

**MODELAGEM DE SÉRIES REPRESENTATIVAS DO
SETOR ENERGÉTICO BRASILEIRO**

LUCIENE RESENDE GONÇALVES

2007

LUCIENE RESENDE GONÇALVES

**MODELAGEM DE SÉRIES REPRESENTATIVAS DO SETOR
ENERGÉTICO BRASILEIRO**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de "Mestre".

Orientadora: Profa. Dra. Thelma Sáfydi

LAVRAS
MINAS GERAIS - BRASIL
2007

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Gonçalves, Luciene Resende

Análise de séries representativas do setor energético brasileiro / Luciene Resende Gonçalves. – Lavras : UFLA, 2007. - Lavras: UFLA, 2007.

106 p. : il.

Orientador: Thelma Sáfyadi

Dissertação (Mestrado) - UFLA.

Bibliografia.

1. Setor energético. 2. Fontes de energia. 3. Séries temporais. 4. Modelo VAR. 5. Modelo SARIMA. I. Universidade Federal de Lavras. II. Título.

CDD-519.55

LUCIENE RESENDE GONÇALVES

**MODELAGEM DE SÉRIES REPRESENTATIVAS DO SETOR
ENERGÉTICO BRASILEIRO**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de "Mestre".

APROVADA em 22 de fevereiro de 2007

Prof. Dr. Augusto Ramalho de Moraes UFLA

Prof. Dr. Pedro Castro Neto UFLA

Prof. Dr. Renato Ribeiro de Lima UFLA

Profa. Dra. Thelma Sáfadi
UFLA
(Orientadora)

LAVRAS
MINAS GERAIS-BRASIL

*A minha querida família:
meus pais, Edson e Márcia;
minha avó Tereza
e meus irmãos: Christiane, Márcio e
Liliane.*

AGRADECIMENTOS

A Deus, por estar sempre presente, me guiando e dando forças para vencer os obstáculos.

A minha família pelos constantes incentivos e pelo perdão nos muitos momentos de ausência.

À Universidade Federal de Lavras, pela oportunidade de aprimoramento acadêmico.

À CAPES, pela concessão da bolsa.

Ao apoio, confiança, dedicação e amizade da orientadora, Thelma Sáfadi.

Aos professores do Departamento de Ciências Exatas, pelos conhecimentos transmitidos.

Aos funcionários do Departamento de Ciências Exatas, sempre bem dispostos a ajudar.

À amizade dos amigos: Ademária, Anderson, Carlos, Claudiney, Elisa, Graziela, Lourdinha, Melissa, Osvaldo, Patrícia Paranaíba, Patrícia Neves, Patrícia Siqueira, Quintiliano, Rafael e Verônica.

Ao plantão "tira-dúvidas" do Andersinho.

A todos, que direta ou indiretamente, contribuíram para a realização deste trabalho.

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	iv
RESUMO	vii
ABSTRACT	viii
1 INTRODUÇÃO	1
1.1 Objetivo geral	3
1.2 Objetivos específicos	3
2 REFERENCIAL TEÓRICO	4
2.1 Energia	4
2.1.1 A Revolução Industrial	4
2.1.2 O petróleo	5
2.1.3 O petróleo e a crise energética	7
2.1.4 Energia alternativa	9
2.1.5 Biomassa	9
2.1.6 A cana-de-açúcar e o álcool	11
2.1.7 O álcool e a gasolina	13
2.1.8 O biodiesel e o óleo diesel	15
2.1.9 Alguns trabalhos desenvolvidos	17
2.2 Séries temporais	19
2.2.1 Modelos para séries temporais	19
2.2.2 Estacionaridade	20
2.2.3 Função de autocovariância e autocorrelação	20
2.2.4 Decomposição clássica	22
2.2.5 Tendência	24

2.2.5.1	Teste do sinal	25
2.2.6	Sazonalidade	25
2.2.6.1	Teste de Fisher	26
2.2.6.2	Estimação da sazonalidade	27
2.2.7	Modelos de Box e Jenkins	29
2.2.8	Modelos lineares estacionários	31
2.2.8.1	Processo auto-regressivo de ordem p - AR(p)	33
2.2.8.2	Processo de médias móveis de ordem q - MA(q)	34
2.2.8.3	Modelos auto-regressivos e de médias móveis de ordem p, q - ARMA (p, q)	35
2.2.8.4	Função de autocorrelação parcial	35
2.2.8.5	Características das funções de autocorrelação e autocorrelação par- cial dos processos lineares estacionários	37
2.2.8.5.1	Função de autocorrelação (fac)	37
2.2.8.5.2	Função de autocorrelação parcial (facp)	37
2.2.9	Modelos não-estacionários	38
2.2.9.1	Modelo auto-regressivo integrado de médias móveis - ARIMA (p, d, q)	38
2.2.9.2	Modelo sazonal auto-regressivo integrado de médias móveis - SA- RIMA (p, d, q) \times (P, D, Q)	39
2.2.10	Estágios do ciclo iterativo	40
2.2.10.1	Identificação	40
2.2.10.2	Estimação	41
2.2.11	Diagnóstico	42
2.2.11.1	Teste da autocorrelação residual	42
2.2.11.2	Teste de Box-Pierce	43
2.2.12	Previsões	44

2.2.13 Critérios para escolha do modelo	44
2.2.13.1 Critério de informação de Akaike (AIC)	44
2.2.13.2 Critério do erro quadrático médio de previsão (EQMP)	45
2.2.14 MAPE	45
2.2.15 Análise de intervenção	45
2.2.15.1 Efeitos da intervenção	47
2.2.16 Modelos lineares multivariados	50
2.2.16.1 Séries estacionárias	51
2.2.16.2 Estimação de médias e covariâncias	53
2.2.16.3 Modelos auto-regressivos vetoriais - VAR(p)	54
2.2.16.4 Construção de modelos VAR	54
2.2.16.4.1 Identificação	54
2.2.16.4.2 Estimação	56
2.2.16.4.3 Diagnóstico	56
3 MATERIAL E MÉTODOS	57
4 RESULTADOS E DISCUSSÃO	60
4.1 Modelos univariados	60
4.1.1 Série de produção de cana-de-açúcar	60
4.1.2 Série de produção de álcool	66
4.1.3 Série de produção de petróleo	73
4.1.4 Série de produção de óleo diesel	79
4.1.5 Série de produção de gasolina	85
4.2 Modelo multivariado	91
5 CONCLUSÕES	96
REFERÊNCIAS BIBLIOGRÁFICAS	98
ANEXOS	100

LISTA DE TABELAS

2.1	Comparações entre as propriedades dos combustíveis álcool e gasolina.	14
4.1	Estimativas dos parâmetros dos modelos ARIMA(2,1,1) e ARIMA(2,1,1) com intervenção para a série de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998.	64
4.2	Estatísticas para o teste de Box-Pierce de modelos ARIMA com e sem intervenção.	65
4.3	Valores reais de produção anual de cana-de-açúcar, em milhões de toneladas, para o período de 1999 a 2004 e valores preditos para o mesmo período.	65
4.4	Estimativas dos critérios de AIC, EQMP e MAPE de modelos ARIMA com e sem intervenção.	66
4.5	Estimativas dos parâmetros dos modelos SARIMA com e sem intervenções para a série de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a dezembro de 2005.	71
4.6	Estatísticas para o teste de Box-Pierce de modelos SARIMA com e sem intervenções.	71
4.7	Valores reais de produção de álcool nível 80 - índice (média 2002=100), para o período de janeiro de 2006 a setembro do mesmo ano e valores preditos para o mesmo período.	72
4.8	Estimativas dos critérios de AIC, EQMP e MAPE de modelos SARIMA com e sem intervenção	72

4.9	Estimativas dos parâmetros dos modelos SARIMA com e sem intervenções para a série de produção mensal de petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005. . . .	77
4.10	Estatísticas para o teste de Box-Pierce dos modelos SARIMA com e sem intervenção.	77
4.11	Valores reais de produção de petróleo e valores preditos com e sem intervenção, para o período de janeiro de 2006 a setembro do mesmo ano.	78
4.12	Estimativas dos critérios de AIC, EQMP e MAPE dos modelos SARIMA com e sem intervenção.	79
4.13	Estimativas dos parâmetros dos modelos SARIMA com e sem intervenção, para a série de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005. .	83
4.14	Estatísticas para o teste de Box-Pierce dos modelos SARIMA com e sem intervenção.	83
4.15	Valores reais de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro a setembro de 2006 e valores preditos para o mesmo período.	84
4.16	Estimativas dos critérios de AIC, EQMP e MAPE dos modelos SARIMA com e sem intervenção	85
4.17	Estimativas dos parâmetros dos modelos ARIMA(2,1,0), ARIMA(0,1,1) e ARIMA(2,1,0) com intervenções para a série de produção mensal de gasolina, em metros cúbicos, no período de janeiro de 2000 a abril de 2006.	89
4.18	Estatísticas para o teste de Box-Pierce dos modelos ARIMA com e sem intervenção.	90

4.19	Valores reais de produção de gasolina, para o período de maio a setembro de 2006 e valores preditos para o mesmo período.	90
4.20	Estimativas dos critérios de AIC, EQMP e MAPE dos modelos ARIMA com e sem intervenção.	91
4.21	Estatísticas resultantes de ajustes de modelos VAR(p), $p=1,\dots,8$, para as séries de produção de álcool, óleo diesel e gasolina	92
4.22	Ajuste de um modelo VAR(2) para as séries de álcool, óleo diesel e gasolina.	95
1.A	Série de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998.	101
2.A	Série de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a setembro de 2006.	102
3.A	Série de produção mensal de petróleo, em m^3 , no período de janeiro de 2000 a setembro de 2006.	103
4.A	Série de produção mensal de óleo diesel, em m^3 , no período de janeiro de 2000 a setembro de 2006.	104
5.A	Série de produção mensal de gasolina, em m^3 , no período de janeiro de 2000 a setembro de 2006.	105
6.A	Série de produção mensal de álcool sem sazonalidade - método de regressão.	106

LISTA DE FIGURAS

2.1	Filtro linear, com entrada a_t , saída Z_t , e função de transferência $\psi(B)$	31
2.2	Efeito gradual permanente (a), gradual temporário (b), abrupto permanente (c) e abrupto temporário (d), sobre uma série temporal Z_t	47
2.3	Estrutura da função de transferência $v_j(B)$ de uma série temporal.	48
4.1	Representação gráfica da série original de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998.	60
4.2	Representação gráfica da função de autocorrelação e periodograma da série de produção anual de cana-de-açúcar em milhões de toneladas no período de 1947 a 1998.	61
4.3	Representação gráfica da série diferenciada de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998.	62
4.4	Representação gráfica das funções de autocorrelação (fac) e autocorrelação parcial (facp) da série diferenciada de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998.	63
4.5	Representação gráfica da série original de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a dezembro de 2005.	67
4.6	Representação gráfica da função de autocorrelação e periodograma da série de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a dezembro de 2005.	67

4.7	Representação gráfica da série diferenciada de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a dezembro de 2005.	69
4.8	Representação gráfica das funções de autocorrelação (fac) e autocorrelação parcial (facp) da série diferenciada de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a dezembro de 2005.	69
4.9	Representação gráfica da série original de produção mensal de petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.	73
4.10	Representação gráfica da função de autocorrelação e periodograma da série de produção mensal de petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.	74
4.11	Representação gráfica da série diferenciada de produção mensal de petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.	75
4.12	Representação gráfica das funções de autocorrelação (fac) e autocorrelação parcial (facp) da série diferenciada de produção mensal petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.	75
4.13	Representação gráfica da série original de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.	79
4.14	Representação gráfica da função de autocorrelação e periodograma da série de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.	80

4.15	Representação gráfica da série diferenciada de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.	82
4.16	Representação gráfica das funções de autocorrelação (fac) e autocorrelação parcial (facp) da série diferenciada de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.	82
4.17	Representação gráfica da série original de produção mensal de gasolina, em metros cúbicos, no período de janeiro de 2000 a abril de 2006.	85
4.18	Representação gráfica da função de autocorrelação e periodograma da série de produção mensal de gasolina, em metros cúbicos, no período de janeiro de 2000 a abril de 2006.	86
4.19	Representação gráfica da série diferenciada de produção mensal de gasolina, em metros cúbicos, no período de janeiro de 2000 a abril de 2006.	87
4.20	Representação gráfica das funções de autocorrelação (fac) e autocorrelação parcial (facp) da série diferenciada de produção mensal de gasolina, em metros cúbicos, no período de janeiro de 2000 a abril de 2006.	87
4.21	Representação gráfica da função de autocorrelação dos resíduos. .	94

RESUMO

GONÇALVES, Luciene Resende. Modelagem de séries representativas do setor energético brasileiro. Lavras: UFLA, 2007. 106 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG, Brasil. *

Neste trabalho foram realizadas modelagens univariadas e multivariada das principais séries que compõem o setor energético brasileiro, constituído por fontes renováveis e não renováveis de energia, utilizando técnicas de séries temporais. A série de produção de cana-de-açúcar, em milhões de toneladas, e a série de produção de álcool nível 80 - índice (média 2002=100) representaram as fontes renováveis. As séries de produção de petróleo, óleo diesel e gasolina, em metros cúbicos, compõem as fontes não renováveis. Na modelagem univariada foram utilizados os modelos de Box e Jenkins, considerando, para previsão, modelos com e sem intervenção. Para a produção de cana-de-açúcar e álcool os modelos com intervenção obtiveram melhores previsões. No ajuste do modelo multivariado, considerou-se o modelo auto-regressivo vetorial (VAR) para as séries de produção dos combustíveis: álcool, óleo diesel e gasolina, no período de janeiro de 2000 a setembro de 2006. Foram observadas relações de dependência nas séries de produção do óleo diesel e da gasolina e a série de produção de álcool depende apenas de seus valores passados.

*Comitê orientador: Profa. Dra. Thelma Sáfiadi - UFLA (Orientadora)

ABSTRACT

GONÇALVES, Luciene Resende. Modeling of Representative Series of the Brazilian Energetic Sector. Lavras: UFLA, 2007. 106 p. Dissertation (Master in Statistics and Agricultural Experimentation) - Federal University of Lavras, Lavras, MG, Brazil.*

In this work, univariate and multivariate modeling of the main series that compound the Brazilian energetic sector, which is constituted by renewable and nonrenewable energy sources, were accomplished using time series techniques. The sugar cane production series, in millions of tons, and the alcohol level 80 production series - index (mean 2002=100) represented the renewable sources. The petroleum, diesel and gasoline production series, in cubic meters, compounded the non-renewable sources. The Box and Jenkins models were used in the univariate modeling, considering the models with and without intervention for forecasting. It was observed that the adjusted models obtained good predictions. For sugar cane and alcohol and production, the models with intervention obtained better predictions. In the adjustment of the multivariate model, the vectorial auto-regressive (VAR) model was considered for the production series of the following fuels: alcohol, diesel and gasoline in the period from january/2000 to september/2006. It was observed relations of dependence in the diesel and gasoline production series, and the alcohol production series depends only on its past values.

*Guidance Committee: Profa. Dra. Thelma Sáfyadi - UFLA (Adviser)

1 INTRODUÇÃO

A importância da energia está, a cada dia, mais presente no cotidiano das pessoas. A dependência é tanta que o consumo pode ser usado para medir o poder aquisitivo da sociedade. Ela faz parte da vida do homem desde os tempos remotos, mas foi o progresso trazido pela Revolução Industrial e pela descoberta do petróleo que selou, de uma vez por todas, a dependência energética do mundo moderno.

O abastecimento, porém, vem passando por uma fase mundial de colapso. O petróleo, que foi visto em outros tempos como sinal de abundância, seus derivados hoje, são considerados alguns dos maiores vilões dos danos ao meio ambiente, por causa das emissões de gás carbônico. Outro fato preocupante em relação ao petróleo é o esgotamento das reservas fósseis conhecidas, que tendem a se exaurir em menos de 40 anos, caso os patamares atuais de consumo persistam. Como se não bastasse, ainda há o fato da localização geográfica, isto é, os maiores poços estão numa região marcada por conflitos, deixando o mundo à mercê dos mandos e desmandos de uma minoria.

Há não muito tempo, a preocupação com as fontes de energia utilizadas era quase que exclusivamente de físicos e ambientalistas. A realidade atual é outra. Essas preocupações, hoje, angariam cada vez mais adeptos em todo o mundo, no sentido de encontrar alternativas viáveis para que o planeta não se transforme numa estufa.

As pesquisas em biomassa e biodiesel vêm mostrando resultados muito bons. A cana-de-açúcar é o tipo de biomassa de maior destaque, da qual se obtém o álcool, combustível suporte da gasolina. O biodiesel é um combustível obtido a partir de oleaginosas e que pode, gradativamente, vir a substituir o óleo diesel. Essas formas alternativas de energia são, portanto, vistas como suprimento ener-

gético de um futuro não muito distante.

Este trabalho consiste em uma modelagem univariada e multivariada com utilização das técnicas de séries temporais. Os modelos univariados são construídos com a finalidade de se obter previsões de produção das principais séries de energia renovável (cana-de-açúcar e álcool) e não renovável (petróleo, óleo diesel e gasolina). O objetivo de se estabelecer relações de dependência entre as séries dos combustíveis álcool, óleo diesel e gasolina levou à utilização de um modelo multivariado.

A previsão de produção dessas fontes favorece uma tomada de decisões, no sentido de planejar o volume de estoque para o consumo ao longo dos períodos subsequentes, principalmente agora, momento de transição da era fóssil para a era alternativa.

Este trabalho está organizado em cinco seções, incluindo esta introdução.

A segunda seção consiste de uma revisão bibliográfica sobre energia fóssil e renovável. Faz-se também uma revisão das técnicas temporais, a serem utilizadas na modelagem uni e multivariada.

Na terceira seção, em material e métodos, são descritas as séries a serem utilizadas no estudo, bem como os passos seguidos no ajuste dos modelos de Box e Jenkins e os passos para a construção do modelo multivariado.

Os resultados e discussões, na quarta seção, se dividem em duas subseções: na primeira, são apresentados, de forma detalhada, a modelagem univariada, as previsões e o melhor modelo para representar cada série. Na segunda, é apresentado o ajuste do vetor auto-regressivo multivariado (VAR).

Finalmente, na quinta seção, são apresentadas as conclusões referentes aos modelos univariados e multivariado ajustados.

1.1 Objetivo geral

Analisar as relações de dependência entre as séries de combustíveis e ajustar modelos de previsão para as séries de produção do setor energético brasileiro, constituído por fontes renováveis e não renováveis de energia. Entre as fontes renováveis, é abordada a biomassa, representada pelas séries de produção de cana-de-açúcar e de álcool; entre as não renováveis, são tomadas as séries de produção de petróleo, óleo diesel e gasolina.

1.2 Objetivos específicos

- Identificar características inerentes às séries em estudo, como tendência e sazonalidade por meio de testes específicos.
- Observar a presença e os efeitos de possíveis intervenções.
- Ajustar modelos que representem os dados adequadamente e fazer previsões de produção.
- Ajustar um modelo multivariado para as séries dos combustíveis álcool, óleo diesel e gasolina.

2 REFERENCIAL TEÓRICO

2.1 Energia

A energia sempre esteve presente na evolução da humanidade, não sendo possível imaginar a existência do homem sem ela. As primeiras formas de energia a serem controladas pelos ancestrais do homem foram a térmica e a luminosa, liberadas pelo fogo; a madeira era o principal combustível na época. As carroças movidas por animais na Suméria, os botes com velas no Egito e a roda d'água na Babilônia, em 3500 a.C., vieram em seguida. O carvão foi adotado em 1000 a.C.

Na era moderna, o setor energético foi marcado por dois fatores: a Revolução Industrial e a descoberta da maior fonte energética de todos os tempos, o petróleo. Esses dois acontecimentos selaram a dependência energética das sociedades modernas, cujo progresso material pode ser equilatado pelo seu consumo de energia (Goldemberg, 1979).

2.1.1 A Revolução Industrial

A Revolução Industrial, caracterizada pela substituição das manufaturas pelas máquinas, nos séculos XVIII e XIX, tinha como núcleo a conquista de uma seqüência de tecnologias, baseada em uma sucessão de combustíveis fósseis - carvão, petróleo, gás natural - e, no século XX, a conversão de combustíveis e energia hidráulica em uma nova forma de energia, a eletricidade.

Para Baumol (1989)*, citado em Leite (1997), a Revolução Industrial foi responsável por uma crescente demanda por energia e matérias-primas que o mundo nunca viu antes. O ritmo fantástico de expansão continuou durante o século XX.

*BAUMOL, William, J. et al. "Productivity and american leadership: the long view". The Mit Press, 1989.

Foi estimado que, nas primeiras décadas do século XX, mais energia foi consumida pela humanidade do que havia sido em todos os séculos anteriores de sua existência e, para cada período subsequente de 20 anos, foi utilizado mais energia do que na totalidade do passado.

2.1.2 O petróleo

Petróleo é o nome genérico e "ouro negro" é a denominação popular para uma das maiores descobertas energéticas de todos os tempos. Trata-se de um líquido oleoso inflamável, formado, principalmente, de hidrocarbonetos e pequenas quantidades de enxofre, oxigênio e nitrogênio. É encontrado na natureza, às vezes em fontes ou depósitos superficiais, mas, geralmente, abaixo da superfície terrestre, de onde pode ser retirado por meio de poços. Quando destilado, pode ser separado em frações principais, como gasolina, óleo combustível e óleo diesel, dentre outras (Goldemberg, 1979).

Na antigüidade, era usado para fins medicinais ou para lubrificação. Como combustível, entrou em cena em 1854, quando houve uma perfuração bem-sucedida no estado da Pensilvânia, Estados Unidos.

Dois fatos fizeram com que o petróleo fosse um recurso estratégico no século XX. O primeiro foi a segunda Revolução Industrial, com a industrialização dos transportes em série, o veículo automotor, aumentando o consumo de gasolina e de óleo diesel, impulsionando a propensão e a busca por mais poços de petróleo no mundo. Outro fator que tornou o petróleo o negócio do século foi a substituição do carvão mineral pelo óleo combustível nos navios, tornando-se a fonte energética da modernidade.

O petróleo consolidou-se como a principal fonte de energia do mundo. Sua extração conheceu uma progressão ininterrupta, ou quase ininterrupta, durante

mais de um século. Às vésperas da I Guerra Mundial, a extração do petróleo era relativamente pequena. Após a II Guerra Mundial, houve um grande crescimento no seu consumo, principalmente entre 1960 e 1973, representando 47% da fonte energética mundial (Brasil, 2002).

Segundo Leite (1997), no Brasil, somente com a crise cafeeira de 1929 e a Revolução de 1930, com o aumento das necessidades por derivados do petróleo, ocasionado pelo desenvolvimento da indústria, que os investimentos na descoberta de novos poços começaram a ser cogitados. O primeiro poço produtor foi aberto em 1939, em Lobato, próximo de Salvador. Portanto, até 1939 nada era extraído em território nacional e, somente houve exportação do óleo entre os anos de 1958 e 1963 e a partir de 1970. De 1930 até meados dos anos 1970, o petróleo constituiu-se na principal escolha para atender às necessidades de combustível líquido no Brasil. No período compreendido entre os anos de 1934 e 1975, o consumo aparente de petróleo bruto aumentou mais de 16.000 vezes. E, neste período, a maior parte do petróleo consumido no país era importado.

O uso de derivados do petróleo disseminou-se para grandes indústrias e para o setor de transportes, tendo como principal ícone a implantação da indústria automobilística no país.

A produção brasileira só foi intensificada após os anos 1970, quando ocorreram as primeiras crises mundiais de abastecimento causadas pelo cartel formado pela Organização dos Países Exportadores de Petróleo (OPEP). A produção chegou a oscilar ao redor de 50% das necessidades nacionais de consumo, até meados da década de 1990 (Brasil, 2002).

2.1.3 O petróleo e a crise energética

A versatilidade do petróleo e seus derivados e a facilidade do seu manuseio e transporte seriam razões suficientes para a sua crescente importância relativa. Mas, além disso, e apesar do aumento da demanda no século XX, foi possível manter o suprimento sempre satisfatório em escala mundial e a preço relativamente estável. Os preços do petróleo, no período de 1900 a 1970, oscilaram entre um máximo de *US\$4.00*/barril, em 1931, e um mínimo pouco inferior a *US\$ 2.00*/barril, em 1910. O preço real ficou praticamente inalterado, em torno de *US\$ 3.00*/barril, de 1948 até 1972 (Leite, 1997).

A abundância e o preço explicam em grande parte, de um lado, a complacência dos usuários com os desperdícios de energia e, de outro, o desestímulo de inovações tecnológicas, tanto na busca de maior eficiência como de novas formas de energia. Apenas a energia nuclear, resultante de pesquisas com outras motivações, viria, nas décadas de 1940 e 50, provocar novos desenvolvimentos tecnológicos.

A fase do comodismo energético se alterou no prazo curtíssimo em que ocorreram os dois choques dos preços do petróleo promovidos pela OPEP, o poderoso cartel que havia alcançado um consenso sobre a forma de agir desde dezembro de 1970 e que, de fato, agiu eficazmente em 1973. O comando das ações, que estivera, durante décadas, em poder de grandes empresas sediadas em países industrializados, passou para as mãos dos países exportadores, detentores de grandes reservas. Tornava-se, então, evidente que os preços do petróleo estavam baixos, comparados aos dos demais energéticos. O que poucos esperavam, no entanto, era que a elevação fosse tão forte e rápida quanto a que ocorreu entre 1973 e 1979 (Leite, 1997).

O primeiro choque do petróleo, em 1973, reduziu a produção e, consequen-

temente, elevou os preços nominais do barril de *US\$3.05* para *US\$10.73* em 1974, provocando um grande impacto na economia mundial (OPEC, 2002). Em contrapartida, os países importadores de petróleo buscavam uma forma de entrar em acordo com a OPEP para baixar os preços, mas todas as tentativas foram em vão. Perdeu-se, em consequência, e em parte, a confiança na garantia do suprimento do petróleo e na estabilidade do seu preço.

Segundo Maciel (1991), na década de 1970, o Brasil atravessava um período de extraordinária expansão de sua economia. A inflação encontrava-se estabilizada em 13% e a balança comercial apresentava-se equilibrada. Diante do contexto mundial das crises petrolíferas, a estratégia político-econômica adotada pelo Brasil foi a de não-comprometimento das até então elevadas taxas de crescimento econômico. Dessa forma, não foi observada uma redução significativa da importação ou do consumo de petróleo, após o choque de 1973.

Segundo Silva (1983), apenas 20% do total de petróleo consumido pelo Brasil, nesta época, era produzido aqui. O restante provinha de importações. Então, era de se esperar que a crise do petróleo viesse a refletir negativamente nas contas externas do país. De fato, a partir de 1974, a balança comercial apresentou-se seriamente deficitária. Os gastos com as importações de petróleo contribuíram sensivelmente para este quadro. Em 1972, o gasto do país com as importações desse produto foi de *US\$409* milhões e, dois anos depois, chegou a *US\$2.84* bilhões. Em suma, os efeitos na economia brasileira, devido à crise, levaram o país ao endividamento e à inflação.

O segundo choque, em 1979, quando os preços se elevaram de *US\$12.26*, em 1978, para *US\$30.57* o barril, acentuou a crise econômica no país, levando as autoridades governamentais a adotarem medidas para reestruturar o abastecimento energético. Uma dessas providências foi o uso de energias alternativas.

2.1.4 Energia alternativa

Energia alternativa ou renovável se refere a toda energia produzida por meio de fontes que ocorrem repetidamente na natureza, como, por exemplo, a energia solar, a eólica (proveniente dos ventos), a energia das marés, a energia hidrelétrica e a biomassa, dentre várias outras. Além de renovável e menos poluente que os derivados do petróleo, está mais bem distribuída geograficamente.

Essas fontes entraram no cenário mundial de pesquisas em razão das conseqüências trazidas pelos choques do petróleo e também pela iminência de esgotamento das reservas fósseis conhecidas. Os acidentes ambientais que vêm assolando de forma intensa e catastrófica a humanidade por causa da poluição também fazem parte dos motivos desse investimento. A Convenção de Mudança de Clima, assinada em 1992 pela maioria das nações industrializadas e o Protocolo de Kyoto prevêem que os países desenvolvidos mantenham a emissão de gases de efeito estufa em menores patamares, o que significa um maior esforço para o desenvolvimento de tecnologias que proporcionem aumento da eficiência no uso de combustíveis e de energias renováveis.

A energia renovável, da qual tanto se fala hoje como uma possível solução aos problemas energéticos, sustentou a humanidade por vários séculos. Porém, vinha sendo utilizada com uma tecnologia diversa das que se espera que possam vir a ser desenvolvidas (Leite, 1997).

Entre as fontes que foram citadas como sendo de energia alternativa, será dado destaque à energia de biomassa.

2.1.5 Biomassa

A biomassa é toda matéria orgânica, de origem vegetal, que, no processo de fotossíntese, captura a energia do sol e a transforma em energia química (ANEEL,

2002). Esta, por sua vez, pode ser convertida em eletricidade, combustível e calor, por meio de diversos processos produtivos.

Os combustíveis mais comuns da biomassa são os resíduos agrícolas; plantas, como a cana-de-açúcar, o eucalipto e a beterraba; o biogás (produzido pela biodegradação anaeróbica existente no lixo e dejetos orgânicos); lenha e carvão vegetal e alguns óleos vegetais (amendoim, soja, dendê), etc., que são colhidos com o objetivo de produzir energia.

A biomassa é formada pela combinação de dióxido de carbono da atmosfera e água na fotossíntese clorofiliana, que produz os hidratos de carbono - a energia solar é armazenada nas ligações químicas dos componentes estruturais da biomassa. Se a biomassa for queimada de modo eficiente, há produção de dióxido de carbono e água. Portanto, o processo é cíclico e renovável.

No início da década de 1940, a biomassa era responsável por cerca de 83% da oferta interna de energia do Brasil, dos quais 81% correspondia à lenha e 2% ao bagaço de cana. Com o uso crescente de derivados de petróleo, logo acompanhado de grande expansão da hidreletricidade, a biomassa passou a diminuir sua participação, década após década (Energia Brasil, 2006).

Porém, com a deteriorização do meio ambiente e o esgotamento iminente da mais palpitante fonte de energia, o petróleo, estima-se que, daqui a 20 anos, cerca de 30% do total de energia consumida pela humanidade seja proveniente da biomassa.

Os recursos renováveis representam cerca de 20% do suprimento total de energia no mundo, sendo 14% proveniente de biomassa e 6% de fonte hídrica. No Brasil, a proporção da energia total consumida é de cerca de 35% de origem hídrica e 25% de origem em biomassa, significando que os recursos renováveis suprem algo em torno de $\frac{2}{3}$ dos requisitos energéticos do país (Ambiente Brasil,

2005).

2.1.6 A cana-de-açúcar e o álcool

Além de ser um dos principais produtos agrícolas do Brasil, cultivada desde a época da colonização, a cana-de-açúcar é considerada o principal tipo de biomassa energética. Dela extraem-se o açúcar, um dos componentes da cesta básica nacional; o álcool, cogitado como o combustível mais indicado para substituir a gasolina e o bagaço, resultante da produção de açúcar e álcool, que pode ser aproveitado nas usinas termelétricas para a geração de energia. É um potencial enorme, pois a quantidade de bagaço produzida a cada safra representa 30% do volume de cana moída. Isso permite que as usinas de cana se tornem auto-suficientes, em termos de energia, podendo mesmo vender a eletricidade excedente. Estima-se que o potencial da cana-de-açúcar seja equivalente à metade da produção gerada em Itaipu (Energia Brasil, 2006).

Atualmente, o país é o maior produtor mundial de cana e açúcar e, de acordo com a UNICA (2005), o setor envolvendo cana, álcool e açúcar é um dos mais competitivos do mundo. Movimenta cerca de R\$ 40 bilhões por ano, o que corresponde a, cerca de, 2,4% do PIB nacional, além de ser um dos setores que mais empregam no país, com a geração de 3,6 milhões de empregos diretos e indiretos.

O Brasil foi um dos pioneiros na implantação de combustível renovável alternativo ao petróleo, em larga escala. O álcool é reconhecido mundialmente pelas suas vantagens ambientais, sociais e econômicas.

Com a experiência acumulada de produção em todo o país, desde a década de 1920, o Brasil apostou no álcool combustível, em 1975, por meio da criação do Programa Nacional do Álcool (Proálcool), como alternativa para dimi-

nuir sua vulnerabilidade energética. A utilização em larga escala do álcool, além de matéria-prima para a indústria química, deu-se em duas etapas: como álcool etílico hidratado carburante em carros movidos exclusivamente a álcool e como álcool anidro adicionado à gasolina, na proporção de 20% a 25%.

O programa de produção de álcool recebeu, no início, muitos incentivos do governo, como descreve Maciel (1991). O principal deles, na sua primeira fase, foram os subsídios implícitos nas condições de financiamentos. O crédito destinado às atividades do Proálcool, tanto para a área agrícola quanto para a industrial, além de subsidiado, era de disponibilidade generosa, quando comparado aos recursos disponíveis destinados a outros tipos de empréstimos.

Em 1979, o Proálcool conheceu sua segunda fase, quando ocorreu uma mudança radical em sua ampliação e objetivos, prevendo-se que, além da produção de álcool anidro para mistura carburante, o programa passaria a produzir álcool hidratado, com o objetivo de abastecer veículos movidos exclusivamente por esse combustível. A nova meta do programa estipulava alcançar a produção de 10,7 bilhões de litros de álcool, até 1985. Esse objetivo exigiria uma expansão anual da produção da ordem de 20,8%. A meta foi alcançada na safra 85/86, quando a produção ficou em torno de quase 12 bilhões de litros.

A partir da década de 1990, com a queda do preço do barril de petróleo, o governo percebeu que o álcool estava saindo caro para as contas públicas e, então, o Proálcool passou a apresentar um perfil diferente daquele verificado no início de sua implantação. As novas medidas econômicas, entre elas o corte dos subsídios, fizeram com que o programa se declinasse.

Nesse contexto, a desaceleração do Proálcool, nos anos 1990, representou significativa diminuição da frota de carros 100% a álcool e a desestabilização conjuntural do modelo. Entretanto, graças à mistura do álcool anidro na gasolina,

a produção do etanol se manteve, compensando a queda no consumo do álcool hidratado.

O contexto mudou de novo e o Brasil tornou-se o maior produtor de álcool mundial. A atividade possui enorme potencial de expansão, graças a fatores como: o combate mundial ao efeito estufa e à poluição local, que levou à substituição de aditivos tóxicos na gasolina; a valorização da segurança energética, buscando-se autonomia pela diversificação das fontes de energia utilizadas; o incremento da atividade agrícola, que permite a criação de empregos e a descentralização econômica.

Outro fato que traz de novo o álcool ao cenário energético é a utilização dos carros bicomustíveis, ou flex, que funcionam a álcool e a gasolina. O álcool também aparece na pauta de exportações brasileiras, com os dados mostrando uma recuperação nas vendas externas de álcool carburante. Observou-se um salto no volume exportado, de 226 milhões de litros para 2,39 bilhões de litros, de 2000 para 2004.

Deve-se ressaltar que a sustentação dinâmica da oferta e o consumo brasileiro do álcool viram-se sempre pressionados pela competição oscilante dos preços internacionais do petróleo e as tendências de atratividade do açúcar, o que exigia um complexo sistema de regulamentação para a garantia de sua estocagem e oferta.

2.1.7 O álcool e a gasolina

O álcool tem ganho reconhecimento, pela comunidade internacional, como uma das possíveis soluções aos problemas ambientais, destacando-se como um dos melhores candidatos a serem apoiados com políticas de financiamento, pois representa um dos mecanismos de desenvolvimento limpo (MDL), como estabelecido no Protocolo de Kyoto (Cana-de-Açúcar no Brasil, 2005).

Para ser um substituto eficaz da gasolina, especifica-se que o preço do litro do álcool deve ser em torno de 70% do preço do litro da gasolina, para que haja uma paridade de troca, considerando que o motor a álcool tem um consumo maior que o motor a gasolina.

Comparações entre as propriedades da gasolina e do álcool encontram-se na Tabela 2.1.

TABELA 2.1: Comparações entre as propriedades dos combustíveis álcool e gasolina.

Característica	Álcool	Gasolina
Evaporação	menos volátil	mais volátil
Composição	álcool + água	gasolina + álcool anidro
Energia Liberada	4800 kcal/L	8050 kcal/L
Disponibilidade	recurso renovável	não renovável
Rendimento	10,7 km/L	13,6 km/L

As principais vantagens do uso do álcool em relação à gasolina são:

- O álcool adicionado à gasolina permite a eliminação do chumbo tetraetila presente na gasolina. A emissão desse composto tem efeitos cancerígenos;
- redução das emissões de monóxido de carbono;
- o Brasil possui terras propícias e clima adequado à cultura da cana, além de possuir mão-de-obra excedente;
- o Protocolo de Kyoto determinou a diminuição gradativa das emissões de gases poluentes. Para substituir o metano presente na gasolina o mais indicado é o álcool;
- possibilidade de capturar recursos internacionais por meio do mecanismo de desenvolvimento limpo (MDL).

2.1.8 O biodiesel e o óleo diesel

O óleo diesel é um combustível derivado do petróleo, constituído basicamente por átomos de carbono, hidrogênio e em baixas concentrações, por enxofre, nitrogênio e oxigênio. É um produto inflamável, medianamente tóxico, volátil, límpido, isento de material em suspensão e com odor forte e característico. O óleo diesel é utilizado em motores de combustão interna e ignição por compressão (motores do ciclo diesel) empregados nas mais diversas aplicações, tais como automóveis, furgões, ônibus, caminhões, pequenas embarcações marítimas, máquinas de grande porte, locomotivas, navios e aplicações estacionárias (geradores elétricos, por exemplo). Em função dos tipos de aplicações, o óleo diesel apresenta características e cuidados diferenciados.

A partir do refino do petróleo obtêm-se, pelo processo inicial de destilação atmosférica, entre outras, as frações denominadas de óleo diesel leve e pesado, básicas para a produção de óleo diesel. A elas podem ser agregadas outras frações, como a nafta, o querosene e o gasóleo leve de vácuo, resultando no produto conhecido como óleo diesel. O atual modelo energético brasileiro é apoiado, entre outros pontos, no transporte de cargas em motores diesel, por via rodoviária, em detrimento do transporte ferroviário, fluvial ou cabotagem. Isso faz com que o óleo diesel seja o derivado propulsor do refino, correspondendo a 34% do volume do barril de petróleo. Na maioria dos outros países do mundo, esta demanda situa-se entre 15% e 25% do volume do barril de petróleo, sendo a gasolina o produto que comanda o refino, situação mais fácil de atender, em função das características dos petróleos e dos esquemas de refino disponíveis mundialmente.

Atualmente, 10% do diesel consumido no Brasil é importado. A comercialização anual é de cerca de 38,2 bilhões de litros, correspondendo a 57,7% do consumo nacional de combustíveis veiculares (Miragaya, 2005).

Em razão da alta demanda por este combustível e os problemas ocasionados pelo seu uso, alternativas de substituição ou como suporte de abastecimento vêm sendo estudados. O biodiesel é uma das soluções mais viáveis.

De acordo com o Miragaya (2005), devido à preocupação com a redução do volume de emissões de gases causadores do efeito estufa até 2012, como determina o Protocolo de Kyoto, a demanda mundial por combustíveis renováveis tem se expandido rapidamente nos últimos anos. Essa demanda também é verificada no Brasil, pela necessidade de diminuir a dependência de derivados de petróleo na matriz energética nacional e pelo incentivo à agricultura e às indústrias locais.

O biodiesel é um combustível de queima limpa, derivado de fontes naturais e renováveis, como soja, colza, palma, babaçu, nabo-forrageiro, girassol e mamona, dentre outros. Pode ser obtido também a partir de óleos de fritura ou gorduras animais. Estes componentes reagem com um álcool de cadeia curta, metanol ou etanol.

Existem registros da utilização do combustível de óleos vegetais ou de seus ésteres, conhecidos como biodiesel, desde a montagem dos primeiros protótipos de motores de ignição por compressão por Rudolph Diesel, no final do século XIX.

É uma alternativa viável, capaz de reduzir em até 78% as emissões de gases poluentes, como o dióxido de carbono e 98% de enxofre na atmosfera. O dióxido de carbono é um gás responsável pelo efeito estufa que está alterando o clima mundial.

Em dezembro de 2004, o governo brasileiro autorizou o uso comercial do biodiesel no Brasil. Inicialmente, foi autorizada a mistura de 2% de biodiesel (B2) ao diesel de petróleo até 2007; a partir de 2008, essa mistura será obrigatória e, a partir de 2013, esse percentual será de 5% (B5). Com a obrigatoriedade, estima-se que serão necessários 900 milhões de litros de biodiesel, em 2008 e 2,65 bilhões

de litros, em 2013.

2.1.9 Alguns trabalhos desenvolvidos

Trabalhos na literatura relativos à produção dos combustíveis álcool, óleo diesel e gasolina são escassos. Existem diversos estudos referentes ao consumo dos derivados de petróleo relacionados a variáveis macroeconômicas.

Burnquist e Bacchi (2002) apresentaram estimativas de elasticidades preço e renda da demanda por gasolina no Brasil, referentes ao período de 1973 a 1998. Os resultados obtidos indicam que, no curto prazo, a demanda por gasolina no país é inelástica a mudanças na renda real, dado que um aumento da ordem de 1% nessa variável resulta em incremento pouco expressivo, da ordem de 0,6% no consumo de gasolina. No longo prazo, conforme esperado, a elasticidade renda obtida apresentou-se relativamente mais elevada, embora pouco inferior à unidade, da ordem de 0,959. No que se refere à elasticidade preço da demanda, os resultados mostraram que o consumo de gasolina, no contexto da economia brasileira, é aparentemente pouco sensível a mudanças nos preços desse combustível, tanto no curto como no longo prazo.

Maciel (1991) analisou os aspectos relacionados à oferta e à demanda dos produtos envolvidos no Proálcool. Desenvolveu-se um modelo constituído por um conjunto de equações de comportamento e identidades para cada produto: álcool, gasolina e açúcar. As equações foram estimadas pelo método de mínimos quadrados ordinários. A partir deste modelo, fizeram-se simulações econométricas. Após a validação do modelo, passou-se à etapa de simulações. Os resultados obtidos mostraram que, a persistirem as tendências dos últimos anos, verificar-se-iam um déficit cada vez mais acentuado no abastecimento de álcool, evidenciando a divergência existente na estrutura que vem determinando a oferta e a demanda do

produto.

Assis & Lopes (1980) avaliaram o comportamento do consumo de gasolina e óleo diesel entre 1970 e 1977, obtendo estimativas de preço de curto e longo prazos, para o Brasil. Utilizaram a técnica "pooled regressions", juntamente com observações de séries temporais e estimaram a equação de consumo de gasolina e óleo diesel por mínimos quadrados ordinários, com variáveis dummies. Os resultados obtidos indicaram tanto baixas elasticidades-preço da demanda de gasolina e óleo diesel como baixa sensibilidade a variações na renda para o consumo de gasolina. O consumo mostrou-se sensível à variação na renda para o óleo diesel, com exceção da região Norte.

Moreira (1996) enfocou o consumo de óleo diesel no Brasil, empregando dois modelos de projeção de longo prazo do consumo, sendo um para o nível nacional e outro para análise regional. O autor utilizou um modelo para a realização de previsões condicionadas ao preço do combustível e ao nível de atividade (PIB) para a análise de abrangência nacional. A metodologia empregada foi a de análise de auto-regressão vetorial com correção de erros (VEC). Essa escolha foi feita por se tratar de um procedimento de estimação que permite testar a existência e avaliar os parâmetros das relações de longo prazo. Foram testadas algumas especificações alternativas do modelo, para a análise regional, não tendo sido avaliada a possibilidade de existirem relações de equilíbrio de longo prazo.

Brown (1980) analisou a utilização de álcool como substituto para os derivados de petróleo. Ele partiu de hipóteses, como crescimento econômico, produção de etanol, produção nacional de petróleo e preço por barril de petróleo. Investigou as diversas formas de suprir a demanda por derivados do petróleo com o uso do álcool. Concluiu que o álcool pode servir para diminuir a importação de petróleo, desde que combinado com uma política de substituição adequada, sendo

talvez necessário alterar a composição da frota de veículos.

2.2 Séries temporais

Uma série temporal é um conjunto de observações coletadas de forma seqüencial, ao longo do tempo. A dependência entre as observações é o que caracteriza as aplicações das técnicas de séries temporais, já que as metodologias estatísticas clássicas, para serem aplicadas, exigem independência dos dados. Vale ressaltar que, além do tempo, uma série pode ser função de outra variável, como, por exemplo, espaço, profundidade, etc.

A análise de uma série temporal pode ser feita no domínio do tempo ou no domínio da frequência, sendo os modelos propostos, respectivamente, paramétricos e não-paramétricos.

O objetivo da análise em séries temporais é a construção de modelos com propósitos determinados, tais como: investigar o mecanismo gerador da série temporal, fazer previsões de valores futuros, descrever o comportamento da série e procurar periodicidades relevantes nos dados.

2.2.1 Modelos para séries temporais

Conforme Morettin & Tolo (2004), os modelos utilizados para descrever séries temporais são processos estocásticos, controlados por leis probabilísticas. Um processo estocástico é definido como sendo uma coleção de variáveis aleatórias seqüenciadas no tempo e definidas em um conjunto de pontos T , que pode ser contínuo ou discreto. A variável aleatória no tempo t é denotada por Z_t , em que $t = 0, \pm 1, \pm 2, \dots \in T$.

2.2.2 Estacionaridade

Uma das suposições mais frequentes que diz respeito a uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Existem, tecnicamente, duas formas de estacionaridade: fraca (ou ampla, ou de segunda ordem) e estrita (ou forte). A estacionaridade forte é uma propriedade muito exigente e, em geral, de difícil verificação. Assim, faz sentido definir um conceito de estacionaridade baseado nos momentos de uma série temporal, mais precisamente nos momentos de primeira e segunda ordens. Isto é dado pela estacionaridade fraca.

Morettin & Toloi (2004) definem um processo estocástico como sendo fracamente estacionário ou estacionário de segunda ordem se:

- (i) $E\{Z(t)\} = \mu(t) = \mu, \forall t \in T$;
- (ii) $E\{Z^2(t)\} < \infty, \forall t \in T$;
- (iii) $\gamma(t_1, t_2) = cov\{Z(t_1), Z(t_2)\}$ é uma função de $|t_1 - t_2|$.

Esta definição confirma que um processo estocástico fica bem descrito por meio das funções média, variância e autocovariância.

2.2.3 Função de autocovariância e autocorrelação

Conforme Morettin & Toloi (2004), a função de autocovariância é definida por:

$$\gamma_\tau = E\{Z_t Z_{t+\tau}\}$$

em que Z_t é um processo estacionário real discreto de média zero e τ é chamado de lag e representa a defasagem no tempo.

A função de autocovariância (facv) satisfaz às seguintes propriedades:

- (i) $\gamma_0 > 0$;
- (ii) $\gamma_{-\tau} = \gamma_\tau$;
- (iii) $|\gamma_\tau| \leq \gamma_0$;
- (iv) γ_τ é não negativa definida, no sentido que

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\tau_j - \tau_k} \geq 0$$

para quaisquer reais a_1, \dots, a_n e τ_1, \dots, τ_n pertencentes aos inteiros.

A função de autocorrelação (fac) de um processo estacionário é definida por:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0},$$

em que $\tau \in Z$. Essa função satisfaz às propriedades:

- (i) $\rho_0 = 1$
- (ii) $\rho_{-\tau} = \rho_\tau$
- (iii) $|\rho_\tau| \leq \rho_0 = 1$
- (iv) ρ_τ é positiva definida.

O estimador da função de autocovariância γ_τ é:

$$c_\tau = \frac{1}{N} \sum_{t=1}^{N-\tau} (Z_t - \bar{Z})(Z_{t+\tau} - \bar{Z}),$$

em que $\tau = 1, \dots, N - 1$, N é o número de observações da série e $\bar{Z} = \frac{1}{N} \sum_{i=1}^N Z_i$ é a média amostral. A estimativa da função de autocorrelação ρ_τ é dada por:

$$r_\tau = \frac{c_\tau}{c_0},$$

sendo c_τ a função de autocovariância no *lag* τ e c_0 a variância.

2.2.4 Decomposição clássica

Uma decomposição clássica de séries temporais permite que a série $\{Z_t, t = 1, 2, \dots, N\}$ seja escrita como uma soma ou multiplicação de componentes não-observáveis, como nas equações 2.1 e 2.2,

$$Z_t = T_t + S_t + a_t \quad (2.1)$$

$$Z_t = T_t S_t a_t \quad (2.2)$$

em que T_t é a tendência, S_t é a sazonalidade e a_t é a componente aleatória de média zero e variância constante σ_a^2 .

Para Morettin & Tolo (2004), a tendência pode ser entendida como aumento ou diminuição gradual das observações ao longo do tempo, a sazonalidade indica possíveis flutuações ocorridas sempre em períodos menores ou iguais a doze meses e a componente aleatória que mostra oscilações aleatórias irregulares.

O modelo 2.1 é dito aditivo, já que a componente sazonal é independente das outras componentes. O modelo 2.2 é utilizado para descrever a dependência das amplitudes sazonais em relação à tendência. Neste caso, é necessário aplicar uma transformação nos dados para se obter uma série com variância estável. Isso é feito tomando-se o logaritmo na equação 2.2, obtendo-se:

$$\log Z_t = \log T_t + \log S_t + \log a_t.$$

Nos primeiros estudos, era comum incluir-se, nos modelos 2.1 e 2.2, a componente cíclica de período longo, C_t , geralmente acima de doze meses. Hoje, o efeito do ciclo não é considerado, pois a sua retirada da série acarreta a perda de muitas observações provocando a redução do seu tamanho, prejudicando e dificultando o ajuste de um modelo.

Segundo Morettin & Toloi (2004), o principal interesse em considerar um modelo do tipo 2.1 e 2.2 é estimar a sazonalidade S_t e a tendência T_t , pois estas duas componentes estão intrinsicamente ligadas e, conforme Pierce (1979), a influência da tendência sobre a componente sazonal pode ser muito forte, por duas razões:

- (i) se a tendência não for levada em conta, métodos de estimação de S_t podem ser bastante afetados;
- (ii) a especificação de S_t depende da especificação de T_t .

Estas componentes fazem com que a série não atinja seu estágio estacionário, condição exigida na metodologia de ajuste dos modelos de Box e Jenkins. Uma vez estimadas T_t e S_t , elas são subtraídas de Z_t , restando apenas uma estimativa da componente aleatória a_t . Quando estas estimativas formam uma seqüência de variáveis identicamente e independentemente distribuídas (i.i.d) com média zero e variância constante, as variáveis são chamadas de choques aleatórios e a seqüência é denominada ruído branco.

2.2.5 Tendência

Supondo-se a ausência de sazonalidade, tem-se o modelo:

$$Z_t = T_t + a_t \quad t = 1, 2, \dots, n$$

em que a_t é um ruído branco, com variância σ_a^2 . Existem vários métodos para estimar tendência. Os mais utilizados são: ajuste polinomial e suavização (filtro), descritos em (Morettin & Tolo, 2004).

Estimando-se a tendência por meio de \hat{T}_t , a série ajustada para tendência ou livre de tendência é dada por:

$$Y_t = Z_t - \hat{T}_t.$$

Um procedimento também utilizado para eliminar a tendência consiste em tomar diferenças sucessivas da série original. A primeira diferença dada por:

$$\Delta Z_t = Z_t - Z_{t-1}$$

geralmente é suficiente para deixar a série estacionária.

A construção de gráfico na análise de séries temporais é uma ferramenta importante. Por meio dele é possível identificar características inerentes aos dados, como variabilidade, observações atípicas, sazonalidade e tendência, dentre outras. Mas, como os procedimentos visuais nem sempre são confiáveis, existem testes para confirmar a presença destas componentes.

Os testes para tendência, citados em Morettin & Tolo (2004), são: teste de seqüências (Wald-Wolfowitz), teste do sinal (Cox-Stuart) e o teste baseado no coeficiente de correlação de Spearman. Abaixo segue a descrição do teste do sinal

de Cox-Stuart, segundo Morettin & Toloi (2004).

2.2.5.1 Teste do sinal

O teste consiste em dividir a série em dois grupos, nos quais são comparadas as observações $(Z_1, Z_{1+c}), (Z_2, Z_{2+c}), \dots, (Z_{N-c}, Z_N)$, em que $c = \frac{N}{2}$ para as situações em que N é um número par e $c = \frac{N+1}{2}$, quando N for um número ímpar. É atribuído sinal "+" sempre que $Z_i < Z_{i+c}$ e sinal "-" para $Z_i > Z_{i+c}$, os empates são eliminados e n é o número de pares em que $Z_i \neq Z_{i+c}$, ou seja, é a soma de sinais "+" com os sinais "-". É testada a hipótese bilateral

$$H_0 : P(Z_i < Z_{i+c}) = P(Z_i > Z_{i+c}), \forall i : \text{não existe tendência;}$$

$$H_1 : P(Z_i < Z_{i+c}) \neq P(Z_i > Z_{i+c}), \forall i : \text{existe tendência.}$$

Definindo T_2 como o número de sinais positivos, rejeita-se H_0 , ou seja, a série apresenta tendência se $T_2 \geq n - t$, em que t é obtido por meio de uma distribuição binomial com parâmetros $(n, p = \frac{1}{2})$ e α um dado nível de significância. Para $n > 20$, utiliza-se a aproximação normal.

2.2.6 Sazonalidade

Da mesma forma que a tendência, a sazonalidade (ou periodicidade) constitui uma outra forma de não-estacionaridade e deve ser estimada e retirada da série. Sendo \hat{S}_t a estimativa de S_t , a série sazonalmente ajustada considerando um modelo aditivo é

$$Z_t^{SA} = Z_t - \hat{S}_t.$$

Os testes encontrados em Morettin & Toloi (2004) para a confirmação desta componente são os de Kruskal-Wallis, de Friedman e o teste F para análise

de variância.

Priestley (1989) propõe o teste de Fisher, usado para testar a presença de sazonalidade determinística, que é baseado na análise de uma quantidade chamada de periodograma, a qual é dependente das funções seno e cos-seno.

O periodograma é uma descrição dos valores observados numa realização de uma série através da sobreposição de ondas sinusoidais com várias frequências. A aplicação prática mais óbvia desta decomposição é a de servir de instrumento à identificação de componentes cíclicas ou periódicas.

Conforme Priestley (1989), a função periódica é dada por:

$$I_p(f_i) = \frac{2}{n} \left[\left(\sum_{t=1}^n a_t \cos \frac{2\pi i}{n} t \right) \left(\sum_{t=1}^n a_t \sin \frac{2\pi i}{n} t \right)^2 \right] \quad (2.3)$$

em que $0 < f_i < \frac{1}{2}$ e $t = 1, 2, \dots, n$. $I_p(f_i)$ é a intensidade da frequência f_i . A periodicidade de período $\frac{1}{f_i}$ pode ser observada pela existência de picos na frequência $f_i = \frac{1}{n}$.

2.2.6.1 Teste de Fisher

O teste de Fisher foi proposto, inicialmente, para testar o maior período. As hipóteses a serem testadas são:

H_0 : não existe sazonalidade

H_1 : existe sazonalidade

A estatística do teste é dada por:

$$g = \frac{\max I_p}{\sum_{p=1}^{N/2} I_p}$$

em que I_p é o valor do periodograma (dado pela equação 2.3) no período p e N é o número de observações da série. A estatística do teste de Fisher, z_α , é dada por:

$$z_\alpha = 1 - \left(\frac{\alpha}{n}\right)^{\frac{1}{n-1}}$$

em que $n = \frac{N}{2}$ e α é o nível de significância do teste.

Se $g > z_\alpha$, rejeita-se H_0 , ou seja, a série apresenta periodicidade p .

2.2.6.2 Estimação da sazonalidade

A sazonalidade encontrada na série pode ser determinística ou estocástica. É determinística quando pode ser prevista a partir de meses anteriores. A sua eliminação da série pode ser feita pela utilização de diferenças, cuja ordem corresponde ao valor do período p encontrado pelo periodograma.

A outra forma de eliminação é pela estimação dada pelo método de regressão. Este método consiste em considerar o modelo $Z_t = T_t + S_t + a_t$, em que $T_t = \sum_{j=0}^m \beta_j t^j$ e $S_t = \sum_{j=1}^{12} \alpha_j d_{jt}$, d_{jt} são variáveis periódicas (senos, cossenos ou variáveis sazonais "dummies") e a_t é ruído branco, com média zero e variância σ_a^2 .

Supondo sazonalidade constante, α_j não depende de t . Então

$$d_{jt} = \begin{cases} 1, & \text{se o período } t \text{ corresponde ao mês } j, j=1, \dots, 12; \\ 0, & \text{caso contrário.} \end{cases}$$

Neste caso,

$$d_{1t} + d_{2t} + \dots + d_{12,t} = 1, \quad t = 1, \dots, N,$$

de modo que a matriz de regressão não é de posto completo, mas de posto $m + 12$ (aqui são $m + 13$ parâmetros: $\alpha_1, \dots, \alpha_{12}, \beta_0, \beta_1, \dots, \beta_m$). Impondo-se a restrição

adicional

$$\sum_{j=1}^{12} \alpha_j = 0$$

obtém-se um modelo de posto completo:

$$Z_t = \sum_{j=0}^m \beta_j t^j + \sum_{j=1}^{11} \alpha_j D_{jt} + a_t,$$

em que

$$D_{jt} = \begin{cases} 1, & \text{se o período } t \text{ corresponde ao mês } j; \\ -1, & \text{se o período } t \text{ corresponde ao mês } 12; \\ 0, & \text{caso contrário, } j=1, \dots, 11. \end{cases}$$

Deste modo, pode-se utilizar a teoria usual de mínimos quadrados e obter os estimadores de α_j e β_j , ou seja, para uma amostra Z_1, \dots, Z_N , obtém-se o modelo:

$$\mathbf{Z} = \mathbf{C}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \mathbf{a}$$

em que:

$$\mathbf{Z}_{N \times 1} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix}, \mathbf{C}_{N \times (m+1)} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & N & \cdots & N^m \end{bmatrix}, \boldsymbol{\beta}_{(m+1) \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix},$$

$$\mathbf{D}_{N \times 11} = \begin{bmatrix} D_{11} & D_{21} & \cdots & D_{11,1} \\ D_{12} & D_{22} & \cdots & D_{11,2} \\ \vdots & \vdots & \ddots & \vdots \\ D_{1N} & D_{2N} & \cdots & D_{11,N} \end{bmatrix}, \boldsymbol{\alpha}_{11 \times 1} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{11} \end{bmatrix}, \mathbf{a}_{N \times 1} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}.$$

A equação pode ser escrita na forma:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{a}$$

em que:

$$\mathbf{X} = [\mathbf{C} : \mathbf{D}] \quad \text{e} \quad \boldsymbol{\gamma} = \begin{bmatrix} \beta \\ \vdots \\ \alpha \end{bmatrix}$$

de modo que:

$$\hat{\boldsymbol{\gamma}} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z},$$

são os estimadores usuais de mínimos quadrados.

Se for estocástica, a estimação é feita pelo método de médias móveis. Detalhes deste método de estimação podem ser encontrados em Morettin & Tolo (2004).

2.2.7 Modelos de Box e Jenkins

Dentre os diversos métodos e modelos de previsão existentes, destacam-se os modelos de Box e Jenkins, cuja metodologia consiste em ajustar modelos auto-regressivos integrados de médias móveis, $\text{ARIMA}(p, d, q)$, a um conjunto de dados. Tais modelos se caracterizam, ainda, por serem simples e parcimoniosos; as previsões são bastante precisas, comparando-se favoravelmente com os demais métodos de previsão. Se classificam em modelos lineares estacionários e não-estacionários.

Segundo Morettin & Tolo (2004), a estratégia para a construção do modelo é baseada em um ciclo iterativo. Os estágios deste ciclo são: especificação, identificação, estimação e verificação. Esses autores ainda salientam que a fase

mais crítica desta iteração é a identificação, pois, se o modelo não for adequado, o ciclo é repetido voltando-se a este estágio. Um procedimento prático é identificar vários modelos e, dentre os que se ajustam, escolhe-se o que fornece o menor erro quadrático médio de previsão (EQMP), se este for o objetivo do ajuste.

A notação de operadores é bastante utilizada nestes modelos. Estes operadores são:

- (i) B - operador translação para o passado: este operador ocasiona uma defasagem de um período m para trás, cada vez que é utilizado. É definido por:

$$B^m Z_t = Z_{t-m};$$

- (ii) F - operador translação para o futuro: responsável por uma defasagem de um período de tempo para frente. É definido por:

$$F^m Z_t = Z_{t+m};$$

- (iii) Δ - operador diferença. Definido por:

$$\Delta Z_t = Z_t - Z_{t-1} = (1 - B)Z_t \implies \Delta = 1 - B. \quad (2.4)$$

De 2.4 segue que:

$$\Delta^n Z_t = (1 - B)^n Z_t \implies \Delta^n = (1 - B)^n;$$

- (iv) S - operador soma. Definido por:

$$S Z_t = \sum_{j=0}^{\infty} Z_{t-j} = (1 + B + B^2 + \dots)Z_t = (1 - B)^{-1}Z_t \implies S = \Delta^{-1}.$$

2.2.8 Modelos lineares estacionários

A metodologia de Box e Jenkins supõe que a série temporal é o resultado da passagem de um processo aleatório (ruído branco) por um filtro ou sistema linear, conforme a Figura 2.1.

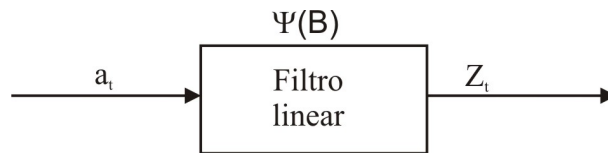


FIGURA 2.1: Filtro linear, com entrada a_t , saída Z_t , e função de transferência $\psi(B)$.

Formalmente, tem-se que:

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t, \quad (2.5)$$

em que:

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

é a função de transferência do filtro e μ é um parâmetro que determina o nível da série.

Uma condição para que o processo Z_t , dado em 2.5, seja estacionário, é que a seqüência de pesos $\{\psi_j, j \geq 1\}$ seja finita ou infinita convergente. Isso produz um filtro estável (somável) com μ representando a média do processo. Caso contrário, Z_t é não estacionária e μ não tem significado específico, a não ser como um ponto de referência para o nível da série. Z_t possui média e variância constantes e a covariância só depende de j , como conseqüências de a_t ser um ruído branco, isto é, $E(a_t) = 0, \forall t; \text{Var}(a_t) = \sigma_a^2, \forall t$ e $E(a_t a_s) = 0, s \neq t$.

Chamando $\tilde{Z}_t = Z_t - \mu$ em 2.5, tem-se que:

$$\tilde{Z}_t = \psi(B)a_t. \quad (2.6)$$

De forma alternativa, \tilde{Z}_t pode ser escrito como uma soma ponderada de valores passados mais um ruído, da seguinte forma:

$$\tilde{Z}_t = \pi_1 \tilde{Z}_{t-1} + \pi_2 \tilde{Z}_{t-2} + \dots + a_t = \sum \pi_j \tilde{Z}_{t-j} + a_t. \quad (2.7)$$

De 2.7 segue que:

$$(1 - \sum \pi_j B^j) \tilde{Z}_j = a_t$$

ou

$$\pi(B) \tilde{Z}_t = a_t, \quad (2.8)$$

em que $\pi(B)$ é o operador:

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$$

De 2.6 e 2.8 tem-se que:

$$\pi(B) \psi(B) a_t = a_t$$

de modo que:

$$\pi(B) = \psi^{-1}(B).$$

Esta relação pode ser usada para se obter os pesos π_j , em função dos pesos

ψ_j e vice-versa.

Um processo linear será estacionário se a série $\psi(B)$ convergir para $|B| \leq 1$ e será invertível se $\pi(B)$ convergir para $|B| \leq 1$.

Existem três casos particulares de processos lineares estacionários. São eles:

1. processo auto-regressivo de ordem p , representado por AR (p);
2. processo de médias móveis de ordem q , representado por MA(q);
3. processos auto-regressivos e de médias móveis de ordens p e q , representados por ARMA(p, q).

2.2.8.1 Processo auto-regressivo de ordem p - AR(p)

Um modelo auto-regressivo de ordem p , AR(p) é escrito em função de seus valores passados e do ruído branco. É denotado por:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t. \quad (2.9)$$

A equação 2.9 pode ser reescrita como:

$$\phi(B) \tilde{Z}_t = a_t$$

em que $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ é o polinômio auto-regressivo de ordem p . Isolando-se a_t em 2.9, tem-se:

$$\tilde{Z}_t - \phi_1 \tilde{Z}_{t-1} - \phi_2 \tilde{Z}_{t-2} - \dots - \phi_p \tilde{Z}_{t-p} = a_t. \quad (2.10)$$

Na notação de operadores

$$\tilde{Z}_{t-m} = B^m \tilde{Z}_t. \quad (2.11)$$

Aplicando-se a equação 2.11 em 2.10, tem-se que:

$$\tilde{Z}_t - \phi_1 B \tilde{Z}_t - \phi_2 B^2 \tilde{Z}_t - \dots - \phi_p B^p \tilde{Z}_t = a_t. \quad (2.12)$$

De 2.12 obtém-se:

$$\tilde{Z}_t(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) = a_t.$$

Sendo $\phi(B)$ finito, o processo $\text{AR}(p)$ é invertível para qualquer p e será estacionário se as raízes do polinômio, $\phi(B) = 0$, estiverem fora do círculo unitário.

2.2.8.2 Processo de médias móveis de ordem q - $\text{MA}(q)$

O processo de médias móveis é uma combinação linear de ruídos brancos ocorridos no período presente e no período passado da seguinte forma:

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (2.13)$$

Colocando a_t em evidência em 2.13 e usando-se a notação de operadores, tem-se:

$$\tilde{Z}_t(1 - \theta_1 B - \dots - \theta_q B^q) a_t = \theta(B) a_t,$$

em que:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

é o operador de médias móveis de ordem q .

A condição de invertibilidade para um modelo $MA(q)$ é que todas as raízes da equação característica, $\theta(B) = 0$, estejam fora do círculo unitário e não há restrições sobre os parâmetros θ_j para que o processo seja estacionário.

2.2.8.3 Modelos auto-regressivos e de médias móveis de ordem p, q - ARMA (p, q)

Os modelos auto-regressivos e de médias móveis são uma combinação linear dos modelos auto-regressivos com médias móveis, podendo, então, serem escritos da forma:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q},$$

ou, ainda, na forma compacta:

$$\phi(B)\tilde{Z}_t = \theta(B)a_t,$$

sendo $\phi(B)$ e $\theta(B)$, respectivamente, os operadores auto-regressivos e de médias móveis.

As condições de estacionaridade e invertibilidade são, respectivamente, que as raízes dos polinômios, $\phi(B) = 0$ e $\theta(B) = 0$ estejam fora do círculo unitário.

2.2.8.4 Função de autocorrelação parcial

Para facilitar a identificação de um modelo, Box, et al. (1994) propuseram a função de autocorrelação parcial (facp). O coeficiente de correlação parcial mede o grau de associação entre as observações Z_t e Z_{t-k} , quando os efeitos das

defasagens até $k - 1$ são fixados.

Seja ϕ_{kj} o j -ésimo coeficiente de um modelo $AR(k)$, de forma que ϕ_{kk} seja o último coeficiente. Sabe-se que:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}, \quad j = 1, \dots, k. \quad (2.14)$$

A partir de 2.14, são obtidas as equações de Yule-Walker

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \cdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdots \\ \rho_k \end{bmatrix}.$$

Resolvendo-se estas equações sucessivamente para $k = 1, 2, 3, \dots$ obtém-se:

$$\begin{aligned} \phi_{11} &= \rho_1, \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \end{aligned}$$

e, em geral:

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|}$$

em que ϕ_{kk} é a função de autocorrelação parcial (facp), P_k é a matriz de autocorrelações e P_k^* é a matriz P_k , com a última coluna substituída pelo vetor de autocorrelações.

2.2.8.5 Características das funções de autocorrelação e autocorrelação parcial dos processos lineares estacionários

2.2.8.5.1 Função de autocorrelação (fac)

- Processo auto-regressivo de ordem p -AR(p): constituído por uma mistura de polinômios, exponenciais e senóides amortecidas.
- Processo de médias móveis de ordem q -MA(q): é igual a zero para "lags" maiores do que q .
- Processo auto-regressivo e de médias móveis de ordem p, q - ARMA(p, q): possui (fac) infinita em extensão, a qual decai de acordo com exponenciais e ou senóides amortecidas após o "lag" $q - p$.

2.2.8.5.2 Função de autocorrelação parcial (facp)

- Processo auto-regressivo de ordem p -AR(p): possui (facp) $\phi_{kk} \neq 0$, para $k \leq p$ e $\phi_{kk} = 0$, para $k > p$.
- Processo de médias móveis de ordem q -MA(q): a (facp) se comporta de maneira similar à (fac) de um processo AR(p), ou seja, exponenciais e ou senóides amortecidas.

- Processo auto-regressivo e de médias móveis de ordem p, q - ARMA(p, q):
a (facp) se comporta como a (facp) de um processo MA puro.

2.2.9 Modelos não-estacionários

Grande parte das séries encontradas na prática apresenta alguma forma de não-estacionaridade e, como a maioria dos procedimentos utilizados em séries temporais é para séries estacionárias, é necessário tomar um número d de diferenças finitas para atingir este estágio. Quando isso é possível, tem-se um caso de séries não-estacionárias homogêneas ou séries portadoras de raízes unitárias.

Os modelos usados para séries com este comportamento são os modelos ARIMA e SARIMA.

2.2.9.1 Modelo auto-regressivo integrado de médias móveis - ARIMA (p, d, q)

Este tipo de modelo é o caso mais geral dos modelos de Box e Jenkins abrangendo todos os outros já vistos. A novidade é o I de integrado indicando a diferenciação da série para torná-la estacionária.

Sendo

$$W_t = \Delta^d Z_t \quad (2.15)$$

estacionário, W_t pode ser representado por um modelo ARMA(p, q), isto é,

$$\phi(B)W_t = \theta(B)a_t. \quad (2.16)$$

Se W_t for uma diferença da série original Z_t , então, Z_t é uma integral de W_t . Dessa forma, Z_t segue um modelo auto-regressivo, integrado, de médias móveis,

ou modelo ARIMA e, substituindo-se 2.15 em 2.16, obtém-se:

$$\phi(B)\Delta^d Z_t = \theta(B)a_t \quad (2.17)$$

de ordem (p, d, q) .

Tomando $\varphi(B) = \phi(B)\Delta^d = \phi(B)(1 - B)^d$ e substituindo em 2.17, tem-se:

$$\varphi(B)Z_t = \theta(B)a_t$$

em que (B) é um operador auto-regressivo não estacionário, de ordem $p + d$, com d raízes iguais a um (sobre o círculo unitário) e as restantes p fora do círculo unitário.

2.2.9.2 Modelo sazonal auto-regressivo integrado de médias móveis - SARIMA $(p, d, q) \times (P, D, Q)$

A componente periódica sazonal que se repete a cada s ($s > 1$) observações está presente em grande parte das séries temporais. Em dados mensais, por exemplo, com período sazonal $s = 12$, espera-se que Z_t dependa de Z_{t-12} e, talvez, de Z_{t-24} , além de Z_{t-1}, Z_{t-2}, \dots . Neste caso, tomar uma diferença não é suficiente para eliminar um padrão sazonal acentuado e, conseqüentemente, deixar a série estacionária. A forma apropriada é tomar diferenças no período sazonal.

Em geral, uma diferença sazonal é denotada por Δ_d .

Box & Jenkins (1976) generalizaram o modelo ARIMA para lidar com sazonalidade e definiram um modelo ARIMA sazonal multiplicativo, denominado SARIMA, representado por:

$$\phi(B)\Phi(B^s)\Delta^d\Delta_s^D Z_t = \theta(B)\Theta(B^s)a_t,$$

em que:

- $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ é o polinômio auto-regressivo de ordem p ;
- $\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_p B^{Ps}$ é o polinômio auto-regressivo sazonal de ordem P ;
- $\Delta^d = (1 - B)^d$ é o operador diferença e d é o número de diferenças necessárias para retirar a tendência da série;
- $\Delta_s^D = (1 - B^s)^D$ é o operador diferença generalizado, quando duas observações estão distantes entre si de s intervalos de tempos que apresentam alguma semelhança e D é o número de diferenças de lags s necessárias para retirar a sazonalidade da série;
- $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ é o polinômio de médias móveis de ordem q ;
- $\Theta(B) = 1 - \Theta_s B^s - \dots - \Theta_p B^{Qs}$ é o polinômio de médias móveis sazonal de ordem Q .

2.2.10 Estágios do ciclo iterativo

Conforme mencionado na subseção 2.2.7, os estágios do ciclo iterativo do método de Box e Jenkins são a identificação, a estimação e a verificação, dado que a classe geral de modelos ARIMA foi especificada.

2.2.10.1 Identificação

O objetivo da identificação é determinar os valores de p, d, q do modelo ARIMA (p, d, q) e (P, D, Q) , quando a série apresentar sazonalidade estocástica. A escolha das ordens é feita com base nas autocorrelações e autocorrelações parciais.

ais estimadas, que espera-se representar adequadamente as respectivas quantidades teóricas, as quais são desconhecidas.

O procedimento de identificação consiste de três etapas:

1. verificar a necessidade de transformação nos dados originais para estabilizar a variância por meio da análise gráfica;
2. tomar diferenças da série, tantas vezes quantas necessárias para se obter uma série estacionária. O número de diferenças, d , é alcançado quando a (fac) decai rapidamente para zero;
3. identificar o processo ARMA(p, q), por meio da análise das autocorrelações e autocorrelações parciais estimadas.

Segundo Morettin & Tolo (2004), esta é a fase mais crítica da utilização da metodologia de Box e Jenkins, devido à forma complicada da (fac) e (facp). Vários pesquisadores podem identificar modelos diferentes para uma mesma série, demandando, portanto, um pouco de experiência.

2.2.10.2 Estimação

Ao se identificar um modelo provisório para a série temporal, o próximo passo é estimar seus parâmetros e o método usado é o método da máxima verossimilhança. Considerando um modelo ARIMA(p, d, q), o número de parâmetros é $p + q + 1$ e $\xi = (\phi, \theta, \sigma_a^2)$ é o vetor desses parâmetros, em que $\phi = (\phi_1, \dots, \phi_p)$ são os parâmetros auto-regressivos, $\theta = (\theta_1, \dots, \theta_q)$ são os parâmetros de médias móveis e σ_a^2 é a variância da componente aleatória.

A estimação de ξ é feita pelo método da máxima verossimilhança. A função de verossimilhança, dadas as N observações Z_1, \dots, Z_N , é $L(\xi)|Z_1, \dots, Z_N$.

Os estimadores de máxima verossimilhança (EMV) de ξ são os valores que maximizam L ou $l = \log L$.

Segundo Box, et al. (1994), o logaritmo da função de verossimilhança não-condicional é dado por:

$$l(\xi) \simeq -n \log \sigma_a - \frac{S(\eta)}{2\sigma_a^2},$$

em que:

$$S(\eta) = S(\phi, \theta) = \sum_{t=-\infty}^n [a_t(\eta, W)]^2 \quad (2.18)$$

é a soma de quadrados não-condicional com $[a_t(\eta, W)] = E(a_t | \eta, W)$ e $W = W_t = \Delta^d Z_t$ é a série diferenciada com $n = N - d$ observações, sendo d o número de diferenças para obter estacionaridade.

Segue-se que os estimadores de mínimos quadrados (EMQ) obtidos minimizando-se 2.18, serão boas aproximações para os EMV.

2.2.11 Diagnóstico

Após identificar a ordem e estimarem-se eficientemente os parâmetros de um modelo, é necessário verificar se ele representa os dados de maneira adequada. Esta verificação é feita pelo teste da autocorrelação residual e pelo teste de Box-Pierce.

2.2.11.1 Teste da autocorrelação residual

Os resíduos podem ser estimados utilizando-se o modelo:

$$\phi(B)W_t = \theta(B)a_t \quad (2.19)$$

Assim:

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)W_t.$$

O modelo 2.19 é adequado quando os resíduos estimados estão próximos dos a_t e, portanto, devem ser aproximadamente não-correlacionados. Sob a suposição de o modelo ajustado ser apropriado, as estimativas das autocorrelações são dadas por:

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

seguinto uma distribuição normal $(0, \frac{1}{n})$.

Pode-se dizer que o modelo é adequado quando \hat{r}_k estiver dentro dos limites de $\pm \frac{2}{\sqrt{n}}$.

2.2.11.2 Teste de Box-Pierce

Box & Jenkins (1976) sugeriram o uso do teste de Box-Pierce para verificar se o resíduo é ruído branco. Tal teste é baseado nas k primeiras estimativas das autocorrelações, \hat{r}_k , dos resíduos. A estatística do teste é dada por:

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

Se o modelo ajustado for apropriado $Q(k) \sim \chi_{k-p-q}^2$, em que k é o número de "lags", p é a ordem da parte auto-regressiva do modelo e q é a ordem da parte de médias móveis, a hipótese de ruído branco será aceita se $Q(k) < \chi^2$ tabelado.

2.2.12 Previsões

Uma das formas de utilização de um modelo ajustado é fazer previsões de valores futuros. A previsão de Z_{t+h} , para $h = 1, 2, \dots$ é denotada por $\hat{Z}_t(h)$ e é definida como a esperança condicional de Z_{t+h} , dados todos os valores passados, isto é:

$$\hat{Z}_t(h) = [Z_{t+h}] = E[Z_{t+h}|Z_t, Z_{t-1}, \dots].$$

O erro de previsão é definido por:

$$e_t(h) = Z_{t+h} - \hat{Z}_t(h)$$

em que Z_{t+h} é o valor real e $\hat{Z}_t(h)$ é o valor predito.

2.2.13 Critérios para escolha do modelo

Conforme 2.2.10.1 vários modelos podem ser identificados para descrever uma série, mas existem critérios para escolha do melhor modelo de acordo com o objetivo do ajuste. Dentre diversos critérios, têm-se o critério de informação de Akaike (AIC) e o critério do erro quadrático médio de previsão (EQMP).

2.2.13.1 Critério de informação de Akaike (AIC)

Akaike (1973) sugere escolher o modelo cujas ordens p e q minimizem o critério.

O critério de Akaike, na comparação de diversos modelos, com N fixo, pode ser expresso por:

$$AIC = N \log \hat{\sigma}_a^2 + 2(k + l + 2),$$

em que $\hat{\sigma}_a^2$ é o estimador de máxima verossimilhança de σ_a^2 , $0 \leq k \leq p$ e $0 \leq l \leq q$.

O melhor modelo será aquele que apresentar menor AIC.

2.2.13.2 Critério do erro quadrático médio de previsão (EQMP)

Sendo o objetivo do ajuste a realização de previsão, o melhor modelo será o que apresentar o menor erro quadrático médio de previsão (EQMP).

As estimativas EQMP são dadas pela média dos quadrados das diferenças entre valores observados e valores preditos. Logo, o EQMP com origem em t é dado por:

$$EQMP_t = \frac{1}{N} \sum_{h=1}^n [Z_{t+h} - Z_t(h)]^2.$$

2.2.14 MAPE

O erro percentual médio absoluto (MAPE) é utilizado para verificar o bom desempenho de ajuste de um modelo. É dado por:

$$MAPE = \frac{1}{h} \sum_1^h \left| \frac{e_t(h)}{Z_{t+h}} \right| \times 100.$$

em que $e_t(h)$ é o erro de previsão, Z_{t+h} é o valor real e h é o número de previsões.

2.2.15 Análise de intervenção

Segundo Morettin & Toloi (2004), uma intervenção pode ser entendida como a ocorrência de algum tipo de evento, em dado instante de tempo T , conhecido a priori. Tal ocorrência pode manifestar-se por um intervalo de tempo subsequente e que afeta, temporária ou permanentemente, a série em estudo. A análise tem por objetivo avaliar o impacto de tal evento no comportamento da sé-

rie. Em geral, o efeito de uma intervenção é mudar o nível da série ou, então, a inclinação.

O efeito da intervenção pode ser obscurecido pela tendência, sazonalidade e erro aleatório.

As séries indicadoras de intervenções podem ser representadas por três tipos de variáveis binárias:

1. função degrau - o efeito da intervenção se mantém após o instante de tempo T , até o final da série

$$X_{j,t} = S_t^{(T)} = \begin{cases} 0 & , t < T; \\ 1 & , t \geq T. \end{cases} \quad (2.20)$$

2. função impulso - o efeito da intervenção é temporário, afetando a série apenas no instante T

$$X_{j,t} = I_t^{(T)} = \begin{cases} 0 & , t \neq T; \\ 1 & , t = T. \end{cases} \quad (2.21)$$

3. função intermediária - o efeito da intervenção afeta a série temporal apenas entre os instantes de tempo T_1 e T_2

$$X_{j,t} = \begin{cases} 0 & , t < T_1 \text{ e } t > T_2; \\ 1 & , T_1 \leq t \leq T_2. \end{cases}$$

Uma classe geral de modelos, que leva em conta a ocorrência de múltiplas intervenções, é dado por:

$$Z_t = \sum_{j=1}^k v_j(B) X_{j,t} + N_t, \quad (2.22)$$

em que:

- $X_{j,t}$, $j = 1, 2, \dots, k$, são variáveis de intervenção do tipo 2.20 e 2.21;
- $v_j(B)$, $j = 1, \dots, k$, são funções racionais da forma $\frac{\omega_j(B)B^{b_j}}{\delta_j(B)}$, em que $\omega_j(B) = \omega_{j,0} - \omega_{j,1}B - \dots - \omega_{j,s}B^s$ e $\delta_j(B) = 1 - \delta_{j,1}B - \dots - \delta_{j,r}B^r$ são polinômios em B , b_j é a defasagem no tempo para o início do efeito da j -ésima intervenção;
- N_t é a série temporal livre do efeito das intervenções e é denominada série residual cuja representação geral é:

$$\varphi(B)N_t = \theta(B)a_t.$$

2.2.15.1 Efeitos da intervenção

Uma intervenção para Morettin & Tolo (2004) pode afetar uma série temporal de várias maneiras. Na sua manifestação, ela pode ser abrupta ou gradual; na sua duração, pode ser permanente ou temporária. Os tipos mais comuns de efeitos de uma intervenção estão resumidos na Figura 2.2.

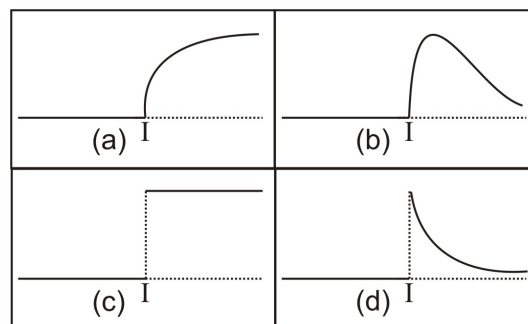


FIGURA 2.2: Efeito gradual permanente (a), gradual temporário (b), abrupto permanente (c) e abrupto temporário (d), sobre uma série temporal Z_t .

Uma mudança na variabilidade da série, após a intervenção, bem como um efeito de evolução pode aparecer: a série decai inicialmente e, depois, retoma o crescimento, até atingir um novo nível.

Para cada efeito de intervenção, tem-se uma forma apropriada para a função de transferência $v_j(B)$. A Figura 2.3 exhibe algumas formas que $v(B)$ pode assumir de acordo com X_t , nas equações 2.20 e 2.21, considerando por simplicidade o caso de uma única função de transferência

$$Z_t = v(B)X_t + N_t,$$

em que:

$$v(B) = \frac{\omega(B)}{\delta(B)}.$$

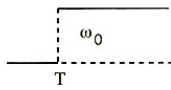
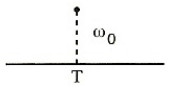
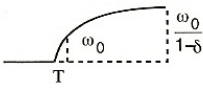
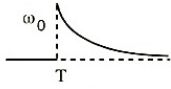
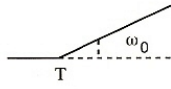
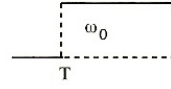
$v(B)$	$X_t = \begin{cases} 0, t < T \\ 1, t \geq T \end{cases}$	$X_t = \begin{cases} 0, t \neq T \\ 1, t = T \end{cases}$
ω_0		
$\frac{\omega_0}{1-\delta B}$ $ \delta < 1$		
$\frac{\omega_0}{1-B}$		

FIGURA 2.3: Estrutura da função de transferência $v_j(B)$ de uma série temporal.

Para os casos da Figura 2.3, têm-se:

- a) um efeito permanente após um início abrupto de mudança de nível.

$$v(B)X_t = \begin{cases} 0 & , t < T; \\ \omega_0 & , t \geq T; \end{cases}$$

- b) uma mudança de nível apenas no instante T.

$$v(B)X_t = \begin{cases} 0 & , t \neq T; \\ \omega_0 & , t = T; \end{cases}$$

- c) um efeito permanente após uma mudança de nível gradual até atingir a assíntota $\frac{\omega_0}{1-\delta}$. Neste caso, $v(B) = \frac{\omega_0}{1-\delta_0}$ e $Z_t = \delta Z_{t-1} + \omega_0 + N_t$. Segue-se que:

$$v(B)X_t = \begin{cases} 0 & , t < T; \\ \omega_0 \sum_{j=0}^k \delta^j & , t = T + k. \end{cases} \quad k = 0, 1, \dots$$

de modo que $\lim_{t \rightarrow \infty} v(B)X_t = \frac{\omega_0}{1-\delta}$

- d) uma mudança de nível apenas abrupta, sendo ω_0 o valor desta mudança e depois decai exponencialmente para zero.

$$v(B)X_t = \begin{cases} 0 & , t < T; \\ \delta^k \omega_0 & , t = T + k; \end{cases} \quad k = 0, 1, \dots$$

- e) uma mudança de direção da série, apresentando uma tendência determinística a partir do instante T. Neste caso, $\delta = 1$ e, após a intervenção, tem-se o

modelo $Z_t = Z_{t-1} + \omega_0 X_t + N_t$.

$$v(B)X_t = \begin{cases} 0 & , t < T; \\ (k+1)\omega_0 & , t = T+k; \end{cases} \quad k = 0, 1, \dots$$

f) um efeito permanente após uma mudança de nível abrupta, seguindo neste novo nível até a assíntota $\frac{\omega_0}{1-B}$.

$$v(B)X_t = \begin{cases} 0 & , t < T; \\ \omega_0 & , t \geq T. \end{cases}$$

2.2.16 Modelos lineares multivariados

O interesse neste tópico é estabelecer modelos para uma série temporal vetorial \mathbf{Z}_t , com n componentes, observadas em $t = 0, \pm 1, \pm 2, \dots$. A notação para esta situação é dada por: $\mathbf{Z}_t = (Z_{1t}, Z_{2t}, \dots, Z_{nt})'$, em que t pertence aos inteiros e Z_{it} ou $Z_{i,t}$ é a i -ésima componente, $i = 1, \dots, n$.

O vetor de médias de \mathbf{Z}_t é denotado por:

$$\boldsymbol{\mu}_t = E(\mathbf{X}_t) = (\mu_{1t}, \mu_{2t}, \dots, \mu_{nt})' \quad (2.23)$$

e a matriz de covariâncias de \mathbf{Z}_t é definida por:

$$\boldsymbol{\Gamma}(t+\tau, t) = E \left\{ (\mathbf{X}_{t+\tau} - \boldsymbol{\mu}_{t+\tau})(\mathbf{X}_t - \boldsymbol{\mu}_t)' \right\} \quad (2.24)$$

que é uma matriz $n \times n$ e que, em geral, depende de t .

As quantidades 2.23 e 2.24 descrevem as propriedades de primeira e segunda ordens das séries $Z_{1t}, Z_{2t}, \dots, Z_{nt}$. Se estas tiverem uma distribuição normal multivariada, as propriedades das séries serão completamente especificadas pelas

médias e covariâncias. Denotando por $\gamma_{ij}(t + \tau, t)$, $i, j = 1, \dots, n$ as componentes da matriz $\Gamma(t + \tau, t)$, então

$$\begin{aligned}\gamma_{ij}(t + \tau, t) &= \text{cov} \{Z_{i,t+\tau}, Z_{j,t}\} \\ &= E \{(Z_{i,t+\tau} - \mu_{i,t+\tau})(Z_{j,t} - \mu_{j,t})\}\end{aligned}$$

é a covariância entre as séries $Z_{i,t+\tau}$, $Z_{j,t}$, $i \neq j$.

Um caso de interesse é quando tanto o vetor de médias quanto a matriz de covariâncias não depende de t . Dessa forma, obtêm-se séries (fracamente) estacionárias.

2.2.16.1 Séries estacionárias

De acordo com Morettin (2006), a série n -variada \mathbf{Z}_t é estacionária se a média $\boldsymbol{\mu}_t$ e a matriz de covariâncias $\Gamma(t + \tau, t)$, t, τ pertencente aos inteiros, não dependem do tempo t . Nesta situação, tem-se:

$$\boldsymbol{\mu} = E(\mathbf{Z}_t) = (\mu_1, \dots, \mu_n)' \quad e$$

$$\Gamma(\tau) = E \left\{ (\mathbf{Z}_{t+\tau} - \boldsymbol{\mu})(\mathbf{Z}_t - \boldsymbol{\mu})' \right\} = [\gamma_{ij}(\tau)]_{i,j=1}^n.$$

Assim, $\gamma_{ii}(\tau)$ é a função de auto-covariância da série estacionária Z_{it} e $\gamma_{ij}(\tau)$ é a função de covariância cruzada da série Z_{it} e Z_{jt} . Em geral, $\gamma_{ij}(\tau) \neq \gamma_{ji}(\tau)$. Fazendo $\tau = 0$, obtém-se que:

$$\Gamma(0) = E \left\{ (\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_t - \boldsymbol{\mu})' \right\},$$

que é a matriz de covariâncias contemporâneas. Em particular, $\gamma_{ii}(0) = \text{Var}(Z_{it})$, $\gamma_{ij}(0) = \text{Cov}\{Z_{it}, Z_{jt}\}$.

O coeficiente de correlação contemporâneo entre Z_{it} e Z_{jt} é, então, dado por:

$$\rho_{ij}(0) = \frac{\gamma_{ij}(0)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}.$$

A matriz de correlações de lag τ é definida por:

$$\boldsymbol{\rho}(\tau) = \mathbf{D}^{-1}\boldsymbol{\Gamma}(\tau)\mathbf{D}^{-1}$$

em que $\mathbf{D} = \text{diag}\{\sqrt{\gamma_{11}(0)}, \dots, \sqrt{\gamma_{nn}(0)}\}$.

O coeficiente de correlação entre $Z_{i,t+\tau}$ e $Z_{j,t}$ é dado por:

$$\rho_{ij}(\tau) = \frac{\gamma_{ij}(\tau)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}.$$

Quando $\tau > 0$, este coeficiente mede a dependência linear de Z_{it} sobre Z_{jt} , que ocorreu antes do instante $t + \tau$.

As matrizes $\boldsymbol{\Gamma}(\tau)$ e $\boldsymbol{\rho}(\tau)$ não são, em geral, simétricas, mas valem as seguintes propriedades:

- (i) $\boldsymbol{\Gamma}(\tau) = \boldsymbol{\Gamma}'(-\tau)$
- (ii) $|\gamma_{ij}(\tau)| \leq [\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}$, $i, j = 1, \dots, n$.
- (iii) $\gamma_{ii}(\tau)$ é uma função de auto-covariância, para todo i .
- (iv) $\sum_{j,k=1}^m \mathbf{a}_j' \boldsymbol{\Gamma}(j-k) \mathbf{a}_k \geq 0$, para quaisquer m e $\mathbf{a}_1, \dots, \mathbf{a}_m$ vetores de \mathbb{R}^n .

A série $\{\mathbf{a}_t, t \in Z\}$ é um ruído branco multivariado $(n \times 1)$, com média $\mathbf{0}$ e matriz de covariâncias $\boldsymbol{\Sigma}$ se \mathbf{a}_t é estacionário com média $\mathbf{0}$ e matriz de covariâncias dada por:

$$\boldsymbol{\Gamma}(\tau) = \begin{cases} \boldsymbol{\Sigma} & \text{se } \tau = 0; \\ \mathbf{0} & \text{se } \tau \neq 0. \end{cases}$$

A notação usada é $\mathbf{a}_t \sim \text{RB}(\mathbf{0}, \Sigma)$. Se, além disso, os vetores \mathbf{a}_t forem independentes e identicamente distribuídos, escreve-se $\mathbf{a}_t \sim \text{IID}(\mathbf{0}, \Sigma)$.

Um processo \mathbf{Z}_t é dito linear se:

$$\mathbf{Z}_t = \sum_{j=0}^{\infty} \Psi_j \mathbf{a}_{t-j},$$

em que Ψ_j é uma seqüência de matrizes cujas componentes são absolutamente somáveis e \mathbf{a}_t é ruído branco multivariado. Segue-se que $E(\mathbf{Z}_t) = \mathbf{0}$ e a matriz de covariâncias de \mathbf{Z}_t é dada por:

$$\Gamma(\tau) = \sum_{j=0}^{\infty} \Psi_{j+\tau} \Sigma \Psi_j', \quad \tau \in Z.$$

2.2.16.2 Estimação de médias e covariâncias

Supondo-se as observações $\{\mathbf{X}_t, t = 1, \dots, T\}$ do processo estacionário, a média pode ser estimada pelo vetor de médias amostrais

$$\bar{\mathbf{Z}} = \frac{\sum_{t=1}^T \mathbf{Z}_t}{T}.$$

$\Gamma(\tau)$ é estimada por:

$$\hat{\Gamma}(\tau) = \begin{cases} \frac{1}{T} \sum_{t=1}^{T-\tau} (\mathbf{Z}_{t+\tau} - \bar{\mathbf{Z}})(\bar{\mathbf{Z}}_t - \bar{\mathbf{Z}})' & , 0 \leq \tau \leq T-1; \\ \frac{1}{T} \sum_{t=-\tau+1}^T (\mathbf{Z}_{t+\tau} - \bar{\mathbf{Z}})(\bar{\mathbf{Z}}_t - \bar{\mathbf{Z}})' & , -T+1 \leq \tau \leq 0. \end{cases}$$

E, finalmente, a matriz de correlações é estimada por:

$$\hat{\rho}(\tau) = \hat{\mathbf{D}}^{-1} \hat{\Gamma} \hat{\mathbf{D}}^{-1},$$

em que $\hat{\mathbf{D}}$ é a matriz diagonal $n \times n$ dos desvios padrões amostrais das séries

individuais.

2.2.16.3 Modelos auto-regressivos vetoriais - VAR(p)

O processo \mathbf{Z}_t , de ordem $n \times 1$, segue um modelo VAR(p) se:

$$\mathbf{Z}_t = \Phi_0 + \Phi_1 \mathbf{Z}_{t-1} + \dots + \Phi_p \mathbf{Z}_{t-p} + \mathbf{a}_t, \quad (2.25)$$

em que $\mathbf{a}_t \sim \text{RB}(\mathbf{0}, \Sigma)$, $\Phi_0 = (\phi_{10}, \dots, \phi_{n0})'$ é um vetor $n \times 1$ de constantes e Φ_k são matrizes $n \times n$ constantes, com elementos ϕ_{ij}^k , $i, j = 1, \dots, n$, $k = 1, \dots, p$. Sendo \mathbf{I}_n a matriz identidade de ordem n , o modelo 2.25 pode ser escrito na forma:

$$\Phi(B)\mathbf{Z}_t = \Phi_0 + \mathbf{a}_t$$

em que $\Phi(B) = \mathbf{I}_n - \Phi_1 B - \dots - \Phi_p B^p$ é o operador auto-regressivo vetorial de ordem p ou, ainda, um polinômio matricial $n \times n$ em B . O elemento genérico de $\Phi(B)$ é $[\delta_{ij} - \phi_{ij}^{(1)} B - \dots - \phi_{ij}^{(p)} B^p]$, para $i, j = 1, \dots, n$ e $\delta_{ij} = 1$, se $i = j$ e igual a zero, caso contrário.

2.2.16.4 Construção de modelos VAR

A construção de modelos VAR segue o mesmo ciclo de identificação, estimação e diagnóstico usado para modelos univariados da classe ARMA.

2.2.16.4.1 Identificação

De acordo com Morettin (2006), uma maneira de identificar a ordem p de um modelo VAR(p) consiste em ajustar seqüencialmente modelos auto-regressivos vetoriais de ordens $1, 2, \dots, k$ e testar a significância dos coeficientes (matrizes).

Considerando, pois, os modelos

$$\begin{aligned}
\mathbf{Z}_t &= \Phi_0^{(1)} + \Phi_1^{(1)} \mathbf{Z}_{t-1} + \mathbf{a}_t^{(1)}; \\
\mathbf{Z}_t &= \Phi_0^{(2)} + \Phi_1^{(2)} \mathbf{Z}_{t-1} + \Phi_2^{(2)} \mathbf{Z}_{t-2} + \mathbf{a}_t^{(2)} \\
&\dots = \dots \\
\mathbf{Z}_t &= \Phi_0^{(k)} + \Phi_1^{(k)} \mathbf{Z}_{t-1} + \Phi_k^{(k)} \mathbf{Z}_{t-k} + \mathbf{a}_t^{(k)},
\end{aligned}$$

os parâmetros podem ser estimados por MQ ordinários, que fornecem estimadores consistentes e eficientes. Testa-se, então,

$$\begin{aligned}
H_0 : \quad & \Phi_k^{(k)} = 0 \\
H_1 : \quad & \Phi_k^{(k)} \neq 0, k = 1, 2, \dots
\end{aligned} \tag{2.26}$$

O teste da razão de verossimilhanças é baseado nas estimativas das matrizes de covariâncias dos resíduos dos modelos ajustados. Para a k-ésima equação, considere

$$\hat{\mathbf{a}}_t^{(k)} = \mathbf{Z}_t - \Phi_0^{(k)} - \Phi_1^{(k)} \mathbf{Z}_{t-1} - \dots - \Phi_k^{(k)} \mathbf{Z}_{t-k}.$$

A matriz de covariância dos resíduos, que estima Σ , é dada, então, por:

$$\hat{\Sigma}_k = \frac{1}{T-k} \sum_{t=k+1}^T \hat{\mathbf{a}}_t^{(k)} (\hat{\mathbf{a}}_t^{(k)})', k \geq 0,$$

em que para $k = 0$, $\hat{\mathbf{a}}_t^{(0)} = \mathbf{Z}_t - \bar{\mathbf{Z}}$. A estatística da razão de verossimilhanças para o teste 2.26 é dada por:

$$RV(k) = (T-k) \ln \frac{|\hat{\Sigma}_{k-1}|}{|\hat{\Sigma}_k|},$$

que tem distribuição qui-quadrado com n^2 graus de liberdade, ou seja, $\chi^2(n^2)$.

Uma outra maneira de identificar a ordem de um VAR é usar algum critério de informação, como o de Akaike (AIC) e Schwarz (BIC), dados por:

$$AIC(k) = \ln(|\hat{\Sigma}_k|) + 2kn^2/T \quad (2.27)$$

$$BIC(k) = \ln(|\hat{\Sigma}_k|) + kn^2 \ln(T)/T. \quad (2.28)$$

Escolhe-se como o melhor modelo o que apresentar menor AIC e menor BIC.

2.2.16.4.2 Estimação

Identificado o valor de p e supondo-se $\mathbf{a}_t \sim N(\mathbf{0}, \Sigma)$, pode-se estimar os coeficientes por máxima verossimilhança. Neste caso, os estimadores de mínimos quadrados são equivalentes a estimadores de máxima verossimilhança condicionais que são obtidos por métodos de maximização numérica.

2.2.16.4.3 Diagnóstico

Os resíduos são usados para construir a versão multivariada da estatística de Box-Ljung-Pierce para testar se o modelo é adequado. Esta estatística é dada por:

$$Q(m) = T^2 \sum_{\tau=1}^m \frac{1}{T-\tau} \text{tr}(\hat{\Gamma}(\tau)' \hat{\Gamma}(0)^{-1} \hat{\Gamma}(\tau) \hat{\Gamma}(0)^{-1}),$$

que sob H_0 : a série \mathbf{a}_t é ruído branco, tem distribuição $\chi^2(n^2(m-p))$. Para que o número de graus de liberdade seja positivo, m deve ser maior do que p .

3 MATERIAL E MÉTODOS

Nesta seção, são descritas as séries utilizadas neste trabalho, bem como os procedimentos para o ajuste dos modelos univariados e também os passos para o ajuste multivariado.

As séries utilizadas neste trabalho foram:

- série de produção anual de cana-de-açúcar, em milhões de toneladas, coletada no período anual de 1947 a 1998, num total de 52 observações. As observações referentes aos anos de 1999 até 2004 foram reservadas para efeito de comparação com as previsões. Esses dados foram obtidos junto ao banco de dados do Instituto de Pesquisas Econômicas Aplicadas (IPEA) (IPEA, 2005) e se encontram na Tabela 1A;
- série de produção mensal de álcool de cana-de-açúcar nível 80 - índice (média 2002=100), coletada no período de janeiro de 1991 a dezembro de 2005, num total de 180 observações. As observações do período de janeiro de 2006 a setembro do mesmo ano foram reservadas para serem comparadas com as previsões. Esses dados também foram obtidos junto ao banco de dados do IPEA e se encontram na Tabela 2A;
- séries de produção de petróleo, óleo diesel e gasolina, todas em metros cúbicos (m^3), com 72 observações cada uma, coletadas mensalmente no período de janeiro de 2000 a dezembro de 2005. As observações de janeiro a setembro de 2006 para as séries de petróleo e óleo diesel foram deixadas para comparar com as previsões. Para a série de gasolina, foram reservadas as observações de maio a setembro do mesmo ano. Essas séries foram obtidas pelo banco de dados da Agência Nacional do Petróleo (ANP) (ANP, 2006)

e se encontram, respectivamente, nas Tabelas 3A, 4A e 5A.

A metodologia aplicada no ajuste dos modelos obedeceu os seguintes passos:

- (i) construção do gráfico da série original para captar indícios de tendência, sazonalidade, variações atípicas e dados discrepantes;
- (ii) construção dos gráficos das funções de autocorrelação (fac) e autocorrelação parcial (facp), também para observar características de tendência e (ou) sazonais;
- (iii) aplicação dos testes do sinal (Cox-Stuart) para verificar existência de tendência e de Fisher para verificar sazonalidade;
- (iv) caso, na etapa (iii), sejam detectadas tendência e (ou) sazonalidade, tomam-se diferenças para eliminar estas componentes. Normalmente, para eliminar tendência, 1 ou 2 diferenças são necessárias e para sazonalidade a diferença é da ordem do período;
- (v) nesta etapa, constroem-se os gráficos das funções de autocorrelação (fac) e autocorrelação parcial (facp) da série diferenciada para estimar as ordens do modelo;
- (vi) ajuste dos modelos ARIMA ou SARIMA pelo método da máxima verossimilhança e verificação da validade das estimativas;
- (vii) aplicação do teste de Box-Pierce para verificar se os resíduos são ruído branco;
- (viii) construção das previsões;

- (ix) aplicação do critério de AIC e do EQMP para escolha do melhor modelo ajustado aos dados. Escolhe-se o modelo que apresentar os menores valores de AIC EQMP;
- (ix) cálculo do MAPE - erro percentual médio absoluto para todos os modelos com a finalidade de mostrar o desempenho dos ajustes;
- (x) equação do modelo ajustado com as respectivas estimativas.

As séries usadas no ajuste do modelo multivariado são as séries de produção de álcool, óleo diesel e gasolina, por terem períodos coincidentes de observação e também porque é interessante conhecer o comportamento produtivo destes três combustíveis, isto é, determinar se existe alguma relação de dependência entre eles.

O ajuste foi feito obedecendo-se aos seguintes passos:

- (i) identificação da ordem p por meio dos critérios de AIC e BIC, conforme equação 2.27 e 2.28;
- (ii) estimação dos parâmetros por mínimos quadrados, os quais são equivalentes a estimadores de máxima verossimilhança condicionais e são obtidos por maximização numérica;
- (iii) verificação do ajuste por meio da versão multivariada da estatística de Box-Ljung-Pierce.

4 RESULTADOS E DISCUSSÃO

Os resultados obtidos neste trabalho são apresentados em duas subseções. Na primeira, são ajustados modelos univariados com e sem intervenção para as séries de cana-de-açúcar, álcool, petróleo, óleo diesel e gasolina. Na segunda, é ajustado um modelo multivariado para as séries dos combustíveis álcool, óleo diesel e gasolina.

4.1 Modelos univariados

4.1.1 Série de produção de cana-de-açúcar

Na Figura 4.1 é apresentado o gráfico da série original de produção de cana-de-açúcar. Visualmente, pode-se intuir a presença da componente tendência.

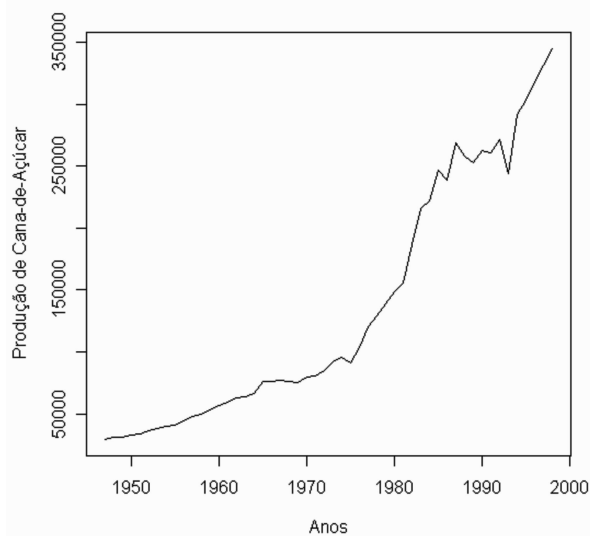


FIGURA 4.1: Representação gráfica da série original de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998.

A função de autocorrelação e o periodograma da série de produção anual de cana-de-açúcar são exibidos na Figura 4.2.

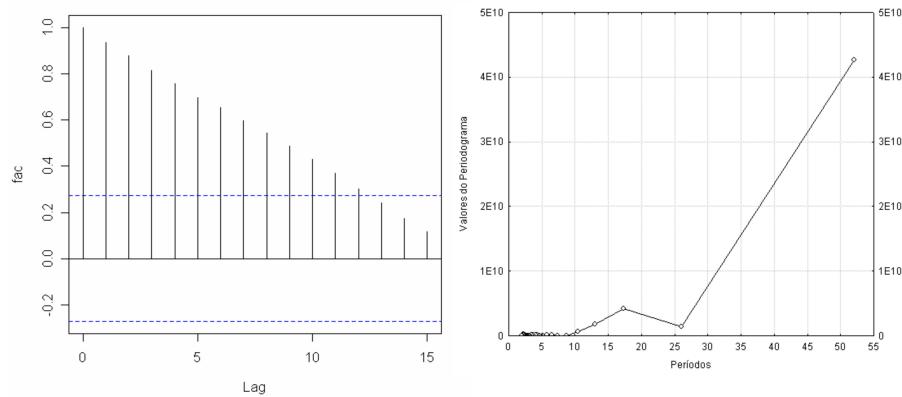


FIGURA 4.2: Representação gráfica da função de autocorrelação e periodograma da série de produção anual de cana-de-açúcar em milhões de toneladas no período de 1947 a 1998.

No gráfico da função de autocorrelação (Figura 4.2), é possível perceber que a série não decai rapidamente para zero, indicando a sua não estacionaridade. O teste do sinal (Cox-Stuart) foi aplicado para confirmar a existência da tendência. Considerando-se um nível de significância $\alpha = 0,05$ e tomando-se as 52 observações, tem-se que: $c = \frac{52}{2} = 26$ e $n=26$ ($Z_t \neq Z_{t+c} \forall t$). O número de sinais positivos T_2 é igual a 26 e $t = 17, 19$. Como $T_2 > 26 - 17, 19$, a série apresenta tendência.

O periodograma, ainda na Figura 4.2, fornece os possíveis períodos cíclicos e sazonais da série. Para o período sazonal, observa-se a não existência de picos em períodos menores ou iguais a 12 meses, indicando que a série não possui sazonalidade.

A tendência caracteriza um estágio não-estacionário e, como o ajuste dos modelos pressupõe estacionaridade, a série foi diferenciada uma vez, para eliminar

esta componente. A série diferenciada é mostrada na Figura 4.3.

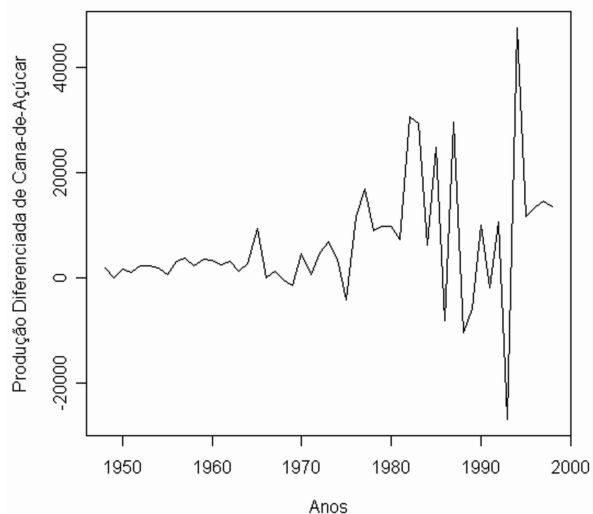


FIGURA 4.3: Representação gráfica da série diferenciada de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998.

Pela Figura 4.3 é possível observar várias oscilações na produção de cana, especialmente a partir de 1975. Essas oscilações podem ser efeito de intervenções que são bem explicadas pela fase de implantação do Proálcool - medida tomada pelo governo para solucionar a crise energética ocasionada pelas crises de abastecimento do petróleo. Iniciou-se um deslocamento do consumo dos combustíveis derivados dos fósseis para o álcool derivado da cana. Dentre as várias medidas tomadas pelo governo, podem-se citar as políticas de incentivo ao crédito para a produção de cana.

Inicialmente, existem indícios de intervenções nas observações 37 e 48, referentes aos anos de 1983 e 1994, respectivamente. Apesar disso, será considerada no ajuste como possível intervenção a referente ao ano de 1983. Para 1994,

não é possível identificar a significância, pois, após esta data, existem poucas observações.

As funções de autocorrelação e autocorrelação parcial da série diferenciada, usadas na identificação das ordens do modelo, são exibidas na Figura 4.4.

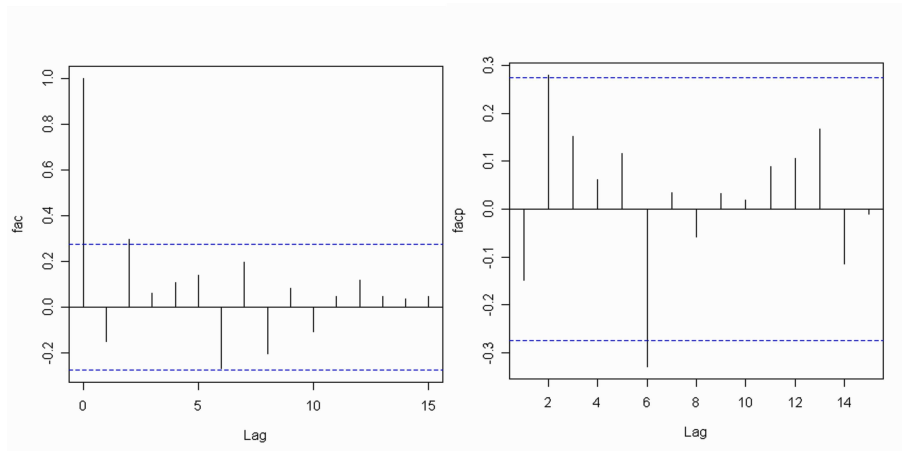


FIGURA 4.4: Representação gráfica das funções de autocorrelação (fac) e autocorrelação parcial (facp) da série diferenciada de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998.

A representação gráfica da fac e facp da série diferenciada na Figura 4.4 mostra que os lags 2 e 6 são significativos. A facp indica a ordem da parte autorregressiva e a fac indica a ordem da parte de médias móveis. A partir disso, um modelo proposto seria ARIMA(6,1,2). Este modelo, porém, não apresentou um bom ajuste, pois o resíduo não foi ruído branco.

Os modelos propostos foram, então:

- ARIMA (2,1,1) dado por:

$$Z_t = \frac{(1 - \theta_1 B)}{(1 - \phi_1 B - \phi_2 B^2)(1 - B)} a_t$$

- ARIMA (2,1,1) com uma intervenção na observação 37, referente ao ano de 1983. Sua representação é dada por:

$$Z_t = \frac{\omega_1}{1 - \delta_1 B} x_{1,t} + \frac{(1 - \theta_1 B)}{(1 - \phi_1 B - \phi_2 B^2)(1 - B)} a_t$$

em que a variável *dummy* $x_{1,t}$ é:

$$x_{1,t} = \begin{cases} 0, & t \neq 1983; \\ 1, & t=1983. \end{cases}$$

Na Tabela 4.1, são apresentadas as estimativas dos parâmetros dos modelos propostos.

TABELA 4.1: Estimativas dos parâmetros dos modelos ARIMA(2,1,1) e ARIMA(2,1,1) com intervenção para a série de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998.

Modelo	Parâmetro	Estimativa	Erro padrão	p-valor
ARIMA (2, 1, 1)	ϕ_1	0,521440	0,221801	0,022881
	ϕ_2	0,419828	0,160778	0,012003
	θ_1	0,619478	0,236612	0,011794
ARIMA (2, 1, 1)-CI	ϕ_1	0,476970	0,000000	0,000000
	ϕ_2	0,522940	0,000000	0,000000
	θ_1	0,858450	0,068150	0,000000
	ω_1	307220,0	10978,00	0,006650
	δ_1	0,776750	0,205640	0,000450

Na Tabela 4.2 são exibidas as estatísticas para o teste de Box-Pierce, o qual identifica se o modelo é ruído branco (independente e identicamente distribuído). Pela análise da Tabela 4.2, observa-se que os dois modelos constituem um ruído branco, já que $Q_{15} < \chi^2$ nos dois modelos.

TABELA 4.2: Estatísticas para o teste de Box-Pierce de modelos ARIMA com e sem intervenção.

Modelo	GL	Q_{15}	χ^2
ARIMA (2, 1, 1)	12	13,98	21,03
ARIMA (2, 1, 1) – CI	10	13,17	18,31

Os valores reais, as estimativas dos valores preditos para a produção de cana-de-açúcar e o erro de previsão, para o período de 1999 a 2004, encontram-se na Tabela 4.3.

TABELA 4.3: Valores reais de produção anual de cana-de-açúcar, em milhões de toneladas, para o período de 1999 a 2004 e valores preditos para o mesmo período.

Modelo	Ano	Z_{t+h}	$\hat{Z}_t(h)$	$e_t(h)$
ARIMA (2, 1, 1)	1999	333847	356816,1	-22969,1
	2000	326121	368572,3	-42451,3
	2001	344292	379556,5	-35264,5
	2002	364389	390219,7	-25830,7
	2003	396012	400391,4	-4379,4
	2004	415205	410172,0	5033,0
ARIMA (2, 1, 1)-CI	1999	333847	355030,3	-21183,3
	2000	326121	366885,1	-40764,1
	2001	344292	377696,8	-33404,8
	2002	364389	389088,0	-24699,0
	2003	396012	400202,3	-4190,30
	2004	415205	411481,4	3723,60

Nota: Z_{t+h} = valor real; $\hat{Z}_t(h)$ = valor predito; $e_t(h)$ = erro de previsão.

A escolha do melhor modelo foi feita com base no critério de informação de Akaike (AIC) e erro quadrático médio de previsão (EQMP). Esses dois critérios e o MAPE, que avalia o desempenho dos ajustes, podem ser vistos na Tabela 4.4

TABELA 4.4: Estimativas dos critérios de AIC, EQMP e MAPE de modelos ARIMA com e sem intervenção.

Modelo	AIC	EQMP	MAPE
ARIMA (2, 1, 1)	957,68	714.168.780	6,57%
ARIMA (2, 1, 1) – CI	955,55	644.631.274	6,19%

O modelo escolhido para representar a série de produção anual de cana-de-açúcar, em milhões de toneladas, é o ARIMA (2, 1, 1) com intervenção na observação 37 referente ao ano de 1983, por apresentar menor AIC e menor EQMP. Este modelo apresentou também um bom desempenho de ajuste.

O modelo substituído por suas respectivas estimativas é dado por:

$$Z_t = \frac{307220,0}{1 - 0,776750B} x_{1,t} + \frac{(1 - 0,858450B)}{(1 - 0,476970B - 0,522940B^2)(1 - B)} a_t$$

4.1.2 Série de produção de álcool

Na Figura 4.5 é mostrado o gráfico da série original de produção de álcool.

Visualmente, é possível intuir a presença de sazonalidade e ausência de tendência. Na Figura 4.6 está representada a função de autocorrelação da série original de produção mensal de álcool e o seu periodograma.

A estrutura da função de autocorrelação na Figura 4.6 é mais um indício de que a série apresenta componente sazonal, e o periodograma na mesma Figura apresenta um pico no período de 12 meses. Aplicando-se o teste de Fisher para testar a existência do efeito sazonal nesse período, obtiveram-se $g=0,90$ e $z_{0,05} = 0,08$. Como $g > z$, a série apresenta sazonalidade de 12 meses, a 5% de significância.

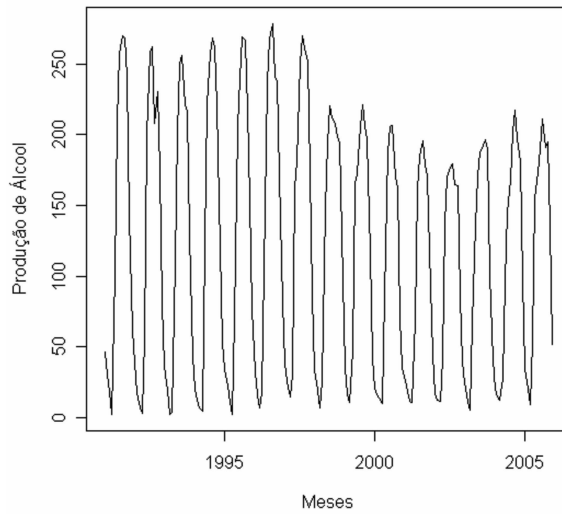


FIGURA 4.5: Representação gráfica da série original de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a dezembro de 2005.

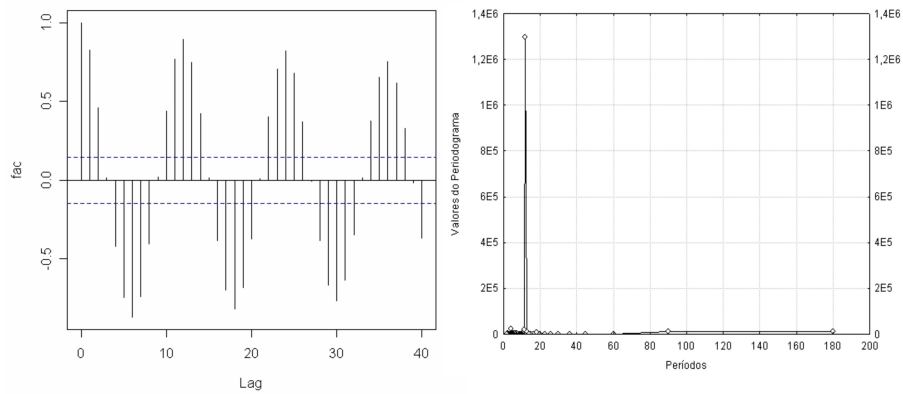


FIGURA 4.6: Representação gráfica da função de autocorrelação e periodograma da série de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a dezembro de 2005.

Segundo Morettin & Toloi (2004), havendo outra componente na série ela deve ser eliminada antes de se aplicar o teste para tendência. Como foi comprovada a componente sazonal, ela foi retirada com uma diferença de 12 e aí então foi aplicado o teste de Cox-Stuart, a 5% de significância. O resultado do teste foi: $c = \frac{168}{2} = 84$ e $n=84$ ($Z_t \neq Z_{t+c} \forall t$). O número de sinais positivos T_2 é igual a 30 e $t = 34,47$. Como $T_2 < 84 - 34,47$, a série não apresenta tendência.

A Figura 4.7 é a representação da série diferenciada de 12, ou seja, em estágio estacionário. Existem indícios de intervenções nas observações 65, 78 e 120, correspondentes, respectivamente, aos períodos de maio de 1996, junho de 1997 e dezembro de 2000. O efeito abrupto no aumento da produção refletido em maio de 1996 explica-se pela fase de redefinição do Proálcool. Nesta época, houve a instituição da chamada "frota verde" que foi o estímulo e a determinação do uso do álcool hidratado em determinadas classes de veículos leves, como os carros oficiais e táxis. Houve também o aumento de 22% para 24% do percentual de adição de álcool etílico anidro à gasolina. Os efeitos refletidos em junho de 1997 e dezembro de 2000, quando ocorreram quedas na produção, são explicados, em parte, pela diminuição do consumo de álcool hidratado, cujo preço encontrava-se acima de 70% do preço da gasolina e também porque houve um aumento no mercado internacional nos preços do açúcar, deslocando a produção de cana para este setor.

As funções de autocorrelação e autocorrelação parcial da série diferenciada que possuem a característica de identificar as ordens do modelo são apresentadas na Figura 4.8.

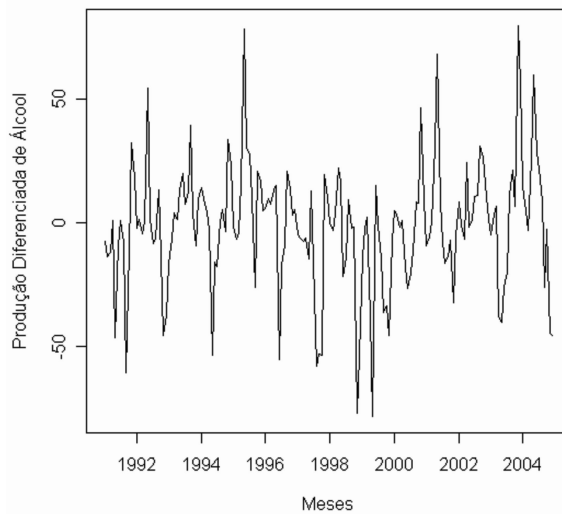


FIGURA 4.7: Representação gráfica da série diferenciada de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a dezembro de 2005.

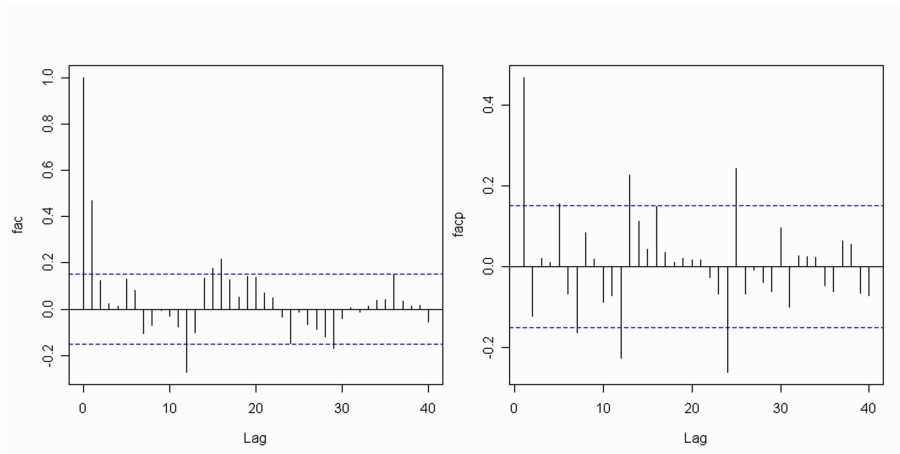


FIGURA 4.8: Representação gráfica das funções de autocorrelação (fac) e autocorrelação parcial (facp) da série diferenciada de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a dezembro de 2005.

Pela observação dessas funções, os modelos sugeridos foram:

- SARIMA (1, 0, 0)(0, 1, 1)₁₂ descrito por:

$$Z_t = \frac{1 - \Theta_1 B^{12}}{(1 - B^{12})(1 - \phi_1 B)} a_t$$

- SARIMA (1, 0, 0)(0, 1, 1)₁₂ com intervenções em maio/96, junho/97 e dezembro/2000. O modelo adequado para esta situação é dado por:

$$Z_t = \omega_1 x_{1,t} + \omega_2 x_{2,t} + \frac{\omega_3}{1 - \delta_3 B} x_{3,t} + \frac{1 - \Theta_1 B^{12}}{(1 - B^{12})(1 - \phi_1 B)} a_t$$

em que os efeitos de intervenção representados por $x_{1,t}$, $x_{2,t}$ e por $x_{3,t}$ são:

$$x_{1,t} = \begin{cases} 0, & t < 65; \\ 1, & t \geq 65; \end{cases}$$

$$x_{2,t} = \begin{cases} 0, & t < 78; \\ 1, & t \geq 78; \end{cases}$$

$$x_{3,t} = \begin{cases} 0, & t \neq 120; \\ 1, & t = 120. \end{cases}$$

As estimativas dos parâmetros dos modelos propostos são apresentadas na Tabela 4.5

Na Tabela 4.6 são exibidas as estatísticas para o teste de Box-Pierce, o qual identifica se o modelo é ruído branco (independente e identicamente distribuído).

Pela análise da Tabela 4.6, os modelos constituem um ruído branco já que $Q_{40} < \chi^2$.

TABELA 4.5: Estimativas dos parâmetros dos modelos SARIMA com e sem intervenções para a série de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a dezembro de 2005.

Modelo	Parâmetro	Estimativa	Erro padrão	p-valor
SARIMA (1, 0, 0)(0, 1, 1) ₁₂	ϕ_1	0,584330	0,065785	0,000000
	Θ_1	0,642768	0,068343	0,000000
SARIMA (1, 0, 0)(0, 1, 1) ₁₂ -CI	ϕ_1	0,540070	0,069640	0,000000
	Θ_1	0,665770	0,059970	0,000000
	ω_1	24,4040	10,5352	0,021788
	ω_2	-29,6270	11,0087	0,007866
	ω_3	-30,7567	12,6068	0,015778
	δ_3	0,966550	0,033200	0,000000

TABELA 4.6: Estatísticas para o teste de Box-Pierce de modelos SARIMA com e sem intervenções.

Modelo	GL	Q_{40}	χ^2
SARIMA (1, 0, 0)(0, 1, 1) ₁₂	38	39,28	53,38
SARIMA (1, 0, 0)(0, 1, 1) ₁₂ - CI	34	37,52	49,80

Os valores reais, as estimativas dos valores preditos para a produção de álcool e o erro de previsão, para o período de janeiro de 2006 a setembro do mesmo ano, encontram-se na Tabela 4.7

A escolha do melhor modelo foi feita com base no critério de informação de Akaike (AIC) e erro quadrático médio de previsão (EQMP). Esses dois critérios e o MAPE, o qual avalia o desempenho do ajuste, são mostrados na Tabela 4.8

TABELA 4.7: Valores reais de produção de álcool nível 80 - índice (média 2002=100), para o período de janeiro de 2006 a setembro do mesmo ano e valores preditos para o mesmo período.

Modelo	Meses	Z_{t+h}	$\hat{Z}_t(h)$	$e_t(h)$
SARIMA (1, 0, 0)(0, 1, 1) ₁₂	jan/06	24,78	20,2492	4,53080
	fev/06	9,43	12,4961	-3,06610
	mar/06	7,48	7,53430	-0,05430
	abr/06	65,66	41,3285	24,3315
	mai/06	169,6	127,9289	41,6711
	jun/06	197,85	166,861	30,9891
	jul/06	213,14	186,249	26,8908
	ago/06	223,15	203,569	19,5801
	set/06	212,04	196,228	15,8123
SARIMA (1, 0, 0)(0, 1, 1) ₁₂ -CI	jan/06	24,78	23,3523	1,42770
	fev/06	9,43	16,6772	-7,24720
	mar/06	7,48	12,1233	-4,64330
	abr/06	65,66	45,1557	20,5043
	mai/06	169,6	131,540	38,0598
	jun/06	197,85	171,941	25,9085
	jul/06	213,14	191,5108	21,6292
	ago/06	223,15	208,179	14,9711
	set/06	212,04	200,491	11,5491

Nota: Z_{t+h} = valor real; $\hat{Z}_t(h)$ = valor predito; $e_t(h)$ = erro de previsão

TABELA 4.8: Estimativas dos critérios de AIC, EQMP e MAPE de modelos SARIMA com e sem intervenção .

Modelo	AIC	EQMP	MAPE
SARIMA (1, 0, 0)(0, 1, 1) ₁₂	1014,58	519,47	17,52%
SARIMA (1, 0, 0)(0, 1, 1) ₁₂ -CI	1007,37	382,41	25,97%

O modelo escolhido para representar a série de produção mensal de álcool foi o SARIMA (1, 0, 0)(0, 1, 1)₁₂ com intervenção. Este modelo, além de possuir o menor AIC, possui o menor EQMP.

O modelo substituído por suas respectivas estimativas é dado por:

$$Z_t = 24,404x_{1,t} - 29,627x_{2,t} - \frac{30,7567}{1 - 0,96655B}x_{3,t} + \frac{1 - 0,66577B^{12}}{(1 - B^{12})(1 - 0,54007B)}a_t$$

4.1.3 Série de produção de petróleo

A representação gráfica da série de produção mensal de petróleo, em metros cúbicos, encontra-se na Figura 4.9.

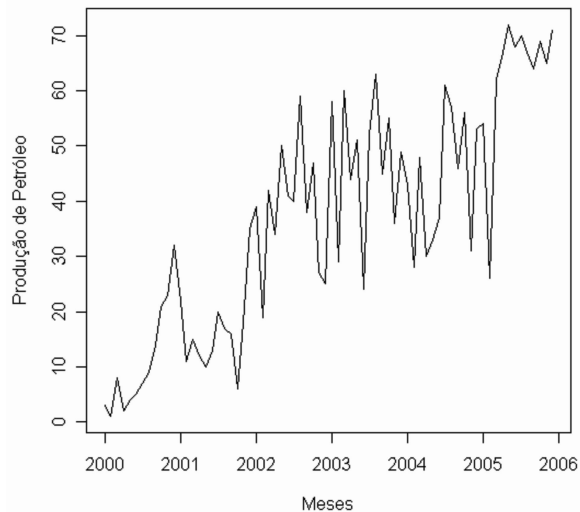


FIGURA 4.9: Representação gráfica da série original de produção mensal de petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.

Por meio de uma análise visual desta Figura, verifica-se a presença de tendência. Na Figura 4.10, está representada a função de autocorrelação da série original de produção mensal de petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005 e o seu periodograma.

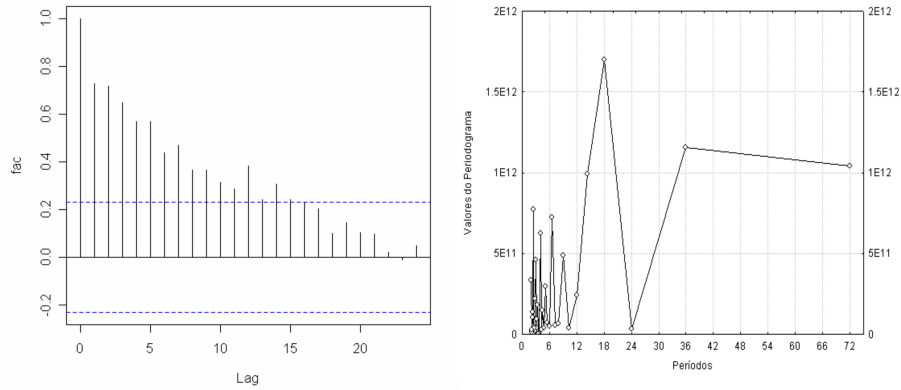


FIGURA 4.10: Representação gráfica da função de autocorrelação e periodograma da série de produção mensal de petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.

A estrutura da função de autocorrelação na Figura 4.10 é mais um indício de que a série apresenta componente de tendência, e o periodograma apresenta o maior pico no 2º período, sendo possível a existência de sazonalidade. As periodicidades observadas acima de 12 meses não foram consideradas.

Por meio do teste para sazonalidade proposto por Fisher, testa-se a existência de efeito sazonal no período de 2 meses. As estatísticas foram $g = 0,07$ e $z_{0,05} = 0,17$. Como $g < z$, a série não apresenta periodicidade 2.

O teste do sinal (Cox-Stuart), considerando um nível de significância $\alpha = 0,05$, para confirmação da tendência, resultou em: $c = \frac{72}{2} = 36$ e $n=36$ ($Z_t \neq Z_{t+c} \forall t$). O número de sinais positivos T_2 é igual a 36 e $t = 22,93$. Como $T_2 > 36 - 22,93$, a série apresenta tendência.

Na Figura 4.11 é exibida a série diferenciada, isto é, estacionária.

As funções de autocorrelação e autocorrelação parcial da série diferenciada que possuem a característica de identificar as ordens do modelo são apresentadas na Figura 4.12.

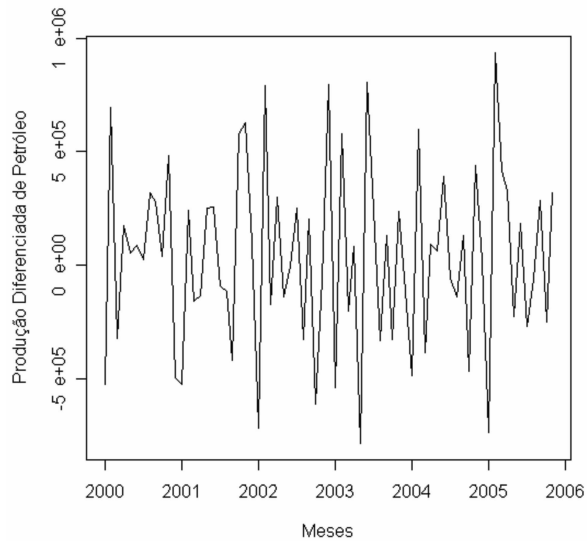


FIGURA 4.11: Representação gráfica da série diferenciada de produção mensal de petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.

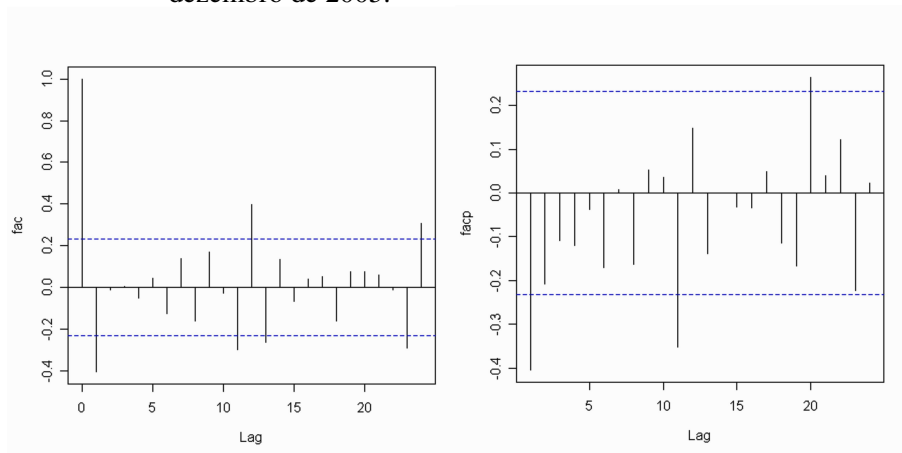


FIGURA 4.12: Representação gráfica das funções de autocorrelação (fac) e autocorrelação parcial (fACP) da série diferenciada de produção mensal petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.

A fac e a $facp$, na Figura 4.12, possuem correlações significativas em *lags* sazonais, sugerindo um modelo SARIMA. A fac possui correlação significativa em *lags* múltiplos de 12 e a $facp$ no *lag* 1.

Dessa maneira, os modelos sugeridos foram:

- SARIMA $(0, 1, 1)(1, 0, 0)_{12}$ descrito por:

$$Z_t = \frac{(1 - \theta_1 B)}{(1 - B)(1 - \Phi_1 B^{12})} a_t$$

- SARIMA $(1, 1, 0)(0, 0, 1)_{12}$ representado por:

$$Z_t = \frac{1 - \Theta_1 B^{12}}{(1 - B)(1 - \phi_1 B)} a_t$$

- SARIMA $(0, 1, 1)(1, 0, 0)_{12}$ com uma intervenção na observação 35 referente a novembro de 2002. O modelo adequado para esta situação é dado por:

$$Z_t = \omega_1 x_{1,t} + \frac{(1 - \theta_1 B)}{(1 - B)(1 - \Phi_1 B^{12})} a_t$$

em que:

$$x_{1,t} = \begin{cases} 0, & t < 35; \\ 1, & t \geq 35. \end{cases}$$

As estimativas dos parâmetros dos modelos propostos são apresentadas na Tabela 4.9

Na Tabela 4.10 são exibidas as estatísticas para o teste de Box-Pierce, o qual identifica se o resíduo é ruído branco (independente e identicamente distribuído).

TABELA 4.9: Estimativas dos parâmetros dos modelos SARIMA com e sem intervenções para a série de produção mensal de petróleo, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.

Modelo	Parâmetro	Estimativa	Erro padrão	p-valor
SARIMA	θ_1	0,399281	0,134200	0,004032
$(0, 1, 1)(1, 0, 0)_{12}$	Φ_1	0,380883	0,134772	0,006158
SARIMA	ϕ_1	-0,335995	0,115471	0,004864
$(1, 1, 0)(0, 0, 1)_{12}$	Θ_1	-0,318573	0,121872	0,010979
SARIMA	θ_1	0,36639	0,100000	0,007123
$(0, 1, 1)(1, 0, 0)_{12}$ -CI	Φ_1	0,446060	0,100000	0,001264
	ω_1	-644.093	293.055	0,031368

TABELA 4.10: Estatísticas para o teste de Box-Pierce dos modelos SARIMA com e sem intervenção.

Modelo	GL	Q_{24}	χ^2
SARIMA $(0, 1, 1)(1, 0, 0)_{12}$	22	21,25	33,92
SARIMA $(1, 1, 0)(0, 0, 1)_{12}$	22	27,01	33,92
SARIMA $(0, 1, 1)(1, 0, 0)_{12} - CI$	21	20,90	32,67

Pela análise da Tabela 4.10, os modelos constituem um ruído branco, já que $Q_{24} < \chi^2$.

Os valores reais, as estimativas dos valores preditos para a produção de petróleo e o erro de previsão, para o período de janeiro a setembro de 2006, encontram-se na Tabela 4.11.

Na Tabela 4.12 são exibidos os critérios de informação de Akaike (AIC), erro quadrático médio de previsão (EQMP) e o MAPE.

O modelo escolhido para representar a série de produção mensal de petróleo foi o SARIMA $(0, 1, 1)(1, 0, 0)_{12}$. Este modelo possui o menor EQMP e um bom desempenho de ajuste.

TABELA 4.11: Valores reais de produção de petróleo e valores preditos com e sem intervenção, para o período de janeiro de 2006 a setembro do mesmo ano.

Modelo	Meses	Z_{t+h}	$\hat{Z}_t(h)$	$e_{t(h)}$
SARIMA (0, 1, 1)(1, 0, 0) ₁₂	jan/06	8321362	8248032	73,330
	fev/06	7533460	7966998	-433,538
	mar/06	8358067	8324075	33992,0
	abr/06	8286098	8481444	-195346,0
	mai/06	8615745	8606988	8757,0
	jun/06	7776547	8521688	-745141,0
	jul/06	8502763	8590848	-88085,0
	ago/06	8393094	8488306	-95212,0
	set/06	8265597	8464548	-198951,0
SARIMA (1, 1, 0)(0, 0, 1) ₁₂	jan/06	8321362	8272821	48541,0
	fev/06	7533460	8099628	-566168,0
	mar/06	8358067	8340188	17879,0
	abr/06	8286098	8508377	-222279,0
	mai/06	8615745	8602301	13444,0
	jun/06	7776547	8501434	-724887
	jul/06	8502763	8546120	-43357,0
	ago/06	8393094	8711107	-78013,0
	set/06	8265597	8457161	-191564
SARIMA (0, 1, 1)(1, 0, 0) ₁₂ - CI	jan/06	8321362	8261516	59846,0
	fev/06	7533460	7932393	-398933,0
	mar/06	8358067	8350571	7496,0
	abr/06	8286098	8534869	-248771,0
	mai/06	8615745	8681896	-66151,0
	jun/06	7776547	8582000	-805453,0
	jul/06	8502763	8662995	-160232,0
	ago/06	8393094	8542905	-149811,0
	set/06	8265597	8515082	-249485,0

Nota: Z_{t+h} = valor real; $\hat{Z}_t(h)$ = valor predito; $e_t(h)$ = erro de previsão

O modelo substituído por suas respectivas estimativas é dado por:

$$Z_t = \frac{(1 - 0,399281B)}{(1 - B)(1 - 0,380883B^{12})} a_t.$$

TABELA 4.12: Estimativas dos critérios de AIC, EQMP e MAPE dos modelos SARIMA com e sem intervenção.

Modelo	AIC	EQMP	MAPE
SARIMA (0, 1, 1)(1, 0, 0) ₁₂	1815,70	$85,68 \times 10^9$	1,89%
SARIMA (1, 1, 0)(0, 0, 1) ₁₂	1820,49	$96,53 \times 10^9$	2,69%
SARIMA (0, 1, 1)(1, 0, 0) ₁₂ -CI	1812,47	99×10^9	2,99%

4.1.4 Série de produção de óleo diesel

A Figura 4.13 é a representação gráfica da série de produção mensal de óleo diesel, em metros cúbicos. Por meio de uma análise visual, verifica-se a presença de tendência.

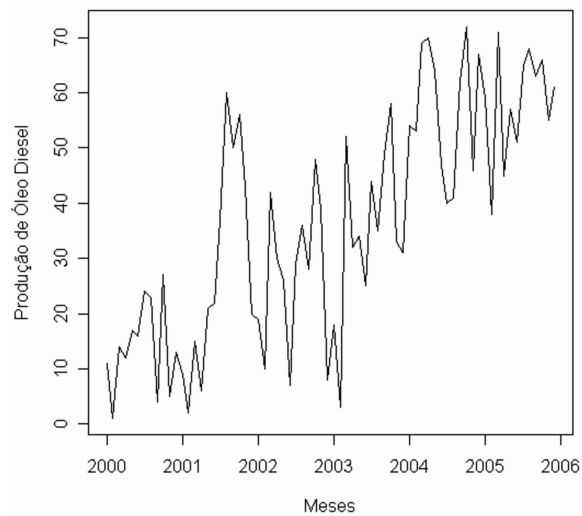


FIGURA 4.13: Representação gráfica da série original de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.

Na Figura 4.14 está representada a função de autocorrelação da série ori-

ginal de produção de óleo diesel, bem como seu periodograma.

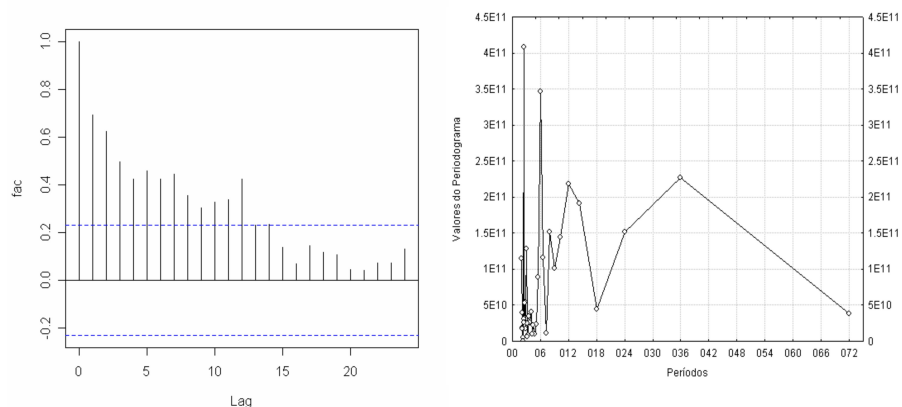


FIGURA 4.14: Representação gráfica da função de autocorrelação e periodograma da série de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.

Estes dois gráficos também auxiliam na identificação das componentes não-estacionárias. Aplicando-se o teste do sinal de Cox-Stuart para tendência, a um nível $\alpha = 0,05$, tem-se que: $c = \frac{72}{2} = 36$ e $n=36$ ($Z_t \neq Z_{t+c} \forall t$). O número de sinais positivos T_2 é igual a 35 e $t = 22,93$. Como $T_2 > 36 - 22,93$, a série apresenta tendência.

O periodograma usado para testar a sazonalidade apresenta um pico no 2º período. Aplicando-se o teste de Fisher, tem-se $g = 0,13$ e $z_{0,05} = 0,17$. Como $g < z$, a série não apresenta periodicidade 2. Os picos acima de 12 meses não foram considerados.

Na Figura 4.15, está representada a série diferenciada, isto é, sem tendência. Existem indícios de intervenções nas observações 30, 58 e 62, correspondentes, respectivamente, aos períodos de julho de 2002, outubro de 2004 e fevereiro de 2005.

As ordens do modelo são identificadas pelas funções de autocorrelação e autocorrelação parcial da série diferenciada na Figura 4.16.

Por meio desta representação, percebe-se que a fac e a $facp$ possuem correlação de lag 4. O modelo sugerido, a princípio, é um ARIMA(4,1,0). O ajuste deste modelo forneceu parâmetros não significativos. Como o lag 1 é significativo em ambas as funções, inclusive na parte sazonal, os modelos propostos foram:

- SARIMA (1, 1, 1)(1, 0, 0)₁₂ descrito por:

$$Z_t = \frac{(1 - \theta_1 B)}{(1 - B)(1 - \phi_1 B)(1 - \Phi_1 B^{12})} a_t$$

- SARIMA (1, 1, 1)(1, 0, 0)₁₂ com intervenções nas observações 30, 58 e 62, referentes, respectivamente, a junho de 2002, outubro de 2004 e fevereiro de 2005. O modelo adequado para esta situação é dado por:

$$Z_t = \frac{\omega_1}{1 - \delta_1 B} x_{1,t} + \frac{\omega_2}{1 - \delta_2 B} x_{2,t} + \frac{\omega_3}{1 - \delta_3 B} x_{3,t} + \frac{(1 - B)(1 - \theta_1 B)}{(1 - \phi_1 B)(1 - \Phi_1 B^{12})} a_t$$

em que:

$$x_{1,t} = \begin{cases} 0, & t \neq 30; \\ 1, & t=30. \end{cases}$$

$$x_{2,t} = \begin{cases} 0, & t \neq 58; \\ 1, & t=58. \end{cases}$$

$$x_{3,t} = \begin{cases} 0, & t \neq 62; \\ 1, & t=62. \end{cases}$$

Na Tabela 4.13 são apresentadas as estimativas dos parâmetros dos modelos propostos.

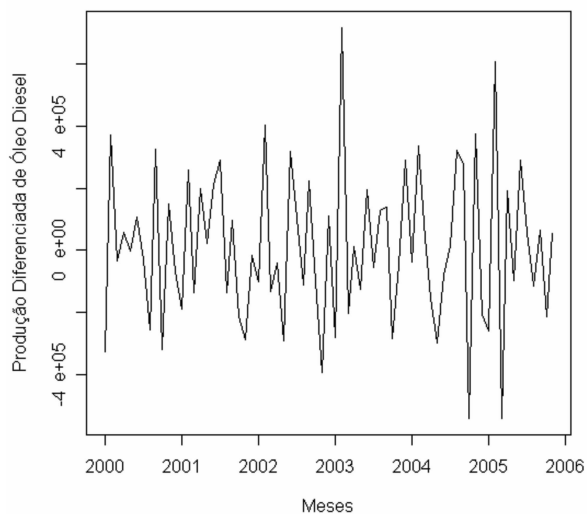


FIGURA 4.15: Representação gráfica da série diferenciada de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.

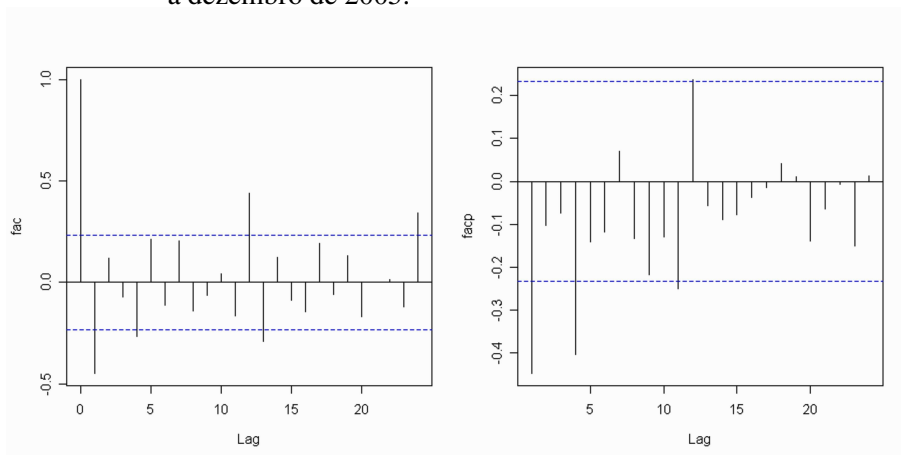


FIGURA 4.16: Representação gráfica das funções de autocorrelação (fac) e autocorrelação parcial (facp) da série diferenciada de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.

TABELA 4.13: Estimativas dos parâmetros dos modelos SARIMA com e sem intervenção, para a série de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro de 2000 a dezembro de 2005.

Modelo	Parâmetro	Estimativa	Erro padrão	p-valor
SARIMA (1, 1, 1)(1, 0, 0) ₁₂ -CI	ϕ_1	0,523430	0,133290	0,000219
	Φ_1	0,562880	0,116710	0,000000
	θ_1	0,913180	0,053720	0,000010
	ω_1	-280000,0	125300,0	0,028619
	δ_1	0,929560	0,071060	0,000000
	ω_2	207700,0	60766,00	0,001117
	δ_2	-0,954000	0,080030	0,000000
	ω_3	-506000,0	112600,0	0,000031
SARIMA (1, 1, 1)(1, 0, 0) ₁₂	ϕ_1	0,398750	0,168150	0,020562
	Φ_1	0,889750	0,087810	0,000000
	θ_1	0,507480	0,130640	0,000235

As estatísticas para o teste de Box-Pierce, o qual identifica se o modelo constitui um ruído branco, encontram-se na Tabela 4.14.

TABELA 4.14: Estatísticas para o teste de Box-Pierce dos modelos SARIMA com e sem intervenção.

Modelo	GL	Q_{24}	χ^2
SARIMA (1, 1, 1)(1, 0, 0) ₁₂	21	10,52	32,67
SARIMA (1, 1, 1)(1, 0, 0) ₁₂ - CI	15	15,66	24,99

Os resíduos, de acordo com a Tabela 4.14, são ruído branco, já que $Q_{24} < \chi^2$, em ambos os modelos.

Os valores reais, as estimativas dos valores preditos para a produção de óleo diesel e o erro de previsão, para o período de janeiro a setembro de 2006, encontram-se na Tabela 4.15

As estimativas do critério de Akaike, EQMP e MAPE encontram-se na

Tabela 4.16, na qual percebe-se que o modelo mais indicado para representar a série de produção mensal, em metros cúbicos, de óleo diesel é o SARIMA $(1, 1, 1)(1, 0, 0)_{12}$. Este modelo fornece previsões mais próximas dos valores reais, menor EQMP, menor AIC e um bom desempenho de ajuste. O modelo, com suas respectivas estimativas, é dado por:

$$Z_t = \frac{(1 - 0,913180B)}{(1 - B)(1 - 0,523430B)(1 - 0,562880B^{12})} + a_t.$$

TABELA 4.15: Valores reais de produção mensal de óleo diesel, em metros cúbicos, no período de janeiro a setembro de 2006 e valores preditos para o mesmo período.

Modelo	Meses	Z_{t+h}	$\hat{Z}_t(h)$	$e_t(h)$
SARIMA $(1, 1, 1)(1, 0, 0)_{12}$	jan/06	3212927	3155719	57206
	fev/06	2902481	3051440	-148959
	mar/06	3437810	3370182	67628
	abr/06	3231419	3099397	132022
	mai/06	3450654	3197733	252921
	jun/06	3245143	3149063	96080
	jul/06	3245143	3296741	-56315
	ago/06	3245143	3326530	-155719
	set/06	3245143	3267995	-125235
SARIMA $(1, 1, 1)(1, 0, 0)_{12}$ - CI	jan/06	3212927	3195357	17570
	fev/06	2902481	3301112	-398631
	mar/06	3437810	3191499	246311
	abr/06	3231419	3255013	-23594
	mai/06	3450654	3082318	368336
	jun/06	3245143	3250547	-5404
	jul/06	3245143	3241702	-1276
	ago/06	3245143	3410356	-239545
	set/06	3245143	3239484	-96724

Nota: Z_{t+h} = valor real; $\hat{Z}_t(h)$ = valor predito; $e_t(h)$ = erro de previsão

TABELA 4.16: Estimativas dos critérios de AIC, EQMP e MAPE dos modelos SARIMA com e sem intervenção

Modelo	AIC	EQMP	MAPE
SARIMA (1, 1, 1)(1, 0, 0) ₁₂	1718,74	$18,20 \times 10^9$	3,76%
SARIMA (1, 1, 1)(1, 0, 0) ₁₂ - CI	1735,13	$20,00 \times 10^9$	4,85%

4.1.5 Série de produção de gasolina

Para que fosse possível o ajuste do modelo da série de produção de gasolina foram tomadas 76 observações e apenas 5 foram reservadas para comparar com as previsões. Na Figura 4.17 é apresentado o gráfico da série original de produção de gasolina.

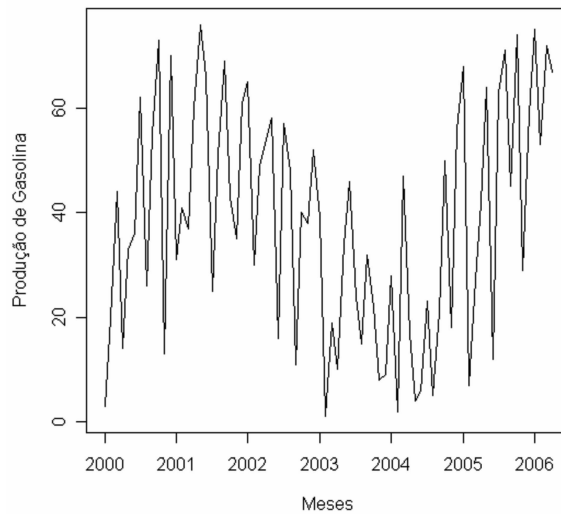


FIGURA 4.17: Representação gráfica da série original de produção mensal de gasolina, em metros cúbicos, no período de janeiro de 2000 a abril de 2006.

A função de autocorrelação e o periodograma da série de produção mensal

de gasolina são exibidos na Figura 4.18. No gráfico da função de autocorrelação é possível perceber que a série não decai rapidamente para zero. O teste do sinal (Cox-Stuart), a um nível $\alpha = 0,05$, foi aplicado para testar a existência da tendência. Tomando-se as 76 observações, tem-se que: $c = \frac{76}{2} = 38$ e $n=38$ ($Z_t \neq Z_{t+c} \forall t$). O número de sinais positivos T_2 é igual a 17 e $t = 24,07$. Como $T_2 > 38 - 24,07$, a série apresenta tendência.

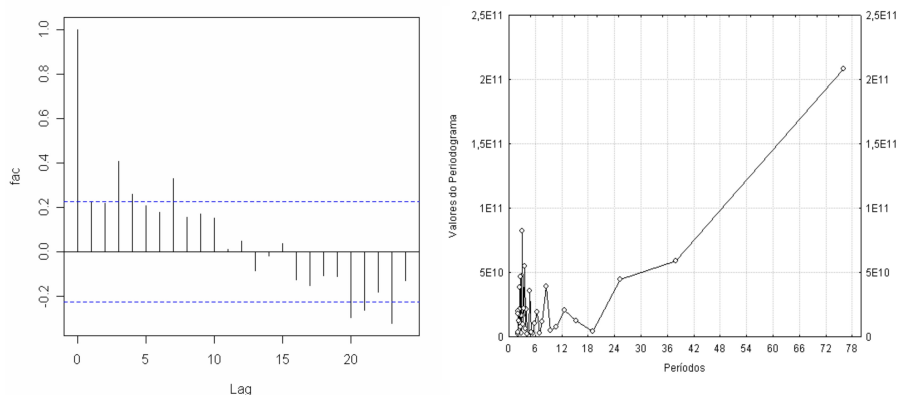


FIGURA 4.18: Representação gráfica da função de autocorrelação e periodograma da série de produção mensal de gasolina, em metros cúbicos, no período de janeiro de 2000 a abril de 2006.

Pelo periodograma, ainda na Figura 4.18, pode-se observar a presença de um pico no 3. O teste de Fisher resultou em: $g = 0,09$ e $z_{0,05} = 0,16$. Como $g < z$, a série não apresenta periodicidade 3.

A tendência caracteriza um estágio não-estacionário da série. Esta componente foi retirada e na Figura 4.19 é exibido o gráfico da série diferenciada, em que se percebe a ocorrência de possíveis intervenções.

A representação gráfica da fac e $facp$ da série diferenciada na Figura 4.20 mostra que os lags 1 da fac e 2 da $facp$ são significativos. A partir disso, os modelos propostos foram:

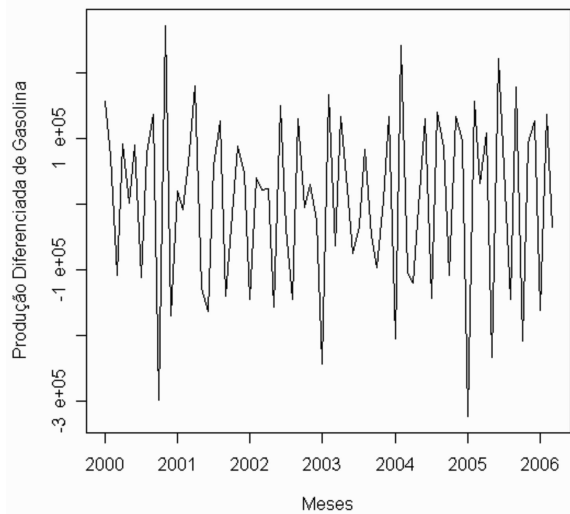


FIGURA 4.19: Representação gráfica da série diferenciada de produção mensal de gasolina, em metros cúbicos, no período de janeiro de 2000 a abril de 2006.

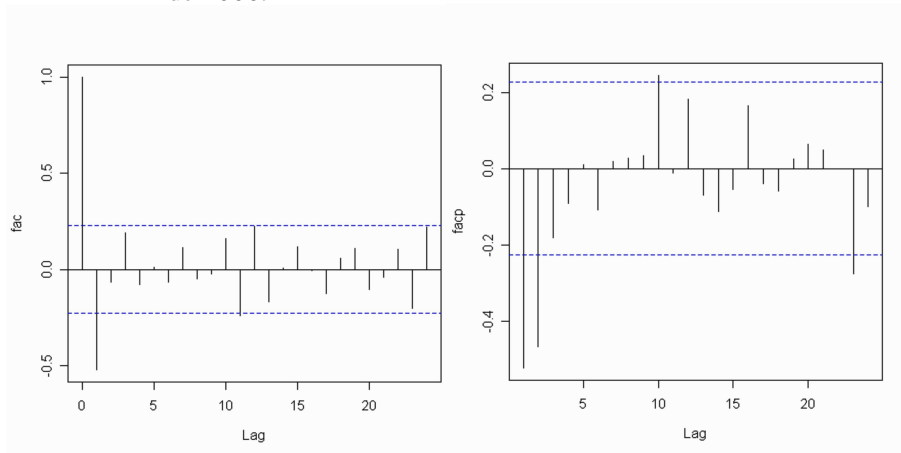


FIGURA 4.20: Representação gráfica das funções de autocorrelação (fac) e autocorrelação parcial (facp) da série diferenciada de produção mensal de gasolina, em metros cúbicos, no período de janeiro de 2000 a abril de 2006.

- ARIMA (2,1,0) dado por:

$$Z_t = \frac{a_t}{(1 - \phi_1 B - \phi_2 B^2)(1 - B)}$$

- ARIMA (0,1,1) dado por:

$$Z_t = \frac{(1 - \theta_1 B)a_t}{1 - B}$$

- ARIMA (2,1,0) com intervenções nas observações 16, 38, 41 e 61 referentes, respectivamente, aos meses de abril de 2001, fevereiro de 2003, maio de 2003 e janeiro de 2005. A representação para esta situação é:

$$Z_t = \omega_1 x_{1,t} + \omega_2 x_{2,t} + \frac{\omega_3}{1 - \delta_3 B} + \frac{\omega_4}{1 - \delta_4 B} + \frac{a_t}{(1 - \phi_1 B - \phi_2 B^2)(1 - B)}$$

em que as variáveis *dummy* são:

$$x_{1,t} = \begin{cases} 0, & t < 16; \\ 1, & t \geq 16. \end{cases}$$

$$x_{2,t} = \begin{cases} 0, & t < 38; \\ 1, & t \geq 38. \end{cases}$$

$$x_{3,t} = \begin{cases} 0, & t \neq 41; \\ 1, & t = 41. \end{cases}$$

$$x_{4,t} = \begin{cases} 0, & t \neq 61; \\ 1, & t = 61. \end{cases}$$

Na Tabela 4.17 são apresentadas as estimativas dos parâmetros dos mode-

los propostos.

TABELA 4.17: Estimativas dos parâmetros dos modelos ARIMA(2,1,0), ARIMA(0,1,1) e ARIMA(2,1,0) com intervenções para a série de produção mensal de gasolina, em metros cúbicos, no período de janeiro de 2000 a abril de 2006.

Modelo	Parâmetro	Estimativa	Erro padrão	p-valor
ARIMA (2, 1, 0)	ϕ_1	-0,764224	0,104477	0,000000
	ϕ_2	-0,467303	0,105209	0,000031
ARIMA (0, 1, 1)	θ_1	0,703180	0,064540	0,000000
	ϕ_1	-0,846400	0,104890	0,000000
ARIMA (2, 1, 0)-CI	ϕ_2	-0,551900	0,104730	0,000002
	ω_1	160300	64361,27	0,015218
	ω_2	162000	64234,60	0,013922
	ω_3	159100	68419,01	0,023054
	δ_3	0,736820	0,224820	0,001663
	ω_4	170500	73196,30	0,022865
	δ_4	-0,616900	0,186130	0,001484

Na Tabela 4.18 são exibidas as estatísticas para o teste de Box-Pierce, o qual identifica se o modelo é ruído branco (independente e identicamente distribuído).

Pela análise da Tabela 4.18, os resíduos constituem um ruído branco, já que $Q_{24} < \chi^2$. Os valores reais, as estimativas dos valores preditos para a produção de gasolina e o erro de previsão, para o período de maio de 2006 a setembro de 2006, encontram-se na Tabela 4.19.

Na Tabela 4.20 são apresentados os resultados do critério de Akaike (AIC), erro quadrático médio de previsão (EQMP) e MAPE.

TABELA 4.18: Estatísticas para o teste de Box-Pierce dos modelos ARIMA com e sem intervenção.

Modelo	GL	Q_{24}	χ^2
ARIMA (2, 1, 0)	22	28,19	33,92
ARIMA (0, 1, 1)	23	28,19	35,17
ARIMA (2, 1, 0) – CI	18	17,42	28,87

TABELA 4.19: Valores reais de produção de gasolina, para o período de maio a setembro de 2006 e valores preditos para o mesmo período.

Modelo	Meses	Z_{t+h}	$\hat{Z}_t(h)$	$e_t(h)$
ARIMA (2, 1, 0)	mai/06	1682265	1654508	27757
	jun/06	1710233	1698715	11518
	jul/06	1739056	1681631	57425
	ago/06	1701896	1674029	27867
	set/06	1631991	1687822	-55831
ARIMA (0, 1, 1)	mai/06	1682265	1671054	11211
	jun/06	1710233	1671054	39179
	jul/06	1739056	1671054	68002
	ago/06	1701896	1671054	30842
	set/06	1631991	1671054	-39063
ARIMA (2, 1, 0) – CI	mai/06	1682265	1646222	36043
	jun/06	1710233	1703305	6928
	jul/06	1739056	1679362	59694
	ago/06	1701896	1668075	33821
	set/06	1631991	1690872	-58881

Nota: Z_{t+h} = valor real; $\hat{Z}_t(h)$ = valor predito; $e_t(h)$ = erro de previsão

O modelo escolhido para representar a série de produção mensal de gasolina, em metros cúbicos, é o ARIMA (2, 1, 0) sem intervenção. Este modelo, apesar de não possuir o menor AIC, possui o menor EQMP e este critério é o que deve ser levado em consideração, já que o objetivo é a previsão. O modelo apresentou um bom desempenho de ajuste.

TABELA 4.20: Estimativas dos critérios de AIC, EQMP e MAPE dos modelos ARIMA com e sem intervenção.

Modelo	AIC	EQMP	MAPE
ARIMA (2, 1, 0)	1735,96	$1,60 \times 10^9$	2,14%
ARIMA (0, 1, 1)	1736,17	$1,70 \times 10^9$	2,21%
ARIMA (2, 1, 0) – CI	1719,16	$1,90 \times 10^9$	2,32%

O modelo substituído por suas respectivas estimativas é dado por:

$$Z_t = \frac{a_t}{(1 + 0,764224B + 0,467303B^2)(1 - B)}. \quad (4.1)$$

4.2 Modelo multivariado

Para o ajuste de um modelo VAR, as séries devem estar no seu estágio estacionário, isto é, sem presença de tendência e sazonalidade. Conforme realizado na subseção anterior, no ajuste dos modelos univariados, as séries de óleo diesel e gasolina apresentam tendência e a série de álcool possui sazonalidade. As séries tomadas, então, são aquelas com uma diferença, de óleo diesel e gasolina. A série de álcool possui sazonalidade de 12 meses. Fazer diferenças nesta série faria com que ela perdesse 12 observações, não sendo possível o ajuste do VAR, que deve contar com o mesmo número de observações. Assim, foi feito o ajuste da sazonalidade determinística pelo método de regressão comentado na subseção 2.2.6.2. A série sazonalmente ajustada encontra-se na Tabela 6A dos anexos. Como as séries de gasolina e óleo diesel perderam uma observação, a série de álcool foi tomada desprezando-se sua primeira observação. O ajuste conta, dessa forma, com 80 observações.

O vetor de variáveis Z_t , constituído pelas séries de produção de álcool,

óleo diesel e gasolina, representado respectivamente por A_t , O_t e G_t , é dado por:

$$\mathbf{Z}_t = \begin{bmatrix} A_t \\ O_t \\ G_t \end{bmatrix}.$$

Na Tabela 4.21 são apresentados os resultados dos ajustes dos modelos auto-regressivos até a ordem 8.

TABELA 4.21: Estatísticas resultantes de ajustes de modelos VAR(p), $p=1,\dots,8$, para as séries de produção de álcool, óleo diesel e gasolina .

Ordem	1	2	3	4	5	6	7	8
AIC	61,79	61,44	61,48	61,47	61,57	61,69	61,87	62,02
BIC	62,08	62,01	62,33	62,61	62,99	63,40	63,86	64,30

Usando-se os valores de AIC e BIC dados pela Tabela 4.21, a ordem selecionada foi $p = 2$, por apresentar menor valor de AIC e BIC.

A equação do modelo VAR(2) pode ser escrita como:

$$\mathbf{Z}_t = \Phi_0 + \Phi_1 \mathbf{Z}_{t-1} + \Phi_2 \mathbf{Z}_{t-2} + \mathbf{a}_t,$$

em que \mathbf{Z}_{t-j} é o vetor das variáveis defasadas de ordem j , $j = 1, 2$ e

$$\mathbf{a}_t = \begin{bmatrix} \epsilon_{At} \\ \epsilon_{Ot} \\ \epsilon_{Gt} \end{bmatrix}$$

é o vetor de resíduos. $\mathbf{a}_t \sim \text{RB}(\mathbf{0}, \Sigma)$.

As estimativas dos parâmetros e seus erros padrões são dados na Tabela 4.22.

Dessa forma, o modelo ajustado é dado por:

$$A_t = 0,739561A_{t-1} - 0,189697A_{t-2} + \epsilon_{At},$$

$$O_t = -0,389558O_{t-1} - 0,754898G_{t-1} - 0,479197G_{t-2} + \epsilon_{Ot},$$

$$G_t = 0,105652O_{t-2} - 0,697854G_{t-1} - 0,590279G_{t-2} + \epsilon_{Gt}.$$

A estatística multivariada de Box-Ljung-Pierce, para verificar se o resíduo é ruído branco, resultou em: $Q(36)=243,7860$ e $\chi^2(306)=347,79$. Como $Q(36) < \chi^2(306)$, o ajuste é adequado.

Na Figura 4.21 é apresentada a função de autocorrelação para os resíduos, confirmando-se a adequação do modelo.

Por estas estimativas, conclui-se que os valores atuais da produção de álcool não são influenciados pelos valores defasados de nenhuma das três séries, ou seja, álcool, óleo diesel e gasolina. Esta conclusão estatística contraria os fatos históricos e econômicos referentes ao mercado de combustíveis. Esperava-se uma dependência do álcool em relação à gasolina e ao óleo diesel porque o uso do álcool como combustível reduz a participação da gasolina na estrutura de refino do petróleo e isso amplia a produção de óleo diesel.

Os valores presentes de produção de óleo diesel são influenciados por seus valores passados e também pelos valores passados da produção de gasolina.

Os valores atuais da gasolina são influenciados por seus valores passados e também pelos valores passados de óleo diesel.

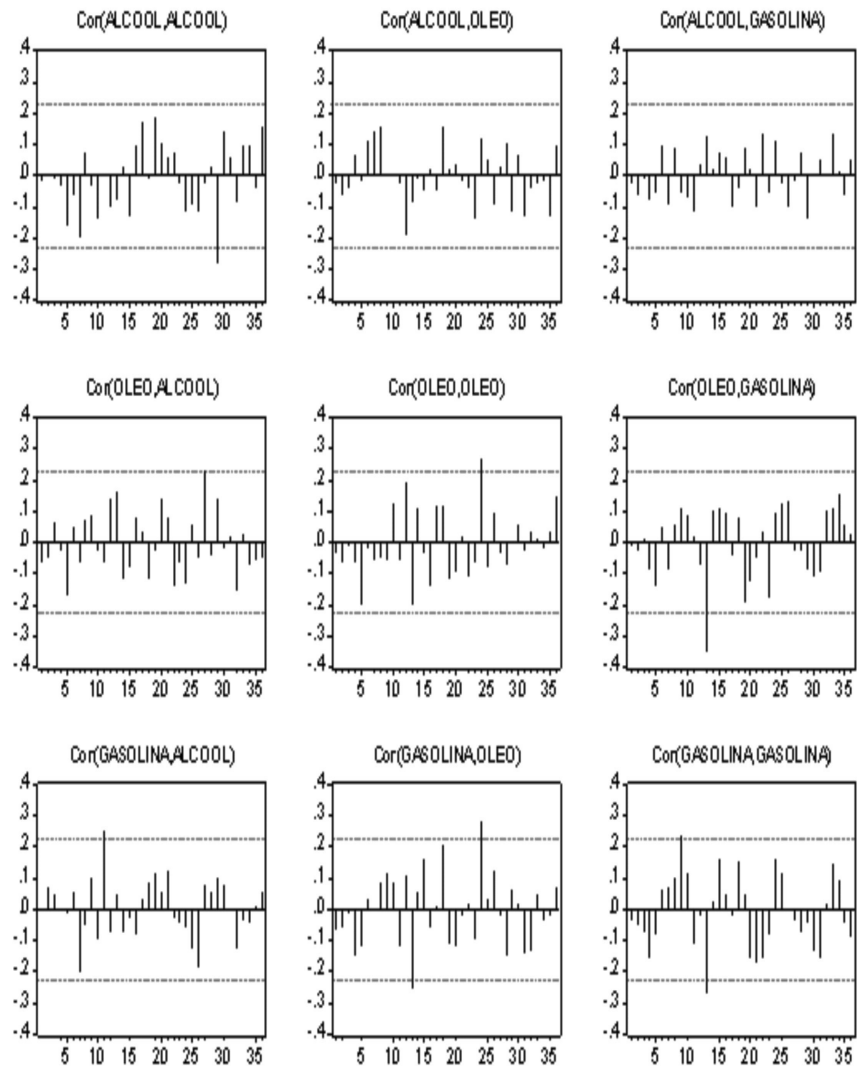


FIGURA 4.21: Representação gráfica da função de autocorrelação dos resíduos.

TABELA 4.22: Ajuste de um modelo VAR(2) para as séries de álcool, óleo diesel e gasolina.

Parâmetros	Estimativa			Erro padrão		
	Φ_1	0,739561*	-0,000008	0,000011	0,11358	0,000009
	1375,86	-0,389558*	-0,754898*	1489,01	0,118350	0,229590
	-190,066	-0,075114	-0,697854*	670,396	0,053290	0,103370
Φ_2	-0,189697*	0,000017	0,000019	0,11372	0,000008	0,000017
	-2173,03	-0,048772	-0,479197*	1490,84	0,113900	0,227870
	775,582	-0,105652*	-0,590279*	671,217	0,051280	0,102590

* estimativas significativas a 5%.

5 CONCLUSÕES

De modo geral, os modelos univariados, propostos por Box e Jenkins, ajustaram-se bem aos dados.

Para a série de produção de cana-de-açúcar, o modelo que melhor se ajustou aos dados foi o ARIMA (2,1,1) com intervenção no ano de 1983. Neste ano, observou-se um aumento na produção de cana em razão da alta demanda desta cultura para a produção de álcool combustível.

A série de produção de álcool apresentou correlações sazonais, sendo o SARIMA (1, 0, 0)(0, 1, 1)₁₂ com intervenção em maio de 1996, junho de 1997 e dezembro de 2000, o modelo de melhor ajuste. O efeito no aumento da produção refletido em maio de 1996, explica-se pela fase de redefinição do Proálcool. Nesta época, houve a instituição da chamada "frota verde", que foi o estímulo e a determinação do uso do álcool hidratado em determinadas classes de veículos leves, como os carros oficiais e táxis. Houve também o aumento de 22% para 24% do percentual de adição de álcool etílico anidro à gasolina. Os efeitos refletidos, respectivamente, em junho de 1997 e dezembro de 2000, quando ocorreram quedas na produção, são explicados em parte pela diminuição do consumo de álcool hidratado, cujo preço encontrava-se acima de 70% do preço da gasolina e também porque houve um aumento no mercado internacional nos preços do açúcar, deslocando a produção de cana para este setor.

As séries de produção de petróleo e óleo diesel apresentaram correlações sazonais, sendo o SARIMA(0, 1, 1)(1, 0, 0)₁₂ o melhor ajuste para o petróleo e o SARIMA(1, 1, 1)(1, 0, 0)₁₂ o melhor ajuste para o óleo diesel.

Para a série de produção de gasolina, o melhor ajuste foi fornecido pelo ARIMA (2,1,0) sem intervenções.

O modelo multivariado que se ajustou melhor à série de combustíveis foi o VAR(2). Para a série de produção de álcool, não foram observadas relações de dependência em relação às séries de produção de óleo diesel e gasolina. Para as séries de produção de óleo diesel e gasolina, observou-se uma relação de "feedback", isto é, os valores atuais do diesel, além de dependerem dos seus próprios valores defasados, dependem também dos valores passados da gasolina. Para os valores atuais da gasolina, observou-se o mesmo comportamento.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. Maximum likelihood identification of gaussian autoregressive moving average models. **Biometrika**, London, v.60, n.21, p. 255-265, 1973.
- AMBIENTE BRASIL. Disponível em
<www.ambientebrasil.com.br/energia/biomassa> Acesso em: 16 dez. 2005.
- ANEEL: **Atlas de energia elétrica do Brasil**. Brasília: Agência Nacional de Energia Elétrica. 2002.
- ANP. **Agência Nacional do Petróleo**. Disponível em
<www.anp.gov.br/doc/dados_estatisticos> Acesso em: 14 set. 2006.
- ASSIS, A. N.; LOPES, L. B. R. A ineficiência da política de preços para conter o consumo de derivados de petróleo. **Revista Brasileira de Economia**, Rio de Janeiro, v. 34, n.3, p.417-428, jul./set.1980.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis: forecasting and control**. 3 ed. New Jersey: Prentice Hall, 1994. 598 p.
- BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: forecasting and control**. San Francisco: Holden Day, 1976.
- BRASIL. Ministério de Minas e Energia. **Balanco Energético Nacional**. Brasília: 2002. 200 p.
- BROWN, R. I. Um esquema para avaliar os impactos de estratégias diversas para etanol, outras substituições e a racionalização na demanda de derivados de petróleo em 1985. **Ciência e Cultura**, Campinas, v. 32, n. 8, p. 1032-1040, ago. 1980.
- BURNQUIST, H. L.; BACCHI, M. R. P. A demanda por gasolina no Brasil: uma análise utilizando técnicas de co-integração. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL, 40., 002, Passo Fundo, RS. Equidade e Eficiência na Agricultura Brasileira, 2002. **Ciência e Cultura**, Campinas, v. 32, n. 8, p. 1032-1040, ago.1980.
- CANA-DE-AÇÚCAR NO BRASIL. Disponível em
<www.infoener.iee.usp.br/scripts/biomassa/br_cana.asp>. Acesso em: 16 dez. 2005.

- ENERGIA BRASIL. Disponível em <www.energiabrasil.gov.br/fontesenergia_biomassa.asp>. Acesso em: 18 maio. 2006.
- GOLDEMBERG, J. **Energia no Brasil**. Rio de Janeiro: Livros Técnicos e Científicos. 1979. 171 p.
- Instituto de Pesquisas Econômicas Aplicadas - IPEA.
Disponível em <www.ipeadata.gov.br/ipeaweb.dll/ipeadata?406640843>. Acesso em: 10 dez. 2005.
- LEITE, A. D. **A energia do Brasil**. Rio de Janeiro: Nova Fronteira, 1997. 598 p.
- MACIEL, T. J. L. **Análise do setor alcooleiro do Brasil: perspectivas do Proálcool**. 1991. 83 p. Tese (Doutorado em Economia Rural) - Universidade Federal de Viçosa, Viçosa, MG.
- MIRAGAYA, J. C. G. Biodiesel: tendências no mundo e no Brasil. **Informe Agropecuário**, Belo Horizonte, v. 26, n. 229, p. 7-13, 2005.
- MOREIRA, R. B. **Modelos para a projeção do consumo nacional e regional de óleo diesel**. Rio de Janeiro: IPEA, 1996. (Texto para Discussão n. 443).
- MORETTIN, P. A. **Econometria financeira - um curso em séries temporais financeiras**. São Paulo: Copyright, 2006. 341 p.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: Edgard Blücher, 2004. 439 p.
- OPEC. **Annual statistical bulletin 2002**. Disponível em <www.opec.org>. Acesso em: 05 jan. 2007.
- PIERCE, D. A. Some recent developments in seasonal adjustment. In: BRILLINGER, D. R.; TIAO, G. C. (Ed.). **Reports on Directions in Time Series**. Institute of Mathematical Statistics, 1979. p.123-140.
- PRIESTLEY, M. B. **Spectral Analysis And Time Series**. London: Academic Press, 1989. 407p.
- SILVA, C. R. L. **A crise Energética e o proálcool: algumas considerações**. São Paulo: Secretaria da Agricultura/IEA, 1983 (Relatório de pesquisa).
- UNICA. **União da Agroindústria Canavieira de São Paulo**. Disponível em <www.unica.com.br>. Acesso em: 25 nov. 2005.

ANEXOS

ANEXO	Páginas
TABELA 1A	Série de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998. 101
TABELA 2A	Série de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a setembro de 2006 102
TABELA 3A	Série de produção mensal de petróleo, em m ³ , no período de janeiro de 2000 a setembro de 2006 103
TABELA 4A	Série de produção mensal de óleo diesel, em m ³ , no período de janeiro de 2000 a setembro de 2006 104
TABELA 5A	Série de produção mensal de gasolina, em m ³ , no período de janeiro de 2000 a setembro de 2006 105
TABELA 6A	Série de produção mensal de álcool sem sazonalidade - método de regressão. 106

TABELA 1.A: Série de produção anual de cana-de-açúcar, em milhões de toneladas, no período de 1947 a 1998.

Anos	Observações	Anos	Observações	Anos	Observações	Anos	Observações	Anos	Observações
1947	28.989,00	1959	53.512,00	1971	80.380,00	1983	216.036,00	1995	303.699,00
1948	30.892,00	1960	56.926,00	1972	85.106,00	1984	222.317,00	1996	317.105,00
1949	30.928,00	1961	59.377,00	1973	91.994,00	1985	247.199,00	1997	331.612,00
1950	32.670,00	1962	62.534,00	1974	95.623,00	1986	239.178,00	1998	345.254,00
1951	33.652,00	1963	63.722,00	1975	91.524,00	1987	268.741,00	1999	333.847,00
1952	36.041,00	1964	66.398,00	1976	103.173,00	1988	258.412,00	2000	326.121,00
1953	38.336,00	1965	75.852,00	1977	120.081,00	1989	252.642,00	2001	344.292,00
1954	40.301,00	1966	75.787,00	1978	129.144,00	1990	262.674,00	2002	364.389,00
1955	40.946,00	1967	77.086,00	1979	138.898,00	1991	260.887,00	2003	396.012,00
1956	43.975,00	1968	76.610,00	1980	148.650,00	1992	271.474,00	2004	415.205,00
1957	47.703,00	1969	75.247,00	1981	155.924,00	1993	244.530,00		
1958	50.020,00	1970	79.752,00	1982	186.646,00	1994	292.101,00		

TABELA 2.A: Série de produção mensal de álcool nível 80 - índice (média 2002=100), no período de janeiro de 1991 a setembro de 2006.

Anos	Meses											
	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
1991	46,19	31,19	19,06	2,38	101,98	209,45	257,16	269,92	268,54	241,34	142,3	62,21
1992	38,55	17,63	7,02	3,36	55,46	200,23	257,9	262,06	208,06	230,37	174,9	80,6
1993	36,61	19,12	2,66	3,72	109,98	201,54	249,53	255,99	221,44	216,55	129,34	43,4
1994	20,1	12,76	6,92	4,93	123,25	221,63	257,34	268,4	260,94	220,45	120,15	53,17
1995	34,25	22,49	11,04	2,17	69,47	205,64	239,72	268,87	266,5	216,9	154,13	76,75
1996	32,87	16,09	7,27	17,81	148,26	236,53	267,59	278,59	240,49	237,82	170,97	81,62
1997	38,71	25,97	14,88	31,25	163,45	181,01	250,22	269,78	261,58	251,85	174,22	86,85
1998	34,48	19,85	7,36	25,29	148,85	194,13	220,34	211,63	208,36	198,37	193,8	97,26
1999	34,71	16,93	10,82	47,62	165,34	172,48	206,26	221,08	206,2	196,46	116,53	43,03
2000	21,22	14,69	13,03	9,97	86,74	187,79	205,29	206,77	169,92	162,75	70,8	35,36
2001	26,14	18,5	11,38	10,71	69,17	161,47	185,02	195,4	178,54	170,8	117,31	44,72
2002	16,82	12,65	12,1	41,84	137,6	169,78	177,39	179,17	164,69	163,75	84,81	39,41
2003	25,1	11,73	5,39	66,34	135,99	171,06	188,07	190,39	196,01	190,29	100,48	40,9
2004	20,26	15,12	12,3	28,24	95,75	145,14	168,02	201,88	217,5	197,06	180,56	97,76
2005	34,78	19,91	9,08	54,16	155,76	173,51	189,98	211,03	191,28	194,6	135,75	52,12
2006	24,78	9,43	7,48	65,66	169,60	197,85	213,14	223,15	212,04			

TABELA 3.A: Série de produção mensal de petróleo, em m³, no período de janeiro de 2000 a setembro de 2006.

Meses	Anos						
	2000	2001	2002	2003	2004	2005	2006
jan	5.690.793,00	6.498.405,00	7.106.895,00	7.409.792,00	7.182.535,00	7.382.254,00	8.321.362,00
fev	5.167.310,00	5.971.240,00	6.389.095,00	6.872.130,00	6.694.564,00	6.644.407,00	7.533.460,00
mar	5.861.504,00	6.211.220,00	7.179.072,00	7.453.190,00	7.294.742,00	7.581.902,00	8.358.067,00
abr	5.538.707,00	6.052.810,00	7.007.921,00	7.250.214,00	6.911.528,00	7.995.070,00	8.286.098,00
mai	5.713.178,00	5.920.527,00	7.305.933,00	7.332.231,00	7.001.749,00	8.324.684,00	8.615.745,00
jun	5.764.513,00	6.173.469,00	7.166.751,00	6.545.205,00	7.066.628,00	8.100.731,00	7.776.547,00
jul	5.851.190,00	6.428.124,00	7.164.729,00	7.349.096,00	7.458.958,00	8.282.310,00	8.502.763,00
ago	5.876.616,00	6.338.025,00	7.413.679,00	7.595.173,00	7.403.277,00	8.013.086,00	8.393.094,00
set	6.196.036,00	6.223.285,00	7.085.702,00	7.261.756,00	7.264.852,00	7.950.711,00	8.265.597,00
out	6.476.597,00	5.806.609,00	7.286.273,00	7.390.239,00	7.396.430,00	8.233.697,00	
nov	6.513.555,00	6.384.420,00	6.676.713,00	7.062.968,00	6.927.263,00	7.985.490,00	
dez	6.993.695,00	7.011.827,00	6.616.203,00	7.297.703,00	7.364.453,00	8.302.393,00	

TABELA 4.A: Série de produção mensal de óleo diesel, em m³, no período de janeiro de 2000 a setembro de 2006.

Meses	Anos						
	2000	2001	2002	2003	2004	2005	2006
jan	2.544.176,00	2.521.585,00	2.628.029,00	2.621.111,00	3.105.798,00	3.153.877,00	3.212.927,00
fev	2.219.266,00	2.333.298,00	2.527.433,00	2.341.505,00	3.067.108,00	2.893.174,00	2.902.481,00
mar	2.591.153,00	2.593.064,00	2.931.149,00	3.058.670,00	3.401.493,00	3.499.249,00	3.437.810,00
abr	2.558.680,00	2.457.261,00	2.799.012,00	2.857.146,00	3.441.698,00	2.956.879,00	3.231.419,00
mai	2.617.471,00	2.655.077,00	2.759.299,00	2.867.457,00	3.287.997,00	3.147.151,00	3.450.654,00
jun	2.614.114,00	2.678.424,00	2.468.991,00	2.741.256,00	2.990.713,00	3.049.850,00	3.245.143,00
jul	2.720.483,00	2.891.551,00	2.786.211,00	2.937.182,00	2.912.456,00	3.340.298,00	3.240.426,00
ago	2.690.493,00	3.181.726,00	2.890.831,00	2.882.735,00	2.926.464,00	3.398.776,00	3.170.811,00
set	2.434.005,00	3.046.258,00	2.780.784,00	3.011.350,00	3.249.463,00	3.283.342,00	3.142.760,00
out	2.759.506,00	3.143.436,00	3.005.089,00	3.150.930,00	3.525.443,00	3.348.417,00	
nov	2.440.754,00	2.931.220,00	2.902.754,00	2.866.691,00	2.984.513,00	3.135.813,00	
dez	2.589.951,00	2.644.980,00	2.511.018,00	2.816.829,00	3.359.120,00	3.189.602,00	

TABELA 5.A: Série de produção mensal de gasolina, em m³, no período de janeiro de 2000 a setembro de 2006.

Meses	Anos						
	2000	2001	2002	2003	2004	2005	2006
jan	1.332.710,00	1.539.926,00	1.680.375,00	1.555.879,00	1.533.658,00	1.697.326,00	1.753.060,00
fev	1.488.925,00	1.560.129,00	1.534.569,00	1.312.381,00	1.327.733,00	1.373.259,00	1.590.795,00
mar	1.561.730,00	1.550.977,00	1.574.775,00	1.477.798,00	1.568.699,00	1.530.198,00	1.726.402,00
abr	1.453.783,00	1.630.224,00	1.595.397,00	1.413.374,00	1.464.942,00	1.561.089,00	1.690.245,00
mai	1.545.238,00	1.809.625,00	1.617.909,00	1.545.413,00	1.345.008,00	1.669.162,00	1.682.265,00
jun	1.546.508,00	1.681.133,00	1.461.294,00	1.568.538,00	1.362.901,00	1.435.033,00	1.710.233,00
jul	1.636.440,00	1.517.270,00	1.610.749,00	1.493.161,00	1.492.318,00	1.656.461,00	1.739.056,00
ago	1.523.980,00	1.575.922,00	1.573.232,00	1.458.990,00	1.349.238,00	1.710.499,00	1.701.896,00
set	1.600.458,00	1.702.046,00	1.428.677,00	1.541.412,00	1.489.388,00	1.565.061,00	1.631.991,00
out	1.737.382,00	1.561.429,00	1.558.176,00	1.492.176,00	1.575.751,00	1.743.541,00	
nov	1.438.743,00	1.545.443,00	1.552.423,00	1.395.085,00	1.467.579,00	1.534.246,00	
dez	1.710.466,00	1.632.647,00	1.581.342,00	1.400.713,00	1.600.593,00	1.626.460,00	

TABELA 6.A: Série de produção mensal de álcool sem sazonalidade - método de regressão.

Meses	Anos									
	2000	2001	2002	2003	2004	2005	2006			
jan	-2,94	1,98	-7,34	0,94	-3,90	10,62	0,62			
fev	0,11	3,92	-1,93	-2,85	0,54	5,33	-5,15			
mar	2,92	1,27	1,99	-4,72	2,19	-1,03	-2,63			
abr	-29,59	-28,85	2,28	26,78	-11,32	14,60	26,10			
mai	-34,78	-52,35	16,08	14,47	-25,77	34,24	48,08			
jun	15,42	-10,90	-2,59	-1,31	-27,23	1,14	25,48			
jul	15,73	-4,54	-12,17	-1,49	-21,54	0,42	23,58			
ago	5,66	-5,71	-21,94	-10,72	0,77	9,92	22,04			
set	-20,08	-11,46	-25,31	6,01	27,50	1,28	22,04			
out	-17,12	-9,07	-16,12	10,42	17,19	14,73	-			
nov	-44,15	2,36	-30,14	-14,47	65,61	20,80	-			
dez	-16,35	-6,99	-12,30	-10,81	46,05	0,41	-			