

Universidade Estadual de Campinas - UNICAMP  
Instituto de Matemática, Estatística e Ciência da Computação - IMECC

**Valores Aberrantes em Séries Temporais:  
Teste de Detecção e  
Efeito na Previsão de Valores Agregados**

Rissa Ota

Orientador: Prof. Dr. Luiz Koodi Hotta ✓

Mestrado em Estatística

Campinas/SP

1996

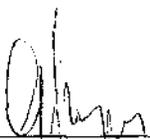
Tese de Mestrado defendida e aprovada em 05 de Julho de 1996

pela Banca Examinadora composta pelos Profs. Drs.



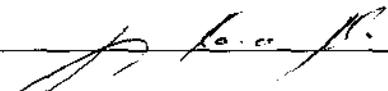
---

Prof (a). Dr (a). MARLI MIKAEL DA COSTA NEVES



---

Prof (a). Dr (a). GABRIELA STANGENHAUS



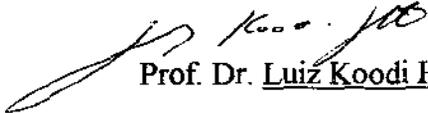
---

Prof (a). Dr (a). LUIZ KOODI HOTTA

Valores Aberrantes em Séries Temporais:  
Teste de Detecção e Efeito na Previsão de Valores Agregados

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Srta. Rissa Ota e aprovada pela Comissão Julgadora.

Campinas, 14 de agosto de 1996

  
Prof. Dr. Luiz Koodi Hotta

Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em Estatística.



*Dedico aos meus pais*

## **AGRADECIMENTOS**

Ao Prof. Dr. Luiz Koodi Hotta pela orientação, apoio e compreensão.

Ao CNPq (Conselho Nacional de Pesquisa), e ao FAEP (Fundo de Apoio ao Ensino à Pesquisa) pelo apoio financeiro e condições para minha formação acadêmica.

À Fatima N. Freires pelo trabalho de revisão de texto.

Aos meus colegas da UNICAMP pelos convívio, intercâmbio e aprendizado.

Aos professores e funcionários do Departamento de Estatística da UNICAMP.

# Sumário

<b>RESUMO</b> .....	<b>i</b>
<b>INTRODUÇÃO</b> .....	<b>1</b>
<b>1. MODELOS DE SÉRIES TEMPORAIS COM OUTLIER</b> .....	<b>4</b>
1.1 - Modelos ARIMA .....	4
1.2 - Modelos Estruturais e a Relação com Modelos ARIMA .....	7
1.3 - Modelos ARIMA com outliers .....	12
1.3.1 - <i>Outlier aditivo (AO)</i> .....	13
1.3.2 - <i>Outlier de Inovação (IO)</i> .....	13
1.3.3 - <i>Mudança de Nível - Level Shift - (LS)</i> .....	16
1.3.4 - <i>Mudança de Nível Temporário - Temporal Change - (TC)</i> .....	17
1.4 - Outlier em Modelos Estruturais .....	18
<b>2. TESTES DE DETECÇÃO DE OUTLIERS</b> .....	<b>21</b>
2.1 - Teste da Razão de Verossimilhança .....	21
2.1.1 - <i>Teste de detecção quando os parâmetros do modelo são conhecidos</i> .....	23
2.1.2 - <i>Teste de detecção quando os parâmetros do modelo são desconhecidos</i> .....	26
2.1.3 - <i>Um procedimento iterativo para detecção de outliers e estimação dos parâmetros</i> .....	29
2.1.4 - <i>O poder do teste e o valor crítico</i> .....	32
2.1.5 - <i>Identificação incorreta do tipo de outlier</i> .....	33
2.2 - Teste do Escore .....	34
2.2.1 - <i>Teste de detecção de outlier de tipo desconhecido</i> .....	35
2.2.2 - <i>Testes de detecção de outlier de tipo conhecido</i> .....	36
2.2.3 - <i>Teste para posição do outlier desconhecido</i> .....	36
2.2.4 - <i>Teste com parâmetros de AR e <math>\sigma^2</math> estimados pelos dados</i> .....	39
2.2.5 - <i>Distinção entre dois tipos de outliers</i> .....	40

2.3 - Deslocamento de Verossimilhança - <i>Likelihood Displacement</i> (LD) .....	41
2.3.1 - <i>Algumas medidas de influência</i> .....	41
2.3.2 - <i>O teste de LD</i> .....	45
<b>3. EFEITOS DA PRESENÇA DO OUTLIER EM SÉRIES TEMPORAIS.....</b>	<b>47</b>
3.1 - Efeitos nos resíduos .....	47
3.1.1 - <i>Efeitos nos resíduos considerando os parâmetros e o modelo conhecidos</i> .....	48
3.1.2 - <i>Efeito no ruído quando os parâmetros são estimados</i> .....	52
3.2 - Efeito na Estimativa dos Parâmetros do Modelo ARIMA .....	53
3.2.1 - <i>Efeito de AO na estimação dos parâmetros</i> .....	53
3.2.2 - <i>Efeito de IO, TC e LS na estimativa dos parâmetros</i> .....	56
3.3 - Efeito na identificação do Modelo .....	57
3.3.1 - <i>Considerações gerais do vício da fac</i> .....	57
3.3.2 - <i>Efeito do outlier na estimação de <math>\sigma^2</math></i> .....	60
3.3.3 - <i>Alguns procedimentos para identificação do modelo</i> .....	61
3.4 - Efeitos na Previsão .....	62
3.4.1 - <i>Previsão através do modelo ARIMA</i> .....	63
3.4.2 - <i>Erro de previsão na presença de um outlier</i> .....	64
3.4.3 - <i>Efeito do outlier na previsão quando o outlier ocorre no meio da série</i> .....	65
3.4.4 - <i>Efeito do outlier na previsão quando o outlier ocorre próximo do final da série</i> .....	67
<b>4. EFEITO DO OUTLIER NA PREVISÃO DE SÉRIES TEMPORAIS AGREGADAS .....</b>	<b>72</b>
4.1 - Modelo agregado .....	73
4.2 - Efeito da agregação nas previsões .....	74
4.2.1 - <i>Perda de eficiência na previsão devida à agregação num processo ARIMA(0,1,1) conhecido</i> .....	75
4.2.2 - <i>Perda de eficiência na previsão devida à agregação num processo ARIMA(1,1,0) conhecido</i> .....	78
4.2.3 - <i>Perda de eficiência na previsão devida à agregação em modelos ARIMA desconhecidos</i> .....	80

4.3 - Efeito de AO na previsão .....	80
4.3.1 - Modelos conhecidos.....	81
4.3.1.1 - ARIMA(1,1,0) desagregado .....	84
4.3.1.2 - ARIMA(0,1,1) desagregado.....	89
4.3.2 - Modelos desconhecidos.....	95
4.3.2.1 - Aproximação pelo modelo ARI(P,1) .....	99
4.3.2.2 - Um outlier de tamanho 1.....	101
4.3.2.3 - Um outlier de tamanho 2.....	103
4.3.2.4 - Dois outliers de tamanho 1 de mesmo sinal.....	110
4.3.2.5 - Dois outliers de tamanho 1 de sinais trocados .....	111
4.4 - Efeito de IO na previsão .....	116
4.4.1 - Modelos conhecidos.....	116
4.4.1.1 - ARIMA(1,1,0) desagregado .....	118
4.4.1.2 - ARIMA(0,1,1) desagregado.....	123
4.4.2 - Modelos desconhecidos.....	127
4.4.2.1 - Um outlier de tamanho 1 .....	128
4.4.2.2 - Um outlier de tamanho 2.....	130
4.4.2.3 - Dois outliers de tamanho 1 de mesmo sinal.....	137
4.4.2.4 - Dois outliers de tamanho 1 de sinais trocados.....	138
4.5 - Conclusões .....	142
<b>Apêndice 1 - Teste de Razão de Verossimilhança.....</b>	<b>143</b>
<b>Apêndice 2 - Teste de Escore.....</b>	<b>147</b>
<b>Apêndice 3 - Critérios de Informação.....</b>	<b>149</b>
<b>Apêndice 4 - Programa.....</b>	<b>150</b>
<b>Apêndice 5 - Tabelas dos resultados de simulação.....</b>	<b>168</b>
<b>Referências Bibliográficas.....</b>	<b>176</b>

## RESUMO

Neste trabalho são discutidos alguns tipos de valores aberrantes (denotado nesse trabalho por outlier) mais citados na literatura de séries temporais e os efeitos que eles podem causar na identificação, estimação e previsão dos modelos, mostrando assim a importância em detectá-los.

Nos primeiros dois capítulos são apresentados os modelos de outliers e alguns testes de detecção existentes na literatura. O Capítulo 3 é dedicado ao estudo dos efeitos dos outliers nas estimações, identificações e previsões.

No Capítulo 4 são apresentados os efeitos dos outliers presentes nas últimas observações na previsão de valores agregados, comparando os efeitos nas previsões calculadas através de modelos desagregados e agregados. No estudo são considerados os casos de modelos conhecido e desconhecido, sendo este último realizado através de simulações. De um modo geral, a previsão através de modelo agregado, na presença de outlier aditivo (AO), é menos afetada do que a previsão pelo modelo desagregado. Quando um outlier de inovação (IO) está presente na série a previsão pelo modelo agregado é geralmente mais afetada. Isto era esperado porque no caso de modelos conhecidos o IO não tem efeito nas previsões do modelo desagregado. São também realizados estudos para verificar o efeito dos testes usuais de detecção de outlier na previsão, mostrando que, embora na maioria dos casos a utilização dos testes diminuam os vícios de previsão devido aos outliers, em alguns casos eles aumentam o erro quadrático médio de previsão. Isto ocorre principalmente na presença de dois IOs, de sinais trocados, devido à incorreta detecção dos outliers, na posição e/ou tipo.

# Introdução

Séries temporais, muitas vezes, medem fenômenos sujeitos a eventos inesperados ou incontroláveis. Esses eventos dão origem às observações espúrias que de alguma forma são inconsistentes com o resto da série. Estas observações podem ser denominadas de várias formas: *outliers*, valores aberrantes, observações aberrantes, dados atípicos, observações discrepantes etc. O uso mais comum tem sido *outlier*, o termo escolhido para este trabalho.

Quando o tempo e a causa do outlier são conhecidos, seus efeitos podem ser estimados usando o modelo de intervenção de Box e Tiao. Na prática, estes dois fatores geralmente são desconhecidos.

Dependendo da sua natureza, os outliers podem não causar nenhum impacto importante como também podem causar um efeito substancial na análise dos dados. É importante detectar os outliers por várias razões:

1. Melhor entendimento da série em estudo. Um outlier detectado pode ser uma evidência da ocorrência de algum fator externo afetando a série. Por exemplo, greves, mudanças políticas, mudanças repentinas na estrutura do mercado e distúrbios inesperados nas condições de um sistema físico são exemplos de eventos que podem explicar por que uma série se comporta como tal. Além disso, podem-se descobrir observações espúrias, como exemplo erros de digitação, que podem mascarar a modelagem.
2. Melhor modelagem e estimação. Eventos externos desconhecidos podem afetar na modelagem e/ou estimação. Detectar outliers pode resultar em simplificar a estrutura do modelo utilizado, além disso, mesmo empregando o modelo “verdadeiro” da série, a não consideração da presença de um evento externo pode afetar seriamente na estimativa dos parâmetros do modelo, tornando pouco confiáveis ou até mesmo inválidos os resultados de inferência.
3. Melhor previsão. Dependendo do tempo de ocorrência e natureza do evento, um evento externo pode afetar o desempenho da previsão de um modelo. Ajustada à presença de um outlier a previsão pode ser melhorada e de um modo geral obtém-se um modelo de previsão com melhor desempenho.

O objetivo desta tese é apresentar a importância em detectar os outliers mostrando os efeitos que eles podem causar na identificação do modelo, na estimação dos parâmetros, na previsão dos valores futuros. O enfoque foi dado principalmente no efeito dos outliers nas previsões de valores agregados. São apresentados os efeitos teóricos exatos e aproximados, neste último os resultados são posteriormente comprovados através de simulações, recurso utilizado quando não é possível obter um resultado teórico, nem mesmo aproximado. A influência da utilização de testes de detecção de outliers na previsão também é estudada via simulação.

No Capítulo 1 são apresentados as notações, as nomenclaturas, os termos utilizados nos outros capítulos, os modelos ARIMA e modelos estruturais e definido os tipos de outliers abordados nesta dissertação. No Capítulo 2 são apresentados alguns testes de detecção e medidas de influência dos tipos de outlier mais utilizados na literatura. Alguns efeitos que um outlier pode causar em séries temporais, tais como, na estimação dos parâmetros, na identificação do modelo e nas previsões são apresentados no Capítulo 3.

Muitas séries temporais econômicas, como renda e consumo são disponíveis em diferentes níveis de agregação, por exemplo mensal ou trimestralmente. No entanto, em muitas situações, os economistas estão interessados em prever dados agregados (dados quadrimestrais ou anuais). Nesse caso a função de previsão pode ser baseada tanto na série desagregada como na série agregada.

Um dos objetivos desta tese, abordado no Capítulo 4, é estudar os efeitos dos outliers nessas previsões de valores agregados. Quando o outlier ocorre longe do início da previsão o efeito na previsão pontual é pequeno (Ledolter, 1989), exceto quando a magnitude do outlier é muito grande. Neste caso, no entanto, ele pode ser detectado sem causar problemas. Desta forma, este trabalho se concentra nos casos em que o outlier ocorre perto da origem de previsão, i.e., nas últimas observações.

Os resultados obtidos mostram que os outliers podem afetar muito as previsões, dependendo da agregação, do modelo, do tipo de outlier presente e do tamanho do outlier. Por exemplo, quando um outlier aditivo está presente na série, o modelo desagregado produz maiores erros de previsão do que o modelo agregado, enquanto a presença de outlier de inovação no modelo desagregado não causa erro de previsão.

Hotta e Pereira (1995) estudaram os efeitos de dois tipos de outliers para modelos conhecidos na previsão de valores agregados quando o outlier ocorre nas últimas observações da série. Eles analisaram os efeitos dos outliers na previsão utilizando modelos

desagregado e agregado conhecidos. Os resultados mostraram que o outlier de inovação não afeta a previsão em modelos desagregados, mas pode ter efeito bastante grande na previsão do modelo agregado. O outlier aditivo pode afetar os erros de previsão em ambos os modelos; o efeito é mais persistente no modelo agregado, mas em geral é menor do que do modelo desagregado.

Para compreender melhor o efeito do outlier na previsão é necessário considerar o efeito dos outliers na identificação do modelo e estimação dos parâmetros. Esse efeito é estudado através de simulações apresentados no Capítulo 4. É também considerada a aplicação do teste de Razão de Verossimilhança de detecção de outliers. Os resultados mostram que, na presença do outlier, a eficiência das previsões realizadas através do modelo agregado pode ser maior que a da realizada através do modelo desagregado.

# Capítulo 1

## Modelos de séries temporais com outlier

O objetivo deste capítulo é apresentar os termos e as notações utilizados na dissertação. Nas duas primeiras seções são apresentados os dois modelos em Séries Temporais citados neste trabalho: o modelo ARIMA na primeira seção e o modelo estrutural e sua relação com o modelo ARIMA na segunda seção. Na seção 1.3 são apresentados os quatro tipos de outliers abordados nesta tese, no contexto de modelos ARIMA. Finalmente, na seção 1.4 são apresentados os modelos estruturais com outliers e uma comparação com os modelos ARIMA com outlier.

### 1.1 Modelos ARIMA

Muitos dos processos estocásticos podem ser aproximados por um dos modelos da classe de modelos auto-regressivo-integrado-média-móvel (ARIMA). Estes modelos popularizaram-se após o lançamento do livro de Box e Jenkins (1970) e por isso muitas vezes são denominados como modelos de Box-Jenkins, ou simplesmente como modelos Box-Jenkins. Eles propuseram um método sistemático para identificação, estimação e avaliação dos modelos.

Este modelo tem como grande vantagem, em relação aos outros modelos, a parcimônia nos parâmetros, além de ajustar tão bem quanto os modelos anteriormente utilizados. A parcimônia aumenta a eficiência na estimação do modelo, obtendo-se assim melhores desempenhos nas previsões em relação aos modelos super-parametrizados.

O modelo ARIMA de ordem  $p$ ,  $d$  e  $q$  (ARIMA( $p$ ,  $d$ ,  $q$ )) é dado por

$$\phi(B)\alpha(B) x_t = \theta(B)a_t, \quad (1.1.1)$$

onde  $t = 1, 2, \dots, T$ ; e

1.  $B$  é o operador de atraso, i.e.,  $Bx_t = x_{t-1}$ ;
2.  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  e  $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  são, respectivamente, polinômio em  $B$  auto-regressivo de ordem  $p$  (AR( $p$ )) e média móvel de ordem  $q$  (MA( $q$ )), com:
  - todas as raízes fora do círculo unitário,
  - sem nenhuma raiz comum em  $\phi(B)$  e  $\theta(B)$ ;
3.  $\alpha(B) = (1-B)^d$ , o operador de diferenças responsável pela não estacionariedade;
4.  $\{a_t\}$  é uma seqüência de ruídos brancos com média zero e variância  $\sigma_a^2$  (geralmente a suposição de normalidade é necessária, pois a maioria dos testes é baseada na função de verossimilhança);
5. caso o processo não seja estacionário assume-se que  $x_t$  começa num tempo finito.

O modelo anterior pode ser representado de duas formas. A primeira é escrever o processo  $x_t$  como uma combinação linear de uma seqüência de variáveis aleatórias não correlacionadas com média zero, i.e.,

$$x_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \sum \psi_i a_{t-i} = \psi(B)a_t. \quad (1.1.2)$$

Essa representação é útil para calcular a variância e as autocovariâncias de um processo estacionário. Por exemplo, é fácil mostrar que

$$\text{Var}(x_t) = \sigma_a^2 \sum \psi_i^2$$

e

$$E(a_t x_{t-j}) = \begin{cases} \sigma_a^2 & \text{se } j = 0; \\ 0 & \text{se } j > 0. \end{cases}$$

Assim as funções de autocovariâncias e autocorrelações são dadas, respectivamente por

$$\gamma_k = \text{Cov}(x_t, x_{t+k}) = E(x_t x_{t+k}) = \sigma_a^2 \sum \psi_i \psi_{i+k}$$

e

$$\rho_k = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}.$$

Pode-se verificar que as funções de autocorrelações e autocovariâncias são funções somente da diferença de tempo,  $k$ . Como elas têm infinitos termos, para ser estacionário é necessário mostrar que  $\gamma_k$  é finito para cada  $k$ . Desta forma

$$|\gamma_k| = |E(x_t x_{t+k})| \leq [\text{Var}(x_t) \text{Var}(x_{t+k})]^{1/2} = \sigma_a^2 \sum \psi_j^2.$$

Assim  $\sum \psi_j^2 < \infty$  é uma condição necessária e suficiente para que o processo (1.1.2) seja estacionário. Esta condição é equivalente às raízes de  $\phi(B)$  estarem fora do círculo unitário,

como no item 2, e a ordem de não estacionariedade ser nula, i.e.,  $d = 0$ , neste caso o modelo é denominado ARMA( $p, q$ ).

A forma (1.1.2) é chamada de representação média-móvel (MA) de um processo. Wold (1954) provou que um processo estacionário que é puramente não determinístico (i.e., um processo não contém nenhum componente determinístico que possa ser previsto ou predito exatamente pelo seu passado) sempre pode ser expresso na forma MA. Assim, essa representação é conhecida como representação de Wold na literatura, e qualquer processo que possa ser representado desta forma é chamado processo não determinístico.

Uma outra forma de escrever o processo  $x_t$  é pela representação auto-regressiva (AR), em que a variável observada no instante  $t$  é uma função linear das observações passadas mais uma inovação, i.e.,

$$x_t = \pi_1 x_{t-1} + \pi_2 x_{t-2} + \dots + a_t,$$

ou equivalentemente,

$$\pi(B)x_t = a_t, \quad (1.1.3)$$

onde  $\pi(B) = \phi(B)\alpha(B)/\theta(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$ , e os coeficientes  $\pi_i$ s são denominados *pesos* do polinômio  $\pi(B)$ . Se  $\sum |\pi_i| < \infty$ ,  $x_t$  é dito ser um processo invertível (Box e Jenkins, 1976). Box e Jenkins alegam que não faz sentido fazer previsão num processo não invertível. Essa condição de invertibilidade é equivalente às raízes de  $\theta(B)$  estarem fora do círculo unitário, como no item 2.

Esta representação é útil para entender o mecanismo de previsão. Note que num modelo ARIMA( $p, d, 0$ ) (ou simplesmente ARI( $p, d$ )), os coeficientes de  $\pi(B)$ ,  $\pi_i \neq 0$  para  $i = 1, \dots, p + d$ ; assim  $\sum_{i=0}^{\infty} |\pi_i| < \infty$ , o que leva a concluir que o processo ARIMA( $p, d, 0$ ) é sempre invertível.

Pelo exemplo anterior pode-se ver que um processo invertível não necessariamente é um processo estacionário. Para que o processo (1.1.3) possa ser estacionário é preciso que se possa representá-lo na forma MA, i.e.,

$$x_t = \pi(B)^{-1} a_t = \psi(B) a_t,$$

onde  $\sum \psi_j^2 < \infty$  seja satisfeita. Para atingir essa restrição, a condição necessária e suficiente é que as raízes de  $\pi(B)$  estejam todas fora do círculo unitário.

## 1.2 Modelos Estruturais e a Relação com Modelos ARIMA

Uma outra forma de modelar uma série temporal é através de modelo estrutural. Este modelo é formulado diretamente em termos de componentes de interesse, com um considerável apelo intuitivo (Harvey, 1989, p.17), i.e., neste modelo os parâmetros e componentes têm interpretação imediata, o que nos modelos ARIMA é mais difícil. Uma outra característica importante desse modelo é a possibilidade de ser representado como modelo ARIMA.

A essência de um modelo de série temporal estrutural é considerar as observações como sendo geradas em termos de componentes que têm interpretação direta. Um modelo estrutural não tem interesse em representar o processo gerador subjacente dos dados como no modelo ARIMA. Ele procura explicar as estruturas de uma série através de uma decomposição em fatores tais como tendência, sazonalidade, ciclos e componentes irregulares.

Considere, por exemplo, uma linha de produção de engrenagem numa fábrica de autopeças. A variável de interesse é o diâmetro da engrenagem. Essa variável pode ser modelada como sendo uma soma do nível da série e de uma perturbação dada à variabilidade natural do processo. Esse nível pode mudar continuamente, por exemplo, de acordo com um passeio ao acaso mais um distúrbio não observável.

O modelo anteriormente descrito é um exemplo de um modelo estrutural conhecido como *modelo de nível local* ou *modelo de passeio ao acaso mais o ruído*. Esse modelo torna-se completamente definido determinando a distribuição dos dois distúrbios. Com suposição usual de que essas distribuições são normais com média zero, restam apenas as variâncias para especificar completamente o modelo. Assim o modelo é dado por

$$x_t = \mu_t + \varepsilon_t \quad \text{e} \quad (1.2.1)$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad (1.2.2)$$

onde  $\varepsilon_t$  e  $\eta_t$  são variáveis aleatórias independentes com variâncias  $\sigma_\varepsilon^2$  e  $\sigma_\eta^2$ , respectivamente. Assume-se que o nível do processo,  $\mu_t$ , é gerado por um passeio ao acaso não observável.

Reescrevendo o modelo de nível local, substituindo (1.2.2) em (1.2.1), tem-se o modelo na forma reduzida

$$\nabla x_t = \eta_t + \nabla \varepsilon_t, \quad (1.2.3)$$

onde  $\nabla = 1 - B$  é o operador de primeira diferença, i.e.,  $\nabla x_t = x_t - x_{t-1}$ .

Por essa equação é fácil ver que

$$E(\nabla x_t) = E(\eta_t) + E(\varepsilon_t) - E(\varepsilon_{t-1}) = 0,$$

e as autocovariâncias

$$\gamma(0) = E[(\eta_t + \nabla \varepsilon_t)^2] = \sigma_\eta^2 + 2\sigma_\varepsilon^2,$$

$$\gamma(1) = E[(\eta_t + \nabla \varepsilon_t)(\eta_{t-1} + \nabla \varepsilon_{t-1})] = -\sigma_\varepsilon^2,$$

$$\gamma(\tau) = 0, \quad \tau \geq 2.$$

Assim a função de autocorrelação de  $\nabla x_t$  dada por

$$\rho(\tau) = \begin{cases} -\sigma_\varepsilon^2 / (\sigma_\eta^2 + 2\sigma_\varepsilon^2), & \tau = 1; \\ 0 & , \tau \geq 2, \end{cases} \quad (1.2.4)$$

exibe um corte na primeira defasagem, característico de um processo MA(1). A forma reduzida é então um modelo ARIMA(0,1,1), i.e., pode ser representada como  $\nabla x_t = (1 - \theta B)a_t$ , onde  $a_t$  é um ruído branco com média zero e variância  $\sigma_a^2$ , pois  $0 \leq \rho(\tau) \leq 1/2$ . Desta forma, tem-se as seguintes relações:

$$\text{Var}(\nabla x_t) = \sigma_\eta^2 + 2\sigma_\varepsilon^2 = (1 + \theta^2)\sigma_a^2 \text{ e}$$

$$\text{Cov}(\nabla x_t, \nabla x_{t-1}) = -\sigma_\varepsilon^2 = -\theta \sigma_a^2$$

que resultam em

$$\theta = \frac{2 + q - (4s - s^2)^{1/2}}{2}, \quad (1.2.5)$$

onde  $s$  é a razão sinal-ruído  $s = \sigma_\eta^2 / \sigma_\varepsilon^2$ .

A outra raiz implica em  $|\theta| > 1$  e foi descartada, pois ela leva a um modelo MA não invertível. Como os valores de  $\theta$  definido em (1.2.5), correspondem a  $0 \leq \theta \leq 1$  ou  $0 \leq \rho(1) \leq 0,5$ , valores negativos de  $\theta$  não podem ser obtidos pela forma estrutural de nível local. Assim, o parâmetro de MA na forma reduzida cobre apenas a metade do espaço paramétrico usual, ou seja, qualquer modelo estrutural de nível local pode ser representado como modelo ARIMA(0,1,1), mas a recíproca não é verdadeira. Esta restrição pode parecer forte em certos casos, mas na prática geralmente o parâmetro  $\theta$  é positivo (Chatfield, 1988).

Verifique que o modelo de forma reduzida é estritamente não invertível quando  $\sigma_\eta^2 = 0$ , pois corresponde a  $\theta = 1$  ( $x_t = \mu + \varepsilon_t$ ), i.e., um modelo com apenas o ruído branco.

Finalmente, se  $\sigma_\varepsilon^2 = 0$ , o modelo se restringe a um modelo de passeio ao acaso, i.e.,  $x_t = x_{t-1} + \eta_t$ .

Como outro exemplo de modelo estrutural considere o consumo de combustível numa certa cidade. Com o aumento da população do município e do nível de vida o consumo do combustível cresce de forma exponencial. Considere que após a transformação logarítmica nas observações, a tendência da série se torna localmente linear. Essa tendência pode variar com o tempo, por exemplo, de acordo com um passeio ao acaso. Este tipo de comportamento pode ser descrito por

$$\begin{aligned}x_t &= \mu_t + \varepsilon_t, \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, \\ \beta_t &= \beta_{t-1} + \zeta_t,\end{aligned}\tag{1.2.6}$$

onde  $\varepsilon_t$ ,  $\eta_t$  e  $\zeta_t$  são processos de ruído branco mutuamente não correlacionados com variâncias  $\sigma_\varepsilon^2$ ,  $\sigma_\eta^2$  e  $\sigma_\zeta^2$ , respectivamente. Esse modelo é denominado *modelo de tendência local*. O efeito de  $\eta_t$  permite a mudança de nível da tendência, enquanto  $\zeta_t$  permite mudar a tendência. Quanto maiores as variâncias maiores são os movimentos estocásticos na tendência. Caso  $\sigma_\zeta^2 = \sigma_\eta^2 = 0$  o modelo se reduz ao caso de tendência determinística

$$\mu_t = \mu_{t-1} + \beta.$$

Quando  $\sigma_\zeta^2 = 0$ , então a inclinação,  $\beta$ , é constante e o processo gerador do nível se reduz a um passeio aleatório mais um deslocamento,

$$\mu_t = \mu_{t-1} + \beta + \eta_t.$$

Se, além disso,  $\beta$  é igual a zero, o modelo é um modelo de nível local (1.2.1) e (1.2.2).

O modelo de tendência local pode ser reescrita na forma reduzida:

$$\nabla^2 x_t = \nabla \eta_t + \zeta_t + \nabla^2 \varepsilon_t.\tag{1.2.7}$$

O primeiro componente do lado direito da equação é a parte da tendência que é derivada pelo movimento no nível. O segundo componente é o ruído na inclinação (tendência). Note que se  $\sigma_\zeta^2$  é zero esta equação se simplifica à equação reduzida de modelo de nível local (1.2.3), i.e., modelo sem tendência.

A função de autocorrelação da segunda diferenciação  $\nabla^2 x_t$  dada por

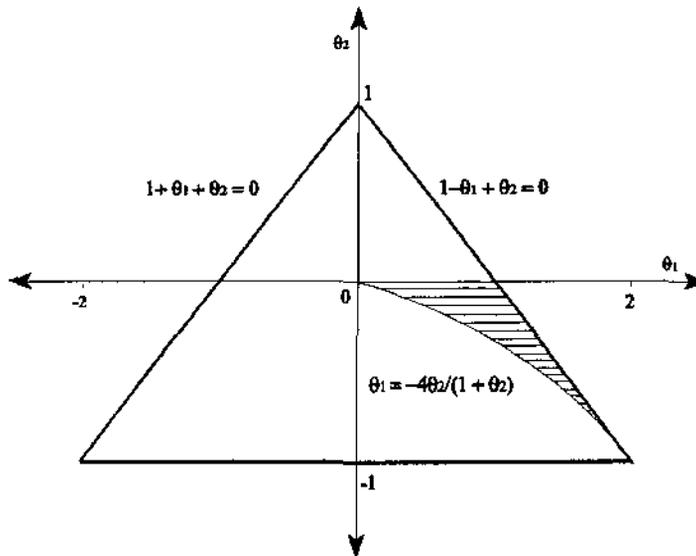


Figura 1.2.1 - Região de admissibilidade dos parâmetros do modelo ARIMA(0,2,2), apresentada em destaque os modelos que podem ser representados na forma estrutural.

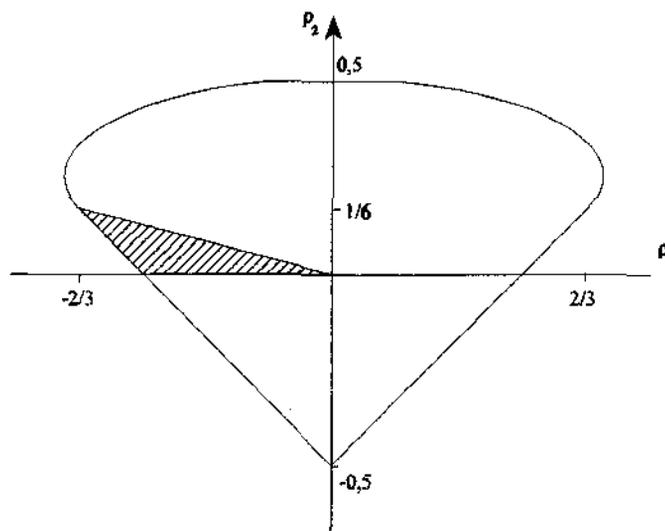


Figura 1.2.2 - Região de admissibilidade dos parâmetros do modelo ARIMA(0,2,2), apresentada em destaque os modelos que podem ser representados na forma estrutural.

$$\rho(\tau) = \begin{cases} -(\sigma_{\eta}^2 + 4\sigma_{\varepsilon}^2)/(2\sigma_{\eta}^2 + \sigma_{\zeta}^2 + 6\sigma_{\varepsilon}^2) & \tau = 1, \\ \sigma_{\varepsilon}^2/(2\sigma_{\eta}^2 + \sigma_{\zeta}^2 + 6\sigma_{\varepsilon}^2) & \tau = 2, \\ 0 & \tau \geq 3, \end{cases}$$

exibe um corte na segunda defasagem, desta forma  $\nabla^2 x_t$  é um processo MA(2). Assim o modelo de forma reduzida é equivalente a um modelo ARIMA(0,2,2), que pode ser representado como  $\nabla^2 x_t = (1 - \theta_1 B - \theta_2 B^2) a_t$ , onde  $a_t$  é um ruído branco com variância  $\sigma_a^2$ . Pelas relações de auto-correlações obtém-se as seguintes restrições  $0 \leq \rho(2) \leq 1/4$  e  $\rho(2) + \rho(1) \geq -1/2$ . Desta forma,  $-2/3 \leq \rho(1) \leq 0$  e  $0 \leq \rho(2) \leq 1/6$ . Esta região é equivalente a  $\theta_1 > 0$ ,  $\theta_1 + \theta_2 \leq 1$  e  $\theta_1 \geq -4\theta_2/(1 + \theta_2)$ ,  $-2/3 \leq \rho(1) \leq 0$  e  $0 \leq \rho(2) \leq 1/6$ . A região onde o modelo ARIMA(0,2,2) pode ser representado como modelo estrutural está ilustrada nas Figuras 1.2.1 e 1.2.2.

### 1.3 Modelos ARIMA com outliers

Nesta seção os quatro tipos de outliers são apresentados sob a estrutura dos modelos ARIMA e na seção seguinte serão apresentados alguns outliers em modelos estruturais.

Os dois primeiros tipos de outliers considerados foram introduzidos por Fox (1972). Para descrever uma série temporal observada,  $\{y_t\}$ , com  $n_{out}$  outliers o modelo é dado por

$$y_t = x_t + \sum_{j=1}^{n_{out}} f_j(t), \quad (1.3.1)$$

onde  $x_t$  representa uma série de um processo ARIMA sem outlier,  $f_j(t)$  é uma função paramétrica representando os distúrbios de  $x_t$ , i.e., outliers na posição  $k_j$ . De forma geral, pode-se considerar que o efeito afeta a observação em que ocorre o outlier e também os valores subsequentes. O seu efeito pode ser escrito como

$$f_j(t) = \omega_{0j} L_j(B) \delta_{t,k_j}, \quad (1.3.2)$$

onde  $\delta_{t,k_j} = 1$  se  $t = k_j$  e  $\delta_{t,k_j} = 0$  se  $t \neq k_j$ , é uma função indicadora. O polinômio  $L_j(B)$  é o modelo dinâmico do efeito de  $\omega_{0j}$ , uma constante denotando o impacto do distúrbio. Na prática  $f_j(t)$  é especificado pelo pesquisador baseado na informação do distúrbio e do processo  $x_t$  (Tsay, 1988).

Caso o tempo, tipo do distúrbio (mas não a magnitude) e o modelo dinâmico sejam conhecidos, então o modelo (1.3.1) é denominado *modelo de intervenção* de Box e Tiao (1975), onde o modelo dinâmico  $L_j(B)$  descreve o comportamento dinâmico da intervenção. Nesse caso, o modelo ARIMA é primeiramente identificado usando as observações anteriores à intervenção.

São considerados neste trabalho quatro modelos específicos para  $L_j(B)$ : outlier de inovação, outlier aditivo, mudança de nível e mudança temporária. Uma discussão mais detalhada sobre a natureza e motivação desses outliers pode ser encontrada em Fox (1972); Chang (1982); Hillmer *et alii* (1983); Tsay (1988), Chang, Tiao e Chen (1988), Chen Liu (1993a,b) e outros.

Para ilustrar os efeitos de cada tipo de outlier, será considerado um processo simulado de AR(1) inserindo um outlier no meio da série. Para esse propósito 60 observações foram simuladas do seguinte modelo

$$x_t = a_t / (1 + 0,6B), \quad \text{com } \sigma_a = 1. \quad (1.3.3)$$

Os dados são apresentados no Gráfico 1.3.1 (a). As mesmas inovações utilizadas para gerar a série do Gráfico 1.3.1 (a) foram usadas para gerar as séries com outlier (todos os outliers de tamanho 5 no tempo  $t = 30$ , Gráfico 1.3.1 (b) a (e)). No caso de mudança temporária o fator de amortecimento,  $\lambda$ , foi estabelecido como 0,7. O valor inicial foi gerado através de uma distribuição normal  $N(0, \sigma_a^2/(1+\phi^2))$ .

### 1.3.1 - Outlier aditivo (AO)

Um *outlier aditivo* (AO) é um evento que afeta a série em apenas uma observação. Um exemplo típico é erro de digitação (por exemplo, o valor verdadeiro 17,5 pode ser erroneamente digitado como 1,75; 175 etc.). Por esta razão um outlier aditivo às vezes é chamado de *erro grosseiro* (*gross error*) ou *outlier de observação*. Esse outlier também é denominado como *outlier aberrante* (Abraham e Box, 1979) ou de *Tipo I* (Fox, 1972). Nesse trabalho será utilizado o termo *outlier aditivo*, também proposto por Fox, ou simplesmente AO.

Assumindo que um único outlier aditivo ocorra no tempo  $t = k$ , a série observada pode ser representada pelo modelo (com  $L(B) = 1$ )

$$y_t = x_t + \omega_A \delta_{t,k}, \quad (1.3.4)$$

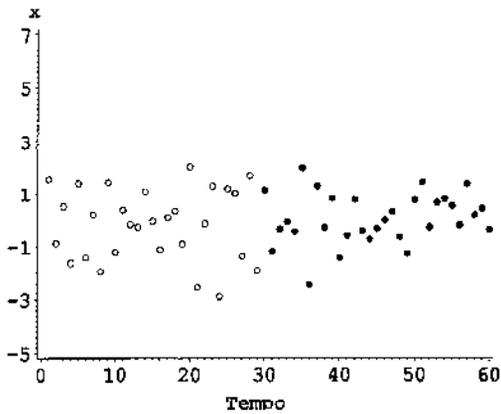
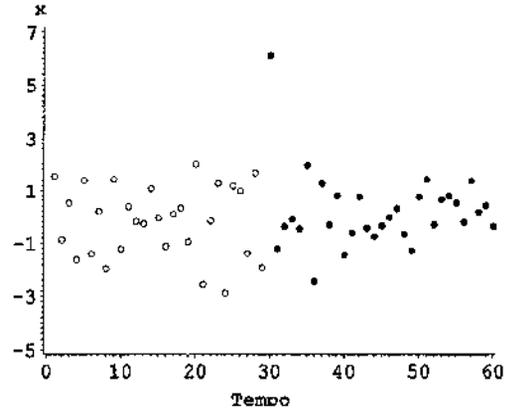
onde  $\omega_A$  representa a magnitude do desvio do “valor verdadeiro” de  $x_t$  e  $\delta_{t,k}$  é a variável indicadora da presença do outlier.

A média do novo processo é dada por  $E(y_t) = \omega_A \delta_{t,k}$ . Para ilustrar o efeito de um AO num modelo AR(1), foi introduzido um no tempo  $t = 30$  de magnitude  $\omega_A = 5$ . A série resultante é apresentada no Gráfico 1.3.1 (b). Pode-se verificar que todas as observações estão inalteradas, exceto no tempo  $t = 30$  em que o valor agora é 6,11, ao invés de 1,11, como seria na ausência do outlier.

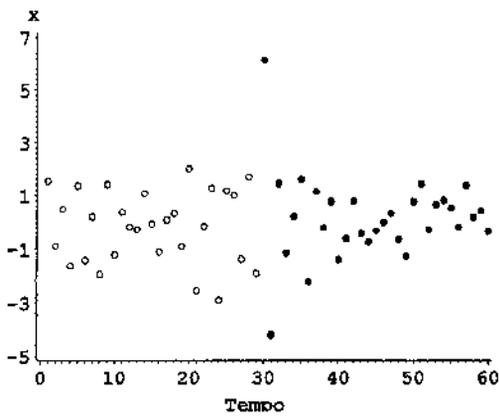
### 1.3.2 - Outlier de Inovação (IO)

Diferente de AO e dos outros tipos de outlier apresentados nesta seção, um *outlier de inovação* (IO), é o único evento cujo efeito é propagado para observações subsequentes de acordo com o modelo ARIMA subjacente.

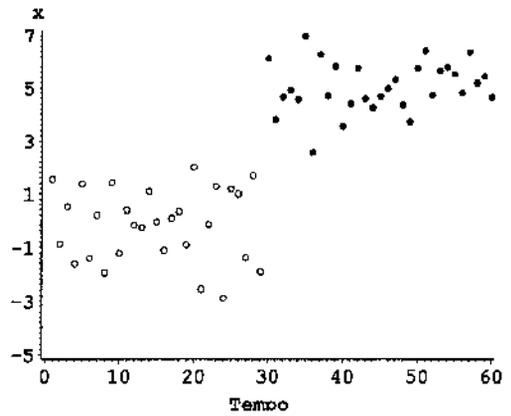
Esse modelo é também chamado de outlier de *Tipo II* (Fox, 1972) ou *inovação aberrante* (Abraham e Box, 1979). Será denominado nesse trabalho *outlier de inovação* ou simplesmente IO. Essas duas últimas terminologias surgem devido à representação dada em (1.3.5a) e pela série  $\{a_t\}$  ser referida como série de inovação.

a) Série  $x_t$  sem outlier

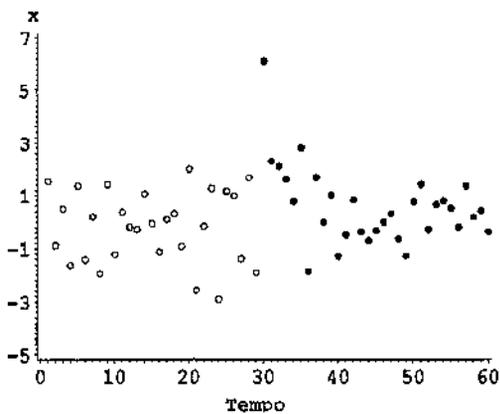
b) Série com AO



c) Série com IO



d) Série com LS

e) Série com TC ( $\lambda = 0,7$ )**Legenda:**

- - observações não afetadas pelo outlier
- - observações afetadas em pelo menos uma das séries.

**Gráfico 1.3.1** - Modelo AR(1) com  $\phi = -0,6$  geradas através das mesmas inovações (com  $\sigma_\varepsilon^2 = 1$ ) sem outlier ou com outlier de tamanho 5 no tempo  $t = 30$ .

Como o outlier afeta as inovações, a série observada é dada por

$$y_t = \frac{\theta(B)}{\phi(B)\alpha(B)} (a_t + \omega_1 \delta_{t,k}), \quad (1.3.5a)$$

ou na forma (1.3.2) como

$$y_t = x_t + \omega_1 \frac{\theta(B)}{\phi(B)\alpha(B)} \delta_{t,k}. \quad (1.3.5b)$$

Esses dois últimos modelos foram considerados por vários autores, entre eles Fox (1972), Chang e Tiao (1983), Hillmer *et alii* (1983) e Tsay (1986).

A diferença entre um AO e um IO pode ser melhor visualizada através das equações (1.3.5a) e (1.3.4). Pode-se ver em (1.3.4) que um AO apenas altera a observação  $x_k$ , enquanto um IO altera o choque  $a_k$ . Como resultado, um AO afeta apenas a observação  $y_k$ , enquanto o IO altera todos os valores de  $y_t$ , posteriores ao tempo  $t = k$ , de acordo com os pesos  $\pi_t$ s do modelo (1.1.3). Assim, as observações  $y_k, y_{k+1}, y_{k+2}, \dots$  e  $\hat{a}_k$  são afetadas pela inovação discrepante. Desta forma, tem-se que o comportamento das observações após a ocorrência do AO fornece informações sobre o outlier.

Um AO “pode ser interpretado como o efeito de uma mudança de alguns erros externos ou exógenos, ... (enquanto o IO) pode ser considerado como o efeito de alguma mudança interna ou efeito endógeno” (Peña, 1990). Assim, um AO pode ser um erro no registro ou algum evento econômico local, tal como greve, quando o seu efeito não é transferido para observações subsequentes.

Muitos autores afirmam que outlier aditivo é mais freqüente que o de inovação (ver, por exemplo, a discussão em Kleiner *et alii*, 1979).

O efeito de um IO num modelo AR(1), com um IO no tempo  $t = 30$  de magnitude  $\omega_1 = 5$  é ilustrado no Gráfico 1.3.1 (c). Os valores representados nos tempos  $t = 30, 31, 32, \dots$  são todos distintos do valor correspondente do Gráfico 1.3.1 (a), mas a diferença se torna difícil de visualizar aproximadamente após a observação 33, fato este já esperado como será discutido a seguir.

As raízes da equação  $\phi(B) = 0$  determinam se a transição tem decaimento geométrico (raízes reais) ou decaimento senoidal (raízes complexas) após o período  $t = k+q$ , onde  $q$  é ordem de MA. Entre o período  $t = k$  e  $t = k+q$ , a transição depende dos parâmetros de MA,  $\theta_j$ . Por exemplo, quando  $p = q = 1$ ,  $E(y_t) = \phi E(y_{t-1}) + \omega_1 \delta_{t,k} - \theta \omega_1 \delta_{t,k}$ ;

equivalentemente,  $E(y_t) = \omega_1 \phi^{t-k}(1-\theta\phi^{-1})$ , para  $t > k$ . Quando  $q = 2$ , o parâmetro  $\theta_2$  não aparece na transição até o tempo  $t = k+2$ .

O efeito de um IO numa série temporal é o mais complexo dentre todos os tipos de outliers (Chen e Liu, 1993b) apresentados neste trabalho. De acordo com a equação (1.3.5b), pode-se ver que o efeito de um IO depende dos pesos  $\psi_j$  do modelo ARIMA de  $x_t$  (1.1.2). Para uma série estacionária um IO deve produzir um efeito temporário dado que  $\psi_j$ s decaem exponencialmente para zero quando  $t \rightarrow \infty$ . No entanto, o padrão de  $\{\psi_j\}$  para uma série não estacionária pode ser bastante diferente. Dependendo do modelo de  $x_t$ , um IO pode produzir (Chen e Liu, 1993a):

1. um efeito inicial no tempo de sua ocorrência e uma mudança de nível num período seguinte quando  $x_t$  segue um modelo ARIMA(0,1,1) (ver os Gráficos 1.4.1 (a) e (b), pois o modelo de nível local de forma reduzida é um caso particular de ARIMA(0,1,1), neste caso, como  $\sigma_\eta = \sigma_\varepsilon = 1$ , o valor do parâmetro  $\theta$  é igual a 0,63, pela equação (1.2.5));
2. um efeito inicial no tempo de sua ocorrência que deve convergir gradualmente para uma mudança de nível permanente quando  $x_t$  segue um modelo ARIMA(1,1,1).

E caso esteja trabalhando com modelo ARIMA sazonal:

3. uma mudança de nível sazonal quando  $x_t$  segue um modelo ARIMA(0,1,1) sazonal puro, ex.: uma mudança de nível no mês janeiro de cada ano; ou
4. uma mudança de tendência quando  $x_t$  segue um modelo sazonal multiplicativo ARIMA(0,1,1) $\times$ (0,1,1).

Isso sugere que os outliers de mudança de nível (LS e TC, apresentados a seguir) podem ser aproximados por um efeito IO.

Ao contrário do outlier AO, devido à sua estrutura, o outlier IO não afeta a previsão dos dados, como será visto no Capítulo 3.

### 1.3.3 - Mudança de Nível - *Level Shift* - (LS)

Uma *mudança de nível* (LS) é um evento que afeta a série num determinado tempo, e este efeito permanece ao longo do tempo. Uma mudança de nível pode refletir numa mudança do processo de mecanismo gerador de dados, do aparelho de registro, da

definição da variável etc. (ver Box e Tiao, 1965). O modelo para a série observada pode ser apresentado por (com  $L(B) = 1/(1-B)$ )

$$y_t = x_t + \frac{\omega_L \delta_{t,k}}{1-B} = \begin{cases} x_t, & t < k, \\ x_t + \omega_L, & t \geq k. \end{cases} \quad (1.3.6a)$$

Esta equação pode ser reescrita como

$$y_t = x_t + \omega_L S_{t,k}, \quad (1.3.6b)$$

onde  $S_{t,k}$  é uma função degrau tal que  $S_{t,k}$  assume valor 0 (zero) para  $t < k$ , e 1 para  $t \geq k$ . Pode-se ver que o modelo para um AO, dado por (1.3.4), e o modelo para a mudança de nível, dado por (1.3.6b), seriam iguais se não fosse o fato de que AO afeta  $y_t$  apenas no tempo  $t = k$ , enquanto que um LS afeta  $y_t$  permanentemente após a sua ocorrência do tempo  $t = k$ .

Para ilustrar o efeito de um LS, foi incluído um LS no tempo  $t = 30$  numa série gerada pelo processo AR(1), como nos casos anteriores, o outlier é  $\omega_L = 5$ . O gráfico da série resultante é apresentado no Gráfico 1.3.1 (d). Pode-se observar que o comportamento da série é exatamente o mesmo da série sem outlier, exceto no nível em que antes da ocorrência do outlier ele é zero e após a ocorrência é 5.

### 1.3.4 - Mudança de Nível Temporário - *Temporaly Change* - (TC)

Um AO e LS representam duas formas em que um evento afeta uma série. Para LS, o nível do processo é afetado para todos os tempos subsequentes, enquanto que um outlier aditivo afeta apenas em um único tempo. É prático considerar um evento que tenha um impacto inicial sobre uma série, mas cujo impacto desaparece com o decorrer do tempo. Uma *mudança temporária* ou *transiente* (TC) é um evento que tem  $\omega_T$  como impacto inicial, onde o seu efeito decai exponencialmente de acordo com um fator de amortecimento ( $\lambda$ ). Na prática, o valor de  $\lambda$  pode ser especificado pelo pesquisador (Chen e Liu, 1993b). Pode-se representar a série observada como (com  $L(B) = 1/(1-\lambda B)$ )

$$y_t = x_t + \frac{\omega_T \delta_{t,k}}{1-\lambda B} = \begin{cases} x_t, & t < k, \\ x_{t+j} + \lambda^j \omega_T, & t + j \geq k, \end{cases} \quad (1.3.7)$$

onde  $0 < \lambda < 1$ . Pode-se ver pelas equações (1.3.4) e (1.3.6) que os outliers AO e LS são os casos limites (extremos) de outlier TC (1.3.7). Em (1.3.4) o fator de amortecimento  $\lambda$  é 0, enquanto que em (1.3.6) este fator é 1. Quanto maior o valor de  $\lambda$  mais lento é o

decaimento do impacto do outlier. No caso de um modelo AR(1) o efeito de um TC é equivalente ao de um IO se  $\lambda = \phi$  (ver as equações (1.3.5) e (1.3.7)).

Para distinguir TC de um AO ou LS, Chen e Liu (1993b) recomendam usar  $\lambda = 0,7$  (ver Tsay, 1986). O desempenho de tal especificação é apresentado no artigo de Chen e Liu.

Para ilustrar o efeito, foi incluído um TC no tempo  $t = 30$  com  $\omega_c = 5$  e  $\lambda = 0,7$ . A série resultante está apresentada no Gráfico 1.3.1 (e). Dado que  $\lambda$  é aproximadamente 1, o efeito de um outlier é significativo para um certo número de períodos (neste caso até aproximadamente  $t = 35$ ).

## 1.4 - Outlier em Modelos Estruturais

Considerando o modelo reduzido de nível local (1.2.3), se os outliers estão presentes no tempo  $k$ , no nível ( $\omega_1$ ) e na observação ( $\omega_2$ ), o modelo pode ser escrito como

$$y_t = \nabla x_t = (\eta_t + \omega_1 \delta_{t,k}) + \nabla(\varepsilon_t + \omega_2 \delta_{t,k}) = (1 - \theta B) \varepsilon_t.$$

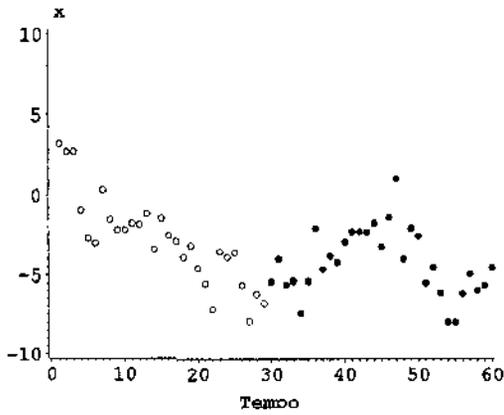
A série diferenciada pode também ser escrita como

$$y_t = (1 - \theta B)a_t + (\omega_1 + \omega_2)\delta_{t,k} - \omega_2\delta_{t-1,k}.$$

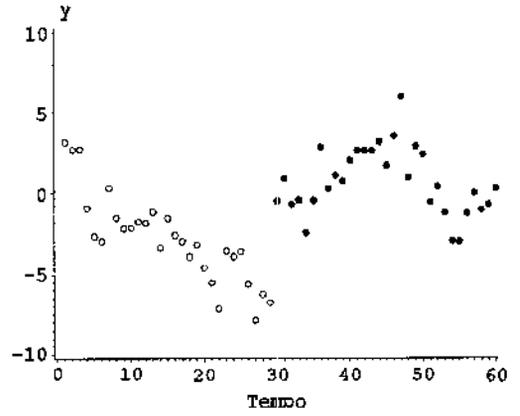
Logo, se  $\omega_1 \neq 0$  e  $\omega_2 = 0$ , o modelo acima é um MA(1) com um outlier AO no tempo  $k$ , e se  $\omega_2 \neq 0$  e  $\omega_1 = 0$  então esse modelo é um MA(1) com outliers AO nos tempos  $k$  e  $k+1$  de tamanhos  $\omega_2$  e  $-\omega_2$ , respectivamente. Caso ambos sejam diferentes de zero este modelo continua sendo um MA(1) com outliers nos tempos  $k$  e  $k+1$ , mas com magnitudes  $\omega_1 + \omega_2$  e  $-\omega_2$ , respectivamente.

Para ilustrar o efeito foram geradas duas séries de tamanho  $n = 60$ , uma sem outlier e a outra com um outlier na inovação do nível da série ( $\eta_t$ ) no tempo  $t = 30$  de tamanho  $\omega = 5\sigma_\eta$ , as duas séries foram geradas através de mesmas inovações. As inovações foram geradas como sendo de uma distribuição normal padrão, independentes. Os valores iniciais utilizados foram zero e para diminuir o efeito de valores iniciais as primeiras 100 observações foram eliminadas. Analisando essas séries resultantes, apresentadas nos Gráficos 1.4.1 (a) e (b), pode-se observar que o efeito do outlier é uma mudança de nível nos períodos seguintes, como foi observado no item 1 da subseção 1.3.2.

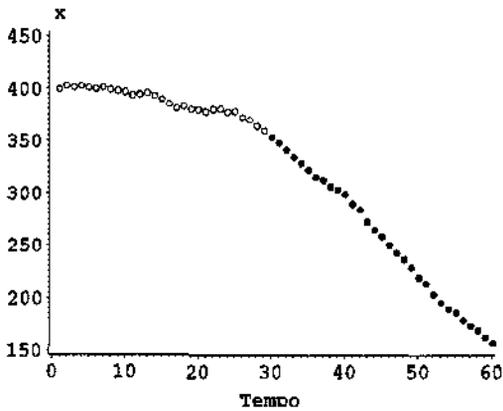
No modelo de tendência local, se os outliers estão presentes no nível ( $\omega_1$ ), tendência ( $\omega_2$ ) e/ou na observação ( $\omega_3$ ), no tempo  $k$ , o modelo pode ser escrito como



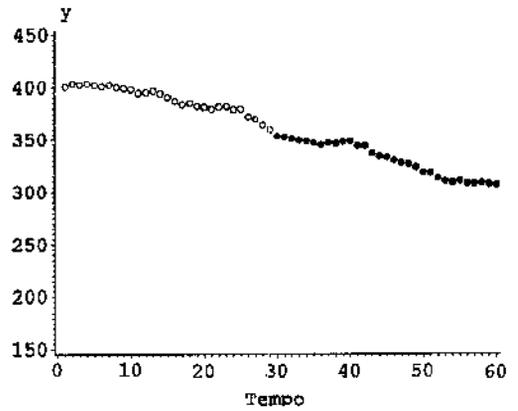
a) Modelo de nível local sem outlier



b) Modelo de nível local com outlier ( $5\sigma_{\eta}$ ) na posição 30 na inovação do nível ( $\eta_t$ )



c) Modelo de tendência local sem outlier



d) Modelo de tendência local com outlier ( $5\sigma_{\zeta}$ ) na posição 30 na inovação da tendência ( $\zeta_t$ )

**Legenda:**

- - observações não afetadas pelo outlier
- - observações a partir da ocorrência do outlier.

**Gráfico 1.4.1** - Modelos estruturais sem e com outlier, gerados com as mesmas inovações.

$$\begin{aligned}
 y_t = \nabla^2 x_t &= (1-B)(\eta_t + \omega_1 \delta_{t,k}) + (\zeta_{t-1} + \omega_2 \delta_{t-1,k}) + (1-2B-B^2)(\varepsilon_t + \omega_3 \delta_{t,k}) \\
 &= (1-\theta_1 B - \theta_2 B^2) a_t + (\omega_1 + \omega_3) \delta_{t,k} + (\omega_2 - \omega_1 - 2\omega_3) \delta_{t-1,k} + \omega_3 \delta_{t-2,k}.
 \end{aligned}$$

Após as diferenças a série pode ser escrita como:

$$y_t = (1-\theta_1 B - \theta_2 B^2) a_t + (\omega_1 + \omega_3) \delta_{t,k} + (\omega_2 - \omega_1 - 2\omega_3) \delta_{t-1,k} + \omega_3 \delta_{t-2,k}.$$

A presença de um outlier apenas no nível, i.e.,  $\omega_1 \neq 0$ , é equivalente a outliers AO nos tempos  $k$  e  $k+1$  de mesmas magnitudes, mas de sinais opostos. A presença de um outlier apenas na tendência  $\omega_3 \neq 0$ , é equivalente a outliers AO nos tempos  $k$ ,  $k+1$  e  $k+2$  de tamanhos  $\omega_1 + \omega_3$ ,  $-2\omega_3$  e  $\omega_3$ , respectivamente. Diferentes combinações são possíveis, mas todas elas resultam em outlier AO nos tempos  $k$ ,  $k+1$  ou  $k+2$ .

Para ilustrar um outlier na inovação da tendência da série foram geradas duas séries com as três inovações de uma normal padrão independentes. A primeira série foi gerada sem outlier e a segunda com um outlier de tamanho  $5\sigma_\zeta$  na inovação da tendência ( $\zeta_t$ ) na posição  $t = 30$ , todas as variâncias foram consideradas de tamanho unitário, i.e.,  $\sigma_\eta^2 = \sigma_\zeta^2 = 1$ . As séries resultantes estão apresentadas nos Gráficos 1.4.1 (c) e (d). O efeito causado pelo outlier é uma mudança brusca da tendência da série.

## Capítulo 2

### Testes de detecção de outliers

São muitos os testes de detecção de outliers e de pontos influentes propostos na literatura. Nesse capítulo são apresentados alguns dos testes mais utilizados, todos baseados na função de verossimilhança. O teste da Razão de Verossimilhança, apresentado na seção 2.1, é o mais antigo e mais utilizado na literatura. O segundo teste, apresentado na seção 2.2, é o teste do Escore também conhecido como teste do Multiplicador de Lagrange. Diferente do teste da Razão de Verossimilhança ele é baseado apenas em estimativas do modelo sob a hipótese nula, que para o caso em estudo é mais fácil de calcular. Na seção 2.3 são apresentados além do teste do Deslocamento da Verossimilhança algumas medidas de influência dos outliers. Essas medidas de influência são análogas à distância de D de Cook utilizada em regressão linear.

#### 2.1 Teste da Razão de Verossimilhança

O teste da Razão de Verossimilhança<sup>1</sup> (RV) para detecção de outliers em séries temporais, desenvolvido por Fox (1972), foi o primeiro teste nesse contexto. Ele desenvolveu o teste para um processo autoregressivo puro com ordem conhecida, considerando a posição do outlier conhecida e desconhecida, para outliers do tipo IO e AO. Este teste foi posteriormente generalizado para processos ARMA por Chang e Tiao (1983), Tsay (1986, 1988), Chang *et alii* (1988), Chen e Liu (1993a) e outros. Essas generalizações não se limitaram apenas aos testes de detecção de outliers, procuraram também melhorar as estimativas dos parâmetros e a identificação do modelo e diminuir os problemas de mascaramento, utilizando procedimentos iterativos.

---

<sup>1</sup> A idéia básica de teste de RV é apresentada no Apêndice 1.

Chang e Tiao (1983) propuseram um procedimento iterativo, com dois estágios, para detectar múltiplos outliers. No primeiro estágio são estimadas as possíveis posições e os tipos dos outliers e no segundo estágio os parâmetros são reestimados considerando essas posições e tipos dos outliers como conhecidos. Tsay (1986) incluiu mais um estágio de especificação nesse procedimento de forma a lidar com o caso de especificação desconhecida. Tsay (1988) acrescentou ao procedimento de Chang e Tiao (1983) outros tipos de outliers (LS e TC) e mudanças de variância na série. Balke (1993) verificou que esse último procedimento nem sempre tem um desempenho satisfatório na presença de LS na série. Para amenizar esse problema ele propôs uma modificação no procedimento acrescentando mais um estágio de detecção de outliers. Chen e Liu (1993a) propuseram uma outra mudança no procedimento de Tsay (1988) procurando diminuir o problema de vício na estimação dos parâmetros e efeitos de mascaramento, através da estimação conjunta dos parâmetros do modelo e das magnitudes dos outliers.

Como observado por Martin (1980) e Chang e Tiao (1983) a carga computacional é um dos maiores problemas ao lidar com outliers em séries temporais. Para tentar contornar esse problema têm sido feitas modificações na estimação dos parâmetros. Por exemplo, ao invés de maximizar a complicada função de verossimilhança, proposto por Fox (1972), Tsay (1986) propôs um procedimento que utiliza o método de mínimos quadrados, ou seja uma técnica de regressão linear para obter a estimativa dos parâmetros. Essa mudança reduz substancialmente a carga computacional envolvida no processo de identificação iterativa e aumenta muito o valor prático dos procedimentos propostos. O método de estimação proposto por Tsay (1986) está apresentado na subseção seguinte.

Os testes considerados nesta seção são:

$$H_0 \text{ vs. } H_{TP},$$

onde  $H_0$  é a hipótese nula de que não existe nenhum outlier, i.e.,

$$H_0: \omega_A = \omega_I = \omega_L = \omega_T = 0$$

e as alternativas, de que há um outlier do tipo TP, i.e.,

$$H_{TP}: \omega_{TP} \neq 0 \text{ e } \omega_j = 0 \text{ para } j \neq TP,$$

TP = A, I, L e T para AO, IO, LS e TC, respectivamente.

Nas duas subseções seguintes (2.1.1 e 2.1.2) estão apresentados os testes de detecção de outliers quando os parâmetros do modelo são conhecidos e desconhecidos,

respectivamente. Posteriormente, na subseção 2.1.3 é apresentado um procedimento iterativo para detectar outlier e estimar os parâmetros. Na subseção 2.1.4 são apresentados alguns resultados de poder do teste dos procedimentos anteriores juntamente com a escolha dos valores críticos. Na última subseção, 2.1.5, são discutidos alguns problemas da identificação incorreta do outlier que acarreta na previsão.

### 2.1.1 Testes de detecção quando os parâmetros do modelo são conhecidos

Considere o modelo com um outlier do tipo TP no tempo k, dado por

$$y_t = \frac{\theta(B)}{\phi(B)\alpha(B)} a_t + f(t) = \frac{\theta(B)}{\phi(B)\alpha(B)} e_t \quad (2.1.1)$$

onde  $f(t) = \omega_{TP}L(B)\delta_{t,k}$  foi definido em (1.3.2). Quando os coeficientes e a posição do outlier são conhecidos, os resíduos para cada tipo de outlier, como apresentado por Chang *et alii* (1988) e Tsay (1988), são dados por

$$e_t = \begin{cases} \omega_I \delta_{t,k} + a_t & \text{para IO;} \\ \omega_A \pi(B) \delta_{t,k} + a_t & \text{para AO;} \\ \frac{\omega_L}{1-B} \pi(B) \delta_{t,k} + a_t & \text{para LS;} \\ \frac{\omega_T}{1-\lambda B} \pi(B) \delta_{t,k} + a_t & \text{para TC;} \end{cases} \quad (2.1.2)$$

onde  $e_t = \phi(B)(1-B)^d \theta^{-1}(B)y_t$ . Devido a condição de invertibilidade o termo do erro  $e_t$  pode ser considerado como dado, sob a hipótese de que os coeficientes são conhecidos e que o outlier ocorre longe do fim da série. Assim o teste da presença do outlier (sob certas condições) é equivalente ao teste de significância dos parâmetros  $\omega$  nos modelos (2.1.2).

Dado que os parâmetros  $\phi_i$ s e  $\theta_j$ s e k são conhecidos e  $f(t)$  é determinístico, os estimadores de  $\omega_{TP}$ ,  $TP = I, A, L$  e  $T$ , do modelo (2.1.1) podem ser obtidos usando uma técnica de regressão linear simples (Chang, 1982).

O modelo (2.1.2) pode ser escrito como uma equação de regressão linear simples

$$e_t = \omega_{TP} w_{TP,t} + a_t \quad (2.1.3)$$

onde  $\omega_{TP}$  é o coeficiente de regressão e

$$w_{TP,t} = \frac{\phi(B)\alpha(B)}{\theta(B)} L(B)\delta_{t,k} = \zeta(B)\delta_{t,k},$$

$$= \begin{cases} 0, & t < k \\ \xi_{t-k} & t \geq k \end{cases}$$

a variável explanatória, onde  $\xi(B)$  pode ser escrito como um polinômio em  $B$  dado por  $\xi(B) = 1 - \xi_1 B - \dots$ . Desta forma tem-se que

$$\hat{\omega}_{TP} = \frac{\sum_{t=1}^n e_t w_{TP,t}}{\sum_{t=1}^n w_{TP,t}^2} \quad \text{e} \quad \text{Var}(\hat{\omega}_{TP}) = \frac{\sigma_a^2}{\sum_{t=1}^n w_{TP,t}^2}, \quad (2.1.4)$$

onde  $TP = I, A, L, T$  e  $w_{TP,t}$  assume valor zero para  $t < k$ ; valor 1 para  $t = k$ ; e

$$w_{TP,k+m} = \begin{cases} 0 & TP = I; \\ -\pi_m & TP = A; \\ 1 - \sum_{j=1}^{m-1} \pi_j & TP = L; \\ \lambda^m - \sum_{j=1}^{m-1} \lambda^{m-j} \pi_j - \pi_m & TP = T, \end{cases} \quad (2.1.5)$$

para  $k + m$  ( $m = 1, 2, \dots, n - k$ ).

Usando esta técnica pode-se obter os estimadores do impacto dos quatro tipos de outliers:

- a) IO:  $\hat{\omega}_{I,k} = e_k$  e  $\text{Var}(\hat{\omega}_{I,k}) = \sigma_a^2$ ,
- b) AO:  $\hat{\omega}_{A,k} = \rho_{A,k}^2 (e_k - \sum_{i=1}^{n-k} \pi_i e_{k+i})$  e  $\text{Var}(\hat{\omega}_{A,k}) = \rho_{A,k}^2 \sigma_a^2$ , onde  $\pi_i$ s são os pesos  $\pi$  de  $x_t$  e  $\rho_{A,k}^2 = (1 + \pi_1^2 + \dots + \pi_{n-k}^2)^{-1}$ ,
- c) LS:  $\hat{\omega}_{L,k} = \rho_{L,k}^2 (e_k - \sum_{i=1}^{n-k} \gamma_i e_{k+i})$  e  $\text{Var}(\hat{\omega}_{L,k}) = \rho_{L,k}^2 \sigma_a^2$ , onde  $\gamma_i$  é o coeficiente de  $B^i$  no polinômio  $\gamma(B) = \gamma_0 - \gamma_1 B - \gamma_2 B^2 - \dots = \pi(B)/(1-B)$  e  $\rho_{L,k}^2 = (1 + \gamma_1^2 + \dots + \gamma_{n-k}^2)^{-1}$ .
- d) TC:  $\hat{\omega}_{T,k} = \rho_{T,k}^2 (e_k - \sum_{i=1}^{n-k} \beta_i e_{k+i})$  e  $\text{Var}(\hat{\omega}_{T,k}) = \rho_{T,k}^2 \sigma_a^2$ , onde  $\beta_i$  é o coeficiente de  $B^i$  no polinômio  $\beta(B) = \beta_0 - \beta_1 B - \beta_2 B^2 - \dots = \pi(B)/(1-\lambda B)$ ,  $\rho_{T,k}^2 = (1 + \beta_1^2 + \dots + \beta_{n-k}^2)^{-1}$  e  $\lambda$  é uma constante pré-especificada satisfazendo  $0 < \lambda < 1$ .

As estatísticas dos testes da RV, dado que a posição do outlier é conhecida são:

$$a) H_0 \text{ vs. } H_I: \lambda_{I,k} = \hat{\omega}_{I,k} / \sigma_a \quad (2.1.6a)$$

$$b) H_0 \text{ vs. } H_A: \lambda_{A,k} = \hat{\omega}_{A,k} / \rho_{A,k} \sigma_a \quad (2.1.6b)$$

$$c) H_0 \text{ vs. } H_L: \lambda_{L,k} = \hat{\omega}_{L,k} / \rho_{L,k} \sigma_a \quad (2.1.6c)$$

$$d) H_0 \text{ vs. } H_T: \lambda_{T,k} = \hat{\omega}_{T,k} / \rho_{T,k} \sigma_a \quad (2.1.6d)$$

Em regressão linear do tipo  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ ,  $i = 1, \dots, n$ , com a suposição de normalidade de  $y_i$ , a estatística  $t$  de Wald para o teste  $H_0: \beta_i = 0$  vs.  $H_1: \beta_i \neq 0$  é dada por

$$t = \frac{\hat{\beta}_i}{\sqrt{\text{Var}(\hat{\beta}_i)}},$$

onde  $\hat{\beta}_i$  é o EMV de  $\beta_i$ . A relação entre esta estatística e o teste de RV ( $\lambda$ ) para as mesmas hipóteses é

$$t^2 = (n-2)(\lambda^{-2/n} - 1).$$

Assim, por essa relação os testes são equivalentes (ver, por exemplo, Mood, Graybill e Boes, 1974, p.495).

No contexto de análise de regressão linear, com distribuição normal, (sem a suposição de normalidade é suficiente que a verossimilhança tenha uma forma quadrática, pelo menos na região de interesse) as três famílias de testes RV, Multiplicador de Lagrange e Wald são equivalentes para testar a significância dos parâmetros  $\omega_{TP}$ ,  $TP = I, A, L$  e  $T$  (Buse, 1982).

Quando a posição do outlier não é conhecida Chang *et alii* (1988) propõem usar as máximas das estatísticas calculadas para cada tempo:

$$\lambda_I = \max_{k=1, \dots, n} |\lambda_{I,k}|, \quad \lambda_A = \max_{k=1, \dots, n} |\lambda_{A,k}|, \quad \lambda_L = \max_{k=1, \dots, n} |\lambda_{L,k}| \quad \text{e} \quad \lambda_T = \max_{k=1, \dots, n} |\lambda_{T,k}|.$$

A hipótese  $H_0$  é rejeitada quando pelo menos uma destas estatísticas é maior que o ponto de corte  $C$ . Este ponto de corte é um valor pré determinado discutido nas próximas subseções. Quando duas ou mais estatísticas são maiores que  $C$ , Chang *et alii* (1988) sugerem considerar apenas a maior estatística em valor absoluto; quando o tipo do outlier é desconhecido estes autores sugerem usar uma regra simples mencionada por Fox (1972). Por exemplo, considera como IO se  $\max\{\lambda_I, \lambda_A, \lambda_L, \lambda_T\} = \lambda_I$ .

### 2.1.2 Teste de detecção quando os parâmetros do modelo são desconhecidos

Na prática os parâmetros do modelo ARMA são geralmente desconhecidos. Estimadores destes parâmetros conjuntamente com  $\omega$  podem ser obtidos maximizando a função de verossimilhança de  $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \omega, \sigma_a^2)$  da mesma forma que é descrito em Box e Jenkins (1976, p.208). A estatística da razão de verossimilhança pode ser calculada baseada nestes estimadores. Segue-se assim um critério para detectar um outlier numa posição desconhecida.

Os estimadores de máxima verossimilhança (EMV) para esses parâmetros não se reduzem a expressões simples devido à natureza não linear nos parâmetros. Quando a posição do outlier for desconhecida a carga computacional se tornará muito pesada e com isso fica inviável usar o EMV neste caso. Desta forma, aparecem naturalmente estatísticas simplificadas como uma aproximação do critério de razão de verossimilhança.

Uma destas estatísticas surge quando os parâmetros das estatísticas dos testes mencionados na subseção anterior são substituídos por estimadores consistentes. Sob a hipótese nula de não ter nenhum distúrbio externo, os estimadores de máxima verossimilhança (EMV) são consistentes e podem ser usados para obter uma justificativa assintótica para o uso das estatísticas apresentadas na subseção anterior. Mais especificamente, os testes (2.1.6.a ~ d), com os parâmetros substituídos por EMV, têm uma distribuição assintótica  $N(0,1)$  sob a hipótese nula e a suposição de que a posição do outlier é conhecida. A consistência de EMV para modelos estacionários ARMA pode ser visto em Chang (1982) e o caso para modelos não estacionários foi estudado por Tsay e Tiao (1984), concluindo, de maneira geral, que o fator não estacionário de  $x_t$  não afeta a propriedade de consistência dos estimadores, pois a taxa de convergência dos parâmetros não estacionários é mais rápida que a dos parâmetros estacionários.

Sejam  $\hat{\phi}_1, \dots, \hat{\phi}_p; \hat{\theta}_1, \dots, \hat{\theta}_q;$  e  $\hat{\sigma}_a^2$  os EMV dos parâmetros  $\phi_1, \dots, \phi_p; \theta_1, \dots, \theta_q; \sigma_a^2$ , respectivamente, obtidos ignorando a presença do outlier na série temporal  $y_t$ . Além disso, seja  $\hat{e}_t$  os resíduos calculados por esse modelo estimado e  $\hat{\pi}(B) = \hat{\phi}(B)\alpha(B) / \hat{\theta}(B)$ . Serão consideradas as seguintes estatísticas:

$$\hat{\lambda}_{I,t} = \hat{\omega}_{I,t} / \hat{\sigma}_a \quad (2.1.7a)$$

$$\hat{\lambda}_{A,t} = \hat{\omega}_{A,t} / (\hat{\rho}_{A,t} \hat{\sigma}_a) \quad (2.1.7b)$$

$$\hat{\lambda}_{L,t} = \hat{\omega}_{L,t} / (\hat{\rho}_{L,t} \hat{\sigma}_a) \quad (2.1.7c)$$

$$\hat{\lambda}_{T,t} = \hat{\omega}_{T,t} / (\hat{\rho}_{T,t} \hat{\sigma}_a) \quad (2.1.7d)$$

onde

$$\begin{aligned}\hat{\omega}_{I,t} &= \hat{e}_t, \\ \hat{\omega}_{A,k} &= \hat{\rho}_{A,k}^2 (\hat{e}_k - \sum_{i=1}^{n-k} \hat{\pi}_i \hat{e}_{k+i}), \\ \hat{\omega}_{L,k} &= \hat{\rho}_{L,k}^2 (\hat{e}_k - \sum_{i=1}^{n-k} \hat{\gamma}_i \hat{e}_{k+i}), \\ \hat{\omega}_{T,k} &= \hat{\rho}_{T,k}^2 (\hat{e}_k - \sum_{i=1}^{n-k} \hat{\beta}_i \hat{e}_{k+i}),\end{aligned}$$

e

$$\begin{aligned}\hat{\rho}_{A,k}^2 &= (1 + \hat{\pi}_1^2 + \dots + \hat{\pi}_{n-k}^2)^{-1}, \\ \hat{\rho}_{L,k}^2 &= (1 + \hat{\gamma}_1^2 + \dots + \hat{\gamma}_{n-k}^2)^{-1}, \text{ onde } \hat{\gamma}_i \text{ é o coeficiente de } B^i \text{ no polinômio} \\ &\hat{\gamma}(B) = \hat{\gamma}_0 - \hat{\gamma}_1 B - \hat{\gamma}_2 B^2 - \dots = \hat{\pi}(B)/(1-B), \\ \hat{\rho}_{T,k}^2 &= (1 + \hat{\beta}_1^2 + \dots + \hat{\beta}_{n-k}^2)^{-1} \text{ onde } \hat{\beta}_i \text{ é o coeficiente de } B^i \text{ no polinômio} \\ &\hat{\beta}(B) = \hat{\beta}_0 - \hat{\beta}_1 B - \hat{\beta}_2 B^2 - \dots = \hat{\pi}(B)/(1-\lambda B), \text{ e } \lambda \text{ é uma constante pré-} \\ &\text{especificada satisfazendo } 0 < \lambda < 1.\end{aligned}$$

Pode-se mostrar que  $\hat{\lambda}_{TP,t}$ , é assintoticamente equivalente ao critério de razão de verossimilhança  $\lambda_{TP,t}$ , numa dada observação  $t$ , para os testes  $H_0$  vs.  $H_{TP}$  para cada  $TP = I, A, L$  e  $T$  (ver Chang, 1982).

Geralmente há pouca informação disponível sobre o tipo de outlier presente, o que dificulta saber qual dos testes é mais apropriado para determinada situação: (2.1.7a), (b), (c) ou (d). Quando um teste do tipo não apropriado é usado o poder de detecção reduz-se substancialmente (Chang *et alii*, 1988). Além disso, mesmo sabendo a posição de ocorrência do outlier, a remoção do mesmo não será possível até que a natureza do outlier seja identificada.

Chang (1982) e Muirhead (1986) desenvolveram testes da RV aproximados para distinguir entre AO e IO. Estes testes requerem muitos cálculos computacionais tornando a aplicação uma tarefa bastante demorada e trabalhosa para utilização prática.

Para simplificar o problema, geralmente usa-se uma regra simples, proposta por Fox (1972) como uma possível forma de distinguir entre os tipos de outlier. Em qualquer observação  $k$  suspeita, o possível outlier é classificado como IO se  $\max_{TP} \{|\lambda_{TP,k}|\} = |\lambda_{L,k}|$ , classificado como AO se  $\max_{TP} \{|\lambda_{TP,k}|\} = |\lambda_{A,k}|$ , e assim por diante.

Para verificar a magnitude da perda de informação decorrente dessa simplificação, Chang (1982) realizou uma simulação para séries AR de tamanho  $n = 50$ . Para um outlier de tamanho  $\omega = 3\sigma_a$ , no meio da série, utilizando um nível de significância de 10%, essa regra detectou corretamente como AO ou IO, em 78,6% dos casos, enquanto que para o teste de RV aproximado para distinguir os tipos de outlier, o sucesso médio foi de 87,3%. Quando  $\omega$  cresce para  $5\sigma_a$ , a regra simples identifica corretamente 92% dos casos, que é quase o mesmo de 92,4% atingido pelo teste de RV aproximados, ao nível de significância 10%.

Para detectar um outlier numa posição desconhecida, pode-se procurar através da seqüência  $\hat{\lambda}_{TP,t}$ ,  $t = 1, \dots, n$ ;  $TP = I, A, L, T$ . Em outras palavras, a possibilidade de um outlier na série pode ser testada verificando se pelo menos um desses estimadores é maior que o ponto de corte  $C$ , onde  $C$  é uma constante positiva escolhida adequadamente. Chang *et alii* (1988) realizaram simulações em modelos AR e MA para estimar os tamanhos destes testes para diferentes magnitudes do valor crítico (para outliers IO e AO). Eles recomendaram, através do resultado das simulações, usar  $C = 3,0$  para alta sensibilidade;  $C = 3,5$  para média sensibilidade e  $C = 4,0$  para baixa sensibilidade, no procedimento de detecção de outlier quando o tamanho da série é menor que 200. Este teste é razoável para tamanhos de amostra moderados e outliers de tamanhos  $3\sigma_a$  e  $5\sigma_a$  (Chang, 1982).

Chang e Tiao (1983) e Tsay (1986a) obtiveram resultados satisfatórios usando esses três valores de  $C$ . As distribuições exatas das estatísticas usadas, como por exemplo  $\lambda_A$ , são intratáveis pois elas são os máximos de variáveis dependentes. No entanto, algumas pesquisas foram feitas para obter boas aproximações dos percentis destas estatísticas, ver, por exemplo, Atwood (1989).

### **2.1.3 Um procedimento iterativo para detecção de outliers e estimação dos parâmetros**

Chang *et alii* (1988) sugeriram um procedimento iterativo de dois estágios, baseado no procedimento de Chang e Tiao (1983) para lidar com situações onde não se conhece o número de outliers. As descrições dos procedimentos são apresentadas a seguir, com algumas alterações em virtude dos outliers LS e TC, não terem sido considerados por esses autores em seus trabalhos.

**I) Estágio de Detecção de Outlier**

- 1 - Pelo modelo estimado calcule os resíduos  $\hat{e}_t$  e estime  $\sigma_a^2$  através de  $\hat{\sigma}_a^2 = \sum_{i=1}^n \hat{e}_i^2 / n$  ou usando algum estimador robusto<sup>1</sup>;
- 2 - calcule  $\hat{\lambda}_{TP,t}$ ,  $t = 1, \dots, n$ ;  $TP = I, A, L, T$  e  $\hat{\lambda} = \max_{TP,t} \{|\hat{\lambda}_{TP,t}|\}$ . Se  $\hat{\lambda} < C$ , onde  $C$  é uma constante positiva pré-determinada, vá para o estágio de estimação de parâmetros, caso contrário remova o efeito do outlier nos resíduos, definindo um novo conjunto de resíduos como

$$\tilde{e}_t = \begin{cases} \hat{e}_t - \hat{\omega}_{L,k} = 0 & t = k, & \text{para IO,} \\ \hat{e}_t - \hat{\omega}_{A,k} \hat{\pi}(B) \delta_{t,k} & t = k, \dots, n - k & \text{para AO,} \\ \hat{e}_t - \hat{\omega}_{L,k} \hat{\gamma}(B) \delta_{t,k} & t = k, \dots, n - k & \text{para LS,} \\ \hat{e}_t - \hat{\omega}_{T,k} \hat{\beta}(B) \delta_{t,k} & t = k, \dots, n - k & \text{para TC;} \end{cases}$$

- 3 - se algum outlier foi identificado no passo 2, recalcule os  $\hat{\lambda}_{TP,t}$  baseado nas mesmas estimativas dos parâmetros do modelo, mas usando os resíduos modificados  $\tilde{e}_t$ 's e a estimativa  $\tilde{\sigma}_a^2$  e repita o passo 2;
- 4 - repita os passos 2 e 3 até não identificar mais nenhum candidato para ser outlier;

**II - Estágio de Estimação dos Parâmetros**

- 5 - Suponha que tenham sido definidos  $b$  tempos  $k_1, \dots, k_b$  como possíveis ocorrências de outliers de tipo  $TP_1, \dots, TP_b$ , respectivamente. Trate esses tempos e os tipos de outliers como conhecidos e estime simultaneamente as magnitudes dos outliers  $\omega_{TP_1}, \dots, \omega_{TP_b}$  e os parâmetros do modelo usando o modelo de intervenção de Box e Tiao (1975), i.e., o modelo definido como

$$y_t = \sum_{j=1}^b \omega_{TP_j} L_j(B) \delta_{t,k_j} + \frac{\theta(B)}{\phi(B)\alpha(B)} a_t, \tag{2.1.8}$$

onde

---

<sup>1</sup> Exemplos de estimadores robustos: *leave-one-out*, MAD.

$$L_j(\mathbf{B}) = \begin{cases} 1 & , \text{ para AO,} \\ 1/\pi(\mathbf{B}) & , \text{ para IO,} \\ 1/(1-\mathbf{B}) & , \text{ para LS,} \\ 1/(1-\lambda\mathbf{B}) & , \text{ para TC,} \end{cases}$$

no tempo  $t = k_j$ . Usando este modelo estimado, retorne ao estágio I. As notações  $\hat{\pi}_j$ s  $\hat{\omega}_{\text{TP}j}$ s e  $\hat{\epsilon}_j$ s representam os valores obtidos pela estimação conjunta de todos os parâmetros do modelo anterior. Se nenhum outlier foi encontrado então pare a iteração. Caso contrário, o estágio de estimação é repetido, com os novos outliers identificados incorporados nesse modelo anterior, até que nenhum outlier possa ser encontrado e todos os efeitos tenham sido simultaneamente estimados com os parâmetros do modelo.  $\square$

Os cálculos no passo 2 podem ser realizados recursivamente para  $t = 1, 2, \dots, n$  pois todas as modificações necessárias de  $t$  para  $t+1$  envolvem apenas a observação  $y_{t+1}$ .

É possível detectar mais de um distúrbio dentro de uma iteração interna (estágio I). No entanto, experiências sugerem que as técnicas que detectam um de cada vez têm melhores desempenhos, em termos de poder, nas aplicações (Tsay, 1988). Uma alternativa é usar teste estatístico para distúrbios múltiplos (ver por exemplo, Wu *et alii*, 1993), mas as possíveis combinações para distúrbios múltiplos podem complicar bastante a aplicação prática.

Quando as ordens do processo não são conhecidas elas devem ser escolhidas no início e as funções de autocorrelação têm uma função-chave durante este estágio. No entanto, a presença do outlier pode afetar as autocorrelações (parciais) amostrais (Chang, 1982) e a função de autocorrelação estendida (Tsay e Tiao, 1984) e levar a uma especificação distorcida. Por exemplo, um AR(1) pode ser facilmente identificado como um processo ARMA(1,1) (Ledolter, 1987). Com o objetivo de superar esses problemas Tsay (1986a) propôs incluir um estágio de especificação no procedimento. Ele testou esse procedimento modificado usando o método de identificação, através da função de autocorrelação estendida, de Tsay e Tiao (1984), no estágio de especificação. Este estágio poderia, no entanto, ter sido feito utilizando qualquer outra técnica, como por exemplo, um critério automático. Uma comparação de diferentes tipos de técnicas de especificação automática pode ser encontrada, por exemplo, em Koreisha e Yoshimoto (1991).

Para uma dada locação as estatísticas dos testes padronizados (2.1.6a ~ d), seguem uma distribuição aproximada a uma normal padrão. Conhecendo o tipo e a locação de um outlier, a série pode ser ajustada usando a equação (2.1.8). Em geral, a observação ajustada no tempo  $t_1$ , denotado por  $\tilde{y}_{t_1}$ , pode ser expressa como uma soma ponderada de todo o resto da série observada. No caso de IO, pode ser visto que a observação ajustada  $\tilde{y}_{t_1}$  é a esperança condicional de  $y_{t_1}$  dadas as observações passadas. No caso de um AO, a observação ajustada é a interpolação baseada em ambos, passado e futuro de  $y_{t_1}$ , sem envolver a própria observação. Isso sugere um possível procedimento para estimar os outliers em séries temporais tratando-os como dados faltantes. Veja alguns procedimentos propostos por Bruce e Martin (1989) e Ljung (1993) para o caso de AO; pode ser visto também um procedimento de estimação usando algoritmo EM de Dempster *et alii* (1977) em Abraham e Chuang (1987) para AO e IO.

Tsay (1988) generalizou o procedimento anterior (de Chang *et alii*, 1988) incluindo os LS e TC, e acrescentando um estágio de detecção de mudança de variância. Embora o procedimento apresentado inclua quatro tipos de outliers, o procedimento original de Chang e Tiao (1983) considera apenas AO e IO.

Balke (1993) verificou que o procedimento de Tsay nem sempre tem um desempenho satisfatório quando LS está presente na série. Primeiro, a presença de LSs causa sérios problemas na especificação e estimação no modelo ARMA inicial, que afeta os passos seguintes de detecção e remoção dos outliers. Segundo, a regra utilizada por Tsay para distinguir entre LS e IO não funciona bem na presença de LSs. Assim, como consequência, o procedimento de Tsay pode identificar erroneamente os LSs.

Para lidar com esse problema Balke sugere uma modificação simples no procedimento de Tsay:

- 1) Fase de detecção de outlier. Identificar duas listas de possíveis outliers. A primeira lista é identificada ajustando-se um modelo ARIMA e aplicando o método de Tsay. A segunda lista é identificada ajustando-se um modelo com uma constante e/ou uma tendência determinística mais um ruído branco, como modelo inicial.
- 2) Fase de redução de outliers. Combinar todos os possíveis outliers encontrados na fase anterior num único modelo de intervenção e depois reduzir passo a passo o modelo, como em regressão passo a passo para trás (*stepwise*).

Na última fase ao eliminar os possíveis resultados conflitantes das duas listas de outliers, diminui-se a chance de mascaramento de outliers espúrios. Balke verificou teoricamente, no caso de AR(1), que as estatísticas para detecção de outlier no novo procedimento têm resultados bastante favoráveis.

Sob a hipótese nula este teste e os outros de Escore e de Wald (apresentados nas próximas seções) são equivalentes (Buse, 1982).

#### **2.1.4 - O poder do teste e o valor crítico**

Como apontado por Chen e Liu (1993a), as estimativas dos parâmetros dos modelos podem ser afetadas pela escolha do valor crítico para detecção de outlier na estimação conjunta dos parâmetros do modelo e efeitos de outlier. Se o valor crítico é muito grande, o poder de detecção deve ser baixo; na situação extrema em que é usado um valor crítico muito grande não se detecta outlier. Quando o valor crítico é pequeno, o poder da detecção de outlier aumenta, no entanto, a chance de erro do Tipo I (detectar outliers espúrios) também aumenta. Como resultado, os estimadores dos parâmetros do modelo baseado no procedimento de estimação não são eficientes, devido ao super ajustamento causado pelos outliers espúrios.

Para o propósito de previsão, é interessante obter estimativas precisas dos parâmetros do modelo, e também ter alto poder de detecção durante o período de previsão. Para atingir esses dois objetivos, Chen e Liu (1993b) recomendaram um procedimento com dois passos: (1) para estimação do modelo, usa-se um valor crítico moderado no procedimento de estimação conjunta; e (2) para previsão, usa-se o modelo estimado em (1), mas usando um valor crítico menor para detecção de outlier durante a previsão.

Chen e Liu (1993b) estudaram o desempenho da estratégia anterior numa certa amplitude de valores críticos.

Para os modelos usados para estudo de simulação de Chen e Liu (1993b) (AR(1), MA(1) e IMA(1,1)) não encontraram maiores diferenças no poder de teste quando um outlier ocorre no meio da série ou próximo do fim da série, sem ser na última observação. O tamanho da série foi 100 com 500 replicações. No caso onde o outlier ocorre no fim da série houve uma notável redução do poder. Por exemplo, com um outlier de tamanho  $3\sigma$  e valor crítico  $C = 2,50$  o poder de detecção de outlier na última observação da série foi aproximadamente de 0,70 e para outlier no meio ou antes do final da série o poder foi em torno de 0,85, com exceção do caso de LS. O poder de detecção no final da série diminui até 0,40 quando o valor crítico cresce a 3,25.

### 2.1.5 - Identificação incorreta do tipo de outlier

Além do estudo do poder de detecção é também desejável obter informação referente à natureza do tipo de identificação incorreta dos outliers, por exemplo, se AO é mais fácil de ser identificado incorretamente como IO do que como LS, etc. O estudo da consequência da identificação incorreta do outlier na previsão será apresentado na seção 3.4.4, no caso onde o outlier está presente perto ou no final da série. Não será analisado o efeito do outlier presente no meio da série, pois estes (grandes) geralmente são detectados e identificados sem muitos problemas. Quando o outlier ocorre no meio da série o efeito do outlier na previsão ocorre somente através do efeito na estimativa dos parâmetros, e embora o efeito dependa do tipo de outlier e do modelo subjacente, em geral ele é pequeno (Chen e Liu, 1993b).

Para aproveitar as vantagens de ajuste de outlier na previsão é necessário identificar a localização e o tipo de outlier corretamente. Chen e Liu (1993b) estudaram este problema utilizando o procedimento por eles propostos e obtiveram em geral os seguintes resultados:

- (1) A probabilidade do tipo de identificação incorreta é maior quando o valor crítico é pequeno;
- (2) A probabilidade de identificação incorreta cresce quando o outlier ocorre perto do fim das séries;
- (3) AO tem menor chance de ser identificado incorretamente em relação aos outros tipos de outlier;
- (4) LS tem uma chance menor de ser identificado incorretamente se ele ocorre no meio da série, e tem uma probabilidade alta de ser identificado como IO ou TC se o outlier ocorre perto do fim da série;
- (5) O poder de detecção é uma função decrescente dos valores críticos e uma função crescente do tamanho do outlier.

Com esses resultados Chen e Liu (1993b) propõem restringir os tipos de outliers a serem considerados no teste de detecção quando realizado para observações perto do fim da série. No caso, eles sugerem sempre considerar o AO, pois o ganho da eficiência na previsão é alto caso ele seja detectado corretamente, e ao mesmo tempo é o que tem menor probabilidade de ser identificado incorretamente. Eles propõem também que o IO seja considerado no teste, devido à sua natureza adaptativa ao modelo subjacente. Outros tipos de outlier somente devem ser considerados caso haja alguma evidência forte para ocorrência de tais tipos de outliers.

## 2.2 Teste do Escore

O teste do Escore, também conhecido como teste de Multiplicador de Lagrange<sup>1</sup> (Silvey, 1959), é um dos três testes mais usados na literatura estatística, os outros são testes da RV e Wald (Hotta e Neves, 1992). Enquanto a aplicação do teste de RV requer estimações do modelo sob as hipóteses nula e alternativa, na aplicação dos outros testes é suficiente a estimação em apenas uma das hipóteses, a hipótese nula no teste do Escore e a alternativa no teste de Wald. No caso em estudo é claro que é mais fácil estimar o modelo sob a hipótese nula e assim é natural usar o teste do Escore, caso não haja uma perda significativa do poder.

O teste do Escore tem sido usado para testar ajuste em modelos de séries temporais (ver, por exemplo, Hosking, 1980). É usado como teste de especificação porque, em muitos casos, ele engloba diferentes hipóteses alternativas, no sentido de que elas levam ao mesmo teste estatístico.

Abraham e Yatawara (1988) propuseram o teste do Escore para testar simultaneamente a presença de um AO e/ou um IO num tempo  $k$ . O trabalho deles foi desenvolvido para um processo autoregressivo puro  $AR(p)$  mas, sob a hipótese de invertibilidade, pode-se transformar qualquer processo ARMA em um AR e aplicar esse teste.

Nesta seção serão apresentados os testes do Escore em modelos AR, apenas para outliers do tipo AO e IO, assim as hipóteses consideradas são:

$$H_0: \omega_I = \omega_A = 0;$$

$$H_{II}: \omega_I = 0;$$

$$H_{IA}: \omega_A = 0 \text{ e}$$

$$H_I: \omega_I \neq 0, \omega_A \neq 0.$$

Na primeira subseção é apresentado o teste de detecção quando o possível tipo de ocorrência é conhecido, mas o tipo do outlier é desconhecido (os outliers podem ocorrer

---

<sup>1</sup>A idéia básica do teste do Escore é apresentado no Apêndice 2.

simultaneamente). Na subseção seguinte são apresentados os testes quando o tipo do outlier é conhecido. Na subseção 2.2.3 é apresentado o caso em que a posição do outlier é desconhecida. Na seção 2.2.4 são apresentados os testes para o caso em que os parâmetros são estimados. Finalmente, na seção 2.2.5 é apresentado um critério para distinguir o tipo de outlier. Já mencionado na seção anterior, as estatísticas resultantes desse teste são equivalentes aos do teste de RV, na suposição de modelo conhecido, como poderá ser comprovado na derivação do teste a seguir.

### 2.2.1 Teste de detecção de outlier de tipo desconhecido

Considere um modelo AR e suponha que  $\phi_i$ s e  $\sigma^2$  são conhecidos a priori e um outlier é suspeito em  $t = k$ . O problema consiste em descobrir se as magnitudes são iguais a zero ou não, isto é, testar a hipótese  $H_0$  contra  $H_{1B}$ ,  $H_{1A}$  ou  $H_1$ .

Seja  $L$  a função de verossimilhança, assumindo que  $k > p$  e sob a hipótese  $H_0$ :  $\omega_I = \omega_A = 0$ , obtém-se

$$\frac{\partial \log L}{\partial \omega_A} = -\frac{1}{\sigma^2} \sum_{i=0}^p \phi_i \left[ x_{k+i} - \sum_{r=0}^p \phi_r x_{k+i-r} \right], \quad \phi_0 \equiv -1, \quad (2.2.1)$$

$$\frac{\partial \log L}{\partial \omega_I} = -\frac{1}{\sigma^2} \left[ x_k - \sum_{r=0}^p \phi_r x_{k-r} \right], \quad (2.2.2)$$

e a matriz de informação

$$\mathbf{I} = \frac{1}{\sigma^2} \begin{pmatrix} 1 + \sum_r^p \phi_r^2 & 1 \\ 1 & 1 \end{pmatrix}. \quad (2.2.3)$$

Assim, assumindo a notação  $\mathbf{D} = \{\partial \log L / \partial \omega_i\}_{2 \times 1}$ , a estatística Escore para testar  $H_0$  é dada por,

$$T_k = \mathbf{D}' \mathbf{I}^{-1} \mathbf{D} = \frac{a_k^2}{\sigma^2} + \frac{\left( \sum_i^p \phi_i a_{k+i} \right)^2}{\sigma^2 \sum_i^p \phi_i^2}, \quad (2.2.4)$$

que, sob  $H_0$  tem uma distribuição assintótica de  $\chi_{2,2}^2$ , qui-quadrado com 2 graus de liberdade (Silvey, 1959).

### 2.2.2 Testes de detecção de outlier quando o tipo é conhecido

Na seção anterior foi discutido um procedimento para testar a presença de outlier de tipo desconhecido. Nesta seção são apresentados testes específicos, onde o tipo de outlier é conhecido. Num processo AR(p) com  $\phi_1, \dots, \phi_p$  e  $\sigma^2$  conhecidos, com outlier do tipo AO no tempo  $t = k$ , pode se mostrar, analogamente à seção anterior, que a estatística Escore para testar  $H_{1A}$  é dada por:

$$T_{Ak} = \frac{\left( a_k - \sum_i^p \phi_i a_{k+i} \right)^2}{\sigma^2 \left( 1 + \sum_i^p \phi_i^2 \right)}. \quad (2.2.5)$$

Essa estatística tem uma distribuição assintótica  $\chi_1^2$ , sob  $H_{1A}$ . Similarmente, para outlier IO, a estatística do teste do Escore para a hipótese  $H_{1I}$ :  $\omega_1 = 0$  é dada por,

$$T_{Ik} = a_k^2 / \sigma^2, \quad (2.2.6)$$

que também tem uma distribuição assintótica  $\chi_1^2$ , sob  $H_{1I}$ . Nos dois casos, rejeita-se a hipótese nula se a estatística for maior que um valor crítico C, sendo o valor C determinado pelo nível de significância estipulado, i.e.,  $P(\chi_1^2 > C) = \alpha$ .

### 2.2.3 Teste para detecção do outlier quando a posição é desconhecida

Muitas vezes a posição k do outlier é desconhecida. Assim considera-se a estatística,

$$T = \max_{p < k < n-p} \{T_k\}, \quad (2.2.7)$$

onde  $T_k$  é dado por (2.2.4), para testar se existe algum outlier na série. As variáveis aleatórias  $T_{p+1}, \dots, T_{n-p-1}$  têm distribuição  $\chi_2^2$ , mas são dependentes, o que dificulta o cálculo da distribuição exata de T. Uma possível alternativa é trabalhar com as estatísticas de valores extremos.

A distribuição assintótica de máximos em processos estocásticos estacionários, pode ser obtida pelo seguinte resultado de Leadbetter (1983), para uma série de tamanho m.

**LEMA** - Suponha que  $\{X_j\}$  seja uma seqüência estacionária com função distribuição marginal  $F$ . Seja  $C$  a seqüência de constantes para os inteiros  $1 \leq t_1 < \dots < t_i < s_1 < \dots < s_j \leq m$ , com  $s_j - t_i \geq 1$ .

$$\left| F_{t_1, \dots, t_i, s_1, \dots, s_j}(C_m) - F_{t_1, \dots, t_i}(C_m) F_{s_1, \dots, s_j}(C_m) \right| \leq b_{m,l}, \quad (2.2.8)$$

onde para alguma seqüência  $l_m = O(m)$ ,

$$\lim_{m \rightarrow \infty} b_{m,l_m} = 0, \quad e$$

$$F_{t_1, \dots, t_i}(C_m) = P(X_{t_1} \leq C_m, \dots, X_{t_i} \leq C_m).$$

Suponha também que para cada  $\xi > 0$  possa ser escolhido um  $C_m = C_m(\xi)$  tal que

$$m[1 - F(C_m(\xi))] \rightarrow \xi \quad \text{quando } m \rightarrow \infty. \quad (2.2.9)$$

Então

$$P\left(\max_i \{x_i\} \leq C_m(\xi)\right) \rightarrow \exp(-v\xi) \quad \text{quando } m \rightarrow \infty, \quad (2.2.10)$$

onde  $v$  é uma constante, tal que  $0 < v < 1$ . □

Pela definição,  $\{T_k\}$  forma uma seqüência estacionária, na qual a distribuição marginal, é uma  $\chi_2^2$  ou uma exponencial de média 2. Dado que  $\{T_k\}$  é uma seqüência que depende de  $p$ , escolhendo  $l > p$  pode-se escolher  $b_{m,l}$  de forma que o mesmo se anule. Galambos (1978) mostrou que (2.2.10) é válida quando  $F$  é uma função distribuição acumulada de uma distribuição Gama. Assim, no caso em questão a aproximação também é válida, uma vez que  $\chi_2^2$  é um caso especial da distribuição Gama.

Sendo  $F(y)$  a função distribuição de  $\chi_2^2$  tem-se que

$$1 - F(y) = P(Y \geq y) = \exp(-y/2),$$

dado que a distribuição de  $\text{Exp}(2)$ , expressa por,

$$P(X \leq x) = F(x) = 1 - \exp(-x/2).$$

A fórmula (2.2.9), pelo resultado anterior, tendo  $m = n - 2p$  e  $n$  grande é dada por:

$$(n-2p) \exp[(-1/2)C_m(\xi)] = \xi$$

$$C_m(\xi) = 2[\ln(n-2p) - \ln(\xi)]. \quad (2.2.11)$$

Assim por (2.2.10), o valor crítico para um nível de significância  $\alpha$  deve ter

$$P(\max_k T_k \geq C_m(\xi)) = \alpha.$$

Tomando o limite em  $n$  dessa probabilidade:

$$\lim_{n \rightarrow \infty} P(\max_k \{T_k\} \leq C_n(\xi)) = \lim F(C_n(\xi)) = \exp(-v\xi) = 1 - \alpha \quad (2.2.12)$$

e, portanto,

$$\exp(-v\xi) = 1 - \alpha \quad \Leftrightarrow \quad \xi = -(\ln(1 - \alpha))/v. \quad (2.2.13)$$

Assim, dado um nível de significância  $\alpha$ , pela equação (3.12),  $C_n(\xi)$  é dado por,

$$C_n(\xi) = 2[\ln(n-2p) - \ln(\xi)]$$

$$C_n(\xi) = 2[\ln(n-2p) - \ln(-\ln(1-\alpha)) + \ln(v)]. \quad (2.2.14)$$

Essa fórmula para encontrar o valor de  $C_n(\xi)$  só é válida para o teste de hipótese  $H_0$  (caso em que o tipo de outlier é desconhecido). Generalizando agora o resultado anterior, para qualquer hipótese por (2.2.9) e (2.2.10) tem-se que:

$$m[1-F(C_n(\xi))] = -(\ln(1-\alpha))/v = \xi$$

$$C_n(\xi) = F^{-1}[1 + (\ln(1-\alpha))/mv]. \quad (2.2.15)$$

Assim pode-se encontrar o valor de  $C_n(\xi)$  para qualquer tipo de função Gama,  $\alpha$  (nível de significância),  $v$  e  $m$ . Em particular, para testes de hipóteses do tipo  $H_{II}$  ou  $H_{IA}$  (casos em que o tipo de outlier é conhecido) a distribuição de  $F$  é uma  $\chi^2_1$ . O exemplo seguinte ilustra esse caso.

**EXEMPLO:** Seja  $F$  a função acumulada de uma distribuição  $\chi^2_1$ ,  $\alpha = 0,10$ ,  $v = 0,8$ ,  $n = 100$  e o modelo utilizado um  $AR(p)$ , com  $p = 1$  ( $m = n-2p = 98$ ). Assim o ponto de corte para esse caso seria:

$$F(C_m(\xi)) = 1 + (\ln(1-0,10))/(98*0,8) = 0,9987$$

$$C_m(\xi) = F^{-1}(0,9987) = 10,3506. \quad \square$$

Leadbetter (1983) mostrou que as expressões para obtenção de  $v$  são complicadas, com os cálculos envolvendo um grande número de integrações. No caso de  $\chi_2^2$  Abraham e Yatawara (1988) verificaram empiricamente que  $v$  geralmente está entre 0,7 e 0,9 e no artigo sugerem a utilização de  $v = 0,8$ . Ota (1993) realizou algumas simulações para verificar o valor adequado de  $v$  para o caso em que a distribuição da estatística é  $\chi_1^2$  e concluiu também que  $v = 0,8$  seria o valor adequado para esse caso.

As estatísticas dos teste das hipóteses  $H_{1A}$  e  $H_{1B}$ , dadas na seção anterior pelas equações (2.2.5) e (2.2.6), respectivamente, têm distribuição assintótica  $\chi_1^2$ . Essa distribuição, ao contrário da distribuição  $\chi_2^2$ , que é uma exponencial de média 2, não tem uma fórmula simplificada para a distribuição acumulada.

Leadbetter (1983) e Galambos (1978) mostraram que a distribuição da estatística  $\max\{T_k\}$  converge assintoticamente para a distribuição  $\chi_2^2$ , quando o tipo de outlier é desconhecido, e para  $\chi_1^2$ , quando conhecido ( $T_A = \max\{T_{Ak}\}$  e  $T_I = \max\{T_{Ik}\}$ , para teste de detecção de outlier do tipo AO e IO, respectivamente).

#### 2.2.4 Teste com parâmetros de AR e $\sigma^2$ estimados pelos dados

Na prática os parâmetros de AR e  $\sigma^2$  não são conhecidos previamente e devem ser estimados pelos dados. Quando os parâmetros de um modelo auto-regressivo de ordem  $p$  (AR( $p$ )),  $\phi = (\phi_1, \phi_2, \dots, \phi_p)'$  e  $\sigma^2$  são desconhecidos é razoável considerar as estatísticas modificadas,  $\hat{T}_{Ak}$ ,  $\hat{T}_{Ik}$  e  $\hat{T}_k$  obtidas pela substituição de  $\phi$ ,  $\sigma^2$  e  $a_t$  pelos seus estimadores  $\hat{\phi}$ ,  $\hat{\sigma}^2$  e  $\hat{a}_t$  (neste último, o erro de previsão um passo à frente é dado por  $\hat{a}_t = x_t - \sum_j \phi_j x_{t-j}$ ) nas equações (2.2.5), (2.2.6) e (2.2.7).

Quando os estimadores são obtidos através do método de máxima verossimilhança, a distribuição assintótica dessas estatísticas sob  $H_0$  é  $\chi_2^2$ , quando o tipo de outlier é desconhecido, e  $\chi_1^2$ , quando conhecido.

Através de simulações, Abraham e Yatawara (1988) estudaram a distribuição de  $T$  para pequenas amostras utilizando o modelo AR(1) ( $\phi = 0,4$ ,  $\sigma = 1$ ) para tamanhos de amostra 50, 100 e 200 com 500 replicações. Os resultados encontrados foram razoáveis, principalmente para tamanhos de amostra maiores,  $n = 100$  e 200.

### 2.2.5 Distinção entre dois tipos de outliers

Na prática nem sempre se sabe o motivo que gerou o outlier, e dessa forma é difícil saber qual o tipo de outlier presente, quando ele é detectado por algum dos testes propostos nessa seção. Nessa subseção será descrita uma forma de identificar o tipo de outlier, caso o teste tenha sido significativo.

Um critério estatístico simples para distinguir entre os dois tipos de outlier, baseado em  $T_{1k}$  e  $T_{2k}$  é dado por

$$U = \frac{\left( a_k - \sum_i^p \phi_i a_{k+i} \right)^2}{1 + \sum_i^p \phi_i^2} - a_k^2. \quad (2.2.16)$$

Se  $U > 0$  o outlier é identificado como AO e caso contrário como IO. Quando  $\phi$ s e  $a_t$ s são estimados, obtém-se o teste estatístico  $\hat{U}$  que converge em probabilidade para  $U$ . Assim, pode-se usar o teste baseado em  $U$ ; esta estatística é a diferença de duas variáveis  $\chi^2$  dependentes, cada uma com um grau de liberdade.

Quando não existe nenhum outlier presente tem-se que  $P(U > 0) = 0,5$ . Mas quando a magnitude do outlier ( $\omega$ ) é diferente de zero,  $P(U > 0)$  depende da razão  $\omega/\sigma$  e dos parâmetros auto-regressivos, para ambos os tipos de outliers.

Abraham e Yatawara (1988) estudaram estatística  $U$  através de simulações, para um modelo AR(1) com o parâmetro  $\phi = 0,4$  e a variância do ruído branco  $\sigma^2 = 1$ , com outlier de tamanho  $4\sigma$  e  $5\sigma$  (500 replicações de séries de tamanho 100). Os resultados obtidos indicaram que a estatística é razoavelmente eficaz em detectar o tipo de outlier, com pelo menos 80% de outliers identificados corretamente.

Como todo processo estacionário ARMA(p,q) pode ser aproximado com um processo AR(p\*), onde  $p^* = p$  se  $q = 0$ ; as estatísticas desenvolvidas nessa seção podem ser generalizados para todo processo ARMA(p,q).

## 2.3 - Deslocamento de Verossimilhança - *Likelihood Displacement* (LD)

O deslocamento da verossimilhança (LD) de Cook (1986, 1987) é uma medida de impacto de uma perturbação do modelo. Uma perturbação comum usada em regressão é a eliminação de uma observação ou de uma equação. No contexto de séries temporais, uma perturbação natural seria a eliminação de uma observação ou um grupo de observações.

Atkinson e Shephard (1996) estudaram a influência da eliminação de uma observação na estimação da transformação Box e Cox no contexto de séries temporais.

São apresentados nessa seção três LDs que medem a influência de cada observação na estimação dos parâmetros (autoregressivos, de média móvel e variância de inovação) propostos por Ledolter (1990) no caso de AO, baseado na suposição de que o modelo ARMA estacionário é conhecido. O primeiro considera todo o vetor de parâmetros, o segundo, os coeficientes autoregressivos e de média móvel, e a última apenas a variância das inovações. Pelos estudos de Bruce e Martins (1989) e Ledolter (1990) a estimativa desta variância das inovações é mais sensível a outliers do que a estimativa dos coeficientes de autoregressão e de média móvel. Por esse motivo o enfoque desta seção será sobre o LD (subseção 2.3.1) considerando apenas a variância da inovação, no qual será descrito o teste de hipótese, este apresentado na subseção 2.3.2.

### 2.3.1 - Algumas medidas de influência

O LD é um diagnóstico útil para detectar a influência de uma perturbação do modelo. Denotando os parâmetros do modelo  $M$  por  $\eta$ , a função log-verossimilhança por  $L(\eta)$  e o EMV de  $\eta$  dada por  $\hat{\eta}$ . Considerando agora um modelo com perturbação  $M(\omega)$ , e denotando a sua função de log-verossimilhança por  $L_{\omega}(\eta)$  e o estimador de máxima verossimilhança associado por  $\hat{\eta}_{\omega}$ . Então o deslocamento de verossimilhança

$$LD_{\omega}(\eta) = 2\{L(\hat{\eta}) - L(\hat{\eta}_{\omega})\}, \quad (2.3.1)$$

expressa a influência da perturbação nas estimativas dos parâmetros medindo a mudança contida na função de log-verossimilhança.

Caso o interesse seja apenas num subconjunto  $\eta_1$  de  $\eta = (\eta_1, \eta_2)$  então o deslocamento de verossimilhança apropriado é dado por

$$LD_{\omega}(\eta_1) = 2[L(\hat{\eta}) - L\{\hat{\eta}_{1\omega}, g(\hat{\eta}_{1\omega})\}], \quad (2.3.2)$$

onde a função  $g(\eta_1)$  maximiza  $L(\eta_1, \eta_2)$  para um dado  $\eta_1$ , e  $\hat{\eta}_{1\omega}$  é obtido pela partição  $\hat{\eta}_\omega = (\hat{\eta}_{1\omega}, \hat{\eta}_{2\omega})$ . Para maiores detalhes e motivações adicionais ver Cook (1986, 1987).

Uma perturbação natural num modelo de regressão  $y_t = \mathbf{x}_t' \boldsymbol{\beta} + a_t$  ( $1 \leq t \leq n$ ) com os erros independentes  $N(0, \sigma^2)$  é a eliminação de uma equação, ou de um grupo de equações. Se apenas o coeficiente de regressão  $\boldsymbol{\beta}$  é de interesse, então o deslocamento de verossimilhança resultante é uma função monótona da estatística  $D$  de Cook (Cook e Weisberg, 1982, p.185, segundo Ledolter, 1990).

No contexto de séries temporais a eliminação da  $i$ -ésima observação  $y_i$  (ARMA estacionário e invertível) define uma perturbação apropriada do modelo (Bruce e Martin, 1989). Para o cálculo de deslocamento de verossimilhança é necessário estimar os coeficientes do modelo,  $\boldsymbol{\beta} = (\boldsymbol{\phi}, \boldsymbol{\theta})$  e a variância da inovação sob o modelo com perturbação. Isso pode ser obtido tratando a  $i$ -ésima observação como um parâmetro desconhecido  $\xi_{(i)}$ , obtendo seu estimador  $\hat{\xi}_{(i)}$  e os estimadores  $\hat{\boldsymbol{\beta}}_{(i)}$  e  $\hat{\sigma}_{(i)}^2$  pelo conjunto de dados em que a observação  $y_i$  é substituída por este  $\hat{\xi}_{(i)}$ . Para um dado coeficiente  $\boldsymbol{\beta} = (\boldsymbol{\phi}, \boldsymbol{\theta})$ , o estimador da observação eliminada é dado por

$$\hat{\xi}_{(i)}(\boldsymbol{\beta}) = \left\{ 1 - \frac{\pi(\mathbf{B})\pi(\mathbf{B}^{-1})}{\sum_{k \geq 0} \pi_k^2} \right\} z_i = \sum_{j \geq 0} \rho_j^* (y_{i+j} + y_{i-j}) \quad (2.3.3)$$

onde  $\pi(\mathbf{B}) = \phi(\mathbf{B})/\theta(\mathbf{B}) = 1 - \pi_1 \mathbf{B} - \pi_2 \mathbf{B}^2 - \dots$  e  $\rho_j^*$  é a autocorrelação inversa de defasagem  $j$  do processo (Brubacher e Wilson, 1976). Esse estimador é uma média ponderada das observações adjacentes pelas autocovariâncias inversas. Dado que o estimador em (2.3.3) é uma função dos coeficientes do modelo, deve ser utilizado um procedimento iterativo para determinar, para cada observação  $i$  eliminada, o estimador de máxima verossimilhança  $\hat{\xi}_{(i)}$ ,  $\hat{\boldsymbol{\beta}}_{(i)}$  e  $\hat{\sigma}_{(i)}^2 = \sum a_t^2/n$ . Os resíduos  $\hat{a}_t$  utilizam  $\hat{\boldsymbol{\beta}}_{(i)}$  e são obtidos após substituir a  $i$ -ésima observação por  $\hat{\xi}_{(i)}$ .

São apresentados três deslocamentos de verossimilhança. O primeiro mede o deslocamento quando ambos os coeficientes da série temporal  $\boldsymbol{\beta}$  e a variância  $\sigma^2$  são de interesse:

$$\begin{aligned} LD_1(\boldsymbol{\beta}, \sigma^2) &= 2\{L(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) - L(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}_{(i)}^2)\} \\ &= -n \log \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} + n \left( \frac{s_{(i)}^2}{\hat{\sigma}_{(i)}^2} - 1 \right) \end{aligned} \quad (2.3.4)$$

A última expressão é obtida após substituir os estimadores na função de log-verossimilhança condicional (i.e., condicional aos valores iniciais)  $L(\beta, \sigma^2) = \text{const.} - (n/2) \log \sigma^2 - (1/2\sigma^2) \sum a_i^2$ . O estimador da variância  $s_{(i)}^2$  é calculado com  $\hat{\beta}_{(i)}$ , mas utiliza todas as observações, incluindo  $y_i$ , para obter os resíduos. Este estimador é diferente de  $\hat{\sigma}_{(i)}^2$ , que usa o mesmo estimador dos coeficientes do modelo em  $\hat{\sigma}_{(i)}^2$ , mas substitui a  $i$ -ésima observação por  $\hat{\xi}_{(i)}$ .

A segunda medida de diagnóstico do deslocamento surge quando apenas o coeficiente do modelo,  $\beta$ , é de interesse:

$$LD_i(\beta) = 2\{L(\hat{\beta}, \hat{\sigma}^2) - L(\hat{\beta}_{(i)}, s_{(i)}^2)\} = -n \log \left( \frac{\hat{\sigma}^2}{s_{(i)}^2} \right). \quad (2.3.5)$$

A terceira medida de diagnóstico mede o deslocamento quando apenas  $\sigma^2$  é de interesse:

$$\begin{aligned} LD_i(\beta, \sigma^2) &= 2\{L(\hat{\beta}, \hat{\sigma}^2) - L(\hat{\beta}, \hat{\sigma}_{(i)}^2)\} \\ &= -n \log \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} + n \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} - 1 \right). \end{aligned} \quad (2.3.6)$$

### Exemplo - Processo AR(1)

As três medidas serão ilustradas para um processo de primeira ordem; suas autocorrelações inversas são dados por  $\rho_1^* = -\phi/(1+\phi^2)$  e  $\rho_j^* = 0$  para  $j \geq 2$ . A  $i$ -ésima observação é substituída por

$$\hat{\xi}_{(i)}(\phi) = \frac{\phi}{1+\phi^2} (y_{i+1} + y_{i-1}) \quad (2.3.7)$$

que é uma soma ponderada das duas observações adjacentes. O estimador  $\hat{\phi}_{(i)}$  é obtido maximizando

$$\hat{\sigma}^2(\phi) = \frac{1}{n} \left[ \sum_{t \neq i, i+1} (y_t - \phi y_{t-1})^2 + \{y_{i+1} - \phi \hat{\xi}_{(i)}(\phi)\}^2 + \{\hat{\xi}_{(i)}(\phi) - \phi y_{i-1}\}^2 \right], \quad (2.3.8)$$

e o estimador de  $\sigma^2$  é dado por  $\hat{\sigma}_{(i)}^2 = \sigma^2(\hat{\phi}_{(i)})$ . Devem ser usados procedimentos iterativos. Os deslocamentos de verossimilhança são obtidos substituindo os estimadores  $\hat{\phi}_{(i)}$  e  $\hat{\sigma}_{(i)}^2$  junto com  $s_{(i)}^2 = \sum (y_t - \hat{\phi}_{(i)} y_{t-1})^2 / n$  e o estimador de máxima verossimilhança

$\hat{\phi} = \Sigma y_t y_{t-1} / \Sigma y_{t-1}^2$  e  $\hat{\sigma}^2 = \Sigma (y_t - \hat{\phi} y_{t-1})^2 / n$  obtidos pelo conjunto de dados completo nas equações (2.3.4) a (2.3.6).  $\square$

Existem alguns métodos de estimação alternativos quando as observações são tratadas como dados faltantes.

Na equação (2.3.3) a equação foi substituída por uma combinação linear das observações adjacentes, onde os pesos eram dados pela autocorrelação inversa e foi preciso usar um procedimento iterativo para obter o estimador  $\hat{\beta}_{(i)}$ . Existem problemas de truncamento como por exemplo em modelos com componentes de médias móveis no corte de autocorrelação inversa em (2.3.3). Esses problemas podem ser evitados estimando os parâmetros pelo modelo de intervenção, de Box e Tiao (1975),  $\phi(B)(y_t - \omega_i \delta_{t,i}) = \theta(B)a_t$  (Ledolter, 1990).

O segundo procedimento para lidar com modelos ARMA com observações faltantes é escrever o modelo na forma de espaço de estado e usar o filtro de Kalman. Peña (1987) e Bruce e Martin (1989) (segundo Ledolter, 1990) mostraram que este procedimento leva essencialmente aos mesmos estimadores obtidos com o procedimento da análise de intervenção.

O terceiro procedimento é utilizar o algoritmo EM (Dempster *et alii*, 1977) que Abraham e Chuang (1993) adaptaram para o contexto de séries temporais, considerando IO e AO. Neste artigo de Abraham e Chuang concentrou-se em trabalhar em processos AR. Os resultados obtidos, através de simulações, foram razoáveis, mas a carga computacional torna-se pesada para  $p$  grande e a maximização para o modelo ARMA é mais complicada.

Os LDs apresentam uma relação entre eles. O deslocamento de verossimilhança conjunto, quando ambos  $\beta$  e  $\sigma^2$  são de interesse é um limite superior da soma dos deslocamentos de verossimilhança individuais, i.e.,

$$LD_i(\beta) + LD_i(\sigma^2) \leq LD_i(\beta, \sigma^2). \quad (2.3.9)$$

Pelas expressões (2.3.4) a (2.3.6) tem-se o seguinte resultado:

$$\log \left( \frac{s_{(i)}^2 / \hat{\sigma}_{(i)}^2}{\hat{\sigma}^2 / \hat{\sigma}_{(i)}^2} \right) \leq \frac{s_{(i)}^2}{\hat{\sigma}_{(i)}^2} - \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2}.$$

dado que  $\hat{\sigma}_{(i)}^2 \leq \hat{\sigma}^2 \leq s_{(i)}^2$ , e assim  $1 \leq a = \hat{\sigma}^2 / \hat{\sigma}_{(i)}^2 \leq s_{(i)}^2 / \hat{\sigma}_{(i)}^2 = b$ , esta desigualdade é verdadeira pois  $\log(b/a) \leq b - a$ , para  $1 \leq a \leq b$ .

Foi observado (Bruce e Martin, 1989; Ledolter, 1989) que o estimador da variância da inovação é mais sensível aos outliers que os estimadores dos coeficientes de ARMA. Assim é razoável considerar deslocamento de verossimilhança em relação a  $\sigma^2$  como possível diagnóstico de outlier. O deslocamento em (2.3.6) também pode ser escrito como

$$\begin{aligned} LD_i(\sigma^2) &= -n \log \left( 1 + \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} - 1 \right) + n \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} - 1 \right) \\ &= -n \left\{ \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} - 1 \right) - \frac{1}{2} \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} - 1 \right)^2 \right\} + n \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} - 1 \right) \\ &= \frac{n}{2} \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} - 1 \right)^2 = DV_i \end{aligned} \tag{2.3.10}$$

que é o diagnóstico para variância da inovação considerado por Bruce e Martin (1989).

### 2.3.2 O teste de LD

As equações (2.4.6) e (2.4.10) mostram que

$$W_{(i)} = n \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} - 1 \right) \tag{2.3.11}$$

é uma componente-chave de deslocamento de verossimilhança para a variância da inovação. Ela compara dois estimadores da variância; um é calculado por todas as observações disponíveis, enquanto o outro substitui a  $i$ -ésima observação por um parâmetro adicional. Agora será investigada a sua distribuição amostral quando a observação  $y_t$  segue um processo ARMA com coeficientes conhecidos e quando os outliers estão presentes. Ledolter (1990) verificou que

$$n\hat{\sigma}_{(i)}^2 = n\hat{\sigma}^2 - \left( \sum_{k=0} \pi_k^2 \right)^{-1} \left\{ \pi(B^{-1})a_i \right\}^2, \tag{2.3.12}$$

onde  $n\hat{\sigma}^2 = \sum a_t^2$  e  $a_t = \pi(B)y_t$ . Assim pode-se escrever (2.3.11) como razão de formas quadráticas:

$$W_{(i)} = n \frac{\mathbf{a}' \mathbf{G} \mathbf{a}}{\mathbf{a}' (\mathbf{I} - \mathbf{G}) \mathbf{a}}, \tag{2.3.13}$$

onde  $\mathbf{a}'\mathbf{G}\mathbf{a} = \left( \sum_{k \geq 0} \pi_k^2 \right)^{-1} \left\{ \boldsymbol{\pi}(\mathbf{B}^{-1})\mathbf{a}_i \right\}^2$ . Dado que  $\mathbf{G}$  é uma matriz idempotente de posto 1 segue, pelos resultados de forma quadrática, que  $\{(n-1)/n\}W_{(i)}$  tem uma distribuição  $F(1, n-1)$ . Se  $n$  é grande,  $W_{(i)}$  é  $\chi^2$  com 1 grau de liberdade.

Na prática os coeficientes devem ser estimados e a posição do outlier não é conhecida. Ledolter (1990) realizou algumas simulações com 5000 replicações num processo AR(1) com 100 observações. Os percentis 90, 95 e 99 para  $\max W_{(k)}$ , para  $\phi = 0,5$  são respectivamente 11,60; 13,01 e 16,04. Esses valores não dependem muito do valor de  $\phi$ . Para  $\phi = -0,9$  eles são iguais a 11,49; 13,02 e 16,93. Baseado nesses resultados ele sugeriu o ponto de corte em torno de 13,0 para  $\alpha = 0,05$ . Usando esse ponto de corte ele simulou 1000 replicações com diferentes valores de AO adicionado na 50ª observação. Para um AO de tamanho  $5\sigma_a$  os máximos de  $W_{(k)}$  foram maiores que o valor de corte em 967 replicações e em 966 casos eles ocorreram na posição correta, mostrando um bom poder. Para AO de tamanho  $3\sigma_a$  os números foram 446 e 426, respectivamente.

## **Capítulo 3**

### **Efeitos da presença do outlier em Séries Temporais**

Para mostrar a importância em detectar os outliers e/ou desenvolver métodos robustos são apresentados os efeitos que eles podem causar em diferentes aspectos, tais como, na modelagem, na estimação dos parâmetros e na previsão.

Os efeitos que cada tipo de outlier causam sobre as observações em séries temporais foram apresentados na seção 1.3. Por isso neste capítulo são apresentados outros tipos de efeitos, tais como os efeitos nos resíduos (seção 3.1), na estimação dos parâmetros supondo conhecida a identificação do modelo ARIMA (seção 3.2), na identificação do modelo (seção 3.3) e na previsão do mesmo (seção 3.4). No Capítulo 4 são apresentados os estudos dos efeitos de outliers na previsão de dados agregados.

#### **3.1 - Efeitos nos resíduos**

A presença de um outlier nem sempre é determinada apenas por uma análise nos dados originais, pois seu efeito pode estar misturado com a própria dinâmica estocástica do processo gerador dos dados. Desta forma, os resíduos do modelo ARIMA ajustado, além de conter informações à respeito do efeito do outlier, podem ser mais úteis na determinação da sua existência. Na subseção 3.1.1 são apresentados os estudos no caso dos parâmetros e modelo conhecidos e na subseção 3.1.2 é discutido o caso onde os parâmetros são desconhecidos.

### 3.1.1 Efeitos nos resíduos considerando os parâmetros e o modelo conhecidos

Considere o modelo com apenas um outlier no tempo  $k$ , apresentado na seção 2.1:

$$y_t = \frac{\theta(B)}{\phi(B)\alpha(B)} a_t + f(t), \quad (3.1.1a)$$

onde  $f(t)$  é uma função do distúrbio definida na equação (1.3.2). Caso a presença do outlier não seja considerada a série observada seria modelada como:

$$y_t = \frac{\theta(B)}{\phi(B)\alpha(B)} e_t. \quad (3.1.1b)$$

Considerando-se os outliers apresentados na seção 2.1 tem-se que:

$$e_t = \begin{cases} \omega_I \delta_{t,k} + a_t & \text{para IO;} \\ \omega_A \pi(B) \delta_{t,k} + a_t & \text{para AO;} \\ \frac{\omega_L}{1-B} \pi(B) \delta_{t,k} + a_t & \text{para LS;} \\ \frac{\omega_T}{1-\lambda B} \pi(B) \delta_{t,k} + a_t & \text{para TC;} \end{cases} \quad (3.1.2)$$

onde  $\delta_{t,k}$  é o indicador do outlier no tempo  $k$ .

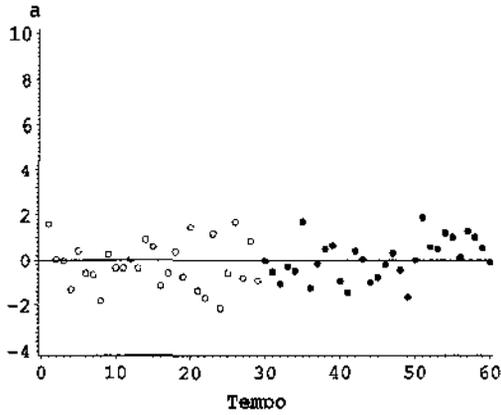
Pode-se ter uma visão mais clara da natureza das séries anteriores se forem apresentadas como

$$e_t = \omega_{TP} w_{TP,t} + a_t, \quad (3.1.3)$$

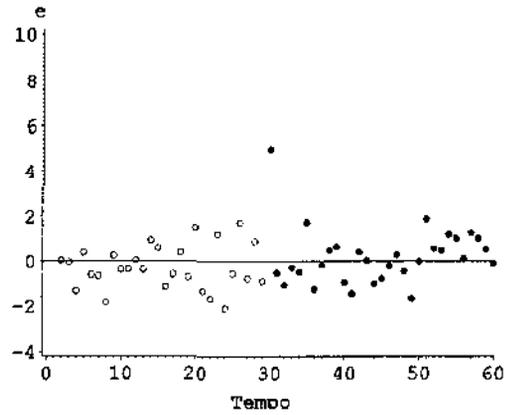
onde  $TP = I, A, L, T$  e  $w_{TP,t}$  assume valor zero para  $t < k$ ; e valor 1 para  $t = k$ ; e

$$w_{TP,k+m} = \begin{cases} 0 & TP = IO; \\ -\pi_m & TP = AO; \\ 1 - \sum_{j=1}^{m-1} \pi_j & TP = LS; \\ \lambda^m - \sum_{j=1}^{m-1} \lambda^{m-j} \pi_j - \pi_m & TP = TC, \end{cases} \quad (3.1.4)$$

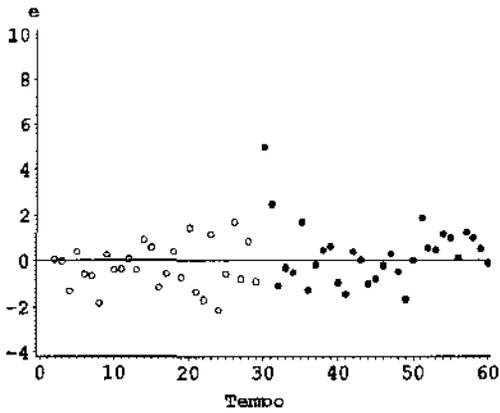
para  $k + m$  ( $m = 1, 2, \dots, n - k$ ). Sendo  $\omega_{TP}$  o tamanho do outlier do tipo  $TP$  que ocorre no tempo  $k$ ,  $w_{TP,t}$  é o efeito de um outlier de tamanho unitário do tipo  $TP$ , que ocorre na observação  $k$ , no resíduo da  $t$ -ésima observação.



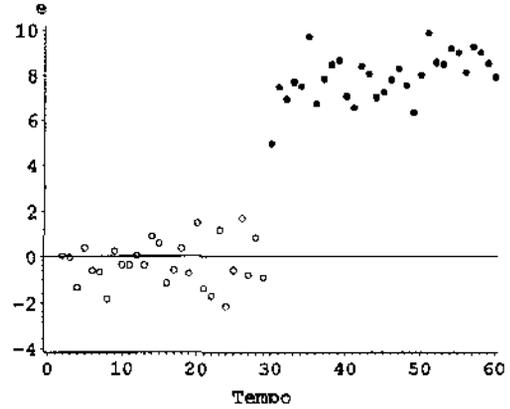
a) Inovação



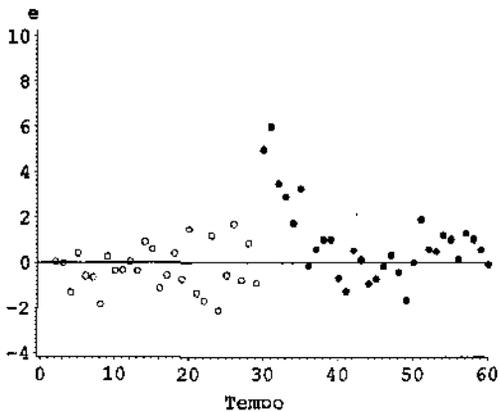
b) Resíduos da série com IO



c) Resíduos da série com AO



d) Resíduos da série com LS



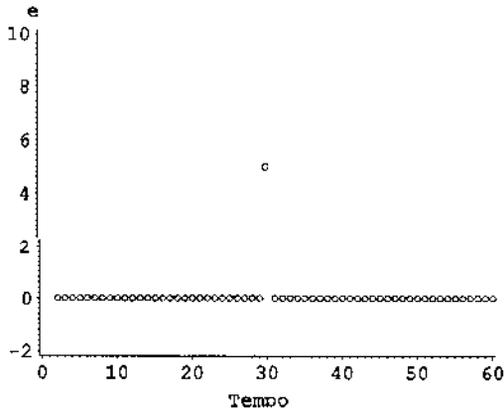
e) Resíduos da série com TC ( $\lambda = 0,7$ )

**Legenda:**

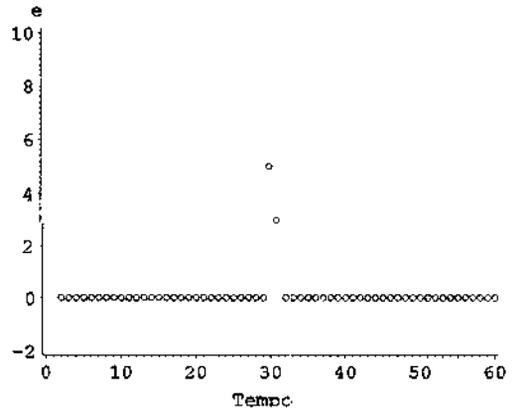
○ - observações não afetadas pelo outlier

● - observações afetadas em pelo menos uma das séries.

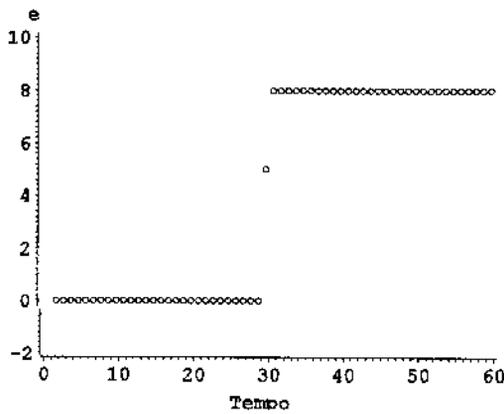
**Gráfico 3.1.1** - Resíduos de modelo AR(1), com  $\phi = -0,6$ , apresentados no Gráfico 1.3.1



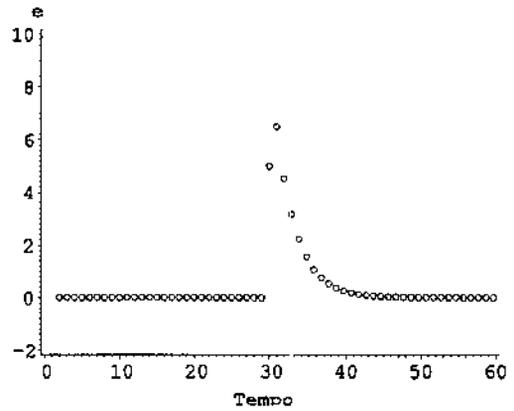
a) Resíduo de uma série com IO, com as inovações removidas



b) Resíduo de uma série com AO, com as inovações removidas



c) Resíduo de uma série com LS, com as inovações removidas



d) Resíduo de uma série com TC, com as inovações removidas

**Gráfico 3.1.2** - Resíduos com as inovações removidas.

### i) Efeito de IO

As equações (3.1.3) e (3.1.4) fornecem as informações sobre o outlier presente na série. Todas as informações começam no tempo  $t = k$ . No caso de IO, toda a informação está contida no tempo  $t = k$ . Para ilustrar isso, foi utilizada a mesma série apresentada no Capítulo 1 (Gráfico 1.3.1), gerada através de simulação de uma série de tamanho 60 de um processo AR(1) com  $\phi = -0,6$  contendo um outlier do tipo IO de tamanho  $5\sigma$  na posição  $t = 30$ . A série observada, está apresentada no Gráfico 1.3.1 (b) e a inovação no Gráfico 3.1.1 (a).

Pelo Gráfico 3.1.1 (b) pode-se ver um impacto no tempo  $t = 30$  dado que esse valor se mostra bem afastado do resto. Os valores restantes são as inovações simuladas. Neste caso, sendo uma série simulada é possível subtrair a inovação de  $e_t$ . A série resultante é apresentada no Gráfico 3.1.2 (a). Agora pode-se ver claramente que o único valor afetado pelo outlier é do tempo  $t = 30$ . Verifique que essa série é o valor esperado do efeito do outlier nos resíduos  $e_t$ .

### ii) Efeito de AO

Diferentemente de um IO, um AO afeta a série numa única observação. É natural suspeitar que a série residual também seja afetada em apenas um tempo. Pelas equações (3.1.3) e (3.1.4) e pela discussão anterior de IO, verifica-se que este não é o caso. A informação de um AO em uma série  $e_t$  começa no período em que ocorre o outlier, e depois decai de acordo com os pesos  $\pi_t$ s do modelo ARIMA. Este fenômeno é ilustrado no Gráfico 3.1.1 (c), que apresenta a série residual de uma série AR(1) contendo um AO. A série observada é apresentada no Gráfico 1.3.1 (b).

Como no caso anterior, pode-se ver um impacto em  $e_t$  no tempo  $t = 30$ . Há também um valor menor no tempo seguinte. A série não aparenta ter sido alterada nesses dois tempos, talvez apenas no primeiro. Para observar melhor os resíduos afetados, a inovação ( $a_t$ ) foi subtraída de  $e_t$ . A série resultante é apresentada no Gráfico 3.1.2 (b). Note que os dois únicos valores diferentes de zero são  $t = 30$  e  $31$ . Pode-se explicar o motivo de apenas esses valores terem sido afetados retornando à equação (3.1.3) e à definição associada a  $w_{it}$ . Os valores de  $w_{it}$  são diferentes de zero apenas para  $t = k$  (neste caso 30) e todos os índices depois de  $t = k$  para os quais valores dos pesos  $\pi_t$ s do modelo ARIMA subjacente são diferentes de zero. Para uma série AR(1), apenas  $\pi_1$  é diferente de zero. Assim, para esse modelo  $e_t$  contém uma informação extra a respeito de AO nos tempos  $t = 30$  e  $31$ . Num processo com o componente MA todos os  $\pi_t$ s são diferentes de zero, embora tenda a zero

quando  $i \rightarrow \infty$ . Quando o processo não é estacionário nem sempre  $|\pi_i| < 1$ ,  $i = 1, 2, \dots$ , assim o impacto de um AO no resíduo é realçado. Por exemplo, num processo ARIMA(1,1,0) com  $\phi > 0$ ,  $\pi_1 = 1 + \phi > 1$  e  $\pi_2 = -\phi$ .

### iii) Efeitos de LS e TC

Como no caso de um AO, a informação de LS ou TC está contida em  $e_t$  e está relacionada aos pesos  $\pi_i$ s do modelo ARIMA subjacente. Pela equação (3.1.3) e (3.1.4) pode-se ver que no caso de LS essa informação é relacionada à soma acumulada dos  $\pi_i$ s. Para TC essa informação é relacionada à soma ponderada acumulada desses mesmos  $\pi_i$ s, com os pesos (da ponderação) reduzidos de acordo com um fator de amortecimento  $\lambda$ .

A informação fornecida será ilustrada pela série simulada de AR(1) com um LS no tempo  $t = 30$ . No caso de série AR(1), pode-se observar que a informação começa no tempo  $t = 30$ . Do tempo  $t = 31$  em diante, essa informação se torna constante dado que  $\pi_k = 0$  para  $k \geq 2$ . Esse comportamento é ilustrado no Gráfico 3.1.2 (c) da série  $e_t$  com o ruído subjacente removido. Desta forma, a informação adicional é similar à combinação de um impacto grande no início, seguido por uma função escada. Pode-se também observar essa informação no gráfico de  $e_t$  com a inovação removida apresentado no Gráfico 3.1.2 (d). Devido ao valor de  $\pi_1 (= \phi)$  ser negativo o efeito de um LS, pela equação (3.1.4) se torna bem evidente.

Diferentemente de LS, o TC não causa deslocamento permanente no nível da série observada, nem na série residual. A informação de  $e_t$  é amortecida após a ocorrência do outlier e diminui com o tempo, como pode ser confirmado no Gráfico 3.1.2 (d).

### 3.1.2 Efeito no ruído quando os parâmetros são estimados

Na seção anterior foram vistos os efeitos causados pelos outliers nos resíduos com as inovações removidas quando os parâmetros e a identificação do modelo eram conhecidos. Assim, o comportamento desses resíduos com as inovações removidas só depende do tipo do outlier. Note-se que analisando essa série a identificação do outlier é quase direta, com exceção de IO e TC que dependendo dos parâmetros e do modelo podem não ser facilmente distinguíveis, ou até mesmo podem ser indistinguíveis, no caso de AR(1) onde  $\phi = \lambda$ , onde  $\lambda$  é o parâmetro de amortecimento de TC.

No entanto, essa suposição não é realística, pois geralmente não se conhece o tipo, a posição e o tamanho do outlier, e a presença do outlier na série pode afetar tanto a

estimação dos parâmetros quanto a identificação do modelo subjacente. Desta forma, os efeitos causados nos ruídos podem ser completamente diferentes dos que foram apresentados na subseção 3.1.1.

Na prática supõe-se que a identificação do modelo seja correta e usam-se os ruídos desse ajuste para verificar a presença dos outliers, como foi visto no Capítulo 2. Dada a diversidade de situações em que o outlier pode afetar no ajuste, afetando o comportamento do ruído não são apresentados nesse trabalho.

## 3.2 - Efeito na Estimativa dos Parâmetros do Modelo ARIMA

Nesta seção serão estudados os efeitos nas estimações dos parâmetros, supondo apenas a identificação do modelo como conhecida, iniciando o estudo com AO.

A importância deste estudo não é apenas analisar como os outliers podem afetar na estimativa dos parâmetros, mas também ver como os vícios na estimação dos parâmetros causados pelos outliers podem afetar as eficiências da detecção dos mesmos. Alguns outliers podem não ser identificados devido ao problema de mascaramento.

### 3.2.1 Efeito de AO na estimação dos parâmetros

Supondo o modelo conhecido, muitos autores (Guttman e Tiao, 1978; Denby e Martin, 1979; Martin e Yohai, 1986; Ledolter, 1987 e 1989; e outros) verificaram que a presença de AOs na série afetam a estimativa dos coeficientes do modelo, quando estes são estimados sem considerar a presença dos outliers. Para ilustrar como os outliers afetam as estimativas dos parâmetros os efeitos são apresentados na forma de exemplos, dos modelos AR(1) e ARIMA(0,1,1).

#### Exemplo 1: Processo AR(1)

O estimador de quadrados mínimos de  $\phi$  em  $x_t = \phi x_{t-1} + a_t$ ,  $t = 1, \dots, n$  é dado por

$$\hat{\phi} = \frac{\sum_{t=2}^n x_t x_{t-1}}{\sum_{t=2}^n x_t^2}. \quad (3.2.1)$$

Um AO na posição  $k$  implica que  $y = x + f$ , onde  $y = (y_1, \dots, y_n)'$ , segue uma distribuição normal multivariada, com vetor de média  $f$  e matriz de covariância

$$\Gamma = \{\gamma_{i-j} = \sigma^2 \phi^{|i-j|} / (1 - \phi^2); i \geq 1, j \leq n\},$$

onde  $\mathbf{f}$  é uma coluna de zeros, exceto na posição  $k$  onde é igual a  $\omega_A$ .

Dado que

$$E(\sum_{t=2}^n y_t y_{t-1}) = (n-1)\gamma_1,$$

$$E(\sum_{t=2}^n y_t^2) = (n-1)\gamma_0 + \omega_A^2,$$

tem-se uma aproximação grosseira de primeira ordem, que não é tão ruim quando  $n$  é razoavelmente grande (Ledolter, 1987), dado por:

$$E(\hat{\phi}) \cong \frac{(n-1)\gamma_1}{(n-1)\gamma_0 + \omega_A^2} \cong \phi - \frac{\phi(1-\phi^2)\omega_A^2}{(n-1)\sigma^2}. \quad (3.2.2)$$

Assim, o vício do estimador de quadrados mínimos é dado por

$$E(\hat{\phi}) - \phi = \frac{\phi(1-\phi^2)\omega_A^2}{(n-1)\sigma^2}.$$

Por exemplo, um outlier de  $3\sigma$  num processo AR(1) de  $n = 100$  observações com  $\phi = 0,5$  leva a um vício de  $-0,034$ ; para um outlier de  $5\sigma$  o vício é de  $-0,095$  (Ledolter, 1989).  $\square$

### Exemplo 2 - Processo ARIMA(0,1,1)

Para estudar o efeito do outlier na estimação do parâmetro média-móvel, primeiramente foi calculada a autocorrelação de defasagem 1 ( $r_1$ ) da primeira diferença da série original ( $z_t = y_t - y_{t-1}$ ,  $t = 2, \dots, n$ ). Essas diferenças seguem um processo de média móvel com vetor de média  $\mathbf{f}$  e estrutura de covariância  $\gamma_0 = \sigma^2(1+\theta^2)$ ,  $\gamma_1 = -\theta\sigma^2$ ,  $\gamma_j = 0$  para  $j > 1$ . Os elementos de  $\mathbf{f}$  são zero, com exceção da posição  $k$  que é igual a  $\omega_A$  e na posição  $k+1$  de  $-\omega_A$ . Assumindo que  $k < n-1$ , i.e., o outlier não ocorre nas duas últimas observações, então:

$$E(\sum_{t=2}^n z_t z_{t-1}) = (n-1)\gamma_1 - \omega_A,$$

$$E(\sum_{t=2}^n z_t^2) = (n-1)\gamma_0 + 2\omega_A^2 \text{ e}$$

$$E(r_1) \cong \frac{(n-1)\gamma_1 - \omega_A}{(n-1)\gamma_0 + 2\omega_A^2} = \frac{-\theta/(1+\theta^2) - (\omega_A/\sigma)^2 / [(n-1)(1+\theta^2)]}{1 + 2(\omega_A/\sigma)^2 / [(n-1)(1+\theta^2)]}. \quad (3.2.3)$$

A relação do parâmetro  $\theta$  de média móvel e de autocorrelação de uma defasagem  $\rho_1$  de primeira diferença de um ARIMA(0,1,1) é da seguinte forma:

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

Pode se obter um estimador pelo método dos momentos substituindo-se  $\rho_1$  na equação anterior por  $r_1$  e resolvendo a equação quadrática resultante para a solução invertível  $\hat{\theta}$ . Se  $r_1$  é menor (maior) que  $-0,5$  ( $0,5$ ),  $\hat{\theta}$  é considerado como  $-1$  ( $1$ ). A equação (3.2.3) mostra como o outlier afeta a esperança de  $r_1$ . Desta forma, uma aproximação simples de  $E(\hat{\theta}) = \theta^*$  pode ser obtida resolvendo a equação quadrática  $E(r_1) = -\theta^*/(1+(\theta^*)^2)$ . A solução invertível (aquela solução que está entre  $-1$  e  $1$ ) de

$$\theta^* = -\frac{1}{2E(r_1)} \pm \sqrt{\frac{1}{4(E(r_1))^2} - 1},$$

permite quantificar o efeito de um outlier na estimativa de média-móvel no processo ARIMA(0,1,1). Por exemplo, um outlier de tamanho  $5\sigma$  numa série de 100 observações num processo ARIMA(0,1,1) com  $\theta = 0,5$  resulta em  $E(r_1) = 0,429$  e  $E(\hat{\theta}) = 0,566$ . Estas aproximações foram verificadas através de simulações por Ledolter (1983) e Hotta (1993).  $\square$

O vício nas estimações dos parâmetros causado pelo(s) outlier(s) pode afetar a eficiência dos testes de detecção do(s) outlier(s). Alguns outliers podem não ser identificados devido ao problema de mascaramento. Chen e Liu (1993a) propuseram um procedimento iterativo de detecção e estimação para amenizar esses problemas. Pelos seus estudos com alguns tipos de modelos e por considerar os quatro tipos de outliers no modelo ARIMA, o procedimento proposto mostrou-se bastante eficiente na detecção de outlier e estimação dos parâmetros, utilizando os valores críticos adequados.

Nesses casos assintoticamente os vícios tendem a zero. Suponha agora um modelo mais geral onde os outliers ocorrem segundo uma distribuição de Bernoulli e os tamanhos dos outliers tenham uma distribuição normal, i.e.,

$$y_t = [\theta(B)/\phi(B)]a_t + \eta_t; \quad (3.2.4)$$

onde  $\eta_t$  é uma variável aleatória igual a zero com probabilidade  $1 - \gamma$  e segue uma distribuição  $N(0, \sigma_\eta^2)$  com probabilidade  $\gamma$ . Assuma ainda que  $\eta_t$  e  $a_t$  são independentes (ver por exemplo, Denby e Martin, 1979).

No caso de AR(1), Ledolter (1987) estudou o vício assintótico do estimador de  $\hat{\phi} = r_1(y)$  neste caso e verificou que é dado por:

$$E(\hat{\phi}) - \phi = -\phi \left[ 1 + \frac{\sigma_a^2}{\gamma \sigma_\eta^2 (1 - \phi^2)} \right]^{-1}.$$

Por exemplo, seja  $\phi = 0,5$ ,  $\gamma = 0,02$  e  $\sigma_\eta/\sigma_a = 4$ . Então o vício chega a  $-0,097$ .

### 3.2.2 Efeito de IO, TC e LS na estimativa dos parâmetros

Embora no modelo com IO,

$$\phi(B)y_t = \theta(B)(a_t + \omega_1 \delta_{t,k}) = \theta(B)e_t,$$

a distribuição de  $e_t$  tenha uma cauda mais pesada que de  $a_t$ , Whitte (1962) e Box e Jenkins (1976) mostraram que a distribuição assintótica dos estimadores de quadrados mínimos de  $\beta$ , dos parâmetros autoregressivos e médias-móveis, não depende da distribuição dos distúrbios. Denotando o estimador por  $\hat{\beta}$  eles mostraram que a distribuição de  $\sqrt{n}(\hat{\beta} - \beta)$  converge à distribuição normal com matriz de covariância que só depende dos parâmetros autoregressivos e médias-móveis. Isso ilustra que, assintoticamente, IO não afeta as estimativas de mínimos quadrados, não leva a vício, e nem afeta a variância da distribuição assintótica, no modelo de outlier com número fixo. No entanto, deve ser destacado que, para pequenas amostras, as estimativas de quadrados mínimos ordinários podem ser ineficientes para distribuição com caudas pesadas (Ledolter, 1987). Martin e Yohai (1985) discutem a situação em que podem ser encontrados estimadores de mínimos quadrados com eficiência bem baixa e estimativas com variância assintótica muito pequena. Chen e Liu (1993b) estudaram através de simulações os efeitos causados por TC e LS no meio da série. Um TC presente no meio da série se comporta de modo muito parecido aos outros outliers (AO e IO) em termos de poder de detecção de outlier, enquanto o LS geralmente tem baixo poder de detecção, pois os parâmetros do modelo podem ser estimados com vício devido à mudança permanente no nível da série o que causa o efeito de mascaramento.

### 3.3 - Identificação do Modelo

Nesta seção são estudados os efeitos de outlier nas estimativas da função de autocorrelação (fac) e variância dos resíduos. Geralmente o modelo é identificado analisando-se o fac. São consideradas significantes as autocorrelações que estão acima de um certo valor, proporcional à estimativa do desvio padrão das inovações.

Guttman e Tiao (1978), Chang (1982) e Ledolter (1987) mostraram o efeito do outlier na estimação da autocorrelação, enquanto Chang (1982) estudou também o efeito na estimação da função de autocorrelação parcial (facp).

Deutsch *et alii* (1990) mostraram através de resultados algébricos e simulações que a presença de apenas um AO ou um IO prejudicam a identificação do modelo ARMA, quando realizada através de estimativas da fac e facp não ajustadas. A magnitude, a locação, o tipo de outlier, e em alguns casos os valores dos parâmetros ARMA influenciam na identificação da série resultante.

Enquanto todos esses autores estudaram o efeito para apenas um outlier, Ledolter (1987) e outros estudaram a situação onde o número de outliers é proporcional ao tamanho da série (ver equação (3.2.4)).

#### 3.3.1 - Considerações gerais do vício da fac

A magnitude e o padrão do vício causado pelo outlier na estimativa da fac influenciam a identificação do modelo. O estimador usual da fac para  $\rho_i = E(x_t x_{t+i}) / \sigma_x^2$  de um processo estacionário  $x_t$  é dado por

$$\hat{\rho}_i = r_i = \frac{\sum_{t=1}^{n-i} x_t x_{t+i}}{(n-i)\hat{\sigma}_x^2}, \quad (3.3.1)$$

onde  $\hat{\sigma}_x^2 = \hat{\text{Var}}(x_t) = \sum_{t=1}^n x_t^2 / n$ . Quando existe um outlier na série  $y_t$ , o estimador viciado usual é dado por,

$$r_i^0 = \frac{\sum_{t=1}^{n-i} y_t y_{t+i}}{(n-i)\hat{\sigma}_y^2}, \quad (3.3.2)$$

onde  $\hat{\sigma}_y^2 = \hat{\text{Var}}(y_t) = \sum_{t=1}^n y_t^2 / n$ . As equações (3.3.1) e (3.3.2) não são iguais quando  $y_t$  é definido como uma série com outlier. A diferença  $\hat{\sigma}_y^2 - \hat{\sigma}_x^2$  mede o aumento na variância, que é o resultado de não se ter filtrado os efeitos dos outliers em  $y_t$ .

Uma relação aproximada entre  $r_i^0$  e  $r_i$  é obtido substituindo  $y_t = x_t + f(t)$ , na equação (3.3.2), após a eliminação do termo do numerador de expectância zero, i.e.,  $\sum_{t=1}^{n-i} x_t f(t+i)$  e  $\sum_{t=1}^{n-i} x_{t+i} f(t)$ ,

$$r_i^0 \cong r_i \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} + \frac{S_f(i)}{n-i} \hat{\sigma}_y^2, \quad (3.3.3)$$

onde  $S_f(i) = \sum_{t=1}^{n-i} f(t)f(t+i)$  e  $f(t)$  representa o efeito do outlier na observação (ver, equação (1.3.2)). Observe que a aproximação utilizada implica  $\hat{\sigma}_y^2 = \hat{\sigma}_x^2 + S_f(0) / n$ .

Quando não existem outliers as equações (3.3.1) e (3.3.2) implicam que  $r_0$  e  $r_0^0$  são iguais a 1, pois neste caso  $S_f(0) = 0$  e  $\hat{\sigma}_y^2 = \hat{\sigma}_x^2$ , e  $r_i^0 = r_i$ ,  $i \geq 0$ .

Suponha, por exemplo, que um AO ocorre na série; então  $S_f(0) = \omega_A^2$ . Quanto maior é o tamanho de AO menor é a semelhança entre  $r_i^0$  e  $r_i$ . Da equação (3.3.3) tem-se que

$$r_i^0 \cong \frac{r_i \hat{\sigma}_x^2}{\hat{\sigma}_x^2 + \omega_A^2 / n}. \quad (3.3.4)$$

Valores altos de  $\omega_A^2$  fazem  $r_i^0$  descrever um processo de ruído branco, pois tornam pequenos os valores de  $\hat{\sigma}_x^2 / (\hat{\sigma}_x^2 + \omega_A^2 / n)$ , mascarando o valor de  $r_i$ .

Num contexto em que o número de outliers é proporcional ao tamanho da série Ledolter (1987) observou que o estimador da autocorrelação de primeira ordem num modelo MA(1) é dado por:

$$\rho_1(y) = -\frac{\theta}{1+\theta^2} [1 - \gamma(\sigma_* / \sigma_x)^2]^{-1}.$$

Por exemplo, para  $\gamma = 0,05$ ,  $\sigma_* / \sigma_x = 3$  e  $\theta = 0,8$  tem-se que  $\rho_1(y) = -0,336$  e pela solução invertível de  $-0,336 = -\theta_*/(1+\theta_*)$  é dado por  $\theta_* = 0,39$  e não por  $\theta = 0,8$ .

Utilizando a mesma aproximação de Deutsch *et alii* (1990), quando ocorre um TC,

$$S_f(i) = \omega_T^2 \sum_{t=k}^{n-i} \lambda^{2(t-k)+i},$$

pode ser relativamente grande. Considerando, por exemplo, uma série de 100 observações com um outlier de tamanho  $\omega_T = 3\sigma_a$  na posição  $t = 50$  com constante de amortecimento  $\lambda = 0,8$  a relação das variâncias das séries sem e com outlier é dada por

$$\hat{\sigma}_y^2 = \hat{\sigma}_x^2 + S_f(0) / n = \hat{\sigma}_x^2 + 0,25(1 - \phi^2) \sigma_x^2,$$

aumentando a diferença para  $0,69(1-\phi^2)\sigma_x^2$  quando ocorre um outlier de tamanho  $5\sigma_a$ .

No caso de um LS na série tem-se que  $S_f(i) = (n-i-k+1)\omega_L^2$ . Essa função  $S_f(i)$  pode ser muito grande, dependendo do valor de  $\omega_L$  e de  $k$ , mostrando que a presença de um LS pode afetar bastante a identificação do modelo. Por exemplo, considerando uma série de tamanho 100 com outlier de magnitude  $3\sigma$  na posição 50, a diferença das variâncias das séries sem e com outlier é de  $4,49(1-\phi^2)\sigma_x^2$  e essa diferença aumenta para  $12,75(1-\phi^2)\sigma_x^2$  quando ocorre um outlier de  $5\sigma$ .

Em geral, a relação entre  $r_i^0$  e  $r_i$  determina se  $r_i^0$  facilita ou impede a identificação correta do modelo ARMA. No entanto, o efeito de IO na equação (3.3.3) não é óbvio como nos outros casos, dado que  $S_f(i)$  depende do tamanho, local, e tipo de outlier, e dos parâmetros do processo ARMA. De um modo geral, pela equação (3.3.4)  $r_i^0$  parece menor que  $r_i$  quando  $S_f(0)/n$  é grande comparado com  $\sigma_x^2$ . Alternativamente,  $r_i^0$  exibe um comportamento correto do processo  $x_t$  se a razão das variâncias  $\hat{\sigma}_x^2 / \hat{\sigma}_y^2$  é próxima de um.

As propriedades da *fac* amostral determinam a sensibilidade da técnica de determinação empírica de ARMA à existência de outlier na série temporal. Em todo o casos, a falha para eliminar um outlier presente nas observações, inovações ou correlação no ruído significa que a estimativa da *fac* é viciada. Assim, segundo Deutsch *et alii* (1990) os componentes de ARMA podem ser identificados erroneamente, com processo AR identificado como modelo MA, processo AR identificado como modelo ARMA, e em geral  $\hat{p} \neq p$  e  $\hat{q} \neq q$ . Além disso, a técnica de identificação-estimação supõe que o modelo ARMA é conhecido, e essa suposição geralmente não é apropriada quando o modelo é identificado empiricamente. O componente AR( $p$ ) em geral deve ser identificado corretamente quando a intervenção ocorre longe do fim da série. O componente MA( $q$ ) em geral não deve ser identificado corretamente quando a magnitude da aberração é grande (AO ou IO).

Sem nenhum conhecimento a priori do tipo de outlier, um outlier grande pode modificar completamente a estimativa da *fac*. Conhecendo-se o seu tipo e sua locação na série é possível diminuir a influência na estimativa da *fac* e conseqüentemente na identificação do modelo. Por exemplo, o modelo e o outlier podem ser estimados conjuntamente pelo procedimento de Chen e Liu (1993a).

### 3.3.2 - Efeito do outlier na estimação de $\sigma^2$

Mesmo que os outliers, em alguns casos, não afetem muito as estimativas dos parâmetros e/ou as previsões dos modelos ARIMA, eles sempre afetam a estimativa de  $\sigma^2$ , a

variância da inovação  $\{a_t\}$ . Como conseqüência eles afetam a largura dos intervalos de predição, dado que esses intervalos são proporcionais à estimativa do desvio padrão da inovação.

Suponha que a série esteja contaminada por um AO na posição  $k$ , os resíduos do modelo ajustado (que é o erro de predição 1 passo à frente) pode ser escrito como

$$y_{k+i} - \hat{y}_{k+i-1}(1) = \hat{a}_{k+i} + \begin{cases} -\omega_A \hat{\pi}_i & i = 1, \dots, n-k \\ \omega_A & i = 0 \\ 0 & i = 1-k, \dots, -1 \end{cases}$$

onde

$$\hat{a}_t = \frac{\hat{\phi}(B)\theta(B)}{\phi(B)\hat{\theta}(B)} a_t = g(B)a_t.$$

Isso implica que o erro quadrático médio esperado (valor esperado do estimador da variância de inovação) é dado por

$$\begin{aligned} \text{EQM} &= \frac{1}{n} \sum_{i=1-k}^{n-k} [y_{k+i} - \hat{y}_{k+i-1}(1)]^2 \\ &\cong \frac{1}{n} \{n\hat{\sigma}_a^2 + \omega_A^2 [1 + \hat{\pi}_1^2 + \dots + \hat{\pi}_{n-k}^2]\} \\ &= \hat{\sigma}_a^2 + \frac{\omega_A^2}{n} [1 + \hat{\pi}_1^2 + \dots + \hat{\pi}_{n-k}^2] \end{aligned}$$

Para um processo AR(1)  $\hat{\pi}_1 = \hat{\phi}$ ,  $\hat{\pi}_j = 0$  para  $j > 1$ ,

$$\hat{a}_t = \frac{1 - \hat{\phi}B}{1 - \phi B} a_t \quad e \quad \hat{\sigma}_a^2 = \frac{\sigma^2 (1 + \hat{\phi}^2 - 2\phi\hat{\phi})}{1 - \phi^2}.$$

No caso onde o outlier não ocorre na última observação

$$E(\text{EQM}) = \sigma^2 \left[ \frac{1 + \hat{\phi}^2 - 2\phi\hat{\phi}}{1 - \phi^2} + \frac{(\omega_A/\sigma)^2}{n} (1 + \hat{\phi}^2) \right].$$

Ledolter (1989) verificou, através de simulações, que no modelo AR(1) um AO de magnitude  $3\sigma$ , na posição  $k = 50$ , para séries de tamanho  $n = 100$ , com  $nr = 10.000$  replicações, o erro quadrático médio aumenta entre 8 e 15%. Considerando agora os modelos AR(1) e ARIMA(0,1,1) com AO de tamanho  $5\sigma$  (com 10.000 e 1.000 replicações, respectivamente) o aumento para ambos os casos foi de 25 a 45%.

### 3.3.3 Alguns procedimentos para identificação do modelo

O primeiro procedimento iterativo desenvolvido para estimar o modelo e os outliers conjuntamente foi de Chang e Tiao (1983). Esse procedimento, apresentado no Capítulo 2 com algumas modificações, foi generalizado por muitos outros autores.

Tsay (1986a) propôs um procedimento iterativo para identificar os outliers (AO ou IO), remover seus efeitos, e tentar identificar o modelo ARIMA no processo subjacente, baseado essencialmente no procedimento de estimação iterativa de Chang e Tiao (1983) e na função de autocorrelação estendida, um método de identificação do modelo de Tsay e Tiao (1984). Posteriormente Chen e Liu (1993a) apresentaram um procedimento baseado também no de Chang e Tiao (1983) procurando amenizar o problema de mascaramento e de efeitos espúrios.

Para aliviar o efeito de outlier(s) na modelagem em séries temporais, Martin (1981) considerou um procedimento robusto (aos valores aberrantes) para estimação dos parâmetros do modelo na presença daqueles. Nesse procedimento, são atribuídos pesos reduzidos para as observações aberrantes no processo de estimação. São citadas na literatura duas desvantagens deste procedimento:

- (1) todos os tipos de outliers são tratados da mesma forma. Isso explica parcialmente por que alguns métodos robustos têm bom desempenho em alguns casos e em outros não, como foi relatado no estudo de simulação de Chuang e Abraham (1989). Por esse motivo o procedimento não é poderoso em identificar IO (Chang e Tiao, 1983).
- (2) nenhuma informação pode ser obtida a respeito da magnitude e/ou natureza do outlier. Isso pode causar uma perda de precisão na previsão, uma vez que o impacto dos outliers não são incorporados nela (Chen e Liu, 1993b).

Smith e West (1983) propuseram um procedimento de monitoramento contínuo para detectar a presença de mudanças abruptas de nível ou de inclinação e AO dentro de uma formulação de um modelo de espaço de estado (mudança de nível, mudança na inclinação e outlier). Para aplicar esse procedimento, é necessário a priori especificar as variâncias e as probabilidades para cada estado. Inferências sobre os outliers nas mudanças estruturais são baseadas em probabilidades a posteriori de vários estados estimados pelo procedimento bayesiano usual. O procedimento de Smith e West é apropriado quando baseado em experiência com o potencial de mudanças estruturais conhecido.

Uma vez que métodos robustos e métodos de detecção não são mutuamente exclusivos, pode-se dizer que em termos de ajuste do modelo, um método robusto “ótimo” pode produzir o mesmo resultado que a aplicação do método de diagnóstico “ótimo”, onde o termo “ótimo” é usado no sentido de menor perda (Hotta e Neves, 1992).

### **3.4 - EFEITOS NA PREVISÃO**

Nas previsões em séries temporais geralmente é dada pouca atenção às observações discrepantes. Os primeiros estudos foram de Hillmer (1984) e Ledolter (1987, 1989). Baseado nesses e em outros trabalhos mais recentes serão estudados nesta seção os efeitos de outliers na previsão através de modelos ARIMA.

Hillmer (1984) estudou o monitoramento e ajuste de previsões na presença de outliers naqueles modelos, verificando que IO não afeta as previsões quando os parâmetros são conhecidos, seu principal interesse concentrou-se no efeito do AO. Ledolter (1989) considerou o efeito de AO na previsão através do mesmo modelo. Ambos concluíram que o intervalo de predição é muito sensível ao AO, como foi visto na seção anterior, no entanto, eles verificaram que a previsão pontual não é muito afetada, a não ser que o outlier ocorra próximo da origem de previsão, i.e., nas últimas observações.

Chen e Liu (1993b) estudaram o efeito de outlier na previsão em séries temporais, quando o outlier ocorre perto ou na origem de previsão e verificaram que a previsão é bastante afetada quando o outlier ocorre perto, e principalmente na origem de previsão. Neste último caso não se pode identificar o tipo do outlier porque o efeito dos quatro tipos de outlier é o mesmo na observação onde ele ocorre. A diferenciação dos efeitos só ocorre nas observações ainda não observadas.

Na prática os procedimentos utilizados para escolha dos modelos para previsão incluem testes de detecção de outliers e modelagem na presença deles. No Capítulo 4 são estudados os efeitos conjuntos na previsão, analisando-se os casos de previsão através das séries desagregada e agregada.

Antes de começar a discutir os efeitos que o outlier causa é apresentada uma breve revisão de previsão com modelos ARIMA na seção 3.4.1 e posteriormente os estudos a respeito dos erros de previsão na seção 3.4.2. Nas duas últimas seções são estudados com mais detalhe os erros de previsão quando o outlier ocorre no meio e no final da série separadamente. É também discutido o efeito nos erros de previsão quando um outlier é

identificado incorretamente e quando um outlier inexistente é “detectado” (outlier espúrio, *swamping*).

### 3.4.1 - Previsão através do modelo ARIMA

As previsões através do modelo ARIMA podem ser escritas como uma combinação linear de observações passadas. Os pesos desta função linear de previsão depende dos parâmetros autoregressivos e de médias móveis e da ordem de diferenças.

A previsão de erros quadráticos médios mínimos (EQMM) de uma futura observação  $x_{n+h}$  realizada pela origem  $n$  é dada pela esperança condicional:

$$x_n(h) = E(x_{n+h} / x_n, x_{n-1}, x_{n-2}, \dots);$$

veja por exemplo, Box e Jenkins (1976, p.128), onde ela pode ser escrita como uma combinação linear dos dados passados  $x_n, x_{n-1}, x_{n-2}, \dots$ ,

$$x_n(h) = \sum_{j \geq 1} \pi_j^{(h)} x_{n+1-j}. \quad (3.4.1)$$

Os pesos de previsão  $\pi_j^{(h)}$  (para  $j = 1, 2, \dots$ ) são calculados por

$$\pi_j^{(h)} = \pi_{j+h-1} + \sum_{i=1}^{h-1} \pi_i \pi_j^{(h-i)} \quad (3.4.2)$$

para  $h > 1$  e  $\pi_j^{(1)} = \pi_j$  são os coeficientes do polinômio

$$\pi(B) = 1 - \sum_{j \geq 1} \pi_j B^j = \phi(B)\alpha(B)/\theta(B).$$

Ver, por exemplo, Box e Jenkins (1976, p.142). Para processos invertíveis esses pesos se aproximam a zero e o seu decaimento para grandes defasagens é exponencial com velocidade igual ao inverso da maior raiz do polinômio média-móvel. Por exemplo, os pesos  $\pi_j$ s para modelo ARIMA(0,1,1) são dados por  $\pi_j^{(h)} = \pi_j = (1-\theta)\theta^{j-1}$ , i.e., previsão é uma média das observações passadas ponderadas exponencialmente e é equivalente a previsão pelo AES (Alisamento Exponencial Simples) com constante de alisamento  $\alpha = 1 - \theta$ . Em geral, para séries suficientemente longas, as previsões de modelos não estacionários ( $d \neq 0$ ) são médias ponderadas de observações passadas, onde a soma dos pesos de previsão é 1.

A previsão  $x_n(h)$  também pode ser escrita como

$$x_n(h) = \psi_h a_n + \psi_{h+1} a_{n-1} + \dots, \quad (3.4.3)$$

onde  $\psi_i$ s são os coeficientes do polinômio  $\psi(B)=\pi(B)^{-1}$ . Esta forma de escrever a previsão é útil para calcular o erro quadrático médio (EQM) da previsão.

Da próxima seção em diante o termo previsão de EQMM será utilizado apenas como previsão.

### 3.4.2 Erro de previsão na presença de um outlier

Seja  $\{x_t, 1 \leq t \leq n\}$  um processo ARIMA(p,d,q) e  $y_t = x_t + f(t)\delta_{t,k}$  a série observada com algum outlier com efeito  $f(t)$  que ocorre no tempo  $t = k$ . Sob a hipótese de que os coeficientes de ARIMA e o efeito de outlier são conhecidos, o erro de previsão  $h$  passos à frente é dado por

$$e_n(h) = y_{n+h} - y_n(h) = x_{n+h} - x_n(h). \quad (3.4.4)$$

Geralmente, as informações concernentes à presença ou tipo do outlier são desconhecidas. Um método para detectar o outlier e obter as estimativas dos parâmetros do modelo seria aplicar o procedimento de Chen e Liu (1993a). A série ajustada sem os efeitos do outlier pode ser utilizada como uma série não contaminada e suas previsões podem ser calculadas baseadas na fórmula recursiva associada ao modelo ARIMA, como na equação (3.4.1). Neste caso, a previsão  $h$  passos à frente pode ser desenvolvida baseada no outlier detectado empiricamente como

$$\hat{y}_n(h) = \hat{x}_n(h) + \hat{\omega}\hat{L}(B)\delta_{n+h,\hat{t}_1}, \quad (3.4.5)$$

onde  $\hat{x}_n(h)$  é a previsão  $h$  passos à frente baseada nas observações ajustadas e  $\hat{\omega}$ ,  $\hat{L}(B)$  e  $\hat{t}_1$  são estimadores de efeito, tipo e localização do outlier, respectivamente. Em outras palavras, previsões na presença de outliers podem ser obtidas somando a previsão da série descontaminada (série com efeito do outlier removido) e a extrapolação do efeito do outlier. É importante enfatizar que  $\hat{x}_n(h)$  representa a previsão da série descontaminada. Em geral, esta série é obtida através do ajuste do outlier baseado na localização, tipo e tamanho do outlier detectado. Assim, o erro de previsão nesse caso é

$$\begin{aligned} \hat{e}_n(h) &= y_{n+h} - \hat{y}_n(h) \\ &= x_{n+h} - \hat{x}_n(h) + (\omega L(B)\delta_{n+h,t_1} - \hat{\omega}\hat{L}(B)\delta_{n+h,\hat{t}_1}). \end{aligned} \quad (3.4.6)$$

Como discutido em Chen e Liu (1993a), geralmente é possível detectar corretamente a posição e o tipo de um outlier (grande) que ocorre no meio da série, no entanto, um outlier que ocorre nas últimas observações da série é difícil de ser detectado corretamente

devido à falta de informação. Em particular, quando um outlier ocorre na última observação da série, os testes estatísticos usados para determinar o seu tipo são idênticos para todos os quatro outliers e sendo impossível identificá-lo através deste teste. Nessa situação, a decisão do tipo do outlier presente é do pesquisador. Em outras situações, um outlier pode ocorrer na penúltima ou antepenúltima observação da série. Nessas situações, um procedimento de detecção de outlier pode não ser capaz de identificar o tipo do outlier corretamente; as consequências da identificação incorreta são estudadas na subseção 3.4.4.

### 3.4.3 Efeito do outlier na previsão quando o outlier ocorre no meio da série

Quando o outlier ocorre  $r = n - k$  períodos antes da origem de previsão, a observação  $y_t$  pode ser expressa como

$$y_{n-i} = \begin{cases} x_{n-i} & i > r \\ x_{n-i} + \omega L(B)\delta_{n-i,n-r} & i \leq r \end{cases} \quad (3.4.7)$$

Os resíduos obtidos filtrando a série  $y_t$  supondo o modelo conhecido podem ser expressos como:

$$\hat{e}_{n-i} = \begin{cases} a_{n-i} & i > r \\ a_{n-i} + \omega \pi(B)L(B)\delta_{n-i,n-r} & 0 \leq i \leq r \end{cases} \quad (3.4.8)$$

Uma identificação errada de um outlier pode resultar de uma detecção correta da locação mas com a especificação do tipo incorreto, ou de uma identificação do tipo correta mas a identificação da locação errada, ou de ambos, tipo e locação incorreta. Nesse estudo, será considerado apenas o caso em que a locação do outlier é corretamente identificada e o tipo identificado incorretamente.

Quando o outlier é detectado à  $r$  períodos da origem de previsão as estimativas para diferentes tipos de outliers são dados por:

$$\hat{\omega}_I = \hat{e}_{n-r}, \quad (3.4.9a)$$

$$\hat{\omega}_A = \hat{\rho}_A^2 (\hat{e}_{n-r} - \sum_{i=1}^r \pi_i \hat{e}_{n-r+i}), \quad (3.4.9b)$$

$$\hat{\omega}_L = \hat{\rho}_L^2 (\hat{e}_{n-r} + (1 - \pi_1)\hat{e}_{n-r+1} + \dots + (1 - \pi_1 - \dots - \pi_r)\hat{e}_n), \quad (3.4.9c)$$

$$\hat{\omega}_T = \hat{\rho}_T^2 (\hat{e}_{n-r} + (\lambda - \pi_1)\hat{e}_{n-r+1} + \dots + (\lambda^r - \lambda^{r-1}\pi_1 - \dots - \pi_r)\hat{e}_n), \quad (3.4.9d)$$

onde

$$\hat{\rho}_{A,k}^2 = (1 + \pi_1^2 + \dots + \pi_r^2)^{-1},$$

$$\hat{\rho}_{L,k}^2 = (1 + (1 - \pi_1)^2 + \dots + (1 - \pi_1 - \dots - \pi_r)^2)^{-1},$$

$$\hat{\rho}_{T,k}^2 = (1 + (\lambda - \pi_1)^2 + \dots + (\lambda^r - \lambda^{r-1}\pi_1 - \dots - \pi_r)^2)^{-1}.$$

Usando (3.4.5), as observações ajustadas de acordo com o tipo TP podem ser expressas como

$$\hat{x}_{n-i} = \begin{cases} y_{n-i} & i > r \\ y_{n-i} + \hat{\omega}_{TP} \hat{L}(B) \delta_{n-i,n-r} & i \leq r \end{cases} \quad (3.4.10)$$

$$= \begin{cases} x_{n-i} & i > r \\ x_{n-i} + (\omega L(B) \delta_{n-i,n-r} - \hat{\omega}_{TP} \hat{L}(B) \delta_{n-i,n-r}) & i \leq r \end{cases}$$

Usando o resultado anterior e a equação (3.4.1), a previsão  $h$  passos à frente da série ajustada do tipo do outlier TP pode ser expressa como

$$\hat{x}_n(h)_{TP} = x_n(h) + \sum_{i=0}^r \pi_{i+1}^{(h)} [\omega L(B) \delta_{n-i,n-r} - \hat{\omega}_{TP} \hat{L}(B) \delta_{n-i,n-r}], \quad (3.4.11)$$

onde  $x_n(h)$  é a previsão  $h$  passos à frente do processo  $\{x_i\}$ . Como na equação (3.4.5), pode-se expressar  $\hat{y}_n(h)_{TP}$  como

$$\hat{y}_n(h)_{TP} = \hat{x}_n(h)_{TP} + \hat{\omega}_{TP} \hat{L}(B) \delta_{n+h,n-r}. \quad (3.4.12)$$

Assim o erro da previsão definido na equação (3.4.6) pode ser reescrito como

$$\begin{aligned} \hat{e}_n(h) &= y_{n+h} - \hat{y}_n(h)_{TP} \\ &= \{x_{n+h} - x_n(h)\} - \left\{ \sum_{i=0}^r \pi_{i+1}^{(h)} (\omega L(B) \delta_{n-i,n-r} - \hat{\omega}_{TP} \hat{L}(B) \delta_{n-i,n-r}) \right\} \\ &\quad + \{\omega L(B) \delta_{n-i,n-r} - \hat{\omega}_{TP} \hat{L}(B) \delta_{n-i,n-r}\}. \end{aligned} \quad (3.4.13)$$

Os três componentes (cada um entre chaves) na expressão anterior representam respectivamente (1) incerteza intrínseca dos erros de previsão associada ao processo subjacente, (2) erros de amostragem no ajuste dos efeitos de outlier nas observações e (3) o erro associado à extrapolação do efeito de outlier. Substituindo  $\hat{e}_{n-i}$  da equação (3.4.8) nas equações (3.4.9) e (3.4.13), pode-se desenvolver o EQM da previsão com função de outlier detectado e do outlier verdadeiro apresentados na Tabela 3.4.2.

### 3.4.4 Efeito do outlier na previsão quando o outlier ocorre próximo do final da série

Chen e Liu (1993b) estudaram o efeito de outlier na previsão em séries temporais usando um procedimento de detecção e ajuste proposto por esses mesmos autores num outro trabalho (1993a). O estudo focalizou o comportamento da previsão quando o outlier ocorre perto ou na origem de previsão, considerando os outliers AO, IO, LS e TC. Verificaram que a previsão é bastante afetada quando o outlier ocorre perto da origem de previsão, principalmente quando ocorre na origem, sendo que nesse caso não se pode identificar o tipo do outlier devido à falta de informação. Demonstraram, através de simulações, que o poder de detecção de outlier se reduz se o outlier ocorre na origem de previsão, i.e., na última observação da série.

Chen e Liu (1993b) investigaram também o potencial da perda de precisão da previsão devido à identificação errada do tipo de outlier. Genericamente, previsão com ajuste de outlier parece ter um melhor desempenho do que a previsão sem ajuste de outlier mesmo que o tipo do outlier não tenha sido identificado corretamente.

Assumindo os coeficientes dos modelos ARIMA como conhecidos, o efeito de um outlier pode ser calculado baseado nos resíduos dados em (3.1.2), pode-se verificar que o efeito do outlier presente na origem de previsão é o mesmo para todos os tipos de outlier. Caso não seja disponível nenhuma informação adicional não é possível distinguir o tipo, mesmo que o outlier seja detectado. No entanto, um efeito nas previsões depende do tipo do outlier. A previsão  $h$  passos à frente  $\hat{y}_n(h)$  pode ser calculada como segue, baseada na equação (3.4.12):

$$\begin{aligned}
 (1) \text{ Sem outlier} & \quad \hat{y}_n(h)_{SO} = \hat{x}_n(h)_{SO} \\
 (2) \text{ IO} & \quad \hat{y}_n(h)_{IO} = \hat{x}_n(h)_{IO} + \hat{\omega}\psi_h \\
 (3) \text{ AO} & \quad \hat{y}_n(h)_{AO} = \hat{x}_n(h)_{AO} \\
 (4) \text{ TC} & \quad \hat{y}_n(h)_{LS} = \hat{x}_n(h)_{LS} + \hat{\omega} \\
 (5) \text{ LS} & \quad \hat{y}_n(h)_{TC} = \hat{x}_n(h)_{TC} + \hat{\omega}\lambda^h
 \end{aligned} \tag{3.4.14}$$

onde  $\hat{x}_n(h)_{TP}$  é a previsão  $h$  passos à frente da série com o outlier do tipo TP removido, com origem de previsão  $n$ ;  $\hat{y}_n(h)_{TP}$  é a previsão  $h$  passos à frente da série com o outlier do tipo TP, com origem de previsão  $n$ . A notação  $\hat{x}_n(h)_{SO}$  (que é igual a  $\hat{y}_n(h)_{SO}$ ) representa a

previsão de  $x_{n+h}$  baseada nas observações com efeito do outlier removido, caso exista outlier. Caso não exista outlier,  $\hat{x}_n(h)_{SO} = \hat{y}_n(h)$ . No caso de um outlier ocorrer na origem de previsão, a estimativa do efeito do outlier para qualquer tipo de outlier é a mesma, e o seu impacto pode ser observado apenas em  $t = n$ . Assim, tomar uma decisão do tipo de outlier presente neste caso apenas afeta a extrapolação do efeito do outlier nas futuras observações na origem de previsão.

Os resultados da equação (3.4.14) mostram que a previsão com ajuste de outlier consiste na soma de previsões de série não contaminada e do efeito do outlier nas futuras observações. O primeiro componente representa o comportamento homogêneo das séries e o último denota o impacto do outlier nas futuras observações. Quando a origem da previsão é detectada como um outlier, o procedimento típico de detecção de outlier é capaz de obter uma estimativa para a primeira componente, mas pode-se ter dificuldade de obter informação completa para o segundo.

Para obter um melhor entendimento do efeito do ajuste de outlier na previsão, a previsão  $h$  passos à frente com o efeito do outlier removido  $\hat{y}_n(h)$  é expressa em função da previsão sem nenhuma remoção do efeito do outlier:

$$\begin{aligned}
 (1) \text{ Sem outlier} & \quad \hat{y}_n(h)_{SO} = \hat{x}_n(h)_{SO}, \\
 (2) \text{ IO} & \quad \hat{y}_n(h)_{IO} = \hat{y}_n(h)_{SO}, \\
 (3) \text{ AO} & \quad \hat{y}_n(h)_{AO} = \hat{y}_n(h)_{SO} - \hat{\omega}\psi_h, \\
 (4) \text{ LS} & \quad \hat{y}_n(h)_{LS} = \hat{y}_n(h)_{SO} + \hat{\omega}(1 - \psi_h), \\
 (5) \text{ TC} & \quad \hat{y}_n(h)_{TC} = \hat{y}_n(h)_{SO} + \hat{\omega}(\lambda^h - \psi_h).
 \end{aligned} \tag{3.4.15}$$

As expressões anteriores podem ser facilmente obtidas usando os resultados das equações (3.4.2), (3.4.3) e (3.4.5).

Os resultados da equação (3.4.15) mostram as diferenças entre as previsões com o outlier acomodado (i.e., considerando a presença do outlier) e com presença do outlier ignorado. É importante que no caso de IO, a previsão  $h$  passos à frente é idêntica à previsão sem ajuste de outlier. Esse resultado é congruente com a obtida por Ledolter (1989), e isso é uma forma adaptativa das previsões ARIMA. Isto é, quando um outlier produz um efeito que segue o comportamento da memória dos pesos  $\psi$ s não precisa conduzir qualquer ajuste para outlier para obter uma previsão mais precisa. Essa é a parte da razão que os modelos

ARIMA podem ainda ter um bom desempenho na previsão, mesmo que a série temporal seja ajustada aos outlier e intervenções.

Para obter propriedades estatísticas dos erros de previsão, é preciso expressar a previsão das séries não contaminadas em termos das previsões dos processos subjacentes como descrito na equação (3.4.1). A série ajustada pode ser obtida substituindo o outlier detectado na equação (3.4.7). Assim pode ser mostrado que

$$\hat{x}_n(h)_{TP} = \begin{cases} x_n(h) + \psi_h(\omega - \hat{\omega}) & \text{quando um outlier ocorre em } t = n, \\ x_n(h) - \psi_h \hat{\omega} & \text{quando um outlier não ocorre em } t = n, \end{cases} \quad (3.4.16)$$

onde TP = AO, TC, IO, LS e

$$\hat{x}_n(h)_{SO} = \begin{cases} x_n(h) + \psi_h \omega & \text{quando um outlier ocorre em } t = n, \\ x_n(h) & \text{quando um outlier não ocorre em } t = n, \end{cases} \quad (3.4.17)$$

onde  $x_n(h)$  é definido na equação (3.4.1). Baseado nas equações (3.4.7), e (3.4.17), pode-se derivar o contingente dos erros de previsão sob a especificação do tipo de outlier em termos de  $e_n(h)$  definido na equação (3.4.4) e o efeito de outlier. Esses erros de previsão são listados na Tabela 3.4.1. Nesta tabela os termos da diagonal apresentam os erros de previsão quando o tipo do outlier é corretamente especificado, e os termos fora da diagonal apresentam os erros de previsão quando o tipo de outlier é especificado incorretamente. Em particular, a primeira linha apresenta os erros de previsão quando as previsões são realizadas sem ajuste do outlier. Isso proporciona um vício devido ao outlier não detectado na origem da previsão.

Exceto no caso de LS, o principal efeito para um outlier é na previsão em curto período, i.e., quando  $h$  é pequeno. O efeito associado com o tipo do outlier especificado incorretamente depende do comportamento da memória subjacente  $\{\psi_h\}$  e do tipo do verdadeiro outlier. Usando o fato de que

$$\hat{\omega} = \begin{cases} \omega + a_n & \text{quando ocorre um outlier em } t = n \\ a_n & \text{quando não ocorre nenhum outlier em } t = n \end{cases} \quad (3.4.18)$$

pode-se obter o EQM da previsão para a previsão apresentados na Tabela 3.4.1. A primeira linha dessa tabela mostra os efeitos de vários outliers quando o outlier de detecção e ajuste não é empregado. Foi observado que exceto para IO, o EQM de previsão cresce se um

**Tabela 3.4.1 - Erro de previsão h passos à frente quando um outlier ocorre na origem da previsão, i.e., na última observação da série.**

Previsão	Verdadeiro				
	Sem outlier	IO	AO	TC	LS
Sem outlier	$e_n(h)$	$e_n(h)$	$e_n(h) - \omega \psi_h$	$e_n(h) - \omega(\psi_h - \delta^h)$	$e_n(h) - \omega(\psi_h - 1)$
IO	$e_n(h)$	$e_n(h)$	$e_n(h) - \omega \psi_h$	$e_n(h) - \omega(\psi_h - \lambda^h)$	$e_n(h) - \omega(\psi_h - 1)$
AO	$e_n(h) + \hat{\omega} \psi_h$	$e_n(h) + \hat{\omega} \psi_h$	$e_n(h) + (\hat{\omega} - \omega) \psi_h$	$e_n(h) + (\hat{\omega} - \omega) \psi_h + \hat{\omega} \lambda^h$	$e_n(h) + (\hat{\omega} - \omega) \psi_h + \hat{\omega}$
TC	$e_n(h) + \hat{\omega} (\psi_h - \lambda^h)$	$e_n(h) + \hat{\omega} (\psi_h - \lambda^h)$	$e_n(h) + (\hat{\omega} - \omega) \psi_h - \hat{\omega} \lambda^h$	$e_n(h) + (\hat{\omega} - \omega) (\psi_h - \lambda^h)$	$e_n(h) + (\hat{\omega} - \omega) \psi_h - (\hat{\omega} - \omega) \lambda^h$
LS	$e_n(h) + \hat{\omega} (\psi_h - 1)$	$e_n(h) + \hat{\omega} (\psi_h - 1)$	$e_n(h) + (\hat{\omega} - \omega) \psi_h - \hat{\omega}$	$e_n(h) + (\hat{\omega} - \omega) \psi_h - (\hat{\omega} - \omega) \lambda^h$	$e_n(h) + (\hat{\omega} - \omega) (\psi_h - 1)$

**Tabela 3.4.2 - EQM de previsão h passos à frente quando um outlier ocorre na origem da previsão, i.e., na última observação da série.**

Previsão	Verdadeiro				
	Sem outlier	IO	AO	TC	LS
Sem	$\sigma_e^2(h)$	$\sigma_e^2(h)$	$\sigma_e^2(h) + \psi_h^2 \omega^2$	$\sigma_e^2(h) + (\psi_h - \lambda^h)^2 \omega^2$	$\sigma_e^2(h) + (\psi_h - 1)^2 \omega^2$
IO	$\sigma_e^2(h)$	$\sigma_e^2(h)$	$\sigma_e^2(h) + \psi_h^2 \omega^2$	$\sigma_e^2(h) + (\psi_h - \lambda^h)^2 \omega^2$	$\sigma_e^2(h) + (\psi_h - 1)^2 \omega^2$
AO	$\sigma_e^2(h) + \psi_h^2 \sigma_a^2$	$\sigma_e^2(h) + \psi_h^2 (\sigma_a^2 + \omega^2)$	$\sigma_e^2(h) + \psi_h^2 \sigma_a^2$	$\sigma_e^2(h) + (\psi_h - \lambda^h)^2 \sigma_a^2 + \omega^2 \lambda^{2h}$	$\sigma_e^2(h) + \psi_h^2 \sigma_a^2 + \omega^2$
TC	$\sigma_e^2(h) + (\psi_h - \lambda^h)^2 \sigma_a^2$	$\sigma_e^2(h) + (\psi_h - \lambda^h)^2 (\sigma_a^2 + \omega^2)$	$\sigma_e^2(h) + (\psi_h - \lambda^h)^2 \sigma_a^2 + \omega^2 \lambda^{2h}$	$\sigma_e^2(h) + (\psi_h - \lambda^h)^2 \sigma_a^2$	$\sigma_e^2(h) + (\psi_h - \lambda^h)^2 \sigma_a^2 + \omega^2 (1 - \lambda^h)^2$
LS	$\sigma_e^2(h) + (\psi_h - 1)^2 \sigma_a^2$	$\sigma_e^2(h) + (\psi_h - 1)^2 (\sigma_a^2 + \omega^2)$	$\sigma_e^2(h) + (\psi_h - 1)^2 \sigma_a^2 + \omega^2$	$\sigma_e^2(h) + (\psi_h - 1)^2 \sigma_a^2 + \omega^2 (1 - \lambda^h)^2$	$\sigma_e^2(h) + (\psi_h - 1)^2 \sigma_a^2$

outlier ocorre perto do fim da série.

Examinando a Tabela 3.4.2 linha por linha é fácil verificar que a detecção de outlier correto produz o menor EQM. Para uma dada linha, a razão dos EQMs com o menor EQM da linha proporciona informação de perda de eficiência devido à decisão incorreta no uso de um tipo específico de outlier. As colunas da Tabela 3.4.2 lista os EQMs de previsão sob cinco diferentes condições. Para os casos de IO e sem outlier, se existe uma especificação errada de um AO, TC ou LS, o EQM aumenta. Quando um IO não é identificado ou uma observação é identificada espuriamente como IO, não há perda de eficiência no EQM. Esse resultado propõe uma base para argumentar que talvez não precise considerar o IO no contexto de previsão (Ledolter, 1989). No entanto, isso não parece uma boa sugestão, pois quando IO é identificado incorretamente como AO, TC ou LS, há sempre uma perda de eficiência no EQM.

Em geral, a perda de eficiência no EQM devido a má especificação do outlier depende (1) dos pesos  $\psi$  do processo subjacente, (2) do tamanho do efeito do outlier e (3) do tipo do outlier. Quando um AO, TC ou LS é identificado incorretamente como um IO ou sem outlier, ou quando um AO é identificado como TC ou um LS, o EQM da previsão é sempre maior do que o EQM com detecção correta, se  $|\omega| > \sigma_a$ . Quando um TC ou um LS é identificado incorretamente como um AO, TC ou LS, EQM cresce se

$$|\omega| > \sigma_a \sqrt{\frac{2(1-\psi_h)}{1-\lambda^h} + 1}.$$

Por exemplo, quando um IO de tamanho  $3\sigma_a$  na última posição é identificado como um AO, pela Tabela 3.4.1, o erro de previsão para um modelo ARIMA(0,1,1) com  $\theta = -0.8$  chega a  $5,4\sigma_a$ .

## Capítulo 4

### Efeito do outlier na previsão de séries temporais agregadas

Muitas séries temporais econômicas, tais como renda e consumo são disponíveis em diferentes níveis de agregação, por exemplo mensal ou trimestralmente. No entanto, em muitas situações, os economistas estão interessados em prever dados agregados (por exemplo, dados quadrimestrais ou anuais). Nesse caso a função de previsão pode ser baseada modelando-se tanto a série desagregada quanto a série agregada.

Na série agregada a perda de eficiência é pequena mas inevitável, então qual a necessidade de agregar e qual o interesse em medir a perda? Hotta e Cardoso (1993) citam algumas razões para trabalhar com a série agregada quando são disponíveis os valores desagregados, entre elas:

- 1) Em alguns casos a série agregada é melhor representada por um modelo linear do que a série desagregada. Uma indicação disso é o fato de que para muitas variáveis apenas as séries semestrais são ajustadas sazonalmente porque a série mensal não tem sazonalidade estável (Hotta e Cardoso, 1993; Teräsvirta, 1995).
- 2) Por outro lado, mesmo que a série desagregada possa ser modelada, por exemplo, por um processo ARIMA, o modelo desagregado, em geral, é mais afetado pelos outliers (Hotta, 1993; Hotta e Pereira, 1995).
- 3) Caso a série apenas seja disponível na forma agregada é interessante conhecer a eficiência relativa para saber se vale a pena investir para conseguir dados com maior nível de desagregação.

- 4) Nos casos onde uma parte da série é agregada e a outra desagregada, qual o melhor procedimento: agregar as observações desagregadas ou procurar as observações desagregadas nas agregadas?

Neste trabalho os estudos foram realizados com ênfase no período de agregação  $m = 4$ , seguindo os resultados teóricos de Hotta e Pereira (1995). Na primeira seção deste capítulo é apresentado o modelo agregado em séries temporais. Na segunda seção são estudados os efeitos de agregação na previsão. Nas seções 4.3 e 4.4 são apresentados os efeitos de outliers (AO e IO, respectivamente) na previsão no caso de modelos conhecidos e desconhecidos; para este último os resultados são obtidos através de simulações. Nos programas de simulações, foi implementado um teste de RV de detecção de outliers (apresentado na seção 2.1), onde são reestimados as ordens, os parâmetros e a magnitude destes outliers. Finalmente, na seção 4.5 são apresentadas as conclusões dos resultados das seções anteriores.

#### 4.1 Modelo agregado

Seja  $\{x_t, t = 1, \dots, n\}$  uma série temporal gerada pelo processo ARIMA(p,d,q):

$$\phi_p(B)(1-B)^d x_t = \theta_q(B) a_t. \quad (4.1.1)$$

A série agregada em  $m$  períodos sem superposição é dada por:

$$X_\tau = x_{\tau m - m + 1} + x_{\tau m - m + 2} + \dots + x_{\tau m}, \quad (4.1.2)$$

onde  $\tau = 1, \dots, N$  e  $N = n/m$ , supondo que  $n$  é múltiplo de  $m$ . Nesse caso,  $X_\tau$  segue um processo ARIMA(p,d,Q) (ver, por exemplo, Stram e Wei, 1986):

$$\Phi_p(B) (1-B)^d X_\tau = \Theta_Q(B) \varepsilon_\tau, \quad (4.1.3)$$

onde  $Q \leq [(m-1)(p+d+1)+q]/m$ ,  $[z]$  é o maior inteiro menor ou igual a  $z$ ,  $B$  é um operador de atraso, tal que,  $BX_\tau = X_{\tau-1}$  e a equação (4.1.3) é chamada de modelo agregado. Quando os parâmetros do modelo desagregado são conhecidos segue-se que os parâmetros do modelo agregado também o são.

As representações AR e MA equivalentes para o modelo (4.1.3) são dados, respectivamente, por:

$$\Pi(B) X_\tau = \varepsilon_\tau$$

e

$$X_\tau = \Lambda(B)\varepsilon_\tau.$$

Supondo que o pesquisador esteja interessado em prever a série agregada e não a série desagregada, então a previsão  $h$  passos à frente calculada pela série agregada é dada por:

$$X_N^*(h) = X_N(h), \quad (4.1.4)$$

enquanto calculada pela série desagregada é dada por:

$$X_N^d(h) = x_n(hm-m+1) + x_n(hm-m+2) + \dots + x_n(hm), \quad (4.1.3)$$

onde  $x_n(hm)$  é a previsão  $hm$  passos à frente de  $x_{n+hm}$  calculada com a origem de previsão  $n$ .

Hotta e Cardoso (1993) analisaram os efeitos de agregação na previsão em alguns modelos ARIMA conhecidos com período de agregação  $m = 2$ , enquanto Gonzalez (1992) estudou os efeitos de agregação em modelos estruturais. Hotta (1993) estudou os efeitos de AO na previsão de valores agregados e desagregados em modelos ARIMA. Hotta e Pereira (1995) generalizaram os estudos anteriores para os outliers AO e IO nas previsões de valores agregados e desagregados em modelos ARIMA e modelos estruturais, com período de agregação  $m = 4$ .

## 4.2 Efeito da agregação nas previsões

Quando o modelo é considerado conhecido (ordem, coeficientes e variância da inovação) o preditor linear ótimo, em termos de erro quadrático médio (EQM) de previsão, calculado com as observações agregadas é no máximo tão bom quanto o calculado com as observações desagregadas (ver, por exemplo, Lütkepohl, 1986 e Pino *et alii*, 1987).

Amemiya e Wu (1972) compararam esses dois preditores para processos AR( $p$ ) conhecidos e Ahsanullah e Wei (1984) para processo ARMA(1,1) conhecido. Gonzalez (1992) apresentou alguns resultados para modelos estruturados de série temporais usando o filtro de Kalman. Lütkepohl (1986) mostrou que em algumas situações particulares esses dois preditores, e conseqüentemente os EQMs, são iguais. Hotta e Cardoso (1993) os estudaram para modelos ARIMA(0,d,d) com  $d = 1$  e 2, com período de agregação 2.

Nas duas subseções seguintes os resultados de Amemiya e Wu (1972), Ahsanullah e Wei (1984), Gonzalez (1992), Hotta e Cardoso (1993) e Hotta e Pereira (1995) são estendidos para avaliar a eficiência do preditor resultante das observações desagregadas em comparação com o preditor proveniente das observações agregadas para processos ARIMA(0,1,1) e ARIMA(1,1,0) conhecidos.

Na subseção 4.2.3 é discutido um procedimento para contornar o problema da estimação das ordens de modelos desconhecidos.

#### 4.2.1 Perda de eficiência na previsão devido à agregação num processo ARIMA(1,1,0) conhecido

Suponha que a série desagregada siga um processo ARIMA(1,1,0) dado por:

$$(1-B)(1-\phi B) x_t = a_t,$$

e que o pesquisador esteja interessado em prever observações agregadas de um período  $m$  igual a 4, por exemplo, previsão anual de uma série trimestral.

O modelo agregado é um processo ARIMA(1,1,2) com o coeficiente autoregressivo  $\Phi = \phi^4$ , os dois coeficientes de média móvel e a variância da nova inovação dados pelo seguinte sistema de equações:

$$\begin{cases} (1 + \Theta_1^2 + \Theta_2^2)\sigma_e^2 = (44 + 80\phi + 106\phi^2 + 120\phi^3 + 106\phi^4 + 80\phi^5 + 44\phi^6)\sigma_a^2, \\ -\Theta_1(1 + \Theta_2)\sigma_e^2 = (10 + 24\phi + 42\phi^2 + 64\phi^3 + 42\phi^4 + 24\phi^5 + 10\phi^6)\sigma_a^2, \\ -\Theta_2\sigma_e^2 = (\phi^2 + 4\phi^3 + \phi^4)\sigma_a^2. \end{cases}$$

Os valores máximos de  $\Theta_1$  e  $\Theta_2$  tendem, respectivamente, a 0,828 e 0,172, quando  $\phi \rightarrow -1$ . No limite, quando  $\phi = -1$  o modelo agregado passa a ser um processo ARIMA(0,1,1).

Os valores de  $\sigma_e$  para cada valor de  $\phi$  estão ilustrados no Gráfico 4.2.1 e é possível ver que o valor de  $\sigma_e$  pode ser muito grande dependendo do valor de  $\phi$ .

Os valores observados podem ser reescritos como

$$X_\tau = \frac{(1 + B + \dots + B^{m-1})}{(1 - \phi B)(1 - B)} a_{tm} = \sum_{j=0}^{\infty} b_j a_{tm-j}$$

onde

$$b_j = \sum_{i=0}^j (j+1-i)\phi^i, \text{ para } j = 0, \dots, m-1$$

$$b_j = m \sum_{i=0}^{j-m+1} \phi^i + \sum_{i=j-m+2}^j (j+1-i)\phi^i$$

$$= m \sum_{i=0}^{j-m+1} \phi^i + \sum_{k=1}^{m-1} k\phi^{j+1-k}, \text{ para } j = m, \dots, hm-1$$

Assim o EQM pelo modelo desagregado é dado por

$$EQM^d(h) = \sum_{j=0}^{hm-1} b_j^2 \sigma_a^2.$$

Analogamente

$$X_\tau = \frac{(1 - \Theta_1 B - \Theta_2 B^2)}{(1 - \Phi B)(1 - B)} \varepsilon_\tau$$

$$= \sum_{j=0}^{\infty} d_j \varepsilon_{\tau-j}$$

onde

$$d_j = \begin{cases} 1 & j = 0 \\ 1 + \Phi - \Theta_1 & j = 1 \\ \phi^j + (1 - \Theta_1)\phi^{j-1} + \left(\sum_{i=0}^{j-2} \Phi^i\right)(1 - \Theta_1 - \Theta_2) & j > 1 \end{cases}$$

Assim, o EQM é

Os EQMs calculados pelos modelos desagregados e agregados são dados respectivamente por

$$EQM^a(h) = \sum_{j=1}^h d_j^2 \sigma_e^2.$$

Uma forma de analisar a perda de eficiência devida à agregação é através da razão dos EQMs, essa razão é denominada *eficiência*, que é definida como

$$ef(h) = EQM^d(h)/EQM^a(h).$$

Utilizando esses EQMs as eficiências foram calculadas para alguns valores de h, e apresentadas no Gráfico 4.2.2. Através desse gráfico pode-se ver que a eficiência aumenta quando h cresce.

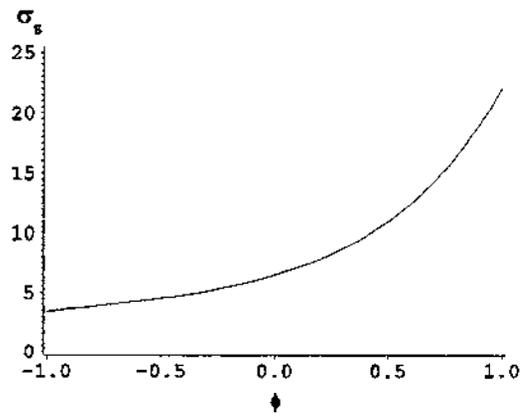


Gráfico 4.2.1 - Valores do desvio padrão da série ARIMA(1,1,0) agregada,  $\sigma_s$ , com agregação  $m = 4$ .

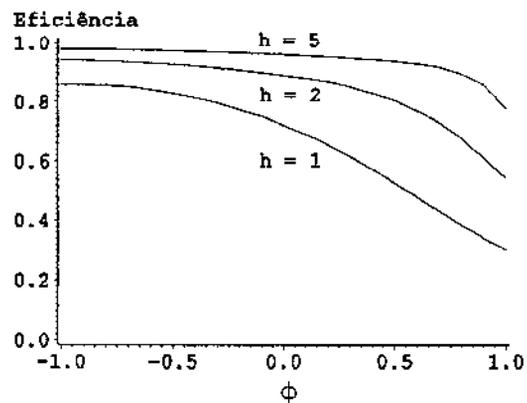


Gráfico 4.2.2 - Eficiência de agregação na previsão  $h$  passos à frente para agregação  $m = 4$ , no modelo ARIMA(1,1,0) conhecido.

### 4.2.2 Perda de eficiência na previsão devido à agregação num processo ARIMA(0,1,1) conhecido

Seja  $x_t$  um processo ARIMA(0,1,1) gerado por:

$$x_t = x_{t-1} + a_t - \theta a_{t-1}, \quad (4.2.1)$$

onde  $a_t$ s são variáveis aleatórias identicamente e independentemente distribuídas (i.i.d.) com  $E(a_t) = 0$ ,  $E(a_t^2) = \sigma_a^2$  e  $|\theta| \leq 1$ .

A observação agregada, com período de agregação  $m = 4$ , é dada por:

$$X_\tau = [1 + (2-\theta)B + (3-2\theta)B^2 + (4-3\theta)B^3 + \sum_{i=4}^{\infty} (4-4\theta)B^i] a_{4\tau}. \quad (4.2.2)$$

Assim, o preditor  $h$  passos à frente, baseado nas observações desagregadas, tem um EQM dado por

$$EQM^d(h) = \{1 + (2-\theta)^2 + (3-2\theta)^2 + (4-3\theta)^2 + 4^3(h-1)(1-\theta)^2\} \sigma_a^2. \quad (4.2.3)$$

Genericamente, para qualquer período de agregação  $m$ ,

$$EQM^d(h) = \left[ \sum_{i=1}^m [i - (i-1)\theta]^2 + m^3(h-1)(1-\theta)^2 \right] \sigma_a^2. \quad (4.2.4)$$

Observe que a equação (4.2.2) pode ser escrita como,

$$X_\tau - X_{\tau-1} = \mu_\tau,$$

onde  $\mu_\tau$  segue um processo MA(1), pois  $E(\mu_\tau \mu_{\tau+j})$  é diferente de zero apenas para  $j = 1$ . (Anderson, 1971, p. 400-6). Assim, as observações agregadas podem ser consideradas como sendo geradas por um processo ARIMA(0,1,1)

$$X_\tau - X_{\tau-1} = \varepsilon_\tau - \Theta \varepsilon_{\tau-1}. \quad (4.2.5)$$

O fato de igualar a função de autocorrelação das diferenças de  $X_\tau$  das equações (4.2.2) e (4.2.5) implica que  $\sigma_\varepsilon^2$  e  $\Theta$  são soluções do seguinte sistema de equações (com  $|\Theta| \leq 1$ ):

$$\begin{aligned} (44 - 80\theta + 44\theta^2)\sigma_a^2 &= (1 + \Theta^2)\sigma_\varepsilon^2, \\ (10 - 24\theta + 10\theta^2)\sigma_a^2 &= -\Theta\sigma_\varepsilon^2. \end{aligned} \quad (4.2.6)$$

Os valores de  $\Theta$  e  $\sigma_\varepsilon$  para cada valor de  $\theta$  estão ilustrados nos Gráficos 4.2.3 (a) e (b), respectivamente.

O EQM de previsão  $h$  passos à frente usando o modelo agregado é dado por

$$EQM^a(h) = \{1 + (h-1)(1-\Theta)^2\} \sigma_\varepsilon^2.$$

No Gráfico 4.2.4 são apresentadas as eficiências de previsão  $h$  passos à frente para  $h = 1, 2$  e  $5$  com períodos de agregação  $m = 2$  e  $4$ . Pode-se verificar que em ambas as agregações a eficiência cresce quando  $\theta$  e/ou  $h$  cresce, como foi observado por Hotta e Cardoso (1993), para  $m = 2$ .

A perda de eficiência para previsão um passo à frente pode ser grande sendo a eficiência mínima em torno de  $0,55$  em ambos os períodos de agregação. Para  $h > 1$  a eficiência é no mínimo  $0,86$  e  $0,82$ , respectivamente, para  $m = 2$  e  $4$ .

Considere agora que o processo  $ARIMA(0,1,1)$  foi gerado por um modelo estrutural de nível local, i.e., que  $\theta$  é positivo. O desempenho da previsão neste caso para  $m = 2$ , mesmo quando  $h = 1$  é bom, sendo a eficiência no mínimo igual a  $0,858$ . A eficiência para  $m = 4$  não é tão boa quanto no caso anterior, sendo no mínimo  $0,721$ , para  $h = 1$  e no mínimo  $0,89$  para  $h > 1$ .

Embora o processo  $ARIMA(0,1,1)$ , ao ser agregado, continue seguindo o mesmo processo  $ARIMA$ , porém com parâmetros diferentes, ele não necessariamente pode ser representado como um modelo estrutural. No caso de período de agregação  $4$  apenas para  $\theta > 0,54$  a série agregada segue um processo de modelo estrutural de nível local (ver Gráfico 4.2.3 (a)).

O método de alisamento exponencial é muito relacionado ao modelo estrutural. Utilizando esse método, muitos autores sugerem usar o parâmetro de alisamento menor que  $0,3$ , que é equivalente a ter  $\theta > 0,7$ . Dentro dessa amplitude, no caso de período de agregação  $m = 2$ , tem-se uma eficiência mínima igual a  $0,994$  para previsão um passo à frente e igual a  $0,996$  para previsão três passos à frente, e de  $0,967$  e  $0,979$ , respectivamente, para  $m = 4$ .

### **4.2.3 Perda de eficiência na previsão devido à agregação em modelos ARIMA desconhecidos**

Na prática, geralmente, não se conhece as ordens do processo. Uma forma de contornar a subjetividade envolvida na escolha dessas ordens é fazê-la de uma forma automática e padronizada. Como um modelo  $ARMA$  é aproximável por um modelo  $AR(P)$ , uma solução é selecionar  $P$  através de um critério estatístico. Lütkepohl (1986) usou esse procedimento e realizou algumas simulações para comparar EQMs usando observações

agregadas. O critério escolhido para esta dissertação foi BIC<sup>1</sup> (Schwartz, 1978) devido à sua eficiência e ao fato de reduzir o problema de super-parametrização existente no critério AIC (Schwartz, 1978).

São introduzidos nesse procedimento mais dois tipos de erro: o erro devido à aproximação do processo ARMA(p, q) por um processo AR (embora o erro seja nulo se  $q = 0$  e o critério selecionar a ordem correta de AR) e o erro devido à estimação dos parâmetros. A influência do erro de estimação deve diminuir com o tamanho da série sendo assintoticamente igual a zero. Uma discussão sobre o efeito da má especificação em modelos autoregressivos pode ser encontrado em Bhansali (1981) e Kunitomo e Yamamoto (1985). No entanto, eles não consideram o efeito da estimação das ordens. Dada a dificuldade de avaliar a influência geral, Hotta e Cardoso (1993) estudaram as influências através de simulações, para período de agregação  $m = 2$ . Eles concluíram que mesmo com o modelo desconhecido, o preditor que usa as observações agregadas tem um desempenho bom comparado com o preditor que usa as observações desagregadas.

### 4.3 Efeito de AO na previsão

Se os parâmetros do modelo são conhecidos, já foi mencionado anteriormente que o preditor que usa valores agregados é no máximo tão bom quanto o preditor que usa observações desagregadas, em termos de EQM. No entanto, em algumas situações a série  $\{x_t\}$  não é observada diretamente, e sim contaminada por um ou mais outliers. Os efeitos causados pelo outlier na previsão de valores desagregados foram estudados na seção 3.4. Nesta seção são comparados os efeitos de AO na previsão de valores agregados através dos modelos agregados e desagregados.

A presença de um outlier  $r$  ( $r \geq 0$ ) observações antes da origem de previsão no modelo desagregado é equivalente à presença de um AO, de mesmo tamanho, à  $R$  observações na série agregada, onde  $R = [r/m]$  ( $[z]$  é o maior inteiro menor ou igual a  $z$ ), e  $m$  o período de agregação. Neste estudo focalizou-se apenas o caso onde os outliers ocorrem nas últimas observações. A presença de dois ou mais AOs nas  $m$  últimas observações na série desagregada é equivalente a um AO na série agregada na última observação, com a magnitude do outlier igual à soma das magnitudes dos outliers do modelo desagregado.

---

<sup>1</sup> As idéias básicas de BIC estão no Apêndice 3

Na primeira subseção são apresentados os efeitos de AO na previsão supondo o modelo conhecido, onde são considerados os casos em que os processos desagregados são dados pelos modelos ARIMA(0,1,1) e ARIMA(1,1,0). Quando o modelo é desconhecido o estudo teórico do efeito se torna inviável, dado o efeito conjunto de identificação, estimação e teste de detecção de outlier. Desta forma, os estudos foram realizados através de simulações considerando separadamente, a utilização ou não do teste automático de detecção de outlier. A ordem de não estacionariedade do modelo ARIMA é considerada conhecida. Os resultados dessas simulações são apresentados na subseção 4.3.2.

### 4.3.1 Modelos conhecidos

Hillmer (1984) e Ledolter (1987 e 1989) estudaram o efeito do outlier na previsão na série desagregada quando o outlier ocorre perto da origem de previsão.

Hotta e Pereira (1995) mostraram que em alguns modelos conhecidos o efeito de AO pode ser maior em  $X_N^d(h)$  que em  $X_N^*(h)$  quando um outlier ocorre perto da origem de previsão. Eles observaram também que, geralmente, é mais difícil detectar um outlier que ocorre perto da origem de previsão do que um outlier que ocorre no meio da série.

Considere o modelo desagregado com um AO de tamanho  $\omega_A$  que ocorre  $r$  ( $r = n - k$ ) observações antes da origem de previsão, i.e., na  $(n-r)$ -ésima (ou  $k$ -ésima) observação da série dado por

$$x_t = \psi(B)a_t + \omega_A \delta_{t,n-r},$$

onde  $\delta_{t,k}$  é indicador da presença do outlier,  $\delta_{t,k} = 1$  se  $t = k$  e zero, no caso contrário.

Se os coeficientes do modelo ARIMA são conhecidos e a presença do outlier não é detectada, o erro de previsão  $h$  passos à frente é dado por (Hillmer, 1984):

$$e_n(h) = y_{n+h} - \hat{y}_n(h) = \sum_{j=0}^{h-1} \psi_j a_{n+h-j} - \pi_{r+1}^{(h)} \omega_A, \quad (4.3.1)$$

onde  $\pi_r^{(h)}$  é dada pela equação (3.4.2).

A previsão ótima de valores agregados  $h$  passos à frente pelo modelo desagregado é dada por:

$$\hat{Y}_N^d(h) = \sum_{i=0}^{m-1} \hat{y}_n(hm-i),$$

com erro de previsão igual a:

$$\begin{aligned}\hat{\varepsilon}_N^d(h) &= \sum_{i=0}^{m-1} e_n(hm-i) = \sum_{i=0}^{m-1} \left\{ \sum_{j=0}^{hm-i-1} \psi_j a_{n+hm-i-j} - \pi_{r+1}^{(hm-i)} \omega_A \right\}, \\ &= \sum_{i=0}^{hm-1} \psi_{h,i} a_{n+hm-i} - \sum_{i=0}^{m-1} \pi_{r+1}^{(hm-i)} \omega_A,\end{aligned}\quad (4.3.2)$$

onde  $\psi_{h,i} = \sum_{j=b}^i \psi_j$ , e  $b = \max(0, i-m+1)$ .

Utilizando o resultado dado pela equação anterior o erro de previsão que usa o modelo agregado é dado por:

$$\hat{\varepsilon}_N^a(h) = \sum_{j=0}^{h-1} \Psi_j \varepsilon_{N+h-j} - \Pi_{R+1}^{(h)} \omega_A \quad (4.3.3)$$

onde  $\Pi_j^{(h)}$  é dado por expressão similar a (3.4.2),  $R = [r/m]$  é a posição do outlier na série agregada, em relação à origem de previsão. Como neste trabalho  $r$  assume valores 0 ou 1 (i.e., ocorre na última ou penúltima observação, respectivamente)  $R$  sempre assume valor zero, i.e., a presença de dois outliers na série desagregada é equivalente à presença de um outlier na última observação da série agregada com o tamanho igual a soma dos dois outliers.

Nesse trabalho os efeitos na previsão serão analisados através de duas estatísticas. Uma delas é o efeito no EQM de previsão e a outra é o efeito na média ou vício na previsão de valores agregados. Este vício é definido como a diferença entre os erros de previsão considerando-se que o outlier é conhecido e desconhecido. Os modelos desagregado e agregado, são dados pelo segundo termo das equações (4.3.2) e (4.3.3), respectivamente, ou seja

$$\text{vício}^d(r,h) = \sum_{i=0}^{m-1} \pi_{r+1}^{(hm-i)} \omega_A, \quad (4.3.4a)$$

$$\text{e} \quad \text{vício}^a(r,h) = \Pi_{R+1}^{(h)} \omega_A. \quad (4.3.4b)$$

Pode-se chamá-los de efeitos na média ou efeitos esperados, já que a média dos primeiros termos de (4.3.2) e (4.3.3) são iguais a zero.

Os EQMs de previsão podem ser calculados por:

$$\begin{aligned}\text{EQM}^d(r,h) &= \text{Var}(\hat{\varepsilon}_N^d(h)) + [\text{vício}^d(r,h)]^2 \\ &= \sigma_a^2 \sum_{i=0}^{hm-1} (\psi_{h,i})^2 + \left( \sum_{i=0}^{m-1} \pi_{r+1}^{(hm-i)} \omega_A \right)^2,\end{aligned}\quad (4.3.5a)$$

$$\begin{aligned}
 e \quad EQM^a(r,h) &= \text{Var}(\hat{e}_N^a(h)) + [\text{vício}^a(r,h)]^2 \\
 &= \sigma_\varepsilon^2 \sum_{j=0}^{h-1} (\Psi_j)^2 + \left( \Pi_{R+1}^{(h)} \omega_A \right)^2, \quad (4.3.5b)
 \end{aligned}$$

para modelos desagregados e agregados, respectivamente. Os EQMs são analisados através da eficiência, definida por:

$$ef(r,h) = EQM^d(r,h)/EQM^a(r,h).$$

No restante do trabalho, o vício, o EQM e a eficiência de previsão de valores agregados através de modelo agregado (desagregado) serão referidos apenas como vício, EQM e eficiência do modelo agregado (desagregado).

Lütkepohl (1986) verificou que sob algumas condições as previsões são iguais e assim o efeito na predição em ambos os modelos são iguais.

Se o modelo desagregado não tem o termo média móvel então  $\psi_j^{(h)}$  é igual a zero para  $j > p+d$ . Assim, um AO não afetará as previsões desagregadas se ele ocorrer antes das últimas  $p+d$  observações, i.e.,  $r > p+d$ .

Mesmo que o modelo desagregado não tenha o termo média-móvel, o modelo agregado sempre tem o componente de média-móvel (isto só não ocorre nos casos especiais quando existe cancelamento de raízes) e, conseqüentemente,  $\Pi_j^{(h)}$  é diferente de zero para quaisquer valores de  $j$  e  $h$ . Isso significa que o efeito de um AO no modelo agregado será sempre diferente de zero, apesar dele passar a ser desprezível com o tempo, devido à condição de invertibilidade. O mesmo acontece com o modelo agregado quando o modelo desagregado tem um termo média-móvel. Assim, o efeito do outlier, em ambos os modelos será transportado para todas as observações posteriores a ocorrência do outlier, embora esse efeito desapareça também com o tempo, exceto quando o modelo não é invertível.

Como o efeito do outlier é uma função linear do tamanho do outlier, o efeito de múltiplos outliers é a soma dos efeitos individuais, na suposição dos modelos serem conhecidos.

A seguir são analisados os casos de ARIMA(1,1,0) e ARIMA(0,1,1) separadamente. Para cada caso são analisados os efeitos na média (vício) e no EQM na previsão um passo à frente, quando a série contém um ou dois outliers, neste último caso com outliers de mesmo sinal ou de sinais trocados. O efeito na média é analisado também para previsões  $h$  ( $h > 1$ ) passos à frente.

### 4.3.1.1 - ARIMA(1,1,0) desagregado

Suponha que a série desagregada siga um processo ARIMA(1,1,0) dado por

$$(1 - \phi B)(1 - B)x_t = a_t,$$

e o interesse seja prever observações agregadas de período  $m$  igual a 4. Sabe-se que o modelo agregado é um processo ARIMA(1,1,2) com os coeficientes apresentados na subseção 4.2.2.

O período de agregação tem um papel duplo na persistência do vício. Quando o período é grande o coeficiente autoregressivo do modelo agregado se torna menor, o que tende a diminuir a persistência, exceto quando  $|\phi| = 1$ . Por outro lado, um outlier presente  $r$  ( $r > 0$ ) observações antes da origem de previsão na série desagregada deve ser localizado a  $R = [r/m]$  ( $< r$ ) observações na série agregada, o que tende a aumentar a persistência.

São analisados a seguir os efeitos causados por um AO num modelo ARIMA(1,1,0) desagregado e agregado. Para facilitar, a análise, são considerados três casos:

- 1) Um outlier de tamanho 1,
- 2) Dois outliers de tamanho 1 de mesmos sinais e
- 3) Dois outliers de tamanho 1 de sinais trocados.

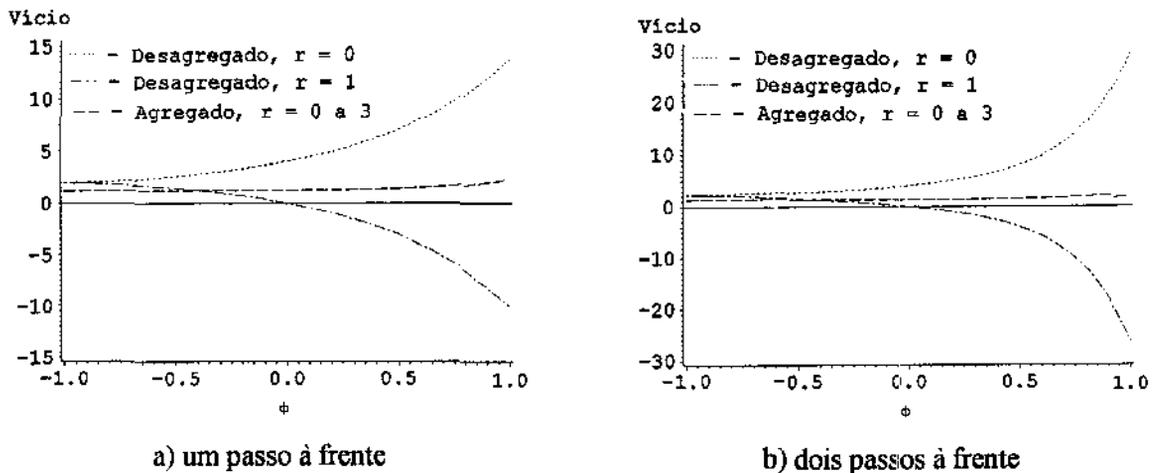
Para cada caso serão analisados os efeitos na média, no EQM e na eficiência.

#### 1) Um AO de tamanho 1

##### a) Efeito na média (vício)

O vício de um AO de tamanho  $\omega_A = \sigma_a = 1$  na previsão quando o modelo é conhecido é apresentado no Gráfico 4.3.1 para  $m = 4$ . Este gráfico mostra os vícios de um AO que ocorre nas duas últimas observações para previsões um e dois passos à frente.

As principais conclusões pela análise desses gráficos são citadas a seguir:



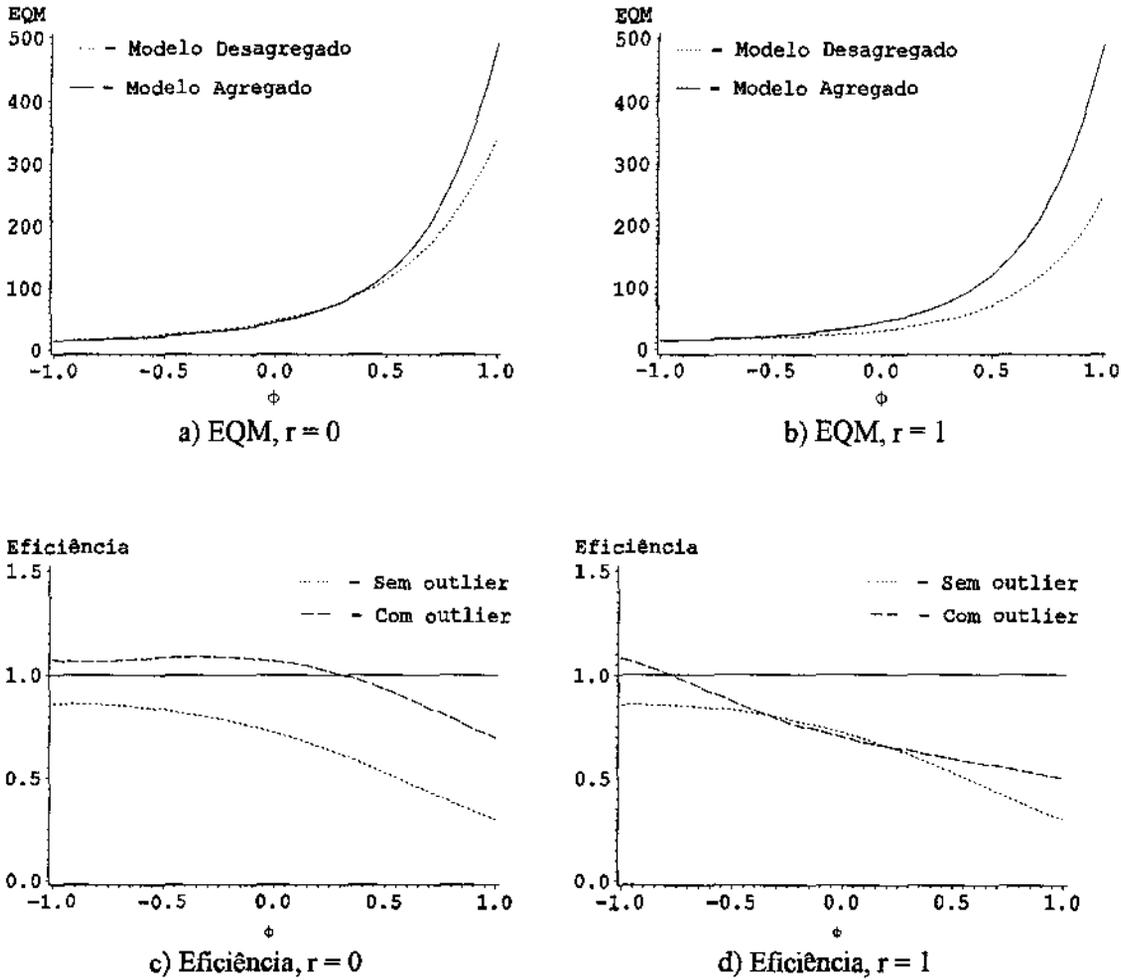
**Gráfico 4.3.1** - Vício causado pela presença de um AO de tamanho 1 na previsão  $h$  ( $h = 1, 2$ ) passos à frente, em modelo ARIMA(1,1,0) conhecido. AO ocorre  $r$  observações antes da origem de previsão.

- no modelo desagregado os vícios são simétricos em torno do valor 2 para  $r = 0$  e 1.
- Se o outlier ocorre na última observação ( $r = 0$ ), o vício no modelo desagregado é sempre maior.
- Quando o outlier ocorre na penúltima observação ( $r = 1$ ) o vício do modelo desagregado pode ser muito grande (máximo de 10 e 30 para um e dois passos à frente, respectivamente), enquanto que para o modelo agregado os valores máximos são 2,44 e 3,89 para um e dois passos à frente, respectivamente.
- O vício é maior no modelo agregado do que no modelo desagregado quando  $r = 1$  e  $\phi$  pertence (aproximadamente) ao intervalo  $[-0,44; 0,26]$  para um e dois passos à frente. No entanto, nesse intervalo o vício no modelo agregado é menor que 1,32.

Considerando que o outlier é mais difícil de ser detectado quando ocorre na última observação, e que neste caso o vício pode ser muito grande no modelo desagregado, pode-se dizer que o vício causado por um AO é mais perigoso neste modelo.

#### b) Efeito no EQM

Os EQMs e as eficiências estão ilustrados no Gráfico 4.3.2. Em geral, a presença do outlier aumenta a eficiência da previsão do modelo agregado, principalmente quando o outlier ocorre na última observação. Neste caso, para a maioria dos valores de  $\phi$  ( $\phi < 0,31$ ), as previsões realizadas através do modelo agregado são mais eficientes. A melhora da

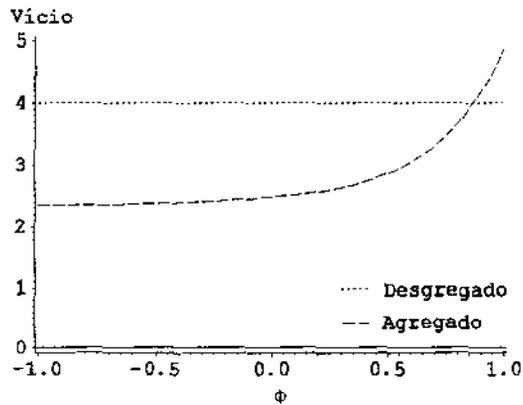


**Gráfico 4.3.2 - EQM e eficiência de previsão um passo à frente de modelos desagregado e agregado, quando o modelo desagregado é um processo ARIMA(1,1,0), com um AO de tamanho um presente  $r$  posições antes da origem de previsão, para  $r = 0$  e  $1$ .**

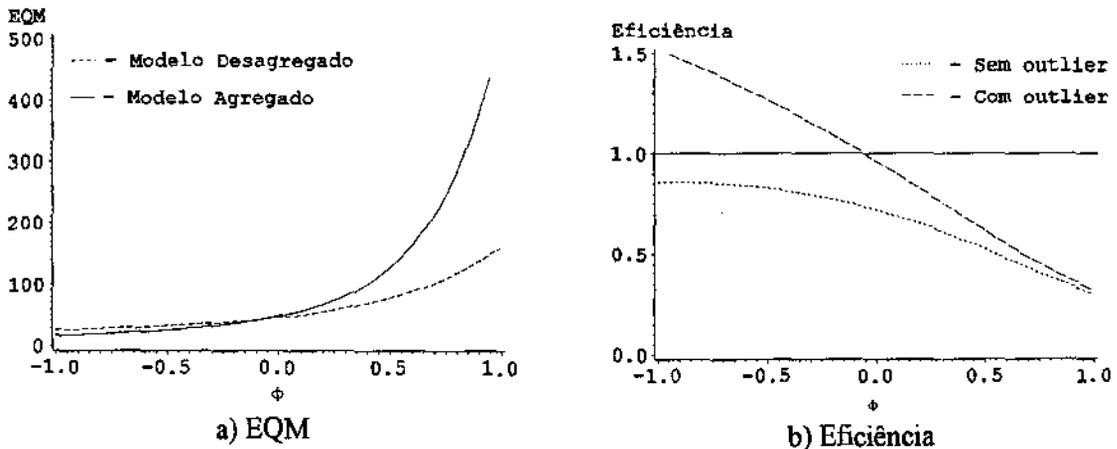
eficiência não é tão marcante quando o outlier ocorre na penúltima observação.

No caso do outlier presente na última observação ( $r = 0$ ), o EQM do modelo agregado é menor do que do desagregado quando  $\phi < 0,31$ , mas neste caso os valores são bem pequenos e próximos (eficiência entre 1 e 1,1), enquanto que para  $\phi > 0,31$ , a eficiência é menor do que 1 e diminui conforme  $\phi \rightarrow +1$ .

Da mesma forma, quando  $r = 1$ , o EQM do desagregado é maior para  $\phi < -0,78$ , sendo a diferença muito pequena (eficiência entre 1 e 1,1), enquanto no modelo agregado para  $\phi > -0,78$  o EQM é maior e crescente em  $\phi$ , chegando a uma eficiência de 0,5 no caso limite,  $\phi = 1$ .



**Gráfico 4.3.3** - Vício causado por dois AOs de tamanho 1, de mesmo sinal, nas duas últimas observações na previsão  $h$  ( $1 \leq h \leq 4$ ) passos à frente no modelo ARIMA(1,1,0).



**Gráfico 4.3.4** - EQM e eficiência de previsão um passo à frente com dois AOs de tamanho um, de mesmo sinal, nas duas últimas observações da série desagregada ARIMA(1,1,0).

## 2) Dois AOs de tamanho 1 com mesmo sinal

### a) Efeito na média

Os vícios causados pelos dois AOs de tamanho 1, de mesmo sinal, nas duas últimas observações estão apresentados no Gráfico 4.3.3, para  $h$  ( $1 \leq h \leq 4$ ) passos à frente. No modelo agregado o vício é menor do que no modelo desagregado quando  $\phi < 0,85$ . O vício no modelo

agregado é crescente em  $\phi$ , atingindo o valor máximo igual a 5,0 quando  $\phi = +1$ . Como no caso anterior, em geral, o vício de um AO é mais perigoso no modelo desagregado.

#### b) Efeito no EQM

Nos Gráfico 4.3.4 estão apresentados os efeitos no EQM e a eficiência de previsão um passo à frente na presença de dois AOs, de mesmo sinal, nas duas últimas observações. A presença dos outliers aumenta a eficiência relativa do modelo agregado para todos os valores de  $\phi$ . O aumento da eficiência é mais acentuado quando  $\phi$  é próximo de  $-1$  e diminui conforme  $\phi \rightarrow +1$ . Para  $\phi < -0,07$  o EQM do modelo desagregado passa a ser maior do que do modelo agregado, i.e., o modelo agregado é mais eficiente, chegando a atingir uma marca de 1,5 no limite quando  $\phi = -1$ , no entanto, no limite os EQMs são pequenos e muito próximos em termos absolutos. Já para  $\phi > -0,07$  o EQM do modelo agregado é maior e a diferença aumenta conforme  $\phi \rightarrow +1$ , chegando a uma eficiência de aproximadamente 0,35.

### 2) Dois AOs de tamanho 1 com sinais trocados

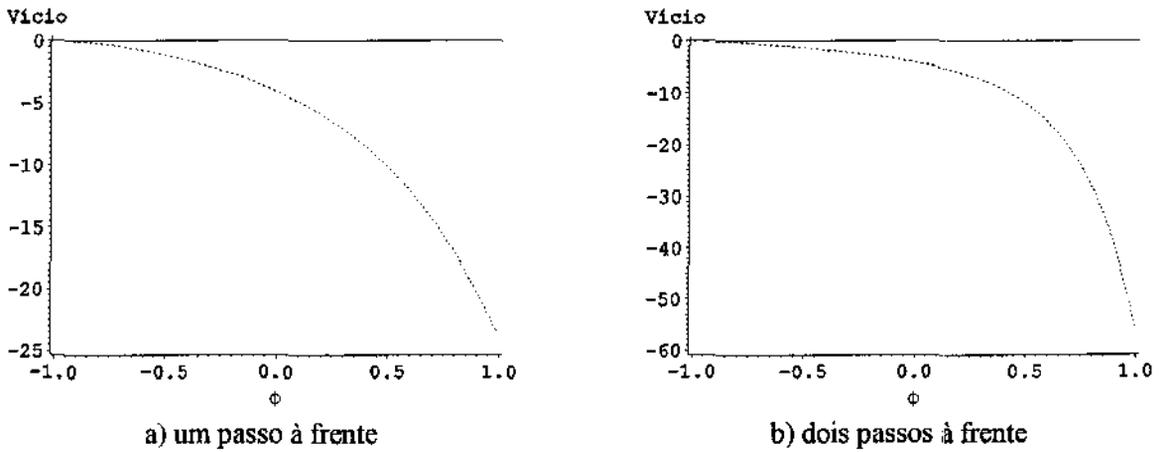
#### a) Efeito na média

No Gráfico 4.3.5 é apresentado o vício causado pelos dois AOs de tamanho 1 e -1, na penúltima e última observações, respectivamente. O vício no modelo agregado é nulo, pois os outliers somados se cancelam. Já no modelo desagregado é crescente em  $\phi$ , atingindo o valor máximo em módulo de 24,0 em  $\phi = +1$  para previsão um passo à frente e de 60,0 para dois passos à frente.

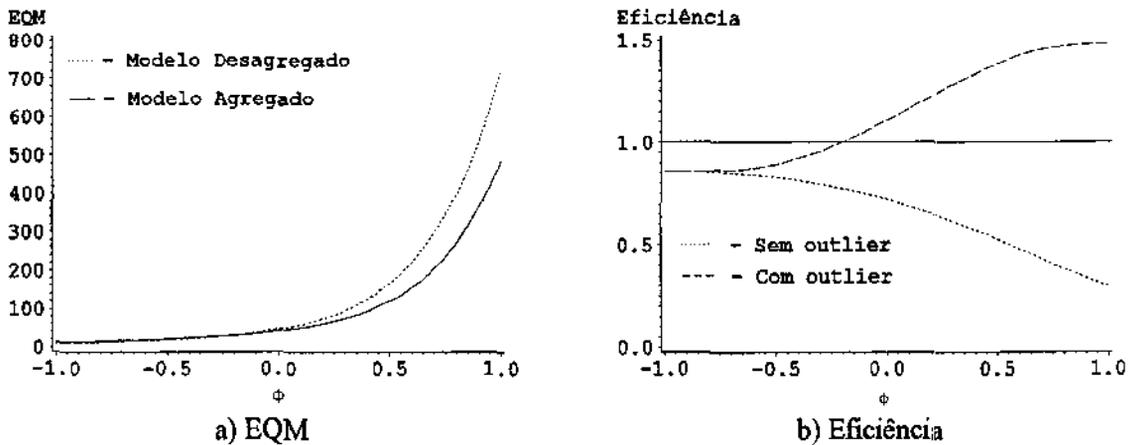
#### b) Efeito no EQM

Como no caso anterior, a presença do outlier sempre aumenta a eficiência do modelo agregado, no entanto esta eficiência só se dá para  $\phi$  maiores.

Analisando-se o Gráfico 4.3.6 pode-se verificar que o EQM do modelo agregado é sempre menor para  $\phi > -0,2$ , com eficiência crescente em  $\phi$ , atingindo 1,5 quando  $\phi = +1$ . Como nos casos anteriores, em alguns valores de  $\phi$  ( $\phi < -0,25$ ) a eficiência é menor que 1, sendo que os valores de EQM nesse intervalo são muito pequenos, para  $\phi > -0,25$  a magnitude nos EQMs são maiores e a eficiência é maior do que 1.



**Gráfico 4.3.5** - Vício causado por dois AOs de tamanho 1 e -1, respectivamente, na penúltima e última observações num modelo ARIMA(1,1,0) desagregado (no modelo agregado o vício é zero).



**Gráfico 4.3.6** - EQM e eficiência de previsão 1 passo à frente de séries com dois AOs de tamanho um de sinais trocados nas duas últimas observações da série no modelo ARI.

#### 4.3.1.2 - ARIMA(0,1,1) desagregado

Considere o processo ARIMA(0,1,1) dado por:

$$(1-B)x_t = a_t - \theta a_{t-1}$$

com período de agregação  $m$  igual a 4.

O modelo agregado é um processo ARIMA(0,1,1) com coeficiente de média móvel  $\Theta$  e variância de inovação  $\sigma_e^2$ , dados pela resolução do sistema de equações (4.2.6).

A previsão  $h$  passos à frente para os modelos desagregado e agregado é constante para qualquer horizonte. O vício do outlier que ocorre  $r$  observações antes da origem de previsão é dado por:

$$\text{vício}^d(r,h) = m\theta^r(1-\theta)\omega_A,$$

$$\text{vício}^a(r,h) = \Theta^{[r/m]}(1-\Theta)\omega_A,$$

para os modelos desagregado e agregado, respectivamente.

Como no caso anterior do modelo ARIMA(1,1,0), os efeitos serão analisados separadamente.

### 1) Um AO de tamanho 1

#### a) Efeito na média

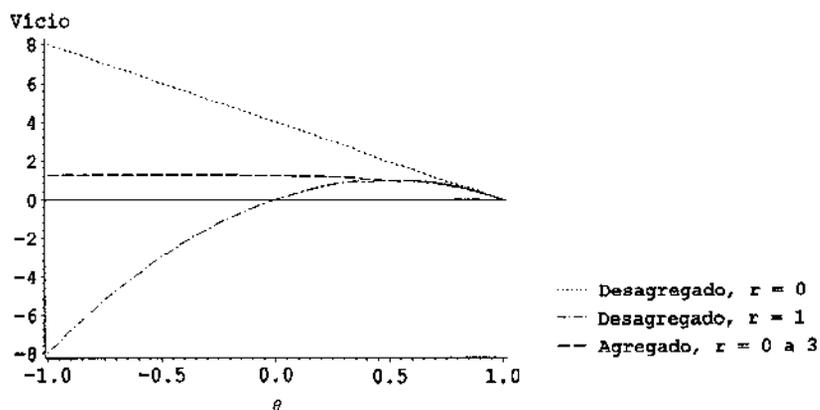
Os vícios causados pela presença de um único outlier, num modelo conhecido, estão apresentado no Gráfico 4.3.7. As principais conclusões são:

- Quando  $r = 0$  o vício é sempre maior no modelo desagregado.
- O vício no modelo desagregado pode ser muito maior (no máximo 8 em módulo para ambos os valores de  $r$ ) quando  $\theta$  está perto de  $-1$ . No modelo agregado o valor máximo é igual a  $(-0,283)^{[r/4]}1,283$ , que neste caso é igual a  $1,283$ .
- O vício no modelo desagregado pode ser menor que no modelo agregado quando  $r = 1$  e  $\theta$  pertence ao intervalo  $[-0,24; 0,5]$ . No entanto, nesse caso o vício é sempre menor que  $1,32$ .

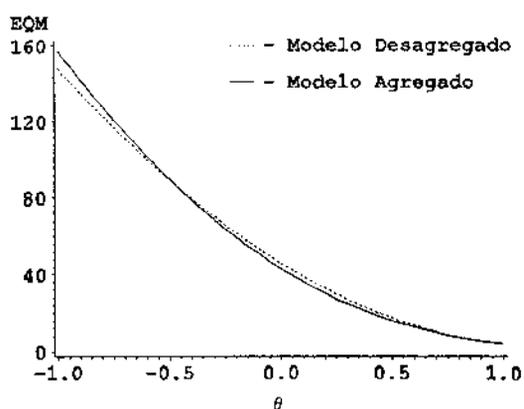
Assim, em geral, o vício de um AO é menor no modelo agregado do que no desagregado.

#### b) Efeito no EQM

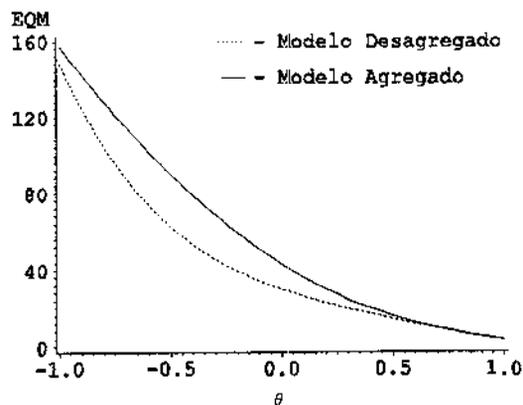
Os EQMs e as eficiências estão apresentados no Gráfico 4.3.8. De um modo geral, pode-se dizer que os EQMs aumentam quando  $\theta \rightarrow -1$ . Quando  $r = 0$ , os EQMs são muito próximos (a eficiência da previsão do modelo agregado com outlier varia entre 0,94 e 1,1). No caso de  $r = 1$ , verifica-se uma diferença maior nos EQMs (a eficiência com outlier varia entre 0,65 e 1).



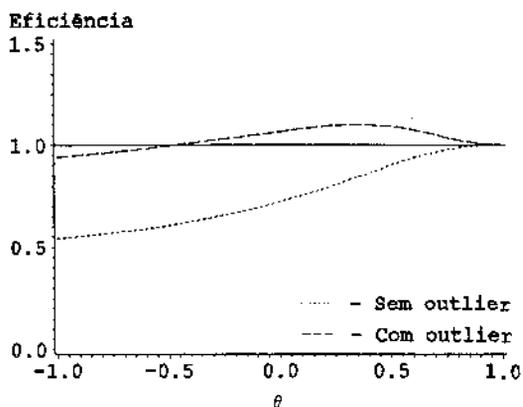
**Gráfico 4.3.7** - Vício causado por um AO, presente  $r$  observações antes da origem de previsão, nas previsões  $h$  ( $\geq 1$ ) passos à frente num modelo ARIMA(0,1,1) conhecido.



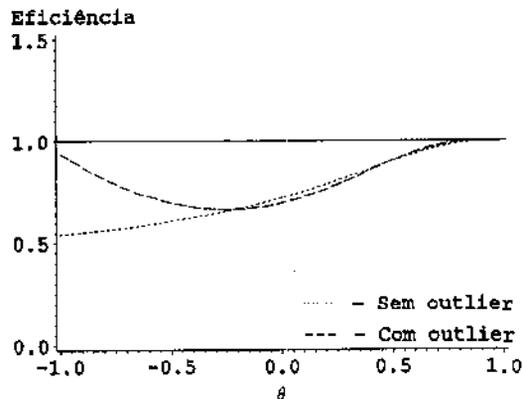
a) EQM,  $r = 0$



b) EQM,  $r = 1$



c) Eficiência,  $r = 0$



d) Eficiência,  $r = 1$

**Gráfico 4.3.8** - EQM e eficiência com um AO de tamanho 1,  $r$  posições antes da origem de previsão, no modelo ARIMA(0,1,1) conhecido, para previsão um passo à frente.

Comparando-se a eficiência na presença de AO de tamanho 1 na última observação pode-se ver que ela é praticamente igual a 1, sendo maior para todos os valores de  $\theta > -0,5$ . Como normalmente o vício é maior no modelo desagregado tem-se que um outlier de tamanho maior fará com que a eficiência do modelo agregado também passe a ser maior.

A presença de um outlier na penúltima observação ou aumenta a eficiência ou a diminui muito pouco. Como no modelo ARIMA(1,1,0), na maioria dos casos a presença do AO aumenta a eficiência da previsão do modelo agregado, principalmente quando  $r = 0$ . A melhora não é tão boa quando o outlier ocorre na penúltima observação ( $r = 1$ ).

## 2) Dois AOs de tamanho 1 de mesmo sinal

### a) Efeito na média

O vício causado pelos dois AOs de tamanho 1, de mesmo sinal, nas duas últimas observações é apresentado no Gráfico 4.3.9, para previsões  $h (\geq 1)$  passos à frente. No modelo desagregado o vício é maior do que no modelo agregado quando  $\theta > -0,6$ . O vício nos modelos desagregado e agregado não é muito grande em toda amplitude de  $\theta$  sendo o máximo de 4 e 2,6 para modelos desagregado e agregado, respectivamente. No modelo desagregado o vício tem uma forma quadrática, enquanto no modelo agregado o vício é decrescente em  $\theta$ .

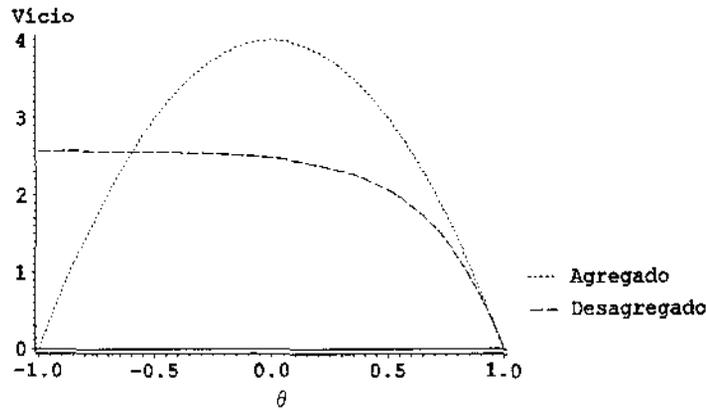
### b) Efeito no EQM

Pelo Gráfico 4.3.10 pode-se ver que a presença dos outliers, em geral, aumenta a eficiência. Quando isto não ocorre a perda de eficiência é muito pequena. A eficiência do modelo agregado passa a ser maior do que 1 para  $\theta > 0,05$ , mas a diferença, absoluta dos EQMs nestes casos é pequena.

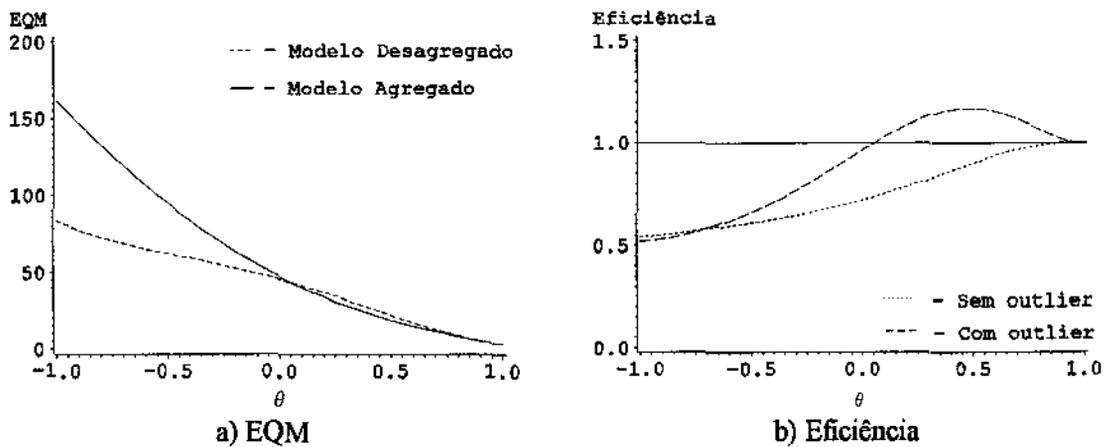
## 3) Dois AOs de tamanho um de sinais trocados

### a) Efeito na média

No Gráfico 4.3.11 é apresentado o vício causado pelos dois AOs de tamanho 1 e -1, na penúltima e última observação, respectivamente. O vício no modelo agregado é nulo, como foi comentado na subseção anterior, pois os outliers se cancelam na última observação da série agregada. No modelo desagregado o vício em módulo é decrescente em  $\theta$  e pode ser bastante grande para  $\theta$  negativo, chegando, em módulo, a 16 quando  $\theta = -1$ .



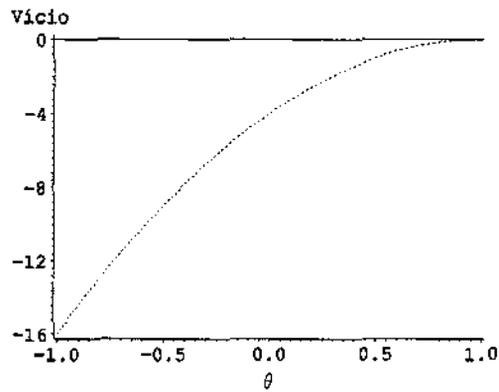
**Gráfico 4.3.9** - Vício causado por dois AOs de tamanho 1, de mesmo sinal, nas previsões  $h$  ( $\geq 1$ ) passos à frente num modelo ARIMA(0,1,1) conhecido.



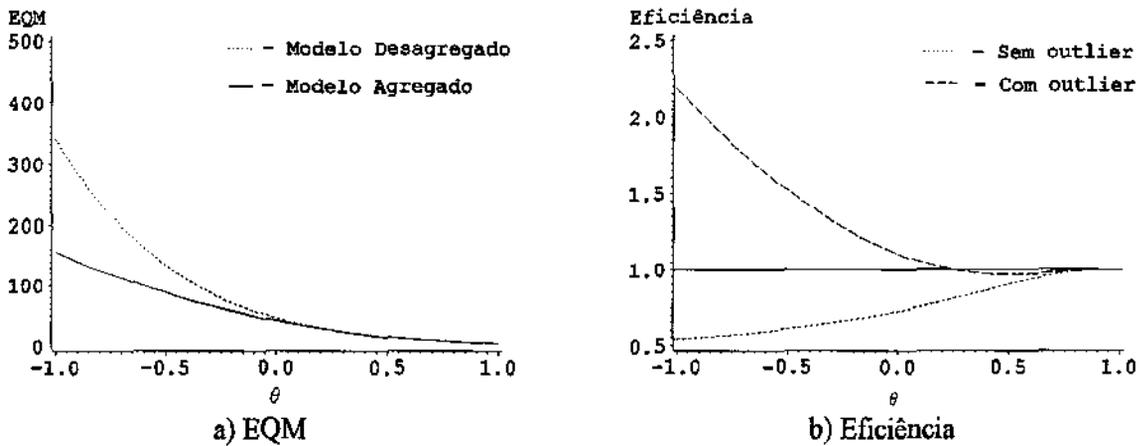
**Gráfico 4.3.10** - EQM e eficiência com dois AOs de tamanho 1, de mesmo sinal, no modelo ARIMA(0,1,1) conhecido, na previsão um passo à frente.

#### b) Efeito no EQM

No Gráfico 4.3.12 são apresentados os EQMs e as eficiências para previsão um passo à frente com os dois AOs, com sinais trocados. Pelos valores das eficiências pode-se notar que a eficiência do modelo agregado é maior do que 1 para  $\theta < 0,25$  e como nos casos anteriores a eficiência aumenta bastante na presença de outliers. A eficiência quando menor do que 1 é próximo a 1, mas pode chegar a ser maior do que 2 quando  $\theta$  está próximo de -1.



**Gráfico 4.3.11** - Vício causado por dois AOs de tamanho um, de sinais trocados, no modelo ARIMA(0,1,1) agregado conhecido, para previsões  $h$  ( $h > 0$ ) passos à frente.



**Gráfico 4.3.12** - EQM e eficiência com dois AOs de tamanho 1 de sinais trocados nas duas últimas observações no modelo ARIMA(0,1,1) conhecido, na previsão um passo à frente.

### 4.3.2 - Modelos desconhecidos

Na prática, geralmente os modelos precisam ser estimados e desta forma, os outliers podem afetar a previsão através dos efeitos de estimação e identificação. Hotta (1993) comparou o efeito de um AO na estimativa dos parâmetros nos modelos desagregados e agregados. Os resultados, num processo ARIMA(0,1,1), mostraram que, geralmente, o efeito é maior no modelo desagregado, pelo menos em termos do aumento da estimativa da variância da inovação. Por outro lado, se o processo ARIMA(0,1,1) for gerado pelo modelo estrutural (quando  $\theta$  é positivo, como foi visto no Capítulo 1), o modelo agregado é mais afetado pelo AO. Neste modelo o tamanho de amostra é menor ( $N = n/4$  ao invés de  $n$ ), o que interfere no sentido de aumentar a influência dos outliers na estimativa; porém o outlier é menor, em termos do desvio padrão da inovação, diminuindo o efeito. O balanço destas influências opostas dirá se o efeito final será maior ou menor na previsão utilizando o modelo agregado e desagregado.

Adicionalmente, é necessário considerar que as ordens do modelo são desconhecidas, exceto a ordem de não estacionariedade. Como foi mencionado na subseção 4.2.3, uma forma de contornar a subjetividade na identificação do modelo é aproximar o processo diferenciado por um AR(P), onde P é selecionado pelo critério BIC. Utilizando este processo de identificação pode-se estudar os efeitos dos outliers na previsão de séries agregadas através de simulações, quando os modelos são desconhecidos.

A seguir é apresentado o algoritmo do programa de simulações utilizado para obter os resultados apresentados posteriormente. O programa foi desenvolvido na linguagem do Procedimento IML do pacote estatístico SAS (*Statistical Analysis System*) e é apresentado no **Apêndice 4**.

O algoritmo consiste basicamente de quatro passos: geração, identificação, teste de detecção de outlier e previsão. Para verificar melhor o efeito da utilização do teste de detecção automático de outliers foram utilizadas simulações com e sem teste de detecção de outliers. Os outliers foram colocados apenas na penúltima e/ou na última observação da série desagregada, sendo aplicados testes automáticos de detecção apenas nessas observações onde foram inseridos os outliers. Seguindo as sugestões apresentadas na subseção 2.1.5, foram utilizados apenas os testes para outlier IO e AO, já que eles foram introduzidos nas duas últimas observações; no caso de inserção na última observação foi realizado apenas o teste de AO. Desta forma, no modelo agregado não existe possibilidade de um outlier ser identificado incorretamente, enquanto no modelo desagregado, esta possibilidade existe apenas quando ocorre na penúltima observação. No resumo dos principais passos dado a seguir, observe que estimação e identificação aparecem em dois passos.

## O algoritmo do programa de simulação

### Passo 1: Geração das observações

Gera-se  $n_0+n+m*h$  observações de um processo ARIMA, onde as primeiras  $n_0$  observações são eliminadas para reduzir a influência dos valores iniciais utilizados. As previsões são baseadas nas  $n$  observações seguintes, e as últimas  $m*h$  observações são utilizadas para estimar o efeito do outlier, sendo que  $m$  é o período de agregação e  $h$  o número de previsões à frente das séries agregadas. O ruído branco foi gerado através de uma distribuição i.i.d. normal  $N(0, \sigma_a^2)$ , onde neste caso  $\sigma_a^2 = 1$ . Tomou-se o valor inicial da série igual a zero. Para o caso de IMA(1,1) a inovação inicial é igual a zero.

Os modelos considerados neste trabalho são IMA(1,1) e ARI(1,1), sendo a ordem de não estacionariedade tratada como conhecida.

### Passo 2: Identificação e estimação

Utilizando o mesmo procedimento de Lütkepohl (1986), os processos diferenciados ARMA(p,q) são aproximados por um processo AR(P), onde a ordem P é estimada pelo critério BIC. O cálculo da estimação dos coeficientes do processo foi efetuado usando procedimento SWEEP (Goodnight, 1979), utilizando as últimas  $n-P_m$  observações (onde  $P_m$  é a ordem máxima estipulada) dentre as  $n$  geradas no passo 1. Obtendo-se os parâmetros do modelo diferenciado,  $(\phi_1, \phi_2, \dots, \phi_p)'$ , pode-se obter os parâmetros do modelo original (não diferenciado) dados por

$$\pi_i = -\phi_{i-1} + \phi_i, \quad i = 1, \dots, P,$$

onde  $\pi_0 = -1$

$$e \quad \pi_{P+1} = -\phi_P.$$

Os valores máximos de ordens de AR(P) estabelecidos para estas simulações foram 5 para modelo desagregado e 4 para modelo agregado.

### Passo 3: Detecção de outliers

Utilizando o modelo ajustado no passo 2, são realizados os testes de detecção de outliers (neste trabalho, as posições dos outliers foram consideradas conhecidas). Os outliers foram inseridos na penúltima e/ou última observação da série, e os tipos considerados foram AO e IO de tamanho unitário.

Na aplicação dos testes de detecção dadas pela equação (2.1.6), são utilizadas as

inovações estimadas pelas observações originais. A variância da inovação foi estimada pela variância amostral, sem considerar os resíduos das posições que ocorre(m) o(s) outlier(s) para torná-la mais robusto.

Considerando dois outliers, eles são testados ao mesmo tempo, mas individualmente e no caso de ambos serem significativos, eles são incorporados no modelo para reestimá-los.

O outlier detectado na última observação é considerado um AO, já que não é possível identificar o tipo.

Chen e Liu (1993b) propõem o uso de ponto de corte  $C$  entre 2,0 e 2,5 quando o outlier ocorre perto ou no final da série, assim, o escolhido para as simulações foi  $C = 2,0$ .

#### Passo 4: Identificação, estimação e previsão

a) Caso algum outlier seja detectado no passo anterior, reestimam-se a ordem do processo, os parâmetros e a magnitude do outlier. Para incorporar a presença do outlier no modelo foi acrescentada uma variável indicadora para cada outlier detectado. Essa variável indicadora é um vetor de zeros com valor um na posição do outlier detectado, e o seu coeficiente estimado é a magnitude estimada do outlier. A reestimação dos parâmetros (autoregressivos e  $a(s)$  magnitude(s) do(s) outlier(s)) é a mesma do passo 2, utilizando o procedimento SWEEP. No caso de ser detectado um outlier AO na penúltima observação, o valor do vetor da variável indicadora na última posição é  $1+\phi_1$ , nesse caso a estimação dos parâmetros se torna não linear. Para realizar essa estimação foi implementado o procedimento de Newton Raphson (ver, por exemplo, Bates e Watts, 1988, p. 40-42), utilizando como valores iniciais os valores verdadeiros. Neste passo a ordem do modelo AR(P) é estimada incorporando o(s) outlier(s) detectado(s). Após a identificação e estimação são realizadas as previsões.

b) Caso nenhum outlier tenha sido detectado no passo 3, não é necessário identificar e estimar o modelo novamente, passa-se diretamente à previsão.

Esse mesmo procedimento se repete no caso de modelos agregados. Como nestes modelos o outlier está sempre presente na última observação, elimina-se a necessidade de estimação não linear, que era necessária nos modelos desagregados, quando um AO era detectado na penúltima observação.

Esse algoritmo pode ser modificado de duas formas:

- após o passo 3 se volte ao passo 2 (incorporando o(s) outlier(s) detectado(s) no

passo 3) até que não se detecte mais nenhum outlier;

- após o passo 4 volte ao passo 3 até não detectar mais nenhum outlier.

No caso de múltiplos outliers, Chang *et alii* (1983), por experiências empíricas, propõem detectar um outlier de cada vez. Neste trabalho, isso não foi considerado, principalmente por que só foram testadas as observações onde realmente foram colocados os outliers.

Foram realizadas 1000 replicações de séries de tamanho 100 para cada processo ARIMA(1,1,0) e ARIMA(0,1,1) com os parâmetros assumindo os seguintes valores -0,8; -0,5; -0,2; 0,2; 0,5 e 0,8. A série contaminada tem um ou dois outliers na última e/ou penúltima observação. Na presença de um outlier a magnitude considerada foi 1 e 2, e na presença de dois outliers ambos de tamanho 1 de mesmo sinal ou com sinais trocados, sendo o primeiro igual a 1 e o segundo -1.

Nas simulações são analisados os efeitos nas previsões em termos de vício, EQM e eficiência. O vício e EQM neste caso não são definidos pelas equações (4.3.4) e (4.3.5), respectivamente, pois adiciona-se o efeito do outlier na especificação e na estimação do modelo. O vício será estimado como a média dos erros de previsão. A eficiência é definida como

$$ef(r,h) = EQM^d(r,h)/EQM^a(r,h),$$

onde  $EQM^d(r,h)$  é o EQM de previsão  $h$  passos à frente de valores agregados através do modelo desagregado com um outlier  $r$  observações antes da origem de previsão e  $EQM^a(r,h)$  é a previsão  $h$  passos à frente do modelo agregado com um outlier  $[r/m]$  observações antes da origem de previsão.

Para facilitar a análise, os resultados foram divididos em 4 casos:

- um outlier de tamanho  $\omega_A = 1$  na penúltima ( $r = 1$ ) ou última ( $r = 0$ ) observação;
- um outlier de tamanho  $\omega_A = 2$  na penúltima ( $r = 1$ ) ou última ( $r = 0$ ) observação;
- dois outliers de tamanho  $\omega_{A1} = \omega_{A2} = 1$  nas duas últimas observações;
- dois outliers de tamanho  $\omega_{A1} = 1$  e  $\omega_{A2} = -1$  na penúltima e última observação, respectivamente,

sendo que em cada caso são apresentados os resultados com e sem teste de detecção de outliers separadamente.

A seguir são apresentados os resultados obtidos através dessas simulações, para previsão um passo à frente. Para simplificar, os modelos ARIMA(1,1,0) e ARIMA(0,1,1) serão denominados apenas como ARI e IMA, respectivamente. Antes, é apresentada a escolha das ordens dos modelos aproximados.

#### 4.3.2.1 - Aproximação pelo modelo ARI(P,1)

Foi realizado um estudo das ordens dos processos diferenciados estimados considerando as séries desagregadas e agregadas sem e com outliers. As séries com outliers tinham um outlier na penúltima e/ou última observação de tamanho 1 e/ou -1 do tipo AO ou IO, como nos 4 casos (i), ..., (iv) apresentados anteriormente. As ordens máximas estabelecidas para o estudo foram de 8 e 6, para os modelos desagregados e agregados, respectivamente.

As ordens estimadas das séries com outlier(s) e sem outlier(s) diferem em menos de 1% dos casos. Na presença do outlier, a ordem estimada tanto pode ser menor como maior que a ordem da série sem outlier.

Quando as séries são geradas através do processo ARI a ordem das séries desagregadas diferenciadas foi corretamente determinada (ordem 1) para aproximadamente 97% das séries e para mais de 99% das séries a ordem foi de até 2 (ver Tabela 4.3.1). Todas as ordens foram menores ou iguais a 5. Na série agregada mais de 98% dos casos são no máximo de ordem 4. Na maioria delas a ordem é 1 (entre 67 e 87% do total, onde a porcentagem diminui conforme  $\phi \rightarrow +1$ ). Ver Tabela 4.3.2.

No caso das séries serem geradas pelo processo IMA, a ordem do modelo aproximado depende do valor de  $|\theta|$ . Quanto menor o valor de  $|\theta|$  menor a ordem P (ver Tabela 4.3.3). No caso do modelo agregado, como esperado, mais de 98,4% dos casos têm ordem até 4, predominando a ordem igual 1 (mais de 85% do total, ver Tabela 4.3.4). Assim, de um modo geral, estipular a ordem máxima de 5 no modelo desagregado não chega a ser uma restrição forte.

**Tabela 4.3.1 - Porcentagem acumulada média da frequência da ordem do processo ajustado em processo ARI(1,1) desagregado**

$\phi$	ordem			
	1	2	3	$\geq 4$
-0,8	97,2	99,7	100	
-0,5	97,6	99,7	99,9	100
-0,2	97,3	99,7	100	
0,2	97,5	99,7	100	
0,5	97,4	99,6	99,9	100
0,8	97,5	99,6	100	

**Tabela 4.3.2 - Porcentagem acumulada média da frequência da ordem do processo ajustado em processo ARI(1,1) agregado.**

$\phi$	ordem					
	1	2	3	4	5	$>5$
-0,8	88	95,9	97,8	98,9	99,3	100
-0,5	87,1	95,5	97,6	98,9	99,6	100
-0,2	87,2	95,7	97,7	99,0	99,6	100
0,2	86,2	95,5	94,9	97,7	98,8	100
0,5	82,3	94,5	97,4	98,4	99,4	100
0,8	66,3	92,0	96,3	97,9	99,3	100

**Tabela 4.3.3 - Porcentagem acumulada média da frequência da ordem do processo ajustado em processo IMA(1,1) desagregado**

$ \theta $	ordem						
	1	2	3	4	5	6	$> 6$
0,8	9,9	51,5	80,4	94,7	98,1	99,8	100
0,5	63,5	95,3	99,6	100			
0,2	95	99,7	100				

**Tabela 4.3.4 - Porcentagem acumulada média da frequência da ordem do processo ajustado em processo ARIMA(0,1,1) agregado.**

$\theta$	ordem					
	1	2	3	4	5	>5
-0,8	85,0	95,1	97,7	98,6	99,3	100
-0,5	86,2	94,8	97,5	98,9	99,2	100
-0,2	86,5	95,3	97,6	98,9	99,6	100
0,2	88,8	95,3	98,1	99,2	99,7	100
0,5	88,3	95,0	97,2	98,6	99,8	100
0,8	81,2	94,4	97,5	98,9	99,6	100

#### 4.3.2.2 - Um AO de tamanho 1

São apresentadas aqui as análises dos resultados de simulação nos modelos ARI e IMA desagregado e agregado, no caso de apenas um outlier na série, podendo esse outlier estar presente na penúltima ou última observação da série. O teste aplicado para detecção de outliers foi de RV, apresentado na seção 2.1. Os resultados sem e com teste de detecção são apresentados separadamente.

##### a) sem o teste de detecção de outlier

Nos Gráficos 4.3.13 (a) e (c) são apresentados os vícios considerando os modelos conhecido e desconhecido, sem teste de detecção. Para modelos conhecidos os resultados são exatos enquanto que para modelos desconhecidos eles só são obtidos através de simulações com 1000 replicações. Os vícios dos modelos conhecidos e desconhecidos são muito próximos, principalmente nos modelos ARI, tanto no modelo agregado como no desagregado. Isso mostra que o processo de identificação e de estimação dos parâmetros têm pouca influência no vício. Já no caso de IMA existe uma diminuição (aumento) do vício ao se considerar o modelo desagregado desconhecido, quando  $\theta \rightarrow -1$  (+1). Isso pode ter sido causado pela aproximação de um modelo IMA(1,1) por um modelo ARI(P,1) e/ou pelo efeito do outlier na estimativa dos parâmetros.

Os EQMs são apresentados no Gráfico 4.3.14. Também para o EQM os valores simulados são bastante próximos de e sempre menores que os valores encontrados para modelos conhecidos, com exceção no modelo IMA, quando  $\theta \rightarrow +1$ . No limite a eficiência estimada (Gráfico 4.3.16) é consideravelmente maior do que a eficiência do modelo conhecido, chegando em torno de 1,4 quando  $\theta = 0,8$  para  $r = 0$  ou 1, quando o valor da eficiência para

modelo conhecido era em torno de 1.

b) com o teste de detecção de outlier

Na Tabela A5.1<sup>1</sup> são apresentadas as porcentagens de outliers detectados nos testes de detecção. No modelo ARI desagregado, com o outlier na penúltima observação ( $r = 1$ ), a porcentagem de detecção do outlier e de identificação correta são crescentes em  $\phi$ , onde a porcentagem de detecção varia entre 15% e 53,6%, para  $\phi = -0,8$  e  $0,8$ , respectivamente. A porcentagem de identificação correta, dentre os que foram detectados, foi de 52% a 88,4%, para  $\phi = -0,8$  e  $0,8$ , respectivamente, i.e., a porcentagem de casos com identificação correta varia entre 7,8% e 47,4%, para  $\phi = -0,8$  e  $0,8$ , respectivamente.

Quando o outlier ocorre na última observação ( $r = 0$ ), no modelo ARI desagregado, a porcentagem de outliers detectados não se altera com o valor do parâmetro do modelo, ficando em torno de 15%. No modelo agregado, a porcentagem de outliers detectados é decrescente em  $\phi$ , variando de 4% a 1,5% (para  $\phi = -0,8$  e  $0,8$ , respectivamente) para outlier presente na penúltima ou última observação da série ( $r = 1$  e  $0$ ).

No modelo IMA desagregado e agregado a relação entre a proporção de outliers detectados e o valor de  $\theta$  é inversa ao do modelo ARI. A proporção de outliers detectados aumenta conforme diminui o valor de  $\theta$  e os valores das proporções são muito próximos.

O vício, com a aplicação do teste, em média diminui bastante, tanto no modelo ARI como no IMA, como pode ser visto no Gráfico 4.3.13, com maior diminuição absoluta no modelo desagregado quando o outlier ocorre na penúltima observação ( $r = 1$ ). No entanto, os EQMs de previsão, principalmente no modelo agregado, aumentam consideravelmente (ver Gráfico 4.3.14). Esse resultado pode ser explicado da seguinte forma:

quando um outlier é detectado, as estatísticas do teste são todas estimadas como maiores do que 2, pois o ponto de corte considerado da estatística é 2. A estatística do teste de detecção tem a seguinte relação com o outlier estimado:

$$\hat{\lambda}_A = \hat{\omega}_A / (1 + \hat{\pi}_1^2)^{1/2} \hat{\sigma}_a \quad \text{se } k = n - 1 \text{ e}$$

$$\hat{\lambda}_A = \hat{\omega}_A / \hat{\sigma}_a \quad \text{se } k = n.$$

onde  $\hat{\omega}_A$  e  $\hat{\sigma}_a$  são os estimadores de  $\omega_A$  (magnitude do outlier) e  $\sigma_a$  (desvio padrão da inovação), respectivamente. Desta forma, detectar um outlier implica em

<sup>1</sup> As tabelas estão apresentadas no Apêndice 5

$$\hat{\lambda}_A > 2 \Rightarrow \hat{\omega}_A > 2(1 + \hat{\pi}_1^2)^{1/2} \hat{\sigma}_a \quad \text{se } k = n - 1,$$

$$\hat{\lambda}_A > 2 \Rightarrow \hat{\omega}_A > 2 \hat{\sigma}_a \quad \text{se } k = n.$$

Assim, quando detectado no modelo desagregado, o outlier é geralmente estimado em torno de 2 ou maior do que 2, pois o valor da variância da inovação estimada é em torno de 1 e  $(1 + \hat{\pi}_1^2)^{1/2} \geq 1$ , enquanto o valor verdadeiro do outlier é igual a 1. Embora os outliers e os parâmetros sejam reestimados após a detecção, as magnitudes dos outliers se mantêm, em geral, maior que o valor real. No caso do outlier ser reestimado por SWEEP (quando  $r = 0$  ou  $r = 1$  e o outlier for identificado como IO) a magnitude do outlier ficou em torno de 2,5 ou mais, dependendo do valor do parâmetro do modelo. Quando o outlier reestimado pelo método não linear (quando  $r = 1$  e o outlier é identificado como AO) ficou em torno de 1,5. Quando o outlier, identificado como AO, é super estimado (em módulo) o vício de previsão troca de sinal, pois ele é proporcional à diferença dos outliers real e estimado. Assim, em média, o vício diminui bastante e o EQM aumenta muito. Quando o outlier é identificado como IO a previsão é muito pouco afetada.

No modelo agregado o erro é maior, já que o desvio padrão das inovações é sempre maior do que 1 nos modelos ARI e IMA desagregados, chegando a ser maior que 20 (no modelo ARI com  $\phi \rightarrow +1$ ), i.e., o outlier estimado do modelo agregado pode ser maior que 40. Nos Gráficos 4.2.1 e 4.2.4, são apresentados os desvios padrão, para modelos desagregados ARI e IMA, respectivamente.

Uma vez que com a aplicação do teste de detecção de outlier o EQM aumenta proporcionalmente mais no modelo agregado, tem-se uma diminuição da eficiência tanto para os modelos ARI como no IMA para  $r = 0$  e 1, como pode ser confirmado no Gráfico 4.3.16.

#### 4.3.2.3 - Um AO de tamanho 2

O objetivo, ao dobrar o tamanho do outlier em relação ao caso anterior, é verificar como os efeitos da especificação, estimação e teste se modificam.

a) sem o teste de detecção de outlier

Nos Gráficos 4.3.13 (b) e (d) são apresentados os vícios para modelos conhecidos e simulados causados por um AO na previsão de valores agregados em modelos desagregados e agregados. Os vícios simulados sem teste de detecção foram muito próximos dos vícios para

modelos conhecidos, principalmente para os modelos ARI, continuando válidos os comentários anteriores. Como foi mencionado no modelo IMA, os vícios simulados se distanciam dos vícios dos modelos conhecidos quando  $|\theta| \rightarrow 1$ .

Analisando os Gráficos 4.3.15 os EQMs simulados são, em geral, bastante próximos dos EQM dos modelos conhecidos. Quando o outlier ocorre na última observação, a eficiência fica entre 1,5 e 2, tanto para o modelo ARI como IMA (Gráfico 4.3.17). No entanto, quando o outlier ocorre na penúltima observação, a eficiência depende fortemente dos valores dos coeficientes, não havendo dominância do modelo agregado ou desagregado.

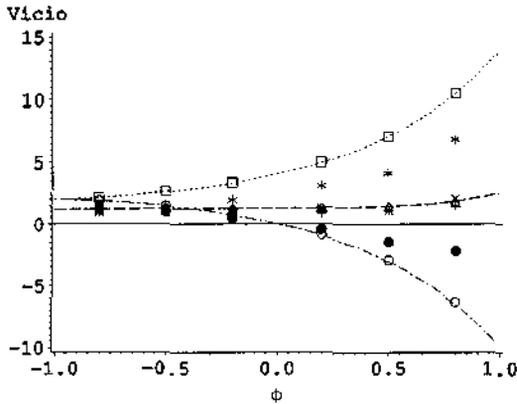
b) com o teste de detecção de outlier

No modelo desagregado, as porcentagens de outliers detectados são apresentadas na Tabela A5.1. Os comentários da relação do valor dos parâmetros com a porcentagem dos outliers detectados são análogos ao caso em que o tamanho do outlier era igual a 1, apenas aumentando as porcentagens de detecção de outlier. O aumento, muitas vezes foi maior que 100%, variando de 83% a 255% (por exemplo, de 15% para 53,3% no modelo ARI com  $\phi = -0,8$ ), mas a porcentagem de identificação correta dentre os detectados não muda muito, apenas aumentando um pouco. No caso de modelos agregados, o aumento percentual de outliers detectados é em geral menor, e a proporção de outliers detectados continua bastante baixa.

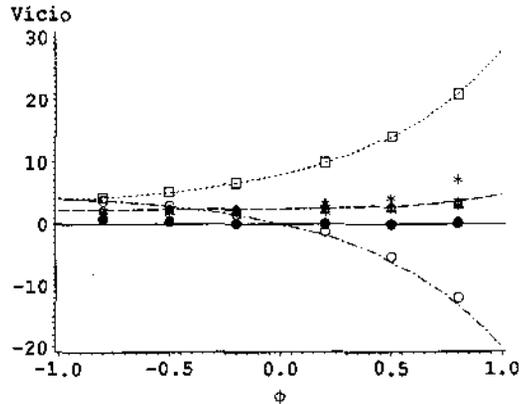
O Gráfico 4.3.13 mostra que a utilização do teste diminui bastante o vício do modelo desagregado, principalmente quando  $r = 1$ . Neste caso, o vício é praticamente zero, chegando a ser menor que o do modelo agregado. No caso de  $r = 0$ , o vício diminui bastante também, mas ainda continua um pouco maior do que o do modelo agregado.

Analisando-se agora os EQMs pelo Gráfico 4.3.15, verifica-se que com a aplicação do teste, em ambos os modelos desagregados, ARI e IMA, o EQM diminui bastante para  $r = 0$  e 1. Esse resultado é coerente, pois o vício praticamente foi zerado e as estimativas dos outliers detectados são agora mais próximas do valor verdadeiro do que no caso anterior, quando o tamanho do outlier era de magnitude 1. No modelo agregado, os EQMs sempre aumentam com a aplicação do teste, porque neste caso os outliers são estimados com valores muito maiores que o valor real.

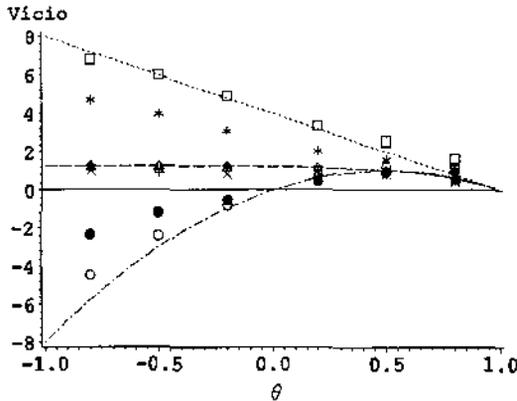
No caso de  $r = 0$ , a eficiência que era em torno de 1,5 e 2 com as aplicações dos testes fica aproximadamente igual a 1 (Gráfico 4.3.17). Para  $r = 1$  a eficiência diminui e em geral fica menor do que sem outlier. Assim, como no caso anterior, a aplicação do teste diminui a



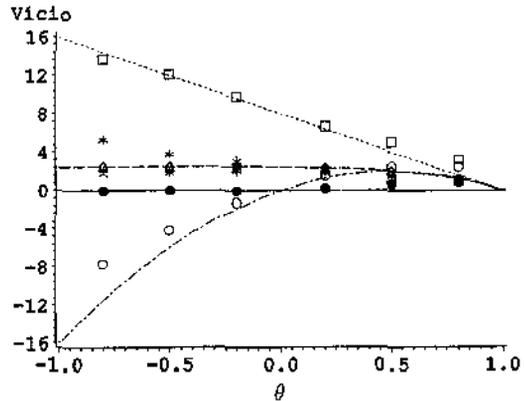
a) Vício de um AO de tamanho 1 no modelo ARIMA(1,1,0).



b) Vício de um AO de tamanho 2 no modelo ARIMA(1,1,0).



c) Vício de um AO de tamanho 1 no modelo ARIMA(0,1,1).



d) Vício de um AO de tamanho 2 no modelo ARIMA(0,1,1).

**Legenda:**

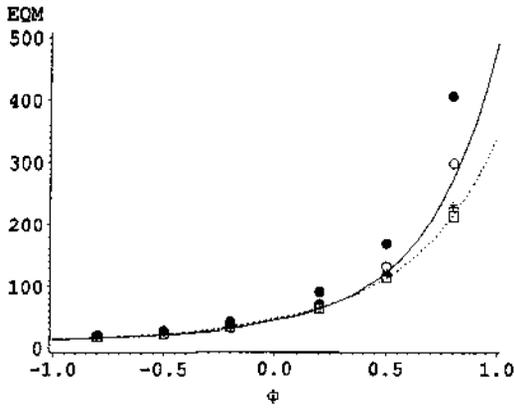
**Modelo desagregado**

- ..... - vício para modelo conhecido,  $r = 0$
- .-.- - vício para modelo conhecido,  $r = 1$
- - vício simulado sem teste,  $r = 0$
- \* - vício simulado com teste,  $r = 0$
- - vício simulado sem teste,  $r = 1$
- - vício simulado com teste,  $r = 1$

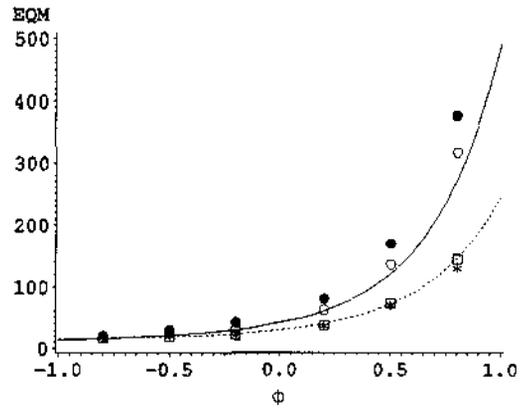
**Modelo agregado**

- - vício para modelo conhecido,  $r = 0$  a 3
- Δ - vício simulado sem teste,  $r = 0$
- + - vício simulado com teste,  $r = 0$
- ◇ - vício simulado sem teste,  $r = 1$
- × - vício simulado com teste,  $r = 1$

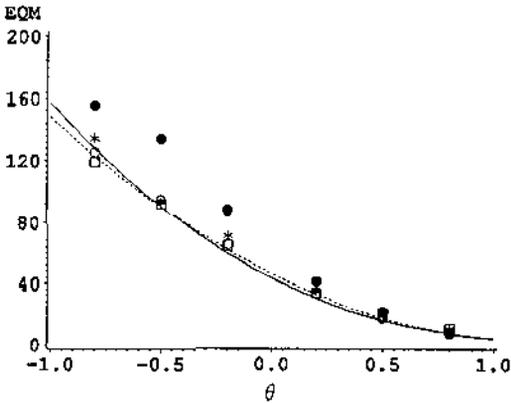
**Gráfico 4.3.13** - Vício de um AO na penúltima ( $r = 1$ ) ou última ( $r = 0$ ) observação de uma série com 100 observações, em 1000 replicações, com e sem teste de detecção de outlier, para previsão um passo à frente.



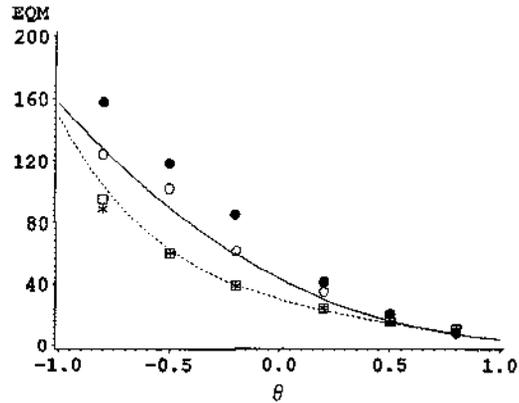
a) ARIMA(1,1,0), r = 0



b) ARIMA(1,1,0), r = 1



c) ARIMA(0,1,1), r = 0



d) ARIMA(0,1,1), r = 1

**Legenda:**

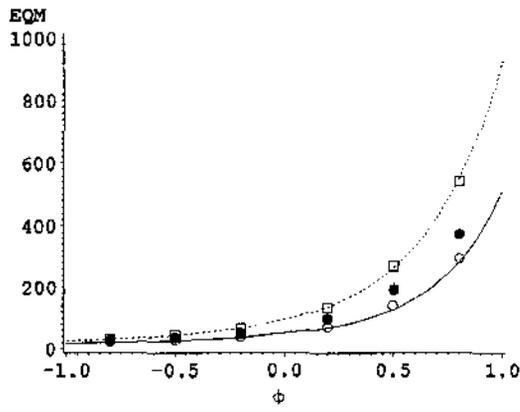
**Modelo desagregado**

- ..... - EQM para modelo conhecido
- - EQM simulado sem teste
- \* - EQM simulado com teste

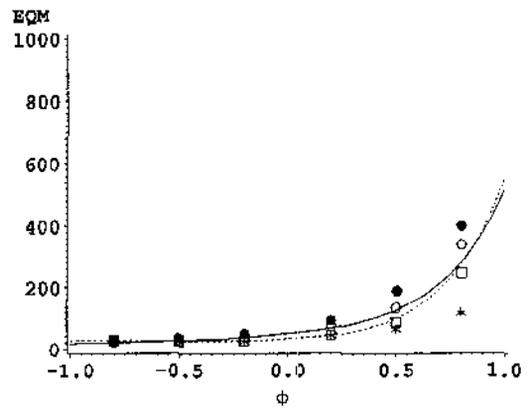
**Modelo agregado**

- - EQM para modelo conhecido
- - EQM simulado sem teste
- - EQM simulado com teste

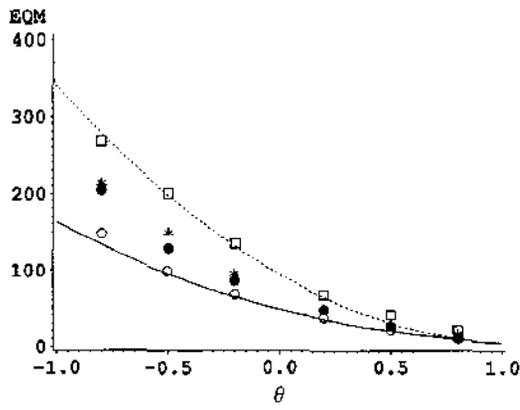
**Gráfico 4.3.14** - EQM com um AO de tamanho 1 r posições antes da origem de previsão, com e sem teste de detecção de outlier, para previsão um passo à frente. Séries de tamanho 100 com 1000 replicações.



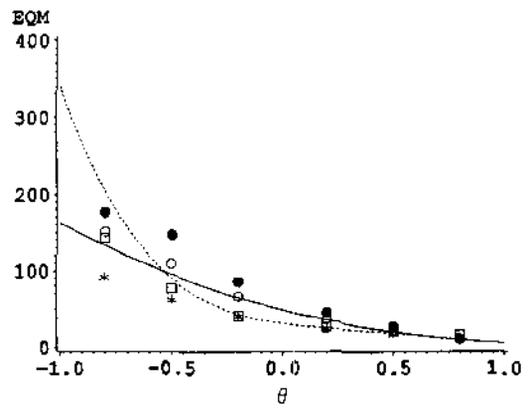
a) ARIMA(1,1,0), r = 0



b) ARIMA(1,1,0), r = 1



c) ARIMA(0,1,1), r = 0



d) ARIMA(0,1,1), r = 1

**Legenda:**

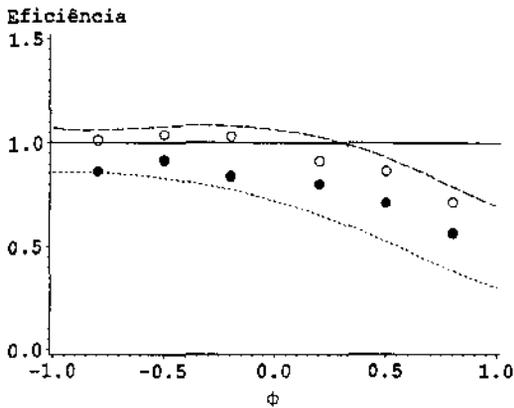
**Modelo desagregado**

- ..... - EQM para modelo conhecido
- - EQM simulado sem teste
- \* - EQM simulado com teste

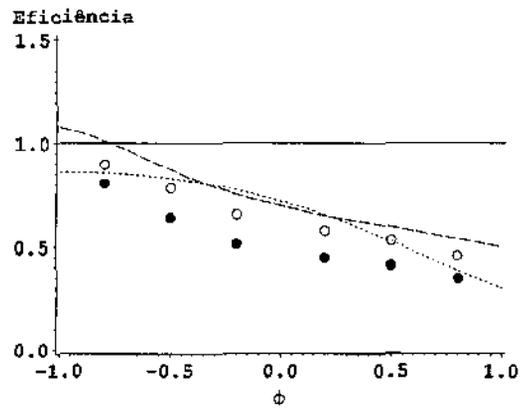
**Modelo agregado**

- - EQM para modelo conhecido
- - EQM simulado sem teste
- - EQM simulado com teste

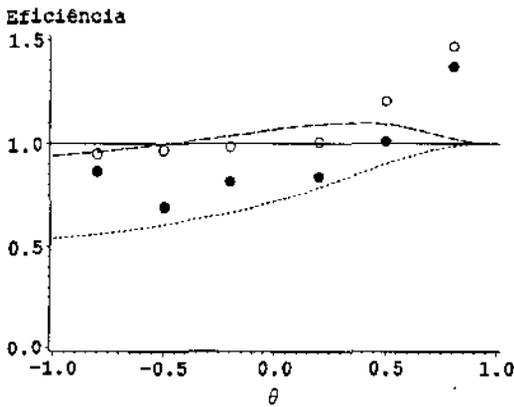
**Gráfico 4.3.15** - EQM com um AO de tamanho 2 presente r posições antes da origem de previsão, com e sem teste de detecção de outlier, para previsão um passo à frente. Séries de tamanho 100 com 1000 replicações.



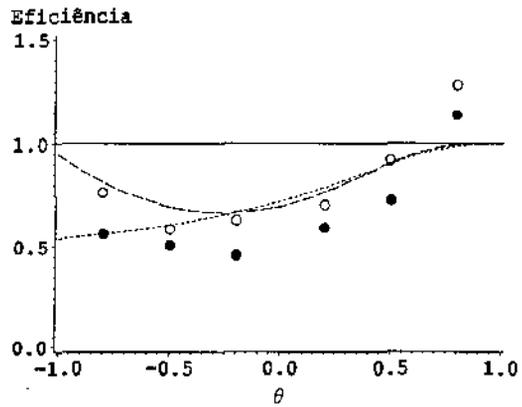
a) ARIMA(1,1,0), r = 0



b) ARIMA(1,1,0), r = 1



c) ARIMA(0,1,1), r = 0

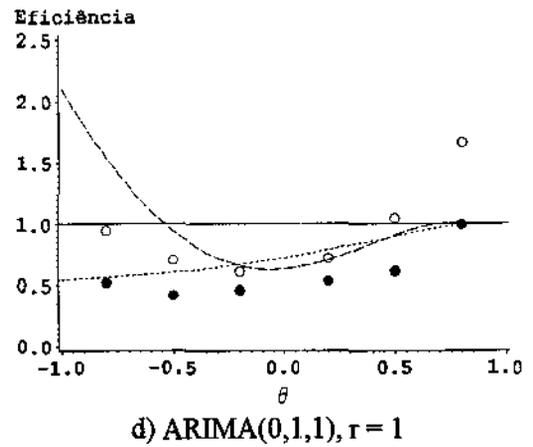
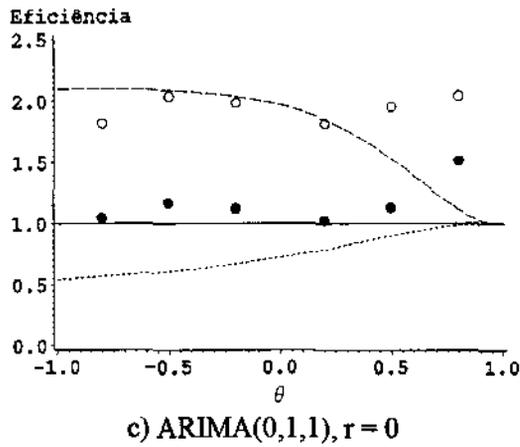
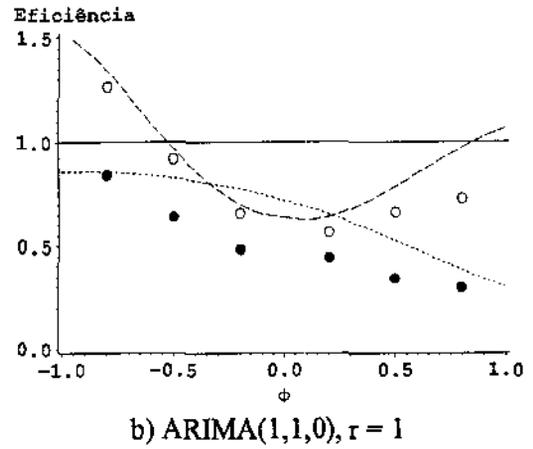
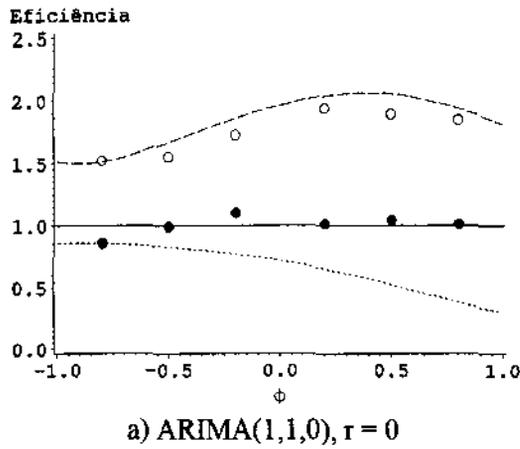


d) ARIMA(0,1,1), r = 1

**Legenda:**

- .... Eficiência considerando os modelos conhecidos, sem outlier
- Eficiência considerando os modelos conhecidos, com outlier
- Eficiência considerando os modelos desconhecidos, sem teste
- Eficiência considerando os modelos desconhecidos, com teste

**Gráfico 4.3.16** - Eficiência de previsão de modelo agregado, considerando o modelo conhecido e desconhecido, com e sem AO de tamanho 1, para previsão um passo à frente. Séries de tamanho 100 em 1000 replicações.



**Legenda:**

- ..... Eficiência considerando os modelos conhecidos, sem outlier
- Eficiência considerando os modelos conhecidos, com outlier
- Eficiência considerando os modelos desconhecidos, sem teste
- Eficiência considerando os modelos desconhecidos, com teste

**Gráfico 4.3.17** - Eficiência de previsão de modelo agregado, considerando o modelo conhecido e desconhecido, com e sem AO de tamanho 2, para previsão um passo à frente. Séries de tamanho 100 com 1000 replicações.

#### 4.3.2.4 - Dois AOs de tamanho 1 de mesmo sinal

São apresentadas aqui as análises dos resultados de simulação nos modelos ARI e IMA, no caso de dois AOs de tamanho 1, de mesmo sinal, presentes nas duas últimas observações da série. Um AO no modelo desagregado na penúltima observação pode ser identificado como AO ou IO, enquanto um AO na última observação pode ser identificado apenas como AO, por isso outliers detectados na última e penúltima observação têm diferentes efeitos na média e no EQM.

##### a) sem teste de detecção de outlier

Na presença de dois AOs, o vício dos modelos ARI conhecido e desconhecido são muito próximos, tanto no modelo desagregado como no agregado (Gráfico 4.3.18 (a)). No entanto, no modelo IMA, como no caso de apenas um AO, os resultados simulados de modelos desconhecidos não são tão próximos dos vícios do modelo conhecido, principalmente quando  $|\theta| \rightarrow +1$  (Gráfico 4.3.18 (c)). Esse resultado pode ser explicado como erro de aproximação e/ou estimação quando o modelo ARI é utilizado para representar o modelo IMA e esse erro é maior quando  $|\theta| \rightarrow +1$ .

Analisando-se o EQM (ver os Gráficos 4.3.19 (a) e (c)) pode-se verificar que os resultados simulados, em geral, são bastante próximos dos EQMs do modelo conhecido. As eficiências simuladas, nos modelos ARI e IMA, são bem próximas das eficiências do modelo conhecido, com exceção no modelo IMA, quando  $\theta \rightarrow +1$ . Enquanto a eficiência considerando o modelo conhecido após  $\theta = 0,45$  decresce para um, a eficiência simulada cresce, chegando a 1,8 quando  $\theta = 0,8$  (ver Gráficos 4.3.20 (a) e (c)).

##### b) com teste de detecção de outlier

A porcentagem de casos onde os dois outliers são detectados corretamente é bastante baixa, tanto no modelo ARI como para IMA (Tabela A5.2). Para o primeiro modelo esta porcentagem é crescente em  $\phi$  variando de 0% quando  $\phi = -0,8$  para 8,3% para  $\phi = 0,8$ , enquanto para o segundo modelo ela é decrescente em  $\theta$ , diminuindo de 5,5% quando  $\theta = -0,8$  para 0,4% quando  $\theta = 0,8$ . No entanto, em ambos os modelos, entre aproximadamente 20% a 25% dos casos pelo menos um outlier é detectado. Nos dois modelos o caso mais frequente é não detectar o último outlier e detectar o outlier da penúltima observação incorretamente como IO (caso (IO,SO)), que ocorre em torno de 12% para modelos ARI e IMA. O erro menos frequente é detectar os dois outliers, mas classificar o penúltimo outlier como IO. Os outros dois casos de detecção de apenas um outlier como AO têm comportamentos diferentes para o

modelo ARI e IMA. Enquanto no modelo ARI a porcentagem de detecção na última observação diminui com  $\phi$  (de 2,0% quando  $\phi = -0,8$  para 9,0% quando  $\phi = 0,8$ ) a de detecção do penúltimo outlier aumenta com  $\phi$  (de 2,6% quando  $\phi = -0,8$  para 7,0% quando  $\phi = 0,8$ ), no modelo IMA ocorre inversamente em termos de  $\theta$ .

Analisando-se agora o vício da previsão após as reestimações (Gráficos 4.3.18 (a) e (c)), nos modelos desagregados, os vícios médios não mudam muito, diminuindo um pouco para ambos os modelos, ARI e IMA. Também nos modelos agregados os vícios médios diminuem, um pouco mais, em todos os casos.

A aplicação do teste de detecção de outlier aumenta o EQM nos modelos ARI e IMA agregados (Gráfico 4.3.19 (a) e (c)). Como nos outros casos, isto ocorre devido a magnitude do outlier estimado ser muito maior que o real, modificando completamente a estrutura de previsão ao "remover" o outlier. Nos modelos desagregados ARI e IMA não houve muita mudança no EQM.

O efeito da aplicação do teste de detecção de outlier na eficiência, no modelo ARI, é em geral inversamente proporcional a  $\phi$ , diminuindo a diferença com a eficiência calculada sem outlier, conforme  $\phi \rightarrow +1$ . No modelo IMA, a diminuição da eficiência é praticamente constante.

#### 4.3.2.5 - Dois AOs de tamanho 1 de sinais trocados

Na subseção anterior foi estudado o caso em que os outliers presentes são de mesmo sinal. Considere agora o caso em que dois AOs de tamanho 1 e -1 estão presentes na penúltima e última observação da série, respectivamente.

##### a) sem teste de detecção de outlier

Os vícios e EQMs estimados pela simulação estão bem próximos dos estimados e EQM em modelos conhecidos (ver os Gráficos 4.3.18 (b) e (d) e Gráficos 4.3.19 (b) e (d)). A única exceção é o modelo IMA, quando  $\theta \rightarrow -1$ , onde os valores dos vícios simulados são um pouco menores que o vício para modelo conhecido, provavelmente devido ao erro de aproximação de IMA para ARI. Desta forma, as conclusões do caso de apenas um outlier de tamanho 1 continuam válidas, i.e., os procedimentos de estimação e identificação têm pouca influência no vício e EQM. Os EQMs no modelo desagregado continuam maiores ou próximos do agregado para ambos os modelos ARI e IMA, assim a eficiência (Gráfico 4.3.20 (b) e (d)) diminui um pouco em relação à eficiência de modelo conhecido.

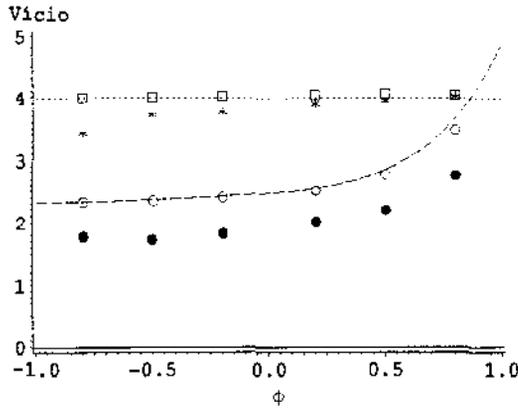
## b) com teste de detecção

Diferentemente do caso anterior de dois AOs de tamanho um, de mesmo sinal, a porcentagem de identificação correta nos dois outliers é alta, variando bastante com o valor dos parâmetros, tanto no modelo ARI como no IMA (ver Tabela A5.3). Para o primeiro modelo essa porcentagem é crescente em  $\phi$ , variando de 5,4% quando  $\phi = -0,8$  para 72,1% quando  $\phi = 0,8$ , enquanto que para o segundo modelo ela é decrescente em  $\theta$ , diminuindo de 60,7% quando  $\theta = -0,8$  para 7,9% quando  $\theta = 0,8$ . A porcentagem de detecção de pelo menos um outlier também varia com o valor dos parâmetros, crescente em ARI (variando de 33,8% quando  $\phi = -0,8$  para 84,4% quando  $\phi = 0,8$ ) e decrescente em IMA (diminuindo de 77,4% quando  $\theta = -0,8$  para 37,4% quando  $\theta = 0,8$ ).

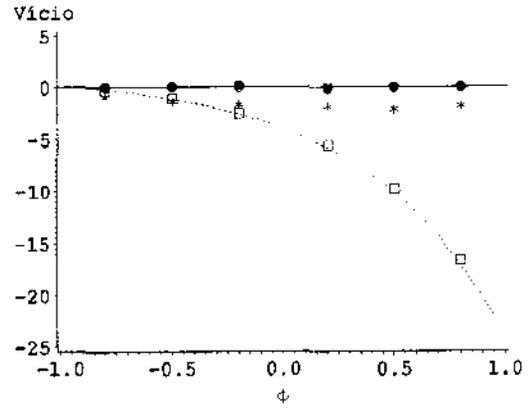
Em ambos os modelos os casos mais frequentes varia conforme o valor dos parâmetros do modelo, no modelo ARI, quando  $\phi \rightarrow -1$ , o caso mais frequente é detectar apenas um outlier na última observação como AO, com 15,2%, quando  $\phi = -0,8$ , diminuindo conforme aumenta o valor de  $\phi$ . Já para  $\phi \rightarrow +1$ , o caso mais frequente é identificar os dois outliers corretamente, com 72,1% quando  $\phi = 0,8$ . O menos frequente é detectar a penúltima observação como IO e a última como AO, sendo praticamente nulo para ambos os modelos. Nos outros casos, onde apenas o penúltimo outlier é detectado, a porcentagem de detecção não varia muito com os valores dos parâmetros, ficando em torno de 11%, quando é identificado como AO e de 2,5%, quando identificado como IO.

Apesar de não ter nenhum outlier presente na série agregada, pois eles se cancelaram, foi aplicado o teste de detecção de outliers, e os resultados mostram que a porcentagem de "outliers detectados" no modelo agregado não diminuiu muito em relação ao caso de apenas um outlier de tamanho 1.

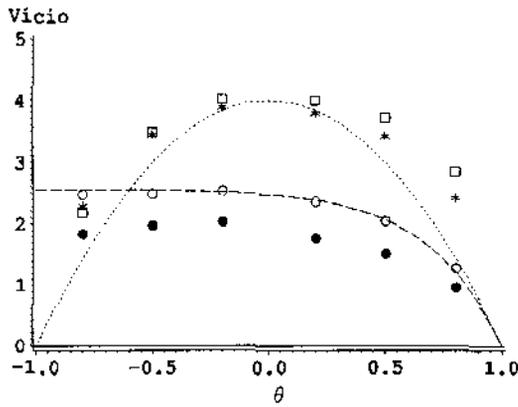
O vício ao aplicar o teste diminuiu bastante, pois as proporções de outliers detectados e identificados corretamente foram altas, para ambos os modelos desagregados ARI e IMA (Gráfico 4.3.18 (b) e (d)). Quando os dois outliers são identificados corretamente os outliers reestimados são bastante próximos do valor real, diminuindo assim o vício. Nos modelos agregados a proporção de outliers detectados é pequena, assim o vício médio continua praticamente inalterado. Com a redução dos vícios os EQMs dos modelos desagregados diminuíram bastante, chegando a ser menor que os EQMs dos modelos agregados, para ARI e IMA (Gráfico 4.3.19 (b) e (d)). Assim, a eficiência passa a ser menor que um, na maioria dos casos (Gráfico 4.3.20 (b) e (d)).



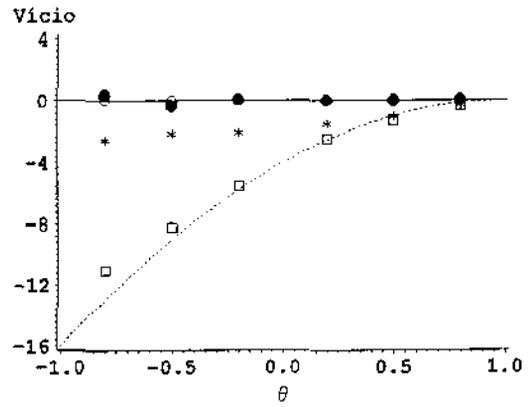
a) Vício causado por dois AOs de tamanho 1 de mesmo sinal, no modelo ARIMA(1,1,0).



b) Vício causado por dois AOs de tamanho 1 de sinais trocados, no modelo ARIMA(1,1,0).



c) Vício causado por dois AOs de tamanho 1 de mesmo sinal, no modelo ARIMA(0,1,1).



d) Vício causado por dois AOs de tamanho 1 de sinais trocados, no modelo ARIMA(0,1,1).

**Legenda:**

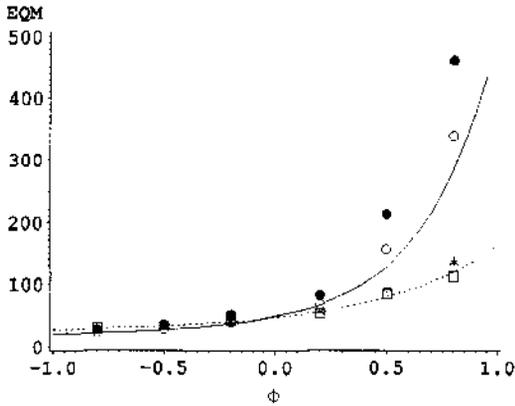
**Modelo desagregado**

- - vício para modelo conhecido
- - vício simulado sem teste
- \* - vício simulado com teste

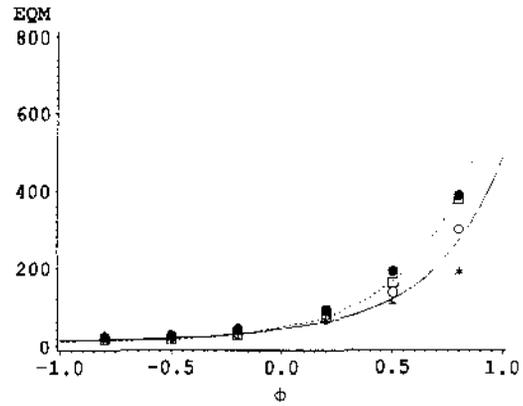
**Modelo agregado**

- - vício para modelo conhecido
- - vício simulado sem teste
- - vício simulado com teste

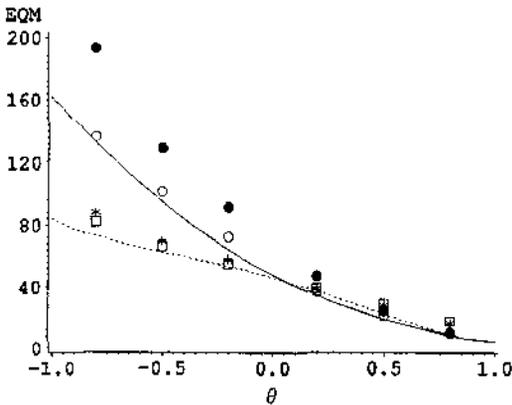
**Gráfico 4.3.18** - Vício na previsão um passo à frente, causado por dois AOs nas duas últimas observações de séries com 100 observações, em 1000 replicações.



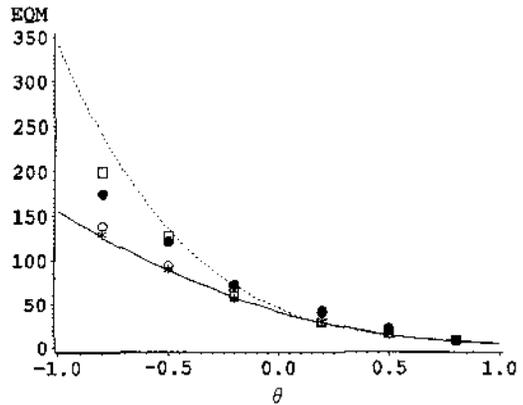
a) dois AOs de tamanho 1, de mesmo sinal, no modelo ARIMA(1,1,0).



b) dois AOs de tamanho 1 de sinais trocados, no modelo ARIMA(1,1,0).



c) dois AOs de tamanho 1 de mesmo sinal, no modelo ARIMA(0,1,1).



d) dois AOs de tamanho 1 de sinais trocados, no modelo ARIMA(0,1,1).

**Legenda:**

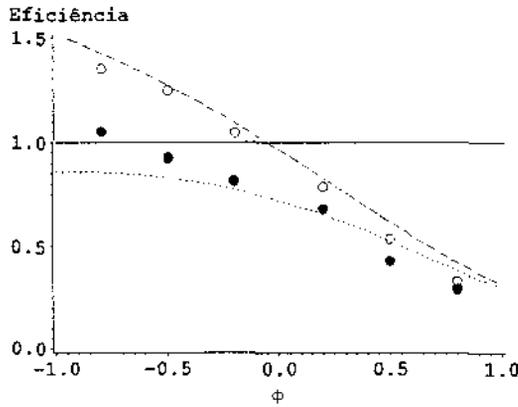
**Modelo desagregado**

- EQM para modelo conhecido
- EQM simulado sem teste
- \* EQM simulado com teste

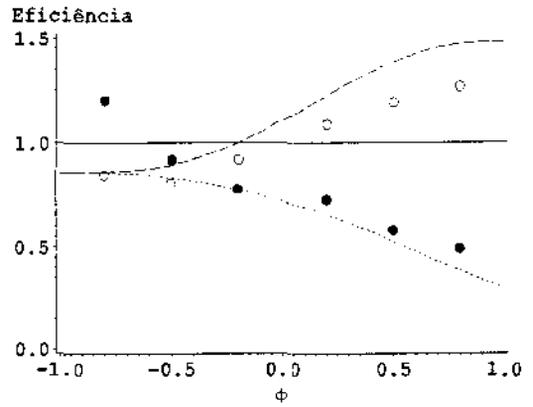
**Modelo agregado**

- EQM para modelo conhecido
- EQM simulado sem teste
- EQM simulado com teste

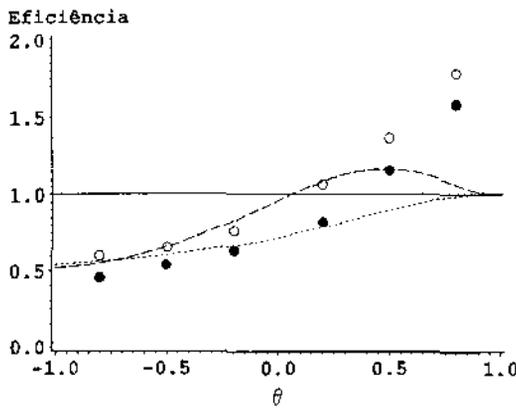
**Gráfico 4.3.19** - EQM na previsão um passo à frente, com dois AOs nas duas últimas observações da série com 100 observações e 1000 replicações.



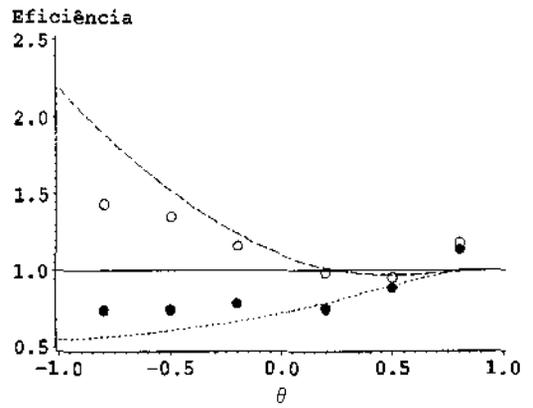
a) dois AOs de tamanho 1 de mesmo sinal, no modelo ARIMA(1,1,0)



b) dois AOs de tamanho 1 de sinais trocados, no modelo ARIMA(1,1,0).



c) dois AOs de tamanho 1 de mesmo sinal, no modelo ARIMA(0,1,1).



d) dois AOs de tamanho 1 de sinais trocados, no modelo ARIMA(0,1,1).

**Legenda:**

- .... Eficiência considerando o modelo conhecido, sem outlier
- Eficiência considerando o modelo conhecido, com outlier
- Eficiência considerando o modelo desconhecido, sem teste
- Eficiência considerando o modelo desconhecido, com teste

**Gráfico 4.3.20** - Eficiência de previsão um passo à frente de modelo agregado, considerando o modelo conhecido e desconhecido, com e sem AO de tamanho 1 de mesmo sinal ou sinais trocados. Para séries de tamanho 100 em 1000 replicações.

## 4.4 Efeito de IO na previsão

Ledolter (1987) mostrou que IO não afeta as previsões pontuais quando o modelo é conhecido. Assim, a presença de IO não deve afetar a previsão pontual de um modelo desagregado, mas pode afetar a previsão quando o modelo é desconhecido através do efeito na identificação e estimação dos parâmetros. No caso do modelo agregado o outlier afeta a previsão, mesmo que o modelo seja conhecido, caso a presença do outlier não seja totalmente conhecida.

Na primeira subseção são apresentados os efeitos de IO na previsão supondo o modelo desagregado ARIMA(0,1,1) e ARIMA(1,1,0) conhecidos. Os resultados das simulações no caso de modelos desconhecidos são apresentados na subseção 4.4.2.

### 4.4.1 Modelo conhecido

Considere o modelo desagregado sem outlier dado por (4.1.1) e o modelo agregado dado por (4.1.3). Então

$$\left[ \prod_{j=1}^p (1 - \xi_j^m B) \right] (1 - B)^d X_\tau = \theta_q(B) \left( \sum_{j=0}^{m-1} B^j \right)^{d+1} \left[ \prod_{j=1}^p \left( 1 + \sum_{i=0}^{m-1} (\xi_j B)^i \right) \right] a_{\tau m}, \quad (4.4.1a)$$

$$\Phi(B)X_\tau = \mu_\tau = \alpha_{Q'}(B) a_{\tau m} = \Theta_Q(B) \varepsilon_\tau \quad (4.4.1b)$$

onde  $\xi_j$  são as raízes de  $\phi_p(B)$ ;  $\Phi(B) = \Phi_p(B)(1-B)^d$ ;  $Q = [Q'/m]$  e  $Q' = (m-1)(p+d+1)+q$ . Assume-se que não há raízes em comum entre os polinômios  $\Theta_Q(B)$  e  $\Phi(B)$  e que as seqüências  $\{a_t\}$  e  $\{\varepsilon_t\}$  são processos de ruído branco.

Considere a presença de um IO na série desagregada na posição  $k$ . Então,

$$\Phi(B)Y_\tau = \mu_\tau \quad (4.4.2a)$$

e nesse caso

$$\begin{aligned} \mu_\tau &= \alpha_{Q'}(B) (a_{\tau m} + \omega_I \delta_{\tau,ka}) \\ &= \Theta_Q(B) \varepsilon_\tau + \alpha_{Q'}(B) \omega_I \delta_{\tau,ka} \end{aligned} \quad (4.4.2b)$$

$$= \Theta_Q(B) e_\tau \quad (4.4.2c)$$

onde

$$\begin{aligned} e_\tau &= \varepsilon_\tau + \Theta_Q^{-1}(B) \alpha_{Q'}(B) \omega_I \delta_{\tau,ka} \\ &= \varepsilon_\tau + \gamma(B) \omega_I \delta_{\tau,ka} \end{aligned} \quad (4.4.3)$$

$$e \quad ka = [(k-1)/m]+1,$$

isto é, a presença de um IO no modelo desagregado não é equivalente a ter um IO ou AO no modelo agregado.

Da equação (4.4.2) tem-se que a função de previsão de modelo agregado é dada por:

$$\hat{Y}_N(h) = \sum_{j=1}^{p+d} \Phi_j \hat{Y}_N(h-j) + \hat{\mu}_{N+h},$$

Da equação (4.4.2c), tem-se que:

$$\hat{\mu}_{\tau+h} = - \sum_{i=0}^Q \Theta_i \hat{e}_{\tau+h-i}, \quad \Theta_0 = -1,$$

onde

$$\hat{e}_{\tau+j} = \begin{cases} 0 & j > 0 \\ e_{\tau+j} = [\Theta_Q(B)]^{-1} \Phi(B) Y_{\tau+j} & j \leq 0 \end{cases}$$

Considere a presença de um IO de tamanho  $\omega_I$ ,  $r$  observações antes da origem de previsão. Se os modelos agregado e desagregado são conhecidos,  $\mu_{\tau+h}$  é conhecido pelo modelo (4.4.2b), i.e.,

$$\hat{\mu}_{\tau+h} = \Theta_Q(B) \hat{e}_{\tau+h} - \omega_I \delta_{\tau+h,ka} \alpha_{hm+r}.$$

O segundo termo é igual a zero se  $(hm+r) > Q'$ , pois o polinômio  $\alpha_{Q'}(B)$  é de ordem  $Q'$ . Tem-se também que  $\hat{e}_{\tau+i}$  é igual a zero para  $i > 0$ , pois  $\hat{e}_{\tau+i} = \varepsilon_{\tau+i}$  para  $i \leq 0$  e  $\varepsilon_{\tau}$  é um processo ruído branco. Por (4.4.3) tem-se que:

$$\varepsilon_{\tau-j} - e_{\tau-j} = \omega_I \delta_{\tau-j,ka} \gamma(B),$$

$$\varepsilon_{\tau-j} - e_{\tau-j} = \omega_I \delta_{\tau-j,ka} \gamma_{1-mj}, \quad \gamma_0 = -1,$$

que é igual a zero para  $j \geq R+1$ , pois o resíduo da série com outlier antes da ocorrência do outlier é igual à inovação. Assim, o preditor levando em consideração a presença do outlier, é dado por:

$$\begin{aligned} \hat{Y}_N^0(h) &= \sum_{j=1}^{p+d} \Phi_j \hat{Y}_N^0(h-j) - \hat{\mu}_{N+h}^0 \\ &= \sum_{j=1}^{p+d} \Phi_j \hat{Y}_N^0(h-j) - \\ &\quad \Theta_h \varepsilon_N - \Theta_{h+1} \varepsilon_{N-1} - \dots - \Theta_R \varepsilon_{N-R} - \Theta_{R+1} \varepsilon_{N-R-1} - \Theta_Q \varepsilon_{N+h-Q} - \omega_1 \alpha_{r+hm}. \end{aligned}$$

Utilizando a definição dada anteriormente, o vício (efeito na média) da previsão é dado por

$$\begin{aligned} \text{vício}^a(r,h) &= \hat{Y}_N(h) - \hat{Y}_N^0(h) \\ &= \Phi_1 \text{vício}^a(r,h-1) + \dots + \Phi_{p+d} \text{vício}^a(r,h-p-d) + \\ &\quad \omega_1 [\Theta_h \gamma_r + \Theta_{h+1} \gamma_{r-m} + \dots + \Theta_{k+R} \gamma_{r-mR} + \Phi_{r+hm}], \end{aligned} \quad (4.4.4)$$

onde  $\text{vício}^a(r,h)$  é igual a zero para  $h \leq 0$ . A segunda parte é igual a zero quando  $h \geq Q+1$ .

O EQM<sup>a</sup>(r,h) é dado por

$$\text{EQM}^a(r,h) = \text{Var}(\mu_{\tau+h}) + [\text{vício}^a(r,h)]^2.$$

Para avaliar esses efeitos serão considerados a seguir alguns casos especiais: ARIMA(1,1,0) e ARIMA(0,1,1).

#### 4.4.1.1 ARIMA(1,1,0) desagregado

Considere o modelo (4.1.1) dado na seção 4.1 com período de agregação igual a 4,  $p = d = 1$  e  $q = 0$ . Então, o modelo agregado é dado por:

$$(1 - \phi^4 B)(1 - B)X_\tau = (1 - \Theta_1 B - \Theta_2 B^2) \varepsilon_\tau;$$

Pelas equações (4.4.1) e (4.4.3) tem-se que:

$$\varepsilon_\tau = \Theta_Q(B)^{-1} \alpha_{Q'}(B) a_{\tau m}$$

$$e_\tau = \varepsilon_\tau + \Theta_Q^{-1}(B) \alpha_{Q'}(B) \omega_1 \delta_{\tau m, k} = \Theta_Q^{-1}(B) \alpha_{Q'}(B) (a_{\tau m} + \omega_1 \delta_{\tau m, k})$$

$$e_\tau = [1 - \Theta_1 B - \Theta_2 B^2]^{-1} \sum_{j=1}^Q \alpha_j B^j [a_{\tau m} + \omega_1 \delta_{\tau m, k}],$$

onde

$$\begin{aligned} \alpha_{Q'}(B) &= \left( \sum_{j=0}^{m-1} B^j \right)^{d+1} \left[ \prod_{j=1}^p (1 + \sum_{i=0}^{m-1} (\xi_j B)^i) \right] = \left( \sum_{j=0}^3 B^j \right)^2 \left( 1 + \sum_{i=0}^3 (\phi B)^i \right) \\ &= (1 + B + B^2 + B^3)^2 (1 + \phi B + \phi^2 B^2 + \phi^3 B^3) = \sum_{j=1}^9 \alpha_j B^j \end{aligned}$$

que resulta nos seguintes valores:

$$\begin{aligned} \Phi_1 &= 1+\phi^4 & \Phi_2 &= -\phi^4 \\ \alpha_1 &= -(2+\phi) & \alpha_2 &= -(3+2\phi+\phi^2) & \alpha_3 &= -(4+3\phi+2\phi^2+\phi^3) \\ \alpha_4 &= -(3+4\phi+3\phi^2+2\phi^3) & \alpha_5 &= -(2+3\phi+4\phi^2+3\phi^3) & \alpha_6 &= -(1+2\phi+3\phi^2+4\phi^3) \\ \alpha_7 &= -(\phi+2\phi^2+3\phi^3) & \alpha_8 &= -(\phi^2+2\phi^3) & \alpha_9 &= -\phi^3 \end{aligned}$$

$\gamma(B) = (1-\Theta_1 B^4 - \Theta_2 B^8)^{-1} (1 - \omega_T \delta_{vm,k} \sum_{j=1}^9 \alpha_j B^j)$ , onde  $\Theta_1$  e  $\Theta_2$  são dados na seção 3.1.

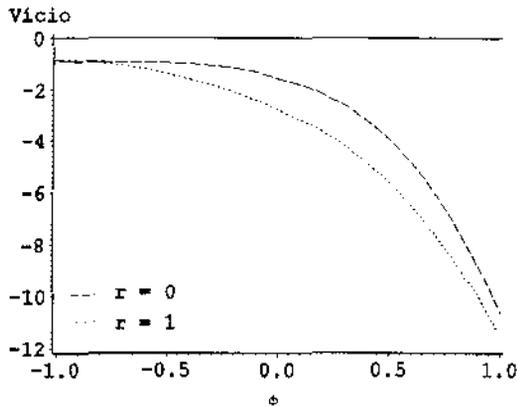
Com esses resultados pode-se calcular os efeitos no caso de modelo conhecido de IO do modelo desagregado num modelo agregado dado pela equação (4.4.4).

A seguir são analisados os efeitos, em termos de vício e EQM nos modelos ARIMA(1,1,0) e ARIMA(0,1,1) separadamente. Para facilitar a análise, para cada modelo, serão apresentados três casos, como na subseção 4.3.1.

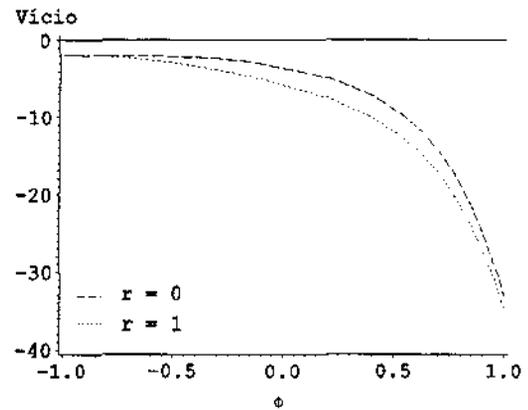
### 1) Um IO de tamanho 1

O Gráfico 4.4.1 mostra o vício nas previsões do modelo agregado causado por um IO de tamanho unitário para  $r = 0$  e 1 para previsões 1 e 2 passos à frente. O vício pode ser muito grande quando  $\phi$  está próximo de +1, chegando em módulo aproximadamente a 11 e 35, para 1 e 2 passos à frente, respectivamente, quando o outlier ocorre na última ou penúltima observação. Quando o parâmetro autoregressivo é negativo o vício não é muito grande. Genericamente, o vício cresce com o horizonte de previsão e com  $\phi$ , e decresce com  $r$ .

Os efeitos no EQM e na eficiência são apresentados no Gráfico 4.4.2. Pode-se verificar que os EQMs do modelo agregado são maiores que do modelo desagregado, aumentando a diferença conforme aumenta o valor de  $\phi$ . Analisando as eficiências com e sem outlier para  $r = 0$  e 1, tem-se que a eficiência do modelo agregado diminui na presença de um IO, o que é esperado, pois no modelo desagregado o EQM não muda, enquanto no modelo agregado ele apenas aumenta. A eficiência mínima ocorre quando  $\phi = 1$  e é em torno de 0,25 para  $r = 0$  e 1. A diferença nas eficiências com e sem outlier diminui conforme  $r$  aumenta, i.e., quando a posição do outlier distancia da origem de previsão, mas depende pouco dos valores de  $\phi$ .

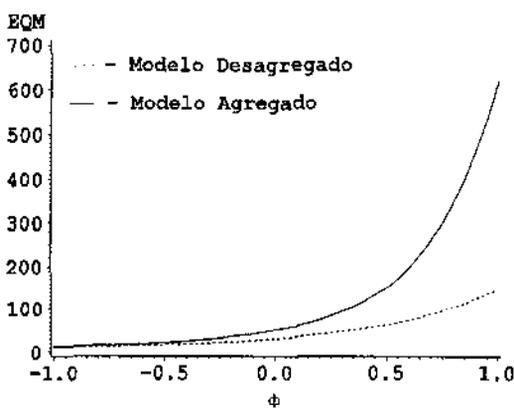
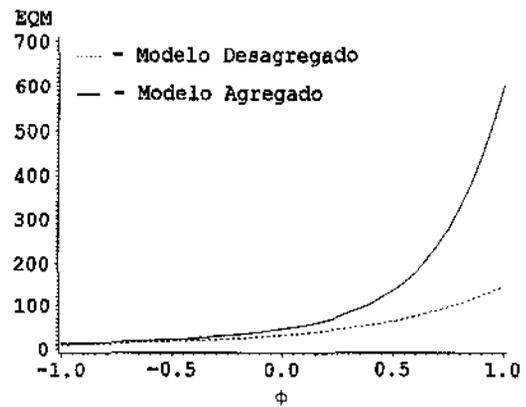
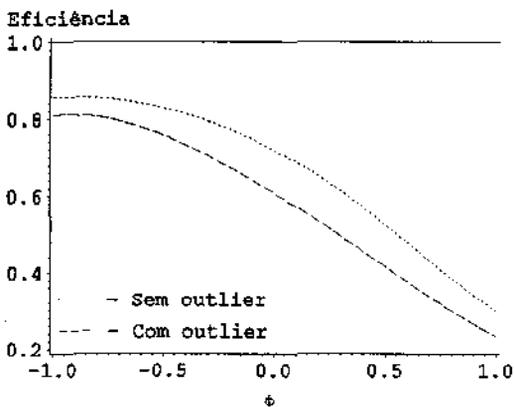
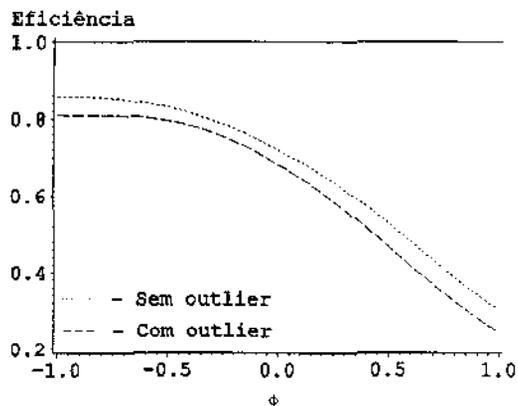


a) um passo à frente



b) dois passos à frente

**Gráfico 4.4.1** - Vício causado por um IO de tamanho 1, presente  $r$  observações antes da origem de previsão, nas previsões um e dois passos à frente, num modelo ARIMA(1,1,0) agregado conhecido (no modelo desagregado o efeito é nulo).

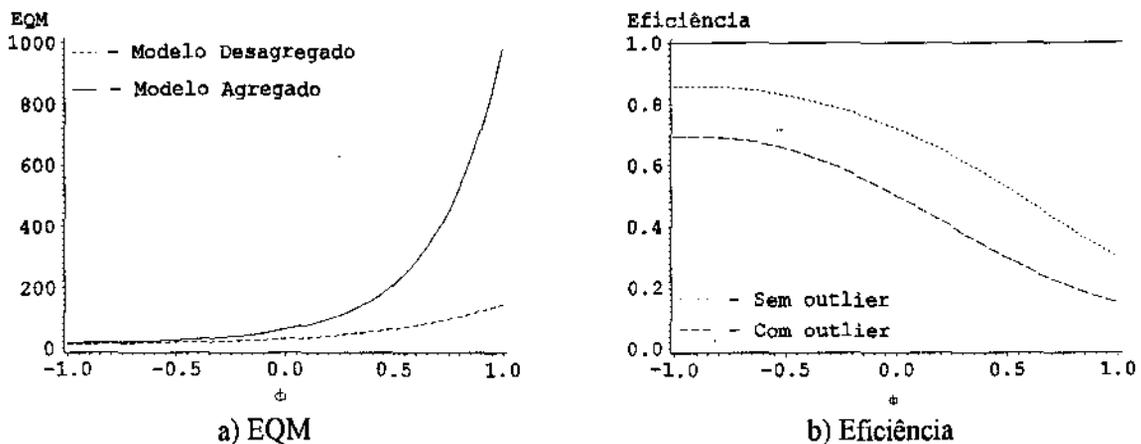
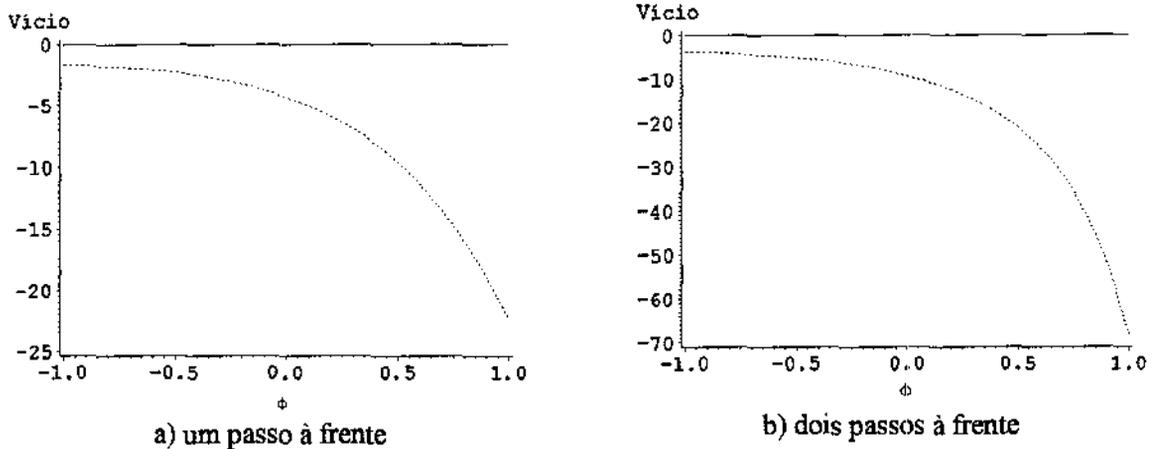
a) EQM,  $r = 0$ b) EQM,  $r = 1$ c) Eficiência,  $r = 0$ d) Eficiência,  $r = 1$ 

**Gráfico 4.4.2** - EQM e eficiência com um IO de tamanho unitário, presente  $r$  posições antes da origem de previsão, no modelo ARIMA(1,1,0) conhecido, na previsão um passo à frente.

**2) Dois IOs de tamanho 1 de mesmo sinal**

O vício de dois IOs de tamanho 1 nas duas últimas observações são apresentados no Gráfico 4.4.3. O vício não é muito grande quando  $\phi < 0$  e é crescente em  $\phi$ .

No Gráfico 4.4.4 são apresentados os efeitos no EQM e na eficiência da previsão um passo à frente. Os comentários em relação ao vício, EQM e eficiência são análogos ao caso anterior. O vício aumenta muito conforme aumenta a horizonte de previsão e o valor de  $\phi$ . O EQM no modelo agregado é maior e a diferença absoluta com o do desagregado aumenta mais ainda conforme aumenta o valor de  $\phi$ . No entanto, embora a eficiência do modelo agregado diminua bastante na presença de dois IOs, a diminuição da eficiência não depende muito de  $\phi$ .

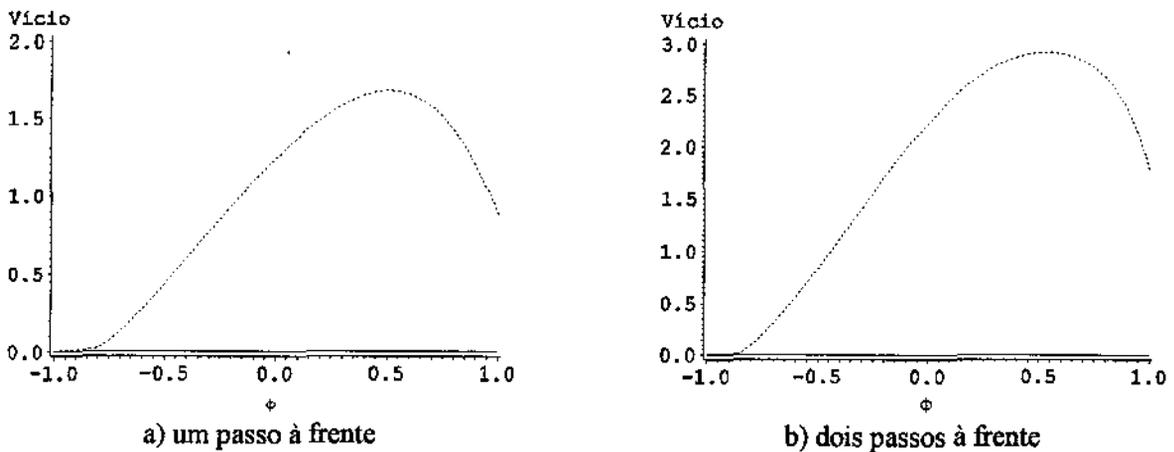


**Gráfico 4.4.4 - EQM e eficiência de previsão um passo à frente com dois IOs de tamanho unitário de mesmo sinal, nas duas últimas observações no modelo ARIMA(1,1,0) conhecido.**

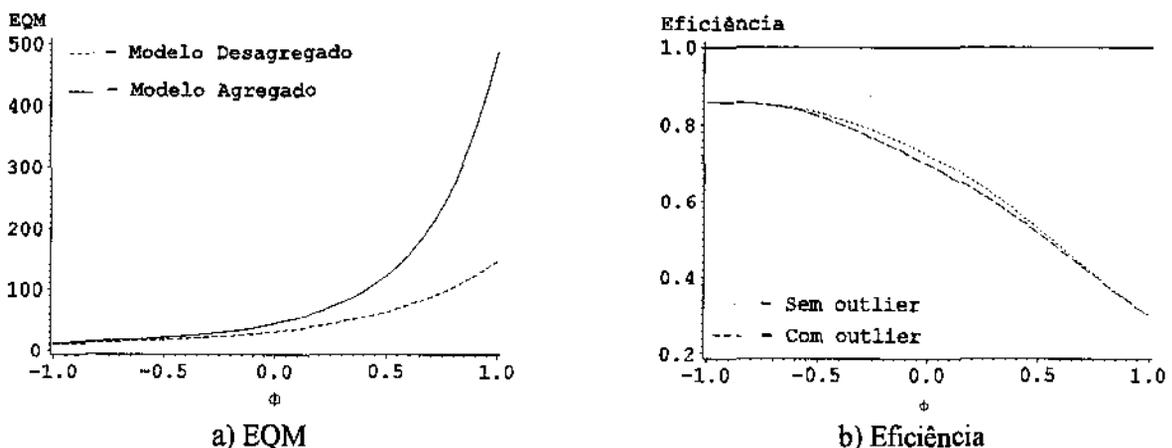
### 3) Dois IOs de tamanho 1 de sinais trocados

Os vícios de dois IOs de tamanho 1 com sinais trocados nas duas últimas observações no modelo ARI não são grandes em toda amplitude de  $\phi$ , todos são menores que 1,7 e 2,9 para previsão um e dois passos à frente, respectivamente. O vício é crescente até  $\phi = 0,5$ , atingindo o máximo nesse ponto para previsão um e dois passos à frente (ver Gráfico 4.4.5).

Os comentários em relação ao EQM são análogos aos casos anteriores, maior no modelo agregado aumentando conforme  $\phi \rightarrow +1$ . No entanto, a eficiência com e sem outlier não muda muito, dado que o vício é relativamente pequeno, como pode ser confirmado no Gráfico 4.4.6b.



**Gráfico 4.4.5** - Vício causado por dois IOs de tamanho 1 e -1, respectivamente, presentes na penúltima e última observação, num modelo ARIMA(1,1,0) agregado conhecido (no modelo desagregado o vício é zero).



**Gráfico 4.4.6** - EQM e eficiência de previsão um passo à frente dos modelos ARIMA(1,1,0) conhecido, quando dois IOs de tamanho 1 e -1 ocorrem nas duas últimas observações da série.

#### 4.4.1.2 ARIMA(0,1,1) desagregado

Considere o modelo ARIMA(0,1,1) desagregado com um IO dado por

$$y_t = y_{t-1} + (1-\theta B)(a_t + \omega_t \delta_{t,ka}).$$

Na subseção 4.2.1 foi visto que um IMA agregado continua sendo um IMA. Assim, usando período de agregação 4, e as equações (4.4.1) e (4.4.3), o modelo agregado deve ser dado por:

$$(1-B)Y_\tau = (1-\Theta B)e_\tau,$$

$$e_\tau = \gamma(B)(a_{\tau m} + \omega_\tau \delta_{\tau,ka}),$$

onde  $\gamma(B) = \Theta(B)^{-1} \alpha_Q(B) = \Theta(B)^{-1} \theta(B) \left( \sum_{j=0}^{m-1} B^j \right)^{d+1} = \Theta(B)^{-1} (1-\theta B) \left( \sum_{j=0}^3 B^j \right)^2$ . Assim

$$e_\tau = (1-\Theta B)^{-1} \left( \sum_{i=1}^7 \alpha_i B^i \right) (a_{\tau m} + \omega_\tau \delta_{\tau,ka}),$$

onde

$$\alpha_0 = -1 \quad \alpha_1 = \theta - 2 \quad \alpha_2 = 2\theta - 3 \quad \alpha_3 = 3\theta - 4$$

$$\alpha_4 = 4\theta - 3 \quad \alpha_5 = 3\theta - 2 \quad \alpha_6 = 2\theta - 1 \quad \alpha_7 = \theta$$

e  $\gamma(B) = \left( -\sum_{i=0}^{\infty} \gamma_i B^i \right) = (1 - \Theta B)^{-1} \left( -\sum_{i=1}^7 \alpha_i B^i \right)$ , onde

$$\gamma_i = \begin{cases} \alpha_i & 0 < i \leq 3 \\ \alpha_4 - \Theta & i = 4 \\ \Theta \alpha_{i-4} + \alpha_i & 5 \leq i \leq 7 \\ \Theta^i \alpha_s + \Theta^{j-1} \alpha_{4+s} & i = 4j + s, j = 2, 3, \dots, s = 0, 1, 2, 3. \end{cases}$$

Supondo apenas um IO o vício é dado por

$$\text{vício}^a(r, 1) = \begin{cases} -\Theta + (4\theta - 3) & r = 0 \\ \Theta(\theta - 2) + (3\theta - 2) & r = 1 \\ \Theta(2\theta - 3) + (2\theta - 1) & r = 2, \\ \Theta(3\theta - 4) + \theta & r = 3 \\ \Theta \text{eff}(r - 4, 1) & r = 4 \end{cases}$$

e  $\text{vício}^a(r, h) = \text{vício}^a(r, 1)$  para  $h \geq 1$ .

A seguir são analisados os efeitos, em termos de vício, EQM e eficiência nos modelos ARIMA(1,1,0) e ARIMA(0,1,1).

### 1) Um IO de tamanho 1

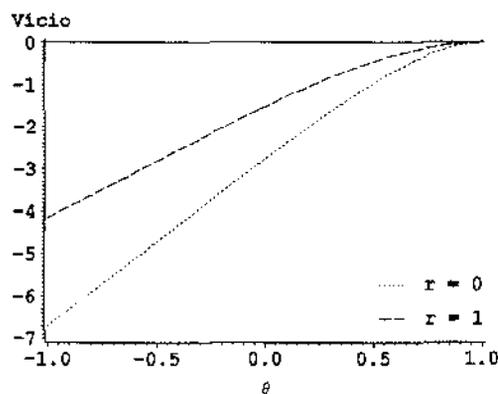
No Gráfico 4.4.7 são apresentados os vícios causados por um IO presente na penúltima ou última observação da série, no modelo agregado. Por esse gráfico pode-se verificar que para  $r = 0$  o vício é sempre maior e ele aumenta conforme diminui o valor de  $\theta$ .

Nos EQMs apresentados no Gráfico 4.4.8, verifica-se que eles são maiores quando  $r$  é menor e aumentam quando  $\theta \rightarrow -1$ . Na presença de outlier a eficiência é sempre menor com as eficiências mínimas de 0,42 e 0,48, para  $r = 0$  e 1, respectivamente. A diferença entre as eficiências com e sem outlier diminui quando  $\theta \rightarrow +1$  e  $r$  é maior.

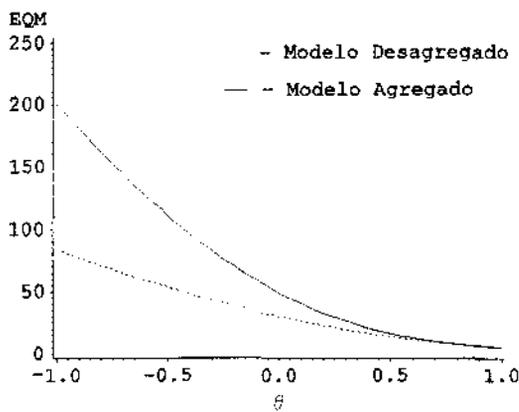
### 2) Dois IOs de tamanho 1 de mesmo sinal

O vício de dois IOs de tamanho 1 nas duas últimas observações é apresentado no Gráfico 4.4.9. Como no caso anterior, sendo o vício resultante a soma dos vícios dos casos onde ocorre apenas um outlier na última ( $r = 0$ ) e na penúltima observação da série ( $r = 1$ ), ele cresce (em módulo) quando  $\theta \rightarrow -1$ .

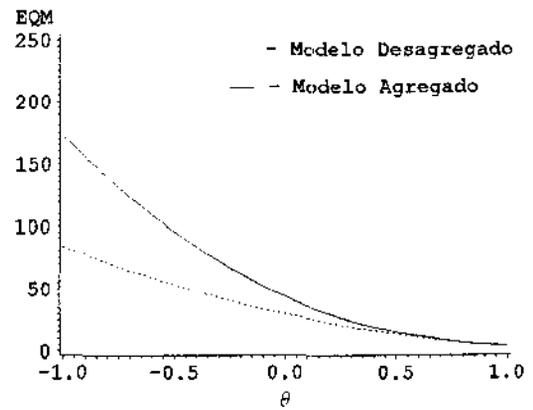
O efeito do outlier no EQM e conseqüentemente na eficiência é análogo ao caso anterior, com a diferença de que a eficiência mínima agora é menor, em torno de 0,3, pois o vício no modelo agregado aumentou bastante, aumentando assim o EQM (Gráfico 4.4.10).



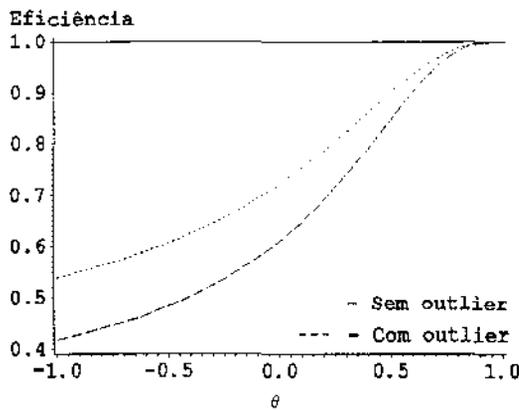
**Gráfico 4.4.7** - Vício causado por um IO presente  $r$  observações antes da origem de previsão, nas previsões  $h$  ( $\geq 1$ ) passos à frente, num modelo ARIMA(0,1,1) agregado conhecido (o vício no modelo desagregado é igual a zero).



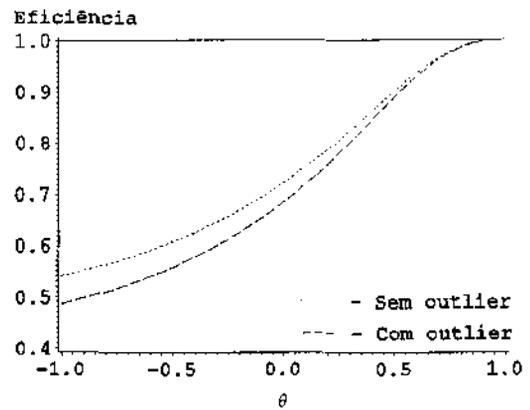
a) EQM,  $r = 0$



b) EQM,  $r = 1$

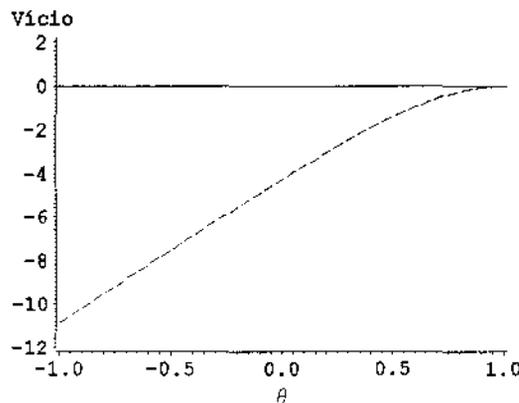


c) Eficiência,  $r = 0$

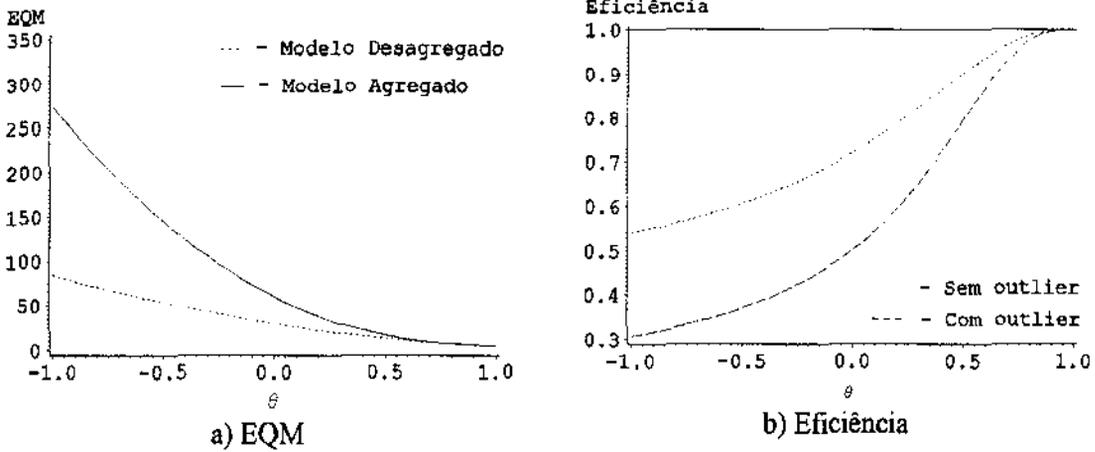


d) Eficiência,  $r = 1$

**Gráfico 4.4.8** - EQM e eficiência de previsão um passo à frente, com um IO de tamanho unitário presente  $r$  posições antes da origem de previsão, no modelo ARIMA(0,1,1) conhecido.



**Gráfico 4.4.9** - Vício causado por dois outliers IO nas previsões  $h (\geq 1)$  passos à frente num modelo ARIMA(0,1,1) agregado conhecido. Os dois outliers estão presentes nas duas últimas observações.

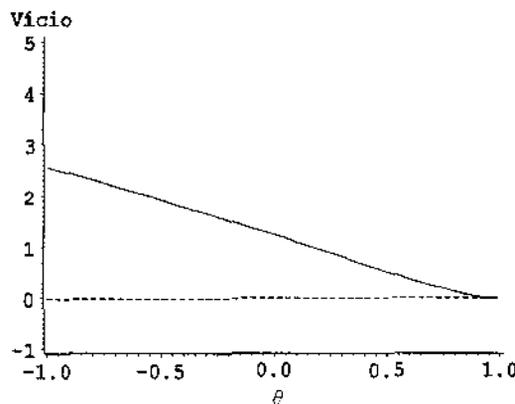


**Gráfico 4.4.10** - EQM e eficiências de modelos agregados e desagregados quando dois IOs, de mesmo sinal, ocorrem nas duas últimas observações da série.

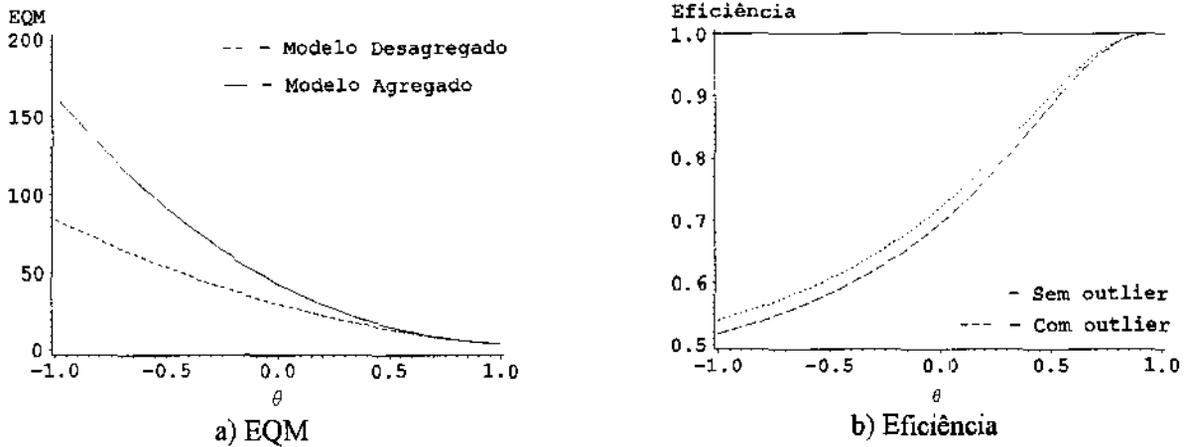
### 3) Dois IOs de tamanho 1 com sinais trocados

Os vícios de dois IOs de tamanho 1 com sinais trocados nas duas últimas observações não são grandes em toda amplitude de  $\theta$ , todos são menores que 2,6 e são positivos. O vício é decrescente em  $\theta$  (ver Gráfico 4.4.11).

Os EQMs do modelo agregado são maiores que do desagregado para todos os valores de  $\theta$ , aumentando quando  $\theta$  tende a -1, como nos casos anteriores. Entretanto a diminuição de eficiência devido ao outlier é bem pequena (ver Gráfico 4.4.12 (b)).



**Gráfico 4.4.11** - Vício causado por dois outliers IOs de tamanho 1, de sinais trocados, respectivamente na penúltima e última observação nas previsões  $h (\geq 1)$  passos à frente num modelo ARIMA(0,1,1) conhecido agregado.



**Gráfico 4.4.12** - EQM e eficiência com dois IOs de tamanho 1, de sinais trocados, nas duas últimas posições da séries de um processo ARIMA(0,1,1) conhecido.

#### 4.4.2 - Modelo desconhecido

Para estudar o caso de modelos desconhecidos foi feita uma análise através de simulações como na subseção 4.3.2. Foi utilizado o mesmo programa (o algoritmo é apresentado na subseção 4.3.2) e a ordem dos processos estimados foi analisada junto aos outros casos nessa mesma subseção.

No modelo desagregado, em se tratando de outlier do tipo IO na última observação nunca se pode detectar o outlier de forma totalmente correta, já que quando detectado na última observação será considerado como AO, conforme comentados na subseção 4.3.2. Assim, no modelo agregado todo outlier detectado será considerado como AO já que o possível outlier ocorre na última observação. De qualquer forma, independentemente da posição, o IO no modelo desagregado não pode ser detectado corretamente no modelo agregado já que o efeito de um outlier IO no modelo agregado não pode ser classificado como nenhum dos outliers considerados neste trabalho (ver início da seção 4.4.1).

A seguir são analisados os efeitos nas previsões em termos de vício e EQM. Para facilitar a análise, os resultados foram divididos em 4 casos, como na subseção anterior.

#### 4.4.2.1 - Um IO de tamanho 1

##### a) sem teste de detecção de outlier

Nos Gráficos 4.4.13 (a) e (c) (para  $r = 0$ ) e Gráficos 4.4.14 (a) e (c) (para  $r = 1$ ) são ilustrados os vícios para modelos conhecidos, e os vícios obtidos pelas simulações. Os vícios simulados para ambos os modelos ARI e IMA foram, de um modo geral, bem próximos dos vícios para modelos conhecidos tanto no modelo agregado como no desagregado, com algumas exceções, como no caso de modelos IMA desagregado com  $|\theta| \rightarrow +1$  quando o vício aumenta um pouco. Isso decorre provavelmente, como nos outros casos, do erro de estimação causado pela aproximação de um IMA para um ARI.

O fato dos vícios, em geral, serem próximos dos vícios para modelos conhecidos mostra que o outlier não teve um grande efeito na especificação e estimação do modelo. Desta forma, continuam válidas as comparações realizadas entre as previsões dadas pelos dois modelos quando o modelo era considerado conhecido, ou seja, que o EQM do modelo agregado pode ser muito maior do que do desagregado. As diferenças são maiores quando  $\phi \rightarrow +1$  e  $\theta \rightarrow -1$ , para os modelos ARI e IMA, respectivamente.

Em todos os casos, nos modelos ARI ou IMA, quando o outlier ocorre na penúltima ou última observação existe uma leve diminuição na eficiência, dada a especificação e estimação.

##### b) com teste de detecção de outlier

As porcentagens de detecção de outliers são apresentadas na Tabela A5.4. Considere inicialmente o caso em que o outlier IO está presente na penúltima observação da série desagregada. Para o modelo desagregado, a porcentagem de outliers detectados é em torno de 19%, tanto no modelo ARI como no modelo IMA, com pequena influência do coeficiente do modelo. No entanto, a porcentagem de casos onde o outlier é classificado corretamente varia com o coeficiente. Enquanto ele cresce com o valor de  $\phi$  no modelo ARI (de 53,7%, quando  $\phi = -0,8$  para 72,5% quando  $\phi = 0,8$ ), esta porcentagem decresce no modelo IMA (de 80%, quando  $\theta = -0,8$  para 50,9% quando  $\theta = 0,8$ ). No modelo agregado a porcentagem de outliers detectados (sempre classificado como AO por ocorrer na última observação) é bastante pequena. Para o modelo IMA é de aproximadamente 4% e praticamente independe do coeficiente  $\theta$ , enquanto para o modelo ARI é decrescente no coeficiente  $\phi$  (diminuiu de 4,6% quando  $\phi = -0,8$  para 1,2% quando  $\phi = 0,8$ ).

Quando o outlier ocorre na última observação, a porcentagem de casos onde o outlier é detectado diminui um pouco em relação ao caso em que ele ocorre na penúltima observação. Para o modelo desagregado fica em torno de 15%, tanto no modelo ARI como no modelo IMA dependendo pouco do modelo e dos coeficientes. Neste caso, todo outlier será classificado como AO por ocorrer na última observação, e portanto, sempre é classificado incorretamente. No modelo agregado, exceto pelo fato da porcentagem ter diminuído um pouco, todos os comentários realizados para o outlier na penúltima observação continuam válidos.

Nos Gráficos 4.4.13 (a) e (c) (para  $r = 0$ ) e Gráficos 4.4.14 (a) e (c) (para  $r = 1$ ) são ilustrados os vícios com e sem teste de detecção de outlier.

Nos modelos ARI e IMA desagregados, quando o outlier está presente na última observação (ver Gráficos 4.4.13 (a) e (c)), dependendo do valor dos parâmetros do modelo, os vícios aumentam bastante (em módulo) quando é aplicado o teste. Esse aumento pode ser explicado pelo fato de todos os outliers detectados serem identificados incorretamente e detectados, em geral, com magnitudes maiores que dois.

Quando o outlier está presente na penúltima observação da série desagregada (ver Gráficos 4.4.14 (a) e (c)) os vícios não alteram muito com a aplicação do teste, apenas aumentam um pouco em ambos os modelos agregados, ARI e IMA. Quando o outlier é detectado nestes modelos, o vício aumenta muito, pois o outlier detectado é muito maior do que o real, mas como são poucos os casos de detecção (em torno de 4%) o vício médio aumenta um pouco com a aplicação do teste. No entanto, o EQM aumenta consideravelmente (ver Gráfico 4.4.15), ao passo que nos modelos desagregados os EQMs praticamente são inalterados.

As eficiências apresentadas no Gráfico 4.4.17 mostram que o teste de detecção melhora a eficiência quando o outlier ocorre na última observação por aumento relativamente maior no EQM do modelo desagregado. Em alguns casos a eficiência chega a ser maior do que a eficiência calculada sem o outlier. Quando o outlier ocorre na penúltima observação, entretanto, a eficiência quase sempre piora.

#### 4.4.2.2 - Um IO de tamanho 2

##### a) sem o teste de detecção de outlier

Nos Gráficos 4.4.13 (b) e (d) (para  $r = 0$ ) e Gráficos 4.4.14 (b) e (d) (para  $r = 1$ ) são ilustrados os vícios causados por um outlier de tamanho 2, na previsão de valores agregados, para modelos ARI e IMA. Os vícios, os EQMs e as eficiências simulados para ambos os modelos foram, em geral, bem próximos dos vícios, EQMs e eficiências dos modelos conhecidos.

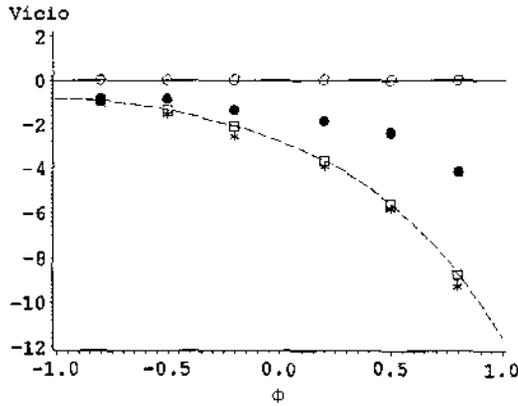
##### b) com o teste de detecção de outlier

Os comentários em termos de porcentagem de outliers são análogos ao caso em que o tamanho do outlier era igual a 1, apenas aumentando a porcentagem de detecção. Essa porcentagem tem um aumento maior do que 100% (por exemplo, de 20,4% para 50,9% para outlier na penúltima posição no modelo ARI com  $\phi = -0,2$ ), mas a porcentagem de detecção correta altera pouco.

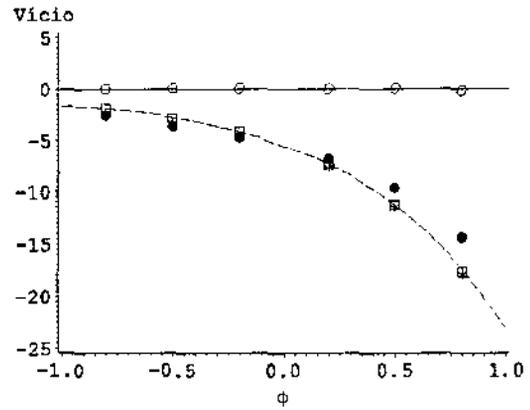
No caso agregado, o aumento percentual de outlier detectado é em geral menor, e a proporção de outliers detectados continua bastante pequena.

O vício com teste se comporta de forma análoga ao caso anterior com outlier de tamanho um, com diferença nas magnitudes dos vícios. No caso de  $r = 0$ , o vício tende a aumentar por três fatores: aumento da porcentagem de outliers detectados, e de outliers identificados incorretamente (pois todos os outliers detectados na última observação são identificados como AO) e aumento do valor do outlier detectado (ver Gráfico 4.4.13). Nesse caso os EQMs aumentam bastante, em alguns casos chegam a ser maiores que do agregado.

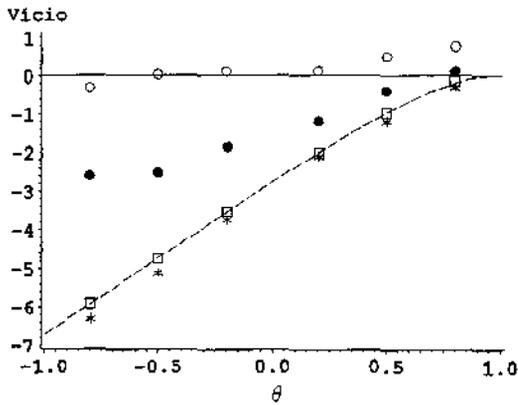
O efeito da aplicação do teste na eficiência é diferente para  $r = 0$  e 1. Quando o outlier ocorre na última observação ( $r = 0$ ) a eficiência é próxima a 1 nos modelos ARI e IMA. No modelo IMA é maior do que 1 para  $\theta > 0$  e menor para  $\theta < 0$ . No caso do outlier ocorrer na penúltima observação ( $r = 1$ ) a aplicação do teste diminui a eficiência, exceto quando  $|\phi|$  é próximo a 1.



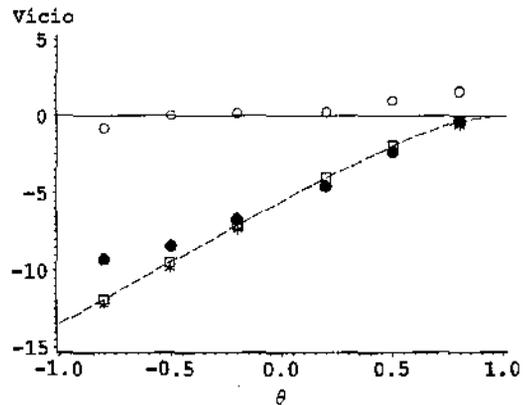
a) Vício causado por um IO de tamanho 1, no modelo ARIMA(1,1,0).



b) Vício causado por um IO de tamanho 2, no modelo ARIMA(1,1,0).



c) Vício causado por um IO de tamanho 1, no modelo ARIMA(0,1,1).



d) Vício causado por um IO de tamanho 2, no modelo ARIMA(0,1,1).

**Legenda:**

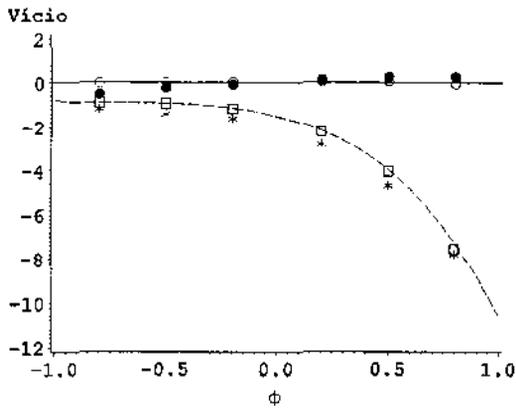
**Modelo desagregado**

- vício para modelos conhecidos
- vício simulado sem teste
- vício simulado com teste

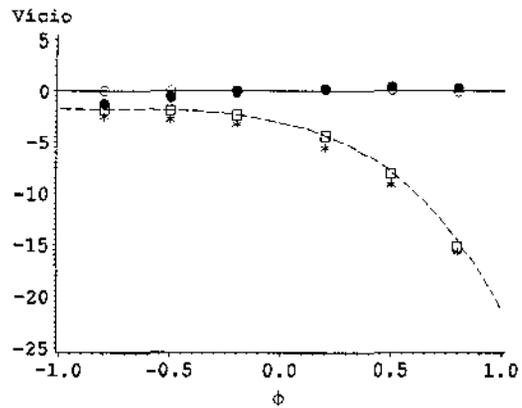
**Modelo agregado**

- vício para modelos conhecidos
- vício simulado sem teste
- \* vício simulado com teste

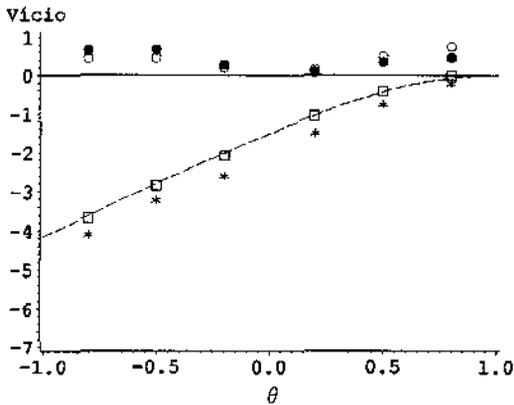
**Gráfico 4.4.13** - Vício causado por um IO na última observação ( $r = 0$ ) de uma série com 100 observações em 1000 replicações, com e sem teste de detecção de outlier.



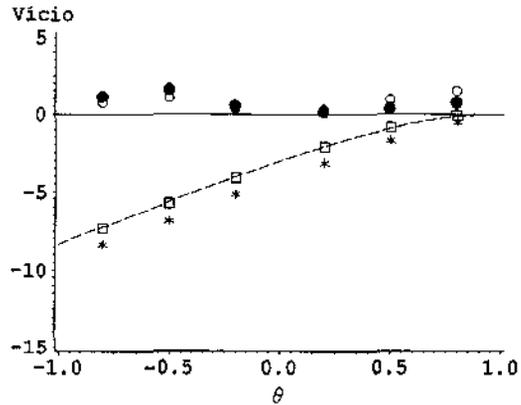
a) Vício causado por um IO de tamanho 1, no modelo ARIMA(1,1,0).



b) Vício causado por um IO de tamanho 2, no modelo ARIMA(1,1,0).



c) Vício causado por um IO de tamanho 1, no modelo ARIMA(0,1,1).



d) Vício causado por um IO de tamanho 2, no modelo ARIMA(0,1,1).

**Legenda:**

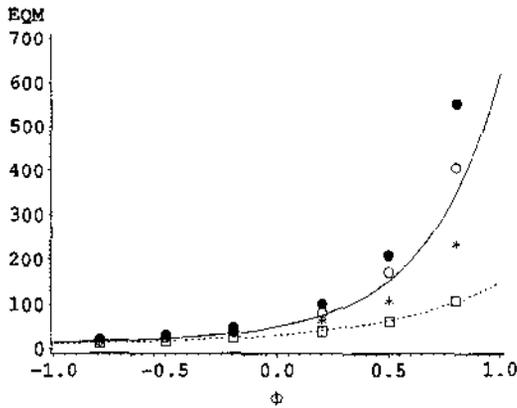
**Modelo desagregado**

- vício para modelos conhecidos
- vício simulado sem teste
- vício simulado com teste

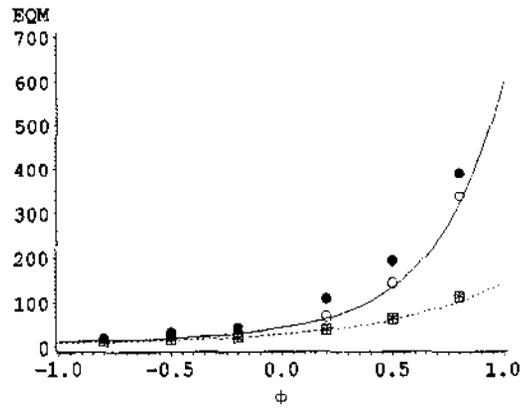
**Modelo agregado**

- vício para modelos conhecidos
- vício simulado sem teste
- \* vício simulado com teste

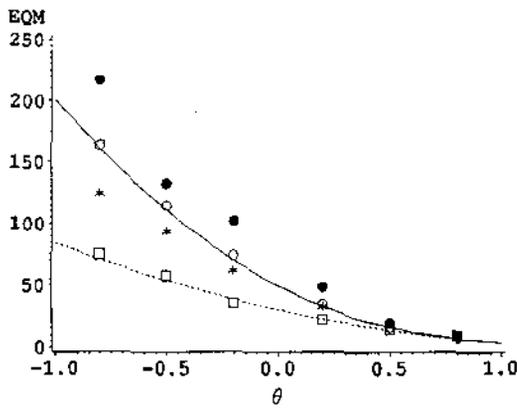
**Gráfico 4.4.14** - Vício causado por um IO na penúltima observação ( $r = 1$ ) de uma série com 100 observações em 1000 replicações, com e sem teste de detecção de outlier.



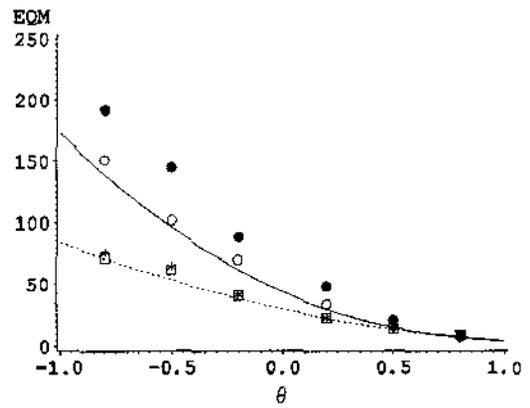
a) ARIMA(1,1,0), r=0



b) ARIMA(1,1,0), r=1



c) ARIMA(0,1,1), r=0



d) ARIMA(0,1,1), r=1

**Legenda:**

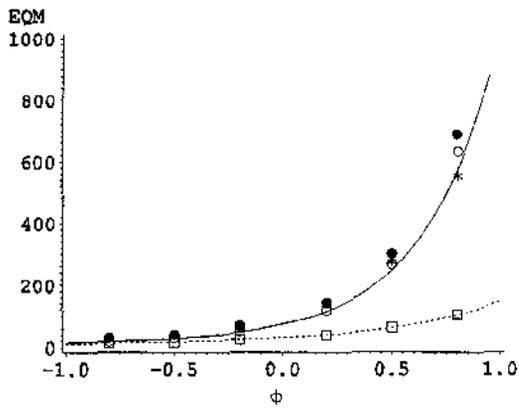
**Modelo desagregado**

- - EQM para modelos conhecidos
- - EQM simulado sem teste
- \* - EQM simulado com teste

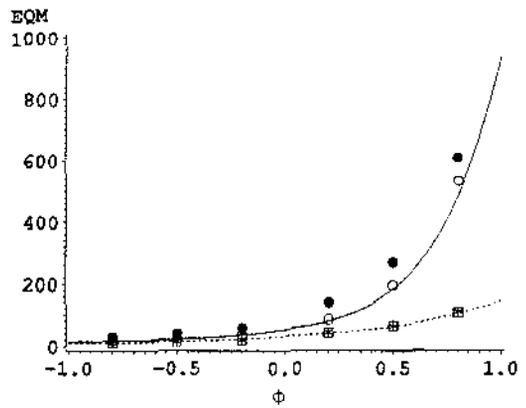
**Modelo agregado**

- - EQM para modelos conhecidos
- - EQM simulado sem teste
- - EQM simulado com teste

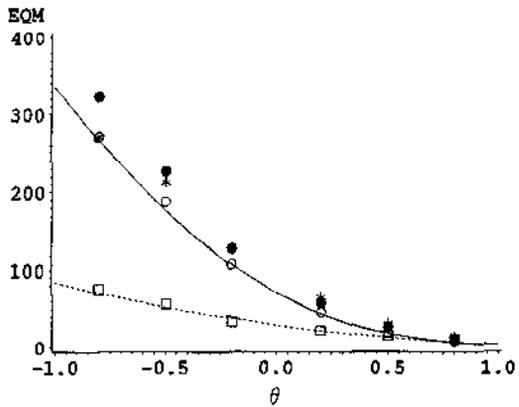
**Gráfico 4.4.15** - EQM com um IO de tamanho 1 presente r posições antes da origem de previsão, com e sem teste de detecção de outlier. Séries de tamanho 100 com 1000 replicações.



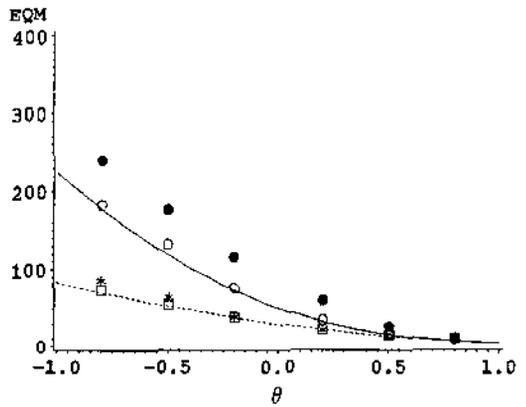
a) ARIMA(1,1,0), r = 0



b) ARIMA(1,1,0), r = 1



c) ARIMA(0,1,1), r = 0



d) ARIMA(0,1,1), r = 1

**Legenda:**

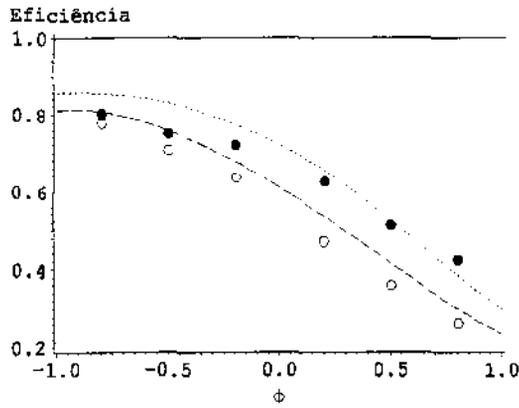
**Modelo desagregado**

- ..... EQM teórico
- EQM simulado sem teste
- \* EQM simulado com teste

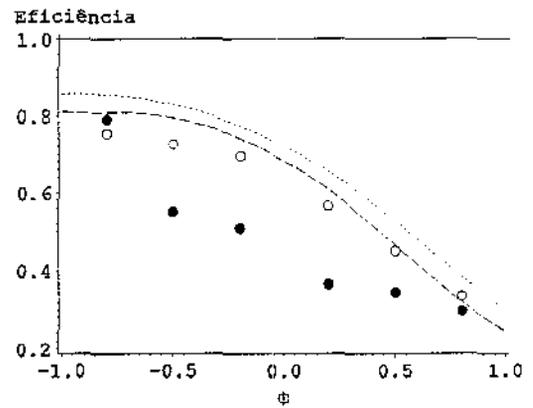
**Modelo agregado**

- EQM teórico
- EQM simulado sem teste
- EQM simulado com teste

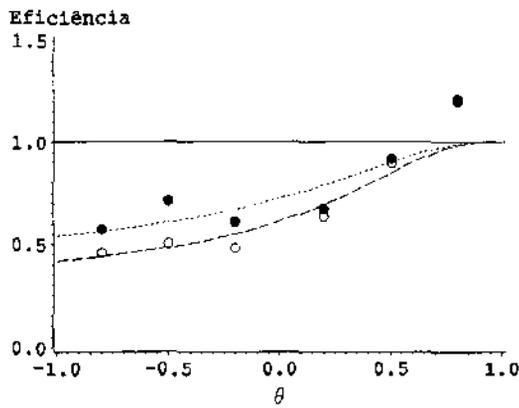
**Gráfico 4.4.16** - EQM com um IO de tamanho 2 presente r posições antes da origem de previsão nas séries de tamanho 100, com e sem teste de detecção de outlier, em 1000 replicações.



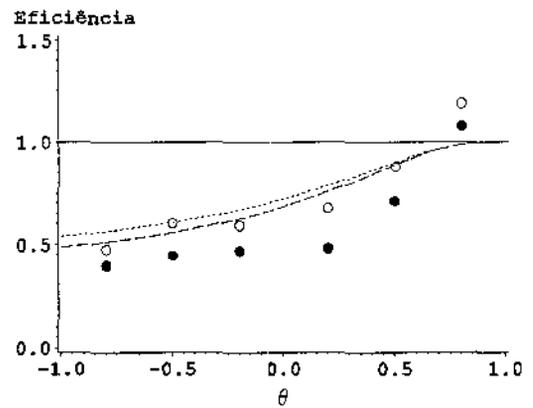
a) ARIMA(1,1,0),  $r = 0$



b) ARIMA(1,1,0),  $r = 1$



c) ARIMA(0,1,1),  $r = 0$

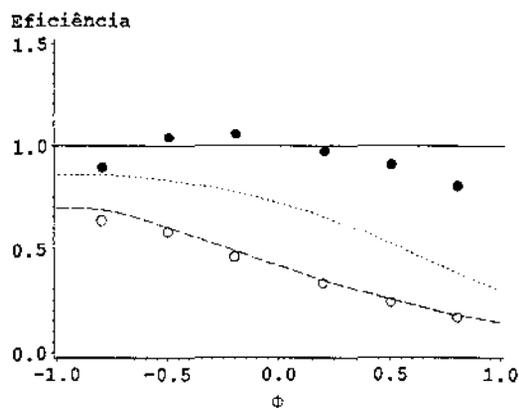
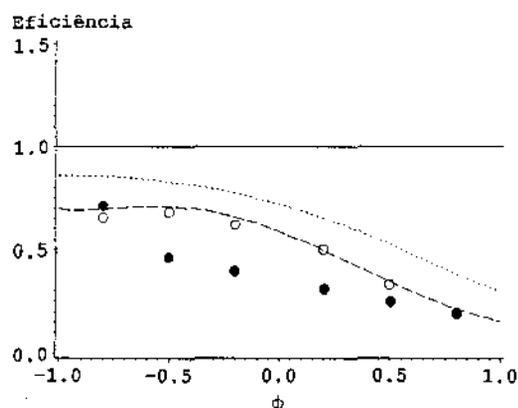
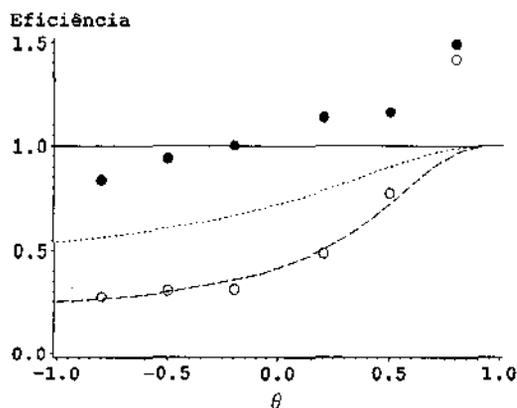
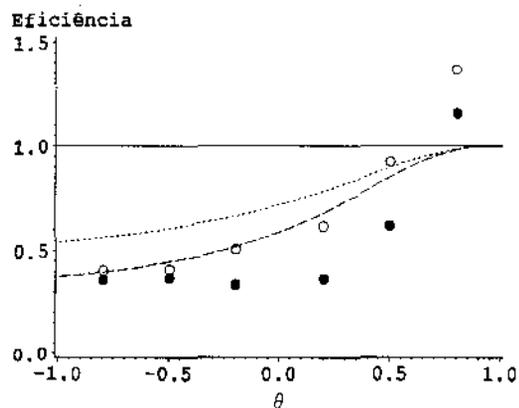


d) ARIMA(0,1,1),  $r = 1$

**Legenda:**

- .... Eficiência considerando o modelo conhecido, sem outlier
- Eficiência considerando o modelo conhecido, com outlier
- Eficiência considerando o modelo desconhecido, sem teste
- Eficiência considerando o modelo desconhecido, com teste

**Gráfico 4.4.17** - Eficiência de previsão um passo à frente de modelo agregado, considerando o modelo conhecido e desconhecido, com e sem IO de tamanho 1. Séries de tamanho 100 com 1000 replicações.

a) ARIMA(1,1,0),  $r = 0$ b) ARIMA(1,1,0),  $r = 1$ c) ARIMA(0,1,1),  $r = 0$ d) ARIMA(0,1,1),  $r = 1$ **Legenda:**

- .... Eficiência considerando o modelo conhecido, sem outlier
- Eficiência considerando o modelo conhecido, com outlier
- Eficiência considerando o modelo desconhecido, sem teste
- Eficiência considerando o modelo desconhecido, com teste

**Gráfico 4.4.18** - Eficiência de previsão 1 passo à frente de modelo agregado, considerando o modelo conhecido e desconhecido, com e sem IO de tamanho 2. Séries de tamanho 100 com 1000 replicações.

#### 4.4.2.3 - Dois IOs de tamanho 1 de mesmo sinal

##### a) sem teste de detecção

Nos Gráficos 4.4.19 (a) e (c) são ilustrados os vícios calculados considerando o modelo conhecido e os simulados da previsão para modelos ARI e IMA. Os comentários em termos de vício e de EQM são análogos ao caso de um outlier de tamanho um. Os vícios simulados, de um modo geral, são próximos dos vício do modelo conhecido, com exceção no modelo IMA com  $\theta \rightarrow +1$  (ver Gráficos 4.4.19 (a) e (c)). Os EQMs simulados são bem próximos dos valores dos EQM de modelos conhecidos.

##### b) com teste de detecção

As porcentagens de detecção de outliers são apresentadas na Tabela A5.5 A porcentagem de séries desagregadas com pelo menos um outlier detectado fica em torno de 28%, no modelo ARI e IMA. No primeiro modelo os casos mais frequentes variam conforme o valor do parâmetro, quando  $\phi \rightarrow -1$  o caso mais frequente é a detecção de apenas um outlier na última observação como AO (caso (SO,AO)), de 14,3% (quando  $\phi = -0,8$ ) e quando  $\phi$  se aproxima de +1 o caso mais frequente é a detecção de apenas um outlier na penúltima observação como IO (caso (IO,SO)). Nos outros casos de detecção a porcentagem também varia com o valor do parâmetro, enquanto a identificação dos dois outliers como AO é crescente (aumenta de 0%, quando  $\phi = -0,8$ , para 4,9%, quando  $\phi = 0,8$ ) a identificação do penúltimo como IO e o último como AO e a detecção de apenas um outlier na penúltima observação como AO são crescentes em  $\phi$ .

No modelo ARI agregado a porcentagem de detecção é decrescente em  $\phi$  (5,1% quando  $\phi = 0,8$  para 1,8% quando  $\phi = -0,8$ ), enquanto no modelo IMA agregado a porcentagem de outliers detectados é crescente no parâmetro (3,6% quando  $\theta = -0,8$  para 9,3% quando  $\theta = 0,8$ ).

No modelos desagregados, os outliers sempre detectados com algum erro, pois não existe identificação correta dos dois outliers (o último, quando detectado sempre é identificado como AO), assim suspeita-se que tenha um aumento no vício ao aplicar o teste, mas isso não acontece, pois apenas quando o outlier da última observação é detectado o vício é grande (em módulo), nos outros casos a estimação dos outliers é muito boa, em geral, não afeta muito o vício. No caso dos outliers detectados apenas na última observação, eles são reestimados por SWEEP, assim o valor dos outliers é bastante alto, em torno de 2 ou mais. Por se tratar de outliers detectados na última observação é sempre identificado como AO.

Devido a isso o EQM do modelo IMA desagregado com  $\theta \rightarrow -1$  aumenta muito, fazendo aumentar a eficiência quando  $\theta < 0$ , mas diminui quando  $\theta > 0$  (ver Gráfico 4.4.21 (b)).

No modelo ARI, os comentários em termos de vício e EQM são análogos ao caso de um outlier de tamanho 1, com algumas modificações na magnitude dos vícios que agora é a soma dos vícios para  $r = 1$  e  $r = 0$ . Em consequência do aumento do vício o EQM também aumenta. O efeito da aplicação dos testes é pequeno na eficiência, diminui um pouco quando  $\phi < 0$  e aumenta um pouco quando  $\phi > 0$  (ver Gráficos 4.4.21 (a)).

#### 4.4.2.4 - Dois IOs de tamanho 1 de sinais trocados

##### a) sem teste de detecção de outlier

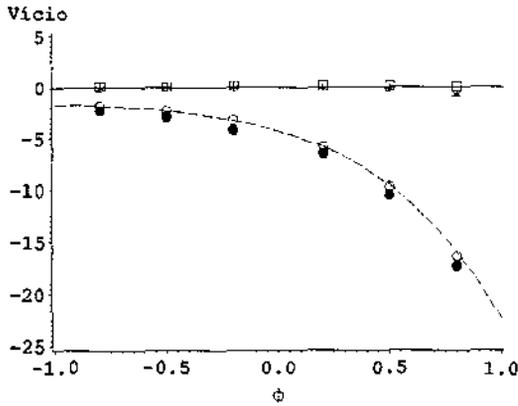
No caso de presença de dois IOs, os vícios simulados estão, em geral, bem pertos dos vícios calculados considerando o modelo conhecido, com exceção no modelo IMA desagregado quando  $\theta \rightarrow -1$ . Assim, os comentários dos casos anteriores continuam válidos, os EQMs considerando o modelo agregado conhecido sempre são maiores que no desagregado (ver Gráficos 4.4.17 (b) e (d) e Gráficos 4.4.21 (b) e (d)) e as eficiências também são bem próximas dos valores calculados considerando o modelo conhecido.

##### b) com teste de detecção de outlier

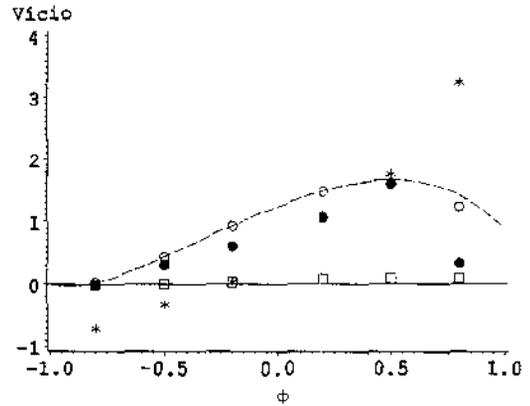
A porcentagem de pelo menos um outlier detectado é em torno de 30% para ambos os modelos desagregados ARI e IMA. Dentre os que detectaram pelo menos um outlier, a porcentagem de outliers detectados na penúltima observação, no modelo ARI, é crescente em  $\phi$ , variando de 69,5% a 93,1%, e a porcentagem de identificação correta na penúltima observação também é crescente em  $\phi$ , variando de 12,1% a 30,6%. No modelo IMA, a relação de porcentagem de detecção e de identificação correta com os parâmetros do modelo é inversa, com as porcentagens bastante próximas. A porcentagem de outliers detectados na última observação é em torno de 40% para ambos os modelos desagregados, ARI e IMA.

No modelo agregado, a porcentagem de detecção foi em torno de 3%, variando pouco com o valor dos parâmetros, nos modelos ARI e IMA.

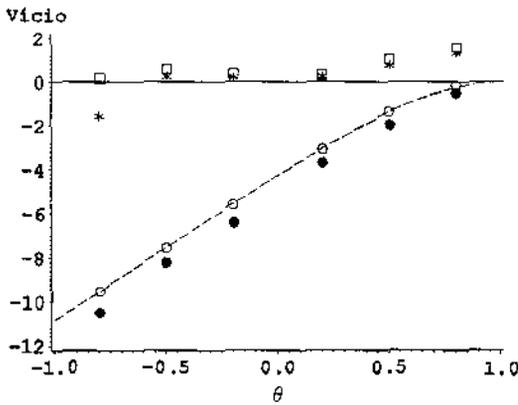
A aplicação do teste de detecção aumenta consideravelmente o EQM nos modelos agregados, principalmente para valores de  $\phi$  próximo a 1 no modelo ARI e  $\theta$  próximo a -1 no modelo IMA, sendo neste último caso o efeito maior no modelo desagregado. Com esses efeitos a eficiência aumenta quando  $|\phi|$  ou  $|\theta|$  tende para 1.



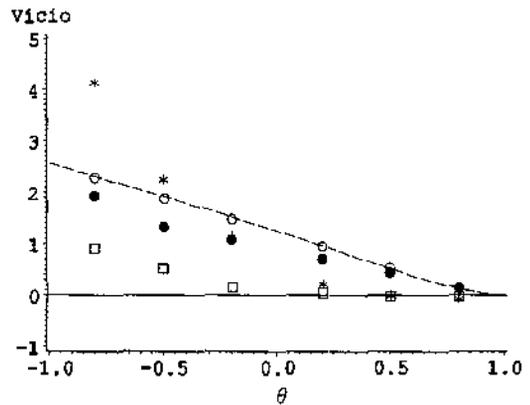
a) Vício causado por dois IOs de tamanho 1 de mesmo sinal no modelo ARIMA(1,1,0)



b) Vício causado dois IOs de tamanho 1 de sinais trocados, no modelo ARIMA(1,1,0).



c) Vício causado por dois IOs de tamanho 1 de mesmo sinal no modelo ARIMA(0,1,1)



d) Vício causado por dois IOs de tamanho 1 de sinais trocados, no modelo ARIMA(0,1,1).

**Legenda:**

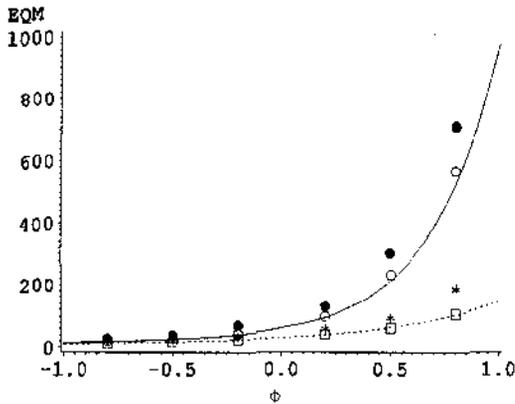
**Modelo desagregado**

- - vício para modelo conhecido
- - vício simulado sem teste
- \* - vício simulado com teste

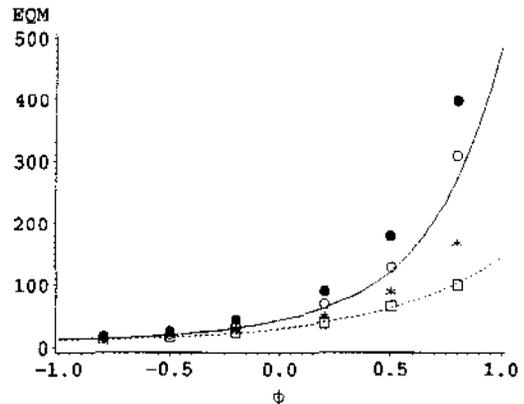
**Modelo agregado**

- - vício para modelo conhecido
- - vício simulado sem teste
- - vício simulado com teste

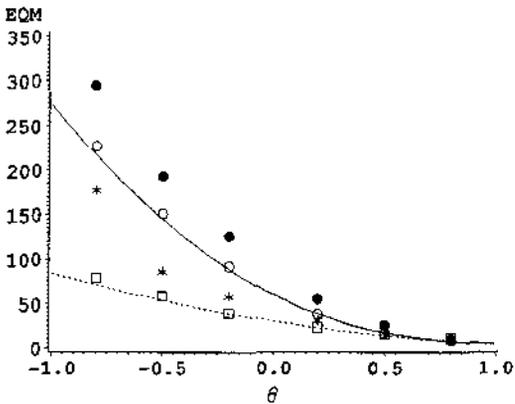
**Gráfico 4.4.19** - Vício causado por dois IOs nas duas últimas observações de séries com 100 observações em 1000 replicações.



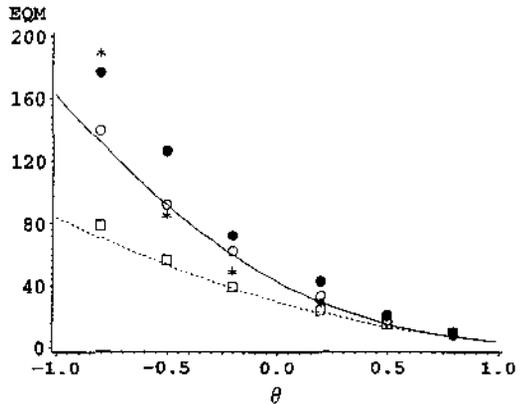
a) dois IOs de tamanho 1 de mesmo sinal, no modelo ARIMA(1,1,0).



b) dois IOs de tamanho 1 de sinais trocados, no modelo ARIMA(1,1,0)



c) dois IOs de tamanho 1 de mesmo sinal, no modelo ARIMA(0,1,1).



d) dois IOs de tamanho 1 de sinais trocados, no modelo ARIMA(0,1,1).

**Legenda:**

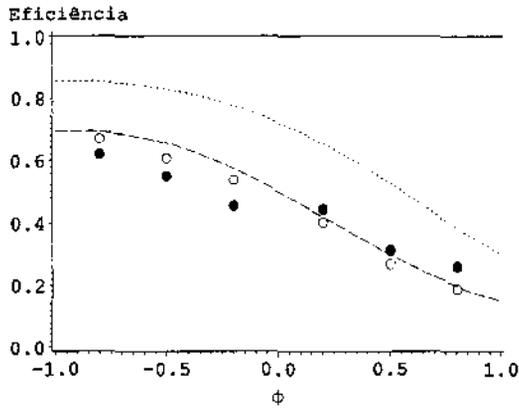
**Modelo desagregado**

- ..... - EQM para modelo conhecido
- - EQM simulado sem teste
- \* - EQM simulado com teste

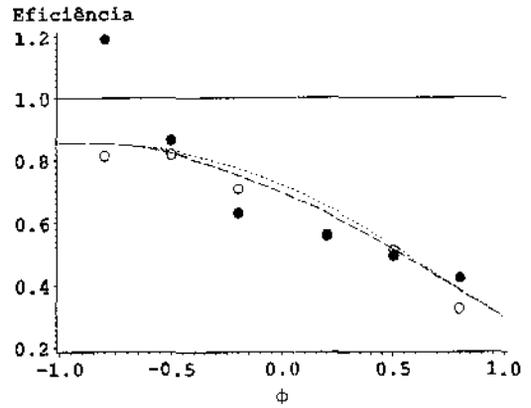
**Modelo agregado**

- - EQM para modelo conhecido
- - EQM simulado sem teste
- - EQM simulado com teste

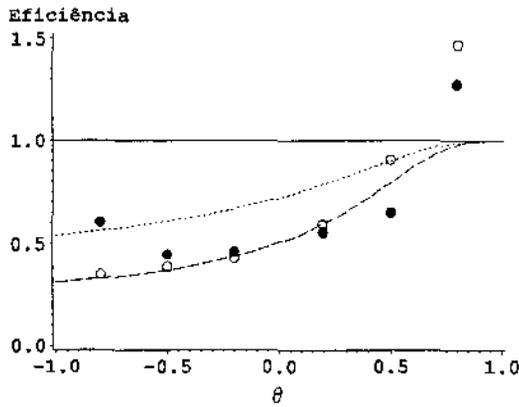
**Gráfico 4.4.20** - EQM de previsão um passo à frente, com dois IOs nas duas últimas observações da série de tamanho 100, em 1000 replicações.



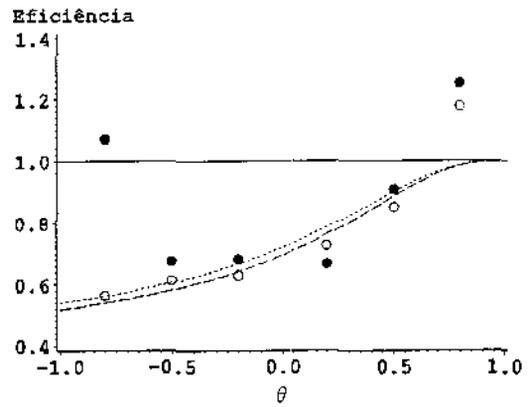
a) dois IOs de tamanho 1, de mesmo sinal, no modelo ARIMA(1,1,0).



b) dois IOs de tamanho 1, de sinais trocados, no modelo ARIMA(1,1,0).



c) dois IOs de tamanho 1 de mesmo sinal, no modelo ARIMA(1,1,0).



d) dois IOs de tamanho 1 de sinais trocados, no modelo ARIMA(1,1,0).

**Legenda:**

- .... Eficiência considerando o modelo conhecido, sem outlier
- Eficiência considerando o modelo conhecido, com outlier
- Eficiência considerando o modelo desconhecido, sem teste
- Eficiência considerando o modelo desconhecido, com teste

**Gráfico 4.4.21** - Eficiência de previsão de modelo agregado (um passo à frente), com e sem IO de tamanho 1 de mesmo sinal ou sinais trocados, nas duas últimas observações de séries de tamanho 100, em 1000 replicações.

## 4.5 - Conclusões

A presença de outliers na série desagregada pode ter um efeito bastante grande na eficiência de previsão de valores agregados.

Os estudos mostram que a presença de outliers aditivos (AO), em geral, são favoráveis no modelo agregado, havendo vários casos em que a eficiência passa a ser maior do que 1, principalmente quando o outlier é grande. A utilização do teste de detecção de outliers diminui o efeito dos outliers em termos de vício nos modelos agregados e desagregados; em geral, ele diminui um pouco o EQM nos modelos desagregados e aumenta nos modelos agregados, diminuindo-se assim a eficiência dos modelos agregados. Mesmo assim, em alguns casos (por exemplo, outlier de tamanho 2 presente na última observação num processo  $ARIMA(0,1,1)$ ) a eficiência é maior do que 1.

Quando o modelo é conhecido a presença de um outlier de inovação (IO) não afeta as previsões no modelo desagregado apenas diminui a eficiência da previsão utilizando o modelo agregado, esta situação é pouco alterada quando não são utilizados testes de detecção de outliers nos modelos desconhecidos. Utilizando-se testes de detecção de outlier a eficiência do modelo agregado aumenta quando o outlier ocorre na última observação, isso decorre do fato de IO ser incorretamente classificado como AO quando detectado. Neste caso a eficiência do modelo agregado passa a ser maior do que 1 em algumas situações. No entanto, quando o outlier ocorre na penúltima observação a porcentagem de casos com classificação incorreta do outlier diminui e o efeito do teste é bastante pequeno no vício e EQM do modelo desagregado. Já que no modelo agregado o efeito ainda pode ser significativo, o EQM é aumentado diminuindo, portanto, a eficiência.

De forma geral, os testes de detecção de outlier diminuem o vício, mas podem aumentar o EQM, principalmente quando o outlier é pequeno. Além disso, o efeito do outlier, com ou sem teste de detecção geralmente aumenta a eficiência do modelo agregado quando o outlier é AO, e diminui a eficiência quando o outlier é IO.

# Apêndice 1

## Teste de Razão de Verossimilhança

Sejam  $X_1, \dots, X_n$ , v.a. i.i.d com função de densidade de probabilidade (f.d.p)  $f(\cdot; \theta)$ ,  $\theta \in \Omega \subseteq \mathbb{R}^r$ , e  $x_1, \dots, x_n$  uma realização das variáveis aleatórias, a função de verossimilhança é definida como

$$L(\theta/x_1, \dots, x_n) = L(\theta/\mathbf{x}) = f(\mathbf{x}/\theta) = \prod_{i=1}^n f(x_i/\theta). \quad (\text{A1.1})$$

O interesse é testar a hipótese  $H_0: \theta \in \Theta_0$  contra  $H_1: \theta \in \Theta_0^c$ , onde  $\Theta_0 \subset \Omega$  e  $\Theta_0^c$  é complementar em relação a  $\Omega$ . Quando ambos  $\Theta_0$  e  $\Theta_0^c$  consistem de apenas uma observação, então  $L(\Theta_0/\mathbf{x})$  e  $L(\Theta_0^c/\mathbf{x})$  são completamente determinadas, o Teste MP<sup>1</sup> rejeita a hipótese nula quando a razão de verossimilhança (RV)  $L(\Theta_0^c)/L(\Theta_0)$  é muito grande (maior que um valor  $c$  especificado pelo nível de significância do teste), este teste é conhecido como Teste de Neyman-Pearson. No entanto, se  $\Theta_0$  e  $\Theta_0^c$  contêm mais que um ponto cada, então tanto  $L(\Theta_0^c)$  como  $L(\Theta_0)$  são determinados por  $H_0$  e  $H_1$  e o método de teste anterior não é aplicado. O problema pode ser reduzido pela seguinte definição.

**Definição 1** - A estatística do teste de razão de verossimilhança para testar  $H_0: \theta \in \Theta_0$  contra  $H_1: \theta \in \Theta_0^c$  é

---

<sup>1</sup>Seja  $C$  uma classe de testes para testar  $H_0: \theta_0$  contra  $H_1: \theta_1$ . Um teste na classe  $C$  com função poder  $\beta(\theta) = P_\theta(X \in R)$ , onde  $R$  é a região de rejeição, é *Teste Mais Poderoso* (TMP) se  $\beta(\theta_1) \geq \beta'(\theta_1)$  para todo  $\beta'(\theta_1)$  que é uma função poder de um teste na classe  $C$ .

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta / \mathbf{x})}{\sup_{\theta \in \Omega} L(\theta / \mathbf{x})} \quad (\text{A1.2})$$

Um teste de *razão de verossimilhança* (RV) é qualquer teste que tem a região de rejeição da forma  $\{\mathbf{x}; \lambda(\mathbf{x}) \leq c\}$ , onde  $c$  é um número que satisfaz  $0 \leq c \leq 1$  e é selecionado de modo que o teste tenha um nível de significância desejado, i.e.,  $P(\lambda(\mathbf{x}) \leq c/H_0) = \alpha$ .

O raciocínio do teste de RV pode ser melhor entendido na situação em que  $f(\mathbf{x}/\theta)$  é uma f.d.p. de uma variável discreta. Neste caso, o numerador de  $\lambda(\mathbf{x})$  é a probabilidade máxima da amostra observada sob parâmetros na hipótese nula. O denominador  $\lambda(\mathbf{x})$  é a probabilidade máxima de observar a amostra sob todos os possíveis parâmetros. A razão desses dois máximos é pequena se existem pontos no parâmetro na hipótese alternativa para a qual a amostra é muito mais provável que qualquer outro ponto na hipótese nula. Nesta situação, o critério de teste de RV diz se  $H_0$  deve ser rejeitado. O valor é encontrado de acordo com o nível de significância especificado.

Pensando-se em fazer maximização sob os dois espaços de parâmetros: inteiros (maximização irrestrita) e um subconjunto do espaço paramétrico (maximização restrita) a correspondência de teste de MV e EMV<sup>2</sup> se torna mais clara. Suponha que  $\theta^*$ , um EMV de  $\theta$ , exista;  $\theta^*$  é obtido fazendo maximização irrestrita de  $L(\theta/\mathbf{x})$ . Considerando-se também o EMV de  $\theta$ , denotando-o por  $\theta_0^*$ , obtido pela maximização restrita, assumindo  $\Theta_0$  o espaço de parâmetros, isto é,  $\theta_0^* = \theta_0^*(\mathbf{x})$  é o valor de  $\theta \in \Theta_0$  que maximiza  $L(\theta/\mathbf{x})$ , então a estatística do teste de RV é

$$\lambda(\mathbf{x}) = \frac{L(\theta_0^*/\mathbf{x})}{L(\theta^*/\mathbf{x})} \quad (\text{A1.3})$$

Sendo  $\lambda$  um quociente de funções não negativas,  $\lambda \geq 0$ . Sendo ainda  $\Theta_0 \in \Omega$  tem-se que  $L(\theta_0^*) \leq L(\theta^*)$  e desta forma  $\lambda \leq 1$ . Assim,  $0 \leq \lambda \leq 1$ . Se o máximo de  $L$  em  $\Theta_0$  é muito menor que em  $\Omega$ , deve parecer que os dados  $x_1, \dots, x_n$  não suportam a hipótese  $H_0: \theta \in \Theta_0$ . Isto é, valores pequenos da razão  $\lambda$  deve levar à rejeição de  $H_0$ . Por outro lado, um valor da razão  $\lambda$  que é perto de 1 deve suportar a hipótese nula  $H_0$ .

---

<sup>2</sup> Para cada ponto amostral  $\mathbf{x}$ , seja  $\theta^*(\mathbf{x})$  um valor do parâmetro em que  $L(\theta/\mathbf{x})$  atinge o seu máximo como uma função de  $\theta$ , em que  $\mathbf{x}$  é mantido fixo. Um Estimador de Máxima Verossimilhança (EMV) de um parâmetro  $\theta$  baseado em uma amostra  $X$  é  $\theta^*(X)$ .

O teste pelo método de RV pela sua explícita definição da estatística do teste (A1.2), e pela sua explícita região de rejeição  $\{X; \lambda(X) \leq c\}$  é um dos métodos mais usados para modelos complicados. Após os dados serem observados  $X = \mathbf{x}$  a função de verossimilhança,  $L(\theta/\mathbf{x})$ , é uma função completamente definida da variável  $\theta$ . Mesmo que os dois supremos de  $L(\theta/\mathbf{x})$ , sobre os conjuntos  $\Theta_0$  e  $\Omega$ , não possam ser obtidos analiticamente, eles geralmente podem ser calculados numericamente. Assim, o teste estatístico  $\lambda(\mathbf{x})$  pode ser obtido para o conjunto de dados observado mesmo que não tenha disponível fórmula conveniente para definir  $\lambda(\mathbf{x})$ .

Para definir um teste de tamanho  $\alpha$ , a constante  $c$  deve ser escolhida de forma que

$$\sup_{\theta \in \Theta_0} P_{\theta}(\lambda(\mathbf{x}) \leq c) \leq \alpha \tag{A1.4}$$

Se não pode ser derivada uma fórmula simples de  $\lambda(\mathbf{x})$ , deve parecer impossível derivar a distribuição amostral de  $\lambda(X)$  e assim encontrar  $c$  para garantir (A1.4). O seguinte Teorema leva a garantir que (A1.4) é verdade, pelo menos para amostras grandes. Uma discussão completa desse tópico pode ser encontrado em Kendall e Stuart (1979).

**Teorema.** Seja  $x_1, \dots, x_n$  uma a.a. de uma f.d.p  $f(\mathbf{x}/\theta)$ . Sob algumas condições de regularidade<sup>3</sup> no modelo  $f(\mathbf{x}/\theta)$ , se  $\theta \in \Theta_0$  então a distribuição da estatística  $-2\log\lambda(\mathbf{x})$  converge à uma distribuição qui-quadrada quando o tamanho da amostra  $n \rightarrow \infty$ . Os graus de liberdade da distribuição limite é a diferença entre o número de parâmetros livres especificados por  $\theta \in \Theta_0$  e o número de parâmetros especificados por  $\theta \in \Omega$ .

As 'condições de regularidades' necessárias para o modelo são condições gerais que são satisfeitas para muitas distribuições. Essas condições são principalmente relacionadas com a existência e comportamento das derivadas (com respeito aos parâmetros) da função de verossimilhança, e o suporte da distribuição (ele não pode depender dos parâmetros).

Rejeição de  $H_0: \theta \in \Theta_0$  para valores pequenos de  $\lambda(\mathbf{x})$  é equivalente à rejeição para valores grandes de  $-2\log\lambda(\mathbf{x})$ . Assim

$$H_0 \text{ é rejeitada se e somente se } -2\log\lambda(X) \geq \chi_{v,\alpha}^2,$$

---

<sup>3</sup> Condições de Regularidade: (a) o espaço de parâmetros,  $\Omega$  tem dimensão finita, é fechado e compacto e o parâmetro verdadeiro está contido no interior de  $\Omega$ ; (b) as distribuições de probabilidade definidas para quaisquer dois diferentes valores de  $\theta$  são distintas; (c) o suporte não depende de  $\theta$ ; (d) as derivadas terças de log-verossimilhança existem.

onde  $\nu$  é o grau de liberdade especificado no Teorema. A probabilidade de Erro do Tipo I<sup>4</sup> deve ser aproximadamente satisfeita para tamanhos de amostra grandes e um teste assintótico de tamanho  $\alpha$  foi definido. Note que o teorema deve apenas implicar que

$$\lim_{n \rightarrow \infty} P_{\theta}(\text{rejeitar } H_0) = \alpha \quad \text{para cada } \theta \in \Theta_0,$$

não que  $\sup_{\theta \in \Theta_0} P_{\theta}(\text{rejeitar } H_0)$  converge para  $\alpha$ . Isso é usualmente o caso de um teste assintótico de tamanho  $\alpha$ .

Geralmente o cálculo do grau de liberdade é direta. Muitas vezes  $\Omega$  pode ser representado num subconjunto de espaço euclidiano  $q$ -dimensional que contém um subconjunto aberto em  $\mathbb{R}^p$ , onde  $p < q$ . Então  $q - p = \nu$  é o grau de liberdade do teste estatístico.

---

<sup>4</sup> $P_{\theta}(X \in RR/H_0) = \text{Erro do Tipo I}$ , onde RR é a região de rejeição.

## Apêndice 2

### Teste de Escore

Será apresentado aqui um breve resumo do teste de escore. Suponha que  $\theta \in \Theta$  seja um vetor de parâmetros  $m \times 1$  e

$$H_0: h_j(\theta) = 0 \quad (j = 1, 2, 3, \dots, s, \theta \in \Theta)$$

são  $s$  restrições aplicadas no espaço de parâmetros  $\Theta$ . Considerando agora a função

$$l(\theta) + \sum_{j=1}^s \lambda_j h_j(\theta), \quad (\text{A2.1})$$

onde  $l(\theta)$  é a função de log-verossimilhança e  $\lambda_j$  ( $j = 1, 2, 3, \dots, s$ ) são  $s$  multiplicadores de Lagrange.

Diferenciando (A2.1) com respeito aos parâmetros desconhecidos  $\theta$  e  $\lambda_j$  ( $j = 1, 2, 3, \dots, s$ ) e igualando as derivadas à zero sob  $H_0$  tem-se que

$$\tilde{\mathbf{D}} + \tilde{\mathbf{H}}\tilde{\boldsymbol{\lambda}} = \mathbf{0} \quad (\text{A2.2})$$

e

$$h_j(\tilde{\theta}) = 0 \quad (j = 1, 2, 3, \dots, s),$$

onde

$$\tilde{\mathbf{D}} = \left\{ \frac{\partial l(\tilde{\theta})}{\partial \theta_i} \right\}_{m \times 1}, \quad \tilde{\mathbf{H}} = \left\{ \frac{\partial h_j(\tilde{\theta})}{\partial \theta_i} \right\}_{m \times s}.$$

$\tilde{\boldsymbol{\lambda}}$  é um vetor  $s \times 1$  de Multiplicadores de Lagrange. O til indica que os parâmetros foram obtidos sob  $H_0$ .

A motivação em usar (A2.1) vem do fato de que quando a hipótese nula é verdadeira, a estimativa restrita  $\tilde{\theta}$  deve se aproximar do estimador de máxima verossimilhança sem restrição, de forma que  $\tilde{\mathbf{D}}$  deve ser zero.

Pela equação (A2.2) tem-se que

$$\tilde{\mathbf{D}}'\tilde{\mathbf{T}}^{-1}\tilde{\mathbf{D}} = \tilde{\lambda}'\tilde{\mathbf{H}}'\tilde{\mathbf{T}}^{-1}\tilde{\mathbf{H}}\tilde{\lambda}, \quad (\text{A2.3})$$

onde  $\tilde{\mathbf{T}}$  é a matriz de informação de Fisher calculada sob  $H_0$ . A expressão do lado esquerdo de (A2.3) é conhecido como “estatística de escore” (Rao, 1973) e o do lado direito é conhecida como estatística de multiplicador de Lagrange (Silvey, 1959). Foi mostrado que o lado direito, e conseqüentemente o lado esquerdo da expressão assintoticamente segue uma distribuição qui-quadrada de  $s$  graus de liberdade quando a hipótese nula é verdadeira (Silvey, 1959). Hosking (1980) e Godfrey (1979) discutiram previamente os testes de multiplicador de Lagrange para testar a adequação de modelos em séries temporais.

## Apêndice 3

### Critérios de Informação

Um procedimento natural para comparara dois modelos é através de suas funções de verossimilhança. Se o modelo 1, com  $k_1$  parâmetros e função de verossimilhança  $L_1$  é conhecida e ele é comparado com o modelo 2, que tem  $(k_2 - k_1)$  parâmetros adicionais (supondo que  $k_2 > k_1$ ) e função de verossimilhança  $L_2$  conhecida (Kendall e Stuart, 1979, seção 24.7) que o negativo de duas vezes a log-verossimilhança da razão segue assintoticamente uma distribuição qui-quadrada com  $(k_2 - k_1)$  graus de liberdade. Assim, para amostras grandes:

$$E(-\ln(L_1/L_2)) = k_2 - k_1. \quad (\text{A3.1})$$

A comparação entre log-verossimilhança deve ser levada em consideração, isso levou Akaike (1974) a propor um critério de informação (AIC) da forma

$$\text{AIC} = -2 \ln(L) + 2 k, \quad (\text{A3.2})$$

onde  $k$  denota os números de parâmetros no modelo. O modelo com menor AIC é considerado melhor no sentido de minimizar o EQMP (Erro Quadrático Médio de Previsão). No entanto, foi verificado por Schwartz (1978) e outros que AIC não é um critério consistente, de forma que ele não seleciona o modelo verdadeiro com probabilidade tendendo a 1 quando  $n \rightarrow \infty$ . Para contornar esse problema, Schwartz propôs uma critério de informação bayesiana (BIC),

$$\text{BIC} = 2 \ln(L) + k \ln(n). \quad (\text{A3.3})$$

O BIC tende a incluir menos termos que AIC dado que o termo de penalidade é maior. Esse critério é também chamado SBC (Schwartz Bayesian Criterion) na literatura.

## Apêndice 4

### Programa

Programa para simular séries ARIMA(1,1,0) ou ARIMA(0,1,1) e fazer previsões h passos a frente de séries desagregadas e agregadas, com período de agregação  $m = 4$ .

As séries podem estar contaminadas no máximo com dois outliers nas duas últimas posições da série.

O primeiro outlier  $o_1$  na posição,  $k_1$  e  $o_2$  na posição  $k_2$  do tipo AO ou IO.

#### **Entrada:**

$\phi, \theta$  → parâmetros do modelo ARIMA

$n$  → tamanho da série

$nr$  → número de replicações

$o_1, o_2$  → tamanhos dos outlier1 e outlier2

$k_1, k_2$  → posições dos outlier1 e outlier2

$t_1, t_2$  → tipos dos outlier1 e outlier2, onde 0 → sem outlier, 1 → IO, 2 → AO

$h$  → número de previsões a frente

$r$  → cálculo da variância das inovações do tipo usual (0) ou robusta (1)

$cd, ca$  → pontos de corte para testes de detecção de outliers nas séries desagregadas e agregadas

**Saída:** dois arquivos ef.txt e nout.txt, onde

1) ef.txt contém os seguintes valores:

theta, phi → os parâmetros do modelo

efds, efas → efeitos dos outlier na previsão pelo modelo desagregado e agregado, respectivamente

eftds, eftas → efeitos dos outlier na previsão com aplicação pelo modelo de testes de detecção de outlier nos modelos desagregado e agregado, respectivamente

eq21s, eq22s → EQMP pelas previsões através dos modelos desagregado e agregado, respectivamente

eqt21s, eqt22s → EQMP pelas previsões com aplicação dos testes de detecção de outlier através dos modelos desagregado e agregado, respectivamente

2) nout.txt contém os seguintes valores:

theta, phi → os parâmetros do modelo

n\_outd, n\_outa → número de séries com pelo menos um outlier detectado nas séries desagregadas e agregadas

nd1 → número de séries desagregadas com outlier detectado na posição k1

nd12 → número de séries com dois outliers detectados

naod1 → número de séries desagregadas com outlier do tipo AO na posição k1

n12 → número de séries desagregadas com dois outliers detectados, o primeiro do tipo IO e o segundo do tipo AO

```

options ps=60 ls=70;
*filename dados 'c:\users\rissa\ef.txt';
*filename nout 'c:\users\rissa\nout.txt';
filename dados disk 'ef.txt al';
filename nout disk 'nout.txt al';
/* os dois primeiros filenames sao para rodar no micro e os dois
seguintes para rodar na estacao IBM */

%MACRO SIMULA(PHI, THETA, N, NR, O1, O2, K1, K2, T1, T2, H, R, Cd, ca);
Cd=&Cd; * ponto de corte desagregado;
Ca=&Ca; * ponto de corte agregado;
phi=&phi; * par. do modelo;
theta=&theta;
o1=&o1; * tamanho do outlier1 real;
o2=&o2; * tamanho do outlier2 real;
t1=&t1; * tipo do outlier1;
t2=&t2; * tipo do outlier2;
r=&r; * calculo de variancia da inovacao;
nr=&nr; * numero de replicacoes;
n=&n; * tamanho da serie;
k1=&k1; * posicao do outlier1;
k2=&k2; * posicao do outlier2;
h=&h; * numero de previsoes a frente;
pm=5; * numero max. de ordem de AR serie desag;
pma=4; * numero max. de ordem de AR serie agreg;
n0=100; * as primeiras n0 obs sao eliminadas;

** criando as matrizes **;
erro=j(2,2,0);
prev=j(2,2,0); * prev sem considerar o(s) outlier(s);
prevd=j(2,2,0); * prev de ser desag testando os out;
prevat=j(2,2,0); * prev de ser agreg testando os out;
var=j(2,2,0); * var dos erros de previsao de prev;
vard=j(2,2,0); * var dos erros de previsao de prevdt;
varat=j(1,2,0); * var dos erros de previsao de prevat;
vart=j(2,1,0);
med=j(2,2,0); * media dos erros de previsao de prev;
meddt=j(2,2,0); * media dos erros de previsao de prevdt;
medat=j(1,2,0); * media dos erros de previsao de prevat;
medts=j(2,1,0);
eqmts=j(2,2,0);

seed=int(abs(&theta)*&n*&k1/&k2+&theta*100+&n*10+&nr+10*&o1+&cd+&ca+
20*&o2+&k1**2+&k2+123*&phi+&t1*1111+&t2*989898-&h*0.1);
* semente inicial;

*-----*
* Inicio do Programa Principal *
*-----*

n_outd=0; * num de séries desag com >= 1 outlier detectado;
nd1=0; * num de séries desag com outlier detectado na pos k1;
nd2=0; * num de séries desag com outlier detectado na pos k2;
nd12=0; * num de séries desag com os dois outliers detectados;
nl2=0; * num de séries desag com IO na pos k1 e AO na pos k2;
n_outa=0; * num de séries agreg com um outlier detectado;
naod1=0; * num de séries desag com AO detectado na pos k1;

```

```

%do l=1 %to &nr;
*-----*
*          SERIE SEM OUTLIER          *
*-----*

*----- SERIE DESAGREGADA -----*
m=4;
hd=h*m;
id1=0;          * indicador de outlier o1;
id2=0;          * indicador de outlier o2;
nod=id1+id2;    * numero de outliers detectados;
run sim(y,z,n0,&n,&phi,&theta,hd,seed);          * SIMULA OS DADOS;
run montas(&n,pm,&k1,&k2,z,s,id1,id2);          * MONTA A MATRIZ PARA;
run bic(&n,s,pm,paryd1,parzd1,nod,out);          * CALCULA P PELO BIC;
run previsao(h,m,y,n,&k1,&k2,0,0,paryd1,parzd1,out,prevd1);
* CALCULA PREVISAO;

*----- SERIE AGREGADA -----*
na=int(n/m);
run agrega(h,na,y,ya,za,m,k1,ka);          * AGREGA EM m PERIODOS;
obs1=ya[na+h];
ia=0;          * INDICADORES DE OUTLIERS O1 E/OU O2;
ta=0;          * INDICADOR DO TIPO DE OUTLIER;
kal=ka;
run montas(na,pma,ka,kal,za,s,ia,0);          * MONTA A MATRIZ PARA SWEEP;
run bic(na,s,pma,paryal,parzal,ia,out);          * CALCULA PA PELO BIC;
run previsao(h,1,ya,na,ka,kal,0,0,paryal,parzal,out,preval);
* CALCULA AS PREVISOES;

*-----*
*          SERIE COM OUTLIER          *
*-----*

*----- SERIE DESAGREGADA -----*
* Inserindo outlier(s);

if t1=2 then do;
  if (t2=2) | (t2=0) then run ao(n,y,z,&k1,&k2,o1,o2);
  if t2=1 then do;
    run ao(n,y,z,&k1,&k2,o1,0);
    run io(n,y,z,&k1,&k2,0,o2,&phi,&theta,h);
  end;
end;
if t1=1 then do;
  if t2=2 then do;
    run ao(n,y,z,&k1,&k2,0,o2);
    run io(n,y,z,&k1,&k2,o1,0,&phi,&theta,h);
  end;
  if t2<2 then run io(n,y,z,&k1,&k2,o1,o2,&phi,&theta,h);
end;
pre2=y[n+1:n+hd];

run agrega(h,na,y,ya,za,ka,k1,ka);          * AGREGA EM M PERIODOS;
obs2=ya[na+h];

*----- SEM TESTE DE DETECCAO DE OUTLIER -----*
run montas(&n,pm,&k1,&k2,z,s,id1,id2);          * MONTA A MATRIZ PARA BIC;
run bic(&n,s,pm,paryd2,parzd2,nod,out);          * CALCULA P PELO BIC;
run previsao(h,m,y,n,&k1,&k2,0,0,paryd2,parzd2,out,prevd2);

```

```

*----- COM TESTES DE DETECCAO DE OUTLIER -----*;

* td=0 (sem outlier)      td=1 (IO)      td=2 (AO);
run sigma(&n, y, paryd2, &k1, &k2, aest, sigma, R);
* CALCULA A VARIANCIA DA INOVACAO;
run teste(cd, &n, y, paryd2, &k1, aest, sigma, ld1, td1);
* TESTE DE DETECCAO DE OUTLIER NA POSICAO K1;
IF (td1>0) then do;
  if td1=2 then naod1=naod1+1;
  id1=1;
  nd1=nd1+1;
  OUT=LD1*SIGMA;
  OUT1=OUT;
end;
td2=0;
if &o2^=0 then do;
  run teste(cd, &n, y, paryd2, &k2, aest, sigma, ld2, td2);
  * TESTE DE DETECCAO DE OUTLIER NA POSICAO K2;
  if td2=2 then do;
    nd2=nd2+1;
    id2=1;
    OUT2=LD2*SIGMA;
  end;
end;
nod=id1+id2;

prevtd=prevd2;
paryd3=paryd2;
parzd3=parzd2;

if nod>0 then do;
  n_outd=n_outd+1;

  /* ESTIMACAO NAO LINEAR DOS PARAMETROS, QUANDO E DETECTADO UM AO NA
  PENULTIMA OBSERVACA */
  if (td1=2) & (&k1=n-1) then do;
    npar=nrow(parzd3);
    if td2=2 then out=out|| (ld2*sigma);
    run matx(k1, z, parzd3, x, id2);
    run nlsfit(z, x, out, parzd3, criterio, 10, 0.001, 0.001, 1);
    parzd3=parzd3[npar:1];
    run parzy(parzd3, paryd3);
    parzd3=parzd3[npar:1];
    paryd3=paryd3[npar+1:1];
  end;
  * FIM DA ESTIMACAO NAO LINEAR */

  if td1<2 | (td1=2 & n=&k1) then do;
    if td1=0 & td2=2 then out=ld2*sigma;
    if td1=1 & td2=2 then do;
      out=out|| (ld2*sigma);
      n12=n12+1;
    end;
    run montas(n, pm, &k1, &k2, z, s, id1, id2);
    run bic(n, s, pm, paryd3, parzd3, nod, out);
  end;
  run previsao(h, m, y, n, &k1, &k2, td1, td2, paryd3, parzd3, out, prevtd);
  pre=prevtd( |+, | );
  if nod=2 then do;
    nd12=nd12+1;
    run medvar(obs2, 1, 1, nd12, pre, meddt, vardt, &l);
  end;
  if id1=1 & id2=0 then do;
    a1=nd1-nd12;
  end;

```

```

        run medvar(obs2,2,1,a1,pre,meddt,vardt,&l);
    end;
    if id1=0 & id2=1 then do;
        a2=nd2-nd12;
        run medvar(obs2,1,2,a2,pre,meddt,vardt,&l);
    end;
end;
else do;
    pre=prevtd(|+,|);
    cont=&l-n_outd;
    run medvar(obs2,2,2,cont,pre,meddt,vardt,&l);
end;

*----- SERIE AGREGADA -----*;
*----- SEM TESTE DE DETECCAO DE OUTLIER -----*;
run montas(na,pma,ka,kal,za,s,0,0);          * MONTA A MATRIZ PARA SWEEP;
run bic(na,s,pma,parya2,parza2,0,out);      * CALCULA PA PELO BIC;
run previsao(h,1,ya,na,ka,kal,ta,0,parya2,parza2,out,preva2); * CALCULA AS PREVISOES;

*----- COM TESTE DE DETECCAO DE OUTLIER -----*;
* ta=0 (sem outlier) ta=1 (IO) ta=2 (AO);
run sigma(na,ya,parya2,ka,kal,aest,sigma,R);
run teste(ca,na,ya,parya2,ka,aest,sigma,la,ta);
if ta=2 then ia=1;
sa=sigma;

prevta=preva2;
parya3=parya2;
parza3=parza2;

* REESTIMANDO OS PARAMETROS;
if ia=1 then do;
    n_outa=n_outa+1;
    run montas(na,pma,ka,kal,za,s,ia,0);
    run bic(na,s,pma,parya3,parza3,ia,out);
    run previsao(h,1,ya,na,ka,kal,ta,0,parya3,parza3,out,prevta);
    run medvar(obs2,1,1,n_outa,prevta,medat,varat,&l);
end;
else do;
    cont=&l-n_outa;
    run medvar(obs2,1,2,cont,prevta,medat,varat,&l);
end;

l=&l;
run erro;

%end;

nr=&nr;
run var;
run eqm_t(n,theta,phi,h,m,k1,k2,o1,o2,t1,t2,eqmt,eft);
run print;

efds=ef[1];
efas=ef[2];
eftds=efts[1];
eftas=efts[2];
eq21s=eqm[2,1];
eq22s=eqm[2,2];
eqt21s=eqmts[2,1];
eqt22s=eqmts[2,2];

```

```

file dados;
put theta ' ' phi ' ' efds ' ' efas ' ' eftds ' ' eftas ' '
    eq21s ' ' eq22s ' ' eqt21s ' ' eqt22s ;

file nout;
put theta phi n_outd n_outa nd1 nd12 naodi nl2;

%mend simula;

proc iml symsize=600;
file dados;

*--- SIMULA UM PROCESSO ARIMA(1,1,1) ---*;
start sim(y,z,n0,n,phi,theta,hd,seed);
  ny=n+n0;
  y=j(ny+hd,1,0);
  z=j(n+hd,1,0);
  a_ant=0;
  do i=3 to ny+hd;
    a=rannor(seed);
    y[i]=(1+phi)*y[i-1]-phi*y[i-2]+a-theta*a_ant;
    a_ant=a; * ARIMA(1,1,1);
    if i>n0 then z[i-n0]=y[i]-y[i-1];
  end;
  y=y[n0+1:ny+hd];
  z=z[1:n];
finish;
*--- Fim sim ---*;

*--- MONTA A MATRIZ PARA FAZER SWEEP ---*;
start montas(n,pm,k1,k2,z,s,ind1,ind2);
  x=j(n,1,0);
  if ind1=1 then do;
    x[k1]=1;
    if k1<n then x[k1+1]=-1;
  end;
  if ind1=0 & ind2=1 then x[k2]=1;
  if ind1*ind2=1 then do;
    dummy2=j(n,1,0);
    dummy2[k2]=1;
    x=x||dummy2;
  end;
  x=x[pm+1:n,];

  do k=0 to pm-1;
    zaux=z[pm-k:n-1-k];
    x=x||zaux;
  end;
  if ind1+ind2=0 then x=x[,2:pm+1];

  z1=z[pm+1:n];
  xx=x`*x;
  xz=x`*z1;
  zz=z1`*z1;

  *** monta a matriz s;
  s= (xx||xz)/(xz`||zz);
* print s;
finish;

*--- TRANSFORMA O VETOR DE PARAMETROS DE Z EM VETOR DE PAR. DE Y ---*;

```

```

start parzy(parz,pary);
  par=parz;
  p=nrow(parz);
  a2=(-1);
  par=a2//par;
  a=(0);
  pary=par[2:p+1,]//a;
  pary=pary-par;
finish;

*--- BIC ---*;
start bic(n,s,pm,pary,parz,d,out);
  bicm=30000;
  p=-1;
  sa=s;

  do i=1 to pm+d;
    sa=sweep(sa,i);
    if i>d then do;
      aux=sa[pm+1+d,pm+1+d];
      xbic=log(aux/(n-i+1+d))+i*log(n)/n;
      if xbic < bicm then do;
        bicm=xbic;
        p=i-d;
        out=0;
        if d=1 then out=sa[1:d,pm+1+d];
        par=sa[1+d:p+d,pm+1+d];
        *print sa par p d out;
      end;
    end;
  end;

parz=par;
run parzy(parz,pary);
pary=pary[nrow(pary):1];
parz=parz[nrow(parz):1];

* OBS: Os parametros estao em ordem inversa !!!*;
finish;

*--- AGREGA A SERIE POR PERIODO DE M ---*;
start agrega(h,na,y,ya,za,m,k1,ka);
  ka=int((k1+m-1)/m);
  ya=j(na+h,1,0);
  za=j(na+h,1,0);
  do i=2 to na+h;
    a=(i-1)*m+1;
    b=i*m;
    yaux=y[a:b];
    ya[i]=yaux(1+,1);
    za[i]=ya[i]-ya[i-1];
  end;
  za=za[1:na];
finish;

*--- CALCULO DE PREVISOES H PASSOS A FRENTE ---*;
start previsao(h,m,y,n,k1,k2,t1,t2,pary,parz,out,prev);
  hd=h*m;
  p=nrow(pary);
  prev=j(p+hd,1,0);

```

```

* retirando os efeitos de outliers;
if t1>0 then do;
  y[k1]=y[k1]-out[1];
  if t2>0 then y[k2]=y[k2]-out[nrow(out)];
end;
if (t1=1) & (k1<n) then y[k1+1]=y[k1+1]-pary[p]*out[1];
if (t1=0) & (t2=2) then y[k2]=y[k2]-out;

* fazendo previsao;
do i=1 to p;
  prev[i]=y[n+i-p];
end;
do i=p+1 to p+hd;
  prev1=prev[i-p:i-1,];
  prev[i]=pary'*prev1;
end;
prev=prev[p+1:p+hd,];

* inserindo efeito de outlier;
ma={1};
ar=1;
ar=ar||(-parz[nrow(parz):1]`);
ma=ratio(ar,ma,20);
ar={1 -1};
psi=-ratio(ar,ma,2+hd);

if t1*k1=n-1 then do;
  do i=1 to hd;
    prev[i]=prev[i]-psi[i+2]*out[1];
  end;
end;

prev=prev[hd-m+1:hd];
finish;

*--- INSERE OUTLIER AO ---*;
start ao(n,y,z,k1,k2,o1,o2);
Y[k1]=Y[k1]+o1;
Z[k1]=Z[k1]+o1;
if k1<n then do;
  Z[k1+1]=Z[k1+1]-o1;
end;
if o2^=0 then do;
  Y[k2]=Y[k2]+o2;
  Z[k2]=Z[k2]+o2;
end;
finish;

```

```

*--- INSERE OUTLIER IO ---*;
START IO(n, y, z, k1, k2, O1, O2, phi, theta, h);
  m=4;
  a=h*m+n-k1+1;
  z[k1]=z[k1]+o1;

  if k1<n then z[k1+1]=z[k1+1]+phi*o1-theta*o1;
  ma=1;
  ar=1;
  if theta^=0 then ma=ma||(-theta);
  if phi^=0 then ar=ar||(-phi);
  ma=ratio(ar, ma, 20);
  ar=(1 -1);
  psi=ratio(ar, ma, a);
  if o2^=0 then do;
    z[k2]=z[k2]+o2;
    psil=psi*o2;
    psil=psil[1:a-1];
    j=j(k2-1, 1, 0);
    psil=j//psil;
    y=y+psil;
  end;

  psi=psi*o1;
  j=j(k1-1, 1, 0);
  psi=j//psi`;
  y=y+psi;
finish;

*--- CALCULO DE INOVACAO E SUA VARIANCIA ---*;
start sigma(n, y, pary, k1, k2, aest, sigma, R);
  p=nrow(pary)-1;
  ma=0;
  sigma=0;
  aest=j(n, 1, 0);

  *----- Estimacao de variancia pelo metodo robusto -----*;
  if r=1 then do;
    do i=p+2 to n;
      aux=y[i-p-1:i-1];
      aest[i]=y[i]-pary`*aux;
      auxa=aest[i]-ma;
    end;
    a=n-k1+1;
    med=sum(aest[p+2:n-a])/(n-p-1-a);
    sigma=sqrt((ssq(aest[p+2:n-a])-(n-p-1-a)*(med**2))/(n-2*(p+1+a)));
  end;
  *----- Fim da estimacao robusta -----*;

  *----- Estimacao da variancia pelo metodo usual -----*;
  if r=0 then do;
    do i=p+2 to n;
      aux=y[i-p-1:i-1];
      aest[i]=y[i]-pary`*aux;
    end;
    med=sum(aest)/(n-p-1);
    sigma=sqrt((ssq(aest)-(n-p-1)*(med**2))/(n-2*(p+1)));
  END;
  *----- Fim da estimacao usual -----*;
finish;

*--- TESTE DE DETECAO DE OUTLIER ---*;

```

```

start teste(c,n,y,pary,k,aest,sigma,lao,tipo);
  tipo=0;
  p=nrow(pary);
  if n=k+1 then do;
    lio=aest[k]/sigma;
    lao=aest[k];
    var=1;
    min=n-k;
    if p<n-k then min=p;
    do i=1 to min;
      lao=lao-pary[p-i+1]*aest[k+i];
      var=var+pary[p-i+1]**2;
    end;
    lao=lao/(sigma*sqrt(var));
  end;
  else do;
    lao=aest[n]/sigma;
    if abs(lao)>c then tipo=2;
  end;

  if n-k>0 then do;
    if max(abs(lio),abs(lao))>c then do;
      if abs(lio)>abs(lao) then do;
        tipo=1;
        lao=lio;
      end;
      else tipo=2;
    end;
  end;
finish;

/* AS SEGUINTE TRÊS SUB-ROTINAS PARA ESTIMACAO NAO LINEAR FORAM
RETIRADAS DO LIVRO BATES E WATTS (1988 PÁGINAS) */

*--- CRIA MATRIZ X PARA ESTIMACAO NAO LINEAR ---*;
start matx(k1,z,par,x,out2);
  n=nrow(z);
  p=nrow(par);
  x=j(n,p+2,0);
  x[k1,1]=1;
  x[n,1]=-1;
  x[n,2]=1;
  do j=1 to p;
    do i=1 to n-j;
      x[i+j,3+p-j]=z[i];
    end;
  end;
  if out2 then do;
    aux=j(n,1,0);
    aux[n]=1;
    x=x[,1:2]||aux||x[,3:p+2];
  end;
  X=X[1+p:n,];
finish;

```

```

*--- INICIALIZACAO PARA ESTIMACAO NAO LINEAR -----*;
start resgrad(z, x, nparz, par, res, grad, derivs);
  n=nrow(z);
  p=nrow(par);
  nout=p-nparz;
  px=ncol(x);
  parz=par[nout+1:p];
  if nout=1 then do;
    res=z[1+nparz:n]-x[,1]*par[1]+x[,2]*par[1]*par[p]-
      x[,2+nout:px]*parz;
    if derivs=1 then do;
      grad=(x[,1]-x[,2]*par[p])|| (x[,px]-x[,2]*par[1]);
      if nparz>1 then do;
        do i=1 to nparz-1;
          grad=grad||x[,px-i];
        end;
      end;
    end;
  end;
  if nout=2 then do;
    res=z[1+nparz:n]-x[,1]*par[1]+x[,2]*par[1]*par[p]+
      x[,3]*par[2]-x[,2+nout:px]*parz;
    if derivs=1 then do;
      grad=(x[,1]-x[,2]*par[p])|| (x[,3])||
        (x[,px]-x[,2]*par[1]);
      if nparz>1 then do;
        do i=1 to nparz-1;
          grad=grad||x[,px-i];
        end;
      end;
    end;
  end;
end;
finish;

*--- ESTIMACAO NAO LINEAR -----*;
start nlsfit(z, x, out, parzd2, criterio, maxiter, minstep, tol, verbose);
  out=out`;
  par=out//parzd2;
  nparz=nrow(parzd2);
  run resgrad(z, x, nparz, par, resid, grad, 1);
  nout=nrow(out);
  p=nrow(par);
  n=nrow(resid);
  ndof=n-p;
  mult=sqrt(p/ndof);
  stepsize=1;
  out0=out;
  phi0=parzd2;
  do iter=1 to maxiter;
    out1=par[1:nout];
    phi=par[nout+1:p];
    oldssq=ssq(resid);
    call gsorth(qhat, rhat, rank, grad);
    if rank=1 then do;
      print 'matriz de derivadas singular';
      stop;
    end;
    tan=qhat`*resid;
    ss_tan=ssq(tan);
    incr=solve(rhat, tan);
    criterio=mult*sqrt(ss_tan/(oldssq-ss_tan));
    *if verbose then
      print iter(1format=2.0) criterio incr;
  end;
end;

```

```

if criterio<tol then do;
  out=par[1:nout,]`;
  parzd2=phi;
  * print iter out0 out1 phi0 phi;
  * put iter ' ' out0 ' ' out1 ' ' phi0 ' ' phi;
  link returns;
end;
do until (newssq<oldssq);
  if stepsize<minstep then do;
    out=par[1:nout,]`;
    parzd2=par[nout+1:p,];
    print 'atingiu o numero minimo de passos';
    * print iter out0 out1 phi0 phi;
    * put iter ' ' out0 ' ' out1 ' ' phi0 ' ' phi;
    link returns;
  end;
  trial=par+stepsize*incr;
  run resgrad(z,x,nparz,trial,resid,grad,0);
  newssq=ssq(resid);
  *if verbose then
    print stepsize newssq oldssq trial;
  stepsize=stepsize/2;
end;
par=trial;
stepsize=min(1||4*stepsize);
run resgrad(z,x,nparz,par,resid,grad,1);
end;
out=par[1:nout,]`;
parzd2=par[nout+1:p,];
* print iter out0 out1 phi0 phi;
* put iter ' ' out0 ' ' out1 ' ' phi0 ' ' phi;
print 'atingiu o numero maximo de iteracoes: fim do programa';
link returns;
returns:finish;

*--- CALCULO DE MEDIA E VARIANCIA DOS ERROS DE PREVISAO ---*;
start medvar(obs,i,j,cont,pre,medt,vart,1);
  e=obs-pre;
  aux=e-medt[i,j];
  vart[i,j]=vart[i,j]+aux**2-(aux**2)/cont;
  medt[i,j]=medt[i,j]*(cont-1)/cont+e/cont;
finish;

*--- CALCULO DE EQM TEORICO ---*;
START EQM_T(n,THETA,phi,h,m,k1,k2,o1,o2,t1,t2,EQMT,eft);
  eqmt=j(2,2,0);
  eft=j(2,1,0);
  ar=j(1,2,1);
  ma=j(1,3,1);
  p=j(1,5,0);
  hd=h*m;

*---- calculo dos parametros da serie agregada ----*;
if phi^=0 then do;
  p2=phi**2;
  p3=phi**3;
  p4=phi**4;
  p5=phi**5;
  p6=phi**6;
  c1=44+80*phi+106*p2+120*p3+106*p4+80*p5+44*p6;
  c2=10+24*phi+ 42*p2+ 64*p3+ 42*p4+24*p5+10*p6;
  c3=
      p2+ 4*p3+      p4;

```

```

p[1]=c3**3;
p[2]=(c1-2*c3)*(c3**2);
p[3]=(2*c3-2*c1)*(c3**2)+(c2**2)*c3;
p[4]=p[2];
p[5]=p[1];

r=polyroot(p);
ra=abs(r);
max=ra[<>];

if abs(max)>00001 then do;
  theta2=100;
  do i=1 to nrow(r);
    if r[i,2]=0 then do;
      if abs(r[i,1])<theta2 then
        theta2=r[i,1];
    end;
  end;

  theta1=theta2*c2/((1-theta2)*c3);
  sigma=-c3/theta2;

  ar[2]=-phi**4;
  ma[2]=-theta1;
  ma[3]=-theta2;

  ma=ratio(ar,ma,8);
  ar={1 -1};
  ef2=ratio(ma,ar,h+1);

  ** calculo dos efeitos de AOs **;
  if t1=2 | t2=2 then do;
    if n-k1=1 & o2=0 then do;
      eft1=j(hd,1,-phi);
      eft1[2]=- (1+phi)*phi;
      do i=3 to hd;
        eft1[i]=(1+phi)*eft1[i-1]-phi*eft1[i-2];
      end;
      eft1=eft1[hd-m+1:hd];
      eft[1]=eft1(+,|);
    end;

    if n-k1=0 & o2=0 then do;
      eft1=j(hd,1,1+phi);
      eft1[2]=eft1[1]**2-phi;
      do i=3 to hd;
        eft1[i]=(1+phi)*eft1[i-1]-phi*eft1[i-2];
      end;
      eft1=eft1[hd-m+1:hd];
      eft[1]=eft1(+,|);
    end;
  end;

```

```

if o2^=0 then do;
  eft1=j(hd,1,-phi);
  eft1[2]=-(1+phi)*phi;
  do i=3 to hd;
    eft1[i]=(1+phi)*eft1[i-1]-phi*eft1[i-2];
  end;
  eft2=eft1[hd-m+1:hd];
  eft[1]=eft2(|+,|)*o1;

  eft1[1]=1+phi;
  eft1[2]=eft1[1]**2-phi;
  do i=3 to hd;
    eft1[i]=(1+phi)*eft1[i-1]-phi*eft1[i-2];
  end;
  eft1=eft1[hd-m+1:hd];
  eft[1]=eft[1]+eft1(|+,|)*o2;
end;

eft[2]=-ef2[h+1]*(o1+o2);
end;

** calculo do efeito com IOs **;
if t1=1 | t2=1 then do;
  alfa=j(1,10,1);
  thet=j(1,9,0);

  alfa[2]=2+ phi;
  alfa[3]=3+2*phi+p2;
  alfa[4]=4+3*phi+2*p2+p3;
  alfa[5]=3+4*phi+3*p2+2*p3;
  alfa[6]=2+3*phi+4*p2+3*p3;
  alfa[7]=1+2*phi+3*p2+4*p3;
  alfa[8]= phi+2*p2+3*p3;
  alfa[9]= p2+2*p3;
  alfa[10]= p3;

  thet[1]=1;
  thet[5]=-theta1;
  thet[9]=-theta2;

  phi1=1+phi**4;
  phi2=-phi**4;

  gama=-ratio(thet,alfa,8);

  if n-k1<2 then do;
    eft[2]=thet1*gama[n-k1+1]-alfa[n-k1+m+1];
    if h>1 then do;
      e1=eft[2];
      eft[2]=phi1*eft[2]+theta2*gama[n-k1+1]-alfa[n-k1+m+1];
      if h=3 then eft[2]=phi1*eft[2]-phi2*e1;
    end;
  end;
  else print 'Os efeitos teoricos soh foram calculados
para 0<n-k1<2';
end;

end;
end;

*--- fim do calculo dos parametros da serie agregada ---*;

*print theta1 theta2 sigma;

```

```

** calculo de alfa **;
if theta^=0 then do;
a=10-24*theta+10*(theta**2);
b=(44-80*theta+44*(theta**2))/a;
d=sqrt(b**2-4);
alfa=(-b-d)/2;
alfa2=(-b+d)/2;
if abs(alfa)>abs(alfa2) then alfa=alfa2;
alfa=alfa+phi**4;
sigma=- (10-24*theta+10*(theta**2))/alfa;

** calculo dos efeitos de AOs **;
if t1=2 | t2=2 then do;
if n-k1=1 & o2=0 then eft[1]=m*theta*(1-theta)*o1;
if n-k1=0 & o2=0 then eft[1]=m*(1-theta)*o1;
if n-k1=1 & o2^=0 then eft[1]=m*(1-theta)*(o1+theta*o2);
eft[2]=(1-alfa)*(o1+o2);
end;
if t1=1 | t2=1 then do;
if n-k1=0 & o2=0 then eft[2]=-alfa+(4*theta-3);
if n-k1=1 & o2=0 then eft[2]=alfa*(theta-2)+(3*theta-2);
end;
end;

** calculo de eqm **;
eqmt[1,1]=30-40*theta+14*(theta**2);
eqmt[2,1]=sigma;
if phi^=0 then do;
eqmt[1,1]=0;
do i=(h-1)*m+1 to h*m;
a=0;
do j=1 to i;
a=a+(i-j+1)*(phi**(j-1));
end;
eqmt[1,1]=eqmt[1,1]+a**2;
end;
if h=2 then eqmt[2,1]=eqmt[2,1]*((1+p4-theta1)**2);
if h=3 then eqmt[2,1]=eqmt[2,1]*((1+p4+p4**2-
theta1*(1+p4)-theta2)**2);
if h=4 then eqmt[2,1]=eqmt[2,1]*((1+p4+p4**2+p4**3-
theta1*(1+p4+p4**2)-theta2*(1+p4))**2);
end;
eqmt[,2]=eqmt[,1]+eft##2;
finish;

*--- CALCULA ERRO E VAR DE ERRO DE PREVISAO ---*;
start erro;
prev[1,1]=prevd1(|+,|);
prev[2,1]=preval;
prev[1,2]=prevd2(|+,|);
prev[2,2]=preva2;

erro[,1]=obs1-prev[,1];
erro[,2]=obs2-prev[,2];

```

```

do i2=1 to 2;
  do i3=1 to 2;
    aux=erro[i2,i3]-med[i2,i3];
    var[i2,i3]=var[i2,i3]+aux**2-(aux**2)/1;
    med[i2,i3]=med[i2,i3]*(1-1)/1+erro[i2,i3]/1;
  end;
end;
* print obs1 obs2 prev erro med var;
finish;

*--- DIVIDE OS SSQ PELOS SEUS GRAUS DE LIBERDADE ---*;
start var;
  ef=med[,1]-med[,2];
  var=var/nr;
  eqm=var;
  eqm[,2]=var[,1]+ef##2;

  m2=meddt##2;
  m3=medat##2;
  if nd12>0 then do;
    medts[1]=meddt[1,1]*nd12;
    if nd12=nd2 then medts[1]=medts[1]+meddt[1,2]*(nd2-nd12);
  end;
  medts[1]=(medts[1]+meddt[2,1]*(nd1-nd12)+meddt[2,2]*(nr-n_outd))/nr;
  if n_outa>0 then
  medts[2]=medat[1]*n_outa/nr;
  medts[2]=medts[2]+(medat[2]*(nr-n_outa))/nr;
  efts=med[,1]-medts;

  varts[1]=sum(vardt)+m2[1,1]*nd12+m2[1,2]*(nd2-nd12)+
    m2[2,1]*(nd1-nd12)+m2[2,2]*(nr-n_outd)-nr*(medts[1]**2);
  varts[2]=sum(varat)+m3[1]*n_outa+m3[2]*(nr-n_outa)-nr*(medts[2]**2);

  varts=varts/nr;
  eqmts[,1]=varts;
  eqmts[,2]=varts+efts##2;

finish;

*--- IMPRIME OS RESULTADOS ---*;
start print;
  print nr n o1 o2 k1 k2 h;
  print t1 t2 phi theta cd ca r;
  print 'Valores teoricos:';
  print eft eqmt;
  print 'Valores observados:';
  print ef eqm med;
  print '1a lin=serie desagregada 2a lin=serie agregada (m=4)';
  print '1a col=serie sem outlier 2a col=serie com outlier';
  print 'Resultados com teste de deteccao de outlier:';
  print efts eqmts medts;
  print 'Numero de outliers detectados: (respect desagr. e agreg.)';
  print n_outd n_outa;
  print naod1 nd1 nd2 nd12;
finish;

```

```
** colocar:
    phi,theta, n,   nr,  o1, o2, k1, k2, t1, t2, h, r, cd, ca ;
    /* r=1 (robusto), r=0 (usual) ,
       t=0 (sem outlier)   t=1 (IO)   t=2 (AO) */
%simula(-0.8,  0 , 100,1000 , 1,-1, 99, 100 ,1 , 1, 1, 0, 2, 2);
%simula(-0.5,  0 , 100,1000 , 1,-1, 99, 100 ,1 , 1, 1, 0, 2, 2);
quit;

data teste;
infile dados;
input theta phi efds efas eftds eftas
       eq21s eq22s eqt21s eqt22s;

data nout;
infile nout;
input theta phi n_outd n_outa nd1 nd12 naod1 n12;

proc print data=teste;
proc print data=nout;
run;
□
```

## **Apêndice 5**

### **Tabelas dos Resultados de Simulação**

**Tabela A5.1** - Porcentagem de outliers detectados para diferentes modelos (ARIMA(1,1,0) e ARIMA(0,1,1)) e tamanhos de outliers ( $\omega_A = 1$  e 2) do tipo AO, presente nas posições  $r = 1$  (penúltima) e  $r = 0$  (última, em negrito). Simulações com 1000 replicações de séries de tamanho 100.

ARIMA(1,1,0)								
$\phi$	$\omega_A = 1$				$\omega_A = 2$			
	D	AO	IO	Agreg	D	AO	IO	Agreg
-0,8	15	52	48	3,9	53,3	56,85	43,15	5,3
	<b>16,8</b>	<b>100</b>	<b>0</b>	<b>4,2</b>	<b>47,4</b>	<b>100</b>	<b>0</b>	<b>6,4</b>
-0,5	22,8	59,2	40,79	4,4	59,6	71,64	28,36	4,9
	<b>16,5</b>	<b>100</b>	<b>0</b>	<b>2,8</b>	<b>45,6</b>	<b>100</b>	<b>0</b>	<b>5,5</b>
-0,2	25,3	67,59	32,41	3,8	72	80,06	19,94	4,9
	<b>17,3</b>	<b>100</b>	<b>0</b>	<b>3,6</b>	<b>48,8</b>	<b>100</b>	<b>0</b>	<b>5,3</b>
0,2	35,4	77,68	22,32	3,1	86,9	89,18	10,82	4,3
	<b>15,3</b>	<b>100</b>	<b>0</b>	<b>3,4</b>	<b>47</b>	<b>100</b>	<b>0</b>	<b>4,4</b>
0,5	42,5	81,18	18,82	2,5	94,8	96,73	3,27	2,8
	<b>16,9</b>	<b>100</b>	<b>0</b>	<b>3</b>	<b>50,4</b>	<b>100</b>	<b>0</b>	<b>3,5</b>
0,8	53,6	88,43	11,57	1,3	98,1	97,35	2,65	1,2
	<b>14,6</b>	<b>100</b>	<b>0</b>	<b>1,7</b>	<b>47,2</b>	<b>100</b>	<b>0</b>	<b>1,3</b>

ARIMA(0,1,1)								
$\theta$	D	AO	IO	Agreg	D	AO	IO	Agreg
-0,8	47,5	82,32	17,68	3,5	93,2	95,92	4,08	3,7
	<b>15,3</b>	<b>100</b>	<b>0</b>	<b>3,3</b>	<b>45,1</b>	<b>100</b>	<b>0</b>	<b>4,2</b>
-0,5	41,3	80,04	19,96	3	92,8	94,07	5,93	4,3
	<b>14</b>	<b>100</b>	<b>0</b>	<b>3,1</b>	<b>46,2</b>	<b>100</b>	<b>0</b>	<b>3,8</b>
-0,2	32,7	78,59	21,41	5,2	88,1	90,35	9,65	3,4
	<b>14,7</b>	<b>100</b>	<b>0</b>	<b>3,7</b>	<b>48,3</b>	<b>100</b>	<b>0</b>	<b>3,3</b>
0,2	27	68,15	31,85	3,8	73,3	77,49	22,51	5,7
	<b>15,5</b>	<b>100</b>	<b>0</b>	<b>3,5</b>	<b>49,9</b>	<b>100</b>	<b>0</b>	<b>4,4</b>
0,5	23,3	66,52	33,48	4,3	60,5	71,24	28,76	6,7
	<b>13,7</b>	<b>100</b>	<b>0</b>	<b>4,7</b>	<b>43</b>	<b>100</b>	<b>0</b>	<b>7,3</b>
0,8	15,1	60,93	39,07	5,7	51	62,16	37,84	10,3
	<b>12,2</b>	<b>100</b>	<b>0</b>	<b>4,2</b>	<b>43,1</b>	<b>100</b>	<b>0</b>	<b>10,7</b>

**Legenda:**

- D      Porcentagem de outliers detectados no modelo desagregado,  
 AO     Porcentagem de AO detectados no modelo desagregado, dentre os D,  
 IO     Porcentagem de IO detectados no modelo desagregado, dentre os D, e  
 Agreg   Porcentagem de outliers detectados no modelo agregado.

**Tabela A5.2 - Porcentagem de outliers detectados para diferentes modelos e outliers do tipo AO nas duas últimas observações de tamanhos  $\omega_{A1} = 1$  e  $\omega_{A2} = 1$ . Simulações com 1000 replicações de séries de tamanho 100.**

<b>Modelo ARIMA(1,1,0)</b>							
$\phi$	(AO,AO)	(IO,AO)	(AO,SO)	(IO,SO)	(SO,AO)	out_d	out_a
-0,8	0	1	2,6	12,2	9	24,8	5,9
-0,5	0,3	1,8	2,4	12,5	5	22	6,1
-0,2	1,3	0,7	3,3	10,6	3,7	19,6	5,5
0,2	3,3	0,4	5,5	12,9	1,4	23,5	3,4
0,5	5,9	0,1	6,2	10,4	1,2	23,8	3,7
0,8	8,3	0,1	7	14,5	2	31,9	0

<b>Modelo ARIMA(0,1,1)</b>							
$\theta$	(AO,AO)	(IO,AO)	(AO,SO)	(IO,SO)	(SO,AO)	out_d	out_a
-0,80	5,50	0,1	6,8	12,6	2,3	27,3	4,3
-0,50	4,2	0	5,4	13,2	2,9	25,7	3,1
-0,20	2,2	0,1	4,3	13,2	2,6	22,4	3,7
0,20	1,6	0,4	4,2	14,9	3,6	24,7	3,1
0,5	0,8	0,3	3,7	11,2	3,7	19,7	4,8
0,80	0,4	0,9	2,7	8,5	6,5	19	8,1

**Legenda:**

(TP1,TP2)

TP1 - Tipo do outlier detectado na penúltima observação

TP2 - Tipo do outlier detectado na última observação

onde, TP1 e TP2 podem ser AO, IO ou SO (sem outlier)

A coluna de (TP1, TP2) representa a porcentagem de outlier do tipo TP1 detectado na penúltima observação e outlier do tipo TP2 detectado na última observação da série.

out\_d - Porcentagem de séries com pelo menos um outlier detectado na série desagregada

out\_a - Porcentagem de séries com um outlier detectado na série agregada

**Tabela A5.3 -** Porcentagem de outliers detectados para diferentes modelos e outliers do tipo AO nas duas últimas observações de tamanhos  $\omega_{A1} = 1$  e  $\omega_{A2} = -1$ . Simulações com 1000 replicações de séries de tamanho 100.

<b>Modelo ARIMA(1,1,0)</b>							
$\phi$	(AO,AO)	(IO,AO)	(AO,SO)	(IO,SO)	(SO,AO)	out_d	out_a
-0,80	5,4	0,2	10,6	2,4	15,2	33,8	4,3
-0,50	14,5	0	14,5	2,9	16	47,9	3,3
-0,20	26,6	0,2	15,5	3	12,1	57,4	2,4
0,20	48,2	0	11,8	3	8,4	71,4	3,6
0,50	61	0	9,8	1,8	5,7	78,3	2,9
0,80	72,1	0,1	6,2	1,8	4,2	84,4	1,2

<b>Modelo ARIMA(0,1,1)</b>							
$\theta$	(AO,AO)	(IO,AO)	(AO,SO)	(IO,SO)	(SO,AO)	out_d	out_a
-0,80	60,7	0,1	9,8	2,8	4	77,4	2,1
-0,50	55,4	0	11,3	2,6	7,7	77	2,6
-0,20	48,4	0,2	11,9	2,2	8,3	71	2,7
0,20	24,8	0,1	14,4	2,2	12,5	54	3,6
0,50	16,4	0,1	13,5	2,4	13,3	45,7	4
0,80	7,9	0,1	11,7	2,8	14,9	37,4	3,6

**Legenda:**

(TP1,TP2)

TP1 - Tipo do outlier detectado na penúltima observação

TP2 - Tipo do outlier detectado na última observação

onde, TP1 e TP2 podem ser AO, IO ou SO (sem outlier)

A coluna de (TP1, TP2) representa a porcentagem de outlier do tipo TP1 detectado na penúltima observação e outlier do tipo TP2 detectado na última observação da série.

out\_d - Porcentagem de séries com pelo menos um outlier detectado na série desagregada

out\_a - Porcentagem de séries com um outlier detectado na série agregada

**Tabela A5.4** - Porcentagem de outliers detectados para diferentes modelos (ARIMA(1,1,0) e ARIMA(0,1,1) e tamanhos de outliers ( $\omega_1 = 1$  e 2) do tipo IO presente nas posições  $r = 1$  (penúltima) e  $r = 0$  (última, em negrito). Simulações com 1000 replicações de séries de tamanho 100.

ARIMA(1,1,0)								
$\phi$	$\omega_1 = 1$				$\omega_1 = 2$			
	D	AO	IO	Agreg	D	AO	IO	Agreg
-0,8	19	46,32	53,68	4,6	49,1	41,34	58,66	7,1
	<b>16,7</b>	<b>100</b>	<b>0</b>	<b>4</b>	<b>44</b>	<b>100</b>	<b>0</b>	<b>7,2</b>
-0,5	17,2	41,28	58,72	4,9026	50,8	30,91	69,09	7,1
	<b>13,5</b>	<b>100</b>	<b>0</b>	<b>3,9</b>	<b>49,3</b>	<b>100</b>	<b>0</b>	<b>5,3</b>
-0,2	20,4	37,75	62,25	4,2	50,9	23,58	76,42	5,8
	<b>16,9</b>	<b>100</b>	<b>0</b>	<b>3,7</b>	<b>50,8</b>	<b>100</b>	<b>0</b>	<b>4,3</b>
0,2	19,7	26,4	73,6	4,3	53,7	15,46	84,54	7,5
	<b>15,2</b>	<b>100</b>	<b>0</b>	<b>3,2</b>	<b>49,1</b>	<b>100</b>	<b>0</b>	<b>3,8</b>
0,5	19,2	28,13	71,87	2,8	52,4	14,12	85,88	4
	<b>13,8</b>	<b>100</b>	<b>0</b>	<b>2,6</b>	<b>49</b>	<b>100</b>	<b>0</b>	<b>2,6</b>
0,8	18,2	27,47	72,53	1,2	51,8	10,81	89,19	1,2
	<b>15,7</b>	<b>100</b>	<b>0</b>	<b>2</b>	<b>49</b>	<b>100</b>	<b>0</b>	<b>1,4</b>

$\theta$ ARIMA(0,1,1)								
-0,8	16,5	20	80	4,5	42,9	8,16	91,84	3,6
	<b>13,5</b>	<b>100</b>	<b>0</b>	<b>4,5</b>	<b>42,5</b>	<b>100</b>	<b>0</b>	<b>5</b>
-0,5	18	23,33	76,67	4,3	49,7	11,07	88,93	5,7
	<b>15,3</b>	<b>100</b>	<b>0</b>	<b>3,2</b>	<b>47,5</b>	<b>100</b>	<b>0</b>	<b>3,8</b>
-0,2	20,9	30,62	69,38	3,9	52,2	13,03	86,97	6,3
	<b>15,4</b>	<b>100</b>	<b>0</b>	<b>4</b>	<b>46,6</b>	<b>100</b>	<b>0</b>	<b>4</b>
0,2	20,8	26,44	73,56	4,7	52,4	25	75	7,3
	<b>18,3</b>	<b>100</b>	<b>0</b>	<b>4,3</b>	<b>49,6</b>	<b>100</b>	<b>0</b>	<b>4,6</b>
0,5	18,7	39,57	60,43	5,3	50,6	35,71	64,29	7,9
	<b>16,3</b>	<b>100</b>	<b>0</b>	<b>4,8</b>	<b>47,6</b>	<b>100</b>	<b>0</b>	<b>5,1</b>
0,8	17,1	49,12	50,88	4,7	50,1	47,11	52,89	11,5
	<b>12,7</b>	<b>100</b>	<b>0</b>	<b>4,4</b>	<b>41,6</b>	<b>100</b>	<b>0</b>	<b>9,2</b>

**Legenda:**

- D        Porcentagem de outliers detectados no modelo desagregado,  
 AO       Porcentagem de AO detectados no modelo desagregado, dentre os D,  
 IO       Porcentagem de IO detectados no modelo desagregado, dentre os D, e  
 Agreg   Porcentagem de outliers detectados no modelo agregado.

**Tabela A5.5 -** Porcentagem de outliers detectados para diferentes modelos e outliers do tipo IO nas duas últimas observações de tamanhos  $\omega_{t1} = 1$  e  $\omega_{t2} = 1$ . Simulações com 1000 replicações de séries de tamanho 100.

<b>Modelo ARIMA(1,1,0)</b>							
$\phi$	(AO,AO)	(IO,AO)	(AO,SO)	(IO,SO)	(SO,AO)	out_d	out_a
-0,80	0	2,6	2,2	13,2	14,3	32,3	5,1
-0,50	0,5	2,6	2,2	11,9	12	29,2	5,5
-0,20	1,4	2,2	1	14,4	12,5	31,5	6,5
0,20	3,1	2,1	0,7	13,1	9,4	28,4	4,5
0,50	3,7	1,2	1,1	11,7	8,6	26,3	3,6
0,80	4,9	1,8	0,7	12,3	8,1	27,8	1,8

<b>Modelo ARIMA(0,1,1)</b>							
$\theta$	(AO,AO)	(IO,AO)	(AO,SO)	(IO,SO)	(SO,AO)	out_d	out_a
-0,8	4,4	2,4	1,1	11,3	7,2	26,4	3,6
-0,5	3,3	2,2	0,6	12,4	9,6	28,1	3,2
-0,2	2,1	2,1	1,3	12,7	11	29,2	5,2
0,2	1,3	2	1,7	13,2	11,3	29,5	7,7
0,5	0,7	1,6	1,3	13,3	11,4	28,3	7,3
0,8	0,1	1,3	2,7	10,6	10,8	25,5	9,3

**Legenda:**

(TP1,TP2)

TP1 - Tipo do outlier detectado na penúltima observação

TP2 - Tipo do outlier detectado na última observação

onde, TP1 e TP2 podem ser AO, IO ou SO (sem outlier)

A coluna de (TP1, TP2) representa a porcentagem de outlier do tipo TP1 detectado na penúltima observação e outlier do tipo TP2 detectado na última observação da série.

out\_d - Porcentagem de séries com pelo menos um outlier detectado na série desagregada

out\_a - Porcentagem de séries com um outlier detectado na série agregada

**Tabela A5.6 -** Porcentagem de outliers detectados para diferentes modelos e outliers do tipo IO nas duas últimas observações de tamanhos  $\omega_{11} = 1$  e  $\omega_{12} = -1$ . Simulações com 1000 replicações de séries de tamanho 100.

<b>Modelo ARIMA(1,1,0)</b>							
$\phi$	(AO,AO)	(IO,AO)	(AO,SO)	(IO,SO)	(SO,AO)	out_d	out_a
-0,80	4,3	0	11,5	3,1	8,3	27,2	2,8
-0,50	9,1	0	12,1	6,1	8,4	35,7	3
-0,20	9,3	0	13,5	8,2	3,6	34,6	4,3
0,20	11,3	0	11,3	8,7	3,5	34,8	4,6
0,50	12	0,1	7,9	9	2,2	31,2	3,4
0,80	13,2	0,2	10,4	11,6	2,7	38,1	23

<b>Modelo ARIMA(0,1,1)</b>							
$\theta$	(AO,AO)	(IO,AO)	(AO,SO)	(IO,SO)	(SO,AO)	out_d	out_a
-0,80	11,6	0	7,5	9,3	2,6	31	4
-0,50	8,7	0	9,1	9,3	3,8	30,9	3
-0,20	11,1	0,1	11	9,1	3,9	35,2	4,2
0,20	10,4	0,1	14,2	7,8	4,8	37,3	4
0,50	8,2	0,1	13	6,3	7,7	35,3	3,8
0,80	6,7	0,1	11,7	3,4	9,3	31,2	3,2

**Legenda:**

(TP1,TP2)

TP1 - Tipo do outlier detectado na penúltima observação

TP2 - Tipo do outlier detectado na última observação

onde, TP1 e TP2 podem ser AO, IO ou SO (sem outlier)

A coluna de (TP1, TP2) representa a porcentagem de outlier do tipo TP1 detectado na penúltima observação e outlier do tipo TP2 detectado na última observação da série.

out\_d - Porcentagem de séries com pelo menos um outlier detectado na série desagregada

out\_a - Porcentagem de séries com um outlier detectado na série agregada

## Referências Bibliográficas

- Abraham, B. (1982). Temporal aggregation in time series. *Internat. Statist. Rev.* **50**: 285-291.
- Abraham, B. e Box, G. E. P. (1979). Bayesian analysis of some outlier problems in time series. *Biometrika*. **66**: 229-236.
- Abraham, B. e Chuang, A. (1989). Outlier detection and time series modeling. *Technometrics*. **31**: 241-248.
- Abraham, B. e Yatawara, N. (1988). A score test for detection of time series outliers. *J. Time Ser. Anal.* **9**: 109-119.
- Ahsanullah, M. E Wei, W. W. S. (1984). Effects of temporal aggregation on forecasts in an ARMA(1,1) process. *ASA Proceedings of the Business Economic Statistical Section*, 297-302.
- Akaike, H. (1974). A new look at statistical identification. *IEEE Trans. Automat. Control*. **19**: 716-723.
- Amemiya, T. e Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model. *J. Amer. Statist. Assoc.* **67**: 628-632.
- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. New York: Wiley.
- Atkinson, A. C. e Shephard, N. (1996). Deletion diagnostics for transformations of time series. *J. Forecast.* **5**: 1-17.
- Balke, N. S. (1993). Detecting level shift in time series. *J. Bus. Econom. Statist.* **11**: 81-92.
- Bhansali, R. J. (1981). The effect of not knowing the order of an autoregressive process on the mean squared error of prediction. - I *J. Am. Statist. Assoc.* **76**: 588-597.
- Bates, D. M. e Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. Ed. John Willey & Sons. Inc.

- Box, G. E. P. e Jenkins, G. M. (1970). *Times Series Analysis: Forecasting and Control*. San Francisco, Holden Day, 1ª ed.
- \_\_\_\_\_ (1976). *Times Series Analysis: Forecasting and Control*. San Francisco, Holden Day, 2ª ed.
- Box, G. E. P. e Tiao, G. C. (1965). A change in level of nonstationary time series, *Biometrika*. **52**: (1 e 2) 181-192.
- \_\_\_\_\_ (1975). Intervention analysis with application to economic and environmental problems. *J. Amer. Statist. Assoc.* **70**: 70-79.
- Bruce, A. G. e Martin R. D. (1989). Leave-k-out diagnostic for time series. *J. Roy. Statist. Soc. Ser. B.* **51**: 363-424.
- Brubacher, S. R. e Wilson, G. T. (1976). Interpolating time series with applications to the estimation of holiday effects on electricity demand. *Appl. Statist.* **25**: 107-116.
- Buse, A. (1982). The Likelihood Ratio, Wald, and Lagrange Multiplier test: an exploratory note. *Amer. Statist.* **36**: 153-157.
- Chang, I. (1982). *Outliers in Time Series*. University of Wisconsin-Madison, Dept. of Statistics.
- Chang, I. e Tiao, G. C. (1983). *Estimation of time series parameters in the presence of outliers*. University of Chicago, Statistics Research Center. (Technical Report).
- Chang, I., Tiao, G. C. e Chen, C. (1988). Estimation of times series parameters in the presence of outliers. *Technometrics*. **30**: 193-204.
- Chatfield (1988). What is the best method for forecasting. *J. Appl. Statist.*, **15**: 19-38.
- Chen, C. e Liu L.-M. (1993a). Joint estimation of model parameters and outliers effects in time series. *J. Amer. Statist. Assoc.* **88**: 284-297.
- \_\_\_\_\_ (1993b). Forecasting time series with outliers. *J. Forecast.* **12**: 13-35.
- Chen, C., Liu L.-M., e Hudak, G. B. (1992). *Modeling and Forecasting Time Series in the Presence of Outliers and Missing Data*. DeKalb, IL: Scientific Computing Associates. (Working Paper Series).
- Chen, C. e Tiao, G. C. (1986). *Some diagnostic statistics in time series*. University of Chicago, Statistical Research Center. (Technical Report).
- \_\_\_\_\_ (1990). Random level shift times series models, ARIMA approximation, and level shift detection *J. Bus. Econom. Statist.* **8**: 170-186.

- Cook, R.P. (1986). Assessment of local influence. *J. Roy. Statist. Soc. Ser. B.* **48**: 133-169.
- \_\_\_\_\_ (1987). Influence assessment. *J. Appl. Statist.* **14**: 117-131.
- \_\_\_\_\_ e Weisberg, S. (1982). *Residual and Influence in Regression*. New York: Chapman Hall.
- Denby, L. e Martin, R. D. (1979). Robust estimation of the first-order autoregressive parameter. *J. Amer. Statist. Assoc.* **74**: 140-146.
- Dempster, A. P., Laird, N.M. e Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B.* **39**: 1-39.
- Deutsch, S. J., Richards, J. E., e Swain, J. J. (1990). Effects of a single outlier on ARMA identification. *Comm. in Stat.-Theory and Methods.* **19**: 2207-2227.
- Fox, A. J. (1972). Outliers in time series. *J. Roy. Statist. Soc. Ser. B.* **34**: 350-363.
- Galambos, J. (1978). *The Asymptotic Theory of the Extreme Order Statistics*. New York: Wiley.
- Gonzalez, P. (1992). Temporal aggregation and systematic sampling in structural time series model. *J. Forecast.* **11**: 271-281.
- Goodnight, J. H. (1979). A tutorial on the sweep operator. *Amer. Statist.* **33**: 149-158.
- Guttman, I. e Tiao, G. C. (1978). Effect of correlation on the estimation of a mean in the presence of spurious observations. *Canad. J. Statist.* **6**: 229-247.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, U.K.: Cambridge University Press.
- Hillmer, S. C. (1984). Monitoring and adjusting forecasts in the presence of additive outliers. *J. Forecast.* **3**: 205-215.
- Hillmer, S. C., Bell, W. R. e Tiao, G. C. (1983). Modelling considerations in the seasonal adjustment of economic time series in *Applied Time Series Analysis of Economic Data*, Washington, DC: US Bureau of Census, 74-100.
- Hotta, L. K. (1993). The effect of additive outliers on the estimates from the aggregate and disaggregate ARIMA Models. *Int. J. Forecast.* **9**: 85-93.
- Hotta, L. K. e Cardoso N., J. (1993). The effect of aggregation on prediction in autoregressive integrated moving average models. *J. Time Ser. Anal.* **14**: 261-269.

- Hotta, L. K., e Neves, M. M. C. (1992). A brief review on tests for detection of time series outliers. *Estatística*. **44**: 103-148.
- Hotta, L. K., e Pereira, P. V. (1996). Effects of outliers on forecasting temporally aggregated flow variables. Atas do Encontro de Econometria.
- Kleiner, B., Martin, R. D. e Thompson, D. J. (1979). Robust estimation of power spectra. *J. Roy. Statist. Soc. Ser. B*. **41**: 313-351.
- Koreisha e Yoshimoto (1991). Comparison among identification procedures for autoregressive moving average models. *Internat. Statist. Rev.* **59**: 37-58.
- Kunimoto, N. e Yamamoto, T. (1985). Properties of predictors in misspecified autoregressive time series models. *J. Am. Statist. Assoc.* **80**: 941-950.
- Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequence. *Z. Wahrscheinlichkeitstheorie vern, Geibiet*. **45**: 291-306.
- Ledolter, J. (1981). Parsimony and its importance in time series forecastings. *Technometrics*. **23**: 411-414.
- \_\_\_\_\_ (1987). The effect of additive outlier on the forecast from ARIMA models. *ASA Proceedings of the Business and Economic Section*, 453-458.
- \_\_\_\_\_ (1989). The effect of additive outlier on the forecast from ARIMA models. *J. Int. Forecast.* **5**: 231-240.
- \_\_\_\_\_ (1990). Outlier diagnostics in time series analysis. *J. Time Ser. Anal.* **11**: 317-324.
- Lütkepohl, H. (1986). Comparison of predictors for temporally and contemporaneously aggregated time series. *Int. J. Forecast.* **2**: 461-475.
- Martin, R. D. (1981). Robust methods in time series. in *Applied Time Series Analysis II*. ed. D. F. Findley, New York: Academic Press. 683-759.
- Martin, R. D. e Yohai, V. J. (1986). Influence functionals for time series. *Ann. Statist.* **14**: 781-818.
- Miller, R. B. (1980). Comentários em Robust estimation of autoregressive models. In *Directions in Time series* eds. D. R. Brillinger and G. C. Tiao, Hayward, CA: Institute of Mathematical Statistics.
- Mood, A. M., Graybill, F.A. e Boes, D. C. (1974). *Introduction to the Theory Statistics*. 3ª ed. Editora McGraw-Hill.
- Morettin, P. A. e Toloi, C.M.C. (1987). *Previsão de Series Temporais*. Ed. Atual. 2ª ed.

- Muirhead, C. R. (1986). Distinguishing outliers types in time series. *J. Roy. Statist. Soc. Ser. B.* **48**: 39-47.
- Peña, D. (1987). Measuring the importance of outliers in ARIMA models in *New Perspective in Theoretical and Applied Statistics*, eds. M.K. Puri, J.P. Vilapalna, e W. Wertz, New York: John Wiley, 109-118.
- \_\_\_\_\_ (1990). Influential observations in time series. *J. Bus. Econom. Statist.* **8**: 235-259.
- Pino, F.A., Morettin, P. A. e Mentz, R. P. (1987). Modelling and forecasting linear combinations of time series. *International Statistical Review*, **55**:295-314.
- Ota, R. (1993). *Testes de Escore para detecção de outliers em modelos ARMA*. (Relatório de Iniciação Científica).
- Schwartz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**: 461-464.
- Silvey, S. D. (1959). The Lagrange multiplier test. *Ann. Math. Statist.* **30**: 389-407.
- Smith, A. F.M. e West, M. (1983). Monitoring renal transplants: an application of the multiprocess Kalman filter. *Biometrika*. **39**: 867-878.
- Stoodley, K. D. C. e Mirnia, M.(1979). The automatic detection of transients, step changes and slope changes, in the monitoring medical time series. *The Statistician*. **28**: 163-170.
- Stram, D. e Wei, W. W. S. (1986). Temporal aggregation in the ARIMA process. *J. Time Ser. Anal.* **7**: 279-292.
- Teräsvirta, Y. (1995). Prof. C. W. J. Granger: an interview for the *Int. J. Forecast.* **11**: 585-590.
- Tsay, R. S. (1986a). Time serie model specification in the presence of outliers. *J. Amer. Statist. Assoc.* **81**: 132-141.
- \_\_\_\_\_ (1986b). Comment on Non-Gaussian state space modeling of nonstationary time series by Kitagawa. *J. Amer. Statist. Assoc.*
- \_\_\_\_\_ (1988). Outliers, level shifts, and variance changes in time series. *J. Forecast.* **7**: 1-20.
- Tsay, R. S. e Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *J. Amer. Statist. Assoc.* **79**: 84-96.

- Whittle, P. (1962). Gaussian estimation in stationary time series. *Bulletin of the International Statistics*. **39**: 105-129.
- Wold, H. O. (1954). *A Study in the Analysis of Stationary Time Series*. 2<sup>a</sup> ed. Almqvist and Wiksell, Uppsala.
- Wu, L. S.-Y., Hosking, J. R. M. e Ravishanker, N.(1993). Reallocation outlier in time series. *J. Roy. Statist. Soc. Ser. B*. **42**: 301-313.