

ADRIANO FERRETI BORGATTO

ANÁLISE DE INTERVENÇÃO EM SÉRIES TEMPORAIS:
APLICAÇÕES EM TRANSPORTE URBANO

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

Orientadora:
Prof.^a. Dr.^a. Thelma Sáfydi

CDD-388.4
-219.22

LAVRAS
MINAS GERAIS – BRASIL
2000

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

Borgatto, Adriano Ferreti

**Análise de intervenção em séries temporais: aplicações em transporte urbano /
Adriano Ferreti Borgatto. -- Lavras : UFLA, 2000.**

102 p. : il.

Orientadora: Thelma Sáfadi.

Dissertação (Mestrado) – UFLA.

Bibliografia.

**1. Série temporal. 2. Análise de intervenção. 3. Tendência. 4. Sazonalidade. 5.
Transporte urbano. I. Universidade Federal de Lavras. II. Título.**

CDD-388.4

-519.55

ADRIANO FERRETI BORGATTO

ANÁLISE DE INTERVENÇÃO EM SÉRIES TEMPORAIS: APLICAÇÕES EM TRANSPORTE URBANO

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

APROVADA em 17 de agosto de 2000

Prof.^a. Eliana Zandonade

UFES

Prof. Augusto Ramalho de Moraes

UFLA

Prof. Eduardo Bearzoti

UFLA


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

LAVRAS
MINAS GERAIS – BRASIL

Aos meus pais Waldyr e Ana Cleide.

Aos meus irmãos Ana Cristina e Vicente.

À minha noiva Fabiana.

AGRADECIMENTOS

A Deus por mais esta etapa vencida.

A Thelma Sáfadi pela orientação, dedicação, apoio e confiança.

Aos professores Eduardo Bearzoti, Augusto de Moraes, Daniel Furtado Ferreira, Ruben Delly, Luís Henrique, Júlio Silvio de Souza Bueno, Mario Vivanco e Lucas Chaves pelos ensinamentos, apoio e dedicação.

A Maria das Dores Correa Santos por sua dedicação.

Ao professor Joel Augusto Muniz pela grande dedicação e esforço para a melhora do curso.

A todos colegas do curso de pós-graduação, em especial a Letícia, Denis, Mônica, Andrea, Ivani, Marcelo, Carlos, Paulo José, Andréa Cristiane, Ana Rita, Cristiane, Iara, Nagib, Moisés, Éverton, Teixeira, Alex e Aladir pela convivência que tivemos, pela amizade e apoio nas horas difíceis.

À CAPES pela colaboração financeira.

SUMÁRIO

	Página
RESUMO.....	i
ABSTRACT	ii
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 Problemas do Transporte Urbano	3
2.2 Definições em Séries Temporais	5
2.3 Modelos de Decomposição	7
2.4 Testes para as Componentes T_t e S_t	9
2.4.1 Teste para Tendência (T_t).....	9
2.4.2 Teste para Sazonalidade (S_t).....	10
2.5 Modelos de Box e Jenkins	12
2.5.1 Introdução.....	12
2.5.2 Modelos ARMA	14
2.5.2.1 Condições de Estacionaridade e Invertibilidade	16
2.5.3 Função de Autocorrelação.....	17
2.5.4 Função de Autocorrelação Parcial	19
2.5.5 Identificação	21
2.5.6 Estimação	22
2.5.7 Verificação	25
2.5.7.1 Teste de Normalidade para os Resíduos.....	27
2.5.7.2 Teste de Ruído Branco	28
2.5.8 Modelos ARIMA e SARIMA.....	29
2.6 Método para Escolha do Modelo.....	32
2.7 Modelos de Intervenção.....	34
2.7.1 Definição	34
2.7.2 Modelo Dinâmico para Intervenção.....	34
2.7.3 Efeitos de Intervenção.....	38
2.7.4 Estrutura da Função de Transferência.....	40
2.7.5 Aplicações de Análise de Intervenção em Leis de Trânsito.....	43
2.7.6 Testes de Intervenção.....	45
2.7.6.1 Teste t de Student.....	46
2.7.6.2 Teste de Qui-Quadrado.....	47
3 MATERIAL E MÉTODOS	49
4 RESULTADOS E DISCUSSÃO	51

4.1 Modelos para a Série: Número de Passageiros por Ônibus.....	51
4.1.1 Identificação	54
4.1.2 Estimação	57
4.1.3 Previsão	62
4.2 Modelos para a Série: Número de Assaltos nos Ônibus	65
4.2.1 Identificação	68
4.2.2 Estimação	71
4.2.3 Previsão	75
4.3 Modelos para a Série: Número de Acidentes com os Ônibus	78
4.3.1 Identificação	81
4.3.2 Estimação	83
4.3.3 Previsão	88
5 CONCLUSÃO	91
REFERÊNCIAS BIBLIOGRÁFICAS	92
ANEXOS	95

RESUMO

BORGATTO, Adriano Ferreti. Análise de Intervenção em Séries Temporais: Aplicações em Transporte Urbano. Lavras: UFLA, 2000. 102p. (Dissertação – Mestrado em Estatística e Experimentação Agropecuária)

O presente trabalho teve como principal objetivo analisar o comportamento do transporte urbano na cidade de São Paulo, utilizando análise de intervenção em séries temporais. Atualmente, 54 empresas operam com uma frota aproximada de 10 mil ônibus, 800 linhas, utilizadas por quase 4 milhões de passageiros por dia. Foram utilizadas três séries: a primeira relata o número médio de passageiros que circulam diariamente nos ônibus urbanos do município de São Paulo, abrangendo o período de janeiro de 1983 a dezembro de 1998, com 192 observações mensais; a segunda série fornece o número de assaltos nos ônibus urbanos que circulam no município de São Paulo, com medições de janeiro de 1992 a agosto de 1999, totalizando 93 observações mensais; a terceira série descreve o número de acidentes de trânsito com os ônibus urbanos do município de São Paulo no período de janeiro de 1988 a setembro de 1999, com 141 observações mensais. Os parâmetros foram estimados pelos modelos de Box e Jenkins, acrescentando-se parâmetros para os fenômenos ocorridos nos períodos de suas medições, denominados intervenção. A identificação dos modelos foi feita através da função de autocorrelação e da função de autocorrelação parcial para a série estacionária. A verificação da adequabilidade dos modelos foi realizada através do teste de Box e Pierce. Para cada série foram propostos mais de um modelo, sendo a escolha do modelo mais adequado realizada através do critério de Akaike e no Quadrado Médio Residual. As 6 últimas observações de cada série foram desconsideradas da análise para serem comparadas com os valores de previsão propostos por cada modelo. Alguns fatores pelos quais ocorreu a intervenção estão diretamente ligados ao acréscimo do número de linhas circulando na cidade de São Paulo e pela privatização da C.M.T.C., em 1993, sendo o modelo de intervenção mais eficiente em relação ao modelo sem intervenção.

Orientadora: Thelma Sáfiadi – UFLA

ABSTRACT

BORGATTO, A.F. Intervention analysis in time series: Applications in urban transportation. Lavras: UFLA, 2000; 102p. (Dissertation – master in Statistical and Agricultural Experimentation)

The present work aimed chiefly to analyse the behavior of urban transportations in the city of São Paulo by using intervention analysis in time series. Currently, 54 companies operate with a fleet of about 10 thousand units, 800 transport lines utilized by almost 4 million passengers a day. Three series were considered; the first reports the mean daily number of passengers, encompassing the period from January of 1983 to December of 1998 with 192 month averages; the second furnishes the number of holdups in urban buses with measurements from January of 1992 to August of 1999, amounting to 93 month observations; and last series reports the number of traffic accidents with buses over the period from January of 1988 to September of 1999, with 141 month observations. Box and Jenkins models were fitted, with additions parameters to account for intervention. Identification of models was accomplished function autocorrelation and function autocorrelation partial for the stationary series. Reasonability of fitted models was evaluated by Box and Pierce test. Amongst reasonable models for each series, one was chosen according to Akaike criterion and residual mean square values. The last 6 observations of each series were disregarded from the analysis to be compared with the forecasted values of each model. Some events showed evidence significative of intervention, as the increase of the number of transport lines in the city of São Paulo and the privatization of C.M.T.C. in 1993.

Adviser: Thelma Sáfaci – UFLA

1 INTRODUÇÃO

As técnicas de análise de séries temporais vêm sendo utilizadas com frequência em várias áreas de pesquisa como a Engenharia, Economia, Ciências Sociais e Políticas, Sociologia, História, Psicologia, Meio Ambiente, entre outras, em que os fenômenos estão geralmente associados ao tempo e seus métodos evidenciam características dinâmicas.

Em estudos de dados ^{analisados no tempo} históricos, nos quais existe uma correlação entre as observações, é preferível utilizar técnicas de séries temporais a técnicas de estatísticas usuais.

Os estudos relacionados com transporte urbano são um campo em que também são apropriadas técnicas de séries temporais, ajustando alguns modelos para os dados, como, por exemplo, os propostos por Box e Jenkins (1976), com seus parâmetros estimados de forma iterativa. Um caso particular do modelo é o ARMA (Autorregressivo – Médias-Móveis), que tem a estrutura baseada nos próprios dados.

Para ajustar os modelos de Box e Jenkins, é necessário que a série seja estacionária, estando livre de tendência e sazonalidade. A tendência ocorre quando a série sofre uma inclinação, linear, quadrática ou de ordens superiores; a sazonalidade é uma periodicidade que acontece em um intervalo máximo de 12 meses, sendo o intervalo acima deste período denominado ciclo.

As etapas da metodologia empregada por Box e Jenkins são a identificação da estrutura do processo gerador da série, a estimação dos parâmetros do modelo e as previsões realizadas com todo embasamento probabilístico e com o menor erro de estimação. É importante salientar que o

principal objetivo da previsão é fornecer informações para uma consequente tomada de decisões.

Existem fenômenos, em algumas séries, que não são levados em conta pelo modelo ARMA, os quais são denominados intervenções. A sua manifestação consiste em uma mudança de nível ou inclinação dos dados num determinado instante do tempo; geralmente este fenômeno ocorre por causa de algum acontecimento conhecido.

A intervenção muitas vezes pode estar obscura por três fontes de “ruídos”: a tendência, sazonalidade e o erro aleatório, fatores que conduzem o pesquisador a falsas conclusões, pois a ocorrência de uma inclinação ou mudança de nível na série não é necessariamente uma intervenção, podendo ser confundida com uma tendência.

Este trabalho teve como principal enfoque estudar os efeitos de intervenção propondo alguns modelos de função de transferência, e como objetivo aplicar a metodologia proposta de séries temporais utilizando os fenômenos de intervenção na área de transporte urbano.

2 REFERENCIAL TEÓRICO

2.1 Problemas do Transporte Urbano

A cidade de São Paulo possui uma população em torno de 10 milhões de habitantes, e a região metropolitana, contendo 38 municípios, acomoda mais de 7 milhões de habitantes. Cerca de 55% das viagens diárias motorizadas na região metropolitana são feitas por transporte coletivo. Na cidade, os ônibus atendem cerca de 70% das viagens de transporte coletivo.

Todas as linhas de ônibus são operadas por empresas privadas, sob a supervisão da São Paulo Transporte S.A. – SPTrans, empresa municipal de planejamento e gerenciamento do transporte coletivo.

Atualmente, 54 empresas operam, com uma frota aproximada de 10 mil ônibus, 800 linhas, utilizadas por quase 4 milhões de passageiros/dia.

O sistema metropolitano de ônibus, sob responsabilidade da Empresa Metropolitana de Transportes Urbanos (E.M.T.U.), subordinada ao governo estadual, liga São Paulo às demais cidades da região, incluindo uma extensa rede de linhas operadas por empresas privadas.

As informações obtidas como referência, citadas no texto, foram fornecidas pelo sistema de transporte coletivo do município de São Paulo e pela cronologia do transporte público de São Paulo, na página <http://www.sptrans.com.br>.

Os principais eventos que ocorreram no período de 1983 a 1999 estão apresentados abaixo:

- a) 1983 – Início da integração ônibus-ferrovia, entre a linha Pinheiros-Largo São Francisco, da CMTC, e os trens metropolitanos da Fepasa, quando o transporte público volta a ser alvo da atenção da administração municipal;

- b) 1984 – Entra em funcionamento a primeira linha operadora a gás metano, Ceasa-Lapa, da CMTC;
- c) 1991 – Em Janeiro é assinada uma lei que determina a total substituição, no prazo de dez anos, da frota de ônibus urbanos movidos a diesel por ônibus a gás natural.

Em abril operam 19 empresas de ônibus de acordo com os regulamentos da municipalização.

Em junho tem início a operação da primeira linha com entradas pela porta dianteira.

Em julho surge uma lei que municipaliza o transporte coletivo dos ônibus, determina licitações para cobrir 42 lotes de áreas de operação, substituindo as 23 áreas exclusivas de operação de ônibus;

- d) 1993 – A nova administração assume e encontra a CMTC e o Sistema Municipalizado em condições precárias. O número de passageiros volta a ter um peso significativo na remuneração das empresas contratadas e a primeira fase de privatização das áreas de operação e manutenção da CMTC é iniciada;
- e) 1994 – O Sistema de Transporte Coletivo por Ônibus passa a ser operado por 47 empresas privadas;
- f) 1995 – A partir do segundo semestre do ano, é adotada a sistemática de coleta automática de dados operacionais do sistema de transporte urbano por ônibus, a Fiscalização Eletrônica;
- g) É implementado o Plano de Recuperação do Desempenho Operacional dos Corredores de Transporte Coletivo.

2.2 Definições em Séries Temporais

Partindo do princípio de que dados são analisados no tempo, pode-se estabelecer alguns fenômenos que interferem na análise, conduzindo a um problema de modelagem estatística e inferência.

É razoável admitir que se os dados são analisados no tempo, existe uma correlação entre a observação no instante t e a observação no instante $(t+h)$, onde h é um período da série temporal posterior ao tempo t , dificultando a aplicabilidade de métodos estatísticos convencionais, pois na utilização desses métodos, parte-se da pressuposição de que as observações sejam independentes e identicamente distribuídas (i.i.d.). A melhor abordagem a ser aplicada em tais casos, em que nenhuma pressuposição é violada, são as técnicas de análise de séries temporais.

A principal utilização da análise de séries temporais é a previsão, mas nem sempre é possível adequar um bom modelo às observações. Um exemplo de uma das primeiras séries registradas em que é difícil adequar um modelo é a série de “manchas solares” estudada por Schuster em 1906, com observações monitoradas anualmente. Constata-se que as manchas solares não têm um comportamento previsível, dificultando a realização da previsão.

Segundo Morettin e Toloí (1985), uma série temporal é um conjunto de observações compreendidas seqüencialmente no tempo. Muitos conjuntos de dados aparecem como séries temporais, e tais seqüências podem ser medidas mensalmente, trimestralmente, semanalmente ou em horas. Se os valores futuros de uma série temporal podem ser determinados exatamente por uma função matemática, como, por exemplo,

$$z_t = \cos(2\pi f t) \quad (1)$$

a série é denominada determinística.

O caso mais comum de série temporal é aquele em que os valores futuros podem ser descritos por uma distribuição de probabilidade, sendo as séries denominadas não-determinísticas. Neste caso, a série pode ser considerada como uma trajetória de um processo estocástico $x(t,\omega)$, em que t é o tempo da série temporal e ω é uma variável aleatória. Este processo, para t fixado $x(t,\omega)$, é uma variável aleatória, e para ω fixado $x(t,\omega)$, é uma série temporal.

As observações de uma série temporal podem ser discretas ou contínuas. Como exemplo de dados discretos, pode-se citar a medida da temperatura diária do município de São Paulo; e como exemplo de dados contínuos, o registro de marés no porto de Santos. Na análise de dados contínuos, os valores observados são divididos em intervalos para que se tornem discretos, a fim de que se possa analisar a série temporal.

Existem dois enfoques na aplicação da análise de séries temporais, sendo que o primeiro consiste na análise realizada no domínio do tempo, em que são propostos modelos paramétricos; e o segundo baseia-se no domínio da frequência, sendo sua análise realizada através de modelos não-paramétricos.

Neste trabalho, será utilizada a análise no domínio do tempo, considerando um modelo proposto por Box e Jenkins. Na análise no domínio da frequência, utiliza-se a análise espectral, que consiste na decomposição da série em componentes de frequência. Este método é citado por Koopmans (1974), Morettin (1979), Priestley (1989) e Pereira (1984).

Segundo Pereira (1984), o estudo de séries temporais no domínio da frequência supõe que a série é melhor analisada como uma soma ou superposição linear de ondas periódicas de senos e cossenos de diferentes períodos ou frequência. Sua justificativa teórica é baseada na representação espectral; essa abordagem é utilizada em Ciências Físicas e Engenharia, pois seus eventos tendem a ser gerados por fenômenos periódicos.

A abordagem no domínio do tempo geralmente pressupõe que a correlação entre as observações adjacentes é melhor explicada em termos de uma regressão. Este método de análise foi proposto por Wold (1938), seguindo uma decomposição da série em que as observações presentes pudessem ser preditas como a soma de uma combinação linear de valores passados de uma série de ruídos e um componente determinístico ortogonal à combinação linear.

2.3 Modelos de Decomposição

Na análise de uma série temporal, primeiramente é necessário fazer uma análise visual do gráfico para verificar a possibilidade da existência de alguma periodicidade, tendência e (ou) “outlier”. Após realizada esta análise, aplicam-se testes apropriados a fim de detectar a existência desses fatores na série.

Um modelo clássico para séries temporais foi proposto por Wold(1938), no qual supõe-se que a série z_1, z_2, \dots, z_n pode ser escrita através de quatro componentes existentes na série, ou seja,

$$z_t = T_t + S_t + C_t + a_t \quad t = 1, 2, \dots, N \quad (2)$$

em que

- z_t é a série discreta;
- T_t é a tendência;
- S_t é a sazonalidade;
- C_t é o ciclo;
- a_t é o residuo do modelo.

Nos primeiros estudos, era comum incluir uma componente cíclica para representar periodicidades muito longas, geralmente maiores que 12 meses. Segundo Granger e Newbold (1977), devido às séries em geral não conterem

observações suficientes para detectar um ciclo, aquilo que parece ser uma tendência pode estar confundido com parte de um ciclo.

Se removidas as componentes T_t (tendência) e S_t (sazonalidade), sobra a_t (erro aleatório), supondo que seja um processo estocástico puramente aleatório.

Segundo Morettin e Toloí (1985), se existir uma seqüência $\{a_t, t \in T\}$ de variáveis i.i.d., com média zero e variância constante, tais variáveis são chamadas choques aleatórios e a seqüência é denominada ruído branco.

O modelo (2) é aditivo, sendo utilizado quando a sazonalidade S_t varia independentemente da componente de tendência T_t ; se estas componentes são dependentes, o melhor modelo proposto é o multiplicativo.

$$z_t = T_t \cdot S_t \cdot a_t ; \quad t = 1, 2, \dots, N \quad (3)$$

Neste caso, a variância não é constante, sendo necessário aplicar uma transformação apropriada aos dados, por exemplo “logaritmo”, para que a variância da série se torne constante; o modelo aditivo é expresso por

$$z_t^* = \log T_t + \log S_t + \log a_t \quad (4)$$

em que z_t^* é a série transformada.

Estudos realizados comprovam que as transformações não melhoram a qualidade da previsão. Granger e Newbold (1976) mostram que as previsões dos antilogaritmos dos dados transformados são estimadores viciados.

O principal objetivo dos modelos de decomposição é o de modelar convenientemente as três componentes, T_t , S_t e a_t , a fim de que gerem boas previsões e ajuste para a série.

2.4 Testes para as Componentes T_i e S_i

2.4.1 Teste para Tendência (T_i)

Este teste foi proposto por Cox-Stuart, apresentado em Conover (1980).

O teste dos sinais recebe este nome pelo fato de utilizar sinais (+) e (-) como forma de medir os dados entre duas amostras.

No caso das séries temporais, este teste é aplicado para verificar a existência de tendência na série, sendo necessário separar a série em duas partes iguais, considerando a primeira parte como uma amostra (X_A) e a outra parte da série como uma amostra (X_B).

A hipótese nula do teste dos sinais é

$$H_0: p(X_A > X_B) = p(X_A < X_B) = 0,5$$

Considera-se $p(X_A > X_B)$ a probabilidade da amostra X_A ser maior que a amostra X_B , e $p(X_A < X_B)$ a probabilidade da amostra X_A ser menor que a amostra X_B .

Toma-se z_i ($i=1,2,\dots,n$), como a observação da amostra X_A , no instante i e z_j ($j=1,2,\dots,n$) como a observação da amostra X_B no instante j , em que n é o tamanho de cada amostra.

Comparam-se as observações z_i e z_j , sendo $i=j$

Se $z_i < z_j$, atribui-se sinal (+)

Se $z_i > z_j$, atribui-se sinal (-)

No caso das observações das duas amostras serem iguais no instante $i=j$, descarta-se esta comparação, não atribuindo os sinais (+) ou (-).

A estatística do teste a ser utilizada é dada por T ,

sendo T = número de sinais positivos (+) atribuídos ao final da amostra.

Esta estatística é comparada com o valor dado por t ,

$$t = \frac{1}{2} (n' + w_{\alpha/2} \sqrt{n'})$$

em que N' é o número total de sinais (+) e (-) atribuídos às comparações entre as duas amostras; $w_{\alpha/2}$ é o valor tabelado e α o nível de significância.

No caso do tamanho da amostra ser menor que 20, utiliza-se o valor tabelado para $w_{\alpha/2}$ através da distribuição binomial, considerando $p=0,5$. Se o tamanho amostral for maior que 20, utiliza-se a distribuição normal para encontrar o valor tabelado $w_{\alpha/2}$.

Utiliza-se a seguinte regra de decisão, independente do tamanho amostral: rejeita-se H_0 se $T \leq t$ ou se $T \geq N' - t$.

2.4.2 Teste para Sazonalidade (S_t)

O teste para sazonalidade foi proposto por Fisher, utilizando valores do periodograma. Este teste é encontrado em Priestley (1989).

O periodograma detecta periodicidades existentes na série temporal, que não podem ser consideradas significativas imediatamente. Uma condição para a utilização deste procedimento é de que a série temporal tenha um grande número de observações para que se possa detectar as periodicidades.

Aplica-se o teste do periodograma para determinar se as periodicidades encontradas na série são significativas, pois caso não forem detectadas, irão provavelmente surgir com componentes periódicas nos modelos.

O procedimento usual para “plotar” o periodograma é através da frequência padrão, dada por

$$w_p = \frac{2\pi p}{N} \quad p = 0, 1, \dots, N/2$$

sendo N o número de observações encontradas na série temporal. Os valores deste teste detectam grandes periodicidades.

Considera-se a hipótese nula deste teste, sendo

H_0 : Não existência de periodicidade

Através do procedimento descrito acima, pode-se construir um teste para a maior periodicidade encontrada no periodograma. A partir do $\max(I_p)$, pode-se construir a estatística g , dada por

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

O teste de Fisher consiste em determinar um z_α , tal que, sob a hipótese nula,

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

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

Para se obter a estatística z , é necessário fixar o nível de significância (α). Esta estatística é comparada com a estatística g , tendo como regra de decisão $g > z$. Rejeita-se a hipótese H_0 , constatando o efeito de sazonalidade no período analisado.

2.5 Modelos de Box e Jenkins

2.5.1 Introdução

Yule, em 1926, introduziu os modelos autorregressivos (AR), enquanto os modelos de médias-móveis (MA) surgiram com Slutsky em 1937. Já Wold (1938) mostrou que qualquer processo estocástico estacionário discreto pode ser representado por modelos autorregressivos e médias-móveis (ARMA), segundo Morettin e Toloí (1985).

Box e Jenkins apresentaram, em 1970, uma metodologia geral para desenvolver modelos de previsões e controle em séries temporais.

A partir das idéias propostas por Wold (1938), Box e Jenkins construíram uma técnica de identificação do modelo, estimação dos parâmetros e verificação da validade do modelo. Essa sofisticada técnica fornece previsões com embasamento probabilístico e com o menor erro de estimativa, baseada na construção de modelos estimados de forma iterativa. Contudo, a estimação com tais modelos tem algumas restrições, tais como o tamanho da amostra, que deve ser de no mínimo de 50 observações devido à existência de alguns resultados assintóticos.

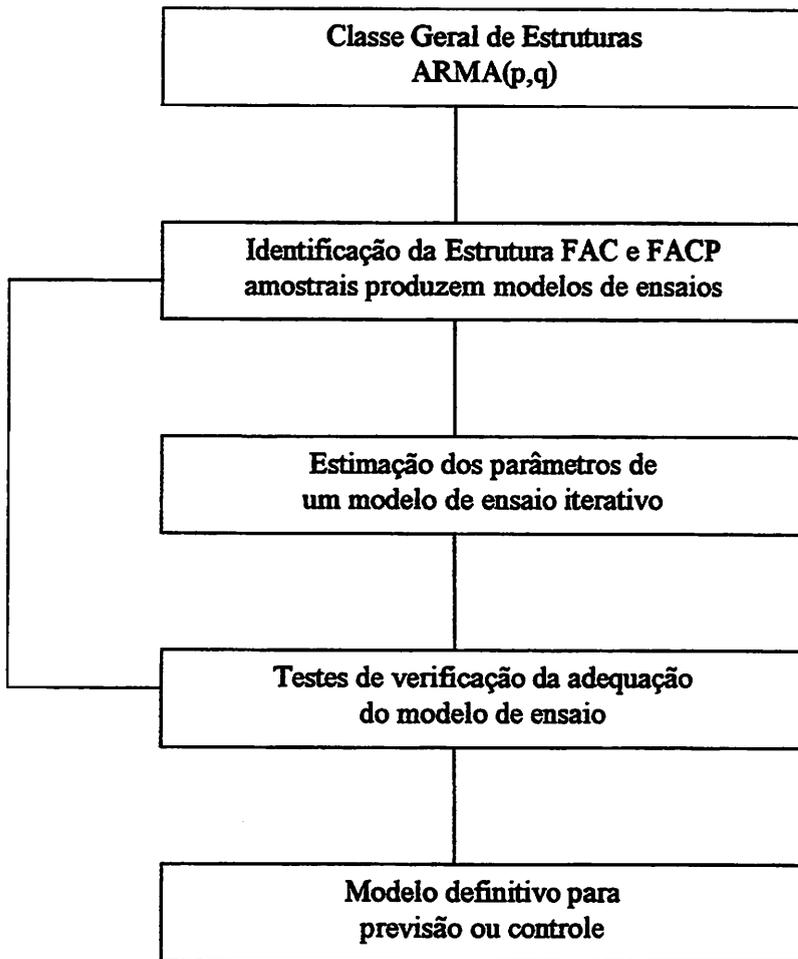


FIGURA 1 – Fluxograma das Fases da Metodologia Box e Jenkins

A estrutura do modelo é baseada em seus próprios dados, em que os estágios para a escolha do melhor modelo são (Figura 1):

- a) considerar uma classe geral de modelos para a análise (especificação);
- b) identificar um modelo com base nas funções de autocorrelações (FAC), autocorrelações parciais (FACP) e outros critérios;

- c) estimar parâmetros do modelo identificado;
- d) e verificar o modelo ajustado, através de uma análise de resíduos, para saber se o modelo escolhido é adequado para o objetivo de fazer previsão ou controle.

Caso o modelo não seja adequado, o ciclo é repetido, retornando-se à fase de identificação.

2.5.2 Modelos ARMA

Segundo Morettin e Tolo (1985), baseados na proposta de Wold (1938), se combinados o modelo autorregressivo (AR) e o modelo médias-móveis (MA), obtém-se o modelo autorregressivo médias-móveis (ARMA).

O modelo ARMA de ordem (p,q), sendo p parâmetros autorregressivos e q parâmetros de médias-móveis, para a série temporal $\{ z_t ; t= 1, 2, \dots, N \}$, tem a seguinte forma

$$\phi(B)\tilde{z}_t = \theta(B)a_t \tag{5}$$

sendo a_t o resíduo do modelo ARMA, $\tilde{z}_t = z_t - \mu$ os desvios em relação à constante μ , onde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ e $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ são polinômios do operador de retardo B, tal que

$$Bz_t = z_{t-1} \quad \text{e} \quad B^m z_t = z_{t-m}$$

Os modelos ARMA envolvem o conceito de sistema dinâmico, que é caracterizado por uma série de entrada denominada $x(t)$, uma série de saída $z(t)$ e uma função de transferência $v(t)$. Este modelo supõe que a série temporal é gerada através de um filtro linear, sendo que a entrada é um ruído branco (Figura 2).

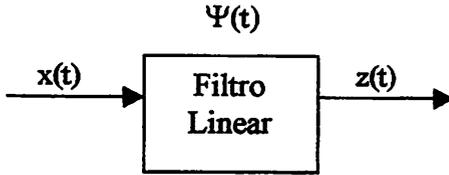


FIGURA 2 – Filtro linear com entrada $x(t)$, saída $z(t)$ e função de transferência $\Psi(t)$

Segundo Morettin e Toloi (1985), a operação de filtro linear consiste simplesmente de uma soma ponderada dos ruídos anteriores

$$\begin{aligned} z_t &= \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \dots \\ &= \mu + \Psi(B)a_t \end{aligned} \quad (6)$$

sendo

$$\Psi(B) = 1 + \Psi_1 B + \Psi_2 B^2 + \dots$$

que é o operador linear chamado função de transferência; μ é um parâmetro que determina o nível da série e a_t é um processo linear discreto, contendo as seguintes características

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

Sendo $\tilde{z}_t = z_t - \mu$, tem-se que

$$\tilde{z}_t = \Psi(B)a_t \quad (7)$$

Quando a seqüência de pesos $\{\Psi_j, j \geq 1\}$ é finita ou infinita e convergente, o filtro é estável (somável) e z_t é estacionária. Neste caso, μ se torna a média do processo; caso contrário, z_t não é estacionária e μ não tem significado específico, a não ser como um ponto de referência para o nível da série.

2.5.2.1 Condições de Estacionaridade e Invertibilidade

A equação (5) define um processo estacionário se as raízes de $\phi(B) = 0$ estiverem fora do círculo unitário, e o processo é invertível se as raízes da função característica $\theta(B) = 0$ estiverem fora da região do círculo unitário.

Considerando, como exemplo, um modelo ARMA(1,1)

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + a_t - \theta_1 a_{t-1}$$

sendo

$$\Psi(B) = (1 - \phi_1 B)^{-1} = \sum_{j=0}^{\infty} \phi_1^j B^j$$

e

$$\pi(B) = (1 - \theta_1 B)^{-1} = \sum_{j=0}^{\infty} \theta_1^j B^j$$

chega-se às seguintes condições de estacionaridade e de invertibilidade, respectivamente,

$$|\phi_1| < 1 \quad \text{e} \quad |\theta_1| < 1$$

A região de variação é representada pela Figura 3

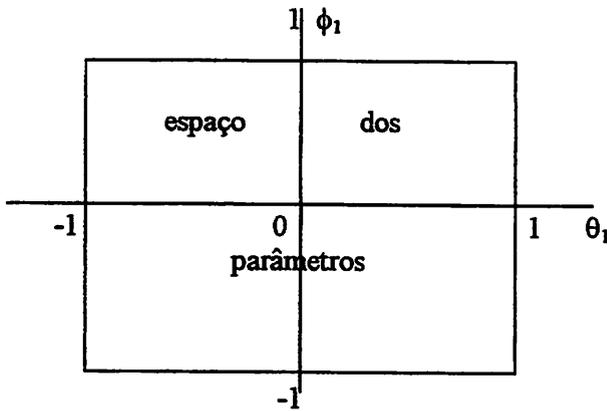


FIGURA 3 – Região de Estacionaridade e Invertibilidade do Processo ARMA (1,1)

2.5.3 Função de Autocorrelação

A função de autocorrelação do processo ARMA(p,q) é definida pela seguinte expressão

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (8)$$

Multiplicando-se ambos os lados da equação por \tilde{z}_{t-k} e tomando-se as esperanças, obtém-se a função de autocovariância

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + \gamma_{za}(k) - \theta_1 \gamma_{za}(k-1) - \dots - \theta_q \gamma_{za}(k-q) \quad (9)$$

onde $\gamma_{za}(k)$ é a covariância cruzada entre z_t e a_t , expressa por

$$\gamma_{za}(k) = E[\tilde{z}_{t-k} a_t]$$

e γ_k é a covariância de valores passados da série temporal expressa por

$$\gamma_k = E(\tilde{z}_t \cdot \tilde{z}_{t-k})$$

Como \tilde{z}_{t-k} só depende dos choques ocorridos até o instante t-k, tem-se

$$\gamma_{za} \begin{cases} = 0, & k > 0 \\ \neq 0, & k \leq 0 \end{cases}$$

Então, para $k \geq q + 1$, a função de autocovariância γ_k é dada por

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} \quad (10)$$

e a função de autocorrelação ρ_k por

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p} \quad (11)$$

em que $\rho_k = \frac{\gamma_k}{\gamma_0}$.

Portanto, no processo ARMA (p,q), as primeiras q autocorrelações dependem dos parâmetros autorregressivos e dos parâmetros de médias-móveis. Entretanto, após o “lag” q, as autocorrelações comportam-se como nos modelos autorregressivos.

Se $q < p$, a função de autocorrelação consiste de uma mistura de exponenciais amortecidas e (ou) de senóides amortecidas.

Se $q > p$, os $q-p+1$ primeiros valores $\rho_0, \rho_1, \dots, \rho_{q-p}$ não seguirão este padrão. Mais detalhes são apresentados por Box e Jenkins (1976).

Para uma maior compreensão, desenvolve-se a FAC para o modelo ARMA (1,1). Por exemplo, tomando-se a função de autocovariância (10)

$$\gamma_1 = \phi_1 \gamma_0 - \theta_1 \gamma_{za}(0) = \phi_1 \gamma_0 - \theta_1 \sigma_a^2 \quad (12)$$

e

$$\gamma_0 = \phi_1 \gamma_1 + \gamma_{za}(0) - \theta_1 \gamma_{za}(-1) = \phi_1 \gamma_1 + \sigma_a^2 - \theta_1 (\phi - \theta) \sigma_a^2 \quad (13)$$

Partindo de (12) e (13), obtém-se a função de autocorrelação

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta}$$

Para o modelo ARMA (1,1), a FAC tem decaimento exponencial a partir da defasagem 1. A autocorrelação nesta defasagem tem o sinal de $(\phi_1 - \theta_1)$. Se $\phi_1 > 0$, todas as autocorrelações terão mesmo sinal; caso contrário, os sinais serão alternados.

2.5.4 Função de Autocorrelação Parcial

Box e Jenkins (1976) propõem a utilização da função de autocorrelação parcial como um critério para facilitar a identificação do modelo.

Na análise de séries temporais, o coeficiente de autocorrelação parcial é utilizado para medir o grau de associação entres as observações z_t e z_{t-k} quando os efeitos das outras defasagens 1, 2, 3, ..., k-1 são fixados.

A maneira usual de estimar os coeficientes de autocorrelação parcial é calculá-los em função dos coeficientes de autocorrelação, de modo que se obtenham as chamadas equações de Yule-Walker, que podem ser expressas matematicamente como:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} \quad (14)$$

nas quais ϕ_{kj} é o j-ésimo coeficiente do modelo autorregressivo e ϕ_{kk} o último coeficiente, substituindo-se por $\hat{\phi}_{11}, \hat{\phi}_{22}, \dots, \hat{\phi}_{kk}$. A correlação ρ_j , $j=1,2,\dots, k$ é substituída por $\hat{\rho}_j$ estimado de modo usual.

Para um processo ARMA(p,q), a FACP se comporta como a FACP de um processo MA(q) puro.

As condições de invertibilidade da FACP em um processo ARMA(p,q) é infinita, sendo que seu comportamento é uma mistura de exponenciais e (ou) senóides amortecidos após p-q primeiros “lags”.

As características da FAC e FACP, nos diversos processos, são usadas no estágio de identificação da estrutura. Procuram-se estas propriedades no correlograma amostral. Dada a presença de aleatoriedade nas estimativas dos valores da FAC e FACP, este processo muitas vezes não é suficientemente claro. Os correlogramas amostrais são construídos com as estimativas dos parâmetros ρ_k e ϕ_{kk} , $k= 1, 2, 3, \dots$, dadas por

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad \text{com} \quad \hat{\gamma}_k = \frac{\sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{N}$$

onde

$$\bar{z} = \frac{\sum_{t=1}^N z_t}{N}$$

e $\hat{\phi}_{kk}$ obtido das equações de Yule-Walker com ρ_k substituído por $\hat{\rho}_k$. As distribuições de probabilidade das variáveis aleatórias $\hat{\rho}_k$ e $\hat{\phi}_{kk}$ são complicadas. Anderson, em 1942, mostrou que se o parâmetro estimado, ρ_k , é nulo e o tamanho da série é de moderado a grande, então o estimador $\hat{\rho}_k$ tem distribuição gaussiana

$$\hat{\rho}_k \sim N[0, V(\hat{\rho}_k)]$$

Bartlett(1946) propôs uma expressão aproximada para a variância de $\hat{\rho}_k$

$$V(\hat{\rho}_k) \cong \frac{1}{N} \sum_{v=-\infty}^{\infty} [\rho_v^2 + \rho_{v+k} \rho_{v-k} - 4 \rho_k \rho_v \rho_{v-k} + 2 \rho_v^2 \rho_k^2] \quad (15)$$

Como ρ_v é desconhecida, utiliza-se um estimador com $\hat{\rho}_v$, substituindo ρ_v em (15). Bartlett mostrou que quando as correlações teóricas ρ_v são nulas para defasagens maiores do que um valor fixo q $v > q$, tem-se

$$V(\hat{\rho}_k) \cong \frac{1}{N} \left[1 + 2 \cdot \sum_{v=1}^q \hat{\rho}_v^2 \right] \quad k > q \quad (16)$$

Da mesma forma, com ρ_v desconhecida, utiliza-se a estimativa desta expressão aproximada por $\hat{\rho}_v$.

Quenouille (1949) mostrou que na hipótese de um processo autorregressivo de ordem p , a variância aproximada da estimativa $\hat{\phi}_{kk}$ é

$$V(\hat{\phi}_{kk}) \cong \frac{1}{N} \quad (17)$$

se N é grande, supõe-se distribuição Gaussiana para $\hat{\phi}_{kk}$.

2.5.5 Identificação

A etapa inicial da metodologia de Box e Jenkins, como comentado, é a identificação da estrutura do processo ARMA, que é uma fase importante do método, pois um erro nesta fase conduzirá a falsos resultados. A estrutura do modelo é identificada comparando as funções amostrais de autocorrelação e de autocorrelação parcial com os correlogramas teóricos das diversas estruturas ARMA(p, q).

A técnica, em geral, é eficiente, mas é importante salientar que pode ser identificado mais de um modelo para a mesma série. Como o esquema segue o modelo de decomposição de Wold dado pela equação (2), é necessário que a série temporal seja um processo estocástico estacionário. Esta condição permite estimar a média e as funções de autocorrelação e autocorrelação parcial do processo, portanto

$$\mu(t) = E[z(t)] = \mu$$

sendo μ constante; a função de autocovariância é expressa por

$$\text{cov}(z_{t_1}, z_{t_2}) = \gamma(t_1, t_2)$$

e depende apenas de $t_1 - t_2$.

Na prática, dificilmente encontram-se séries estacionárias, portanto retiram-se diferenças nos dados para que a série satisfaça à condição de estacionaridade. Por exemplo, se a série apresentar sazonalidade no tempo k , faz-se $z_t - z_{t-k}$, sendo z_t a observação no instante t e z_{t-k} a observação no instante $t-k$.

A identificação de uma estrutura ARMA(p,q) é realizada comparando-se os “lags” dos correlogramas teóricos da FAC e FACP das várias estruturas com os correlogramas obtidos da série observada $\{z_t; t = 1, 2, \dots, N\}$. Portanto, a função de autocorrelação e a função de autocorrelação parcial são fundamentais na escolha do modelo apropriado.

Após feita a análise de identificação da provável estrutura de modelagem da série, passa-se para o estágio de estimação. No caso de existir mais de um modelo para a série, utiliza-se a última fase do processo para a análise dos resíduos.

2.5.6 Estimação

Segundo Box e Jenkins (1976), quando identificado o modelo da estrutura ARMA(p,q) da forma

$$\phi(B)z_t = \delta + \theta(B)a_t \quad (18)$$

sendo

$$z_t = \nabla^d x_t \quad d = 0, 1, 2, \dots$$

$$\nabla = 1 - B$$

x_t são as t observações da série;

z_t os valores da série com diferença, onde $t= 1, 2, \dots, N$;

a_t o termo do resíduo;

estima-se o vetor de parâmetros

$$\zeta = [\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q; \sigma_a^2]$$

utilizando-se o processo iterativo de determinação de mínimos quadrados não-lineares.

Se necessário, para realizar a estimação dos parâmetros, utiliza-se alguma transformação. Os $p+q+1$ parâmetros do modelo da estrutura identificada são analisados partindo-se da suposição de que o ruído a_t tem distribuição gaussiana, ou seja,

$$a_t \sim N(0, \sigma_a^2) \quad t = 1, 2, \dots, N$$

O princípio da máxima verossimilhança afirma que “identificado o modelo correto, toda informação sobre os parâmetros está contida na função de verossimilhança”. Assim, o método da máxima verossimilhança, que foi proposto por Gauss em 1821 e redescoberto por Fisher em 1922, é o método iterativo mais utilizado para a estimação.

No estudo de uma série $z = (z_1, z_2, \dots, z_N)$, determinam-se os parâmetros do modelo que maximizam a probabilidade desta função, ou seja, utiliza-se a função de verossimilhança. No procedimento não-condicional de estimação, o logaritmo da função de verossimilhança para as $(N+d)$ observações da série é dado por

$$\ell(\xi) = f(\eta) - N \cdot \ln(\sigma_a) - S(\eta) / 2\sigma_a^2 \quad (19)$$

onde

$$\eta = [\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q]$$

e

$$S(\eta) = \sum_{t=1}^N [a_t(\eta, z)]^2$$

é denominada função soma de quadrados não-condicional Box e Jenkins (1976) afirmam que $f(\eta)$ é importante somente para pequenos valores de η , e ainda que para os tamanhos da amostra moderados a grandes, $\ell(\xi)$ depende somente de $\frac{S(\eta)}{2\sigma_s^2}$.

As curvas de níveis geradas pela função da soma de quadrados não-condicional no espaço paramétrico de η são aproximadamente as curvas de nível da função logaritmo de verossimilhança. Portanto, maximizar $\ell(\xi)$ corresponde a minimizar $S(\eta)$. As estimativas de mínimos quadrados obtidas são consideradas boas aproximações para as estimativas de máxima verossimilhança, determinando-as através das técnicas de mínimos quadrados não-lineares.

As segundas derivadas da função logaritmo de verossimilhança fornecem medidas do espalhamento da função de verossimilhança, sendo possível usá-las como medidas aproximadas de variância do estimador. Considerando que a função logaritmo de verossimilhança é aproximadamente quadrática, tem-se

$$\ell(\eta, \sigma_s) \cong \ell(\hat{\eta}, \sigma_s) + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \ell_{ij} (\eta_i - \hat{\eta}_i) (\eta_j - \hat{\eta}_j) \quad k=p+q \quad (20)$$

com as derivadas $\ell_{ij} = \frac{\partial^2 \ell(\eta, \sigma_s)}{\partial \eta_i \partial \eta_j}$ constantes. Considerando η grande, a parcela

$\ell(\hat{\eta}, \sigma_s)$ pode ser desprezada, tomando-se quadrática. As variâncias e covariâncias dos estimadores podem ser obtidas considerando a matriz de informação de Fisher para o parâmetro η .

A matriz de covariância tem a seguinte expressão:

$$V(\hat{\eta}) = \begin{bmatrix} v(\hat{\eta}_1) & \text{cov}(\hat{\eta}_1, \hat{\eta}_2) & \dots & \text{cov}(\hat{\eta}_1, \hat{\eta}_k) \\ \text{cov}(\hat{\eta}_1, \hat{\eta}_2) & v(\hat{\eta}_2) & \dots & \text{cov}(\hat{\eta}_2, \hat{\eta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\hat{\eta}_k, \hat{\eta}_2) & \text{cov}(\hat{\eta}_k, \hat{\eta}_2) & \dots & v(\hat{\eta}_k) \end{bmatrix}$$

e fazendo $\ell_{ij} \equiv \frac{-s_{ij}}{2\sigma_s^2}$, sendo $s_{ij} = \frac{\partial^2 S(\eta|w)}{\partial \eta_i \partial \eta_j}$,

tem-se $V(\hat{\eta}) \equiv [-E(\ell_{ij})]^{-1} \equiv 2\sigma_s^2 [S_{ij}]^{-1}$

ou ainda

$$V(\hat{\eta}) \equiv 2\sigma_s^2 \begin{bmatrix} \frac{\partial^2 S(\eta)}{\partial \eta_1^2} & \frac{\partial^2 S(\eta)}{\partial \eta_1 \partial \eta_2} & \dots & \frac{\partial^2 S(\eta)}{\partial \eta_1 \partial \eta_k} \\ & \frac{\partial^2 S(\eta)}{\partial \eta_2^2} & \dots & \frac{\partial^2 S(\eta)}{\partial \eta_2 \partial \eta_k} \\ & & \ddots & \vdots \\ & & & \frac{\partial^2 S(\eta)}{\partial \eta_k^2} \end{bmatrix}$$

que é uma matriz simétrica, e o valor de σ_s^2 é estimado pelo quadrado médio dos resíduos e as derivadas calculadas por método numérico.

2.5.7 Verificação

Tendo-se estimado o modelo e obtidas as estimativas dos parâmetros, deve-se aplicar alguns testes nos componentes do vetor $\hat{\eta} = [\hat{\eta}_1, \dots, \hat{\eta}_k]$ a fim de fornecer validade ao modelo. O principal procedimento para a verificação do ajuste do modelo é a análise de resíduos.

O teste de validade de um modelo de série temporal pode ser aplicado aos resíduos, e quando existe dúvida sobre a ordem do modelo, deve-se aplicar o

teste da sobrefixação. Este teste consiste basicamente na construção de um modelo com número de parâmetros de ordem superior ao modelo fixado. Através das estimativas dos novos parâmetros, aplica-se o teste “t” sobre a hipótese de o parâmetro adicional ser nulo; caso o parâmetro seja significativo, acrescenta-se este novo parâmetro ao modelo.

A análise de resíduos abrange vários procedimentos, primeiramente verifica-se a hipótese de independência de a_t por meio da função de autocorrelação residual. Sendo verdadeira esta suposição, tem-se que $\hat{\rho}(\hat{a}_t) \cong 0$.

A verificação da distribuição gaussiana dos resíduos pode ser analisada observando o gráfico da distribuição normal. Este procedimento consiste em construir o gráfico dos resíduos ordenados $\hat{a}(t)$ contra os escores padronizados

$$z_i = \phi^{-1} \left[\frac{(i - 0,5)}{N} \right], \quad i = 1, 2, \dots, N, \quad \text{com } \phi \text{ correspondendo à função de}$$

distribuição da variável normal padrão z_i :

$$\phi(z_j) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{z_j} e^{-\frac{x^2}{2}} dx$$

Se os resíduos \hat{a}_t têm distribuição aproximadamente normal, então o gráfico da normal deve produzir aproximadamente uma reta. Quando o modelo ajustado não é satisfatório, e portanto a parte sistemática não está convenientemente modelada, os resíduos costumam apresentar componentes periódicas.

O teste do periodograma acumulado aplicado à série dos resíduos, a fim de detectar a existência de possíveis componentes periódicos que apontam uma não-aleatoriedade periódica, foi estudado por Bartlett (1946).

O mais utilizado dos testes aplicados aos resíduos foi proposto por Box e Pierce (1970), detalhado na seção a seguir (2.5.7.2).

Neste ponto da metodologia Box e Jenkins dispõem de um modelo com estrutura identificada, parâmetros estimados e com o ajuste do modelo analisado de forma que se pode partir para previsões, usando-o com razoável segurança.

2.5.7.1 Teste de Normalidade para os Resíduos

O teste utilizado para verificar normalidade foi proposto por Shapiro-Wilk, conforme Conover (1989), sendo as hipóteses a serem testadas descritas abaixo:

H_0 : $F(x)$ é uma função de distribuição normal com média e variância não especificadas

H_1 : $F(x)$ não segue uma distribuição normal

A estatística do teste é realizada pelo denominador D

$$D = \sum_{i=1}^N (x_i - \bar{x})^2$$

sendo que \bar{x} é a média amostral. Tem-se que a ordem amostral é realizada do menor para o maior valor observado, e x_i representa o i -ésimo valor observado.

A estatística T_s é dada por

$$T_s = \frac{1}{D} \left[\sum_{i=1}^k a_i (x^{(N-i+1)} - x^{(i)}) \right]^2$$

Através do coeficiente do teste de Shapiro-Wilk, com o tamanho amostral n , obtêm-se os coeficientes a_1, \dots, a_k , onde k é aproximadamente $N/2$.

Rejeita-se a hipótese H_0 no nível de significância α se T_s for menor que o quantil α dado pela estatística de Shapiro-Wilk. Se um nível crítico para um valor observado de T_s é desejado, utiliza-se a conversão de T_s para uma variável aleatória aproximadamente normal, o qual pode ser comparado com α da distribuição Normal.

2.5.7.2 Teste de Ruído Branco

Utilizado para verificar se o resíduo do modelo é um ruído branco.

O mais utilizado dos testes, proposto por Box e Pierce (1970), foi aplicado aos resíduos para verificar a existência de correlação, com uma distribuição aproximadamente qui-quadrado, com graus de liberdade $m=k-(p+q)$, em que

k representa o número de lags, em geral utiliza-se o tamanho 25;

p é a ordem do modelo autorregressivo;

q é a ordem do modelo de médias-móveis.

A estatística de Box e Pierce é dada por

$$Q_c = N \sum_{k=1}^K \hat{\rho}_k^2$$

sendo N o número de observações da série temporal.

A estatística de Box e Pierce foi posteriormente modificada por Box e Ljung (1978) e passou a ser utilizada devido à distribuição amostral ser mais próxima da distribuição qui-quadrado.

$$Q_c = N(N+2) \sum_{k=1}^K (N-k)^{-1} \hat{\rho}_k^2$$

Quando o valor da estatística $Q_c < \chi_m^2$, conclui-se que o modelo está bem ajustado e seu resíduo não é correlacionado, ou seja, o resíduo é independente e identicamente distribuído.

2.5.8 Modelos ARIMA e SARIMA

O modelo ARIMA é um caso geral dos modelos propostos por Box e Jenkins, o qual é apropriado para descrever séries não-estacionárias, ou seja, séries que não possuem média constante no período de análise, nas quais os parâmetros quase sempre são pequenos. Na prática, geralmente as séries encontradas apresentam tendência e (ou) sazonalidade.

Quando a série não apresenta o componente sazonal, mas apresenta o componente de tendência, ou esses componentes são homogêneos não-estacionários, ela pode ser representada por um modelo ARIMA (Autorregressivo Integrado Médias-Móveis), ou seja, o ajuste é feito por um modelo sem a componente sazonal.

O modelo ARIMA considera a tendência da série temporal, tem ordem (p,d,q) e pode ser representado por:

$$\phi(B)\Delta^d z_t = \theta(B)a_t \quad (21)$$

sendo Δ o operador das diferenças e d é o número de diferenças necessárias para que a série se torne estacionária. Em geral, com o uso de no máximo duas diferenças, a série obtida não apresenta mais a componente de tendência.

Quando uma série temporal apresenta comportamento periódico em um período máximo de 12 meses, é necessário acrescentar uma componente sazonal no modelo. Esse intervalo periódico é chamado de sazonalidade. Se o intervalo periódico for superior a 12 meses, não é acrescentada nenhuma componente no modelo. Em geral o número de observações é pequeno, não sendo viável o ajuste desta periodicidade, que é denominada ciclo.

A sazonalidade normalmente ocorre no período de 6 ou 12 meses. Esta variação ocorre devido a vários fatores, como o clima, datas festivas, fatores institucionais, entre outros. Um exemplo desta ocorrência é o número de

acidentes ocorridos nas rodovias federais e estaduais brasileiras. Na época de férias e feriados, este índice aumenta consideravelmente.

Um fator não muito comum é a existência de efeito sazonal dentro de cada mês. Tome-se como exemplo o volume de negócios de um banco, que tem sazonalidade de 12 meses, ou seja, o volume de negócios se concentra em uma determinada época do ano. Mas também existe uma sazonalidade nos dias de pagamento, que tendem a se concentrar no 5º dia útil ou todo dia 10 do mês. A sazonalidade, neste caso, ocorre no período de 4 semanas; portanto, uma série pode ter mais de um componente sazonal.

Pode-se definir um operador diferença generalizado, quando duas observações distantes entre si de s intervalos de tempo apresentam alguma semelhança, por

$$\nabla_s z_t = (1 - B^s)z_t = z_t - z_{t-s} \quad (22)$$

Obtendo-se uma série z_t , com efeitos sazonais no período de 12 meses, então espera-se que ocorra:

- a) relações entre as observações dos meses do mesmo ano (variação dentro do período).
- b) relação entre as observações de um mesmo mês em diferentes anos (variação entre períodos).

Seja uma série temporal observada no período de janeiro de 1995 a dezembro de 1999. A primeira observação, janeiro de 1995, está relacionada com as observações deste mesmo ano e com as observações de janeiro dos anos de 1996, 1997, 1998 e 1999.

Para representar esta dependência, pode-se usar o seguinte modelo

$$\Phi_P(B^s)\nabla_s^D z_t = \Theta_Q(B^s)\alpha_t \quad (23)$$

o qual é escrito somente para as séries com sazonalidade e sem tendência, sendo D o número de diferenças sazonal, $\Phi_P(B^s)$ e $\Theta_Q(B^s)$ polinômios de graus P

(autorregressivo sazonal) e Q (médias -móveis sazonal), respectivamente, que satisfaçam as condições de estacionaridade e invertibilidade.

É comum supor que os parâmetros em Φ e Θ são os mesmos para todos os meses, ou seja, a ordem do modelo não se altera de mês em mês.

O termo α_t é o resíduo do modelo, que em geral consiste em um ruído branco, satisfazendo as condições de média zero, variância constante e (i.i.d.). Se α_t não satisfaz as condições acima descritas, pode-se representar o resíduo através do seguinte modelo

$$\phi_p(B)\nabla^d\alpha_t = \theta_q(B)a_t \quad (24)$$

onde o termo a_t é o ruído branco.

Box e Jenkins (1976) propõem o modelo sazonal multiplicativo, representado pela expressão

$$\phi_p(B)\Phi_p(B^s)\nabla^d\nabla_s^D z_t = \theta_q(B)\Theta_Q(B^s)a_t \quad (25)$$

e denominado SARIMA, de ordem $(p,d,q)\times(P,D,Q)$.

Em algumas séries, o modelo multiplicativo não é adequado, sendo necessário transformar o operador de médias-móveis $\theta(B)\Theta(B^s)$ por um operador $\theta^*(B)$, e o operador autorregressivo $\phi(B)\Phi(B^s)$ por um operador $\phi^*(B)$.

A expressão abaixo mostra um exemplo com um operador multiplicativo, utilizando componentes de médias-móveis

$$\theta(B)\Theta(B) = (1 - \theta_1 B)(1 - \Theta_1 B^{12})$$

pode ser também expresso pela forma

$$\text{SARIMA}(0,0,1)\times(0,0,1)$$

e seja o operador não multiplicativo

$$\theta^*(B) = 1 - \theta_1 B - \theta_{12} B^{12} - \theta_{13} B^{13}$$

É importante observar, no modelo multiplicativo, que o componente de médias-móveis era representado por dois parâmetros, sendo o parâmetro sazonal

e o de médias-móveis. Já no operador não-multiplicativo, existe um só parâmetro, sendo os polinômios de ordens superiores.

Se θ_{13} diferir muito de $\theta_{1\oplus 1}$, o modelo não-multiplicativo será melhor, mas se as estimativas dos parâmetros forem próximas, o modelo multiplicativo será mais parcimonioso no sentido de um menor número de parâmetros.

Para a identificação do modelo SARIMA, utiliza-se a FAC e a FACP, observando as correlações nos “lags” sazonais de interesse; por exemplo, pode existir correlação no “lag” múltiplo de 12, ou seja, existe um período sazonal de 12 meses.

A estimação dos parâmetros dos modelos sazonais, é realizada, da mesma forma que os parâmetros autorregressivos e de médias-móveis, calculando-se a soma de quadrados para diversos conjuntos de valores dos parâmetros e tomando-se as estimativas que resultem na menor soma de quadrados do resíduo.

Para verificar se o modelo proposto é adequado, utiliza-se os testes de autocorrelação residual, ou aplica-se o teste de Box e Pierce, ou o periodograma acumulado.

No modelo SARIMA, o teste de Box e Pierce faz uso da distribuição qui-quadrado acrescentando o número de coeficientes sazonais nos graus de liberdade, tendo então $k-(p+q+P+Q)$ graus de liberdade.

Existem outros critérios para escolha do melhor modelo, entre eles o critério de Akaike.

2.6 Método para Escolha do Modelo

Um método para a seleção do modelo é o critério de Akaike, conhecido como AIC (Littell et al, 1996).

A implementação deste método pode ser feita para um modelo ARMA(p,q), cujo parâmetros são estimados através do método da máxima verossimilhança, e a fórmula utilizada pelo critério de AIC é dada por

$$AIC_{p,q} = \frac{-2 \ln(\text{máxima verossimilhança}) + 2r}{N} \approx \ln(\hat{\sigma}_s^2) + r \frac{2}{N} + c \quad (26)$$

onde $\hat{\sigma}_s^2$ é estimado pelo método da máxima verossimilhança, $r=p+q+1$ é dado pelo número de parâmetros do modelo estimado e c é uma constante.

No critério AIC, o primeiro termo corresponde essencialmente a $-2/N$ do $\ln(\text{máxima verossimilhança})$, enquanto o segundo termo é conhecido como “fator de penalidade”, para inclusão de um parâmetro no modelo.

O modelo que tiver o menor valor de AIC será escolhido, sendo esse valor comparado entre os vários modelos propostos, como critério de seleção do modelo desejado.

Este critério, quando a estimação é feita através do método da máxima verossimilhança, evita um modelo com muitos parâmetros. Tal procedimento e o uso do critério de informação na seleção do modelo pode ser utilizado, mas eles devem ser verificados como um complemento para o processo de seleção do modelo. Este critério não deve ser utilizado como uma substituição da verificação da função de autocorrelação e a função de autocorrelação parcial da série, e da verificação do resíduo \hat{a}_t para um modelo inadequado, deve sempre ser incluída como um método de seleção do modelo global.

2.7 Modelos de Intervenção

2.7.1 Definição

Seja uma série temporal para a qual estima-se um modelo ARMA, que vem sendo analisado há algum tempo. Num determinado instante de tempo ocorre um evento independente do fenômeno que originou a série temporal, cujos efeitos são significativos sobre a série. Por exemplo, seja uma série temporal sobre o número de acidentes ocasionados dentro de um município, para o qual se dispõe de um modelo para fazer previsões ou ajuste. A implantação de radares eletrônicos ou uma nova lei de trânsito mais rígida ocasiona uma diminuição no número de acidentes num instante ou intervalo de tempo. Este fenômeno é estimado por um modelo que possa representá-lo. A esse evento, cujo efeito deve ser introduzido ao modelo, é chamado de intervenção.

2.7.2 Modelo Dinâmico para Intervenção

O termo intervenção foi introduzido por Glass (1972), baseado em Box e Tiao (1965), que já utilizava estes métodos, mas não com o termo intervenção.

Segundo Jenkins (1979), “os modelos de intervenção representam generalizações de métodos usados para análise de dados, usualmente não expresso na forma de séries temporais, aos quais os estatísticos referem-se pelo título geral Delineamento e Análise de Experimentos.”. Apesar de o modelo parecer simples, ele descreve um grande número de efeitos simultaneamente.

O modelo de intervenção é dado pela expressão

$$z_t = \sum_{i=1}^k v_i(B)x_{i,t} + n_t \tag{27}$$

- z_t é a variável resposta;
- $v_i(B)$ é a função de transferência;
- $x_{i,t}$ é uma variável binária;
- n_t é o resíduo do modelo.

onde cada variável $v_i(B)$ é da forma $\frac{w_i(B)}{\delta_i(B)}$, sendo $w_i(B)$ os parâmetros que determinam o efeito de intervenção e $\delta_i(B)$ o efeito de intervenção até atingir um novo nível na série temporal.

A expressão $v_i(B)$ chamada de função de transferência, é encontrada com detalhes na seção (2.7.4).

Na análise de intervenção, algumas variáveis $x_{i,t}$ são classificadas como binárias, fazendo o mesmo papel das variáveis “dummy” em regressão, e o resíduo do modelo de intervenção é explicado por um modelo de Box e Jenkins.

A intervenção é o evento que ocorre em determinado instante da série afetando a sua seqüência, temporariamente ou permanentemente. Portanto, a idéia global da análise de intervenção é avaliar o impacto desta mudança de nível ou inclinação da série, sendo representada pela variável $x_{i,t}$.

Existem três diferentes tipos de variáveis binárias, onde as observações da série serão denotadas por 0, as intervenções por 1 e o instante ou intervalo de tempo em que ocorre a intervenção, por T . Os efeitos de intervenção no modelo estão exemplificados nas Figuras 4, 5 e 6, ilustrando três tipos de função.

a) Função Degrau

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

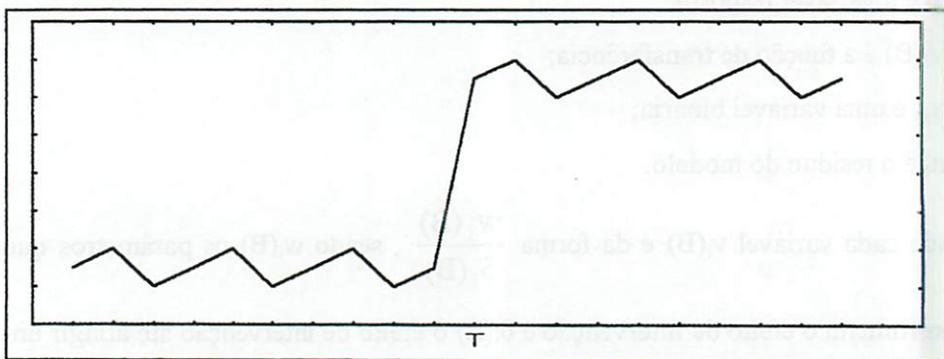


FIGURA 4 – Função Degrau

Esta função muda de nível após sofrer efeito de intervenção no instante T , e continua neste novo nível até o fim da série.

b) Função Impulso

$$x_{i,t} = \begin{cases} 0, & t \neq T \\ 1 & t = T \end{cases}$$

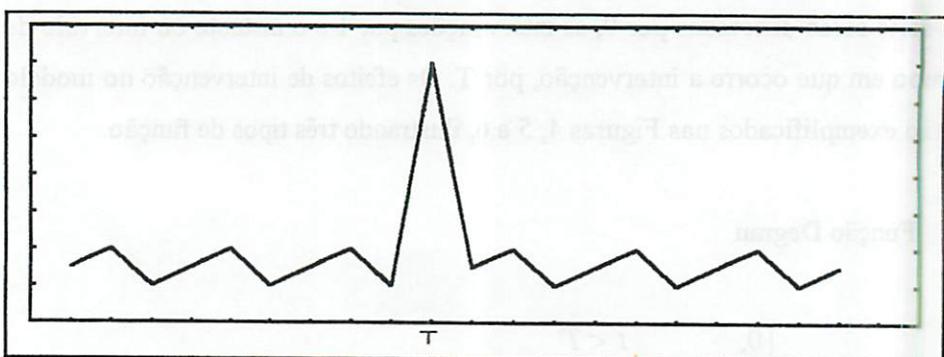


FIGURA 5 – Função Impulso

A série muda de nível no instante em que ocorre o efeito de intervenção T , mas em seguida volta ao nível pré-intervenção.

c) Função Intermediária

$$x_{i,t} = \begin{cases} 0, & t < T_1 \quad t > T_2 \\ 1, & T_1 \leq t \leq T_2 \end{cases}$$

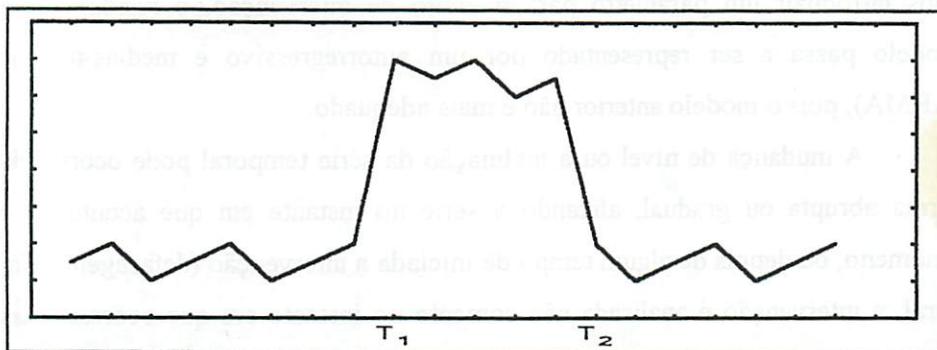


FIGURA 6 – Função Intermediária

A série muda de nível ao sofrer efeito de intervenção T_1 , continua neste novo nível durante um período e volta ao nível pré-intervenção T_2 .

2.7.3 Efeitos de Intervenção

Existem várias formas de uma intervenção afetar uma série temporal. As alterações podem ser afetadas por uma mudança de nível ou por uma inclinação da série. A inclinação ocorre quando a intervenção é complexa e seu efeito é gradativo.

A intervenção pode condicionar alterações no resíduo do modelo. Por exemplo, antes de incluir a estimativa para o efeito de intervenção, o resíduo poderia ser apresentado adequadamente por um modelo autorregressivo (AR); após introduzir um parâmetro para o efeito de intervenção, o resíduo deste modelo passa a ser representado por um autorregressivo e médias-móveis (ARMA), pois o modelo anterior não é mais adequado.

A mudança de nível ou a inclinação da série temporal pode ocorrer de forma abrupta ou gradual, afetando a série no instante em que aconteceu o fenômeno, ou depois de algum tempo de iniciada a intervenção (defasagem). Em geral, a intervenção é analisada não somente no instante em que ocorreu, mas dentro de um intervalo de tempo. O efeito de intervenção pode ser temporário ou permanente.

A Figura 7 apresenta os tipos mais comuns de efeitos de intervenção sobre uma série temporal.

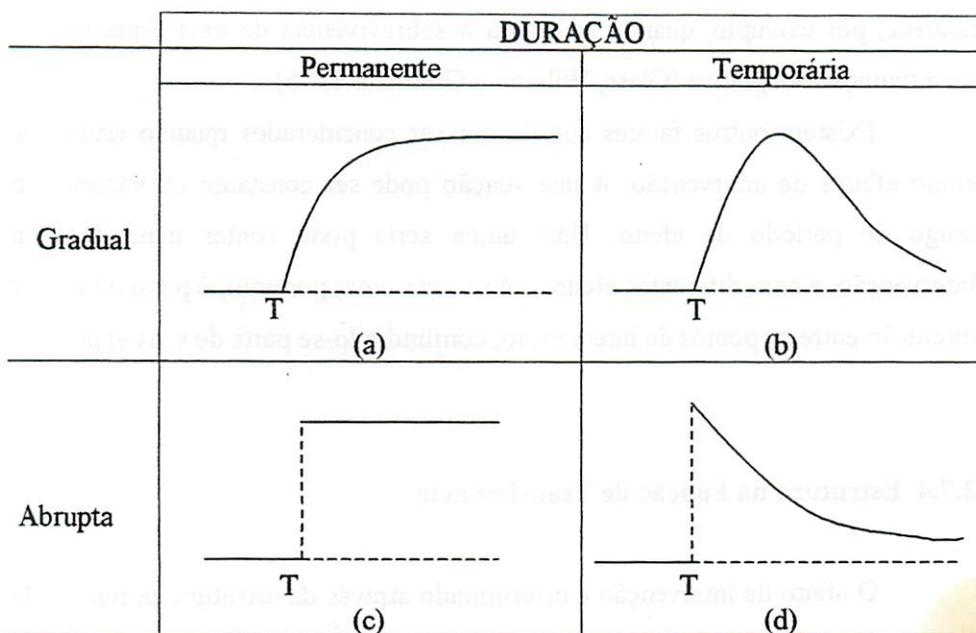


FIGURA 7 – Efeitos de Intervenção

Seguindo a Figura 7, tem-se:

- (a) A série temporal apresenta mudança de nível gradual com duração permanente.
- (b) Ocorre uma mudança de nível gradual na série temporal, mas depois de um intervalo ou instante de tempo, esta volta a se estabilizar.
- (c) A série temporal muda de nível abruptamente e continua permanente neste novo nível.
- (d) Ocorre uma intervenção abrupta, mas depois de um instante ou intervalo de tempo, esta volta a se estabilizar de forma exponencial.

Após a ocorrência da intervenção, pode haver uma mudança de variabilidade na série, bem como um efeito de evolução, no qual a série cai inicialmente e retoma o crescimento, até atingir um novo nível. Este efeito

aparece, por exemplo, quando se estuda a sobrevivência de uma espécie após uma mutação adaptativa (Glass, Wilsson e Gottman, 1975).

Existem outros fatores que devem ser considerados quando estudados como efeitos de intervenção. A intervenção pode ser constante ou variável ao longo do período de efeito. Uma única série pode conter mais de uma intervenção, e com diferentes efeitos sobre cada uma; portanto, é possível existir interação entre os pontos de intervenção, confundindo-se parte de seus efeitos.

2.7.4 Estrutura da Função de Transferência

O efeito de intervenção é determinado através da estrutura da função de transferência. Se já for conhecida a forma da função de transferência do modelo e estimados os seus parâmetros, conhece-se o tipo de efeito de intervenção.

No caso de uma série temporal, com seu fenômeno conhecido, sabe-se o que pode ocasionar um efeito de intervenção, o que facilitará a identificação da estrutura da função de transferência.

Na Figura 8 estão apresentados os tipos mais comuns de função de transferência e os tipos de efeito de intervenção. Mais detalhes, consultar Box, Jenkins e Reinsel (1994).

Os modelos estudados, para a função de transferência, têm a forma da equação (27).

Para cada tipo de função de transferência, ajusta-se um modelo $v_i(B)$.

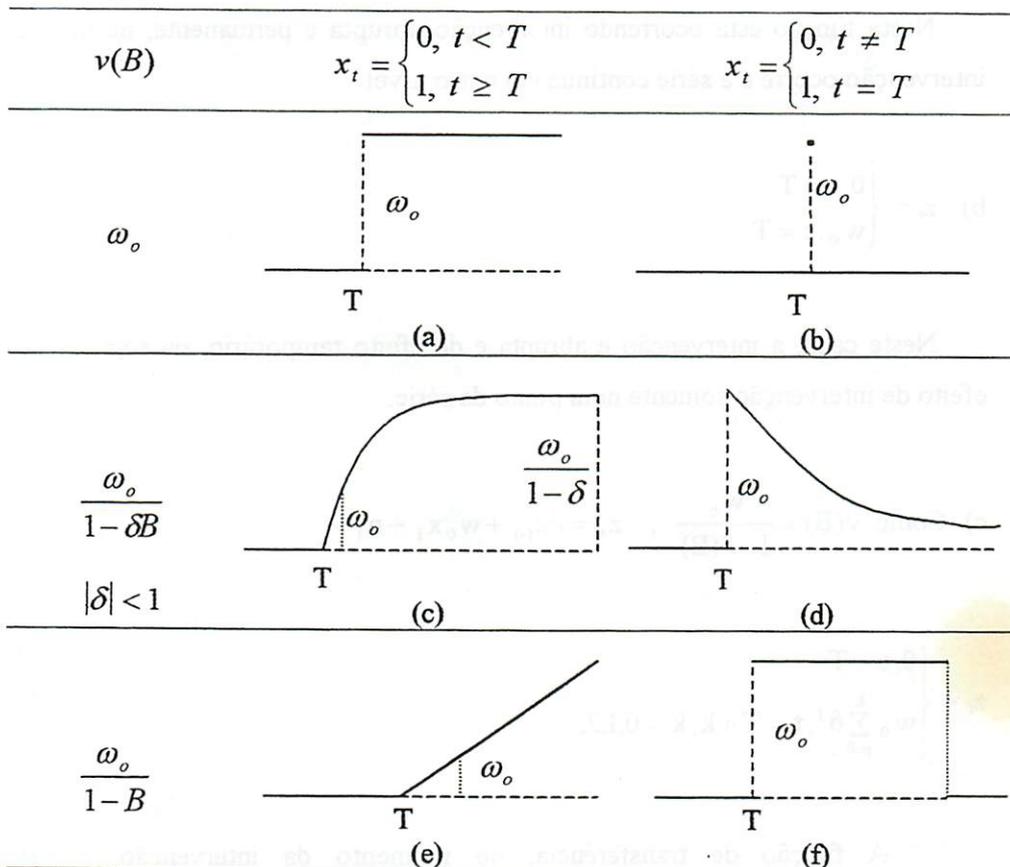


FIGURA 8 – Estrutura da função de transferência

A seguir descrevem-se os casos mais comuns de efeitos de intervenção e suas funções de transferência, as quais estão apresentadas na Figura 8, considerando-se $z_t = v(B) x_t$

$$a) z_t = \begin{cases} 0, & t < T \\ \omega_o, & t \geq T \end{cases}$$

Nesta função está ocorrendo intervenção abrupta e permanente, na qual a intervenção ocorre e a série continua em outro nível.

$$b) z_t = \begin{cases} 0, t \neq T \\ w_0, t = T \end{cases}$$

Neste caso, a intervenção é abrupta e de efeito temporário, ou seja, existe efeito de intervenção somente num ponto da série.

$$c) \text{ Como } v(B) = \frac{w_0}{1-\delta(B)}, \quad z_t = \delta z_{t-1} + w_0 x_t + n_t$$

$$z_t = \begin{cases} 0, t < T \\ w_0 \sum_{j=0}^k \delta^j, t = T+k, k = 0,1,2,\dots \end{cases}$$

A função de transferência, no momento da intervenção, atinge gradualmente um nível e continua permanentemente neste nível.

$$d) z_t = \begin{cases} 0, t < T \\ \delta^k w_0, t = T+k, k = 0,1,2,\dots \end{cases}$$

A série muda abruptamente de nível após a intervenção, e decai exponencialmente para o nível de antes da intervenção.

$$d) z_t = \begin{cases} 0, t < T \\ (k+1)w_0, t = T+k, k = 0,1,2,\dots \end{cases}$$

A série após a intervenção tornou-se não estacionária, com manifestação gradual e tendência determinística.

$$f) z_t = \begin{cases} 0, & t < T \\ w_0, & t \geq T \end{cases}$$

Este caso é o mesmo que ocorreu no item c), no qual a série sofre intervenção abrupta, e segue neste novo nível até a assíntota $\frac{w_0}{1-B}$.

2.7.5 Aplicação de Análise de Intervenção em Leis de Trânsito

Bhattacharrya e Layton (1979) utilizaram a análise de intervenção para avaliar o efeito da legislação do uso de cinto de segurança, em automóveis, sobre o número de mortes por acidentes rodoviários, no estado de Queensland (Austrália).

A série temporal foi analisada no período de janeiro de 1950 a abril de 1976, com dados coletados trimestralmente, totalizando 108 observações.

O modelo proposto considerou três fatores que podem ser considerados intervenção, apresentados abaixo:

- I₁) 01/01/1969, tomou-se obrigatória a instalação de cintos de segurança dianteiros nos veículos novos;
- I₂) 01/01/1971, tomou-se obrigatória a instalação de cinto de segurança traseiros e dianteiros nos veículos novos;
- I₃) 01/01/1972, tomou-se obrigatória a utilização do cinto de segurança nos veículos em que tivessem sido instalados previamente.

Através da série temporal (Figura 9), pode-se observar os pontos que foram considerados como intervenção pelos autores, representados por I₁, I₂ e I₃.

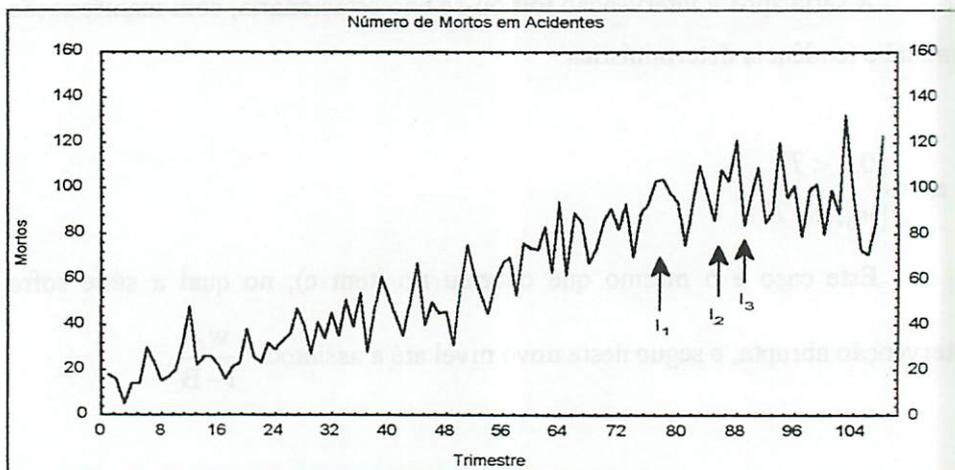


FIGURA 9 – Número de mortos por acidentes nas rodovias de Queensland

Ajustou-se o resíduo do modelo univariado para a série

$$(1 - B^4)z_t = \theta_0 + (1 + \theta_4 B^4)(1 + \theta_3 B^3 + \theta_5 B^5)a_t$$

Foram considerados dois pontos de intervenção na série, sendo a função de transferência escrita conforme a Figura 8, tendo a seguinte estrutura

$$v_1(B) = \frac{w_1(1 - B^4)}{1 - \delta B} \quad ; \quad v_2 = w_2$$

e os efeitos de intervenção dado pela Figura 7, tendo a forma

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

Verifica-se que os efeitos de intervenção propostos são de efeito gradual e duração permanente, mas podendo ajustar várias outras possibilidades.

Neste modelo, \hat{w}_1 foi considerado não significativo, excluindo este efeito do modelo, e \hat{w}_2 é estatisticamente significativo, conforme a Tabela 1.

TABELA 1 – Estimativas e Desvios-padrão da Série de Uso de Cinto de Segurança, com Erro Quadrático Médio 11,00.

Parâmetro	Estimativa	Desvio-Padrão
θ_0	0,20	0,22
θ_3	-0,17	0,10
θ_4	-0,89	0,05
θ_5	0,32	0,10
w_2	-5,49	1,15

Conclui-se que a nova lei de trânsito, tornando obrigatória ao motorista a utilização do cinto de segurança, diminui o número de acidentes fatais nas rodovias da Austrália.

Outros modelos foram considerados pelos autores, mas esse foi o que melhor se ajustou aos dados.

Existem outros trabalhos na literatura utilizando análise de intervenção. Pino (1980) faz aplicações em economia agrícola.

2.7.6 Testes de Intervenção

Existem alguns testes estatísticos, paramétricos ou não-paramétricos, que podem ser aplicados para verificar o efeito de intervenção. Apresentam-se os seguintes testes: teste t de Student e o teste de qui-quadrado de Box e Tiao (1976). Pode-se consultar o modelo linear geral apresentado por Glass, Willson e Gottman (1975).

2.7.6.1 Teste t de Student

Se as observações forem independentes e a série não apresentar tendência ou sazonalidade, pode ser aplicado o teste t de Student.

$$t_{\text{calc}} = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \quad (28)$$

onde

\bar{x}_1 : a média da série antes da intervenção;

\bar{x}_2 : a média da série após a intervenção;

s : desvio-padrão amostral;

N_1 : o número de observações antes da intervenção;

N_2 : o número de observações após a intervenção;

t : a estatística do teste

A hipótese a ser testada é

$$H_0 : \bar{x}_1 = \bar{x}_2$$

$$H_1 : \bar{x}_1 \neq \bar{x}_2$$

Se $t_{\text{calc}} > t_{\text{tab}}$, a série tem intervenção no instante t.

Se a hipótese H_0 é aceita, o efeito de intervenção no instante t não é significativo, ou seja, não existe intervenção neste instante da série.

Se as observações não forem independentes, as situações para verificar a intervenção podem ser aplicadas aleatoriamente; portanto, a distribuição t de Student pode ser aplicada normalmente como sendo uma aproximação, tendo base válida para a inferência.

Na prática, as observações de uma série temporal são quase sempre não-estacionárias e dependentes, não sendo possível aleatorizar, invalidando os procedimentos usuais do teste t de Student.

2.7.6.2 Teste de Qui-Quadrado

É sugerida a utilização deste procedimento em séries não estacionárias.

Este método segue os seguintes passos:

- Identificar e estimar o modelo para a parte anterior à intervenção da série;
- Fazer previsões para o modelo estimado, que serão comparadas com os valores reais da série, para verificação da intervenção;
- Calculam-se funções apropriadas dos erros de previsões, que possam indicar possíveis mudanças no modelo postulado.

Para um modelo ARIMA escrito na forma

$$z_t = \sum_{j=0}^{\infty} \Psi_j a_{t-j} \quad , \Psi_0 = 1$$

a previsão do erro quadrático médio mínimo é dada por

$$\hat{z}_t(\ell) = \sum_{j=0}^{\infty} \Psi_{\ell+j} a_{t-j}$$

e o erro de previsão dado por

$$e_t(\ell) = \sum_{j=0}^{\ell-1} \Psi_j a_{t+\ell-j}$$

com a_t sendo o erro de previsão um passo a frente, ou seja, os erros de previsão a um passo adiante são não-correlacionados.

Considere-se a seguinte estatística

$$\hat{Q}_q = \sum_{j=1}^m \frac{a_j^2}{\sigma_a^2} \quad (29)$$

em que

$$a_j = y_j - \hat{y}_{j-1}(1) \quad , j = 1, \dots, m;$$

σ_a^2 : variância amostral;

\hat{Q}_q : estatística do teste de Qui-Quadrado

\hat{Q}_q tem distribuição χ^2 com m graus de liberdade.

Se $\hat{Q}_q > \chi^2_m$, o ponto de intervenção testado é estatisticamente significativo, ou seja, existe efeito de intervenção na série temporal.

Este teste só é válido para tamanhos de amostra grandes; se o número de observações é pequeno, o teste torna-se inadequado.

3 MATERIAL E MÉTODOS

As séries que foram utilizadas para análise foram obtidas através da internet, pela página www.sptrans.com.br.

A primeira série é relacionada com o número de passageiros por ônibus urbanos que circulam no município de São Paulo, com medições mensais no período analisado de janeiro de 1983 a dezembro de 1998, totalizando 192 observações.

A segunda série utilizada neste trabalho é sobre o número de assaltos nos ônibus urbanos que circulam no município de São Paulo, com medições mensais no período analisado de janeiro de 1992 a agosto de 1999, totalizando 93 observações.

A terceira série analisada está associada ao número de acidentes de trânsito nos ônibus urbanos que circulam no município de São Paulo, com medições mensais no período de janeiro de 1988 a setembro de 1999, totalizando 141 observações.

Os parâmetros foram estimados considerando modelos de Box e Jenkins, acrescentando-se parâmetros para os fenômenos ocorridos nos períodos de suas medições, denominados intervenção. Para cada série foram propostos mais de um modelo, para que possam ser analisadas as vantagens e desvantagens de cada modelo, baseadas no critério de Akaike e no quadrado médio residual.

A metodologia aplicada para ajustar modelos a séries reais, relacionadas aos ônibus urbanos do município de São Paulo, foi a seguinte:

1. Descrição do comportamento do conjunto de dados através da construção de gráficos e diagramas de dispersão, obtenção de estatísticas descritivas,

verificação da existência de tendência, sazonalidade e intervenção através de testes específicos;

- 2. Fazer as estimativas da tendência, sazonalidade e intervenção através do método da Máxima Verossimilhança;**
- 3. Ajuste conveniente de modelos a partir das características da série;**
- 4. Fazer previsões para a série a partir de valores passados, comparando a eficiência da previsão através do erro quadrático médio.**

Os “softwares” utilizados para realizar as análises foram os seguintes: STATISTICA® (1995) e SAS® (1999).

4 RESULTADOS E DISCUSSÃO

Nas séries de transporte urbano é importante avaliar o efeito de eventos que ocorreram no período de 1983 a 1999, utilizando os dados das três séries: a primeira série é o “número de passageiros por ônibus”; a segunda série observa o “número de assaltos nos ônibus”; e a terceira série registra o “número de acidentes com os ônibus”.

Para analisar as séries de transporte urbano, aplicam-se as metodologias de análise de intervenção e de séries temporais utilizando os modelos de Box e Jenkins.

4.1 Modelos para a Série: Número de Passageiros por Ônibus

Esta série relata o número médio mensal de passageiros por ônibus que circulam dentro da cidade de São Paulo, no período de janeiro de 1983 a dezembro de 1998, e seus dados estão apresentados no Anexo A1. A análise foi realizada retirando-se as observações no período de julho a dezembro de 1998, para que se pudesse comparar a previsão com os valores reais dentro deste período, sendo que a série analisada contém 186 observações mensais e está apresentada na Figura 10.

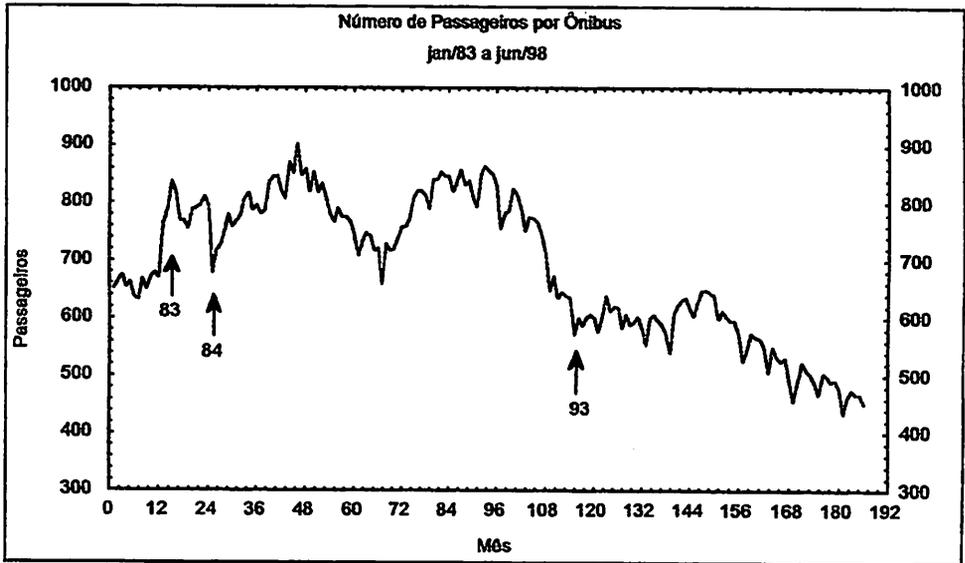


FIGURA 10 – Série do número médio mensal de passageiros por ônibus

Na Figura 10, pode-se observar que a partir da 115ª observação, referente a janeiro de 1993, a série decaiu para um novo nível. Esta mudança de nível pode ter ocorrido por causa de uma intervenção ou estar confundida com a tendência. Para uma análise mais detalhada, é necessário verificar se a série apresenta tendência e (ou) sazonalidade, a fim de obter uma série estacionária, para que se ajustem modelos e se analisem os possíveis efeitos de intervenção da série.

Para verificar a sazonalidade e a tendência, é necessário aplicar o teste de Fisher e o teste dos sinais de Cox-Stuart, respectivamente.

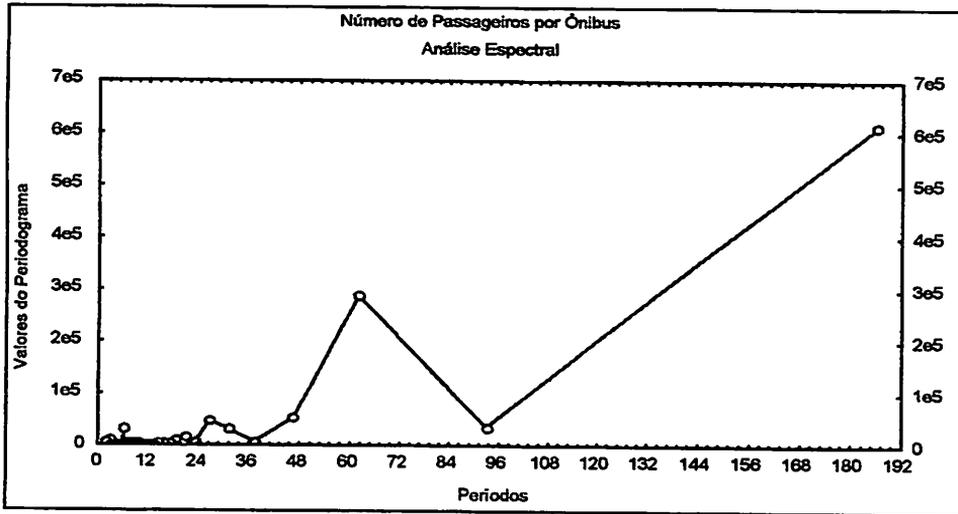


FIGURA 11 – Análise Espectral da série do número de passageiros por ônibus

A Figura 11 fornece os possíveis períodos cíclicos e sazonais que a série possui. Toda periodicidade acima do 12º período é considerada ciclo, sendo uma componente de difícil ajuste no modelo, nota-se uma possível periodicidade no 62º período; mas retirando-se o efeito desse ciclo, a série perderia 62 observações para o ajuste do modelo, reduzindo bastante o tamanho da série; quanto ao período sazonal, pode-se observar uma pequena periodicidade no 6º período, ou seja, pode existir um aumento ou decréscimo no número de passageiros no período de férias. Portanto, o teste da sazonalidade é aplicado no período semestral.

Através do teste do periodograma proposto por Fisher, testa-se a existência de efeito sazonal no período de 6 meses, considerando as estatísticas

$$g = 0,026 \qquad z = 1,957$$

Como $g < z$, não existem evidências de sazonalidade no período de 6 meses, testadas no nível de 5% de significância.

Através do teste dos sinais, testa-se o efeito de tendência na série.
Considerando as estatísticas

$$c = \frac{186}{2} = 93 \qquad n' = 93$$

$$T = 13 \qquad t = \frac{1}{2}(93 + 1,96 \cdot \sqrt{93}) = 55,95$$

Como $T < t$, pode-se afirmar que a série apresenta tendência, sendo necessário utilizar o método das diferenças para que se retire esta componente da série, a fim de que se obtenha uma série estacionária e que para esta possa ser identificado um melhor modelo.

4.1.1 Identificação

Para identificar o modelo da série, é necessário fazer a FAC e FACP para a série estacionária; como foi constatada tendência na série, é necessário retirar a tendência da série aplicando a primeira diferença. Apresenta-se na Figura 12 a série diferenciada FAC e FACP.

Na Figura 12, observa-se que após retirar a tendência, a série apresenta média próxima de zero e variância constante, tendo fortes indícios de estacionaridade. Verificando as FAC e FACP, observa-se que a série não tem comportamento sazonal, comprovado através do teste de Fisher, apresentando apenas correlações no período semestral, sendo necessário acrescentar os efeitos sazonais autorregressivos e de médias-móveis na série.

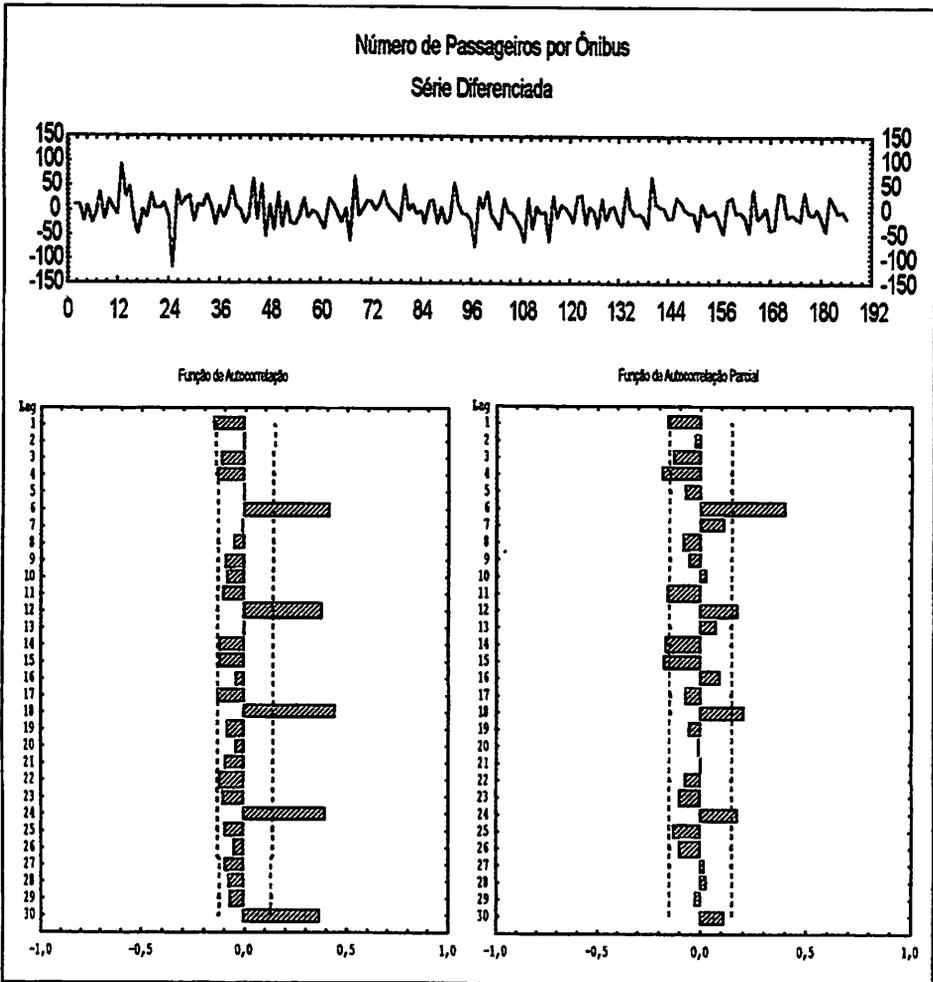


FIGURA 12 – Série diferenciada e as FAC e FACP, da série do número de passageiros por ônibus

Através da série diferenciada (Figura 12), observam-se dois picos que podem ter efeitos de intervenção, sendo o primeiro no “lag” 13, representado por $x_{1,t}$, em Janeiro de 1984, e o segundo no “lag” 25, representado por $x_{2,t}$, em Janeiro de 1985. Considera-se também o possível efeito de intervenção na

privatização da C.M.T.C., ocorrida em janeiro de 1993, sendo representado por $x_{3,t}$, correspondendo ao “lag” 115.

A partir de 1983, a gestão do serviço de transporte público voltou a ser alvo da atenção da administração municipal, melhorando a qualidade do transporte urbano, conduzindo a um incentivo na utilização dos ônibus e aumentando o número de passageiros. O efeito de intervenção ocorreu no primeiro mês de 1984.

No segundo semestre de 1984 houve um acréscimo no número de ônibus em circulação, em decorrência da circulação da linha operadora de gás metano, provocando uma queda no número de passageiros por ônibus, tendo esta queda efeito de intervenção nas férias de janeiro de 1985.

No período de privatização, em janeiro de 1993, representada pelo “lag” 115, observada através da Figura 10, verifica-se que existe um decréscimo no número de passageiros por ônibus, o qual pode ter ocorrido pelo fato da inclusão de várias empresas privadas, que entram em circulação na cidade, aumentando o número da frota de ônibus.

Verificando a FAC e a FACP, propõem-se dois modelos univariados:

- a) Como a FACP apresenta correlação no “lag” 1 e a FAC apresenta correlação sazonais nos “lags” 6, 12, 18 e 24, indica-se o modelo SARIMA(1,1,0)x(0,0,4)₆

$$(1 - \phi_1 B)(1 - B)z_t = (1 - \Theta_1 B^6 - \Theta_2 B^{12} - \Theta_3 B^{18} - \Theta_4 B^{24})a_t$$

- b) Também é interessante propor o modelo SARIMA(1,1,0)x(2,0,0)₆, pois estes “lags” apresentam correlação significativa.

$$(1 - \phi_1 B)(1 - \Phi_1 B^6 - \Phi_2 B^{12})(1 - B)z_t = a_t$$

Para propor os modelos de intervenção, é necessário verificar se os três pontos considerados como possíveis intervenções têm efeito significativo.

A série diferenciada nos mostra que possivelmente as intervenções propostas para $x_{1,t}$ e $x_{2,t}$ têm efeito significativo, não ocorrendo o mesmo com $x_{3,t}$, sendo que a mudança de nível da série está confundida com a tendência. Portanto, foram considerados, no modelo, os efeitos de intervenção ocorridos em janeiro de 1984 $x_{1,t}$ e janeiro de 1985 $x_{2,t}$.

Os modelos de intervenção propostos têm efeito abrupto e duração temporária para $x_{1,t}$ e $x_{2,t}$, sendo a função de transferência de efeito significativo apenas no primeiro "lag". Com referência à Figura 7, item (d), descreveu-se o efeito de intervenção, e segundo a Figura 8, item (b), propõe-se o modelo de função de transferência,

$$v = w_0$$

$$x_{1,t} = \begin{cases} 1, & T = 13 \\ 0, & T \neq 13 \end{cases} \quad x_{2,t} = \begin{cases} 1, & T = 25 \\ 0, & T \neq 25 \end{cases}$$

Os modelos de intervenção representados por c) e d) têm seus resíduos descritos através dos modelos univariado dos itens a) e b), respectivamente.

$$c) \quad z_t = w_{1,0}x_{1,t} + w_{2,0}x_{2,t} + \frac{(1 - \Theta_1 B^6 - \Theta_2 B^{12} - \Theta_3 B^{18} - \Theta_4 B^{24})}{(1 - B)(1 - \phi_1 B)} a_t$$

$$d) \quad z_t = w_{1,0}x_{1,t} + w_{2,0}x_{2,t} + \frac{1}{(1 - B)(1 - \phi_1 B)(1 - \Phi_1 B^6 - \Phi_2 B^{12})} a_t$$

4.1.2 Estimação

A estimação dos parâmetros para os modelos identificados acima, sendo dois univariados e dois de intervenção, representados através dos itens a), b), c)

e d) respectivamente, estão nas Tabelas 2, 3, 4 e 5. O programa para realizar as estimativas é apresentado no Anexo B1.

TABELA 2 – Estimativas dos Parâmetros do Primeiro Modelo Univariado

Modelo: SARIMA(1,1,0)x(0,0,4)₆				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
ϕ_1	-0,15699	0,07405	-2,12	0,035
Θ_1	-0,26717	0,07436	-3,59	< 0,001
Θ_2	-0,17419	0,07220	-2,41	0,017
Θ_3	-0,30155	0,07589	-3,97	< 0,001
Θ_4	-0,17944	0,07668	-2,34	0,020
Quadrado Médio Residual			590,94	
AIC			1712,95	
Teste de Box e Pierce		Q=22,38 < $\chi_{20}^2=31,41$ (ruído branco)		

TABELA 3 – Estimativas dos Parâmetros do Segundo Modelo Univariado

Modelo: SARIMA(1,1,0)x(2,0,0)₆				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
ϕ_1	-0,11669	0,06455	-1,81	0,072
Φ_1	0,29937	0,07052	4,25	< 0,001
Φ_2	0,25496	0,07167	3,55	< 0,001
Quadrado Médio Residual			603,74	
AIC			1714,49	
Teste de Box e Pierce		Q=40,22 > $\chi_{22}^2=33,93$ (não é ruído branco)		

TABELA 4 – Estimativas dos Parâmetros do Primeiro Modelo de Intervenção

Modelo: SARIMA(1,1,0)x(0,0,4) ₆ w _{1,0} =Janeiro 1984 e w _{2,0} = Janeiro 1985				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
ϕ_1	-0,23402	0,07266	-3,22	0,002
Θ_1	-0,29081	0,07363	-3,95	< 0,001
Θ_2	-0,32570	0,07320	-4,45	< 0,001
Θ_3	-0,33616	0,07540	-4,46	< 0,001
Θ_4	-0,20368	0,07724	-2,64	0,009
w _{1,0}	100,19149	19,58362	5,12	< 0,001
w _{2,0}	-111,55992	18,85788	-5,92	< 0,001
Quadrado Médio Residual			433,85	
AIC			1658,68	
Teste de Box e Pierce		Q=24,82 < χ_{18} =28,87 (ruído branco)		

TABELA 5 – Estimativas dos Parâmetros do Segundo Modelo de Intervenção

Modelo: SARIMA(1,1,0)x(2,0,0) ₆ w _{1,0} =Janeiro 1984 e w _{2,0} = Janeiro 1985				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
ϕ_1	-0,12344	0,05843	-2,11	0,036
Φ_1	0,31161	0,06740	4,62	< 0,001
Φ_2	0,35849	0,06870	5,22	< 0,001
w _{1,0}	101,38419	20,32137	4,99	< 0,001
w _{2,0}	-109,86347	19,26203	-5,70	< 0,001
Quadrado Médio Residual			424,33	
AIC			1652,71	
Teste de Box e Pierce		Q=30,43 < χ_{20} =31,41 (ruído branco)		

As estimativas dos parâmetros para os modelos das Tabelas 2, 3, 4 e 5, testadas através do teste t de Student, foram significativas ao nível de 5% de significância.

Os critérios adotados para a escolha do melhor modelo foram os critérios de Akaike, quadrado médio residual, sendo o modelo que apresentou o menor valor para os dois critérios o modelo SARIMA(1,1,0)x(2,0,0)₆, $w_{1,0}$ =Janeiro 1984 e $w_{2,0}$ = Janeiro 1985, sendo seu resíduo um ruído branco, comprovado através do teste de Box e Pierce. As estimativas deste modelo são as seguintes:

$$z_t = 101,384 x_{1,t} - 109,863 x_{2,t} + \frac{1}{(1-B)(1+0,123B)(1-0,312B^6-0,358B^{12})} a_t$$

Baseado no teste de Box e Pierce, averiguou-se que o segundo modelo univariado, Tabela 3, não é um ruído branco, e a estimativa do parâmetro autorregressivo de ordem 1 não é significativa, não sendo um bom modelo para a série. Os resíduos dos modelos das Tabelas 2, 4 e 5 são i.i.d., sendo então ruído branco.

Através da função de autocorrelação residual do modelo de intervenção escolhido, observa-se o valor da estatística Q, dado pelo teste de Box e Pierce para os modelos propostos. Utilizando o teste de Shapiro-Wilk, testa-se a distribuição residual do modelo de intervenção. Estes aspectos estão detalhados a seguir.

Teste para Ruído Branco

O teste de Box e Pierce é baseado na estrutura da função de autocorrelação residual representada na Figura 13.

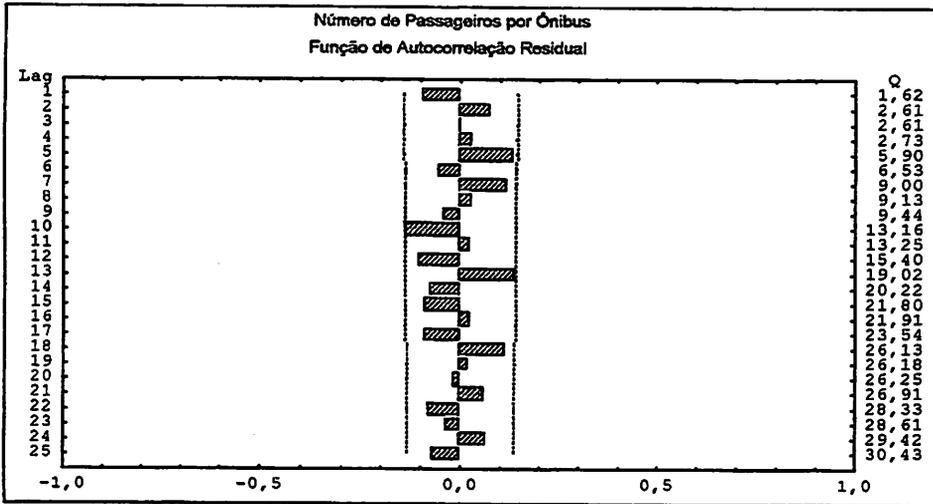


FIGURA 13 – Função de Autocorrelação Residual da série do número de passageiros por ônibus.

Através da análise gráfica da Figura 13, pode-se observar que todos os “lags” são não-correlacionados, tendo fortes evidências de ruído branco.

A estatística de Box e Pierce é dada pelo “lag” 25 da Figura 13, cuja estimativa é

$$Q_c = 30,43$$

O valor da estatística de Box e Pierce é comparado com o valor crítico da distribuição qui-quadrado, no nível de 5% de significância, com 20 graus de liberdade, dado por

$$\chi_{20} = 31,41$$

Como a estimativa da $\chi_{20} > Q_c$, o resíduo do modelo ajustado é um ruído branco, indicando que o resíduo é i.i.d., com média zero e variância constante.

Teste para Normalidade dos Resíduos

Partindo-se da pressuposição de que o resíduo tem distribuição Gaussiana, é preciso aplicar o teste de Shapiro-Wilk a fim de que se verifique o resíduo provém de distribuição normal. Testam-se, então, as hipóteses

H_0 : O resíduo tem distribuição normal

H_1 : O resíduo não tem distribuição normal

A estatística T_s de Shapiro-Wilk e sua probabilidade são:

Shapiro-Wilk	Estimativa	Probabilidade
T_s	0,975365	0,1255

Através da probabilidade da estimativa da estatística de Shapiro-Wilk, pode-se concluir que o resíduo do modelo ajustado provém de uma distribuição normal.

4.1.3 Previsão

As previsões estimadas do número de passageiros por ônibus e respectivos intervalos de confiança de 95%, para o período de julho a dezembro de 1998, são encontradas nas Tabelas 6 e 7.

São propostos o melhor modelo de intervenção (Tabela 6) e o melhor modelo univariado (Tabela 7) para que se possa comparar qual modelo apresenta melhor previsão, utilizando o critério do Erro Quadrático Médio.

TABELA 6 – Previsão para o modelo SARIMA(1,1,0)x(2,0,0)₆ w_{1,0}=Janeiro 1984 e w_{2,0}= Janeiro 1985

Data	Valor Real	Previsão	LI:95%	LS:95%
julho 1998	417	432	392	473
agosto 1998	450	456	403	510
setembro 1998	442	456	391	521
outubro 1998	427	451	377	525
novembro 1998	433	452	369	534
dezembro 1998	408	440	351	531
Erro Quadrático Médio			2418	

TABELA 7 – Previsão para o modelo SARIMA(1,1,0)x(0,0,4)₆

Data	Valor Real	Previsão	LI:95%	LS:95%
julho 1998	417	435	388	483
agosto 1998	450	456	393	518
setembro 1998	442	463	388	537
outubro 1998	427	459	373	544
novembro 1998	433	457	362	552
dezembro 1998	408	449	345	552
Erro Quadrático Médio			4082	

Observa-se que os valores reais da série estão dentro do intervalo de confiança, para todos os meses, nos dois modelos. Através das estimativas do erro quadrático médio, nota-se que o modelo de intervenção fornece melhores previsões, pois apresenta um menor erro quadrático médio, logo seus intervalos de confiança terão amplitude menor do que o do modelo sem intervenção.

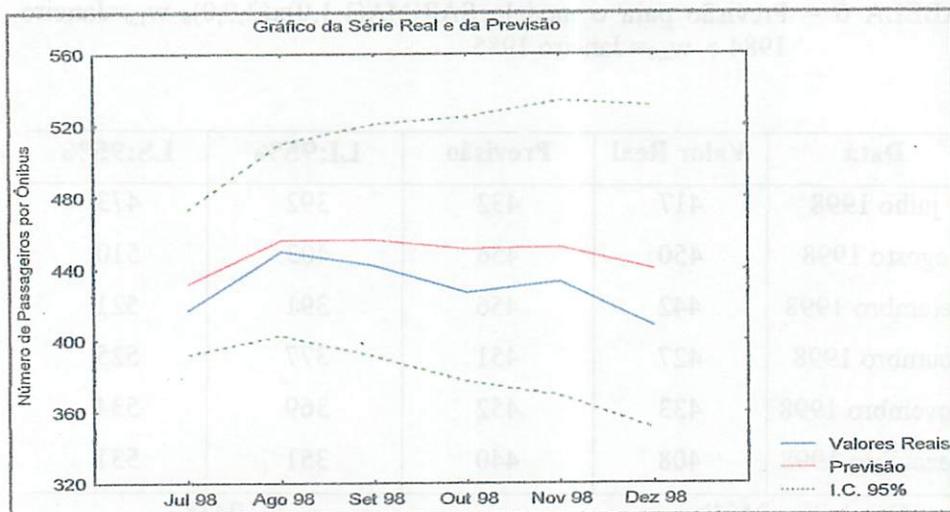


FIGURA 14 – Série Real e de Previsão com os respectivos IC 95%, no período de Julho de 1998 a Dezembro de 1998

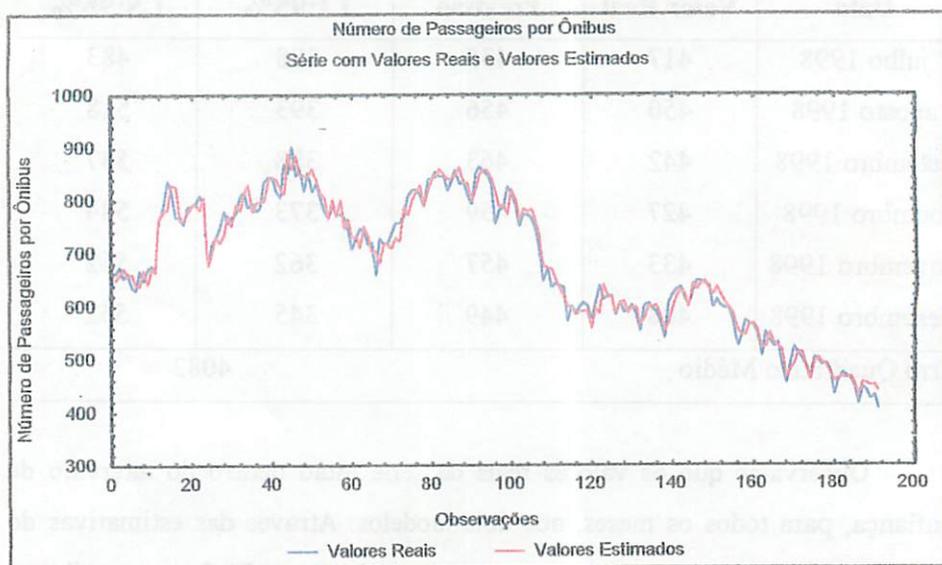


FIGURA 15 – Série Real e Série Estimada para o período de Janeiro de 1983 a Dezembro de 1998

A Figura 14 representa o comportamento da previsão para o modelo de intervenção SARIMA(1,1,0)x(2,0,0)₆ w_{1,0}=Janeiro 1984 e w_{2,0}= Janeiro 1985, em relação aos valores reais da série, no período de Julho a Dezembro de 1998.

A Figura 15 ilustra a série real e a série obtida a partir do modelo SARIMA (1,1,0)x(2,0,0)₆ w_{1,0} e w_{2,0}. Pode-se observar que o modelo escolhido ajustou bem os dados.

Através da análise gráfica das Figuras 14 e 15, observa-se que a previsão superestima os valores reais da série, porém mantém o mesmo comportamento da série real.

4.2 Modelos para a Série: Número de Assaltos nos Ônibus

Esta série fornece o número de assaltos que ocorreram nos ônibus urbanos de São Paulo no período de janeiro de 1992 a fevereiro de 1999, totalizando 86 observações mensais; os dados estão apresentados no Anexo A2. As observações de março de 1999 a agosto de 1999 foram retiradas da série a fim de comparar seus valores com os valores de previsão, estimados dentro desse período.

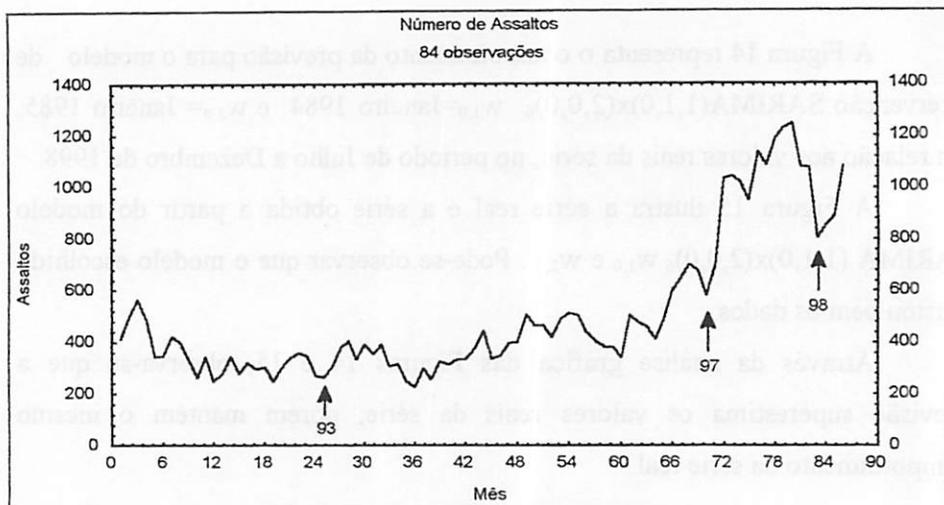


FIGURA 16 – Série do número de assaltos mensais

Na Figura 16 observa-se que a partir da 66ª observação, referente a junho de 1997, o número de assaltos cresce consideravelmente, podendo ser considerada como um possível efeito de intervenção. É preciso salientar que esta mudança de nível pode estar confundida com outras componentes da série.

Em janeiro de 1993, referente à 13ª observação, foi o período em que ocorreu a privatização da C.M.T.C.; aparentemente não houve efeito significativo diminuindo ou aumentando o número de assaltos, pois neste período a série apresenta-se estacionária, ou seja, não muda de nível.

A série temporal aparentemente apresenta uma componente de tendência, sendo necessário aplicar o teste dos sinais para verificar sua presença. O teste de Fisher é proposto para analisar o efeito de sazonalidade da série.

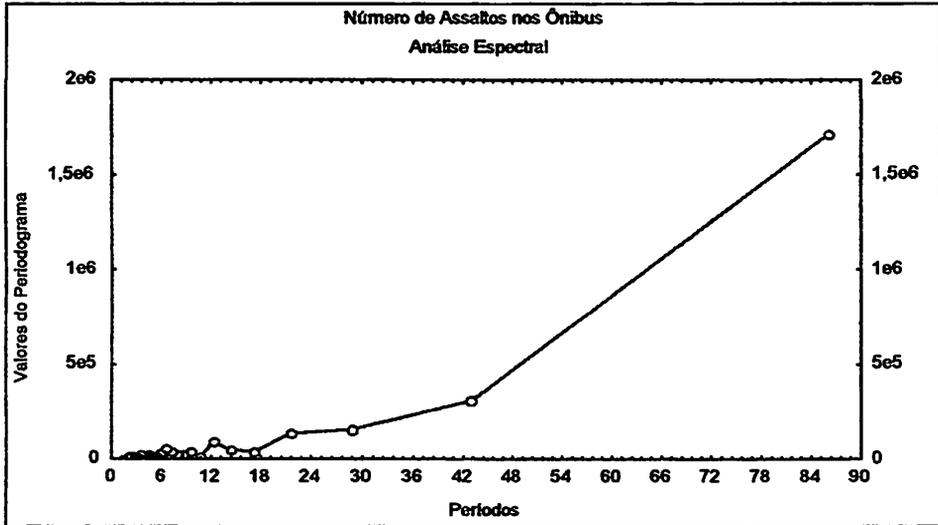


FIGURA 17 – Análise Espectral para a série do número de assaltos nos ônibus

Através da Figura 17, pode-se considerar que a série não apresenta nenhuma periodicidade, tendo fortes evidências de que a série temporal não possua sazonalidade.

No entanto, é possível identificar uma pequena periodicidade em 12 meses, sendo necessário aplicar o teste de Fisher neste período.

Através do teste do periodograma proposto por Fisher, testa-se a existência de efeito sazonal no período de 12 meses, considerando as estatísticas

$$g = 0,033 \qquad z = 1,916$$

Como $g < z$, não existem evidências de sazonalidade no período de 12 meses, testadas ao nível de 5% de significância.

Através do teste dos sinais, testa-se o efeito de tendência na série. Considerando as estatísticas

$$c = \frac{86}{2} = 43 \qquad n' = 43$$

$$T=40 \qquad t = \frac{1}{2}(43 + 1,96 \cdot \sqrt{43}) = 27,93$$

Como $T > n \cdot t$, pode-se concluir que a série apresenta tendência, sendo necessário utilizar o método das diferenças para que se retire esta componente da série, a fim de identificar os modelos a serem propostos.

4.2.1 Identificação

Através da FAC e da FACP da Figura 18, sugerem-se quatro modelos para a série temporal, sendo dois deles univariado e dois de intervenção.

Na Figura 18 observa-se que após retirar a tendência, a série apresenta média em torno de zero e variância constante, sendo que a partir da 70ª observação, referente a outubro de 1997, aumenta a variabilidade da série, podendo ser uma possível intervenção.

Verificam-se as FAC e a FACP e observa-se que não é necessário acrescentar nenhum componente sazonal para os modelos, pois não existe correlação nos “lags” múltiplo de 12.

A FAC apresenta correlação no “lag” 3, sendo necessário compor um modelo com efeito de médias-móveis de ordem 3. A FACP também apresenta correlação no “lag” 3, sendo necessário incluir o efeito autorregressivo de ordem 3.

Considera-se o pico existente na 72ª observação, dezembro de 1997, como um possível ponto de intervenção. É importante salientar que através da Figura 16, levanta-se a hipótese de que a intervenção ocorre na 66ª observação, referente a junho de 1997, porém seu efeito ocorre na 72ª observação, referente a dezembro de 1997. Neste período houve um aumento no número de assaltos nos

ônibus, sendo constatada somente a introdução dos “perueiros” como uma opção para utilizar um meio de transporte.

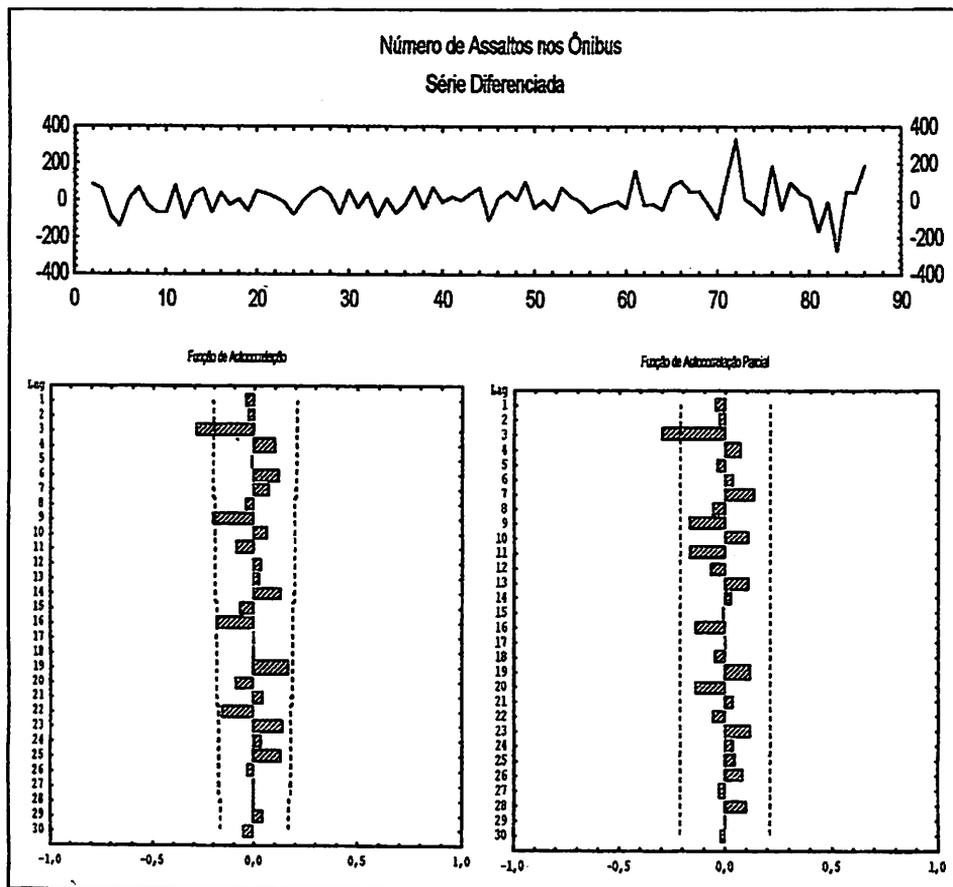


FIGURA 18 – Série diferenciada e as FAC e FACP da série do número de assaltos nos ônibus

No período referente a 83ª observação, novembro de 1998, ocorre uma queda no número de assaltos, sendo proposto como segundo efeito de intervenção. Como não se constatou nenhum evento que possa ter reduzido o

número de assaltos, não foi possível identificar o fenômeno que ocasionou a queda no número de assaltos.

Na época de privatização, em janeiro de 1993, o qual poderia ter ocasionado um acréscimo no número de assaltos nos ônibus, por causa do aumento da frota, não existe efeito de intervenção.

Através da FAC e da FACP, propõem-se dois modelos univariados.

a) Considerando o efeito da FACP, com correlação no “lag” 3, indica-se o modelo ARIMA(3,1,0)

$$(1 - \phi_3 B^3)(1 - B)z_t = a_t$$

b) Verificando a FAC, com correlação no “lag” 3, indica-se o modelo ARIMA(0,1,3)

$$(1 - B)z_t = (1 - \theta_3 B^3)a_t$$

Os dois modelos de intervenção propostos têm efeitos significativo em dezembro de 1997 e em novembro de 1998, que serão representados por $x_{1,t}$ e $x_{2,t}$, respectivamente.

A intervenção ocorrida em dezembro de 1997, representada por $x_{1,t}$, tem efeito abrupto e duração temporária. Em novembro de 1998, representada por $x_{2,t}$, sofre efeito abrupto e duração temporária. Estes efeitos têm a forma do item (d) da Figura 7.

Os dois modelos de intervenção propostos apresentam correlação significativa no primeiro “lag”, sendo que o primeiro torna-se não estacionário após a intervenção. A função de transferência para este modelo é representada pelo item (e) da Figura 8. O segundo modelo é igual ao item (b) da Figura 8 e os dois modelos são apresentados respectivamente.

$$v = \frac{w_0}{1 - B} \quad , \quad v = w_0$$

Os efeitos de intervenção estão apresentados por $x_{1,t}$ e $x_{2,t}$ e dados por

$$x_{1,t} = \begin{cases} 1, & T < 72 \\ 0, & T \geq 72 \end{cases} \quad x_{2,t} = \begin{cases} 1, & T = 83 \\ 0, & T \neq 83 \end{cases}$$

Os modelos de intervenção descritos a seguir têm seus respectivos resíduos descritos através dos modelos univariados dos itens a) e b), respectivamente.

$$c) \quad z_t = \frac{w_{1,0}}{1-B} x_{1,t} + w_{2,0} x_{2,t} + \frac{1}{(1-B)(1-\phi_3 B^3)} a_t$$

$$d) \quad z_t = \frac{w_{1,0}}{1-B} x_{1,t} + w_{2,0} x_{2,t} + \frac{(1-\theta_3 B^3)}{(1-B)} a_t$$

4.2.2 Estimação

As estimativas dos parâmetros para os modelos univariados e de intervenção propostos nos itens a), b), c) e d), respectivamente, estão apresentados nas Tabelas 8, 9, 10 e 11. O programa para realizar as estimativas é apresentado no Anexo B2.

TABELA 8 – Estimativas dos Parâmetros do Primeiro Modelo Univariado

Modelo: ARIMA(3,1,0)				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
ϕ_3	-0,30603	0,10756	-2,85	0,005
Quadrado Médio Residual			6269,83	
AIC			985,71	
Teste de Box e Pierce		Q=22,07 < $\chi_{22}^2=33,93$ (ruído branco)		

TABELA 9 – Estimativas dos Parâmetros do Segundo Modelo Univariado

Modelo: ARIMA(0,1,3)				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
θ_3	0,27531	0,10582	2,60	0,011
Quadrado Médio Residual			6350,68	
AIC			986,74	
Teste de Box e Pierce		Q=22,21 < $\chi_{22}^2=33,93$ (ruído branco)		

TABELA 10 – Estimativas dos Parâmetros do Primeiro Modelo de Intervenção

Modelo: ARIMA(3,1,0) $w_{0,1}$=Dezembro 1997 e $w_{0,2}$=Novembro 1998				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
ϕ_3	-0,25838	0,10930	-2,36	0,021
$w_{1,0}$	40,13787	16,59346	2,42	0,018
$w_{2,0}$	-280,95673	75,37933	-3,73	< 0,001
Quadrado Médio Residual			5420,77	
AIC			975,20	
Teste de Box e Pierce		Q=31,17 < $\chi_{22}^2=33,93$ (ruído branco)		

TABELA 11 – Estimativas dos Parâmetros do Segundo Modelo de Intervenção

Modelo: ARIMA(0,1,3) $w_{0,1}$ =Dezembro 1997 e $w_{0,2}$ =Novembro 1998				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
θ_3	0,30,846	0,11124	2,77	0,007
$w_{1,0}$	41,86364	15,39717	2,72	0,008
$w_{2,0}$	-288,69826	76,14887	-3,79	< 0,001
Quadrado Médio Residual			5383,60	
AIC			974,71	
Teste de Box e Pierce		$Q=32,22 < \chi_{22}^2=33,93$ (ruído branco)		

Aplicando-se o teste t de Student nos modelos apresentados nas Tabelas 8, 9, 10 e 11, verifica-se que os efeitos das estimativas são todas significativas. Os modelos são incompletos, sendo as estimativas dos parâmetros autorregressivos e de médias móveis de ordens 1 e 2 não significativas, e devido a este fato foram excluídas do modelo.

Os critérios para a escolha do melhor modelo são baseados no QMR e no critério de Akaike, sendo escolhido o que obtiver a menor estimativa. De acordo com os critérios adotados, o melhor modelo é o ARIMA(0,1,3) $w_{1,0}$ =Dezembro 1997 e $w_{2,0}$ =Novembro 1998.

O modelo proposto para a série é

$$z_t = \frac{41,864}{1-B} x_{1,t} - 288,698 x_{2,t} + \frac{(1-0,308 B^3)}{(1-B)} a_t$$

Através do teste de Box e Pierce, realizados conforme Tabelas 8, 9, 10 e 11, os quatro modelos apresentam independência residual no modelo, portanto são considerados ruído branco.

Através do valor de Q dado pelo Teste de Box e Pierce, observa-se a função de autocorrelação residual. Utilizando o teste de Shapiro-Wilk, testa-se a distribuição residual do modelo de intervenção. Estes aspectos estão detalhados a seguir.

Teste para Ruído Branco

Figura 19 apresenta a função de autocorrelação residual do modelo escolhido.

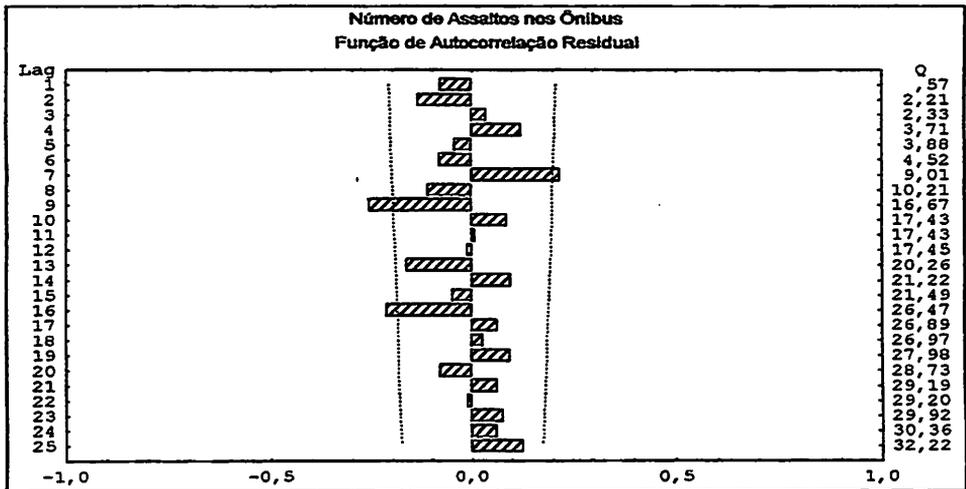


FIGURA 19 – Função de Autocorrelação Residual para a série do número de assaltos nos ônibus

O valor da estatística de Box e Pierce é dado através do “lag” 25 da Figura 19, cuja estimativa foi

$$Q_c = 32,22$$

Esta estatística é comparada com o valor crítico encontrado da distribuição qui-quadrado, no nível de 5% de significância, com 22 graus de liberdade, $\chi_{22}=33,93$.

Como a estimativa da $\chi_{22} > Q_c$, o resíduo do modelo ajustado é um ruído branco, sendo o resíduo considerado i.i.d., com média zero e variância constante.

Teste para Normalidade dos Resíduos

Para conhecer a distribuição residual do modelo, aplica-se o teste de Shapiro-Wilk.

Sob a hipótese

H_0 : Distribuição Normal

H_1 : Não é Distribuição Normal

A estatística do teste é representada por T_s , e sua respectiva probabilidade é fornecida abaixo

Shapiro-Wilk	Estimativa	Probabilidade
T_s	0,9772	0,4415

Pode-se concluir, através da probabilidade da estatística do teste, que o resíduo do modelo provém de uma distribuição normal.

4.2.3 Previsão

O melhor modelo univariado e o melhor modelo de intervenção foram submetidos à previsão para que se possa analisar, através do erro quadrático médio, qual dos modelos fornece melhor previsão.

As previsões realizadas no período de março a agosto de 1999, fornecidas através das Tabelas 11 e 12, são submetidas a um intervalo de confiança de 95%, comparando se os valores reais estão dentro deste padrão de confiança. A Tabela 11 está representada pelo modelo de intervenção, e a Tabela 12 através do modelo univariado.

TABELA 11 – Previsão para o modelo ARIMA(0,1,3) $w_{1,0}$ =Dezembro 1997 e $w_{2,0}$ =Novembro 1998

Data	Valor Real	Previsão	LI:95%	LS:95%
março 1999	1094	1102	959	1246
abril 1999	1255	1106	902	1309
maio 1999	1214	1063	814	1312
junho 1999	1042	1063	795	1331
julho 1999	933	1063	777	1349
agosto 1999	856	1063	760	1366
Erro Quadrático Médio			105256	

TABELA 12 – Previsão para o modelo ARIMA(3,1,0)

Data	Valor Real	Previsão	LI:95%	LS:95%
março 1999	1094	1071	916	1227
abril 1999	1255	1058	839	1278
maio 1999	1214	1000	731	1269
junho 1999	1042	1004	715	1294
julho 1999	933	1008	699	1371
agosto 1999	856	1026	699	1353
Erro Quadrático Médio			121103	

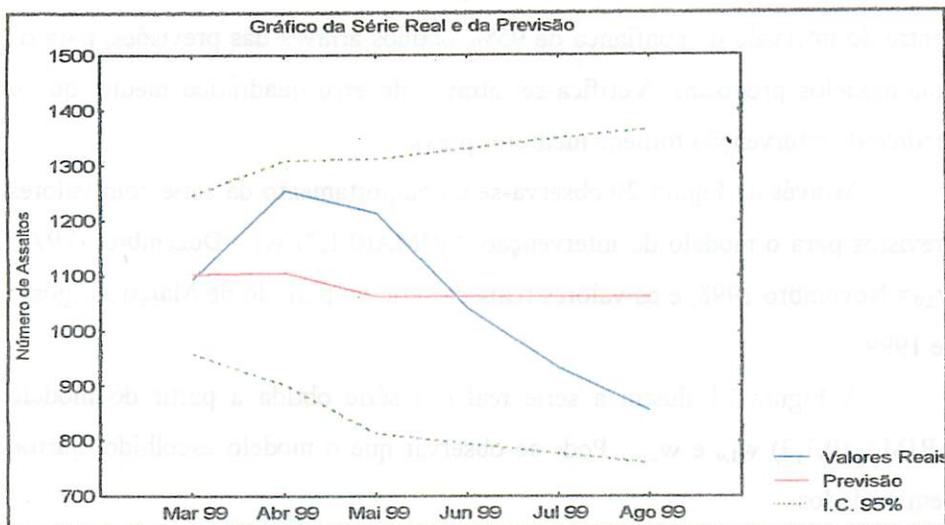


FIGURA 20 – Série Real e de Previsão com seus respectivos IC 95%, no período de Março de 1999 a Agosto de 1999

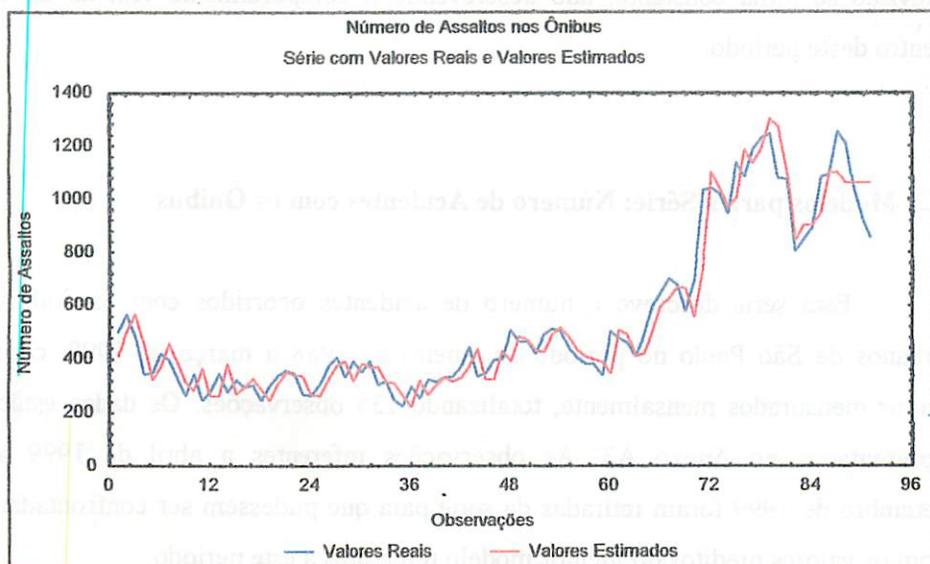


FIGURA 21 – Série Real e Série Estimada para o período de Janeiro de 1992 a Agosto de 1999

Nas Tabelas 11 e 12, observa-se que os valores reais da série estão todos dentro do intervalo de confiança de 95%, obtidos através das previsões, para os dois modelos propostos. Verifica-se, através do erro quadrático médio, que o modelo de intervenção fornece melhores previsões.

Através da Figura 20 observa-se o comportamento da série com valores previstos para o modelo de intervenção $ARIMA(0,1,3)$ $w_{1,0}$ =Dezembro 1997 e $w_{2,0}$ = Novembro 1998, e os valores reais da série no período de Março a Agosto de 1999.

A Figura 21 ilustra a série real e a série obtida a partir do modelo $ARIMA(0,1,3)$ $w_{1,0}$ e $w_{2,0}$. Pode-se observar que o modelo escolhido ajustou bem os dados.

A Figura 21 apresenta a série original e a série obtida a partir do modelo ajustado.

Através das Figuras 20 e 21, observa-se que a partir de maio de 1999 a previsão se torna constante, não descrevendo o comportamento real da série dentro deste período.

4.3 Modelos para a Série: Número de Acidentes com os Ônibus

Esta série descreve o número de acidentes ocorridos com os ônibus urbanos de São Paulo no período de janeiro de 1988 a março de 1999, com dados mensurados mensalmente, totalizando 135 observações. Os dados estão apresentados no Anexo A3. As observações referentes a abril de 1999 a setembro de 1999 foram retiradas da série para que pudessem ser confrontadas com os valores preditos do melhor modelo referentes a este período.

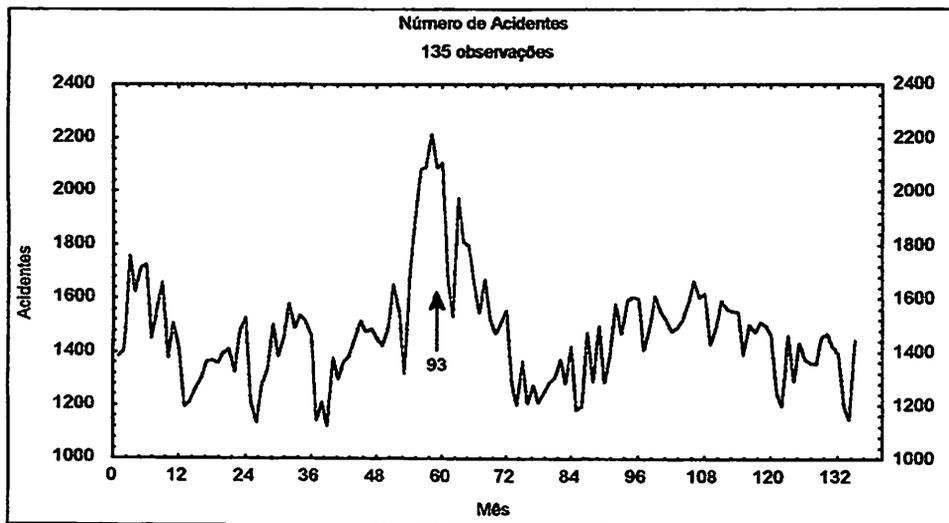


FIGURA 22 – Série do número de acidentes mensais com os ônibus urbanos

Na Figura 22, pode-se observar que em torno da 58ª observação, outubro de 1992, ocorreu um acréscimo no número de acidentes. A série apresentava em torno de 1400 acidentes mensais em janeiro de 1988, e no início de 1993 este índice aumentou para mais de 2000 acidentes mensais. Portanto é necessário verificar ou identificar o motivo deste acréscimo.

Através da análise gráfica da Figura 22, não é possível afirmar a presença de tendência e sazonalidade na série, sendo necessário verificar, através da análise espectral, a existência de periodicidades, e se existir, em que período ocorrem.

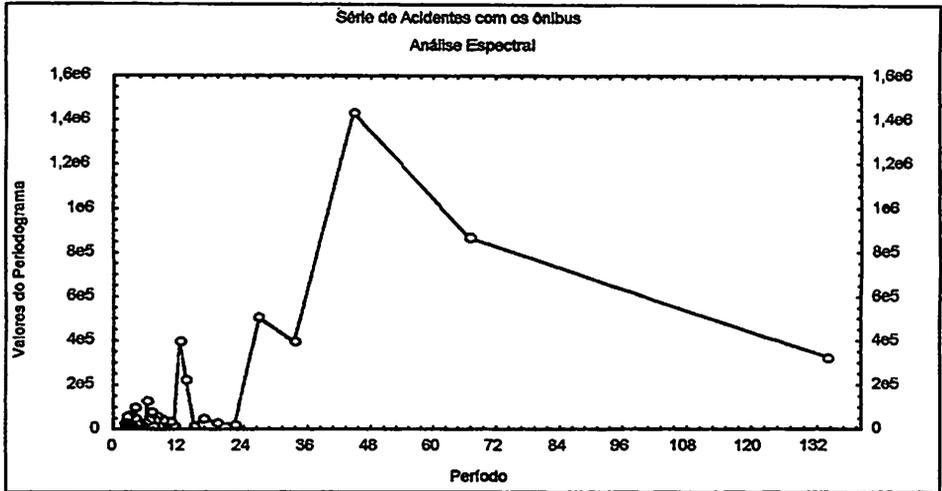


FIGURA 23 – Análise Espectral para a série de acidentes com os ônibus

A Figura 23 apresenta-se com um período elevado em 12 meses, sendo possível a existência de sazonalidade. As periodicidades observadas acima deste intervalo não serão consideradas para análise.

Através do teste do periodograma proposto por Fisher, testa-se a existência de efeito sazonal no período de 12 meses, considerando as estatísticas

$$g = 0,078 \qquad z = 1,90$$

Como $g < z$, não existem evidências de sazonalidade no período de 12 meses, testadas ao nível de 5% de significância.

Através do teste dos sinais, testa-se o efeito de tendência na série. Considerando as estatísticas

$$c = \frac{134}{2} = 67 \qquad n' = 67$$

$$T = 31 \qquad t = \frac{1}{2}(67 + 1,96 \cdot \sqrt{67}) = 41,52$$

Como $T > n^{\frac{1}{2}}$, a série apresenta tendência, sendo necessário utilizar o método das diferenças para retirar esta componente da série.

4.3.1 Identificação

A Figura 24 apresenta a série diferenciada, da qual foram retiradas a tendência, a FAC e a FACP.

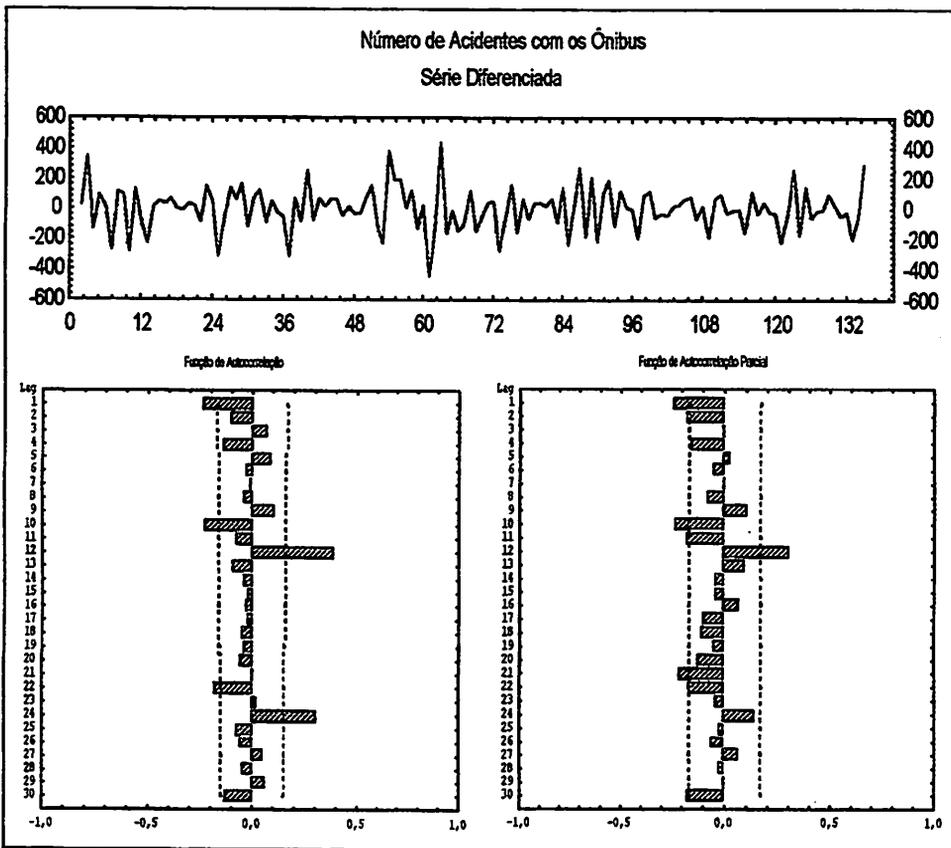


FIGURA 24 – Série diferenciada e as FAC e FACP para a série do número de acidentes com os ônibus

Através da Figura 24, observa-se que a série apresenta média em torno de zero e variância constante, podendo ser estacionária.

Analisando a FAC e a FACP, observa-se que a série apresenta correlação nos “lags” múltiplos de 12, sendo necessário acrescentar uma componente sazonal nos modelos que serão propostos. Além das componentes sazonais, existe correlação no primeiro “lag” das FAC e FACP, apresentando estruturas autorregressivas e de médias-móveis de ordem 1.

Considera-se como possível efeito de intervenção a 62ª observação, referente a março de 1993, sendo representada por $x_{1,t}$. Este fato ocorre no período da privatização da C.M.T.C., ou seja, na época em que ocorreu a privatização, o número de acidentes aumentou consideravelmente, como pode ser visto na Figura 22. Com os ônibus das empresas privadas circulando nas ruas de São Paulo, aumentou o número das frotas de ônibus.

Propõem-se dois modelos univariados, com referência na FAC e FACP.

- a) O primeiro modelo tem uma estrutura de médias-móveis, com correlação nos “lags” 12 e 24; portanto indica-se o modelo SARIMA(0,1,1)x(0,0,2)₁₂

$$(1 - B)z_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12} - \Theta_2 B^{24})a_t$$

- b) Um modelo com estrutura autorregressiva também é indicado, SARIMA(1,1,0)x(0,0,2)

$$(1 - \phi_1 B)(1 - B)z_t = (1 - \Theta_1 B^{12} - \Theta_2 B^{24})a_t$$

O efeito de intervenção que ocorreu em março de 1993, representado por $x_{1,t}$, no período de privatização da C.M.T.C., tem efeito abrupto e duração temporária, pode ser melhor visualizado através do item d) da Figura 7 e definido por

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

A estrutura da função de transferência tem a forma descrita abaixo, (item b) Figura 8);

$$v = w_0$$

Através das estruturas conhecidas, propõem-se dois modelos de intervenção.

- c) O primeiro modelo de intervenção tem o resíduo representado através do modelo univariado proposto através do item b)
- d) Este modelo de intervenção tem o resíduo SARIMA(1,1,0)x(0,0,1)

$$z_t = w_0 x_{1,t} + \frac{(1 - \Theta_1 B^{12})}{(1 - B)(1 - \phi_1 B)} a_t$$

4.3.2 Estimação

As estimativas dos parâmetros, realizadas através do método da máxima verossimilhança, para os modelos dos itens a), b), c) e d), respectivamente, estão apresentados nas Tabelas 13, 14, 15 e 16. O programa para realizar as estimativas está no Anexo B3.

TABELA 13 – Estimativas dos Parâmetros do Primeiro Modelo Univariado

Modelo: SARIMA(0,1,1)x(0,0,2)₁₂				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
θ_1	0,23654	0,08110	2,92	0,004
Θ_1	-0,32125	0,08498	-3,78	< 0,001
Θ_2	-0,21943	0,08735	-2,51	0,013
Quadrado Médio Residual			15905,65	
AIC			1681,86	
Teste de Box e Pierce		Q=24,99 < $\chi_{22}^2=33,93$ (ruído branco)		

TABELA 14 – Estimativas dos Parâmetros do Segundo Modelo Univariado

Modelo: SARIMA(1,1,0)x(0,0,2)₁₂				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
ϕ_1	-0,24298	0,08510	-2,86	0,005
Θ_1	-0,34092	0,08553	-3,99	< 0,001
Θ_2	-0,26746	0,09058	-2,95	0,004
Quadrado Médio Residual			15678,75	
AIC			1680,43	
Teste de Box e Pierce		Q=22,49 < $\chi_{22}^2=33,93$ (ruído branco)		

TABELA 15 – Estimativas dos Parâmetros do Primeiro Modelo de Intervenção

Modelo: SARIMA(1,1,0)x(0,0,2)₁₂ w₀= Março de 1993				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
ϕ_1	-0,21998	0,08619	-2,55	0,012
Θ_1	-0,34747	0,08634	-4,02	< 0,001
Θ_2	-0,25814	0,09144	-2,82	0,006
w ₀	292,09531	111,09721	2,63	0,010
Quadrado Médio Residual			15022,69	
AIC			1675,59	
Teste de Box e Pierce		Q=16,83 < $\chi_{21}^2=32,68$ (ruído branco)		

TABELA 16 – Estimativas dos Parâmetros do Segundo Modelo de Intervenção

Modelo: SARIMA(1,1,0)x(0,0,1)₁₂ w₀= Março de 1993				
Parâmetro	Estimativa	Erro-Padrão	Teste t	p-value
ϕ_1	-0,22964	0,08603	-2,67	0,009
Θ_1	-0,30876	0,08748	-3,53	0,001
w ₀	310,24258	118,69779	2,61	0,010
Quadrado Médio Residual			16072,60	
AIC			1682,27	
Teste de Box e Pierce		Q=28,88 < $\chi_{22}^2=33,93$ (ruído branco)		

As estimativas dos parâmetros dos modelos propostos nas Tabelas 13, 14, 15 e 16 são todas estatisticamente significativas ao nível de 5% de significância, conforme o teste t de Student.

Os critérios adotados para escolher o melhor modelo consistem em escolher o modelo que apresentar menor QMR e menor AIC, sendo então o que se ajustou melhor a série temporal SARIMA(1,1,0)x(0,0,2) com efeito de intervenção em $w_0 =$ Março de 1993.

Portanto, escolhe-se o modelo de intervenção

$$z_t = 292,095 x_{1,t} + \frac{(1 + 0,347 B^{12} + 0,258 B^{24})}{(1 - B)(1 + 0,22 B)} a_t$$

Utilizando a modelo de intervenção, verifica-se que em março de 1993, referente à privatização da C.M.T.C., aumentou em média 292 acidentes.

Através da função de autocorrelação residual do modelo de intervenção escolhido, observa-se o valor de Q dado pelo teste de Box e Pierce. Utilizando o teste de Shapiro-Wilk, testa-se a distribuição residual do modelo de intervenção. Estes aspectos estão detalhados a seguir.

Teste para Ruído Branco

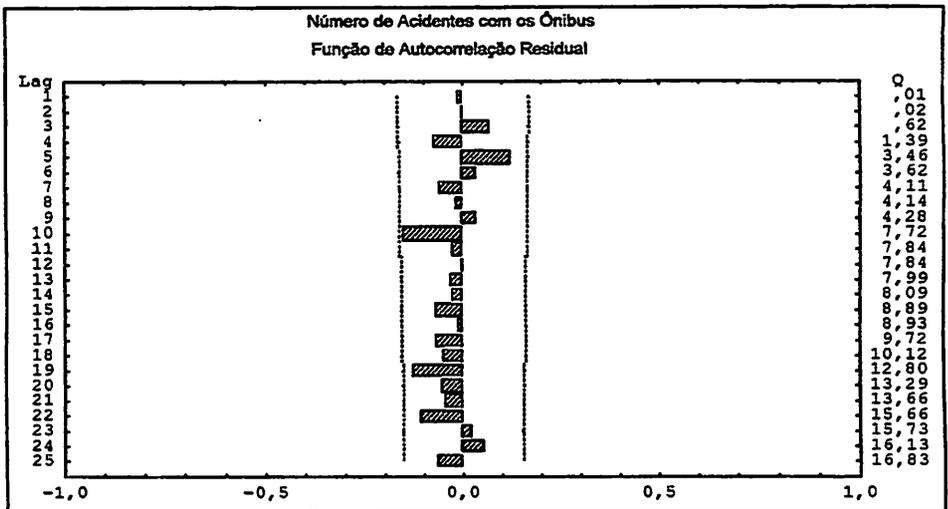


FIGURA 25 – Função de Autocorrelação Residual para a série do número de acidentes com os ônibus

Através da análise gráfica representada pela Figura 25, observa-se que todos os “lags” são não-correlacionados, sendo provavelmente um ruído branco.

A estatística de Box e Pierce é representada no “lag” 25 da Figura 25
 $Q_c = 16,83$

O valor da estatística de Box e Pierce é confrontada com o valor da distribuição Qui-Quadrado encontrada, ao nível de 5% de significância, com 21 graus de liberdade.

$$\chi_{21} = 32,68$$

Como a estimativa da $\chi_{21} > Q_c$, o resíduo do modelo ajustado é um ruído branco, sendo i.i.d. com média zero e variância residual constante.

Teste para Normalidade dos Resíduos

Aplicando-se o teste de Normalidade proposto por Shapiro-Wilk, é verificada se o resíduo provém de um distribuição normal, considerando a hipótese

H_0 : Distribuição Normal

H_1 : Não tem Distribuição Normal

A estatística T_s de Shapiro-Wilk e sua probabilidade são fornecidas abaixo:

Shapiro-Wilk	Estimativa	Probabilidade
w	0,982408	0,5872

Pode-se concluir, através da probabilidade da estatística de Shapiro-Wilk, que o resíduo do modelo proposto provém de uma distribuição normal.

4.3.3 Previsão

Escolhe-se, através dos critérios de Akaike e do QMR, o melhor modelo univariado e de intervenção para fazer previsão.

As previsões realizadas no período de abril a setembro de 1999 são fornecidas pelas Tabelas 17 e 18, propondo um intervalo de confiança de 95% para os valores de previsão, comparando qual modelo proposto fornece melhor previsão para a série temporal.

TABELA 17 – Previsão para o modelo SARIMA(1,1,0)x(0,0,2)₁₂ w₀= Março de 1993

Data	Valor Real	Previsão	LI:95%	LS:95%
abril 1999	1370	1340	1100	1581
maio 1999	1384	1408	1103	1713
junho 1999	1379	1387	1023	1751
julho 1999	1320	1362	948	1775
agosto 1999	1350	1374	915	1832
setembro 1999	1477	1396	898	1895
Erro Quadrático Médio			10441	

TABELA 18 – Previsão para o modelo SARIMA(1,1,0)x(0,0,2)₁₂

Data	Valor Real	Previsão	LI:95%	LS:95%
abril 1999	1370	1336	1090	1581
maio 1999	1384	1405	1097	1713
junho 1999	1379	1384	1016	1751
julho 1999	1320	1356	939	1773
agosto 1999	1350	1368	907	1829
setembro 1999	1477	1390	889	1892
Erro Quadrático Médio			10811	

Os valores reais da série estão todos dentro do intervalo de confiança, estruturados através dos valores preditos nos dois modelos propostos. Conclui-se que o modelo de intervenção apresenta erro quadrático médio menor que o modelo univariado, fornecendo melhores previsões.

Através da Figura 26, observa-se o comportamento da série prevista para o modelo de intervenção SARIMA(1,1,0)x(0,0,2)₁₂ w₀=março 1993, com os valores reais da série no período de abril a setembro de 1999.

A Figura 27 ilustra a série real e a série obtida a partir do modelo SARIMA (1,1,0)x(0,0,2)₁₂ w₀. Pode-se observar que o modelo escolhido ajustou bem os dados.

Através da Figura 27, pode-se observar que o modelo de intervenção proposto teve o mesmo comportamento que a série real no período de previsão, sendo, neste caso, indicado o uso do modelo de intervenção.

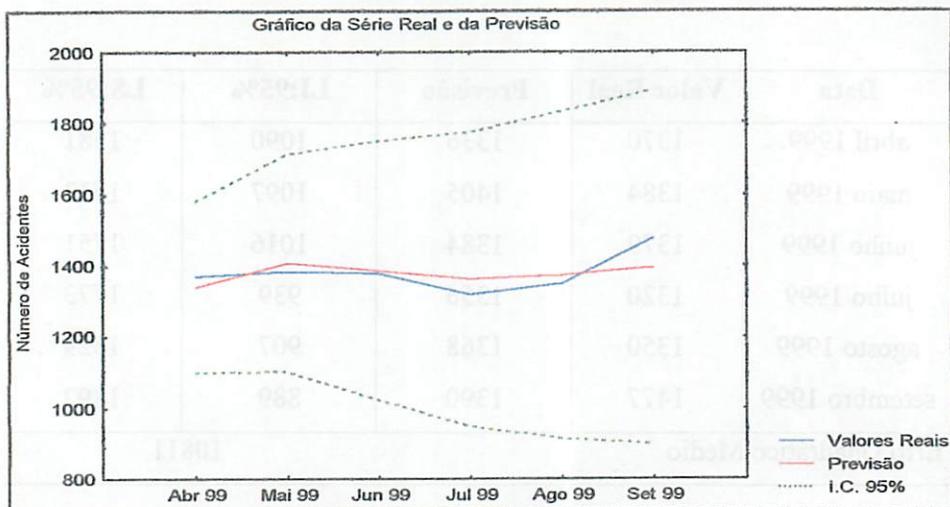


FIGURA 26 – Série Real e de Previsão com os respectivos IC 95%, no período de Abril a Setembro de 1999

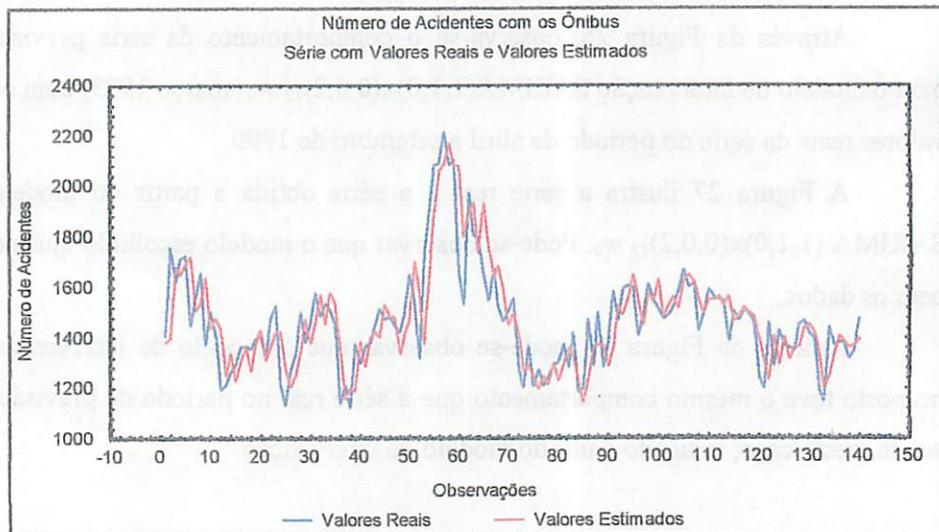


FIGURA 27 – Série Real e Série Estimada para o período de Janeiro de 1988 a Setembro de 1999

5 CONCLUSÃO

Os modelos univariados, propostos por Box e Jenkins, quando aplicados às séries de transporte urbano, tiveram bom ajuste.

As séries de acidentes e de número médio de passageiros apresentaram correlações sazonais, sendo que o modelo SARIMA se ajustou melhor aos dados. Para a série de assaltos, o melhor modelo foi um ARIMA.

Através do modelo de intervenção, inclui-se o efeito de eventos extraordinários sobre as séries, ocasionando melhores previsões em todas as séries analisadas.

Nas três séries estudadas, foram propostos quatro modelos, sendo dois univariados e dois de intervenção. Através do teste do ruído branco, dos critérios de Akaike, quadrado médio residual e do erro quadrático médio, conclui-se que os modelos de intervenção apresentaram-se mais eficientes do que os modelos univariados, para as três séries.

A série referente ao número médio de passageiros por ônibus apresentou efeito de intervenção significativo em janeiro de 1984, decorrente dos incentivos à utilização do transporte público da administração. Em janeiro de 1985 a intervenção ocorreu pelo acréscimo do número de ônibus, ocorrido no segundo semestre de 1984.

Na série referente ao número de assaltos, houve um aumento significativo em dezembro de 1997 e novembro de 1998, não sendo possível identificar os fatores que influenciaram esta intervenção.

A série do número de acidentes teve efeito de intervenção significativo no período da privatização da C.M.T.C., referente a março de 1993, aumentando o número de acidentes neste período.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARTLETT, M.S. On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society*, London, v.8, n.1, p.27-41, Jan. 1946.
- BHATTACHARYYA, M.N.; LAYTON, A.P. Effectiveness of seat belt legislation on the Queensland road toll – An Australian case study in intervention analysis. *Journal of the American Statistical Association*, Alexandria, v.74, n.367, p.596-603, Sept. 1979.
- BOX, G.E.P.; JENKINS, G.M. *Time series analysis, forecasting and control*. San Francisco: Holden-Day, 1976. 575 p.
- BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. *Time series analysis: forecasting and control*. 3. ed. New Jersey: Prentice Hall, 1994. 598 p.
- BOX, G.E.P.; LJUNG, G.M. On a measure of lack of fit in time series models. *Biometrika*, Great Britain, v.65, n.2, p.297-303, Aug. 1978.
- BOX, G.E.P.; PIERCE, D.A. Distribution of residual auto-correlations in autorregressive-integrated moving average time series models. *Journal of the American Statistical Association*, Alexandria, v.65, n.332, p.1509-1529, Dec.1970.
- BOX, G.E.P.; TIAO, G.C. A change in level of a non-stationary time series. *Biometrika*, London, v.52, n.1/2, p.181-192, June 1965.
- BOX, G.E.P.; TIAO, G.C. Comparison of forecast and actuality. *Applied Statistics*, London, v.25, n.3, p.195-200, Oct. 1976.
- CONOVER, W.J. *Practical nonparametric statistics*. New York: Wiley, 1980. 493 p.
- GLASS, G.V. Estimating the effects of intervention into a non-stationary time series. *American Educational Research Journal*, Washington, v.9, n.3, p.463-477, 1972.

- GLASS, G.V.; WILLSON, V.L.; GOTTMAN, J.M.** Design and analysis of time series experiments. Boulder: Colorado Associated University Press, 1975. 241 p.
- GRANGER, C.W.J.; NEWBOLD, J.P.** Forecasting transformed series. *Journal of the Royal Statistical Society*, London, v. B38, n.2, p.189-203, May 1976.
- GRANGER, C.W.J.; NEWBOLD, J.P.** Forecasting economic time series. New York: Academic Press, 1977. 333 p.
- JENKINS, G.M.** Pratical experiences with modelling and forecasting time series. Jerrey: Channel Islands, G JP Publications, 1979. 146 p.
- KOOPMANS, L.H.** The spectral analysis of time series. New York: Academic Press, 1974. 366 p.
- LITTELL, R.C.; MILLEKEN, G.A.; STROUP, W.W.; WOLFEINGER, R.D.** SAS[®] System for mixed models. Cary: SAS Institute, 1996. 633 p.
- MORETTIN, P.A.** Análise harmônica de processos estocásticos. Rio de Janeiro: IMPA-CNPq, 1979. 176 p.
- MORETTIN, P.A.; TOLOI, C.M.C.** Previsão de séries temporais. São Paulo: Atual, 1985. 439 p.
- PEREIRA, B.B.** Séries temporais multivariadas. VI Simpósio Nacional de Probabilidade e Estatística. São Paulo: Associação Brasileira de Estatística, 1984.
- PINO, F.A.** Análise de Intervenção em Séries Temporais: Aplicações em Economia Agrícola. São Paulo: USP, 1980. 253 p. (Dissertação – Mestrado em Estatística).
- PRIESTLEY, M.B.** Spectral analysis and time series – (Probability and Mathematical Statistical). 6. ed. New York: Academic Press, 1989. 890 p.
- QUENOUILLE, M.H.** Approximate tests of correlation in time series. *Journal of the Royal Statistical Society*, London, v.11, n.1, p.68-84, Jan. 1949.
- SAS INSTITUTE.** SAS/STAT[®]: user's guide. North Carolina, 1999.
- STATISTICA for Windows.** Realese 5. Copyright Stat Soft. 1984-1995.

**WOLD, H.O. A study in the analysis of stationary time series. Sweden:
Uppsala, 1938. 214 p.**

ANEXOS

ANEXO A	Página
TABELA A1 Série do Número de Passageiros Mensal por Ônibus no Município de São Paulo no período de jan/83 a dez/98	96
TABELA A2 Série de Assaltos nos Ônibus Urbanos do Município de São Paulo no período de jan/92 a ago/99	97
TABELA A3 Série de Acidentes de Trânsito nos Ônibus Urbanos do Município de São Paulo no período de jan/88 a set/99	98

TABELA A1 - Série do Número de Passageiros Mensal por Ônibus no Município de São Paulo, no período de jan/83 a dez/98

Ano	Jan.	Fev.	Mar.	Abr.	Maio	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
1983	651	663	675	654	663	638	632	668	650	671	678	670
1984	763	790	837	818	769	770	755	788	792	796	810	796
1985	678	717	728	751	780	758	770	778	808	817	788	796
1986	781	788	836	845	846	820	807	870	852	903	849	859
1987	819	853	818	834	809	780	768	791	776	775	769	746
1988	709	732	748	742	718	721	659	729	718	720	740	758
1989	759	772	810	821	821	812	790	841	842	854	847	846
1990	819	837	858	832	838	812	794	849	864	857	850	830
1991	756	782	787	825	815	791	752	775	773	767	748	712
1992	648	673	636	644	639	636	572	601	588	600	607	601
1993	577	607	639	613	621	619	584	607	589	594	603	584
1994	554	601	606	596	589	575	541	609	623	631	636	619
1995	604	630	648	649	644	639	599	613	603	596	596	573
1996	526	546	573	568	565	552	507	551	534	525	532	492
1997	456	492	524	510	502	488	467	504	500	490	493	477
1998	435	462	475	469	467	451	417	450	442	427	433	408

TABELA A2 – Série de Assaltos nos Ônibus Urbanos do Município de São Paulo no período de jan/92 a ago/99

Ano	Jan.	Fev.	Mar.	Abr.	Maio	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
1992	412	501	565	484	343	349	421	397	330	264	344	247
1993	282	346	278	321	296	304	247	300	340	358	347	270
1994	263	307	376	407	334	392	354	392	305	317	247	225
1995	299	255	325	314	334	336	373	444	336	348	398	403
1996	507	464	468	420	491	514	506	442	407	385	382	341
1997	504	478	460	412	489	598	648	701	680	581	698	1034
1998	1045	1022	948	1139	1088	1188	1231	1248	1082	1076	804	852
1999	894	1086	1094	1255	1214	1042	933	856				

TABELA A3 – Série de Acidentes de Trânsito nos Ônibus Urbanos do Município de São Paulo no período de jan/88 a set/99

Ano	Jan.	Fev.	Mar.	Abr.	Mai	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
1988	1383	1409	1757	1621	1713	1725	1449	1564	1658	1376	1506	1422
1989	1194	1211	1263	1297	1366	1368	1357	1395	1411	1326	1477	1527
1990	1214	1134	1274	1336	1503	1383	1455	1580	1488	1538	1512	1459
1991	1143	1211	1123	1376	1298	1366	1383	1452	1516	1474	1486	1452
1992	1423	1494	1650	1548	1319	1702	1893	2084	2091	2215	2087	2107
1993	1661	1532	1973	1811	1796	1649	1548	1670	1522	1468	1507	1556
1994	1277	1201	1364	1208	1274	1209	1248	1288	1307	1373	1282	1421
1995	1184	1198	1471	1289	1499	1284	1389	1582	1470	1593	1606	1600
1996	1407	1494	1611	1552	1519	1476	1493	1523	1588	1667	1602	1619
1997	1429	1496	1592	1564	1554	1549	1389	1504	1473	1514	1498	1465
1998	1246	1200	1461	1289	1432	1372	1359	1355	1454	1470	1420	1397
1999	1195	1148	1444	1370	1384	1379	1320	1350	1477			

Anexo B1 Estrutura do Programa SAS para estimar o modelo de intervenção para a Série do Número de Passageiros Mensal por Ônibus no Município de São Paulo 100

Anexo B2 Estrutura do Programa SAS para estimar o modelo de intervenção para a Série de Assaltos nos Ônibus Urbanos do Município de São Paulo..... 101

Anexo B3 Estrutura do Programa SAS para estimar o modelo de intervenção para a Série de Acidentes de Trânsito nos Ônibus Urbanos do Município de São Paulo 102

Anexo B1

Estrutura do Programa SAS para estimar o modelo de intervenção para a Série do Número de Passageiros Mensal por Ônibus no Município de São Paulo

```
data passag;
input date monyy. dados;
format date monyy.;
label x1='Intervention for jan1985 period';
month=month(date);
year=year(date);
x1=(month=1)*(year=1984);
x2=(month=1)*(year=1985);
cards;
jan83 651
:
jun98 451
;
proc print data=passag;
run;
proc Arima data=passag;
identify var=dados(1,0) crosscorr=(x1 x2) noprint nlag=30;
estimate
p=(1 6 12)
q=(0)
input=(x1 x2)
noconstant
plot
method=ml;
forecast lead=0 id=date interval=month printall;
run;
quit;
```

Anexo B2

Estrutura do Programa SAS para estimar o modelo de intervenção para a Série de Assaltos nos Ônibus Urbanos do Município de São Paulo

```
data assalto;
input date monyy. dados;
format date monyy.;
label x1='Intervention for dec97 period'
      x2='Intervention for nov98 period';
month=month(date);
year=year(date);
x1=(month=12)*(year=1997);
x2=(month=11)*(year=1998);
cards;
jan92 412
:
feb99 1086
;
proc print data=assalto;
run;
proc Arima data=assalto;
identify var=dados(1,0) crosscorr=(x1 x2) noprint nlag=30;
estimate
p=(0)(0)
q=(3)(0)
input=(x1 x2)
noconstant
plot
method=ml;
forecast lead=6 id=date interval=month;
run;
quit;
```

de Acidentes de Trânsito nos Ônibus Urbanos do Município de São Paulo

Estrutura do Programa SAS para estimar o modelo de intervenção para a Série

Anexo B3

```

data acidente;
input date monyy. dados;
format date monyy.;
label x1='Intervention for mar93 period';
month=month(date);
year=year(date);
x1=(month=3)*(year=1993);
cards;
:
jan88 1383
:
mar99 1444
;
proc print data=acidente;
run;
proc arima data=acidente;
identify var=dados(1,0) crosscorr=(x1) noprint nlag=30;
estimate
p=(1)(0)
q=(12 24)
input=(x1)
nocconstant
plot
method=ml;
forecast lead=6 id=date interval=month;
run;
quit;

```