



ESTIMAÇÃO E DIAGNÓSTICO EM MODELOS DE

REGRESSÃO SIMPLES COM ERROS SLASH

Taísa A. Angolini¹, Víctor H. Lachos²

taisa.angolini@gmail.com¹, hlachos@ime.unicamp.br²



IMECC – UNICAMP

PIBIC/CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

Palavras-chave: Algoritmo EM - Influência Local - Distribuição Slash

Introdução

Por mais rotineiro que seja assumir a normalidade dos dados, nem sempre esta suposição é realista, assim, é conveniente considerar famílias paramétricas de distribuições com comportamentos simétricos, produzindo uma estimação robusta no modelo considerado. Este trabalho tem como base a análise de influência local para modelos com erros de medição quando o erro aleatório e o valor observado da covariável seguem conjuntamente uma distribuição slash multivariada, permitindo modelar dados com certa proporção de *outliers*. Foi considerado um método alternativo para a obtenção de influência local, o algoritmo EM, baseado na log-verossimilhança aumentada para a obtenção das estimativas dos parâmetros.

Metodologia

Considera-se o modelo de regressão com erro de medida

$$X_i = x_i + u_i$$

$$Y_i = \alpha + \beta x_i + e_i,$$

onde tanto os erros aleatórios (e_i e u_i) como a variável latente x (não observável) seguem uma distribuição slash. Para isso, a distribuição slash é construída a partir de um modelo representado por um vetor aleatório normal \mathbf{Z} com vetor de médias $\mathbf{0}$, matriz de variância – covariância Σ e uma v.a. de mistura $\kappa^{1/2}(U)$ dado por

$$\mathbf{S} = \boldsymbol{\mu} + \kappa^{1/2}(U)\mathbf{Z},$$

onde $\boldsymbol{\mu}$ é um vetor de locação e U segue uma distribuição Beta($v,1$), em que v é um parâmetro adicional que regula o comportamento caudal da distribuição.

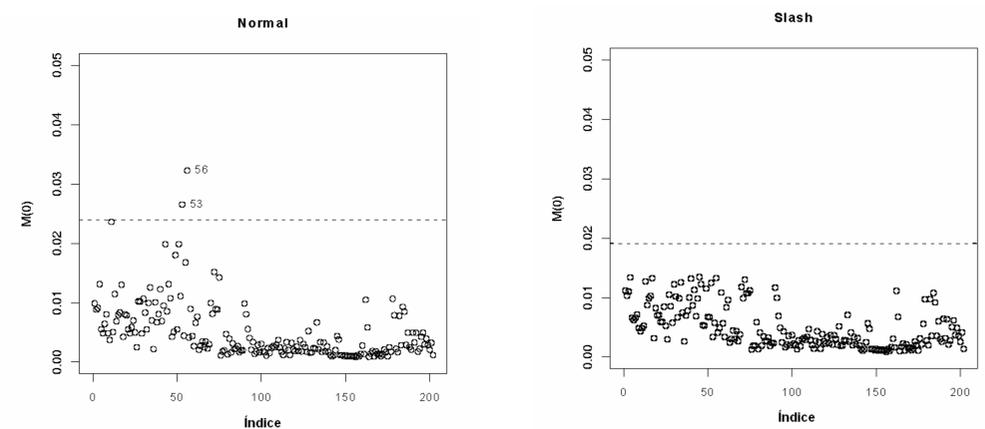
Para a estimação dos parâmetros usa-se o algoritmo EM, que divide-se em duas partes: cálculo da Q-function (passo E) e sua maximização com respeito aos parâmetros da regressão (passo M).

A fim de detectar observações que, sob menores perturbações do modelo, exercem grande influência sobre as estimativas dos parâmetros, utilizou-se a abordagem de influência local de Zhu e Lee (2001) e alguns tipos de perturbação.

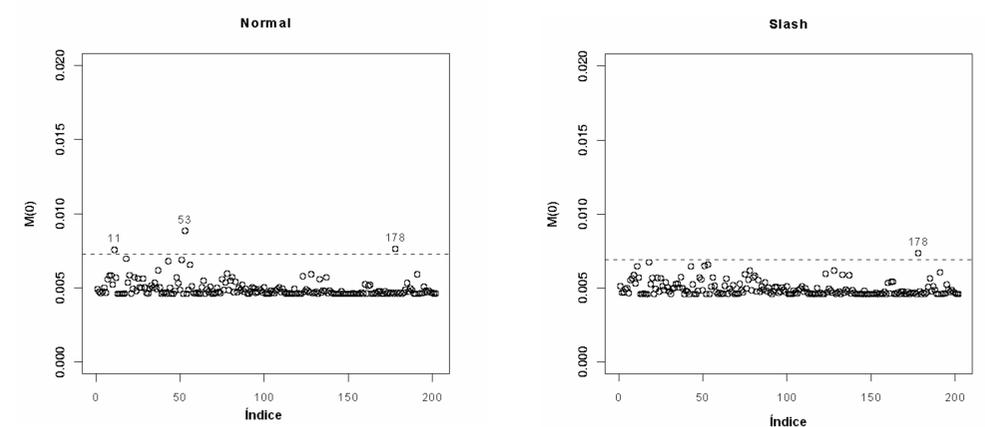
Resultados

Para a ilustração da teoria utilizou-se uma parte do conjunto de dados AIS (disponível no arquivo de dados do programa R (pacote SN)), e foi possível verificar a redução de pontos influentes quando utilizada a distribuição slash se comparada com a normal.

Perturbação da variável resposta: usada para detectar observações com grande influência nos seus próprios valores preditos.



Perturbação na razão de variância: usada para detectar observações com grande influência na suposição de homocedasticidade.



Conclusão

Foi possível reduzir a influência de pontos discrepantes utilizando uma distribuição mais robusta que a normal, a distribuição slash, através de diferentes técnicas de perturbação no modelo.