

PREVISÃO PARA AS VENDAS DE GASOLINA COMUM NO BRASIL: UMA APLICAÇÃO DO MODELO ARIMA COM SAZONALIDADE¹.

Alisson Rocha da Silva²

Anderson Luiz R. Mól.³

RESUMO

O presente trabalho tem como objetivo a construção de um modelo ARIMA com sazonalidade (SARIMA) para fazer previsões das vendas de gasolina comum no Brasil. Os dados foram coletados no site da Agência Nacional do Petróleo e Biocombustíveis (ANP), para o período de janeiro do ano 2000 a dezembro do ano 2010. Visando mensurar a qualidade da estimação, foi feita a previsão das vendas para o período de janeiro a junho de 2011. O modelo foi capaz de prever, com seis meses de antecedência – e 6,39% de erro em relação ao verdadeiro valor observado – as vendas do primeiro semestre de 2011. Para trabalhos futuros, fica a sugestão de prever esta série com outros modelos de séries temporais e utilizar um menor nível de significância na construção dos intervalos de confiança da previsão como forma de estreitar o intervalo de valores previstos.

Palavras-chave: SARIMA; previsões; gasolina comum.

ABSTRACT

This work aims the building of an ARIMA model with seasonality (SARIMA) to make forecasting of brazilian gasoline sales. Data were collected from the National Agency of Petroleum and Biofuels (ANP) for the period January 2000 to December 2010. In order to measure the quality of the estimation was made the forecasting of sales for the period January to June 2011. The model was able to predict six months before - and 6.39% error compared to the actual observed value – the sales of the first half of 2011. For future work, the suggestion is to predict this serie with other time series models and use a lower level of significance in the construction of confidence intervals of prediction as a way of narrowing the range of forecasted values.

Key words: SARIMA, forecasting, gasoline

¹ Trabalho de Conclusão do curso de especialização MBA em Administração Financeira da Faculdade Natalense para o Desenvolvimento do Rio Grande do Norte (FARN).

² Bacharel em Economia pela Universidade Federal do Rio Grande do Norte, campus Natal. Contato: alisson_rocha2003@yahoo.com.br

³ Orientador do Trabalho de Conclusão do curso. Doutor em Finanças Estocásticas, Derivativos e Riscos. Mestre em Finanças Corporativas e Mercados. Professor universitário.

1. INTRODUÇÃO

Há muito tempo os seres humanos tem interesse em prever acontecimentos futuros. Não por acaso, séculos atrás, surgiram as ciências para tentar estudar e prever fenômenos. Essa curiosidade, por assim dizer, possivelmente tenha sido um dos principais motivos para o desenvolvimento das ciências como um todo. Ao longo do tempo, o conhecimento humano foi se desenvolvendo cada vez mais e as próprias previsões foram ficando mais precisas.

Podemos citar vários exemplos de profissionais que lidam freqüentemente com a necessidade de tentar de alguma forma antecipar o futuro: os meteorologistas procuram prever quanto – e se – choverá; astrônomos tentam, por exemplo, prever rotas de cometas; agrônomos se interessam em calcular qual a produtividade esperada quando se adota uma determinada técnica agrícola, e assim por diante.

Na economia e nos negócios, governos e empresários procuram fazer estimativas que os auxiliem na obtenção dos seus objetivos, seja a maximização do bem-estar da população – no caso do governo – ou da maximização do lucro, no caso do empresário.

Desta maneira, as empresas procuram prever, por exemplo, as suas receitas futuras, custos, etc. Quanto mais precisas forem estas previsões, mais eficientemente serão utilizados os recursos disponíveis e maior poderá ser o lucro do negócio. Isso ocorre porque uma superestimativa das vendas pode, dentre outras coisas, resultar em altos estoques e, como bem lembra a Microeconomia, uma utilização ineficiente⁴ dos fatores de produção, com impactos óbvios sobre a rentabilidade do negócio. Por outro lado, uma subestimativa das vendas pode resultar, na “melhor” das hipóteses, na geração de *backorders*⁵.

Com o avanço da matemática e da estatística, muito progresso foi feito na área de previsão de valores futuros de variáveis, desde os modelos preditivos iniciais de regressão linear com mínimos quadrados ordinários até o grande desenvolvimento da econometria de séries temporais, tanto com modelos lineares quanto com modelos não lineares, bem como

⁴ Imagine uma empresa que estima – no período t – as suas vendas, para o próximo ano ($t+1$), em 100 milhões e, em cima desta previsão, escolhe uma determinada curva de isoquanta para atender à demanda. Supondo agora que, no período $t+1$, as vendas foram de apenas 50 milhões, ou seja, 50 milhões a menos do que o previsto; com isso, e supondo retornos constantes de escala, a empresa utilizou duas vezes mais fatores de produção do que utilizaria, caso fizesse a previsão das vendas corretamente. Assim, os custos subiriam e o lucro cairia.

⁵ As *backorders* são os pedidos não atendidos pelas empresas, ou seja, há uma oferta do produto menor do que a demanda pelo mesmo. Isso significa que além de vender – e crescer – menos do que poderia, a empresa deixará parte dos seus clientes insatisfeitos com a falta do produto. Em tempos de redes sociais, a insatisfação de um cliente pode rapidamente deixar de ser um problema individual e virar uma enorme “mancha” na reputação da empresa.

modelos paramétricos e não-paramétricos⁶ e, com o desenvolvimento da informática, a utilização intensiva de simulações.

Outros estudos também já se utilizaram da mesma abordagem deste trabalho para fazer previsões de séries temporais. Fiel (2005) modela a previsão de vendas de combustível de um determinado posto de gasolina da cidade de Baião, no Estado do Pará. Lopes (2003) constrói modelos preditivos – dentre eles um SARIMA – para prever a produção de veículos automotores, chegando à conclusão de que, a série estudada, tem forte componente de sazonalidade e que, para a previsão da produção de veículos automotores, o modelo ARIMA com sazonalidade se mostrou melhor do que um modelo de regressão dinâmica. Outro trabalho na área é o de Sabry *et alli* (2007), no qual eles comparam modelos ARIMA – inclusive modelando a sazonalidade - a uma regressão logística, no intuito de prever o volume de tráfego em rodovias egípcias e chegam à conclusão de que o ARIMA mostrou-se mais adequado para prever a variável estudada.

Este trabalho procura fazer previsões mensais para as vendas de gasolina comum⁷, no Brasil. Para isso serão utilizados os dados disponibilizados no site⁸ da Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP). O modelo aqui utilizado será um modelo auto-regressivo integrado e de médias móveis – ARIMA – com sazonalidade ou, simplesmente, SARIMA.

Assim, além desta introdução, o artigo está organizado da seguinte forma: na segunda seção, é feita uma visão geral sobre o mercado de combustíveis no Brasil; na terceira seção, tem-se uma explanação sobre a econometria⁹ de séries temporais, bem como, a descrição matemática dos modelos aqui utilizados; na quarta seção, são apresentados os resultados da modelagem econométrica e; na quinta seção, são feitas as conclusões.

⁶ Segundo Morettin e Toloi (2004, p.01), “os modelos paramétricos são aqueles com um número finito de parâmetros e sua análise é feita no domínio temporal. Nos modelos não-paramétricos, a análise é conduzida no domínio de frequências”.

⁷ A gasolina comum “gasolina C” – no Brasil – é composta de etanol anidro combustível e de gasolina tipo “A” (aquela que é derivada do petróleo). Vez por outra, a proporção de etanol (“álcool”) anidro x gasolina é alterada. Atualmente, essa proporção é de 20% de etanol anidro e 80% de gasolina tipo “A”, conforme resolução do Conselho Interministerial do Açúcar e do Álcool (CIMA), nº 1, de 31 de agosto de 2011, publicada no Diário Oficial da União em 1º de setembro de 2011, sendo admitida uma margem de erro de mais ou menos 1%, conforme disposto no Art. 1º, da Lei 10.203/2001.

⁸ <www.anp.gov.br>. Acesso em: 3 de setembro. 2011.

⁹ Segundo Hill *et alli* (2003) a Econometria pode ser definida como a utilização da teoria e dados da economia, negócios e ciências sociais, juntamente com recursos da estatística, para responder questões do tipo *quanto*.

2. O MERCADO DE COMBUSTÍVEIS

2.1 Visão geral

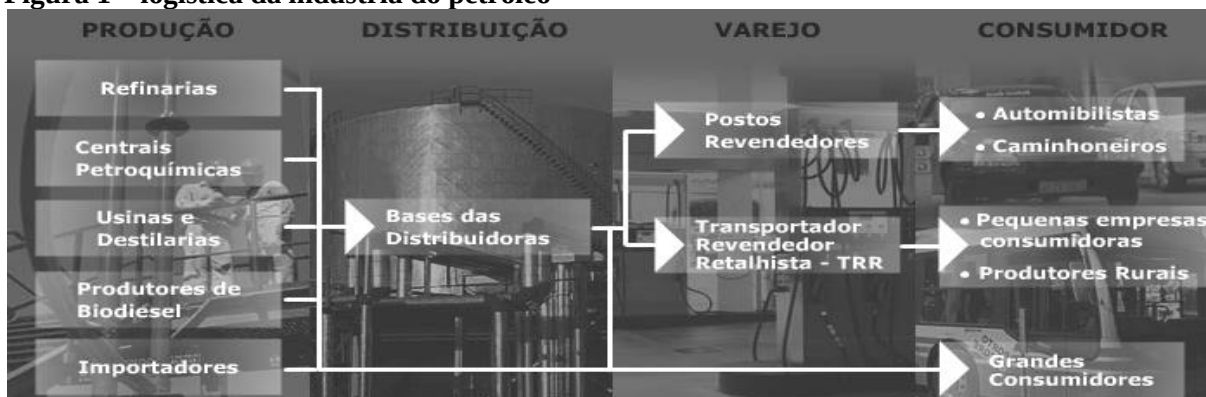
A indústria do petróleo abrange o conjunto de atividades relacionadas com a exploração, desenvolvimento, produção, refino, processamento, transporte, importação, exportação de petróleos, gás natural, outros hidrocarbonetos fluidos e seus derivados (ANP, 2011).

Em 06 de agosto de 1997, iniciou-se o que pode ser tido como uma nova era na indústria brasileira, através da aprovação da Lei 9.478 (“Lei do Petróleo”). Dentre outras importantes mudanças, tal legislação acabou com o monopólio da Petrobras na exploração do petróleo e criou a ANP¹⁰.

Porém, a abertura total do mercado de combustíveis se consolidou em janeiro de 2002, data a partir da qual as distribuidoras puderam usufruir da flexibilidade de importação de derivados de petróleo – via importadores.

A indústria do petróleo requer uma extensa logística. No Brasil, segundo Leal *et alli* (2003), esta logística pode ser subdividida em dois grupos: o setor de *upstream* e o de *downstream*. O primeiro diz respeito às atividades de exploração e produção de petróleo, ao passo que, o segundo engloba a aquisição, armazenagem, comercialização e distribuição de derivados. Uma representação desta estrutura é apresentada na figura a seguir:

Figura 1 – logística da indústria do petróleo



Fonte: SINDICOM (Sindicato Nacional das Empresas Distribuidoras de Combustíveis e Lubrificantes).

¹⁰ A Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), autarquia federal, implantada pelo Decreto nº 2.455, de 14 de janeiro de 1998, é responsável pela execução da política nacional para o setor energético do petróleo, gás natural e biocombustíveis, de acordo com a Lei do Petróleo (Lei nº9478/1997).

Da figura 1 vemos que, dependendo do tipo de combustível, as distribuidoras recorrem ao “canal” apropriado para adquirir o produto. Assim, compram etanol das usinas e destilarias, biodiesel dos respectivos produtores e gasolina das refinarias. Feito isto, revendem o produto aos postos revendedores, ao transportador revendedor retalhista ou para grandes consumidores¹¹.

Este artigo foca no **elo de distribuição** (grifo nosso) da cadeia logística, ou seja, a previsão feita no trabalho serão as vendas – no Brasil – das distribuidoras para os seus consumidores, e não as vendas dos postos revendedores (“postos de combustíveis”) para o consumidor final.

2.2. Distribuição de combustíveis

Em 2010, estavam autorizadas pela ANP 501 bases de distribuição de combustíveis líquidos, sendo que 197 estavam na região Sudeste, 108 no Sul, 67 no Centro-Oeste, 76 no Nordeste e 53 no Norte. A capacidade de armazenamento desta infraestrutura foi de aproximadamente 3,8 milhões m³, sendo que 77,4% se destinaram aos derivados de petróleo (exceto GLP), 18,7% para o etanol e 3,9% para o GLP.

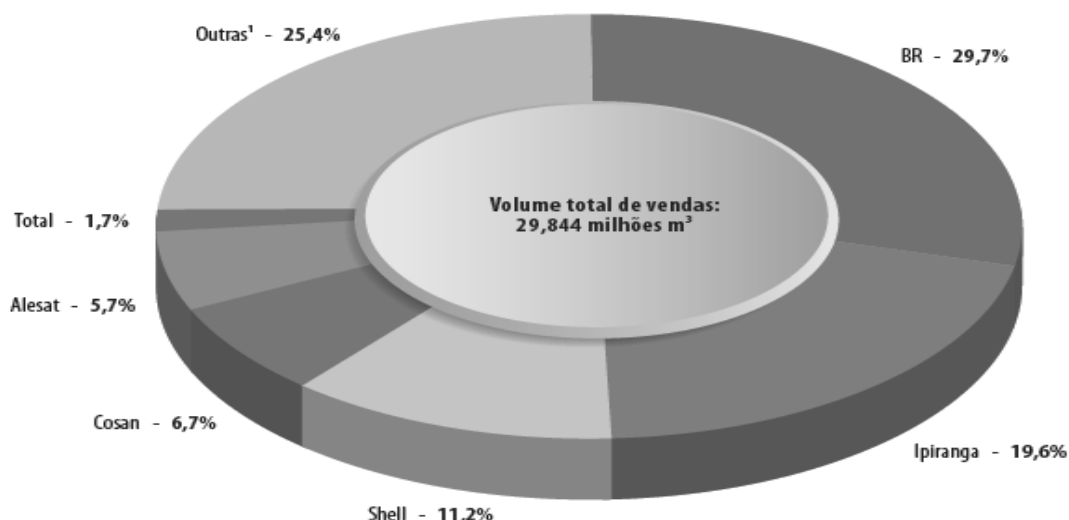
Ainda em relação a 2010 o Produto Interno Bruto (PIB) cresceu 7,5%. Isso se refletiu no crescimento de 8,4% das vendas dos principais combustíveis. Se observarmos as vendas nacionais dos principais derivados de petróleo apenas pelas distribuidoras em 2010, a alta foi de 11,4%, quando comparadas a 2009, atingindo 102,9 milhões m³.

A venda de gasolina comum, foco deste trabalho, cresceu 17,5% entre 2009 e 2010, passando de 25.409 mil m³ para 29.844 mil m³. Todas as regiões tiveram alta nas vendas, com destaque para a Região Nordeste, que cresceu 24,8%. Apenas para constar, no Estado do Rio Grande do Norte a alta foi de 20,71%.

Há uma grande concentração das vendas de combustíveis no Brasil, e com a gasolina comum ocorre o mesmo fenômeno. Isso é fácil de ser percebido no gráfico da página a seguir:

¹¹ As distribuidoras podem realizar contratos individuais com grandes consumidores. Um exemplo disto está no contrato – de exclusividade – da BR Distribuidora com a empresa de mineração Vale (à época Vale do Rio Doce) para o fornecimento do combustível utilizado pela Vale. Este contrato tem um valor aproximado de 11 bilhões de reais. Em um setor com um produto tão homogêneo, o atendimento às necessidades de combustíveis de grandes clientes pode ser um grande filão para as distribuidoras de combustíveis. Mais detalhes a esse respeito ver <<http://www.br.com.br/wps/wcm/connect/3deea48043a7de07a012afecc2d0136c/solucoesBR28.pdf?MOD=AJPERES>> Acesso em: 25 de Out. 2011.

Gráfico 1 – Participação das distribuidoras nas vendas de gasolina comum - 2010



Fonte e elaboração: ANP

¹ Inclui outras 134 distribuidoras

Em 2010, 140 distribuidoras estavam autorizadas a comercializar gasolina comum. Como se observa, as cinco maiores distribuidoras concentraram 72,9% das vendas totais de gasolina comum: BR Distribuidora (29,7%), Ipiranga (19,6%), Shell (11,2%), Cosan (6,7%) e Alesat (5,7%). O restante das vendas se concentra nas demais 135 distribuidoras.

O comércio varejista de combustíveis obteve um faturamento¹² de 223,1 bilhões de reais no ano de 2010, o que demonstra a importância desse setor para o Brasil. Considerando a alta carga tributária dos combustíveis no nosso país, também é fácil perceber o quanto é importante o comércio de combustíveis na arrecadação de impostos e contribuições para os cofres públicos.

Em termos de perspectivas futuras do setor de distribuição de combustíveis, elas foram resumidas pelo presidente da BR distribuidora – José Eduardo Dutra – em uma entrevista¹³ dada em 2008, mas que continua atual. Os principais pontos desta entrevista foram: até 2013, o crescimento do mercado de combustíveis no Brasil deve atingir – pelo menos – uma taxa anual média de 3,6% e; o consumo de álcool anidro, somado ao de álcool hidratado, deve ultrapassar o de gasolina comum.

¹² Fonte: Relatório Anual de Revenda de Combustíveis da FECOMÉRCIO. Disponível em <<http://www.fecombustiveis.org.br/relatorio2011>>. Acesso em: 15 de Out. 2011.

¹³ Esta entrevista está disponível em <http://oglobo.globo.com/economia/mat/2008/04/09/mercado_de_combustivel_no_pais_crescera_3_6_ao_ano_ate_2013-426757294.asp>. Acesso em: 14 de Out. 2011.

Cabe uma observação importante a respeito do último ponto, qual seja a previsão que o consumo do álcool anidro somado ao álcool hidratado ultrapassará o de gasolina comum. No ano da entrevista, ou seja, em 2008, o setor de combustíveis como um todo vinha enfrentando uma mudança estrutural importante, qual seja, a enorme expansão do consumo de etanol. Para termos uma idéia, entre o ano de 2006 e 2008, o consumo de etanol hidratado havia subido 114%, passando de 6.186.553 m³ para 13.290.096 m³. Talvez isso explique o otimismo do presidente da BR Distribuidora ao prever – em 2008 – que o consumo do etanol ultrapassaria o consumo de gasolina. Embora ainda falem dois anos para 2013, no cenário atual é pouco provável que isso aconteça, pois a indústria não conseguiu atender à demanda de etanol, ocasionando grande alta no preço do mesmo e retirando¹⁴ boa parte da sua competitividade frente à gasolina comum.

3. SÉRIES TEMPORAIS

3.1. Estacionariedade

Para Morettin e Tolo (2004), uma série temporal é qualquer conjunto de observações ordenadas no tempo. Assim, os dados de séries temporais são aqueles coletados para uma única variável em múltiplos pontos no tempo. Estas séries podem ser regulares ou irregulares.

O presente artigo analisará uma série regular, ou seja, a diferença temporal entre duas observações é sempre a mesma (mensal). Caso a diferença temporal entre duas observações não fosse sempre a mesma, teríamos uma série irregular. Em um modelo de séries temporais, como por exemplo, um ARIMA, podemos utilizar o comportamento passado da variável para prever o seu valor futuro. Isso simplifica a previsão, já que o modelo terá apenas uma variável (univariado). De acordo com Pindyck e Rubinfeld (2004) ao invés de prevermos os movimentos futuros de uma variável qualquer apenas relacionando-a a um conjunto de outras variáveis em uma estrutura de causalidade, utilizamos o comportamento passado da variável para prever o seu comportamento futuro. De certa forma, podemos dizer que é um método sofisticado de extrapolação.

Quando se estudam as séries temporais, o conceito de estacionariedade toma grande importância. Caso as variáveis (ou a variável) analisadas não sejam estacionárias, os testes de

¹⁴ Uma reportagem muito interessante a esse respeito encontra-se em <<http://www.infomoney.com.br/carros/noticia/2277235-etanol+nao+vantajoso+nemhum+estado+pais>>. Acesso em: 27 de Nov. 2011.

hipótese convencionais, os intervalos de confiança e as previsões podem não ser confiáveis. Poderá ocorrer a chamada regressão espúria¹⁵, que significa a aparente relação entre duas variáveis quando, na verdade, estas não estão relacionadas.

Hill *et alli* (2003, p.389) definem estacionariedade¹⁶ como:

Um processo estocástico (série temporal) Y_t é estacionário se sua média e sua variância são constantes ao longo do tempo, e a covariância entre dois valores da série depende apenas da distância no tempo que separa os dois valores, e não dos tempos reais em que as variáveis são observadas.

Stock e Watson (2004) lembram que a estacionariedade requer que o futuro seja igual ao passado, pelo menos em um sentido probabilístico.

Para saber se uma série é estacionária (ou não), aplicam-se os testes de raiz unitária¹⁷. A implicação de uma série temporal possuir raiz unitária é explorada no exemplo de Madalla (2003, p.292) a seguir:

Considere, por exemplo, o modelo:

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

onde ε_t é um processo estacionário de média zero. Suponha que em algum período, digamos, Y_t , há um aumento C em ε_t . Então, todos, $Y_t, Y_{t+1}, Y_{t+2}, \dots$, aumentam em C . Logo, o efeito do choque C é permanente. No entanto, se tivermos o modelo:

$$Y_t = \alpha Y_{t-1} + \varepsilon_t \quad |\alpha| < 1$$

o efeito do choque termina com o tempo. Começando com Y_t , que aumenta em C , valores sucessivos de Y_t se elevarão em $C\alpha, C\alpha^2, C\alpha^3, (\dots)$ Portanto, a

¹⁵ Stock e Watson (2004) ilustram um caso interessante de regressão espúria envolvendo a inflação nos EUA e o PIB japonês. De 1965 a 1981, uma regressão linear simples, tendo como variável endógena a inflação americana e como variável exógena o PIB japonês, apresentou uma forte relação positiva entre as variáveis. Já para o período de 1982 a 1999 houve uma relação negativa entre as mesmas. Estes resultados conflitantes são atribuídos a não-estacionariedade de ambas as séries. Em termos de teoria econômica, não há motivos para o PIB japonês ter um impacto significativo na inflação americana.

¹⁶ Na verdade, este é o conceito de estacionariedade fraca, ou de segunda ordem. A distinção entre estacionariedade de primeira e segunda ordem é encontrada, detalhadamente, em Cryer e Chan (2006).

¹⁷ Dizer que uma série temporal possui uma raiz unitária é análogo a dizer que a série é não-estacionária.

questão de se, em auto-regressão, $Y_t = \alpha Y_{t-1} + \varepsilon_t$, a raiz α é igual a 1 ou <1 , isto é, se há uma raiz unitária ou não, é muito importante (...).

Existem diversos testes com a finalidade de detectar a presença de raiz unitária em uma série, mas os utilizados neste trabalho serão os testes dickey-fuller aumentado (ADF) e o teste KPSS¹⁸.

O teste Dickey-Fuller Ampliado é explicitado por Stock e Watson (2004, p.316) a seguir:

O teste ampliado de Dickey-Fuller (ADF) para detectar a presença de uma raiz auto-regressiva unitária testa a hipótese nula $H_0 : \delta = 0$ contra a alternativa $H_1 : \delta < 0$ na regressão

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \dots + \gamma_p \Delta Y_{t-p} + u_t$$

Sob a hipótese nula, Y_t possui uma tendência estocástica; sob a hipótese alternativa, Y_t é estacionário. A estatística ADF é a estatística t de MQO testando $\delta = 0$ na equação acima.

Se, em vez disso, a hipótese alternativa é de que Y_t é estacionário em torno de uma tendência temporal determinista, t (o número da observação) deve ser acrescentado como um regressor adicional; nesse caso, a regressão de Dickey-Fuller torna-se

$$\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \dots + \gamma_p \Delta Y_{t-p} + u_t$$

onde α é um coeficiente desconhecido e a estatística ADF é a estatística t de MQO testando $\delta = 0$ na equação acima (segunda equação).

Na prática, os testes de detecção de raiz unitária são feitos utilizando-se de softwares estatísticos.

¹⁸ Para informações mais aprofundadas sobre o teste KPSS recomenda-se a leitura do original Kwiatkowski *et alli* (1992), "Testing the null hypothesis of Stationarity Against the Alternative of a Unit Root", *Journal of Econometrics*, Vol. 54, pp. 159-178.

3.2. O modelo ARMA

O modelo misto auto-regressivo e de médias móveis (ARMA) é descrito por Pindyck e Rubinfeld (2004) como uma “junção” de um modelo auto-regressivo e um modelo de médias móveis, pois, nem sempre, um processo aleatório pode ser modelado apenas como de médias móveis ou unicamente como auto-regressivos. Um modelo ARMA (p,q) é denotado matematicamente pela equação (3.1):

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad [3.1]$$

Se este processo for estacionário, e assim possuir uma média constante ao longo do tempo:

$$\mu = \phi_1 \mu + \dots + \phi_p \mu + \delta \quad [3.2]$$

Para que o processo seja estacionário, é necessário que:

$$\phi_1 + \phi_2 + \dots + \phi_p < 1 \quad [3.3]$$

Às vezes, as séries trabalhadas possuem raiz unitária. Quando isto ocorre, é necessário transformá-las em séries estacionárias, e uma dessas transformações é através do operador de diferença. Uma série Y_t será estacionária de ordem d se

$$W_t = \Delta^d Y_t \quad [3.4]$$

é uma série estacionária. O “ Δ ” na equação (3.4) denota o *operador de diferença*, ou seja, um operador de primeira diferença será representado por $\Delta Y_t = Y_t - Y_{t-1}$. Ao se diferenciar a série original Y_t para produzir uma série estacionária W_t , será possível modelar W_t como um processo ARMA. Se $W_t = \Delta^d Y_t$ e W_t segue um processo ARMA(p,q), então Y_t será um processo auto-regressivo integrado e de médias móveis de ordem (p,d,q) também chamado de ARIMA(p,d,q). Reescrevendo a equação para o processo ARIMA(p,d,q) usando o operador de defasagem, tem-se que :

$$\phi(\beta) \Delta^d Y_t = \delta + \theta(\beta) \varepsilon_t \quad [3.5]$$

$$\text{com} \quad \varnothing(\beta) = 1 - \varnothing_1 \beta - \varnothing_2 \beta^2 - \dots - \varnothing_p \beta^p \quad [3.6]$$

$$\text{e} \quad \theta(\beta) = 1 - \theta_1 \beta - \theta_2 \beta^2 - \dots - \theta_q \beta^q \quad [3.7]$$

sendo o $\varnothing(\beta)$ o operador auto-regressivo e $\theta(\beta)$ o operador de médias móveis.

3.3. Modelos Sazonais

Para Morettin e Toloi (2004), é possível que, mesmo após a eliminação da componente sazonal determinística de uma série temporal, ainda reste correlação significativa em *lags* sazonais. Então, será necessário considerar a existência de uma sazonalidade estocástica e ajustar à série original um modelo ARIMA sazonal – SARIMA. Supondo uma série temporal Z_t que exhibe um comportamento sazonal determinístico com período 12, um modelo SARIMA de ordem $(p,d,q) \times (P,D,Q)$ é definido por Morettin e Toloi (2004) como:

$$\varnothing(\beta)\phi(\beta^{12})(1 - \beta^{12})^D (1 - \beta)^d Z_t = \theta(\beta)\Theta(\beta^{12})\alpha_t \quad [3.8]$$

onde:

$$\theta(\beta) = 1 - \theta_1 \beta - \dots - \theta_q \beta^q;$$

$$\varnothing(\beta) = 1 - \varnothing_1 \beta - \dots - \varnothing_p \beta^p;$$

$\phi(\beta^{12}) = 1 - \phi_1 \beta^{12} - \dots - \phi_P \beta^{12P}$ é o operador auto-regressivo sazonal de ordem P , estacionário;

$\Theta(\beta^{12}) = 1 - \Theta_1 \beta^{12} - \dots - \Theta_Q \beta^{12Q}$ é o operador de médias móveis sazonal de ordem Q , invertível;

$(1 - \beta^{12})^D$ é o operador de diferença sazonal, com D indicando o número de “diferenças sazonais”;

$(1 - \beta)^d$ é o operador de diferença, com d significando o número de diferenças.

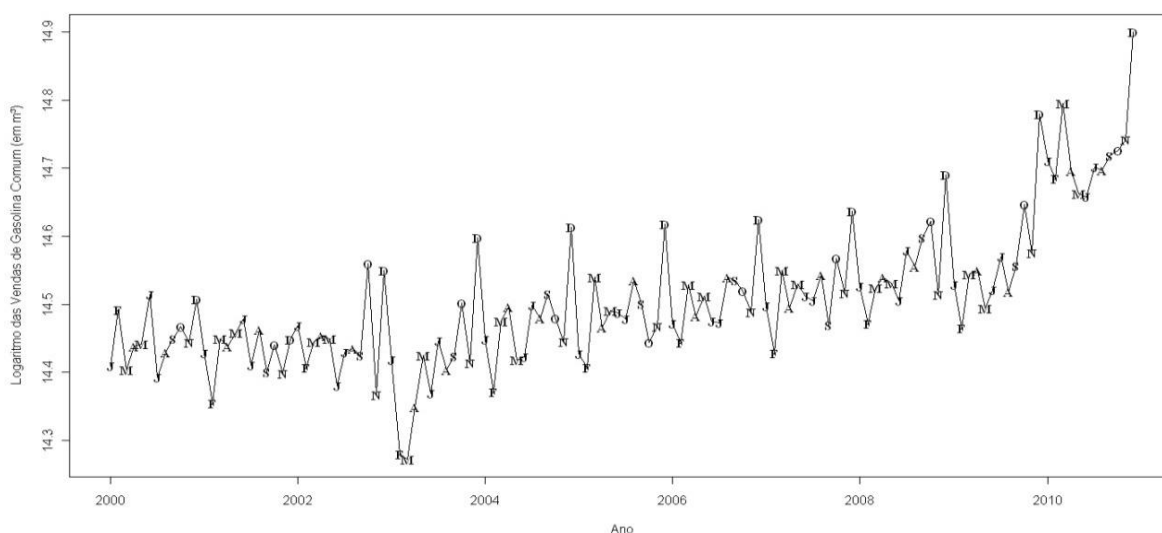
α_t é o erro (independente e identicamente distribuído).

O modelo da equação (3.8) é chamado de ARIMA sazonal – SARIMA – de ordem $(p,d,q) \times (P,D,Q)$.

4. RESULTADOS

Os dados¹⁹ aqui analisados vão do período de janeiro do ano 2000 até dezembro de 2010. Porém, para testar o grau de ajustamento do modelo, serão utilizados as vendas dos meses de janeiro, fevereiro, março, abril, maio e junho de 2011.

Gráfico 2 – Inspeção visual da sazonalidade no Brasil – 2000-2010



Fonte: Elaboração própria a partir do software R²⁰

As vendas de gasolina comum no Brasil apresentam, visualmente, sazonalidade. Se observarmos com atenção, em algum mês do último trimestre é onde, normalmente, localiza-se o mês com maiores vendas do ano. Já no primeiro trimestre, freqüentemente, ocorrem as vendas mais fracas. Isso é ilustrado no gráfico 2 no qual as maiores vendas estão, quase sempre, associadas à letra O (de outubro) ou D (de dezembro). Ainda no gráfico 2, as vendas mais baixas costumam estar associadas a algum dos três meses do primeiro trimestre (janeiro, fevereiro, março).

¹⁹ Os dados utilizados neste trabalho encontram-se disponíveis no site da ANP <http://www.anp.gov.br/doc/dados_estatisticos/Vendas_de_Combustiveis_m3.xls> Acesso em: 10 de Set. 2011.

²⁰ Disponível – gratuitamente – em <<http://www.r-project.org>> Acesso em: 6 de Set. 2011.

Para tentar observar melhor a sazonalidade, construíram-se duas tabelas. A tabela 1 contém o mês de menor venda de gasolina comum – no Brasil – em cada ano e a tabela 2 possui o mês de maior venda em cada ano:

Tabela 1 – Meses de menores vendas de gasolina comum, no Brasil – 2000-2010

mês/ano	Vendas (em m³)
Julho/2000	1.780.834
Fevereiro/2001	1.714.464
Novembro/2002	1.736.800
Março/2003	1.578.290
Fevereiro/2004	1.742.412
Fevereiro/2005	1.805.672
Fevereiro/2006	1.874.940
Fevereiro/2007	1.844.983
Fevereiro/2008	1.926.803
Fevereiro/2009	1.914.531
Junho/2010	2.323.568

Fonte: Elaboração própria

Da tabela 1, observa-se que apenas em três anos – 2000, 2002 e 2010 – as menores vendas não se localizaram em algum mês do primeiro trimestre.

Tabela 2 – Meses de maiores vendas de gasolina comum, no Brasil – 2000-2010

mês/ano	Vendas (em m³)
junho/2000	2.011.256
junho/2001	1.940.547
outubro/2002	2.104.499
dezembro/2003	2.187.311
dezembro/2004	2.221.234
dezembro/2005	2.229.875
dezembro/2006	2.246.973
dezembro/2007	2.273.262
dezembro/2008	2.398.995
dezembro/2009	2.623.497
dezembro/2010	2.959.459

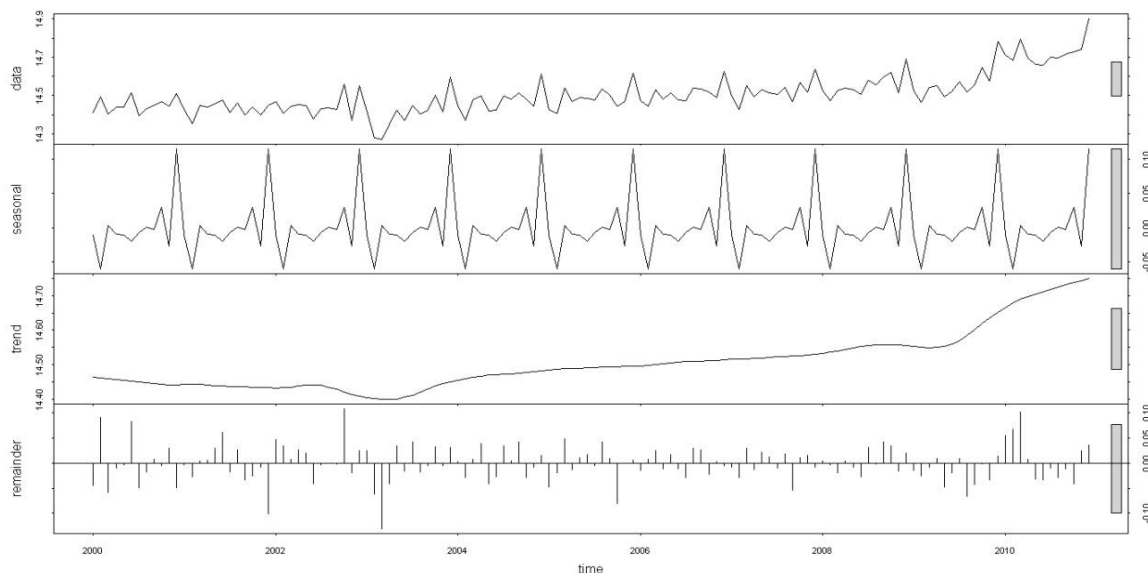
Fonte: Elaboração própria

Da tabela 2, observa-se que apenas em dois anos – 2000 e 2001 – as maiores vendas não se localizaram em algum mês do último trimestre do ano.

Porém, até o presente momento, a possível existência de sazonalidade só foi “analisada” visualmente (no caso do gráfico) ou usando uma lógica econômica informal (no caso da tabela). Dada a necessidade de uma metodologia científica deste trabalho, a

confirmação da existência de sazonalidade passa por um filtro estatístico²¹, conforme demonstrado no gráfico, a seguir:

Gráfico 3 – sazonalidade das vendas de gasolina comum, no Brasil



Fonte: Elaboração própria a partir do software R.

O componente *seasonal*, do gráfico 3 ilustra, estatisticamente, a existência da sazonalidade nas vendas de gasolina comum, no Rio Grande do Norte. Como é possível observar, há uma regularidade no comportamento da sazonalidade, com topos e fundos apresentando um padrão de comportamento relativamente uniforme. Isso significa que na hora de estimar o modelo de previsão podemos, também, modelar essa sazonalidade. A componente *data* é o gráfico da série original (em logaritmo neperiano) e *trend* é a tendência da série – onde é possível observar que, desde 2003, as vendas de gasolina comum no Brasil apresentam tendência de alta, com uma leve reversão dessa tendência em fins²² de 2008 e o início de 2009, seguida de uma forte retomada em 2010.

Dada a sazonalidade da série, ao invés de um modelo ARIMA (p,d,q), será usado um modelo ARIMA com sazonalidade – ou SARIMA – (p,d,q)x(P,D,Q).

Partindo agora para a estimação do modelo para fazer as previsões, já alertamos na seção 3 que é necessário primeiro saber se a série possui raiz unitária. Assim, vejamos os

²¹ A sazonalidade exposta no gráfico 3 foi extraída – da série – através da decomposição sazonal de séries temporais de Loess. Informações adicionais em Cleveland *et al.* (1990).

²² Lembrar que nesse período de fins de 2008 e no ano de 2009 a economia brasileira sofreu forte desaceleração no seu crescimento devido à crise mundial.

resultados dos testes de raiz unitária, para a série logaritimizada²³ das vendas de gasolina comum no Brasil, descritos na tabela a seguir:

Tabela 3 – testes de raiz unitária

Variável	ADF (com constante)		ADF (com constante e tendência)		KPSS
	Estat. T	Prob.	Estat. t	Prob.	Estat. LM
Vendas gasolina (ln)	0,169385	0,9707	-1,87671	0,6600	2,02833

Valores críticos (KPSS): 0,739(1%); 0,463(5%); 0,347(10%).

* Resultados estimados com o uso do software Gretl²⁴.

** Na análise dos resultados será considerado um nível de significância de 5%.

Os testes ADF e KPSS detectaram²⁵ a presença de raiz unitária nos dados, pois o primeiro aceitou²⁶ a hipótese nula e o segundo rejeitou²⁷ a mesma. Falamos anteriormente que a presença de raiz unitária nos dados pode trazer várias implicações indesejáveis para as estimativas.

Para tentar evitar essas implicações, será necessário utilizar o operador de diferença para tentar tornar as séries não-estacionárias em séries estacionárias. Para esse artigo, isso significa que se trabalhará com a diferença da série logaritmicada. Aplicando os testes à série modificada (diferenciada), têm-se os seguintes resultados, na tabela 4:

Tabela 4 – testes de raiz unitária (após diferenciação)

Variável	ADF (com constante)		ADF (com constante e tendência)		KPSS
	Estat. T	Prob.	Estat. t	Prob.	Estat. LM
Diferença de Vendas gasolina (ln)	-5,96015	1,493e-007	-6,17025	4,393e-007	0,155249

Valores críticos (KPSS): 0,739(1%); 0,463(5%); 0,347(10%).

* Resultados estimados com o uso do software Gretl.

** Na análise dos resultados será considerado um nível de significância de 5%.

²³ A série original foi logaritmicada para a base natural (logaritmo neperiano).

²⁴ Disponível – gratuitamente – em <<http://gretl.sourceforge.net/>>. Acesso em: 6 de Out. 2011.

²⁵ Lembrar que, enquanto o teste ADF possui como hipótese nula a presença de raiz unitária (não-estacionariedade), o teste KPSS possui como hipótese nula a ausência de raiz unitária (estacionariedade).

²⁶ No teste ADF, a aceitação ou rejeição da hipótese nula seguirá a metodologia de análise do *prob*. O *prob*. é a probabilidade de se cometer o erro tipo I (rejeitar a hipótese nula e ela ser verdadeira). Logo, no teste ADF, a hipótese nula será rejeitada quando *prob*. ≤ 0,05.

²⁷ No teste KPSS, quando o módulo da estatística LM for maior do que o seu valor crítico a 5%, rejeita-se a hipótese nula – frisando mais uma vez que no teste KPSS a hipótese nula é a ausência de raiz unitária.

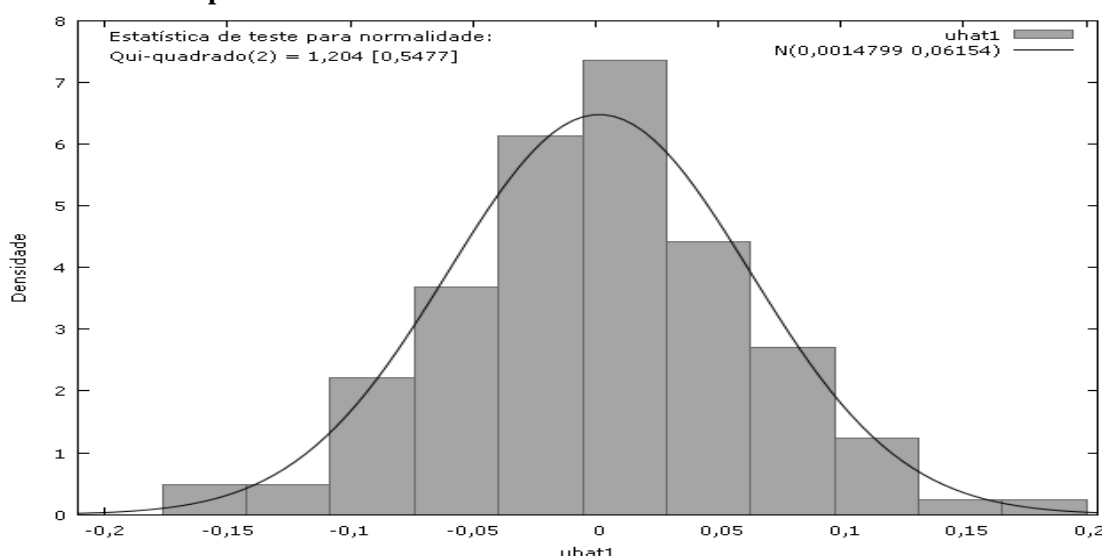
Da tabela 4, observando-se os resultados dos testes, percebe-se que ao tomar a diferença da série logaritmizada, gerou-se uma série estacionária, pois a hipótese nula do teste KPSS foi aceita e a hipótese nula do teste ADF foi rejeitada.

Depois da obtenção de séries estacionárias, o passo seguinte será o de escolher os parâmetros $(p,d,q) \times (P,D,Q)$ do modelo SARIMA. Essa escolha será feita tendo em mente duas características desejáveis em um modelo linear de série temporal: minimização dos critérios de informação e normalidade dos resíduos.

Há um dilema na escolha do tamanho da defasagem (parâmetros p e q), pois, a escolha de poucas defasagens pode diminuir a precisão da previsão – informações valiosas podem estar sendo perdidas –, mas o acréscimo de defasagens aumenta a incerteza da estimação, por isso escolhe-se a defasagem que minimize algum critério de informação. Já para a construção dos intervalos de confiança em torno das previsões é importante fazer uma suposição sobre como os erros, do modelo estimado, estão distribuídos. Assim como o fazem Morettin e Toloi (2004), a suposição, para este trabalho, será a de que os erros do modelo seguem uma distribuição normal (gausiana).

Dada as ponderações feitas anteriormente, o modelo escolhido para os dados deste trabalho foi um SARIMA²⁸ $(0,1,0) \times (0,1,1)$. Antes de fazer as previsões sobre as vendas de gasolina comum pelas distribuidoras no Brasil, cabe fazer um teste²⁹ de normalidade nos resíduos. Este teste está representado no gráfico 4, a seguir:

Gráfico 4 – teste para normalidade nos resíduos



Fonte: Elaboração própria. Estimativas com o Gretl.

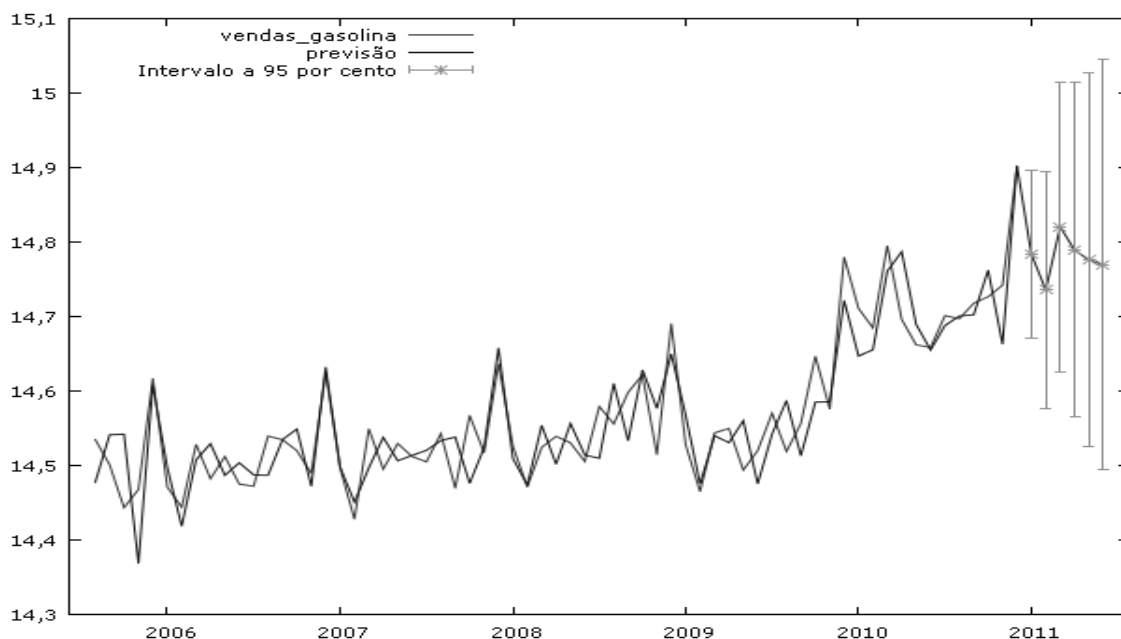
²⁸ A estimação do modelo foi feita com o software Gretl.

²⁹ O teste de normalidade implementado pelo software Gretl – neste caso – foi o teste Doornik–Hansen. A hipótese nula deste teste é que o erro tem distribuição normal. Para mais detalhes ver Doornik e Hansen (1994).

O teste para normalidade nos resíduos informa que a hipótese nula de normalidade é aceita³⁰ para o nível de significância de 5%, pois, o *p*-valor (0,5477) é maior do que 0,05.

Agora que os dados transformados não possuem raiz unitária e o erro do modelo SARIMA estimado tem distribuição normal, pode-se partir para a previsão dos valores futuros da série estudada. Tal previsão está representada no gráfico 5, logo abaixo:

Gráfico 5 – previsão de vendas futuras de gasolina comum no Brasil (em ln)



Fonte: Elaboração própria. Estimativas com o Gretl.

Conforme exposto nas legendas do gráfico 5, a linha mais escura (cor mais sólida) representa os valores observados para o logaritmo das vendas de gasolina comum no Brasil, a linha mais clara representa a previsão feita pelo modelo e as barras cinza correspondem ao intervalo de confiança para a previsão.

Para visualizar melhor, os resultados do gráfico 5 são transferidos para uma tabela, e comparados com as vendas divulgadas pela ANP.

Isso está feito na tabela 5 da página a seguir:

³⁰ Segundo Hill *et alli.* (2003, p.119) “quando o valor-p de um teste de hipótese é menor do que o valor escolhido de α , o procedimento de teste conduz à rejeição da hipótese nula”. Neste trabalho está se usando $\alpha=0,05$, ou seja, 5% de nível de significância.

Tabela 5 – previsão de vendas* de gasolina comum no Brasil

Data	Valor previsto	limite inferior do intervalo de confiança (a 95%) para o valor previsto	limite superior do intervalo de confiança (a 95%) para o valor previsto	Valor observado
Jan/2011	2.633.114,7	2.353.431,42	2.946.035,71	2.506.194,00
Fev/2011	2.509.836,07	2.141.298,54	2.941.805,3	2.517.826,00
Mar/2011	2.731.894,82	2.249.028,71	3.318.432,19	3.077.410,00
Abr/2011	2.647.803,53	2.115.189,57	3.314.528,7	3.134.541,00
Mai/2011	2.612.713,77	2.032.558,69	3.358.463,04	2.822.025,00
Jun/2011	2.596.401,48	1.972.035,86	3.418.447,3	2.747.819,00

Fonte: Elaboração própria. Estimativas com o Gretl.

* valores em m³

Da tabela 5, é possível perceber que a diferença entre o valor previsto e o valor observado foi, para 2011, de: 5,06% para o mês de janeiro; -0,31% para o mês de fevereiro; -11,22% para o mês de março; de -15,52% para abril; -7,41% para o mês de maio e de -5,51% para o mês de junho. Além disso, o valor efetivamente observado ficou – em todos os casos – dentro do intervalo de confiança dos valores previstos.

Como foram feitas previsões para os primeiros seis meses do ano de 2011, podemos somar estes valores previstos para chegarmos à previsão semestral. Fazendo este cálculo, temos uma venda prevista – no primeiro semestre de 2011 – de 15.731.768,37 m³ de gasolina comum no Brasil, contra uma venda efetivamente observada de 16.805.815,00 m³. Com isso, houve uma diferença de apenas -6,39% entre a previsão para o primeiro semestre de 2011 e o valor efetivamente observado.

Constata-se que, a partir dos dados de vendas de gasolina comum no Brasil para o período de jan/2000 a dez/2010, foi possível construir um modelo estatístico preditivo para as vendas dos próximos seis meses – ou seja, para janeiro/2011 a jun/2011.

Desta maneira, esta quarta seção demonstrou como um modelo de séries temporais – modelo SARIMA – se adequou bem à previsão da variável aqui estudada, principalmente se levarmos em consideração o Princípio da Parcimônia³¹.

³¹ Este princípio foi proposto, primeiramente, pelo filósofo inglês Guilherme de Occam (1285-1347) que disse: “(...) não multiplique as coisas desnecessariamente (*entia non multiplicanda praeter necessitatem*)”. Interpretando de outra forma: “se existe mais de uma explicação para uma dada observação, devemos adotar aquela mais simples”. Em trabalhos de previsões de variáveis é comum se observar abordagens altamente complexas e avançadas que, no final das contas, se mostram iguais ou – apenas – um pouco melhores do que modelos mais simples.

5. CONSIDERAÇÕES FINAIS

Um SARIMA é só um modelo, dentre inúmeros outros, que podem ser utilizados para fazer previsões de valores futuros de séries temporais. Exemplos de outras abordagens que poderiam ser utilizadas são: o VAR (vetor autoregressivo); VECM (vetores com correção de erros); BVAR (vetores autoregressivos bayesianos); ARCH (modelo autoregressivo com heterocedasticidade condicional) e; inúmeros outros.

Não foi escopo deste trabalho aplicar os diversos modelos em busca de um que se ajustasse melhor aos dados, já que, conforme descrito na introdução, o objetivo aqui foi apenas o de mostrar uma aplicação de modelos ARIMA com sazonalidade (SARIMA) na previsão das vendas de gasolina comum no Brasil. A grande vantagem da modelagem aqui utilizada é a sua, razoavelmente fácil, implantação *vis-à-vis* outras abordagens.

Para trabalhos futuros sobre este assunto, fica a sugestão de se utilizar outros modelos de séries temporais no intuito de testar se a capacidade preditiva destes é maior do que o SARIMA aqui utilizado. Em trabalhos – ou consultorias – para o mercado, as previsões também podem ser estimadas com um intervalo de confiança menor do que os 95% aqui utilizados, a fim de estreitar o intervalo de valores previstos e facilitar o planejamento econômico da empresa – ou mercado.

Por fim, seria interessante que cada distribuidora de combustível também se utilizassem de modelos econométricos para prever as suas próprias vendas, ao invés de se basearem em uma previsão para todo o Estado. Valores individuais podem, por exemplo, auxiliar na escolha de metas de vendas **plausíveis** (grifo nosso) de serem atingidas. Em suma, a modelagem econométrica de dados econômicos e financeiros tem uma ampla gama de aplicações e este trabalho expôs apenas uma destas possibilidades.

6. REFERÊNCIAS

ANP, Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. **Anuário Estatístico Brasileiro do Petróleo, Gás Natural e Biocombustíveis**. Rio de Janeiro: ANP, 2011.

CLEVELAND, R. B. ... [et al.] STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6, 3–73, 1990.

CRYER, Jonathan D. ; CHAN, Kung-Sik. **Time Series Analysis with application in R**. 2ed. USA: Springer, 2006.

DOORNIK, J. A. ; HANSEN, H. “**An Omnibus Test for Univariate and Multivariate Normality**”, working paper, Nuffield College, Oxford, 1994.

FIEL, Rair Tocantins. **Modelagem de previsão de venda de combustível no posto santos dumont – Baião/PA**. Monografia de conclusão de curso, Universidade Federal do Pará. Departamento de Estatística. Ano de conclusão: 2005.

HILL, Carter...[et al.]. **Econometria**. 2 ed. São Paulo: Saraiva, 2003.

KWIATKOWSKI, D. ... [et al.], “**Testing the null hypothesis of Stationarity Against the Alternative of a Unit Root**”, *Journal of Econometrics*, 1992.

LEAL, José Eugênio ... [et al.], **Diagnóstico da rede de distribuição de derivados de petróleo no Brasil e sua representação em um SIG**. XXIII Encontro Nac. de Eng. De Produção – Ouro Preto, MG, Brasil, 21 a 24 de out de 2003.

LOPES, Ricardo Dantas. **Modelos de previsão para uma série temporal de produção de veículos automotores**. Revista de Administração Nobel, Nº 01, p. 27-36, jan./jun.2003.

MADALLA, G. S. **Introdução à Econometria** . 3 ed. Rio de Janeiro 2003.

MORETTIN, Pedro A.; TOLOI, Clélia M. C. **Análise de Séries Temporais**. São Paulo: Edgard Blücher, 2004.

PINDYCK, Robert S. ; RUBINFELD , Daniel L. **Econometria: Modelos e Previsões** . 1 ed. Campus, 2004.

SABRY, M. ... [et al.], **Comparison Between Regression and Arima Models in Forecasting Traffic Volumes**. Australian Journal of Basic and Applied Sciences, 1(2):126 – 136, 2007.

STOCK, James H. ; WATSON, Mark W. **Econometria** . 1 ed. Makron, 2004.