

**ESTIMAÇÃO L_1
EM MODELOS ARMA**

Francisco Alberto Pino

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE DOUTOR
EM
CIÊNCIAS

Área de Concentração: **Estatística**
Orientador: **Prof. Dr. Pedro Alberto Morettin**

-São Paulo, Fevereiro de 1990-

*A Daphne e Marina
musas*

AGRADECIMENTOS

Agradecemos ao Prof. Dr. Pedro Alberto Morettin e à Prof. Carmen Diva Saldiva, do Instituto de Matemática e Estatística, da Universidade de São Paulo; ao Prof. Dr. Subhash C. Narula, da School of Business, Virginia Commonwealth University; à Prof. Dra. Gabriela Stangenhans, do Instituto de Matemática, Estatística e Ciências da Computação, da Universidade Estadual de Campinas; a Irene Roque de Oliveira e Vera Lúcia Ferraz dos Santos Francisco, do Instituto de Economia Agrícola. Agradecemos, também, à Empresa Agropecuária Brasileira pela bolsa durante o curso de doutoramento.

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO	1
1.1. Desenvolvimento Histórico	2
1.2. Estimação L_1 e Distribuição dos Erros	5
1.3. Estimação L_1 em Regressão e Auto-regressão	7
1.3.1. Estimação L_2 em auto-regressão com variância infinita	8
1.3.2. Estimação L_1 em regressão linear	8
1.3.3. Estimação L_1 em auto-regressão com variância infinita	9
1.4. Objetivos e Plano do Trabalho	10
CAPÍTULO 2 - CONSISTÊNCIA E OUTRAS PROPRIEDADES	12
2.1. Séries Temporais e Modelos ARMA	13
2.2. O Problema de Estimação	24
2.3. Consistência	30
2.4. Observações sobre a Distribuição Assintótica	38
CAPÍTULO 3 - ALGORITMOS	39
3.1. Introdução	39
3.2. Algoritmos para a Estimação L_1 de Modelos ARMA	41
3.2.1 Algoritmo de Osborne e Watson	42
3.2.2 Algoritmo de Anderson e Osborne	43
3.2.3 Algoritmo de El-Attar, Vidyasagar e Dutta	44
3.2.4 Algoritmo de Tishler e Zang	45
3.2.5 Algoritmo de Schlossmacher	46
CAPÍTULO 4 - ESTUDO DE MONTE CARLO	47
4.1. Planejamento dos Experimentos	48
4.1.1. Geração das séries temporais	48

4.1.2. Primeiro experimento	49
4.1.3. Segundo experimento	49
4.2. Resultados e Discussão	50
4.2.1. Curvas de nível	50
4.2.2. Estimação	52
4.2.3. Eficiência relativa	53
CAPÍTULO 5 - CONCLUSÕES E RECOMENDAÇÕES	87
REFEFÊNCIAS	89

L_1 -NORM ESTIMATION OF ARMA MODELS

ABSTRACT

The least absolute deviations (L_1 -norm) estimator for the parameters of a stationary and invertible autoregressive - moving average model is considered. It is shown to be strongly consistent via a recent result on statistical estimators that minimize a non differentiable function, based on nonsmooth analysis. Asymptotic distribution is also discussed. Finally, a review of the available algorithms and a Monte Carlo study are presented.

Keywords: time series analysis; ARMA models; L_1 -norm estimation; consistency; infinite variance; non-differentiability; non-convexity; non-linearity.

RESUMO

Considera-se o estimador de mínimos desvios absolutos (estimador L_1) para os parâmetros de um modelo auto-regressivo e de médias móveis. Mostra-se que ele é consistente por meio de um resultado recente sobre estimadores que minimizam uma função não diferenciável. Discute-se também a distribuição assintótica. Finalmente, apresenta-se uma revisão dos algoritmos disponíveis e um estudo de Monte Carlo.

Palavras chave: análise de séries temporais; modelos ARMA; estimação L_1 ; consistência; variância infinita; não diferenciabilidade; não convexidade; não linearidade.

CAPÍTULO 1

INTRODUÇÃO

Nos últimos dois ou três séculos o problema de ajustar modelos a um conjunto de observações minimizando somas de resíduos do tipo

$$S = \sum_i |r_i|^p, \quad (1.1)$$

para $p = 1, 2, 3, \dots$ fixado, tem sido freqüentemente discutido. Cada valor de p define um método ou critério de ajustamento conhecido por estimação L_p . Tais métodos foram inicialmente usados em problemas de astronomia e geodésia, mas, atualmente, são largamente empregados em outras áreas, como a econometria. Embora a estimação L_p tenha sido desenvolvida tendo em vista o problema de ajustamento de modelos lineares — inicialmente apenas uma linha reta — a um conjunto de observações, nas últimas duas décadas outros modelos, como os auto-regressivos, passaram a ser considerados. Neste trabalho consideramos a estimação L_1 no caso de um modelo não linear, a saber, o modelo auto-regressivo e de médias móveis (ARMA), de grande importância na análise de séries temporais.

1.1. DESENVOLVIMENTO HISTÓRICO

A seguir apresentamos resumidamente o desenvolvimento da estimação L_1 no contexto mais geral da estimação L_p .

Ruggero Giuseppe Boscovich, um abade italiano de extração croata, estudando a medição “de um arco de meridiano próximo de Roma”, para “corrigir o mapa do Estado Papal”, propôs pela primeira vez dois critérios para ajustar uma linha reta $y = a + bx$ a um conjunto de observações: a) as somas dos resíduos positivos e negativos deveriam ser numericamente iguais, o que equivale a exigir que a reta ajustada passe pelo centróide (\bar{x}, \bar{y}) ; b) a soma dos valores absolutos dos resíduos deveria ser mínima (Boscovich, 1757; Harter, 1974). O próprio Boscovich(1760) esboçou um método geométrico para resolver este problema, enquanto Laplace(1793) propôs um método analítico.

Pierre Simon Laplace(1786) propôs o critério de minimizar o desvio absoluto máximo para o mesmo problema. Esse princípio minimax também é creditado a Leonhard Euler(1749).

Carl Friedrich Gauss foi o primeiro a utilizar na estimação de modelos lineares o critério de minimizar a soma de quadrados dos resíduos, desde 1795 (Gauss, 1806), embora tenha sido Adrien Marie Legendre(1805) o primeiro a publicar o princípio de mínimos quadrados. O uso do método de mínimos quadrados generalizou-se rapidamente, suplantando os outros métodos. Uma das razões foi a facilidade dos cálculos. Outra foi que ele segue naturalmente da distribuição normal (Gauss, 1809), cujo uso também se generalizou. Além disso, embora o melhor método dependa da distribuição dos erros quando o tamanho da amostra é pequeno, o método de mínimos quadrados é assintoticamente melhor para qualquer distribuição desde que assintoticamente a média seja normalmente distribuída, como mostrado por Laplace(1812) e outros.

Quando os desvios são tomados em relação à mediana, ao invés da média, metade deles é positiva e metade negativa (Delambre, 1813). Neste caso, a restrição de que as somas dos desvios positivos e negativos seja igual, exigida por Boscovich e Laplace, é removida. De

fato, “a soma de quadrados dos desvios é um mínimo quando tomados em relação à média aritmética, enquanto que a soma de desvios absolutos é um mínimo quando tomados em relação à mediana”, como mostrado por Gustav Theodor Fechner(1874). Edgeworth(1887) removeu essa restrição e discutiu as vantagens do método de mínimos desvios absolutos em relação ao método de mínimos quadrados. Turner(1887) mostrou que o algoritmo de Edgeworth podia permitir soluções não únicas, mas, Edgeworth(1888) mostrou como tratar esse problema. Uma parte significativa da discussão sobre os métodos de mínimos quadrados e de mínimos desvios absolutos recaiu na discussão sobre o uso da média aritmética ou da mediana, que por sua vez está ligada à questão da existência de valores estranhos (*outliers*) entre as observações. Como mostrado por Maurice Fréchet(1935), na maioria dos casos a média é mais precisa, como no caso de distribuição normal, mas, em alguns casos a mediana deve ser preferida.

Técnicas para a estimação dos parâmetros de um modelo de regressão linear minimizando a soma de desvios absolutos dos resíduos foram propostas por Rhodes(1930) e Singleton(1940). Apesar de todos esses trabalhos, o método de desvios absolutos mínimos “nunca despertou muita atenção” (Taylor, 1974) até Charnes, Cooper e Ferguson(1955), que demonstraram que esse problema poderia ser resolvido usando-se as técnicas de programação linear. Karst(1958) e Wagner(1959) trabalharam na mesma linha.

Desde então, muitos autores têm-se dedicado ao estudo do método de desvios absolutos mínimos para estimar parâmetros de regressão linear. Vejam-se Bloomfield e Steiger(1983), Dielman e Pfaffenberger(1982), Dielman(1984), Gentle(1977), Harter(1974), Narula e Wellington(1982), Sposito, Smith e McCormick(1978), Taylor(1974). Algoritmos cada vez mais eficientes foram apresentados por Amemiya(1982), Armstrong e Frome(1976, 1977), Armstrong e Kung(1978), Armstrong, Frome e Kung(1979), Barrodale e Roberts (1973, 1977), Bloomfield e Steiger(1983), McCormick e Sposito(1975), Narula e Wellington(1977a, 1977b, 1984), Sadovsky(1974), Schlossmacher(1973), Sposito, Kennedy e Gentle(1977), Sposito(1976). As propriedades do estimador foram estudadas por Gentle,

Kennedy e Sposito(1977) e sua convergência foi estudada por Basset e Koenker(1978), Rosenberg e Karlson(1977), Balet-Lawrence(1975). Desenvolvimentos recentes foram apresentados na Conferência Internacional sobre Análise de Dados Estatísticos Baseada na Norma L_1 e Métodos Relacionados, acontecida em Neuchâtel, Suíça, em 1987 (Dodge, 1987).

O problema de estimação L_1 em modelos de regressão não linear foi considerado por Oberhofer(1982) e por Dupačová(1987).

Finalmente, as possibilidades desse método na estimação dos parâmetros de modelos de séries temporais foram discutidas por Cogger(1979). A convergência das estimativas em modelos auto-regressivos foi estudada por Gross e Steiger(1979), An e Chen(1982) e por Bloomfield e Steiger(1983).

A contribuição do presente trabalho consiste na extensão de alguns resultados obtidos em modelos auto-regressivos para o caso mais geral de modelos ARMA.

1.2. ESTIMAÇÃO L_1 E DISTRIBUIÇÃO DOS ERROS

Uma das suposições mais usuais em modelos de séries temporais é a de que os erros do modelo tenham sido gerados por uma distribuição com variância finita. Na verdade, boa parte da teoria clássica de séries temporais baseia-se nessa suposição (ver, por exemplo, Anderson, 1970, e Grenander e Rosenblat, 1957). Entretanto, alguns autores têm mostrado que a variância infinita é mais realista em alguns contextos econômicos (Granger e Orr, 1972; Mandelbrot, 1963 e 1967; Nyquist, 1983). A título de ilustração, consideremos o exemplo da variação de preço da batata no Estado de São Paulo, que é considerado especulativo. Tomemos o preço médio mensal de venda de batata lisa especial no mercado atacadista de São Paulo, no período de janeiro de 1966 a agosto de 1989, levantado pelo Instituto de Economia Agrícola e deflacionado pelo índice geral de preços (disponibilidade interna) da Fundação Getúlio Vargas. A primeira diferença (variação de preço) tem média 0,2082, próxima de zero. O histograma mostra certo afastamento em relação à distribuição normal equivalente (figura 1.1). Semelhantemente aos casos apresentados por Mandelbrot(1963), o histograma dos dados é mais pontiagudo que a normal correspondente e parece haver valores estranhos ou aberrantes (*outliers*).

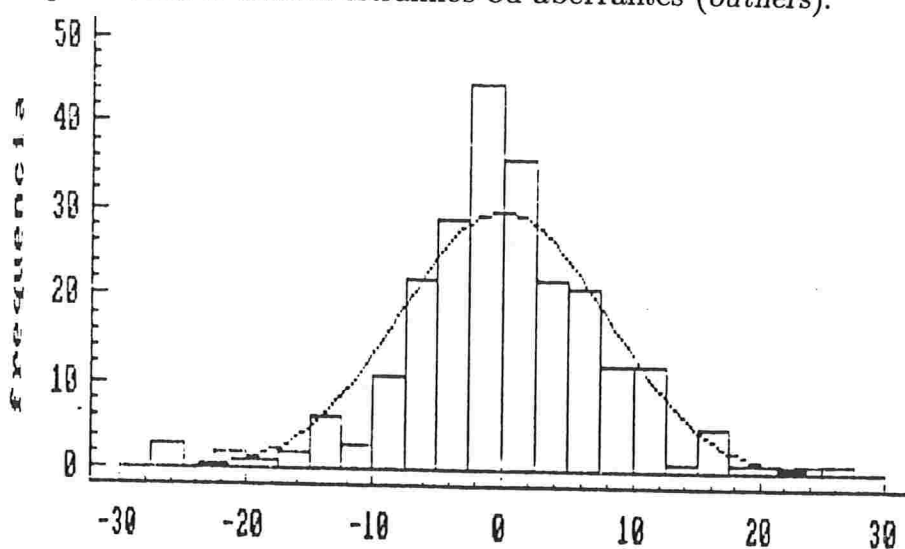


FIGURA 1.1 - Afastamento da normalidade da primeira diferença do preço médio mensal deflacionado de venda de batata lisa especial no mercado atacadista de São Paulo, janeiro de 1966 a agosto de 1989.

Como será visto adiante, quando os erros do modelo seguem a distribuição de Laplace, o estimador L_1 é equivalente ao de máxima verossimilhança. Entretanto, mesmo essa distribuição não se ajusta ao caso da figura 1.1 (a estatística de qui-quadrado de ajustamento correspondente é igual a 213,58), continuando a aparecer valores estranhos, o que parece indicar a necessidade de se utilizar uma distribuição com caudas pesadas para ajustar-se a tais observações.

Uma forma de tratar com muitos valores estranhos consiste em permitir que a variância seja infinita, o que significa uma distribuição com caudas pesadas (Taylor, 1974). A variância infinita aparece, por exemplo, quando estamos tratando com distribuições estáveis com expoente característico $\alpha \in (0, 2]$. A distribuição normal é estável com $\alpha = 2$, enquanto que a distribuição de Cauchy é estável com $\alpha = 1$ (Feller, 1966, v.2, p.167 e 170). Uma distribuição estável com expoente α tem o α -ésimo momento finito (Feller, 1966, v.2, p.215) e assim, distribuições estáveis com $\alpha \in (0, 2)$ têm variância infinita. Portanto, a questão da variância infinita também está relacionada à questão da não normalidade.

Por outro lado, independentemente do modelo, o estimador L_1 é robusto (no sentido de resistência a valores estranhos), porque, ao contrário do que acontece com o estimador L_2 , ele não leva em conta o valor de cada erro, mas, seu sinal. Isto significa que se espera que o estimador L_1 funcione melhor do que o L_2 quando os erros seguem uma distribuição com variância infinita.

Para se obter alguma indicação sobre a relação entre a estimação L_p e a distribuição dos erros do modelo, consideremos o caso mais simples em que as observações y_i possam ser modeladas por $y_i = \mu + e_i, (i = 1, 2, \dots, n)$, onde os e_i sejam variáveis aleatórias e μ seja o parâmetro a ser estimado. Neste caso, Rice e White(1964), baseados em Cramer(1946) e Siddiqui(1962) mostraram que a eficiência (no sentido de menor variância assintótica) da estimação L_p depende da distribuição dos erros. O estimador L_∞ (também chamado minimax ou de Tchebisheff) é o melhor (no sentido de menor variância) quando a distribuição dos erros tem extremos bem definidos, como é o caso da distribuição

uniforme. Por outro lado, o estimador L_1 é o melhor quando a distribuição dos erros tem caudas pesadas, como é o caso da distribuição de Cauchy. O estimador L_2 (ou de mínimos quadrados) é o melhor em casos intermediários, como o da distribuição normal (Rice e White, 1964). Essa relação entre a eficiência da estimação L_p e a distribuição dos erros também aparece em casos mais gerais quando a estimação L_p é comparada à estimação de máxima verossimilhança (Turner, 1960; Nyquist, 1987).

Em resumo, há indicações de que em modelos de séries temporais o estimador L_1 parece funcionar melhor do que o L_2 em certas situações, como quando os erros não são normais, ou quando há valores estranhos, ou quando os erros seguem uma distribuição com variância infinita.

1.3. ESTIMAÇÃO L_1 EM REGRESSÃO E AUTO-REGRESSÃO

Existe considerável interesse em saber se as estimativas L_p convergem em algum sentido. Seja $\hat{\beta}_n^p$ o vetor de estimativas L_p , baseadas em n observações, do vetor de parâmetros β de algum modelo dado e consideremos a expressão

$$n^\delta (\hat{\beta}_n^p - \beta). \quad (1.2)$$

Podemos estudar a consistência das estimativas, isto é, se (1.2) converge para zero quase certamente (consistência forte), ou em probabilidade (consistência fraca), ou em L_p , para algum δ , quando $n \rightarrow \infty$. Também a convergência em distribuição da expressão em (1.2) pode ser investigada, particularmente a convergência para a distribuição normal.

Nesta seção apresentamos resultados existentes na literatura para modelos de regressão e para processos auto-regressivos.

1.3.1. Estimação L_2 em Auto-regressão com Variância Infinita

Neste caso, Kanter e Steiger(1974) e Hannan e Kanter(1977) mostraram que, se os erros tiverem distribuição simétrica no domínio de atração de uma lei estável de índice $\alpha \in (0, 2)$, então, para $p = 2$, a convergência (1.2) vale quase certamente e em probabilidade, para qualquer $\delta < 1/\alpha$.

Yohai e Maronna(1977) mostraram que se os erros tiverem distribuição simétrica não concentrada na origem e se valer $E(\log^+ |a_t|) \leq \infty$, onde $\{a_t\}$ é o termo de erro, então, $n^{1/2}(\hat{\beta}_n^2 - \beta)$ será limitada em probabilidade. Além disso, se $E(a_t^2) < \infty$, então,

$$n^{1/2}(\hat{\beta}_n^2 - \beta) \xrightarrow{\mathcal{D}} N(0, W), \quad (1.3)$$

isto é, ela converge em distribuição para uma normal multivariada, onde a matriz de covariância W depende de β , mas, não da distribuição do erro.

1.3.2. Estimação L_1 em Regressão Linear

Para regressão linear, Balet-Lawrence(1975) mostrou que, num espaço paramétrico compacto, e sob fortes condições de regularidade, as estimativas L_p são consistentes, isto é,

$$(\hat{\beta}_n^p - \beta) \longrightarrow 0 \quad (1.4)$$

e, para $1 < p \leq 2$, a expressão em (1.2) converge em distribuição para a normal.

Para estimativas L_1 , Rosenberg e Carlson(1977) mostraram que, se os erros tiverem distribuição simétrica, então, $(\hat{\beta}_n^1 - \beta)$ terá distribuição amostral aproximada normal com média zero e matriz de covariância $\lambda^2(X'X)^{-1}$, onde X é a matriz de variáveis explicativas e λ^2/n é a variância da mediana de uma amostra de tamanho n da distribuição dos erros.

Basset e Koenker(1978) mostraram que, se os erros tiverem distribuição contínua, com densidade f contínua e positiva na mediana, e se $\lim n^{-1}X'X = Q$, uma matriz

positiva definida, então,

$$n^{1/2}(\hat{\beta}_n^1 - \beta) \xrightarrow{\mathcal{D}} N(\mathbf{0}, w^2 \mathbf{Q}^{-1}), \quad (1.5)$$

quando $n \rightarrow \infty$, onde w^2 é a variância assintótica da mediana amostral, isto é, $w = [2 f(0)]^{-1}$.

Boomfield e Steiger(1983) mostraram que, se as variáveis forem integráveis e se os erros tiverem mediana única nula, então,

$$(\hat{\beta}_n^1 - \beta) \longrightarrow \mathbf{0} \text{ q.c.} \quad (1.6)$$

quando $n \rightarrow \infty$. Além disso, se a matriz \mathbf{C} dos segundos momentos de \mathbf{X} for positiva definida, se a variável dependente tiver densidade f contínua e positiva na mediana (única e nula), e se a amostra for estacionária e ergódica, então,

$$n^{1/2}(\hat{\beta}_n^1 - \beta) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{C}^{-1}/[2 f(0)]^2). \quad (1.7)$$

1.3.3. Estimação L_1 em Auto-regressão com Variância Infinita

Neste caso, Gross e Steiger(1979) mostraram que, se os erros tiverem mediana única nula e média finita, então,

$$(\hat{\beta}_n^1 - \beta) \longrightarrow \mathbf{0} \text{ q.c.} \quad (1.8)$$

quando $n \rightarrow \infty$.

An e Chen(1982) mostraram que, se os erros tiverem mediana nula e distribuição no domínio de atração de uma lei estável de índice $\alpha \in (1, 2)$, então, a convergência de (1.2) para zero vale em probabilidade, quando $n \rightarrow \infty$, para qualquer $\delta < 1/\alpha$. Além disso, essa convergência vale para qualquer $\delta < 1$ se o erro tiver distribuição de Cauchy centrada na origem.

1.4. OBJETIVOS E PLANO DO TRABALHO

Nesta introdução procuramos mostrar que os resultados existentes na literatura sugerem algumas situações (como não normalidade dos erros do modelo, valores estranhos e erros seguindo distribuição com variância infinita) nas quais o estimador L_1 pode ser melhor do que outros da classe dos estimadores L_p (especialmente o L_2) para estimar certos modelos de séries temporais, como os auto-regressivos e os ARMA. Tais resultados sugerem, ainda, que essas situações ocorrem na prática e sobre isso apresentamos um exemplo brasileiro em nossa área de atuação, a Economia Agrícola (figura 1.1). Uma vez definido o interesse teórico e prático do problema, e um modelo no qual ele ainda não foi estudado (no caso, o modelo ARMA, que tem aplicações em áreas como a Economia Agrícola), podemos estabelecer nosso objetivo e o plano de trabalho.

Nosso objetivo é estabelecer algumas propriedades do estimador L_1 no contexto do modelo auto-regressivo e de médias móveis (ARMA), utilizado na análise de séries temporais, particularmente a consistência forte das estimativas.

Para estabelecer as limitações do trabalho e para facilitar a pesquisa futura sobre o assunto, é válido assinalar as principais dificuldades a serem removidas no estudo desse problema, que são as seguintes:

- a) a função objetivo não é diferenciável;
- b) a função objetivo não é linear nos parâmetros;
- c) a função objetivo não é convexa no espaço paramétrico;
- d) a variância dos erros pode ser infinita;
- e) o estimador não pode ser escrito de forma explícita.

A não convexidade da função objetivo impede que utilizemos técnicas de demonstração análogas às de Gross e Steiger(1979) e de An e Chen(1982). A dificuldade em escrever o estimador de forma explícita impede que utilizemos técnicas de demonstração semelhantes às de Basset e Koenker(1978), por exemplo. A não linearidade, a não diferen-

ciabilidade e a não convexidade da função objetivo levaram-nos a utilizar técnicas baseadas na análise de funções não diferenciáveis (Rockafellar, 1980, 1981 e 1983; Clarke, 1983; Ekeland e Turnbull, 1983).

No capítulo 2 apresentamos um resultado de consistência forte das estimativas. No capítulo 3 apresentamos algoritmos que podem ser utilizados no trabalho computacional de estimação. Finalmente, no capítulo 4 apresentamos um estudo de Monte Carlo e no capítulo 5 as conclusões e recomendações.

CAPÍTULO 2

CONSISTÊNCIA E OUTRAS PROPRIEDADES

Neste capítulo caracterizamos a estimação L_1 dos parâmetros de uma série temporal gerada por um processo ARMA. Apresentamos um resultado de consistência forte das estimativas baseado em Dupačová e Wets(1988). Para chegar a tal, mostramos algumas propriedades dos resíduos, da função objetivo e do espaço paramétrico.

2.1. SÉRIES TEMPORAIS E MODELOS ARMA

Seja $(\Omega', \mathcal{A}', P')$ um espaço de probabilidade, (Ω, \mathcal{A}) e $(\Omega_a, \mathcal{A}_a)$ espaços mensuráveis e T um conjunto arbitrário (usualmente T é o conjunto dos números inteiros). Suponhamos que $\Omega = \Omega_a$ seja a reta real e que $\mathcal{A} = \mathcal{A}_a$ seja a sigma-álgebra de Borel na reta. Seja

$$a := \{a_t(\omega), t \in T, \omega \in \Omega'\} \quad (2.1)$$

um processo estocástico em $(\Omega', \mathcal{A}', P')$ com espaço de estados $(\Omega_a, \mathcal{A}_a)$ e conjunto de índices T . Para cada $t \in T$, $a_t(\omega)$ é uma variável aleatória. A suposição a seguir valerá daqui por diante, mesmo sem menção explícita.

Suposição 2.1. *O processo estocástico definido em (2.1) tem mediana única nula $\text{med}(a_t) = 0$, qualquer que seja $t \in T$, bem como variância constante, ainda que não necessariamente finita, isto é, $\text{Var}(a_t) = \sigma_a^2$, qualquer que seja $t \in T$. Além disso, as variáveis aleatórias a_t são independentes e identicamente distribuídas.*

O processo estocástico definido em (2.1), com a suposição 2.1 satisfeita, é chamado *processo de ruído*, ou simplesmente *ruído*.

Seja

$$z := \{z_t(\omega), t \in T, \omega \in \Omega'\} \quad (2.2)$$

um processo estocástico em $(\Omega', \mathcal{A}', P')$ com espaço de estados (Ω, \mathcal{A}) e conjunto de índices T . A suposição a seguir valerá daqui por diante, exceto onde for explicitado em contrário.

Suposição 2.2. *O processo estocástico z definido em (2.2) foi gerado a partir do processo de ruído definido em (2.1) por meio de um modelo auto-regressivo e de médias móveis de ordens p e q , abreviadamente $ARMA(p, q)$, isto é,*

$$\phi(B) z_t(\omega) = \theta(B) a_t(\omega), \quad (2.3)$$

para $\omega \in \Omega'$ e $t \in T$, e onde

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

é um operador auto-regressivo de ordem p ,

$$\theta(B) := 1 - \theta_1 B - \dots - \theta_q B^q$$

é um operador de médias móveis de ordem q e B é o operador de atraso tal que

$$B z_t(\omega) := z_{t-1}(\omega).$$

Os modelos de equações de diferenças estocásticas conhecidos como modelos ARMA formam uma classe geral de modelos para séries temporais introduzida por Yule(1927) e Slutsky(1937) e popularizados por Box e Jenkins(1976).

Denotemos o espaço paramétrico do modelo em (2.3) por Υ e o vetor de parâmetros do modelo por $\beta' = (\phi' \theta')$, onde $\phi = (\phi_1 \dots \phi_p)'$ e $\theta = (\theta_1 \dots \theta_q)'$, com $\beta \in \Upsilon \subset \mathbf{R}^{p+q}$. Denotemos, também, $\xi' = (\beta' \sigma_a^2)$. Às vezes, a seguinte notação é útil:

$$\phi^*(B) = \phi(B) - 1$$

$$\theta^*(B) = \theta(B) - 1$$

$$\bar{\phi}_0 = \phi_0 = 1$$

$$\bar{\phi}_i = -\phi_i, \quad (i = 1, 2, \dots, p),$$

$$\bar{\theta}_0 = \theta_0 = 1$$

$$\bar{\theta}_j = -\theta_j, \quad (j = 1, 2, \dots, q).$$

No processo estocástico definido em (2.1), fixando-se $\omega \in \Omega'$, obtemos uma trajetória ou realização do processo de ruído, chamada *série de ruído*:

$$a(\omega) = \{a_t(\omega), t \in T\} := \{a_t, t \in T\}, \quad (2.4)$$

para cada $\omega \in \Omega'$ fixado. Analogamente, no processo estocástico definido em (2.2), fixando-se $\omega \in \Omega'$, obtemos uma trajetória ou realização do processo, chamada *série temporal*:

$$z(\omega) = \{z_t(\omega), t \in T\} := \{z_t, t \in T\}, \quad (2.5)$$

para cada $\omega \in \Omega'$ fixado.

Sob certas condições (Box e Jenkins, 1976), a série temporal em (2.5) pode ser representada como a saída de um filtro linear $\psi(B)$, cuja entrada é a série de ruído:

$$z_t = \psi(B) a_t, \quad (2.6)$$

onde

$$\psi(B) := \sum_{r=0}^{\infty} \psi_r B^r,$$

com $\psi_0 = 1$.

Do mesmo modo, sob certas condições podemos escrever:

$$\pi(B) z_t = a_t, \quad (2.7)$$

onde

$$\pi(B) := 1 - \sum_{r=1}^{\infty} \pi_r B^r := \sum_{r=0}^{\infty} \bar{\pi}_r B^r,$$

com $\bar{\pi}_0 = \pi_0 = 1$ e $\bar{\pi}_r = -\pi_r$, ($r = 1, 2, \dots$).

Os pesos ψ e π dependem de β e devem satisfazer a algumas condições, como será mostrado a seguir. Estamos especialmente interessados em encontrar condições sobre os parâmetros β e sobre o ruído a_t que assegurem o que se chama estacionariedade e inversibilidade (Box e Jenkins, 1976). Na teoria clássica de séries temporais (ver Anderson, 1970, ou Grenander e Rosenblat, 1957), supomos que as raízes das equações polinomiais

$$\bar{\phi}(x) = \sum_{i=0}^p \bar{\phi}_i x^i = 0 \quad (2.8)$$

e

$$\bar{\theta}(y) = \sum_{j=0}^q \bar{\theta}_j y^j = 0 \quad (2.9)$$

caiam fora do círculo unitário e que $E(a_t^2) < \infty$. Entretanto, tem se mostrado que alguns fenômenos, como é o caso de certas variáveis econômicas, são mais bem modelados por processos com variância infinita (Mandelbrot, 1963 e 1967; Granger e Orr, 1972; Gross e Steiger, 1979). Para um processo auto-regressivo puro, Yohai e Maronna(1977) mudaram a condição de variância finita para uma mais geral, a saber, $E(\log^+ |a_t|) < \infty$. Esta condição vale também para processos ARMA, como será mostrado no teorema que segue.

Teorema 2.3. *Suponhamos que valham as suposições 2.1 e 2.2 e que as raízes das equações polinomiais (2.8) e (2.9) caiam fora do círculo unitário. Então,*

a) *O processo de ruído ter distribuição simétrica satisfazendo*

$$E(\log^+ |a_t|) < \infty, \quad (2.10)$$

onde $\log^+ x$ denota $\max\{0, \log x\}$, é condição suficiente para que as séries $\psi(B)a_t$ e $\pi(B)z_t$ dadas em (2.6) e (2.7) converjam quase certamente;

b) *Se estas séries convergirem quase certamente, existirá um processo estacionário e inversível, quase certamente único, satisfazendo (2.3).*

Prova.

Podemos estudar a estacionariedade no processo

$$\bar{\phi}(B) z_t = e_t, \quad (2.11)$$

onde $e_t = \bar{\theta}(B)a_t$ é um processo MA(q) estacionário (Box e Jenkins, 1976, p.67), e a inversibilidade no processo

$$w_t = \bar{\theta}(B) a_t, \quad (2.12)$$

onde $w_t = \bar{\phi}(B)z_t$ é um processo inversível (Box e Jenkins, 1976, p.54), porque a condição de inversibilidade é independente da condição de estacionariedade (Box e Jenkins, 1976, p.50).

Iniciemos a demonstração pela parte (b), supondo que $\psi(B)a_t$ e $\pi(B)z_t$ convirjam quase certamente.

Para estudar a estacionariedade, usamos (2.11) recursivamente, obtendo após m substituições:

$$z_t = \sum_{i=1}^p \phi_i^*(m) z_{t-m-i} + \sum_{r=0}^m \delta_r e_{t-r}, \quad (2.13)$$

onde

$$\delta_0 = 1,$$

$$\delta_r = \phi_1^*(r-1) = - \sum_{i=1}^H \bar{\phi}_i \delta_{r-i}, \quad (r = 1, 2, \dots), \quad (2.14)$$

com $H = \min\{p, r\}$, e

$$\phi_i^*(m) = \phi_{i+1}^*(m-1) - \bar{\phi}_i \phi_1^*(m-1), \quad (i = 1, 2, \dots, p), \quad (2.15)$$

com $\phi_{p+1}^*(m) = 0$, para todo m .

Para simplificar a demonstração vamos supor que a equação $\phi(x) = 0$ não tenha raízes múltiplas. O resultado também é válido no caso de raízes múltiplas, fazendo-se algumas modificações nas provas (ver Anderson, 1971, p.172 e 249). Sejam x_1, x_2, \dots, x_p tais raízes. Então, para $\bar{\phi}_p \neq 0$,

$$\bar{\phi}(x) = \sum_{i=0}^p \bar{\phi}_i x^i = \prod_{i=1}^p (1 - x/x_i). \quad (2.16)$$

Por hipótese essas raízes caem fora do círculo unitário, isto é, $|x_i| > 1, (i = 1, 2, \dots, p)$.

Seja $\{\delta_r, r \geq 0\}$ uma seqüência tal que

$$\sum_{i=0}^p (\bar{\phi}_i x^i)^{-1} = \sum_{r=0}^{\infty} \delta_r x^r = \delta(x). \quad (2.17)$$

De (2.16) e (2.17) obtemos

$$\sum_{r=0}^{\infty} \delta_r x^r = \prod_{i=1}^p (1 - x/x_i)^{-1} = \prod_{i=1}^p \sum_{j=0}^{\infty} (x/x_i)^j. \quad (2.18)$$

As séries $\sum_j (x/x_i)^j$, ($i = 1, 2, \dots, p$), convergem se $|x/x_i| < 1$ para cada i . Então, elas convergem se $|x| < \min |x_i|$ e, pelo teorema de Abel (Hönig, 1971, p.49), para qualquer x tal que $|x| < \min |x_i|$ elas convergem absolutamente.

Podemos mostrar que os δ_r em (2.13) e em (2.17) são os mesmos (Anderson, 1971, p.169). Então, a primeira soma do lado direito de (2.13) converge para zero quando $m \rightarrow \infty$ e, portanto, $\phi_i^*(m)$ converge para zero para cada i .

Segue-se diretamente que a convergência de $\psi(B)a_t$ é equivalente à convergência de $\delta(B)e_t$.

Agora, vamos mostrar que existe uma seqüência estacionária que satisfaz (2.11). Seja \tilde{z}_t tal que

$$\tilde{z}_t = \sum_{r=0}^{\infty} \delta_r e_{t-r}, \quad (2.19)$$

onde $\delta(B)$ seja convergente, isto é, $\{\tilde{z}_t, t \in T\}$ seja gerada por um processo estacionário.

Usando (2.19) e depois (2.14), obtemos:

$$\begin{aligned} \bar{\phi}(B)\tilde{z}_t &= \sum_{i=0}^p \bar{\phi}_i \sum_{r=0}^{\infty} \delta_r e_{t-i-r} \\ &= e_t + \sum_{r=1}^{\infty} \delta_r e_{t-r} + \sum_{i=1}^p \bar{\phi}_i \sum_{j=0}^{\infty} \delta_j e_{t-i-j} \\ &= e_t + \sum_{r=1}^{\infty} \left(\delta_r + \sum_{i=1}^H \bar{\phi}_i \delta_{r-i} \right) e_{t-r} \\ &= e_t, \end{aligned} \quad (2.20)$$

onde H é como em (2.14).

Logo, a seqüência $\{z_t, t \in T\}$ é uma solução estacionária de (2.11) e, portanto, de (2.3).

Por outro lado, se $\{z_t^*, t \in T\}$ for uma seqüência estacionária que satisfaça (2.11), então, a primeira soma em (2.13) convergirá para zero quase certamente, de modo que a segunda soma deverá convergir quase certamente.

Quanto à inversibilidade, a prova da parte (b) é inteiramente análoga. Da mesma forma que em (2.13) escrevemos

$$a_t = \sum_{j=1}^q \theta_j^*(m) a_{t-m-j} + \sum_{r=0}^m \gamma_r w_{t-r}, \quad (2.21)$$

onde a primeira soma converge para zero quando $m \rightarrow \infty$, bem como $\theta_j^*(m)$ para cada j . A convergência de $\pi(B)z_t$ ou de $\bar{\pi}(B)z_t$ é equivalente à convergência de $\gamma(B)w_t$ para pesos γ adequados.

Em resumo, provamos que, sob a suposição de que as raízes das equações polinomiais (2.8) e (2.9) caíam fora do círculo unitário, a convergência quase certa de $\delta(B)e_t$ e de $\gamma(B)w_t$, ou, equivalentemente, de $\psi(B)a_t$ e de $\pi(B)z_t$, é necessária e suficiente para a existência de uma solução estacionária e inversível de (2.3). A unicidade da solução é dada pela convergência em probabilidade (Yohai e Maronna, 1977).

A fim de provar a convergência em probabilidade, consideremos

$$\begin{aligned} \psi(B)a_t &= \delta(B)e_t \\ &= \delta(B)\bar{\theta}(B)a_t \\ &= \sum_{r=0}^{\infty} \delta_r \sum_{j=0}^q \bar{\theta}_j a_{t-r-j} \\ &= \sum_{r=0}^{\infty} \sum_{j=0}^H \delta_{r-j} \bar{\theta}_j a_{t-r}, \end{aligned} \quad (2.22)$$

onde $H = \min\{q, r\}$. Igualando os coeficientes de a_t em (2.22), obtemos

$$\psi_r = \sum_{j=0}^H \delta_{r-j} \bar{\theta}_j, \quad (r = 0, 1, \dots) \quad (2.23)$$

Então,

$$\begin{aligned} \sum_{r=0}^m \psi_r a_{t-r} &= \sum_{r=0}^m \sum_{j=0}^H \delta_{r-j} \bar{\theta}_j a_{t-r} \\ &= \sum_{r=0}^m \delta_r \sum_{j=0}^q \bar{\theta}_j a_{t-r-j} - \sum_{j=1}^q \sum_{r=m-j+1}^m \delta_r \bar{\theta}_j a_{t-r-j}. \end{aligned} \quad (2.24)$$

Agora, seja $\varepsilon > 0$ qualquer. Então, usando (2.24) e (2.13),

$$\begin{aligned} P(|z_t - \sum_{r=0}^m \psi_r a_{t-r}| > \varepsilon) &= P(|z_t - \sum_{r=0}^{\infty} \delta_r e_{t-r} + \sum_{j=1}^q \sum_{r=m-j+1}^m \delta_r \bar{\theta}_j a_{t-r-j}| > \varepsilon) \\ &= P(|\sum_{i=1}^p \phi_i^*(m) z_{t-m+i} + \sum_{j=1}^q \sum_{r=m-j+1}^m \delta_r \bar{\theta}_j a_{t-r-j}| > \varepsilon). \end{aligned} \quad (2.25)$$

A primeira soma no lado direito da última igualdade em (2.25) converge para zero quanto $m \rightarrow \infty$, uma vez que $\phi_i^*(m)$ vai para zero para cada i . Além disso,

$$\sum_{r=0}^m \delta_r e_{t-r} = \sum_{j=0}^q \bar{\theta}_j \sum_{r=0}^m \delta_r a_{t-r-j}. \quad (2.26)$$

Esta soma convergirá quando $m \rightarrow \infty$ se e somente se $\sum_{r=0}^m \delta_r a_{t-r-j}$ convergir para cada $(j = 0, 1, \dots, q)$.

Aplicando o critério de Cauchy,

$$P(|\sum_{j=1}^q \sum_{r=m-j+1}^m \delta_r a_{t-r-j}| > \varepsilon) = P[|\sum_{j=1}^q (\sum_{r=0}^m \delta_r a_{t-r-j} - \sum_{r=0}^{m-j} \delta_r a_{t-r-j})| > \varepsilon] \quad (2.27)$$

converge para zero quando $m \rightarrow \infty$.

A prova de que

$$P(|a_t - \sum_{r=0}^m \bar{\pi}_r z_{t-r}| > \varepsilon)$$

converge para zero quando $m \rightarrow \infty$ é exatamente análoga.

Portanto,

$$z_t = \sum_{r=0}^{\infty} \psi_r a_{t-r}, \quad (2.28)$$

e

$$a_t = \sum_{r=0}^{\infty} \bar{\pi}_r z_{t-r} = z_t - \sum_{r=1}^{\infty} \pi_r z_{t-r} \quad (2.29)$$

no sentido de convergência em probabilidade. Isso completa a prova da parte (b).

Para demonstrar a parte (a), vejamos que a solução geral da equação de diferenças homogênea

$$\bar{\phi}(B)z_t = 0 \quad (2.30)$$

é dada por

$$\delta_r = \sum_{i=1}^p k_i x_i^r, (r = 0, 1, \dots), \quad (2.31)$$

onde os k_i 's dependem de ϕ e x_1, x_2, \dots, x_p são as raízes de $\bar{\phi}(x) = 0$ (Fuller, 1976, p.41-45).

Logo,

$$\begin{aligned} \sum_{r=0}^{\infty} \delta_r e_{t-r} &= \sum_{r=0}^{\infty} \left(\sum_{i=1}^p k_i x_i^r \right) \left(\sum_{j=0}^q \bar{\theta}_j a_{t-r-j} \right) \\ &= \sum_{j=0}^q \bar{\theta}_j \sum_{i=1}^p k_i \sum_{r=0}^{\infty} x_i^r a_{t-r-j} \\ &= \sum_{s=t-q}^t \bar{\theta}_{t-s} \sum_{i=1}^p k_i \sum_{r=0}^{\infty} x_i^r a_{s-r}. \end{aligned} \quad (2.32)$$

Para mostrar que (2.32) converge é suficiente mostrar que a série $\sum_{r=0}^{\infty} |a_r| |x^r|$ converge quase certamente para todo $|x| < 1$. Conforme Yohai e Maronna(1977), isso acontecerá se o processo de ruído tiver distribuição simétrica e satisfizer (2.10).

Quanto à inversibilidade, do mesmo modo que na estacionariedade, a solução geral da equação de diferenças homogênea

$$\bar{\theta}(B)a_t = 0 \quad (2.33)$$

é dada por

$$\gamma_r = \sum_{i=1}^q g_i y_i^r, (r = 0, 1, \dots), \quad (2.34)$$

onde os g_i 's dependem de θ e y_1, y_2, \dots, y_q são as raízes de $\bar{\theta}(y) = 0$.

Logo,

$$\begin{aligned} \sum_{r=0}^{\infty} \gamma_r w_{t-r} &= \sum_{r=0}^{\infty} \left(\sum_{i=1}^q g_i y_i^r \right) \left(\sum_{j=0}^q \bar{\theta}_j a_{t-r-j} \right) \\ &= \sum_{s=t-q}^t \bar{\theta}_{t-s} \sum_{i=1}^q g_i \sum_{r=0}^{\infty} y_i^r a_{s-r}. \end{aligned} \quad (2.35)$$

Mostrar que (2.35) converge é exatamente análogo a mostrar a convergência de (2.32), que segue de Yohai e Maronna(1977). Isso completa a prova da parte (a).

△

A suposição de que as raízes das equações polinomiais (2.8) e (2.9) caiam fora do círculo unitário delimita o espaço paramétrico Υ ao que se chama região de estacionariedade e inversibilidade.

Seja z um processo estacionário e inversível para o qual valha a suposição 2.2. Definimos, então, um processo de resíduo associado ao processo estocástico z por

$$r(\mathbf{b}) := r(\mathbf{b}, z) := \{r_t(\omega, \mathbf{b}), \omega \in \Omega', \mathbf{b} \in \Upsilon\}, \quad (2.36)$$

como sendo um processo estocástico em $(\Omega', \mathcal{A}', P')$ com espaço de estados $(\Omega_a, \mathcal{A}_a)$ e conjunto de índices T , gerado a partir do processo z por meio de um modelo ARMA(p,q), isto é,

$$c(B) z_t(\omega) = d(B) r_t(\omega, \mathbf{b}), \quad (2.37)$$

onde $\mathbf{b}' = (\mathbf{c}', \mathbf{d}')$. Como $\mathbf{b} \in \Upsilon$, em condições análogas às do teorema 2.3 podemos escrever

$$\begin{aligned} r_t(\omega, \mathbf{b}) &= c(B) d(B)^{-1} \theta(B) \phi(B)^{-1} a_t(\omega) \\ &= \eta(B) a_t(\omega), \end{aligned} \quad (2.38)$$

onde os pesos η dependem de \mathbf{b} e de $\boldsymbol{\beta}$, e o processo $r(\mathbf{b})$ será quase certamente único. Nestas condições, fixado $t \in T$,

$$r_t(\boldsymbol{\beta}) = a_t \quad q.c. \quad (2.39)$$

Por outro lado, fixando-se $\omega \in \Omega'$, obtemos uma trajetória ou realização do processo, chamada *série de resíduo*:

$$r(\omega) = \{r_t(\omega, \mathbf{b}), t \in T\} := \{r_t(\mathbf{b}), t \in T\}, \quad (2.40)$$

para cada $\omega \in \Omega'$ fixado.

Uma propriedade fundamental na construção do nosso problema é a ergodicidade, dada no lema seguinte.

Lema 2.4. *Sob as suposições 2.1 e 2.2 e sob a suposição de estacionariedade e inversibilidade, os processos estocásticos a , z e $r(\mathbf{b})$, definidos, respectivamente, em (2.1), (2.2) e (2.36), são ergódicos.*

Prova.

O processo de ruído é ergódico porque os a_t 's são independentes e identicamente distribuídos, o que segue da lei 0-1 de Kolmogorov (Gnedenko, 1962, p.120).

Como z é estacionário, vale (2.6). Uma vez que a ergodicidade é preservada quando se tomam funções do processo (Breiman, 1968, p.119), segue-se que z é ergódico. Analogamente, como vale (2.38), o processo de resíduo também é ergódico.

\triangle

Uma consequência da estacionariedade e da ergodicidade do processo z é que podemos estimar o vetor de parâmetros β usando apenas uma trajetória, isto é, podemos concentrar nossas atenções sobre uma série temporal.

2.2. O PROBLEMA DE ESTIMAÇÃO

Uma suposição natural na estimação L_1 é que o processo definido em (2.2) seja um processo L_1 , isto é, que

$$\begin{aligned} \|z_t\| &:= \int_{\Omega'} |z_t(\omega)| P'(d\omega) \\ &= \int_{\Omega} |z_t| P(dz_t) \\ &= E|z_t| < \infty, \end{aligned} \tag{2.41}$$

para cada $t \in T$, e onde $\|\cdot\|$ é a norma L_1 .

Então, dado um processo L_1 definido como em (2.2) e satisfazendo a suposição 2.2, o problema de estimação L_1 dos parâmetros do modelo ARMA consiste em encontrar $\hat{\beta} \in \Upsilon$ que minimize a norma L_1 do processo de resíduo definido em (2.36):

$$\begin{aligned} \|r_t(\mathbf{b})\| &= E|r_t(\mathbf{b})| \\ &= \int_{\Omega} |r_t(\mathbf{b})| P(dz_t), \end{aligned} \tag{2.42}$$

para cada $t \in T$, isto é,

$$\|r_t(\hat{\beta})\| = \inf_{\mathbf{b} \in \Upsilon} \|r_t(\mathbf{b})\|. \tag{2.43}$$

Vamos supor que $\hat{\beta}$ seja quase certamente o único ponto de mínimo de (2.42).

Na prática, a medida de probabilidade P é desconhecida e o problema de estimação dado em (2.42) deve ser aproximado pelo problema de encontrar $\hat{\beta}_n \in \Upsilon$ que minimize

$$\begin{aligned} E^n|r_t(\mathbf{b})| &:= \int_{\Omega} |r_t(\mathbf{b})| P^n(dz_t) \\ &= \sum_{t=1}^n |r_t(\mathbf{b})|/n. \end{aligned} \tag{2.44}$$

Nestas condições, definimos a soma em L_1 dos resíduos amostrais, ou soma de desvios absolutos, ou *função objetivo*, como sendo a função dada por

$$S_n(\mathbf{b}) = \sum_t |r_t(\mathbf{b})|, \mathbf{b} \in \Upsilon \tag{2.45}$$

e o erro médio em L_1 do resíduo amostral, ou erro médio absoluto dos desvios, com sendo a função dada por

$$f_n(\mathbf{b}) = S_n(\mathbf{b})/n, \mathbf{b} \in \Upsilon. \tag{2.46}$$

Então, minimizar (2.44) é equivalente a minimizar (2.45) ou (2.46).

O conjunto de minimizadores é o subconjunto do espaço paramétrico dado por

$$\begin{aligned} M_n(\beta) &:= \operatorname{argmin} f_n(.) \\ &:= \{\hat{\beta}_n \in \Upsilon : f_n(\hat{\beta}_n) = \inf_{\mathbf{b} \in \Upsilon} f_n(\mathbf{b})\} \\ &= \operatorname{argmin} S_n(.) \\ &= \operatorname{argmin} E^n |r_t(\mathbf{b})|. \end{aligned} \tag{2.47}$$

O problema de estimação L_1 dado em (2.44) é equivalente ao seguinte problema de otimização : queremos encontrar um vetor $\mathbf{b}' = (\mathbf{c}' \mathbf{d}')$ que minimize a função objetivo (2.45) sujeita às restrições

$$c(B)z_t - d(B)r_t(\mathbf{b}) = 0, (t = 1, 2, \dots, n), \tag{2.48}$$

onde

$$\begin{aligned} \mathbf{b} &\in \Upsilon, \\ r_t(\mathbf{b}) &= r_t(\mathbf{b})^+ - r_t(\mathbf{b})^- \\ |r_t(\mathbf{b})| &= r_t(\mathbf{b})^+ + r_t(\mathbf{b})^- \\ r_t(\mathbf{b})^+, r_t(\mathbf{b})^- &\geq 0. \end{aligned}$$

Num processo puramente auto-regressivo os resíduos amostrais são lineares nos parâmetros, exceto pelo efeito de valores iniciais. Entretanto, quando há termos de médias móveis, os resíduos amostrais são sempre funções não lineares dos parâmetros (Box e Jenkins, 1976, p.232). Portanto, o problema de estimação L_1 dos parâmetros de um processo ARMA é equivalente a um problema de programação não linear.

A convexidade da função objetivo desempenha papel importante na solução de problemas de programação não linear, bem como na demonstração de propriedades assintóticas na estimação L_1 dos parâmetros de regressões e auto-regressões (An e Chen, 1982; Bloomfield e Steiger, 1983; Gross e Steiger, 1979). Todavia, a não linearidade dos resíduos conduz à não convexidade da função objetivo no caso de processos ARMA (Rockafellar, 1983, p.369). O caso de um modelo puramente auto-regressivo é considerado na seguinte proposição.

Proposição 2.5. *Suponhamos que o processo definido em (2.2) tenha sido gerado por um processo $AR(p)$ e que se disponham de $n + p$ observações de uma trajetória do processo. Então, se a função objetivo em (2.45) for calculada sobre as n últimas observações, ela será uma função convexa em Υ .*

Prova.

Seja $\lambda \in [0, 1]$, $\mathbf{b} \in \Upsilon$ e $\mathbf{x} \in \Upsilon$. Usando a desigualdade triangular, temos

$$\begin{aligned}
 S_n[\lambda \mathbf{b} + (1 - \lambda)\mathbf{x}] &= \sum_{t=1}^n |[\lambda \mathbf{b} + (1 - \lambda)\mathbf{x}](B)z_t| \\
 &= \sum_t |(\lambda \mathbf{b})(B)z_t + [(1 - \lambda)\mathbf{x}](B)z_t| \\
 &\leq \lambda \sum_t |\mathbf{b}(B)z_t| + (1 - \lambda) \sum_t |\mathbf{x}(B)z_t| \\
 &= \lambda S_n(\mathbf{b}) + (1 - \lambda)S_n(\mathbf{x})
 \end{aligned} \tag{2.49}$$

e, portanto, S_n é convexa em Υ .

△

Entretanto, se no lema 2.5 as $n + p$ observações forem utilizadas no cálculo de S_n , os p primeiros resíduos amostrais precisarão ser obtidos de outro modo que não diretamente e S_n deixará de ser convexa devido a esses valores iniciais.

Consideremos o caso em que o ruído tenha distribuição de Laplace (Damsleth e El-Shaarawi, 1989). A proposição a seguir relaciona a estimação L_1 com a estimação de máxima verossimilhança condicional aos valores iniciais da série temporal e da série de ruído.

Proposição 2.6. *Seja o processo definido em (2.2) sob as suposições 2.1 e 2.2. Suponhamos, ainda, que o processo de ruído tenha distribuição de Laplace (exponencial dupla) e que estejam disponíveis p valores $\mathbf{z}_* = (z_{1-p} \dots z_0)'$ da série temporal, bem como q valores $\mathbf{a}_* = (a_{1-q} \dots a_0)'$ da série de ruído, anteriores ao começo das n observações da amostra $\mathbf{z}_n = (z_1 \dots z_n)'$. Nestas condições, a estimação L_1 será equivalente à estimação de máxima verossimilhança condicional aos valores iniciais.*

Prova.

Seja

$$f(x) = [\exp(-|x|/\sigma)]/(2\sigma), \quad (2.50)$$

com $x \in \mathbf{R}$ e $\sigma > 0$, a função densidade de probabilidade do processo de ruído, onde

$$\text{med}(a_t) = E(a_t) = 0 \quad (2.51)$$

e

$$\text{Var}(a_t) = 2\sigma^2. \quad (2.52)$$

O logaritmo da verossimilhança condicional à escolha de $(\mathbf{z}'_* \mathbf{a}'_*)'$ é dado por

$$L(\boldsymbol{\beta}, \sigma) = -n \log 2 - n \log \sigma - S_*(\boldsymbol{\beta})/\sigma, \quad (2.53)$$

onde

$$S_*(\boldsymbol{\beta}) = \sum_{t=1}^n |r_t(\boldsymbol{\beta})| \quad (2.54)$$

e $r_t(\boldsymbol{\beta})$ depende de \mathbf{z}_* , de \mathbf{a}_* e de \mathbf{z}_n .

Diferenciando-se (2.53) em relação a σ e igualando-se a zero obtemos o estimador de máxima verossimilhança condicional de σ :

$$\hat{\sigma} = S_*(\beta)/n. \quad (2.55)$$

Substituindo-se em (2.53), obtemos

$$L(\beta) = -n \log 2 - n - n \log[S_*(\beta)/n]. \quad (2.56)$$

Portanto, minimizar a soma em L_1 dos resíduos $S_*(\beta)$ é equivalente a maximizar o logaritmo da função de verossimilhança $L(\beta, \sigma)$.

△

Na prática, vamos usar $r_t(\mathbf{b})$ no lugar de a_t , para $(t = 1 - q, \dots, 0)$. Para obter esses valores iniciais de $r_t(\mathbf{b})$ podemos adotar dois procedimentos:

a) tomar

$$r_{nt}(\mathbf{b}) = \text{med}(a_t) = 0, \quad (2.57)$$

como sugerido por Åström e Bohlin(1966) para o caso da estimação L_2 (Osborn, 1976);

b) tomar

$$\begin{aligned} r_{nt}(\mathbf{b}) &= E[r_t(\mathbf{b}) \mid \mathbf{z}_n] \\ &= E[r_t(\mathbf{b}) \mid \mathcal{F}^n], \end{aligned} \quad (2.58)$$

onde \mathcal{F}^n é a sigma-álgebra gerada pela amostra \mathbf{z}_n . Neste caso, os resíduos amostrais podem ser calculados pelo procedimento de *backforecasting* sugerido por Box e Jenkins(1976) para o caso da estimação L_2 .

Para um modelo com termos de médias móveis, (2.58) é melhor do que (2.57), embora com gasto computacional maior, porque a aproximação (2.57) pode ser pobre se algumas das raízes de (2.8) estiverem próximas da fronteira do espaço paramétrico (Box e

Jenkins, 1976). Um procedimento econômico consiste em usar (2.57) nos estágios iniciais da estimação e (2.58) somente no último estágio.

Por outro lado, num modelo puramente auto-regressivo é mais interessante usar (2.57), reservando as primeiras p observações como os valores iniciais z_* , porque a utilização do procedimento de *backforecasting* tornaria a função objetivo não convexa e o modelo não linear. É interessante notar que os resultados de consistência em auto-regressões com variância infinita obtidos por Kanter e Steiger(1974) e por Yohai e Maronna(1977) para o caso de estimação L_2 , bem como os obtidos por Gross Steiger(1979) e por An e Chen(1982) para o caso de estimação L_1 , não valerão se utilizarmos o procedimento de *backforecasting*.

Finalmente, embora não seja necessário no presente trabalho, convém mencionar que a proposição 2.6 pode ser estendida para o caso mais geral (Nyquist, 1987) de estimação L_p em que o processo de ruído tenha distribuição com função densidade de probabilidade dada por

$$f(x) = p[\exp(-|x - \mu|^p/\sigma^p)]/[2\sigma\Gamma(1/p)]. \quad (2.59)$$

2.3. CONSISTÊNCIA

Nesta seção mostraremos que o problema da estimação L_1 dos parâmetros de um modelo ARMA satisfaz as suposições de Dupačová e Wets(1988) e, portanto, a consistência do estimador pode ser demonstrada. Iniciaremos mostrando a continuidade de algumas funções.

Lema 2.7. *Suponhamos que o processo definido em (2.2) seja estacionário e inversível e que satisfaça as suposições 2.1 e 2.2. Então, fixados $\omega \in \Omega$ e $t \in T$, as seguintes funções são inferiormente semicontínuas em Υ :*

- a) $r_t(\omega, \mathbf{b})$, dada em (2.36);
- b) $g_t(\mathbf{b}) = |r_t(\mathbf{b})|$, dada em (2.42);
- c) $S_n(\mathbf{b})$, dada em (2.45);
- d) $f_n(\mathbf{b})$, dada em (2.46).

Prova.

Basta mostrar que essas funções são contínuas em Υ , uma vez que uma função é contínua num ponto se e somente se ela for semicontínua inferiormente e superiormente nesse ponto (Royden, 1968, p.49, item b).

Seja $\mathbf{b}' = (\mathbf{c}' \ \mathbf{d}') \in \Upsilon$ e tomemos $\boldsymbol{\delta}' = (\mathbf{u}' \ \mathbf{v}')$ tal que $\mathbf{b} + \boldsymbol{\delta} \in \Upsilon$. Como o processo é estacionário e inversível, fixado $t \in T$ na série de resíduo, temos:

$$|r_t(\mathbf{b} + \boldsymbol{\delta}) - r_t(\mathbf{b})| = |(c + u)(B) (d + v)(B)^{-1} z_t - c(B) d(B)^{-1} z_t| \quad (2.60)$$

e isso vai para zero quando $\boldsymbol{\delta} \rightarrow \mathbf{0}$. Portanto, para cada $\varepsilon > 0$ é possível encontrar um número M_ε tal que

$$|r_t(\mathbf{b} + \boldsymbol{\delta}) - r_t(\mathbf{b})| < \varepsilon \quad (2.61)$$

para todo $\boldsymbol{\delta}$ tal que $|u_i| < M_\varepsilon$ e $|v_j| < M_\varepsilon$, ($i = 1, 2, \dots, p$) e ($j = 1, 2, \dots, q$) e segue-se a continuidade de $r_t(\omega, \mathbf{b})$.

A continuidade de $g_t(\mathbf{b})$ segue-se diretamente da continuidade de $r_t(\omega, \mathbf{b})$.

A continuidade da função objetivo segue-se da desigualdade triangular e da continuidade de $r_t(\mathbf{b})$:

$$\begin{aligned} |S_n(\mathbf{b} + \delta) - S_n(\mathbf{b})| &= \left| \sum_t |r_t(\mathbf{b} + \delta)| - \sum_t |r_t(\mathbf{b})| \right| \\ &\leq \sum_t \left| |r_t(\mathbf{b} + \delta)| - |r_t(\mathbf{b})| \right|. \end{aligned} \quad (2.62)$$

A continuidade de $f_n(\mathbf{b})$ segue-se diretamente da continuidade de $S_n(\mathbf{b})$.

△

A continuidade quase certa de $g_t(\mathbf{b}) = |r_t(\mathbf{b})|$ em Ω é trivial.

O teorema a seguir trata de algumas propriedades da região de estacionariedade e inversibilidade.

Teorema 2.8. *A região de estacionariedade e inversibilidade de um processo ARMA(p,q) é um conjunto limitado e aberto.*

Prova.

Inicialmente, vamos mostrar que Υ é limitado. Se $G_i^{-1}, (i = 1, 2, \dots, p)$, forem as raízes de $\phi(B) = 0$ e $H_j^{-1}, (j = 1, 2, \dots, q)$, forem as raízes de $\theta(B) = 0$, então,

$$\phi(B) = \prod_{i=1}^p (1 - G_i B) \quad (2.63)$$

e

$$\theta(B) = \prod_{j=1}^q (1 - H_j B) \quad (2.64)$$

A região de estacionariedade e inversibilidade é dada por

$$\begin{aligned} \Upsilon &= \Upsilon(p, q) \\ &= \{ \mathbf{b} \in \mathbf{R}^{p+q} : |G_i| < 1 \text{ e } |H_j| < 1, (i = 1, \dots, p) \text{ e } (j = 1, \dots, q) \} \end{aligned} \quad (2.65)$$

Igualando-se os coeficientes das potências de B em (2.63) obtemos

$$\phi_i = (-1)^{i+1} \sum_{j_1=1}^{p-i+1} \sum_{j_2=j_1+1}^{p-i+2} \dots \sum_{j_i=j_{i-1}+1}^p G_{j_1} \dots G_{j_i}, \quad (2.66)$$

para $(i = 1, 2, \dots, p)$, que é uma soma de $\binom{p}{i}$ parcelas. Como $|G_{j_i}| < 1, (i = 1, 2, \dots, p)$, então,

$$|\phi_i| \leq \sum_{j_1} \sum_{j_2} \dots \sum_{j_i} |G_{j_1} \dots G_{j_i}| < \binom{p}{i}, \quad (2.67)$$

para $(i = 1, 2, \dots, p)$.

Analogamente, obtemos

$$|\theta_j| < \binom{q}{j}, (j = 1, 2, \dots, q) \quad (2.68)$$

Então, existe uma bola de dimensão $p + q$, com centro na origem e raio igual a $\max\{\max_i \binom{p}{i}, \max_j \binom{q}{j}\}$ que contém Υ . Logo, Υ é limitado.

Mostremos, agora, que Υ é aberto. Em (2.66), chamemos

$$g_i(G_1, \dots, G_p) = \sum_{j_1} \dots \sum_{j_i} G_{j_1} \dots G_{j_i} \quad (2.69)$$

e escrevamos

$$\phi_i = (-1)^{i+1} g_i(G_1, \dots, G_p), \quad (2.70)$$

de onde resulta o sistema de equações

$$\phi = T'G, \quad (2.71)$$

com

$$\phi = (\phi_1 \dots \phi_p)'$$

$$T = \{(-1)^i\}$$

$$G = \{g_i(G_1, \dots, G_p)\}.$$

Dados os valores de $p - 1$ parâmetros ϕ , estarão determinadas $p - 1$ raízes em (2.70). Logo, dados ϕ_2, \dots, ϕ_p em Υ , quaisquer, obtemos os correspondentes valores para as raízes $G_2 = g_2, \dots, G_p = g_p$. Então, em (2.70),

$$|\phi_1| = |G_1 + \sum_{i=2}^p g_i|, \quad (2.72)$$

isto é, dados os demais parâmetros, ϕ_1 variará num intervalo aberto, simétrico ao redor da origem, contido no intervalo $(-p, p)$.

Analogamente, dados $\theta_2, \dots, \theta_q$, então, θ_1 variará num intervalo aberto, simétrico ao redor da origem, contido no intervalo $(-q, q)$.

De (2.65) segue-se que

$$\Upsilon(p, q) = \Upsilon_\phi(p) \times \Upsilon_\theta(q), \quad (2.73)$$

onde

$$\Upsilon_\phi(p) = \{\phi \in \mathbf{R}^p : |G_i| < 1, (i = 1, \dots, p)\} \quad (2.74)$$

e

$$\Upsilon_\theta(q) = \{\theta \in \mathbf{R}^q : |H_j| < 1, (j = 1, \dots, q)\} \quad (2.75)$$

$\Upsilon_\phi(p)$ é a reunião dos intervalos abertos gerados por (2.72) e, portanto, é aberto. Analogamente, $\Upsilon_\theta(q)$ é aberto e segue-se o resultado.

△

Uma consequência imediata do teorema 2.8 é que o fecho $\tilde{\Upsilon}$ do espaço paramétrico é um conjunto compacto. Como β é sempre um ponto interior de $\tilde{\Upsilon}$, evitamos qualquer falta de regularidade introduzida por restrições com desigualdades (Dupačová, 1987; Oberhofer, 1982).

Lema 2.9. *Suponhamos que o processo definido em (2.2) seja um processo L_1 , estacionário e inversível, e que satisfaça as suposições 2.1 e 2.2. Então, a função $g_t(\mathbf{b}) = |r_t(\mathbf{b})|$ dada em (2.42) é localmente inferiormente lipschitziana em Υ .*

Prova.

Tomemos $\mathbf{b} \in \Upsilon$ qualquer e δ tal que $\mathbf{b} + \delta \in \Upsilon$. Então,

$$\begin{aligned} |g_t(\mathbf{b}) - g_t(\mathbf{b} + \delta)| &\leq |g_t(\mathbf{b})| \\ &= g_t(\mathbf{b}) \\ &\leq \sup_{\mathbf{b} \in \Upsilon} g_t(\mathbf{b}). \end{aligned} \tag{2.76}$$

De (2.67) e (2.68),

$$|b_i| < \binom{r}{i} \leq \binom{r}{[r/2]}, \quad (i = 1, \dots, p + q) \tag{2.77}$$

onde $r = \max\{r, p\}$ e $[r/2]$ indica o maior inteiro contido em $r/2$. O mesmo vale para cada $b_i + \delta_i$. Logo,

$$\begin{aligned} \|\delta\| &= \sum_{i=1}^{p+q} |\delta_i| \\ &= \sum_i |(b_i + \delta_i) - b_i| \\ &\leq \sum_i (|b_i + \delta_i| + |b_i|) \\ &< 2 \sum_i \binom{r}{[r/2]} \\ &= 2(p + q) \binom{r}{[r/2]}. \end{aligned} \tag{2.78}$$

Fazendo

$$K_n(z) = \sup_{\mathbf{b} \in \Upsilon} g_t(\mathbf{b}) [2(p + q) \binom{r}{[r/2]}]^{-1}, \tag{2.79}$$

temos que

$$\sup_{\mathbf{b} \in \Upsilon} g_t(\mathbf{b}) \leq K_n(z) \|\delta\| \tag{2.80}$$

e, portanto,

$$|g_t(\mathbf{b}) - g_t(\mathbf{b} + \delta)| \leq K_n(z) \|\delta\|. \tag{2.81}$$

A função $K_n(z)$ é integrável, uma vez que $r_t(\mathbf{b})$ o é, o que, por sua vez, segue-se de $z \in L_1$. Então, segue-se o resultado.

△

A convergência do erro médio em L_1 dos resíduos é mostrada a seguir.

Teorema 2.10. *Suponhamos que o processo definido em (2.2) seja um processo L_1 , estacionário e inversível, e que satisfaça as suposições 2.1 e 2.2. Sejam as funções f_n e r_t dadas por (2.46) e (2.42), respectivamente. Então,*

$$f_n(\mathbf{b}) \longrightarrow E|r_1(\mathbf{b})| \equiv f(\mathbf{b}) \text{ q.c.} \quad (2.82)$$

quando $n \rightarrow \infty$.

Prova.

O espaço das trajetórias ou funções amostrais do processo estocástico z é dado pelo espaço infinito-dimensional \mathbf{R}^∞ que consiste em todas as seqüências infinitas de reais $\mathbf{x} = (x_1, x_2, \dots)$, sendo \mathcal{B}_∞ a sigma-álgebra de Borel correspondente. Consideremos, inicialmente, as séries temporais $z(\omega)$ dadas em (2.5) e as séries de resíduos $r(\omega)$ dadas em (2.40), com $\omega \in \Omega'$. Para todo $X \in \mathcal{B}_\infty$ temos que

$$A_X = \{\omega : r(\omega) \in X\} \in \mathcal{A}'$$

(Breiman, 1968, proposição 2.13). Porém como

$$\{z(\omega) : \omega \in A_X\} \in \mathcal{B}_\infty,$$

segue-se que a transformação do processo z para o processo $r(\mathbf{b})$ dada em (2.37) é mensurável em relação a \mathcal{B}_∞ . Logo, o processo de resíduo $r(\mathbf{b})$ é estacionário (Breiman, 1968, proposição 6.6), sendo fácil mostrar que $|r(\mathbf{b})|$ também o é e, portanto, pelo teorema

ergódico (Breiman, 1968, p.118),

$$\begin{aligned} f_n(\mathbf{b}) &= n^{-1} \sum_t |r_t(\mathbf{b})| \longrightarrow E[|r_1(\mathbf{b})| | \mathcal{I}] q.c. \\ &= E|r_1(\mathbf{b})| \text{ q.c.} \end{aligned} \tag{2.83}$$

onde \mathcal{I} é a sigma-álgebra dos eventos invariantes e a última igualdade segue-se do fato de os processos serem ergódicos.

△

Então, em (2.42) e (2.44) temos

$$E^n|r_t(\mathbf{b})| \xrightarrow{q.c.} E|r_t(\mathbf{b})| < \infty. \tag{2.84}$$

Este resultado é equivalente à integrabilidade uniforme de $|r_t(\mathbf{b})|$ em P^n , para $\mathbf{b} \in \Upsilon$, o que, por sua vez, significa que, quaisquer que sejam $\mathbf{b} \in \Upsilon$ e $\varepsilon > 0$, existe um conjunto compacto $K_\varepsilon \subset \Omega$ tal que

$$\int_{\Omega \setminus K_\varepsilon} |r_t(\mathbf{b})| P^n(dz_t) < \varepsilon, \quad (n = 1, 2, \dots), \tag{2.85}$$

e

$$\int_{\Omega \setminus K_\varepsilon} |r_t(\mathbf{b})| P(dz_t) < \varepsilon, \tag{2.86}$$

(Dupačová, 1987, p.5).

A seguir apresenta-se o resultado principal deste capítulo. Seja $(\mathbf{R}^\infty, \mathcal{B}_\infty, \mu)$ o espaço de probabilidade das funções amostrais do processo z . Uma função amostral ζ leva-nos a uma seqüência crescente $\{\mathcal{F}^n(\cdot, \zeta), n = 1, 2, \dots\}$ de sigma-álgebras contidas em \mathcal{B}_∞ tais que para todo $A \in \mathcal{A}$ temos $P^n(A, \zeta)$ mensurável em relação a \mathcal{F}^n .

Teorema 2.11. *Suponhamos que o processo definido em (2.2) seja um processo L_1 , estacionário e inversível, e que satisfaça as suposições 2.1 e 2.2. Então,*

$$\hat{\beta}_n \longrightarrow \beta \quad (2.87)$$

quando $n \rightarrow \infty$, para quase todas (μ) as amostras ζ em \mathbf{R}^∞ , e para toda seleção $\{\hat{\beta}_n, n \geq 1\}$, mensurável em relação a \mathcal{F}^n , de $\{\operatorname{argmin} E^n g, n \geq 1\}$.

Prova.

De (2.47) segue-se que $\operatorname{argmin} E^n g \subset \Upsilon$ e $\operatorname{argmin} Eg \subset \Upsilon$. Como consequência do teorema 2.8 podemos tomar $D = \tilde{\Upsilon} \subset \mathbf{R}^{p+q}$ compacto e tal que

$$(\operatorname{argmin} E^n g) \cap D = \operatorname{argmin} E^n g \neq \emptyset \quad (2.88)$$

e

$$\{\beta\} = (\operatorname{argmin} Eg) \cap D = \operatorname{argmin} Eg, \quad (2.89)$$

isto é, β é o único ponto de $\operatorname{argmin} Eg$.

Os lemas 2.7 e 2.9 e o teorema 2.10 completam as suposições de Dupačová e Wets (1988). Logo, o teorema da consistência desses autores vale em nosso caso para toda seleção mensurável e segue-se o resultado.

\triangle

2.4. OBSERVAÇÕES SOBRE A DISTRIBUIÇÃO ASSINTÓTICA

Um resultado genérico de normalidade assintótica das estimativas foi apresentado por Dupačová e Wets (1988). O fato de em nosso caso as estimativas serem pontos interiores do espaço amostral (teorema 2.8) permite que nosso problema seja tratado como um problema de estimação sem restrições. Entretanto, dadas as dificuldades envolvidas na demonstração no caso não linear (Dupačová, 1987), consideremos apenas uma simplificação do problema.

Como será visto no capítulo 3, na prática as estimativas são obtidas iterativamente num modelo linearizado. Assim, na j -ésima iteração obtemos a estimativa

$$\hat{\beta}_{nj} = \hat{\beta}_{n,j-1} + \hat{\delta}_{nj} \quad (2.93)$$

sendo $\hat{\delta}_{nj}$ a estimativa obtida no modelo de regressão linear

$$r_{nt}(\hat{\beta}_{n,j-1}) = \mathbf{x}'_{tj} \delta_{nj} + r_{nt}(\beta), \quad (2.94)$$

onde \mathbf{x} é um vetor de derivadas dadas por (3.2). Podemos escrever na forma matricial

$$\mathbf{r} = \mathbf{X}'\delta + \mathbf{e} \quad (2.95)$$

Mesmo nesse caso simplificado, o resultado de Dupačová (1987) não pode ser aplicado porque a densidade do resíduo $r_{nt}(\beta)$ não é independente da densidade de \mathbf{X} , uma vez que \mathbf{X} é a matriz das derivadas do resíduo.

Outra possibilidade a ser considerada é o caso particular de normalidade assintótica quando as estimativas L_1 forem equivalentes às de máxima verossimilhança (proposição 2.6).

Assim, voltaremos a considerar a questão da distribuição assintótica no estudo de Monte Carlo do capítulo 4.

CAPÍTULO 3

ALGORITMOS

3.1. INTRODUÇÃO

Uma vez estudadas algumas propriedades do estimador nosso interesse volta-se para a obtenção de algoritmos eficientes para o cálculo das estimativas. O aumento do uso de computadores nas últimas décadas possibilitou a difusão de métodos estatísticos que envolvem longas seqüências de cálculos repetitivos, como é o caso da estimação L_1 em modelos ARMA.

Gonin e Money(1987) apresentaram uma revisão dos algoritmos para estimação L_1 não linear, classificando-os em três categorias:

a)“Aqueles que usam somente informação da derivada primeira. Estes métodos são do tipo Gauss-Newton. O problema não linear original é reduzido a uma seqüência de problemas L_1 lineares, cada um dos quais pode ser resolvido eficientemente como um problema de programação linear usual.” Os autores incluíram nesta categoria os algoritmos de Osborne e Watson, de Anderson e Osborne, de Anderson-Osborne-Levenberg-Marquardt, e de McLean e Watson.

b)“Aqueles que usam informação da derivada segunda. A não diferenciabilidade da função objetivo é superada transformando-se o problema L_1 original numa seqüência

de problemas de minimização sem restrições, cada um dos quais requer solução. Isto é conhecido como método de funções de penalidade da programação linear". Os autores incluíram nesta categoria os algoritmos de El-Attar-Vidyasagar-Dutta e de Tishler e Zang.

c) "Aqueles que linearizam as funções do modelo, mas, incorporam aproximações quadráticas para levar em conta efeitos de curvatura". Os autores incluíram nesta categoria os algoritmos de Murray e Overton e o de Bartels e Conn, bem como o enfoque de Busovača.

Uma condição necessária para um ponto em Υ ser ponto de mínimo é dado no lema seguinte.

Suposição 3.1. Os resíduos $r_{nt}(\mathbf{b}^*)$, $\mathbf{b}^* \in \Upsilon$, são continuamente diferenciáveis em \mathbf{b}^* .

Lema 3.2. Sob as suposições 2.1, 2.2 e 3.1, uma condição necessária para $\mathbf{b}^* \in \Upsilon$ ser um mínimo local é que existam constantes (multiplicadores) $v_t \in [-1, 1]$, $t \in A$, tais que

$$\sum_{t \notin A} \text{sign}[r_{nt}(\mathbf{b}^*)] \nabla r_{nt}(\mathbf{b}^*) + \sum_{t \in A} v_t \nabla r_{nt}(\mathbf{b}^*) = 0, \quad (3.1)$$

onde

$$A = A(\mathbf{b}^*) = \{t \mid r_{nt}(\mathbf{b}^*) = 0, t \in T, \mathbf{b}^* \in \Upsilon\}$$

e $\nabla r_{nt}(\mathbf{b}^*)$ é o gradiente de r_{nt} em \mathbf{b}^* .

A prova consiste numa aplicação das condições de Kuhn-Tucker e segue-se imediatamente dos teoremas 1 e 2 de Charalambous(1979) ou do lema 2.1 de El-Attar et alii(1979). Ver também Gonin e Money(1987). Note-se que se r_{nt} fosse uma função linear em Υ , então esta condição seria também suficiente.

3.2. ALGORITMOS PARA A ESTIMAÇÃO L_1 EM MODELOS ARMA

Nesta seção discutimos o uso de alguns algoritmos iterativos disponíveis para a estimação L_1 dos parâmetros de um processo ARMA. Em todos estes algoritmos um critério de convergência a fim de parar as iterações tem de ser estabelecido.

Os resíduos podem ser estimados por procedimentos condicionais ou, mais eficientemente, pelos procedimentos de *backforecasting*, como em Box e Jenkins(1976). Como foi visto no capítulo 2, as estimativas dos parâmetros baseadas nos resíduos de *backforecasting* são consistentes. Além disso, tais resíduos satisfazem a suposição 3.1.

Sob as suposições 2.1 e 2.2, a linearização aproximada do modelo pode ser feita expandindo-se o resíduo numa série de Taylor ao redor de seu valor correspondente a algum conjunto de valores iniciais dos parâmetros. Consideremos os vetores β_j , δ_j e x_{tj} , todos de dimensão $p + q$, onde β_j contém as estimativas iniciais na j -ésima iteração, $\delta_j = \beta - \beta_j$ e x_{tj} contém as derivadas do resíduo em relação aos parâmetros:

$$x_{itj} = \left. \frac{\partial r_{nt}(\beta)}{\partial \beta_i} \right|_{\beta = \beta_j}, \quad (3.2)$$

para $(i = 1, \dots, p + q)$.

Expandindo o resíduo ao redor de $r_{nt}(\beta_j)$, temos que

$$r_{nt}(\beta) = r_{nt}(\beta_j) - x'_{tj} \delta_j. \quad (3.3)$$

Segundo Box e Jenkins(1976) o cálculo numérico das derivadas pode ser feito diretamente ou usando

$$x_{itj} = [r_{nt}(\beta_j) - r_{nt}(\beta_j + \alpha)] / \alpha_i, \quad (3.4)$$

para $(i = 1, \dots, p + q)$ e uma escolha apropriada de $\alpha = (0 \dots 0 \alpha_i 0 \dots 0)'$. Logo, δ_j pode ser estimado por regressão de $r_{nt}(\beta_j)$ sobre x_{tj} em (3.3).

Podemos, agora, aplicar alguns algoritmos para estimação L_1 não linear ao nosso caso.

3.2.1. Algoritmo de Osborne e Watson

O algoritmo proposto por Osborne e Watson(1971) é do tipo Gauss-Newton (Gonin e Money, 1987). Sua adaptação ao nosso problema nos dá na j -ésima iteração :

Passo 1 - Calcular $\delta = \delta_j$ que minimize

$$\sum_t |r_{nt}(\beta_j) + x'_{tj}\delta|; \quad (3.5)$$

Passo 2 - Calcular $\gamma = \gamma_j > 0$ que minimize

$$\sum_t |r_{nt}(\beta_j + \gamma\delta_j)|; \quad (3.6)$$

Passo 3 - Tomar

$$\beta_{j+1} = \beta_j + \gamma_j\delta_j \quad (3.7)$$

e retornar ao passo 1 para a iteração seguinte.

O problema L_1 linear no passo 1 pode ser resolvido, por exemplo, pelo algoritmo de Barrodale e Roberts(1973). No passo 2 calculamos a direção a seguir.

Este é um método descendente ou de gradiente (El-Attar et alii, 1979) e, portanto, sua convergência é lenta após umas poucas iterações (Marquardt, 1963). Se suas suposições não forem satisfeitas na prática, ele pode não convergir de modo algum (El-Attar et alii, 1979).

3.2.2. Algoritmo de Anderson e Osborne

O algoritmo proposto por Anderson e Osborne(1977a) é também do tipo Gauss-Newton, e um método descendente ou de gradiente. Em relação ao algoritmo de Osborne e Watson, ao invés de escolher γ no passo 2 que minimize (3.6), este determina um γ que assegure convergência, embora sem requerer a minimização unidimensional. Sua adaptação ao nosso problema nos dá o seguinte:

Passo 0 - Escolher $\alpha \in (0, 1)$ e $\lambda > 0$ independente de j (Gonin e Money, 1987, sugerem $\alpha = 0,1$ e $\lambda = 0,0001$);

Passo 1 - Na j -ésima iteração, calcular $\delta = \delta_j$ que minimize (3.5); seja \hat{S}_j esse mínimo;

Passo 2 - Escolher $\gamma = \gamma_j$ como o maior número em $\{1, \alpha, \alpha^2, \dots\}$ para o qual

$$\frac{\sum |r_{nt}(\beta_j)| - \sum |r_{nt}(\beta_j + \gamma\delta_j)|}{\gamma[\sum |r_{nt}(\beta_j)| - \hat{S}_j]} \geq \lambda; \quad (3.8)$$

Passo 3 - Tomar (3.7) e retornar ao passo 1 para a iteração seguinte.

Este algoritmo parece ser mais rápido que o de Osborne e Watson, isto é, como o quadrático oposto ao linear (Gonin e Money, 1987).

3.2.3. Algoritmo de El-Attar, Vidyasagar e Dutta

El-Attar et alii(1979) propuseram um algoritmo que converte o problema L_1 não linear numa seqüência de problemas de minimização sem restrições. Ao invés de minimizar uma função objetivo não diferenciável, ele minimiza funções de penalidade diferenciáveis. Mostra-se que a seqüência de soluções ótimas converge para uma solução do problema original. Sua adaptação ao nosso problema nos dá o seguinte:

Passo 0 - escolher números pequenos α e δ (dependendo da precisão desejada), um número pequeno $\varepsilon_1 > 1$ e um número $L > 1$. Os autores recomendam

$$\varepsilon_1 = \max_t |r_{nt}(\beta_1)|/10, \quad (3.9)$$

ou

$$\varepsilon_1 = (10n)^{-1} \sum_{t=1}^n |r_{nt}(\beta_1)|; \quad (3.10)$$

Passo 1 - Na j -ésima iteração, calcular $\mathbf{b} = \beta_j$ que minimize

$$P(\mathbf{b}, \varepsilon_j) = \sum_t \{|r_{nt}(\mathbf{b})|^2 + \varepsilon_j\}^{1/2}; \quad (3.11)$$

Passo 2 - Tomar $\varepsilon_{j+1} = \varepsilon_j/L$;

Passo 3 - se $\varepsilon_{j+1} \leq \delta$ e/ou se $\|\beta_j - \beta_{j-1}\|_1 \leq \alpha$, parar. Caso contrário, retornar ao passo 1 para a próxima iteração.

Gonin e Money(1987) sugerem o método quase-Newton de Fletcher(1970) para a minimização no passo 1.

À medida que ε_j aproxima-se de zero, a função $P(\mathbf{b}, \varepsilon)$ pode ficar não diferenciável na prática, e o algoritmo pode convergir lentamente ou não convergir. Para evitar este problema, os autores melhoraram o algoritmo com:

Passo 4 - Se $j = 1$, tomar $\hat{\beta}_2 = \beta_1$; se $j > 1$, achar a estimativa $\hat{\beta}_{j+1}$ pela técnica de extrapolação de Fiacco e McCormick(1966). Então, retornar ao passo 1 para a iteração seguinte.

Nessa técnica de extrapolação temos que

$$\hat{\beta}_{j+1} = \sum_{i=0}^{j-1} a_i (\epsilon_1 / L^j)^{i/2}, \quad (3.12)$$

onde os vetores \mathbf{b}_i são calculados de

$$\beta_k = \sum_{i=0}^{j-1} a_i \epsilon_j^{i/2}, \quad (k = 1, \dots, j). \quad (3.13)$$

3.2.4. Algoritmo de Tishler e Zang

Tishler e Zang(1982) propuseram um algoritmo que aproxima o problema L_1 não linear por uma seqüência de problemas continuamente diferenciáveis que podem ser resolvidos por métodos de gradiente. A função objetivo não diferenciável é substituída por uma função uma ou duas vezes continuamente diferenciável. Sua adaptação ao nosso problema nos dá o seguinte:

Passo 0 - Escolher números pequenos ϵ_1 e $\delta > 0$;

Passo 1 - Na j -ésima iteração, calcular $\mathbf{b} = \beta_j$ que minimize

$$\sum_t A[\epsilon_j, r_{nt}(\mathbf{b})]$$

onde A é a aproximação uma vez continuamente diferenciável

$$A[\epsilon_j, r_{nt}(\mathbf{b})] = \begin{cases} -r_{nt}(\mathbf{b}) & \text{se } r_{nt}(\mathbf{b}) \leq -\epsilon_j; \\ \frac{r_{nt}(\mathbf{b})^2 + \epsilon_j^2}{2\epsilon_j} & \text{se } -\epsilon_j \leq r_{nt}(\mathbf{b}) \leq \epsilon_j; \\ r_{nt}(\mathbf{b}) & \text{se } r_{nt}(\mathbf{b}) \geq \epsilon_j. \end{cases} \quad (3.14)$$

ou A é a aproximação duas vezes continuamente diferenciável

$$A[\epsilon_j, r_{nt}(\mathbf{b})] = \begin{cases} -r_{nt}(\mathbf{b}) & \text{se } r_{nt}(\mathbf{b}) \leq -\epsilon_j; \\ -r_{nt}(\mathbf{b})^4 / 8\epsilon_j^3 + 3r_{nt}(\mathbf{b})^2 / 4\epsilon_j + 3\epsilon_j / 8 & \text{se } -\epsilon_j \leq r_{nt}(\mathbf{b}) \leq \epsilon_j; \\ r_{nt}(\mathbf{b}) & \text{se } r_{nt}(\mathbf{b}) \geq \epsilon_j. \end{cases} \quad (3.15)$$

Passo 2 - Se $\|\beta_j - \beta_{j-1}\| < \delta$, parar. Caso contrário, escolher $\epsilon_{j+1} < \epsilon_j$ e retornar ao passo 1 para a próxima iteração.

Os autores sugerem $\varepsilon_1 = 0,1$. Gonin e Money(1987) sugerem o esquema de El-Attar et alii(1979) para o decréscimo de ε_j .

À medida que ε_j se aproxima de zero o mesmo problema do algoritmo de El-Attar et alii pode ocorrer (Gonin e Money, 1987).

3.2.5. Algoritmo de Schlossmacher

Schlossmacher(1973) propôs um algoritmo que aproxima o problema de estimação L_1 por uma seqüência de problemas de estimação L_2 ponderada. Sua adaptação ao nosso problema nos dá o seguinte:

Passo 0 - escolher um número pequeno $\delta > 0$;

Passo 1 - Calcular $\mathbf{b} = \beta_0$ que minimize $\sum_t [r_{nt}(\mathbf{b})]^2$;

Passo 2 - Na j -ésima iteração, calcular $\mathbf{b} = \beta_j$ que minimize

$$\sum_t w_{tj} [r_{nt}(\mathbf{b})]^2, \quad (3.16)$$

onde

$$w_{tj} = \begin{cases} 0 & \text{se } r_{nt}(\beta_{j-1}) \simeq 0; \\ 1/r_{nt}(\beta_{j-1}) & \text{caso contrário;} \end{cases} \quad (3.17)$$

Passo 3 - Se $|r_{nt}(\beta_j) - r_{nt}(\beta_{j-1})| \leq \delta$, parar. Caso contrário, retornar ao passo 2.

Este algoritmo pressupõe a existência de outro algoritmo para a estimação de mínimos quadrados ponderados em modelos ARMA no passo 2. Cogger(1979) indica este algoritmo como alternativa à programação linear quando o modelo contiver parâmetros de médias móveis. Por outro lado, para modelos auto-regressivos puros, os algoritmos de programação linear podem ser usados.

CAPÍTULO 4

ESTUDO DE MONTE CARLO

Neste capítulo apresentamos um estudo de Monte Carlo a fim de visualizar melhor o comportamento da estimativas L_1 . Os objetivos deste estudo são três. Primeiro, estudamos o comportamento da função objetivo dentro do espaço paramétrico, principalmente quanto ao aspecto da não convexidade. Segundo, procuramos verificar o resultado de consistência das estimativas obtido no teorema 2.11. Terceiro, procuramos lançar alguma luz sobre a questão da distribuição amostral das estimativas. Fizemos, também, uma comparação entre as estimativas L_1 e L_2 no caso de um modelo de médias móveis.

4.1. PLANEJAMENTO DOS EXPERIMENTOS

Estudamos três modelos de séries temporais: AR(2), ARMA(1,1) e MA(2). Os modelos AR(1) e MA(1) foram, eventualmente, tratados como casos especiais do AR(2) e do MA(2), onde $\phi_2 = 0$ e $\theta_2 = 0$, respectivamente. Para cada modelo, três distribuições do ruído foram consideradas: a distribuição normal padrão, a distribuição de Laplace com mediana zero e variância 2 e a distribuição de Cauchy com mediana zero e parâmetro de escala igual a um. Séries temporais de tamanho $N = 400$ observações foram geradas. As n primeiras observações de cada série gerada formaram uma sub-amostra de n observações, para $n = 25, 50, 100, 200$ e 400 .

4.1.1. Geração das Séries Temporais

Primeiramente, variáveis aleatórias com distribuição uniforme $(0,1)$ foram geradas. Os processos de ruído branco com distribuição normal padrão foram gerados usando o método polar de Marsaglia(1962). Os processos de ruído branco com distribuição de Laplace e de Cauchy foram gerados usando o método da transformação inversa (Abramowitz e Stegun, 1972, p.950). Dados um conjunto de parâmetros e um processo de ruído branco, a série temporal foi gerada diretamente usando (2.3). Uma amostra de $N + 100$ observações foi gerada, sendo as 100 primeiras descartadas para evitar o efeito de valores iniciais.

4.1.2. Primeiro Experimento

Antes mesmo de se obterem os resultados teóricos do capítulo 2, um primeiro experimento foi realizado para se saber quão irregular seria o comportamento da função objetivo dentro da região de estacionariedade e inversibilidade. Para cada um dos três modelos e cada uma das três distribuições de ruído tomou-se uma amostra sistemática de pares de pontos distribuídos dentro da respectiva região de estacionariedade e inversibilidade. Desta maneira, obtiveram-se pontos próximos à origem, próximos à fronteira e intermediários, num total de 1020 conjuntos de observações. Para cada conjunto de observações os resíduos foram calculados pelo procedimento de *backforecasting*, para valores distribuídos sistematicamente pela região de estacionariedade e inversibilidade, obtendo-se com eles os erros médios em L_1 e em L_2 e as correspondentes funções objetivo. Com tais dados construíram-se curvas de nível para os erros médios.

4.1.3. Segundo Experimento

Um segundo experimento foi realizado a fim de determinar a partir de que tamanho de amostra as estimativas poderiam ser consideradas normalmente distribuídas (Dielman e Pfaffenberger, 1988). Geraram-se 100 réplicas para cada modelo e distribuição, num total de 900 séries. Tomaram-se os modelos MA(2) com parâmetros $\theta_1 = 0,4$ e $\theta_2 = -0,5$, AR(2) com parâmetros $\phi_1 = 0,4$ e $\phi_2 = -0,5$ e ARMA(1,1) com parâmetros $\phi_1 = 0,4$ e $\theta_1 = -0,5$.

Os parâmetros foram estimados pelo algoritmo de Schlossmacher(1973). Em cada passo foi usado o algoritmo de Marquardt(1963), de modo semelhante ao descrito em Box e Jenkins(1976,p.504). Estes algoritmos foram escolhidos apenas por facilidade de uso, já que não era nosso objetivo comparar a eficácia dos diversos existentes nem elaborar um programa para ser comercializado. A cada iteração dos algoritmos os resíduos foram calculados pelo procedimento de *backforecasting*.

Calcularam-se algumas estatísticas descritivas sobre as estimativas obtidas: mínimo, média aritmética, máximo, mediana, desvio padrão, bem como erro médio L_1 e erro médio L_2 , dados por:

$$e_{pi} = \sum_{j=1}^{100} |\hat{\beta}_{ij}|^p, \quad (4.1)$$

para $(i = 1, 2)$ e $(p = 1, 2)$, respectivamente.

As distribuições amostrais construídas a partir das estimativas foram testadas para normalidade usando o teste de Lilliefors(1967), sendo a estatística D de Kolmogorov-Smirnov calculada pelo algoritmo de Gonzalez, Sahni e Franta(1977).

4.2. RESULTADOS E DISCUSSÃO

4.2.1. Curvas de Nível

Para um estudo gráfico do comportamento da função objetivo e do erro médio dentro da região de estacionariedade e inversibilidade, alguns casos do primeiro experimento foram selecionados (figuras 4.1 a 4.20). Nas demais séries estudadas, para as quais não apresentamos figuras, o comportamento é semelhante.

O exame do comportamento dos erros médios em L_1 e L_2 mostra que o problema da não convexidade aparentemente não é muito sério, no sentido de o número de mínimos locais não ser grande e de a função objetivo não ser representada por uma função muito complexa, existindo, mesmo, alguma regularidade nas curvas de nível, particularmente nas proximidades dos valores verdadeiros. Nos processos ARMA(1,1) estudados não se detectou a não convexidade (figuras 4.1 a 4.6, 4.19 e 4.20). Existe, entretanto, a possibilidade de que ela ocorra fora da região de estacionariedade e inversibilidade ou de que o intervalo de amostragem utilizado na construção das curvas de nível não tenha sido suficientemente

pequeno para detectá-la. Já nos processos MA(2) a não convexidade é nítida (figuras 4.7 a 4.12). Nos exemplos mostrados, com $\theta_1 = 0,9$ e $\theta_2 = -0,5$ aparecem mínimos locais próximos ao valor $\hat{\theta}_1 = -1,0$. Surpreendentemente, nos processos AR(2) a não convexidade também aparece, semelhantemente aos processos MA(2) analisados (figuras 4.13 a 4.18). O motivo está no uso do procedimento de *backforecasting* para obtenção dos resíduos amostrais. Como visto na proposição 2.5 isso não deve ocorrer se as duas primeiras observações forem reservadas para o cálculo inicial dos resíduos. Portanto, é na prática da estimação computacional que pode ocorrer a não convexidade em modelos auto-regressivos. Como é usual, mesmo na estimação L_2 , o uso de *backforecasting* na estimação final dos parâmetros (Box e Jenkins, 1976), resultados como os de An e Chen(1982), Gross e Steiger(1979), Hannan e Kanter(1977), Yohai e Maronna(1977), podem não se aplicar às estimativas obtidas em modelos auto-regressivos. Entretanto, os resultados mais gerais obtidos no capítulo 2 continuam sendo válidos.

Na maior parte dos casos apresentados nas figuras o comportamento das curvas de nível L_1 e L_2 é muito semelhante. Note-se, entretanto, que a estimação L_2 do processo ARMA(1,1) com distribuição de Cauchy é muito inferior à estimação L_1 (figuras 4.5 e 4.6), no sentido de as curvas de menor nível incluírem o valor verdadeiro no caso L_1 , mas, não no caso L_2 .

É importante observar que os modelos ARMA(p,q) com $p = q$ e $\theta_i = \phi_i$, ($i = 1, \dots, p$) reduzem-se a $z_t = a_t$. Isso explica os valores constantes dos erros médios ao longo da diagonal $\phi_1 = \theta_1$ que aparecem nas figuras 4.1 a 4.6, 4.19 e 4.20. Se os valores verdadeiros dos parâmetros estiverem sobre essa diagonal, a estimativa poderá ser ruim, como no exemplo da figura 4.20.

4.2.2. Estimação

As estatísticas descritivas (tabelas 4.1 a 4.3) mostram que, em todos os casos, à medida que o tamanho da amostra aumenta, as estimativas aproximam-se do valor verdadeiro, pois, a média e a mediana das estimativas aproximam-se do valor verdadeiro, enquanto diminuem as medidas de dispersão, como o desvio padrão, os erros L_1 e L_2 e o intervalo entre o mínimo e o máximo. Este resultado reforça o resultado teórico de consistência, mesmo no caso de distribuição de Cauchy, em que o processo não é L_1 e, portanto, não se aplica o teorema 2.11. Convém lembrar que, em situações reais onde se dispõe apenas de uma série, a variância das estimativas poderá, eventualmente, ser obtida pela técnica de *bootstrap*.

Já o teste de Lilliefors (tabelas 4.4 a 4.6) mostra que nenhum dos casos apresentados evidenciou normalidade das estimativas mesmo para 400 observações. A pior situação foi com a distribuição de Cauchy, como esperado. Uma explicação parcial para esse resultado pobre pode estar no pequeno número de réplicas utilizado, o que se deveu ao alto custo da simulação, já que o presente estudo de Monte Carlo é antes ilustrativo que objetivo final do trabalho.

O exame dos gráficos (figuras 4.21 a 4.38) mostra que:

a) a distribuição das estimativas tende a ser simétrica com pico próximo à mediana ou à média;

b) com poucas exceções, o histograma é mais pontiagudo que a normal correspondente, principalmente quando a distribuição é de Cauchy (e dentro deste caso, principalmente no modelo auto-regressivo);

c) na maioria dos casos aparecem valores estranhos (*outliers*), principalmente quando a distribuição é de Cauchy.

Assim, boa parte dos gráficos assemelha-se ao da figura 1.1 e aos de Mandelbrot (1963). Ainda que para conclusões definitivas sejam necessárias simulações de maior fôlego, o presente estudo sugere que inicialmente uma distribuição estável de índice $\alpha < 2$ poderia

ser considerada para as estimativas. O exame dos gráficos também reforça o resultado de consistência obtido no teorema 2.11.

4.2.3. Eficiência Relativa

A eficiência relativa da estimação L_1 em relação à L_2 foi estudada no segundo experimento somente para o modelo MA(2) com 400 observações. Tomaram-se como estimativas L_2 as obtidas na primeira iteração do algoritmo de Schlossmacher.. Comparando-se os resultados das tabelas 4.1 e 4.7 verificamos que em todos os casos a média das estimativas L_2 esteve mais próxima do valor verdadeiro que a média das estimativas L_1 . Entretanto, a variância das estimativas L_1 foi menor do que a das estimativas L_2 no caso do parâmetro θ_2 , distribuição de Laplace, e do parâmetro θ_1 , distribuição normal e bem menor no caso da distribuição de Cauchy, o que parece indicar que o estimador L_1 é melhor exatamente nos casos mais extremos.

TABELA 4.1 – Modelo MA(2), com parâmetros $\theta_1 = 0,4$ e $\theta_2 = -0,5$, resultados de estimação L_1 sobre 100 séries geradas.

Distribuição do ruído	Parâmetro	Tamanho da amostra	Mínimo	Média	Máximo	Mediana	Desvio padrão	Erro L_1	Erro L_2
Laplace	θ_1	25	-0,1388	0,3404	0,7631	0,3653	0,1564	0,1249	0,0278
		50	0,0276	0,3461	0,5061	0,3826	0,1029	0,0845	0,0134
		100	0,1552	0,3537	0,4966	0,3731	0,0755	0,0650	0,0078
		200	0,2396	0,3713	0,4860	0,3826	0,0578	0,0487	0,0041
		400	0,2669	0,3786	0,4852	0,3863	0,0449	0,0377	0,0025
	θ_2	25	-0,6512	-0,3664	-0,0350	-0,3828	0,1395	0,1544	0,0371
		50	-0,6216	-0,4131	-0,0040	-0,4241	0,1112	0,1059	0,0198
		100	-0,6388	-0,4388	-0,2786	-0,4457	0,0726	0,0733	0,0090
		200	-0,5647	-0,4598	-0,3437	-0,4594	0,0552	0,0538	0,0046
		400	-0,5690	-0,4703	-0,3612	-0,4757	0,0433	0,0401	0,0027
Normal	θ_1	25	0,0704	0,3478	0,8560	0,3556	0,1278	0,1071	0,0189
		50	0,1697	0,3525	0,5531	0,3634	0,0969	0,0885	0,0116
		100	0,2133	0,3565	0,5159	0,3565	0,0670	0,0658	0,0063
		200	0,2512	0,3732	0,5034	0,3698	0,0563	0,0509	0,0039
		400	0,2746	0,3757	0,4603	0,3813	0,0371	0,0348	0,0020
	θ_2	25	-0,6544	-0,3685	0,1961	-0,4052	0,1693	0,1554	0,0457
		50	-0,7252	-0,4301	-0,0038	-0,4498	0,1123	0,0991	0,0174
		100	-0,6652	-0,4549	-0,1776	-0,4760	0,0952	0,0778	0,0110
		200	-0,5994	-0,4589	-0,1820	-0,4746	0,0686	0,0582	0,0063
		400	-0,5815	-0,4729	-0,3208	-0,4787	0,0549	0,0463	0,0037
Cauchy	θ_1	25	-0,8472	0,3094	1,0286	0,3170	0,2179	0,1625	0,0552
		50	0,0167	0,3370	0,8476	0,3401	0,1164	0,0995	0,0174
		100	0,1756	0,3644	0,9661	0,3595	0,0978	0,0724	0,0107
		200	-0,0363	0,3759	0,5175	0,3864	0,0615	0,0392	0,0043
		400	0,2623	0,3817	0,4543	0,3908	0,0354	0,0265	0,0016
	θ_2	25	-1,3224	-0,3806	0,0063	-0,3678	0,1626	0,1577	0,0404
		50	-0,5661	-0,3979	-0,1055	-0,4074	0,0896	0,1087	0,0184
		100	-0,5871	-0,4265	0,0291	-0,4479	0,0867	0,0796	0,0128
		200	-0,5781	-0,4650	-0,1859	-0,4791	0,0592	0,0447	0,0047
		400	-0,5753	-0,4826	-0,3152	-0,4940	0,0391	0,0257	0,0018

TABELA 4.2 – Modelo AR(2), com parâmetros $\phi_1 = 0,4$ e $\phi_2 = -0,5$, resultados de estimação L_1 sobre 100 séries geradas.

Distribuição do ruído	Parâmetro	Tamanho da amostra	Mínimo	Média	Máximo	Mediana	Desvio padrão	Erro L_1	Erro L_2
Laplace	ϕ_1	25	0,0046	0,4005	0,7779	0,4076	0,1723	0,1380	0,0294
		50	0,1195	0,4128	0,6900	0,4229	0,1182	0,0948	0,0140
		100	0,1942	0,4103	0,5582	0,4244	0,0812	0,0672	0,0066
		200	0,2169	0,3967	0,5337	0,3950	0,0588	0,0460	0,0034
		400	0,2802	0,3884	0,5130	0,3933	0,0461	0,0371	0,0022
	ϕ_2	25	-0,7967	-0,5481	-0,0857	-0,5812	0,1423	0,1237	0,0224
		50	-0,7399	-0,5189	-0,2534	-0,5225	0,1128	0,0923	0,0129
		100	-0,7132	-0,5093	-0,2937	-0,5010	0,0865	0,0694	0,0075
		200	-0,6210	-0,5090	-0,3954	-0,5087	0,0531	0,0423	0,0029
		400	-0,5969	-0,4983	-0,3954	-0,4977	0,0418	0,0332	0,0017
Normal	ϕ_1	25	0,1590	0,4178	0,6440	0,4191	0,0770	0,0598	0,0062
		50	0,2670	0,3972	0,5186	0,3978	0,0540	0,0445	0,0029
		100	0,3240	0,4024	0,4872	0,4036	0,0368	0,0288	0,0013
		200	0,3490	0,4029	0,4551	0,4059	0,0238	0,0191	0,0006
		400	0,3622	0,3985	0,4374	0,3949	0,0166	0,0138	0,0003
	ϕ_2	25	-0,8565	-0,4990	0,0729	-0,5322	0,2011	0,1623	0,0400
		50	-0,8331	-0,5053	0,0476	-0,5213	0,1539	0,1184	0,0235
		100	-0,7350	-0,5116	-0,2360	-0,5007	0,0944	0,0728	0,0090
		200	-0,6298	-0,4957	-0,3141	-0,4993	0,0679	0,0554	0,0046
		400	-0,6068	-0,4959	-0,3658	-0,5029	0,0471	0,0365	0,0022
Cauchy	ϕ_1	25	0,0674	0,4208	1,9503	0,4006	0,2024	0,1076	0,0410
		50	0,1994	0,4066	0,8238	0,3988	0,0800	0,0474	0,0064
		100	-0,5254	0,3969	0,5595	0,4002	0,1033	0,0372	0,0106
		200	0,2879	0,4010	0,6320	0,3997	0,0370	0,0173	0,0014
		400	0,3143	0,4017	0,5191	0,4002	0,0211	0,0096	0,0004
	ϕ_2	25	-1,2940	-0,5050	0,7792	-0,5112	0,1845	0,0914	0,0337
		50	-0,7066	-0,5052	-0,2169	-0,5050	0,0770	0,0496	0,0059
		100	-0,6710	-0,5067	-0,0148	-0,5033	0,0712	0,0332	0,0051
		200	-0,6493	-0,5040	-0,3844	-0,5010	0,0335	0,0174	0,0011
		400	-0,6256	-0,5024	-0,3992	-0,5007	0,0235	0,0101	0,0006

TABELA 4.3 – Modelo ARMA(1,1), com parâmetros $\phi_1 = 0,4$ e $\theta_1 = -0,5$, resultados de estimação L_1 sobre 100 séries geradas.

Distribuição do ruído	Parâmetro	Tamanho da amostra	Mínimo	Média	Máximo	Mediana	Desvio padrão	Erro L_1	Erro L_2
Laplace	ϕ_1	25	-0,0398	0,4646	0,9185	0,4503	0,2018	0,1652	0,0445
		50	0,0158	0,4558	0,8529	0,4366	0,1563	0,1287	0,0273
		100	0,1888	0,4352	0,6635	0,4262	0,1086	0,0922	0,0129
		200	0,2407	0,4234	0,5853	0,4303	0,0715	0,0617	0,0056
		400	0,2885	0,4134	0,5638	0,4072	0,0489	0,0377	0,0025
	θ_1	25	-0,8181	-0,3648	0,2476	-0,3563	0,2224	0,2117	0,0672
		50	-0,7092	-0,3967	-0,0509	-0,3988	0,1688	0,1619	0,0389
		100	-0,7026	-0,4262	-0,1360	-0,4239	0,1169	0,1116	0,0190
		200	-0,7226	-0,4519	-0,2177	-0,4583	0,0827	0,0740	0,0091
		400	-0,5859	-0,4656	-0,3048	-0,4748	0,0551	0,0482	0,0042
Normal	ϕ_1	25	0,0700	0,4634	0,9276	0,4260	0,2075	0,1715	0,0466
		50	0,0641	0,4621	0,7746	0,4331	0,1626	0,1347	0,0300
		100	0,1463	0,4377	0,7046	0,4040	0,1133	0,0841	0,0141
		200	0,2388	0,4159	0,5828	0,4056	0,0721	0,0538	0,0054
		400	0,2582	0,4167	0,6127	0,4067	0,0566	0,0408	0,0035
	θ_1	25	-0,8338	-0,3883	0,0757	-0,3852	0,2141	0,1942	0,0578
		50	-0,8786	-0,4091	-0,0739	-0,4377	0,1723	0,1524	0,0377
		100	-0,7597	-0,4418	-0,1344	-0,4680	0,1174	0,0940	0,0170
		200	-0,6604	-0,4706	-0,2921	-0,4849	0,0731	0,0579	0,0062
		400	-0,6435	-0,4748	-0,2715	-0,4871	0,0578	0,0444	0,0039
Cauchy	ϕ_1	25	0,0408	0,4629	1,9547	0,4354	0,1747	0,1287	0,0342
		50	0,1223	0,4384	0,9371	0,4157	0,1074	0,0709	0,0129
		100	0,2365	0,4203	0,6226	0,4067	0,0592	0,0401	0,0039
		200	0,2430	0,4113	0,5864	0,4017	0,0469	0,0260	0,0023
		400	0,3256	0,4139	0,5710	0,4017	0,0365	0,0185	0,0015
	θ_1	25	-0,7078	-0,3987	0,0226	-0,4495	0,1611	0,1349	0,0360
		50	-0,7084	-0,4328	-0,1061	-0,4803	0,1184	0,0882	0,0184
		100	-0,6948	-0,4594	-0,1237	-0,4895	0,0838	0,0555	0,0086
		200	-0,6244	-0,4759	-0,1252	-0,4949	0,0693	0,0338	0,0053
		400	-0,6700	-0,4831	-0,1294	-0,4959	0,0580	0,0257	0,0036

TABELA 4.4 – Modelo MA(2), estatística D de Kolmogorov-Smirnov, resultados de estimação L_1 sobre 100 séries geradas.

Distribuição do ruído	Parâmetro	Tamanho da Amostra				
		25	50	100	200	400
Normal	θ_1	0,770	0,733	0,852	0,716	0,823
	θ_2	1,170	0,863	1,038	0,998	0,795
Laplace	θ_1	0,692	1,471	1,180	1,121	0,821
	θ_2	0,740	0,870	0,926	0,721	0,860
Cauchy	θ_1	1,161	1,372	1,430	1,506	1,451
	θ_2	1,070	0,826	1,316	1,522	1,435

Valores críticos: 0,103(1%); 0,089(5%); 0,081(10%).

TABELA 4.5 – Modelo AR(2), estatística D de Kolmogorov-Smirnov, resultados de estimação L_1 sobre 100 séries geradas.

Distribuição do ruído	Parâmetro	Tamanho da Amostra				
		25	50	100	200	400
Normal	ϕ_1	0,579	0,552	0,852	0,846	1,048
	ϕ_2	1,214	0,759	0,838	0,593	0,693
Laplace	ϕ_1	0,647	0,567	0,873	0,695	0,569
	ϕ_2	1,082	0,488	0,519	0,426	0,375
Cauchy	ϕ_1	1,802	1,908	2,694	2,769	2,702
	ϕ_2	2,170	1,964	2,852	2,497	3,324

Valores críticos: 0,103(1%); 0,089(5%); 0,081(10%).

TABELA 4.6 – Modelo ARMA(1,1), estatística D de Kolmogorov-Smirnov, resultados de estimação L_1 sobre 100 séries geradas.

Distribuição do ruído	Parâmetro	Tamanho da Amostra				
		25	50	100	200	400
Normal	ϕ_1	0,815	0,804	1,575	1,305	1,321
	θ_1	0,587	0,759	1,366	1,254	1,275
Laplace	ϕ_1	0,552	0,658	0,661	0,625	0,671
	θ_1	0,688	0,683	0,367	0,592	0,834
Cauchy	ϕ_1	1,455	1,675	1,392	2,185	3,186
	θ_1	1,495	1,771	1,920	3,093	3,186

Valores críticos: 0,103(1%); 0,089(5%); 0,081(10%).

TABELA 4.7 – Modelo MA(2), com parâmetros $\theta_1 = 0,4$ e $\theta_2 = -0,5$, resultados de estimação L_2 sobre 100 séries geradas, com 400 observações cada uma.

Distribuição do ruído	Parâmetro	Média	Desvio padrão
Laplace	θ_1	0,4032	0,0424
	θ_2	-0,5004	0,0446
Normal	θ_1	0,3966	0,0380
	θ_2	-0,4969	0,0479
Cauchy	θ_1	0,3990	0,0493
	θ_2	-0,4912	0,0820

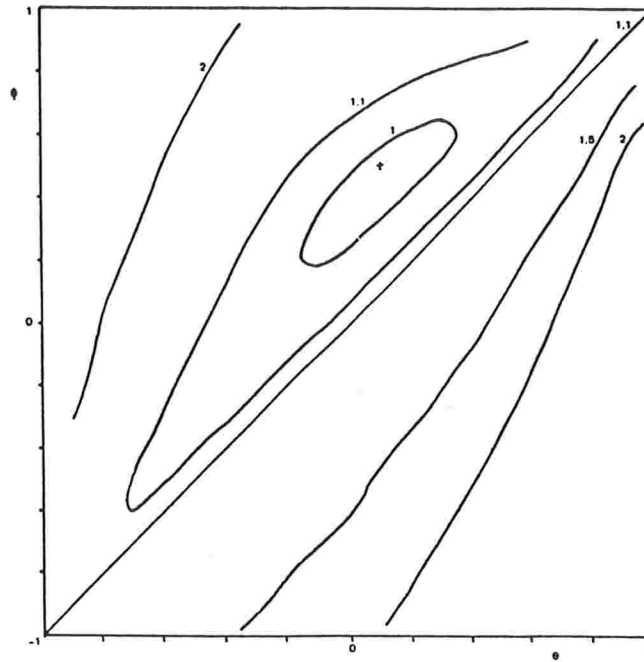


FIGURA 4.1 — Curvas de nível para o erro médio L_1 , modelo ARMA(1,1) com $\phi_1 = 0,5$ e $\theta_1 = 0,1$, distribuição de Laplace, 400 observações.

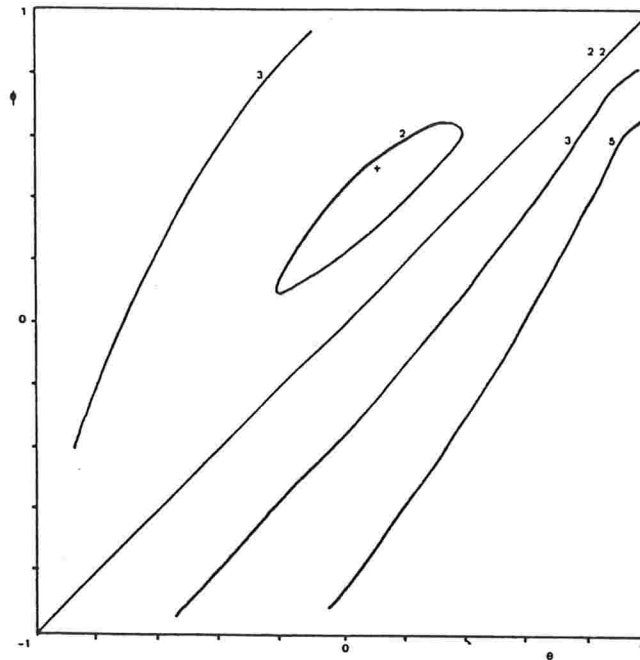


FIGURA 4.2 — Curvas de nível para o erro médio L_2 , modelo ARMA(1,1) com $\phi_1 = 0,5$ e $\theta_1 = 0,1$, distribuição de Laplace, 400 observações.

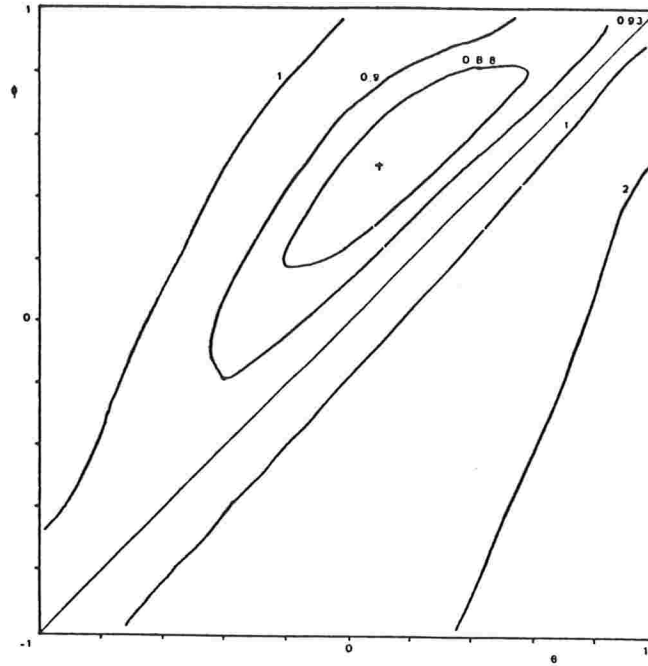


FIGURA 4.3 — Curvas de nível para o erro médio L_1 , modelo ARMA(1,1) com $\phi_1 = 0,5$ e $\theta_1 = 0,1$, distribuição normal, 400 observações.

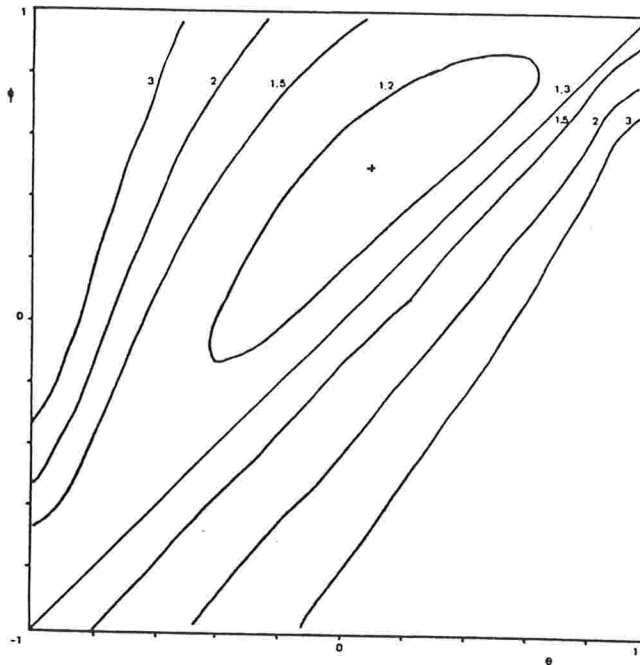


FIGURA 4.4 — Curvas de nível para o erro médio L_2 , modelo ARMA(1,1) com $\phi_1 = 0,5$ e $\theta_1 = 0,1$, distribuição normal, 400 observações.

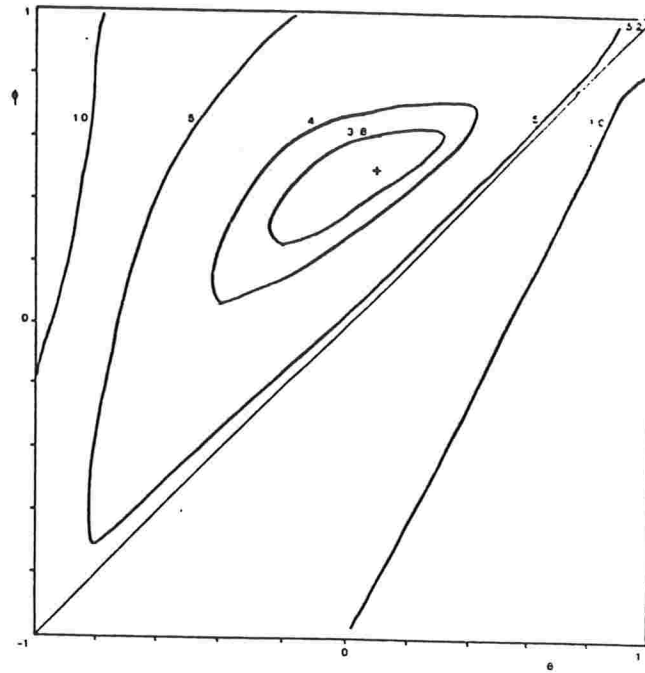


FIGURA 4.5 — Curvas de nível para o erro médio L_1 , modelo ARMA(1,1) com $\phi_1 = 0,5$ e $\theta_1 = 0,1$, distribuição de Cauchy, 400 observações.

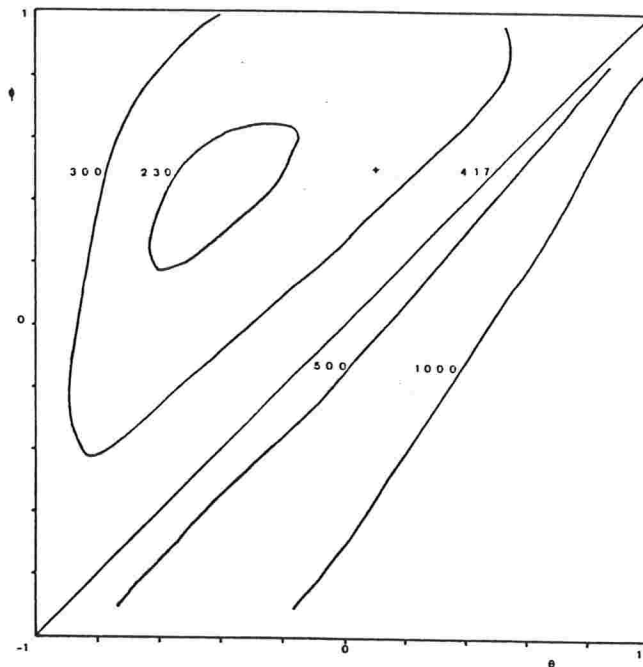


FIGURA 4.6 — Curvas de nível para o erro médio L_2 , modelo ARMA(1,1) com $\phi_1 = 0,5$ e $\theta_1 = 0,1$, distribuição de Cauchy, 400 observações.

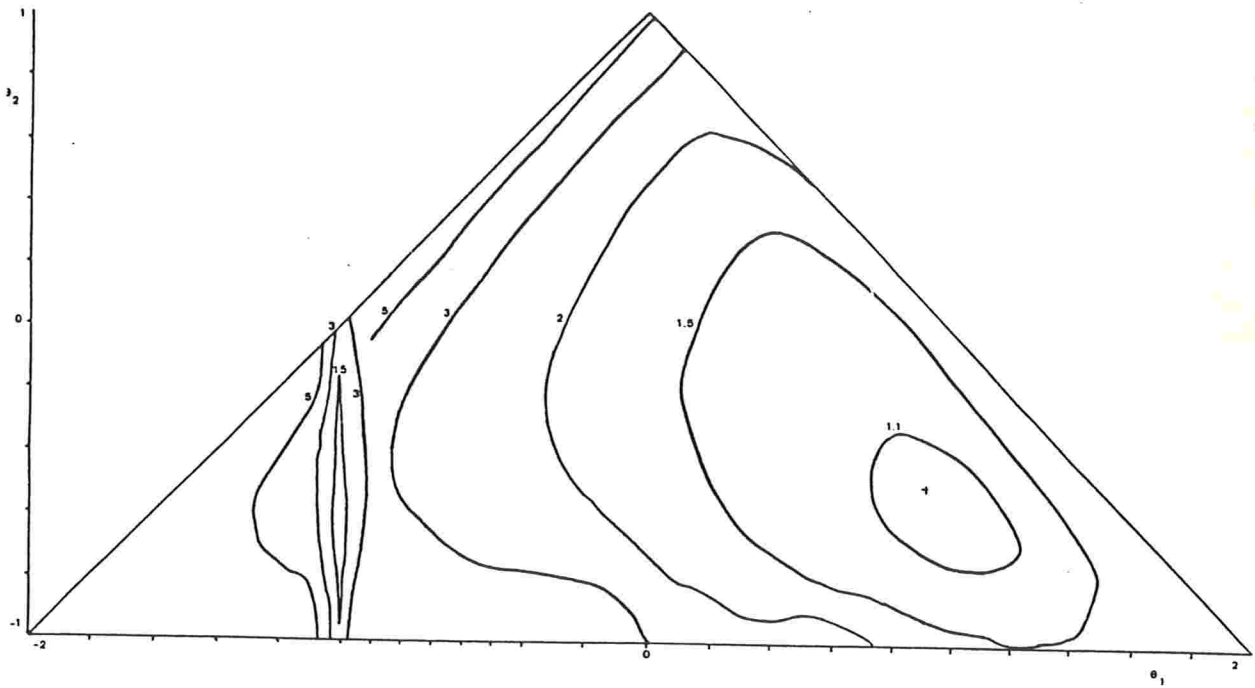


FIGURA 4.7 — Curvas de nível para o erro médio L_1 , modelo MA(2) com $\theta_1 = 0,9$ e $\theta_2 = -0,5$, distribuição de Laplace, 400 observações.

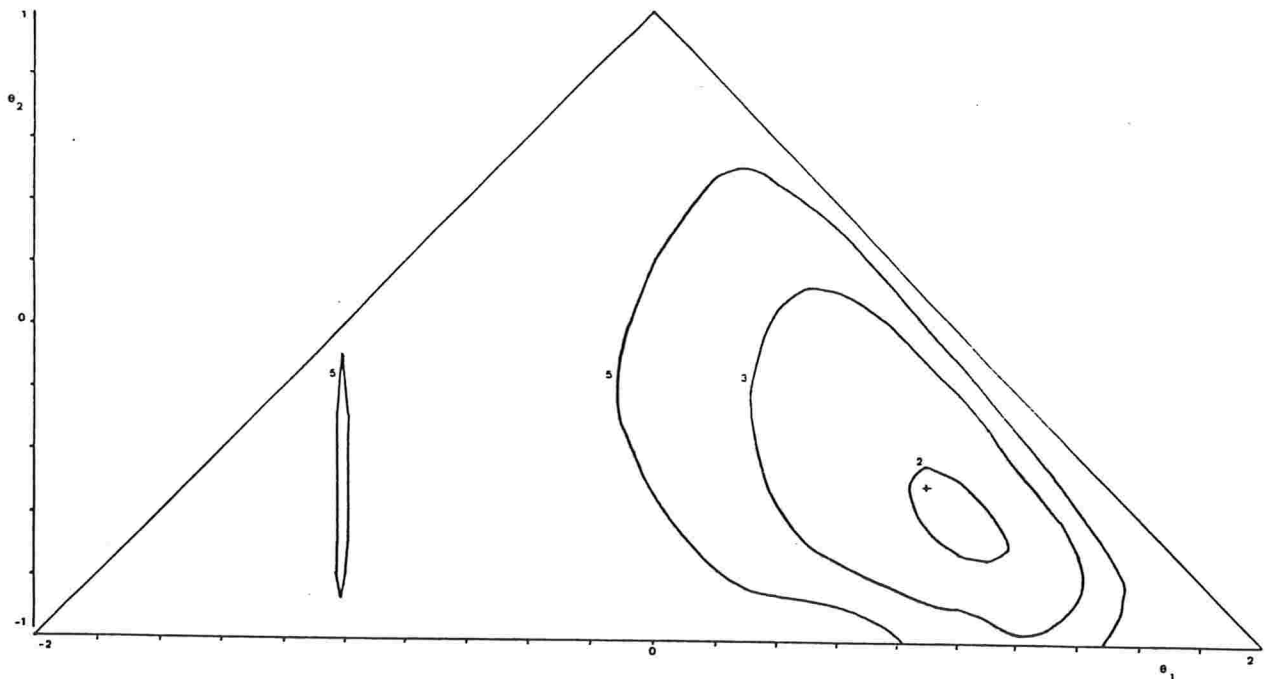


FIGURA 4.8 — Curvas de nível para o erro médio L_2 , modelo MA(2) com $\theta_1 = 0,9$ e $\theta_2 = -0,5$, distribuição de Laplace, 400 observações.

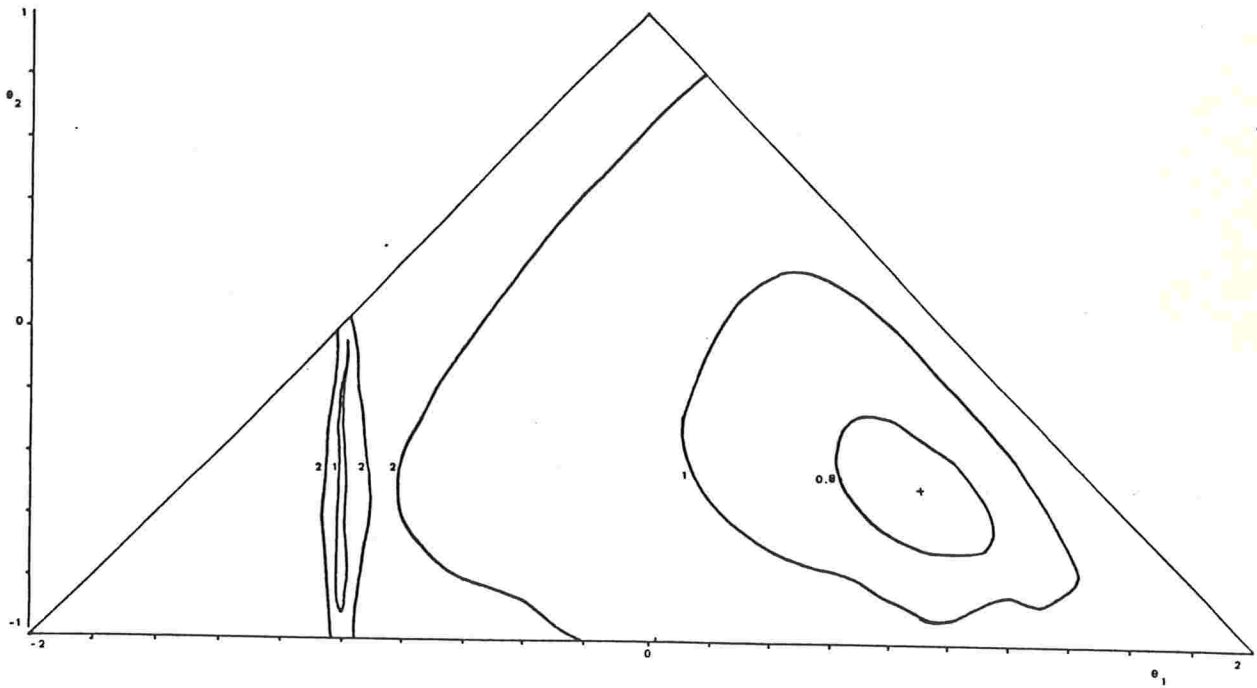


FIGURA 4.9 — Curvas de nível para o erro médio L_1 , modelo MA(2) com $\theta_1 = 0,9$ e $\theta_2 = -0,5$, distribuição normal, 400 observações.

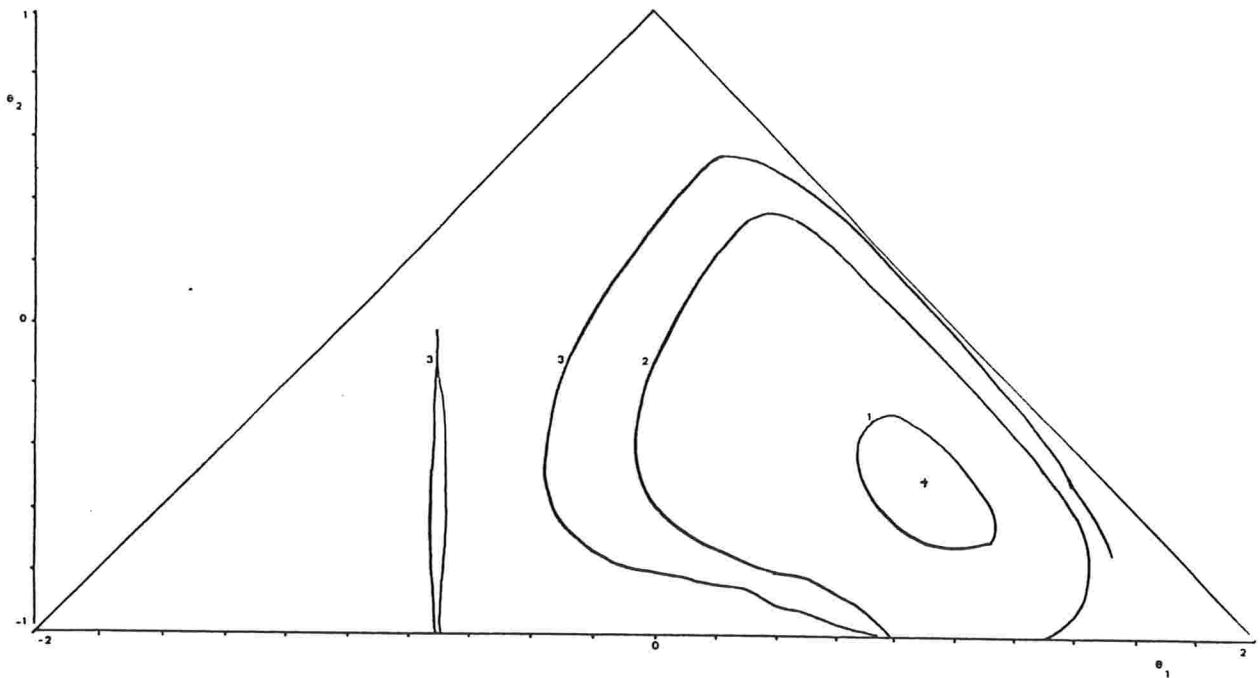


FIGURA 4.10 — Curvas de nível para o erro médio L_2 , modelo MA(2) com $\theta_1 = 0,9$ e $\theta_2 = -0,5$, distribuição normal, 400 observações.

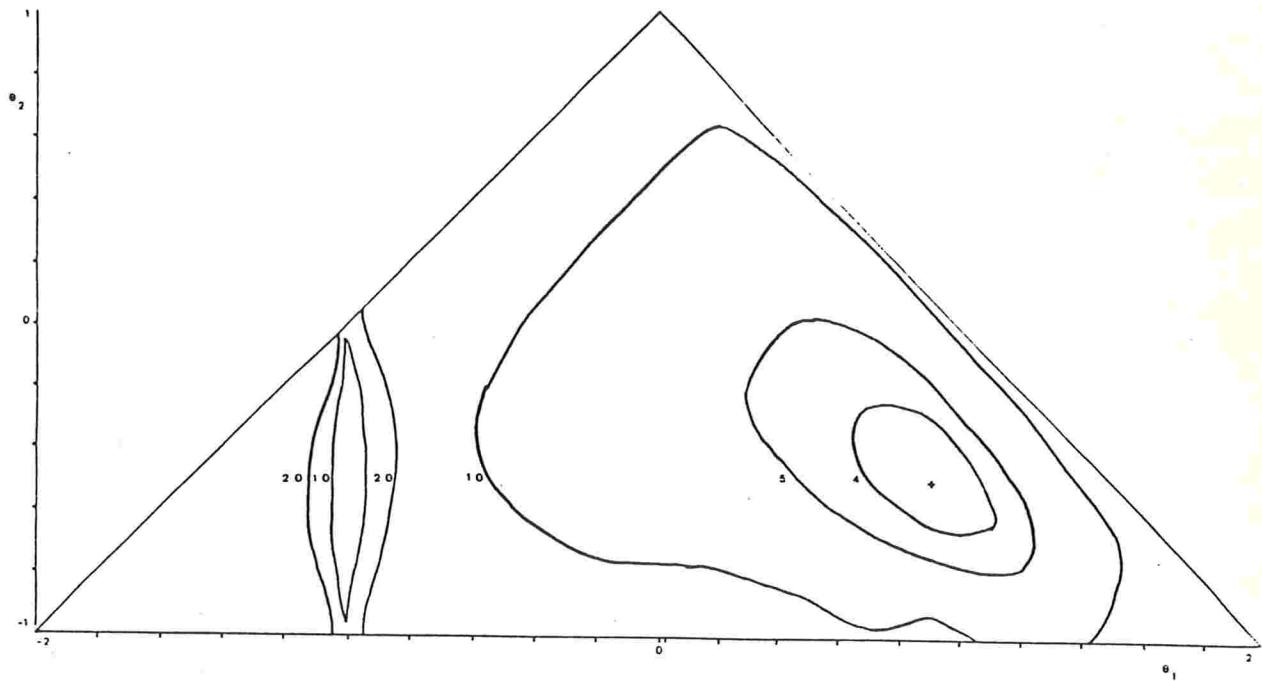


FIGURA 4.11 — Curvas de nível para o erro médio L_1 , modelo MA(2) com $\theta_1 = 0,9$ e $\theta_2 = -0,5$, distribuição de Cauchy, 400 observações.

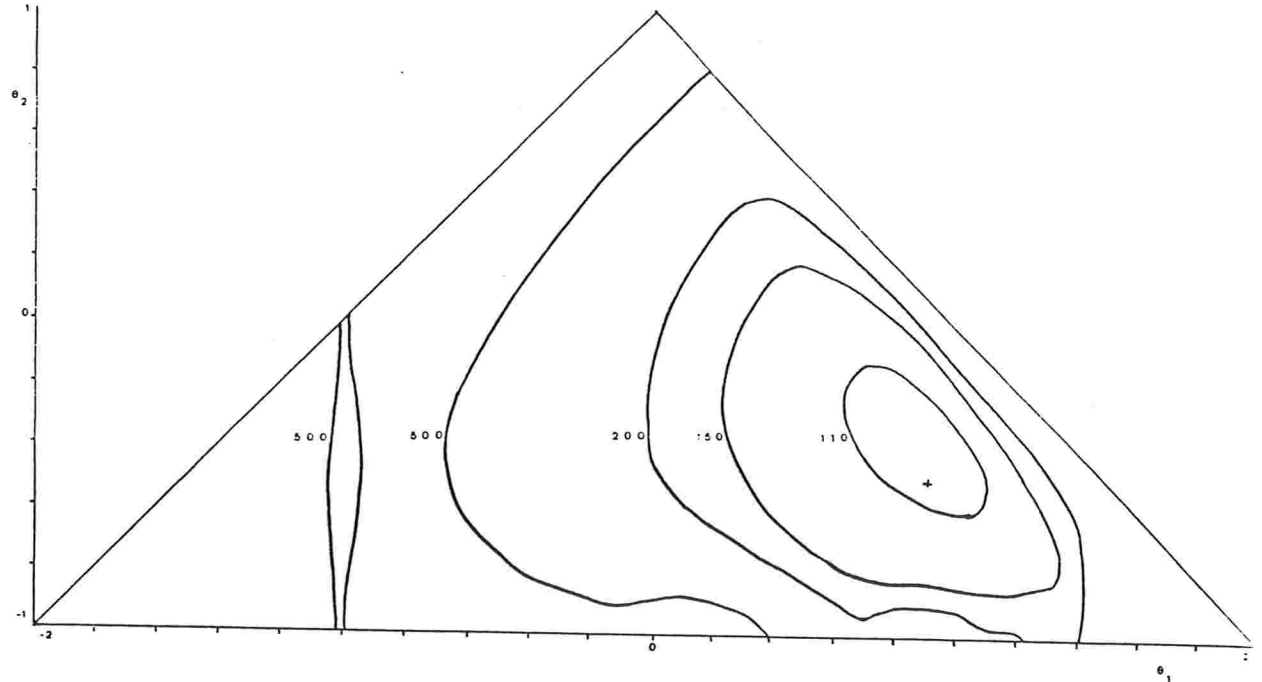


FIGURA 4.12 — Curvas de nível para o erro médio L_2 , modelo MA(2) com $\theta_1 = 0,9$ e $\theta_2 = -0,5$, distribuição de Cauchy, 400 observações.

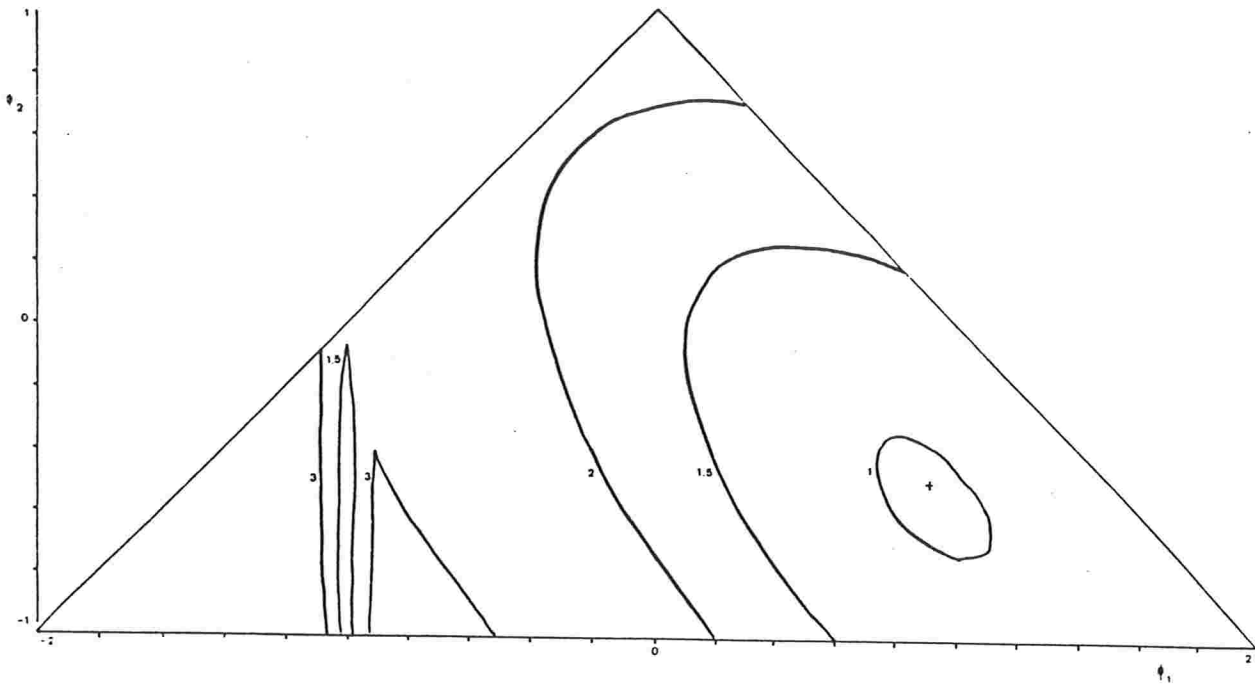


FIGURA 4.13 — Curvas de nível para o erro médio L_1 , modelo AR(2) com $\phi_1 = 0,9$ e $\phi_2 = -0,5$, distribuição de Laplace, 400 observações.

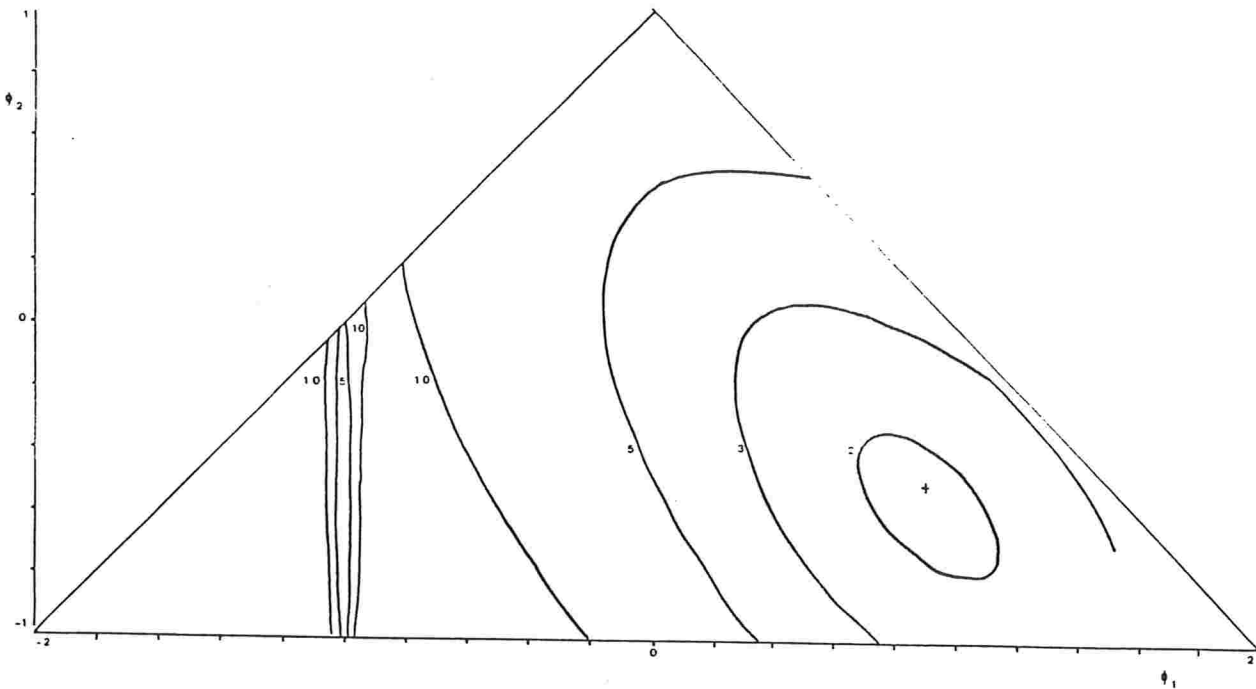


FIGURA 4.14 — Curvas de nível para o erro médio L_2 , modelo AR(2) com $\phi_1 = 0,9$ e $\phi_2 = -0,5$, distribuição de Laplace, 400 observações.

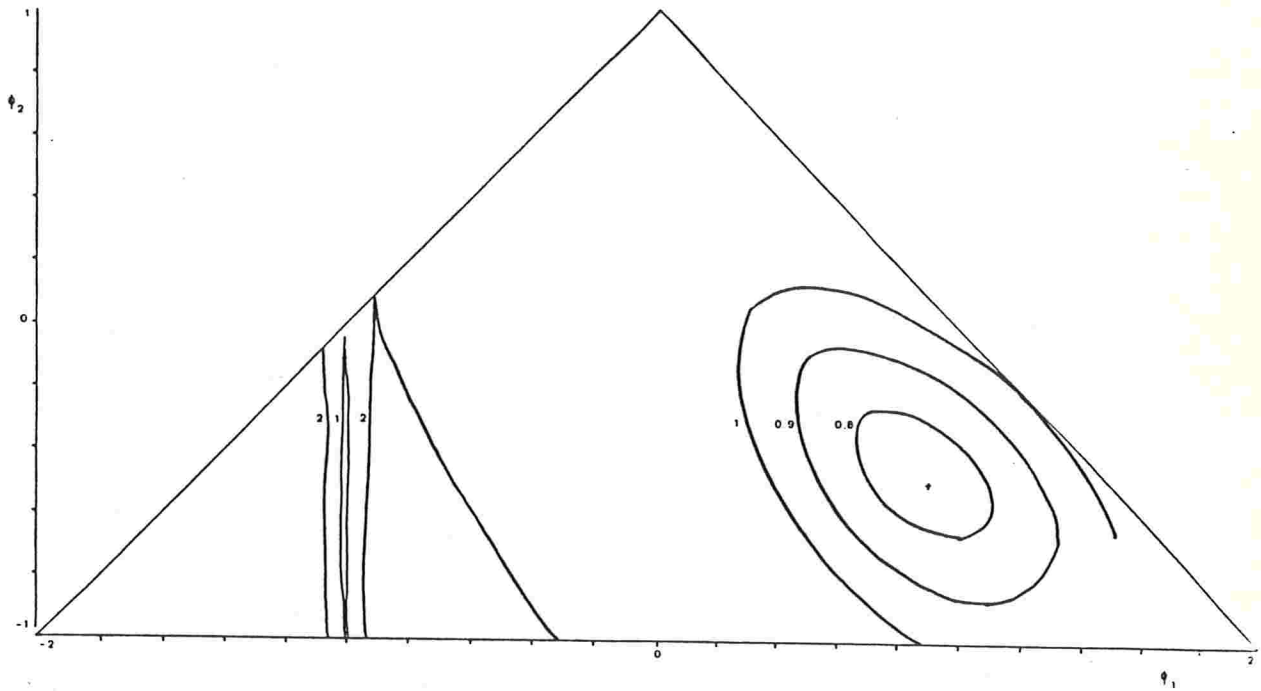


FIGURA 4.15 — Curvas de nível para o erro médio L_1 , modelo AR(2) com $\phi_1 = 0,9$ e $\phi_2 = -0,5$, distribuição normal, 400 observações.

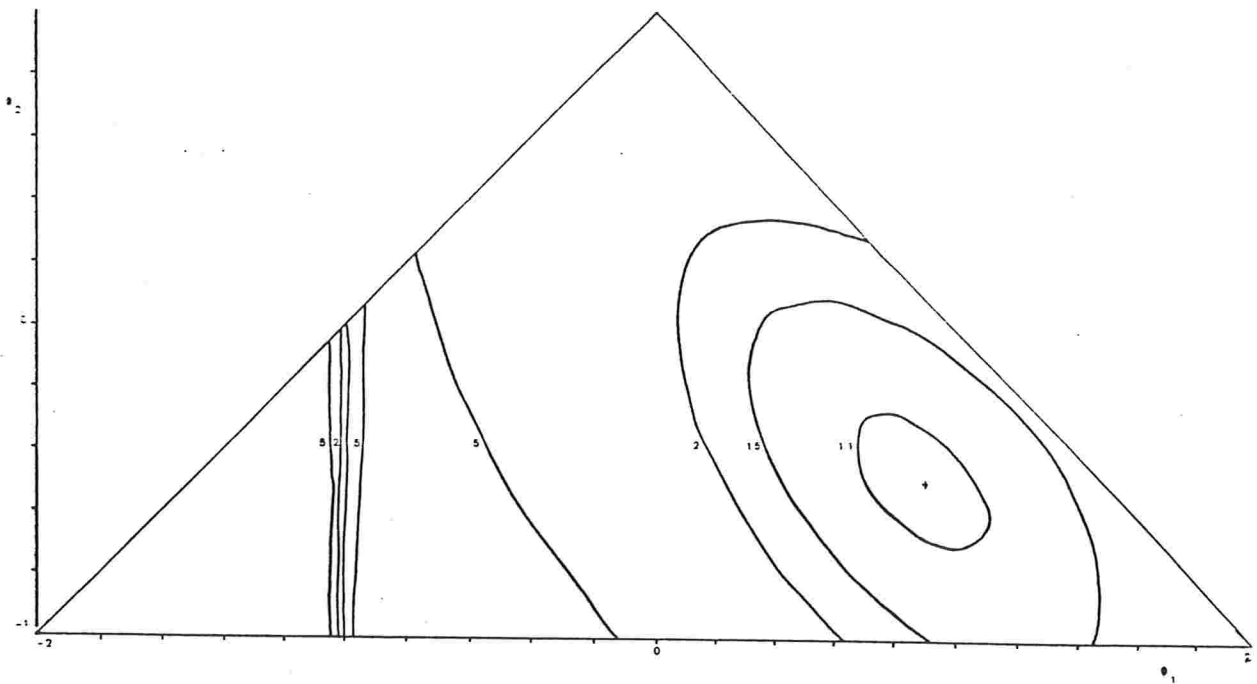


FIGURA 4.16 — Curvas de nível para o erro médio L_2 , modelo AR(2) com $\phi_1 = 0,9$ e $\phi_2 = -0,5$, distribuição normal, 400 observações.

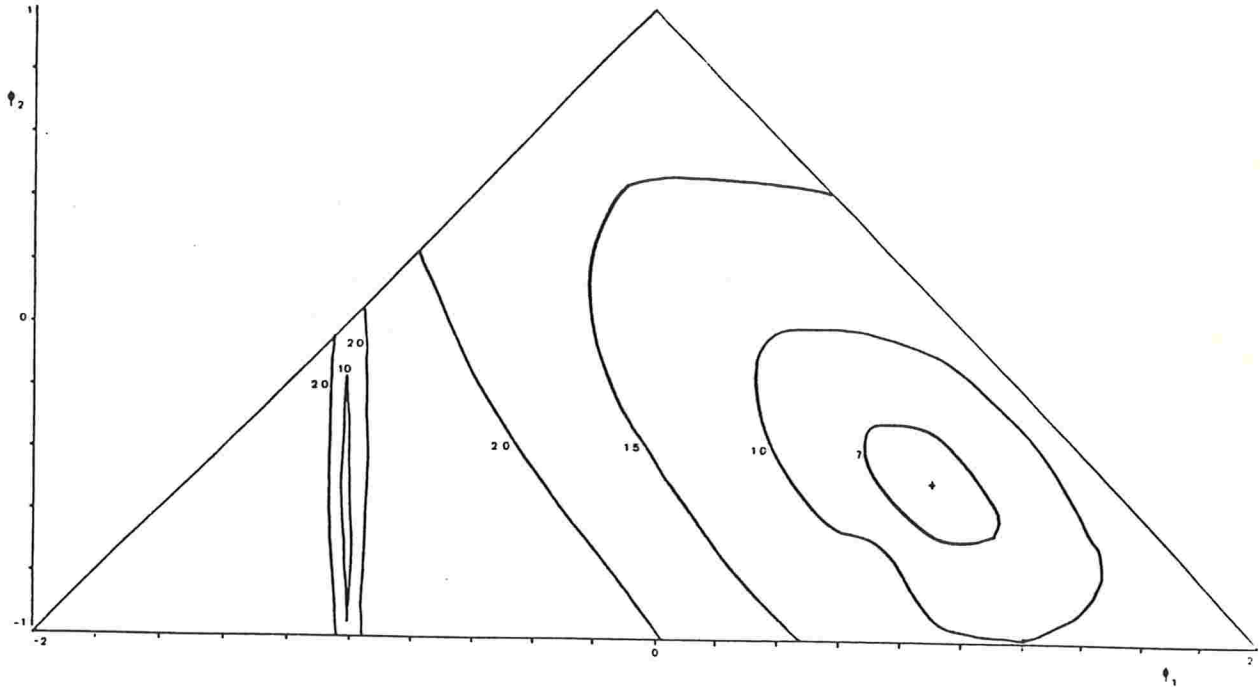


FIGURA 4.17 — Curvas de nível para o erro médio L_1 , modelo AR(2) com $\phi_1 = 0,9$ e $\phi_2 = -0,5$, distribuição de Cauchy, 400 observações.

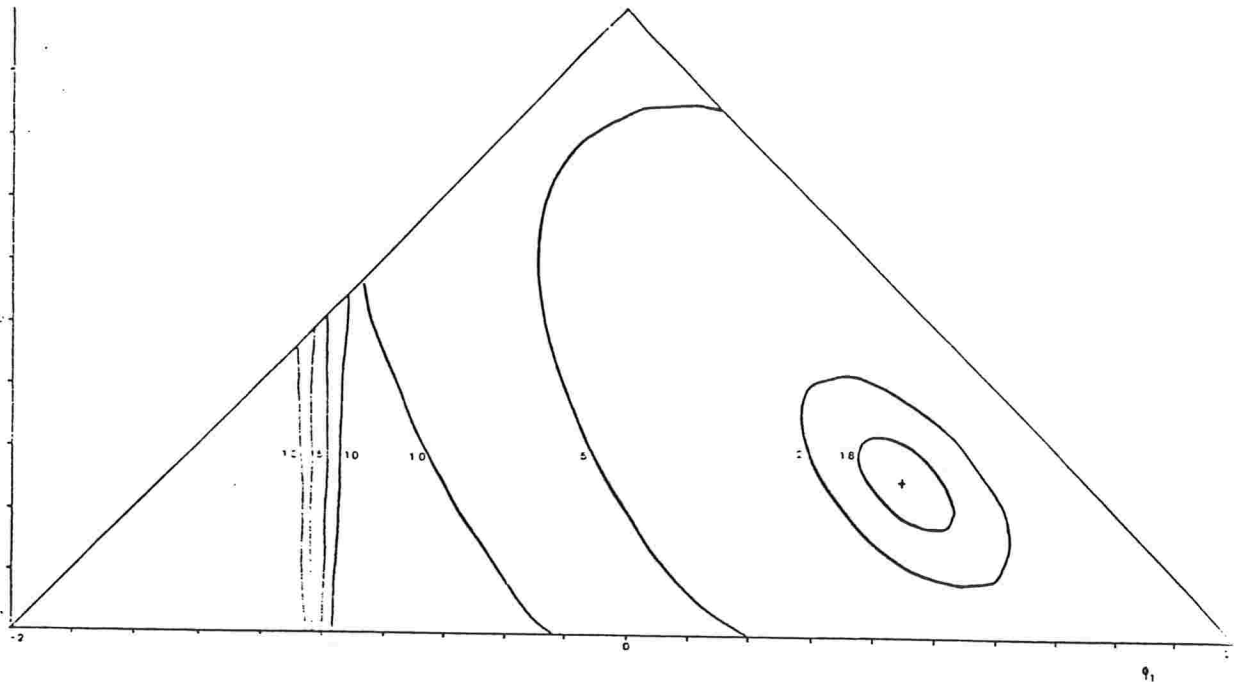


FIGURA 4.18 — Curvas de nível para o erro médio L_2 , modelo AR(2) com $\phi_1 = 0,9$ e $\phi_2 = -0,5$, distribuição de Cauchy, 400 observações.

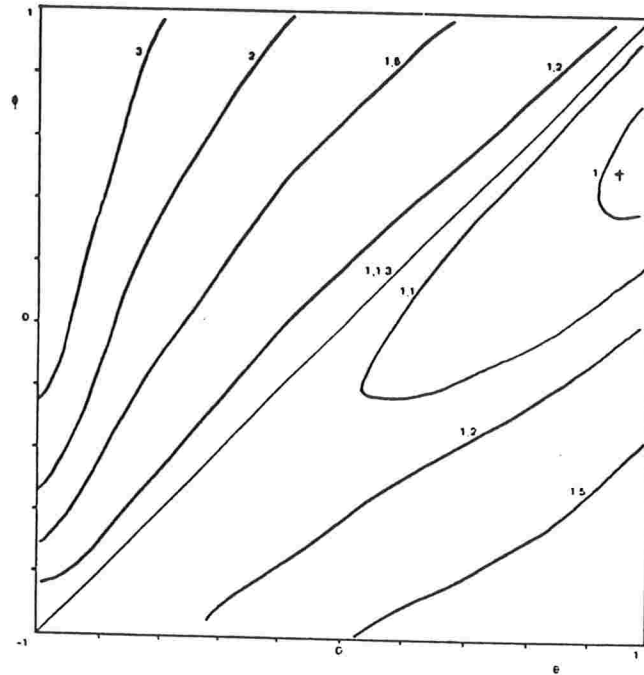


FIGURA 4.19 — Curvas de nível para o erro médio L_1 , modelo ARMA(1,1) com $\phi_1 = 0,5$ e $\theta_1 = 0,9$, distribuição de Laplace, 400 observações.

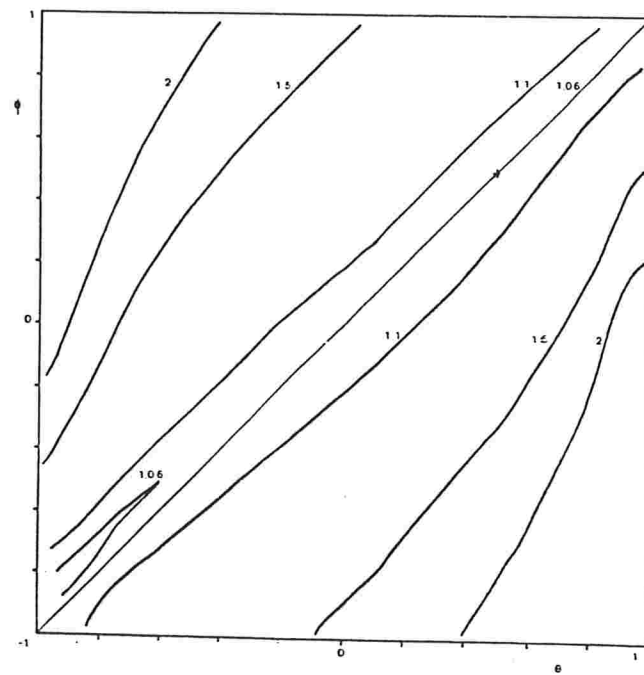
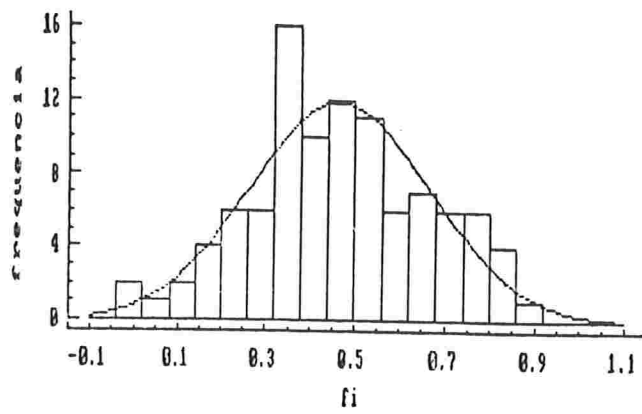
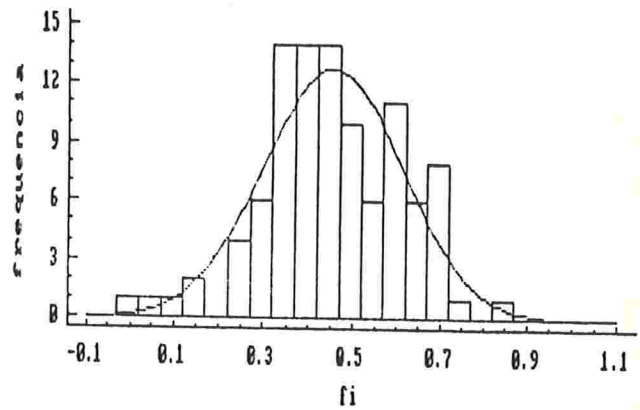


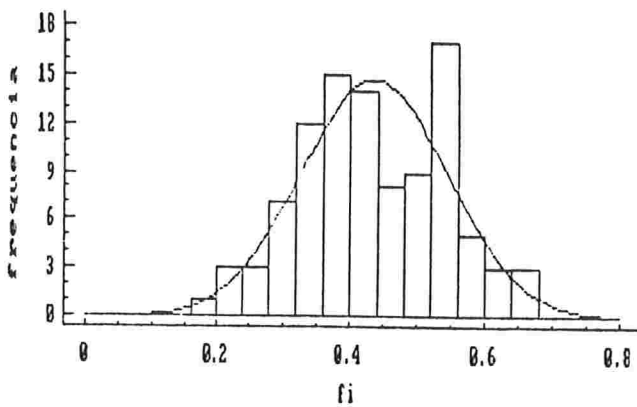
FIGURA 4.20 — Curvas de nível para o erro médio L_1 , modelo ARMA(1,1) com $\phi_1 = 0,5$ e $\theta_1 = 0,5$, distribuição de Laplace, 400 observações.



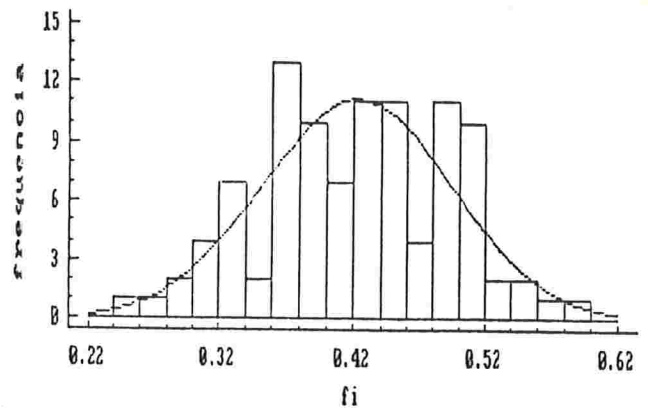
25



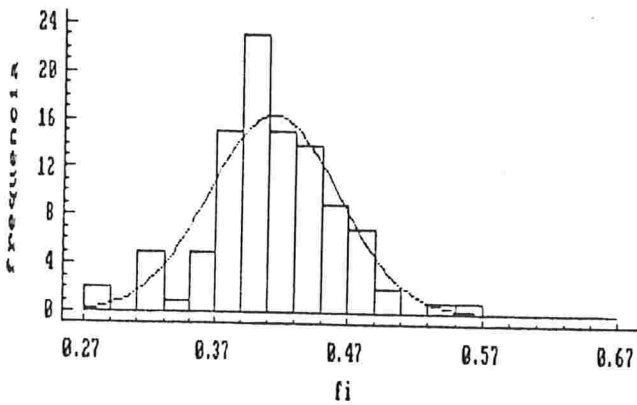
50



100



200



400

FIGURA 4.21 — Afastamento da normalidade das estimativas do parâmetro $\phi = 0,4$, modelo ARMA(1,1), distribuição de Laplace, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

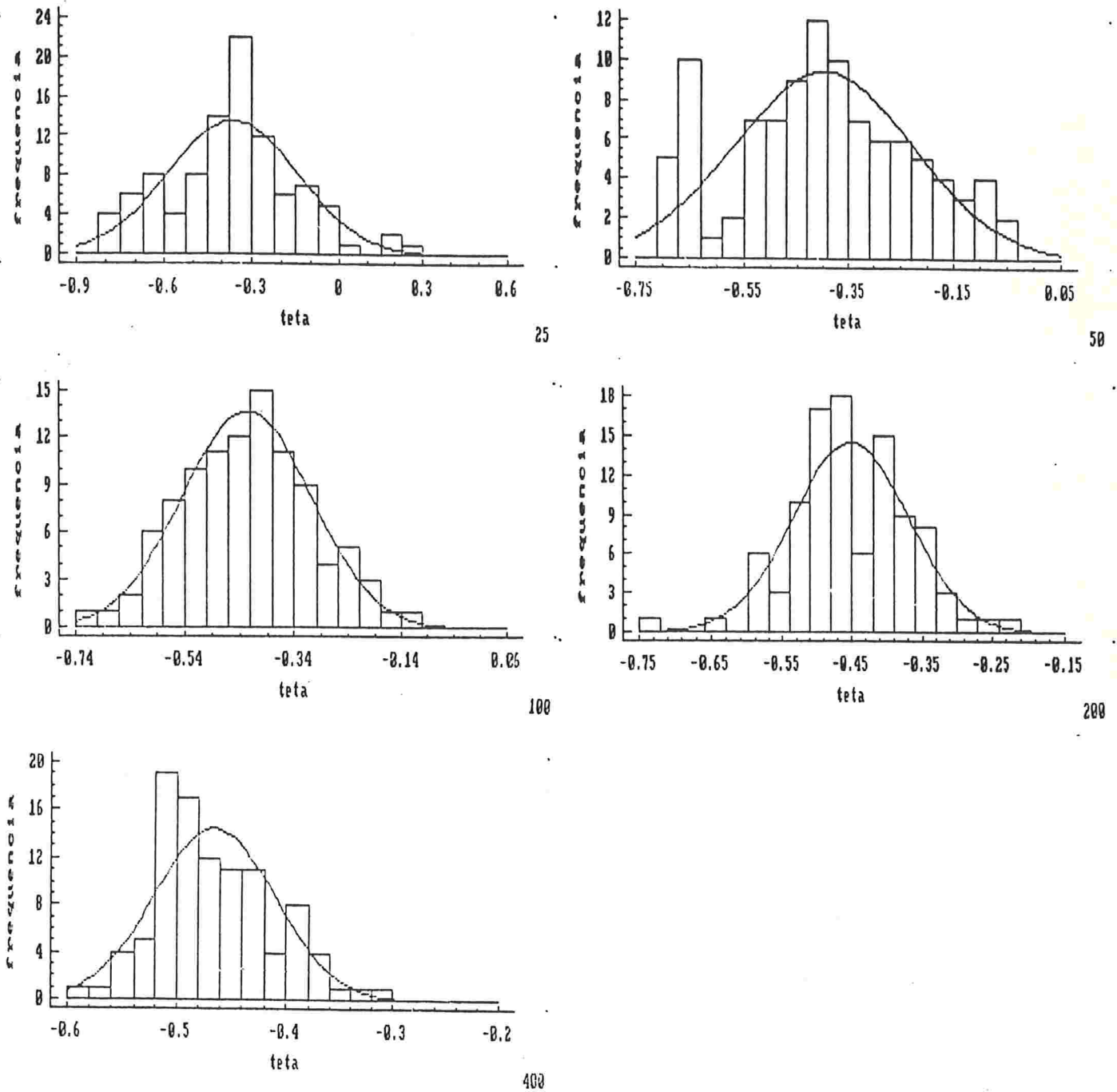


FIGURA 4.22 — Afastamento da normalidade das estimativas do parâmetro $\theta = -0,5$, modelo ARMA(1,1), distribuição de Laplace, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

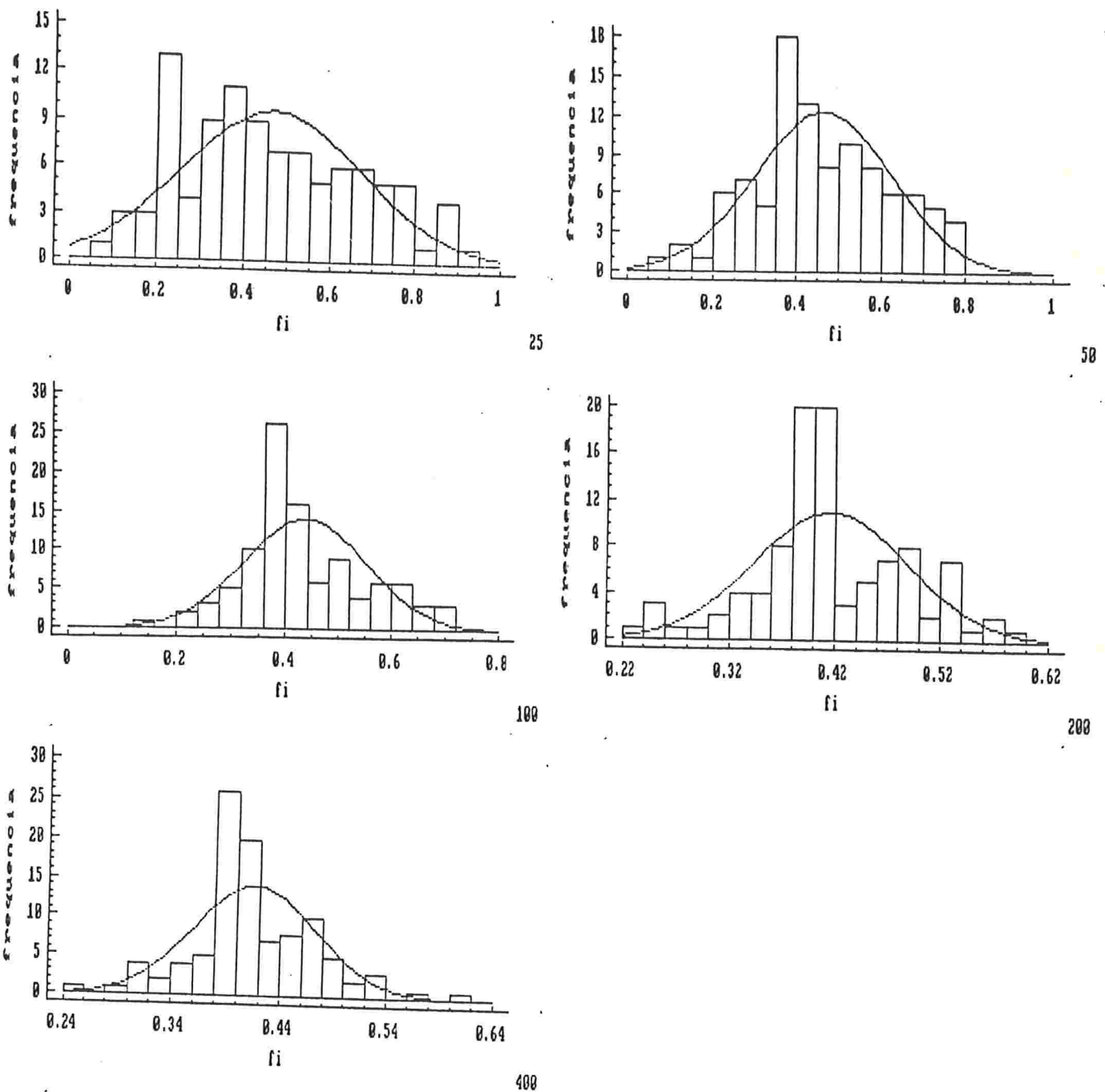
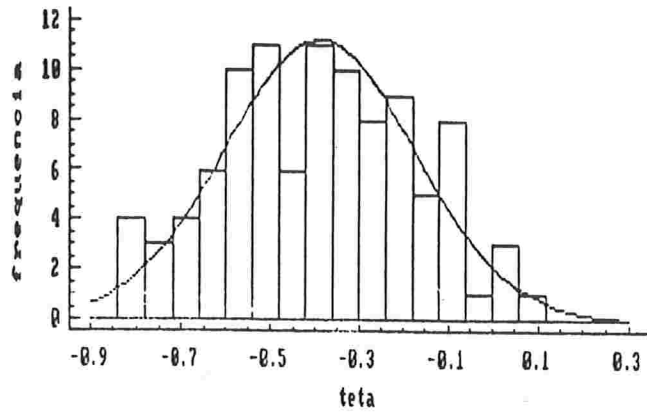
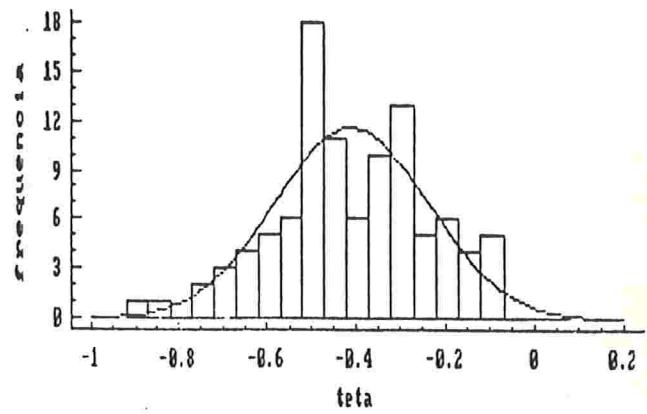


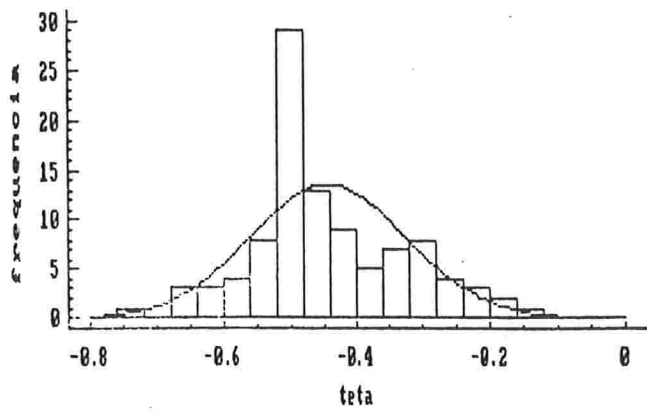
FIGURA 4.23 — Afastamento da normalidade das estimativas do parâmetro $\phi = 0,4$, modelo ARMA(1,1), distribuição normal, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.



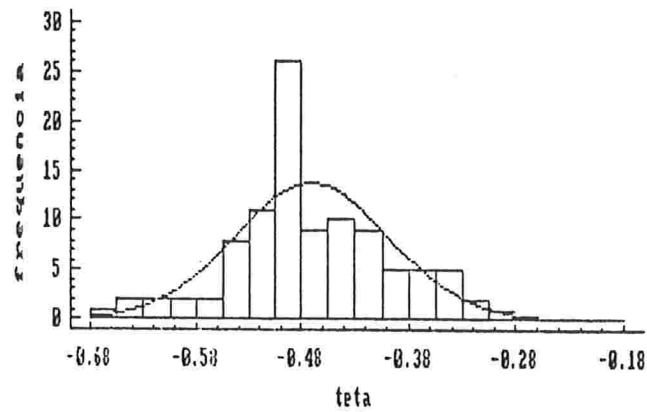
25



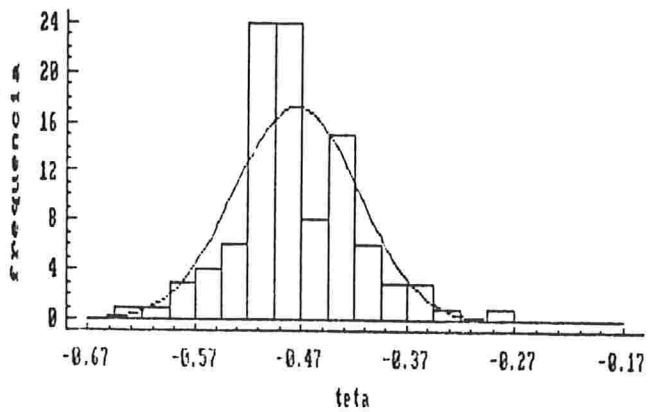
50



100



200



400

FIGURA 4.24 — Afastamento da normalidade das estimativas do parâmetro $\theta = -0,5$, modelo ARMA(1,1), distribuição normal, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

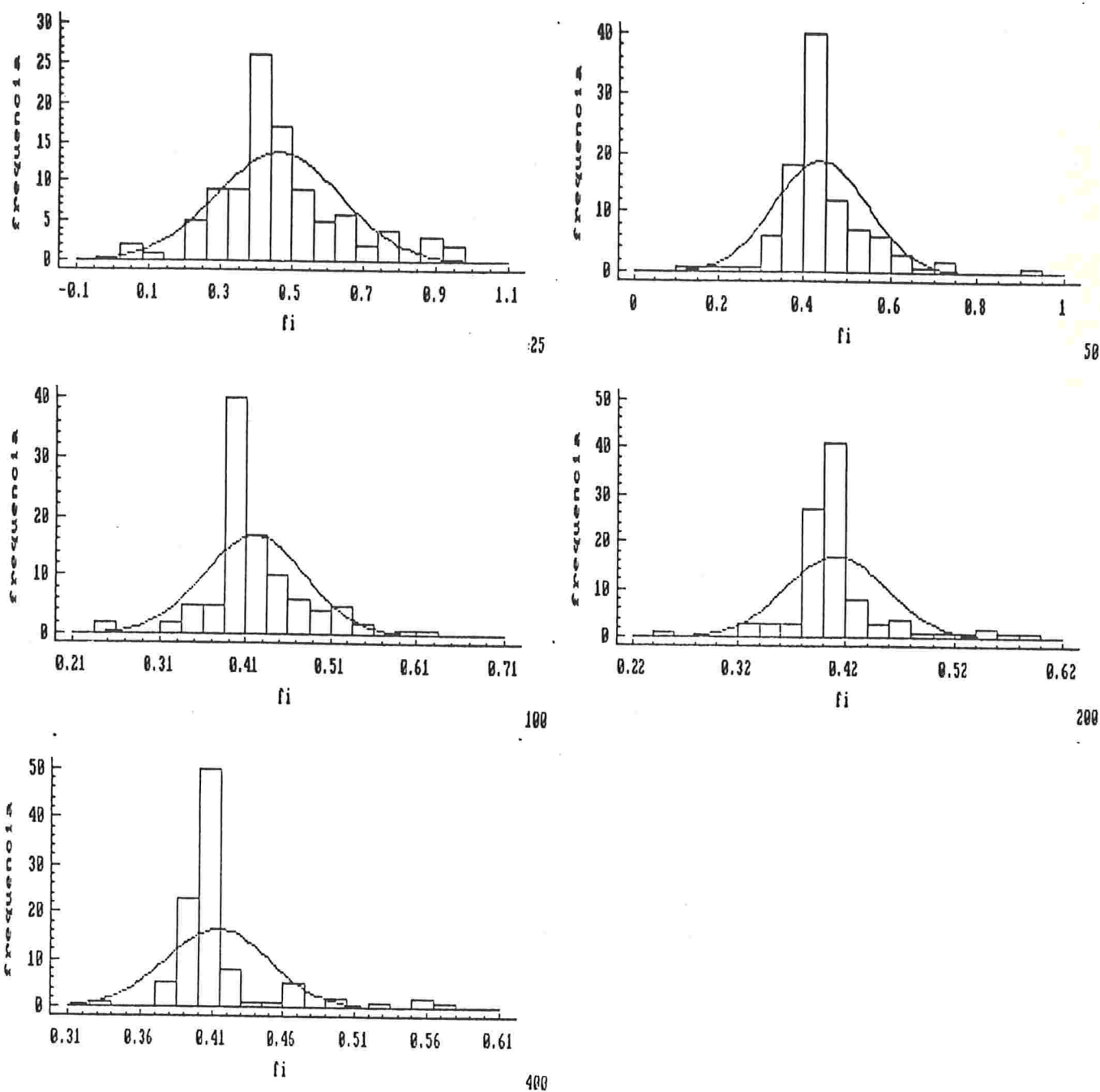


FIGURA 4.25 — Afastamento da normalidade das estimativas do parâmetro $\phi = 0,4$, modelo ARMA(1,1), distribuição de Cauchy, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

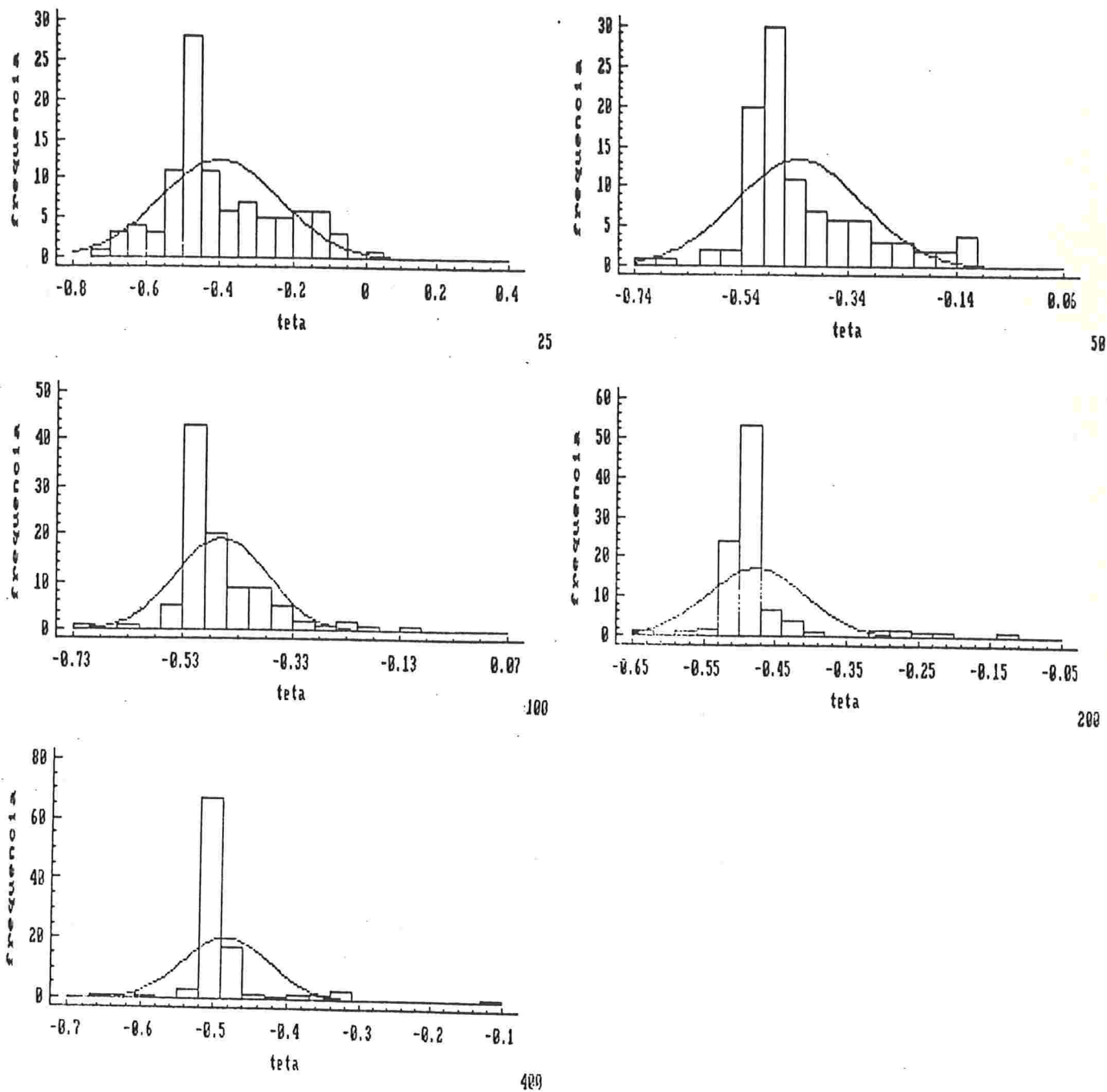


FIGURA 4.26 — Afastamento da normalidade das estimativas do parâmetro $\theta = -0,5$, modelo ARMA(1,1), distribuição de Cauchy, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

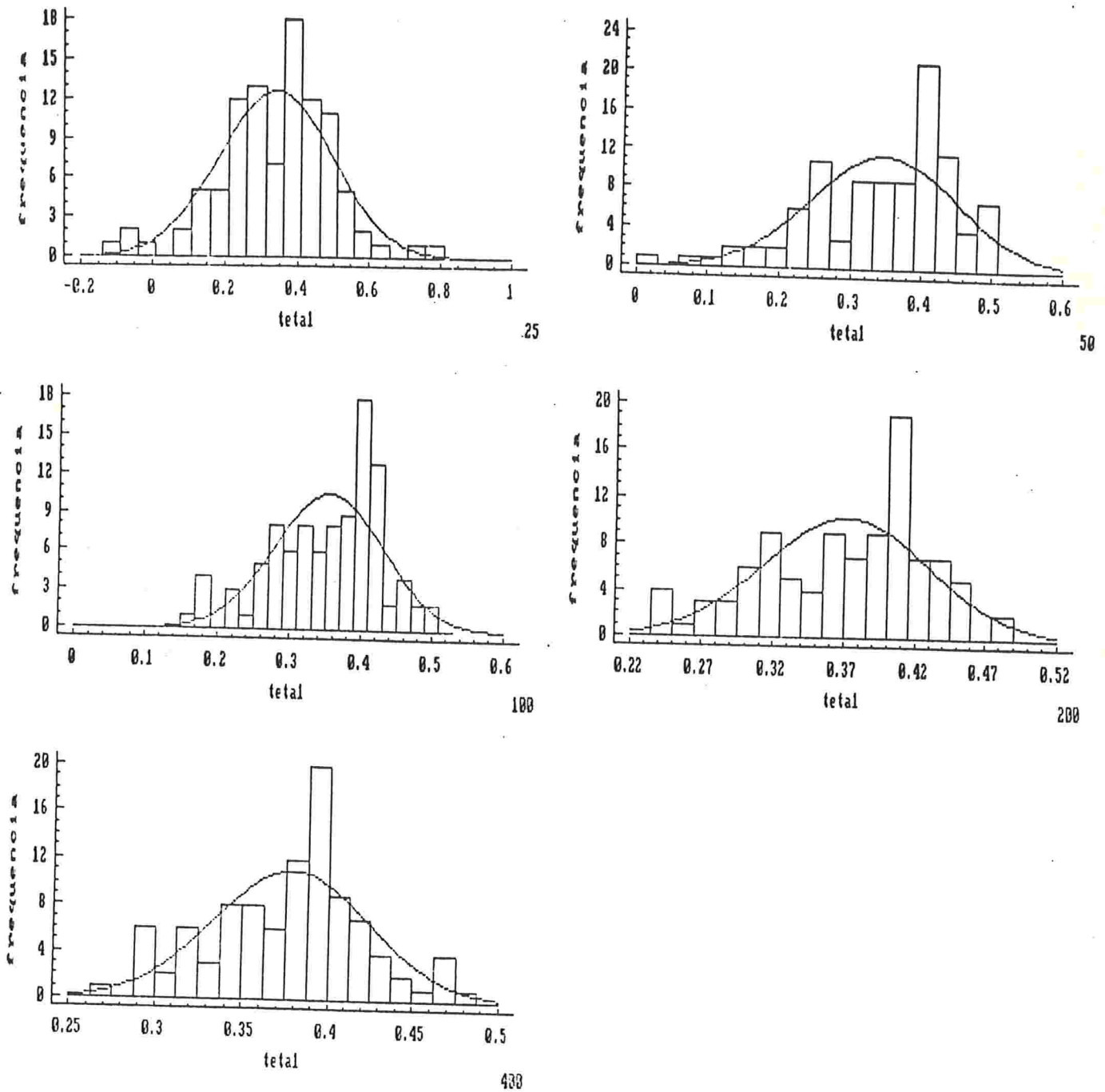


FIGURA 4.27 — Afastamento da normalidade das estimativas do parâmetro $\theta_1 = 0,4$, modelo MA(2), distribuição de Laplace, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

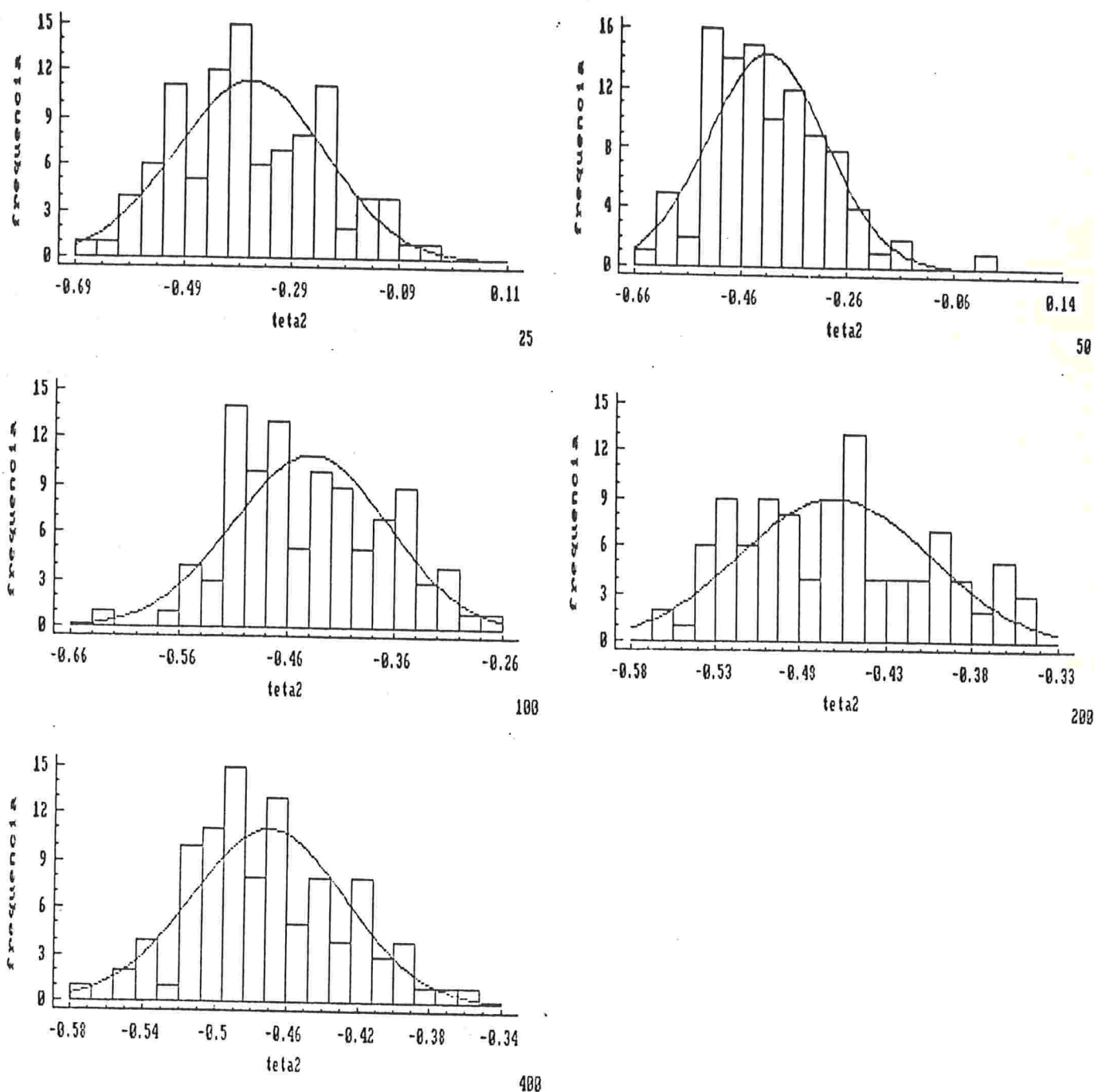


FIGURA 4.28 — Afastamento da normalidade das estimativas do parâmetro $\theta_2 = -0,5$, mo- delo MA(2), distribuição de Laplace, 100 réplicas de séries geradas, amostras de tama- nho 25, 50, 100, 200 e 400 observações.

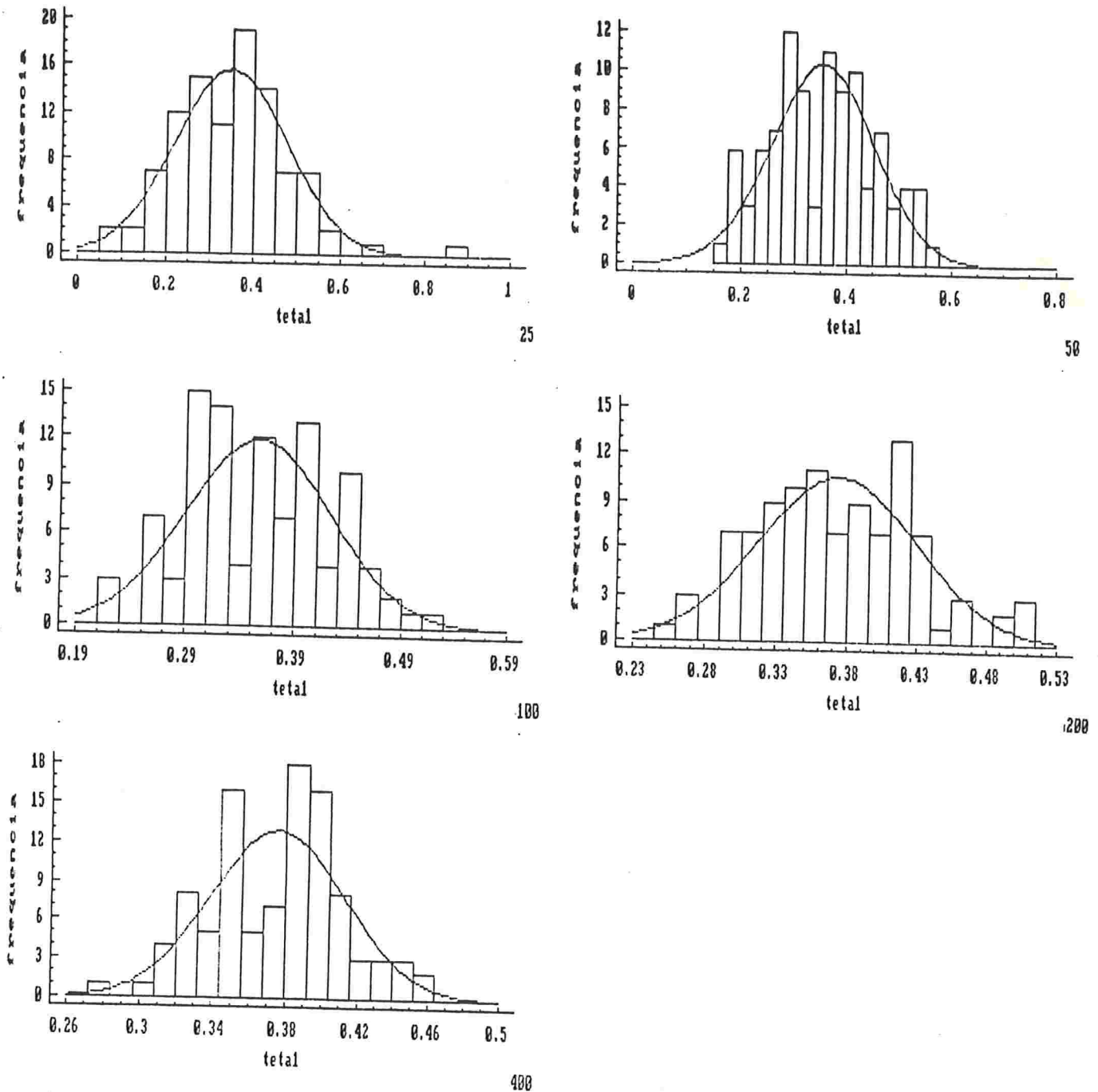


FIGURA 4.29 — Afastamento da normalidade das estimativas do parâmetro $\theta_1 = 0,4$, modelo MA(2), distribuição normal, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

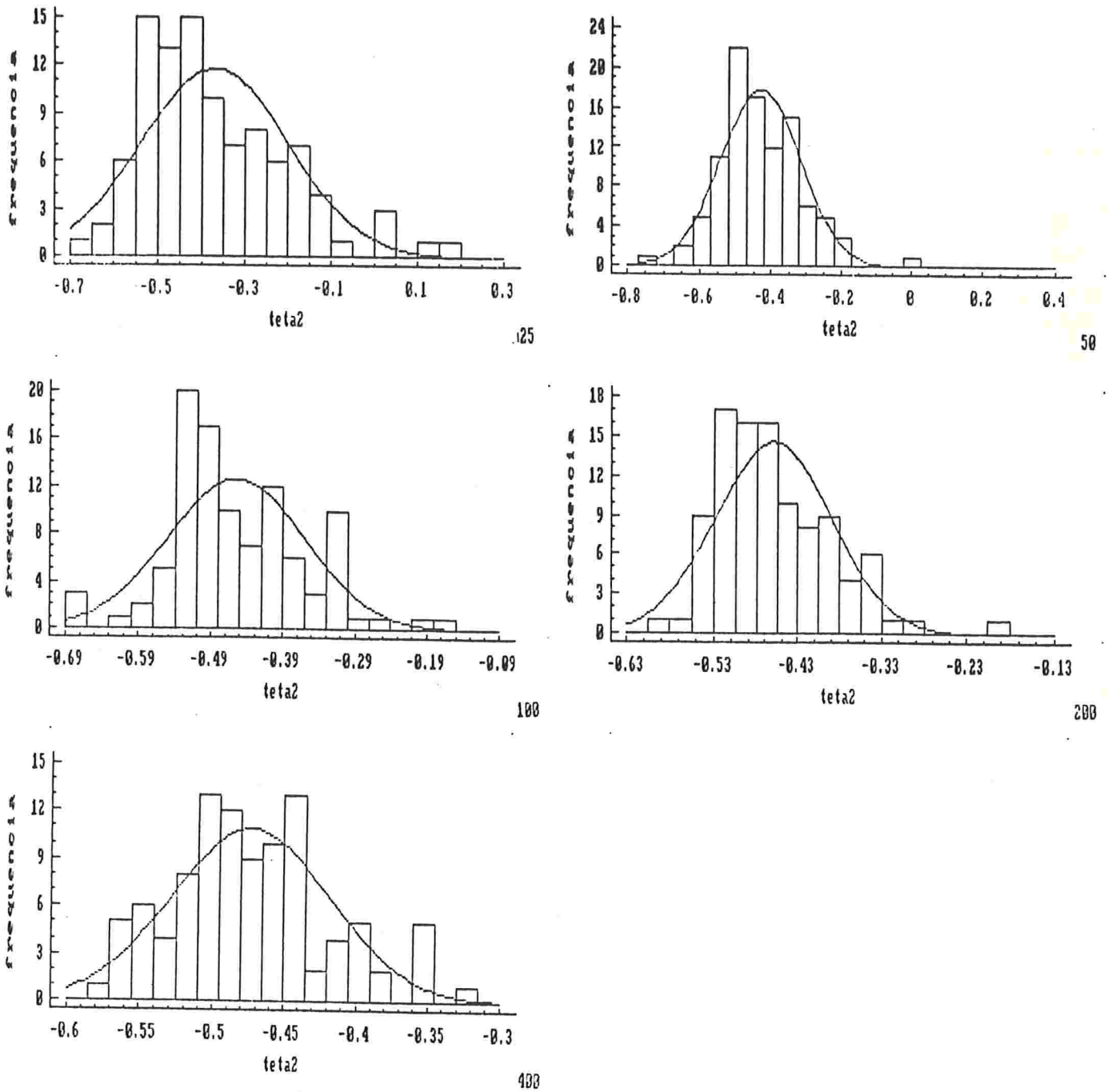


FIGURA 4.30 — Afastamento da normalidade das estimativas do parâmetro $\theta_2 = -0,5$, modelo MA(2), distribuição normal, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

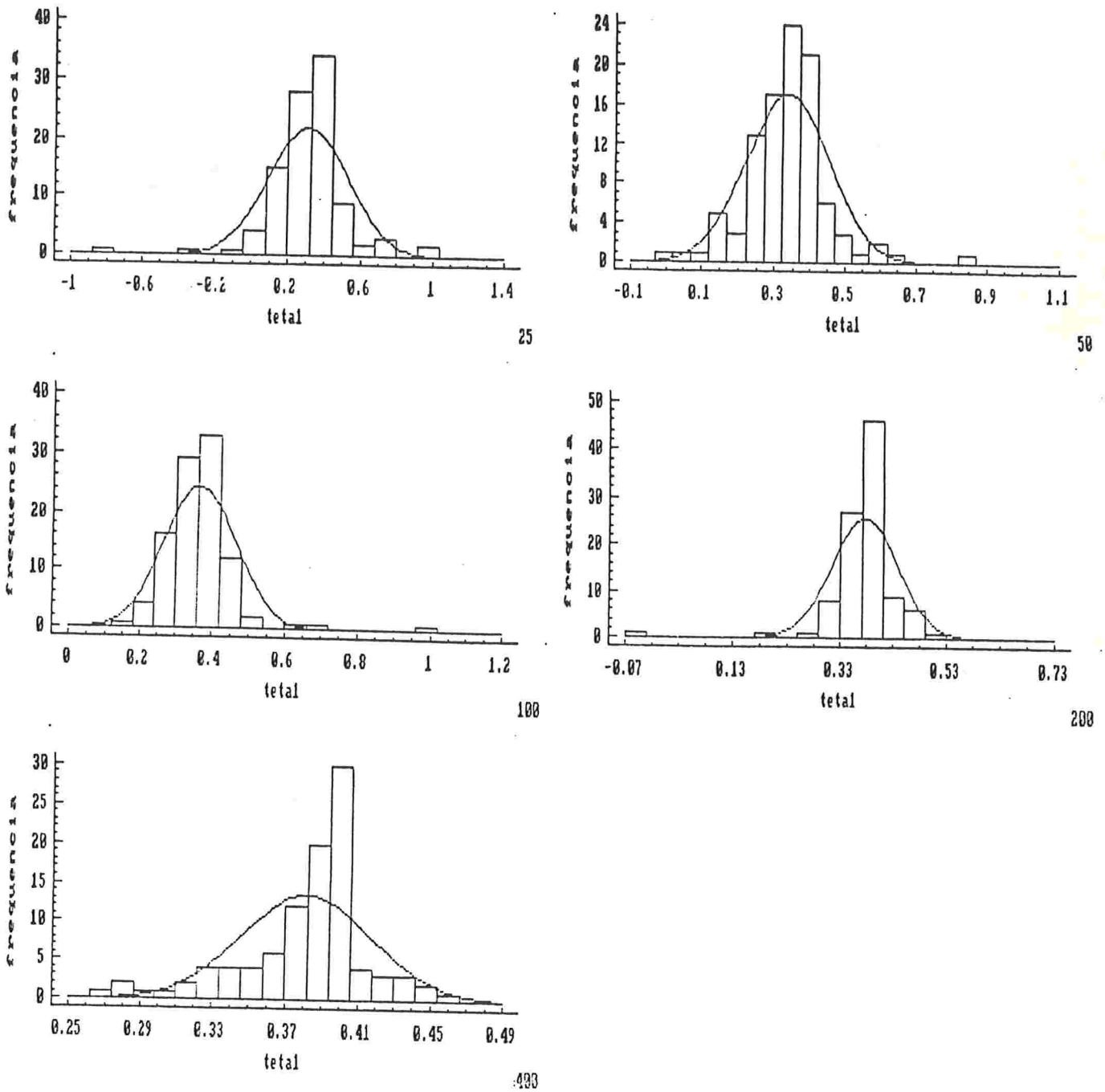


FIGURA 4.31 — Afastamento da normalidade das estimativas do parâmetro $\theta_1 = 0,4$, modelo MA(2), distribuição de Cauchy, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

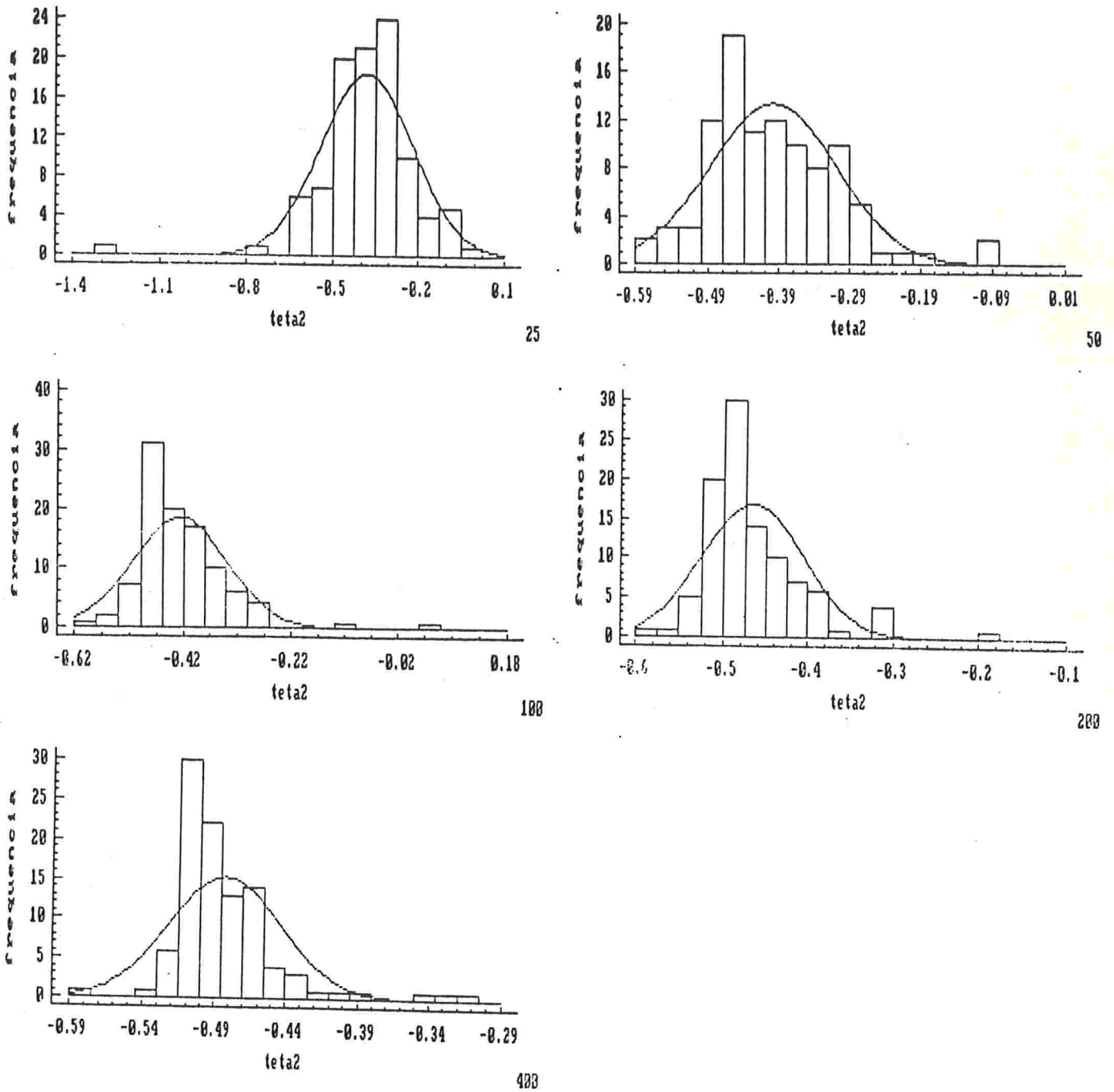


FIGURA 4.32 — Afastamento da normalidade das estimativas do parâmetro $\theta_2 = -0,5$, modelo MA(2), distribuição de Cauchy, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

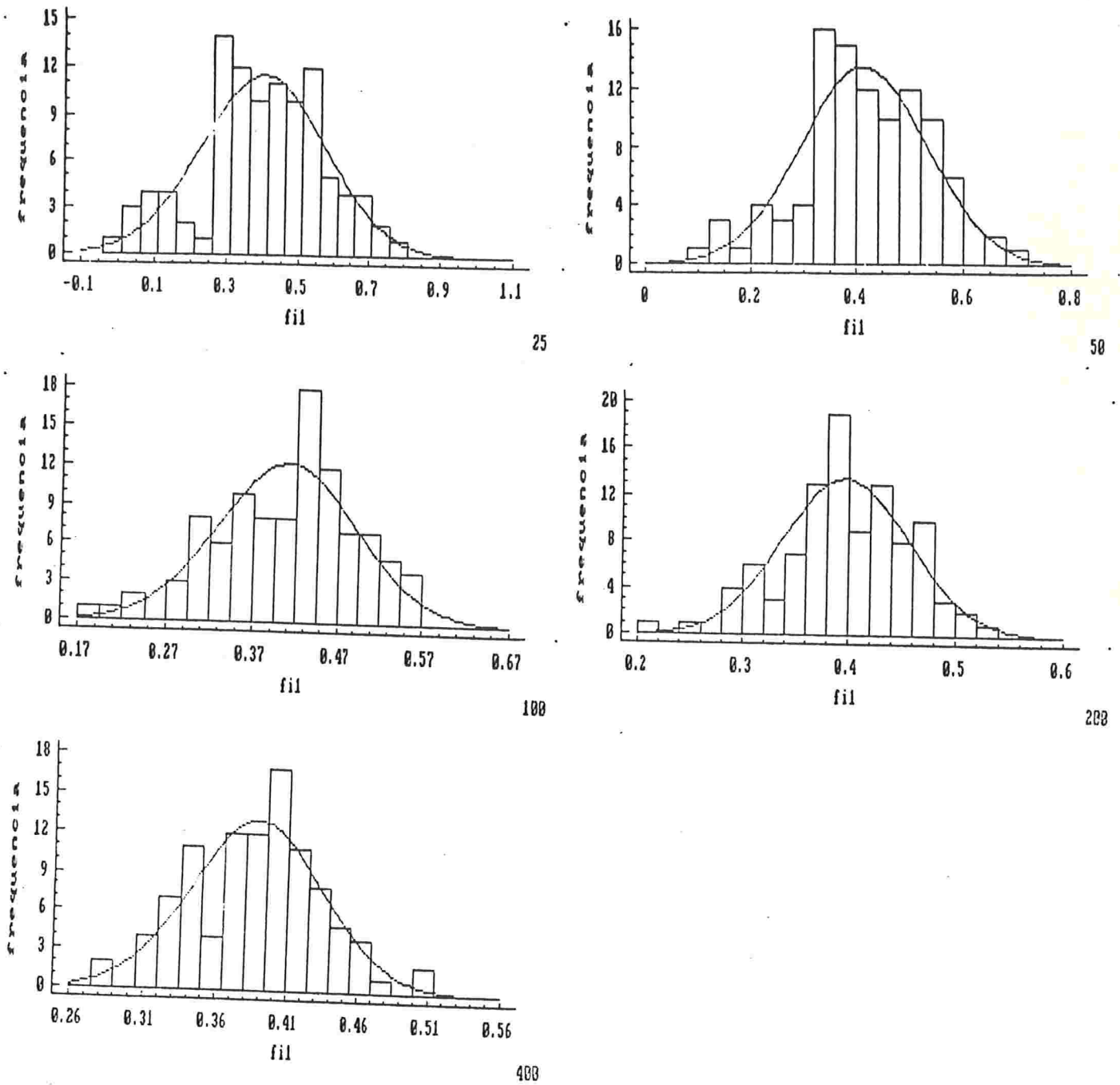


FIGURA 4.33 — Afastamento da normalidade das estimativas do parâmetro $\phi_1 = 0,4$, modelo AR(2), distribuição de Laplace, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

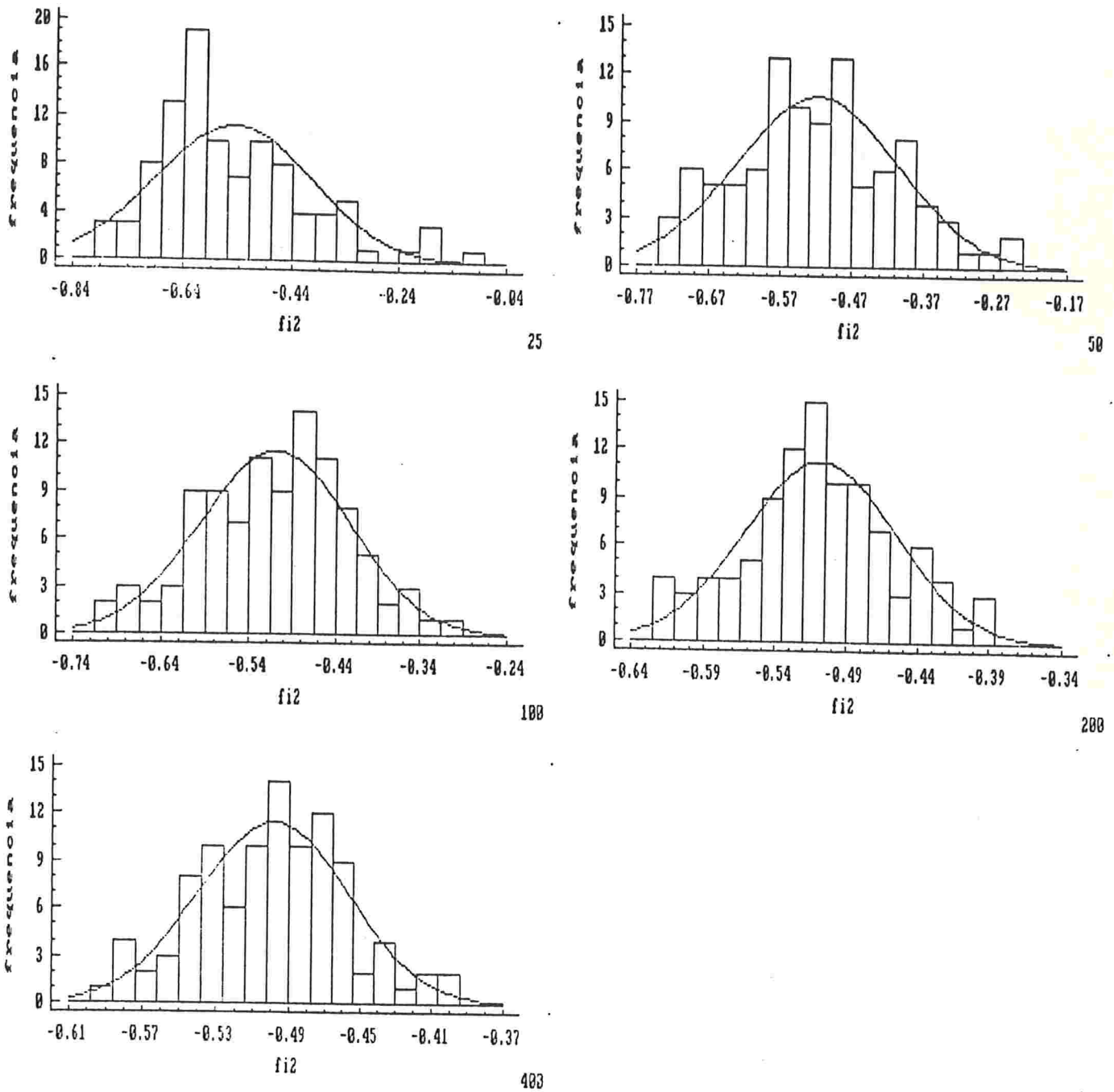


FIGURA 4.34 — Afastamento da normalidade das estimativas do parâmetro $\phi_2 = -0,5$, modelo AR(2), distribuição de Laplace, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

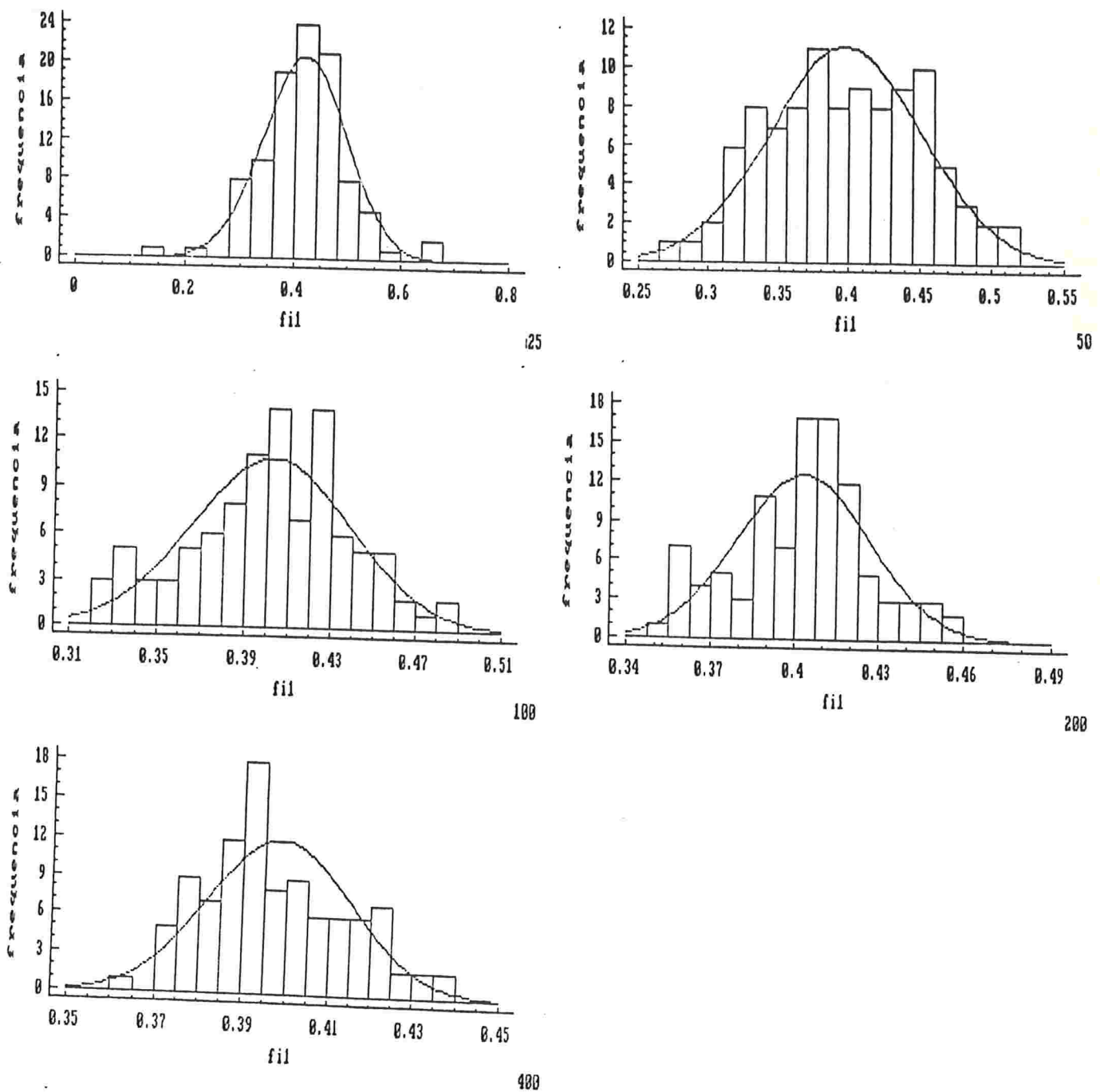


FIGURA 4.35 — Afastamento da normalidade das estimativas do parâmetro $\phi_1 = 0,4$, modelo AR(2), distribuição normal, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

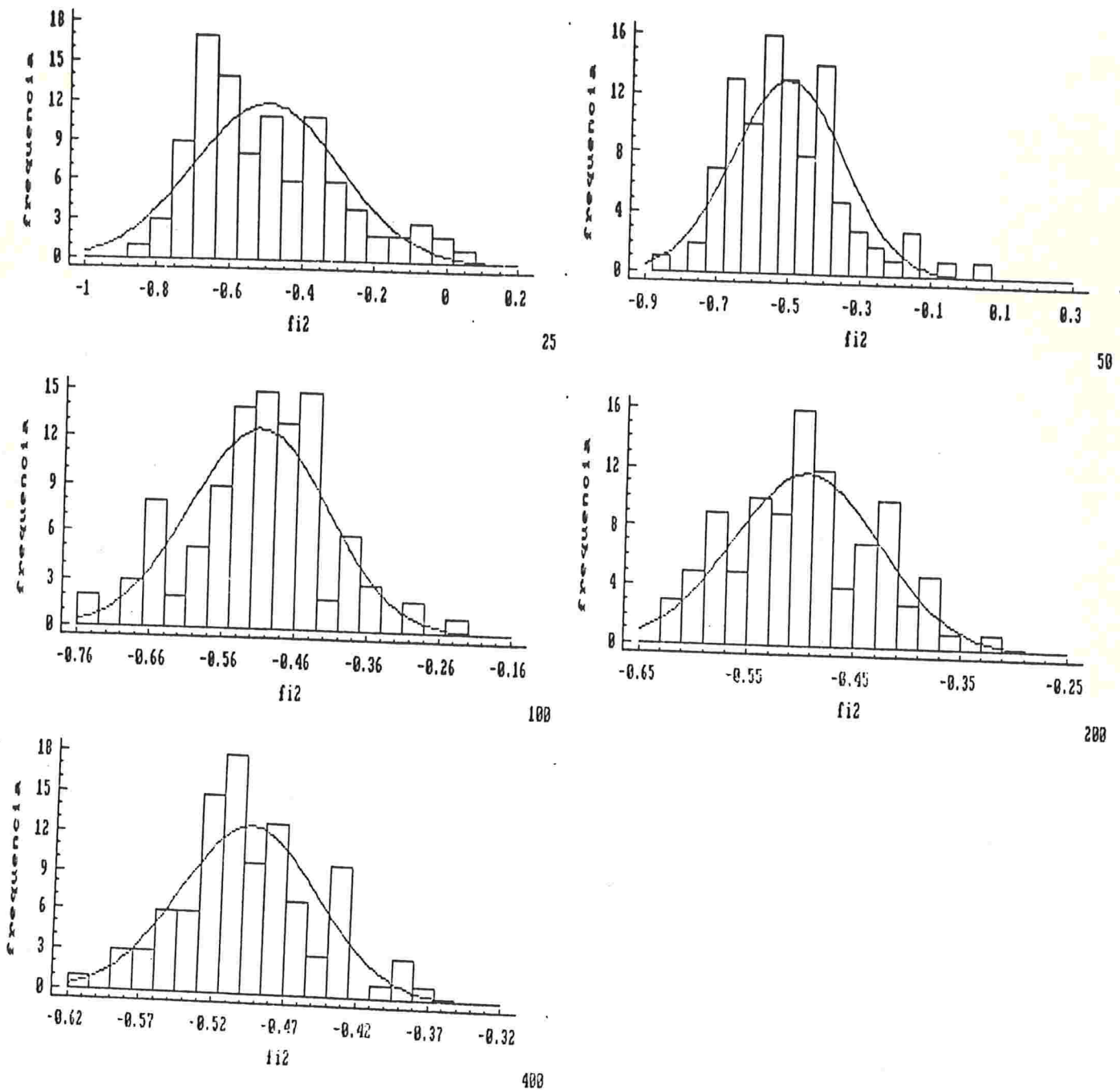


FIGURA 4.36 — Afastamento da normalidade das estimativas do parâmetro $\phi_2 = -0,5$, modelo AR(2), distribuição normal, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

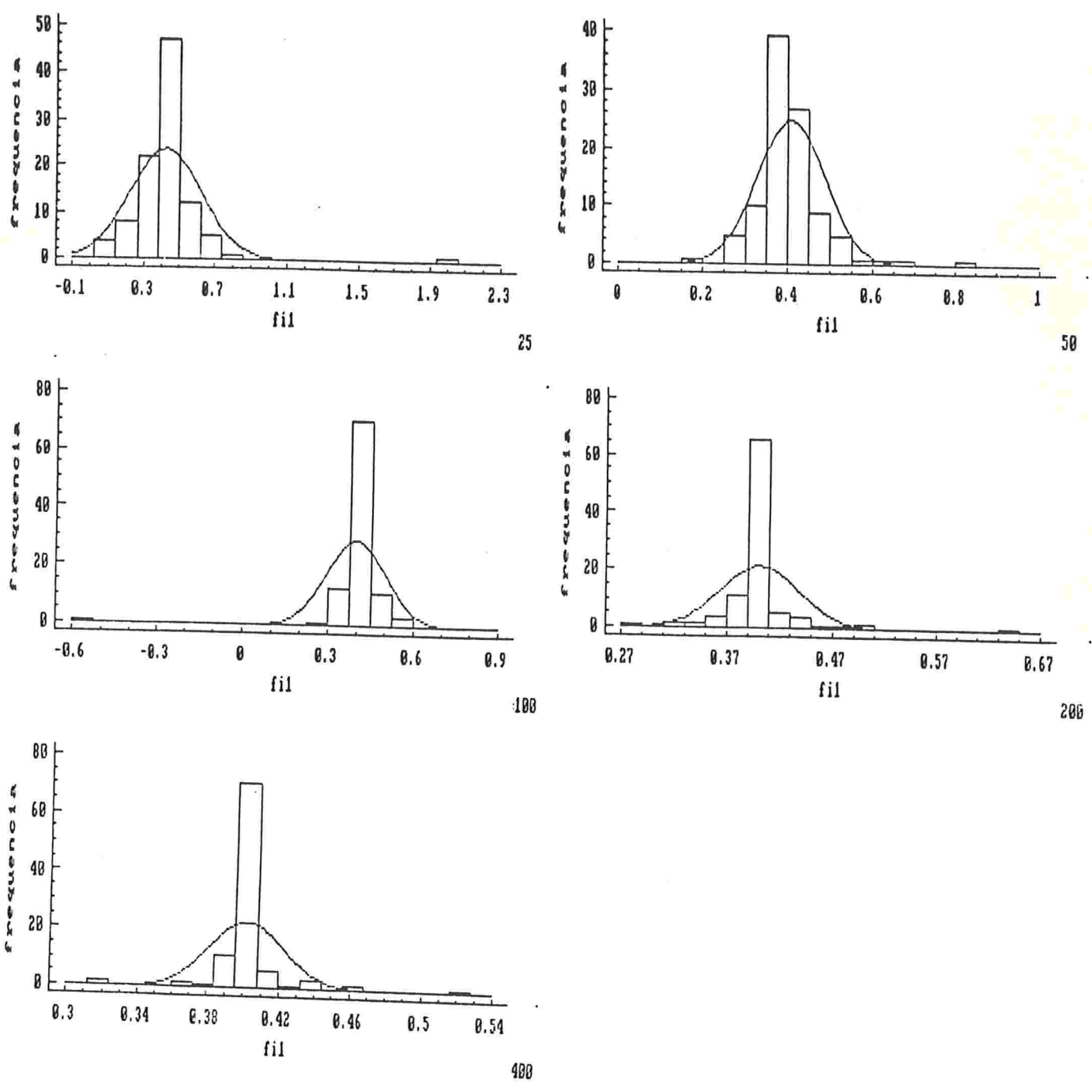


FIGURA 4.37 — Afastamento da normalidade das estimativas do parâmetro $\phi_1 = 0,4$, modelo AR(2), distribuição de Cauchy, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

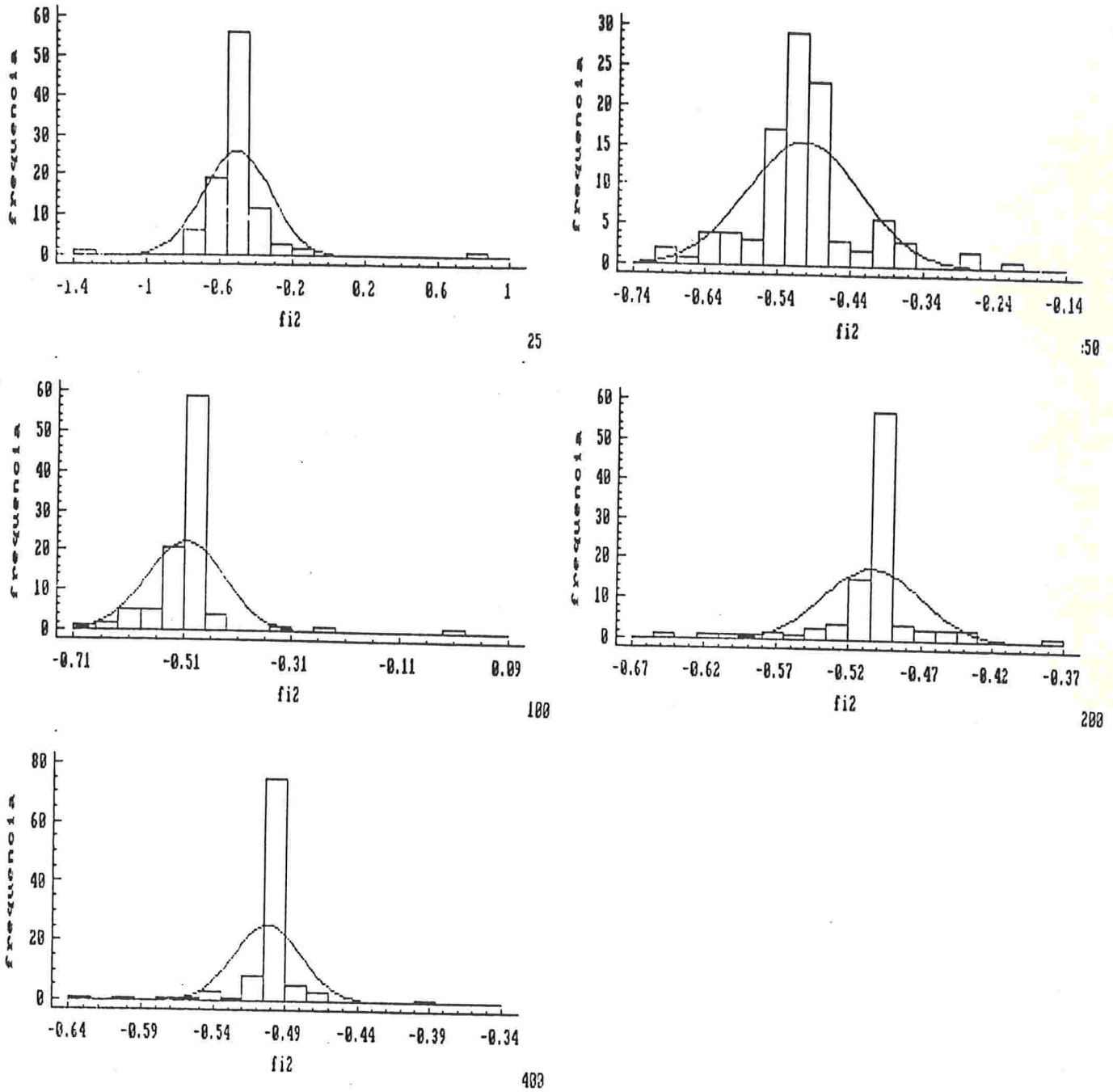


FIGURA 4.38 — Afastamento da normalidade das estimativas do parâmetro $\phi_2 = -0,5$, modelo AR(2), distribuição de Cauchy, 100 réplicas de séries geradas, amostras de tamanho 25, 50, 100, 200 e 400 observações.

CAPÍTULO 5

CONCLUSÕES E RECOMENDAÇÕES

O presente estudo mostra que as estimativas L_1 podem ser úteis no contexto de análise de séries temporais geradas por processos ARMA. Concluimos que:

- a) é possível considerar processos ARMA estacionários e inversíveis com variância infinita;
- b) as estimativas L_1 são equivalentes às de máxima verossimilhança quando o ruído tem distribuição de Laplace;
- c) o problema de estimação L_1 dos parâmetros de um modelo ARMA é equivalente a um problema de programação não linear;
- d) o uso do procedimento de *backforecasting* no cálculo dos resíduos amostrais torna a função objetivo não convexa;
- e) as estimativas L_1 admissíveis são pontos interiores de um conjunto compacto;
- f) as estimativas L_1 são fortemente consistentes;
- g) as estimativas L_1 podem não ser normalmente distribuídas, baseado em resultados de simulação.

Finalizando, como este estudo não tem a pretensão de ser um tratado completo

sobre o assunto, mas, apenas uma etapa de uma longa escalada, destacamos a seguir alguns pontos que a nosso ver merecem vir a ser estudados:

- a) a distribuição assintótica das estimativas;
- b) a estimação L_1 com resíduos calculados por outros procedimentos que não o de *backforecasting*;
- c) a previsão no contexto da norma L_1 , já que essa é uma das principais finalidades da análise de séries temporais;
- d) a identificação do modelo no contexto da norma L_1 ;
- e) simulação com maior número de réplicas, com maior número de observações e com outras distribuições do ruído;
- f) comparação entre estimadores;
- g) comparação entre algoritmos de estimação ;
- h) extensão dos estudos para modelos sazonais, modelo de função de transferência e modelo ARMA vetorial.

REFERÊNCIAS

- [1] ABRAMOWITZ, M. & STEGUN, I. A. (1972). *Handbook of mathematical functions: with formulas, graphs and mathematical tables*. New York, Dover.
- [2] AMEMYIA, T. (1982). Two stage least absolute deviations estimators. *Econometrica*, **50**: 689–711.
- [3] AN, H. -Z. & CHEN, Z. -G. (1982). On convergence of LAD estimates in autoregression with infinite variance. *Journal of Multivariate Analysis*, **12**:335–345.
- [4] ANDERSON, T. W. (1971). *The statistical analysis of time series*. New York, John Wiley.
- [5] ANDERSON, D. H. & OSBORNE, M. R. (1977). Discrete, nonlinear approximation problems in polyhedral norms. *Numerische Mathematik*, **28**:143–156.
- [6] ARMSTRONG, R. D. & FROME, E. L. (1976). A comparison of two algorithms for absolute deviation curve fitting. *Journal of the American Statistical Association*, **71**:328–330.
- [7] ARMSTRONG, R. D. & FROME, E. L. (1977). A special purpose linear programming algorithm for obtaining least absolute value estimators in a linear model with dummy variables. *Communications in Statistics*, **B6**:383–398.
- [8] ARMSTRONG, R. D. ; FROME, E. L. & KUNG, D. S. (1979). A revised simplex algorithm for the absolute deviations curve fitting problem. *Communications in Statistics*, **B8**:175–190.

- [9] ARMSTRONG, R. D. & KUNG, M. T. (1978). Algorithm AS 132: least absolute value estimates for a simple linear regression problem. *Applied Statistics*, **27**:363–366.
- [10] ÅSTRÖM, K. J. & BOHLIN, T. (1966). Numerical identification of linear dynamic systems from normal operating records. In: HAMMOND, P. H. (ed.) *Theory of self-adaptive control systems*. New York, Plenum. ⁽¹⁾
- [11] BALET-LAWRENCE, S. (1975). *Estimation of the parameters in an implicit model by minimizing the sum of absolute values of order p*. Raleigh, Department of Statistics, North Carolina State University. (PhD thesis).
- [12] BARRODALE, I. & ROBERTS, F. D. K. (1973). An improved algorithm for discrete L_1 linear approximation. *SIAM Journal of Numerical Analysis*, **10**:839–848.
- [13] BARRODALE, I. & ROBERTS, F. D. K. (1977). Algorithms for restricted least absolute value estimation. *Communications in Statistics*, **B6**:353–363.
- [14] BASSET, G. & KOENKER, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, **73**:618–622.
- [15] BLOOMFIELD, P. & STEIGER, W. L. (1983). *Least absolute deviations: theory, applications and algorithms*. Boston, Birkhäuser.
- [16] BOSCOVICH, R. G. (1757). De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa. *Bononiensi Scientiarum et Artum Instituto Atque Academia Commentarii*, **4**:353–396. ⁽²⁾
- [17] BOSCOVICH, R. G. (1760). De recentissimis graduum dimensionibus, et figura, ac magnitudine terrae inde derivanda. *Philosophiae Recentioris, a Benedicto Stay in Romano Archigynasis Publico Eloquentare Professore, versibus traditae, Libri X, cum adnotianibus et Supplementis P. Rogerii Joseph Boscovich, S. J.*, Tomus II, p.p.406–426, esp.420–425. Romae. ⁽²⁾
- [18] BOX, G. E. P. & JENKINS, G. M. (1976). *Time series analysis: forecasting and control*. San Francisco, Holden-Day.

- [19] BREIMAN,L. (1968). *Probability*. Reading, Addison-Wesley.
- [20] CHARALAMBOUS,C. (1979). On conditions for optimality of the nonlinear ℓ_1 problem. *Mathematical Programming*, 17:123–135.
- [21] CHARNES,A. ; COOPER,W. W. & FERGUSON,R. O. (1955). Optimal estimation of executive compensation by linear programming. *Management Sciences*, 1:138–151.
- [22] CLARKE,F. H. (1983). *Optimization and nonsmooth analysis*. New York, John Wiley.
- [23] COGGER,K. O. (1979). Time series analysis and forecasting with an absolute error criterion. In: MAKRIDAKIS,S. & WHEELWRIGHT, S. C. (eds.). *Forecasting studies in the management sciences*. v.12. p.189–201.
- [24] CRAMER,H. (1946). *Mathematical methods of Statistics*. Princeton.
- [25] DAMSLETH,E. & EL-SHAARAWI,A. H. (1989). ARMA models with double - exponentially distributed noise. *Journal of the Royal Statistical Society B*, 51:61–69.
- [26] DELAMBRE,J. B. (1813). *Abrégé d'astronomie, ou leçons élémentaire d'astronomie theorique et pratique*. Paris, Courcier. (2)
- [27] DIELMAN,T. (1984). Least absolute value estimation in regression models: an annotated bibliography. *Communications in Statistics*, 13:513–541.
- [28] DIELMAN,T. & PFAFFENBERGER,R. (1982). LAV (least absolute value) estimation in linear regression: a review. *TIMS Studies in the Management Sciences* 19:31–52.
- [29] DIELMAN,T. & PFAFFENBERGER,R. (1988). Least absolute value regression: necessary sample sizes to use normal theory inference procedures. *Decision Sciences*, 19:734–743.
- [30] DODGE,Y. (1987). *Statistical data analysis based on the L_1 -norm and related methods*. Amsterdam, Elsevier.
- [31] DUPAČOVÁ,J. (1987). Asymptotic properties of restricted L_1 -estimates of regression. In: DODGE,Y. (ed.) *Statistical data analysis based on the L_1 -norm and related methods*. Amsterdam, Elsevier. p.263–274.

- [32] DUPAČOVÁ, J. & WETS, R. (1988). Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *The Annals of Statistics*, **16**:1517–1549.
- [33] EDGEWORTH, F. Y. (1887a). On observations relating to several quantities. *Hermathena*, **6**:279–285. ⁽²⁾
- [34] EDGEWORTH, F. Y. (1887b). A new method of reducing observations relating to several quantities. *Philosophical Magazine*, **24**:222–223. ⁽²⁾
- [35] EDGEWORTH, F. Y. (1888). On a new method of reducing observations relating to several quantities. *Philosophical Magazine*, **25**:184–191. ⁽²⁾
- [36] EKELAND, I. & TURNBULL, T. (1983). *Infinite-dimensional optimization and convexity*. Chicago, Univ. Chicago Press.
- [37] EL-ATTAR, R. A. ; VIDYASAGAR, M. & DUTTA, S. R. K. (1979). An algorithm for ℓ_1 -norm minimization with application to nonlinear ℓ_1 -approximation. *SIAM Journal of Numerical Analysis*, **16**:70–86.
- [38] EULER, L. (1749). *Pièce qui remporté le Prix de l'Academie Royale des Sciences en 1748, sur les inegalités du mouvement de Saturn et de Jupiter*. Paris. ⁽²⁾
- [39] FECHNER, G. T. (1874). Ueber den Ausgangsweth der kleinsten Abweichungssumme, dessen Bestimmung, Verwendung und Verallgemeinerung. *Abhandlungen der Königl. Sächsische Gesellschaft der Wissenschaft zu Leipzig*, **18** (Math. Phys. Classe 11):1–76. ⁽²⁾
- [40] FELLER, W. (1968). *An introduction to probability theory and its applications*. New York, John Wiley.
- [41] FIACCO, A. V. & McCORMICK, G. P. (1966). Extensions of SUMT for nonlinear programming: equality constraints and extrapolation. *Management Sciences*, **12**:816–829.
- [42] FLETCHER, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, **13**:317–322. ⁽³⁾

- [43] FRÉCHET, M. (1935). Sur les précisions comparées de la moyenne et de la médiane. *Adtuárske Vědy*, 5:29–34. (2)
- [44] FULLER, W. A. (1976). *Introduction to statistical time series*. New York, John Wiley.
- [45] GAUSS, C. F. (1806). II comet vom Jahr 1805. *Monatliche Correspondenz zur Beförderung der Erd- und Himmels-Kunde*, 14:181–186. (2)
- [46] GAUSS, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Hamburgi, Frid. Perthes et I.H.Besser. (2)
- [47] GENTLE, J. E. (1977). Least absolute values estimation: an introduction. *Communications in Statistics*, B6:313–328.
- [48] GENTLE, J. E. ; KENNEDY, W. J. & SPOSITO, V. A. (1977). On least absolute values estimation. *Communications in Statistics*, A6:839–845.
- [49] GNEDENKO, B. V. (1962). *The theory of probability*. New York, Chelsea.
- [50] GONIN, E. & MONEY, A. H. (1987). A review of computational methods for solving the nonlinear L_1 -norm estimation problem. In: DODGE, Y. (ed.) *Statistical data analysis based on the L_1 -norm and related methods*. Amsterdam, Elsevier. p.117–129.
- [51] GONZALEZ, T. ; SAHNI, S. & FRANTA, W. R. (1977). An efficient algorithm for the Kolmogorov-Smirnov and Lilliefors tests. *ACM Transactions on Mathematical Software*, 3:60–64.
- [52] GRANGER, C. W. J. & ORR, D. (1972). “Infinite variance” and research strategy in time series analysis. *Journal of the American Statistical Association*, 67:275–285.
- [53] GRENANDER, U. & ROSENBLATT, M. (1957). *Statistical analysis of stationary time series*. New York, John Wiley.
- [54] GROSS, S. & STEIGER, W. L. (1979). Least absolute deviation estimates in autoregression with infinite variance. *Journal of Applied Probability*, 16:104–116.
- [55] HANNAN, E. J. & KANTER, M. (1977). Autoregressive processes with infinite variance. *Journal of Applied Probability*, 14:441–445.
- [56] HARTER, L. H. (1974–76). The method of least squares and some alternatives: parts

- I-IV. *International Statistical Review*, 42:147-174, 235-264, 43:1-44, 125-190, 269-278, 44:113-159.
- [57] HÖNIG, C. S. (1971). *Introdução às funções de uma variável complexa*. São Paulo, IME-USP.
- [58] KANTER, M. & STEIGER, W. L. (1974). Regression and autoregression with infinite variance. *Advances in Applied Probability*, 6:768-783.
- [59] KARST, O. J. (1958). Linear curve fitting using least deviations. *Journal of the American Statistical Association*, 53: 118-132.
- [60] LAPLACE, P. S. (1886). Mémoire sur la figure de la terre. *Mémoires de l'Académie Royale des Sciences de Paris, Année 1783*. 17-46. (2)
- [61] LAPLACE, P. S. (1793). Sur quelque points du système du monde. *Mémoires de l'Académie Royale des Sciences de Paris, Année 1789*, 1-87. (2)
- [62] LAPLACE, P. S. (1812). *Théorie analytique des probabilités*. Paris, Courcier. (2)
- [63] LEGENDRE, A. M. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris, Courcier. (2)
- [64] LILLIEFORS, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *American Statistical Association Journal*, 62:399-402.
- [65] MANDELBROT, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36:394-419.
- [66] MANDELBROT, B. (1967). The variation of some other speculative prices. *The Journal of Business*, 40:303-413.
- [67] MARQUARDT, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.*, 11:431-441.
- [68] MARSAGLIA, G. (1962). Random variables and computers. In: KOZESNIK, J. (ed.) *Information theory statistical decision functions random processes: Transactions of the Third Prague Conference*. Prague, Czechoslovak Academy of Sciences. p.499-510.
- [69] McCORMICK G. F. & SPOSITO, V. A. (1975). A note on the L_1 estimation based

- on the median positive quotient. *Applied Statistics*, **24**:347–350.
- [70] NARULA, S. C. & WELLINGTON, J. F. (1977a). Algorithm AS 108: multiple linear regression with minimum sum of absolute errors. *Applied Statistics*, **26**:106–111.
- [71] NARULA, S. C. & WELLINGTON, J. F. (1977b). An algorithm for the minimum sum of weighted absolute errors regression. *Communications in Statistics*, **B6**:341–352.
- [72] NARULA, S. C. & WELLINGTON, J. F. (1982). The minimum sum of absolute errors regression: a state of the art survey. *International Statistical Review*, **50**:317–326.
- [73] NARULA, S. C. & WELLINGTON, J. F. (1984). *An efficient algorithm for the MSAE and the MMAE regression problems*. Richmond, Institute of Statistics, Virginia Commonwealth University. (Working Paper Series, 8).
- [74] NYQUIST, H. (1983). *Robust procedures in time series analysis: a review*. Umeå, Sweden, Department of Statistics, University of Umeå. Presented at the Third International Symposium on Forecasting, Philadelphia.
- [75] NYQUIST, H. (1987). *Recent studies on L_p -norm estimation*. Umeå, Sweden, Institute of Statistics, University of Umeå. (Statistical Research Report, 901).
- [76] OBERHOFER, W. (1982). The consistency of nonlinear regression minimizing the L_1 -norm. *The Annals of Statistics*, **10**:316–319.
- [77] OSBORN, D. R. (1976). Maximum likelihood estimation of moving average processes. *Annals of Economic and Social Measurement*, **5**:75–87.
- [78] OSBORNE, M. R. & WATSON, G. A. (1971). On an algorithm for discrete nonlinear L_1 approximation. *The Computer Journal*, **14**:184–188.
- [79] RHODES, E. C. (1930). Reducing observations by the method of minimum deviations. *Philosophical Magazine*, **9**:974–992.
- [80] RICE, J. R. & WHITE, J. S. (1964). Norms for smoothing and estimation. *SIAM Review*, **6**:243–256.
- [81] ROCKAFELLAR, R. T. (1980). Generalized directional derivatives and subgradients of nonconvex functions. *Can. J. Math.*, **32**:257–280.

- [82] ROCKAFELLAR, R. T. (1981). *The theory of subgradients and its applications to problems of optimization: convex and nonconvex functions*. Berlin, Heldermann Verlag.
- [83] ROCKAFELLAR, R. T. (1983). Generalized subgradients in mathematical programming. In: *Mathematical programming: the state-of-the-art*. Berlin, Springer Verlag. p.368–390.
- [84] ROSENBERG, B. & CARLSON, D. (1977). A simple approximation of the sampling distribution of least absolute residuals regression estimates. *Communications in Statistics*, **B6**:421–437.
- [85] ROYDEN, H. L. (1968). *Real analysis*. London, MacMillan.
- [86] SADOVSKI, A. N. (1974). Algorithm AS 74: L_1 -norm fit of a straight line. *Applied Statistics*, **23**:244–248.
- [87] SCHLOSSMACHER, E. J. (1973). An iterative technique for absolute deviations curve fitting. *Journal of the American Statistical Association*, **68**:857–859.
- [88] SIDDIQUI, M. M. (1962). Approximations to the moments of the sample median. *Annals of Mathematical Statistics*, **33**:157–168.
- [89] SINGLETON, R. R. (1940). A method for minimizing the sum of absolute errors. *Annals of Math. Statistics*, **11**:301–310.
- [90] SLUTZKY, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica*, **5**:105–146.
- [91] SPOSITO, V. A. ; KENNEDY, W. J. & GENTLE, J. E. (1977). Algorithm AS 110: L_p norm fit of a straight line. *Applied Statistics*, **26**:114–118.
- [92] SPOSITO, V. A. ; SMITH, W. & McCORMICK, G. (1978). *Minimizing the sum of absolute deviations*. Göttingen, Vandenhoeck & Ruprecht.
- [93] TAYLOR, L. (1974). Estimation by minimizing the sum of absolute errors. In: ZAREMBKA, P. *Frontiers in econometrics*. New York, Academic Press. p.169–190.
- [94] TISHLER, A. & ZANG, I. (1982). An absolute deviations curve-fitting algorithm for

- nonlinear models. In: ZANAKIS, S. H. & RUSTAGI, J. S. (eds.) *Optimization in Statistics, TIMS Studies in the Management Sciences*, 19, North Holland. p.81-103.
- [95] TURNER, H. H. (1887). On Mr. Edgeworth's method of reducing observations relating to several quantities. *Philosophical Magazine*, 24:466-470. ⁽²⁾
- [96] TURNER, M. (1960). On heuristic estimation methods. *Biometrics*, 16:299-301.
- [97] WAGNER, H. M. (1959). Linear programming techniques for regression analysis. *Journal of the American Statistical Association*, 54:206-212.
- [98] YOHAJ, V. J. & MARONNA, R. A. (1977). Asymptotic behavior of least-squares estimates for autoregressive processes with infinite variances. *The Annals of Statistics*, 5:554-560.
- [99] YULE, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wölfer's sunspot numbers. *Phil. Trans.*, A226:267-298. ⁽⁴⁾

⁽¹⁾ Citado por Osborn(1976).

⁽²⁾ Citado por Harter(1974-76).

⁽³⁾ Citado por Gonin e Money(1987).

⁽⁴⁾ Citado por Box e Jenkins(1976).