Saddlepoint tests for quantile regression

Journal Item

How to cite:

For guidance on citations see FAQs.

© 2016 Statistical Society of Canada

Version: Accepted Manuscript

Link(s) to article on publisher’s website: http://dx.doi.org/doi:10.1002/cjs.11288

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
Saddlepoint tests for quantile regression

Elvezio Ronchetti and Radka Sabolová

1 Research Center for Statistics and Geneva School of Economics and Management, University of Geneva, Switzerland
2 Department of Mathematics and Statistics, The Open University, Milton Keynes, UK

Key words and phrases: Saddlepoint techniques; quantile regression; M-estimator; robust tests.
MSC 2010: Primary 62J05; secondary 62F05

Abstract: Quantile regression is a flexible and powerful technique which allows to model the quantiles of the conditional distribution of a response variable given a set of covariates. Regression quantile estimators can be viewed as M-estimators and standard asymptotic inference is readily available based on likelihood-ratio, Wald, and score-type test statistics. However, these statistics require the estimation of the sparsity function $s(\alpha) = [g(G^{-1}(\alpha))]^{-1}$, where $G$ and $g$ are the cumulative distribution function and the density of the regression errors respectively, and this can lead to nonparametric density estimation. Moreover, the asymptotic $\chi^2$ distribution for these statistics can provide an inaccurate approximation of tail probabilities and this can lead to inaccurate p-values, especially for moderate sample sizes. Alternative methods which do not require the estimation of the sparsity function, include rank techniques and resampling methods to obtain confidence intervals, which can be inverted to test hypotheses. These are typically more accurate than the standard M-tests.

In this paper we show how accurate tests can be obtained by using a nonparametric saddlepoint test statistic. The proposed statistic is asymptotically $\chi^2$ distributed, does not require the specification of the error distribution, and does not require the estimation of the sparsity function. The validity of the method is demonstrated through a simulation study, which shows both the robustness and the accuracy of the new test compared to the best available alternatives.
1. INTRODUCTION

Introduced in a seminal paper by Koenker & Bassett (1978), quantile regression has become a standard tool in statistical methodology and practice. Instead of modelling the conditional expectation of the response given the covariates, it models the $\alpha$-quantiles of the conditional distribution and provides a richer information on the underlying relationship between the response and the covariates. From the original formulation for the standard regression model, many extensions have been provided, including generalized linear models, survival data, autoregressive models, penalized methods, and nonparametric regression. Moreover, many applications in various fields ranging from economics and finance to biology and ecology have been developed. An excellent overview on theoretical, computational, and applied aspects is given in the book Koenker (2005).

Let $Y_1, \ldots, Y_n$ be observations following the regression model

$$Y_i = x_i^\top \beta + u_i, \quad i = 1, \ldots, n$$

where $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \in \mathbb{R}^p$, $x_{i1} \equiv 1$, $\beta \in \mathbb{R}^p$, and $u_i \sim G$ with density $g$. Notice that our test will not require to specify $G$. Later we will denote by $X$ the design matrix with $i$-th row $x_i$. Although a more general model with non-iid errors would be more useful in the context of regression quantiles, we derive our results in the simpler model given above, but we provide some numerical results for a location-scale model in Table 10, where the standard deviation of the errors depends linearly on the $x$'s.

The regression quantile estimator $\hat{\beta}_\alpha$ for $\beta$ is the solution of the minimization problem

$$\hat{\beta}_\alpha = \arg\min_{\beta} \sum_{i=1}^n \rho_\alpha(Y_i - x_i^\top \beta),$$

(1)

where

$$\rho_\alpha(u) = |u| \{(1 - \alpha) I[u < 0] + \alpha I[u > 0]\}.$$ 

It is an M-estimator defined by (9) with score function

$$\psi(y; \beta) = \psi_\alpha(y - x^\top \beta),$$

(2)

where

$$\psi_\alpha(u) = \alpha I[u > 0] - (1 - \alpha) I[u < 0] = \alpha - I[u < 0].$$

(3)

The estimator $\hat{\beta}_\alpha$ is consistent for $\beta_\alpha = (\beta_1 + G^{-1}(\alpha), \beta_2, \ldots, \beta_p)$. Since its influence function (Hampel (1974), Hampel et al. (1986)), which is proportional to the score function (2), is bounded, the regression quantile estimator is robust against moderate deviations from the underlying error distribution when the $x$’s are not too discordant; see He et al. (1990).

From the inferential point of view, an exact formula for the joint density of the regression quantile estimator is available; see Koenker (2005), Theorem 3.1, p. 70, Jurečková (2010), Portnoy (2012). However, from an operational point of view it requires the computation of $\binom{n}{p}$ terms and this becomes unfeasible in many applications. Accurate finite sample and saddlepoint approximations to the density of regression quantiles are available; see Spady (1991), De Jongh et al. (2004). However, the joint density would have to be marginalized to make inference on

*Author to whom correspondence may be addressed.
E-mail: radka.sabolova@open.ac.uk
single components, and this would be numerically very challenging. Therefore, inference on the parameters is typically carried out through the approximation provided by the asymptotic normal distribution of the estimator. Notice that the asymptotic variance (8) depends on the so-called sparsity function $s(\alpha) = [g(G^{-1}(\alpha))]^{-1}$, whose estimation is challenging and can lead to nonparametric density estimation. A variety of methods to construct confidence intervals not requiring density estimation are available. They include the inversion of rank-tests (Gutenbrunner & Jurečková (1992) and Gutenbrunner et al. (1993)), a range of resampling methods (Parzen, Wei, & Ying (1994)), Markov Chain Marginal Bootstrap (Hu & He (2012)), jackknife (Portnoy (2014)), and a “direct” method by Zhou & Portnoy (1996).

$M$–tests, i.e. tests based on $M$–estimators can also be used. They are the natural counterpart of standard Wald-, score-, and likelihood-ratio tests, where - loglikelihood is replaced by the objective function $\rho_\alpha(\cdot)$ in (1), and the score function by $\psi_\alpha(\cdot)$ in (3). Their robustness and asymptotic properties have been studied in Heritier & Ronchetti (1994). It turns out that a bounded score function guarantees robustness of validity and robustness of efficiency for these tests. They all require the estimation of the sparsity function. Finally, notice that in this case score tests are a class of generalized rank tests. Therefore, rank-based inference can be carried out, but in practice one still has to rely on the asymptotic distribution of the test statistic.

$P$–values based on the asymptotic distribution of the regression quantiles estimator or of the $M$–test statistics discussed above can be misleading when the sample size is moderate and/or when small tail probabilities are required.

In this paper we focus on hypotheses testing and develop a so-called saddlepoint test, which exhibits several desirable properties especially in small samples. The test statistic is asymptotically $\chi^2$ distributed under the null hypothesis and is therefore first-order equivalent to the standard $M$–tests. However, it exhibits a better finite sample behavior than the latter by combining excellent accuracy even in small samples and robustness. The test statistic is given by an explicit formula, is nonparametric, and it does not require the estimation of the sparsity function. It is derived from the results in Robinson, Ronchetti, & Young (2003) for $M$–estimators, which were obtained using saddlepoint techniques (Daniels (1954)) and can be viewed as an empirical likelihood procedure based on tilted exponential weights; cf. the discussion in Ma & Ronchetti (2011), p. 148. The corresponding weights are different from those obtained by the standard empirical likelihood approach by Owen (1988), Owen (2001).

The paper is organized as follows. In Section 2. we consider the case of a simple hypothesis and we derive the saddlepoint test statistic in the parametric case. This is useful to understand the construction of the new statistic. Then, we obtain the test statistic in its nonparametric version, i.e. when we do not specify the errors distribution. Section 3. is devoted to the composite hypothesis case. Compared to the simple hypothesis case where the formula is explicit, here we need an additional numerical minimization over the nuisance parameters. The simulation study of Section 4. shows the excellent accuracy and robustness properties of the nonparametric saddlepoint test statistic in finite samples. Comparisons with Wald, likelihood-ratio type, rank tests, resampling techniques, and a “direct” method are provided. Overall the $\chi^2_0$ quantiles are very close to the quantiles of the distribution of the nonparametric saddlepoint test statistic. Even in the extreme situation of 21 observations, 6 parameters, and under a spectrum of distributions for the errors, the accuracy of the new test is still good. While the standard $M$– tests break down, the saddlepoint test has good accuracy comparable to the best resampling or rank methods. Finally, in Section 5. we provide some conclusions and discuss possible extensions of this work. In the Appendix, we summarize for completeness the definition of the saddlepoint test statistic for $M$–estimators and its properties as developed in Robinson, Ronchetti, & Young (2003) and we give the assumptions and the proofs of the Propositions.
2. SIMPLE HYPOTHESIS

Consider the simple hypothesis \( H_0 : \beta_\alpha = \beta_{\alpha_0} \). Although we will mostly use only its nonparametric version, it is useful to provide first the derivation of the test statistic in the parametric setup.

2.1. Parametric case

To derive the test statistic in this case, we assume for convenience that \((Y_i, x_i)\) are independent identically distributed with density \( g(y_i - x_i^T \beta)k(x_i) \), where \( k(\cdot) \) is the density of \( x_i \). We will not have to specify the latter, because the final test statistic will be an expectation with respect to \( k(\cdot) \) and it will simply be replaced by the average over the \( x_i \)’s.

We now proceed to compute the test statistic \( 2nh(\hat{\beta}_\alpha) = 2n \sup_{\lambda} \{-K_\psi(\lambda; \hat{\beta}_\alpha)\} \) (see Appendix), where \( K_\psi(\lambda; \hat{\beta}_\alpha) \) is the cumulant generating function of the score function \( \psi(Y_i; \beta) \) defined by (2) and (3) corresponding to the regression quantile and given by

\[
K_\psi(\lambda; \beta_\alpha) = \log \mathbb{E}e^{\lambda^T \psi(Y_i; \beta_\alpha)X_i} = \log \mathbb{E}e^{\lambda^T X_i (\alpha - [Y_i - x_i^T \beta_\alpha] \beta_\alpha < 0)} \\
= \log \int \int e^{\lambda^T X_i} e^{-\lambda^T X_i [y_i - x_i^T \beta_\alpha] \beta_\alpha < 0} g(y_i - x_i^T \beta)k(x_i) dy_i \, dx_i \\
= \log \int \left[ \int_{-\infty}^{\alpha^T X_i} e^{\lambda^T X_i} g(y_i - x_i^T \beta)k(x_i) dy_i + \int_{\alpha^T X_i}^{\infty} e^{\lambda^T X_i} g(y_i - x_i^T \beta)k(x_i) dy_i \right] dx_i \\
= \log \left\{ e^{\alpha^T X_i k(x_i)} \left[ e^{-\lambda^T X_i G(x_i^T (\beta_\alpha - \beta)) + 1 - G(x_i^T (\beta_\alpha - \beta))} \right] \right\} dx_i.
\]

In order to compute the saddlepoint, we need the derivative of \( K_\psi(\lambda; \beta_\alpha) \) with respect to \( \lambda \):

\[
\frac{\partial K_\psi(\lambda; \beta_\alpha)}{\partial \lambda} = \frac{\partial}{\partial \lambda} \log \left\{ e^{\alpha^T X_i k(x_i)} \left[ e^{-\lambda^T X_i G(x_i^T (\beta_\alpha - \beta)) + 1 - G(x_i^T (\beta_\alpha - \beta))} \right] \right\} dx_i \\
= \int \left\{ \frac{\partial}{\partial \lambda} e^{\alpha^T X_i k(x_i)} \left[ e^{-\lambda^T X_i G(x_i^T (\beta_\alpha - \beta)) + 1 - G(x_i^T (\beta_\alpha - \beta))} \right] \right\} dx_i \\
= \left\{ \int \left[ \alpha x_i e^{\alpha^T X_i k(x_i)} \left[ e^{-\lambda^T X_i G(x_i^T (\beta_\alpha - \beta)) + 1 - G(x_i^T (\beta_\alpha - \beta))} \right] + \\
e^{\alpha^T X_i k(x_i)} \left[ -x_i e^{-\lambda^T X_i G(x_i^T (\beta_\alpha - \beta))} \right] \right\} dx_i \right\} \times \\
\left\{ \int \left[ e^{\alpha^T X_i k(x_i)} \left[ e^{-\lambda^T X_i G(x_i^T (\beta_\alpha - \beta)) + 1 - G(x_i^T (\beta_\alpha - \beta))} \right] \right\} dx_i \right\}^{-1}.
\]
By solving the equation
\[
\frac{\partial K_\psi(\lambda, \beta_\alpha)}{\partial \lambda} = 0
\]
we obtain that \( \lambda(\hat{\beta}_\alpha)^\top x_i \) must satisfy
\[
\lambda(\hat{\beta}_\alpha)^\top x_i = -\log \left\{ \frac{\alpha}{1 - \alpha} \frac{1 - G \left( x_i^\top (\hat{\beta}_\alpha - \beta) \right)}{G \left( x_i^\top (\hat{\beta}_\alpha - \beta) \right)} \right\}, \quad i = 1, \ldots, n.
\]

Therefore, under the null hypothesis \( \beta_\alpha = \beta_{\alpha 0} \), we get
\[
h(\hat{\beta}_\alpha) = -K_\psi(\lambda(\hat{\beta}_\alpha); \hat{\beta}_\alpha)
\]
\[
= -\log \int \left( \frac{G \left( x_i^\top (\hat{\beta}_\alpha - \beta_0) \right)}{\alpha} \right)^\alpha \left( \frac{1 - G \left( x_i^\top (\hat{\beta}_\alpha - \beta_0) \right)}{1 - \alpha} \right)^{1 - \alpha} k(x_i) \, dx_i
\]
and
\[
2nh(\hat{\beta}_\alpha) = 2\log \left\{ \left( \frac{G \left( x_i^\top (\hat{\beta}_\alpha - \beta_0) \right)}{\alpha} \right)^\alpha \left( \frac{1 - G \left( x_i^\top (\hat{\beta}_\alpha - \beta_0) \right)}{1 - \alpha} \right)^{1 - \alpha} \right\},
\]
where \( \beta_0 \) is a regression parameter corresponding to \( \beta_{\alpha 0} \), i.e. \( \beta_0 = \beta_{\alpha 0} - (G^{-1}(\alpha), 0, \ldots, 0)^\top \).

We now replace the expectation over \( x \) by the average over the observed \( x_i \)'s and the next Proposition shows that the resulting test statistic is asymptotically \( \chi^2_p \) under the null hypothesis.

**Proposition 2.1.** Under the Assumptions given in the Appendix,
\[
2nh(\hat{\beta}_\alpha) = -2n \log \frac{1}{n} \sum_{i=1}^{n} \left( \frac{G \left( x_i^\top (\hat{\beta}_\alpha - \beta_0) \right)}{\alpha} \right)^\alpha \left( \frac{1 - G \left( x_i^\top (\hat{\beta}_\alpha - \beta_0) \right)}{1 - \alpha} \right)^{1 - \alpha} \overset{D}{\rightharpoonup} \chi^2_p, \quad (4)
\]
under the null hypothesis.

The proof is given the Appendix.

The test statistic for the special case of simple quantiles can be easily obtained from (4) by setting \( p = 1, x \equiv 1, \hat{\beta}_\alpha = F_n^{-1}(\alpha) \), the empirical quantile, and \( G(\hat{\beta}_\alpha - \beta_0) = F_0(\hat{\beta}_\alpha) \), the cumulative distribution of the observations under the null hypothesis. In this case the test statistic is simply
\[
2nh(\hat{\beta}_\alpha) = -2n \log \left\{ \left( \frac{F_0(\hat{\beta}_\alpha)}{\alpha} \right)^\alpha \left( \frac{1 - F_0(\hat{\beta}_\alpha)}{1 - \alpha} \right)^{1 - \alpha} \right\}. \quad (5)
\]
2.2. Nonparametric case

If we do not want to specify the distribution \( G \) of the errors, we can derive a nonparametric version of the saddlepoint test statistic. Its derivation is provided in the Appendix and is given by

\[
2n \hat{h}(\hat{\beta}_\alpha) = -2 \sum_{j=1}^{n} \log \left\{ \left( \sum_{i=1}^{n} w_{ij} \right)^{\frac{\alpha}{\alpha}} \left( \sum_{i=1}^{n} w_{ij} (1 - I_i) \right)^{1-\alpha} \right\},
\]

where

\[
w_{ij} = \frac{\left( \frac{\alpha}{1-\alpha} \right) F_{jn}(\hat{\beta}_{\alpha}) \tilde{I}_{ij}(\alpha, \beta) \alpha}{\sum_{k=1}^{n} \left( \frac{\alpha}{1-\alpha} \right) F_{kn}(\hat{\beta}_{\alpha})},
\]

\[
r_i = y_i - x_i^\top \hat{\beta}_\alpha,
\]

\[
I_i = I[r_i < 0],
\]

\[
I_{ij}(\beta) = I[y_i + x_j^\top (\hat{\beta}_\alpha - \beta) < 0] = I[y_i - x_j^\top \beta < 0],
\]

\[
F_{jn}(\beta) = \frac{1}{n} \sum_{i=1}^{n} I_{ij}(\beta), \quad i, j = 1, \ldots, n.
\]

The nonparametric saddlepoint test statistic for the special case of simple quantiles \( p = 1, x_i \equiv 1 \) is given by

\[
2n \hat{h}(\hat{\beta}_\alpha) = -2n \log \left\{ \left( \sum_{i=1}^{n} w_{i} I_i \right)^{\alpha} \left( \sum_{i=1}^{n} w_{i} (1 - I_i) \right)^{1-\alpha} \right\},
\]

where

\[
w_{i} = \frac{\left( \frac{\alpha}{1-\alpha} \right) F_i(\hat{\beta}_{\alpha}) \tilde{I}_i(\alpha, \beta) \alpha}{\sum_{k=1}^{n} \left( \frac{\alpha}{1-\alpha} \right) F_k(\hat{\beta}_{\alpha})},
\]

\[
I_i = I[r_i < 0] = I[y_i < \hat{\beta}_\alpha],
\]

\[
\tilde{I}_i(\beta) = I[y_i < \beta],
\]

\[
F_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \tilde{I}_i(\beta),
\]

\[
\hat{\beta}_\alpha = F_n^{-1}(\alpha).
\]

We prove the asymptotic distribution of the test statistic in the case of simple quantiles. The proof shows the structure of the test statistic (7), which is based on weighted sums of the \( I_i^\prime \)s and
(1 − I_j)'s, where the weights are given by

\[ w_i = \begin{cases} 
\frac{1}{n} \left( \frac{n}{F_n(\beta_{\alpha 0})} \right) : y_i < \beta_{\alpha 0} \\
\frac{1}{n} \left( \frac{1 - \alpha}{1 - F_n(\beta_{\alpha 0})} \right) : y_i > \beta_{\alpha 0}, 
\end{cases} \]

which are close to \( \frac{1}{n} \) when \( n \) increases. The proof for regression quantiles is similar with a more complicated notation.

**Proposition 2.2** Under the Assumptions given in the Appendix, the test statistic \( 2n h(\hat{\beta}_\alpha) \) defined by (7) converges in distribution to a \( \chi_1^2 \) as \( n \to \infty \) under the null hypothesis.

The proof is given the Appendix.

### 3. COMPOSITE HYPOTHESIS

Suppose we now want to perform a test only for the first subvector of the regression quantile, \( H_0 : \beta_{\alpha 1} = \beta_{\alpha 10} \in \mathbb{R}^{p_1} \), where \( \beta_{\alpha} = (\beta_{\alpha 1, \beta_{\alpha 2}})^T \) and \( \hat{\beta}_{\alpha} = (\hat{\beta}_{\alpha 1}, \hat{\beta}_{\alpha 2})^T \). We consider this hypothesis for simplicity of notation, but more general hypotheses such as those on functions of the parameters can be treated, as presented in Robinson, Ronchetti, & Young (2003).

Denote by \( x_{i1} \in \mathbb{R}^{p_1} \) the subvector of \( x_i \) consisting of the first \( p_1 \) components of \( x_i \). Here we derive directly the nonparametric test. Let

\[
\beta_{\alpha 2} = \arg\min_{\beta_{\alpha 2}} \left\{ -\log \left( \frac{1}{n} \sum_{i=1}^{n} \exp \left[ \lambda^T \psi(\beta_{\alpha 10}, \beta_{\alpha 2}) \right] \right) \right\} \\
= \arg\min_{\beta_{\alpha 2}} \left\{ -\log \left( \frac{1}{n} \sum_{i=1}^{n} \exp \left[ \alpha - I[y_i - x_{i1}^T \beta_{\alpha 10} - x_{i2} \beta_{\alpha 2} < 0] \right] x_i \right) \right\}
\]

and \( \beta^* = (\beta_{\alpha 10}, \beta_{\alpha 2}) \). Then, by following the same development as in the simple hypothesis case, we can see that

\[
x_j^T \mu = \log \frac{1 - \alpha}{\alpha} \frac{F_{\alpha 1}^2(\beta^*)}{1 - F_{\alpha 1}^2(\beta^*)}
\]

solves the equation

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha - I[r_i + x_j^T (\hat{\beta}_\alpha - \beta^*)]) x_j e^{(\alpha - I[r_i + x_j^T (\hat{\beta}_\alpha - \beta^*)] x_j^T \mu} = 0,
\]

and the weights are given by

\[
w_{ij} = \frac{\left( \frac{\alpha - 1 - F_{\alpha 1}^2(\beta^*)}{1 - F_{\alpha 1}^2(\beta^*)} \right) I_{ij}(\beta^*)}{\sum_{k=1}^{n} \left( \frac{\alpha - 1 - F_{\alpha 1}^2(\beta^*)}{1 - F_{\alpha 1}^2(\beta^*)} \right) I_{kj}(\beta^*)}.
\]

Hence

\[
x_j^T \lambda(\hat{\beta}_{\alpha 1}, \beta_{\alpha 2}) = \log \left\{ \frac{1 - \alpha}{\alpha} \frac{\sum_{i=1}^{n} w_{ij} I_i(\beta_{\alpha 2})}{\sum_{i=1}^{n} w_{ij} (1 - I_i(\beta_{\alpha 2}))} \right\},
\]

DOI: 7
where
\[ I_i(\beta_{a2}) = I[y_i - x_{i1}\hat{\beta}_{a1} - x_{i2}\beta_{a2} < 0]. \]

Finally, we obtain
\[
K^J(\hat{\beta}_{a1}, \beta_{a2}), \lambda(\hat{\beta}_{a1}, \beta_{a2})) = \log \left\{ \left( \sum_{i=1}^{n} w_{ij} I_i(\beta_{a2}) \right)^{\alpha} \left( \sum_{i=1}^{n} w_{ij} (1 - I_i(\beta_{a2})) \right)^{1-\alpha} \right\},
\]
\[
h(\hat{\beta}_{a1}) = \inf_{\beta_{a2}} - \frac{1}{n} \sum_{j=1}^{n} \log \left\{ \left( \sum_{i=1}^{n} w_{ij} I_i(\beta_{a2}) \right)^{\alpha} \left( \sum_{i=1}^{n} w_{ij} (1 - I_i(\beta_{a2})) \right)^{1-\alpha} \right\}
\]
and under null hypothesis
\[
2nh(\hat{\beta}_{a1}) = -2 \sum_{j=1}^{n} \log \left\{ \left( \sum_{i=1}^{n} w_{ij} I_i(\beta_{a2}) \right)^{\alpha} \left( \sum_{i=1}^{n} w_{ij} (1 - I_i(\beta_{a2})) \right)^{1-\alpha} \right\}
\]

Notice that the minimization over the nuisance parameters can be computationally challenging, especially in moderate to high dimensions.

4. SIMULATION STUDY

In order to demonstrate the accuracy and the robustness of the saddlepoint tests for regression quantiles, we performed a simulation study in two setups and we compared the new test with a variety of other available tests.

In the first set of simulations, we considered a fixed balanced design, i.e. we set the i-th row of the matrix \(X\) to be \((1, \frac{i-1}{n}).\) To study the behavior of the different tests across a spectrum of different distributions, we simulated the errors from a normal, a contaminated normal (obtained as a mixture of a standard normal with another normal with larger variance), a Laplace, and a logistic distribution. The true parameter value for \(\beta\) was \((3, 2)^\top\), \(p = 2\), the sample sizes for the parametric case \(n = 5, 10, 20, 50, 100, 300, 1000, 10000\) and for the nonparametric case \(n = 21, 51, 101\). We tested the null hypothesis \(H_0 : \beta_a = (3 + G_{N(0,1)}(\alpha), 2)^\top\). We performed 50000 simulations and we considered the quantile \(\alpha = .25\) both for the parametric and the nonparametric case. We computed the saddlepoint test defined by (4) in the parametric case and by (6) in the nonparametric case and we compared them with the Wald test. In the parametric case, the tests are calibrated at the normal model, i.e. the asymptotic variance for the Wald test is computed at the normal model and the parametric saddlepoint test is computed by (4), where \(G(\cdot)\) is the normal distribution. In the nonparametric case, the asymptotic covariance matrix of \(\hat{\beta}_{a1}\) used in the Wald test was estimated using the formula from Koenker & Bassett (1978) (implemented in the function `summary.rq` in R).

The results of the simulations are summarized in Table 1 and Figures 1, 2.

We plotted the percentage (out of 50000) of simulated test statistics (dots for Wald, stars for saddlepoint) that were smaller than the \(\chi^2_{2,0.9}, \chi^2_{2,0.95}\) and \(\chi^2_{2,0.99}\) respectively versus the logarithm of the sample size. The exact figures for the parametric case can be found in Table 1. The Monte Carlo standard error is always smaller than .001.

In the parametric case (Figure 1) both tests behave similarly with the saddlepoint test a bit better for small sample sizes. The figure also shows the robustness of both tests which are calibrated at the normal model, but behave reasonably well across a spectrum of long-tailed distributions (with the exception of the Laplace, last row). In the nonparametric case (Figure 2) the Wald
test lacks accuracy everywhere. On the other side, the nonparametric saddlepoint test exhibits excellent accuracy and robustness across all the distributions and even down to small sample sizes.

\begin{table}
\centering
\begin{tabular}{ccccccccc}
\hline
 & \multicolumn{3}{c}{$n = 5$} & \multicolumn{3}{c}{$n = 10$} & \multicolumn{3}{c}{$n = 20$} \\
\hline
 & 0.90 & 0.95 & 0.99 & 0.90 & 0.95 & 0.99 & 0.90 & 0.95 & 0.99 \\
\hline
\text{N} & \text{SAD} & 0.9524 & 0.9842 & 0.9980 & 0.9284 & 0.9700 & 0.9963 & 0.9142 & 0.9614 & 0.9941 \\
& \text{Wald} & 0.9187 & 0.9599 & 0.9907 & 0.9077 & 0.9533 & 0.9896 & 0.9028 & 0.9515 & 0.9887 \\
& \text{Lap} & \text{cont} & \text{Log} & \text{LR} & 0.8331 & 0.8650 & 0.9552 & 0.8434 & 0.8963 & 0.9522 \\
& \text{SAD} & 0.9225 & 0.9655 & 0.9949 & 0.8974 & 0.9490 & 0.9903 & 0.8880 & 0.9434 & 0.9873 \\
& \text{Wald} & 0.8730 & 0.9221 & 0.9902 & 0.8692 & 0.9233 & 0.9721 & 0.8728 & 0.9270 & 0.9771 \\
& \text{Lap} & \text{cont} & \text{Log} & \text{LR} & 0.8406 & 0.9130 & 0.9765 & 0.8292 & 0.8933 & 0.9613 \\
& \text{SAD} & 0.7849 & 0.8405 & 0.9085 & 0.7940 & 0.8531 & 0.9210 & 0.7959 & 0.8605 & 0.9320 \\
& \text{Wald} & 0.7849 & 0.8405 & 0.9085 & 0.7940 & 0.8531 & 0.9210 & 0.7959 & 0.8605 & 0.9320 \\
\hline
\end{tabular}
\caption{Parametric case: $H_0 : \beta = (3 + G_{N(0,1)}^{-1}(\alpha), 2)^T$, $\alpha = 0.25$, $N = 50000$; data generated from: N(0,1), $8 \times N(0,1) + .2 \times N(0,9)$, logistic, and Laplace. The frequencies of accepting $H_0$ under the null hypothesis are reported.}
\end{table}

In the second set of simulations, we focussed only on the nonparametric case, with $p = 6$, $n = 21, 31, 51$, and 50000 replicates. The $i$-th row of the design matrix $X$ was set to $x_i = (x_{i1}, \ldots, x_{i6})$, where $x_{i1} = 1$ and $x_{ij} \sim U(0, 1)$, $j = 2, 3, \ldots, 6$. The true value of the parameter $\beta$ was set to $\beta = (3, 1, 2, 3, 4, 5)^T$. We test the null hypothesis $H_0 : \beta = (3 + G_{N(0,1)}^{-1}(\alpha), 1, 2, 3, 4, 5)^T$. The errors $u_i$, $i = 1, \ldots, n$ were generated from two distributions with the same $\alpha$-quantile: normal distribution $N(0,1)$ and contaminated normal distribution $N(0,\epsilon)$ with $N(0,9)$ ($\epsilon = 0.2$) and the simulations were carried out for different values of $\alpha$: $0.1, 0.15, 0.25$ and 0.5.

In addition to the nonparametric saddlepoint test, we considered seven alternative tests.

(i) The nonparametric version of the Wald test (Wald), as in the previous simulation.
(ii) The likelihood ratio type test (LR) defined by the test statistic

$$LR_n = 2 \frac{B}{A} \sum_{i=1}^{n} \left\{ \rho_\alpha(y_i - \beta_{\alpha}^T \hat{\beta}) - \rho_\alpha(y_i - \beta_{\alpha}^T \hat{\beta}) \right\},$$

where $A = \text{E}[\psi_\alpha^2] = \alpha(1 - \alpha)$, $B = \text{E}[\psi_\alpha^2] = g(G^{-1}(\alpha)) = s^{-1}(\alpha)$. The inverse $B$ of the sparsity function was estimated using the following relationship between the asymptotic vari-

DOI:

The Canadian Journal of Statistics / La revue canadienne de statistique
Figure 1: Parametric case, $H_0 : \beta_\alpha = (3 + G_{N(0, 1)}^{-1}(\alpha), 2)^\top$, $\alpha = 0.25$, $N = 50000$; data generated from: $N(0, 1), 0.8 \times N(0, 1) + 0.2 \times N(0, 9)$, logistic, and Laplace; dots: Wald, stars: sad. The frequencies of accepting $H_0$ under the null hypothesis are reported.
Figure 2: Nonparametric case, $H_0 : \beta_\alpha = (3 + G_{N(0,1)}^{-1}(\alpha), 2)^\top$, $\alpha = 0.25$, $N = 50000$; data generated from: $N(0,1), .8 * N(0,1) + .2 * N(0,9)$, logistic, and Laplace; dots: Wald, stars: sad. The frequencies of accepting $H_0$ under the null hypothesis are reported.

DOI: The Canadian Journal of Statistics / La revue canadienne de statistique
\[
\text{var}(\hat{\beta}_\alpha) = \frac{A}{B^2} (X^\top X)^{-1},
\]

where asymptotic covariance matrix of \( \hat{\beta}_\alpha \) was estimated using the formula from Koenker & Bassett (1978) (implemented in the function `summary.rq` in R).

(iii) The test denoted by `rank-Koenker` as implemented in R in the package `quantreg` using the command `rq(..., se = "rank")`. This command returns confidence intervals as described in Koenker (1994), which can be inverted to get the test.

(iv) The test denoted by `rank-Sen`, a rank test defined by the test statistic

\[
\sum_{i=1}^n x_{ij} \text{sign}(Y_i - x_i^\top \beta_{\alpha 0}) \left[ \varphi_\alpha \left( \frac{R_i(Y_i - x_i^\top \beta_{\alpha 0})}{n + 1} \right) - \bar{\varphi}_\alpha \right],
\]

where

\[
\varphi_\alpha(u) = \begin{cases} 0 : 0 < u < 1 - \alpha \\ 1 - \alpha : u < 1, \end{cases}
\]

\( R_i \) denotes the rank, and \( \bar{\varphi}_\alpha = \int_0^1 \varphi_\alpha(u) du = \alpha \).

(v) The test denoted by `asymp` based on the asymptotic distribution of regression quantiles

\[
\sqrt{n}(\hat{\beta}_\alpha - \beta_{\alpha 0}) = \frac{1}{\sqrt{ng(G^{-1}(\alpha))}} Q_n^{-1} \sum_{i=1}^n x_i^\top \psi_\alpha(E_{i\alpha}) + O_p(n^{-1/4}),
\]

where \( E_{i\alpha} = e_i - G^{-1}(\alpha) \) and

\[
\psi_\alpha(x) = \begin{cases} \alpha : x > 0 \\ \alpha - 1 : x \leq 0. \end{cases}
\]

Then

\[
\sqrt{n}(\hat{\beta}_\alpha - \beta_{\alpha 0}) \xrightarrow{p} N_p \left( 0, Q^{-1} \frac{\alpha(1 - \alpha)}{g^2(G^{-1}(\alpha))} \right)
\]

and the test statistic is given by

\[
n^{-\alpha/(1 - \alpha)} (\hat{\beta}_\alpha - \beta_{\alpha 0})^\top Q_n(\hat{\beta}_\alpha - \beta_{\alpha 0}) \sim \chi^2_p,
\]

where \( Q_n = \frac{1}{n} X^\top X \). The sparsity function \( s(\alpha) = [g(G^{-1}(\alpha))]^{-1} \) was estimated by the kernel estimator

\[
\hat{\beta}_{n1}(\alpha + \nu_n) - \hat{\beta}_{n1}(\alpha - \nu_n)
\]

\[
\frac{1}{2\nu_n}
\]

where the value of \( \nu_n \) was set to \( \frac{1}{2} \alpha \) (as the optimal choice of bandwidth recommended in Dodge & Jurečková (2000) would yield negative values of \( \alpha - \nu_n \) for small values of \( \alpha \)).
Hence, the difference between this test and Wald test lies in the choice of bandwidth $\nu_n$.

(vi) The test denoted by $\text{direct}$ obtained by inverting confidence intervals constructed by using directly the empirical quantile function; see Zhou & Portnoy (1996).

(vii) The test denoted by $\text{pivot-resam}$ obtained by inverting confidence intervals constructed by resampling a pivotal estimating function; see Parzen, Wei, & Ying (1994).

Under the null hypothesis, all the tests have asymptotically a $\chi^2_1$ distribution.

Tables 2, 3, 4, 5 summarize the results. Similar results are obtained for different combinations of the simulation’s parameters.

<table>
<thead>
<tr>
<th>norm, 21, 0.1</th>
<th>norm, 21, 0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9 0.95 0.99</td>
<td>0.9 0.95 0.99</td>
</tr>
<tr>
<td>sad 0.84102   0.88694 0.88870</td>
<td>sad 0.94076 0.95946 0.96288</td>
</tr>
<tr>
<td>Wald 0.10452 0.14114 0.23052</td>
<td>Wald 0.28230 0.51500 0.47054</td>
</tr>
<tr>
<td>LR 0.21942 0.43646 0.63936</td>
<td>LR 0.50968 0.61882 0.78202</td>
</tr>
<tr>
<td>rank-Sen 0.70316 0.88876 0.99278</td>
<td>rank-Sen 0.90752 0.98194 0.99986</td>
</tr>
<tr>
<td>rank-Koenker 0.99464 0.99828 1</td>
<td>rank-Koenker 0.99172 0.99776 1</td>
</tr>
<tr>
<td>asymp 0.29802 0.31960 0.35678</td>
<td>asymp 0.46778 0.49726 0.54654</td>
</tr>
<tr>
<td>direct 0.78489 0.81609 0.85392</td>
<td>direct 0.78849 0.81609 0.85392</td>
</tr>
<tr>
<td>pivot-resam 0.90026 0.96562 0.99426</td>
<td>pivot-resam 0.92590 0.97704 0.99814</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>norm, 31, 0.1</th>
<th>norm, 31, 0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9 0.95 0.99</td>
<td>0.9 0.95 0.99</td>
</tr>
<tr>
<td>sad 0.95166 0.96882 0.96166</td>
<td>sad 0.97106 0.98692 0.99288</td>
</tr>
<tr>
<td>Wald 0.19896 0.25186 0.35566</td>
<td>Wald 0.43368 0.49620 0.60226</td>
</tr>
<tr>
<td>lr 0.42900 0.53526 0.71360</td>
<td>LR 0.61584 0.70600 0.81724</td>
</tr>
<tr>
<td>rank-Sen 0.64038 0.80546 0.97888</td>
<td>rank-Sen 0.93480 0.97796 0.99918</td>
</tr>
<tr>
<td>rank-Koenker 0.98356 0.99548 0.99984</td>
<td>rank-Koenker 0.99516 0.99863 0.99994</td>
</tr>
<tr>
<td>asymp 0.49138 0.52192 0.57022</td>
<td>asymp 0.57120 0.60272 0.65146</td>
</tr>
<tr>
<td>direct 0.73465 0.75330 0.77826</td>
<td>direct 0.73465 0.75330 0.77826</td>
</tr>
<tr>
<td>pivot-resam 0.91912 0.97194 0.99628</td>
<td>pivot-resam 0.93758 0.98090 0.99834</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>norm, 51, 0.1</th>
<th>norm, 51, 0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9 0.95 0.99</td>
<td>0.9 0.95 0.99</td>
</tr>
<tr>
<td>sad 0.97528 0.98814 0.99472</td>
<td>sad 0.97372 0.99034 0.99866</td>
</tr>
<tr>
<td>Wald 0.42490 0.48230 0.58250</td>
<td>Wald 0.58316 0.63694 0.72626</td>
</tr>
<tr>
<td>lr 0.60702 0.60722 0.62658</td>
<td>LR 0.71036 0.79106 0.88792</td>
</tr>
<tr>
<td>rank-Sen 0.40648 0.60894 0.80980</td>
<td>rank-Sen 0.90140 0.96410 0.99532</td>
</tr>
<tr>
<td>rank-Koenker 0.95330 0.98482 0.99982</td>
<td>rank-Koenker 0.93392 0.97336 0.99900</td>
</tr>
<tr>
<td>asymp 0.63534 0.66484 0.70734</td>
<td>asymp 0.67036 0.70202 0.74906</td>
</tr>
<tr>
<td>direct 0.72123 0.73592 0.75360</td>
<td>direct 0.82138 0.84340 0.84858</td>
</tr>
<tr>
<td>pivot-resam 0.91490 0.97134 0.99702</td>
<td>pivot-resam 0.93578 0.98102 0.99870</td>
</tr>
</tbody>
</table>

Table 2: Nonparametric case: $H_0: \beta_\alpha = (3 + G_{N(0,1)}^{-1}(\alpha), 1, 2, 3, 4, 5)^T$, $\alpha = 0.1$ and $\alpha = 0.15$, $N = 50000$; data generated from $N(0,1)$. The frequencies of accepting $H_0$ under the null hypothesis are reported.

In general the Wald and the asymp tests are very inaccurate even under normality and should be avoided. The likelihood-ratio type test is better, but still too inaccurate, except for some sample sizes and for some $\alpha$. The test based on the direct method has a performance somewhere in the middle. The saddlepoint test, the especially the rank-koenker test and the pivot-resam test are the most reliable across distributions, different values of $\alpha$, and even down to very small sample sizes. Notice that the pivot-resam is accurate, but more computational intensive than the saddlepoint test. For instance in our simulation, its computing time is 180, 100, 40 times higher than that of the saddlepoint test for $n = 21$, $n = 31$, $n = 51$, respectively.

In addition we conducted a power study to compare the proposed parametric and nonparametric saddlepoint test to other alternatives. The setup was the following.

DOI: The Canadian Journal of Statistics / La revue canadienne de statistique
in the nonparametric case and we tested the simple null hypothesis $H_0: \beta = 0$ at significance level $\alpha = 0.05$. Under the null hypothesis, $t_3$ were generated from a standardized normal distribution $N(0,1)$, a Student $t_3$ and a contaminated normal distribution. We considered values $\alpha = 0.2$ and $\alpha = 0.4$ for sample sizes $n = 25$ and $n = 45$ for the parametric case, $n = 25$ for the nonparametric case, and the number of replications was set to 50000. We compared the performance of the parametric saddlepoint test to the Wald test calibrated at the normal distribution of the errors, whereas in the nonparametric case we compared the performance of the tests considered in the simulations under the null hypothesis.

Results are summarized in the Tables 6 to 9, where we present the proportion of rejections under the alternative at significance level 0.05. In the case of a simple hypothesis, the saddlepoint test shows good power, with the Wald test improving rapidly when moving toward the centre of the distribution. In the case of composite hypothesis, the power was smaller across all the tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald</td>
<td>0.6068</td>
<td>0.9240</td>
</tr>
<tr>
<td>LR</td>
<td>0.5780</td>
<td>0.9114</td>
</tr>
<tr>
<td>rank-Sen</td>
<td>0.5908</td>
<td>0.9111</td>
</tr>
<tr>
<td>rank-Koenker</td>
<td>0.5892</td>
<td>0.9111</td>
</tr>
<tr>
<td>asymp</td>
<td>0.6068</td>
<td>0.9111</td>
</tr>
<tr>
<td>direct</td>
<td>0.6068</td>
<td>0.9111</td>
</tr>
<tr>
<td>pivot-resam</td>
<td>0.6068</td>
<td>0.9111</td>
</tr>
</tbody>
</table>

Table 3: Nonparametric case: $H_0 : \beta = (3 + G_{N(0,1)}^{-1}(1), 1, 2, 3, 4, 5)^\top$, $\alpha = 0.25$ and $\alpha = 0.5$, N = 50000; data generated from $N(0,1)$. The frequencies of accepting $H_0$ under the null hypothesis are reported.
TABLE 4: Nonparametric case: $H_0: \beta_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 1, 2, 3, 4, 5)^T$, $\alpha = 0.1$ and $\alpha = 0.15$, $N = 50000$; data generated from a contaminated distribution $0.8 * N(0, 1) + 0.2 * N(0, 9)$. The frequencies of accepting $H_0$ under the null hypothesis are reported.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>cont, 0.1</th>
<th>cont, 0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.95</td>
<td>0.99</td>
</tr>
<tr>
<td>sad</td>
<td>0.70392</td>
<td>0.55750</td>
</tr>
<tr>
<td>Wald</td>
<td>0.03148</td>
<td>0.05012</td>
</tr>
<tr>
<td>LR</td>
<td>0.23376</td>
<td>0.33826</td>
</tr>
<tr>
<td>rank-Sen</td>
<td>0.50876</td>
<td>0.75861</td>
</tr>
<tr>
<td>rank-Koenker</td>
<td>0.99106</td>
<td>0.99757</td>
</tr>
<tr>
<td>direct</td>
<td>0.79168</td>
<td>0.81515</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>cont, 0.1</th>
<th>cont, 0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.95</td>
<td>0.99</td>
</tr>
<tr>
<td>sad</td>
<td>0.73562</td>
<td>0.81704</td>
</tr>
<tr>
<td>Wald</td>
<td>0.07012</td>
<td>0.09888</td>
</tr>
<tr>
<td>LR</td>
<td>0.29808</td>
<td>0.39914</td>
</tr>
<tr>
<td>rank-Sen</td>
<td>0.41010</td>
<td>0.54102</td>
</tr>
<tr>
<td>rank-Koenker</td>
<td>0.95018</td>
<td>0.95060</td>
</tr>
<tr>
<td>direct</td>
<td>0.73112</td>
<td>0.74922</td>
</tr>
<tr>
<td>pivot-resam</td>
<td>0.91552</td>
<td>0.97066</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>cont, 0.1</th>
<th>cont, 0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.95</td>
<td>0.99</td>
</tr>
<tr>
<td>sad</td>
<td>0.71662</td>
<td>0.74393</td>
</tr>
<tr>
<td>Wald</td>
<td>0.18766</td>
<td>0.23534</td>
</tr>
<tr>
<td>LR</td>
<td>0.43034</td>
<td>0.52372</td>
</tr>
<tr>
<td>rank-Sen</td>
<td>0.15522</td>
<td>0.28156</td>
</tr>
<tr>
<td>rank-Koenker</td>
<td>0.94996</td>
<td>0.98268</td>
</tr>
<tr>
<td>direct</td>
<td>0.73562</td>
<td>0.75911</td>
</tr>
<tr>
<td>pivot-resam</td>
<td>0.91476</td>
<td>0.97114</td>
</tr>
</tbody>
</table>

DOI:

The Canadian Journal of Statistics / La revue canadienne de statistique
Table 5: Nonparametric case: $H_0: \beta_0 = (3 + G_{N(0,1)}^{-1}(\alpha), 1, 2, 3, 4, 5)^T$, $\alpha = 0.25$ and $\alpha = 0.5$, $N = 50000$; data generated from $0.8 * N(0, 1) + 0.2 * N(0, 9)$. The frequencies of accepting $H_0$ under the null hypothesis are reported.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha$</th>
<th>norm</th>
<th>$t_3$</th>
<th>cont</th>
<th>$\alpha = 0.4$</th>
<th>norm</th>
<th>$t_3$</th>
<th>cont</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.2</td>
<td>sad</td>
<td>0.91320</td>
<td>0.82402</td>
<td>0.90748</td>
<td>sad</td>
<td>0.94418</td>
<td>0.91766</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>0.10292</td>
<td>0.26656</td>
<td>0.20402</td>
<td>Wald</td>
<td>0.75394</td>
<td>0.73732</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha$</th>
<th>norm</th>
<th>$t_3$</th>
<th>cont</th>
<th>$\alpha = 0.4$</th>
<th>norm</th>
<th>$t_3$</th>
<th>cont</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>0.2</td>
<td>sad</td>
<td>0.99164</td>
<td>0.93094</td>
<td>0.98166</td>
<td>sad</td>
<td>0.99794</td>
<td>0.99200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>0.13824</td>
<td>0.28180</td>
<td>0.21940</td>
<td>Wald</td>
<td>0.94422</td>
<td>0.91688</td>
</tr>
</tbody>
</table>

Table 6: Parametric case, simple hypothesis: $H_0: \beta_0 = (0, 0, 0)^T$, $\alpha = 0.2$ and $\alpha = 0.4$, $\beta = (1, 1, 1)^T$, $N = 50000$; data generated from $N(0,1)$, Student $t_3$-distribution and contaminated normal distribution. Powers are reported.
TABLE 7: Parametric case, composite hypothesis: $H_0 : \beta_{\alpha2} = (0, 0)^\top$, $\alpha = 0.2$ and $\alpha = 0.4$, $\beta = (1, 1, 1)^\top$, $N = 50000$; data generated from $N(0,1)$, Student $t_3$-distribution and contaminated normal distribution. Powers are reported.

<table>
<thead>
<tr>
<th>$n = 25$</th>
<th>$\alpha = 0.2$</th>
<th>$\alpha = 0.4$</th>
<th>$n = 45$</th>
<th>$\alpha = 0.2$</th>
<th>$\alpha = 0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sad</td>
<td>0.06362</td>
<td>0.47420</td>
<td>sad</td>
<td>0.07422</td>
<td>0.51504</td>
</tr>
<tr>
<td>Wald</td>
<td>0.08254</td>
<td>0.51522</td>
<td>Wald</td>
<td>0.09120</td>
<td>0.55610</td>
</tr>
</tbody>
</table>

TABLE 8: Nonparametric case, simple hypothesis: $H_0 : \beta_{\alpha2} = (0, 0)^\top$, $\alpha = 0.2$ and $\alpha = 0.4$, $\beta = (1, 1, 1)^\top$, $N = 50000$; data generated from $N(0,1)$, Student $t_3$ and a contaminated normal distribution. Powers are reported.

<table>
<thead>
<tr>
<th>$n = 25$</th>
<th>$\alpha = 0.2$</th>
<th>$\alpha = 0.4$</th>
<th>$n = 45$</th>
<th>$\alpha = 0.2$</th>
<th>$\alpha = 0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sad</td>
<td>0.09560</td>
<td>0.71156</td>
<td>sad</td>
<td>0.11360</td>
<td>0.81138</td>
</tr>
<tr>
<td>Wald</td>
<td>0.11064</td>
<td>0.72944</td>
<td>Wald</td>
<td>0.12894</td>
<td>0.82760</td>
</tr>
</tbody>
</table>

TABLE 9: Nonparametric case, composite hypothesis: $H_0 : \beta_{\alpha2} = (0, 0)^\top$, $\alpha = 0.2$ and $\alpha = 0.4$, $\beta = (1, 1, 1)^\top$, $N = 50000$; data generated from $N(0,1)$, Student $t_3$ and a contaminated normal distribution. Powers are reported.

<table>
<thead>
<tr>
<th>$n = 25$</th>
<th>$\alpha = 0.2$</th>
<th>$\alpha = 0.4$</th>
<th>$n = 45$</th>
<th>$\alpha = 0.2$</th>
<th>$\alpha = 0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sad</td>
<td>0.88102</td>
<td>0.54892</td>
<td>sad</td>
<td>0.99996</td>
<td>0.99196</td>
</tr>
<tr>
<td>Wald</td>
<td>0.87582</td>
<td>0.72790</td>
<td>Wald</td>
<td>0.99852</td>
<td>0.98074</td>
</tr>
<tr>
<td>LR</td>
<td>0.88612</td>
<td>0.66406</td>
<td>LR</td>
<td>0.99982</td>
<td>0.98960</td>
</tr>
<tr>
<td>rank-Sen</td>
<td>0.06882</td>
<td>0.04376</td>
<td>rank-Sen</td>
<td>0.08920</td>
<td>0.07386</td>
</tr>
<tr>
<td>rank-Koenker</td>
<td>0.07218</td>
<td>0.05154</td>
<td>rank-Koenker</td>
<td>0.18124</td>
<td>0.13412</td>
</tr>
<tr>
<td>asymp</td>
<td>0.67578</td>
<td>0.43348</td>
<td>asymp</td>
<td>0.94112</td>
<td>0.77542</td>
</tr>
<tr>
<td>direct</td>
<td>0.21448</td>
<td>0.52321</td>
<td>direct</td>
<td>0.13628</td>
<td>0.33350</td>
</tr>
<tr>
<td>pivot-resam</td>
<td>0.54452</td>
<td>0.46506</td>
<td>pivot-resam</td>
<td>0.68614</td>
<td>0.61708</td>
</tr>
</tbody>
</table>

DOI: The Canadian Journal of Statistics / La revue canadienne de statistique
Finally, Table 10 presents the frequencies of accepting $H_0$ under the null hypothesis in a location-scale model (non-iid case).

5. CONCLUSIONS AND OUTLOOK

We introduced a new test for quantile regression. It is derived using saddlepoint methods and it shows good accuracy in small sample sizes as well as good robustness properties. Although in theory the test can be inverted to obtain confidence intervals, this seems to be possible at present only by brute computational force. It would be interesting to find a direct way to obtain confidence intervals, but this is an open problem. Finally, the structure of the test is quite general and extensions can be worked out easily, including e.g. extreme regression quantiles; see Smith (1994), Portnoy & Jurečková (1999), Knight (2001), Chernozhukov (2005), and Jurečková (2007).

ACKNOWLEDGEMENTS

The authors would like to thank the Editor, the Associate Editor, and two Referees for very helpful comments and references and Jana Jurečková for stimulating discussions. A grant of the Swiss National Science Foundation Sciex Program 10.172 is gratefully acknowledged. The work of the second author was also partially supported by the Czech Republic Grant 201/12/0083.

BIBLIOGRAPHY


**APPENDIX**

**Assumptions**

**A1** The distribution function \( G \) is absolutely continuous, with continuous density \( g \) differentiable at the point \( G^{-1}(\alpha) \) and uniformly bounded away from 0 and \( \infty \) at the point \( G^{-1}(\alpha) \).

**A2** There exist positive definite matrices \( D_0 \) and \( D_1(\alpha) \) such that

(i) \( \lim_{n \to \infty} \frac{1}{n} X^T X = D_0 \),

(ii) \( \lim \frac{1}{\sqrt{n}} \sum_{i=1}^n g(G^{-1}(\alpha)x_i)x_i^T = D_1(\alpha) \),

(iii) \( \max_{i=1, \ldots, n} ||x_i||/\sqrt{n} \to 0 \)

DOI: *The Canadian Journal of Statistics / La revue canadienne de statistique*
We assume $\sqrt{n}(a_n(\varepsilon) - \alpha) \to \infty$ and $\sqrt{n}(b_n(\varepsilon) - \alpha) \to \infty$, where

$$a_n(\varepsilon) = \frac{1}{n} \sum_{i=1}^{n} G_{ni}(x_i^\top \beta_{\alpha 0} - \varepsilon)$$

and

$$b_n(\varepsilon) = \frac{1}{n} \sum_{i=1}^{n} G_{ni}(x_i^\top \beta_{\alpha 0} + \varepsilon)$$

and $G_{ni}$ denotes the conditional distribution function of $Y_i$, $i = 1, 2, \ldots, n$.

There exists $d > 0$ such that

$$\lim \inf_{n \to \infty} \inf_{||u||=1} \frac{1}{n} \sum_{i=1}^{n} 1[|x_i^\top u| < d] = 0.$$  

There exists $D > 0$ such that

$$\lim \sup_{n \to \infty} \sup_{||u||=1} \frac{1}{n} \sum_{i=1}^{n} (x_i^\top u)^2 \leq D.$$  

Assumptions A1 - A2 are conditions for asymptotic normality of regression quantiles; see Koenker (2005), section 4.2. The assumption A1 is a strengthened version of the condition A1 in Koenker (2005) in that we require the differentiability of the density function, which is necessary for the Taylor expansion of the test statistic. Assumptions A3 - A5 are conditions for the consistency of a regression quantile; see Koenker (2005), section 4.1.2.

Saddlepoint Test for M-estimators

For completeness we summarize here the definition of the saddlepoint test statistic for $M$-estimators and its properties as developed in Robinson, Ronchetti, & Young (2003).

Let $Y_1, \ldots, Y_n$ be an independent sample from a distribution $F$. Consider an $M$-estimator $\hat{\theta}$ of a parameter $\theta = \theta(F)$, defined as a solution of the equation

$$\sum_{i=1}^{n} \psi(Y_i; \theta) = 0.$$  

Consider the composite hypothesis

$$H_0 : \theta_1 = \theta_{10} \in \mathbb{R}^{p_1}, \theta_2 \in \mathbb{R}^{p_2}$$

where $\theta = (\theta_1^\top, \theta_2^\top)$, $\hat{\theta} = (\hat{\theta}_1^\top, \hat{\theta}_2^\top)$.

**Parametric case**

Define the one-dimensional statistic

$$b(\hat{\theta}_1) = \inf_{\theta_2} \sup_{\lambda} \{-K_\psi(\lambda; (\hat{\theta}_1, \theta_2))\},$$

where

$$K_\psi(\lambda; \theta) = \log E_{F_0}[e^{\lambda^\top \psi(Y_i; \theta)}]$$

The Canadian Journal of Statistics / La revue canadienne de statistique

DOI:
is the cumulant generating function of the score \( \psi(Y_i; \theta) \) and the expectation is computed with respect to the distribution \( F_0 \) of the observations under the null hypothesis. Notice that the \( \sup \) part in (10) can be rewritten as

\[
\sup_{\lambda} \{-K_\psi(\lambda; t)\} = -K_\psi(\lambda(t); t),
\]

where \( \lambda(t) \) is the saddlepoint satisfying

\[
\frac{\partial}{\partial \lambda} K_\psi(\lambda; t) = 0,
\]

i.e.

\[
E_{F_0}[\psi(Y_i; t)e^{\lambda^T \psi(Y_i; t)}] = 0.
\]

Then under the null hypothesis,

\[
2nh(\hat{\theta}_1) \xrightarrow{D} \chi^2_{p_1}
\]

with relative error of order \( O(n^{-1}) \). This test is first-order equivalent to the three classical tests, but exhibits better second-order properties, i.e. has better small sample properties.

In the case of a simple hypothesis, the statistic simplifies to

\[
h(\hat{\theta}) = -K_\psi(\lambda(\hat{\theta}); \hat{\theta}).
\]

Moreover, if the observations \( Y_1, \ldots, Y_n \) are independent but not identically distributed (as in the regression case), the cumulant generating function becomes

\[
K_\psi(\lambda; \theta) = \frac{1}{n} \sum_{i=1}^{n} K^i_\psi(\lambda; \theta),
\]

where \( K^i_\psi(\lambda; \theta) = \log E[F_i(e^{\lambda^T \psi(Y_i; \theta)})] \) and \( F_i \) is the distribution function of \( Y_i \); see Lô & Ronchetti (2009).

**Nonparametric case**

When \( F \) is unspecified, an empirical version of the test may be used. Let \( \hat{F}_0 = (w_1, \ldots, w_n) \) be the empirical distribution which satisfies the null hypothesis and is closest to \( (1/n, \ldots, 1/n) \) in the sense of the backward Kullback-Leibler divergence.

Then, the saddlepoint test statistic is given by

\[
\hat{h}(\hat{\theta}_1) = \inf_{\theta_2} \sup_{\lambda} \left\{ -K^w_0(\lambda; (\hat{\theta}_1, \theta_2)) \right\},
\]

where

\[
K^w_0(\lambda; \theta) = \log \left( \sum_{i=1}^{n} w_i e^{\lambda^T \psi(Y_i; \theta)} \right),
\]

DOI: *The Canadian Journal of Statistics / La revue canadienne de statistique*
and the weights $w_i$ are computed as

$$w_i = e^{\mu(\theta^*)^\top \psi(y_i; \theta^*)}/\sum_{j=1}^n e^{\mu(\theta^*)^\top \psi(y_j; \theta^*)},$$

where

$$\theta^* = (\theta_{10}, \theta_2^*)$$

$$\theta_2^* = \arg \min \left\{ -\kappa(\mu(\theta_{10}, \theta_2); (\theta_{10}, \theta_2)) \right\}$$

$$\mu(\theta) = \arg \max \left\{ -\kappa(\mu; \theta) \right\},$$

$$\kappa(\mu; \theta) = \log \left( \frac{1}{n} \sum_{i=1}^n e^{\mu(\theta)^\top \psi(y_i; \theta)} \right).$$

When $n \to \infty$, the p-value satisfies

$$P_{H_0} \{|2n\hat{h}(\hat{\theta}_1)| \geq 2n\hat{h}(\hat{\theta}_{1obs})\} = \{1 - Q_{p_1}(2n\hat{h}(\hat{\theta}_{1obs}))\{1 + O_P(n^{-1})\}\},$$

where $Q_{p_1}$ denotes the cumulative distribution function of the $\chi^2$ distribution with $p_1$ degrees of freedom.

**Proof of Proposition 2.1**

**Proof.**

By Taylor expansion we obtain,

$$G \left( x_i^\top (\hat{\beta}_a - \beta_0) \right) = G \left( x_i^\top (\hat{\beta}_a - \beta_{a0} + (G^{-1}(\alpha), 0, \ldots, 0)^\top) \right)$$

$$= G(x_i^\top (G^{-1}(\alpha), 0, \ldots, 0)^\top)$$

$$+ x_i^\top (\hat{\beta}_a - \beta_{a0}) g(x_i^\top (G^{-1}(\alpha), 0, \ldots, 0)^\top)$$

$$+ \frac{1}{2} (x_i^\top (\hat{\beta}_a - \beta_{a0}))^2 g'(x_i^\top (G^{-1}(\alpha), 0, \ldots, 0)^\top)$$

$$+ O_P((x_i^\top (\hat{\beta}_a - \beta_{a0}))^3)$$

$$= \alpha + x_i^\top (\hat{\beta}_a - \beta_{a0}) g(G^{-1}(\alpha)) +$$

$$\frac{1}{2} (x_i^\top (\hat{\beta}_a - \beta_{a0}))^2 g'(x_i^\top (G^{-1}(\alpha), 0, \ldots, 0)^\top)$$

$$+ O_P((x_i^\top (\hat{\beta}_a - \beta_{a0}))^3)$$
and by further Taylor expansion of \((1 + x)^\alpha\) we get

\[
\left( \frac{G \left( x^\top (\hat{\beta}_a - \beta_0) \right)}{\alpha} \right)^\alpha = \left( 1 + \frac{x^\top (\hat{\beta}_a - \beta_0) g(G^{-1}(\alpha))}{\alpha} \right)
\]

\[
= 1 + \frac{1}{2\alpha} (x^\top (\hat{\beta}_a - \beta_0))^2 g'(G^{-1}(\alpha))
\]

\[
+ O_P((x^\top (\hat{\beta}_a - \beta_0))^3)\alpha
\]

\[
= 1 + \frac{1}{2\alpha} (x^\top (\hat{\beta}_a - \beta_0))^2 g'(G^{-1}(\alpha))
\]

\[
+ \frac{\alpha - 1}{2\alpha} (x^\top (\hat{\beta}_a - \beta_0)) g(G^{-1}(\alpha))^2
\]

\[
+ O_P((x^\top (\hat{\beta}_a - \beta_0))^3).
\]

Similarly,

\[
\left( \frac{1 - G \left( x^\top (\hat{\beta}_a - \beta_0) \right)}{1 - \alpha} \right)^{1-\alpha} = \left( 1 - \frac{x^\top (\hat{\beta}_a - \beta_0) g(G^{-1}(\alpha))}{1 - \alpha} \right)
\]

\[
= 1 - \frac{1}{2(1 - \alpha)} (x^\top (\hat{\beta}_a - \beta_0))^2 g'(G^{-1}(\alpha))
\]

\[
+ O_P((x^\top (\hat{\beta}_a - \beta_0))^3)\alpha
\]

\[
= 1 - \frac{1}{2\alpha} (x^\top (\hat{\beta}_a - \beta_0))^2 g'(G^{-1}(\alpha))
\]

\[
- \frac{\alpha - 1}{2\alpha} (x^\top (\hat{\beta}_a - \beta_0)) g(G^{-1}(\alpha))^2
\]

\[
+ O_P((x^\top (\hat{\beta}_a - \beta_0))^3),
\]

and finally

\[
\left( \frac{G \left( x^\top (\hat{\beta}_a - \beta_0) \right)}{\alpha} \right)^\alpha \left( 1 - \frac{G \left( x^\top (\hat{\beta}_a - \beta_0) \right)}{1 - \alpha} \right)^{1-\alpha}
\]

\[
= 1 - \frac{1}{2\alpha(1 - \alpha)} (x^\top (\hat{\beta}_a - \beta_0))^2 g(G^{-1}(\alpha))^2
+ O_P((x^\top (\hat{\beta}_a - \beta_0))^3).\]
Therefore,

\[-2n \log \left( \frac{1}{n} \sum_{i=1}^{n} \left( \frac{G(x_i^T(\hat{\beta}_\alpha - \beta_0))}{\alpha} \right)^{\alpha} \left( \frac{1 - G(x_i^T(\hat{\beta}_\alpha - \beta_0))}{1 - \alpha} \right)^{1-\alpha} \right) = -2n \log \left( \frac{1}{n} \sum_{i=1}^{n} \left( 1 - \frac{1}{2\alpha(1 - \alpha)}(x_i^T(\hat{\beta}_\alpha - \beta_0)g(G^{-1}(\alpha)))^2 \right) + O_P((x_i^T(\hat{\beta}_\alpha - \beta_0))^3) \right) = -2n \log \left( 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2\alpha(1 - \alpha)}(x_i^T(\hat{\beta}_\alpha - \beta_0)g(G^{-1}(\alpha)))^2 \right) + O_P((x_i^T(\hat{\beta}_\alpha - \beta_0))^3) = \sum_{i=1}^{n} \frac{1}{\alpha(1 - \alpha)}(x_i^T(\hat{\beta}_\alpha - \beta_0))^2 + O_P((x_i^T(\hat{\beta}_\alpha - \beta_0))^3) = n\sum_{i=1}^{n} \frac{1}{\alpha(1 - \alpha)}(x_i^T(\hat{\beta}_\alpha - \beta_0))^2 + O_P((x_i^T(\hat{\beta}_\alpha - \beta_0))^3),\]

The proof can be completed by using the consistency and asymptotic normality of the regression quantile estimator

\[\sqrt{n}(\hat{\beta}_\alpha - \beta_{00}) \xrightarrow{D} \lim_{n \to \infty} N_p \left( 0, \frac{\alpha(1 - \alpha)}{g(G^{-1}(\alpha))^2} D \right),\]

which hold under the Assumptions of the Proposition. Thus, the test statistic 2nh(\hat{\beta}_\alpha) is asymptotically \(\chi^2_p\).

Proof of Proposition 2.2

Proof.
Let us rewrite the test statistic (7) and let us first consider the denominator of $w_i$:

\[
\sum_{k=1}^{n} \left( \frac{\alpha}{1 - \alpha} \frac{1 - F_n(\beta_{\alpha0})}{F_n(\beta_{\alpha0})} \right) \hat{I}_k(\beta_{\alpha0}) = \sum_{k=1}^{n} (1 - \hat{I}_k(\beta_{\alpha0})) + \left( \frac{\alpha}{1 - \alpha} \frac{1 - F_n(\beta_{\alpha0})}{F_n(\beta_{\alpha0})} \right) \sum_{k=1}^{n} \hat{I}_k(\beta_{\alpha0})
\]

\[
= n + \left\{ \left( \frac{\alpha}{1 - \alpha} \frac{1 - F_n(\beta_{\alpha0})}{F_n(\beta_{\alpha0})} \right) - 1 \right\} \sum_{k=1}^{n} \hat{I}_k(\beta_{\alpha0})
\]

\[
= n + \left\{ \frac{\alpha - F_n(\beta_{\alpha0})}{(1 - \alpha)F_n(\beta_{\alpha0})} \right\} nF_n(\beta_{\alpha0})
\]

\[
= n \left( \frac{1 - F_n(\beta_{\alpha0})}{1 - \alpha} - 1 \right).
\]

Thus,

\[
w_i I_i = \left( \frac{\alpha}{nF_n(\beta_{\alpha0})} \right) \hat{I}_i(\beta_{\alpha0}) \left( \frac{1 - \alpha}{n(1 - F_n(\beta_{\alpha0}))} \right)^{1 - \hat{I}_i(\beta_{\alpha0})} I_i
\]

\[
= \frac{1}{n} \left( \frac{\alpha}{F_n(\beta_{\alpha0})} \right) \hat{I}_i(\beta_{\alpha0}) \left( \frac{1 - \alpha}{1 - F_n(\beta_{\alpha0})} \right)^{1 - \hat{I}_i(\beta_{\alpha0})} I_i
\]

and

\[
\sum_{i=1}^{n} w_i I_i = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\alpha}{F_n(\beta_{\alpha0})} \right) \hat{I}_i(\beta_{\alpha0}) \left( \frac{1 - \alpha}{1 - F_n(\beta_{\alpha0})} \right)^{1 - \hat{I}_i(\beta_{\alpha0})} I_i.
\]

Similarly for $\sum_{i=1}^{n} w_i (1 - I_i)$. Now

\[
E\{\hat{I}_i(\beta_{\alpha0})\} = E\{I[y_i < \beta_{\alpha0}]\} = E\{I[y_i < \beta_0 + G^{-1}(\alpha)]\}
\]

\[
= P[u_i < G^{-1}(\alpha)] = G(G^{-1}(\alpha)) = \alpha
\]

and

\[
E\{I_i\} = E\{I[y_i < \hat{\beta}_\alpha]\}
\]

\[
= E\{I[y_i - \hat{\beta}_\alpha < \hat{\beta}_\alpha - \beta_0]\}
\]

\[
= P[u_i < \hat{\beta}_\alpha - \beta_0] = G(\hat{\beta}_\alpha - \beta_0).
\]

By the law of large numbers $F_n(\beta_{\alpha0}) \to E\{\hat{I}_i(\beta_{\alpha0})\} = \alpha$ as $n \to \infty$ and the test statistic (7) is asymptotically equivalent to the test statistic (5).

Derivation of the nonparametric saddlepoint test statistic

To compute the saddlepoint test statistic, we follow the result given for $M$–estimators in the Appendix. Here the $n$ distributions $F^i$ are estimated through the empirical distribution of the residuals as in Ronchetti & Welsh (1994).
Let us define

\[ r_i = y_i - x_i^\top \hat{\beta}_\alpha \]

\[ I_{ij}(\beta) = I[r_i + x_j^\top (\hat{\beta}_\alpha - \beta) < 0] = I[y_i - x_j^\top \beta < 0] \]

\[ F_n^j(\beta) = \frac{1}{n} \sum_{i=1}^{n} I_{ij}(\beta), \quad i, j = 1, \ldots, n. \]

As in the parametric case,

\[ x_j^\top \mu = \log \frac{1 - \alpha}{\alpha} \frac{F_n^j(\beta_{\alpha 0})}{1 - F_n^j(\beta_{\alpha 0})} \]

solves

\[ \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha - I_{ij}(\beta_{\alpha 0})) \left( \frac{1 - \alpha}{\alpha} \frac{F_n^j(\beta_{\alpha 0})}{1 - F_n^j(\beta_{\alpha 0})} \right)^{\alpha - I_{ij}(\beta_{\alpha 0})} x_j^\top \mu = 0, \]

because the following equalities are equivalent.

\[ \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha - I_{ij}(\beta_{\alpha 0})) \left( \frac{1 - \alpha}{\alpha} \frac{F_n^j(\beta_{\alpha 0})}{1 - F_n^j(\beta_{\alpha 0})} \right)^{\alpha - I_{ij}(\beta_{\alpha 0})} x_j^\top \mu = 0 \]

\[ \iff \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha \left( \frac{1 - \alpha}{\alpha} \frac{F_n^j(\beta_{\alpha 0})}{1 - F_n^j(\beta_{\alpha 0})} \right)^{\alpha - I_{ij}(\beta_{\alpha 0})} x_j \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{n} I_{ij}(\beta_{\alpha 0}) \left( \frac{1 - \alpha}{\alpha} \frac{F_n^j(\beta_{\alpha 0})}{1 - F_n^j(\beta_{\alpha 0})} \right)^{\alpha - I_{ij}(\beta_{\alpha 0})} x_j \]

\[ \iff \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha I_{ij}(\beta_{\alpha 0}) \left( \frac{1 - \alpha}{\alpha} \frac{1 - F_n^j(\beta_{\alpha 0})}{1 - F_n^j(\beta_{\alpha 0})} \right)^{1-\alpha} x_j \]

\[ + \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha (1 - I_{ij}(\beta_{\alpha 0})) \left( \frac{1 - \alpha}{\alpha} \frac{F_n^j(\beta_{\alpha 0})}{1 - F_n^j(\beta_{\alpha 0})} \right)^{\alpha} x_j \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{n} I_{ij}(\beta_{\alpha 0}) \left( \frac{1 - \alpha}{\alpha} \frac{1 - F_n^j(\beta_{\alpha 0})}{1 - F_n^j(\beta_{\alpha 0})} \right)^{1-\alpha} x_j \]

\[ \iff \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha (1 - I_{ij}(\beta_{\alpha 0})) \left( \frac{1 - \alpha}{\alpha} \frac{F_n^j(\beta_{\alpha 0})}{1 - F_n^j(\beta_{\alpha 0})} \right)^{\alpha} x_j \]
Therefore, the test statistic is given by

\[ 2n \hat{h}(\hat{\beta}_\alpha) = -2n K_\psi(\lambda; \hat{\beta}_\alpha; \hat{\beta}_\alpha), \]

where \( \lambda(\hat{\beta}_\alpha) \) satisfies \( \frac{\partial K_\psi'(\lambda; \beta)}{\partial \lambda} = 0 \), i.e.

\[ \sum_{i=1}^{n} (\alpha - 1 [r_i < 0]) x_j e^{(\alpha - 1 [r_i < 0]) \lambda + (\alpha - I_{ij}(\beta_{\alpha})) x_j^\top \mu} = 0. \]
Now let
\[ I_i = I[r_i < 0], \quad i = 1, \ldots, n. \]

Then \( x_j^T \lambda \) is a solution of the equation
\[
\sum_{i=1}^{n} (\alpha - I_i) e^{\alpha - I_i} x_j^T \lambda \left( \frac{\alpha}{1 - \alpha} \frac{1 - F_j^i(\beta_{ao})}{F_j^i(\beta_{ao})} \right)^{I_{ij}(\beta_{ao})} = 0
\]
i.e.
\[
(1 - \alpha) \sum_{i=1}^{n} I_i e^{\alpha - I_i} x_j^T \lambda \left( \frac{\alpha}{1 - \alpha} \frac{1 - F_j^i(\beta_{ao})}{F_j^i(\beta_{ao})} \right)^{I_{ij}(\beta_{ao})}
\]
\[ = \alpha \sum_{i=1}^{n} (1 - I_i) e^{\alpha} x_j^T \lambda \left( \frac{\alpha}{1 - \alpha} \frac{1 - F_j^i(\beta_{ao})}{F_j^i(\beta_{ao})} \right)^{I_{ij}(\beta_{ao})}
\]
and this is equivalent to
\[
x_j^T \lambda(\hat{\beta}_a) = \log \left\{ \frac{1 - \alpha}{\alpha} \frac{\sum_{i=1}^{n} w_{ij} I_i}{\sum_{i=1}^{n} w_{ij} (1 - I_i)} \right\},
\]
where
\[
w_{ij} = \frac{\left( \frac{\alpha}{1 - \alpha} \frac{1 - F_j^i(\beta_{ao})}{F_j^i(\beta_{ao})} \right)^{I_{ij}(\beta_{ao})}}{\sum_{k=1}^{n} \left( \frac{\alpha}{1 - \alpha} \frac{1 - F_j^k(\beta_{ao})}{F_j^k(\beta_{ao})} \right)^{I_{kj}(\beta_{ao})}}.
\]

Thus, we obtain
\[
K^a_V(\lambda(\hat{\beta}_a); \hat{\beta}_a) = \log \sum_{l=1}^{n} \left( \frac{1 - \alpha}{\alpha} \frac{\sum_{i=1}^{n} w_{ij} I_i}{\sum_{i=1}^{n} w_{ij} (1 - I_i)} \right)^{\alpha - 1} w_{lj}
\]
\[ = \log \left\{ \frac{1 - \alpha}{\alpha} \frac{\sum_{i=1}^{n} w_{ij} I_i}{\sum_{i=1}^{n} w_{ij} (1 - I_i)} \right\}^\alpha
\]
\[ \times \sum_{l=1}^{n} \left( \frac{1 - \alpha}{\alpha} \frac{\sum_{i=1}^{n} w_{ij} (1 - I_i)}{\sum_{i=1}^{n} w_{ij} I_i} w_{lj} I_i + w_{lj} (1 - I_i) \right)^\alpha
\]
\[ = \log \left\{ \frac{1 - \alpha}{\alpha} \frac{\sum_{i=1}^{n} w_{ij} I_i}{\sum_{i=1}^{n} w_{ij} (1 - I_i)} \right\}^\alpha \sum_{i=1}^{n} w_{ij} (1 - I_i)
\]
\[ = \log \left\{ \frac{\sum_{i=1}^{n} w_{ij} I_i}{\alpha} \right\}^\alpha \left( \frac{\sum_{i=1}^{n} w_{ij} (1 - I_i)}{1 - \alpha} \right)^{1 - \alpha}
\]
and

\[
\hat{h}(\hat{\beta}_\alpha) = - K_\psi(\lambda(\hat{\beta}_\alpha); \hat{\beta}_\alpha) = - \frac{1}{n} \sum_{j=1}^{n} K_{\psi j}(\lambda(\hat{\beta}_\alpha); \hat{\beta}_\alpha) \\
= - \frac{1}{n} \sum_{j=1}^{n} \log \left\{ \left( \frac{\sum_{i=1}^{n} w_{ij} I_i}{\alpha} \right)^\alpha \left( \frac{\sum_{i=1}^{n} w_{ij} (1 - I_i)}{1 - \alpha} \right)^{1-\alpha} \right\}
\]

(13) and this concludes the computation of the saddlepoint test statistic.

The nonparametric test statistic for the special case of simple quantiles can be easily obtained from (14) by putting \( p = 1, x_i \equiv 1, r_i = y_i - \hat{\beta}_\alpha, \hat{\beta}_\alpha = F_n^{-1}(\alpha) \), the empirical quantile, \( I_i = 1[r_i < 0] = 1[y_i < \hat{\beta}_\alpha], I_{ij}(\beta) = 1[y_i < \beta] = \tilde{I}_i(\beta) \), and by observing that in this case the weights \( w_{ij} \equiv w_i \) are independent of \( j \).

Received 2014

Accepted

DOI: