,1,0) para ambas situações de avaliação como melhor modelo de representação dos dados. Foi percebido que o modelo em questão se caracteriza por representar dados que são aleatórios e foi realizada uma discussão sobre o que pode causar tamanha aleatoriedade para a oferta e demanda do gás natural liquefeito.

O uso do Software R foi adequado para o processo de entendimento das características das séries em questão, possibilitando o uso das ferramentas necessárias para a criação de um modelo que representa bem os dados discutidos neste artigo.

Os autores recomendam a realização de estudos para compreensão das características temporais de outras fontes de energia elétrica para o sistema de abastecimento brasileiro, de modo a realizar comparações entre os modelos encontrados e investigar as relações de aleatoriedade de cada uma dessas fontes.



Referências

BOX, G.E.P. and JENKINS, G.M. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day. 1976.

DICKEY, D.A. and FULLER, W.A. **Distribution of the estimators for autoregressive time series with a unit root**, Journal of the American Statistical Association, 1979.

GOMES, I. **The vanishing LNG Market in Brazil**. University of Oxford, 2017.

GUJARATI, D. **Basic Econometrics**. 4 ed. The McGraw-Hill Companies, 2004.

HALLACK, M. **Os tempos e os desafios das escolhas atuais da indústria de gás natural no Brasil**. Blog Infopetro, 2016. Acessado em 15 de Agosto de 2017.

INTERNACIONAL ENERGY AGENCY – IEA. **Market Analysis and Forecasts to 2020**. Gas Medium-Term Market Report, 2015.

MINISTÉRIO DE MINAS E ENERGIA – MME. Acessado em Janeiro de 2018.

MORETTIN, P. A., TOLOI, C. M. **Análise de Séries Temporais**. São Paulo: Edgard Blucher. 2004.

PETROBRAS. **“Petrobras inicia operação de seu terceiro Terminal de Regaseificação de GNL”**, 2014.

PORTAL ACTION. Estatcamp- Consultoria em estatística e qualidade, São Carlos - SP, Brasil. Acessado em Março de 2018.

PRADE, Y. C. **O mercado de GNL do futuro: risco ou oportunidade para o Brasil?**. Blog Infopetro, 2015. Acessado em 20 de Fevereiro de 2018.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>, 2017.

WOOLDRIDGE, J. M., **Introductory Econometrics: A modern approach**. 4 ed. Nelson Education, 201

Anexo

Script

```
##### PACKAGES #####
# 01 - Inserindo pacotes
if (!require("dplyr")) install.packages("dplyr") ; library(dplyr)
if (!require("zoo")) install.packages("zoo") ; library(zoo)
if (!require("tseries")) install.packages("tseries") ; library(tseries)
if (!require("forecast")) install.packages("forecast") ; library(forecast)
if (!require("reshape2")) install.packages("reshape2") ; library("reshape2")
if (!require("scales")) install.packages("scales") ; library("scales")
if (!require("ggplot2")) install.packages("ggplot2") ; library("ggplot2")
if (!require("gridExtra")) install.packages("gridExtra") ; library("gridExtra")

# 02 - Carregando Dados #####
dados <- read.delim2(file = "gnv.txt") #read.delim2 <- carrega os dados
gnv <- dados %>%
  mutate(data = as.Date(as.yearmon(dados$data, "%b/%y"))) %>%
  as_tibble()
# 03 - Transformando em Série Temporal #####
gnv <- gnv %>%
  ts(frequency = 12) #transformar em série temporal
# Teste de Autocorrelação (ESTUDAR autocorrelação total e parcial)

# Para Oferta

a <- acf(diff(gnv1), main = "ACF")
b <- pacf(diff((gnv1)), main = "ACF Parcial")
### Para Demanda

c <- acf(diff(gnv2), main = "ACF")
d <- pacf(diff((gnv2)), main = "ACF Parcial")
# Auto Arima
```



```
auto.arima(gnv1, max.order = 3, trace = TRUE, ic = c("aicc"))
```

```
auto.arima(gnv2, max.order = 3, trace = TRUE, ic = c("aicc"))
```

```
# Para os Resíduos
```

```
fit<-arima(gnv1,c(0,1,0))
```

```
lags<-c(seq(from = 1, to = 12, by = 1))
```

```
res1<-resid(fit)
```

```
fit<-arima(gnv2,c(0,1,0))
```

```
res2<-resid(fit)
```

```
l <- LjungBox(res1, lags, order=1) %>%
```

```
l1 <- LjungBox(res2, lags, order=1) %>%
```

Leia também o “*Regulamento para submissão de artigos*” (que resume as informações apresentadas neste Template) disponibilizados no site do evento (www.ser.uff.br).