

46449

131904FN

MARCELO INÁCIO FERREIRA FERRAZ

**USO DE MODELOS DE SÉRIES TEMPORAIS NA PREVISÃO DA
SÉRIE DE PRECIPITAÇÕES PLUVIAIS MENS AIS
NO MUNICÍPIO DE LAVRAS - MG**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

Orientador
Prof.^a Dr.^a THELMA SÁFADI

LAVRAS
MINAS GERAIS - BRASIL
1999

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

Ferraz, Marcelo Inácio Ferreira

Uso de modelos de séries temporais na previsão da séries de precipitações
pluviais mensais no município de Lavras – MG. / Marcelo Inácio Ferreira Ferraz.
– Lavras : UFLA, 1999.

97 p. : il.

Orientador: Thelma Sáfadi.

Dissertação (Mestrado) – UFLA.

Bibliografia.

1. Tendência. 2. Sazonalidade. 3. Serie temporal. 4. Ruído branco. 5. Modelo de
regressão. 6. Modelo SARINA. I. Universidade Federal de Lavras. II. Título.

CDD-311.2

-551.577

630.724

MARCELO INÁCIO FERREIRA FERRAZ

**USO DE MODELOS DE SÉRIES TEMPORAIS NA PREVISÃO DE
SÉRIE DE PRECIPITAÇÕES PLUVIAIS MENSAIS
NO MUNICÍPIO DE LAVRAS - MG**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

APROVADA em 11 de fevereiro de 1999.

Prof. Dr. Augusto Ramalho de Moraes UFLA

Prof. Dr. Eduardo Bearzoti UFLA


Prof.^a. Dr.^a. Thelma Sáfyadi
UFLA
(Orientadora)

LAVRAS
MINAS GERAIS - BRASIL

1999 - UFLA

*À minha noiva Diokleiany que muito
me ajudou nessa caminhada*

*Aos meus pais Elpidio e Magaly pela
formação moral e intelectual que permitiram
meu desenvolvimento pessoal e profissional.*

DEDICO

AGRADECIMENTOS

À Deus pela saúde concedida.

À Universidade Federal de Lavras (UFLA), pela oportunidade de realização do curso.

À Coordenadoria de Aperfeiçoamento de Pessoal de Ensino Superior /CAPES.

Aos professores e funcionários pelo incentivo.

Aos professores membros da banca examinadora, Thelma, Eduardo e Augusto pelas sugestões.

Aos professores do curso, que além de transmitirem os conhecimentos, se tornaram grandes amigos.

A minha orientadora, Thelma, pela oportunidade de aprender um pouco com sua vasta experiência, pela paciência, pela troca de idéias, pela relação amiga com que me permitiu desenvolver esse trabalho.

A todos os colegas do curso, Janaína, Dileta, Hugo e Telde pelo convívio e amizade.

SUMÁRIO

	Página
RESUMO.....	i
ABSTRACT.....	ii
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	3
2.1 SÉRIES TEMPORAIS.....	3
2.1.1 Considerações Gerais.....	3
2.1.2 Decomposição dos Elementos de uma Série.....	6
2.1.3 A Estimação do Componente Tendência.....	8
2.1.4 A Estimação do Componente Sazonal.....	10
2.1.5 Funções de Autocorrelação.....	13
2.2 MODELOS DE BOX & JENKINS.....	16
2.2.1 Considerações Gerais.....	16
2.2.2 Modelos Auto-regressivos - AR.....	21
2.2.3 Modelos de Médias Móveis - MA.....	24
2.2.4 Modelos Mistos - ARMA.....	26
2.2.5 Função de Autocorrelação Parcial.....	28
2.2.6 Modelo Auto-Regressivo Integrado - ARIMA.....	32
2.2.7 Modelos ARIMA Sazonal - SARIMA.....	38
2.2.7.1 Sazonalidade Estacástica.....	38
2.2.7.2 Sazonalidade Determinística.....	39
2.3 PREVISÃO.....	41
2.3.1 Considerações Preliminares.....	41
2.3.2 Previsão com Modelos Suavizados.....	41
2.3.2.1 Alisamento Exponencial Simples (AES).....	42
2.3.2.2 Alisamento Exponencial Biparamétrico de Holt (AEH).....	44
2.3.2.3 Alisamento Exponencial Sazonal de Holt e Winters (HW).....	45
2.3.3 Previsão com Modelo ARIMA.....	47
2.3.3.1 Representação e Notação.....	47
2.3.3.2 Variância da Previsão e Intervalo de Confiança.....	48
2.3.3.3 Cálculo e Atualização das Previsões.....	50
2.3.4 Previsão com Modelos SARIMA.....	51
2.3.5 Previsão com Modelos ARMA.....	52

2.4 PERIODOGRAMA (REPRESENTAÇÃO ESPECTRAL)	55
2.5 APLICAÇÕES	58
2.6 OUTRAS METODOLOGIAS	61
3 MATERIAL E MÉTODOS.....	63
3.1 MATERIAL	63
3.2 MÉTODOS	63
4 RESULTADOS E DISCUSSÃO	66
4.1 FAMILIARIZAÇÃO	66
4.2 AJUSTE DE MODELOS	70
4.2.1 Modelo de Regressão	70
4.2.2 Modelo SARIMA	76
4.3 PREVISÃO	82
5 CONCLUSÕES	86
6 BIBLIOGRAFIA.....	87
ANEXOS	91

RESUMO

FERRAZ, Marcelo Inácio Ferreira. **Uso de modelos de séries temporais na previsão da série de precipitação pluviais mensais no município de Lavras - MG.** Lavras: UFLA, 1999. 97p. (Dissertação - Mestrado em Estatística e Experimentação Agropecuária)

O presente trabalho foi dirigido no sentido de apresentar um estudo detalhado sobre a aplicação de modelos de séries temporais na análise de dados pluviométricos para a região de Lavras, MG. Foram utilizados dados de precipitações pluviais diárias expressas em altura de lâmina d'água (mm), abrangendo o período janeiro de 1966 a dezembro de 1997, as observações foram agrupados em registros mensais num total de 32 anos. Considerou-se o período de janeiro de 1966 a dezembro de 1996 para ajustar os modelos e as observações do período de janeiro de 1997 a dezembro de 1997 foram utilizadas para verificação da previsão. Para o ajuste dos modelos inicialmente foi feita a familiarização da série através da construção de gráficos, aplicação de testes específicos e, a construção do periodograma da série. Foram ajustados o modelo de regressão e o modelo SARIMA determinístico. Para verificação da adequabilidade dos modelos foi realizado a análise dos resíduos através das funções de autocorrelações e autocorrelações parciais. Aos resíduos foram ajustados modelos de Box e Jenkins para obtenção do ruído branco. Foram feitas previsões com o modelo de Alisamento Exponencial de Holt e Winters e com os modelos de regressão e SARIMA. O desempenho dos modelos de previsão foi verificado através da análise gráfica e da soma dos erros de previsão. As principais conclusões foram: a série não apresenta tendência com uma periodicidade de ordem 12 que é correspondente a uma sazonalidade de ordem 12; o ruído branco foi obtido após o ajuste do modelos misto ARMA (4, 4) ao resíduo dos modelos de regressão e SARIMA determinístico; o modelo SARIMA foi o que forneceu as melhores previsões vindo em seguida o modelo de regressão e por último, o modelo de alisamento sazonal aditivo de Holt e Winters.

Orientadora: Thelma Sáfydi – UFLA

ABSTRACT

FERRAZ, M. I. F. *Use of time series models in forecasting the series of monthly rainfall in the city of Lavras – State of Minas Gerais.* Lavras: UFLA; 1999. 97p. (Dissertation – Master in Statistics and Agricultural Experimentation)

The present work was run in the sense of presenting a detailed study upon the application of time series models in the rainfall data analysis for the region of Lavras, MG. Data of daily rainfalls expressed in water sheet height (mm), encompassing the period of January, 1966 to December 1997, the observations were grouped into monthly records in a total of 32 years. The period of January 1966 to December 1996 was considered to fit the models and the observations of the period of January, 1997 to December, 1997 were utilized for verifying the forecast. To the fitting of the models, at first a familiarization of the series was done through the construction of plots, application of particular tests and construction of the periodogram of the series. Both the regression model and SARIMA deterministic model were adjusted. To verify the suitability of the models the residual analysis through the functions of autocorrelations and partial autocorrelations was performed. To the residues were adjusted Box and Jenkins models to obtain the white noise. Forecasts with the Holt Winters Exponential Smoothing model and with the regression and SARIMA models. The performed of forecast models was verified through the graphical plot and forecast errors sums. The chief conclusions were: the series does not present trend, there is a periodicity of order 12 which is correspondent to a seasonality of order 12, the white noise was obtained after the fitting of the ARMA mixed model (4, 4) to the residue of the regression models and deterministic SARIMA, the SARIMA model was the one that one yielded the best forecastings coming next the regression model and last Holt and Winters additive seasonal smoothing.

Major Professor: Thelma Sáfaci - UFLA

1 INTRODUÇÃO

O sucesso ou fracasso de um empreendimento, quer em uma atividade agrícola ou não, pode estar intimamente ligado às características do regime de chuvas. A importância da climatologia nas diversas áreas da agronomia é fator inegável por exercer influência no fator ambiental, ou seja, em variáveis como: balanço de água, temperatura e umidade do ar, que, em conjunto, atuam como fatores preponderantes para o sucesso da produção agrícola.

A distribuição e o comportamento das precipitações pluviais são fatores básicos para o planejamento das atividades agrícolas, tais como: definição de datas de plantio e preparo de solo, dimensionamento de vertedouros de redes pluviais, erosão, enchentes e elaboração de projetos de irrigação suplementar. Tendo em conta esses aspectos, pode-se dizer que o volume de chuva que cai em uma localidade é fator determinante da atividade agrícola a ser desenvolvida.

O município de Lavras e região circunvizinha caracterizam-se por possuir uma intensa atividade agropecuária. Nesse sentido, o estudo das precipitações na região vem também merecendo esforços de alguns pesquisadores, como Botelho (1998), Castro Neto e Silveira (1981), que utilizaram a distribuição gama para estudar a probabilidade da precipitação na região. Com o intuito de contribuir para a caracterização do regime de chuvas e auxiliar no planejamento da atividade agropecuária, objetivou-se, aqui, abordar a série de precipitação através dos modelos de séries temporais.

Dessa forma, este trabalho focalizou a análise de séries temporais no domínio do tempo e teve como objetivos:

- a) obter modelos que melhor se ajustem à série climatológica de precipitações pluviiais mensais no município de Lavras, identificando seus componentes;
- b) caracterizar o regime mensal de chuvas no município de Lavras;
- c) propor modelos de previsão para o regime mensal de chuvas.

2 REFERENCIAL TEÓRICO

2.1 Séries Temporais

2.1.1 Considerações Gerais

A partir da possibilidade de mensuração do tempo, pode-se estabelecer algumas relações entre a passagem do tempo e a ocorrência de fenômenos. Na observação de um fenômeno na natureza, pode-se perceber que este é estimulado por uma infinidade de outros fenômenos, com diferentes graus de interferência. Uma das variáveis que podem estar interferindo em um fenômeno é o transcurso do tempo. Dessa forma, o tempo pode ser utilizado como uma variável auxiliar para o seu entendimento.

A possível correlação introduzida pela amostragem de observações em instantes de tempo adjacentes pode severamente restringir a aplicabilidade de alguns métodos estatísticos convencionais, os quais dependem da suposição que observações adjacentes são independentes e identicamente distribuídas. A abordagem sistemática pela qual responderemos à questão de termos dados correlacionados no tempo é chamada comumente como *análise de séries temporais*.

A idéia de obter previsões (projeções) de um determinado fenômeno não é absolutamente recente, constituindo-se objeto de estudo de diversas áreas do conhecimento, como Economia, Engenharia, etc. Muitos trabalhos foram desenvolvidos propiciando a evolução de teorias para estudar os fenômenos que se desenvolvem ao longo do tempo. Surgiram, assim, metodologias de análise

de séries temporais com o objetivo de fazer estimativas de valores futuros para os fenômenos.

A análise de séries temporais tem sido de grande utilidade em diversas áreas, como, por exemplo, Economia, Ciências Sociais, Demografia, Epidemiologia e Ciências Ambientais.

Algumas das mais intensivas e sofisticadas aplicações na análise de séries temporais têm sido em problemas de Ciências Físicas e Ambientais. Uma das primeiras séries registradas é a série de manchas solares estudada por Schuster (1906). Mesmo hoje, o fracasso na previsão desta série pode ter implicações nefastas pertinentes; por exemplo, para estimar o tempo de vida de um satélite. Na Ciência Ambiental, pode-se estar interessado em verificar a correlação entre a atividade industrial e os níveis de poluição da água e do ar medidos ao longo do tempo. Estudos têm investigado mortalidade como função da poluição e temperatura.

Estas séries são exemplos de dados experimentais em que a metodologia estatística clássica pode ser substituída pela de séries temporais. O primeiro passo na investigação de uma série temporal deve sempre ser um cuidadoso escrutínio dos dados registrados num gráfico ao longo do tempo. Isso sugere o método de análise, assim como o “summary” estatístico para a informação dos dados.

De acordo com Morettim e Tolo (1987), o estudo de séries temporais tem por filosofia a idéia básica de que o passado é o melhor elemento para se estimar o futuro; assim, os valores passados e presentes da série serão projetados para o futuro. Nesse sentido, entende-se por previsões a tentativa de predizer o comportamento de um sistema em um instante $t + h$, com base nas informações até o momento t , ou seja, passados e presentes.

Uma série temporal é um conjunto de observações ordenadas em intervalos equidistantes no tempo; entretanto, tempo pode ser substituído por qualquer variável como espaço, profundidade, etc. As observações apresentam uma dependência serial entre elas, e o estudo de uma série temporal consiste em analisar e modelar esta dependência. O consumo de energia elétrica de uma residência, valores diários do preço das ações de uma empresa, valores mensais de temperatura da cidade de Lavras, valores diários do índice de produto industrial no Brasil, eletrocardiograma de uma pessoa e registro do movimento da crosta terrestre são alguns exemplos de série temporal.

Para Stevenson (1981), o estudo de tais dados tem por objetivo determinar se eles apresentam algum padrão não aleatório. Por vezes, o que se deseja é, realmente, localizar esse padrão não aleatório que pode então, ser usado para predição quanto ao futuro. Dessa forma, todos os métodos de predição de séries temporais baseiam-se na idéia de que as observações passadas contêm informações de comportamento no futuro. Assim, pode-se dizer que quando se analisa uma série temporal, tem-se como objetivos básicos: modelagem do fenômeno sob consideração; obtenção de conclusões em termos estatísticos e avaliação da adequação do modelo em termos de previsão.

É apropriado mencionar que existem dois métodos para análise de séries temporais distintos, mas não necessariamente exclusivos; eles são identificados como abordagem no *domínio da freqüência* e no *domínio do tempo*.

Segundo Pereira (1984), a abordagem no domínio da freqüência supõe que a série temporal é melhor considerada como uma soma ou superposição linear de ondas periódicas de senos e cossenos de diferentes períodos ou freqüências. Sua justificativa teórica é derivada da representação espectral; essa abordagem é boa em Ciências Físicas e Engenharia, nas quais eventos tendem a ser gerados por fenômenos periódicos.

Por outro lado, a abordagem no domínio do tempo geralmente pressupõe que a correlação entre observações adjacentes é melhor explicada em termos de uma regressão de valores presentes sobre valores passados. Essa pressuposição depende, em teoria, da decomposição de Wold, a qual afirma que observações presentes podem ser preditas como a soma de uma combinação linear de valores passados de uma série de ruídos e um componente determinístico ortogonal àquela combinação linear (Wold 1938). A abordagem no domínio do tempo considera a estimação de coeficientes de regressão e a determinação do número de coeficientes como sendo os primeiros problemas estatísticos a serem resolvidos na modelagem de uma série temporal. Os chamados ARIMA, ou modelos auto-regressivos integrados de médias móveis (Morettin e Toloi, 1987), têm sido aplicados principalmente na análise de séries econômicas nas quais o número de observações é limitado e o principal interesse é fazer previsão. As duas abordagens produzem resultados similares em séries longas, mas a comparação não é clara em séries pequenas.

2.1.2 Decomposição dos Elementos de uma Série

Wold (1938) mostrou que qualquer série temporal poderia ser representada por modelos auto-regressivos e de médias móveis. O teorema de Wold demonstrou que todo processo estacionário admite a representação

$$Z_t = D_t + Y_t \quad (1)$$

em que:

- a) D_t e Y_t são não correlacionados
- b) D_t é determinístico, no sentido de que valores futuros são previstos com precisão baseados em função linear de valores passados D_{t-1}, D_{t-2}, \dots
- c) Y_t é um processo linear discreto (ou médias móveis infinito, $MA(\infty)$)

$$Y_t = \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \dots$$

sendo :

$E\{\varepsilon_t\} = 0$, $E\{\varepsilon_t \varepsilon_{t+k}\} = 0$ e $E\{\varepsilon_t^2\} = \sigma^2 \quad \forall t, k$, isto é, ε_t é um ruído branco.

Assim, segundo Morettin e Tolo (1987), o modelo clássico para séries temporais pressupõe que a série temporal possa ser expressa em termo de seus componentes determinísticos e probabilísticos a partir da decomposição de Wold. Expandindo as partes determinísticas e probabilísticas D_t e Y_t em termo de seus componentes isolados, pode-se escrever a série Z_t , com $t = 1, 2, \dots, n$, como compostas por três componentes não-observáveis: tendência, sazonalidade e uma variação aleatória.

Considerando um modelo aditivo, tem-se:

$$Z_t = T_t + S_t + a_t \quad (2)$$

na qual o componente tendência (T_t) pode ser entendido como um aumento ou uma diminuição gradual das observações ao longo de um período. O componente sazonal (S_t) mostra as flutuações ocorridas em subperíodos (como anos), ocorrendo, por exemplo, mensalmente, trimestralmente ou diariamente. O componente aleatório (a_t) mostra as oscilações irregulares causadas por fenômenos climatológicos excepcionais, intervenções governamentais, etc. A suposição usual é que a_t seja uma série puramente aleatória ou ruído branco, com média zero e variância constante.

De acordo com Gonçalves (1983), quando se observa um acréscimo (ou decréscimo) gradual na série, ou melhor, um comportamento que em grande parte das vezes pode ser descrito por polinômios até de 3º grau, o que se verifica é a existência de tendência. No caso de flutuações periódicas, de periodicidade constante bem definida, é verificada a presença de sazonalidade.

Segundo Morettim e Tolo (1987), o modelo aditivo (2) é adequado, por exemplo, quando S_t não depende dos outros componentes, como T_t . Quando as amplitudes sazonais variarem com a tendência, tem-se o modelo multiplicativo e, apesar da decomposição de Wold traduzir aditividade, pode-se decompor o modelo multiplicativo de maneira similar, conforme:

$$Z_t = T_t \cdot S_t \cdot a_t \quad (3)$$

A previsão de valores futuros da série pode ser obtida modelando-se os três componentes T_t , S_t e a_t de forma isolada. Assim, os métodos de decomposição buscam isolar um ou mais componentes, trabalhando com o restante. Com essa decomposição, são feitas previsões individuais de cada uma das partes, reunindo-as, a seguir, para previsão final.

2.1.3 A Estimação do Componente Tendência

O estudo da tendência tem por objetivo atender a duas finalidades básicas: uma, é identificar a tendência e usá-la, por exemplo, em previsões, e a outra, é remover a tendência, de modo a permitir o estudo de outros componentes da série.

Existem vários procedimentos para estimação do componente tendência em uma série temporal. Os mais utilizados são: tendência polinomial, suavização e o método das diferenças.

Dentre os procedimentos apresentados, o método de tendência polinomial merece enfoque especial.

Supondo que o componente sazonal não esteja presente, e que o modelo seja o aditivo, tem-se:

$$Z_t = T_t + a_t.$$

Estimando-se a tendência através de um estimador \hat{T}_t , pode-se obter a série ajustada livre de tendência, dada por:

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

O método da tendência polinomial, segundo Morettin e Toloi (1987), é um procedimento paramétrico que consiste em ajustar uma curva aos valores da série para estimar e fazer previsões, conforme:

$$T_t = B_0 + B_1 t + \dots + B_m t^m.$$

na qual o grau m do polinômio deve ser bem menor que o número de observações N . Para estimar os parâmetros B_j , pode-se utilizar o método dos mínimos quadrados, ou seja, minimizando:

$$f(B_0, \dots, B_m) = \sum_{t=1}^N (Z_t - B_0 - B_1 t + \dots + B_m t^m)^2,$$

obtendo-se os estimadores de mínimos quadrados usuais $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m$.

A análise gráfica da série é um procedimento que possibilita a identificação de algumas características importantes, como a tendência. Porém, existem na literatura alguns testes estatísticos para verificar a existência e a presença da tendência em uma série (Morettin e Toloi, 1987).

Após a estimação de T_t , pelo método polinomial, pode-se utilizar um procedimento paramétrico para testar a existência da tendência. Assim, com base na teoria usual dos mínimos quadrados, testam-se as seguintes hipóteses:

$$\begin{cases} H_0: \beta_1 = 0 \text{ (não existe tendência)} \\ H_1: \beta_1 \neq 0 \text{ (existe tendência)} \end{cases}$$

É importante ressaltar que se existe outro componente (como S_t) na série, além de T_t , há que eliminá-lo antes de testar a presença de T_t .

Para Hannan e Vieira (1975), a remoção da tendência através de regressão nem sempre é eficiente, pois, a estimação de um componente por vezes é afetada por efeitos de outro.

2.1.4 A Estimação do Componente Sazonal

Existem vários procedimentos para estimação da componente sazonal numa série temporal. Os mais utilizados são: método de regressão, método de médias móveis e o método de diferença sazonal.

Dentre os procedimentos citados, o método de regressão merece enfoque especial, pois oferece a vantagem da utilização de toda a teoria estatística disponível para o modelo linear geral.

O método de regressão, segundo Morettin e Tolo (1986), é ótimo para séries que apresentam sazonalidade determinística, ou seja, que podem ser previstas perfeitamente a partir de meses anteriores.

Este método permite estimar o componente sazonal, conforme descrito por Morettin e Tolo (1986), sendo apresentado resumidamente conforme a seguir.

Se a série Z_t pode ser descrita pela expressão (2), tem-se que as equações representativas da tendência e sazonalidade, admitindo que esta última ocorra dentro de intervalo de tempo correspondente a anos, podem ser modeladas respectivamente por:

$$T_t = \sum_{j=0}^m B_j t^j, \quad t = 1, 2, \dots, N \quad (4)$$

$$S_t = \sum_{j=1}^{12} \alpha_j d_{jt}, \quad t = 1, 2, \dots, N \quad (5)$$

nas quais: m é o grau do polinômio a ser ajustado e d_{jt} são variáveis periódicas (seno, cossenos ou variáveis sazonais dummies) e a_t é ruído branco, com média zero e variância σ_a^2 .

Supondo sazonalidade constante, pode-se afirmar que o índice sazonal α_j não depende de t , pois conforme aumenta a tendência, a sazonalidade acompanha. Assim, tem-se:

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

Impondo a restrição

$$\sum_{j=1}^{12} \alpha_j = 0 \quad (6)$$

tem-se um o modelo de posto completo, e reestruturado de modo que:

$$Z_t = \sum_{j=0}^m \beta_j t^j + \sum_{j=1}^{11} \alpha_j D_{jt} + a_t \quad (7)$$

na qual

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

Através da teoria usual de mínimos quadrados, para a obtenção dos estimadores α_j e β_j , ou seja, para uma amostra Z_1, \dots, Z_N , obtém-se o modelo

$$\underline{Z} = \underline{C} \underline{\beta} + \underline{D} \underline{\alpha} + \underline{a} \quad (8)$$

sendo

$$Z_{N \times 1} = \begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ Z_N \end{bmatrix}, C_{N \times (m+1)} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & 1 \\ 1 & 2 & \cdot & \cdot & 2^m \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & N & \cdot & \cdot & N^m \end{bmatrix}, \beta_{(m+1) \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_m \end{bmatrix},$$

$$D_{N \times 11} = \begin{bmatrix} D_{11} & D_{21} & \cdot & \cdot & D_{11,1} \\ D_{12} & D_{22} & \cdot & \cdot & D_{11,2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ D_{1N} & D_{2N} & \cdot & \cdot & D_{11,N} \end{bmatrix}, \alpha_{11 \times 1} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \alpha_{11} \end{bmatrix} \text{ e } a_{N \times 1} = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_N \end{bmatrix}.$$

O modelo (8) também pode ser escrito de forma particionada, como:

$$\underline{Z} = \underline{X} \gamma + \underline{a}$$

no qual: $\underline{X} = [C | D]$ e $\gamma = \begin{bmatrix} \beta \\ - \\ \alpha \end{bmatrix}$.

Os estimadores de γ podem ser obtidos através do método de mínimos quadráticos usuais, e são:

$$\hat{\gamma} = [\underline{X}'\underline{X}]^{-1} \underline{X}'\underline{Z}.$$

Para o modelo (7), diz-se que existe sazonalidade se a hipótese

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_{11} = 0$$

for rejeitada.

Para um modelo sem tendência (ou trabalhando com dados ajustados para este componente), pode-se aplicar um teste F rotineiro ao considerar o seguinte modelo:

$$Y_{ij} = S_j + e_{ij}, \quad i = 1, \dots, n; j = 1, \dots, s$$

supondo $e_{ij} \sim N(0, \sigma^2)$ independentes.

São testadas as seguintes hipóteses:

$$\begin{cases} H_0 : S_1 = S_2 = \dots = S_s \text{ (não existe sazonalidade);} \\ H_1 : S_i \neq S_j, \text{ para algum } i \text{ e } j \text{ (existe sazonalidade).} \end{cases}$$

A estatística do teste é:

$$T = \frac{N - s}{s - 1} \cdot \frac{\sum_{j=1}^s n_j (\bar{Y}_j - \bar{Y})^2}{\sum_{j=1}^s \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2} \quad (9)$$

que tem, sob H_0 , uma distribuição $F(s - 1, N - 1)$.

Alternativamente, pode-se aplicar um teste nos moldes de análise de variância para experimentos em blocos casualizados para identificar a existência de tendência e sazonalidade. Nesse caso, o efeito de blocos é representado pelos anos e os tratamentos correspondem aos meses.

2.1.5 Funções de Autocorrelação

O procedimento de obtenção da função de autocorrelação, conforme apresentado na obra de Box, Jenkins e Reinsel (1994), é apresentado resumidamente a seguir.

A autocovariância de “lag” k pode ser definida como a covariância entre Z_t e Z_{t+k} , sendo k o número de intervalos de tempo defasados, ou seja:

$$\gamma_k = \text{cov} [Z_t, Z_{t+k}] = E[(Z_t - \mu)(Z_{t+k} - \mu)]. \quad (10)$$

Assim, a autocovariância entre dois pontos é o valor esperado do produto do desvio de cada ponto em relação à média do processo.

Em um processo dito estacionário , a função de autocovariância apresenta as seguintes propriedades:

- i) $\gamma_0 > 0$
- ii) $\gamma_{-k} = \gamma_k$
- iii) $\gamma_0 \geq |\gamma_k|$

Para fins de comparação entre diferentes séries, é utilizada uma medida padronizada do coeficiente de autocovariância, que é obtida através do coeficiente de autocorrelação, que se caracteriza por ser o coeficiente entre a autocovariância, de “lag” k e a autocovariância de “lag” zero, ou seja, a autocorrelação de “lag” k é definida por:

$$\rho_k = \frac{E[(Z_t - \mu)(Z_{t+k} - \mu)]}{\sqrt{E[(Z_t - \mu)^2]E[(Z_{t+k} - \mu)^2]}} = \frac{E[(Z_t - \mu)(Z_{t+k} - \mu)]}{\sigma_Z^2}.$$

Observa-se que, se o processo for estacionário, a variância $\sigma_Z^2 = \gamma_0$ é a mesma no tempo t e no tempo t + k.

Então, a autocorrelação de “lag” k é:

$$\rho_k = \frac{\gamma_k}{\gamma_0}. \quad (11)$$

E, conseqüentemente, com k = 0, tem-se:

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1.$$

Porém, na prática não se conhece os verdadeiros valores das autocorrelações, tem-se apenas amostras de observações. Assim, são necessárias estimativas.

As estimativas das autocovariâncias (γ_k) são dadas por:

$$\hat{\gamma}_k = c_k = \frac{1}{N} \sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z}) \quad k = 0, 1, 2, \dots, N-1 \quad (12)$$

e as estimativas das autocorrelações (ρ_k) por:

$$\hat{\rho}_k = r_k = \frac{c_k}{c_0} \quad k = 0, 1, 2, \dots, n-1 \quad (13)$$

A função que associa cada valor de “k” com o seu respectivo coeficiente de autocorrelação é chamado de função de autocorrelação (f.a.c.). A representação gráfica da função de autocorrelação é chamada de correlograma, sendo obtida colocando-se em um gráfico os valores de ρ_k em função dos “lag” k.

Através da análise do correlograma, é possível identificar diversas características da série em estudo. A seqüência ρ_k , $k = 1, 2, 3, \dots$, é utilizada para medir o número de defasagens e a intensidade da “memória” do processo, isto é, indica o grau de intensidade da correlação existente entre as observações vizinhas da série. É, portanto, uma medida de interdependência entre as observações.

A função de autocorrelação também é importante para verificar se a série é estacionária. Se ao construir o correlograma, o coeficiente de autocorrelação não diminui rapidamente para zero à medida que “k” cresce, então está indicada a não-estacionaridade da série. Em caso contrário, se o coeficiente de autocorrelação diminui rapidamente, estar-se-á em presença de uma série estacionária.

Pode-se também utilizar o correlograma para identificar a presença do componente sazonal. As séries sazonais se caracterizam-se pela regularidade dentro do período de tempo da sazonalidade, apresentando um grau de correlação. Essas correlações serão necessariamente mantidas no correlograma. Assim, pode-se identificar a sazonalidade da série pela observação de picos regulares no correlograma, mesmo que na série original esses picos não possam ser percebidos. Para eliminar a sazonalidade, sugere-se que seja gerada uma nova série a partir de diferenciações. No caso de uma série de período sazonal de 12 meses, por exemplo, será obtida uma nova série dada por $w_t = Z_t - Z_{t-12}$.

2.2 Modelos de Box & Jenkins

2.2.1 Considerações Gerais

O estudo matemático na área de previsão e controle teve seu grande impulso no início da década de setenta, após os trabalhos de Box & Jenkins que, a partir dos resultados de Wold, propuseram uma classe geral de modelos lineares conhecidos por modelos Box & Jenkins, baseados em operações de filtros lineares. Tal método consiste em ajustar modelos auto-regressivos integrados e de médias móveis, ARIMA, a um conjunto de dados.

Aqui, apresentaremos um resumo desta metodologia com as diversas modificações e aperfeiçoamentos sugeridos, desde o trabalho original de Box e Jenkins, tendo por base o livro de Box, Jenkins e Reinsel (1994).

Bustos (1986) apresenta alguns operadores que são extensivamente utilizados na manipulação dos modelos de Box & Jenkins. Estes operadores são introduzidos a seguir.

O operador de defasagem, retardo ou translação para o passado impõe uma defasagem de um período de tempo para trás a cada vez que é usado. É denotado por B e definido por:

$$BZ_t = Z_{t-1}, \quad B^m Z_t = Z_{t-m}$$

Analogamente, o operador translação para o futuro impõe uma defasagem de um período para frente a cada vez que é usado. É denotado por F e definido por:

$$F^m Z = Z_{t+m}$$

O operador diferença (Δ) impõe a diferença de um período para trás a cada vez que é usado. Pode-se escrever:

$$\Delta Z_t = Z_t - Z_{t-1} = (1 - B)Z_t$$

Já o operador soma (S) é o inverso do operador diferença. Então:

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

Uma série temporal pode ser considerada como um caso particular de modelos de filtro linear. Assim, Z_t é gerada por uma série de “choques” independentes e aleatoriamente tomados de uma distribuição fixa; usualmente assume-se a distribuição normal $N(0, \sigma^2)$. Essa série de ruídos será denominada a_t .

O ruído branco a_t é, suposto, transformado na observação Z_t , através do filtro linear $\psi(B)$, conforme apresentado na Figura 1.

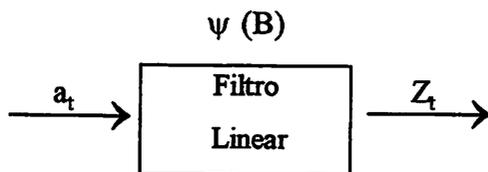


FIGURA 1 Diagrama de um filtro linear, com entrada a_t , saída Z_t e função de transferência $\psi(B)$.

A operação de filtro linear consiste simplesmente de uma soma ponderada dos ruídos anteriores, isto é:

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t \quad (14)$$

na qual

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

é o operador linear chamado função de transferência do filtro; μ é um parâmetro que determina o nível da série, e Z_t é um processo linear (discreto), e ainda:

$$\begin{aligned} E(a_t) &= 0, & \forall t \\ \text{Var}(a_t) &= \sigma_a^2, & \forall t \\ E(a_t a_s) &= 0, & s \neq t. \end{aligned}$$

Diz-se que Z_t é um passeio casual se seu valor no instante t é uma “soma” de choques aleatórios que “entram” no sistema (Figura 1) desde o passado remoto até o instante t ; por outro lado, a primeira diferença é ruído branco.

Uma das suposições mais frequentes que se faz a respeito de 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. Na prática, a maioria das séries apresentam alguma forma de não estacionaridade. Pode-se também ter uma forma de não estacionaridade explosiva, como o crescimento de uma colônia de bactérias. Intuitivamente, um processo Z_t é estacionário se ele se desenvolve no tempo, de modo que a escolha de uma origem dos tempos não seja importante, isto é, as características de $Z(t+h)$ são semelhantes às de $Z(t)$, para todo h .

Tecnicamente, há duas formas de estacionaridade: fraca (ou segunda ordem) e estrita (forte).

Um processo estocástico $Z = \{ Z(t), t \in T \}$ diz-se estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob translação do tempo, ou seja

$$F(Z(t_1), Z(t_2), \dots, Z(t_n)) = F(Z(t_1 + \tau), \dots, Z(t_n + \tau))$$

para quaisquer $t_1, t_2, \dots, t_n, \tau \in T$.

Isto significa, em particular, que todas as distribuições unidimensionais são invariantes sob translações do tempo, ou seja, média $\mu(t) = \mu$ e a variância $V(t) = \sigma^2$ são constantes para todo $t \in T$.

Um processo estocástico $Z = \{ Z(t), t \in T \}$ diz-se fracamente estacionário (ou estacionário de segunda ordem) se e somente se:

- i) $E(Z(t)) = \mu$, constante para todo $t \in T$;
- ii) $E(Z^2(t)) < \infty$, para todo $t \in T$;
- iii) $\gamma(t_1, t_2) = \text{Cov}(Z(t_1), Z(t_2))$ é uma função de $t_1 - t_2$.

Fazendo $\tilde{Z}_t = Z_t - \mu$, a equação (14) pode ser escrita como:

$$\tilde{Z}_t = \psi(B) a_t. \quad (15)$$

A série será estacionária se a seqüência de pesos $\{\psi_j, j \geq 1\}$ é finita ou infinita e convergente, neste caso, o filtro é estável (somável) e μ será a média do processo. Caso contrário, quando a seqüência de pesos é não convergente, o filtro não é estável (Z_t não é estacionária) e, μ é apenas um ponto de referência para o nível da série.

Assim, se média e a variância de Z_t são constantes e a covariância só depende de j , tem-se que Z_t é estacionária e, quando $\sum_{i=0}^{\infty} \psi_i^2$ e $\sum_{i=0}^{\infty} \psi_i \psi_{i+j}$ para $j = 0, \pm 1, \pm 2, \dots$ convergirem, a variância e a autocovariância são dadas por:

$$V(Z_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2, \quad (16)$$

$$\gamma_j = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+j}. \quad (17)$$

Uma maneira alternativa de escrever (15) constitui em representá-la como uma soma ponderada de valores passados $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots$ mais um ruído a_t :

$$\tilde{Z}_t = \pi_1 \tilde{Z}_{t-1} + \pi_2 \tilde{Z}_{t-2} + \dots + a_t = \sum_{j=1}^{\infty} \pi_j \tilde{Z}_{t-j} + a_t,$$

e fazendo:

$$\left(1 - \sum_{j=1}^{\infty} \pi_j B^j \right) \tilde{Z}_t = a_t,$$

obtem-se

$$\pi(B) \tilde{Z}_t = a_t, \quad (18)$$

na qual

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots = \sum_{j=0}^{\infty} \theta^j B^j$$

e $\pi_j = -\theta^j, j \geq 1$.

A sequência de pesos π_j será convergente se $|\theta| < 1$ e, neste caso, o processo é dito invertível. Assim, o processo é invertível se $\pi(B)$ convergir para $|B| \leq 1$, e

$$\tilde{Z}_t = -\theta \tilde{Z}_{t-1} - \theta^2 \tilde{Z}_{t-2} + \dots + a_t.$$

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

2.2.2 Modelos Auto-regressivos - AR

Os modelos auto-regressivos de Box e Jenkins são apresentados resumidamente a seguir, conforme descrito por Morettin e Toloi (1987).

O processo auto-regressivo de ordem p é denotado por AR (p) e representado por

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t \quad (19)$$

na qual $\tilde{Z}_t = Z_t - \mu$, $\phi_1, \phi_2, \dots, \phi_p$ são parâmetros e a_t é o ruído branco.

A expressão (19) pode ser reescrita como:

$$\phi(B)\tilde{Z}_t = a_t, \quad (20)$$

sendo:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p.$$

Assim, no processo auto-regressivo de ordem “ p ”, considerando-se as “ p ” primeiras observações diferentes de zero, a observação Z_t é gerada pela média ponderada das “ p ” primeiras observações próximas passadas da variável, acrescida do termo a_t .

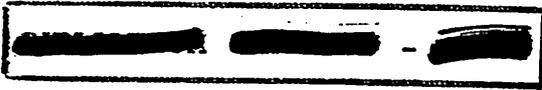
No processo AR(p), a média independe do tempo, se e somente se o processo for estacionário. Nesse caso, tem-se $E(Z_t) = E(Z_{t-1}) = \dots = E(Z_{t-p}) = \mu$. Portanto, se o processo é estacionário, a média do processo é uma constante.

Uma condição necessária (porém não suficiente) para que o processo seja estacionário é que sua média seja finita e determinada, ou seja, necessariamente deve ocorrer

$$\phi_1 + \phi_2 + \dots + \phi_p < 1.$$

Considerando-se o caso mais simples de modelo auto-regressivo de ordem $p=1$, AR(1), tem-se que:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + a_t$$



ou

$$(1 - \phi_1 B) \tilde{Z}_t = a_t$$

ou, ainda,

$$\tilde{Z}_t = (1 - \phi_1 B)^{-1} a_t = (1 + \phi_1 B + \phi_1^2 B^2 + \dots) a_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \dots$$

Como a_t é um ruído branco, tem-se que a variância de \tilde{Z}_t

$$V(\tilde{Z}_t) = (1 + \phi_1^2 + \phi_1^4 + \dots) \sigma_a^2$$

é finita para $|\phi_1| < 1$. Diz-se, então, que o processo auto-regressivo, AR(1), é estacionário se $|\phi_1| < 1$. Considerando-se que

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

é finito, não há restrições sobre os parâmetros para assegurar a invertibilidade de Z_t .

Se G_i^{-1} , $i = 1, \dots, p$, são as p raízes da equação $\phi(B) = 0$, pode-se escrever

$$\phi(B) = (1 - G_1 B)(1 - G_2 B) \dots (1 - G_p B)$$

e, expandindo em frações parciais, tem-se:

$$\psi(B) = \phi^{-1}(B) = \sum_{i=1}^p \frac{A_i}{1 - G_i B}$$

Se $\psi(B)$ deve convergir para $|B| \leq 1$, devemos ter $|G_i| < 1$, $i = 1, \dots, p$. Portanto, um AR(p) é estacionário somente se as raízes de $\phi(B) = 0$ forem, em módulo, maiores que 1, ou seja, é necessário que as raízes de $\phi(B)$ estejam fora do círculo unitário.

Considerando(19), pode-se representar o modelo AR(2), como:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + a_t,$$

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \cdot \\ \cdot \\ \phi_3 \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \cdot \\ \rho_p \end{bmatrix}$$

2.2.3 Modelos de Médias Móveis - MA

Os modelos de médias móveis de Box e Jenkins são apresentados resumidamente a seguir, conforme descrito por Morettin e Toloí (1987).

Seja o processo linear estabelecido pela expressão (14). Considerando-se que somente as primeiras “q” ponderações são diferentes de zero, então o processo

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (24)$$

ou

$$\tilde{Z}_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t = \theta(B) a_t \quad (25)$$

é chamado de processo de média móveis de ordem “q”, simbolizado por MA(q).

Assim, no processo de médias móveis de ordem “q”, cada observação da série é gerada por uma média ponderada do presente, e “q” são os valores passados de um processo de ruído branco.

O processo médias móveis de ordem “q”, MA(q), apresenta a média independente do tempo. Cada a_t é gerado pelo mesmo processo de ruído branco, com $E(a_t) = 0$, $E(a_t^2) = \sigma_a^2$ e $E(a_t, a_{t+k}) = 0$, ou seja, o processo é descrito por q +2, parâmetros que são a média μ , a variância do ruído σ_a^2 e os pesos $\theta_1, \theta_2, \dots, \theta_q$.

Historicamente, adquirem especial importância os processo de médias móveis de primeira e segunda ordem, MA(1) e MA(2).

No caso particular do MA(1), tem-se:

$$\tilde{Z}_t = a_t - \theta a_{t-1}$$

ou

$$\tilde{Z}_t = (1 - \theta_1 B)a_t$$

de modo que $\theta(B) = 1 - \theta_1 B$. Como $\psi(B) = 1 - \theta B$ é finito, o processo é sempre estacionário.

Como

$$a_t = (1 - \theta_1 B)^{-1} \tilde{Z}_t$$

ou

$$a_t = (1 + \theta_1 B + \theta_1^2 B^2 + \dots) \tilde{Z}_t$$

os coeficientes $1, \theta_1, \theta_1^2, \dots$ dessa expressão formam uma seqüência convergente se $|\theta_1| < 1$. Nesse caso, diz-se que MA(1) é invertível, e ele é equivalente a um processo auto-regressivo de ordem infinita.

Considerando-se $\theta(B) = 1 - \theta_1 B$ como um polinômio em B, verifica-se que a raiz de $\theta(B) = 0$ é $1/\theta_1$. Se $|\theta_1| < 1$, essa raiz será, em módulo, maior que 1. Assim, um processo MA(1) será sempre invertível se a raiz de $\theta(B) = 0$ tiver módulo maior que 1, ou seja, estiver fora do círculo unitário.

Considere-se, agora, um modelo MA(2):

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2},$$

o processo é de ordem finita; assim, é sempre estacionário. As raízes da equação característica $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 = 0$ devem estar fora do círculo unitário. Então devem satisfazer:

$$\theta_1 + \theta_2 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$-1 < \theta_2 < 1,$$

que são equivalentes às condições de estacionaridade para um AR(2).

De uma maneira geral, pode-se verificar que a condição de invertibilidade para um modelo MA(q) é a de que as raízes da equação característica $\theta(B) = 0$ tenham módulo maior que 1 (estejam fora do círculo unitário). Nessas condições, diz-se que o modelo MA(q) é equivalente a um processo auto-regressivo AR de ordem infinita.

A função de autocorrelação (f.a.c.) do processo MA(q) é dada por:

$$\rho_j = \frac{-\theta + \theta_1\theta_{j+1} + \theta_2\theta_{j+2} + \dots + \theta_q - j\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}, \quad j = 1, \dots, q \quad (26)$$

$$\rho_j = 0, \quad j > q$$

Observa-se, então, que a fac de um processo MA(q) é igual a zero para “lags” maiores do que q, ao contrário do que ocorre com um processo auto-regressivo AR.

2.2.4 Modelos Mistos - ARMA

Em muitas séries, na prática, o uso de processos AR e MA puros com um número muito grande de parâmetros é insuficiente para representar o comportamento de um sistema. Assim, segundo Kassouf (1988), combinando-se os dois processos AR e MA, obtém-se um modelo auto-regressivo de média móvel ARMA (p, q), da forma:

$$\tilde{Z}_t = \phi_1\tilde{Z}_{t-1} + \dots + \phi_p\tilde{Z}_{t-p} + a_t - \theta_1a_{t-1} - \dots - \theta_qa_{t-q} \quad (27)$$

na qual os valores correntes da série temporal dependem dos valores passados da série e dos erros correntes e passados.

Segundo Morettin e Tolo (1987), se $\phi(B)$ e $\theta(B)$ são os operadores auto-regressivos e de médias móveis, pode-se escrever (27) na forma compacta:

$$\phi(B)\tilde{Z}_t = \theta(B)a_t \quad (28)$$

ou

$$\tilde{Z}_t = \phi^{-1}(B)\theta(B)a_t.$$

Para o modelo mais simples, ou seja, ARMA(1, 1), em que $p = q = 1$, o modelo (27) se reduz a

$$\tilde{Z}_t = \phi_1\tilde{Z}_{t-1} + a_t - \theta_1a_{t-1}.$$

A equação (28) define um processo estacionário se as raízes de $\phi(B) = 0$ caírem fora do círculo unitário e a condição de inversibilidade é a de que as raízes de $\theta(B) = 0$ estejam fora da região do círculo unitário.

A função de autocorrelação de um processo ARMA é obtida multiplicando-se ambos os membros de (27) por \tilde{Z}_{t-j} e tomando as esperanças; assim:

$$\gamma_j = E[\tilde{Z}_t\tilde{Z}_{t-j}]$$

$$\begin{aligned} \gamma_j = & \phi_1 E[\tilde{Z}_{t-1}\tilde{Z}_{t-p}] + \dots + \phi_p E[\tilde{Z}_{t-p}\tilde{Z}_{t-j}] + E[a_t\tilde{Z}_{t-j}] - \theta_1 E[a_{t-1}\tilde{Z}_{t-j}] - \\ & - \dots - \theta_q E[a_{t-q}\tilde{Z}_{t-j}] \end{aligned}$$

$$\gamma_j = \phi_1\gamma_{j-1} + \dots + \phi_p\gamma_{j-p} + \gamma_{za}(j) - \theta_1\gamma_{za}(j-1) - \dots - \theta_q\gamma_{za}(j-q) \quad (29)$$

na qual $\gamma_{za}(j)$ é a covariância cruzada entre Z_t e a_t , definida por

$$\gamma_{za}(j) = E(a_t\tilde{Z}_{t-j}) \Rightarrow \begin{cases} = 0 & \text{se } j > 0 \\ \neq 0 & \text{se } j \leq 0. \end{cases} \quad (30)$$

Consequentemente, a função de autocovariância do processo será

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \dots + \phi_p\gamma_{j-p}, \quad j > q. \quad (31)$$

Assim, conforme (11), obtém-se a f.a.c.

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \dots + \phi_p\rho_{j-p}, \quad j > q \quad (32)$$

Observando a expressão (30), é possível deduzir que as autocorrelações de “lags” 1, 2, ..., q serão afetadas pelos parâmetros do processo MA, mas para $j > q$, as mesmas comportam-se como nos modelos AR.

2.2.5 Função de Autocorrelação Parcial

As funções de autocorrelação dos processos AR(p), MA(q) e ARMA(p, q) possuem características especiais e de grande importância na identificação do modelo. Essas características são apresentadas adiante, conforme encontradas em Morettin e Tolo (1987):

- a) para um processo auto-regressivo AR(p), a função de autocorrelação decai de acordo com exponenciais e/ou senóides amortecidos, sendo infinitas em extensão.
- b) para um processo de médias móveis MA(q), a função de autocorrelação é finita, no sentido que apresenta um corte após o “lag” q (após q todas as fac são iguais a zero).
- c) para um processo misto ARMA(p, q), a função de autocorrelação é infinita em extensão, e decai de acordo com exponenciais e/ou senóides amortecidos após “lag” q - p.

Uma forma alternativa para a identificação de modelos é proposta por Box e Jenkins através das funções de autocorrelações parciais (f.a.c.p). Este procedimento é apresentado adiante tendo por base Box, Jenkins e Reinsel (1994).

Considerando-se um modelo AR conforme definido em (19), com função de autocorrelação dada por (22), e indicando por ϕ_{kj} o j-ésimo coeficiente do processo AR(p), então ϕ_{kk} será o último coeficiente. Então, segundo Box, Jenkins e Reinsel (1994), podemos escrever:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}, \quad j = 1, \dots, k, \quad (33)$$

a partir das quais, obtemos as equações de Yule-Walker:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \cdot \\ \cdot \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \cdot \\ \rho_k \end{bmatrix} \quad (34)$$

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

$$\phi_{11} = \rho_1,$$

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

AR para (2,2)

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}},$$

Handwritten scribbles

e, em geral;

$$\phi_{kk} = \frac{|\rho_k^*|}{|\rho_k|}, \quad (35)$$

em que ρ_k refere-se à matriz de autocorrelação de ordem k , e ρ_k^* , à matriz ρ_k , com a última coluna trocada pelo vetor de autocorrelações.

Assim, o valor ϕ_{kk} será a função de autocorrelação parcial entre as variáveis Z_t e Z_{t-k} , dadas as variáveis intermediárias Z_{t-1} e Z_{t-k+1} .

Para obter-se a estimativa da correlação parcial, existem duas opções: a primeira consiste em estimar, sucessivamente, modelos auto-regressivos de ordens $p = 1, 2, \dots$ por mínimos quadrados e tomar as estimativas do último coeficiente de cada ordem; a outra consiste em substituir ρ_k pelas suas estimativas r_k .

Para Kassouf (1987), as funções de autocorrelação parcial dos processos AR(p), MA(q) e ARMA(p, q) possuem as seguintes características que são úteis na identificação dos modelos:

- a) para um processo auto-regressivo AR(p), a função de autocorrelação parcial é tal que: $\phi_{kk} \neq 0$, para $k \leq p$ e $\phi_{kk} = 0$, para $k > p$;
- b) para um processo de médias móveis MA(q), a facp se comporta de maneira similar à fac de um processo AR(p) com exponenciais e/ou senóides amortecidos;
- c) para um processo ARMA(p,q), a facp se comporta como a facp de um processo MA(q) puro.

Para facilitar o processo de identificação dos processos AR, MA e ARMA, estão apresentados na Tabela 1, conforme Gonçalves (1983), as características de cada um desses processos quanto à condição de invertibilidade, estacionaridade, função de autocorrelação e função de autocorrelação parcial.

Na Figura A.1 do anexo, são apresentados exemplos de funções de autocorrelação e autocorrelação parcial para modelos AR(1), MA(1) e ARMA(1,1).

TABELA 1 Condição de estacionaridade, invertibilidade, fac e facp para os modelos AR(p), MA (q) e ARMA (p, q)

Modelo	AR (p)	MA (q)	ARMA (p, q)
Forma	$\phi(B)\tilde{Z}_t = a_t$ $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$	$\tilde{Z}_t = \theta(B)a_t$ $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$	$\phi(B)\tilde{Z}_t = \theta(B)a_t$
Condição de Estacionaridade	Raízes de $\phi(B)=0$ fora do círculo unitário	Sempre estacionário	Raízes de $\phi(B)=0$ fora do círculo unitário
Condição de Invertibilidade	Sempre invertível	Raízes de $\theta(B)=0$ fora do círculo unitário	Raízes de $\theta(B)=0$ fora do círculo unitário
f.a.c.	Infinita (mistura de exponenciais e ou senoides amortecidos)	$\rho_k = 0 \quad k > q$ Apenas q primeiros valores não nulos.	Infinita (mistura de exponenciais e ou senoides amortecidos após q - p primeiros "lags")
f.a.c.p.	$\phi_{kk} = 0 \quad k > p$ Apenas p primeiros valores não nulos.	Infinita (mistura de exponenciais e ou senoides amortecidos)	Infinita (mistura de exponenciais e/ou senoides amortecidos após p - q primeiros "lags")

2.2.6 Modelo Auto-Regressivo Integrado - ARIMA

Os modelos apresentados até aqui são apropriados para descrever séries estacionárias, que se desenvolvem no tempo aleatoriamente ao redor de uma média constante. Quando as séries são não estacionárias, porém homogêneas, ou seja, processos cujo nível e/ou inclinação mudam com o decorrer do tempo, pode-se obter a estacionaridade da série pela aplicação do operador diferença entre valores consecutivos, isto é, calculando-se $W_t = \Delta^d Z_t$. Diante dessa nova série estacionária W_t , pode-se ajustar um modelo ARMA (p, q), ou seja,

$$\phi(B)W_t = \theta(B)a_t. \quad (36)$$

Uma vez que $W_t = \Delta^d Z_t = (1 - B)^d Z_t$, obtém-se

$$\phi(B)\Delta^d Z_t = \theta(B)a_t, \quad (37)$$

que é chamado de modelo auto-regressivo-integrado-médias móveis, ou simplesmente, modelo ARIMA de ordem (p, d, q).

Considerando o modelo (37), então:

- i) $\phi(B)\Delta^d = \phi(B)(1 - B)^d$ é um polinômio que tem d raízes iguais a 1, ou seja, pertencentes ao círculo unitário;
- ii) tirando-se estas d raízes, por diferença, sobram p raízes, ou seja, com as raízes do polinômio $\phi(B)$, podendo, assim, tratá-lo como se fosse um modelo ARMA.

Escrever (37) é equivalente a escrever

$$\xi(B)Z_t = \theta(B)a_t,$$

em que $\xi(B)$ é o operador auto-regressivo não estacionário homogêneo, de ordem $p + q$, tal que d raízes da equação característica $\xi(B) = 0$ estão sobre o círculo unitário e as demais, fora.

Dessa maneira, pode-se escrever o modelo ARIMA(p, d, q) como:

$$\xi(B)Z_t = \phi(B)\Delta^d Z_t = \phi(B)(1 - B)^d Z_t = \theta(B)a_t, \quad (38)$$

em que:

- $\phi(B)$ é o operador auto-regressivo, supondo que seja estacionário;
- $\xi(B) = \Delta^d \phi(B)$ é um operador não estacionário, em que as d raízes da equação característica $\xi(B) = 0$ são iguais a um e as restantes fora do círculo unitário;
- $\theta(B)$ é o operador média móvel. Supõe-se que seja invertível.

A estimativa da f.a.c. é dada por:

$$r_j = \frac{c_j}{c_0}, \quad j = 0, 1, \dots, N-1,$$

em que:

$$c_j = \frac{1}{N} \sum_{i=1}^{N-j} [(Z_i - \bar{Z})(Z_{i+j} - \bar{Z})], \quad j = 0, 1, \dots, N-1.$$

Para N suficientemente grande e sob hipótese que $\rho_j = 0$, para $j > q$, a distribuição é aproximadamente normal, com média zero e variância dada por

$$\sigma^2(r_j) = \frac{1}{N} \left[1 + 2 \sum_{v=1}^q r_v^2 \right], \quad j > q.$$

Assim, um intervalo de confiança aproximado para as autocorrelações é:

$$r_j \pm t_\gamma \hat{\sigma}^2(r_j),$$

na qual t_γ é o valor da estatística t de Student, com $N-1$ graus de liberdade. Na prática, usa-se $t_\gamma = 2$, considerando ρ_j significativamente diferente de zero, quando:

$$|r_j| > 2\hat{\sigma}(r_j), \quad j > q. \quad (39)$$

Para a função de autocorrelação parcial, sob hipótese que o processo é $AR(p)$

$$Var(\phi_{jj}) \cong \frac{1}{N}, \quad j > p, \quad (40)$$

de modo que

$$\hat{\sigma}(\phi_{jj}) \cong \frac{1}{\sqrt{N}}, \quad j > p.$$

Assim, ϕ_{jj} terá distribuição aproximadamente normal, com média zero e variância dada por (40), de modo que consideraremos ϕ significativamente diferente zero se:

$$|\phi_{jj}| > \frac{2}{\sqrt{N}}, \quad j > p. \quad (41)$$

Os modelos ARIMA, segundo Gonçalves (1983), trabalham com séries estacionárias ou aproximadas, mediante diferenciações ou transformações como as de Box-Cox.

A construção dos modelos é baseada em um ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados. Box & Jenkins sugerem que se “deixe que os dados falem por si mesmo”.

Para Morettin e Tolo (1987), este ciclo iterativo envolve a fase de identificação, estimação, verificação e obtenção das previsões. Elas estão descritas a seguir.

Durante a fase de identificação, procura-se estabelecer as características básicas da série, tais como: se a série é estacionária ou não estacionária; há existência ou não de sazonalidade e tendência. O objetivo desta fase é identificar e determinar os valores de p , d e q do modelo ARIMA (p , d , q) e estimar preliminarmente os parâmetros a serem utilizados na fase de estimação.

O procedimento de identificação envolve o estudo das funções de autocorrelação que, inicialmente, informarão se a série é ou não estacionária.

Para obter estacionaridade, diferencia-se a série através do operador diferença. O número de diferenças d necessárias para que o processo se torne estacionário, é alcançado quando a função de autocorrelação da série $W_t = \Delta^d Z_t$, decresce rapidamente para zero.

Após estabelecer d , obtendo a estacionaridade, trabalha-se com a série W_t , que pode ser representada por um processo ARMA(p , q). Para se estabelecer a ordem (p , q) deve-se proceder à análise das funções de autocorrelação e autocorrelação parcial, obtendo estimativas preliminares das ordens p , d e q do modelo ARIMA.

Na maioria das séries, o número de diferenças d é pequeno, em geral $d \leq 2$. Na prática, para tais casos, é suficiente inspecionar somente as primeiras 15 ou 20 autocorrelações da série diferenciada.

Por fim, nesta fase, através das funções de autocorrelações amostrais da série $W_t = \Delta^d Z_t$, também obtém-se estimativas iniciais dos parâmetros.

Para um AR(p), as estimativas iniciais de ϕ podem ser extraídas a partir das equações de Yule-Walker (23), com ρ_j substituído por suas estimativas r_j .

No caso do processo MA(q), utiliza-se (26), que fornece q equações com q incógnitas, $\theta_1, \dots, \theta_q$, substituindo-se ρ_j por r_j .

Após a etapa de identificação, o passo seguinte é a estimação propriamente dita dos parâmetros. Será necessário usar um procedimento iterativo de estimação de mínimos quadrados não-linear e as estimativas preliminares encontradas na fase de identificação serão usadas como valores iniciais para esse procedimento. Considerando-se $\phi = (\phi_1, \dots, \phi_p)$ e $\theta = (\theta_1, \dots, \theta_q)$, o objetivo consiste em minimizar $S(\phi, \theta)$, sendo

$$S(\phi, \theta) = \sum a_t^2$$

na qual

$$a_t = \theta^{-1}(B)\phi(B) W_t.$$

Após estimar o modelo, a fase de verificação pode ser feita analisando-se os resíduos. Suponha-se que o modelo ajustado é definido por (36). Os erros são, em geral, assumidos normais e independentes, ou seja, constituem um ruído branco. Assim, se o modelo for verdadeiro, os resíduos deverão seguir estas propriedades.

Se o modelo for adequado, os resíduos estimados

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

deverão estar próximos dos a_t e aproximadamente não correlacionados. Se \hat{r}_k indicarem as autocorrelações dos resíduos \hat{a}_t , então deve-se ter $\hat{r}_k \cong 0$. Particularmente, deve-se ter, aproximadamente,

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^n \hat{a}_t^2} \sim N\left(0, \frac{1}{n}\right). \quad (42)$$

De uma maneira geral, pode-se dizer que o modelo é adequado quando se tem \hat{r}_k dentro dos limites $\pm \frac{2}{\sqrt{n}}$.

Segundo Morettin e Toloi (1987), a comparação de \hat{r}_k com os limites $\pm \frac{2}{\sqrt{n}}$ fornece uma indicação geral de possível quebra de comportamento de ruído branco em a_t , com a condição de que seja lembrado que, para pequenos valores de k , estes limites subestimarão a significância de qualquer discrepância.

Box e Pierce (1970), buscando sanar um dos grandes problemas enfrentados para modelos ARIMA, a saber, o da determinação de sua ordem p e q , propuseram a estatística x^2 . Já Ozaki (1977), formulou o procedimento de obtenção do estimador de mínimos quadrados para o critério de informação de Akaike, chegando ao critério de informação de Akaike normalizado.

O teste de Box & Pierce, segundo Kassouf (1988), é baseado nas primeiras k autocorrelações dos \hat{a}_t , isto é, nos valores de $r_k(\hat{a})$ para ($k = 1, 2, \dots, K$). Para um processo ARIMA(p, d, q), se o modelo ajustado é apropriado, então

$$Q = n(n+2) \sum_{k=1}^K \frac{\hat{r}_k^2}{(n-k)} \quad (43)$$

tem distribuição aproximadamente quiquadrado (χ^2) com $(k - p - q)$ graus de liberdade. A hipótese de ruído branco é rejeitada para valores $Q < \chi^2(k - p - q)$. Em geral, basta tomar as primeiras 20 ou 25 primeiras \hat{r}_k .

Para Gonçalves (1983), uma forma alternativa de determinar a ordem de um modelo, dentro de uma classe identificada de modelos, é selecionar aquele que minimize o critério Akaike (AIC) normalizado, expresso por

$$AIC = N \cdot \log \hat{\sigma}_a^2 + \frac{N}{N-d} 2 \cdot (p+q+1+\delta_{d0}) + N \cdot \log 2\pi + N \quad (44)$$

sendo:

$$\delta_{d0} = \begin{cases} 1, & d = 0 \\ 0, & d \neq 0 \end{cases}$$

Na comparação de diversos modelos, com N fixo, pode-se eliminar os dois últimos termos da expressão (44). Assim, o critério de Akaike pode ser expresso como:

$$AIC = N \log \hat{\sigma}_a^2 + 2(p+q+2) \quad (45)$$

2.2.7 Modelos ARIMA Sazonal - SARIMA

As séries sazonais são caracterizadas por mostrarem alta correlações da variável distanciada pelo período da sazonalidade e pela correlação entre observações próximas. Muitas vezes, mesmo após eliminar a componente sazonal determinística, ainda resta correlação significativa em “lags” de baixa ordem ou em “lags” sazonais, isto é, múltiplos. Assim, pode-se ter dois tipos de modelos sazonais: os determinísticos e os estocásticos.

2.2.7.1 Sazonalidade Estocástica

Segundo Morettin e Tolo (1987), para as séries com sazonalidade estocástica, pode-se aplicar um modelo ARIMA sazonal, ou seja, o modelo SARIMA, sem nenhuma dificuldade adicional. Inicialmente, há que se calcular diferenças com lag 1 e com lag s, ou seja, tem-se que tomar d diferenças simples e D diferenças sazonais de série Z_t , a fim de produzir estacionaridade. Com isso obtém-se:

$$w_t = \Delta^d \Delta_s^D Z_t = (1-B)^d (1-B^s)^D Z_t \quad (46)$$

O próximo passo é observar as funções de autocorrelação e autocorrelação parcial amostrais da série W_t nos lags 1, 2,... para obter os valores de p e q e nos lags s, 2s, ..., e assim obter P e Q, selecionando-se, então, um modelo prévio SARIMA (p, d, q) x (P, D, Q).

O procedimento seguinte é estimar os valores dos parâmetros identificados, utilizando-se os estimadores de máxima verossimilhança, de maneira idêntica ao exposto para o processo ARIMA; porém, agora com um modelo inicial SARIMA.

2.2.7.2 Sazonalidade Determinística

Segundo Pereira (1984), para as séries que possuem sazonalidade determinística, o que geralmente ocorre em séries de temperatura e outras variáveis meteorológicas, o modelo SARIMA a ser ajustado tem características específicas. Esse procedimento é apresentado resumidamente a seguir, conforme encontrado em Morettin e Tolo (1987).

Quando Z_t exibe um comportamento determinístico com período 12 (por simplicidade de exposição, considera-se $s = 12$). Um modelo que pode ser útil é

$$Z_t = \mu_t + N_t \quad (47)$$

em que: μ_t é uma função determinística periódica, satisfazendo $\mu_t - \mu_{t-12} = 0$, ou

$$(1 - B^{12}) \mu_t = 0 \quad (48)$$

e N_t é um processo estacionário que pode ser expresso por um modelo ARMA (p, q).

Portanto, N_t satisfaz à equação

$$\phi(B)N_t = \theta(B)\alpha_t \quad (49)$$

sendo a_t um ruído branco e μ_t é dado por:

$$\mu_t = \mu + \sum_{j=1}^6 \left[\alpha_j \cos\left(\frac{2\pi j t}{12}\right) + \beta_j \sin\left(\frac{2\pi j t}{12}\right) \right] \quad (50)$$

na qual $\mu, \alpha_j, \beta_j, j = 1, \dots, 6$ são constantes desconhecidas.

Assim, para um modelo sazonal determinístico (47), aplicando a diferença sazonal $(1 - B^{12})$, obtém-se:

$$(1 - B^{12})Z_t = (1 - B^{12})\mu_t + (1 - B^{12})N_t$$

e de acordo com (47) e (48), tem-se

$$\phi(B) (1 - B^{12})Z_t = \theta(B) (1 - B^{12})a_t$$

ou

$$\phi(B) w_t = \theta(B) (1 - B^{12})a_t \quad (51)$$

em que $w_t = (1 - B^{12})Z_t$.

O ajuste do modelo com sazonalidade determinística passa pelos passos de identificação, estimação e previsão.

O processo de identificação é feito em duas etapas. Primeiramente obtêm-se as estimativas preliminares $\hat{\mu}, \hat{\alpha}_j, \hat{\beta}_j$ de μ, α_j e β_j respectivamente, $j = 1, \dots, 6$, em (51) através da análise de regressão Z_t sobre $1, \sin\frac{2\pi j t}{12}$ e $\cos\frac{2\pi j t}{12}$, $j = 1, \dots, 6$. O passo seguinte consiste em calcular os resíduos

$$\tilde{N}_t = Z_t - \hat{\mu} - \sum_{j=1}^6 \left[\tilde{\alpha}_j \cos\left(\frac{2\pi j t}{12}\right) + \tilde{\beta}_j \sin\left(\frac{2\pi j t}{12}\right) \right] \quad (52)$$

procedendo, em seguida, ao exame das funções de autocorrelação e autocorrelação parcial para identificar um modelo ARMA(p, q) para N_t .

A estimação de máxima verossimilhança dos parâmetros μ , α_i , β_i , ϕ_i e θ_k com $i = 1, \dots, 6$, $j = 1, \dots, p$ e $k = 1, \dots, q$ é obtida de maneira similar à utilizada na estimação dos parâmetros do modelo ARMA.

2.3 Previsão

2.3.1 Considerações Preliminares

Uma das principais razões para decompor uma série temporal em seus componentes é prever seus valores futuros.

Porém, um modelo que descreve bem uma série amostral não conduz, necessariamente, a previsões eficazes. Será necessário especificar uma função perda, além do modelo, para se chegar a um procedimento ótimo.

Morettin e Toloi (1986) salientam que a previsão não constitui um fim em si, mas apenas um meio de fornecer informações para uma conseqüente tomada de decisões, visando a determinar objetivos.

Há interesse em se prever Z_{t+h} , $h \geq 1$, em função das observações $\dots, Z_{t-2}, Z_{t-1}, Z_t$, até o instante t , que é chamada de origem das previsões.

A previsão Z_{t+h} é denotada por $\hat{Z}_t(h)$, e é a esperança matemática condicional de Z_{t+h} , dado Z_t, Z_{t-1}, \dots , isto é,

$$\hat{Z}_t(h) = [Z_{t+h}] = E[Z_{t+h} / Z_t, Z_{t-1}, Z_{t-2}, \dots].$$

2.3.2 Previsão com Modelos Suavizados

Os modelos de amortecimento ou suavizados caracterizam por sua flexibilidade para tratar conjuntamente as causas determinísticas e probabilísticas de flutuações de séries temporais. Essa técnica assume que os

valores extremos da série representam a aleatoriedade, e assim, através do alisamento desses extremos, pode-se identificar o padrão básico de variação.

Estes modelos, do ponto de vista da estimação de parâmetros, são mais simples que os modelos de Box & Jenkins, que possuem uma formulação estatística rebuscada; porém, podem em muitas situações gerar projeções mais sofisticadas, embora do ponto de vista estatístico, sejam formulados com menor rigor.

2.3.2.1 - Alisamento Exponencial Simples (AES)

O Alisamento Exponencial Simples apresenta a idéia de que o valor de uma série temporal no ponto t (Z_t) pode ser expresso como uma média ponderada dos valores anteriores ($Z_{t-1}; Z_{t-2}; Z_{t-3}; \dots Z_1$) em que os pesos decrescem de forma exponencial à medida que se afastam para trás do ponto t . O procedimento de obtenção destes modelos é apresentado a seguir conforme exposto em Galber (1995). Desse modo, tem-se:

$$\bar{Z}_t = \alpha Z_t + (1 - \alpha) \bar{Z}_{t-1}, \quad \bar{Z}_0 = Z_1, \quad t = 1, 2, \dots, N$$

ou

$$\bar{Z}_t = \alpha \sum_{k=0}^{t-1} (1 - \alpha)^k Z_{t-k} + (1 - \alpha)^t \bar{Z}_0, \quad t = 1, 2, \dots, N \quad (53)$$

sendo \bar{Z}_t denominado valor exponencial alisado e α é a constante de alisamento, $0 \leq \alpha \leq 1$.

Desenvolvendo-se a expansão (53), tem-se que

$$\bar{Z}_t = \alpha Z_t + \alpha(1 - \alpha) Z_{t-1} + \alpha(1 - \alpha)^2 Z_{t-2} + \dots$$

através da qual pode-se perceber que:

1. A projeção para o momento t é uma ponderação dos valores anteriores.

2. O peso decresce exponencialmente à medida que se afasta do momento t .
3. Se ocorre uma constante de amortecimento exponencial perto de 1, está-se valorizando as observações mais perto do ponto de projeção do modelo. Por outro lado, se a constante está perto de 0, está-se valorizando as observações mais antigas da série para elaboração das projeções.

A forma de se obter o melhor valor para α é através de método numérico, não existindo uma forma analítica de otimizar a constante de amortecimento. O que se faz são varreduras no intervalo $[0, 1]$, escolhendo-se o modelo que apresenta o menor erro quadrático médio ($EQM_{\text{ajustamento}}$).

A previsão dos valores futuros é dada pelo último valor exponencialmente alisado, isto é:

$$\hat{Z}_t(h) = \bar{Z}_t, \quad h = 1, 2, 3, \dots \quad (54)$$

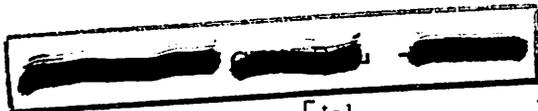
$$\hat{Z}_t = \alpha Z_t + (1 - \alpha) \hat{Z}_{t-1}(h + 1).$$

Quanto menor o valor de α , mais estáveis serão as previsões finais, uma vez que a utilização de baixo valor de α implica que pesos finais maiores são dados às observações passadas, e, conseqüentemente, qualquer flutuação aleatória no presente exercerá um peso menor no cálculo da previsão.

Assim como ocorre nos modelos de regressão linear simples, as estimativas do erro quadrático médio (EQM) são dadas pela média dos quadrados das diferenças entre os valores observados e os valores projetados, ou seja, a projeção um passo a frente:

$$EQM = \frac{1}{N} \sum_{i=1}^N (Z_i - \hat{Z}_i)^2.$$

A variância de Z_{t+h} é dada por



$$\begin{aligned} \text{Var}(\hat{Z}_t(h)) &= \text{Var} \left[\sum_{k=0}^{t-1} \alpha(1-\alpha)^k Z_{t-k} \right] \\ &= \frac{\alpha \sigma_a^2 [1 - (1-\alpha)^{2t}]}{2 - \alpha} \end{aligned}$$

e, supondo $a_t \sim N(0, \sigma_a^2)$, o intervalo de confiança assintótico para Z_{t+h} é

$$\left[\hat{Z}_t(h) - z_\gamma \sqrt{\frac{\alpha}{2-\alpha} \sigma_a^2}; \hat{Z}_t(h) + z_\gamma \sqrt{\frac{\alpha}{2-\alpha} \sigma_a^2} \right]$$

em que z_γ é o coeficiente da $N(0, 1)$ e $\hat{Z}_t(h)$ é dado por (54).

2.3.2.2 Alisamento Exponencial Biparamétrico de Holt (AEH)

Segundo Morettin e Tolo (1987), trata-se de um método não sazonal para séries temporais, que apresentam tendência, em princípio similar ao AES, apresentando a seguinte diferença: ao invés de alisar só o nível, ele utiliza uma nova constante de alisamento para modelar a tendência da série.

O AEH permite grande flexibilidade, na utilização de constantes de alisamento com valores diferentes para o nível e tendência, e tem por desvantagem, a dificuldade na escolha das constantes de suavização e inferências sobre média e variância da previsão. Este procedimento é abordado por Morettin e Tolo (1987), conforme descrito a seguir.

Os valores do nível e da tendência, no instante t , são fornecidos pelas expressões.

$$\begin{aligned} \bar{Z}_t &= \alpha Z_t + (1-\alpha)(\bar{Z}_{t-1} - \hat{T}_{t-1}), & 0 < \alpha < 1 & \text{ e } & t = 2, \dots, N \\ \hat{T}_t &= \beta(\bar{Z}_t - \bar{Z}_{t-1}) + (1-\beta)\hat{T}_{t-1}, & 0 < \beta < 1 & \text{ e } & t = 2, \dots, N \end{aligned} \quad (55)$$

nas quais α e β são constantes de alisamento.

Para esse modelo de suavização, o nível \bar{Z}_t é ajustado diretamente através da tendência no período anterior \hat{T}_{t-1} , juntando-se o último valor

suavizado. Dessa forma, \bar{Z}_t é trazido a valores mais recentes; a seguir, atualiza-se a tendência, colocando-se na forma de diferença entre os últimos valores observados.

O procedimento de determinação das constantes de alisamento é análogo ao AES, só que ao invés de escolher um valor que minimiza a soma dos erros quadráticos de previsão, escolhe-se o valor do vetor (α, β) para que isso ocorra.

A equação de previsão é dada por:

$$\hat{Z}_t(h) = \hat{Z}_t + h\hat{T}_t \quad h = 1, 2, \dots,$$

ou seja, a previsão é feita adicionando-se ao valor básico (\bar{Z}_t) a tendência multiplicada pelo número de passos à frente que se deseja prever.

Previsões tendo uma nova observação é dada por:

$$\hat{Z}_{t+1}(h-1) = \hat{Z}_{t+1} + (h-1)\hat{T}_{t+1}.$$

com:

$$\begin{aligned} \bar{Z}_{t+1} &= AZ_{t+1} + (1-A)(\bar{Z}_t - \hat{T}_t) \\ \hat{T}_{t+1} &= C(\bar{Z}_{t+1} - \bar{Z}_t) + (1-C)\hat{T}_t \end{aligned}$$

Para os valores iniciais, são consideradas as seguintes hipóteses

$$\hat{T}_2 = Z_2 - Z_1 \text{ e } \bar{Z}_2 = Z_2.$$

2.3.2.3 Alisamento Exponencial Sazonal de Holt-Winters - (HW)

Alisamento Exponencial Sazonal de Holt-Winters, segundo Gonçalves (1983), é adequado para séries que apresentam, além da tendência, um componente sazonal. É baseado em equações de suavização, cada uma associada a um componente do padrão da série: nível, tendência e sazonalidade. Existem

duas abordagens distintas, sendo uma para as séries com sazonalidades multiplicativas e outra para as com sazonalidades aditivas.

A abordagem com sazonalidade aditiva é apresentada por Morettin e Toloi e é descrita a seguir.

Quando se supõe aditividade, considera-se o fator sazonal F_t e a tendência como fatores aditivos, assim:

$$Z_t = \mu_t + F_t + T_t + a_t, \quad t = 1, \dots, N.$$

As estimativas do fator sazonal, do nível e da tendência são dadas pelas seguintes equações de alisamento,

$$\begin{aligned} \hat{F}_t &= \gamma(Z_t - \bar{Z}_t) + (1 - \gamma)\hat{F}_{t-s}, & 0 < \gamma < 1 \\ \bar{Z}_t &= \alpha(Z_t - \hat{F}_{t-s}) + (1 - \alpha)\bar{Z}_{t-1} + \hat{T}_{t-1}, & 0 < \alpha < 1 \\ \hat{T}_t &= \beta(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - \beta)\hat{T}_{t-1}, & 0 < \beta < 1 \end{aligned} \quad (56)$$

respectivamente. Os termos α , β e γ são constantes de alisamento determinadas de modo a tornar mínima a soma dos quadrados dos erros de ajustamento.

A estimativa do modelo é dada por:

$$\begin{aligned} \hat{Z}_t(h) &= \bar{Z}_t + h\hat{T}_t + \hat{F}_{t+s}, \quad h = 1, 2, \dots, s \\ \hat{Z}_t(h) &= \bar{Z}_t + h\hat{T}_t + \hat{F}_{t+2s}, \quad h = s+1, 2, \dots, 2s \\ &\vdots \\ &\vdots \\ &\text{etc.} \end{aligned}$$

Ainda segundo Morettin e Toloi (1987), quando se supõe efeito sazonal multiplicativo, o modelo pode ser considerado como misto, pois considera-se o fator sazonal F_t como multiplicativo, enquanto a tendência permanece aditiva, de modo que a série é representada por

$$Z_t = \mu_t \cdot F_t + T_t + a_t, \quad t = 1, \dots, N.$$

As estimativas do fator sazonal, do nível e da tendência são dadas pelas seguintes equações de alisamento, respectivamente:

$$\begin{aligned}\hat{F}_t &= \gamma \left\{ \frac{Z_t}{\bar{Z}_t} \right\} + (1-\gamma)\hat{F}_{t-s}, & 0 < \gamma < 1, & \quad t = s + 1, \dots, N, \\ \bar{Z}_t &= \alpha \left\{ \frac{Z_t}{\hat{F}_{t-s}} \right\} + (1-\alpha)(\bar{Z}_{t-1} + \hat{T}_{t-1}), & 0 < \alpha < 1, & \quad t = s + 1, \dots, N, \\ \hat{T}_t &= \beta(\bar{Z}_t - \bar{Z}_{t-1}) + (1-\beta)\hat{T}_{t-1}, & 0 < \beta < 1, & \quad t = s + 1, \dots, N,\end{aligned} \quad (57)$$

sendo α , β e γ constantes de alisamento determinadas de modo a tornar mínima a soma dos quadrados dos erros de ajustamento.

As previsões são dadas por:

$$\begin{aligned}\hat{Z}_t(h) &= (\bar{Z}_t + h\hat{T}_t)\hat{F}_{t+s}, \quad h = 1, 2, \dots, s, \\ \hat{Z}_t(h) &= (\bar{Z}_t + h\hat{T}_t)\hat{F}_{t+2s}, \quad h = 1, 2, \dots, 2s, \\ &\vdots \\ &\text{etc.}\end{aligned}$$

2.3.3 Previsão com Modelo ARIMA

2.3.3.1 Representação e Notação

Para Morettin e Tolo (1987), um modelo ARIMA (p, d, q) pode ser escrito de três formas alternativas:

i) diretamente em forma de equação de diferença

$$\begin{aligned}Z_{t+h} &= \xi_1 Z_{t+h-1} + \dots + \xi_p Z_{t+h-p} - \theta_1 a_{t+h-1} - \theta_2 a_{t+h-2} - \dots \\ &\quad \dots - \theta_q a_{t+h-q} + a_{t+h}\end{aligned} \quad (58)$$

ii) uma soma de choques aleatórios

$$Z_{t+h} = \sum_{j=-\infty}^{t+h} \psi_{t+h-j} a_j = \sum_{j=0}^{\infty} \psi_j a_{t+h-j}; \quad (59)$$

em que $\psi_0 = 1$ e os demais pesos são obtidos de $\xi(B)\psi(B) = \theta(B)$.

iii) forma invertida

$$Z_{t+h} = \sum_{j=1}^{\infty} \pi_j Z_{t+h-j} + a_{t+h}, \quad (60)$$

em que os pesos π_j são obtidos de $\xi(B) = \theta(B) \pi(B)$ e são tais que

$$\sum_{j=1}^{\infty} \pi_j = 1.$$

2.3.3.2 Variância da Previsão e Intervalo de Confiança

Segundo Pereira (1983), pode-se estar interessado em prever Z_{t+h} como função linear dos valores presentes e passados $Z_t, Z_{t-1}, Z_{t-2}, \dots$, que, por sua vez, podem ser escritos como funções lineares dos resíduos presentes e passados a_t, a_{t-1}, \dots (pois $Z_t = \psi(B) a_t$). Logo, se for indicada a melhor previsão por

$$\hat{Z}_t(h) = \psi_h^* a_t + \psi_{h+1}^* a_{t-1} + \psi_{h+2}^* a_{t-2} + \dots,$$

os pesos que minimizam EQM de previsão são dados por:

$$E[Z_{t+h} - \hat{Z}_t(h)]^2 = E\left[\sum_{j=0}^{\infty} \psi_j a_{t+h-j} - \sum_{j=0}^{\infty} \psi_{h+j}^* a_{t-j}\right]^2.$$

A primeira soma pode ser escrita como $\sum_{j=h}^{\infty} \psi_{h+j} a_{t-j}$; portanto, o erro de

previsão é dado por:

$$e_t(h) = Z_{t+h} - \hat{Z}_t(h) \quad (61)$$

ou

$$e_t(h) = \psi_0 a_{t+h} + \psi_1 a_{t+h-1} + \dots + \psi_{h-1} a_{t+1} - \sum (\psi_{h-1} - \psi_{h+j}^*) a_{t-j}.$$

Assim,

$$E[e_t(h)]^2 = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2) \sigma_a^2 + \sum_{j=0}^{\infty} (\psi_{h+j} - \psi_{h+j}^*) \sigma_a^2, \quad (62)$$

como os a_t são não correlacionados, minimiza-se $E[e_t(h)]^2$, se $\psi_{h+j}^* = \psi_{h+j}$.

Portanto, a previsão de EQM mínimo é dada por:

$$\hat{Z}_t(h) = \psi_h a_t + \psi_{h+1} a_{t-1} + \psi_{h+2} a_{t-2} + \dots = \sum_{j=0}^{\infty} \psi_{h+j} a_{t-j}$$

e o erro de previsão (61) fica:

$$e_t(h) = a_{t+h} + \psi_1 a_{t+h-1} + \dots + \psi_{h-1} a_{t+1}. \quad (63)$$

Tem-se, então, que $E[e_t(h)] = 0$ e a variância de erro de previsão é:

$$V(h) = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2) \sigma_a^2.$$

Os erros de previsão um passo a frente são não correlacionados. Por outro lado, os erros de previsão para um intervalo maior que um serão correlacionados, o mesmo ocorrendo com os erros de previsão para o mesmo horizonte h , de diferentes origens t e $t-j$.

A obtenção de intervalos de previsão é apresentada por Morettin e Toloi (1987), e são detalhadas adiante.

Para podermos determinar um intervalo de confiança para Z_{t+h} , será necessário fazer as seguintes suposições para os resíduos:

- i) $E(a_t) = 0$;
- ii) $E(a_t a_s) = 0$, $t \neq s$;
- iii) $\text{Var}(a_t) = \sigma_a^2$ para todo t ;
- iv) $a_t \sim N(0, \sigma_a^2)$, para cada t .

Assim, a distribuição condicional de Z_{t+h} , dados os valores passados e presentes da série Z_t, Z_{t-1}, \dots , será $N(\hat{Z}_t(h), V(h))$.

Logo:

$$u = \frac{Z_{t+h} - \hat{Z}_t(h)}{[V(h)]^{1/2}} \sim N(0, 1)$$

e fixando o coeficiente γ , pode-se encontrar um valor u_γ tal que $P(-u_\gamma < U < u_\gamma) = \gamma$. Substituindo $V(h)$ por sua estimativa $\hat{\sigma}_a^2$ obtida no estágio de estimação do modelo, pode-se construir o intervalo de previsão:

$$\hat{Z}_t(h) - u_\gamma \hat{\sigma}_a \left[1 + \sum_{j=1}^{h-1} \psi_j^2 \right]^{1/2} \leq Z_{t+h} \leq \hat{Z}_t(h) + u_\gamma \hat{\sigma}_a \left[1 + \sum_{j=1}^{h-1} \psi_j^2 \right]^{1/2}.$$

É importante ressaltar que a variância aumenta com h ; logo as amplitudes desses intervalos aumentarão à medida que se afastam da origem t , caracterizando o aumento da incerteza das previsões para h passos a frente.

2.3.3.3 Cálculo e Atualização das Previsões

Segundo Kassouf (1988), o mais conveniente modo de fazer previsões com modelos ARIMA é escrever o modelo na forma de equações de diferença. Tomando esperanças condicionais em (58) obtém-se:

$$\hat{Z}_t(h) = \zeta_1 [Z_{t+h-1}] + \dots + \zeta_{p+d} [Z_{t+h-p-d}] - \theta_1 [a_{t+h-1}] - \dots - \theta_q [a_{t+h-q}] + [a_{t+h}], \quad (64)$$

para $h \geq 1$. Foram considerados os seguintes fatos:

$$\begin{aligned} [Z_{t+k}] &= \hat{Z}_t(k), & k > 0, \\ [Z_{t+k}] &= Z_{t+k}, & k \leq 0, \\ [a_{t+k}] &= 0, & k > 0, \\ [a_{t+k}] &= a_{t+k}, & k \leq 0. \end{aligned}$$

A equação de previsão, considerada como uma função de h , com origem t fixa, satisfaz a equação de diferenças:

$$\hat{Z}_t(h) = \sum_{i=1}^{p+d} \xi_i \hat{Z}_t(h-i), \quad h > q$$

ou

$$\xi(B)\hat{Z}_t(h) = (1-B)^d \phi(B)\hat{Z}_t(h) = 0, \quad h > q$$

com $\xi(B)$ operando sobre h .

Morettin e Toloi (1987), demonstram que para $h > q - p - d$, a função $\hat{Z}_t(h)$ consistirá em uma mistura de polinômios exponenciais e senóides amortecidos, com sua forma exata determinada pelas raízes G_i^{-1} do operador $\xi(B) = 0$.

Para Kassouf (1988), quando observa-se Z_{t+1} , as previsões $\hat{Z}_t(1), \dots, \hat{Z}_t(h)$ podem ser facilmente atualizadas para a nova origem $t+1$, usando-se a relação:

$$\hat{Z}_{t+1}(h) = \hat{Z}_t(h+1) + \psi_h a_{t+1}, \quad j = 1, 2, \dots, h$$

na qual:

$$a_{t+1} = Z_{t+1} - \hat{Z}_t(1) = a_t(1).$$

Este resultado é obtido lembrando-se que:

$$i) \quad t+1: \hat{Z}_{t+1}(h) = \psi_h a_{t+1} + \psi_{h+1} a_t + \psi_{h+2} a_{t-1} + \dots$$

$$ii) \quad t: \hat{Z}_t(h+1) = \psi_{h+1} a_t + \psi_{h+2} a_{t-1} + \dots$$

2.3.4 - Previsão com Modelos SARIMA

Para Pereira (1984), o cálculo das previsões com os modelos SARIMA(p, d, q)(P, D, Q) $_s$ com sazonalidade estocásticas, o procedimento é

análogo ao procedimento exposto para o processo ARIMA (p, d, q), utilizando-se da equação de diferença expressa como (64).

Por outro lado, para série com sazonalidades determinísticas, conforme definidas na seção 2.2.7.2, as previsões de valores futuros Z_{t+h} , dados Z_1, Z_2, \dots, Z_t , são dadas por:

$$\hat{Z}_t(h) = \mu_{t+h} + \hat{N}_t(h).$$

sendo μ_{t+h} e $N_t(h)$ calculados utilizando-se (50) e (49) respectivamente.

2.3.5 - Previsão com Modelos ARMA

Para Pereira (1984), os modelos ARIMA (p, d, q) e SARIMA (p, d, q)(P, D, Q)_s podem ser considerados respectivamente como modelos ARMA (p+d, q) e ARMA(p + d + D + P_s, q + Q_s). O procedimento de obtenção das previsões, conforme descrito por Pereira, é apresentado de forma resumida a seguir.

Um modelo ARMA (p, q) estacionário e invertível pode ser escrito de três formas alternativas:

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

ii) $\tilde{Z}_t = \psi(B)a_t$, em que $\psi(B) = \phi^{-1}(B)\theta(B)$

iii) $\pi(B)\tilde{Z}_t = a_t$, em que $\pi(B) = \theta^{-1}(B)\phi(B)$.

Para o cálculo da equação de previsão, considera-se inicialmente

$$\begin{aligned} E[Z_{t+h}] &= E[a_{t+h}] + \psi_1 E[a_{t+h-1}] + \dots + \psi_h E[a_t] + \psi_{h+1} E[a_{t-1}] + \dots \\ &= \psi_h a_t + \psi_{h+1} a_{t-1} + \dots \\ &= \hat{Z}_t(h) \end{aligned}$$

pois,

$$E[a_{t+j}] = \begin{cases} 0 & j = 1, 2, \dots \\ a_{t+j} & j = 0, -1, -2, \dots \end{cases}$$

A equação de previsão para

$$Z_{t+h} = \phi_1 Z_{t+h-1} + \dots + \phi_p Z_{t+h-p} + a_{t+h} + \theta_1 a_{t+h-1} + \dots + \theta_q a_{t+h-q}$$

é

$$\hat{Z}_{t+h} = \phi_1 E[Z_{t+h-1}] + \dots + \phi_p E[Z_{t+h-p}] + E[a_{t+h}] + \theta_1 E[a_{t+h-1}] + \dots + \theta_q E[a_{t+h-q}]$$

fazendo-se

$$E[Z_{t+j}] = \begin{cases} Z_{t+j} & j = 0, -1, -2, \dots \\ \hat{Z}_{t+j} & j = 1, 2, \dots \end{cases}$$

$$E[a_{t+j}] = \begin{cases} 0 & j = 1, 2, \dots \\ a_{t+j} = Z_{t+j} - \hat{Z}_{t+j-1}(1) & j = 0, -1, -2, \dots \end{cases}$$

Assim, para um modelo ARMA(p, 0), ou seja, AR(p) tem-se:

$$\hat{Z}_t(1) = \phi_1 Z_t + \dots + \phi_p Z_{t-p+1}$$

$$\hat{Z}_t(2) = \phi_1 \hat{Z}_t(1) + \dots + \phi_p Z_{t-p+2}$$

.....

$$\hat{Z}_t(p) = \phi_1 \hat{Z}_t(p-1) + \dots + \phi_p Z_t$$

$$\hat{Z}_t(p+1) = \phi_1 \hat{Z}_t(p) + \dots + \phi_p \hat{Z}_t(1)$$

em geral

$$\hat{Z}_t(p+j) = \phi_1 \hat{Z}_t(p+j-1) + \dots + \phi_p \hat{Z}_t(j)$$

Para um modelo ARMA(0, q), ou seja, MA(q) tem-se:

$$\begin{aligned} \hat{Z}_t(1) &= \theta_1 a_t + \theta_2 a_{t-1} + \dots + \theta_q a_{t-q+1} \\ \hat{Z}_t(2) &= \theta_2 a_t + \theta_3 a_{t-1} + \dots + \theta_q a_{t-q+2} \\ &\dots\dots\dots \\ \hat{Z}_t(q) &= \theta_q a_t \\ \hat{Z}_t(q+j) &= 0 \quad j = 1, 2, \dots \end{aligned}$$

E, finalmente, para o modelo ARMA (p, q) tem-se:

$$\begin{aligned} \hat{Z}_t(1) &= \phi_1 Z_t + \dots + \phi_p Z_{t-p+1} + \theta_1 a_t + \dots + \theta_q a_{t-q+1} \\ \hat{Z}_t(2) &= \phi_1 \hat{Z}_t(1) + \dots + \phi_p Z_{t-p+2} + \theta_2 a_t + \dots + \theta_q a_{t-q+2} \end{aligned}$$

Em geral, para $j = 1, 2, \dots$, tem-se:

$$\begin{aligned} \hat{Z}_t(q+j) &= \phi_1 \hat{Z}_t(q+j-1) + \dots + \phi_p \hat{Z}_t(q+j-p), \quad q \geq p \\ \hat{Z}_t(q+j) &= \phi_1 \hat{Z}_t(q+j-1) + \dots + \phi_p \hat{Z}_t(j), \quad q \geq p \end{aligned}$$

Quando observa-se Z_{t+1} , as previsões $\hat{Z}_t(1), \dots, \hat{Z}_t(h)$ podem ser facilmente atualizadas para a nova origem $t+1$ (como no modelo ARIMA) usando a relação:

$$\hat{Z}_{t+1}(j) = \hat{Z}_t(j+1) + \psi_j a_{t+1}, \quad j = 1, 2, \dots, h$$

sendo:

$$a_{t+1} = Z_{t+1} - \hat{Z}_t(1) = a_t(1).$$

Esse resultado é obtido lembrando-se que

- i) $\hat{Z}_t(j+1) = \psi_{j+1} a_t + \psi_{j+2} a_{t-1} + \psi_{j+3} a_{t-2} + \dots$
- ii) $\hat{Z}_{t+1}(h) = \psi_j a_{t+1} + \psi_{j+1} a_t + \psi_{j+2} a_{t-1} + \dots$

A variância e o intervalo de previsão com o modelo ARMA podem ser obtidos de modo análogo ao procedimento utilizado para o modelo ARIMA.

2.4 - Periodograma (Representação Espectral)

Uma série temporal, segundo Pereira (1984), também pode ser analisada no domínio da frequência; nestes casos, interessado na frequência com que certos eventos ocorrem em um período de tempo (movimentos cíclicos, por exemplo). A ferramenta utilizada é o espectro (transformação da função de autocorrelação) e a análise é baseada em modelos não paramétricos.

Os estudos da análise espectral são bastante antigos, podendo-se citar Jenkins e Watts (1968), que expõem amplamente a teoria espectral. Para Pereira (1984), o estudo das séries temporais através da análise espectral é baseada no seguinte par de teoremas:

- a) Teorema de Bochner-Khintchin - A função de autocorrelação $\rho(\tau)$ de um processo estacionário discreto no tempo admite representação:

$$\rho(\tau) = \int_{-\pi}^{\pi} e^{i w \tau} dF(w) = \int_0^{\pi} [\sin(w\tau) dQ(w) + \cos(w\tau) dC(w)]$$

na qual $F(w)$ tem as propriedades de uma função de distribuição de probabilidade no intervalo $(-\pi, \pi)$, isto é, $F(-\pi) = 0$ e $F(\pi) = 1$. A notação $dF(w)$ significa integração ou soma conforme $F(w)$ seja contínuo ou uma função escada. A função $F(w)$ é denominada distribuição espectral.

- b) Teorema de Cramer - Todo processo estacionário discreto no tempo com função de distribuição espectral $F(w)$ admite a representação

$$Y_t = \int_{-\pi}^{\pi} e^{i w \tau} dF(w) = \int_0^{\pi} [\sin(w\tau) \alpha(w)dw + \cos(w\tau) \beta(w)dw]$$

em que $\alpha(w)$ e $\beta(w)$ são funções estocásticas com propriedades análogas a α_i e β_i do processo harmônico. Da mesma forma, a variância de Y_t é decomposta na soma:

$$\int dF(w) = F(\pi).$$

Assim, este outro caminho para analisar uma série temporal está baseado nas ondas de seno e cosseno para diferentes frequências. Essa análise no domínio da frequência constitui uma ótima ferramenta na procura de características determinísticas de uma série temporal.

Segundo Shumway (1992), o periodograma foi originalmente usado para detectar e estimar a amplitude do componente senóide de frequência conhecida. Posteriormente foi usado por Box & Jenkins para verificar a aleatoriedade da série (geralmente uma série de resíduos após o ajuste de um modelo), em que se considera a possibilidade de que componentes periódicos de frequência desconhecida possam ainda permanecer na série. O estudo da periodicidade de uma série, conforme Priestley (1989), é apresentado resumidamente a seguir.

O periodograma da série a_t , $t = 1, 2, \dots, n$, é definido por:

$$I_a(f_i) = \frac{2}{n} \left[\left(\sum_{t=1}^n a_t \cos \frac{2\pi i}{n} t \right)^2 + \left(\sum_{t=1}^n a_t \sin \frac{2\pi i}{n} t \right)^2 \right] \quad (65)$$

com $0 < f_i < \frac{1}{2}$, em que $I(f_i)$ indica a intensidade da frequência f_i .

A periodicidade da série pode ser verificada observando a existência de picos na frequência $f_i = \frac{i}{n}$, o que indica uma periodicidade de período $\frac{1}{f_i}$.

O procedimento inicial é fazer o gráfico do periodograma tendo na ordenada as frequências.

Se for observado que no periodograma existe mais de um pico, não se pode concluir imediatamente se esses picos correspondem a genuínos

componentes periódicos da série. Deve-se, portanto, aplicar um teste para verificar se o pico é um componente periódico genuíno.

O teste de Fisher para verificar a existência de periodicidade foi proposto inicialmente para o maior período, conforme (66):

$$g^* = \frac{\max(I_p)}{\sum_{p=1}^{N/2} I_p}. \quad (66)$$

Porém, foi estendido para os períodos de menor intensidade, sendo a estatística do teste dada por:

$$g = \frac{I_p}{\left\{ \sum_{p=1}^{N/2} I_p \right\} - I_p} \quad (67)$$

Sendo que a distribuição exata para g é dada por:

$$P(g > z) = \alpha = n(1 - z)^{n-1} \quad (68)$$

em que $n = \frac{N}{2}$ e α é o nível de significância. São testadas as seguintes hipóteses

$$\begin{cases} H_0: \text{existe periodicidade} \\ H_1: \text{não existe periodicidade} \end{cases}$$

Assim, se $g \geq z$, aceita-se H_0 , ou seja, a série apresenta período p . Para maiores detalhes, pode-se consultar Pristley (1989).

Foi mostrado por Bartlett que o periodograma acumulado proporciona um meio eficaz de detectar a não randomicidade periódica.

A chamada função espectral $p(l)$ para o ruído branco tem um valor constante $2\sigma_a^2$ no domínio da frequência de 0 a 0,5 ciclos; conseqüentemente, a função espectral acumulada para o ruído branco é:

$$p(f) = P_a(f) = \int_0^f p_a(g)dg = \begin{cases} 0, & f < 0 \\ 2\sigma_a^2 f, & 0 \leq f < 1/2 \\ \sigma_a^2, & f \geq 1/2. \end{cases}$$

ou seja, para ruído branco, o espectro varia linearmente com a frequência. Então, quando a série \hat{a}_t se aproxima do ruído branco, a função $C(f_j) \times f_j$ tem comportamento linear entre os pontos (0, 0) e (0.5, 1) e aceita-se o modelo fixo como válido.

2.5 Aplicações

A metodologia das séries temporais vem sendo utilizada ao longo do tempo para ajustar modelos a um conjunto de dados e, assim, realizar previsões. A seguir, citam-se alguns.

Morettin e Toloí (1981) fazem sugestões para a determinação das ordens dos processos ARIMA, incluindo os modelos bayesianos, fazendo comparações para séries reais, especialmente as econômicas. Em 1989, os mesmos autores apresentam os modelos de funções de transferências baseados na metodologia de Box e Jenkins para analisar modelos ARIMA.

Pereira (1984) apresentou a metodologia para séries temporais multivariadas e séries temporais com informações concomitantes. Em particular, mostra diferentes enfoques para a identificação da função de transferência.

Para Couto (1996), “os modelos de séries temporais possibilitam o estudo e a elaboração de previsões a partir de séries de dados, em que se tem pouco conhecimento dos fatores que determinam suas variações, desde que se disponha de informações em um horizonte ($t = 1, \dots, T$) que permita o ajuste satisfatório dos modelos.”

Schmitz e Watts (1970), citados por Gonçalves (1983), utilizaram a técnica de Box e Jenkins, mais especificamente o modelo ARIMA e a técnica do

alisamento exponencial, para obter previsões da produção de trigo em quatro grandes países exportadores. Os autores concluíram que o método de previsão escolhido para cada país depende do conjunto de dados.

Pino e Morettin (1981) utilizaram modelos sazonais e modelos de análise de intervenção para a série de produção de leite no período (1975-1980), para a série de produção (1882-1980) e para a série de preços do café tipo exportação(1954-1980) do Estado de São Paulo. Foi possível concluir que fortes geadas e eventos históricos afetaram significativamente as séries.

Kassouf (1988) utilizou o modelo ARIMA, entre outros, para obter previsões dos preços da arroba do boi magro e do boi gordo no Estado de São Paulo, levando-se em consideração a tendência e as variações cíclicas. As previsões obtidas foram comparadas aos preços reais do mercado físico, para avaliar a capacidade preditiva do modelo. Para a série do boi gordo, obteve-se como melhor modelo um ARIMA (0,1,1).

Os modelos ARIMA e estrutural foram comparados por Vicente (1992) no estudo do comportamento das séries de produção de carne de frango e ovos. As melhores previsões foram obtidas com o modelo estrutural.

Mais recentemente, pode-se citar os trabalhos de Pino e Rocha (1994), que se utilizaram do modelo ARIMA como função de transferência, para detectar a influência dos preços praticados na Bolsa de Chicago nos preços diários recebidos pelos produtores de soja e pelas indústrias brasileiras.

A previsão de preços na pecuária também foi alvo de estudo de Couto (1996), que trabalhou com uma série de preços da arroba do boi gordo no Estado de São Paulo. Para tanto, utilizou os modelos estruturais de séries temporais, empregando o filtro de Kalman e, ainda, considerou efeitos de transferência (variáveis explicativas) e efeitos de intervenção (dos cinco planos econômicos ocorridos no período em estudo).

E, finalmente, para as séries de dados climatológicos, pode-se citar o trabalho de Gonçalves (1983), que comparou diversos métodos para previsão de séries temporais, dentre os quais: Naive, Suavizados, Auto-regressivos e Box & Jenkins. Esses modelos foram estudados na série de precipitações pluviiais mensais no Município de Pindorama - SP, no período de janeiro de 1950 a abril de 1981. Pesquisou-se a acurácia dos modelos em casos excessivamente aleatórios. O modelo que obteve melhor desempenho foi o de Box & Jenkins, auto-regressivo de medias móveis (ARMA). Executando-se o modelo ARMA, o melhor dentre os estudados neste trabalho, foi o de suavização sazonal aditivo de Holt e Winters.

Segundo Morettin (1984), muitos autores, incluindo Markham (1974), Girardi e Texeira (1978), Almeida et al (1980), Jones e Kearns (1976), Nobre et al (1982), analisaram a série de precipitação de Fortaleza, e disponível para o período de 1849 até 1979. Exceto Jones e Kearns (1976), todos concluíram que a periodicidade de 13 e 26 anos estava presente nos dados, enquanto que dois autores concluíram que os dados eram ruído branco.

Morettin (1984) reanalisou a série em questão, utilizando um conjunto de testes estatísticos (Fisher, Whittle, Bartlett, Hannan, Priestley, Siegel e Fase), os quais sugeriram a existência de periodicidade de 13 e 26 anos na série de precipitações atmosféricas de Fortaleza. Contudo, ressalta que não se pode tirar conclusões definitivas, especialmente no que se refere a extrapolações para todo o polígono da seca.

Nobre et al (1984) também trabalharam com a série de precipitações de Fortaleza, com especial enfoque para modelos de previsão de secas no Nordeste pelo método da periodicidade. Foi possível verificar que a série de precipitações de Fortaleza não foi representativa da chuva de áreas mais amplas do Nordeste, somente para anos de grande seca, impondo uma limitação à previsão mais

geral. Foram ajustados modelos auto-regressivos de ordem 1, AR(1), e o modelo senóide, que obteve melhor resultado, mas explicou apenas 24% da variância da série original. Contudo, o modelo senóide se mostrou-se bastante limitado para previsão de valores que se afastam consideravelmente da média; por conseguinte, o modelo se mostrou de utilidade bastante limitada para a previsão de grandes secas no Nordeste, visto que, pelo histórico da série, nos anos das grandes secas, o total anual de precipitação em Fortaleza sempre foi menor que 1000mm, e para esses valores, o nível de confiança foi de apenas 50% ou menor.

2.6 Outras Metodologias

Barger & Thom (1949) verificaram que, em períodos pequenos, com até quatro semanas, o aspecto da curva se ajusta a dados de pluviometria, sugerindo uma distribuição exponencial negativa para ajuste e, à medida que aumenta o período, a tendência da curva é ser unimodal assimétrica, para depois se aproximar da curva normal. E, a partir dessas características, os autores utilizaram a distribuição gama por admitir maior flexibilidade.

A função potência foi utilizada por Castro Neto et al (1980a) para estudar os dados diários de chuva na região de Lavras-MG e para identificar os meses que apresentam maiores probabilidades de ocorrência de períodos secos. Usando esse mesmo procedimento, Castro Neto et al (1980 b) verificaram que as maiores probabilidades de ocorrência de períodos chuvosos em Lavras ocorreram nos meses de novembro a março. Já Castro Neto e Silveira (1981) utilizaram-se da distribuição gama para estudar os dados de Lavras-MG em períodos mensais, concluindo que menores precipitações prováveis foram encontradas nos meses de junho, julho e agosto e as maiores nos meses de outubro e março.

Para estudar a quantidade de chuva diária em Pelotas-RS, no período de 1893 a 1991, Assis (1991) construiu modelos teóricos a partir da distribuição binominal negativa truncada e da distribuição gama. Ao analisar os totais semanais de chuva, concluiu que a quantidade de chuva nas semanas com chuva pode ser representada pela função densidade de probabilidade gama.

Piccinini (1993), para determinar a distribuição de probabilidade das alturas pluviométricas de intensidade máxima para Piracicaba - SP, testou os modelos de distribuição normal e gama com dados de 33 anos. Os resultados demonstraram que as duas distribuições podem ser utilizadas, sendo que a distribuição gama apresentou melhor resultado quando se considerou períodos mensais, e a distribuição normal para períodos anuais.

A distribuição gama, aplicada a dados diários de chuva, segundo Almeida (1995), tem sido utilizada por muitos, e tem se mostrado adequada para descrever a distribuição de quantidade de chuva sob as mais variadas condições climáticas.

Botelho (1998) estimou os parâmetros da distribuição gama para modelar o regime de chuva em Lavras. Foram utilizados dados de precipitações diárias (mm) relativas ao período de 1966 a 1996. As observações referentes aos dados foram agrupados em períodos de 1-3, 1-6, 1-9, 1-12, 1-15, 1-18, 1-21, 1-24, 1-27 dias e em totais mensais. Os resultados demonstraram que em 65% dos períodos estudados os dados se ajustaram a uma distribuição gama, sendo mais eficiente nos períodos maiores e nos meses de maiores precipitações. Além disso, observou-se que os dados diários não se ajustaram à distribuição gama. Constatou-se, ainda, que alturas prováveis podem ser obtidas considerando-se um certo nível de probabilidade. Assim, por exemplo, existe uma probabilidade de 90% de chover até 281mm no mês de março e 10% de chover até 79,8 mm.

3 MATERIAL E MÉTODOS

3.1 Material

A base de dados para o ajustamento do modelo de séries temporais foi obtida da Área de Agrometeorologia do Departamento de Engenharia da Universidade de Lavras e oriunda da Estação Climatológica Principal de Lavras-MG, situada no campus da Universidade Federal de Lavras, em Lavras, Estado de Minas Gerais, em convênio com o Instituto Nacional de Meteorologia (INMET). O município de Lavras está situado na latitude de 21°14'S, longitude de 45°00'W e altitude média 918 metros, Brasil (1992). O clima da região segundo a classificação Köppen, é considerado de transição entre Cwa e Cwb, variando de subtropical a temperado propriamente dito, com chuvas predominantemente no verão e tendo o inverno considerado como seco. As observações referem-se às precipitações pluviais diárias expressas em altura de lâmina d'água (mm), abrangendo o período de janeiro de 1966 a dezembro de 1997, num total de 32 anos.

3.2 Métodos

Os dados foram ajustados a modelos de séries temporais, levando em conta a acentuada variabilidade dos fenômenos climatológicos. Nesse sentido, os registros das precipitações diárias foram agrupados em registros mensais, e a série foi constituída de 384 totais mensais sucessivos de precipitação pluviométrica de Lavras; seus valores estão expostos no quadro B.1 em anexo.

Considerou-se o período de janeiro de 1966 a dezembro de 1996 para ajustar os modelos, e as observações, do período de janeiro de 1997 a dezembro de 1997, foram utilizadas para verificação da previsão.

A metodologia usada para ajustar modelos de séries temporais à série de precipitações pluviiais mensais no município de Lavras foi a seguinte:

1. Familiarização:

- construção de gráficos para descrição do comportamento do conjunto de dados através da inspeção visual;
- verificação da existência de tendência e de sazonalidade através dos seguintes testes: análise de variância, considerando-se o delineamento em blocos casualizados, método de regressão, e através da análise da função de autocorrelação;
- verificação da periodicidade através da análise do periodograma da série e a aplicação da estatística g de Fisher, conforme descrito no item 2.4.

2. Estimativas da tendência e sazonalidade:

- a tendência foi estimada pelo método de ajuste polinomial, conforme descrito no item 2.1.3;
- a sazonalidade foi estimada pelo método de regressão, conforme descrito no item 2.1.4, considerando-se a expressão 7.

3. Ajuste de modelos:

- foi ajustado o modelo de regressão do item 2.1.4 na forma:

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

- foi ajustado o modelo SARIMA do item 2.2.8 na forma:

$$Z_t = \mu_t + N_t ,$$

sendo
$$\mu_t = \mu + \sum_{j=1}^6 \left[\alpha_j \cos \frac{(2\pi j t)}{12} + \beta_j \sin \frac{(2\pi j t)}{12} \right];$$

- as estimativas das funções de autocorrelações e autocorrelações parciais foram utilizadas para testar a adequabilidade dos modelos, ou seja, se os resíduos constituíam um ruído branco;
- aos resíduos estimados, foram ajustados modelos de Box e JenKins através de um ciclo iterativo composto dos seguintes estágios: especificação, identificação, estimação e verificação conforme descrito no item 2.2.7;
- a escolha entre os diversos modelos propostos se deu através do procedimento AIC (minimum Akaike information criterium estimation), conforme expressão 45.

4. Previsões

- foram obtidas previsões com modelos suavizados e com os modelos ajustados (regressão e SARIMA);
- os modelos foram testados quanto ao desempenho através da análise gráfico e da soma dos quadrados dos erros de previsão.

Para a realização deste estudo, foram utilizados como suporte computacional os seguintes softwares: STATISTICA e SAS(SAS Institute, 1998).

4 RESULTADOS E DISCUSSÃO

4.1 Familiarização

A Figura 2 apresenta a análise da série em estudo, sugerindo, inicialmente, a presença do componente sazonal significativa com periodicidade anual e ausência do componente tendência. A presença da sazonalidade já era esperada, pois em séries climatológicas, este aspecto é extremamente comum.

Quanto à aditividade, aparentemente não existe forma de dependência entre as causas de variação, ou seja, a variação sazonal mantém-se aproximadamente inalterada, não demonstrando a presença de efeitos multiplicativos, sugerindo que a série possa assumir a forma aditiva.

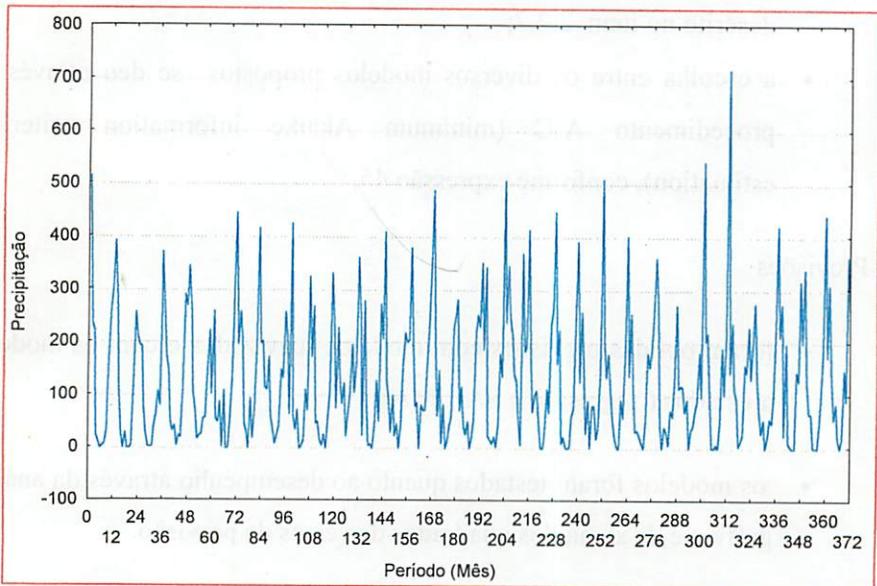


FIGURA 2 Representação gráfica das precipitações pluviométricas mensais, em milímetros d'água, no município de Lavras - MG, no período de 1966 a 1996.

Na Tabela 2, tem-se a análise da variância (experimentos em blocos casualizados) para a série de precipitações, através da qual é possível verificar que o efeito dos anos (tendência) não foi significativo, ou seja, não existe tendência, o mesmo não acontece com o efeito dos meses, que foi altamente significativo, revelando a presença da sazonalidade.

Deve-se ressaltar que a análise de variância é um procedimento restrito, pois não há aleatorização, podendo não haver independência entre as observações.

TABELA 2 ANAVA (experimento em blocos casualizados) para a série de precipitações pluviiais mensais no município de Lavras - MG, no período de 1966 a 1996, em milímetros d'água.

Causa de Variação	GL	S.Q	Q.M	F	Pr > F
Anos (tendência)	30	219303,61995	7310,12067	1,238	0,1869
Meses (sazonalidade)	11	3394826,90029	308620,62730	52,278	0,0000
Erro	330	1948137,63555	5903,44738		
Total	371	5562268,15579			

A análise de variância relativa ao teste de tendência e sazonalidade revelou um F significativo (Tabela 3). Os coeficientes α_j não são todos nulos (Tabela 4); portanto, rejeita-se a hipótese $H_1: \alpha_1 = \dots = \alpha_{12} = 0$, ou seja, aceita-se a existência da sazonalidade determinística na série de precipitações pluviiais. Por outro lado, o coeficiente para a tendência β_1 é não-significativo, indicando a não-existência de tendência na série.

TABELA 3 ANAVA para o modelo de regressão ajustado para a série de precipitações pluviiais mensais no município de Lavras -MG, no período de 1966 a 1996, em milímetros d'água.

C. V.	GL	S.Q	Q.M	F	Pr > F
Modelo	12	3397126,7165	283093,8930	46,940	0,0001
Erro	359	2165141,4393	6031,0346		
Total	371	5562268,1558			

TABELA 4 Estimativas dos parâmetros, erro padrão, teste t e respectivas probabilidades obtidas para o modelo de regressão para a série de precipitações pluviais mensais no município de Lavras - MG, no período de 1966 a 1996, em milímetros d'água.

	GL	Estimativa	Erro padrão	T	Prob > T
β_0	1	123,623380	8,07233636	15,314	0,0001
β_1	1	0,023166	0,03751442	0,618	0,5373
α_1	1	155,131982	13,35588327	11,615	0,0001
α_2	1	76,005591	13,35535640	5,691	0,0001
α_3	1	41,405005	13,35493489	3,100	0,0021
α_4	1	-58,611709	13,35461875	-4,389	0,0001
α_5	1	-78,492939	13,35440798	-5,878	0,0001
α_6	1	-101,645137	13,35430260	-7,611	0,0001
α_7	1	-107,736045	13,35430260	-8,068	0,0001
α_8	1	-106,197921	13,35440798	-7,952	0,0001
α_9	1	-54,043667	13,35461875	-4,047	0,0001
α_{10}	1	-6,431349	13,35493489	-0,482	0,6304
α_{11}	1	80,077743	13,35535640	5,996	0,0001

Através da análise da função de autocorrelação da série de precipitações (Figura 3), percebe-se facilmente que existem picos nos “lags” múltiplos de 12, o que é indicativo de uma série com sazonalidade.

Assim, a partir desses resultados, conclui-se que o efeito dos anos foi não-significativo, indicando a ausência de tendência, enquanto ocorreu contrário com os meses, notando a presença da sazonalidade.

Através do periodograma da série de precipitações (Figura 4), é possível notar a existência de apenas um período, e conforme Tabela 5, observa-se que o maior componente da série de precipitações pluviais possui um período de 12 meses, sendo os demais períodos observados não-significativos segundo a estatística g de Fisher. Portanto, a série apresenta um período de ordem 12, que é igual ao componente sazonal identificado anteriormente.

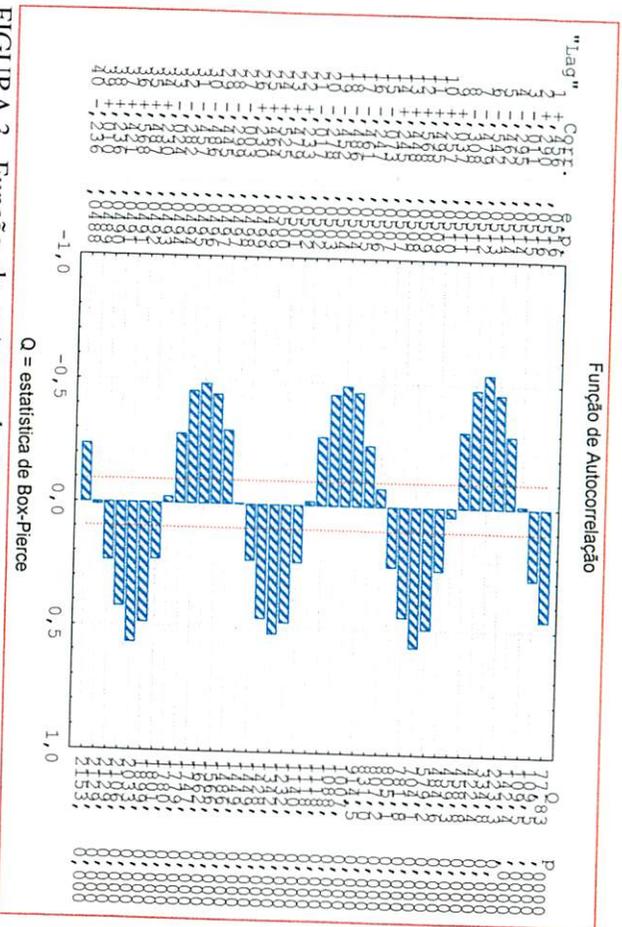


FIGURA 3 Função de autocorrelação para a série de precipitações pluviiais mensais no município de Lavras - MG, no período de 1966 a 1996 em milímetros d'água.

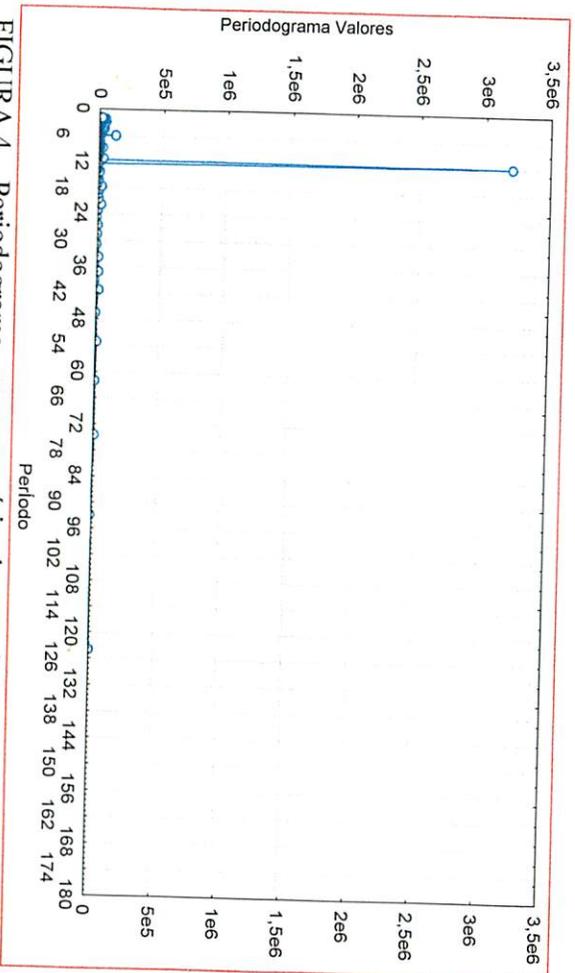


FIGURA 4 Periodograma para a série de precipitações pluviiais mensais no município de Lavras- MG, no período de 1966 a 1996, em milímetros d'água.

TABELA 5 Análise Spectral (Periodograma) para a série de precipitações pluviiais mensais no município de Lavras - MG, no período de 1966 a 1996, em milímetros d'água.

Frequência	Período	Coseno	Senso	Periodograma	Teste g de Fisher ¹	
31	0,0833	12	129,12	-24,48	3212434,56	1,3604 *
62	0,1667	6	22,05	-14,91	131814,64	0,0236 n/s
127	0,3414	2,93	-17,40	4,26	59660,78	0,0107 n/s
163	0,4382	2,28	17,67	-0,46	58086,10	0,0104 n/s
84	0,2258	4,43	-9,34	-12,78	46589,60	0,0084 n/s
164	0,4409	2,27	-8,14	13,18	44599,61	0,0080 n/s
167	0,4489	2,23	14,21	-5,83	43874,11	0,0079 n/s
33	0,0887	11,27	14,71	-2,97	41873,86	0,0075 n/s
160	0,4301	2,33	6,67	13,25	40954,44	0,0073 n/s
158	0,4247	2,35	-14,36	2,99	40000,83	0,0072 n/s

1- A distribuição g ao nível de significância de 5% é 0,04347

4.2 Ajuste de Modelos

4.2.1 Modelo de Regressão

Como a tendência foi considerada não-significativa, considerou-se

o modelo (7) como:

$$Z_t = \beta_0 + \sum_{j=1}^{11} \alpha_j D_{jt} + a_t$$

no qual β_0 é a média do processo. Obtem-se então:

$$\hat{\beta} = \beta_0 = [127,9438] \quad \hat{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9 \\ \alpha_{10} \\ \alpha_{11} \\ \alpha_{12} \end{bmatrix} = \begin{bmatrix} 148,279 \\ 75,796 \\ 40,321 \\ -56,400 \\ -76,522 \\ -100,102 \\ -106,256 \\ -104,867 \\ -55,284 \\ -4,019 \\ 76,165 \\ 162,890 \end{bmatrix} \quad (69)$$

sendo α_{12} obtido de $\sum_{i=1}^{12} \alpha_i = 0$.

Na Figura 5, têm-se a série de precipitações pluviais e a série sazonalmente ajustada através do método de regressão. Já na figura 6, têm-se os resíduos do ajuste sazonal pelo método da regressão. Observa-se que as precipitações pluviais são razoavelmente ajustadas pelo modelo de regressão acompanhado o movimento sazonal da série; porém, não é capaz de acompanhar os picos mais altos de precipitações.

Através da inspeção da função de autocorrelação do resíduo do ajuste da série, pelo método de regressão (Figura 7), verifica-se, mais uma vez, um rápido declínio do coeficiente de autocorrelação, apresentando pequenas flutuações em torno do ponto de estabilidade zero. A rápida queda da f.a.c. indica que a série é estacionária, e as pequenas variações em torno de zero indicam a eliminação dos picos sazonais. Assim, a nova série é estacionária, não possuindo sazonalidade.

Os valores destas autocorrelações ainda não são significativamente diferentes de zero, pois se encontram fora da faixa de $\pm 2 \frac{1}{\sqrt{372}} = \pm 0,104$, e a estatística de Q de Box-Pierce, conforme definida pela expressão (42), apresenta-se significativa ao longo da função de autocorrelação, indicando que o resíduo não é um ruído branco.

Assim, a análise das funções de autocorrelação e autocorrelação parcial parece indicar que os resíduos assumem um modelo ARMA, conforme definido no item 2.2.5.

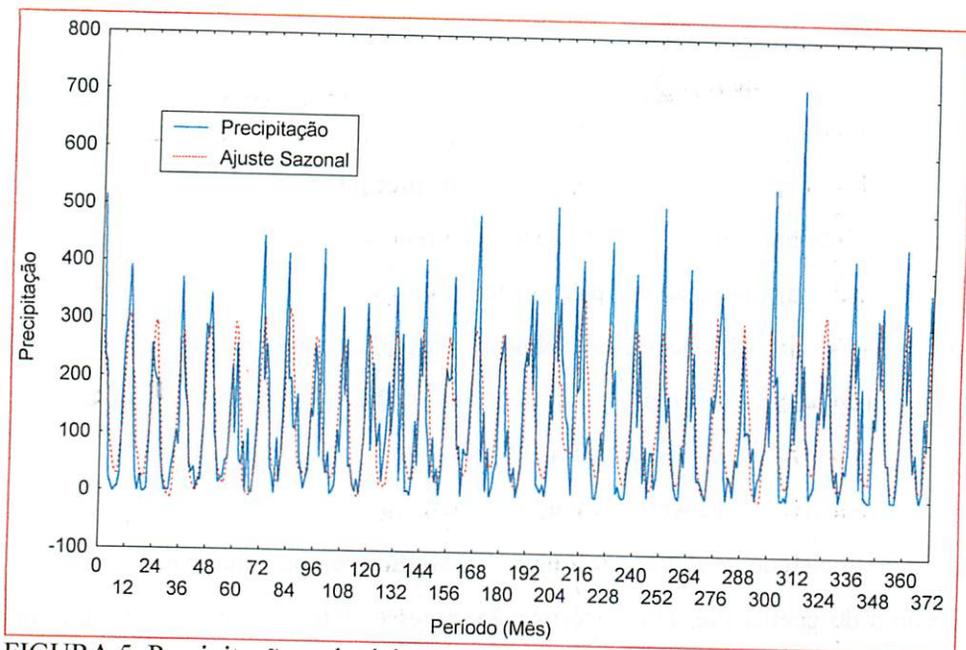


FIGURA 5 Precipitações pluviiais mensais no município de Lavras-MG no período de 1966 a 1996 em milímetros d'água e série sazonalmente ajustada.

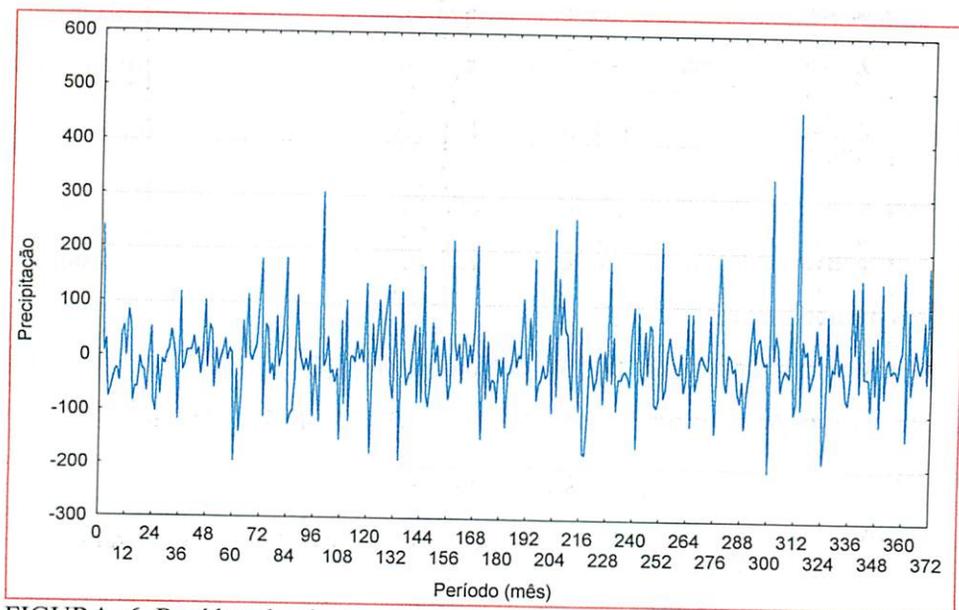
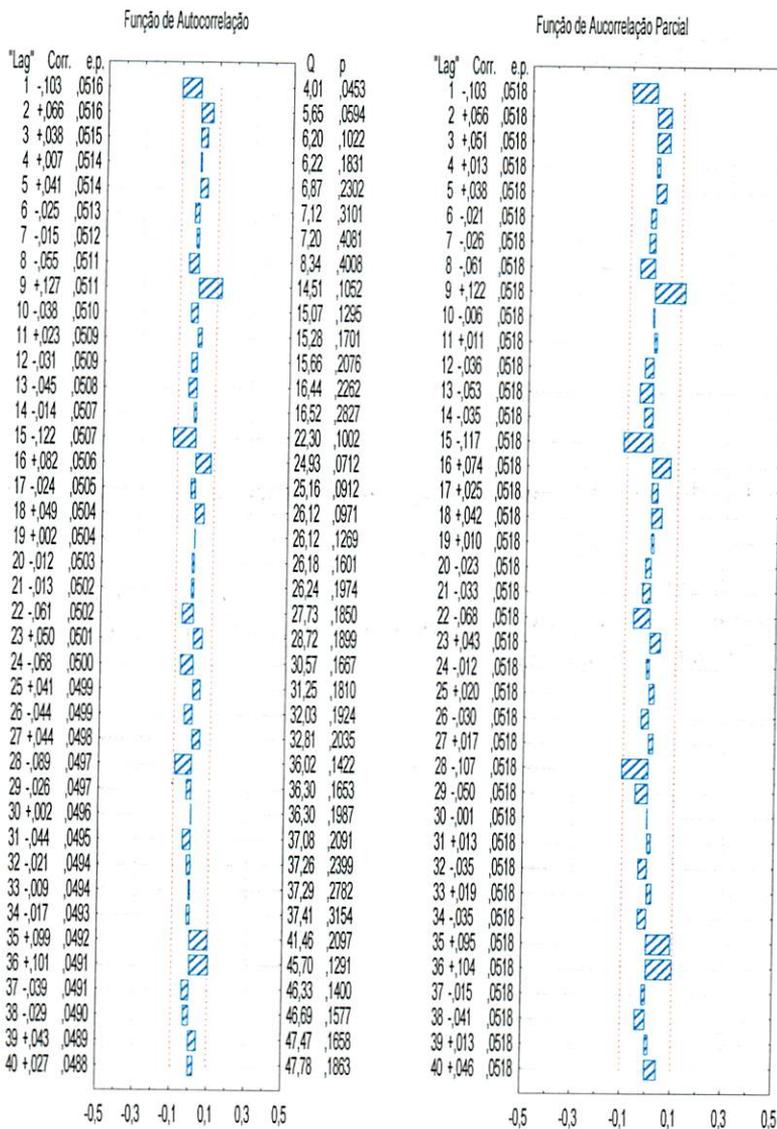


FIGURA 6 Resíduo do ajuste sazonal da série de precipitações pluviiais mensais no município de Lavras - MG.



Q = estatística Q de Box-Pierce

FIGURA 7 Função de Autocorrelação e autocorrelação parcial do resíduo do ajuste sazonal.

Como a série apresenta grande aleatoriedade, procurou-se incluir parâmetros um a um nos modelos testados. Alternativamente, também foram propostos modelos auto-regressivos e de média móvel. Na Tabela B.2 do anexo, está apresentado o resumo das principais informações de alguns dos modelos ajustados.

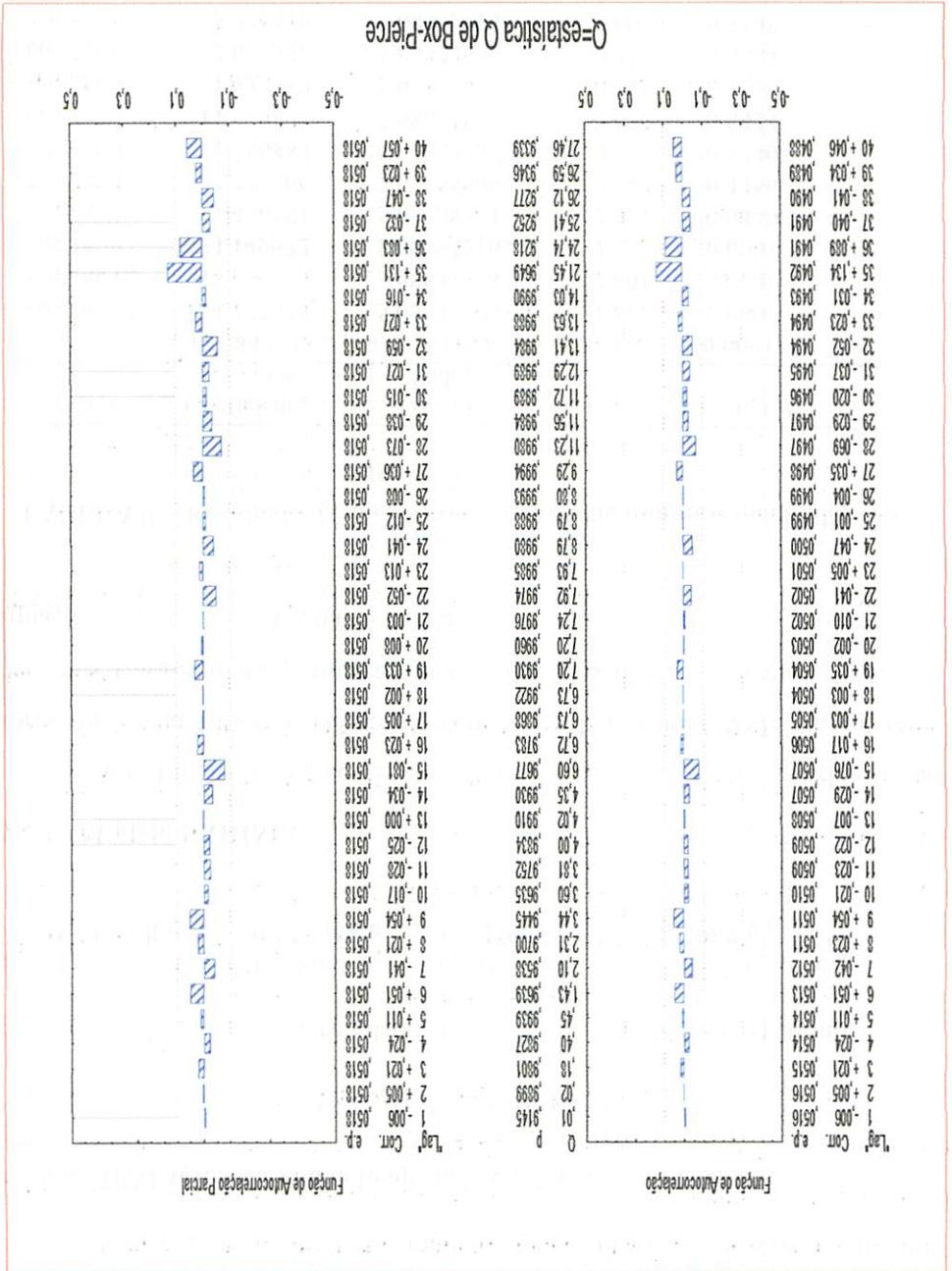
De uma maneira geral, todos os modelos apresentaram defasagens significativamente diferentes de zero. Porém, considerando como ruído branco os modelos com no máximo 5% das correlações fora do intervalo de ± 104 , foi possível identificar os modelos ARMA (2, 2), ARMA (3,3), ARMA (5, 5) e ARMA (5, 4). Dentre os quais, o de menor variância é o ARMA (4, 4), com todas as estimativas dos parâmetros significativas, conforme pode ser observado na Tabela 6.

TABELA 6 Modelo ARMA (4, 4) ajustado ao modelo de regressão -(variância residual = 5654,5).

	Param.	Erro Padrão	t(364)	p	L. Inferior 95% Conf	L. Superior 95% Conf
ϕ_1	-0,690165	0,085327	-8,0885	0,00000	-0,85796	-0,522369
ϕ_2	0,208658	0,080151	2,6033	0,00961	0,05104	0,366275
ϕ_3	-0,535788	0,088372	-6,0628	0,00000	-0,70957	-0,362003
ϕ_4	-0,842871	0,076643	-10,9974	0,00000	-0,99359	-0,692153
θ_1	-0,599008	0,080094	-7,4788	0,00000	-0,75651	-0,441503
θ_2	0,215448	0,054825	3,9297	0,00010	0,10763	0,323261
θ_3	-0,641905	0,064143	-10,0074	0,00000	-0,76804	-0,515767
θ_4	-0,863963	0,071704	-12,0490	0,00000	-1,00497	-0,722957

Na Figura 8, têm-se as funções de autocorrelação e autocorrelação parcial para o modelo ARMA (4, 4), que apresentam-se com mais de 95% das correlações dentro da faixa $[- 0,104; 0,104]$, e como a estatística Q de Box-Pierce para as vinte primeiras autocorrelações é 7,20 , sendo este valor inferior ao $\chi^2_{12} (0,05) = 21,03$, pode-se concluir que o resíduo é um ruído branco.

FIGURA 8 Função de autocorrelação e autocorrelação parcial para o modelo ARMA (4,4), ajustado ao resíduo do ajuste modelo de regressão.



Portanto, a escolha do modelo mais adequado à série recai no modelo ARMA(4, 4). O modelo ajustado é fornecido a seguir

$$Z_t = \beta_0 + \sum_{j=1}^{11} \alpha_j D_{jt} + a_t$$

em que α_j e β_0 , $j = 1, \dots, 12$ são dados por (69) e $a_t \sim \text{ARMA}(4, 4)$, assim:

$$a_t = -0,690165 a_{t-1} + 0,208658 a_{t-2} - 0,535788 a_{t-3} - 0,842871 a_{t-4} + e_t - 0,599008 e_{t-1} + 0,215448 e_{t-2} - 0,641905 e_{t-3} - 0,863963 e_{t-4}.$$

4.2.3 - Modelo SARIMA

As estimativas dos coeficientes e seus respectivos desvios padrão estão apresentados na Tabela 7. De acordo com Morettin e Tolo (1987), são menores que seus valores reais, porque a hipótese de erros não autocorrelacionados na equação $Z_t = \mu_t + N_t$, não é válida.

TABELA 7 Resultado da análise de regressão de mínimos quadrados do modelo $Z_t = \mu_t + N_t$.

Variável	Coefficiente Estimado	Desvio Padrão	T para H0	Prob > T
μ	127.843543	4.02331169	31.776.	0.0001
cos $2\pi t/12$	124.132618	5.68846200	21.822	0.0001
sen $2\pi t/12$	43.236056	5.68841308	7.601	0.0001
cos $2\pi t/6$	24.149622	5.69094216	4.244	0.0001
sen $2\pi t/6$	11.498581	5.68908481	2.021	0.0440
cos $2\pi t/4$	4.348766	5.69555006	0.764	0.4456
sen $2\pi t/4$	1.266884	5.69089487	0.223	0.8240
cos $2\pi t/3$	11.529823	5.68851804	2.027	0.0434
sen $2\pi t/3$	1.622517	5.68847922	0.285	0.7756
cos $5\pi t/6$	2.677676	5.69324656	0.470	0.6384
sen $5\pi t/6$	5.726516	5.68902674	-1.007	0.3148
cos πt	-8.369093	4.57219308	-1.830	0.0680
sen πt	-9.397518E13	8.7929206E13	-1.069	0.2859

Analisando a Tabela 7, observa-se que apenas as variáveis $\cos 2\pi t/12$, $\sin 2\pi t/12$, $\cos 2\pi t/6$ e $\sin 2\pi t/6$ são significativamente diferentes de zero, ao nível de 5%. Assim, reestimado o modelo apenas com as variáveis significativas, obtém-se o seguinte modelo μ_t :

$$\hat{\mu}_t = 127,94 + 124,08 \cos 2\pi t/12 + 43,28 \sin 2\pi t/12 + 23,96 \cos 2\pi t/6 + 11,60 \sin 2\pi t/6 + 11,60 \cos 2\pi t/3. \quad (70)$$

Na Figura 9, é mostrada a série de precipitações pluviais e o ajuste do modelo $\hat{\mu}_t$, e na Figura 10, tem-se o gráfico para os resíduos $\hat{N}_t = Z_t - \hat{\mu}_t$. Observa-se que as precipitações pluviais são razoavelmente ajustadas pelo modelo $\hat{\mu}_t$, acompanhando o movimento sazonal da série; porém, em alguns períodos, por exemplo 98, 168 e 312, os resíduos foram bastante discrepantes, como também, os picos das precipitações foram altos.

Através da inspeção da função de autocorrelação do resíduo $\hat{N}_t = Z_t - \hat{\mu}_t$ (Figura 11), verifica-se, mais uma vez, um rápido declínio do coeficiente de autocorrelação, apresentando pequenas flutuações em torno do ponto de estabilidade zero, indicando a eliminação de picos sazonais. Porém, os valores destas autocorrelações ainda não são significativamente diferentes de zero (mais de 5 % fora da faixa ± 104) e, a estatística Q de Box-Pierce apresenta-se significativa ao longo da função de autocorrelação, indicando que o resíduo não é um ruído branco.

As funções de autocorrelação e autocorrelação parcial dos resíduos sugerem um ARMA. Na Tabela B.3 do anexo estão apresentados alguns dos modelos propostos.

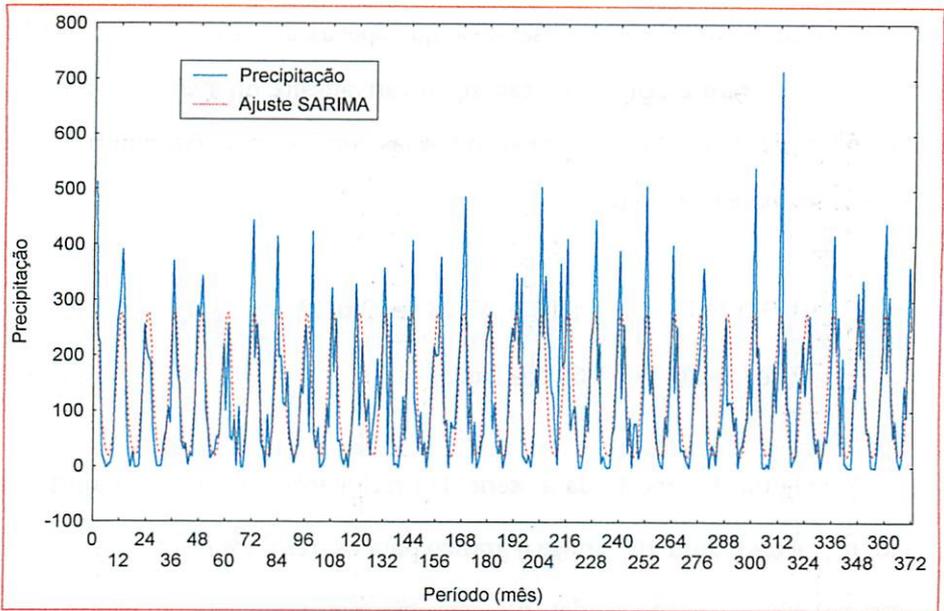


FIGURA 9 O modelo $\hat{\mu}_t$, e precipitações pluviais mensais no município de Lavras - MG, no período 1966 a 1996, em milímetros d'água.

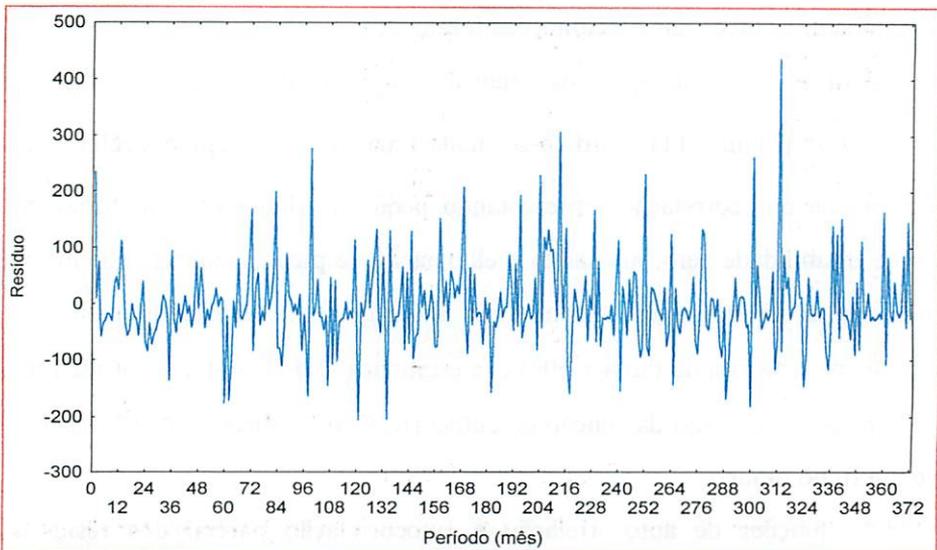
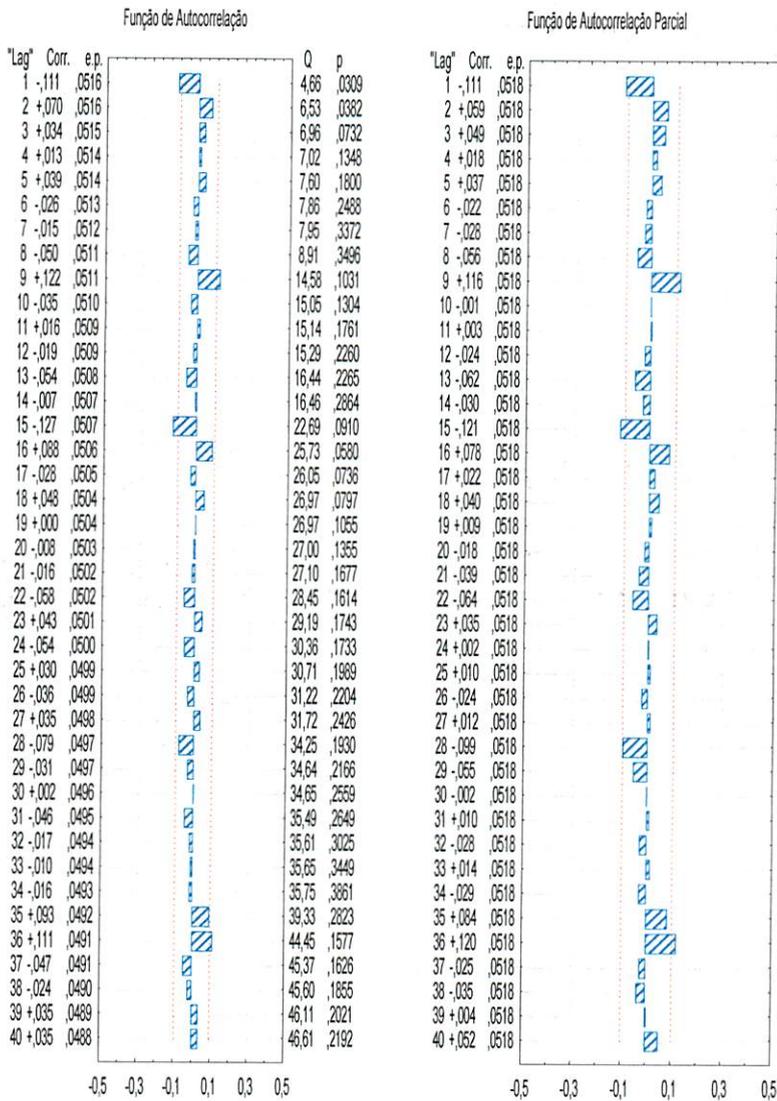


FIGURA 10 Resíduo do ajuste do modelo $\hat{\mu}_t$, da série de precipitações pluviais mensais no município de Lavras - MG.



Q= estatística Q de Box-Pierce

Figura 11 Função de autocorrelação e autocorrelação parcial do resíduo do ajuste modelo $\hat{\mu}_t$.

Porém, o modelo que se mostrou mais adequado foi o ARMA(4, 4) conforme tabela 8, por apresentar as funções de autocorrelações (figura 12) residuais dentro do intervalo de dois desvios padrões e com estatística Q de Box-Pierce para as vinte primeiras autocorrelações igual a 7,22, sendo este valor inferior ao $\chi^2_{16}(0,05)=21,03$, indicando que o resíduo é um ruído branco. Assim o resíduo do ajuste do modelo ARMA (4, 4) é um processo de ruído branco.

TABELA 8 Modelo ARMA (4, 4) ajustado a \hat{N}_t - resultados obtidos (variância residual = 5711,9).

	Parâmetro	Erro Padrão	t(363)	P	L. Inferior 95% Conf	L. Superior 95% Conf
ϕ_1	-0,6897	0,0934	-7,3821	0,0000	-0,8735	-0,5060
ϕ_2	0,2038	0,0823	2,4751	0,0138	0,0419	0,3657
ϕ_3	-0,5309	0,0920	-5,7738	0,0000	-0,7117	-0,3501
ϕ_4	-0,8357	0,0810	-10,3203	0,0000	-0,9949	-0,6764
θ_1	-0,5948	0,0894	-6,6559	0,0000	-0,7706	-0,4191
θ_2	0,2116	0,0557	3,7994	0,0002	0,1021	0,3211
θ_3	-0,6387	0,0662	-9,6411	0,0000	-0,7689	-0,5084
θ_4	-0,8581	0,0766	-11,2012	0,0000	-1,0088	-0,7075

O modelo ajustado é fornecido a seguir

$$Z_t = \mu_t + N_t \quad \text{EST}$$

Ex 61

em que μ_t é dado por (72) e $N_t \sim \text{ARMA}(4, 4)$, assim:

$$N_t = -0,6897 N_{t-1} + 0,2038 N_{t-2} - 0,5309 N_{t-3} - 0,8357 N_{t-4} + e_t - 0,5948 e_{t-1} + 0,2116 e_{t-2} - 0,6387 e_{t-3} - 0,8581 e_{t-4}.$$

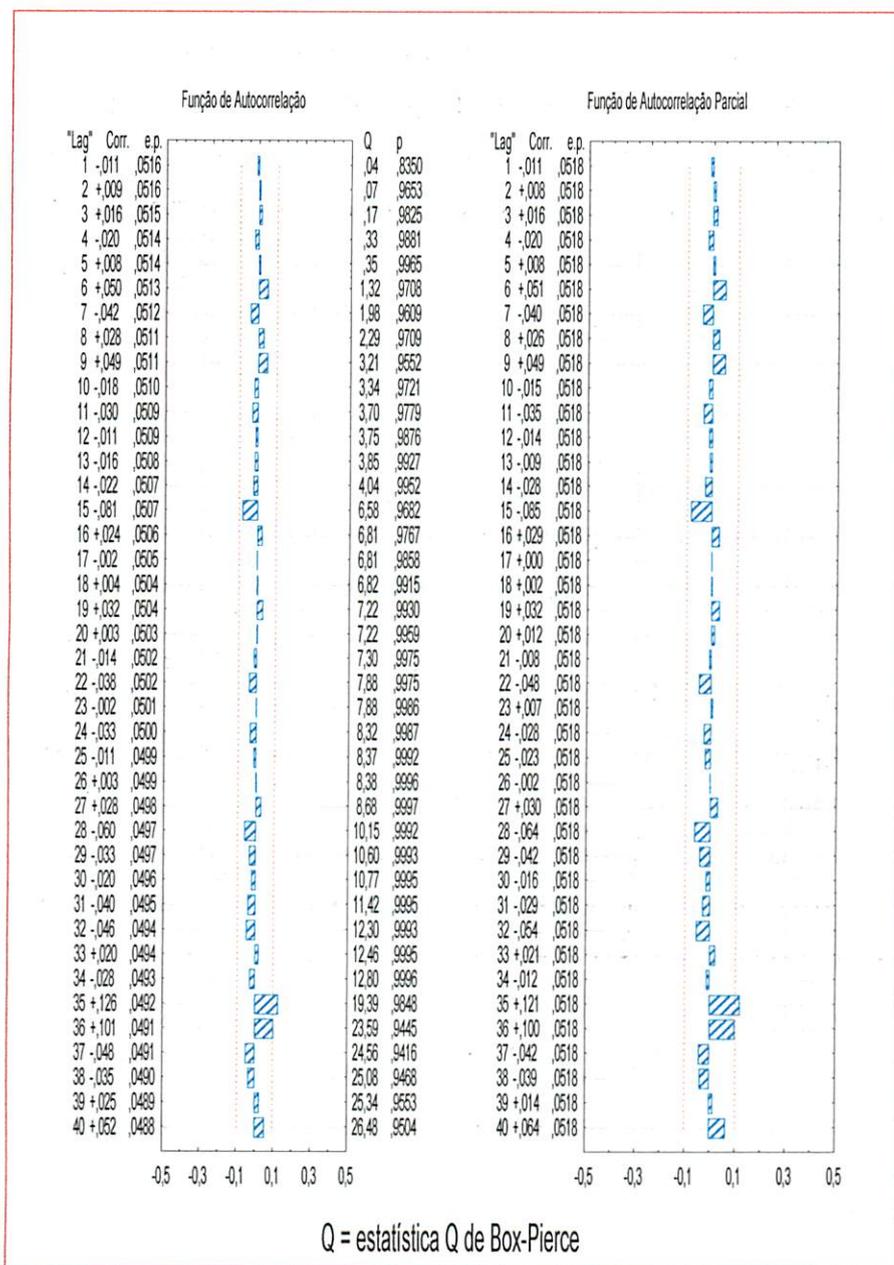


FIGURA 12 Função de autocorrelação e autocorrelação parcial para o modelo ARMA(4, 4) ajustado ao resíduo do ajuste do modelo \hat{N}_t .

4.3 Previsão

A obtenção de valores previstos para um horizonte previamente estabelecido é uma das finalidades da utilização de modelos de séries temporais. Assim, os testes de desempenho do modelo fora do período amostral podem ser decisivos na escolha do melhor modelo. Para tanto, utilizou-se o período final da série não utilizada na etapa de estimação dos modelos como horizonte de previsão, ou seja, o período de janeiro de 1997 a dezembro de 1997.

Inicialmente, procurou-se ajustar um modelo de suavização exponencial conforme definido no item 2.3.2, como forma alternativa de obter os previsões. Considerando as características da série, resolveu-se aplicar o método de Alisamento Exponencial Sazonal de Holt-Winters à série de precipitações, que é periódica com sazonalidade aditiva $s = 12$. As estimativas obtidas por esse procedimento podem ser visualizadas na Figura 13.

Os modelos selecionados anteriormente como de melhor ajuste, foram testados para obter valores previstos para 12 períodos adiante, a partir da observação 372 (dezembro de 1996).

A Figura 14 sintetiza os resultados das previsões para os modelos propostos, na forma gráfica, na qual é possível notar que as diferenças dos valores previstos entre os diversos modelos se apresentam bastante pequenas, e de uma maneira geral, tendem a obter estimativas superiores às precipitações observadas.

Na Tabela 9 estão apresentadas as precipitações observadas e as previsões para os modelos propostos ou seja, o Alisamento Exponencial, o Modelo de Regressão e o SARIMA.

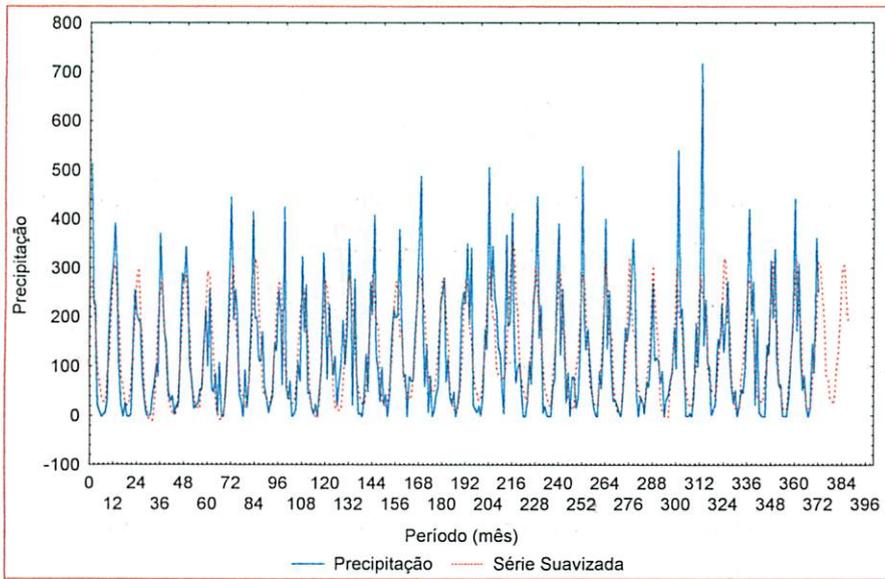


FIGURA 13 Precipitação observada e Alisamento Exponencial Sazonal de Holt-Winters.

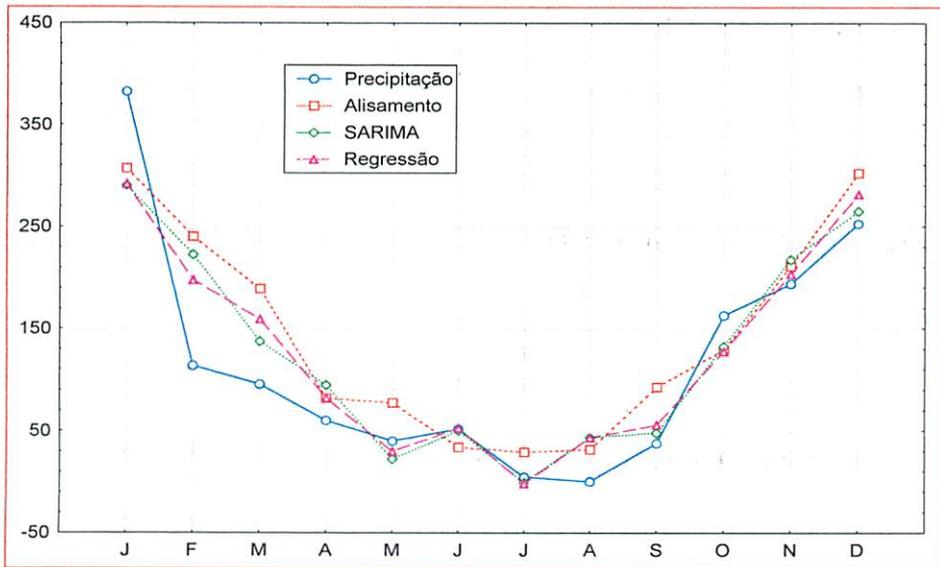


FIGURA 14 Precipitação observada e previsões segundo o modelo de Alisamento Exponencial Sazonal de Holt-Winters, modelo de Regressão e o Modelo SARIMA - jan/97 a dez/97.

TABELA 9 Precipitação observada e previsões segundo o Modelo de Alisamento Exponencial Sazonal de Holt e Winters, Modelo de Regressão e o Modelo SARIMA - jan/97 a dez/97.

Período (mês)	Precipitação Observada	Previsão Alisamento	Previsão Modelo Regressão	Previsão Modelo SARIMA
Jan/97	383,300	307,9093	292,559	291,271
Fev/97	114,500	241,162	198,477	223,612
Mar/97	96,500	190,1609	160,172	138,594
Abril/97	61,100	82,6487	83,192	95,508
Mai/97	41,000	78,4694	31,054	23,355
Jun/97	52,600	34,8637	53,526	51,217
Jul/97	5,600	30,1467	-,865	0,229
Ago/97	1,200	32,8096	44,481	44,679
Set/97	38,800	93,6242	57,184	48,790
Out/97	164,100	130,3049	128,886	133,871
Nov/97	194,800	212,2097	204,275	218,546
Dez/97	253,600	302,9781	282,385	265,855

Para avaliar o desempenho dos modelos ajustados, foi calculada a soma dos quadrados dos erros de previsão para cada modelo. Os resultados encontram-se na Tabela 10, e mostram ser o modelo SARIMA o que forneceu as melhores previsões, vindo, em seguida, o modelo de Regressão, e por último, o modelo de Alisamento.

No geral, verificou-se que as hipóteses foram alcançadas. Os modelos propostos ajustaram-se bem à série de precipitações pluviiais de Lavras, captando bem o componente sazonal da série; porém, não sendo capaz de modelar os períodos com precipitações muito elevadas. Os modelos SARIMA e o de Regressão mostraram-se bastante semelhantes quanto ao ajuste e, principalmente, por obterem o ruído branco após o ajuste de um ARMA(4, 4).

Quanto à previsão, ficou claro a superioridade dos modelos de regressão e o SARIMA sobre o modelo de suavização, tendo em vista a existência na série de picos de precipitações muito elevados.

TABELA 10 Erro de previsão e soma dos quadrados dos erros de previsão segundo o modelo de Alisamento Exponencial Sazonal de Holt-Winters, Modelo de Regressão e o Modelo SARIMA - jan/97 a dez/97.

Período (mês)	Erro dePrevisão Modelo de Alisamento	Erro dePrevisão Modelo de Regressão	Erro dePrevisão Modelo SARIMA
Jan/97	75,39	90,74	84,60
Fev/97	-126,66	-83,98	-84,00
Mar/97	-93,66	-63,67	-65,50
Abril/97	-21,55	-22,09	-18,50
Mai/97	-37,47	9,95	10,20
Jun/97	17,74	-0,93	3,90
Jul/97	-24,55	6,47	3,10
Ago/97	-31,61	-43,28	-36,70
Set/97	-54,82	-18,38	-23,90
Out/97	33,80	35,21	40,80
Nov/97	-17,41	-9,47	-18,00
Dez/97	-49,38	-28,79	-23,10
SQ Erro	41173,07	24339,42	23414,87

De uma maneira geral, em climatologia, os fenômenos não apresentam um desenvolvimento rigoroso, ocorrendo frequentemente, flutuações extremas, porém; os resultados desta pesquisa demonstram que os modelos de séries temporais são adequados para prever séries de precipitações pluviais.

5 CONCLUSÕES

A partir das análises feitas neste trabalho, chegou-se às seguintes conclusões:

1. Existe apenas uma periodicidade na série, de ordem 12, que é correspondente a uma sazonalidade de 12 meses;
2. Os modelos de regressão e SARIMA, quando aplicados à série de precipitação, mostraram bom ajuste; porém, foram constatadas autocorrelações significativas nos resíduos;
3. O ruído branco é obtido após o ajuste do modelo misto ARMA (4, 4) ao resíduo dos modelos de regressão e SARIMA;
4. O modelo SARIMA foi o que forneceu as melhores previsões, vindo, em seguida o modelo de regressão, e por último, o modelo de alisamento sazonal aditivo de Holt e Winters;
5. Ficou claro, neste trabalho, que os modelos de séries temporais podem ser usados para prever as precipitações pluviais no município de Lavras -MG.

6. BIBLIOGRAFIA

- ALMEIDA, R. M. de. *Características Climatológicas do regime de chuvas em Minas Gerais*. Viçosa, 1995. 64p. (Mestrado - Universidade Federal de Viçosa).
- AMARAL, E. *Análise harmônica e previsão a longo prazo*. Anais da Reunião Internacional de Biometria. Piracicaba, p.232-271, 1975.
- ANSLEY, C. F., APIVEY, W. A. e WROBLESKI, W. J. "Comments" In *Seasonal Analysis of Economic Time Series*. U.S. Department of Commerce, Bureau of the Census. 1979.
- ASSIS, F. N. de. *Modelagem da ocorrência e da quantidade de dias secos em Piracicaba-SP e Pelotas-RS*. Piracicaba: ESALQ / USP, 1991. 134p. (Tese Doutorado - Estatística e Experimentação Agronômica).
- BARGER, G. L.; THOM, H. C. S. Evaluation of drought hazard. *Agronomy Journal*, Madison: v. 41, n.11, p. 519-526, 1949.
- BOTELHO, V. A.V. A. *Ajuste da distribuição gama aos dados de precipitação pluvial do município de Lavras - MG*. Lavras: UFLA, 1998 (Dissertação - Mestrado em Estatística e Experimentação Agropecuária).
- BOX, G. E. ; JENKINS, G. M; REINSEL, G. C. *Time series analysis: forecasting and control*. 3 ed. New Jersey: Prentice Hall, 1994. 598p.
- BOX, G. E. P. e PIERCE, D. A. Distribution of residual auto-correlations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, v.65, 1509-1529, 1970.
- BRASIL, Ministério da Agricultura. *Normas climatológicas*. Rio de Janeiro: MA, 1961-1990. Brasília: 1992. 84p.

- BUSTOS, O. Algumas idéias de robustez aplicada à estimação paramétrica em séries temporais. Campinas. VII Simpósio Nacional de Probabilidade e Estatística Universidade Federal do Rio de Janeiro, 1986.
- CASTRO NETO, P. ; SEDIYAMA, J. V.; VILLELA. E.A. Probabilidade de ocorrência de períodos secos em Lavras, Minas Gerais. I. *Ciência e Prática*, Lavras, v.4, n.1, p.46-56, 1980a.
- CASTRO NETO, P. ; SEDIYAMA, J. V.; VILLELA. E. A. Probabilidade de ocorrência de períodos chuvoso em Lavras, Minas Gerais. I. *Ciência e Prática*, Lavras, v4, n.1, p.56-65, 1980b.
- CASTRO NETO, P. ; SILVEIRA, J. V. Precipitação provável para Lavras, Região Sul de Minas Gerais, baseada na função de probabilidade gama. I. Períodos mensais. *Ciência e Prática*, Lavras, v4, n.1, p144-151. 1981.
- COUTO, M. T. *Modelos estruturais de séries temporais: o caso dos preços da bovinocultura de corte do Estado de São Paulo*. Piracicaba: ESALQ/USP, 1996. 96p. (Dissertação de Mestrado em Economia Agrária).
- CRAMER, D. M. e LEADBETTER, E. F. Stationary and related stochastic process: Sample function properties and their applications. New York:John Wiley, 1966.
- JENKINS, G. M. e WATTS, D. G. *Spectral Analysis and its applications*. San Francisco: Holden-Day, 1968.
- GARBER, R. Análise de séries temporais. *Oecologia Brasiliensis*. Rio de Janeiro: UFRJ, v 2., 1995.
- GARCIA, E. A.C. Os preços da pecuária bovina do pantanal Matogrossense. *Pesquisa Agropecuária Brasileira*. Brasília, v.19, n.2. p.123-148, 1984.

- GONÇALVES, E. B. *Comparação de modelos para previsão de séries temporais - série de precipitações pluviiais mensais no Município de Pindorama-SP*. Piracicaba:ESALQ/USP, 1983.(Dissertação de Mestrado em Estatística e Experimentação Agronômica).
- HOFFMANN, R. e VIEIRA, S. *Análise de regressão: uma introdução à econometria*. 2 ed. São Paulo: Hucitec, 1975. 379p.
- KASSOUF, A.L. *Previsões de preços na pecuária de corte do estado de São Paulo*. Piracicaba:ESALQ/USP, 1988. (Mestrado em Economia Rural)
- MORETTIN, P.A. Periodicidades aparentes na série de precipitações atmosféricas de Fortaleza., Ceará. 6^o Simpósio Nacional de Probabilidade e Estatística Universidade Federal do Rio de Janeiro, 1984.
- MORETTIN, P.A. e TOLOI, C.M.C. *Séries temporais*. São Paulo: Atual, 1986. 136p.
- *Previsão de séries temporais*. 2 ed. São Paulo: Atual, 1987. 439p.
- *Modelos de funções de transferência*. São Paulo. 3^a Escola de Séries Temporais e Econometria. USP, 1989.
- NOBRE, C. A.; YANASSE, H. H.; YANASSE, C. C. F. *Previsão de secas no Nordeste pelo método das periodicidades: usos e abusos*. 6^o Simpósio Nacional de Probabilidade e Estatística. Universidade Federal do Rio de Janeiro, 1984.
- OZAKI, T. *On the order determination of ARIMA models*, Journal of the Royal Statistical Society, Serie C, Vol. 26, p.290-301. 1977.
- PEREIRA, B. B. *Séries Temporais Multivariadas*. 6^o Simpósio de Probabilidade e Estatística, Rio de Janeiro, 1984.

PICCININI, M. R. D. Distribuição de probabilidade de precipitação de intensidade máxima para Piracicaba, SP. Piracicaba, 1993.

PINO, F.A. e MORETTIN, P.A. *Intervention analysys applied to brasilian coffee and milk time series*. s.l., s. ed, 1981. 36p.

PINO, F.A. e ROCHA, M.B. *Transmissão de preços de soja no Brasil*. Revista de Economia e Sociedade Rural, Brasília, 32 (4): 345 - 361, out/dez. 1994.

PRIESTLEY, M. B. *Spectral analysis and time series. volume1: Univariate series; volume 2; Multivariate series, Prediction and control*. New York: Academic Press. 1981. 890p.

SAS INSTITUTE, INC. SAS user's guide statistics. 6.ed. SAS Institute, Inc. Cary, N. C.,1998.

STATISTICA for Windows. Realese 5. Copyright Stat Soft, Inc. 1984 -1995.

STEVENSON, W. J. *Estatística aplicada à administração*. São Paulo, Harbra, 1981

SHUMWAY, H. S. Applied statistical time series analysis. New Jersey. Prentice Hall, 1988.

VICENTE, J. R. *Modelos estruturais para previsão das produções brasileiras de carne de frango e ovos*. Revista de Economia e Sociedade Rural, Brasília, 30(4):305-319, out/dez.1992.

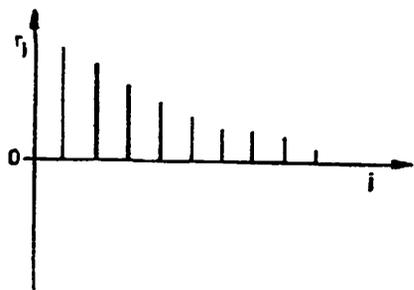
WOLD, H. *A large sample test for MA Processes*. Journal of The Roysl Statistics Societ., B, 11,237- 305. 1938.

ANEXOS

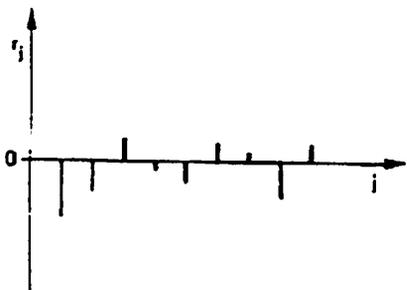
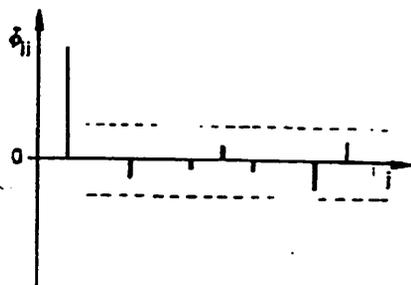
ANEXO A

Página

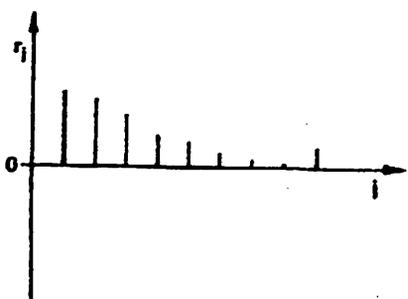
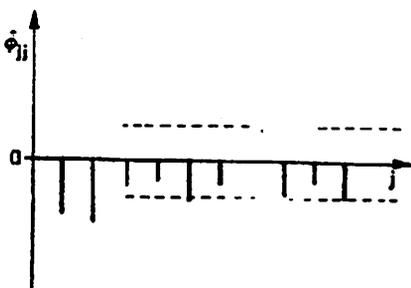
Figura A.1 Exemplos de função de autocorrelação e autocorrelação parcial para modelos AR(1), MA(1) e ARMA(1,1) Extraído de Morettin e Toloi(1985).....	92
--	----



AR(1)



MA(1)



ARMA(1, 1)

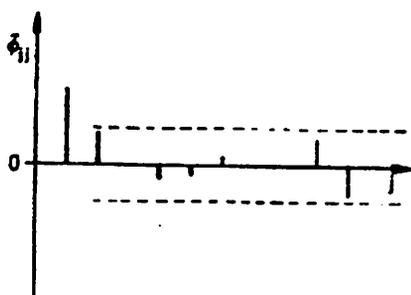


FIGURA A.1 Exemplos de função de autocorrelação e autocorrelação parcial para modelos AR(1), MA(1) e ARMA(1,1). Extraídos de Moretlin e Toloí(1985).

TABELA B.1	Série de Precipitações Pluviais Mensais do Município de Lavras-MG, Jan.,66 / Dez., 97.....	94
TABELA B.2	Principais modelos ajustados para o resíduo do método de regressão.....	96
TABELA B.3	Principais modelos ajustados para o resíduo do métodoSARIMA.....	97

TABELA B.1 Série de Precipitações Pluviais Mensais do Município de Lavras-MG, Jan.,66 / Dez., 97

Ano	Jan.	Fev.	Mar.	Abr.	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
→ 1966	515,00	237,10	224,20	24,40	11,60	0,00	4,90	8,70	32,30	168,00	265,10	302,70
→ 1967	393,00	283,70	115,20	26,10	0,60	28,40	0,00	0,00	5,100	123,00	257,60	210,00
→ 1968	196,00	191,50	76,40	28,60	2,00	2,00	3,30	42,80	66,80	108,00	81,000	372,00
→ 1969	254,00	175,30	147,60	50,00	33,20	43,00	4,70	24,50	20,20	115,00	291,50	269,40
→ 1970	345,00	255,50	106,00	72,20	16,80	23,60	26,10	55,40	57,40	144,00	222,80	102,40
→ 1971	259,00	56,10	49,70	87,30	2,70	108,90	0,60	0,00	48,90	131,00	294,60	446,40
→ 1972	197,00	258,80	204,00	45,20	32,60	0,00	94,40	18,00	72,80	188,00	417,10	199,30
→ 1973	202,00	116,50	111,70	172,30	60,40	35,00	8,20	27,20	42,20	149,00	133,70	257,40
→ 1974	212,00	64,10	426,20	62,00	35,40	72,20	0,00	4,500	15,20	111,00	71,900	324,50
→ 1975	173,00	268,60	47,20	49,80	17,80	5,10	25,40	0,00	50,00	107,00	331,70	225,40
→ 1976	76,00	228,20	126,90	84,30	123,40	22,40	68,60	107,00	196,0	105,00	166,40	360,90
1977	250,00	23,80	279,40	88,40	5,10	7,40	0,00	42,00	128,0	51,200	273,20	207,30
1978	410,00	129,20	90,10	30,60	100,20	23,20	46,10	0,20	43,00	156,00	218,60	201,20
1979	205,00	380,30	204,20	80,40	86,00	0,00	81,80	71,30	73,00	166,00	235,60	349,60
1980	489,00	201,80	61,00	147,60	8,80	82,60	0,00	13,20	45,00	60,600	229,00	250,50

CONTINUAÇÃO TABELA B I

	1981	282,00	70,60	117,20	41,40	25,40	50,60	0,00	22,00	57,00	220,00	253,20	230,60
	1982	352,00	187,60	343,20	23,00	16,00	8,50	21,80	2,00	44,00	179,00	136,00	507,20
	1983	238,40	346,50	246,00	215,80	140,80	127,40	54,60	6,00	369,0	184,00	193,60	413,60
	1984	172,00	68,60	101,20	109,70	46,60	0,00	0,00	33,40	113,0	66,500	232,80	267,90
	1985	448,00	160,60	225,10	8,30	21,20	2,80	0,00	2,00	60,70	76,000	262,60	392,00
	1986	126,00	258,70	147,70	28,10	89,90	1,20	80,40	81,10	17,60	37,600	140,20	509,20
	1987	205,00	135,50	179,00	109,20	56,60	20,60	4,00	0,00	91,90	57,400	169,00	402,30
26 -->	1988	138,00	255,60	111,70	31,60	33,80	23,20	9,50	0,00	53,20	181,00	153,80	189,50
	1989	270,00	361,20	272,80	45,20	0,00	40,50	31,40	6,20	71,40	59,400	123,40	271,30
->	1990	114,00	120,00	115,00	68,40	92,60	0,60	34,80	42,60	71,40	91,400	181,40	97,000
->	1991	543,00	203,20	218,90	101,40	2,00	0,00	7,40	0,00	46,60	190,00	101,50	214,90
	1992	718,00	144,00	237,80	111,00	93,90	3,50	14,10	25,40	158,0	144,00	230,00	131,60
	1993	194,00	274,90	134,20	62,70	28,00	54,00	0,00	21,60	55,40	47,800	102,00	229,20
	1994	421,00	211,90	273,80	23,80	198,30	10,00	3,50	0,00	0,000	146,00	127,40	316,80
	1995	200,00	339,50	124,80	64,60	65,60	1,20	1,00	0,00	38,60	116,00	192,00	442,10
	1996	174,00	310,30	129,10	54,10	84,60	17,00	,200	18,10	149,0	90,500	363,20	252,70
	1997	383,3	114,5	96,5	61,1	41	52,6	5,6	1,2	38,8	164,1	194,8	253,6

TABELA B. 2 - Principais modelos ajustados para o resíduo do método de regressão

Modelo Proposto	Variância	f.a.p. ≠ 0	f.a.c. ≠ 0	AIC	Estimativas não significantivas
AR(1)	5802,7	9 15 28 35	9 15 35 36	1406,07	0
MA(1)	5810,0	9 15 28 35	9 15 35 36	1406,27	1
ARMA(1, 1)	5809,3	9 15 28 35	9 15 35 36	1408,25	2
ARMA(2, 2)	5800,7	15 35	15 35	1412,01	0
ARMA(4, 4)	5664,5	35	35	1416,18	0
ARMA(5, 5)	5700,8	35	15 35	1421,21	5
ARMA(5, 4)	5789,6	35	35	1421,71	1

96

ARMA (3,3) → 6181,3
 ARMA (3,4) → 6232,8
 ARMA (4,3) → 6111,0

TABELA B3 - Principais modelos ajustados para o resíduo do método SARIMA

Modelo Proposto	Variância	f.a.c. $\neq 0$	f.a.p. $\neq 0$	AIC	Estimativas não significativas
AR (1)	5854,0	9 15 28 36	9 15 35 36	1407,49	0
MA (1)	5862,9	9 15 36	9 15 36	1407,74	0
AR (2)	5843,3	9 15 28 36	9 15 35 36	1409,20	1
MA (2)	5849,2	9 15 28 36	9 15 35 36	1409,36	1
ARMA (2, 2)	5878,2	9 15 28 36	9 15 35 36	1414,16	4
ARMA (4, 4)	5727,6	35	35	1417,97	0
ARMA (4, 3)	5813,1	15 35	15 35 36	1418,36	2
ARMA (5, 4)	5742,1	35	35	1420,37	1
ARMA (4, 5)	5742,1	35	35	1420,37	1
ARMA (5, 5)	5757,8	36	35 36	1422,82	5