UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

Application of robust residuals in matrices of consistent covariance

Natália Silva Lourenço

João Pessoa - PB
2019

Natália Silva Lourenço

Application of robust residuals in matrices of consistent covariance

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Estatística da Universidade Federal da Paraíba, como requisito para obtenção do grau de Bacharel em Estatística.

Orientador: Prof[a]. Dr[a]. Tatiene Correia de Souza

João Pessoa - PB
2019

Catalogação na publicação
Seção de Catalogação e Classificação

L892a Lourenço, Natalia Silva.
Application of robust residuals in matrices of consistent covariance / Natalia Silva Lourenço. - João Pessoa, 2019.
52 f. : il.

Orientação: Tatiene Correia de Souza.
Monografia (Graduação) - UFPB/CCEN.

1. Heterokesdasticity. 2. Regression models. 3. Robust residuals. I. de Souza, Tatiene Correia. II. Título.

UFPB/CCEN

UNIVERSIDADE FEDERAL
DA PARAÍBA

COORDENAÇÃO DE ESTATÍSTICA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
CAMPUS I – Cidade Universitária
Telefone: (83) 3216-7018/7785
Cep: 58051-900 – João Pessoa, PB, Brasil

ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

"Application of robust residual in matrices official consistent covariance"

Natália Silva Lourenço

No décimo terceiro dia do mês de maio de 2019, ás 09:00h na sala 19 do Departamento de Estatística, a Banca Examinadora do Trabalho de Conclusão de Curso do(a) aluno(a) Natália Silva Lourenço, mat. 11502277, foi composta pelos professores: Tatiene Correia de Souza, Presidente/Orientador(a) (Departamento de Estatística - UFPB), Luiz Medeiros de Araújo Lima Filho (Departamento de Estatística - UFPB) e Tarciana Liberal Pereira (Departamento de Estatística - UFPB). Dando início aos trabalhos, o presidente da banca cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra ao candidato para que se fizesse, oralmente, a exposição do trabalho de conclusão de curso intitulado **"Application of robust residual in matrices official consistent covariance"**. Concluída a apresentação, o(a) candidato(a) foi arguido(a) pela Banca Examinadora que sugeriu que o(a) aluno(a) fizesse algumas alterações até o dia 17 de maio de 2019. Uma vez entregue a versão final do Trabalho de Conclusão de Curso à Coordenação do Bacharelado em Estatística com as alterações solicitadas pela banca examinadora dentro do prazo que o aluno recebeu, o(a) aluno(a) será aprovado com a nota 9.5 (Nove inteiros e Cinco decimos), que é a média aritmética das notas atribuídas pelos membros da Banca Examinadora.

Tatiene C. Souza
Professor(a) Orientador(a) Tatiene Correia de Souza

Luiz Araujo Filho
Professor(a) Luiz Medeiros de Araújo Lima Filho

Tarciana Liberal Pereira
Professor(a) Tarciana Liberal Pereira

Natália Silva Lourenço
Aluno(a) Natália Silva Lourenço

João Pessoa, 13 de maio de 2019.

Natália Silva Lourenço

Application of robust residuals in matrices of consistent covariance

Trabalho de Conclusão de Curso apresentado e aprovado na data 13/05/2019 ao Curso de Bacharelado em Estatística da Universidade Federal da Paraíba, como requisito para obtenção do grau de Bacharel em Estatística.

Orientador: Prof^a. Dr^a. Tatiene Correia de Souza

Banca Examinadora

Tatiene C Souza

Prof^a. Dr^a. Tatiene Correia de Souza
Orientador - UFPB

Tarciana L Oliveira Pereira

Prof^a. Dr^a. Tarciana Liberal Pereira
Examinador - UFPB

Luiz Araujo Filho

Prof. Dr. Luiz Medeiros de Araujo Lima Filho
Examinador - UFPB

Dedico este trabalho aos meus pais e irmão,

Leonicio, Sandra e Rafael Lourenço.

Agradecimentos

Agradeço primeiramente à Deus, que me deu o dom da vida, pelo seu imenso amor e misericórdia, pelo seu carinho e cuidado comigo e minha família, pela força que nos tem dado para atravessar os obstáculos e continuar crescendo.

Agradeço aos meus pais por tudo. Pelo amor, educação, confiança e carinho dedicados incondicionalmente à mim ao longo da minha vida. Vocês são o maior exemplo de força, coragem e caráter que segui durante todo esse tempo. Obrigada por estar comigo em todos os momentos e por tantas vezes ter me motivado a seguir em frente.

Ao meu irmão por toda amizade, companheirismo e compreensão. Obrigada por estar tão perto de mim.

A Profª. Tatiene pela orientação, ensinamentos, conselhos, paciência e disponibilidade, sendo ela, uma das principais responsáveis pela minha formação acadêmica e profissional e pela realização desse trabalho de conclusão de curso. A Profª. Tarciana que esteve comigo durante essa caminhada e pelo auxílio com suas orientações prestadas, me incentivando e colaborando no meu desenvolvimento. Agradeço de coração a vocês duas, que talvez nem imagine o quanto me ajudou, principalmente neste percurso final. Muito obrigada por me orientar, me incentivar e dar forças para continuar.

Ao Prof. Hemílio, que me aceitou como orientanda no PIBIC, obrigada pela sua confiança, orientação, disponibilidade, incentivo e apoio durante o período da iniciação científica e ao longo da minha guaduação.

À todos os professores do Departamento de Estatística, pela importância durante a minha formação acadêmica e por engrandecer meus conhecimentos.

Aos participantes da banca examinadora, aos professores Tarciana e Luiz, por compartilhar deste momento tão especial.

Agradeço aos querido amigos e colegas com quem convivi durante a graduação, em especial, Arthur, Mateus, Fernanda e Giovani, pela amizade, companheirismo, descontração, incentivo e paciência, vocês sempre tornam meus dias mais alegres.

Por todo o aprendizado, amizade, carinho e troca de experiências, meu muito obrigada.

”O mundo está nas mãos daqueles que tem a coragem de sonhar e de correr o risco de viver seus sonhos.”

Paulo Coelho.

RESUMO - Português

A estimação dos parâmetros, pelo modelo de regressão linear, é feita pelo método dos mínimos quadrados ordinários (OLS). Este método fornece estimativas não-viesadas, consistentes e eficientes. No entanto, sob heterocedasticidade, os estimadores OLS tornam-se ineficientes e o estimador comum de sua matriz de covariância não é consistente. Foram propostos estimadores de matrizes de covariância consistentes sob heterocedasticidade (HCCME) para resolver o problema heteroscedástico na regressão linear. Neste trabalho, quatro tipos de resíduos robustos foram aplicados às matrizes HC3, HC4, HC4m e HC5 para avaliar seus desempenhos. Também apresentamos uma aplicação empírica que usa dados reais.

Palavras-chave: Heterocedasticidade, Modelos de regressão, Resíduos robustos.

RESUMO - Inglês

The estimation of the parameters, by the linear regression model, is made by the method of ordinary least squares (OLS). This method provides estimates unbiased, consistent and efficient. However, under heteroscedasticity, the OLS estimators become inefficient and the common estimator of their covariance matrix is not consistent. Heteroskedasticity consistent covariance matrix estimator (HCCME) were proposed to solve the heteroscedastic problem in linear regression. In this work, four types of robust residuals were applied in the HC3, HC4, HC4m and HC5 estimators to evaluate their performances. We also present an empirical application that uses real data.

Key words: Heterokesdasticity, Regression models, Robust residuals.

Lista de Abreviações e Siglas

OLS = Ordinary Least Squares
LMS = Least Median of Squares
LTS = Least Trimmed Squares
LQS = Least Quantile of Squares
S = S-Estimator

Sumário

1 Introdução

O modelo de regressão linear é uma das ferramentas estatísticas mais populares para analisar a relação entre variáveis na pesquisa científica. Uma suposição comum é que a variância dos erros deve ser constante para todas as observações, conhecidas como homocedasticidade. Na prática, essa suposição pode ser frequentemente violada e erros com variâncias não constantes são conhecidos como erros heteroscedásticos. Sob heterocedasticidade, o estimador de mínimos quadrados ordinários (OLS) do vetor de parâmetros de regressão é ineficiente e sua matriz de covariância não é consistente. Assim, os resultados do teste de hipóteses usando este estimador podem ser questionáveis. Diversos autores propuseram estimadores consistentes para a matriz de covariância sob heterocedasticidade (HCCME) baseada em resíduos de mínimos quadrados ordinários. O mais conhecido HCCME é o proposto White (1980). Estes testes baseados em HCCME, são conhecidos como testes quase-t, geralmente tendem a ser substancialmente enviesados em amostras finitas, e, portanto, testes quase-t associados tendem a ser liberais.

A classe de HCCMEs também tem suas limitações. O teste baseado em HC3 produz um comportamento pobre de amostra finita quando os dados contêm pontos de alta alavancagem. O estimador HC4 leva em conta o impacto de pontos de alta alavancagem no comportamento da amostra finita do estimador da matriz de covariância. Quando os dados são altamente aproveitados, o estimador de teste baseado em HC5, proposto por Cribari–Neto e Souza (2012), é mais confiável do que os testes baseados em HC3 e HC4. Ao contrário dos estimadores concorrentes, o estimador HC5 leva em conta não apenas os graus individuais de alavancagem, mas também a alavancagem máxima. Cribari–Neto e Zarkos (2001) sugerem que a presença de pontos de alta alavancagem nos dados é mais decisiva para os comportamentos de amostras finitas dos diferentes testes baseados em HCCMEs do que o grau de heteroscedasticidade em si. Frequentemente os testes tendem a ser superdimensionados em dados de alta alavancagem, levando a inferências não confiáveis.

Como o método dos mínimos quadrados ordinários é afetado quando no conjunto de dados há observações que se destacam dos demais, dados atípicos, e essas observações afetam o ajuste do modelo, influenciando as estimativas dos coeficientes de regressão e, consequentemente, as demais quantidades que são funções dessas estimativas o principal objetivo deste trabalho é apresentar alguns métodos alternativos de detecção de observações por métodos de estimação que forneçam estimadores resistentes a tais observações.

APPLICATION OF ROBUST RESIDUALS IN MATRICES OF CONSISTENT COVARIANCE

NATÁLIA S. LOURENÇO

ABSTRACT. The estimation of the parameters, by the linear regression model, is made by the method of ordinary least sqares (OLS). This method provides estimates unbiased, consistent and efficient. However, under heteroscedasticity, the OLS estimators become inefficient and the common estimator of their covariance matrix is not consistent. Heteroskedasticity consistent covariance matrix estimator (HCCME) were proposed to solve the heteroscedastic problem in linear regression. In this work, four types of robust residuals were applied in the HC3, HC4, HC4m and HC5 estimators to evaluate their performances. We also present an empirical application that uses real data.

1. INTRODUCTION

The linear regression model is one of the most popular statistical tools for analyzing the relationship between variables in scientific research. A common held assumption is that the variance of errors must be constant for all observations, known as homoskedasticity. In practice, this assumption can be frequently violated and errors with of non-constant variances are known as heteroskedastic errors. Under heteroscedasticity, the ordinary least squares estimator (OLS) of the regression parameter vector is inefficient and its covariance matrix is not consistent. Thus, the results of hypothesis testing using this estimator can be questionables. Several authors have proposed consistent estimators for the covariance matrix under heterocedasticity (HCCME) based on ordinary least squares residuals. The best known HCCME is the one proposed [19]. These HCCME-based tests, are known as quasi-t tests, generally tend to be substantially biased in finite samples, and therefore associated quasi-t tests tend to be liberal.

The class of HCCMEs also have their limitations. The test based on HC3 yields a poor finite-sample behavior when the data contain high leverage points. The HC4 estimator takes into account the impact of high leverage points on the finite-sample behaviour of the covariance matrix estimator. When the data are high leveraged, the HC5-based test estimator, proposed by [3], is more reliable than HC3-based and HC4-based tests. [2] reports that HC4m-based inference from tests is more reliable than that based on standard HC4 errors under normal and non-normal random errors. Unlike the competing estimators, the HC5 estimator takes into account not only the individual degrees of leverage but also the maximal leverage. [7] suggest that the presence of points of high leverage in the data is more decisive for the finite sample behaviors of the different HCCMEs-based tests than the degree of heteroskedasticity itself. Frequently tests tend to be oversized in high leverage data, leading to unreliable inferences.

As the ordinary least squares method is affected when in the dataset there are observations that stand out from the others, atypical data, and these observations affect

Key words and phrases. Heterokesdasticity, Regressions models, Robust residuals.

the adjustment of the model, influencing the estimates of the regression coefficients and, consequently, the other quantities that are functions of these estimates the main. The objective of this work is four types of robust residuals were applied in the HC3, HC4, HC4m and HC5 estimators to evaluate their performances.

Our paper is organized as follows. In the Section 2, we describe the linear regression model and some HCCMEs. In the Section 3, we explain how HCCMEs can be constructed using robust residuals. The results of the simulation are presented and discussed in the Section 4 and the results of the application in the Section 5. Finally, the Section 6 concludes the article.

2. The model and estimators

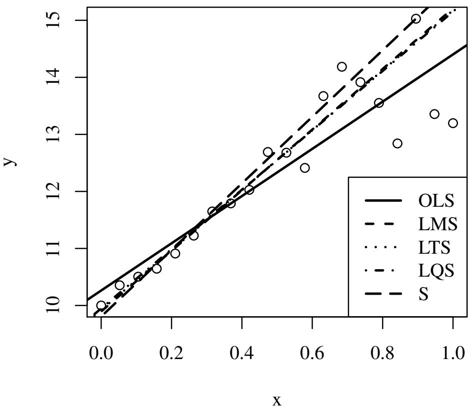
The model of interest is the linear regression model, i.e., $y = X\beta + e$, where y is an $n \times 1$ vector of responses, X is an $n \times p$ (full column rank) matrix of independent variables ($p < n$), assumed fixed, β is an p-vector of unknown regression parameters, and e is an n-vector of random errors. The errors under heteroscedasticity, are such that $\mathbb{E}(e_i) = 0$ and $\text{var}(e_i) = \sigma_i^2$ $(0 < \sigma_i^2 < \infty)$, for $i = 1, \ldots, n$; also, $\mathbb{E}(e_i e_j) = 0$ $\forall i \neq j$. The covariance matrix of e is $\Psi = \text{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$.

The OLS estimator of β can be expressed in closed-form as $b = (X'X)^{-1}X'y$. It is unbiased (i.e., $\mathbb{E}(b) = \beta$ for all β) and has covariance structure given by $\text{cov}(b) = (X'X)^{-1}X'\Psi X(X'X)^{-1}$. Under homoskedasticity (i.e., common error variances), $\sigma_i^2 = \sigma^2 > 0$ and hence $\Psi = \sigma^2 I_n$, where I_n is the $n \times n$ identity matrix. The covariance of b thus becomes $\text{cov}(b) = \sigma^2(X'X)^{-1}$, which can be easily estimated by $\widehat{\sigma}^2(X'X)^{-1}$, where $\widehat{\sigma}^2 = \widehat{e}'\widehat{e}/(n-p)$, $\widehat{e} = (I_n - H)y$ being the n-vector of OLS residuals and $H = X(X'X)^{-1}X'$ is a symmetric and idempotent matrix.

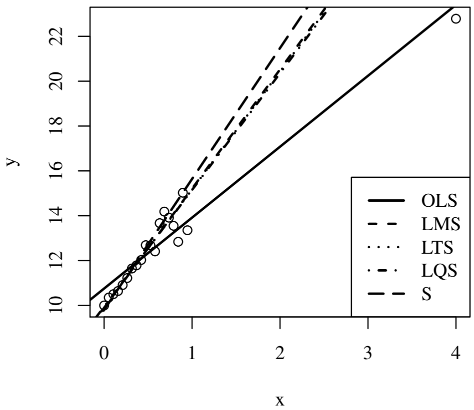
Under heteroskedasticity, however, the usual covariance matrix estimator $\widehat{\sigma}^2(X'X)^{-1}$ is biased and inconsistent for $\text{cov}(b)$. The estimation of the consistent covariance matrix takes into account the error dispersion. The most well-known HCCME was proposed by [19] and is usually referred to as HC0. It is obtained by simply replacing σ_i^2 by $\widehat{e}_i^2$ in the expression for the covariance structure of b, that is, $\text{HC0} = (X'X)^{-1}X'\widehat{\Psi}_0 X(X'X)^{-1}$, where $\widehat{\Psi}_0 = \text{diag}\{\widehat{e}_1^2, \ldots, \widehat{e}_n^2\}$. White's estimator is consistent under both homoskedasticity and heteroskedasticity of unknown form. A shortcoming of the HC0 estimator is that it is typically biased in finite samples. The simulation results in [10] suggest that HC0 estimates should not be used when $n \leq 250$. [6] and [11] have also shown that White's estimator can be severely biased in small to moderately large samples. In particular, the HC0 fails to provide reliable estimates of variance when the regressors include high leverage points.

The either disadvantage of the HC0 estimator is that OLS residuals tend to underestimate the true errors [8]. This can make HC0 an biased estimator when the sample size is small. To illustrate this fact, Figure 1 presents the adjustments of the regression model $y_i = \beta_0 + \beta_1 x_i + e_i, i = 1, ..., 20$, obtained through Ordinary Least of Squares estimator (OLS), Least Median of Squares estimator (LMS), Least Trimmed Squares estimator (LTS), Least Quantile of Squares estimator (LQS) and S-Estimator (S). Two scenarios were considered, one without points (Scenario 1) and one with high leverage points (Scenario 2). The values of the covariant x correspond to equally spaced values between 0 and 1, when we have the Scenario 1. In the second case, Scenario 2, we replace the last observation of the covariable by 4 in order to introduce a high point in the data, ie to the last diagonal element of the matrix $H = X(X'X)^{-1}X'$ exceeds $3p/n = 0.30$, the threshold value commonly used to identify high leverage points. In both scenarios we take $\beta_0 = 10$ and $\beta_1 = 5$. We adjusted the model considering the OLS, LMS, LTS, LQS and S estimators, as shown in Figure 1.

It is observed that in Scenario 1, the adjustments of the models using the LMS, LTS and LQS estimators are almost identical and closer to the adjustment with the S estimator, same with the residuals of the last observations obtained through this estimator being slightly larger. In Scenario 2, there is a high leverage point ($h_{\max} = 0.888$) and the predicted value of the regression using the OLS estimator is the closest to the high leverage observation and the residue is less than the residues obtained via LMS, LTS, LQS and S. It is important to note that the LMS, LTS, and LQS adjustment differ little.



(a) Scenario 1



(b) Scenario 2

FIGURE 1. Adjustments of the regression model $y_i = \beta_0 + \beta_1 x_i + e_i, i = 1, ..., 20$, considering the estimators MQO, LMS, LTS, LQS and S in Scenarios 1 and 2.

A variant of HC0 is the HC3 estimator ([12]); given by $\text{HC3} = (X'X)^{-1}X'\widehat{\Psi}_3 X(X'X)^{-1}$ where $\widehat{\Psi}_3 = \text{diag}\left\{\frac{\widehat{e}_1^2}{(1-h_1)^2}, \ldots, \frac{\widehat{e}_n^2}{(1-h_n)^2}\right\}$, where $h_1, \ldots, h_n$ are the diagonal elements of H. [1] proposed another HC0 variant, that is based on HC3. The main idea behind his estimator, known as HC4, is to replace the exponent of $(1 - h_i)$, which is a fixed value in HC3, by quantities that take into account the ratio between each leverage measure and the mean leverage. The HC4 estimator is defined as follows: $\text{HC4} = (X'X)^{-1}X'\widehat{\Psi}_4 X(X'X)^{-1}$, where $\widehat{\Psi}_4 = \text{diag}\left\{\frac{\widehat{e}_1^2}{(1-h_1)^{\delta_1}}, \ldots, \frac{\widehat{e}_n^2}{(1-h_n)^{\delta_n}}\right\}$ and $\delta_i = \min\left\{4, \frac{h_i}{\overline{h}}\right\}$ with $\overline{h} = n^{-1}\sum_{i=1}^n h_i$, i.e., $\overline{h}$ is the average value of h_i's.

[2] proposed the HC4m, and in the article the authors show that test inference based on the robust standard error of heteroscedasticity is typically more reliable than when based on alternative standard errors, even under non-normal errors. The HC4m is defined as

HC4m $= (X'X)^{-1}X'\widehat{\Psi}_{4m}X(X'X)^{-1}$, where $\widehat{\Psi}_{4m} = \text{diag}\left\{\widehat{e}_1\sqrt{\frac{1}{(1-h_1)^{\eta_1}}}, \ldots, \widehat{e}_n\sqrt{\frac{1}{(1-h_n)^{\eta_n}}}\right\}$, where $\eta_i = \min\left\{\gamma_1, \frac{h_i}{\bar{h}}\right\} + \min\left\{\gamma_2, \frac{h_i}{\bar{h}}\right\}, i = 1, \ldots, n$, with $\gamma_1 = 1.0$ and $\gamma_2 = 1.5$.

[3] proposed yet another variant of White's estimator: HC5. The HC5 estimator is defined as HC5 $= (X'X)^{-1}X'\widehat{\Psi}_5X(X'X)^{-1}$, where $\widehat{\Psi}_5$ can be found by the diagonal $\widehat{\Psi}_5 = \text{diag}\left\{\widehat{e}_1\sqrt{\frac{1}{(1-h_1)^{\alpha_1}}}, \ldots, \widehat{e}_n\sqrt{\frac{1}{(1-h_n)^{\alpha_n}}}\right\}$. The exponent α_i is given by $\alpha_i = \min\left\{\frac{h_i}{\bar{h}}, \max\left\{4, \frac{kh_{\max}}{\bar{h}}\right\}\right\}$ where $h_{\max}$ denotes the maximal leverage and $0 < k < 1$ is a pre-defined constant. Since $0 < 1 - h_i < 1$ and $\alpha_i > 0$, it follows that $0 < (1 - h_i)^{\alpha_i} < 1$, $i = 1, ..., n$. When $h_i/\bar{h} \leq 4$ it follows that $\alpha_i = h_i/\bar{h}$. Based on a number of simulation results, [4], with errata [5], suggested using $k = 0.7$, which leads to reliable quasi-t inference.

3. Robust estimator

Diagnostic techniques for the detection of atypical data commonly used in linear regression analysis are based on the method of ordinary least squares estimation. However, if there are atypical observations in the data set, these observations will influence the adjustment, affecting the estimator obtained by this method.

In this work, methods of estimation with high rupture point are approached, that is, methods whose estimators are not affected when there are atypical observations in the data set. As an alternative to the residual ordinary least squares method are LMS, LTS, LQS and S. Diagnostic techniques based on in such methods are resistant to atypical data, being able to identify them. Our proposal is to use residuals obtained from LMS, LTS, LQS and S estimations when constructing the HC3, HC4, HC4m and HC5 HCCMEs.

[15] show the main idea of the Least median of squares (LMS) is to minimize the dispersion of the residuals. The least median of squares (LMS) estimator, given by minimize med $\hat{e}_i^2$. There is also another way to obtain this, by using a different objective function. [15] also presents an alternative that is the least trimmed squares estimator (LTS) that is given by *minimize* $\sum_i^s(\widehat{e}^2)_{i:n}$ where $(\widehat{e}_i^2)_{i:n} \leq \ldots \leq (\widehat{e}_i^2)_{n:n}$ are the ordered squared residuals and $s = \frac{n}{2} + \frac{p+1}{2}$.

[16] explain that the LMS estimator can be viewed as a special case of a larger family of estimators, namely the least quantize of squares estimators (LQS), which are defined by *minimize* $(\widehat{e}_i^2)_{([(1-\theta)n]+[\theta(p+1)]):n}$ where $0 \leq \theta \leq 50\%$. For θ tending to 50%, the LQS is asymptotically equivalent to the LMS. The breakdown point of the LQS is equal to θ for $n \to \infty$. *minimize max* $\widehat{e}_i^2(\beta)$, which is also referred to as minimax regression because the largest (absolute) residual is minimized.

[14] define the S-estimator is given by *minimize* $s(\widehat{e}_1(\beta), \ldots, \widehat{e}_n(\beta))$, where $s(\widehat{e}_1(\beta), \ldots, \widehat{e}_n(\beta))$ is the dispersion calculation, with final scale estimate $\hat{\sigma} = s(\widehat{e}_1(\widehat{\beta}), \ldots, \widehat{e}_n(\widehat{\beta}))$. The dispersion $s(\widehat{e}_1(\beta), ..., \widehat{e}_n(\beta))$ is defined as the solution of $\frac{1}{n}\sum_{i=1}^n \rho(\frac{\widehat{e}_i}{s}) = K$. K is often considered equal to $E_\phi[\rho]$, where ϕ is the standard normal. The function ρ must satisfy the following conditions: ρ is symmetric, continuously differentiable and $\rho(0) = 0$ and there exists $c > 0$ such that ρ is strictly increasing on $[0, c)$ and constant on $[c, \infty)$.

4. Numerical results

In this section, we will report results acquired through simulation for the performance of finite samples under heteroscedasticity in an unknown manner. Will be evaluated numerically based on the size of the associated quasi-t test of the null hypothesis H_0 : $\beta_2 = 0 \times H_1 : \beta_2 \neq 0$, coverage rate for $\widehat{\beta}_2$ and total relative bias.

Quasi-t test consider the null hypothesis test $H_0 : \beta = 0 \times H_1 : \beta \neq 0$, under the assumptions of the linear regression model, a statistic

$$\tau = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\widehat{var}(\beta_i)}}$$

has, under H_0, Student's t distribution with $n - k$ degrees of freedom. The standard error estimate is obtained considered differents heterokedasticity consistent standard errors.

The coverage rate is defined by the percentage of times that the true value of β_2 appears within the confidence interval constructed with 95%. Coverage rates are shown in Table 3.

The total relative bias is defined as the sum of the absolute values of the relative biases of the estimated variances of β_0, β_1 and β_2. This is for each estimator was estimated

$$\frac{|E\{\widehat{var}(\beta_0)\} - var(\beta_0)|}{var(\beta_0)} + \frac{|E\{\widehat{var}(\beta_1)\} - var(\beta_1)|}{var(\beta_1)} + \frac{|E\{\widehat{var}(\beta_2)\} - var(\beta_2)|}{var(\beta_2)}$$

where $\widehat{var}$ is the estimator of interest.

The Monte Carlo simulation of the numerical evaluation is based on the following linear regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$$

with $i = 1, ..., n$.

The data were extracted from [18] and the covariates x_1 and x_2 are the per capita income and the per capita income squared, respectively, the two being divided by 10^4. Each observation of x_1 and x_2 is repeated one, two and four times, forming samples of sizes $n = 50, 100, 200$, which are kept constant throughout the experiment. In this simulation, $\beta_0 = \beta_1 = 1$ and $\beta_2 = 0$ was used. In the homocedastic scenario $\lambda = 1$ and under heteroscedastic scenario the value of λ depends on the specification of the cedastic function, taking $\lambda > 1$. For heteroscedastic data, we consider a degree of heteroscedasticity close to 50.

The simulation experiments are based on 5,000 Monte Carlo replicas. All simulations were performed using the programming language `R` [17]. We used residuals from ordinary least squares estimators and robust estimators LMS, LTS, LQS and S. We performed eight Monte Carlo simulations under homoscedasticity and eight under heteroscedasticity for each set of residuals. The nomenclatures used in these tables are: HCt, $t = 3, 4, 4m$ and 5: residuals from the OLS, LMS, LTS, LQS or S estimator.

- HCt-OLS, $t = 3, 4, 4m$ e 5: residuals based in the OLS estimator;
- HCt-LMS, $t = 3, 4, 4m$ e 5: residuals based in the LMS estimator;
- HCt-LTS, $t = 3, 4, 4m$ e 5: residuals based in the LTS estimator;
- HCt-LQS, $t = 3, 4, 4m$ e 5: residuals based in the LQS estimator;
- HCt-S, $t = 3, 4, 4m$ e 5: residuals based in the S estimator.

The main results can be summarized as follows. We will start by reporting the results with leverage points. First, the HC5-based test, under leveraged data and using OLS (HC5-OLS), does not display reliable performance, both under homoskedasticity and heteroskedasticity. Indeed its null rejection rates are considerably larger than those of the HC3-OLS, HC4-OLS and HC4m-OLS tests. For example (Table 1), under homoskedasticity, $n = 50$ and at 5% nominal level, the null rejection rate of the HC5-based test is 12.83%, whereas textHC3-OLS), HC4-OLS) and HC4m-OLS) tests display sizes of 5.89%, 11.08% and 10.23%, respectively. The null rejection rates of the HC5-based test constructed using LMS, LTS, LQS and S robust residuals were closer to the nominal level than the tests constructed usig OLS; these tests, however, still display considerable size distortions under both homoskedasticity and heteroskedasticity. When based on robust residuals, the HC5 tests became more reliable. For example, under heteroskedasticity, $n = 50$ and at the 1% nominal level, the HC5-based test with OLS residuals, its null rejection rate is 33.47%, whereas using robust residual, its null rejection rates are 8.45%, 9.17%, 8.63% and 9.75%, respectively, as presented in Table 1. Under the same circumstances, but with $n = 200$, its null rejection rates are nearest to the nominal levels considered.

Second, under heteroskedasticity, the HC3, HC4, HC4m, HC5-based test also displays better performance when robust residuals are used. Consider, for example, $n = 100$ and at the 5% nominal level, the HC4-based test using OLS residuals rejects the null hypothesis 24.18% of the time, whereas using the robust residuals, the null rejection rates are 5.48%, 5.74%, 5.57% and 5.34%, respectively.

Third, when compared to other HCCME, HC3 for both homoskedasticity and heteroskedasticity, for sizes $n = 50$ and $n = 100$ we can verify that at the level of 1% the LMS residual presents the values nearest to the level of expected significance.

Fourth, for HC4-based tests, robust residuals presents results that are nearest to the level of sifnificannce assumed. For $n = 200$ and $\alpha = 5\%$, under homocedasticity the HCS-S showed the nearest percentage (5.17%) compared to other HCs based on the OLS residual or other robust residuals. Also for $n = 200$ and $\alpha = 1\%$, under heteroskedasticity, the HC4-based estimator LMS obtained 1.02 the nearest percentage being the 1% level. The HC4m-LQS produced two good results for rejection rate. The first was 4.97 for $n = 50$, $\alpha = 5\%$ and under homoscedasticity and 4.68 for $n = 100, \alpha = 1\%$ and under heteroscedasticity. Comparing the performance of the residuals applied in the HC4 and HC4m estimator, the rejection rates of the HC4m-based tests are lower and nearest to the significance levels tested. That is, the different residuals applied in the HC4m estimator present more reliable results than when applied in the HC4-based tests.

TABLE 1. Rejection percentages of quasi-t tests in the model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i, e_i \sim \mathcal{N}(0,1)$, with leverage points.

scenario	homokedasticy						heteroskedasticity					
n	$n = 50$		$n = 100$		$n = 200$		$n = 50$		$n = 100$		$n = 200$	
α	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
HC3-OLS	5.89	2.03	5.57	1.71	5.17	1.58	16.34	10.55	14.50	7.48	9.71	4.09
HC3-LMS	2.62	0.71	3.10	0.99	3.29	0.79	7.21	5.13	2.88	1.38	2.17	0.65
HC3-LTS	4.70	1.59	3.32	1.20	3.25	0.85	8.43	6.39	2.93	1.42	2.13	0.47
HC3-LQS	2.45	0.62	3.29	1.04	3.28	0.88	7.50	5.52	2.95	1.42	2.09	0.51
HC3-S	7.29	2.49	5.03	1.80	4.15	1.29	9.55	7.21	2.73	1.38	2.10	0.51
HC4-OLS	11.88	4.96	8.32	2.94	6.53	2.06	42.99	29.34	24.18	13.98	13.39	6.23
HC4-LMS	5.62	2.04	4.55	1.52	4.15	1.15	11.38	7.65	5.48	2.59	4.12	1.02
HC4-LTS	6.76	2.77	4.64	1.71	4.08	1.14	11.99	8.58	5.74	2.63	3.86	0.97
HC4-LQS	5.92	1.94	4.80	1.64	4.10	1.23	11.28	8.10	5.57	2.61	3.89	0.95
HC4-S	10.06	4.35	6.81	2.56	5.17	1.74	12.65	9.19	5.34	2.39	3.90	0.95
HC4m-OLS	10.23	4.03	7.59	2.50	6.11	1.94	33.43	22.11	21.14	11.82	12.21	5.63
HC4m-LMS	4.82	1.61	4.16	1.38	3.97	1.04	9.50	6.72	4.62	2.13	3.55	0.91
HC4m-LTS	6.21	2.47	4.27	1.53	3.80	1.05	10.31	7.74	4.66	2.24	3.24	0.79
HC4m-LQS	4.97	1.52	4.32	1.43	3.86	1.10	9.75	7.13	4.68	2.11	3.22	0.82
HC4m-S	9.33	3.85	6.25	2.33	4.88	1.59	11.14	8.43	4.36	1.91	3.35	0.80
HC5-OLS	12.83	5.46	8.78	3.13	6.73	2.14	47.77	33.47	25.86	15.07	13.93	6.58
HC5-LMS	5.95	2.25	4.87	1.58	4.27	1.19	12.49	8.45	6.19	2.75	4.52	1.07
HC5-LTS	6.99	2.85	4.81	1.81	4.15	1.20	13.24	9.17	6.33	2.85	4.19	1.06
HC5-LQS	6.36	2.18	5.04	1.76	4.22	1.30	12.48	8.63	6.11	2.87	4.25	1.03
HC5-S	10.35	4.57	7.03	2.65	5.33	1.79	13.70	9.75	5.97	2.65	4.29	1.02

An analysis of the influence of the observations corresponding to the states reveals that the observation corresponding to the State of Alaska is the highest point since the reference value $3\frac{p}{n}$ is equal to 0.180 and the diagonal element of the H matrix for this observation is 0.651. The others two points correspond to the states of Mississippi and Washington DC, where the diagonal element of matrix H are equal to 0.200 and 0.208, respectively.

In Table 2, with the data without the three points of leverage, we have that, firstly, the HC5-based test using OLS residuals produces the highest estimates for all the tests performed under heteroscedasticity. Under homoscedasticity, for $n = 47$ and $n = 94$, HC5 also produces the worst estimates. Only when $n = 188$ the HC3-based test using LTS residuals produce the most distant estimates (3.67 and 0.62, respectively, for 5% and 1%, respectively).

Second, for $n = 47$, under heteroscedasticity and 1% nominal level the HC3 using LMS residual produced produces nearest estimates (0.92) when compared to other HCCMES and residuals used.

Third, HC5 using S residual provided the best estimates for $n = 94$, $n = 188$, both nominals levels and consider heteroskedasticity. The other HCCME using residual S, considering $n = 188$ and 5%, presents estimates 3.39, 4.02 and 3.78 for HC3, HC4 and HC4m, respectively. Considering $n = 94$ and 5% nominal level, presents estimates 2.96, 4.22 and 3.61 for HC3, HC4 and HC4m-based test, respectively. Also, under these same conditions, HC3-based test and HC4-based test using S residual are the best estimates of their group for $\alpha = 5\%$ nominal level when compared to other residuals applied in HC3 and HC4.

TABLE 2. Rejection percentages of quasi-t tests in the model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i, e_i \sim \mathcal{N}(0,1)$, without leverage points.

scenario	homokedasticy						heteroskedasticity					
n	$n = 47$		$n = 94$		$n = 188$		$n = 47$		$n = 94$		$n = 188$	
α	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
HC3-OLS	5.33	1.28	5.47	1.36	5.21	1.11	9.31	3.60	7.44	2.45	5.81	1.66
HC3-LMS	2.92	0.56	3.52	0.69	3.89	0.74	3.14	0.92	2.87	0.58	3.29	0.69
HC3-LTS	3.00	0.55	3.44	0.64	3.67	0.62	3.01	0.86	2.89	0.55	3.42	0.61
HC3-LQS	2.97	0.55	3.67	0.80	3.93	0.73	3.07	0.79	2.96	0.60	3.32	0.71
HC3-S	4.00	0.83	4.50	1.03	4.58	0.91	3.19	0.91	2.96	0.64	3.39	0.71
HC4-OLS	7.20	1.94	6.23	1.57	5.79	1.26	12.89	5.72	9.06	3.20	6.57	1.91
HC4-LMS	4.16	0.97	4.07	0.92	4.27	0.85	4.70	1.64	4.04	0.83	3.96	0.80
HC4-LTS	4.15	0.92	3.99	0.85	3.98	0.81	4.58	1.48	4.01	0.89	4.04	0.73
HC4-LQS	4.09	0.95	4.19	0.95	4.35	0.83	4.74	1.52	4.13	0.86	3.93	0.82
HC4-S	5.39	1.34	5.13	1.22	5.01	1.05	4.91	1.60	4.22	0.89	4.02	0.79
HC4m-OLS	6.60	1.69	5.98	1.50	5.59	1.20	11.46	4.74	8.35	2.89	6.31	1.82
HC4m-LMS	3.61	0.77	3.77	0.82	4.11	0.81	3.99	1.29	3.56	0.71	3.67	0.74
HC4m-LTS	3.67	0.73	3.74	0.78	3.86	0.74	3.90	1.21	3.59	0.74	3.75	0.69
HC4m-LQS	3.62	0.77	4.03	0.89	4.18	0.77	3.92	1.21	3.70	0.76	3.63	0.74
HC4m-S	4.83	1.15	4.91	1.14	4.81	0.98	4.07	1.29	3.61	0.79	3.78	0.73
HC5-OLS	7.68	2.04	6.36	1.63	5.88	1.30	13.76	6.12	9.37	3.36	6.77	1.97
HC5-LMS	4.40	1.03	4.18	0.98	4.33	0.87	5.13	1.80	4.23	0.94	4.15	0.82
HC5-LTS	4.37	0.99	4.22	0.90	4.05	0.82	5.01	1.74	4.34	0.93	4.17	0.79
HC5-LQS	4.40	1.06	4.35	1.00	4.43	0.86	5.25	1.74	4.38	0.91	4.10	0.85
HC5-S	5.71	1.52	5.23	1.29	5.10	1.07	5.36	1.80	4.44	0.97	4.19	0.85

The numbers in Table 3 show that, first, under homoscedasticity ($\lambda = 1$), the coverage rate for HC3 using LMS, LTS and LQS residual have empirical coverage nearest to the nominal level (95%) for all sizes samples. The same happens for the HC3, HC4m and HC5-based test using LMS, LTS and LQS residual for $n = 100$ and $n = 200$. The application of the robust residuals in the HC4, HC4m and HC5-based test exhibit good coverage when the sample size is small ($n = 50$). Under heteroscedasticity, robust residuals for HC3, HC4 and HC4m-based test for $n = 100$ and $n = 200$, and HC5-based test for $n = 200$ exhibit coverage rates nearest to nominal coverage (95%). For heteroscedasticity, when $n = 50$, do not exhibit good coverage when the sample size is small ($n = 50$). For example, the empirical coverages for HC5-based test using OLS residual and from the S robust residual to β_2 when n = 50 are respectively 52.23% and 86.30%. We can also to punctuate the following observations:

Second, it was observed that HC5 using OLS residual, for both levels of significance, both sample sizes and for both heteroscedasticity and homoscedasticity, offers the worst coverage rates when compared to other HC using the other residuals. The HC5-OLS presents the following rates under homoscedasticity $87.17, 91.22$ and 93.27 for $n = 50, n = 100$ and $n = 200$, respectively. Under heteroscedasticity, rates are $52.23, 74.15$ and 86.07, respectively.

Third, HC3-based confidence interval produces the highest coverage rates. For homoscedasticity, considering $n = 50$, the HC3-based confidence interval using LQS residual provides 97.55 being the highest rate of all in this sample size and for $n = 100$ and $n = 200$ the rates are 96.90 , provided by HC3-LMS and 96.75 provided by HC3-LTS, respectively.

Fourth, considering HC4, HC4m and HC5-based confidence interval, for $n = 200$ and homoscedasticity, the confidence interval based on LTS residual provides the highest rates among its groups, that is, for HC4-based test the rate is 95.92 being the largest compared to the other residuals and sample sizes of HC4. In these same comparisons, the highest HC4m rate is 96.20 and for HC5 it is 95.85.

Fifth, to heteroskedasticity, the highest rates are, considering $n = 50$ to the HC3 using LMS residual provides 92.79 and $n = 100$ the HC3 using S residual provides 97.27. For $n = 200$ the highest rate is produced by HC3-based confidence interval using LQS residuals providing 97.91 being the highest rate of all in this sample size.

Sixth, for HC4 and HC5, considering $n = 200$ and under heteroskedasticity, the LTS residual provides the highest rates when compared to other residuals applied in HC4 and HC5.; for HC4 the rate is 96.14 and for HC5 is 95.81.

TABLE 3. Empirical converages for β_2, with leverage points and considering nominal level of 95%.

scenario	homokedasticy			heteroskedasticity		
n	50	100	200	50	100	200
HC3-OLS	94.11	94.43	94.83	83.66	85.50	90.29
HC3-LMS	97.38	96.90	96.71	92.79	97.12	97.83
HC3-LTS	95.30	96.68	96.75	91.57	97.07	97.87
HC3-LQS	97.55	96.71	96.72	92.50	97.05	97.91
HC3-S	92.71	94.97	95.85	90.45	97.27	97.90
HC4-OLS	88.12	91.68	93.47	57.01	75.82	86.62
HC4-LMS	94.39	95.45	95.85	88.62	94.52	95.88
HC4-LTS	93.24	95.36	95.92	88.01	94.26	96.14
HC4-LQS	94.08	95.20	95.90	88.72	94.43	96.11
HC4-S	89.94	93.19	94.83	87.35	94.66	96.10
HC4m-OLS	89.77	92.41	93.89	66.57	78.86	87.79
HC4m-LMS	95.18	95.84	96.04	90.50	95.38	96.45
HC4m-LTS	93.79	95.73	96.20	89.69	95.34	96.76
HC4m-LQS	95.03	95.68	96.14	90.25	95.32	96.78
HC4m-S	90.67	93.75	95.12	88.36	95.64	96.65
HC5-OLS	87.17	91.22	93.27	52.23	74.15	86.07
HC5-LMS	94.05	95.13	95.73	87.51	93.81	95.49
HC5-LTS	93.01	95.19	95.85	86.76	93.67	95.81
HC5-LQS	93.65	94.96	95.78	87.52	93.89	95.75
HC5-S	89.66	92.97	94.67	86.30	94.03	95.71

Table 4 shows the total relative bias. The individual relative bias of an estimator is defined as the average of the estimates minus the true value of the parameter, this difference being divided by the true value of the parameter. The total relative bias thus measures the aggregate bias of the variance estimates.

Under homoskedasticity the worst biases are given by HC3 using LMS residual for $n = 50$ (135.890) and for $n = 100$ and $n = 200$ by HC3 using LTS residual (15.386 and 4.686, respectively). Under heteroskedasticity and $n = 50$ the highest total relative bias was give by HC3 using LMS residual (27.579), for $n = 100$ by HC3 using LTS residual (5.478) and for $n = 200$ the highest total relative bias was given by HC3-LQS (2.496).

The best relative bias presented under homoskedastic and for all sample sizes is obtained by HC4 using OLS residual with $1.133, 0.917$ and 0.678 for $n = 50, n = 100$ and $n = 200$, in order. Under heteroskedasticity, HC4-OLS was also the best for $n = 100$ and $n = 200$ with 2.072 and 1.537, respectively. For $n = 50$ the lowest total relative bias was presented by HC5-OLS with 2.206. The HC4m using OLS residual when compared to the other residuals applied in HC4m or applied in other HCCMES, in all sample sizes under homoscedasticity and for $n = 100$ and $n = 200$ under heteroscedasticity, have the lowest total relative bias.

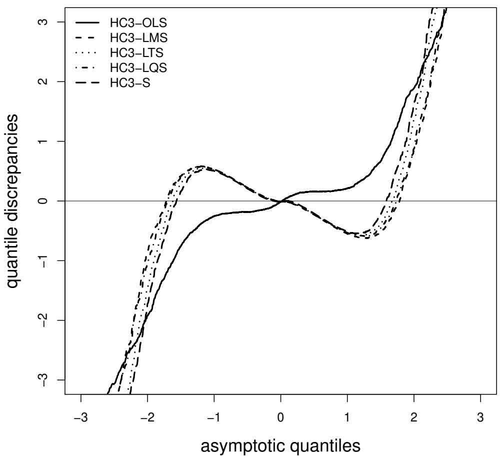
TABLE 4. Total relative bias in the model $y_t = \beta_0+\beta_1 x_{t1}+\beta_2 x_{t2}+e_t, e_t \sim \mathcal{N}(0, 1)$, with leverage points.

scenario	homokedasticy			heteroskedasticity		
n	50	100	200	50	100	200
HC3-OLS	3.1691	1.3497	0.8184	3.4273	2.9157	1.8289
HC3-LMS	135.8903	14.2593	3.94935	27.5795	5.4021	2.4874
HC3-LTS	129.0705	15.3863	4.68632	27.0827	5.4784	2.4902
HC3-LQS	116.0278	13.3991	3.82678	26.9672	5.3848	2.4965
HC3-S	57.8785	5.7687	1.79141	25.5908	5.2210	2.4054
HC4-OLS	1.1440	0.93179	0.68503	2.2769	2.1100	1.5539
HC4-LMS	27.4088	7.75063	2.92777	5.0659	2.7579	1.7978
HC4-LTS	26.1345	8.40304	3.49065	5.0222	2.8009	1.7974
HC4-LQS	23.3869	7.29274	2.83278	4.9671	2.7483	1.8032
HC4-S	10.0920	2.72361	1.16096	4.1153	2.3363	1.5392
HC4m-OLS	1.1331	0.91716	0.67823	2.3489	2.0724	1.5374
HC4m-LMS	21.9071	7.11083	2.80769	4.0822	2.5469	1.7308
HC4m-LTS	20.9030	7.7158	3.3494	4.0552	2.5848	1.7301
HC4m-LQS	18.6988	6.6939	2.7167	4.0094	2.5376	1.7357
HC4m-S	9.7213	2.9684	1.2937	3.8666	2.4800	1.6857
HC5-OLS	1.2893	0.9896	0.7073	2.2068	2.2345	1.6027
HC5-LMS	43.1061	9.2182	3.1860	8.1006	3.2903	1.9543
HC5-LTS	41.0528	9.9780	3.79407	8.0035	3.3414	1.9546
HC5-LQS	36.7710	8.6668	3.0834	7.9302	3.2778	1.9611
HC5-S	18.6736	3.7759	1.4497	7.5673	3.1865	1.8963

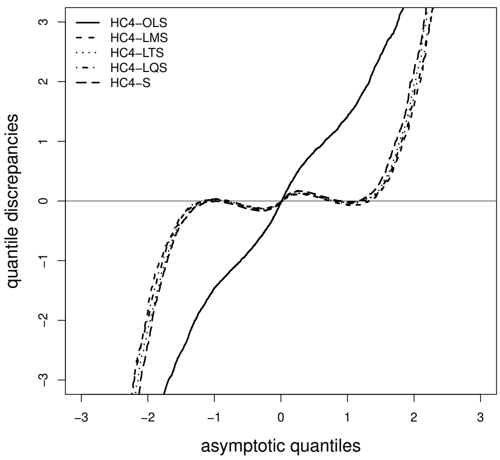
Figure 2 shows four graphs corresponding the data (high leverage); The images are related to heteroscedastic errors. O sample size is $n = 50$. We present the differences between exact quantile (estimated by simulation) and asymptotic quantile (from the normal standard distribution) against asymptotic quantiles of four select test statistics: their formulations HC3, HC4, HC4m and HC5.

The closer the lines are to the horizontal line ($y = 0$), the more reliable the inference. Figure 2(a) is related to the HC3-based covariance matrix, Figure 2(b) is HC4-based covariance matrix, Figure 2(c) is HC4m-based covariance matrix and, finally, matrix based on HC5 is presented in Figure 2(d).

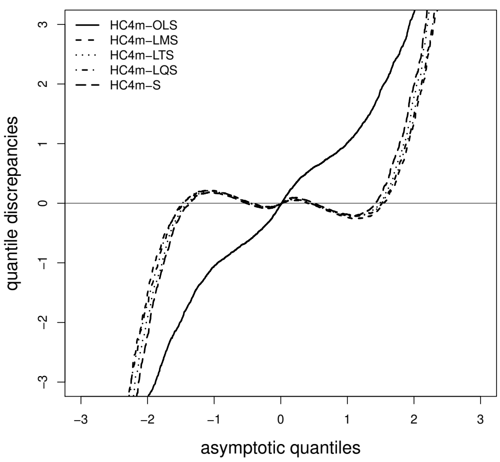
For both cases we conclude that the OLS residual is the most distant from line solid drawn at zero. LMS estimator, LTS estimator, LMS estimator and S estimator waste have similar behavior. In case 3, HC4m, the robust residuals have behavior closer to solid line ($y = 0$). In case 1, HC3-OLS is shown to be closer to the solid line.



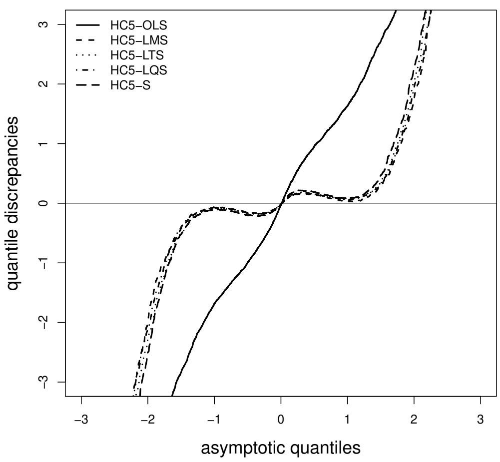
(a) HC3



(b) HC4



(c) HC4m



(d) HC5

FIGURE 2. Relative quantile discrepancy plots, $n = 50$: regression models with high leverage.

5. Empirical illustration

In what follows present one empirical applications that use real data. The dependent variable y is the number of monthly man-hours for manning installations in the U.S. Navy in Bachelor Office Quarters, and the independent variables are average daily occupancy x_1 and number of building wings x_2. The source of the data is [13] (Table 5.2, p. 218), the data contain 25 observations.

The model proposed to describe the relationship between the dependent variable and the explanatory variables is of the form

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i,$$

for $i = 1, ..., 25$.

In the Table 5, the mean of the dependent variable is 2109.39, the median is 1845.89 and has an amplitude of 8102.39. The data appear to be tilted to the right, which explains why the mean is greater than the median. The standard deviation is 1946.25. For the first independent variable (x_1), we have the median equal to 95.00, the mean equal to 118.36 and the standard deviation of 169.80. For the second independent variable (x_2), we have the median equal to 9.00, the mean equal to 11.12 and the standard deviation of 12.04. Comparing the coefficient of variation, the variable x_1 has greater variability than the other variables.

Table 5. Descriptive measures

Measure	y	x_1	x_2
Minimum	164.38	2.00	1.00
1° Quartile	931.80	25.89	3.00
Mean	2109.39	118.36	11.12
Median	1845.89	95.00	9.00
3° Quartile	3036.60	113.88	14.00
Maximum	8266.77	811.08	58.00
Standard deviation	1946.25	169.80	12.04
Coefficient of variation	0.92	1.43	1.08

We tested for the presence of heteroskedasticity using the test proposed by [9]. The null hypothesis of equal error variances was rejected at the 5% significance level (p-value$= 0.0213$). Hence, there is evidence that the data are heteroskedastic.

The linear parameters estimated by the OLS method presents point estimates of β_0, β_1 and β_2 equal $\widehat{\beta}_0 = 610.83$, $\widehat{\beta}_1 = 4.23$ and $\widehat{\beta}_2 = 89.71$. The coefficient of determination, R^2, that measures how much the independent variables can explain the dependent variable is $R^2 = 0.6124$.

The standard error is a measure of variation of a sample mean relative to the population mean. Thus, it is a measure that helps to verify the reliability of the calculated sample mean. To arrive at an estimate of the standard error, simply divide the deviation by the square root of the sample size. The result obtained will also be in the same unit of measurement of the sample value.

In Table 6, the standard error of $\widehat{\beta}_1$ was higher than when compared to $\widehat{\beta}_0$ and $\widehat{\beta}_2$. For HC3- based test, in all types of residuals, presented the biggest standard error than the other HC's. The HC5-based standard errors has the lowest among all standard errors.

Additionally, the HC5-based standard error using residuals LMS has the lowest among all HC5-based standard error.

We can observe that HC3-based test OLS residual produces standard error of $\widehat{\beta}_1$ that are more than 4.6 times larger than estimates of standard error of $\widehat{\beta}_1$ produced by HC4-based test OLS residual, 3.4 times larger than HC5-OLS and 5.3 times larger than the estimate of HC4m-based test OLS residual.

TABLE 6. Standard errors for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ in linear regression model of the form $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$, $i = 1, ..., 25$, with high leverage.

Estimator	Residual	HC3	HC4	HC4m	HC5
$\sqrt{\widehat{var}(\hat{\beta}_0)}$	OLS	691.5883	263.3244	305.1799	247.1552
$\sqrt{\widehat{var}(\hat{\beta}_1)}$		17.7897	3.8041	5.1049	3.3085
$\sqrt{\widehat{var}(\hat{\beta}_2)}$		127.4618	35.4533	45.0417	31.6079
$\sqrt{\widehat{var}(\hat{\beta}_0)}$	LMS	3041.0790	607.8197	839.9366	517.7016
$\sqrt{\widehat{var}(\hat{\beta}_1)}$		123.3402	24.4340	33.9112	20.7406
$\sqrt{\widehat{var}(\hat{\beta}_2)}$		681.0262	134.9863	187.2969	114.6038
$\sqrt{\widehat{var}(\hat{\beta}_0)}$	LTS	3006.2481	600.8021	830.2500	511.7245
$\sqrt{\widehat{var}(\hat{\beta}_1)}$		121.9372	24.1560	33.5254	20.5046
$\sqrt{\widehat{var}(\hat{\beta}_2)}$		673.2534	133.4441	185.1551	113.2959
$\sqrt{\widehat{var}(\hat{\beta}_0)}$	LQS	2892.1099	578.7600	799.2914	493.1872
$\sqrt{\widehat{var}(\hat{\beta}_1)}$		117.2939	23.2363	32.2489	19.7239
$\sqrt{\widehat{var}(\hat{\beta}_2)}$		647.6385	128.3824	178.1236	109.0020
$\sqrt{\widehat{var}(\hat{\beta}_0)}$	S	2978.6039	595.3609	822.6737	507.1179
$\sqrt{\widehat{var}(\hat{\beta}_1)}$		120.8152	23.9338	33.2169	20.3160
$\sqrt{\widehat{var}(\hat{\beta}_2)}$		667.0601	132.2237	183.4569	112.2621

In analyzing possible points of influence, we find that points 22 and 23 are strong candidates to be influential. These observations are points of high leverage since, the reference value $3p/n$ equals 0.360 and the elements of the main diagonal of the H matrix referring to points 22 and 23 are, respectively, 0.724 and 0.857.

In Table 7, we observe that, the HC3-based standard error produce the largest estimates of the standard error when compared to the other HC's, but the use of the OLS residual produces the lowest estimates. HC5-OLS produced the lowest estimates of the standard error. When we compared HC3 and HC5-based standard error using OLS residual, we have that estimates of HC5-based standard error using OLS residual are higthes less than 1.2 times the HC3-based standard error using OLS residual.

TABLE 7. Standard errors for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ in linear regression model of the form $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$, $i = 1, ..., 25$, without high leverage.

Estimator	Residual	HC3	HC4	HC4m	HC5
$\sqrt{\widehat{var}(\hat{\beta}_0)}$	OLS	123.7933	102.4562	106.0359	100.7859
$\sqrt{\widehat{var}(\hat{\beta}_1)}$		1.9652	1.4095	1.4824	1.3779
$\sqrt{\widehat{var}(\hat{\beta}_2)}$		25.9332	20.2269	21.1963	19.7754
$4\sqrt{\widehat{var}(\hat{\beta}_0)}$	LMS	129.8354	106.6675	110.6456	104.8049
$\sqrt{\widehat{var}(\hat{\beta}_1)}$		1.8241	1.3577	1.4220	1.3292
$\sqrt{\widehat{var}(\hat{\beta}_2)}$		27.2908	21.0297	22.1123	20.5238
$\sqrt{\widehat{var}(\hat{\beta}_0)}$	LTS	140.5977	113.7067	118.4583	111.4727
$\sqrt{\widehat{var}(\hat{\beta}_1)}$		1.7020	1.3836	1.4375	1.3584
$\sqrt{\widehat{var}(\hat{\beta}_2)}$		31.1082	23.6275	24.9490	23.0073
$\sqrt{\widehat{var}(\hat{\beta}_0)}$	LQS	249.0100	119.8854	135.5368	113.5280
$\sqrt{\widehat{var}(\hat{\beta}_1)}$		9.3625	3.5823	4.3477	3.2592
$\sqrt{\widehat{var}(\hat{\beta}_2)}$		46.2470	21.2605	24.3703	19.9832
$\sqrt{\widehat{var}(\hat{\beta}_0)}$	S	144.7718	116.5108	121.5096	114.1606
$\sqrt{\widehat{var}(\hat{\beta}_1)}$		1.7737	1.4490	1.5051	1.4228
$\sqrt{\widehat{var}(\hat{\beta}_2)}$		32.6523	24.7769	26.1694	24.1232

Consider the test of $H_0 : \beta_2 = 0 \times H_1 : \beta_2 \neq 0$. The estimated standard error of $\widehat{\beta}_2$ is obtained using the different estimators considerede here. This estimated standard error are used to construed quasi-t statistics test. The interest lies in determining whether x_2 should be removed from the regression model given that average daily occupancy x_1 is already in the model.

Table 8 contains the p-values tests for the complete data and for the data without the two leverage points. When all observations are used, the only tests that produce rejection of the null hypothesis at the usual significance levels are the test whose statistic uses the HC4, HC4m, and HC5-based test using OLS residual, that is, suggest that the number of building wings is important in the sense that variations in this regressor lead to significant variations, on mean, in the number of monthly man-hours. The quasi-t test based on the HC3 estimator using the LMS, LTS, LQS and S residues has the highest p-value among all the tests, with p-values above 0.88.

It is interesting to note that when the two leverage data points are removed from the data, all tests generate p-values above 0.5, thus, indicating that the evidence against the null hypothesis is very small. The OLS residuals when applied to the HC4, HC4m and HC5 estimators are unreliable, since the associated tests lead to misleading conclusion. This is not the case when robust residues are applied in the HC3, HC4, HC4m and HC5 estimators, whose associated tests present the same conclusions independent of to have or not influential observations. That is, the inferences made from tests constructed on

TABLE 8. P-values for β_2 in linear regression model of the form $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$, $i = 1, ..., 25$, with and without high leverage.

With high leverage ($n = 25$)		Without high leverage ($n = 23$)	
Test	p-value	Test	p-value
HC3-OLS	0.4815	HC3-OLS	0.6121
HC3-LMS	0.8952	HC3-LMS	0.5211
HC3-LTS	0.8940	HC3-LTS	0.5248
HC3-LQS	0.8898	HC3-LQS	0.5206
HC3-S	0.8930	HC3-S	0.5302
HC4-OLS	0.0114	HC4-OLS	0.6721
HC4-LMS	0.5063	HC4-LMS	0.5387
HC4-LTS	0.5014	HC4-LTS	0.5447
HC4-LQS	0.4847	HC4-LQS	0.5376
HC4-S	0.4975	HC4-S	0.5550
HC4m-OLS	0.0464	HC4m-OLS	0.6514
HC4m-LMS	0.6320	HC4m-LMS	0.5316
HC4m-LTS	0.6280	HC4m-LTS	0.5367
HC4m-LQS	0.6145	HC4m-LQS	0.5307
HC4m-S	0.6248	HC4m-S	0.5450
HC5-OLS	0.0045	HC5-OLS	0.6822
HC5-LMS	0.4337	HC5-LMS	0.5428
HC5-LTS	0.4285	HC5-LTS	0.5492
HC5-LQS	0.4105	HC5-LQS	0.5416
HC5-S	0.4242	HC5-S	0.5607

the basis of robust residues applied in the estimators HC3, HC4, HC4m and HC5 are not sensitive to the presence of high leverage observations.

In order to examine the impact of observations on inference, the model considered was estimated without point 22, without point 23 and without point 22 and 23. The estimates of the parameter β_i, for $i = 0, 1, 2$, are presented in Table 9 .

When the observations are not in the sample, the estimate of $\widehat{\beta}_0$ is much lower than when compared to the model estimate with all points. Without points 22 and 23 the estimate of $\widehat{\beta}_1$ is greater nearly 5 times than the estimate with all points. Without point 23, the estimate of $\widehat{\beta}_2$ changes sign.

TABLE 9. Estimates of β_i, with and without leverage points.

Points Removed	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
none	610.83	4.23	89.71
22	107.91	2.59	171.15
23	176.04	21.88	-7.70
22 and 23	105.00	20.91	10.30

6. CONCLUSIONS

We evaluated, through the quasi-t test sizes, coverage rates, relative total bias, the behaviors of the estimators HC3, HC4, HC4m and HC5 of the covariance matrix of the estimator of OLS when residuals from robust regressions (LMS, LTS, LQS and S) are used to replace OLS residuals in the linear regression model.

The results showed that HC3-based test showed better performance when residuals from robust regression are used instead of OLS residuals, in the presence or in the absence of leverage points.

The test based on the HC4 estimator has superior performance when the residuals from the LMS estimators are used for large sample sizes. The performance of the test based on the HC4m and HC5 estimator improves significantly when using robust residuals. When there are no leverage points, the best results for the quasi-t test were provided by residual S.

In the application, we observed that the null hypothesis of the quasi- t test is rejected for tests based on HC4, HC4m and HC5 using the OLS method at the level of 5% when there are leverage points. When there are no leverage points, the p -values of the null hypothesis ($H_0 : \beta_0 = 0$) of the quasi- t tests do not are rejected. The coefficients of β_i, for $i = 0, 1, 2$ are much smaller, larger, or invert the signal in the presence of high leverage points.

We recommend that practitioners use residuals from robust regressions to construct the HC'S-based test statistics.

References

[1] Cribari-Neto, F. Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis 45*, 2 (2004), 215–233.

[2] Cribari-Neto, F., and da Silva, W. B. A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model. *AStA Advances in Statistical Analysis 95*, 2 (2010), 129–146.

[3] Cribari-Neto, F., and Souza, T. C. Testing inference in variable dispersion beta regressions. *Journal of Statistical Computation and Simulation 82*, 12 (2012), 1827–1843.

[4] Cribari-Neto, F., Souza, T. C., and Vasconcellos, K. L. Inference under heteroskedasticity and leveraged data. *Communications in Statistics Theory and Methods 36*, 10 (2007), 1877–1888.

[5] Cribari-Neto, F., Souza, T. C., and Vasconcellos, K. L. Errata: Inference under heteroskedasticity and leveraged data, communications in statistics, theory and methods, 36, 1877–1888, 2007.

[6] Cribari-Neto, F., and Zarkos, S. G. Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. *Econometric Reviews 18*, 2 (1999), 211–228.

[7] Cribari-Neto, F., and Zarkos, S. G. Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. *Journal of Statistical Computation and Simulation 68*, 4 (2001), 391–411.

[8] Judge, G. C.; Hill, R. C. G. W. E. L. H. L. T. C. The theory and practice of econometrics.

[9] Koenker, R. A note on studentizing a test for heteroscedasticity. *Journal of Econometrics 17*, 1 (1981), 107–112.

[10] Long, J. S., and Ervin, L. H. Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician 54*, 3 (2000), 217–224.

[11] MacKinnon, J. G., and White, H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of econometrics 29*, 3 (1985), 305–325.

[12] Murphy, K. M., and Topel, R. H. Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics 20*, 1 (2002), 88–97.

[13] Myers, R. H., and Myers, R. H. *Classical and modern regression with applications*, vol. 2. Duxbury press Belmont, CA., 1990.

[14] Rousseeuw, P., and Yohai, V. Robust regression by means of s-estimators. In *Robust and nonlinear time series analysis*. Springer, 1984, pp. 256–272.

[15] Rousseeuw, P. J. Least median of squares regression. *Journal of the American statistical association 79*, 388 (1984), 871–880.

[16] Rousseeuw, P. J., and Leroy, A. M. *Robust regression and outlier detection*, vol. 1. Wiley Online Library, 1987.

[17] Team, R. C., et al. R: A language and environment for statistical computing.

[18] W.H., G. *Econometric Analysis*, vol. 3nd edition. Upper Sanddler River: Prentice Hall., 1997.

[19] White, H., et al. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*, 4 (1980), 817–838.

Departamento de Estatística, Universidade Federal da Paraíba, Cidade Universitária, João Pessoa/PB, 58089–900, Brazil

E-mail address, N. S. Lourenço: natalialourenco2811@hotmail.com

2 Considerações finais

Avaliamos, através dos tamanhos de teste quase-t, as taxas de cobertura, o viés total relativo, os comportamentos do estimadores HC3, HC4, HC4m e HC5 da matriz de covariância do estimador de OLS quando os resíduos de regressões robustas (LMS, LTS, LQS e S) são utilizado para substituir resíduos OLS no modelo de regressão linear.

Os resultados mostraram que o teste baseado em HC3 mostrou melhor desempenho quando os resíduos de regressão robusta são utilizados em vez de resíduos OLS, na presença ou no ausência de pontos de alavancagem.

O teste baseado no estimador HC4 possui desempenho superior quando os resíduos dos estimadores LMS são usado para amostras grandes. O desempenho do teste baseado no estimador HC4m e HC5 melhora significativamente ao usar resíduos robustos. Quando não há pontos de alavancagem, os melhores resultados para o teste quasi- t foram fornecidos pelo resíduo S.

Na aplicação, observamos que a hipótese nula do teste quasi- t é rejeitado para testes baseados em HC4, HC4m e HC5 usando o método OLS no nível de 5% quando houver pontos de alavancagem. Quando não há pontos de alavancagem, os p-valores da hipótese nula ($H_0 : \beta_2 = 0$) dos testes quasi- t não são rejeitados. Os coeficientes de β_2 são muito menores ou iguais à mediana das estimativas quando pontos de alavancagem estão presentes.

Recomendamos que os profissionais usem resíduos de regressões robustas para construir as estatísticas de teste baseadas no HC.

3 Referências

Referências

[1] Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics and Data Analysis*, 45, 215–233.

[2] Cribari-Neto F, da Silva WB. A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model. AStA Adv Statist Anal. 2011;95(2):129–146.

[3] Cribari-Neto, F.; Souza, T.C. (2012). Testing inference in variable dispersion beta regressions. *Journal of Statistical Computation and Simulation*, 82, 1827–1843.

[4] Cribari-Neto, F.; Souza, T.C. & Vasconcellos, K. L. P. (2007) Inference Under Heteroskedasticity and Leveraged Data, Communications in Statistics - Theory and Methods, 36:10, 1877-1888

[5] Cribari-Neto, F.; Souza, T.C. & Vasconcellos, K. L. P. (2007) Errata: Inference Under Heteroskedasticity and Leveraged Data, Communications in Statistics - Theory and Methods, 36:10, 1877-1888

[6] Cribari-Neto, F.; Zarkos, S. (1999). Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. Econometric Reviews 18, 211–228.

[7] Cribari-Neto F, Zarkos SG (2001). Heteroskedasticity-consistent covariancematrix estimation:White's estimator and the bootstrap. J Stat Comput Simul. 68(4):391–411.

[8] Judge, G. C.; Hill, R. C.; Griffiths, W. E.; Lutkepohl, H.; Lee, T. C. The Theory and Pratice of Econometrics. New York: Wiley, 1985.

[9] Koenker,R. (1981). Anote on Studentizing a test for heteroscedasticity. Journal of Econometrics 17,107–112.

[10] Long, J.; Ervin, L. (2000). Using heteroskedasticity-consistent standard errors in the linear regression model. American Statistician 54, 217–224.

[11] MacKinnon, J. G.; White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite-sample properties. Journal of Econometrics 29, 305–325.

[12] Murphy, Kevin M and Topel, Robert H. (2002) Estimation and inference in two-step econometric models, Journal of Business & Economic Statistics, v. 20, n. 1, p. 88-97, Taylor & Francis.

[13] Myers, R. H. (1990). Classical and Modern Regression with Applications. Belmont: Duxbury Press.

[14] Rousseeuw, P. J. (1984). Least Median of Squares Regression. Journal of the American Statistical Association, 79(388), 871–880.

[15] Rousseeuw, Peter J. and Leroy, Annick M. (1987). Robust regression and outlier detection. *Wiley*.

[16] Rousseeuw, P. J., and Yohai, V. (1984), Robust regression by means of S-estimators, in Robust and Nonlinear Time Series Analysis, edited by J. Franke, W. Hiirdle, and R. D. Martin, Lecture Notes in Statistics No. 26, Springer Verlag, New York, pp. 256-272.

[17] White, H. (1980). A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838.

[18] White, H. (1994). *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.

[19] Ihaka, Ross and Gentleman, Robert. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

4 Apêndice

4.1 Simulação

```
rm(list=ls())
library(MASS)
library(car)  # hc
library(sandwich)
library(xtable)
#library(xlsx)
#library(xtable)
library(carData)
library(openxlsx)
library(sandwich)
library(car)
data("PublicSchools", package = "sandwich")

bd = PublicSchools

# quantis da normal
vc10 = qnorm(0.95)
vc5 = qnorm(0.975)
vc1 = qnorm(0.995)

nr=nrow(bd)-1
p=3
Vr=3*p/nr

y=c(bd$Expenditure)[-50]
x1=c(bd$Income)[-50]/10000
x2=c(bd$Income^2)[-50]/10000


X=cbind(1,x1,x2)

invX = solve(t(X)%*%X)
invXX = invX%*%t(X)
Xinv = X%*%invX

# matrix de alavanca
hvalues = hat(X)
hmean = mean(hvalues)
sort(hvalues)
```

```
# ajustando o modelo - lm
ajustelm=lm(y~x1+x2)
resid_lm= as.numeric(ajustelm$resid)
slm<-summary(ajustelm)
# ajustando o modelo - lms
ajustelms=lqs(y~x1+x2, method = "lms")
resid_lms= as.numeric(ajustelms$resid)
# ajustando o modelo - lts
ajustelts=lqs(y~x1+x2, method = "lts")
resid_lts= as.numeric(ajustelts$resid)
# ajustando o modelo - lqs
ajustelqs=lqs(y~x1+x2, method = "lqs")
resid_lqs= as.numeric(ajustelqs$resid)
# ajustando o modelo - lts
ajusteS=lqs(y~x1+x2, method = "S")
resid_S= as.numeric(ajusteS$resid)

#####HC3
deltaHC3=2

meatHC3_lm=diag(resid_lm^2/((1-hvalues)^deltaHC3))
HC3_lm = invXX%*%meatHC3_lm%*%Xinv

meatHC3_lts=diag(resid_lts^2/((1-hvalues)^deltaHC3))
HC3_lts = invXX%*%meatHC3_lts%*%Xinv

meatHC3_lms=diag(resid_lms^2/((1-hvalues)^deltaHC3))
HC3_lms = invXX%*%meatHC3_lms%*%Xinv

meatHC3_lqs=diag(resid_lqs^2/((1-hvalues)^deltaHC3))
HC3_lqs = invXX%*%meatHC3_lqs%*%Xinv

meatHC3_S=diag(resid_S^2/((1-hvalues)^deltaHC3))
HC3_S = invXX%*%meatHC3_S%*%Xinv


##HC4
deltaHC4 = min(4,(hvalues/hmean))

meatHC4_lm=diag(resid_lm^2/((1-hvalues)^deltaHC4))
HC4_lm = invXX%*%meatHC4_lm%*%Xinv

meatHC4_lts=diag(resid_lts^2/((1-hvalues)^deltaHC4))
HC4_lts = invXX%*%meatHC4_lts%*%Xinv

meatHC4_lms=diag(resid_lms^2/((1-hvalues)^deltaHC4))
HC4_lms = invXX%*%meatHC4_lms%*%Xinv
```

```
meatHC4_lqs=diag(resid_lqs^2/((1-hvalues)^deltaHC4))
HC4_lqs = invXX%*%meatHC4_lqs%*%Xinv

meatHC4_S=diag(resid_S^2/((1-hvalues)^deltaHC4))
HC4_S = invXX%*%meatHC4_S%*%Xinv

###HC4m
gamma1=1;gamma2=1.5
deltaHC4m=min(gamma1,(hvalues/hmean))+min(gamma2,
(hvalues/hmean))

meatHC4m_lm=diag(resid_lm^2/((1-hvalues)^deltaHC4m))
HC4m_lm = invXX%*%meatHC4m_lm%*%Xinv

meatHC4m_lts=diag(resid_lts^2/((1-hvalues)^deltaHC4m))
HC4m_lts = invXX%*%meatHC4m_lts%*%Xinv

meatHC4m_lms=diag(resid_lms^2/((1-hvalues)^deltaHC4m))
HC4m_lms = invXX%*%meatHC4m_lms%*%Xinv

meatHC4m_lqs=diag(resid_lqs^2/((1-hvalues)^deltaHC4m))
HC4m_lqs = invXX%*%meatHC4m_lqs%*%Xinv

meatHC4m_S=diag(resid_S^2/((1-hvalues)^deltaHC4m))
HC4m_S = invXX%*%meatHC4m_S%*%Xinv

#HC5
k=0.7;hmax=max(hvalues)
deltaHC5<-min((hvalues/hmean),max(4,k*hmax/hmean))

meatHC5_lm=diag(resid_lm^2/((1-hvalues)^(deltaHC5/2)))
HC5_lm = invXX%*%meatHC5_lm%*%Xinv

meatHC5_lts=diag(resid_lts^2/((1-hvalues)^(deltaHC5/2)))
HC5_lts = invXX%*%meatHC5_lts%*%Xinv

meatHC5_lms=diag(resid_lms^2/((1-hvalues)^(deltaHC5/2)))
HC5_lms = invXX%*%meatHC5_lms%*%Xinv

meatHC5_lqs=diag(resid_lqs^2/((1-hvalues)^(deltaHC5/2)))
HC5_lqs = invXX%*%meatHC5_lqs%*%Xinv

meatHC5_S=diag(resid_S^2/((1-hvalues)^(deltaHC5/2)))
HC5_S = invXX%*%meatHC5_S%*%Xinv

##ERRO PADRÃO

summary(ajustelm) #Erro padrão
```

```
#HC3

eplmHC3b0=sqrt(HC3_lm[1,1])
eplmHC3b1=sqrt(HC3_lm[2,2])
eplmHC3b2=sqrt(HC3_lm[3,3])

eplmsHC3b0=sqrt(HC3_lms[1,1])
eplmsHC3b1=sqrt(HC3_lms[2,2])
eplmsHC3b2=sqrt(HC3_lms[3,3])

epltsHC3b0=sqrt(HC3_lts[1,1])
epltsHC3b1=sqrt(HC3_lts[2,2])
epltsHC3b2=sqrt(HC3_lts[3,3])

eplqsHC3b0=sqrt(HC3_lqs[1,1])
eplqsHC3b1=sqrt(HC3_lqs[2,2])
eplqsHC3b2=sqrt(HC3_lqs[3,3])

epSHC3b0=sqrt(HC3_S[1,1])
epSHC3b1=sqrt(HC3_S[2,2])
epSHC3b2=sqrt(HC3_S[3,3])

#Hc4
eplmHC4b0=sqrt(HC4_lm[1,1])
eplmHC4b1=sqrt(HC4_lm[2,2])
eplmHC4b2=sqrt(HC4_lm[3,3])

eplmsHC4b0=sqrt(HC4_lms[1,1])
eplmsHC4b1=sqrt(HC4_lms[2,2])
eplmsHC4b2=sqrt(HC4_lms[3,3])

epltsHC4b0=sqrt(HC4_lts[1,1])
epltsHC4b1=sqrt(HC4_lts[2,2])
epltsHC4b2=sqrt(HC4_lts[3,3])

eplqsHC4b0=sqrt(HC4_lqs[1,1])
eplqsHC4b1=sqrt(HC4_lqs[2,2])
eplqsHC4b2=sqrt(HC4_lqs[3,3])

epSHC4b0=sqrt(HC4_S[1,1])
epSHC4b1=sqrt(HC4_S[2,2])
epSHC4b2=sqrt(HC4_S[3,3])

#HC4m

eplmHC4mb0=sqrt(HC4m_lm[1,1])
eplmHC4mb1=sqrt(HC4m_lm[2,2])
```

```
eplmHC4mb2=sqrt(HC4m_lm[3,3])

eplmsHC4mb0=sqrt(HC4m_lms[1,1])
eplmsHC4mb1=sqrt(HC4m_lms[2,2])
eplmsHC4mb2=sqrt(HC4m_lms[3,3])

epltsHC4mb0=sqrt(HC4m_lts[1,1])
epltsHC4mb1=sqrt(HC4m_lts[2,2])
epltsHC4mb2=sqrt(HC4m_lts[3,3])

eplqsHC4mb0=sqrt(HC4m_lqs[1,1])
eplqsHC4mb1=sqrt(HC4m_lqs[2,2])
eplqsHC4mb2=sqrt(HC4m_lqs[3,3])

epSHC4mb0=sqrt(HC4m_S[1,1])
epSHC4mb1=sqrt(HC4m_S[2,2])
epSHC4mb2=sqrt(HC4m_S[3,3])

#Hc5

eplmHC5b0=sqrt(HC5_lm[1,1])
eplmHC5b1=sqrt(HC5_lm[2,2])
eplmHC5b2=sqrt(HC5_lm[3,3])

eplmsHC5b0=sqrt(HC5_lms[1,1])
eplmsHC5b1=sqrt(HC5_lms[2,2])
eplmsHC5b2=sqrt(HC5_lms[3,3])

epltsHC5b0=sqrt(HC5_lts[1,1])
epltsHC5b1=sqrt(HC5_lts[2,2])
epltsHC5b2=sqrt(HC5_lts[3,3])

eplqsHC5b0=sqrt(HC5_lqs[1,1])
eplqsHC5b1=sqrt(HC5_lqs[2,2])
eplqsHC5b2=sqrt(HC5_lqs[3,3])

epSHC5b0=sqrt(HC5_S[1,1])
epSHC5b1=sqrt(HC5_S[2,2])
epSHC5b2=sqrt(HC5_S[3,3])


# eplmHC5mb0=sqrt(HC5m_lm[1,1])
# eplmHC5mb1=sqrt(HC5m_lm[2,2])
#eplmHC5mb2=sqrt(HC5m_lm[3,3])


M=c("OLS","HC3","HC4","HC4m","HC5","HC5m")
MOLSlm=c(340.206,1.669,23.535)
```

```
MHC3lm=c(round(eplmHC3b0,4),round(eplmHC3b1,4),
round(eplmHC3b2,4))
MHC4lm=c(round(eplmHC4b0,4),round(eplmHC4b1,4),
round(eplmHC4b2,4))
MHC4mlm=c(round(eplmHC4mb0,4),round(eplmHC4mb1,4),
round(eplmHC4mb2,4))
MHC5lm=c(round(eplmHC5b0,4),round(eplmHC5b1,4),
round(eplmHC5b2,4))

MHC3lms=c(round(eplmsHC3b0,4),round(eplmsHC3b1,4),
round(eplmsHC3b2,4))
MHC4lms=c(round(eplmsHC4b0,4),round(eplmsHC4b1,4),
round(eplmsHC4b2,4))
MHC4mlms=c(round(eplmsHC4mb0,4),round(eplmsHC4mb1,4),
round(eplmsHC4mb2,4))
MHC5lms=c(round(eplmsHC5b0,4),round(eplmsHC5b1,4),
round(eplmsHC5b2,4))

MHC3lts=c(round(epltsHC3b0,4),round(epltsHC3b1,4),
round(epltsHC3b2,4))
MHC4lts=c(round(epltsHC4b0,4),round(epltsHC4b1,4),
round(epltsHC4b2,4))
MHC4mlts=c(round(epltsHC4mb0,4),round(epltsHC4mb1,4),
round(epltsHC4mb2,4))
MHC5lts=c(round(epltsHC5b0,4),round(epltsHC5b1,4),
round(epltsHC5b2,4))

MHC3lqs=c(round(eplqsHC3b0,4),round(eplqsHC3b1,4),
round(eplqsHC3b2,4))
MHC4lqs=c(round(eplqsHC4b0,4),round(eplqsHC4b1,4),
round(eplqsHC4b2,4))
MHC4mlqs=c(round(eplqsHC4mb0,4),round(eplqsHC4mb1,4),
round(eplqsHC4mb2,4))
MHC5lqs=c(round(eplqsHC5b0,4),round(eplqsHC5b1,4),
round(eplqsHC5b2,4))

MHC3S=c(round(epSHC3b0,4),round(epSHC3b1,4),
round(epSHC3b2,4))
MHC4S=c(round(epSHC4b0,4),round(epSHC4b1,4),
round(epSHC4b2,4))
MHC4mS=c(round(epSHC4mb0,4),round(epSHC4mb1,4),
round(epSHC4mb2,4))
MHC5S=c(round(epSHC5b0,4),round(epSHC5b1,4),
round(epSHC5b2,4))


E=c("SQRTvarb0","SQRTvarb1","SQRTvarb2")
lm=rep("OLS",3)
```

```
lms=rep("LMS",3)
lts=rep("LTS",3)
lqs=rep("LQS",3)
S=rep("S",3)

TR20 = cbind(E,lm,MHC3lm,MHC4lm,MHC4mlm,MHC5lm)
TR21 = cbind(E,lms,MHC3lms,MHC4lms,MHC4mlms,MHC5lms)
TR22 = cbind(E,lts,MHC3lts,MHC4lts,MHC4mlts,MHC5lts)
TR23 = cbind(E,lqs,MHC3lqs,MHC4lqs,MHC4mlqs,MHC5lqs)
TR24 = cbind(E,S,MHC3S,MHC4S,MHC4mS,MHC5S)


TR2=rbind(TR20,TR21,TR22,TR23,TR24)
colnames(TR2)<-c("Estimador", M)
TR2

xtable(TR2)


#################################


###TESTE QUASI-T
summary(ajustelm) #p-valor
coeflm1 =  ajustelm$coef[3]

# considerando o estimador HC3
testOLSHC3 = coeflm1 / eplmHC3b2
# considerando o estimador HC4
testOLSHC4 = coeflm1/ eplmHC4b2
# considerando o estimador HC4m
testOLSHC4m = coeflm1/ eplmHC4mb2
# considerando o estimador HC5
testOLSHC5 = coeflm1/ eplmHC5b2


# considerando o estimador HC3
testLMSHC3 = coeflm1 / eplmsHC3b2
# considerando o estimador HC4
testLMSHC4 = coeflm1/ eplmsHC4b2
# considerando o estimador HC4m
testLMSHC4m = coeflm1/ eplmsHC4mb2
# considerando o estimador HC5
testLMSHC5 = coeflm1/ eplmsHC5b2

# considerando o estimador HC3
testLTSHC3 = coeflm1 / epltsHC3b2
# considerando o estimador HC4
```

```
testLTSHC4 = coeflm1/ epltsHC4b2
# considerando o estimador HC4m
testLTSHC4m = coeflm1/ epltsHC4mb2
# considerando o estimador HC5
testLTSHC5 = coeflm1/ epltsHC5b2

# considerando o estimador HC3
testLQSHC3 = coeflm1 / eplqsHC3b2
# considerando o estimador HC4
testLQSHC4 = coeflm1/ eplqsHC4b2
# considerando o estimador HC4m
testLQSHC4m = coeflm1/ eplqsHC4mb2
# considerando o estimador HC5
testLQSHC5 = coeflm1/ eplqsHC5b2


# considerando o estimador HC3
testSHC3 = coeflm1 / epSHC3b2
# considerando o estimador HC4
testSHC4 = coeflm1/ epSHC4b2
# considerando o estimador HC4m
testSHC4m = coeflm1/ epSHC4mb2
# considerando o estimador HC5
testSHC5 = coeflm1/ epSHC5b2



p1lm=pnorm(testOLSHC3,lower.tail = F)
p2lm=pnorm(testOLSHC4,lower.tail = F)
p3lm=pnorm(testOLSHC4m,lower.tail = F)
p4lm=pnorm(testOLSHC5,lower.tail = F)
#p5lm=pnorm(testOLSHC5m,lower.tail = F)

p1lms=pnorm(testLMSHC3,lower.tail = F)
p2lms=pnorm(testLMSHC4,lower.tail = F)
p3lms=pnorm(testLMSHC4m,lower.tail = F)
p4lms=pnorm(testLMSHC5,lower.tail = F)

p1lts=pnorm(testLTSHC3,lower.tail = F)
p2lts=pnorm(testLTSHC4,lower.tail = F)
p3lts=pnorm(testLTSHC4m,lower.tail = F)
p4lts=pnorm(testLTSHC5,lower.tail = F)

p1lqs=pnorm(testLQSHC3,lower.tail = F)
p2lqs=pnorm(testLQSHC4,lower.tail = F)
p3lqs=pnorm(testLQSHC4m,lower.tail = F)
p4lqs=pnorm(testLQSHC5,lower.tail = F)
```

```
p1S=pnorm(testSHC3,lower.tail = F)
p2S=pnorm(testSHC4,lower.tail = F)
p3S=pnorm(testSHC4m,lower.tail = F)
p4S=pnorm(testSHC5,lower.tail = F)


M1=c("HC3","HC4","HC4m","HC5")
M20=c(round(p1lm[[1]],4),round(p2lm[[1]],4),
round(p3lm[[1]],4),round(p4lm[[1]],4))
M21=c(round(p1lms[[1]],4),round(p2lms[[1]],4),
round(p3lms[[1]],4),round(p4lms[[1]],4))
M22=c(round(p1lts[[1]],4),round(p2lts[[1]],4),
round(p3lts[[1]],4),round(p4lts[[1]],4))
M23=c(round(p1lqs[[1]],4),round(p2lqs[[1]],4),
round(p3lqs[[1]],4),round(p4lqs[[1]],4))
M24=c(round(p1S[[1]],4),round(p2S[[1]],4),
round(p3S[[1]],4),round(p4S[[1]],4))



TR3 = cbind(M1,M20,M21,M22,M23,M24)

colnames(TR3)<-c("teste", "p-valor OLS",
"p-valor LMS", "p-valor LTS",
"p-valor LQS","p-valor S")
TR3

xtable(TR3)



#####Pontos de alavanca #
X=cbind(1,x1,x2)
hvalues = hat(X)
sort(hvalues)


###SEM pontos de alavanca
y=c(bd$Expenditure)[-c(2,24,48,50)]
x1=c(bd$Income)[-c(2,24,48,50)]/10000
x2=c(bd$Income^2)[-c(2,24,48,50)]/10000

####
```

4.2 Aplicação

```
rm(list=ls())
library(MASS)
library(car)  # hc
library(sandwich)
library(xtable)
#library(xlsx)
#library(xtable)
library(carData)
library(openxlsx)
library(sandwich)
library(car)



dados=read.table(file = "TCC/marinha.txt",
header=F, dec=".")
attach(dados)
#bd = PublicSchools

y=dados[-c(22,23),3]
x1=dados[-c(22,23),1]
x2=dados[-c(22,23),2]

summary(dados)
sd(y);sd(x1);sd(x2)


col1=c(min(y),min(x1),min(x2))
col2=c(max(y),max(x1),max(x2))
col3=c(mean(y),mean(x1),mean(x2))
col4=c(median(y),median(x1),median(x2))
col5=c(sd(y),sd(x1),sd(x2))

m=matrix(c(col1,col2,col3,col4,col5),nrow =5,byrow = T)
colnames(m)=c("y","x_1","x_2")
rownames(m)=c("Minimum","Maximum","Mean", "Median",
"Standard deviation")
xtable(m)


ajustelm=lm(y~x1+x2)
summary(ajustelm)
library(lmtest)
bptest(ajustelm)
#
```

```
# quantis da normal
vc10 = qnorm(0.95)
vc5 = qnorm(0.975)
vc1 = qnorm(0.995)

nr=nrow(dados)
p=3
Vr=3*p/nr

#y=c(bd$Expenditure)[-50]
#x1=c(bd$Income)[-50]/10000
#x2=c(bd$Income^2)[-50]/10000
#x2=x1^2

X=cbind(1,x1,x2)

invX = solve(t(X)%*%X)
invXX = invX%*%t(X)
Xinv = X%*%invX


# matrix de alavanca
hvalues = hat(X)
hmean = mean(hvalues)
sort(hvalues)


# ajustando o modelo - lm
ajustelm=lm(y~x1+x2)
resid_lm= as.numeric(ajustelm$resid)
resid_lm2=(resid_lm)^2
slm<-summary(ajustelm)
# ajustando o modelo - lms
ajustelms=lqs(y~x1+x2, method = "lms")
resid_lms= as.numeric(ajustelms$resid)
resid_lms2=(resid_lms)^2

# ajustando o modelo - lts
ajustelts=lqs(y~x1+x2, method = "lts")
resid_lts= as.numeric(ajustelts$resid)
resid_lts2=(resid_lts)^2

# ajustando o modelo - lqs
ajustelqs=lqs(y~x1+x2, method = "lqs")
resid_lqs= as.numeric(ajustelqs$resid)
resid_lqs2=(resid_lqs)^2
```

```
# ajustando o modelo - lts
ajusteS=lqs(y~x1+x2, method = "S")
resid_S= as.numeric(ajusteS$resid)
resid_S2=(resid_S)^2


#####HC3
deltaHC3=2

meatHC3_lm=diag(resid_lm2/((1-hvalues)^deltaHC3))
HC3_lm = invXX%*%meatHC3_lm%*%Xinv

meatHC3_lts=diag(resid_lts2/((1-hvalues)^deltaHC3))
HC3_lts = invXX%*%meatHC3_lts%*%Xinv

meatHC3_lms=diag(resid_lms2/((1-hvalues)^deltaHC3))
HC3_lms = invXX%*%meatHC3_lms%*%Xinv

meatHC3_lqs=diag(resid_lqs2/((1-hvalues)^deltaHC3))
HC3_lqs = invXX%*%meatHC3_lqs%*%Xinv

meatHC3_S=diag(resid_S2/((1-hvalues)^deltaHC3))
HC3_S = invXX%*%meatHC3_S%*%Xinv


##HC4
deltaHC4 = min(4,(hvalues/hmean))

meatHC4_lm=diag(resid_lm2/((1-hvalues)^deltaHC4))
HC4_lm = invXX%*%meatHC4_lm%*%Xinv

meatHC4_lts=diag(resid_lts2/((1-hvalues)^deltaHC4))
HC4_lts = invXX%*%meatHC4_lts%*%Xinv

meatHC4_lms=diag(resid_lms2/((1-hvalues)^deltaHC4))
HC4_lms = invXX%*%meatHC4_lms%*%Xinv

meatHC4_lqs=diag(resid_lqs2/((1-hvalues)^deltaHC4))
HC4_lqs = invXX%*%meatHC4_lqs%*%Xinv

meatHC4_S=diag(resid_S2/((1-hvalues)^deltaHC4))
HC4_S = invXX%*%meatHC4_S%*%Xinv

###HC4m
gamma1=1;gamma2=1.5
deltaHC4m=min(gamma1,(hvalues/hmean))+min(gamma2,
(hvalues/hmean))
```

```
meatHC4m_lm=diag(resid_lm2/((1-hvalues)^deltaHC4m))
HC4m_lm = invXX%*%meatHC4m_lm%*%Xinv

meatHC4m_lts=diag(resid_lts2/((1-hvalues)^deltaHC4m))
HC4m_lts = invXX%*%meatHC4m_lts%*%Xinv

meatHC4m_lms=diag(resid_lms2/((1-hvalues)^deltaHC4m))
HC4m_lms = invXX%*%meatHC4m_lms%*%Xinv

meatHC4m_lqs=diag(resid_lqs2/((1-hvalues)^deltaHC4m))
HC4m_lqs = invXX%*%meatHC4m_lqs%*%Xinv

meatHC4m_S=diag(resid_S2/((1-hvalues)^deltaHC4m))
HC4m_S = invXX%*%meatHC4m_S%*%Xinv

#HC5
k=0.7;hmax=max(hvalues)
deltaHC5<-min((hvalues/hmean),max(4,k*hmax/hmean))

meatHC5_lm=diag(resid_lm^2/((1-hvalues)^(deltaHC5/2)))
HC5_lm = invXX%*%meatHC5_lm%*%Xinv

meatHC5_lts=diag(resid_lts^2/((1-hvalues)^(deltaHC5/2)))
HC5_lts = invXX%*%meatHC5_lts%*%Xinv

meatHC5_lms=diag(resid_lms^2/((1-hvalues)^(deltaHC5/2)))
HC5_lms = invXX%*%meatHC5_lms%*%Xinv

meatHC5_lqs=diag(resid_lqs^2/((1-hvalues)^(deltaHC5/2)))
HC5_lqs = invXX%*%meatHC5_lqs%*%Xinv

meatHC5_S=diag(resid_S^2/((1-hvalues)^(deltaHC5/2)))
HC5_S = invXX%*%meatHC5_S%*%Xinv




##ERRO PADRÃO

summary(ajustelm) #Erro padrÃ$o

#HC3

eplmHC3b0=sqrt(HC3_lm[1,1])
eplmHC3b1=sqrt(HC3_lm[2,2])
eplmHC3b2=sqrt(HC3_lm[3,3])
```

```
eplmsHC3b0=sqrt(HC3_lms[1,1])
eplmsHC3b1=sqrt(HC3_lms[2,2])
eplmsHC3b2=sqrt(HC3_lms[3,3])

epltsHC3b0=sqrt(HC3_lts[1,1])
epltsHC3b1=sqrt(HC3_lts[2,2])
epltsHC3b2=sqrt(HC3_lts[3,3])

eplqsHC3b0=sqrt(HC3_lqs[1,1])
eplqsHC3b1=sqrt(HC3_lqs[2,2])
eplqsHC3b2=sqrt(HC3_lqs[3,3])

epSHC3b0=sqrt(HC3_S[1,1])
epSHC3b1=sqrt(HC3_S[2,2])
epSHC3b2=sqrt(HC3_S[3,3])

#Hc4
eplmHC4b0=sqrt(HC4_lm[1,1])
eplmHC4b1=sqrt(HC4_lm[2,2])
eplmHC4b2=sqrt(HC4_lm[3,3])

eplmsHC4b0=sqrt(HC4_lms[1,1])
eplmsHC4b1=sqrt(HC4_lms[2,2])
eplmsHC4b2=sqrt(HC4_lms[3,3])

epltsHC4b0=sqrt(HC4_lts[1,1])
epltsHC4b1=sqrt(HC4_lts[2,2])
epltsHC4b2=sqrt(HC4_lts[3,3])

eplqsHC4b0=sqrt(HC4_lqs[1,1])
eplqsHC4b1=sqrt(HC4_lqs[2,2])
eplqsHC4b2=sqrt(HC4_lqs[3,3])

epSHC4b0=sqrt(HC4_S[1,1])
epSHC4b1=sqrt(HC4_S[2,2])
epSHC4b2=sqrt(HC4_S[3,3])

#HC4m

eplmHC4mb0=sqrt(HC4m_lm[1,1])
eplmHC4mb1=sqrt(HC4m_lm[2,2])
eplmHC4mb2=sqrt(HC4m_lm[3,3])

eplmsHC4mb0=sqrt(HC4m_lms[1,1])
eplmsHC4mb1=sqrt(HC4m_lms[2,2])
eplmsHC4mb2=sqrt(HC4m_lms[3,3])

epltsHC4mb0=sqrt(HC4m_lts[1,1])
```

```
epltsHC4mb1=sqrt(HC4m_lts[2,2])
epltsHC4mb2=sqrt(HC4m_lts[3,3])

eplqsHC4mb0=sqrt(HC4m_lqs[1,1])
eplqsHC4mb1=sqrt(HC4m_lqs[2,2])
eplqsHC4mb2=sqrt(HC4m_lqs[3,3])

epSHC4mb0=sqrt(HC4m_S[1,1])
epSHC4mb1=sqrt(HC4m_S[2,2])
epSHC4mb2=sqrt(HC4m_S[3,3])

#Hc5

eplmHC5b0=sqrt(HC5_lm[1,1])
eplmHC5b1=sqrt(HC5_lm[2,2])
eplmHC5b2=sqrt(HC5_lm[3,3])

eplmsHC5b0=sqrt(HC5_lms[1,1])
eplmsHC5b1=sqrt(HC5_lms[2,2])
eplmsHC5b2=sqrt(HC5_lms[3,3])

epltsHC5b0=sqrt(HC5_lts[1,1])
epltsHC5b1=sqrt(HC5_lts[2,2])
epltsHC5b2=sqrt(HC5_lts[3,3])

eplqsHC5b0=sqrt(HC5_lqs[1,1])
eplqsHC5b1=sqrt(HC5_lqs[2,2])
eplqsHC5b2=sqrt(HC5_lqs[3,3])

epSHC5b0=sqrt(HC5_S[1,1])
epSHC5b1=sqrt(HC5_S[2,2])
epSHC5b2=sqrt(HC5_S[3,3])


# eplmHC5mb0=sqrt(HC5m_lm[1,1])
# eplmHC5mb1=sqrt(HC5m_lm[2,2])
#eplmHC5mb2=sqrt(HC5m_lm[3,3])


M=c("OLS","HC3","HC4","HC4m","HC5","HC5m")
MOLSlm=c(ajustelm$coefficients[1],ajustelm$coefficients[2]
,ajustelm$coefficients[3])
MHC3lm=c(round(eplmHC3b0,4),round(eplmHC3b1,4),
round(eplmHC3b2,4))
MHC4lm=c(round(eplmHC4b0,4),round(eplmHC4b1,4),
round(eplmHC4b2,4))
MHC4mlm=c(round(eplmHC4mb0,4),round(eplmHC4mb1,4),
round(eplmHC4mb2,4))
```

```
MHC5lm=c(round(eplmHC5b0,4),round(eplmHC5b1,4),
round(eplmHC5b2,4))

MHC3lms=c(round(eplmsHC3b0,4),round(eplmsHC3b1,4),
round(eplmsHC3b2,4))
MHC4lms=c(round(eplmsHC4b0,4),round(eplmsHC4b1,4),
round(eplmsHC4b2,4))
MHC4mlms=c(round(eplmsHC4mb0,4),round(eplmsHC4mb1,4),
round(eplmsHC4mb2,4))
MHC5lms=c(round(eplmsHC5b0,4),round(eplmsHC5b1,4),
round(eplmsHC5b2,4))

MHC3lts=c(round(epltsHC3b0,4),round(epltsHC3b1,4),
round(epltsHC3b2,4))
MHC4lts=c(round(epltsHC4b0,4),round(epltsHC4b1,4),
round(epltsHC4b2,4))
MHC4mlts=c(round(epltsHC4mb0,4),round(epltsHC4mb1,4),
round(epltsHC4mb2,4))
MHC5lts=c(round(epltsHC5b0,4),round(epltsHC5b1,4),
round(epltsHC5b2,4))

MHC3lqs=c(round(eplqsHC3b0,4),round(eplqsHC3b1,4),
round(eplqsHC3b2,4))
MHC4lqs=c(round(eplqsHC4b0,4),round(eplqsHC4b1,4),
round(eplqsHC4b2,4))
MHC4mlqs=c(round(eplqsHC4mb0,4),round(eplqsHC4mb1,4),
round(eplqsHC4mb2,4))
MHC5lqs=c(round(eplqsHC5b0,4),round(eplqsHC5b1,4),
round(eplqsHC5b2,4))

MHC3S=c(round(epSHC3b0,4),round(epSHC3b1,4),
round(epSHC3b2,4))
MHC4S=c(round(epSHC4b0,4),round(epSHC4b1,4),
round(epSHC4b2,4))
MHC4mS=c(round(epSHC4mb0,4),round(epSHC4mb1,4),
round(epSHC4mb2,4))
MHC5S=c(round(epSHC5b0,4),round(epSHC5b1,4),
round(epSHC5b2,4))



E=c("SQRTvarb0","SQRTvarb1","SQRTvarb2")
lm=rep("OLS",3)
lms=rep("LMS",3)
lts=rep("LTS",3)
lqs=rep("LQS",3)
S=rep("S",3)
```

```
TR20 = cbind(E,lm,MHC3lm,MHC4lm,MHC4mlm,MHC5lm)
TR21 = cbind(E,lms,MHC3lms,MHC4lms,MHC4mlms,MHC5lms)
TR22 = cbind(E,lts,MHC3lts,MHC4lts,MHC4mlts,MHC5lts)
TR23 = cbind(E,lqs,MHC3lqs,MHC4lqs,MHC4mlqs,MHC5lqs)
TR24 = cbind(E,S,MHC3S,MHC4S,MHC4mS,MHC5S)


TR2=rbind(TR20,TR21,TR22,TR23,TR24)
colnames(TR2)<-c("Estimador", M)
TR2

xtable(TR2)


#################################




###TESTE QUASI-T
summary(ajustelm) #p-valor
coeflm1 =  ajustelm$coef[3]

# considerando o estimador HC3
testOLSHC3 = coeflm1 / eplmHC3b2
# considerando o estimador HC4
testOLSHC4 = coeflm1/ eplmHC4b2
# considerando o estimador HC4m
testOLSHC4m = coeflm1/ eplmHC4mb2
# considerando o estimador HC5
testOLSHC5 = coeflm1/ eplmHC5b2


# considerando o estimador HC3
testLMSHC3 = coeflm1 / eplmsHC3b2
# considerando o estimador HC4
testLMSHC4 = coeflm1/ eplmsHC4b2
# considerando o estimador HC4m
testLMSHC4m = coeflm1/ eplmsHC4mb2
# considerando o estimador HC5
testLMSHC5 = coeflm1/ eplmsHC5b2

# considerando o estimador HC3
testLTSHC3 = coeflm1 / epltsHC3b2
# considerando o estimador HC4
testLTSHC4 = coeflm1/ epltsHC4b2
# considerando o estimador HC4m
```

```
testLTSHC4m = coeflm1/ epltsHC4mb2
# considerando o estimador HC5
testLTSHC5 = coeflm1/ epltsHC5b2

# considerando o estimador HC3
testLQSHC3 = coeflm1 / eplqsHC3b2
# considerando o estimador HC4
testLQSHC4 = coeflm1/ eplqsHC4b2
# considerando o estimador HC4m
testLQSHC4m = coeflm1/ eplqsHC4mb2
# considerando o estimador HC5
testLQSHC5 = coeflm1/ eplqsHC5b2


# considerando o estimador HC3
testSHC3 = coeflm1 / epSHC3b2
# considerando o estimador HC4
testSHC4 = coeflm1/ epSHC4b2
# considerando o estimador HC4m
testSHC4m = coeflm1/ epSHC4mb2
# considerando o estimador HC5
testSHC5 = coeflm1/ epSHC5b2


p1lm=1-pchisq((testOLSHC3)^2,1)
p2lm=1-pchisq((testOLSHC4)^2,1)
p3lm=1-pchisq((testOLSHC4m)^2,1)
p4lm=1-pchisq((testOLSHC5)^2,1)
#p5lm=pnorm(testOLSHC5m,lower.tail = F)

p1lms=1-pchisq((testLMSHC3)^2,1)
p2lms=1-pchisq((testLMSHC4)^2,1)
p3lms=1-pchisq((testLMSHC4m)^2,1)
p4lms=1-pchisq((testLMSHC5)^2,1)


p1lts=1-pchisq((testLTSHC3)^2,1)
p2lts=1-pchisq((testLTSHC4)^2,1)
p3lts=1-pchisq((testLTSHC4m)^2,1)
p4lts=1-pchisq((testLTSHC5)^2,1)

p1lqs=1-pchisq((testLQSHC3)^2,1)
p2lqs=1-pchisq((testLQSHC4)^2,1)
p3lqs=1-pchisq((testLQSHC4m)^2,1)
p4lqs=1-pchisq((testLQSHC5)^2,1)

p1S=1-pchisq((testSHC3)^2,1)
p2S=1-pchisq((testSHC4)^2,1)
```

```
p3S=1-pchisq((testSHC4m)^2,1)
p4S=1-pchisq((testSHC5)^2,1)


M1=c("HC3","HC4","HC4m","HC5")
M20=c(round(p1lm[[1]],4),round(p2lm[[1]],4),
round(p3lm[[1]],4),round(p4lm[[1]],4))
M21=c(round(p1lms[[1]],4),round(p2lms[[1]],4),
round(p3lms[[1]],4),round(p4lms[[1]],4))
M22=c(round(p1lts[[1]],4),round(p2lts[[1]],4),
round(p3lts[[1]],4),round(p4lts[[1]],4))
M23=c(round(p1lqs[[1]],4),round(p2lqs[[1]],4),
round(p3lqs[[1]],4),round(p4lqs[[1]],4))
M24=c(round(p1S[[1]],4),round(p2S[[1]],4),
round(p3S[[1]],4),round(p4S[[1]],4))



TR3 = cbind(M1,M20,M21,M22,M23,M24)

colnames(TR3)<-c("teste", "p-valor OLS","p-valor LMS",
"p-valor LTS",
"p-valor LQS","p-valor S")
round(TR3, 4)

xtable(TR3)



#####Pontos de alavanca
X=cbind(1,x1,x2)
hvalues = hat(X)
sort(hvalues)


##################################################
rm(list=ls())
library(MASS)
library(car)  # hc
library(sandwich)
library(xtable)
#library(xlsx)
#library(xtable)

dados=read.table(file = "TCC/marinha.txt", header=F, dec=".")

y=dados[,3]
x1=dados[,1]
```

```
x2=dados[,2]

X=cbind(1,x1,x2)
hvalues = hat(X)
sort(hvalues)

LM=LMS=LTS=LQS=S=matrix(c(0),nrow=25,ncol=4)

# ajustando o modelo - lm
ajustelm=lm(y~x1+x2)
# ajustando o modelo - lms
ajustelms=lqs(y~x1+x2, method = "lms")
# ajustando o modelo - lts
ajustelts=lqs(y~x1+x2, method = "lts")
# ajustando o modelo - lqs
ajustelqs=lqs(y~x1+x2, method = "lqs")
# ajustando o modelo - lts
ajusteS=lqs(y~x1+x2, method = "S")

Matriz=matrix(c(ajustelm$coefficients[1],ajustelm$coefficients[2],
ajustelm$coefficients[3],
ajustelms$coefficients[1],ajustelms$coefficients[2],
ajustelms$coefficients[3],
ajustelts$coefficients[1],ajustelts$coefficients[2],
ajustelts$coefficients[3],
ajustelqs$coefficients[1],ajustelqs$coefficients[2],
ajustelqs$coefficients[3],
ajusteS$coefficients[1],ajusteS$coefficients[2],
ajusteS$coefficients[3]),
ncol=3, byrow = T)
dimnames(Matriz) = list(c("LM","LMS","LTS","LQS","S"),
c("b0","b1","b2"))
round(Matriz,4)
xtable(round(Matriz,4))


for(i in 1:length(y)){
y1<-y[-i]
x11<-x1[-i]
x22<-x2[-i]
ajustelm1=lm(y1~x11+x22)
ajustelms1=lqs(y1~x11+x22, method = "lms")
ajustelts1=lqs(y1~x11+x22, method = "lts")
ajustelqs1=lqs(y1~x11+x22, method = "lqs")
ajusteS1=lqs(y1~x11+x22, method = "S")
LM[i,1]=LMS[i,1]=LTS[i,1]=LQS[i,1]=S[i,1]=i

LM[i,2]=ajustelm1$coefficients[1]
```

```
LM[i,3]=ajustelm1$coefficients[2]
LM[i,4]=ajustelm1$coefficients[3]


LMS[i,2]=ajustelms1$coefficients[1]
LMS[i,3]=ajustelms1$coefficients[2]
LMS[i,4]=ajustelms1$coefficients[3]


LTS[i,2]=ajustelts1$coefficients[1]
LTS[i,3]=ajustelts1$coefficients[2]
LTS[i,4]=ajustelts1$coefficients[3]


LQS[i,2]=ajustelqs1$coefficients[1]
LQS[i,3]=ajustelqs1$coefficients[2]
LQS[i,4]=ajustelqs1$coefficients[3]


S[i,2]=ajusteS1$coefficients[1]
S[i,3]=ajusteS1$coefficients[2]
S[i,4]=ajusteS1$coefficients[3]


}
dimnames(LM) =dimnames(LMS)= dimnames(LTS) = dimnames(LQS)=
dimnames(S) = list(c(1:25),c("Obs. retirada","b0","b1","b2"))



#beta0
a=cbind(1:25,round(LM[,2],4),round(LMS[,2],4),round(LTS[,2],4),
round(LQS[,2],4),round(S[,2],4))
colnames(a)=c("i", "OLS","LMS", "LTS", "LQS", "S")
xtable(a)

#beta1
b=cbind(1:25,round(LM[,3],4),round(LMS[,3],4),round(LTS[,3],4),
round(LQS[,3],4),round(S[,3],4))
colnames(b)=c("i", "OLS","LMS", "LTS", "LQS", "S")
xtable(b)

#beta2
c=cbind(1:25,round(LM[,4],4),round(LMS[,4],4),round(LTS[,4],4),
round(LQS[,4],4),round(S[,4],4))
colnames(b)=c("i", "OLS","LMS", "LTS", "LQS", "S")
xtable(c)
```