Fundação de Economia e Estatística

séries univariantes de tempo – metodologia de Box & Jenkins

PORTO ALEGRE, RS - N.4 - AGOSTO 1982

SÉRIES UNIVARIANTES DE TEMPO - METODOLOGIA DE BOX & JENKINS

FUNDAÇÃO DE ECONOMIA E ESTATÍSTICA

séries univariantes de tempo – metodologia de Box & Jenkins

Sergio Fischer

PORTO ALEGRE, RS - N.4 - AGOSTO 1982

F529 Fischer, Sérgio

Séries univariantes de tempo — metodologia de Box & Jenkins. Porto Alegre, FEE, 1982.

186p. ilust. (Teses, 4)

1. Estatística. 2. Econometria. I. Título

CDU 311:330.115

Tiragem: 600 exemplares

Toda a correspondência para esta publicação deverá ser endereçada à:
FUNDAÇÃO DE ECONOMIA E ESTATÍSTICA
Rua Gen. Vitorino, 77 - 2º andar
90.000 - Porto Alegre-RS

Administração Amaral de Souza

SUMÁRIO

LISTA DE FIGURAS 5

LISTA DE QUADROS 8

APRESENTAÇÃO 11

1 – INTRODUÇÃO 13
1.1 – A Problemática 13
1.2 – A Importância das Séries Econômicas de Tempo 17
1.3 – O Conceito de Série Temporal 19
1.4 – O Significado da Análise de Séries Temporais 21

2 – SOBRE SÉRIES TEMPORAIS E PROCESSOS ESTOCÁSTICOS 23
2.1 – Processos Estocásticos e suas Propriedades 23
2.2 – Caracterização de um Processo Estocástico 24
2.3 – Processos Estacionários e Não Estacionários 25
2.4 – Propriedades dos Processos Estacionários 27
2.5 – Processos Evolutivos Homogêneos 28
2.6 – A Função de Autocorrelação 29

3 – OS MODELOS DE BOX & JENKINS 39
3.1 – Fundamentos 39
3.2 – Modelo Média Móvel 46
3.3 – Modelo Auto-Regressivo 51
3.4 – Modelo Auto-Regressivo – Média Móvel 57
3.5 – Modelo Auto-Regressivo Integrado – Média Móvel 62

4 – O PROCESSO DE MODELAGEM 69
4.1 – Modelagem de Séries Não Sazonais: Modelo ARIMA (p, d, q) . . . 69
4.2 – Modelagem de Séries Sazonais: Modelo SARIMA (p, d, q) x x (P, D, Q) 99

5 – O MÉTODO DE PREVISÃO DE BOX & JENKINS 101
5.1 – Introdução 101
5.2 – O Processo de Previsão 104
5.3 – A Atualização das Previsões 111

5.4 – Previsões por Intervalo 113
5.5 – Previsões para Séries Sazonais 114
5.6 – Previsão com Dados Transformados 115

6 – APLICAÇÕES PRÁTICAS DO MÉTODO DE BOX & JENKINS 117
6.1 – Nota Introdutória 117
6.2 – Previsões para o Índice Geral de Preços (Custo de Vida) para a Cidade de Porto Alegre 118
6.3 – Previsões para o Consumo de Energia Elétrica no Rio Grande do Sul 139

7 – CONSIDERAÇÕES FINAIS 159

ANEXO I – Manual dos Comandos Necessários para a Operação das Rotinas Computacionais 165

ANEXO II – Relatórios de Saída do Computador: Previsões para o Índice de Preços ao Consumidor, Usando Diversos Modelos Supostos na Fase de Identificação 175

ABSTRACT 181

BIBLIOGRAFIA 183

LISTA DE FIGURAS

CAPÍTULO 2

Figura 1 – Correlograma característico de um processo de ruído branco . . 32
Figura 2 – Correlograma característico de uma série não estacionária 34
Figura 3 – Correlograma característico de uma série estacionária 35
Figura 4 – Produção suína nos Estados Unidos (em milhões de cabeças por mês) – jan./62-dez./71 36
Figura 5 – Correlograma da série temporal da produção suína nos Estados Unidos, exposta na Figura 4 36
Figura 6 – Produção suína nos Estados Unidos – jan./62-dez./71. Função de autocorrelação de $(Y_t - Y_{t-12})$ 37
Figura 7 – Produção suína nos Estados Unidos – jan./62-dez./71. Função de autocorrelação de $\nabla(Y_t - Y_{t-12})$ 37

CAPÍTULO 3

Figura 1 – Diagrama de um modelo de filtro linear para séries temporais . . 45
Figura 2 – Diagrama do processo ARIMA para modelagem de séries temporais 45
Figura 3 – Regiões de admissão para que um MA(2) seja inversível 50
Figura 4 – Gráfico de uma série não estacionária quanto ao nível 64
Figura 5 – Gráfico de uma série não estacionária quanto ao nível e à inclinação 65

CAPÍTULO 4

Figura 1 – Gráfico "Amplitude x Média" para alguns valores possíveis de λ 71
Figura 2 – Correlograma de um processo não estacionário não sazonal 72

Figura 3 – Correlograma de um processo não estacionário sazonal 72
Figura 4 – Correlograma de um processo não estacionário sazonal 72

CAPÍTULO 6

Figura 1 – Gráfico da série IPC – jan./75-dez./81 119
Figura 2 – Gráfico "Amplitude x Média" para a série Índice de Preços ao Consumidor – jan./75-out./81 120
Figura 3 – Gráfico da série $Y_t = \ln X_t$, onde X_t é a série observada do IPC . 121
Figura 4 – Gráfico da função de autocorrelação da série Y_t 122
Figura 5 – Gráfico da função de autocorrelação da série ∇Y_t 123
Figura 6 – Gráfico da função de autocorrelação da série $\nabla^2 Y_t$ 124
Figura 7 – Gráfico da função de autocorrelação da série $\nabla^3 Y_t$ 125
Figura 8 – Gráfico da função de autocorrelação parcial da série $\nabla^2 Y_t$ 127
Figura 9 – Gráfico da função de autocorrelação da série de resíduos para o modelo $\nabla^2 \ln X_t = (1 - 0{,}79113B)a_t$ 135
Figura 10 – Gráfico da PACF da série de resíduos para o modelo $\nabla^2 \ln X_t = (1 - 0{,}79113B)a_t$ 135
Figura 11 – Gráfico da série de resíduos para o modelo $\nabla^2 \ln X_t = (1 - 0{,}79113B)a_t$ 136
Figura 12 – Relatório de saída do computador: gráfico dos valores previstos, para o ano de 1982, do IPC 137
Figura 13 – Gráfico da série consumo de energia elétrica no Rio Grande do Sul – jan./70-out./79 140
Figura 14 – Gráfico "Amplitude x Média" para a série consumo de energia elétrica no Rio Grande do Sul – jan./70-out./81 141
Figura 15 – Gráfico da série $T_t = \ln Z_t$, onde Z_t é a série observada do consumo de energia elétrica 142
Figura 16 – Gráfico da função de autocorrelação da série T_t 143
Figura 17 – Gráfico da função de autocorrelação da série ∇T_t 144
Figura 18 – Gráfico da função de autocorrelação da série $\nabla^2 T_t$ 145
Figura 19 – Gráfico da função de autocorrelação da série $\nabla\nabla_{12} T_t$ 146
Figura 20 – Gráfico da função de autocorrelação parcial da série ∇T_t 148
Figura 21 – Gráfico da ACF da série de resíduos para o modelo $(1 - 0{,}94055B^{12})\nabla \ln Z_t = (1 - 0{,}57944B)(1 - 0{,}63588B^{12})a_t$ 153

Figura 22 – Gráfico da PAC da série de resíduos para o modelo $(1-0{,}94055B^{12})\nabla \ln Z_t = (1-0{,}57944B)(1-0{,}63588B^{12})a_t$ 154

Figura 23 – Gráfico da série de resíduos para o modelo $(1-0{,}94055B^{12})\nabla \ln Z_t = (1-0{,}57944B)(1-0{,}63588B^{12})a_t$ 155

Figura 24 – Relatório de saída do computador: gráfico dos valores previstos para nov./81-out./82 do consumo de energia elétrica gaúcho . . . 157

LISTA DE QUADROS

CAPÍTULO 4

Quadro 1 – Comportamento das funções de autocorrelação e autocorrelação parcial para o processo ARIMA (p, d, q) de ordens mais comumente encontradas na prática 75

Quadro 2 – Obtenção dos valores de a_t em um modelo $W_t = (1 - 0{,}4B)a_t$. 90

Quadro 3 – Obtenção dos valores de a_t em um modelo $W_t = (1 - 0{,}4F)e_t$. 92

CAPÍTULO 6

Quadro 1 – Índice de Preços ao Consumidor (custo de vida) em Porto Alegre – jan./75-dez./81 118

Quadro 2 – Valores numéricos da ACF para as séries Y_t e ∇Y_t 125

Quadro 3 – Valores numéricos da ACF para as séries $\nabla^2 Y_t$ e $\nabla^3 Y_t$ 126

Quadro 4 – Valores numéricos da PACF para a série $\nabla^2 Y_t$ 127

Quadro 5 – Principais informações estatísticas dos modelos supostos para a série Índice de Preços ao Consumidor para a cidade de Porto Alegre 132

Quadro 6 – Relatório de saída do computador: previsão para os valores mensais do IPC – 1982 136

Quadro 7 – Previsão para os valores mensais do IPC – 1982 138

Quadro 8 – Consumo mensal de energia elétrica no Rio Grande do Sul – jan./70-out./81 139

Quadro 9 – Valores numéricos da ACF para as séries T_t e ∇T_t 147

Quadro 10 – Valores numéricos da ACF para as séries $\nabla^2 T_t$ e $\nabla\nabla_{12} T_t$ 147

Quadro 11 – Valores numéricos da PACF para as séries T_t e ∇T_t 149

Quadro 12 – Principais informações estatísticas dos modelos supostos para a série consumo de energia elétrica no Rio Grande do Sul – jan./70-out./81 151

Quadro 13 – Relatório de saída do computador: previsão para os valores mensais do consumo de energia elétrica no Rio Grande do Sul – nov./81-out./82 156
Quadro 14 – Previsão para os valores mensais do consumo de energia elétrica gaúcho – nov./81-out./82 157

APRESENTAÇÃO

Este trabalho de dissertação de mestrado consiste em uma análise da metodologia desenvolvida pelos professores George E. P. Box e Gwilym M. Jenkins para a análise de séries temporais univariantes e sua aplicação a variáveis da realidade econômica gaúcha.

A escolha desse tema surgiu de uma conversa com a colega Economista Sara Brumer que, após experiência profissional em Madrid, onde havia iniciado seu curso de doutorado em Economia, salientava a importância que a metodologia de Box & Jenkins tivera no estudo da Análise de Séries Temporais.

A curiosidade sobre a metodologia e a infrutífera busca de material bibliográfico no Rio Grande do Sul, que apresentasse o assunto de forma acessível e capaz de proporcionar condições de operacionalização, notadamente aos economistas pouco habilitados a um tratamento estatístico mais sofisticado dos dados da realidade econômica, levaram-me a desenvolver esse tema.

A idéia central que norteou o estudo foi expor o assunto com uma estrutura que, além de desenvolver a metodologia de Box & Jenkins propriamente dita, incluísse a análise dos pré-requisitos estatísticos, proporcionando, assim, condições para que o documento tivesse um detalhamento técnico que suprisse a deficiência bibliográfica encontrada e servisse de manual aos usuários da Análise de Séries Temporais.

Para a realização deste trabalho, contei com a colaboração de colegas e professores a quem sou grato. Devo um especial agradecimento à Fundação de Economia e Estatística que, dentro de uma política de incentivo ao desenvolvimento científico de sua equipe técnica, proporcionou as condições materiais para minha formação acadêmica a nível de mestrado.

A presente dissertação, orientada pelo Prof. Carlos Augusto Crusius, foi defendida no Centro de Estudos e Pesquisas Econômicas da Universidade Federal do Rio Grande do Sul (IEPE), para a obtenção do título de Mestre em Economia, no dia 12 de maio de 1982.

A banca examinadora foi composta pelos professores Herbert Guarini Calhau – Presidente da Banca –, Carlos Augusto Crusius e Emani Hickmann, todos do IEPE.

1 – INTRODUÇÃO

1.1 – A Problemática

O desenvolvimento científico do século XX tem mostrado profundas transformações no que tange aos métodos de análise das ciências sociais.

Nos séculos anteriores, as pesquisas nessas áreas caracterizavam-se, basicamente, pela acumulação, organização e classificação dos fatos observados em investigações empíricas. A pobreza conceitual não permitia a total compreensão dos movimentos das variáveis sociais, fato que resultava, comumente, em uma falta de clareza e precisão em suas análises.

A partir do final do século passado, uma verdadeira revolução no modo de pesquisa dessas ciências começou a tomar vulto. Iniciou-se uma etapa de conscientização de que o objetivo da pesquisa social não é simplesmente o acúmulo e a organização de informações, mas, principalmente, a sua compreensão.

A necessidade de entender os mecanismos envolvidos nas relações do mundo real induziu o pensamento científico a ressaltar o valor das teorias, notadamente aquelas formuladas com o auxílio da Matemática, e a busca de modelos teóricos específicos tornou-se elemento natural das pesquisas na análise social.

A necessidade de "modelização"[1] do mundo real tem estimulado, de forma intensa, o desenvolvimento de novos métodos, cada vez mais potentes, capazes de descrever com maior grau de adequação as inter-relações entre as variáveis sociais.

Na Ciência Econômica, a busca incessante de sistemas esquemáticos simplificados, capazes de simular e quantificar a complexidade das relações existentes na atividade econômica, tem sido uma das características mais marcantes em seu desenvolvimento neste século. A integração ocorrida entre essa Ciência e os instrumentos quantitativos proporcionou condições de confrontação de seu avanço teórico com os dados numéricos do mundo real, de tal modo que se pode considerar o avanço da Teoria Econômica do último século intimamente vinculado aos aperfeiçoamentos dos seus métodos quantitativos, onde o progresso de um é estímulo e necessidade exigida pelo outro.

O marco inicial da utilização de métodos numéricos na análise econômica pode ser situado no século passado[2], com as investigações teóricas

---

[1] Para fins deste trabalho, o conceito de modelo é o mesmo utilizado por ZEITOUN (1971): "Um modelo é uma representação de uma realidade provável ou concreta, equivalente em sua expressão a um conjunto de variáveis e de relações entre essas variáveis".

[2] Deve-se lembrar, entretanto, que já em épocas anteriores os métodos numéricos foram utilizados de forma a descrever os movimentos econômicos, embora sem uma elaboração teórica consistente como a que se verificou no século passado a partir dos trabalhos de Cournot e Walras. Os mercantilistas, no século XV, descreviam precariamente, de forma contábil, seus movimentos comerciais de além-mar. Posteriormente, no período fisiocrata, Quesnay, através de seu famoso "Tableau économique", proporciona o primeiro trabalho elaborado numericamente para descrever o funcionamento global de um sistema econômico.

de Cournot e Walras, entre outros[3], para assuntos microeconômicos.

Cournot, através das suas "Recherches sur les principes mathématiques de la théorie des richesses", de 1838, proporcionou um trabalho pioneiro que estimulou a elaboração de todo um novo tratamento analítico em Economia, não apenas em termos de teoria, mas também, e principalmente, em termos de método. Cournot foi o primeiro pensador a empregar o cálculo integral na análise econômica. Foi ele também o precursor dos métodos gráficos, utilizando diagramas cartesianos para expressar as relações entre variáveis econômicas[4].

Seu trabalho pioneiro significou o início da Economia Matemática, que teve seguidores em todo o mundo, passando a Matemática a ser vista como algo indispensável ao pensamento econômico, como um instrumento para a exposição de idéias e como um meio para expressar hipóteses para a análise das relações no sistema econômico.

Posteriormente, Walras, em Lausanne, com seu "Éléments d'économie politique pure", de 1874, sistematizou a análise matemática na Ciência Econômica, passando a ser, conjuntamente com Cournot, o fundador da escola matemática de economistas. A Universidade de Lausanne tornou-se conhecida no campo da Economia Matemática através dos trabalhos de Walras e seu mais próximo sucessor, Vilfredo Pareto. Este, em seu "Cours d'économie politique", sistematizou as leis da produção e da distribuição, atualmente conhecidas como "Leis de Pareto", sendo um dos primeiros (depois de Edgeworth) a trabalhar com curvas de indiferença para explicar a procura e eliminar a crítica, segundo a qual a utilidade marginal não pode ser medida quando comparada ao custo marginal.

Foram os economistas de Lausanne os grandes representantes do desenvolvimento do equilíbrio geral, usando o tratamento de equações simultâneas.

Desde então, a quantificação passa a ser vista como uma das bases da descrição da realidade econômica, e a Matemática torna-se instrumento indispensável para o economista.

Nas três primeiras décadas do século XX, ganha realce a pesquisa com vistas ao estabelecimento de modelos matemáticos para representar fenômenos macroeconômicos. Esses destaques aos métodos quantitativos culminaram com a fundação da Sociedade de Econometria, no ano de 1933, criada com o imperativo científico de estimular, orientar e divulgar trabalhos teóricos e práticos de Economia Matemática para a verificação das idéias econômicas.

O desenvolvimento dos métodos matemáticos em Economia proporciona um alargamento de suas potencialidades na "modelização" da atividade econômica. Um grande número de linhas de estudos com instrumental quantitativo é pesquisado e divulgado. A preocupação com a análise dinâmica das variáveis macroeconômicas torna-se constante.

---

[3] O florescimento das investigações matemáticas na Economia teve vários seguidores: Jevons, Menger, Cassel, Edgeworth, Wicksell e tantos outros, preocupados, notadamente, com a mensuração da utilidade na Teoria do Consumidor. Para maiores detalhes sobre o assunto ver, entre outros, MALANOS (1969) e BELL (1961).

[4] Sua postulação de que a oferta e a demanda determinam o preço está totalmente expressa em termos de diagramas cartesianos bidimensionais (BELL, 1961).

Já na época da 1ª Guerra Mundial, muitos cientistas estavam preocupados com o estudo de fenômenos que mostravam variação no tempo, notadamente aqueles que pudessem proporcionar a compreensão da natureza, da intensidade e do controle do comportamento cíclico da atividade econômica, na esperança de que talvez fosse possível prever esses fenômenos e, em conseqüência, diminuir os efeitos nefastos que suas crises causavam ou, até mesmo, evitar sua repetição.

Nas primeiras décadas deste século, um número elevado de estudos sobre o acompanhamento dos negócios foi elaborado. A constatação de certas regularidades dos fenômenos econômicos, justaposta à sua inerente dificuldade de mensuração, e o crescente interesse das instituições públicas e privadas em obterem previsões sobre o comportamento dos negócios em vários setores da atividade econômica estimularam o estabelecimento e o desenvolvimento de novos rumos sobre o assunto.

Os estudos de acompanhamento das flutuações dos negócios em seus vários setores econômicos vão proporcionar subsídios para uma época de grande euforia em construções dos chamados "barômetros de conjuntura"[5]. Através desse instrumental, pela primeira vez, via-se estabelecida uma ligação estreita entre a Estatística e a Economia Política.

As idéias que norteavam a constituição desses barômetros tinham origem nos trabalhos de Juglar[6], e se caracterizavam como uma tentativa de estabelecimento de relações entre os movimentos do mercado de mercadorias (industriais e comerciais), do mercado monetário e do mercado de aplicações financeiras com as oscilações da atividade econômica. A contínua observação das variações das séries desses três mercados permitia constatar que, embora descrevessem as mesmas variações cíclicas, essas ocorriam em momentos distintos de tempo. Assim, o conhecimento de uma dessas séries proporcionava os meios para prever a evolução das outras duas.

Na década de vinte, os estudos das estatísticas econômicas e dos negócios refletiam um pensamento homogêneo sobre a análise das variáveis no tempo. Os barômetros construídos na Inglaterra e na Bélgica eram muito parecidos com os de Harvard e Berlim. As séries temporais eram, em geral, agregados macroeconômicos e incluíam a maioria dos registros estatísticos disponíveis que pudessem, de alguma forma, clarear o fenômeno dos ciclos da economia.

A grande depressão do final dos anos vinte trouxe consigo o questionamento de todas as áreas de pensamento da Ciência Econômica. O Comitê de Pesquisas Econômicas da Universidade de Harvard não só não previu o "crack" de Wall Street, como também o negou enquanto foi possível.

Em 1936, a publicação da "Teoria Geral do Emprego, Juros e Moedas" de Lord John M. Keynes revolucionou a forma de pensar a Teoria Econômica.

Entretanto a época não era propícia ao florescimento de discussões e idéias sobre o pensamento econômico. A 2ª Guerra Mundial não permitia

---

[5] Entre os principais "barômetros", destacavam-se o Barômetro de Harvard, criado em 1917 pelo Harvard Committee for Economic Research, e o do Institut für Konjunturforschung, fundado em 1925, em Berlim. Ver a respeito, entre outros, PLATIER (1967).

[6] Clément Juglar, em 1857, publica "Les crises commerciales et leur retour périodique", no qual, pela primeira vez, é mostrado o caráter cíclico das crises, até então consideradas acidentais.

que as atividades de pesquisa alcançassem a intensidade que a revolução keynesiana havia proporcionado à Teoria Econômica, em termos de discussão acadêmica da validade de sua base teórica e da comprovação, através de testes empíricos, do valor de seus métodos de análise.

Mesmo assim, a resposta insatisfatória dos barômetros aos problemas econômicos da época tinha sedimentado a idéia de que a forma de pensar os ciclos econômicos nos anos vinte também deveria sofrer drásticas modificações para atender aos requisitos da nova síntese econômica.

Ligado ao objetivo básico de obtenção de informações sobre os ciclos da economia, o campo da análise das séries temporais veio, no decorrer dos primeiros quarenta anos do século, sedimentando-se sobre uma metodologia cuja característica tem sido norteada no sentido de buscar separar a variação apresentada pela série de observações no tempo em quatro componentes distintos: tendência, sazonalidade, ciclo e aleatoriedade.

O procedimento comumente aceito, à época, para o tratamento das séries temporais, com vistas a proporcionar informações sobre aspectos cíclicos das atividades econômicas, era o seguinte[7]:

- estimar a componente sazonal com a ajuda de uma média móvel de comprimento correspondente ao período do ciclo da sazonalidade e posterior eliminação dessa componente da série original;
- estimar a tendência da série sem a componente sazonal, por meio de uma função de regressão, cujos parâmetros poderiam ser obtidos por mínimos quadrados;
- analisar, através da visualização gráfica da série residual (posterior à eliminação das componentes sazonal e tendencial), seu respectivo comportamento em comparação às demais séries econômicas.

A lógica implícita nesse tipo de análise era que, embora esse procedimento propiciasse visualizar comportamentos distintos com respeito à tendência e às flutuações sazonais entre as diversas séries econômicas, se essas peculiaridades individuais fossem eliminadas, os padrões de variação remanescentes em qualquer série lançariam alguma luz sobre o objetivo maior que era o ciclo dos negócios.

Nos anos quarenta, os analistas econômicos dedicados aos estudos dos ciclos, após refletirem sobre o comportamento dessas séries de tempo e os novos rumos alcançados pela Teoria Econômica, constataram que esse tipo de enfoque pouco estava colaborando com a teoria em sua busca de explicação para o funcionamento da economia.

Em parte, os analistas estavam constatando seu limitado instrumental de trabalho: método de pesquisa quantitativa pouco sofisticado, pequeno número de variáveis e de equações com as quais os modelos macroeconômicos eram formulados, séries pequenas, em geral com periodicidade de tempo que não satisfazia às necessidades da análise. Mesmo assim, os defensores desses modelos macroeconômicos mantinham a convicção de que

[7] Atualmente, a bibliografia especializada sobre séries temporais denomina esse procedimento de análise clássica ou método de decomposição. Para uma descrição detalhada, ver, entre outros: FOX (1968), HADLEY (1968), HAMBURG (1970).

suas concepções básicas estavam corretas. Concordavam, entretanto, que seus modelos agregativos não atendiam satisfatoriamente às finalidades (FOX, 1968).

A partir do final da década de cinqüenta, o desenvolvimento da tecnologia de computação eletrônica permitiu aos economistas, interessados na descrição quantitativa do mundo dos negócios, intenso intercâmbio de idéias, tecnologias e conhecimento de fatores institucionais do funcionamento da atividade econômica com pesquisadores de outras áreas do conhecimento científico. Além do maior número de informações numéricas disponíveis e do maior entendimento da teoria estatística, que estimulam o desenvolvimento de novas técnicas quantitativas, a Economia Matemática adquire expressivo avanço, tendo em vista sua maior integração com os demais campos da ciência.

Dentre os muitos exemplos de novos métodos quantitativos criados recentemente para simular a realidade e fazer previsões sobre o futuro, destaca-se a metodologia que os professores George E. P. Box e Gwilym M. Jenkins desenvolveram para analisar o comportamento de variáveis através de séries de tempo.

Essa metodologia foi desenvolvida no decorrer dos anos sessenta e divulgada no ano de 1970, na obra intitulada "Time series analysis: Forecasting and control" (BOX & JENKINS, 1976).

O fundamento do trabalho elaborado por esses dois autores está assentado na possibilidade de obtenção de alguns modelos lineares que se apresentam potencialmente capazes de descrever, com relativa precisão e de forma parcimoniosa, o comportamento do processo estocástico gerador da série temporal em análise, proporcionando, por conseguinte, previsões acuradas de valores futuros.

O objeto deste trabalho de dissertação para a obtenção do grau de Mestre em Ciências Econômicas é uma análise da metodologia desenvolvida por Box & Jenkins e sua aplicação a dados da realidade econômica do Rio Grande do Sul.

A importância desse tipo de estudo reside na divulgação de uma metodologia que tem-se mostrado extremamente potente para a "modelização" de variáveis dinâmicas. O estudo visa a apresentar a metodologia de forma acessível e capaz de proporcionar condições de operacionalização desse instrumento estatístico relativamente complexo e, portanto, de difícil acesso aos economistas pouco habilitados a um tratamento estatístico mais sofisticado dos dados da realidade econômica.

1.2 – A Importância das Séries Econômicas de Tempo

A Econometria é o ramo da Ciência Econômica que, através da análise estatística da realidade, busca estabelecer proposições econômicas, de caráter quantitativo, que possibilitam não só compreender e analisar o comportamento das variáveis sob uma determinada base econômica teórica, como também permitem a previsão de seus comportamentos futuros.

O método econométrico utiliza-se de procedimentos estatístico-matemáticos, com os quais proporciona ao economista condições de abstrair o

funcionamento da realidade econômica, buscando sintetizá-la através da formulação de modelos teóricos.

Dentre os instrumentos estatísticos usados em Econometria para o estabelecimento das estimativas das relações funcionais propostas em seus modelos teóricos, podem-se destacar dois grandes grupos: aqueles que procuram estimar relações entre variáveis econômicas, seja através de modelos de uma única equação, seja através de modelos mais complexos, como os de múltiplas equações simultâneas; e aqueles que se valem de equações de série de tempo.

Embora ambos os grupos de modelos se caracterizem por suas finalidades de descrever a realidade, o modelo de série de tempo é bastante distinto em sua natureza. Ao contrário dos modelos de regressão, não analisa seus movimentos os relacionando com uma ou um conjunto de outras variáveis causais que permitam explicar seu comportamento. O modelo de série temporal caracteriza-se por ser um estudo onde a previsão e o estabelecimento de seu modelo gerador são obtidos somente baseados no comportamento passado da própria variável em estudo, sem a relacionar com nenhuma outra, a não ser o tempo.

Assim, a característica básica da análise das séries temporais, que a distingue das outras análises estatísticas, é o reconhecimento explícito da importância da ordem em que as observações são feitas: enquanto em muitos outros tratamentos estatísticos as observações podem ser consideradas independentes, na análise das séries temporais as sucessivas observações são vistas como estatisticamente dependentes.

Esse tipo de tratamento torna-se extremamente importante uma vez que, em muitas áreas do conhecimento, existem fenômenos cujo desenvolvimento e variação assumem elevada importância com o passar do tempo. A análise das séries temporais permite que tanto os objetivos de previsão do futuro, baseados nos dados passados, como os de controle do processo que gera a série, possam ser conhecidos independentemente do comportamento de outras variáveis.

De outra parte, a análise de séries temporais pode-se tornar o único meio de tratamento estatístico disponível. É o caso em que se apresenta difícil (ou até mesmo impossível) a explicação do comportamento de uma variável através de um modelo causal que busca explicar esses movimentos através de outras variáveis econômicas. Isso pode ocorrer quando:

- os movimentos da variável são atribuíveis a fatores outros que, simplesmente, não são passíveis de caracterização em modelos causais ou, pelo menos, quando existe a impossibilidade de avaliar quantitativamente o efeito que apresentam para explicar o comportamento da variável em estudo. É o caso, por exemplo, de se querer estabelecer o comportamento do volume de vendas de um determinado artigo. Sua variação pode ocorrer em resposta a variações de preços, a variações de renda da população, a variação de gostos e hábitos ou, simplesmente, por ciclos de sazonalidade que caracterizam o consumo desses artigos. A maioria dessas causas apresenta efeitos nas variações de comportamento temporal da variável em análise. Porém, a intensidade dos efeitos de cada uma dessas causas é bastante difícil (às vezes impossível) de ser caracterizada e quantificada;
- mesmo sendo possível quantificar seus efeitos, os coeficientes dos modelos estimados possuem erros padrões relativamente grandes, levando a erros de estimativas bastante insatisfatórios;

- os custos do processo, no caso de ser possível a quantificação dos efeitos das variáveis, sejam elevados;
- é desejável a previsão futura do valor da variável e não se tem os valores das variáveis explicativas para aquele determinado período que se quer estimar. No caso dessas variáveis serem também estimadas, mesmo tendo-se um modelo que se ajuste bastante bem, com erros padrões de previsão pequenos, os erros padrões das variáveis explicativas podem não ser suficientemente pequenos, o que levará a estimativas com níveis de confiança inaceitáveis.

Assim, a análise das séries temporais, ao se caracterizar como o estudo da variável, baseado unicamente no conhecimento dos seus valores passados, independentemente do comportamento das demais variáveis, proporciona um meio bastante útil para o tratamento numérico dos dados econômicos.

1.3 – O Conceito de Série Temporal

Define-se uma série temporal como um conjunto de observações de uma variável dispostas seqüencialmente no tempo[8]. Conforme o conjunto gerado, pode-se classificar a série em contínua ou discreta. Diz-se que a série cronológica é discreta quando o conjunto de observações no tempo for finito ou infinito enumerável[9]. Caso contrário, isto é, se o conjunto gerado for infinito não enumerável, diz-se que a série é contínua. A grande maioria dos estudos econômicos que se valem do tratamento de séries temporais utiliza séries discretas, onde as observações são geradas em um intervalo de tempo com amplitude constante.

De outra parte, uma série temporal pode ser classificada como determinística ou estocástica. Diz-se que é determinística quando os futuros valores da série podem ser estabelecidos precisamente por alguma relação funcional matemática do tipo Y = f (Tempo). Será dita estocástica quando seus futuros valores só puderem ser expostos em termos probabilísticos, uma vez que a série está descrita por meio de uma relação funcional que envolve não só o tempo, mas também uma variável aleatória do tipo Y = f (Tempo, a), onde "a" é o termo aleatório residual, cuja inclusão se torna necessária quando não se consegue explicar completamente algum movimento irregular da série através unicamente da relação matemática.

Para fins deste trabalho, somente se considerarão séries temporais estocásticas discretas, cujas observações são geradas em intervalos de tempo constantes.

Na análise clássica, costuma-se classificar os movimentos da série econômica de tempo em quatro tipos básicos de variações: a tendência, as variações sazonais, as variações cíclicas e as variações aleatórias. Assim, as séries podem ser consideradas como constituídas por essas

---

[8] A definição dada aqui é de série temporal univariada. A generalização da definição deve conter séries multivariadas. Optou-se por essa definição tendo em vista os objetivos específicos do trabalho.

[9] Um conjunto é dito enumerável quando é possível existir uma bijeção entre ele e o conjunto dos números naturais

quatro componentes e tem-se como objetivo especificar a magnitude de cada um desses movimentos para possibilitar descrever conjuntamente o movimento da série.

A tendência, também chamada por alguns autores de tendência secular, é caracterizada como aquele movimento regular e contínuo de longo prazo, refletindo um movimento ascendente ou descendente em longo período de tempo. De certa forma, pode ser vista como aquela componente que descreve as variações graduais que se mantêm em um longo período de observação da variável no tempo. Estatisticamente, pode-se caracterizá-la pelo fato de que a expectância de Y, E(Y), varia no tempo.

As variações sazonais são aquelas variações periódicas (cíclicas) que ocorrem com certa regularidade dentro de um curto período de tempo. Embora o próprio nome dê a entender que esses movimentos ocorrem por um período anual, de acordo com as estações climáticas, esses movimentos podem ser estendidos a qualquer intervalo de curto prazo, como diário, horário, semanal, mensal, trimestral etc.

As variações cíclicas são aquelas variações que se referem às oscilações de longo prazo que caracterizam, em geral, os ciclos econômicos. São as flutuações de longo prazo em torno da curva de tendência.

As variações aleatórias, também chamadas residuais, referem-se não só àqueles movimentos esporádicos ocasionados por eventos aleatórios imprevisíveis, tais como as calamidades da natureza, mas também ao conjunto de todos aqueles movimentos da série que não foram passíveis de identificação em seus demais componentes, uma vez que não obedecem a nenhuma lei comportamental capaz de ser descrita de forma determinística, através de relações funcionais exclusivamente matemáticas.

Deve-se salientar, entretanto, que, modernamente, a inclusão da componente cíclica é assunto bastante controverso entre os pesquisadores estatísticos[10]. As variações cíclicas são pouco comuns. Basicamente, sua existência resume-se a algumas séries econômicas particulares, as quais nem sempre são passíveis de caracterização. Uma parte de um ciclo de amplitude bastante grande, por exemplo, pode ser confundida com a componente tendencial, dado o conjunto relativamente pequeno de valores observáveis em contraste com a amplitude que o ciclo pode ter. Também pode ser muito difícil diferenciar as flutuações que são devidas à componente cíclica daquelas estritamente aleatórias.

Por essas razões, é bastante comum, nas análises estatísticas de séries temporais, decompor os movimentos da série excluindo a componente cíclica. Assim, a série é usualmente constituída de três componentes: a tendência, que descreverá os movimentos regulares de longo prazo, onde se incluirão as flutuações cíclicas de longo prazo e que não podem ser percebidas e explicitadas com os dados da série que se encontram disponíveis[11]; a sazonalidade, que descreverá as flutuações cíclicas de curto prazo; e o termo aleatório residual.

Para algumas séries econômicas, porém, julga-se que deva ser considerada a possibilidade de terem suas flutuações influenciadas por algum

---

[10] Ver a respeito, entre outros: GRANGER & NEWBOLD (1977), MORETTIN (1981).

[11] Alguns autores preferem, por essa razão, denominar essa componente de tendência-ciclo.

movimento cíclico da atividade econômica. Nesse caso, deve ser buscado instrumental adequado específico para sua análise. Por exemplo, a utilização da análise econométrica dinâmica no estudo do ciclo.

1.4 – O Significado da Análise de Séries Temporais

A base do tratamento requerido pelas séries de tempo está assentada na possibilidade de serem tiradas conclusões sobre o comportamento passado da variável e que poderão ser úteis para proporcionar informações sobre o seu comportamento futuro provável. Os problemas fundamentais que norteiam o estudo das séries temporais dizem respeito, basicamente, a questões como: a série exibiu, no passado, algum tipo de tendência que possa influenciar o seu comportamento futuro? A série exibe algum tipo de comportamento cíclico, seja de curto ou longo prazo, que poderá ser extrapolado para o seu comportamento futuro?

Se tais tipos de causas comportamentais estão presentes, poder-se-á buscar construir um modelo para os valores da variável seqüencialmente dispostos no tempo, de maneira que a variável possa ser estudada de forma a recaírem sobre o seu próprio comportamento passado os meios a serem utilizados para a avaliação de seu comportamento futuro provável, sem necessidade de serem estabelecidas relações explicativas dos efeitos da influência de outras variáveis.

Portanto a análise de um modelo de série de tempo significa um estudo que busca obter as características comportamentais sistemáticas da série, capazes de propiciar a construção de um modelo que descreva os movimentos passados de uma variável, com o que se poderá predizer os futuros movimentos da mesma.

Sua base teórica está assentada na suposição de que a série tenha sido gerada por um processo estocástico, através do qual sua estrutura pode ser caracterizada e descrita. Vale dizer que a análise do modelo de série temporal proporciona uma descrição dos mecanismos e da natureza aleatória do processo estocástico que gerou a amostra de observações em estudo. Sua descrição é dada, não em termos de relação entre variáveis, como no modelo tradicional de regressão, mas sim em termos de como a aleatoriedade está embutida no processo.

Basicamente, a análise de série temporal pressupõe que exista um processo estocástico gerador da série, ou seja, que, a cada possível realização aleatória da variável, esteja associada uma probabilidade de ocorrência da observação. Assim, o que se busca é descrever os mecanismos do processo e as características de sua aleatoriedade, porquanto isso irá fornecer os meios para que se chegue a conclusões sobre as probabilidades associadas aos valores futuros alternativos da série.

Em síntese, uma série temporal significa um conjunto de variáveis aleatórias conjuntamente distribuídas no tempo. Sua análise baseia-se na suposição da existência de alguma função que assume probabilidades para todas as possíveis combinações dos valores da variável. Conseqüentemente, se é possível descrever numericamente como é a estrutura probabilística da variável no tempo, então poder-se-á inferir sobre a probabilidade de ocorrência de um outro futuro valor.

2 – SOBRE SÉRIES TEMPORAIS E PROCESSOS ESTOCÁSTICOS

2.1 – Processos Estocásticos e suas Propriedades

O objeto da teoria dos processos estocásticos é o estudo daqueles mecanismos dinâmicos que proporcionam meios de análise de uma seqüência de observações, vista conjunta e interdependentemente em uma sucessão de momentos de tempo[1], as quais são influenciadas por fatores aleatórios.

Assim, um processo estocástico deve ser entendido como um modelo que descreve a estrutura probabilística de uma seqüência de observações. Formalmente, pode-se definir um processo estocástico como uma família de variáveis aleatórias $\{X(t), t\in T\}$, classificada mediante um parâmetro "t" que varia em um intervalo de tempo T. Portanto um processo estocástico caracteriza-se por ser uma função aleatória de "t". Uma seqüência de observações X_t, t=1, 2, 3, 4, ..., t, que corresponde a uma amostra de pontos no tempo, é chamada uma realização parcial do processo estocástico.

Ao observar o comportamento de uma série temporal, verifica-se que essa pode ser considerada como uma particular realização de uma seqüência de observações produzida por um mecanismo probabilístico. Dessa maneira, uma série temporal pode ser vista como uma realização de um processo estocástico. Ou seja, uma série temporal $x_1, x_2, x_3, \ldots, x_t$ pode ser considerada como uma realização amostral de uma população infinita constituída por uma infinidade de outros possíveis resultados para o conjunto $\{X(t), t\in T\}$, realização que pode ter sido gerada pelo processo.

A análise das séries temporais supõe que o conjunto de observações esteja sendo gerado por um processo estocástico e, como tal, possua uma estrutura probabilística que possa ser caracterizada e descrita. O estudo das séries temporais proporciona, então, a descrição da natureza aleatória do processo que gerou a amostra de observações em estudo. Sua análise supõe que cada ocorrência $x_1, x_2, x_3, \ldots, x_t$ da série seja obtida aleatoriamente com base em uma estrutura probabilística. Conseqüentemente, a série temporal X_t é uma seqüência de variáveis aleatórias conjuntamente distribuídas, vale dizer, existe alguma função de probabilidade $P(x_1, x_2, x_3, \ldots, x_t)$ que assume valores para todas as possíveis combinações $x_1, x_2, x_3, \ldots, x_t$.

Assim, o objetivo básico da análise estatística de séries de tempo é buscar, a partir da realização amostral do processo (isto é, a série $x_1, x_2, x_3, \ldots, x_t$), descrever as características de sua aleatorieda-

[1] Embora, em termos práticos, seja comum tratar a sucessão de observações em um domínio temporal, nada impede que a sucessão de observações seja analisada em qualquer outro domínio, tal como volume, distância etc.

de, com o fito de proporcionar os instrumentos para a inferência sobre as probabilidades associadas com o conjunto de valores futuros alternativos da série. Ao se conseguir especificar numericamente como é a função de probabilidade da série, torna-se viável inferir a probabilidade de um ou outro futuro valor ocorrer.

Portanto estudar modelos de séries temporais significa buscar obter meios capazes de inferir as características de seu processo gerador, bem como buscar modelos estocásticos que sejam capazes de descrever as situações particulares que ocorrem na realidade.

2.2 – Caracterização de um Processo Estocástico

Um processo estocástico pode ser caracterizado como uma família de variáveis aleatórias definidas em um espaço de probabilidades (Ω, η, P),[2] o qual estará perfeitamente especificado se se conseguir obter a sua função de probabilidade conjunta para a série.

Assim, considerando $t_1, t_2, t_3, \ldots, t_n$ um vetor de elementos de T e

$$F(x_1, \ldots, x_n; t_1, \ldots, t_n) = P(X(t_1) \leq x_1; \ldots; X(t_n) \leq x_n) \qquad (2.1)$$

então o processo estará completamente especificado se se conseguir obter as distribuições n-dimensionais estabelecidas pela equação (2.1). Vale dizer que, se, por exemplo, n=1, se deve conhecer a distribuição unidimensional de $X(t_1)$; se n=2, deve-se conhecer a distribuição bidimensional $(X(t_1), X(t_2))$; se n=k, deve-se conhecer a distribuição k-dimensional $(X(t_1), X(t_2), \ldots, X(t_k))$.

Para que a equação (2.1) defina um processo estocástico, deve satisfazer as condições de simetria e compatibilidade, onde:

a) **condição de simetria**: para qualquer permutação $i_1, i_2, \ldots, i_n$ dos números $1, 2, \ldots, n$, tem-se que

$$F(x_{i_1}, x_{i_2}, \ldots, x_{i_n}; t_{i_1}, t_{i_2}, \ldots, t_{i_n}) = F(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n)$$

b) **condição de compatibilidade**: se $m < n$, então, para qualquer $t_{m+1}, t_{m+2}, \ldots, t_n$

$$F(x_1, \ldots, x_m; t_1, \ldots, t_n) =$$

$$= F(x_1, \ldots, x_m, \infty, \ldots, \infty; t_1, \ldots, t_m, t_{m+1}, \ldots),$$

---

[2] A nível de análise de séries cronológicas desenvolvidas neste trabalho, supõe-se que a família de variáveis aleatórias esteja definida no mesmo espaço de probabilidade (Ω, η, P), onde Ω é o espaço amostral; η a σ-álgebra de subconjuntos de Ω; e P, a probabilidade definida em η.

onde o segundo membro dessa equação deve ser entendido como

$$\lim_{\substack{X_{m+1} \to \infty \\ \vdots \\ X_n \to \infty}} F(x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_n; t_1, t_2, \ldots, t_n)$$

Contudo a especificação completa da função de probabilidade n-dimensional, que rege o processo gerador da série temporal, é bastante improvável de se obter. Por isso é usual, nesses tipos de análises, estudar algumas características paramétricas associadas com sua estrutura probabilística, capazes de proporcionar meios alternativos mais simplificados de descrição daquela estrutura do que propriamente a função de probabilidade.

Uma maneira equivalente de especificar a distribuição n-dimensional, geradora da série temporal, é determinar todos os seus momentos-produtos de ordem $(r_1, r_2, \ldots, r_n)$ das variáveis $X(t_1), X(t_2), \ldots, X(t_n)$, ou seja,

$$E(X^{r_1}(t_1) \cdot X^{r_2}(t_2) \ldots X^{r_n}(t_n)) =$$

$$= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} x_1^{r_1} \ldots x_n^{r_n} f(x_1, \ldots, x_n; t_1, \ldots, t_n) dx_1 \ldots dx_n$$

que, como se pode verificar pela sua formulação, apresenta as mesmas restrições do conhecimento da estrutura probabilística.

A busca de meios capazes de descrever o processo gerador da série impõe a necessidade de se restringir o estudo a momentos de baixa ordem. Para as finalidades específicas dos processos estocásticos que se estudarão neste trabalho, serão úteis os momentos de primeira e segunda ordem, ou seja, média, variância e covariância, os quais podem ser definidos como segue.

Seja um processo estocástico $\{X(t), t \in T\}$, então:

a) **função valor médio**: é definida por $E(X_t) = \mu_t$, qualquer $t \in T$.

Observe-se que, a nível de série temporal, a função valor médio de seu processo gerador significa a componente tendencial da série.

b) **função variância**: é definida por $Var(X_t) = E(X_t - \mu_t) = \sigma_t^2$, qualquer $t \in T$.

c) **função covariância**: é definida por $Cov(X_t, X_s) = E(X_t - \mu_t)(X_s - \mu_s) = \gamma_{t,s}$, qualquer $t, s \in T$.

2.3 – Processos Estacionários e Não Estacionários

Se o processo estocástico que gerou a série de observações é invariante com respeito ao tempo, diz-se que o mesmo é estacionário. Se as características do processo se alteram no decorrer do tempo, diz-se que é não estacionário.

Segundo PAPOULIS (1965), um processo estacionário pode ser classificado em:

a) **estritamente estacionário**: quando suas estatísticas não são afetadas por variações devido à escolha da origem dos tempos, ou seja, quando as séries X_t e X_{t+k} estão distribuídas identicamente, qualquer que seja "k".

Dessa definição segue que a função densidade de um processo estritamente estacionário deve ser tal que permaneça idêntica quando se varia a origem dos tempos, ou seja:

$$P(x_1, x_2, \ldots, x_t; t_1, t_2, \ldots, t_t) = P(x_1, x_2, \ldots, x_t; t_{1+k}, t_{2+k}, \ldots, t_{t+k})$$

para quaisquer $t_1, t_2, \ldots, t_t \in T$ e k.

b) **estritamente estacionária de ordem finita**: diz-se que um processo é estritamente estacionário de ordem "i" se a estacionariedade no item (a) é válida não para todo $t_j \in T$, mas somente para $j \leq i$;

c) **estacionária em sentido lato (ou fracamente estacionário ou, ainda, estacionário de segunda ordem)**: quando a sua função valor médio é constante e sua função de covariância depende somente da diferença, em valor absoluto, de $t_s - t_j$.[3]

Atente-se para o fato de que um processo estacionário em sentido lato pode não ser estritamente estacionário, pois os momentos até segunda ordem não garantem condições sobre a estacionariedade da função de probabilidade. Porém, se o processo é estritamente estacionário de ordem dois, então ele é estacionário em sentido lato.

Deve-se salientar, entretanto, que, se o processo for gaussiano[4] e estacionário em sentido lato, ele será estritamente estacionário, devido ao fato de a distribuição normal ser determinada unicamente em termos do primeiro e do segundo momento.

Para fins deste trabalho, onde a suposição usual é a de um processo gaussiano, é suficiente o requerimento das condições da definição de estacionário em sentido lato, o qual será, a partir desse instante, denominado simplesmente estacionário.

Um processo estacionário, assim definido, satisfaz as seguintes condições:

$$E(x_t) = \mu, \text{ qualquer } t \in T;$$

$$Var(X_t) = \gamma_o, \text{ qualquer } t \in T;$$

$$Cov(X_t, X_{t-k}) = \gamma_k, \text{ isto é, depende somente de "k".}$$

---

[3] Observe-se que o processo estacionário em sentido lato requer condições envolvendo somente os momentos de primeira e segunda ordem, ao invés de sua função finito-dimensional. Daí porque é chamado por muitos autores de estacionário de segunda ordem.

[4] Um processo é dito gaussiano se a distribuição de probabilidade da série X_t é multinormal.

A importância do conhecimento de a série ser ou não estacionária reside no fato de que, quando se trabalha com uma série estacionária, se está em presença de uma função amostral do processo que tem a mesma forma em todos os instantes de tempo t∈T, o que acarreta possibilidades de obtenção de estimativas das características do processo de forma bastante simples, o que, em caso contrário, não seria tarefa fácil.

De outra parte, um processo estacionário, que se caracteriza por possuir um comportamento geral de sua estrutura probabilística invariante no tempo, impõe que uma realização amostral x_t, x_{t+1}, x_{t+2}, ..., x_{t+n}, para qualquer "t", embora não necessariamente seja igual a, por exemplo, x_{t+100}, x_{t+101}, x_{t+102}, ..., $x_{t+100+n}$, apresente sua forma geral e suas futuras observações de maneira similar, o que implica facilidades de se fazerem previsões acuradas com esse tipo de processo.

2.4 – Propriedades dos Processos Estacionários

Uma série temporal pode ser vista como sendo gerada por um conjunto de variáveis aleatórias conjuntamente distribuídas. A série x_1, x_2, x_3, ..., x_t, portanto, representa um específico resultado da função de probabilidade conjunta $P(x_1, x_2, x_3, \ldots, x_t)$. Conseqüentemente, uma futura observação, x_{t+h}, pode ser considerada como sendo gerada pela função de probabilidade condicional $P(x_{t+h}/x_1, x_2, x_3, \ldots, x_t)$.

Conforme foi visto no item anterior, um processo estacionário pode ser definido como um processo onde a distribuição conjunta e a distribuição condicional são invariantes aos deslocamentos no tempo, ou seja,

$$P(x_1, x_2, x_3, \ldots, x_t) = P(x_{1+k}, x_{2+k}, x_{3+k}, \ldots, x_{t+k}), \ \forall k.$$

Assim, se a série é estacionária, sua função valor médio mantém-se invariante no tempo, ou seja,

$$E(X_t) = E(X_{t+k}) = \mu \text{ (constante)},$$

uma vez que:

$$E(X_t) = \int_{-\infty}^{\infty} x f(x_t)dx = \int_{-\infty}^{\infty} x f(x_{t+k})\, dx = E(x_{t+k}), \ \forall k.$$

Significa dizer que, se o processo é estacionário, $E(X_t) = \mu$, qualquer "t", a série não apresenta a componente tendencial.

Sua função variância também é constante:

$$Var\ (X_t) = E(X_t - \mu)^2 = E(X_{t+k} - \mu)^2 = Var\ (X_{t+k}), \ \forall k,$$

pois,

$$Var(X_t) = E(X_t^2) - E(X_t)^2 = \int_{-\infty}^{\infty} x^2 f(x_t)dx - \mu^2 = \int_{-\infty}^{\infty} x^2 f(x_{t+k})dx - \mu^2 =$$
$$= E(X^2_{t+k}) - E(X_{t+k})^2 = Var(x_{t+k}), \ \forall k.$$

Sua função de covariância é função somente da distância "t-s", independente da origem do período de tempo:

$Cov(X_t, X_s) = Cov(X_{t+n}, X_{s+n}) = \gamma_{t-s}$;

$Cov(X_t, X_s) = Cov(X_{t+k}, X_{s+h})$. Fazendo n = -s, tem-se que:

$Cov(X_{t+h}, X_{s+h}) = Cov(X_{t-s}, X_o) = \gamma_{t-s}$.

Note-se que, se o processo é estacionário, então tem média e variância que não variam no tempo e a covariância entre observações existentes em dois períodos distintos de tempo depende somente da distância entre esses pontos e não do específico período de tempo das observações.

Na prática, entretanto, o que se dispõe é de uma realização amostral do processo. Assim, as funções valor médio, variância e covariância são obtidas através de suas estimativas. Como a função de probabilidade $P(x_{1+k}, x_{2+k}, x_{3+k}, \ldots, x_{t+k})$ é a mesma para todo "k", se o processo é estacionário, então uma estimativa da função valor médio do processo, $E(X_{t+k}) = \mu$, qualquer "k", pode ser obtida pela função valor médio amostral da série

$$\bar{X} = \frac{1}{T} \Sigma X_t.$$

Da mesma forma, a estimativa da função variância pode ser obtida pela variância amostral,

$$\widehat{Var}(X_t) = \frac{1}{T} \Sigma (X_t - \bar{X})^2 = \hat{\sigma}^2,$$

e a estimativa da função de autocovariância,

$$\widehat{Cov}(X_t, X_s) = \frac{1}{T} \Sigma (X_k - \bar{X})(X_s - \bar{X}) = \hat{\gamma}_{t-s}.$$

2.5 – Processos Evolutivos Homogêneos

Embora a teoria mostre a conveniência prática do uso de séries estacionárias, no mundo real, infelizmente, muito poucas séries encontradas podem ser classificadas como estacionárias.

A estacionariedade é uma condição bastante restritiva imposta à série temporal e, no que diz respeito a séries econômicas, somente um reduzidíssimo número satisfaz essa condição.

Um simples exame de algumas das variáveis mais comuns em Economia, como por exemplo, preço, produto nacional, renda, vendas, consumo, investimento e tantas outras, mostra que suas variações apresentam um comportamento não estacionário evolutivo no decorrer do tempo.

O tipo de não estacionariedade mostrado por essas séries pode ser caracterizado como homogêneo. Ainda que as séries se movimentem livremente sem se fixarem em torno de um particular valor médio, seu comportamento em diferentes períodos de tempo é essencialmente similar.

Felizmente, séries não estacionárias homogêneas são aquelas que se caracterizam por apresentarem propriedades tais que, diferenciando-as uma ou mais vezes, resultam em séries estacionárias.

O número de vezes que a série original deve ser diferenciada para se tornar uma série estacionária é chamado de ordem de homogeneidade. Assim, por exemplo, se X_t é uma série não estacionária homogênea de primeira ordem, então a série $Y_t = X_t - X_{t-1} = \nabla X_t$ é estacionária. Da mesma forma, se X_t é não estacionária homogênea de segunda ordem, então a série $Y_t = \nabla^2 X_t = \nabla X_t - \nabla X_{t-1}$ será estacionária[5].

2.6 – A Função de Autocorrelação

Com base nas características paramétricas definidas no item 2.2, pode-se definir o coeficiente de autocovariância com defasagem "k" da série como sendo a covariância entre X_t e X_{t+k}, onde "k" é o número de intervalos de tempo defasados, ou seja,

$$Cov(X_t, X_{t+k}) = E(X_t - \mu_t)\ (X_{t+k} - \mu_{t+k}).$$

Se a série é estacionária, tanto a média quanto a variância não variam no tempo, e a autocovariância entre duas observações quaisquer depende somente do número de intervalos defasados. Assim,

$$Cov(X_t, X_{t+k}) = E(X_t - \mu)\ (X_{t+k} - \mu) = \gamma_k.$$

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

A análise dessa fórmula proporciona importantes interpretações sobre o comportamento da série. Por exemplo, se uma observação que se encontra acima do valor médio tende a ser seguida por outra observação acima da média, "k" períodos defasados ou, semelhantemente, para observações X_t e X_{t+k} situadas abaixo da média, então a autocovariância será positiva. Caso contrário, será negativa.

Assim, é importante analisar os valores dos coeficientes de autocovariância, pois eles irão desempenhar papel fundamental no estudo do comportamento das séries temporais. Por exemplo, se o sinal da autocovariância, para um período defasado (k=1), é negativo para algum processo, significa que a série se caracteriza por ser uma seqüência de observações situadas alternadamente acima e abaixo do valor médio.

---

[5] O símbolo ∇ é chamado de operador de retardo ou de diferença de retardo e será especificamente analisado no capítulo subseqüente.

Para um processo estacionário, a função de autocovariância, $\gamma_k = E(X_{t+k} - \mu)\ (X_t - \mu)$, apresenta as seguintes propriedades:

$$* \ \gamma_0 > 0 \tag{2.2}$$

$$* \ \gamma_{-k} = \gamma_k \tag{2.3}$$

$$* \ \gamma_0 \geq |\gamma_k| \tag{2.4}$$

* γ_k é tal que

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \ \gamma_{j-k} \geq 0, \forall_j, \ r \in \mathbb{R} . \tag{2.5}$$

As propriedades (2.2) e (2.3) são imediatas, a partir da definição de autocovariância do processo estocástico.

A propriedade (2.4) prova-se a partir do fato de $E\left[(X_{t+k} - \mu) \pm (X_t - \mu)\right]^2 \geq 0$.

$$0 \leq E\left[(X_{t+k} - \mu) \pm (X_t - \mu)\right]^2 = E\left[(X_{t+k} - \mu)^2 \pm 2\ (X_{t+k} - \mu)\ (X_t - \mu) + \right.$$

$$\left. + (X_t - \mu)^2\right] = E(X_{t+k} - \mu)^2 + 2\ E\left[(X_{t+k} - \mu)\ (X_t - \mu)\right] + E(X_t - \mu)^2 = \gamma_o \pm$$

$$\pm 2\gamma_k + \gamma_o = 2\gamma_o \pm 2\gamma_k .$$

Logo $2\gamma_o \pm 2\gamma_k \geq 0 \iff \gamma_o \geq |\gamma_k|$

Para provar a propriedade (2.5) vê-se que:

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \ \gamma_{j-k} = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \ E(X_j - \mu)(X_k - \mu) =$$

$$= E\left(\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \ (X_j - \mu)(X_k - \mu)\right) = E\left(\sum_{j=1}^{n} a_j a_j (X_j - \mu)(X_j - \mu)\right) =$$

$$= E\left(\sum_{j=1}^{n} a_j^2 \ (X_j - \mu)^2\right) \geq 0.$$

Para fins de comparação entre diferentes séries, é conveniente trabalhar com uma medida padronizada do coeficiente de autocovariância, sem unidade de medida, uma vez que diferenças em escala e/ou medida podem levar a diferenças acentuadas nos coeficientes de autocovariância, dificultando a comparação entre séries distintas.

Essa padronização é obtida através do coeficiente de autocorrelação, que se caracteriza por ser o quociente entre a autocovariância de defasagem "k" e a autocovariância de defasagem zero. Ou seja, o coeficiente de autocorrelação de defasagem "k" é definido por:

$$\rho_k = \frac{E(x_t - \mu_t)\ (x_{t+k} - \mu_{t+k})}{\sqrt{Var(x_t)\ .\ Var(x_{t+k})}}\ ;$$

$$\rho_k = \frac{Cov\ (x_t\ x_{t+k})}{DP(x_t)\ .\ DP(x_{t+k})}\ ,\ \text{onde DP é o desvio padrão.}$$

Observe-se que, se o processo for estacionário, existirá a igualdade $Var(X_t) = Var(X_{t+k})$. Então, o coeficiente de autocorrelação para processos estacionários é dado por:

$$\rho_k = \frac{Cov(X_t, X_{t+k})}{Var(X_t)} = \gamma_k/\gamma_o\ .$$

Conseqüentemente,

$$\rho_o = \gamma_o/\gamma_o = 1\ .$$

Como na prática se tem somente uma amostra de observações, o coeficiente de autocorrelação é obtido através de sua estimativa, que é dada por:

$$r_k = \hat{\rho}_k = \frac{\Sigma(X_k - \bar{X})\ (X_{t+k} - \bar{X})}{\Sigma\ (X_t - \bar{X})^2}$$

A função que associa cada valor de "k" com o seu respectivo coeficiente de autocorrelação é chamada função de autocorrelação. Seu gráfico é comumente chamado de correlograma.

Deve-se notar que a função de autocorrelação é simétrica em relação a $k = 0$, ou seja, os valores da função para as defasagens positivas de "k" são os mesmos que os valores para as defasagens negativas correspondentes: $\rho_k = \rho_{-k}$, $\forall\ k$.[6]

Além das características interpretativas que se obtêm com a análise da autocovariância, a autocorrelação proporciona maiores informações. A seqüência ρ_k, $k = 0, 1, 2, 3, \ldots$, é uma indicação da extensão para a qual um valor do processo é correlacionado com seus valores defasados "k" períodos de tempo. Pode, então, para alguma extensão, ser usada pa-

---

[6] Essa propriedade é uma conseqüência natural de (2.3): se

$$\gamma_k = \gamma_{-k} \Rightarrow \rho_k = \frac{\gamma_k}{\gamma_o} = \frac{\gamma_{-k}}{\gamma_o} = \rho_{-k}.$$

ra medir o número de defasagens e a intensidade da "memória" do processo, isto é, a extensão para a qual o valor tomado no tempo "t" depende daqueles tomados no tempo "t-k".

A análise da função de autocorrelação da série, portanto, torna-se importante à medida que proporciona um grau de intensidade da correlação existente entre observações vizinhas da série. Indica, por conseguinte, uma medida de interdependência entre as observações e será um instrumento de extrema importância para caracterizar propriedades de processos geradores de séries temporais.

Assim, por exemplo, estando em presença de um processo caracterizado por uma seqüência de variáveis independentes e identicamente distribuídas[7], a função de autocorrelação terá seu correlograma facilmente descrito, pois,

$$\rho_k = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

e, conseqüentemente, seu correlograma será dado por:

Figura 1 – Correlograma característico de um processo de ruído branco

De outra parte, a função de autocorrelação também é um instrumento bastante útil para ser usado como teste inicial de verificação de estacionariedade da série temporal, à medida que satisfaz um conjunto de condições que estabelecerá restrições aos diversos valores de seus coeficientes.

Essas condições podem ser assim deduzidas: seja uma série temporal estacionária, $x_1, x_2, \ldots, x_t$; suas funções de autocorrelação e autocovariância, expostas na forma matricial, são:

---

[7] Esse processo, quando possui média zero e variância constante, recebe o nome de processo de ruído branco e será intensamente utilizado nos capítulos subseqüentes.

$$\mathbb{P}_t = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{t-1} \\ \rho_1 & 1 & \cdots & \rho_{t-2} \\ \vdots & & & \\ \rho_{t-1} & \rho_{t-2} & \cdots & 1 \end{bmatrix} \quad \text{e} \quad \Gamma_t = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{t-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{t-2} \\ \vdots & & & \\ \gamma_{t-1} & \cdots & & \gamma_0 \end{bmatrix}$$

onde $\Gamma_t = \gamma_0 \mathbb{P}_t$.

Seja L_t uma funçao linear de x_t e seus valores defasados, ou seja:

$$L_t = a_1 x_t + a_2 x_{t-1} + \ldots + a_n x_{t-n+1} .$$

Como X_t é estacionário, $Cov(X_i, X_j) = \gamma_{i-j}$, então a variância de L_t é

$$Var(L_t) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \gamma_{i-j}$$

e, se a_i e a_j não são todos zeros, por (2.5) e (2.2) tem-se que:

$$Var(L_t) > 0.$$

Vale dizer que, para um processo estacionário, a matriz de autocovariância e a de autocorrelação se apresentam positivamente definidas, com determinante e todos os menores principais positivos.

Conseqüentemente, as restrições aos valores dos coeficientes de autocorrelação podem ser assim obtidas, se o número de observações da série for:

$$t = 2 , \det \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} > 0 \Rightarrow 1 - \rho_1^2 > 0 \Leftrightarrow -1 < \rho_1 < 1 ;$$

$$t = 3, \quad \det \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} > 0; \ \det \begin{bmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{bmatrix} > 0 ;$$

$$\det \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix} > 0 ,$$

o que implica

$$-1 < \rho_1 < 1; \quad -1 < \rho_2 < 1 \; ; \; 1 < \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} < 1,$$

e assim sucessivamente para t = 4, 5, ...

Como se vê, essas restrições impostas aos valores da função de autocorrelação tornar-se-ão gradativamente de mais difícil obtenção à medida que "t" cresce.

Embora exista complexidade analítica em sua dedução, essas restrições apresentam como característica principal o fato de ρ_k diminuir rapidamente para zero à medida que "k" cresce[8]. Como o cálculo desses valores pode ser obtido sem maiores dificuldades via pacotes computacionais de processamento eletrônico, tem-se, através de um simples exame visual do correlograma, uma avaliação da possibilidade de a série ser ou não estacionária.

Assim, por exemplo, se, construindo o correlograma, se constata que o coeficiente de autocorrelação não diminui rapidamente quando o valor de "k" aumenta (conforme se visualiza na Figura 2), então está indicada a não estacionariedade da série. Em caso contrário, se o coeficiente de autocorrelação diminuir rapidamente, estar-se-á em presença de uma série estacionária (Figura 3).

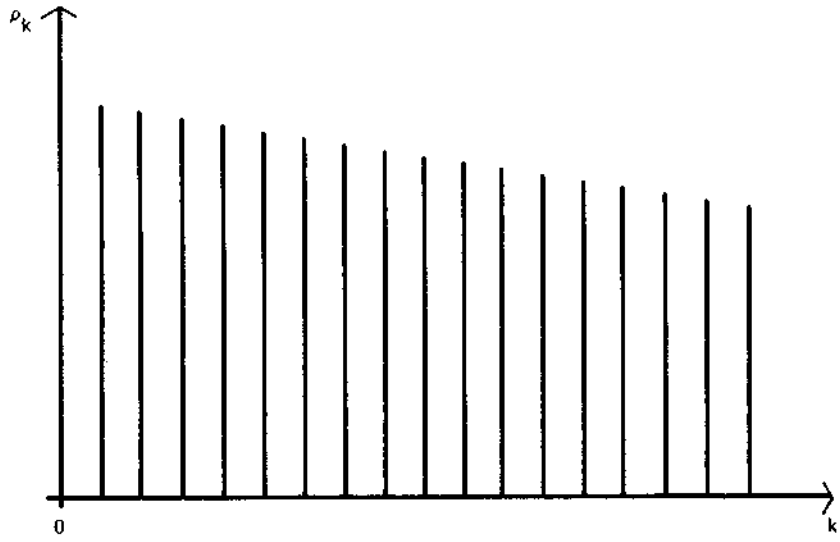


Figura 2 – Correlograma característico de uma série não estacionária

[8] Da mesma forma, pode-se provar que, se um processo é não estacionário, sua função de autocorrelação tende a declinar lentamente. Ver a respeito: ANDERSON (1979); BOX & JENKINS (1976) e NELSON (1973).

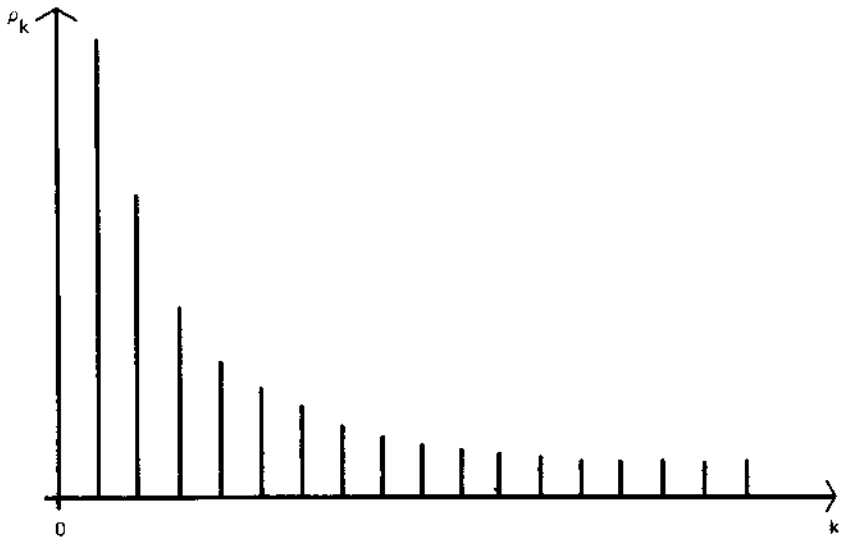


Figura 3 – Correlograma característico de uma série estacionária

Pode-se também utilizar o correlograma para verificar o grau de homogeneidade das séries não estacionárias. Para tanto, basta diferenciar, sucessivamente, a série original e construir os respectivos gráficos da função de autocorrelação. Se, na k-ésima vez que se diferencia, resulta que a função de autocorrelação tende rapidamente para zero, então pode-se determinar que a série original é homogênea de k-ésima ordem.

A função de autocorrelação e a sazonalidade: seja uma série que exibe algum comportamento sazonal. Por exemplo, seja uma série cujo ciclo sazonal tem periodicidade anual. Como a sazonalidade é uma componente oscilatória da série, que se caracteriza pela regularidade dentro de um período curto de tempo (no caso 12 meses), então as observações originais da série x_t e x_{t-12} deverão mostrar algum grau de correlação. Da mesma forma, para x_{t-24}, x_{t-36}, x_{t-48}, ... Essas correlações serão necessariamente mantidas na função de autocorrelação. Assim, pode-se identificar a sazonalidade da série pela observação de picos regulares no correlograma (no caso exemplificado, quando k = 12, 24, 36, ...), mesmo que na série original esses picos não possam ser percebidos.

No caso de série que apresenta ciclo sazonal de 12 meses, o meio de que se dispõe para sua eliminação é gerar uma nova série a partir da diferença de 12 meses, ou seja, obtendo-se a série $Z_t = X_t - X_{t-12}$, qualquer $t \varepsilon T$.

A análise através do correlograma: exemplo: seja a série temporal graficamente apresentada na Figura 4 representativa de 120 observações mensais da produção suína nos Estados Unidos, no período de janeiro de 1962 a dezembro de 1971.[9]

[9] O exemplo aqui apresentado foi retirado de PINDYDK & RUBINFELD (1976). A fonte dos dados brutos utilizados por aqueles autores foi o Survey of Current Bussiness.

Deseja-se saber se a série em questão apresenta algum componente cíclico ou tendencial. Através da apresentação gráfica da série original é impreciso concluir algo. Porém, quando se constrói o correlograma (Figura 5), pode-se perceber facilmente que existem picos em k = 12, 24, 36, o que é indicativo de uma série com oscilação sazonal anual. De outra parte, verifica-se, também, que o coeficiente de autocorrelação não decresce rapidamente, o que é indicativo de uma série não estacionária (isto é, série que apresenta a componente tendencial).

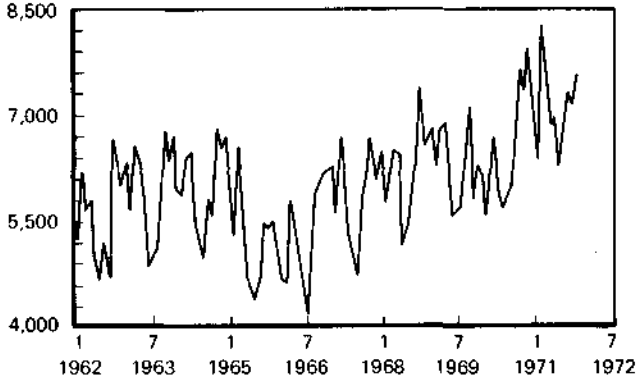


Figura 4 — Produção suína nos Estados Unidos (em milhões de cabeças por mês) — jan./62-dez./71

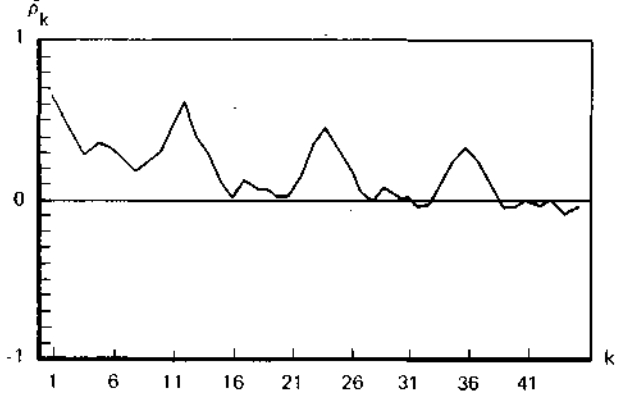


Figura 5 — Correlograma da série temporal da produção suína nos Estados Unidos, exposta na Figura 4

Para retirar a sazonalidade, cria-se uma nova série $Z_t = X_t - X_{t-12}$, conforme o explicado no decorrer deste capítulo. O correlograma dessa nova série pode ser visto na Figura 6. Nota-se que esse já não apresenta picos tão intensos, característicos da sazonalidade. Verifica-se, também, que o declínio do coeficiente de autocorrelação é lento, o que é um indicativo da permanência de componente tendencial. Constrói-se uma nova série $W_t = \nabla Z_t = \nabla(X_t - X_{t-12})$ e seu respectivo correlograma. A Figura 7 mostra um rápido declínio do coeficiente de autocorrelação, apresentando pequena flutuação em torno do ponto de estabilidade zero. A rápida queda do coeficiente de autocorrelação indica que essa nova série W_t é estacionária (sem tendência), e as pequenas variações em torno do valor zero indicam a eliminação dos picos sazonais. Assim, W_t é estacionária, não possuindo sazonalidade.

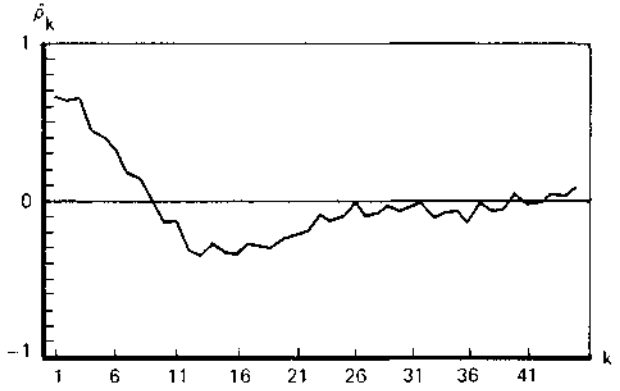


Figura 6 – Produção suína nos Estados Unidos – jan./62-dez./71. Função de autocorrelação de $(y_t - y_{t-12})$

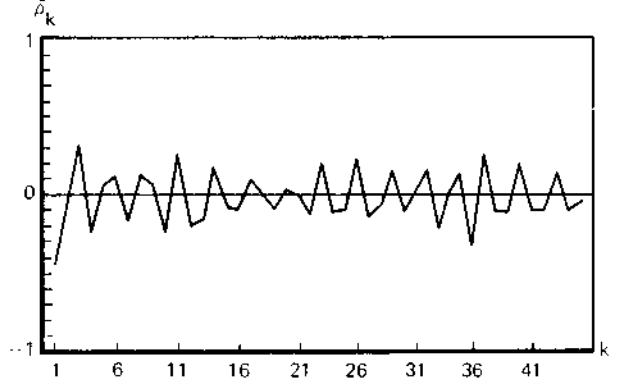


Figura 7 – Produção suína nos Estados Unidos – jan./62-dez./71. Função de autocorrelação de $\nabla(y_t - y_{t-12})$

3 – OS MODELOS DE BOX & JENKINS

3.1 – Fundamentos

Embora esteja-se tornando comum na literatura especializada denominar os modelos aqui tratados, seus processos construtivos e a inferência de previsões como de Box & Jenkins, deve-se salientar que a maioria dos modelos utilizados por esses dois autores eram bastante conhecidos já no segundo quartel deste século, bem antes, portanto, de suas pesquisas. A justificativa para o uso do nome Box & Jenkins associado a esses processos está na contribuição representativa que deram ao estudo de séries temporais, através de uma minuciosa e fundamentada integração desses processos, que lhes permitiu um tratamento analítico de inferência estatística nas previsões de valores futuros das variáveis dinâmicas.

O tratamento desenvolvido por Box & Jenkins para a análise das séries estocásticas de tempo está baseado no fato de que, embora seus respectivos valores no tempo, x_t, apresentem correlação serial, cada um deles pode ser considerado como gerado por uma seqüência de choques "a_t", $t \in T$, aleatórios e independentes entre si, cada um possuindo uma determinada distribuição[1], com média zero e variância constante σ_a^2.[2]

A seqüência de choques aleatórios "a_t", $t \in T$, com as características acima mencionadas é denominada de **processo de ruído branco**[3].

O fundamento da metodologia de Box & Jenkins, para a busca de uma classe geral de modelos capazes de representarem o processo gerador da série estocástica, está assentado no teorema da decomposição de Wold para séries estacionárias (WOLD, 1938). Esse teorema demonstra que todo o processo estocástico estacionário pode ser decomposto em um modelo linear de tipo **média móvel**, cuja formulação é

$$x_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \ldots,$$

onde μ e ψ_j, $j = 1, 2, 3, \ldots$ são parâmetros estabelecidos pelos valores da série, x_t é a observação no período "t", a_t o ruído branco no período "t" e os ψ_j devem formar uma série absolutamente convergente para que X_t seja estável (isto é, tenha variância finita e capaz de ser somável e mantenha a condição de estacionariedade.

---

[1] Em geral, cada choque "a_t" é considerado como possuindo distribuição normal.

[2] Vale dizer: $E(a_t) = 0$; $Var(a_t) = \sigma_a^2$; $Cov(a_t, a_{t-k}) = \begin{cases} \sigma_a^2, & \text{se } k = 0 \\ 0, & \text{se } k \neq 0 \end{cases}$

Conseqüentemente, a função de autocorrelação de um processo de choques "a_t", $t \in T$, é

$$\rho_k = \begin{cases} 1 & \text{se } k = 0 \\ 0 & \text{se } k \neq 0 \end{cases}$$

[3] Dá-se o nome de ruído branco a esse processo pelo fato de sua representação espectral ser bastante similar ao espectro físico da luz branca.

Nesse caso,

$$E(X_t) = E(\mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots) =$$
$$= \mu + E(a_t + \sum_{j=1}^{\infty} \psi_j a_{t-1}) = \mu \quad [4]$$

$$Var(X_t) = Var(\mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots) =$$
$$= Var(a_t) + \psi_1^2 Var(a_{t-1}) + \psi_2^2 Var(a_{t-2}) + \ldots =$$
$$= \sigma_a^2 \sum_{i=0}^{\infty} \psi_j^2 \quad , \quad \text{onde } \psi_o = 1.$$

$$Cov(X_t, X_{t-k}) = E(X_t - E(X_t))(X_{t-k} - E(X_{t-k})) =$$
$$= E\left[(a_t + \psi_1 a_{t-1} + \ldots)(a_{t-k} + \psi_1 a_{t-k-1} + \ldots)\right] =$$
$$= E\left[(\psi_k a_{t-k}^2) + (\psi_1 \psi_{k+1} a_{t-k-1}^2) + \ldots\right] =$$
$$= \sigma_a^2(\psi_k + \psi_1\psi_{k+1} + \ldots) =$$
$$= \sigma_a^2 \sum_{k=0}^{\infty} \psi_t \psi_{t+k}$$ que tem significado se a soma $\sum_{k=0}^{\infty} \psi_t \psi_{t+k}$ existe[5].

Para maior simplicidade, pode-se utilizar uma notação condensada para esse modelo linear:

$$x_t = \mu + \sum_{j=1}^{\infty} \psi_j a_{t-j} + a_t$$
$$x_t - \mu = \sum_{j=1}^{\infty} \psi j\, a_{t-j} + a_t$$
$$\tilde{x}_t = x_t - \mu = \psi(B)\, a_t, \qquad (3.1)$$

onde $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \psi_4 B^4 + \ldots$ é chamado de operador linear, função de transferência ou, simplesmente, filtro do sistema[6].

Em termos gerais, podem-se definir os seguintes operadores que serão úteis no desenvolvimento deste e dos demais capítulos:

---

[4] Deve-se ressaltar que, se a série não é estacionária, o parâmetro μ é somente um ponto de referência para o nível da série, sem ter um significado especial.

[5] Observe-se que nas condições impostas aos coeficientes ψ_j (isto é, $\Sigma\psi_j^2 < \infty$), a média e a variância são constantes e sua autocovariância depende somente de k, condições essas que definem a estacionariedade do processo.

[6] Para que um processo linear seja estacionário é necessário não só que a série $\psi(B)$ seja convergente para que se tenha variância finita. Além disso, a função de autocorrelação deve satisfazer um conjunto de condições, conforme se viu no capítulo anterior, para assegurar a estacionariedade do processo. Essas condições podem ser englobadas na condição única de que a série $\psi(B)$ deve convergir para o interior ou sobre o círculo unitário, ou seja, $|B| \leqslant 1$. (BOX & JENKINS, 1976, p. 49 e Anexo A3.1).

a) **operador de retardo**: esse operador impõe uma defasagem de um período de tempo para trás a cada vez que é usado. Assim, definindo o operador de retardo por "B", pode-se dizer que:

$$B\, x_t = x_{t-1}$$

$$B^2 x_t = x_{t-2}$$

.....

.....

$$B^m x_t = x_{t-m}$$

Note-se que esse operador foi utilizado na formulação condensada do operador linear da equação (3.1).

b) **operador de avanço**: é o operado $F = B^{-1}$, inverso de B. Ou seja, é o operador que impõe uma defasagem de um período para a frente a cada vez que é usado:

$$F\, x_t = x_{t+1}$$

$$F^2 x_t = x_{t+2}$$

.....

.....

$$F^m x_t = x_{t+m}$$

c) **operador diferença de retardo**: é o operador que impõe a diferença de um período para trás a cada vez que é usado. Seja ∇ esse operador. Então:

$$\nabla\, x_t = x_t - x_{t-1}$$

$$\nabla^2 x_t = \nabla\, x_t - \nabla\, x_{t-1}$$

.....

.....

.....

d) **operador somatório**: é o operador inverso do operador diferença $S = \nabla^{-1}$. Conforme foi visto no Capítulo II, diz-se que uma série é evolutiva homogênea de ordem "d" se $Z_t = \nabla^d X_t$ é uma série estacionária. A operação inversa, consiste em tendo-se a série Z_t, transformá-la em X_t somando Z_t "d" vezes. Assim, $X_t = \Sigma^d Z_t = S^d Z_t$, onde $\Sigma^d = S^d$ é o operador somatório.

Portanto,

$$Sz_t = \nabla z_t = \sum_{i=-\infty}^{t} z_i = \sum_{j=0}^{\infty} z_{t-j} = z_t + z_{t-1} + z_{t-2} + \ldots =$$

$$= (1 + B + B^2 + B^3 + \ldots)\, z_t =$$

$$= (1 - B - B^2 - B^3 - \ldots)^{-1} z_t = \nabla^{-1} z_t.$$

Logo, $S = \nabla^{-1}$.

$$S^2 z_t = S\, z_t + S\, z_{t-1} + S\, z_{t-2} + \ldots =$$

$$= \sum_{j=-\infty}^{t} \sum_{i=-\infty}^{j} z_i.$$

.....

.....

.....

O modelo de filtro linear exposto acima, através da equação (3.1), apresenta como característica básica o fato de, sob determinadas condições, ser equivalente a outro, denominado auto-regressivo, onde x_t é descrito como uma soma ponderada de seus valores passados mais um choque aleatório a_t. Ou seja, dada a equação (3.1),

$$x_t = \mu + a_t + \psi_1\, a_{t-1} + \psi_2\, a_{t-2} + \ldots$$

Então, pode-se obter

$$a_t = x_t - \mu - \psi_1\, a_{t-1} - \psi_2\, a_{t-2} - \ldots \qquad (3.2)$$

$$a_{t-1} = x_{t-1} - \mu - \psi_1\, a_{t-2} - \psi_2\, a_{t-3} - \ldots \qquad (3.3)$$

Substituindo (3.3) em (3.2), tem-se

$$x_t = \mu(1 - \psi_1) + \psi_1\, x_{t-1} + a_t + (\psi_2 - \psi_1^2) a_{t-2} + (\psi_3 - \psi_1\, \psi_2)\, a_{t-3} + \ldots$$

Continuando o mesmo processo de substituição para $a_{t-2}, a_{t-3}, a_{t-4}, \ldots$ em (3.2), obtém-se a equação linear composta pela soma ponderada de seus valores passados mais um choque aleatório.

$$x_t = \pi_1\, x_{t-1} + \pi_2\, x_{t-2} + \pi_3\, x_{t-3} + \ldots + \delta + a_t, \qquad (3.4)$$

onde $\delta = \mu/(1 - \pi_1 - \pi_2 - \pi_3 - \ldots)$ e os pesos π_j são funções de ψ_j.

Essa equação pode ser escrita condensadamente utilizando o operador "B":

$$x_t = \sum_{j=1}^{\infty} \pi_j\, x_{t-j} + \delta + a_t$$

$$x_t - \sum_{j=1}^{\infty} \pi_j\, x_{t-j} = \delta + a_t$$

$$\pi(B) x_t = \delta + a_t,$$

ou trabalhando com variáveis centradas na média, $\tilde{x}_t = x_t - \mu$:

$$\tilde{x}_t = \sum_{j=1}^{\infty} \pi_j \tilde{x}_{t-j} + a_t$$

$$\tilde{x}_t - \sum_{j=1}^{\infty} \pi_j \tilde{x}_{t-j} = a_t$$

$$\pi(B)\, \tilde{x}_t = a_t.$$

Nesse caso, as restrições impostas aos coeficientes devem ser tais que $\sum_{j=0}^{\infty} \pi_j$ converge. Ou seja, a série $\pi(B)$ deve convergir para o interior ou fronteira do círculo unitário, isto é, $|B| \leq 1$. Processos lineares com as condições restritivas citadas recebem o nome de processos inversíveis.

As condições de estacionariedade e inversibilidade dos modelos (3.1) e (3.4) serão analisadas de forma específica mais adiante, no decorrer deste mesmo capítulo, ao se estudar cada um dos modelos de descrição do processo estocástico original. Por ora, é importante ter presente que:

- um processo linear é estacionário se a série $\psi(B)$ converge para $|B| \leq 1$; é inversível se a série $\pi(B)$ converge para $|B| \leq 1$;
- dadas as formas condensadas das equações (3.1) e (3.4),

$$\pi(B)\, \tilde{x}_t = a_t \quad \text{e} \quad \tilde{x}_t = \psi(B)\, a_t,$$

a relação entre esses dois operadores é

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

- a condição de inversibilidade é independente da condição de estacionariedade. Esse fato pode ser visto através do seguinte exemplo:

seja um modelo linear finito do tipo (3.1):

$$x_t = \mu + a_t - \theta a_{t-1}.$$

Esse modelo apresenta $\psi_0 = 1$, $\psi_1 = -\theta$ e $\psi_j = 0$, $j>1$. Como $\sum_{j=0}^{\infty} \psi_j = 1 - \theta$, a série é convergente para qualquer valor de θ. Logo, é estacionária. Substituindo os valores de a_t e a_{t-1}, conforme o exposto anteriormente para a obtenção de um processo do tipo (3.4), obtém-se:

$$x_t = \mu - \theta\mu + a_t - \theta x_{t-1} - \theta^2 a_{t-2}.$$

Continuando a substituir a_{t-2} e seus sucessivos resultados a_{t-3}, a_{t-4}, a_{t-5}, ..., tem-se:

$$x_t = -\theta x_{t-1} - \theta^2 x_{t-2} - \theta^3 x_{t-3} - \ldots + \mu/(1-\theta) + a_t.$$

Note-se que, para esse processo, os pesos π_j de (3.4) são $\pi_j = -\theta^j$. Ora, se $|\theta| > 1$ os pesos crescem exponencialmente com o deslocamento do tempo, a série $\Sigma\pi_j$ é divergente, a variância do modelo "explode" (não é finita), e os maiores pesos são dados para os valores mais distantes no tempo. Para evitar essas contradições, impõe-se a condição de inversibilidade $|\theta| < 1$.

A proposta de Box & Jenkins para a modelagem do processo utiliza o princípio da "parcimônia"[7], buscando a descrição aproximada do modelo linear de decomposição de Wold.

Box & Jenkins, percebendo a relação dual existente entre os processos média móvel e auto-regressivo, propuseram-se a trabalhar com a postulação de que os dados são gerados por processos lineares do tipo equação média móvel (3.1), do tipo equação auto-regressiva (3.4) ou por uma classe de processos mistos, compostos pela utilização conjunta dos dois processos, denominados **modelo ARMA** (iniciais de "autoregressive-moving average").

A justificativa principal para a inclusão de ambos os processos na descrição do modelo está em que, muitas vezes, as características do processo estocástico não permitem uma descrição satisfatória através de um processo puramente auto-regressivo ou puramente média móvel, de poucos parâmetros, porque a série original incorpora, em sua forma parcimoniosa, ambos os tipos de processos. A existência da dualidade entre eles se dá de forma que um processo média móvel finito (de qualquer ordem) pode ser descrito como um processo auto-regressivo infinito, e vice-versa[8]. Isso implica, em termos práticos, que a obtenção de estimativas de seus parâmetros, de forma parcimoniosa, muitas vezes requer a inclusão de ambos os processos no modelo.

---

[7] Isto é, que o modelo contenha um número de parâmetros tão pequeno quanto possível.

[8] Seja um processo auto-regressivo. No momento "t" pode-se defini-lo por

$$\tilde{x}_t = \Sigma\phi_j\tilde{x}_{t-j} + a_t.$$

Da mesma forma, um processo média móvel, no momento "t" vale

$$\tilde{x}_t = -\Sigma\theta a_{t-j} + a_t.$$

A dualidade entre eles pode ser percebida de imediato, utilizando o operador "B". Seja, por exemplo, um processo auto-regressivo finito de ordem 1. Então

$$\tilde{x}_t = \phi x_{t-1} + a_t.$$

Conseqüentemente,

$$(1 - \phi B)\tilde{x}_t = a_t.$$

Ou, o que é o mesmo:

$$\tilde{x}_t = (1 - \phi B)^{-1} a_t = (1 + \phi B + \phi B^2 + \ldots)a_t = \Sigma\phi^j a_{t-j} + a_j,$$

o que equivale a um processo média móvel de ordem infinita (tipo 3.1). Semelhantemente, conforme foi dito no exemplo da p.43, verifica-se a recíproca: um processo média móvel finito equivale a um processo auto-regressivo infinito.

Essa classe de modelos pode ampliar sua capacidade de ação para processos não estacionários homogêneos, bastando para tanto ser acrescida de um operador que seja capaz de transformar processos evolutivos homogêneos de ordem "d" em estacionários, o qual, como vimos no capítulo anterior (seção 2.1.4), é o operador somatório S^d, inverso do operador diferença de retardo, ∇^d. A essa classe de modelos não estacionários, que utilizam os princípios desenvolvidos no modelo ARMA, acrescidos de um operador de transformação de processos evolutivos homogêneos em estacionários, dá-se a denominação de **modelo ARIMA** (iniciais de "autoregressive-integrated moving average").

A razão para a inclusão da palavra "integrated" no modelo ARIMA está no fato de que a existência de um processo não estacionário exige o operador somatório. Como se sabe, o operador somatório para variáveis contínuas é o operador integral, de onde surgiu a inclusão do termo "integrated".

Sinteticamente, pode-se assim especificar a proposta de Box & Jenkins: enquanto o teorema da decomposição de Wold supõe que o processo de ruído branco é transformado na observação x_t, através de um filtro linear, conforme esquema apresentado na Figura 1, a metodologia de Box & Jenkins supõe que o processo de filtragem do ruído branco pode ser decomposto em três etapas, conforme a Figura 2 está a indicar: um processo de filtragem "média móvel", através do qual o ruído branco é transformado em e_t; um processo de filtragem "auto-regressivo estacionário", transformando e_t em w_t; e, caso a série seja não estacionária, um processo de filtragem "somatório" transformando w_t na observação da série temporal x_t.

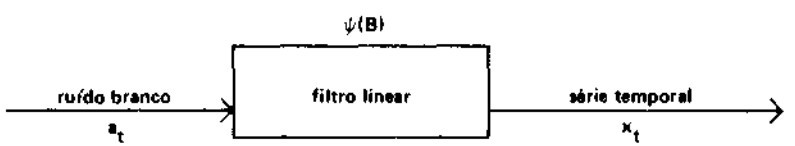


Figura 1 – Diagrama de um modelo de filtro linear para séries temporais

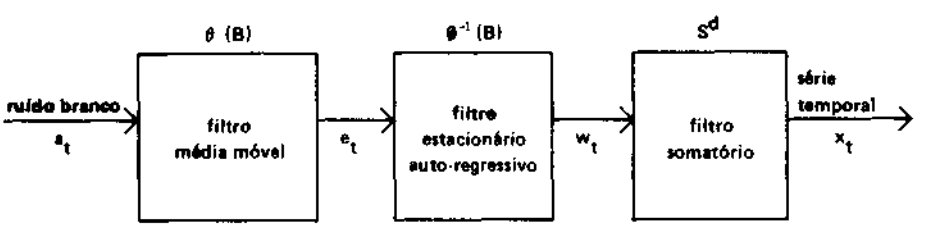


Figura 2 – Diagrama do processo ARIMA para modelagem de séries temporais

3.2 – Modelo Média Móvel

Seja a decomposição do teorema de Wold, dado através do processo linear estabelecido pela equação (3.1). Considerando que somente as primeiras "q" ponderações são diferentes de zero, então o processo

$$\begin{aligned}\tilde{x}_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} = \\ &= (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) a_t = \\ &= \theta(B)\, a_t\end{aligned}$$

é chamado de processo média móvel de ordem "q", simbolicamente definido por MA(q).

Assim, em um processo média móvel de ordem "q", cada observação da série é gerada por uma média ponderada do presente e "q" valores passados de um processo de ruído branco. Em outras palavras, um processo média móvel é um processo de filtro linear do tipo

$$\tilde{x}_t = a_t - \sum_{j=1}^{q} \theta_j a_{t-j}.$$

Em geral, em sua postulação parcimoniosa adquirem particular importância os processos de primeira e segunda ordem: MA(1) e MA(2).

O processo média móvel de ordem "q", MA(q), apresenta as seguintes características:

- a média é independente do tempo, $E(x_t) = \mu$;
- como cada a_t é gerado pelo mesmo processo de ruído branco, $E(a_t) = 0$, $E(a_t^2) = \sigma_a^2$ e $E(a_t, a_{t-k}) = 0$, para $k \neq 0$, então o processo é descrito por q + 2 parâmetros que são a média μ, a variância do ruído σ_a^2 e os pesos $\theta_1, \theta_2, \theta_3, \ldots, \theta_q$;
- a variância do processo é:

$$\begin{aligned}\mathrm{Var}(x_t) = \mathrm{Var}(\tilde{x}_t) &= E(a_t^2 + \theta_1^2 a_{t-1}^2 + \ldots + \theta_q^2 a_{t-q}^2 - 2\theta_1 a_t a_{t-1} - \ldots) = \\ &= E(a_t^2) + E(\theta_1^2 a_{t-1}^2) + \ldots + E(\theta_q^2 a_{t-q}^2) - E(2\theta_1 a_t a_{t-1}) - \ldots = \\ &= \sigma_a^2 + \theta_1^2 \sigma_a^2 + \ldots + \theta_q^2 \sigma_a^2 - 0 - \ldots - 0) = \\ &= \sigma_a^2 (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2);\end{aligned}$$

- a função de autocorrelação é dada por:

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2}, & \text{se } k = 1,2,\ldots,q \\ 0, & \text{se } k > q. \end{cases}$$

Deve-se observar que um processo média móvel de ordem "q" apresenta coeficiente de autocorrelação igual a zero para k>q. Ou seja, o processo estocástico no período "t", x_t, só está correlacionado com $x_{t\pm1}$, $x_{t\pm2}$, $x_{t\pm3}$, ..., $x_{t\pm q}$. Para os demais valores da série não há correlação. Isso significa que o processo não sofre influência de eventos que ocorram além de "q" períodos de distância. Diz-se, então, que o processo MA(q) possui uma "memória" de "q" períodos, pois desconsidera, ou melhor dizendo, esquece o que ocorreu a distâncias superiores a "q" períodos. Deve-se notar, também, que a obtenção dos valores da função de autocorrelação se apresenta útil à medida que permite especificar a ordem "q" do processo média móvel.

Condições de inversibilidade e estacionariedade

Seja um processo média móvel de ordem "q", então:

$$\tilde{x}_t = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} =$$

$$= (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) a_t =$$

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

$$\theta^{-1}(B)\, \tilde{x}_t = a_t$$

$$\pi(B)\, \tilde{x}_t = a_t, \text{ onde } \pi(B) = \theta^{-1}(B).$$

Conforme se viu no item 3.1, a condição de inversibilidade do processo linear é que $\pi(B)$ convirja para o interior ou fronteira do círculo unitário. No caso de um MA(q),

$$\pi(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)^{-1} = \theta^{-1}(B).$$

Se $\theta^{-1}(B)$ deve convergir, então necessariamente as raízes da equação característica $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q = 0$ devem estar todas fora do círculo unitário. Ou seja, os valores solução B_i, i = 1, 2, ..., "q" da equação acima devem ser todos maiores que um, em valor absoluto.

Para que haja estacionariedade, note-se que um MA de ordem "q" implica que a série $\Psi(B) = \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ seja finita. Então, as condições impostas para estacionariedade de um processo linear, expostas no item 3.1, não são necessárias para um processo MA(q) garantir sua estacionariedade. Um MA(q), com "q" finito, **será sempre estacionário.**

Os modelos média móvel usuais em séries econômicas, MA(1) e MA(2), apresentam as seguintes características:

Processo Média Móvel de Primeira Ordem: MA (1)

Modelo

$$x_t = \mu + a_t - \theta_1 a_{t-1}$$

Condições de estacionariedade

Como o processo é de ordem finita, o MA(1) é sempre estacionário.

Condições de inversibilidade

Para que o processo seja inversível, a série $\pi(B)$ deve convergir para $|B| \leq 1$. Como $\pi(B) = \theta^{-1}(B)$, as raízes da equação característica $\theta(B) = 0$ devem cair fora do círculo unitário.

Para um MA(1), $\theta(B) = (1 - \theta_1 B)$. Como a raiz vale θ_1^{-1}, então $|\theta_1| < 1$.

Média

$$E(X_t) = E(\mu + a_t - \theta_1 a_{t-1}) =$$

$$= E(\mu) + E(a_t) - E(\theta_1 a_{t-1}) =$$

$$= \mu + 0 - 0 =$$

$$= \mu$$

Variância

$$Var(X_t) = E(X_t - u)^2 = E(a_t - \theta_1 a_{t-1})^2 =$$

$$= E(a_t^2 - 2a_t\theta_1 a_{t-1} + \theta^2 a_{t-1}^2) =$$

$$= \sigma_a^2 - 0 + \theta_1^2 \sigma_a^2 =$$

$$= (1 - \theta_1^2)\, \sigma_a^2$$

Coeficiente de autocovariância

$$\gamma_o = Var(X_t) = (1 - \theta_1^2)\, \sigma_a^2$$

$$\gamma_1 = E(a_t - \theta_1 a_{t-1})(a_{t-1} - \theta_1 a_{t-2}) =$$

$$= - \theta_1$$

$$\gamma_j = E(a_t - \theta_1 a_{t-1})(a_{t-j} - \theta_1 a_{t-j}) =$$

$$= 0, \text{ quaisquer } j > 1.$$

Coeficiente de autocorrelação

$$\rho_1 = - \theta_1/(1 + \theta_1^2)$$

$$\rho_j = 0 \quad , \quad \text{qualquer } j > 1.$$

Processo Média Móvel de Segunda Ordem: MA (2)

Modelo

$$x_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

Condições de estacionariedade

Como o processo é de ordem finita, o MA(2) é sempre estacionário.

Condições de inversibilidade

As raízes da equação característica $\theta(B)=1-\theta_1 B-\theta_2 B^2=0$ devem cair fora do círculo unitário. Então devem satisfazer a:

$$\theta_2 + \theta_1 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$-1 < \theta_2 < 1$$

Média

$$E(X_t) = E(\mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}) =$$

$$= \mu + E(a_t) - \theta_1 E(a_{t-1}) - \theta_2 E(a_{t-2}) =$$

$$= \mu + 0 - 0 - 0 =$$

$$= \mu$$

Variância

$$Var(X_t) = E(X_t - \mu)^2 = E(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2})^2 =$$

$$= E(a_t^2 + \theta_1^2 a_{t-1}^2 + \theta_2^2 a_{t-2}^2 - 2\theta_1 a_t a_{t-1} -$$

$$- 2\theta_2 a_t a_{t-2} - 2\theta_1 \theta_2 a_{t-1} a_{t-2}) =$$

$$= \sigma_a^2 + \theta_1^2 \sigma_a^2 + \theta_2^2 \sigma_a^2 - 0 - 0 + 0 =$$

$$= (1 + \theta_1^2 + \theta_2^2)\, \sigma_a^2$$

Coeficientes de autocovariância

$$\gamma_0 = Var(X_t) = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2$$

$$\gamma_1 = E(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2})(a_{t-1} - \theta_1 a_{t-2} - \theta_2 a_{t-3}) =$$

$$= -\theta_1 \sigma_a^2 + \theta_2 \theta_1 \sigma_a^2 =$$

$$= -\theta_1 (1 - \theta_2)\, \sigma_a^2$$

$$\gamma_2 = E(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2})(a_{t-2} - \theta_1 a_{t-3} - \theta_2 a_{t-4}) =$$

$$= - \theta_2 \sigma_a^2$$

$$\gamma_j = 0 \quad , \quad \text{qualquer } j > 2.$$

Coeficientes de autocorrelação

$$\rho_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_2 = \frac{- \theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_j = 0 \quad , \quad \text{qualquer } j > 2.$$

A figura a seguir ilustra as várias possibilidades para θ_1 e θ_2 que satisfazem a condição de inversibilidade de um MA(2).

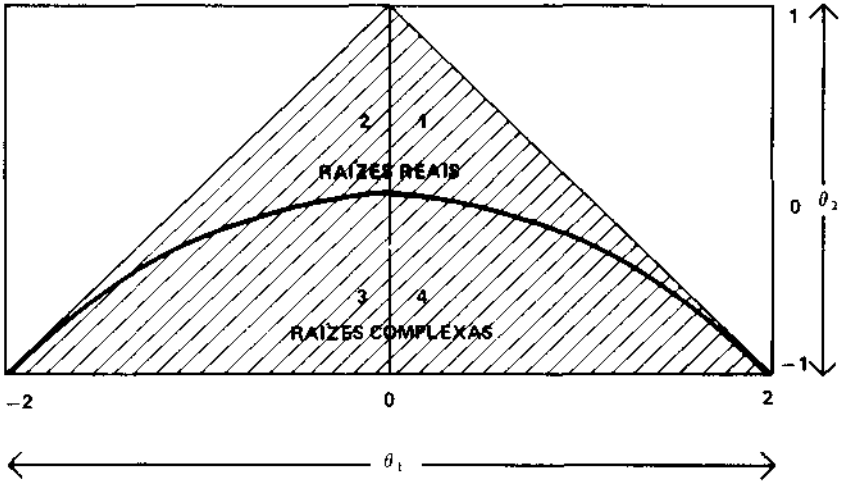


Figura 3 – Regiões de admissão para que um MA(2) seja inversível

3.3 – Modelo Auto-Regressivo

Seja a decomposição do teorema de Wold, dado através do processo linear estabelecido pela equação (3.4). Considerando unicamente as primeiras "p" ponderações diferentes de zero, então a observação x_t é gerada pela média ponderada de, somente, as "p" primeiras observações próximas passadas da variável, acrescidas de uma "disturbância" aleatória do período presente a_t. Então, o processo

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \ldots + \phi_p x_{t-p} + \delta + a_t$$

é chamado de processo auto-regressivo de ordem "p", simbolicamente definido por AR(p).

Na postulação parcimoniosa de modelos, adquirem especial importância os processos auto-regressivos de primeira e segunda ordem, AR(1) e AR(2).

O modelo auto-regressivo de ordem "p", AR(p), apresenta as seguintes características:

- a média independe do tempo, se e somente se, o processo for estacionário. Nesse caso, como $E(x_t) = E(x_{t-1}) = \ldots = E(x_{t-p}) = \mu$, então,

$$E(x_t) = E(\phi_1 x_{t-1}) + E(\phi_2 x_{t-2}) + \ldots + E(\phi_p x_{t-p}) + E(\delta) + E(a_t)$$

$$\mu = \phi_1 \mu + \phi_2 \mu + \phi_3 \mu + \ldots + \phi_p \mu + \delta + 0$$

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \ldots - \phi_p} .$$

Se o processo é estacionário, a média $E(x_t) = \mu$ deve ser uma constante. Portanto, 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

$$\phi_1 + \phi_2 + \ldots + \phi_p < 1$$

Conseqüentemente, se o processo auto-regressivo for estacionário, a média será zero somente se o termo constante δ for igual a zero.

Condições de inversibilidade e estacionariedade

Seja um processo auto-regressivo de ordem "p". Então,

$$\tilde{x}_t = \phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + \ldots + \phi_p \tilde{x}_{t-p} + at.$$

Conforme se viu no item 3.1, para que o processo linear definido pela equação (3.1) seja estacionário, é necessário que $\Psi(B)$ convirja para $|B| \leq 1$. Ou seja, como $\Psi(B) = \phi^{-1}(B)$, sejam G_i^{-1}, $i = 1, 2, 3, \ldots, p$, as raízes de $\phi(B) = 0$. Então, $\phi(B) = (1-G_1B)\ (1-G_2B)\ \ldots\ (1-G_pB)$ e, expandindo em frações parciais, tem-se o modelo AR(p):

$$X_t = \Psi(B)a_t = \phi^{-1}(B)a_t = \sum_{i=1}^{p} \frac{k_i}{1 - G_iB}\ a_t .$$

Como para $\Psi(B) = \phi^{-1}(B)$ convergir deve-se ter, necessariamente, $|B| \leq 1$, então $|G_i| < 1$, $i = 1, 2, 3, \ldots, p$, o que é equivalente a dizer que as raízes de $\phi(B) = 0$ devem cair fora do círculo unitário.

Para a condição de inversibilidade, note-se que um AR(p), com "p" finito, implica que

$$\pi(B) = \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_2 B^2$$

seja finito. Então, as condições impostas para a inversibilidade do processo linear, expostas no item 3.1, não são requeridas para um AR(p) com "p" finito. O processo AR(p) sempre será inversível.

Deve-se ressaltar que as duas formas equivalentes de apresentação do modelo linear geral,

$$\pi(B)\tilde{x}_t = a_t \quad \text{e} \quad \tilde{x}_t = \Psi(B)a_t,$$

onde

$$\Psi(B) = \pi^{-1}(B),$$

refletem a dualidade existente entre esses modelos lineares. Pode-se perceber a dualidade entre a estacionariedade e a inversibilidade do processo, as quais são caracterizadas em função da convergência das séries $\Psi(B)$ e $\pi(B)$ respectivamente. A estacionariedade funciona para os processos auto-regressivos, assim como a inversibilidade funciona para os processos média móveis.

Processo Auto-Regressivo de Primeira Ordem: AR (1)

Modelo

$$x_t = \phi_1 x_{t-1} + \delta + a_t$$

Condições de estacionariedade

A raiz da equação característica $\phi(B)=0$ deve ser tal que $|\phi_1| < 1$.

Condições de inversibilidade

Como o processo é de ordem finita, o AR(1) é sempre inversível.

Média

$$\begin{aligned} E(x_t) &= E(\phi_1 x_{t-1} + \delta + a_t) = \\ &= \phi_1 \mu - \delta - E(a_t) = \\ &= \delta/(1 - \phi_1) \end{aligned}$$

Variância

$$\begin{aligned} E(x_t - \mu)^2 &= E(\phi_1 x_{t-1} + \delta + a_t - \delta/(1-\phi_1))^2 = \\ &= E(\phi_1 x_{t-1} + a_t)^2 = \\ &= \phi_1^2 E(x_{t-1})^2 + E(a_t^2) + 2\phi_1 E(x_{t-1} a_t) = \\ &= \phi_1^2 \sigma_a^2 + \sigma_a^2 - 0 = \\ &= \sigma_a^2/(1 - \phi_1^2) \end{aligned} \qquad (3.5)$$

Coeficientes de covariância

$$\begin{aligned} \gamma_k &= E(\tilde{x}_t \cdot \tilde{x}_{t-k}) = E\left[(\phi_1 \tilde{x}_{t-1} + a_t)\, \tilde{x}_{t-k}\right] = \\ &= E(\phi_1 \tilde{x}_{t-1} \cdot \tilde{x}_{t-k} + a_t \tilde{x}_{t-k}) = \phi_1 E(\tilde{x}_{t-1} \cdot \tilde{x}_{t-k}) + \\ &+ E(a_t \cdot \tilde{x}_{t-k}) = \\ &= \phi_1 E(\tilde{x}_{t-1} \cdot \tilde{x}_{t-k}) + 0 = \\ &= \phi_1 \gamma_{k-1} \end{aligned}$$

Assim, para $k=1$, $\gamma_1 = \phi_1 \gamma_0$

$k=2$, $\gamma_2 = \phi_1 \gamma_1 = \phi_1(\phi_1^2 \gamma_0) = \phi_1^2 \gamma_0$

$k=3$, $\gamma_3 = \phi_1 \gamma_2 = \phi_1(\phi_1^2 \gamma_0) = \phi_1^3 \gamma_0$.

Genericamente, $\gamma_k = \phi_1^k \gamma_0$

ou, substituindo em (3.5),

$$\gamma_k = \phi_1^k \gamma_0 = \frac{\phi_1^k \sigma_a^2}{1 - \phi_1^2} .$$

Coeficiente de autocorrelação

$$\rho_k = \frac{\gamma_k}{\gamma_o} = \frac{\phi_1^k \gamma_o}{\gamma_o} = \phi_1^k \text{, qualquer } k .$$

Note-se que, como $\rho_k = \phi_1^k$, a observação presente depende de todos os valores passados, embora seu grau de dependência decline com o tempo. Conseqüentemente, o processo auto-regressivo estacionário de primeira ordem tem memória infinita.

Processo Auto-Regressivo de Segunda Ordem: AR (2)

Modelo

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \delta + a_t$$

Condições de estacionariedade

As raízes da equação característica $\phi(B) = 0$ devem ser tais que caiam fora do círculo unitário. Isso é equivalente a

$$\phi_1 + \phi_2 < 1$$

$$\phi_2 - \phi_1 < 1$$

$$|\phi_2| < 1$$

Condições de inversibilidade

Como o processo é de ordem finita, o AR(2) é sempre inversível.

Média

$$E(x_t) = E(\phi_1 x_{t-1} + \phi_{t-2} + \delta + a_t) =$$

$$= \delta/(1 - \phi_1 - \phi_2)$$

Variância

$$E(x_t - \mu)^2 = E(\tilde{x}_t)^2 = E(\phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + a_t)^2$$

$$= E(\phi_1^2 \tilde{x}_{t-1}^2 + \phi_2^2 \tilde{x}_{t-2}^2 + a_t^2 + 2\phi_1 \tilde{x}_{t-1} \phi_2 \tilde{x}_{t-2} + 2\phi_1 \tilde{x}_{t-1} a_t +$$

$$+ 2\phi_2 \tilde{x}_{t-2} a_t) =$$

$$= \phi_1^2 E(\tilde{x}_{t-1}^2) + \phi_2^2 E(\tilde{x}_{t-2}^2) + E(a_t^2) + 2\phi_1 \phi_2 E(\tilde{x}_{t-1} \tilde{x}_{t-2}) =$$

$$= \phi_1^2 \gamma_o + \phi_2^2 \gamma_o + 2\phi_1 \phi_2 \gamma_1 + \sigma_a^2 \qquad (3.6)$$

Coeficientes de autocovariância

$$\gamma_k = E(\tilde{x}_t . \tilde{x}_{t-k}) = E\left[(\phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + a_t) x_{t-k}\right] =$$
$$= \phi_1 E(\tilde{x}_{t-1} . \tilde{x}_{t-k}) + \phi_2 E(\tilde{x}_{t-2} . \tilde{x}_{t-k}) + E(a_t . x_{t-k}) =$$
$$= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}$$

Note-se que:

$$\text{se } k = 0, \ \gamma_o = E(\tilde{x}_t . \tilde{x}_t) = E\left[\tilde{x}_t (\phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + a_t)\right] =$$
$$= \phi_1 E(\tilde{x}_t . \tilde{x}_{t-1}) + \phi_2 E(\tilde{x}_t . \tilde{x}_{t-2}) + E(a_t . \tilde{x}_t) =$$
$$= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_a^2 ;$$

$$\text{se } k = 1, \ \gamma_1 = E(\tilde{x}_t . \tilde{x}_{t-1}) = E\left[\tilde{x}_{t-1} (\phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + a_t)\right] =$$
$$= \phi_1 E(\tilde{x}_{t-1} . \tilde{x}_{t-1}) + \phi_2 E(\tilde{x}_{t-1} . \tilde{x}_{t-2}) + E(a_t . x_{t-1}) =$$
$$= \phi_1 \gamma_o + \phi_2 \gamma_1 \text{ ou } \gamma_1 = \frac{\phi_1 \gamma_o}{1 - \phi_2} ; \text{ e}$$

$$\text{se } k = 2, \ \gamma_2 = E(\tilde{x}_t . \tilde{x}_{t-2}) = E\left[\tilde{x}_{t-2} (\phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + a_t)\right] =$$
$$= \phi_1 E(x_{t-1} . x_{t-2}) + \phi_2 E(x_{t-2} . x_{t-2}) + E(a_t . x_{t-2}) =$$
$$= \phi_1 \gamma_1 + \phi_2 \gamma_o .$$

Genericamente,

$$\text{se } k \geq 2, \ \gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} . \qquad (3.7)$$

Coeficiente de autocorrelação

$$\rho_k = \frac{\gamma_k}{\gamma_o} \qquad \gamma_k = \rho_k \gamma_o$$

Substituindo em (3.5),

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} .$$

Note-se que:

$$\text{se } k = 0, \ \rho_o = \frac{\gamma_o}{\gamma_o} = 1;$$

$$\text{se } k = 1, \ \rho_1 = \phi_1 \rho_o + \phi_2 \rho_{-1} = \phi_1 \rho_o + \phi_2 \rho_1 = \frac{\phi_1}{1 - \phi_2} ;$$

$$\text{se } k = 2, \quad \rho_2 = \phi_1\rho_1 + \phi_2\rho_o =$$

$$= \phi_1 \frac{\phi_1}{1-\phi_2} + \phi_2 =$$

$$= \frac{\phi_1^2}{1-\phi_2} + \phi_2 \; ; \; e$$

$$\text{se } k > 2, \quad \rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} .$$

Em geral, para um processo AR(p) estacionário, tem-se que sua covariância vale

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \ldots + \phi_p\gamma_{k-p}, \quad k > 0 ,$$

e seu coeficiente de autocorrelação vale

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots \phi_p\rho_{k-p}, \quad k > 0 \quad ^{9} ,$$

onde os parâmetros $\phi_1, \phi_2, \ldots, \phi_p$ podem ser obtidos através das chamadas equações de Yule-Walker. Ou seja, se $\rho_1, \rho_2, \rho_3, \ldots, \rho_p$ são conhecidos (e, em geral, o são através da função amostral de autocorrelação), então o conjunto de equações ρ_k, para k = 1, 2, 3, ..., p:

$$\begin{aligned}
\rho_1 &= \phi_1\rho_o + \phi_2\rho_1 + \ldots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2\rho_o + \ldots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots + \phi_p\rho_o,
\end{aligned}$$

chamadas equações de Yule-Walker, pode oferecer os valores dos parâmetros através da resolução simultânea do sistema.

---

[9] Para se obter essas relações, basta tomar o processo

$$\tilde{x}_t = \phi_1\tilde{x}_{t-1} + \phi_2\tilde{x}_{t-2} + \ldots + \phi_p\tilde{x}_{t-p} + a_t$$

e multiplicar ambos os membros por $\tilde{x}_{t-k}$ e utilizar o operador expectância:

$$E(\tilde{x}_t \cdot \tilde{x}_{t-k}) = \phi_1 E(\tilde{x}_{t-1}\tilde{x}_{t-k}) + \phi_2 E(\tilde{x}_{t-2}\tilde{x}_{t-k}) + \ldots + \phi_p E(\tilde{x}_{t-p}\tilde{x}_{t-k}) + E(a_t \cdot x_{t-k})$$

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \ldots + \phi_p\gamma_{k-p}$$

como

$$\rho_k = \frac{\gamma_k}{\gamma_o} .$$

basta dividir a equação acima por γ_o para se obter o coeficiente de autocorrelação ρ_k:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} .$$

Sinteticamente, podem-se expressar as equações do Yule-Walker por

$$\Phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}; \quad \rho_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}; \quad P_p = \begin{bmatrix} \rho_o & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & \rho_o & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & \rho_o \end{bmatrix}$$

e a solução do sistema é dada, conseqüentemente, por

$$\Phi = P_p^{-1} \cdot \rho_p$$

3.4 – Modelo Auto-Regressivo – Média Móvel

Seja um processo estocástico estacionário. Se o processo apresenta características que não permitem sua descrição através de uma parametrização parcimoniosa de um processo puramente auto-regressivo ou puramente média móvel, pelo fato de seu comportamento incluir características de ambos os tipos de processos, então, considerando um modelo misto, obtido pela agregação desses dois processos, auto-regressivo e média móvel, tem-se:

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \delta + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \qquad (3.8)$$

$$\text{ou } \phi(B)x_t = \delta + \theta(B)a_t \qquad (3.9)$$

que é chamado de processo misto auto-regressivo-média móvel de ordem "p" e "q", simbolicamente definido por ARMA (p,q)[10].

Note-se que as equações (3.8) e (3.9) podem ser expressas de forma simplificada, através do uso de variáveis centradas na média. Então,

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

ou

$$\varphi(B)\tilde{x}_t = \theta(B)a_t .$$

[10] A justificativa da utilização desse modelo está em que, para muitas séries encontradas na vida real, a inclusão de ambos os termos (auto-regressivo e média móvel) resulta em um modelo com menor número de parâmetros do que seria necessário para descrever satisfatoriamente através de modelos puros auto-regressivos ou média móvel.

Portanto,

$$\tilde{x}_t = \phi_1^{-1}(B)\ \theta(B)\ a_t$$

$$\tilde{x}_t = \frac{\theta(B)}{\phi(B)}\ a_t = \frac{1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p}\ a_t .$$

Assim, um processo misto auto-regressivo estacionário-média móvel pode ser suposto como sendo gerado de várias maneiras:

- cada observação da série $\tilde{x}_t$ é gerada por um filtro linear cuja função de transferência é a razão de dois polinômios, $\theta(B)$ de grau "q" e $\phi(B)$ de grau "p", quando o "input" é um processo de ruído branco;
- como um processo auto-regressivo de ordem "p",

$$\phi(B)\tilde{x}_t = e_t$$

  onde, posteriormente, e_t segue um processo média móvel de ordem "q",

$$e_t = \theta(B)a_t ;$$

- como um processo média móvel de ordem "q",

$$\tilde{x}_t = \theta(B)b_t$$

  onde, posteriormente, b_t segue um processo auto-regressivo de ordem "p".

$$\phi(B)\ b_t = a_t .$$

  Portanto,

$$\phi(B)\ \tilde{x}_t = \theta(B)\ \phi(B)\ b_t = \theta(B)\ a_t .$$

O processo misto auto-regressivo estacionário-média móvel, ARMA (p, q), apresenta as seguintes características.

Média

$$E(x_t) = E(\phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \delta + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}) =$$

$$= \phi_1 E(x_{t-1}) + \ldots + \phi_p E(x_{t-p}) + \delta + E(a_t) - \theta_1 E(a_{t-1}) - \ldots - \theta_q E(a_{t-q}) =$$

$$= \phi_1 \mu + \ldots + \phi_p \mu + \delta + 0 + 0 - \ldots - 0 =$$

$$= \delta/(1 - \phi_1 - \phi_2 - \ldots - \phi_p)$$

Condições de inversibilidade e estacionariedade

Como os termos do processo média móvel da equação (3.8) não impõem nenhuma exigência para a estacionariedade do processo misto ARMA (p,q), então esse terá suas condições de estacionariedade estabelecidas somente pela parcela do processo auto-regressivo. Assim, a equação

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

será originária de um processo estacionário se $\phi(B) = 0$ tiver todas as suas raízes fora do círculo unitário. De forma semelhante, para o processo ser inversível, exige-se que as raízes de $\theta(B) = 0$ fiquem fora do círculo unitário.

Funções de autocovariância e autocorrelação

Obtêm-se de forma semelhante à utilizada para um AR(p). Seja o processo ARMA(p,q), dado por

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

Multiplicando-se ambos os membros por $\tilde{x}_{t-k}$ e, após, utilizando o operador expectância, tem-se

$$E(\tilde{x}_t.\tilde{x}_{t-k}) = \phi_1 E(\tilde{x}_{t-1}.\tilde{x}_{t-k}) + \ldots + \phi_p E(\tilde{x}_{t-p}.\tilde{x}_{t-k}) +$$

$$+ E(a_t.\tilde{x}_{t-k}) - \theta_1 E(a_{t-1}.\tilde{x}_{t-k}) - \ldots - \theta_q E(a_{t-q}\tilde{x}_{t-q})$$

$$\gamma_k = \phi_1\gamma_{k-1} + \ldots + \phi_p\gamma_{k-p} + \gamma_{xa}(k) - \theta_1\gamma_{xa}(k-1) -$$

$$- \ldots - \theta_q\gamma_{xa}(k-q), \qquad (3.10)$$

onde $\gamma_{xa}(k) = E(\tilde{x}_{t-k}.a_t) \Rightarrow \begin{cases} = 0 & \text{se } k > 0 \\ \neq 0 & \text{se } k \leq 0\,. \end{cases}$

Conseqüentemente,

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \ldots + \phi_p\gamma_{k-p} \qquad \text{se } k > q.$$

Similarmente, como $\rho_k = \dfrac{\gamma_k}{\gamma_o}$, então,

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} \qquad \text{se } k > q \qquad (3.11)$$

ou $\qquad \phi(B)\rho_k = 0 \qquad \text{se } k > q\,.$

Para $k \leq q$, o processo misto auto-regressivo estacionário-média móvel terá "q" coeficientes de autocorrelação, ρ_q, ρ_{q-1}, ..., ρ_{q-2}, ..., ρ_1, cujos valores dependerão tanto dos parâmetros do processo média móvel θ_1, θ_2, θ_3, ..., θ_q, como dos "p" parâmetros auto-regressivos ϕ_1, ϕ_2, ϕ_3, ..., ϕ_p.

Na prática, os modelos ARMA(p,q) conseguem descrever satisfatoriamente as séries de tempo com um número bastante reduzido de parâmetros. Em geral satisfazem à condição de $p + q \leq 2$.

O processo misto auto-regressivo-média móvel mais freqüentemente encontrado, o ARMA(1,1), apresenta as seguintes características.

Modelo

$$x_t = \phi_1 x_{t-1} + \delta + a_t - \theta_1 a_{t-1} \qquad (3.12)$$

ou

$$x_t - \phi_1 x_{t-1} = a_t - \theta_1 a_{t-1} + \delta$$

$$(1 - \phi_1 B)x_t = (1 - \theta_1 B)a_t + \delta$$

Condição de estacionariedade

Para um processo ser estacionário é necessário que a série $\Psi(B)$ convirja. Observe-se que, por (3.12), o valor de "x" no tempo "t-1" vale

$$x_{t-1} = \phi_1 x_{t-2} + \delta + a_t - \theta_1 a_{t-2}$$

Substituindo x_{t-1} e seus sucessivos resultados em (3.12), obtém-se o processo escrito segundo a forma média móvel infinita (tipo equação 3.1):

$$x_t = \phi_1(\phi_1 x_{t-2} + \delta + a_t - \theta_1 a_{t-2}) = \phi_1^2 x_{t-2} + \phi_1\delta + \phi_1 a_1 - \phi_1\theta_1 a_{t-2} + \delta + a_t - \phi_1 a_{t-2} =$$

$$= \phi^3 x_{t-3} + (\delta + \phi_1\delta + \phi_1^2\delta) + a_t + (\phi_1 - \theta_1)a_{t-1} + \phi_1(\phi_1 - \theta_1)a_{t-2} - \phi_1^2\theta_1 a_{t-3}$$

$$\vdots$$

$$= \frac{\delta}{1-\phi_1} + a_t + (\phi_1 - \theta_1)a_{t-1} + \phi_1(\phi_1 - \theta_1)a_{t-2} + \phi_1^2(\phi_1 - \theta_1)a_{t-3} + \ldots,$$

onde os $\psi_j = \phi_1^{j}(\phi_1 - \theta_1)$.

Dessa maneira, a condição de estacionariedade torna-se satisfeita se $\psi_j = \sum_{j=1}^{\infty} \phi_1^{j-1}(\phi_1 - \theta_1)$ converge. Para tanto, $|\phi_1| < 1$, que é a condição de estacionariedade requerida para um AR(1).

Condição de inversibilidade

Para um processo ser inversível é necessário que $\pi(B)$ convirja.

Semelhantemente ao que foi desenvolvido para o estabelecimento da condição de estacionariedade, pode-se escrever a equação (3.12) como um processo auto-regressivo infinito, bastando para tanto substituir a seqüência de valores a_{t-i}, $i = 1, 2, 3, \ldots$ em (3.12). Nesse caso, os pesos $\pi_j = \theta_1^{j-1}(\phi_1 - \theta_1)$, para satisfazerem a condição de inversibilidade, devem ter $|\theta_1| < 1$.

Média

$$E(x_t) = \delta/(1-\phi_1)$$

Variância e coeficientes da autocovariância

$$\gamma_o = E(\tilde{x}_t.\tilde{x}_t) = E(\phi_1\tilde{x}_{t-1} + a_t - \theta_1 a_{t-1})^2 =$$

$$= \phi_1^2\gamma_o - 2\,\phi_1\theta_1 E(\tilde{x}_{t-1}a_{t-1}) + \sigma_a^2 + \theta_1^2\sigma_a^2 =$$

$$= \phi_1^2\gamma_o - 2\,\phi_1\theta_1\,\sigma_a^2 + \sigma_a^2 + \theta_1^2\,\sigma_a^2 =$$

$$= \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2}\;\sigma_a^2$$

$$\gamma_1 = E(\tilde{x}_t.\tilde{x}_{t-1}) = E\left[(\tilde{x}_{t-1}(\phi_1\tilde{x}_{t-1} + a_t - \theta_1 a_{t-1})\right] =$$

$$= \phi_1\,\gamma_o - \theta_1\,\sigma_a^2 =$$

$$= \frac{(1 - \phi_1\theta)(\phi_1 - \theta_1)}{1-\phi_1^2}\;\sigma_a^2$$

$$\gamma_2 = E(\tilde{x}_{t-2}.\tilde{x}_t) = E\left[\tilde{x}_{t-2}(\phi_1\tilde{x}_{t-1} + a_t - \theta_1 a_{t-1})\right]$$

$$= \phi_1\gamma_1$$

$$\vdots$$

$$\gamma_k = \phi_1\,\gamma_{k-1} \quad , \quad \text{qualquer } k \geq 2.$$

Coeficientes de autocorrelação

$$\rho_o = \frac{\gamma_o}{\gamma_o} = 1$$

$$\rho_1 = \frac{\gamma_1}{\gamma_o} = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}$$

$$\rho_2 = \frac{\gamma_2}{\gamma_o} = \frac{\phi_1\gamma_1}{\gamma_o} = \phi_1\rho_1$$

$$\vdots$$

$$\rho_k = \phi_1 \rho_{k-1}, \text{ qualquer } k \geq 2.$$

Note-se que, como o processo é um ARMA(1,1), é função de ϕ_1 e θ_1. A função de autocorrelação apresenta um decréscimo exponencial a partir de ρ_1. A presença do termo média móvel entra na determinação de ρ_1 somente. As restantes autocorrelações são determinadas pela parte auto-regressiva do modelo. Conseqüentemente, a função de autocorrelação combina características de MA(1) e AR(1).

3.5 – Modelo Auto-Regressivo Integrado – Média Móvel

Seja um processo estocástico não estacionário. Se esse apresenta características tais que sua não estacionariedade é do tipo homogêneo, isto é, que permite sua transformação em série estacionária pela aplicação do operador diferença, então o processo não estacionário homogêneo, descrito pela transformação da série em estacionária e, posteriormente, pela utilização de um processo misto auto-regressivo-média móvel, é chamado de processo Auto-Regressivo Integrado-Média Móvel de ordem (p, d, q) ou, simplesmente, definido por ARIMA (p, d, q). Simbolicamente,

$$\phi(B)\,\nabla^d \tilde{x}_t = \theta(B)a_t .$$

Note-se que, conforme foi visto nos itens anteriores deste capítulo, a exigência para que o processo ARMA (p,q) seja estacionário está somente na parcela do modelo descrito pelo processo auto-regressivo. Ou seja, as raízes da equação característica $\phi(B) = 0$ devem todas, necessariamente, cair fora do círculo unitário. Porém, se houver raízes que caiam dentro ou na fronteira do círculo unitário, o processo deixa de ter comportamento estacionário, passando a possuir um comportamento não estacionário explosivo ou não estacionário homogêneo.

Define-se uma série não estacionária homogênea de ordem "d" como aquela onde "d" raízes da equação característica $\phi(B) = 0$ caem sobre a fronteira do círculo unitário e as demais fora dele.

Conforme o exposto no item 3.1, o operador diferença é tal que sua aplicação "d" vezes é capaz de transformar o processo homogêneo evolutivo em estacionário. Portanto, considerando o modelo

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

onde $\phi(B)$ é um operador auto-regressivo não estacionário homogêneo, tal que "d" raízes da equação característica $\phi(B) = 0$ estão sobre o círculo unitário e as demais fora, então pode-se estabelecer que

$$\phi(B)\tilde{x}_t = \varphi(b)(1-B)^d\tilde{x}_t = \theta(B)a_t . \qquad (3.13)$$

Repare-se que, como $\nabla^d x_t = \nabla^d\tilde{x}_t$, d = 1, 2, ..., é indiferente escrever $\varphi(B)x_t$ ou $\phi(B)\tilde{x}_t$ no modelo (3.13). Portanto, o modelo ARIMA (p, d, q) vale

$$\varphi(B)x_t = \varphi(B)(1-B)^d x_t = \theta(B)a_t \qquad (3.14)$$

ou

$$\varphi(B)w_t = \theta(B)a_t, \text{ com } w_t = \nabla^d x_t ,$$

onde $\varphi(B)$ é um operador auto-regressivo estacionário.

Deve-se observar que $w_t = \nabla^d x_t \Rightarrow x_t = S^d w_t$, o que permite estabelecer a equação (3.13) por

$$\varphi(B)\, S^d w_t = \theta(B)a_t , \qquad (3.15)$$

onde $x_t = S^d w_t$ é uma série não estacionária homogênea de ordem "d".

Dessa maneira, o modelo definido pela equação (3.15) permite que, aplicando o operador diferença "d" vezes à série homogênea não estacionária, essa se torne capaz de ser representada pelo modelo estacionário auto-regressivo-média móvel.

O modelo ARIMA (p,d,q), assim definido para descrever uma série temporal, é

$$\varphi(B)x_t = \varphi(B)\nabla^d x_t = \theta(B)a_t ,$$

onde:

- $\varphi(B)$ é chamado operador auto-regressivo. Supõe-se que seja estacionário, isto é, que as raízes de $\varphi(B)=0$ caiam fora do círculo unitário;
- $\phi(B) = \nabla^d\varphi(B)$ é chamado operador auto-regressivo generalizado. É um operador não estacionário, com "d" raízes da equação característica $\phi(B) = 0$ iguais a um e as restantes fora do círculo unitário;
- $\theta(B)$ é chamado operador média móvel. Supõe-se que é inversível, ou seja, as raízes de $\theta(B)=0$ devem, todas, cair fora do círculo unitário.

Observe-se que, se d=0, o modelo ARIMA (p,0,q) representa um processo estacionário descrito por um ARMA (p,q).

Da mesma forma, é possível descrever a série estacionária w_t unicamente por um processo auto-regressivo ou unicamente por um processo média móvel. Assim, se w_t é somente um AR(p), então o modelo ARIMA (p,d,0) se torna somente um modelo Auto-Regressivo Integrado de ordem (p,d), denotado por ARI (p,d). Semelhantemente, se w_t é descrito somente por um MA(q), então o modelo ARIMA(0,d,q) se torna um modelo Integrado Média Móvel de ordem (d,q), definido sinteticamente por IMA(d,q).

Em geral, o número de vezes "d" de diferenças que se deve obter para que a série $\nabla^d x_t$ possa ser representada por um modelo estacionário inversível ARMA(p,q) é, no máximo, igual a dois.

Os casos onde d=1 e d=2 se caracterizam por descrever não estacionariedade quanto ao nível e/ou quanto à inclinação, os quais abrangem a grande maioria dos comportamentos das séries encontradas na vida real.

Necessita-se de d=1 quando a série não é estacionária, somente porque ocorrem trocas aleatórias em relação ao nível. Ou seja, quando o comportamento da série oscila ao redor de um nível durante um certo período de tempo e após passa para outro nível, sem que haja troca significativa de direção da série. Nesse caso, diz-se que a série é não estacionária homogênea de grau um ou que apresenta uma tendência estocástica em relação ao nível da série. Essa situação é visualizada graficamente na Figura 4 e, na prática, esse tipo de comportamento é característico da maioria das séries econômicas de tempo.

A não estacionariedade homogênea de ordem um pode ser percebida a partir das equações (3.13) e (3.14), que comprovam o fato de a série não ser influenciada pelo nível do processo. Nesse caso, para qualquer constante K, tem-se

$$\phi(B)(x_t + K) = \phi(B)x_t,$$

ou seja, $\phi(B)K = 0$. Isso implica que $\phi(1) = 0$, vale dizer que $\phi(B)$ tem um fator $(1 - B)$ e, portanto, ao se diferenciar uma vez, obtém-se uma série estacionária.

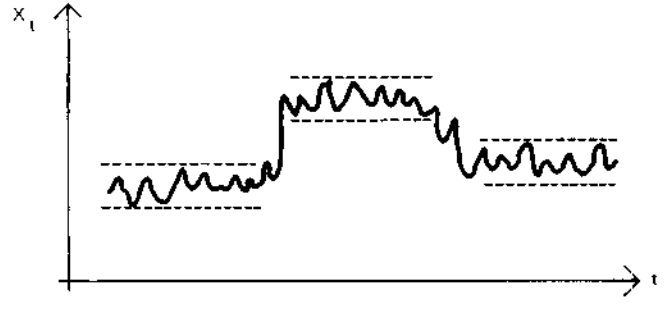


Figura 4 – Gráfico de uma série não estacionária quanto ao nível

Necessita-se de d=2, quando a série é não estacionária também quanto à inclinação, ou seja, quando o comportamento da série oscila em uma direção por um certo período e, após, muda para outra direção, conforme se visualiza na Figura 5.

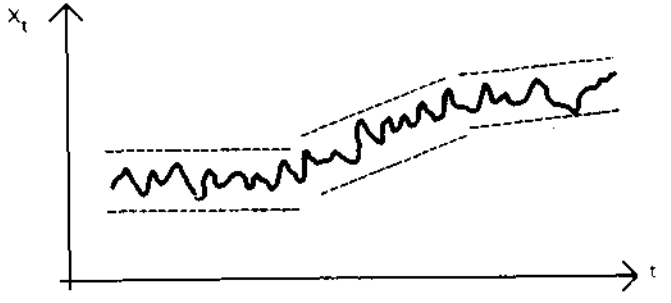


Figura 5 – Gráfico de uma série não estacionária quanto ao nível e à inclinação

A componente tendencial e o modelo ARIMA

O modelo ARIMA, até aqui exposto através das equações (3.13) e (3.14), é tal que a série não é influenciada pelo nível nem apresenta persistência em relação a alguma direção que possa caracterizar tendências determinísticas.

Em termos da equação (3.14),

$$\varphi(B)w_t = \theta(B)a_t, \text{ onde } w_t = \nabla^d x_t,$$

significa que a série estacionária w_t apresenta o termo constante $\delta = 0$ e, conseqüentemente, possui $E(w_t) = \delta/(1-\phi_1-\ldots-\phi_p) = 0$.

A inexistência do termo constante no modelo ARIMA significa que esse é capaz de representar somente aquelas tendências estocásticas que se caracterizam por oscilações com respeito ao nível e/ou à inclinação da série.

Se, no modelo ARIMA, o termo constante for diferente de zero, a média do processo $w_t = \nabla^d x_t$ será dada por $E(w_t) = \delta/(1 - \phi_1 - \ldots \phi_p) \neq 0$.

Isso implica que a série x_t possa possuir a componente tendencial embutida em seu modelo descritivo. Por exemplo, considerando d=1 e $\delta>0$, tem-se que $x_t = \Sigma w_t$ cresce linearmente com o tempo. Considerando d=2 e $\delta>0$, tem-se que $\nabla x_t = \Sigma w_t \Rightarrow x_t = \Sigma\Sigma w_t$, cuja tendência possui uma taxa de crescimento cumulativa constante no tempo. Se, ao contrário, $\delta < 0$, então a tendência será decrescente.

Portanto, o modelo descrito pela equação (3.14) pode ser generalizado pela inclusão de um termo constante, o qual permite a descrição de possível tendência determinística.

Então,

$$\phi(B)\ \nabla^{d} x_t = \theta_o + \theta(B) a_t . \qquad (3.16)$$

A componente sazonal e o modelo ARIMA

Em grande número de séries temporais, principalmente em Economia, é comum o aparecimento de algum comportamento cíclico de curto prazo (de até um ano), chamado de sazonalidade.

Em vista disso, para um tratamento completo sobre séries de tempo, torna-se necessário caracterizar e eliminar essa função cíclica do tempo para se obter a condição de estacionariedade.

Sazonalidade significa uma tendência de repetição de um determinado comportamento da variável que ocorre com certa regularidade no tempo. Séries sazonais são, então, aquelas séries que apresentam variações similares de um espaço de tempo a outro, caracterizando-se por mostrarem alta correlação serial entre observações da variável distanciadas pelo período da sazonalidade, além, é claro, da correlação serial existente entre observações próximas.

O ajuste sazonal necessário requer alguma forma de eliminação dessa componente cíclica de curto prazo, ou seja, de eliminar a correlação entre valores sazonais periodicamente defasados.

Box & Jenkins sugerem a aplicação de um modelo ARIMA sazonal para descrever a série possuidora de correlação serial nos períodos sazonalmente defasados do tipo

$$\Phi(B^{s}) \nabla_{s}^{D} x_t = \theta_o + \Theta(B^{s})\ \alpha_t , \qquad (3.17)$$

onde

$\Phi(B^{s}) = 1 - \Phi_1 B^{s} - \Phi_2 B^{2s} - \ldots - \Phi_p B^{Ps}$ é o operador auto-regressivo sazonal de ordem "p", estacionário;

$\Theta(B^{s}) = 1 - \Theta_1 B^{s} - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs}$ é o operador média móvel sazonal de ordem "Q", inversível;

$\nabla_{s}^{D} = (1 - B^{s})^{D}$ é o operador diferença sazonal, tal que $\nabla_s x_t = x_t - x_{t-s}$ e "D" indicam o número de diferenças sazonais.

Deve-se notar que (3.17) busca descrever somente o comportamento sazonal. Nessa equação, α_t não é um processo de ruído branco, pois ainda permanecem as correlações entre observações próximas.

Para descrever essas correlações para os a_t, cuja sazonalidade já está descrita por (3.17), usa-se o modelo comum ARIMA

$$\phi(B)\ \nabla^d \alpha_t = \theta(B) a_t \ . \qquad (3.18)$$

Substituindo (3.18) em (3.17), tem-se

$$\phi(B)\ \Phi(B^S) \nabla^{\ d} \nabla_s^D\ x_t = \theta_0 + \theta(B)\ \Theta(B^S) a_t \ , \qquad (3.19)$$

o qual Box & Jenkins denominaram modelo ARIMA sazonal multiplicativo de ordem $(p,d,q)x(P,D,Q)_s$, simbolicamente chamado de SARIMA $(p,d,q)x(P,D,Q)_s$.

Note-se que, para obter uma série estacionária, w_t, deve-se tomar "d" diferenças simples e "D" diferenças sazonais da variável x_t, ou seja,

$$w_t = \nabla^{\ d} \nabla_s^D\ x_t \ .$$

4 – O PROCESSO DE MODELAGEM

Os modelos completos ARIMA (p,d,q) e sua variante sazonal SARIMA (p,d,q) x (P,D,Q), definidos no capítulo anterior, são tidos como alguns dos meios mais poderosos e úteis para representar modelos de séries econômicas de tempo e, conseqüentemente, para inferir previsões sobre os movimentos futuros da variável.

O processo de construção de modelos proposto por Box & Jenkins está assentado em um ciclo iterativo, composto de três estágios: identificação, estimação e checagem do diagnóstico.

A identificação consiste no estabelecimento da ordem (p,d,q) – ou, no caso sazonal, (p,d,q)x(P,D,Q) – e deve ser determinada inicialmente. Essa fase é a mais crucial, requerendo certo grau de conhecimento teórico e experiência sobre o assunto, uma vez que a definição do tipo de modelo que se ajusta aos dados do mundo real é precária, de difícil caracterização, feita através de medidas amostrais, o que implica uma identificação geralmente inexata. É comum, inclusive, que a avaliação feita nessa etapa sugira a viabilidade de a série ser ajustada por mais de um modelo.

O grau de acurácia, entretanto, poderá ser aumentado no decorrer das etapas do ciclo iterativo de modelagem.

Concluídas as fases de identificação da ordem e de estimação dos parâmetros do modelo, que fornecem as estimativas dos parâmetros auto-regressivos e médias móveis, deve-se realizar um processo de checagem do diagnóstico, isto é, de verificação da acurácia do modelo estimado na etapa anterior, com o fito de avaliar quão bem o modelo matemático está ajustado aos dados da realidade. Caso se verifique que um modelo não é adequado, o ciclo será repetido.

Para facilitar processos computacionais, é bastante comum trabalhar-se com mais de um modelo durante o ciclo iterativo de modelagem. No final do ciclo, a etapa de checagem do diagnóstico permitirá avaliar a viabilidade do ajustamento para cada um dos modelos sugeridos na fase de identificação.

De outra parte, se mais de um modelo se apresenta viável, têm-se meios para escolher aquele que melhor atende às finalidades do estudo. Assim, por exemplo, para fins de previsão, a escolha deve recair naquele que apresenta menor erro de previsão, vale dizer, menor erro quadrático médio.

O processo de construção de modelos pode ser estabelecido como segue.

4.1 – Modelagem de Séries Não Sazonais: Modelo ARIMA (p, d, q)

4.1.1 – A Identificação

A identificação compreende um conjunto de procedimentos com o objetivo de proporcionar uma idéia aproximada da estrutura do modelo. Assim, a identificação é a etapa onde se estabelecem as características básicas

da série, tais como o tipo de processo gerador (estacionário ou não estacionário); a existência ou não das componentes sazonal e tendencial, e que capacita o pesquisador a estabelecer a ordem do modelo que se ajusta à série observada; e as estimativas iniciais para os valores dos parâmetros envolvidos.

Como primeiro passo na identificação do modelo, é conveniente plotar os dados para se obter, "a priori", uma idéia do comportamento das observações. Já neste momento, pode-se detectar informações úteis que poderão abreviar alguns procedimentos posteriores. A grosso modo, pode-se perceber se a série é ou não estacionária, é ou não sazonal, e alguma idéia sobre a necessidade de transformações.

A possibilidade de transformação é salientada por Box & Jenkins pelo fato de a modelagem ser sempre melhor se os dados são homoscedásticos. Assim, se a série apresenta heteroscedasticidade, é conveniente fazer-se uma transformação a fim de tornar sua variância constante.

Para facilitar a tomada de decisão sobre qual a transformação que melhor se adapta, é sugerida a utilização de um gráfico cartesiano do tipo amplitude x média, o qual apresenta em um eixo a média de segmentos de observações da série original (no caso de dados sazonais, o comprimento dos segmentos deve ser de um ciclo completo; caso os dados não sejam sazonais, 8 a 12 observações satisfazem) e no outro a amplitude de cada um desses segmentos. A idéia básica desse gráfico é colocar em um eixo uma medida de posição e, no outro, uma da variação. Se a amplitude for independente da média, não há necessidade de transformação. Se for diretamente proporcional[1], a transformação sugerida é a logarítmica natural.

Representando a transformação por $Z_t(\lambda)$, onde λ é um vetor paramétrico a ser estimado, Box & Cox sugerem as seguintes alternativas, estando algumas ilustradas na Figura 1.

$$Z_t^{(\lambda)} = \begin{cases} (Z_t^{(\lambda)} - C)/\lambda & , \text{para } \lambda \neq 0 \\ \log_e Z_t & , \text{para } \lambda = 0 \end{cases}$$

Deve-se notar, entretanto, que o problema da conveniência de se fazer transformações é ainda um campo bastante controverso. Embora Box & Jenkins sugiram sua feitura, outros autores questionam a sua utilidade e a qualidade das previsões assim obtidas, as quais podem estar ocasionando tendenciosidade em seus estimadores[2].

A identificação do modelo, entretanto, envolve o estudo da função de autocorrelação. Essa função irá, conforme visto no Capítulo 2, diagnosticar as principais características da série, permitindo ao pesquisador inferir informações sobre a estrutura do modelo gerador.

---

[1] Isso é o caso mais comum em séries econômicas, onde a componente tendencial induz o crescimento da variância à medida que o tempo passa.

[2] Ver a respeito, entre outros, NELSON (1973), GRANGER & NEWBOLD (1977).

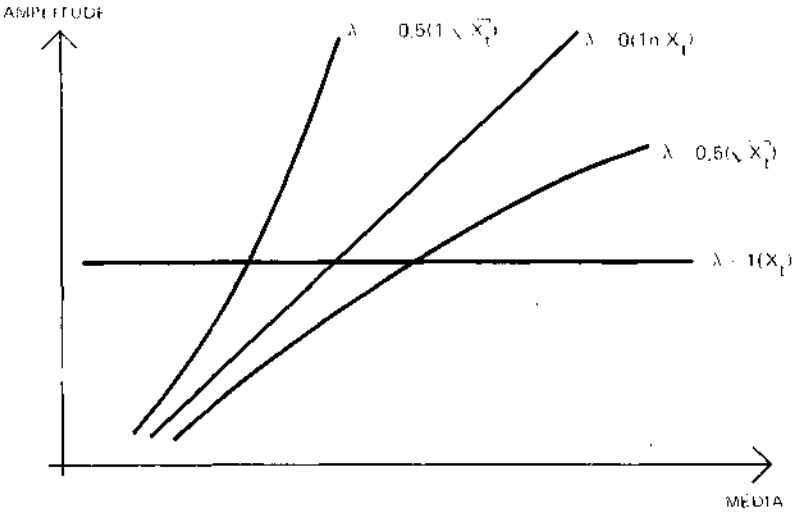


Figura 1 – Gráfico "Amplitude x Média" para alguns valores possíveis de λ

A função de autocorrelação, inicialmente, informará se o processo gerador é ou não estacionário. Para tanto, deve-se lembrar a característica dessa função em uma série estacionária, a qual, conforme exposto no Capítulo 2, deve aproximar-se rapidamente de zero quando "k" cresce. Se os coeficientes de autocorrelação possuírem valores grandes e tiverem uma diminuição lenta, enquanto "k" cresce, indicam que a série é não estacionária.

Esse fato pode ser comprovado a partir da equação (3.11), onde a função de autocorrelação de um modelo ARMA estacionário satisfaz a equação $\phi(B)\rho_k = 0$, $k>q$. Seja $\phi(B) = (1-G_1B)(1-G_2B) \ldots (1-G_pB)$. Então a solução para os coeficientes de autocorrelação, supondo raízes distintas, é dada por $\rho_k = A_1G_1^k + A_2G_2^k + \ldots + A_pG_p^k$, $k>q-p$. Se o modelo é estacionário, $G_i<0$, $i = 1, 2, \ldots, p$. Se nenhuma raiz está próxima do círculo unitário, as autocorrelações tenderão rapidamente a zero para valores moderados e grandes de "k". Por outro lado, se existir G_j real, tal que $G_j \simeq 1$, ou seja, $G_j = 1-\lambda$, $\lambda>0$, pequeno, então, como $G_j^k = (1-\lambda)^k \simeq 1-k\lambda$, tem-se que $\rho_k = A_j(1-k\lambda)$, fato que indica um declínio lento e linear para zero da função de autocorrelação[3]. Nesse caso, se a queda demorada ocorre de forma regular e contínua, conforme Figura 2, a série é não sazonal; se a queda demorada ocorre de forma descontínua, apresentando picos regulares pelo período "s" de sazonalidade (Figura 3), ou em um padrão cíclico, com período igual a "s" (Figura 4), a série é sazonal de ordem "s".

[3] A grosso modo, pode-se dizer que a distinção do decréscimo da função de autocorrelação entre processos estacionários e não estacionários se caracteriza pelo fato de o primeiro apresentar um decréscimo exponencial, enquanto que o segundo tem sua queda de forma linear.

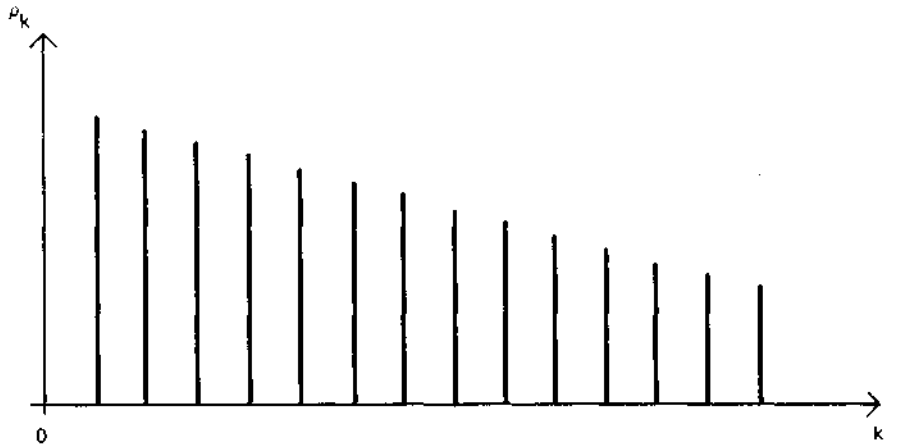


Figura 2 – Correlograma de um processo não estacionário não sazonal

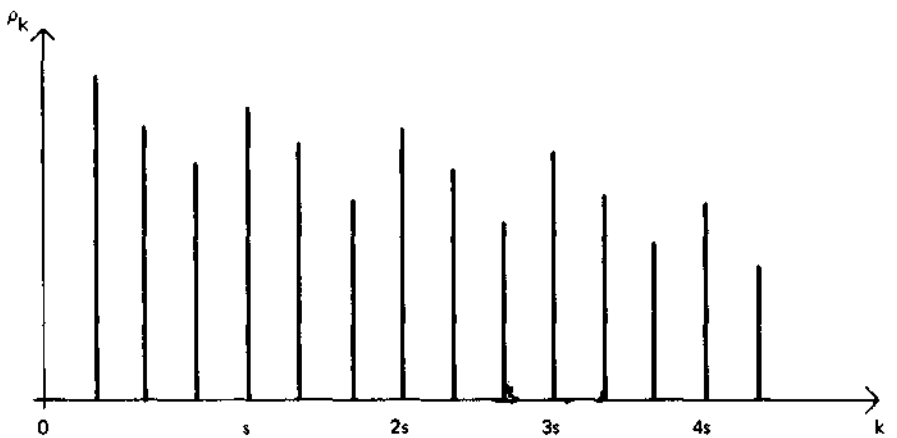


Figura 3 – Correlograma de um processo não estacionário sazonal

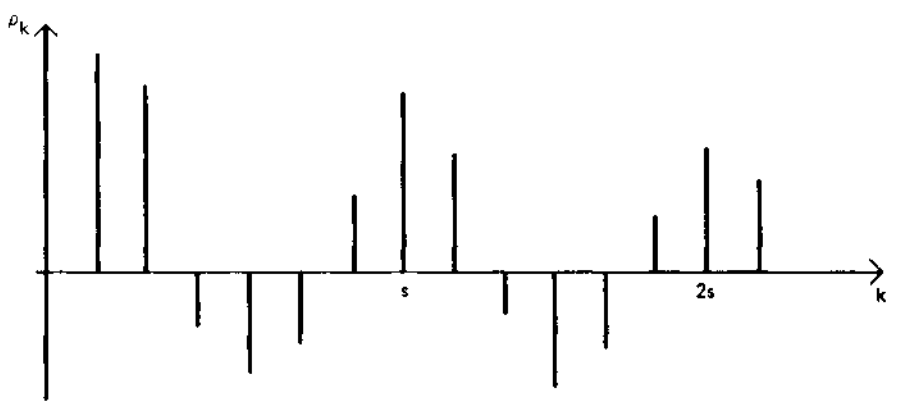


Figura 4 – Correlograma de um processo não estacionário sazonal

O processo de transformação da série não estacionária em estacionária é feito através do operador diferença, $\nabla^d=(1 - B)^d$. O número "d" de diferenças necessárias para se obter a estacionariedade da série é alcançado quando a função de autocorrelação da série $W_t = \nabla^d X_t$ decresce rapidamente para zero.

Na maioria das séries, o número de diferenças "d" é pequeno, com $d \leq 2$. Para tais casos, a análise das autocorrelações é suficiente, em geral, com uma inspeção nos 15 ou 20 coeficientes iniciais de autocorrelação (isto é, ρ_k, k=1, 2, 3, ..., 20).

Deve-se notar, entretanto, que essa forma de obter "d" se baseia em uma avaliação subjetiva do pesquisador. O julgamento do fato de a função de autocorrelação decrescer rapidamente para zero poderá apresentar situações de dúvida.

Como foi visto, o número de diferenças necessárias para tornar a série estacionária é igual ao número de raízes da equação característica $\phi(B) = 0$ que caiam sobre a fronteira do círculo unitário.

A dificuldade de julgamento é bastante comum naquelas séries onde o modelo é definido por parâmetros que, embora satisfazendo a condição de estacionariedade, se apresentam com valores próximos à fronteira do círculo unitário. A queda, nesse caso, da função de autocorrelação para sucessivas diferenças da série pode não apresentar alterações significativas.

Para facilitar a escolha de "d", ANDERSON (1979) sugere a análise de mais um instrumento, que é o estudo do comportamento da variância das séries sucessivamente diferenciadas. A análise baseia-se no fato de a variância de uma série não estacionária ser maior que a variância da série transformada em estacionária, através do operador diferença, e, quando essa é conseguida, uma sobrediferenciação (isto é, uma operação diferença a mais do que o necessário para a obtenção da estacionariedade) apresenta um acréscimo substancial da variabilidade da série.

Após estabelecer "d", tal que $W_t = \nabla^d X_t$ é estacionária, trabalha-se com a série W_t, que pode ser representada por um processo ARMA, para estabelecer a ordem (p,q). Como a maioria das séries econômicas se caracterizam por serem passíveis de descrição por um modelo de baixa ordem, onde tanto "p" quanto "q" assumem valores inferiores ou iguais a dois, o estabelecimento da ordem (p,q) torna-se bastante simples, tendo em vista o conhecimento que se tem das características da função de autocorrelação das séries descritas por modelos AR(p), MA(q) e ARMA(p,q)[4]: um MA(q) tem memória de "q" períodos e um AR(p), embora possuindo memória infinita, tem função de autocorrelação que declina geometricamente para zero[5]. Além disso, para uma definição precisa da ordem "p" de um processo AR(p), existe o instrumento da função de autocorrelação parcial do processo, a qual, para um modelo auto-regressivo, se apresenta com memória finita de ordem "p".

---

[4] Deve-se salientar que, se a série não se caracteriza por ser representada por um processo de baixa ordem, a especificação de "p" e "q" torna-se bastante difícil, pois o meio disponível é o de tentar adivinhar seus valores, fato que, obviamente, só terá algum valor prático se houver experiência suficiente (e, às vezes, sorte!) no processo de modelagem de séries temporais.

[5] Isso é o caso geral, onde as raízes de $\phi(B)$ são distintas e reais. Se houver raízes complexas, o declínio das autocorrelações terá a forma senoidal; para maiores discussões a respeito, inclusive demonstração, ver BOX & JENKINS (1976).

A função de autocorrelação parcial, denotada por ϕ_{kk}, é um instrumento analítico que, partindo do fato de um processo AR(p) ter uma função de autocorrelação com memória infinita, permite sua obtenção através de "p" funções de autocorrelação diferentes de zero[6].

Define-se a função de autocorrelação parcial por:

$$\phi_{kk} = \begin{cases} \rho_k & \text{se } k = 1 \\ \dfrac{|\mathbb{P}^*_k|}{|\mathbb{P}_k|} & \text{se } k \geq 2 \end{cases}$$

onde $\mathbb{P}_k$ é a matriz de autocorrelação e $\mathbb{P}^*_k$ é a matriz que difere de $\mathbb{P}_k$ somente pela última coluna, a qual é substituída pelo vetor coluna $(\rho_1, \rho_2, \rho_3, \ldots, \rho_k)$.

Para um processo auto-regressivo de ordem "p", a função de autocorrelação parcial ϕ_{kk} será diferente de zero para $k \leq p$ e zero para $k > p$. Dessa maneira, embora função de autocorrelação de um processo AR(p) tenha memória infinita, a função de autocorrelação parcial, assim definida, possui memória de "p" períodos, permitindo, conseqüentemente, a identificação da ordem do modelo auto-regressivo.

Em geral, o comportamento da função de autocorrelação de um processo auto-regressivo tende a imitar o comportamento da função de autocorrelação parcial de um processo média móvel e, inversamente, a função de

---

[6] As funções de autocorrelação parciais podem ser assim obtidas: seja ϕ_{kj} o j-ésimo coeficiente de um AR(K), tal que ϕ_{kk} seja o último coeficiente. Como se viu no capítulo anterior, ϕ_{kj} deve satisfazer a

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

ou, utilizando as equações de Yule-Walker,

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad \text{ou} \quad \mathbb{P}_k \cdot \Phi_k = \rho_k$$

Resolvendo para os diversos valores de k, tem-se:

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

e assim por diante.

autocorrelação de um processo média móvel tende a imitar o comportamento da função de autocorrelação parcial de um processo auto-regressivo.

Assim,a função de autocorrelação parcial de um processo auto-regressivo de ordem "p" tem memória de "p" períodos, enquanto sua função de autocorrelação tem memória infinita. De forma inversa, para um processo média móvel de ordem "q", a função de autocorrelação parcial tem memória infinita, enquanto sua função de autocorrelação tem memória de "q" períodos. Se ambas as funções, autocorrelação e autocorrelação parcial, apresentam memória infinita, possivelmente um processo misto seja o mais conveniente para descrever a série temporal.

No Quadro 1, estão sintetizadas as principais propriedades das funções de autocorrelação e autocorrelação parcial, para a identificação dos processos auto-regressivo, média móvel e misto (auto-regressivo-média móvel), para as ordens comumente encontradas na prática. Nas páginas subseqüentes, apresentam-se os gráficos dos padrões comportamentais dessas funções.

Quadro 1

Comportamento das funções de autocorrelação e autocorrelação parcial para o processo ARIMA (p,d,q) de ordens mais comumente encontradas na prática

DISCRIMINAÇÃO	ORDEM				
	(1,d,0)	(0,d,1)	(2,d,0)	(0,d,2)	(1,d,1)
Comportamento de ρ_k	decai exponencialmente	somente ρ_1 é diferente de zero	mistura de exponencial e senoidal amortecida	somente ρ_1 e ρ_2 são diferentes de zero	decai exponencialmente a partir da primeira defasagem
Comportamento de ϕ_{kk}	somente ϕ_{11} é diferente de zero	dominado por uma queda exponencial	somente ϕ_{11} e ϕ_{22} são diferentes de zero	dominado por uma mistura de exponencial e senoidal amortecida	dominado pela queda exponencial a partir da primeira defasagem
Estimativas preliminares	$\phi_1 = \rho_1$	$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$	$\phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}$	$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$	$\rho_1 = \frac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1+\theta_1-2\theta_1\phi_1}$
			$\phi_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}$	$\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$	$\rho_2 = \rho_1\phi_1$
Regiões admissíveis	$-1 < \phi_1 < 1$	$-1 < \rho_1 < 1$	$-1 < \phi_2 < 1$	$-1 < \theta_2 < 1$	$-1 < \phi_1 < 1$
			$\phi_2 + \phi_1 < 1$	$\theta_2 + \theta_1 < 1$	
			$\phi_2 - \phi_1 < 1$	$\theta_2 - \theta_1 < 1$	$-1 < \theta_1 < 1$

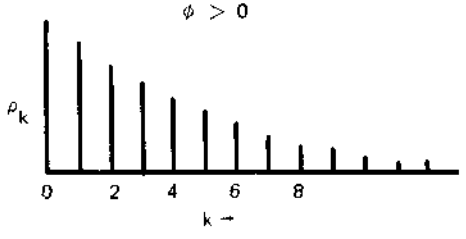
Padrões de ρ_k e ϕ_{kk} para um AR(1)

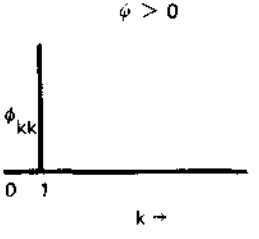
Autocorrelações: $\rho_k = \phi^k$

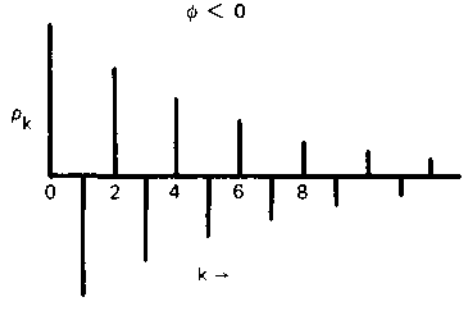
Autocorrelações parciais:

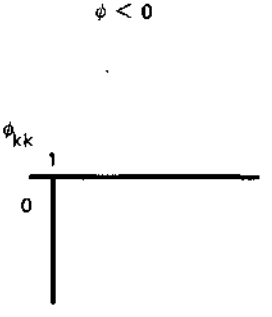
$\phi_{11} = \rho_1 = \phi$

$$\phi_{kk} = 0, \quad k \geq 2, \text{ pois } \phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\phi_1^2 - \phi^2}{1 - \phi^2} = 0, \quad (\phi \neq 1)$$







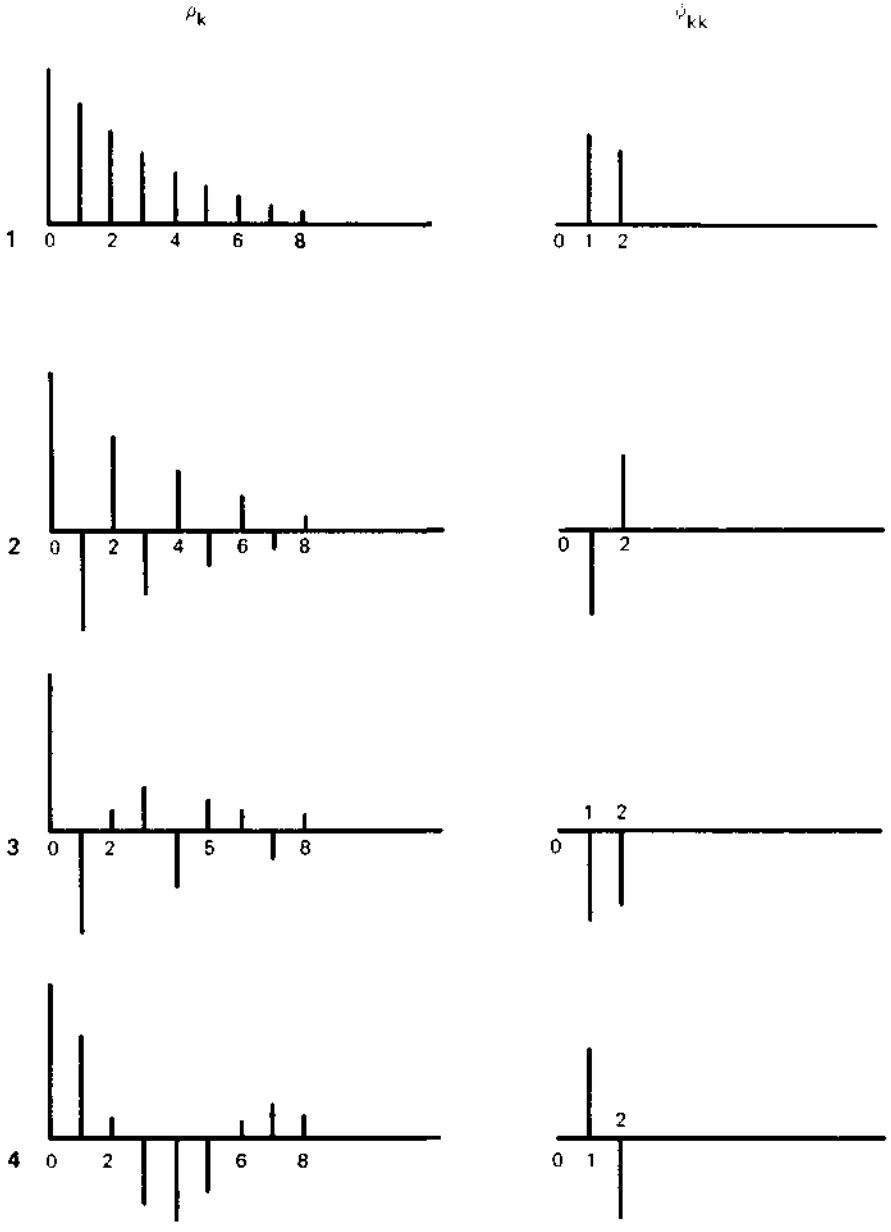


Padrões de ρ_k e ϕ_{kk} para um AR(2)

Autocorrelações: $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$, $k>0$

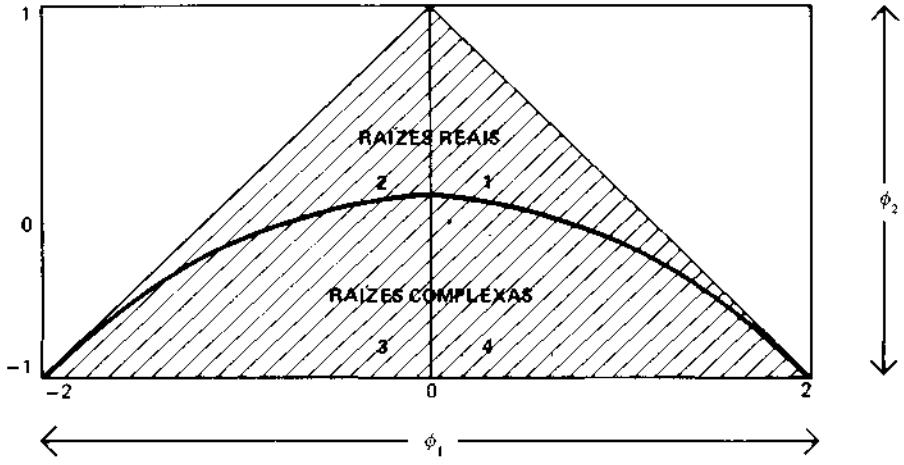
Autocorrelações parciais:

$\phi_{11} = \rho_1 = \dfrac{\phi_1}{1-\phi_2}$; $\phi_{22} = \dfrac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \phi_2$, $\phi_{kk} = 0$, $k \geq 3$

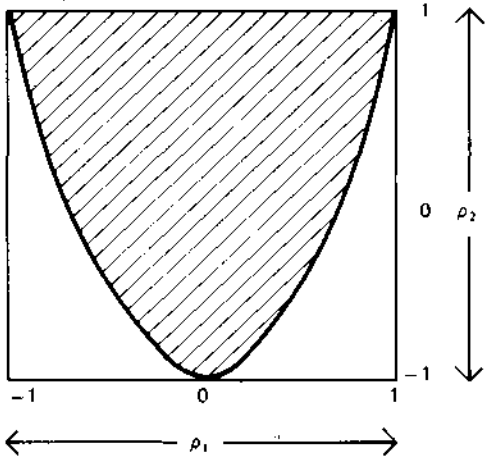


Cada um desses quatro tipos de comportamento dependerá das condições de estacionariedade requeridas por esse modelo.

Assim, as regiões admissíveis são as apresentadas a seguir.



$-1 < \rho_1 < 1$
$-1 < \rho_2 < 1$
$\rho_1^2 < 1/2(\rho_2 + 1)$

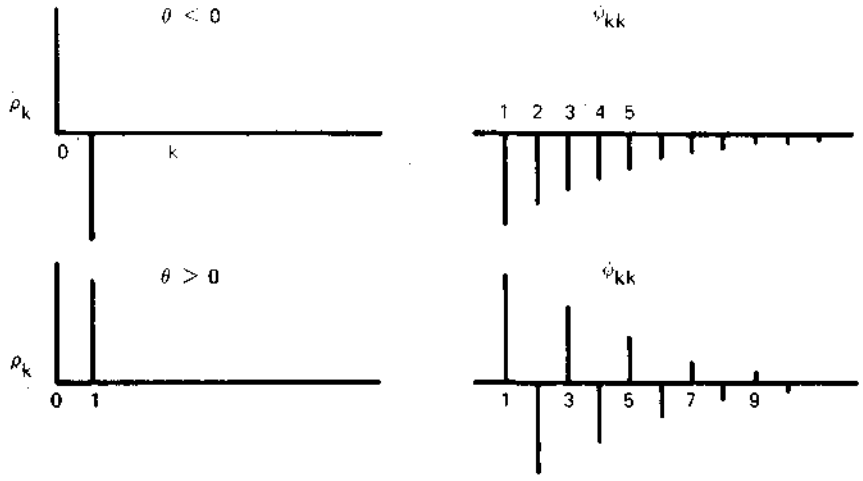


Padrões de ρ_k e ϕ_{kk} para um MA(1)

Autocorrelações: $\rho_1 = \dfrac{\theta 1}{1 + \theta_1^2}$; $\rho_k = 0$, $k \geq 2$

Autocorrelações Parciais:

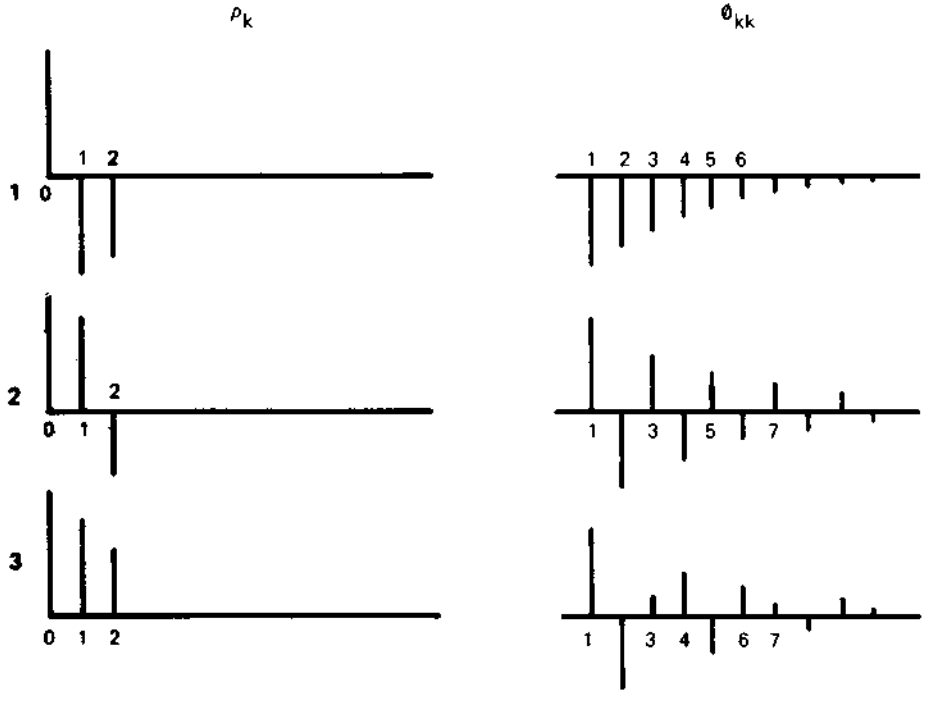
$$\phi_{kk} = \frac{-\theta_1^k(1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}$$

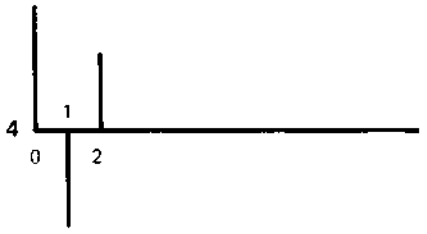


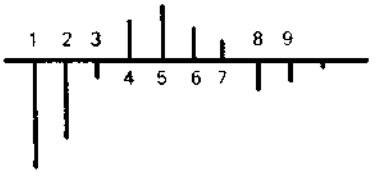
Padrões de ρ_k e ϕ_{kk} para um MA(2):

Autocorrelações: $\rho_1 = \dfrac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2}$; $\rho_2 = \dfrac{\theta 2}{1+\theta_1^2+\theta_2^2}$; $\rho_k = 0,\ k \geq 3$

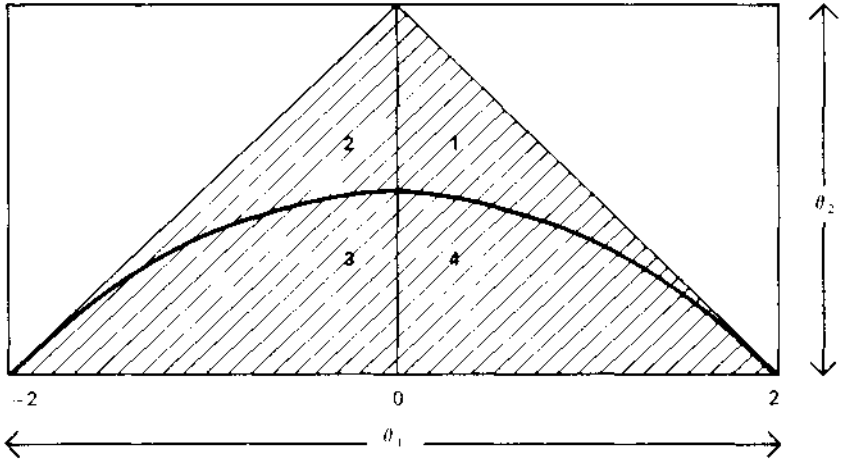
Autocorrelações Parciais: semelhantes a AR(2)



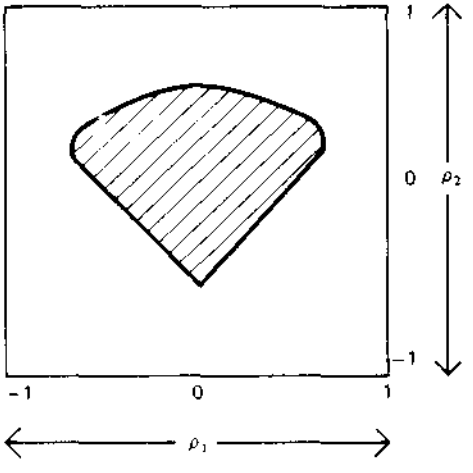




Condições de admissibilidade para o MA(2)



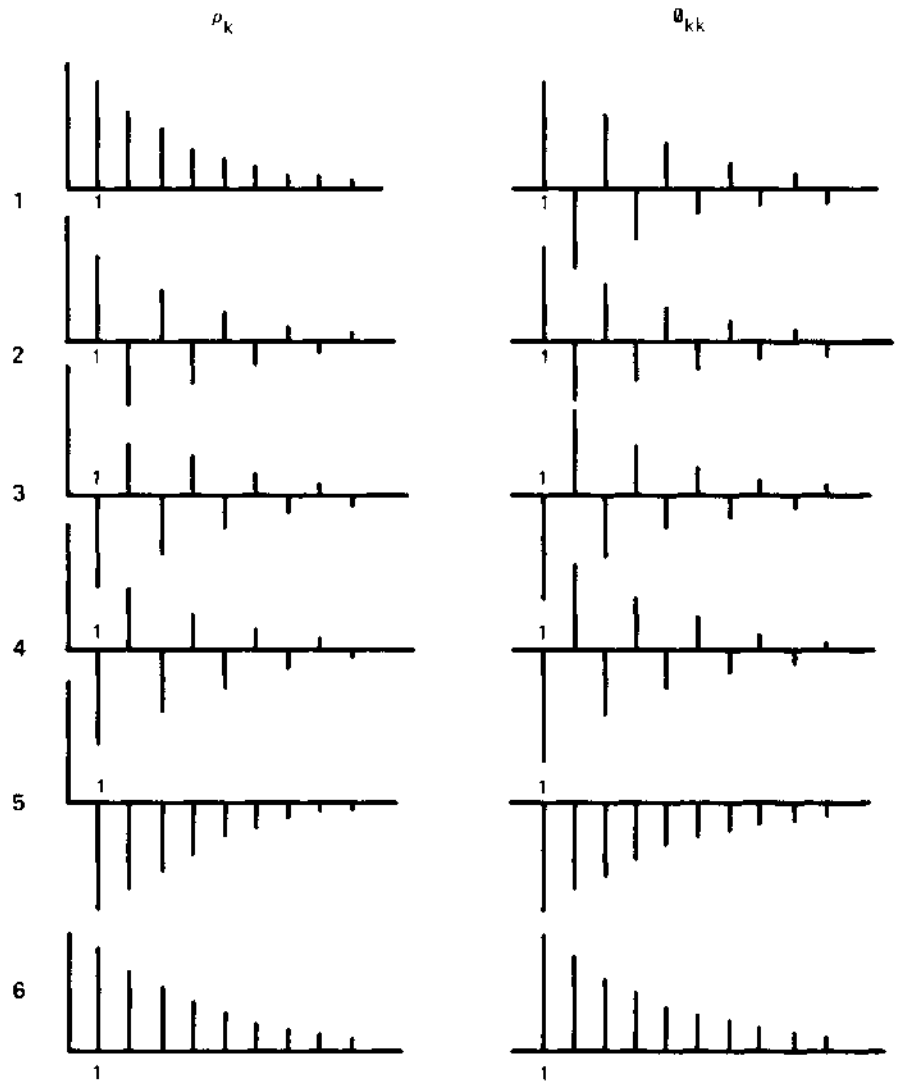
$\rho_2 + \rho_1 = -0{,}5$
$\rho_2 - \rho_1 = -0{,}5$
$\rho_1^2 = 4\rho_2(1 - 2\rho_2)$



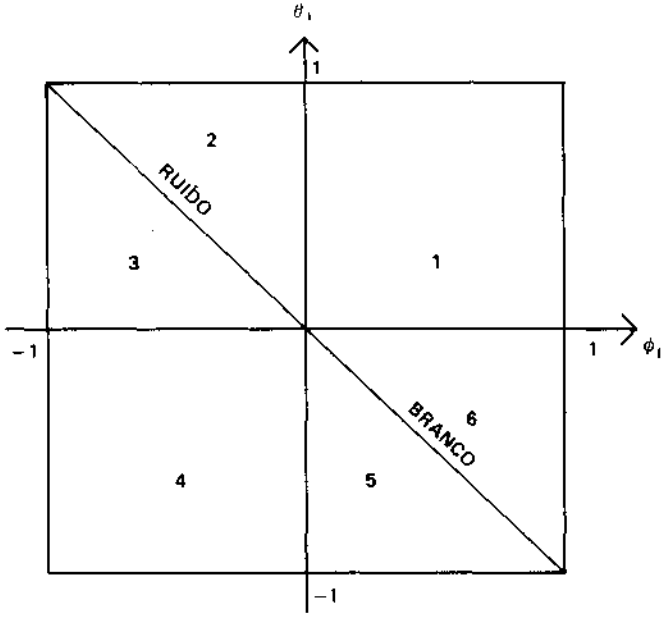
Padrões de ρ_k e ϕ_{kk} para um ARMA(1,1)

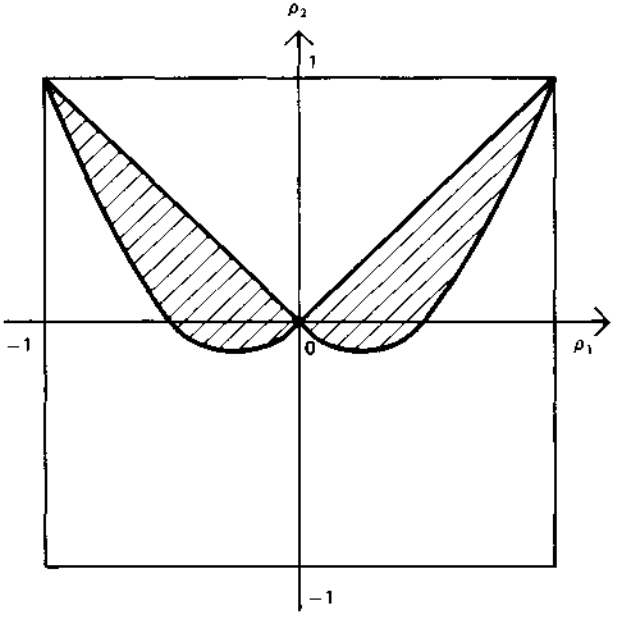
Autocorrelações: $\rho_1 = \dfrac{(1+\theta)(1+\phi_1\theta_1)}{1+2\phi_1\theta_1+\theta_1^2}$; $\rho_2 = \phi\rho_1$; $\rho_k = \phi^{k-1}\rho_1$, $k \geq 2$

Autocorrelações Parciais: ϕ cai exponencialmente (alternando ou não de sinal) a partir de $\phi_{11} = \rho_1$.



Condições de admissibilidade para um ARMA(1,1):





$|\rho_2| < |\rho_1|$

$\rho_2 > \rho_1(2\rho_1 + 1),\ \rho_1 < 0$

$\rho_2 > \rho_1(2\rho_1 - 1),\ \rho_1 > 0$

Na prática, porém, surge outro problema, à medida que não se dispõe das funções de autocorrelação e autocorrelação parcial teóricas, necessitando-se trabalhar com suas funções estimadas. Isso porque é importante ter alguma indicação de quanto um valor estimado pode diferir do valor teórico correspondente. Ou seja, torna-se necessário usar algum instrumento para detectar se, de fato, os coeficientes de autocorrelação e autocorrelação parcial são efetivamente zero após alguma determinada defasagem "k".

Box & Jenkins sugerem, como instrumento, uma simplificação da fórmula de Bartlett para grandes amostras[7], através da qual podem-se calcular erros padrão das autocorrelações e autocorrelações parciais estimadas, denominadas **erros padrão de grande defasagem**, onde as estimativas amostrais substituem as autocorrelações teóricas. Essa fórmula é dada por:

$$\hat{\sigma}\ (r_k) \cong \frac{1}{\sqrt{n}}\ \sqrt{1 + 2(r_1^2 + r_2^2 + \ldots + r_q^2)},\ k > q$$

Para as autocorrelações parciais, o erro padrão de grande defasagem é deduzido com a hipótese de que o processo é auto-regressivo de ordem "p". Sua fórmula é dada por:

$$\hat{\sigma}\ (\hat{\phi}_{kk}) \cong \frac{1}{\sqrt{n}}\ ,\ k > p$$

A distribuição dos coeficientes de autocorrelação e autocorrelação parcial estimados, onde o valor teórico respectivo é zero, é aproximadamente uma Normal. Então, o valor estimado r_k, dividido pelo seu erro padrão, será uma variável normal padronizada. Resultado similar obtém-se para as autocorrelações parciais.

Esses aspectos proporcionam meios bastante bons para verificar se as funções de autocorrelação e de autocorrelação parcial teóricas são iguais a zero, após alguma determinada defasagem. Como se sabe, em uma distribuição Normal, distâncias que excedem um erro padrão, em módulo, têm probabilidade de ocorrência em torno de 1/3, e as mesmas para dois erros padrões têm probabilidade de 1/20 (5%).

Assim, para avaliações com as funções de autocorrelação e autocorrelação parcial estimadas, é conveniente plotar linhas de controle em $\pm\ \hat{\sigma}$ ou $\pm\ 2\hat{\sigma}$.[8] Se os valores dessas funções caírem na região limitada por

---

[7] A expressão para a variância do coeficiente de autocorrelação estimado de um processo estacionário normal, dado por Bartlett, pode ser assim definida:

$$VAR(r_k) \cong \frac{1}{n} \sum (\rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2\rho_v^2\rho_k^2)$$

Como as autocorrelações $\rho_v = 0$ para $v > q$, todos os termos da equação, exceto o primeiro, são zero quando $k > q$. Então, a variância da autocorrelação estimada de defasagem k pode ser obtida usando a aproximação de Bartlett:

$$VAR(r_k) \cong \frac{1}{n}\ \{1 + 2 \sum_{v=1}^{q} \rho_v^2\},\ k > q$$

[8] Em geral, para maior confiança, é usada uma região limitada por dois desvios padrão estimados.

essas linhas de controle, então proporcionam ao pesquisador uma boa indicação de que as hipóteses formuladas sobre o tipo de processo empregado são aproximadamente corretas, isto é, $\hat{\phi}_{kk}$ e r_k são significativamente diferentes de zero se $|r_k| > 2DP(r_k)$, $j>q$ e $|\hat{\phi}_{kk}| > 2DP(\hat{\phi}_{kk})$, $j>p$.

Após a identificação de "d" e da ordem (p,q) da série estacionária W_t, pode-se, então, verificar a existência do parâmetro θ_o. Para tanto, testa-se a hipótese de $E(W_t)$ ser igual a zero, através da comparação da média amostral $\bar{W}$ com seu desvio padrão estimado $DP(\bar{W})$.

Os valores aproximados desses desvios padrão para os modelos usuais, estão sintetizados a seguir, onde n=N-d, sendo N o número de valores da série[9].

$$AR(1) \rightarrow DP(\bar{W}) = \sqrt{\frac{\gamma_o(1 + r_1)}{n(1 - r_1)}}$$

$$AR(2) \rightarrow DP(\bar{W}) = \sqrt{\frac{\gamma_o(1+r_1)(1-2r_1^2+r_2)}{n(1-r_1)(1-r_2)}}$$

$$MA(1) \rightarrow DP(\bar{W}) = \sqrt{\frac{\gamma_o(1+2r_1+2r_2)}{n}}$$

$$MA(2) \rightarrow DP(\bar{W}) = \sqrt{\frac{\gamma_o(1+2r_1+2r_2)}{n}}$$

$$ARMA(1,1) \rightarrow DP(\bar{W}) = \sqrt{\frac{\gamma_o}{n}\left(1+\frac{2r_1^2}{r_1-r_2}\right)}$$

Se o desvio padrão $DP(\bar{W}) > |\bar{W}|$, a média pode ser considerada igual a zero. Conseqüentemente, $\theta_o = 0$.

Se o desvio padrão $DP(\bar{W}) < |\bar{W}|$, a média pode ser considerada diferente de zero. Nesse caso, como $E(w_t) = \bar{W} = \frac{\theta_o}{1-\phi_1-\ldots-\phi_p}$, o valor de θ_o é obtido após a obtenção dos parâmetros auto-regressivos $\phi_1, \phi_2, \ldots, \phi_p$.

---

[9] Para a dedução do $DP(\bar{W})$ de cada modelo, ver BOX & JENKINS (1976: 193-95).

Por fim, o estudo das funções de autocorrelação e autocorrelação parcial proporciona as estimativas iniciais para os parâmetros, as quais serão os instrumentos básicos para, no ítem subseqüente deste capítulo, produzir as estimativas convenientes dos parâmetros do modelo.

No Capítulo 3, as análises dos processos média móvel, auto-regressivo e misto auto-regressivo-média móvel, para ordem inferior ou igual a dois, proporcionaram as expressões dos coeficientes de autocorrelação em termos dos parâmetros do processo. As estimativas iniciais podem ser obtidas substituindo valores estimados.

Deve-se salientar que, no processo de obtenção dos valores paramétricos, surgirão, muito freqüentemente, múltiplas soluções para estimativas dos parâmetros, porém somente uma dessas soluções irá satisfazer às condições restritivas de inversibilidade e/ou estacionariedade.

Na página 75, o Quadro 1 expõe os valores calculados para os parâmetros ϕ e θ em função dos coeficientes de autocorrelação, para as ordens comumente encontradas nas séries econômicas de tempo.

Da mesma forma, as estimativas iniciais para σ_a^2 são obtidas pelos valores da autocovariância, conforme análise do capítulo anterior. Sinteticamente, suas fórmulas são:

Para um AR(p): $\hat{\sigma}_a^2 = C_o(1-\hat{\phi}_1 r_1 - \hat{\phi}_2 r_2 - \ldots - \hat{\phi}_p r_p)$;

para um MA(q): $\hat{\sigma}_a^2 = \dfrac{C_o}{1+\hat{\theta}_1^2+\hat{\theta}_2^2+\ldots+\hat{\theta}_p^2}$;

para um ARMA(1,1): $\hat{\sigma}_a^2 = \dfrac{1-\hat{\phi}_1^2}{1+\hat{\theta}_1^2-2\hat{\phi}_1\hat{\theta}_1} c_o$,

onde c_o é uma estimativa de γ_o.

4.1.2 – A Estimação dos Parâmetros

Após a etapa de identificação da ordem do modelo que melhor representa a série observada, torna-se necessária a obtenção de estimativas para os parâmetros do modelo identificado.

O processo de estimação do modelo geral ARIMA (p,d,q),

$$\phi(B)\,\nabla^d x_t = \phi(B)\,W_t = \theta(B)\,a_t,$$

onde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \ldots - \phi_p B^p$

e $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \ldots - \theta_q B^q$,

significa, portanto, a obtenção de "p" estimativas para os parâmetros ϕ_1, ϕ_2, ϕ_3, ... ϕ_p, e "q" estimativas para os parâmetros θ_1,θ_2,θ_3, ..., θ_q, além da variância do ruído, σ_a^2.[10]

Embora os parâmetros auto-regressivos sejam lineares, permitindo o uso de métodos simples de estimação como, por exemplo, o de mínimos quadrados ordinários, os parâmetros média móvel não são lineares, o que torna bastante mais complexos os processos de estimação. Por essa razão, a escolha da técnica usualmente utilizada recai no **método de máxima verossimilhança.**

Esse método é uma técnica de estimação bastante usual na inferência estatística, à medida que as estatísticas assim obtidas apresentam muitas das propriedades aconselháveis para a escolha dos estimadores.

O princípio da verossimilhança sustenta que, se o modelo está corretamente identificado, todas as informações que os dados observados fornecem sobre os parâmetros estão contidas na função de verossimilhança, e todos os demais aspectos informativos das observações são irrelevantes.

Para grandes amostras, as estimativas da máxima verossimilhança possuem as propriedades assintóticas de:

- eficiência, com menor variância que qualquer outro estimador;
- consistência;
- no caso de haver um estimador suficiente, ele será o produzido pelo estimador de máxima verossimilhança;
- ser normalmente distribuído, com facilidade de cálculo para a obtenção dos parâmetros média e variância.

A idéia básica do método de máxima verossimilhança é achar grandezas populacionais que gerem os valores que mais se assemelhem aos da amostra observada, ou seja, o método busca estabelecer os valores populacionais hipotéticos que maximizam a verossimilhança da amostra observada. Dito de outra maneira, o método consiste em selecionar aqueles estimadores que maximizam a probabilidade de se obter a amostra realmente observada.

No caso de um modelo de regressão, o objetivo do processo de estimação de um modelo ARIMA(p,d,q) passa a ser achar um vetor dos parâmetros auto-regressivos $\underline{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)$ e um vetor dos parâmetros médias móveis $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)$ tais que minimizem a soma das diferenças quadradas entre os pontos observados na amostra e os esperados pela estimativa obtida com esses parâmetros estimados: $\underline{\hat{\phi}} = (\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p)$, $\underline{\hat{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q)$ e os resíduos $\hat{a}_t$ associados com os valores desses parâmetros. Ou seja, simbolicamente, deve-se achar $\underline{\phi}$, $\underline{\theta}$, tais que

$$S(\underline{\phi},\underline{\theta}) = \Sigma\, a_t^2 \text{ seja um mínimo.}$$

[10] Por simplificação de exposição, optou-se, aqui, por considerar um modelo onde $\theta_0 = 0$. Caso $\theta_0 \neq 0$, então, também esse parâmetro deve ser estimado.

Esses vetores serão $\underline{\hat{\phi}}, \underline{\hat{\theta}}$, e $\hat{a}_t$, de tal forma que, para a equação $\hat{\phi}(B)W_t = \hat{\theta}(B)\,\hat{a}_t$, onde $W_t = \nabla^d X_t$, $S(\underline{\hat{\phi}}, \underline{\hat{\theta}}) = \Sigma\, \hat{a}_t^2$ é um mínimo.

Como se supõe que o processo de ruído branco é normalmente distribuído com $E(a_t)=0$ e $Var(a_t) = \sigma_a^2$, para qualquer t, então as estimativas de máxima verossimilhança são assintoticamente equivalentes aos estimadores de mínimos quadrados.

O processo de estimação desses parâmetros é bastante complexo. A minimização da equação $S(\underline{\phi},\underline{\theta}) = \Sigma a_t^2$ requer um método iterativo de cálculo, à medida que os valores de a_t dependem dos valores passados e não observáveis de W_t e a_t. Assim, por exemplo, o primeiro termo do ruído branco, a_1, depende dos valores passados e não observáveis $w_o, w_{-1}, w_{-2}, \ldots, w_{-p+1}$ e a_o, a_{-1}, a_{-2}, ..., a_{-q+1}. Conseqüentemente, algum critério deve ser usado para estabelecer esses valores não observáveis e assim iniciar a obtenção dos valores de a_t, antes que o processo de estimação, propriamente, seja aplicado.

Box & Jenkins destacam dois procedimentos para a obtenção dos a_t. O primeiro, chamado de condicional, onde, partindo de suposições razoáveis do ponto de vista teórico, são atribuídos valores aos a_t não observáveis; o segundo, incondicional, onde os valores não observáveis são estimados a partir da amostra de dados.

Critério condicional

Para um processo ARIMA(p,d,q), a série X_t possui n + d observações: x_{-d+1}, x_{-d+2}, x_{-d+3}, ..., x_o, x_1, ..., X_n. Através da aplicação do operador diferença "d" vezes, ∇^d, pode-se transformar a série X_t em $W_t = \nabla^d X_t$, com "n" observações w_1, w_2, ..., w_n, onde W_t é especificado por um ARIMA (p,q). Assim, o problema de estimação de $\underline{\phi}$ e $\underline{\theta}$ do modelo ARIMA(p,d,q), especificado para a série X_t, é equivalente a estimar $\underline{\phi}$ e $\underline{\theta}$ do modelo ARMA(p,q) especificado por:

$$a_t = \tilde{w}_t - \phi_1\tilde{w}_{t-1} - \ldots - \phi_p\tilde{w}_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q a_{t-q}, \qquad (4.2.1)$$

onde $\tilde{w}_t = w_t - u$, $E(w_t) = u$.

Como a_t é um processo de ruído branco tal que, para qualquer "i", a_i é normalmente distribuído com média zero e variância σ_a^2, a função de verossimilhança para uma dada série W_t será log condicional, associada com os valores paramétricos $(\underline{\phi},\underline{\theta},\sigma_a^2)$, condicionadas as escolhas dos vetores passados e não observáveis

$w_* = (w_o, w_{-1}, w_{-2}, \ldots, w_{-p+1})$ e $a_* = (a_o, a_{-1}, a_{-2}, \ldots, a_{-q+1})$.

Assim, a função log condicional de verossimilhança, associada aos vetores paramétricos $\underline{\phi}$, $\underline{\theta}$ e ao desvio padrão residual σ_a, é

$$L_*(\phi,\theta,\sigma_a) = -n \ln \sigma_a - \frac{S_*(\underline{\phi},\underline{\theta})}{2\sigma_a^2}$$

Como tanto a função de verossimilhança $L_*(\phi,\theta,\sigma_a)$ como a função soma dos quadrados $S_*(\underline{\phi},\underline{\theta})$ são funções condicionadas aos valores passados não observáveis de $\underline{w}_t$ e $\underline{a}_t$, as estimativas obtidas pela função de verossimilhança dependem dos valores que são fixados para esses vetores. De outra parte, a maximização da função de verossimilhança é obtida minimizando a função soma dos quadrados $S_*(\underline{\phi},\underline{\theta})$.

Entre as alternativas disponíveis para a fixação dos valores iniciais para os elementos dos vetores não observáveis de w_t e de a_t, a mais simples e que proporciona uma aproximação satisfatória é considerar o conjunto de elementos w_* e a_* igual aos seus valores esperados incondicionais. Os valores esperados incondicionais para os elementos a_* são todos zero e, se o modelo não contém a parte determinística (ou seja, $\theta_0 = 0$), a média do processo também é zero, o que implica ter todos os elementos de w_*, em valor esperado incondicional, também iguais a zero.

Embora essa mecânica de iniciação da série seja bastante boa, gerando estimativas suficientemente aproximadas aos valores reais para a maioria das séries usualmente analisadas na prática, essa aproximação pode não ser satisfatória se as raízes da equação $\phi(B) = 0$ são próximas do círculo unitário e o número de observações não é grande em relação à ordem (p,q) do modelo.

Nesse caso, existem outros procedimentos que estabelecem valores esperados condicionados para w_* e a_*. Esses procedimentos são, em geral, extensos e trabalhosos em termos de operações de cálculo necessárias e, muitas vezes, sua eficiência em melhoria da precisão das estimativas dos parâmetros pode não ser significativa.

A sensibilidade dos valores de iniciação da série dependerá do tamanho da amostra que será modelada relativamente aos valores estabelecidos para "p" e "q". Se a série é pequena em relação a "p" e "q", haverá, possivelmente, algum ganho de eficiência nas estimativas. Porém, se a série é relativamente grande, a função condicional soma dos quadrados será, aproximadamente, a mesma que a função incondicional.

No caso de se necessitar do ganho de precisão originado pelo uso da função condicional, um procedimento usual para estabelecer valores esperados condicionais é iniciar um processo iterativo tomando w_* e a_* como sendo zero. Estima-se o modelo ARMA por minimização de $S_*(\underline{\phi},\underline{\theta})$ condicional a esses valores zero, gerando novos valores passados para w_*.[11]

[11] Note-se que, se o modelo não contém a parte auto-regressiva, os dois procedimentos são equivalentes. Em modelos sazonais, entretanto, o primeiro procedimento não apresenta aproximação satisfatória. Ver detalhes em BOX & JENKINS (1976).

Um exemplo prático permite maior clareza no exposto.

Seja uma série descrita por um modelo ARIMA (0,d,1). Então,

$$W_t = a_t - \theta_1 a_{t-1}, \text{ onde } W_t = \nabla^d X_t$$

Nesse caso, a função soma dos quadrados pode ser expressa por:

$$S_*(\theta_1) = \Sigma a_t^2(\theta_1) = \sum_{t=1}^{n} (w_t + \theta_1 a_{t-1}(\theta_1))^2.$$

Em concordância com o que foi exposto, tomando-se $a_o = 0$, tem-se:

$$a_1(\theta_1) = w_1 + \theta_1 a_o(\theta_1) = w_1 + \theta_1(0) = w_1 \ ;$$

$$a_2(\theta_1) = w_2 + \theta_1 a_1(\theta_1) = w_2 + \theta_1 w_1 \ ;$$

$$a_3(\theta_1) = w_3 + \theta_1 a_2(\theta_1) = w_3 + \theta_1(w_2 + \theta_1 w_1) = w_3 + \theta_1 w_2 + \theta_1^2 w_1 \ ;$$

.....

.....

.....

$$a_t(\theta_1) = w_t + \ldots + \theta_1^{t-1} w_1 + \theta_1^t a_o \ .$$

Como $|\theta_1|<1$ (condição de inversibilidade do modelo MA(1)), para "n" suficientemente grande, a escolha do a_o terá pouca importância no valor de $a_t(\theta_1)$. Entretanto, se "n" for pequeno e θ_1 estiver próximo do círculo unitário, a escolha do a_o poderá tornar-se significativa.

Na prática, os a_t podem ser calculados recursivamente.

Seja o modelo acima tal que $\theta_1=0,4$ e os valores W_t da série expostos no Quadro 2, então $a_t = w_t + 0,4\ a_{t-1}$. Para iniciar o processo, supõe-se que $a_o=0$. O Quadro 2 apresenta os valores obtidos para os a_t.

Quadro 2

Obtenção dos valores de a_t em um modelo $W_t=(1-0,4B)a_t$

t	W_t	a_t	$0,4a_{t-1}$
0	-	0	-
1	-3,0	-3,0000	0
2	-4,0	-5,2000	-1,2000
3	4,0	1,9200	-2,0800
4	5,0	5,7680	0,7680
5	-4,0	-1,6928	2,3072
6	6,0	5,3229	-0,6771
7	-3,0	-0,8708	2,1292
8	2,0	1,6517	-0,3483
9	5,0	5,6607	0,6607
10	-2,0	-0,2643	2,2643

A soma condicional dos quadrados é

$$S_*(0,4) = \sum_{t=1}^{10} a_t^2 \ (0,4/a_o=0) = 139,7948 \qquad (4.2.2)$$

Critério incondicional

Um outro procedimento proposto por Box & Jenkins para a obtenção da função soma dos quadrados é através da função log incondicional de verossimilhança.

Pode-se demonstrar[12] que a função log incondicional de verossimilhança, para uma série com número de observações suficientemente grande em relação à ordem (p,q) do modelo, é, aproximadamente,

$$L(\phi,\theta,\sigma_a) = -n \ln \sigma_a - \frac{S(\underline{\phi},\underline{\theta})}{2\,\sigma_a^2},$$

onde a soma incondicional dos quadrados vale

$$S(\underline{\phi},\underline{\theta}) = \sum_{t=-\infty}^{n} E(a_t/\underline{\phi},\underline{\theta},\underline{W})^2 = \Sigma[a_t]^2.$$ [13]

[12] BOX & JENKINS, (1976).

[13] Para simplificação de notação, daqui para diante considerar-se-á $[a_t] = E(a_t/\underline{\phi},\underline{\theta},\underline{W})$.

Para o cálculo de $S(\underline{\phi},\underline{\theta})$, os $[a_t]$ são obtidos aplicando o operador expectância condicional em (4.2.1). Nesse caso, é necessário começar um processo recursivo de previsão para trás, para calcular os w_{-j}, j=0, 1, 2, ...

Note-se que a série estacionária e inversível W_t tem seu modelo ARMA descrito tanto por

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

como por

$$\phi(F)\ W_t = \theta(F)\ e_t,$$

onde F é o operador para diante, tal que $FW_t = W_{t+1}$; $F^2W_t = W_{t+2}$; ..., conforme o definido no Capítulo 3, e e_t é um processo de ruído branco com $E(e_t) = 0$ e $Var(e_t) = Var(a_t) = \sigma_a^2$.

Assim, pode-se obter $\hat{w}_o$, $\hat{w}_{-1}$, $\hat{w}_{-2}$, ..., $\hat{w}_{-p+1}$, a partir de um processo de previsão para trás.

Tomando-se o mesmo modelo exemplificado anteriormente, o processo pode ser exposto como segue.

Um ARIMA(0,d,1) pode ser expresso de duas maneiras:

$W_t = (1=\theta_1 B)a_t$ ou $W_t = (1-\theta_1 F)e_t$

Portanto deve-se calcular

$$\left[a_t\right] = \left[w_t\right] + \theta_1 \left[a_{t-1}\right] \qquad (4.2.3)$$

$$\left[e_t\right] = \left[w_t\right] + \theta_1 \left[e_{t+1}\right] \qquad (4.2.4)$$

Nesse caso, a relação (4.2.4) proporciona as previsões para trás e (4.2.3) os $[a_t]$.

Numericamente, o exemplo pode ser assim ilustrado:

seja $\theta_1 = 0,4$. Então $W_t = (1 - 0,4\ F)e_t$.

$\left[e_t\right] = \left[w_t\right] + 0,4\ \left[e_{t+1}\right]$ e $\left[a_t\right] = \left[w_t\right] + 0,4\ \left[a_{t-1}\right]$, e o processo recursivo de cálculo está exposto no Quadro 3.

Quadro 3

Obtenção dos valores de a_t em um modelo $W_t=(1-0,4F)e_t$

t	$[w_t]$	$[a_t]$	$0,4\,[a_{t-1}]$	$[e_t]$	$0,4\,[e_{t+1}]$
-1	0	0	0	0	0
0	1,4749	1,4749	0	0	-1,4749
1	-3,0	-2,4100	0,5900	-3,6872	-0,6872
2	-4,0	-4,9640	-0,9640	-1,7180	2,2820
3	4,0	2,0144	-1,9856	5,7049	1,7049
4	5,0	5,8058	0,8058	4,2622	-0,7378
5	-4,0	-1,6777	2,3223	-1,8445	2,1555
6	6,0	5,3289	-0,6711	5,3888	-0,6112
7	-3,0	-0,8684	2,1316	-1,5280	1,4720
8	2,0	1,8684	-0,3474	3,6800	1,6800
9	5,0	5,6610	0,6610	4,2000	-0,8000
10	-2,0	0,2644	2,2644	-2,0000	0

A soma incondicional dos quadrados é

$$S(0,4) = \sum_{t=1}^{10} (a_t/0,4;\ \underline{W})^2 = 135,0284\ . \qquad (4.2.5)$$

Note-se que, para esse exemplo, utilizando um MA(1), tem-se:

$[W_t] = w_t$, t=1, 2, ..., n é previsto para trás se $t \leq 0$;

$[e_o]$, $[e_{-1}]$, $[e_{-2}]$, ... são nulos, pois e_o, e_{-1}, e_{-2}, ..., são independentes de $\underline{w}$;

$[a_{-1}]$, $[a_{-2}]$, ... são nulos, pois, em um MA(q), a_{-q}, a_{-q+1}, ... são independentes de $\underline{w}$;

para iniciar o processo que se arbitrar $\left[e_{11} \right] = 0$. Disso segue, por (4.2.4), que

$$\left[e_{10} \right] = \left[w_{10} \right] + 0,4 \left[e_{11} \right] = -2,0 + 0 = -2,0$$

$$\left[e_{9} \right] = \left[w_{9} \right] + 0,4 \left[e_{10} \right] = 5,0 + 0,4\ (-2,0) = 5,0 - 0,8 = 4,2$$

.....
.....
.....

$$\left[e_{0} \right] = \left[w_{o} \right] + 0,4 \quad e_{1} <-> 0 = \left[w_{o} \right] + 0,4(3,6872) <-> w_{o} = 1,4749$$

$$\left[w_{-j} \right] = 0,\ j = 1, 2, \ldots$$

e, por (4.2.3), obtém-se os $\left[a_{t} \right]$.

Observe-se que, mesmo para uma série pequena, a soma condicional dos quadrados (4.2.2) é próxima da soma incondicional dos quadrados (4.2.5).

Após a obtenção dos valores de a_t, que permitem calcular a soma dos quadrados, está-se apto a atender aos propósitos da estimação, ou seja, achar estimativas dos parâmetros que minimizem a soma dos quadrados.

No exemplo dado, uma maneira de obtenção do parâmetro θ_1 que minimize $S(\theta_1)$ seria calcular o processo para vários valores de θ_1, acompanhando graficamente a evolução dos respectivos $S(\theta_1)$, para então concluir pelo mais eficiente (isto é, por aquele que minimize $S(\theta_1)$).

Na prática, entretanto, para um modelo mais complexo, envolvendo mais de um parâmetro, tornam-se necessários procedimentos metodológicos mais sofisticados para a obtenção precisa dos parâmetros mais eficientes.

O processo de cálculo sugerido por Box & Jenkins emprega uma rotina iterativa baseada na expansão da série de Taylor[14] para a linearização da equação geral do modelo ARIMA, a partir de valores iniciais para $\underline{\phi}$ e $\underline{\theta}$.

Esse procedimento pode ser assim exposto:

Seja _ o vetor que representa os k = p + q parâmetros $(\underline{\phi}, \underline{\theta})$ que se quer estimar. Assim, a equação a ser minimizada,

$$S(\underline{\phi}, \underline{\theta}) = \Sigma \left[a_t / \underline{\phi}, \underline{\theta}, \underline{w} \right]^2 = \Sigma \left[a_t \right]^2 ,$$

pode ser escrita por:

$$S(\underline{\beta}) = \Sigma \left[a_t / \underline{\beta}, \underline{w} \right]^2 = \Sigma \left[a_t \right]^2 .$$

[14] Essa rotina iterativa para o cálculo dos valores expandidos da série de Taylor aqui utilizada é a desenvolvida por MARQUARDT (1963).

Expandindo $\left[a_t/\underline{\beta},\underline{w}\right]$ em série de Taylor (com aproximação para somente as duas primeiras parcelas), a partir dos valores iniciais de $\beta_o = (\beta_{1,0}, \beta_{2,0}, \ldots, \beta_{k,0})$, tem-se que:

$$\left[a_t/\underline{\beta},\underline{w}\right] = \left[a_t/\underline{\beta}_o,\underline{w}\right] - \sum_{i=1}^{p+q} (\beta_i - \beta_{i,o})\, X_{i,t}, \text{ onde}$$

$$X_{i,t} = - \left.\frac{\partial (a_t/\underline{\beta}_o,\underline{w})}{\partial\, \underline{\beta}_i}\right|_{\beta = \beta_o}$$

ou

$$\left[a_t/\underline{\beta}_o,\underline{w}\right] + \sum_{i=1}^{p+q} \beta_{i,o} X_{i,t} = \sum_{i=1}^{p+q} \beta_i X_{i,t} + \left[a_t/\underline{\beta},\underline{w}\ .\right]$$

Essa última equação pode ser vista como uma regressão linear múltipla, onde o lado esquerdo é o vetor coluna $\mathbb{Y}$, composto por t = 1, 2, ..., n variáveis aleatórias dependentes,

$$\mathbb{Y} = \begin{bmatrix} \left[a_1/w_1\beta_1\right] - \sum_{i=1}^{p+q} \beta_{i,o}\, X_{i,1} \\ \left[a_2/w_1\beta_2\right] - \sum_{i=1}^{p+q} \beta_{i,o}\, X_{i,2} \\ \vdots \\ \left[a_T/w_1\beta_n\right] - \sum_{i=1}^{p+q} \beta_{i,o}\, X_{i,n} \end{bmatrix}$$

e o lado direito composto pela matriz n x (p+q), de p+q variáveis independentes, multiplicadas pelo vetor p+q x 1 de ... p+q parâmetros desconhecidos que se desejam estimar e mais o vetor de termos residuais:

$$\mathbb{X} = \begin{bmatrix} X_{1,1} & X_{2,1} & \cdots & X_{p+q,1} \\ X_{1,2} & X_{2,2} & \cdots & X_{p+q,2} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ X_{1,n} & X_{2,n} & \cdots & X_{p+q,n} \end{bmatrix}; \quad \mathbb{B} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{p+q} \end{bmatrix}; \quad \mathbb{a} = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix}$$

Sinteticamente, $\mathbb{Y} = \mathbb{X}\mathbb{B} + \mathbb{a}$.

Porque os $[a_t/\underline{\beta},\underline{w}]$ não são exatamente lineares em relação aos $\underline{\beta}$, o ajustamento a partir de $\underline{\beta}_o$ não produz estimativas de mínimos quadrados imediatamente. Essas estimativas iniciais servem para iniciar um mecanismo iterativo de aproximações sucessivas, a partir dos valores $\underline{\beta}_o$ anteriores. O processo é repetido n vezes até a ocorrência de convergência, ou seja, até que $\underline{\beta}_n - \underline{\beta}_{n-1} \cong 0$.

A velocidade de convergência dependerá dos valores iniciais tomados para $\underline{\beta}_o$. Se os $\underline{\beta}_o$ são próximos dos $\underline{\beta}_n$ finais, a convergência ocorre após poucas etapas iterativas. Porém, se os $\underline{\beta}_o$ são muito distantes dos $\underline{\beta}_n$ finais, o processo pode necessitar de um número bastante grande de etapas iterativas, podendo, inclusive, em alguns casos, não convergir totalmente. Em termos práticos, os valores iniciais obtidos no processo de identificação, conforme o estudo feito no item 4.1, são, em geral, bons valores iniciais.

Felizmente, as dificuldades de cálculo desse processo iterativo apresentam-se diminutas, à medida que as rotinas programadas para elaboração por meios eletrônicos estão bastante desenvolvidas, proporcionando rapidamente estimativas satisfatórias.

Uma avaliação da precisão dos parâmetros assim estimados pode ser feita a partir da variância desses estimadores e da construção dos respectivos intervalos de confiança e testes de hipótese.

Da teoria da inferência estatística sabe-se que os estimadores de máxima verossimilhança têm distribuição aproximadamente normal, se "n" é grande, com expectância igual aos verdadeiros parâmetros, e sua matriz de variância-covariância é dada por:

$$\mathrm{Var}(\underline{\beta}) \cong 2\,\sigma_a^2 \begin{bmatrix} \dfrac{\partial^2 S_{(\underline{\beta})}}{\partial \beta_1^2} & \cdots\cdots\cdots\cdots & \dfrac{\partial^2 S_{(\underline{\beta})}}{\partial \beta_1\ \partial \beta_{p+q}} \\ \vdots & & \vdots \\ \dfrac{\partial^2 S_{(\underline{\beta})}}{\partial \beta_{p+q}\partial \beta_1} & \cdots\cdots\cdots\cdots & \dfrac{\partial^2 S_{(\underline{\beta})}}{\partial \beta^2_{p+q}} \end{bmatrix}^{-1}$$

Como $\hat{\sigma}_a^2 = \dfrac{S(\beta)}{n}$, substituindo σ_a^2 por sua estimativa $\hat{\sigma}_a^2$ e calculando os valores da matriz, obtém-se estimativas para a variância de seus parâmetros.

Para os modelos geralmente utilizados na prática, as variâncias estimadas estão sintetizadas a seguir.

$$\mathrm{AR(1)} \rightarrow \mathrm{VAR}(\hat{\phi}_1) \cong \frac{1-\phi_1^2}{n}$$

$$\mathrm{AR(2)} \rightarrow \mathrm{VAR}(\hat{\phi}_2) = \mathrm{VAR}(\phi_1) \cong \frac{1-\phi_2^2}{n}$$

$$MA(1) \rightarrow VAR(\hat{\theta}_1) \cong \frac{1-\theta_1^2}{n}$$

$$MA(2) \rightarrow VAR(\hat{\theta}_2) = VAR(\hat{\theta}_1) \cong \frac{1-\theta_2^2}{n}$$

$$ARMA(1,1) \rightarrow VAR(\hat{\phi}_1) \cong \frac{1-\phi_1^2}{n} \frac{(1-\phi_1\theta_1)^2}{(\phi_1 - \theta_1)^2}$$

$$VAR(\hat{\theta}_1) \cong \frac{1-\theta_1^2}{n} \frac{(1-\phi_1\theta_1)^2}{(\phi_1 - \theta_1)^2}$$

4.1.3 – A Checagem do Diagnóstico

Após o modelo ter sido identificado e seus parâmetros eficientemente estimados, chega-se à última etapa do processo de modelagem, denominada de checagem do diagnóstico, que consiste em averiguar a adequação do modelo utilizado para ajustar os dados observados na vida real.

Essa etapa apresenta-se importante, pois, se a verificação mostrar alguma evidência de que o modelo ajustado é inadequado para descrever os valores reais da série, é necessário que ela própria (a etapa da verificação) aponte as causas da inadequação e sugira as modificações apropriadas.

As técnicas existentes atualmente são ainda pouco evoluídas, portanto, precárias, não permitindo que se obtenham respostas científicas plenamente conclusivas acerca da adequação do modelo identificado.

Uma das técnicas recomendadas por Box & Jenkins para a checagem do modelo é o chamado **sobreajustamento**. É baseada no princípio de que o modelo identificado na primeira etapa do processo, embora seja capaz de descrever a série, está sob a suspeição de que necessita um modelo mais elaborado, com um número maior de parâmetros. Justifica-se esse procedimento pelo fato de o modelo mais elaborado conter parâmetros adicionais que permitem descrever melhor as dúvidas e discrepâncias não abrangidas pelo modelo inicialmente identificado. Assim, se a análise de um modelo estimado com parâmetros extras não indicar que os parâmetros são significativos e a variância residual não diminuir, pode-se supor que as adições de novos parâmetros não são necessárias, e que o modelo identificado consegue descrever razoavelmente a série observada.

O método do sobreajustamento, portanto, é uma técnica bastante precária, pois somente estende o modelo para ordens superiores, assumindo que se conhece o tipo de discrepância cuja identificação ainda proporciona algumas dúvidas.

Outro procedimento menos dependente do conhecimento inicial do modelo e de uso mais corrente no processo de checagem do diagnóstico está baseado na análise dos resíduos.

Esse processo de checagem, em geral, envolve dois estágios. Em primeiro lugar, a função de autocorrelação para a série gerada pelo modelo deve ser comparada com a função de autocorrelação da série de dados realmente observados. Se as duas funções de autocorrelação se apresentam bastante distintas entre si, pode-se imaginar que o modelo especificado necessita de uma melhor identificação. Por outro lado, se as duas funções de autocorrelação apresentam comportamentos semelhantes, então se passa para o segundo estágio, que é uma análise quantitativa dos resíduos gerados pelo modelo identificado.

A análise dos resíduos está baseada no fato de testar-se se realmente os resíduos formam um processo de ruído branco. Fundamentalmente, busca-se testar se os resíduos não estão correlacionados entre si.

Assim, se o modelo está bem especificado, espera-se que os $\hat{a}_t$ ($\hat{a}_t = \theta^{-1}(B)\phi(B)w_t$) sejam não correlacionados entre si, e que a função de autocorrelação dos resíduos

$$\hat{r}_k \quad (\hat{r}_k = \frac{\Sigma \hat{a}_t . \hat{a}_{t-k}}{\Sigma \hat{a}_t}) \text{ se aproxime de zero para deslocamentos } k \geq 1.$$

O teste de independência dos resíduos, sugerido por Box & Jenkins, é o desenvolvido por BOX & PIERCE (1970) a partir das constatações de ANDERSON (1942) para as funções de autocorrelação. Esse teste comprova que, se o modelo está corretamente especificado, então, para grandes deslocamentos de "k", os coeficientes de autocorrelação $\hat{r}_k$ estão não correlacionados entre si, e estão normalmente distribuídos com média zero e variância 1/n, onde n é o número de observações da série W_t.

Box & Pierce definiram a estatística $Q = n \Sigma \hat{r}_k^2$, constituída por uma soma de variáveis normais de média zero e variância 1/n elevada ao quadrado, a qual, portanto, possui distribuição qui-quadrado[15], com k-p-q graus de liberdade. A regra de decisão consiste em que um valor de "Q" abaixo de 10% da cauda direita na tabela qui-quadrado indica que não é, necessariamente, aceitável a hipótese de os resíduos serem processos de ruído não branco, uma vez que a probabilidade dessa hipótese ser verdadeira é menor que 90%. O teste do qui-quadrado é, portanto, um teste fraco, uma vez que testa somente de forma indireta a hipótese dos resíduos formarem um processo de ruído branco.

De outra parte, se o valor de "Q" calculado estiver entre os pontos compreendidos pelas áreas de 10% e 5% da tabela, alguma dúvida sobre a adequação da identificação do modelo permanecerá, necessitando-se buscar outros meios para dirimir as incertezas sobre a especificação do modelo.

Por último, após ter decidido aceitar a adequação do modelo identificado e estimado os valores ótimos dos respectivos parâmetros, é, ainda, aconselhável mais avaliação sobre a representatividade dos parâmetros estimados. É possível que o modelo apresente um número não conveniente de parâmetros, os quais podem caracterizar uma sub ou uma superparametrização do modelo.

---

[15] Na verdade, "Q" é aproximadamente qui-quadrado, pois poderá haver correlação entre as primeiras $\hat{r}_k$ e sua variância ser menor que 1/n.

Para o caso de se objetivar a verificação da existência de parâmetros em demasia no modelo ajustado (modelos superparametrizados), o teste utilizado é comparar os valores dos parâmetros com o dobro de seus respectivos desvios padrão (vale dizer, α=5%). Se houver parâmetros inferiores ao dobro de seus desvios padrão, em módulo, esses são possíveis candidatos a serem eliminados.

Se o parâmetro não significante é o de ordem mais elevada,o modelo pode ser simplificado, reduzindo-se de um a ordem do operador.

Exemplo: seja o modelo $\nabla x_t=(1 - 0,47B - 0,05B^2)a_t$ com desvios de 0,02 e 0,07, respectivamente, para $\hat{\theta}_1$ e $\hat{\theta}_2$.

O parâmetro 0,47 é significante, pois $\hat{\theta}_1 > 2DP(\hat{\theta}_1)$,isto é, 0,47 > 2 x 0,02.

O parâmetro 0,05 é não significativo, pois é menor que 2 x $DP(\hat{\theta}_2)$, isto é, 2 x 0,07.

Sugere-se, então, o modelo alternativo $\nabla X_t = (1 - 0,47B)a_t$.

Se o parâmetro não significante não é o de ordem mais elevada, deve-se verificar o grau de correlação entre os parâmetros através de sua matriz de correlação. Se a matriz indica correlação entre os parâmetros, pode-se eliminar a de maior ordem; se não houver correlação entre os parâmetros, então,mantêm-se todos os parâmetros.

Exemplo: seja o modelo $Z_t - Z_{t-1} = (1 - 0,08B - 0,56B^2)a_t$, com desvios padrão dos parâmetros de 0,09 e 0,01 respectivamente.

O parâmetro 0,08 é não significante, pois é menor que 2 x 0,09.

O parâmetro 0,56 é significante, pois é maior que 2 x 0,01.

Se a correlação entre os dois parâmetros for, por exemplo, de 0,95, ou seja, se existe correlação entre eles,então se sugere o modelo $Z_t - Z_{t-1} = (1 - 0,08B)a_t$.

Um último caso é aquele em que todos os parâmetros se apresentam com significância, mas altamente correlacionados. Nesse caso, deve-se optar pela redução do número de parâmetros, baixando a ordem do modelo. Após cada simplificação, os valores ótimos devem ser reestimados.

Para o caso de modelos subparametrizados, onde o modelo especificado apresenta um número de parâmetros menor do que o necessário para descrever adequadamente o comportamento dos dados, o processo de análise de sua subparametrização é baseado no fato de os resíduos obtidos após a estimação dos valores dos parâmetros terem autocorrelações significantes. Sua avaliação é feita através dos seguintes procedimentos:

- se os valores dos coeficientes de autocorrelação são significantes nas defasagens iniciais e/ou nos períodos múltiplos de sazonalidade,isso indica que os comportamentos regular e/ou sazonal não foram considerados na íntegra. Devem-se buscar as causas da deficiência através da comparação dos coeficientes significantes desse modelo com os coeficientes conhecidos para os operadores AR e MA. Obtidas as razões,elimina-se a deficiência pela agregação de tais operadores ao modelo;

- se os coeficientes de autocorrelação significantes ocorrem em defasagens distintas das expostas acima, não apresentando então a necessidade de operadores regulares e/ou sazonais adicionais e, além disso,

aparecem esporadicamente no correlograma (por exemplo, 5% das autocorrelações são significantes), então, a série de resíduos deve ser aceita como aleatória e o modelo considerado satisfatório.

Exemplo: seja um modelo MA(1): $w_t = (1-\theta_1 B)a_t$.

Se os resíduos apresentam autocorrelação significante na defasagem 1, isso está a indicar a necessidade de um operador MA adicional de ordem 1. Sugere-se o modelo MA(2): $w_t = (1 - \theta_1 B - \theta_2 B^2)a_t$.

Exemplo: seja o modelo sazonal $\nabla_{12} \ln w_t = (1 - 0.4B)(1 - 0.5B^{12})a_t$ com desvios padrão dos parâmetros 0,06 e 0,04 repectivamente.

No caso, ambos os parâmetros são significantes e devem permanecer no modelo (a correlação entre eles é zero) e não existe qualquer evidência que indique superparametrização. Fazendo uma análise das autocorrelações dos resíduos após o ajustamento, essa não mostra melhoria do modelo.

Poderá, ainda, ocorrer um caso especial, onde a supere a subparametrização estão presentes simultaneamente. Isso acontece quando se necessita tanto uma simplificação como uma melhoria do modelo. Ocorre quando estão presentes tanto parâmetros sazonais que se desejam simplificar, quanto parâmetros regulares, que se pretendem incorporar.

4.2 – Modelagem de Séries Sazonais: Modelo SARIMA (p, d, q) x (P, D, Q)

O processo de modelagem para séries de tempo que apresentam variações cíclicas de curto prazo é bastante similar àqueles em que a sazonalidade é inexistente. Fundamentalmente, essas diferenças residem em se precisar estabelecer um número maior de parâmetros e identificar valores para a ordem (p,d,q)x(P,D,Q). Como o processo de estimação dos parâmetros de um modelo SARIMA segue os mesmos princípios de um modelo não sazonal, e essa técnica de cálculo encontra-se embutida nos pacotes computacionais, o problema do modelador reduz-se somente a estabelecer valores para a ordem (p,d,q)x(P,D,Q).

Assim, para se obter "d" e "D", deve-se utilizar, conjuntamente, os operadores $\nabla^d = (1 - B)^d$ e $\nabla_S^D = (1 - B^s)^D$, de maneira a tornar $W_t = \nabla^d \nabla_S^D X_t$ estacionária.

A obtenção das ordens "p", "P", "q", "Q" é feita, como no modelo não sazonal, através das funções de autocorrelação e autocorrelação parcial da série $W_t = \nabla^d \nabla_S^D X_t$. Para o modelo multiplicativo, as conclusões obtidas pela existência de correlações diferentes de zero nas primeiras defasagens devem ser agora vistas conjuntamente com as defasagens sazonais.

As funções de autocorrelação e autocorrelação parcial do processo multiplicativo são geradas pelo produto das funções de seus componentes não sazonais e sazonais. Dessa maneira, as conclusões obtidas pela exis-

tência de correlações diferentes de zero nas primeiras defasagens são indicações dos valores das ordens "p" e "q" e nas defasagens sazonais das ordens "P" e "Q".

Por exemplo, para um processo $(0,1,1)x(0,1,1)_{12}$, os coeficientes de autocorrelações da série $\nabla\nabla_{12}X_t$ devem ser diferentes de zero somente nas defasagens 1, 11, 12, 13.

5 – O MÉTODO DE PREVISÃO DE BOX & JENKINS

5.1 – Introdução

Após o término do processo iterativo de identificação, estimação e checagem do diagnóstico, o qual proporciona uma estimativa do modelo gerador da série que se ajusta aos dados da realidade, pode-se utilizá-lo para prever valores futuros da variável.

Entretanto deve-se sempre ter presente que, na prática, nunca se tem o processo gerador verdadeiro da realização amostral. Dispõe-se apenas de uma aproximação desse processo gerador, a qual está sujeita a erros tanto no que se refere à identificação quanto à estimação. É claro, então, que o sucesso do processo de previsão dependerá, fundamentalmente, da qualidade representativa do modelo teórico estimado, a qual depende do trabalho nas etapas de identificação e estimação.

Os processos de previsão com modelos de séries de tempo são procedimentos que visam a estender a valores futuros o modelo descrito e ajustado aos valores passados e ao valor presente da variável. Dessa maneira, o estudo dos processos de previsão que aqui se analisam envolve um conjunto de técnicas, que busca estabelecer a base teórica que norteia o estudo de previsões de séries univariantes de tempo, proposto por Box & Jenkins, bem como os fundamentos da inferência estatística dos erros de previsão e os conseqüentes intervalos de confiança para aqueles procedimentos, além, é claro, da etapa posterior de cálculo das previsões propriamente ditas.

Ao estender a valores futuros prováveis o modelo de série temporal construído de forma a descrever os valores passados e o valor presente da variável, está-se diante de uma análise que busca obter as características comportamentais sistemáticas da série, vale dizer, que busca apresentar uma descrição dos mecanismos e da natureza aleatória do processo estocástico que gerou a série de valores passados e do valor presente. O estabelecimento da estrutura probabilística da realização amostral da variável proporcionará os meios para inferir conclusões sobre as probabilidades associadas a cada possível valor futuro da variável.

Portanto a previsão se torna o cálculo do valor esperado de uma futura observação, condicionado aos valores passados e ao valor presente da variável. Ou seja, chamando de $\hat{x}_t(h)$ o valor previsto estimado para um horizonte de "h" períodos de tempo futuros e "t" o período de origem da previsão, então,

$$\hat{x}_t(h) = E(x_{t+h}/x_t, x_{t-1}, \ldots).$$

O valor de x no tempo t+h é obtido pela equação do modelo ARIMA.

Como foi visto em capítulos anteriores, uma observação x_{t+h} gerada por um modelo ARIMA $\S(B)x_t = \theta(B)a_t$, onde $\S(B) = \phi(B)\nabla^d$, pode ser expressa de três formas distintas:

- em termos de equações a diferenças

$$x_{t+h} = \xi_1 x_{t+h-1} + \ldots + \xi_{p+d} x_{t+h-p-d} - \theta_1 a_{t+h-1} - \ldots - \theta_q a_{t+h-q} + a_{t+h} \quad ; \qquad (5.1)$$

- em termos de uma soma ponderada do presente e dos passados choques aleatórios

$$x_{t+h} = \sum_{j=-\infty}^{t+h} \psi_{t+h-j} a_j = \sum_{j=0}^{\infty} \psi_j a_{t+h-j} \quad , \qquad (5.2)$$

onde $\psi_o = 1$ e ψ_j são tais que $(1 - \xi_1 B - \ldots - \xi_{p+d} B^{p+d})(1 + \psi_1 B + \psi_2 B^2 + \ldots) =$

$= (1 - \theta_1 B - \ldots - \theta_q B^q)$;

- em termos de uma soma ponderada das observações passadas mais um choque aleatório no período presente

$$x_{t+h} = \sum_{j=1}^{\infty} \pi_j x_{t+h-j} + a_{t+h} \quad , \qquad (5.3)$$

onde os pesos π_j são tais que $\xi(B) = (1 - \pi_1 B - \pi_2 B^2 - \ldots)\ \theta(B)$.

A base do estudo que norteia a teoria das previsões é que essa inferência deve estar sujeita ao menor erro possível. Matematicamente, significa dizer que a previsão ótima deve apresentar um erro quadrático médio de previsão mínimo, ou seja, $\hat{x}_t(h)$ deve ser tal que

$E(e_t(h)^2) = E(x_{t+h} - \hat{x}_t(h))^2$ seja um mínimo.

O desenvolvimento dessa expressão estabelece os valores que a satisfazem. Seja o modelo ARIMA $\xi(B) x_t = \theta(B) a_t$, então,

$$x_t = \xi^{-1}(B)\ \theta(B)\ a_t = \psi(B) a_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad .$$

Portanto o valor de x no tempo t+h vale

$$x_{t+h} = \sum_{j=0}^{\infty} \psi_j a_{t+h-j} = \sum_{j=-\infty}^{t+h} \psi_{t+h-j} a_j =$$

$$= \psi_o a_{t+h} + \psi_1 a_{t+h-1} + \ldots + \psi_h a_t + \psi_{h+1} a_{t-1} + \ldots \quad , \qquad (5.4)$$

o que estabelece o valor de x_{t+h} através de uma função linear de valores do processo de ruído branco, do tipo da equação (5.2).

De maneira semelhante, pode-se obter os valores de $\hat{x}_t(h)$.

$\hat{x}_t(h) = E(x_{t+h}/x_t, x_{t-1}, \ldots)$ deve ser tal que dependa somente dos valores passados e do valor presente.

$$\text{Seja } \hat{x}_t(h) = \psi^*_h a_t + \psi^*_{h+1} a_{t-1} + \psi^*_{h+2} a_{t-2} + \cdots , \qquad (5.5)$$

dado também através de uma função linear de valores do processo de ruído branco, onde os pesos $\psi^*_{h+j}, j=0,1,2,\ldots$ são determinados de forma que $\hat{x}_t(h)$ seja uma previsão ótima, então,

$$E(e_t(h))^2 = E(x_{t+h} - \hat{x}_t(h)^2 = (\psi_o^2 + \psi_1^2 + \psi_2^2 + \ldots +$$

$$+ \psi_{h-1}^2) \sigma_a^2 + \sum_{j=o}^{\infty} (\psi_{h+j} - \psi^*_{h+j})^2 \sigma_a^2.$$

A expressão acima será minimizada quando os ψ^*_{h+j} forem iguais aos verdadeiros pesos ψ_{h+j}, $j = 0, 1, 2, \ldots$

Então, por (5.4), $\hat{x}_t(h)$ deve ser igual a

$$\hat{x}_t(h) = \psi_h a_t + \psi_{h+1} a_{t-1} + \psi_{t+2} a_{t-2} + \cdots \qquad (5.6)$$

ou, deduzindo pela equação (5.5),

$$E(x_{t+h}/x_t, x_{t-1}, x_{t-2}, \ldots) = E(\psi_o a_{t+h} + \psi_1 a_{t+h-1} + \ldots + \psi_h a_t + \psi_{h+1} a_{t-1} + \ldots) =$$

$$= \psi_o E(a_{t+h}) + \psi_1 E(a_{t+h-1}) + \ldots + \psi_h E(a_t) + \psi_{h+1} E(a_{t-1}) + \ldots =$$

$$= 0 + 0 + \ldots + \psi_h a_t + \psi_{h+1} a_{t-1} + \ldots = \sum_{j=o}^{\infty} \psi_{h+j} a_{t-j} .$$

Portanto a função ótima de previsão é

$$\hat{x}_t(h) = E(x_{t+h}/x_t, x_{t-1}, \ldots) = \sum_{j=o}^{\infty} \psi_{h+j} a_{t-j} .$$

Note-se, por outro lado, que o erro no instante "t" para o avanço "h" pode ser obtido por (5.4) e (5.6):[1]

$$e_t(h) = x_{t+h} - \hat{x}_t(h) = a_{t+h} + \psi_1 a_{t+h-1} + \ldots + \psi_{h-1} a_{t+1} .$$

[1] Note-se que, para $h = 1$, $e_t(1) = x_{t+1} - \hat{x}_t(1) = a_{t+1}$.

Conseqüentemente,

$$E(e_t(h)) = E(x_{t+h} - \hat{x}_t(h)) = 0 \ ;$$

$$Var(e_t(h)) = (1 + \psi_1^2 + \ldots + \psi_{h-1}^2)\,\sigma_a^2 \ .$$

Vale dizer, a previsão é não tendenciosa e sua variância é função da variância do processo de ruído branco, σ_a^2. (5.7)

5.2 – O Processo de Previsão

A obtenção da estimativa do valor futuro provável $\hat{x}_t(h) = E(x_{t+h}/x_t, x_{t-1}, \ldots)$, proposta por Box & Jenkins, é feita através de um processo iterativo onde, primeiro, é calculada a previsão para um período adiante, para posterior utilização desse resultado no cálculo da previsão subseqüente.

Dado o modelo ARIMA (p,d,q), o qual pode ser escrito de três formas distintas, de acordo com (5.1), (5.2) e (5.3), as respectivas expressões para a previsão são apresentadas a seguir.

Previsão utilizando equações a diferenças

O valor de x no tempo t+h é dado por:

$$x_{t+h} = \varphi_1 x_{t+h-1} + \ldots + \varphi_{p+d} x_{t+h-p-d} - \theta_1 a_{t+h-1} - \ldots - \theta_q a_{t+h-q} + a_{t+h} \ .$$

Então a previsão é

$$\hat{x}_t(h) = E(x_{t+h}/x_t, x_{t-1}, \ldots) = \varphi_1 E(x_{t+h-1}/x_t, x_{t-1}, \ldots) + \ldots + \varphi_{p+d} E(x_{t+h-p-d}/x_t,$$

$$x_{t-1}, \ldots) - \theta_1 E(a_{t+h-1}/x_t, x_{t-1}, \ldots) - \ldots - \theta_q E(a_{t+h-q}/x_t, x_{t-1}, \ldots) -$$

$$- E(a_{t+h}/a_t, a_{t-1}, \ldots) \ .$$

Deve-se ter presente que:

$$E(x_{t+h}/x_t, x_{t-1}, \ldots) = x_{t+h} \qquad \text{para qualquer } h \leq 0;$$

$$E(a_{t+h}/x_t, x_{t-1}, \ldots) = a_{t+h} = x_{t+h} - \hat{x}_{t+h-1}(1) \qquad \text{se } h \leq 0;$$

$$E(a_{t+h}/x_t, x_{t-1}, \ldots) = 0 \qquad \text{para qualquer } h > 0.$$

Previsão utilizando equação da soma ponderada do presente e dos passados choques aleatórios

O valor de x no tempo t+h é dado por:

$$x_{t+h} = \sum_{j=-\infty}^{t+1} \psi_{t+h-j} a_j = \sum_{j=0}^{\infty} \psi_j a_{t+h-j} .$$

Então a previsão é

$$\hat{x}_t(h) = \psi_o E(a_{t+h}) + \psi_1 E(a_{t+h-1}) + \ldots + \psi_{h-1} E(a_{t+1}) + \psi_h E(a_t) +$$

$$+ \psi_{h+1} E(a_{t-1}) + \ldots + E(a_{t-h}) .$$

Previsão utilizando equação da soma ponderada das observações passadas mais um choque aleatório

O valor de x no tempo t+h é dado por:

$$x_{t+h} = \sum_{j=1}^{\infty} \pi_j x_{t+h-j} + a_{t+h} .$$

Então a previsão é

$$\hat{x}_t(h) = \pi_1 E(x_{t+h-1}/x_t, x_{t-1} \ldots) + \pi_2 E(x_{t+h-2}/x_t, x_{t-1}, \ldots) + \ldots + E(a_{t+h}) .$$

Exemplo 01: previsão de um MA(1): $X_t = (1-\theta_1 B)a_t + \mu$

O valor de x no tempo t+h é dado por

$$x_{t+h} = (1 - \theta_1 B)a_{t+h} + \mu ,$$

e a previsão é

$$\hat{x}_{t(h)} = E((1-\theta_1 B)a_{t+h} + \mu) = E(a_{t+h} - \theta_1 a_{t+h-1} + \mu) = E(a_{t+h}) - \theta_1 E(a_{t+h-1}) + \mu$$

Então, para h=1,

$$\hat{x}_t(1) = E(a_{t+1}) - \theta_1 E(a_t) + \mu = 0 - \theta_1 a_t + \mu = \mu - \theta_1 a_t ;$$

para $h \geq 2$,

$$\hat{x}_t(h) = E(a_{t+h}) - \theta_1 E(a_{t+h-1}) + \mu = 0 - 0 + \mu = \mu .$$

Como o processo MA(1) tem memória de somente um período, os dados só condicionam as previsões para um período adiante. Para dois ou mais períodos adiante, a melhor previsão é a média do processo.

A variância do erro de previsão vale,

para h=1,

$$E(e_t(1)^2) = E(x_{t+1}-\hat{x}_t(1)^2) = E(a_t^2) = \sigma_a^2 \ ;$$

para $h \geq 2$,

$$E(e_t(h)^2)=E(x_{t+h}-\hat{x}_t(h)^2) = E(a_{t+h}-\theta_1 a_{t+h-1})^2 = (1 - \theta_1^2)\ \sigma_a^2 \ .$$

Vale dizer que a variância do erro de previsão cresce até dois períodos adiante. A partir deste horizonte, torna-se constante.

Exemplo 02: previsão de um AR(1): $(1-\phi_1 B)x_t=a_t+\delta$

Como $(1-\phi_1 B)\ x_t = a_t + \delta \Leftrightarrow x_t - \phi_1 x_{t-1} = a_t + \delta \Leftrightarrow x_t =$

$= \phi_1 x_{t-1} + a_t + \delta$, o valor de x no tempo t+h é dado por

$$x_{t+h} = \phi_1 x_{t+h-1} + a_{t+h} + \delta \ ,$$

e a previsão é

$$\hat{x}_t(h) = E(\phi_1 x_{t+h-1}) + E(a_{t+h}) + \delta = \phi_1 E(x_{t+h-1}) + E(a_{t+h}) + \delta \ .$$

Então, para h=1,

$$\hat{x}_t(1) = \phi_1 E(x_t) + E(a_{t+1}) + \delta = \phi_1 x_t + 0 + \delta = \phi_1 x_t + \delta \ ;$$

para h=2,

$$\hat{x}_t(2) = \phi_1 E(x_{t+1}) + E(a_{t+2}) + \delta = \phi_1 \hat{x}_t(1) + 0 + \delta = \phi_1(\phi_1 x_t + \delta) + \delta =$$

$$= \phi_1^2\ x_t + \phi_1 \delta + \delta = \phi_1^2\ x_t + (\phi_1+1)\delta \ ,$$

ou seja,

para $h \geq 2$,

$$\hat{x}_t(h) = \phi_1 \hat{x}_t(h-1) + \delta = \phi_1^h x_t + (\phi_1^{h-1} + \phi_1^{h-2} + \ldots + \phi_1 + 1)\,\delta \ .$$

Note-se que

$$\lim_{h \to \infty} \hat{x}_t(h) = \delta \sum_{j=0}^{\infty} \phi^j = \frac{\delta}{1-\phi_1} = \mu \ ,$$

isto é, para horizontes muito distantes do valor de origem, a previsão tende à média do processo.

A variância do erro de previsão vale

$$e_t(h) = x_{t+h} - \hat{x}_t(h) = \phi_1 x_{t+h-1} + \delta + a_{t+h} - \hat{x}_t(h) =$$

$$= \phi_1^h x_t + (\phi_1^{h-1} + \phi_1^{h-2} + \ldots + \phi_1 + 1)\,\delta + a_{t+h} + \phi_1 a_{t+h-1} + \ldots +$$

$$+ \phi_1^{h-1} a_{t+1} - \hat{x}_t(h) = a_{t+h} + \phi_1 a_{t+h-1} + \ldots + \phi_1^{h-1} a_{t+1} \ .$$

Logo, $E(e_t(h)^2) = (1 + \phi_1^2 + \phi_1^4 + \ldots + \phi_1^{2h-2})\,\sigma_a^2$.

Assim, a variância do erro de previsão de um modelo AR(1) cresce para horizontes distantes.

Exemplo 03: previsão de um ARMA(1,1):

$$(1-\phi_1 B)x_t = (1-\theta_1 B)a_t + \delta$$

$$(1-\phi_1 B)x_t = (1-\theta_1 B)a_t + \delta \Leftrightarrow x_t - \phi_1 x_{t-1} = a_t - \theta_1 a_t + \delta \Leftrightarrow x_t = \phi_1 x_{t-1} + a_t - \theta_1 a_{t-1} + \delta$$

O valor de x no tempo t+h é dado por

$$x_{t+h} = \phi_1 x_{t+h-1} + a_{t+h} - \theta_1 a_{t+h-1} + \delta \ ,$$

e a previsão vale

$$\hat{x}_t(h) = \phi_1 E(x_{t+h-1}) + E(a_{t+h}) - \theta_1 E(a_{t+h-1}) + \delta \ .$$

Então, para h=1,

$$\hat{x}_t(1) = \phi_1 E(x_t) + E(a_{t+1}) - \theta_1 E(a_t) + \delta = \phi_1 x_t + 0 - \theta_1 a_t + \delta =$$

$$= \phi_1 x_t - \theta_1 a_t + \delta \ ;$$

para h=2,

$$\hat{x}_t(2) = \phi_1 E(x_{t+1}) + E(a_{t+2}) - \phi_1 E(a_{t+1}) + \delta = \phi_1 \hat{x}_t(1) + 0 - 0 + \delta =$$

$$= \phi_1 \hat{x}_t(1) + \delta = \phi_1(\phi_1 x_t - \theta_1 a_t + \delta) + \delta = \phi_1^2 x_t - \theta_1 \phi_1 a_t + \phi_1 \delta + \delta =$$

$$= \phi_1^2 x_t - \theta_1 \phi_1 a_t + (\phi_1 + 1)\delta \ ;$$

e, para $h \geq 2$,

$$\hat{x}_t(h) = \phi_1^h x_t + (\phi_1^{h-1} + \phi_1^{h-2} + \ldots + \phi_1 + 1)\delta - \phi_1^{h-1} \theta_1 a_t \ .$$

Note-se que, semelhantemente a um AR(1),

$$\lim_{h \to \infty} \hat{x}_t(h) = \frac{\delta}{1-\phi_1} = \mu \ .$$

Deve-se salientar que, tanto o AR quanto o MA e o ARMA são processos que apresentam a característica comum de suas previsões tenderem à média, quando o horizonte de previsão se torna distante. Esse fato implica uma limitação significativa no uso desses processos para a obtenção de previsões de valores distantes. A potencialidade dos métodos é elevada somente para pequenos horizontes. Para valores distantes, sua utilidade é questionável e, certamente, outros métodos que não apresentam essa limitação seriam mais adequados.

Exemplo 04: previsão de um ARIMA (1,1,0) = ARI(1,1):

$$(1-\phi_1 B)\nabla x_t = \delta + a_t$$

$$(1-\phi_1 B)(1-B)x_t = \delta + a_t$$

$$(1-(1+\phi_1)B+\phi_1 B^2)\, x_t = \delta + a_t$$

$$x_t - (1+\phi_1)\, x_{t-1} + \phi_1 x_{t-2} = \delta + a_t$$

O valor de x no período t+h é

$$x_{t+h} = (1+\phi_1)\, x_{t+h-1} - \phi_1 x_{t+h-2} + a_{t+h} + \delta \ .$$

Portanto, as previsões são:

para h=1,

$$\hat{x}_t(1) = E((1+\phi_1)x_t - \phi_1 x_{t-1} + a_{t+1} + \delta) = (1+\phi_1)\, E(x_t) - \phi_1 E(x_{t-1}) +$$

$$+ E(a_{t+1}) + \delta = (1+\phi_1)x_t - \phi_1 x_{t-1} + \delta \ ;$$

para h=2,

$$\hat{x}_t(2) = E((1+\phi_1)x_{t+1} - \phi_1 x_t + a_{t+2} + \delta) = (1+\phi_1)\, \hat{x}_t(1) - \phi_1 x_t - 0 + \delta;$$

para $h \geq 2$,

$$\hat{x}_t(h) = E((1+\phi_1)x_{t+h-1} - \phi_1 x_{t+h-2} + a_{t+h} + \delta) =$$

$$= (1+\phi_1)\, \hat{x}_t(h-1) - \phi_1 \hat{x}_t(h-2) + \delta \ .$$

Exemplo 05: previsão de um ARIMA(0,1,1) = IMA(1,1):

$$\nabla x_t = (1-\theta_1 B)\, a_t + \delta$$

$$(1-B)x_t = a_t - \theta_1 a_{t-1} + \delta$$

$$x_t - x_{t-1} = a_t - \theta_1 a_{t-1} + \delta$$

$$x_t = x_{t-1} + a_t - \theta_1 a_{t-1} + \delta$$

O valor de x no tempo t-h é

$$x_{t+h} = x_{t+h-1} + a_{t+h} - \theta_1\, a_{t+h-1} + \delta \ .$$

Portanto as previsões são:

para h=1,

$$\hat{x}_t(1) = E(x_t + a_{t+1} - \theta_1 a_t + \delta) = x_t + 0 - \theta_1 a_t + \delta = x_t - \theta_1 a_t + \delta ;$$

para h=2,

$$\hat{x}_t(2) = E(x_{t+1} + a_{t+2} - \theta_1 a_{t+1} + \delta) = \hat{x}_t(1) + 0 - 0 + \delta .$$

Logo, para $h \geq 2$,

$$\hat{x}_t(h) = \hat{x}_t(h-1) + \delta ,$$

ou seja, para qualquer horizonte, a previsão é uma linha horizontal no nível de $\hat{x}_t(h)$, caso o termo δ não esteja presente ou, se estiver, será uma reta com declividade (aclividade) definida por δ.

Exemplo 06: previsão de um ARIMA (1,1,1):

$$(1-\phi_1 B)\ \nabla x_t = (1-\theta_1 B)\ a_t + \delta$$

$$(1-\phi_1 B)\ (1-B) x_t = (1-\theta_1 B) a_t + \delta$$

$$(1-(\phi_1+1)B+\phi_1 B^2) x_t = (1-\theta_1 B) a_t + \delta$$

$$x_t - (1+\phi_1) x_{t-1} + \phi_1 x_{t-2} = a_t - \theta_1 a_{t-1} + \delta$$

$$x_t = (1+\phi_1) x_{t-1} - \phi_1 x_{t-2} + a_t - \theta_1 a_{t-1} + \delta$$

O valor de x no tempo t+h é

$$x_{t+h} = (1+\phi_1) x_{t+h-1} - \phi_1 x_{t+h-2} + a_{t+h} - \theta_1 a_{t+h-1} + \delta .$$

Portanto as previsões são:

para h=1,

$$\hat{x}_t(1) = E((1+\phi_1) x_t - \phi_1 x_{t-1} + a_{t+1} - \theta_1 a_t + \delta) =$$

$$= (1-\phi_1) x_t - \phi_1 x_{t-1} + 0 - \theta_1 a_t + \delta =$$

$$= (1+\phi_1) x_t - \phi_1 x_{t-1} - \theta_1 a_t + \delta ;$$

para h=2,

$$\hat{x}_t(2) = E((1+\phi_1)x_{t+1} - \phi_1 x_t + a_{t+2} - \theta_1 a_{t+1} + \delta) =$$

$$= (1+\phi_1)\hat{x}_t(1) - \phi_1 x_t + 0 - 0 + \delta =$$

$$= (1+\phi_1)\hat{x}_t(1) - \phi_1 x_t + \delta \ .$$

Logo, para $h \geq 2$,

$$\hat{x}_t(h) = (1+\phi_1)\hat{x}_t(h-1) - \phi_1 \hat{x}_t(h-2) + \delta \ .$$

5.3 – A Atualização das Previsões

Em geral, o processo de previsão para vários períodos de tempo pode requerer que se atualize o período de origem das previsões. Nesse caso, é conveniente a montagem de um processo onde os valores condicionados utilizados como base do cálculo da expectância de x_{t+h} possam ser atualizados de período a período, eliminando a necessidade de repetir o complexo processo de cálculo elaborado para a obtenção das previsões $\hat{x}_t(h)$.

Assim, o processo de atualização das previsões consiste em gerar instrumentos que permitam o cálculo de uma previsão $\hat{x}_{t+n}(h) = E(x_{t+h}/x_{t+n}, x_{t+n-1}, \ldots)$, a partir da previsão de $\hat{x}_t(h) = E(x_{t+h}/x_t, x_{t-1}, \ldots)$, quando são obtidos novos valores $x_{t+1}, x_{t+2}, \ldots x_{t+n}$, através de um processo iterativo, sem necessidade de recalcular o modelo.

Esse processo de atualização é denominado **previsão adaptada** e consiste em tomar o conjunto de valores realmente observados (ou mesmo suas previsões) para os períodos anteriores, juntamente com o erro de previsão do período da previsão desejada.

Ou seja, para atualizar a origem da previsão de x_{t+h+1} quando se obtém o valor x_{t+1}, deve-se calcular a previsão de x_{t+h+1} na origem t+1, através do valor de $\hat{x}_t(h+1)$ acrescido de um múltiplo do erro de previsão a_{t+1}. Para constatar a validade dessa forma de atualização, pode-se verificar que as previsões de x_{t+h+1}, feitas nas origens t+1 e t, são, respectivamente, as que seguem.

Seja um modelo ARIMA dado por (5.2), então:

$$\hat{x}_{t+1}(1) = \psi_h a_{t+1} + \psi_{h+1} a_t + \psi_{h+2} a_{t-1} + \ldots$$

$$\hat{x}_t(h+1) = \psi_{h+1} a_t + \psi_{h+2} a_{t-1} + \ldots$$

subtraindo membro a membro,

$$\hat{x}_{t+1}(h) - \hat{x}_t(h+1) = \Psi_h a_{t+1}$$

ou

$$\hat{x}_{t+1}(h) = \hat{x}_t(h+1) - \Psi_h a_{t+1} ,$$

onde o estabelecimento do valor de Ψh é obtido pela relação entre os pesos da equação (5.2), $\S(B)\ \Psi(B) = \theta(B)$.

Ou seja,

$$(1-\S_1 B-\S_2 B^2- \ldots - \S_{p+d} B^{p+d})(1+\Psi_1 B+\Psi_2 B^2+ \ldots) = (1 - \theta_1 B - \ldots - \theta_q B^q),$$

então:

$$\Psi_1 = \S_1 - \theta_1$$

$$\Psi_2 = \S_1\Psi_1 + \S_2 - \theta_2$$

$$\vdots$$

$$\Psi_j = \S_1 \Psi_{j-1} + \ldots + \S_{p+d} \Psi_{j-p-d} - \theta_j ,$$

onde $\Psi_o = 1$, $\Psi_j = 0$, $\Psi_j < 0$ e $\theta_j = 0$, $\Psi_j > q$.

Se $K = \max(p+d-1;\ q)$, então, para $j > k$,

$$\Psi_j = \S_1\Psi_{j-1} + \S_2\Psi_{j-2} + \ldots + \S_{p+d} \Psi_{j-p-d} . \qquad (5.8)$$

Note-se que os Ψ_j são calculados facilmente de forma recursiva.

Exemplo: seja a série modelada por $(1-1,8B+0,8B^2)x_t=a_t$,

onde $\S_1 = 1,8$, $\S_2 = -0,8$,

então:

$$\Psi_o = 1$$

$$\Psi_1 = 1,8$$

$$\Psi_2 = \S_1\Psi_1 + \S_2\Psi_{j-2} = 1,8(1,8)+(-0,8)1 = 2,44$$

$$\Psi_3 = \S_1\Psi_2 + \S_2\Psi_1 = 1,8(2,44)+(-0,8).1,8 = 2,95$$

$$\vdots$$

Se as previsões para um, dois e três períodos adiante, com origem "t", são $\hat{x}_t(1) = 200$, $\hat{x}_t(2) = 210$ e $\hat{x}_t(3) = 220$, então, no momento em que se obtém o valor x_{t+1}, pode-se atualizar a origem das previsões.

Seja $x_{t+1} = 201,$

então: $a_{t+1} = x_{t+1} - \hat{x}_t(1) = 201 - 200 = 1$

e $$\hat{x}_{t+1}(1) = \hat{x}_t(2) + \psi_1 a_{t+1} =$$

$$= 210 - (1{,}8).1 = 211{,}8$$

$$\hat{x}_{t+1}(2) = \hat{x}_t(3) + \psi_2 a_{t+1}$$

$$= 220-(2{,}44)(1) = 222{,}44$$

e assim sucessivamente.

5.4 – Previsões por Intervalo

Conforme (5.7), a variância do erro de previsão é dada por:

$$\text{Var}(e_t(h)) = (1 + \psi_1^2 + \ldots + \psi_{h-1}^2)\ \sigma_a^2 \ .$$

Se cada elemento do processo de ruído branco é considerado como tendo distribuição normal, toda a distribuição dos futuros erros e, conseqüentemente, as futuras observações x_{t+h} terão também distribuição normal. Ou seja,

$$e_t(h) \sim N(0, \text{Var}(e_t(h)))$$

e $$(x_{t+h}/x_t, x_{t-1}, \ldots) \sim N(\hat{x}_t(h),\ \text{Var}(e_t(h))) \ .$$

Então, a estatística

$$U = \frac{x_{t+h} - \hat{x}_t(h)}{\sqrt{\text{Var}(e_t(h))}} \sim N(0,1) \ .$$

Conseqüentemente, pode-se obter previsões por intervalo, fixado um determinado nível de confiança $(1 - \alpha)$, tal que $P(-z_c < U < z_c) = 1 - \alpha$,

ou seja, a previsão de x_{t+h} por intervalo é

$$P(\hat{x}_t(h) - z_c \sqrt{Var(e_t(h))} < x_{t+h} < \hat{x}_t(h) + z_c \sqrt{Var(e_t(h))}) = 1 - \alpha.$$

Note-se que $Var(e_t(h)) = (1+\psi_1^2 + \psi_2^2 + \ldots + \psi_{h-1}^2)\,\sigma_a^2$.

Como, na prática, os erros de previsão são valores estimados, o intervalo de previsão possível de ser obtido é.

$$P(\hat{x}_t(h) - z_c\ \hat{\sigma}_a \sqrt{\sum_{j=0}^{h-1} \psi_j^2} < x_{t+h} < \hat{x}_t(h) - z_c\ \hat{\sigma}_a \sqrt{\sum_{j=0}^{h-1} \psi_j^2}) = 1 - \alpha.^2$$

5.5 — Previsões para Séries Sazonais

O processo de previsão de um modelo SARIMA (p,d,q) x (P,D,Q) é semelhante ao processo desenvolvido para o ARIMA (p,d,q).

Desenvolve-se o modelo em uma das três formas básicas de descrição de cada valor de x_t, e aplica-se os mesmos procedimentos desenvolvidos para um modelo onde a componente sazonal não está presente.[3] Após estabelecer o valor da variável no período t+h, aplica-se o operador expectância. Erros de previsão, intervalos de confiança e atualização são tratados similarmente ao modelo ARIMA.

Exemplo: seja um modelo SARIMA $(0,1,1)\times(0,1,1)_{12}$.

Então:

$$\nabla\nabla_{12} x_t = (1 - \theta B)(1 - \Theta B^{12})a_t$$

$$(1-B)(1-B^{12})x_t = (1 - \theta B)(1 - \Theta B^{12})a_t$$

$$(1-B-B^{12} + BB^{12})x_t = (1-\theta B - \Theta B^{12} + \theta\Theta BB^{12})a_t$$

$$x_t - x_{t-1} - x_{t-12} + x_{t-13} = a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\,\Theta a_{t-13}$$

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} + a_t - \theta a_{t-1} - \Theta a_{t-1} - \Theta a_{t-12} + \theta\,\Theta a_{t-13}$$

---

[2] Observe-se que, se o erro de previsão é estimado, a estatística U deixa de ser exatamente normal. Porém, dado que o estudo de previsão com séries de tempo pressupõe um número relativamente grande de observações da variável, pode-se considerar U aproximadamente normal em todos os estudos de previsão.

[3] Para maior facilidade de operacionalização descreve-se comumente o valor x_t pela forma clássica de equações a diferenças.

Portanto o valor de x no tempo t+h vale

$$x_{t+h} = x_{t+h-1} + x_{t+h-12} - x_{t+h-13} + a_{t+h} - \theta a_{t+h-1} - \Theta a_{t+h-12} + \theta\,\Theta a_{t+h-13} \ ,$$

e a previsão $\hat{x}_t(h)$ é dada pela expectância condicional

$$\hat{x}_t(h) = E(x_{t+h}/M) = E(x_{t+h-1}/M) + E(x_{t+h-12}/M) - E(x_{t+h-13}/M) +$$

$$+ E(a_{t+h}/M) - \theta E(a_{t+h-1}/M) - \Theta E(a_{t+h-12}) + \theta\Theta E(a_{t+h-13}/M) \quad ,$$

onde a condição "M" são os valores realmente observados da variável X: x_t, x_{t-1}, ...

5.6 – Previsão com Dados Transformados

Conforme foi visto no processo de modelagem de séries não estacionárias, quando a magnitude das flutuações da variável X não se mantém constante no decorrer dos períodos de tempo, Box & Jenkins indicam a conveniência de se fazer alguma transformação de variável do tipo $Y_t = f(X_t)$, com o fito de tornar suas flutuações mais homogêneas. Nesse caso, a previsão gerada pelo modelo ARIMA transformado propicia estimativas de futuros valores transformados $\hat{y}_t(h)$.

Para obter a previsão de $\hat{x}_t(h)$ em função de $\hat{y}_t(h)$, deve-se salientar que essa pode não ser simplesmente

$$\hat{x}_t(h) = f^{-1}(\hat{y}_t(h))$$

no caso de "f" admitir inversa.

Para séries econômicas, por exemplo, a transformação geralmente viável é do tipo logarítmica natural, tal que $Y_t = \ln X_t$, onde X_t é a variável realmente observada. Para obter a previsão de $\hat{x}_t(h)$ em função de $\hat{y}_t(h)$, deve-se ressaltar que essa não é necessariamente $\hat{x}_t(h) = \exp\hat{y}_t(h)$, uma vez que a convexidade da função exponencial implica que, se a série Y_t é normal,

$$E(\exp x_{t+h}/x_t, x_{t-1}, \ldots) \neq \exp E(x_{t+h}/x_t, x_{t-1}, \ldots) \ ,$$

pois,

$$E(\exp x_{t+h}/x_t, x_{t-1}, \ldots) = \exp\left[E(x_{t+h}/x_t, x_{t-1}, \ldots) + 1/2 \mathrm{Var}(x_{t+h}/x_t, x_{t-1})\right] =$$

$$= \exp E(x_{t+h}/x_t, x_{t-1}, \ldots) \exp 1/2 \mathrm{Var}(x_{t+h}/x_t, x_{t-1}, \ldots) >$$

$$> \exp E(x_{t+h}/x_t, x_{t-1}, \ldots).$$

Isto é,

$$\hat{x}_t(h) = \exp\left[\hat{Y}_t(h) + 1/2 \ \mathrm{Var}(Y_{t+h})\right], \text{ onde } \mathrm{Var}(Y_{t+h}) \cong \mathrm{Var}(e_t(h)),$$

e $e_t(h) = Y_{t+h} - \hat{Y}_t(h)$.

Então, como a distribuição de Y_{t+h} é normal, com $E(Y_{t+h}) = \hat{y}_t(h)$ e $\mathrm{Var}(y_{t+h})$, sua previsão por intervalo será

$$P(\hat{Y}_t(h) - z_c \sqrt{\mathrm{Var}\ e_t(h)} \leq Y_{t+h} \leq \hat{Y}_t(h) + z_c \sqrt{\mathrm{Var}\ e_t(h)}) = 1 - \alpha,$$

e a previsão de x_{t+h} será dada por

$$P(\exp(\hat{Y}_t(h) - z_c \sqrt{\mathrm{Var}\ e_t(h)}) \leq X_{t+h} \leq \exp(\hat{Y}_t(h) + z_c \sqrt{\mathrm{Var}\ e_t(h)})) = 1 - \alpha.$$

6 – APLICAÇÕES PRÁTICAS DO MÉTODO DE BOX & JENKINS

6.1 – Nota Introdutória

Este capítulo constitui-se de um trabalho prático de previsão de valores futuros de determinadas séries de tempo da realidade econômica gaúcha, utilizando a metodologia proposta por Box & Jenkins.

Foram escolhidas duas variáveis: Índice de Preços ao Consumidor (custo de vida) para a Cidade de Porto Alegre e consumo de energia elétrica no Rio Grande do Sul.

A escolha dessas variáveis não se deu ao acaso. Pelo contrário, chegou-se a elas tendo em vista não só a importância que apresentam na atividade econômica rio-grandense, mas também através de consultas a técnicos ligados ao setor público estadual que as sugeriram em função do interesse prático de suas análises. De outra parte, foi fundamental para a escolha o fato de essas variáveis terem particularidades distintas umas das outras, em suas flutuações temporais, situação que permite, em conjunto, mostrar as principais características do processo de construção de modelos com o método de Box & Jenkins.

A finalidade deste capítulo é, além de oferecer um estudo detalhado de previsões de variáveis econômicas importantes, apresentar um acompanhamento prático de todas as etapas da metodologia, propiciando um entendimento melhor da teoria exposta nos capítulos anteriores e capacitando o leitor a melhores condições de manuseio e utilização da mesma.

O "pacote" computacional utilizado foi o de PACK (1977), que se encontra à disposição dos usuários no Centro de Processamento de Dados da Universidade Federal do Rio Grande do Sul[1]. Esse "pacote" é composto por duas rotinas, denominadas "Identificação" e "Estimação", estando incluídos, nessa última, os processos de cálculo das previsões e os instrumentos analíticos da etapa de checagem do diagnóstico[2].

No final do trabalho, no Anexo I, juntam-se cópias do manual do usuário, onde se encontram os comandos necessários para cada sub-rotina.

---

[1] Deve-se ressaltar, entretanto, que existe um grande número de outros "pacotes" contendo o método de Box & Jenkins, embora nenhum deles, até o momento, se encontre à disposição no mercado gaúcho. Entre eles pode-se citar: ECONOMETRIC SOFTWARE PACKAGE (1974), MEEKER (s.d.) e NIE & HULL (1980).

[2] Infelizmente, no desenvolvimento do trabalho foi constatado que o "pacote" utilizado apresenta alterações significativas em relação ao seu manual, implicando a necessidade de obtenção de algumas fases importantes do processo de forma extra "pacote".

6.2 – Previsões para o Índice Geral de Preços (Custo de Vida) para a Cidade de Porto Alegre

A série escolhida para essa variável perfaz um período de sete anos, com informações mensais desde janeiro de 1975 até dezembro de 1981, e seus valores estão expostos no Quadro 1.

Quadro 1

Índice de Preços ao Consumidor (custo de vida) em Porto Alegre — jan./75-dez./81

(base: jun./75)

MESES	1975	1976	1977	1978	1979	1980	1981
Jan.	91	117	171	243	335	587	1 235
Fev.	94	127	178	248	347	614	1 339
Mar.	97	133	185	257	365	647	1 436
Abr.	97	137	192	264	383	673	1 500
Maio	99	139	200	269	390	711	1 584
Jun.	100	141	206	277	401	763	1 642
Jul.	102	144	210	287	424	819	1 760
Ago.	106	151	213	294	445	876	1 937
Set.	109	155	217	300	480	945	2 013
Out.	112	159	220	304	496	1 024	2 080
Nov.	114	161	225	308	528	1 091	2 166
Dez.	116	165	234	317	557	1 144	2 286

FONTE: Centro de Estudos e Pesquisas Econômicas da Universidade Federal do Rio Grande do Sul (IEPE).

6.2.1 – A Etapa da Identificação

A utilização da rotina "Identificação" permite o estabelecimento das principais características do modelo.

6.2.1.1 – Avaliação preliminar

Em primeiro lugar, é conveniente plotar os valores da série observada. Seu gráfico (Figura 1) permite a visualização de uma série com tendência crescente, geométrica, sugerindo, inicialmente, a inexistência de componente sazonal significativa[3]. Por outro lado, percebe-se que a sé-

[3] No decorrer da fase de estabelecimento da ordem do modelo, essa suposição sobre a inexistência da componente sazonal deverá ser mais precisamente analisada (e sua suposição confirmada ou não) através do comportamento da função de autocorrelação.

rie se caracteriza por apresentar maior dispersão de seus valores em torno da linha de tendência, à medida que o tempo cresce, sugerindo a possibilidade de conveniência de transformação da variável para estabilizar a variância.

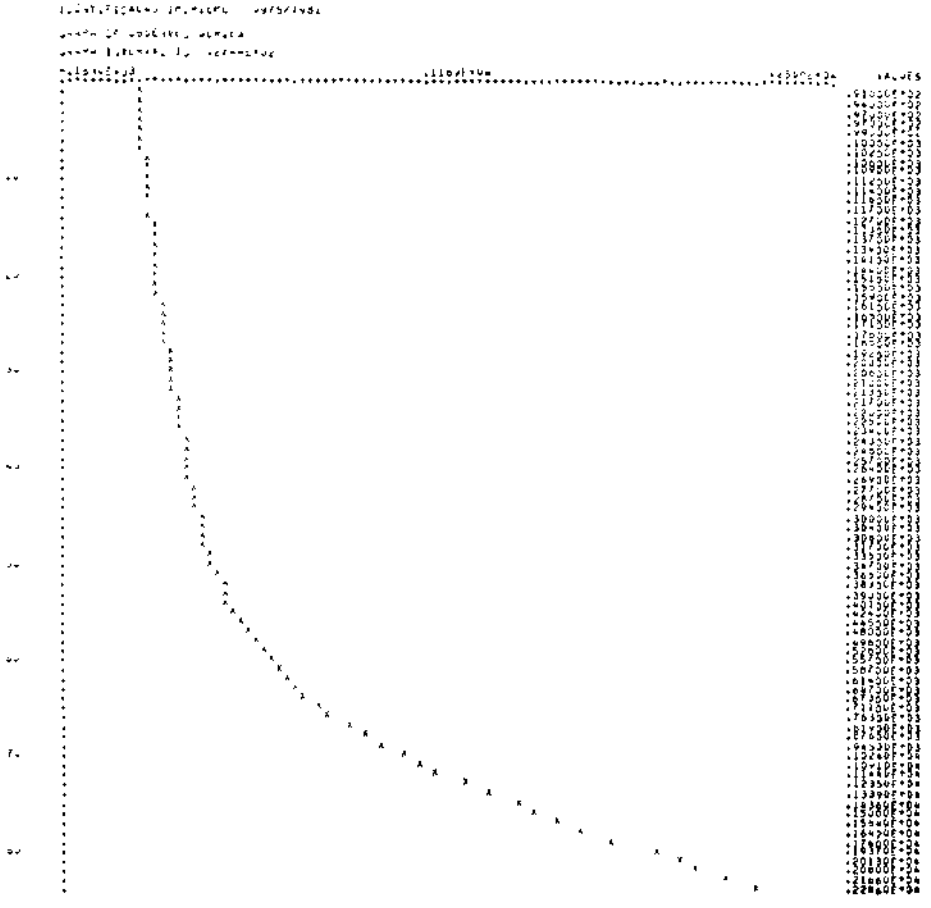


Figura 1 – Gráfico da série IPC – jan./75 - dez./81

Para melhor avaliação da existência ou não de heteroscedasticidade na série, constrói-se o Gráfico "Amplitude X Média"[4].

Para a construção desse diagrama, optou-se por dividir a seqüência de observações em segmentos de tamanho doze, uma vez que esse número, além

[4] Deve-se ressaltar que o programa utilizado não apresenta nenhuma sub-rotina para esse fim, o que implicou a necessidade de programação à parte para a geração desse diagrama.

de satisfazer à conveniência de tamanho dos segmentos[5], permite a obtenção de segmentos de mesmo tamanho.

O Gráfico "Amplitude X Média", assim construído, está apresentado na Figura 2. Pode-se perceber, através dele, uma tendência crescente da amplitude em relação à média dos segmentos, o que sugere a possibilidade da transformação logarítmica natural, $Y_t = \ln X_t$, onde X_t representa o Índice Geral de Preços da Cidade de Porto Alegre no período de janeiro de 1975 a dezembro de 1981.

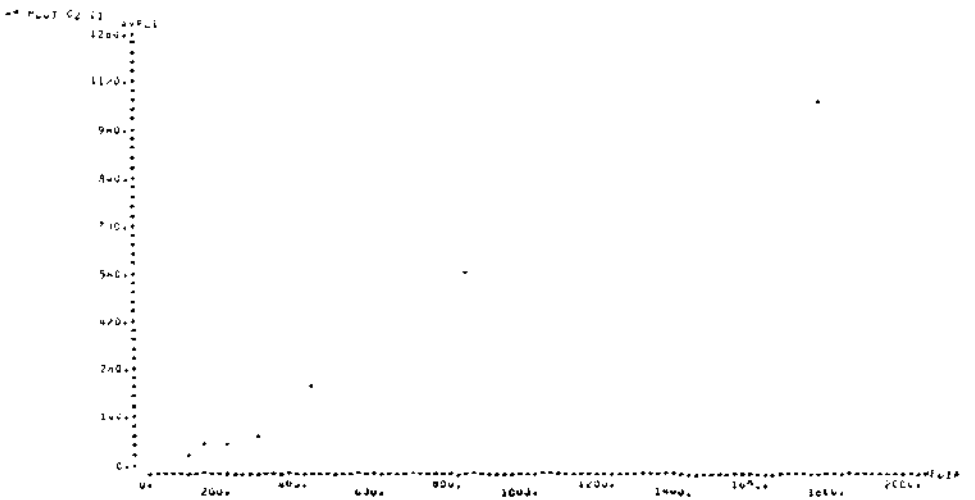


Figura 2 — Gráfico "Amplitude x Média" para a série Índice de Preços ao Consumidor — jan./75 - out./81

A análise do gráfico da série de dados logaritmados, exposto na Figura 3, permite verificar que a dispersão dos valores em torno da curva de tendência é menor, o que a torna mais conveniente para fins de previsão. Em vista disso, a série escolhida para o trabalho é a logarítmica natural.

6.2.1.2 – O estabelecimento da ordem do modelo

Obtenção do valor da ordem "d"

A análise da Figura 3 permite verificar a existência de uma série com tendência geometricamente crescente. Significa dizer que a série é não estacionária, ($d \neq 0$), e torna-se necessário diferenciá-la para se obter uma série estacionária. A grosso modo, como a tendência é geome-

[5] A experiência aconselha, conforme o exposto nos capítulos anteriores, segmentos de tamanho entre oito e 12 observações para séries sazonais e de tamanho igual ao período do ciclo sazonal para séries que contenham essa componente.

tricamente crescente, pode-se inferir que é bastante provável a obtenção de série estacionária com a aplicação do operador diferença duas vezes.

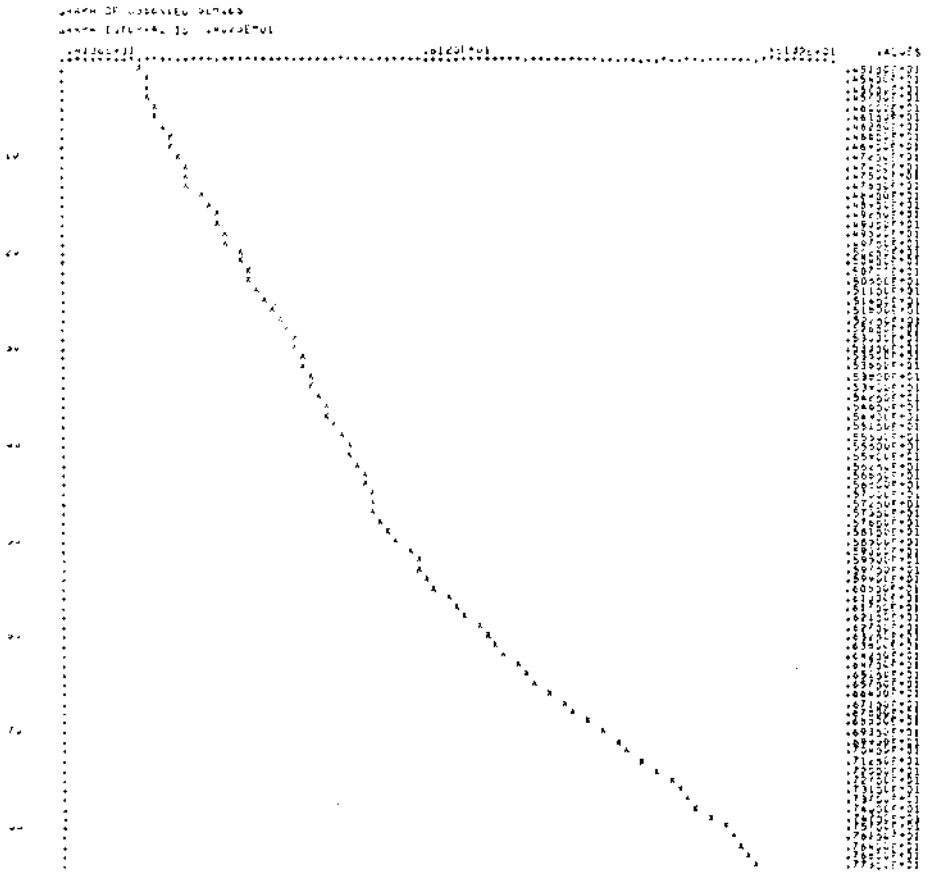


Figura 3 – Gráfico da série $Y_t = \ln X_t$, onde X_t é a série observada do IPC

Para o estabelecimento mais preciso da ordem "d", o rigor analítico sugere o estudo da função de autocorrelação, a qual, para simplificação, será chamada de ACF. Para tanto, buscou-se analisar a ACF das séries Y_t, ∇Y_t, $\nabla^2 Y_t$ e $\nabla^3 Y_t$[6]. Os gráficos da ACF para essas quatro séries estão expostos, respectivamente, nas Figuras 4, 5, 6 e 7, e seus valores numéricos, nos Quadros 2 e 3.

[6] Embora nem sempre esteja explícito, deve-se ter presente que $Y_t = \ln X_t$.

Pode-se observar, na Figura 4, que a ACF da série Y_t não apresenta nem uma queda rápida para zero, nem flutuações cíclicas em seus valores, revelando a correção da suposição inicial de inexistência da componente sazonal e da não estacionariedade tendencial da série, quando da avaliação da Figura 3.

Figura 4 — Gráfico da função de autocorrelação da série Y_t

A ACF da série ∇Y_t (Figura 5) também não apresenta uma queda rápida para zero. A ACF da série $\nabla^2 Y_t$ (Figura 22), ao contrário das anteriores, apresenta uma queda rápida para zero. Deixando de considerar o valor da defasagem seis, significativamente diferente de zero[7], somente o valor da ACF de defasagem um é diferente de zero, de acordo com o teste de significância de 2 $DP(r_k)$, $k > 0$.

[7] O que é razoável de se supor, uma vez que esse valor se mostra distinto de seus próximos, incoerente com a teoria que mostra a diminuição dos r_k, à medida que k cresce, permitindo supor que esse fato é devido às características do diagrama ser obtido através de estatísticas amostrais.

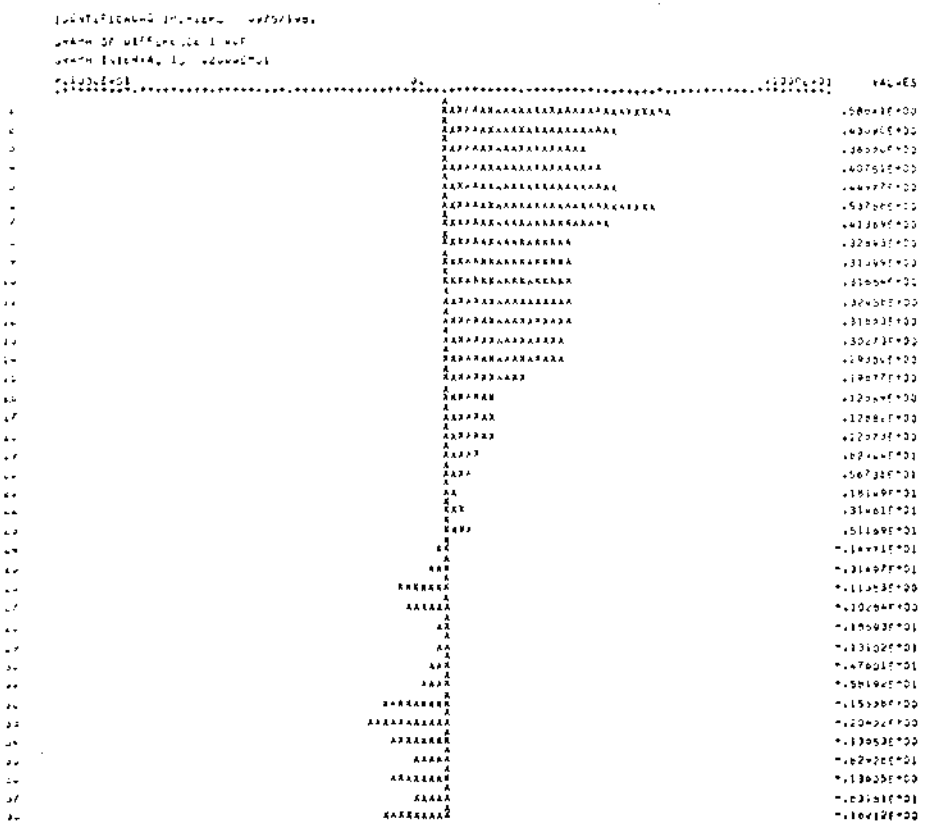


Figura 5 – Gráfico da função de autocorrelação da série ∇Y_t

A ACF da série $\nabla^3 Y_t$ apresenta também uma queda rápida para zero, porém com maior flutuação em seus valores. De acordo com o critério de ANDERSON (1979), a $Var(\nabla^2 Y_t) < Var(\nabla Y_t)$ e a $Var(\nabla^3 Y_t) > Var(\nabla^2 Y_t)$, o que torna preferível a série $\nabla^2 Y_t$. Portanto, pode-se concluir que d = 2.

Obtenção dos valores da ordem (p, q)

Para obter a ordem (p,q), a análise é efetuada através das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) da série estacionária $\nabla^2 Y_t$.

Pode-se perceber, pela Figura 6, que a ACF da série $\nabla^2 Y_t$ apresenta a primeira defasagem significativamente diferente de zero (o valor crítico, a partir do qual é traçada a "linha de controle", é $2\,DP(r_k), k > 0$), embora seu valor seja bastante pequeno, voltando a apresentar valor significativamente diferente de zero somente para a defasagem seis.

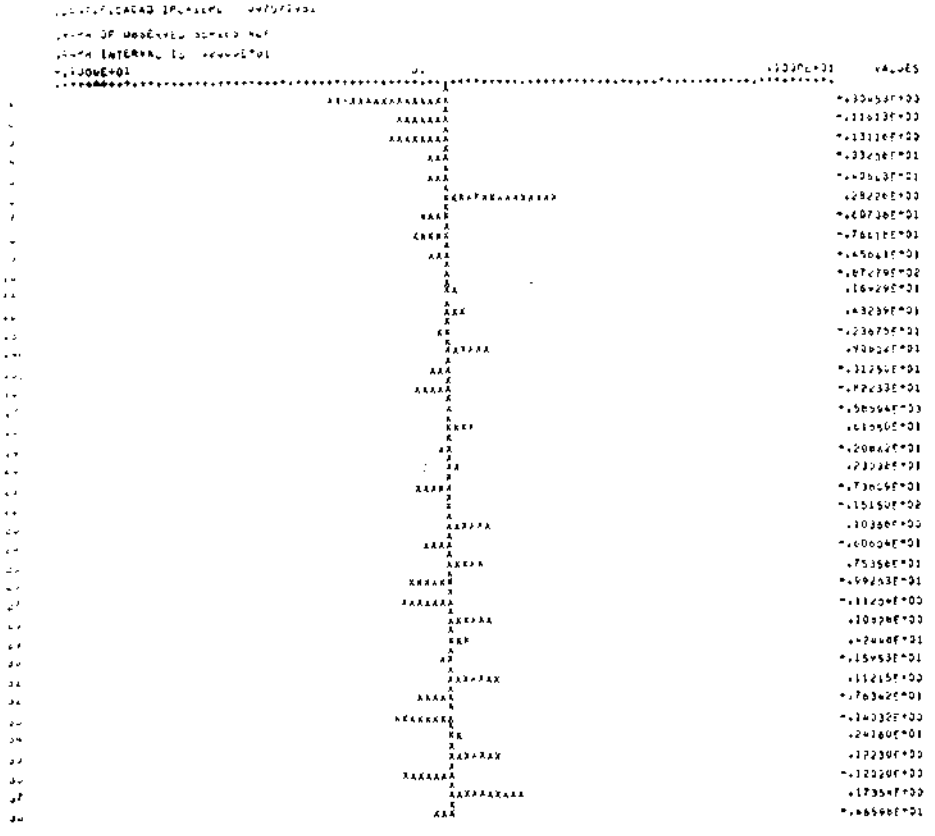


Figura 6 — Gráfico da função de autocorrelação da série $\nabla^2 Y_t$

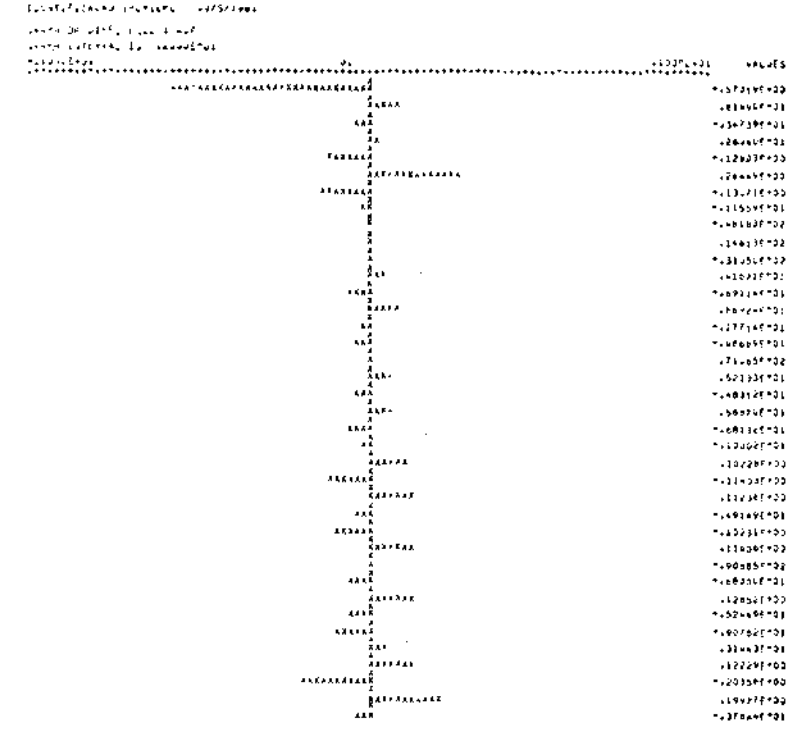


Figura 7 — Gráfico da função de autocorrelação da série $\nabla^3 Y_t$

Quadro 2

Valores numéricos da ACF para as séries Y_t e ∇Y_t

Quadro 3

Valores numéricos da ACF para as séries $\nabla^2 Y_t$ e $\nabla^3 Y_t$

```
AUTOCORRELATION FUNCTION

DATA = IDENTIFICACAO IPC-IEPE   0975/1981
DIFFERENCING = ORIGINAL SERIES IS NOW DATA DIFFERENCED BY
               2) 1 OF ORDER  1   3) 1 OF ORDER  1
               DIFFERENCES BELOW ARE OF ORDER   1
YOUR DATA TRANSFORMED DATA = LOG(Z(T)+0.          )


ORIGINAL SERIES
MEAN OF THE SERIES = .26E03E-03
ST. DEV. OF SERIES = .13699E-01
NUMBER OF OBSERVATIONS =    82

  1- 12    -0.30   -0.12   -0.13   -0.03   -0.09    0.08   -0.06   -0.00   -0.05   -0.01    0.02    0.04
 ST.E.      0.11    0.12    0.12    0.12    0.12    0.12    0.13    0.13    0.13    0.13    0.13    0.13

 13- 24    -0.06    0.09   -0.04   -0.00   -0.00    0.06   -0.02    0.02   -0.07   -0.00    0.10   -0.06
 ST.E.      0.13    0.13    0.13    0.13    0.13    0.13    0.13    0.13    0.13    0.13    0.13    0.14

 25- 36     0.06   -0.16   -0.11    0.11    0.04   -0.02    0.11   -0.08   -0.14    0.02    0.12   -0.12
 ST.E.      0.14    0.14    0.14    0.14    0.14    0.14    0.14    0.14    0.14    0.14    0.14    0.14

 37- 38     0.17   -0.05
 ST.E.      0.15    0.15

MEAN DIVIDED BY ST. ERROR =  .12069E+00


DIFFERENCE 1
MEAN OF THE SERIES = .17910E-03
ST. DEV. OF SERIES = .30534E-01
NUMBER OF OBSERVATIONS =    81

  1- 12    -0.57    0.00   -0.04    0.03   -0.14    0.08   -0.13   -0.01   -0.00    0.00   -0.00    0.04
 ST.E.      0.11    0.14    0.14    0.14    0.14    0.15    0.15    0.15    0.15    0.15    0.15    0.15

 13- 24    -0.07    0.05   -0.03   -0.05    0.05    0.05   -0.03    0.08   -0.07    0.04    0.13   -0.11
 ST.E.      0.15    0.15    0.15    0.15    0.15    0.15    0.15    0.15    0.15    0.15    0.15    0.15

 25- 36     0.11   -0.02   -0.10    0.12   -0.01   -0.07    0.13   -0.02   -0.09    0.03    0.16   -0.12
 ST.E.      0.16    0.16    0.16    0.16    0.16    0.16    0.16    0.16    0.16    0.16    0.16    0.16

 37- 38     0.20   -0.04
 ST.E.      0.17    0.17

MEAN DIVIDED BY ST. ERROR =  .52807E-01
```

A Figura 8 mostra que a PACF dessa série não apresenta uma queda rápida para zero (o valor crítico, a partir do qual é traçada a "linha de controle", nesse caso, é $2DP(\hat{\phi}_{kk})$. Pelo contrário, as cinco primeiras defasagens mantêm-se relativamente estáveis e significativamente diferentes de zero e, só a partir de então, a PACF apresenta queda rápida para zero.

A análise conjunta das ACF e PACF permite estabelecer algumas hipóteses de modelos:

- considerando que a ACF "morre" a partir da primeira defasagem (isto é, não se considera a defasagem seis significativa) e que a PACF não apresenta queda rápida para zero, a partir das primeiras defasagens, tem-se um modelo ARIMA (0,2,1);
- as oscilações e a defasagem seis, significativamente distinta de zero no correlograma da ACF (Figura 6), podem estar indicando a existência de uma parte auto-regressiva na série. Assim, hipóteses possíveis de modelos seriam aquelas que também contivessem parcelas auto-regressivas. O modelo ARIMA (1,2,1), nesse caso, não pode ser descartado;
- como o valor da defasagem um da ACF é pequeno, o modelo poderá apresentar o parâmetro média móvel não significativo. Nesse caso, só permaneceria a parte auto-regressiva. Dado o fato de as oscilações se apresentarem, em alguns momentos, com periodicidade superior a uma defasagem, é possível que haja necessidade de ordem superior a um na parcela auto-regressiva. Assim, deve-se levar em consideração modelos tais como ARIMA (1,2,0) e ARIMA (2,2,0).

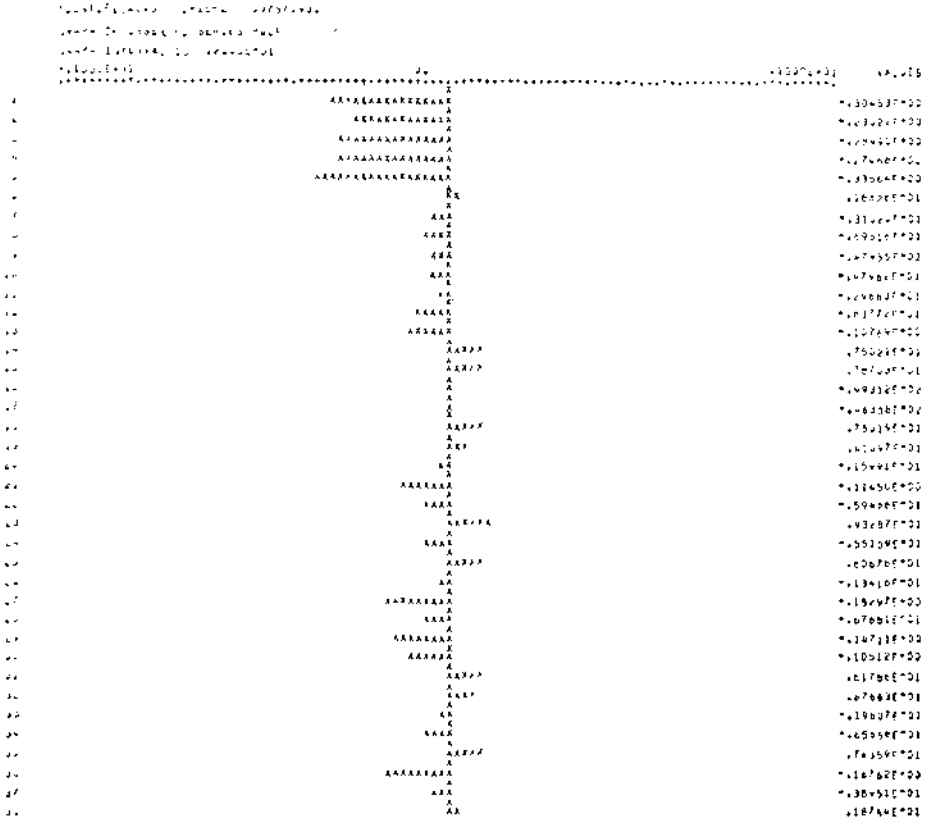


Figura 8 — Gráfico da função de autocorrelação parcial da série $\nabla^2 Y_t$

Quadro 4

Valores numéricos da PACF para a série $\nabla^2 Y_t$

Resumindo, o processo de identificação indicou a possibilidade dos seguintes modelos, todos com a variável transformada $Y_t = \ln X_t$.

Modelo suposto nº 1: ARIMA (0,2,1)
Modelo suposto nº 2: ARIMA (1,2,1)
Modelo suposto nº 3: ARIMA (1,2,0)
Modelo suposto nº 4: ARIMA (2,2,0)

6.2.1.3 – O estabelecimento das estimativas iniciais dos parâmetros

Para completar o processo de identificação, devem-se estabelecer as estimativas iniciais dos parâmetros dos modelos supostos, fase onde também é feita a avaliação da existência ou não da componente tendencial determinística.

O "pacote" computacional utilizado não apresenta sub-rotinas para esses fins, e torna-se necessário obter esses resultados manualmente, o que é feito a seguir[8].

Primeiro modelo suposto: ARIMA (0, 2, 1)

Primeiramente, deve-se verificar a existência ou não da componente tendencial determinística, ou seja, deve-se avaliar se θ_o é significativamente diferente de zero ou não. Para tanto, testa-se se $\overline{\nabla^2 Y_t}$, em módulo, é maior do que DP $(\nabla^2 Y_t)$.

Para esse modelo,

$$\overline{\nabla^2 Y_t} = 0{,}00026203$$

$$\text{e } DP(\overline{\nabla^2 Y_t}) = \sqrt{\frac{Co(1+2r_1)}{n}}, \text{ com } n = N - d$$

$$= \sqrt{\frac{0{,}0003496(1+2(-0{,}3))}{82}}$$

$$= 0{,}001306$$

Como $\overline{\nabla^2 Y_t} < DP(\overline{\nabla^2 Y_t})$, a média pode ser considerada igual a zero e o modelo não contém θ_o.

O modelo suposto é $\nabla^2 \ln X_t = (1 - \theta_1 B)a_t$ e a estimativa inicial do parâmetro θ_1 é tal que $r_1 = \frac{-\theta_1}{1+\theta_1^2}$, $-1 < \theta_1 < 1$.

$$\text{Logo, } -\theta_1 = r_1(1+\theta_1^2) \Rightarrow \begin{cases} \theta_1 = 3 \text{ (incompatível)} \\ \theta_1 = 0{,}33 \end{cases}$$

[8] Essa deficiência é relativa no que tange à obtenção das estimativas iniciais dos parâmetros, uma vez que o processo de aproximações sucessivas de MARQUARDT é bastante potente. Mesmo que sejam dados valores aleatórios, dentro das regiões admissíveis, para estimativas iniciais dos parâmetros, a convergência para as estimativas eficientes é rápida. No item seguinte, quando do processo de estimação dos parâmetros dos modelos, voltar-se-á a abordar, com melhores condições, esse assunto.

Portanto o modelo suposto inicialmente é:

$$\nabla^2 \ln X_t = (1 - 0{,}33B)a_t = a_t - 0{,}33a_{t-1}$$

Segundo modelo suposto: ARIMA (1, 2, 1)

Para esse modelo,

$$DP(\overline{\nabla^2 Y_t}) = \sqrt{\frac{Co}{n}\left(1 - \frac{2r_1^2}{r_1 - r_2}\right)}, \quad n = N - d.$$

$$= \sqrt{\frac{0{,}0003496}{82}\left(1 - \frac{2(-0{,}3)^2}{-0{,}3 - (-0{,}12)}\right)}$$

$$= 0{,}00289827$$

Como $\overline{\nabla^2 Y_t} = 0{,}00026203 < DP(\overline{\nabla^2 Y_t}) = 0{,}00289827$, a média pode ser considerada igual a zero e o modelo não contém θ_0.

O modelo suposto é $(1-\phi_1 B)\nabla^2 \ln X_t = (1-\theta_1 B)a_t$ e as estimativas iniciais dos parâmetros ϕ_1 e θ_1 são tais que $r_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}$ e $r_2 = r_1\phi_1$, $-1<\phi_1<1$ e $-1<\theta_1<1$.

Logo, $\phi_1 = 0{,}4$

$\theta_1 = 1{,}0$

Portanto o segundo modelo suposto inicialmente é:

$(1-0{,}4B)\nabla^2 \ln X_t = (1-B)a_t = a_t - a_{t-1}$ [9]

Terceiro modelo suposto: ARIMA (1, 2, 0)

Para esse modelo,

$$DP(\overline{\nabla^2 Y_t}) = \sqrt{\frac{Co(1+r_1)}{n(1-r_1)}}, \quad n = N - d = \sqrt{\frac{0{,}0003496(1-0{,}3)}{82(1+0{,}3)}} = \sqrt{\frac{0{,}0002447}{106{,}6}} =$$

$$= 0{,}00148323$$

Como $\overline{\nabla^2 Y_t} = 0{,}00026203 < DP(\overline{\nabla^2 Y_t}) = 0{,}00148323$, a média não é significativamente diferente de zero e o modelo não contém θ_0.

O modelo suposto é $(1-\phi_1 B)\nabla^2 \ln X_t = a_t$ e a estimativa inicial de ϕ_1 é r_1.

Logo $\phi_1 = -0{,}3$

Portanto o terceiro modelo suposto inicialmente é:

$(1+0{,}3B)\nabla^2 \ln X_t = a_t$.

---

[9] Deve-se chamar a atenção para o fato de θ_2 inicial ser igual a 1, o que não satisfaz a região de admissibilidade. Isso ocorre em conseqüência de r_1 e r_2 serem tomados com valores aproximados a uma casa decimal e não irá causar dificuldades na construção do modelo, uma vez que a estimativa eficiente desse parâmetro se deverá encontrar dentro da região de admissibilidade.

Quarto modelo suposto: ARIMA (2, 2, 0)

Para esse caso,

$$DP(\overline{\nabla^2 Y_t}) = \sqrt{\frac{Co(1+r_1)(1-2r_1^2+r_2)}{n(1-r_1)(1+r_2)}} = \sqrt{\frac{0,0003496(0,7)(1-0,18-0,12)}{82(1,3)(1,12)}}$$

$$= \sqrt{0,0000014} = 0,0011832$$

Como $\overline{\nabla^2 Y_t} < DP(\overline{\nabla^2 Y_t})$, a média pode ser considerada igual a zero e o modelo não contém θ_o.

O modelo suposto é $(1-\phi_1 B-\phi_2 B^2)\nabla^2 \ln X_t = a_t$ e as estimativas iniciais de ϕ_1 e ϕ_2 são

$$\phi_1 = \frac{r_1(1-r_2)}{1-r_1^2} \quad \text{e} \quad \phi_2 = \frac{r_2-r_1^2}{1-r_1^2}$$

Logo,

$$\phi_1 = 0,3692$$

$$\phi_2 = -0,2308$$

Portanto o quarto modelo suposto inicialmente é

$$(1-0,3692B+0,2308B^2)\nabla^2 \ln X_t = a_t.$$

Com a obtenção dos valores iniciais dos parâmetros dos modelos supostos, está encerrada a etapa de identificação.

6.2.2 – A Etapa da Estimação

Após a etapa de identificação de um ou mais modelos para descrever a série temporal e o estabelecimento das estimativas preliminares de seus parâmetros, pode-se trabalhar com a outra rotina do "pacote".

A rotina "estimação", fazendo uso das técnicas do algoritmo de Marquardt, gera um processo iterativo de aproximações sucessivas, a partir das estimativas preliminares obtidas na etapa anterior. Essa rotina proporcionará as estimativas eficientes[10] dos parâmetros de cada modelo suposto, o restante dos instrumentos estatísticos das demais etapas do processo de construção de modelos e a fase final de previsão de valores futuros da variável.

Nas páginas seguintes, estão expostos os sumários das saídas do computador para a etapa da estimação, para os modelos supostos. Após, no Quadro 5, estão resumidas as principais informações para a avaliação desses modelos.

Deve-se ressaltar que o "pacote" não oferece, diretamente, valores dos desvios padrões dos estimadores e a estatística $Q = n\Sigma r_k^2$, os quais tiveram de ser calculados à parte, com base nas informações que o programa apresenta no sumário do relatório de saída.

[10] O termo "estimadores eficientes" está aqui empregado no sentido de especificar aqueles parâmetros que proporcionam um $S(\phi, \theta)$ mínimo.

Quadro 5

Principais informações estatísticas dos modelos supostos para a série Índice de Preços ao Consumidor para a cidade de Porto Alegre

DISCRIMINAÇÃO	ESTIMATIVAS DOS PARÂMETROS (1)	DESVIO PADRÃO DOS ESTIMADORES	σ_a^2	Q (2)	NÚMERO DE ITERAÇÕES A PARTIR DAS ESTIMATIVAS INICIAIS	NÚMERO DE ITERAÇÕES A PARTIR DE ZERO
ARIMA(1,2,0) sem θ_0 $(1-\phi_1 B)\nabla^2 \ln X_t = a_t$	ϕ_1 =-0,30632 (S)	DP(ϕ_1)=0,108301	0,00032103	25,129	2	2
ARIMA(0,2,1) sem θ_0 $\nabla^2 \ln X_t = (1-\theta_1 B)a_t$	θ_1 = 0,79113 (S)	DP(θ_1)=0,0697	0,00026259	21,338	5	5
ARIMA(1,2,1) sem θ_0 $(1-\phi_1 B)\nabla^2 \ln X_t = (1-\theta_1 B)a_t$	ϕ_1 = 0,24358 θ_1 = 0,85552 (S)	DP(ϕ_1)=0,14255 DP(θ_1)=0,07521	0,00025929	13,861	5	4
ARIMA(2,2,0) sem θ_0 $(1-\phi_1 B-\phi_2 B^2)\nabla^2 \ln X_t = a_t$	ϕ_1 =-0,38239 (S) ϕ_2 =-0,24489 (S)	DP(ϕ_1)=0,11088 DP(ϕ_2)=0,10932	0,00030659	26,738	2	2
ARIMA(1,2,0) com θ_0 $(1-\phi_1 B)\nabla^2 \ln X_t = \theta_0 + a_t$	ϕ_1 =-0,32118 (S) θ_0 = 0,00028055	DP(ϕ_1)=0,1082142 DP(θ_0)=0,0021515	0,00036777	25,357	-	3
ARIMA(0,2,1) com θ_0 $\nabla^2 \ln X_t = \theta_0 + (1-\theta_1 B)a_t$	θ_1 = 0,83557 (S) θ_0 = 0,00034417	DP(θ_1)=0,0634948 DP(θ_0)=0,0003082	0,00028753	18,165	-	7
ARIMA(1,2,1) com θ_0 $(1-\phi_1 B)\nabla^2 \ln X_t = \theta_0 + (1-\theta_1 B)a_t$	ϕ_1 = 0,25165 θ_1 = 0,91993 (S) θ_0 = 0,00031330	DP(ϕ_1)=0,1264285 DP(θ_1)=0,0475102 DP(θ_0)=0,0001797	0,00028984	12,611	-	1
ARIMA(2,2,0) com θ_0 $(1-\phi_1 B-\phi_2 B^2)\nabla^2 \ln X_t = \theta_0 + a_t$	ϕ_1 =-0,51789 (S) ϕ_2 =-0,42455 (S) θ_0 = 0,00054322	DP(ϕ_1)=0,1038520 DP(ϕ_2)=0,1098061 DP(θ_0)=0,0020438	0,00031665	24,215	-	4
ARIMA(0,2,2) com θ_0 $\nabla^2 \ln X_t = (1-\theta_1 B-\theta_2 B^2)a_t + \theta_0$	θ_1 = 0,77134 (S) θ_2 = 0,17711 θ_0 = 0,00033700	DP(θ_1)=0,0927142 DP(θ_2)=0,1401326 DP(θ_0)=0,0003741	0,00028655	14,323	-	7
ARIMA(0,2,2) sem θ_0	θ_1 = 0,60013 (S) θ_2 = 0,20872	DP(θ_1)=0,11301 DP(θ_2)=0,113194	0,00025342	14,406	-	4

(1) O sinal (S) na frente do estimador informa que esse é significante, de acordo com o critério de compará-lo com 2 vezes seu desvio padrão. (2) Para todos os modelos, $Q = n \sum_{k=1}^{K} r_k^2$.

O Quadro 5 permite algumas considerações importantes para a escolha do modelo mais adequado ao estudo.

- Na etapa de identificação, concluiu-se pela não inclusão do termo constante. Porém, para fins exemplificativos, optou-se por estimar todos os modelos supostos, considerando também a existência de θ_0. Percebe-se que essas inclusões não alteraram substancialmente as estimativas dos demais parâmetros em relação aos modelos supostos sem o termo constante. Para todos os modelos, θ_0 é bastante pequeno e se apresenta não significativamente distinto de zero. A melhora do ajustamento é desprezível, conforme se constata pelo ínfimo acréscimo da variância residual, comprovando a correção da decisão de não incluir θ_0 e a potencialidade do teste de significância feito na etapa inicial de identificação.

- Assim como foram estimados, para exemplificar, os modelos com o termo constante, optou-se em estimar todos os cinco modelos mais parcimoniosos capazes de descrever a série observada, embora, de antemão, se saiba que alguns não satisfazem às características para um bom ajustamento àqueles valores.

- As estimativas eficientes dos parâmetros dos modelos supostos foram obtidas não só a partir de valores iniciais calculados na etapa de identificação, como também supondo estimativas preliminares iguais a zero. O número de iterações necessárias para se obter estimativas eficientes foi bastante semelhante em ambos os casos, revelando a capacidade de rápida convergência àqueles valores que o processo de aproximações sucessivas do algoritmo de Marquardt possui, independentemente dos valores iniciais considerados. Por essa razão, torna-se desnecessário preocupar-se com a obtenção de boas estimativas preliminares.

- O modelo suposto ARIMA(1,2,1) apresenta ϕ_1 não significativo, o que induz à opção por um ARIMA(0,2,1). Da mesma forma, o modelo suposto ARIMA(0,2,2) apresenta o parâmetro de ordem dois não significante. Ao reduzir a ordem desse modelo, recai-se novamente em um ARIMA(0,2,1). Assim, pode-se reduzir os cinco modelos analisados a apenas três: ARIMA(0,2,1), ARIMA(1,2,0) e ARIMA(2,2,0).

- Os modelos supostos ARIMA(0,2,1), ARIMA(1,2,0) e ARIMA(2,2,0) apresentam seus parâmetros significativos. A estatística "Q", para todos os três, apresenta-se bastante inferior ao valor crítico (para 24 graus de liberdade, $\chi^2_{0,10}=36,4$ e para 23 graus de liberdade, $\chi^2_{0,10}=35,2$), o que indica que os resíduos são não significativamente não-brancos, atendendo às exigências teóricas do método.

- A escolha do melhor modelo, entre o ARIMA(0,2,1) e o ARIMA(1,2,0) e o ARIMA(2,2,0), depende do critério utilizado. Para fins de previsão, é usual o critério de menor erro quadrático médio. Para fins de ajustamento do modelo estimado aos dados observados, pode-se adotar o critério da variância residual mínima. Como, para o cálculo do erro quadrático médio, se necessita dos valores realmente observados, x_{85}, x_{86},..., x_{96}, e esses não são disponíveis, optou-se por adotar o critério da variância residual mínima. Por esse critério, o modelo escolhido é o ARIMA (0,2,1).

6.2.3 – A Etapa da Checagem do Diagnóstico

A etapa de verificação da correção ou não da escolha do modelo, efetuada no item anterior, consiste em avaliar se os resíduos daquele modelo formam um processo de ruído branco. Para tanto, um dos testes mais comumente utilizados e sugerido por Box & Jenkins é o "portmanteau", através da estatística "Q", além da análise gráfica das ACF e PACF.

Conforme foi visto no item anterior, a estatística "Q", para o modelo escolhido, apresenta-se inferior ao valor crítico da tabela qui-quadrado, satisfazendo, portanto, às exigências do teste.

A análise das ACF e PACF revela, conforme pode-se verificar nas Figuras 9 e 10, poucos coeficientes dessas funções com valores superiores à linha de controle. As funções ACF e PACF não apresentam comportamento que revele má identificação do modelo. O gráfico da série de resíduos (Figura 11) não mostra nenhum comportamento sistemático. Os resíduos podem ser considerados aleatórios.

Portanto o modelo escolhido nas etapas anteriores atende às exigências da teoria. Está encerrada a fase de construção. O modelo é o ARIMA (0,2,1):

$$\nabla^2 \ln X_t = (1 - 0{,}79113B)a_t = a_t - 0{,}79113a_{t-1} .$$

6.2.4 – Previsões

A fase final do processo é o cálculo dos valores previstos para um horizonte previamente estabelecido.

Essa é a fase que requer menores considerações, uma vez que o trabalho se constitui somente de cálculos, os quais são totalmente estabelecidos pelo computador.

Optou-se em obter valores previstos para 12 períodos adiante, a partir da observação 84 (dezembro de 1981). Nas páginas seguintes, apresentam-se os resultados computacionais.

A seguir, no Quadro 7, apresenta-se a previsão do Índice de Preços ao Consumidor para o ano de 1982.[11]

Para demonstrar as potencialidades do método, foram calculadas também as previsões para os demais modelos supostos, cujos resultados estão expostos no Anexo II. Deve-se reparar que as diferenças dos valores previstos entre os diversos modelos supostos se apresentam bastante pequenas.

[11] Repare-se que, como a série $\nabla^2 \ln X_t$ é descrita por um MA(1), a partir de h = 1, sua previsão estabiliza-se em μ. Vale dizer que a previsão de X_t apresenta, a partir de h = 1, taxa constante de crescimento. No caso, 5,4%.

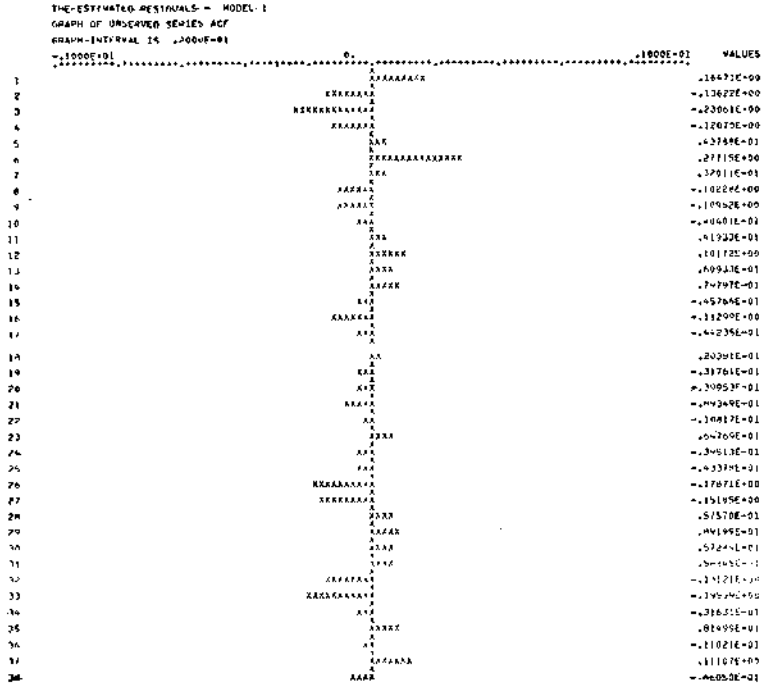


Figura 9 — Gráfico da função de autocorrelação da série de resíduos para o modelo $\nabla^2 \ln X_t = (1 - 0{,}79113B)a_t$

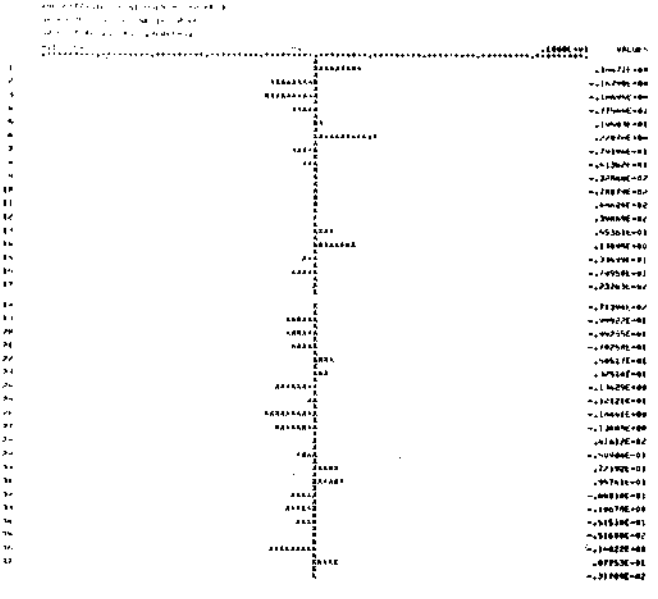


Figura 10 — Gráfico da PACF da série de resíduos para o modelo $\nabla^2 \ln X_t = (1 - 0{,}79113B)a_t$

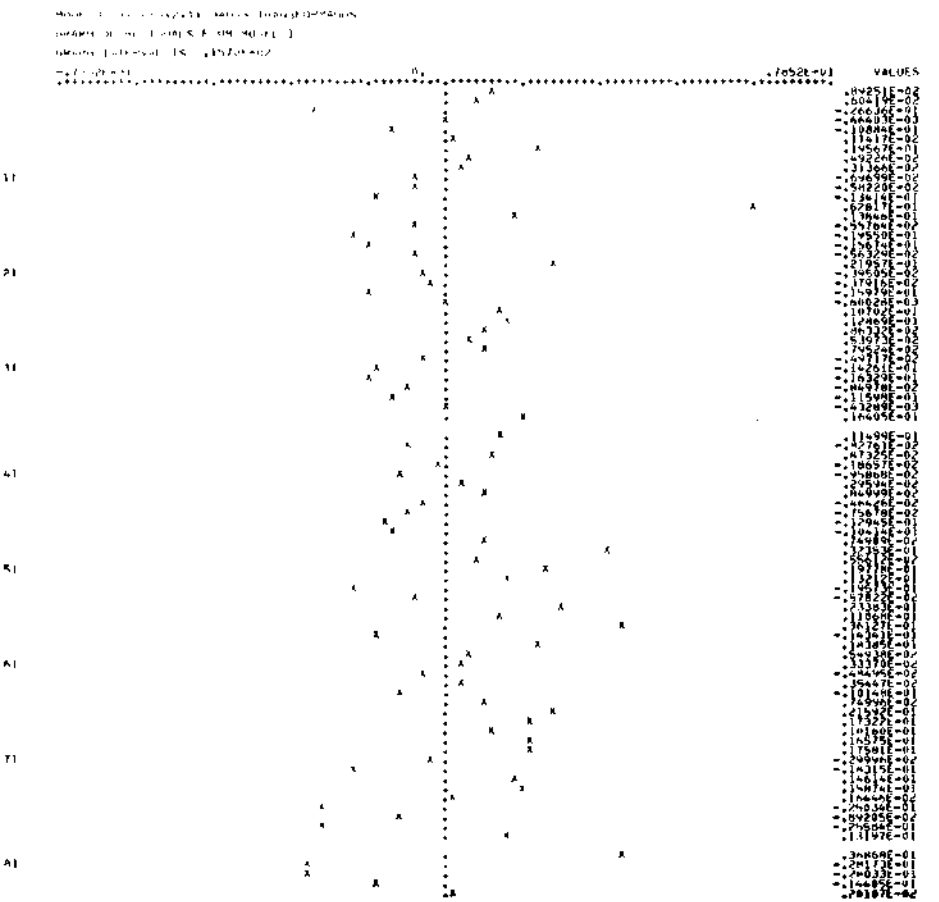


Figura 11 — Gráfico da série de resíduos para o modelo ∇^2 1n X_t = (1 − 0,79113B) a_t

Quadro 6

Relatório de saída do computador: previsão para os valores mensais do IPC — 1982

```
TIME SERIES FORECASTING FOR MODEL  1
*****************************************************************************
DATA =   Z = MODELO ARIMA(0,2,1) DADOS TRANSFORMADOS
DIFFERENCING ON   Z = 1) 1 OF ORDER  1  2) 1 OF ORDER  1
MODEL DEVELOPED WITH TRANSFORMED DATA = LOG(  Z(T)+0.          )
*****************************************************************************
UNIVARIATE MODEL PARAMETERS
*****************************************************************************
PARAMETER          PARAMETER          PARAMETER        BEGINNING
 NUMBER              TYPE              ORDER            VALUE
*****************************************************************************
    1             MOVING AVERAGE 1        1            .79113E+00
*****************************************************************************
NUMBER OF TIME ORIGINS FOR FORECASTS =        1
NUMBER OF FORECASTS AT EACH TIME ORIGIN =    12
FORECAST TIME ORIGINS ARE T=                 84
```

Quadro 6

Relatório de saída do computador: previsão para os valores mensais do IPC — 1982

```
REGULAR FORECAST RESULTS IN TERMS OF  TRANSFORMED DATA
**********************************************************

MODEL  1 FORECASTS AT BASE PERIOD  84 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD      LO. CONF. LIMIT          FORECAST          UP. CONF. LIMIT      ACTUAL, IF KNOWN

        1          .7755d300E+01          .7756d889E+01          .7761d479E+01
        2          .7739961E+01           .7763v220E+01          .7788d779E+01
        3          .7824734E+01           .7871b531E+01          .7955d487E+01
        4          .7859068E+01           .7943b81E+01           .8032d155E+01
        5          .7893v20E+01           .7996d212E+01          .8099d504E+01
        6          .7927d71E+01           .8049d543E+01          .8169d914E+01
        7          .7960d16E+01           .8103b873E+01          .8241d30E+01
        8          .7992739E+01           .8157d204E+01          .8321d069E+01
        9          .8024d406E+01          .8210d534E+01          .8388d829E+01
       10          .8055v027E+01          .8257b805E+01          .8468d703E+01
       11          .8085d112E+01          .8311d198E+01          .8553d279E+01
       12          .8114d306E+01          .8362d526E+01          .8611d545E+01


REGULAR FORECAST RESULTS IN TERMS OF THE ORIGINAL DATA
**********************************************************

MODEL  1 FORECASTS AT BASE PERIOD  84 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD      LO. CONF. LIMIT          FORECAST          UP. CONF. LIMIT      ACTUAL, IF KNOWN

        1          .2333v10E+04           .2436b813E+04          .2548d120E+04
        2          .2415v98E+04           .2534d24E+04           .2667d00E+04
        3          .2511771E+04           .2674b68E+04           .2855d34E+04
        4          .2574b054E+04          .2811b27E+04           .3068d07E+04
        5          .2644v324E+04          .2944b67E+04           .3268d44E+04
        6          .2772d84E+04           .3129d31E+04           .3534d27E+04
        7          .2886b51E+04           .3277d46E+04           .3794d60E+04
        8          .2954v91E+04           .3477d43E+04           .4037d51E+04
        9          .3034d00E+04           .3691b57E+04           .4382d77E+04
       10          .3144v26E+04           .3827d69E+04           .4726d78E+04
       11          .3245d84E+04           .4066d10E+04           .5096d53E+04
       12          .3342d12E+04           .4233b03E+04           .5486d38E+04


      MODELO ARIMA(0,2,2) DADOS TRANSFORMADOS
      GRAPH OF FORECASTS AT ORIGIN 1
      GRAPH INTERVAL IS  .3746E+02
        .1e71E+04                    .3346E+04                    .5221E+04      VALUES
      +------------+------------+------------+------------+------------+
      :                                                                 :      .2436B3E+04
      :                A                                                :      .2534B2E+04
      :                 X                                               :      .2674B6E+04
      :                  A                                              :      .2811B2E+04
      :                   X                                             :      .2944B6E+04
      :                    X                                            :      .3129D3E+04
      :                     X                                           :      .3277D4E+04
      :                      X                                          :      .3477D4E+04
      :                       X                                         :      .3691B5E+04
      :                        A                                        :      .3827D6E+04
      :                         X                                       :      .4066D1E+04
VA    :                          A                                      :      .4233B0E+04
      :                           X                                     :      .4486B3E+04
      :                            A                                    :      .4806B1E+04
      :                             X                                   :      .5203E+04
      +------------+------------+------------+------------+------------+

WEIGHTS USED IN CALCULATING CONFIDENCE LIMITS
          J          PSI(J)

          0          .1000000E+01
          1          .1299995E+01
          2          .1617711E+01
          3          .1826184E+01
          4          .1983746E+01
          5          .2044126E+01
          6          .2070922E+01
          7          .2077487E+01
          8          .2086653E+01
          9          .2097518E+01
         10          .2104483E+01
         11          .2106403E+01
```

Figura 12 — Relatório de saída do computador: gráfico dos valores previstos, para o ano de 1982 do IPC

Quadro 7

Previsão para os valores mensais do IPC — 1982

MESES	VALOR PREVISTO	INTERVALO DE CONFIANÇA (de 95%)	VALOR REALMENTE OCORRIDO	PERCENTUAL DE ERRO
Jan.	2 408,8	(2 333,9;2 486,1)	2 390,00	0,79
Fev.	2 538,2	(2 415,5;2 667,2)	2 556,00	0,69
Mar.	2 674,6	(2 501,8;2 859,3)	2 785,00	3,96
Abr.	2 818,3	(2 590,5;3 066,1)	2 934,00	3,94
Maio	2 969,7	(2 680,9;3 289,5)		
Jun.	3 129,2	(2 772,7;3 531,6)		
Jul.	3 297,3	(2 855,5;3 794,2)		
Ago.	3 474,5	(2 959,4;4 079,3)		
Set.	3 661,2	(3 054,1;4 388,9)		
Out.	3 857,8	(3 149,6;4 725,4)		
Nov.	4 065,1	(3 245,8;5 091,2)		
Dez.	4 284,5	(3 342,6;5 489,2)		

NOTA: O percentual de erro é obtido através da diferença entre o valor previsto e o valor realmente ocorrido dividida pelo valor realmente ocorrido.

6.3 – Previsões para o Consumo de Energia Elétrica no Rio Grande do Sul

A série observada para essa variável está exposta no Quadro 8 e apresenta 142 valores mensais, desde janeiro de 1970 até outubro de 1981.

6.3.1 – A Etapa da Identificação

6.3.1.1 – Avaliação preliminar

O gráfico da série observada para essa variável (Figura 13) permite visualizar uma série com tendência crescente e linear, sugerindo a existência de uma componente sazonal com periodicidade anual. Também pode-se perceber que a série apresenta maior dispersão em torno da linha de tendência, à medida que o tempo cresce.

Quadro 8

Consumo mensal de energia elétrica no Rio Grande do Sul — jan./70-out./81

(Mwh)

MESES	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981
Jan.	140 866	150 669	171 120	203 556	234 811	256 496	298 065	340 917	394 825	443 272	509 209	570 082
Fev.	129 002	135 280	162 714	189 489	225 277	249 877	289 977	322 201	390 210	390 138	492 248	541 401
Mar.	128 058	148 725	167 062	186 167	223 794	261 441	285 732	330 762	383 574	456 950	485 479	538 021
Abr.	129 069	148 278	167 215	186 382	214 992	252 986	297 526	328 211	403 584	431 662	487 579	516 306
Maio	135 350	152 817	168 532	200 570	228 349	260 978	296 963	338 541	351 228	413 552	497 852	522 933
Jun.	130 320	146 482	172 731	199 134	229 950	259 829	301 691	336 645	356 376	427 158	481 153	505 254
Jul.	139 719	147 126	177 512	207 022	231 310	263 384	296 390	347 089	364 498	418 580	495 582	506 583
Ago.	131 466	153 034	176 084	206 068	233 175	263 176	309 156	346 286	365 965	442 714	497 444	513 147
Set.	135 722	151 492	176 608	203 899	229 948	267 423	309 341	352 923	370 888	400 832	488 005	511 258
Out.	137 995	152 059	170 160	204 555	234 523	268 849	314 935	354 125	391 820	432 796	492 743	501 628
Nov.	138 114	151 645	168 059	205 399	230 502	268 935	313 229	348 506	386 556	433 931	488 446	
Dez.	144 942	164 310	177 155	219 610	242 348	277 466	329 040	407 152	394 231	456 001	533 441	

FONTE: CEEE – Departamento de Estatística.

NOTA: Os dados referem-se aos municípios atendidos pela CEEE, totalizando cerca de 98% do consumo total do Estado.

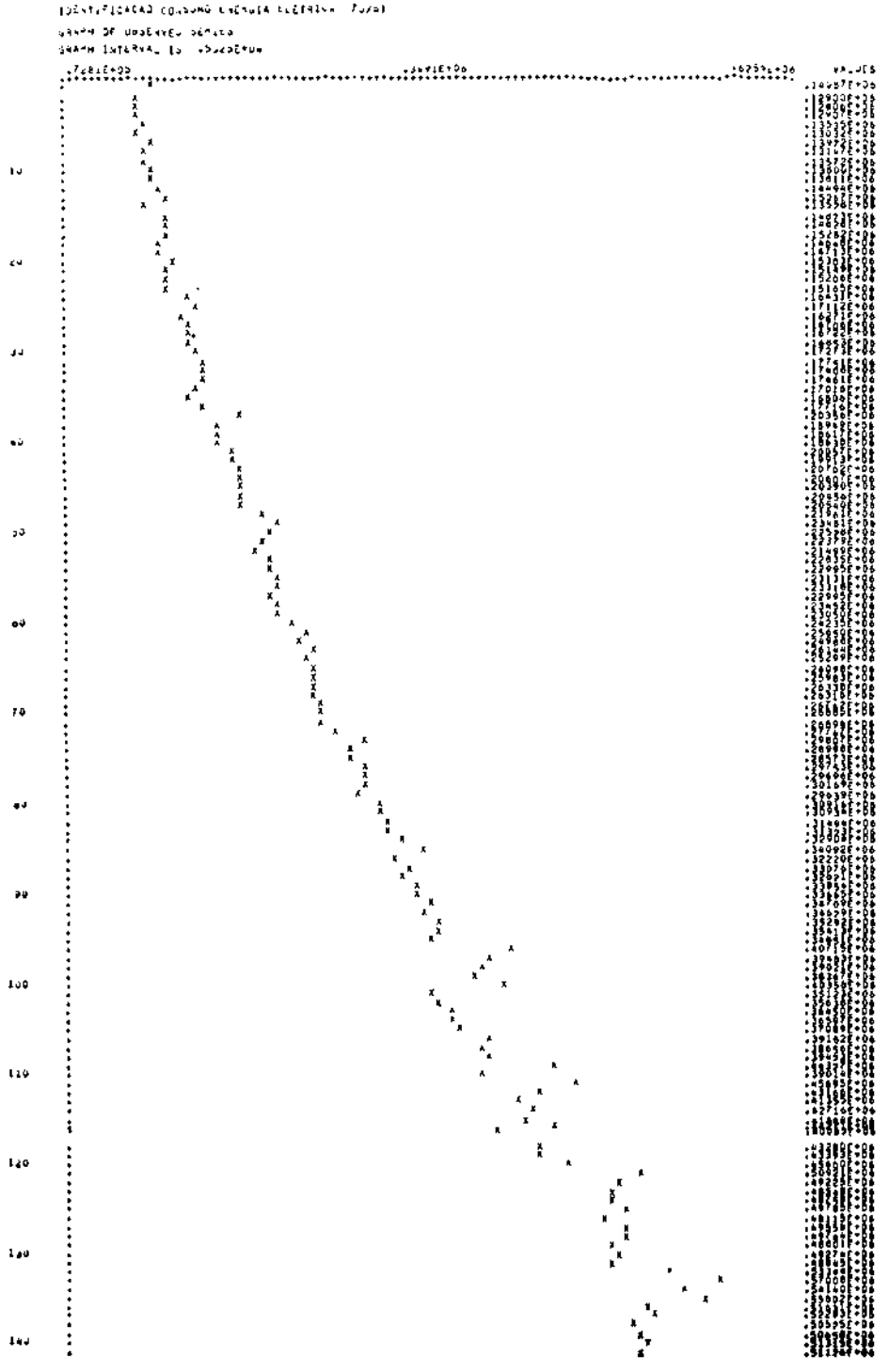


Figura 13 — Gráfico da série consumo de energia elétrica no Rio Grande do Sul — jan./70-out./79

A avaliação da necessidade de transformação da variável é feita através do Gráfico "Amplitude X Média", exposto na Figura 14.[12] Esse diagrama mostra uma tendência crescente da amplitude, embora com oscilações, em relação à média dos segmentos, o que indica, de acordo com o critério de Box & Cox visto em capítulos anteriores, a conveniência do uso da transformação logarítmica natural, $T_t = \ln Z_t$, onde Z_t representa o consumo de energia elétrica gaúcho, no período de janeiro de 1970 a outubro de 1981, o qual será, a partir de agora, a variável escolhida para trabalho.

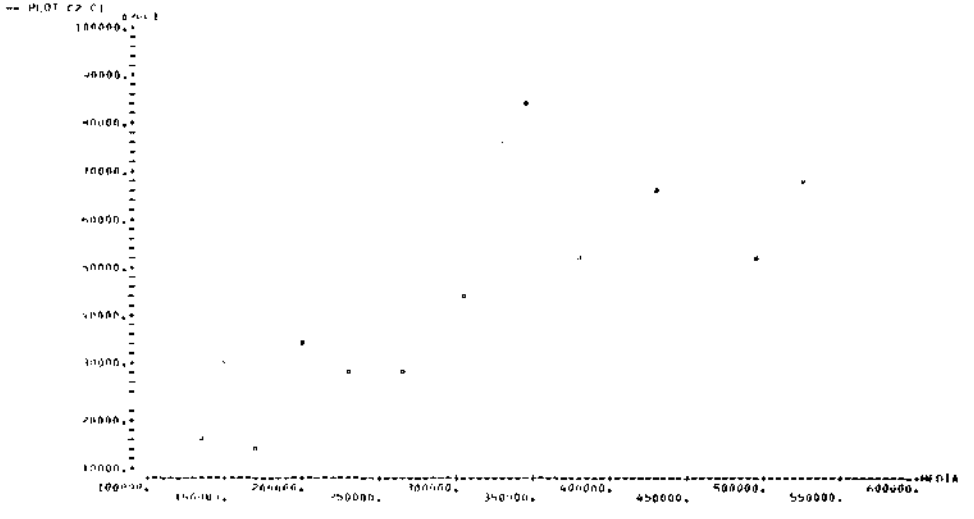


Figura 14 – Gráfico "Amplitude x Média" para a série consumo de energia elétrica no Rio Grande do Sul – jan./70-out./81

6.3.1.2 – O estabelecimento da ordem (p, d, q) x (P, D, Q)

Obtenção dos valores de ordem "d" e "D"

Na análise do gráfico da série $T_t = \ln Z_t$ (Figura 15), visualiza-se a existência de uma série com tendência linear, crescente, com pequenos picos de periodicidade anual. Significa dizer que a série é não estacionária, tanto no que se refere à parte tendencial quanto à parte sazonal, e torna-se necessário diferenciá-la para obtenção de uma série estacionária.

No estabelecimento das ordens "d" e "D" dos operadores diferença, utiliza-se o estudo da função de autocorrelação. Para tanto, buscou-se analisar a ACF das séries T_t, ∇T_t, $\nabla^2 T_t$ e $\nabla\nabla_{12} T_t$. Os gráficos da ACF para essas quatro séries estão expostos, respectivamente, nas Figuras 16, 17, 18 e 19 e seus valores numéricos nos Quadros 9 e 10.

[12] Dado que a periodicidade do ciclo sazonal é anual, construiu-se este gráfico dividindo a seqüência de observações em segmentos de tamanho doze.

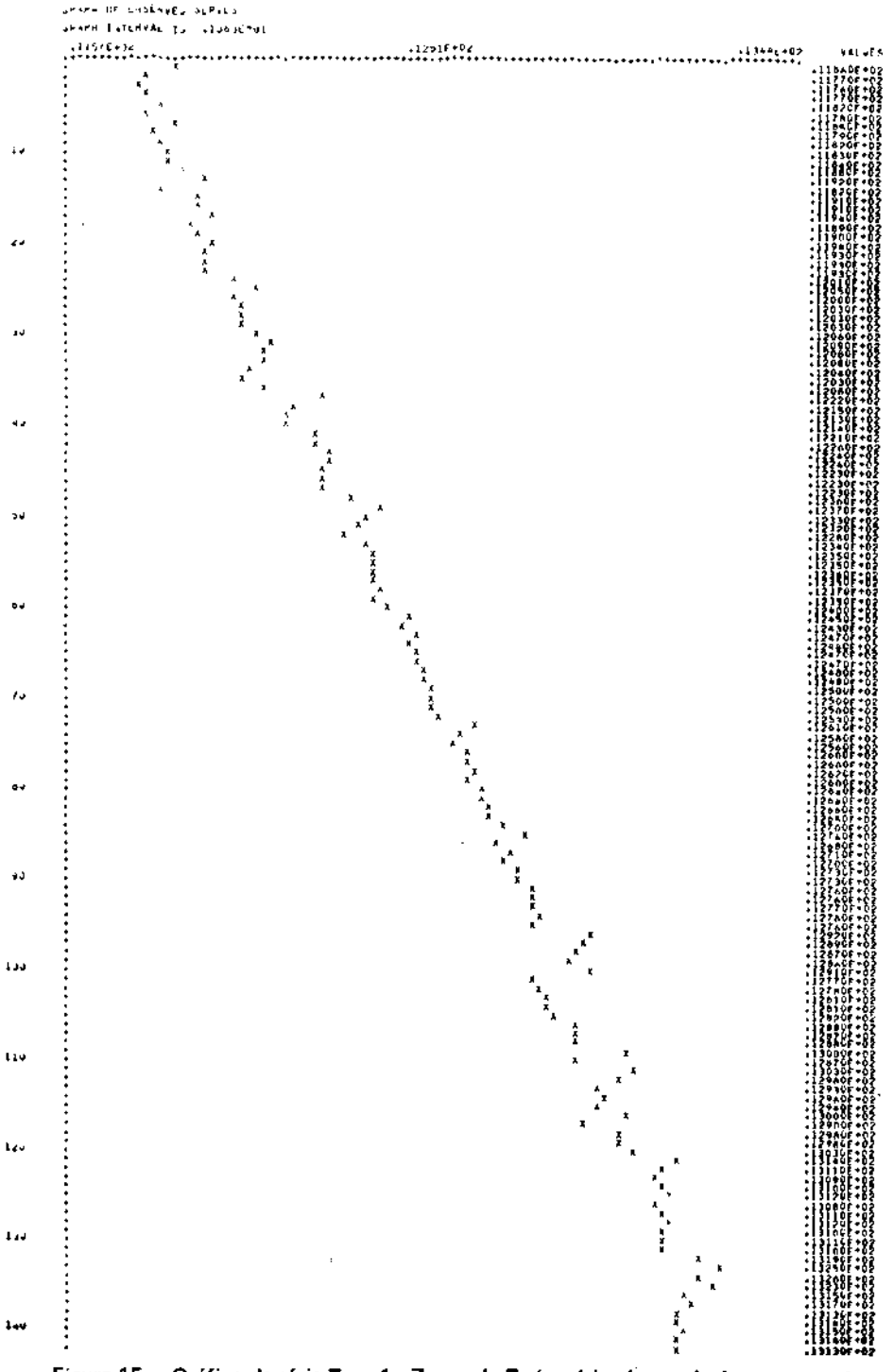


Figura 15 – Gráfico da série $T_t = 1n\ Z_t$, onde Z_t é a série observada do consumo de energia elétrica

A Figura 16 permite observar que a ACF da série T_t não apresenta uma queda rápida para zero, revelando a correção da suposição inicial da não estacionariedade tendencial da série, quando da avaliação da Figura 15. Por outro lado, esse gráfico não permite certificar a correção da hipótese da existência da sazonalidade, uma vez que, dada a escala do gráfico, não é visível uma flutuação cíclica em seus valores.

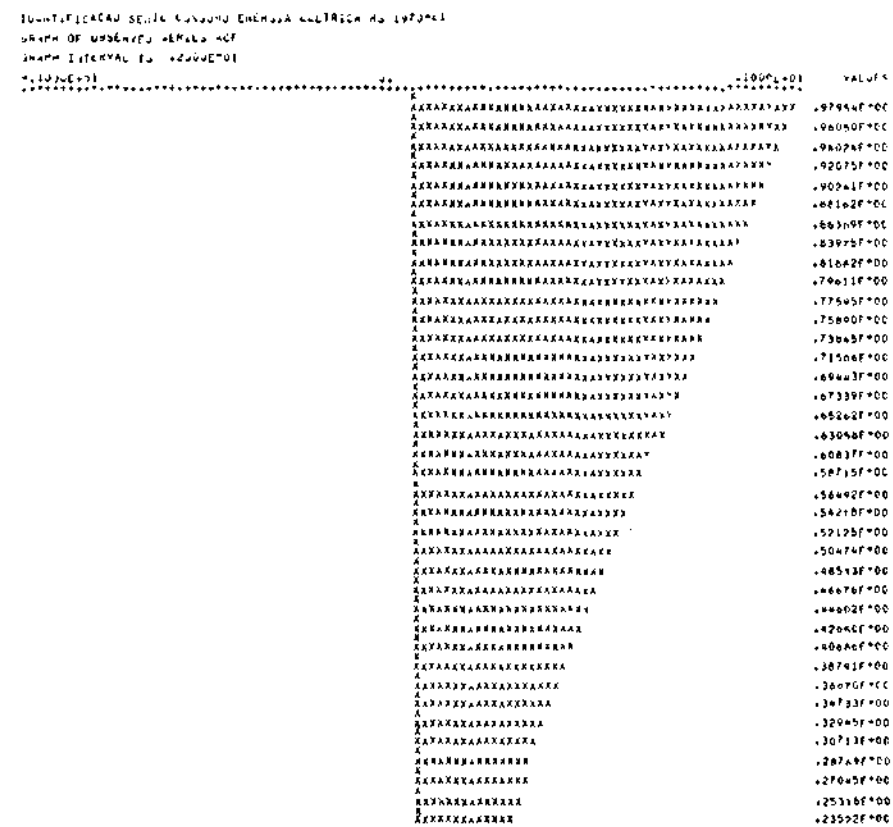


Figura 16 – Gráfico da função de autocorrelação da série T_t

A ACF da série ∇T_t (Figura 17) apresenta uma queda rápida para zero, voltando a ter valores significativamente distintos de zero nos períodos defasados 12, 24 e 36, o que comprova a existência da componente sazonal. Considerando o teste de significância de $2\,DP(r_k)$, $k>0$, somente o valor da ACF de defasagem um e dos pontos sazonais são diferentes de zero[13].

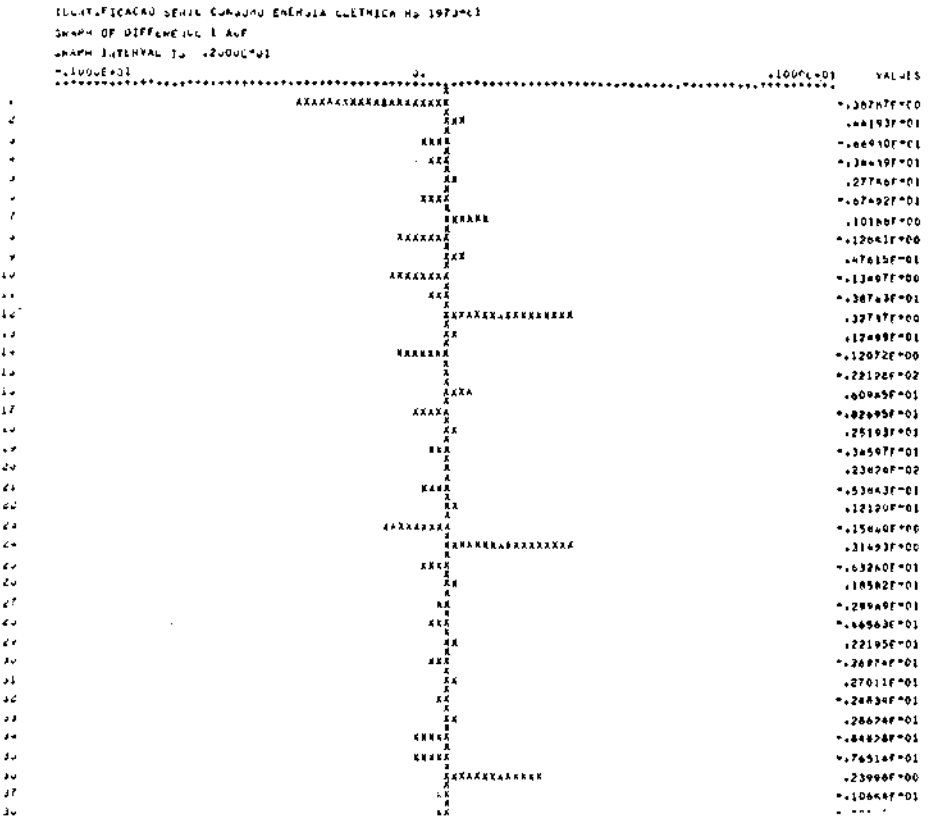


Figura 17 — Gráfico da função de autocorrelação da série ∇T_t

[13] Deve-se salientar, porém, que para o teste de $2DP(r_k)$, $k \geq 12$, os valores de ACF nos pontos sazonais não são significantes, fato que pode estar indicando a não necessidade de diferença de ordem 12.

A ACF da série $\nabla^2 T_t$ (Figura 18) também apresenta uma queda rápida para zero (exceto nos pontos sazonais), porém com maior flutuação em seus valores, o que torna preferível a série ∇T_t. De acordo com o critério de ANDERSON(1979) a $Var(\nabla T_t) < Var(T_t)$ e a $Var(\nabla^2 T_t) > Var(\nabla T_t)$, o que confirma a preferência pela série ∇T_t. Portanto pode-se concluir que d=1.

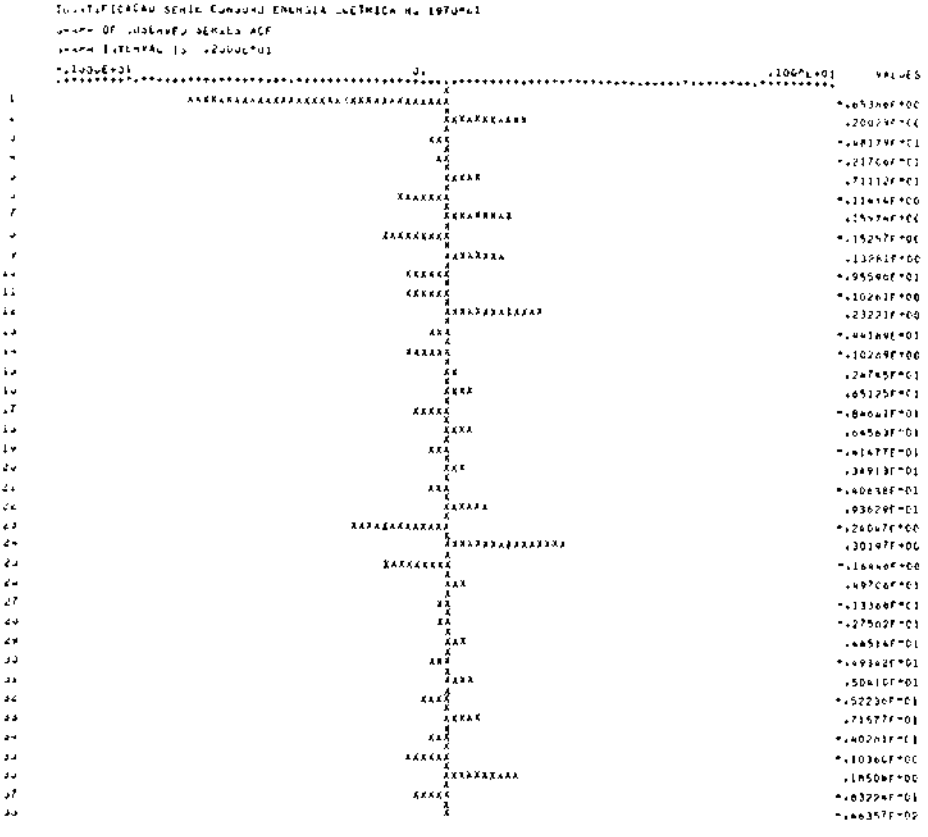


Figura 18 – Gráfico da função de autocorrelação da série $\nabla^2 T_t$

Para a obtenção do valor da ordem "D", devem-se analisar as ACF das séries ∇T_t diferenciadas de ordem doze. A Figura 19 apresenta o gráfico da ACF da série $\nabla\nabla_k \ln Z_t$. Percebe-se que os valores dos coeficientes de autocorrelação se apresentam maiores, mesmo nos pontos de defasagem múltiplos do período sazonal. A série $\nabla\nabla_{12} T_t$ tem variância superior à da série ∇T_t. Portanto não é conveniente o uso do operador diferença de ordem 12. O valor de D é zero.

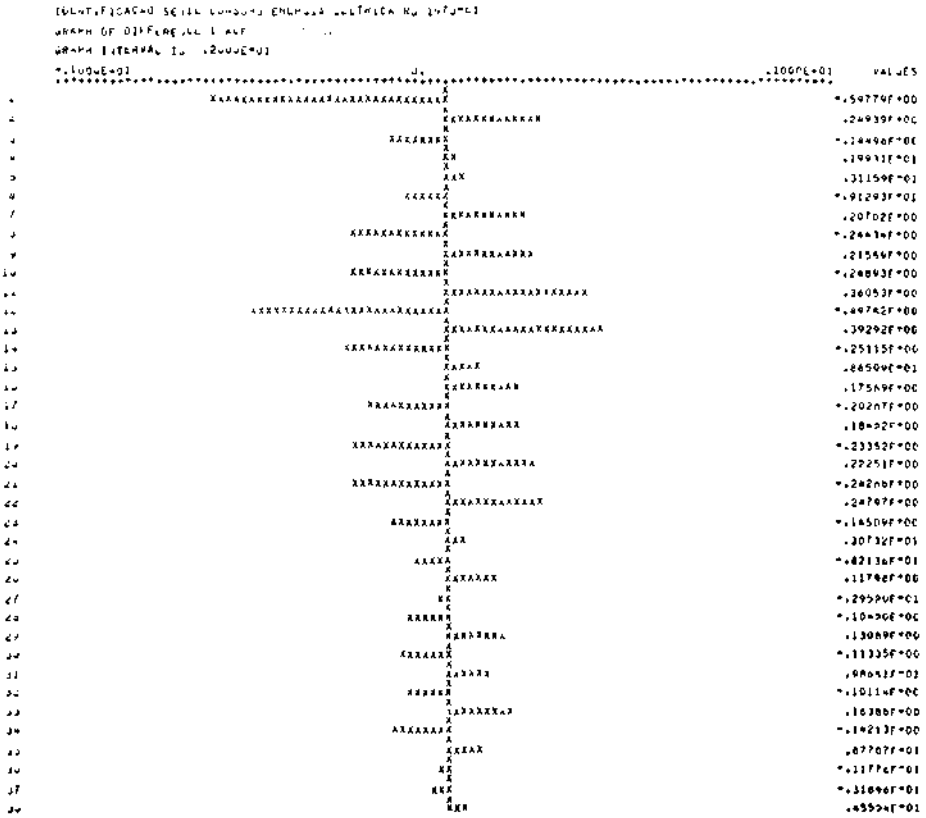


Figura 19 — Gráfico da função de autocorrelação da série $\nabla\nabla_{12} T_t$

Quadro 9

Valores numéricos da ACF para as séries T_t e ∇T_t

AUTOCORRELATION FUNCTION

DATA = IDENTIFICACAO SERIE CONSUMO ENERGIA ELETRICA RS 1970*01

DIFFERENCING = ORIGINAL SERIES IS YOUR DATA.
DIFFERENCING BELOW ARE OF ORDER 1

YOUR DATA TRANSFORMED DATA = LOG(Z(T)*0.

ORIGINAL SERIES
MEAN OF THE SERIES = .1401E+06
ST. DEV. OF SERIES = .4631E+05
NUMBER OF OBSERVATIONS = 144

1 TO 12	0.96	0.91	0.84	0.77	0.72	0.68	0.66	0.64	0.62	0.60	0.58	0.56
ST.ER.	0.08	0.14	0.18	0.21	0.24	0.26	0.28	0.30	0.32	0.33	0.35	0.36
13 TO 24	0.54	0.52	0.49	0.47	0.45	0.43	0.41	0.40	0.39	0.38	0.37	0.36
ST.ER.	0.37	0.38	0.39	0.40	0.40	0.41	0.42	0.42	0.43	0.43	0.44	0.44
25 TO 36	0.35	0.34	0.33	0.33	0.32	0.32	0.31	0.31	0.31	0.31	0.30	0.30
ST.ER.	0.45	0.45	0.46	0.46	0.46	0.46	0.47	0.47	0.47	0.47	0.47	0.47
37 TO 48	0.29	0.28										
ST.ER.	0.48	0.48										

MEAN DIVIDED BY ST. ERROR = .3017E+01

TO TEST WHETHER THIS SERIES IS WHITE NOISE THE VALUE .1497E+02
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 12 DEGREES OF FREEDOM

DIFFERENCE 1

MEAN OF THE SERIES = -.2477E+04
ST. DEV. OF SERIES = .1407E+05
NUMBER OF OBSERVATIONS = 143

1 TO 12	-0.39	0.04	-0.07	-0.03	0.03	-0.07	0.10	-0.13	0.05	-0.13	-0.04	0.23
ST.ER.	0.08	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
13 TO 24	0.31	-0.12	-0.04	0.06	-0.03	0.03	-0.03	0.00	-0.05	0.01	-0.16	0.21
ST.ER.	0.11	0.11	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
25 TO 36	-0.26	0.02	-0.03	-0.05	0.04	-0.08	0.03	-0.04	0.03	-0.04	-0.08	0.14
ST.ER.	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
37 TO 48	-0.01	-0.05										
ST.ER.	0.12	0.12										

MEAN DIVIDED BY ST. ERROR = .1227E+01

TO TEST WHETHER THIS SERIES IS WHITE NOISE THE VALUE .4504E+02
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 12 DEGREES OF FREEDOM

Quadro 10

Valores numéricos da ACF para as séries $\nabla^2 T_t$ e $\nabla\nabla_{12} T_t$

DATA = IDENTIFICACAO SERIE CONSUMO ENERGIA ELETRICA RS 1970*01

DIFFERENCING = ORIGINAL SERIES IS YOUR DATA DIFFERENCED BY
(1 - B**1)*(1 - B**12)
DIFFERENCES BELOW ARE OF ORDER 1

YOUR DATA TRANSFORMED DATA = LOG(Z(T)*0.

ORIGINAL SERIES
MEAN OF THE SERIES = .1976E+03
ST. DEV. OF SERIES = .7524E+01
NUMBER OF OBSERVATIONS = 140

1 TO 12	-0.05	0.13	-0.12	-0.12	0.12	-0.11	0.10	-0.12	0.12	-0.13	-0.10	0.23
ST.ER.	0.08	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
13 TO 24	-0.04	-0.14	0.11	0.07	-0.03	0.09	-0.04	0.03	-0.04	0.04	-0.24	0.20
ST.ER.	0.13	0.13	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
25 TO 36	-0.10	0.05	-0.01	-0.03	0.04	-0.05	0.05	-0.03	0.07	-0.04	-0.10	0.19
ST.ER.	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
37 TO 48	-0.08	-0.04										
ST.ER.	0.14	0.14										

MEAN DIVIDED BY ST. ERROR = .4744E+01

TO TEST WHETHER THIS SERIES IS WHITE NOISE THE VALUE .1079E+02
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 12 DEGREES OF FREEDOM

DIFFERENCE 1

MEAN OF THE SERIES = -.1367E+03
ST. DEV. OF SERIES = .1532E+02
NUMBER OF OBSERVATIONS = 139

1 TO 12	-0.60	0.32	-0.14	0.05	0.03	-0.09	0.21	-0.44	0.22	-0.25	0.38	-0.50
ST.ER.	0.09	0.12	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
13 TO 24	0.39	-0.23	0.07	0.13	-0.03	0.18	-0.23	0.06	-0.24	0.25	-0.15	0.13
ST.ER.	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
25 TO 36	-0.08	0.12	-0.06	-0.16	0.14	-0.11	0.10	-0.10	0.16	-0.14	0.07	-0.01
ST.ER.	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
37 TO 48	-0.03	0.07										
ST.ER.	0.19	0.19										

MEAN DIVIDED BY ST. ERROR = -.1815E+01

TO TEST WHETHER THIS SERIES IS WHITE NOISE THE VALUE .1346E+03
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 12 DEGREES OF FREEDOM

Obtenção dos valores da ordem (p, q) x (P, Q)

O estabelecimento da ordem (p,q) x (P,Q) é feito a partir da análise das funções de autocorrelação e autocorrelação parcial da série estacionária $\nabla \ln Z_t$.

Percebe-se pela Figura 17 que a ACF da série ∇T_t tem um comportamento oscilatório, com a primeira defasagem significativamente diferente de zero (o valor crítico, a partir do qual é traçada a linha de controle, é $2DP(r_k)$, $k>0$), voltando a apresentar valores significativamente distintos de zero para as defasagens 12, 24, e 36.

A Figura 20 mostra que a PACF desta série tem a primeira defasagem significativamente diferente de zero (o valor crítico, a partir do qual é traçada a linha de controle, neste caso, é $2\ DP(\hat{\phi}_{kk})$), voltando a apresentar valores significativos em torno da defasagem 12 (mais especificamente, nas defasagens 10, 11, 12 e 13) e na defasagem 23.

Figura 20 — Gráfico da função de autocorrelação parcial da série ∇T_t

Quadro 11

Valores numéricos da PACF para as séries T_t e ∇T_t

```
PARTIAL AUTOCORRELATIONS
DATA - IDENTIFICACAO CONSUMO ENERGIA ELETRICA RS
DIFFERENCING - ORIGINAL SERIES IS YOUR DATA.
              DIFFERENCES BELOW ARE OF ORDER   1
YOUR DATA TRANSFORMED DATA = LOG(Z(T)+0.          )

ORIGINAL SERIES
MEAN OF THE SERIES = .12515E+02
ST. DEV. OF SERIES = .44131E+00
NUMBER OF OBSERVATIONS =   142
  1- 12      0.98   0.02  -0.04   0.01   0.02  -0.07   0.04  -0.12   0.04  -0.05   0.05   0.06
 13- 24      0.08  -0.11   0.09  -0.04  -0.01  -0.05  -0.02   0.02  -0.02  -0.05   0.06   0.07
 25- 36     -0.08   0.03  -0.07   0.00  -0.02   0.01  -0.04  -0.05   0.04  -0.07   0.03   0.03
 37- 38     -0.02  -0.03

DIFFERENCE  .
MEAN OF THE SERIES = .90074E-02
ST. DEV. OF SERIES = .47073E-01
NUMBER OF OBSERVATIONS =   141
  1- 12     -0.39  -0.13  -0.11  -0.12  -0.05  -0.11   0.03  -0.11  -0.07  -0.20  -0.25   0.21
 13- 24      0.30  -0.00  -0.03   0.11   0.01  -0.01  -0.09  -0.04  -0.05   0.07  -0.16   0.08
 25- 36     -0.02   0.10   0.05  -0.14  -0.10   0.04  -0.02   0.01   0.03  -0.02  -0.04   0.14
 37- 38      0.07  -0.04
```

A análise conjunta da ACF e PACF da série $\nabla \ln Z_t$ permite estabelecer algumas hipóteses de modelos:

- considerando que a ACF apresenta a primeira defasagem distinta de zero, voltando a ter coeficientes significativos somente nas defasagens sazonais, e que a PACF apresenta queda pouco representativa de seus valores até em torno dos coeficientes de defasagem sazonal 12, tem-se um modelo SARIMA $(0,1,1) \times (1,0,0)_{12}$;
- como o valor da defasagem 23 da PACF é significativo, deve-se considerar também um modelo com componente sazonal auto-regressivo de ordem dois. Nesse caso, sugere-se um SARIMA $(0,1,1) \times (2,0,0)_{12}$;
- a PACF tem coeficientes significativos nas defasagens 10, 11, 12, 13 e 23, e o coeficiente de defasagem 36, embora não significativo, apresenta-se próximo da "linha de controle". Esses fatos podem indicar que, tanto a ACF quanto a PACF, não apresentam queda rápida para zero, sugerindo modelos cuja componente sazonal seja ARMA. Assim, deve-se examinar também o modelo SARIMA $(0,1,1) \times (1,0,1)_{12}$;
- o fato de a PACF se apresentar oscilatória e somente com o coeficiente da primeira defasagem distinto de zero (exceto os coeficientes de defasagem sazonal) sugere que se estude, também, modelos que contenham processos auto-regressivos na componente não sazonal.

Assim, o processo de identificação permitiu a suposição de vários modelos. Uma estratégia de ação conveniente, com vistas ao barateamento dos custos de processamento eletrônico, é começar a análise pelo modelo que se supõe mais provável (no caso, SARIMA $(0,1,1) \times (1,0,0)_{12}$ ou SARIMA $(0,1,1) \times (1,0,1)_{12}$), acrescentando ou alterando parâmetros conforme sugerirem as informações oriundas do sumário dos relatórios de saída para cada modelo e suas respectivas ACF e PACF residuais.

De outra parte, deve ser salientado que, conforme se pode verificar no processo de construção do modelo para a série de Índice de Preços ao Consumidor, o processo iterativo de aproximação sucessiva de Marquardt apresenta convergência bastante rápida para as estimativas eficientes dos parâmetros. Por essa razão, não se considera necessário calcular boas estimativas iniciais para os parâmetros dos modelos supostos, bastando indicar valores aleatórios, dentro das respectivas regiões de admissibilidade, para alimentar a rotina "Estimação".

Da mesma forma, a avaliação preliminar da existência do parâmetro θ_0 também não será feita. A análise dos modelos supostos será efetuada considerando a existência do termo constante. Posteriormente, através do teste de significância deste parâmetro, decidir-se-á pela conveniência ou não da manutenção de θ_0 nos modelos supostos.

6.3.2 – A Etapa da Estimação

O Quadro 12 expõe o resumo das principais informações estatísticas proporcionadas pelos sumários dos relatórios de saída do computador para a etapa da estimação.

Por esse quadro pode-se concluir que:

- dos três primeiros modelos apresentados nesse quadro, o terceiro, SARIMA $(0,1,1) \times (1,0,1)_{12}$, é o que possui menor variância residual[14]. Seus parâmetros são significativos, exceto o termo constante. Ao estimar o modelo sem θ_0, obtém-se o quarto modelo;
- o quarto modelo apresenta todos os parâmetros significativos, com variância residual levemente superior ao terceiro, permitindo supor que seja, até o momento, o melhor;
- ao incluir θ_2 nos modelos três e quatro, obtêm-se os modelos SARIMA $(0,1,2) \times (1,0,1)_{12}$ com e sem θ_0, que apresentam variância residual superior à do quarto modelo;
- se, ao invés de acrescentar θ_2, for incluído um parâmetro auto-regressivo na componente não sazonal, obtém-se o sétimo modelo, SARIMA $(1,1,1) \times (1,0,1)_{12}$. Esse apresenta ϕ_1 e θ_0 não significativos, além de maior variância residual do que aquele;
- eliminando ϕ_1, volta-se novamente ao terceiro modelo. Eliminando θ_1, tem-se o oitavo modelo, SARIMA $(1,1,0) \times (1,0,1)_{12}$. Esse também apresenta maior variância residual que o quarto modelo.

Portanto a escolha do modelo mais adequado à série observada recai no quarto modelo, SARIMA $(0,1,1) \times (1,0,1)_{12}$, sem o termo constante:

$$(1 - 0{,}94055\ B^{12})\ \nabla \ln Z_t = (1 - 0{,}57944\ B)(1 - 0{,}63588\ B^{12})a_t .$$

[14] A escolha do critério da variância residual mínima, aqui utilizado, prende-se às mesmas justificativas expostas para a escolha do modelo da série anterior.

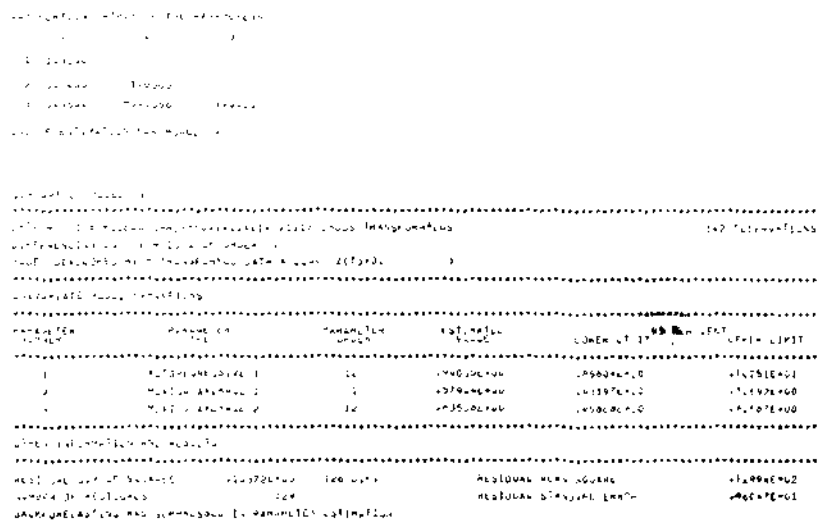
Quadro 12

Principais informações estatísticas dos modelos supostos para a série consumo de energia elétrica no Rio Grande do Sul — jan./70-out./81

MODELO SUPOSTO	ESTIMATIVAS DOS PARÂMETROS	DESVIO PADRÃO DOS PARÂMETROS	VARIÂNCIA RESIDUAL
(1) SARIMA $(0,1,1)x(1,0,0)_{12}$ com θ_o	ϕ_{12} = 0,48650	$DP(\phi_k)$ = 0,0815561	
	θ_1 = 0,63979	$DP(\theta_1)$ = 0,0713367	0,0014654
	θ_o = 0,0048753	$DP(\theta_o)$ = 0,0015066	
(2) SARIMA $(0,1,1)x(2,0,0)_{12}$ com θ_o	ϕ_{12} = 0,37640	$DP(\phi_{12})$ = 0,00355867	
	ϕ_{24} = 0,26775	$DP(\phi_{24})$ = 0,0934642	0,0013444
	θ_1 = 0,60960	$DP(\theta_1)$ = 0,0773061	
	θ_o = 0,0030087 (N)	$DP(\theta_o)$ = 0,0016962	
(3) SARIMA $(0,1,1)x(1,0,1)_{12}$ com θ_o	ϕ_{12} = 0,89408	$DP(\phi_{12})$ = 0,0605918	
	θ_1 = 0,59426	$DP(\theta_1)$ = 0,0747091	0,0012933
	θ_{12} = 0,59914	$DP(\theta_{12})$ = 0,1046428	
	θ_o = 0,00099844 (N)	$DP(\theta_o)$ = 0,0008091	
(4) SARIMA $(0,1,1)x(1,0,1)$ sem θ_o	ϕ_{12} = 0,94055	$DP(\phi_{12})$ = 0,0432653	
	θ_1 = 0,57945	$DP(\theta_1)$ = 0,0755204	0,0012994
	θ_{12} = 0,63589	$DP(\theta_{12})$ = 0,0920153	
(5) SARIMA $(0,1,2)x(1,0,1)$ com θ_o	ϕ_{12} = 0,89820	$DP(\phi_{12})$ = 0,0593928	
	θ_{12} = 0,60090	$DP(\theta_{12})$ = 0,1039234	
	θ_1 = 0,60758	$DP(\theta_1)$ = 0,093204	0,0013012
	θ_2 = 0,60090	$DP(\theta_2)$ = 0,0934433	
	θ_0 = 0,00094837 (N)	$DP(\theta_o)$ = 0,0008368	
(6) SARIMA $(0,1,2)x(1,0,1)$ sem θ_o	ϕ_{12} = 0,93694	$DP(\phi_{12})$ = 0,0446530	
	θ_{12} = 0,63133	$DP(\theta_{12})$ = 0,0932448	0,0013045
	θ_1 = 0,60242	$DP(\theta_1)$ = 0,0925051	
	θ_2 = -0,64007	$DP(\theta_2)$ = 0,0922361	
(7) SARIMA $(1,1,1)x(1,0,1)_{12}$ com θ_o	ϕ_1 = -0,27481 (N)	$DP(\phi_1)$ = 0,1614183	
	ϕ_{12} = 0,87872	$DP(\phi_{12})$ = 0,0654489	
	θ_{12} = 0,56196	$DP(\theta_{12})$ = 0,1084336	0,0013494
	θ_1 = 0,31175 (N)	$DP(\theta_1)$ = 0,1607653	
	θ_o = 0,0015206 (N)	$DP(\theta_o)$ = 0,0013764	
(8) SARIMA $(1,1,0)x(1,0,1)_{12}$ com θ_o	ϕ_1 = 0,49386	$DP(\phi_1)$ = 0,0805408	
	ϕ_{12} = 0,88036	$DP(\phi_{12})$ = 0,0589744	0,0013588
	θ_{12} = 0,59841	$DP(\theta_{12})$ = 0,1006377	
	θ_o = 0,0017472 (N)	$DP(\theta_o)$ = 0,0017257	

NOTA: O sinal (N) à direita do parâmetro estimado indica que esse é não significativo, de acordo com o critério de compará-lo com duas vezes seu desvio padrão.

Relatório de saída do computador: sumário de informações estatísticas para o modelo escolhido



6.3.3 – A Etapa da Checagem do Diagnóstico

A ACF, a PACF e a série de resíduos para o modelo escolhido estão apresentadas nas Figuras 21, 22 e 23. Pode-se perceber que nem as ACF e PACF, nem a série de resíduos apresentam comportamento que indique má identificação. A série de resíduos pode ser considerada aleatória.

O teste de "portmanteau" apresenta a estatística "Q" = $n\Sigma r_k^2$ = 35,77. A um nível de significância de 10%, o valor tabelado qui-quadrado, para 27 graus de liberdade, é 36,7. Logo, pode-se aceitar a hipótese de que os resíduos não são não-brancos[15].

Portanto o modelo escolhido nas etapas anteriores está bem identificado, atendendo às exigências da teoria. Está encerrada a fase de construção do modelo.

---

[15] Para séries sazonais, é comum a utilização de outros tipos de testes mais específicos, a fim de avaliar a permanência de algum comportamento cíclico nos resíduos. Entre eles, é usual o Periodograma Acumulado. Como, no problema em questão, não aparece nenhuma dúvida sobre a não existência de comportamento sistemático nos resíduos, julgou-se desnecessária sua aplicação. Para detalhes sobre o teste do Periodograma Acumulado, ver, entre outros, BOX & JENKINS (1976), *MORETTIN & TOLOI (1981) e Nelson* (1973).

6.3.4 – Previsões

Para essa série optou-se em inferir valores previstos para 12 períodos adiante, a partir da observação 142, correspondente ao mês de outubro de 1981. Nas páginas seguintes apresentam-se os resultados computacionais e, no Quadro 14 os valores previstos para a série original.

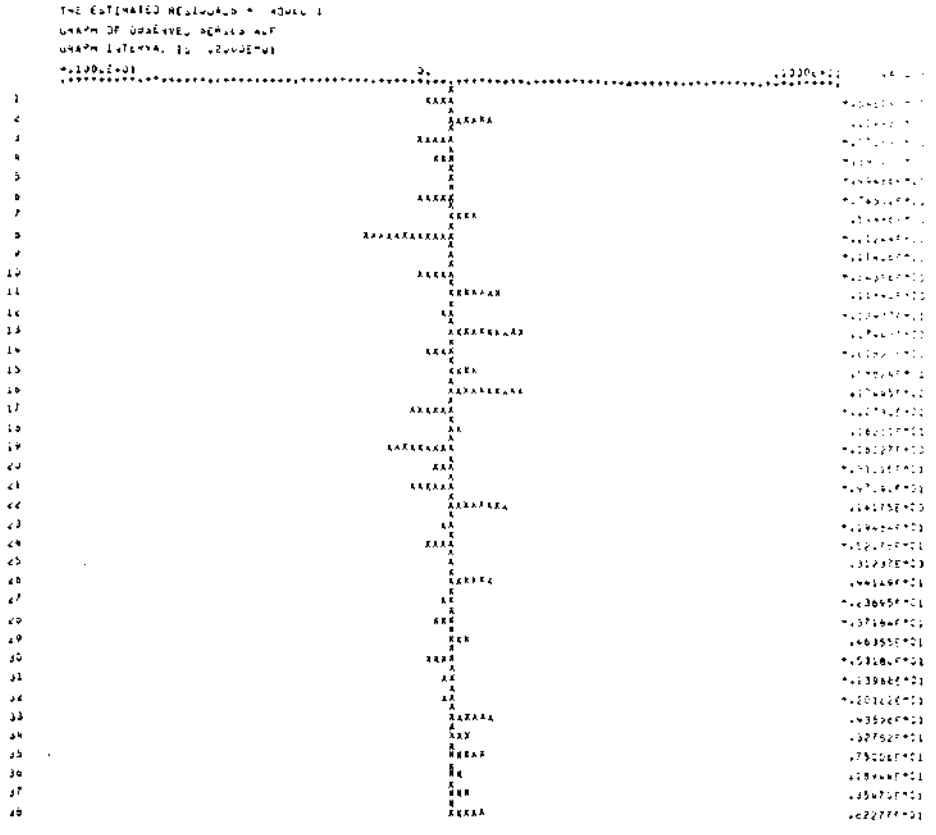


Figura 21 — Gráfico da ACF da série de resíduos para o modelo
$(1 - 0{,}94055B^{12})\nabla \ln Z_t = (1 - 0{,}57944B)(1 - 0{,}63588B^{12})a_t$

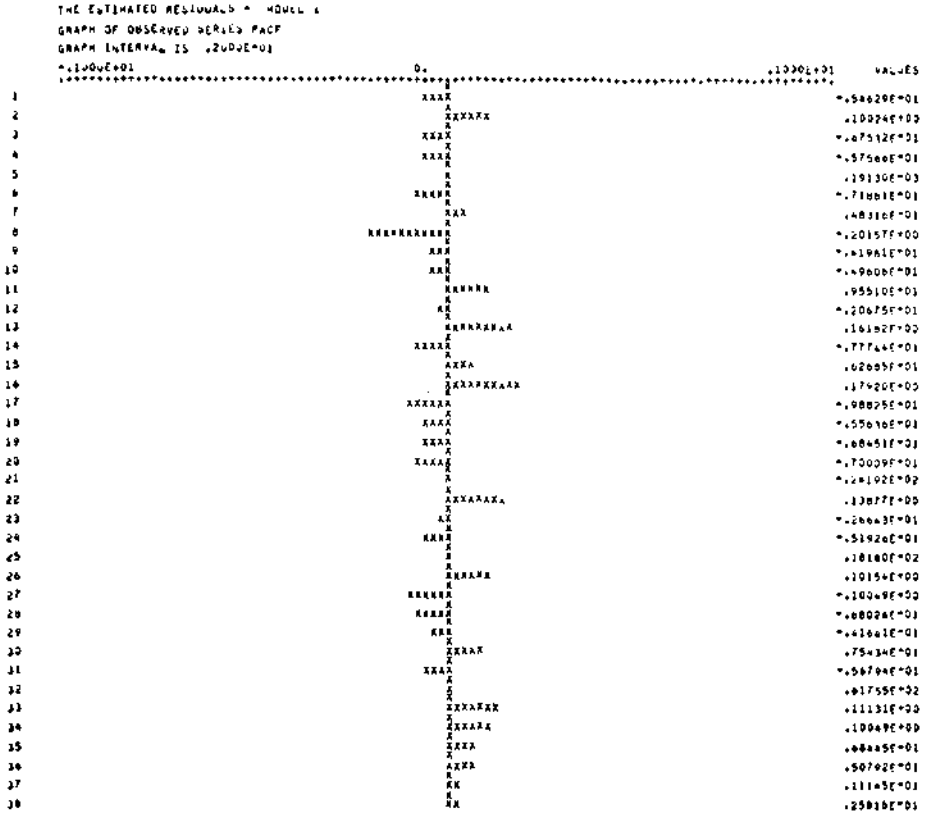


Figura 22 — Gráfico da PAC da série de resíduos para o modelo
$(1 - 0{,}94055B^{12})\nabla 1n\ Z_t = (1 - 0{,}57944B)(1 - 0{,}63588B^{12})a_t$

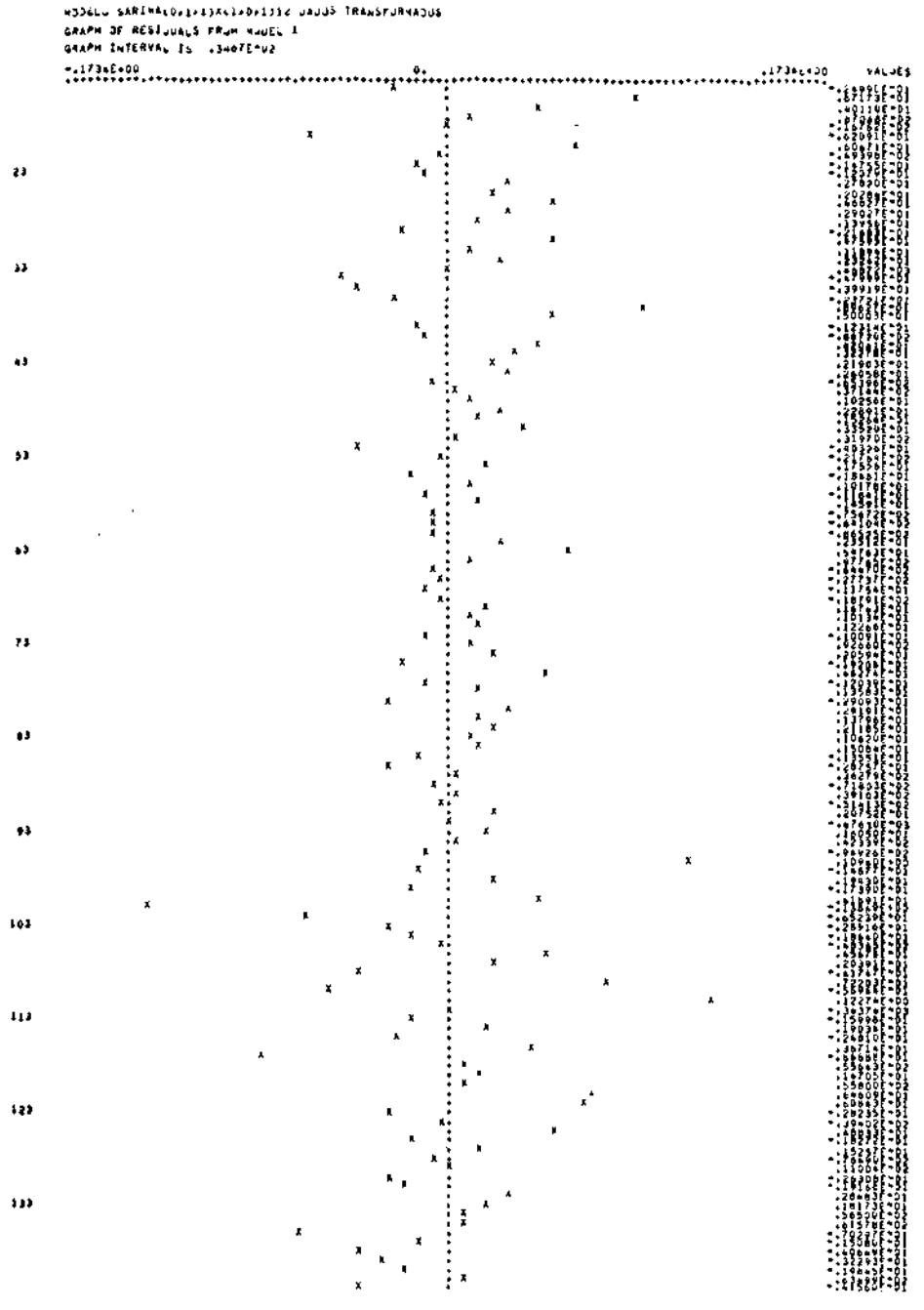


Figura 23 — Gráfico da série de resíduos para o modelo
$(1 - 0{,}94055B^{12}) \nabla \ln Z_t = (1 - 0{,}57944B)(1 - 0{,}63588B^{12})\, a_t$

Quadro 13

Relatório de saída do computador: previsão para os valores mensais do consumo de energia elétrica no Rio Grande do Sul — nov./81-out./82

```
TIME SERIES FORECASTING FOR MODEL  1
**********************************************************************
DATA =   Z = MODELO SARIMA(0,1,1)X(1,0,1)12 DADOS TRANSFORMADOS
DIFFERENCING ON   Z = 1) 1 OF ORDER  1
MODEL DEVELOPED WITH TRANSFORMED DATA = LOG( Z(T)+0.          )
**********************************************************************
UNIVARIATE MODEL PARAMETERS
**********************************************************************
PARAMETER NUMBER   PARAMETER TYPE          PARAMETER ORDER   BEGINNING VALUE
**********************************************************************
      1            AUTOREGRESSIVE 1            12            .940353E+00
      2            MOVING AVERAGE 1             1            .579046E+00
      3            MOVING AVERAGE 2            12            .635906E+00
**********************************************************************
NUMBER OF TIME ORIGINS FOR FORECASTS =          1
NUMBER OF FORECASTS AT EACH TIME ORIGIN =       12
FORECAST TIME ORIGINS ARE T=                    142
BACKFORECASTING WAS SUPPRESSED IN THIS ANALYSIS


REGULAR FORECAST RESULTS IN TERMS OF TRANSFORMED DATA
*****************************************************

MODEL  1 FORECASTS AT BASE PERIOD 142 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD     LO. CONF. LIMIT          FORECAST          UP. CONF. LIMIT     ACTUALS IF KNOWN

        1        .132788E+03          .133199E+03          .133610E+03
        2        .131259E+03          .132018E+03          .132777E+03
        3        .131015E+03          .132009E+03          .133003E+03
        4        .131308E+03          .132176E+03          .133044E+03
        5        .131569E+03          .132444E+03          .133320E+03
        6        .131218E+03          .132181E+03          .133144E+03
        7        .131111E+03          .132138IE+03         .133165E+03
        8        .130968E+03          .132017E+03          .133066E+03
        9        .131009E+03          .132097E+03          .133185E+03
       10        .131099E+03          .132226IE+03         .133355IE+03
       11        .130741E+03          .132077E+03          .133413E+03
       12        .130963E+03          .132186IE+03         .133590IE+03


REGULAR FORECAST RESULTS IN TERMS OF THE ORIGINAL DATA
******************************************************

MODEL  1 FORECASTS AT BASE PERIOD 142 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD     LO. CONF. LIMIT          FORECAST          UP. CONF. LIMIT     ACTUALS IF KNOWN

        1        .471381E+06          .519911E+06          .573488E+06
        2        .501013E+06          .559166E+06          .624479E+06
        3        .530975E+06          .597221E+06          .671924E+06
        4        .504769E+06          .594978E+06          .701826E+06
        5        .494926E+06          .599406E+06          .725887E+06
        6        .474736E+06          .595024E+06          .746781E+06
        7        .469JJ8E+06          .596758E+06          .760544E+06
        8        .487JJ5E+06          .613159E+06          .771687E+06
        9        .485J381E+06         .615J97E+06          .775087E+06
       10        .489J790E+06         .622724E+06          .787967E+06
       11        .485J163E+06         .619459E+06          .791916E+06
       12        .486J230E+06         .620003E+06          .793776E+06
```

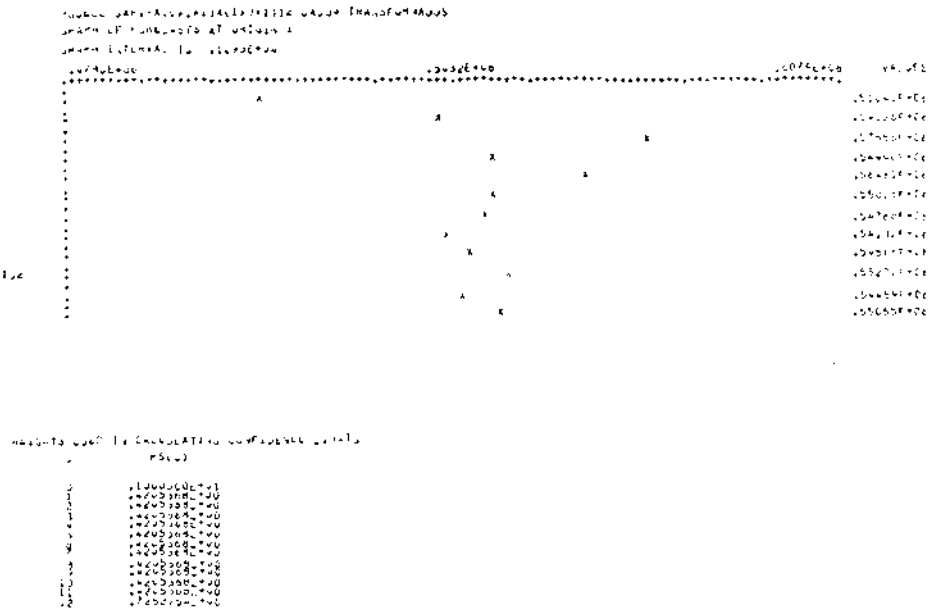


Figura 24 — Relatório de saída do computador: gráfico dos valores previstos para nov./81-out./82 do consumo de energia elétrica gaúcho

Quadro 14

Previsão para os valores mensais do consumo de energia elétrica gaúcho — nov./81-out./82

ANOS E MESES	VALOR PREVISTO	INTERVALO DE CONFIANÇA (de 95%)	VALOR REALMENTE OBSERVADO (se conhecido)	PERCENTUAL DE ERRO
1981				
Nov.	510 911,2	(476 311,7;548 023,7)	516 655	1,1
Dez.	541 261,5	(501 613,7;584 043,0)	559 779	3,1
1982				
Jan.	575 551,3	(530 457,5;624 478,5)	589 137	2,3
Fev.	549 978,7	(504 276,5;599 822,8)		
Mar.	564 805,6	(515 352,6;619 004,1)		
Abr.	550 248,6	(499 751,6;605 848,1)		
Maio	547 880,6	(495 408,2;605 910,8)		
Jun.	541 315,9	(487 405,9;601 188,7)		
Jul.	545 676,2	(489 338,1;608 500,6)		
Ago.	552 724,0	(493 719,0;618 780,7)		
Set.	544 592,6	(484 616,3;611 991,6)		
Out.	550 550,3	(488 123,6;620 960,9)		

NOTA: O percentual de erro é obtido através da diferença entre o valor previsto e o valor realmente observado (se conhecido) dividida pelo valor realmente observado.

7 – CONSIDERAÇÕES FINAIS

Este trabalho procurou expor, em detalhe, a metodologia criada pelos professores George E.P. Box e Gwilym M. Jenkins para a análise de séries univariantes de tempo. Foram abrangidos tanto os aspectos teóricos quanto questões relativas a seu manuseio prático.

Conforme se pode perceber no decorrer deste documento, a metodologia de Box & Jenkins emprega técnicas estatísticas relativamente sofisticadas e proporciona previsões bastante precisas, em comparação com outros métodos de previsão. Contudo o método apresenta um conjunto de restrições operacionais e pressupõe que a série observada tenha algumas especificidades para gerar boas previsões.

O fundamento dessa metodologia está inserido na Teoria dos Processos Estocásticos, onde as relações entre consecutivas observações assumem especial importância, estando o método incluído no conjunto de instrumentos de Análise de Séries Temporais classificados como **não automáticos**[1].

O método de Box & Jenkins para a Análise de Séries Temporais caracteriza-se por seu objetivo de estabelecer modelos do tipo auto-regressivo integrado-média móvel, de forma a se ajustar ao conjunto de pontos observados no tempo, partindo de um ciclo iterativo de cálculo.

A estratégia desse processo de "modelização" consiste em identificar um modelo tentativo capaz de descrever adequadamente os dados disponíveis, estimar eficientemente seus parâmetros, validar o modelo proposto através de uma avaliação de quão bem este é capaz de descrever a realidade e, em caso de considerar que a representação do modelo é adequada, usá-lo para fins de análise de séries temporais ou, em caso de rejeição, selecionar um novo modelo, repetindo-se o ciclo.

Verifica-se que a etapa crítica desse processo é a de identificação do modelo tentativo. Sua análise é feita a partir das funções amostrais de autocorrelação e autocorrelação parcial e, por se tratarem de dados amostrais, caracteriza-se pela falta de precisão. É comum, por essa razão, que mais de um pesquisador, trabalhando com a mesma série de dados observados, identifique modelos distintos. Nessa etapa é, então, usual trabalhar com mais de um modelo tentativo para, na etapa final de verificação, escolher aquele que, de acordo com algum critério previamente estabelecido, se mostra mais conveniente para a descrição dos dados de que se dispõe.

Quando da divulgação dessa metodologia, muitas críticas foram levantadas, fato que não permitiu sua utilização intensa no início dos anos setenta, embora os modelos assim obtidos sejam bastante simples e parcimoniosos, contendo um número pequeno de parâmetros, além de propor-

---

[1] São classificados, na maioria dos livros técnicos sobre o assunto, como **método automático** de análise de série temporal aqueles instrumentos para os quais, em geral, não existe um tratamento estatístico completo mas que, por serem *de fácil manuseio e apresentarem resultados satisfatórios em muitas situações, são intensamente utilizados. Entre* esses instrumentos destacam-se os modelos de decomposição (também chamada Análise Clássica de Séries Temporais) e os de alisamento.

cionarem previsões bastante satisfatórias quando comparados com os demais métodos existentes para a Análise de Séries Temporais.

Essa situação pode ser creditada, em muitos aspectos, às características da metodologia, cuja utilização apresenta algumas restrições de ordem operacional. Entre elas, podem-se destacar as seguintes:

- a necessidade de um conhecimento teórico de estatística bastante superior ao usualmente ministrado nas disciplinas de cursos, onde a estatística é vista como meio e não fim;
- a necessidade de um número mínimo relativamente grande de observações para produzir previsões acuradas: em torno de 40 para séries não sazonais e, para séries sazonais, 30 mais três vezes o tamanho do ciclo de sazonalidade;
- a possibilidade de acesso a pacotes computacionais, dado o grande volume de cálculos necessários;
- a necessidade de alguma experiência e sensibilidade técnica para o estabelecimento da ordem do modelo identificado, além do mero uso mecânico de programas computacionais "enlatados";
- a possibilidade de proporcionar a identificação de mais de um modelo que satisfaça às condições exigidas pelos dados, de acordo com o critério interpretativo de cada usuário;[2]
- os parâmetros dos modelos estimados não são atualizados a cada nova observação obtida, pressupondo, por conseguinte, que o processo que está sendo analisado mantém a mesma estrutura para todos os horizontes de previsão, o que nem sempre é verdade;
- as previsões obtidas tendem à média do processo para horizontes distantes, em todos os modelos propostos por Box & Jenkins, indicando uma importante limitação do método;
- apresenta custo operacional alto, quando comparado com outros métodos de Análise de Séries Temporais.

Entretanto a experiência tem mostrado, cada vez mais, a importância desse método. As previsões assim obtidas são, em geral, melhores do que as proporcionadas pela maioria dos métodos baseados em modelos econométricos correntes de ajustamento.

Nesse contexto, é necessário salientar a grande preocupação dos pesquisadores de séries de tempo em estabelecer os limites de conveniência para a utilização de cada metodologia de Análise de Séries Temporais.

Na aplicação prática da metodologia às duas séries analisadas neste documento, podem-se verificar algumas características do processo de "modelização" das séries temporais e constatar as potencialidades do método através da obtenção de previsões bastante precisas.

---

[2] Essa restrição e a anterior foram as mais criticadas no decorrer da década passada. Como conseqüência, vários autores propuseram formas alternativas de identificação de modelos, entre as quais se destacam os procedimentos de CLEVELAND (1972): Função de autocorrelação inversa; de AKAIKE (1973): Minimum Akaike Information Criterium Estimation (MAICE); de PARZEN (1976): Criterion Autoregressive Transferfunction (CAT); de GRAY, KELLEY & McINTIRE (1978); de NERLOVE, GRETHER & CARVALHO (1979). Para detalhes sobre estes procedimentos, ver MORETTIN & TOLOI (1981).

Para ambas as variáveis, foram escolhidas amostras de tamanho suficientemente grande para cumprir as exigências de precisão do método, apresentando previsões acuradas para os horizontes que, até o momento, são passíveis de confrontação com a realidade.

Embora as séries tenham sido modeladas por processos diversos, a primeira por um ARIMA (p,d,q) e a segunda por um multiplicativo SARIMA (p,d,q) x (P,D,Q), esse fato não implicou graus de precisão significativamente distintos. A diferença entre esses dois processos reside somente no fato de o modelo multiplicativo apresentar maior dificuldade operacional em seu processo de modelagem, à medida que as ACF e PACF devem descrever maior número de movimentos da série. Em geral, a etapa de identificação de um processo SARIMA (p,d,q) x (P,D,Q) sugere um número bem maior de modelos capazes de serem descritos pela série realmente observada do que a do processo simples ARIMA (p,d,q), o que a torna passível de maior probabilidade de um ajustamento não adequado, além de apresentar maior custo operacional.

Os modelos descrevem, de modo simples, o relacionamento estatístico que existe entre os dados observados, não requerendo um conhecimento "a priori" de como a série é gerada, permitindo que as observações "falem por si mesmas" e gerem previsões para um horizonte tão distante quanto se queira. Para cada um dos valores previstos, são estimados limites de confiança que, no fundo, tornam preferencial o seu uso em estudos de curto prazo, à medida que, quanto mais o futuro se afasta do presente, maior é o intervalo de confiança e menor a precisão das estimativas.

O método oferece um meio essencialmente lógico de descrição dos dados, onde a inerente subjetividade existente em alguns modelos econométricos não é requerida, proporcionando meios de previsão apropriados para as situações específicas examinadas.

Fundamentalmente, o desempenho de um determinado método aplicado à análise de uma série temporal depende de um conjunto de condições, que irá proporcionar os meios de avaliação da acuidade e conveniência do método. Em geral, esse conjunto de condições envolve, entre outros fatores:

- o modelo ao qual a série obedece;
- a habilidade do pesquisador em identificar e ajustar corretamente o modelo;
- o critério escolhido para medir a precisão das previsões;
- a presença de componentes periódicas na série;
- o tamanho da série;
- a variabilidade da componente aleatória em comparação com a variabilidade de outras componentes constitutivas da série;
- a utilização de variáveis transformadas;
- o horizonte de previsão desejado.

Vários autores têm abordado o problema da escolha do método mais indicado a cada série observada, procurando expor de forma comparativa a utilidade de cada um. Destacam-se MORETTIN & TOLOI (1981) que, além de seus próprios estudos, expõem uma síntese de alguns trabalhos representativos sobre comparação de resultados obtidos pelo emprego de diferentes métodos de análise de série temporal. Entre esses estudos, incluem-se os desenvolvidos pelos autores a seguir relacionados.

- KIRBY (1966), comparando os métodos Médias Móveis, Alisamento Exponencial e Regressão (linear), concluiu que, para previsões a um passo adiante, o melhor método (pelo critério de menor percentagem de erro médio absoluto[3]) é o de Alisamento Exponencial. Para previsões de horizonte igual a seis períodos, os métodos de Alisamento Exponencial e Médias Móveis apresentam acurácia semelhante e para previsão de longo prazo, h ≥ 12, o modelo de regressão é o mais indicado.

- GRANGER & NEWBOLD (1974), comparando especificamente métodos de Alisamento e modelos ARMA, concluíram que, para a maioria das séries analisadas, as previsões de Box & Jenkins proporcionam melhores resultados do que as obtidas pelos procedimentos automáticos, tais como os métodos de Ajustamento Exponencial de Holt & Winters e Regressão "Stepwise".

- REID (1969) obteve conclusões semelhantes às de Granger & Newbold para as séries que analisou. Em sua quase totalidade, as séries apresentaram resultados mais satisfatórios através do método de Box & Jenkins do que no de Alisamento Exponencial de Brown, mesmo naqueles casos em que esse é modificado para erros serialmente correlacionados.

- GROFF (1973), ao contrário dos outros autores, chegou a conclusões diferentes: "Para a maioria das séries, os erros de previsão do melhor modelo de Box & Jenkins são iguais ou maiores do que os erros do correspondente modelo de Alisamento Exponencial".

- GEURTZ & IBRAHIM (1975), conforme exemplificam MAKRIDAKIS & HIBON (1979), examinando uma única série, chegaram à conclusão de que tanto os modelos de Alisamento Exponencial, como o de Brown e os de Box & Jenkins apresentam resultados igualmente satisfatórios.

- MAKRIDAKIS & WEELWRIGHT (1977) concluíram que não se pode optar indiscriminadamente por um único método, uma vez que a melhor metodologia varia não só de uma série a outra, como também depende do horizonte de previsão que se deseja.

- MAKRIDAKIS & HIBON (1979) apresentam conclusões semelhantes aos anteriores. O melhor método depende não só do critério de precisão que se requer, mas também se o objetivo é medi-lo na fase de ajustamento ou de previsão. Porém, dão ênfase ao fato de obterem bons resultados utilizando métodos simples ao invés dos métodos não automáticos.

- MORETTIN & TOLOI (1981), comparando os métodos de Alisamento Exponencial Simples, Alisamento de Holt & Winters, Alisamento Exponencial Geral, Regressão Stepwise, Bayesiano e Box & Jenkins aplicados a um conjunto de dez séries temporais, concluíram que o método de Box & Jenkins apresenta um desempenho geral melhor que os demais, exceto para previsões de horizonte distante, onde a regressão mostrou ser mais adequada.

Portanto, conforme se pode verificar por essa síntese de estudos comparativos, embora a metodologia de Box & Jenkins seja de indubitável valor, não se pode esquecer a existência de outros métodos de Análise de Séries Temporais, os quais podem proporcionar previsões de boa qualidade. Em certos casos, o método de Box & Jenkins poderá ser inaceitável, extenso e com resultados inferiores aos dos outros métodos.

---

[3] Uma exposição sobre este critério de escolha de método encontra-se em MORETTIN & TOLOI (1981).

Ao longo do desenvolvimento deste documento, ainda que não se tenha testado na prática outros métodos de Análise de Séries Temporais, pode-se constatar a boa qualidade e a precisão das previsões obtidas no estudo das variáveis Índice de Preços ao Consumidor para a Cidade de Porto Alegre e Consumo de Energia Elétrica no Rio Grande do Sul, através da aplicação da metodologia de Box & Jenkins.

As séries foram escolhidas com tamanhos compatíveis com as exigências do modelo e proporcionaram valores previstos para os primeiros horizontes que já estão disponíveis, extremamente próximos à realidade: para o IPC-Porto Alegre, o percentual de erro entre o valor previsto e o valor real foi de 0,79% para o horizonte 1 e de 0,69% para o horizonte 2.

Na aplicação à série Consumo de Energia Elétrica no Rio Grande do Sul, cujo último valor real disponível se refere ao horizonte 3, os percentuais de erro foram 1,1%, 3,1% e 2,3% para os horizontes 1, 2 e 3, respectivamente, comprovando, nesses exemplos, o excelente desempenho da metodologia analisada, pelo menos para horizontes pequenos.

ANEXO I

Manual dos Comandos Necessários para a Operação das Rotinas Computacionais

Identificação

READ Statement Explanations for Univariate Analysis
Note: All numbers should be right-justified (i.c., as far to the right as possible) within the columns allowed unless otherwise indicated.
cc = card column.

1. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

2. 1 2 3 4 5

3. 1 2 3 4 5

   cc 1-5 No. of observations in time series (NOB).

4. (1 2 3 4 5 ..... ??)

   cc 1 Left parenthesis
   cc 2-?? Regular FORTRAN format describing typed cards in 5. below.
   cc ??+1 Right parenthesis.

5. Block of one or more data cards with observed series typed on them, following the format given in 4. above.

6. One card with title describing time series. Title may be typed anywhere within the 80 columns on the card.

7. 1 2 3 4 5

   cc 1-5. Enter a zero if wish to suppress listing of time series data on output (ILDID).

8. 1 2 3 4 5

   cc 1-5 Enter a zero if wish to suppress plot of time series data on output (IPDID).

9. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

   cc 1-8 Number with four decimals (TLAMID). 1.0000 if no data transformation desired. .0000 if want natural log transformation. xxx.xxxx if want power transformation with exponent xxx.xxxx.

cc 9-16 Number with four decimals (TMID).
.0000 if no data transformation desired.
xxx.xxxx if want the amount xxx.xxxx added to each time series observation before data transformation.

10.
1 2 3 4 5

cc 1-5 No. of differencing factors (or types). An autocorrelation function is calculated for the original series and each requested difference of type 1. Differencing factors beyond the first are used to difference the original series to obtain a new "original series" (NDIFID).

11. (Card required only if entry in 10. above is non-zero.)

1 2 3 4 5 . . . . . ??

cc 1-5 No. of differences of first type (NDID(1)).
cc 6-10 Order of differences of first type (IODID(1)).

Repeat in cc 11-15, 16-20, etc., for number of different types of differences indicated in 10. above.

13.
1 2 3 4 5 . . . . . 27 28 29 30

cc 1-5 Maximum lag desired in calculating sample autocorrelations (NAC).
cc 6-10 Maximum lag desired in calculating sample partial autocorrelations (NPAC). (Maximum is NAC.)
cc 11-15 No. of autocorrelations to be printed per line on output (NAPL). ($1 \leq$ NAPL ≤ 12.)
cc 16-20 No. of autocorrelations to be used in calculating chi-square statistic to test zero-correlation hypothesis. Enter a zero to suppress statistic (NCHI). (Maximum is NAC.)
cc 21-25 Enter a zero to suppress calculation of standard errors of individual autocorrelations (MCSE).
cc 26-30 Enter a zero to suppress plotting of autocorrelations calculated (IWTPA).

Estimação

READ Statement Explanations for Univariate Analysis
Note: All numbers should be right-justified (i.e., as far to the right as possible) within the columns allowed unless otherwise indicated.
cc = card column.
Note: The values of variables NMODEL, IEYON, IFYON are set in a DATA statement early in the program. You may change the values by changing the numbers between the slashes in the DATA statement. Initially all are set equal to one. These variables represent the following:

NMODEL — Number of models to be estimated-checked and/or forecast for a given time series.

IEYON — Enter a zero if wish to suppress estimation-checking stages (i.e., do only forecasting).

IFYON — Enter a zero if wish to suppress forecasting stage (i.e., do only estimation-checking).

1.
1 2 3 4 5

2.
1 2 3 4 5

cc 1-5 Number of observations in time series (NOB).

3. (. . . . .)
1 2 3 4 5 ??

cc 1 Left parenthesis
cc 2-?? Regular FORTRAN format describing typed cards in 4. below
cc ??+1 Right parenthesis

4. Block of one or more data cards with observed series typed on them following the format given in 3. above.

5. One card with title describing time series. Title may be typed anywhere within the 60 columns on the card.

Repeat 6. through 1.2. for each of the models to be examined (NMODEL times).

6.
1 2 3 4 5

cc 1-5 Number of values of the time series data transformation parameter λ to be considered. Enter zero if wish analysis only in terms of original time series data (NLAM).

7. (Card required only if entry in 6. above is non-zero.)

1 2 3 4 5 6 7 8 . . . . . ??

cc 1-8 Number with four decimals (TM (1)). This number to be added to each time series observation prior to any data transformation.
cc 9-16 Number with four decimals (PLAM (1)). First value of transformation parameter λ to be considered.
1.0000 means no real transformation.
.0000 means desire natural log transformation.
xxx.xxxx means desire power transformation with exponent xxx.xxxx.

Repeat in cc 17-24, 25-32, etc. for number of different values of λ indicated in 6. above.

8.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

cc 1-5 Number of autoregressive factors in the univariate time series model (MFAC (1,1)).
cc 6-10 Number of differencing factors in the univariate time series model (MFAC (2,1)).
cc 11-15 Number of moving average factors in the univariate time series model (MFAC (3,1)).

9. (Card required only if entry in cc 6-10 of 8. above is non-zero.)

1 2 3 4 5 . . . . . ??

cc 1-5 Number of differences of first type in the univariate time series model (ND (1,1)).
cc 6-10 Order of differences of first type in the univariate time series model (IOD (1,1)).

Repeat in cc 11-15, 16-20, etc. for number of different types of differences indicated in cc 6-10 in 8. above.

12. (Card required only if entry in cc. 6-10 of 8. above is zero).

1 2 3 4 5 6 7 8

cc 1-8 Enter number with four decimals, xxx.xxxx, the initial estimate of the mean of your time series (AVEPA(1)).

13.
1 2 3 4 5 . . . . . ??

In cc 1-5, 6-10, 11-15, etc., enter the number of individual parameters in each autoregressive factor, and then the number of individual parameters in each moving average factor. Total number of values entered on this card should equal the sum of values in cc 1-5 and cc 11-15 in 9. above. If number of values is zero, delete card (INC).

14.
1 2 3 4 5 . . . . . ??

In cc 1-5, 6-10, 11-15, etc., for each individual autoregressive or moving average parameter in your univariate time series model (first autoregressive, then moving average), enter the "order" of the parameter. "Order" is the power of the B operator which multiplies

the parameter in the model equation. Total number of values entered on this card should equal the sum of values in 13. above. If number of values is zero, delete card (IOPA).

15.
1 2 3 4 5 6 7 8 . . . . . ??

In cc 1-8, 9-16, 17-24, etc., for each individual autoregressive or moving average parameter in your univariate time series model (first autoregressive then moving average), enter an initial estimate for the parameter. Enter each with four decimals, xxx.xxxx. Total number of values entered on this card should equal the sum of values in 13. above. If number of values is zero, delete card (UPA).

16.
1 2 3 4 5

cc 1-5 Enter a zero if the univariate time series model does not contain a trend parameter (ITREND(1)).

17- (Card required only if entry in 16. above is non-zero).

1 2 3 4 5 6 7 8

cc 1-8 Enter number with four decimals, xxx.xxxx, the initial estimate of the trend parameter in the univariate time series model.

37.
1 2 3 4 5

cc 1-5 Enter a zero to supress the use of the backforecasting procedure in the estimation of univariate time series model parameters (see [2], pp. 215-220) (IWBF).

38.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

cc 1-8 If relative change in residual sum of squares less than this value, the iterative change in parameter values stops, at presumed optimum. Number with four decimals,.xxxx (EPSI).
cc 9-16 If relative change in each individual parameter less than this value, the iterative change in parameter values stops, at presumed optimum. Number with four decimals, .xxxx (EPS2).

39.
1 2 3 4 5 6 7 8 9 10

cc 1-5 Maximum number of iterations to be allowed for parameter estimation process (MIT).
cc 6-10 Enter a zero to suppress the plotting of the univariate time series model residuals (IPRES).

40.
1 2 3 4 5

cc 1-5 Enter a zero if wish to suppress listing of observations in your time series on output (ILDEST(1)).

41.
1 2 3 4 5

cc 1-5 Enter a zero if wish to suppress plotting of observations in your time series on output (IPDEST(1)).

42.
1 2 3 4 5 . . . . . 27 28 29 30

cc 1-5 Maximum lag desired in calculating sample autocorrelations of the time series model residuals (NAC).
cc 6-10 Maximum lag desired in calculating sample partial autocorrelation of the time series model residuals (NPAC). (Maximum is NAC)
cc 11-15 No. of autocorrelations to be printed per line on output (NAPL). ($1 \leq$ NAPL ≤ 12.)

cc 16-20 No. of autocorrelations to be used in calculating chi-square statistics to test zero-correlation hypothesis. Enter a zero to suppress statistic (NCHI). (Maximum is NAC.)
cc 21-25 Enter a zero to suppress calculation of standard errors of individual autocorrelations (MCSE).
cc 26-30 Enter a zero to suppress plotting of the autocorrelations calculated (IWTPA).

44.
1 2 3 4 5 . . . . . 22 23 24 25

cc 1-5 No. of forecasts desired at each time origin (NF).
cc 6-10 No. of forecast time origins desired (NTO).
cc 11-15 No. of new observations (beyond the original NOB used in estimation) to be read in to produce updated forecasts (NU).
cc 16-20 Enter 1,2,3,4, or 5, for 50,75,90,95,99 percent confidence limits, respectively, on forecasts (ICI).
cc 21-25 Enter a zero to suppress plotting of forecasts at all time origins (IWTPF).

45.
1 2 3 4 5 . . . . . ??

cc 1-5 First time origin for forecasts (NT(1)). Repeat in cc 6-10, 11-15, etc., for number of forecast time origins indicated in cc 6-10 in 44. above. (The only "true forecasting origin" is equal to the value of NOB from 2. above.)

48. (Cards required only if entry in cc. 11-15 in 44. above is non-zero.)

Block of one or more data cards with new observations for time series (beyond the original NOB) typed on them, following the format supplied in 3. above. Used to determine updated forecasts.

NOTE: Cards 38. through 42. not required if IEYON=0.

NOTE: Cards 44., 45., and 48. not required if IFYON=0.

ANEXO II

Relatórios de Saída do Computador: Previsões para o Índice de Preços ao Consumidor, Usando Diversos Modelos Supostos na Fase de Identificação

TIME SERIES FORECASTING FOR MODEL 1

DATA = Z = MODELO ARIMA(1,2,0)DADOS TRANSFORMADOS

DIFFERENCING ON Z = 1) 1 OF ORDER 1 2) 1 OF ORDER 1

MODEL DEVELOPED WITH TRANSFORMED DATA = CON(Z(T)*0.)

UNIVARIATE MODEL PARAMETERS

PARAMETER NUMBER	PARAMETER TYPE	PARAMETER ORDER	BEGINNING VALUE
1	AUTOREGRESSIVE 1	1	-.3068J2E+00

NUMBER OF TIME ORIGINS FOR FORECASTS = 1

NUMBER OF FORECASTS AT EACH TIME ORIGIN = 12

FORECAST TIME ORIGINS ARE T= 84

REGULAR FORECAST RESULTS IN TERMS OF TRANSFORMED DATA

MODEL 1 FORECASTS AT BASE PERIOD 84 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD	LO. CONF. LIMIT	FORECAST	UP. CONF. LIMIT
1	.7749268E+01	.7934737E+01	.8119407E+01
2	.7700382E+01	.7935446E+01	.8170510E+01
3	.7774950E+01	.7996133E+01	.8217316E+01
4	.7777066E+01	.7996394E+01	.8215722E+01
5	.7775328E+01	.7996770E+01	.8218212E+01
6	.7767957E+01	.8030488E+01	.8293019E+01
7	.7756094E+01	.8089265E+01	.8422436E+01
8	.7740049E+01	.8140042E+01	.8539980E+01
9	.7720067E+01	.8170820E+01	.8621573E+01
10	.7696355E+01	.8241598E+01	.8786840E+01
11	.7669092E+01	.8292375E+01	.8915658E+01
12	.7638432E+01	.8343153E+01	.9047873E+01

REGULAR FORECAST RESULTS IN TERMS OF THE ORIGINAL DATA

MODEL 1 FORECASTS AT BASE PERIOD 84 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD	LO. CONF. LIMIT	FORECAST	UP. CONF. LIMIT
1	.2319055E+04	.2402760E+04	.2488028E+04
2	.2359919E+04	.2522663E+04	.2702474E+04
3	.2360639E+04	.2660138E+04	.2972650E+04
4	.2367182E+04	.2798779E+04	.3281344E+04
5	.2381125E+04	.2944539E+04	.3641267E+04
6	.2363037E+04	.3097925E+04	.4066327E+04
7	.2333762E+04	.3259290E+04	.4554797E+04
8	.2290585E+04	.3429004E+04	.5115352E+04
9	.2253410E+04	.3607079E+04	.5777016E+04
10	.2200314E+04	.3795599E+04	.6554730E+04
11	.2141137E+04	.3993300E+04	.7448963E+04
12	.2076486E+04	.4201315E+04	.8504442E+04

TIME SERIES FORECASTING FOR MODEL 1

****

DATA - Z = MODELO ARIMA(2,2,0) DADOS TRANSFORMADOS

DIFFERENCING ON Z - 1) 1 OF ORDER 1 2) 1 OF ORDER 1

MODEL DEVELOPED WITH TRANSFORMED DATA = LOG(Z(T)+0.)

****

UNIVARIATE MODEL PARAMETERS

****

PARAMETER NUMBER	PARAMETER TYPE	PARAMETER ORDER	BEGINNING VALUE
1	AUTOREGRESSIVE 1	1	-.38239E+00
2	AUTOREGRESSIVE 1	2	-.24469E+00

****

NUMBER OF TIME ORIGINS FOR FORECASTS = 1

NUMBER OF FORECASTS AT EACH TIME ORIGIN = 12

FORECAST TIME ORIGINS ARE T= 84

REGULAR FORECAST RESULTS IN TERMS OF TRANSFORMED DATA

****

MODEL 1 FORECASTS AT BASE PERIOD 84 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD	LO. CONF. LIMIT	FORECAST	UP. CONF. LIMIT
1	.7747140E+01	.7781450E+01	.7815760E+01
2	.7762498E+01	.7827746E+01	.7892995E+01
3	.7777854E+01	.7875992E+01	.7974129E+01
4	.7786621E+01	.7923637E+01	.8060654E+01
5	.7790839E+01	.7971035E+01	.8151231E+01
6	.7791895E+01	.8018674E+01	.8245454E+01
7	.7789399E+01	.8066262E+01	.8343165E+01
8	.7783587E+01	.8113843E+01	.8444099E+01
9	.7774790E+01	.8161429E+01	.8548069E+01
10	.7763110E+01	.8209017E+01	.8654925E+01
11	.7748067E+01	.8256598E+01	.8764530E+01
12	.7731596E+01	.8304182E+01	.8876767E+01

REGULAR FORECAST RESULTS IN TERMS OF THE ORIGINAL DATA

****

MODEL 1 FORECASTS AT BASE PERIOD 84 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD	LO. CONF. LIMIT	FORECAST	UP. CONF. LIMIT
1	.2314943E+04	.2395746E+04	.2479370E+04
2	.2350769E+04	.2509268E+04	.2676454E+04
3	.2387147E+04	.2633296E+04	.2904026E+04
4	.2406167E+04	.2761798E+04	.3167360E+04
5	.2416347E+04	.2895853E+04	.3467044E+04
6	.2420700E+04	.3037149E+04	.3810265E+04
7	.2414066E+04	.3185238E+04	.4201367E+04
8	.2400870E+04	.3340390E+04	.4647967E+04
9	.2379843E+04	.3503190E+04	.5156786E+04
10	.2352208E+04	.3673930E+04	.5735339E+04
11	.2318479E+04	.3852966E+04	.6403052E+04
12	.2279238E+04	.4040735E+04	.7163594E+04

TIME SERIES FORECASTING FOR MODEL 1

****

DATA = Z = MODELO ARIMA(0,2,2) DADOS TRANSFORMADOS

DIFFERENCING ON Z = 1) 1 OF ORDER 1 2) 1 OF ORDER 1

MODEL DEVELOPED WITH TRANSFORMED DATA = LOG(Z(T)+0.)

****

UNIVARIATE MODEL PARAMETERS

****

PARAMETER NUMBER	PARAMETER TYPE	PARAMETER ORDER	BEGINNING VALUE
1	MOVING AVERAGE 1	1	.6001[illegible]E+00
2	MOVING AVERAGE 1	2	.2087[illegible]E+00

****

NUMBER OF TIME ORIGINS FOR FORECASTS = 1

NUMBER OF FORECASTS AT EACH TIME ORIGIN = 12

FORECAST TIME ORIGINS ARE T= 84

REGULAR FORECAST RESULTS IN TERMS OF TRANSFORMED DATA

****

MODEL 1 FORECASTS AT BASE PERIOD 84 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD	LO. CONF. LIMIT	FORECAST	UP. CONF. LIMIT
1	.7730[illegible]E+01	.7783[illegible]E+01	.7[illegible]E+01
2	.7790187E+01	.7843476E+01	.7896765E+01
3	.7825982E+01	.7897667E+01	.7969[illegible]E+01
4	.7866966E+01	.7951658E+01	.8036[illegible]E+01
5	.7898[illegible]E+01	.8000[illegible]E+01	.8[illegible]E+01
6	.7931603E+01	.8002402E+01	.8073[illegible]E+01
7	.7960341E+01	.8031[illegible]E+01	.8102[illegible]E+01
8	.8001907E+01	.8158[illegible]E+01	.8[illegible]E+01
9	.8034964E+01	.8242813E+01	.8[illegible]E+01
10	.8060[illegible]E+01	.8277004E+01	.8[illegible]E+01
11	.8101481E+01	.8311195E+01	.8[illegible]E+01
12	.8133630E+01	.8365386E+01	.8[illegible]E+01

REGULAR FORECAST RESULTS IN TERMS OF THE ORIGINAL DATA

****

MODEL 1 FORECASTS AT BASE PERIOD 84 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD	LO. CONF. LIMIT	FORECAST	UP. CONF. LIMIT
1	.2340[illegible]E+04	.2414590E+04	.2495[illegible]E+04
2	.2416769E+04	.2549049E+04	.2688[illegible]E+04
3	.2506[illegible]E+04	.2670999E+04	.2[illegible]E+04
4	.2593247E+04	.2804847E+04	.3[illegible]E+04
5	.2687[illegible]E+04	.2999043E+04	.3347[illegible]E+04
6	.2783887E+04	.3156049E+04	.3[illegible]E+04
7	.2886[illegible]E+04	.3342394E+04	.3[illegible]E+04
8	.2983[illegible]E+04	.3526477E+04	.4[illegible]E+04
9	.3087[illegible]E+04	.3744965E+04	.4[illegible]E+04
10	.3191964E+04	.3952394E+04	.4[illegible]E+04
11	.3298692E+04	.4151374E+04	.5[illegible]E+04
12	.3407144E+04	.4392546E+04	.5[illegible]E+04

TIME SERIES FORECASTING FOR MODEL 1

DATA = Z = MODELO ARIMA(1,0,1) DADOS TRANSFORMADOS

DIFFERENCING ON Z = 1) 1 OF ORDER 1 2) 1 OF ORDER 1

MODEL DEVELOPED WITH TRANSFORMED DATA = CON+ Z(T)+0.

UNIVARIATE MODEL PARAMETERS

PARAMETER NUMBER	PARAMETER TYPE	PARAMETER ORDER	BEGINNING VALUE
1	AUTOREGRESSIVE 1	1	.2430[illegible]E+00
2	MOVING AVERAGE 1	1	.8552[illegible]E+00

NUMBER OF TIME ORIGINS FOR FORECASTS = 1

NUMBER OF FORECASTS AT EACH TIME ORIGIN = 12

FORECAST TIME ORIGINS ARE T= 84

BACKFORECASTING WAS SUPPRESSED IN THIS ANALYSIS

REGULAR FORECAST RESULTS IN TERMS OF TRANSFORMED DATA

MODEL 1 FORECASTS AT BASE PERIOD 84 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD	LO. CONF. LIMIT	FORECAST	UP. CONF. LIMIT
1	.7757373E+01	.7798793E+01	.7820613E+01
2	.7709352E+01	.7843103E+01	.7896555E+01
3	.7823252E+01	.7997432E+01	.7971613E+01
4	.7857716E+01	.7951766E+01	.8046166E+01
5	.7892741E+01	.8006101E+01	.8119460E+01
6	.7927412E+01	.8060435E+01	.8193459E+01
7	.7961760E+01	.8114770E+01	.8267781E+01
8	.7995089E+01	.8169105E+01	.8342521E+01
9	.8029148E+01	.8223440E+01	.8417733E+01
10	.8062105E+01	.8277775E+01	.8493446E+01
11	.8094345E+01	.8332110E+01	.8569676E+01
12	.8126459E+01	.8386445E+01	.8646431E+01

REGULAR FORECAST RESULTS IN TERMS OF THE ORIGINAL DATA

MODEL 1 FORECASTS AT BASE PERIOD 84 WITH 95 PER CENT CONFIDENCE LIMITS

PERIODS AHEAD	LO. CONF. LIMIT	FORECAST	UP. CONF. LIMIT
1	.2338754E+04	.2413403E+04	.2490435E+04
2	.2414752E+04	.2548101E+04	.2688013E+04
3	.2498017E+04	.2690366E+04	.2897226E+04
4	.2586125E+04	.2840587E+04	.3129087E+04
5	.2677775E+04	.2999199E+04	.3359208E+04
6	.2772243E+04	.3166669E+04	.3617212E+04
7	.2869117E+04	.3343490E+04	.3896294E+04
8	.2968136E+04	.3530184E+04	.4196063E+04
9	.3069125E+04	.3727303E+04	.4520029E+04
10	.3171960E+04	.3935429E+04	.4868260E+04
11	.3276244E+04	.4155177E+04	.5241421E+04
12	.3382799E+04	.4387194E+04	.5640636E+04

ABSTRACT

The present dissertation is concerned with the method developed by George E. P. Box and Gwilym M. Jenkins.

That methodology was developed during the sixties and published by the authors in 1970 under the title "Time series analysis: forecasting and control".

The significance of Box and Jenkins' work is given by the possibility of constructing linear models of the autoregressive integrated moving average type that could describe with precision and parsimony the stochastic process that generates a time series, offering therefore accurate forecasts.

The strategy in that methodology consists in identifying models able to describe the data available, to efficiently estimate its parameters and to evaluate the model "vis-à-vis" the phenomenon being investigated. If this representation of the real world is adequate, the model can be applied in time series analysis or, on the contrary, if that representation is not adequate, a new model should be designed and a new round of tests run.

The present dissertation is structured in seven chapters.

Chapter I consists of not only the general question being posed and our concern about it justified, but we also develop the fundamental ideas to deal with time series analysis: a short history of the subject, the importance and the statiscal meaning of the technique and the concept proper of time series.

In the chapter II we develop the basic concepts of Stochastic Process that will be useful for the understanding and working of Box and Jenkins' methodology.

In chapters III, IV and V we deal in detail with Box and Jenkins' time series analysis: the models they use, the process of modeling and the method for forecasting.

In chapter VI, we apply the proceding developments to two sets of data, which are significant to the understanding of Rio Grande do Sul economy: the Consumer Price Index (cost of living index) in the city of Porto Alegre and Electric Energy Comsumption in the State. It can be observed that the forecasts obtained through Box and Jenkins' method show values very much close to the real data, making it clear about the usefulness of the author's methodology.

The last chapter is a summary of the method discussed. We call attention to its qualities but also to its operational shortcomings and we refer to some other methods for time series analysis, trying to locate aspects for applying each one of them.

BIBLIOGRAFIA

AKAIKE, H. A new look at the statistical model identification. IEEE Transactions on Automatic Control, p. 716-23, 1974.

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. Second Internacional Symposium on Information Theory, Budapest, Akademia Kaido, 1973.

ANDERSON, O. D. Time series analysis and forecasting: the Box-Jenkins approach. London, Butterworths, 1979.

ANDERSON, R. L. Distribution of the serial correlation coefficient. Ann. Math. Statist., Baltimore, S.S.Wilks, 13:1-13, Mar. 1942.

ANDERSON, T. W. The statistical analysis of time series. New York, J. Wiley, 1971.

BAILEY, Norman T. The elements of stochastic processes with applications to the natural sciences. New York, J. Wiley, 1964.

BARBER, William J. Uma história do pensamento econômico. Rio de Janeiro, Zahar, 1979.

BELL, John Fred. História do pensamento econômico. Rio de Janeiro, Zahar, 1961.

BOX, G. E. P. & COX, D. R. An analysis of transformations. Journal of the Royal Statistical Society, London, B26:211-252, 1964.

BOX, George E. P. & JENKINS, Gwilym M. Time series analysis: forecasting and control. San Francisco, Holden-Day, 1976.

BOX, G. E. P. & PIERCE, D. A. Distribution of residual autocorrelations in autoregressive integrated moving average time series models. Journal of the American Statistical Association, Washington, 65:1509-26, Dec.1970.

BUNGE, Mário. Teoria e realidade. São Paulo, Perspectiva, 1974.

CHAO, Lincoln L. Estadística para las ciencias administrativas. Bogotá, McGraw-Hill, 1975.

CHURCHMAN, C. W. Introdução à teoria dos sistemas. Petrópolis, Vozes, 1972.

CLARKE, A. Bruce & DISNEY, Ralph, L. Probability and random processes for engineers and scientists. New York, J. Wiley, 1970.

CLEVELAND, W. P. The inverse autocorrelations of a time series and their applications. Technometrics, 14(2):277-98, 1972.

DAGUM, Camilo & DAGUM, Estela M. Bee. Introducción a la econometría. México, Siglo Veintiuno, 1971.

ECHENIQUE, Marcial et alli. Modelos matemáticos de la estructura espacial urbana: aplicaciones en América Latina. Buenos Aires, SIAP, 1975.

ECONOMETRIC SOFTWARE PACKAGE. Washington, Synergy Inc., University of Chicago Version, 1974.

FERGUSON, C. E. Teoría microeconômica. México, Fondo de Cultura Econômica, 1974.

FERNANDEZ, Pedro Jesus. Introdução aos processos estocásticos. s.l.p., IMPA, s.d.

FOX, Karl A. Intermediate economic statistics. New York, J. Wiley, 1968.

FREUND, John E. & WILLIAMS, Frank J. Modern business statistics. Englewood Cliffs, Prentice-Hall, 1958.

GNEDENKO, B. V. The theory of probability. Moscow, MIR, 1969.

GRANGER, C. W. J. & NEWBOLD, J. P. Experience with forecasting univariate time series and the combination of forecasting. Journal of the Royal Statistical Society, London, A137: 131-46, Jan. 1974.

GRANGER, C. W. J. & NEWBOLD, Paul. Forecasting economic time series. New York, Academic Press, 1977.

GRAY, H. L. et alli. A new approach to ARMA modeling. Communications in Statistics. 87(1): 1-78, 1978.

GROFF, G. K. Empirical comparison of models for short range forecasting. Management Science, Baltimore, The Institute of Management Sciences, 20(1): 22-31, Sep. 1973.

GEURTZ, M. D. & IBRAHIM, J. B. Comparing the Box-Jenkins approach with exponentially smooted forecasting model application to Hawaii tourists. Journal of Marketing Research, Chicago, American Marketing Association, 12: 182-88, May 1975.

HADLEY, G. Introduction to business statistics. San Francisco, Holden-Day, 1968.

HAMBURG, Morris. Statistical analysis for decision Making, New York, Harcourt, Brace and World, 1970.

HOFFMANN, Rodolfo & VIEIRA Sônia. Análise de regressão: uma introdução à econometria. São Paulo, HUCITEC, 1977.

JOHNSTON, J. Métodos econométricos. São Paulo, Atlas, 1871.

KENDALL, Maurice and STUART, Alan. The advanced theory of statistics. London, Charles Griffin, 1976, v.3.

KEYNES, John M. Teoria geral do emprego, juros e moeda. Rio de Janeiro, Fundo de Cultura, 1964.

KIRBY, R. M. A comparison of short and medium range statistical forecasting methods. Management Science, Baltimore, The Institute of Management Sciences, 13(4). 1966.

KLEIN, Lawrence L. Manual de Econometría. Madrid, Aguilar, 1958.

KRUPP, Sherman Roy. La estructura de la ciencia economica; ensaios sobre metodologia. Madrid, Aguilar, 1973.

LAJUGIE, J. As doutrinas econômicas. São Paulo, Difel, 1976.

MADDALA, G. S. Econometrics. New York, McGraw-Hill, 1977.

MAKRIDAKIS, S. & HIBON, M. Accuracy of forecasting and empirical investigation. The Journal of the Royal Statistical Society, London, 142: 97-145, 1979.

MAKRIDAKIS, S. & WHEELWRIGHT, S. C. Adaptive filtering: an integrated autoregressive moving average filter for time series forecasting. Operational Research Quarterly, 28: 425-37, 1977.

MALANOS, George. Teoria econômica. Rio de Janeiro, Forum, 1962.

MARQUARDT, D. W. An algorithm for least-squares estimations of non linear parameters. Journal Soc. Ind. Appl. Math., Philadelphia, 11 (2): 431-41, Jun. 1963.

MATTHEWS, R. C. O. O ciclo econômico. Rio de Janeiro, Zahar, 196[illegible].

MEEKER, W. Q. T-SERIES - A user-oriented computer program for identifying, fitting and forecasting ARIMA time series models. Ames, Iowa State University, s.d.

MERRIL, William C. & FOX Karl A. Introducción a la estadística económica. Buenos Aires, Amorrortu, 1972.

MILLS, Richard L. Statistics for applied economics and business. New York, McGraw-Hill, 1977.

MONTELLO, Jessé. Estatística para economistas. Rio de Janeiro, APEC, 1970.

MOOD, Alexander M. & GRAYBILL, Franklin A. Introducción la teoria de la estadística. Madrid, Aguilar, 1976.

MORETTIN, Pedro Alberto. Análise harmônica de processos estocásticos. Rio de Janeiro, IMPA, 1979.

MORETTIN, Pedro Alberto & TOLOI, Clélia Maria de Castro. Modelos para previsão de séries temporais. Rio de Janeiro, IMPA, 1981.

MORGENBESSER, Sidney et alli. Filosofia da ciência. São Paulo, Cultrix, 1972.

MURPHY, James L. Introductory econometrics. Homewood, Richard D. Irwin, 1973.

NELSON, Charles R. Applied time series analysis for managerial forecasting. San Francisco, Holden-Day, 1973.

NERLOVE, M. et alli. Analysis of economic time series; a synthesis. New York, Academic Press, 1979.

NIE, Norman H. & HULL, C. H. SPSS: Statistical package for the social sciences; update 7-9, new procedures and facilities for releases 7-9. New York, McGraw-Hill, 1981.

PACK, D. J. A computer program for the analysis of time series models using the Box-Jenkins philosophy. Columbus, The Ohio State University, College of Administrative Sciences, Data Center, 1977.

PAPOULIS, Athanasios. Probability, random variables, and stochastic processes. Tokio, McGraw-Hill Kogakusha, 1965.

PARZEN, Emanuel. An approach to time series modeling and forecasting illustrated by hourly electricity demands. Technical Report, Buffalo, State University of New York, 37, 1976.

PARZEN, Emanuel. Processos estocasticos. Madrid, Paraninfo, 1972.

PIATIER, André. Estadística y observación económica. Barcelona, Ariel, 1967, v.1.

PINDYCK, Robert S. & RUBINFELD, Daniel L. Econometric models and economic forecasts. New York, McGraw-Hill, 1976.

REID, D. J. A comparative study of time series prediction techniques on economics data. Nottingham, University of Nottingham, 1969. (Department of Mathematics, Ph.D.Thesis).

RODRIGUES, Flávio Wagner. Tópicos da teoria das probabilidades. s.l.p. IMPA, s.d.

TINTNER, Gerhard. Econometrics. New York, J. Wiley, 1963.

WOLD, H.O. A study in the analysis of stationary time series. Uppsala, Almquist and Wicksell, 1938.

WONNACOTT, Thomas H. & WONACOTT, Ronald J. Introductory statistics for business and economics. New York, J. Wiley, 1972.

ZEITTOUN, Jean. Modeles en urbanisme: une étude critique. Paris, Centre de Recherche d'Urbanisme, 1971.