Correção tipo-Bartlett nos modelos não-lineares simétricos heteroscedásticos

Kátia Pires do Nascimento, Audrey Helen Mariz de Aquino Cysneiros.
Departamento de Estatística, UFPE, Departamento de Matemática computacional, UFPE.
(katia1pires@hotmail.com.br, audrey@de.ufpe.br)

Resumo

Na última década, diversos resultados de natureza teórica e aplicados surgiram como alternativas à modelagem com erros normais como, por exemplo, o uso de distribuições simétricas (ou elípticas). Grande parte desses resultados podem ser encontrados em Fang et al. (1990) e Fang e Anderson (1990). Esta classe de distribuições contempla distribuições de caudas leves e pesadas, tais como, t de Student, Logística tipo I e II, Normal, Normal Contaminada, dentre outras. Na classe dos modelos não–lineares simétricos, abordamos a situação em que os parâmetros de dispersão não são constantes para todas as observações, havendo assim uma estrutura heteroscedástica. Neste trabalho, desenvolvemos um fator de correção tipo–Bartlett para a estatística escore nos modelos não lineares simétricos heteroscedásticos, com funções de ligação quaisquer para a média e para o parâmetro de dispersão. A fórmula da correção é dada em notação matricial e pode ser implementada em um sistema de computação algébrica para se obter expressão em forma fechada quando aplicada a modelos especiais. Este fator de correção obtido generaliza o resultado em Cysneiros et al. (2008), já que estes autores desenvolveram um fator de correção tipo–Bartlett para a estatística escore nos modelos não–lineares simétricos homoscedásticos. Comparação do desempenho do teste escore com suas respectivas versões corrigidas foram feitas em termos de tamanho e poder, em amostras pequenas.

Palavras-chave: Correção tipo-Bartlett, Distribuições simétricas, Modelos heteroscedásticos, Modelos não–lineares, Teste escore.

1 Modelo não–linear simétrico heteroscedástico

As variáveis aleatórias $Y_1, \cdots, Y_n$ são assumidas como sendo independentes e cada observação Y_l tem uma distribuição simétrica, com parâmetros de locação $\mu_l \in R$ e de escala $\phi_l > 0$, dado por

$$\pi(y_l; \mu_l, \phi_l) = \frac{1}{\sqrt{\phi_l}} g\left(u_l\right), y_l \in R, \tag{1}$$

sendo $g : R \to [0, \infty)$ tal que $\int_0^\infty g(u)du < \infty$ e $u_l = \phi_l^{-1}(y - \mu_l)^2$. A função $g(\cdot)$ é tipicamente conhecida como função geradora de densidades, com $g(u) \geq 0$. Denotamos $Y_l \sim S(\mu_l, \phi_l, g)$.

A função densidade de probabilidade de $Z_l = (Y_l - \mu_l)/\sqrt{\phi_l}$ é dada por $\pi(u;0,1) = g(u^2)$, $u \in R$, isto é $Z_l \sim S(0,1,g)$.

Os modelos não–lineares simétricos heteroscedásticos são definidos por (1) e pelas componentes sistemáticas

$$\mu_l = f(x_l;\beta) \quad \text{e} \quad \phi_l = h(\tau_l), \quad l = 1, \cdots, n \tag{2}$$

em que μ e ϕ são vetores de dimensão $n\times 1$, $\tau_l = z_l^\top \gamma$ sendo $z_l = (z_{l1}, \cdots, z_{lq})^\top$ um vetor de dimensão $q\times 1$ de variáveis explicativas, γ e β são vetores de parâmetros desconhecidos a serem estimados de dimensões $p\times 1$ e $q\times 1$, respectivamente, $x_l = (x_{l1}, \cdots, x_{lm})^\top$ é um vetor de dimensão $m\times 1$ de variáveis explicativas, associadas com a i-ésima resposta. As funções $f(.;.)$ e $h(\cdot)$ são conhecidas, contínuas e duplamente diferenciáveis e denominadas funções de ligação da média e de dispersão, respectivamente. Além disso, a matriz de derivadas, de dimensão $n \times p$, denotada por $\widetilde{X} = \partial\mu/\partial\beta$ é de posto completo, isto é, posto$(\tilde{X}) = p$, para todo β. A matriz $\tilde{X}$ tem elementos que são, em geral, funções do vetor de parâmetros β desconhecidos.

Seja $\ell(\theta)$ o logaritmo da função de verossimilhança, definido como

$$\ell(\theta) = -\frac{1}{2}\sum_{l=1}^{n} log\phi_l + \sum_{l=1}^{n} t\left(z_l\right),$$

em que $t\left(z_l\right) = \log\, g\left(z_l^2\right)$, com $z_l = \sqrt{u_l} = \frac{y_l - \mu_l}{\sqrt{\phi_l}}$.

A função $\ell(\theta)$ é assumida regular (vide Cox e Hinkley, 1974, cap. 9) com respeito às derivadas dos componentes de β e ϕ até a quarta ordem. Condições de regulares, também, são encontradas em Serfling (1980, p. 144).

Para o modelo definido em (1) e (2), estamos interessados em fazer inferências que envolvessem alguns, mas não todos os parâmetros do modelo. Considere a seguinte partição dos vetores de parâmetros: $\beta = (\beta_1^\top, \beta_2^\top)^\top$ e $\gamma = (\gamma_1^\top, \gamma_2^\top)^\top$, sendo $\beta_1 = (\beta_1, \cdots, \beta_{p_1})$, $\beta_2 = (\beta_{p_1}, \cdots, \beta_p)$, $\gamma_1 = (\gamma_1, \cdots, \gamma_{q_1})$ e $\gamma_2 = (\gamma_{q_1}, \cdots, \gamma_q)$. Neste modelo, estamos interessados em testar a hipótese $\mathcal{H}_0^1 : \beta_1 = \beta_1^{(0)}$, $\gamma_1 = \gamma_1^{(0)}$, contra $\mathcal{H}_1^1$: pelo menos uma das igualdades é violada, em que $\beta_1^{(0)}$ e $\gamma_1^{(0)}$ são vetores de parâmetros de interesse, fixos de dimensões $p_1 \times 1$ e $q_1 \times 1$, respectivamente e β_2 e γ_2 são vetores de parâmetros de perturbação de dimensões $(p-p_1)\times 1$ e $(q-q_1)\times 1$. Além disso, considere $X = (X_1, X_2)$ e $P = (P_1, P_2)$ as matrizes particionadas correspondentes ao modelo, em que X_1, X_2, P_1 e P_2 são respectivamente, de dimensões $n\times p_1$, $n\times(p-p_1)$, $n\times q_1$ e $n\times(q-q_1)$. Utilizamos também a seguinte notação: $\delta\left(a,b,c,d,e\right) = \mathrm{E}\left\{t^{(1)a}t^{(2)b}t^{(3)c}t^{(4)d}z_l^e\right\}$, com $a,b,c,d,e = 1,2,3,4$.

A matriz de informação total de Fisher $K = K(\beta,\gamma)$ é bloco diagonal com submatrizes $K_{\beta_{11}} = \delta_{(2,0,0,0,0)}$ $X_1^\top \Lambda X_1$, $K_{\beta_{12}} = K_{\beta_{21}}^\top = \delta_{(2,0,0,0,0)}\, X_1^\top \Lambda X_2$, $K_{\beta_{22}} = \delta_{(2,0,0,0,0)}\, X_2^\top \Lambda X_2$, $K_{\gamma_{11}} = (\delta_{(2,0,0,0,2)} - 1)/4\, P_1^\top \Lambda_5 P_1$ e $K_{\gamma_{12}} = K_{\gamma_{21}}^\top = (\delta_{(2,0,0,0,2)} - 1)/4\, P_1^\top \Lambda_5 P_2$, $K_{\gamma_{22}} = (\delta_{(2,0,0,0,2)} - 1)/4\, P_2^\top \Lambda_5 P_2$, sendo Λ e Λ_5 matrizes disgonais cujos l-ésimos elementos são dados por $\lambda_l = \frac{h_l'^2}{\phi_l^2}$ e $\lambda_{5l} = \frac{1}{\phi_1}$, respectivamente, para $l = 1, \cdots, n$. Aqui, "$\prime$" denota a derivada com respeito a τ.

A estatística escore para testar a hipótese $\mathcal{H}_0^1 : \beta_1 = \beta_1^{(0)}$, $\gamma_1 = \gamma_1^{(0)}$ pode ser escrita como a soma de duas formas quadráticas, a saber:

$$S_R = \widetilde{r}^\top \widetilde{X}_1(\widetilde{X}_1^\top \Lambda \widetilde{X}_1)^{-1}\widetilde{X}_1^\top \widetilde{r} \,+\, \widetilde{\zeta}^\top\, \widetilde{P}_1 \left(\widetilde{P}_1^\top \Lambda_5 \widetilde{P}_1\right)^{-1} \widetilde{P}_1^\top\, \widetilde{\zeta},$$

em que $\tilde{P} = \frac{\partial\phi}{\partial\gamma}$ e $\tilde{X} = \frac{\partial\mu}{\partial\beta}$ são matrizes de derivadas, $r = (r_1, \cdots, r_n)^\top$ com $r_l = \Lambda^{1/2}\, s_l\, z_l \delta_{(2,0,0,0,0)}^{-1/2}$, $S = \mathrm{diag}\left\{\mathrm{s_1}, \cdots, \mathrm{s_n}\right\}$ com $s_l = \frac{-2g^{'}(u_l)}{g(u_l)}$ sendo $u_l = \frac{(y_l-\mu_l)^2}{\phi_l}$, $\zeta = (\zeta_1, \cdots, \zeta_n)^\top$ com $\zeta_l = (u^\top F_1 S^\top - F_1^\top \iota)(\delta_{(2,0,0,0,2)} - 1)^{-1/2}\Lambda$, $F_1 = \mathrm{diag}\left\{\mathrm{h}_1', \cdots, \mathrm{h}_n'\right\}$, $u = (u_1, \cdots, u_n)^\top$ e ι é um vetor $n\times 1$ de uns.

2 Correção Tipo–Bartlett

Sabe-se que em pequenas amostras a aproximação da distribuição das estatística escore pela distribuição χ^2 pode não ser satisfatória. No entanto, para melhorar a qualidade da aproximação da distribuição da estatística escore pela distribuição χ^2 utiliza-se a correção tipo–Bartlett.

Cordeiro e Ferrari (1991) mostraram que, em problemas regulares, a estatística escore, S_R, pode ser melhorada por uma correção tipo–Bartlett que não é exatamente a correção de Bartlett porque envolve um polinômio de segundo grau na estatística original, produzindo uma estatística escore modificada ajustada com distribuição χ^2 até ordem n^{-1}, segundo $\mathcal{H}_0$. Cordeiro e Ferrari (1991) propuseram a estatística escore modificada, dada por

$$S_R^* = S_R\left[1 - \frac{A_3}{12u(u+2)(u+4)}S_R^2 - \frac{A_2 - 2A_3}{12(u+2)}S_R - \frac{A_1 - A_2 + A_3}{12u}\right],$$

sendo A_1, A_2, e A_3 funções de cumulantes conjuntos de derivadas do logaritmo da função de verossimilhança. Entretanto, a estatística escore aperfeiçoada (S_R^*) nem sempre é uma transformação monótona, assim, para solucionar esse problema Kakizawa (1996) sugeriu uma transformação monótona dada por $S_{R1}^* = S_R^* + P(S_R)$, envolvendo a própria estatística escore e os coeficientes a, b e c, sendo $P(S_R)$ dada por

$$P(S_R) = \frac{1}{4}\left\{c^2 S_R + 2bcS_R^2 + \left(2ac + \frac{4}{3}b^2\right)S_R^3 + 3abS_R^4 + \frac{9}{5}S_R^5\right\}.$$

Posteriormente, Cordeiro et al. (1998) também apresentaram uma fórmula para a estatística escore aperfeiçoada, de modo que também fosse uma transformação monótona de S_R dada da seguinte forma

$$S_{R2}^* = \begin{cases} \sqrt{\frac{\pi}{3a}}\exp\left(\frac{b^2}{3a} - c\right)\times \\ \left\{\Phi\left(\sqrt{6a}S_R + \sqrt{\frac{2}{3a}b}\right) - \Phi\left(\sqrt{\frac{2}{3a}b}\right)\right\}, \; se \; a > 0, \\ \frac{1}{2b}\exp(-c)\left\{1 - \exp(-2bS_R)\right\}, \; se \, a = 0 \; e \; b \neq 0. \end{cases}$$

Neste trabalho, derivamos o fator de correção tipo–Bartlett nos modelos não–lineares simétricos heteroscedásticos. Os termos A_1, A_2 e A_3 são expressos, em notação matricial, e serão omitidos aqui por motivo de espaço. Somente serão apresentado os termos de A_1, A_2 e A_3, em notação matricial, no caso do teste da hipótese nula $\mathcal{H}_0^2 : \gamma = \gamma^{(0)}$ contra a alternativa $\mathcal{H}_1^2 : \gamma \neq \gamma^{(0)}$, em que β é o vetor de parâmetros de perturbação e $\gamma^{(0)}$ é um valor especificado para γ.

Os termos para A_1, A_2 e A_3 são dados por:

$$\begin{aligned} A_1 \;\; = \;\; & 3(\delta_{(2,0,0,0,0)})^{-1}(\iota^\top Q_2\Lambda_6(Z_\beta - Z_{2\beta})\Lambda_6\, Q_2\iota) \\ & -b_1\left(\delta_{(0,0,1,0,1)} + 2\delta_{(0,1,0,0,0)}\right)\iota^\top\Lambda_4 Z_{2\beta d}Z_\gamma Z_{2\beta d}\Lambda_4\iota \\ & -6\iota^\top\Lambda_6(Z_\beta - Z_{2\beta})(Z_\beta - Z_{2\beta})\Lambda_6 \odot J\iota \\ & +\, 3(\delta_{(3,0,0,0,1)} - \delta_{(1,0,0,0,1)})\,\iota^\top\Lambda_4 Z_{2\beta}\odot Z_\gamma \odot Z_{2\beta}\Lambda_4,\iota \\ & +b_{15}tr\left\{\Lambda_6 Z_{2\beta d}(Z_{\beta d} - Z_{2\beta d})\right\} \\ & +b_{14}\,tr\{\Lambda_7 Z_{2\beta d}Z_{\gamma d}\}, \end{aligned}$$

$$A_2 \;\; = \;\; +b_9\,(1 - 3\delta_{(0,1,0,0,2)} - \delta_{(0,0,1,0,3)})1^\top\Lambda_1 Z_{\gamma d}Z_\gamma Z_{\gamma d}\Lambda_1\iota$$

$$
\begin{aligned}
&-b_{10}\,(\delta_{(2,0,0,0,2)}-1)(1-3\delta_{(0,1,0,0,2)}-\delta_{(0,0,1,0,3)})\iota^\top\Lambda_1 Z_{\gamma d}Z_\gamma Z_{\beta d}\Lambda_4\iota\\
&+2b_1\,(\delta_{(3,0,0,0,1)}-\delta_{(0,1,0,0,0)})\iota^\top\Lambda_4 Z_{2\beta d}Z_\gamma(Z_{\beta d}-Z_{2\beta d})\Lambda_4\iota\\
&-\frac{4\,b_1\,\delta_{(2,0,0,0,0)}}{(\delta_{(2,0,0,0,2)}-1)}(1-3\delta_{(0,1,0,0,2)}-\delta_{(0,0,1,0,3)})\iota^\top\Lambda_4 Z_{2\beta d}Z_\gamma Z_{\gamma d}\Lambda_1\iota\\
&+2\,\iota^\top\Lambda_4(Z_\beta-Z_{2\beta})\odot Z_{2\beta}(Z_\gamma-Z_{2\gamma})\Lambda_4\iota\\
&+\frac{3(\delta_{(4,0,0,0,0)}-3\delta^2_{(0,1,0,0,0)})}{\delta^2_{(2,0,0,0,0)}}tr\left\{\Lambda_6(Z_{\beta d}-Z_{2\beta d})^2\right\}\\
&+\frac{3\left(\delta_{(0,1,0,0,2)}+2\delta_{(3,0,0,0,1)}+\delta_{(4,0,0,0,2)}\right)}{(\delta_{(2,0,0,0,2)}-1)\delta_{(2,0,0,0,0)}}tr\left\{\Lambda_7(Z_{\beta d}-Z_{2\beta d})Z_{\gamma d}\right\}\\
&+\frac{3}{(\delta_{(2,0,0,0,2)}-1)^2}\{-6+\delta_{(2,0,0,0,2)}(12-3\delta_{(2,0,0,0,2)})\\
&+4\delta_{(3,0,0,0,3)}+\delta_{(4,0,0,0,4)}\}tr\{\Lambda_9(Z_{\gamma d})^2\},
\end{aligned}
$$

$$
\begin{aligned}
A_3 \;=\; &-\frac{b_{11}}{2\delta_{(2,0,0,0,2)}}(\delta_{(3,0,0,0,1)}-\delta_{(0,1,0,0,0)})\iota^\top\Lambda_5(Z_{\beta d}-Z_{2\beta d})Z_\gamma(Z_{\beta d}-Z_{2\beta d})\Lambda_5\iota\\
&-\frac{b_{11}\delta_{(2,0,0,0,2)}}{(\delta_{(2,0,0,0,2)}-1)}(-1+3\delta_{(0,1,0,0,2)}+\delta_{(0,0,1,0,3)})\iota^\top\Lambda_1 Z_{\gamma d}Z_\gamma(Z_{\beta d}-Z_{2\beta d})\Lambda_4\iota\\
&-\frac{b_{10}\,\delta_{(2,0,0,0,2)}}{18\,(\delta_{(2,0,0,0,2)}-1)}(1-3\delta_{(0,1,0,0,2)}-\delta_{(0,0,1,0,3)})^2\iota^\top\Lambda_1(Z_{\beta d}-Z_{2\beta d})Z_\gamma Z_{\gamma d}\Lambda_4\iota\\
&-b_9\left(1-3\delta_{(0,1,0,0,2)}-\delta_{(0,0,1,0,3)}\;\right)\iota^\top\Lambda_1 Z_{\gamma d}Z_\gamma Z_{\gamma d}\Lambda_1\iota\\
&-b_{11}(\delta_{(3,0,0,0,1)}-\delta_{(0,1,0,0,0)})\,\iota^\top\Lambda_4(Z_\beta-Z_{2\beta})\odot Z_\gamma\odot(Z_\beta-Z_{2\beta})\iota\\
&-\frac{2\,b_9}{3}(1-3\delta_{(0,1,0,0,2)}-\delta_{(0,0,1,0,3)})\iota^\top\Lambda_1 Z_\gamma\odot Z_\gamma\odot Z_\gamma\iota,
\end{aligned}
$$

em que b_1, b_9, b_{11}, b_{14} e b_{15} são:

$$
b_1=\frac{3(\delta_{(1,1,0,0,1)}-\delta_{(0,1,0,0,0)})}{(\delta_{(2,0,0,0,2)}-1)\delta^2_{(2,0,0,0,0)}},\;b_9=\frac{-12\left(1-3\delta_{(0,1,0,0,2)}-\delta_{(0,0,1,0,3)}\;\right)}{(\delta_{(2,0,0,0,2)}-1)^3},\;b_{11}=\frac{-6\left(\delta_{(3,0,0,0,1)}-\delta_{(0,1,0,0,0)}\right)}{\left(\delta_{(2,0,0,0,2)}-1\right)\delta^2_{(2,0,0,0,0)}},
$$

$$
b_{14}=\frac{-6(\delta_{(2,0,0,0,2)}\delta_{(0,1,0,0,2)}+2\delta_{(3,0,0,0,1)}+\delta_{(4,0,0,0,2)})}{(\delta_{(2,0,0,0,2)}-1)\delta_{(2,0,0,0,2)}},\;b_{15}=12\left(\frac{\delta_{(2,1,0,0,0)}}{\delta^2_{(2,0,0,0,0)}}+1\right).
$$

3 Estudo de Simulação

Apresentaremos alguns resultados de simulações para avaliar a eficácia da correção tipo–Bartlett para os testes escore, em modelos não lineares simétricos heteroscedásticos. Comparamos os desempenhos de quatro estatísticas de testes, isto é, o teste S_R, $S_R{}^*$, S^*_{R1} e S^*_{R2}. Os desempenhos são avaliados em função da proximidade das probabilidades de rejeição da hipótese nula, sendo verdadeira (probalidade de ocorrer o erro tipo I), aos respectivos níveis nominais dos testes. A hipótese nula considerada e $\mathcal{H}_0^2 : \gamma_1 = \gamma_1^{(0)}$ e a hipótese alternativa é $\mathcal{H}_1^2 : \gamma_1 \neq \gamma_1^{(0)}$, para $\gamma_1^{(0)}$ o vetor de parâmetros de interesse de dimensão $q_1 \times 1$. Sem perda de generalidade, tomamos os seguintes valores para os parâmetros da regressão: $\beta_1 = 5$, $\beta_2 = 2$, $\beta_3 = 1$, $\gamma_1 = 0$, $\gamma_2 = 0,3$ e $\gamma_3 = 0,5$. Todas as variáveis em X e Z são independentes e geradas de uma distribuição $U(0,1)$.

Utilizamos o modelo exponencial potência não–linear ($k = 0.1$), $p = 1$ e $q = 5$. O número de réplicas foi fixado em 10.000 para o tamanho de amostra $n = 40$ e consideramos os seguintes níveis nominais $\alpha = 10\%, 5\%, 1\%, 0,5\%$. As simulações foram realizadas utilizando a linguagem de programação matricial `Ox` (Doornik, 2001).

A Tabela 1 apresenta as taxas de rejeição dos diferentes testes. Podemos observar que, na maioria dos casos, as taxas de rejeição dos testes baseados nas estatísticas corrigidas tiveram valores mais próximas dos correspondentes tamanhos nominais do que as taxas de rejeição do teste escore usual. Na Tabela 2 consideramos a hipótese alternativa e diferentes valores para γ para o nível nominal 5%. A análise desta tabela mostra que comparando o poder dos testes baseados nas estatísticas S_R, S_R^*, S_{R1}^* e S_{R2}^* os resultados mostram que não há nenhuma perda de poder decorrente de se utilizar o fator de correção tipo–Bartlett. Todavia, o poder dos testes para as diferentes estatísticas, em análise, parecem semelhantes, com pequenas distorções de tamanho.

Tabela 1: Tamanho dos testes – modelo não–linear exponencial potência, com $k = 0,1$ $p = 1$ $n = 40$ e α.

	10%	5%	1%	0.5%
S_R	8,6	4,2	0,7	0,5
${S_R}^*$	9,6	5,0	1,3	0,7
S_{R1}^*	10,0	5,0	1,3	0,7
S_{R2}^*	10,0	5,0	1,4	0,7

Tabela 2: Poder dos testes – modelo exponencial potência não–linear com $k = 0,1$ $p = 1$ $n = 40$, $\alpha = 5\%$.

γ	S_R	${S_R}^*$	S_{R1}^*	S_{R2}^*
$0,0$	$4,2$	$5,0$	$5,0$	$5,0$
$1,0$	$15,1$	$16,9$	$16,8$	$26,9$
$2,0$	$49,4$	$54,5$	$54,1$	$54,3$
$3,0$	$84,6$	$84,6$	$85,5$	$84,5$
$4,0$	$97,6$	$97,8$	$97,8$	$97,8$

Referências

Cysneiros, A. H. M. A., Rodrigues, K.S.P., Cordeiro, G.M e Ferrari, S.L.P. (2008). Three-Bartlett-Type Correction for Score Statistics in Symmetric Nonlinear Regression Models. *Statistical Papers*, **00**, p. 1-13.

Cordeiro, G.M. e Ferrari, S.L.P. (1991). A modified score test statistic having chi-squared distribution to order n^{-1}. *Biometrika*, **78**, 573–582.

Cordeiro, G.M. and Ferrari, S.L.P. and Cysneiros, A.H.M.A. (1998). A formula to improve score test statistics. Journal of Statistical Computatio

Doornik, J. A. (2001). *Ox: An Object-Oriented Matrix Language.* 4th ed. Timberlake Consultants Press, London; Oxford, `http://www.doornik.com`.

Fang, K.T. e Anderson, T.W. (1990). *Statistical Inference in Elliptical Contoured and Related Distributions.* New York: Allerton Press.

Fang, K.T., Kotz, S. e Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions.* London: Chapman & Hall.

Kakizawa, Y. (1996), Higher order monotone Bartlett-type adjustment for some multivariate test statistic,*Biometrika*, **83**, 923–659.

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics. New York: John Wiley.