

**FIGURA A5.2** Série de preços de ações da IBM com comparação das previsões do lead 3 obtidas das melhores Modelo IMA(0, 1, 1) e previsão quadrática de Brown para um período a partir de 11 de julho de 1960.

que, argumentando como na Seção 4.3.3, pode ser escrito como

$$\hat{y}_3 = \hat{y}_3 + 0\hat{y}_2\hat{y}_1 + 1\hat{y}\hat{y}_1 + 2\hat{y}_1$$

No entanto, mostramos no Capítulo 7 que se este modelo for ajustado corretamente, as estimativas de mínimos quadrados dos parâmetros são  $\hat{y}_1 = 0,02$ . Assim,  $\hat{y} = 0$  ( $1\hat{y}$ ) com  $\hat{y}$  próximo de zero, é o modelo estocástico apropriado, o polinômio de previsão apropriado é  $\hat{y}(\cdot) = \hat{y}$  que é de grau 0 em  $\hat{y}$  não de grau 2.

2. A escolha da função de peso em (A5.3.8) deve ser decidida de forma correspondente pelo modelo estocástico, e não arbitrariamente. O uso dos mínimos quadrados descontados procedimento de ajuste produziria previsões de erro quadrado médio mínimo no próprio caso restrito, onde

uma. o processo foi de ordem (0, 1, 1), então  $\hat{y} = (1\hat{y})$  b.

um polinômio de grau 0 foi ajustado, e

c. a constante de suavização foi definida igual a nosso  $\hat{y} = 1\hat{y}$ .

No presente exemplo, mesmo se o modelo polinomial correto de grau 0 tivesse sido escolhido, o valor  $\hat{y} = 0,1$ , realmente usado por Brown, teria sido bastante inadequado. O valor correto para esta série está próximo da unidade.

3. O procedimento de mínimos quadrados ponderados com desconto exponencial força todos os + coeficientes no vetor de atualização  $\mathbf{h}$  para serem funções do único parâmetro de suavização  $\hat{y}$ . Na verdade, eles devem ser funções dos + parâmetros independentes ( $\hat{y}_1, \hat{y}_2, \dots$ ).

Assim, as diferenças entre os dois métodos não são triviais, e é interessante comparar seus desempenhos nos dados da IBM. A previsão do erro quadrado médio mínimo é

TABELA A5.3 Comparação do Erro Quadrado Médio das Previsões Obtidas em Vários Leads  
Veze usando o melhor IMA( , ) Modelo e Previsões Quadráticas de Brown

	Tempo de espera									
	1	2	3	4	5	6	7	8	9	10
MSE (Marrom)	102	158 218		256	363	452 554 669			799	944
MSE ( = 0,9)	42	91	136	180	282 266 317			371	427	483

$\hat{y}(t) = 0(t)$ , com atualização igual  $\hat{y}_0^{(t)} = \hat{y}_0^{(-1)} + \dots$ , onde  $\hat{y} \geq 1,0$ . Se for considerado exatamente à unidade, isso equivale a usar

$\hat{y}(t) =$

o que implica que a melhor previsão do preço das ações para todo o tempo futuro é o presente preço.<sup>1</sup> A sugestão de que os preços das ações se comportam dessa maneira, obviamente, não é nova e vai de volta a Bachelier (1900). Como  $\hat{y} = 1$ , isso implica que é um passeio aleatório.

Para comparar a previsão do erro quadrado médio mínimo com as previsões quadráticas de Brown, uma comparação direta foi feita usando a série de preços das ações da IBM de 11 de julho de 1960 a 10 de fevereiro de 1961, para 150 observações. Para este trecho da série, o MSE mínimo a previsão é obtida usando o modelo  $\hat{y} = \hat{y}$  com  $\hat{y} = 0,1$ , ou  $\hat{y} = 0,9$ .

A Figura A5.2 mostra as previsões mínimas de MSE para o lead time 3 e o correspondente valores das previsões quadráticas de Brown. Vê-se que as previsões MSE mínimas, que são virtualmente equivalentes a usar o preço de hoje para prever que 3 dias à frente, são consideravelmente melhores do que as obtidas pelo procedimento mais complicado de Brown.

Os erros quadráticos médios para a previsão em vários prazos de entrega, calculados por comparação do valor das séries e suas previsões de liderança, são mostrados na Tabela A5.3 para os dois tipos de previsões. Vê-se que as previsões quadráticas de Brown têm quadrado médio erros que são muito maiores do que aqueles obtidos pelo método do quadrado médio mínimo.

EXERCÍCIOS

5.1. Para os modelos

- (1)  $\hat{y} = \hat{y} 0,5 \hat{y}_1 =$
- (2)  $\hat{y} = \hat{y} 0,5$  (3)  $(1 \hat{y} \hat{y}_1$
- 0,6 )  $\hat{y} =$  anotar as
- previsões para prazos de entrega = 1 e = 2:
- (a) Da equação diferencial
- (b) De forma integrada (usando os pesos)
- (c) Como média ponderada de observações anteriores

<sup>1</sup> Este resultado é aproximadamente verdadeiro supondo que nenhuma informação relevante, exceto valores passados da própria série, seja disponíveis e que períodos de previsão bastante curtos estão sendo considerados. Para períodos mais longos, o crescimento e a inflação fatores se tornariam importantes.

5.2. As seguintes observações representam os valores do  $y_1, y_2, \dots, y_{100}$  de uma série montada por modelo  $\hat{y}_t = \hat{y}_1 + 0,28 \hat{y}_2$ :

166, 172, 172, 169, 164, 168, 171, 167, 168, 172

- (a) Gere as previsões  $\hat{y}_{100}(t)$  para  $t = 1, 2, \dots, 12$  e desenhe um gráfico da série valores e as previsões (suponha  $y_0 = 0, y_1 = 0$ ).
- (b) Com  $\hat{y}_2 = 1,103$ , calcule os desvios padrão estimados  $\hat{y}(t)$  da previsão erros e usá-los para calcular limites de probabilidade de 80% para as previsões. Inserir esses limites de probabilidade no gráfico, em ambos os lados das previsões.

5.3. Suponha que os dados do Exercício 5.2 representem vendas mensais.

- (a) Calcule as previsões de erro quadrado médio mínimo para vendas trimestrais para 1, 2, 3, 4 trimestres à frente, usando os dados até  $t = 100$ .
- (b) Calcule os limites de probabilidade de 80% para essas previsões.

5.4. Usando os dados e previsões do Exercício 5.2, e dada a observação adicional  $y_{101} = 174$ :

- (a) Calcule as previsões  $\hat{y}_{101}(t)$  para  $t = 1, 2, \dots, 11$  usando a fórmula de atualização  $\hat{y}_{t+1}(t) = \hat{y}_t(t) + \frac{1}{t+1}$ .
- (b) Verifique essas previsões usando a equação das diferenças diretamente.

5.5. Para o modelo  $\hat{y}_t = \hat{y}_1 + 0,28 \hat{y}_2$  do Exercício 5.2:

- (a) Escreva expressões para os erros de previsão  $(1), (2), \dots, (6)$ , a partir do mesma origem.
- (b) Calcule e plote as autocorrelações da série de erros de previsão  $(3)$ .
- (c) Calcule e plote as correlações entre os erros de previsão  $(2)$  e  $(t)$  para  $t = 1, 2, \dots, 6$ .

5.6. Seja o vetor  $(y_1, y_2, \dots, y_T)$  tenha para seus elementos os erros de previsão feitos  $1, 2, \dots, T$  passos à frente, todos da mesma origem. Então se  $(y_1, y_2, \dots, y_T)$  são os choques aleatórios não correlacionados correspondentes, mostre que

$$= \text{onde} = \begin{matrix} & y_1 & 100 y_2 & 0 & y_3 \\ y_1 & 1 & 1 & 0 & y_4 \\ y_2 & 2 & 1 & 1 & y_5 \\ y_3 & y_4 & y_5 & y_6 & y_7 \\ y_4 & y_5 & y_6 & y_7 & y_8 \end{matrix}$$

Além disso, mostre que (por exemplo, Box e Tião, 1976; Tião et al., 1975), a matriz de covariância dos  $y_t$ 's, é  $\Sigma_y = \frac{1}{2} \Sigma_y$ , portanto, um teste  $\chi^2$  para determinar se um conjunto de os valores realizados das séries tomadas em conjunto diferem significativamente dos as previsões feitas na origem são obtidas consultando

$$\hat{y}_{t-1} = \frac{\hat{y}_t - \hat{y}_{t-1}}{2} = \frac{\hat{y}_t}{2} = \frac{1}{2} \hat{y}_t + \frac{1}{2} \hat{y}_{t-1}$$

## 176 PREVISÃO

para uma distribuição qui-quadrado com graus de liberdade. Observe que o  $\chi^2$  é o passo único erro de previsão à frente calculado a partir de  $\hat{y}_{t+1|t}$ .

**5.7.** Suponha que uma série temporal econômica trimestral seja bem representada pelo modelo

$$\tilde{y}_t = 0,5 + (1 - \tilde{\gamma}) 1,0 + 0,5 \tilde{\gamma}_t$$

com  $\tilde{\gamma}^2 = 0,04$ .

**(a)** Dado  $\tilde{y}_0 = 0,3248$  para  $t = 1$ ,  $\tilde{\gamma}_0 = 0,2$ , calcule e plote as previsões  $\tilde{y}_{48|t}$  ( )

**(b)** Calcule e insira os limites de probabilidade de 80% no gráfico.

**(c)** Expresse as séries e previsões de forma integrada.

**5.8.** Considere os números anuais de manchas solares de Wolfer para o período 1770-1869 listados como Série E na Parte Cinco deste texto. A mesma série está disponível para o período mais longo 1700--1988 como "sunspot.year" no pacote de conjuntos de dados do R. Você pode usar conjunto de dados. Suponha que a série possa ser representada por um modelo autorregressivo de ordem 3.

**(a)** Trace a série temporal e comente. A série parece estacionária?

**(b)** Gerar previsões e limites de probabilidade associados para até 20 períodos de tempo à frente da série.

**(c)** Realize uma transformação de raiz quadrada dos dados e repita (a) e (b) acima.

**(d)** Use a função `BoxCox.ar()` no pacote TSA de R para mostrar que o quadrado transformação de raiz é apropriada para esta série; veja `help(BoxCox.ar)` para detalhes. (Nota: Adicionar uma pequena quantidade, por exemplo, 1/2, à série, elimina zero valores e permite que o programa considere uma transformação de log como uma opção).

**5.9.** Uma série temporal que representa um índice global de temperatura média terrestre-oceânica de 1880 para 2009 está disponível em um arquivo chamado "gtemp" no pacote `astsa` do R. Os dados são desvios de temperatura, medidos em graus centígrados, da média de 1951-1980 temperatura, conforme descrito por Shumway e Stoffer (2011, p. 5). Suponha que um modelo autorregressivo de terceira ordem seja apropriado para as primeiras diferenças  $(1 - \tilde{\gamma})$  de esta série.

**(a)** Trace a série temporal e a série diferenciada usando R.

**(b)** Gerar previsões e limites de probabilidade associados para até 20 períodos de tempo adiante para esta série usando a função `sarima.for()` sem incluir uma constante termo no modelo.

**(c)** Gere as mesmas previsões e limites de probabilidade que na parte (b), mas com um termo constante agora adicionado ao modelo. Discuta suas descobertas e comente as implicações de incluir uma constante neste caso.

**5.10.** Para o modelo  $(1 - \tilde{\gamma} 0,6)(1 - \tilde{\gamma}) = (1 + 0,3)$ , expresse explicitamente na forma de espaço de estados de (5.5.2) e (5.5.3), e escreva precisamente as relações recursivas de o filtro de Kalman para este modelo. Indique como as previsões (exatas)  $\tilde{y}$  e suas variações de erro de previsão  $\tilde{\gamma}$  são determinados a partir dessas recursões.

## PARTE DOIS

---

### EDIFÍCIO DE MODELO ESTOCÁSTICO

Vimos que um modelo ARIMA de ordem  $(p, d, q)$  fornece uma classe de modelos capaz de representar séries temporais que, embora não necessariamente estacionárias, são homogêneas e em equilíbrio estatístico em muitos aspectos.

O modelo ARIMA é definido pela equação

$$\phi(B)(1 - B)^d \psi(B)Y_t = \theta(B) + \epsilon_t$$

onde  $\phi(B)$  e  $\psi(B)$  são operadores de grau  $p$  e  $q$ , respectivamente, cujos zeros estão fora do círculo unitário. Observamos que o modelo é muito geral, incluindo como casos especiais modelos autorregressivos, modelos de média móvel, modelos mistos autorregressivos--móveis, modelos médios e as formas integradas de todos os três.

**Abordagem Iterativa à Construção de Modelos.** O desenvolvimento de um modelo deste tipo para descrever a estrutura de dependência em uma série temporal observada geralmente é melhor alcançada por um método de três estágios procedimento iterativo baseado na identificação, estimativa e verificação diagnóstica.

1. Por *identificação* entendemos o uso dos dados e de qualquer informação sobre como o série foi gerada, para sugerir uma subclasse de modelos parcimoniosos dignos de serem entretido.
2. Por *estimativa* entendemos o uso eficiente dos dados para fazer inferências sobre o parâmetro eters condicionada à adequação do modelo entretido.
3. Por *verificação diagnóstica* queremos dizer verificar o modelo ajustado em sua relação com os dados com a intenção de revelar inadequações do modelo e, assim, alcançar a melhoria do modelo.

No Capítulo 6, a seguir, discutimos a identificação do modelo, no Capítulo 7 a estimativa de parâmetros e no Capítulo 8 a verificação de diagnóstico do modelo ajustado. No Capítulo 9 expandimos a classe de modelos desenvolvidos nos Capítulos 3 e 4 para os modelos ARIMA *saizonais*, e todas as técnicas de construção de modelos dos capítulos anteriores são ilustradas aplicando-as à modelagem e previsão de séries temporais sazonais. No Capítulo 10, consideramos alguns tópicos adicionais que representam extensões além da classe de modelos lineares ARIMA, como modelos de séries temporais heterocedásticas condicionais, modelos de séries temporais não lineares e processos de memória longa fracionalmente integrados, que permitem certos recursos mais gerais nas séries temporais do que são possíveis nos modelos lineares ARIMA. O teste de raiz unitária também é discutido neste capítulo.

---

# 6

---

## IDENTIFICAÇÃO DO MODELO

Neste capítulo, discutimos métodos para identificar modelos de séries temporais de média móvel integrada autorregressiva não sazonal (ARIMA). Os métodos de identificação são procedimentos grosseiros aplicados a um conjunto de dados para indicar o tipo de modelo que merece investigação adicional. O objetivo específico aqui é obter alguma ideia dos valores e necessidades do modelo ARIMA linear geral e obter estimativas iniciais para os parâmetros. O modelo experimental especificado fornece um ponto de partida para a aplicação dos métodos de estimativa mais formais e eficientes descritos no Capítulo 7. Os exemplos usados para demonstrar o processo de construção do modelo incluirão as Séries A-F que foram discutidas nos capítulos anteriores e estão listadas na Coleção de Séries Temporais na Parte Cinco deste livro. As séries também estão disponíveis eletronicamente em <http://pages.stat.wisc.edu/reinsel/bjr-data/>.

### 6.1 OBJETIVOS DE IDENTIFICAÇÃO

Em primeiro lugar, deve-se dizer que a identificação e a estimativa necessariamente se sobrepõem. Assim, podemos estimar os parâmetros de um modelo, que é mais elaborado do que aquele que esperamos usar, para decidir *em que ponto* a simplificação é possível. Aqui empregamos o procedimento de estimação para realizar parte da identificação. Também deve ser explicado que a identificação é necessariamente inexata. É inexata porque a questão de quais tipos de modelos ocorrem na prática e em quais casos específicos depende do comportamento do mundo físico e, portanto, não pode ser decidida por argumentos puramente matemáticos. Além disso, como na fase de identificação não está disponível uma formulação precisa do problema, métodos estatisticamente “ineficientes” devem necessariamente ser usados. É um estágio em que os métodos gráficos são particularmente úteis e o julgamento deve ser exercido. No entanto, deve-se ter em mente

---

*Análise de Séries Temporais: Previsão e Controle*, Quinta Edição. George EP Box, Gwilym M. Jenkins, Gregory C. Reinsel e Greta M. Ljung © 2016 John Wiley & Sons, Inc. Publicado em 2016 por John Wiley & Sons, Inc.

que a identificação preliminar não nos compromete a nada, exceto a consideração provisória de uma classe de modelos que mais tarde serão ajustados e verificados de forma eficiente.

### 6.1.1 Etapas do Procedimento de Identificação

Nossa tarefa, então, é identificar uma subclasse apropriada de modelos do ARIMA geral família

$$\hat{y}_t = \mu + \epsilon_t \quad (6.1.1)$$

que pode ser usado para representar uma determinada série temporal. Nossa abordagem será a seguinte:

1. Avaliar a estacionariedade do processo e, se necessário, diferenciar quantos vezes conforme necessário para produzir estacionariedade, reduzindo o processo em estudo para o processo de média móvel autoregressivo misto:

$$\hat{y}_t = \mu + \epsilon_t$$

Onde

$$\epsilon_t = (1 - B)\hat{y}_t = \hat{y}_t$$

2. Identificar o modelo de média móvel autoregressiva (ARMA) resultante para

Nossas principais ferramentas para colocar em prática as etapas 1 e 2 serão a autocorrelação de amostra função e a função de autocorrelação parcial da amostra. Eles são usados não apenas para ajudar adivinhar a forma do modelo, mas também obter estimativas aproximadas dos parâmetros. Tais aproximações são frequentemente úteis no estágio de estimativa para fornecer valores iniciais para procedimentos iterativos empregados nessa fase. Algumas ferramentas adicionais de identificação de modelo podem também podem ser empregados e são discutidos na Seção 6.2.4.

## 6.2 TÉCNICAS DE IDENTIFICAÇÃO

### 6.2.1 Uso das Funções de Autocorrelação e Autocorrelação Parcial em Identificação

**Identificando o Grau de Diferença.** Vimos na Seção 3.4.2 que para um estacionário misto autoregressivo - processo de média móvel de ordem  $(p, d, q)$ , o  $\hat{y}_t = \mu + \epsilon_t$ , função de autocorrelação satisfaz a equação de diferença

$$\hat{\gamma}_t = 0, \quad t > 0$$

Além disso, se  $\hat{\gamma}_t = 1(1 - B)\hat{y}_t$ , a solução desta equação diferencial para a autocorrelação é, assumindo raízes distintas, da forma

$$\hat{\gamma}_t = \sum_{i=1}^p \alpha_i \lambda_i^t + \sum_{j=1}^q \beta_j \lambda_j^t + \gamma \quad (6.2.1)$$

O requisito de estacionariedade de que os zeros de  $\hat{\gamma}_t$  fiquem fora do círculo unitário implica que as raízes  $\lambda_1, \lambda_2, \dots, \lambda_{p+q}$  situar-se dentro do círculo unitário.



Esta expressão mostra que no caso de um modelo estacionário em que nenhuma das raízes está próxima ao limite do círculo unitário, a função de autocorrelação irá rapidamente “apagar” para moderado e grande. Suponha agora que uma única raiz real, digamos 1, se aproxime da unidade, de modo que

$$= 1 - \gamma$$

onde  $\gamma$  é uma pequena quantidade positiva. Então, uma vez que para grandes

$$\gamma(1 - \gamma)^k$$

a função de autocorrelação não morrerá rapidamente e cairá lentamente e quase linearmente. Um argumento semelhante pode ser aplicado se mais de uma das raízes se aproximar da unidade.

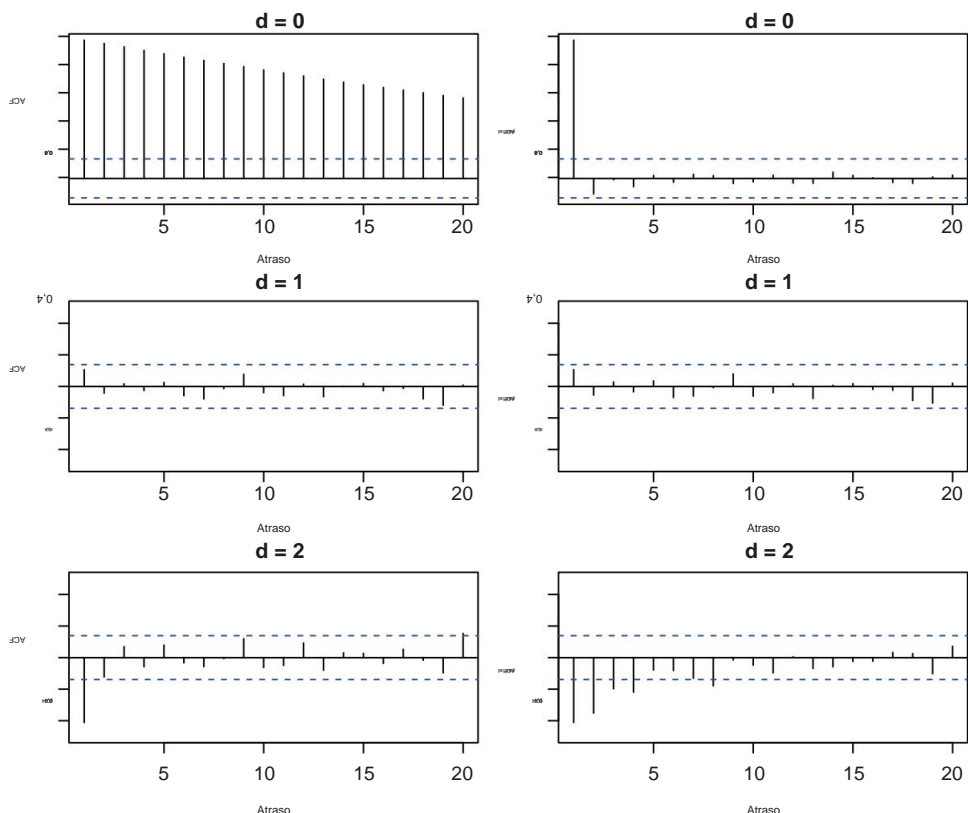
Portanto, uma tendência para a função de autocorrelação não morrer rapidamente é tomada como uma indicação de que uma raiz próxima da unidade pode existir. A função de autocorrelação estimada tende a seguir o comportamento da função de autocorrelação teórica. Portanto, a falha da função de autocorrelação estimada em morrer rapidamente pode sugerir logicamente que devemos tratar o processo estocástico subjacente como não estacionário em, mas possívelmente estacionário em  $\gamma$  ou em alguma diferença maior.

No entanto, mesmo que a falha da função de autocorrelação estimada em desaparecer rapidamente sugira não estacionariedade, as autocorrelações estimadas não precisam ser extremamente altas, mesmo em defasagens baixas. Isso é ilustrado no Apêndice A6.1, onde o comportamento esperado da função de autocorrelação estimada é considerado para o processo não estacionário  $(0, 1 - \gamma, 1)$  dos valores esperados cai apenas lentamente, mas depende inicialmente do valor de  $\gamma$  e do número de observações na série, e não precisa estar próximo da unidade se estiver próximo de 1. Ilustramos este ponto novamente na Seção 6.3.4 para Série A.

Pelas razões apresentadas, assume-se que o grau de diferenciação, necessário para atingir a estacionariedade, foi alcançado quando a função de autocorrelação de  $\gamma$  se extinguiu rapidamente. Na prática, é normalmente 0, 1 ou 2, e geralmente é suficiente inspecionar as primeiras 20 autocorrelações estimadas da série original e de sua primeira e segunda diferenças, se necessário.

**Superdiferenciação.** Uma vez que a estacionariedade é alcançada, outras diferenças devem ser evitadas. A superdiferenciação introduz uma correlação serial extra e aumenta a complexidade do modelo. Para ilustrar este ponto, suponha que a série segue um passeio aleatório de modo que a série diferenciada  $(1 - \gamma) = \epsilon$  é ruído branco e, portanto, estacionária. Diferenciação adicional de leva para  $(1 - \gamma)^2 = (1 - \gamma)$ , que é um modelo  $AR(1)$  com  $\gamma = 1$ . Assim, o modelo de ordem  $AR(1)$  com  $\gamma = 1$  é mais simples. O modelo com  $\gamma = 1$  é não inversível e a representação autoregressiva pura não existe. A não-invertibilidade também causa problemas no estágio de estimação de parâmetros, pois métodos de máxima verossimilhança aproximados tendem a produzir estimativas tendenciosas neste caso.

A Figura 6.1 mostra as funções de autocorrelação e autocorrelação parcial de uma série temporal de comprimento 200 gerada a partir de um modelo de passeio aleatório com variância de inovações igual a 1. As primeiras 1.000 observações foram descartadas para eliminar potenciais efeitos de inicialização. As autocorrelações estimadas até a defasagem 20 da série original e suas primeiras e segundas diferenças são mostradas no gráfico. As autocorrelações da série original não conseguem amortecer rapidamente,



**FIGURA 6.1** Autocorrelação estimada e funções de autocorrelação parcial para um processo de caminhada aleatória simulada e sua primeira ( $d = 1$ ) e segunda ( $d = 2$ ) diferenças.

indicando a necessidade de diferenciação. As autocorrelações de  $\hat{\gamma}_h$  por outro lado, são todos pequenos, demonstrando que a estacionariedade já foi alcançada. A autocorrelação função das segundas diferenças  $\hat{\gamma}_h^2$  também indica estacionariedade, mas tem um pico na defasagem 1 mostrando a correlação extra que surgiu por causa da superdiferença. o o valor de  $\hat{\gamma}_1^2$  é próximo a  $-0,5$ , o que é consistente com o coeficiente de autocorrelação lag 1 para um modelo MA(1) com  $\theta = 1$ . A Figura 6.1 pode ser reproduzida em R da seguinte forma:

```
> RW=arima.sim(lista(ordem=c(0,1,0)),n=200,n.start=1000)
> acf0=acf(RW,20)
> pacf0=pacf(RW,20)
> acf1=acf(dif(RW),20)
> pacf1=pacf(dif(RW),20)
> acf2=acf(dif(dif(RW)),20)
> pacf2=pacf(dif(dif(RW)),20)
> par(mfrow=c(3,2))
> plot(acf0,main='d=0')
> plot(pacf0,main='d=0')
> plot(acf1,ylim=c(-0.5,0.5),main='d=1')
> plot(pacf1,ylim=c(-0.5,0.5),main='d=1')
```

```
> plot(acf2,ylim=c(-0.5,0.5),main='d=2') >
plot(pacf2,ylim=c(-0.5,0.5),main='d=2')
```

**Identificando um Modelo ARMA Estacionário para a Série Diferenciada.** Tendo decidido provisoriamente o grau de diferenciação, examinamos os padrões de autocorrelação e autocorrelação parcial das funções de  $(1 - \hat{\gamma})$  determinamos uma escolha adequada para as ordens e dos operadores autoregressivo e de média móvel. Aqui lembramos o comportamento característico das funções de autocorrelação teórica e de autocorrelação parcial para processos de média móvel, autorregressivo e misto, discutidos no Capítulo 3.

Resumidamente, enquanto a função de autocorrelação de um processo autorregressivo de ordem  $p$  diminui, sua função de autocorrelação parcial tem um corte após a defasagem. Por outro lado, a função de autocorrelação de um processo de média móvel de ordem  $q$  tem um corte após lag  $q$ , enquanto sua função de autocorrelação parcial diminui. Se ambas as autocorrelações e as autocorrelações parciais cessarem, um processo misto é sugerido. Além disso, a função de autocorrelação para um processo misto, contendo um componente autorregressivo de  $p$  ordem e um componente de média móvel de  $q$  ordem, é uma mistura de exponenciais e ondas senoidais amortecidas após as primeiras  $q$  defasagens. Por outro lado, a função de autocorrelação parcial para um processo misto é dominada por uma mistura de exponenciais e ondas senoidais amortecidas após as primeiras defasagens (ver Tabela 3.2).

Em geral, o comportamento autorregressivo (média móvel), medido pela função de autocorrelação, tende a imitar o comportamento da média móvel (autorregressivo) medido pela função de autocorrelação parcial. Por exemplo, a função de autocorrelação de um processo autorregressivo de primeira ordem decai exponencialmente, enquanto a função de autocorrelação parcial corta após a primeira defasagem. Correspondentemente, para um processo de média móvel de primeira ordem, a função de autocorrelação é interrompida após a primeira defasagem. Embora não seja exatamente exponencial, a função de autocorrelação parcial é dominada por termos exponenciais e tem a aparência geral de uma exponencial.

De particular importância são os processos autorregressivo e de média móvel de primeira e segunda ordem e o processo misto simples  $(1, 1)$ . As propriedades das funções de autocorrelação teórica e de autocorrelação parcial para esses processos estão resumidas na Tabela 6.1, que requer um estudo cuidadoso e fornece uma tabela de referência conveniente. O leitor também deve consultar as Figuras 3.2, 3.7 e 3.10, que mostram o comportamento típico da função de autocorrelação e da função de autocorrelação parcial para o processo autorregressivo de segunda ordem, o processo de média móvel de segunda ordem e o ARMA misto simples  $(1, 1)$  processo.

### 6.2.2 Erros Padrão para Autocorrelações Estimadas e Autocorrelações Parciais

Autocorrelações estimadas podem ter variações bastante grandes e podem ser altamente autocorrelacionadas entre si. Por esta razão, não se pode esperar aderência *detalhada* à função de autocorrelação teórica na função estimada. Em particular, autocorrelações estimadas moderadamente grandes podem ocorrer após a função de autocorrelação teórica ter amortecido, e ondulações e tendências aparentes podem ocorrer na função estimada que não têm base na função teórica. Ao empregar a função de autocorrelação estimada como ferramenta de identificação, geralmente é possível ter bastante certeza sobre características amplas, mas indicações mais sutis podem ou não representar efeitos reais. Por essas razões, dois ou mais

TABLE 6.1 Behavior of the Autocorrelation Functions for the  $d$ th Difference of an ARIMA Process of Order  $(p, d, q)^a$ 

	Order				
	$(1, d, 0)$	$(0, d, 1)$	$(2, d, 0)$	$(0, d, 2)$	$(1, d, 1)$
Behavior of $\rho_k$	Decays exponentially	Only $\rho_1$ nonzero	Mixture of exponentials or damped sine wave	Only $\rho_1$ and $\rho_2$ nonzero	Decays exponentially from first lag
Behavior of $\phi_{kk}$	Only $\phi_{11}$ nonzero	Exponential dominates decay	Only $\phi_{11}$ and $\phi_{22}$ nonzero	Dominated by mixture of exponential or damped sine wave	Dominated by exponential decay from first lag
Preliminary estimates from	$\phi_1 = \rho_1$	$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2}$	$\phi_1 = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2}$	$\rho_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}$	$\rho_1 = \frac{(1 - \theta_1\phi_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}$
Admissible region	$-1 < \phi_1 < 1$	$-1 < \theta_1 < 1$	$\phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$ $-1 < \phi_2 < 1$ $\phi_2 + \phi_1 < 1$ $\phi_2 - \phi_1 < 1$	$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$ $-1 < \theta_2 < 1$ $\theta_2 + \theta_1 < 1$ $\theta_2 - \theta_1 < 1$	$\rho_2 = \phi_1\rho_1$ $-1 < \phi_1 < 1$ $-1 < \theta_1 < 1$

<sup>a</sup> Table A and Charts B–D are included at the end of this book to facilitate the calculation of approximate estimates of the parameters for first-order moving average, second-order autoregressive, second-order moving average, and the mixed ARMA(1, 1) processes.

modelos podem precisar ser entretidos e investigados mais a fundo na estimativa e diagnóstico verificando as etapas da construção do modelo.

Na prática, é importante ter alguma indicação de até que ponto um valor estimado pode diferir do valor teórico correspondente. Em particular, precisamos de alguns meios para julgar se as autocorrelações e autocorrelações parciais são efetivamente zero após algum atraso específico ou, respectivamente. Para *defasagens maiores*, na hipótese de que o processo é média móvel de ordem podemos calcular erros padrão de autocorrelações estimadas da forma simplificada da fórmula de Bartlett (2.1.15), com estimativas amostrais substituindo autocorrelações teóricas. Desta forma,

$$\hat{\gamma}[k] \pm \frac{1}{2(1-\hat{\gamma}_2^2)} \left[ 1 + \frac{2}{1 + \hat{\gamma}_2^2} + \hat{\gamma}_2^2 \right]^{1/2} \quad (6.2.2)$$

Para as autocorrelações parciais, utilizamos o resultado citado em (3.2.36) que, na hipótese de que o processo é autorregressivo de ordem o erro padrão para estimativas parciais autocorrelações de ordem + 1 e superior é

$$\hat{\gamma}[k] \pm \frac{1}{1-\hat{\gamma}_2^2} \quad (6.2.3)$$

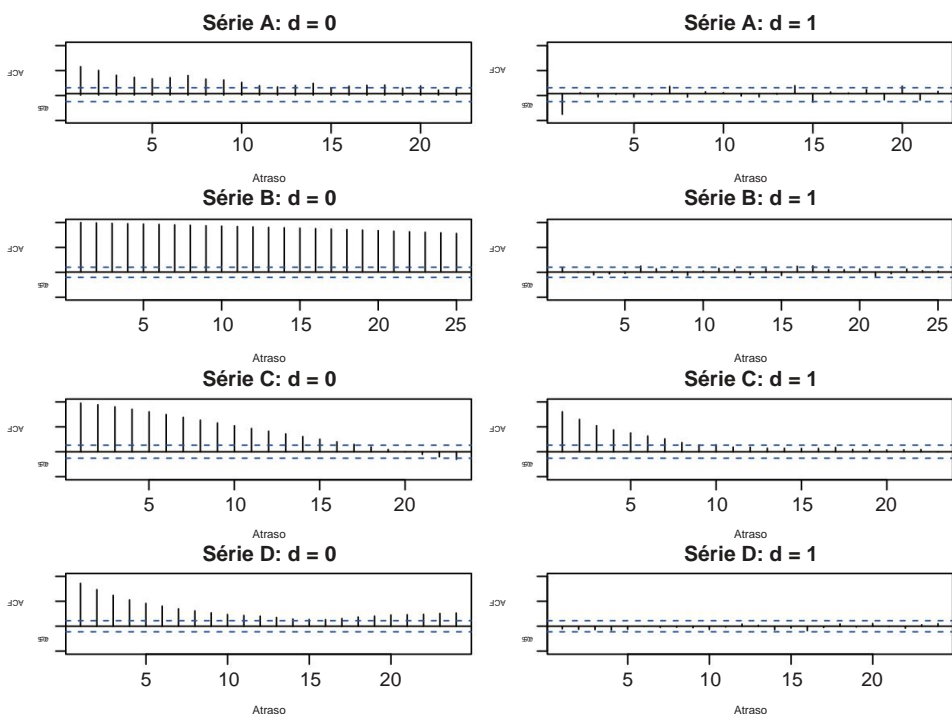
Foi demonstrado por Anderson (1942) que para , a distribuição de uma estimativa o coeficiente de autocorrelação, cujo valor teórico é zero, é aproximadamente normal. Desta forma, na hipótese de que a autocorrelação teórica é zero, a estimativa dividida pelo seu erro padrão será aproximadamente distribuído como uma variável normal unitária. Um similar resultado é verdadeiro para as autocorrelações parciais. Esses fatos fornecem um guia informal para se autocorrelações teóricas e autocorrelações parciais além de um atraso específico são essencialmente zero.

### 6.2.3 Identificação de Modelos para Algumas Séries Temporais Reais

**Série A-D.** Nesta seção, as ferramentas de especificação de modelo descritas acima são aplicadas a algumas das séries temporais reais que encontramos nos capítulos anteriores. Discutimos primeiro modelos potenciais para as Séries A a D plotados na Figura 4.1. Conforme observado no Capítulo 4 sobre não estacionaridade, esperamos que as séries A, C e D possuam características não estacionárias, uma vez que eles representam o comportamento "descontrolado" de certas saídas do processo. Da mesma forma, gostaríamos de esperar que o preço das ações da IBM Série B não tenha um nível fixo e seja não estacionário.

As autocorrelações estimadas e as primeiras diferenças  $\hat{\gamma}$  para a Série A-D são mostrados na Figura 6.2. A Figura 6.3 mostra as correspondentes autocorrelações parciais estimadas. As duas figuras foram geradas em R usando comandos semelhantes aos usados para produzir Figura 6.1. Para as leituras de concentração do processo químico na Série A, as autocorrelações para  $\hat{\gamma}$  são pequenos após a primeira defasagem. Isso sugere que essa série temporal pode ser descrita por um modelo IMA(0, 1, 1). No entanto, a partir da função de autocorrelação dele é visto que após lag 1 as correlações diminuem bastante regularmente. Portanto, uma alternativa é que a série segue um modelo misto ARMA(1, 0, 1). A função de autocorrelação parcial de parece apoiar esta possibilidade. Veremos mais adiante que as duas alternativas resultam virtualmente na mesmo modelo. Para o preço das ações Série B, os resultados confirmam a não estacionariedade do série original e sugerem que um modelo de passeio aleatório  $(1 - \hat{\gamma}) = 1$  é apropriado para isso Series.

As autocorrelações estimadas da Série C de temperatura também indicam não estacionaridade. O decaimento aproximadamente exponencial nas autocorrelações para a primeira diferença sugere um



**FIGURA 6.2** Funções de autocorrelação estimadas da série original ( $d = 0$ ), e seus primeiros diferenças ( $d = 1$ ) para a Série A-D.

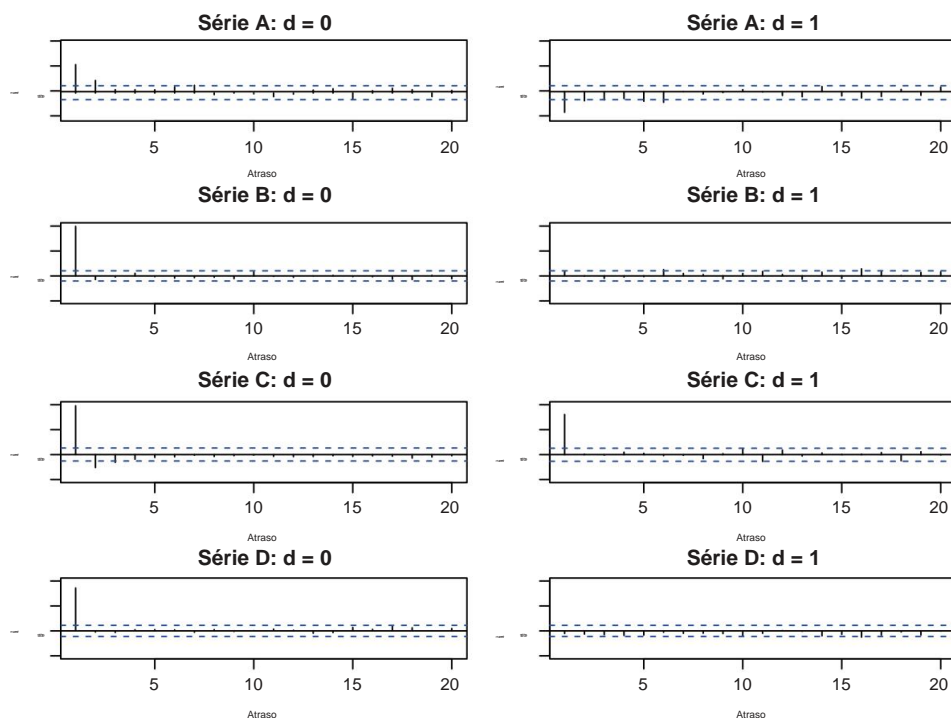
processo de ordem  $(1, 1, 0)$ , com parâmetro autorregressivo em torno de 0,8. Alternativamente, notamos que as autocorrelações de  $\hat{y}$  decaem a uma taxa relativamente lenta, sugerindo que uma diferenciação adicional pode ser necessária. A autocorrelação e a autocorrelação parcial funções das segundas diferenças  $\hat{y}_2$  (não mostradas) foram bastante pequenas, sugerindo um branco processo de ruído para as segundas diferenças. Isso implica que um modelo IMA(0, 2, 0) pode também ser apropriado para esta série. Assim, as possibilidades são

$$(1 - 0,8B)(1 - B^2) =$$

O segundo modelo é muito semelhante ao primeiro, diferindo apenas na escolha de 0,8 em vez de 1,0 para o coeficiente autorregressivo.

Finalmente, as funções de autocorrelação e autocorrelação parcial para a viscosidade A série D sugere que um modelo AR(1)  $(1 - B)$  com cerca de 0,8 pode ser adequado para esta série. Alternativamente, uma vez que os coeficientes de autocorrelação decaem em um relativamente lenta, também consideraremos o modelo  $(1 - B)$  para esta série.

**Série E e F.** Série E mostrada no gráfico superior da Figura 6.4 representa o "Wolfes anual" números de manchas solares no período de 1770 a 1869. Esta série é provavelmente estacionária, uma vez que espera-se que o número de manchas solares permaneça em equilíbrio por longos períodos de tempo. As funções de autocorrelação e autocorrelação parcial na Figura 6.4 mostram características semelhantes aos de um processo AR(2). No entanto, como será visto mais adiante, um sistema marginalmente melhor



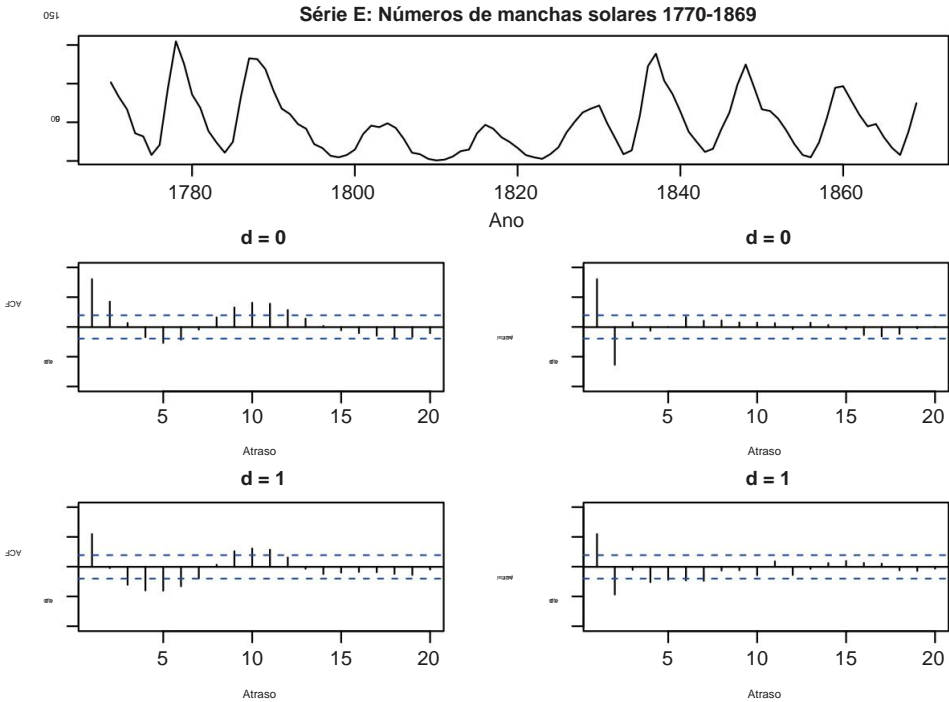
**FIGURA 6.3** Funções de autocorrelação parcial estimadas da série original ( $d = 0$ ), e suas primeiras diferenças ( $d = 1$ ) para a Série A--D.

o ajuste é obtido usando um modelo AR(3). O ajuste pode ser melhorado ainda mais usando uma raiz quadrada ou transformação logarítmica da série. Um modelo autorregressivo de ordem nove é sugerido por a rotina de seleção de ordem `ar()` no pacote R que seleciona a ordem AR com base no Akaike critério de informação (AIC) a ser discutido na Seção 6.2.4. Outras opções consideradas na literatura incluem modelos de séries temporais não lineares, como bilinear ou autorregressivo de limiar modelos, discutidos brevemente na Seção 10.3.

A série F introduzida no Capítulo 2 representa os rendimentos de um processo químico descontínuo. Espera-se que a série seja estacionária, uma vez que os lotes são processados sob condições uniformemente condições controladas. A estacionariedade é confirmada pela Figura 6.5 que mostra um gráfico de a série juntamente com as funções de autocorrelação e autocorrelação parcial da série e suas primeiras diferenças. Os resultados para a série indiferenciada sugerem que uma modelo autorregressivo pode ser apropriado para esta série.

Um resumo dos modelos provisoriamente identificados para as Séries A a F é apresentado na Tabela 6.2. Observe que para as Séries C e F, os modelos alternativos sugeridos acima foram feitos um pouco mais geral para mais ilustrações mais tarde.

**Notas sobre o procedimento de identificação.** Os gráficos das funções de autocorrelação e autocorrelação parcial mostrados acima foram gerados usando R. Na avaliação da estimativa funções de correlação, é muito útil traçar um ou dois limites de erro padrão em torno de zero para os coeficientes estimados. Os limites do pacote R estão incluídos nos gráficos exibidos acima. Esses limites são aproximadamente dois limites de erro padrão,  $\pm 2\hat{\sigma}(\hat{\rho}_k)$ , determinados



**FIGURA 6.4** Autocorrelação estimada e funções de autocorrelação parcial da série de manchas solares (Série E) e suas primeiras diferenças.

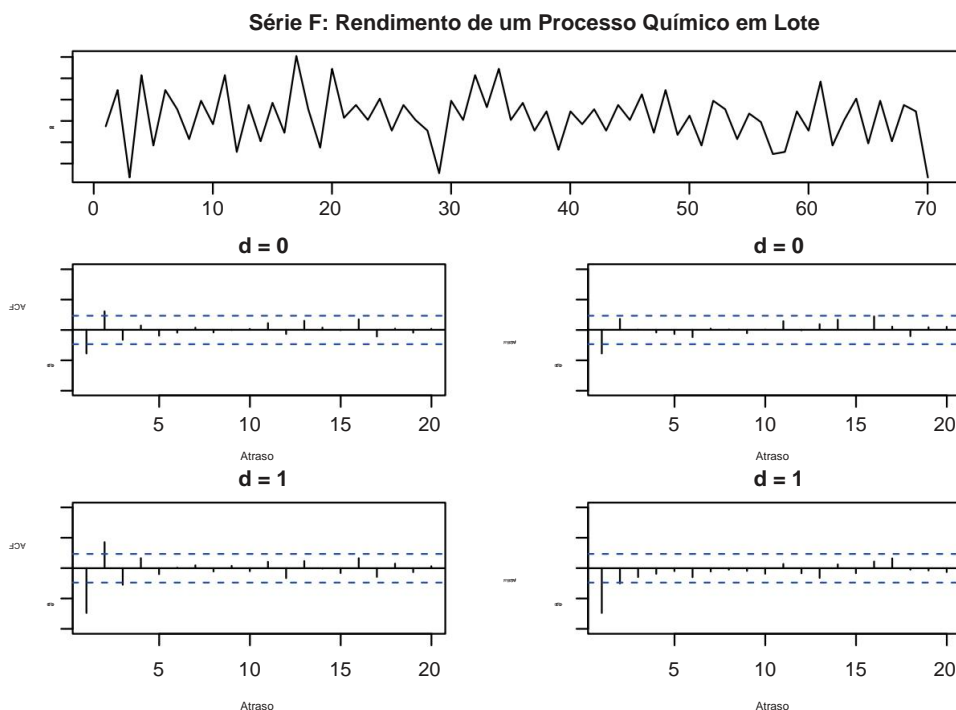
sob a suposição de que todos os coeficientes teóricos de autocorrelação são zero, de modo que a série é ruído branco. Se uma hipótese sobre um modelo específico for postulada, alternativas limites podem ser determinados a partir da fórmula de Bartlett como discutido acima. Quando os cálculos são realizados em R, inclusão do argumento `ci.type='ma'` na função `acf()`

**TABELA 6.2** Identificação Provisória de Modelos para Série A--F

Series	Grau de Diferenciando	Natureza Aparente da Série Diferenciada	Identificação por
UMA	Ou 0	AR misto de primeira ordem com MA de primeira ordem	(1, 0, 1)
	ou 1	MA de primeira ordem	(0, 1, 1)
B	1	MA de primeira ordem	(0, 1, 1)
C	Ou 1	AR de primeira ordem	(1, 1, 0)
	ou 2	Ruído não correlacionado	(0, 2, 2)
D	Ou 0	AR de primeira ordem	(1, 0, 0)
	ou 1	Ruído não correlacionado	(0, 1, 1)
E	Ou 0	AR de segunda ordem	(2, 0, 0)
	ou 0	AR de terceira ordem	(3, 0, 0)
F	0	AR de segunda ordem	(2, 0, 0)

A ordem do operador de média móvel parece ser zero, mas a forma mais geral é mantida para consideração posterior.





**FIGURA 6.5** Autocorrelação estimada e funções de autocorrelação parcial para o rendimento de um lote processo químico (Série F) e suas primeiras diferenças.

produz limites de confiança calculados com base na suposição de que o modelo verdadeiro é  $E(\tilde{y} | 1)$ .

Três outros pontos relativos a este procedimento de identificação precisam ser mencionados:

1. Diferenciação simples do tipo que usamos não produzirá estacionariedade em série contendo componentes sazonais. No Capítulo 9, discutimos as modificações apropriadas para essas séries temporais sazonais.
2. Conforme discutido no Capítulo 4, um valor diferente de zero para  $\tilde{y}$  em (6.1.1) implica a existência de uma tendência polinomial sistemática de grau  $\tilde{y}$ . Para os modelos não estacionários na Tabela 6.2, um valor de  $\tilde{y} = 0$  pode perfeitamente explicar o comportamento da série. Ocasionalmente, no entanto, haverá algum fenômeno físico real exigindo a fornecimento de tal componente. Em outros casos, pode ser incerto se tal disposição deve ser feita. Alguma indicação das provas fornecidas pelo dados, para inclusão no modelo, podem ser obtidos na fase de identificação comparando a média de  $\tilde{y}$  com seu erro padrão aproximado, usando
 
$$2(\overline{\tilde{y}}) = \frac{\tilde{y}^2}{[1 + 2(1) + 2(2) + \tilde{y}]}$$
3. Foi observado na Seção 3.4.2 que, para qualquer processo  $ARMA(\tilde{y}, \tilde{y})$  com  $\tilde{y} > 0$ , o *metade positiva inteira* da função de autocorrelação será uma mistura de seno amortecido ondas e exponenciais. Isso não nos impede, é claro, de identificar provisoriamente  $\tilde{y}$ , porque (a) a função de autocorrelação parcial mostrará  $\tilde{y}$  "anômalo" valores antes de se comportar como um processo  $MA(\tilde{y})$ , e (b) deve ser tal que o

função de autocorrelação pode tomar, como valores iniciais seguindo o padrão geral, de volta a  $\hat{\gamma}(\hat{\gamma} - 1)$ .

#### 6.2.4 Algumas Ferramentas Adicionais de Identificação de Modelo

Embora as funções de autocorrelação de amostra e autocorrelação parcial sejam extremamente útil na identificação de modelos, às vezes há casos envolvendo modelos mistos em que eles podem fornecer resultados ambíguos. Isso pode não ser um problema sério, uma vez que, como foi enfatizado, a especificação do modelo é sempre provisória e sujeita a um exame mais aprofundado, verificação de diagnóstico e modificação, se necessário. No entanto, tem havido um interesse considerável no desenvolvimento de ferramentas adicionais para uso na fase de identificação do modelo. Esses incluem a abordagem de matriz R e S proposta por Gray et al. (1978), a parcial generalizada função de autocorrelação estudada por Woodward e Gray (1981), a função de autocorrelação inversa considerada por Cleveland (1972) e Chatfield (1979), a amostra estendida função de autocorrelação de Tsay e Tiao (1984), e o uso de análise de correlação canônica como examinado por Akaike (1976), Cooper e Wood (1982) e Tsay e Tiao (1985). Critérios de seleção de modelos, como o critério AIC introduzido por Akaike (1974a) e o O Critério de Informação Bayesiano (BIC) de Schwarz (1978) também são úteis complementares

Ferramentas.

**Métodos de Correlação Canônica.** Para ilustração, discutimos brevemente o uso de análise de correlação para identificação do modelo. Em geral, para dois conjuntos de variáveis aleatórias,

$\mathbf{y}_1 = (y_{11}, y_{12}, \dots, y_{1p})^T$  e  $\mathbf{y}_2 = (y_{21}, y_{22}, \dots, y_{2q})^T$ , de dimensões  $p$  e  $q$  (suponha  $p \geq q$ ), a análise de correlação canônica envolve a determinação de combinações lineares  $\mathbf{e}_1 = e_{11}y_{11} + \dots + e_{1p}y_{1p}$  e  $\mathbf{e}_2 = e_{21}y_{21} + \dots + e_{2q}y_{2q}$  e correlações correspondentes  $\rho = \text{corr}[\mathbf{e}_1, \mathbf{e}_2]$  com  $(1) \hat{\gamma}$

$(2) \hat{\gamma} \hat{\gamma}^T (\hat{\gamma} - 1) \hat{\gamma} \geq 0$ . As combinações lineares são escolhidas de modo que  $\mathbf{e}_1$  e  $\mathbf{e}_2$  sejam mutuamente não correlacionado para  $\hat{\gamma}$ ,  $\mathbf{e}_1$  ter a máxima correlação possível (1) entre todas as combinações lineares de  $\mathbf{y}_1$  e  $\mathbf{e}_2$  ter a máxima correlação possível (2) entre todas as combinações lineares de  $\mathbf{y}_2$  que não são correlacionados com  $\mathbf{e}_1$  e assim de on. As correlações resultantes  $(\hat{\gamma})$  são chamadas de *correlações canônicas* entre

$\mathbf{y}_1$  e  $\mathbf{y}_2$  e as variáveis  $\mathbf{e}_1$  e  $\mathbf{e}_2$  são as variáveis canônicas correspondentes. Se  $\Sigma = \text{cov}[\mathbf{y}]$  denota a matriz de covariância de  $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)^T$ , com  $\Sigma = \text{cov}[\mathbf{y}]$ , então é conhecido

os valores  $2(\hat{\gamma})$  são os autovalores ordenados da matriz  $\begin{bmatrix} \hat{\gamma}_1 & \hat{\gamma}_2 \\ 11 & 12 & 22 & 21 \end{bmatrix}$  e a vetores  $\mathbf{v}_1, \dots, \mathbf{v}_q$ , tal que  $\mathbf{v}_i^T \Sigma \mathbf{v}_i = 1$ , são os autovetores correspondentes (normalizados); isso é,  $\mathbf{v}_i^T \Sigma \mathbf{v}_i = 1$  e satisfaz

$$[\mathbf{v}_1, \dots, \mathbf{v}_q]^T \begin{bmatrix} \hat{\gamma}_1 & \hat{\gamma}_2 \\ 11 & 12 & 22 & 21 \end{bmatrix} = \mathbf{0} \quad \mathbf{v}_i^T \Sigma \mathbf{v}_i = 1, \dots, q \quad (6.2.4)$$

com  $2(1) \hat{\gamma} \hat{\gamma}^T (\hat{\gamma} - 1) \hat{\gamma} \geq 0$  (por exemplo, Anderson (1984), p. 490). Da mesma forma, pode-se definir a noção de *correlações canônicas parciais* entre  $\mathbf{y}_2$ , dado outro conjunto

das variáveis  $\mathbf{y}_3$ , como as correlações canônicas entre  $\mathbf{y}_2$  e  $\mathbf{y}_3$  depois de terem sido "ajustado" para os efeitos de  $\mathbf{y}_1$  por regressão linear em  $\mathbf{y}_3$ , análoga à definição de

correlações parciais conforme discutido na Seção 3.2.5. Uma propriedade útil a ser observada é que, se houver existem (pelo menos)  $q$  combinações lineares linearmente independentes de  $\mathbf{y}_2$  que são completamente não correlacionado com  $\mathbf{y}_1$ , então existem (em  $q$ ) combinações lineares linearmente independentes de  $\mathbf{y}_2$  que são completamente não correlacionado com  $\mathbf{y}_1$  e  $\mathbf{y}_3$  e, portanto, existem  $q$  correlações canônicas entre  $\mathbf{y}_2$  e  $\mathbf{y}_3$  que seguem de (6.2.4) vez que haverá (pelo menos) autovetores linearmente independentes satisfazendo (6.2.4) com

correspondente  $\gamma_{\tilde{y}\tilde{y}}(0) = 0$ . Com efeito, então, o número de zero correlações canônicas é igual para  $\tilde{y}$  e  $\tilde{y}_1$ , onde  $\tilde{y}_1 = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_p)$ .

No contexto do modelo de séries temporais ARMA, seguindo a abordagem de Tsay e Tiao (1985), consideramos  $\tilde{y}$  e examinamos a estrutura de correlação canônica entre as variáveis  $\tilde{y}$  e  $\tilde{y}_1$ .

$$\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_p)$$

para várias combinações de  $p = 0, 1, \dots$  e  $q = 0, 1, \dots$ .

Uma característica importante a ser lembrada é que o

função de autocovariância de um processo ARMA( $p, q$ ) satisfaz (3.4.2), e, em particular,

$$\gamma_{\tilde{y}\tilde{y}}(k) = 0 \quad k > p + q$$

Assim, por exemplo, se  $\tilde{y}$  existe (pelo menos) uma combinação linear de

$$\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_p, 0, \dots, 0) \quad (6.2.5)$$

de tal modo que

$$\tilde{y} = \tilde{y}_1$$

que não está correlacionado com  $\tilde{y}_1$  para  $\tilde{y}$ . Em particular, então, para  $\tilde{y} = \tilde{y}_1$ , lá uma correlação canônica zero entre  $\tilde{y}$  e  $\tilde{y}_1$ , bem como entre  $\tilde{y}$  e  $\tilde{y}_1$ ,  $\gamma_{\tilde{y}\tilde{y}}(k) = 0$  para  $k > 1$ . É interessante observar que existem  $\min(p+1, q+1)$  zero correlações canônicas entre  $\tilde{y}$  e  $\tilde{y}_1$ .

Assim, percebe-se que a determinação da estrutura das correlações canônicas zero entre  $\tilde{y}$  e  $\tilde{y}_1$  para vários valores de  $p$  e  $q$  para caracterizar as ordens  $p$  e  $q$  do modelo ARMA, e assim as correlações canônicas serão úteis na identificação do modelo. Observamos os casos especiais de essas correlações canônicas são as seguintes. Primeiro, quando  $p = 0$ , estamos simplesmente examinando as autocorrelações que serão todas iguais a zero em um processo MA( $q$ ) para  $\tilde{y}$ . Em segundo lugar, quando  $q = 0$ , estamos examinando as autocorrelações parciais entre  $\tilde{y}_1$ , dado  $\tilde{y}_1, \dots, \tilde{y}_p$ , e todas serão iguais a zero em um AR( $p$ ) processo para  $\tilde{y}$ . Assim, a análise de correlação canônica pode ser vista como uma extensão da análise das funções de autocorrelação e autocorrelação parcial do processo.

Na prática, com base em (6.2.4), leva-se a considerar as correlações canônicas da amostra  $\hat{\gamma}(\cdot)$ , que são determinados a partir dos autovalores da matriz:

$$\begin{pmatrix} \hat{\gamma}_{\tilde{y}\tilde{y}}(0) & \hat{\gamma}_{\tilde{y}\tilde{y}_1}(0) \\ \hat{\gamma}_{\tilde{y}_1\tilde{y}}(0) & \hat{\gamma}_{\tilde{y}_1\tilde{y}_1}(0) \end{pmatrix} \times \begin{pmatrix} \hat{\gamma}_{\tilde{y}\tilde{y}}(1) & \hat{\gamma}_{\tilde{y}\tilde{y}_1}(1) \\ \hat{\gamma}_{\tilde{y}_1\tilde{y}}(1) & \hat{\gamma}_{\tilde{y}_1\tilde{y}_1}(1) \end{pmatrix} \quad (6.2.6)$$

para vários valores de atraso  $k = 0, 1, \dots$  e  $l = 0, 1, \dots$ . Tsay e Tiao (1985) usam uma abordagem estatística de teste qui-quadrado com base no menor autovalor (amostra quadrada canônica correlação)  $\hat{\gamma}(\cdot)$  de (6.2.6). Eles propõem a estatística  $(\hat{\gamma}(\cdot)) = \hat{\gamma}(\tilde{y}\tilde{y}) \ln[1 - \hat{\gamma}(\tilde{y}\tilde{y})]$

$\tilde{y}(k), \tilde{y}(k-1), \dots, \tilde{y}(k-m+1)$ , onde  $m = 1+2\tilde{y}$   $\tilde{y}^2(k-1), > 0, (\tilde{y}(k-1))$  denota a amostra au  $\tilde{y}(k)$

a correlação na defasagem de são estimativas de  $\tilde{y}_1$  e a

os  $\tilde{y}$ 's obtidos do autovetor (veja, por exemplo, equação (6.2.5)) correspondente

para  $\tilde{y}(k)$ . A estatística  $(\tilde{y}, k)$  tem uma assintótica  $\tilde{y}_1^2$  distribuição quando  $\tilde{y} = e$  e  $\tilde{y}$

ou quando  $\tilde{y} = e$  e pode ser usado para testar se existe um zero canônico

correlação em teoria. Portanto, se as estatísticas da amostra exibem um padrão tal que são todas insignificantes, em relação a uma  $\tilde{y}_1^2$  distribuição, para  $\tilde{y}$  e  $\tilde{y}$  para alguns e valores, então

o modelo pode ser razoavelmente identificado como um ARMA( $\tilde{y}, k$ ) para os menores valores  $(\tilde{y}, k)$

tal que este padrão se mantenha. Tsay e Tião (1985) também mostram que este procedimento é válido para modelos ARIMA não estacionários  $(\tilde{y}, k) = (\tilde{y}, k)$  do operador  $\tilde{y}$ , no sentido de que a ordem geral +

AR generalizado  $(\tilde{y}, k)$  podem ser determinados pelo procedimento, sem inicialmente

decidir sobre a diferenciação da série original Os

métodos de correlação canônica também foram previamente propostos para modelagem ARMA por

Akaike (1976) e Cooper e Wood (1982). Sua abordagem é realizar uma análise de correlação canônica

entre o vetor de valores presentes e passados,  $\tilde{y}(k-1)$  e o vetor de valores futuros, Na prática,  $\tilde{y}(k) =$

$\tilde{y}_1, \dots, \tilde{y}_m$   $\tilde{y}(k-1)$   $\tilde{y}_1 = (\tilde{y}_1 + 1, \tilde{y}_2 + 1, \dots) \tilde{y}$ .

o atraso finito usado para construir o vetor de valores presentes e passados pode ser fixado por uso de um critério de determinação de pedido, como os critérios de informação Akaike a serem discutidos

um pouco mais adiante nesta seção, aplicado ao ajuste de modelos AR de várias ordens. O canônico

a análise de correlação é realizada sequencialmente adicionando elementos a  $\tilde{y}_1 + 1$  um de cada vez,

começando com  $\tilde{y}_1 = (\tilde{y}_1 + 1)$ , até a primeira correlação canônica zero entre e o

$\tilde{y}_1 + 1$  está determinado. Akaike (1976) usa um critério do tipo AIC chamado informação de desvio critério (DIC) para julgar se a menor correlação canônica da amostra pode ser tomada

ser zero, enquanto Cooper e Wood (1982) usam uma abordagem estatística tradicional de qui-quadrado para avaliar a significância da menor correlação canônica, embora como apontado por

Tsay (1989a), para ser válida na presença de um componente de média móvel, esta estatística precisa ser modificado.

Em um determinado estágio do procedimento, quando a menor correlação canônica da amostra entre e é a variável mais recente a ser julgado como 0 e  $\tilde{y}_1 + 1$

incluído em  $\tilde{y}_1$  uma combinação linear de  $\tilde{y}_1 + 1$   $\tilde{y}_1 + 1$ , em relação aos restantes elementos de

$\tilde{y}_1 + 1$  é identificado que não está correlacionado com o passado. Especificamente, a combinação linear

$\tilde{y}_1 + 1$   $\tilde{y}_1 \tilde{y}_1 = 1$   $\tilde{y}_1 + 1$  dos elementos no vetor  $\tilde{y}_1$  de valores futuros é (em teoria)

determinado a ser não correlacionado com o passado. Portanto, esta análise de correlação canônica procedimento determina que as previsões  $\tilde{y}(k+1)$  do processo satisfazem

$$\tilde{y}(k+1) - \tilde{y}_1 \tilde{y}_1 = 0$$

Por referência à relação (5.3.2) na Seção 5.3, para um processo estacionário, isso implica que um modelo ARMA é identificado para o processo, com  $\tilde{y} = \max\{\tilde{y}_1, \tilde{y}_2\}$ .

Como pode ser visto, na notação de Tsay e Tiao (1985), os métodos de Akaike e

Cooper e Wood representam a análise de correlação canônica entre para vários  $\tilde{y}_1 = 1, \tilde{y}_2 = 2, \dots$  e

Como o método de Tsay e Tiao considera a correlação canônica

análise entre para várias combinações de  $\tilde{y}_1 = 0, 1, \dots$  e  $\tilde{y}_2 =$

$0, 1, \dots$ , é mais geral e, em princípio, é capaz de fornecer informações sobre o

pedidos e das partes AR e MA do modelo separadamente, em vez de apenas o máximo

desses dois valores. Na prática, ao usar os métodos de Akaike e Cooper e Wood,

as informações mais detalhadas sobre os pedidos individuais e seria determinado no etapa de estimativa de máxima verossimilhança dos parâmetros do ARMA( , ) modelo.

**Uso de Critérios de Seleção de Modelos.** Outra abordagem para a seleção de modelos envolve o uso de critérios de informação como o AIC proposto por Akaike (1974a) ou o critério de informação Bayesiano de Schwarz (1978). Na implementação desta abordagem, uma série de potenciais modelos ARMA são estimados por métodos de máxima verossimilhança a serem discutidos no Capítulo 7, e para cada modelo, um critério como AIC (normalizado pelo tamanho da amostra), dado por

$$AIC = \frac{\bar{y}^2 \ln(\text{probabilidade maximizada}) + 2}{\bar{y}^2} \ln(\bar{y}^2) + \frac{2}{\bar{y}^2} + \text{constante}$$

ou o BIC relacionado fornecido por

$$BIC = \ln(\bar{y}^2) + \frac{\ln(\bar{y}^2)}{\bar{y}^2}$$

é avaliado. Aqui,  $\bar{y}^2$  é a estimativa de máxima verossimilhança de  $\sigma^2$ , número de parâmetros estimados, incluindo um termo constante. Nos critérios acima, o primeiro termo corresponde essencialmente a  $\bar{y}^2$  vezes o logaritmo da verossimilhança maximizada, enquanto o segundo termo é um "fator de penalidade" para inclusão de parâmetros adicionais no modelo. Dentro na abordagem dos critérios de informação, os modelos que geram um valor mínimo para o critério são a ser preferido, e os valores de AIC ou BIC são comparados entre vários modelos como o base para a escolha do modelo. Assim, uma vez que o critério BIC impõe uma penalidade maior para o número de parâmetros de modelo estimados do que AIC, uso de BIC mínimo para seleção de modelo sempre resultaria em um modelo escolhido cujo número de parâmetros não é maior do que o escolhido pela AIC.

Hannan e Rissanen (1982) propuseram um procedimento de seleção de modelo em duas etapas que evita a necessidade de maximizar a função de verossimilhança para múltiplas combinações de  $p$  e  $q$ . No primeiro passo, ajusta-se um modelo AR de ordem suficientemente alta  $p$  do modelo AR(  $p$  ) ajustado fornece estimativas das inovações no ARMA(  $p$ ,  $q$  ) modelo. Na segunda etapa, regride-se por vários  $q$  sobre  $\hat{y}_1, \dots, \hat{y}_p$  e  $\hat{y}_1, \dots, \hat{y}_p$ . Ou seja, ajustam-se modelos aproximados da forma

$$\hat{y}_t = \sum_{i=1}^p \hat{y}_{t-i} + \epsilon_t \quad (6.2.7)$$

usando mínimos quadrados ordinários, e a variância de erro estimada, não corrigida para graus de liberdade, é denotado por  $\hat{\sigma}^2$ . Então, usando o critério BIC, a ordem (  $p$ ,  $q$  ) do ARMA o modelo é escolhido como aquele que minimiza  $\ln(\hat{\sigma}^2) + \ln(\hat{\sigma}^2)$ . Hannan e Rissanen mostram que, em condições muito gerais, os estimadores de  $p$  e  $q$  escolhidos dessa maneira tendem quase certamente aos valores verdadeiros. O recurso deste procedimento é que o cálculo de estimativas de máxima verossimilhança em uma ampla gama de modelos ARMA possíveis são evitadas.

Embora esses procedimentos de seleção de pedidos sejam úteis, eles devem ser vistos como ferramentas complementares para auxiliar no processo de seleção de modelos. Em particular, não devem ser usado como um substituto para o exame cuidadoso da autocorrelação estimada e parcial funções de autocorrelação da série e exame crítico dos resíduos  $\hat{y}$  de

um modelo ajustado deve sempre ser incluído como uma parte importante da seleção geral do modelo processo.

### 6.3 ESTIMATIVAS INICIAIS PARA OS PARÂMETROS

#### 6.3.1 Unicidade das Estimativas Obtidas da Função de Autocovariância

Embora um determinado modelo ARMA tenha uma estrutura de autocovariância única, a recíproca não é verdadeira sem convenções adicionais impostas para exclusividade, como discutiremos posteriormente. No à primeira vista, isso parece descartar o uso das autocovariâncias estimadas como um meio de identificação. No entanto, mostramos na Seção 6.4 que a função de autocovariância estimada pode realmente ser usado para este fim. A razão é que, embora exista uma multiplicidade dos modelos ARMA possuindo a mesma função de autocovariância, existe apenas um que expressa o valor atual de  $\tilde{y}$ , exclusivamente em termos de histórico anterior e em forma inversível estacionária.

#### 6.3.2 Estimativas iniciais para processos de média móvel

Como mostrado no Capítulo 3, as primeiras autocorrelações de um processo MA( ) são diferentes de zero e podem ser escrito em termos dos parâmetros do modelo como

$$= \frac{\tilde{y} + 1 + 1 + 2 + 2 + \tilde{y} + \dots}{1 + \frac{2}{1} + \frac{2}{1} + \tilde{y} + 2 + \frac{2}{1} + \dots} = 1, 2, \dots, \quad (6.3.1)$$

A expressão (6.3.1) para  $1, 2, \dots$ , em termos de  $1, 2, \dots$ , fornece equações em desconhecidos. As estimativas preliminares dos 's podem ser obtidas substituindo as estimativas para em (6.3.1) e resolvendo as equações não lineares resultantes. Uma estimativa preliminar do  $^2$  pode então ser obtido de

$$_0 = \frac{2}{(1 + \frac{2}{1} + \tilde{y} + 1 + \frac{2}{1})}$$

substituindo as estimativas preliminares de 's e substituindo  $_0 = \frac{2}{1}$  por sua estimativa  $_0$ . Os valores numéricos dos coeficientes de autocorrelação estimados  $_0$  para a série z

são convenientemente obtidos de R como se segue:

```
>ac=acf(z)
>e
```

**Estimativas preliminares para um processo (0, d, 1).** A Tabela A na Parte Cinco relaciona e, por  $_1$ , substituindo  $1( )$  por  $_1$  pode ser usado para fornecer estimativas iniciais para qualquer processo  $(0, , 1)$   $= (1 \tilde{y})$  onde  $= \tilde{y}$ .

**Estimativas preliminares para um processo (0, d, 2).** O Quadro C na Parte Cinco relaciona  $_1$  e  $_2$  para  $_1$  e  $_2$ , e substituindo  $1( )$  e  $2( )$  por e  $_1$  podem ser usados para fornecer estimativas para qualquer  $(0, , 2)$  processo.

Ao obter estimativas preliminares desta forma, os seguintes pontos devem ser mantidos em mente:

1. As autocovariâncias são segundos momentos da distribuição conjunta dos 's. Desta forma, as estimativas dos parâmetros são obtidas igualando os momentos amostrais aos seus valores teóricos valores. Sabe-se que o *método dos momentos* não é necessariamente eficiente e pode produzir estimativas ruins para modelos que incluem termos de média móvel. No entanto, as estimativas aproximadas obtidas podem ser úteis na obtenção de estimativas totalmente eficientes, porque eles fornecem uma ideia aproximada de "onde no espaço do parâmetro procurar" para as estimativas mais eficientes.
2. Em geral, a equação (6.3.1), obtida por equacionamento de momentos, terá múltiplos soluções. Por exemplo, quando  $\gamma_1 = 1$ ,

$$\gamma_1 = \frac{1}{1 + \gamma_2} \quad (6.3.2)$$

e daí de  $\gamma_1^2 + (1 - \gamma_1) \gamma_1 + 1 = 0$ , vemos que ambos

$$\gamma_1 = \frac{1}{2} \quad \text{e} \quad \gamma_1 = \frac{1}{1 + [1 - (2 - 1)^2 - \gamma_1] \gamma_2}$$

e

$$\gamma_1 = \frac{1}{2} \quad \text{e} \quad \gamma_1 = \frac{1}{1 - \gamma_1 [1 - (2 - 1)^2 - \gamma_1] \gamma_2} \quad (6.3.3)$$

são soluções possíveis. Para ilustração, a primeira autocorrelação de defasagem da primeira diferença da Série A é de cerca de -0,4. A substituição em (6.3.3) produz o par de soluções  $\gamma_1 = 0,5$  e  $\gamma_1 = 2,0$ . No entanto, o valor escolhido  $\gamma_1 = 0,5$  é o único valor que encontra-se dentro do intervalo de invertibilidade  $-1 < \gamma_1 < 1$ . Na verdade, é mostrado na Seção 6.4.1 que é sempre verdade que apenas uma das múltiplas soluções de (6.3.1) pode satisfazer a condição de invertibilidade.

**Exemplos.** As séries A, B e D foram todas identificadas na Tabela 6.2 como possíveis processos IMA de ordem (0, 1, 1). Vimos na Seção 4.3.1 que este modelo pode ser escrito da seguinte forma as formas alternativas:

$$\begin{aligned} \tilde{y} &= (1 - \gamma_1) \\ \tilde{y} &= 0 - \gamma_1 + \gamma_2 \quad (0 = 1 - \gamma_1) \\ &= 0 - \gamma_1 + (1 - 0) - \gamma_1 + \gamma_2 \end{aligned}$$

Usando a Tabela A na Parte Cinco, as estimativas aproximadas dos parâmetros mostrados na Tabela 6.3 foram obtidos.

A Série C foi provisoriamente especificada na Tabela 6.2 como um processo IMA(0, 2, 2):

$$\tilde{y}_2 = (1 - \gamma_1 - \gamma_2)$$

ou equivalente,

$$\tilde{y}_2 = (0 - \gamma_1 + 1) - \gamma_1 + \gamma_2$$

TABELA 6.3 Estimativas iniciais de parâmetros para as séries A, B e D

Series	1	1	= 1y0	1
UMA	-0,41	0,5	0,5	
B	0,09	-0,1	1.1	
D	-0,05	0,1	0,9	

Como as duas primeiras autocorrelações amostrais de  $\hat{y}_2$  são muito próximas de zero, o Gráfico C na Parte Cinco dá  $= 0, = 0_1$  de modo que  $= 1$  e  $= 1$ . Nesta base,  $= 1\hat{y}_1$   $= 1$   $= 2$  2 a série seria representada por

$$\hat{y}_2 = \tag{6.3.4}$$

Isso significaria que a segunda diferença,  $\hat{y}_2$  series. , foi quase um aleatório (ruído branco)

6.3.3 Estimativas Iniciais para Processos Autoregressivos

Para um processo AR assumido de ordem 1 ou 2, as estimativas iniciais para e 2 podem ser calculadas 1 substituindo estimativas para as autocorrelações teóricas nas fórmulas da Tabela 6.1, que são obtidas a partir das equações de Yule-Walker (3.2.6). Em particular, para um AR(1),  $11 = 1$ , e para um AR(2),

$$\begin{aligned} 21 &= \frac{1(1 - \hat{y}_2)}{1 - \hat{y}_1^2} \\ 22 &= \frac{2\hat{y}_2}{1 - \hat{y}_1^2} \end{aligned} \tag{6.3.5}$$

onde o  $\hat{y}_1$  denota o parâmetro autoregressivo estimado em um processo de ordem 1. fórmulas correspondentes dadas pelas equações de Yule-Walker para esquemas de ordem superior podem ser obtido substituindo o  $\hat{y}_1$  pelo  $\hat{y}_k$  em (3.2.7). Desta forma,

$$\hat{y}_k = \hat{y}_1 \tag{6.3.6}$$

onde  $\hat{y}_k$  é uma estimativa das autocorrelações da matriz  $\mathbf{x}$  até a ordem  $\hat{y}_1$ ,  $\mathbf{e} = (1, 2, \dots, k)$ , conforme descrito a seguir (3.2.6) em 3.2.2, de  $\hat{y}_k$ . Por exemplo, se  $k = 3$ , (6.3.6.) torna-se

$$\begin{aligned} \hat{y}_1 &= \hat{y}_1 \\ \hat{y}_2 &= \hat{y}_2 \\ \hat{y}_3 &= \hat{y}_3 \end{aligned} \tag{6.3.7}$$

Um método recursivo simples devido a Levinson e Durbin para obter as estimativas para um AR( ) daqueles de um AR(  $\hat{y}_1$  ) foi discutido no Apêndice A3.2. Será mostrado no Capítulo 7 que, em contraste com a situação dos processos de MA, a parâmetros autorregressivos obtidos de (6.3.6) aproximam o máximo totalmente eficiente estimativas de probabilidade.



**Exemplo.** A série E representando os dados de manchas solares se comporta em sua forma indiferenciada como um processo autoregressivo de segunda ordem:

$$(1 - \alpha_1 B - \alpha_2 B^2) Y_t = \epsilon_t$$

Substituindo as estimativas  $\alpha_1 = 0,81$  e  $\alpha_2 = 0,43$ , obtido usando R, em (6.3.5), temos

12

Como segundo exemplo, considere novamente a Série C identificada como de ordem (1, 1, 0) ou possivelmente (0, 2, 2). A primeira possibilidade daria

$$(1 - \alpha_1 B) \ddot{Y}_t = \epsilon_t$$

com  $\alpha_1 = 0,81$ , pois  $\alpha_1$  para  $\ddot{Y}$  é 0,81.

Este exemplo é interessante porque deixa claro que os dois modelos alternativos que foram identificados para esta série estão intimamente relacionados. Na suposição de que a série é de ordem (0, 2, 2), encontramos em (6.3.4) que isso simplifica para

$$(1 - \ddot{Y})(1 - \ddot{Y}) = \epsilon_t \quad (6.3.8)$$

A alternativa

$$(1 - 0,81 B)(1 - \ddot{Y}) = \epsilon_t \quad (6.3.9)$$

é muito semelhante.

### 6.3.4 Estimativas Iniciais para Processos Autoregressivos Mistos - Processos de Média Móvel

É frequentemente encontrado, inicialmente ou após uma diferenciação adequada, que  $\ddot{Y} = \epsilon_t$  é mais economicamente representado por um processo ARMA misto:

$$(\ddot{Y}) = (\epsilon_t)$$

Conforme observado na Seção 6.2.1, um processo misto é indicado se tanto a autocorrelação quanto a parcial funções de autocorrelação diminuem em vez de ter um recurso de corte. Outra útil fato na identificação do processo misto é que após lag  $\ddot{Y}$  as autocorrelações teóricas do processo misto se comportam como as autocorrelações de um processo autoregressivo puro

$(\ddot{Y}) = (\text{ver (3.4.3)})$ . Em particular, se a função de autocorrelação da  $a$  diferença parece estar caindo exponencialmente de um primeiro valor aberrante 1, suspeitaríamos que temos um processo de ordem (1, , 1) ou seja,

$$(1 - \alpha_1 B) \ddot{Y}_t = (1 - \beta_1 B) \epsilon_t \quad (6.3.10)$$

onde  $\ddot{Y} = \ddot{Y}$ .

A série de manchas solares tem sido objeto de muita investigação. As primeiras referências incluem Schuster (1906), Yule (1927) e Moran (1954). A série não parece ser adequadamente representada por um processo autorregressivo de segunda ordem. Um modelo relacionado ao mecanismo subjacente em funcionamento seria, naturalmente, o mais satisfatório. Trabalhos mais recentes sugeriram empiricamente que um modelo autorregressivo de segunda ordem forneceria um melhor ajuste se uma transformação adequada, como log ou raiz quadrada, foi aplicada pela primeira vez à inclusão de um termo de ordem superior, pelo menos lag 9, no modelo AR também melhora o ajuste. Outras possibilidades incluem o uso de modelos de séries temporais não lineares, como modelos bilineares ou autorregressivos de limiar (por exemplo, consulte a Seção 10.3), como foi investigado por Subba Rao e Gabr (1984), Tong e Lim (1980) e Tong (1983,1990).

Valores aproximados para os parâmetros do processo (6.3.10) são obtidos por substituição na expressão ajustando as estimativas  $\hat{\gamma}_1$  e  $\hat{\gamma}_2$  para

$$\begin{aligned} \hat{\gamma}_1 &= \frac{(1 - \hat{\gamma}_1)(1 - \hat{\gamma}_2)}{1 + \hat{\gamma}_1^2} \\ \hat{\gamma}_2 &= \hat{\gamma}_1 \end{aligned} \quad (3.4.8)$$

O Quadro D na Parte Cinco relaciona  $\hat{\gamma}_1$  e  $\hat{\gamma}_2$  para  $\hat{\gamma}_1$  e  $\hat{\gamma}_2$  pode ser usado para fornecer estimativas iniciais parâmetros para qualquer processo  $(1, \gamma, 1)$ .

Por exemplo, usando a Figura 6.2, a Série A foi identificada como de ordem  $(0, 1, 1)$ , com  $\hat{\gamma}_1 \approx 0,5$ . Olhando para a função de autocorrelação de  $\hat{\gamma}_1$  em vez de  $\hat{\gamma}_2$ , Nós vemos  $\hat{\gamma}_1$  em diante, as autocorrelações decaem aproximadamente exponencialmente, embora lentamente. Desta forma, uma especificação alternativa para a Série A é que ela é gerada por um processo estacionário de ordem  $(1, 0, 1)$ . As autocorrelações estimadas e as estimativas iniciais correspondentes dos parâmetros são então

$$\hat{\gamma}_1 = 0,57 \quad \hat{\gamma}_2 = 0,50 \quad \hat{\gamma}_1 = 0,87 \quad \hat{\gamma}_2 = 0,48$$

Essa identificação produz o modelo aproximado de ordem  $(1, 0, 1)$ :

$$(1 - \hat{\gamma}_1 B) \hat{\gamma}_1 = (1 - \hat{\gamma}_2 B)$$

Considerando que o modelo de ordem previamente identificado  $(0, 1, 1)$ , dado na Tabela 6.5, é

$$(1 - \hat{\gamma}_1 B) = (1 - \hat{\gamma}_2 B)$$

Novamente vemos que os modelos "alternativos" são quase os mesmos.

**Compensação entre Operadores Autoregressivos e de Média Móvel.** A alternativa dos modelos identificados acima são ainda mais parecidos do que parecem. Isso porque pequenos mudanças no operador autorregressivo de um modelo misto podem ser quase compensadas por alterações correspondentes no operador de média móvel. Em particular, se tivermos um modelo

$$[1 - \hat{\gamma}_1 (1 - \hat{\gamma}_2 B)] \hat{\gamma}_1 = (1 - \hat{\gamma}_2 B)$$

onde  $\hat{\gamma}_1$  é pequeno e positivo, podemos escrever

$$\begin{aligned} &= [1 - \hat{\gamma}_1 (1 - \hat{\gamma}_2 B)] \hat{\gamma}_1 (1 - \hat{\gamma}_2 B) (1 - \hat{\gamma}_2 B) \\ &= \{1 - \hat{\gamma}_1 [1 + (1 - \hat{\gamma}_2 B) + (1 - \hat{\gamma}_2 B)^2 + \dots]\} \hat{\gamma}_1 (1 - \hat{\gamma}_2 B) \\ &= [1 - \hat{\gamma}_1 (1 + \hat{\gamma}_2 B)] + \text{termos em } \hat{\gamma}_2^2, \hat{\gamma}_2^3, \dots, \text{ de ordem} \end{aligned}$$

### 6.3.5 Estimativa Inicial da Variação do Erro

Para comparação com os métodos mais eficientes de estimativa a serem descritos no Capítulo 7, é interessante ver quanta informação adicional sobre o modelo pode ser extraída em a fase de identificação. Já mostramos como obter estimativas iniciais  $(\hat{\gamma}_1, \hat{\gamma}_2)$  dos parâmetros  $(\gamma, \gamma)$  no modelo ARMA, identificados para uma diferença apropriada  $= \hat{\gamma}$  da série. Nesta seção mostramos como obter estimativas preliminares do erro variância  $\hat{\sigma}^2$ , e na Seção 6.3.6 mostramos como obter um erro padrão aproximado para a média amostral da série apropriadamente diferenciada.

Uma estimativa inicial da variância do erro pode ser obtida substituindo uma estimativa na expressão  $\hat{\sigma}_0^2$  pela estimativa inicial da variância de  $\hat{\sigma}_0^2$  dado no Capítulo 3. Assim, substituindo em (3.2.8), um  $\hat{\sigma}_0^2$  para um processo AR pode ser obtido de

$$\hat{\sigma}_0^2 = 0(1 - \hat{\phi}_1^2 - \hat{\phi}_2^2 - \dots) \quad (6.3.11)$$

Da mesma forma, de (3.3.3), uma estimativa inicial para um processo MA pode ser obtida de

$$\hat{\sigma}_0^2 = \frac{\hat{\sigma}_0^2}{1 + \hat{\theta}_1^2 + 1} \quad (6.3.12)$$

A forma da estimativa para um processo misto é, em geral, mais complicada. No entanto, para o importante processo ARMA(1,1), ele assume a forma (ver (3.4.7))

$$\hat{\sigma}_0^2 = \frac{1 - \hat{\phi}_1^2}{1 + \hat{\theta}_1^2} \hat{\sigma}_0^2 \quad (6.3.13)$$

Por exemplo, considere o modelo (1, 0, 1) identificado para a Série A. Usando (6.3.13) com  $\hat{\phi}_1 = 0,87$ ,  $\hat{\theta}_1 = 0,48$  e  $\hat{\sigma}_0^2 = 0,098$ , obtemos a estimativa  $\hat{\sigma}_0^2 = 0,1586$ .

### 6.3.6 Erro Padrão Aproximado para

O modelo ARIMA geral, para o qual a média de  $\hat{\sigma}_0^2$  não é necessariamente zero, pode ser escrito em qualquer uma das três formas:

$$\hat{\sigma}_0^2 = \hat{\sigma}_0^2 \quad (6.3.14)$$

$$\hat{\sigma}_0^2 = \hat{\sigma}_0^2 + \hat{\sigma}_0^2 \quad (6.3.15)$$

$$\hat{\sigma}_0^2 = \hat{\sigma}_0^2 + \hat{\sigma}_0^2 \quad (6.3.16)$$

Onde

$$\hat{\sigma}_0^2 = \frac{\hat{\sigma}_0^2}{1 - \hat{\phi}_1^2 - \hat{\phi}_2^2 - \dots} = \frac{(1 - \hat{\phi}_1^2 - \hat{\phi}_2^2 - \dots) \hat{\sigma}_0^2}{1 - \hat{\phi}_1^2 - \hat{\phi}_2^2 - \dots}$$

Portanto, se  $1 - \hat{\phi}_1^2 - \hat{\phi}_2^2 - \dots \neq 0$  e  $1 - \hat{\phi}_1^2 - \hat{\phi}_2^2 - \dots \neq 0$ ,  $\hat{\sigma}_0^2 = 0$  e  $\hat{\sigma}_0^2 = 0$ . Agora, em geral, quando  $\hat{\sigma}_0^2 = 0$ ,  $\hat{\sigma}_0^2$  não será nulo. No entanto, considere a função de previsão eventual associada ao modelo geral (6.3.14) quando  $\hat{\sigma}_0^2 > 0$ .

Com  $\hat{\sigma}_0^2 = 0$ , esta função de previsão já contém um componente polinomial adaptativo de grau  $\hat{\sigma}_0^2 - 1$ . O efeito de permitir ser diferente de zero é introduzir um polinômio *fixo* termo nesta função de grau.

Por exemplo, se  $\hat{\sigma}_0^2 = 2$  e for diferente de zero, a previsão função  $\hat{\sigma}_0^2(\cdot)$  inclui um componente quadrático em que o coeficiente da quadrática prazo é fixo e não se adapta à série. Porque modelos deste tipo são frequentemente inaplicável quando  $\hat{\sigma}_0^2 > 0$ , a hipótese de que  $\hat{\sigma}_0^2 = 0$  frequentemente não será contrariada por os dados. De fato, como indicamos, geralmente assumimos que  $\hat{\sigma}_0^2 = 0$ , a menos que haja evidência para o contrário se apresenta.

Nesta fase, a fase de identificação da construção do modelo, uma indicação de se um valor diferente de zero para  $\hat{\sigma}_0^2$  é necessário pode ser obtido por comparação de  $\hat{\sigma}_0^2$  com  $\hat{\sigma}_0^2$ .

seu erro padrão aproximado (consulte a Seção 2.1.5). Com  $\gamma = 0$  - diferenças disponíveis,

$$\sigma^2(\gamma) = \frac{\gamma^2}{1 - \gamma^2} = \frac{\gamma^2}{1 - \gamma^2}$$

isso é,

$$\sigma^2(\gamma) = \frac{\gamma^2}{1 - \gamma^2} \quad (1) \quad (6.3.17)$$

onde  $\gamma$  é a função geradora de autocovariância definida em (3.1.10) e  $\gamma(1)$  é seu valor quando  $\gamma = 1$  é substituído.

Para ilustração, considere o processo de pedido  $(1, \gamma, 0)$ :

$$(1 - \gamma)(\gamma) =$$

com  $\gamma = 0$ . De (3.1.11), obtemos

$$\gamma(\gamma) = \frac{\gamma^2}{(1 - \gamma)(1 + \gamma)}$$

assim

$$\sigma^2(\gamma) = \frac{\gamma^2}{(1 - \gamma)^2}$$

Mas  $\gamma^2 = \gamma^2(1 - \gamma)$ , então

$$\sigma^2(\gamma) = \frac{\gamma^2}{(1 - \gamma)^2} = \frac{\gamma^2}{1 - \gamma}$$

e

$$\gamma(\gamma) = [1 + \frac{\gamma^2}{(1 - \gamma)^2}]^{-1}$$

Agora  $\gamma$  e  $\sigma^2$  são estimados por 1 e 0, respectivamente, conforme definido em (2.1.11) e (2.1.12). Assim, para um processo  $(1, \gamma, 0)$ , o erro padrão requerido é dado por

$$\sigma(\gamma) = [0(1 + \frac{\gamma^2}{(1 - \gamma)^2})]^{-1/2}$$

Procedendo desta forma, as expressões para  $\sigma^2(\gamma)$  fornecidos na Tabela 6.4 podem ser obtidos.

**Identificação provisória dos modelos A-F.** A Tabela 6.5 resume os modelos provisoriamente identificadas para as Séries A a F, com as estimativas preliminares dos parâmetros inseridas. Esses os valores dos parâmetros são usados como estimativas iniciais para os métodos de estimativa mais eficientes a serem descrito no Capítulo 7.

### 6.3.7 Escolha entre modelos estacionários e não estacionários em casos duvidosos

Como sugerem os resultados das Tabelas 6.2 e 6.5, a identificação preliminar da necessidade para a diferenciação e do grau de diferenciação nem sempre é facilmente determinado. o aparente ambiguidade na identificação de modelos para as Séries A, C e D (particularmente no que diz respeito

TABELA 6.4 Erro padrão aproximado para o  $\hat{\theta}$ , onde  $\epsilon$  é um processo ARIMA pedido  $(p, d, q)$

$(1, , 0)$	$(0, , 1)$
$[ 0 (1 + \hat{\theta}_1) ] 1 \hat{\theta}_2$	$[ 0 (1 + 2 \hat{\theta}_1) ] 1 \hat{\theta}_2$
$(2, , 0)$	$(0, , 2)$
$\frac{2}{1 + 2}$	$\frac{1}{+ 2 \hat{\theta}_2}$
$1) (1 \hat{\theta}_1 \hat{\theta}_2) (1 \hat{\theta}_2) [ 1 \hat{\theta}_2 [ 0 (1 +$	$[ 0 (1 + 2 \hat{\theta}_1 + 2 \hat{\theta}_2) ] 1 \hat{\theta}_2$
$(1, , 1)$	
$\frac{2}{1 + 2}$	
$[ 0 (1 + \hat{\theta}_1 \hat{\theta}_2) ] 1 \hat{\theta}_2$	

ao grau de diferenciação) é, é claro, mais aparente do que real. Surge sempre que o raízes de  $(\lambda) = 0$  aproximam-se da unidade. Quando isso acontece, torna-se cada vez menos importante se uma raiz próxima da unidade está incluída em  $(\lambda)$  ou uma diferença adicional está incluída correspondente a uma raiz unitária. Uma avaliação mais precisa é possível usando a estimativa procedimentos discutidos no Capítulo 7 e, em particular, o teste de raiz unitária mais formal procedimentos a serem discutidos no Capítulo 10. No entanto, deve-se ter em mente o seguinte:

1. A partir de séries temporais que são necessariamente de comprimento finito, nunca é possível *provar* que um zero do operador autorregressivo é exatamente igual à unidade.
2. É claro que não há transição repentina de comportamento estacionário para não estacionário comportamento. Isso pode ser entendido considerando o comportamento da mistura simples

TABELA 6.5 Resumo dos Modelos Identificados para a Série A--F, com as Estimativas Iniciais Inseridas

Series	Diferenciando	$\hat{\theta} \pm \hat{\theta}(\lambda)$	$\hat{\theta}_2 = 0$	Modelo Identificado	$\hat{\theta}_2$
UMA	Ou 0	17,06 ± 0,10	0,1586	- 0,87 $\hat{\theta}_1 = 2,45$ + $\hat{\theta}_1 0,48$	0,098
	ou 1	0,002 ± 0,011	0,1364 $\hat{\theta} = \hat{\theta}_1 0,53$	$\hat{\theta}_1$	0,107
B	1	$\hat{\theta}_1 0,28 \pm 0,41$	52,54	$\hat{\theta} = + 0,09$ $\hat{\theta}_1$	52.2
C	Ou 1	$\hat{\theta}_1 0,035 \pm 0,047$	0,0532 $\hat{\theta} \hat{\theta}_1 0,81 \hat{\theta}_2 = \hat{\theta}_1 0,09$	$\hat{\theta}_1 =$	0,019
	ou 2	$\hat{\theta}_1 0,003 \pm 0,008$	0,0198	$\hat{\theta}_1$	0,020
D	Ou 0	9,13 ± 0,04	0,3620 $\hat{\theta} 0,86 = 1,32 + 0,09 \hat{\theta}_1 \hat{\theta}_2 = \hat{\theta}_1 0,05$	$\hat{\theta}_1$	0,093
	ou 1	0,004 ± 0,017	1382,2 $\hat{\theta}_1 1,32 + 0,63 = 14,9 + \hat{\theta}_1 1,37$	$\hat{\theta}_1$	0,096
E	Ou 0	46,9 ± 5,4	$\hat{\theta}_1 0,08 + 0,32 = 58,3 +$	$\hat{\theta}_1$ $\hat{\theta}_2$	289,0
	ou 0	46,9 ± 5,4	1382,2	$\hat{\theta}_1 + 0,74$ $\hat{\theta}_2$	287,0
F	0	51,1 ± 1,1	139,80	$\hat{\theta}_3 = 13,7 +$	$\hat{\theta}_2$ 115,0
				$\hat{\theta}_1 \hat{\theta}_2 0,18$	

Quando  $\lambda = 0$ , leia para  $\lambda = 1$ .

modelo

$$(1 - \alpha_1)(\ddot{y}) = (1 - \alpha_1)\ddot{y}$$

As séries geradas por tal modelo se comportam de maneira mais não estacionária à medida que  $\alpha_1$  aumenta em direção à unidade. Por exemplo, uma série com  $\alpha_1 = 0,99$  pode desviar-se de sua média e não retornar por períodos muito longos. É como se a atração que a média exerce na série torna-se cada vez menor à medida que  $\alpha_1$  se aproxima da unidade e, finalmente, quando é igual à unidade, o comportamento da série é completamente

independente de  $\ddot{y}$ . Em casos duvidosos, pode haver vantagem em empregar o modelo não estacionário em vez da alternativa estacionária (por exemplo, ao tratar um  $\alpha_1$ , cuja estimativa é próxima da unidade, como sendo *igual* à unidade). Isso é particularmente verdadeiro em problemas de previsão e controle. Onde  $\alpha_1$  está próximo da unidade, não sabemos realmente se a média da série tem significado ou não. Portanto, pode ser vantajoso empregar o modelo não estacionário, que não inclui uma média fixa. Se usarmos tal modelo, as previsões de comportamento futuro não dependem de uma média estimada, calculada a partir de um período anterior, que pode não ter relevância para o nível futuro da série.

## 6.4 MULTIPLICIDADE DE MODELO

### 6.4.1 Multiplicidade de Autoregressivos - Modelos de Média Móvel

Com a suposição de distribuição normal, o conhecimento do primeiro e segundo momentos de uma distribuição de probabilidade implica o conhecimento completo da distribuição. Em particular, o conhecimento da média  $\ddot{y}$  e de sua função de autocovariância exclusivamente determina a estrutura de probabilidade ou  $\gamma$ . Agora mostramos que, embora essa estrutura de probabilidade única possa ser representada por uma *multiplicidade* de modelos ARMA lineares, a unicidade é alcançada no modelo quando introduzimos a estacionariedade e a invertibilidade apropriadas restrições.

Suponha que  $\ddot{y}$ , tendo função geradora de autocovariância  $\gamma$ , é representada pela o modelo linear ARMA

$$(\ddot{y}) = (\gamma) \quad (6.4.1)$$

onde os zeros de  $(\ddot{y})$  e de  $(\gamma)$  estão fora do círculo unitário. Então, este modelo também pode ser escrito como

$$\ddot{y} = \sum_{i=1}^n (1 - \alpha_i) \ddot{y} \quad (6.4.2)$$

onde está  $\alpha_i$  são as raízes de  $(\ddot{y})=0$  e  $\alpha_i$  são as raízes de  $(\gamma)=0$ , e  $\alpha_i$  dentro do círculo unitário. Usando (3.1.11), a função geradora de autocovariância para é

$$(\gamma) = \ddot{y} \sum_{i=1}^n (1 - \alpha_i) \ddot{y} (1 - \alpha_i) \ddot{y} \sum_{i=1}^n (1 - \alpha_i) \ddot{y} (1 - \alpha_i) \ddot{y}^2$$

**Múltipla escolha de parâmetros de média móvel.** Desde

$$(1 - \lambda)(1 - \lambda^*) = (1 - \lambda^*)^2 (1 - \lambda)$$

segue que qualquer um dos modelos estocásticos

$$\sum_{j=1}^p \lambda_j (1 - \lambda_j) = \sum_{j=1}^p (1 - \lambda_j^*)^2$$

pode ter a mesma função geradora de autocovariância se a constante for escolhida apropriadamente. No acima, entende-se que para raízes complexas, recíprocos de ambos os membros de o par conjugado será tomado (de modo a obter sempre coeficientes *de valor real* no MA operador). No entanto, se uma raiz real estiver dentro do círculo unitário, ficará fora, ou se um par complexo, digamos 1 e 2, estão dentro, então o par e ficará do lado de fora. Isto segue que haverá apenas *um modelo invertível estacionário* da forma (6.4.2), que tem uma determinada função de autocovariância.

**Representações para trás.** Agora ( ) também permanece inalterado se em (6.4.2) substituirmos  $1 - \lambda$  por  $1 - \lambda^*$  ou  $1 - \lambda$ . Assim, todos os modelos estocásticos

$$\sum_{j=1}^p \lambda_j (1 - \lambda_j^*) = \sum_{j=1}^p (1 - \lambda_j^*)^2$$

têm estrutura de autocovariância idêntica. No entanto, representações contendo o operador  $\lambda^*$  = referem-se a futuros e/ou futuros, de modo que, embora estacionários e inversíveis existem representações nas quais é expandida em termos de 's e 's futuros, apenas uma dessas representações, (6.4.2), existe que se relaciona inteiramente com a história *passada*.

Uma forma modelo que, surpreendentemente, é de interesse prático é aquela na qual *todos* 's são substituídos por 's em (6.4.1), de modo que

$$( ) = ( )$$

onde é uma sequência de variáveis aleatórias distribuídas independentemente com média zero e variância 1. Está então é uma representação estacionária inversível na qual é expresso *inteiramente* em termos de 's e 's futuros. Nós nos referimos a ela como a forma *inversa* de o processo, ou mais simplesmente como o *processo inverso*.

A Equação (6.4.2) não é a forma mais geral de um ARMA linear inversível estacionário modelo com a função geradora de autocovariância ( ). Por exemplo, o modelo (6.4.2) pode ser multiplicado em ambos os lados por qualquer fator  $1 - \lambda$ . Assim, o processo

$$(1 - \lambda) \sum_{j=1}^p \lambda_j (1 - \lambda_j) = (1 - \lambda) \sum_{j=1}^p (1 - \lambda_j^*)^2$$

tem a mesma estrutura de autocovariância que (6.4.2). Este fato não apresentará nenhuma dificuldade particular na fase de identificação, pois seremos naturalmente levados a escolher o mais simples representação e, portanto, para exclusividade, exigimos que não *haja fatores comuns* entre os operadores AR e MA no modelo. No entanto, como discutido no Capítulo 7, precisamos estar alerta para a possibilidade de fatores comuns nos operadores estimados de AR e MA quando adequando ao processo.

Por fim, chegamos à conclusão de que um modelo estacionário-invertível, no qual um valor atual é expresso apenas em termos de história *anterior* e que contém *nenhum fator comum* entre os operadores AR e MA, é determinado exclusivamente pela estrutura de autocovariância.

A compreensão adequada da multiplicidade de modelos é importante por várias razões:

1. Estamos tranquilizados pelo argumento anterior de que a função de autocovariância pode logicamente ser usado para identificar um modelo ARMA linear estacionário-invertível que expressa em termos de história anterior.
2. A natureza das múltiplas soluções para parâmetros de média móvel obtidos por equacionar momentos é esclarecida.
3. O processo de retrocesso

$$\hat{\gamma}(k) = \gamma(k)$$

obtido substituindo por no modelo ARMA linear, é útil para estimar valores da série que ocorreram antes da primeira observação ser feita.

Agora consideramos as razões 2 e 3 com mais detalhes.

#### 6.4.2 Soluções de Momento Múltiplo para Parâmetros de Média Móvel

Ao estimar os parâmetros  $\theta_1, \theta_2, \dots$ , no modelo MA, equacionando autocovariâncias, vimos que várias soluções são obtidas. Para cada combinação de raízes, haverá uma representação linear correspondente, mas para apenas uma dessas combinações haverá uma representação inversível em termos de história passada.

Por exemplo, considere o processo MA(1) em:

$$y_t = (1 - \theta_1) y_{t-1} + \epsilon_t$$

e suponha que  $\gamma(0)$  e  $\gamma(1)$  são conhecidos e queremos deduzir os valores de  $\theta_1$  e  $\theta_2$ . Desde

$$\gamma_0 = (1 + \theta_1^2) \sigma_{\epsilon}^2, \quad \gamma_1 = -\theta_1 \sigma_{\epsilon}^2, \quad \gamma_2 = 0, \quad \gamma_3 = 0, \quad \dots > 1 \quad (6.4.3)$$

então

$$\frac{\gamma_1}{\gamma_0} = \frac{-\theta_1 \sigma_{\epsilon}^2}{(1 + \theta_1^2) \sigma_{\epsilon}^2} = -\frac{\theta_1}{1 + \theta_1^2}$$

e se  $\left( \theta_1 = \frac{-\gamma_1}{\gamma_0}, \theta_2 = 2 \right)$  é uma solução para dado  $0 < \theta_1 < 1$ , assim é  $\theta_1 = -1, 2$  e  $\theta_2 = 2, 2$ .  
Aparentemente, então, para valores dados de  $\gamma_0$  e  $\gamma_1$ , (há um *par* de modelos possíveis:

$$y_t = (1 - \theta_1) y_{t-1} + \epsilon_t$$

e

$$y_t = (1 - \theta_1) y_{t-1} + \epsilon_t \quad (6.4.4)$$



com  $\text{var}[\tilde{y}] = \sigma^2$  e  $\text{var}[\tilde{y}^2] = 2\sigma^2$  representação. Se  $\gamma_1 < 1$ , então (6.4.4) não é inversível. No entanto, este modelo pode ser escrito como

$$= [(1 - \gamma_1)\tilde{y} + \gamma_1\tilde{y}^2](\tilde{y} + \gamma_1\tilde{y}^2)$$

Assim, após definir  $\tilde{y} = \tilde{y} + \gamma_1\tilde{y}^2$ , o modelo se torna

$$= (1 - \gamma_1)\tilde{y} \quad (6.4.5)$$

onde tem a mesma variância que  $\tilde{y}$ . Assim, (6.4.5) é simplesmente o processo "para trás", que é dual para o processo forward:

$$= (1 - \gamma_1)\tilde{y} \quad (6.4.6)$$

Assim como o choque em (6.4.6) é expresso como uma soma convergente de valores de  $\tilde{y}$ ,

$$= \tilde{y}_1 + \gamma_1\tilde{y}_2 + \gamma_1^2\tilde{y}_3 + \dots$$

o choque em (6.4.5) é expresso como uma soma convergente dos valores atuais e futuros de:

$$= \tilde{y}_t + \gamma_1\tilde{y}_{t+1} + \gamma_1^2\tilde{y}_{t+2} + \dots$$

Assim, a raiz  $\gamma_1$  produziria um processo "invertível", mas somente se uma representação do choque em termos de valores futuros de  $\tilde{y}$  eram permitidos. As regiões de invertibilidade mostrados na Tabela 6.1 delimitam os valores aceitáveis dos parâmetros, *dado* que expressamos o choque em termos de história *anterior*.

### 6.4.3 Uso do processo retroativo para determinar os valores iniciais

Suponha que uma série temporal  $y_1, y_2, \dots$  está disponível a partir de um processo

$$y_t = (1 - \gamma_1)y_{t-1} + \gamma_1 y_{t+1} + \epsilon_t \quad (6.4.7)$$

No Capítulo 7, surgem problemas onde precisamos estimar os valores  $y_1, y_2, \dots$ , da série que ocorreu *antes* da primeira observação ser feita. Isso acontece porque "valores iniciais" são necessários para certos cálculos recursivos básicos usados para estimar o parâmetros no modelo. Agora, suponha que precisamos estimar  $y_1$ , dado  $y_2, \dots$ . A discussão da Seção 6.4.1 mostra que a estrutura de probabilidade de  $y_1$  é igualmente explicado pelo modelo forward (6.4.7), ou pelo modelo backward

$$y_t = (1 - \gamma_1)y_{t+1} + \gamma_1 y_{t-1} + \epsilon_t \quad (6.4.8)$$

O valor que  $y_1$ , portanto, tem exatamente a mesma relação de probabilidade com o de valores  $y_2, y_3, \dots$ , assim como a sequência  $y_{t+1}, y_{t+2}, \dots$  para a sequência  $y_t, y_{t+1}, \dots$ . Assim, para estimar um valor  $+1$  períodos antes do início das observações, podemos primeiro considere qual seria a estimativa ou previsão ótima  $+1$  períodos após a série

terminou e, em seguida, aplique este procedimento à série *invertida*. Em outras palavras, nós "prevemos" a série invertida. Chamamos isso de "previsão de retorno".

#### APÊNDICE A6.1 COMPORTAMENTO ESPERADO DO ESTIMADO FUNÇÃO DE AUTOCORRELAÇÃO PARA UM PROCESSO NÃO ESTACIONÁRIO

Suponha que uma série de observações  $y_1, y_2, \dots, y_n$  é gerado por um não estacionário  $(0, 1, 1)$  processe

$$\tilde{y} = (1 - \alpha) \tilde{y}$$

e as autocorrelações estimadas são computadas, onde

$$r_0 = \frac{\sum_{t=1}^n \tilde{y}_t (\tilde{y}_t - \bar{\tilde{y}})}{\sum_{t=1}^n (\tilde{y}_t - \bar{\tilde{y}})^2}$$

Alguma idéia do comportamento dessas autocorrelações estimadas pode ser obtida derivando valores esperados para o numerador e denominador desta expressão e considerando o Razão. Escreveremos, seguindo Wichern (1973),

$$E[r_0] = \frac{E[\sum_{t=1}^n \tilde{y}_t (\tilde{y}_t - \bar{\tilde{y}})]}{E[\sum_{t=1}^n (\tilde{y}_t - \bar{\tilde{y}})^2]}$$

Depois de uma álgebra simples, mas tediosa, descobrimos que

$$E[r_0] = \frac{(1 - \alpha)^2 (1 + \alpha) (1 - \alpha^2)}{1 - \alpha^2} \quad (A6.1.1)$$

Para perto de zero,  $E[r_0]$  será próximo da unidade, mas para valores grandes pode ser consideravelmente menor que a unidade, mesmo para valores pequenos de  $\alpha$ . A Figura A6.1 ilustra isso fato mostrando valores de  $E[r_0]$  para  $\alpha = 0,8$  com  $n = 100$  e  $n = 200$ . Embora, como antecipado para um processo não estacionário, as razões  $E[r_0]$  dos valores esperados não conseguem amortecer rapidamente, será visto que eles não se aproximam do valor 1 mesmo para pequenas defasagens.

Efeitos semelhantes podem ser demonstrados sempre que os parâmetros se aproximam de valores onde o cancelamento em ambos os lados do modelo produziria um processo estacionário. Por exemplo, no exemplo acima podemos escrever o modelo como

$$(1 - \alpha) \tilde{y} = [(1 - \alpha) \tilde{y}] + \alpha \tilde{y}$$

onde  $\alpha = 0,2$ . À medida que  $\alpha$  tende a zero, espera-se que o comportamento do processo venha cada vez mais próximo do processo de ruído branco  $\tilde{y} = a$ , para o qual a autocorrelação função é zero para atrasos  $> 0$ .

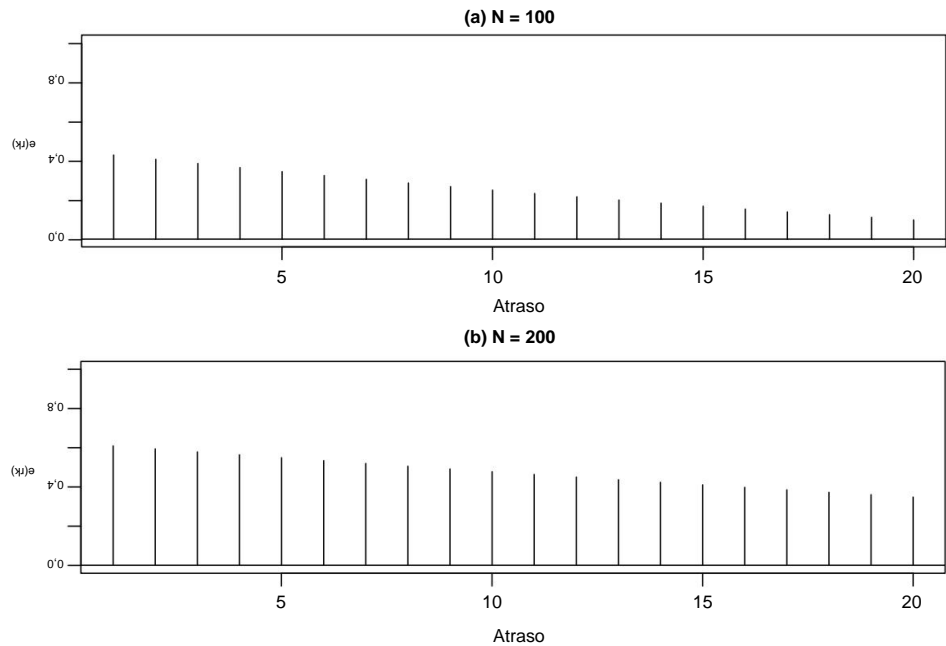


FIGURA A6.1  $E[\hat{\gamma}_k] = [\hat{\gamma}_k | 0]$  para séries geradas por  $\tilde{y} = (1 \ \tilde{y} \ 0,8)$

EXERCÍCIOS

6.1. Dados os cinco modelos identificados e os valores correspondentes do auto estimado correlações de  $\tilde{y}$  na tabela a seguir:

Modelo Identificado				Autocorrelações estimadas
(1)	1	1	0	$\hat{\gamma}_1 = 0,72$
(2)	0	1	1	$\hat{\gamma}_1 = -0,41$
(3)	1	0	1	$\hat{\gamma}_1 = 0,40, \hat{\gamma}_2 = 0,32$
(4)	0	2	2	$\hat{\gamma}_1 = 0,62, \hat{\gamma}_2 = 0,13$
(5)	2	1	0	$\hat{\gamma}_1 = 0,93, \hat{\gamma}_2 = 0,81$

- (a) Obtenha estimativas preliminares dos parâmetros analiticamente.
- (b) Verifique essas estimativas usando os gráficos e tabelas da Parte Cinco do livro.
- (c) Escreva os modelos identificados na notação do operador de deslocamento para trás com o estimativas preliminares inseridas.
- 6.2. Para o processo  $(2, 1, 0)$  considerado na linha (5) do Exercício 6.1, a média amostral e variância de  $\tilde{y}$  são  $= 0,23$  e  $= 0,25$ . Se a série contém  $n = 101$  observações,
- (a) mostrar que um termo constante precisa ser incluído no modelo,

**(b)** expresse o modelo na forma  $\hat{y}_t = \mu + \delta_1 t + \delta_2 t^2 + \epsilon_t$  com 0 numérico inseridos para os parâmetros, incluindo uma estimativa de 2.

**6.3.** Considere as leituras de temperatura do processo químico referidas como Série C neste livro.

**(a)** Plote a série original e a série de primeiras diferenças usando R. **(b)** Use o pacote R para calcular e plotar o ACF e o PACF desta série. Repetir o cálculo para a primeira e segunda diferenças da série.

**(c)** Especifique um modelo ou modelos adequados para esta série. Use o método dos momentos para obter estimativas preliminares de parâmetros para a série.

**6.4.** Medições trimestrais do produto interno bruto (PIB) no Reino Unido durante o período de 1955 a 1969 estão incluídos na Série P da Parte Cinco deste livro. **(a)** Calcule e plote o ACF e o PACF desta série. **(b)** Repita a análise da parte (a) para as primeiras diferenças da série. **(c)** Identifique um modelo para a série. Uma transformação de log dos dados seria útil? **(d)** Obtenha estimativas preliminares para os parâmetros e seus erros padrão. **(e)** Obter estimativas preliminares para a série.

**6.5.** A taxa de desemprego trimestral do Reino Unido (em milhares) faz parte da Série P analisada no Exercício 6.4. Repita as partes (a) a (e) do Exercício 6.4 para esta série.

**6.6.** Uma série temporal definida por  $y_t = 1000 \log_{10}(x_t)$ , onde  $x_t$  é o preço dos suínos registrado anualmente pelo Censo de Agricultura dos EUA em 1º de janeiro para cada um dos 82 anos, de 1867 a 1948, é listado como Série Q na Coleção de Séries Temporais na Parte Cinco. Esta é uma série temporal bem conhecida analisada por Quenouille (1957), e outros. **(a)** Trace a série. Calcule e plote o ACF e o PACF da série. **(b)** Identifique um modelo de série temporal para a série.

**6.7.** As medições do fluxo anual do rio Nilo em Ashwan de 1871 a 1970 estão disponíveis como série "Nilo" no pacote de conjuntos de dados em R; digite `help(Nile)` para detalhes. **(a)** Plote a série e calcule o ACF e o PACF para a série. **(b)** Repita a análise da parte (a) para a série diferenciada. **(c)** Identifique um modelo para a série. Existem características incomuns dignas de nota.

**6.8.** O arquivo "EuStockMarkets" no pacote de conjuntos de dados R contém os preços de fechamento diários dos quatro principais índices de ações europeus: Alemanha DAX (Ibis), Suíça SMI, França CAC e UK FTSE. Os dados são amostrados em horário comercial, portanto, fins de semana e feriados são omitidos. **(a)** Plote cada uma das quatro séries e calcule o ACF e o PACF para a série. **(b)** Repita a análise da parte (a) para a série diferenciada. **(c)** Identifique um modelo para a série. Existem características incomuns dignas de nota.

**6.9.** Baixe uma série temporal de sua escolha da Internet. Trace a série temporal e identifique um modelo adequado para a série.

---

# 7

---

## ESTIMATIVA DE PARÂMETRO

Este capítulo trata da estimação dos parâmetros em modelos ARIMA e fornece uma descrição geral da verossimilhança e métodos Bayesianos para estimação de parâmetros. Assume-se que um modelo adequado desta forma foi selecionado usando as ferramentas de especificação de modelo descritas no Capítulo 6. Após os parâmetros terem sido estimados, o modelo ajustado será submetido a verificações de diagnóstico e testes de ajuste a serem descritos em o próximo capítulo.

Conforme apontado por RA Fisher, para que os testes de bondade de ajuste sejam relevantes, é necessário que o uso eficiente dos dados tenha sido feito no processo de ajuste. Se não for assim, a inadequação do ajuste pode simplesmente surgir devido ao ajuste ineficiente e não porque a forma do modelo é inadequada. Este capítulo examina em detalhes a estimativa de máxima verossimilhança sob a suposição de normalidade e descreve aproximações de mínimos quadrados que são adequadas para muitas séries.

Supõe-se que o leitor esteja familiarizado com certas ideias básicas da teoria da estimativa. Os apêndices A7.1 e A7.2 resumem alguns resultados importantes da teoria da distribuição normal e dos mínimos quadrados lineares que são úteis para este capítulo. Ao longo do capítulo, o tipo negrito é usado para denotar vetores e matrizes. Assim,  $\mathbf{X} = \{ \}$  é uma matriz com  $n$  linhas e  $p$  colunas, e  $\mathbf{y}$  é um vetor com  $n$  elementos.

### 7.1 ESTUDO DAS FUNÇÕES DE PROBABILIDADE E SOMA DE QUADRADOS

#### 7.1.1 Função de probabilidade

Suponha que temos uma amostra de  $n$  observações,  $\mathbf{z}$  com as quais associamos uma variável aleatória dimensional, cuja distribuição de probabilidade conhecida  $(\mathbf{z})$  depende de alguns parâmetros desconhecidos. Usamos o vetor para denotar um conjunto geral de parâmetros e, em particular, pode se referir aos parâmetros  $\phi, \theta, \sigma^2$  do modelo ARIMA.

---

*Análise de Séries Temporais: Previsão e Controle*, Quinta Edição. George EP Box, Gwilym M. Jenkins, Gregory C. Reinsel e Greta M. Ljung © 2016 John Wiley & Sons. Inc. Publicado em 2016 por John Wiley & Sons. Inc.

## 210 ESTIMATIVA DE PARÂMETROS

Antes que os dados estejam disponíveis,  $(\mathbf{z})$  associará uma densidade a cada resultado diferente  $\mathbf{z}$  do experimento, para fixo. Após os dados estarem disponíveis, somos levados a contemplar os vários valores de  $\mathbf{z}$  que podem ter dado origem ao conjunto fixo de observações  $\mathbf{z}$  realmente obtido. A função apropriada para este propósito é a *função de verossimilhança*  $(\mathbf{z})$ , que é da mesma forma que  $(\mathbf{z})$ , mas na qual  $\mathbf{z}$  agora é fixo, mas é variável. É apenas o valor relativo de  $(\mathbf{z})$  que interessa, de modo que a função de verossimilhança é geralmente considerado como contendo uma *constante multiplicativa arbitrária*.

Muitas vezes é conveniente trabalhar com a função de probabilidade logarítmica  $\ln[(\mathbf{z})] = (\mathbf{z})$ , que contém uma *constante aditiva arbitrária*. Uma razão pela qual a função de verossimilhança é de importância fundamental na teoria da estimativa é por causa do *princípio da verossimilhança*, instado em bases um tanto diferentes por Fisher (1956), Barnard (1949) e Birnbaum (1962). Este princípio diz que, dado que o modelo assumido está correto, tudo o que os *dados* precisam conte-nos sobre os parâmetros está contido na função de verossimilhança, todos os outros aspectos dos dados irrelevantes. Do ponto de vista bayesiano, a função de verossimilhança é igualmente importante, pois é o componente na distribuição posterior dos parâmetros que vem dos dados.

Para um entendimento completo da estimação de parâmetros em um caso específico, é necessário estudar cuidadosamente a função de verossimilhança, ou no framework bayesiano, a distribuição posterior dos parâmetros, que nos casos que consideramos, é dominada por a probabilidade. Em muitos exemplos, para amostras moderadas e grandes, a função de probabilidade logarítmica será unimodal e pode ser aproximada adequadamente ao longo de uma extensão suficientemente extensa. região próxima ao máximo por uma função quadrática. Nesses casos, a função de probabilidade logarítmica pode ser descrito por seu máximo e suas segundas derivadas no máximo. Os valores dos parâmetros que maximizam a função de verossimilhança, ou equivalentemente o log-verossimilhança função, são chamadas de estimativas de *máxima verossimilhança (ML)*. As segundas derivadas do A função de probabilidade logarítmica fornece medidas de "spread" da função de verossimilhança e pode ser usado para calcular os erros padrão aproximados para as estimativas.

As propriedades limitantes das estimativas de máxima verossimilhança são geralmente estabelecidas para observações independentes. Mas como foi mostrado por Whittle (1953), eles podem ser estendidos para cobrir séries temporais estacionárias. Outra literatura inicial sobre a estimativa de parâmetros no tempo modelos de série inclui Barnard et al. (1962), Bartlett (1955), Durbin (1960), Grenander e Rosenblatt (1957), Hannan (1960) e Quenouille (1942, 1957).

### 7.1.2 Probabilidade Condicional para um Processo ARIMA

Suponhamos que as  $y_t$  observações originais  $\mathbf{z}$  formam uma série temporal que denotar por  $y_{t+1}, \dots, 0, 1, 2, \dots$ . Assumimos que esta série é gerada por um Modelo ARIMA( $p, d, q$ ). A partir dessas observações, podemos gerar uma série  $w_t$  de  $y_t$  diferenças onde  $w_t = y_t - y_{t-1}$ . Assim, o problema geral de ajuste dos parâmetros e do modelo ARIMA (6.1.1) é equivalente ao ajuste aos

modelo ARMA( $p, q$ ), estacionário e invertível<sup>1</sup>, que pode ser escrito como

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_q y_{t-q} + \epsilon_t \quad (7.1.1)$$

onde  $y_t$  são as observações centradas na média.

<sup>1</sup>É necessário um cuidado especial para garantir que a estimativa esteja na região invertível. Consulte o Apêndice A7.7.

Para  $\sigma^2 > 0$ , muitas vezes é apropriado assumir que  $\sigma^2 = 0$ . Quando isso não é apropriado, nós suponhamos que a série significa  $\bar{y} = 1$   $\bar{y}$  é substituído por  $\bar{y}$ . Para muitos tamanhos de amostra comum na prática, esta aproximação será adequada. No entanto, se desejar, pode ser incluído como parâmetro adicional a ser estimado.

Os  $\bar{y}$ 's não podem ser calculados imediatamente a partir de (7.1.1) devido à dificuldade de iniciando a equação diferencial. No entanto, suponhamos que os valores  $w\bar{y}$  de  $\bar{y}$ 's e foram dados os valores  $a\bar{y}$  dos  $\bar{y}$ 's anteriores ao início da série. Então, para qualquer escolha de parâmetros  $(\sigma^2, \bar{y})$ , poderíamos calcular sucessivamente um conjunto de valores  $(\bar{y}_1, w\bar{y}_1, a\bar{y}_1, w) = 1, 2, \dots$ . Agora, supondo que os  $\bar{y}$ 's sejam normalmente distribuídos, sua probabilidade densidade é

$$(\bar{y}_1, 2, \dots, \bar{y}_n) \propto \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 \right]$$

Dados os dados  $w$ , a probabilidade logarítmica associada aos valores dos parâmetros  $(\sigma^2, \bar{y})$ , com *ditional* na escolha de  $(w\bar{y}, a\bar{y})$ , seria então

$$\bar{y}(\sigma^2, \bar{y}) = \bar{y} \ln \left( \frac{1}{2\pi\sigma^2} \right) - \frac{\bar{y}(\sigma^2, \bar{y})}{2\sigma^2} \quad (7.1.2)$$

Onde

$$\bar{y}(\sigma^2, \bar{y}) = \bar{y} \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 \quad (7.1.3)$$

Nas equações acima, um asterisco subscrito é usado na probabilidade e na soma dos quadrados funções para enfatizar que estão condicionadas à escolha dos valores iniciais. Nós observe que a probabilidade logarítmica condicional envolve os dados apenas por meio da probabilidade condicional *função soma de quadrados*. Segue-se que os contornos de  $\bar{y}$  para qualquer valor fixo de  $\sigma^2$  no espaço  $(\sigma^2, \bar{y})$ , que essas *estimativas de máxima verossimilhança* são as estimativas de mínimos quadrados, e que, em geral, podemos, na suposição normal, estudar a comportamento da verossimilhança condicional estudando a função condicional da soma dos quadrados. Em particular para qualquer fixo  $\sigma^2$ ,  $\bar{y}$  é uma função linear de  $\bar{y}$ . Os valores dos parâmetros obtidos minimizando a função de soma de quadrados condicional  $\bar{y}(\sigma^2, \bar{y})$  será chamada de *condicional estimativas de mínimos quadrados*.

### 7.1.3 Escolha de Valores Iniciais para Cálculo Condicional

Discutiremos brevemente o cálculo da verossimilhança incondicional, que, estritamente, é o que precisamos para a estimativa de parâmetros. No entanto, quando é moderado ou grande, uma quantidade suficiente A aproximação da verossimilhança incondicional é muitas vezes obtida usando o condicional probabilidade com valores adequados substituindo os elementos de  $w\bar{y}$  e  $a\bar{y}$  em (7.1.3). Um O procedimento é igualar os elementos de  $w\bar{y}$  e de  $a\bar{y}$  às suas expectativas incondicionais. As expectativas incondicionais dos elementos de  $a\bar{y}$  são zero, e se o modelo não contiver parte determinística e, em particular, se  $\sigma^2 = 0$ , as expectativas incondicionais dos elementos

TABELA 7.1 Funções de soma de quadrados para o modelo = ( ŷ ) Ajustado ao IBM  
Dados

	=1ŷ	ŷ( )	( )		=1ŷ	ŷ( )	( )
ŷ0,5	1,5	23.929	23.928	0,1	0,9	19.896	19.896
ŷ0,4	1,4	21.595	21.595	0,2	0,8	20.851	20.851
ŷ0,3	1,3	20.222	20.222	0,3	0,7	22.315	22.314
-0,2	1,2	19.483	19.483	0,4	0,6	24.471	24.468
ŷ0,1	1,1	19.220	19.220	0,5	0,5	27.694	27.691
0,0	1,0	19.363	19.363				

de wŷ também será zero2. No entanto, esta aproximação pode ser pobre se algumas das raízes de ( )=0 estão próximos ao limite do círculo unitário, de modo que o processo se aproxima não estacionaridade. Isso também é verdade se algumas das raízes de ( )=0 estiverem próximas ao limite da região de invertibilidade. Definir os valores de pré-amostra iguais a zero pode nesses casos introduzir um grande transiente, que é lento para morrer. Para um modelo AR( ) puro, uma procedimento de aproximação, e um que empregamos às vezes, é usar (7.1.1) para calcular o 's de +1 em diante, usando assim os valores reais dos 's por toda parte. Usando este método, temos apenas ŷ = ŷ ŷ valores de mas a ligeira perda.de informação será sem importância para séries longas.

Para séries sazonais, discutidas no Capítulo 9, a aproximação condicional nem sempre é satisfatório e o cálculo incondicional torna-se necessário. A inclusão do determinante na função de verossimilhança incondicional também pode ser importante para o tempo sazonal Series.

**Exemplo: Processo IMA(0, 1, 1).** Para ilustrar o cálculo recursivo da condicional soma dos quadrados consideramos o modelo IMA(0, 1, 1) provisoriamente identificado na Seção 6.4 para os dados da IBM na Série B. O modelo é

$$\ddot{y} = (1 \ddot{y}) \qquad \ddot{y} \ 1 < < 1 \qquad (7.1.4)$$

de modo que = + ŷ1, onde = ŷ e [ ]=0. Assim, para o parâmetro específico value = 0,5, os 's são calculados recursivamente a partir de

$$= + 0,5 \qquad \ddot{y}_1$$

definindo o valor inicial 0 igual a zero. Procedendo desta forma, encontramos que

$$\ddot{y}(0,5) = \ddot{y}^{368}_{=1} \quad ^2 ( = 0,5 \mid 0 = 0) = 27.694$$

As somas condicionais dos quadrados ŷ( ) são mostradas na Tabela 7.1 para valores de -0,5 a +0,5 em passos de 0,1. Notamos que ŷ( ) tem seu mínimo para = ŷ0,1. Isso é consistente com a estimativa de momento preliminar de -0,09 derivada para esta série no Capítulo 6.

2Se a suposição | = ŷ 0 é apropriado, podemos substituir ŷ para cada um dos elementos de wŷ.



### 7.1.4 Probabilidade Incondicional, Função Soma dos Quadrados e Mínimos Quadrados Estimativas

Assumindo que as observações  $y_t$  são geradas por um modelo ARIMA, a probabilidade logarítmica não condicional é dada por

$$L(\theta) = -\frac{1}{2} \ln |C| - \frac{1}{2} (y - \hat{y})' C^{-1} (y - \hat{y}) \quad (7.1.5)$$

onde  $|C|$  envolve o determinante na densidade conjunta dos  $y_t$  e é uma função de  $\theta$ . A função soma de quadrados incondicional é dada por

$$Q(\theta) = (y - \hat{y})' [I - \Phi(L)]' \Sigma^{-1} [I - \Phi(L)] (y - \hat{y}) \quad (7.1.6)$$

onde  $[\Phi(L)] = [\Phi(L)]$  denota a expectativa de condicional em  $w_t$ , e

. Quando o significado é claro a partir do contexto, vamos abreviar ainda mais esta condicional expectativa para  $[\cdot]$ . Em (7.1.6),

$$e\tilde{y} = (y_1, \dots, y_T, 0, 1\tilde{y}, \dots, 0)'$$

representa os valores iniciais  $y_1$  e antes de  $t=1$ , matriz de covariância  $\Sigma = \text{cov}[e\tilde{y}]$  é o de  $e\tilde{y}$ , e

$$[e\tilde{y}] = ([y_1, \dots, 0], [0, 1\tilde{y}, \dots, 0])'$$

denota o vetor de expectativas condicionais ("back-forecasts") dos valores iniciais, dado  $w_t$ , e . Uma forma alternativa de representar  $(y, \hat{y})$  é como

$$(y, \hat{y}) = \tilde{y} \begin{bmatrix} I \\ \Phi(L) \end{bmatrix}$$

que em comparação com (7.1.6) indica que  $\tilde{y}' \tilde{y} = [e\tilde{y}]' \Sigma^{-1} [e\tilde{y}]$ .

Normalmente,  $(y, \hat{y})$  é importante apenas para valores pequenos. Para valores moderados e grandes de  $T$ , (7.1.5) é dominado por  $(y, \hat{y})'$  e, portanto, os contornos do incondicional função de soma de quadrados no espaço dos parâmetros  $(\theta)$  são quase contornos da verossimilhança e da probabilidade logarítmica. Segue-se, em particular, que as estimativas de parâmetros obtido minimizando a soma dos quadrados (7.1.6), que chamamos de (*incondicional* ou *exata*) *estimativas de mínimos quadrados*, geralmente fornecerão aproximações muito próximas ao máximo estimativas de probabilidade. Do ponto de vista bayesiano, em suposições discutidas na Seção 7.5, para todos AR( $p$ ) e MA( $q$ ), essencialmente a densidade posterior é uma função apenas de  $(y, \hat{y})$ . Assim, quase as estimativas de mínimos quadrados são aquelas com densidade posterior máxima. No restante desta seção e na Seção 7.1.5, a ênfase principal será na função soma de quadrados incondicional  $(y, \hat{y})$  em (7.1.6), e seu uso no cálculo de estimativas de mínimos quadrados. Um método alternativo para o cálculo da soma incondicional dos quadrados e funções de verossimilhança baseadas no modelo de espaço de estados e na abordagem de inovações serão discutido na Seção 7.4.

No cálculo da soma incondicional dos quadrados, os  $[\cdot]$ 's são calculados recursivamente tomando expectativas condicionais em (7.1.1). Um cálculo retroativo preliminar fornece a

## 214 ESTIMATIVA DE PARÂMETROS

valores  $\hat{y}_t$  e  $\hat{y}_{t-1}$ ,  $t = 0, 1, 2, \dots$  (ou seja, as previsões anteriores) necessários para iniciar a recursão para frente.

**Cálculo da Soma Incondicional dos Quadrados para um Processo de Média Móvel.** Por

ilustração, reconsideramos o exemplo do preço das ações da IBM usando apenas os primeiros 10 valores de a série.<sup>3</sup> Para o modelo IMA(0, 1, 1), a única previsão de retorno necessária para  $\hat{y}_0$  é

$\hat{y}_0$ . Começamos descrevendo um método *aproximado*, mas preciso, para obter

$\hat{y}_0$ . Lembre-se da Seção 6.4.3 que o modelo para  $\hat{y}_t$  pode ser escrito tanto no ou formas invertidas:

$$\hat{y}_t = (1 - \theta) \hat{y}_{t-1} + \theta \hat{y}_0 \quad \text{ou} \quad \hat{y}_t = (1 - \theta) \hat{y}_{t-1} + \theta \hat{y}_0$$

e onde novamente  $\hat{y}_0$  é assumido igual a zero. Assim, podemos escrever

$$\hat{y}_t = \hat{y}_{t-1} + \theta (\hat{y}_0 - \hat{y}_{t-1}) \quad (7.1.7)$$

$$\hat{y}_t = \hat{y}_{t-1} + \theta (\hat{y}_0 - \hat{y}_{t-1}) \quad (7.1.8)$$

onde  $\hat{y}_t = \hat{y}_{t-1}$  for  $t = 1, 2, \dots$ , e  $\hat{y}_0$  é a previsão de retorno de  $t = 0$ . Estes são as duas equações básicas que precisamos nos cálculos. Um formato conveniente para o cálculos é mostrado na Tabela 7.2. Começamos inserindo na tabela o que sabemos:

1. Os valores de dados  $y_0, y_1, \dots, y_9$ , a partir do qual podemos calcular as primeiras diferenças  $y_1 - y_0, y_2 - y_1, \dots, y_9 - y_8$ .
2. Os valores  $\hat{y}_0, \hat{y}_1, \dots$ , que são zero, pois  $\hat{y}_t = 0$  para  $t = 0, -1, \dots$  são distribuídos independentemente.
3. Os valores  $\hat{y}_1, \hat{y}_2, \dots$ , que são zero, pois para qualquer MA(1)  $\hat{y}_t = 0$  para  $t = 0, -1, \dots$  independentemente de  $w$ . No entanto, note que  $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{t-1}$  será diferente de zero e pode ser obtido por back-forecast. Assim, no presente exemplo,  $\hat{y}_0$  é calculado desta forma.

Começando no final da série, (7.1.7) agora é usado para calcular os  $\hat{y}_t$ 's para  $t = 9, 8, 7, \dots, 1$ . Iniciamos o processo de retrocesso definindo  $\hat{y}_{10} = 0$ . O efeito deste A aproximação será introduzir um transiente no sistema. No entanto, para séries de duração moderada, o efeito normalmente será insignificante no momento em que o início da série é atingido e, portanto, não afetará o cálculo dos  $\hat{y}_t$ 's. Se desejado, a adequação deste a aproximação pode ser verificada em qualquer caso, realizando um segundo ciclo iterativo.

Assim, para iniciar a recursão na Tabela 7.2, na linha correspondente a  $t = 9$ , entramos um zero na sexta coluna para o valor desconhecido  $0.5[\hat{y}_{10}]$ . Então, usando (7.1.7), obtemos

$$\begin{aligned} \hat{y}_9 &= [\hat{y}_9] + 0.5[\hat{y}_{10}] \\ &= y_9 + 0 = -3 \end{aligned}$$

<sup>3</sup>Na prática, é claro, estimativas de parâmetros úteis não podem ser obtidas com apenas 10 observações. Nós utilizar este subconjunto de dados meramente para ilustrar os cálculos.

TABELA 7.2 Cálculo dos [ ]'s dos 10 primeiros valores da série B, usando =

		[ ]	0,5[ -1]	[ ]	0,5[ +1]	[ ]	
$\tilde{y}_1$	[458,4]					0	
0	460	0		0	0		$\tilde{y}_{2,1}$
1	457	1,6	0 0	1,6	$\tilde{y}_{1,6}$	0	$\tilde{y}_{4,1}$
2	452	$\tilde{y}_{2,2} \tilde{y}_{6,1}$	0,8 $\tilde{y}_{1,1}$	$\tilde{y}_{3,0} \tilde{y}_{5,0}$	$\tilde{y}_{0,1} 4,8$	$\tilde{y}_{3,1} \tilde{y}_{0,2}$	$\tilde{y}_{2,3}$
3	459	3,9	-3,0	7,0	2,6	9,6	8,5
4	462	5,0	2,0	3,0	2,3	5,3	9,5
5	459	-0,5	2,5	-3,0	7,6	4,6	9,2
6	463	3,7	-0,2	4,0	11,1	15,1	19,4
7	479	17,9	1,9	16,0	6,2	22,2	31,4
8	493	22,9	9,0	14,0	$\tilde{y}_{1,5}$	12,5	27,5
9	490	8,5	11,5	-3,0	0	-3,0	8,5

então  $0,5[9] = \tilde{y}_{1,5}$  pode ser inserido na linha = 8, o que nos permite calcular [ 8], e em breve. Por fim, obtemos

$$[0] = [0] + [1]$$

isto é,  $0 = [0] \tilde{y}_{1,6}$ , que dá  $[0] = 1,6$ , e depois  $[\tilde{y}\tilde{y}] = 0$ ,  $\tilde{y} = 1, 2, 3, \dots$ .  
Agora, usando (7.1.8) com  $= 0$ , obtemos

$$[0] = [0] + [\tilde{y}_1] = 1,6 + (0,5)(0) = 1,6$$

e podemos então continuar os cálculos para frente dos restantes [ ]'s, levando a  
 $(0,5) = \tilde{y}_9 = 0[0,5, w] = 1016,406$ .

Um método alternativo que produz estimativas exatas dos valores de pré-amostra é apresentado no Apêndice A7.3. Para o modelo considerado acima, este método envolve primeiro calcular os valores ( $0 = 0$ ), que abreviamos como  $\tilde{y}_0$ , pelo método condicional como  $\tilde{y}_0 = + \tilde{y}_1$ ,  $\tilde{y}_1 = 1, 2, \dots$ , usando  $\tilde{y}_0 = 0$  como o valor inicial. Então um retrocesso recursivo é realizado para obter  $=$  com  $+1$ , começando de  $=$  para baixo até  $= 0$ ,  $+1 = 0$  como valor inicial. Finalmente, então, a estimativa exata de  $[0]$  é dada por  $[0] = \tilde{y}_0(1 - \tilde{y}_2)\tilde{y}_1(1 - \tilde{y}_2(+1))$ . Usando este valor inicial, os [ ] são calculados a partir do recursão direta  $[ ] = + [\tilde{y}_1]$ ,  $= 1, 2, \dots$ , como em (7.1.8) e a soma exata de quadrados se torna  $( ) = \tilde{y}_0[ ]^2$ .

No exemplo acima, calculando primeiro o  $\tilde{y}_0$  usando uma configuração de recursão para frente  $0 = 0$ , obtemos os valores de pela recursão para trás para  $= 9, 8, \dots, 0$ , exibido na coluna final da Tabela 7.2. Assim, obtemos a estimativa exata de como  $[0] = \tilde{y}_0 \tilde{y}_0(1 - \tilde{y}_2)\tilde{y}_1(1 - \tilde{y}_2(+1)) = 1,549$ . Este valor está muito próximo do valor aproximado de 1,545 obtido pela abordagem do modelo para trás, e a pequena diferença tem essencialmente nenhum efeito no cálculo dos valores restantes [ ]. Usando o método exato para a toda a série, descobrimos que a soma incondicional dos quadrados para  $= 0,5$  é

368

$$(0,5) = \tilde{y}_0 [0,5, \text{pol}]^2 = 27.691$$

## 216 ESTIMATIVA DE PARÂMETROS

que para este exemplo em particular está muito próximo do valor condicional  $\hat{y}(0,5) = 27,694$ .

A soma incondicional dos quadrados ( ), para valores entre -0,5 e +0,5, foi

adicionados à Tabela 7.1 e estão muito próximos dos valores condicionais  $\hat{y}(\cdot)$  calculados anteriormente.

### 7.1.5 Procedimento Geral para Cálculo da Soma Incondicional dos Quadrados

No exemplo acima, foi um processo de média móvel de primeira ordem, com média zero. Isto seguiu-se que todas as previsões para lead times maiores que 1 eram zero e, consequentemente, que apenas um valor preliminar (o back-forecast [0]=1,6) foi necessário para iniciar os cálculos recursivos usando a abordagem aproximada, e apenas um valor [0] na abordagem exata. Por

um processo de média móvel de  $a$  ordem, valores preliminares diferentes de zero [0], [1], ..., [1] seriam necessários, ou equivalentemente, os valores [0], [1], ..., [1] na abordagem exata,

com ( ) =  $\hat{y}$  capaz de estimar parâmetros do processo autorregressivo na Seção 7.3, então, não é possível

A7.3 que o procedimento descrito nesta seção pode fornecer a soma incondicional de quadrados para qualquer modelo ARIMA.

Especificamente, suponha que os  $\hat{y}$ 's sejam gerados pelo modelo direto estacionário

$$(\hat{y}) = (\hat{y}) \quad (7.1.9)$$

onde  $\hat{y} = \hat{y}$  e  $\hat{y}$  o modelo para  $\hat{y}$ . Então, eles poderiam igualmente ter sido gerados por trás

$$(\hat{y}) = (\hat{y}) \quad (7.1.10)$$

Como antes, no método aproximado que utiliza o modelo para trás, poderíamos primeiro

empregar (7.1.10) para fornecer retrovisões [1], ..., [1]. Teoricamente, a presença de

o operador autoregressivo assegura uma série de tais estimativas de extensão infinita.

No entanto, assumindo estacionariedade, as estimativas [1], ..., [1] convergem e além de algum ponto  $\hat{y}$ ,

de tamanho moderado, tornam-se essencialmente iguais a zero. Assim, com uma aproximação suficiente, nós podemos escrever

$$= \hat{y}(1) = \hat{y} \quad \hat{y} = 0$$

Isso significa que o processo misto original pode ser substituído por um processo de média móvel de ordem  $a$  e o procedimento para médias móveis descritos na Seção 7.1.4 podem ser usados.

Assim, em geral, o conjunto dual de equações para gerar as expectativas condicionais

[1], ..., [1] é obtido tomando as expectativas condicionais em (7.1.10) e (7.1.9). Aquilo é,

$$(\hat{y}) = (\hat{y}) \quad (7.1.11)$$

é usado primeiro para gerar as previsões para trás e, em seguida,

$$(\hat{y}) = (\hat{y}) \quad (7.1.12)$$

$$\begin{aligned} [\ddot{y}|, w] &= 0 && = 0, 1, 2, \dots \\ [\ddot{y}|, w] &= 0 && > -1 \end{aligned} \quad (7.1.13)$$

Alternativamente, para o modelo geral (7.1.9), o método exato discutido no Apêndice A7.3 pode ser usado para obter a soma dos quadrados como

$$(\quad, \quad) = \ddot{y} \quad [ \quad ]^2 + ([w\ddot{y}] \ddot{y} C\ddot{y} [a\ddot{y}]) \ddot{y} K\ddot{y}1([w\ddot{y}] \ddot{y} C\ddot{y} [a\ddot{y}]) \quad (7.1.14)$$

Aqui, os vetores  $[w\ddot{y}]^{\bar{y}} = ([\ddot{y} \ 1\ddot{y}], \dots, [\ddot{y} \ 0])$  e  $[a\ddot{y}]^{\bar{y}} = ([1\ddot{y}], \dots, [0])$  são os exatos valores retroprevistos obtidos como em (A7.3.12). Eles são dados por  $[e\ddot{y}] = ([w\ddot{y}]^{\bar{y}}, [a\ddot{y}]^{\bar{y}})^{\bar{y}} = \ddot{D}\ddot{y}1\ddot{F}\ddot{y} \ u$ , onde os valores  $= 1, \dots$ , recursão  $=$  do vetor  $u$  são obtidos através do retrocesso com valores iniciais zero  $= 0, \varphi_1^{-\bar{y}-}$  +  $0 \ 0 \ 0$  são os valores condicionais do calculado de (7.1.12) usando valores iniciais zero,  $\frac{0}{1-} = 0$  e  $\ddot{y} \ 0 \ \frac{\ddot{y} \ 0}{1-} = 0$ . Após resolver as equações  $D[e\ddot{y}] = \ddot{F}\ddot{y} \ u$ , como descrito em (A7.3.12), os  $[ \ ]$ 's exatos são então calculados através da recursão

$$[\ddot{U}] = [\ddot{U}_1] \ddot{V}_1 [1] \quad [\ddot{V}_1] \ddot{V} \ddot{V} \ddot{V} [1] \quad \ddot{V} [1] + [1] \ddot{V}_1 [1] + \ddot{V} + [\ddot{V}] \quad (7.1.15)$$

for  $i = 1, 2, \dots$ , usando as previsões exatas como valores iniciais, com  $\tilde{y} = y$ . As matrizes  $C$ ,  $K$ ,  $D$  e  $\tilde{y}$  para  $i = 1$  são necessárias para o cálculo em (7.1.14) são definidas explicitamente no Apêndice A7.3.

**Comente sobre a aproximação.** Vimos que para o modelo IMA(0, 1, 1) ajustado ao IBM Série B, as somas de quadrados *condicionais* fornecem uma aproximação muito próxima de o valor incondicional. Este será geralmente o caso de períodos não sazonais suficientemente longos série temporal. No entanto, como é discutido mais adiante no Capítulo 9, para séries sazonais, em particular, a aproximação condicional torna-se menos satisfatória e a soma incondicional de quadrados devem ser normalmente calculados. Além disso, incluir o determinante na função de probabilidade similar para obter estimativas de máxima verossimilhança exata dos parâmetros pode ser benéfico se as raízes do operador de média móvel estiverem próximas do círculo unitário.

Estudos de simulação foram realizados por Dent e Min (1978) e Ansley e Newbold (1980) para investigar e comparar empiricamente o desempenho da condicional mínimos quadrados, mínimos quadrados incondicionais e estimadores de máxima verossimilhança para ARMA modelos. Geralmente, os estimadores de mínimos quadrados condicionais e incondicionais servem como aproximações satisfatórias para o estimador de máxima verossimilhança para tamanhos de amostras grandes. No entanto, a evidência da simulação sugere uma preferência pelo estimador de máxima verossimilhança para tamanhos de amostra pequenos ou moderados, especialmente se o operador de média móvel tiver uma raiz próxima ao limite da região de invertibilidade. Algumas informações adicionais sobre o desempenho relativo dos diferentes estimadores foi fornecido por Hillmer e Tião (1979) e Osborn (1982), que examinou os *valores esperados* da soma condicional dos quadrados, a soma incondicional dos quadrados e a probabilidade logarítmica para um modelo MA(1), como funções do parâmetro desconhecido, para diferentes tamanhos de amostra. Esses estudos fornecem uma idéia de

## 218 ESTIMATIVA DE PARÂMETROS

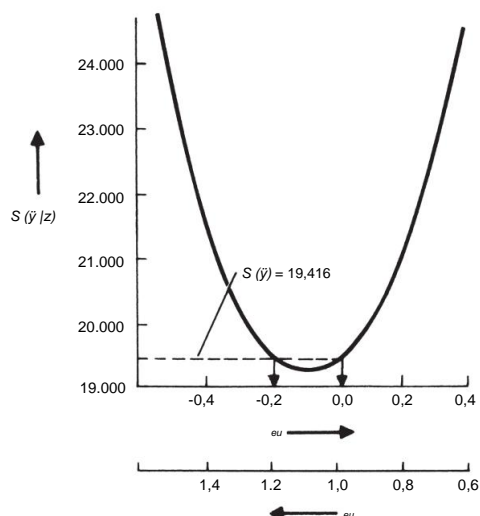


FIGURA 7.1 Gráfico de ( ) para Série B.

como os estimadores correspondentes se comportarão para vários tamanhos de amostra, e os resultados são consistente com os obtidos a partir de estudos de simulação.

## 7.1.6 Estudo Gráfico da Função Soma dos Quadrados

A função de soma de quadrados ( ) para os dados IBM fornecidos na Tabela 7.1 é plotada em Figura 7.1. A soma mínima geral dos quadrados é aproximadamente  $-0,09$  ( $= 1,09$ ), o que é a estimativa de mínimos quadrados e, na suposição de normalidade, uma aproximação próxima a a *estimativa* de máxima verossimilhança do parâmetro .

O estudo gráfico das funções de soma de quadrados é prontamente estendido para dois parâmetros, avaliando a soma de quadrados sobre uma grade adequada de valores de parâmetros e plotando contornos. Como discutido anteriormente, na suposição de normalidade, os contornos são muito próximos contornos de probabilidade. A Figura 7.2 mostra uma grade de valores ( 0, 1) para a Série B equipada com o Modelo IMA(0, 2, 2):

$$\begin{aligned} \tilde{y}_2 &= (1 \tilde{y}_1 + 2 \tilde{y}_2) \\ &= [1 \tilde{y}_1 (2 \tilde{y}_2 + 1) \tilde{y}_1 (0 \tilde{y}_1 + 2) \tilde{y}_2] \end{aligned} \quad (7.1.16)$$

ou na forma

$$\tilde{y}_2 = (0 \tilde{y}_1 + 1) \tilde{y}_1 + \tilde{y}_2$$

A soma mínima dos quadrados na Figura 7.2 é aproximadamente  $= 0,0$ . ~~em 1,09~~ <sup>em 1,09</sup> confirma assim que o modelo preferido neste caso é um processo IMA(0, 1, 1). O dispositivo ilustrado aqui, de encaixar um modelo um pouco mais elaborado do que o esperado necessário, pode fornecer uma confirmação útil da identificação original. A elaboração de o modelo deve ser feito, é claro, na direção "temida" ser necessária.

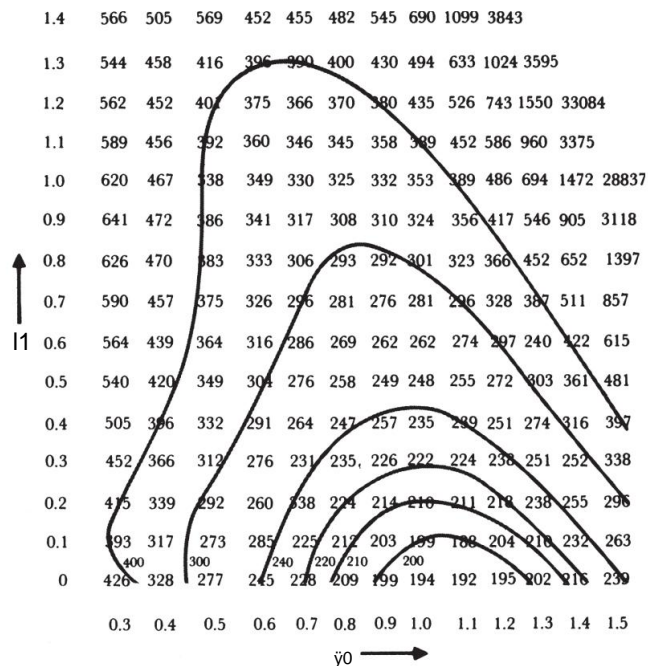


FIGURA 7.2 Valores de ( contornos,  $0, 1 \times 10^2$  para Série B em uma grade de (  $0, 1$ ) valores e aproximados

**Três Parâmetros.** Quando desejamos estudar modelos com três parâmetros, modelos bidimensionais diagramas de contorno para vários valores do terceiro parâmetro podem ser desenhados. Para ilustração, parte de tal série de diagramas é mostrada na Figura 7.3 para as Séries A, C e D. Em cada caso, o modelo "elaborado"

$$\begin{aligned} \ddot{y}_2 &= (1 \ddot{y}_1^2 + 2 \ddot{y}_1 \ddot{y}_2 + \ddot{y}_2^2) \\ &= [1 \ddot{y}_1^2 + 2 \ddot{y}_1 \ddot{y}_2 + \ddot{y}_2^2] \end{aligned}$$

ou

$$\ddot{y}_2 = (\ddot{y}_1 \ddot{y}_2 + 0 \ddot{y}_1 + \ddot{y}_2)$$

foi ajustado, levando à conclusão de que os modelos de melhor ajuste deste tipo<sup>4</sup> são tão mostrados na Tabela 7.3.

A inclusão de parâmetros adicionais (particularmente  $-1$ ) neste processo de ajuste não é estritamente necessário, mas nós os incluímos para ilustrar o efeito do overfitting e para mostram quão perto nossa identificação parece ser confirmada para essas séries.

<sup>4</sup>Mostramos mais adiante na Seção 7.2.5 que ajustes ligeiramente melhores são obtidos em alguns casos com modelos intimamente relacionados contendo termos autorregressivos "estacionários".

220 ESTIMATIVA DE PARÂMETROS

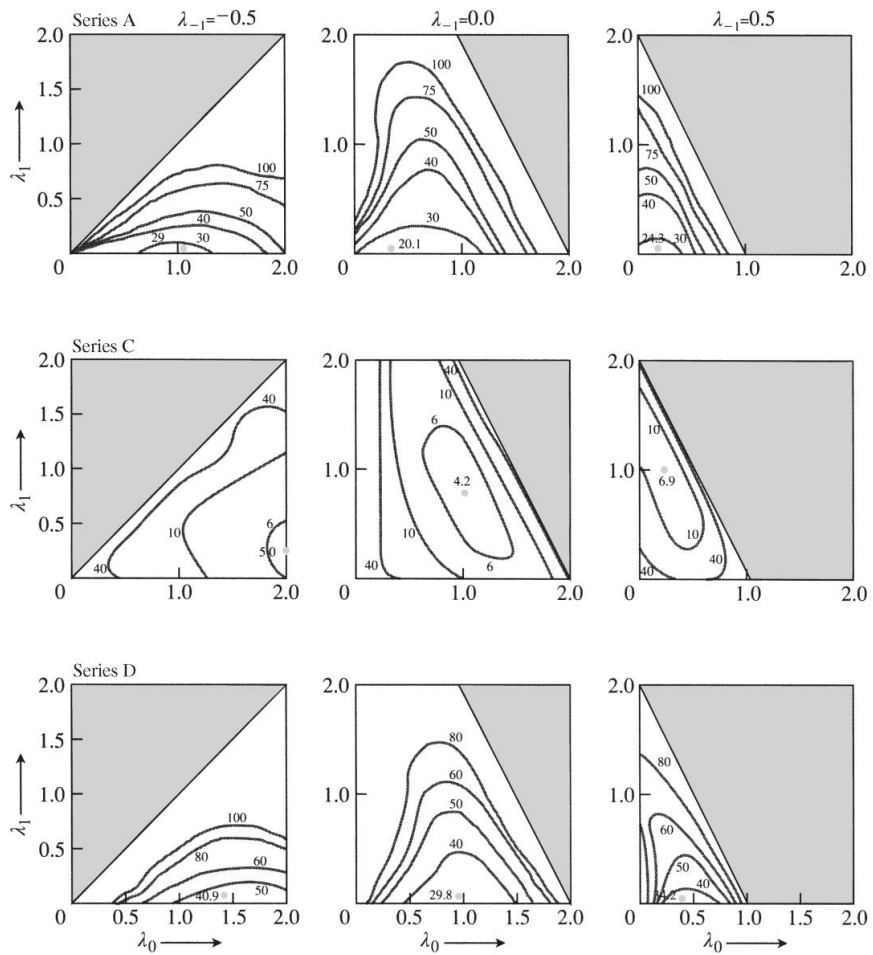


FIGURA 7.3 Contornos de soma de quadrados para as Séries A, C e D (linhas sombreadas indicam limites de as regiões de invertibilidade).

TABELA 7.3 Modelos IMA Ajustados às Séries A, C e D

Series	$\hat{\gamma}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	Série ajustada
UMA	0	0,3	0,0	$\hat{\gamma} = 0,3 \hat{\gamma}_2$
C	0	1,1	0,8	$= 1,1 \hat{\gamma}$
D	0	0,9	0,0	$\hat{\gamma} = 0,9$

7.1.7 Exame da Função de Probabilidade e Regiões de Confiança

A função de verossimilhança não é, obviamente, plotada apenas para indicar a máxima verossimilhança valores. O gráfico desta função contém a totalidade das informações que vem do dados. Em alguns campos de estudo, podem ocorrer casos em que a função de verossimilhança tem duas ou mais picos e também onde a função de verossimilhança contém cristas e picos nítidos. Em cada tal



Nesse caso, a função de verossimilhança está tentando nos dizer algo que precisamos saber. Assim, a existência de dois picos de alturas aproximadamente iguais implica que existem dois conjuntos de valores de parâmetros que podem explicar os dados. A existência de cristas orientadas obliquamente significa que um valor de um parâmetro, consideravelmente diferente de sua máxima verossimilhança valor, poderia explicar os dados se acompanhados por um valor do outro parâmetro, que desviado adequadamente. Para entender completamente a estimativa, é útil examinar a função de verossimilhança tanto analítica quanto graficamente.

**Necessidade de Cuidado na Interpretação da Função de Probabilidade.** É preciso cuidado na interpretação a função de verossimilhança. Por exemplo, os resultados discutidos posteriormente, que assumem que a probabilidade logarítmica é aproximadamente quadrática perto de seu máximo, claramente não se aplicam aos casos de três parâmetros descritos na Figura 7.3. No entanto, esses exemplos são excepcionais porque aqui estamos deliberadamente *superajustando* o modelo. Se o modelo mais simples for justificado, devemos *esperar* encontrar os contornos da função de verossimilhança truncados perto de seu máximo por um limite no espaço de parâmetros de dimensão superior. No entanto, aproximações quadráticas *poderia* ser usado se o modelo *identificado* mais simples, em vez do modelo superparametrizado, fosse montado.

Cuidado especial é necessário quando o máximo da função de verossimilhança pode estar próximo ou um limite. Considere a situação mostrada na Figura 7.4 e suponha que sabemos a priori que  $\beta_0 > 0$ . Será notado que a máxima verossimilhança está no intervalo permitido de  $\beta$  se  $\beta_0 > 0$ , neste caso não em  $\beta = 0$ . A aproximação quadrática é diferente de zero no valor de máxima verossimilhança e que a aproximação quadrática certamente não é uma representação adequada da probabilidade.

Quando uma classe de problemas de estimação é examinada inicialmente, é importante traçar o função de probabilidade para identificar problemas potenciais. Depois que o comportamento de um modelo potencial é bem compreendido, e o conhecimento da situação indica que é apropriado fazê-lo, pode tomar certos atalhos, que agora consideramos. Começamos considerando expressões para as variâncias e covariâncias das estimativas de máxima verossimilhança, apropriadas quando a probabilidade logarítmica é aproximadamente quadrática e o tamanho da amostra é moderadamente grande.

A seguir, é conveniente definir um vetor cujos elementos  $\beta_0$  e  $\beta_1$  são os parâmetros autoregressivos e de média móvel e  $\beta_2$ . Assim, o conjunto completo de

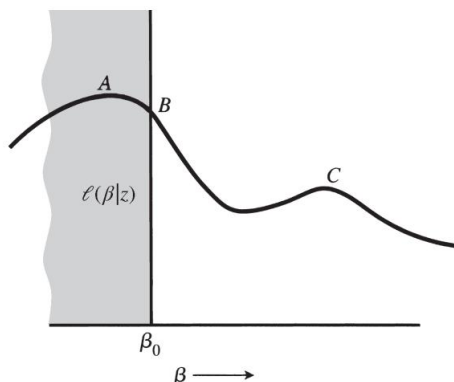


FIGURA 7.4 Função de verossimilhança hipotética com uma restrição  $\beta_0 > 0$ .

0.

## 222 ESTIMATIVA DE PARÂMETROS

$\phi + 1 = \phi + 1$  parâmetros do processo ARMA podem ser escritos como ou  $\phi, \theta$ ; ou como  $\phi, \theta$ ; simplesmente como

**Variações e Covariâncias de Estimativas de ML.** Para o ARMA devidamente parametrizado modelo, muitas vezes acontecerá que sobre a região relevante<sup>5</sup> do espaço de parâmetros, a probabilidade logarítmica é aproximadamente quadrática nos elementos de  $(\phi, \theta)$ , de modo que

$$l(\phi, \theta) \approx l(\phi_0, \theta_0) + \frac{1}{2} (\phi - \phi_0, \theta - \theta_0)' I(\phi_0, \theta_0) (\phi - \phi_0, \theta - \theta_0) \quad (7.1.17)$$

onde, para a aproximação considerada, as derivadas

$$I(\phi, \theta) = - \frac{\partial^2 l(\phi, \theta)}{\partial \phi \partial \phi'} \quad (7.1.18)$$

são constantes. Para grandes  $n$ , a influência do termo  $(\phi, \theta)$  em (7.1.5) pode ser ignorada na maioria dos casos. Portanto,  $(\phi, \theta)$  será essencialmente quadrático em  $(\phi, \theta)$  se isso for verdade para  $(\phi, \theta)$ . Alternativamente,  $(\phi, \theta)$  será essencialmente quadrático se as expectativas condicionais  $E[\phi | w]$  em (7.1.6) são aproximadamente localmente lineares nos elementos de  $w$ . Assim, para amostras moderadas e grandes tamanhos quando a aproximação quadrática local (7.1.17) é adequada, aproximações úteis às variâncias e covariâncias das estimativas e regiões de confiança aproximadas podem ser obtido.

**Matriz de Informação dos Parâmetros.** A matriz  $(\phi, \theta)' I(\phi, \theta) (\phi, \theta)$  é referida (Fisher, 1956; Whittle, 1953) como a *matriz de informação* para os parâmetros  $(\phi, \theta)$ , onde a expectativa é assumida sobre a distribuição de  $w$ . Para um determinado valor de  $(\phi, \theta)$ , a matriz de *variância-covariância*  $V(\phi, \theta)$  para as estimativas de ML é, para grandes amostras, dada por o inverso desta matriz de informação, ou seja,

$$V(\phi, \theta) = I(\phi, \theta)^{-1} \quad (7.1.19)$$

Por exemplo, se  $n = 2$ , a matriz de variância-covariância de amostra grande é

$$V(\phi, \theta) = \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n} \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n} \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n} \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n} \end{bmatrix}$$

Além disso, as estimativas de ML  $\hat{\phi}, \hat{\theta}$  obtido a partir de um processo ARMA inversível estacionário mostraram-se distribuídas assintoticamente como *normal multivariada* com vetor médio e matriz de covariância  $I(\phi, \theta)^{-1}$  (por exemplo, Mann e Wald, 1943; Whittle, 1953; Hannan, 1960; Walker, 1964) no sentido de que  $\sqrt{n}(\hat{\phi} - \phi, \hat{\theta} - \theta)$  converge na distribuição para o multivariado normal  $\{0, \text{eu } I(\phi, \theta)^{-1}\}$  como  $\sqrt{n}(\hat{\phi} - \phi, \hat{\theta} - \theta)$ , onde  $I(\phi, \theta) = \lim_{n \rightarrow \infty} \frac{1}{n} I(\phi, \theta)$ . A forma específica do matriz de informação  $I(\phi, \theta)$  e a matriz limitante  $I(\phi, \theta)$  para modelos ARMA  $(\phi, \theta)$  são descritas na Seção 7.2.6, e detalhes sobre a normalidade assintótica do estimador são examinados para o caso especial de modelos AR no Apêndice A7.5.

<sup>5</sup>Diga sobre uma região de confiança de 95%.

Agora, usando (7.1.5), temos

$$\bar{y} = \frac{\sum y_i}{n} \quad (7.1.20)$$

Onde

$$= \frac{\sum (y_i - \bar{y})^2}{n-1}$$

Além disso, se para grandes amostras, aproximamos os valores esperados de  $\bar{y}$  ou de  $\bar{y}^2$  por os valores realmente observados, então, usando (7.1.19), obtemos

$$V(\bar{y}) = \frac{\sum (y_i - \bar{y})^2}{n-1} = \frac{\sum y_i^2}{n} - \bar{y}^2 \quad (7.1.21)$$

Assim, para  $n = 2$ ,

$$V(\bar{y}) = \frac{\sum y_i^2}{n} - \bar{y}^2 = \frac{y_1^2 + y_2^2}{2} - \left(\frac{y_1 + y_2}{2}\right)^2$$

Se  $(y_1, y_2)$  fosse exatamente quadrático sobre a região relevante do espaço de parâmetros, então todas as derivadas seriam constantes nessa região. Na prática, o valor varia um pouco, e normalmente assumiremos que as derivadas são determinadas no ou próximo ao ponto  $\bar{y}$ . Agora, é mostrado nos Apêndices A7.3 e A7.4 que uma estimativa<sup>6</sup> de  $\bar{y}^2$  é fornecido por

$$\bar{y}^2 = \frac{(\sum y_i)^2}{n^2} \quad (7.1.22)$$

e que para amostras grandes,  $\bar{y}^2$  e  $\bar{y}$  são não correlacionados. Finalmente, os elementos de (7.1.21) podem ser estimado a partir de

$$a = \frac{1}{n-1} \sum (y_i - \bar{y})^2 \quad (7.1.23)$$

onde a matriz  $(x_i)$  é dada por  $\{y_i\}$  e a expressão (7.1.23) é para definir a variância  $(a)$  quando  $n = 2$ .

**Regiões de confiança aproximadas para os parâmetros.** Em particular, esses resultados permitem para obter as variações aproximadas de nossas estimativas. Tirando a raiz quadrada desses variâncias, obtemos *erros padrão* aproximados (SE) das estimativas. O erro padrão de uma estimativa]. Quando se considera vários parâmetros simultaneamente, precisamos de alguns meios de julgar a precisão das estimativas *em conjunto*. Um meio de fazer isso é determinar uma *região de confiança*. Se, para dado  $\bar{y}^2$ ,  $(\bar{y}, \bar{y}^2)$  é aproximadamente quadrática na vizinhança de  $\bar{y}$ , então usando (7.1.19) (veja também

<sup>6</sup>Argumentos podem ser avançados para usar o divisor  $\bar{y} = \bar{y} - \bar{y}$  em vez de  $\bar{y}$  em (7.1.22), mas para tamanhos de amostra moderados, esta modificação não faz muita diferença.

## 224 ESTIMATIVA DE PARÂMETROS

**TABELA 7.4 ( ) e seu primeiro e segundo  
Diferenças para Vários Valores da Série B**

$=1\ddot{y}$	( )	$\ddot{y}(\ )$	$\ddot{y}2(\ )$
1,5	23.928	2.333	960
1,4	21.595	1.373	634
1,3	20.222	739	476
1,2	19.483	263	406
1,1	19.220	$\ddot{y}143$	390
1,0	19.363	$\ddot{y}533$	422
0,9	19.896	$\ddot{y}955$	508
0,8	20.851	$\ddot{y}1.463$	691
0,7	22.314	$\ddot{y}2.154$	1069
0,6	24.468	$\ddot{y}3.223$	
0,5	27.691		

Apêndice A7.1), uma região de confiança aproximada de 1  $\ddot{y}$  será definida por

$$\ddot{y} \ddot{y} [ \ddot{y} ( \ddot{y} ) ] < \ddot{y}^2 ( ) \quad (7.1.24)$$

onde  $\ddot{y}^2 ( )$  é o ponto de significância superado por uma proporção do  $\chi^2$  distribuição, tem graus de liberdade.

Alternativamente, usando a aproximação (7.1.21) e substituindo a estimativa de (7.1.22) por  $\ddot{y}^2$ , a região de confiança aproximada é dada por 7

$$\ddot{y} \ddot{y} ( - ) ( \ddot{y} ) < 2 \ddot{y}^2 ( ) \quad (7.1.25)$$

No entanto, para uma superfície quadrática ( )

$$( ) \ddot{y} ( \ddot{y} ) = \frac{1}{2} \ddot{y} \ddot{y} ( - ) ( \ddot{y} ) \quad (7.1.26)$$

Assim, usando (7.1.22) e (7.1.25), finalmente obtemos o resultado de que a região de confiança aproximada de 1  $\ddot{y}$  é limitada pelo contorno na superfície da soma de quadrados, para a qual

$$( ) = ( \ddot{y} ) \left[ 1 + \frac{\ddot{y}^2}{( )} \right] \quad (7.1.27)$$

### **Exemplos de Cálculo de Intervalos e Regiões de Confiança Aproximados.**

- 1. Exemplo: Série B.** Para a Série B, os valores de ( ) e suas diferenças são mostrados na Tabela 7.4. A segunda diferença de ( ) não é constante e, portanto, ( ) não é estritamente quadrático. No entanto, na faixa de  $\ddot{y} = 0,85$  a  $\ddot{y} = 1,35$ ,  $\ddot{y}2( )$  não muda muito, de modo que (7.1.27) pode ser esperado para fornecer uma aproximação razoavelmente próxima

7Aproximação um pouco mais próxima com base na distribuição  $F$ , que leva em conta a amostragem aproximada distribuição de  $\ddot{y}^2$ , pode ser empregado. Para tamanhos de amostra moderados, esse refinamento não é muito prático diferença.

## ESTUDO DAS FUNÇÕES DE PROBABILIDADE E SOMA DE QUADRADOS 225

imação. Com um valor mínimo ( $\hat{y}$ ) = 19, 216, o valor crítico ( $t_{\alpha/2}$ ), definindo um intervalo de confiança aproximado de 95%, é então dado por

$$(\hat{y}) = 19, 216 \left( 1 + \frac{3,84}{368} \right) = 19, 416$$

Lendo os valores correspondentes a ( $\hat{y}$ ) = 19.416 na Figura 7.1, obtemos um intervalo de confiança aproximado  $0,98 < 1,19$ .

Alternativamente, podemos empregar (7.1.25). Usando a segunda diferença em  $\hat{y} = 1,1$ , dado na Tabela 7.4, para aproximar a derivada, obtemos

$$\hat{y}_{11} = \frac{2}{2} \hat{y} - \frac{390}{(0,1)^2}$$

Além disso, usando (7.1.22),  $\hat{y}^2 = 19, 216 \hat{y} / 368 = 52,2$ . Assim, o intervalo de confiança de 95%, definido por (7.1.25), é

$$\frac{390}{(0,1)^2} (\hat{y} - 1,09)^2 < 2 \times 52,2 \times 3,84$$

isto é,  $|\hat{y} - 1,09| < 0,10$ . Assim, o intervalo é  $0,99 < 1,19$ , o que concorda estreitamente com o cálculo anterior.

Neste exemplo, onde há apenas um único parâmetro, o uso de (7.1.24) e (7.1.25) é equivalente a usar um intervalo é o valor  $\hat{y}_{11}$  ( $\hat{y}$ ), onde  $\hat{y}_2$  exclui uma proporção  $\hat{y}_2$  na cauda superior da distribuição normal padrão. Um erro padrão aproximado para o  $\hat{y}$ , ( $\hat{y}) = \hat{y}^2 - \hat{y}_2^2 - \hat{y}_{11}^2$ , é obtido de (7.1.23). No presente exemplo,

$$(\hat{y}) = \hat{y}^2 - \hat{y}_2^2 - \hat{y}_{11}^2 = \frac{2 \times 52,2 \times 0,12}{390} = 0,00268$$

e o erro padrão aproximado é  $\hat{y}(\hat{y}) = \hat{y} \cdot 0,00268 = 0,052$ . Assim,  $\pm 1,96 \hat{y}(\hat{y}) = 1,09 \pm 0,10$ , mate intervalo de confiança de 95% é como antes.

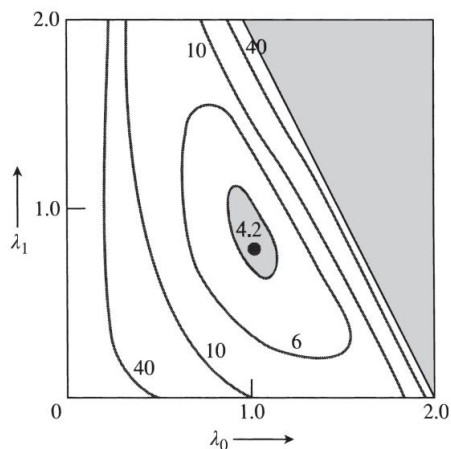
Finalmente, mostramos mais adiante na Seção 7.2.6 que é possível avaliar (7.1.19) analiticamente, para grandes amostras de um processo MA(1), produzindo

$$(\hat{y}) \hat{y} = \frac{(2 - \hat{y})}{2}$$

Para o presente exemplo, substituindo  $\hat{y} = 1,09$  para que  $\hat{y}$ , encontramos que  $(\hat{y}) \hat{y} = 0,00269$ , concorda de perto com a estimativa anterior e, portanto, produz o mesmo erro padrão de 0,052 e o mesmo intervalo de confiança.

2. *Exemplo: Série C.* Na identificação da Série C, um modelo que foi atendido foi um processo (0, 2, 2). Para ilustrar a aplicação de (7.1.27) para mais de um parâmetro, a Figura 7.5 mostra uma região de confiança aproximada de 95% (sombreada) para o  $\hat{y}$  da Série C. Para este exemplo,  $(\hat{y}) = 4,20$ ,  $\hat{y} = 224$  e

## 226 ESTIMATIVA DE PARÂMETROS



**FIGURA 7.5** Contornos de soma de quadrados com região de confiança de 95% sombreada para Série C, assumindo um modelo de ordem (0, 2, 2).

de modo que a região de confiança aproximada de 95% seja limitada pelo contorno para que

$$(0, 1) = 4,20 \left( 1 + \frac{5,99}{224} \right) = 4,31$$

## 7.2 ESTIMATIVA NÃO LINEAR

### 7.2.1 Método Geral de Abordagem

A plotagem da função soma de quadrados é de particular importância no estudo de novos problemas de estimativa porque garante que quaisquer peculiaridades na situação de estimativa mostrar-se. Quando estivermos convencidos de que as anomalias são improváveis, outros métodos podem ser usados.

Vimos que, para a maioria dos casos, as estimativas de máxima verossimilhança estão próximas aproximado pelas estimativas de mínimos quadrados, que minimizam

$$Q(\beta) = \sum_{i=1}^n \tilde{y}_i^2 + [\mathbf{e}\tilde{\mathbf{y}}] \tilde{\mathbf{y}} \mathbf{1} [\mathbf{e}\tilde{\mathbf{y}}]$$

e na prática, esta função pode ser aproximada por uma soma finita  $\tilde{y} = \sum_{i=1}^n \tilde{y}_i^2$ .

Em geral, uma simplificação considerável ocorre na minimização em relação a uma soma de quadrados  $\tilde{y} = \sum_{i=1}^n \tilde{y}_i^2$ , se cada  $\tilde{y}_i$  ( $i = 1, 2, \dots, n$ ) é uma função *linear* dos parâmetros. Mostramos agora que os modelos autorregressivo e de média móvel diferem com em relação à linearidade do  $\tilde{y}_i$ . Para o processo puramente autorregressivo,  $\tilde{y}_i = \sum_{j=1}^p \tilde{y}_{i-j} + \epsilon_i$

$$\tilde{y}_i = \sum_{j=1}^p \tilde{y}_{i-j} + \epsilon_i$$

Agora para  $\gamma > 0$ , Assim, exceto pelo efeito de valores parâmetros,  $[Y]$  é linear e  $[X]$  são ambas as funções de

's. Por outro lado, para o processo de média móvel pura,

$$[Y] = \gamma_1([X]) \quad \frac{[Y]}{\gamma} = \gamma_2([X]) + \gamma_1([X]) \quad \frac{[Y]}{\gamma}$$

de modo que os  $[Y]$  são sempre funções não lineares dos parâmetros de média móvel.

Veremos na Seção 7.3 que simplificações especiais ocorrem na obtenção de mínimos quadrados e estimativas de máxima verossimilhança para o processo autorregressivo. Mostramos no presente seção como, por aplicação iterativa de mínimos quadrados lineares, as estimativas podem ser obtidas para qualquer processo ARMA.

**Linearização do Modelo.** No que se segue, continuamos a usar como um símbolo geral para os parâmetros  $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ . Precisamos, então, minimizar

$$Q(\theta) = \sum_{t=1}^n [Y_t - \gamma(\theta)]^2 = \sum_{t=1}^n [Y_t - \gamma(\theta)]^2$$

Expandindo  $[Y_t - \gamma(\theta)]$  em uma série de Taylor sobre seu valor correspondente a algum conjunto de  $[Y_t]$  valores de parâmetro  $\theta_0 = (\theta_{01}, \theta_{02}, \dots, \theta_{0n})$ , temos aproximadamente

$$[Y_t - \gamma(\theta)] \approx [Y_t - \gamma(\theta_0)] + \sum_{i=1}^n \frac{\partial \gamma(\theta)}{\partial \theta_i} \bigg|_{\theta=\theta_0} (\theta_i - \theta_{0i}) \quad (7.2.1)$$

onde  $[Y_t - \gamma(\theta_0)] = [w_t]$  e

$$\frac{\partial \gamma(\theta)}{\partial \theta_i} \bigg|_{\theta=\theta_0} = 0$$

Agora, se  $X$  é a matriz  $(n \times n)$ , então as equações (7.2.1) podem ser expressas

Como

$$[a_0] = X(\gamma(\theta_0) - \gamma(\theta_0)) + [a]$$

onde  $[a_0]$  e  $[a]$  são vetores coluna com  $n$  elementos.

Os ajustes  $\gamma$  que minimizam  $Q(\theta) = \sum_{t=1}^n [Y_t - \gamma(\theta)]^2$   $[a]$ , agora pode ser obtido por mínimos quadrados lineares, isto é, "regredindo" os  $[Y_t]$ 's para os  $\theta$ 's. Isso dá a estimativas lineares usuais de mínimos quadrados, conforme apresentado no Apêndice A7.2.1, dos ajustes  $\theta_0 = (X^T X)^{-1} X^T Y$   $[a_0]$ , portanto,  $\theta_0 = (X^T X)^{-1} X^T Y$   $[a_0]$ . Porque os  $[Y_t]$  não ser exatamente linear nos parâmetros, um único ajuste não produzirá imediatamente o valores finais de mínimos quadrados. Em vez disso, os valores ajustados são substituídos por novas estimativas e o processo é repetido até que a convergência ocorra. A convergência é mais rápida se razoavelmente boas suposições, como as que podem ser obtidas no estágio de identificação, são usadas inicialmente. Se suficientemente estimativas iniciais ruins são usadas, o processo pode não convergir.

## 7.2.2 Estimativas Numéricas dos Derivativos

Os derivados  $\frac{\partial \gamma(\theta)}{\partial \theta_i}$  podem ser obtidos diretamente, como ilustraremos mais adiante. Eles também podem ser computado numericamente usando uma rotina geral de mínimos quadrados não linear. Isso está feito

## 228 ESTIMATIVA DE PARÂMETROS

perturbando os parâmetros "um de cada vez". Assim, para um determinado modelo, os valores  $[w, 1, 0, 2, 0, \dots]$ , são calculados recursivamente, usando qualquer "back-forecast" apropriado, se necessário. O cálculo

$[w, 1, 0 + 1, 2, 0, \dots, 0]$ , então para  $[w, 1, 0, 2, 0 + 2, \dots, 0]$ , e assim por diante. O negativo da derivada requerida é então dado com precisão suficiente usando

$$= \frac{[w, 1, 0, \dots, 0, \dots, 0] \tilde{y} [w, 1, 0, \dots, 0 + \dots, 0]}{\dots} \quad (7.2.2)$$

O método numérico descrito acima tem a vantagem de aplicabilidade universal e exige que programemos o cálculo dos  $\tilde{y}$ 's apenas, não suas derivadas. Em geral rotinas de estimativa não linear, que essencialmente requerem apenas instruções de entrada sobre como calcular os  $\tilde{y}$ 's, estão geralmente disponíveis. Em algumas versões, é necessário escolher o  $\tilde{y}$  é com antecedência. Em outros, o próprio programa realiza uma iteração preliminar para encontrar adequados. Muitos programas incluem recursos especiais para evitar overshoot e acelerar convergência.

Desde que a solução de mínimos quadrados não esteja dentro ou perto de um limite restritivo, o valor de  $X = X \tilde{y}$  da iteração final pode ser usado para calcular variâncias aproximadas, covariâncias e intervalos de confiança. Assim, semelhante aos resultados usuais de mínimos quadrados lineares no Apêndice A7.2.3,

$$(\tilde{y} \tilde{y})^{-1} \tilde{y}^2$$

irá aproximar a matriz de variância-covariância dos  $\tilde{y}$ 's, e  $\tilde{y}^2$  será estimado por  $\tilde{y}^2 = (\tilde{y} \tilde{y})$ .

### 7.2.3 Avaliação Direta dos Derivativos

Mostramos agora que também é possível obter derivadas diretamente, mas são necessários cálculos recursivos adicionais. Para ilustrar o método, basta considerar um processo ARMA(1, 1), que pode ser escrito em qualquer uma das formas como

$$\begin{aligned} \tilde{y}_t &= \theta_1 \tilde{y}_{t-1} + \epsilon_t \\ \tilde{y}_t &= \phi_1 \tilde{y}_{t-1} + \theta_1 \epsilon_t \end{aligned}$$

Vimos na Seção 7.1.4 como as duas versões do modelo podem ser usadas alternadamente, uma fornecendo valores iniciais com os quais iniciar uma recursão com a outra.

Assumimos que um primeiro cálculo já foi feito produzindo valores de  $\tilde{y}$ , de  $\epsilon$ , e de  $[0, \tilde{y}_1, \dots, \tilde{y}_t]$ , como na Seção 7.1.5, e que  $\tilde{y}$ ,  $[\tilde{y} \tilde{y}_1, \dots]$  e, portanto,  $[\tilde{y}]$ ,  $[\tilde{y} \tilde{y}_1]$ ,  $\dots$  são desprezíveis. Mostramos agora que um cálculo duplo semelhante pode ser usado no cálculo das derivadas.

Usando a notação  $\frac{\partial}{\partial \theta_1}$  para denotar a derivada parcial  $\frac{\partial}{\partial \theta_1}$ , obtemos

$$\frac{\partial}{\partial \theta_1} \tilde{y}_t = \frac{\partial}{\partial \theta_1} \tilde{y}_{t-1} + \frac{\partial}{\partial \theta_1} \epsilon_t \quad (7.2.3)$$

$$\frac{\partial}{\partial \theta_1} \tilde{y}_t = \frac{\partial}{\partial \theta_1} \tilde{y}_{t-1} + \frac{\partial}{\partial \theta_1} \tilde{y} [\tilde{y}_1] \quad (7.2.4)$$

$$\frac{\partial}{\partial \theta_1} \tilde{y}_t = \frac{\partial}{\partial \theta_1} \tilde{y}_{t-1} + \frac{\partial}{\partial \theta_1} \tilde{y} [\tilde{y}_1] + \frac{\partial}{\partial \theta_1} \epsilon_t \quad (7.2.5)$$



$$\hat{y}_t = \hat{y}_{t-1} + \hat{y}_{t-1} + [-1] \quad (7.2.6)$$

Agora,

$$\hat{y}_t = \hat{y}_{t-1} = 0, \quad t = 1, 2, \dots, \quad (7.2.7)$$

e

$$\hat{y}_t = 0, \quad t = 0, 1, \dots, \quad (7.2.8)$$

Considere as equações (7.2.3) e (7.2.4). Definindo a  $\hat{y}_{t+1} = 0$  em (7.2.3), podemos iniciar um recursão, que usando (7.2.7) e (7.2.8) eventualmente nos permite calcular  $\hat{y}_t$  por  $t = 0, 1, \dots, \hat{y} - 1$ . Desde  $\hat{y}_t, \hat{y}_{t-1}, \dots$  pode ser tomado como zero, agora podemos usar (7.2.4) para calcular recursivamente as derivadas necessárias  $\hat{y}_t$ . De maneira semelhante, (7.2.5) e (7.2.6) pode ser usado para calcular as derivadas  $\hat{y}_t$ .

## 7.2.4 Algoritmo Geral de Mínimos Quadrados para o Modelo Condicional

Uma aproximação que às vezes usamos com séries longas é definir valores iniciais para os  $\hat{y}_t$ 's e, portanto, para os derivados nos  $\hat{y}_t$ 's, iguais às suas expectativas incondicionais de zero e então prosseguir diretamente com as recursões para frente. O efeito é introduzir um transiente tanto na série quanto na série, sendo a última mais lenta para morrer desde o  $\hat{y}_t$ 's depende do  $\hat{y}_t$ 's. Em alguns casos, onde há uma abundância de dados (digamos, 200 ou mais observações), o efeito da aproximação pode ser anulado à custa de alguns perda de informação, descartando, digamos, os primeiros 10 valores calculados.

Se adotarmos a aproximação, um algoritmo geral interessante para esta condicional resultados do modelo. O modelo ARMA( , ) pode ser escrito como

$$\hat{y}_t = \hat{y}_{t-1} + \hat{y}_{t-1}$$

onde  $\hat{y}_t = \hat{y}_{t-1}$ ,  $\hat{y}_t = -e$

$$\hat{y}_t = 1\hat{y}_{t-1} = 1\hat{y}_{t-1}$$

Se as primeiras suposições para os parâmetros  $\hat{y}_t = ( , )$  são  $\hat{y}_0 = (0, 0)$ , então

$$\hat{y}_0 = \hat{y}_0 + \hat{y}_0$$

e

$$\hat{y}_t = \hat{y}_{t-1} + \hat{y}_{t-1} = \hat{y}_{t-1} + \hat{y}_{t-1}$$

Onde

$$\hat{y}_t = \hat{y}_0 + \hat{y}_0 = \hat{y}_0 + \hat{y}_0 \quad (7.2.9)$$

$$\hat{y}_t = \hat{y}_0 + \hat{y}_0 = \hat{y}_0 + \hat{y}_0 \quad (7.2.10)$$

## 230 ESTIMATIVA DE PARÂMETROS

Os 's, 's e 's podem ser calculados recursivamente, com valores iniciais para 's, de, e 's igual a zero, como segue:

$$\hat{y}_{1,0} = -1,0 \hat{y}_1 \hat{y}_0 + 1,0 \hat{y}_{1,0} + \hat{y}_0 \quad (7.2.11)$$

$$= 1,0 \hat{y}_1 + \hat{y}_0 + 1,0 \hat{y}_1 + \hat{y}_0 \quad (7.2.12)$$

$$= \hat{y}_0 + \hat{y}_0 + \hat{y}_0 \quad (7.2.13)$$

$$= \hat{y}_1 + \hat{y}_0 + 1,0 \hat{y}_0 \quad (7.2.14)$$

Correspondendo a (7.2.1), a equação de regressão linear aproximada torna-se

$$\hat{y}_{1,0} = \hat{y}_0 \sum_{i=1}^n (\hat{y}_i, 0) \hat{y}_i + \hat{y}_1 \quad (7.2.15)$$

Os ajustes são então os coeficientes de regressão de no . Por e a ,0 adicionando os ajustes aos primeiros palpites ( 0, 0), um conjunto de "segundos palpites" é formado e estes agora tomam o lugar de ( 0, 0) em uma segunda iteração, na qual novos valores de ,0, , e são computados, até que eventualmente ocorra convergência.

**Forma Alternativa para o Algoritmo.** A expansão linear aproximada (7.2.15) pode ser escrito na forma

$$\begin{aligned} \hat{y}_{1,0} &= \hat{y}_0 \sum_{i=1}^n (\hat{y}_i, 0) \hat{y}_i + \hat{y}_1 \\ &= \hat{y}_0 \sum_{i=1}^n (\hat{y}_i, 0) \hat{y}_i + \hat{y}_1 \end{aligned}$$

isso é,

$$\hat{y}_{1,0} = \hat{y}_0 \sum_{i=1}^n (\hat{y}_i, 0) \hat{y}_i + \hat{y}_1 \quad (7.2.16)$$

que apresenta o algoritmo de uma forma interessante.

**Aplicação a um processo IMA(0, 2, 2).** Para ilustrar o cálculo com a condicional , aproximação, considere a estimativa de valores de mínimos quadrados para a Série C usando o modelo de pedido (0, 2, 2):

$$= (1 \hat{y}_1 - 2 \hat{y}_2)$$

com =  $\hat{y}_2$  ,

$$= +,0 \quad 1,0 \hat{y}_{1,0} + 2,0 \hat{y}_{2,0} - \hat{y}_0 \quad ,0 + 1,0 \hat{y}_1 + 2,0 \hat{y}_2$$

Usando os valores iniciais = 0,1 e 1,0 e  $\hat{y}_{2,0} = 0,1$ , os primeiros ajustes e 1,0  $\hat{y}_{2,0}$  são encontrados por "regredindo" ,0 em -1  $\hat{y}_2$ . para O processo é repetido até a convergência ocorre. Estimativas sucessivas de parâmetros são mostradas na Tabela 7.5.

TABELA 7.5 Convergência de Parâmetro  
Estimativas para o processo IMA(0, 2, 2)

Iteração	1	2
0	0,1000	0,1000
1	0,1247	0,1055
2	0,1266	0,1126
3	0,1286	0,1141
4	0,1290	0,1149
5	0,1292	0,1151
6	0,1293	0,1152
7	0,1293	0,1153
8	0,1293	0,1153

7.2.5 Modelos ARIMA instalados na Série A--F

Na Tabela 7.6, resumimos os modelos ajustados pelo procedimento iterativo de mínimos quadrados de Seções 7.2.1 e 7.2.2 da Série A--F. Os modelos ajustados foram identificados no Capítulo 6 e resumidos nas Tabelas 6.2 e 6.5. No caso das Séries A, C e D, dois modelos possíveis foram identificados e posteriormente montados. Para as Séries A e D, os modelos alternativos envolvem o uso de um operador autoregressivo estacionário (1  $\hat{y}$ ) em vez do operador de raiz unitária (1  $\hat{y}$ ). O exame da Tabela 7.6 mostra que em ambos os casos os resultados do modelo autorregressivo em uma variância residual um pouco menor, embora os modelos sejam muito semelhantes. Apesar de um ajuste ligeiramente melhor é possível com um modelo estacionário, o modelo IMA(0, 1, 1) pode ser

TABELA 7.6 Resumo dos Modelos Ajustados à Série A--F

Número da série de Observações		Modelos ajustados	Varição residual
UMA	197	$\hat{y} = 0,92 + \hat{y}_1 = 1,45 + \hat{y} 0,58 (\pm 0,08)$	0,097
		$(\pm 0,04) \hat{y} = \hat{y} 0,70 (\pm 0,05)$	0,101
B	369	$\hat{y} = + 0,09 (\pm 0,05)$	52,2
C	226	$\hat{y} \hat{y} 0,82 \hat{y}_1 = (\pm 0,04)$	0,018
		$\hat{y}_2 = \hat{y} 0,13 \hat{y} 0,12 (\pm 0,07) \hat{y} 0,87 \hat{y}_2 (\pm 0,07)$	0,019
D	310	$\hat{y}_1 = 1,17 + (\pm 0,03)$	0,090
		$\hat{y} = \hat{y} 0,06 \hat{y}_1 (\pm 0,06)$	0,096
E	100	$= 14,35 + 1,42 - 0,73 \hat{y}_1 \hat{y}_2 + (\pm 0,07) (\pm 0,07)$	227,8
		$= 11,31 + 1,57 \hat{y} 1,02 \hat{y}_1 \hat{y}_2 + 0,21 \hat{y}_3 + (\pm 0,10) (\pm 0,15) (\pm 0,10)$	218,1
F	70	$= 58,87 \hat{y} 0,342 \hat{y}_1 + 0,19 \hat{y}_2 + (\pm 0,12) (\pm 0,12)$	112,7

Os valores (±) sob cada estimativa denotam os erros padrão dessas estimativas.  
Obtido de (  $\hat{y}$ ,  $\hat{y}$ )  $\hat{y}$ .

## 232 ESTIMATIVA DE PARÂMETROS

preferível nestes casos porque, ao contrário do modelo estacionário, não assume que a série tem uma média fixa. Isso é especialmente importante na previsão de valores futuros da série. Pois se o nível mudar, um modelo com  $\mu > 0$  continuará a rastrear-lo, Considerando que um modelo para o qual  $\mu = 0$  será vinculado a um nível médio que pode ter ficado fora de encontro. Deve-se notar, no entanto, que para o teste formal de raiz unitária da Série D a ser discutido mais adiante na Seção 10.1 não suporta a necessidade de diferenciação e sugere uma preferência para o modelo estacionário AR(1). Além disso, o teste de raiz unitária para a Série C indica uma preferência para o modelo ARIMA(1, 1, 0) sobre um modelo em termos de segundas diferenças. Raiz da unidade testar para a Série A dentro do modelo ARMA(1, 1), no entanto, não rejeita a necessidade do operador não estacionário  $(1 - \gamma)$  para a parte autorregressiva.

Os limites sob os coeficientes na Tabela 7.6 representam os erros padrão das estimativas obtido da matriz de covariância  $(\hat{\gamma}_1, \hat{\gamma}_2)$ , conforme descrito na Seção 7.2.1. Observe que a estimativa no modelo AR(3), ajustado à mancha solar Série E, é 2,1 vezes o padrão erro, indicando que um ajuste marginalmente melhor é obtido pelo método autorregressivo de terceira ordem processo, em comparação com o processo autoregressivo de segunda ordem. Isso está de acordo com a conclusão de Moran (1954).

**Estimativa de parâmetros usando R.** Estimativa de parâmetros para modelos ARIMA com base no métodos descritos acima estão disponíveis no pacote de software R. As ferramentas relevantes incluem o comando `arima()` no pacote `stats` e o comando `sarima()` no `astsa` pacote. Os detalhes dos comandos são obtidos digitando `help(arima)` e `help(sarima)` em R. Usando o comando `arima()`, a ordem do modelo é especificada usando o argumento `order=c(p,d,q)`, e o método de estimação é especificado por `method=c("CSS")` para mínimos quadrados condicionais e `method=c("ML")` para o método de máxima verossimilhança total. o `sarima()` ajusta o modelo ARIMA( , , ) a uma série  $z$  por máxima verossimilhança usando o comando `sarima(z,p,d,q)`.

Para ilustração, primeiro usamos a rotina `arima()` no pacote `stats` para estimar o parametriza o modelo ARIMA(3, 0, 0) para os dados de manchas solares na Série E. comando e uma saída de modelo parcial são fornecidos abaixo.

```
> arima(ts(seriesE),order=c(3,0,0),method=c("CSS"))
```

Coefficientes:

ar1	ar2	interceptação ar3
1,5519 -1,0069 0,2076		46,7513
com 0,0980 0,1540 0,0981		5,9932

sigma^2 estimado como 219,3: log-probabilidade = -411,42, aic = NA

Vemos que as estimativas dos parâmetros autorregressivos estão muito próximas dos valores fornecido na Tabela 7.6. No entanto, usando essa rotina, a interceptação informada na saída é a média da série, de modo que o termo constante no modelo precisa ser calculado como  $\mu = \bar{y}(1 - \gamma_3)$ . Isso dá uma estimativa para a constante de 11,57.

Os comandos e uma saída parcial da execução da análise usando `sarima()` são como segue:

```
> biblioteca(astsa)
> sarima(ts(série E),3,0,0)
```

Coeficientes:

	ar1	ar2	ar3	xmean
1,5531 -1,0018 0,2063 48,4443				
com 0.0981 0.1544 0.0989 6.0706				

$\sigma^2$  estimado como 218,2: log-probabilidade=-412,49, aic 834,99

\$AIC: [1] 6.465354. \$AICc: [1] 6.491737. \$BIC: [1] 5.569561

Os resultados são próximos dos anteriores. O comando `sarima()` tem a vantagem de diagnósticos de modelo do tipo discutido no Capítulo 8 abaixo são fornecidos automaticamente como parte da saída (veja, por exemplo, Figuras 8.2 e 8.3). Isso permite que o usuário com eficiência avaliar a adequação de um modelo ajustado e fazer comparações entre modelos alternativos. Por exemplo, ajustando os modelos AR(2) e AR(3) à série de manchas solares, é prontamente visto que o modelo AR(3) fornece um melhor ajuste aos dados. Além disso, o ajuste pode ser melhorado usando uma raiz quadrada ou transformação logarítmica da série, embora um gráfico Q-Q ainda indica um afastamento da normalidade dos resíduos padronizados.

### 7.2.6 Matrizes de Informações de Grande Amostra e Estimativas de Covariância

Nesta seção, examinamos com mais detalhes a matriz de informação e a matriz de covariância das estimativas dos parâmetros. Denote por  $\mathbf{X} = [\mathbf{U} : \mathbf{V}]$ , a matriz  $\times ( + )$  do tempo defasado e definido em (7.2.13) e (7.2.14), quando os elementos de  $\mathbf{V}$  são os verdadeiros valores de os parâmetros, para um tamanho de amostra  $n$  suficientemente grande para que os efeitos finais sejam ignorados. Então, desde  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  e usando (7.1.20),

$$[\ddot{y}]_{2 \times 2} = \frac{1}{2} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \ddot{y} \quad [\ddot{y}_1] = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \ddot{y}_1$$

a matriz de informações para  $(\theta, \sigma^2)$  para o modelo ARMA misto é

$$(\ddot{y}, \ddot{y}) = \begin{bmatrix} \ddot{y} & \ddot{y}^2 \\ \ddot{y} & \ddot{y} \end{bmatrix} = \begin{bmatrix} \ddot{y} & \ddot{y} \\ \ddot{y} & \ddot{y} \end{bmatrix} \quad (7.2.17)$$

isso é,

$$= \frac{1}{\sqrt{2}} \begin{array}{c|c} \begin{array}{cc} \tilde{y} & (0) \quad (1) \tilde{y} \quad (\tilde{y} \ 1) \quad (0) \tilde{y} \quad (\tilde{y}) \\ (1) & 2) \\ \tilde{y}\tilde{y} \ \tilde{y} \ \tilde{y}\tilde{y} \end{array} & \begin{array}{cc} (0) & (\tilde{y}1) \tilde{y} \quad (1 \ \tilde{y}) \\ (1) & (0) \tilde{y} \quad (2 \ \tilde{y}) \end{array} \\ \hline \begin{array}{cc} (-1) & (\tilde{y} \ 2) \tilde{y} \quad (0) \\ (0) & (1) \tilde{y} \quad (\tilde{y} \ 1) \quad (0) \tilde{y} \quad (\tilde{y}) \\ (\tilde{y}1) & 2) \\ \tilde{y}\tilde{y} \ \tilde{y} \ \tilde{y}\tilde{y} \end{array} & \begin{array}{cc} (-1) & (\tilde{y} \ 2) \tilde{y} \quad (\tilde{y}) \\ (0) & (1) \tilde{y} \quad (\tilde{y} \ 1) \\ (1) & (0) \tilde{y} \quad (\tilde{y} \ 2) \end{array} \\ \hline \begin{array}{cc} \tilde{y} & (1 \ \tilde{y}) \quad (2 \ \tilde{y}) \tilde{y} \quad (\tilde{y}) \end{array} & \begin{array}{cc} (-1) & (\tilde{y} \ 2) \tilde{y} \quad (0) \end{array} \end{array} \quad \tilde{y} \quad (7.2.18)$$

## 234 ESTIMATIVA DE PARÂMETROS

onde  $\gamma(k)$  e  $\gamma(k, l)$  são as autocovariâncias para os  $y_t$ 's e os  $y_t$ 's, e  $\gamma(k, l)$  são as covariâncias cruzadas definidas por

$$\gamma(k, l) = \text{Cov}(y_{t+k}, y_{t+l}) = \text{Cov}(y_t, y_{t-l+k})$$

A matriz de covariância de amostra grande para as estimativas de máxima verossimilhança pode ser obtida usando

$$V(\hat{\theta}) = \frac{1}{n} \text{Cov}(\hat{\theta})$$

As estimativas de  $\hat{\theta}$  e, portanto, de  $V(\hat{\theta})$  podem ser obtidas avaliando os  $y_t$ 's e  $\gamma(k)$  com  $\gamma(k) = \frac{1}{n} \sum_{t=1}^n (y_{t+k} - \bar{y})(y_t - \bar{y})$  e omitindo o sinal de expectativa em (7.2.17) levando a  $\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^n (y_{t+k} - \bar{y})(y_t - \bar{y})$  substituindo as estimativas de amostra padrão das autocovariâncias e covariâncias cruzadas em (7.2.18). Resultados teóricos de grandes amostras podem ser obtidos observando que, com os elementos de  $\hat{\gamma}(k)$  iguais aos valores verdadeiros dos parâmetros, as equações (7.2.13) e (7.2.14) implicam que as séries derivadas e seguem processos *autoregressivos* definidos por

$$\hat{y}_t = \phi_1 \hat{y}_{t-1} + \epsilon_t$$

Segue que as autocovariâncias que aparecem em (7.2.18) são aquelas para autoregressividade pura processos, e as covariâncias cruzadas são o negativo daqueles entre dois desses processos gerado pelos mesmos  $\epsilon_t$ 's.

Ilustramos o uso deste resultado com alguns exemplos.

**Matriz de Covariância de Estimativas de Parâmetros para Processos AR(p) e MA(q).** Deixar  $\gamma(k)$  seja a matriz de autocovariância  $n \times n$  de  $p$  observações sucessivas de um processo AR( $p$ ) com parâmetros  $\theta$ . Então, usando (7.2.18), a matriz de covariância  $n \times n$  de as estimativas  $\hat{\theta}$  é dado por

$$V(\hat{\theta}) = \frac{1}{n} \text{Cov}(\hat{\theta}) = \frac{1}{n} \text{Cov}(\hat{\theta}) \quad (7.2.19)$$

Seja  $\gamma(k)$  a matriz de autocovariância  $n \times n$  de observações sucessivas de um AR( $p$ ) processo com parâmetros  $\theta$ . Então, usando (7.2.18), a covariância  $n \times n$  matriz das estimativas  $\hat{\theta}$  em um modelo MA( $q$ ) é

$$V(\hat{\theta}) = \frac{1}{n} \text{Cov}(\hat{\theta}) = \frac{1}{n} \text{Cov}(\hat{\theta}) \quad (7.2.20)$$

**Covariâncias para os Zeros de um Processo ARMA.** É ocasionalmente útil parametrizar um processo ARMA em termos dos zeros de  $\phi(z)$  e  $\theta(z)$ . Neste caso, um procedimento particularmente simples forma é obtida para a matriz de covariância das estimativas dos parâmetros.

Considere o processo ARMA( $p, q$ ) parametrizado em termos de seus zeros (supostamente real e distinto), de modo que

$$\hat{y}_t = \sum_{i=1}^p \phi_i \hat{y}_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t$$

ou

$$= \ddot{y}_{=1} (1 - \ddot{y}_{=1}) \ddot{y}_{=1} (1 - \ddot{y}_{=1}) \ddot{y}^1$$

As derivadas de  $\ddot{y}$ 's são tais que

$$\ddot{y} = \ddot{y} - (1 - \ddot{y}) \ddot{y}^1$$

$$\ddot{y} - \ddot{y} = \ddot{y} (1 - \ddot{y}) \ddot{y}^1$$

Portanto, usando (7.2.18), para grandes amostras, a matriz de informação para as raízes é tal que

**1I(G,H)**

$$= \begin{pmatrix} \ddot{y} (1 - \ddot{y}_1) \ddot{y}^1 & (1 - \ddot{y}_{12}) \ddot{y}^1 & \ddot{y} (1 - \ddot{y}_1) \ddot{y}^1 \\ \ddot{y} \ddot{y} \ddot{y} & & \ddot{y} \ddot{y} \ddot{y} \\ (1 - \ddot{y}_1) \ddot{y}^1 & (1 - \ddot{y}_2) \ddot{y}^1 & \ddot{y} (1 - \ddot{y}_2) \ddot{y}^1 \\ \ddot{y} (1 - \ddot{y}_{11}) \ddot{y}^1 & \ddot{y} (1 - \ddot{y}_{21}) \ddot{y}^1 & \ddot{y} \ddot{y} (1 - \ddot{y}) \ddot{y}^1 \\ \ddot{y} \ddot{y} \ddot{y} & & \ddot{y} \ddot{y} \ddot{y} \\ \ddot{y} (1 - \ddot{y}_1) \ddot{y}^1 & \ddot{y} (1 - \ddot{y}_2) \ddot{y}^1 & \ddot{y} \ddot{y} (1 - \ddot{y}) \ddot{y}^1 \\ (1 - \ddot{y}_1) \ddot{y}^1 & \ddot{y} (1 - \ddot{y}_2) \ddot{y}^1 & \ddot{y} (1 - \ddot{y}_2) \ddot{y}^1 \end{pmatrix} \quad (7.2.21)$$

**Exemplos:** Para um processo AR(2)  $(1 - \ddot{y}_1)(1 - \ddot{y}_2)$  = , temos

$$= \frac{1}{(1 - \ddot{y}_1)(1 - \ddot{y}_2)} \begin{pmatrix} \ddot{y} (1 - \ddot{y}_1) \ddot{y}^1 & (1 - \ddot{y}_2) \ddot{y}^1 \\ \ddot{y} (1 - \ddot{y}_1) \ddot{y}^1 & (1 - \ddot{y}_2) \ddot{y}^1 \\ \ddot{y} (1 - \ddot{y}_1) \ddot{y}^1 & (1 - \ddot{y}_2) \ddot{y}^1 \\ \ddot{y} (1 - \ddot{y}_1) \ddot{y}^1 & (1 - \ddot{y}_2) \ddot{y}^1 \end{pmatrix} \quad (7.2.22)$$

Resultados exatamente paralelos serão obtidos para um processo de média móvel de segunda ordem.

Da mesma forma, para o processo ARMA(1,1)  $(1 - \ddot{y})(1 - \ddot{y}) = (1 - \ddot{y})$ , na configuração =  $\ddot{y}_1$  e em (7.2.21), obtemos

$$= \frac{1}{(1 - \ddot{y})^2} \begin{pmatrix} \ddot{y} (1 - \ddot{y}) \ddot{y}^1 & \ddot{y} (1 - \ddot{y}) \ddot{y}^1 \\ \ddot{y} (1 - \ddot{y}) \ddot{y}^1 & \ddot{y} (1 - \ddot{y}) \ddot{y}^1 \\ \ddot{y} (1 - \ddot{y}) \ddot{y}^1 & \ddot{y} (1 - \ddot{y}) \ddot{y}^1 \\ \ddot{y} (1 - \ddot{y}) \ddot{y}^1 & \ddot{y} (1 - \ddot{y}) \ddot{y}^1 \end{pmatrix} \quad (7.2.23)$$

Os resultados para esses dois processos ilustram uma propriedade de dualidade entre as informações matrizes para o modelo autoregressivo e o modelo geral ARMA( , ). Ou seja, suponha

## 236 ESTIMATIVA DE PARÂMETROS

que a matriz de informação para os parâmetros  $(\mathbf{G}, \mathbf{H})$  do modelo ARMA( , )

$$\ddot{y}_{=1} (1 - \ddot{y}) = \ddot{y}_{=1} (1 - )$$

é denotado como  $I\{ , \}(\ , )$ , e suponha, correspondentemente, que a matriz de informação para os parâmetros  $(\mathbf{G}, \mathbf{H})$  no modelo AR(+) puro

$$\ddot{y}_{=1} (1 - \ddot{y}) \ddot{y}_{=1} (1 - \ddot{y}) =$$

é indicado como

$$\{ , \}(\ , ) \ddot{y}_{=1} (1 - \ddot{y}) \ddot{y}_{=1} (1 - \ddot{y}) = [ \ddot{y}_{=1} (1 - \ddot{y}) \ddot{y}_{=1} (1 - \ddot{y}) ]$$

onde a matriz é particionada após a <sup>a</sup> linha e coluna. Em seguida, para moderados e grandes amostras, podemos ver diretamente de (7.2.21) que

$$\{ , \}(\ , ) \ddot{y}_{=1} (1 - \ddot{y}) \ddot{y}_{=1} (1 - \ddot{y}) = [ \ddot{y}_{=1} (1 - \ddot{y}) \ddot{y}_{=1} (1 - \ddot{y}) ] \quad (7.2.24)$$

Assim, uma vez que para amostras moderadas e grandes, o inverso da matriz de informação fornece uma aproximação próxima à matriz de covariância ( ) das estimativas dos parâmetros, temos tem, correspondentemente,

$$\{ , \}(\ , ) \ddot{y}_{=1} (1 - \ddot{y}) \ddot{y}_{=1} (1 - \ddot{y}) = [ \ddot{y}_{=1} (1 - \ddot{y}) \ddot{y}_{=1} (1 - \ddot{y}) ] \quad (7.2.25)$$

### 7.3 ALGUNS RESULTADOS DE ESTIMATIVA PARA MODELOS ESPECÍFICOS

Nos Apêndices A7.3, A7.4 e A7.5, alguns resultados de estimativas para casos especiais são derivados. Estes, e os resultados obtidos anteriormente neste capítulo, são resumidos aqui para referência.

#### 7.3.1 Processos Autoregressivos

É possível obter estimativas dos parâmetros de um processo autoregressivo puro por resolvendo *certas* equações lineares. Mostramos no Apêndice A7.4:

1. Como as estimativas exatas de mínimos quadrados podem ser obtidas resolvendo um sistema linear de equações (veja também a Seção 7.5.3).
2. Como, por ligeira modificação dos coeficientes nestas equações, pode ser obtida uma aproximação das equações de máxima verossimilhança exatas.
3. Como as estimativas condicionais de mínimos quadrados, conforme definido na Seção 7.1.3, podem ser obtidas resolvendo um sistema de equações lineares da forma da regressão linear padrão modelar equações normais.



4. Como as estimativas que são aproximações das estimativas dos mínimos quadrados e das estimativas de máxima verossimilhança podem ser obtidas usando as autocorrelações estimadas como coeficientes nas equações lineares de Yule-Walker.

As estimativas obtidas no item 1 são, obviamente, idênticas às fornecidas por minimização de ( ), conforme descrito em termos gerais na Seção 7.2. As estimativas em 4 são as conhecidas aproximações devido a Yule e Walker. Eles são úteis como primeiras estimativas na fase de identificação, mas pode diferir sensivelmente das estimativas 1, 2 ou 3, em alguns casos. Por exemplo, podem ocorrer diferenças para um modelo AR(2) se o parâmetro  $\rho_1$  e  $\rho_2$  são altamente correlacionados, como é o caso do modelo AR(2) ajustado à Série E na Tabela 7.6.

**Yule-Walker Estimativas.** As estimativas de Yule-Walker (6.3.6) são

$$\hat{\rho}_1 = \hat{\gamma}_1$$

Onde

$$\mathbf{R} = \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \hat{\gamma}_2 \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \hat{\gamma}_1 \\ \hat{\gamma}_2 & \hat{\gamma}_1 & \hat{\gamma}_0 \end{pmatrix} \quad \mathbf{r} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_1 \end{pmatrix} \quad (7.3.1)$$

Em particular, as estimativas para os processos AR(1) e AR(2) são

$$\begin{aligned} \text{AR(1)} \quad \hat{\rho}_1 &= \hat{\gamma}_1 \\ \text{AR(2)} \quad \hat{\rho}_1 &= \frac{1(\hat{\gamma}_1 \hat{\gamma}_2)}{1 - \hat{\gamma}_1^2} \quad \hat{\rho}_2 = \frac{\hat{\gamma}_2}{1 - \hat{\gamma}_1^2} \end{aligned} \quad (7.3.2)$$

É mostrado no Apêndice A7.4 que uma aproximação para  $(\hat{\gamma})$  é fornecida por

$$(\hat{\gamma}) = \hat{\gamma} \begin{pmatrix} \hat{\gamma}_2 \\ 1 \end{pmatrix} \quad (7.3.3)$$

de modo a

$$\hat{\gamma}_2 = \frac{(\hat{\gamma})}{\hat{\gamma}_1} = 0(1 \hat{\gamma} \hat{\gamma}_1) \quad (7.3.4)$$

onde  $\hat{\gamma}_1$  é a variância amostral de 's. Uma expressão paralela relaciona a variância teórica dos  $\gamma_2$  e  $\gamma_1$ , a 's [ver (3.2.8)], ou seja,

$$\gamma_2 = 0(1 \gamma \gamma_1)$$

onde os elementos de  $\gamma$  e de  $\gamma_1$  são os valores teóricos. Assim, de (7.2.19) e Apêndice A7.5, a matriz de covariância para as estimativas é

$$(\hat{\gamma}) \hat{\gamma} \begin{pmatrix} \hat{\gamma}_1 & \hat{\gamma}_2 \\ \hat{\gamma}_2 & \hat{\gamma}_1 \end{pmatrix} = \hat{\gamma}_1 (1 \hat{\gamma} \hat{\gamma}_1) \quad (7.3.5)$$

onde  $\gamma = (1 \gamma 0)$  são as matrizes de autocovariância e autocorrelação de valores sucessivos do processo AR( ).

## 238 ESTIMATIVA DE PARÂMETROS

Em particular, para os processos AR(1) e AR(2), descobrimos que

$$AR(1) \quad \hat{\gamma} = \hat{\gamma}_1(1 - \hat{\gamma}_1) \quad (7.3.6)$$

$$AR(2) \quad \hat{\gamma} = \begin{pmatrix} \hat{\gamma}_1 & \hat{\gamma}_2 \end{pmatrix} \quad \hat{\gamma}_1 \begin{bmatrix} 1 - \frac{\hat{\gamma}_2}{2} & \hat{\gamma}_2 \\ \hat{\gamma}_2 & 1 - \frac{\hat{\gamma}_2}{2} \end{bmatrix} \quad (7.3.7)$$

As estimativas das variâncias e covariâncias são obtidas substituindo as estimativas dos parâmetros em (7.3.5). Desta forma,

$$V(\hat{\gamma}) = \hat{\gamma}_1(1 - \hat{\gamma}_1) \hat{\gamma}_1 \quad (7.3.8)$$

Usando (7.3.7) é facilmente mostrado que a correlação entre as estimativas do AR(2) parâmetros é aproximadamente igual a  $\hat{\gamma}_2$ . Isso implica, em particular, que uma grande correlação lag 1 na série pode dar origem a estimativas instáveis, o que pode explicar o diferenças entre o Yule-Walker e as estimativas de mínimos quadrados mencionadas acima.

### 7.3.2 Processos de média móvel

As estimativas de máxima verossimilhança para processos de média móvel podem, em casos simples, ser obtido graficamente, conforme ilustrado na Seção 7.1.6, ou mais geralmente, pelo iterativo cálculo descrito na Seção 7.2.1. De (7.2.20), segue que para moderado e grande amostras, a matriz de covariância para as estimativas dos parâmetros de um movimento de ordem  $m$  processo médio é da mesma forma que a matriz correspondente para um processo autoregressivo da mesma ordem. Assim, para os processos MA(1) e MA(2), encontramos, correspondentes para (7.3.6) e (7.3.7)

$$MA(1) \quad V(\hat{\gamma}) = \hat{\gamma}_1(1 - \hat{\gamma}_1) \quad (7.3.9)$$

$$MA(2) \quad V(\hat{\gamma}) = \begin{pmatrix} \hat{\gamma}_1 & \hat{\gamma}_2 \end{pmatrix} \quad \hat{\gamma}_1 \begin{bmatrix} 1 - \frac{\hat{\gamma}_2}{2} & \hat{\gamma}_2 \\ \hat{\gamma}_2 & 1 - \frac{\hat{\gamma}_2}{2} \end{bmatrix} \quad (7.3.10)$$

### 7.3.3 Processos Mistos

Estimativas de máxima verossimilhança  $(\hat{\gamma}, \hat{\gamma})$  para processos mistos, como para processos de média móvel, pode ser obtido graficamente em casos simples e, mais geralmente, por cálculo iterativo. Para amostras moderadas e grandes, a matriz de covariância pode ser obtida avaliando e invertendo a matriz de informação (7.2.18). No importante caso especial do ARMA(1, 1) processo

$$(1 - \hat{\gamma}_1 B) X_t = (1 - \hat{\gamma}_2 B) a_t$$

obtemos, como em (7.2.23),

$$V(\hat{\gamma}, \hat{\gamma}) = \hat{\gamma}_1 \frac{1 - \hat{\gamma}_1}{(1 - \hat{\gamma}_1)^2 + (1 - \hat{\gamma}_2)^2 + (1 - \hat{\gamma}_1)(1 - \hat{\gamma}_2)} \quad (7.3.11)$$

Observe que quando  $\hat{\gamma}_1 \rightarrow 1$ , as variações de  $\hat{\gamma}_1$  e  $\hat{\gamma}_2$  são infinitos. Isto é esperado, pois neste caso o fator  $(1 - \hat{\gamma}_1) = (1 - \hat{\gamma}_1)$  se cancela em ambos os lados do modelo, o que torna-se

$$=$$

Este é um caso particular de *redundância de parâmetros*, que discutiremos mais adiante na Seção 7.3.5.

### 7.3.4 Separação de componentes lineares e não lineares na estimativa

É ocasionalmente interessante fazer uma análise em que a estimação dos parâmetros de o modelo misto é separado em suas partes básicas lineares e não lineares. Considere o geral modelo misto  $(\hat{y}) = (\hat{y})$ , que escrevemos como  $(\hat{y}) = (\hat{y})_1(\hat{y})$ , ou

$$(\hat{y}) = (\hat{y})_1(\hat{y}) \quad (7.3.12)$$

Onde

$$(\hat{y}) = \hat{y}_1(\hat{y})$$

isso é,

$$(\hat{y}) = (\hat{y})_1(\hat{y}) \quad (7.3.13)$$

Para qualquer conjunto de  $\hat{y}$ 's, os  $\hat{y}$ 's podem ser calculados recursivamente de (7.3.13), o que pode ser escrito como

$$= \hat{y}_1 + \hat{y}_2 + \hat{y}_3 + \dots$$

A recursão pode ser iniciada definindo desconhecido  $\hat{y}$ 's igual a zero. Tendo calculado os  $\hat{y}$ 's, as estimativas condicionais podem ser prontamente obtidas. Estas são as estimativas parâmetros autorregressivos no modelo linear (7.3.12), que podem ser escritos como

$$\hat{y}_1 = \hat{y}_1, \hat{y}_2 = \hat{y}_2, \dots \quad (7.3.14)$$

Conforme discutido na Seção 7.3.1, as estimativas de mínimos quadrados dos parâmetros autorregressivos podem ser encontradas por solução direta de um conjunto de equações lineares. Em casos simples, podemos examinar o comportamento de  $(\hat{y})$  em um argumento sobre seu mínimo calculando  $(\hat{y})$  de valores e contornos de plotagem.

**Exemplo Usando a Série C.** Um modelo possível para a Série C considerado anteriormente é o ARIMA(1, 1, 0) modelo  $(1 - \hat{y}) = \hat{y}$  com  $\hat{y} = 0$ . Considere agora um pouco mais elaborado  $(1 - \hat{y}) = (1 - \hat{y})$  argumento dado acima, o processo pode ser seguido como resultado de uma combinação dos modelo não linear  $\hat{y} = \hat{y}_1 + \hat{y}_2$  e o modelo linear  $\hat{y}_1 = \hat{y}_1$ .

Para cada escolha dos parâmetros não lineares  $\hat{y} = (1, 2)$  dentro da região de invertibilidade, um conjunto de  $\hat{y}$ 's foi calculado recursivamente. Usando a aproximação Yule-Walker, uma estimativa  $\hat{y} = 1()$  agora pode ser obtido junto com

$$(\hat{y}, \hat{y}) = \sum_{i=1}^2 [1 - \hat{y}_i^2(\hat{y})]$$

Esta soma de quadrados foi plotada para uma grade de valores de  $\hat{y}$  e seus contornos são mostrados Figura 7.6. Vemos que um mínimo próximo de  $\hat{y} = 0$  é indicado, portanto

$\hat{y} = 0,805$ . Assim, dentro de toda a classe de modelos de ordem (1, 1, 2), o simples (1, 1, 0) modelo  $(1 - \hat{y}) = \hat{y}$  é confirmado para fornecer uma representação adequada.

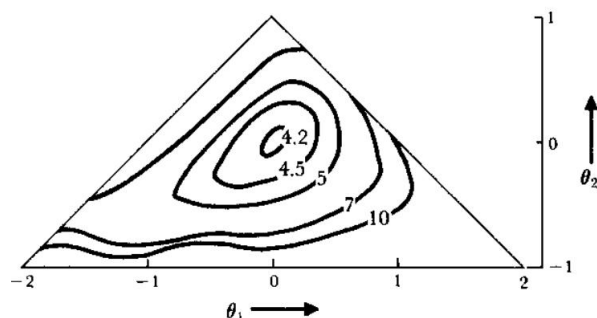


FIGURA 7.6 Contadores de ( os 's. , ) para a Série C plotada sobre o espaço de parâmetros admissível para

### 7.3.5 Redundância de Parâmetros

O modelo ( ) = ( ) é idêntico ao modelo

$$(1 - \tilde{y})(1 - \tilde{y}) = (1 - \tilde{y})(1 - \tilde{y})$$

em que os operadores de média móvel e autorregressivo são multiplicados pelo mesmo fator,  $1 - \tilde{y}$ . Sérias dificuldades no procedimento de estimação surgirão se um modelo for ajustado que contém um fator redundante. Portanto, é necessário cuidado para evitar a situação em que ocorrem fatores *comuns* redundantes ou quase redundantes. A existência de redundância não é sempre óbvio. Por exemplo, pode-se ver o fator comum no modelo ARMA(2, 1)

$$(1 - \tilde{y} - 1,3 + 0,42) \tilde{y} = (1 - \tilde{y} - 0,5)$$

somente depois de fatorar o lado esquerdo para obter

$$(1 - \tilde{y} - 0,5)(1 - \tilde{y} - 0,8) = (1 - \tilde{y} - 0,5)$$

isto é,  $(1 - \tilde{y} - 0,8) =$ .

Na prática, não é apenas o cancelamento exato que causa dificuldades, mas também quase cancelamento. Por exemplo, suponha que o modelo verdadeiro foi

$$(1 - \tilde{y} - 0,4)(1 - \tilde{y} - 0,8) = (1 - \tilde{y} - 0,5) \quad (7.3.15)$$

Se foi feita uma tentativa de ajustar este modelo como ARMA(2, 1), instabilidade extrema no parâmetro estimativas podem surgir devido ao quase cancelamento dos fatores  $(1 - \tilde{y} - 0,4)$  e  $(1 - \tilde{y} - 0,5)$ , nos lados esquerdo e direito. Neste caso, combinações de valores de parâmetros que produzem semelhantes [ ] e probabilidades semelhantes podem ser encontradas, e uma mudança de valor de parâmetro em a esquerda pode ser quase compensada por uma mudança adequada na direita. A soma dos quadrados superfícies de contorno no espaço de parâmetros tridimensional se aproximarão obliquamente cilindros orientados e uma linha de soluções de "quase mínimos quadrados" em vez de um ponto mínimo será encontrado.

De um ponto de vista ligeiramente diferente, podemos escrever o modelo (7.3.15) em termos de um operador autoregressivo infinito. Fazendo a expansão necessária, encontramos que

$$(1 - \tilde{y} - 0,700 \tilde{y} - 0,030 \quad 2 - 0,015 \quad 3 - 0,008 \quad 4 - \tilde{y} - \tilde{y}) =$$

Assim, muito próximo, o modelo é

$$(1 - \hat{y} \ 0,7) = \quad (7.3.16)$$

A instabilidade das estimativas, obtidas ao tentar ajustar um modelo ARMA(2, 1), ocorrer porque estaríamos tentando encaixar três parâmetros em uma situação que quase poderia ser representado por um.

A principal razão para passar pelo procedimento de identificação antes de encaixar o modelo é evitar dificuldades decorrentes da redundância de parâmetros e obter *parcimônia* na parametrização.

**Redundância no Modelo ARMA(1,1).** O modelo mais simples onde ocorre a possibilidade para cancelamento direto de fatores é o processo ARMA(1, 1):

$$(1 - \hat{y}) = (1 - \hat{y})$$

Em particular, se  $\theta = 0$  de modo que  $\theta$  é qualquer valor comum que eles tenham,  $\theta = 0$ , gerado por um processo de ruído branco. Os dados, então, não podem fornecer informações sobre o parâmetro comum, e usando (7.3.11), e tem variâncias infinitas. Além disso, quaisquer que sejam os valores de  $\phi$  e  $\theta$ ,  $\theta = 0$  deve ser constante na linha  $\theta = 0$  ilustrado na Figura 7.7, que mostra um gráfico de soma de quadrados para os dados da Série A. No entanto, para esses dados, os valores de mínimos quadrados  $\theta = 0,92$  correspondem a um ponto que não está particularmente próximo da linha  $\theta = 0$  e não ocorrem dificuldades no estimativa desses parâmetros.

Na prática, se for adotada a técnica de identificação que recomendamos, essas dificuldades serão evitadas. Um processo ARMA(1, 1) no qual  $\theta$  é quase igual a normalmente será identificado como ruído branco, ou se a diferença não for desprezível, como um AR(1) ou processo MA(1) com um único coeficiente pequeno.

Resumindo:

1. Devemos evitar modelos mistos contendo fatores quase comuns e devemos ser alerta para as dificuldades que podem resultar.

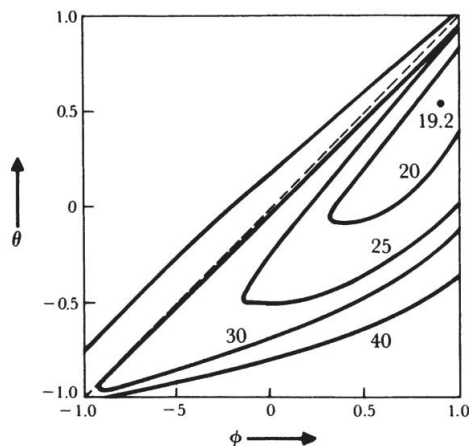


FIGURA 7.7 Gráfico de soma de quadrados para Série A.

## 242 ESTIMATIVA DE PARÂMETROS

2. Evitaremos automaticamente esses modelos se usarmos identificação e estimativa procedimentos de forma inteligente.

### 7.4 FUNÇÃO DE PROBABILIDADE COM BASE NO MODELO DE ESTADO-ESPAÇO

Na Seção 5.5, introduzimos a formulação do modelo de espaço de estados do processo ARMA junto com com filtragem de Kalman e descreveu seu uso para previsão. Essa abordagem também fornece uma método conveniente para avaliar a função de verossimilhança exata para um modelo ARMA. O uso dessa abordagem foi sugerido por Jones (1980), Gardner et al. (1980), e outros.

A forma de modelo de espaço de estados do modelo ARMA( , ) dada na Seção 5.5 é

$$\mathbf{y}_t = \mathbf{H} \mathbf{x}_t + \mathbf{e}_t \quad (7.4.1)$$

Onde  $\mathbf{y}_t = (y_t, \dots, y_t)^T$ ,  $\mathbf{H} = (1, 0, \dots, 0)$ ,  $\mathbf{e}_t = (e_t, \dots, e_t)^T$ ,  $\mathbf{e}_t \sim N(0, \mathbf{V})$ ,  $\mathbf{V} = \text{diag}(V_1, \dots, V_p)$ ,  $\mathbf{V}_1 = \sigma^2$ ,  $\mathbf{V}_2 = \dots = \mathbf{V}_p = 0$ .

$$\mathbf{V} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

$\mathbf{e}_t = (e_t, \dots, e_t)^T$ . As equações do filtro de Kalman (5.5.6)–(5.5.9) fornecem previsões um passo à frente  $\hat{\mathbf{x}}_{t+1|t}$  e a covariância de erro  $\mathbf{P}_{t+1|t}$ .

matriz  $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_p]$ . Especificamente, para a forma de espaço de estado do modelo ARMA(1,1) as equações recorrentes são

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{A} \hat{\mathbf{x}}_{t|t-1} + \mathbf{B} \mathbf{y}_t + \mathbf{K} (\mathbf{y}_t - \mathbf{H} \hat{\mathbf{x}}_{t|t-1}) \quad \mathbf{P}_{t+1|t} = \mathbf{A} \mathbf{P}_{t|t-1} \mathbf{A}^T + \mathbf{B} \mathbf{V}_1 \mathbf{B}^T + \mathbf{K} \mathbf{V}_1 \mathbf{K}^T \quad (7.4.2)$$

Onde  $\hat{\mathbf{x}}_{t|t-1} = \mathbf{H} \hat{\mathbf{x}}_{t|t-1}$ ,  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{x}}_{t|t-1}$ .

$$\mathbf{P}_{t|t-1} = \mathbf{P}_{t-1|t-1} - \mathbf{K} \mathbf{V}_1 \mathbf{K}^T \quad \mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} + \mathbf{K} \mathbf{V}_1 \mathbf{K}^T \quad (7.4.3)$$

com

$$\mathbf{E}[\mathbf{e}_t \mathbf{e}_t^T] = \mathbf{V}_1 \quad (7.4.4)$$

para  $t = 1, 2, \dots$ . Em particular, então, o primeiro componente do vetor de previsão é

$$\hat{y}_{t+1|t} = \mathbf{H} \hat{\mathbf{x}}_{t+1|t} = \mathbf{H} \mathbf{A} \hat{\mathbf{x}}_{t|t-1} + \mathbf{H} \mathbf{B} \mathbf{y}_t + \mathbf{H} \mathbf{K} (\mathbf{y}_t - \mathbf{H} \hat{\mathbf{x}}_{t|t-1})$$

ance.

Para obter a função de verossimilhança exata do vetor de  $(y_1, y_2, \dots, y_T)$  usando observações  $\mathbf{y}_t = (y_t, \dots, y_t)^T$  os resultados acima, notamos que a distribuição conjunta de  $\mathbf{w}$  pode ser fatorado como

$$p(\mathbf{w} | \mathbf{y}) = \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{w}_{t-1}) \quad (7.4.5)$$

onde  $(\tilde{y}_1, \dots, \tilde{y}_T)$ , dado  $(\tilde{y}_1, \dots, \tilde{y}_{t-1})$ , denota a distribuição condicional de  $\tilde{y}_t, \dots, \tilde{y}_T$ . Sob normalidade desta distribuição condicional é normal com condicional significância a distribuição conjunta de  $\tilde{y}_1, \dots, \tilde{y}_T$  pode ser convenientemente expressa como

$$(em \tilde{y}_1, \dots, \tilde{y}_T) = \tilde{y}_1' \tilde{y}_2' \exp \left[ -\frac{1}{2} \tilde{y}_1' \tilde{y}_2' \right] \quad (7.4.6)$$

onde as quantidades são facilmente determinadas recursivamente a partir do Kalman procedimento de filtro. Os valores iniciais necessários para iniciar as recursões do filtro de Kalman são fornecidos por  $\tilde{y}_0 = 0$ , um vetor dimensional de zeros, e  $V_0 = cov[\tilde{y}_0]$ . Os elementos de  $V_0$  pode ser facilmente determinado em função das autocovariâncias e dos pesos de o processo ARMA( $p, q$ ), fazendo uso da relação do Capítulo 5. Ver Jones (1980) para mais detalhes. Por exemplo, no caso de um Modelo ARMA(1, 1) para  $\tilde{y}_t$ , temos

$$\tilde{y}_0 = 0, \quad \tilde{y}_1 = 0, \quad \tilde{y}_2 = 2\tilde{y}_1$$

Também é geralmente o caso que as previsões um passo à frente  $\tilde{y}_{t+1|t}$  e o correspondente variâncias de erro  $\tilde{y}_{t+1|t}^2$  aproximam-se rapidamente de suas formas de estado estacionário, caso em que o Os cálculos do filtro de Kalman em algum estágio (além do tempo 0, digamos) podem ser alterados para o forma mais simples  $\tilde{y}_{t+1|t} = \tilde{y}_t$  para  $\tilde{y}_{t+1|t}^2 = \tilde{y}_t^2$ , onde  $\tilde{y}_t$  é a variância de  $\tilde{y}_t$ . Na comparação de (7.4.6) com expressões dadas anteriormente em (7.1.5) e (7.1.6), e também (A7.3.11) e (A7.3.13), a função soma de quadrados incondicional pode ser representada em duas formas equivalentes como

$$(\tilde{y}_1, \dots, \tilde{y}_T) = \tilde{y}_1' \tilde{y}_2' \exp \left[ -\frac{1}{2} \tilde{y}_1' \tilde{y}_2' \right]$$

Onde  $\tilde{y}_1 = \tilde{y}_1$  e também  $\tilde{y}_2 = \tilde{y}_1$ .

**Método de Inovações.** A função de verossimilhança expressa na forma de (7.4.6) é geralmente referido como o *formulário de inovações* e as quantidades  $\tilde{y}_1 = 1, \dots, \tilde{y}_T$  são as *inovações* (de amostra finita). Cálculo da função de verossimilhança neste formulário, baseado na representação em espaço de estados do processo ARMA e Kalman associado algoritmos de filtragem, tem sido proposto por muitos autores, incluindo Gardner et al. (1980), Harvey e Phillips (1979) e Jones (1980). A forma de inovação da probabilidade pode também pode ser obtido sem usar diretamente a representação no espaço de estados através do uso de um "algoritmo de inovações" (por exemplo, ver Ansley, 1979; Brockwell e Davis, 1991). este método envolve essencialmente uma decomposição de Cholesky de uma matriz de covariância  $\times$  banda do processo MA( $q$ ) derivado:

$$\tilde{y}_1 = 1, \quad \tilde{y}_2 = 2, \quad \dots, \quad \tilde{y}_T = T$$

Mais especificamente, usando a notação do Apêndice A7.3, escrevemos o modelo ARMA relações para observações como  $\tilde{y}_t = \tilde{y}_t + \tilde{y}_t$ , onde  $\tilde{y}_t = (1, 2, \dots, T)$  e  $\tilde{y}_t =$

## 244 ESTIMATIVA DE PARÂMETROS

$(1\tilde{y}, \dots, 0, 1\tilde{y}, \dots, 0)$  é o vetor  $(+)$ -dimensional dos valores pré-amostra. Então, a matriz de covariância do vetor de variáveis derivadas  $\mathbf{L} \mathbf{w}$  é

$$\tilde{\mathbf{y}} = \text{cov}[\mathbf{L} \mathbf{w}] = \text{cov}[\mathbf{L} \mathbf{a} + \mathbf{F} \tilde{\mathbf{y}}] = (\mathbf{L} \mathbf{L}' + \mathbf{F} \mathbf{F}') \quad (7.4.7)$$

que é uma *matriz de bandas*. Ou seja, é uma matriz com elementos não nulos apenas em uma banda sobre a diagonal principal da largura de banda máxima  $= \max(\dots)$ , e da largura de banda após as primeiras linhas desde  $\text{cov}[\dots] = 0$  para  $>$ . O algoritmo de inovações obtém a

(sem raiz quadrada) Decomposição de Cholesky da matriz de banda  $\mathbf{L} \mathbf{L}' + \mathbf{F} \mathbf{F}'$  como  $\mathbf{G} \mathbf{D} \mathbf{G}'$ , onde  $\mathbf{G}$  é uma matriz de banda triangular inferior com largura de banda correspondente àquela de e com uns na diagonal, e  $\mathbf{D}$  é uma matriz diagonal com elementos diagonais positivos,  $= 1, \dots$ . Portanto,  $\text{cov}[\mathbf{w}] = \mathbf{G} \mathbf{D} \mathbf{G}'$  e a forma quadrática no expoente da função de verossimilhança (7.4.6) é

$$\begin{aligned} \mathbf{w}' \{\text{cov}[\mathbf{w}]\}^{-1} \mathbf{w} &= \frac{1}{2} \mathbf{w}' (\mathbf{L}' \mathbf{L} + \mathbf{F}' \mathbf{F})^{-1} \mathbf{w} \\ &= \frac{1}{2} \mathbf{e}' \mathbf{D}^{-1} \mathbf{e} = \frac{1}{2} \sum_{i=1}^T \tilde{y}_i^2 \end{aligned} \quad (7.4.8)$$

onde  $\mathbf{e} = \mathbf{G}' \mathbf{L}' \mathbf{w} = (1|0, 2|1, \dots)$  é o vetor de inovações, que são computado recursivamente de  $\mathbf{G} \mathbf{e} = \mathbf{L} \mathbf{w}$ . Assim, as inovações podem ser obtidas

$$\text{recursivamente como } 1, 2|1 = 2^{-1} (1 + 1, 1|0, \dots) = \tilde{y}_1 | \tilde{y}_1 + \dots$$

$$\tilde{y}_1 = 1, \quad \tilde{y}_1 | \tilde{y}_1, \quad \mathbf{e}$$

$$|\tilde{y}_1| = \tilde{y}_1^2 + \tilde{y}_1^2, \quad \tilde{y}_1 | \tilde{y}_1 \quad (7.4.9)$$

para  $>$ , onde a  $t$ -ésima linha da matriz  $\mathbf{G}$  tem a forma

$$[0, \dots, 0, \tilde{y}_1, \dots, 1, 0, \dots, 0]$$

com o 1 na  $a$  posição (ou seja, diagonal). Além disso, os coeficientes em (7.4.9) e os elementos diagonais (variância) são obtidos recursivamente através do método de Cholesky procedimento de decomposição. Em particular, os são dados pela recursão

$$= \frac{0(\tilde{y})}{2} \tilde{y}_1^2, \quad \tilde{y}_1^2 \quad \text{para } > \quad (7.4.10)$$

$$\text{onde } 0(\tilde{y}) = \text{var}(\tilde{y}) = 1 + \tilde{y}_1^2.$$

de espaço de estados "inovações" para avaliar a função de verossimilhança exata tem também se mostrou bastante útil para lidar com problemas de estimação para modelos ARMA quando a série tem valores omissos; ver, por exemplo, Jones (1980), Harvey e Pierse (1984) e Wincek e Reinsel (1986).

A função de verossimilhança exata calculada usando a abordagem de filtragem de Kalman pode ser maximizado usando algoritmos de otimização numérica. Estes normalmente requerem a primeira parcial derivadas da probabilidade logarítmica em relação aos parâmetros desconhecidos, e muitas vezes é benéfico usar derivados analíticos. Da forma da verossimilhança em (7.4.6), é visto que isso envolve a obtenção de derivadas parciais das previsões de uma etapa  $|\tilde{y}_1$  e



das variações de erro  $\sigma^2$  para cada  $i = 1, \dots, n$ . Wincek e Reinsel (1986) mostram como as derivadas exatas de  $\hat{\beta}$  e  $\text{var}[\hat{\beta}]$  em relação ao modelo  $\beta$  podem ser obtidos recursivamente a partir das equações de atualização e equações de previsão. Isso, por sua vez, leva a uma forma explícita de cálculos iterativos para a estimativa de máxima verossimilhança associada à verossimilhança (7.4.6), semelhante aos procedimentos de mínimos quadrados não lineares detalhados na Seção 7.2.

## 7.5 ESTIMATIVA UTILIZANDO O TEOREMA DE BAYES

### 7.5.1 Teorema de Bayes

Nesta seção, novamente usamos o símbolo  $\beta$  para representar um vetor geral de parâmetros. Bayes' teorema nos diz que se  $\pi(\beta)$  é a distribuição de probabilidade para antes da coleta de dados, então  $\pi(\beta|y)$ , a distribuição de posterior aos dados é obtida combinando os distribuições prévias  $\pi(\beta)$  e a probabilidade  $\pi(y|\beta)$  da seguinte forma:

$$\pi(\beta|y) = \frac{\pi(y|\beta)\pi(\beta)}{\int \pi(y|\beta)\pi(\beta) d\beta} \quad (7.5.1)$$

O denominador apenas garante que  $\int \pi(\beta|y) d\beta = 1$ . A parte importante da expressão é o numerador, a partir do qual vemos que a distribuição posterior é proporcional à distribuição anterior multiplicada pela probabilidade. Savage (1962) mostrou que antes e probabilidades posteriores podem ser interpretadas como probabilidades subjetivas. Em particular, muitas vezes antes que os dados estejam disponíveis, temos muito pouco conhecimento e estaríamos preparados para concordar que na região relevante, teria parecido a priori tão provável que tinha um valor como outro. Neste caso,  $\pi(\beta)$  poderia ser tomado como *localmente* uniforme, e portanto  $\pi(\beta|y)$  seria proporcional à probabilidade.

Deve-se notar que, para que esse argumento seja válido, não é necessário que a densidade anterior de ser uniforme em toda a sua faixa (que para alguns parâmetros pode ser infinita). Por exigindo que seja *localmente uniforme*, queremos dizer que seja aproximadamente uniforme na região em que a probabilidade é apreciável e que não é necessário um grande valor fora dessa região.

Assim, se fosse o peso de uma cadeira, certamente poderíamos dizer a priori que pesava mais de uma onça e menos de uma tonelada. Também é provável que quando obtivemos uma observação de pesando a cadeira em uma máquina de pesagem, que tinha um desvio padrão de erro poderia dizer  $\pm 3$ , nós honestamente que teríamos ficado igualmente felizes com valores a priori na faixa  $\pm 3$ . A exceção seria se a balança dissesse que uma cadeira aparentemente pesada pesava, digamos, 10 onças. Nesse caso, a verossimilhança e a anterior seriam incompatíveis, e não devemos, é claro, usar o teorema de Bayes para combiná-los, mas verificar o balanço e, se isso for preciso, inspecione a cadeira mais de perto.

Há, é claro, alguma arbitrariedade nessa ideia. Suponha que assumimos a distribuição prévia de ser localmente uniforme. Isso implica, então, que a distribuição de qualquer função linear de  $\beta$  também é localmente uniforme. No entanto, a distribuição prévia de alguns transformações não lineares  $\pi(\beta)$  (como  $\log \beta$ ) não pode ser exatamente uniforme localmente. Essa arbitrariedade geralmente terá muito pouco efeito se formos capazes de obter informações bastante precisas.

## 246 ESTIMATIVA DE PARÂMETROS

as estimativas. Estaremos então considerando apenas um pequeno intervalo, e em tal de alcance da transformação de para, digamos, log seriam quase sempre lineares.

Jeffreys (1961) argumentou que é melhor escolher a métrica  $(\cdot)$  para que a da informação  $= \ddot{y} [ \cdot ]$  e, portanto,  $\ddot{y}^2 [ \cdot ]$  é independente do valor da medida  $\cdot$ , de Isso é equivalente a escolher  $(\cdot)$  para que a variância limite de seu máximo estimativa de probabilidade é independente e é alcançada escolhendo a distribuição anterior de ser proporcional a  $\ddot{y}$ .

Jeffreys justificou essa escolha de prior com base em sua invariância à parametrização empregado. Especificamente, com esta escolha, as distribuições a posteriori para  $(\cdot)$  e para onde  $(\cdot)$  e são conectados por uma transformação um-para-um, são tais que  $(\cdot) = (\cdot)$

$\ddot{y}$ . O mesmo resultado pode ser obtido (Box e Tião, 1973) pelo seguinte argumento.

Se para grandes amostras, a função de verossimilhança esperada para  $(\cdot)$  se aproxima de uma curva normal, então a média e a variância da curva resumem as informações a serem esperadas do dados. Suponha, agora, que uma transformação  $(\cdot)$  pode ser encontrada na qual a aproximação curva normal tem variância quase constante, quaisquer que sejam os valores verdadeiros do parâmetro. Então, nesta parametrização, a *única* informação em perspectiva a partir dos dados é veiculada pelo *localização* da função de verossimilhança esperada. Dizer que não sabemos essencialmente nada priori em relação a essa informação observacional prospectiva é dizer que consideramos diferentes *locais* igualmente prováveis a priori. De forma equivalente, dizemos que deve ser tomado como localmente uniforme.

A generalização da regra de Jeffreys para lidar com vários parâmetros é que a priori conjunta distribuição de parâmetros seja proporcional a

$$| \cdot | \ddot{y}^2 = \left| \ddot{y} [ \cdot ] \right| \frac{\ddot{y}^2}{2} \quad (7.5.2)$$

Insistiu-se (por exemplo, Jenkins, 1964) que a probabilidade em si é melhor considerada e traçada nessa métrica para a qual é independente de  $\ddot{y}$ . Se isso for feito, será observado que o a função de verossimilhança e a função de densidade posterior com a priori uniforme são proporcionais.

### 7.5.2 Estimativa Bayesiana de Parâmetros

Consideramos agora a estimação dos parâmetros em um modelo ARIMA a partir de um modelo Bayesiano ponto de vista. É mostrado no Apêndice A7.3 que a probabilidade exata de uma série temporal  $\mathbf{z}$  de length =  $n$  de um processo ARIMA( $p, d, q$ ) é da forma

$$(\cdot, \cdot) = \frac{(\cdot, \cdot)}{\ddot{y}(\cdot, \cdot) \exp[\ddot{y} \frac{(\cdot, \cdot)}{2}]} \quad (7.5.3)$$

Onde

$$(\cdot, \cdot) = \ddot{y} [ \cdot ]^2 + [ \ddot{y} ] \ddot{y} [ \cdot ] \quad (7.5.4)$$

Se não tivermos informações anteriores  $\cdot$ ,  $\cdot$ , ou  $\cdot$ , e uma vez que as informações sobre sobre não fornecer informações sobre  $\cdot$  for sensato, seguindo Jeffreys, empregar um distribuição para  $\cdot$  e do formulário

$$(\cdot, \cdot) \ddot{y} [ \cdot ] \ddot{y}^2 [ \cdot ]$$

Segue que a distribuição posterior é

$$(\beta, \sigma^2 | y) \propto y^{(n+1)} | (\beta, \sigma^2) \propto \frac{1}{\sigma^{2n+2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \beta)^2\right] \quad (7.5.5)$$

Se agora integrarmos (7.5.5) de zero ao infinito em relação à distribuição  $\sigma^2$ , obtemos o exato conjunta posterior dos parâmetros e como

$$(\beta | y) \propto \int_0^\infty \frac{1}{\sigma^{2n+2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \beta)^2\right] d\sigma^2 \quad (7.5.6)$$

### 7.5.3 Processos Autoregressivos

Se segue um processo ARIMA(1, 0), então  $y_t$  segue um processo AR(1) puro. Isto é mostrado no Apêndice A7.4 que para tal processo, os fatores  $|(\beta)| \propto 1/\sigma^2$  e  $(\beta)$ , que em qualquer caso são dominados pelo termo em  $(\beta)$ , essencialmente cancelam. Isso produz notavelmente resultado simples que, dadas as suposições, os parâmetros do processo AR(1) têm a distribuição posterior

$$(\beta | y) \propto \exp\left[-\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \beta)^2\right] \quad (7.5.7)$$

Por este argumento, então, os contornos de soma de quadrados, que são probabilidades aproximadas contornos, são, quando nada se sabe a priori, também contornos de probabilidade posterior.

**Distribuição Conjunta dos Parâmetros Autoregressivos.** É mostrado no Apêndice A7.4 que para o processo AR puro, as estimativas de mínimos quadrados dos  $\beta$ 's que minimizam  $(\beta)$  = são dadas por

$$\hat{\beta} = y_1 \quad (7.5.8)$$

Onde  $y = (1, y_1, \dots, y_n)$ ,

$$= \begin{bmatrix} y_1 & y_2 & \dots & y_n \\ y_2 & y_3 & \dots & y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{n+1} & \dots & y_{n+1} \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \\ y_2 & y_3 & \dots & y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{n+1} & \dots & y_{n+1} \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \\ y_2 & y_3 & \dots & y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{n+1} & \dots & y_{n+1} \end{bmatrix} \quad (7.5.9)$$

e

$$= \begin{bmatrix} y_1 & y_2 & \dots & y_n \\ y_2 & y_3 & \dots & y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{n+1} & \dots & y_{n+1} \end{bmatrix} + \begin{bmatrix} y_1 & y_2 & \dots & y_n \\ y_2 & y_3 & \dots & y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{n+1} & \dots & y_{n+1} \end{bmatrix} + \begin{bmatrix} y_1 & y_2 & \dots & y_n \\ y_2 & y_3 & \dots & y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{n+1} & \dots & y_{n+1} \end{bmatrix} + \begin{bmatrix} y_1 & y_2 & \dots & y_n \\ y_2 & y_3 & \dots & y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{n+1} & \dots & y_{n+1} \end{bmatrix} + \begin{bmatrix} y_1 & y_2 & \dots & y_n \\ y_2 & y_3 & \dots & y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{n+1} & \dots & y_{n+1} \end{bmatrix} \quad (7.5.10)$$

Segue que

$$(\beta) = \frac{1}{2} \sum_{t=1}^n (y_t - \beta)^2 + \frac{1}{2} \sum_{t=1}^n (y_t - \beta)^2 \quad (7.5.11)$$

## 248 ESTIMATIVA DE PARÂMETROS

Onde

$$s^2 = \frac{(\ddot{y})}{n-2} \quad (7.5.12)$$

e

$$(\ddot{y}) = \frac{\ddot{y}\ddot{y}}{n-2} = \frac{11}{11} \cdot \frac{\ddot{y}\ddot{y}}{11} = \frac{11}{11} \cdot \ddot{y}^{-1} \quad (7.5.13)$$

Assim, podemos escrever

$$(\ddot{y}) \ddot{y} [1 + \frac{(\ddot{y} \ddot{y}) \ddot{y}}{2} \frac{(\ddot{y} \ddot{y})}{\ddot{y}} \ddot{y}^2] \quad (7.5.14)$$

Equivalentemente,

$$\ddot{y} \ddot{y} (\ddot{y}) \ddot{y} [1 + \ddot{y} \ddot{y} \frac{\ddot{y} \ddot{y} \ddot{y} (-\ddot{y})}{2} (\ddot{y}) \ddot{y} \ddot{y}^2] \quad (7.5.15)$$

Onde

$$= \frac{2}{n-2} (\ddot{y}) = 2 \quad +1, +1$$

Segue-se que, a posteriori, os parâmetros de um processo autoregressivo têm uma distribuição múltipla (A7.1.13), com  $n-2$  graus de liberdade.

Em particular, para o caso especial  $n=1$ ,  $(\ddot{y} \ddot{y}) \ddot{y}$  é distribuído *exatamente* em um Student distribuição com  $n-1$  graus de liberdade onde, usando os resultados gerais dados acima, e  $\ddot{y}$  são dados por

$$\ddot{y} = \frac{12}{22} \quad \ddot{y} = [1 \quad \ddot{y} \quad 1] \frac{11}{22} \frac{2 \cdot 12}{11 \cdot 22} \ddot{y}^2 \quad (7.5.16)$$

A quantidade  $\ddot{y}$ , para amostras grandes, tende a  $[(1-\ddot{y})\ddot{y}]^{-1}\ddot{y}^2$  e na estrutura da teoria de amostragem é idêntica ao "erro padrão" para amostras grandes para  $\ddot{y}$ . No entanto, ao usar esta e outras expressões semelhantes dentro da estrutura Bayesiana, deve-se lembrar que são os parâmetros (neste caso) que são variáveis aleatórias. Quantidades como  $\ddot{y}$  e  $\ddot{y}^2$ , que são de dados que já ocorreram, são consideradas fixas.

**Aproximação normal.** Para amostras de tamanho  $n > 50$ , nas quais normalmente estamos interessados, a aproximação normal da distribuição é adequada. Assim, muito próximo, tem uma distribuição normal conjunta-variada  $\{\ddot{y}, \ddot{y}^2\}$  tendo vetor médio e variância-covariância matriz  $\ddot{y}^{-1} \ddot{y}^2$ .

**Regiões Bayesianas de Maior Densidade de Probabilidade.** Ao resumir o que a distribuição a posteriori tem a nos dizer sobre a probabilidade de vários valores, é útil indicar uma região de *maior densidade de probabilidade*, chamada abreviadamente de região HPD (Box e Tião, 1965). Uma região Bayesiana  $1-\alpha$  HPD tem as seguintes propriedades:

1. Qualquer ponto de parâmetro dentro da região tem maior densidade de probabilidade do que qualquer ponto fora.
2. A massa de probabilidade posterior total dentro da região é  $1 - \alpha$ .

Como tem uma distribuição múltipla, segue, usando o resultado (A7.1.4), que

$$\Pr\left\{\left(\tilde{y} - \frac{\tilde{y}}{\sqrt{2}}\right) \leq \left(\tilde{y} - \frac{\tilde{y}}{\sqrt{2}}\right) < \left(\tilde{y} + \frac{\tilde{y}}{\sqrt{2}}\right)\right\} = 1 - \alpha \quad (7.5.17)$$

define a região 1 - HPD *exata* para  $\tilde{y}$ . Agora, para  $\tilde{y} > 100$ ,

$$\left(\tilde{y} - \frac{\tilde{y}}{\sqrt{2}}\right) \approx \left(\tilde{y} - \frac{\tilde{y}}{\sqrt{2}}\right)$$

Também,

$$\left(\tilde{y} - \frac{\tilde{y}}{\sqrt{2}}\right) \approx \left(\tilde{y} - \frac{\tilde{y}}{\sqrt{2}}\right) = \frac{1}{2} \tilde{y} \left(1 - \frac{1}{\sqrt{2}}\right)$$

Assim, aproximadamente, a região HPD definida em (7.5.17) é tal que

$$\left(\tilde{y} - \frac{\tilde{y}}{\sqrt{2}}\right) \leq \left(\tilde{y} - \frac{\tilde{y}}{\sqrt{2}}\right) < \left(\tilde{y} + \frac{\tilde{y}}{\sqrt{2}}\right) \quad (7.5.18)$$

que se definirmos  $\tilde{y}^2 = \tilde{y}^2$  é idêntica à região de confiança definida por (7.1.25).

Embora essas regiões aproximadas sejam idênticas, será lembrado que suas interpretação é diferente. Do ponto de vista da teoria da amostragem, dizemos que se uma confiança região é calculada de acordo com (7.1.25), então para cada um de um conjunto de amostras repetidas, um a proporção  $1 - \alpha$  dessas regiões incluirá o ponto de parâmetro verdadeiro. Do Bayesiano ponto de vista, estamos preocupados apenas com a amostra única, que na verdade foi observado. Assumindo a relevância da distribuição a priori não informativa que temos tomada, a região HPD inclui aquela proporção  $1 - \alpha$  da distribuição de probabilidade resultante de  $\tilde{y}$ , dado  $\tilde{y}$ , que tem a maior densidade. Em outras palavras, a probabilidade de que o valor dos quais deu origem aos dados  $\tilde{y}$ , está na região HPD é  $1 - \alpha$ .

Usando (7.5.11), (7.5.12) e (7.5.18), para grandes amostras o valor aproximado de  $1 - \alpha$  Bayesiano A região HPD é delimitada por um contorno para o qual

$$\left(\tilde{y} - \frac{\tilde{y}}{\sqrt{2}}\right) \leq \left(\tilde{y} - \frac{\tilde{y}}{\sqrt{2}}\right) < \left(\tilde{y} + \frac{\tilde{y}}{\sqrt{2}}\right) \quad (7.5.19)$$

que corresponde exatamente à região de confiança definida por (7.1.27).

#### 7.5.4 Processos de média móvel

Se segue um processo ARIMA(0, 1, 0), então  $\tilde{y}$  segue um processo MA(1) puro. Devido à dualidade nos resultados das estimativas e nas matrizes de informação, em particular, entre o modelo autorregressivo e o modelo de média móvel, segue que no caso de média móvel os fatores  $|\tilde{y}|^2$  e  $(\tilde{y})$  em (7.5.6), que em qualquer caso são dominado por  $(\tilde{y})$ , também se cancelam para amostras grandes. Assim, correspondendo a (7.5.7), encontramos que os parâmetros do processo MA(1) têm a distribuição posterior

$$(\tilde{y}) \propto (\tilde{y}) \tilde{y}^2 \quad (7.5.20)$$

## ESTIMATIVA DE 250 PARÂMETROS

Novamente, os contornos de soma de quadrados são, para amostras moderadas, contornos essencialmente *exatos* de densidade posterior. No entanto, porque [ ] não é uma função linear de 's, ( ) [ não é exatamente quadrático em  $\lambda$  ], embora para grandes amostras muitas vezes seja quase isso dentro do relevantes. Nesse caso, temos aproximadamente

$$Q(\lambda) = \frac{1}{2} (\lambda - \bar{\lambda})' \Sigma^{-1} (\lambda - \bar{\lambda}) + \text{constante}$$

aproximação exponencial de  $Q(\lambda)$ . Segue, depois de substituir ( ) em (7.5.20) e usar seguinte vale:

1. Para amostras grandes,  $\lambda$  é *aproximadamente* distribuído em uma distribuição normal multivariada  $N(\bar{\lambda}, \Sigma)$ .
2. Uma região HPD aproximada é definida por (7.5.18) ou (7.5.19), substituindo  $Q(\lambda)$  e substituindo.

**Exemplo: Distribuição Posterior de  $\lambda$  para um Processo IMA(0, 1, 1).** Ilustrar, A Figura 7.8 mostra a distribuição aproximada de densidade posterior  $p(\lambda|z)$  dos dados de Série B. É visto como aproximadamente normal com sua moda em  $\lambda = 1,09$  e tendo um desvio padrão de cerca de 0,05. Um intervalo HPD Bayesiano de 95% cobre essencialmente o mesmo intervalo,  $0,98 < \lambda < 1,19$ , assim como o intervalo de confiança de 95%. Observe que a densidade foi normalizado para ter área unitária sob a curva.

## 7.5.5 Processos Mistos

Se  $\lambda$  segue um processo ARIMA( , , ), então  $\lambda$  segue um processo ARMA( , )

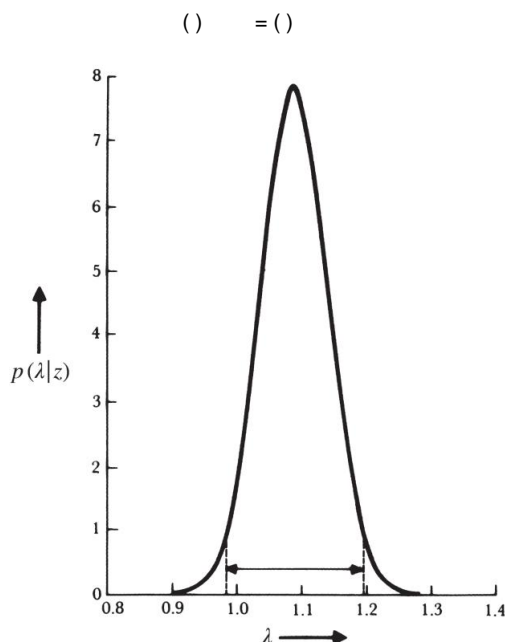


FIGURA 7.8 Densidade posterior  $p(\lambda|z)$  para Série B.

Pode-se mostrar que para tal processo os fatores  $|(\cdot)|^{-1/2}$  e  $(\cdot)$  em (7.5.5) não exatamente cancelar. Em vez disso, podemos mostrar, com base em (7.2.24), que

$$|(\cdot)|^{-1/2}(\cdot) = (\cdot)^{-1/2} \quad (7.5.21)$$

Em (7.5.21), os  $\gamma$ 's são os + parâmetros obtidos pela multiplicação do valor autoregressivo e operadores de média móvel:

$$(1 - \gamma_1 - \gamma_2 - \dots - \gamma_p + \dots) = (1 - \gamma_1 - \dots) \times (1 - \gamma_1 - \dots)$$

e é o jacobiano da transformação de  $\gamma$  para  $(\cdot)$ , ou seja,

$$(\cdot) \gamma (\cdot) = (\cdot) \gamma^2 \quad (7.5.22)$$

Em particular, para o processo ARMA(1, 1),  $\gamma_1 = +$ ,  $\gamma_2 = -\gamma_1$ ,  $\dots = |\gamma|$ , e

$$(\cdot) \gamma (\cdot) = (\cdot) \gamma^2 \quad (7.5.23)$$

Neste caso, vemos que o jacobiano dominará em uma região próxima à linha  $\gamma = 0$  e produzirá densidade zero na linha. Isso é sensato porque a soma dos quadrados  $(\cdot)$  tomará o valor finito  $\gamma$  para qualquer  $\gamma^2 = 0$  e corresponde ao nosso entretenimento possibilidade de que  $\gamma$  seja ruído branco. No entanto, em nossa derivação, não restringimos a intervalo dos parâmetros. A possibilidade de estar assim associada a intervalos ilimitados para os parâmetros (iguais). O efeito de limitar o espaço de parâmetros, por exemplo, introduzindo os requisitos para estacionaridade e invertibilidade ( $|\gamma| < 1$ ,  $|\gamma| < 1$ ) seria produzir um pequeno valor positivo para a densidade, mas esse refinamento parece pouco vale a pena.

A análise Bayesiana reforça o ponto feito na Seção 7.3.5 de que as dificuldades de estimação serão encontradas com o modelo misto e, em particular, com soluções iterativas, quando há quase redundância nos parâmetros (ou seja, fatores quase comuns entre os peças AR e MA). Já vimos que o uso da identificação preliminar geralmente garantem que essas situações sejam evitadas.

## APÊNDICE A7.1 REVISÃO DA TEORIA DA DISTRIBUIÇÃO NORMAL

### A7.1.1 Particionamento de uma forma quadrática positiva-definida

Considere a forma quadrática positiva-definida  $\mathbf{y}^T \mathbf{S}^{-1} \mathbf{y}$ . Suponha que o  $n \times 1$  particionado após o 1º elemento, de modo que  $\mathbf{y} = \begin{pmatrix} y_1 \\ \mathbf{y}_2 \end{pmatrix} = (y_1, y_2, \dots, y_n)^T$ , suponha que a matriz  $\mathbf{S}$  também seja particionada após a 1ª linha e coluna, de modo que

$$\mathbf{S} = \begin{bmatrix} s_{11} & \mathbf{s}_{12}^T \\ \mathbf{s}_{12} & \mathbf{S}_{22} \end{bmatrix}$$

## 252 ESTIMATIVA DE PARÂMETROS

Verifica-se prontamente que  $\tilde{y}^1$  pode ser representado como

$$\tilde{y}^1 = \begin{bmatrix} \tilde{y}^1 \\ 11 \end{bmatrix} \begin{bmatrix} 11 \\ (22 \quad \tilde{y}^1_{12} \quad \tilde{y}^1_{11} \quad 12) \tilde{y}^1 \end{bmatrix} \begin{bmatrix} \tilde{y}^1_{12} \quad \tilde{y}^1_{11} \end{bmatrix}$$

Então, desde

$$\tilde{y}^{-1} = \begin{pmatrix} \tilde{y}^1_{12} & \tilde{y}^1_{11} \\ 1 & 11 \end{pmatrix} \begin{bmatrix} \tilde{y}^1_{12} \\ 11 \end{bmatrix} \begin{bmatrix} 1 \\ (22 \quad \tilde{y}^1_{12} \quad \tilde{y}^1_{11} \quad 12) \tilde{y}^1 \end{bmatrix} \begin{bmatrix} 1 \\ \tilde{y}^1_{12} \quad \tilde{y}^1_{11} \end{bmatrix}$$

=  $\tilde{y}^{-1}$  sempre pode ser escrito como uma soma de duas formas quadráticas contendo 1 e elementos, respectivamente, onde

$$\begin{aligned} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \tilde{y}^1_{12} & \tilde{y}^1_{11} \\ 1 & 11 \end{pmatrix} \begin{bmatrix} 1 \\ 11 \end{bmatrix} \begin{bmatrix} 1 \\ 11 \end{bmatrix} \quad (A7.1.1) \\ &= (2 \quad \tilde{y}^1_{12} \quad \tilde{y}^1_{11} \quad 1) \begin{bmatrix} 1 \\ 11 \end{bmatrix} (22 \quad \tilde{y}^1_{12} \quad \tilde{y}^1_{11} \quad 12) \tilde{y}^1 \begin{bmatrix} 1 \\ 11 \end{bmatrix} \begin{bmatrix} 1 \\ 11 \end{bmatrix} \end{aligned}$$

Também podemos escrever para o determinante de

$$| \tilde{y}^{-1} | = | 11 | | 22 \quad \tilde{y}^1_{12} \quad \tilde{y}^1_{11} \quad 12 | \quad (A7.1.2)$$

## A7.1.2 Duas Integrais Úteis

Let  $\tilde{y}$  Ser uma forma quadrática positiva-definida em  $(1, \dots, \tilde{y})$ , que tem elementos, de modo que  $\tilde{y} = 2, \dots$ , números. Então, pode ser mostrado que e seja real positivo  $\tilde{y}$ ,  $\tilde{y}$ ,

$$\tilde{y} \left( 1 + \frac{\tilde{y}^2}{\tilde{y}^2 + \tilde{y}^2} \right) = \frac{(\tilde{y})^2 \tilde{y}^2}{\tilde{y}^2 | 1 + \tilde{y}^2 |} \quad (A7.1.3)$$

onde a integral -fold se estende por todo o espaço  $\tilde{y}$ , e

$$\frac{\tilde{y}^2}{\tilde{y}^2 (1 + \tilde{y}^2)} \frac{\tilde{y}^2 (1 + \tilde{y}^2)}{\tilde{y}^2 (1 + \tilde{y}^2)} = \tilde{y}^2 \quad (1, \tilde{y}) \quad (A7.1.4)$$

onde a função  $(1, \tilde{y})$  é a densidade de probabilidade da distribuição com e graus de liberdade e é definida por

$$(1, \tilde{y}) = \frac{(\tilde{y})^2 \tilde{y}^2 (1 + \tilde{y}^2)}{C(\tilde{y}^2) C(\tilde{y}^2)} \quad (1 + \tilde{y}^2) \tilde{y}^2 > 0 \quad (A7.1.5)$$

Se tende ao infinito, então

$$(1 + \frac{\tilde{y}^2}{\tilde{y}^2 (1 + \tilde{y}^2)}) \quad \text{tende a} \quad \tilde{y}^2 (1 + \tilde{y}^2)$$



e escrevendo  $\gamma = 2$ , obtemos de (A7.1.4) que

$$\frac{\gamma}{\gamma - 2} > \frac{\gamma}{2} \quad (2) \quad \gamma^2 \quad (A7.1.6)$$

onde a função  $(2)$  é a densidade de probabilidade da  $\gamma^2$  distribuição com graus de liberdade, e é definida por

$$(2) = \frac{1}{2 \gamma^2 \Gamma(\gamma/2)} \left( \frac{\gamma}{2} \right)^{\gamma/2} e^{-\gamma/2} \quad \gamma > 0 \quad (A7.1.7)$$

Aqui e em outros lugares  $( )$  é usado como uma notação *geral* para denotar uma densidade de probabilidade função para uma variável aleatória.

### A7.1.3 Distribuição Normal

A variável aleatória tem distribuição normal com média e desvio padrão  $(\mu, \sigma)$ , se sua densidade de probabilidade for

$$(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (A7.1.8)$$

Assim, a variável normal unitária  $z = (x - \mu)/\sigma$  tem uma distribuição  $(0, 1)$ . Tabela E na Parte Cinco mostra ordenadas  $(z)$  e valores tais que  $\Pr\{Z > z\} = \alpha$  para valores escolhidos de  $\alpha$ .

**Distribuição Multinormal.** O vetor  $\mathbf{y} = (y_1, y_2, \dots, y_p)$  de variáveis aleatórias tem um conjunto de distribuição normal variável  $\{\mathbf{y}\}$  se sua função densidade de probabilidade for

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})} \quad (A7.1.9)$$

A variável multinormal tem vetor médio  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]$  e matriz de variância-covariância  $\Sigma = \text{cov}[\mathbf{y}]$ . Os contornos de densidade de probabilidade são elipsóides definidos por  $(\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) = \text{constante}$ . Para ilustração, os contornos elípticos para uma distribuição normal bivariada  $(p = 2)$  são mostrados na Figura A7.1.

No ponto  $\mathbf{y} = \boldsymbol{\mu}$ , a distribuição normal multivariada tem sua densidade máxima

$$\max_{\mathbf{y}} f(\mathbf{y}) = f(\boldsymbol{\mu}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}}$$

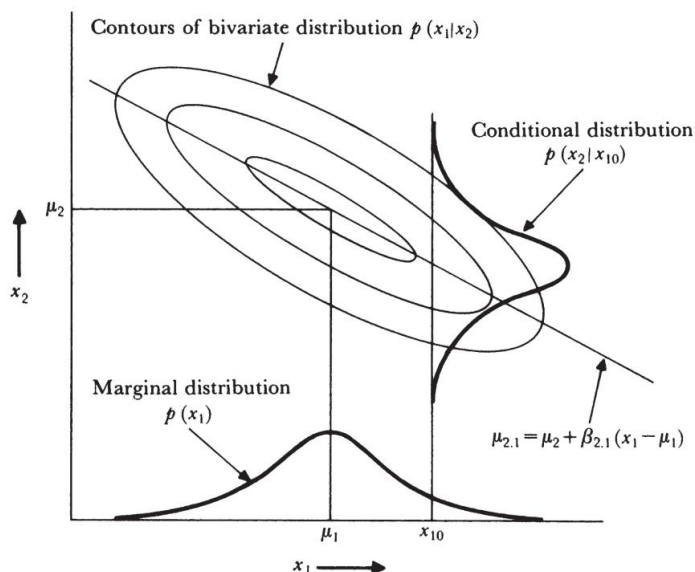
**A Distribuição como Massa de Probabilidade Fora de um Contorno de Densidade da Normal Multivariada.** Para a distribuição normal variável, (A7.1.9), a massa de probabilidade fora do contorno de densidade definido por

$$\gamma = \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})$$

é dado pelo  $\gamma^2$  integral com graus de liberdade:

$$\Pr\{\gamma > \gamma_0\} = \int_{\gamma_0}^{\infty} f(\gamma) d\gamma$$

onde os  $\gamma^2$  função de densidade é definida como em (A7.1.7). A Tabela F na Parte Cinco mostra os valores do  $\gamma^2$ , tal que  $\Pr\{\gamma > \gamma_0\} = \alpha$  para valores escolhidos de  $\alpha$ .



**FIGURA A7.1** Contornos de uma distribuição normal bivariada mostrando a distribuição marginal (1) e a distribuição condicional (2|10) em  $x_1 = 10$ .

### **Distribuições marginais e condicionais para a distribuição normal multivariada.**

Suponha que o vetor de  $n$  variáveis aleatórias seja particionado após os primeiros  $p_1$  elementos, para que

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} y_1, \dots, y_{p_1} \\ y_{p_1+1}, \dots, y_n \end{pmatrix}$$

e que a matriz de variância-covariância é

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Então, usando (A7.1.1) e (A7.1.2), podemos escrever a distribuição normal multivariada para  $\mathbf{y}$  como a distribuição *marginal* de  $\mathbf{y}_1$  multiplicada pela *condicional* distribuição de  $\mathbf{y}_2$  dado  $\mathbf{y}_1$ , ou seja,

$$\begin{aligned} f(\mathbf{y}) &= f(\mathbf{y}_1, \mathbf{y}_2) = f(\mathbf{y}_1) f(\mathbf{y}_2 | \mathbf{y}_1) \\ &= \frac{1}{(2\pi)^{p_1/2} |\Sigma_{11}|^{1/2}} \exp \left[ -\frac{1}{2} \mathbf{y}_1' \Sigma_{11}^{-1} \mathbf{y}_1 \right] \\ &\quad \times \frac{1}{(2\pi)^{(n-p_1)/2} |\Sigma_{22.1}|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{y}_2 - \mathbf{y}_1' \Sigma_{12}^{-1} \Sigma_{21})' \Sigma_{22.1}^{-1} (\mathbf{y}_2 - \mathbf{y}_1' \Sigma_{12}^{-1} \Sigma_{21}) \right] \end{aligned} \quad (\text{A7.1.10})$$

Onde

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \quad (\text{A7.1.11})$$

e  $2 + 2,1(\bar{y} - 1) = [2 | 1]$  definir hiperplanos de regressão em  $(1 + 2)$ -espaço dimensional, traçando os loci das médias (condicionais) de à medida que  $\bar{y}$  varia. O plano de  $2 \times 1$  de coeficientes de regressão é dado por

$$2,1 = \frac{1}{12} \frac{\bar{y} - 1}{11}.$$

Ambas as distribuições marginal e condicional para a normal multivariada são, portanto, distribuições normais multivariadas. É visto que para a *distribuição normal multivariada*, a distribuição condicional  $(2 | 1)$  é, *exceto para localização* (ou seja, valor médio), idêntica qualquer que seja o valor de  $\bar{y}$  (ou seja, normal multivariado com variância idêntica - covariância matriz 2.11).

**Marginais Univariadas.** Em particular, a densidade marginal para um único elemento  $( = 1, 2, \dots, )$  é  $( , )$ , uma normal univariada com média igual ao  $^o$  elemento de  $e$  variação  $^2$  igual ao elemento diagonal de  $e$ .

**Normal bivariado.** Para ilustração, as distribuições marginais e condicionais para uma distribuição bivariada normais são mostrados na Figura A7.1. Neste caso, a distribuição marginal de  $y_1$  é  $(1, \frac{2}{1})$  enquanto a distribuição condicional de  $y_2$  dado  $y_1$  é

$$\{ 2 + \frac{2}{1}(\bar{y}_1 - 1), \frac{2}{2}(\bar{y}_2 - 2) \}$$

onde  $\bar{y}_2 = (1 \bar{y}_2) 2,1$  é o coeficiente de correlação entre  $y_1$  e  $y_2$  e  $2,1 = 12 \bar{y}_1 \frac{2}{1}$  é o coeficiente de regressão de  $y_2$  em  $y_1$ .

#### A7.1.4 Distribuição do Aluno

A variável aleatória é distribuída como uma distribuição em escala com média e escala parâmetro e com graus de liberdade, denotados como  $( , é 2, )$ , se sua densidade de probabilidade

$$f(y) = \frac{1}{2} \bar{y}_1 \bar{y}_2 (2) \bar{y}_1 \bar{y}_2 ( \bar{y}_2 ) \bar{y}_1 \bar{y}_2 ( \bar{y}_2 ) \bar{y}_1 ( \bar{y}_2 ) [ 1 + \frac{(-) 2}{2 \bar{y}_1} + 1 ) \bar{y}_2^2 ]^{2} \quad (A7.1.12)$$

Assim, a variável padronizada  $z = (\bar{y} - \bar{y}) / \bar{y}$  tem distribuição  $(0, 1, )$ . Tabela G na Parte Cinco mostra valores tais que  $Pr\{z > \} = \alpha$  para valores escolhidos de  $\alpha$ .

**Abordagem à Distribuição Normal.** Para grandes  $n$ , o produto

$$(\bar{y} - \bar{y}_1) \bar{y}_2 \bar{y} ( \bar{y}_2 ) \bar{y}_1 ( \bar{y}_2 )$$

tende à unidade, enquanto o colchete à direita em (A7.1.12) tende a  $\bar{y}_1 \bar{y}_2 (2) \bar{y}_1 \bar{y}_2 ( \bar{y}_2 )$ . Assim, se para grande  $n$  escrevemos  $2 = 2$ , a distribuição tende à distribuição normal (A7.1.8).

**Distribuição t múltipla.** Seja  $\mathbf{y} = (y_1, 2, \dots, y_n)$  um vetor  $n \times 1$  e  $\mathbf{S}$  uma matriz definida positiva  $n \times n$ . Então, a variável aleatória vetorial tem uma distribuição multivariada em escala  $( \mathbf{y}, \mathbf{S} )$ , com vetor médio  $\mathbf{\bar{y}}$ , matriz de escala  $\mathbf{S}$  e graus de liberdade se sua probabilidade

## 256 ESTIMATIVA DE PARÂMETROS

densidade é

$$\begin{aligned} f(\mathbf{y}) = & \frac{1}{(2\pi)^{p/2} |\mathbf{S}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \bar{\mathbf{y}})' \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{y}}) \right\} \\ & \times \left[ 1 + \frac{(\mathbf{y} - \bar{\mathbf{y}})' \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{y}})}{p} \right]^{-\frac{p+2}{2}} \end{aligned} \quad (\text{A7.1.13})$$

Os contornos de probabilidade da distribuição múltipla são elipsóides definidos por  $(\mathbf{y} - \bar{\mathbf{y}})' \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{y}}) = \text{constante}$ .

**Abordagem da Forma Multinormal.** Para grandes, o produto

$$\left( \frac{1}{2\pi} \right)^{p/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \bar{\mathbf{y}})' \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{y}}) \right\}$$

Assim, tende à unidade; também, o colchete direito em (A7.1.13) tende a  $(\mathbf{y} - \bar{\mathbf{y}})' \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{y}})$ . Se para grande escrevemos  $\mathbf{S} = \sigma^2 \mathbf{I}$ , o múltiplo tende à distribuição normal multivariada (A7.1.9).

## APÊNDICE A7.2 REVISÃO DA TEORIA LINEAR DOS MÍNIMOS QUADRADOS

### A7.2.1 Equações Normais e Mínimos Quadrados

O modelo de regressão linear é assumido como

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i \quad (\text{A7.2.1})$$

onde  $i = 1, 2, \dots, n$  são observações sobre uma resposta ou variável dependente obtido de um experimento no qual as variáveis independentes assumem valores  $x_{i1}, x_{i2}, \dots, x_{ip}$ .

Os  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  são parâmetros desconhecidos a serem estimados. Os  $e_i$  são erros aleatórios, independentes, com média zero e o mesmo comum variação  $\sigma^2$ .

As relações (A7.2.1) podem ser expressas em forma de matriz como

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

ou

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (\text{A7.2.2})$$

onde a matriz  $\mathbf{X}$  é assumida como de posto completo. O teorema de Gauss dos mínimos quadrados pode ser enunciado (Barnard, 1963) da seguinte forma: As estimativas  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$  dos parâmetros  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  que são observações e imparciais, tais que qualquer função quadrática dos parâmetros é minimizada quando a soma dos quadrados

$$\mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{e}'\mathbf{e} \quad (\text{em } \boldsymbol{\beta}) \quad (\text{A7.2.3})$$

Para estabelecer o mínimo de  $(\hat{y})$ , notamos que o vetor  $\mathbf{w} - \hat{\mathbf{y}}$  pode ser decomposto em dois vetores  $\mathbf{w} - \hat{\mathbf{y}}$  e  $(\hat{\mathbf{y}} - \hat{\mathbf{y}})$  de acordo com

$$\mathbf{w} - \hat{\mathbf{y}} = \mathbf{w} - \hat{\mathbf{y}} + (\hat{\mathbf{y}} - \hat{\mathbf{y}}) \quad (\text{A7.2.4})$$

Assim, desde que escolhamos de modo a  $(\mathbf{w} - \hat{\mathbf{y}}) = 0$ , ou seja,

$$(\hat{\mathbf{y}}) = \hat{\mathbf{y}} \quad (\text{A7.2.5})$$

segue que

$$(\hat{y}) = (\hat{y}) + (\hat{y} - \hat{y}) \quad (\text{A7.2.6})$$

e os vetores  $\mathbf{w} - \hat{\mathbf{y}}$  e  $(\hat{\mathbf{y}} - \hat{\mathbf{y}})$  são ortogonais. Como o segundo termo do lado direito de (A7.2.6) é uma forma quadrática positiva-definida, segue-se que o mínimo é atingido quando  $\hat{\mathbf{y}} = \hat{\mathbf{y}}$ , onde

$$(\hat{y}) = (\hat{y}) \hat{\mathbf{y}} \quad (\text{A7.2.5})$$

é a estimativa de *mínimos quadrados* de dada pela solução da *equação normal* (A7.2.5).

### A7.2.2 Estimativa da Variação do Erro

Usando (A7.2.3) e (A7.2.5), a soma dos quadrados no mínimo é

$$(\hat{y}) = (\mathbf{w} - \hat{\mathbf{y}}) \hat{\mathbf{y}} \quad (\mathbf{w} - \hat{\mathbf{y}}) \hat{\mathbf{y}} = \mathbf{w} \hat{\mathbf{y}} - \hat{\mathbf{y}} \hat{\mathbf{y}} \quad (\text{A7.2.7})$$

Além disso, se definirmos

$$s^2 = \frac{(\hat{y})}{n-2} \quad (\text{A7.2.8})$$

pode ser mostrado que  $[s^2] = 2$  e, portanto, variância  $s^2$  fornece uma estimativa imparcial do erro 2.

### A7.2.3 Matriz de Covariância de Estimativas de Mínimos Quadrados

A matriz de covariância das estimativas de mínimos quadrados  $(\hat{\mathbf{y}})$  é definido por

$$\begin{aligned} V(\hat{\mathbf{y}}) &= \text{cov}[\hat{\mathbf{y}}, \hat{\mathbf{y}}] \\ &= \frac{1}{n-2} [(\mathbf{w} - \hat{\mathbf{y}}) \hat{\mathbf{y}} \quad \mathbf{w}, \mathbf{w}' ((\hat{\mathbf{y}} - \hat{\mathbf{y}}) \hat{\mathbf{y}})] \\ &= \frac{1}{n-2} (\hat{\mathbf{y}}) \hat{\mathbf{y}} \quad \text{cov}[\mathbf{w}, \mathbf{w}] \quad (\hat{\mathbf{y}}) \hat{\mathbf{y}} \end{aligned} \quad (\text{A7.2.9})$$

uma vez que  $\text{cov}[\mathbf{w}, \mathbf{w}] = \mathbf{I} 2$ .

### A7.2.4 Regiões de Confiança

Assumindo normalidade, as formas quadráticas  $(\hat{y})$  e  $(\hat{y} - \hat{y})$  distribuídas  $(\hat{y})$  e  $(\hat{y} - \hat{y})$  em (A7.2.6) são independentemente como  $2$  vezes variáveis aleatórias qui-quadrado com  $-2$

## 258 ESTIMATIVA DE PARÂMETROS

graus de liberdade, respectivamente. Por isso,

$$\frac{(\tilde{y} - \hat{y})'(\tilde{y} - \hat{y})}{n - k} \sim F_{k, n-k}$$

é distribuído como  $F_{k, n-k}$ . Usando (A7.2.8), segue que

$$(\tilde{y} - \hat{y})'(\tilde{y} - \hat{y}) = (\tilde{y} - \hat{y})' \mathbf{V}^{-1} (\tilde{y} - \hat{y}) \quad (A7.2.10)$$

define uma região de confiança de  $1 - \alpha$  para

## A7.2.5 Erros Correlacionados

Suponha que os erros  $\mathbf{e}$  em (A7.2.2) tenham uma matriz de covariância *conhecida*  $\mathbf{V}$ , e seja  $\mathbf{P}$  uma matriz  $n \times n$  não singular tal que  $\mathbf{V}\tilde{\mathbf{P}} = \mathbf{P}\tilde{\mathbf{P}}' = \mathbf{I}$ , de modo que  $\tilde{\mathbf{P}}\mathbf{V}\mathbf{P} = \mathbf{I}$ . Então, (A7.2.2) pode ser transformado em

$$\tilde{\mathbf{P}}\mathbf{w} = \tilde{\mathbf{P}}\mathbf{y} + \tilde{\mathbf{P}}\mathbf{e}$$

ou

$$\tilde{\mathbf{w}} = \tilde{\mathbf{y}} + \mathbf{e}\tilde{\mathbf{P}} \quad (A7.2.11)$$

onde  $\tilde{\mathbf{w}} = \tilde{\mathbf{P}}\mathbf{w}$  e  $\tilde{\mathbf{y}} = \tilde{\mathbf{P}}\mathbf{y}$ . A matriz de covariância de  $\mathbf{e}\tilde{\mathbf{P}}$  em (A7.2.11) é

$$\text{cov}[\tilde{\mathbf{P}}\mathbf{e}, \mathbf{e}\tilde{\mathbf{P}}] = \tilde{\mathbf{P}} \text{cov}[\mathbf{e}, \mathbf{e}] \mathbf{P} = \tilde{\mathbf{P}}\mathbf{V}\mathbf{P} = \mathbf{I}$$

Assim, podemos aplicar a teoria ordinária dos mínimos quadrados com  $\mathbf{V} = \mathbf{I}$  ao modelo *transformado* (A7.2.11), em que  $\mathbf{w}$  é substituído por  $\tilde{\mathbf{w}} = \tilde{\mathbf{P}}\mathbf{w}$  e por  $\tilde{\mathbf{y}}$ . Assim, obtemos as estimativas

$$\hat{\tilde{\mathbf{w}}} = (\tilde{\mathbf{y}}'\tilde{\mathbf{y}})^{-1} \tilde{\mathbf{y}}'\tilde{\mathbf{w}}$$

com  $\mathbf{V}(\hat{\tilde{\mathbf{w}}}) = \text{cov}[\hat{\tilde{\mathbf{w}}}] = (\tilde{\mathbf{y}}'\tilde{\mathbf{y}})^{-1}$ . Em termos das variáveis *originais*  $\mathbf{w}$  do modelo (A7.2.2), a estimativa é

$$\hat{\mathbf{w}} = (\mathbf{P}\mathbf{P}')^{-1} \mathbf{P}\hat{\tilde{\mathbf{w}}} = (\mathbf{P}\mathbf{V}^{-1}\mathbf{P}')^{-1} \mathbf{P}\tilde{\mathbf{y}}'\tilde{\mathbf{w}} \quad (A7.2.12)$$

com

$$\mathbf{V}(\hat{\mathbf{w}}) = \text{cov}[\hat{\mathbf{w}}] = (\mathbf{P}\mathbf{V}^{-1}\mathbf{P}')^{-1}$$

O estimador em (A7.2.12) é geralmente referido como o estimador *de mínimos quadrados generalizados* (GLS), e segue que esta é a estimativa de obtida minimizando a função de soma de quadrados *generalizada*

$$(\hat{\mathbf{w}} - \mathbf{w})' \mathbf{V}^{-1} (\hat{\mathbf{w}} - \mathbf{w})$$

### APÊNDICE A7.3 FUNÇÃO DE PROBABILIDADE EXATA PARA MOVIMENTO PROCESSOS MÉDIOS E MISTOS

Para obter a função de verossimilhança necessária para um modelo  $MA(1)$ , temos que derivar a função de densidade de probabilidade para uma série  $w_t = (1, 2, \dots, n)$  assumida como sendo gerada por um modelo de média móvel inversível de ordem:

$$w_t = \sum_{j=0}^{p-1} \theta_j w_{t-j} + \epsilon_t \quad (A7.3.1)$$

onde  $\epsilon_t$  com  $\epsilon_t \sim N(0, \sigma^2)$  [normalmente]. Sob a suposição de que os  $\epsilon_t$ 's e os  $\theta_j$ 's são distribuído, a densidade da junta pode ser escrita como

$$(w | \theta, \sigma^2) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} w' M(0, \theta) w\right\} \quad (A7.3.2)$$

onde  $M(0, \theta)$   $n \times n$  denota a matriz de covariância  $\times$  dos  $w$ 's para um  $MA(1)$  processo. Consideramos agora uma maneira conveniente de calcular  $M(0, \theta)$ , e por simplicidade, supomos que  $\theta_0 = 0$ , de modo que  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ . Usando o modelo (A7.3.1), podemos escrever as equações:

$$w_t = \sum_{j=1}^p \theta_j w_{t-j} + \epsilon_t \quad (t = 1, 2, \dots, n)$$

Essas equações podem ser convenientemente expressas em forma de matriz em termos de vetores de dimensão  $w = (1, 2, \dots, n)$  e  $a = (1, 2, \dots, n)$ , e o vetor dimensional de valores preliminares  $a_0 = (1, 2, \dots, 0)$  como

$$w = L a + F a_0$$

onde  $L$  é uma matriz triangular inferior  $n \times n$  com 1's na diagonal principal,  $\theta_1$  na primeira subdiagonal,  $\theta_2$  na segunda subdiagonal, e assim por diante, com  $\theta_j = 0$  para  $j > p$ . Mais longe,  $F$  é uma matriz  $n \times p$  com a forma  $F = (B, 0)$  onde  $B$  é  $n \times p$  igual a

$$B = \begin{pmatrix} \theta_1 & \theta_2 & \dots & \theta_p \\ 0 & \theta_1 & \dots & \theta_p \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_p \end{pmatrix}$$

Agora a distribuição conjunta dos valores  $a$ , que são os elementos de  $(a, a_0)$ , é

$$(a, a_0 | \theta, \sigma^2) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} (a, a_0)' a + a' \theta a_0\right\}$$

Observando que a transformação de  $(a, a_0)$  para  $(w, a_0)$  tem unidade Jacobiana e  $a = L^{-1} w - F a_0$ , a distribuição conjunta de  $w = L a + F a_0$  e  $a_0$  é

$$(w, a_0 | \theta, \sigma^2) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} (w, a_0)' w\right\}$$

## ESTIMATIVA DE 260 PARÂMETROS

Onde

$$(\hat{a} - \tilde{a}) = (w - F\tilde{a}) - \tilde{y} L \tilde{y}^{-1} L \tilde{y} (w - F\tilde{a}) + \tilde{a} \tilde{y} \tilde{a} \quad (A7.3.3)$$

Agora, seja  $\hat{a}$  o vetor de valores que minimizam  $(\hat{a} - \tilde{a})$ , que da teoria dos mínimos quadrados generalizados pode ser mostrado igual a  $\hat{a} = D^{-1} F \tilde{y} L \tilde{y}^{-1} L \tilde{y} w$ , onde  $D = I + F \tilde{y} L \tilde{y}^{-1} L \tilde{y} F$ . Então, usando o resultado (A7.2.6), temos

$$(\hat{a} - \tilde{a}) = (\hat{a} - \tilde{a}) + (\tilde{a} - \tilde{a}) = D^{-1} F \tilde{y} L \tilde{y}^{-1} L \tilde{y} w$$

Onde

$$(\hat{a} - \tilde{a}) = (\hat{a} - \tilde{a}) = (w - F\tilde{a}) - \tilde{y} L \tilde{y}^{-1} L \tilde{y} (w - F\tilde{a}) + \tilde{a} \tilde{y} \tilde{a} \quad (A7.3.4)$$

é uma função das observações  $w$  mas não dos valores preliminares  $\tilde{a}$ . Desta forma,

$$(w, \tilde{a} | \tilde{y}) = (2^{-1/2}) \exp \left\{ -\frac{1}{2} [(\tilde{a} - \tilde{a}) - \tilde{y} L \tilde{y}^{-1} L \tilde{y} (w - F\tilde{a})]^2 \right\} D(\tilde{a} - \tilde{a})$$

No entanto, como a distribuição conjunta de  $w$  e  $\tilde{a}$  pode ser fatorada como

$$(w, \tilde{a} | \tilde{y}) = (w | \tilde{y}) (\tilde{a} | w, \tilde{y})$$

segue, semelhante a (A7.1.10), que

$$(\tilde{a} | w, \tilde{y}) = (2^{-1/2}) \exp \left\{ -\frac{1}{2} (\tilde{a} - \tilde{a})^T D(\tilde{a} - \tilde{a}) \right\} \quad (A7.3.5)$$

$$(w | \tilde{y}) = (2^{-1/2}) \exp \left\{ -\frac{1}{2} (w - F\tilde{a})^T D(w - F\tilde{a}) \right\} \quad (A7.3.6)$$

Podemos agora deduzir o seguinte:

1. De (A7.3.5), vemos que  $\hat{a}$  é a esperança condicional de  $\tilde{a}$  dado  $w$  e  $\tilde{y}$ . Assim, usando a notação introduzida na Seção 7.1.4, obtemos

$$\hat{a} = [a | w, \tilde{y}] = [a | \tilde{y}]$$

onde  $[a] = L \tilde{y}^{-1} (w - F[a])$  é a expectativa condicional de  $um$  dado  $w$  e usando (A7.3.4):

$$(\hat{a} - \tilde{a}) = [a] - \tilde{a} = [a] - \tilde{a} = \tilde{y} L \tilde{y}^{-1} L \tilde{y} (w - F\tilde{a}) \quad (A7.3.7)$$

Para calcular  $(\hat{a} - \tilde{a})$ , as quantidades  $[a] = [w, \tilde{y}]$  podem ser obtidas usando as estimativas  $[a] = ([1 \tilde{y}], [1 \tilde{y}])$  e obtidas como os elementos  $[a]$  da previsão para valores da relação  $L[a] = w - F[a]$  como

$$[a] = +1[\tilde{y}_1] + \dots + [\tilde{y}_n] \quad (= 1, 2, \dots, n)$$



Observe que se a expressão para  $\hat{a}_t$  é utilizada em (A7.3.4), após rearranjar obtemos

$$(\hat{a}_t) = w_t' L_t^{-1} (I - L_t^{-1} F_t) \hat{a}_{t-1} + w_t' a_0 \quad (A7.3.5)$$

onde  $a_0 = L_1^{-1} w$  denota o vetor cujos elementos podem ser calculados recursivamente

a partir de  $a_0 = +1$ ,  $a_1 = \hat{y}_1 + \hat{y}_1$ ,  $a_2 = \hat{y}_2 + \hat{y}_2$ , ..., definindo os valores iniciais  $\hat{a}_t$

igual a zero. Assim, o primeiro termo descrito acima,  $\hat{y}(t) = a_0 \hat{y} a_0 = \hat{y} = 1$  (a  $a_0$ )<sup>2</sup>, é função de soma de quadrados condicional, dado  $\hat{a}_t =$ , conforme discutido na Seção 7.1.2.

2. Além disso, descobrimos que

$$M(0, t) = L_t^{-1} (I - L_t^{-1} F_t) M(0, t-1)$$

e  $(t) = w' M(0, t) w$ . Além disso, comparando (A7.3.6) e (A7.3.2), temos

$$|D|^{y_1} = |M(0, t)|$$

3. As previsões anteriores  $\hat{a}_t = [a_t]$  podem ser calculadas mais convenientemente a partir de  $\hat{a}_t = D_t^{-1} F_t' u$  (ou seja, resolvendo  $D_t \hat{a}_t = F_t' u$ ), onde  $u = L_t^{-1} L_t^{-1} w = L_t^{-1} a_0 = (1, 2, \dots, y_t)$ . Observação que os elementos de  $u$  são calculados através de uma recursão inversa como

$$u_t = a_0 + 1 + 1 + \dots +$$

de  $a_0$  até  $a_1 = 1$ , usando valores iniciais zero  $a_1 = \dots = a_{y_t} = 0$ , onde

$a_0$  denotam as estimativas da condicional nos valores iniciais zero  $\hat{a}_t =$ . Além

disso, o vetor  $h = F_t' u$  consiste nos elementos  $\hat{y}_t = \hat{y}_t = 1, \dots, \hat{y}_t$ .

4. Finalmente, usando (A7.3.6) e (A7.3.7), a verossimilhança incondicional é dada exatamente por

$$(y_t, a_t | w) = \left( \frac{1}{2\pi} \right)^{y_t/2} \exp \left\{ -\frac{1}{2} (y_t - \hat{y}_t)' D_t^{-1} (y_t - \hat{y}_t) \right\} \quad (A7.3.8)$$

Por exemplo, no modelo MA(1) com  $\theta = 1$ , temos  $F_t = (y_t, 0, \dots, 0)$ , um  $n$ -dimensional vetor, e  $L$  é tal que  $L_t^{-1}$  tem a primeira coluna igual a  $(1, 1, \dots, 1)$ , de modo a

$$D = 1 + F_t' L_t^{-1} L_t^{-1} F_t = 1 + y_t + y_t + \dots + y_t = \frac{1 - 2^{y_t+1}}{1 - 2}$$

Além disso, os valores condicionais  $a_0$  são calculados recursivamente  $a_0 = +$ ,  $a_1 =$ ,  $a_2 =$ , ..., usando o valor inicial zero  $a_0$  como  $a_0 = 0$ , e os valores do vetor  $u = L_t^{-1} a_0$  são calculado na recursão para trás como  $a_0 =$ ,  $a_1 = a_1 = 1$ , com  $a_1 = +1$ ,  $a_2 = 0$ . Então,

$$\hat{a}_t = [0] = \hat{y}_t D_t^{-1} 1 = \hat{y}_t D_t^{-1} a_0 = \frac{0(1 - 2)}{1 - 2^{y_t+1}}$$

## 262 ESTIMATIVA DE PARÂMETROS

Onde  $\phi_0 = \phi_0 + 1 \phi_0 = 1$ , e a probabilidade exata para o processo MA(1) é

$$(\phi_0^2 | w) = \binom{2}{\phi_0^2} \frac{\phi_0^2 (1 - \phi_0^2)^{1-\phi_0^2}}{(1 - \phi_0^2)^{1+\phi_0^2}} \exp\left\{ -\frac{\phi_0^2}{2} \right\} \quad (A7.3.9)$$

**Extensão aos Processos Autoregressivos e Mistos.** O método descrito acima pode ser prontamente estendido para fornecer a probabilidade incondicional para o modelo misto geral

$$(\phi_0^2 | w) = (\phi_0^2) \quad (A7.3.10)$$

onde  $\phi_0^2$  foram apresentados por Ansley (1979) e por Ljung e Box (1979), a qual derivação pode ser uma alternativa para obter a probabilidade exata que usa a decomposição de Cholesky de uma matriz de covariância de bandas (ou seja, o método de inovações como discutido na Seção 7.4) foi dada por Ansley (1979). Primeiro, assumindo uma média zero para o processo, as relações para o modelo ARMA podem ser escritas na forma matricial, semelhante ao anterior, como

$$L w = L a + F e y$$

onde  $L$  é uma matriz  $n \times n$  da mesma forma que  $L$ , mas com  $s$  no lugar de  $s$ ,  $e' = (1, \dots, 0, 1, \dots, 0)$  é o vetor  $(n+1)$ -dimensional dos valores iniciais, e

$$F = [A B]$$

com

$$\begin{aligned} & \begin{bmatrix} \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & \dots & 2A \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & 0 & \dots \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & 0 & \dots \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & 0 & \dots \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & 0 & \dots \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & 0 & \dots \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & 0 & \dots \end{bmatrix} & \begin{bmatrix} \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & \dots \\ 2\phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & \dots \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & \dots \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & \dots \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & \dots \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & \dots \\ \phi_0^2 & \phi_0^2 & \phi_0^2 & \phi_0^2 & 0 & \dots \end{bmatrix} \\ & \phi_0^2 & \phi_0^2 \end{aligned}$$

Deixar  $\phi_0^2 = [e y e y]$  denota a matriz de covariância de  $e y$ . Esta matriz tem a forma

$$\begin{aligned} & \begin{bmatrix} \phi_0^2 & \phi_0^2 \\ \phi_0^2 & \phi_0^2 \end{bmatrix} \\ & = [ \phi_0^2 \phi_0^2 ] \end{aligned}$$

onde  $\phi_0^2 = [w y w y]$  é uma matriz  $n \times n$  com  $(i, j)$ -ésimo elemento  $\phi_0^2 = [a y w y]$  tem  $n^2$  elementos definidos por  $[ \phi_0^2 ] = \phi_0^2$  para  $y y + y 0$  e  $0$  de outra forma. Para o caso MA(1)  $(\phi_0^2) = \phi_0^2$   $\phi_0^2 = \phi_0^2 = 1$ , e são facilmente determinados recursivamente através das equações na Seção 3.4. As autocorrelações em  $\phi_0^2$  podem ser determinadas a partir das equações (6.12) (ver, por exemplo, Ljung e Box, 1979).

Semelhante ao resultado em (A7.3.3), como  $\mathbf{a} = \mathbf{L}^{-1}(\mathbf{L} \mathbf{w} - \mathbf{F} \mathbf{e})$  e  $\mathbf{e}$  são independentes, a distribuição conjunta de  $\mathbf{w}$  e  $\mathbf{e}$  é

$$(\mathbf{w}, \mathbf{e} | \mathbf{y}, \mathbf{y}_1, \mathbf{y}_2) = (2\pi)^{-\frac{n}{2}} \frac{1}{|\mathbf{D}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (\mathbf{w}, \mathbf{e}) \mathbf{D}^{-1} (\mathbf{w}, \mathbf{e})'\right]$$

Onde

$$(\mathbf{w}, \mathbf{e}) = (\mathbf{L} \mathbf{w} - \mathbf{F} \mathbf{e})' \mathbf{L}^{-1} (\mathbf{L} \mathbf{w} - \mathbf{F} \mathbf{e}) + \mathbf{e}' \mathbf{e}$$

Novamente, pela teoria dos mínimos quadrados generalizada, podemos mostrar que

$$(\mathbf{w}, \mathbf{e}) = (\mathbf{L} \mathbf{w} - \mathbf{F} \mathbf{e})' \mathbf{L}^{-1} (\mathbf{L} \mathbf{w} - \mathbf{F} \mathbf{e}) + \mathbf{e}' \mathbf{e}$$

Onde

$$(\mathbf{w}, \mathbf{e}) = (\mathbf{L} \mathbf{w} - \mathbf{F} \mathbf{e})' \mathbf{L}^{-1} (\mathbf{L} \mathbf{w} - \mathbf{F} \mathbf{e}) + \mathbf{e}' \mathbf{e} \quad (A7.3.11)$$

é a função soma de quadrados incondicional e

$$\mathbf{y} = [\mathbf{e} | \mathbf{w}, \mathbf{y}_1] = [\mathbf{e} | \mathbf{y}_1] = \mathbf{D}^{-1} \mathbf{F}' \mathbf{L}^{-1} \mathbf{L} \mathbf{w} \quad (A7.3.12)$$

representa a expectativa condicional dos valores preliminares  $\mathbf{e}$ , com  $\mathbf{D} = \mathbf{F}' \mathbf{L}^{-1} \mathbf{L} \mathbf{F} + \mathbf{I}$  e  $\hat{\mathbf{a}} = [\mathbf{a}] = \mathbf{L}^{-1}(\mathbf{L} \mathbf{w} - \mathbf{F} \mathbf{e})$ . Por fatoração da distribuição conjunta de  $\mathbf{w}$  e  $\mathbf{e}$ , podemos obter

$$(\mathbf{e} | \mathbf{y}, \mathbf{y}_1, \mathbf{y}_2) = (2\pi)^{-\frac{n}{2}} \frac{1}{|\mathbf{D}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (\mathbf{e} | \mathbf{y}, \mathbf{y}_1, \mathbf{y}_2) \mathbf{D}^{-1} (\mathbf{e} | \mathbf{y}, \mathbf{y}_1, \mathbf{y}_2)'\right] \quad (A7.3.13)$$

como a probabilidade incondicional. Segue imediatamente de (A7.3.13) que o máximo  $Q(\mathbf{y}, \mathbf{y}_1, \mathbf{y}_2) = \frac{1}{2} (\mathbf{y}, \mathbf{y}_1, \mathbf{y}_2) \mathbf{D}^{-1} (\mathbf{y}, \mathbf{y}_1, \mathbf{y}_2)'$  e denota estimativa de máxima verossimilhança para

Novamente, notamos que  $(\mathbf{w}, \mathbf{e}) = \mathbf{y}' \mathbf{D}^{-1} \mathbf{y}$  de  $\hat{\mathbf{a}} = [\mathbf{a}]$  são todos os elementos de  $\mathbf{a}$  a partir da relação  $\mathbf{L} [\mathbf{a}] = \mathbf{L} \mathbf{w} - \mathbf{F} \mathbf{e}$  como

$$[\mathbf{a}] = \mathbf{y}' \mathbf{D}^{-1} \mathbf{y} = \mathbf{y}' \mathbf{D}^{-1} \mathbf{y} + \mathbf{y}' \mathbf{D}^{-1} \mathbf{y}$$

for  $i = 1, 2, \dots, n$ , usando os valores retroprevistos  $[\mathbf{e}]$  para os valores preliminares, com  $[\mathbf{e}] = \mathbf{e}$  for  $i = 1, 2, \dots, n$ .

Além disso, as previsões *retroativas*  $\hat{\mathbf{e}} = [\mathbf{e}]$  podem ser calculadas a partir de  $\hat{\mathbf{e}} = \mathbf{D}^{-1} \mathbf{F}' \mathbf{u}$ , onde  $\mathbf{u} = \mathbf{L}^{-1} \mathbf{L} \mathbf{w} - \mathbf{F} \mathbf{a}$ , e os elementos de  $\mathbf{u}$  são calculados através da recursão para trás como

$$u_i = -\mathbf{F}_i' \mathbf{a}_i + \mathbf{L}_i^{-1} \mathbf{L}_i \mathbf{w}_i + \mathbf{F}_i' \mathbf{e}_i$$

com valores iniciais  $u_1 = 0$ , e  $\mathbf{e}_1 = 0$ . Assim, as estimativas condicionais em valores iniciais zero  $\mathbf{e}_1 = 0$

## 264 ESTIMATIVA DE PARÂMETROS

vetor  $\mathbf{h} = \mathbf{F}'\mathbf{u}$  consiste nos elementos +:

$$\mathbf{h} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Finalmente, usando (A7.1.1) e (A7.1.2), em  $(\hat{y}, \epsilon)$  podemos escrever  $\hat{y} = \mathbf{A}\hat{y} + \mathbf{F}\epsilon$  (A7.3.13)  $\hat{y} = \mathbf{K}\hat{y} + \mathbf{F}\epsilon$ , de modo que temos

$$(\hat{y}, \epsilon) = \begin{bmatrix} \hat{y} \\ \epsilon \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{F} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{y} \\ \epsilon \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{F} \end{bmatrix} \epsilon \quad (\text{A7.3.14})$$

onde  $\mathbf{K} = \mathbf{A} - \mathbf{F}\mathbf{C}$ , bem como  $\hat{y} = \mathbf{K}\hat{y}$ .

Portanto, em geral, a probabilidade associada a uma série  $\mathbf{z}$  de  $n$  valores gerados por qualquer processo ARIMA é dado por

$$(\hat{y}, \epsilon | \mathbf{z}) = (2\pi)^{-n/2} |\mathbf{M}|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{z}' \mathbf{M}^{-1} \mathbf{z}\right] \quad (\text{A7.3.15})$$

Onde

$$(\hat{y}, \epsilon) = \begin{bmatrix} \hat{y} \\ \epsilon \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{F} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{y} \\ \epsilon \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{F} \end{bmatrix} \epsilon$$

e  $|\mathbf{M}| = |\mathbf{A} - \mathbf{F}\mathbf{C}| |\mathbf{I}| = |\mathbf{A} - \mathbf{F}\mathbf{C}|$ . Além disso, expressando o modelo ARMA misto como uma média móvel infinita  $\hat{y} = (1 + \mathbf{A}_1 + \mathbf{A}_2 + \dots)\epsilon$ , e referindo-se aos resultados para o modelo MA puro, segue que na função soma de quadrados incondicional para o modelo misto, temos a relação que  $\hat{y} = \mathbf{A}\hat{y} + \mathbf{F}\epsilon$  (A7.3.13)  $\hat{y} = \mathbf{K}\hat{y} + \mathbf{F}\epsilon$ . Assim, temos também a representação  $(\hat{y}, \epsilon) = \begin{bmatrix} \hat{y} \\ \epsilon \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{F} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{y} \\ \epsilon \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{F} \end{bmatrix} \epsilon$  e na prática os valores  $\hat{y}$  podem ser calculados recursivamente com a soma procedendo de algum ponto  $\hat{y}_1$ , além do qual o  $\hat{y}$  são insignificantes.

**Caso Especial: AR(1).** No caso especial de um modelo AR(1) puro, os resultados descritos acima simplificar um pouco. Temos então  $\hat{y} = \mathbf{A}\hat{y} + \mathbf{F}\epsilon$ ,  $\mathbf{D} = \mathbf{I} - \mathbf{A}$ ,  $\mathbf{C} = \mathbf{I}$ , e  $\hat{y} = \mathbf{D}\hat{y} + \mathbf{F}\epsilon$ , onde  $\hat{y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$  e submatriz esquerda superior de  $\mathbf{L}$  são determinados a partir das relações  $\hat{y}_1 = \mathbf{D}\hat{y}_1 + \mathbf{F}\epsilon_1$ , anteriores  $\hat{y}_1 = 0$ , e, portanto, estes são os mesmos valores obtidos a partir do uso de a abordagem do modelo para trás, conforme discutido na Seção 7.1.4, para o caso especial do AR modelo. Assim, obtemos a soma exata dos quadrados como  $(\hat{y}, \epsilon) = \begin{bmatrix} \hat{y} \\ \epsilon \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{F} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{y} \\ \epsilon \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{F} \end{bmatrix} \epsilon$ . Para ilustrar, considere o processo autorregressivo de primeira ordem em  $\hat{y}_1$ ,

$$\hat{y}_1 = \mathbf{A}\hat{y}_1 + \mathbf{F}\epsilon_1 \quad (\text{A7.3.16})$$

onde pode ser a <sup>a</sup> diferença  $\tilde{y}$  das observações reais e uma série  $\mathbf{z}$  de comprimento  $n$  observações está disponível. Para calcular a probabilidade (A7.3.15), exigimos

$$\begin{aligned} \dot{y}_1 &= \ddot{y}_1 \quad \dot{y}_2 = \ddot{y}_2 \\ \dot{y}_3 &= \ddot{y}_3 \quad \dot{y}_4 = \ddot{y}_4 \end{aligned}$$

uma vez que  $\ddot{y}_1 = 2(1 - \ddot{y}_2) \ddot{y}_1$ . Agora, porque  $\mathbf{D} = 2 \ddot{y}_1 \mathbf{1}_1 + \mathbf{A}^T \mathbf{1} \mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ ,  
 = portanto  $\mathbf{D} = \mathbf{1}$ , substituindo isso nos dois últimos termos de ( ) acima, reduz-se a

$$(\cdot) = \ddot{y} \quad (- \quad \ddot{y}_1)^2 + (1 \ddot{y}_2)^2 \quad 1 \quad (A7.3.17)$$

como resultado que pode ser obtido mais diretamente pelos métodos discutidos no Apêndice A7.4.

**Caso especial: ARMA(1,1).** Como exemplo para o modelo misto, considere o ARMA(1, 1) modelo

$$\tilde{v}_1 = \bar{y} \quad \tilde{v}_1 \quad (\text{A7.3.18})$$

Então, temos  $e\ddot{y} = (0, 0)$ ,  $A1 =$  ,  $1^{-\gamma}$  , e

$$2 = \begin{matrix} & & 0 & 1 \\ & & 1 & \\ 2[-2 & & & 1] \end{matrix}$$

com  $\gamma_0^2 = (1 + \gamma^2)(1 - \gamma^2)$ . Assim, temos

$$\mathbf{D} = \begin{bmatrix} \ddot{y}_1 + \ddot{y}_2 \\ \ddot{y}_1 - \ddot{y}_2 \end{bmatrix} = \frac{1}{\ddot{y}_2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \ddot{y}_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \ddot{y}_2$$

e as estimativas dos valores iniciais são obtidas como  $\ddot{y} = D\dot{y}_1h$ , onde  $\ddot{y} = (\ddot{y}_1, \ddot{y}_2) = (\ddot{y}, \ddot{y})^T$ ,  $\ddot{y}_1 = 1$ ,  $\ddot{y}_2 = 1$ ,  $\ddot{y}_3 = 0$  são obtidos da recursão para trás  $\ddot{y}_i = 0$ , e são obtidos usando o zero inicial  $\ddot{y}_i = 0$ ,  $i = 1, 2, \dots, n$ . Assim, a soma exata dos quadrados é obtida para  $\ddot{y}_i = 0$ ,  $i = 1, 2, \dots, n$ .

$$(\cdot, \cdot) = \ddot{y} \left[ \cdot \right]_{y=0}^{y=1} + \frac{\ddot{y}_0 \ddot{y} \ddot{y}_0}{10} \quad (A7.3.19)$$

com  $[ ] = \ddot{y} [ \ddot{y}1 ] + [ \ddot{y}1 ] = 1, 2, \dots, 2$ ). Além disso,  $\ddot{y}^2_0 \ddot{y}1 = \mathbf{K} = ( \ddot{y} ) 2 \ddot{y}1 ( \ddot{y}$   
temos  $\ddot{y} M(1,1) \ddot{y} = \{ \ddot{y} \mathbf{K} \ddot{y} \mathbf{D} \ddot{y} \} 1$ , com

$$|K||D| = 1 + \frac{1 - 2}{1 - 2} \frac{(-)}{1 - 2} 2$$

## 266 ESTIMATIVA DE PARÂMETROS

APÊNDICE A7.4 FUNÇÃO DE PROBABILIDADE EXATA PARA UM  
PROCESSO AUTOREGRESSIVO

Suponhamos agora que uma dada série  $w_t = (w_1, w_2, \dots)$  é gerado pela ordem  $p$ th modelo autoregressivo estacionário:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + \epsilon_t$$

onde, temporariamente, supõe-se que os  $\epsilon_t$ 's tenham média = 0, mas, como antes, o argumento pode ser estendido para o caso em que  $\epsilon_t \neq 0$ . Assumindo normalidade para os  $\epsilon_t$ 's e, portanto, para os  $w_t$ 's, a função densidade de probabilidade conjunta dos  $w_t$ 's é

$$(w_1, \dots, w_n) = (2\pi)^{-n/2} |M|^{-1/2} \exp\left[-\frac{1}{2} w' M^{-1} w\right] \quad (A7.4.1)$$

e devido ao caráter reversível do processo geral, a matriz  $M$  é simétrica em relação a ambas as suas diagonais principais. Diz-se que tal matriz é *duplamente* simétrica. Agora,

$$(w_1, \dots, w_n) = (w_1, w_2, \dots, w_n) \quad (w_1, \dots, w_n)$$

onde  $w_t = (w_1, w_2, \dots, w_n)$ . O primeiro fator à direita pode ser obtido fazendo uso distribuição

$$(w_1, \dots, w_n) = (2\pi)^{-n/2} \exp\left[-\frac{1}{2} w' M^{-1} w\right] \quad (A7.4.2a)$$

Para  $w$  fixo,  $(w_1, \dots, w_n)$  e  $(w_1, \dots, w_n)$  estão relacionados pela transformação

$$w_1 = w_1, w_2 = w_2, \dots, w_n = w_n$$

que tem unidade jacobiana. Assim, obtemos

$$(w_1, \dots, w_n) = (2\pi)^{-n/2} \exp\left[-\frac{1}{2} w' M^{-1} w\right] \quad (A7.4.2b)$$

Também,

$$(w_1, \dots, w_n) = (2\pi)^{-n/2} |M|^{-1/2} \exp\left[-\frac{1}{2} w' M^{-1} w\right]$$

Desta forma,

$$(w_1, \dots, w_n) = (2\pi)^{-n/2} |M|^{-1/2} \exp\left[-\frac{1}{2} w' M^{-1} w\right] \quad (A7.4.3)$$

Onde

$$M(\gamma, 0) = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} + \sum_{i=1}^{\infty} \gamma^i \sum_{j=1}^{\infty} \gamma^j \gamma^{i+j} = \sum_{i=1}^{\infty} \gamma^i \sum_{j=1}^{\infty} \gamma^j \gamma^{i+j} \quad (A7.4.4)$$

Também,

$$M(\gamma, 0) = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} \quad (A7.4.5)$$

Onde  $\gamma_0, \gamma_1, \dots, \gamma_{j-1}$  são as autocovariâncias teóricas do processo, e

$$|M(\gamma, 0)| = |M(\gamma, 0)|.$$

Agora, seja  $\gamma = +1$ , de modo que

$$w(\gamma) + 1M(\gamma, 0) = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} + \sum_{i=1}^{\infty} \gamma^i \sum_{j=1}^{\infty} \gamma^j \gamma^{i+j} = \sum_{i=1}^{\infty} \gamma^i \sum_{j=1}^{\infty} \gamma^j \gamma^{i+j} \quad (1)2$$

Então,

$$M(\gamma) = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} \quad (1)$$

e os elementos de  $M(\gamma) = M(\gamma, 0)$  pode agora ser deduzido da consideração de que ambos

$M(\gamma)$  e  $M(\gamma)$  são duplamente simétricas. Assim, por exemplo,

$$M_{11}^{(1)} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} \quad (1)$$

e depois de igualar os elementos nas duas matrizes, temos

$$M_{11}^{(1)} = M_{11}^{(1)} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} \quad (1)$$

Procedendo desta forma, encontramos para os processos das ordens 1 e 2:

$$M_{11}^{(1)} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} = \sum_{j=1}^{\infty} \gamma^j \sum_{i=1}^{\infty} \gamma^i \gamma^{j+i} \quad (1)$$

## 268 ESTIMATIVA DE PARÂMETROS

Por exemplo, quando  $\gamma = 1$ ,

$$(y | \beta, \gamma) = (2^{-\gamma/2}) \exp\left\{-\frac{1}{2} \sum_{i=1}^n y_i^2\right\} \quad (A7.4.5)$$

que confere com o resultado obtido em (A7.3.17). O processo de geração deve levar a matrizes cujos elementos são *quadráticos* nos  $\beta$ 's.

Assim, fica claro de (A7.4.4) que não apenas  $(y) = w$  uma forma quadrática em o  $\beta$ 's, mas também é quadrático nos parâmetros. Escrevendo  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ , é claramente verdade que para alguma matriz  $(p+1) \times (p+1)$  cujos elementos são funções quadráticas dos

$$y_i = \sum_{j=1}^p \beta_j x_{ij} + \beta_{p+1}$$

Agora escreva

$$y_i = \sum_{j=1}^p \beta_j x_{ij} + \beta_{p+1} \quad (A7.4.6)$$

A inspeção de (A7.4.4) mostra que os elementos de  $\Sigma$  são somas de quadrados "simétricas" e produtos defasados, definidos por

$$\Sigma_{ij} = \sum_{k=1}^n x_{ik} x_{jk} + \beta_{p+1}^2 \delta_{ij} \quad (A7.4.7)$$

onde a soma contém  $\sum_{k=1}^n x_{ik} x_{jk}$  termos.

Finalmente, podemos escrever a densidade de probabilidade *exata* e, portanto, a probabilidade exata, como

$$(y | \beta, \gamma) = (2^{-\gamma/2}) \exp\left\{-\frac{1}{2} \sum_{i=1}^n y_i^2\right\} \quad (A7.4.8)$$

Onde

$$(y) = \sum_{i=1}^n y_i^2 = \sum_{i=1}^n \left( \sum_{j=1}^p \beta_j x_{ij} + \beta_{p+1} \right)^2 \quad (A7.4.9)$$

e a probabilidade logarítmica é

$$(y | \beta, \gamma) = -\frac{\gamma}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n y_i^2 \quad (A7.4.10)$$

Por exemplo, quando  $\gamma = 1$ , temos

$$(y) = \sum_{i=1}^n y_i^2 = \sum_{i=1}^n \left( \sum_{j=1}^p \beta_j x_{ij} + \beta_{p+1} \right)^2 \quad (A7.4.11)$$



**Estimativas de Probabilidade Máxima.** Diferenciando em relação a  $\gamma$  e cada um dos  $\beta$ 's em (A7.4.10), obtemos

$$\frac{\partial \ln L}{\partial \gamma} = \frac{1}{2} \sum_{t=1}^T \frac{y_t^2}{1 + y_t^2} + \frac{1}{2} \sum_{t=1}^T \frac{y_t^2}{1 + y_t^2} \quad (\text{A7.4.11})$$

$$\frac{\partial \ln L}{\partial \beta_j} = \sum_{t=1}^T \frac{y_t}{1 + y_t^2} \beta_j = 0, \quad j = 1, 2, \dots, p \quad (\text{A7.4.12})$$

Onde

$$L = \frac{1}{2} \ln |D| - \frac{1}{2} (y - D\beta)' D^{-1} (y - D\beta)$$

Portanto, as estimativas de máxima verossimilhança podem ser obtidas igualando essas expressões a zero e resolvendo as equações resultantes.

Temos imediatamente de (A7.4.11)

$$\hat{\gamma}^2 = \frac{1}{2} \sum_{t=1}^T \frac{y_t^2}{1 + y_t^2} \quad (\text{A7.4.13})$$

**Estimativas de  $\beta$ .** Ocorre uma dificuldade em lidar com a equação (A7.4.12), pois, em geral, as quantidades  $(y_t = 1, 2, \dots, T)$  são funções complicadas dos  $\beta$ 's. Consideramos brevemente quatro aproximações alternativas.

1. **Estimativas de mínimos quadrados.** Como o valor esperado de  $(y_t)$  é proporcional a  $(\beta)$ , enquanto o valor de  $|D|$  é independente de  $\beta$ , (A7.4.8) é para tamanhos de amostra moderados ou grandes dominado pelo termo em  $(y_t)$  e pelo termo em  $|D|$ . Se  $(y_t)$  é, em comparação, pequeno. ignorarmos a influência deste termo, então

$$\ln L \approx \ln \left( \frac{1}{2} \right) - \frac{1}{2} (y - D\beta)' D^{-1} (y - D\beta) \quad (\text{A7.4.14})$$

e as estimativas de  $\beta$  obtidas pela maximização de (A7.4.14) são os mínimos quadrados estimativas obtidas por minimização  $(y - D\beta)' D^{-1} (y - D\beta)$ . Agora, de (A7.4.9),  $(y - D\beta)' D^{-1} (y - D\beta)$  é

é uma matriz  $(T + 1) \times (T + 1)$  de somas simétricas de quadrados e produtos, definida em (A7.4.7). Assim, na diferenciação, os valores minimizantes são

$$\begin{aligned} \frac{\partial}{\partial \beta_1} &= \frac{1}{2} \sum_{t=1}^T \frac{y_t}{1 + y_t^2} \beta_1 + \frac{1}{2} \sum_{t=1}^T \frac{y_t}{1 + y_t^2} \beta_1 \\ \frac{\partial}{\partial \beta_2} &= \frac{1}{2} \sum_{t=1}^T \frac{y_t}{1 + y_t^2} \beta_2 + \frac{1}{2} \sum_{t=1}^T \frac{y_t}{1 + y_t^2} \beta_2 \\ &\vdots \\ \frac{\partial}{\partial \beta_p} &= \frac{1}{2} \sum_{t=1}^T \frac{y_t}{1 + y_t^2} \beta_p + \frac{1}{2} \sum_{t=1}^T \frac{y_t}{1 + y_t^2} \beta_p \end{aligned} \quad (\text{A7.4.15})$$

que, em uma notação matricial óbvia, pode ser escrita como

$$d = D\beta$$

## 270 ESTIMATIVA DE PARÂMETROS

de modo a

$$\hat{\beta} = \hat{\beta}_1 \mathbf{d}$$

Essas estimativas de mínimos quadrados também maximizam a densidade posterior (7.5.15).

2. *Estimativas de Probabilidade Máxima Aproximadas.* Recordamos agora um resultado anterior (3.2.3), que pode ser escrito como

$$\hat{\beta}_1 - 1, \hat{\beta}_2 - 2, \dots, \hat{\beta}_p - p = 0 \quad \text{ou} \quad \hat{\beta}_1 > 0 \quad \text{(A7.4.16)}$$

Além disso, tomando as expectativas em (A7.4.12) e usando o fato de que  $E[\hat{\beta}] = 0$ , temos

$$\begin{aligned} & \hat{\beta}_1^2 + (\hat{\beta}_2 - 1) \hat{\beta}_1 + (\hat{\beta}_3 - 2) \hat{\beta}_1 + \dots + (\hat{\beta}_p - p) \hat{\beta}_1 = 0 \\ & \hat{\beta}_1 \hat{\beta}_2 + (\hat{\beta}_2 - 1) \hat{\beta}_2 + (\hat{\beta}_3 - 2) \hat{\beta}_2 + \dots + (\hat{\beta}_p - p) \hat{\beta}_2 = 0 \end{aligned} \quad \text{(A7.4.17)}$$

Depois de multiplicar (A7.4.16) por  $\hat{\beta}_1$  e subtrair o resultado de (A7.4.17), obtemos

$$\hat{\beta}_1^2 = \hat{\beta}_1 (\hat{\beta}_2 - 1) + (\hat{\beta}_3 - 2) \hat{\beta}_1 + \dots + (\hat{\beta}_p - p) \hat{\beta}_1$$

Portanto, ao usar  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$  como uma estimativa de  $\beta_1, \beta_2, \dots, \beta_p$ , uma estimativa natural de  $\beta_1$  é

$$\frac{\hat{\beta}_1^2}{\hat{\beta}_1 (\hat{\beta}_2 - 1) + (\hat{\beta}_3 - 2) \hat{\beta}_1 + \dots + (\hat{\beta}_p - p) \hat{\beta}_1}$$

Substituindo esta estimativa em (A7.4.12) produz

$$\begin{aligned} & \frac{\hat{\beta}_1^2}{\hat{\beta}_1 (\hat{\beta}_2 - 1) + (\hat{\beta}_3 - 2) \hat{\beta}_1 + \dots + (\hat{\beta}_p - p) \hat{\beta}_1} = \frac{1}{\hat{\beta}_1 (\hat{\beta}_2 - 1) + (\hat{\beta}_3 - 2) \hat{\beta}_1 + \dots + (\hat{\beta}_p - p) \hat{\beta}_1} \\ & = 1, 2, \dots, p \end{aligned} \quad \text{(A7.4.18)}$$

levando a um conjunto de equações lineares da forma (A7.4.15), mas agora com

$$\hat{\beta}_1 = \frac{\hat{\beta}_1^2}{\hat{\beta}_1 (\hat{\beta}_2 - 1) + (\hat{\beta}_3 - 2) \hat{\beta}_1 + \dots + (\hat{\beta}_p - p) \hat{\beta}_1}$$

substituindo

3. *Estimativas Condicionais de Mínimos Quadrados.* Para moderado e relativamente grande, podemos considerar também a função soma de quadrados condicional, obtida adotando-se a procedimento na Seção 7.1.3. Isso produz a soma dos quadrados dada no expoente de a expressão em (A7.4.2),

$$\hat{\beta}_1^2 = \hat{\beta}_1^2 - (\hat{\beta}_1 - 1) \hat{\beta}_1 - (\hat{\beta}_2 - 2) \hat{\beta}_1 - \dots - (\hat{\beta}_p - p) \hat{\beta}_1$$

e é a soma dos quadrados associados à distribuição condicional de  $\hat{\beta}_1$ . As estimativas condicionais de mínimos quadrados são obtidas por dado  $\hat{\beta}_1 = (1, 2, \dots, p)$ , que é um problema de regressão linear de mínimos quadrados padrão

associado ao modelo linear =  $\frac{1}{2} \ddot{y}_1 + \frac{1}{2} \ddot{y}_2 + \ddot{y} + \ddot{y} + d\ddot{y}$ , como em  $\ddot{y} = \ddot{y}^{-1}$ , =  $+1, \dots$ , Isso resulta nas estimativas familiares de mínimos quadrados (A7.2.5), onde  $D\ddot{y}$  tem  $(\ddot{y})^0$  elemento e  $d\ddot{y}$  tem ele =  $\ddot{y} = +1 \ddot{y}$  mento =  $+1 \ddot{y}$ .

4. *Estimativas de Yule-Walker.* Por fim, se for moderado ou grande, como aproximação, pode substituir as somas simétricas de quadrados e produtos em (A7.4.15) por vezes a estimativa de autocovariância apropriada. Por exemplo, , onde  $|\ddot{y}|$  = seria ser substituído por  $\ddot{y} \ddot{y} +$ . Ao dividir por 0 em toda a resultante equações, obtemos as seguintes relações expressas em termos da estimativa autocorrelações =  $\ddot{y}$  0:

$$\begin{aligned} \ddot{y}_1 &= \frac{1}{1 + \ddot{y}} \frac{1}{2} \ddot{y}_1 + \ddot{y}_1 \\ \ddot{y}_2 &= \frac{1}{1 + \ddot{y}} \frac{1}{2} \ddot{y}_2 + \ddot{y}_2 \\ &= \frac{1}{1 + \ddot{y}} \frac{1}{2} \ddot{y}_1 + \frac{1}{2} \ddot{y}_2 + \ddot{y} + \ddot{y} \end{aligned}$$

Estas são as bem conhecidas equações de Yule-Walker.

Na notação matricial (7.3.1), eles podem ser escritos  $r = \ddot{y}$ , de modo que

$$r = -1r \quad (\text{A7.4.19})$$

que corresponde às equações (3.2.7), com  $r$  substituído por  $e$  e por

Para ilustrar as diferenças entre as quatro estimativas, tome o caso  $\ddot{y} = 1$ . Então, e, correspondendo  $\ddot{y}_1^2 =$  a (A7.4.12), a estimativa de máxima verossimilhança exata de  $\ddot{y}$  é a 1 solução de

$$\ddot{y}_1^2 = 0 \quad \ddot{y}_1^2 = 0 \quad \ddot{y}_1^2 = 0 \quad \ddot{y}_1^2 = 0 \quad \ddot{y}_1^2 = 0 \quad \ddot{y}_1^2 = 0 \quad \ddot{y}_1^2 = 0 \quad \ddot{y}_1^2 = 0 \quad \ddot{y}_1^2 = 0 \quad \ddot{y}_1^2 = 0$$

Observe  $\ddot{y}_1^2 = \ddot{y}_1^2 (1 - \ddot{y}_1^2)$  e a solução de máxima verossimilhança para  $\ddot{y} = (1 - \ddot{y}_1^2)$  na equação de que de (A7.4.13), pode ser substituído na expressão para onde  $(1 - \ddot{y}_1^2)$  verossimilhança acima, = 12 + cuja solução produz a estimativa de máxima verossimilhança exata. A equação cúbica resulta em uma equação cúbica em  $\ddot{y}$ . Ao rearranjar, o cubo como

$$(-1) \ddot{y}_1^2 \ddot{y} (\ddot{y}_1^2 - 1) \ddot{y}_1^2 \ddot{y} (22 + 11) + 12 = 0 \quad (\text{A7.4.20})$$

e há uma única solução única para esta equação cúbica tal que  $\ddot{y}_1 < \ddot{y}_1$  (por exemplo, Anderson, 1971, pág. 354).

A aproximação 1 corresponde a ignorar o termo  $\ddot{y}_1$  ao todo, rendendo

$$\ddot{y} = \frac{\ddot{y}_1^2}{\ddot{y}_1^2} = \frac{12}{22}$$

## 272 ESTIMATIVA DE PARÂMETROS

A aproximação 2 corresponde a substituir a estimativa  $\hat{y}_t$  rendendo  $\hat{y}_t = \hat{y}_t(1)$  para  $t = 1, 2, \dots$

$$\hat{y}_t = \frac{\hat{y}_t - 2}{\hat{y}_t - 2} \frac{\hat{y}_t(1) - 1}{\hat{y}_t(1) - 2} = \frac{-2}{-2} \frac{12}{22} = \frac{12}{22}$$

A aproximação 3 corresponde à estimativa de mínimos quadrados do modelo linear padrão obtida por regressão de  $\hat{y}_t$  para  $t = 2, 3, \dots$ , de modo que

$$\hat{y}_t = \frac{\hat{y}_t - 2}{\hat{y}_t - 2} \frac{\hat{y}_t(1) - 1}{\hat{y}_t(1) - 2} = \frac{12}{22 + 2} = \frac{12}{24}$$

Com efeito, isso pode ser visto como obtido substituindo a  $\hat{y}_t$  por  $\hat{y}_t(1)$  na probabilidade equação acima pela Aproximação 4 substitui o numerador e o denominador pela autocovariância padrão estimativas (2.1.12), rendendo

$$\hat{y}_t = \frac{\hat{y}_t - 2}{\hat{y}_t - 2} \frac{\hat{y}_t(1) - 1}{\hat{y}_t(1) - 2} = \frac{1}{0} = \frac{12}{11}$$

Normalmente, como neste exemplo, para amostras moderadas e grandes, as diferenças entre as estimativas dadas pelas várias aproximações serão pequenas. Muitas vezes empregamos as estimativas de mínimos quadrados dadas pela aproximação 1 que podem ser calculadas diretamente de (A7.4.15). No entanto, para cálculos de computador, muitas vezes é mais simples, mesmo quando o ajuste modelo é autorregressivo, para usar o algoritmo iterativo geral descrito na Seção 7.2.1, que calcula estimativas de mínimos quadrados para qualquer processo ARMA.

**Estimativa de  $\hat{y}_t$**  Usando a aproximação 4 com (A7.4.9) e (A7.4.13),

$$\hat{y}_t = \frac{(\hat{y}_t)}{0} = \frac{\hat{y}_t}{1} = \frac{\hat{y}_t}{1} = \frac{\hat{y}_t}{1}$$

Ao multiplicar o lado direito e lembrar que  $r = 0$ , encontramos que

$$\hat{y}_t^2 = 0(1 - r\hat{y}_t) = 0(1 - r\hat{y}_t) = 0(1 - r\hat{y}_t) = 0(1 - r\hat{y}_t) \quad (\text{A7.4.21a})$$

É facilmente mostrado  $\hat{y}_t^2$  pode ser escrito de forma semelhante em termos da autocorrelação teórica que tions:

$$\hat{y}_t^2 = 0(1 - r\hat{y}_t) = 0(1 - r\hat{y}_t) = 0(1 - r\hat{y}_t) = 0(1 - r\hat{y}_t) \quad (\text{A7.4.21b})$$

concordando com o resultado (3.2.8).

Expressões paralelas para  $\hat{y}_t$  pode ser obtido para as aproximações 1, 2 e 3.

**Matriz de Informações.** Diferenciando pela segunda vez em (A7.4.11) e (A7.4.18), temos obtivemos

$$\frac{\partial^2}{\partial \theta^2} = \frac{\partial^2}{\partial \theta^2} + \frac{(\partial)}{\partial \theta^3} \quad (\text{A7.4.22a})$$