Applications of Mathematics

Radim Navrátil; A. K. Md. Ehsanes Saleh

Aligned rank tests in measurement error model

Applications of Mathematics, Vol. 61 (2016), No. 1, 47–59

Persistent URL: `http://dml.cz/dmlcz/144811`

Terms of use:

© Institute of Mathematics AS CR, 2016

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* `http://dml.cz`

ALIGNED RANK TESTS IN MEASUREMENT ERROR MODEL

Radim Navrátil, Brno, A. K. Md. Ehsanes Saleh, Ottawa

(Received December 27, 2014)

Abstract. Aligned rank tests are introduced in the linear regression model with possible measurement errors. Unknown nuisance parameters are estimated first and then classical rank tests are applied on the residuals. Two situations are discussed: testing about an intercept in the linear regression model considering the slope parameter as nuisance and testing of parallelism of several regression lines, i.e. whether the slope parameters of all lines are equal. Theoretical results are derived and the simulation study is also made to illustrate good performance of introduced tests.

Keywords: aligned test; measurement error; rank test

MSC 2010: 62G10, 62J05

1. Introduction

Consider the classical linear regression model

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + e_i, \quad i = 1, \dots, n, \tag{1.1}$$

where $\beta_0 \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^p$ are unknown parameters, $\mathbf{x}_i$ are vectors of known regressors, model errors e_i are assumed to be independent identically distributed with an unknown distribution function F with finite Fisher information with respect to the location, i.e.

$$I(f) = \int_{-\infty}^{\infty} \Big(\frac{f'(x)}{f(x)}\Big)^2 f(x)\,\mathrm{d}x < \infty, \quad f(x) = F'(x).$$

This paper is based in part of Radim Navrátil's dissertation thesis at Charles University in Prague, the research originated and was supported by NSERC Grant A3088, Canada, during Radim Navrátil's visit at Carleton University, Ottawa. His research was also supported by Student Project Grant at MU (specific research, rector's programme) MUNI/A/1441/2014.

Measurement error model assumes that the regressors $\mathbf{x}_i$ are not observed accurately, but only with an additive, unobservable, error $\mathbf{v}_i$ (i.i.d. random variables independent with e_i), i.e. we observe $\mathbf{w}_i = \mathbf{x}_i + \mathbf{v}_i$ instead of $\mathbf{x}_i$. Hence, we may write our model as

$$\begin{aligned} Y_i &= \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + e_i, \\ \mathbf{w}_i &= \mathbf{x}_i + \mathbf{v}_i. \end{aligned} \tag{1.2}$$

There exists a rich literature about measurement error models and there have been developed a lot of different methods for dealing with measurement errors during last century. Most of them are interested in estimation problem, there is a lack of the literature about testing, although this problem might be as important as estimation. The bulk of the little literature about tests uses parametric approach with its restrictive normality assumptions or a knowledge of some additional information about error distribution (see e.g. [1]). We will avoid this and introduce a class of rank tests that will be valid even if measurement errors are present.

First attempts in this area were made by [7], who showed that some rank tests stay valid even if measurement errors are present, they only cause a decrease of tests' power. The papers [10], [11] and [14] generalized these results for other models and tests.

Let us start with an example that should warn you about thoughtless use of (rank) tests in measurement error models. The paper [7] showed a solution to the problem of testing the hypothesis $\mathbf{H}_{0,0}\colon \boldsymbol{\beta} = \mathbf{0}$, where the classical rank test for regression was extended to the measurement error model (1.2).

However, the problem may arise when we want to test the hypothesis $\mathbf{H}_{0,1}\colon \boldsymbol{\beta} = \boldsymbol{\beta}^* \neq \mathbf{0}$, with $\boldsymbol{\beta}^* \in \mathbb{R}^p$ known. In the model (1.1) without measurement errors we may transform it into a previous case of testing $\mathbf{H}_{0,0}\colon \boldsymbol{\beta} = \mathbf{0}$ by subtracting $\mathbf{x}_i^\top \boldsymbol{\beta}^*$ from both sides of (1.1):

$$Y_i^* = Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^* = \beta_0 + \mathbf{x}_i^\top (\boldsymbol{\beta} - \boldsymbol{\beta}^*) + e_i.$$

Using the same technique in the measurement error model (1.2), i.e. subtracting the term $\mathbf{w}_i^\top \boldsymbol{\beta}^*$, we get

$$Y_i^* = Y_i - \mathbf{w}_i^\top \boldsymbol{\beta}^* = \beta_0 + \mathbf{x}_i^\top (\boldsymbol{\beta} - \boldsymbol{\beta}^*) - \mathbf{v}_i^\top \boldsymbol{\beta}^* + e_i. \tag{1.3}$$

Unlike previous case we did not get rid of $\boldsymbol{\beta}^*$ (under $\boldsymbol{\beta} = \boldsymbol{\beta}^*$) from the right-hand side of the equation (1.3) and the test will not work. We may illustrate this on the following simulation example. Consider the model of regression line passing through the origin

$$Y_i = x_i \beta + e_i, \quad i = 1, \ldots, 50 \text{ with true } \beta = 2.$$

The regressors x_i were once generated from a sample of size $n = 50$ from uniform $(-2, 10)$ distribution and then considered fixed, the model errors e_i were generated from standard normal distribution. The empirical power of the Wilcoxon test for regression was computed as a percentage of rejections of $\mathbf{H}_{0,1}: \beta = 2$ among 10 000 replications, at significance level $\alpha = 0.05$. The results are summarized in Table 1.

β	v_i: 0	$\mathcal{N}(0,1)$	$\mathcal{N}(0,0.5)$	$\mathcal{U}(-1,1)$	$\mathcal{U}(-0.5,0.5)$	$\mathcal{U}(-2,2)$
2.00	5.06	39.73	19.92	13.48	5.98	53.69
1.80	99.41	97.15	96.84	97.01	98.80	97.44
1.85	93.02	91.41	88.51	87.96	90.62	93.72
1.90	63.20	79.06	69.15	65.66	62.93	85.28
1.95	22.06	60.73	42.89	36.52	25.17	72.34
2.05	21.71	21.86	7.84	5.14	10.81	35.80
2.10	63.94	9.38	5.81	9.22	39.16	20.00
2.15	92.98	4.72	14.30	25.75	76.05	9.56
2.20	99.50	2.20	31.96	51.81	94.92	4.66

Table 1. Percentage of rejections of hypothesis $\mathbf{H}_{0,1}: \beta = 2$ for various measurement errors v_i for Wilcoxon test for regression.

The previous example illustrates that we have to be very careful when dealing with measurement errors and not only recklessly without thinking use methods for the model without measurement errors.

In the followin two sections we will introduce an aligned ranktest about an intercept and an aligned rank test of parallelism forseveral regression lines, both with possible measurement errors. In both cases we first estimate the nuisance parameter in the model (1.2) and then apply standard rank test on residuals. Although such estimates are biased, this inconsistency disappears when considering residuals.

First, the test statistics are introduced, their distribution is derived both under null hypothesis and local alternatives. Finally, in Section 4 the simulation study is made to illustrate good behavior of these tests, influence of measurement errors is identified and the power of the tests is compared with the model without measurement errors.

2. Test about an intercept

Consider the measurement error model (1.2), where β_0 is an unknown intercept parameter of our interest and $\boldsymbol{\beta}$ is p-dimensional vector of unknown nuisance parameters, and both the model errors e_i and the measurement errors $\mathbf{v}_i$ are assumed to be symmetric. The symmetry assumption of $\mathbf{v}_i$ is very natural, because it means that the measurement is not affected by a systematic error. In case of systematic error it would be impossible to distinguish which part of the regressor belongs to the original one a which to the errors. To deal with this situation we would need some additional information about the measurement errors.

Our aim is to test the null hypothesis $\mathbf{H}_1\colon\ \beta_0 = 0$ against the alternative $\beta_0 > 0$. Without any further information about the regressors $\mathbf{x}_i$ it is impossible to make statistical inference about the parameter β_0 (problem of identifiability). To be able to test $\mathbf{H}_1$, we will assume that the regressors $\mathbf{x}_i$ are centered, i.e. $\sum\limits_{i=1}^{n} x_{i,j} = 0$ for all $j = 1, \dots, p$.

Remark. The assumption that $\sum\limits_{i=1}^{n} x_{i,j} = 0$ for all $j = 1, \dots, p$ is quite strong, but essential. In the model without measurement errors (1.1) aligned rank tests work without this assumption. One may use the following reparametrization:

$$Y_i = \beta_0 + (\mathbf{x}_i - \bar{\mathbf{x}})^\top \widetilde{\boldsymbol{\beta}} + e_i, \quad i = 1, \dots, n.$$

However, this parametrization cannot be used in the measurement error model (1.2) due to the fact that we do not observe the original regressors $\mathbf{x}_i$. Even if we considered a different parametrization of the model (1.2), the intercept still could not be correctly identified.

If $\boldsymbol{\beta}$ is known, then we may rewrite (1.2) as

$$Y_i^* = Y_i - \mathbf{w}_i^\top \boldsymbol{\beta} = \beta_0 + e_i^*,$$

where $e_i^* = e_i - \mathbf{v}_i^\top \boldsymbol{\beta}$ are i.i.d. model errors with symmetric density f_β^*.

We will test the hypothesis $\mathbf{H}_1\colon\ \beta_0 = 0$ with the aid of signed rank test (see for instance [2]). Choose a square integrable score function $\varphi\colon\ (0,1) \mapsto \mathbb{R}$ and define $\varphi^+(u) = \varphi((u+1)/2)$, the approximate scores $a_n^+(i) = \varphi^+(i/(n+1))$ and the signed rank statistic

$$S_n^+(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_n^+(R_i^+(\boldsymbol{\beta}))\operatorname{sign}(Y_i^*), \tag{2.1}$$

where $R_i^+(\boldsymbol{\beta})$ is the rank of $|Y_i^*|$ among $|Y_1^*|, \dots, |Y_n^*|$.

The distribution of $S_n^+(\boldsymbol{\beta})$ under $\mathbf{H}_1$ is distribution-free and for small n can be computed directly; for large n normal distribution approximation holds:

$$T_n(\boldsymbol{\beta}) = A^{-1}(\varphi^+) S_n^+(\boldsymbol{\beta}), \quad \text{with} \quad A^2(\varphi) = \int_0^1 \varphi^2(t)\,\mathrm{d}t - \Big(\int_0^1 \varphi(t)\,\mathrm{d}t\Big)^2,$$

has asymptotically standard normal distribution as $n \to \infty$.

Under the local alternative

$$\mathbf{K}_{1n}\colon\ \beta_0 = n^{-1/2}\beta_0^*, \quad \beta_0^* \in \mathbb{R} \text{ fixed},$$

the test statistic $T_n(\boldsymbol{\beta})$ has asymptotically normal distribution with mean μ a variance 1, where

$$\mu = \beta_0^*\gamma(\varphi^+, f_\beta^*), \quad \gamma(\varphi, f) = -\int_0^1 \varphi(u)\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}\,\mathrm{d}u. \tag{2.2}$$

For more details see [2].

However, in our situation $\boldsymbol{\beta}$ is unknown, hence we have to estimate it first and then consider the signed rank test based on aligned ranks of the residuals.

In general, as an estimator of $\boldsymbol{\beta}$ we may take any $\sqrt{n}$-consistent estimate of $\boldsymbol{\beta}$. Anyway, we want to preserve the robust properties that rank tests have, hence as an estimator $\widehat{\boldsymbol{\beta}}_n$ we take the R-estimator based on the hypothetical model affected by the measurement errors

$$Y_i = \mathbf{w}_i^\top\boldsymbol{\beta} + e_i^*.$$

The paper [3] introduced a class of estimators of the location parameter in one- and two- sample location models, by inverting a class of rank tests for the location. This methodology was then extended to linear regression models without measurement error by [5].

For $\mathbf{b} \in \mathbb{R}^p$ denote by $\widetilde{R}_i(\mathbf{b})$ the rank of the residual $(Y_i - \mathbf{w}_i^\top\mathbf{b})$ among $(Y_1 - \mathbf{w}_1^\top\mathbf{b}), \ldots, (Y_n - \mathbf{w}_n^\top\mathbf{b})$. Choose a square integrable, skew-symmetric score function $\psi\colon\ (0,1) \mapsto \mathbb{R}$ and define the approximate scores $a_n(i) = \psi(i/(n+1))$ and the vector of aligned rank statistics

$$\mathbf{L}_n(\mathbf{b}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n(\mathbf{w}_i - \overline{\mathbf{w}}_n)a_n(\widetilde{R}_i(\mathbf{b})).$$

Then $L_{n,j}(\mathbf{b})$ is stepwise, nonincreasing, symmetrically distributed around 0 for all $j = 1, \ldots, p$ provided $\mathbf{b} = \boldsymbol{\beta}$. The R-estimator of $\boldsymbol{\beta}$ is then defined as

$$\widehat{\boldsymbol{\beta}}_n = \operatorname{argmin}\{\|\mathbf{L}_n(\mathbf{b})\|,\ \mathbf{b} \in \mathbb{R}^p\}, \tag{2.3}$$

where $\|\cdot\|$ stands for any norm on $\mathbb{R}^p$.

Originally, the estimate (2.3) was defined in [5] with the aid of l^1-norm, then [9] used l^2-norm and finally [8] proved that for any norm the corresponding R-estimates are asymptotically equivalent. In [4] the author defined his estimate as a minimizer of a measure of rank dispersion of residuals

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^{n}(Y_i - \mathbf{w}_i^\top \mathbf{b})a_n(\widetilde{R}_i(\mathbf{b})), \tag{2.4}$$

with respect to $\mathbf{b} \in \mathbb{R}^p$. He also showed that $-n^{1/2}\mathbf{L}_n(\mathbf{b})$ is the subgradient of $\mathcal{D}_n(\mathbf{b})$; hence the estimator defined as a minimizer of $\mathcal{D}_n$ exists and is equivalent to the above estimators based on $\mathbf{L}_n$ (see [8]).

Note that the estimate $\widehat{\boldsymbol{\beta}}_n$ is not a consistent estimate of the parameter $\boldsymbol{\beta}$, in fact it is asymptotically biased—see [6], or [14]. However, for the testing procedure we will introduce this does not matter and this "inconsistency" disappears because of multiplying $\widehat{\boldsymbol{\beta}}_n$ by $\mathbf{w}_i$.

Now, consider the residuals $\widehat{e}_1 = Y_1 - \mathbf{w}_1^\top \widehat{\boldsymbol{\beta}}_n$, …, $\widehat{e}_n = Y_n - \mathbf{w}_n^\top \widehat{\boldsymbol{\beta}}_n$ and insert $\widehat{\boldsymbol{\beta}}_n$ into (2.1) to get aligned signed rank statistic

$$S_n^+(\widehat{\boldsymbol{\beta}}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} a_n^+(R_i^+(\widehat{\boldsymbol{\beta}}_n))\operatorname{sign}(\widehat{e}_i), \tag{2.5}$$

or we may also use the standardized version

$$T_n(\widehat{\boldsymbol{\beta}}_n) = A^{-1}(\varphi^+)S_n^+(\widehat{\boldsymbol{\beta}}_n). \tag{2.6}$$

The distribution of $S_n^+(\widehat{\boldsymbol{\beta}}_n)$ under $\mathbf{H}_1$ is no longer distribution-free because of the inserted estimate $\widehat{\boldsymbol{\beta}}_n$. Anyway, the asymptotic distribution remains the same. To prove this, we need to add some assumptions on the regressors. Suppose that there exist positive definite matrices $\mathbf{Q}$, $\mathbf{V}$ such that as $n \to \infty$:

$$\mathbf{Q}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}}_n)(\mathbf{x}_i - \overline{\mathbf{x}}_n)^\top \longrightarrow \mathbf{Q}, \tag{2.7}$$

$$\mathbf{V}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}_i - \overline{\mathbf{v}}_n)(\mathbf{v}_i - \overline{\mathbf{v}}_n)^\top \xrightarrow{p} \mathbf{V}, \tag{2.8}$$

$$\frac{1}{n}\max_{1\leqslant i\leqslant n}(\mathbf{x}_i - \overline{\mathbf{x}}_n)^\top \mathbf{Q}_n^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}_n) \longrightarrow 0, \tag{2.9}$$

$$\frac{1}{n}\max_{1\leqslant i\leqslant n}(\mathbf{v}_i - \overline{\mathbf{v}}_n)^\top \mathbf{V}_n^{-1}(\mathbf{v}_i - \overline{\mathbf{v}}_n) \xrightarrow{p} 0. \tag{2.10}$$

Theorem 2.1. *Let the conditions* (2.7)–(2.10) *be satisfied. Then the test statistic* $T_n(\widehat{\boldsymbol{\beta}}_n)$ *in the model* (1.2) *has asymptotically standard normal distribution under* $\mathbf{H}_1$ *and under the local alternative*

$$\mathbf{K}_{1n}\colon\ \beta_0 = n^{-1/2}\beta_0^*, \quad \beta_0^* \in \mathbb{R} \text{ fixed},$$

it has asymptotically $\mathcal{N}(\mu, 1)$ *distribution with* μ *defined in* (2.2).

P r o o f. According to Theorem 7.2.1 in [13] we have under the assumptions of Theorem 2.1 the following asymptotic representation of the R-estimate:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \frac{1}{\gamma(\psi, f_\beta^*)}(\mathbf{Q}+\mathbf{V})^{-1}\mathbf{L}_n(\boldsymbol{\beta}) + o_p(1) \quad \text{as } n \to \infty.$$

In addition, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ is asymptotically normally distributed (see [6]), hence in particular bounded in probability, i.e. for any $\varepsilon > 0$ there exists $K > 0$ such that

$$P(\|\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\| > K) < \varepsilon \quad \forall\, n. \tag{2.11}$$

The signed rank statistic S_n^+ is uniformly asymptotically linear on any compact set, i.e. for any fixed $K > 0$ and as $n \to \infty$ (see [8]):

$$\sup_{\|\mathbf{t}\| \leqslant K} \Big\{ \Big| S_n^+\Big(\boldsymbol{\beta} + \frac{\mathbf{t}}{\sqrt{n}}\Big) - S_n^+(\boldsymbol{\beta}) \Big| \Big\} \xrightarrow{p} 0. \tag{2.12}$$

Inserting $\mathbf{t} = \sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ into (2.12), together with (2.11) we get

$$|S_n^+(\widehat{\boldsymbol{\beta}}) - S_n^+(\boldsymbol{\beta})| \xrightarrow{p} 0 \quad \text{as } n \to \infty.$$

Hence the asymptotic distribution of $S_n^+(\widehat{\boldsymbol{\beta}})$ is the same as that of $S_n^+(\boldsymbol{\beta})$ discussed at the beginning of this section, which implies that $T_n(\widehat{\boldsymbol{\beta}}_n) = A^{-1}(\varphi^+)S_n^+(\widehat{\boldsymbol{\beta}}_n)$ has under $\mathbf{H}_1$ asymptotically standard normal distribution as $n \to \infty$ and under the local alternative $\mathbf{K}_{1n}$ asymptotically normal distribution with mean μ and variance 1. □

R e m a r k. As far as practical applications are concerned, there arises a natural question how to choose the score function φ. According to [2], Theorem 3.4.9, the locally most powerful rank test for $\mathbf{H}_1$ against $\mathbf{K}_1$ is based on the test statistic

$$\sum_{i=1}^{n} \operatorname{sign}(Y_i)\varphi^+\Big(\frac{R_i^+}{n+1}, f\Big),$$

where

$$\varphi^+(u, f) = \varphi\Big(\frac{u+1}{2}, f\Big), \quad \varphi(u, f) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}.$$

For normal model errors $\varphi(u,f) = \Phi^{-1}(u)$, where Φ^{-1} is the quantile function of standard normal distribution; this choice leads to the van der Waerden test. The Wilcoxon test ($\varphi(u,f) = 2u-1$) is the locally most powerful rank test for logistic and sign test for double exponential (Laplace) distribution of model errors.

The optimal φ could be chosen based on the estimate of unknown model errors. Anyway, the simplest choice of the Wilcoxon test provides very reasonable results (see the simulations). The choice of the ψ function does not affect the asymptotic properties of the test statistic.

3. Test of parallelism

In this section we extend the results of [15] to the measurement error model. Consider the regression model of several lines

$$Y_i^{(j)} = \beta_0^{(j)} + \beta^{(j)} x_i^{(j)} + e_i^{(j)}, \quad i = 1,\dots,n_j,\ j = 1,\dots,p, \tag{3.1}$$

where $\beta_0^{(1)},\dots,\beta_0^{(p)}$ are unknown (nuisance) intercept parameters, $\beta^{(1)},\dots,\beta^{(p)}$ are unknown slope parameters of our interest, $x_i^{(j)}$ are known (fixed) or stochastic regressors, mutually uncorrelated for all $j = 1,\dots,p$. The model errors $e_1^{(j)},\dots,e_{n_j}^{(j)}$ are assumed to be independent identically distributed with an unknown joint distribution function F with finite Fisher information with respect to the location and mutually independent for all $j = 1,\dots,p$.

Our problem is to test the null hypothesis

$$\mathbf{H}_2\colon\ \beta^{(1)} = \ldots = \beta^{(p)}$$

against the alternative that $\beta^{(1)},\dots,\beta^{(p)}$ are not all equal.

Again, the regressors are subject to measurement errors, i.e. we do not observe $x_i^{(j)}$, instead we observe $w_i^{(j)} = x_i^{(j)} + v_i^{(j)}$; $v_i^{(j)}$ are (unobservable) additive measurement errors mutually uncorrelated for all $j = 1,\dots,p$ and uncorrelated with $x_i^{(j)}$ (for stochastic regressors); for fixed regressors let for all $j = 1,\dots,p$

$$\frac{1}{n_j}\sum_{i=1}^{n_j}(x_i^{(j)} - \bar{x}_{n_j}^{(j)})(v_i^{(j)} - \overline{v}_{n_j}^{(j)}) \xrightarrow{p} 0 \quad \text{as } n_j \to \infty.$$

Let us now define

$$\bar{x}_{n_j}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i^{(j)}, \quad Q_{n_j}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j}(x_i^{(j)} - \bar{x}_{n_j}^{(j)})^2,$$

$$\overline{v}_{n_j}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j} v_i^{(j)}, \quad V_{n_j}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j}(v_i^{(j)} - \overline{v}_{n_j}^{(j)})^2,$$
$$\overline{w}_{n_j}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j} w_i^{(j)}, \quad W_{n_j}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j}(w_i^{(j)} - \overline{w}_{n_j}^{(j)})^2.$$

We shall assume that there exist positive numbers $Q^{(1)}, \dots, Q^{(p)}$ and $V^{(1)}, \dots, V^{(p)}$ such that as $n_j \to \infty$

$$Q_{n_j}^{(j)} \longrightarrow Q^{(j)}, \quad \frac{1}{n_j}\max_{i=1,\dots,n_j}\frac{(x_i^{(j)} - \bar{x}_{n_j}^{(j)})^2}{Q_{n_j}^{(j)}} \longrightarrow 0, \tag{3.2}$$

$$V_{n_j}^{(j)} \xrightarrow{p} V^{(j)}, \quad \frac{1}{n_j}\max_{i=1,\dots,n_j}\frac{(v_i^{(j)} - \overline{v}_{n_j}^{(j)})^2}{V_{n_j}^{(j)}} \xrightarrow{p} 0. \tag{3.3}$$

For stochastic regressors let the convergence in (3.2) hold in probability.

In the model (3.1) while testing $\mathbf{H}_2\colon \beta^{(1)} = \ldots = \beta^{(p)} = \beta$ the problem is that the hypothetical common value β is unknown. In this case we have to first estimate it and then use an aligned test. For simplification of notation, denote $n = \sum_{j=1}^{p} n_j$ and further denote the pooled sample of responses by $Y_1^{(1)}, \dots, Y_{n_1}^{(1)} \dots, Y_1^{(p)}, \dots, Y_{n_p}^{(p)}$ and the corresponding regressors by $x_1, \dots, x_n$, $w_1, \dots, w_n$, respectively.

The following setup was proposed in [15] for the model without measurement errors. We will generalize it and show that it also works for the measurement error model. Again, first we have to estimate the parameter β in the same way as in the previous section as a minimizer of $L_n(b)$—see (2.3). In particular the R-estimator of β may be then defined as

$$\widehat{\beta}_n = \frac{1}{2}\{\sup\{b\colon\ L_n(b) > 0\} + \inf\{b\colon\ L_n(b) < 0\}\}. \tag{3.4}$$

For each sample $j = 1, \dots, p$ we may choose different square integrable score function $\varphi^{(j)}\colon\ (0,1) \mapsto \mathbb{R}$ and define the approximate scores $a_{n_j}^{(j)}(i) = \varphi^{(j)}(i/(n_j+1))$ and the corresponding aligned statistic

$$\widehat{S}_{n_j}^{(j)} = \frac{1}{\sqrt{W_{n_j}^{(j)} A(\varphi^{(j)})}}\sum_{i=1}^{n_j}(w_i^{(j)} - \overline{w}_{n_j}^{(j)})a_{n_j}^{(j)}(\widehat{R}_i^{(j)}),$$

where $\widehat{R}_1^{(j)}, \dots, \widehat{R}_{n_j}^{(j)}$ are the ranks of $Y_1^{(j)} - w_1^{(j)}\widehat{\beta}_n, \dots, Y_{n_j}^{(j)} - w_{n_j}^{(j)}\widehat{\beta}_n$.

For testing $\mathbf{H}_2$ we use the statistic

$$\widehat{T}_n^2 = \sum_{j=1}^{p}(\widehat{S}_{n_j}^{(j)})^2. \tag{3.5}$$

Theorem 3.1. *Let the conditions* (3.2)–(3.3) *be satisfied. Then in the model* (3.1) *with measurement errors* $v_i^{(j)}$, *the test statistic* $\widehat{T}_n^2$ *has asymptotically, as* $n_1 \to \infty, \ldots, n_p \to \infty$, χ^2*-distribution with* $p-1$ *degrees of freedom.*

Under the local alternative

$$\text{(3.6)} \quad \mathbf{K}_{2n}\colon\ \beta_j = \beta + n_j^{-1/2}\theta_j,\ j = 1, \ldots, p,\ \theta_j \in \mathbb{R} \text{ fixed such that } \sum_{j=1}^{p} \Delta_j \theta_j = 0 \\ \text{and } \Delta_j = \lim n_j/n \text{ as } n_1 \to \infty, \ldots, n_p \to \infty,$$

$\widehat{T}_n^2$ *has asymptotically* χ^2*-distribution with* $p-1$ *degrees of freedom and noncentrality parameter*

$$\delta = \sum_{j=1}^{p} I(f_{\beta_j}^*)\Delta_j \theta_j^2 \gamma^2(\varphi^{(j)}, f_{\beta_j}^*)(Q_j + V_j).$$

P r o o f. The proof of the asymptotic distributions is analogous to the previous one for the test about an intercept. According to Theorem 3.2 in [15] for the test of parallelism in the model without measurement errors we have that under $\mathbf{H}_2$ test statistic $\widehat{T}_n^2$ has in the model (3.1) without measurement error asymptotically χ^2-distribution with $p-1$ degrees of freedom and under $\mathbf{K}_{2n}$ asymptotically χ^2-distribution with $p-1$ degrees of freedom with noncentrality parameter δ.

The formulas (3.2)–(3.3) imply that that there exists a limit (in probability) of

$$W_n = \frac{1}{n}\sum_{j=1}^{n}(w_i - \overline{w}_n)^2 \to Q + V = \sum_{i=1}^{p} \Delta_i (Q_i + V_i).$$

Next, for every fixed $\beta \in \mathbb{R}$

$$\text{(3.7)} \qquad \sqrt{n}(\widehat{\beta}_n - \beta) = \frac{1}{\gamma(\psi, f_\beta^*)(Q+V)} L_n(\beta) + o_p(1) \quad \text{as } n \to \infty,$$

where f_β^* is the density of $e_i^* = e_i - v_i\beta$ (see [8]). This result also holds for stochastic regressors, see [12]. Then inserting this result into Theorem 3.2 in [15] completes the proof. □

The previous result may be extended in a straightforward manner to multidimensional parameters $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(p)}$ and testing that all the parameters are equal (as vectors).

4. Numerical illustration

We made an extensive simulation study to illustrate how the proposed test procedures work in finite sample situation and indicate influence of the measurement errors both for the test about an intercept and the test of parallelism. Because of the lack of space we will present here only the first one. However, the corresponding simulation results for the test of parallelism are very similar to those for the test about an intercept.

All the simulations were performed in the statistical software `R` using standard tools and libraries, the random numbers generator was set up with the initial value `set.seed(15)`.

Consider the model of regression line

$$Y_i = \beta_0 + x_i\beta + e_i, \quad i = 1, \ldots, 50,$$

and test $\mathbf{H}_1\colon\ \beta_0 = 0$ against $\beta_0 > 0$. The regressors x_i were once generated from a sample of size $n = 50$ from uniform $(-6, 6)$ distribution, centered and then considered as fixed design points, the model errors e_i were generated from standard normal distribution. We considered the Wilcoxon aligned signed rank test that corresponds to the score function $\varphi(u) = 2u-1$. For the estimation of the nuisance parameter the score function $\psi(u) = 2u - 1$ was used. The empirical powers of the tests were computed as a percentage of rejections of $\mathbf{H}_1$ among 10 000 replications, at significance level $\alpha = 0.05$.

Empirical powers of the Wilcoxon aligned signed rank test for various measurement errors v_i are summarized in Table 2 (the value of the nuisance parameter β was taken $\beta = 1$).

β_0	v_i:	0	$\mathcal{N}(0,1)$	$\mathcal{N}(0,2)$	$\mathcal{U}(-1,1)$	$\mathcal{U}(-2,2)$	t(4)
0		5.59	5.53	5.51	5.41	5.70	5.48
0.1		17.40	13.36	11.91	15.58	12.54	11.97
0.2		38.74	25.40	20.74	32.28	23.54	23.09
0.3		65.10	43.71	34.55	55.30	39.20	38.93
0.4		86.22	62.51	50.14	76.65	56.23	54.04
0.5		96.36	79.60	66.81	90.68	73.87	70.86
0.6		99.14	90.72	80.27	96.76	85.69	83.70

Table 2. Percentage of rejections of the hypothesis $\mathbf{H}_1\colon\ \beta_0 = 0$ for various measurement errors v_i for the Wilcoxon aligned signed rank test; $\beta = 1$.

Empirical power of the Wilcoxon aligned signed rank test for various measurement errors v_i and for various values of nuisance parameter β are summarized in Table 3 (the true value of β_0 was taken $\beta_0 = 0.3$).

β	v_i:	0	$\mathcal{N}(0,1)$	$\mathcal{N}(0,2)$	$\mathcal{U}(-1,1)$	$\mathcal{U}(-2,2)$	t(4)
0		66.14	65.17	65.29	65.91	65.79	66.21
-0.5		65.85	58.56	53.02	62.79	56.11	54.70
0.5		65.68	57.92	52.71	62.44	55.57	54.65
-1		65.21	44.66	35.64	55.36	38.95	37.66
1		66.40	43.08	34.69	55.10	38.90	37.16
-2		66.05	24.72	18.43	37.65	20.83	21.42
2		66.08	24.62	18.36	38.09	20.94	21.30

Table 3. Percentage of rejections of the hypothesis $\mathbf{H}_1\colon \beta_0 = 0$ for various measurement errors v_i for the Wilcoxon aligned signed rank test; $\beta_0 = 0.3$.

We performed more simulations for other choices of regressors x_i, model errors e_i, measurement errors v_i, score functions φ and ψ, sample size n and model parameters β and β_0. However, corresponding results are similar to those in Tables 2 and 3. Our simulation shows that the proposed test actually works, the error of the first kind is under control (it is around the prescribed $\alpha = 0.05$); only its power decreases with increasing variance of measurement errors. Unlike the model without measurement errors, the power of the proposed test does depend on the nuisance parameter β—the greater value of β the smaller power. This is not surprising, because greater value of β means greater influence of measurement errors.

References

[1] *W. A. Fuller*: Measurement Error Models. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1987.

[2] *J. Hájek, Z. Šidák, P. K. Sen*: Theory of Rank Tests. Probability and Mathematical Statistics, Academic Press, Orlando, 1999.

[3] *J. L. Hodges,* Jr.*, E. L. Lehmann*: Estimates of location based on rank tests. Ann. Math. Stat. *34* (1963), 598–611.

[4] *L. A. Jaeckel*: Estimating regression coefficients by minimizing the dispersion of the residuals. Ann. Math. Stat. *43* (1972), 1449–1458.

[5] *J. Jurečková*: Nonparametric estimate of regression coefficients. Ann. Math. Stat. *42* (1971), 1328–1338.

[6] *J. Jurečková, H. L. Koul, R. Navrátil, J. Picek*: Behavior of R-estimators under measurement errors. Accepted to Bernoulli, arXiv:1411.3609.

[7] *J. Jurečková, J. Picek, A. K. Md. E. Saleh*: Rank tests and regression and rank score tests in measurement error models. Comput. Stat. Data Anal. *54* (2010), 3108–3120.

[8] *J. Jurečková, P. K. Sen*: Robust Statistical Procedures. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1996.

[9] *H. L. Koul*: Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. Ann. Math. Stat. *40* (1969), 1950–1979.

[10] *R. Navrátil*: Rank tests of symmetry in measurement errors models. WDS2010 Proceedings of Contributed Papers: Part I, Mathematics and Computer Sciences (2010), 177–182.

[11] *R. Navrátil*: Rank tests of symmetry and R-estimation of location parameter under measurement errors. Acta Univ. Palacki. Olomuc., Fac. Rerum Nat., Math. *50* (2011), 95–102.
[12] *J. Picek*: Statistical Procedures Based on Regression Rank Scores. Ph.D. thesis, Charles University in Prague, 1996.
[13] *M. L. Puri, P. K. Sen*: Nonparametric Methods in General Linear Models. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1985.
[14] *A. K. Md. E. Saleh, J. Picek, J. Kalina*: R-estimation of the parameters of a multiple regression model with measurement errors. Metrika *75* (2012), 311–328.
[15] *P. K. Sen*: On a class of rank order tests for the parallelism of several regression lines. Ann. Math. Stat. *40* (1969), 1668–1683.

Authors' addresses: *Radim Navrátil*, Masaryk University, Faculty of Science, Department of Mathematics and Statistics, Kotlářská 2, 611 37 Brno, Czech Republic, e-mail: `navratil@math.muni.cz`; *A. K. Md. Ehsanes Saleh*, Carleton University, School of Mathematics and Statistics, 1125 Colonel By Drive, Ottawa, Canada, e-mail: `saleh@math.carleton.ca`.