

Estudo do modelo de regressão polinomial com erros nas variáveis quando existe heterogeneidade na razão das variâncias

RESUMO: O presente trabalho pretende fazer um estudo dos modelos de regressão com erros nas variáveis. Este problema está presente nos modelos econométricos, e tem repercussão maior quando as variâncias dos erros de medição da variável dependente são heterogêneas e alguma correção deve ser sugerida em ela. Para este estudo considerou-se o modelo de correção da função escore com a finalidade de obter estimadores não viciados no modelo.

Arturo Alejandro Z. Zavala*

* Departamento de Economia – Faculdade de Economia – FE – Universidade Federal de Mato Grosso – UFMT

1 Introdução

Na atualidade, muitos autores vêm discutindo o problema da estimação dos modelos de regressão polinomial com erros nas variáveis, variados casos são apresentados na literatura estatística, entre elas temos a Wolter e Fuller (1982) que encontraram estimadores consistentes a partir da construção de uma matriz de momentos, eles só trabalharam com o modelo quadrático. Carroll, Rupert e Stefanski (1995) obtiveram estimadores consistentes ao considerar o algoritmo SIMEX, desenvolvido por Cook e Stefanski. Cheng e Schneeweis (1998) consideraram a técnica de mínimos quadrados ajustados, para a obtenção dos estimadores consistentes, em sua metodologia criaram uma nova variável denominada “T” que se relacionava diretamente com a variável observável (W) e relacionada a traves do valor esperado com a variável não observável (X). Esta metodologia foi chamado por eles como método de mínimos quadrados ajustados, Cheng e Van Ness, estudaram de uma forma geral o modelo polinomial de Berkson.

Nakamura (1990) estudou os modelos lineares generalizados e obteve como resultado que uma função escore pode ser corrigida a partir da correção da função de verossimilhança “naive”, para isto, é necessário que exista permutabilidade entre a derivada da função e a esperança condicional da função com a finalidade de que a função escore corrigida seja gradiente da função de verossimilhança corrigida, o problema principal deste método é justamente a necessidade da permutabilidade. Uma definição expressa no artigo de Nakamura diz

Definição 1 Seja θ o espaço paramétrico e $\theta = (\beta^T, \Sigma_u^T)^T$, um vetor de parâmetros que desejamos estimar, onde $\Sigma_u = \text{diag}(\sigma_{u1}^2, \sigma_{u2}^2, \dots, \sigma_{un}^2)$, $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$, e $\mathfrak{S}$, um conjunto aberto e conexo de θ , se $E[\bullet | X, Y]$ e $\frac{\partial}{\partial \theta}$, foram operadores permutáveis, então a gradiente de logaritmo da verossimilhança corrigida, cumpre a seguinte condição:

$$E[U^*(\theta; W, Y) | X, Y] = U(\theta; X, Y), \quad \forall \theta \in \mathfrak{S}$$

Quando existe a falta de permutabilidade outra solução é apresentada, Gimenez e Bolfarine (1997) desenvolveram uma metodologia onde eles não se preocuparam em corrigir a função de verossimilhança, mas sim em corrigir a função escore, eles

trabalharam a função de verossimilhança “naive” e a partir dela obtiveram a função escore “naive”, corrigindo esta ultima fazendo uso da metodologia proposta por Nakamura.

Uma nova metodologia foi apresentada definindo do modo seguinte

Definição 2. Uma função $U^*(\theta; W, Y)$, é dita função escore corrigida se:

$$E[U^*(\theta; W, Y) | X, Y] = U(\theta; X, Y), \quad \forall \theta \in \mathfrak{S}$$

ambos métodos apresentam uma variância de tipo “Sandwich”.

O objetivo do presente trabalho é apresentar o modelo polinomial com erros nas variáveis quando $\lambda_i = \sigma_{e_i}^2 / \sigma_u^2$ é conhecido, sendo que σ_u^2 é desconhecido, num modelo funcional, assim como apresentar os estimadores consistentes e os testes de hipóteses apropriados.

2 Descrição do processo de estimação

Consideremos o seguinte modelo polinomial com erros nas variáveis, sendo as variáveis x_i fixas (modelo funcional)

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_p X_i^p + e_i \quad \dots(1)$$

$$W_i = X_i + u_i \quad \dots(2)$$

para $i = 1, \dots, n$ onde Y_i é conhecida como variável resposta, os β_j são parâmetros desconhecidos que desejamos estimar, e_i é uma variável aleatória que supomos que se distribui normalmente com média zero e variância $\sigma_{e_i}^2$, isto indica, que por cada observação existirá uma variância diferente, X_i é uma variável não observável, dado um modelo funcional esta variável não tem distribuição própria, esta variável pode ser observado através de W_i , e u_i que é o erro presente na variável W_i , desta forma a variável u_i é uma variável aleatória que assumiremos que tem distribuição normal com média zero e variância σ_u^2 .

Em resumo as variáveis que representam os erros de medição têm a seguinte distribuição

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda_i & 0 \\ 0 & 1 \end{pmatrix} \sigma_u^2 \right] \quad \dots(3)$$

Para o presente estudo consideramos como ponto de partida o modelo que associa as variáveis observáveis, isto é, Y_i e W_i (modelo “naive”) sendo esta

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 W_i^2 + \beta_p W_i^p + e_i \quad \dots (4)$$

este modelo é conhecida na literatura estatística como modelo “naive” ou “ingênua”, já que W_i não é a variável regressora real de Y_i . O logaritmo da função de verossimilhança é definido por

$$\ell(\beta, \sigma_u^2; Y_i, W_i) = -\frac{1}{2} \log(\sigma_u^2) - \frac{1}{2\lambda_i \sigma_u^2} \left(Y_i - \sum_{i=1}^p \beta_i W_i^{i+s} \right)^2 \quad \dots (5)$$

Como λ_i é conhecido claramente, nossos parâmetros de interesse são β_j e σ_u^2 .

Para este trabalho considera-se o método de Gimenez e Bolfarine (1997), já que o método de Nakamura (1990) não consegue apresentar estimadores consistentes para σ_u^2 , isto é devido a exigência de permutabilidade entre a derivada e a esperança condicional não é satisfeita.

2.1 Metodologia

Aplicando as derivadas parciais a expressão (5) com respeito a β_j , temos que:

$$\frac{\partial \ell}{\partial \beta_j} = \frac{1}{\sigma_u^2} \sum_{i=1}^p \frac{1}{\lambda_i} \left(Y_i W_i^j - \sum_{s=0}^p \beta_s W_i^{j+s} \right) \quad j = 0, 1, \dots, p \quad \dots (6)$$

e com respeito a σ_u^2 , temos que

$$\frac{\partial \ell}{\partial \sigma_u^2} = -\frac{1}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} \sum_{i=1}^p \frac{1}{\lambda_i} \left(Y_i^2 - 2Y_i \sum_{j=0}^p \beta_j W_i^j - \sum_{j=1}^p \sum_{s=0}^p \beta_j \beta_s W_i^{j+s} \right) \quad (7)$$

Nas expressões (6) e (7) foram aplicados os esperados condicionais a Y_i , β_j e σ_u^2 , isto é, $E[\bullet | X_i, Y_i; \beta, \sigma_u^2]$, para facilidade de notação, nas expressões numéricas poderemos considerar $h = X_i, Y_i; \beta, \sigma_u^2$ e $r = W_i, Y_i; \beta, \sigma_u^2$, obtendo um esperado condicional que tem a forma seguinte $E[\bullet | h]$, desta forma temos

$$E \left[\frac{\partial \ell}{\partial \beta_j} \right] = \frac{1}{\sigma_u^2} \sum_{i=1}^p \frac{1}{\lambda_i} \left(Y_i E[W_i^j | h] - \sum_{s=0}^p \beta_s E[W_i^{j+s} | h] \right) \quad j = 0, 1, \dots, p \quad \dots (8)$$

e

$$E \left[\frac{\partial \ell(r | h)}{\partial \sigma_u^2} \right] = -\frac{1}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} \sum_{i=1}^p \frac{1}{\lambda_i} \left(Y_i^2 - 2Y_i \sum_{j=0}^p \beta_j E[W_i^j | h] - \sum_{j=1}^p \sum_{s=0}^p \beta_j \beta_s E[W_i^{j+s} | h] \right) \quad \dots (9)$$

para obter os valores esperados nas expressões acima, consideremos o seguinte lema:

Lema 1 Se u_i é distribuído normalmente com média zero e variância σ_u^2 , então

$$E[W_i^{m+1} | h] = X_i E[W_i^m | h] + m \sigma_u^2 E[W_i^{m-1} | h]$$

para $m = 1, 2, \dots, p-1$.

$$\text{Como } E \left[\frac{\partial \ell(r | h)}{\partial \beta_j} \right] = E[U_j(r | h)]$$

temos que

$$E[U_j(r | h)] = U_j(h) + E[\varphi(r | h)]$$

onde $\varphi(r) = \varphi(\beta, \sigma_u^2; Y_i, W_i)$ é a função que corresponde ao vício do valor esperado condicional da função escore para o β_j . Como a função escore verdadeira que relaciona X_i, Y_i é $U_j(h) = U_j(\beta, \sigma_u^2; Y_i, X_i)$, então temos que

$$U_j(h) = E[U_j(r | h)] - E[\varphi(r | h)]$$

Baseado neste ultimo fato a função escore corrigida é dado por

$$U_j^*(\beta_j; Y_i, W_i, \sigma_u^2) = U_j(\beta_j; Y_i, W_i, \sigma_u^2) - \varphi_j(\beta_j; Y_i, W_i, \sigma_u^2) \quad j = 0, 1, \dots, p \quad \dots (10)$$

De forma que

$$E[U_j^*(\beta_j; Y_i, W_i, \sigma_u^2)] = U_j(\beta_j; Y_i, X_i, \sigma_u^2) \quad j = 0, 1, \dots, p$$

o mesmo acontece com a expressão (9), obtendo-se

$$U^*(\sigma_u^2; Y_i, W_i, \beta_j) = U(\sigma_u^2; Y_i, W_i, \beta_j) - \psi(\sigma_u^2; Y_i, W_i, \beta_j) \quad \dots (11)$$

onde as funções $\varphi(\beta_j, Y_i, W_i, \sigma_u^2)$ e $\psi(\sigma_u^2; Y_i, W_i, \beta_j)$ correspondem aos vícios do valor esperado condicional das funções escores para β_j e σ_u^2 , respectivamente. Organizando adequadamente as expressões (10) e (11), temos que

$$U^*(\beta_j; Y_i, W_i, \sigma_u^2) = \frac{1}{\sigma_u^2} \sum_{i=1}^p \frac{1}{\lambda_i} \left(Y_i t_{i,j} - \sum_{s=0}^p \beta_s t_{i,j+s} \right) \quad j = 0, 1, \dots, p \quad \dots (12)$$

$$U^*(\sigma_u^2; Y_i, W_i, \beta_j) = -\frac{1}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} \sum_{i=1}^p \frac{1}{\lambda_i} \left(Y_i^2 - 2Y_i \sum_{j=0}^p \beta_j t_{i,j} - \sum_{j=1}^p \sum_{s=0}^p \beta_j \beta_s t_{i,j+s} \right) \quad \dots (13)$$

onde $t_{i,(m)}$ é definido segundo o seguinte lema

Lema 2 Se u_i é distribuído normalmente com média zero e variância σ_u^2 , então

$$t_{i,(m+1)} = W_i t_{i,(m)} - m \sigma_u^2 t_{i,(m-1)}$$

para $m = 1, 2, \dots, p-1$, sendo que $t_{i,(0)} = 1$ e $t_{i,(m)} = W_i$ a sua vez $E(t_{i,(s)}) = X_i^s$.

A partir disto temos o seguinte sistema de equações lineares

$$\sum_{i=1}^n \frac{1}{\lambda_i} \begin{pmatrix} 1 & t_{i,(1)} & \dots & t_{i,(p)} \\ t_{i,(1)} & t_{i,(2)} & \dots & t_{i,(p+1)} \\ \vdots & \vdots & \ddots & \vdots \\ t_{i,(p)} & t_{i,(p+1)} & \dots & t_{i,(2p)} \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} = \begin{pmatrix} Y_i \\ Y_i t_{i,(1)} \\ \vdots \\ Y_i t_{i,(p)} \end{pmatrix} \quad \dots (14)$$

$$\hat{\sigma}_u^2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i} \left(Y_i^2 - 2Y_i \sum_{j=0}^p \hat{\beta}_j t_{i,(j)} - \sum_{j=0}^p \sum_{s=0}^p \hat{\beta}_j \hat{\beta}_s t_{i,(j+s)} \right) \dots (15)$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n \frac{Y_i}{\lambda_i} - \sum_{i=1}^n \frac{W_i}{\lambda_i} \hat{\beta}_1}{\sum_{i=1}^n \frac{1}{\lambda_i}} \dots (16)$$

Como se pode observar nas expressões (14) e (15), não se pode encontrar uma expressão fechada para estimar β_j e σ_u^2 simultaneamente, porém se faz necessário de um processo computacional para suas estimativas. O processo computacional considera os seguintes passos

- 1 Escolher apropriadamente o valor de σ_u^2 .
- 2 Atualizar a expressão (14) considerando o Lema 2, estimar os valores de β_j .
- 3 Substituir os valores de β_j e as atualizações feitas no Lema 2, na expressão (15).
- 4 Retornar ao passo 2, caso a convergência não foi obtida.

O sistema de equações apresentados em (14) e (15), tem raízes múltiplas. Portanto a escolha da raiz apropriada deve ser feita cuidadosamente, devido a que podem existir raízes que não são consistentes. Duas possíveis soluções para o problema de raízes múltiplas são propostas em Stefanski e Carroll (1987) e Stefanski (1989). A primeira recomenda a raiz mais próxima do estimador “naive” e com o mesmo sinal que este. Uma segunda solução é desenvolver o sistema em forma iterativa, considerando como valores iniciais os obtidos pelo estimador “naive”.

2.2 Exemplo

Consideremos a equação da linha reta

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

$$W_i = X_i + u_i \quad i = 1, 2, \dots, n$$

sendo $t_{i,(0)} = 1$, $t_{i,(1)} = W_i$ e considerando o resultado do Lema 2, $t_{i,(2)} = W_i^2 - \sigma_u^2$, o sistema de equações lineares está composta por

$$\sum_{i=1}^n \frac{1}{\lambda_i} \begin{pmatrix} 1 & W_i \\ W_i & W_i^2 - \sigma_u^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \sum_{i=1}^n \frac{1}{\lambda_i} \begin{pmatrix} Y_i \\ Y_i W_i \end{pmatrix}$$

$$\hat{\sigma}_u^2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i}$$

$$(Y_i^2 - 2Y_i \hat{\beta}_0 - 2Y_i \hat{\beta}_1 W_i + \hat{\beta}_0^2 + \hat{\beta}_1^2 (W_i^2 - \sigma_u^2) + 2\hat{\beta}_0 \hat{\beta}_1 W_i)$$

após uns desenvolvimentos algébricos temos:

$$S_{WY} - S_{WW} \hat{\beta}_1 + \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i} \hat{\sigma}_u^2 \hat{\beta}_1 = 0 \dots (17)$$

$$\hat{\sigma}_u^2 = \left(\frac{1}{1 + \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i} \hat{\beta}_1^2} \right) [S_{YY} - 2S_{YW} \hat{\beta}_1 + S_{WW} \hat{\beta}_1^2] \dots (18)$$

onde

$$S_{WW} = \sum_{i=1}^n \frac{W_i^2}{\lambda_i} - \frac{\left(\sum_{i=1}^n \frac{W_i}{\lambda_i} \right)^2}{\sum_{i=1}^n \frac{1}{\lambda_i}}$$

$$S_{WY} = \sum_{i=1}^n \frac{W_i Y_i}{\lambda_i} - \frac{\left(\sum_{i=1}^n \frac{W_i}{\lambda_i} \right) \left(\sum_{i=1}^n \frac{Y_i}{\lambda_i} \right)}{\sum_{i=1}^n \frac{1}{\lambda_i}}$$

$$S_{YY} = \sum_{i=1}^n \frac{Y_i^2}{\lambda_i} - \frac{\left(\sum_{i=1}^n \frac{Y_i}{\lambda_i} \right)^2}{\sum_{i=1}^n \frac{1}{\lambda_i}}$$

substituindo a expressão (18) na expressão (17), temos que:

$$\hat{\beta}_1 = \frac{\left(\frac{S_{YW}}{n} \sum_{i=1}^n \frac{1}{\lambda_i} - S_{WW} \right) \pm \sqrt{\left(\frac{S_{YW}}{n} \sum_{i=1}^n \frac{1}{\lambda_i} - S_{WW} \right)^2 - 4 \frac{S_{WY}^2}{n} \sum_{i=1}^n \frac{1}{\lambda_i}}}{2 \frac{S_{YW}}{n} \sum_{i=1}^n \frac{1}{\lambda_i}}$$

a raiz apropriada é obtida como aquele mais próxima do estimador “naive” que neste caso corresponde a

$$\hat{\beta}_1 = \frac{\left(\frac{S_{YW}}{n} \sum_{i=1}^n \frac{1}{\lambda_i} - S_{WW} \right) + \sqrt{\left(\frac{S_{YW}}{n} \sum_{i=1}^n \frac{1}{\lambda_i} - S_{WW} \right)^2 - 4 \frac{S_{WY}^2}{n} \sum_{i=1}^n \frac{1}{\lambda_i}}}{2 \frac{S_{YW}}{n} \sum_{i=1}^n \frac{1}{\lambda_i}}$$

O resultado obtido em (19) será substituído em (16) e (18), obtendo desta forma os estimadores de β_0 , β_1 e σ_u^2 .

No caso homogêneo, isto é, $\lambda = \lambda$ os resultados obtidos a partir das expressões acima seriam similares as obtidas por Gimenez e Bolfarine (1997).

O estimador dos β_j é bom para amostras grandes, mas é instável para amostras pequenas, Cheng et. al. (2000), propuseram um estimador assintoticamente equivalente para amostras pequenas, a partir de

$$\begin{vmatrix} \sum_{i=1}^n \frac{Y_i^2}{\lambda_i} & \mathbf{K}_t^T \\ \mathbf{K}_t & \mathbf{M}_t - \rho \mathbf{V}_t \end{vmatrix} = 0$$

onde ρ é uma raiz positiva pequena da

expressão acima, $t_i = (1, t_{i(1)}, \dots, t_{i(p)})^T$,

$$M_t = \sum_{i=1}^n \frac{t_i t_i^T}{\lambda_i}$$

$$K_t = \sum_{i=1}^n \frac{1}{\lambda_i} (Y_i, Y_i t_{i(1)}, \dots, Y_i t_{i(p)})$$

$$T_t = \sum_{i=1}^n T_i = \sum_{i=1}^n \frac{1}{\lambda_i} \begin{pmatrix} 1 & t_{i(1)} & \dots & t_{i(p)} \\ t_{i(1)} & t_{i(2)} & \dots & t_{i(p+1)} \\ \vdots & \vdots & \ddots & \vdots \\ t_{i(p)} & t_{i(p+1)} & \dots & t_{i(2p)} \end{pmatrix}$$

e $V_t = M_t - T_t$, com estes componentes pode-se obter o estimador não viciado para pequenas amostras, sendo esta:

$$(M_t - aV_t)\beta = K_t \quad \dots (20)$$

onde

$$a = \begin{cases} (n-\alpha)/n & \text{se } \rho > 1+1/n \\ \rho(n-\alpha)/(n-1) & \text{se } \rho \leq 1+1/n \end{cases}$$

com $\alpha = \rho + 2$, como o sugerido em Cheng et. al. (2000).

2.3 Matriz de Variância Assintóticas

O vetor do Escore Corrigido para cada observação é definido pela seguinte expressão:

$$U_i^*(\beta; Y_i, W_i) = \begin{pmatrix} \frac{1}{\lambda_i \sigma_u^2} \left[Y_i - \sum_{s=0}^p \beta_s t_{i(s)} \right] \\ \frac{1}{\lambda_i \sigma_u^2} \left[Y_i t_{i(1)} - \sum_{s=0}^p \beta_s t_{i(s+1)} \right] \\ \vdots \\ \frac{1}{\lambda_i \sigma_u^2} \left[Y_i t_{i(p)} - \sum_{s=0}^p \beta_s t_{i(s+p)} \right] \end{pmatrix} \quad \dots (21)$$

onde se cumpre que $E(U_i^*(\beta; Y_i, W_i)) = 0$. Pelo método do "Sandwich", temos que a matriz de covariâncias assintóticas pode ser definida pelo seguinte teorema:

Teorema 1 Seja $\hat{\beta}_n$ uma raiz consistente da equação $\bar{U}_n^*(\beta; Y, W) = 0$ e β_0 é o verdadeiro valor do parâmetro. Sob as condições de regularidade C1-C5 apresentada por Gimenez e Bolfarine, $\hat{\beta}_n$ é assintoticamente normal com vetor de médias β_0 e a matriz de covariâncias assintóticas é dada por $\hat{V} = n^{-1} \hat{\Omega}_n$ onde

$$\hat{\Omega}_n(\beta) = \{\bar{I}_n^*(\beta)\}^{-1} \{\bar{S}_n^*(\beta)\} \{\bar{I}_n^*(\beta)\}^{-1}$$

sendo que

$$\bar{I}_n^*(\beta) = \frac{1}{n} \sum_{i=1}^n E \left[-\frac{\partial}{\partial \beta} U_i^*(\beta; Y_i, W_i) \right]$$

$$\bar{S}_n^*(\beta) = \frac{1}{n} \sum_{i=1}^n E[U_i^*(\beta; Y_i, W_i) U_i^*(\beta; Y_i, W_i)^T]$$

Prova

A Demonstração deste teorema esta explicado no trabalho de Gimenez e Bolfarine (1997), no Teorema 2.2.

Outra forma de apresentar a $\hat{\Omega}_n(\beta)$, é considerando a matriz T_t na expressão acima

$$\hat{\Omega}(\beta) = n^2 \sigma_u^2 T_t^{-1} \bar{S}^* T_t^{-1}$$

2.4 Teste de Hipótese

Considerando o método de escore corrigido em modelos polinomiais apresentado neste trabalho e considerando a matriz de covariância assintótica apresentada na equação (18), podemos desejar testar a hipóteses seguintes:

$$H_0 : \beta = \beta^0 \quad \text{vs} \quad H_1 : \beta \neq \beta^0$$

onde β^0 é o vetor de constantes conhecidas.

Gimenez et.al.(2000), propus os testes de Escore e Wald baseados na metodologia de escore corrigida.

O teste de Wald, então é definido por:

$$W_t = (\hat{\beta} - \beta^0)^T (\hat{\Omega}(\hat{\beta}))^{-1} (\hat{\beta} - \beta^0)$$

onde $\hat{\Omega}(\hat{\beta})$ é o estimador da matriz de covariâncias assintótica.

O teste de Escore, também conhecido como teste de Rao, é definido quando $U^*(\hat{\beta}) = 0$, é definida pela seguinte expressão

$$R_t = U(\beta^0)^T (\hat{\Omega}(\hat{\beta}))^{-1} U(\beta^0)$$

Segundo Gimenez et.al.(1997), as estatísticas de Wald e Escore têm a mesma distribuição assintótica sob a hipótese nula, isto é,

$$W_t = R_t \xrightarrow{D} X_{(p+1)}^2$$

Para uma hipótese individual, isto é, testar os parâmetros estruturais do modelo polinomial

$$H_0 : \beta_j - \beta_j^0 \quad \text{vs} \quad H_1 : \beta_j \neq \beta_j^0 \quad j = 0, 1, 2, \dots, p$$

o teste a ser considerado é definido por:

$$Z_j = \frac{\hat{\beta}_j - \beta_j^0}{\hat{\Omega}_{jj}^{1/2}} \quad j = 0, 1, 2, \dots, p$$

onde $\hat{\Omega}_{jj}^{1/2}$ corresponde ao j-ésimo componente da diagonal principal da matriz de covariância assintótica apresentada em (18).

3 Um estudo de Simulação

Para o presente estudo de simulação foi considerado um vetor de $\beta = (\beta_0, \beta_1, \beta_2)^T$ todos iguais a um, assumimos que $u_i \sim U(0; \sigma_u^2)$, com a finalidade de não ter muita perturbação consideramos que $\lambda_i \sim U(0,90;1,10)$, assim que $e_i \sim N(0; \lambda_i \sigma_u^2)$, além disso $X_i \sim \text{Bino}(1000;0,009)$, de modo que $W_i = X_i + u_i$, sendo nosso modelo $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i$. Para a simulação foi considerada 1.000 iterações para amostras de 10, 30, 50, 100 e 500.

Nos resultados, apresentaremos as estimativas $\hat{\beta}_j$, os desvios padrões (D.P.), os Erros Quadráticos Médios (EQM).

Na tabela 1 (abaixo), é o resultado de um programa feito no Linguagem Ox (Doornik (2001)), para a construção da tabela foi considerado os valores seguintes $\beta_0 = \beta_1 = \beta_2 = 1$, a partir deste, observa-se que os estimadores apresentados na tabela se aproximam aos valores sugeridos na simulação, por outro lado observa-se que a medida que se acrescenta o valor de σ_u^2 a variabilidade de cada parâmetro aumenta. Segundo esta simulação os parâmetros se aproximam aos valores sugeridos se $\sigma_u^2 \leq 0,07$, após este valor os valores de $\hat{\beta}$'s se vem influenciados pelo valor de s_u^2 , afastando-se dos valores sugeridos.

Tabela 1 – Resultados da Simulação para um problema de erros nas variáveis no modelo quadrático quando, $\lambda_i = \sigma_{e_i}^2 / \sigma_u^2$ é heterogênea.

σ_u^2	Estimativas			Desvio Padrão			Erro Quadrático Médio		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
Tamanho da amostra -> n=10									
0,01	1,02140	0,99659	1,00010	0,0000770	0,0002460	0,0000269	0,3992200	0,0152950	0,0000319
0,02	1,04260	0,99329	1,00020	0,0006682	0,0009576	0,0001094	1,3469000	0,0784080	0,0002598
0,03	1,06360	0,99012	1,00030	0,0018685	0,0023746	0,0002739	2,4409000	0,2385800	0,0011450
0,04	1,08600	0,98705	1,00040	0,0051493	0,0063084	0,0007302	2,4668000	0,7956300	0,0060333
Tamanho da amostra -> n=30									
0,01	1,00330	1,00030	0,99993	0,0005382	0,0000689	0,0000012	0,0020511	0,0002140	0,0000047
0,02	1,00630	1,00070	0,99985	0,0017370	0,0001961	0,0000010	0,0012331	0,0017861	0,0000181
0,03	1,00900	1,00110	0,99977	0,0036248	0,0003800	0,0000010	0,0596780	0,0068547	0,0000378
0,04	1,01150	1,00160	0,99969	0,0062864	0,0006236	0,0000055	0,3568500	0,0187160	0,0000602
0,05	1,01390	1,00230	0,99960	0,0099115	0,0009377	0,0000135	1,2685000	0,0425780	0,0000795
0,06	1,01620	1,00300	0,99951	0,0149550	0,0013524	0,0000269	3,6408000	0,0882590	0,0000867
0,07	1,01880	1,00380	0,99942	0,0230250	0,0019852	0,0000514	10,3410000	0,1856000	0,0000642
Tamanho da amostra -> n=50									
0,01	0,99986	1,00080	0,99994	0,0003182	0,0000351	0,0000000	0,0040839	0,0010225	0,0000029
0,02	0,99999	1,00170	0,99989	0,0011796	0,0001220	0,0000008	0,0573150	0,0056073	0,0000104
0,03	1,00040	1,00250	0,99983	0,0026339	0,0002657	0,0000023	0,2892900	0,0168660	0,0000205
0,04	1,00120	1,00330	0,99978	0,0047810	0,0004761	0,0000047	0,9607200	0,0396520	0,0000314
0,05	1,00240	1,00400	0,99972	0,0078183	0,0007725	0,0000082	2,5853000	0,0823210	0,0000409
0,06	1,00420	1,00480	0,99967	0,0121900	0,0011979	0,0000133	6,3238000	0,1622300	0,0000458
0,07	1,00730	1,00550	0,99962	0,0194300	0,0019006	0,0000219	16,1930000	0,3358100	0,0000396
Tamanho da amostra -> n=100									
0,01	0,99558	1,00080	0,99998	0,0001497	0,0000030	0,0000021	0,0352980	0,0010225	0,0000029
0,02	0,99136	1,00160	0,99995	0,0005974	0,0000180	0,0000090	0,2210700	0,0056073	0,0000104
0,03	0,98738	1,00250	0,99993	0,0013921	0,0000463	0,0000213	0,7435800	0,0168660	0,0000205
0,04	0,98375	1,00330	0,99990	0,0026473	0,0000913	0,0000409	1,9585000	0,0396520	0,0000314
0,05	0,98074	1,00420	0,99988	0,0046766	0,0001642	0,0000726	4,7482000	0,0823210	0,0000409
Tamanho da amostra -> n=500									
0,01	1,00410	0,99892	1,00010	0,0000459	0,0000009	0,0000004	0,0096891	0,0011256	0,0000053
0,02	1,00850	0,99783	1,00010	0,0001826	0,0000032	0,0000017	0,0193510	0,0044035	0,0000273
0,03	1,01310	0,99673	1,00020	0,0004169	0,0000068	0,0000040	0,0138920	0,0096867	0,0000781
0,04	1,01800	0,99561	1,00030	0,0007621	0,0000120	0,0000073	0,0007925	0,0168130	0,0001754
0,05	1,02320	0,99449	1,00030	0,0012441	0,0000188	0,0000120	0,0242450	0,0255660	0,0003468
0,06	1,02890	0,99335	1,00040	0,0019160	0,0000278	0,0000185	0,2094100	0,0355810	0,0006430
0,07	1,03530	0,99220	1,00050	0,0029188	0,0000407	0,0000284	0,9485900	0,0460000	0,0011856

4 Aplicação

Para poder avaliar a consistência da metodologia proposta aplicou a teoria a um conjunto de dados reais, que se encontram no Exemplo 4 no trabalho de Galea-Rojas et.al. (2003), os dados correspondem ao conteúdo de Ouro (em gramas/ton.) para 501 amostras as quais foram resultado de

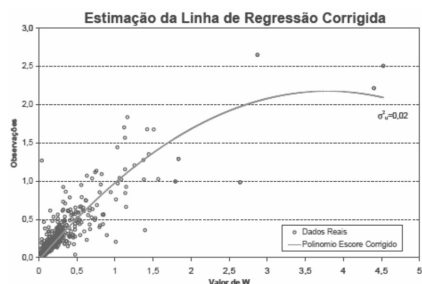
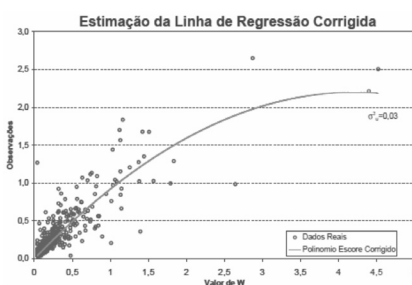
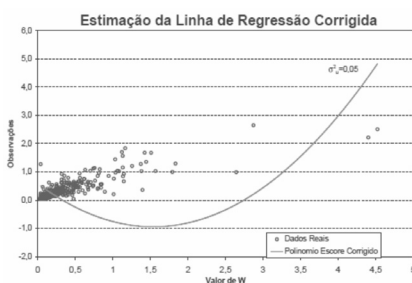
a análise de duas metodologias o método clássico e o método fogo dividido.

Os resultados são apresentados na seguinte tabela

Tabela 2 – Resultados dos dados do exemplo 4 no Galea-Rojas et.al. (2003), quando $\lambda_i = \sigma_{e_i}^2 / \sigma_u^2$ heterogênea.

		β_0	β_1	β_2
$\sigma_u^2 = 0,01$	Estimativa	-0,0986970	1,4092000	-0,2135900
	Desvio Padrão	0,0053739	0,0245750	0,0060950
$\sigma_u^2 = 0,02$	Estimativa	-0,0350320	1,1669000	-0,1539600
	Desvio Padrão	0,0002806	0,0037988	0,0009262
$\sigma_u^2 = 0,03$	Estimativa	-0,0012376	1,0453000	-0,1241000
	Desvio Padrão	0,0020642	0,0062340	0,0015537
$\sigma_u^2 = 0,04$	Estimativa	-0,4278100	3,0222000	-0,6157100
	Desvio Padrão	0,0133090	0,0546390	0,0133530
$\sigma_u^2 = 0,05$	Estimativa	0,5855900	-1,9903000	0,6473300
	Desvio Padrão	0,0089673	0,0464750	0,0117940
$\sigma_u^2 = 0,06$	Estimativa	0,1156100	0,5723500	-0,0087483
	Desvio Padrão	0,0021400	0,0108310	0,0018394

Como no processo de simulação anterior concluímos que as melhores estimativas de um modelo de regressão, se encontram quando $\sigma_u^2 \leq 0,07$ consideramos no modelo vários valores iniciais de $\sigma_u^2 = (0,01; 0,02; 0,03; 0,04; 0,05; 0,06)$ nos dados reais, uma interpretação gráfica da tabela acima pode-se apresentar a seguir:

(a) $\sigma_u^2 = 0,01$ (b) $\sigma_u^2 = 0,02$ (c) $\sigma_u^2 = 0,03$ (d) $\sigma_u^2 = 0,04$ (e) $\sigma_u^2 = 0,05$ (f) $\sigma_u^2 = 0,06$

Observamos nos gráficos apresentados acima que as curvas do modelo polinomial de grau 2 que melhor se ajustam aos dados estariam definidos quando σ_u^2 é igual a 0,01, 0,02, 0,03 e 0,06, os valores de σ_u^2 igual a 0,04 e 0,05 se ajustam também, só que seu ajuste não passa pela maioria da nuvens de pontos da amostra, o que faz pouco eficiente para o modelo.

5 Conclusões

O presente trabalho pretendeu abordar um problema de falta de estimação da função de verossimilhança em modelos com erros nas variáveis, a obtenção de uma solução considerando uma função escore corrigida num modelo de regressão funcional e polinomial, observou-se que frente a este problema a metodologia apresentada e sua aplicação não são difíceis de realizar, a sua vez esta metodologia apresenta estimadores consistentes e de mínima variância, a

contribuição de este trabalho é que em virtude a falta de conhecimento da variância do erro da medida σ_u^2 , e ao conhecimento da relação entre a variabilidade do erro σ_u^2 do modelo e da variabilidade do erro σ_u^2 da variável independente (W), a dificuldade principal é estimar a variável latente σ_u^2 , muitos autores foram dando diversas alternativas, a maioria solicita informação adicional da variabilidade, neste estudo também se fez um estudo do valor de σ_u^2 adequado para dar início as estimativas do modelo.

6 Referências

- CARROLL, R. J.; RUPPERT, D. and STEFANSKI, L. A. **Measurement Error in Nonlinear Models**. Chapman & Hall, New York. 1995.
- CHAN, L. K. and MAK, T. K. **On the polynomial functional relationship**. J. Roy. Statist. Soc. Ser. B, 47, p. 510-518. 1985.
- CHENG, C.-L. and SCHNEEWEISS, H. **Polynomial regression with errors in the variables**. J. Roy. Statist. Soc. Ser. B, 60, p. 189-199. 1998.
- CHENG, C.-L., SCHNEEWEISS, H. and THAMERUS, M. **A small sample estimator for a polynomial regression with errors in the variables**. J. Roy. Statist. Soc. Ser. B, 62, p. 699-709. 2000.
- CHENG, C.-L. and VAN NESS, J. W. **Statistical Regression with Measurement Error**. Arnold, London, 1999.
- DOORNIK, J. A. **Ox: An Object-Oriented Matrix Language**. fourth edition, Timberlake Consultants Press, London. 2001.
- FULLER, W. A. **Measurement Error Models**. Wiley, New York. 1987.
- GALEA-ROJAS, M.; DE CASTILHO, M. V.; BOLFARINE, H. and DE CASTRO, M. **Detection of analytical bias**. Analyst, 128, 1073-1081. 2003.
- GIMENEZ, P. and BOLFARINE, H. **Corrected score functions in classical error-in-variables and incidental parameter models**. Austral. J. Statist., 39, p. 325-344. 1997.
- GIMENEZ, P.; BOLFARINE, H. and COLOSIMO, E. A. **Hypotheses testing for error-in-variables models**. Ann. Inst. Statist. Math., 52, p. 698-711. 2000.
- KUKUSH, A.; SCHNEEWEISS, H. and WOLF, R. **Relative efficiency of three estimators in a polynomial regression with measurement errors**. J. Statist. Plann. Infer., 127, p. 179-203. 2005.
- MOON, M.-S. and GUNST, R. F. **Polynomial measurement error modeling**. Comp. Statist. & Data Anal., 19, p. 1-21. 1993.
- NAKAMURA, T. **Corrected score function of errors-in-variables models: methodology and applications to generalized linear models**. Biometrika, 77, p. 127-137. 1990.
- STEFANSKI, L. A. **Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models**. Comm. Statist. Theory Methods, 18, p. 4335-4358. 1989.
- SCHNEEWEISS, H. and NITTNER, T. **Estimating a polynomial regression with measurement errors in the structural and in the functional case A comparison**. Data Analysis from Statistical Foundations, (ed. A. K. M. E. Saleh), p. 195-205, Huntington, Nova Science, 2001.