XXIII CONIC
VII CONITI
IV ENIC

INFERÊNCIA PARAMÉTRICA EM MODELOS NÃO LINEARES DA FAMÍLIA EXPONENCIAL

France Evellyn Gomes de Oliveira[1]; Audrey Helen Mariz de Aquino Cysneiros[2]

[1]Estudante do Curso de Estatística - CCEN – UFPE; E-mail: franceoliveirac@gmail.com,
[2]Docente/pesquisador do Depto de Estatística - CCEN – UFPE. E-mail: audrey@de.ufpe.br .

Sumário: Este trabalho primeiramente introduz os modelos não-lineares da família exponencial e apresenta alguns testes de heteroscedasticidade. Comparamos o desempenho do teste baseado na estatística da razão de verossimilhanças bootstrap corrigida (via Bartlett) com os diferentes testes propostos na literatura, baseados nas seguintes estatísticas, a saber: razão de verossimilhança original, razão de verossimilhanças perfiladas modificadas, razão de verossimilhanças perfiladas modificadas corrigida (via Bartlett) e razão de verossimilhanças bootstrap, em termos de tamanho e poder, em pequenas amostras, através de simulações de Monte Carlo, usando a linguagem de programação matricial Ox (Doornik, 2001). Os resultados favorecem aos testes da razão de verossimilhanças perfilado modificado corrigido, ${LR_m}^*$, e razão de verossimilhanças bootstrap corrigido, ${LR_{boot}}^*$, propostos.

Palavras–chave: bootstrap, correção de bartlett, modelos não lineares da família exponencial, razão de verossimilhanças, verossimilhança perfilada.

INTRODUÇÃO

Os modelos não-lineares da família exponencial 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 apresentamos alguns testes de heteroscedasticidade.

Definição do modelo

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

$$\pi(y_l;\theta_l,\phi_l) = exp\left\{\phi_l[y_l\theta_l - b(\theta_l) - c(y_l)] - \frac{1}{2}e(y;\theta_l)\right\}, l = 1, \dots, n, \qquad (1)$$

em que $e(.,.)$ $b(.)$ $c(.)$ são funções conhecidas e θ_l e ϕ_l são chamados de parâmetros canônico e de precisão, respectivamente.
Para o modelo (1), valem as relações:

$$E(y_l) = \mu_l = \frac{db(\theta_l)}{d\theta_l} \; e \; Var(y_l) = {\phi_l}^{-1}V_l, l = 1, \dots, n$$

Os modelos não-lineares da família exponencial são definidos por (1) e pelo componente sistemático

$$g(\mu_l) = \eta_l = f(x_l;\beta), \qquad (2)$$

em que $g(.)$ é uma função conhecida monótona e duplamente diferenciável, denominada função de ligação, relacionando a média μ_l de uma observação com o preditor não-linear η_l do modelo, $\beta = (\beta_1, \dots, \beta_k)^\top$ é um vetor de $k(k < n)$ parâmetros desconhecidos a serem estimados, $x_l = (x_{l1}, \dots, x_{lK})^\top$ é um vetor de dimensão K de covaráveis conhecidas associadas à l-ésima observação e $f(.;.)$ é uma função possivelmente não-linear no segundo argumento, contínua e diferenciável com respeito aos componentes de β 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 β Supomos indentificabilidade no sentido de que diferentes $\beta's$ implicam em diferentes $\eta's$. A matriz X^* tem elementos que são, em geral, funções do vetor de parâmetros β desconhecidos.

Alguns testes de heteroscedasticidade

O interesse é testar a hipótese nula $H_0: \delta = \delta_0$ (homoscedasticidade) contra a hipótese alternativa bilateral $H_1: \delta \neq \delta_0$ em que δ_0 é um vetor de dimensão p x *1* de constantes especificado tal que $m(z_l, \delta_0) = 1$ para $l = 1, \dots, n$. Assim, os números de parâmetros de interesse e de perturbação são p e $(k + 1)$, respectivamente.
O logaritmo da função de verossimilhança do vetor de parâmetros $\omega = (\delta, \beta, \gamma)$, dado o vetor de observações $(y_1, \dots, y_n)$ é dado por

$$L^* = L^*(y, \delta, \beta, \gamma) = -\frac{1}{2}\sum_{l=1}^{n}\left\{\frac{q_l}{\gamma}d_l + t(y_l) + \log\left(\frac{\gamma}{q_l}\right)\right\}$$

em que,

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

A estatística da razão de verossimilhança para o teste de H_0 pode ser escrita como

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

em que $\hat{\delta}$ é o valor de δ que maximiza L_p e que L_p depende apenas do parâmetro de interesse, obtida da verossimilhança original em que os parâmetros de perturbação são substituídos pelos seus respectivos estimadores de máxima verossimilhança obtidos para cada valor fixado dos parâmetros de interesse. Note que LR envolve a maximização da função de verossimilhança perfilada L_p.

Cordeiro (1993) obteve o fator de correção de Bartlett para a estatística da razão de verossimilhanças do teste de H_0 para H_1 no modelo heteroscedástico multiplicativo, isto é, $\omega(z_l, \delta) = exp\{z_l{}^\top\delta\}$, 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

$$h(X,R) = -\frac{1}{2}tr\left({A_d}^2\right) + \frac{1}{3}l^{\intercal}A^{(3)}l + 12l^{\intercal}A_dAA_dl + tr(A_dB_d) - l^{\intercal}\left(B^{(2)} \odot A\right)l + \frac{1}{2}l^{\intercal}B_dAB_dl + l^{\intercal}A_dAB_dl,$$

com $A = R(R^{\intercal}\ R)^{-1}R^{\intercal}$sendo $R = [lZ]$, $A_d = diag(a_{11}, \dots, a_{nn})$, $B = X(X'X)^{-1}X^{\intercal}$, $B_d = diag(b_{11}, \dots, b_{nn})$, e ι um vetor de uns.
A estatística da razão de verossimilhanças baseada na função de verossimilhança perfilada modificada para o teste de H_0 contra H_1 é dada por

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

em que $\tilde{\delta}$ é o valor de δ que maximiza a ${L_{mp}}^*$. Note que LR_m envolve maximização da função de verossimilhança perfilada modificada ${L_{mp}}^*$.
Cysneiros (2004) obteve o fator de correção de Bartlett para estatística do modelo, dado por:

$$c_m = -\frac{1}{2}tr(H_dH_d) + \frac{1}{2n}p^2 + \frac{1}{2}l^{\intercal}H_dHH_dl + \frac{1}{3}l^{\intercal}H^{(3)}l - \frac{2}{n}p + \frac{1}{n}l^{\intercal}H^{(2)}l,$$

em que $H_d = diag(h_{11}, \dots, h_{nn})$, $H^{(2)} = (h_{ls}^2)$ e $H^{(3)} = (h_{ls}^3)$.
Rocke (1989) propôs um fator de correção de Bartlett para a estatística da razão de verossimilhanças bootstrap dado por

$${LR_{boot}}^* = \frac{LR}{\overline{LR^{*b}}}q_l,$$

em que LR é o valor da estatística da razão de verossimilhanças com base na amostra original, q_1 representa o numero de restrições sob a hipótese nula sendo $\overline{LR}^{*b} = 1/B\sum_{b=1}^{B} LR^{*b}$ sendo LR^{*b} é a estatística da razão de verosimilhanças bootstrap.
.

RESULTADOS

Apresentamos uma simulação para avaliar a eficácia das correções de Bartlett nos modelos de regressão não lineares heteroscedásticos. Comparamos os desempenhos de seis testes, baseados nas seguintes estatísticas, a saber: razão de verossimilhanças original (LR) razão de verossimilhanças perfiladas modificadas (LR_m), razão de verossimilhanças bootstrap (LR_{boot}) 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. A simulação realizada é baseada no modelo de regressão não-linear dado por:

$$y_l = \beta_1 + exp\{\beta_2 x_{l2}\} + \sum_{j=3}^{k} \beta_j x_{lj} + u_l, l = 1, \dots, n,$$

Tabela 1: Tamanho dos testes – Modelo normal linear com p = 2, k = 5 e n = 30,50 e 100

n	$\alpha(\%)$	k = 5					
		LR	LR^*	LR_m	${LR_m}^*$	LR_{boot}	${LR_{boot}}^*$
30	10	25,0	16,4	9,1	10,2	11,1	10,9
	5	16,7	9,9	4,2	5,0	5,4	5,2
	1	6,6	2,8	0,8	1,0	1,2	0,9
50	10	16,2	11,2	9,2	9,8	10,4	10,2
	5	9,2	5,8	4,3	4,7	5,2	5,1
	1	2,5	1,3	0,9	1,0	1,2	1,0
100	10	12,0	9,8	9,1	9,4	9,6	9,9
	5	6,2	5,0	4,6	4,9	4,9	4,9
	1	1,6	1,0	0,9	1,0	0,9	1,0

Como uma ilustração, quando k = 5, $\alpha = 5\%$ e n = 100, as taxas de rejeição dos testes da razão de verossimilhanças, razão de verossimilhanças perfiladas modificadas e de sua versão corrigida são, respectivamente, 6,2%, 4,6%, 4,9%, 5%, 4,9% e 4,9%. Podemos observar que, para tamanho de amostra pequeno os testes LR e ${LR_{boot}}^*$ são liberais enquanto que o teste LR_m é conservativo. Conforme o tamanho da amostra cresce as taxas desses testes tornam-se próximas dos níveis nominais. A aplicação da correção de Bartlett para estes testes produziu resultados favoráveis trazendo as taxas de rejeição mais próximas dos níveis nominais.

CONCLUSÕES

O teste baseado na estatística LR é, em geral, liberal. 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 tornando-o conservativo. Os testes baseados nas estatísticas ${LR_m}^*$ e ${LR_{boot}}^*$ levam as taxas de rejeição para valores ainda mais próximos dos níveis nominais.

AGRADECIMENTOS

Agradeço primeiramente a Deus, a minha família e amigos pela força e compreensão dada em todo período do estudo cientifico, ao CNPq e à UFPE pelo investimento na pesquisa e à minha orientadora Audrey Cysneiros que contribuiu para a realização deste trabalho.

REFERÊNCIAS

Cysneiros, A.H.M.A e S.L.P Ferrari (2006). An improved likelihood ratio test for varying dispersion in exponential family nonlinear models. Statistics e probability letters 76(3), 255–265.

Cordeiro, G.M. (1993). Bartlett corrections and bias correction for two heteroscedastic regression models. Communications in Statistics, Theory and Methods, 22, 169–188.

Rocke, D. (1989). Bootstrap Bartlett adjustment in seemingly unrelated regression. Journal of the American Statistical Association 84(406), 598–601.