Robust tests for linear regression models based on τ-estimates

Matias Salibian-Barrera[a,1], Stefan Van Aelst[c,1], Victor Yohai[b,1]

[a]*Department of Statistics, University of British Columbia, Vancouver, BC, Canada.*
[b]*Department of Mathematics, University of Buenos Aires*
[c]*Department of Mathematics, KU Leuven, Belgium and Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium.*

Abstract

ANOVA tests are the standard tests to compare nested linear models fitted by least squares. These tests are equivalent to likelihood ratio tests, so they have high power. However, least squares estimators are very vulnerable to outliers in the data, and thus the related ANOVA type tests are also extremely sensitive to outliers. Therefore, robust estimators can be considered to obtain a robust alternative to the ANOVA tests. Regression τ-estimators combine high robustness with high efficiency which makes them suitable for robust inference beyond parameter estimation. Robust likelihood ratio type test statistics based on the τ-estimates of the error scale in the linear model are a natural alternative to the classical ANOVA tests. The higher efficiency of the τ-scale estimates compared with other robust alternatives is expected to yield tests with good power. Their null distribution can be estimated using either an asymptotic approximation or the fast and robust bootstrap. The robustness and power of the resulting robust likelihood ratio type tests for nested linear models is studied.

Keywords: robust statistics, robust tests, linear regression

1. Introduction

An important step in regression analysis is determining which of the available explanatory variables are relevant in the proposed model. One approach is to test whether some of the regression coefficients are different from zero or not. The standard test for linear hypotheses of this type is the well-known F-test based on least squares estimates. It is also the likelihood ratio test when the errors are normally distributed. Unfortunately, small deviations from this assumption may seriously affect both the least squares estimates and the corresponding F-test, invalidating the resulting inference conclusions. Such small perturbations in the data are very common in real applications, and as the number of variables and the complexity of the models increase, they become much more difficult to detect using diagnostic methods based on non-robust estimators.

To overcome this problem robust estimators have been proposed and studied extensively in the literature. These estimators yield reliable point estimates for the model parameters even when the ideal distributional assumptions are not satisfied. Robustness properties of such estimators have been investigated via their influence function and breakdown point (see e.g. Hampel et al. 1986, Maronna et al. 2006). The influence function provides information on the effect of a small amount of contamination on the estimator while, intuitively speaking, the breakdown point is the largest fraction of arbitrary contamination that can be present in the data without driving the bias of the estimator to infinity. Another robustness criterion to compare estimators is the maximum asymptotic bias which measures the effect of a positive (non-infinitesimal) fraction of contamination (Martin *et al.*, 1989; Berrendero *et al.*, 2007). Robust high-breakdown estimators of the regression parameters include the least median of squares and least trimmed squares estimators (Rousseeuw 1984), S-estimators (Rousseeuw and Yohai 1984), MM-estimators (Yohai 1987), τ-estimators (Yohai and Zamar 1988) and CM-estimators (Mendes and Tyler 1996).

In this paper we consider the problem of performing inference for a linear regression model using robust estimators. Specifically, we are interested in obtaining robust and efficient tests for linear hypotheses on the regression

Email address: matias@stat.ubc.ca (Matias Salibian-Barrera)

coefficients. Robust hypothesis tests have received much less attention in the literature than point estimators. A natural approach to obtain robust tests is to use a robust point estimator of the model parameters. Robust Wald-, scores- and likelihood-ratio-type tests based on M- and GM-estimators have been proposed in the literature by Markatou and Hettmansperger (1990), Markatou, Stahel and Ronchetti (1991), Markatou and He (1994), and Heritier and Ronchetti (1994). Unfortunately, the breakdown point of these estimators is less than $1/p$, where p is the number of regression coefficients. In other words, the more explanatory variables in the model the less robust these estimators are. This lack of robustness in turn affects the associated test statistics.

Alternatively, one can note that the classical F-test compares the residual sum of squares obtained under the null and alternative hypotheses. Since the residual sum of squares can also be thought of as (non-robust) residual scale estimates, it is natural to consider test statistics of the form

$$\left(\frac{\hat{\sigma}_0^2 - \hat{\sigma}_a^2}{\hat{\sigma}_a^2} \right) \tag{1}$$

where $\hat{\sigma}_0$ is a robust residual scale estimate obtained under the null hypothesis, and $\hat{\sigma}_a$ is the scale estimate for the unrestricted model. τ-estimators (Yohai and Zamar, 1988) are a class of high-breakdown and highly efficient regression estimators that are naturally accompanied by an associated estimator of the error scale which is also highly robust and highly efficient. This is an advantage compared to other classes of high-breakdown, highly efficient regression estimators such as MM-estimators or CM-estimators. The τ-estimators are defined as the minimizers of a robust and efficient scale estimator of the regression errors. These estimators can be tuned to simultaneously have a high-breakdown point (50%) and achieve high-efficiency (e.g. 85% or 95%) at the central model with normal errors. Good robustness properties of τ-estimators have been shown for both the estimator of the regression coefficients (Berrendero and Zamar, 2001) and the estimator of the error scale (Van Aelst, Willems and Zamar, 2013). We expect that the good robustness properties of the τ-scale estimates compared with other scale estimators to yield tests with good robustness properties as well. Until recently, the main drawback of τ-estimators was the lack of a good algorithm for their computation. However, Salibian-Barrera, Willems and Zamar (2008) proposed an efficient algorithm for these estimators and implementations in `R` and `MATLAB` / `OCTAVE` are publicly available on-line at `http://www.stat.ubc.ca/~matias`.

In what follows we study ANOVA-type tests of the intuitively appealing test in (1) using τ-scale estimators which we call ANOVA τ-tests. We show that under certain regularity conditions the test statistics proposed in this paper are asymptotically central chi-squared distributed under the null hypothesis, and non-central chi-squared distributed under sequences of contiguous alternatives. Note that these ANOVA-type test statistics thus have a much simpler asymptotic distribution than several robust likelihood ratio type test statistics based on M-estimators whose asymptotic distribution is a linear combination of χ_1 distributions (Ronchetti 1982, Heritier and Ronchetti, 1994). Furthermore, we derive the influence functions of these tests, which show that the tests are robust against vertical outliers and bad leverage points, although good leverage points may have a larger influence on the test statistic and corresponding level and power.

Since the finite-sample distribution of test statistics of this form is generally unknown, p-values are usually approximated using the asymptotic distribution of the test statistic. However, numerical experiments show that in some cases these approximations are reliable only for relatively large sample sizes. Moreover, some of the required regularity assumptions may not hold when the data contain outliers, which compromises the validity of the asymptotic approximation. To obtain better p-value approximations one can consider using the bootstrap (Efron, 1979). However, bootstrapping robust estimators when the data may contain outliers presents two important challenges. First, re-calculating many times the estimator is highly computationally demanding. Second, the bootstrap results may not be reliable due to bootstrap samples containing many more outliers than the original sample. In fact, this might cause the bootstrapped estimate to break down even if the original estimate did not. Salibian-Barrera and Zamar (2002) proposed a fast and robust bootstrap (FRB) method for MM-regression estimates that solves both of these problems by calculating a fast approximation to the bootstrapped MM-estimators. This FRB method has been extended to other settings (see e.g. Van Aelst and Willems, 2005; Salibian-Barrera, Van Aelst and Willems, 2006; and Samanta and Welsh, 2013).

In this paper we also extend the FRB methodology to the class of τ-estimators. Salibian-Barrera (2005) showed that the FRB also works well as a way to obtain p-value estimates for robust scores-type test statistics. However, because the likelihood ratio type tests have a higher order of convergence than the scores-type tests, the consistency of

the FRB estimator for the null distribution of the robust likelihood ratio type tests proposed here needs to be studied carefully. This problem is discussed in Section 3, where we show that the test statistics satisfy a sufficient condition given in Van Aelst and Willems (2011) that guarantees that the FRB is a consistent estimator for their null distribution. This result allows us to propose a computationally feasible and reliable p-value estimate based on the FRB that is consistent under weaker regularity assumptions than those required by the corresponding asymptotic approximation. Our simulation results show that, regardless of the presence of outliers in the sample, the FRB approximation yields tests with empirical levels closer to the nominal one. The advantage of using the FRB approximation is more noticeable for smaller sample sizes.

The rest of the paper is organized as follows. In Section 2 we introduce the ANOVA-type tests based on τ-estimators and derive their asymptotic distribution and robustness properties. A fast and robust bootstrap estimator for the null distribution of the test statistics is discussed in Section 3, where we also show that this bootstrap approximation is consistent. In Section 4 we investigate the behavior of the ANOVA τ-tests by means of simulations, while Section 5 illustrates our proposal on a real data set. Concluding remarks are included in Section 6 while all technical details and proofs are relegated to the Appendix.

2. ANOVA tests based on τ-estimators

In what follows let $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, denote a random sample, with $y_i \in \mathbb{R}$ and $\mathbf{x}_i \in \mathbb{R}^p$. The model of interest is

$$y_i \;=\; \boldsymbol{\theta}_0' \mathbf{x}_i + \epsilon_i\,, \quad i = 1, \ldots, n\,, \tag{2}$$

where $\boldsymbol{\theta}_0 \in \mathbb{R}^p$ denotes the vector of regression coefficients. The errors ϵ_i are assumed to be iid according to a symmetric distribution F with center zero and scale σ_0, and independent from the covariates $\mathbf{x}_i$. The distribution F is often assumed to be Gaussian. We consider regression τ-estimators (Yohai and Zamar, 1988), which are defined by

$$\hat{\boldsymbol{\theta}}_n \;=\; \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \tau_n^2(\boldsymbol{\theta})\,, \tag{3}$$

where, for each $\boldsymbol{\theta} \in \mathbb{R}^p$,

$$\tau_n^2(\boldsymbol{\theta}) \;=\; s_n^2(\boldsymbol{\theta})\, \frac{1}{b_2\, n} \sum_{i=1}^{n} \rho_2\left(\frac{r_i(\boldsymbol{\theta})}{s_n(\boldsymbol{\theta})}\right), \tag{4}$$

and $s_n(\boldsymbol{\theta})$ is an M-scale estimator of the residuals $r_i(\boldsymbol{\theta}) = y_i - \boldsymbol{\theta}'\mathbf{x}_i$, $1 \le i \le n$, satisfying

$$\frac{1}{n} \sum_{i=1}^{n} \rho_1\left(\frac{r_i(\boldsymbol{\theta})}{s_n(\boldsymbol{\theta})}\right) \;=\; b_1\,. \tag{5}$$

The parameters b_1 and b_2 are chosen such that $E_\Phi\left(\rho_j(u)\right) = b_j$, $j = 1, 2$, where Φ denotes the standard normal distribution. These conditions ensure consistency of the scale estimators $\hat{\tau}_n = \tau_n(\hat{\boldsymbol{\theta}}_n)$ and $\hat{\sigma}_n = s_n(\hat{\boldsymbol{\theta}}_n)$ for normally distributed errors. The functions ρ_j, $j = 1, 2$, are supposed to be continuous, non-decreasing and even on the positive real numbers, with $\rho_j(0) = 0$, $j = 1, 2$. Furthermore, if $a = \sup_u \rho_j(u)$, then $0 < a < \infty$ and, for $0 \le u < v$ such that $\rho_j(u) < a$ we have $\rho_j(u) < \rho_j(v)$. Following Maronna, Martin and Yohai (2006), we refer to functions that satisfy these conditions as "ρ-functions".

Note that the choice of loss functions ρ_1 and ρ_2 can have important practical and theoretical consequences (see e.g. Berrendero and Zamar (2001), Van Aelst, Willems and Zamar (2013)). In this paper we use the family of functions proposed in Muler and Yohai (2002), which has been shown to yield τ-estimators with better robustness properties than those obtained when using the Tukey bisquare family (Van Aelst *et al.*, 2013). This family is indexed with a tuning parameter c and given by

$$\rho_c(t) = \begin{cases} 1.38\left(\frac{t}{c}\right)^2 & |\frac{t}{c}| \le \frac{2}{3} \\ 0.55 - 2.69\left(\frac{t}{c}\right)^2 + 10.76\left(\frac{t}{c}\right)^4 - 11.66\left(\frac{t}{c}\right)^6 + 4.04\left(\frac{t}{c}\right)^8 & \frac{2}{3} < |\frac{t}{c}| \le 1 \\ 1, & |\frac{t}{c}| > 1. \end{cases}$$

To obtain consistency and a maximal break-down point, we select the tuning parameters $c_1 = 1.214$ and $b_1 = 0.5$ for $\rho_1 = \rho_{c_1}$ in (5). The choice $c_2 = 3.270$ and $b_2 = 0.128$ for $\rho_2 = \rho_{c_2}$ results in an estimator with 95% efficiency compared to the least-squares estimator when the errors in (2) are normally distributed (see Yohai and Zamar, 1988). In this case, the associated τ-scale estimate $\tau_n(\boldsymbol{\theta}_n)$ for the residual scale has 50% breakdown point and a Gaussian efficiency of 97.7%.

Yohai and Zamar (1998) showed that any local minimum $\hat{\boldsymbol{\theta}}_n$ of (3) satisfies

$$\sum_{i=1}^{n} \left[W_n(\hat{\boldsymbol{\theta}}_n)\psi_1\left(\tilde{r}_i(\hat{\boldsymbol{\theta}}_n)\right) + \psi_2\left(\tilde{r}_i(\hat{\boldsymbol{\theta}}_n)\right) \right] \mathbf{x}_i \;=\; \mathbf{0}\,, \tag{6}$$

where $\psi_j = \rho_j'$, $j = 1, 2$, $\tilde{r}_i(\hat{\boldsymbol{\theta}}_n) = (y_i - \mathbf{x}_i'\hat{\boldsymbol{\theta}}_n)/\hat{\sigma}_n$, $1 \le i \le n$, with $\hat{\sigma}_n = s_n(\hat{\boldsymbol{\theta}}_n)$, and

$$W_n(\hat{\boldsymbol{\theta}}_n) \;=\; \frac{\sum_{i=1}^{n} \left[2\,\rho_2(\tilde{r}_i(\hat{\boldsymbol{\theta}}_n)) - \psi_2(\tilde{r}_i(\hat{\boldsymbol{\theta}}_n))\,\tilde{r}_i(\hat{\boldsymbol{\theta}}_n) \right]}{\sum_{i=1}^{n} \psi_1(\tilde{r}_i(\hat{\boldsymbol{\theta}}_n))\,\tilde{r}_i(\hat{\boldsymbol{\theta}}_n)} \tag{7}$$

An efficient algorithm to solve (3) have been developed by Salibian-Barrera, Willems and Zamar (2008). `R` and `MATLAB` / `OCTAVE` code can be obtained from `http://www.stat.ubc.ca/~matias`.

In this paper we study tests for linear hypotheses on the vector of regression parameters $\boldsymbol{\theta}_0$ in (2). In general, these hypotheses can be written in the following form:

$$H_0 : \boldsymbol{\theta}_0 \in V \qquad \text{vs} \qquad H_a : \boldsymbol{\theta}_0 \notin V\,, \tag{8}$$

where $V \subset \mathbb{R}^p$ denotes a linear subspace of dimension $q < p$. Equivalently (see Appendix), we may consider hypotheses of the form

$$H_0 : \boldsymbol{\theta}_{(2)} = \mathbf{0} \qquad \text{vs} \qquad H_a : \boldsymbol{\theta}_{(2)} \neq \mathbf{0}\,, \tag{9}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_{(1)}', \boldsymbol{\theta}_{(2)}')'$, with $\boldsymbol{\theta}_{(1)} \in \mathbb{R}^q$ and $\boldsymbol{\theta}_{(2)} \in \mathbb{R}^{p-q}$. The usual F-test rejects H_0 if the residual sum of squares under the null hypothesis is sufficiently larger than under the alternative. Noting that these sums (or averages) of squares can also be thought of as (non-robust) residual scale estimates, it is natural to replace the sums of squared residuals by robust scale estimators in order to obtain a robust test. More specifically, let

$$\hat{\boldsymbol{\theta}}_{n,0} \;=\; \arg\min_{\boldsymbol{\theta} \in V} \tau_n^2(\boldsymbol{\theta})\,,$$

be the τ-regression estimator for the null model, i.e. the restricted model under the null hypothesis, and let $\hat{\tau}_{n,0}^2 = \tau_n^2(\hat{\boldsymbol{\theta}}_{n,0})$ be the corresponding scale estimator. Similarly, let $\hat{\boldsymbol{\theta}}_n$ and $\hat{\tau}_n^2 = \tau_n^2(\hat{\boldsymbol{\theta}}_n)$ denote the estimators for the full model, i.e. the larger model under the alternative. To test the hypotheses in (8) consider the following ANOVA-type test statistic:

$$L_n(\,\hat{\boldsymbol{\theta}}_{n,0}\,, \hat{\boldsymbol{\theta}}_n\,) \;=\; n\left(\frac{\hat{\tau}_{n,0}^2 - \hat{\tau}_n^2}{\hat{\tau}_n^2}\right). \tag{10}$$

Note that this test statistic can immediately be obtained once the τ-estimates of both models have been calculated.

The next theorem shows that the asymptotic distribution of L_n is χ^2_{p-q} if the null hypothesis H_0 in (8) holds. Moreover, under a sequence of contiguous alternatives the asymptotic distribution becomes a non-central χ^2_{p-q} distribution.

Theorem 1 *Assume that ρ_1 and ρ_2 are ρ-functions, and that ρ_2 satisfies $2\rho_2(u) - u\rho_2'(u) \ge 0$. F, the distribution of the errors ϵ_i in* (2)*, is assumed to have a unimodal density function that is symmetric around zero. Let $s_0 \in (0, +\infty)$ be such that*

$$E_F\left(\rho_1\left(\frac{u}{s_0}\right)\right) \;=\; b_1\,,$$

with $b_1 \in (0, 1/2]$. For each $j = 1, 2$, let

$$B_j \;=\; E_F\left(\psi_j\left(\frac{u}{s_0}\right)\frac{u}{s_0}\right), \quad M_j \;=\; E_F\left(\rho_j\left(\frac{u}{s_0}\right)\right), \;\; \textit{and} \;\; D_j \;=\; E_F\left(\psi_j'\left(\frac{u}{s_0}\right)\right).$$

Moreover, define $W = (2M_2 - B_2)/B_1$, $\psi_0 = W\psi_1 + \psi_2$, *where* $\psi_j = \rho_j'$, $j = 1, 2$, *and let*

$$D_0 = E_F\left(\psi_0'\left(\frac{u}{s_0}\right)\right), \quad \text{and} \quad K_0 = E_F\left(\psi_0^2\left(\frac{u}{s_0}\right)\right).$$

Consider the hypotheses in (8)*, then for the test statistic* $L_n(\hat{\boldsymbol{\theta}}_{n,0}, \hat{\boldsymbol{\theta}}_n)$ *defined in* (10) *the following holds:*

(a) Under the null hypothesis, as $n \to \infty$*,*

$$\frac{(2D_0M_2)}{K_0} L_n \xrightarrow{\mathcal{D}} \chi^2_{p-q}. \tag{11}$$

(b) Let $\boldsymbol{\theta}_0 \in V$ *and* $\mathbf{a} \in V^+$*, the orthogonal complement of* V*, then under the sequence of contiguous alternative hypotheses*

$$H_n : \boldsymbol{\theta} = \boldsymbol{\theta}_0 + \frac{\mathbf{a}}{\sqrt{n}}. \tag{12}$$

it holds that, as $n \to \infty$

$$\frac{(2D_0M_2)}{K_0} L_n \xrightarrow{\mathcal{D}} \chi^2_{p-q}(\delta),$$

where the non-centrality parameter $\delta^2 = (\mathbf{a}'\Sigma\mathbf{a})/\vartheta$*, with* $\Sigma = E(\mathbf{X}\mathbf{X}')$ *and* $\vartheta = s_0^2 K_0/D_0^2$.

Note that the class of τ-estimators includes the S-estimators (Rousseeuw and Yohai, 1984), when we choose $\rho_1 = \rho_2$. Hence, the theorem above also provides the asymptotic distribution of the test statistic

$$SL_n = n\left(\frac{\hat{\mathbf{s}}^2_{n,0} - \hat{\mathbf{s}}^2_n}{\hat{\mathbf{s}}^2_n}\right), \tag{13}$$

where $\hat{\mathbf{s}}_{n,0}$ and $\hat{\mathbf{s}}_n$ denote the S-scale estimators of the models under the null and alternative hypotheses, respectively.

2.1. Robustness properties

To study the robustness properties of the ANOVA τ-tests, following Heritier and Ronchetti (1994) we examine the effect of a small amount of contamination on the asymptotic level of the test. Let us consider a statistical functional $\Lambda(H)$ corresponding to some test statistic Λ_n, i.e. $\Lambda_n = \Lambda(H_n)$ with H_n the empirical distribution of the data. Now, consider a sequence of contaminated distributions

$$H_{\epsilon,n} = (1 - \frac{\epsilon}{\sqrt{n}})H + \frac{\epsilon}{\sqrt{n}}G,$$

where H is the distribution of the vector $(y, \mathbf{x}')'$ when no outliers are present, and G is an arbitrary contaminating distribution. Following e.g. Heritier and Ronchetti (1994), Wang and Qu (2007), Van Aelst and Willems (2011), we obtain the following general result.

Theorem 2 *Consider the statistical functional* $\Lambda(H)$ *and let*

$$\xi_2(\mathbf{z}_1, \mathbf{z}_2) = \frac{\partial}{\partial\epsilon_1\partial\epsilon_2}\Lambda(H_{\epsilon_1,\mathbf{z}_1,\epsilon_2,\mathbf{z}_2})|_{\epsilon_1=0,\epsilon_2=0},$$

with $H_{\epsilon_1,\mathbf{z}_1,\epsilon_2,\mathbf{z}_2} = (1 - \epsilon_1 - \epsilon_2)H + \epsilon_1\Delta_{\mathbf{z}_1} + \epsilon_2\Delta_{\mathbf{z}_2}$ *where* $\mathbf{z}_i = (y_i, \mathbf{x}_i')'$*, and* $\Delta_{\mathbf{z}}$ *denotes a point-mass distribution at* $\mathbf{z}$*. Assume that*

$$n(\Lambda(H_{\epsilon,n}) - \Lambda_n) \xrightarrow{\mathcal{D}} \chi^2_q \tag{14}$$

uniformly over distributions in the contamination set $H_{\epsilon,n}$*. Denote the asymptotic level of the test based on* Λ_n *by* $\alpha(K)$ *when the underlying distribution is* K*, and denote the nominal level* $\alpha(H)$ *by* α_0*. Furthermore, denote by* $H_q(., \delta)$ *the cumulative distribution function of a* $\chi^2_q(\delta)$ *distribution, and by* $\eta_{1-\alpha_0}$ *the* $1 - \alpha_0$ *quantile of the central* χ^2_q *distribution. Then, we have*

$$\lim_{n\to\infty} \alpha(H_{\epsilon,n}) = \alpha_0 + \frac{\epsilon^2}{2}\kappa \iint \xi_2(\mathbf{z}_1, \mathbf{z}_2)\, dG(\mathbf{z}_1)\, dG(\mathbf{z}_2) + o(\epsilon^2), \tag{15}$$

where $\kappa = -(\partial/\partial\delta)H_q(\eta_{1-\alpha_0};\delta)|_{\delta=0}$. *For the special case of point-mass contamination* $G = \Delta_{y,\mathbf{x}}$ *this reduces to*

$$\lim_{n\to\infty} \alpha(H_{\epsilon,n}) = \alpha_0 + \epsilon^2 q\,\kappa\, SIF(\mathbf{x}, y, \Lambda, H) + o(\epsilon^2), \tag{16}$$

where $SIF(\mathbf{x}, y, \Lambda, H)$ *is the self standardized influence function of the test statistic* Λ *under the null hypothesis.*

Note that condition (14) is stronger than requiring the existence of the influence function of the test statistic, but is guaranteed for functionals that are Fréchet differentiable (see Heritier and Ronchetti 1994, Ronchetti and Trojani, 2001). Fréchet differentiability is fulfilled for M-estimators and τ-estimators, which are asymptotically equivalent to M-estimators, that satisfy conditions (A.1)-(A.9) of Heritier and Ronchetti (1994).

The self standardized influence function of a test statistic Λ corresponding to an asymptotically quadratic test is given by

$$\mathrm{SIF}(\mathbf{x}, y, \Lambda, H) \;=\; \mathrm{ASV}(\Lambda, H)^{-1}\,\mathrm{IF2}(\mathbf{x}, y, \Lambda, H)\,,$$

where $\mathrm{IF2}(\mathbf{x}, y, \Lambda, H)$ is the second-order influence function of the test statistic Λ, i.e.

$$\mathrm{IF2}(\mathbf{x}, y, \Lambda, H) \;=\; \frac{\partial^2}{\partial\epsilon^2}\Lambda(H_{\epsilon,y,\mathbf{x}})\bigg|_{\epsilon=0}\,,$$

with $H_{\epsilon,y,\mathbf{x}} = (1-\epsilon)H + \epsilon\Delta_{(y,\mathbf{x})}$ (see e.g. Croux et al. 2008). We can apply Theorem 2 to our test based on τ-scales by rescaling the test statistic L_n as in (11) so that the test has an asymptotic chi-square distribution with $p-q$ degrees of freedom, i.e. $\Lambda_n = d\,L_n$ with $d = 2\,D_0\,M_2/H_0$. It then follows that

$$\lim_{n\to\infty} \alpha(H_{\epsilon,n}) = \alpha_0 + \epsilon^2 d\,(p-q)\,\kappa\,\mathrm{SIF}(\mathbf{x}, y, L, H) + o(\epsilon^2) \tag{17}$$

Hence, the stability of the level of the test is guaranteed if the test statistic L has a bounded self-standardized influence function. A result similar to Theorem 2 can be obtained for the power of a test, which will be stable when the self-standardized influence function is bounded (see Ronchetti and Trojani, 2001).

Denote by $\tau_0(H)$ and $\tau(H)$ the functionals corresponding to the τ-scale estimators of the models under the null and alternative, respectively. Then, the corresponding test functional is given by $L(H) = (\tau_0(H)^2 - \tau(H)^2)/(\tau(H)^2)$. It can easily be shown that under the null hypothesis

$$\mathrm{SIF}(\mathbf{x}, y, L, H_0) \;=\; \frac{\psi_0^2(y - \boldsymbol{\theta}'_{(1)}\mathbf{x}_{(1)})}{(p-q)E[\psi_0^2]}\,(\mathbf{x}'\Sigma^{-1}\mathbf{x} - \mathbf{x}'_{(1)}\Sigma_{11}^{-1}\mathbf{x}_{(1)})\,, \tag{18}$$

where $\mathbf{x} = (\mathbf{x}'_{(1)}, \mathbf{x}'_{(2)})'$, with $\mathbf{x}_{(1)} \in \mathbb{R}^{p-q}$ and $\mathbf{x}_{(2)} \in \mathbb{R}^q$. It can immediately be seen that for bounded ψ_0 functions this self-standardized influence function is bounded in the response y, but not in $\mathbf{x}$. Hence, good leverage points may have a high influence on the test statistic and corresponding level and power of the test. However, the effect of vertical outliers and bad leverage points remains bounded. Figure 1 shows the effect of response outliers by plotting the residual part of the influence function (18), i.e. the factor $\psi_0^2(y - \boldsymbol{\theta}'_{(1)}\mathbf{x}_{(1)})/(E[\psi_0^2])$. The solid line corresponds to the 95% efficient τ-estimators while the dashed line corresponds to the 50% breakdown point S-estimator obtained by setting $\psi_0 = \psi_1$ in (18). These influence functions show that the redescending behavior of the estimators is passed on to the test statistics. Intermediate outliers can affect the test statistics to some extend, but far outliers are effectively downweighted and thus do not harm the test statistics. The higher efficiency of τ-estimators makes the corresponding test statistic more sensitive to intermediate outliers in the sense that their effect can be larger and outliers need to lie further away before their effect disappears.

3. Bootstrap

The asymptotic approximation to the null distribution of the test statistic given in Theorem 1 requires regularity conditions that may be hard to verify in practice, or that might be violated by the presence of outliers (e.g. the symmetric distribution of the errors). Furthermore, numerical experiments show that relatively large sample sizes are required for this approximation to be reliable (see Section 4). The bootstrap (Efron, 1979) provides an alternative

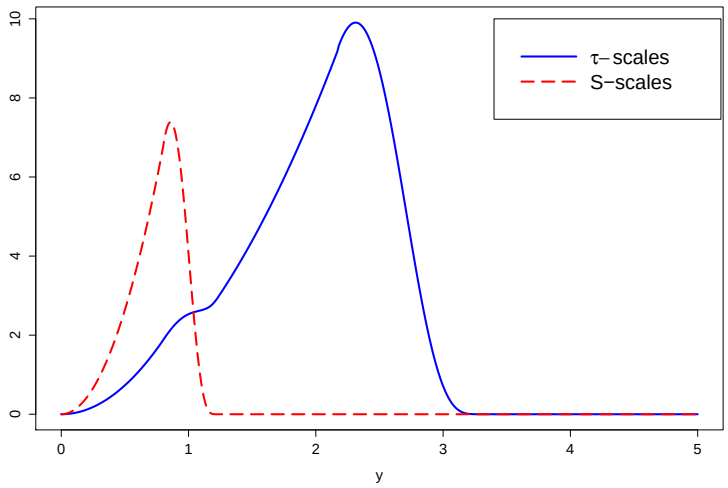


Figure 1: Response part of the standardized second-order influence function of the robust test statistic for a normal error distribution.

estimator for the null distribution of our test statistic that in many cases requires fewer regularity conditions to be consistent. However, the presence of outliers in the data may have a serious damaging effect on the usual bootstrap estimators. Moreover, the computational complexity of robust estimators might make the standard bootstrap unfeasible, particularly for moderate to high-dimensional problems. In this section we investigate the use of the fast and robust bootstrap (Salibian-Barrera and Zamar, 2002) to obtain approximate p-values for the ANOVA τ-tests discussed above. The fast and robust bootstrap (FRB) successfully overcomes both problems mentioned above. Although it was initially proposed for linear regression models, the FRB has been extended to several other models (Van Aelst and Willems, 2002, 2005; Salibian-Barrera, Van Aelst, 2008; Roelant, Van Aelst and Croux, 2009; Khan, Van Aelst and Zamar, 2010; and Samanta and Welsh, 2013).

The basic idea of the FRB can be outlined as follows. Let $\hat{\boldsymbol{\gamma}}_n$ denote a vector of estimated parameters. Many robust estimators satisfy a set of fixed-point equations of the form $\hat{\boldsymbol{\gamma}}_n = \mathbf{g}_n(\hat{\boldsymbol{\gamma}}_n)$, where the function $\mathbf{g}_n : \mathbb{R}^k \to \mathbb{R}^k$ typically depends on the sample. Given a bootstrap sample, the re-calculated estimator $\hat{\boldsymbol{\gamma}}_n^*$ solves $\hat{\boldsymbol{\gamma}}_n^* = \mathbf{g}_n^*(\hat{\boldsymbol{\gamma}}_n^*)$. To avoid solving this new system of nonlinear equations for each bootstrap sample, the FRB computes instead

$$\hat{\boldsymbol{\gamma}}_n^{R*} = \hat{\boldsymbol{\gamma}}_n + [\mathbf{I} - \nabla \mathbf{g}_n(\hat{\boldsymbol{\gamma}}_n)]^{-1} \left(\mathbf{g}_n^*(\hat{\boldsymbol{\gamma}}_n) - \hat{\boldsymbol{\gamma}}_n\right) . \tag{19}$$

The above equation is derived from a first-order Taylor approximation to the set of nonlinear equations that needs to be solved for each bootstrap sample. Note that the matrix $[\mathbf{I} - \nabla \mathbf{g}_n(\hat{\boldsymbol{\gamma}}_n)]^{-1}$ is computed only once with the original sample, and that evaluating $\mathbf{g}_n^*(\hat{\boldsymbol{\gamma}}_n)$ is typically very fast (it only requires finding the solution of a linear system of equations compared to solving a nonlinear optimization problem in p dimensions).

For τ-regression estimators, we set $\hat{\boldsymbol{\gamma}}_n = (\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n)$, then we obtain from (5) and (6) that

$$g_n(\boldsymbol{\theta}, \sigma) = \begin{pmatrix} \left(\sum_{i=1}^n \frac{\psi_n(\frac{r_i(\boldsymbol{\theta})}{\sigma})}{\frac{r_i(\boldsymbol{\theta})}{\sigma}} \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \sum_{i=1}^n \frac{\psi_n(\frac{r_i(\boldsymbol{\theta})}{\sigma})}{\frac{r_i(\boldsymbol{\theta})}{\sigma}} \mathbf{x}_i y_i \\ \sigma \frac{1}{b_1} \frac{1}{n} \sum_{i=1}^n \rho_1\left(\frac{r_i(\boldsymbol{\theta})}{\sigma}\right) \end{pmatrix}, \tag{20}$$

where $\psi_n = W_n \psi_1 + \psi_2$ and W_n is given in (7). Simple calculations show that

$$\mathbf{I} - \nabla g_n(\boldsymbol{\theta}, \sigma) = \begin{pmatrix} \mathbf{A}^{-1}\mathbf{B} & \mathbf{A}^{-1}\mathbf{v} \\ \mathbf{b}' & a \end{pmatrix}, \tag{21}$$

where

$$\mathbf{A} = \frac{1}{\sigma} \sum_{i=1}^n \frac{\psi_n(\frac{r_i(\boldsymbol{\theta})}{\sigma})}{\frac{r_i(\boldsymbol{\theta})}{\sigma}} \mathbf{x}_i \mathbf{x}_i' , \qquad \mathbf{B} = \frac{1}{\sigma} \sum_{i=1}^n \psi_n'(\frac{r_i(\boldsymbol{\theta})}{\sigma}) \mathbf{x}_i \mathbf{x}_i' ,$$

$$\mathbf{v} \;=\; \frac{1}{\sigma}\sum_{i=1}^{n}\frac{\psi_n(\frac{r_i(\boldsymbol{\theta})}{\sigma})}{\frac{r_i(\boldsymbol{\theta})}{\sigma}}\mathbf{x}_i\,, \qquad \mathbf{b} \;=\; \frac{1}{b_1}\,\frac{1}{n}\,\sum_{i=1}^{n}\psi_1(\frac{r_i(\boldsymbol{\theta})}{\sigma})\mathbf{x}_i\,,$$

and $a = (1/(b_1\, n))\ \sum_{i=1}^{n}\psi_1'(r_i(\boldsymbol{\theta})/\sigma)\, r_i(\boldsymbol{\theta})/\sigma$.

In order to show that the FRB can be used to find a consistent estimator of the null distribution of the test statistic L_n in (10), we first need to show that the FRB provides a consistent estimator for the distribution of the τ-regression estimators in (3).

Theorem 3 *Let $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ be a random sample satisfying the linear regression model in* (2)*. Let $\hat{\boldsymbol{\theta}}_n$ and $\hat{\sigma}_n$ be the τ-regression and scale estimators, respectively, defined in* (3) *and* (4)*, such that $\hat{\boldsymbol{\theta}}_n \xrightarrow[n\to\infty]{P} \boldsymbol{\theta}_0$ and $\hat{\sigma}_n \xrightarrow[n\to\infty]{P} \sigma_0$. Assume that ρ_j, $j = 1, 2$ are ρ-functions with continuous third derivatives, $E[\rho_1'(r)r] \neq 0$ and finite, and such that $\rho_j'(u)/u$, $j = 1, 2$ are continuous. Furthermore, assume that $E[\rho_0'(r)/r\,\mathbf{X}\mathbf{X}']$ is non-singular, and that the following quantities exist and are finite: $E[\rho_0''(r)\mathbf{X}\mathbf{X}']$, $E[\rho_0''(r)r\mathbf{X}]$, and $E[\rho_1'(r)\mathbf{X}]$; where $\rho_0(r) = W\rho_1(r) + \rho_2(r)$ and W is defined in Theorem 1. Then, along almost all sample sequences, the distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n)$ converges weakly to the same limit as $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$.*

In order to obtain correct estimates of the sampling distribution of the estimators under the null hypothesis, we follow Hall and Wilson (1990) and Fisher and Hall (1991). As before, let $\hat{\boldsymbol{\theta}}_{n,0}$ and $\hat{\boldsymbol{\theta}}_n$ be the regression estimators computed under the null and full model, respectively. Let $r_i^{(a)} = y_i - \mathbf{x}_i'\,\hat{\boldsymbol{\theta}}_n$ and $f_i^{(0)} = \mathbf{x}_i'\,\hat{\boldsymbol{\theta}}_{n,0}$, $i = 1, \ldots, n$, be the residuals in the full model and the fitted values in the null model, respectively. We build a data set that mimics samples from the null hypothesis as follows:

$$\tilde{y}_i = f_i^{(0)} + r_i^{(a)} = \mathbf{x}_i'\,\hat{\boldsymbol{\theta}}_{n,0} + r_i^{(a)}\,,\ i = 1, \ldots, n\,. \tag{22}$$

Note that these "null data" approximately satisfy the null hypothesis even if the original sample does not. Let $r_1^{*(a)}, \ldots, r_n^{*(a)}$ be a bootstrap sample from the residuals $r_1^{(a)}, \ldots, r_n^{(a)}$ and construct our "null" bootstrap observations

$$\tilde{y}_i^* = \mathbf{x}_i'\hat{\boldsymbol{\theta}}_{n,0} + r_i^{*(a)}\,, \quad i = 1, \ldots, n\,.$$

We can then use the FRB with these bootstrap samples to obtain recomputed τ-regression estimates and evaluate the test statistics with them.

It is important to note that even after obtaining consistent FRB estimates of the sampling distribution of the regression estimators, the consistency of the FRB for the null distribution of the test statistic in (10) requires special care. As noted in Van Aelst and Willems (2011), an additional regularity condition is required for the FRB distribution estimator to be consistent for statistics with rate of convergence higher than $\sqrt{n}$. The idea can be outlined as follows. Let $\hat{\boldsymbol{\Theta}}_n = (\hat{\boldsymbol{\theta}}_n', \hat{\boldsymbol{\theta}}_{n,0}')' \in \mathbb{R}^{2p}$ denote the vector containing the regression estimators with and without the null hypothesis restriction, and let the test statistic of interest be $L_n = h_n(\hat{\boldsymbol{\Theta}}_n)$, a function of these combined estimates. Note that the function h_n may depend on the sample. Then, the FRB re-computed test statistic is $h_n^*(\hat{\boldsymbol{\Theta}}_n^{R*})$. The Taylor expansion leading to (19) implies that $\hat{\boldsymbol{\Theta}}_n^{R*} = \hat{\boldsymbol{\Theta}}_n^* + O_p(1/n)$ (Salibian-Barrera, Van Aelst and Willems, 2006). This order of approximation is enough to estimate the distribution of statistics that converge with order $O_p(1/\sqrt{n})$, but more is needed for the FRB to work with test statistics that are of order $O_p(1/n)$, like L_n in (10). More specifically, we require that the FRB re-calculation of the test statistic $h_n^*(\hat{\boldsymbol{\Theta}}_n^{R*})$ satisfies

$$h_n^*(\hat{\boldsymbol{\Theta}}_n^{R*}) \;=\; h_n^*(\hat{\boldsymbol{\Theta}}_n^*) + o_p(1/n)\,. \tag{23}$$

Using a Taylor expansion of the test statistic h_n around $\hat{\boldsymbol{\Theta}}_n^*$ we have

$$h_n(\hat{\boldsymbol{\Theta}}_n^{R*}) \;=\; h_n^*(\hat{\boldsymbol{\Theta}}_n^*) + \nabla h_n^*(\hat{\boldsymbol{\Theta}}_n^*)\,(\hat{\boldsymbol{\Theta}}_n^{R*} - \hat{\boldsymbol{\Theta}}_n^*) \;+\; o_p(1/n)\,.$$

Since $\hat{\boldsymbol{\Theta}}_n^{R*} - \hat{\boldsymbol{\Theta}}_n^* = O_p(1/n)$, a sufficient condition for (23) to hold is that

$$\nabla h_n^*(\hat{\boldsymbol{\Theta}}_n^*) \;=\; o_p(1)\,. \tag{24}$$

Note that the test statistic L_n in (10) can be written as a smooth function of the regression estimators $\hat{\boldsymbol{\theta}}_{n,0}$, $\hat{\boldsymbol{\theta}}_n$:

$$L_n(\hat{\boldsymbol{\theta}}_{n,0}\,, \hat{\boldsymbol{\theta}}_n) \;=\; h_n(\hat{\boldsymbol{\theta}}_{n,0}\,, \hat{\boldsymbol{\theta}}_n)\,, \tag{25}$$

for which it can be seen that condition (24) above is satisfied. Details are included in the Appendix.

Note that in the representation (25) above, we are expressing our test statistic as a function of the regression estimators only, and do not include the scale estimators. This implies that when we use the FRB to approximate the distribution of (10), we need to compute the scale estimates exactly using the residuals of the corresponding bootstrapped regression estimators. Although this implies a slight increase in computational complexity over simply using the bootstrapped scale estimators, it is negligible compared to the cost of re-computing the regression estimators for each bootstrap sample.

4. Finite-sample behavior

To investigate the finite-sample properties of our robust ANOVA τ-tests, we conducted a simulation study. We generated 500 samples of sizes n = 50, 100, 200 and 500 according to the linear regression model

$$Y = \boldsymbol{\theta}_0' \mathbf{Z} + \varepsilon\,, \tag{26}$$

where $\boldsymbol{\theta}_0 \in \mathbb{R}^5$ and $\mathbf{Z} = (1, \mathbf{X}')'$, with $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_4)$, independent from $\varepsilon \sim N(0,1)$. The null hypothesis is $H_0 \,:\, \boldsymbol{\theta}_{0,4} = \boldsymbol{\theta}_{0,5} = 0$. The vector of regression parameters was set to $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}, \ldots, \boldsymbol{\theta}_{0,5})' = (1,1,1,d,0)'$, and we varied the parameter d to obtain samples from the null ($d = 0$) and alternative hypotheses (d = 0.25, 0.50 and 1).

We used both the asymptotic approximation and the FRB to estimate the p-values of all robust tests in this comparison. In addition to the robust ANOVA tests based on L_n in (10) and SL_n in (13), we also consider the robust Wald test (Hampel *et al.*, 1986) and the robust scores test (Markatou and Hettmansperger, 1990). Let $\hat{\boldsymbol{\theta}}_{n,0}$ and $\hat{\boldsymbol{\theta}}_n$ be the estimators for the null and full model, respectively. Then, the Wald test statistic is given by

$$W_n \;=\; \hat{\boldsymbol{\theta}}_n^{(2)\,\prime}\; (V_{n,(2,2)})^{-1}\; \hat{\boldsymbol{\theta}}_n^{(2)}\,, \tag{27}$$

where $V_{n,(2,2)}$ is the lower right submatrix of $V_n \;=\; M_n^{-1}\, Q_n\, M_n^{-1}$, $M_n = \sum_{i=1}^n \psi_0'(r_i/\hat{\sigma}_n)\mathbf{x}_i\mathbf{x}_i'/n$, $Q_n = \sum_{i=1}^n \psi_0^2(r_i/\hat{\sigma}_n)\mathbf{x}_i\mathbf{x}_i'/n$, $r_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\theta}}_n$, $\hat{\sigma}_n$ is the scale estimator in the full model, and $\psi_0 = W\psi_1 + \psi_2$ as before. In practice, we estimate the constant W in ψ_0 using its empirical counterpart.

The scores test statistic is given by

$$R_n \;=\; n\; S_{(2)}^{(0)\,\prime}\; U_n^{-1}\, S_{(2)}^{(0)}\,, \tag{28}$$

where

$$U_n \;=\; Q_{n,(2,2)} - M_{n,(2,1)}\, M_{n,(1,1)}^{-1}\, Q_{n,(1,2)} - Q_{n,(2,1)}\, M_{n,(1,1)}^{-1}\, M_{n,(1,2)} + M_{n,(2,1)}\, M_{n,(1,1)}^{-1}\, Q_{n,(1,1)}\, M_{n,(1,1)}^{-1}\, M_{n,(1,2)}\,,$$

$S_{(2)}^{(0)}$ is the corresponding subvector of the scores function computed under the null hypothesis,

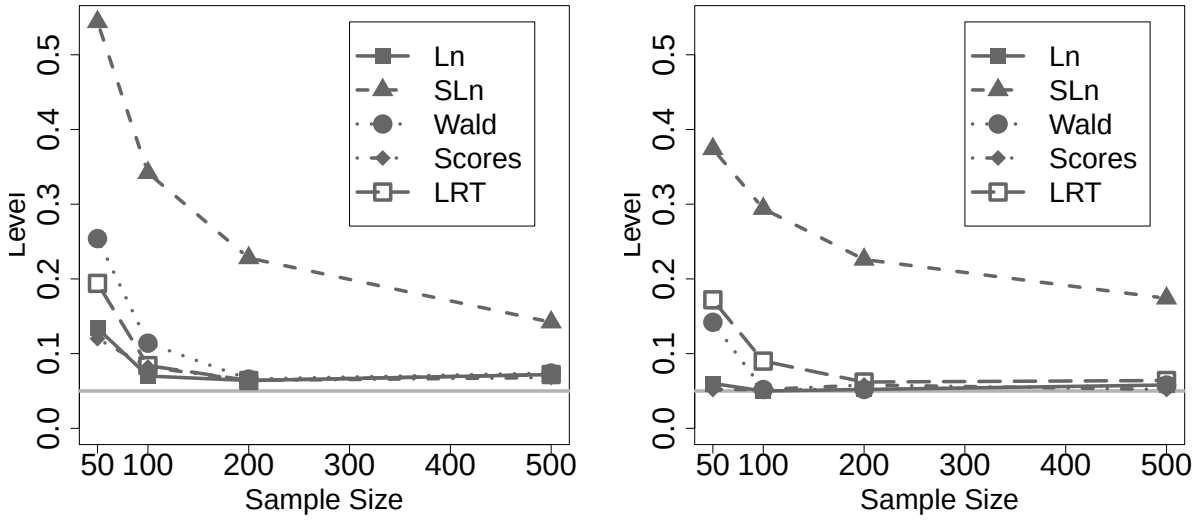
$$S^{(0)} = \sum_{i=1}^{n} \psi_0(r_i^{(0)}/\hat{\sigma}_n)\mathbf{x}_i\mathbf{x}_i'/n\,,$$

and $r_i^{(0)} = y_i - \mathbf{x}_i'\hat{\boldsymbol{\theta}}_{n,0}$. Under H_0, both W_n and R_n have an asymptotic χ^2_{p-q} distribution.

We also included in our study a robust likelihood-ratio type test statistic, also called drop-in dispersion test (Ronchetti, 1982; Markatou *et al.*, 1991). :

$$H_n \;=\; \frac{2}{n}\, \frac{\sum_{i=1}^n \psi_0'(r_i/\hat{\sigma}_n)}{\sum_{i=1}^n \psi_0^2(r_i/\hat{\sigma}_n)} \left(\sum_{i=1}^{n} \left[\rho_0\Big(\frac{r_i^{(0)}}{\hat{\sigma}_n}\Big) - \rho_0\left(\frac{r_i}{\hat{\sigma}_n}\right) \right] \right), \tag{29}$$

where $\rho_0 = W\rho_1 + \rho_2$. For τ-estimators it can be shown that the statistic nH_n has an asymptotic χ^2_{p-q} distribution under the null hypothesis.



(a) Asymptotic approximation

(b) FRB

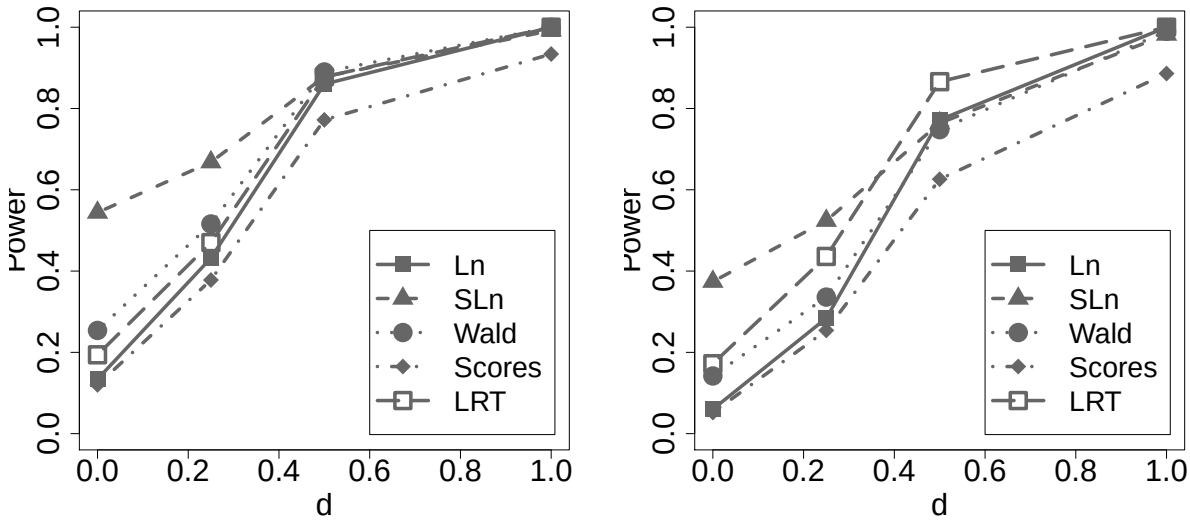
Figure 2: Accuracy of level. Empirical levels as a function of the sample size, when no contamination is present. The left panel contains the results obtained by using the asymptotic χ^2 approximation. Estimating the p-values using FRB yields the empirical levels displayed in the right panel.

First we investigate the level of these tests. Figure 2 contains the empirical levels for tests with a nominal level of 5%, when the data follow model (26) with $\boldsymbol{\theta}_{0,4} = \boldsymbol{\theta}_{0,5} = 0$, i.e. the null model. We see that when the sample size increases the asymptotic approximation becomes more accurate and the observed levels approach the nominal level, as expected. Also note that the FRB approximation consistently yields tests with empirical levels closer to the nominal one. Moreover, the scores test and the ANOVA τ-test based on L_n are the most accurate tests while the others are too liberal, especially for small sample sizes. In particular, note that SL_n, the ANOVA test based on S-estimators, is not reliable at all due to the low efficiency of the S-scale estimators.

To investigate the power of the tests when the p-values are estimated using either the asymptotic approximation or the FRB, we varied the values of the parameter d in (26). The results for samples of sizes $n = 50$ and $n = 100$ when the data do not contain outliers are displayed in Figures 3 and 4, respectively. We can see that the power of the tests increases quickly with increasing distance d, especially for larger samples. For close-by alternatives, the power is generally higher when using the asymptotic distribution than with FRB. However, this is a consequence of the fact that the asymptotic tests are also more liberal, as can again be seen for the case $d = 0$ in these plots. For $n = 100$ we can see from Figure 4 that there is little difference between the tests, except for the unreliable SL_n test. For $n = 50$ Figure 3 shows that the LRT and Wald tests yield higher power than the ANOVA τ-test and the scores test, but this is again a consequence of these tests being too liberal under the null hypothesis. Comparing the ANOVA τ-test with the scores test, we can see that the ANOVA τ-test has higher power.

To investigate the robustness of the tests, samples were then contaminated with bad leverage points. We first ran a small numerical experiment to identify the least favourable outlier configuration for these tests. We replaced 10% of the values of X_2 with observations following a $\mathcal{N}(5, 0.1^2)$ distribution, and the corresponding response Y had a $\mathcal{N}(\eta, 0.1^2)$ distribution, with $\eta = 0$, 2, 5, 7, 10, 12, 15, 20, 25, 30, 35, 40, and 45. Figure 5 shows the observed empirical level and powers (for the closest alternative with $d = 0.25$) as a function of η, for $n = 200$, and p-values obtained with the FRB. The results for other sample sizes and p-values computed with the asymptotic approximation are very similar. We see that the least favourable outlier contamination appears to be around $\eta = 12$.

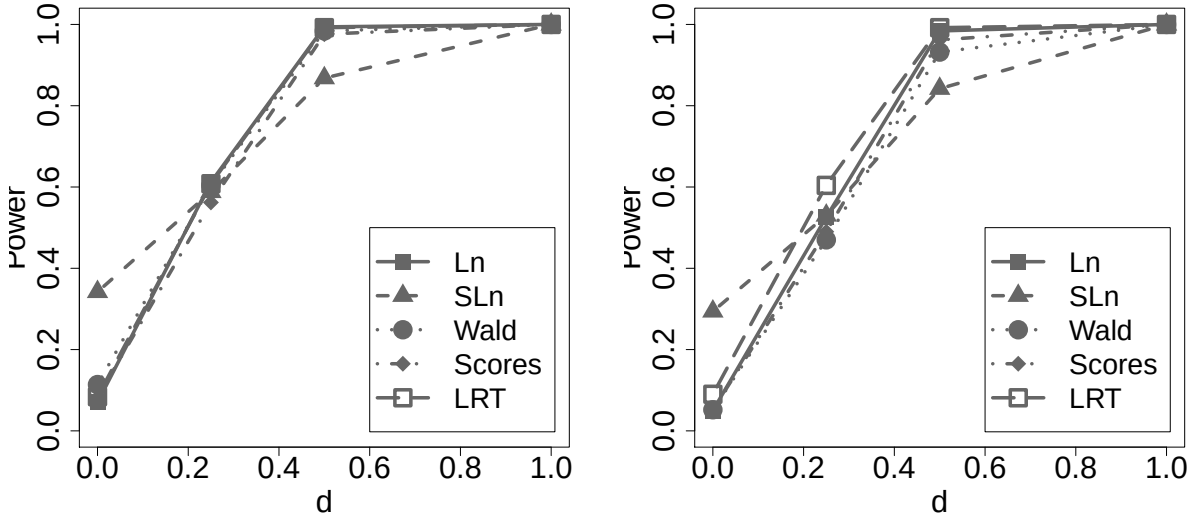
To investigate the effect of particularly damaging outliers on the level of these tests we placed outliers with $X_2 \sim \mathcal{N}(5, 0.1^2)$ and $Y \sim \mathcal{N}(12, 0.1^2)$. The results are shown in Figure 6. Note that p-values estimated by FRB again yield



(a) Asymptotic approximation

(b) FRB

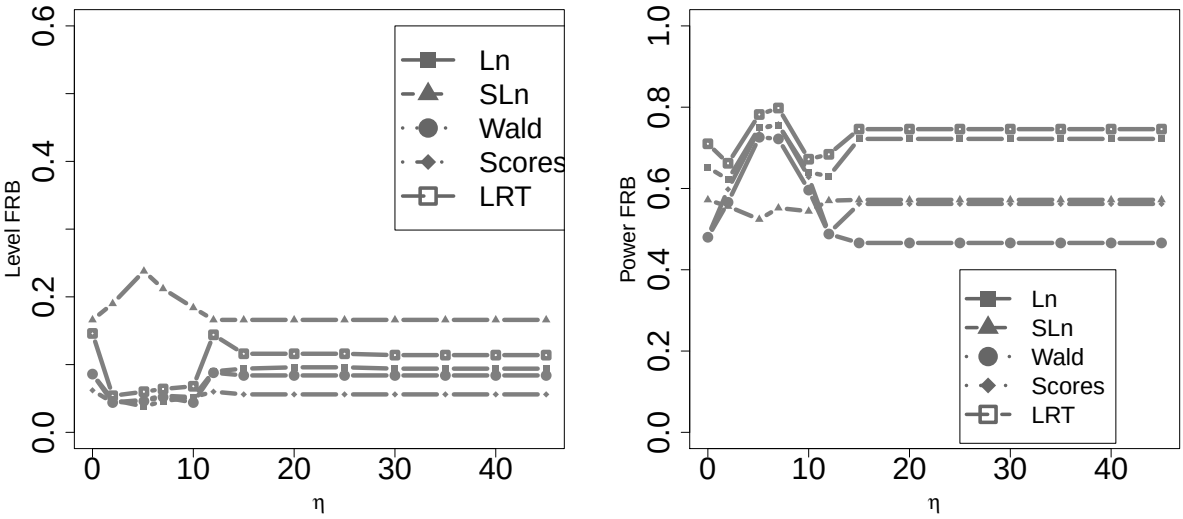
Figure 3: Empirical power. Empirical power levels for uncontaminated samples of size $n = 50$ as a function of the parameter d in (26). The left and right panels contain the results obtained with the asymptotic χ^2 approximation and with the FRB, respectively.



(a) Asymptotic approximation

(b) FRB

Figure 4: Empirical power. Empirical power levels for uncontaminated samples of size $n = 100$ as a function of the parameter d in (26). The left and right panels contain the results obtained with the asymptotic χ^2 approximation and with the FRB, respectively.



(a) Levels

(b) Powers

Figure 5: Empirical levels and powers (for $d = 0.25$) for different outlier configurations, for samples of size $n = 200$.

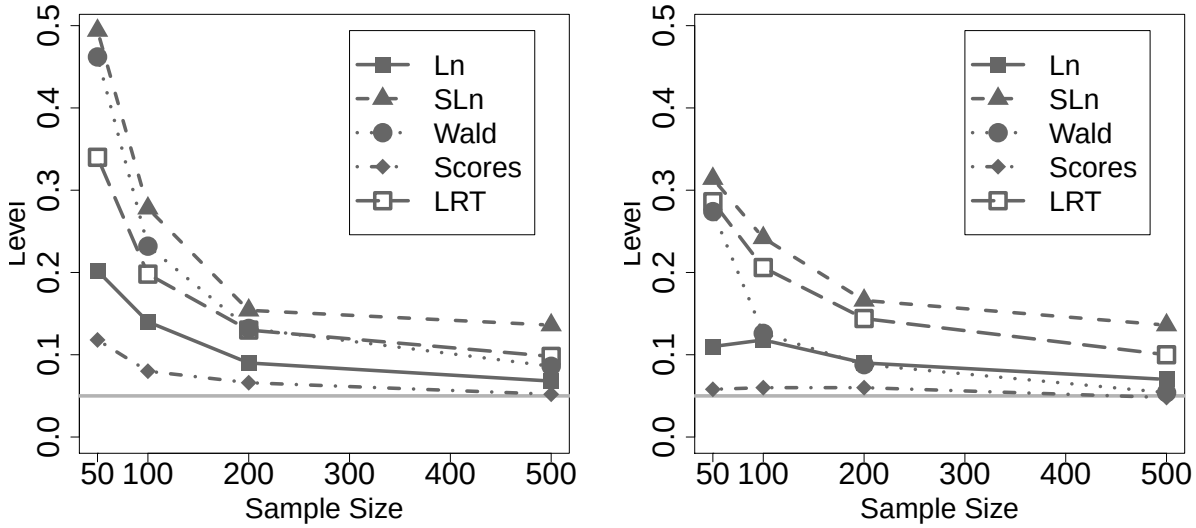
empirical levels closer to the nominal level than asymptotic p-value approximations. In both cases (samples with and without outliers) the advantage of using the FRB is more noticeable for smaller sample sizes. The scores test clearly best maintains the level in this case. The ANOVA τ-test also gives reasonable results, especially when using FRB. The Wald test based on FRB performs well for larger samples, but not for small samples.

To investigate the robustness of the power, Figures 7 and 8 show the results for contaminated samples of size $n = 50$ and $n = 100$ respectively. From these plots we can see that of the tests with level closest to the nominal value, the power curve for the ANOVA τ-test increases fastest, and hence the test rapidly achieves power comparable to the other tests while maintaining one of the best empirical levels. Note that this experiment was done using the most-damaging outlier configuration for the ANOVA τ-test. The scores test does not seem to perform well, especially for smaller samples ($n = 50$). The good (robustness of) level of this test thus comes with a price in (robustness of) power. Overall, we can conclude that the ANOVA τ-test shows a good performance both in terms of accuracy under the null and power under the alternative. Moreover, in the presence of contamination this test gives a good compromise between robustness of level and robustness of power, compared to the other robust tests.

5. Example

To illustrate our method, we consider data from a linguistics experiment, designed to study whether a reader's processing time when he or she encounters new and complex words (in this case: neologisms in Dutch) is affected by having been previously exposed to it or its base. The data contains 656 observations obtained from 32 subjects. The response variable is the logarithm of the reading latency, and there are 13 possible covariates: the logarithm of the frequency of the lowest-level base of the neologism (RootFrequency), the estimated frequency of the word (Rating), the frequency of the base adjective (BaseFrequency), the number of letters in the word (LengthInLetters), a measure of the number of meanings of the words base: its number of synonym sets in WordNet (NumberOfSynsets) (Miller, 1990), the logarithm of the size of the morphological family of the word (FamilySize), the logarithm of reading latency when exposed 1, 2, 3 and 4 trials back (RT1WordBack, RT2WordsBack, RT3WordsBack, RT4WordsBack), 1 and 2 trials later (RT1WordLater, RT2WordsLater), and for the prime word (RTtoPrime).

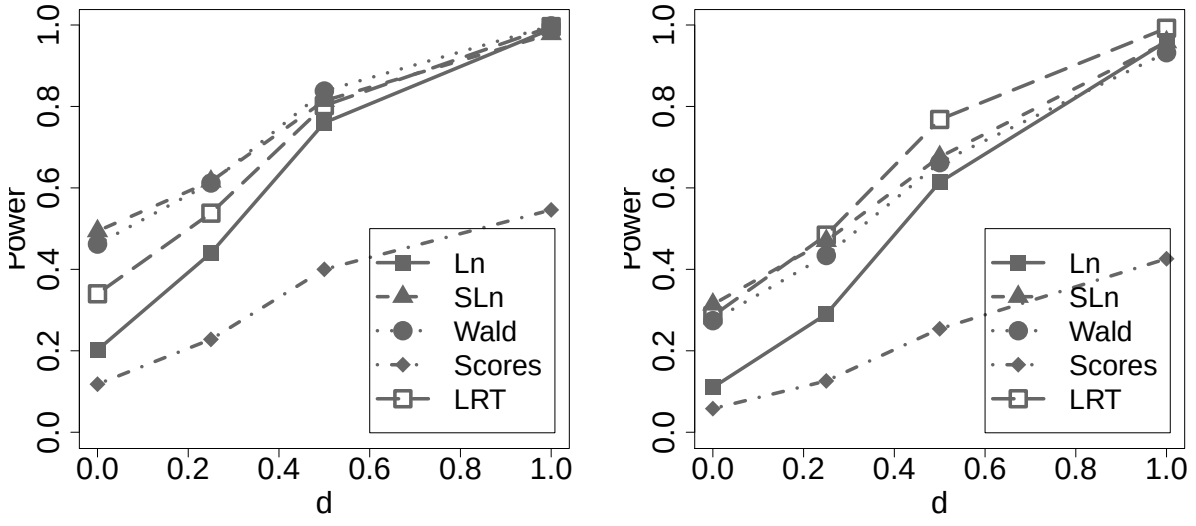
The data set is available in R (`selfPacedReadingHeid` in package `languageR`) and was originally analyzed by De Vaan *et al.* (2007) (see also Baayen, 2008) who identified 5 observations with "extreme" values of the response



(a) Asymptotic approximation

(b) FRB

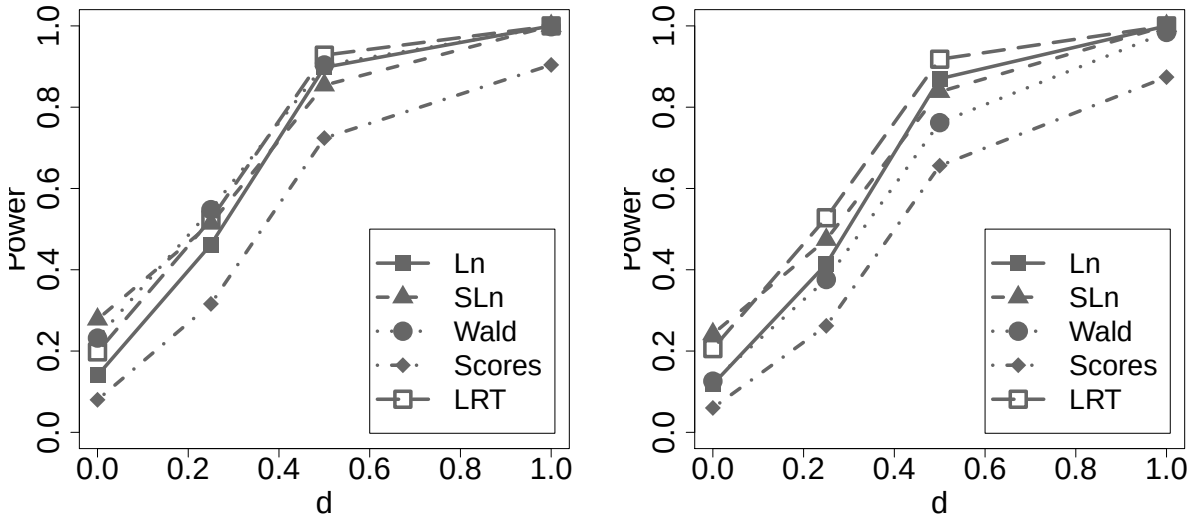
Figure 6: Robustness of level. Empirical levels as a function of the sample size, when samples contain 10% of high-leverage outliers. The left panel contains the results obtained by using the asymptotic χ^2 approximation. Estimating the p-values using FRB yields the empirical levels displayed in the right panel.



(a) Asymptotic approximation

(b) FRB

Figure 7: Empirical power for samples with 10% of high-leverage outliers. Empirical power levels for contaminated samples of size $n = 50$ as a function of the parameter d in (26). The left and right panels contain the results obtained with the asymptotic χ^2 approximation and with the FRB, respectively.



(a) Asymptotic approximation

(b) FRB

Figure 8: Empirical power for samples with 10% of high-leverage outliers. Empirical power levels for contaminated samples of size $n = 100$ as a function of the parameter d in (26). The left and right panels contain the results obtained with the asymptotic χ^2 approximation and with the FRB, respectively.

variable (less than 5 and larger than 7), and flagged them as potential outliers. We follow a more principled approach and first fit a linear model using a τ-regression estimator. The robust estimator identifies 6 potential outliers (approximately 10% of the observations) with residuals larger than 3.5 standard deviations in absolute value. Figure 9 shows the corresponding outlier diagnostic plot which plots the residuals vs robust distances of the observations in the design space, based on MM-location and scatter estimators. This plot reveals that there are 5 vertical outliers and one extreme bad leverage point that may largely affect any nonrobust analysis. Table 1 displays the summary of the τ-regression estimate, tuned to attain a breakdown point of 50% and 95% efficiency compared to least squares when the errors are normal. The standard errors of the regression estimates were computed using the FRB. Interestingly, potentially significant covariates (individual p-values less than 0.05) are the length of the word and most of the reading latencies for previous exposure to the word or its prime, which seems to be in agreement with the researchers' hypothesis.

We now test the hypothesis that the coefficients for all the other covariates are zero. We consider the same tests as in the comparison in the previous section. The estimated p-values of the tests using both the asymptotic approximation and the FRB are given in Table 2. Comparing the asymptotic χ^2 with the FRB approximation for the null distribution of the test statistics, we see that the p-values estimated by FRB are larger for all but one of the tests (the unreliable SL_n test which is not significant in either case). Note that the asymptotic χ^2 distribution does not seem to give a good approximation for the null distribution of the scores test in this example, leading to a clearly deviating p-value. Since the sample size is reasonably large, next to the ANOVA type τ-test, also the robust Wald and Scores test based on FRB can be expected to give reliable results. In Table 2 we can see that the FRB based p-values of these tests are indeed fairly close to each other and all lead to the same conclusion.

Overall, the test results suggest that there is not enough evidence to include variables in the model other than those related to previous exposure to the word, or its length. To verify that this is a sensible conclusion, we run a small cross-validation (CV) experiment to compare the prediction power of the full and null models. We used 500 runs of 5-fold CV. Given the presence of potential outliers in the data, to measure the quality of the predicted values we used

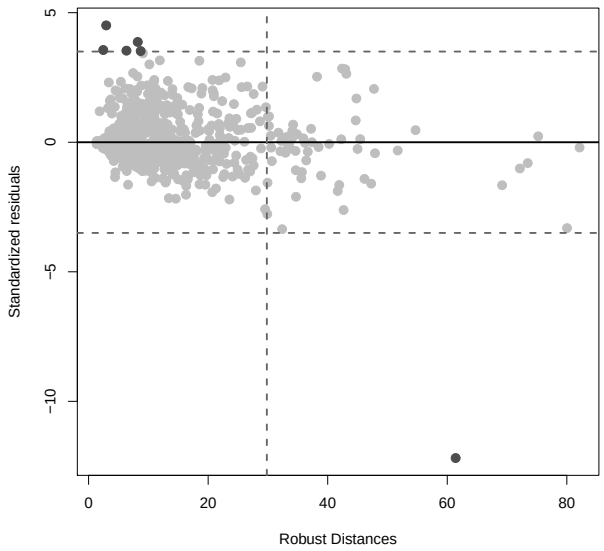


Figure 9: Plot of standardized residuals versus robust Mahalanobis distances for the linguistic data. The horizontal dashed lines indicate ±3.5 standard deviations. The vertical line corresponds to the 99.5% quantile of a χ^2_{13} distribution. Observations that are beyond this line can be considered high-leverage points.

	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	-1.454	0.299	-4.856	0.00
RT4WordsBack	0.193	0.062	3.090	0.00
RT3WordsBack	0.305	0.068	4.467	0.00
RT1WordBack	0.344	0.058	5.906	0.00
RTtoPrime	0.123	0.049	2.520	0.01
RT2WordsLater	0.139	0.048	2.877	0.00
LengthInLetters	0.032	0.015	2.056	0.04
RT2WordsBack	0.069	0.059	1.175	0.24
FamilySize	-0.015	0.011	-1.369	0.17
RT1WordLater	0.032	0.044	0.723	0.47
Rating	-0.005	0.035	-0.131	0.90
NumberOfSynsets	0.016	0.017	0.927	0.35
BaseFrequency	0.004	0.010	0.367	0.71
RootFrequency	-0.002	0.009	-0.190	0.85

Table 1: Summary of the τ-regression estimator for the linguistics data. Standard errors were estimated using the fast and robust bootstrap.

Test	Asymptotic	FRB
L_n	0.22	0.32
SL_n	0.23	0.10
Wald	0.32	0.41
Scores	0.00	0.26
LRT	0.22	0.27

Table 2: Estimated p-values for the linguistics example using the asymptotic approximation and the fast and robust boostrap.

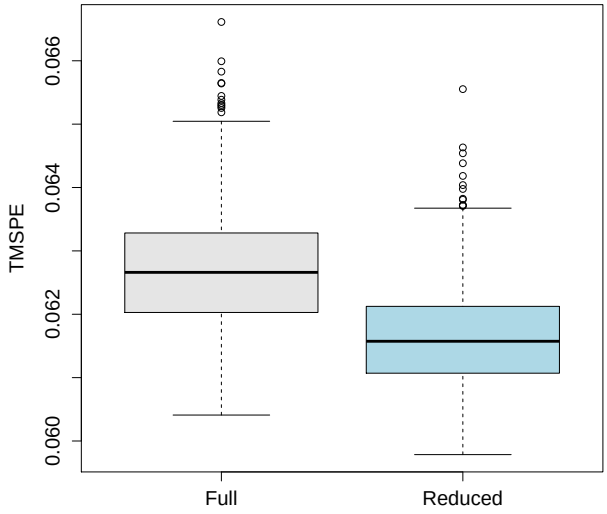


Figure 10: Boxplots of 500 replicates of 5-fold cross-validation estimates of the trimmed mean squared prediction error (10% trimming proportion) for the full and reduced models.

trimmed mean squared prediction errors of the form

$$TMSPE = \frac{1}{n - \lfloor \gamma n \rfloor} \sum_{i=1}^{n-\lfloor \gamma n \rfloor} r_{(i)}^2 ,$$

for $0 \leq \gamma \leq 0.2$, where $r_{(1)}^2 \leq r_{(2)}^2 \leq \ldots \leq r_{(n)}^2$ are the ordered squared residuals $r_i = y_i - \hat{y}_i$. In this way, we do not unduly penalize a fit that does not accommodate possible outliers. Another way to interpret this is by saying that we prefer a model that produces better predictions for a large majority of the data, possibly at the expense of predicting a small minority of observations not well. Figure 10 contains the boxplots of the 500 trimmed mean squared prediction errors with $\gamma = 0.10$ (based on the proportion of outliers identified above). Note that the smaller model gives better predictions for the majority (98%) of the observations. Furthermore, the corresponding trimmed mean squared prediction errors are more stable than those from the full model. The results for $\gamma = 0.00$, 0.01, 0.02, 0.05, 0.15 and 0.20 are very similar and lead to the same conclusion. This finding thus supports the conclusion obtained from the p-values estimated with FRB that the discarded covariates may indeed not be necessary.

6. Conclusion

The illustrative example together with the empirical investigations show that ANOVA τ-tests for testing linear hypotheses in regression models provide both good level and power, even in the presence of outliers, in comparison to the available alternative robust tests. In particular, the SL_n test turned out to be very unreliable and also the Wald test can better be avoided as it is not really reliable. The scores test showed excellent behavior in terms of the level of the test, but this may come at a high price in terms of power, especially for smaller samples. The L_n test shows a power behavior similar to the LRT test while being more stable in terms of level.

The Fast and Robust Bootstrap (FRB) yields a reliable estimate for the null distribution of these tests. Its advantage over the asymptotic approximation is most noticeable for small sample sizes. These results constitute another example of the versatility of the FRB as a feasible and robust distribution estimator, that in many cases outperforms traditional

asymptotic approximations, while typically requiring fewer regularity assumptions. In this paper we have used a model based bootstrap procedure by bootstrapping the residuals obtained from the fit on the original data. However, once 'pseudo responses' have been obtained as in (22), bootstrapping pairs can also be used. Bootstrapping pairs typically requires less assumptions about the distribution of the errors in the linear model, but we expect the power of the resulting tests to be somewhat lower. Finally, note the sample sizes considered in our experiments are not very small compared to the dimension of the data. Robust inference beyond point estimation generally requires sufficiently large sample sizes so that the data contain enough information to obtain stable, useful inference results.

Appendix

Appendix A.1. Equivalence of (8) *and* (9)

Consider the null hypothesis in (8). Let $\mathbf{Q}$ be such that $\mathbf{Q}\mathbf{Q}' = \mathbf{\Sigma}$ where $\mathbf{\Sigma} = E(\mathbf{X}\mathbf{X}')$. Moreover, define $\tilde{V} = \{\boldsymbol{\beta} : \boldsymbol{\beta} = \mathbf{Q}'\boldsymbol{\theta}, \boldsymbol{\theta} \in V\}$ and let $\mathbf{P} \in \mathbb{R}^{p\times p}$ be an orthogonal matrix $\mathbf{P} = (\mathbf{P}_1, \mathbf{P}_2)$ such that the columns of $\mathbf{P}_1 \in \mathbb{R}^{p\times q}$ and $\mathbf{P}_2 \in \mathbb{R}^{p\times(p-q)}$ contain orthogonal bases for $\tilde{V}$ and $\tilde{V}^{\perp}$, respectively. Now transform the linear model in (2) as follows:

$$\begin{aligned} y_i &= \boldsymbol{\theta}_0'\mathbf{x}_i + \epsilon_i \\ &= \boldsymbol{\theta}_0'\mathbf{Q}\,\mathbf{P}\,\mathbf{P}'\,\mathbf{Q}^{-1}\mathbf{x}_i + \epsilon_i \\ &= \boldsymbol{\gamma}_0'\mathbf{z}_i + \epsilon_i\,, \end{aligned}$$

where $\boldsymbol{\gamma}_0 = \mathbf{P}'\mathbf{Q}'\boldsymbol{\theta}_0$ and $\mathbf{z}_i = \mathbf{P}'\mathbf{Q}^{-1}\mathbf{x}_i$. Write $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_0^{(1)'}, \gamma_0^{(2)'})'$, where $\boldsymbol{\gamma}_0^{(1)}$ contains the first p coordinates of $\boldsymbol{\gamma}_0$.

With this change of variables the null hypothesis $\boldsymbol{\theta}_0 \in V$ is equivalent to $\mathbf{Q}'\boldsymbol{\theta}_0 \in \tilde{V}$, which is equivalent to $\mathbf{P}_2'\mathbf{Q}'\boldsymbol{\theta}_0 = \mathbf{0}$ and to

$$H_0^* : \boldsymbol{\gamma}_0^{(2)} \;=\; \mathbf{0}\,. \tag{A.1}$$

Furthermore, the alternative hypothesis H_n in (12) is equivalent to

$$H_n^* \;:\; \boldsymbol{\gamma} \;=\; (\boldsymbol{\gamma}_0^{(1)}, \mathbf{b}/\sqrt{n})\,, \tag{A.2}$$

where $\boldsymbol{\gamma}_0^{(1)} = \mathbf{P}_1'\mathbf{Q}'\boldsymbol{\theta}_0$, and $\mathbf{b} = \mathbf{P}_2'\mathbf{Q}'\mathbf{a}$. Note that

$$E(\mathbf{z}_i\,\mathbf{z}_i) \;=\; \mathbf{P}'\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}'\;\mathbf{Q}'^{-1}\,\mathbf{P} \;=\; \mathbf{I}_p\,,$$

where $\mathbf{I}_p$ denotes the $p \times p$ identity matrix. Moreover, since $\mathbf{P}_1'\mathbf{Q}'\mathbf{a} = \mathbf{0}$ we have

$$\begin{aligned} \mathbf{b}'\mathbf{b} &= \|\mathbf{P}'\mathbf{Q}'\mathbf{a}\|^2 \\ &= \mathbf{a}'\mathbf{Q}\mathbf{Q}'\mathbf{a} \\ &= \mathbf{a}'\mathbf{\Sigma}\mathbf{a}\,. \end{aligned} \tag{A.3}$$

Using the affine equivariance of the τ-estimates, it is easy to show that the test statistic L_n based on the sample $(y_i, \mathbf{x}_i)$, $1 \le i \le n$, for the hypothesis H_0 coincides with the statistic L_n based on the transformed sample $(y_i, \mathbf{z}_i)$, $1 \le i \le n$, for the hypothesis H_0^*. Hence, without loss of generality, we can work with the transformed model. ■

Appendix A.2. Proof of Theorem 1

Without loss of generality (see Section Appendix A.1), assume that the regression model is

$$y_i = \boldsymbol{\gamma}_0'\mathbf{z}_i + \epsilon_i\,,$$

where $\boldsymbol{\gamma}_0 \in \mathbb{R}^p$ and $E(\mathbf{z}_i\,\mathbf{z}_i) = \mathbf{I}_p$, and the null hypothesis is given by (A.1). Let the scale estimators $s_n(\boldsymbol{\gamma})$ and $\tau_n(\boldsymbol{\gamma})$ be as in (5) and (4) with $r_i(\boldsymbol{\gamma}) = y_i - \boldsymbol{\gamma}'\mathbf{z}_i$. The τ-estimator for $\boldsymbol{\gamma}_0$ is

$$\widehat{\boldsymbol{\gamma}}_n \;=\; \arg\min_{\boldsymbol{\gamma}\in R^p} \tau_n^2(\boldsymbol{\gamma})\,,$$

and the restricted τ-estimate under (A.1) is given by $\widehat{\boldsymbol{\gamma}}_{n,0} = (\widehat{\boldsymbol{\gamma}}^{(1)}_{n,0}, \mathbf{0}')'$, where

$$\widehat{\boldsymbol{\gamma}}^{(1)}_{n,0} \;=\; \arg\min_{\boldsymbol{\gamma}^{(1)} \in R^q} \tau^2_n((\boldsymbol{\gamma}^{(1)}, \mathbf{0}))\,.$$

Given a function $f(\boldsymbol{\gamma}) : \mathbb{R}^p \to \mathbb{R}$, we denote by $f'(\boldsymbol{\gamma})$ the p-dimensional vector of derivatives with j-th element $f'(\boldsymbol{\gamma})_j = \partial f(\boldsymbol{\gamma})/\partial \boldsymbol{\gamma}_j$, and by $f''(\boldsymbol{\gamma})$ the $p \times p$ matrix of second derivatives, with (i, j) element $f''(\boldsymbol{\gamma})_{i,j} = \partial^2 f(\boldsymbol{\gamma})/(\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_j)$. If $g(\boldsymbol{\gamma}) : \mathbb{R}^p \to \mathbb{R}^p$, then $g'(\boldsymbol{\gamma})$ denotes the the $p \times p$ matrix of first derivatives, with (i, j) element $g'(\boldsymbol{\gamma})_{i,j} = \partial g_i(\boldsymbol{\gamma})/\partial \boldsymbol{\gamma}_j$.

For $j = 1, 2$, define the following quantities:

$$\begin{aligned}
A_{nj}(\gamma) &= \sum_{i=1}^n \psi_j\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right)\mathbf{z}_i\,, & B_{nj}(\boldsymbol{\gamma}) &= \sum_{i=1}^n \psi_j\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right) r_i(\boldsymbol{\gamma})\,,\\
C_{nj}(\boldsymbol{\gamma}) &= \sum_{i=1}^n \psi'_j\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right) r_i(\boldsymbol{\gamma})\mathbf{z}_i\,, & D_{nj}(\boldsymbol{\gamma}) &= \sum_{i=1}^n \psi'_j\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right) \mathbf{z}_i\,\mathbf{z}'_i\,,\\
F_{nj}(\boldsymbol{\gamma}) &= \sum_{i=1}^n \psi'_j\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right) r^2_i(\boldsymbol{\gamma})\,, & M_{nj}(\boldsymbol{\gamma}) &= \sum_{i=1}^n \rho_j\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right)\,.
\end{aligned}$$

The following lemma gives the gradient $s'_n(\boldsymbol{\gamma})$ and Hessian matrix $s''_n(\boldsymbol{\gamma})$ of the function $s_n(\boldsymbol{\gamma})$ in (5).

Lemma 1 *With the above notation, it holds that*

(a) $s'_n(\boldsymbol{\gamma}) = -s_n(\boldsymbol{\gamma})\, A_{n1}(\boldsymbol{\gamma})/B_{n1}(\boldsymbol{\gamma})$,

$$\begin{aligned}
(b)\;\; s''_n(\boldsymbol{\gamma}) = {}& \frac{s_n(\boldsymbol{\gamma})A_{n1}(\boldsymbol{\gamma})A^T_{n1}(\boldsymbol{\gamma}) + D_{n1}(\boldsymbol{\gamma})B_{n1}(\boldsymbol{\gamma}) \;- A_{n1}(\boldsymbol{\gamma})C^T_{n1}(\boldsymbol{\gamma})}{B^2_{n1}(\boldsymbol{\gamma})}\\
&+ \frac{C_{n1}(\boldsymbol{\gamma})A^T_{n1}(\boldsymbol{\gamma}) \;- F_{n1}(\boldsymbol{\gamma})A_{n1}(\boldsymbol{\gamma})A^T_{n1}(\boldsymbol{\gamma})/B_{n1}(\boldsymbol{\gamma}) + s_n(\boldsymbol{\gamma})A_{n1}(\boldsymbol{\gamma})A^T_{n1}(\boldsymbol{\gamma})}{B^2_{n1}(\boldsymbol{\gamma})}.
\end{aligned}$$

Proof: Part (a) follows by differentiating both sides of (5). We obtain

$$\frac{1}{n}\sum_{i=1}^n \psi_1\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right)(-\mathbf{z}_i s_n(\boldsymbol{\gamma}) - s'_n(\boldsymbol{\gamma})r_i(\boldsymbol{\gamma})) = 0\,.$$

Hence,

$$s'_n(\boldsymbol{\gamma}) \;=\; -\frac{s_n(\boldsymbol{\gamma})\, A_{n1}(\boldsymbol{\gamma})}{B_{n1}(\boldsymbol{\gamma})}\,. \tag{A.4}$$

To show (b), we first find the derivatives of $A_{n1}(\boldsymbol{\gamma})$ and $B_{n1}(\boldsymbol{\gamma})$ which yields

$$\begin{aligned}
A'_{n1}(\boldsymbol{\gamma}) &= \sum_{i=1}^n \psi'_1\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right)\frac{\mathbf{z}_i}{s^2_n(\boldsymbol{\gamma})}[-\mathbf{z}^{\mathrm{T}}_i s_n(\boldsymbol{\gamma}) - s'_n(\boldsymbol{\gamma})r_i(\boldsymbol{\gamma})]\\
&= -\sum_{i=1}^n \psi'_1\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right)\frac{\mathbf{z}_i\mathbf{z}'_i}{s_n(\boldsymbol{\gamma})} - \frac{1}{s^2_n(\boldsymbol{\gamma})}\sum_{i=1}^n \psi'_1\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right)\mathbf{z}_i s'^{\mathrm{T}}_n(\boldsymbol{\gamma})r_i(\boldsymbol{\gamma})\\
&= -\frac{D_{n1}(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})} + \frac{A_{n1}(\boldsymbol{\gamma})C'^{\mathrm{T}}_{n1}(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})B_{n1}(\boldsymbol{\gamma})}.
\end{aligned} \tag{A.5}$$

and

$$B'_{n1}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \psi'_1\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right) \frac{r_i(\boldsymbol{\gamma})}{s_n^2(\boldsymbol{\gamma})}[-\mathbf{z}_i s_n(\boldsymbol{\gamma}) - s'_n(\boldsymbol{\gamma}) r_i(\boldsymbol{\gamma})]$$
$$- \sum_{i=1}^{n} \psi_1\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right)\mathbf{z}_i$$
$$= -\frac{C_{n1}(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})} + \frac{F_{n1}(\boldsymbol{\gamma})A_{n1}(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})B_{n1}(\boldsymbol{\gamma})} - A_{n1}(\boldsymbol{\gamma}). \tag{A.6}$$

Moreover, differentiating (A.4) yields

$$s''_n(\boldsymbol{\gamma}) = -\frac{B_{n1}(s'_n A^{\mathrm{T}}_{n1}(\boldsymbol{\gamma}) + s_n(\boldsymbol{\gamma})A'_{n1}(\boldsymbol{\gamma})) - s_n(\boldsymbol{\gamma})A_{n1}(\boldsymbol{\gamma})B'_{n1}(\boldsymbol{\gamma})}{B^2_{n1}(\boldsymbol{\gamma})},$$

and by inserting (A.5) and (A.6) we obtain (b). ■

The following lemma gives the derivatives of the τ-scale function $\tau_n^2(\boldsymbol{\gamma})$ in (4).

Lemma 2 *(a)* $\tau_n^{2\prime}(\boldsymbol{\gamma}) = 2s_n(\boldsymbol{\gamma})s'_n(\boldsymbol{\gamma})M_{n2}(\boldsymbol{\gamma}) - s_n(\boldsymbol{\gamma})A_{n2}(\boldsymbol{\gamma}) - B_{n2}(\boldsymbol{\gamma})s'_n(\boldsymbol{\gamma})$,

$$(b)\ \tau_n^{2\prime\prime}(\boldsymbol{\gamma}) = 2s'_n s'^T_n M_{n2}(\boldsymbol{\gamma}) + 2s_n s''_n M_2(\boldsymbol{\gamma}) + 2s_n s'_n\left(-\frac{A^T_{n2}}{s_n} - \frac{B_{n2}s'^T_n}{s_n^2}\right)$$
$$- s'_n A^T_{n2} - s_n A'_{n2} - B'_{n2}s'^T_n - B_{n2}s''_n.$$

Proof: By differentiating (4) we get

$$\tau_n^{2\prime}(\boldsymbol{\gamma}) = 2s_n s'_n \sum_{i=1}^{n} \rho_2\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right) + s_n^2 \sum_{i=1}^{n} \psi_2\left(\frac{r_i(\boldsymbol{\gamma})}{s_n(\boldsymbol{\gamma})}\right)\left[\frac{-\mathbf{z}_i s_n(\boldsymbol{\gamma}) - r_i(\boldsymbol{\gamma})s'_n}{s_n^2}\right]$$
$$= 2s_n(\boldsymbol{\gamma})s'_n(\boldsymbol{\gamma})M_{n2}(\boldsymbol{\gamma}) - s_n(\boldsymbol{\gamma})A_{n2}(\boldsymbol{\gamma}) - B_{n2}(\boldsymbol{\gamma})s'_n(\boldsymbol{\gamma}), \tag{A.7}$$

which shows (a). By differentiating (A.7) we obtain (b). ■

Lemma 3 *If* $\boldsymbol{\gamma}_n^* \to \boldsymbol{\gamma}_0$ *a.s., then* $\tau_n^{2'}(\boldsymbol{\gamma}_n^*) \to D_0\,\mathbf{I}_p$.

Proof: Using Lemma 4.2 of Yohai (1985) we immediately obtain that $s_n(\boldsymbol{\gamma}_n^*) \to s_0$ a.s., $A_{nj}(\boldsymbol{\gamma}_n^*) \to 0$ a.s., and $C_{nj}(\boldsymbol{\gamma}_n^*) \to 0$ a.s.. Moreover, for $j = 1, 2$ we have that

$$B_{nj}(\boldsymbol{\gamma}_n^*) \to E\left(\psi_j\left(\frac{u}{s_0}\right)u\right) \quad \text{a.s.},$$
$$D_{nj}(\boldsymbol{\gamma}_n^*) \to E\left(\psi'_j\left(\frac{u}{s_0}\right)\right)E\left(\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\right) \quad \text{a.s.}, \tag{A.8}$$
$$F_{nj}(\boldsymbol{\gamma}_n^*) \to E\left(\psi'_j\left(\frac{u}{S_0}\right)u^2\right) \text{ a.s.}, \tag{A.9}$$
$$M_{nj}(\boldsymbol{\gamma}_n^*) \to E\left(\rho_j\left(\frac{u}{s_0}\right)\right) \text{ a.s..} \tag{A.10}$$

Using Lemma 1 we have $s'_n(\boldsymbol{\gamma}_n^*) \to 0$ a.s., and $s''_n(\boldsymbol{\gamma}_n^*) \to D_1/(s_0B_1)\mathbf{I}_p$ a.s. We now use Lemma 2 (b) to obtain

$$\tau_n^{2\prime\prime}(\boldsymbol{\gamma}_n^*) \longrightarrow \frac{2s_0D_1M_2}{s_0B_1}\mathbf{I}_p + D_2\mathbf{I}_p + \frac{B_2D_1}{B_1}\mathbf{I}_p \quad \text{a.s.}$$
$$= (WD_1 + D_2)\mathbf{I}_p = D_0\mathbf{I}_p.$$

■

Lemma 4 *Under H_0^* we have that $n^{1/2}(\boldsymbol{\gamma}_n^{(1)} - \boldsymbol{\gamma}_{n,0}^{(1)}) \to_P 0$.*

Proof: Following the proof of Theorem 5.1 in of Yohai and Zamar (1986) and using that $E(\mathbf{z}_i\mathbf{z}_i') = \mathbf{I}_p$, we have

$$n^{1/2}(\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0) = \frac{1}{D_0 n^{1/2}} \sum_{i=1}^{n} \psi_0\left(\frac{u}{s_0}\right)\mathbf{z}_i + o_p(1)\,.$$

Therefore, if $\mathbf{z}_i^{(1)}$ is the vector formed with the first q components of $\mathbf{z}_i$ we obtain that

$$n^{1/2}(\widehat{\boldsymbol{\gamma}}_n^{(1)} - \boldsymbol{\gamma}_0^{(1)}) = \frac{1}{D_0 n^{1/2}} \sum_{i=1}^{n} \psi_0\left(\frac{u}{s_0}\right)\mathbf{z}_i^{(1)} + o_p(1)\,.$$

When H_0 holds, using the same Lemma we also find that

$$n^{1/2}(\widehat{\boldsymbol{\gamma}}_{n0}^{(1)} - \boldsymbol{\gamma}_0^{(1)}) = \frac{1}{D_0 n^{1/2}} \sum_{i=1}^{n} \psi_0\left(\frac{u}{s_0}\right)\mathbf{z}_i^{(1)} + o_p(1)\,,$$

and the result follows immediately. ■

Lemma 5 *Under H_0^* we have that*

$$L_n \;=\; n\,\frac{D_0}{2s_0^2 M_2}\,\widehat{\boldsymbol{\gamma}}_n^{(2)T}\widehat{\boldsymbol{\gamma}}_n^{(2)} \;+\; o_p(1)\,. \tag{A.11}$$

Proof: Since $\tau^{2\prime}(\widehat{\boldsymbol{\gamma}}_n) = 0$, we have

$$\tau_n^2(\widehat{\boldsymbol{\gamma}}_{n,0}) - \tau_n^2(\widehat{\boldsymbol{\gamma}}_n) = \frac{1}{2}(\widehat{\boldsymbol{\gamma}}_{n,0} - \widehat{\boldsymbol{\gamma}}_n)^{\mathrm{T}}\tau_n^{2\prime\prime}(\tilde{\boldsymbol{\gamma}}_n)(\widehat{\boldsymbol{\gamma}}_{n,0} - \widehat{\boldsymbol{\gamma}}_n), \tag{A.12}$$

where $\tilde{\boldsymbol{\gamma}}_n$ is an intermediate point between $\widehat{\boldsymbol{\gamma}}_{n,0}$ and $\widehat{\boldsymbol{\gamma}}_n$. Theorem 4.1 in Yohai and Zamar (1988) shows that under H_0 both sequences are consistent to γ_0, and hence, by Lemma 3 we have that

$$\tau_n^{2\prime\prime}(\tilde{\boldsymbol{\gamma}}_n) \to\; D_0\,\mathbf{I}_p \quad \text{a.s.} \tag{A.13}$$

By Lemma 4 we have that

$$n^{1/2}(\widehat{\boldsymbol{\gamma}}_{n,0} - \widehat{\boldsymbol{\gamma}}_n) - n^{1/2}(0, \widehat{\boldsymbol{\gamma}}_n^{(2)}) \to_P 0\,, \tag{A.14}$$

and, finally, using (A.10) we obtain

$$\tau_n^2(\widehat{\boldsymbol{\gamma}}_n) \to s_0^2 M_2 \quad \text{a.s.} \tag{A.15}$$

Equations (A.12)-(A.15) imply (A.11). ■

Proof of Theorem 1: Theorem 5.1 in Yohai and Zamar (1988) shows that

$$n^{1/2}(\widehat{\boldsymbol{\gamma}}_n^{(2)} - \boldsymbol{\gamma}_0^{(2)}) \to^D \mathrm{N}_{p-q}(0, \vartheta\mathbf{I}_{p-q}),$$

where $\vartheta = s_0^2 K_0 / D_0^2$. Then, under H_0^* in (A.1) we obtain

$$\frac{n^{1/2}\widehat{\boldsymbol{\gamma}}_n^{(2)}}{\vartheta^{1/2}} \to^D \mathrm{N}_{p-q}(0, \mathbf{I}_{p-q})\,,$$

which implies that

$$\frac{n\widehat{\boldsymbol{\gamma}}_n^{(2)\mathrm{T}}\widehat{\boldsymbol{\gamma}}_n^{(2)}}{\vartheta} \to^D \chi_{p-q}^2\,, \tag{A.16}$$

where χ_{p-q}^2 is the central chi-square distribution with $p-q$ degrees of freedom. Using Lemma 5 we have that

$$\frac{2\,D_0\,M_2}{K_0}\,L_n \;=\; n\,\frac{\widehat{\boldsymbol{\gamma}}_n^{(2)\mathrm{T}}\widehat{\boldsymbol{\gamma}}_n^{(2)}}{\vartheta} + o_p(1)\,, \tag{A.17}$$

which, together with (A.16), shows part (a) of the Theorem.

Under H_n^* we have that

$$\frac{n^{1/2}(\widehat{\boldsymbol{\gamma}}_n^{(2)} - \mathbf{b}/n^{1/2})}{\vartheta^{1/2}} = \frac{n^{1/2}\widehat{\boldsymbol{\gamma}}_n^{(2)} - \mathbf{b}}{\vartheta^{1/2}} \to^D \mathrm{N}_{p-q}(0, \mathbf{I}_{p-q}),$$

where $\mathbf{b}$ is as in (A.2), and then

$$\frac{n^{1/2}\widehat{\boldsymbol{\gamma}}_n^{(2)}}{\vartheta^{1/2}} \to^D \mathrm{N}_{p-q}\left(\frac{\mathbf{b}}{\vartheta}, \mathbf{I}_{p-q}\right),$$

which implies

$$\frac{n\widehat{\boldsymbol{\gamma}}_n^{(2)\mathrm{T}}\widehat{\boldsymbol{\gamma}}_n^{(2)}}{\vartheta} \to^D \chi^2_{p-q}(\delta),$$

where $\chi^2_{p-q}(\delta)$ is the non-central chi-square distribution with $p-q$ degrees of freedom and non-centrality parameter $\delta = \sqrt{\mathbf{b}'\mathbf{b}/\vartheta}$. Since the sequence of hypothesis H_n^* is contiguous to H_0^*, (A.17) still holds, and therefore we have that

$$\frac{2\,D_0\,M_2}{K_0}\,L_n \to^D \chi^2_{p-q}(\delta).$$

Finally, note that by (A.3) $\delta^2 = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}/\vartheta$, which completes the proof of part (b). ■

Appendix A.3. Proof of Theorem 2

The test statistic Λ is asymptotically χ^2_q according to (13) and up to $O(1/n)$ we have that $\alpha(H_{\epsilon,n}) = 1 - H_q(\eta_{1-\alpha_0}, \delta(\epsilon))$ where $\delta(\epsilon) = n\Lambda(H_{\epsilon,n})$. Let $b(\epsilon) = -H_q(\eta_{1-\alpha_0}, \delta(\epsilon))$, then we have up to $O(1/n)$ that

$$\alpha(H_{\epsilon,n}) - \alpha_0 = b(\epsilon) - b(0) = \epsilon\, b'(0) + \frac{\epsilon^2}{2} b''(0) + o(\epsilon^2).$$

A second order von Mises expansion of $\Lambda(H_{\epsilon,n})$ yields

$$\Lambda(H_{\epsilon,n}) = \Lambda(H) + \frac{\epsilon}{\sqrt{n}} \int \xi_1(\mathbf{x})\, dG(\mathbf{x}) + \frac{1}{2}\frac{\epsilon^2}{n} \iint \xi_2(\mathbf{x}, \mathbf{y})\, dG(\mathbf{x})\, dG(\mathbf{y}) + o(\epsilon^2/n), \tag{A.18}$$

with $\xi_1(\mathbf{x}) = IF(\mathbf{x}, \Lambda, H) = 0$ (see e.g. Fernholz 2001, Gatto and Ronchetti 1996). From (A.18) we immediately obtain $b'(0) = \kappa \frac{\partial \delta}{\partial \epsilon}|_{\epsilon=0} = n\kappa \frac{\partial \Lambda(H_{\epsilon,n})}{\partial \epsilon}|_{\epsilon=0} = 0$, and

$$b''(0) = \kappa \frac{\partial^2 \delta}{\partial \epsilon^2}|_{\epsilon=0} = n\kappa \frac{\partial^2 \Lambda(H_{\epsilon,n})}{\partial \epsilon^2}|_{\epsilon=0} = \kappa \iint \xi_2(\mathbf{x}, \mathbf{y})\, dG(\mathbf{x})\, dG(\mathbf{y}).$$

For $G = \Delta_{\mathbf{y}}$ this expression reduces to $b''(0) = \kappa\xi_2(\mathbf{y}, \mathbf{y}) = \kappa\, IF2(\mathbf{y}, \Lambda, H)$. ■

Appendix A.4. Proof of Theorem 3

The proof follows the same lines as that of Theorem 1 in Salibian-Barrera and Zamar (2002). The basic idea is to note that τ-regression estimators are asymptotically equivalent to an M-estimator as in (6). More specifically, note that $\hat{\boldsymbol{\theta}}_n$ and $\hat{\sigma}_n$ satisfy the fixed-point equation $g_n(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n) = (\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n)$, where g_n is given in (20). A Taylor expansion around the limiting values $\boldsymbol{\theta}_0$ and σ_0 yields

$$(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n) = g_n(\boldsymbol{\theta}_0, \sigma_0) + \nabla g_n(\boldsymbol{\theta}_0, \sigma_0)\,(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n) + R_n.$$

The assumed regularity conditions on the derivatives of ρ_1 and ρ_2 suffice to show that $R_n = o_p(1/\sqrt{n})$, from which we obtain

$$\sqrt{n}((\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n) - (\boldsymbol{\theta}_0, \sigma_0)) = (\mathbf{I} - \nabla g_n(\boldsymbol{\theta}_0, \sigma_0))^{-1}\, \sqrt{n}\,(g_n(\boldsymbol{\theta}_0, \sigma_0) - (\boldsymbol{\theta}_0, \sigma_0)) + o_p(1).$$

Simple calculations along the same lines as in Salibian-Barrera and Zamar (2002) show that the matrix $\mathbf{I} - \nabla g_n(\boldsymbol{\theta}_0, \sigma_0)$corresponds to (21). We only need to show that the bootstrap distribution of $\sqrt{n}\,(g_n(\boldsymbol{\theta}_0, \sigma_0) - (\boldsymbol{\theta}_0, \sigma_0))$ is asymptotically the same as that of $\sqrt{n}\,(g_n^*(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n) - (\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n))$. The argument is the same as that used in Salibian-Barrera and Zamar (2002), which relies on bounding the distance between the corresponding distribution functions using the fact that $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ and $\hat{\sigma}_n \to \sigma_0$ almost surely (see Bickel and Freedman, 1981). ■

Appendix A.5. Proof of (24)

In this section we show that (24) holds for the test statistic (10). It is easy to see that this also holds for (13) and for the Wald, Scores and LRT-type tests in (27), (28) and (29), respectively. Let $\hat{\boldsymbol{\theta}}^*_{n,0}$ and $\hat{\boldsymbol{\theta}}^*_n$ be the bootstrapped regression estimators under the null hypothesis and without restrictions, respectively. Then,

$$\begin{aligned} L^*_n &= n\left(\frac{\hat{\tau}^{2,*}_{n,0} - \hat{\tau}^{2,*}_n}{\hat{\tau}^{2,*}_n}\right) \\ &= n\left(\frac{\hat{\tau}^2_n(\hat{\boldsymbol{\theta}}^*_{n,0}) - \hat{\tau}^2_n(\hat{\boldsymbol{\theta}}^*_n)}{\hat{\tau}^2_n(\hat{\boldsymbol{\theta}}^*_n)}\right) = h_n(\hat{\boldsymbol{\Theta}}^*_n)\,, \end{aligned}$$

where $\hat{\boldsymbol{\Theta}}^*_n = (\hat{\boldsymbol{\theta}}^*_{n,0}, \hat{\boldsymbol{\theta}}^*_n)$ are the estimators fully computed with the bootstrap sample. To simplify the notation, in what follows we will drop the index *. Note that

$$\nabla h_n(\hat{\boldsymbol{\Theta}}_n) = \begin{pmatrix} \frac{\partial h_n(\hat{\boldsymbol{\Theta}}_n)}{\partial \hat{\boldsymbol{\theta}}_{n,0}} \\ \frac{\partial h_n(\hat{\boldsymbol{\Theta}}_n)}{\partial \hat{\boldsymbol{\theta}}_n} \end{pmatrix}.$$

Let $\tau^2_a = \hat{\tau}^2_n(\hat{\boldsymbol{\theta}}_n)$, $r_i(0) = y_i - \mathbf{x}_i'\hat{\boldsymbol{\theta}}_{n,0}$ and $\tilde{r}_i(0) = r_i(0)/s_n(\hat{\boldsymbol{\theta}}_{n,0})$. Then we have

$$\begin{aligned} \frac{\partial h_n(\hat{\boldsymbol{\Theta}}_n)}{\partial \hat{\boldsymbol{\theta}}_{n,0}} &= \frac{1}{\tau^2_a}\nabla\tau^2_n(\hat{\boldsymbol{\theta}}_{n,0}) = \frac{1}{\tau^2_a}\frac{1}{n\,b_2}\Bigg[2\,s_n(\hat{\boldsymbol{\theta}}_{n,0})\,\nabla s_n(\hat{\boldsymbol{\theta}}_{n,0})\sum_{i=1}^n \rho_2\left(\tilde{r}_i(0)\right) \\ &\qquad + s^2_n(\hat{\boldsymbol{\theta}}_{n,0})\sum_{i=1}^n \psi_2\left(\tilde{r}_i(0)\right)\left\{\frac{-\mathbf{x}_i s_n(\hat{\boldsymbol{\theta}}_{n,0}) - r_i(0)\nabla s_n(\hat{\boldsymbol{\theta}}_{n,0})}{s^2_n(\hat{\boldsymbol{\theta}}_{n,0})}\right\}\Bigg]. \end{aligned}$$

By differentiating (5) it is easy to see that

$$\nabla s_n(\hat{\boldsymbol{\theta}}_{n,0}) = -\frac{\sum_{i=1}^n \psi_1(\tilde{r}_i(0))\mathbf{x}_i}{\sum_{i=1}^n \psi_1(\tilde{r}_i(0))\tilde{r}_i(0)}\,.$$

Replacing this in the above expression we obtain

$$\frac{\partial h_n(\hat{\boldsymbol{\Theta}}_n)}{\partial \hat{\boldsymbol{\theta}}_{n,0}} = -\frac{s_n(\hat{\boldsymbol{\theta}}_{n,0})}{\tau^2_a\,n\,b_2}\sum_{i=1}^n\Big[W_n(\hat{\boldsymbol{\theta}}_{n,0})\psi_1(\tilde{r}_i(0)) + \psi_2(\tilde{r}_i(0))\Big]\mathbf{x}_i\,.$$

Note that the equation above is proportional to the estimating equation for $\hat{\boldsymbol{\theta}}_{n,0}$. Similarly,

$$\frac{\partial h_n(\hat{\boldsymbol{\Theta}}_n)}{\partial \hat{\boldsymbol{\theta}}_n} = -\frac{2\,\hat{\tau}^2_n(\hat{\boldsymbol{\theta}}_{n,0})\,\hat{\tau}_n(\hat{\boldsymbol{\theta}}_n)\,\nabla\hat{\tau}_n(\hat{\boldsymbol{\theta}}_n)}{\hat{\tau}^2_n(\hat{\boldsymbol{\theta}}_n)^2} = -\frac{\hat{\tau}^2_n(\hat{\boldsymbol{\theta}}_{n,0})}{\hat{\tau}^2_n(\hat{\boldsymbol{\theta}}_n)^2}\,\nabla\hat{\tau}^2_n(\hat{\boldsymbol{\theta}}_n)\,.$$

As before, we have that $\nabla\hat{\tau}^2_n(\hat{\boldsymbol{\theta}}_n)$ is proportional to the score equations for the unrestricted estimator $\hat{\boldsymbol{\theta}}_n$. Hence $\nabla h^*_n(\hat{\boldsymbol{\Theta}}^*_n) = o_p(1)$. ■

Acknowledgements

Matias Salibian-Barrera was partially supported by a Discovery Grant of the Natural Sciences and Engineering Research Council of Canada and a grant of the Fund for Scientific Research-Flanders. Stefan Van Aelst was partially supported by a grant of the Fund for Scientific Research-Flanders and by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy).

Bibliography

[1] Baayen, R.H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge University Press.

[2] Berrendero, J.R. and Zamar, R.H. (2001). Maximum bias curves for robust regression with non-elliptical regressors. *The Annals of Statistics*, **29**, 224–251.

[3] Berrendero, J.R., Mendes, B.V.M. and Tyler, D.E. (2007). On the maximum bias functions of MM-estimates and constrained M-estimates of regression. *The Annals of Statistics*, **35**, 13–40.

[4] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**, 1196–1217.

[5] Croux, C., Filzmoser, P., and Joossens, K. (2008). Classification efficiencies for robust linear diiscriminant analysis. *Statistica Sinica*, **18**, 581–599.

[6] De Vaan, L., Schreuder, R. and Baayen, R. H. (2007) Regular morphologically complex neologisms leave detectable traces in the mental lexicon, *The Mental Lexicon*, **2**, 1-23.

[7] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, **7**, 1-26.

[8] Fernholz, L.T. (2001). On multivariate higher order von Mises expansions. *Metrika,* **53**, 123-140.

[9] Fisher, N.I. and Hall, P. On bootstrap hypothesis testing. (1990). *The Australian Journal of Statistics*, **32**, 177-190.

[10] Gatto, R. and Ronchetti, E. (1996). General saddlepoint approximations of marginal densities and tail probabilities. *Journal of the American Statistical Association,* **91**, 666-673.

[11] Hall, P. and Wilson, S.R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, **47**, 757-762.

[12] Hampel, F.R., Ronchetti, E.Z., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics. The approach based on influence functions*. New York: Wiley.

[13] Heritier, S. and Ronchetti, E. (1994), Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association*, **89**, 897–904.

[14] Khan, J.A., Van Aelst, S., and Zamar, R.H. (2010). Fast Robust Estimation of Prediction Error Based on Resampling. *Computational Statistics and Data Analysis*, **54** (12), 3121-3130.

[15] Markatou, M. and He, X. (1994). Bounded influence and high breakdown point testing procedures in linear models. *Journal of the American Statistical Association*, **89**, 543-549.

[16] Markatou, M. and Hettmansperger, T.P. (1990). Robust bounded-influence tests in linear models. *Journal of the American Statistical Association*, **85**, 187-190.

[17] Markatou, M., Stahel, W.A., and Ronchetti, E. (1991). Robust M-type testing procedures for linear models. In *Directions in Robust Statistics and Diagnostics*, Part I, Stahel, W., Weisberg, S., eds. Springer-Verlag. pp. 201-220.

[18] Maronna, R.A., Martin, D.R., and Yohai, V.J. (2006), *Robust statistics: theory and methods*, New York: John Wiley and Sons.

[19] Martin, R.D., Yohai, V.J. and Zamar, R.H. (1989). Min-max bias robust regression. *The Annals of Statistics*, **17**, 1608–1630.

[20] Mendes, B. and Tyler, D.E. (1996). Constrained *M*-estimation for regression, in *Robust statistics, Data analysis, and Computer Intensive Methods*, ed. H. Rieder, Lecture Notes in Statistics, 109, 299-320. Springer, New York.

[21] Muler, N. and Yohai, V.J. (2002). Robust estimates for ARCH processes. Journal of Time Series Analysis, **23**(3), 341-375.

[22] Miller, G. A. (1990). Wordnet: An on-line lexical database. *International Journal of Lexicography*, **3**, 235312.

[23] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

[24] Roelant, E., Van Aelst, S., and Croux, C. (2009). Multivariate Generalized S-estimators. *Journal of Multivariate Analysis*, **100** (5), 876-887.

[25] Ronchetti, E. (1982). Robust Testing in Linear Models: the Infinitesimal Approach," Unpublished Ph.D. thesis, ETH Zürich.

[26] Ronchetti, E., and Trojani, F. (2001). Robust Inference with GMM Estimators. *Journal of Econometrics,* **101**, 36–79.

[27] Rousseeuw, P.J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, **79**, 871-880.

[28] Rousseeuw, P.J. and Yohai, V.J. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series*. (J. Franke, W. Hardle and D. Martin, eds.). *Lecture Notes in Statist.*, **26** 256-272. Berlin: Springer-Verlag.

[29] Salibian-Barrera, M. (2005). Estimating the p-values of robust tests for the linear model. *Journal of Statistical Planning and Inference,* 128 (1), 241–257.

[30] Salibian-Barrera, M., and Van Aelst, S. (2008). Robust Model Selection Using Fast and Robust Bootstrap. *Computational Statistics and Data Analysis*, **52**, 5121-5135.

[31] Salibian-Barrera, M., Van Aelst, S., and Willems, G. (2006). PCA based on multivariate MM-estimators with fast and robust bootstrap. *Journal of the American Statistical Association,* 101 (475), 1198 - 1211.

[32] Salibian-Barrera, M., Willems, G. and Zamar, R.H. (2008). The fast-τ estimator for regression. To appear in the *Journal of Computational and Graphical Statistics*.

[33] Salibian-Barrera, M. and Zamar, R.H. (2002). Bootstrapping robust estimates of regression. *The Annals of Statistics*, **30**, 556-582.

[34] Samanta, M. and Welsh, A.H. (2013). Bootstrapping for highly unbalanced clustered data. *Computational Statistics and Data Analysis*, **59**, 70-81. DOI: 10.1016/j.csda.2012.09.004.

[35] Van Aelst, S. and Willems, G. (2002). Robust Bootstrap for S-Estimators of Multivariate regression. In *Statistical Data Analysis Based on the L1 Norm and Related Methods*. Y. Dodge, Ed., Basel: Birkhauser, pp. 201-212.

[36] Van Aelst, S., and Willems, G. (2005). Multivariate regression S-Estimators for robust estimation and inference. *Statistica Sinica,* 15 (4), 981-1001.

[37] Van Aelst, S., and Willems, G. (2011). Robust and efficient one-way MANOVA tests. *Journal of the American Statistical Association*, **106**, 706–718.

[38] Van Aelst, S., Willems, G., and Zamar, R.H. (2013). Robust and efficient estimation of the residual scale in linear regression. *Journal of Multivariate Analysis*, **116**, 278–296.

[39] Wang, L. and Qu, A. (2007), "Robust Tests in Regression Models With Omnibus Alternatives and Bounded Influence," *Journal of the American Statistical Association,* **102**, 347–358.

[40] Yohai, V.J. (1985). High breakdown-point and high efficiency robust estimates for regression. Technical Report No 66, Department of Statistics, University of Washington. Available at http://www.stat.washington.edu/research/reports/1985/tr066.pdf

[41] Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, **15**, 642-656.

[42] Yohai,V.J. and Zamar, R.H. (1986). High breakdown point and high efficiency estimates of regression by means of the minimization of an efficient scale. Technical Report No 84, Department of Statistics, University of Washington. Available at http://www.stat.washington.edu/research/reports/1986/tr084.pdf

[43] Yohai,V.J. and Zamar, R.H. (1988) High breakdown point and high efficiency estimates of regression by means of the minimization of an efficient scale. J. Amer. Statist. Assoc. 83 406–413.

[44] Yohai, V.J. and Zamar, R.H. (1997). Optimal locally robust M-estimates of regression *Journal of Statistical Planning and Inference,* **64**, 309–323.