

Refinamentos para Testes de Hipóteses  
em Modelos de Regressão Lineares e  
Não-Lineares Heteroscedásticos

Audrey Helen Mariz de Aquino Cysneiros

Tese Apresentada ao  
Instituto de Matemática e Estatística  
da Universidade de São Paulo  
para Obtenção do Grau de  
Doutor em Estatística

Área de Concentração: **Estatística**

Orientadora: **Prof<sup>a</sup> Dr<sup>a</sup> Silvia Lopes de Paula Ferrari**

São Paulo, 16 de abril de 2004

**Refinamentos para Testes de Hipóteses  
em Modelos de Regressão Lineares e  
Não-Lineares Heteroscedásticos**

AUDREY HELEN MARIZ DE AQUINO CYSNEIROS

Este exemplar corresponde à redação final da tese de doutorado devidamente corrigida e defendida por **Audrey Helen Mariz de Aquino Cysneiros** e aprovada pela comissão julgadora.

São Paulo, 16 de abril de 2004

**Comissão julgadora:**

- Prof<sup>ª</sup> Dr<sup>ª</sup> Silvia Lopes de Paula Ferrari (Orientadora) IME/USP
- Prof. Dr. Francisco Cribari-Neto UFPE
- Prof. Dr. Gauss Moutinho Cordeiro UFRPE
- Prof. Dr. Klaus Vasconcelos UFPE
- Prof. Dr. Miguel Angel Uribe Opazo UNIOESTE/PR

Ao meu marido Francisco José

e ao meu filho Rafael

**dedico com muito amor e carinho.**

---

## Agradecimentos

- À professora **Silvia Ferrari**, a confiança, o incentivo e a excelente orientação dedicada na elaboração desta tese.
- Ao meu marido **Francisco** e ao meu filho **Rafael**, o grande apoio, amor e carinho oferecidos e também a convivência alegre, saudável, harmoniosa e feliz.
- Aos meus pais, **Lauricy** e **Washington** (in memorian), que me forneceram princípios básicos e fundamentais para minha formação moral e dedicação integral ao meu objetivo.
- A tia **Cleide**, vovô **Hermano** (in memorian), minhas irmãs **Mary** e **Walma**, minha sogra **Gilvanete**, meu sogro **Gilberto** (in memorian), meus cunhados **Aldo**, **Gilberto**, **Jorge** e **Júnior**, minhas amigas **Maria de Lourdes** e **Anunciada**, que me apoiaram e ajudaram permitindo que este passo da minha vida fosse dado.
- Aos meus sobrinhos (**Diogo**, **Igor**, **Matheus** e **Víctor**) e sobrinhas (**Catarina** e **Maria Eduarda**), o carinho e amor que sempre tive.
- Aos professores do Instituto de Matemática e Estatística da Universidade de São Paulo, que ajudaram na minha formação acadêmica.
- Aos colegas do Departamento de Estatística da Universidade Federal da Pernambuco, o apoio dado durante este curso e em especial, aos amigos **Enivaldo Rocha**, **Jacira Guiro**, **Gauss Cordeiro**, **Francisco Cribari**, **Cláudia Lima** e **Manoel Sena**, que sempre me incentivaram nesta caminhada.
- Aos amigos **Paulo de Tarso**, **Maria Paula Chicarino**, **Raquel Valle**, **Jô Hayashi**, **Érika Fukunaga**, **Silvana Meira**, **Iracema Arashiro**, **Jacqueline David**, **Regina Ishimoto**, **Lourdes Montenegro**, **Márcia Mizoi**, **Márcia Branco**, **Olga Yoshida** e **Fábio Machado**, a agradável convivência e momentos de descontração.
- À CAPES, o apoio financeiro através do PICDT (Programa Institucional de Capacitação Docente e Técnico).

## Resumo

Nesta tese, tratamos de refinamentos para testes de hipóteses em modelos de regressão lineares e não-lineares heteroscedásticos. Na primeira parte da tese, obtemos uma correção de Bartlett para um teste de heteroscedasticidade baseado em verossimilhança perfilada modificada proposta por Cox e Reid (1987) no modelo de regressão normal linear considerando que o parâmetro que define a estrutura de heteroscedasticidade é possivelmente multidimensional. Na segunda parte, estendemos esses resultados para a classe dos modelos não-lineares da família exponencial com parâmetros de dispersão que não são constantes para todas as observações. Finalmente, na terceira parte da tese, desenvolvemos novos ajustes para estatísticas da razão de verossimilhanças nesta classe de modelos, desta vez com base nos trabalhos de Skovgaard (1996, 2001). Os resultados assim obtidos são bem gerais e produzem, como caso particular, ajustes para o teste da razão de verossimilhanças de heteroscedasticidade. Desenvolvemos estudos de simulação de Monte Carlo para avaliar e comparar numericamente os desempenhos dos testes em amostras finitas.

## Abstract

In this dissertation, we deal with refinements for hypothesis tests in linear and non-linear heteroskedastic regression models. Firstly, we obtain a Bartlett correction for a test of heteroskedasticity based on the modified profile likelihood proposed by Cox and Reid (1987) in the normal linear regression model assuming that the parameter that defines the heteroskedasticity structure is possibly multidimensional. Secondly, we extend these results to the class of the exponential family non-linear models with dispersion parameters that are not constant over the observations. Thirdly, we develop new adjustments for likelihood ratio statistics in this class of models. These adjustments, which are based on Skovgaard's work (see Skovgaard, 1996, 2001), are very general and lead, as a particular case, to adjusted likelihood ratio tests of heteroskedasticity. The finite-sample performances of the various tests considered in this dissertation are evaluated and compared using Monte Carlo simulations.

---

## Conteúdo

---

Lista de Figuras	x
Lista de Tabelas	xii
<b>1 Introdução</b>	<b>1</b>
<b>2 Refinamento de um teste de heteroscedasticidade baseado em verossimilhança perfilada modificada no modelo normal linear</b>	<b>9</b>
2.1 Introdução	9
2.2 Modelo normal linear e alguns testes de heteroscedasticidade	10
2.3 Inferência assintótica de segunda ordem	14
2.4 Resultados numéricos	21
2.5 Aplicações	27
2.5.1 Exemplo 1	27
2.5.2 Exemplo 2	29
2.6 Comentários	31
<b>3 Refinamento de um teste de heteroscedasticidade baseado em verossimilhança perfilada modificada em modelos não-lineares da família exponencial</b>	<b>32</b>
3.1 Introdução	32
3.2 Definição	33
3.3 Alguns testes de heteroscedasticidade	35
3.4 Fator de correção de Bartlett	40
3.5 Resultados numéricos	43

CONTEÚDO	viii
3.6 Aplicações	49
3.6.1 Exemplo 1	49
3.6.2 Exemplo 2	50
3.7 Comentários	52
<b>4 Novos ajustes para a estatística da razão de verossimilhanças em modelos não-lineares da família exponencial</b>	<b>54</b>
4.1 Introdução	54
4.2 Ajustes para a estatística da razão de verossimilhanças	56
4.3 Modelos não-lineares da família exponencial	60
4.4 Resultados numéricos	67
4.4.1 Modelo normal linear	68
4.4.2 Modelo normal não-linear	83
4.5 Aplicações	93
4.5.1 Exemplo 1	93
4.5.2 Exemplo 2	94
4.5.3 Exemplo 3	94
4.5.4 Exemplo 4	95
4.6 Comentários	95
<b>5 Conclusões e propostas para pesquisas futuras</b>	<b>98</b>
<b>A Alguns cumulantes no modelo normal linear heteroscedástico</b>	<b>102</b>
<b>B Matriz de informação total de Fisher</b>	<b>105</b>
<b>C Multiplicação de “array”</b>	<b>108</b>
<b>D Derivadas no modelo não-linear heteroscedástico reparametrizado.</b>	<b>110</b>
<b>E Obtenção das quantidades propostas por Skovgaard</b>	<b>113</b>

CONTEÚDO

	ix
F Serviços de manutenção	128
G Taxas mensais de retorno de ações	129
H Coelhos europeus na Austrália	130
I Solução de cálcio radioativo	131
Referências	132

---

## Lista de Figuras

---

- 2.1 (a) Gráfico dos resíduos studentizados  $t_{s_i}$  contra percentis normais com envelope simulado; (b) Gráfico dos resíduos studentizados  $t_{s_i}$  contra os valores ajustados; (c) e (d) Gráficos dos resíduos studentizados  $t_{s_i}$  contra variáveis explicativas para o modelo homoscedástico - Serviços de manutenção. 28
- 2.2 (a) Gráfico dos resíduos studentizados  $t_{s_i}$  contra percentis normais com envelope simulado; (b) Gráfico de dispersão de  $y$  contra  $x$ ; (c) Gráfico dos resíduos studentizados  $t_{s_i}$  contra os valores ajustados e (d) Gráfico dos resíduos studentizados  $t_{s_i}$  contra  $x$  para o modelo homoscedástico - Taxas mensais de retorno de ações. 30
- 3.1 (a) Gráfico dos resíduos padronizados contra percentis normais com envelope simulado; (b) Gráfico de dispersão do peso das lentes dos olhos contra a idade; (c) Gráfico dos resíduos studentizados  $t_{s_i}$  contra os valores ajustados e (d) Gráfico dos resíduos studentizados  $t_{s_i}$  contra  $x$  para o modelo homoscedástico - Coelhos europeus na Austrália. 50
- 3.2 (a) Gráfico dos resíduos padronizados contra percentis normais com envelope simulado; (b) Gráfico de dispersão da quantidade absorvida de cálcio radioativo contra o tempo; (c) Gráfico dos resíduos studentizados  $t_{s_i}$  contra os valores ajustados e (d) Gráfico dos resíduos studentizados  $t_{s_i}$  contra  $x$  para o modelo homoscedástico - Solução de cálcio radioativo. 51

- |     |   |    |
|-----|---|----|
| 4.1 | Gráfico das discrepâncias relativas de quantis - Modelo normal linear com $n = 40$ , $p = 2$ e $k = 5$ .            | 82 |
| 4.2 | Gráfico das discrepâncias relativas de níveis descritivos - Modelo normal linear com $n = 40$ , $p = 2$ e $k = 5$ . | 82 |
| 4.3 | Gráfico das discrepâncias relativas de quantis - Modelo normal não-linear com $n = 40$ , $p = 2$ e $k = 3$ .        | 91 |
| 4.4 | Gráfico das discrepâncias relativas de níveis descritivos - Modelo normal linear com $n = 40$ , $p = 2$ e $k = 3$ . | 91 |

---

## Lista de Tabelas

---

2.1	Tamanho dos testes - Modelo normal linear com $n = 35$ , $p = 2$ e diversos valores para $k$ .	25
2.2	Tamanho dos testes - Modelo normal linear com $n = 35$ , $k = 5$ e diversos valores para $p$ .	25
2.3	Tamanho dos testes - Modelo normal linear com $p = 2$ , $k = 5, 8$ e diversos valores para $n$ .	26
2.4	Tamanho dos testes - Modelo de análise de variância com um fator e três níveis, $k = 3$ e $p = 2$ .	26
2.5	Poder dos testes - Modelo normal linear com $n = 30$ , $p = 2$ e $\alpha = 10\%$ .	27
3.1	Tamanho dos testes - Modelo normal não-linear com $n = 35$ , $p = 2$ , diversos valores para $k$ , covariáveis uniformes e erros normais.	47
3.2	Tamanho dos testes - Modelo normal não-linear com $n = 35$ , $k = 5$ , diversos valores para $p$ , covariáveis uniformes e erros normais.	47
3.3	Tamanho dos testes - Modelo normal não-linear com $p = 2$ , $k = 2, 5$ , covariáveis uniformes e erros normais.	48
3.4	Poder dos testes - Modelo normal não-linear com $n = 35$ , $p = 2$ , $\alpha = 10\%$ , covariáveis uniformes e erros normais.	48
4.1	Tamanhos dos testes - Modelo normal linear com $n = 40$ , $p = 1$ e diversos valores para $k$ .	75
4.2	Tamanhos dos testes - Modelo normal linear com $n = 40$ , $p = 2$ e diversos valores para $k$ .	76
4.3	Tamanhos dos testes - Modelo normal linear com $n = 40$ , $k = 5$ e diversos valores para $p$ .	77

- 4.4 Tamanho dos testes - Modelo normal linear com  $p = 2$ ,  $k = 5$  e diversos valores para  $n$ . 78
- 4.5 Média e variância da  $\chi_2^2$ ,  $LR$ ,  $LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  - Modelo normal linear com  $n = 40$ ,  $p = 2$  e  $k = 5$ . 78
- 4.6 Quantis da  $\chi_2^2$ ,  $LR$ ,  $LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  - Modelo normal linear com  $n = 40$ ,  $p = 2$  e  $k = 5$ . 78
- 4.7 Tamanhos dos testes - Modelo normal linear com  $n = 35$ ,  $p = 2$ ,  $k = 5$  e diversas distribuições de covariáveis. 79
- 4.8 Poder dos testes - Modelo normal linear com  $n = 40$ ,  $p = 2$ ,  $k = 5$  e  $\alpha = 5\%$ . 80
- 4.9 Poder dos testes - Modelo correto: normal linear com  $n = 35$ ,  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2} + \delta_3 z_{\ell 3})$ ,  $k = 5$  e  $\alpha = 5\%$ . 80
- 4.10 Poder dos testes - Modelo postulado: normal linear com  $n = 35$ ,  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2})$ ,  $k = 5$  e  $\alpha = 5\%$ . 81
- 4.11 Número de rejeições em 100.000 amostras de Monte Carlo - Modelo normal linear com  $n = 40$ ,  $p = 2$  e  $k = 5$ . 81
- 4.12 Tamanhos dos testes - Modelo normal não-linear com  $n = 40$ ,  $p = 2$  e diversos valores para  $k$ . 86
- 4.13 Tamanhos dos testes - Modelo normal não-linear com  $n = 40$ ,  $k = 5$  e diversos valores para  $p$ . 87
- 4.14 Média e variância de  $\chi_2^2$ ,  $LR$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  - Modelo normal não-linear com  $n = 40$ ,  $k = 3$  e  $p = 2$ . 88
- 4.15 Quantis de  $\chi_2^2$ ,  $LR$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  - Modelo normal não-linear com  $n = 40$ ,  $k = 3$  e  $p = 2$ . 88
- 4.16 Tamanho dos testes - Modelo normal não-linear com  $p = 2$  e  $k = 3, 5$ . 88
- 4.17 Tamanhos dos testes - Modelo normal não-linear com  $n = 35$ ,  $p = 2$ ,  $k = 5$  e diversas distribuições das covariáveis. 89
- 4.18 Poder dos testes - Modelo correto: normal não-linear com  $n = 35$ ,  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2} + \delta_3 z_{\ell 3})$ ,  $k = 3$  e  $\alpha = 5\%$ . 89
- 4.19 Poder dos testes - Modelo postulado: normal não-linear com  $n = 35$ ,  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2})$ ,  $k = 3$  e  $\alpha = 5\%$ . 90

4.20	Número de rejeições em 100.000 amostras de Monte Carlo - Modelo normal não-linear com $n = 40$ , $p = 2$ e $k = 3$ .	90
F.1	Tempo gasto no serviço ( $y$ ), em minutos, número de bebidas estocadas ( $x_1$ ) e distância percorrida ( $x_2$ ), em pés, numa amostra de 25 observações.	128
G.1	Taxas mensais de retorno de ações de mercado ( $x$ ) e da Corporação Acme Cleveland ( $y$ ), no período de janeiro de 1986 a dezembro de 1990.	129
H.1	Pesos das lentes dos olhos de coelhos europeus ( $y$ ), em miligramas e idade ( $x$ ), em dias, numa amostra de 71 observações.	130
I.1	Quantidade absorvida de cálcio radioativo ( $y$ ), em nmoles/mg, e tempo ( $x$ ), em minutos, numa amostra de 27 observações.	131

---

## CAPÍTULO 1

# Introdução

---

Pesquisas em teoria assintótica de alta ordem têm crescido a passos largos nos últimos 20 anos. Esta teoria tem como objetivo principal o refinamento de métodos inferenciais a fim de torná-los mais precisos em amostras finitas. A teoria assintótica de alta ordem utiliza aproximações baseadas em expansões de Edgeworth, ponto de sela e Laplace, enquanto que na teoria assintótica de primeira ordem os resultados decorrem de técnicas de linearização local baseadas em expansões de série de Taylor e no teorema do limite central. Nesta tese enfocaremos duas vertentes da teoria assintótica de alta ordem: a primeira corresponde à obtenção de ajustes para verossimilhanças perfiladas e a segunda visa à obtenção de ajustes para estatísticas de teste.

Na presença de parâmetros de perturbação, inferências podem ser baseadas na função de verossimilhança perfilada. Esta depende apenas do parâmetro de interesse, uma vez que é obtida da verossimilhança original em que os parâmetros de perturbação são substituídos por seus estimadores de máxima verossimilhança obtidos para cada valor fixado dos parâmetros de interesse. Embora a verossimilhança perfilada goze de algumas propriedades semelhantes às que são válidas para a verossimilhança genuína, ela pode apresentar, por exemplo, vícios na função score e na informação. A presença de um número considerável de parâmetros de perturbação também deteriora a qualidade das aproximações envolvidas nas inferências que se baseiam em resultados assintóticos. Desta forma, torna-se importante obter ajustes para a verossimilhança perfilada que reduzam estes problemas. Esta é a primeira vertente da teoria assintótica de alta ordem a ser abordada nesta tese.

Vários ajustes para a função de verossimilhança perfilada e para estatísticas da razão de verossimilhanças derivadas de tal função têm sido propostos recente-

mente. Como exemplos de trabalhos recentes neste sentido, podemos mencionar Barndorff-Nielsen (1983, 1994), Cox e Reid (1987), DiCiccio e Stern (1994), Simonoff e Tsai (1994), Ferrari e Cribari-Neto (2002), Stern (1997), Wei, Shi, Fung e Hu (1998), Severini (1998), Bellio e Brazzale (2000) e Ferrari, Lucambio e Cribari-Neto (2003).

Barndorff-Nielsen (1983) definiu uma versão modificada da função de verossimilhança perfilada que é invariante sob reparametrizações da forma  $(\nu, \psi) \longrightarrow (\nu, \zeta(\nu, \psi))$ , sendo  $\nu$  o vetor de parâmetros de interesse,  $\psi$  o vetor de parâmetros de perturbação e  $\zeta$  uma função de  $\nu$  e  $\psi$ . Esta versão é baseada na fórmula  $p^*$ , uma aproximação para a densidade condicional do estimador de máxima verossimilhança do vetor de parâmetros do modelo, dada uma estatística ancilar. Pode ser difícil obter a função de verossimilhança perfilada modificada de Barndorff-Nielsen em problemas particulares, uma vez que esta depende da determinação de uma estatística ancilar e envolve derivadas com respeito ao espaço amostral. Barndorff-Nielsen (1994) propõe duas aproximações para a verossimilhança perfilada modificada para o caso em que o parâmetro de interesse é escalar que não requerem a determinação de uma estatística ancilar. Severini (1998) derivou uma aproximação para a verossimilhança perfilada modificada de Barndorff-Nielsen (1983, 1994) que não exige a determinação de uma estatística ancilar e não envolve derivada com respeito ao espaço amostral.

Cox e Reid (1987) definiram uma versão modificada da verossimilhança perfilada explorando as conseqüências da ortogonalidade entre os parâmetros de interesse e os de perturbação em que um termo de ajuste é incorporado à verossimilhança perfilada anteriormente à estimação. Esta é uma versão penalizada do logaritmo da função de verossimilhança perfilada em que a penalização leva em conta a informação sobre o parâmetro de perturbação para cada valor do parâmetro de interesse. Esta pseudo-verossimilhança requer uma parametrização que só é garantida existir quando o parâmetro de interesse é escalar. Embora qualquer função dos parâmetros de perturbação seja também ortogonal ao parâmetro de interesse, a verossimilhança perfilada modificada não é invariante sob reparametrizações da

forma  $(\nu, \psi) \longrightarrow (\nu, \zeta(\nu, \psi))$ . A modificação de Cox e Reid para a verossimilhança perfilada atenua o efeito do número de parâmetros de perturbação e é baseada em expansão de ponto de sela. Wei, Shi, Fung e Hu (1998) utilizaram o ajuste de Cox e Reid (1987) para desenvolver o teste da razão de verossimilhanças perfiladas modificadas e um teste escore ajustado em modelos não-lineares da família exponencial. Vale salientar que a estatística da razão de verossimilhanças perfiladas modificadas foi derivada somente para os modelos não-lineares normal e normal inverso.

A segunda vertente da teoria assintótica de alta ordem a ser abordada nesta tese visa a obter ajustes para estatísticas de teste. Mais especificamente, enfocaremos a estatística da razão de verossimilhanças ( $LR$ ). Em problemas regulares, esta estatística tem, sob a hipótese nula ( $\mathcal{H}_0$ ), distribuição  $\chi_p^2$  aproximadamente, em grandes amostras, onde  $p$  é o número de restrições impostas por  $\mathcal{H}_0$  e o erro dessa aproximação é de ordem  $n^{-1}$ . A idéia de modificar a estatística por um fator de correção, visando a produzir uma estatística modificada com o primeiro momento igual ao da distribuição  $\chi^2$  de referência é devida a Bartlett (1937). Sob a hipótese nula, a média da estatística  $LR$  pode ser expandida, em problemas regulares, como  $E(LR) = p\{1 + c/p + O(n^{-2})\}$ , onde  $n$  é o tamanho da amostra. Assim, a média da estatística modificada  $LR^* = LR/(1 + c/p)$  é dada por  $p + O(n^{-2})$  e é “mais próxima” daquela da distribuição  $\chi_p^2$  do que a média de  $LR$ . O fator de correção  $1/(1 + c/p)$  tornou-se conhecido como correção de Bartlett.

Lawley (1956) desenvolveu um método geral de obtenção de  $c$  em termos de cumulantes do logaritmo da função de verossimilhança, que são simplesmente valores esperados de produtos de derivadas do logaritmo da função de verossimilhança. Além disso, através de uma demonstração extremamente complexa, concluiu que a estatística da razão de verossimilhanças modificada tem todos os momentos concordando com os respectivos da distribuição  $\chi^2$  de referência, ignorando os termos de ordem  $n^{-2}$ . Deve-se destacar que o fator de correção de Bartlett não depende do valor da estatística da razão de verossimilhanças, mas pode depender de parâmetros desconhecidos e, neste caso, estes devem ser substituídos por suas respectivas estimativas de máxima verossimilhança sob a hipótese nula. Barndorff-Nielsen e

Hall (1988) mostraram que a distribuição nula de  $LR^*$ , substituindo ou não  $c$  por seu estimador de máxima verossimilhança sob  $\mathcal{H}_0$ , coincide com a distribuição  $\chi_p^2$  com erro de ordem  $n^{-2}$ . Barndorff-Nielsen e Cox (1984) apresentam um método indireto de cálculo das correções de Bartlett. Barndorff-Nielsen e Blaesild (1986) desenvolvem um método alternativo para obtenção de correção de Bartlett, enquanto que McCullagh e Cox (1986) apresentam fórmulas alternativas para obter as correções.

DiCiccio e Stern (1994) mostraram que é possível obter correção de Bartlett para a estatística da razão de verossimilhanças proveniente de uma verossimilhança perfilada modificada de tal forma que a aproximação por  $\chi_p^2$  para a distribuição da estatística corrigida seja de ordem  $n^{-2}$ . Note-se que as duas vertentes da teoria assintótica de alta ordem citadas anteriormente se unem aqui: inicialmente a função de verossimilhança perfilada é modificada e uma estatística da razão de verossimilhanças é obtida; posteriormente, esta estatística é corrigida. Espera-se que a modificação na verossimilhança perfilada atenuar o efeito dos parâmetros de perturbação e a correção à estatística melhore a aproximação de sua distribuição pela distribuição assintótica.

Usando os resultados de DiCiccio e Stern (1994), Ferrari e Cribari-Neto (2002) corrigiram o teste de heteroscedasticidade em modelos de regressão normais lineares proposto por Simonoff e Tsai (1994) e obtiveram um teste da razão de verossimilhanças perfiladas modificadas corrigido, que é preciso até segunda ordem, para o caso em que o parâmetro de interesse é unidimensional. Stern (1997) derivou um ajuste aditivo ao logaritmo da função de verossimilhança perfilada que reduz os vieses da função score e da matriz de informação obtidas desta função. Ferrari, Lucambio e Cribari-Neto (2003) obtiveram a expressão correta para o ajuste de Stern (1997), a qual continha um erro, e derivaram uma correção de Bartlett para a estatística da razão de verossimilhanças perfiladas ajustada na classe dos modelos lineares generalizados.

Atenção tem sido dada também à obtenção de ajustes para a estatística da razão de verossimilhanças sinalizada ( $R$ ) que pode ser utilizada quando o parâmetro de

interesse é unidimensional. O ajuste de Barndorff-Nielsen (1986, 1991) envolve derivadas com respeito ao espaço amostral, requer a especificação de uma estatística ancilar e pode, portanto, ser muito difícil de se obter em problemas particulares. Skovgaard (1996) obteve uma aproximação para o ajuste de Barndorff-Nielsen que evita tal dificuldade. Uma extensão desses resultados para englobar o caso da estatística da razão de verossimilhanças na situação em que o parâmetro de interesse é multidimensional foi obtida por Skovgaard (2001) e será objeto de estudo desta tese.

Os resultados de Skovgaard (1996) foram utilizados por Lyon e Peters (2000) para obter um teste da razão de verossimilhanças ajustado para testar um dos componentes da variância em modelos lineares mistos. Esses resultados foram utilizados também por Bellio e Brazzale (2003) para desenvolver e implementar computacionalmente no software *S-Plus* (Becker, Chambers e Wilk, 1988 e Chambers e Hastie, 1992) testes da razão de verossimilhanças ajustados em modelos de regressão não-lineares heteroscedásticos. No entanto, os métodos somente estão implementados para o caso em que há uma única covariada (com réplicas). Na prática, quando há heteroscedasticidade, as variâncias dos erros freqüentemente dependem de mais de uma covariada, então é importante estender os resultados obtidos na literatura para abordar o caso em que o parâmetro que define o comportamento de heteroscedasticidade é multidimensional. Yi, Wu e Liu (2002) implementaram rotinas computacionais em *S-Plus* com base nos resultados propostos por Skovgaard (1996) para testes sobre os parâmetros de regressão em modelos de regressão normal linear e não-linear. Novamente, esta implementação somente está desenvolvida para o caso unidimensional.

Esta tese tem três objetivos principais. O primeiro é obter um fator de correção de Bartlett para o teste de heteroscedasticidade de Simonoff e Tsai (1994) no modelo normal linear. Este teste é baseado em uma verossimilhança perfilada modificada obtida segundo o enfoque de Cox e Reid (1987). O teste proposto tipicamente exhibe inferências mais confiáveis em amostras finitas, mas a estatística do teste ainda é distribuída como  $\chi_p^2$  sob a hipótese nula com erro de ordem  $n^{-1}$ . Já para a estatística

corrigida este erro é de ordem  $n^{-2}$ . O segundo objetivo é estender os resultados de Simonoff e Tsai (1994) para os modelos não-lineares da família exponencial e obter a correção de Bartlett para o teste de heteroscedasticidade baseado na verossimilhança perfilada modificada de Cox e Reid (1987). Deve ser mencionado que heteroscedasticidade nesse contexto significa que o parâmetro de dispersão (não necessariamente a variância) não é o mesmo para todas as observações. Como já mencionado, a modificação de Cox e Reid (1987) requer que os parâmetros de interesse sejam ortogonais aos de perturbação. No entanto, no modelo gama, um dos membros da classe de modelos considerada, esta ortogonalização não foi possível de ser obtida. O terceiro objetivo da tese é obter ajustes para testes da razão de verossimilhanças nos modelos não-lineares da família exponencial baseados na proposta de Skovgaard (2001), a qual não requer ortogonalidade de parâmetros. Uma aplicação particular dos resultados refere-se ao ajuste para o teste de heteroscedasticidade.

Apresentamos no Capítulo 2 uma breve introdução ao modelo de regressão normal linear e alguns testes de heteroscedasticidade. Desenvolvemos uma expressão simples para o fator de correção de Bartlett que é bastante geral e que pode ser usada para aperfeiçoar o teste da razão de verossimilhanças perfiladas modificadas em diferentes classes de modelos. Este resultado cobre a situação na qual o parâmetro de interesse é multidimensional, generalizando assim os resultados apresentados por Ferrari e Cribari-Neto (2002). Comportamentos em amostras finitas dos diferentes testes são apresentados em relação a tamanho e poder. Analisamos a influência do número de parâmetros de perturbação e de parâmetros de interesse no desempenho dos testes. Apresentamos também aplicações a dados reais e algumas considerações finais são apresentadas.

No Capítulo 3, apresentamos uma breve introdução aos modelos não-lineares da família exponencial que são extensões naturais dos modelos lineares generalizados e dos modelos de regressão normais não-lineares. Apresentamos também alguns testes de heteroscedasticidade, incluindo um teste baseado em verossimilhança perfilada modificada. Desenvolvemos um fator de correção de Bartlett que pode ser

aplicado à estatística da razão de verossimilhanças perfiladas modificadas nesta classe de modelos. Este resultado estende, portanto, o resultado desenvolvido no Capítulo 2, que é restrito ao modelo normal linear. Desenvolvemos um ajuste para o teste proposto por Wei, Shi, Fung e Hu (1998), obtendo assim um teste da razão de verossimilhanças perfiladas modificadas corrigido que é preciso até segunda ordem. Notamos que o ajuste de Cox e Reid (1987) não pode ser aplicado a todos os modelos não-lineares da família exponencial, por exemplo, o modelo gama. Resultados numéricos sobre o comportamento em amostras finitas dos diferentes testes são apresentados. Aplicações a dados reais são também apresentadas e no final do capítulo são feitos alguns comentários.

O Capítulo 4 trata da obtenção de ajustes para a estatística da razão de verossimilhanças com base na proposta de Skovgaard (2001) que podem ser utilizados, sem exceção, sob todos os modelos não-lineares da família exponencial. Estes ajustes são mais gerais, pois são válidos para testar qualquer hipótese de interesse e, em particular, serão feitos ajustes para o teste de heteroscedasticidade. Estes novos ajustes apresentam uma vantagem em relação ao ajuste de Cox e Reid (1987) por serem invariantes segundo reparametrizações da forma  $(\nu, \psi) \longrightarrow (\nu, \zeta(\nu, \psi))$ , sendo  $\nu$  o vetor de parâmetros de interesse,  $\psi$  o vetor de parâmetros de perturbação e  $\zeta$  uma função de  $\nu$  e  $\psi$ . Outra vantagem é a aplicação direta à estatística da razão de verossimilhanças, enquanto que o ajuste de Cox e Reid, ao contrário, é aplicado à função de verossimilhança perfilada, a qual deve ser maximizada sob o modelo restrito (sob  $\mathcal{H}_0$ ) e irrestrito para, posteriormente, obter-se a estatística ajustada. Quando os estimadores de máxima verossimilhanças restritos dos parâmetros de perturbação não têm forma fechada estas maximizações podem ser computacionalmente mais complexas. Outra vantagem adicional é que não é necessária a ortogonalização entre os parâmetros de interesse e os de perturbação. Estes novos ajustes apresentam uma vantagem em relação ao fator de correção de Bartlett, uma vez que são mais simples de serem obtidos pois requerem a obtenção de cumulantes do logaritmo da função de verossimilhanças até segunda ordem, que são simplesmente valores esperados de produtos de derivadas do logaritmo da função

de verossimilhança. Já, para o cálculo do fator de correção de Bartlett, é necessária a obtenção de cumulantes do logaritmo da função de verossimilhanças até quarta ordem e derivadas destes cumulantes.

Poucos resultados são encontrados na literatura acerca do desempenho dos ajustes de Skovgaard (2001) no caso em que o parâmetro de interesse é multidimensional. Nosso trabalho veio a preencher parcialmente esta lacuna. Resultados numéricos sobre o comportamento em amostras finitas de diferentes testes de heteroscedasticidade nos modelos de regressão linear e nos modelos não-lineares da família exponencial são apresentados em relação a tamanho e poder. Consideramos também aspectos de robustez dos testes apresentados. Aplicações a dados reais são apresentadas e considerações são feitas no final do capítulo.

No Capítulo 5, conclusões são apresentadas e são dadas sugestões de pesquisa que podem ser desenvolvidas a partir das idéias desta tese.

Vale salientar aqui que esta tese foi escrita de tal forma que os Capítulos 2, 3 e 4 sejam auto-suficientes, significando que alguns resultados básicos e notações são apresentados mais de uma vez, porém, deste modo, cada capítulo pode ser lido separadamente, em qualquer ordem.

Todos os procedimentos inferenciais derivados dos resultados teóricos obtidos na tese são avaliados e comparados através de simulações de Monte Carlo. Todas as simulações apresentadas foram desenvolvidas a partir de programas construídos usando a linguagem de programação matricial  $\text{Ox}$  (Doornik, 2001). As maximizações não-lineares necessárias para o cálculo de estimativas de máxima verossimilhança foram feitas utilizando-se o algoritmo quasi-Newton BFGS (Nocedal e Wright, 1999, Capítulo 8) e o algoritmo SQP (Nocedal e Wright, 1999, Capítulo 18, similar ao Algoritmo 18.7), disponíveis em funções pré-definidas da linguagem  $\text{Ox}$ .

---

## CAPÍTULO 2

# Refinamento de um teste de heteroscedasticidade baseado em verossimilhança perfilada modificada no modelo normal linear

---

### 2.1 Introdução

Modelos de regressão lineares são frequentemente usados para modelar o comportamento de uma variável de interesse condicionada a um conjunto de variáveis explicativas. Em aplicações práticas, é comum para estes modelos incluir uma estrutura de heteroscedasticidade, indicando assim que as variâncias condicionais não são constantes ao longo das observações. Quando tais variâncias não são constantes ao longo das observações, as estratégias de modelagem são diferentes, portanto, é importante testar a existência de heteroscedasticidade nos dados.

Os testes mais comumente usados para testar heteroscedasticidade são baseados em valores críticos obtidos de uma distribuição limite conhecida. As aproximações assintóticas usadas podem não ser precisas quando o tamanho de amostra não é grande. Em particular, o teste da razão de verossimilhanças para igualdade de variâncias em modelos de regressão lineares tende a ser liberal, apresentando taxas de rejeição bem maiores do que o nível nominal considerado em amostras de tamanho pequeno ou moderado. Então, é importante desenvolver estratégias de inferência que sejam mais precisas e exibam comportamento superior em amostras finitas.

Cordeiro (1993) propôs um fator de correção de Bartlett no teste da razão de verossimilhanças para testar a heteroscedasticidade, em que o teste estatístico é modificado usando um fator de correção que tipicamente reduz a distorção do tamanho em amostras finitas. A estatística da razão de verossimilhanças original é distribuída sob a hipótese nula como  $\chi_p^2$  com um erro de ordem  $n^{-1}$ ,

enquanto que a estatística corrigida segue a mesma distribuição limite com um erro de ordem  $n^{-2}$ .

Simonoff e Tsai (1994) utilizaram a abordagem de verossimilhança perfilada modificada proposta por Cox e Reid (1987) para desenvolver um teste de heteroscedasticidade visando a reduzir o efeito de parâmetros de perturbação na inferência resultante. O teste proposto tipicamente exhibe inferências mais confiáveis em amostras finitas, mas, a estatística do teste é ainda distribuída como  $\chi_p^2$  sob a hipótese nula com erro de ordem  $n^{-1}$ .

Ferrari e Cribari-Neto (2002) corrigiram o teste proposto por Simonoff e Tsai (1994) e obtiveram um teste corrigido da razão de verossimilhanças perfiladas modificada que é preciso até segunda ordem. Seus resultados, porém, somente são válidos em situações em que o parâmetro que define o comportamento da heteroscedasticidade é escalar. Na prática, quando a heteroscedasticidade ocorre, as variâncias freqüentemente dependem de mais de uma covariada, e é portanto importante estender os resultados de Ferrari e Cribari-Neto (2002) para abordar o caso em que tal parâmetro é multidimensional.

Na Seção 2.2, apresentamos uma breve introdução do modelo de regressão linear e alguns testes de heteroscedasticidade acima mencionados. Na Seção 2.3, desenvolvemos o principal resultado do presente capítulo, isto é, uma correção de Bartlett que pode ser aplicada a estatísticas da razão de verossimilhanças perfiladas modificadas. Evidências numéricas sobre o comportamento em amostras finitas dos diferentes testes são apresentadas na Seção 2.4. A evidência numérica favorece o teste proposto neste capítulo. Na Seção 2.5, apresentamos aplicações a dados reais e alguns comentários são feitos na Seção 2.6.

## 2.2 Modelo normal linear e alguns testes de heteroscedasticidade

Consideramos o modelo normal linear

$$y = X\beta + u, \tag{2.1}$$

em que  $y$  é um vetor de dimensão  $n \times 1$  de variáveis aleatórias,  $X$  é uma matriz modelo de dimensão  $n \times k$  de regressores conhecidos e fixados, de posto completo,  $\beta$  é um vetor de dimensão  $k \times 1$  de parâmetros desconhecidos e  $u$  é um vetor de dimensão  $n \times 1$  de erros não observáveis. Para efeito de estimação por máxima verossimilhança e testes de hipóteses, assumimos que  $u \sim N(0, \sigma^2 M)$  em que  $M$  é uma matriz diagonal de dimensão  $n \times n$  cujos elementos são dados por  $m_\ell = m(z_\ell, \delta) > 0$ ,  $\ell = 1, \dots, n$ ,  $z_\ell^\top$  sendo a  $\ell$ -ésima linha da matriz  $Z$  de dimensão  $n \times p$  de covariadas usadas para modelar as flutuações na variância,  $\sigma^2$  é uma constante desconhecida finita e estritamente positiva e  $\delta$  é um vetor de dimensão  $p \times 1$  de parâmetros desconhecidos.

Nosso interesse é testar a hipótese nula  $\mathcal{H}_0 : \delta = \delta_0$  (homoscedasticidade<sup>(1)</sup>) contra a hipótese alternativa bilateral  $\mathcal{H}_1 : \delta \neq \delta_0$ , em que  $\delta_0$  é um vetor de dimensão  $p \times 1$  de constantes especificado de tal forma que  $m(z_\ell, \delta_0) = 1$  para  $\ell = 1, \dots, n$ . Então, o número de parâmetros de interesse é  $p$  e o número de parâmetros de perturbação é  $k + 1$ .

O logaritmo da função de verossimilhança do modelo (2.1) é dado por

$$L(\delta, \beta, \sigma^2; y) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \sum_{\ell=1}^n \log m(z_\ell, \delta) - \frac{1}{2\sigma^2} (y - X\beta)^\top M^{-1} (y - X\beta) + \text{constante.}$$

Quando o modelo a ser estimado envolve parâmetros de perturbação, um procedimento comumente utilizado é a eliminação de tais parâmetros, substituindo-os na função de verossimilhança por suas respectivas estimativas de máxima verossimilhança para valores fixados dos parâmetros de interesse. A função resultante, chamada de função de verossimilhança perfilada, depende, portanto, somente dos parâmetros de interesse. O logaritmo desta função é dado por

$$L_p(\delta; y) = L(\delta, \hat{\beta}_\delta, \hat{\sigma}_\delta^2; y). \quad (2.2)$$

<sup>(1)</sup> Homoscedasticidade. Esta palavra tem origem em duas raízes: *homós*, que significa “semelhante” ou “igual”, e *skedánnummi*, que quer dizer “espalhar” ou “dispersar”.

Aqui,

$$\widehat{\beta}_\delta = (X^\top M^{-1} X)^{-1} X^\top M^{-1} y \quad \text{e} \quad \widehat{\sigma}_\delta^2 = \frac{1}{n} (y - X \widehat{\beta}_\delta)^\top M^{-1} (y - X \widehat{\beta}_\delta).$$

O estimador de máxima verossimilhança (EMV) de  $\delta$  pode ser obtido maximizando o logaritmo da função de verossimilhança perfilada dado em (2.2), que, neste caso, reduz-se a

$$L_p(\delta; y) = -\frac{n}{2} \log(\widehat{\sigma}_\delta^2) - \frac{1}{2} \sum_{\ell=1}^n \log m(z_\ell, \delta) + \text{constante}.$$

A estatística da razão de verossimilhanças para o teste de  $\mathcal{H}_0$  pode ser escrita como

$$LR = -2\{L_p(\delta_0; y) - L_p(\widehat{\delta}; y)\},$$

em que  $\widehat{\delta}$  é o EMV de  $\delta$ . Para grandes amostras, a distribuição da estatística da razão de verossimilhanças,  $LR$ , pode ser aproximada pela distribuição  $\chi_p^2$ . No entanto, a aproximação resultante pode ser bastante pobre para amostras pequenas, conduzindo assim a taxas de rejeição bastante distorcidas.

Explorando as consequências da ortogonalidade entre os parâmetros de interesse e os de perturbação, Cox e Reid (1987) definiram uma versão modificada da verossimilhança perfilada, em que um termo de ajuste é incorporado à verossimilhança perfilada anteriormente à estimação. A verossimilhança perfilada modificada é uma aproximação para a função densidade condicional das observações dado o estimador de máxima verossimilhança do parâmetro de perturbação. Esta versão apresenta algumas desvantagens: (i) requer uma parametrização ortogonal que somente é garantida existir quando o parâmetro de interesse é escalar; (ii) não é invariante sob reparametrização da forma  $(\nu, \psi) \rightarrow (\nu, \zeta(\nu, \psi))$ .

Simonoff e Tsai (1994) utilizaram a abordagem proposta por Cox e Reid (1987) para definir uma estatística da razão de verossimilhanças perfiladas modificadas. A matriz de informação total de Fisher para o vetor  $\theta = (\delta^\top, \beta^\top, \sigma^2)^\top$  é dada por

$$I = -E \left( \frac{\partial^2 L}{\partial \theta \partial \theta^\top} \right) = \begin{pmatrix} \frac{1}{2} \Psi^\top \Psi & 0 & \frac{1}{2\sigma^2} \Psi^\top \iota \\ 0 & X^\top \Phi X & 0 \\ \frac{1}{2\sigma^2} \iota^\top \Psi & 0 & \frac{n}{2\sigma^4} \end{pmatrix}, \quad (2.3)$$

sendo  $\Psi$  uma matriz  $n \times p$  com  $\ell$ -ésima linha dada por  $\partial \log m(z_\ell, \delta) / \partial \delta^\top$ ,  $\Phi$  uma matriz diagonal de dimensão  $n \times n$  cujo  $\ell$ -ésimo elemento é dado por  $\phi_\ell = 1/(\sigma^2 m_\ell)$ ,  $\iota$  é um vetor de uns de dimensão  $n$  e  $0$  denotando matrizes ou vetores nulos de dimensões apropriadas. A principal restrição a ser observada é que  $\delta$  deve ser ortogonal aos parâmetros restantes. Observe que  $(\delta^\top, \sigma^2)^\top$  é ortogonal a  $\beta$  e, então, para que a transformação  $(\delta^\top, \beta^\top, \sigma^2)^\top \longrightarrow (\delta^\top, \beta^\top \gamma)^\top$  seja tal que  $\delta$  e  $\gamma$  sejam ortogonais, é necessário e suficiente que  $E(-\partial^2 L / \partial \delta_a \partial \gamma) = 0$ ,  $a = 1, \dots, p$ . Seguindo Cox e Reid (1987, equação (4)), chegamos à transformação desejada através do sistema de equações diferenciais

$$\frac{n}{2\sigma^4} \frac{\partial \sigma^2}{\partial \delta^a} = -\frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{\partial m_\ell}{\partial \delta^a} \frac{1}{m_\ell}, \quad a = 1, \dots, p,$$

que tem como solução

$$\sigma^2 = \frac{\gamma}{(\prod_{\ell=1}^n m_\ell)^{1/n}}.$$

O logaritmo da função de verossimilhança para o modelo reparametrizado é dado por

$$L^* = L^*(\delta, \beta, \gamma; y) = -\frac{n}{2} \log \gamma - \frac{1}{2\gamma} (y - X\beta)^\top Q (y - X\beta) + \text{constante},$$

em que  $Q = \text{diag}\{q_1, \dots, q_n\}$ , com

$$q_\ell = q_\ell(\delta) = \frac{(\prod_{s=1}^n m_s)^{1/n}}{m_\ell}. \quad (2.4)$$

O logaritmo da função de verossimilhança perfilada correspondente para  $\delta$  é

$$L_p^* = L_p^*(\delta; y) = -\frac{n}{2} \log \hat{\gamma}_\delta + \text{constante},$$

em que  $\hat{\gamma}_\delta = \hat{\sigma}_\delta^2 (\prod_{\ell=1}^n m_\ell)^{1/n}$ .

O logaritmo da função de verossimilhança perfilada modificada (Simonoff e Tsai, 1994) correspondente para  $\delta$  é

$$L_{mp}^*(\delta; y) = L_p^*(\delta; y) - \frac{1}{2} \log[\det\{j^*(\delta; \hat{\beta}_\delta, \hat{\gamma}_\delta)\}],$$

em que  $j^*$  é o bloco da matriz de informação observada para os parâmetros de

perturbação  $(\beta^\top, \gamma)^\top$  no modelo reparametrizado avaliado em  $(\delta^\top, \widehat{\beta}_\delta^\top, \widehat{\gamma}_\delta^\top)^\top$ . É fácil mostrar que

$$L_{mp}^*(\delta; y) = L_p^*(\delta; y) - \frac{1}{2} \log \left\{ \frac{n}{2\widehat{\gamma}_\delta^{k+2}} \det(X_m^\top X_m) \right\},$$

em que  $X_m = Q^{1/2}X$ ,  $Q = Q(\delta)$  sendo uma matriz diagonal de dimensão  $n \times n$  cujo  $\ell$ -ésimo elemento da diagonal é dado por  $\{\prod_{j=1}^n m(z_j, \delta)\}^{1/n}/m(z_\ell, \delta)$ .

A estatística da razão de verossimilhanças baseada na função de verossimilhança perfilada modificada para o teste de  $\mathcal{H}_0$  versus  $\mathcal{H}_1$  é dada por

$$LR_m = -2\{L_{mp}^*(\delta_0; y) - L_{mp}^*(\tilde{\delta}; y)\}, \quad (2.5)$$

em que  $\tilde{\delta}$  é valor de  $\delta$  que maximiza a função  $L_{mp}^*(y; \delta)$ . Note que  $LR_m$  envolve a maximização da função de verossimilhança perfilada modificada. É fácil mostrar que a estatística resultante pode ser escrita como

$$LR_m = \frac{n-k-2}{n} LR + \log \left\{ \frac{\det(X^\top X)}{\det(\tilde{X}_m^\top \tilde{X}_m)} \right\},$$

em que  $\tilde{X}_m = \tilde{Q}^{1/2}X$  e  $\tilde{Q} = Q(\tilde{\delta})$ . Simonoff e Tsai (1994) propuseram uma estatística alternativa que evita tal maximização substituindo  $\tilde{\delta}$  por  $\widehat{\delta}$  na expressão de  $LR_m$  obtendo-se assim a mesma estatística do teste dada acima, mas com  $\tilde{X}_m$  substituído por  $\widehat{X}_m$ , em que  $\widehat{X}_m = \widehat{Q}^{1/2}X$  com  $\widehat{Q} = Q(\widehat{\delta})$ .

### 2.3 Inferência assintótica de segunda ordem

Para melhorar a qualidade da aproximação da distribuição da estatística  $LR$  pela distribuição qui-quadrado, pode ser utilizado um fator de correção de Bartlett  $(1 + c/p)^{-1}$ , que é multiplicado à estatística  $LR$  produzindo a estatística da razão de verossimilhanças corrigida  $LR^*$ , dada por

$$LR^* = \frac{LR}{1 + c/p}, \quad (2.6)$$

em que  $c$  é de ordem  $n^{-1}$  e tal que  $E(LR) = p\{(1 + c/p) + O(n^{-3/2})\}$ . Em problemas regulares, a estatística  $LR^*$  tem distribuição assintótica  $\chi_p^2$  com erro de ordem  $n^{-2}$ ;

ver Barndorff-Nielsen e Hall (1988). Assim, enquanto  $P(LR > x_\alpha) = \alpha + O(n^{-1})$ , tem-se  $P(LR^* > x_\alpha) = \alpha + O(n^{-2})$ , em que  $x_\alpha$  é o quantil  $(1 - \alpha)$  da distribuição  $\chi_p^2$ . Ou seja, o erro da aproximação por  $\chi_p^2$  para a distribuição nula de  $LR$  é de ordem  $n^{-1}$ , enquanto que o erro desta aproximação para a distribuição de  $LR^*$  é reduzido à ordem  $n^{-2}$ .

Um método geral de obtenção da constante  $c$  foi desenvolvido por Lawley (1956) e uma revisão de correção de Bartlett para a estatística da razão de verossimilhanças em modelos de regressão pode ser encontrada em Cribari-Neto e Cordeiro (1996).

Cordeiro (1993) obteve o fator de correção de Bartlett para a estatística da razão de verossimilhanças do teste de  $\mathcal{H}_0$  contra  $\mathcal{H}_1$  no modelo heteroscedástico multiplicativo, isto é,  $w(z_\ell, \delta) = \exp\{z_\ell^\top \delta\}$ , com  $\delta = 0$  correspondendo ao modelo homoscedástico. A constante  $c$  é dada por

$$c = h(X, R) - \frac{3(k+1)^2 - 1}{6n},$$

em que

$$\begin{aligned} h(X, R) &= -\frac{1}{2} \text{tr}(A_d^2) + \frac{1}{3} \iota^\top A^{(3)} \iota + 12 \iota^\top A_d A A_d \iota + \text{tr}(A_d B_d) - \iota^\top (B^{(2)} \odot A) \iota \\ &+ \frac{1}{2} \iota^\top B_d A B_d \iota + \iota^\top A_d A B_d \iota, \end{aligned}$$

com  $A = R(R^\top R)^{-1}R^\top$  sendo  $R = [\iota Z]$ ,  $A_d = \text{diag}(a_{11}, \dots, a_{nn})$ ,  $B = X(X'X)^{-1}X^\top$ ,  $B_d = \text{diag}(b_{11}, \dots, b_{nn})$ . Denotamos  $B^{(2)} = B \odot B$ ,  $A^{(3)} = A^{(2)} \odot A$ , em que  $\odot$  é o produto Hadamard (produto direto); ver Rao (1973, p. 30).

A estatística da razão de verossimilhanças baseada na verossimilhança perfilada modificada,  $LR_m$ , citada na Seção 2.2, tem, em problemas regulares, distribuição  $\chi_p^2$  assintoticamente, sob a hipótese nula. Mesmo com a modificação, a aproximação pela distribuição  $\chi_p^2$  tem erro de ordem  $n^{-1}$ . No entanto, DiCiccio e Stern (1994) mostraram que é possível obter uma correção de Bartlett para a estatística da razão de verossimilhanças proveniente de uma verossimilhança perfilada ajustada. O objetivo desta correção é a redução na distorção de tamanho do teste em amostras finitas através de uma aceleração da taxa de convergência do tamanho verdadeiro

para o tamanho nominal (assintótico). De fato, o erro da aproximação por  $\chi_p^2$  passa a ser de ordem  $n^{-2}$ . A seguir, encontraremos um fator de correção de Bartlett  $c_m$  que define uma estatística da razão de verossimilhanças perfiladas modificadas corrigida dada por

$$LR_m^* = \frac{LR_m}{1 + c_m/p} \quad (2.7)$$

de tal forma que, em problemas regulares,  $P(LR_m^* > x_\alpha) = \alpha + O(n^{-2})$ .

Para evitar ambiguidade de índices entre os componentes do vetor de parâmetros de interesse e os componentes do vetor de parâmetros de perturbação, introduzimos a seguinte notação: os índices  $a, b, \dots$  variam em  $1, \dots, p$ , os índices  $i, j, \dots$  variam em  $p + 1, \dots, p + k$  e os índices  $r, s, t, \dots$  variam em  $1, \dots, p + k + 1$ . Seja  $\theta = (\delta^\top, \beta^\top, \gamma)^\top$ , em que  $\theta^r$  é o  $r$ -ésimo elemento de  $\theta$ . Adotamos a notação tensorial para os cumulantes de derivadas da função de log-verossimilhança:  $\lambda_{rs} = E(\partial^2 L / \partial \theta^r \partial \theta^s)$ ,  $\lambda_{rst} = E(\partial^3 L / \partial \theta^r \partial \theta^s \partial \theta^t)$ , etc. Definimos as derivadas dos cumulantes por  $(\lambda_{rs})_t = \partial \lambda_{rs} / \partial \theta^t$ ,  $(\lambda_{rs})_{tu} = \partial^2 \lambda_{rs} / \partial \theta^t \partial \theta^u$ , etc. A matriz de informação de Fisher tem elementos  $-\lambda_{rs}$ , sendo  $-\lambda^{rs}$  os correspondentes elementos de sua inversa. Adicionalmente, definimos  $\tau^{rs} = \lambda^{ra} \lambda^{sb} \sigma_{ab}$ , em que  $(\sigma_{ab})$  é a inversa da matriz  $(\lambda^{ab})$  de dimensão  $p \times p$ . Note que  $(\lambda^{ab})$  é o bloco da inversa da matriz de informação que corresponde ao vetor de parâmetros de interesse. Quando os índices aparecem repetidos como subscritos e sobrescritos, uma soma sobre toda a amplitude desses índices está subentendida. Definimos ainda  $\nu^{rs} = \lambda^{rs} - \tau^{rs}$ . Observe que as entradas da matriz  $\nu^{rs}$  de dimensão  $(p + k + 1) \times (p + k + 1)$  são todas zero exceto para a submatriz inferior direita de dimensão  $(k + 1) \times (k + 1)$ , que é a matriz de informação de Fisher para os parâmetros de perturbação  $(\beta^\top, \gamma)^\top$  permanecendo  $\delta$  fixo.

DiCiccio e Stern (1994, p. 404, equação (25)) definiram uma expressão geral para  $c_m$ , que pode ser escrita como

$$\begin{aligned} c_m = & \frac{1}{4} \tau^{ru} \tau^{st} \lambda_{rstu} - \lambda^{ru} \tau^{st} (\lambda_{rst})_u + (\lambda^{ru} \lambda^{st} - \nu^{ru} \nu^{st}) (\lambda_{rs})_{tu} \\ & - \left( \frac{1}{4} \lambda^{ru} \tau^{st} \tau^{vw} + \frac{1}{2} \lambda^{ru} \tau^{sw} \tau^{tv} - \frac{1}{3} \tau^{ru} \tau^{sw} \tau^{tv} \right) \lambda_{rst} \lambda_{uvw} \end{aligned}$$

$$\begin{aligned}
 & + (\lambda^{ru}\tau^{st}\lambda^{vw} + \lambda^{ru}\lambda^{sw}\lambda^{tv} - \nu^{ru}\lambda^{sw}\nu^{tv})\lambda_{rst}(\lambda_{uv})_w \\
 & - (\lambda^{ru}\lambda^{st}\lambda^{vw} - \nu^{ru}\nu^{st}\nu^{vw} + \lambda^{ru}\lambda^{sw}\lambda^{tv} - \nu^{ru}\nu^{sw}\nu^{tv})(\lambda_{rs})_t(\lambda_{uv})_w.
 \end{aligned}$$

Considere o vetor de parâmetros  $\theta = (\delta^\top, \beta^\top, \gamma)^\top$ , em que  $\gamma$  é escalar. Por definição, temos que  $\nu^{ab} = \nu^{a\gamma} = \nu^{ai} = \nu^{i\gamma} = 0$ ,  $\tau^{ab} = \lambda^{ab}$ ,  $\tau^{a\gamma} = \lambda^{a\gamma}$ ,  $\tau^{ai} = \lambda^{ai}$  e  $\tau^{i\gamma} = \lambda^{i\gamma}$ , em que  $\gamma$  é usado para representar o índice correspondente ao último componente do vetor de parâmetros  $(\delta^\top, \beta^\top, \gamma)^\top$ . Admitindo ortogonalidade entre  $\delta$ ,  $\beta$  e  $\gamma$ , pode ser mostrado que  $\tau^{ij} = \tau^{\gamma\gamma} = \lambda_{ai} = \lambda_{a\gamma} = \lambda_{i\gamma} = \lambda^{ai} = \lambda^{a\gamma} = \lambda^{i\gamma} = 0$ ,  $\nu^{ij} = \lambda^{ij}$  e  $\lambda^{\gamma\gamma} = \nu^{\gamma\gamma}$ . Após algumas manipulações algébricas, obtemos a seguinte expressão para  $c_m$ :

$$\begin{aligned}
 c_m & = \frac{1}{4}\lambda^{ab}\lambda^{cd}\lambda_{abcd} - \lambda^{ab}\lambda^{cd}(\lambda_{acd})_b + \lambda^{ab}\lambda^{cd}(\lambda_{ac})_{db} - \lambda^{ij}\lambda^{ab}(\lambda_{iab})_j \\
 & - \lambda^{\gamma\gamma}\lambda^{ab}(\lambda_{ab\gamma})_\gamma - \left( \frac{1}{4}\lambda^{ab}\lambda^{cd}\lambda^{ef} + \frac{1}{2}\lambda^{ab}\lambda^{cf}\lambda^{de} - \frac{1}{3}\lambda^{ab}\lambda^{cf}\lambda^{de} \right) \lambda_{acd}\lambda_{bef} \\
 & + (\lambda^{ab}\lambda^{cd}\lambda^{ef} + \lambda^{ab}\lambda^{cf}\lambda^{de})\lambda_{acd}(\lambda_{be})_f - (\lambda^{ab}\lambda^{cd}\lambda^{ef} + \lambda^{ab}\lambda^{cf}\lambda^{de})(\lambda_{ac})_d(\lambda_{be})_f \\
 & - \left( \frac{1}{4}\lambda^{ij}\lambda^{ab}\lambda^{cd} + \frac{1}{2}\lambda^{ij}\lambda^{ad}\lambda^{bc} \right) \lambda_{iab}\lambda_{jcd} + (\lambda^{ij}\lambda^{ab}\lambda^{kl})\lambda_{iab}(\lambda_{jk})_l \\
 & - \left( \frac{1}{4}\lambda^{\gamma\gamma}\lambda^{ab}\lambda^{cd} + \frac{1}{2}\lambda^{\gamma\gamma}\lambda^{ad}\lambda^{bc} \right) \lambda_{ab\gamma}\lambda_{cd\gamma} + (\lambda^{\gamma\gamma}\lambda^{ab}\lambda^{\gamma\gamma})\lambda_{ab\gamma}(\lambda_{\gamma\gamma})_\gamma. \quad (2.8)
 \end{aligned}$$

É importante observar que a expressão (2.8), que derivamos, pode ser usada para obter expressões em forma fechada para o fator de correção de Bartlett em qualquer classe de modelos que utiliza a partição do vetor de parâmetros da mesma forma considerada aqui e em que a ortogonalidade entre os parâmetros  $\delta$ ,  $\beta$  e  $\gamma$  se verifica. Assim, a expressão (2.8) é bastante geral e pode ser usada para aperfeiçoar o teste da razão de verossimilhanças perfiladas modificadas em diferentes classes de modelos.

No modelo de regressão normal linear heteroscedástico considerado aqui, pode ser mostrado facilmente que

$$\frac{\partial L^*}{\partial \delta^a} = -\frac{1}{2\gamma}(y - X\beta)^\top Q_{R_1}(y - X\beta).$$

A expressão para  $Q_{R_1}$  segue de

$$Q_{R_s} = \text{diag} \left\{ \frac{\partial^s q_1}{\partial \delta^{a_1} \dots \partial \delta^{a_s}}, \dots, \frac{\partial^s q_n}{\partial \delta^{a_1} \dots \partial \delta^{a_s}} \right\}$$

com  $s = 1$ , em que  $q_1, \dots, q_n$  são dados em (2.4). Usando o fato de que se  $E(z) = \mu$  e  $\text{cov}(z) = \Sigma$  para um vetor aleatório  $z$  de dimensão  $n \times 1$ , então  $E(z^\top Az) = \text{tr}(A\Sigma) + \mu^\top A\mu$ , com  $A$  sendo uma matriz não aleatória de dimensão  $n \times n$ , obtemos

$$E\left(\frac{\partial L^*}{\partial \delta^a}\right) = -\frac{1}{2}\text{tr}(Q_{R_1}Q^{-1}) = -\frac{1}{2}\sum_{\ell=1}^n \frac{q_{\ell_a}}{q_\ell},$$

em que  $q_{\ell_a} = \partial q_\ell / \partial \delta^a$ . Contudo,

$$\begin{aligned} \sum_{\ell=1}^n \frac{q_{\ell_a}}{q_\ell} &= \sum_{\ell=1}^n \frac{\partial}{\partial \delta^a} \log q_\ell = \frac{\partial}{\partial \delta^a} \left\{ \sum_{\ell=1}^n \log q_\ell \right\} = \frac{\partial}{\partial \delta^a} \left\{ \sum_{\ell=1}^n \log \left[ \frac{(\prod_{s=1}^n m_s)^{1/n}}{m_\ell} \right] \right\} \\ &= \frac{\partial}{\partial \delta^a} \left\{ \sum_{\ell=1}^n \left[ \frac{1}{n} \sum_{s=1}^n \log m_s - \log m_\ell \right] \right\} = \frac{\partial}{\partial \delta^a} \left\{ \sum_{s=1}^n \log m_s - \sum_{\ell=1}^n \log m_\ell \right\} \\ &= \frac{\partial}{\partial \delta^a} \{0\} = 0. \end{aligned}$$

Assim, como esperado,  $E(\partial L^* / \partial \delta^a) = 0$ . Observe que

$$E\left(\frac{\partial^2 L^*}{\partial \delta^a \partial \gamma}\right) = \frac{1}{2\gamma^2} E\left\{(y - X\beta)^\top Q_{R_1}(y - X\beta)\right\} = \frac{1}{2\gamma} \text{tr}(Q_{R_1}Q^{-1}) = \frac{1}{2\gamma} \sum_{\ell=1}^n \frac{q_{\ell_a}}{q_\ell} = 0,$$

de forma que a ortogonalidade entre  $\delta$  e  $\gamma$  é verificada, como desejado.

A segunda derivada de  $L^*$  com respeito aos elementos de  $\delta$  é

$$\frac{\partial^2 L^*}{\partial \delta^a \partial \delta^b} = -\frac{1}{2\gamma}(y - X\beta)^\top Q_{R_2}(y - X\beta).$$

Assim,

$$\lambda_{ab} = -\frac{1}{2}\text{tr}(Q_{R_2}Q^{-1}) = -\frac{1}{2}\sum_{\ell=1}^n \frac{q_{\ell_{ab}}}{q_\ell},$$

em que  $q_{\ell_{ab}} = \partial^2 q_\ell / (\partial \delta^a \partial \delta^b)$ . Usando ainda a notação  $q_{\ell_{abc}} = \partial^3 q_\ell / (\partial \delta^a \partial \delta^b \partial \delta^c)$  e  $q_{\ell_{abcd}} = \partial^4 q_\ell / (\partial \delta^a \partial \delta^b \partial \delta^c \partial \delta^d)$ , obtemos, após alguma álgebra (ver Apêndice A),

$$\begin{aligned} \lambda_{abc} &= -\frac{1}{2}\sum_{\ell=1}^n \frac{q_{\ell_{abc}}}{q_\ell}, \quad \lambda_{abcd} = -\frac{1}{2}\sum_{\ell=1}^n \frac{q_{\ell_{abcd}}}{q_\ell}, \quad \lambda_{ab\gamma} = \frac{1}{2\gamma}\sum_{\ell=1}^n \frac{q_{\ell_{ab}}}{q_\ell}, \\ (\lambda_{ab\gamma})_\gamma &= -\frac{1}{2\gamma^2}\sum_{\ell=1}^n \frac{q_{\ell_{ab}}}{q_\ell}, \quad \lambda_{\gamma\gamma} = -\frac{n}{2\gamma^2}, \quad \lambda^{\gamma\gamma} = -\frac{2\gamma^2}{n}, \quad (\lambda_{\gamma\gamma})_\gamma = \frac{n}{\gamma^3}, \\ (\lambda_{abc})_d &= -\frac{1}{2}\sum_{\ell=1}^n \frac{\partial}{\partial \delta^d} \left( \frac{q_{\ell_{abc}}}{q_\ell} \right), \quad (\lambda_{ab})_{cd} = -\frac{1}{2}\sum_{\ell=1}^n \frac{\partial^2}{\partial \delta^c \partial \delta^d} \left( \frac{q_{\ell_{ab}}}{q_\ell} \right), \\ (\lambda_{ab})_c &= -\frac{1}{2}\sum_{\ell=1}^n \frac{\partial}{\partial \delta^c} \left( \frac{q_{\ell_{ab}}}{q_\ell} \right), \quad \lambda_{iab} = 0. \end{aligned}$$

Como  $\lambda_{iab} = 0$ , é possível simplificar a expressão de  $c_m$  dada em (2.8) para

$$\begin{aligned}
 c_m = & \lambda^{ab}\lambda^{cd} \left\{ \frac{1}{4}\lambda_{abcd} - (\lambda_{acd})_b + (\lambda_{ac})_{db} - \frac{1}{4}\lambda^{\gamma\gamma}\lambda_{ab\gamma}\lambda_{cd\gamma} \right\} \\
 & - \lambda^{ab}\lambda^{cd}\lambda^{ef} \left\{ \frac{1}{4}\lambda_{acd}\lambda_{bef} + \lambda_{acd}(\lambda_{be})_f - (\lambda_{ac})_d(\lambda_{be})_f \right\} \\
 & - \lambda^{ab}\lambda^{cf}\lambda^{de} \left\{ \frac{1}{6}\lambda_{acd}\lambda_{bef} + \lambda_{acd}(\lambda_{be})_f - (\lambda_{ac})_d(\lambda_{be})_f \right\} \\
 & - \lambda^{ab} \left\{ \lambda^{\gamma\gamma}(\lambda_{ab\gamma})_\gamma - (\lambda^{\gamma\gamma})^2\lambda_{ab\gamma}(\lambda_{\gamma\gamma})_\gamma \right\} - \frac{1}{2}\lambda^{ad}\lambda^{bc}\lambda^{\gamma\gamma}\lambda_{ab\gamma}\lambda_{cd\gamma}. \quad (2.9)
 \end{aligned}$$

Considerando o caso de heteroscedasticidade multiplicativa, isto é, o caso em que  $m_\ell = \exp\{z_\ell^\top \delta\}$ , obtemos  $q_\ell = \exp\{-(z_\ell - \bar{z})^\top \delta\}$ ,  $q_{\ell a} = -(z_\ell - \bar{z})_a \exp\{-(z_\ell - \bar{z})^\top \delta\} = -(z_\ell - \bar{z})_a q_\ell$ ,  $q_{\ell ab} = (z_\ell - \bar{z})_{ab} q_\ell$ ,  $q_{\ell abc} = -(z_\ell - \bar{z})_{abc} q_\ell$  e  $q_{\ell abcd} = (z_\ell - \bar{z})_{abcd} q_\ell$ , com  $\bar{z} = (\bar{z}_1, \dots, \bar{z}_p)^\top$  e  $\bar{z}_a = n^{-1} \sum_{\ell=1}^n z_{\ell a}$  para  $a = 1, \dots, p$ . Seja  $(z_\ell - \bar{z})_a = z_{\ell a} - \bar{z}_a$ ,  $(z_\ell - \bar{z})_{ab} = (z_{\ell a} - \bar{z}_a)(z_{\ell b} - \bar{z}_b)$ , e assim por diante. Substituindo os  $q$ 's nas expressões dos cumulantes  $\lambda$ 's, temos

$$\begin{aligned}
 \lambda_{abcd} &= -\frac{1}{2} \sum_{\ell=1}^n (z_\ell - \bar{z})_{abcd}, \quad \lambda_{abc} = \frac{1}{2} \sum_{\ell=1}^n (z_\ell - \bar{z})_{abc}, \quad \lambda_{ab} = -\frac{1}{2} \sum_{\ell=1}^n (z_\ell - \bar{z})_{ab}, \\
 \lambda_{ab\gamma} &= \frac{1}{2\gamma} \sum_{\ell=1}^n (z_\ell - \bar{z})_{ab}, \quad (\lambda_{ab\gamma})_\gamma = -\frac{1}{2\gamma^2} \sum_{\ell=1}^n (z_\ell - \bar{z})_{ab}, \quad \lambda^{\gamma\gamma} = -\frac{2\gamma^2}{n}, \\
 (\lambda_{acd})_b &= (\lambda_{ac})_{db} = (\lambda_{be})_f = 0.
 \end{aligned}$$

Substituindo estas quantidades na equação (2.9), é possível escrever  $c_m$  como

$$\begin{aligned}
 c_m = & -\frac{1}{8} \sum_{\ell=1}^n (z_\ell - \bar{z})_a \lambda^{ab} (z_\ell - \bar{z})_b (z_\ell - \bar{z})_c \lambda^{cd} (z_\ell - \bar{z})_d \\
 & + \frac{1}{8n} \sum_{\ell=1}^n (z_\ell - \bar{z})_a \lambda^{ab} (z_\ell - \bar{z})_b \sum_{s=1}^n (z_s - \bar{z})_c \lambda^{cd} (z_s - \bar{z})_d \\
 & - \frac{1}{16} \sum_{\ell=1}^n \sum_{s=1}^n (z_\ell - \bar{z})_a \lambda^{ab} (z_\ell - \bar{z})_b (z_\ell - \bar{z})_c \lambda^{cd} (z_s - \bar{z})_d (z_s - \bar{z})_e \lambda^{ef} (z_s - \bar{z})_f \\
 & - \frac{1}{24} \sum_{\ell=1}^n \sum_{s=1}^n (z_\ell - \bar{z})_a \lambda^{ab} (z_s - \bar{z})_b (z_\ell - \bar{z})_c \lambda^{cd} (z_s - \bar{z})_d (z_\ell - \bar{z})_e \lambda^{ef} (z_s - \bar{z})_f
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{n} \sum_{\ell=1}^n (z_{\ell} - \bar{z})_a \lambda^{ab} (z_{\ell} - \bar{z})_b \\
 & + \frac{1}{4n} \sum_{\ell=1}^n \sum_{s=1}^n (z_{\ell} - \bar{z})_a \lambda^{ab} (z_s - \bar{z})_b (z_{\ell} - \bar{z})_c \lambda^{cd} (z_s - \bar{z})_d.
 \end{aligned}$$

Seja  $H = \{h_{\ell s}\} = (Z - \bar{Z})[(Z - \bar{Z})^{\top}(Z - \bar{Z})]^{-1}(Z - \bar{Z})^{\top}$ , com  $(Z - \bar{Z}) = (z_1 - \bar{z}, \dots, z_n - \bar{z})^{\top}$ . Observe que  $(z_{\ell} - \bar{z})_a \lambda^{ab} (z_s - \bar{z})_b = -2h_{\ell s}$ . Então, é possível reduzir a expressão anterior de  $c_m$  a

$$\begin{aligned}
 c_m & = -\frac{1}{2} \sum_{\ell=1}^n h_{\ell\ell}^2 + \frac{1}{2n} \left( \sum_{\ell=1}^n h_{\ell\ell} \right)^2 + \frac{1}{2} \sum_{\ell=1}^n \sum_{s=1}^n h_{\ell\ell} h_{\ell s} h_{s s} + \frac{1}{3} \sum_{\ell=1}^n \sum_{s=1}^n h_{\ell s}^3 \\
 & \quad - \frac{2}{n} \sum_{\ell=1}^n h_{\ell\ell} + \frac{1}{n} \sum_{\ell=1}^n \sum_{s=1}^n h_{\ell s}^2.
 \end{aligned}$$

Como  $\sum_{\ell=1}^n h_{\ell\ell} = \text{tr}(H) = p$ , temos que  $c_m$  pode ser escrito em forma matricial como

$$c_m = -\frac{1}{2} \text{tr}(H_d H_d) + \frac{1}{2n} p^2 + \frac{1}{2} \iota^{\top} H_d H H_d \iota + \frac{1}{3} \iota^{\top} H^{(3)} \iota - \frac{2}{n} p + \frac{1}{n} \iota^{\top} H^{(2)} \iota, \quad (2.10)$$

em que  $H_d = \text{diag}\{h_{11}, \dots, h_{mm}\}$ ,  $H^{(2)} = (h_{\ell s}^2)$  e  $H^{(3)} = (h_{\ell s}^3)$ .

É importante salientar alguns pontos relativos à expressão (2.10). Primeiro, ela é uma expressão matricial simples para o fator de correção de Bartlett da estatística da razão de verossimilhanças perfiladas modificadas para o teste de heteroscedasticidade no modelo de regressão normal linear. Ela somente envolve operações simples de matrizes e pode ser facilmente implementada em pacotes de computação simbólica e softwares que permitam executar operações simples de álgebra linear, tais como `Ox`, `MAPLE`, `MATHEMATICA`, `S-Plus`, `R`, etc. Segundo, o fator de correção  $c_m$  dado em (2.10) depende somente da matriz  $Z$  de covariadas usadas para modelar as flutuações na variância condicional da resposta, do número de parâmetros desconhecidos que definem o comportamento de heteroscedasticidade e do número de observações. Em particular, é importante notar que o fator de correção de Bartlett não depende do número de parâmetros de perturbação (nem depende de qualquer parâmetro desconhecido). Isto pode ser interpretado levando em conta o fato de

que o ajuste de Cox e Reid para a verossimilhança perfilada reduz o impacto de parâmetros de perturbação na inferência resultante, e este ajuste foi usado antes da aplicação da correção de Bartlett. Finalmente, observamos que a expressão (2.10) generaliza o resultado obtido por Ferrari e Cribari-Neto (2002, p. 358), que só é válido para  $p = 1$ , como havíamos mencionado anteriormente. É importante observar que estes autores aplicaram a correção de Bartlett à estatística do teste usada por Simonoff e Tsai (1994), que é uma aproximação à forma original da estatística de teste obtida da função de verossimilhança perfilada modificada de Cox e Reid (1987).

## 2.4 Resultados numéricos

Apresentamos alguns resultados de simulação para avaliar a eficácia das correções de Bartlett para os testes da razão de verossimilhanças original e da razão de verossimilhanças perfiladas modificadas. Comparamos os desempenhos de quatro testes, isto é, o da razão de verossimilhanças original ( $LR$ ), o teste da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) e suas respectivas versões corrigidas. Estes desempenhos são avaliados em função da proximidade das probabilidades de rejeição da hipótese nula, sendo esta verdadeira (probabilidade do erro tipo I), aos respectivos níveis nominais dos testes. As simulações realizadas são baseadas no modelo de regressão linear

$$y_\ell = \beta_1 + \beta_2 x_{\ell 2} + \cdots + \beta_k x_{\ell k} + u_\ell, \quad \ell = 1, \dots, n,$$

em que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{\delta_1 z_{\ell 1} + \cdots + \delta_p z_{\ell p}\})$ ,  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$  e a hipótese nula considerada é  $\mathcal{H}_0 : \delta_1 = \cdots = \delta_p = 0$  indicando que o modelo é homoscedástico. A variável resposta foi gerada assumindo que  $\beta_1 = \cdots = \beta_k = 1$ ,  $\sigma^2 = 1$  e diferentes valores de  $p$  e  $k$  foram considerados.

As covariadas  $x_2, \dots, x_k$  foram geradas como amostras aleatórias da distribuição  $\mathcal{U}(0, 1)$ . Quando  $p < k$ , a matriz  $Z$  foi construída usando as colunas  $2, \dots, p + 1$  de  $X$ ; quando  $p \geq k$ , as colunas extras de  $Z$  foram criadas usando amostras independentes da distribuição  $\mathcal{U}(0, 1)$ .

O número de réplicas foi fixado em 10.000 e foram considerados os seguintes níveis nominais:  $\alpha=10\%$ ,  $5\%$ ,  $1\%$ ,  $0,5\%$ . As simulações foram realizadas usando a linguagem de programação matricial Ox (Doornik, 2001).

Para cada tamanho de amostra e cada nível considerado, calculamos as taxas de rejeição de cada teste, isto é, estimamos, via simulação,  $P(LR \geq x_\alpha)$ ,  $P(LR^* \geq x_\alpha)$ ,  $P(LR_m \geq x_\alpha)$  e  $P(LR_m^* \geq x_\alpha)$  em que  $x_\alpha$  é o quantil  $(1 - \alpha)$  apropriado da distribuição  $\chi_p^2$ . Todas as entradas das tabelas apresentadas são porcentagens.

A Tabela 2.1 apresenta resultados de simulação para  $n = 35$ ,  $p = 2$  e diferentes valores de  $k$ . Variamos  $k$  para analisar o efeito do número de parâmetros de perturbação nos diferentes testes. Observamos que o teste da razão de verossimilhanças original é liberal, especialmente quando o número de parâmetros de perturbação aumenta. Por exemplo, quando  $k = 7$  e  $\alpha = 5\%$ , a taxa de rejeição do teste excede  $16\%$ . A tendência do teste em rejeitar com frequência demasiada a hipótese nula é atenuada pela correção de Bartlett de forma que o teste da razão de verossimilhanças corrigido apresenta distorções de tamanho menores. Por exemplo, sua taxa de rejeição para a mesma situação é  $9,0\%$ . O teste da razão de verossimilhanças perfiladas modificadas tende a corrigir a tendência liberal do teste original, exibindo taxas de rejeição menores que o nível nominal do teste. Novamente, quando  $k = 7$  e  $\alpha = 5\%$ , sua taxa de rejeição é  $4,2\%$ . A correção de Bartlett aplicada a este teste leva as taxas de rejeição para valores mais próximos do tamanho nominal do teste. Para a situação acima, a taxa de rejeição do teste da razão de verossimilhanças perfiladas modificadas corrigido sob  $\mathcal{H}_0$  é  $4,6\%$ . As conclusões anteriores também se verificam para outros níveis nominais, até mesmo para níveis nominais muito pequenos.

A Tabela 2.2 contém resultados para a situação em que  $n = 35$ ,  $k = 5$  e  $p = 1, \dots, 8$ . Novamente, o teste da razão de verossimilhanças original é consideravelmente liberal, especialmente quando o número de parâmetros que definem o comportamento de heteroscedasticidade aumenta; por exemplo, quando  $p = 5$  e  $\alpha = 5\%$ , a taxa de rejeição do teste é  $20\%$ , isto é, quatro vezes maior que o nível nominal selecionado. As outras conclusões obtidas a partir da Tabela 2.1 também

se verificam aqui. Para a mesma situação ( $p = 5$  e  $\alpha = 5\%$ ), a taxa de rejeição do teste da razão de verossimilhanças corrigido, do teste da razão de verossimilhanças perfiladas modificadas e de sua versão corrigida são 9,9%, 4,5% e 5,0%, respectivamente.

Na Tabela 2.3, fixamos o valor de  $p$  em 2 e variamos o tamanho da amostra em  $n = 35, 50$  e  $100$  para  $k = 5$  e  $8$ . Com aumento do tamanho da amostra, as taxas de rejeição de todos os quatro testes ficam mais próximas do nível nominal, como era esperado. Observamos também que até mesmo para  $n = 100$  o teste da razão de verossimilhanças original apresenta taxas de rejeição sob  $\mathcal{H}_0$  bem acima dos níveis nominais considerados. Como ilustração, quando  $k = 8$ ,  $\alpha = 5\%$  e  $n = 50$ , as taxas de rejeição do teste da razão de verossimilhanças, de sua versão corrigida, do teste da razão de verossimilhanças perfiladas modificadas e de sua versão corrigida são, respectivamente, 13,6%, 7,7%, 4,6% e 5,0%.

O teste de igualdade de variâncias para o modelo com um fator e três níveis é considerado na Tabela 2.4. Aqui, há três populações normais com médias  $\mu_1, \mu_2, \mu_3$  e variâncias  $\sigma_1^2, \sigma_2^2, \sigma_3^2$  e a hipótese nula de interesse a ser testada é  $\mathcal{H}_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ . Observe que  $p = 2$  e  $k = 3$ . Os números de observações de cada uma das três populações são  $n_1, n_2$  e  $n_3$  com  $n_1 + n_2 + n_3 = n$ . A simulação foi realizada para  $n_1 = n_2 = n_3 = n/3$ . As conclusões das taxas de rejeição dos quatro testes estão de acordo com as conclusões tiradas das Tabelas 2.1, 2.2 e 2.3.

A Tabela 2.5 apresenta resultados de simulação obtidos levando em consideração a hipótese alternativa (heteroscedasticidade) para  $n = 30$ ,  $p = 2$  e valores diferentes de  $\delta_1 = \delta_2 = \delta$  foram considerados ao nível nominal  $\alpha = 10\%$ . É importante observar que estas simulações de poder correspondem à situação abordada na Tabela 2.3 para  $n = 30$ . Somente comparamos os poderes dos testes da razão de verossimilhanças perfiladas modificadas e de sua versão corrigida, uma vez que o teste da razão de verossimilhanças original e sua versão corrigida são consideravelmente liberais, isto é, apresentam taxas de rejeição bem acima dos níveis nominais e, portanto, não podem ser recomendados. Os resultados indicam que não há nenhuma perda de poder derivada do fato de usar o fator de correção de Bartlett derivado

neste capítulo. Os poderes dos dois testes são semelhantes, com leve vantagem do teste corrigido, vantagem esta que se origina de sua menor distorção de tamanho.

Os resultados para  $LR_m^*$  apresentados aqui são obtidos corrigindo a forma original da estatística da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ), que envolve a maximização da função de verossimilhança perfilada modificada ( $L_{mp}^*$ ), enquanto que Ferrari e Cribari–Neto (2002) corrigiram o teste proposto por Simonoff e Tsai (1994), que é uma aproximação à forma original e não envolve tal maximização. Portanto, é interessante ver o que ocorre quando se corrige esta aproximação ao invés de corrigir a estatística  $LR_m$  na sua forma original. Observamos que os desempenhos em amostras finitas de ambos os testes, baseados em  $LR_m$  e  $LR_m^*$ , são melhores quando não usamos a aproximação proposta por Simonoff e Tsai (1994). Por exemplo, considere o resultado de simulação na Tabela 2.3 com  $k = 8$ ,  $\alpha = 5\%$  e  $n = 50$ . Como podemos observar, as taxas de rejeição para o teste da razão de verossimilhanças perfiladas modificadas e sua versão corrigida são 4,6% e 5,0%, respectivamente. As taxas de rejeição correspondentes para os dois testes que empregam a aproximação usada por Simonoff e Tsai (1994) são 3,7% e 4,1%.

Em linhas gerais, o teste da razão de verossimilhanças original pode ser severamente liberal, rejeitando a hipótese nula muito mais freqüentemente do que o esperado com base no nível nominal selecionado. O fator de correção de Bartlett para o teste da razão de verossimilhanças faz com que a taxa de rejeição empírica fique mais próxima do nível nominal do teste, mas não corrige sua tendência liberal completamente. O teste da razão de verossimilhanças perfiladas modificadas, por outro lado, introduz uma correção na taxa de rejeição produzindo taxas de rejeição abaixo dos níveis nominais. O fator de correção de Bartlett para este teste reduz a distorção de tamanho e, dentre os quatros testes apresentados, o teste que o emprega é o mais recomendável.

Tabela 2.1 *Tamanho dos testes - Modelo normal linear com  $n = 35$ ,  $p = 2$  e diversos valores para  $k$ .*

$k$	$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$			$\alpha = 0,5\%$		
	$LR$	$LR^*$	$LR_m$	$LR$	$LR^*$	$LR_m$	$LR$	$LR^*$	$LR_m$	$LR$	$LR^*$	$LR_m$
2	13,4	10,6	8,9	7,5	5,4	4,3	1,6	1,1	0,7	0,9	0,5	0,4
3	14,5	10,9	8,7	8,2	5,9	4,2	2,3	1,2	0,8	1,0	0,7	0,3
4	16,3	11,2	8,3	9,5	5,7	4,1	2,7	1,3	0,8	1,0	0,8	0,4
5	19,2	12,8	8,7	11,9	7,2	4,3	3,8	1,6	0,7	0,9	0,9	0,4
6	22,3	14,4	8,1	14,6	7,8	4,0	5,2	2,1	0,7	1,0	1,3	0,4
7	24,9	15,6	8,9	16,6	9,0	4,2	6,3	2,5	0,8	1,0	1,5	0,4
8	30,4	18,9	8,4	21,8	11,7	3,9	9,7	3,7	0,8	0,9	2,3	0,3

Tabela 2.2 *Tamanho dos testes - Modelo normal linear com  $n = 35$ ,  $k = 5$  e diversos valores para  $p$ .*

$p$	$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$			$\alpha = 0,5\%$		
	$LR$	$LR^*$	$LR_m$	$LR$	$LR^*$	$LR_m$	$LR$	$LR^*$	$LR_m$	$LR$	$LR^*$	$LR_m$
1	16,3	11,6	9,1	9,6	6,1	4,4	3,0	1,6	0,9	1,1	1,8	0,9
2	19,2	12,8	8,7	11,9	7,2	4,3	3,8	1,6	0,7	0,9	2,4	0,9
3	21,8	13,9	8,8	14,0	7,5	4,2	4,7	1,9	0,8	1,0	2,9	1,0
4	26,6	15,8	8,7	17,2	9,1	4,1	6,3	2,4	0,9	1,0	4,1	1,4
5	29,6	17,3	9,0	20,0	9,9	4,5	8,0	2,8	0,8	1,0	5,3	1,7
6	36,0	21,2	9,1	25,9	13,0	4,7	11,2	4,4	1,1	1,3	8,0	2,8
7	46,1	27,6	9,4	34,9	18,0	4,4	17,3	6,9	0,8	1,0	12,9	4,2
8	49,4	28,8	9,2	37,4	19,2	4,2	18,9	7,2	0,9	1,0	14,1	4,7

Tabela 2.3 *Tamanho dos testes - Modelo normal linear com  $p = 2$ ,  $k = 5, 8$  e diversos valores para  $n$ .*

$n$	$\alpha$ (%)	$k = 5$				$k = 8$			
		$LR$	$LR^*$	$LR_m$	$LR_m^*$	$LR$	$LR^*$	$LR_m$	$LR_m^*$
30	10,0	25,0	16,4	9,1	10,2	36,6	23,4	8,3	9,2
	5,0	16,7	9,9	4,2	5,0	27,5	14,9	3,7	4,3
	1,0	6,6	2,8	0,8	1,0	13,4	4,8	0,6	0,8
	0,5	4,2	1,5	0,4	0,6	9,5	3,1	0,3	0,4
50	10,0	16,2	11,2	9,2	9,8	21,6	13,8	9,4	10,1
	5,0	9,2	5,8	4,3	4,7	13,6	7,7	4,6	5,0
	1,0	2,5	1,3	0,9	1,0	4,7	2,1	1,0	1,1
	0,5	1,5	0,7	0,4	0,5	3,0	1,1	0,5	0,4
100	10,0	12,0	9,8	9,1	9,4	14,5	10,8	9,8	10,0
	5,0	6,2	5,0	4,6	4,9	8,1	5,7	4,9	5,2
	1,0	1,6	1,0	0,9	1,0	2,2	1,3	1,0	1,1
	0,5	0,8	0,5	0,4	0,4	1,3	0,6	0,5	0,6

Tabela 2.4 *Tamanho dos testes - Modelo de análise de variância com um fator e três níveis,  $k = 3$  e  $p = 2$ .*

$n$	$\alpha$ (%)	$LR$	$LR^*$	$LR_m$	$LR_m^*$
9	10,0	28,2	14,4	5,4	6,8
	5,0	18,6	8,1	2,4	3,2
	1,0	7,4	2,2	0,3	0,5
	0,5	5,0	1,2	0,1	0,2
18	10,0	16,6	10,6	8,5	9,2
	5,0	9,6	5,8	4,2	4,7
	1,0	3,1	1,3	0,8	0,9
	0,5	1,9	0,7	0,4	0,5
27	10,0	14,6	10,6	9,3	9,9
	5,0	8,2	5,5	4,7	5,0
	1,0	2,1	1,1	0,9	1,0
	0,5	1,2	0,6	0,4	0,5
36	10,0	12,8	10,1	9,4	9,8
	5,0	7,2	5,3	4,7	5,0
	1,0	1,8	1,1	0,9	1,1
	0,5	1,0	0,5	0,4	0,5
45	10,0	12,3	10,1	9,3	9,7
	5,0	6,5	5,2	4,7	5,0
	1,0	1,5	1,1	1,0	1,1
	0,5	0,9	0,5	0,5	0,5

Tabela 2.5 Poder dos testes - Modelo normal linear com  $n = 30$ ,  $p = 2$  e  $\alpha = 10\%$ .

$\delta$	$k = 5$		$k = 8$	
	$LR_m$	$LR_m^*$	$LR_m$	$LR_m^*$
0,5	11,1	12,3	11,2	12,6
1,0	26,3	28,2	25,0	26,9
1,5	45,2	47,2	48,7	51,0
2,0	63,8	65,9	70,1	72,2
2,5	73,9	75,5	75,7	77,3
3,0	91,6	92,3	83,8	85,6
3,5	99,9	99,9	92,2	92,9

## 2.5 Aplicações

Apresentamos a seguir duas ilustrações numéricas dos testes considerados nas seções anteriores.

### 2.5.1 Exemplo 1

Consideremos os dados analisados por Montgomery, Peck e Vining (2001, p. 76) referentes a serviços de manutenção e reposição de latas e garrafas de bebidas em máquinas de venda automáticas (ver Apêndice F, Tabela F.1). A variável resposta  $y$  é o tempo (em minutos) gasto no serviço, as covariadas  $x_2$  e  $x_3$  utilizadas são, respectivamente, o número de bebidas estocadas e a distância percorrida (em pés). O modelo usado é

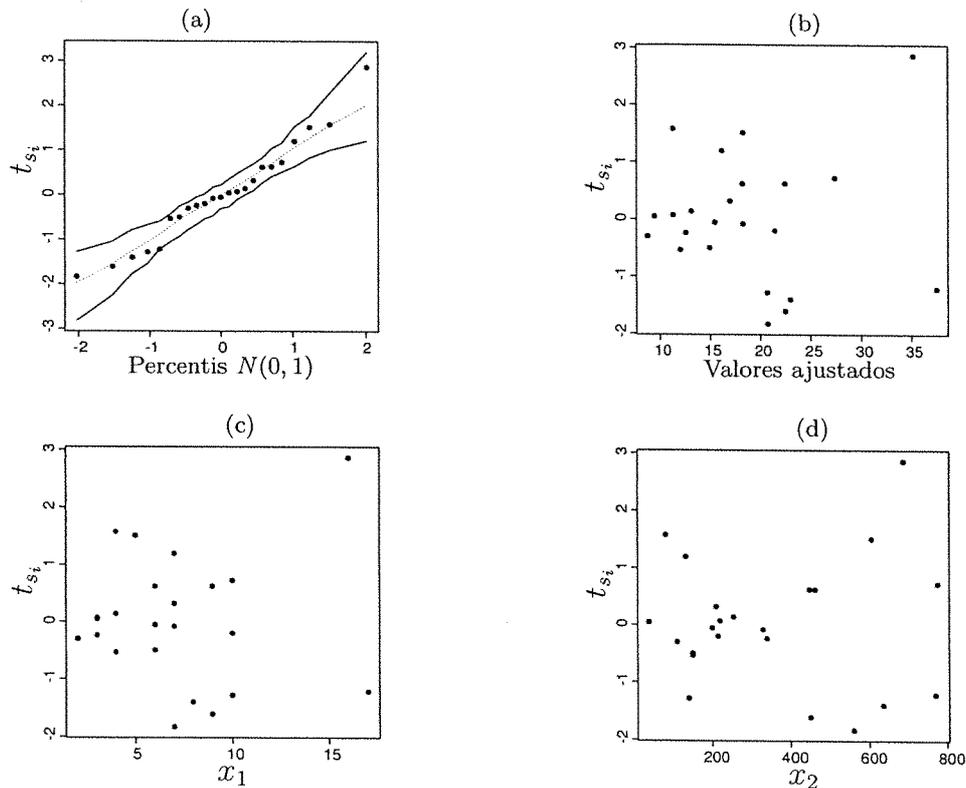
$$y_\ell = \beta_1 + x_{\ell 2}\beta_2 + x_{\ell 3}\beta_3 + u_\ell, \quad \ell = 1, \dots, 25.$$

Assumimos que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{x_{\ell 2}\delta_1 + x_{\ell 3}\delta_2\})$  e  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$ .

Todas as análises de diagnóstico feitas e omitidas aqui destacaram as observações 9 e 22. Estas observações aparecem como pontos de alavanca, sendo que a observação 9 também é classificada como observação influente; ver Montgomery, Peck e Vining (2001, pp. 210, 213, 215, 216, 217). No que segue, trabalhamos com o conjunto de dados incompleto, ou seja, excluimos as observações 9 e 22 dos dados. As estimativas de máxima verossimilhanças dos parâmetros são  $\hat{\beta}_1 = 4,643$ ,

$\hat{\beta}_2 = 1,456$  e  $\hat{\beta}_3 = 0,011$ . A estimativa de máxima verossimilhança de  $\sigma^2$  é  $\hat{\sigma}^2 = 6,163$  e o coeficiente de determinação,  $R^2$ , vale 0,9072. Estas estimativas foram obtidas impondo  $\delta_1 = \delta_2 = 0$ . Análises de diagnóstico para os dados incompletos não mostram evidência de heteroscedasticidade e o gráfico normal de probabilidades com envelope para o resíduo studentizado não apresenta nenhum comportamento não usual (ver Figuras 2.1 (a) - 2.1 (d)).

Figura 2.1 (a) Gráfico dos resíduos studentizados  $t_{s_i}$  contra percentis normais com envelope simulado; (b) Gráfico dos resíduos studentizados  $t_{s_i}$  contra os valores ajustados; (c) e (d) Gráficos dos resíduos studentizados  $t_{s_i}$  contra variáveis explicativas para o modelo homoscedástico - Serviços de manutenção.



Nosso principal interesse é testar  $\mathcal{H}_0 : \delta_1 = \delta_2 = 0$  (homoscedasticidade) contra uma alternativa bilateral (heteroscedasticidade). Para este teste, temos  $LR = 4,825$ ,  $LR^* = 3,717$ ,  $LR_m = 4,126$  e  $LR_m^* = 4,352$ , que conduzem aos

respectivos níveis descritivos 0,090, 0,156, 0,127 e 0,113. Segue então que o teste da razão de verossimilhanças original leva à rejeição da hipótese nula (sugerindo assim a presença de heteroscedasticidade) ao nível nominal de 10%, enquanto que a não rejeição da hipótese nula (homoscedasticidade), ao mesmo nível nominal, é obtida quando os testes são baseados em  $LR^*$ ,  $LR_m$  e  $LR_m^*$ . Em outras palavras, os testes baseados na estatística da razão de verossimilhanças corrigida, na razão de verossimilhanças perfiladas modificadas e em sua versão corrigida conduzem a conclusão diferente daquela obtida através do teste da razão de verossimilhanças original.

Todas as rotinas computacionais necessárias para o cálculo dessas estatísticas, bem como o processo de estimação, estando os dados disponíveis em <http://www.de.ufpe.br/~cribari/FerrariCysneirosCribari.zip>, foram desenvolvidos na linguagem de programação matricial `Ox`.

### 2.5.2 Exemplo 2

O conjunto de dados que apresentamos agora é obtido de Simonoff e Tsai (1994) (ver Apêndice G, Tabela G.1). Os dados representam as taxas mensais de retorno de ações de mercado ( $x$ ) e da Corporação Acme Cleveland ( $y$ ), no período de janeiro de 1986 a dezembro de 1990. O modelo usado é

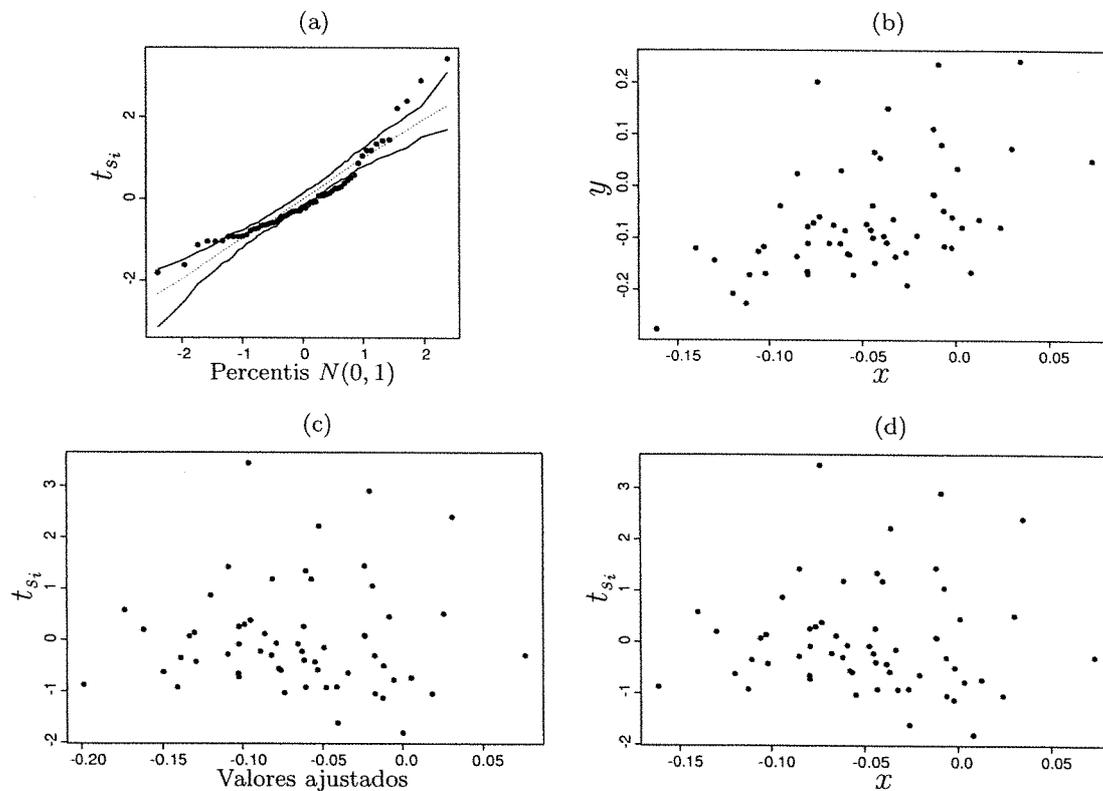
$$y_\ell = \beta_1 + \beta_2 x_\ell + u_\ell, \quad \ell = 1, \dots, 59.$$

Assumimos que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{\delta x_\ell\})$  e  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$ .

Todas as análises de diagnóstico feitas e omitidas aqui destacaram a observação 22, referente ao mês de outubro de 1987. Esta observação aparece como um ponto aberrante, refletindo que o mercado perdeu aproximadamente um terço de seu valor durante um período de duas semanas naquele mês (ver Simonoff e Tsai (1994, p. 359)). No que segue, trabalhamos com o conjunto de dados incompletos, ou seja, eliminamos a observação 22 dos dados. As estimativas de máxima verossimilhança dos parâmetros são  $\hat{\beta}_1 = -0,0095$ ,  $\hat{\beta}_2 = 1,1724$  e  $\hat{\sigma}^2 = 0,0087$ . Estas estimativas foram obtidas impondo  $\delta = 0$ . Análises de diagnóstico para os dados incompletos

mostram evidência de heteroscedasticidade e o gráfico normal de probabilidades com envelope para o resíduo studentizado mostra um leve afastamento da suposição da normalidade (ver Figuras 2.2 (a) - 2.2 (d)).

Figura 2.2 (a) Gráfico dos resíduos studentizados  $t_{s_i}$  contra percentis normais com envelope simulado; (b) Gráfico de dispersão de  $y$  contra  $x$ ; (c) Gráfico dos resíduos studentizados  $t_{s_i}$  contra os valores ajustados e (d) Gráfico dos resíduos studentizados  $t_{s_i}$  contra  $x$  para o modelo homoscedástico - Taxas mensais de retorno de ações.



Nosso objetivo é testar  $\mathcal{H}_0 : \delta = 0$  contra  $\mathcal{H}_1 : \delta \neq 0$ , ou seja, se o modelo é homoscedástico. Para este teste, temos  $LR = 3,329$ ,  $LR^* = 3,119$ ,  $LR_m = 2,968$  e  $LR_m^* = 3,071$ , que conduzem aos respectivos níveis descritivos 0,068, 0,077, 0,085 e 0,080. Segue então que todos os testes indicam uma heteroscedasticidade marginal, ao nível descritivo  $\alpha = 10\%$ . Aqui, o tamanho de amostra é grande ( $n = 59$ ) em

relação ao número de covariadas ( $k = 2$ ) e do número de covariadas utilizadas na especificação do parâmetro que define o comportamento da heteroscedasticidade ( $p = 1$ ). Conseqüentemente, como esperado, as correções para a estatística da razão de verossimilhanças não tiveram impacto significativo.

## 2.6 Comentários

É importante, na prática, testar a presença de heteroscedasticidade nos dados, isto é, verificar se as variâncias são constantes para decidir qual é a estratégia de modelagem mais apropriada. Então, é importante elaborar testes confiáveis, precisos, para detectar a heteroscedasticidade das variâncias.

Derivamos neste capítulo uma correção de Bartlett que pode ser aplicada a estatísticas da razão de verossimilhanças perfiladas modificadas. Os resultados cobrem a situação em que o parâmetro de interesse que define o comportamento de heteroscedasticidade nos dados é multidimensional, generalizando assim resultados disponíveis na literatura.

Os resultados numéricos apresentados comparam o desempenho em amostras finitas de quatro testes: o teste da razão de verossimilhanças original ( $LR$ ), sua versão corrigida através da correção de Bartlett ( $LR^*$ ), o teste da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) e o teste da razão de verossimilhanças perfiladas modificadas corrigido ( $LR_m^*$ ). Em geral, os resultados numéricos favorecem o teste que propusemos. A modificação da verossimilhança perfilada atenua o efeito de parâmetros de perturbação na inferência e a correção de Bartlett acelera a taxa de convergência da estatística de teste para sua distribuição assintótica de primeira ordem. Portanto, em aplicações práticas, sugerimos a pesquisadores empíricos o uso do teste proposto.

---

## CAPÍTULO 3

# Refinamento de um teste de heteroscedasticidade baseado em verossimilhança perfilada modificada em modelos não-lineares da família exponencial

---

### 3.1 Introdução

Os modelos não-lineares da família exponencial (Cordeiro e Paula, 1989) são extensões naturais dos modelos lineares generalizados e dos modelos de regressão normais não-lineares. São definidos por um conjunto de variáveis aleatórias independentes tendo distribuição na família exponencial e admitem que uma função monótona da média da variável resposta seja definida por um preditor não-linear envolvendo regressores e parâmetros desconhecidos.

Nesta classe de modelos, abordamos a situação em que os parâmetros de dispersão não são constantes para todas as observações, havendo assim uma estrutura heteroscedástica. Analogamente à estrutura estabelecida para as variâncias no modelo linear, admitimos a determinação de uma forma funcional que relaciona os parâmetros de dispersão com alguns parâmetros desconhecidos, que não dependem do vetor de parâmetros de regressão, e algumas variáveis auxiliares.

Wei, Shi, Fung e Hu (1998) utilizaram a abordagem da verossimilhança perfilada modificada proposta por Cox e Reid (1987) para desenvolver testes de heteroscedasticidade em modelos não-lineares da família exponencial. No Capítulo 2, obtivemos uma expressão matricial que possibilita o cálculo do fator de correção de Bartlett para a estatística da razão de verossimilhanças perfiladas modificadas associada ao teste de heteroscedasticidade no modelo de regressão linear. Neste capítulo, trabalharemos na classe dos modelos não-lineares da família exponencial estendendo, portanto, o trabalho desenvolvido no Capítulo 2, que é restrito ao modelo normal linear. Em outras palavras, desenvolvemos uma correção de Bartlett para o teste

proposto por Wei, Shi, Fung e Hu (1998), obtendo assim um teste da razão de verossimilhanças perfiladas modificadas corrigido que é preciso até segunda ordem.

Na Seção 3.2, apresentamos a definição dos modelos de regressão não-lineares da família exponencial. Na Seção 3.3, apresentamos alguns testes de heteroscedasticidade, incluindo um teste baseado em uma verossimilhança perfilada modificada. Na Seção 3.4 desenvolvemos um fator de correção de Bartlett que pode ser aplicado à estatística da razão de verossimilhanças perfiladas modificadas. Resultados numéricos sobre o comportamento em amostras finitas dos diferentes testes são apresentados na Seção 3.5. Na Seção 3.6, apresentamos aplicações a dados reais e, na Seção 3.7, alguns comentários são feitos.

### 3.2 Definição

Consideremos  $n$  variáveis aleatórias independentes  $y_1, \dots, y_n$ , cada qual com função densidade na família exponencial da forma

$$\pi(y_\ell; \theta_\ell, \phi_\ell) = \exp \left\{ \phi_\ell [y_\ell \theta_\ell - b(\theta_\ell) - c(y_\ell)] - \frac{1}{2} e(y, \phi_\ell) \right\}, \quad \ell = 1, \dots, n, \quad (3.1)$$

em que  $e(\cdot, \cdot)$ ,  $b(\cdot)$  e  $c(\cdot)$  são funções conhecidas e  $\theta_\ell$  e  $\phi_\ell$  são chamados de parâmetros canônico e de precisão, respectivamente. Assume-se que  $\phi_\ell$  ( $\phi_\ell > 0$ ) é desconhecido e  $\phi_\ell^{-1}$  pode ser visto como um parâmetro de dispersão. Admite-se que  $\phi_\ell^{-1} = \sigma^2 m_\ell$ , sendo  $m_\ell = m(z_\ell, \delta) > 0$  o  $\ell$ -ésimo elemento da matriz diagonal  $M$  de dimensão  $n \times n$ ,  $z_\ell^\top$  é a  $\ell$ -ésima linha da matriz  $Z$  de dimensão  $n \times p$  de covariadas usadas para modelar as flutuações na variância,  $\sigma^2$  é uma constante desconhecida finita e estritamente positiva e  $\delta$  é um vetor de dimensão  $p \times 1$  de parâmetros desconhecidos. Aqui, mesmo se os parâmetros de dispersão forem iguais para todas as observações, suas variâncias poderão diferir, pois estas podem depender das médias desconhecidas. Assim o conceito de heteroscedasticidade no presente contexto difere daquele empregado em modelos de regressão normais em que, sob homoscedasticidade, as variâncias são constantes. Dizemos então que o modelo é

“homoscedástico<sup>(1)</sup>” quando os parâmetros de dispersão  $\phi_1^{-1}, \dots, \phi_n^{-1}$  forem todos iguais; caso contrário, dizemos que o modelo é “heteroscedástico<sup>(2)</sup>”.

Para o modelo em (3.1), valem as relações:

$$E(y_\ell) = \mu_\ell = db(\theta_\ell)/d\theta_\ell, \quad \text{Var}(y_\ell) = \phi_\ell^{-1}V_\ell, \quad \ell = 1, \dots, n,$$

em que  $V = d\mu/d\theta$  é chamada função de variância e  $\theta = \int V^{-1}d\mu = q(\mu)$ , sendo  $q(\mu)$  uma função biunívoca. A função de variância caracteriza a distribuição e desempenha um papel importante no contexto da família exponencial. Dada a função de variância, tem-se a classe de distribuições correspondente e vice-versa.

Os modelos não-lineares da família exponencial são definidos por (3.1) e pelo componente sistemático

$$g(\mu_\ell) = \eta_\ell = f(x_\ell; \beta), \quad (3.2)$$

em que  $g(\cdot)$  é uma função conhecida monótona e duplamente diferenciável, denominada função de ligação, relacionando a média  $\mu_\ell$  de uma observação com o preditor não-linear  $\eta_\ell$  do modelo,  $\beta = (\beta_1, \dots, \beta_k)^\top$  é um vetor de  $k$  ( $k < n$ ) parâmetros desconhecidos a serem estimados,  $x_\ell = (x_{\ell 1}, \dots, x_{\ell K})^\top$  é um vetor de dimensão  $K$  de covariáveis conhecidas associadas à  $\ell$ -ésima observação e  $f(\cdot; \cdot)$  é uma função possivelmente não-linear no segundo argumento, contínua e diferenciável com respeito aos componentes de  $\beta$  tal que a matriz de derivadas  $X^* = X^*(\beta) = \partial\eta/\partial\beta^\top$ , com  $\eta = (\eta_1, \dots, \eta_n)^\top$ , tenha posto  $k$  para todo  $\beta$ . Supomos identificabilidade no sentido de que diferentes  $\beta$ 's implicam diferentes  $\eta$ 's. A matriz  $X^*$  tem elementos que são, em geral, funções do vetor de parâmetros  $\beta$  desconhecidos.

Sem perda de generalidade, poderíamos ter assumido que  $g(\mu_\ell) = \mu_\ell$ . No entanto, preferimos especificar o componente sistemático do modelo da forma estabelecida em (3.2), uma vez que tomando  $k = K$  e  $f(x_\ell; \beta) = x_\ell^\top \beta$  obtemos a importante classe dos modelos lineares generalizados; ver McCullagh e Nelder (1989) e Firth

(1) Segundo Houaiss (2001), “homoscedástico” é um adjetivo que significa “que apresenta a mesma variação ou dispersão” ou ainda “que tem a mesma variância”. No modelo normal, os dois significados se equivalem uma vez que a medida usual de dispersão é a própria variância. No entanto, nos modelos não lineares da família exponencial, homoscedasticidade indicará homogeneidade de dispersão (não necessariamente variância).

(2) Analogamente, dizemos que o modelo é heteroscedástico se os parâmetros de dispersão não são iguais para todas as observações.

(1991). Em particular, se  $g(\mu_\ell) = \mu_\ell$  com  $b(\theta_\ell) = \theta_\ell^2/2$ , então (3.1)–(3.2) definem o modelo de regressão normal não-linear. Há uma grande variedade de trabalhos com aplicações de modelos normais não-lineares. Vários exemplos são apresentados em Ratkowsky (1983, 1990), Bates e Watts (1988), Seber e Wild (1989).

Assim como Smyth (1989) e Cordeiro e McCullagh (1991), assumimos que  $e(y_\ell; \phi_\ell)$  em (3.1) é escrita como

$$e(y_\ell; \phi_\ell) = s(\phi_\ell) + t(y_\ell).$$

Estamos aqui considerando somente as distribuições contínuas da família exponencial, tais como gama, normal e normal inversa. Para estas distribuições, a função  $e(\cdot; \cdot)$  tem a forma acima e, tipicamente, as aplicações envolvendo estas distribuições assumem dispersão desconhecida. Então, podemos reescrever (3.1) como

$$\pi(y_\ell; \theta_\ell, \phi_\ell) = \exp\left(-\frac{1}{2}\{\phi_\ell d(y_\ell) + s(\phi_\ell) + t(y_\ell)\}\right), \quad \ell = 1, \dots, n, \quad (3.3)$$

sendo  $d(y_\ell) = d_\ell = -2[y_\ell\theta_\ell - b(\theta_\ell) - c(y_\ell)]$ , que corresponde a um modelo na família exponencial de distribuições na forma natural com parâmetros canônicos  $\phi_\ell$  e  $\phi_\ell\theta_\ell$ . Supomos que  $s(\phi_\ell)$  tem as quatro primeiras derivadas. Para as distribuições normal e normal inversa, temos  $d_\ell = (y_\ell - \mu_\ell)^2$  e  $d_\ell = (y_\ell - \mu_\ell)^2/\mu_\ell^2 y_\ell$ , respectivamente, e para ambas as distribuições  $s(\phi_\ell) = -\log(\phi_\ell)$ . Para a distribuição gama, temos  $s(\phi_\ell) = -2\{\phi_\ell \log(\phi_\ell) - \log\Gamma(\phi_\ell)\}$  e  $d_\ell = 2\{y_\ell/\mu_\ell - \log(y_\ell/\mu_\ell)\}$ .

### 3.3 Alguns testes de heteroscedasticidade

O interesse aqui reside testar a hipótese nula  $\mathcal{H}_0 : \delta = \delta_0$  (homoscedasticidade) contra a hipótese alternativa bilateral  $\mathcal{H}_1 : \delta \neq \delta_0$ , em que  $\delta_0$  é um vetor de dimensão  $p \times 1$  de constantes especificado tal que  $m(z_\ell, \delta_0) = 1$  para  $\ell = 1, \dots, n$ . Assim, os números de parâmetros de interesse e de perturbação são  $p$  e  $(k + 1)$ , respectivamente.

Seja  $L = L(\delta, \beta, \sigma^2)$  o logaritmo da função de verossimilhança de um modelo não-linear na família exponencial definido por (3.2) e (3.3) em algum espaço paramétrico, dado o vetor de observações  $y = (y_1, \dots, y_n)$ . Assume-se que a função

$L$  seja regular com respeito às derivadas em relação aos parâmetros até quarta ordem (Cox e Hinkley, 1974, Capítulo 9).

O logaritmo da função de verossimilhança do vetor de parâmetros  $\omega = (\delta^\top, \beta^\top, \sigma^2)^\top$ , dado o vetor de observações  $(y_1, \dots, y_n)$ , do modelo definido por (3.3) é

$$L(\delta, \beta, \sigma^2; y) = -\frac{1}{2} \sum_{\ell=1}^n \left\{ \frac{1}{\sigma^2 m_\ell} d_\ell + t(y_\ell) + s(\phi_\ell) \right\}, \quad \ell = 1, \dots, n.$$

A função escore total  $U = U(\delta, \beta, \sigma^2) = (U_\delta^\top, U_\beta^\top, U_{\sigma^2})^\top$  tem componentes (ver Apêndice B) dadas por

$$U_\delta = \frac{\partial L(\delta, \beta, \sigma^2; y)}{\partial \delta} = \frac{1}{2} \Psi^\top \Phi (d + \dot{S}), \quad U_\beta = \frac{\partial L(\delta, \beta, \sigma^2; y)}{\partial \beta} = X^{*\top} \Phi T V^{-1} (y - \mu)$$

e

$$U_{\sigma^2} = \frac{\partial L(\delta, \beta, \sigma^2; y)}{\partial \sigma^2} = \frac{1}{2\sigma^2} \iota^\top \Phi (d + \dot{S}),$$

respectivamente, em que  $\Phi$ ,  $T$  e  $V$  são matrizes diagonais de dimensão  $n \times n$  cujos respectivos elementos são dados por  $\phi_\ell = 1/(\sigma^2 m_\ell)$ ,  $T_\ell = (d\mu/d\eta)_\ell$  e  $V_\ell = (d\mu/d\theta)_\ell$ ,  $\Psi$  é uma matriz  $n \times p$  com a  $\ell$ -ésima linha dada por  $\partial \log m(z_\ell, \delta) / \partial \delta^\top$ ,  $\dot{S} = (\dot{s}(\phi_1), \dots, \dot{s}(\phi_n))^\top$  sendo  $\dot{s}(\phi_\ell) = ds(\phi_\ell)/d\phi_\ell$ ,  $\ell = 1, \dots, n$ ,  $\iota$  é um vetor de uns de dimensão  $n$ ,  $d = (d_1, \dots, d_n)^\top$  e  $\mu = (\mu_1, \dots, \mu_n)^\top$ .

A matriz de informação total de Fisher (ver Apêndice B) é dada por

$$I = -E \left( \frac{\partial^2 L}{\partial \omega \partial \omega^\top} \right) = \begin{pmatrix} \frac{1}{2} \Psi^\top \Phi \ddot{S} \Phi \Psi & 0 & \frac{1}{2\sigma^2} \Psi^\top \ddot{S} \Phi^2 \iota \\ 0 & X^{*\top} \Phi^{1/2} W \Phi^{1/2} X^* & 0 \\ \frac{1}{2\sigma^2} \iota^\top \Phi^2 \ddot{S} \Psi & 0 & \frac{1}{2\sigma^4} \iota^\top \Phi^2 \ddot{S} \iota \end{pmatrix},$$

em que  $\ddot{S}$  e  $W$  são matrizes diagonais de dimensão  $n \times n$  cujos respectivos elementos são dados por  $\ddot{s}(\phi_\ell) = d^2 s(\phi_\ell) / d\phi_\ell^2$  e  $w_\ell = (d\mu/d\eta)_\ell^2 / V_\ell$ ,  $\ell = 1, \dots, n$  e 0 denota matrizes ou vetores nulos de dimensões apropriadas.

O logaritmo da função de verossimilhança perfilada para o parâmetro de interesse  $\delta$  é dado por

$$L_p(\delta; y) = L(\delta, \hat{\beta}_\delta, \hat{\sigma}_\delta^2; y), \quad (3.4)$$

em que  $\widehat{\beta}_\delta$  e  $\widehat{\sigma}_\delta^2$  são os estimadores de máxima verossimilhança de  $\beta$  e  $\sigma^2$ , respectivamente, para  $\delta$  fixado. Sob condições de regularidade,  $\widehat{\beta}_\delta$  é a solução da equação  $U_\beta = 0$  e  $\widehat{\sigma}_\delta^2$  é a solução da equação  $U_{\sigma^2} = 0$ . Como, neste caso, não é possível expressar os estimadores de máxima verossimilhança restritos em uma forma fechada, é necessário utilizar técnicas iterativas de otimização restrita; ver Nocedal e Wright (1999, Capítulo 18) e Luenberger (1973, Capítulo 12).

O estimador de máxima verossimilhança (EMV) de  $\delta$  pode ser obtido maximizando o logaritmo da função de verossimilhança dado em (3.4) sujeito às restrições  $U_\beta = 0$  e  $U_{\sigma^2} = 0$ .

A estatística da razão de verossimilhanças para o teste de  $\mathcal{H}_0$  pode ser escrita da seguinte forma

$$LR = -2\{L_p(\delta_0; y) - L_p(\widehat{\delta}; y)\},$$

em que  $\widehat{\delta}$  é o EMV de  $\delta$ . Assintoticamente e sob a hipótese nula, a estatística da razão de verossimilhanças ( $LR$ ) tem distribuição  $\chi_p^2$ . No entanto, a aproximação resultante pode não ser boa para amostras pequenas ou mesmo de tamanho moderado, conduzindo a taxas de rejeição bastante distorcidas.

Nosso objetivo é determinar uma estatística da razão de verossimilhanças baseada na função de verossimilhança perfilada modificada de Cox e Reid (1987) (e sua correspondente correção de Bartlett) para o teste de  $\mathcal{H}_0 : \delta = \delta_0$  (homoscedasticidade). A modificação de Cox e Reid requer, no entanto, que o parâmetro de interesse,  $\delta$ , seja ortogonal aos demais. Da matriz de informação de Fisher,  $I$ , segue que  $(\delta^\top, \sigma^2)^\top$  é ortogonal a  $\beta$ , entretanto  $\delta$  e  $\sigma^2$  não são ortogonais. É necessário obter uma transformação  $(\delta^\top, \beta^\top, \sigma^2)^\top \longrightarrow (\delta^\top, \beta^\top, \gamma)^\top$  de modo que  $E(-\partial^2 L / \partial \delta^a \partial \gamma) = 0$ ,  $a = 1, \dots, p$ . Seguindo Cox e Reid (1987, 1993) e Wei (1998, equação (6.72)) chega-se à transformação desejada através de uma solução do sistema de equações diferenciais

$$\frac{\partial \sigma^2}{\partial \delta^\top} = -i_{\sigma^2 \sigma^2}^{-1} i_{\sigma^2 \delta}. \quad (3.5)$$

Somente em alguns casos especiais é viável a obtenção de solução do sistema de equações diferenciais acima, a saber, nas distribuições normal e normal inversa.

Daqui até o final deste capítulo, consideraremos apenas essas duas distribuições. Para essas distribuições, vimos que  $s(\phi_\ell) = -\log(\phi_\ell)$ ,  $\ell = 1, \dots, n$ , e, portanto, temos que  $\ddot{S} = \Phi^{-2}$ . Então, a matriz de informação total de Fisher reduz-se a

$$I = -E \left( \frac{\partial^2 L}{\partial \omega \partial \omega^\top} \right) = \begin{pmatrix} \frac{1}{2} \Psi^\top \Psi & 0 & \frac{1}{2\sigma^2} \Psi^\top \iota \\ 0 & X^{*\top} \Phi^{1/2} W \Phi^{1/2} X^* & 0 \\ \frac{1}{2\sigma^2} \iota^\top \Psi & 0 & \frac{n}{2\sigma^4} \end{pmatrix}.$$

Em particular, no caso normal temos que  $W = I_n$ , a matriz identidade de ordem  $n$ , e, portanto,  $I$  coincide com a expressão (2.3) da matriz de informação de Fisher do modelo normal linear heteroscedástico (2.1) (ver Seção 2.2), como esperado. Para as distribuições normal e normal inversa, o sistema de equações (3.5) fica dado por

$$\frac{n}{2\sigma^4} \frac{\partial \sigma^2}{\partial \delta^a} = -\frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{\partial m_\ell}{\partial \delta^a} \frac{1}{m_\ell}, \quad a = 1, \dots, p,$$

e tem como uma possível solução

$$\sigma^2 = \frac{\gamma}{(\prod_{\ell=1}^n m_\ell)^{1/n}}.$$

O logaritmo da função de verossimilhança para o modelo reparametrizado é dado por

$$L^* = L^*(\delta, \beta, \gamma; y) = -\frac{1}{2} \sum_{\ell=1}^n \left\{ \frac{q_\ell}{\gamma} d_\ell + t(y_\ell) + \log \left( \frac{\gamma}{q_\ell} \right) \right\}, \quad \ell = 1, \dots, n,$$

em que

$$q_\ell = q_\ell(\delta) = \frac{(\prod_{s=1}^n m_s)^{1/n}}{m_\ell}. \quad (3.6)$$

O logaritmo da função de verossimilhança perfilada correspondente para  $\delta$  é

$$L_p^* = L_p^*(\delta; y) = -\frac{n}{2} \log \hat{\gamma}_\delta - \frac{1}{2\hat{\gamma}_\delta} (y - \hat{\mu})^\top Q \hat{N}^{-1} (y - \hat{\mu}) + \text{constante},$$

em que  $\hat{\gamma}_\delta = \hat{\sigma}_\delta^2 (\prod_{\ell=1}^n m_\ell)^{1/n}$ ,  $Q = Q(\delta)$  é uma matriz diagonal de dimensão  $n \times n$  cujo  $\ell$ -ésimo elemento da diagonal é dado por  $\{\prod_{j=1}^n m(z_j, \delta)\}^{1/n} / m(z_\ell, \delta)$ ,

$N = \text{diag}(\mu_1^2 y_1, \dots, \mu_n^2 y_n)$  para a distribuição normal inversa e  $N = I_n$  para a distribuição normal.

O logaritmo da função de verossimilhança perfilada modificada (Cox e Reid, 1987) para  $\delta$  é

$$L_{mp}^*(\delta; y) = L_p^*(\delta; y) - \frac{1}{2} \log[\det\{j^*(\delta; \hat{\beta}_\delta, \hat{\gamma}_\delta)\}],$$

em que  $j^*$  é o bloco da matriz de informação observada para os parâmetros de perturbação  $(\beta^\top, \gamma)^\top$  no modelo reparametrizado avaliado em  $(\delta, \hat{\beta}_\delta^\top, \hat{\gamma}_\delta)^\top$ . Aqui obtemos

$$j^*(\delta; \hat{\beta}_\delta, \hat{\gamma}_\delta) = \begin{pmatrix} \frac{1}{\hat{\gamma}_\delta} \{X^{*\top} W^{1/2} Q(\delta) W^{1/2} X^* - [(y - \mu)^\top Q(\delta)] [X^{**}]\} & 0 \\ 0 & \frac{n}{2\hat{\gamma}_\delta^2} \end{pmatrix},$$

em que  $X^{**} = X^{**}(\beta) = \partial^2 \eta / \partial \beta \partial \beta^\top$  é um “array” de dimensão  $n \times k \times k$  e  $[\cdot][\cdot]$  representa multiplicação de uma matriz por um “array”, chamado de produto colchete e definido em Wei (1998, p. 188), Bates e Watts (1988) e Seber e Wild (1989, p. 691); ver Apêndice C. Como uma alternativa para implementação computacional, podemos escrever  $j^*(\delta; \hat{\beta}_\delta, \hat{\gamma}_\delta)$  como

$$j^*(\delta; \hat{\beta}_\delta, \hat{\gamma}_\delta) = \begin{pmatrix} \frac{1}{\hat{\gamma}_\delta} R^\top(\delta) (I_k - B(\delta)) R(\delta) & 0 \\ 0 & \frac{n}{2\hat{\gamma}_\delta^2} \end{pmatrix},$$

em que  $R(\delta)$  é uma matriz triangular superior que vem da seguinte decomposição de Cholesky

$$X^{*\top} W^{1/2} Q(\delta) W^{1/2} X^* = R^\top(\delta) R(\delta)$$

e

$$B(\delta) = \{[(y - \mu)^\top Q(\delta)] [(R^{-1}(\delta))^\top X^{**} R^{-1}(\delta)]\}_{(\delta; \hat{\beta}_\delta, \hat{\gamma}_\delta)}.$$

Observe que para a ligação canônica, temos que  $W = V$ . Então,  $L_{mp}^*$  pode ser escrita da seguinte forma:

$$L_{mp}^*(\delta; y) = L_p^*(\delta; y) - \frac{1}{2} \log \left( \frac{n}{2\hat{\gamma}_\delta^{k+2}} \det\{R^\top(\delta) (I_k - B(\delta)) R(\delta)\} \right).$$

A estatística da razão de verossimilhanças baseada na função de verossimilhança perfilada modificada para o teste de  $\mathcal{H}_0$  contra  $\mathcal{H}_1$  é dada por

$$LR_m = -2\{L_{mp}^*(\delta_0; y) - L_{mp}^*(\tilde{\delta}; y)\}, \quad (3.7)$$

em que  $\tilde{\delta}$  é valor de  $\delta$  que maximiza a função  $L_{mp}^*(y; \delta)$ . Note que  $LR_m$  envolve maximização da função de verossimilhança perfilada modificada. Pode-se mostrar que a estatística resultante pode ser escrita como

$$LR_m = \frac{n - k - 2}{n} LR + \log \left( \frac{\det\{R^\top(\delta_0)(I_k - B(\delta_0))R(\delta_0)\}}{\det\{R^\top(\tilde{\delta})(I_k - B(\tilde{\delta}))R(\tilde{\delta})\}} \right).$$

Para o modelo de regressão normal linear (caso em que  $B(\delta) = 0$ ), a expressão  $LR_m$  dada acima é consistente com a expressão de  $LR_m$  apresentada na Seção 2.2, como esperado.

Wei (1998, Teorema 6.17, eq. (6.90)) propôs uma estatística alternativa que evita tal maximização substituindo  $\tilde{\delta}$  por  $\hat{\delta}$  na expressão para  $LR_m$ , obtendo-se assim a mesma estatística de teste dada acima, mas com  $R(\tilde{\delta})$  substituído por  $R(\hat{\delta})$ .

### 3.4 Fator de correção de Bartlett

As estatísticas da razão de verossimilhanças ( $LR$ ) e razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) dadas na Seção 3.3 têm, em problemas regulares, distribuição  $\chi_p^2$  assintoticamente, em que  $p$  é o número de restrições impostas sob a hipótese nula. A aproximação pela distribuição  $\chi_p^2$  para as distribuições de  $LR$  e  $LR_m$  tem erro de ordem  $n^{-1}$ . Nosso objetivo agora é desenvolver uma correção de Bartlett para a estatística  $LR_m$  de tal forma que a aproximação por  $\chi_p^2$  para a estatística corrigida seja de ordem  $n^{-2}$ .

A suposição de uma estrutura não-linear no componente sistemático dos modelos aumenta consideravelmente a complexidade do cálculo dos cumulantes das derivadas do logaritmo da função de verossimilhança, em relação aos modelos com estrutura linear e, em particular, dificulta enormemente o desenvolvimento das correções de Bartlett para a estatística da razão de verossimilhanças ( $LR$ ). Surpreendentemente, nenhuma dificuldade adicional é observada na obtenção da correção

de Bartlett para  $LR_m$ . Assim, a partir dos resultados de DiCiccio e Stern (1994) encontraremos um fator de correção de Bartlett  $c_m$  que define uma estatística da razão de verossimilhanças perfiladas modificadas corrigida dada por

$$LR_m^* = \frac{LR_m}{1 + c_m/p} \quad (3.8)$$

de tal forma que, em problemas regulares,  $P(LR_m^* > x_\alpha) = \alpha + O(n^{-2})$ , em que  $x_\alpha$  é o quantil  $(1 - \alpha)$  da distribuição  $\chi_p^2$ , enquanto que  $P(LR_m > x_\alpha) = \alpha + O(n^{-1})$ .

As notações referentes aos cumulantes de derivadas do logaritmo da função de verossimilhança, às derivadas dos cumulantes, às funções dos cumulantes e à matriz de informação de Fisher são as mesmas empregadas no Capítulo 2, Seção 2.3.

No Capítulo 2, derivamos uma expressão geral para o fator de correção de Bartlett ( $c_m$ ), que pode ser aplicada à estatística da razão de verossimilhanças perfiladas modificada em diferentes classes de modelos. A expressão de  $c_m$  é dada na Seção 2.3, equação (2.8), e pode ser usada para obter expressões em forma fechada para o fator de correção de Bartlett em modelos em que há a ortogonalidade dos parâmetros e que utilizam a mesma partição do vetor de parâmetros considerada aqui.

Pode ser mostrado facilmente que

$$\frac{\partial L^*}{\partial \delta^a} = -\frac{1}{2} \sum_{\ell=1}^n \left[ \frac{q_{\ell_a}}{\gamma} d_\ell - \frac{q_{\ell_a}}{q_\ell} \right],$$

sendo  $q_{\ell_a} = \partial q_\ell / \partial \delta^a$  e  $q_\ell$  é dado em (3.6). Mas

$$\begin{aligned} \sum_{\ell=1}^n \frac{q_{\ell_a}}{q_\ell} &= \sum_{\ell=1}^n \frac{\partial}{\partial \delta^a} \log q_\ell = \frac{\partial}{\partial \delta^a} \left\{ \sum_{\ell=1}^n \log q_\ell \right\} = \frac{\partial}{\partial \delta^a} \left\{ \sum_{\ell=1}^n \log \left[ \frac{(\prod_{s=1}^n m_s)^{1/n}}{m_\ell} \right] \right\} \\ &= \frac{\partial}{\partial \delta^a} \left\{ \sum_{\ell=1}^n \left[ \frac{1}{n} \sum_{s=1}^n \log m_s - \log m_\ell \right] \right\} = \frac{\partial}{\partial \delta^a} \left\{ \sum_{s=1}^n \log m_s - \sum_{\ell=1}^n \log m_\ell \right\} \\ &= \frac{\partial}{\partial \delta^a} \{0\} = 0, \end{aligned}$$

e, portanto,

$$\frac{\partial L^*}{\partial \delta^a} = -\frac{1}{2\gamma} \sum_{\ell=1}^n q_{\ell_a} d_\ell.$$

Além disso,

$$E(d_\ell) = E(-2[y_\ell\theta_\ell - b(\theta_\ell) - c(y_\ell)]) = -\dot{s}\left(\frac{q_\ell}{\gamma}\right) = -\left(-\frac{\gamma}{q_\ell}\right) = \frac{\gamma}{q_\ell}$$

e, conseqüentemente,  $E(\partial L^*/\partial\delta^a) = 0$ . Temos também que

$$\frac{\partial^2 L^*}{\partial\delta^a\partial\gamma} = \frac{1}{2\gamma^2} \sum_{\ell=1}^n q_{\ell a} d_\ell$$

e

$$\lambda_{a\gamma} = E\left(\frac{\partial^2 L^*}{\partial\delta^a\partial\gamma}\right) = \frac{1}{2\gamma} \sum_{\ell=1}^n \frac{q_{\ell a}}{q_\ell} = 0,$$

de modo que a ortogonalidade entre  $\delta$  e  $\gamma$  é verificada, como esperado.

Podemos observar que a expressão de  $\partial L^*/\partial\delta^a$  dada acima coincide com a do modelo normal linear (ver Apêndice A, equação (A.1)). Portanto, é fácil verificar que as quantidades  $\lambda_{ab}$ ,  $\lambda_{abc}$ ,  $\lambda_{abcd}$ ,  $\lambda_{ab\gamma}$ ,  $(\lambda_{ab\gamma})_\gamma$ ,  $(\lambda_{abc})_d$ ,  $(\lambda_{ab})_{cd}$ ,  $(\lambda_{ab})_c$  e  $\lambda_{iab}$  para os modelos não-lineares normal e normal inverso coincidem com as quantidades correspondentes do modelo normal linear. Em particular,  $\lambda_{iab} = 0$  e, conseqüentemente, a obtenção de  $c_m$  pode ser feita a partir da equação (2.9). Esta envolve o cumulante  $\lambda_{\gamma\gamma}$  e sua derivada  $(\lambda_{\gamma\gamma})_\gamma$ . Temos

$$\frac{\partial L^*}{\partial\gamma} = -\frac{1}{2\gamma} \sum_{\ell=1}^n q_{\ell a} d_\ell - \frac{n}{\gamma},$$

que coincide com a expressão correspondente para o modelo normal linear (ver Apêndice A, equação A.2), o mesmo ocorrendo com  $\lambda_{\gamma\gamma}$  e  $(\lambda_{\gamma\gamma})_\gamma$ . Conclui-se que o fator de correção de Bartlett  $c_m$  obtido para o modelo normal linear é o mesmo que para os modelos não-lineares normal e normal inverso. Desta forma, se  $m_\ell = \exp\{z_\ell^\top \delta\}$ , ou seja, se é admitida uma estrutura multiplicativa para a dispersão, temos

$$c_m = -\frac{1}{2}\text{tr}(H_d H_d) + \frac{1}{2n}p^2 + \frac{1}{2} \iota^\top H_d H H_d \iota + \frac{1}{3} \iota^\top H^{(3)} \iota - \frac{2}{n}p + \frac{1}{n} \iota^\top H^{(2)} \iota, \quad (3.9)$$

sendo  $H = \{h_{\ell s}\} = (Z - \bar{Z})[(Z - \bar{Z})^\top(Z - \bar{Z})]^{-1}(Z - \bar{Z})^\top$ , com  $(Z - \bar{Z}) = (z_1 - \bar{z}, \dots, z_n - \bar{z})^\top$ ,  $\bar{z} = (\bar{z}_1, \dots, \bar{z}_p)^\top$ ,  $\bar{z}_a = n^{-1} \sum_{\ell=1}^n z_{\ell a}$  para  $a = 1, \dots, p$ ,  $H_d = \text{diag}\{h_{11}, \dots, h_{nn}\}$ ,  $H^{(2)} = (h_{\ell s}^2)$  e  $H^{(3)} = (h_{\ell s}^3)$ .

Podemos observar que a expressão (3.9) coincide com a expressão (2.10) para o caso normal linear. É importante salientar que a expressão acima vale para a distribuição normal inversa também. Pode-se notar que a não linearidade no componente sistemático do modelo não influencia o fator de correção de Bartlett  $c_m$  para a estatística da razão de verossimilhanças perfiladas modificadas no teste de heteroscedasticidade. A expressão matricial do fator de correção de Bartlett em (3.9) é bastante simples uma vez que requer apenas operações com vetores e matrizes.

Pacotes de computação simbólica ou software que façam operações simples com matrizes, como Ox, S-Plus, R, MATHEMATICA ou MAPLE, podem ser utilizados para o cálculo de  $c_m$  em aplicações práticas. É importante observar também que a expressão de  $c_m$  em (3.9) não depende do número de parâmetros de perturbação, nem depende de qualquer parâmetro desconhecido, sendo função apenas do número de parâmetros de interesse, do número de observações e da matriz  $H$  através da matriz  $Z$  de variáveis auxiliares usadas para modelar as flutuações na dispersão da resposta. Vale lembrar aqui que aplicamos o fator de correção de Bartlett à forma exata da estatística da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ), que envolve a maximização da função de verossimilhança perfilada modificada ( $L_{mp}^*$ ), e esperamos que o teste baseado na estatística da razão de verossimilhanças perfiladas modificadas corrigida produza inferências mais precisas que o teste da razão de verossimilhanças original.

### 3.5 Resultados numéricos

Nesta seção, apresentamos alguns resultados de simulação comparando o desempenho dos testes da razão de verossimilhanças original ( $LR$ ), da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) e da sua versão corrigida ( $LR_m^*$ ). Estes desempenhos são avaliados em função da proximidade das probabilidades de rejeição da hipótese nula, sendo esta verdadeira (probabilidade do erro tipo I), aos respectivos níveis nominais dos testes. As simulações realizadas são baseadas no

modelo de regressão não-linear

$$y_\ell = \beta_1 + \exp\{\beta_2 x_{\ell 2}\} + \sum_{j=3}^k \beta_j x_{\ell j} + u_\ell, \quad \ell = 1, \dots, n,$$

em que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{\delta_1 z_{\ell 1} + \dots + \delta_p z_{\ell p}\})$ ,  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$ . A hipótese nula considerada é  $\mathcal{H}_0 : \delta_1 = \dots = \delta_p = 0$  indicando assim que o modelo é homoscedástico. A variável resposta foi gerada assumindo que  $\beta_1 = \dots = \beta_k = 1$ ,  $\sigma^2 = 1$  e diferentes valores de  $p$  e  $k$  foram considerados. As covariadas  $x_2, \dots, x_k$  foram geradas como amostras aleatórias da distribuição  $\mathcal{U}(0, 1)$ . A matriz  $Z$  foi construída usando as colunas  $2, \dots, p$  da matriz  $X$ . O número de réplicas foi fixado em 10.000 e foram considerados os seguintes níveis nominais:  $\alpha=10\%$ ,  $5\%$ ,  $1\%$ ,  $0,5\%$ . As simulações foram realizadas usando a linguagem de programação matricial `0x` (Doornik, 2001).

Para cada tamanho de amostra e cada nível considerado, calculamos as taxas de rejeição de cada teste, isto é, estimamos, via simulação,  $P(LR \geq x_\alpha)$ ,  $P(LR_m \geq x_\alpha)$  e  $P(LR_m^* \geq x_\alpha)$ , em que  $x_\alpha$  é o quantil  $(1 - \alpha)$  apropriado da distribuição  $\chi_p^2$ . Todas as entradas das tabelas apresentadas são porcentagens.

A Tabela 3.1 apresenta resultados de simulação dos tamanhos dos testes para  $n = 35$ ,  $p = 2$  e diferentes valores para  $k$ . Variamos  $k$  para analisar o efeito do número de parâmetros de perturbação nos diferentes testes. Observamos que o teste da razão de verossimilhanças original é muito liberal, apresentando taxas de rejeição bem maiores do que os níveis nominais, especialmente quando o número de parâmetros de perturbação aumenta. Por exemplo, quando  $k = 5$  e  $\alpha = 5\%$ , a taxa de rejeição do teste é superior a  $10\%$ , o dobro do nível nominal. É interessante notar que o impacto do número de parâmetros de perturbação é bem menos marcante no caso do teste da razão de verossimilhanças perfiladas modificadas e sua versão corrigida. Para estes testes, as taxas de rejeição permanecem mais estáveis em relação aos respectivos níveis nominais do que o teste da razão de verossimilhanças original. O teste da razão de verossimilhanças perfiladas modificadas tende a corrigir a tendência liberal do teste original e exhibe taxas de rejeição menores que o nível

nominal correspondente. Novamente, quando  $k = 5$  e  $\alpha = 5\%$ , sua taxa de rejeição é 4,5%. A correção de Bartlett aplicada a este teste leva as taxas de rejeição para valores mais próximos do tamanho nominal do teste. Para a situação acima, a taxa de rejeição do teste da razão de verossimilhanças perfiladas modificadas corrigido sob  $\mathcal{H}_0$  é 5,0%. As conclusões anteriores também se verificam para outros níveis nominais, até mesmo para níveis nominais muito pequenos.

A Tabela 3.2 contém resultados de simulação obtidos para o caso em que  $n = 35$ ,  $k = 5$  e  $p = 1, \dots, 5$ . Mais uma vez, o teste da razão de verossimilhanças original apresentou-se muito liberal, especialmente quando o número de parâmetros que definem o comportamento de heteroscedasticidade é grande; por exemplo, quando  $p = 5$  e  $\alpha = 5\%$ , a taxa de rejeição do teste é 18,8%, ou seja, aproximadamente 4 vezes maior que o nível nominal selecionado. Novamente o impacto do número de parâmetros de interesse é bem menos marcante no caso do teste da razão de verossimilhanças perfiladas modificadas e sua versão corrigida relativamente ao teste da razão de verossimilhanças original. Ainda na situação em que  $p = 5$  e  $\alpha = 5\%$ , a taxa de rejeição do teste da razão de verossimilhanças perfiladas modificadas e de sua versão corrigida são 4,5% e 5,0%, respectivamente. As outras conclusões obtidas a partir da Tabela 3.1 também se verificam aqui.

A Tabela 3.3 apresenta os tamanhos dos testes para diversos tamanhos da amostra ( $n = 25, 30, 35, 40$  e  $45$ ), fixando-se o valor de  $p$  em 2 para  $k = 2$  e 5. Observamos que as taxas de rejeição de todos os três testes aproximam-se dos níveis nominais com o aumento do tamanho da amostra, conforme esperado. Destacamos também o fato de que o teste da razão de verossimilhanças original apresenta taxas de rejeição sob  $\mathcal{H}_0$  bem acima dos níveis nominais considerados. No caso em que  $n = 25$ ,  $k = 5$  e  $\alpha = 5\%$ , por exemplo, a taxa de rejeição do teste é 16,0%, isto é, mais do que 3 vezes maior que o nível nominal selecionado. O teste mantém-se liberal até mesmo para  $n = 45$ . O teste da razão de verossimilhanças perfiladas modificadas foi eficaz em reduzir as taxas de rejeição do teste da razão de verossimilhanças original exibindo taxas de rejeição menores que o nível nominal considerado. A aplicação da correção de Bartlett a esse teste produziu resultados bastante favoráveis, de

forma que as taxas de rejeição são próximas do nível nominal. Como uma ilustração, quando  $k = 5$ ,  $\alpha = 5\%$  e  $n = 45$ , as taxas de rejeição do teste da razão de verossimilhanças, do teste da razão de verossimilhanças perfiladas modificadas e de sua versão corrigida são, respectivamente, 9,6%, 4,4% e 4,9%.

A Tabela 3.4 apresenta resultados de simulação obtidos levando em consideração a hipótese alternativa (heteroscedasticidade) para  $n = 35$ ,  $p = 2$ ; diferentes valores de  $\delta_1 = \delta_2 = \delta$  com  $\delta$  variando entre 0,0 e 3,5 foram considerados ao nível  $\alpha = 10\%$ . É importante observar que estas simulações de poder correspondem à situação abordada na Tabela 3.1 para  $n = 35$ . Foi desconsiderado o teste da razão de verossimilhanças original pois, nas simulações de tamanho, mostrou-se liberal e, portanto, não pode ser recomendado. Assim, somente comparamos o poder do teste da razão de verossimilhanças perfiladas modificadas e da sua versão corrigida. Os resultados indicam que não há nenhuma perda de poder derivada do fato de usar o fator de correção de Bartlett. O poder dos dois testes é semelhante, com uma leve vantagem do teste corrigido, vantagem esta que se origina de sua menor distorção de tamanho.

Concluimos destes resultados que o teste da razão de verossimilhanças original pode ser muito liberal levando à rejeição de  $\mathcal{H}_0$  com uma frequência muito maior do que esperado. Por outro lado, o teste da razão de verossimilhanças perfiladas modificadas apresenta taxas de rejeição mais próximas dos níveis nominais mas ainda exibe alguma distorção de tamanho. O fator de correção de Bartlett para este teste é eficaz em reduzir ainda mais esta distorção. Dos três testes avaliados, o teste da razão de verossimilhanças perfiladas modificadas corrigido é o mais recomendável.

Tabela 3.1 Tamanho dos testes - Modelo normal não-linear com  $n = 35$ ,  $p = 2$ , diversos valores para  $k$ , covariáveis uniformes e erros normais.

$k$	$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$			$\alpha = 0,5\%$		
	LR	LR <sub>m</sub>	LR* <sub>m</sub>	LR	LR <sub>m</sub>	LR* <sub>m</sub>	LR	LR <sub>m</sub>	LR* <sub>m</sub>	LR	LR <sub>m</sub>	LR* <sub>m</sub>
2	13,2	9,0	9,8	6,7	3,9	4,4	1,6	0,7	0,9	0,8	0,3	0,4
3	14,4	9,1	10,0	8,1	4,3	4,9	2,0	0,8	0,9	1,1	0,4	0,5
4	17,2	8,9	9,8	9,9	4,2	4,9	2,9	0,8	0,9	1,8	0,5	0,5
5	20,2	9,4	10,0	10,7	4,5	5,0	3,0	0,9	1,0	1,8	0,4	0,5
6	22,6	9,5	10,3	14,9	4,6	5,2	5,4	0,8	1,0	3,5	0,4	0,5

Tabela 3.2 Tamanho dos testes - Modelo normal não-linear com  $n = 35$ ,  $k = 5$ , diversos valores para  $p$ , covariáveis uniformes e erros normais.

$p$	$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$			$\alpha = 0,5\%$		
	LR	LR <sub>m</sub>	LR* <sub>m</sub>	LR	LR <sub>m</sub>	LR* <sub>m</sub>	LR	LR <sub>m</sub>	LR* <sub>m</sub>	LR	LR <sub>m</sub>	LR* <sub>m</sub>
1	17,6	9,2	9,8	9,1	4,5	4,9	2,6	0,8	1,1	1,4	0,4	0,4
2	20,2	9,4	10,0	10,7	4,5	5,0	3,0	0,9	1,0	1,8	0,4	0,5
3	22,2	9,1	9,9	13,6	4,3	4,7	4,5	0,7	0,9	2,8	0,3	0,4
4	24,9	8,9	9,8	15,6	4,7	5,0	5,5	0,8	1,0	3,7	0,4	0,5
5	28,9	9,0	10,0	18,8	4,9	5,0	7,2	0,9	1,0	5,3	0,5	0,5

Tabela 3.3 *Tamanho dos testes - Modelo normal não-linear com  $p = 2$ ,  $k = 2, 5$ , covariáveis uniformes e erros normais.*

$n$	$\alpha(\%)$	$k = 2$			$k = 5$		
		$LR$	$LR_m$	$LR_m^*$	$LR$	$LR_m$	$LR_m^*$
25	10,0	12,7	8,9	9,8	24,6	7,0	8,2
	5,0	6,9	4,2	4,9	16,0	3,0	3,7
	1,0	1,8	0,9	1,1	5,9	0,4	0,5
	0,5	1,0	0,5	0,5	4,0	0,2	0,2
30	10,0	12,2	9,0	9,8	23,1	8,1	9,2
	5,0	6,3	4,2	4,7	14,3	3,8	4,3
	1,0	1,5	0,8	1,0	4,9	0,5	0,7
	0,5	0,8	0,4	0,5	3,0	0,2	0,3
35	10,0	11,6	9,3	9,8	20,2	8,6	9,5
	5,0	6,3	4,5	4,9	12,0	4,3	4,7
	1,0	1,4	0,9	1,0	4,0	0,7	0,8
	0,5	0,8	0,4	0,5	2,5	0,3	0,4
40	10,0	11,3	9,0	9,6	17,4	8,4	9,1
	5,0	5,7	4,3	4,7	10,2	4,1	4,5
	1,0	1,3	0,8	0,9	2,9	0,8	0,9
	0,5	0,6	0,4	0,4	1,7	0,3	0,4
45	10,0	11,3	9,4	10,0	16,4	9,2	9,9
	5,0	6,2	4,9	5,0	9,6	4,4	4,9
	1,0	1,5	0,9	1,0	2,6	0,7	0,9
	0,5	0,8	0,4	0,5	1,6	0,4	0,5

Tabela 3.4 *Poder dos testes - Modelo normal não-linear com  $n = 35$ ,  $p = 2$ ,  $\alpha = 10\%$ , covariáveis uniformes e erros normais.*

$\delta$	$k = 5$		$k = 8$	
	$LR_m$	$LR_m^*$	$LR_m$	$LR_m^*$
0,0	9,3	9,8	8,6	9,5
0,5	13,4	14,1	14,9	16,0
1,0	28,1	29,1	33,9	35,7
1,5	48,6	50,0	58,9	62,0
2,0	69,7	70,9	83,3	84,5
2,5	85,3	86,1	94,5	95,0
3,0	93,8	94,1	96,7	97,3
3,5	95,8	96,0	97,1	98,5

### 3.6 Aplicações

Apresentamos a seguir duas ilustrações numéricas dos testes apresentados nas seções anteriores.

#### 3.6.1 Exemplo 1

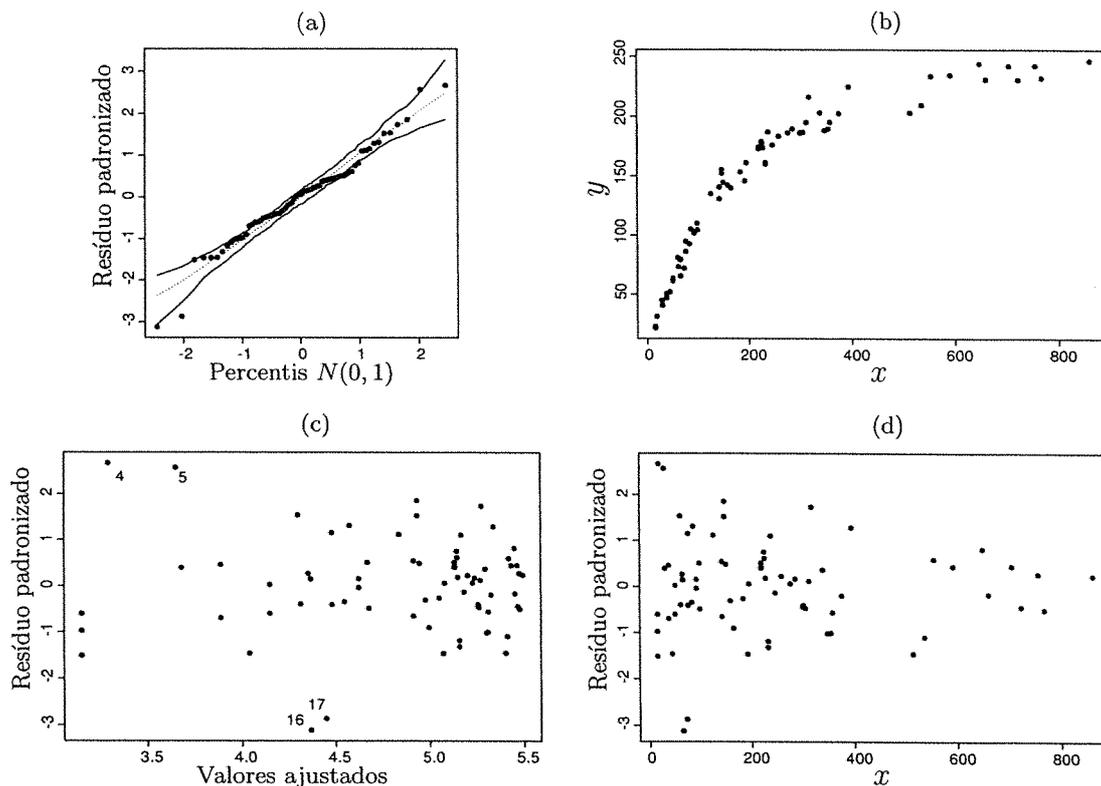
Os dados analisados por Wei (1998, p. 155) e apresentados no Apêndice H representam os pesos  $y$  das lentes dos olhos de coelhos europeus (*Oryctolagus Cuniculus*) (em mg) e à idade  $x$  do animal (em dias) numa amostra de 71 observações. O modelo usado é

$$\log(y_\ell) = \beta_1 - \frac{\beta_2}{x_\ell + \beta_3} + u_\ell, \quad \ell = 1, \dots, 71,$$

em que  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$ . Inicialmente, assumimos que  $u_\ell \sim \mathcal{N}(0, \sigma^2)$  e estimamos os parâmetros do modelo por máxima verossimilhança, obtendo  $\hat{\beta}_1 = 5,640$ ,  $\hat{\beta}_2 = 130,58$ ,  $\hat{\beta}_3 = 37,603$  e  $\hat{\sigma}^2 = 0,004$ . Os gráficos na Figura 3.1 mostram que as observações 4, 5, 16 e 17 apresentam resíduos grandes (em valor absoluto) e dão alguma evidência de heteroscedasticidade.

O próximo passo é considerar um modelo mais geral, heteroscedástico, em que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{\delta x_\ell\})$ . Agora o nosso principal interesse é testar  $\mathcal{H}_0 : \delta = 0$  (homoscedasticidade) contra  $\mathcal{H}_1 : \delta \neq 0$ . Para este teste, temos  $LR = 10,250$ ,  $LR_m = 9,556$ ,  $LR_m^* = 9,776$ , que conduzem aos respectivos níveis descritivos 0,001, 0,001 e 0,002. Segue então que os testes baseados nas estatísticas da razão de verossimilhanças ( $LR$ ), razão de verossimilhanças perfiladas modificada ( $LR_m$ ) e sua versão corrigida ( $LR_m^*$ ) conduzem à rejeição da hipótese nula aos níveis nominais usuais.

Figura 3.1 (a) Gráfico dos resíduos padronizados contra percentis normais com envelope simulado; (b) Gráfico de dispersão do peso das lentes dos olhos contra a idade; (c) Gráfico dos resíduos studentizados  $t_{s_i}$  contra os valores ajustados e (d) Gráfico dos resíduos studentizados  $t_{s_i}$  contra  $x$  para o modelo homoscedástico - Coelhos europeus na Austrália.



### 3.6.2 Exemplo 2

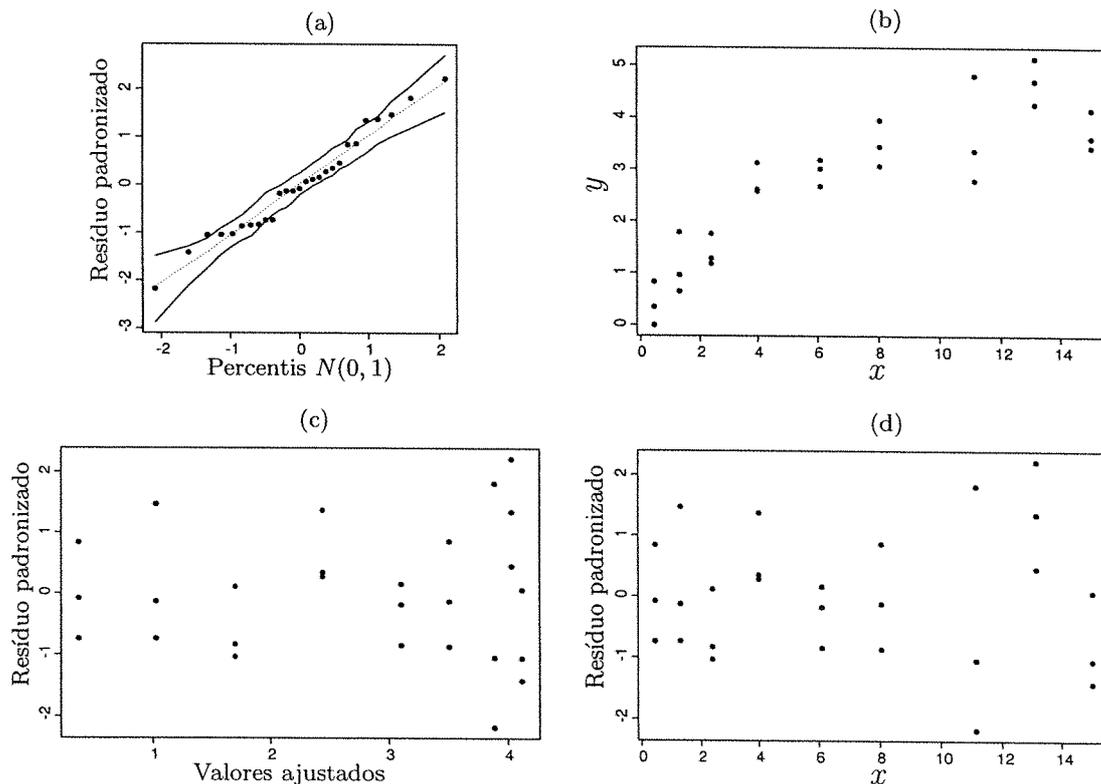
Consideremos os dados analisados por Rawlings (1988) e apresentados no Apêndice I. Howard Grimes, do Departamento de Botânica, North Carolina State University, conduziu um experimento para análise bioquímica de armazenamento intracelular e transporte de cálcio através da membrana plasmática. Células ficavam suspensas em uma solução de cálcio radioativo por um certo período de tempo  $x$  (em minutos) e então a quantidade de cálcio radioativo  $y$  (em nmoles/mg) absorvida pelas células

foi medida numa amostra de 27 observações. O modelo usado é

$$y_\ell = \beta_0\{1 - \exp(-\beta_1 x_\ell)\} + u_\ell, \quad \ell = 1, \dots, 27,$$

em que  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$ . Inicialmente, assumimos que  $u_\ell \sim \mathcal{N}(0, \sigma^2)$  e estimamos os parâmetros do modelo por máxima verossimilhanças obtendo  $\hat{\beta}_0 = 4,309$ ,  $\hat{\beta}_1 = 0,208$  e  $\hat{\sigma}^2 = 0,276$ . Os gráficos na Figura 3.2 mostram alguma evidência de heteroscedasticidade.

Figura 3.2 (a) Gráfico dos resíduos padronizados contra percentis normais com envelope simulado; (b) Gráfico de dispersão da quantidade absorvida de cálcio radioativo contra o tempo; (c) Gráfico dos resíduos studentizados  $t_{s_i}$  contra os valores ajustados e (d) Gráfico dos resíduos studentizados  $t_{s_i}$  contra  $x$  para o modelo homoscedástico - Solução de cálcio radioativo.



O próximo passo é considerar um modelo mais geral, heteroscedástico, em que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{\delta x_\ell\})$ . Agora o nosso principal interesse é testar  $\mathcal{H}_0 : \delta = 0$

(homoscedasticidade) contra  $\mathcal{H}_1 : \delta \neq 0$ . Para este teste, temos  $LR = 3,097$ ,  $LR_m = 2,914$ ,  $LR_m^* = 3,056$ , que conduzem aos respectivos níveis descritivos 0,078, 0,088 e 0,080. Segue então que os testes baseados nas estatísticas da razão de verossimilhanças ( $LR$ ), razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) e sua versão corrigida ( $LR_m^*$ ) levam à rejeição da hipótese nula ao nível nominal de 10% e sugerem evidência marginal de heteroscedasticidade.

### 3.7 Comentários

Trabalhamos neste capítulo com os modelos não-lineares da família exponencial para os quais abordamos a situação em que os parâmetros de dispersão não são constantes para todas as observações, havendo assim uma estrutura heteroscedástica. Admitimos uma forma funcional que relaciona os parâmetros de dispersão com alguns parâmetros desconhecidos, que não dependem do vetor de parâmetros de regressão, e algumas variáveis auxiliares.

Derivamos um fator de correção de Bartlett para a estatística da razão de verossimilhanças perfiladas modificadas do teste de heterogeneidade de dispersões (homoscedasticidade). Os resultados cobrem a situação em que o parâmetro de interesse é multidimensional. Notamos que uma possível não-linearidade no componente sistemático do modelo não influencia o fator de correção de Bartlett ( $c_m$ ). Vale salientar aqui que consideramos apenas as distribuições normal e normal inversa, pois para a distribuição gama o ajuste de Cox e Reid não pôde ser aplicado.

Os resultados numéricos apresentados comparam o desempenho em amostras finitas de três testes: o teste da razão de verossimilhanças original ( $LR$ ), o teste da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) e o teste da razão de verossimilhanças perfiladas modificadas corrigido ( $LR_m^*$ ). As simulações feitas conduziram a diversas conclusões importantes a respeito do comportamento, em amostras finitas, dos testes de heteroscedasticidade em estudo.

Em primeiro lugar, o estudo dos tamanhos dos testes levou a concluir que o teste baseado na estatística  $LR$  é, em geral, liberal, conduzindo à rejeição da hipótese de homoscedasticidade erroneamente com uma probabilidade maior do que o nível

de significância nominal do teste. Por outro lado, o teste baseado na estatística  $LR_m$  mostrou-se eficiente corrigindo a tendência do teste original em rejeitar com frequência demasiada a hipótese nula, exibindo taxas de rejeição menores que o nível nominal do teste. O teste baseado na estatística  $LR_m^*$  leva as taxas de rejeição para valores ainda mais próximos dos níveis nominais.

Em segundo lugar, em relação aos estudos de simulação realizados com o objetivo de analisar a influência dos números de parâmetros de perturbação e de interesse no desempenho dos testes, concluímos que os números de parâmetros de perturbação e de interesse têm um impacto considerável sobre a aproximação por  $\chi^2$  para a distribuição da estatística da razão de verossimilhanças original. É importante ressaltar que o mesmo impacto é bem menos marcante no caso dos testes baseados nas estatísticas  $LR_m$  e  $LR_m^*$ . Para estes testes, as taxas de rejeição permanecem mais estáveis em relação aos respectivos níveis nominais do que o teste baseado na estatística  $LR$ .

Nas simulações envolvendo os poderes dos testes, o teste baseado na estatística  $LR_m^*$  apresentou uma leve vantagem em relação ao teste baseado na estatística  $LR_m$ .

Em suma, entre as três estatísticas, aquela que produz melhor teste de heteroscedasticidade é  $LR_m^*$ .

---

## CAPÍTULO 4

# Novos ajustes para a estatística da razão de verossimilhanças em modelos não-lineares da família exponencial

---

### 4.1 Introdução

Os modelos não-lineares da família exponencial (Cordeiro e Paula, 1989) podem ser interpretados como uma generalização dos modelos lineares generalizados e dos modelos de regressão normais não-lineares. A motivação básica destes modelos é abrir o leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor não-linear. Nesta classe de modelos, mesmo se os parâmetros de dispersão forem iguais para todas as observações, suas variâncias poderão diferir, pois estas podem depender das médias desconhecidas. Assim, homoscedasticidade indicará homogeneidade de dispersão (não necessariamente variância). Admitimos para as variâncias uma estrutura que relacione os parâmetros de dispersão com alguns parâmetros desconhecidos que não dependem do vetor de parâmetros de regressão e algumas variáveis auxiliares.

No Capítulo 3, obtivemos a correção de Bartlett para um teste de heteroscedasticidade baseado em verossimilhança perfilada modificada (Cox e Reid, 1987) em modelos não-lineares da família exponencial. Notamos que o ajuste de Cox e Reid (1987) não pôde ser aplicado a todos os modelos não-lineares da família exponencial, como por exemplo, para o modelo gama.

Neste capítulo, tratamos da obtenção de ajustes para a estatística da razão de verossimilhanças ( $LR$ ) com base nos resultados de Skovgaard (2001) na classe dos modelos não-lineares da família exponencial. Estes ajustes têm ampla aplicabili-

dade, pois são válidos para testar quaisquer hipóteses nulas que estabeleçam um vetor de valores fixados para uma parte do vetor de parâmetros do modelo. Em particular, serão feitos ajustes para o teste de heteroscedasticidade. Estes ajustes podem ser utilizados, sem exceção, sob todos os modelos não-lineares da família exponencial. Estes novos ajustes apresentam uma vantagem em relação ao ajuste de Cox e Reid (1987) por serem invariantes segundo reparametrizações da forma  $(\nu, \psi) \longrightarrow (\nu, \zeta(\nu, \psi))$ , sendo  $\nu$  o vetor de parâmetros de interesse,  $\psi$  o vetor de parâmetros de perturbação e  $\zeta$  uma função de  $\nu$  e  $\psi$ . Outra vantagem é a aplicação direta à estatística da razão de verossimilhanças, enquanto que o ajuste de Cox e Reid, ao contrário, é aplicado à função de verossimilhança perfilada, a qual deve ser maximizada segundo as hipóteses nula e alternativa para, posteriormente, obter-se a estatística ajustada. Outra vantagem adicional é que não é necessária a ortogonalização entre os parâmetros de interesse e os de perturbação. Estes novos ajustes apresentam uma vantagem em relação ao fator de correção de Bartlett uma vez que são mais simples de serem obtidos pois requerem cálculos dos cumulantes de derivadas do logaritmo da função de verossimilhanças até segunda ordem, enquanto que o cálculo do fator de correção de Bartlett é mais complexo e requer a obtenção dos cumulantes de derivadas do logaritmo da função de verossimilhanças até quarta ordem.

Resultados de simulação em relação a tamanho e poder dos testes de heteroscedasticidade obtidos neste capítulo no caso em que o parâmetro de interesse é multidimensional são considerados. Adicionalmente, aspectos de robustez dos testes também foram avaliados.

Na Seção 4.2, apresentamos resultados gerais sobre ajustes para a estatística da razão de verossimilhanças. Na Seção 4.3, apresentamos a definição dos modelos não-lineares da família exponencial; um teste baseado na estatística da razão de verossimilhanças ajustada e sua versão assintoticamente equivalente são propostos nesta classe de modelos. Na Seção 4.4, são apresentados resultados numéricos sobre o comportamento em amostras finitas de diferentes testes nos modelos de regressão normal linear e nos modelos não-lineares da família exponencial. Na Seção 4.5,

apresentamos algumas aplicações a dados reais e alguns comentários são feitos na Seção 4.6.

## 4.2 Ajustes para a estatística da razão de verossimilhanças

Consideremos  $n$  variáveis aleatórias independentes  $y_1, \dots, y_n$ , com função densidade conjunta  $\pi(y; \omega)$ , sendo  $\omega \in \Omega \subseteq \mathbb{R}^{p+k+1}$  o vetor de parâmetros desconhecidos de dimensão  $p + k + 1$  e  $y = (y_1, \dots, y_n)$ . Seja  $\omega^r$  o  $r$ -ésimo elemento do vetor  $\omega$ .

Assuma que o vetor de parâmetros  $\omega$  seja decomposto como  $\omega = (\nu^\top, \psi^\top)^\top$  sendo  $\nu = (\nu^1, \dots, \nu^p)^\top$  o vetor de parâmetros de interesse e  $\psi = (\psi^{p+1}, \dots, \psi^{p+k+1})^\top$  o vetor de parâmetros de perturbação, de dimensões  $p$  e  $k + 1$ , respectivamente.

O objetivo é testar a hipótese nula  $\mathcal{H}_0 : \nu = \nu_0$  contra a hipótese alternativa bilateral  $\mathcal{H}_1 : \nu \neq \nu_0$ , em que  $\nu_0$  é um vetor de constantes especificado de dimensão  $p$ . Denotamos por  $\hat{\omega} = (\hat{\nu}^\top, \hat{\psi}^\top)^\top$  o estimador de máxima verossimilhança irrestrito de  $\omega$  e por  $\tilde{\omega} = (\nu_0^\top, \tilde{\psi}^\top)^\top$  o estimador de máxima verossimilhança de  $\omega$  sob  $\mathcal{H}_0$ .

O logaritmo da função de verossimilhança do vetor de parâmetros  $\omega$  dado o vetor de observações  $y$  é

$$L(\omega) = L(\omega; y) = \log \pi(y; \omega).$$

A função escore total  $U$  é dada por

$$U = U(\omega) = \frac{\partial L(\nu, \psi)}{\partial \omega} = (U_\nu^\top, U_\psi^\top)^\top,$$

em que  $U_\nu = \partial L(\nu, \psi) / \partial \nu$  e  $U_\psi = \partial L(\nu, \psi) / \partial \psi$ .

A matriz de informação total de Fisher ( $I$ ) e a matriz de informação observada ( $J$ ) são dadas por

$$I = I(\omega) = -\mathbb{E} \left( \frac{\partial^2 L(\nu, \psi)}{\partial \omega \partial \omega^\top} \right) \quad \text{e} \quad J = J(\omega) = -\frac{\partial^2 L(\nu, \psi)}{\partial \omega \partial \omega^\top},$$

respectivamente. A notação  $J_{\psi\psi}$  denota a submatriz inferior à direita de  $J$  cuja dimensão é  $(k+1) \times (k+1)$ . No que segue, usaremos as seguintes notações:  $\hat{I} = I(\hat{\omega})$ ,  $\tilde{I} = I(\tilde{\omega})$ ,  $\hat{J} = J(\hat{\omega})$ ,  $\tilde{J} = J(\tilde{\omega})$  e  $\tilde{J}_{\psi\psi} = J_{\psi\psi}(\tilde{\omega})$ .

A estatística da razão de verossimilhanças para o teste de  $\mathcal{H}_0$  pode ser escrita

da seguinte forma

$$LR = -2\{L(\nu_0, \tilde{\psi}) - L(\hat{\nu}, \hat{\psi})\}.$$

Em amostras grandes, a distribuição da estatística da razão de verossimilhanças ( $LR$ ) sob a hipótese nula pode ser aproximada pela distribuição  $\chi_p^2$  com erro de ordem  $n^{-1}$ .

No caso do parâmetro de interesse ser escalar, a estatística da razão de verossimilhanças sinalizada para o teste de  $\mathcal{H}_0$  é dada por

$$R = \text{sign}(\hat{\nu} - \nu_0) \sqrt{LR}$$

e, neste caso, a hipótese alternativa pode ser unilateral ou bilateral. Em amostras grandes, a distribuição da estatística da razão de verossimilhanças sinalizada ( $R$ ) sob a hipótese nula pode ser aproximada pela distribuição  $\mathcal{N}(0, 1)$  com erro de ordem  $n^{-1/2}$ .

Vários ajustes têm sido propostos para melhorar a aproximação da distribuição da estatística da razão de verossimilhanças sinalizada ( $R$ ) pela distribuição normal padrão. No entanto, a maioria das aproximações envolve uma derivada relativa ao espaço amostral e a determinação de uma estatística ancilar. Nem sempre esta derivada pode ser facilmente calculada. Em particular, para os modelos da família exponencial de posto completo e os modelos de transformação, ela pode ser calculada explicitamente.

Suponha que uma estatística suficiente para o modelo pode ser escrita como  $(\hat{\omega}, a)$ , em que  $\hat{\omega}$  denota o estimador de máxima verossimilhança de  $\omega$  e  $a$  é uma estatística ancilar. Então, o logaritmo da função de verossimilhança é  $L(\omega; \hat{\omega}, a)$  em que a dependência nos dados está explicitada através de  $(\hat{\omega}, a)$ . Sua derivada com respeito a  $\hat{\omega}$  (uma derivada com respeito ao espaço amostral) é denotada por  $L'(\omega)$ . Utilizaremos o símbolo  $\prime$  para indicar uma derivada com respeito ao espaço amostral. A derivada do vetor escore com respeito a  $\hat{\omega}$  com  $a$  mantido fixo avaliada em  $\tilde{\omega}$  é denotada por  $\tilde{U}'$ .

Barndorff-Nielsen (1986, 1991) propôs a estatística da razão de verossimilhanças

sinalizada modificada dada por

$$R^* = R - \frac{1}{R} \log \gamma,$$

em que a expressão geral para a quantidade  $\gamma$  pode ser expressa como

$$\gamma = |\hat{J}|^{1/2} |\tilde{U}'|^{-1} |\tilde{J}_{\psi\psi}|^{1/2} \frac{R}{[(\hat{L}' - \tilde{L}')^\top (\tilde{U}')^{-1}]_\nu}, \quad (4.1)$$

que envolve as derivadas com respeito ao espaço amostral em  $\tilde{U}'$  e  $\hat{L}' - \tilde{L}'$ . As notações  $|\dots|$  e  $[(\hat{L}' - \tilde{L}')^\top (\tilde{U}')^{-1}]_\nu$  denotam, respectivamente, o determinante de uma matriz quadrada e o elemento do vetor  $(\hat{L}' - \tilde{L}')^\top (\tilde{U}')^{-1}$  referente ao parâmetro (escalar)  $\nu$ . Em amostras grandes, a distribuição da estatística da razão de verossimilhanças sinalizada modificada ( $R^*$ ) sob a hipótese nula pode ser aproximada pela distribuição  $\mathcal{N}(0, 1)$  com um erro de ordem  $n^{-3/2}$ . Note-se que esta aproximação para a estatística  $R$  tem erro de ordem  $n^{-1/2}$ .

Para contornar a dificuldade de obtenção das derivadas relativas ao espaço amostral, várias aproximações têm sido propostas. Uma das aproximações foi proposta por Skovgaard (1996), que fornece aproximações explícitas para  $(\hat{L}' - \tilde{L}')^\top$  e  $\tilde{U}'$  dadas por

$$(\hat{L}' - \tilde{L}')^\top \approx \hat{q} \hat{I}^{-1} \hat{J} \quad \text{e} \quad \tilde{U}' \approx \hat{Y} \hat{I}^{-1} \hat{J}, \quad (4.2)$$

sendo

$$\hat{q} = \text{cov}_{\omega_0} \{U(\omega_0), L(\omega_0) - L(\omega)\} |_{\omega_0=\hat{\omega}, \omega=\tilde{\omega}} \quad (4.3)$$

e

$$\hat{Y} = \text{cov}_{\omega_0} \{U(\omega_0), U(\omega)\} |_{\omega_0=\hat{\omega}, \omega=\tilde{\omega}}. \quad (4.4)$$

Desta forma, uma aproximação geral para a quantidade  $\gamma$  dada em (4.1) é

$$\tilde{\gamma} = |\hat{J}|^{-1/2} |\hat{I}| |\hat{Y}|^{-1} |\tilde{J}_{\psi\psi}|^{1/2} \frac{R}{[\hat{Y}^{-1} \hat{q}]_\nu}. \quad (4.5)$$

Novamente, a notação  $[\hat{Y}^{-1} \hat{q}]_\nu$  denota o elemento do vetor  $\hat{Y}^{-1} \hat{q}$  correspondente

ao parâmetro (escalar)  $\nu$ . Quando não for possível obter  $\gamma$  em (4.1) pode ser utilizado  $\tilde{\gamma}$  em (4.5) para obter a estatística da razão de verossimilhanças sinalizada modificada,

$$\tilde{R} = R - \frac{1}{R} \log \tilde{\gamma}.$$

Observe que a aproximação dada em (4.5) pode ser obtida explicitamente e envolve cumulantes de derivadas do logaritmo da função de verossimilhança até segunda ordem e duas quantidades não usuais ( $\hat{Y}$  e  $\hat{q}$ ) relativas ao vetor escore e a uma diferença de logaritmos da função de verossimilhança, avaliados em dois pontos diferentes.

A teoria é bastante desenvolvida para o caso em que o parâmetro de interesse é unidimensional. Bellio e Brazzale (1999) realizaram estudos numéricos bastante abrangentes confirmando a excelente aproximação normal para a distribuição de  $\tilde{R}$  no caso em que o parâmetro de interesse é unidimensional.

Uma generalização da estatística da razão de verossimilhanças sinalizada modificada de Barndorff-Nielsen para o caso em que o parâmetro de interesse é multidimensional com propriedades similares ao caso unidimensional é proposta por Skovgaard (2001). A estatística da razão de verossimilhanças ajustada para o teste de  $\mathcal{H}_0$  é dada por

$$LR_a^* = LR \left( 1 - \frac{1}{LR} \log \gamma \right)^2, \quad (4.6)$$

ou sua versão assintoticamente equivalente

$$LR_a^{**} = LR - 2 \log \gamma, \quad (4.7)$$

sendo aqui

$$\gamma = |\hat{J}|^{1/2} |\tilde{U}'|^{-1} |\tilde{J}_{\psi\psi}|^{1/2} |\tilde{J}_{\psi\psi}|^{-1/2} |\tilde{J}|^{1/2} \frac{\{\tilde{U}^\top \tilde{J}^{-1} \tilde{U}\}^{p/2}}{LR^{p/2-1} (\hat{L}' - \tilde{L}')^\top (\tilde{U}')^{-1} \tilde{U}}, \quad (4.8)$$

em que  $\tilde{J} = J(\tilde{\omega}; \tilde{\omega}, a)$ . Sob  $\mathcal{H}_0$ , as estatísticas  $LR_a^*$  e  $LR_a^{**}$  são aproximadamente distribuídas como  $\chi_p^2$  com alto grau de precisão (Skovgaard, 2001, p. 7). A estatística  $LR_a^{**}$  surge naturalmente do desenvolvimento teórico da generalização e

parece ser também numericamente melhor do que a sua versão equivalente  $LR_a^*$ , mas a versão  $LR_a^*$  tem a vantagem de ser sempre não-negativa e de reduzir-se a  $(R^*)^2$  quando  $p = 1$ . Para uma discussão mais detalhada do desenvolvimento teórico para a obtenção de  $LR_a^*$  e  $LR_a^{**}$  ver Skovgaard (2001, Seção 5.7).

Skovgaard (2001) propõe uma aproximação para  $\tilde{J}$  dada por

$$\tilde{J} \approx \tilde{I} \hat{\Upsilon}^{-1} \hat{J} \hat{I}^{-1} \hat{\Upsilon}. \quad (4.9)$$

Utilizando as aproximações dadas em (4.2) e (4.9), temos a seguinte aproximação para  $\gamma$  em (4.8):

$$\tilde{\gamma} = |\tilde{I}|^{1/2} |\hat{I}|^{1/2} |\hat{\Upsilon}|^{-1} |\tilde{J}_{\psi\psi}|^{1/2} |[\tilde{I} \hat{\Upsilon}^{-1} \hat{J} \hat{I}^{-1} \hat{\Upsilon}]_{\psi\psi}|^{-1/2} \frac{\{\tilde{U}^\top \hat{\Upsilon}^{-1} \hat{I} \hat{J}^{-1} \hat{\Upsilon} \tilde{I}^{-1} \tilde{U}\}^{p/2}}{LR^{p/2-1} \tilde{U}^\top \hat{\Upsilon}^{-1} \hat{q}}. \quad (4.10)$$

Vale salientar aqui que os testes baseados na estatística da razão de verossimilhanças original ( $LR$ ), em sua versão corrigida por correção de Bartlett ( $LR^*$ ) e na estatística da razão de verossimilhanças modificada sinalizada ( $R^*$ ) são invariantes segundo reparametrizações da forma  $(\nu, \psi) \rightarrow (\nu, \zeta(\nu, \psi))$ . O mesmo ocorre para as estatísticas da razão de verossimilhanças ajustadas  $LR_a^*$  e  $LR_a^{**}$ .

Lyon e Peters (2000) obtiveram um teste da razão de verossimilhanças ajustado para testar um dos componentes da variância em modelos lineares mistos. Seus resultados numéricos confirmaram a excelente aproximação normal para a distribuição de  $\tilde{R}$ .

Skovgaard (2001), através de resultados numéricos, constatou que os testes baseados nas estatísticas  $LR_a^{**}$  e  $LR_a^*$  têm melhores desempenhos do que os testes baseados na estatística  $LR$  e na sua versão corrigida, em especial no caso em que o vetor de parâmetros de interesse é escalar. Há indícios de que a estatística  $LR_a^{**}$  tem desempenho numericamente melhor que a estatística  $LR_a^*$ .

### 4.3 Modelos não-lineares da família exponencial

(1) Sejam  $y_1, \dots, y_n$  variáveis aleatórias independentes, cada qual com densidade na forma

(1) Como esta tese foi escrita de forma a que todos os capítulos sejam autosuficientes e possam ser lidos em qualquer ordem, a definição dos modelos não-lineares da família exponencial e de algumas quantidades relevantes introduzidas no Capítulo 3 são repetidas aqui.

$$\pi(y_\ell; \theta_\ell, \phi_\ell) = \exp \left\{ \phi_\ell [y_\ell \theta_\ell - b(\theta_\ell) - c(y_\ell)] - \frac{1}{2} e(y, \phi_\ell) \right\}, \quad \ell = 1, \dots, n, \quad (4.11)$$

em que  $E(y_\ell) = \mu_\ell = db(\theta_\ell)/d\theta_\ell$ ,  $\text{Var}(y_\ell) = \phi_\ell^{-1} V_\ell = \phi_\ell^{-1} d\mu/d\theta$ ,  $\phi$  ( $\phi_\ell > 0$ ) é desconhecido e chamado de parâmetro de precisão, seu recíproco ( $\phi_\ell^{-1}$ ) é chamado de parâmetro de dispersão,  $e(\cdot, \cdot)$  é uma função conhecida e  $V_\ell$  é a função de variância. Nesta seção, estamos considerando somente as distribuições contínuas da família exponencial, ou seja, as distribuições normal, normal inversa e gama. Admitimos que  $\phi_\ell^{-1} = \sigma^2 m_\ell$  com  $m_\ell = m(z_\ell, \delta) > 0$ , sendo  $z_\ell^\top$  a  $\ell$ -ésima linha da matriz de covariadas  $Z$ , de dimensão  $n \times p$ ,  $\delta$  um vetor de dimensão  $p \times 1$  de parâmetros desconhecidos e  $\sigma^2$  uma constante desconhecida finita e estritamente positiva. A média de  $y_\ell$  está relacionada com o componente sistemático através de uma função de ligação da forma

$$g(\mu_\ell) = \eta_\ell = f(x_\ell; \beta), \quad (4.12)$$

em que  $g(\cdot)$  é uma função conhecida monótona e duplamente diferenciável,  $\beta = (\beta^1, \dots, \beta^k)^\top$  é um vetor de  $k$  ( $k < n$ ) parâmetros desconhecidos,  $x_\ell = (x_{\ell 1}, \dots, x_{\ell K})^\top$  representa os valores de  $K$  variáveis explicativas e  $f(\cdot; \cdot)$  é uma função possivelmente não-linear no segundo argumento, contínua e diferenciável com respeito aos componentes de  $\beta$  tal que a matriz de derivadas  $X^* = X^*(\beta) = \partial\eta/\partial\beta^\top$ , com  $\eta = (\eta_1, \dots, \eta_n)^\top$ , tem posto  $k$  para todo  $\beta$ . A matriz  $X^*$  tem elementos que são, em geral, funções do vetor de parâmetros  $\beta$  desconhecidos.

Admitindo que  $e(y, \phi_\ell) = s(\phi_\ell) + t(y_\ell)$ , a equação (4.11) pode ser reescrita como

$$\pi(y_\ell; \theta_\ell, \phi_\ell) = \exp \left( -\frac{1}{2} \{ \phi_\ell d(y_\ell) + s(\phi_\ell) + t(y_\ell) \} \right), \quad \ell = 1, \dots, n, \quad (4.13)$$

em que  $d(y_\ell) = d_\ell = -2[y_\ell \theta_\ell - b(\theta_\ell) - c(y_\ell)]$ . Supomos que  $s(\phi_\ell)$  possui as quatro primeiras derivadas. Para a distribuição gama,  $d_\ell = 2\{y_\ell/\mu_\ell - \log(y_\ell/\mu_\ell)\}$  e  $s(\phi_\ell) = -2\{\phi_\ell \log(\phi_\ell) - \log\Gamma(\phi_\ell)\}$ . Já para as distribuições normal e normal inversa, temos  $d_\ell = (y_\ell - \mu_\ell)^2$  e  $d_\ell = (y_\ell - \mu_\ell)^2/(\mu_\ell^2 y_\ell)$ , respectivamente, e para ambas as distribuições  $s(\phi_\ell) = -\log(\phi_\ell)$ .

Seja  $L = L(\delta, \beta, \sigma^2)$  o logaritmo da função de verossimilhança de um modelo não-linear na família exponencial definido por (4.13) e (4.12) em algum espaço paramétrico dado o vetor de observações  $y = (y_1, \dots, y_n)^\top$ . Assume-se que a função  $L$  seja regular com respeito às derivadas em relação aos parâmetros até quarta ordem (Cox e Hinkley, 1974, Capítulo 9).

O logaritmo da função de verossimilhança do vetor de parâmetros  $\omega = (\delta^\top, \beta^\top, \sigma^2)^\top$ , dado o vetor de observações  $(y_1, \dots, y_n)$ , do modelo definido por (4.13) é

$$L(\delta, \beta, \sigma^2; y) = -\frac{1}{2} \sum_{\ell=1}^n \left\{ \frac{1}{\sigma^2 m_\ell} d_\ell + t(y_\ell) + s(\phi_\ell) \right\}$$

e a função escore total é dada por

$$U = U(\delta, \beta, \sigma^2) = (U_\delta^\top, U_\beta^\top, U_{\sigma^2})^\top, \quad (4.14)$$

em que

$$U_\delta = \frac{\partial L(\delta, \beta, \sigma^2; y)}{\partial \delta} = \frac{1}{2} \Psi^\top \Phi (d + \dot{S}), \quad U_\beta = \frac{\partial L(\delta, \beta, \sigma^2; y)}{\partial \beta} = X^{*\top} \Phi T V^{-1} (y - \mu)$$

e

$$U_{\sigma^2} = \frac{\partial L(\delta, \beta, \sigma^2; y)}{\partial \sigma^2} = \frac{1}{2\sigma^2} \iota^\top \Phi (d + \dot{S}),$$

respectivamente, sendo  $\Phi$ ,  $T$  e  $V$  matrizes diagonais de dimensão  $n \times n$  cujos  $(\ell, \ell)$ -ésimos elementos são dados, respectivamente, por  $\phi_\ell = 1/(\sigma^2 m_\ell)$ ,  $T_\ell = d\mu_\ell/d\eta_\ell$  e  $V_\ell = d\mu_\ell/d\theta_\ell$ ,  $\Psi$  é uma matriz  $n \times p$  cuja  $\ell$ -ésima linha é dada por  $\partial \log m(z_\ell, \delta) / \partial \delta^\top$ ,  $\dot{S} = (\dot{s}(\phi_1), \dots, \dot{s}(\phi_n))^\top$  com  $\dot{s}(\phi_\ell) = ds(\phi_\ell)/d\phi_\ell$ ,  $\ell = 1, \dots, n$ ,  $\iota$  é um vetor de uns de dimensão  $n$ ,  $d = (d_1, \dots, d_n)^\top$  e  $\mu = (\mu_1, \dots, \mu_n)^\top$ .

A matriz de informação total de Fisher (ver Apêndice B) é dada por

$$I = -E \left( \frac{\partial^2 L}{\partial \omega \partial \omega^\top} \right) = \begin{pmatrix} \frac{1}{2} \Psi^\top \Phi \ddot{S} \Phi \Psi & 0 & \frac{1}{2\sigma^2} \Psi^\top \ddot{S} \Phi^2 \iota \\ 0 & X^{*\top} \Phi^{1/2} W \Phi^{1/2} X^* & 0 \\ \frac{1}{2\sigma^2} \iota^\top \Phi^2 \ddot{S} \Psi & 0 & \frac{1}{2\sigma^4} \iota^\top \Phi^2 \ddot{S} \iota \end{pmatrix} \quad (4.15)$$

em que  $\ddot{S}$  é uma matriz diagonal de dimensão  $n \times n$  cujos elementos são dados por  $\ddot{s}(\phi_\ell) = d^2 s(\phi_\ell)/d\phi_\ell^2$ ,  $W$  é uma matriz diagonal de dimensão  $n \times n$  cujos elementos

são dados por  $w_\ell = (d\mu_\ell/d\eta_\ell)^2/V_\ell$ ,  $\ell = 1, \dots, n$  e 0 denota matrizes ou vetores nulos de dimensões apropriadas.

A matriz de informação observada é

$$J = -\frac{\partial^2 L}{\partial \omega \partial \omega^\top} = \begin{pmatrix} J_{11} & J_{12} & J_{13} \\ J_{12}^\top & J_{22} & J_{23} \\ J_{13}^\top & J_{23}^\top & J_{33} \end{pmatrix}, \quad (4.16)$$

em que

$$\begin{aligned} J_{11} &= \frac{1}{2} \Psi^\top \Phi \ddot{S} \Phi \Psi + \Psi^\top N_1 \Psi - \frac{1}{2} [(d + \dot{S})^\top M^{-1} \Phi][\Delta^{**}], \\ J_{12} &= \Psi^\top \Phi V^{-1} T N_2 X^*, \\ J_{13} &= \frac{1}{2\sigma^2} \Psi^\top \Phi N_1 \iota + \frac{1}{2\sigma^2} \Psi^\top \Phi \ddot{S} \Phi \iota, \\ J_{22} &= X^{*\top} W^{-1/2} \Phi W^{-1/2} X^* - X^{*\top} N_3 \Phi X^* - X^{*\top} N_4 \Phi X^* \\ &\quad - [(y - \mu)^\top \Phi V^{-1} T][X^{**}], \\ J_{23} &= X^{*\top} T V^{-1} \Phi (y - \mu), \\ J_{33} &= -\frac{1}{(\sigma^2)^2} (d + \dot{S})^\top \Phi \iota + \frac{1}{2(\sigma^2)^2} \iota^\top \Phi \ddot{S} \Phi \iota, \end{aligned}$$

com  $X^{**} = X^{**}(\beta) = \partial^2 \eta / \partial \beta \partial \beta^\top$  e  $\Delta^{**} = \Delta^{**}(\delta) = \partial^2 m / \partial \delta \partial \delta^\top$ , sendo  $m = (m_1, \dots, m_n)^\top$ , representando “arrays” de dimensões  $n \times k \times k$  e  $n \times p \times p$ , respectivamente. As matrizes  $M, N_1, N_2, N_3$  e  $N_4$  são matrizes diagonais de dimensão  $n \times n$  com  $(\ell, \ell)$ -ésimos elementos dados por  $m_\ell = m(z_\ell, \delta)$ ,  $n_{1\ell} = (d_\ell + \dot{s}(\phi_\ell))\phi_\ell$ ,  $n_{2\ell} = (y_\ell - \mu_\ell)/m_\ell$ ,  $n_{3\ell} = (y_\ell - \mu_\ell)g_\ell$  e  $n_{4\ell} = (y_\ell - \mu_\ell)f_\ell$ , respectivamente, em que  $g_\ell$  e  $f_\ell$  são escalares definidos por

$$g_\ell = \frac{1}{V_\ell} \frac{d\mu_\ell}{d\eta_\ell} \frac{d\mu_\ell^2}{d\eta_\ell^2} - \frac{1}{V_\ell^2} \frac{dV_\ell}{d\mu_\ell} \left( \frac{d\mu_\ell}{d\eta_\ell} \right)^3 \quad \text{e} \quad f_\ell = \frac{1}{V_\ell} \frac{d\mu_\ell}{d\eta_\ell} \frac{d\mu_\ell^2}{d\eta_\ell^2},$$

$\ell = 1, \dots, n$ . Considerando ligação canônica ( $\theta_\ell = \eta_\ell$ ), para a distribuição normal, temos que  $w_\ell = V_\ell = T_\ell = 1$ ,  $\theta_\ell = \mu_\ell$  e  $f_\ell = g_\ell = 0$ , enquanto que para a distribuição normal inversa temos que  $w_\ell = V_\ell = \mu_\ell^3$ ,  $T_\ell = -1/2\mu_\ell^3$ ,  $\theta_\ell = -(2\mu_\ell^2)^{-1}$ ,  $f_\ell = -3/8\mu_\ell^5$  e  $g_\ell = 0$ . Já para a distribuição gama, temos que  $w_\ell = V_\ell = \mu_\ell^2$ ,  $T_\ell = -\mu_\ell^2$ ,  $\theta_\ell = -\mu_\ell^{-1}$ ,  $f_\ell = -2\mu_\ell^3$  e  $g_\ell = 0$ . Vale salientar que cada elemento das submatrizes em (4.16) foi obtido no Apêndice B e a notação  $[\cdot][\cdot]$  representa

multiplicação de uma matriz por um “array”, chamado de produto colchete e definido em Wei (1998, p. 188), Bates e Watts (1988) e Seber e Wild (1989, p. 691); ver Apêndice C.

Nosso interesse é testar a hipótese nula  $\mathcal{H}_0 : \nu = \nu_0$  contra a hipótese alternativa bilateral  $\mathcal{H}_1 : \nu \neq \nu_0$ , em que  $\nu_0$  é um vetor de dimensão  $p \times 1$  de constantes especificado. Uma situação particular de interesse é aquela em que a hipótese nula indica que o modelo é homoscedástico no sentido de que todas as observações têm a mesma dispersão ( $\mathcal{H}_0 : \delta = \delta_0$ ). Será dada ênfase a esta situação na Seção 4.4.

Para obter as estatísticas da razão de verossimilhanças ajustadas ( $LR_a^*$ ) e ( $LR_a^{**}$ ), propostas por Skovgaard (2001) e dadas em (4.6) e (4.7), é necessário obter duas quantidades não usuais relativas ao vetor escore e ao logaritmo da função de verossimilhança. Estas quantidades são dadas por

$$q = E_{\omega_0}[U(\omega_0) (L(\omega_0) - L(\omega))] \quad (4.17)$$

e

$$\Upsilon = E_{\omega_0}[U(\omega_0) U^\top(\omega)]. \quad (4.18)$$

As quantidades em (4.17) e (4.18) são obtidas no Apêndice E e podem ser escritas em notação matricial como

$$q = \begin{pmatrix} \frac{1}{2} \Psi_0^\top \ddot{S}_0 \Phi_0 (\Phi - \Phi_0) \iota \\ X_0^{*\top} T_0 \Phi (\Theta_0 - \Theta) \iota \\ \frac{1}{2\sigma_0^2} \iota^\top (\Phi - \Phi_0) \Phi_0 \ddot{S}_0 \iota \end{pmatrix}$$

e

$$\Upsilon = \begin{pmatrix} \frac{1}{2} \Psi_0^\top \Phi_0 \ddot{S}_0 \Phi \Psi & 0 & \frac{1}{2\sigma_0^2} \Psi_0^\top \Phi_0 \ddot{S}_0 \Phi \iota \\ X_0^{*\top} T_0 (\Theta_0 - \Theta) \Phi \Psi & X_0^{*\top} T_0 \Phi V^{-1} T X^* & X_0^{*\top} T_0 \Phi (\Theta_0 - \Theta) \iota \\ \frac{1}{2\sigma_0^2} \iota^\top \Phi_0 \ddot{S}_0 \Phi \Psi & 0 & \frac{1}{2\sigma_0^2 \sigma^2} \iota^\top \Phi_0 \ddot{S}_0 \Phi \iota \end{pmatrix},$$

respectivamente, com  $\Theta = \text{diag}(\theta_1, \dots, \theta_n)$  e o subscrito 0 indica que as quantidades são avaliadas em  $\omega_0$ . Consequentemente, as quantidades  $\hat{q}$  e  $\hat{\Upsilon}$  em (4.3) e (4.4) tornam-se

$$\hat{q} = \begin{pmatrix} \frac{1}{2} \hat{\Psi}^\top \hat{S} \hat{\Phi} (\tilde{\Phi} - \hat{\Phi})_\iota \\ \hat{X}^{*\top} \hat{T} \tilde{\Phi} (\hat{\Theta} - \tilde{\Theta})_\iota \\ \frac{1}{2\hat{\sigma}^2} \iota^\top (\tilde{\Phi} - \hat{\Phi}) \hat{\Phi} \hat{S} \iota \end{pmatrix} \quad (4.19)$$

e

$$\hat{\Upsilon} = \begin{pmatrix} \frac{1}{2} \hat{\Psi}^\top \hat{\Phi} \hat{S} \tilde{\Phi} \tilde{\Psi} & 0 & \frac{1}{2\tilde{\sigma}^2} \hat{\Psi}^\top \hat{\Phi} \hat{S} \tilde{\Phi} \iota \\ \hat{X}^{*\top} \hat{T} (\hat{\Theta} - \tilde{\Theta}) \tilde{\Phi} \tilde{\Psi} & \hat{X}^{*\top} \hat{T} \tilde{\Phi} \tilde{V}^{-1} \tilde{T} \tilde{X}^* & \hat{X}^{*\top} \hat{T} \tilde{\Phi} (\hat{\Theta} - \tilde{\Theta})_\iota \\ \frac{1}{2\hat{\sigma}^2} \iota^\top \hat{\Phi} \hat{S} \tilde{\Phi} \tilde{\Psi} & 0 & \frac{1}{2\hat{\sigma}^2 \tilde{\sigma}^2} \iota^\top \hat{\Phi}_0 \hat{S} \tilde{\Phi} \iota \end{pmatrix}, \quad (4.20)$$

respectivamente. Para as distribuições normal e normal inversa, temos  $s(\phi_\ell) = -\log(\phi_\ell)$  e, portanto,  $\dot{S} = (-1/\phi_1, \dots, -1/\phi_n)^\top$  e  $\ddot{S} = \Phi^{-2}$ , respectivamente. Neste caso, a matriz de informação total de Fisher ( $I$ ),  $\hat{q}$ ,  $\hat{\Upsilon}$  e a matriz de informação observada ( $J$ ) reduzem-se, respectivamente, a

$$I = -E \left( \frac{\partial^2 L}{\partial \omega \partial \omega^\top} \right) = \begin{pmatrix} \frac{1}{2} \Psi^\top \Psi & 0 & \frac{1}{2\sigma^2} \Psi^\top \iota \\ 0 & X^{*\top} \Phi^{1/2} W \Phi^{1/2} X^* & 0 \\ \frac{1}{2\sigma^2} \iota^\top \Psi & 0 & \frac{n}{2\sigma^4} \end{pmatrix},$$

$$\hat{q} = \begin{pmatrix} \frac{1}{2} \hat{\Psi}^\top \hat{\Phi}^{-1} \tilde{\Phi} \iota - \frac{1}{2} \hat{\Psi}^\top \iota \\ \hat{X}^{*\top} \hat{T} \tilde{\Phi} (\hat{\Theta} - \tilde{\Theta})_\iota \\ \frac{1}{2\hat{\sigma}^2} (\iota^\top \hat{\Phi}^{-1} \tilde{\Phi} \iota - n) \end{pmatrix},$$

$$\hat{\Upsilon} = \begin{pmatrix} \frac{1}{2} \hat{\Psi}^\top \hat{\Phi}^{-1} \tilde{\Phi} \tilde{\Psi} & 0 & \frac{1}{2\hat{\sigma}^2} \hat{\Psi}^\top \hat{\Phi}^{-1} \tilde{\Phi} \iota \\ \hat{X}^{*\top} \hat{T}(\hat{\Theta} - \tilde{\Theta}) \tilde{\Phi} \tilde{\Psi} & \hat{X}^{*\top} \hat{T} \tilde{\Phi} \tilde{X}^* & \frac{1}{\hat{\sigma}^2} \hat{X}^{*\top} \hat{T} \tilde{\Phi} (\hat{\Theta} - \tilde{\Theta}) \iota \\ \frac{1}{2\hat{\sigma}^2} \iota^\top \hat{\Phi}^{-1} \tilde{\Phi} \tilde{\Psi} & 0 & \frac{1}{2\hat{\sigma}^2 \tilde{\sigma}^2} \iota^\top \hat{\Phi}^{-1} \tilde{\Phi} \iota \end{pmatrix}$$

e

$$J = \begin{pmatrix} J_{11} & J_{12} & J_{13} \\ J_{12}^\top & J_{22} & J_{23} \\ J_{13}^\top & J_{23}^\top & J_{33} \end{pmatrix},$$

sendo

$$\begin{aligned} J_{11} &= \frac{1}{2} \Psi^\top \Psi + \frac{1}{\sigma^2} \Psi^\top N_1 \Psi - \frac{1}{2} [(d + \dot{S})^\top M^{-1} \Phi][\Delta^{**}], \\ J_{12} &= \Psi^\top \Phi N_2 X^*, \\ J_{13} &= \frac{1}{2\sigma^2} \Psi^\top \Phi N_1 \iota + \frac{1}{2\sigma^2} \Psi^\top \iota, \\ J_{22} &= X^{*\top} W^{-1/2} \Phi W^{-1/2} X^* - [(y - \mu)^\top \Phi][X^{**}], \\ J_{23} &= X^{*\top} \Phi (y - \mu) \end{aligned}$$

e

$$J_{33} = -\frac{1}{(\sigma^2)^2} (d + \dot{S})^\top \Phi \iota + \frac{n}{2(\sigma^2)^2}.$$

Para a distribuição gama vimos que  $s(\phi_\ell) = -2\{\phi_\ell \log(\phi_\ell) - \log \Gamma(\phi_\ell)\}$  e, portanto, temos que  $\dot{s}(\phi_\ell) = -2\{\log(\phi_\ell) + 1 - \psi(\phi_\ell)\}$  e  $\ddot{s}(\phi_\ell) = -2\{\phi_\ell^{-1} + \psi'(\phi_\ell)\}$ , em que  $\psi(\cdot)$  é função digama. Nessa situação, a matriz de informação total de Fisher ( $I$ ),  $\hat{q}$ ,  $\hat{\Upsilon}$  e a matriz de informação observada ( $J$ ) pouco se simplificam.

Em suma, para obtermos a estatística da razão de verossimilhanças ajustada ( $LR_a^*$ ) e sua versão equivalente ( $LR_a^{**}$ ) na classe dos modelos não-lineares da família exponencial, devemos substituir  $\tilde{I}$ ,  $\hat{I}$ ,  $\hat{J}$ ,  $\tilde{U}$ ,  $\tilde{J}_{\psi\psi}$  e as quantidades  $\hat{q}$  e  $\hat{\Upsilon}$  dadas em (4.19) e (4.20) respectivamente, na expressão de  $\tilde{\gamma}$  dada em (4.10). Temos que as matrizes  $U$ ,  $I$  e  $J$  são dadas, respectivamente, pelas equações (4.14), (4.15) e (4.16).

#### 4.4 Resultados numéricos

Nesta seção, nosso objetivo é comparar, através de simulações, os desempenhos de vários testes baseados na estatística da razão de verossimilhanças usual e em suas versões modificadas. Enfocaremos apenas a situação em que a hipótese nula de interesse indica que o modelo é homoscedástico (no sentido de que todas as observações têm a mesma dispersão).

As estatísticas a serem avaliadas são: estatística da razão de verossimilhanças usual ( $LR$ ), estatística da razão de verossimilhanças corrigida via correção de Bartlett<sup>(2)</sup> ( $LR^*$ ), estatística da razão de verossimilhanças ajustada ( $LR_a^*$ ) dada em (4.6), sua versão assintoticamente equivalente ( $LR_a^{**}$ ) dada em (4.7), estatística da razão de verossimilhanças perfiladas modificadas baseada na proposta de Cox e Reid (1987) ( $LR_m$ ) e a estatística da razão de verossimilhanças perfiladas modificadas corrigida via correção de Bartlett ( $LR_m^*$ ). Para o teste de homoscedasticidade em modelos normais lineares, as estatísticas  $LR_m$ ,  $LR^*$  e  $LR_m^*$  são dadas no Capítulo 2 em (2.5), (2.6) e (2.7), respectivamente. Já para os modelos normais não-lineares, as estatísticas  $LR_m$  e  $LR_m^*$  são dadas em (3.7) e (3.8), respectivamente. Note-se que a correção de Bartlett para a estatística  $LR_m$  tem a mesma forma quer seja o modelo linear ou não-linear.

Inicialmente, temos como objetivo analisar a influência do número de parâmetros de perturbação ( $k + 1$ ), do número de parâmetros de interesse ( $p$ ), do tamanho da amostra ( $n$ ) e da distribuição das covariadas nos desempenhos dos testes e compará-los entre si. Os desempenhos são avaliados em função da proximidade da probabilidade de rejeição da hipótese nula, sendo esta verdadeira (probabilidade do erro tipo I), aos respectivos níveis nominais. Avaliamos também o poder dos testes em estudo sob algumas situações. Adicionalmente, foram considerados aspectos de robustez dos testes de heteroscedasticidade. Comparamos a média, a variância e quantis amostrais das estatísticas citadas acima com as quantidades

<sup>(2)</sup> Esta estatística somente será considerada para o modelo linear.

correspondentes da distribuição de referência e avaliamos os tamanhos empíricos dos testes para valores bem pequenos do nível nominal.

Os resultados numéricos baseiam-se nos seguintes modelos:

$$y_\ell = \beta_1 + \beta_2 x_{\ell 2} + \dots + \beta_k x_{\ell k} + u_\ell, \quad \ell = 1, \dots, n, \quad (\text{modelo 1})$$

e

$$y_\ell = \beta_1 + \exp\{\beta_2 x_{\ell 2}\} + \sum_{j=3}^k \beta_j x_{\ell j} + u_\ell, \quad \ell = 1, \dots, n, \quad (\text{modelo 2})$$

em que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{\delta_1 z_{\ell 1} + \dots + \delta_p z_{\ell p}\})$ ,  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$ . A hipótese nula considerada é  $\mathcal{H}_0: \delta_1 = \dots = \delta_p = 0$  indicando assim que o modelo é homoscedástico. A variável resposta foi gerada assumindo que  $\beta_1 = \dots = \beta_k = 1$ ,  $\sigma^2 = 1$  e diferentes valores de  $p$  e  $k$  foram considerados. Note-se que o modelo 1 é linear, ao contrário do modelo 2.

As covariadas  $x_2, \dots, x_k$  foram geradas como amostras aleatórias da distribuição  $\mathcal{U}(0, 1)$ . Quando  $p < k$ , a matriz  $Z$  foi construída usando as colunas 2,  $\dots$ ,  $p + 1$  de  $X$ ; quando  $p \geq k$ , as colunas extras de  $Z$  foram criadas usando amostras independentes da distribuição  $\mathcal{U}(0, 1)$ .

O número de réplicas foi fixado em 10.000<sup>(3)</sup> e foram considerados os seguintes níveis nominais:  $\alpha = 10\%$ ,  $5\%$ ,  $1\%$  e  $0,5\%$ . As simulações foram realizadas usando a linguagem de programação matricial `0x` (Doornik, 2001).

Para cada tamanho da amostra e cada nível considerado, calculamos as taxas de rejeição de cada teste, isto é, estimamos, via simulação,  $P(LR \geq x_\alpha)$ ,  $P(LR^* \geq x_\alpha)$ ,  $P(LR_m \geq x_\alpha)$ ,  $P(LR_m^* \geq x_\alpha)$ ,  $P(LR_a^* \geq x_\alpha)$  e  $P(LR_a^{**} \geq x_\alpha)$ , em que  $x_\alpha$  é o quantil  $(1 - \alpha)$  da distribuição  $\chi_p^2$ . Todas as entradas das tabelas apresentadas correspondem a porcentagens, com exceção das Tabelas 4.5, 4.6, 4.11, 4.14, 4.15 e 4.20.

#### 4.4.1 Modelo normal linear

Os resultados de simulação apresentados aqui baseiam-se no modelo de regressão normal linear (modelo 1).

<sup>(3)</sup> As Tabelas 4.11 e 4.20 apresentam os resultados de simulação em que o número de réplicas foi fixado em 100.000, ou seja, 10 vezes mais réplicas do que nas simulações anteriores.

A Tabela 4.1 apresenta os tamanhos observados dos testes para a situação em que  $n = 40$ ,  $p = 1$  e diferentes valores para  $k$  foram considerados ( $k = 2, \dots, 8$ ). Claramente, podemos observar que o desempenho do teste baseado na estatística da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) é bem melhor do que o do teste baseado na estatística da razão de verossimilhanças ( $LR$ ), considerando ambos sem correção de Bartlett. Notamos ainda que o número de parâmetros de perturbação ( $k + 1$ ) tem impacto considerável sobre a aproximação por  $\chi^2$  para as distribuições das estatísticas  $LR$  e  $LR^*$ . Para  $k = 8$  e  $\alpha = 5\%$ , por exemplo, ambos os testes são liberais apresentando taxas de rejeição de 13,5% e 7,8%, respectivamente, bem maiores portanto que o nível nominal. Este fato é bem mais marcante no caso da estatística  $LR$  do que no caso da estatística  $LR^*$ , pois a correção de Bartlett aplicada à estatística da razão de verossimilhanças atenua a tendência do teste em rejeitar demais a hipótese nula, porém não o bastante. O teste baseado na estatística da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) é conservativo pois exhibe taxas de rejeição menores que os níveis nominais, enquanto que o teste baseado na estatística da razão de verossimilhanças perfiladas modificadas corrigida ( $LR_m^*$ ) introduz uma correção na taxa de rejeição produzindo taxas de rejeição mais próximas do nível nominal do teste. Os testes baseados nas estatísticas da razão de verossimilhanças ajustadas  $LR_a^*$  e  $LR_a^{**}$  tiveram melhor desempenho do que os testes baseados nas estatísticas  $LR$  e  $LR^*$ , apresentando taxas de rejeição mais próximas dos níveis nominais. É interessante notar que o impacto do número de parâmetros de perturbação é bem menos marcante nos testes baseados nas estatísticas  $LR_m$ ,  $LR_m^*$ ,  $LR_a^{**}$  e  $LR_a^*$ . Para estes testes, quando o número de parâmetros de perturbação aumenta, as taxas de rejeição permanecem mais estáveis em relação aos respectivos níveis nominais do que quando os testes são baseados na estatística da razão de verossimilhanças original ( $LR$ ) e na sua versão corrigida ( $LR^*$ ). Para a mesma situação ( $k = 8$  e  $\alpha = 5\%$ ), as taxas de rejeição do teste da razão de verossimilhanças perfiladas modificadas, de sua versão corrigida, da razão de verossimilhanças ajustada e de sua versão assintoticamente equivalente são 4,5%, 4,9%, 5,4% e 4,9%, respectivamente. Entre as estatísticas

$LR_m$  e  $LR_m^*$ , aquela que produz melhor teste é  $LR_m^*$  e, entre as estatísticas  $LR_a^*$  e  $LR_a^{**}$ , aquela que produz melhor teste é  $LR_a^{**}$ . Já entre as estatísticas  $LR_m^*$  e  $LR_a^{**}$ , uma ordenação não é tão evidente, pois as mesmas produzem testes com desempenhos muito semelhantes. As conclusões anteriores também se verificam para outros níveis nominais, até mesmo para níveis nominais muito pequenos.

A Tabela 4.2 contém resultados para a situação em que  $p = 2$ ,  $n = 40$  e  $k = 2, \dots, 8$ . Notamos que os comportamentos dos testes foram bastante similares aos obtidos no caso unidimensional ( $p = 1$ ) e as conclusões obtidas na Tabela 4.1 também se verificam aqui. Como uma ilustração, quando  $k = 7$  e  $\alpha = 5\%$ , as taxas de rejeição dos testes baseados em  $LR$ ,  $LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  são, respectivamente, 16,9%, 9,2%, 4,5%, 5,0%, 5,6% e 4,9%.

Na Tabela 4.3 fixamos  $n = 40$ ,  $k = 5$  e diferentes valores para  $p$ , o número de parâmetros de interesse, foram considerados ( $p = 1, \dots, 8$ ). Os desempenhos dos testes foram bastante similares aos obtidos no caso do aumento de  $k$ . Os testes que tiveram suas taxas de rejeição mais afetadas pelo aumento do número de parâmetros que definem o comportamento de heteroscedasticidade foram o teste da razão de verossimilhanças e o teste da razão de verossimilhanças corrigido por correção de Bartlett. Os demais testes apresentaram desempenhos razoáveis, possuindo tamanhos empíricos mais estáveis em relação aos respectivos níveis nominais. Para  $p = 5$  e  $\alpha = 5\%$ , a taxa de rejeição do teste da razão de verossimilhanças original excede 16,0%, isto é, aproximadamente três vezes maior que o nível nominal considerado, enquanto que as taxas de rejeição dos testes baseados nas estatísticas  $LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  são 8,5%, 4,4%, 5,0%, 5,5% e 4,9%, respectivamente. As conclusões anteriores obtidas a partir das Tabelas 4.1 e 4.2 também se verificam aqui.

Na Tabela 4.4, fixamos o valor de  $p$  em 2 e variamos o tamanho da amostra ( $n = 30, 35, 40, 45, 100$ ) para  $k = 5$ . Claramente, conforme cresce o tamanho da amostra, as taxas de rejeição de todos os testes aproximam-se dos respectivos níveis nominais. Além disto, observamos que o teste baseado na estatística da razão de

verossimilhanças ( $LR$ ) é liberal até mesmo para  $n = 100$ . As conclusões anteriores tiradas a partir das Tabelas 4.1 a 4.3 também podem ser verificadas aqui.

Na Tabela 4.5 apresentamos comparações da média e da variância das estatísticas  $LR$ ,  $LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$ ,  $LR_a^{**}$  e da distribuição  $\chi_2^2$ , para o caso em que  $n = 40$ ,  $p = 2$  e  $k = 5$ . Os resultados desta tabela mostram que as estatísticas  $LR_a^*$ ,  $LR_a^{**}$  e  $LR_m^*$  apresentam médias e variâncias mais próximas da média e da variância da distribuição de referência ( $\chi_2^2$ ) do que a estatística  $LR_m$ . Por outro lado, essas medidas para  $LR$  e  $LR^*$  excedem as da distribuição  $\chi_2^2$ . Este mesmo comportamento foi observado na Tabela 4.6 ao comparar os quantis amostrais de  $LR$ ,  $LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  com os da distribuição  $\chi_2^2$ . Por exemplo, para o quantil de ordem 90% temos que as estatísticas  $LR_m^*$ ,  $LR_a^{**}$ ,  $LR_a^*$  e  $LR_m$ , nesta ordem, apresentam quantis mais próximos aos quantis da distribuição  $\chi_2^2$ , enquanto que os quantis amostrais de  $LR$  e  $LR^*$  excederam os da distribuição  $\chi_2^2$ .

Na Tabela 4.7 são apresentados resultados de simulação obtidos com diferentes configurações para as covariáveis. Fixamos  $n = 35$ ,  $p = 2$ ,  $k = 5$  e geramos valores para as covariadas como amostras aleatórias das seguintes distribuições:  $\mathcal{U}(0, 1)$ ,  $LN(0, 1)$  (log-normal padrão),  $t_3$  e  $\chi_5^2$ . Podemos observar que as configurações amostrais das covariadas parecem ter impacto considerável sobre os tamanhos empíricos dos testes. Os testes baseados nas estatísticas da razão de verossimilhanças ( $LR$ ) e em sua versão corrigida ( $LR^*$ ) apresentaram-se ainda mais liberais do que quando as covariadas foram obtidas de uma distribuição  $\mathcal{U}(0, 1)$ , enquanto que os testes baseados nas estatísticas da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ), na razão de verossimilhanças perfiladas modificadas corrigida ( $LR_m^*$ ) e na razão de verossimilhanças ajustada ( $LR_a^{**}$ ) mostraram-se conservativos e o teste baseado na estatística ( $LR_a^*$ ) apresentou taxas de rejeição ligeiramente acima do nível nominal. Por exemplo, para  $p = 5$ ,  $\alpha = 5\%$ , covariadas  $LN(0, 1)$ , temos que a taxa de rejeição do teste da razão de verossimilhanças original é 30,6%, ou seja, é seis vezes maior o nível nominal considerado, enquanto que as taxas de rejeição dos testes baseados nas estatísticas  $LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  são 9,2%, 4,5%, 4,9%, 5,0% e 4,7%, respectivamente.

A Tabela 4.8 apresenta resultados de simulação obtidos sob a hipótese alternativa (heteroscedasticidade) para  $n = 40$ ,  $p = 2$ ,  $k = 5$ ,  $\alpha = 5\%$  e diferentes valores de  $\delta_1 = \delta_2 = \delta$  com  $\delta$  variando de 0,0 a 3,5 foram considerados. São desconsiderados os testes que nas simulações de tamanho mostraram-se liberais, ou seja, aqueles que apresentaram, sob  $\mathcal{H}_0$ , taxas de rejeição superiores aos níveis de significância nominal. São eles os testes baseados nas estatísticas da razão de verossimilhanças original ( $LR$ ) e em sua versão corrigida ( $LR^*$ ). Então, comparamos os poderes dos testes baseados nas estatísticas  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$ . Podemos observar que os poderes dos testes baseados em  $LR_m$  e  $LR_m^*$  são bem semelhantes, o mesmo ocorrendo com os poderes dos testes baseados em  $LR_a^*$  e  $LR_a^{**}$ . Notamos uma leve vantagem para os testes baseados nas estatísticas  $LR_m^*$  e  $LR_a^{**}$ , uma vez que apresentaram menor distorção de tamanho. Comparando os poderes dos testes baseados nas estatísticas  $LR_m^*$  e  $LR_a^{**}$ , podemos afirmar que têm desempenhos bem semelhantes.

Avaliamos também a robustez dos testes de heteroscedasticidade sob má especificação da função cedástica. Inicialmente, consideramos a omissão de covariáveis em  $Z$ , porém, com a escolha correta da função cedástica. Geramos dados para o caso em que  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2} + \delta_3 z_{\ell 3})$  com  $n = 35$ ,  $k = 5$ , covariáveis uniformes e erros normais. Para efeito de se estimar os parâmetros e obter as estatísticas dos testes de heteroscedasticidade, assumimos (incorretamente) que  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2})$  ( $p = 2$ ). Em segundo lugar, consideramos a escolha incorreta de uma covariável em  $Z$ , porém, com a escolha correta da função cedástica. Geramos dados para os casos em que  $m_\ell = \exp(\delta_1 z_{\ell 1})$ ,  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2}^2)$  e  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 \sqrt{z_{\ell 2}})$  com  $n = 35$ ,  $k = 5$ , covariáveis uniformes e erros normais. Novamente postulamos, erroneamente, que  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2})$ . As Tabelas 4.9 e 4.10 apresentam resultados de simulação obtidos sob a hipótese nula ( $\delta = 0$ ) sob a hipótese nula e sob a hipótese alternativa (heteroscedasticidade) para alguns valores de  $\delta$  e  $\alpha = 5\%$ .

Analisando as Tabelas 4.9 e 4.10, observamos que não há indícios de que a omissão de uma covariável prejudique o desempenho dos testes de heteroscedasticidade. Os poderes dos testes tendem a ser ligeiramente maiores no caso em que

se omite uma covariável em  $Z$  do que quando não se omite. No caso em que a covariável  $z_2$  em  $Z$  é mal especificada, observamos que os poderes dos testes são ligeiramente maiores quando utilizamos o quadrado desta covariável, enquanto que os poderes dos testes tornam-se ligeiramente menores quando utilizamos a raiz quadrada da mesma variável. Todavia, no caso em que não utilizamos a covariável ( $z_2$ ) observamos que ocorre uma diminuição razoável nos poderes dos testes.

Na Tabela 4.11 apresentamos os resultados de simulação de 100.000 amostras de Monte Carlo, ou seja, 10 vezes mais réplicas do que nas simulações anteriores. O uso de um número maior de réplicas permitirá avaliar os tamanhos empíricos dos testes para valores bem pequenos do nível nominal. A Tabela 4.11 mostra o número de rejeições para os diferentes testes (baseados em  $LR$ ,  $LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$ ) com níveis nominais variando de 0,01% a 50% para o modelo normal linear em que  $n = 40$ ,  $p = 2$  e  $k = 5$ .

Da Tabela 4.11, observa-se que no extremo da cauda da distribuição  $\chi_2^2$  (que corresponde a tomar níveis nominais bem pequenos), as diferenças relativas entre as taxas de rejeição empíricas do teste da razão de verossimilhanças e os níveis nominais correspondentes são extremamente elevadas, ao contrário do que ocorre na região mediana da distribuição. Por exemplo, se  $\alpha = 0,01$ , essa diferença relativa é de 1050%  $((115 - 10)/10 \times 100\%)$ , enquanto que, para  $\alpha = 0,5$  esta diferença é de apenas 18%  $((58929 - 50000)/50000 \times 100\%)$ . Tal diferença é menos acentuada para a estatística da razão de verossimilhanças corrigida ( $LR^*$ ). Para as duas situações citadas acima, as diferenças relativas são de 180%  $((28 - 10)/10 \times 100\%)$  e 5%  $((52507 - 50000)/50000 \times 100\%)$ , respectivamente. É interessante notar que esta discrepância é bem menor para as outras estatísticas consideradas. A estatística  $LR_a^{**}$ , por exemplo, apresenta diferenças relativas para  $\alpha = 0,01\%$  e  $\alpha = 50\%$  de -10%  $((9 - 10)/10 \times 100\%)$  e -2%  $((48847 - 50000)/50000 \times 100\%)$ , respectivamente.

A Figura 4.1 apresenta o gráfico das discrepâncias relativas entre os quantis amostrais e o quantis assintóticos das estatísticas dos testes contra os quantis assintóticos. Mais precisamente, denotando por  $ST$  a estatística do teste ( $LR$ ,

$LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  ou  $LR_a^{**}$ , conforme o caso), o quantil amostral de ordem  $(1 - \alpha)$  do conjunto de valores simulados da estatística do teste sendo denotado por  $ST(1 - \alpha)$  e o corresponde quantil da distribuição  $\chi_2^2$  sendo denotado por  $\chi_2^2(1 - \alpha)$ ; a discrepância relativa é definida como

$$\frac{ST(1 - \alpha) - \chi_2^2(1 - \alpha)}{\chi_2^2(1 - \alpha)}.$$

Quanto mais próxima da ordenada zero a curva estiver, melhor é a aproximação assintótica utilizada no teste. Analisando a Figura 4.1, é interessante notar que, para todos os testes, as curvas de discrepâncias de quantis são próximas de linhas retas paralelas ao eixo horizontal de ordenada zero. Tais discrepâncias permanecem, portanto, aproximadamente estáveis, no entanto, em patamares diferentes. As discrepâncias relativas de quantis para a estatística da razão de verossimilhanças usual ( $LR$ ) e sua versão corrigida ( $LR^*$ ) apresentam-se em torno de 32% e 9%, respectivamente, o que confirma que tais testes são liberais. Por outro lado, a estatística da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) apresenta discrepância de quantil em torno de -4% (teste conservativo). Finalmente, as estatísticas  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  apresentam curvas de discrepâncias de quantis bem próximas da ordenada nula.

A Figura 4.2 apresenta o gráfico das discrepâncias relativas entre níveis descritivos ( $p$ -valores) dos testes e os níveis descritivos assintóticos contra os níveis descritivos assintóticos. Tal discrepância relativa, para um valor  $x$  observado da estatística  $ST$ , é definida por

$$\frac{P(ST > x) - P(\chi_2^2 > x)}{P(\chi_2^2 > x)}.$$

A Figura 4.2 mostra claramente que, quando se caminha em direção ao extremo da cauda da distribuição  $\chi^2$  de referência, as discrepâncias relativas de níveis descritivos para o teste da razão de verossimilhanças tendem a crescer rapidamente. Tal discrepância no extremo da cauda é bem menos acentuada, mas ainda considerável, para o teste baseado na estatística da razão de verossimilhanças corrigida ( $LR^*$ ). A discrepância é bem menor para os outros testes.

Tabela 4.1 Tamanhos dos testes - Modelo normal linear com  $n = 40$ ,  $p = 1$  e diversos valores para  $k$ .

$k$	$\alpha = 10\%$						$\alpha = 5\%$					
	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **
2	11,8	10,4	9,5	10,1	10,3	10,2	6,3	5,4	4,8	5,2	5,3	5,3
3	12,5	10,5	9,4	9,9	10,1	10,0	6,8	5,4	4,7	5,1	5,3	5,2
4	14,2	11,1	9,7	10,3	10,6	10,4	8,1	5,6	4,5	4,9	5,3	5,2
5	15,3	11,4	9,4	9,9	10,5	10,1	8,7	6,1	4,7	5,1	5,5	5,2
6	16,3	11,5	9,1	9,7	10,3	9,7	9,5	6,0	4,5	4,8	5,2	4,9
7	19,2	13,4	9,6	10,2	11,0	10,2	12,1	7,7	4,8	5,3	6,0	5,2
8	21,1	14,4	9,1	9,8	10,8	9,8	13,5	7,8	4,5	4,9	5,4	4,9

$k$	$\alpha = 1\%$						$\alpha = 0,5\%$					
	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **
2	1,4	1,1	0,9	1,0	1,0	1,0	0,8	0,5	0,5	0,5	0,5	0,5
3	1,6	1,0	0,7	0,9	0,9	0,9	0,8	0,5	0,3	0,4	0,5	0,4
4	2,0	1,1	0,8	0,9	1,1	1,0	1,1	0,6	0,4	0,5	0,5	0,5
5	2,5	1,2	0,8	0,9	1,0	0,9	1,4	0,7	0,4	0,4	0,5	0,5
6	2,9	1,5	1,0	1,1	1,2	1,1	1,9	0,9	0,5	0,6	0,7	0,6
7	4,5	2,1	0,8	0,9	1,3	0,9	2,8	1,1	0,4	0,5	0,5	0,4
8	4,6	2,2	0,8	1,0	1,2	0,9	3,2	1,3	0,5	0,6	0,6	0,5

Tabela 4.2 Tamanhos dos testes - Modelo normal linear com  $n = 40$ ,  $p = 2$  e diversos valores para  $k$ .

$k$	$\alpha = 10\%$						$\alpha = 5\%$					
	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **
2	12,5	10,0	8,8	9,4	9,6	9,6	6,8	5,3	4,4	4,9	5,0	5,0
3	13,8	10,7	9,2	9,9	10,2	10,1	7,5	5,4	4,4	4,9	5,1	5,0
4	16,0	11,4	9,3	10,2	10,6	10,3	9,2	6,0	4,4	5,1	5,4	5,1
5	18,2	12,3	9,3	10,1	10,4	9,8	10,8	6,3	4,5	4,9	5,3	4,9
6	19,5	13,1	9,3	9,9	10,5	9,8	12,3	7,0	4,4	4,9	5,2	4,9
7	24,8	16,4	9,4	10,3	10,7	10,3	16,9	9,2	4,5	5,0	5,6	4,9
8	26,9	16,4	9,0	9,7	11,4	9,7	17,6	9,5	4,5	5,0	5,8	4,9

$k$	$\alpha = 1\%$						$\alpha = 0,5\%$					
	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **
2	1,6	1,1	0,8	1,0	1,0	1,0	0,9	0,6	0,4	0,5	0,5	0,5
3	1,8	1,0	0,8	0,8	0,9	0,9	0,9	0,5	0,3	0,4	0,5	0,5
4	2,3	1,1	0,6	0,8	0,8	0,8	1,3	0,5	0,3	0,4	0,4	0,4
5	3,1	1,5	0,9	1,0	1,1	1,1	1,9	0,8	0,5	0,5	0,6	0,5
6	4,0	1,7	0,9	1,0	1,2	1,1	2,5	1,0	0,4	0,5	0,7	0,5
7	6,2	2,5	0,7	0,9	1,2	0,9	4,1	1,4	0,3	0,4	0,5	0,4
8	6,8	2,8	0,8	1,0	1,2	0,8	4,6	1,7	0,4	0,5	0,6	0,4

Tabela 4.3 Tamanhos dos testes - Modelo normal linear com  $n = 40$ ,  $k = 5$  e diversos valores para  $p$ .

$p$	$\alpha = 10\%$						$\alpha = 5\%$					
	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **
1	15,3	11,4	9,4	9,9	10,5	10,1	8,7	6,1	4,7	5,1	5,5	5,2
2	18,2	12,3	9,3	10,1	10,4	9,8	10,8	6,3	4,5	4,9	5,3	4,9
3	19,9	12,5	9,2	9,9	10,1	9,5	11,9	7,0	4,4	5,0	5,3	4,7
4	22,5	14,0	9,2	10,1	10,3	9,4	14,2	7,6	4,5	5,0	5,5	4,8
5	25,1	14,9	9,3	10,1	10,4	9,8	16,3	8,5	4,4	5,0	5,5	4,8
6	28,9	16,7	9,0	10,0	9,8	9,7	19,1	9,8	4,2	4,9	5,1	4,8
7	34,7	20,2	9,2	10,2	9,7	9,8	24,4	11,6	4,4	5,0	4,6	4,9
8	39,8	23,2	9,3	10,1	10,6	10,1	28,3	13,9	4,6	5,2	5,3	5,0

$p$	$\alpha = 1\%$						$\alpha = 0,5\%$					
	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **
1	2,5	1,2	0,8	0,9	1,0	0,9	1,4	0,7	0,4	0,4	0,5	0,5
2	3,1	1,5	0,9	1,0	1,1	1,1	1,9	0,8	0,5	0,5	0,6	0,5
3	3,9	1,8	1,0	1,1	1,2	1,1	2,4	1,0	0,5	0,6	0,7	0,6
4	4,8	2,1	0,9	1,1	1,2	1,1	3,1	1,1	0,4	0,5	0,6	0,5
5	5,9	2,3	1,0	1,1	1,2	1,0	3,7	1,3	0,4	0,5	0,7	0,5
6	7,2	2,6	0,9	1,1	1,4	1,2	4,8	1,6	0,4	0,5	0,8	0,8
7	9,7	3,3	0,7	0,9	1,3	1,0	6,3	1,9	0,3	0,4	0,7	0,6
8	12,3	4,3	0,8	1,0	2,1	1,7	8,6	2,5	0,4	0,5	1,4	1,2

Tabela 4.4 *Tamanho dos testes - Modelo normal linear com  $p = 2$ ,  $k = 5$  e diversos valores para  $n$ .*

$n$	$\alpha$ (%)	$LR$	$LR^*$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
30	10,0	20,9	13,3	8,6	9,5	10,4	9,5
	5,0	13,2	7,5	4,1	4,5	5,3	4,6
	1,0	4,6	1,7	0,7	0,8	1,0	0,8
	0,5	2,8	0,9	0,3	0,4	0,5	0,4
35	10,0	20,6	13,9	8,7	9,8	10,8	10,0
	5,0	12,8	7,4	4,1	4,7	5,5	5,0
	1,0	4,2	1,9	0,7	0,8	1,1	1,0
	0,5	2,5	1,0	0,3	0,4	0,6	0,5
40	10,0	18,2	12,5	9,3	10,1	10,4	9,8
	5,0	10,8	6,3	4,5	4,9	5,3	4,9
	1,0	3,1	1,5	0,9	1,0	1,1	1,1
	0,5	1,9	0,8	0,5	0,5	0,6	0,5
45	10,0	17,6	11,9	9,3	9,8	10,2	9,8
	5,0	10,5	6,6	4,4	4,8	5,0	4,8
	1,0	3,1	1,4	0,7	0,9	1,0	0,9
	0,5	1,9	0,8	0,4	0,5	0,5	0,4
100	10,0	12,0	9,8	9,1	9,8	10,1	9,8
	5,0	6,2	5,0	4,6	4,8	4,8	4,8
	1,0	1,6	1,0	0,9	1,0	1,0	1,0
	0,5	0,8	0,5	0,4	0,4	0,4	0,4

Tabela 4.5 *Média e variância da  $\chi_2^2$ ,  $LR$ ,  $LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  - Modelo normal linear com  $n = 40$ ,  $p = 2$  e  $k = 5$ .*

momentos	$\chi_2^2$	$LR$	$LR^*$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
média	2,000	2,673	2,187	1,941	2,005	2,030	1,971
variância	4,000	7,350	4,919	3,838	4,096	4,254	4,092

Tabela 4.6 *Quantis da  $\chi_2^2$ ,  $LR$ ,  $LR^*$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  - Modelo normal linear com  $n = 40$ ,  $p = 2$  e  $k = 5$ .*

quantis (%)	$\chi_2^2$	$LR$	$LR^*$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
85,0	3,8	5,1	4,2	3,7	3,8	3,9	3,8
90,0	4,6	6,1	5,0	4,5	4,6	4,7	4,6
95,0	6,0	8,0	6,5	5,8	6,0	6,1	6,0
97,0	7,0	9,3	7,6	6,7	6,9	7,0	6,9
99,0	9,2	12,3	10,1	8,9	9,2	9,5	9,3
99,5	10,6	14,5	11,8	10,5	10,9	10,9	10,8

Tabela 4.7 Tamanhos dos testes - Modelo normal linear com  $n = 35$ ,  $p = 2$ ,  $k = 5$  e diversas distribuições de covariáveis.

$\alpha$ (%)	$\mathcal{U}(0,1)$						$LN(0,1)$					
	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **
10,0	18,8	12,2	8,7	9,3	10,0	9,3	41,0	30,6	9,0	9,4	11,6	9,3
5,0	11,3	6,7	4,0	4,6	5,1	4,6	30,6	9,2	4,5	4,9	5,0	4,7
1,0	3,5	1,4	0,6	0,8	0,9	0,8	15,8	2,9	0,8	0,9	1,2	0,9
0,5	2,1	0,7	0,3	0,4	0,4	0,3	11,9	1,8	0,4	0,5	0,6	0,5
	$t_3$						$\chi_5^2$					
	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **	LR	LR*	LR <sub>m</sub>	LR <sub>m</sub> *	LR <sub>a</sub> *	LR <sub>a</sub> **
10,0	28,2	17,5	7,2	7,8	10,2	7,9	23,1	14,7	8,6	9,3	10,6	8,7
5,0	18,1	9,7	3,3	3,6	4,5	3,5	15,2	8,4	4,0	4,4	5,4	4,4
1,0	6,5	2,2	0,3	0,7	1,6	0,6	5,9	2,5	0,7	0,8	1,1	0,9
0,5	4,0	1,2	0,2	0,3	0,6	0,3	3,9	1,5	0,3	0,4	0,6	0,5

Tabela 4.8 Poder dos testes - Modelo normal linear com  $n = 40$ ,  $p = 2$ ,  $k = 5$  e  $\alpha = 5\%$ .

$\delta$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
0,0	4,4	4,9	5,2	4,9
0,5	7,7	8,5	9,3	8,7
1,0	21,4	22,4	24,6	23,6
1,5	47,2	48,9	50,3	48,7
2,0	73,2	74,4	75,5	73,9
2,5	88,7	89,6	90,4	90,0
3,0	93,0	93,6	92,5	91,5
3,5	97,6	97,9	97,8	97,5

Tabela 4.9 Poder dos testes - Modelo correto: normal linear com  $n = 35$ ,  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2} + \delta_3 z_{\ell 3})$ ,  $k = 5$  e  $\alpha = 5\%$ .

$\delta$	modelo postulado					
	incorreto			correto		
	$m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2})$			$m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2} + \delta_3 z_{\ell 3})$		
	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_m^{**}$	$LR_a^*$	$LR_a^{**}$
0,0	3,8	4,5	5,0	4,4	3,8	4,5
0,5	9,3	10,3	11,2	10,6	10,3	11,3
1,0	18,6	20,2	18,5	17,0	27,7	29,5
1,5	46,4	48,2	50,8	49,7	45,6	47,3
2,0	69,7	71,6	71,9	70,2	69,7	71,0
2,5	87,7	88,5	87,5	85,7	89,7	90,2
3,0	98,5	98,7	98,4	98,1	98,3	98,5
3,5	99,1	99,2	99,2	99,1	99,7	99,7

Tabela 4.10 Poder dos testes - Modelo postulado: normal linear com  $n = 35$ ,  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2})$ ,  $k = 5$  e  $\alpha = 5\%$ .

$\delta$	especificação correta $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2})$				má especificação $m_t = \exp(\delta_1 z_{t 1})$			
	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
0,0	3,8	4,5	5,0	4,4	3,8	4,5	5,0	4,4
0,5	9,3	10,3	11,2	10,6	6,1	7,0	7,4	6,7
1,0	18,6	20,2	18,5	17,0	9,3	10,2	10,1	9,0
1,5	46,4	48,2	50,8	49,7	27,7	29,2	31,2	30,3
2,0	69,7	71,6	71,9	70,2	42,0	44,1	44,4	42,2
2,5	87,7	88,5	87,5	85,7	75,3	76,7	75,7	74,1
3,0	98,5	98,7	98,4	98,1	81,7	82,9	82,6	80,9
3,5	99,1	99,2	99,2	99,1	84,9	90,2	92,2	91,9

$\delta$	má especificação $m_t = \exp(\delta_1 z_{t 1} + \delta_2 \sqrt{z_{t 2}})$				má especificação $m_t = \exp(\delta_1 z_{t 1} + \delta_2 z_{t 2}^2)$			
	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
0,0	3,8	4,5	5,0	4,4	3,8	4,5	5,0	4,4
0,5	8,2	9,2	10,1	9,4	9,4	10,5	11,5	10,8
1,0	14,3	15,5	14,5	12,9	21,0	22,7	21,1	19,3
1,5	37,5	39,5	41,9	41,0	48,9	50,5	52,8	52,0
2,0	60,2	62,1	62,7	60,8	71,9	73,3	73,8	72,2
2,5	83,6	84,5	83,1	81,1	87,9	88,7	87,8	85,9
3,0	96,2	96,6	96,0	95,3	99,1	99,2	99,1	98,9
2,5	97,5	97,7	97,9	97,6	99,6	99,6	99,6	99,6

Tabela 4.11 Número de rejeições em 100.000 amostras de Monte Carlo - Modelo normal linear com  $n = 40$ ,  $p = 2$  e  $k = 5$ .

$\alpha$ (%)	exato	$LR$	$LR^*$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
50	50000	58929	52507	48482	49647	49979	48847
15	15000	23789	17414	13934	14858	15332	14630
10	10000	17655	12002	9252	9931	10327	9785
5	5000	10521	6358	4536	5003	5262	4938
3	3000	7124	4043	2676	3022	3237	2998
1	1000	3131	1510	850	1009	1132	1031
0,5	500	1942	804	444	513	572	523
0,1	100	580	198	83	106	125	107
0,05	50	358	109	40	50	60	54
0,01	10	115	28	8	12	14	9

Figura 4.1 *Gráfico das discrepâncias relativas de quantis - Modelo normal linear com  $n = 40$ ,  $p = 2$  e  $k = 5$ .*

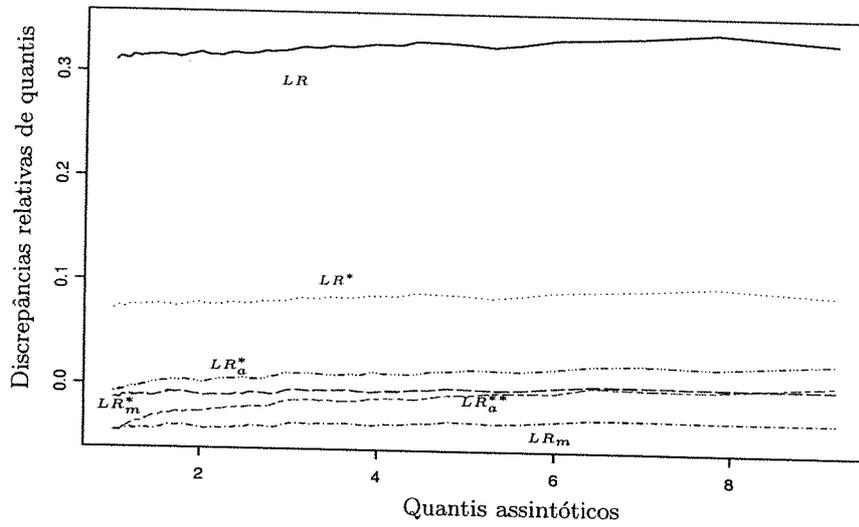
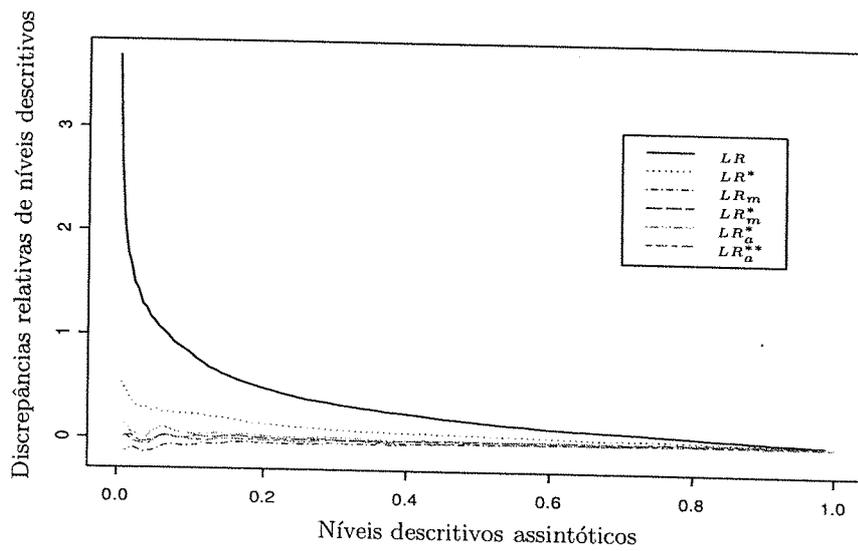


Figura 4.2 *Gráfico das discrepâncias relativas de níveis descritivos - Modelo normal linear com  $n = 40$ ,  $p = 2$  e  $k = 5$ .*



#### 4.4.2 Modelo normal não-linear

Os resultados de simulação apresentados aqui baseiam-se no modelo de regressão normal não-linear (modelo 2) dado no início da Seção 4.4.

As Tabelas 4.12 e 4.13 apresentam os tamanhos dos testes para a situação em que  $n = 40$  com  $p = 2$  e  $k = 2, \dots, 5$  e com  $k = 5$  e  $p = 1, \dots, 5$ , respectivamente. Os comportamentos dos testes foram similares aos observados no caso do modelo normal linear. De forma bem resumida, concluímos que os testes baseados na estatística da razão de verossimilhanças perfiladas modificadas corrigida ( $LR_m^*$ ) e nas estatísticas da razão de verossimilhanças ajustadas ( $LR_a^{**}$  e  $LR_a^*$ ) produzem taxas de rejeição sob  $\mathcal{H}_0$  mais próximas do nível nominal do que os testes baseados na estatística da razão de verossimilhanças ( $LR$ ) e na estatística da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ). Novamente, o teste da razão de verossimilhanças apresenta-se consideravelmente liberal.

Na Tabela 4.14 apresentamos as comparações da média e da variância das estatísticas  $LR$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  e as mesmas medidas da distribuição  $\chi_2^2$ , para o caso em que  $n = 40$ ,  $p = 2$  e  $k = 3$ . Os resultados desta tabela mostram que as médias das estatísticas  $LR_a^*$  e  $LR_m^*$  são bem mais próximas da média de uma variável  $\chi_2^2$  do que as estatísticas  $LR_a^{**}$  e  $LR_m$ , nesta ordem. Adicionalmente, as variâncias das estatísticas  $LR_a^{**}$  e  $LR_m^*$  são mais próximas da variância de uma variável  $\chi_2^2$  do que as estatísticas  $LR_a^*$  e  $LR_m$ , nesta ordem. Similarmente ao caso linear, essas medidas para  $LR$  excedem as da distribuição  $\chi_2^2$ . Este mesmo comportamento foi observado na Tabela 4.15 ao se comparar quantis amostrais de  $LR$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  com os da distribuição  $\chi_2^2$ .

Na Tabela 4.16, fixamos o valor de  $p$  em 2 e variamos o tamanho da amostra ( $n = 30, 35, 40, 45$ ) para  $k = 3, 5$ . Claramente, conforme cresce o tamanho da amostra, as taxas de rejeição de todos os testes aproximam-se dos respectivos níveis nominais. Porém, mesmo para  $n = 45$ , observamos que o teste da razão de verossimilhanças original é liberal. Por exemplo, para este tamanho de amostra com  $k = 5$  e  $\alpha = 10\%$  temos que sua taxa de rejeição excede 15%, enquanto que

as taxas de rejeição dos testes, que se baseiam nas estatísticas  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$ , são 9,1%, 9,6%, 10,0% e 9,7%, respectivamente. As conclusões anteriores tiradas a partir das Tabelas 4.12 e 4.13 também se verificam aqui.

Na Tabela 4.17 são apresentados resultados de simulação para  $n = 35$ ,  $p = 2$  e  $k = 5$ . As covariadas foram obtidas das seguintes distribuições:  $\mathcal{U}(0, 1)$ ,  $t_3$  e  $\chi_5^2$ . Novamente, podemos observar que configurações amostrais das covariadas parecem ter impacto considerável sobre os tamanhos empíricos dos testes.

Da mesma forma que no caso normal linear, também avaliamos a robustez dos testes de heteroscedasticidade à escolha inadequada da função cedástica. As Tabelas 4.18 e 4.19 apresentam resultados de simulação obtidos sob a hipótese alternativa (heteroscedasticidade) para alguns valores de  $\delta$  e  $\alpha = 5\%$ . Analisando as tabelas, concluímos que os desempenhos dos testes foram semelhantes ao caso do modelo de regressão normal linear, ou seja, observamos que não há indícios de que a omissão e a má especificação de uma covariável em  $Z$  prejudique o desempenho dos testes de heteroscedasticidade.

Analogamente ao caso normal linear, avaliamos os tamanhos empíricos dos testes para uma ampla gama de valores do nível nominal (de 0,01% a 50%) para o modelo normal não-linear em que  $n = 40$ ,  $p = 2$  e  $k = 3$ . As simulações para esta situação foram baseadas em 100.000 amostras. Da Tabela 4.20, observa-se que na cauda da distribuição  $\chi_2^2$ , as discrepâncias relativas entre as taxas de rejeição empíricas do teste da razão de verossimilhanças e os níveis nominais correspondentes são muito elevadas, ao contrário do que ocorre na região mediana da distribuição. Por exemplo, se  $\alpha = 0,01\%$ , essa diferença relativa é de 580%  $((68 - 10)/10 \times 100\%)$ , enquanto que, para  $\alpha = 50\%$  esta diferença é de apenas 12%  $((55768 - 50000)/50000 \times 100\%)$ . É interessante notar que esta discrepância é bem menor para as outras estatísticas consideradas. A estatística  $LR_a^{**}$ , por exemplo, apresenta diferenças relativas para  $\alpha = 0,01\%$  e  $\alpha = 50\%$  de 20%  $((12 - 10)/10 \times 100\%)$  e  $-3\%$   $((48743 - 50000)/50000 \times 100\%)$ , respectivamente. Por outro lado, a estatística  $LR_m^*$ , por exemplo, apresenta diferenças

relativas para  $\alpha = 0,01\%$  e  $\alpha = 50\%$  de  $20\%$   $((12 - 10)/10 \times 100\%)$  e  $-0.4\%$   $((49778 - 50000)/50000 \times 100\%)$ , respectivamente.

A Figura 4.3 apresenta o gráfico das discrepâncias relativas entre os quantis amostrais e o quantis assintóticos das estatísticas dos testes contra os quantis assintóticos. Analisando a Figura 4.3, novamente observamos que, para todos os testes, as curvas de discrepâncias de quantis estão próximas de linhas retas paralelas ao eixo horizontal de ordenada zero; tais discrepâncias permanecem, portanto, aproximadamente estáveis, no entanto, em patamares diferentes. A discrepância de quantil para a estatística da razão de verossimilhanças original ( $LR$ ) apresenta-se em torno de  $20\%$ , o que confirma que este teste é liberal. Por outro lado, a estatística da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) apresenta discrepância de quantil em torno de  $-4\%$ , o que confirma que este teste é conservativo. Finalmente, as estatísticas  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  apresentam curvas de discrepâncias de quantis bem próximas da ordenada nula.

A Figura 4.4 apresenta o gráfico das discrepâncias relativas entre níveis descritivos ( $p$ -valores) dos testes e os níveis descritivos assintóticos contra os níveis descritivos assintóticos. Analisando a Figura 4.4, novamente observamos que, quando se caminha em direção ao extremo da cauda da distribuição  $\chi^2$  de referência, as discrepâncias relativas de níveis descritivos para o teste da razão de verossimilhanças tendem a crescer rapidamente. Tal discrepância é bem menor para os outros testes.

Tabela 4.12 Tamanhos dos testes - Modelo normal não-linear com  $n = 40$ ,  $p = 2$  e diversos valores para  $k$ .

$k$	$\alpha = 10\%$					$\alpha = 5\%$				
	$LR$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$	$LR$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
2	12,4	9,0	9,8	9,8	9,7	6,4	4,2	4,7	4,8	4,7
3	14,3	9,0	9,8	10,0	9,6	8,2	4,3	4,8	5,0	4,8
4	16,6	9,2	10,1	10,6	10,2	9,8	4,4	4,9	5,1	4,9
5	17,1	9,2	9,8	10,2	9,7	10,2	4,8	5,1	5,2	4,9

$k$	$\alpha = 1\%$					$\alpha = 0,5\%$				
	$LR$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$	$LR$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
2	1,6	0,8	1,0	1,0	0,9	0,9	0,3	0,4	0,5	0,5
3	2,2	0,8	0,9	1,0	1,0	1,1	0,5	0,5	0,5	0,5
4	2,6	0,8	1,0	1,0	0,9	1,6	0,4	0,4	0,5	0,5
5	3,1	0,9	1,1	1,1	1,0	1,8	0,4	0,5	0,5	0,5

Tabela 4.13 Tamanhos dos testes - Modelo normal não-linear com  $n = 40$ ,  $k = 5$  e diversos valores para  $p$ .

$p$	$\alpha = 10\%$			$\alpha = 5\%$		
	$LR$	$LR_m^*$	$LR_a^{**}$	$LR$	$LR_m^*$	$LR_a^{**}$
1	15,9	8,6	9,5	9,6	9,5	9,5
2	17,1	9,2	9,8	10,2	9,7	9,7
3	20,8	9,1	10,0	10,6	9,8	9,8
4	23,4	9,0	9,9	10,1	9,9	9,9
5	25,4	10,3	10,9	11,2	10,6	10,6

$p$	$\alpha = 1\%$			$\alpha = 0,5\%$		
	$LR$	$LR_m^*$	$LR_a^{**}$	$LR$	$LR_m^*$	$LR_a^{**}$
1	2,9	0,7	0,9	1,0	0,9	0,9
2	3,1	0,9	1,1	1,1	1,0	1,0
3	3,6	0,8	1,0	1,1	1,0	1,0
4	4,4	0,8	0,9	1,1	0,9	0,9
5	5,8	1,0	1,2	1,4	1,2	1,2

Tabela 4.14 Média e variância de  $\chi_2^2$ ,  $LR$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  - Modelo normal não-linear com  $n = 40$ ,  $k = 3$  e  $p = 2$ .

momentos	$\chi_2^2$	$LR$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
média	2,000	2,371	1,907	1,977	1,986	1,947
variância	4,000	5,970	3,727	4,003	4,104	4,056

Tabela 4.15 Quantis de  $\chi_2^2$ ,  $LR$ ,  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$  - Modelo normal não-linear com  $n = 40$ ,  $k = 3$  e  $p = 2$ .

quantis (%)	$\chi_2^2$	$LR$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
85,0	3,8	4,5	3,6	3,7	3,8	3,7
90,0	4,6	5,5	4,4	4,6	4,6	4,5
95,0	6,0	7,1	5,7	5,9	6,0	5,9
97,0	7,0	8,4	6,7	7,0	7,0	6,9
99,0	9,2	10,8	8,8	9,1	9,2	9,2
99,5	10,6	12,9	10,4	10,8	10,8	10,7

Tabela 4.16 Tamanho dos testes - Modelo normal não-linear com  $p = 2$  e  $k = 3, 5$ .

$n$	$\alpha$ (%)	$k = 3$					$k = 5$				
		$LR$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$	$LR$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
30	10,0	18,6	8,5	9,7	10,5	9,8	20,5	8,3	9,2	9,8	9,1
	5,0	11,1	4,1	4,9	5,5	5,1	12,9	4,0	4,5	5,1	4,5
	1,0	3,4	0,6	0,9	1,1	0,9	4,5	0,7	0,9	1,0	0,8
	0,5	2,0	0,3	0,3	0,4	0,4	2,8	0,4	0,5	0,5	0,4
35	10,0	14,9	8,6	9,5	9,8	9,5	18,7	8,9	9,7	10,1	9,6
	5,0	8,5	4,2	4,9	5,1	4,8	11,5	4,5	4,8	5,1	4,8
	1,0	2,4	0,7	1,0	0,9	0,9	3,8	0,8	1,0	1,1	1,0
	0,5	1,3	0,3	0,4	0,4	0,4	2,4	0,4	0,4	0,5	0,5
40	10,0	14,3	9,0	9,8	10,0	9,6	17,1	9,2	9,8	10,2	9,7
	5,0	8,2	4,3	4,8	5,0	4,8	10,2	4,8	5,1	5,2	4,9
	1,0	2,2	0,8	0,9	1,0	1,0	3,1	0,9	1,1	1,1	1,0
	0,5	1,1	0,5	0,5	0,5	0,5	1,8	0,4	0,5	0,5	0,5
45	10,0	13,7	9,2	9,9	10,0	9,8	15,9	9,1	9,6	10,0	9,7
	5,0	7,4	4,2	4,8	4,8	4,7	9,3	4,4	4,8	5,1	4,9
	1,0	1,8	0,9	0,9	1,0	0,9	2,5	0,6	0,7	0,8	0,8
	0,5	1,0	0,5	0,5	0,5	0,5	1,3	0,5	0,5	0,5	0,5

Tabela 4.17 Tamanhos dos testes - Modelo normal não-linear com  $n = 35$ ,  $p = 2$ ,  $k = 5$  e diversas distribuições das covariáveis.

Distribuição	$\alpha$ (%)	LR	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
$t_3$	10,0	26,2	7,5	9,0	11,8	9,4
	5,0	17,5	3,6	4,4	7,0	5,4
	1,0	7,3	0,6	0,8	1,8	1,4
	0,5	4,9	0,3	0,4	1,1	0,7
$\chi_5^2$	10,0	25,9	8,5	9,3	10,2	7,8
	5,0	17,0	4,0	4,6	5,3	4,1
	1,0	6,6	0,7	0,8	1,2	0,9
	0,5	4,3	0,3	0,5	0,6	0,4
$\mathcal{U}(0, 1)$	10,0	19,6	9,1	9,9	10,6	9,5
	5,0	12,0	4,2	4,8	5,6	4,8
	1,0	3,9	0,7	0,8	1,2	0,8
	0,5	2,3	0,3	0,4	0,6	0,4

Tabela 4.18 Poder dos testes - Modelo correto: normal não-linear com  $n = 35$ ,  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2} + \delta_3 z_{\ell 3})$ ,  $k = 3$  e  $\alpha = 5\%$ .

$\delta$	modelo postulado										
	incorreto			correto							
	$m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2})$			$m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2} + \delta_3 z_{\ell 3})$							
$LR_m$	$LR_m^*$	$LR_a^*$	$LR_m^{**}$	$LR_a^*$	$LR_a^{**}$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_m^{**}$	$LR_a^*$	$LR_a^{**}$
0,0	4,5	5,1	5,4	5,1	5,1	5,1	4,5	5,1	5,4	5,1	5,1
0,5	11,8	12,6	12,8	12,5	12,5	12,5	13,4	17,6	17,7	17,7	17,4
1,0	22,6	24,1	23,7	23,1	23,1	23,1	24,8	24,1	23,7	23,7	23,1
1,5	52,6	54,2	54,5	54,2	54,2	54,2	53,4	55,3	55,8	55,8	55,2
2,0	78,0	78,9	80,6	80,6	80,6	80,6	78,6	78,9	81,6	81,6	81,6
2,5	91,2	91,9	91,6	90,6	90,6	90,6	93,2	93,9	93,6	93,6	93,6

Tabela 4.19 Poder dos testes - Modelo postulado: normal não-linear com  $n = 35$ ,  $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2})$ ,  $k = 3$  e  $\alpha = 5\%$ .

$\delta$	especificação correta $m_\ell = \exp(\delta_1 z_{\ell 1} + \delta_2 z_{\ell 2})$				má especificação $m_t = \exp(\delta_1 z_{t 1})$			
	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
0,0	4,5	5,1	5,4	5,1	4,5	4,9	5,3	5,1
0,5	11,8	12,6	12,8	12,5	8,9	9,8	9,5	9,2
1,0	22,6	24,1	23,7	23,1	20,1	21,2	21,6	21,4
1,5	52,6	54,2	54,5	54,2	43,1	45,0	46,3	45,9
2,0	78,0	78,9	80,6	80,6	62,9	65,0	65,4	64,5
2,5	91,2	91,9	91,6	90,6	89,9	90,4	90,5	90,6

$\delta$	má especificação $m_t = \exp(\delta_1 z_{t 1} + \delta_2 \sqrt{z_{t 2}})$				má especificação $m_t = \exp(\delta_1 z_{t 1} + \delta_2 z_{t 2}^2)$			
	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
0,0	4,5	5,0	5,2	5,0	4,6	5,0	5,3	5,1
0,5	9,4	10,4	10,4	10,1	32,1	33,5	29,3	27,9
1,0	27,3	28,7	30,3	30,1	54,4	56,2	58,1	57,0
1,5	36,9	39,0	41,3	40,5	92,5	93,3	90,8	88,8
2,0	60,5	62,6	66,6	66,2	98,9	99,1	98,5	98,2
2,5	89,7	90,4	90,5	90,0	100,0	100,0	100,0	100,0

Tabela 4.20 Número de rejeições em 100.000 amostras de Monte Carlo - Modelo normal não-linear com  $n = 40$ ,  $p = 2$  e  $k = 3$ .

$\alpha \%$	exato	$LR$	$LR_m$	$LR_m^*$	$LR_a^*$	$LR_a^{**}$
50	50000	55768	48536	49778	48744	48743
15	15000	20333	13944	14887	14765	14764
10	10000	14475	9191	10016	9944	9943
5	5000	8170	4457	4935	4989	4988
3	3000	5307	2647	2965	2963	2962
1	1000	2138	834	990	997	996
0,5	500	1200	428	531	523	522
0,1	100	342	77	98	100	99
0,05	50	204	40	51	52	51
0,01	10	68	10	12	13	12

Figura 4.3 Gráfico das discrepâncias relativas de quantis - Modelo normal não-linear com  $n = 40$ ,  $p = 2$  e  $k = 3$ .

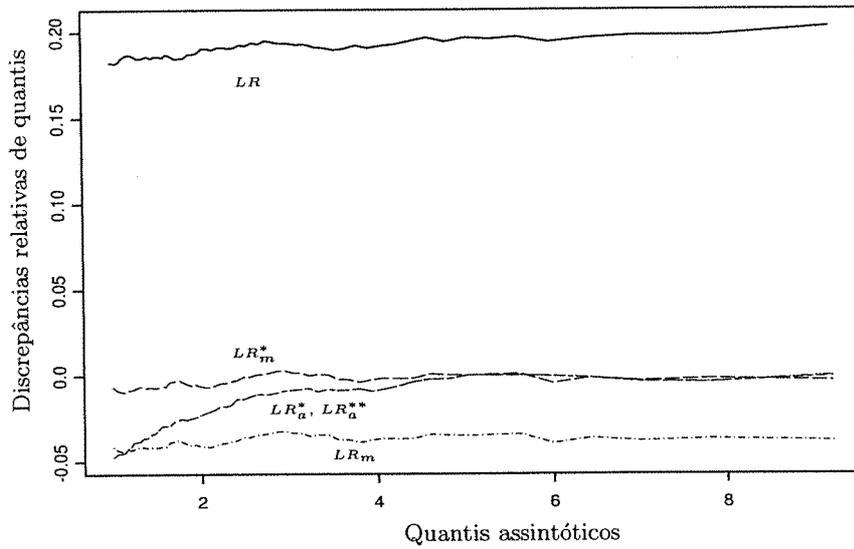
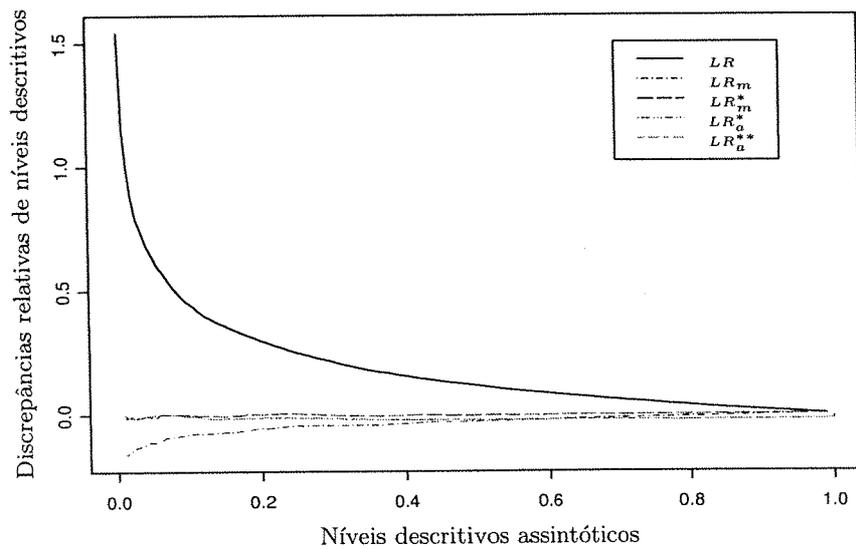


Figura 4.4 Gráfico das discrepâncias relativas de níveis descritivos - Modelo normal linear com  $n = 40$ ,  $p = 2$  e  $k = 3$ .



Finalmente, cabe um comentário acerca da robustez dos testes à não-normalidade da variável resposta. Um breve estudo de simulação para os dois modelos, omitido aqui, sugere que todos os testes considerados são sensíveis à ausência de normalidade, apresentando tamanhos empíricos bem superiores aos correspondentes níveis nominais. No entanto, as distorções de tamanho são bem menores para os testes ajustados do que para o teste da razão de verossimilhanças original.

Resumindo, em todas as situações observamos que o teste da razão de verossimilhanças (baseado em  $LR$ ) é muito liberal, apresentando taxas de rejeição sob  $\mathcal{H}_0$  bem acima dos níveis nominais, enquanto que o teste da razão de verossimilhanças corrigido (baseado em  $LR^*$ ) apresenta taxas de rejeição mais próximas dos níveis nominais, porém ainda acima deste. Por outro lado, o teste da razão de verossimilhanças perfiladas modificadas (baseado em  $LR_m$ ) foi eficiente em corrigir a distorção de tamanho do teste original mas apresenta taxas de rejeição consideravelmente abaixo dos níveis nominais. A correção de Bartlett aplicada à estatística  $LR_m$  proporciona ao teste resultante (baseado em  $LR_m^*$ ) uma maior redução na distorção do tamanho, produzindo taxas próximas dos níveis nominais. Os testes baseados nas estatísticas  $LR_a^{**}$  e  $LR_a^*$  produzem taxas de rejeição mais próximas dos níveis nominais do que os testes baseados nas estatísticas  $LR$  e  $LR^*$ . Comparando os testes baseados em  $LR_a^*$  e  $LR_a^{**}$  com o teste da razão de verossimilhanças perfiladas modificadas corrigido (baseado em  $LR_m^*$ ), podemos observar que os testes que usam as estatísticas  $LR_m^*$  e  $LR_a^{**}$  têm desempenhos bem semelhantes e ligeiramente melhores que o teste baseado em  $LR_a^*$ . Dentre todos os testes apresentados, os que mais se destacaram foram os testes da razão de verossimilhanças perfiladas modificadas corrigido (baseado em  $LR_m^*$ ) e o teste da razão de verossimilhanças ajustado (baseado em  $LR_a^{**}$ ).

Concluindo, os resultados das simulações obtidos para os dois modelos considerados sugerem que a estatística da razão de verossimilhanças original e sua versão corrigida utilizadas em conjunto com uma aproximação por  $\chi^2$  podem produzir inferências inadequadas em amostras pequenas ou mesmo de tamanho moderado, enquanto que as estatísticas  $LR_m$ ,  $LR_m^*$ ,  $LR_a^{**}$  e  $LR_a^*$  podem ser utilizadas para

obter testes mais precisos em amostras finitas. Destacamos o excelente desempenho dos testes da razão de verossimilhanças perfiladas modificadas corrigido (baseado em  $LR_m^*$ ) e da razão de verossimilhanças ajustado (baseado em  $LR_a^{**}$ ). Vale salientar que o teste baseado em  $LR_a^{**}$  apresenta algumas vantagens em relação ao teste baseado em  $LR_m^*$ , a saber: é invariante segundo reparametrizações da forma  $(\nu, \psi) \rightarrow (\nu, \zeta(\nu, \psi))$ , não necessita maximizar uma função de verossimilhança perfilada modificada e também não requer ortogonalidade entre os parâmetros de interesse e os de perturbação. Enfatizamos também que, para a estatística  $LR_a^{**}$ , o ajuste é aplicado diretamente à estatística da razão de verossimilhanças enquanto que o ajuste de Cox e Reid (1987), ao contrário, é aplicado à função de verossimilhança perfilada, a qual deve ser maximizada sob o modelo restrito e irrestrito para, posteriormente, obter-se a estatística ajustada. A estatística  $LR_a^{**}$  envolve cumulantes de derivadas do logaritmo da função de verossimilhanças até segunda ordem, enquanto que a estatística  $LR_m^*$  requer cumulantes de derivadas do logaritmo da função de verossimilhanças até quarta ordem.

#### 4.5 Aplicações

Apresentamos a seguir aplicações dos resultados das seções anteriores a quatro conjuntos de dados<sup>(4)</sup>.

##### 4.5.1 Exemplo 1

Este exemplo refere-se a serviços de manutenção e reposição de latas e garrafas de bebidas em máquinas de venda automáticas. A variável resposta  $y$  é o tempo (em minutos) gasto no serviço, as covariadas  $x_2$  e  $x_3$  utilizadas são, respectivamente, o número de bebidas estocadas e a distância percorrida (em pés); ver Apêndice F. Utilizamos o seguinte modelo:

$$y_\ell = \beta_1 + x_{\ell 2}\beta_2 + x_{\ell 3}\beta_3 + u_\ell, \quad \ell = 1, \dots, 23.$$

<sup>(4)</sup> Os exemplos considerados aqui foram tratados também nas Seções 2.5 e 3.6. Nos exemplos 1 e 2, algumas observações foram omitidas conforme descrito nessas seções.

Assumimos que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{x_{\ell 2}\delta_1 + x_{\ell 3}\delta_2\})$  e  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$ . Testamos a hipótese  $\mathcal{H}_0 : \delta_1 = \delta_2 = 0$  (indicando homoscedasticidade) contra uma alternativa bilateral (heteroscedasticidade). Para este teste, temos que  $LR = 4,825$ ,  $LR^* = 3,717$ ,  $LR_m = 4,126$ ,  $LR_m^* = 4,352$ ,  $LR_a^* = 3,396$  e  $LR_a^{**} = 3,271$  conduzem aos respectivos níveis descritivos 0,090, 0,156, 0,127, 0,113, 0,183 e 0,195. A hipótese nula é rejeitada ao nível nominal de 10% quando o teste é baseado na estatística da razão de verossimilhanças ( $LR$ ), ao contrário do que ocorre quando o teste usa qualquer das outras estatísticas.

#### 4.5.2 Exemplo 2

Este exemplo refere-se a taxas mensais de retorno de ações de mercado ( $x$ ) e da Corporação Acme Cleveland ( $y$ ) no período de janeiro de 1986 a dezembro de 1990 (ver Apêndice G). Consideramos o seguinte modelo

$$y_\ell = \beta_1 + \beta_2 x_\ell + u_\ell, \quad \ell = 1, \dots, 59.$$

Assumimos que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{\delta x_\ell\})$  e  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$ . O objetivo é testar a hipótese  $\mathcal{H}_0 : \delta = 0$  contra  $\mathcal{H}_1 : \delta \neq 0$ , ou seja, se o modelo é homoscedástico. Para este teste, temos que  $LR = 3,329$ ,  $LR^* = 3,119$ ,  $LR_m = 2,968$ ,  $LR_m^* = 3,071$ ,  $LR_a^* = 3,095$  e  $LR_a^{**} = 3,091$  conduzem aos respectivos níveis descritivos 0,068, 0,077, 0,085, 0,080, 0,078 e 0,079. Assim, todos os testes considerados levam à rejeição da hipótese nula ao nível nominal de 10% mas não rejeitam  $\mathcal{H}_0$  ao nível de 5%, sugerindo evidência marginal de heteroscedasticidade.

#### 4.5.3 Exemplo 3

Este exemplo refere-se aos pesos  $y$  das lentes dos olhos de coelhos europeus (em mg), e à idade  $x$  do animal (em dias), numa amostra de 71 observações (ver Apêndice H). O modelo utilizado é dado por

$$\log(y_\ell) = \beta_1 - \frac{\beta_2}{x_\ell + \beta_3} + u_\ell, \quad \ell = 1, \dots, 71.$$

Assumimos que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{\delta x_\ell\})$  e  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$ . O objetivo é testar a hipótese nula  $\mathcal{H}_0 : \delta = 0$  contra a alternativa  $\mathcal{H}_1 : \delta \neq 0$ , ou

seja, verificar se o modelo é homoscedástico. Aqui, as estatísticas  $LR = 10,250$ ,  $LR_m = 9,556$ ,  $LR_m^* = 9,776$ ,  $LR_a^* = 10,839$  e  $LR_a^{**} = 10,831$  conduzem aos respectivos níveis descritivos 0,001, 0,001, 0,002, 0,001 e 0,001. Portanto, todos os testes conduzem à mesma conclusão, isto é, levam à rejeição da hipótese nula aos níveis de significância usuais.

#### 4.5.4 Exemplo 4

Neste exemplo a variável resposta é a quantidade de cálcio radioativo  $y$  (em nmoles/mg) e a variável explicativa é o tempo  $x$  (em minutos) após as células terem sido suspensas na solução de cálcio, numa amostra de 27 observações (ver Apêndice I). Consideramos o seguinte modelo

$$y_\ell = \beta_0 \{1 - \exp(-\beta_1 x_\ell)\} + u_\ell, \quad \ell = 1, \dots, 27.$$

Assumimos que  $u_\ell \sim \mathcal{N}(0, \sigma^2 \exp\{\delta x_\ell\})$  e  $\text{Cov}(u_\ell, u_m) = 0$  para todo  $\ell \neq m$ . O objetivo é testar a hipótese  $\mathcal{H}_0 : \delta = 0$  contra  $\mathcal{H}_1 : \delta \neq 0$ , ou seja, verificar se o modelo é homoscedástico. Os valores observados para as estatísticas são  $LR = 3,097$ ,  $LR_m = 2,914$ ,  $LR_m^* = 3,056$ ,  $LR_a^* = 3,169$  e  $LR_a^{**} = 3,168$ . Estes conduzem aos respectivos níveis descritivos 0,078 0,088 0,080, 0,075 e 0,075. Todos os testes considerados levam então à rejeição da hipótese nula ao nível nominal de 10% mas não ao nível nominal de 5%, sugerindo evidência marginal de heteroscedasticidade.

## 4.6 Comentários

Neste capítulo, obtivemos dois ajustes para a estatística da razão de verossimilhanças que podem ser aplicados a todos os modelos não-lineares da família exponencial, ao contrário da modificação de Cox e Reid (1987). Estes ajustes apresentam uma vantagem em relação ao ajuste de Cox e Reid (1987) por serem invariantes segundo reparametrizações da forma  $(\nu, \psi) \longrightarrow (\nu, \zeta(\nu, \psi))$ . Outra vantagem é que esses novos ajustes são aplicados diretamente à estatística da razão de verossimilhanças. O ajuste de Cox e Reid (1987), ao contrário, é aplicado à função de verossimilhança perfilada a qual deve ser maximizada sob o modelo restrito (sob

$\mathcal{H}_0$ ) e irrestrito para, posteriormente, obter-se a estatística ajustada. Quando os estimadores de máxima verossimilhança restritos dos parâmetros de perturbação não têm forma fechada, estas maximizações podem ser computacionalmente mais complexas. Vale salientar ainda uma vantagem adicional de se utilizar os testes baseados nas estatísticas da razão de verossimilhanças ajustadas  $LR_a^*$  e  $LR_a^{**}$ : não é preciso obter uma parametrização ortogonal, enquanto que para os testes baseados na estatística da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) e sua versão corrigida ( $LR_m^*$ ) a ortogonalização é imprescindível.

Resultados de simulação foram obtidos para os modelos de regressão normal linear e não-linear e conduziram a diversas conclusões importantes a respeito do comportamento dos testes de heteroscedasticidade em amostras pequenas e de tamanho moderado. Concluímos destes resultados que ao utilizarmos o teste baseado na estatística da razão de verossimilhanças perfiladas modificadas corrigida ( $LR_m^*$ ) e na estatística da razão de verossimilhanças ajustada ( $LR_a^{**}$ ) e sua versão equivalente ( $LR_a^*$ ), as taxas de rejeição obtidas sob  $\mathcal{H}_0$  são bem mais próximas do nível nominal do que se usarmos o teste da razão de verossimilhanças original, este sendo notavelmente liberal já que, em todas as situações consideradas (tanto no modelo normal linear quanto no modelo normal não-linear), as taxas de rejeição excedem consideravelmente os níveis nominais correspondentes.

No caso do modelo de regressão normal linear, vimos que o fator de correção de Bartlett aplicado à estatística da razão de verossimilhanças ( $LR$ ) produz testes com taxas de rejeição mais próximas dos níveis nominais, embora ainda sejam consideravelmente liberais.

A modificação de Cox e Reid para a verossimilhança perfilada atenua o efeito do número de parâmetros de perturbação, mas a aproximação pela distribuição  $\chi_p^2$  continua sendo de ordem  $n^{-1}$ . O desempenho do teste da razão de verossimilhanças perfiladas modificadas corrigido (baseado em  $LR_m^*$ ) é semelhante ao do teste da razão de verossimilhanças ajustado (baseado em  $LR_a^{**}$ ). Devemos ressaltar ainda que os testes baseados nas estatísticas  $LR_a^{**}$  e  $LR_a^*$  são mais simples de serem realizados do que aqueles que envolvem correção de Bartlett, uma vez que

estes só dependem de cumulantes do logaritmo da função de verossimilhanças até segunda ordem, enquanto que o fator de correção de Bartlett envolve cumulantes do logaritmo da função de verossimilhanças até quarta ordem.

Dentre todos os testes apresentados na Seção 4.4, os que apresentaram melhores desempenhos foram os testes baseados nas estatísticas  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$ .

Finalmente, deve ser enfatizado que poucos resultados acerca da eficácia dos ajustes de Skovgaard (2001) no caso em que o parâmetro de interesse é multidimensional são encontrados na literatura. Nosso trabalho veio a preencher parcialmente esta lacuna. Deve ainda ser destacado que, nem mesmo no caso unidimensional, estudos de poder dos testes são encontrados. Nosso trabalho, por outro lado, envolve não só estudos de simulação de tamanho dos testes, mas também de poder. Adicionalmente, foram considerados também aspectos de robustez dos testes.

---

## CAPÍTULO 5

### Conclusões e propostas para pesquisas futuras

---

Resumimos as principais contribuições teóricas desta tese nos seguintes itens:

- (i) No Capítulo 2, derivamos uma expressão bastante geral que pode ser usada para aperfeiçoar o teste da razão de verossimilhanças perfiladas modificadas ( $LR_m$ ) em diferentes classes de modelos em que há a ortogonalidade dos parâmetros e que utilizam a mesma partição do vetor de parâmetros considerada nesta tese. Obtivemos um fator de correção de Bartlett para um teste de heteroscedasticidade baseado em uma verossimilhança perfilada modificada proposta por Cox e Reid (1987), no modelo de regressão normal linear. Os resultados cobrem a situação em que o parâmetro de interesse que define o comportamento de heteroscedasticidade é multidimensional, generalizando assim o artigo de Ferrari e Cribari-Neto (2002).
- (ii) No Capítulo 3, desenvolvemos uma correção de Bartlett para o teste de heteroscedasticidade baseado na estatística da razão de verossimilhanças perfiladas modificadas em modelos não-lineares da família exponencial, mais especificamente nos modelos normal e normal inverso. Estendemos portanto, o trabalho citado no item anterior, que é restrito ao modelo de regressão normal linear. Em outras palavras, desenvolvemos uma correção para o teste proposto por Wei, Shi, Fung e Hu (1998), obtendo assim um teste da razão de verossimilhanças perfiladas modificadas corrigido, que é preciso até segunda ordem.
- (iii) Nos trabalhos citados nos dois itens acima, foi utilizada a abordagem da verossimilhança perfilada modificada proposta por Cox e Reid (1987), que requer uma parametrização ortogonal. Devido à dificuldade de se obter tal parametrização para o modelo gama, os resultados se restringiram a testes de heteroscedastici-

dade nos modelos não-lineares normais e normais inversos. Esta limitação não existe se utilizarmos a proposta de Skovgaard (2001). É este o enfoque abordado no Capítulo 4, onde desenvolvemos estatísticas da razão de verossimilhanças ajustadas para os modelos não-lineares da família exponencial. É de se destacar que, neste contexto, nossos resultados não se limitam a testes de heteroscedasticidade, mas abrangem testes de quaisquer hipóteses nulas que estabelecem um vetor de valores fixados para uma parte do vetor de parâmetros do modelo.

Além dessas contribuições, podemos tirar as seguintes conclusões:

- (a) O estudo dos tamanhos dos testes indicou que o teste da razão de verossimilhanças original (baseado em  $LR$ ) é, em geral, anticonservativo, conduzindo a taxas de rejeição da hipótese nula de homoscedasticidade erroneamente com uma probabilidade maior que o nível de significância nominal do teste. O teste da razão de verossimilhanças corrigido através de correção de Bartlett (baseado em  $LR^*$ ) traz a taxa de rejeição empírica para um valor mais próximo do nível nominal do teste, porém, não corrige totalmente a tendência liberal do teste da razão de verossimilhanças em rejeitar demasiadamente a hipótese nula. Por outro lado, o teste da razão de verossimilhanças perfiladas modificado (baseado  $LR_m$ ) mostrou-se eficiente corrigindo a tendência do teste original de rejeitar com frequência demasiada a hipótese nula, exibindo taxas de rejeição menores que o nível nominal do teste. O teste da razão de verossimilhanças perfiladas modificadas corrigido (baseado em  $LR_m^*$ ) leva as taxas de rejeição para valores ainda mais próximos dos níveis nominais. O teste da razão de verossimilhanças ajustado (baseado em  $LR_a^*$ ) e sua versão assintoticamente equivalente ( $LR_a^{**}$ ) mostraram-se eficientes corrigindo a tendência do teste original em rejeitar com maior frequência a hipótese nula, exibindo taxas de rejeição próximas do nível nominal do teste. Comparando os testes baseados nas estatísticas  $LR_a^*$ ,  $LR_a^{**}$  e  $LR_m^*$  observamos que os testes que usam as estatísticas  $LR_m^*$  e  $LR_a^{**}$  apresentaram taxas de rejeição bem semelhantes e mais próximas do nível nominal,

enquanto que o teste baseado na estatística  $LR_a^*$  apresenta taxas de rejeição ligeiramente acima do nível nominal.

(b) Em relação aos estudos de simulação realizados com o objetivo de analisar a influência do número de parâmetros de perturbação e de interesse no desempenho dos testes, concluímos que o número de parâmetros de perturbação e de interesse têm impacto considerável sobre a aproximação por  $\chi^2$  para a distribuição das estatísticas da razão de verossimilhanças ( $LR$ ) e da razão de verossimilhanças corrigida ( $LR^*$ ). É importante ressaltar que o impacto é bem menos marcante no caso dos testes baseados nas estatísticas  $LR_m$ ,  $LR_m^*$ ,  $LR_a^*$  e  $LR_a^{**}$ . Para estes testes, as taxas de rejeição permanecem mais estáveis do que para os testes baseados nas estatísticas  $LR$  e  $LR^*$ .

(c) Nas simulações envolvendo o poder dos testes, os testes da razão de verossimilhanças perfiladas modificadas corrigido ( $LR_m^*$ ) e da razão de verossimilhanças ajustado ( $LR_a^{**}$ ) apresentaram vantagens em relação aos testes da razão de verossimilhanças perfiladas modificado ( $LR_m$ ) e da razão de verossimilhanças ajustado ( $LR_a^*$ ), uma vez que apresentaram menor distorção de tamanho.

(d) Em suma, entre todas as estatísticas apresentadas nesta tese, aquelas que produziram melhores testes de heteroscedasticidade foram  $LR_m^*$  e  $LR_a^{**}$ . Vale ressaltar que esses dois testes apresentam comportamentos bem semelhantes. A vantagem de se utilizar o teste baseado na estatística da razão de verossimilhanças ajustada ( $LR_a^{**}$ ) é que não é preciso obter uma parametrização ortogonal, enquanto que, para o teste baseado na estatística da razão de verossimilhanças perfiladas modificadas corrigida ( $LR_m^*$ ) a ortogonalização é imprescindível. Adicionalmente, o teste baseado em  $LR_a^{**}$  pôde ser aplicado a todos os modelos não-lineares da família exponencial, enquanto que o teste baseado em  $LR_m^*$  somente pôde ser aplicado para os modelos não-lineares normal e normal inverso.

Várias linhas de pesquisa devem ser desenvolvidas no futuro, tais como:

1. Obter ajustes para verossimilhanças perfiladas e, se possível, obter fatores de correção de Bartlett para testes de heteroscedasticidade baseados nas estatísticas

- da razão de verossimilhanças derivadas de tais funções em modelos lineares generalizados com superdispersão (Cordeiro e Botter, 2001).
2. Obter ajustes para verossimilhanças perfiladas em modelos de regressão não-lineares heteroscedásticos que permitam que tanto a média quanto a variância da variável aleatória  $y$  dependam do parâmetro de interesse.
  3. Desenvolver subrotinas em S-Plus (Becker, Chambers e Wilk, 1988 e Chambers e Hastie, 1992) e R (Ihaka e Gentleman, 1996) para o cálculo das estatísticas propostas nos modelos não-lineares heteroscedásticos, abordando o caso em que o parâmetro que define o comportamento de heteroscedasticidade é multidimensional.
  4. Elaborar estudos de simulação abordando os modelos não-lineares da família exponencial assumindo uma distribuição gama para a variável resposta.

Ao finalizar este trabalho, esperamos ter dado uma contribuição relevante à área de pesquisa que trata de refinamentos de métodos assintóticos em inferência estatística.

---

APÊNDICE A

**Alguns cumulantes no modelo normal linear heteroscedástico**

---

Neste apêndice, apresentamos a obtenção de alguns cumulantes conjuntos de derivadas do logaritmo da função de verossimilhanças do modelo normal linear heteroscedástico necessários ao cálculo da quantidade  $c_m$  que define a correção de Bartlett para a estatística  $LR_m$ .

O logaritmo da função de verossimilhanças do vetor de parâmetros  $\theta = (\delta^\top, \beta^\top, \gamma)^\top$  do modelo normal linear heteroscedástico definido por (2.1) tem a forma

$$L^*(y; \delta, \beta, \gamma) = -\frac{n}{2} \log(\gamma) - \frac{1}{2\gamma} (y - X\beta)^\top Q (y - X\beta) + \text{constante},$$

em que  $Q = \text{diag}\{q_1, \dots, q_n\}$ , com  $q_\ell$  dado em (2.4).

Por simples diferenciação em relação aos componentes do parâmetro  $\delta$ , temos que para  $a = 1, \dots, p$

$$\frac{\partial L^*}{\partial \delta^a} = -\frac{1}{2\gamma} \sum_{\ell=1}^n \frac{q_{\ell a}}{\gamma} d_\ell, \quad (\text{A.1})$$

em que  $d_\ell = (y_\ell - \mu_\ell)^2$  ou, em forma matricial,

$$\frac{\partial L^*}{\partial \delta^a} = -\frac{1}{2\gamma} (y - X\beta)^\top Q_{R_1} (y - X\beta),$$

sendo

$$Q_{R_s} = \text{diag} \left\{ \frac{\partial^s q_1}{\partial \delta^{a_1} \dots \partial \delta^{a_s}}, \dots, \frac{\partial^s q_n}{\partial \delta^{a_1} \dots \partial \delta^{a_s}} \right\},$$

$s = 1, \dots, 4$ . De forma análoga, temos que para  $a, b, c, d = 1, \dots, p$

$$\frac{\partial^2 L^*}{\partial \delta^a \partial \delta^b} = -\frac{1}{2\gamma} (y - X\beta)^\top Q_{R_2} (y - X\beta),$$

$$\frac{\partial^3 L^*}{\partial \delta^a \partial \delta^b \partial \delta^c} = -\frac{1}{2\gamma} (y - X\beta)^\top Q_{R_3} (y - X\beta)$$

e

$$\frac{\partial^4 L^*}{\partial \delta^a \partial \delta^b \partial \delta^c \partial \delta^d} = -\frac{1}{2\gamma} (y - X\beta)^\top Q_{R_4} (y - X\beta).$$

Tomando as esperanças das quatro expressões acima, encontram-se os cumulantes

$$\begin{aligned} \lambda_a &= E\left(\frac{\partial L^*}{\partial \delta^a}\right) = -\frac{1}{2} \text{tr}(Q_{R_1} Q^{-1}) = -\frac{1}{2} \sum_{\ell=1}^n \frac{q_{\ell a}}{q_\ell} = 0, \\ \lambda_{ab} &= E\left(\frac{\partial^2 L^*}{\partial \delta^a \partial \delta^b}\right) = -\frac{1}{2} \text{tr}(Q_{R_2} Q^{-1}) = -\frac{1}{2} \sum_{\ell=1}^n \frac{q_{\ell ab}}{q_\ell}, \\ \lambda_{abc} &= E\left(\frac{\partial^3 L^*}{\partial \delta^a \partial \delta^b \partial \delta^c}\right) = -\frac{1}{2} \text{tr}(Q_{R_3} Q^{-1}) = -\frac{1}{2} \sum_{\ell=1}^n \frac{q_{\ell abc}}{q_\ell}, \\ \lambda_{abcd} &= E\left(\frac{\partial^4 L^*}{\partial \delta^a \partial \delta^b \partial \delta^c \partial \delta^d}\right) = -\frac{1}{2} \text{tr}(Q_{R_4} Q^{-1}) = -\frac{1}{2} \sum_{\ell=1}^n \frac{q_{\ell abcd}}{q_\ell}, \end{aligned}$$

em que

$$q_{\ell a} = \frac{\partial q_\ell}{\partial \delta^a}, \quad q_{\ell ab} = \frac{\partial^2 q_\ell}{\partial \delta^a \partial \delta^b}, \quad q_{\ell abc} = \frac{\partial^3 q_\ell}{\partial \delta^a \partial \delta^b \partial \delta^c} \quad \text{e} \quad q_{\ell abcd} = \frac{\partial^4 q_\ell}{\partial \delta^a \partial \delta^b \partial \delta^c \partial \delta^d}.$$

Diferenciando  $\lambda_{ab}$  em relação a  $\delta^c$ ,  $\lambda_{ab}$  em relação a  $\delta^c \delta^d$  e  $\lambda_{abc}$  em relação a  $\delta^d$  verificamos, respectivamente, que

$$\begin{aligned} (\lambda_{ab})_c &= -\frac{1}{2} \sum_{\ell=1}^n \frac{\partial}{\partial \delta^c} \left( \frac{q_{\ell ab}}{q_\ell} \right), \quad (\lambda_{ab})_{cd} = -\frac{1}{2} \sum_{\ell=1}^n \frac{\partial^2}{\partial \delta^c \partial \delta^d} \left( \frac{q_{\ell ab}}{q_\ell} \right) \quad \text{e} \\ (\lambda_{abc})_d &= -\frac{1}{2} \sum_{\ell=1}^n \frac{\partial}{\partial \delta^d} \left( \frac{q_{\ell abc}}{q_\ell} \right). \end{aligned}$$

Para obter os cumulantes em relação ao parâmetro  $\gamma$ , temos

$$\frac{\partial L^*}{\partial \gamma} = \frac{1}{2\gamma^2} \sum_{\ell=1}^n q_\ell d_\ell - \frac{n}{2\gamma} \tag{A.2}$$

ou, em notação matricial,

$$\frac{\partial L^*}{\partial \gamma} = \frac{1}{2\gamma^2} (y - X\beta)^\top Q (y - X\beta) - \frac{n}{2\gamma}.$$

Analogamente, obtemos

$$\frac{\partial^2 L^*}{\partial \gamma^2} = -\frac{1}{\gamma^3} (y - X\beta)^\top Q (y - X\beta) + \frac{n}{2\gamma^2}.$$

Tomando as esperanças das duas expressões acima, encontram-se

$$\begin{aligned}\lambda_\gamma &= E\left(\frac{\partial L^*}{\partial \gamma}\right) = \frac{1}{2\gamma} \text{tr}(QQ^{-1}) - \frac{n}{2\gamma} = 0 \quad \text{e} \\ \lambda_{\gamma\gamma} &= E\left(\frac{\partial^2 L^*}{\partial \gamma^2}\right) = -\frac{1}{\gamma^2} \text{tr}(QQ^{-1}) + \frac{n}{2\gamma^2} = -\frac{n}{2\gamma^2}.\end{aligned}$$

Invertendo  $\lambda_{\gamma\gamma}$  e diferenciando  $\lambda_{\gamma\gamma}$  em relação a  $\gamma$ , verificamos que

$$\lambda^{\gamma\gamma} = -\frac{2\gamma^2}{n} \quad \text{e} \quad (\lambda_{\gamma\gamma})_\gamma = \frac{n}{\gamma^3},$$

respectivamente.

Para obter os cumulantes mistos observamos que, para  $a, b = 1, \dots, p$  e  $i = 1, \dots, k$ ,

$$\begin{aligned}\frac{\partial^3 L^*}{\partial \delta^a \partial \delta^b \partial \gamma} &= \frac{1}{2\gamma^2} (y - X\beta)^\top Q_{R_2} (y - X\beta) \quad \text{e} \\ \frac{\partial^3 L^*}{\partial \delta^a \partial \delta^b \partial \beta^i} &= \frac{1}{2\gamma} X^\top Q_{R_2} (y - X\beta).\end{aligned}$$

Tomando a esperança das duas últimas expressões acima, obtemos

$$\begin{aligned}\lambda_{ab\gamma} &= E\left(\frac{\partial^3 L^*}{\partial \delta^a \partial \delta^b \partial \gamma}\right) = \frac{1}{2\gamma} \text{tr}(Q_{R_2} Q^{-1}) = \frac{1}{2\gamma} \sum_{\ell=1}^n \frac{q_{\ell ab}}{q_\ell} \quad \text{e} \\ \lambda_{abi} &= E\left(\frac{\partial^3 L^*}{\partial \delta^a \partial \delta^b \partial \beta^i}\right) = 0.\end{aligned}$$

Diferenciando  $\lambda_{ab\gamma}$  em relação a  $\gamma$  temos, finalmente,

$$(\lambda_{ab\gamma})_\gamma = -\frac{1}{2\gamma^2} \sum_{\ell=1}^n \frac{q_{\ell ab}}{q_\ell}.$$

---

## APÊNDICE B

# Matriz de informação total de Fisher em modelos não-lineares heteroscedásticos da família exponencial

---

Neste apêndice, apresentamos a obtenção da matriz de informação total de Fisher através dos cálculos da primeira e segunda derivadas do logaritmo da função de verossimilhanças do vetor de parâmetros  $\omega = (\delta^\top, \beta^\top, \sigma^2)^\top$  dos modelos não-lineares heteroscedásticos da família exponencial definidos nas Seções 3.2 e 4.3, cuja forma é dada por

$$L(y; \delta, \beta, \sigma^2) = -\frac{1}{2} \sum_{\ell=1}^n \left\{ \frac{1}{\sigma^2 m_\ell} d_\ell + t(y_\ell) + s(\phi_\ell) \right\}.$$

Por diferenciação em relação aos componentes do parâmetro  $\delta$ , temos que

$$\frac{\partial L}{\partial \delta^a} = \frac{1}{2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \left\{ d_\ell + \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) \right\}$$

e

$$\begin{aligned} \frac{\partial^2 L}{\partial \delta^a \partial \delta^b} &= -\frac{1}{2} \sum_{\ell=1}^n \frac{2}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^b} \left\{ d_\ell + \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) \right\} \\ &\quad + \frac{1}{2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{\partial^2 m_\ell}{\partial \delta^a \partial \delta^b} \frac{1}{m_\ell} \left\{ d_\ell + \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) \right\} \\ &\quad - \frac{1}{2} \sum_{\ell=1}^n \left( \frac{1}{\sigma^2 m_\ell} \right)^2 \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^b} \ddot{s} \left( \frac{1}{\sigma^2 m_\ell} \right), \end{aligned}$$

para  $a, b = 1, \dots, p$ .

Tomando as esperanças das duas expressões acima, encontram-se os cumulantes

$$\lambda_a = E \left( \frac{\partial L^*}{\partial \delta^a} \right) = 0$$

e

$$\lambda_{ab} = E \left( \frac{\partial^2 L^*}{\partial \delta^a \partial \delta^b} \right) = -\frac{1}{2} \sum_{\ell=1}^n \left( \frac{1}{\sigma^2 m_\ell} \right)^2 \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^b} \ddot{s} \left( \frac{1}{\sigma^2 m_\ell} \right),$$

pois

$$E(d_\ell) = -\dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right).$$

Para obter os cumulantes em relação aos componentes do parâmetro  $\beta$ , temos agora que

$$\frac{\partial L}{\partial \beta^i} = \sum_{\ell=1}^n (y_\ell - \mu_\ell) \frac{1}{\sigma^2 m_\ell} \frac{d\theta_\ell}{d\mu_\ell} \frac{d\mu_\ell}{d\eta_\ell} x_{\ell i}^*$$

e

$$\begin{aligned} \frac{\partial^2 L}{\partial \beta^i \partial \beta^j} &= - \sum_{\ell=1}^n x_{\ell i}^* \frac{1}{\sigma^2 m_\ell} w_\ell x_{\ell j}^* + \sum_{\ell=1}^n x_{\ell i}^* (y_\ell - \mu_\ell) g_\ell \frac{1}{\sigma^2 m_\ell} x_{\ell j}^* \\ &+ \sum_{\ell=1}^n x_{\ell i}^* (y_\ell - \mu_\ell) f_\ell \frac{1}{\sigma^2 m_\ell} x_{\ell j}^* + \sum_{\ell=1}^n (y_\ell - \mu_\ell) \frac{1}{\sigma^2 m_\ell} \frac{d\theta_\ell}{d\mu_\ell} \frac{d\mu_\ell}{d\eta_\ell} x_{\ell ij}^{**}, \end{aligned}$$

em que

$$w_\ell = \left( \frac{d\mu_\ell}{d\eta_\ell} \right)^2 \frac{d\theta_\ell}{d\mu_\ell}, \quad g_\ell = \frac{d^2 \theta_\ell}{d\mu_\ell^2} \left( \frac{d\mu_\ell}{d\eta_\ell} \right)^2, \quad f_\ell = \frac{d^2 \mu_\ell}{d\eta_\ell^2} \frac{d\theta_\ell}{d\mu_\ell}, \quad x_{\ell i}^* = \frac{\partial \eta_\ell}{\partial \beta^i} \text{ e } x_{\ell ij}^{**} = \frac{\partial^2 \eta_\ell}{\partial \beta^i \partial \beta^j},$$

para  $i, j = 1, \dots, k$ . Tomando as esperanças das duas expressões acima, encontram-se os cumulantes

$$\lambda_j = E \left( \frac{\partial L^*}{\partial \beta^j} \right) = 0 \quad \text{e} \quad \lambda_{jk} = E \left( \frac{\partial^2 L}{\partial \beta^j \partial \beta^k} \right) = - \sum_{\ell=1}^n x_{\ell j}^* \frac{1}{\sigma^2 m_\ell} w_\ell x_{\ell k}^*,$$

pois

$$E(y_\ell) = \mu_\ell.$$

Para obter os cumulantes em relação ao parâmetro  $\sigma^2$ , temos

$$\frac{\partial L}{\partial \sigma^2} = \frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left\{ d_\ell + \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) \right\}$$

e

$$\frac{\partial^2 L}{\partial (\sigma^2)^2} = \frac{1}{(\sigma^2)^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left\{ d_\ell + \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) \right\} - \frac{1}{2(\sigma^2)^2} \sum_{\ell=1}^n \left( \frac{1}{\sigma^2 m_\ell} \right)^2 \ddot{s} \left( \frac{1}{\sigma^2 m_\ell} \right).$$

Tomando as esperanças das duas expressões acima, encontram-se os cumulantes

$$\lambda_{\sigma^2} = E\left(\frac{\partial L}{\partial \sigma^2}\right) = 0$$

e

$$\lambda_{\sigma^2 \sigma^2} = E\left(\frac{\partial^2 L}{\partial (\sigma^2)^2}\right) = -\frac{1}{2(\sigma^2)^2} \sum_{\ell=1}^n \left(\frac{1}{\sigma^2 m_\ell}\right)^2 \ddot{s}\left(\frac{1}{\sigma^2 m_\ell}\right),$$

pois

$$E(d_\ell) = -\dot{s}\left(\frac{1}{\sigma^2 m_\ell}\right).$$

Para obter os cumulantes mistos, temos agora que

$$\begin{aligned} \frac{\partial^2 L}{\partial \delta^a \partial \sigma^2} &= -\frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \left\{ d_\ell + \dot{s}\left(\frac{1}{\sigma^2 m_\ell}\right) \right\} \\ &\quad - \frac{1}{2\sigma^2} \sum_{\ell=1}^n \left(\frac{1}{\sigma^2 m_\ell}\right)^2 \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \ddot{s}\left(\frac{1}{\sigma^2 m_\ell}\right), \\ \frac{\partial^2 L}{\partial \delta^a \partial \beta^i} &= -\frac{1}{2} \sum_{\ell=1}^n \frac{2}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \frac{d\theta_\ell}{d\mu_\ell} \frac{d\mu_\ell}{d\eta_\ell} (y_\ell - \mu_\ell) x_{\ell i}^* \end{aligned}$$

e

$$\frac{\partial^2 L}{\partial \sigma^2 \partial \beta^i} = -\frac{1}{2} \sum_{\ell=1}^n \frac{2}{\sigma^2 m_\ell} \frac{d\theta_\ell}{d\mu_\ell} \frac{d\mu_\ell}{d\eta_\ell} (y_\ell - \mu_\ell) x_{\ell i}^*,$$

para  $a = 1, \dots, p$  e  $i = 1, \dots, k$ .

Tomando as esperanças das três expressões acima, encontram-se os cumulantes

$$\begin{aligned} \lambda_{a\sigma^2} &= E\left(\frac{\partial^2 L}{\partial \delta^a \partial \sigma^2}\right) = -\frac{1}{2\sigma^2} \sum_{\ell=1}^n \left(\frac{1}{\sigma^2 m_\ell}\right)^2 \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \ddot{s}\left(\frac{1}{\sigma^2 m_\ell}\right), \\ \lambda_{aj} &= E\left(\frac{\partial^2 L}{\partial \delta^a \partial \beta^j}\right) = 0 \quad \text{e} \quad \lambda_{\sigma^2 j} = E\left(\frac{\partial^2 L}{\partial \sigma^2 \partial \beta^j}\right) = 0, \end{aligned}$$

respectivamente, pois

$$E(d_\ell) = -\dot{s}\left(\frac{1}{\sigma^2 m_\ell}\right) \quad \text{e} \quad E(y_\ell) = \mu_\ell.$$

Finalmente, obtemos a matriz de informação total de Fisher, que é dada por

$$I = -E\left(\frac{\partial^2 L}{\partial \omega \partial \omega^\top}\right) = \begin{pmatrix} \frac{1}{2} \Psi^\top \Phi \ddot{S} \Phi \Psi & 0 & \frac{1}{2\sigma^2} \Psi^\top \ddot{S} \Phi^2 \iota \\ 0 & X^{*\top} \Phi^{1/2} W \Phi^{1/2} X^* & 0 \\ \frac{1}{2\sigma^2} \iota^\top \Phi^2 \ddot{S} \Psi & 0 & \frac{1}{2\sigma^4} \iota^\top \Phi^2 \ddot{S} \iota \end{pmatrix}.$$

---

## APÊNDICE C

### Multiplicação de “array”

---

Neste apêndice, apresentamos a definição de um array tri-dimensional e uma propriedade de multiplicação de array que nos auxiliará no cálculo de  $j^*$ , o bloco da matriz de informação observada para os parâmetros de perturbação  $(\beta^\top, \gamma)^\top$  no modelo não-linear heteroscedástico reparametrizado avaliado em  $(\delta^\top, \widehat{\beta}_\delta^\top, \widehat{\gamma}_\delta)^\top$ , definido na Seção 3.4. Mais especificamente, iremos calcular a submatriz inferior direita de dimensão  $k \times k$ . A multiplicação de “array” tri-dimensional foi introduzida por Bates e Watts (1980). Desde então, muitos autores têm discutido e utilizado esta multiplicação; ver Bates e Watts (1988), Seber (1999, pp. 691-692) e Wei (1999, pp. 188-190).

Um “array” tri-dimensional de dimensão  $n \times p \times q$  é denotado por  $X = (X_{\ell ij})$ , onde os índices  $\ell, i$  e  $j$  indicam a face, a linha e a coluna, respectivamente. Um “array” pode ser visto da seguinte forma:  $A = (A_\ell)$  e cada  $A_\ell$  é uma matriz  $A_\ell = (A_{\ell ij})$  de dimensão  $p \times q$  para algum  $\ell$  fixo e  $A_\ell$  é chamada de  $\ell$ -ésima face de  $A$ .

**Definição C.1** *Se  $X$  é um “array” de dimensão  $n \times p \times q$ ,  $A$  e  $B$  são matrizes de dimensões  $r \times p$  e  $q \times s$ , respectivamente, então  $Y = AXB$  é definido como um “array” de dimensão  $n \times r \times s$  com elementos*

$$Y_{\ell kt} = \sum_{i=1}^p \sum_{j=1}^q A_{ki} X_{\ell ij} B_{jt}.$$

**Definição C.2** *Se  $X$  é um “array” de dimensão  $n \times p \times q$ ,  $A$  é uma matriz de dimensão  $m \times n$ , então  $Y = [A][X]$  é chamado de produto colchete de  $A$  e  $X$ , isto*

é, um “array” de dimensão  $m \times p \times q$  com elemento

$$Y_{lij} = \sum_{t=1}^n A_{\ell k} X_{tij}.$$

**Definição C.3** (*Decomposição de Cholesky*) Se  $A$  é uma matriz positiva definida, existe uma única matriz triangular superior,  $U$ , com elementos diagonais positivos tal que

$$A = U^T U.$$

A seguinte propriedade pode ser deduzida diretamente das definições acima.

**Propriedade C.1** Sejam  $A, L, M$  matrizes e  $X$  um “array”, então, temos que

$$[A][LXM] = L[A][X]M.$$

Utilizando a definição (C.2) e a propriedade (C.1), pode-se mostrar que

$$\begin{aligned} -\frac{\partial^2 L^*}{\partial \beta^i \partial \beta^j} &= \frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell i}^* q_{\ell} w_{\ell} x_{\ell j}^* - \frac{1}{\gamma} \sum_{\ell=1}^n (y_{\ell} - \mu_{\ell}) q_{\ell} x_{\ell ij}^{**} \\ &= \frac{1}{\gamma} \{X^{*T} W^{1/2} Q(\delta) W^{1/2} X^* - [(y - \mu)^T Q(\delta)][X^{**}]\} \\ &= \frac{1}{\gamma} R^T(\delta)(I_k - B(\delta))R(\delta), \end{aligned}$$

para  $i, j = 1, \dots, k$  em que  $R(\delta)$  é dada pela definição (C.3) sendo

$$B(\delta) = \{[(y - \mu)^T Q(\delta)][(R^{-1}(\delta))^T X^{**} R^{-1}(\delta)]\}_{(\delta; \hat{\beta}_{\delta}, \hat{\gamma}_{\delta})}$$

e

$$X^{**} = X^{**}(\beta) = \frac{\partial^2 \eta}{\partial \beta \partial \beta^T}.$$

---

## APÊNDICE D

# Derivadas no modelo não-linear heteroscedástico reparametrizado.

---

Neste apêndice, apresentamos o cálculo da primeira e segunda derivadas do logaritmo da função de verossimilhanças do vetor de parâmetros  $\omega = (\delta^\top, \beta^\top, \sigma^2)^\top$  para o modelo não-linear heteroscedástico da família exponencial (reparametrizado) definido na Seção 3.3, cuja forma é dada por

$$L^* = L^*(y; \delta, \beta, \gamma) = -\frac{1}{2} \sum_{\ell=1}^n \left\{ \frac{q_\ell}{\gamma} d_\ell + t(y_\ell) + \log\left(\frac{\gamma}{q_\ell}\right) \right\}$$

em que

$$q_\ell = q_\ell(\delta) = \left( \prod_{s=1}^n m_s \right)^{1/n} / m_\ell.$$

Vale salientar aqui que estamos considerando somente as distribuições normal e normal inversa. Por simples diferenciação em relação aos componentes do parâmetro  $\delta$ , temos que

$$\frac{\partial L^*}{\partial \delta^a} = -\frac{1}{2\gamma} \sum_{\ell=1}^n q_{\ell a} d_\ell,$$
$$\frac{\partial^2 L^*}{\partial \delta^a \partial \delta^b} = -\frac{1}{2\gamma} \sum_{\ell=1}^n q_{\ell ab} d_\ell,$$

em que

$$d_\ell = -2[y_\ell \theta_\ell - b(\theta_\ell) - c(y_\ell)], \quad q_{\ell a} = \frac{\partial q_\ell}{\partial \delta^a} \quad \text{e} \quad q_{\ell ab} = \frac{\partial^2 q_\ell}{\partial \delta^a \partial \delta^b},$$

para  $a, b = 1, \dots, p$ .

Por diferenciação em relação aos componentes do parâmetro  $\beta$ , temos que

$$\frac{\partial L^*}{\partial \beta^i} = \frac{1}{\gamma} \sum_{\ell=1}^n (y_\ell - \mu_\ell) q_\ell \frac{d\theta_\ell}{d\mu_\ell} \frac{d\mu_\ell}{d\eta_\ell} x_{\ell i}^*$$

e

$$\begin{aligned} \frac{\partial^2 L^*}{\partial \beta^i \partial \beta^j} &= -\frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell i}^* q_{\ell} w_{\ell} x_{\ell j}^* + \frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell i}^* q_{\ell} (y_{\ell} - \mu_{\ell}) g_{\ell} x_{\ell j}^* \\ &\quad + \frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell i}^* q_{\ell} (y_{\ell} - \mu_{\ell}) f_{\ell} x_{\ell j}^* + \frac{1}{\gamma} \sum_{\ell=1}^n (y_{\ell} - \mu_{\ell}) q_{\ell} \frac{d\theta_{\ell}}{d\mu_{\ell}} \frac{d\mu_{\ell}}{d\eta_{\ell}} x_{\ell ij}^{**}, \end{aligned}$$

em que

$$\begin{aligned} w_{\ell} &= \left( \frac{d\mu_{\ell}}{d\eta_{\ell}} \right)^2 \frac{d\theta_{\ell}}{d\mu_{\ell}}, \quad g_{\ell} = \frac{d^2\theta_{\ell}}{d\mu_{\ell}^2} \left( \frac{d\mu_{\ell}}{d\eta_{\ell}} \right)^2, \quad f_{\ell} = \frac{d^2\mu_{\ell}}{d\eta_{\ell}^2} \frac{d\theta_{\ell}}{d\mu_{\ell}}, \\ x_{\ell i}^* &= \frac{\partial \eta_{\ell}}{\partial \beta^i} \quad \text{e} \quad x_{\ell ij}^{**} = \frac{\partial^2 \eta_{\ell}}{\partial \beta^i \partial \beta^j}, \end{aligned}$$

para  $i, j = 1, \dots, k$ .

Por diferenciação em relação ao parâmetro  $\gamma$ , temos

$$\frac{\partial L^*}{\partial \gamma} = \frac{1}{2\gamma^2} \sum_{\ell=1}^n q_{\ell} d_{\ell} - \frac{n}{2\gamma}$$

e

$$\frac{\partial^2 L^*}{\partial \gamma^2} = -\frac{1}{\gamma^3} \sum_{\ell=1}^n q_{\ell} d_{\ell} + \frac{n}{2\gamma^2}.$$

Diferenciando  $\partial L^*/\partial \delta^a$  em relação aos parâmetros  $\gamma$  e  $\beta^i$ , temos

$$\frac{\partial^2 L}{\partial \delta^a \partial \gamma} = \frac{1}{2\gamma^2} \sum_{\ell=1}^n q_{\ell a} d_{\ell}$$

e

$$\frac{\partial^2 L^*}{\partial \delta^a \partial \beta^i} = \frac{1}{\gamma} \sum_{\ell=1}^n q_{\ell a} (y_{\ell} - \mu_{\ell}) \frac{d\theta_{\ell}}{d\mu_{\ell}} \frac{d\mu_{\ell}}{d\eta_{\ell}} x_{\ell i}^*,$$

respectivamente.

Finalmente, diferenciando  $\partial L^*/\partial \gamma$  em relação a  $\beta^i$ , temos

$$\frac{\partial^2 L}{\partial \gamma \partial \beta^i} = -\frac{1}{\gamma^2} \sum_{\ell=1}^n q_{\ell} (y_{\ell} - \mu_{\ell}) \frac{d\theta_{\ell}}{d\mu_{\ell}} \frac{d\mu_{\ell}}{d\eta_{\ell}} x_{\ell i}^*.$$

Observe que para a ligação canônica  $\theta = \eta$  e  $d\theta_{\ell}/d\eta_{\ell} = 1$ , então, temos que

$$\frac{\partial L^*}{\partial \beta^i} = \frac{1}{\gamma} \sum_{\ell=1}^n (y_{\ell} - \mu_{\ell}) q_{\ell} x_{\ell i}^*,$$

$$\frac{\partial^2 L^*}{\partial \delta^a \partial \beta^i} = \frac{1}{\gamma} \sum_{\ell=1}^n q_{\ell a} (y_{\ell} - \mu_{\ell}) x_{\ell i}^*,$$

e

$$\frac{\partial^2 L}{\partial \gamma \partial \beta^i} = -\frac{1}{\gamma^2} \sum_{\ell=1}^n q_{\ell} (y_{\ell} - \mu_{\ell}) x_{\ell i}^*,$$

para  $a = 1, \dots, p$  e  $i = 1, \dots, k$ .

---

APÊNDICE E

**Obtenção das quantidades propostas por Skovgaard**

---

Neste apêndice, apresentamos os cálculos de esperanças que são necessárias para a obtenção das quantidades

$$\hat{q} = \text{cov}_{\omega_0} \{U(\omega_0), L(\omega_0) - L(\omega)\} |_{\omega_0=\hat{\omega}, \omega=\tilde{\omega}}$$

e

$$\hat{\Upsilon} = \text{cov}_{\omega_0} \{U(\omega_0), U(\omega)\} |_{\omega_0=\hat{\omega}, \omega=\tilde{\omega}}$$

onde  $L$  é o logaritmo da função de verossimilhanças do vetor de parâmetros  $\omega = (\delta^\top, \beta^\top, \sigma^2)^\top$  do modelo não-linear da família exponencial definido na Seção 4.3 (expressões (4.12) e (4.13)) e  $U$  é a função escore, cujas formas são dadas por

$$L(y; \delta, \beta, \sigma^2) = -\frac{1}{2} \sum_{i=1}^n \left\{ \frac{1}{\sigma^2 m_i} d_i + t(y_i) + s \left( \frac{1}{\sigma^2 m_i} \right) \right\},$$

onde  $d_i = d(y_i) = -2 [y_i \theta_i - b(\theta_i) - c(y_i)]$  e  $U = U(\delta, \beta, \sigma^2) = (U_\delta^\top, U_\beta^\top, U_{\sigma^2})^\top$  com

$$U_\delta = \frac{\partial L(\delta, \beta, \sigma^2)}{\partial \delta} = \frac{1}{2} \Psi^\top \Phi (d + \dot{S}), \quad U_\beta = \frac{\partial L(\delta, \beta, \sigma^2)}{\partial \beta} = X^{*\top} \Phi T V^{-1} (y - \mu)$$

e

$$U_{\sigma^2} = \frac{\partial L(\delta, \beta, \sigma^2)}{\partial \sigma^2} = \frac{1}{2\sigma^2} \iota^\top \Phi (d + \dot{S}),$$

respectivamente, sendo  $\Phi$ ,  $T$  e  $V$  matrizes diagonais de dimensão  $n \times n$  cujos  $(\ell, \ell)$ -ésimos elementos são dados, respectivamente por  $\phi_\ell = 1/(\sigma^2 m_\ell)$ ,  $T_\ell = d\mu_\ell/d\eta_\ell$  e  $V_\ell = d\mu_\ell/d\theta_\ell$ ,  $\Psi$  é uma matriz  $n \times p$  cuja  $\ell$ -ésima linha é dada por  $\partial \log m(z_\ell, \delta)/\partial \delta^\top$ ,  $\dot{S} = (\dot{s}(\phi_1), \dots, \dot{s}(\phi_n))^\top$  com  $\dot{s}(\phi_\ell) = ds(\phi_\ell)/d\phi_\ell$ ,  $m_\ell = m(z_\ell, \delta)$  sendo  $z_\ell^\top$  a  $\ell$ -ésima linha da matriz de covariadas  $Z$  de dimensão  $n \times p$ ,  $\ell = 1, \dots, n$ ,  $\delta$  é um vetor de

dimensão  $p \times 1$  de parâmetros desconhecidos,  $\iota$  é um vetor de 1's de dimensão  $n$ ,  $d = (d_1, \dots, d_n)^\top$  e  $\mu = (\mu_1, \dots, \mu_n)^\top$ .

Para a obtenção de  $\hat{q}$  é necessário obter  $E_\omega[c(y_\ell)]$ ,  $E_\omega[c(y_\ell) U_\delta(\omega)]$ ,  $E_\omega[c(y_\ell) U_{\sigma^2}(\omega)]$ ,  $E_\omega[c(y_\ell) U_\beta(\omega)]$ ,  $E_\omega[y_\ell U_\delta(\omega)]$ ,  $E_\omega[y_\ell U_{\sigma^2}(\omega)]$  e  $E_\omega[y_\ell U_\beta(\omega)]$ . Segue de condições de regularidade que

$$E_\omega[U_\delta(\omega)] = 0, \quad E_\omega[U_\beta(\omega)] = 0 \quad \text{e} \quad E_\omega[U_{\sigma^2}(\omega)] = 0. \quad (\text{E.1})$$

Sabemos que

$$E_\omega[y_\ell] = \mu_\ell = \dot{b}(\theta_\ell), \quad \text{Var}_\omega(y_\ell) = \phi_\ell^{-1} V_\ell, \quad (\text{E.2})$$

e

$$E_\omega[c(y_\ell)] = \theta_\ell \dot{b}(\theta_\ell) - b(\theta_\ell) - \frac{1}{2} \ddot{s} \left( \frac{1}{\sigma^2 m_\ell} \right). \quad (\text{E.3})$$

Por definição

$$E_\omega[c(y_\ell)] = \int c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell, \quad (\text{E.4})$$

sendo  $f(y_\ell; \theta_\ell, \phi_\ell)$  a função de densidade de  $y_\ell$ . Igualando (E.3) e (E.4), temos que

$$\theta_\ell \dot{b}(\theta_\ell) - b(\theta_\ell) - \frac{1}{2} \ddot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) = \int c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell. \quad (\text{E.5})$$

Derivando ambos os lados da expressão (E.5) em relação ao  $a$ -ésimo componente de  $\delta$ ,  $\delta^a$  digamos, chegamos a

$$\frac{1}{2} \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) \ddot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) = \frac{\partial}{\partial \delta^a} \left( \int c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \right),$$

para  $a = 1, \dots, p$ . Desenvolvendo o lado direito da expressão acima, obtemos

$$\begin{aligned} \frac{\partial}{\partial \delta^a} \left( \int c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \right) &= \int \frac{\partial}{\partial \delta^a} \{c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell \\ &= \int c(y_\ell) \frac{\partial}{\partial \delta^a} \{f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell \\ &= \int c(y_\ell) U_{\delta^a}(\omega) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \\ &= E_\omega[c(y_\ell) U_{\delta^a}(\omega)], \end{aligned}$$

para  $a = 1, \dots, p$ . Portanto, o vetor  $E_\omega[c(y_\ell) U_\delta(\omega)]$  de dimensão  $p \times 1$  tem o  $a$ -ésimo elemento dado por

$$\frac{1}{2} \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) \ddot{s} \left( \frac{1}{\sigma^2 m_\ell} \right),$$

para  $a = 1, \dots, p$ . Em notação matricial, temos

$$E_\omega[c(y_\ell) U_\delta(\omega)] = \frac{1}{2} \Psi \Phi \ddot{S}_L. \quad (\text{E.6})$$

Derivando ambos os lados da expressão (E.5) em relação a  $\sigma^2$ , chegamos a

$$\frac{1}{2\sigma} \frac{1}{\sigma^2 m_\ell} \ddot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) = \frac{\partial}{\partial \sigma^2} \left( \int c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \right).$$

Desenvolvendo o lado direito da expressão acima, obtemos

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} \left( \int c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \right) &= \int \frac{\partial}{\partial \sigma^2} \{c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell \\ &= \int c(y_\ell) \frac{\partial}{\partial \sigma^2} \{f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell \\ &= \int c(y_\ell) U_{\sigma^2}(\omega) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \\ &= E_\omega[c(y_\ell) U_{\sigma^2}(\omega)]. \end{aligned}$$

Portanto,

$$E_\omega[c(y_\ell) U_{\sigma^2}(\omega)] = \frac{1}{2\sigma^2} \frac{1}{\sigma^2 m_\ell} \ddot{s} \left( \frac{1}{\sigma^2 m_\ell} \right).$$

Em notação matricial a expressão acima fica dada por

$$E_\omega[c(y_\ell) U_{\sigma^2}(\omega)] = \frac{1}{2\sigma^2} \iota \Phi \ddot{S}_L. \quad (\text{E.7})$$

Finalmente, derivando ambos os lados da expressão (E.5) em relação ao  $i$ -ésimo componente de  $\beta$ ,  $\beta^i$  digamos, chegamos a

$$\theta_\ell V_\ell \frac{d\theta_\ell}{d\mu_\ell} \frac{d\mu_\ell}{d\eta_\ell} \frac{\partial \eta_\ell}{\partial \beta^i} = \frac{\partial}{\partial \beta^i} \left( \int c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \right),$$

para  $i = 1, \dots, k$ . Desenvolvendo o lado direito da expressão acima, obtemos

$$\begin{aligned} \frac{\partial}{\partial \beta^i} \left( \int c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \right) &= \int \frac{\partial}{\partial \beta^i} \{c(y_\ell) f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell \\ &= \int c(y_\ell) \frac{\partial}{\partial \beta^i} \{f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell \\ &= \int c(y_\ell) U_{\beta^i}(\omega) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \\ &= E_\omega[c(y_\ell) U_{\beta^i}(\omega)], \end{aligned}$$

para  $i = 1, \dots, k$ . Portanto, o vetor  $E_\omega[c(y_\ell) U_\beta(\omega)]$  de dimensão  $k \times 1$  tem o  $i$ -ésimo elemento dado por

$$\theta_\ell V_\ell \frac{d\theta_\ell}{d\mu_\ell} \frac{d\mu_\ell}{d\eta_\ell} x_{\ell i}^*,$$

para  $i = 1, \dots, k$ . Em notação matricial temos

$$E_\omega[c(y_\ell) U_\beta(\omega)] = X^{*\top} T\theta. \quad (\text{E.8})$$

Vimos que

$$E_\omega[y_\ell] = \mu_\ell = \dot{b}(\theta_\ell) \quad (\text{E.9})$$

e, por definição,

$$E_\omega[y_\ell] = \int y_\ell f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell. \quad (\text{E.10})$$

Igualando (E.9) e (E.10), temos que

$$\dot{b}(\theta_\ell) = \int y_\ell f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell. \quad (\text{E.11})$$

Derivando ambos os lados da expressão (E.11) em relação a  $\delta^a$ , a  $\sigma^2$  e a  $\beta^i$  chegamos, respectivamente, a

$$\begin{aligned} 0 &= \frac{\partial}{\partial \delta^a} \left( \int y_\ell f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \right) = \int \frac{\partial}{\partial \delta^a} \{y_\ell f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell \\ &= \int y_\ell \frac{\partial}{\partial \delta^a} \{f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell = \int y_\ell U_{\delta^a}(\omega) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell = E_\omega[y_\ell U_{\delta^a}(\omega)], \end{aligned}$$

$$\begin{aligned}
0 &= \frac{\partial}{\partial \sigma^2} \left( \int y_\ell f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \right) = \int \frac{\partial}{\partial \sigma^2} \{y_\ell f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell \\
&= \int y_\ell \frac{\partial}{\partial \sigma^2} \{f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell = \int y_\ell U_{\sigma^2}(\omega) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell = E_\omega[y_\ell U_{\sigma^2}(\omega)]
\end{aligned}$$

$$\begin{aligned}
e \quad V_\ell \frac{d\theta_\ell}{d\mu_\ell} \frac{d\mu_\ell}{d\eta_\ell} \frac{\partial \eta_\ell}{\partial \beta^i} &= \frac{\partial}{\partial \beta^i} \left( \int y_\ell f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell \right) \\
&= \int \frac{\partial}{\partial \beta^i} \{y_\ell f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell = \int y_\ell \frac{\partial}{\partial \beta^i} \{f(y_\ell; \theta_\ell, \phi_\ell)\} dy_\ell \\
&= \int y_\ell U_{\beta^i}(\omega) f(y_\ell; \theta_\ell, \phi_\ell) dy_\ell = E_\omega[y_\ell U_{\beta^i}(\omega)].
\end{aligned}$$

Portanto, temos que

$$E_\omega[y_\ell U_\delta(\omega)] = 0, \quad E_\omega[y_\ell U_{\sigma^2}(\omega)] = 0 \quad \text{e} \quad E_\omega[y_\ell U_\beta(\omega)] = X^{*\top} T_L. \quad (\text{E.12})$$

Para a obtenção da quantidade  $\hat{\Upsilon}$  apresentada no Capítulo 4, equação (4.4), precisamos obter  $E_\omega[y_\ell c(y_\ell)]$ ,  $E_{\omega_0}[U_\delta(\omega)]$ ,  $E_{\omega_0}[U_{\sigma^2}(\omega)]$ ,  $E_{\omega_0}[y_\ell U_\delta(\omega)]$  e  $E_{\omega_0}[y_\ell U_{\sigma^2}(\omega)]$ . Utilizamos também algumas esperanças já calculadas para a obtenção da quantidade  $\hat{q}$ , tais como:  $E_\omega[y_\ell U_\beta(\omega)]$ ,  $E_\omega[c(y_\ell) U_\delta(\omega)]$ ,  $E_\omega[c(y_\ell) U_{\sigma^2}(\omega)]$  e  $E_\omega[c(y_\ell) U_\beta(\omega)]$ . Primeiro, vamos obter a esperança  $E_\omega[y_\ell c(y_\ell)]$ , usando a relação  $E_\omega[y_\ell U_{\delta^a}(\omega)] = 0$ , para  $a = 1, \dots, p$ . Ou seja, substituiremos  $U_{\delta^a}(\omega)$  na esperança  $E_\omega[y_\ell U_{\delta^a}(\omega)]$  obtendo assim

$$\begin{aligned}
0 &= -\frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) \theta_\ell E_\omega[y_\ell^2] + \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) b(\theta_\ell) E_\omega[y_\ell] \\
&\quad + \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) E_\omega[y_\ell c(y_\ell)] + \frac{1}{2} \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) E_\omega[y_\ell].
\end{aligned}$$

Então,

$$\begin{aligned}
E_\omega[y_\ell c(y_\ell)] &= \theta_\ell E_\omega[y_\ell^2] - b(\theta_\ell) E_\omega[y_\ell] - \frac{1}{2} \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) E_{\omega_0}[y_\ell] \\
&= \theta_\ell \phi_\ell^{-1} V_\ell + \theta_\ell \mu_\ell^2 - b(\theta_\ell) \mu_\ell - \frac{1}{2} \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) \mu_\ell. \quad (\text{E.13})
\end{aligned}$$

O vetor  $E_{\omega_0}[U_\delta(\omega)]$  de dimensão  $p \times 1$  tem o  $a$ -ésimo elemento dado por

$$\begin{aligned}
&- \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) \theta_\ell E_{\omega_0}[y_\ell] + \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) b(\theta_\ell) \\
&+ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) E_{\omega_0}[c(y_\ell)] + \frac{1}{2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right),
\end{aligned}$$

para  $a = 1, \dots, p$ . Substituindo  $E_{\omega_0}[c(y_\ell)]$  e  $E_{\omega_0}[y_\ell]$  (dadas em (E.3) e (E.9)) na expressão acima chegamos a

$$\begin{aligned} & \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) (\theta_{0\ell} - \theta_\ell) \dot{b}(\theta_{0\ell}) - \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) [b(\theta_{0\ell}) - b(\theta_\ell)] \\ & - \frac{1}{2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) \left[ \dot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) - \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) \right], \end{aligned} \quad (\text{E.14})$$

para  $a = 1, \dots, p$ . Em notação matricial, temos

$$E_{\omega_0}[U_\delta(\omega)] = \Psi^\top \Phi(\Theta_0 - \Theta)\mu_0 - \Psi^\top \Phi[b(\theta_0) - b(\theta)] - \frac{1}{2} \Psi^\top \Phi(\dot{S}_0 - \dot{S}),$$

com  $\Theta = \text{diag}(\theta_1, \dots, \theta_n)$  e o subscrito 0 indica que as quantidades são avaliadas em  $\omega_0$ . A esperança  $E_{\omega_0}[U_{\sigma^2}(\omega)]$  pode ser expressa na forma

$$\begin{aligned} E_{\omega_0}[U_{\sigma^2}(\omega)] &= -\frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \theta_\ell E_{\omega_0}[y_\ell] + \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} b(\theta_\ell) \\ &+ \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} E_{\omega_0}[c(y_\ell)] + \frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) \\ &= -\frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} (\theta_\ell - \theta_{0\ell}) \dot{b}(\theta_{0\ell}) + \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} [b(\theta_\ell) - b(\theta_{0\ell})] \\ &+ \frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left[ \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) - \dot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) \right]. \end{aligned} \quad (\text{E.15})$$

Em notação matricial, temos que

$$E_{\omega_0}[U_{\sigma^2}(\omega)] = \frac{1}{\sigma^2} \iota^\top \Phi(\Theta_0 - \Theta)\mu_0 + \frac{1}{\sigma^2} \iota^\top \Phi[b(\theta) - b(\theta_0)] + \frac{1}{2\sigma^2} \iota^\top \Phi(\dot{S} - \dot{S}_0).$$

O vetor  $E_{\omega_0}[y_\ell U_\delta(\omega)]$  de dimensão  $p \times 1$  tem o  $a$ -ésimo elemento dado por

$$\begin{aligned} & - \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) \theta_\ell E_{\omega_0}[y_\ell^2] \\ & + \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) b(\theta_\ell) E_{\omega_0}[y_\ell] \\ & + \frac{1}{2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) E_{\omega_0}[y_\ell], \\ & + \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) E_{\omega_0}[y_\ell c(y_\ell)], \end{aligned}$$

para  $a = 1, \dots, p$ . Substituindo  $E_{\omega_0}[y_\ell^2]$ ,  $E_{\omega_0}[y_\ell]$  e  $E_{\omega_0}[y_\ell c(y_\ell)]$  (dadas em (E.2), (E.9) e (E.13)) na expressão acima, chegamos a

$$\begin{aligned} & \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) (\theta_{0\ell} - \theta_\ell) \left[ \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right)^{-1} V_{0\ell} + \{\dot{b}(\theta_{0\ell})\}^2 \right] \\ & - \frac{1}{2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) \left[ \dot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) - \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) \right] \dot{b}(\theta_{0\ell}) \\ & - \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left( \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \right) [b(\theta_{0\ell}) - b(\theta_\ell)] \dot{b}(\theta_{0\ell}), \end{aligned} \quad (\text{E.16})$$

para  $a = 1, \dots, p$ . Em notação matricial, temos

$$\begin{aligned} E_{\omega_0}[y_\ell U_\delta(\omega)] &= \Psi^\top \Phi \Phi_0^{-1} V_0 (\Theta_0 - \Theta)_\ell + \Psi^\top \Phi (\Theta_0 - \Theta) \Xi_\ell \\ &\quad - \frac{1}{2} \Psi^\top \Phi \Lambda \mu_0 - \Psi^\top \Phi \Sigma \mu_0, \end{aligned}$$

sendo  $\Xi = \text{diag}(\mu_{01}^2, \dots, \mu_{0n}^2)$ ,  $\Lambda = \text{diag}([\dot{s}(\phi_{01}) - \dot{s}(\phi_1)], \dots, [\dot{s}(\phi_{0n}) - \dot{s}(\phi_n)])$  e  $\Sigma = \text{diag}([b(\theta_{01}) - b(\theta_1)], \dots, [b(\theta_{0n}) - b(\theta_n)])$ . Por último, a esperança  $E_{\omega_0}[y_\ell U_{\sigma^2}(\omega)]$  é expressa na seguinte forma

$$\begin{aligned} E_{\omega_0}[y_\ell U_{\sigma^2}(\omega)] &= -\frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \theta_\ell E_{\omega_0}[y_\ell^2] + \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} b(\theta_\ell) E_{\omega_0}[y_\ell] \\ &\quad + \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} E_{\omega_0}[y_\ell c(y_\ell)] + \frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) E_{\omega_0}[y_\ell] \\ &= -\frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} (\theta_\ell - \theta_{0\ell}) \left[ \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right)^{-1} V_{0\ell} + \{\dot{b}(\theta_{0\ell})\}^2 \right] \\ &\quad - \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} [b(\theta_\ell) - b(\theta_{0\ell})] \dot{b}(\theta_{0\ell}) \\ &\quad - \frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left[ \dot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) - \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) \right] \dot{b}(\theta_{0\ell}). \end{aligned} \quad (\text{E.17})$$

Em notação matricial, temos que

$$\begin{aligned} E_{\omega_0}[y_\ell U_{\sigma^2}(\omega)] &= \frac{1}{\sigma^2} \iota^\top \Phi \Phi_0^{-1} V_0 (\Theta_0 - \Theta)_\ell + \frac{1}{\sigma^2} \iota^\top \Phi (\Theta_0 - \Theta) \Xi_\ell \\ &\quad + \frac{1}{\sigma^2} \iota^\top \Phi \Sigma \mu_0 - \frac{1}{2\sigma^2} \iota^\top \Phi \Lambda \mu_0. \end{aligned}$$

Para obter as estatísticas da razão de verossimilhanças ajustadas,  $LR_a^*$  e  $LR_a^{**}$ , propostas por Skovgaard (2001) e dadas em (4.6) e (4.7), é necessário calcular as quantidades

$$\begin{aligned} q &= E_{\omega_0}[U(\omega_0) (L(\omega_0) - L(\omega))] \\ &= \begin{pmatrix} E_{\omega_0}[U_\delta(\omega_0)L(\omega_0)] - E_{\omega_0}[U_\delta(\omega_0)L(\omega)] \\ E_{\omega_0}[U_\beta(\omega_0)L(\omega_0)] - E_{\omega_0}[U_\beta(\omega_0)L(\omega)] \\ E_{\omega_0}[U_{\sigma^2}(\omega_0)L(\omega_0)] - E_{\omega_0}[U_{\sigma^2}(\omega_0)L(\omega)] \end{pmatrix} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \end{aligned} \quad (\text{E.18})$$

e

$$\begin{aligned} \Upsilon &= E_{\omega_0}[U(\omega_0) U^\top(\omega)] \\ &= \begin{pmatrix} E_{\omega_0}[U_\delta(\omega_0)U_\delta^\top(\omega)] & E_{\omega_0}[U_\delta(\omega_0)U_\beta^\top(\omega)] & E_{\omega_0}[U_\delta(\omega_0)U_{\sigma^2}(\omega)] \\ E_{\omega_0}[U_\beta(\omega_0)U_\delta^\top(\omega)] & E_{\omega_0}[U_\beta(\omega_0)U_\beta^\top(\omega)] & E_{\omega_0}[U_\beta(\omega_0)U_{\sigma^2}(\omega)] \\ E_{\omega_0}[U_{\sigma^2}(\omega_0)U_\delta^\top(\omega)] & E_{\omega_0}[U_{\sigma^2}(\omega_0)U_\beta^\top(\omega)] & E_{\omega_0}[U_{\sigma^2}(\omega_0)U_{\sigma^2}(\omega)] \end{pmatrix}. \end{aligned} \quad (\text{E.19})$$

Iremos primeiro calcular  $E_{\omega_0}[U_\delta(\omega_0) L(\omega_0)]$  e  $E_{\omega_0}[U_\delta(\omega_0) L(\omega)]$ , dadas na expressão (E.18). Substituindo  $L(\omega_0)$  em  $E_{\omega_0}[U_\delta(\omega_0) L(\omega_0)]$  e  $L(\omega)$  em  $E_{\omega_0}[U_\delta(\omega_0) L(\omega)]$  chegamos, respectivamente, a

$$\begin{aligned} E_{\omega_0}[U_\delta(\omega_0) L(\omega_0)] &= E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} y_\ell \theta_{0\ell} U_\delta(\omega_0) \right] - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} b(\theta_{0\ell}) U_\delta(\omega_0) \right] \\ &\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} c(y_\ell) U_\delta(\omega_0) \right] - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n t(y_\ell) U_\delta(\omega_0) \right] \\ &\quad - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) U_\delta(\omega_0) \right] \end{aligned}$$

e

$$\begin{aligned} E_{\omega_0}[U_\delta(\omega_0) L(\omega)] &= E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} y_\ell \theta_\ell U_\delta(\omega_0) \right] - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} b(\theta_\ell) U_\delta(\omega_0) \right] \\ &\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} c(y_\ell) U_\delta(\omega_0) \right] - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n t(y_\ell) U_\delta(\omega_0) \right] \\ &\quad - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma^2 m_\ell} \right) U_\delta(\omega_0) \right]. \end{aligned}$$

Segundo, vamos obter  $E_{\omega_0}[U_\beta(\omega_0) L(\omega_0)]$  e  $E_{\omega_0}[U_\beta(\omega_0) L(\omega)]$  da mesma forma

como foram obtidas as esperanças dadas acima. Temos

$$\begin{aligned} E_{\omega_0}[U_{\beta}(\omega_0) L(\omega_0)] &= E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} y_{\ell} \theta_{0\ell} U_{\beta}(\omega_0) \right] - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} b(\theta_{0\ell}) U_{\beta}(\omega_0) \right] \\ &\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} c(y_{\ell}) U_{\beta}(\omega_0) \right] - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n t(y_{\ell}) U_{\beta}(\omega_0) \right] \\ &\quad - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) U_{\beta}(\omega_0) \right] \end{aligned}$$

e

$$\begin{aligned} E_{\omega_0}[U_{\beta}(\omega_0) L(\omega)] &= E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_{\ell}} y_{\ell} \theta_{\ell} U_{\beta}(\omega_0) \right] - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_{\ell}} b(\theta_{\ell}) U_{\beta}(\omega_0) \right] \\ &\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_{\ell}} c(y_{\ell}) U_{\beta}(\omega_0) \right] - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n t(y_{\ell}) U_{\beta}(\omega_0) \right] \\ &\quad - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma^2 m_{\ell}} \right) U_{\beta}(\omega_0) \right]. \end{aligned}$$

Por último, vamos obter  $E_{\omega_0}[U_{\sigma^2}(\omega_0) L(\omega_0)]$  e  $E_{\omega_0}[U_{\sigma^2}(\omega_0) L(\omega)]$  da mesma maneira chegando a

$$\begin{aligned} E_{\omega_0}[U_{\sigma^2}(\omega_0) L(\omega_0)] &= E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} y_{\ell} \theta_{0\ell} U_{\sigma^2}(\omega_0) \right] - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} b(\theta_{0\ell}) U_{\sigma^2}(\omega_0) \right] \\ &\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} c(y_{\ell}) U_{\sigma^2}(\omega_0) \right] - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n t(y_{\ell}) U_{\sigma^2}(\omega_0) \right] \\ &\quad - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) U_{\sigma^2}(\omega_0) \right] \end{aligned}$$

e

$$\begin{aligned} E_{\omega_0}[U_{\sigma^2}(\omega_0) L(\omega)] &= E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_{\ell}} y_{\ell} \theta_{\ell} U_{\sigma^2}(\omega_0) \right] - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_{\ell}} b(\theta_{\ell}) U_{\sigma^2}(\omega_0) \right] \\ &\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_{\ell}} c(y_{\ell}) U_{\sigma^2}(\omega_0) \right] - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n t(y_{\ell}) U_{\sigma^2}(\omega_0) \right] \\ &\quad - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma^2 m_{\ell}} \right) U_{\sigma^2}(\omega_0) \right]. \end{aligned}$$

Note que

$$\begin{aligned}
q_1 &= E_{\omega_0}[U_\delta(\omega_0) L(\omega_0)] - E_{\omega_0}[U_\delta(\omega_0) L(\omega)] \\
&= E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} y_\ell \theta_{0\ell} U_\delta(\omega_0) \right] - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} b(\theta_{0\ell}) U_\delta(\omega_0) \right] \\
&\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} c(y_\ell) U_\delta(\omega_0) \right] - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) U_\delta(\omega_0) \right] \\
&\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} y_\ell \theta_\ell U_\delta(\omega_0) \right] + E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} b(\theta_\ell) U_\delta(\omega_0) \right] \\
&\quad + E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} c(y_\ell) U_\delta(\omega_0) \right] + \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma^2 m_\ell} \right) U_\delta(\omega_0) \right],
\end{aligned}$$

$$\begin{aligned}
q_2 &= E_{\omega_0}[U_\beta(\omega_0) L(\omega_0)] - E_{\omega_0}[U_\beta(\omega_0) L(\omega)] \\
&= E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} y_\ell \theta_{0\ell} U_\beta(\omega_0) \right] - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} b(\theta_{0\ell}) U_\beta(\omega_0) \right] \\
&\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} c(y_\ell) U_\beta(\omega_0) \right] - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) U_\beta(\omega_0) \right] \\
&\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} y_\ell \theta_\ell U_\beta(\omega_0) \right] + E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} b(\theta_\ell) U_\beta(\omega_0) \right] \\
&\quad + E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} c(y_\ell) U_\beta(\omega_0) \right] + \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma^2 m_\ell} \right) U_\beta(\omega_0) \right]
\end{aligned}$$

e

$$\begin{aligned}
q_3 &= E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} y_\ell \theta_{0\ell} U_{\sigma^2}(\omega_0) \right] - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} b(\theta_{0\ell}) U_{\sigma^2}(\omega_0) \right] \\
&\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} c(y_\ell) U_{\sigma^2}(\omega_0) \right] - \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) U_{\sigma^2}(\omega_0) \right] \\
&\quad - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} y_\ell \theta_\ell U_{\sigma^2}(\omega_0) \right] + E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} b(\theta_\ell) U_{\sigma^2}(\omega_0) \right] \\
&\quad + E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} c(y_\ell) U_{\sigma^2}(\omega_0) \right] + \frac{1}{2} E_{\omega_0} \left[ \sum_{\ell=1}^n s \left( \frac{1}{\sigma^2 m_\ell} \right) U_{\sigma^2}(\omega_0) \right].
\end{aligned}$$

Substituindo  $E_{\omega_0}[U_\delta(\omega_0)]$  e  $E_{\omega_0}[y_\ell U_\delta(\omega_0)]$  (dadas em (E.1) e (E.12)) em  $q_1$ ,  $E_{\omega_0}[U_\beta(\omega_0)]$  (dada em (E.1)) em  $q_2$  e  $E_{\omega_0}[U_{\sigma^2}(\omega_0)]$  e  $E_{\omega_0}[y_\ell U_{\sigma^2}(\omega_0)]$  (dadas em (E.1) e (E.12)) em  $q_3$ , temos que

$$q_1 = -E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} c(y_\ell) U_\delta(\omega_0) \right] + E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} c(y_\ell) U_\delta(\omega_0) \right],$$

$$q_2 = -E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} c(y_\ell) U_\beta(\omega_0) \right] + E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} c(y_\ell) U_\beta(\omega_0) \right] \\ - E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \theta_\ell y_\ell U_\beta(\omega_0) \right] + E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} \theta_{0\ell} y_\ell U_\beta(\omega_0) \right]$$

e

$$q_3 = -E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} c(y_\ell) U_{\sigma^2}(\omega_0) \right] + E_{\omega_0} \left[ \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} c(y_\ell) U_{\sigma^2}(\omega_0) \right].$$

Substituindo  $E_{\omega_0}[c(y_\ell) U_\delta(\omega_0)]$  (dada na equação (E.6)) em  $q_1$ ,  $E_{\omega_0}[c(y_\ell) U_\beta(\omega_0)]$  (dada na equação (E.8)) e  $E_{\omega_0}[y_\ell U_\beta(\omega_0)]$  (dada na equação (E.12)) em  $q_2$  e  $E_{\omega_0}[c(y_\ell) U_{\sigma^2}(\omega_0)]$  (dada na equação (E.7)) e  $E_{\omega_0}[y_\ell U_{\sigma^2}(\omega_0)]$  (dada na equação (E.12)) em  $q_3$ , temos que o  $a$ -ésimo elemento do vetor  $q_1$  de dimensão  $p \times 1$  é dado por

$$\sum_{\ell=1}^n \left\{ \frac{1}{2} \frac{1}{m_{0\ell}} \frac{\partial m_{0\ell}}{\partial \delta_0^a} \frac{1}{\sigma^2 m_\ell} \ddot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) \frac{1}{\sigma_0^2 m_{0\ell}} \right\} \\ - \sum_{\ell=1}^n \left\{ \frac{1}{2} \frac{1}{m_{0\ell}} \frac{\partial m_{0\ell}}{\partial \delta_0^a} \frac{1}{\sigma_0^2 m_{0\ell}} \ddot{s} \left( \frac{1}{\sigma_0^2 m_0} \right) \frac{1}{\sigma_0^2 m_{0\ell}} \right\},$$

para  $a = 1, \dots, p$ , o  $i$ -ésimo elemento do vetor  $q_2$  de dimensão  $k \times 1$  é dado por

$$- \sum_{\ell=1}^n \left\{ \frac{1}{\sigma^2 m_\ell} \theta_\ell V_{0\ell} \frac{d\theta_{0\ell}}{d\mu_{0\ell}} \frac{d\mu_{0\ell}}{d\eta_{0\ell}} \frac{\partial \eta_{0\ell}}{\beta_0^i} \right\} + \sum_{\ell=1}^n \left\{ \frac{1}{\sigma^2 m_\ell} \theta_{0\ell} V_{0\ell} \frac{d\theta_{0\ell}}{d\mu_{0\ell}} \frac{d\mu_{0\ell}}{d\eta_{0\ell}} \frac{\partial \eta_{0\ell}}{\beta_0^i} \right\},$$

para  $i = 1, \dots, k$  e

$$q_3 = \sum_{\ell=1}^n \left\{ \frac{1}{2\sigma_0^2} \frac{1}{\sigma^2 m_\ell} \ddot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) \frac{1}{\sigma_0^2 m_{0\ell}} \right\} - \sum_{\ell=1}^n \left\{ \frac{1}{2\sigma_0^2} \frac{1}{\sigma_0^2 m_{0\ell}} \ddot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) \frac{1}{\sigma_0^2 m_{0\ell}} \right\}.$$

Então, a quantidade em (E.18) é dada em notação matricial por

$$q = \begin{pmatrix} \frac{1}{2} \Psi_0^\top \ddot{S}_0 \Phi_0 (\Phi - \Phi_0) \iota \\ X_0^{*\top} T_0 \Phi (\Theta_0 - \Theta) \iota \\ \frac{1}{2\sigma_0^2} \iota^\top (\Phi - \Phi_0) \Phi_0 \ddot{S}_0 \iota \end{pmatrix}.$$

Conseqüentemente,

$$\hat{q} = \begin{pmatrix} \frac{1}{2} \hat{\Psi}^\top \hat{S} \hat{\Phi} (\hat{\Phi} - \hat{\Phi}) \iota \\ \hat{X}^{*\top} \hat{T} \hat{\Phi} (\hat{\Theta} - \hat{\Theta}) \iota \\ \frac{1}{2\hat{\sigma}^2} \iota^\top (\hat{\Phi} - \hat{\Phi}) \hat{\Phi} \hat{S} \iota \end{pmatrix}.$$

Agora, obteremos a quantidade  $\Upsilon$  dada em (E.19). Observamos que as esperanças  $E_{\omega_0}[U_\delta(\omega_0) U_\beta^\top(\omega)]$  e  $E_{\omega_0}[U_{\sigma^2}(\omega_0) U_\beta^\top(\omega)]$  se anulam, pois  $E_{\omega_0}[U_\delta(\omega_0)] = 0$  e  $E_{\omega_0}[y_\ell U_{\sigma^2}(\omega_0)] = 0$  (dada em (E.12)).

Temos que a  $a$ -ésima coluna da matriz  $E_{\omega_0}[U_\delta(\omega_0) U_\delta^\top(\omega)]$  de dimensão  $p \times p$  é

$$\begin{aligned} & - \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \theta_\ell E_{\omega_0}[y_\ell U_\delta(\omega_0)] \\ & + \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} b(\theta_\ell) E_{\omega_0}[U_\delta(\omega_0)] \\ & + \frac{1}{2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} E_{\omega_0}[c(y_\ell) U_\delta(\omega_0)], \\ & + \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) E_{\omega_0}[U_\delta(\omega_0)] \end{aligned}$$

para  $a = 1, \dots, p$ . Substituindo  $E_{\omega_0}[U_\delta(\omega_0)]$ ,  $E_{\omega_0}[c(y_\ell) U_\delta(\omega_0)]$  e  $E_{\omega_0}[y_\ell U_\delta(\omega_0)]$  (dadas em (E.1), (E.8) e (E.12)) na expressão acima, pode-se mostrar que o  $(a, b)$ -ésimo elemento da matriz  $E_{\omega_0}[U_\delta(\omega_0) U_\delta^\top(\omega)]$  é dado por

$$\frac{1}{2} \sum_{\ell=1}^n \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \frac{1}{\sigma^2 m_\ell} \dot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) \frac{1}{\sigma_0^2 m_{0\ell}} \frac{1}{m_{0\ell}} \frac{\partial m_{0\ell}}{\partial \delta_0^b},$$

para  $a, b = 1, \dots, p$ . Em notação matricial, temos que

$$E_{\omega_0}[U_\delta(\omega_0) U_\delta^\top(\omega)] = \frac{1}{2} \Psi_0^\top \Phi_0 \ddot{S}_0 \Phi \Psi.$$

Analogamente,  $E_{\omega_0}[U_\delta(\omega_0) U_{\sigma^2}(\omega)]$  é dada por

$$\begin{aligned} & - \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \theta_\ell E_{\omega_0}[y_\ell U_\delta(\omega_0)] + \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} b(\theta_\ell) E_{\omega_0}[U_\delta(\omega_0)] \\ & + \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} E_{\omega_0}[c(y_\ell) U_\delta(\omega_0)] + \frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \dot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) E_{\omega_0}[U_\delta(\omega_0)]. \end{aligned}$$

Substituindo as devidas esperanças na expressão acima, temos que o  $a$ -ésimo elemento do vetor  $E_{\omega_0}[U_\delta(\omega_0) U_{\sigma^2}(\omega)]$  de dimensão  $p \times 1$  é dado por

$$\frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{m_{0\ell}} \frac{\partial m_{0\ell}}{\partial \delta_0^a} \frac{1}{\sigma^2 m_\ell} \dot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) \frac{1}{\sigma_0^2 m_{0\ell}},$$

para  $a = 1, \dots, p$ . Em notação matricial, temos

$$E_{\omega_0}[U_\delta(\omega_0) U_{\sigma^2}^\top(\omega)] = \frac{1}{2\sigma^2} \Psi_0^\top \Phi_0 \ddot{S}_0 \Phi \iota.$$

Obtemos ainda a  $a$ -ésima coluna da matriz  $E_{\omega_0}[U_\beta(\omega_0) U_\delta^\top(\omega)]$  de dimensão  $k \times p$  dada por

$$\begin{aligned} & - \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \theta_\ell E_{\omega_0}[y_\ell U_\beta(\omega_0)] \\ & + \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} b(\theta_\ell) E_{\omega_0}[U_\beta(\omega_0)], \\ & + \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} E_{\omega_0}[c(y_\ell) U_\beta(\omega_0)] \\ & + \frac{1}{2} \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) E_{\omega_0}[U_\beta(\omega_0)], \end{aligned}$$

para  $a = 1, \dots, p$  onde  $E_{\omega_0}[U_\beta(\omega_0)]$ ,  $E_{\omega_0}[y_\ell U_\beta(\omega)]$  e  $E_{\omega_0}[c(y_\ell) U_\beta(\omega_0)]$  são dadas em (E.1), (E.8) e (E.12). Substituindo as expressões dessas esperanças, chegamos ao  $(i, a)$ -ésimo elemento da matriz  $E_{\omega_0}[U_\beta(\omega_0) U_\delta^\top(\omega)]$  dado por

$$\sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} (\theta_{0\ell} - \theta_\ell) \frac{d\mu_{0\ell}}{d\eta_{0\ell}} \frac{\partial \theta_{0\ell}}{\partial \beta_0^j},$$

para  $a = 1, \dots, p$  e  $i = 1, \dots, k$ . Em notação matricial, temos que

$$E_{\omega_0}[U_\delta(\omega_0) U_{\sigma^2}^\top(\omega)] = X_0^{*\top} T_0(\Theta_0 - \Theta)\Phi\Psi.$$

Da mesma forma, mostramos que a matriz  $E_{\omega_0}[U_\beta(\omega_0) U_\beta^\top(\omega)]$  de dimensão  $k \times k$  é dada por

$$E_{\omega_0}[U_\beta(\omega_0) U_\beta^\top(\omega)] = X_0^{*\top} T_0\Phi V^{-1} T X^*.$$

Podemos escrever  $E_{\omega_0}[U_\beta(\omega_0) U_{\sigma^2}(\omega)]$  da seguinte maneira

$$\begin{aligned} & - \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \theta_\ell E_{\omega_0}[y_\ell U_\beta(\omega_0)] + \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} b(\theta_\ell) E_{\omega_0}[U_\beta(\omega_0)] \\ & + \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} E_{\omega_0}[c(y_\ell) U_\beta(\omega_0)] + \frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) E_{\omega_0}[U_\beta(\omega_0)]. \end{aligned}$$

Substituindo as devidas esperanças na expressão acima, pode-se mostrar que o  $i$ -ésimo elemento da matriz  $E_{\omega_0}[U_\beta(\omega_0) U_{\sigma^2}(\omega)]$  de dimensão  $k \times 1$  é dado por

$$\frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} (\theta_{0\ell} - \theta_\ell) \frac{d\mu_{0\ell}}{d\eta_{0\ell}} \frac{\partial \theta_{0\ell}}{\partial \beta_0^i},$$

para  $i = 1, \dots, k$ . Em notação matricial, temos que

$$E_{\omega_0}[U_\beta(\omega_0) U_{\sigma^2}^\top(\omega)] = X_0^{*\top} T_0\Phi(\Theta_0 - \Theta) \iota.$$

Obtemos o  $a$ -ésimo elemento do vetor  $E_{\omega_0}[U_{\sigma^2}(\omega_0) U_\delta^\top(\omega)]$  de dimensão  $1 \times p$  da seguinte forma

$$\begin{aligned} & - \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \theta_\ell E_{\omega_0}[y_\ell U_{\sigma^2}(\omega_0)] \\ & + \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} b(\theta_\ell) E_{\omega_0}[U_{\sigma^2}(\omega_0)] \\ & + \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} E_{\omega_0}[c(y_\ell) U_{\sigma^2}(\omega_0)] \\ & + \frac{1}{2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) E_{\omega_0}[U_{\sigma^2}(\omega_0)], \end{aligned}$$

para  $a = 1, \dots, p$ . Substituindo  $E_{\omega_0}[U_{\sigma^2}(\omega_0)]$ ,  $E_{\omega_0}[c(y_\ell) U_{\sigma^2}(\omega_0)]$  e  $E_{\omega_0}[y_\ell U_{\sigma^2}(\omega_0)]$  (dadas em (E.1), (E.7) e (E.12)) na expressão acima, chegamos a

$$\frac{1}{2\sigma_0^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \ddot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) \frac{1}{\sigma_0^2 m_{0\ell}} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a},$$

para  $a = 1, \dots, p$ . Em notação matricial, temos

$$E_{\omega_0}[U_{\sigma^2}(\omega_0) U_\delta^\top(\omega)] = \frac{1}{2\sigma_0^2} \iota^\top \Phi_0 \ddot{S}_0 \Phi \Psi.$$

Finalmente, a esperança  $E_{\omega_0}[U_{\sigma^2}(\omega_0) U_{\sigma^2}(\omega)]$  é dada por

$$\begin{aligned} & - \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \theta_\ell E_{\omega_0}[y_\ell U_{\sigma^2}(\omega_0)] + \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} b(\theta_\ell) E_{\omega_0}[U_{\sigma^2}(\omega_0)] \\ & + \frac{1}{\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} E_{\omega_0}[c(y_\ell) U_{\sigma^2}(\omega_0)] + \frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \dot{s} \left( \frac{1}{\sigma^2 m_\ell} \right) E_{\omega_0}[U_{\sigma^2}(\omega_0)]. \end{aligned}$$

Substituindo  $E_{\omega_0}[y_\ell U_{\sigma^2}(\omega_0)]$ ,  $E_{\omega_0}[U_{\sigma^2}(\omega_0)]$  e  $E_{\omega_0}[c(y_\ell) U_{\sigma^2}(\omega_0)]$  na expressão acima pode-se mostrar que esta se reduz a

$$\frac{1}{2\sigma^2 \sigma_0^2} \sum_{\ell=1}^n \frac{1}{\sigma_0^2 m_{0\ell}} \ddot{s} \left( \frac{1}{\sigma_0^2 m_{0\ell}} \right) \frac{1}{\sigma^2 m_\ell}.$$

Em notação matricial, temos que

$$E_{\omega_0}[U_{\sigma^2}(\omega_0) U_{\sigma^2}(\omega)] = \frac{1}{2\sigma_0^2 \sigma^2} \iota^\top \Phi_0 \ddot{S}_0 \Phi \iota.$$

Portanto,

$$\Upsilon = \begin{pmatrix} \frac{1}{2} \Psi_0^\top \Phi_0 \ddot{S}_0 \Phi \Psi & 0 & \frac{1}{2\sigma^2} \Psi_0^\top \Phi_0 \ddot{S}_0 \Phi \iota \\ X_0^{*\top} T_0 (\Theta_0 - \Theta) \Phi \Psi & X_0^{*\top} T_0 \Phi V^{-1} T X^* & X_0^{*\top} T_0 \Phi (\Theta_0 - \Theta) \iota \\ \frac{1}{2\sigma_0^2} \iota^\top \Phi_0 \ddot{S}_0 \Phi \Psi & 0 & \frac{1}{2\sigma_0^2 \sigma^2} \iota^\top \Phi_0 \ddot{S}_0 \Phi \iota \end{pmatrix}$$

e, conseqüentemente,

$$\hat{\Upsilon} = \begin{pmatrix} \frac{1}{2} \hat{\Psi}^\top \hat{\Phi}_0 \hat{\ddot{S}} \hat{\Phi} \hat{\Psi} & 0 & \frac{1}{2\hat{\sigma}^2} \hat{\Psi}^\top \hat{\Phi} \hat{\ddot{S}} \hat{\Phi} \iota \\ \hat{X}^{*\top} \hat{T} (\hat{\Theta} - \tilde{\Theta}) \hat{\Phi} \hat{\Psi} & \hat{X}^{*\top} \hat{T} \hat{\Phi} \hat{V}^{-1} \hat{T} \hat{X}^* & \hat{X}^{*\top} \hat{T} \hat{\Phi} (\hat{\Theta} - \tilde{\Theta}) \iota \\ \frac{1}{2\hat{\sigma}_0^2} \iota^\top \hat{\Phi}_0 \hat{\ddot{S}} \hat{\Phi} \hat{\Psi} & 0 & \frac{1}{2\hat{\sigma}_0^2 \hat{\sigma}^2} \iota^\top \hat{\Phi}_0 \hat{\ddot{S}} \hat{\Phi} \iota \end{pmatrix}.$$

---

APÊNDICE F

Serviços de manutenção

---

Tabela F.1 *Tempo gasto no serviço ( $y$ ), em minutos, número de bebidas estocadas ( $x_1$ ) e distância percorrida ( $x_2$ ), em pés, numa amostra de 25 observações.*

$y$	$x_1$	$x_2$
16,68	7	560
11,50	3	220
12,03	3	340
14,88	4	80
13,75	6	150
18,11	7	330
8,00	2	110
17,83	7	210
79,24	30	1460
21,50	5	605
40,33	16	688
21,00	10	215
13,50	4	255
19,75	6	462
24,00	9	448
29,00	10	776
15,35	6	200
19,00	7	132
9,50	3	36
35,10	17	770
17,90	10	140
52,32	26	810
18,75	9	450
19,83	8	635
10,75	4	150

Fonte: Montgomery, Peck e Vining (2001, p. 76)

APÊNDICE G

Taxas mensais de retorno de ações

Tabela G.1 *Taxas mensais de retorno de ações de mercado (x) e da Corporação Acme Cleveland (y), no período de janeiro de 1986 a dezembro de 1990.*

Mês	x	y	Mês	x	y
Janeiro 1986	-0,061134	0,030160	Julho 1988	-0,061718	-0,110515
Fevereiro 1986	0,008220	-0,165457	Agosto 1988	-0,101710	-0,168769
Março 1986	-0,007381	0,080137	Setembro 1988	-0,032705	-0,135585
Abril 1986	-0,067561	-0,109917	Outubro 1988	-0,045334	-0,084077
Mai 1986	-0,006238	-0,114853	Novembro 1988	-0,079288	-0,164550
Junho 1986	-0,044251	-0,099254	Dezembro 1988	-0,036233	0,150269
Julho 1986	-0,112070	-0,226846	Janeiro 1989	-0,011494	-0,015672
Agosto 1986	0,030226	0,073445	Fevereiro 1989	-0,093729	-0,037860
Setembro 1986	-0,129556	-0,143064	Março 1989	-0,065215	-0,074712
Outubro 1986	0,001319	0,034776	Abril 1989	-0,037113	-0,108530
Novembro 1986	-0,033679	-0,063375	Mai 1989	-0,044399	-0,036769
Dezembro 1986	-0,072795	-0,058735	Junho 1989	-0,084412	0,023912
Janeiro 1987	0,073396	0,050214	Julho 1989	0,003444	-0,078430
Fevereiro 1987	-0,011618	0,111165	Agosto 1989	-0,056760	-0,132199
Março 1987	-0,026852	-0,127492	Setembro 1989	-0,078970	-0,110141
Abril 1987	-0,040356	0,054522	Outubro 1989	-0,105367	-0,126302
Mai 1987	-0,047539	-0,072918	Novembro 1989	-0,038634	-0,095730
Junho 1987	-0,001732	-0,058979	Dezembro 1989	-0,043261	0,065740
Julho 1987	-0,008899	0,236147	Janeiro 1990	-0,139773	-0,120056
Agosto 1987	-0,020837	-0,094778	Fevereiro 1990	-0,059094	-0,085205
Setembro 1987	-0,084811	-0,135669	Março 1990	-0,057736	-0,130433
Outubro 1987	-0,262077	-0,284796	Abril 1990	-0,102524	-0,116728
Novembro 1987	0,110167	-0,171494	Mai 1990	0,023881	-0,078039
Dezembro 1987	0,034955	-0,242616	Junho 1990	-0,079116	-0,170322
Janeiro 1988	0,012688	-0,063518	Julho 1990	-0,078965	-0,077727
Fevereiro 1988	-0,002170	-0,117677	Agosto 1990	-0,161359	-0,277035
Março 1988	-0,073462	0,201674	Setembro 1990	-0,119376	-0,207595
Abril 1988	-0,043419	-0,147728	Outubro 1990	-0,076008	-0,070515
Mai 1988	-0,054730	-0,170885	Novembro 1990	-0,006444	-0,046274
Junho 1988	-0,011755	-0,014893	Dezembro 1990	-0,026401	-0,190834

Fonte: Simonoff e Tsai (1994, p. 358)

---

APÊNDICE H

Coelhos europeus na Austrália

---

Tabela H.1 *Pesos das lentes dos olhos de coelhos europeus (y), em miligramas e idade (x), em dias, numa amostra de 71 observações.*

<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
15	21,66	98	104,30	285	189,66
15	22,75	125	134,90	300	186,09
15	22,30	142	130,68	301	186,70
18	31,25	142	140,58	305	186,80
28	44,79	147	155,30	312	195,10
29	40,55	147	152,20	317	216,41
37	50,25	150	144,50	338	203,23
37	46,88	159	142,15	347	188,38
44	52,03	165	139,81	354	189,70
50	63,47	183	153,22	357	195,31
50	61,13	192	145,72	375	202,63
60	81,00	195	161,10	394	224,82
61	73,09	218	174,18	513	203,30
64	79,09	218	173,03	535	209,70
65	79,51	219	173,54	554	233,90
65	65,31	224	178,86	591	234,70
72	71,90	225	177,68	648	244,30
75	86,10	227	173,73	660	231,00
75	94,60	232	159,98	705	242,40
82	92,50	232	161,29	723	230,77
85	105,00	237	187,07	756	242,57
91	101,70	246	176,13	768	232,12
91	102,90	258	183,40	860	246,70
97	110,00	276	186,26		

Fonte: Wei (1998, p. 155)

---

## APÊNDICE I

### Solução de cálcio radioativo

---

Tabela I.1 *Quantidade absorvida de cálcio radioativo (y), em nmoles/mg, e tempo (x), em minutos, numa amostra de 27 observações.*

<i>x</i>	<i>y</i>
0,45	0,34170
0,45	-0,00438
0,45	0,82531
1,30	1,77967
1,30	0,95384
1,30	0,64080
2,40	1,75136
2,40	1,27497
2,40	1,17332
4,00	3,12273
4,00	2,60958
4,00	2,57429
6,10	3,17881
6,10	3,00782
6,10	2,67061
8,05	3,05959
8,05	3,94321
8,05	3,43726
11,15	4,80735
11,15	3,35583
11,15	2,78309
13,15	5,13825
13,15	4,70274
13,15	4,25702
15,00	3,60407
15,00	4,15029
15,00	3,42484

Fonte: Rawlings (1988)

---

## Referências

---

- Barndorff-Nielsen, O.E. (1983). On a formula to the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.
- Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters, based on the standardized signed log likelihood ratio. *Biometrika*, **73**, 307–322.
- Barndorff-Nielsen, O.E. (1991). Modified signed log likelihood ratio. *Biometrika*, **78**, 557–563.
- Barndorff-Nielsen, O.E. (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *Journal of the Royal Statistical Society B*, **56**, 125–140.
- Barndorff-Nielsen, O.E. e Blaesild, P. (1986). A note on the calculation of Bartlett adjustments. *Journal of the Royal Statistical Society B*, **48**, 353–358.
- Barndorff-Nielsen, O.E. e Cox, D.R. (1984). Bartlett adjustments to the likelihood ratio statistics and the distribution of the maximum likelihood estimator. *Journal of the Royal Statistical Society B*, **46**, 483–495.
- Barndorff-Nielsen, O.E. e Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*, **75**, 374–378.
- Bartlett, M.S. (1937). Properties of sufficiency and statistical test. *In Proceeding of the Royal Society A*, **160**, 268–282.
- Bates, D.M. e Watts, D.G. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.
- Becker, R.A., Chambers, J.M. e Wilks, A.R. (1998). *The New S Language*. New York: Chapman and Hall.
- Bellio, R. e Brazzale, A.R. (1999). Higher-order likelihood-based inference in nonlinear regression. In Proceedings of the 14th international workshop on statistical

- modelling (eds. H. Friedl, A. Berghold & G. Kauermann), 440–443. Technical University, Graz.
- Bellio, R. e Brazzale, A.R. (2003). Higher-order asymptotics unleashed: software design for nonlinear heteroscedastic models. *Journal of Computational and Graphical Statistics*, **12**, 682–697.
- Chambers, J.M. e Hastie, T.J. (1992). *Statistical Models in S*. New York: Chapman and Hall.
- Cordeiro, G.M. (1993). Bartlett corrections and bias correction for two heteroscedastic regression models. *Communications in Statistics, Theory and Methods*, **22**, 169–188.
- Cordeiro, G.M. e Botter, D. A. (2001). Second-order biases of maximum likelihood estimates in overdispersed linear models statistics. *Statistics and Probability Letters*, **55**, 269–280.
- Cordeiro, G.M. e McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B*, **53**, 629–634.
- Cordeiro, G.M. e Paula, G.A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika*, **76**, 93–100.
- Cox, D.R. e Hinkley, D.V. (1974). *Theoretical Statistics*. New York, John Wiley.
- Cox, D.R. e Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, **49**, 1–39.
- Cox, D.R. e Reid, N. (1993). A note on the calculation of adjusted profile likelihood. *Journal of the Royal Statistical Society B*, **55**, 467–471.
- Cribari-Neto, F. e Cordeiro, G.M. (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews*, **15**, 339–367.
- DiCiccio, T.J. e Stern, S.E. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *Journal of the Royal Statistical Society B*, **56**, 397–408.
- Doornik, J.A. (2001). *Ox: an Object-oriented Matrix Programming Language*, 4th ed. London: Timberlake Consultants and Oxford: <http://www.nuff.ox.ac.uk/Users/Doornik/>.

- Ferrari, S.L.P. e Cribari-Neto, F. (2002). Corrected modified profile likelihood heteroskedasticity tests. *Statistics and Probability Letters*, **57**, 353–361.
- Ferrari, S.L.P., Lucambio, F. e Cribari-Neto, F. (2003). Adjusted and Bartlett-adjusted profile likelihood ratio tests. (Submetido para publicação).
- Firth, D. (1991). Generalized linear models. In *Statistical Theory and Modelling*. Hinkley, D.V., Reid, N. and Snell, E.J. (eds.) 55–82. London: Chapman and Hall.
- Houaiss (2001). Dicionário Eletrônico Houaiss 1.0 (BR).
- Ihaka, R. e Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational Graphics and Statistics*, **5**, 299–314.
- Lawley, D.N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika*, **71**, 233–244.
- Luenberger, D.G. (1973). *Introduction to Linear and Nonlinear Programming*. Reading: Addison-Wesley.
- Lyon, J.D. e Peters, D. (2000). Applying Skovgaard's modified directed likelihood statistic to mixed linear models. *Journal of Statistical Computation and Simulation*, **65**, 225–242.
- McCullagh, P. e Cox, D.R. (1986). Invariants and likelihood ratio statistics. *The Annals of Statistics*, **10**, 65–80.
- McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- Montgomery, D.C., Peck, E. A. e Vining, G.G. (2001). *Introduction to Linear Regression Analysis*, 3rd ed. New York: Wiley.
- Nocedal, J. & Wright, S.J. (1999). *Numerical Optimization*. New York: Springer-Verlag.
- Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. 2.ed. New York: Wiley. 625p.
- Ratkowsky, D.A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach*. New York: Marcel Dekker.
- Ratkowsky, D.A. (1990). *Handbook of Nonlinear Regression Models*. New York: Marcel dekker.

- Rawlings, J.O. (1988). *Applied Regression Analysis*. Wadsworth and Brooks, Cole Statistics/ Probability Series.
- Seber, G.A.F. e Wild, C.J. (1989). *Nonlinear Regression*. New York: Wiley.
- Severini, T.A. (1998). An approximation to the modified profile likelihood function. *Biometrika*, **85**, 403-411.
- Simonoff, J.S. e Tsai, C.-H. (1994). Use of modified profile likelihood for improved tests of constancy of variance in regression. *Applied Statistics*, **43**, 357-370.
- Skovgaard, I.M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, **2**, 145-165.
- Skovgaard, I.M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics*, **28**, 3-32.
- Smyth, G.K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society B*, **51**, 47-60.
- Stern, S.E. (1997). A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *Journal of the Royal Statistical Society B*, **59**, 653-665.
- Wei, B.C. (1998). *Exponential Family Nonlinear Models*. Singapore: Springer-Verlag.
- Wei, B.C., Shi, J.Q., Fung, W.K. e Hu, Y.Q. (1998). Testing for varying dispersion in exponential family nonlinear models. *Annals of the Institute of Statistical Mathematics*, **50**, 2, 277-294.
- Yi, G.Y., Wu J, e Liu, Y. (2002). Implementation of higher-order asymptotics to S-plus. *Computational Statistics & Data Analysis*, **40**, 775-800.