Rank tests and regression rank scores tests in measurement error models

J. Jurečková and A.K.Md.E. Saleh

Charles University in Prague and *Carleton University in Ottawa*

Abstract

The rank and regression rank score tests of linear hypothesis in the linear regression model are modified for measurement error models. In some situations the modified tests are still distribution free; their asymptotic relative efficiencies with respect to tests in model without errors are evaluated.

1 Introduction

Consider the linear regression model

$$Y_i = \beta_0 + \mathbf{x}'_{ni}\boldsymbol{\beta} + e_i, \quad i = 1, \ldots, n \tag{1.1}$$

with observations $Y_1, \ldots, Y_n$, independent errors $e_1, \ldots, e_n$, identically distributed according to an unknown distribution function F; $\mathbf{x}_{ni} = (x_{i1}, \ldots, x_{in})'$ is the vector of covariates, $i = 1, \ldots, n$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, and $\boldsymbol{\beta}^* = (\beta_0, \boldsymbol{\beta}')'$ are unknown parameters. We shall suppress the subscript n whenever it does not cause a confusion.

If the covariates are only measured with random errors, then instead of $\mathbf{x}_i$ we only observe $\mathbf{w}_i = \mathbf{x}_i + \mathbf{u}_i,\ i = 1, \ldots, n$. Several authors studied the behavior of the (eventually modified) regression quantiles or of the conditional median in error-in-variables (EV) model; let us mention He and Liang (2000), where are other references cited.

Our primary interest is to consider how the eventual measurement errors in the covariates affect the behavior and the efficiency of the rank tests of the linear hypothesis in the linear regression model. In a more general context, when the linear regression is considered as a nuisance in the model, but is not affected by measurement errors, and we have another hypothesis of interest in mind, we can use the tests based on *regression rank scores* proposed in Gutenbrunner

AMS 2000 subject classifications. 62G10, 62J05

Key words and phrases: rank test of linear hypothesis, regression rank score test of linear subhypothesis, measurement error.

Research of J. Jurečková was supported by Research Projects LC024, MSM 0021620839 and by Czech Republic Grant 201/05/2340.

Research of A.K.Md.E.Saleh was supported by NSERC Grant of Canada.

et al. (1993). Their criteria are based on the regression rank scores under the hypothesis, i.e. under the model (1.1), and are asymptotically equivalent to the ordinary rank tests pertaining to the situation where $\boldsymbol{\beta}$ is known.

We shall consider two models:

$$Y_i = \beta_0 + \mathbf{x}'_{ni}\boldsymbol{\beta} + e_i, \quad i = 1, \ldots, n \tag{1.2}$$

where we want to test the hypothesis

$$\mathbf{H}: \ \boldsymbol{\beta} = \mathbf{0}, \tag{1.3}$$

and

$$Y_i = \beta_0 + \mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{z}'_{ni}\boldsymbol{\delta} + e_i, \quad i = 1, \ldots, n \tag{1.4}$$

with unknown parameters $\boldsymbol{\beta} \in \mathbb{R}^p, \ \ \boldsymbol{\delta} \in \mathbb{R}^q$, and the hypothesis

$$\mathbf{H}: \ \boldsymbol{\delta} = \mathbf{0}, \tag{1.5}$$

considering $\boldsymbol{\beta}$ as a nuisance parameter.

In the models without measurement errors, the hypothesis (1.3) can be tested by a rank test, while for the hypothesis (1.5) we can use the regression rank score test. The model (1.4) e.g. appears in the polynomial regression model, when we want to test the hypothesis on the order of the polynomial and we do not know the coefficients of lower order terms. The regression rank scores enable us to avoid estimating the nuisance parameters.

We shall study how the measurement errors affect the performance of tests of the above hypotheses. In Section 2 we shall study the loss of efficiency of the ordinary rank tests of hypothesis (1.3) if the regressors are affected by random measurement errors. Section 3 considers the tests of hypothesis (1.5) in the presence of nuisance regression, based on the regression rank scores, in the situation that the regressors $\mathbf{x}_{ni}$ are precisely observable, while the $\mathbf{z}_{ni}$ are affected by random errors. In turns out that even in such situation we can construct distribution free tests, that only loose efficiency with respect to tests in models without errors. The question whether we can construct a distribution free test in model (1.4) when the $\mathbf{x}_{ni}$ or both the $\mathbf{x}_{ni}$ and $\mathbf{z}_{ni}$ are affected by random errors is still an open problem.

2 Linear rank test of randomness against regression alternative

To see the effect of the measurement errors in the covariates, let us first consider how do they affect the simple linear rank test of the hypothesis of randomness against a regression alternative.

Let $Y_1, \ldots, Y_n$ be independent observations, Y_i distributed according to a d.f. $F(y - x_{ni}\beta)$, $i = 1, \ldots, n$. We assume that F has an absolutely continuous density f, otherwise it is generally unknown; $(x_{n1}, \ldots, x_{nn})$ is a vector of known regressors, satisfying

$$\lim_{n\to\infty} \left[\max_{1\le i\le n} (x_{ni} - \bar{x}_n)^2 \Big(\sum_{j=1}^{n} (x_{nj} - \bar{x}_n)^2 \Big)^{-1} \right] = 0, \tag{2.1}$$

$\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_{ni}$; moreover, for the sake of simplicity and concentration on the main idea, we assume that

$$\lim_{n\to\infty} C_n^2 = C^2 > 0, \quad \text{where} \quad C_n^2 = \frac{1}{n}\sum_{j=1}^n (x_{nj} - \bar{x}_n)^2. \tag{2.2}$$

The problem is that of testing the hypothesis $\mathbf{H}_0 : \ \beta = 0$ in the situation when the x_{ni} are not exactly observable; instead of x_{ni} we can only observe $w_{ni} = x_{ni} + u_{ni}, \ i = 1, \ldots, n$, where $u_{n1}, \ldots, u_{nn}$ are i.i.d. random errors. Their distribution (say G) is unknown, but we assume that it has the finite second moment and

$$\begin{aligned} &D_n^2 \xrightarrow{p} D^2 > 0, \\ &D_n^2 = \frac{1}{n}\sum_{i=1}^n (u_{ni} - \bar{u}_n)^2, \quad \bar{u}_n = \frac{1}{n}\sum_{i=1}^n u_{ni}. \end{aligned} \tag{2.3}$$

Moreover, we assume that Y_i and u_i are independent, $i = 1, \ldots, n$.

Let $R_{n1}, \ldots, R_{nn}$ be the ranks of $Y_1, \ldots, Y_n$. In the models without measurement errors, the rank test of $\mathbf{H}_0$ is based on the criterion

$$S_n = n^{-1/2}\sum_{i=1}^n (x_{ni} - \bar{x}_n) a_n(R_{ni}) \tag{2.4}$$

with the scores $a_n(1), \ldots, a_n(n)$ generated by a nondecreasing, square integrable score function $\varphi : (0,1) \mapsto \mathbb{R}^1$ in either of the following two ways:

$$\begin{aligned} &a_n(i) = \mathbb{E}\varphi(U_{n:i}), \\ &a_n(i) = \varphi\left(\frac{i}{n+1}\right), \ i = 1, \ldots, n, \end{aligned} \tag{2.5}$$

where $U_{n:1} \leq \ldots \leq U_{n:n}$ are the order statistics corresponding to the sample of size n from the $R(0,1)$ distribution.

The asymptotic distributions of the linear rank statistics are studied in Hájek and Šidák (1967). Under the above conditions on the x_{ni} and on F, the asymptotic distribution of S_n under $\mathbf{H}_0$ is normal $\mathcal{N}\left(0, \ C^2 \ A^2(\varphi)\right)$, $A^2(\varphi) = \int_0^1 (\varphi(t) - \bar{\varphi})^2 dt, \ \bar{\varphi} = \int_0^1 \varphi(t)dt$. Under the local alternative

$$\mathbf{H}_n : \ \beta = \beta_n = n^{-1/2}\beta^*, \ \beta^* \in \mathbb{R}^1, \tag{2.6}$$

S_n is asymptotically normally distributed

$$\mathcal{N}\left(C^2 \ \gamma(\varphi, f) \ \beta^*, \ C^2 \ A^2(\varphi)\right), \tag{2.7}$$

$$\gamma(\varphi, f) = \int_0^1 \varphi(t)\varphi(t,f)dt, \quad \varphi(t,f) = -\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}, \ 0 < t < 1. \tag{2.8}$$

We reject $\mathbf{H}_0$ in favor of alternative $\mathbf{K} : \ \beta \neq 0$ on the asymptotic significance level α provided

$$A^{-1}(\varphi)\Big(\frac{1}{n}\sum_{i=1}^n (x_{ni} - \bar{x})^2\Big)^{-1/2} \ |S_n| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \tag{2.9}$$

Then it follows from (2.7) that the asymptotic power of the test against $\mathbf{H}_n$ is

$$1-\Phi\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)-C\ \frac{\gamma(\varphi,f)\ \beta^*}{A(\varphi)}\right)+\Phi\left(\Phi^{-1}\left(\frac{\alpha}{2}\right)-C\ \frac{\gamma(\varphi,f)\ \beta^*}{A(\varphi)}\right). \tag{2.10}$$

However, we can only measure $w_{ni}=x_{ni}+u_{ni}$ instead of $x_{ni},\ i=1,\dots,n$ and hence our test would be based on

$$\widetilde{S}_n=n^{-1/2}\sum_{i=1}^{n}(w_{ni}-\bar{w}_n)a_n(R_{ni})$$

instead of S_n. The asymptotic null distribution of the criterion $A^{-1}(\varphi)\ \left(\frac{1}{n}\sum_{i=1}^{n}(w_{ni}-\bar{w})^2\right)^{-1/2}\ \widetilde{S}_n$ is normal $\mathcal{N}(0,1)$, while that under $\mathbf{H}_n$ is

$$\mathcal{N}\left(\frac{C^2}{(C^2+D^2)^{1/2}}\ \frac{\gamma(\varphi,f)\ \beta^*}{A(\varphi)},\ 1\right).$$

Hence, in the measurement error model we reject $\mathbf{H}_0$ provided

$$A^{-1}(\varphi)\ \left(\frac{1}{n}\sum_{i=1}^{n}(w_{ni}-\bar{w})^2\right)^{-1/2}\ |\widetilde{S}_n|\geq\Phi\left(1-\frac{\alpha}{2}\right) \tag{2.11}$$

and the asymptotic power of this test against $\mathbf{H}_n:\ \beta=n^{-1/2}\beta^*\neq 0$ is then

$$\begin{aligned}1-\Phi&\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)-\frac{C^2}{(C^2+D^2)^{1/2}}\ \frac{\gamma(\varphi,f)\ \beta^*}{A(\varphi)}\right)\\ +\Phi&\left(\Phi^{-1}\left(\frac{\alpha}{2}\right)+\frac{C^2}{(C^2+D^2)^{1/2}}\ \frac{\gamma(\varphi,f)\ \beta^*}{A(\varphi)}\right).\end{aligned} \tag{2.12}$$

Hence, there is a positive loss of the asymptotic power of the test (2.11) with respect to the test (2.9); it is bounded from above by

$$\frac{2C}{\sqrt{2\pi}}\ \frac{\gamma(\varphi,f)\ |\beta^*|}{A(\varphi)}.$$

The asymptotic relative efficiency of the test (2.11) with respect to the test (2.9) is

$$e_{w,x}=\left(1+\frac{D^2}{C^2}\right)^{-1}\quad\left[\downarrow 0\ \text{ as }\ D\uparrow\infty\right]. \tag{2.13}$$

More generally, consider the multiple regression model

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{e} \tag{2.14}$$

with the vector of independent observations $\mathbf{Y}=(Y_1,\dots,Y_n)'$, vector $\mathbf{e}=(e_1,\dots,e_n)'$ of i.i.d. errors, the regression matrix $\mathbf{X}=\mathbf{X}_n$ of order $n\times p$ with the rows $\mathbf{x}_{ni},\ i=1,\dots,n$ and an unknown parameter $\boldsymbol{\beta} \in \mathbb{R}^p$. We want to test the hypothesis $\mathbf{H}_0 : \ \boldsymbol{\beta} = \mathbf{0}$ against the alternative $\mathbf{K} : \ \boldsymbol{\beta} \neq \mathbf{0}$. Let $\bar{\mathbf{X}}_n$ denote the matrix with the rows $\mathbf{x}_{ni} - \bar{\mathbf{x}}_n, \ i = 1, \ldots, n$, and let it satisfy

$$\mathbf{Q}_n = \frac{1}{n}\bar{\mathbf{X}}_n'\bar{\mathbf{X}}_n \to \mathbf{Q} \quad \text{as} \quad n \to \infty, \tag{2.15}$$

$$n^{-1} \max_{1 \le i \le n} \left\{(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)'\mathbf{Q}_n^{-1}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\right\} \to 0 \quad \text{as} \quad n \to \infty$$

where $\mathbf{Q}$ is a positively definite $p \times p$ matrix.

In the model without measurement errors, the rank test of $\mathbf{H}_0 : \ \boldsymbol{\beta} = \mathbf{0}$ is based on the vector of linear rank statistics $\mathbf{S}_n \in \mathbb{R}^p$,

$$\mathbf{S}_n = n^{-1/2} \sum_{i=1}^{n} (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) a_n(R_{ni}) \tag{2.16}$$

where $R_{n1}, \ldots, R_{nn}$ are the ranks of $Y_1, \ldots, Y_n$ and $a_n(i)$ are again the scores generated by φ. The asymptotic distribution of $\mathbf{S}_n$ under $\mathbf{H}_0$ has normal $\mathcal{N}_p(\mathbf{0}, \ A^2(\varphi) \ \mathbf{Q})$. The test criterion for $\mathbf{H}_0$ has the form

$$\mathcal{T}_n^2 = A^{-2}(\varphi) \ \mathbf{S}_n'\mathbf{Q}_n^{-1}\mathbf{S}_n, \tag{2.17}$$

and its asymptotic null distribution is χ^2 with p degrees of freedom.
Under the local alternative

$$\mathbf{H}_n : \ \boldsymbol{\beta} = \boldsymbol{\beta}_n = n^{-1/2}\boldsymbol{\beta}^* \in \mathbb{R}^p, \tag{2.18}$$

$\mathcal{T}_n^2$ has the asymptotic noncentral χ^2 distribution with p degrees of freedom and with the noncentrality parameter

$$\eta^2 = \boldsymbol{\beta}^{*\prime}\mathbf{Q}\boldsymbol{\beta}^* \ \frac{\gamma^2(\varphi, F)}{A^2(\varphi)}. \tag{2.19}$$

However, while we observe Y_i's and their ranks, instead of $\mathbf{x}_{ni}$ we can observe only $\mathbf{w}_{ni} = \mathbf{x}_{ni} + \mathbf{u}_{ni}, \ i = 1, \ldots, n$, where $\mathbf{u}_{n1}, \ldots, \mathbf{u}_{nn}$ are i.i.d. p-dimensional errors, independent of the e_i. Their distribution is generally unknown, but we assume that they satisfy

$$\frac{1}{n} \sum_{i=1}^{n} (\mathbf{u}_{ni} - \bar{\mathbf{u}}_n)(\mathbf{u}_{ni} - \bar{\mathbf{u}}_n)' \xrightarrow{p} \mathbf{U} \tag{2.20}$$

as $n \to \infty$, where $\mathbf{U}$ is a positively definite $p \times p$ matrix. In the model with measurement errors, the test of $\mathbf{H}_0 : \ \boldsymbol{\beta} = \mathbf{0}$ can be only based on the observable statistics; hence we construct the test based on the vector of linear rank statistics $\widetilde{\mathbf{S}}_n \in \mathbb{R}^p$,

$$\widetilde{\mathbf{S}}_n = n^{-1/2} \sum_{i=1}^{n} (\mathbf{w}_{ni} - \bar{\mathbf{w}}_n) a_n(R_{ni}). \tag{2.21}$$

Under the above conditions, its asymptotic null distribution is normal $\mathcal{N}_p \left(\mathbf{0}, \ A^2(\varphi) \ (\mathbf{Q} + \mathbf{U})\right)$, while under $\mathbf{H}_n$ it has the asymptotic normal distribution $\mathcal{N}_p \left(\gamma(\varphi, F) \ \mathbf{Q}\boldsymbol{\beta}^*, \ A^2(\varphi) \ (\mathbf{Q} + \mathbf{U})\right)$.

The test criterion for $\mathbf{H}_0$ now would have the form

$$\widetilde{\mathcal{T}}_n^2 = A^{-2}(\varphi)\ \widetilde{\mathbf{S}}_n'(\mathbf{Q}_n + \mathbf{U_n})^{-1}\widetilde{\mathbf{S}}_n. \tag{2.22}$$

Under $\mathbf{H}_0$, the asymptotic distribution of $\widetilde{\mathcal{T}}_n^2$ is χ^2 with p degrees of freedom. Its asymptotic distribution under $\mathbf{H}_n$ will be noncentral χ^2 with with p degrees of freedom and with the noncentrality parameter

$$\tilde{\eta}^2 = \boldsymbol{\beta}^{*\prime}\ (\mathbf{Q}+\mathbf{U})^{-1/2}\ \mathbf{Q}\ (\mathbf{Q}+\mathbf{U})^{-1/2}\ \boldsymbol{\beta}^*\ \frac{\gamma^2(\varphi, F)}{A^2(\varphi)}. \tag{2.23}$$

Hence, the asymptotic relative efficiency of the test of $\mathbf{H}_0$ based on $\widetilde{\mathcal{T}}_n^2$ with respect to the test based on $\mathcal{T}_n^2$ is

$$e_{\widetilde{\mathcal{T}},\mathcal{T}} = \frac{\boldsymbol{\beta}^{*\prime}\mathbf{Q}\boldsymbol{\beta}^*}{\boldsymbol{\beta}^{*\prime}\ (\mathbf{Q}+\mathbf{U})^{-1/2}\ \mathbf{Q}\ (\mathbf{Q}+\mathbf{U})^{-1/2}\ \boldsymbol{\beta}^*}. \tag{2.24}$$

Conclusion

We can construct a distribution free rank test of hypothesis $\mathbf{H}_0 :\ \boldsymbol{\beta} = \mathbf{0}$ in the model (2.14) even when the regression matrix $\mathbf{X}$ can be only measured with additive errors. The test looses the efficiency with respect to the rank test under completely known $\mathbf{X}$.

3 Regression rank score tests with measurement errors under the alternative

Consider the model (1.4). The hypothesis $\mathbf{H}_0$ of interest is (1.5). Gutenbrunner et al. (1993) constructed a class of tests of $\mathbf{H}_0$ based on regression rank scores (3.1), that are invariant to the nuisance parameters $\beta_0, \boldsymbol{\beta}$. Gutenbrunner et al. (1993) showed that these tests are asymptotically equally efficient as the corresponding rank tests of $\mathbf{H}_0$ under known $\beta_0, \boldsymbol{\beta}$, under some conditions on the tails of the distribution F of the errors e_i.

The regression rank scores in the model (1.1) without errors are defined as the vector $\widehat{\mathbf{a}}_n(\tau) = (\hat{a}_{n1}(\tau), \ldots, \hat{a}_{nn}(\tau))'$ of solutions of the parametric linear programming problem

$$\begin{aligned}
&\sum_{i=1}^n Y_i\ \hat{a}_{ni}(\tau) := \max \\
\text{subject to}\quad & \\
&\sum_{i=1}^n \hat{a}_{ni}(\tau) = (1-\tau)n, \\
&\sum_{i=1}^n x_{ij}\ \hat{a}_{ni}(\tau) = (1-\tau)\sum_{i=1}^n x_{ij}, \quad j = 1\ldots, p, \\
&\widehat{\mathbf{a}}_n(\tau) \in [0,1]^n, \quad 0 \le \tau \le 1.
\end{aligned} \tag{3.1}$$

The regression rank scores are dual to the regression quantiles of model (1.1), and it follows from the duality that

$$\hat{a}_i(\tau) = \begin{cases} 1 & \text{if} \quad Y_i > \mathbf{x}_i' \, \widehat{\boldsymbol{\beta}}(\tau) \\ 0 & \text{if} \quad Y_i < \mathbf{x}_i' \, \widehat{\boldsymbol{\beta}}(\tau), \end{cases} \tag{3.2}$$

where $\widehat{\boldsymbol{\beta}}(\tau)$ is the regression τ-quantile for the model (1.1). The other values $\hat{a}_i(\tau)$ corresponding to $Y_i = \mathbf{x}_i'\widehat{\boldsymbol{\beta}}(\tau)$ are determined by the restrictions in (3.1) (notice that the number of such components is $p+1$).

It follows from the restrictions in (3.1) that $\widehat{\mathbf{a}}_n(\tau)$ is invariant to the transformations

$$Y_i \mapsto Y_i + b_0 + \mathbf{x}_i'\mathbf{b}, \quad b_0 \in \mathbb{R}^1, \ \mathbf{b} \in \mathbb{R}^p. \tag{3.3}$$

Gutenbrunner et al. (1993) constructed the tests under some conditions on the tail of F and on the $\mathbf{x}_{ni}$; for the sake of simplicity, we shall consider slightly stronger but simpler conditions. We shall assume that

(F.1) F has an absolutely continuous density f, that is positive for $A < x < B$ and decreases monotonically as $x \to A+$ and $x \to A-$, where

$$-\infty \leq A \equiv \sup\{x : \ F(x) = 0\} \ \text{ and } \ +\infty \geq B \equiv \inf\{x : \ F(x) = 1\}.$$

The derivative f' is bounded *a.e.* and

$$\left| \frac{f'(x)}{f(x)} \right| \leq c|x| \ \text{ for } \ |x| \geq K \geq 0, \ c > 0.$$

(F.2) $|F^{-1}(t)| \leq c(t(1-t))^{-\frac{1}{4}+\varepsilon}$ for $0 < t \leq t_0, \ 1 - t_0 \leq t < 1, \ \varepsilon > 0, \ c > 0.$

(F.3) $1/f(F^{-1}(t)) \leq c(t(1-t))^{-\frac{5}{4}+\varepsilon}$ for $0 < t \leq t_0, \ 1 - t_0 \leq t < 1, \ \varepsilon > 0, \ c > 0.$

Condition **(F.2)** demands that $\mathbb{E}|e_1|^{4+\delta} < \infty$ for some $\delta > 0$, and jointly with condition **(F.1)** it implies that F has finite Fisher information. Conditions **(F.1)**–**(F.3)** are satisfied for distributions with the tails of t distribution with 5 degrees of freedom or lighter; hence they are satisfied e.g. for the normal, logistic and Laplace distributions.

For the simplicity of notation, denote

$$\widetilde{\mathbf{X}}_n = \left[\mathbf{1}_n \vdots \ \mathbf{X}_n \right] \tag{3.4}$$

the matrix of order $n \times (p+1)$, where $\mathbf{1}_n = (1, \ldots, 1)' \in \mathbb{R}^n$, and denote $\tilde{\mathbf{x}}_{ni}$ the i-th row of $\widetilde{\mathbf{X}}_n$. We shall assume that

(X.1) $\lim_{n\to\infty} \mathbf{Q}_n = \mathbf{Q}$ where $\mathbf{Q}_n = n^{-1}\widetilde{\mathbf{X}}_n'\widetilde{\mathbf{X}}_n$ and $\mathbf{Q}$ is a positively definite $(p+1) \times (p+1)$ matrix.

(X.2) $\max_{1\leq i\leq n} \|\tilde{\mathbf{x}}_{ni}\| = O(1)$ and $n^{-1} \sum_{i=1}^n \|\tilde{\mathbf{x}}_{ni}\|^4 = O(1)$ as $n \to \infty$.

Let $\varphi:\ (0,1)\mapsto\mathbb{R}^1$ be a nondecreasing square integrable score-generating function such that $\varphi'(t)$ exists for $0<t<t_0,\ 1-t_0<t<1$ and satisfies

$$|\varphi'(t)|\leq c(t(1-t))^{-1-\delta},\ \delta>0. \tag{3.5}$$

Notice that the condition (3.5) covers the inverse normal distribution function, among others. Calculate the scores $\hat{b}_{ni}$ generated by φ in the form

$$\hat{b}_{ni}=-\int_0^1\varphi(t)\ d\hat{a}_{ni}(t),\ i=1,\ldots,n \tag{3.6}$$

and the q-dimensional vector of *linear regression rank scores statistics*

$$\mathbf{S}_n=n^{-1/2}(\mathbf{Z}_n-\widehat{\mathbf{Z}}_n)'\widehat{\mathbf{b}}_n,\quad \widehat{\mathbf{b}}_n=(\hat{b}_{n1},\ldots,\hat{b}_{nn})' \tag{3.7}$$

where

$$\widehat{\mathbf{Z}}_n=\widehat{\mathbf{H}}_n\mathbf{Z}_n,\ \widehat{\mathbf{H}}_n=\widetilde{\mathbf{X}}_n(\widetilde{\mathbf{X}}_n'\widetilde{\mathbf{X}}_n)^{-1}\widetilde{\mathbf{X}}_n'$$

is the projection of $\mathbf{Z}_n$ on the space spanned by the columns of $\widetilde{\mathbf{X}}_n$.

The test criterion for the hypothesis $\mathbf{H}_0$ is

$$\mathcal{T}_n^2=A^{-2}(\varphi)\ \mathbf{S}_n'\ \mathbf{D}_n^{-1}\ \mathbf{S}_n \tag{3.8}$$

where

$$\mathbf{D}_n=n^{-1}(\mathbf{Z}_n-\widehat{\mathbf{Z}}_n)'(\mathbf{Z}_n-\widehat{\mathbf{Z}}_n)\to\mathbf{D},\quad \text{as } n\to\infty\quad \text{(we assume)}, \tag{3.9}$$

where $\mathbf{D}$ is a positively definite $q\times q$ matrix.

Under $\mathbf{H}_0$ has $\mathcal{T}_n^2$ asymptotically χ^2 distribution with q degrees of freedom. Under the local alternative

$$\mathbf{H}_n:\ \boldsymbol{\delta}_n=n^{-1/2}\boldsymbol{\delta}^*\in\mathbb{R}^q, \tag{3.10}$$

$\mathcal{T}_n^2$ has asymptotic noncentral χ^2 distribution with q degrees of freedom and with the noncentrality parameter

$$\eta^2=\boldsymbol{\delta}^{*\prime}\ \mathbf{D}\ \boldsymbol{\delta}^*\ \frac{\gamma^2(\varphi,F)}{A^2(\varphi)}. \tag{3.11}$$

Let us now consider the situation that we still want to test the hypothesis $\mathbf{H}_0$, but the vectors $\mathbf{z}_{ni}$ can be only observed with errors, i.e. we observe $\mathbf{w}_{ni}=\mathbf{z}_{ni}+\mathbf{u}_{ni},\ i=1,\ldots,n$, where the $\mathbf{u}_{ni}\in\mathbb{R}^q$ are random errors. Denote

$$\mathbf{W}_n=\begin{bmatrix}\mathbf{w}_{n1}\\ \cdots\\ \mathbf{w}_{nn}\end{bmatrix}\quad\text{and}\quad \mathbf{U}_n=\begin{bmatrix}\mathbf{u}_{n1}\\ \cdots\\ \mathbf{u}_{nn}\end{bmatrix} \tag{3.12}$$

the $n\times q$ matrices and $\widehat{\mathbf{W}}_n=\widehat{\mathbf{H}}_n\mathbf{W}_n$ and $\widehat{\mathbf{U}}_n=\widehat{\mathbf{H}}_n\mathbf{U}_n$ their projections on the space spanned by the columns of $\widetilde{\mathbf{X}}_n$. We shall assume that

$$\mathbf{G}_n=n^{-1}(\mathbf{U}_n-\widehat{\mathbf{U}}_n)'(\mathbf{U}_n-\widehat{\mathbf{U}}_n)\xrightarrow{p}\mathbf{G}\quad\text{as}\quad n\to\infty \tag{3.13}$$

where $\mathbf{G}$ is a positively definite $q \times q$ matrix.

In this situation, we shall replace the statistic $\mathbf{S}_n$ by

$$\widetilde{\mathbf{S}}_n = n^{-1/2}(\mathbf{W}_n - \widehat{\mathbf{W}}_n)'\widehat{\mathbf{b}}_n \tag{3.14}$$

and the test criterion $\mathcal{T}_n^2$ by

$$\widetilde{\mathcal{T}}_n^2 = A^{-2}(\varphi)\ \widetilde{\mathbf{S}}_n'\ (\mathbf{D}_n + \mathbf{G}_n)^{-1}\ \widetilde{\mathbf{S}}_n. \tag{3.15}$$

Under $\mathbf{H}_0$, it has asymptotic χ^2 distribution with q degrees of freedom, while under the local alternative $\mathbf{H}_n$ in (3.10) it has asymptotic noncentral χ^2 distribution with q degrees of freedom and with the noncentrality parameter

$$\tilde{\eta}^2 = \boldsymbol{\delta}^{*\prime}\ (\mathbf{D} + \mathbf{G})\ \boldsymbol{\delta}^*\ \frac{\gamma^2(\varphi, F)}{A^2(\varphi)}. \tag{3.16}$$

The relative asymptotic efficiency of $\widetilde{\mathcal{T}}_n^2$ with respect to $\mathcal{T}_n^2$ then is

$$e_{\widetilde{\mathcal{T}},\mathcal{T}} = \frac{\boldsymbol{\delta}^{*\prime}\ \mathbf{D}\ \boldsymbol{\delta}^*}{\boldsymbol{\delta}^{*\prime}\ (\mathbf{D} + \mathbf{G})\ \boldsymbol{\delta}^*}. \tag{3.17}$$

Conclusion

If only regressors under the alternative in model (1.4) are affected by additive errors, we can still construct the distribution free tests based on regression rank scores under the hypothesis, provided the distribution of errors has the tails of t-distribution or lighter. The loss of efficiency is given in (3.17).

4 Some open problems

Let us turn back to model (1.4); consider the problem that the $\mathbf{x}_{ni}$, too, are observed only with errors, and we observe $\mathbf{w}_{ni} = \mathbf{x}_{ni} + \mathbf{u}_{ni}$ instead of $\mathbf{x}_{ni}$, $i = 1, \ldots, n$. Then the regression rank scores can be modified in the following way:

$$\begin{aligned} &\sum_{i=1}^{n} Y_i\ \tilde{a}_{ni}(\tau) := \max \\ \text{subject to}\quad &\\ &\sum_{i=1}^{n} \tilde{a}_{ni}(\tau) = (1-\tau)n, \\ &\sum_{i=1}^{n} w_{ij}\ \tilde{a}_{ni}(\tau) = (1-\tau)\sum_{i=1}^{n} w_{ij}, \quad j = 1\ldots, p, \\ &\widetilde{\mathbf{a}}_n(\tau) \in [0,1]^n, \quad 0 \le \tau \le 1. \end{aligned} \tag{4.1}$$

Similarly as in the model without errors, the regression rank scores of (4.1) are dual to the regression quantiles in the linear programming sense; however, not to the regression quantiles of the model without errors in the $\mathbf{x}_{ni}$, but rather to those of the model

$$Y_i = \beta_0 + \mathbf{w}_i'\boldsymbol{\beta} + e_i,\ i = 1, \ldots, n \tag{4.2}$$

with the covariates $\mathbf{w}_{ni},\ i = 1, \ldots, n$. Again, it follows from the duality that

$$\tilde{a}_{ni}(\tau) = \begin{cases} 1 & \text{if} \quad Y_i > \mathbf{w}_{ni}'\ \widetilde{\boldsymbol{\beta}}(\tau) \\ 0 & \text{if} \quad Y_i < \mathbf{w}_{ni}'\ \widetilde{\boldsymbol{\beta}}(\tau), \end{cases} \tag{4.3}$$

where $\widetilde{\boldsymbol{\beta}}(\tau)$ is the regression τ-quantile for the model (4.2). The other values $\tilde{a}_{ni}(\tau)$ corresponding to $Y_i = \mathbf{w}_{ni}'\ \widetilde{\boldsymbol{\beta}}(\tau)$ are determined by the restrictions in (4.1), and the number of such components is $p + 1$.

It follows from the restrictions in (4.1) that the vector of regression rank scores $\widetilde{\mathbf{a}}_n(\tau)$ is invariant to the transformations

$$Y_i \mapsto Y_i + b_0 + \mathbf{w}_{ni}'\mathbf{b} = Y_i + b_0 + \mathbf{x}_{ni}'\mathbf{b} + \mathbf{u}_{ni}'\mathbf{b}, \quad b_0 \in \mathbb{R}^1,\ \mathbf{b} \in \mathbb{R}^p; \tag{4.4}$$

but it is not generally invariant to transformations (3.3). In this setup, if we replace $\mathbf{S}_n$ by

$$\widetilde{\mathbf{S}}_n = n^{-1/2}(\mathbf{Z}_n - \widetilde{\mathbf{Z}}_n)'\ \widetilde{\mathbf{b}}_n \tag{4.5}$$

where the scores $\widetilde{\mathbf{b}}_n = (\tilde{b}_{n1}, \ldots, \tilde{b}_{nn})'$ are generated by φ from the $\tilde{a}_{ni}$ similarly as in (3.6) and where $\widetilde{\mathbf{Z}}_n$ is the projection of the matrix $\mathbf{Z}_n$ onto the space spanned by the columns of the matrix $[\mathbf{1}_n \vdots\ \mathbf{W}_n]$ with

$$\mathbf{W}_n = \begin{bmatrix} \mathbf{w}_{n1} \\ \cdots \\ \mathbf{w}_{nn} \end{bmatrix},$$

then the test based on $\widetilde{\mathbf{S}}_n$ is not asymptotically distribution free without additional conditions on the model. A possible construction of a distribution free test is still an open problem.

References

[1] Gutenbrunner, C. and Jurečková, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20**, 305–330.

[2] Gutenbrunner, C., Jurečková, J., Koenker, R. and Portnoy, S. (1993). Tests of linear hypotheses based on regression rank scores. *Nonpar. Statist.* **2**, 307–331.

[3] He, X. and Liang, H. (2000). Quantile regression estimate for a class of linear and partially linear errors-in-variables models. *Statistica Sinica* **10**, 129–140

[4] Hájek, J. and Šidák, Z. (1967). *Theory of rank tests.* Academia, Prague.

[5] Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica* **46**, 33–50.