Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS

Radim Navrátil

Tests of statistical hypotheses in measurement error models

Department of Probability and Mathematical Statistics

Supervisor of the doctoral thesis:	prof. RNDr. Jana Jurečková, DrSc.
Study programme:	Mathematics
Specialization:	Probability and Mathematical Statistics

Prague 2014

First of all, I would like to thank my supervisor Professor Jana Jurečková for her patient guidance throughout this work. Without her help I would never finish the thesis, she was always able to encourage me every time when I was stuck.

I would also like to thank Professor Jana Jurečková for giving me the opportunity to join Jaroslav Hájek center for theoretical and applied statistics and Professor Jan Picek for incorporating me into the Project Klimatext at Technical University of Liberec.

Finally, I am very thankful to Professor Ehsanes Saleh from Carleton University, Ottawa, Professor Hira L. Koul from Michigan State University, East Lansing and Professor Hanu Oja from University of Turku, that they allowed me to stay for a few weeks at their universities. They also helped me a lot with writing my thesis.

This work was financially supported by Charles University Grant GAUK 105610, Research Project LC06024, the Grants SVV 265315, SVV-2013-267315, SVV-2014-260105 and the Project Klimatext CZ.1.07/2.3.00/20.0086.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

Prague, June 2, 2014 ....

Název práce: *Testy statistických hypotéz za přítomnosti chyb měření*

Autor: *Radim Navrátil*

Katedra: *Katedra pravděpodobnosti a matematické statistiky*

Vedoucí disertační práce: *prof. RNDr. Jana Jurečková, DrSc., Katedra pravděpodobnosti a matematické statistiky*

Abstrakt: V práci bylo studováno chování pořadových testů a odhadů v modelech s chybami měření – jestli zůstanou platné a použitelné i za přítomnosti chyb měření, případně jak mají být modifikovány. Byl navržen pořadový test o regresním parametru založený na odhadu minimalizujícím vzdálenost a pořadový test o absolutním členu. Bylo také zkoumáno (asymptotické) vychýlení R-odhadů v modelech s chybami měření. Heteroskedasticita v regresním modelu byla také zkoumána – byly navrženy testy heteroskedasticity s rušivou regresí a testy významnosti regrese s rušivou heteroskedasticitou založené na regresních pořadových skórech. Konečně, v modelu polohy bylo studováno chování testů a odhadů parametru posunutí pro různé chyby měření. Všechny výsledky byly odvozeny teoreticky a poté ilustrovány numericky příklady a simulacemi.

Klíčová slova: modely s chybami měření, pořadové testy, R-odhady, regresní pořadové skóry

Title: *Tests of statistical hypotheses in measurement error models*

Author: *Radim Navrátil*

Department: *Department of Probability and Mathematical Statistics*

Supervisor: *prof. RNDr. Jana Jurečková, DrSc., Department of Probability and Mathematical Statistics*

Abstract: The behavior of rank procedures in measurement error models was studied – if tests and estimates stay valid and applicable when there are some measurement errors involved and if not how to modify these procedures to be able to do some statistical inference. A new rank test for the slope parameter in regression model based on minimum distance estimator and an aligned rank test for an intercept were proposed. The (asymptotic) bias of R-estimator in measurement error model was also investigated. Besides measurement errors the problem of heteroscedastic model errors was considered – regression rank score tests of heteroscedasticity with nuisance regression and tests of regression with nuisance heteroscedasticity were proposed. Finally, in location model tests and estimates of shift parameter for various measurement errors were studied. All the results were derived theoretically and then demonstrated numerically with examples or simulations.

Keywords: measurement error models, rank tests, R-estimates, regression rank scores

Contents

Introduction

Measurement error models (also called *errors-in-variables models*) are regression models that account for measurement errors in the independent variables (regressors). These observed regressors are called *manifest*, *indicator*, or *proxy* variables, while the original unobserved regressors are called *latent* or *true* variables. They may be regarded either as unknown constants (in which case the model is called a *functional model*), or as random variables (correspondingly a *structural model*).

These models occur very commonly in practical data analysis, where some variables cannot be observed exactly, usually due to instrument or sampling error. Sometimes ignoring measurement error may lead to correct conclusions, however in some situations it may have dramatic consequences.

Let us mention some motivational examples to illustrate where one may meet measurement error models and what problems one has to solve when dealing with such models. This first example is from Nummi and Möttönen (2004).

Example. *Finland is Europe's most heavily-forested country, they cover 74.2% of the land area. Forests have been Finland's most important natural resource for centuries. In the harvesting technique used in Finland, tree stems are converted into smaller logs (usually 2–3) immediately at harvest.*

One of the tasks harvesters have to deal with is to find optimal cutting points on the stems. They are given data from high class measuring and computing equipment - length and diameter data from a sensor. In 1980s techniques for the situation if the measurements of the entire stem are known were developed. However, in a real harvesting situation it is not possible to measure the whole stem before crosscutting. In practice the first cutting point must be made under incomplete information about the stem. Therefore some prediction or estimation method is needed for the unknown part of the stem. In addition both stem diameter and stem height are measured only with measurement errors.

Example (Misclassification)**.** *Fleiss (1981) conducted a study to estimate the proportion of heavy smokers in a population. 200 respondents were randomly selected and asked if they would classify themselves as a heavy smoker, 88 of them did. However, such a classification was based only on a subjective opinion of a respondent. Hence a random subsample of 50 subjects was selected and the individuals were subjected to a blood test that is able to determine the true status (heavy smoker or not).*

We know for 50 observations both the true and reported status, but for the rest of 150 observations only the reported. Why do we not know it for all the observations? Since the blood test is very expensive and the testing procedure might be time demanding. Anyway, we would like to combine both types of the information to estimate the proportion of heavy smokers in a population.

Example (Simple linear regression)**.** *DeGracie and Fuller (1972) examined the relationship between corn yield (response variable) and soil nitrogen content (regressor) measured on eleven sites in Marshall County, Iowa. The yields are assumed to be measured without error, but the nitrogen content associated with the* *spatial area that comprises a unit is measured with error caused by spatial subsampling and instrument error in assessing nitrogen content of the soil. Measurement errors are assumed to be additive with a constant known variance 57.*

We still want to describe the relation between corn yield and soil nitrogen content with a simple linear regression model, although the regressors are measured with an additive error. We want to estimate the slope parameter and test if it is zero (if there is a linear relation between corn yield and soil nitrogen content).

Importance of measurement error models was first considered by Adcock at the end of the nineteenth century. In Adcock (1877) and Adcock (1878) he showed that in the model of regression line with measurement error least squares estimate of the slope parameter is downwarded in magnitude. One year later Kummel (1879) generalized his work, unlike Adcock that considered model and measurement errors to be equal, Kummel assumed their ratio to be known. Pearson (1901) extended previous results into multiple linear regression model. Deming (1931) and Lindley (1947) used two different methods for minimizing projections other than orthogonal.

Wald (1940) suggested completely different approach. He did not need any knowledge of the error structure, he split the observations into two groups, where the first group contained the first half of ordered observations and the second group the rest of the observations. Bartlett (1949) further developed this idea by adding one more group. Theoretically the splitting should be based on true values that are unobservable. However, Wald showed that approximate splitting with respect to the observed values gives the same results as splitting with respect to the true values. This method is in the literature known as grouping and was further developed and generalized. Pakes (1982) criticized this approach and showed that Wald's estimate underestimates the true value of the parameter. On the other hand, if the measurement errors are not too large, this grouping method may provide a reasonable estimator of the slope parameter. Wald's grouping method may be considered as a special case of instrumental value method that is very popular in econometrics.

Theil (1950) and Durbin (1954) were probably the first who used ranks of observed values for constructing an estimate. Anyway, there is the same problem as for grouping method, because the ranks of the observed values may differ from the real ones and the estimate will be inconsistent. Teissier (1948) suggested a method of geometric mean, for implementing this estimate it was necessary to have some knowledge about the measurement errors, moreover this estimate may have in some situations infinite variance. An interesting way how to look at geometric mean regression was done by Barker et al. (1988). He considered least triangles method, where minimized the sum of right angled triangles similarly as in the least squares method. Finally he proved a connection between his method and method of geometric mean.

Some authors tried to implement the method of moments into measurement error models, see for instance Scott (1950), Drion (1951). Presence of measurement errors will increase the number of equations needed for estimation and hence higher moments need to be considered. An approach closely related to the method of moments was proposed by Geary (1942) using so called cumulants.

The most popular method for dealing with measurement errors is a maximum likelihood approach, the normality assumption or any other knowledge of the error distribution is crucial for this method. Lindley (1947) was one of the first authors that used maximum likelihood method and pointed out that it is necessary to have some prior information about measurement error to be able to construct a consistent estimator.

Recently, total least squares method has became very popular. Originally it was introduced by Golub and Van Loan (1980) and has grown rapidly in many areas such as engineering and computed science.

There are several books devoted entirely to measurement error models, let us mention among others Fuller (1987), Cheng and Van Ness (1999), Carroll et al. (2006), Buonaccorsi (2010). The literature about measurement error models is very rich, we could not cover all the relevant paper, we just wanted to give an overview about the history and various methods that can be used for dealing with measurement error models.

The majority of the papers about measurement error model deals with the estimation problem. There is almost no mention of testing in these models in the literature, although this problem might be as important as estimation and sometimes we are more interested in identification of regressors that have influence on the response variable rather than the actual magnitude of regression coefficients. Hence our main aim will be an introduction of testing procedures in measurement error models. The most 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. Fuller (1987)). We will avoid this and introduce a class of rank tests that will be valid even if measurement errors are present.

Rank tests form a class of statistical procedures which have the advantage of simplicity combined with surprising power. The first application of rank tests mentioned in the literature may have been done by Arbuthnot (1710) who first used the sign test. Modern development of rank tests began in the 1930's. Let us mention articles of Hotteling and Pabst (1936), Friedman (1937) and Kendall (1938). Well known is also Wilcoxon (1945) who introduced popular Wilcoxon test for comparing two treatments. At first, it was believed that we have to pay a heavy price in loss of efficiency when using rank tests. However, it turned out that efficiency of rank tests behaves quite well under the classical assumption of normality. In addition these tests remain valid and have high efficiency when the assumption of normality is not satisfied. These facts were first brought out by Pitman (1948).

Several books about rank tests were written, among others Hájek and Šidák (1967), Puri and Sen (1971), Lehmann (1975) and Puri and Sen (1985). For location and shift parameters the idea of using rank tests to derive some estimators was done by Hodges and Lehmann (1963), this concept was later generalized by Jurečková (1971) into linear regression model. Gutenbrunner and Jurečková (1992) and Gutenbrunner et al. (1993) introduced ranks into regression models as so called regression rank scores and tests and estimates based on them.

The idea to use rank tests and estimates in measurement error models is relatively new. Only a few years ago Jurečková et al. (2010) first used rank tests and regression rank score tests for testing about regression parameters when both response and regressors are subject to measurement errors. Sen and Saleh (2010) and Saleh et al. (2012) considered some R-estimates of regression parameters under measurement errors.

Main goal of this thesis is to investigate the behavior of standard rank procedures in measurement error models - both tests and estimates. We will also try to modify these procedures to stay valid in measurement error models. This thesis is like a cookbook proposing rank tests and estimates for various models affected by measurement errors.

In the first chapter we summarize existing results about rank tests and R-estimates in linear regression model under measurement errors, especially based on articles Jurečková et al. (2010) and Saleh et al. (2012). These results will be then generalized in Chapter 2, where we propose a new rank test for slope parameter in regression and an aligned rank test for intercept, where the nuisance slope parameter is replaced with its estimate. However, these rank estimates are biased as it is shown in Chapter 3; their asymptotic bias corresponds to that for least square estimate in normal measurement error model.

In Chapter 4 the problem of heteroscedasticity in linear models is discussed. Regression rank score tests of homoscedasticity with nuisance regression and tests for regression under nuisance heteroscedasticity are proposed. Chapters 5 and 6 deal with location model. Rank tests and R-estimates of the shift parameter are investigated under various measurement errors.

All the theoretical results are accompanied by numerical examples, where the performance of tests and estimates for finite sample situation is studied.

1. Rank tests in regression

This chapter is an overview of rank-based methods in the linear model, with or without presence of measurement errors. The main results are cited without proofs, the illustrating examples and simulations are original.

1.1 Tests about slope parameter

1.1.1 Model without measurement errors

Consider classical linear regression model

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

where $\beta_0 \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^p$ are unknown parameters, model errors e_i are assumed to be independent identically distributed (i.i.d.) with an unknown distribution function F and density f, $\mathbf{x}_i$ are vectors of known regressors, such that

$$\mathbf{Q}_n = \frac{1}{n}\sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\top, \quad \text{with} \quad \overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$$

is a positive definite matrix (further on we will tacitly assume that this holds). Our aim is to test the hypothesis

$$\mathbf{H}_0 : \ \boldsymbol{\beta} = \mathbf{0} \quad \text{against} \quad \mathbf{K}_0 : \ \boldsymbol{\beta} \neq \mathbf{0}.$$

Choose a nondecreasing, nonconstant, square integrable score function $\varphi : (0,1) \mapsto \mathbb{R}$ and define

$$a_n(i) = \varphi\left(\frac{i}{n+1}\right), \quad i = 1, \ldots, n, \tag{1.2}$$

$$A^2(\varphi) = \int_0^1 (\varphi(t) - \overline{\varphi})^2 dt, \quad \overline{\varphi} = \int_0^1 \varphi(t) dt. \tag{1.3}$$

Remark. *The scores $a_n(i)$ are called approximate scores, instead of them one may use so-called exact scores*

$$\widetilde{a}_n(i) = \mathbb{E}\varphi\left(U_{(i)}\right), \quad i = 1, \ldots, n,$$

where $U_{(1)} \leq \ldots \leq U_{(n)}$ are the ordered statistics from a sample of size n from uniform $\mathcal{U}(0,1)$ distribution.

Let R_i be the rank of Y_i among $Y_1, \ldots, Y_n$ and define vector of linear rank statistics

$$\mathbf{S}_n = n^{-1/2}\sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}}) a_n(R_i).$$

Test criterion for $\mathbf{H}_0$ is

$$T_n^2 = A^{-2}(\varphi)\mathbf{S}_n^\top \mathbf{Q}_n^{-1} \mathbf{S}_n. \tag{1.4}$$

Lemma 1.1. *Let $Y_1, \dots, Y_n$ be i.i.d. with absolutely continuous distribution and ranks $R_1, \dots, R_n$. Then for each permutation $(r_1, \dots, r_n)$ of numbers $(1, \dots, n)$ it holds*

$$P(R_1 = r_1, \dots, R_n = r_n) = \frac{1}{n!}.$$

Proof. See Hájek et al. (1999). □

The previous lemma gives us a clue how to find critical value for test based on T_n^2. Under $\mathbf{H}_0$ model (1.1) reduces to

$$Y_i = \beta_0 + e_i, \quad i = 1, \dots, n, \tag{1.5}$$

hence we may compute for all permutations $(r_1, \dots, r_n)$ values of T_n^2 and order these $n!$ values in the increasing magnitude. The critical region is then formed by $k = \lfloor \alpha n! \rfloor$ largest values of T_n^2 (α being prescribed level of significance). (k+1)-st largest value might be randomized to achieve exactly level α.

However, this approach becomes computationally demanding for large values of n, hence we shall use asymptotic approximation. For that we will need to add some assumptions on model errors and regressors. Assume that f has finite Fisher information with respect to the location

$$0 < I(f) = \int \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx < \infty \tag{1.6}$$

and there exists a positive definite matrix $\mathbf{Q}$, such that as $n \to \infty$

$$\mathbf{Q}_n \to \mathbf{Q}, \tag{1.7}$$

$$\frac{1}{n} \max_{i=1,\dots,n} (\mathbf{x}_i - \overline{\mathbf{x}})^\top \mathbf{Q}_n^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}) \to 0. \tag{1.8}$$

Theorem 1.1. *Assume that (1.6) – (1.8) hold. Then in model (1.1) under $\mathbf{H}_0$ test statistic T_n^2 has asymptotically as $n \to \infty$ χ^2 distribution with p degrees of freedom.*

Proof. See Puri and Sen (1985, Theorem 5.3.1). □

We are able to describe asymptotic behavior of T_n^2 under alternatives close to the hypothesis. Hence consider a sequence of local alternatives

$$\mathbf{K}_{0,n} : \ \boldsymbol{\beta} = n^{-1/2} \boldsymbol{\beta}^*, \quad \mathbf{0} \neq \boldsymbol{\beta}^* \in \mathbb{R}^p \text{ fixed.} \tag{1.9}$$

Theorem 1.2. *Assume that (1.6) – (1.8) hold. Then in model (1.1) under $\mathbf{K}_{0,n}$ test statistic T_n^2 has asymptotically as $n \to \infty$ noncentral χ^2 distribution with p degrees of freedom and noncentrality parameter*

$$\eta^2 = \boldsymbol{\beta}^{*\top} \mathbf{Q} \boldsymbol{\beta}^* \frac{\gamma^2(\varphi, f)}{A^2(\varphi)},$$

$$\gamma(\varphi, f) = \int_0^1 \varphi(t) \widetilde{\varphi}(t, f) dt, \quad \widetilde{\varphi}(t, f) = -\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}. \tag{1.10}$$

Proof. See Puri and Sen (1985, Theorem 5.5.2). □

1.1.2 Model with errors in responses

Suppose that we observe regressors $\mathbf{x}_i$ accurately, while instead of Y_i we observe $Z_i = Y_i + W_i$, where W_i are i.i.d. random variables independent with e_i with (unknown) distribution function G and density g. We have

$$\begin{aligned} Y_i &= \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + e_i, \\ Z_i &= Y_i + W_i, \quad i = 1, \dots, n. \end{aligned} \tag{1.11}$$

We may rewrite model (1.11) for observed responses as

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

where $e_i^* = e_i + W_i$ are i.i.d. random variables with density

$$h(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy \tag{1.12}$$

and finite Fisher information as it will be proven in the following lemma.

Lemma 1.2. *Let X, Y be independent absolutely continuous random variables, X with finite Fisher information with respect to the location. Then $X + Y$ has finite Fisher information with respect to the location.*

Proof. According to Hölder's inequality

$$\int_{-\infty}^{\infty} |f'(x)|dx \leq \left\{ \int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)} \right)^2 f(x)dx \right\}^{1/2} = [I(f)]^{1/2} < \infty.$$

Hence

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f'(x-y)|dxg(y)dy < \infty$$

and

$$\begin{aligned} h(x) &= \int_{-\infty}^{\infty} f(x-y)g(y)dy \\ h'(x) &= \int_{-\infty}^{\infty} f'(x-y)g(y)dy \end{aligned}$$

Finally, by Cauchy-Schwarz inequality

$$\frac{[h'(x)]^2}{h(x)} \leq \int_{-\infty}^{\infty} \frac{[f'(x-y)]^2}{f(x-y)} g(y)dy$$

and integrating both sides dx we get $I(h) \leq I(f) < \infty$. □

Denote $T_{Z,n}^2$ test statistic (1.4) based on observed values $(Z_i, \mathbf{x}_i)$. The presence of measurement errors may change ranks R_i, fortunately this change does not affect hypothetical distribution of $T_{Z,n}^2$, because under $\mathbf{H}_0$ we have

$$Z_i = \beta_0 + e_i^*$$

and according to Lemma 1.1 distribution of $(R_1, \ldots, R_n)$ remains uniform over all permutations. Hence the exact distribution of $T^2_{Z,n}$ under $\mathbf{H}_0$ is the same as T^2_n. Similarly, as a direct consequence of (1.12) and Lemma 1.2, for asymptotic distribution of $T^2_{Z,n}$ we have the following theorem.

Theorem 1.3. *Assume that (1.6) – (1.8) hold. Then in model (1.11) under* $\mathbf{H}_0$ *test statistic* $T^2_{Z,n}$ *has asymptotically as* $n \to \infty$ χ^2 *distribution with* p *degrees of freedom and under* $\mathbf{K}_{0,n}$ *asymptotically noncentral* χ^2 *distribution with* p *degrees of freedom and noncentrality parameter*

$$\eta^2 = \boldsymbol{\beta}^{*\top}\mathbf{Q}\boldsymbol{\beta}^* \frac{\gamma^2(\varphi, h)}{A^2(\varphi)}.$$

Remark (Asymptotic relative efficiency)**.** *Complete information on the limiting properties of tests is provided by their asymptotic power. If all considered statistics have asymptotically* χ^2 *distribution under both hypothesis and alternative, a number called asymptotic relative efficiency independent of the level of significance may be defined for comparing powers of two various tests.*

Definition. *Let* T_1 *and* T_2 *be two tests for* $\boldsymbol{H}$ *against* $\boldsymbol{K}$ *that are asymptotically* χ^2 *distributed with* p *degrees of freedom under* $\boldsymbol{H}$ *and under* $\boldsymbol{K}$ *asymptotically* χ^2 *distributed with* p *degrees of freedom and noncentrality parameters* δ_1^2, δ_2^2 *respectively. Then the number*

$$ARE(T_1, T_2) = \frac{\delta_1^2}{\delta_2^2}$$

will be called asymptotic relative efficiency of T_1*-test relative to* T_2*-test.*

Remark. *The number* $\lfloor (1 - ARE(T_1, T_2)) \cdot n \rfloor$ *can be interpreted as a number of observations "wasted" when using* T_1*-test instead of* T_2*-test (Pitman's interpretation of ARE). In other words, the number* $\lfloor ARE(T_1, T_2) \cdot n \rfloor$ *gives us a number of observations needed for reaching asymptotically the same power when using* T_2*-test instead of* T_1*.*

In our situation asymptotic relative efficiency of the test with measurement errors relative to the test without measurement errors is

$$\mathrm{ARE}(T^2_{Z,n}, T^2_n) = \left(\frac{\gamma(\varphi, h)}{\gamma(\varphi, f)}\right)^2.$$

1.1.3 Model with errors in regressors

This model assumes that regressors $\mathbf{x}_i$ are not observed accurately, but only with an additive, unobservable, error $\mathbf{v}_i$, i.e. we observe $\mathbf{w}_i = \mathbf{x}_i + \mathbf{v}_i$ instead of $\mathbf{x}_i$, where $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are i.i.d. random vectors independent of $\mathrm{e}_1, \ldots, e_n$. In other words, we may write

$$\begin{aligned} Y_i &= \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + e_i, \\ \mathbf{w}_i &= \mathbf{x}_i + \mathbf{v}_i, \quad i = 1, \ldots, n. \end{aligned} \tag{1.13}$$

Assume that observed regressors $\mathbf{w}_i$ satisfy (similarly as in model without measurement errors):

$$\mathbf{D}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}_i - \overline{\mathbf{w}})(\mathbf{w}_i - \overline{\mathbf{w}})^\top, \quad \text{with} \quad \overline{\mathbf{w}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i$$

is a positive definite matrix.

Denote $T^2_{w,n}$ test statistic (1.4) based on observed values $(Y_i, \mathbf{w}_i)$. The presence of measurement errors does not change ranks R_i, hence nor hypothetical distribution of $T^2_{w,n}$, because under $\mathbf{H}_0$ we have

$$Z_i = \beta_0 + e_i.$$

For deriving asymptotic properties we need to put more assumptions on $\mathbf{w}_i$. Assume that there exists a positive definite matrix $\mathbf{C}$, such that as $n \to \infty$

$$\mathbf{C}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}_i - \overline{\mathbf{v}})(\mathbf{v}_i - \overline{\mathbf{v}})^\top \to \mathbf{C}, \tag{1.14}$$

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}_i - \overline{\mathbf{v}})(\mathbf{x}_i - \overline{\mathbf{x}})^\top \to \mathbf{0}, \tag{1.15}$$

$$\frac{1}{n}\max_{i=1,\ldots,n}(\mathbf{w}_i - \overline{\mathbf{w}})^\top \mathbf{D}_n^{-1}(\mathbf{w}_i - \overline{\mathbf{w}}) \to 0. \tag{1.16}$$

Then we state a theorem about asymptotic properties of $T^2_{w,n}$.

Theorem 1.4. *Assume that (1.6), (1.14) – (1.16) hold. Then in model (1.13) under* $\mathbf{H}_0$ *test statistic* $T^2_{w,n}$ *has asymptotically as* $n \to \infty$ χ^2 *distribution with* p *degrees of freedom and under* $\mathbf{K}_{0,n}$ *has asymptotically noncentral* χ^2 *distribution with* p *degrees of freedom and noncentrality parameter*

$$\eta^2 = \boldsymbol{\beta}^{*\top}\mathbf{Q}(\mathbf{Q}+\mathbf{C})^{-1}\mathbf{Q}\boldsymbol{\beta}^*\frac{\gamma^2(\varphi, f)}{A^2(\varphi)}.$$

Proof. See Jurečková et al. (2009). □

Hence we have the expression for asymptotic relative efficiency of the test with measurement errors relative to the test without measurement errors:

$$\text{ARE}(T^2_{w,n}, T^2_n) = \frac{\boldsymbol{\beta}^{*\top}\mathbf{Q}(\mathbf{Q}+\mathbf{C})^{-1}\mathbf{Q}\boldsymbol{\beta}^{*\top}}{\boldsymbol{\beta}^{*\top}\mathbf{Q}\boldsymbol{\beta}^{*\top}}.$$

Remark. *Numerical illustration of the performance of the above tests may be found in Navrátil (2010).*

1.1.4 Possible complications due to measurement errors

In model (1.1) without measurement errors it is easy to see that when we want to test the hypothesis $\mathbf{H}_{0,*} : \boldsymbol{\beta} = \boldsymbol{\beta}^0 \neq \mathbf{0}, \boldsymbol{\beta}^0 \in \mathbb{R}^p$ known, we may transform this problem into problem of testing $\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ subtracting $\mathbf{x}_i^\top \boldsymbol{\beta}^0$ from both sides of (1.1):

$$Y_i^* = Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^0 = \beta_0 + e_i. \tag{1.17}$$

Testing $\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ in model (1.1) is then equivalent to testing $\mathbf{H}_{0,*} : \boldsymbol{\beta} = \boldsymbol{\beta}^0$ in model (1.17).

Will this approach work also in measurement error models (1.11) and (1.13)? For model (1.11) it is evident that it will work, but in model (1.13) it does not work. Using the same technique, i.e. subtracting $\mathbf{w}_i^\top \boldsymbol{\beta}^0$ from both sides of (1.13), we get

$$Y_i^* = Y_i - \mathbf{w}_i^\top \boldsymbol{\beta}^0 = \beta_0 - \mathbf{v}_i^\top \boldsymbol{\beta}^0 + e_i. \tag{1.18}$$

Unlike the previous case we did not get rid of $\boldsymbol{\beta}^0$ from the right side of equation (1.18) and the test will not work. We may illustrate it with the following simulation example.

Example. *Consider model of regression line passing through the origin*

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

The regressors x_i were once generated from independent sample of size $n = 50$ from uniform $\mathcal{U}(-2, 10)$ distribution and then taken as fixed design points, 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 $\mathbf{H}_{0,*} : \beta = 2$ among 10 000 replications, at significance level $\alpha = 0.05$. The results are summarized in Table 1.1.

$\beta \setminus v_i$	0	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1/2)$	$\mathcal{U}(-1,1)$	$\mathcal{U}(-0.5,0.5)$	$\mathcal{U}(-2,2)$
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 1.1: Percentage of rejections of hypothesis $\mathbf{H}_{0,*} : \beta = 2$ for various measurement errors v_i for Wilcoxon test for regression; $n = 50$.

To be able to deal with this problem one would need some additional knowledge about distribution of measurement errors.

1.2 Aligned rank tests

We are often more interested in testing hypothesis only about a component of the parameter $\boldsymbol{\beta}$, identify regressors that have influence on response variable. Denote

$$\begin{aligned}\boldsymbol{\beta} &= (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top,\\ \mathbf{x}_i &= (\mathbf{x}_{1,i}^\top, \mathbf{x}_{2,i}^\top)^\top,\\ \mathbf{v}_i &= (\mathbf{v}_{1,i}^\top, \mathbf{v}_{2,i}^\top)^\top,\\ \mathbf{w}_i &= (\mathbf{w}_{1,i}^\top, \mathbf{w}_{2,i}^\top)^\top,\end{aligned}$$

where $\boldsymbol{\beta}_1 \in \mathbb{R}^{p-q}$, $\boldsymbol{\beta}_2 \in \mathbb{R}^q$, $\mathbf{x}_{1,i} \in \mathbb{R}^{p-q}$, $\mathbf{x}_{2,i} \in \mathbb{R}^q$, $\mathbf{v}_{1,i} \in \mathbb{R}^{p-q}$, $\mathbf{v}_{2,i} \in \mathbb{R}^q$, $\mathbf{w}_{1,i} \in \mathbb{R}^{p-q}$, $\mathbf{w}_{2,i} \in \mathbb{R}^q$, $1 \le q < p$. Then model (1.13) can be rewritten as

$$\begin{aligned}Y_i &= \beta_0 + \mathbf{x}_{1,i}^\top\boldsymbol{\beta}_1 + \mathbf{x}_{2,i}^\top\boldsymbol{\beta}_2 + e_i,\\ \mathbf{w}_{1,i} &= \mathbf{x}_{1,i} + \mathbf{v}_{1,i},\\ \mathbf{w}_{2,i} &= \mathbf{x}_{2,i} + \mathbf{v}_{2,i}, \quad i = 1, \dots, n.\end{aligned} \tag{1.19}$$

Our goal is to test the hypothesis

$$\mathbf{H}_1 : \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{against} \quad \mathbf{K}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0},$$

considering β_0 and $\boldsymbol{\beta}_1$ as nuisance parameters.

Rank tests are invariant with respect to the location, but not to the nuisance regression. That is why we have to first estimate the nuisance parameter $\boldsymbol{\beta}_1$ and then apply the standard test on residuals. Due to the absence of knowledge of distribution of model errors and to preserve robust properties we use an R-estimator of parameter $\boldsymbol{\beta}_1$.

Model (1.19) under $\mathbf{H}_1$ reduces to

$$Y_i = \beta_0 + \mathbf{w}_{1,i}^\top\boldsymbol{\beta}_1 + e_i^*,$$

where $e_i^* = e_i - \mathbf{v}_{1,i}^\top\boldsymbol{\beta}_1$ are i.i.d. random variables with density $f^* = f^*_{\beta_1}$.

Choose a nondecreasing, nonconstant, square integrable score function $\psi : (0,1) \mapsto \mathbb{R}$ that is skew-symmetric, i.e.

$$\psi(1-t) = -\psi(t), \quad \forall\, 0 < t < 1$$

and define as in (1.2)

$$\widetilde{a}_n(i) = \psi\left(\frac{i}{n+1}\right), \quad i = 1, \dots, n.$$

Following Jaeckel (1972) we define the rank (pseudo)estimator $\widehat{\boldsymbol{\beta}}_{1,n}$ of $\boldsymbol{\beta}_1$ as a minimizer of

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^n \left(Y_i - \mathbf{w}_{1,i}^\top\mathbf{b}\right) \widetilde{a}_n(R_i(\mathbf{b})) \tag{1.20}$$

with respect to $\mathbf{b} \in \mathbb{R}^{p-q}$, where $R_i(\mathbf{b})$ is the rank of $(Y_i - \mathbf{w}_{1,i}^\top\mathbf{b})$ among $(Y_1 - \mathbf{w}_{1,1}^\top\mathbf{b}), \dots, (Y_n - \mathbf{w}_{1,n}^\top\mathbf{b})$.

Remark. *By Jaeckel (1972), the function $\mathcal{D}_n(\mathbf{b})$ is convex and piecewise linear in $\mathbf{b} \in \mathbb{R}^{p-q}$, with the subgradient*

$$-n^{1/2}\mathbf{S}_n(\mathbf{b}) = -\sum_{i=1}^{n}(\mathbf{w}_{1,i} - \overline{\mathbf{w}}_1)\widetilde{a}_n(R_i(\mathbf{b})),$$

hence its minimizer exists.

Jurečková (1969) and Koul (1969) defined an estimate $\widehat{\boldsymbol{\beta}}_{1,n}$ of $\boldsymbol{\beta}_1$ in a different way, as

$$\widehat{\boldsymbol{\beta}}_{1,n} = \mathit{argmin}\{\|\mathbf{S}_n(\mathbf{b})\|, \mathbf{b} \in \mathbb{R}^{p-q}\},$$

where $\|\cdot\|$ might be L^1 or L^2 norm in $\mathbb{R}^{p-q}$. However, all these estimates admit the same asymptotic representation and hence they have under very mild assumptions the same asymptotic distribution.

Now, consider residuals

$$\widehat{e}_i = Y_i - \mathbf{w}_{1,i}^\top \widehat{\boldsymbol{\beta}}_{1,n}, \quad i = 1, \dots, n$$

and proceed the same way as in Section 1.1.1, when applying the test on residuals $\widehat{e}_1, \dots, \widehat{e}_n$. Note that unlike the situation in Section 1.1.1 residuals $\widehat{e}_i$ are not independent, because they depend on the R-estimate of nuisance parameter $\boldsymbol{\beta}_1$. However under some assumptions this fact does not affect the asymptotic distribution.

Hence choose a nondecreasing, nonconstant, square integrable score function $\varphi : (0,1) \mapsto \mathbb{R}$ (it may differ from ψ) and define

$$a_n(i) = \varphi\left(\frac{i}{n+1}\right), \quad i = 1, \dots, n$$

and compute

$$\widehat{\mathbf{S}}_n = n^{-1/2}\sum_{i=1}^{n}(\mathbf{w}_{2,i} - \overline{\mathbf{w}}_2)a_n(R_i(\widehat{\boldsymbol{\beta}}_{1,n})),$$

where $R_i(\widehat{\boldsymbol{\beta}}_{1,n})$ is the rank of $\widehat{e}_i$ among $\widehat{e}_1, \dots, \widehat{e}_n$.

Denote

$$\mathbf{D}_{1,n} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}_{1,i} - \overline{\mathbf{w}}_1)(\mathbf{w}_{1,i} - \overline{\mathbf{w}}_1)^\top,$$

$$\mathbf{D}_{2,n} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}_{2,i} - \overline{\mathbf{w}}_2)(\mathbf{w}_{2,i} - \overline{\mathbf{w}}_2)^\top.$$

Finally, consider test statistic

$$\widehat{T}_n^2 = A^{-2}(\varphi)\widehat{\mathbf{S}}_n^\top \mathbf{D}_{2,n}^{-1}\widehat{\mathbf{S}}_n.$$

Assume that there exist positive definite matrices $\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{C}_1, \mathbf{C}_2$, such that as $n \to \infty$

$$\mathbf{Q}_{1,n} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_{1,i} - \overline{\mathbf{x}}_1)(\mathbf{x}_{1,i} - \overline{\mathbf{x}}_1)^\top \to \mathbf{Q}_1, \tag{1.21}$$

$$\mathbf{C}_{1,n} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}_{1,i} - \overline{\mathbf{v}}_1)(\mathbf{v}_{1,i} - \overline{\mathbf{v}}_1)^\top \to \mathbf{C}_1, \tag{1.22}$$

$$\frac{1}{n}\max_{i=1,\ldots,n}(\mathbf{w}_{1,i} - \overline{\mathbf{w}}_1)^\top \mathbf{D}_{1,n}^{-1}(\mathbf{w}_{1,i} - \overline{\mathbf{w}}_1) \to 0, \tag{1.23}$$

$$\mathbf{Q}_{2,n} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_{2,i} - \overline{\mathbf{x}}_2)(\mathbf{x}_{2,i} - \overline{\mathbf{x}}_2)^\top \to \mathbf{Q}_2, \tag{1.24}$$

$$\mathbf{C}_{2,n} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}_{2,i} - \overline{\mathbf{v}}_2)(\mathbf{v}_{2,i} - \overline{\mathbf{v}}_2)^\top \to \mathbf{C}_2, \tag{1.25}$$

$$\frac{1}{n}\max_{i=1,\ldots,n}(\mathbf{w}_{2,i} - \overline{\mathbf{w}}_2)^\top \mathbf{D}_{2,n}^{-1}(\mathbf{w}_{2,i} - \overline{\mathbf{w}}_2) \to 0. \tag{1.26}$$

Finally, we are able to describe asymptotic distribution of $\widehat{T}_n^2$.

Theorem 1.5. *Assume that (1.6), (1.21) – (1.26) hold. Then in model (1.19) under* $\mathbf{H}_1$ *test statistic* $\widehat{T}_n^2$ *has asymptotically as* $n \to \infty$ χ^2 *distribution with* q *degrees of freedom and under local alternative*

$$\mathbf{K}_{1,n}: \ \boldsymbol{\beta}_2 = n^{-1/2}\boldsymbol{\beta}_2^*, \quad \mathbf{0} \neq \boldsymbol{\beta}_2^* \in \mathbb{R}^q \text{ fixed}$$

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

$$\widehat{\eta}^2 = \boldsymbol{\beta}_2^{*\top}\mathbf{Q}_2(\mathbf{Q}_2 + \mathbf{C}_2)^{-1}\mathbf{Q}_2\boldsymbol{\beta}_2^*\frac{\gamma^2(\varphi, f^*)}{A^2(\varphi)}.$$

Proof. See Jurečková et al. (2010). □

Remark. *Recall that* f^* *depends on unknown nuisance parameter* $\boldsymbol{\beta}_1$ *and distribution of measurement errors* $v_{1,i}$*, hence the asymptotic power of the test does depend on the nuisance parameter* $\boldsymbol{\beta}_1$ *unlike the situation without measurement errors.*

Remark. *Although it may seem at first glance that the previous procedure could be used for estimation of regression parameters in measurement error models, but it could not. As pointed out in Saleh et al. (2012) estimate* $\widehat{\boldsymbol{\beta}}_{1,n}$ *is not consistent estimator of* $\boldsymbol{\beta}_1$*, but in fact it estimates* $(\mathbf{Q}_1 + \mathbf{C}_1)^{-1}\mathbf{Q}_1\boldsymbol{\beta}_1$*.*

For the testing procedure it does not matter and when considering residuals this "inconsistency" disappears because of multiplying $\widehat{\boldsymbol{\beta}}_{1,n}$ *by* $\mathbf{w}_{1,i}$*.*

1.3 Regression rank score tests

Under certain assumptions on measurement errors Jurečková et al. (2010) solved the previous problem without an estimation of nuisance parameter $\boldsymbol{\beta}_1$ with the aid of regression rank scores. They are invariant to the nuisance regression and hence we can avoid estimation of $\boldsymbol{\beta}_1$ and use directly test based on them. To be able to do it we have to assume that corresponding regressors $\mathbf{x}_{1,i}$ are measured exactly, while $\mathbf{x}_{2,i}$ may be affected by some measurement errors:

$$\begin{aligned} Y_i &= \beta_0 + \mathbf{x}_{1,i}^\top \boldsymbol{\beta}_1 + \mathbf{x}_{2,i}^\top \boldsymbol{\beta}_2 + e_i, \\ \mathbf{w}_{2,i} &= \mathbf{x}_{2,i} + \mathbf{v}_{2,i}, \quad i = 1, \ldots, n. \end{aligned} \tag{1.27}$$

Under $\mathbf{H}_1$ it reduces to

$$Y_i = \beta_0 + \mathbf{x}_{1,i}^\top \boldsymbol{\beta}_1 + e_i. \tag{1.28}$$

Hence we compute regression rank scores in model (1.28). Regression rank scores were introduced in Gutenbrunner and Jurečková (1992) as a solution of the dual form of the linear program required for computing regression quantiles.

Remark. *Koenker and Basset (1978) introduced regression quantiles in model (1.28), α - regression quantile is any solution of the minimization*

$$\min \sum_{i=1}^{n} \rho_\alpha (Y_i - t_0 - \mathbf{x}_{1,i}^\top \mathbf{t})$$

with respect to $(t_0, \mathbf{t}) \in \mathbb{R} \times \mathbb{R}^{p-q}$, where

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

Koenker and Basset (1978) also showed that α - regression quantile $\widehat{\boldsymbol{\beta}}_n(\alpha)$ can be computed as a component $\boldsymbol{\beta}$ of optimal solution $(\boldsymbol{\beta}, \mathbf{r}^+, \mathbf{r}^-)$ of the linear program

$$\min \ \alpha \mathbf{1}_n^\top \mathbf{r}^+ + (1-\alpha) \mathbf{1}_n^\top \mathbf{r}^-$$

with respect to

$$\begin{aligned} &\mathbf{X}_{1,n}^* \boldsymbol{\beta} + \mathbf{r}^+ - \mathbf{r}^- = \mathbf{Y}_n \\ &(\boldsymbol{\beta}, \mathbf{r}^+, \mathbf{r}^-) \in \mathbb{R}^{p-q+1} \times \mathbb{R}_+^{2n}, \end{aligned} \tag{1.29}$$

where $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^\top, \mathbf{1}_n = (1, \ldots, 1)^\top \in \mathbb{R}^n$ and $\mathbf{X}_{1,n}^* = (\mathbf{1}_n, \mathbf{X}_{1,n})$ with $\mathbf{X}_{1,n} = (\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,n})^\top$.

Hence regression rank score $\widehat{\mathbf{a}}_n(\alpha) = (\widehat{a}_{n,1}(\alpha), \ldots, \widehat{a}_{n,n}(\alpha))^\top$ may be found as a solution of the maximization

$$\max \ \mathbf{Y}_n^\top \mathbf{a}$$

with respect to

$$\mathbf{X}_{1,n}^{*\top} \mathbf{a} = (1-\alpha) \mathbf{X}_{1,n}^{*\top} \mathbf{1}_n, \quad \mathbf{a} \in [0,1]^n.$$

Gutenbrunner et al. (1993) pointed out that $\widehat{\mathbf{a}}_n(\alpha)$ are regression invariant with respect to $\mathbf{X}_{1,n}^*$, i.e. $\widehat{\mathbf{a}}_n(\alpha)$ is unchanged if $\mathbf{Y}_n$ is transformed to $\mathbf{Y}_n + \mathbf{X}_{1,n}^*\gamma$ for all $\gamma \in \mathbb{R}^{p-q+1}$. This property is crucial for constructing the test statistic.

As in the classical rank tests theory, we shall consider a score function $\varphi : (0,1) \mapsto \mathbb{R}$ that is supposed to be nondecreasing, nonconstant, square integrable. And similarly define scores

$$\hat{b}_{n,i} = -\int_0^1 \varphi(t) d\widehat{a}_{n,i}(t), \quad \widehat{\mathbf{b}}_n = (\hat{b}_{n,1}, \dots, \hat{b}_{n,n})^\top,$$

vector of linear regression rank scores statistics and the test statistic, respectively:

$$\begin{aligned}
&\mathbf{S}_n = n^{-1/2}\sum_{i=1}^n (\mathbf{w}_{2,i} - \widehat{\mathbf{w}}_{2,i})\hat{b}_{n,i} = n^{-1/2}(\mathbf{W}_{2,n} - \widehat{\mathbf{W}}_{2,n})^\top \widehat{\mathbf{b}}_n,\\
&\mathbf{W}_{2,n} = (\mathbf{w}_{2,1}, \dots, \mathbf{w}_{2,n})^\top,\\
&\widehat{\mathbf{W}}_{2,n} = (\widehat{\mathbf{w}}_{2,1}, \dots, \widehat{\mathbf{w}}_{2,n})^\top = \mathbf{X}_{1,n}^*(\mathbf{X}_{1,n}^{*\top}\mathbf{X}_{1,n}^*)^{-1}\mathbf{X}_{1,n}^{*\top}\mathbf{W}_{2,n},\\
&\widehat{\mathbf{D}}_{2,n} = n^{-1}(\mathbf{W}_{2,n} - \widehat{\mathbf{W}}_{2,n})^\top(\mathbf{W}_{2,n} - \widehat{\mathbf{W}}_{2,n}),\\
&T_n^2 = A^{-2}(\varphi)\mathbf{S}_n^\top \widehat{\mathbf{D}}_{2,n}^{-1}\mathbf{S}_n.
\end{aligned}$$

Again, to prove some asymptotic properties we have to put some assumptions on distribution of model errors and observed regressors. Let us start with the imposed conditions on F:

(F.1) $f(x) > 0$ is absolutely continuous, bounded and monotonically decreasing as $x \downarrow A$ and $x \uparrow B$, where

$$-\infty \le A = \sup\{x : F(x) = 0\}, \qquad \infty \ge B = \inf\{x : F(x) = 1\},$$

and

$$\sup_{0<u<1} u(1-u)\frac{|f'(F^{-1}(u))|}{f^2(F^{-1}(u))} = \alpha, \tag{1.30}$$

where $1 \le \alpha \le 1 + \frac{1}{4} - \varepsilon,\ \varepsilon > 0$.

(F.2) $\left|\frac{f'(x)}{f(x)}\right| \le c|x|$, for $|x| \ge K \ge 0,\ c > 0$.

Remark. *Distributions satisfying (1.30) were studied by Csörgő and Révész (1978) and Parzen (1979), among others. The class covers the tail monotone distributions not lighter than normal and lighter than the t-distribution with 4 degrees of freedom. Parzen (1979) showed that then* $f(F^{-1}(u)) = (1-u)^\alpha L(1-u)$ *as* $u \uparrow 1$, *where the function* $L(\cdot)$ *is slowly varying at 0. Moreover, then also* $|F^{-1}(u)| \le c(1-u))^{1-\alpha}L(1-u)$ *for* $1-u_0 \le u < 1,\ c > 0$. *Numerical studies show that the tests work well even under more general conditions than those considered here. The conditions which are the most general for the asymptotic distribution of regression rank score tests is still an open question.*

Further denote

$$\mathbf{X}_{2,n} = (\mathbf{x}_{2,1}, \ldots, \mathbf{x}_{2,n})^\top, \quad \widehat{\mathbf{X}}_{2,n} = (\widehat{\mathbf{x}}_{2,1}, \ldots, \widehat{\mathbf{x}}_{2,n})^\top = \mathbf{X}^*_{1,n}(\mathbf{X}^{*\top}_{1,n}\mathbf{X}^*_{1,n})^{-1}\mathbf{X}^{*\top}_{1,n}\mathbf{X}_{2,n},$$

$$\mathbf{V}_{2,n} = (\mathbf{v}_{2,1}, \ldots, \mathbf{v}_{2,n})^\top, \quad \widehat{\mathbf{V}}_{2,n} = (\widehat{\mathbf{v}}_{2,1}, \ldots, \widehat{\mathbf{v}}_{2,n})^\top = \mathbf{X}^*_{1,n}(\mathbf{X}^{*\top}_{1,n}\mathbf{X}^*_{1,n})^{-1}\mathbf{X}^{*\top}_{1,n}\mathbf{V}_{2,n},$$

$$\widehat{\mathbf{Q}}_{2,n} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_{2,i} - \widehat{\mathbf{x}}_{2,i})(\mathbf{x}_{2,i} - \widehat{\mathbf{x}}_{2,i})^\top = \frac{1}{n}(\mathbf{X}_{2,n} - \widehat{\mathbf{X}}_{2,n})^\top(\mathbf{X}_{2,n} - \widehat{\mathbf{X}}_{2,n}),$$

$$\widehat{\mathbf{C}}_{2,n} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}_{2,i} - \widehat{\mathbf{v}}_{2,i})(\mathbf{v}_{2,i} - \widehat{\mathbf{v}}_{2,i})^\top = \frac{1}{n}(\mathbf{V}_{2,n} - \widehat{\mathbf{V}}_{2,n})^\top(\mathbf{V}_{2,n} - \widehat{\mathbf{V}}_{2,n})$$

and the corresponding conditions on regressors

(X1) There exists a positive definite matrix $\widehat{\mathbf{Q}}_2$ such that $\lim_{n\to\infty} \widehat{\mathbf{Q}}_{2,n} = \widehat{\mathbf{Q}}_2$.

(X2) There exists a positive definite matrix $\widehat{\mathbf{C}}_2$ such that $\lim_{n\to\infty} \widehat{\mathbf{C}}_{2,n} = \widehat{\mathbf{C}}_2$.

(X3) There exists a positive definite matrix $\mathbf{M}_1$ such that $\lim_{n\to\infty} n^{-1}\mathbf{X}^{*\top}_{1,n}\mathbf{X}^*_{1,n} = \mathbf{M}_1$.

(X4) $\max_{i=1,\ldots,n} \|x_{1,i}\| = O(1)$ and $n^{-1}\sum_{i=1}^n \|x_{1,i}\|^4 = O(1)$ as $n \to \infty$.

Theorem 1.6. *Assume that (F1) – (F2) and (X1) – (X4) hold. Then in model (1.27) under* $\mathbf{H}_1$ *test statistic* T_n^2 *has asymptotically as* $n \to \infty$ χ^2 *distribution with* q *degrees of freedom and under* $\mathbf{K}_{1,n}$ *asymptotically noncentral* χ^2 *distribution with* q *degrees of freedom and noncentrality parameter*

$$\eta^2 = \boldsymbol{\beta}_2^{*\top}\widehat{\mathbf{Q}}_2(\widehat{\mathbf{Q}}_2 + \widehat{\mathbf{C}}_2)^{-1}\widehat{\mathbf{Q}}_2\boldsymbol{\beta}_2^*\frac{\gamma^2(\varphi, f)}{A^2(\varphi)}.$$

Proof. See Jurečková et al. (2010). □

Remark. *In this situation the asymptotic distribution of the test statistic does not depend on the nuisance parameter* $\boldsymbol{\beta}_1$ *unlike the aligned rank test considered in the previous section.*

Jurečková et al. (2010) derived the previous test in model (1.27), where some regressors are measured accurately. What if they are not? What will happen with this test in model (1.19)?

Consider the model

$$Y_i = \beta_0 + x_{1,i}\beta_1 + x_{2,i}\beta_2 + e_i, \quad i = 1, \ldots, 30,$$

where $\beta_0, \beta_1, \beta_2 \in \mathbb{R}$ and test the hypothesis

$$\mathbf{H}_1 : \ \beta_2 = 0 \quad \text{against} \quad \mathbf{K}_1 : \ \beta_2 \neq 0.$$

Since the nuisance parameter β_1 is one-dimensional we will consider test statistic

$$\widetilde{S}_n = \frac{S_n}{\sqrt{\widehat{D}_{2,n}A^2(\varphi)}} \tag{1.31}$$

that has under assumptions of Theorem 1.6 under $\mathbf{H}_1$ asymptotically standard normal distribution.

Now, we will add measurement errors to both regressors $x_{1,i}$ and $x_{2,i}$ and use the previous test (1.31). We will illustrate here some simulation results how the test based on Wilcoxon scores ($\varphi(u) = u - 1/2$) perform in general measurement error model (1.19). For simplicity $\mathbf{x}_i = (x_{1,i}, x_{2,i})^\top$ were once generated from two-dimensional normal distribution $\mathcal{N}_2(\mathbf{0}, \mathbf{B}_k)$ and then considered fixed, model errors e_i were generated from standard normal distribution and measurement errors $\mathbf{v}_i$ were generated also from two-dimensional normal distribution $\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_k)$, where

$$\mathbf{B}_k = \begin{pmatrix} 3 & \frac{k}{10} \\ \frac{k}{10} & 2 \end{pmatrix}, \quad \widetilde{\mathbf{B}}_k = \begin{pmatrix} 1 & \frac{k}{10} \\ \frac{k}{10} & 1 \end{pmatrix}.$$

In Table 1.2 empirical error of the first kind is computed as percentage of rejections of $\mathbf{H}_1$ at significance level $\alpha = 0.05$ (when $\beta_0 = 1,\ \beta_1 = 2,\ \beta_2 = 0$) based on 10 000 replications. Mean of the test statistic (1.31) is also added in parentheses.

$\mathbf{x}_i \backslash \mathbf{v}_i$	$\mathbf{0}$	$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_0)$	$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_1)$	$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_2)$	$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_5)$	$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_9)$
$\mathcal{N}_2(\mathbf{0}, \mathbf{B}_0)$	5.42(0.026)	4.68(0.144)	4.34(−0.077)	5.14(−0.300)	14.81(−0.980)	52.35(−1.954)
$\mathcal{N}_2(\mathbf{0}, \mathbf{B}_1)$	5.47(−0.024)	4.62(0.178)	4.02(−0.039)	4.58(−0.261)	14.41(−0.942)	50.46(−1.919)
$\mathcal{N}_2(\mathbf{0}, \mathbf{B}_2)$	5.34(−0.021)	4.97(0.222)	4.03(0.008)	4.63(−0.214)	13.22(−0.894)	48.22(−1.873)
$\mathcal{N}_2(\mathbf{0}, \mathbf{B}_5)$	5.36(−0.015)	5.89(0.380)	4.31(0.166)	3.86(−0.056)	10.26(−0.744)	42.36(−1.750)
$\mathcal{N}_2(\mathbf{0}, \mathbf{B}_7)$	5.31(−0.012)	7.36(0.513)	5.13(0.298)	3.90(0.074)	8.31(−0.625)	38.62(−1.658)
$\mathcal{N}_2(\mathbf{0}, \mathbf{B}_9)$	5.24(−0.011)	9.35(0.662)	6.27(0.446)	4.55(0.220)	6.61(−0.488)	34.16(−1.556)

Table 1.2: Empirical error of the first kind of RRS test and mean of $\widetilde{S}_n$ (in parentheses) for various measurement errors $\mathbf{v}_i$ and regressors $\mathbf{x}_i$; $n = 30$.

According to our simulation results if components of $\mathbf{x}_i$ or $\mathbf{v}_i$ are (highly) correlated, then the test is biased. In addition, in Table 1.3 we will show that the error of the first kind depends on unknown nuisance parameter β_1. Simulation design remains the same, $\mathbf{x}_i$ once generated from $\mathcal{N}_2(\mathbf{0}, \mathbf{B}_7)$ and then considered fixed. We may conclude this example that in general RRS test in not applicable in measurement error model (1.19). However, it is still an open problem to find the asymptotic distribution of the test in model (1.19), at least for some local values of nuisance parameter.

$\beta_1 \backslash \mathbf{v}_i$	$\mathbf{0}$	$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_0)$	$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_1)$	$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_5)$	$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_7)$	$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{B}}_9)$
0	5.31(−0.012)	5.60(0.003)	5.39(−0.011)	5.47(−0.015)	5.55(−0.016)	5.62(−0.017)
1	5.50(−0.011)	5.33(0.239)	4.52(0.073)	8.64(−0.609)	15.17(−0.965)	26.06(−1.336)
−1	5.91(−0.014)	7.63(−0.500)	6.17(−0.353)	5.93(0.285)	9.05(0.627)	16.10(0.993)
2	5.29(−0.008)	4.91(0.288)	4.22(0.059)	11.84(−0.850)	24.94(−1.326)	45.74(−1.826)
−2	5.35(−0.017)	9.18(−0.677)	6.63(−0.443)	7.08(0.487)	15.30(0.994)	33.07(1.538)
3	5.28(−0.003)	4.06(0.025)	4.81(−0.234)	22.32(−1.219)	41.75(−1.732)	67.38(−2.264)
−3	5.96(−0.003)	13.83(−0.994)	9.27(−0.781)	2.87(0.121)	6.15(0.617)	16.61(1.157)
4	5.55(−0.008)	8.37(0.631)	5.36(0.360)	11.42(−0.801)	28.98(−1.430)	60.35(−2.109)
−4	5.31(−0.001)	5.60(−0.343)	3.91(−0.062)	17.13(1.067)	38.78(1.667)	69.37(2.301)

Table 1.3: Empirical error of the first kind of RRS test and mean of $\widetilde{S}_n$ (in parentheses) for various measurement errors $\mathbf{v}_i$, regressors $\mathbf{x}_i$ fixed from $\mathcal{N}_2(\mathbf{0}, \mathbf{B}_7)$; $n = 30$.

2. Another rank tests in measurement error models

2.1 Tests based on minimum distance estimates

Koul (2002) considered a class of estimates in linear regression model based on minimization of certain type of distances, he also proved their asymptotic properties and derived their asymptotic representation. However, he did not considered the problem of testing hypotheses about these parameters.

Consider model of regression line

$$Y_i = \beta_0 + x_i\beta + e_i, \quad i = 1, \dots, n, \tag{2.1}$$

where β_0 and β are unknown parameters, $x_1, \dots, x_n$ are stochastic regressors, $e_1, \dots, e_n$ are assumed to be i.i.d. with an unknown distribution function F and uniformly continuous density f independent with $x_1, \dots, x_n$ and our aim is to test the hypothesis

$$\mathbf{H}_0 : \ \beta = 0 \quad \text{against} \quad \mathbf{K}_0 : \ \beta \neq 0.$$

The situation here is just a special case of that considered in Section 1.1.1. The motivation for the following approach is that the corresponding test may have greater power than classical rank test for some model errors (Koul (2002) proved similar property for corresponding estimates). In addition, these tests might be robust to departures in the assumed design variables.

Let us introduce a class of test statistics based on Cramér - von Mises type of distance involving various weighted empirical processes. Define

$$T_{g,n}(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(x_i)\mathbb{I}\{R_i \leq ns\}, \quad 0 \leq s \leq 1, \tag{2.2}$$

$$K^*_{g,n} = \int_0^1 T^2_{g,n}(s)dL(s), \tag{2.3}$$

where R_i is the rank of Y_i among $Y_1, \dots, Y_n$, L a distribution function on $[0, 1]$ and g a real (weight) function such that $\sum_{i=1}^{n} g(x_i) = 0$.

Discuss some computation aspects of (2.3). First, have a look at the formula (2.3) for $K^*_{g,n}$. Inserting (2.2) into (2.3) we have

$$\begin{aligned} K^*_{g,n} &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} g(x_i)g(x_j) \int_0^1 \mathbb{I}\{R_i \leq ns\}\mathbb{I}\{R_j \leq ns\}dL(s) \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} g(x_i)g(x_j) \int_{\max\left\{\frac{R_i}{n}, \frac{R_j}{n}\right\}}^1 1dL(s). \end{aligned}$$

L is a distribution function, hence $L(\max\{a, b\}) = \max\{L(a), L(b)\}$, it also remains true for limits from the left, hence

$$K^*_{g,n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} g(x_i)g(x_j) \left(1 - \max\left\{L\left(\frac{R_i}{n}-\right), L\left(\frac{R_j}{n}-\right)\right\}\right).$$

Using $\sum_{i=1}^{n} g(x_i) = 0$ we get

$$K^*_{g,n} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g(x_i)g(x_j)\max\left\{L\left(\frac{R_i}{n}-\right), L\left(\frac{R_j}{n}-\right)\right\}.$$

Using the fact that

$$2\max\{a,b\} = a + b + |a-b|, \quad \forall\, a,b \in \mathbb{R}$$

and $\sum_{i=1}^{n} g(x_i) = 0$ we get

$$K^*_{g,n} = -\frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n} g(x_i)g(x_j)\left|L\left(\frac{R_i}{n}-\right) - L\left(\frac{R_j}{n}-\right)\right|,$$

which is much more convenient for practical computations.

Under $\mathbf{H}_0$ model (2.1) reduces to

$$Y_i = \beta_0 + e_i, \quad i = 1,\ldots,n. \tag{2.4}$$

As distribution of model errors e_i is absolutely continuous, there cannot be any ties in ranks with probability 1 and thanks to invariance of ranks with respect to the location, distribution of $R_1,\ldots,R_n$ under hypothesis is uniform over all $n!$ permutations of numbers $\{1,\ldots,n\}$ (see Lemma 1.1). Therefore distribution of $K^*_{g,n}$ given $x_1,\ldots,x_n$ under $\mathbf{H}_0$ will be distribution-free and may be even computed directly. To do it we have to compute all values of test statistic $K^*_{g,n}$ for each of $n!$ permutations of numbers $\{1,\ldots,n\}$ and order these values in the increasing magnitude. The critical region is then formed by $M = \lfloor \alpha n! \rfloor$ largest values and the combination which leads to the $(M+1)$-st largest value can be possibly randomized.

However, for large sample size n computation of exact (conditional) distribution may be computationally demanding, that is why we will investigate asymptotic distribution of $K^*_{g,n}$.

Let us introduce for $s \in [0,1]$ some empirical processes

$$V_{g,n}(s) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(x_i)\mathbb{I}\{e_i \le F_n^{-1}(s)\},$$

$$\widehat{V}_{g,n}(s) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(x_i)\mathbb{I}\{e_i \le F^{-1}(s)\},$$

where $F_n(s) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{e_i \le s)\}$ is empirical distribution function.

Now, we state assumptions under which we will be able to prove asymptotic properties of $K^*_{g,n}$

$$\sum_{i=1}^{n}(x_i - \bar{x}) > 0 \text{ a.s.} \quad \forall\, n > 1, \tag{2.5}$$

$$\max_{i=1,\dots,n} \frac{(x_i - \bar{x})^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \xrightarrow{p} 0, \tag{2.6}$$

$$g(x_i) \neq 0 \text{ a.s. for some} \quad i = 1,\dots,n, \tag{2.7}$$

$$0 < |\mathbb{E}g(X_1)(X_1 - \mathbb{E}X_1)| < \infty, \tag{2.8}$$

$$x_i g(x_i) \geq 0 \text{ a.s.} \quad \forall i = 1,\dots,n \text{ or } x_i g(x_i) \leq 0 \text{ a.s.} \quad \forall i = 1,\dots,n, \tag{2.9}$$

$$\max_{i=1,\dots,n} g^2(x_i) \xrightarrow{p} 0, \tag{2.10}$$

$$\sup_{n\in\mathbb{N}} \max_{i=1,\dots,n} |g(x_i)| \leq c \text{ a.s.} \quad \text{for some} \quad 0 < c < \infty, \tag{2.11}$$

$$0 < \gamma = \sqrt{\mathbb{E}g^2(X_1)} < \infty. \tag{2.12}$$

Lemma 2.1. *Under (2.5) – (2.8) it holds*

$$\left| K^*_{g,n} - \int V^2_{g,n}(s)dL(s) \right| = o_p(1), \quad as \quad n \to \infty.$$

Proof. For convenience we will drop off an index g in $K^*_{g,n}$ and $V_{g,n}$. Adding and subtracting $V_n(s)$ in the first integral, squaring and using Cauchy-Schwarz inequality we get

$$\begin{aligned}
&\Big| \int T_n^2(s)dL(s) - \int V_n^2(s)dL(s) \Big| \\
&= \Big| \int [T_n(s) - V_n(s)]^2\, dL(s) + 2\int V_n(s)(T_n(s) - V_n(s))dL(s) \Big| \\
&\leq \sup_{0\leq s\leq 1} |T_n(s) - V_n(s)|^2 + 2\sqrt{\int V_n^2(s)dL(s) \int (T_n(s) - V_n(s))^2 dL(s)}.
\end{aligned}$$

The fact that

$$\sup_{0\leq s\leq 1} |T_n(s) - V_n(s)| \leq 2 \max_{i=1,\dots,n} |g(x_i)| = o_p(1)$$

together with $\int V_n^2(s)dL(s) = O_p(1)$ proves the Lemma. □

Lemma 2.2. *Under (2.5) – (2.8) it holds*

$$\left| K^*_{g,n} - \int \widehat{V}^2_{g,n}(s)dL(s) \right| = o_p(1), \quad as \quad n \to \infty.$$

Proof.

$$\begin{aligned}
&\Big| \int T_n^2(s)dL(s) - \int \widehat{V}_n^2(s)dL(s) \Big| \\
&= \Big| \int \left[T_n(s) - \widehat{V}_n(s)\right]^2 dL(s) + 2\int \widehat{V}_n(s)(T_n(s) - \widehat{V}_n(s))dL(s) \Big|.
\end{aligned} \tag{2.13}$$

Using Minkowski inequality

$$\int [T_n(s) - \widehat{V}_n(s)]^2 dL(s) = \int [T_n(s) - V_n(s) + V_n(s) - \widehat{V}_n(s)]^2 dL(s)$$
$$\leq 2\int [T_n(s) - V_n(s)]^2 dL(s) + 2\int [V_n(s) - \widehat{V}_n(s)]^2 dL(s). \qquad (2.14)$$

By Cauchy-Schwarz inequality

$$\begin{aligned}\Big|\int \widehat{V}_n(s)(T_n(s) - \widehat{V}_n(s))dL(s)\Big| &\leq \sqrt{\int \widehat{V}_n^2(s)dL(s)\int [T_n(s) - \widehat{V}_n(s)]^2 dL(s)} \\ &= o_p(1), \qquad (2.15)\end{aligned}$$

because $\int \widehat{V}_n^2(s)dL(s) = O_p(1)$ and $\int [T_n(s) - \widehat{V}_n(s)]^2 dL(s) = o_p(1)$.

Observe that

$$\widehat{V}_n(FF_n^{-1}(s)) = \sum_{i=1}^n g(x_i)\mathbb{I}\{e_i \leq F^{-1}FF_n^{-1}(s)\} = \sum_{i=1}^n g(x_i)\mathbb{I}\{e_i \leq F_n^{-1}(s)\} = V_n(s).$$

Therefore

$$\sup_{0\leq s\leq 1} |V_n(s) - \widehat{V}_n(s)| = \sup_{0\leq s\leq 1} |\widehat{V}_n(FF_n^{-1}(s)) - \widehat{V}_n(s)| = o_p(1),$$

because

$$\sup_{0\leq s\leq 1} |FF_n^{-1}(s) - s| = \sup_{0\leq s\leq 1} |FF^{-1}(s) - F_nF_n^{-1}(s) + F_nF_n^{-1}(s) - s|$$
$$\leq \sup_{x\in\mathbb{R}} |F(x) - F_n(x)| + \sup_{0\leq s\leq 1} |F_nF_n^{-1}(s) - s| = o_p(1).$$

Now, combining previous result, Lemma 2.1 and (2.13), (2.14) and (2.15) we have proven the Lemma. □

The previous lemma states that the asymptotic distribution of $K_{g,n}^*$ will be the same as $\int \widehat{V}_{g,n}^2(s)dL(s)$ that is easier to investigate. Hence we are now able to state the theorem about asymptotic null distribution of $K_{g,n}^*$.

Theorem 2.1. *Under (2.5) – (2.12) in model (2.1) under* $\mathbf{H}_0$

$$K_{g,n}^* \xrightarrow{d} \gamma^2 \cdot Y_L, \quad \textit{with} \quad Y_L = \int_0^1 B^2(s)dL(s),$$

where $B(s)$ *is a Brownian bridge in* $\mathcal{C}[0,1]$.

Proof. Recall that

$$\begin{aligned}\widehat{V}_{g,n}(s) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n g(x_i)\mathbb{I}\{e_i \leq F^{-1}(s)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n g(x_i)\mathbb{I}\{F(e_i) \leq s\} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n g(x_i)\mathbb{I}\{U_i \leq (s)\},\end{aligned}$$

where $U_1, \dots, U_n$ are i.i.d. random variables with uniform $\mathcal{U}(0,1)$ distribution.

By Koul (2002) we have

$$\widehat{V}_{g,n}(s) \Rightarrow \gamma \cdot B(s) \quad \text{in} \quad \mathcal{D}[0,1]$$

and therefore $\int \widehat{V}^2_{g,n}(s)dL(s) \xrightarrow{d} \gamma^2 \int B^2(s)dL(s)$. That together with Lemma 2.2 proves the Theorem. □

Distribution of random variable Y_L for $L(s) = s$ was first investigated by Smirnov (1936), values of its distribution function may be found for example in Anderson and Darling (1952) or in Tolmatz (2002) and Tolmatz (2003). For other choices of function L one has to use simulated values of quantiles. That gives us a clue how to use asymptotic version of proposed test. Since γ is unknown we have to estimate it. The most natural way is to estimate it by $\sqrt{\sum_{i=1}^n g^2(x_i)}$. Finally compute test statistic

$$\widetilde{K}^*_{g,n} = \frac{1}{\sqrt{\sum_{i=1}^n g^2(x_i)}} K^*_{g,n}$$

and reject $\mathbf{H}_0$ if $\widetilde{K}^*_{g,n}$ is greater than $(1-\alpha)$ - quantile of distribution Y_L. Some quantiles for distribution function $L(s) = s$ are listed in Table 2.1.

$(1-\alpha)$	0.90	0.95	0.99	0.999
$(1-\alpha)$ - quantile	0.34730	0.46136	0.74346	1.16786

Table 2.1: Quantiles of distribution Y_L for $L(s) = s$.

Now, we will investigate behavior of $K^*_{g,n}$ under local alternative

$$\mathbf{K}_{0,n}: \ \beta = n^{-1/2}\beta^*, \quad 0 \neq \beta^* \in \mathbb{R} \text{ fixed.}$$

For $t \in \mathbb{R}$ define

$$K^*_{g,n}(t) = \int_0^1 \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(x_i)\mathbb{I}\{R_{i,t} \leq ns\} \right)^2 dL(s), \tag{2.16}$$

$$\widehat{K}^*_{g,n}(t) = \\ = \int_0^1 \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(x_i))\mathbb{I}\{U_i \leq s\} - \frac{t}{\sqrt{n}} \sum_{i=1}^n g(x_i)(x_i - \bar{x}) f(F^{-1}(s)) \right)^2 dL(s), \tag{2.17}$$

where $R_{i,t}$ is the rank of $Y_i - x_i t$ among $Y_1 - x_1 t, \dots, Y_n - x_n t$ and $U_1, \dots, U_n$ are i.i.d. random variables with uniform $\mathcal{U}(0,1)$ distribution.

Remark. *Koul (2002) defined an estimator of β as a minimizer of (2.16). Hence proposed test statistic $K^*_{g,n}$ is $K^*_{g,n}(t)$ computed in the hypothetical value $t = 0$, i.e. $K^*_{g,n} = K^*_{g,n}(0)$ is the test statistic under $\mathbf{H}_0$, while $K^*_{g,n}(n^{-1/2}\beta^*)$ is the test statistic under $\mathbf{K}_{0,n}$.*

Lemma 2.3. *Assume that (2.5) – (2.8) hold. Then for every* $0 < b < \infty$

$$\sup_{|u|\le b} |K^*_{g,n}(n^{-1/2}u) - \widehat{K}^*_{g,n}(n^{-1/2}u)| = o_p(1), \quad \textit{as} \quad n \to \infty.$$

Proof. See Koul (2002, Theorem 5.5.5). □

Remark. *Particularly, if* $u = 0$ *in Lemma 2.3, we get* $K^*_{g,n}(0) = \widehat{K}^*_{g,n}(0) + o_p(1)$, *i.e.*

$$K^*_{g,n} = \int_0^1 \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(x_i)) \mathbb{I}\{U_i \le s\} \right)^2 dL(s) + o_p(1).$$

Hence Lemma 2.2 is a special case of Lemma 2.3.

Now, let (2.5) – (2.12) be satisfied. Rewrite (2.17) as

$$\widehat{K}^*_{g,n}(t) = \widehat{K}^*_{g,n}(0) - \frac{2t}{n} \sum_{j=1}^n g(x_j)(x_j - \bar{x}) \sum_{i=1}^n g(x_i) \int_0^1 \mathbb{I}\{U_i \le s\} f(F^{-1}(s)) dL(s)$$

$$+ \frac{t^2}{n} \left(\sum_{i=1}^n g(x_i)(x_i - \bar{x}) \right)^2 \int_0^1 f^2(F^{-1}(s)) dL(s)$$

$$= \widehat{K}^*_{g,n}(0) + \frac{2t}{n} \sum_{j=1}^n g(x_j)(x_j - \bar{x}) \sum_{i=1}^n g(x_i)\varphi(U_i) + \frac{t^2}{n} \sigma_{f,L} \left(\sum_{i=1}^n g(x_i)(x_i - \bar{x}) \right)^2,$$

where $\varphi(u) = \int_0^u f(F^{-1}(s))dL(s)$ and $\sigma_{f,L} = \int_0^1 f^2(F^{-1}(s))dL(s)$. From Lemma 2.3, (2.16) and (2.17) we finally get

$$\begin{aligned} K^*_{g,n}(n^{-1/2}\beta^*) &= K^*_{g,n}(0) + 2\beta^* \frac{1}{n} \sum_{j=1}^n g(x_j)(x_j - \bar{x}) \frac{1}{\sqrt{n}} \sum_{i=1}^n g(x_i)\varphi(U_i) \\ &+ (\beta^*)^2 \sigma_{f,L} \left(\frac{1}{n} \sum_{i=1}^n g(x_i)(x_i - \bar{x}) \right)^2 + o_p(1). \end{aligned} \tag{2.18}$$

Right hand side of (2.18) converges to convolution of two (dependent) random variables $\gamma^2 \cdot Y_L$ and $Z \sim \mathcal{N}(a, b)$, where

$$\begin{aligned} a &= (\beta^*)^2 \sigma_{f,L} (\mathbb{E}\{g(X_1) \cdot (X_1 - \mathbb{E}X_1)\})^2, \\ b &= 4(\beta^*)^2 \left[\mathbb{E}\{g(X_1) \cdot (X_1 - \mathbb{E}X_1)\}\right]^2 \mathbb{E}g^2(X_1) \operatorname{var} \varphi(U_i). \end{aligned}$$

Hence under (2.5) – (2.12) asymptotic distribution of $K^*_{g,n}$ under local alternative $\mathbf{K}_{0,n}$ is the above convolution of dependent random variables.

For practical application it arises a natural question how to choose functions g and L. The function g is in fact a weight function for regressors, so it can downweight outlying observations to robustify these tests against extreme values of x_i (if g is bounded for example). The function L has similar interpretation as score-function φ in standard rank tests theory, hence L should be chosen based on the distribution of unknown model errors. Anyway, the simplest choice $L(s) = s$ gives very reasonable results (see the simulations).

Now, return back to our problem with measurement errors in model

$$\begin{aligned} Y_i &= \beta_0 + \beta x_i + e_i, \\ w_i &= x_i + v_i, \quad i = 1, \ldots, n, \end{aligned} \tag{2.19}$$

where $x_1, \ldots, x_n$ are unobserved stochastic regressors and $v_1, \ldots, v_n$ are i.i.d. measurement errors (independent with x_i and e_i).

We would like to use our test although actual regressors are not observable. We apply our test based on observed regressors w_i and will show that the test will still work. Actually, it does work, because under $\mathbf{H}_0 : \beta = 0$ measurement error model (2.19) reduces to model (2.4) – the same model as in the case without measurement errors. That means that we may apply exactly the same setup and all the results will remain valid (under $\mathbf{H}_0$ considering use of w_i instead of x_i). Similarly, if (2.5) – (2.12) hold for w_i, then asymptotic null distribution will be the same as in model without measurement errors. The only difference is in the asymptotic distribution under alternative, that namely depends on both regressors x_i and w_i. Intuitively and supported by simulation study the presence of measurement errors will only decrease the power of our test.

Remark. *Extension of this test into multiple regression model is straightforward, the conditional distribution under null hypothesis is again distribution-free and may be computed in similar way. Unfortunately, asymptotic null distribution is difficult to express, it is a convolution of dependent random variables with distribution same as Y_L multiplied by some positive constants.*

To support previous theoretical results we conducted a large simulation study, let us present a few interesting results. Start with model without measurement errors (2.1) for moderate sample size $n = 30$. We have compared empirical power of our test based on the test statistic $\widetilde{K}^*_{g,n}$ with $g(x_i) = x_i - \bar{x}$ and $L(s) = s$ (call it *minimum distance test*) with Wilcoxon test for regression (see Section 1.1.1) and standard t-test for regression.

The regressors $x_1, \ldots, x_{30}$ were generated from uniform $\mathcal{U}(-2, 10)$ distribution. The model errors e_i were generated from normal, logistic, Laplace and t-distribution with 6 degrees of freedom, respectively, always with 0 mean a variance 3/2. The empirical powers of the tests were computed as a percentage of rejections of $\mathbf{H}_0$ among 10 000 replications, at significance level $\alpha = 0.05$. The results are summarized in Table 2.2.

Now, let us compare the three previous tests in measurement error model (2.19) with the same simulation setup as before. Empirical error of the first kind for various measurement errors are summarized in Table 2.3, empirical power for various measurement errors are summarized in Table 2.4 (with a true value of parameter $\beta = 0.2$).

According to Table 2.3 minimum distance test preserves error of the first kind at prescribed α even if measurement errors are present. For normal model errors t-test achieves the largest power, for distributions with lighter tails than

$\beta \setminus e_i$	$\mathcal{N}\left(0,\frac{3}{2}\right)$			$Log\left(0,\frac{\sqrt{2}\pi}{3}\right)$			$Lap\left(0,\frac{\sqrt{3}}{2}\right)$			$t(6)$		
0	4.98	4.42	5.00	5.06	4.55	5.00	5.00	4.55	5.04	5.00	4.32	4.93
0.1	28.7	28.3	31.5	32.7	31.4	32.0	42.4	39.0	33.5	34.6	33.1	32.9
−0.1	28.3	28.2	30.9	32.7	31.2	32.2	42.5	39.0	33.7	33.3	32.1	31.9
0.2	78.2	78.8	82.3	82.5	81.8	81.9	88.3	86.6	82.0	84.5	83.9	82.6
−0.2	78.3	78.7	82.9	83.3	82.7	82.9	89.2	87.5	83.1	84.0	83.4	82.5

Table 2.2: Percentage of rejections of hypothesis $\mathbf{H}_0 : \ \beta = 0$ of minimum distance test, Wilcoxon test for regression and t-test for regression; $n = 30$.

$v_i \setminus e_i$	$\mathcal{N}\left(0,\frac{3}{2}\right)$			$Log\left(0,\frac{\sqrt{2}\pi}{3}\right)$			$Lap\left(0,\frac{\sqrt{3}}{2}\right)$			$t(6)$		
0	4.98	4.42	5.00	5.06	4.55	5.00	5.00	4.55	5.04	5.00	4.32	4.93
$\mathcal{N}(0,4)$	4.45	3.94	4.39	4.90	4.29	5.03	4.90	4.29	5.04	4.99	4.66	4.92
$\mathcal{N}(0,6)$	4.53	3.97	4.44	4.81	4.41	5.05	4.81	4.41	5.06	4.77	4.59	4.95
$\mathcal{U}(-\sqrt{18},\sqrt{18})$	5.49	4.78	5.36	5.13	4.53	4.97	5.13	4.53	4.81	4.51	3.85	4.34
$2t(6)$	5.09	4.63	5.04	5.11	4.59	4.94	5.11	4.59	4.96	5.17	4.51	4.81
$\mathcal{U}(-6,6)$	5.50	4.73	5.42	5.18	4.62	5.12	5.18	4.62	4.85	4.87	4.19	4.55

Table 2.3: Percentage of rejections of hypothesis $\mathbf{H}_0 : \ \beta = 0$ of minimum distance test, Wilcoxon test for regression and t-test for regression; true $\beta = 0$, $n = 30$.

normal our test has the largest power even for logistic distribution due to slow convergence of rank test statistic to its asymptotic distribution. The presence of measurement errors only decreases the power of all tests – the larger variance of measurement errors the smaller power.

2.2 Tests about an intercept

Consider linear regression model with possible measurement errors (1.13):

$$\begin{aligned} Y_i &= \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + e_i, \\ \mathbf{w}_i &= \mathbf{x}_i + \mathbf{v}_i, \quad i = 1, \ldots, n, \end{aligned}$$

where $\beta_0 \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^p$ are unknown parameters, model errors e_i are assumed to be independent identically distributed with an unknown distribution function F, symmetric density f and finite Fisher information $I(f)$ and $\mathbf{x}_i$ are vectors of known (unobserved) regressors and $\mathbf{w}_i$ are observed regressors affected by some measurement errors $\mathbf{v}_i$, such that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are i.i.d. random vectors independent of $\mathrm{e}_1, \ldots, e_n$ with an unknown, but symmetric distribution.

Now, our aim is to test the hypothesis

$$\mathbf{H}_0 : \ \beta_0 = 0 \quad \text{against} \quad \mathbf{K}_0 : \ \beta_0 > 0,$$

while $\boldsymbol{\beta}$ is vector of nuisance parameters.

$v_i \setminus e_i$	$\mathcal{N}\left(0, \frac{3}{2}\right)$			$Log\left(0, \frac{\sqrt{2}\pi}{3}\right)$			$Lap\left(0, \frac{\sqrt{3}}{2}\right)$			$t(6)$		
0	78.2	78.8	82.3	82.5	81.8	81.9	88.3	86.6	82.0	84.5	83.9	82.6
$\mathcal{N}(0,4)$	64.1	63.8	68.2	69.1	67.9	68.2	76.4	74.0	68.8	71.5	70.4	69.6
$\mathcal{N}(0,6)$	58.1	57.8	61.7	63.0	62.2	62.6	71.0	68.0	63.5	65.9	64.7	63.5
$\mathcal{U}(-\sqrt{18}, \sqrt{18})$	58.4	58.4	62.5	62.7	61.9	62.8	70.7	67.8	63.8	65.8	64.3	63.6
$2t(6)$	59.0	59.0	62.5	63.8	62.5	63.0	71.4	68.6	64.1	66.8	65.8	64.6
$\mathcal{U}(-6,6)$	45.2	44.7	48.0	49.8	48.4	49.4	57.0	54.1	50.3	51.6	50.2	49.6

Table 2.4: Percentage of rejections of hypothesis $\mathbf{H}_0: \ \beta = 0$ of minimum distance test, Wilcoxon test for regression and t-test for regression; true $\beta = 0.2$, $n = 30$.

Without any further information about regressors $\mathbf{x}_i$ it is impossible to make statistical inference about parameter β_0 (problem of identifiability). To be able to test $\mathbf{H}_0$, we will assume that regressors $\mathbf{x}_i$ are centered, i.e. $\sum_{i=1}^n x_{i,j} = 0$ for all $j = 1, \dots, p$.

Similarly as in Section 1.2 we shall take a recourse to the aligned rank test, replacing the nuisance slope parameter $\boldsymbol{\beta}$ with its estimator $\widehat{\boldsymbol{\beta}}_n$ and then constructing the signed rank test based on aligned ranks and signs of the residuals.

As an estimator $\widehat{\boldsymbol{\beta}}_n$ we take the R-estimator based on the hypothetical model affected by the measurement errors

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

where $e_i^* = e_i^*(\boldsymbol{\beta}) = e_i - \mathbf{v}_i^\top \boldsymbol{\beta}$ are model errors – i.i.d. random variables with symmetric distribution function F_β^* and density f_β^* and finite Fisher information. Hence we are in the same situation as in Section 1.2 and we will just repeat the estimation procedure. Recall briefly that we choose a nondecreasing, nonconstant, square integrable skew-symmetric score function $\psi : (0,1) \mapsto \mathbb{R}$ and define scores

$$\widetilde{a}_n(i) = \psi\left(\frac{i}{n+1}\right), \quad i = 1, \dots, n.$$

The R-estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ is then a minimizer of

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^n \left(Y_i - \mathbf{w}_i^\top \mathbf{b}\right) \widetilde{a}_n(R_i(\mathbf{b}))$$

with respect to $\mathbf{b} \in \mathbb{R}^p$, where $R_i(\mathbf{b})$ is the rank of $(Y_i - \mathbf{w}_i^\top \mathbf{b})$ among $(Y_1 - \mathbf{w}_1^\top \mathbf{b}), \dots, (Y_n - \mathbf{w}_n^\top \mathbf{b})$.

Now, consider residuals $\widehat{e}_i = \widehat{e}_i(\widehat{\boldsymbol{\beta}}_n) = Y_i - \mathbf{w}_i^\top \widehat{\boldsymbol{\beta}}_n, \quad i = 1, \dots, n$. The signed-rank test about β_0 will be then based on these residuals. Note that $\widehat{e}_1, \dots, \widehat{e}_n$ are not independent (the classical case with i.i.d. errors will be further discussed in Chapter 5).

Having chosen a nondecreasing, nonconstant, square integrable score function $\varphi : (0,1) \mapsto \mathbb{R}$ (it may differ from ψ) define $\varphi^+(u) = \varphi(\frac{u+1}{2})$, approximate scores

$$a_n^+(i) = \varphi^+\left(\frac{i}{n+1}\right)$$

and compute

$$\widehat{S}_n^+(\widehat{\boldsymbol{\beta}}_n) = n^{-1/2} \sum_{i=1}^{n} a_n^+(R_i^+(\widehat{\boldsymbol{\beta}}_n)) \operatorname{sign}(\widehat{e}_i),$$

where $R_i^+(\widehat{\boldsymbol{\beta}}_n)$ is the rank of $|\widehat{e}_i|$ among $|\widehat{e}_1|, \dots, |\widehat{e}_n|$.

Now, state the assumptions needed for the proof of asymptotic normality of $\widehat{S}_n^+(\widehat{\boldsymbol{\beta}}_n)$. Suppose that there exist positive definite matrices $\mathbf{Q}, \mathbf{V}$ such that as $n \to \infty$

$$\mathbf{Q}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\top \to \mathbf{Q}, \quad \mathbf{V}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}_i - \overline{\mathbf{v}})(\mathbf{v}_i - \overline{\mathbf{v}})^\top \to \mathbf{V} \quad (2.20)$$

and moreover

$$\frac{1}{n}\max_{1\le i\le n}(\mathbf{x}_i - \overline{\mathbf{x}})^\top \mathbf{Q}_n^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}) \to 0, \quad \frac{1}{n}\max_{1\le i\le n}(\mathbf{v}_i - \overline{\mathbf{v}})^\top \mathbf{V}_n^{-1}(\mathbf{v}_i - \overline{\mathbf{v}}) \to 0. \quad (2.21)$$

Theorem 2.2. *Assume that (2.20)–(2.21) hold. Let f be symmetric with finite Fisher information. Then in model (1.13) under $\mathbf{H}_0$ test statistic $\widehat{S}_n^+(\widehat{\boldsymbol{\beta}}_n)$ has asymptotically as $n \to \infty$ normal distribution $\mathcal{N}(0, \sigma^2)$ with $\sigma^2 = A^2(\varphi^+)$ and under local alternative*

$$\mathbf{K}_{0,n}: \ \beta_0 = n^{-1/2}\beta_0^*, \quad 0 < \beta_0^* \quad \textit{fixed}$$

$\widehat{S}_n^+(\widehat{\boldsymbol{\beta}}_n)$ has asymptotically normal distribution $\mathcal{N}(\mu, \sigma^2)$ with

$$\mu = \beta_0^* \cdot \gamma(\varphi^+, f_\beta^*). \quad (2.22)$$

Proof. As in Jurečková et al. (2009) we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \frac{1}{\gamma(\psi, f_\beta^*)}(\mathbf{Q} + \mathbf{V})^{-1}\mathbf{L}_n(\boldsymbol{\beta}) + o_p(1), \quad \text{as} \quad n \to \infty, \quad (2.23)$$

where

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

The signed rank statistic $\widehat{S}_n^+$ is uniformly asymptotically linear, i.e. for any fixed $K > 0$ and as $n \to \infty$

$$\sup_{\|\mathbf{t}\|\le K}\left\{|\widehat{S}_n^+(n^{-1/2}\mathbf{t}) - \widehat{S}_n^+(0)|\right\} \xrightarrow{p} 0. \quad (2.24)$$

Inserting $\mathbf{t} = \sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ into (2.24) and together with (2.23) and the fact that $\widehat{S}_n^+(\widehat{\boldsymbol{\beta}}_n) = O_p(1)$ we get

$$|\widehat{S}_n^+(\widehat{\boldsymbol{\beta}}_n) - \widehat{S}_n^+(\boldsymbol{\beta})| \xrightarrow{p} 0, \quad \text{as} \quad n \to \infty. \quad (2.25)$$

Now, asymptotic distribution of $\widehat{S}_n^+(\boldsymbol{\beta})$ under $\mathbf{H}_0$ is $\mathcal{N}(0, \sigma^2)$ (see Hájek et al. (1999, Theorem 1, Section 6.1.7)) and under $\mathbf{K}_{0,n}$ is $\mathcal{N}(\mu, \sigma^2)$ (see Hájek et al. (1999, Theorem 1, Section 7.2.5)). That together with (2.25) completes the proof. □

Remark. *Recall that f^*_β depends on unknown nuisance parameter $\boldsymbol{\beta}$ and distribution of measurement errors $\mathbf{v}_i$, hence the asymptotic power of the test does depend on the nuisance parameter $\boldsymbol{\beta}$ unlike the situation without measurement errors.*

Again, we made an extensive simulation study to illustrate how the proposed procedures work in finite sample situation and indicate influence of the measurement errors for test about an intercept.

Consider model of regression line

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

and test $\mathbf{H}_0: \ \beta_0 = 0$ against $\beta_0 > 0$. The regressors x_i were once generated from independent sample of size $n = 50$ from uniform $\mathcal{U}(-6, 6)$ distribution and further on considered fixed and the model errors e_i were generated from standard normal distribution. We considered Wilcoxon aligned signed rank test that corresponds to the score function $\varphi(u) = u - 1/2$. For the estimation of nuisance parameter score function $\psi(u) = u - 1/2$ was used. The empirical powers of the tests were computed as a percentage of rejections of $\mathbf{H}_0$ among 10 000 replications, at significance level $\alpha = 0.05$.

Empirical powers of Wilcoxon aligned signed rank test for various measurement errors v_i are summarized in Table 2.5 (value of nuisance parameter β was taken $\beta = 1$). Empirical power of Wilcoxon aligned signed rank test for various measurement errors v_i and for various values of nuisance parameter β are summarized in Table 2.6 (the true value of β_0 was taken $\beta_0 = 0.3$).

$\beta_0 \setminus v_i$	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 2.5: Percentage of rejections of hypothesis $\mathbf{H}_0: \ \beta_0 = 0$ for various measurement errors v_i for Wilcoxon aligned signed rank test; $\beta = 1$, $n = 50$.

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 2.5 and 2.6. Our simulation shows that proposed test actually works, error of the first kind is under control (it is about prescribed $\alpha = 0.05$); only its power decreases with increasing variance of measurement errors. Unlike the model without measurement errors, power of 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.

$\beta \setminus v_i$	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 2.6: Percentage of rejections of hypothesis $\mathbf{H}_0 : \ \beta_0 = 0$ for various measurement errors v_i for Wilcoxon aligned signed rank test; $\beta_0 = 0.3$, $n = 50$.

3. Rank estimates in measurement error models

In Sections 1.2 and 2.2 we already needed to "somehow" estimate the slope parameter $\boldsymbol{\beta}$. In fact we used a naive estimator based on observed regressors $\mathbf{w}_i$ ignoring the fact that there are some measurement errors involved. As pointed out in the Introduction classical least squares estimate (and others) will be biased. However, nobody has proven similar fact for R-estimates, yet. Although Sen and Saleh (2010) and Saleh et al. (2012) considered some estimates based on ranks, they did not computed the formula for asymptotic bias of R-estimates. In this chapter we will show that the asymptotic bias of R-estimates coincides with bias of LSE in classical measurement error model.

3.1 Model and preliminary considerations

Recall that we deal with the model (1.13):

$$\begin{aligned} Y_i &= \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + e_i, \\ \mathbf{w}_i &= \mathbf{x}_i + \mathbf{v}_i, \quad i = 1, \dots, n, \end{aligned}$$

where $\beta_0 \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^p$ are unknown parameters, model errors e_i are assumed to be i.i.d. with an unknown distribution function F and density f, original regressors $\mathbf{x}_i$ are either deterministic or random and affected by additive random measurement errors $\mathbf{v}_1, \dots \mathbf{v}_n$ that are assumed to be i.i.d. with an unknown distribution and independent of the errors $e_1, \dots, e_n$.

We are interested in R-estimator of the slope vector $\boldsymbol{\beta}$, considering β_0 as nuisance parameter. Anyway, it might be considered without loos of generality to be equal 0 (due to invariance of ranks with respect to the location). Remind briefly the approach already used in Sections 1.2 and 2.2. Let $R_i(\mathbf{b})$ be the rank of the residual

$$Y_i - \mathbf{w}_i^\top \mathbf{b} = e_i + \mathbf{x}_i^\top \boldsymbol{\beta} - \mathbf{w}_i^\top \mathbf{b} = e_i - \mathbf{w}_i^\top \mathbf{b}^0 - \mathbf{v}_i^\top \boldsymbol{\beta}, \quad i = 1, \dots, n,$$

where $\mathbf{b}^0 = \mathbf{b} - \boldsymbol{\beta}$ and denote the vector of linear rank statistics

$$\mathbf{S}_n(\mathbf{b}) = \mathbf{S}_n(\mathbf{b}^0, \boldsymbol{\beta}) = n^{-1/2} \sum_{i=1}^{n} (\mathbf{w}_i - \overline{\mathbf{w}}) a_n(R_i(\mathbf{b})), \tag{3.1}$$

where the scores $a_n(i) = \varphi\left(\frac{i}{n+1}\right)$ are generated by score function φ that is skew-symmetric on $(0, 1)$. Jaeckel (1972) defined the rank estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ as a minimizer of

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^{n} \left(Y_i - \mathbf{w}_i^\top \mathbf{b}\right) a_n(R_i(\mathbf{b})) \tag{3.2}$$

with respect to $\mathbf{b} \in \mathbb{R}^p$.

Remark. *In the absence of measurement errors the estimator* $\widehat{\boldsymbol{\beta}}_n$ *is consistent and asymptotically normal. However, unlike the rank test of hypothesis* $\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$, *studied in Section 1.1.3,* $\widehat{\boldsymbol{\beta}}_n$ *is biased in the presence of measurement errors.*

We are able to study asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$ *in the presence of measurement errors and find its local asymptotic bias only in a neighborhood of true value of the parameter* $\boldsymbol{\beta}$, *i.e. under local alternative* $\boldsymbol{\beta}_n = n^{-1/2}\boldsymbol{\beta}^*$ *with a fixed* $\boldsymbol{\beta}^* \in \mathbb{R}^p$.

In the sequel, all limits are taken as $n \to \infty$, unless mentioned otherwise. We shall now describe the needed assumptions on the underlying entities.

A.1 The score generating function $\varphi : (0,1) \mapsto \mathbb{R}$ is nonconstant, nondecreasing, square-integrable and skew-symmetric on (0,1), i.e. satisfies $\varphi(1-t) = -\varphi(t),\ 0 < t < 1$.

F.1 F has an absolutely continuous density f and derivative f' a.e. and has positive and finite Fisher information $I(f)$.

F.2 For every $u \in \mathbb{R}$, $\int \big(|f'(x-tu)|^j/f^{j-1}(x)\big)dx \to \int \big(|f'(x)|^j/f^{j-1}(x)\big)dx < \infty$, as $t \to 0$, $j = 2, 3$.

V.1 The measurement errors $\mathbf{v}_i$ are independent of e_i and have p-dimensional distribution function $\mathbf{G}$ with a continuous density $\mathbf{g}$.

V.2 $\mathbb{E}\mathbf{C}_n \to \mathbf{C}$, where $\mathbf{C}_n = n^{-1}\sum_{i=1}^n (\mathbf{v}_i - \overline{\mathbf{v}})(\mathbf{v}_i - \overline{\mathbf{v}})^\top$ and $\mathbf{C}$ is a positive definite matrix. Moreover, $\sup_{n\geq 1} \mathbb{E}\big(\|\mathbf{v}_n\|^3 + \|\mathbf{x}_n\|^3\big) < \infty$.

V.3 $\mathbb{E}\left[n^{-1}\sum_{i=1}^n (\mathbf{v}_i - \overline{\mathbf{v}})(\mathbf{x}_i - \overline{\mathbf{x}})^\top\right] \to \mathbf{0}$.

X.1 If the regressors $\mathbf{x}_i$ are nonrandom, then assume that $\mathbf{Q}_n \to \mathbf{Q}$, where

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

and $\mathbf{Q}$ is a positive definite matrix. Moreover,

$$\frac{1}{n}\max_{1\leq i\leq n} (\mathbf{x}_i - \overline{\mathbf{x}})^\top \mathbf{Q}_n^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}) \to 0.$$

X.2 If the regressors $\mathbf{x}_i$ are random, then assume that they are independent of $e_i,\ \mathbf{v}_i,\ i = 1, \dots, n$, and

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}}_n)^\top\right] \to \mathbf{Q},$$

where $\mathbf{Q}$ is a positive definite matrix.

3.2 Asymptotic distribution of R-estimate

The following theorem gives the asymptotic distribution of the estimator $\widehat{\boldsymbol{\beta}}_n$ under the local alternative to the hypothesis $\boldsymbol{\beta} = \mathbf{0}$:

$$\boldsymbol{\beta}_n = n^{-1/2}\boldsymbol{\beta}^*, \ \boldsymbol{\beta}^* \in \mathbb{R}^p \ \text{ fixed.} \tag{3.3}$$

Theorem 3.1. *Under the conditions* $\mathbf{A.1}$, $\mathbf{F.1}-\mathbf{F.2}$, $\mathbf{V.1}-\mathbf{V.3}, \mathbf{X.1}-\mathbf{X.2}$ *and under the local alternative (3.3), the R-estimator* $\widehat{\boldsymbol{\beta}}_n$ *in model (1.13) is asymptotically normally distributed with the bias* $\mathbf{B} = -(\mathbf{Q}+\mathbf{C})^{-1}\mathbf{C}\ \boldsymbol{\beta}^*$, *i.e.*

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) \xrightarrow{d} \mathcal{N}_p\left(\mathbf{B}, (\mathbf{Q}+\mathbf{C})^{-1}\frac{A^2(\varphi)}{\gamma^2(\varphi, f)}\right). \tag{3.4}$$

We shall prove Theorem 3.1 in several steps:

(1) Asymptotic representation of the linear rank statistic

$$\mathbf{S}_n(\mathbf{0},\mathbf{0}) = n^{-1/2}\sum_{i=1}^{n}(\mathbf{w}_i - \overline{\mathbf{w}}_n)a_n(R_i(\mathbf{0})) \tag{3.5}$$

with the sum of independent summands.

(2) Contiguity of the sequence $\{Q_n\}$ of distributions of $(e_i - (\mathbf{w}_i - \overline{\mathbf{w}})^\top \mathbf{b}_n^0 - (\mathbf{v}_i - \overline{\mathbf{v}})^\top \boldsymbol{\beta}_n)$, with $\mathbf{b}_n^0 = n^{-1/2}\mathbf{b}^{0*}$, $\boldsymbol{\beta}_n = n^{-1/2}\boldsymbol{\beta}^*$ for $\mathbf{b}^{0*}$, $\boldsymbol{\beta}^* \in \mathbb{R}^p$ fixed, with respect to the sequence $\{P_n\}$ of distributions of e_i, $i = 1, \ldots, n$.

(3) Asymptotic representation of the linear rank statistic (3.1) under contiguous sequence of distribution $\{Q_n\}$, and the resulting asymptotic linearity of (3.1) in parameters $\mathbf{b}^*, \boldsymbol{\beta}^*$.

(4) Uniform asymptotic quadraticity of $\mathcal{D}_n$ in parameters $\mathbf{b}^*, \boldsymbol{\beta}^*$ under $\{Q_n\}$, as a result of **(3)** and of the convexity of $\mathcal{D}_n$.

(5) Resulting asymptotic distribution and bias of $\widehat{\boldsymbol{\beta}}_n$.

We will start with the following lemma.

Lemma 3.1. *Under the conditions of Theorem 3.1, the statistic* $\mathbf{S}_n(\mathbf{0},\mathbf{0})$ *admits the asymptotic representation*

$$\mathbf{S}_n(\mathbf{0},\mathbf{0}) = \mathbf{Z}_n + o_p(1), \quad \textit{as} \quad n \to \infty, \tag{3.6}$$

where

$$\mathbf{Z}_n = n^{-1/2}\sum_{i=1}^{n}(\mathbf{w}_i - \overline{\mathbf{w}})\varphi(F(e_i)).$$

Proof. The proof is adapted from Picek (1996). If $\mathbf{b} = \boldsymbol{\beta} = \mathbf{0}$, then $(Y_1, \ldots, Y_n) = (e_1, \ldots, e_n)$. Let $R_1, \ldots, R_n$ denote their ranks. Further denote $r_i = a_n(R_i) - \varphi(F(e_i)),\ i = 1, \ldots, n$. Let σ_j^2 be the variance of $w_{i,j},\ i = 1, \ldots, n$, for $j = 1, \ldots, p$ and let $s^2 = \sum_{j=1}^{p} \sigma_j^2$.

Notice that $(r_1, \ldots, r_n)$ and $(\mathbf{w}_1, \ldots, \mathbf{w}_n)$ are independent. Consider the conditional squared distance

$$
\begin{aligned}
& \mathbb{E}_{\mathbf{G}} \left\{ (\mathbf{S}_n(\mathbf{0},\mathbf{0}) - \mathbf{Z}_n)^\top (\mathbf{S}_n(\mathbf{0},\mathbf{0}) - \mathbf{Z}_n) \middle| e_1, \ldots, e_n \right\} \\
= \ & n^{-1}\, \mathbb{E}_{\mathbf{G}} \left\{ \sum_{i=1}^{n} \sum_{k=1}^{n} (\mathbf{w}_i - \overline{\mathbf{w}})^\top (\mathbf{w}_k - \overline{\mathbf{w}}) r_i r_k \middle| e_1, \ldots, e_n \right\} \\
= \ & n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{n} r_i r_k\, \mathbb{E}_G \left\{ \sum_{j=1}^{p} (w_{i,j} - \bar{w}_j)(w_{k,j} - \bar{w}_j) \middle| e_1, \ldots, e_n \right\} \\
= \ & n^{-1} \left\{ \sum_{i=1}^{n} \sum_{k=1}^{n} r_i r_k \sum_{j=1}^{p} (x_{i,j} - \bar{x}_j)(x_{k,j} - \bar{x}_j) + s^2 \sum_{i=1}^{n} (r_i - \bar{r})^2 \right\} \\
= \ & \sum_{j=1}^{p} \left[n^{-1/2} \sum_{i=1}^{n} (x_{i,j} - \bar{x}_j) r_i \right]^2 + s^2 \sum_{i=1}^{n} (r_i - \bar{r})^2.
\end{aligned}
$$

Then (3.6) follows from Hájek and Šidák (1967, Theorems V.1.4.a,b, V.1.6.a). □

Definition. *For any two probability measures P and Q, absolutely continuous with respect to a σ-finite measure μ with $p = dP/d\mu$, $q = dQ/d\mu$, let*

$$
H(P,Q) = \left[\int (\sqrt{p} - \sqrt{q})^2\, d\mu \right]^{1/2} = \left[2 \int (1 - \sqrt{pq})\, d\mu \right]^{1/2}
$$

denote the the Hellinger distance between P and Q.

Definition. *Let $\{P_{n,1}, \ldots, P_{n,n}\}$ and $\{Q_{n,1}, \ldots, Q_{n,n}\}$ be two triangular arrays of probability measures, $P_{n,i}$ and $Q_{n,i}$ defined on measurable space $(\mathcal{X}_{n,i}, \mathcal{A}_{n,i})$ with densities $p_{n,i}, q_{n,i}$ with respect to σ-finite measures μ_i. Denote $P_n^{(n)} = \prod_{i=1}^{n} P_{n,i}$ and $Q_n^{(n)} = \prod_{i=1}^{n} Q_{n,i}$ the product measures, $n = 1, 2, \ldots$. The sequence $\{Q_n^{(n)}\}$ is said to be contiguous with respect to the sequence $\{P_n^{(n)}\}$ if*

$$
\lim_{n\to\infty} P_n^{(n)}(A_n) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} Q_n^{(n)}(A_n) = 0
$$

for any sequence of measurable sets A_n.

The previous definition of contiguity of two sequences is difficult to verify. Anyway, Oosterhoff and van Zwet (1979) proved that it is sufficient to verify more convenient conditions.

Lemma 3.2. *The sequence $\{Q_n^{(n)}\}$ is contiguous with respect to $\{P_n^{(n)}\}$ if and only if*

$$
\limsup_{n\to\infty} \sum_{i=1}^{n} H^2(P_{n,i}, Q_{n,i}) < \infty, \tag{3.7}
$$

$$
\lim_{n\to\infty} \sum_{i=1}^{n} Q_{n,i} \left\{ \frac{q_{n,i}(X_{n,i})}{p_{n,i}(X_{n,i})} \geq c_n \right\} = 0, \quad \forall\ c_n \to \infty, \tag{3.8}
$$

where $X_{n,i}$ is the identity map from $(\mathcal{X}_{n,i}, \mathcal{A}_{n,i})$ onto $(\mathcal{X}_{n,i}, \mathcal{A}_{n,i})$.

Note that in the case $P_{n,i} \equiv P_n, p_{n,i} \equiv p_n$, and $Q_{n,i} \equiv Q_n, q_{n,i} \equiv q_n$, not depending on i,

$$\sum_{i=1}^{n} H^2(P_{n,i}, Q_{n,i}) = n \int \left[\sqrt{q_n(z)} - \sqrt{p_n(z)}\right]^2 dz$$

$$= n \int \frac{(q_n(z) - p_n(z))^2}{[\sqrt{q_n(z)} + \sqrt{p_n(z)}]^2} dz \leq n \int \frac{\big(q_n(z) - p_n(z)\big)^2}{p_n(z)} dz. \tag{3.9}$$

Moreover, for $c_n > 1$ and with $d_n = c_n - 1$,

$$\sum_{i=1}^{n} Q_{n,i} \left\{ \frac{q_{n,i}(X_{n,i})}{p_{n,i}(X_{n,i})} \geq c_n \right\} = nQ_n \left\{ \frac{q_n(X_{n,1}) - p_n(X_{n,1})}{p_n(X_{n,1})} \geq d_n \right\}$$

$$\leq d_n^{-2} n \int \frac{\big|q_n(x) - p_n(x)\big|^2}{p_n^2(x)} q_n(x) dx$$

$$\leq d_n^{-2} n \int \frac{\big|q_n(x) - p_n(x)\big|^3}{p_n^2(x)} dx + d_n^{-2} n \int \frac{\big|q_n(x) - p_n(x)\big|^2}{p_n(x)} dx. \tag{3.10}$$

Now, get back to the residuals (here add the subscript n to $\mathbf{b}^0$ and $\boldsymbol{\beta}$)

$$\begin{aligned} Y_i - (\mathbf{w}_i - \overline{\mathbf{w}})^\top \mathbf{b}_n &= e_i + (\mathbf{x}_i - \overline{\mathbf{x}})^\top \boldsymbol{\beta}_n - (\mathbf{w}_i - \overline{\mathbf{w}})^\top \mathbf{b}_n \\ &= e_i - (\mathbf{w}_i - \overline{\mathbf{w}})^\top \mathbf{b}_n^0 - (\mathbf{v}_i - \overline{\mathbf{v}})^\top \boldsymbol{\beta}_n, \end{aligned}$$

where $\mathbf{b}_n = n^{-1/2}\mathbf{b}^*$, $\boldsymbol{\beta}_n = n^{-1/2}\boldsymbol{\beta}^*$, $\mathbf{b}_n^0 = n^{-1/2}\mathbf{b}^{0*}$, $\mathbf{b}^{0*} = \mathbf{b}^* - \boldsymbol{\beta}^*$, with fixed $\mathbf{b}^*$, $\boldsymbol{\beta}^* \in \mathbb{R}^p$ and with the aid of Lemma 3.2 we shall prove the following lemma.

Lemma 3.3. *Under the conditions of Theorem 3.1, the sequence $\{Q_n^{(n)}\}$ is contiguous with respect to $\{P_n^{(n)}\}$, where $Q_n^{(n)} = \prod_{i=1}^n Q_{n,i}$, $P_n^{(n)} = \prod_{i=1}^n P_{n,i}$, where $P_{n,i}$ is the distribution of e_i and $Q_{n,i}$ is the distribution of $(e_i - (\mathbf{w}_i - \overline{\mathbf{w}})^\top \mathbf{b}_n^0 - (\mathbf{v}_i - \overline{\mathbf{v}})^\top \boldsymbol{\beta}_n)$, $i = 1, \ldots, n$.*

Proof. We shall distinguish the two cases: the $\mathbf{x}_i$ are either i.i.d. random vectors or nonrandom vector components.

We start with the first case, where $\mathbf{w}_1, \ldots, \mathbf{w}_n$ are i.i.d. random vectors. Note that

$$U_i = (\mathbf{w}_i - \overline{\mathbf{w}})^\top \mathbf{b}^{0*} + (\mathbf{v}_i - \overline{\mathbf{v}})^\top \boldsymbol{\beta}^*, \; i = 1, \ldots, n,$$

are i.i.d. random variables. Let k_1 denote the common density of U_i. Then, $Q_{n,i}, P_{n,i}$ do not depend on i and $q_n(x) \equiv \int f(x - n^{-1/2}u)k_1(u)du$, $p_n(x) \equiv f(x)$.

Hence, by the Cauchy-Shwarz inequality and the Fubini Theorem, $\forall\, n \geq 1$

$$\begin{aligned} n \int \frac{\big(q_n(x) - p_n(x)\big)^2}{p_n(x)} dx &= n \int \Big\{ \int \left[f(x - n^{-1/2}u) - f(x) \right] k_1(u) du \Big\}^2 \frac{dx}{f(x)} \\ &\leq n \int \int \left[f(x - n^{-1/2}u) - f(x) \right]^2 \frac{k_1(u)}{f(x)} du dx \\ &\leq n \int \int \Big[\int_{-n^{-1/2}}^{n^{-1/2}} |u\, f'(x - tu)| dt \Big]^2 \frac{k_1(u)}{f(x)} du dx \\ &\leq 2n^{1/2} \int \int \int_{-n^{-1/2}}^{n^{-1/2}} |f'(x - tu)|^2 dt\, u^2 \frac{k_1(u)}{f(x)} du dx \\ &\leq 2n^{1/2} \int \int_{-n^{-1/2}}^{n^{-1/2}} \int \frac{|f'(x - tu)|^2}{f(x)} dx\, u^2 k_1(u) du\, dt. \end{aligned}$$

Hence, by (3.9), (**F.2**) applied with $j = 2$, and by (**V.2**), which guaranteed $\int u^2 k_1(u) du < \infty$,

$$\limsup_{n\to\infty} \sum_{i=1}^n H^2(P_{n,i}, Q_{n,i}) \leq 2I(f) \int u^2 k_1(u) du < \infty. \tag{3.11}$$

Similarly, $\forall\, n \geq 1$ the bound

$$n \int \frac{\big(q_n(x) - p_n(x)\big)^3}{p_n^2(x)} dx \;\leq\; 2n^{1/2} \int \int_{-n^{-1/2}}^{n^{-1/2}} \int \frac{|f'(x - tu)|^3}{f^2(x)} dx\, |u|^3 k_1(u) du\, dt$$

together with (3.10), (**F.2**) applied with $j = 3$, and (**V.2**), which guaranteed $\int |u|^3 k_1(u) du < \infty$, yield

$$\begin{aligned} &\lim_{n\to\infty} \sum_{i=1}^n Q_{n,i} \left\{ \frac{q_{n,i}(Y_i)}{p_{n,i}(Y_i)} \geq c_n \right\} \\ &\leq 2 \lim_{n\to\infty} d_n^{-2} \Big\{ \int \Big(\frac{|f'(x)|}{f(x)} \Big)^3 f(x) dx \int |u|^3 k_1(u) du + I(f) \int u^2 k_1(u) du \Big\} = 0. \end{aligned}$$

This ensures the validity of (3.8), and completes the proof of the contiguity in present case.

Next, consider the case where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are nonrandom, and we observe $\mathbf{w}_i = \mathbf{x}_i + \mathbf{v}_i,\ i = 1, \ldots, n$. Let k_2 denote the density of $(\mathbf{v}_i - \overline{\mathbf{v}})^\top \mathbf{b}^0,\ i = 1, \ldots, n$. Again, to prove (3.7) we have

$$\begin{aligned} &\sum_{i=1}^n H^2(P_{n,i}, Q_{n,i}) \\ \leq\;& \sum_{i=1}^n \int \Big\{ \int \left[f(e - n^{-1/2}u) - f(e) \right] k_2(u + (\mathbf{x}_i - \overline{\mathbf{x}})^\top \mathbf{b}^{0*}) du \Big\}^2 \frac{de}{f(e)} \\ \leq\;& \sum_{i=1}^n \int \Big\{ \int \left[f(e - n^{-1/2}u) - f(e) \right]^2 k_2 \left(u - (\mathbf{x}_i - \overline{\mathbf{x}})^\top \mathbf{b}_0^* \right) du \Big\} \frac{de}{f(e)} \\ \leq\;& 2n^{1/2} \int \int_{-n^{-1/2}}^{n^{-1/2}} \int \frac{|f'(e - tu)|^2}{f(e)} de\, dt\, n^{-1} \sum_{i=1}^n u^2\, k_2 \left(u - (\mathbf{x}_i - \overline{\mathbf{x}})^\top \mathbf{b}^{0*} \right) du. \end{aligned}$$

Hence, by (F.2) and by the change of variable formula,

$$\limsup_{n\to\infty}\sum_{i=1}^{n} H^2(P_{n,i},Q_{n,i}) \le K\Big[\int u^2 k_2(u)du + \mathbf{b}^{0*\top}\mathbf{C}\mathbf{b}^{0*}\Big] < \infty.$$

Similarly one verifies (3.8) here. □

Lemmas 3.1 and 3.3 enable us to extend the approximation of the rank statistic $\mathbf{S}_n(\mathbf{b}_n^0,\boldsymbol{\beta}_n)$ by a sum of independent random variables under the contiguous sequence of distributions. Let

$$\mathbf{T}_n(\mathbf{b}_n^0,\boldsymbol{\beta}_n) = n^{-1/2}\sum_{i=1}^{n}(\mathbf{w}_i-\overline{\mathbf{w}})\varphi\left(F(e_i-(\mathbf{w}_i-\overline{\mathbf{w}})^\top\mathbf{b}_n^0-(\mathbf{v}_i-\overline{\mathbf{v}})^\top\boldsymbol{\beta}_n)\right).$$

We have the following corollary.

Corollary 3.1. *Under the assumptions of Theorem 3.1 and under* $\{Q_n^{(n)}\}$

$$\mathbf{S}_n(\mathbf{b}_n^0,\boldsymbol{\beta}_n) = n^{-1/2}\sum_{i=1}^{n}(\mathbf{w}_i-\overline{\mathbf{w}})a_n\left(R_i(\mathbf{b}_n)\right) = \mathbf{T}_n(\mathbf{b}_n^0,\boldsymbol{\beta}_n)+o_p(1). \tag{3.12}$$

Hence,

$$\mathbf{S}_n(\mathbf{b}_n^0,\boldsymbol{\beta}_n)-\mathbf{S}_n(\mathbf{0},\mathbf{0}) = \mathbf{T}_n(\mathbf{b}_n^0,\boldsymbol{\beta}_n)-\mathbf{T}_n(\mathbf{0},\mathbf{0})+o_p(1).$$

Lemma 3.4. *Under the assumptions of Theorem 3.1*

$$\|\mathbf{S}_n(\mathbf{b}_n^0,\boldsymbol{\beta}_n)-\mathbf{S}_n(\mathbf{0},\mathbf{0})+\gamma(\varphi,f)[(\mathbf{Q}+\mathbf{C})\mathbf{b}^{0*}+\mathbf{C}\boldsymbol{\beta}^*]\| \xrightarrow{p} 0, \tag{3.13}$$

Proof. Consider the sequence of functions $\left\{\varphi^{(k)}(\cdot)\right\}_{k=1}^{\infty}$ defined on $(0,1)$

$$\varphi^{(k)}(u) = \varphi\left(\frac{1}{k+1}\right)\mathbb{I}\left[u<\frac{1}{k}\right]+\varphi(u)\mathbb{I}\left[\frac{i-1}{k+1}<u\le\frac{i}{k+1}\right],\quad i=2,\dots,k. \tag{3.14}$$

Then, by Hájek and Šidák (1967, Lemma V.1.6.a) $\varphi^{(k)}$ is nondecreasing and bounded on (0,1) and

$$\lim_{n\to\infty}\int_0^1[\varphi^{(k)}(u)-\varphi(u)]^2du = 0. \tag{3.15}$$

The function $\varphi^{(k)}$ has at most countable set B_k of discontinuity points. The convergence

$$F(e-n^{-1/2}(\mathbf{w}_i-\overline{\mathbf{w}})^\top\mathbf{b}^{0*}-n^{-1/2}(\mathbf{v}_i-\overline{\mathbf{v}})^\top\boldsymbol{\beta}^*)\to F(e)$$

is uniform in $e\in(-\infty,\infty)$ and $i=1,\dots,n,\ n\to\infty$. Hence the convergence

$$\lim_{n\to\infty}\varphi^{(k)}\left(F(e-n^{-1/2}(\mathbf{w}_i-\overline{\mathbf{w}})^\top\mathbf{b}^{0*}-n^{-1/2}(\mathbf{v}_i-\overline{\mathbf{v}})^\top\boldsymbol{\beta}^*\right) = \varphi^{(k)}\left(F(e)\right)$$

holds uniformly for $i=1,\dots,n$ almost surely with respect to F. It implies that the conditional expectation

$$\mathbb{E}\Big[\Big(\varphi^{(k)}\left(F(e_i-n^{-1/2}(\mathbf{w}_i-\overline{\mathbf{w}})^\top\mathbf{b}^{0*}-n^{-1/2}(\mathbf{v}_i-\overline{\mathbf{v}})^\top\boldsymbol{\beta}^*)\right)-\varphi^{(k)}\left(F(e_i)\right)\Big)^2\Big|\mathbf{v}_i,\mathbf{x}_i\Big]$$

converges to 0 as $n \to \infty$ uniformly for $\mathbf{v}_i, \mathbf{x}_i$, $i = 1, \ldots, n$ and k fixed.

Let $\mathbf{S}_n^{(k)}(\mathbf{b}^0, \boldsymbol{\beta})$ and $\mathbf{T}_n^{(k)}(\mathbf{b}^0, \boldsymbol{\beta})$ be analogous to $\mathbf{S}_n(\mathbf{b}^0, \boldsymbol{\beta})$, $\mathbf{T}_n(\mathbf{b}^0, \boldsymbol{\beta})$ respectively, with φ replaced with $\varphi^{(k)}$. Then we can bound the norm of the covariance matrix of $\mathbf{T}_n^{(k)}(\mathbf{b}^0, \boldsymbol{\beta}) - \mathbf{T}_n^{(k)}(\mathbf{0}, \mathbf{0})$ for any fixed k in the following way. Denote

$$\mathbf{A}_n^{(k)} = \mathbb{E}\left\{[\mathbf{T}_n^{(k)}(\mathbf{b}_n^0, \boldsymbol{\beta}_n) - \mathbf{T}_n^{(k)}(\mathbf{0}, \mathbf{0})][\mathbf{T}_n^{(k)}(\mathbf{b}_n^0, \boldsymbol{\beta}_n) - \mathbf{T}_n^{(k)}(\mathbf{0}, \mathbf{0})]^\top\right\}.$$

Then

$$\begin{aligned}
\mathbf{A}_n^{(k)} &= \mathbb{E}\Big\{ n^{-1} \sum_{i=1}^n (\mathbf{w}_i - \overline{\mathbf{w}})(\mathbf{w}_i - \overline{\mathbf{w}})^\top \cdot \\
&\cdot \left[\varphi^{(k)}\left(F(e_i - (\mathbf{w}_i - \overline{\mathbf{w}})^\top \mathbf{b}_n^0 - (\mathbf{v}_i - \overline{\mathbf{v}})^\top \boldsymbol{\beta}_n)\right) - \varphi^{(k)}(F(e_i))\right]^2 \Big\} \qquad (3.16) \\
&= n^{-1} \sum_{i=1}^n \mathbb{E}\Big\{ (\mathbf{w}_i - \overline{\mathbf{w}})(\mathbf{w}_i - \overline{\mathbf{w}})^\top \cdot \\
&\cdot \mathbb{E}\Big[\Big(\varphi^{(k)}\left(F(e_i - (\mathbf{w}_i - \overline{\mathbf{w}})^\top \mathbf{b}_n^0 - (\mathbf{v}_i - \overline{\mathbf{v}})^\top \boldsymbol{\beta}_n)\right) - \varphi^{(k)}\left(F(e_i)\right)\Big)^2 \Big| \mathbf{v}_i, \mathbf{x}_i\Big]\Big\}
\end{aligned}$$

Hence,

$$\begin{aligned}
\left\|\mathbf{A}_n^{(k)}\right\| &\le \left\{ \left\| n^{-1} \sum_{i=1}^n (\mathbf{w}_i - \overline{\mathbf{w}})(\mathbf{w}_i - \overline{\mathbf{w}})^\top - (\mathbf{Q} + \mathbf{C}) \right\| + \|\mathbf{Q} + \mathbf{C}\| \right\} \cdot o(1) \\
&= \{\|\mathbf{Q} + \mathbf{C}\| + o(1)\} \cdot o(1), \quad \text{as } n \to \infty.
\end{aligned}$$

This implies

$$\|\mathbf{T}_n^{(k)}(\mathbf{b}_n^0, \boldsymbol{\beta}_n) - \mathbf{T}_n^{(k)}(\mathbf{0}, \mathbf{0}) - \mathbb{E}\mathbf{T}_n^{(k)}(\mathbf{b}_n^0, \boldsymbol{\beta}_n)\| \xrightarrow{p} 0 \qquad (3.17)$$

when taking into account that $\mathbb{E}\mathbf{T}_n^{(k)}(\mathbf{0}, \mathbf{0}) = \mathbf{0}$.

Further, for any fixed k and for fixed $\mathbf{b}^{0*}, \boldsymbol{\beta}^*$,

$$\mathbf{T}_n^{(k)}(\mathbf{b}_n^0, \boldsymbol{\beta}_n) - \mathbf{T}_n^{(k)}(\mathbf{0}, \mathbf{0}) + \gamma_k[(\mathbf{Q} + \mathbf{C})\mathbf{b}^{0*} + \mathbf{C}\boldsymbol{\beta}^*] \xrightarrow{p} \mathbf{0}, \qquad (3.18)$$

where

$$\gamma_k = \gamma(\varphi^{(k)}, f) = -\int_0^1 \varphi^{(k)}(u) \frac{f'(F^{-1}(u))}{f(F^{-1}(u))} du.$$

Indeed, (we put $\overline{\mathbf{x}} = \overline{\mathbf{v}} = \mathbf{0}$, without loos of generality for the sake of brevity)

$$\begin{aligned}
& n^{-1/2} \sum_{i=1}^n \mathbb{E}\Big\{ \mathbf{w}_i \Big(\mathbb{E}\Big[\varphi^{(k)}(F(e_i - n^{-1/2}(\mathbf{w}_i^\top \mathbf{b}^{0*} - n^{-1/2}\mathbf{v}_i^\top \boldsymbol{\beta}^*) - \varphi^{(k)}(F(e_i)) \\
& -\gamma_k(n^{-1/2}[\mathbf{w}_i^\top \mathbf{b}^{0*} + \mathbf{v}_i^\top \boldsymbol{\beta}^*] \Big| \mathbf{v}_i, \mathbf{x}_i\Big]\Big)\Big\} \\
&= n^{-1/2} \sum_{i=1}^n \mathbb{E}\Big\{ \mathbf{w}_i \Big(\int_{-\infty}^{\infty} \varphi^{(k)}(F(z)) d[F(z + n^{-1/2}\mathbf{w}_i^\top \mathbf{b}^{0*} + n^{-1/2}\mathbf{v}_i^\top \boldsymbol{\beta}^*) - F(z)] \\
& -n^{-1/2}[\mathbf{w}_i^\top \mathbf{b}^{0*} + \mathbf{v}_i^\top \boldsymbol{\beta}^*] \int_{-\infty}^{\infty} \varphi^{(k)}(F(z)) f'(z) dz \Big)\Big\} \\
&= n^{-1/2} \sum_{i=1}^n \mathbb{E}\Big\{ \mathbf{w}_i \Big(\int_{-\infty}^{\infty} \varphi^{(k)}(F(z)) \\
& \cdot d\Big[F(z + n^{-1/2}\mathbf{w}_i^\top \mathbf{b}^{0*} + n^{-1/2}\mathbf{v}_i^\top \boldsymbol{\beta}^*) - F(z) - n^{-1/2}(\mathbf{w}_i^\top \mathbf{b}^{0*} + \mathbf{v}_i^\top \boldsymbol{\beta}^*) f(z)\Big]\Big)\Big\},
\end{aligned}$$

that converges to $\mathbf{0}$ as $n \to \infty$. Moreover, we have

$$(\gamma(\varphi^{(k)}, f) - \gamma(\varphi, f))^2 = \left\langle (\varphi^{(k)} - \varphi), -\frac{f'(F^{-1}(\cdot))}{f(F^{-1}(\cdot))} \right\rangle^2 \tag{3.19}$$
$$\leq \|\varphi^{(k)} - \varphi\|^2 \left\| -\frac{f'(F^{-1}(\cdot))}{f(F^{-1}(\cdot))} \right\|^2 = I(f)\ \|\varphi^{(k)} - \varphi\|^2 \to 0 \ \text{ as } \ k \to \infty.$$

Using (3.18), (3.19), Lemma 3.1, Lemma 3.3, Corollary 3.1 and Lemma 3.5 in Jurečková (1969), we obtain that

$$P\left(\|\mathbf{S}_n(\mathbf{b}_n^0, \boldsymbol{\beta}_n) - \mathbf{S}_n^{(k)}(\mathbf{b}_n^0, \boldsymbol{\beta}_n)\| > \varepsilon\right) < \varepsilon,$$

$\forall \varepsilon > 0,\ \forall k > k(\varepsilon),\ \forall n > n(k)$ and finally we arrive at (3.13). □

Now, rewrite the Jaeckel dispersion

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^{n} \left(Y_i - \mathbf{w}_i^\top \mathbf{b}\right) a_n(R_i(\mathbf{b}))$$

in the presence of measurement errors in the form

$$\mathcal{D}_n(\mathbf{b^0}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left[e_i - (\mathbf{w}_i - \overline{\mathbf{w}})^\top \mathbf{b}^0 - (\mathbf{v}_i - \overline{\mathbf{v}})^\top \boldsymbol{\beta}\right] a_n(R_i(\mathbf{b}^0 + \boldsymbol{\beta})). \tag{3.20}$$

By Jaeckel (1972), the partial derivatives of $\mathcal{D}_n(\mathbf{b})$ exist for almost all $\mathbf{b}$, and where they exist, are equal to

$$\frac{\partial}{\partial b_j} \mathcal{D}_n(\mathbf{b}) = -n^{1/2} S_{n,j}(\mathbf{b}) = -\sum_{i=1}^{n} (w_{i,j} - \bar{w}_j) a_n(R_i(\mathbf{b})),\ j = 1, \ldots, p.$$

Otherwise speaking,

$$\nabla \mathcal{D}_n(\mathbf{b}) = -n^{1/2} \mathbf{S}_n(\mathbf{b}) = -\sum_{i=1}^{n} (\mathbf{w}_i - \overline{\mathbf{w}}) a_n(R_i(\mathbf{b})),$$
$$\nabla \mathcal{D}_n(\mathbf{b}^0, \boldsymbol{\beta}) = -n^{1/2} \mathbf{S}_n(\mathbf{b}^0, \boldsymbol{\beta}) = -\sum_{i=1}^{n} (\mathbf{w}_i - \overline{\mathbf{w}}) a_n(R_i(\mathbf{b}^0 + \boldsymbol{\beta})),$$

where ∇ denotes the subgradient.

Consider the quadratic function

$$\mathcal{C}_n(\mathbf{b}) = \mathcal{C}_n(\mathbf{b}^0, \boldsymbol{\beta}) = \frac{1}{2} \gamma(\varphi, f) \mathbf{b}^{0\top} (\mathbf{Q} + \mathbf{C}) \mathbf{b}^0 - \mathbf{b}^{0\top} \mathbf{S}_n(\mathbf{0}) + \gamma(\varphi, f) \mathbf{b}^0 \mathbf{C} \boldsymbol{\beta} + \mathcal{D}_n(\mathbf{0}). \tag{3.21}$$

Then $\mathcal{D}_n(\mathbf{b})$ and $\mathcal{C}_n(\mathbf{b})$ are both convex functions and $\mathcal{D}_n(\mathbf{0}) = \mathcal{C}_n(\mathbf{0})$. Moreover, for $\mathbf{b}_n^0 = n^{-1/2} \mathbf{b}^{0*},\ \boldsymbol{\beta}_n = n^{-1/2} \boldsymbol{\beta}^*$ we have

$$\nabla[\mathcal{D}_n(n^{-1/2}\mathbf{b}^{0*}, n^{-1/2}\boldsymbol{\beta}^*) - \mathcal{C}_n(n^{-1/2}\mathbf{b}^{0*}, n^{-1/2}\boldsymbol{\beta}^*)]$$
$$= \mathbf{S}_n(\mathbf{b}_n^0, \boldsymbol{\beta}_n) - \mathbf{S}_n(\mathbf{0}, \mathbf{0}) + \gamma(\varphi, f)(\mathbf{Q} + \mathbf{C})\mathbf{b}^{0*} + \gamma(\varphi, f)\mathbf{C}\boldsymbol{\beta}^*.$$

Hence it follows from (3.13) for $\mathbf{b}^{0*}, \boldsymbol{\beta}^* \in \mathbb{R}^p$ fixed that

$$\|\nabla[\mathcal{D}_n(n^{-1/2}\mathbf{b}^{0*}, n^{-1/2}\boldsymbol{\beta}^*) - \mathcal{C}_n(n^{-1/2}\mathbf{b}^{0*}, n^{-1/2}\boldsymbol{\beta}^*)]\| \xrightarrow{p} 0.$$

Using the convexity arguments in Appendix of Heiler and Willers (1988) and Convexity lemma in Pollard (1991), we conclude that $\forall K > 0$

$$\sup\Big\{\Big|\mathcal{D}_n(n^{-1/2}\mathbf{b}^{0*}, n^{-1/2}\boldsymbol{\beta}^*) - \frac{1}{2}\gamma(\varphi, f)\mathbf{b}^{0*\top}(\mathbf{Q}+\mathbf{C})\mathbf{b}^{0*} + \mathbf{b}^{0*\top}\mathbf{S}_n(\mathbf{0})$$
$$-\gamma(\varphi, f)\mathbf{b}^{0*}\mathbf{C}\boldsymbol{\beta}^* + \mathcal{D}_n(\mathbf{0})\Big| : \|\mathbf{b}^{0*}\| \le K,\ \|\boldsymbol{\beta}^*\| \le K\Big\} = o_p(1) \text{ as } n \to \infty.$$

Hence, following the arguments in the proof of Theorem 1 in Pollard (1991), we conclude that, under the local alternative $\boldsymbol{\beta}_n = n^{-1/2}\boldsymbol{\beta}^*$,

$$\arg\min_{\mathbf{b}^{0*}} \mathcal{D}_n(n^{-1/2}\mathbf{b}^{0*}, n^{-1/2}\boldsymbol{\beta}^*)$$

is asymptotically equivalent to

$$\arg\min_{\mathbf{b}^{0*}} \left[\frac{1}{2}\gamma(\varphi, f)\mathbf{b}^{0*\top}(\mathbf{Q}+\mathbf{C})\mathbf{b}^{0*} - \mathbf{b}^{0*\top}\mathbf{S}_n(\mathbf{0}) + \gamma(\varphi, f)\mathbf{b}^{0*}\mathbf{C}\boldsymbol{\beta}^*\right] \quad (3.22)$$

The minimizer of (3.22) equals to

$$\mathbf{b}^{0*} = \mathbf{b}^* - \boldsymbol{\beta}^* = n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) = \gamma^{-1}(\varphi, f)(\mathbf{Q}+\mathbf{C})^{-1}\mathbf{S}_n(\mathbf{0}) - (\mathbf{Q}+\mathbf{C})^{-1}\mathbf{C}\boldsymbol{\beta}^*.$$

Hence, in the linear model with local value of regression parameter β, when

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_n + e_i,\ \boldsymbol{\beta}_n = n^{-1/2}\boldsymbol{\beta}^*,$$

when we observe only $\mathbf{w}_i = \mathbf{x}_i + \mathbf{v}_i$ instead of $\mathbf{x}_i,\ i = 1, \dots, n$, the R-estimator is asymptotically normally distributed with a bias $\mathbf{B} = -(\mathbf{Q}+\mathbf{C})^{-1}\mathbf{C}\ \boldsymbol{\beta}^*$, i.e.

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_n - n^{-1/2}\boldsymbol{\beta}^*) \xrightarrow{d} \mathcal{N}_p\left(\mathbf{B}, (\mathbf{Q}+\mathbf{C})^{-1}\frac{A^2(\varphi)}{\gamma^2(\varphi, f)}\right).$$

This completes the proof of Theorem 3.1.

3.3 Generalization for errors in response variables

The previous result may be generalized for model where there are also additive measurement errors in response variables:

$$\begin{aligned} Y_i &= \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + e_i, \\ \mathbf{w}_i &= \mathbf{x}_i + \mathbf{v}_i, \quad i = 1, \dots, n, \\ Z_i &= Y_i + W_i, \end{aligned} \quad (3.23)$$

where W_i are i.i.d. random variables independent with e_i (and $\mathbf{x}_i$ if they are random) with (unknown) distribution function H and density h.

Let m, M be the density and distribution function of $e_i + W_i$, $i = 1, \dots, n$, i.e. $m(z) = \int f(z-t)h(t)dt$. The density is absolutely continuous and has finite Fisher information $I(m)$ (see Lemma 1.2). Notice that if we observe $Z_i = Y_i + W_i$ instead of Y_i, then $e_i^* = e_i + W_i$, $i = 1, \dots, n$ are still i.i.d. random variables (new model errors) with density m. Hence the measurement errors in responses Y_i affect only the asymptotic variance, not the bias, as stated in the following corollary of Theorem 3.1.

Corollary 3.2. *Under the conditions* **A.1**, **F.1** $-$ **F.2**, **V.1** $-$ **V.3**, **X.1** $-$ **X.2** *and under the local alternative (3.3), the R-estimator* $\widehat{\boldsymbol{\beta}}_n$ *in model (3.23) is asymptotically normally distributed with the bias* $\mathbf{B} = -(\mathbf{Q}+\mathbf{C})^{-1}\mathbf{C}\,\boldsymbol{\beta}^*$, *i.e.*

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) \xrightarrow{d} \mathcal{N}_p\left(\mathbf{B}, (\mathbf{Q}+\mathbf{C})^{-1}\frac{A^2(\varphi)}{\gamma^2(\varphi, m)}\right). \tag{3.24}$$

3.4 Numerical illustration

The following simulation study illustrates the effect of measurement errors in regressors on the finite-sample performance of R-estimates. Empirical bias (and variance) of R-estimates are computed and compared for various measurement error models. For the sake of comparison, the bias and variance are computed for the least squares estimate (LSE) and 0.5 - regression quantiles (L_1-estimates) under the same setup.

The results illustrate that the bias of R-estimate is surprisingly stable with respect to the sample size; the bias corresponding to small n is comparable to the asymptotic one derived in Theorem 3.1. Not only that the bias of R-estimator depends neither on the choice of the rank score function nor on the distribution of the model errors; generally it only slightly differs from the bias of LSE and L_1-estimators (if the expectation of the estimator exists). Moreover, the deterministic and random regressors are compared.

3.4.1 Regression line

Consider first the model of regression line

$$Y_i = \beta_0 + x_i\beta + e_i, \quad i = 1, \dots, n,$$

where the Y_i are measured accurately, while instead of x_i we observe only $w_i = x_i + v_i, i = 1, \dots, n$. The R-estimator of parameter β is based on Wilcoxon scores generated by score function $\varphi(u) = u - 1/2$.

All the simulation results are based on 10 000 replications, parameters were chosen as $\beta_0 = 1, \beta = 2$, and model errors e_i follow the logistic distribution. In Tables 3.1 and 3.2 the empirical bias of R-estimator based on Wilcoxon scores is compared for various sample sizes ($n = 10, \dots, 1000$) and with the theoretical asymptotic result ($n = \infty$). The regressors x_i are deterministic in Table 3.1; they were generated from uniform $\mathcal{U}(-3, 9)$ distribution once for all experiment and then considered as fixed. The regressors in Table 3.2 are random; each time they were generated also from uniform distribution $\mathcal{U}(-3,9)$. This enables to see the difference between deterministic and random regressors: The bias differs more from its asymptotic value in case of deterministic regressors than in case of random regressors; it can be caused by the slower rate of convergence. The measurement errors v_i are either uniformly or normally distributed.

$v_i \backslash n$	10	20	50	100	200	500	1000	∞
0	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
$\mathcal{U}(-5,0)$	−0.264	−0.295	−0.305	−0.297	−0.302	−0.306	−0.307	−0.296
$\mathcal{U}(0,9)$	−0.684	−0.727	−0.732	−0.714	−0.719	−0.727	−0.728	−0.720
$\mathcal{U}(-3,9)$	−0.982	−1.013	−1.006	−0.983	−0.986	−0.995	−0.995	−1.000
$\mathcal{N}(0,1)$	−0.128	−0.148	−0.150	−0.146	−0.148	−0.151	−0.152	−0.154
$\mathcal{N}(0,2)$	−0.440	−0.483	−0.488	−0.476	−0.480	−0.487	−0.488	−0.500
$\mathcal{N}(0,3)$	−0.790	−0.836	−0.837	−0.819	−0.822	−0.832	−0.833	−0.857

Table 3.1: Empirical bias of R-estimator for various n and measurement errors v_i; nonrandom regressors x_i.

$v_i \backslash n$	10	20	50	100	200	500	1000	∞
0	0.004	−0.001	0.000	0.000	0.000	0.000	0.000	0.000
$\mathcal{U}(-5,0)$	−0.283	−0.297	−0.305	−0.306	−0.307	−0.309	−0.309	−0.296
$\mathcal{U}(0,9)$	−0.711	−0.722	−0.728	−0.730	−0.730	−0.732	−0.732	−0.720
$\mathcal{U}(-3,9)$	−0.998	−1.000	−0.999	−1.000	−0.999	−1.000	−1.000	−1.000
$\mathcal{N}(0,1)$	−0.138	−0.149	−0.150	−0.153	−0.153	−0.153	−0.153	−0.154
$\mathcal{N}(0,2)$	−0.462	−0.481	−0.487	−0.489	−0.491	−0.492	−0.492	−0.500
$\mathcal{N}(0,3)$	−0.813	−0.830	−0.833	−0.835	−0.837	−0.837	−0.838	−0.857

Table 3.2: Empirical bias of R-estimator for various n and measurement errors v_i; random regressors x_i.

Table 3.3 compares empirical bias and variance (in parenthesis) of R-estimator based on Wilcoxon scores, of LSE and L_1-estimator under sample size $n = 50$ and under random regressors x_i generated from uniform $\mathcal{U}(-3,9)$ distribution; model errors e_i generated from normal, logistic, Laplace, Pareto with parameter $\alpha = 0.9$ and Cauchy distributions. The measurement errors v_i follow various distributions, similarly as in Tables 3.1 and 3.2.

3.4.2 Model of two regressors

Consider the model

$$Y_i = \beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 + e_i, \quad i = 1,\dots,n,$$

where again the Y_i are measured accurately, but instead of $\mathbf{x}_i$ we observe only $\mathbf{w}_i = \mathbf{x}_i + \mathbf{v}_i, i = 1,\dots,n$. The R-estimator of parameter $\boldsymbol{\beta} = (\beta_1,\beta_2)^\top$ is based on Wilcoxon scores generated by score function $\varphi(u) = u - 1/2$.

$v_i \setminus e_i$	$\mathcal{N}(0,1)$	$Log(0,1)$	$Lap(0,1)$	$\mathcal{P}(0.9,0)$	$\mathcal{C}(0,1)$
	0.002(0.182)	0.008(0.527)	−0.002(0.254)	0.000(0.416)	0.018(0.672)
0	0.004(0.172)	0.009(0.567)	−0.003(0.355)	5.047(89200)	−3.289(85700)
	0.002(0.274)	0.010(0.708)	−0.002(0.249)	0.000(1.191)	0.018(0.526)
	−0.399(0.155)	−0.398(0.438)	−0.396(0.214)	−0.401(0.422)	−0.404(0.568)
$\mathcal{U}(-3,3)$	−0.395(0.147)	−0.396(0.466)	−0.394(0.283)	−7.137(604000)	22.62(4000000)
	−0.401(0.232)	−0.400(0.591)	−0.400(0.235)	−0.422(0.932)	−0.405(0.456)
	−0.995(0.101)	−1.006(0.278)	−0.997(0.142)	−1.001(0.309)	−1.010(0.397)
$\mathcal{U}(-6,6)$	−0.995(0.096)	−1.009(0.294)	−0.998(0.182)	−7.259(401000)	0.933(36400)
	−0.995(0.151)	−1.006(0.376)	−0.995(0.157)	−1.001(0.587)	−1.014(0.320)
	−0.153(0.174)	−0.161(0.493)	−0.145(0.243)	−0.149(0.439)	−0.147(0.638)
$\mathcal{N}(0,1)$	−0.152(0.163)	−0.158(0.523)	−0.145(0.328)	−2.136(281000)	−7.380(739000)
	−0.153(0.261)	−0.159(0.675)	−0.147(0.259)	−0.165(1.092)	−0.138(0.510)

Table 3.3: Empirical bias (variance) of R-estimator, LSE and L_1-estimator for various measurement errors v_i and model errors e_i; $n = 50$.

Let $n = 50$, parameters $\beta_0 = 1, \beta_1 = 2, \beta_2 = 1$, random regressors $\mathbf{x}_i = (x_{i,1}, x_{i,2})^\top$ were generated from 2-dimensional normal distribution $\mathcal{N}_2(\boldsymbol{\mu}, \mathbf{S}_1)$ and $\mathbf{v}_i = (v_{i,1}, v_{i,2})^\top$ were generated from distributions $\mathcal{N}_2(\boldsymbol{\mu}, \mathbf{S}_\nu)$, $\nu = 1, 2, 3$, where $\boldsymbol{\mu} = (0, 1)^\top$ and

$$\mathbf{S}_1 = \begin{pmatrix} 4 & 0.5 \\ 0.5 & 2 \end{pmatrix}, \quad \mathbf{S}_2 = \begin{pmatrix} 2 & 0.2 \\ 0.2 & 2 \end{pmatrix}, \quad \mathbf{S}_3 = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}.$$

Table 3.4 compares empirical bias and variance (in parenthesis) of R-estimator based on Wilcoxon scores, of the least squares estimator and of L_1-estimator for various measurement errors $\mathbf{v}_i$ and model errors e_i.

We have also computed R-estimates generated by other score functions, e.g. van der Waerden, median; also another simulation design was considered – various sample sizes n, values of the parameters, distributions of regressors, measurement errors $\mathbf{v}_i$ and model errors. It is of interest that the results for corresponding R-estimates are quite similar to those presented in the previous tables.

The simulation study confirms that R-estimates in measurement error models are biased, as well as other usual estimates. The bias is relatively stable with respect to the sample size and to distribution of model errors. The R-estimates provide meaningful results as long as the e_i have a finite Fisher information; even under the normal errors are their empirical variances only slightly greater than that of LSE. The bias and other properties of R-estimates are comparable with those of the least squares and of L_1-estimates unless the distribution of model errors e_i is heavy-tailed, where the LSE fails. Generally, the reduction of the bias is rather a matter of measurement precision, of calibration and repeated measurements. See Jurečková et al. (submitted) for more details and simulation results.

$\mathbf{v}_i \backslash e_i$		$\mathcal{N}(0,1)$	$Log(0,1)$	$Lap(0,1)$	$\mathcal{P}(0.9,0)$	$\mathcal{C}(0,1)$
$\mathbf{0}$	$\widehat{\beta}_1$	0.000(0.600)	−0.015(1.688)	0.002(0.821)	−0.002(0.017)	0.026(2.326)
		−0.001(0.569)	−0.019(1.789)	0.006(1.120)	21.73(5330000)	−3.698(48700)
		−0.007(0.864)	−0.008(2.295)	0.002(0.857)	−0.003(0.034)	0.035(1.843)
	$\widehat{\beta}_2$	0.020(1.176)	0.026(3.497)	0.001(1.678)	−0.003(0.033)	0.037(4.634)
		0.014(1.117)	0.027(3.725)	−0.001(2.250)	−29.28(9770000)	0.695(66200)
		0.031(1.744)	0.017(4.588)	−0.001(1.758)	−0.005(0.067)	0.007(3.618)
$\mathcal{N}_2(\boldsymbol{\mu}, \mathbf{S}_3)$	$\widehat{\beta}_1$	−0.362(0.554)	−0.379(1.603)	−0.379(0.798)	−0.372(0.087)	−0.347(2.161)
		−0.359(0.528)	−0.385(1.693)	−0.376(1.030)	22.81 (6270000)	−4.118(48500)
		−0.365(0.823)	−0.369(2.134)	−0.375(0.881)	−0.370(0.102)	−0.338(1.758)
	$\widehat{\beta}_2$	−0.774(0.936)	−0.738(2.662)	−0.754(1.295)	−0.770(0.136)	−0.738(3.618)
		−0.776(0.881)	−0.734(2.832)	−0.757(1.696)	−26.76(7920000)	−0.268(78500)
		−0.769(1.402)	−0.754(3.366)	−0.754(1.409)	−0.769(0.164)	−0.752(2.935)
$\mathcal{N}_2(\boldsymbol{\mu}, \mathbf{S}_2)$	$\widehat{\beta}_1$	−0.643(0.419)	−0.652(1.155)	−0.648(0.579)	−0.649(0.076)	−0.626(1.622)
		−0.640(0.399)	−0.655(1.216)	−0.653(0.750)	14.66(3070000)	−3.987(47200)
		−0.647(0.615)	−0.645(1.573)	−0.642(0.657)	−0.650(0.089)	−0.623(1.317)
	$\widehat{\beta}_2$	−0.495(0.643)	−0.474(1.730)	−0.477(0.866)	−0.494(0.107)	−0.466(2.426)
		−0.497(0.605)	−0.469(1.843)	−0.471(1.139)	−17.59(3660000)	−0.646(55900)
		−0.489(0.948)	−0.477(2.282)	−0.486(0.944)	−0.492(0.129)	−0.478(1.992)
$\mathcal{N}_2(\boldsymbol{\mu}, \mathbf{S}_1)$	$\widehat{\beta}_1$	−0.997(0.329)	−1.013(0.879)	−1.010(0.448)	−1.005(0.071)	−0.987(1.264)
		−0.999(0.311)	−1.015(0.931)	−1.011(0.577)	9.435(1260000)	−2.474(40500)
		−0.994(0.484)	−1.009(1.192)	−1.011(0.509)	−1.005(0.086)	−0.998(1.037)
	$\widehat{\beta}_2$	−0.505(0.670)	−0.493(1.797)	−0.496(0.917)	−0.499(0.141)	−0.482(2.489)
		−0.505(0.630)	−0.486(1.883)	−0.487(1.168)	−21.75(5680000)	0.106(49600)
		−0.501(0.980)	−0.520(2.374)	−0.504(1.024)	−0.500(0.170)	−0.501(2.064)

Table 3.4: Empirical bias (variance) of R-estimator, LSE and L_1-estimator for various measurement errors $\mathbf{v}_i$ and model errors e_i; $n = 50$.

4. Heteroscedasticity in linear models

In previous chapters we dealt with linear models where the regressors were measured only with an additive error. Now, we will consider different situation that might be also considered in a broader sense as some measurement error model (model where some classical assumptions for liner models are not satisfied).

Homoscedasticity is often tacitly assumed in the analysis of linear models, both classical and robust. To avoid a negative consequence of ignored heteroscedasticity, we should either analyze its possibility before starting an inference on the parameters of the model, or look for an approach invariant to heteroscedasticity, if there is one.

Testing the heteroscedasticity in linear model was studied by a host of authors. Koenker (1981) studied the efficiency of the Breusch and Pagan (1979) test, and extended it to more general distributions. Carroll and Ruppert (1981) constructed robust test for heteroscedasticity, extending test of Bickel (1978). Koenker and Bassett (1982) constructed a test of heteroscedasticity in linear model based on regression quantiles. Dette and Munk (1998) proposed a simple test for heteroscedasticity in nonparametric regression model based on an estimate of the best L^2-approximation of the variance function. Lyon and Tsai (1996) compared eight likelihood ratio and score tests for heteroscedasticity, mainly their sizes, powers and sensitivities to high leverage and outliers. Lin and Qu (2012) constructed a consistent test for heteroscedasticity for nonlinear semiparametric regression models with nonparametric variance function based on the kernel method. Lewbel (2012) used the heteroscedasticity to identify and estimate mismeasured and endogenous regression models.

4.1 Preliminaries

The usual heteroscedastic linear model has the form

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \sigma_i U_i, \quad i = 1, \ldots, n, \tag{4.1}$$

where $\beta_0 \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^p$ and $(\sigma_1, \ldots, \sigma_n)^\top \in \mathbb{R}_+^n$ are unknown parameters, and $U_1, \ldots, U_n$ are the i.i.d. errors with (unknown) distribution function F.

To test for significance of $\boldsymbol{\beta}$ or for heteroscedasticity in model (4.1), we should assume some structure of $\sigma_i, \ i = 1, \ldots, n$, as an alternative to the homoscedasticity. Various scale models were considered in the literature. Ali and Giaccotto (1984) considered the model (4.1) with $\sigma_1 = \ldots = \sigma_n = \sigma$ and

$$\begin{aligned}(\sigma U_i)^2 &= 1 + \theta g(\mathbf{z}_i^T \mathbf{a}) + W_i, \\ \sigma U_i &= [\exp\{\theta g(\mathbf{z}_i^T \mathbf{a})\}] W_i, \quad i = 1, \ldots, n\end{aligned}$$

with $(\mathbf{z}_i, \ \mathbf{a})$ known, unknown parameters $\sigma, \ \theta$, and random $W_i, \ i = 1, \ldots, n$. In this model they studied rank tests of $\mathbf{H}: \ \theta = 0$, with eventual estimation of unknown entities. The heteroscedasticity is often modeled as regression in scale in the form

$$\sigma_i = \exp\{\mathbf{z}_i^\top \boldsymbol{\gamma}\}, \quad i = 1, \ldots, n \tag{4.2}$$

with known or observable $\mathbf{z}_i \in \mathbb{R}^q$, $i = 1, \ldots, n$ and unknown parameter $\boldsymbol{\gamma} \in \mathbb{R}^q$. Such model was considered by Akritas and Albers (1993), who constructed an aligned rank test on some components of parameter $\boldsymbol{\beta}$, with $\boldsymbol{\gamma}$ replaced with a suitable estimator (further not specified). Gutenbrunner (1994) considered testing homoscedasticity $\mathbf{H}: \ \boldsymbol{\gamma} = \mathbf{0}$ when $\mathbf{z}_i$ in (4.2) is partitioned as $\mathbf{z}_i^\top = (1, \mathbf{x}_i^\top, \boldsymbol{\xi}_i^\top)^\top$, with $\mathbf{x}_i$ from (4.1) and $\boldsymbol{\xi}_i$ an external vector, $i = 1, \ldots, n$. His test was based on a combination of regression rank scores for the $\boldsymbol{\xi}_i$ and regression quantile estimator of $\boldsymbol{\beta}$. This test was then modified in Gutenbrunner et al. (1995); see also the review paper by Koenker (1997). Dixon and McKean (1996) considered an estimation problem in model (4.1) with $\sigma_i = \exp\{\theta h(\mathbf{x}_i^\top \boldsymbol{\beta})\}, \ i = 1, \ldots, n$ with a known function h; they estimated $\boldsymbol{\beta}$ and θ iteratively by means of suitable R-estimates.

In testing either for $\boldsymbol{\beta}$ with nuisance $\boldsymbol{\gamma}$ or for $\boldsymbol{\gamma}$ with nuisance $\boldsymbol{\beta}$, the typical approach is to replace the nuisance parameter with an estimator. However, in some cases we are able to find an ancillary statistic for the nuisance parameter and avoid its estimation; then we reduce a risk of an inconvenient estimator.

We will benefit by the ancillarity of regression rank scores in model (4.1) with scales (4.2), and construct the tests of hypotheses

$$\mathbf{H}_1: \ \boldsymbol{\gamma} = \mathbf{0}, \ \boldsymbol{\beta} \text{ unspecified},$$
$$\mathbf{H}_2: \ \boldsymbol{\beta} = \mathbf{0}, \ \boldsymbol{\gamma} \text{ unspecified}.$$

The tests are asymptotically equivalent to the pertinent rank tests with the same score functions, used in the situation where the value of the nuisance parameter is known.

4.2 Rank tests of homoscedasticity

We shall start with testing of homoscedasticity

$$\mathbf{H}_1: \ \boldsymbol{\gamma} = \mathbf{0} \quad \text{against} \quad \mathbf{K}_1: \ \boldsymbol{\gamma} \neq \mathbf{0}.$$

We propose a test based on regression rank scores for $\boldsymbol{\beta}$ in the hypothetic model (see (4.5) below). The concept of regression rank scores was studied in Section 1.3. Write the model (4.1) in the form

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \exp\{\mathbf{z}_i^\top \boldsymbol{\gamma}\} U_i, \quad i = 1, \ldots, n, \tag{4.3}$$

where $\beta_0 \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\gamma} \in \mathbb{R}^q$ and model errors $U_1, \ldots, U_n$ are i.i.d. with distribution function F that has an absolutely continuous density f and finite Fisher information with respect to the location and scale

$$\begin{aligned} &0 < I(f) < \infty, \\ &0 < I_1(f) = \int \left[-1 - x\frac{f'(x)}{f(x)}\right]^2 f(x)dx < \infty. \end{aligned} \tag{4.4}$$

For brevity of notation, we shall further denote

$$\begin{aligned} \mathbf{X}_n &= (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top, & \mathbf{X}_n^* &= (\mathbf{1}_n, \mathbf{X}_n), \\ \mathbf{Z}_n &= (\mathbf{z}_1, \dots, \mathbf{z}_n)^\top, & \mathbf{Z}_n^* &= (\mathbf{1}_n, \mathbf{Z}_n). \\ \mathbf{Y}_n &= (Y_1, \dots, Y_n)^\top \end{aligned}$$

Under $\mathbf{H}_1$, model (4.3) reduces to

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + U_i, \ i = 1, \dots, n. \tag{4.5}$$

The main tool for testing $\mathbf{H}_1$ is the vector $\widehat{\mathbf{a}}_n(\alpha) = (\hat{a}_{n,1}(\alpha), \dots, \hat{a}_{n,n}(\alpha))^\top$ of regression rank scores corresponding to model (4.5). This is an optimal solution of the parametric linear programming problem

$$\widehat{\mathbf{a}}_n(\alpha) = \arg\max\{\mathbf{Y}_n^\top \mathbf{a} \mid \mathbf{X}_n^{*\top} \mathbf{a} = (1-\alpha)\mathbf{X}_n^{*\top} \mathbf{1}_n, \mathbf{a} \in [0,1]^n\}, \quad 0 < \alpha < 1. \tag{4.6}$$

Unlike the score-generating functions used in previous chapters for testing for regression, the score-generating function, suitable for testing the scale, is typically a "U-shaped" function $\varphi : (0,1) \mapsto \mathbb{R}$, square-integrable on $(0,1)$.

Remark. *Remind that the locally optimal score function for model errors U_i with distribution function F and density f (satisfying some regularity conditions) is*

$$\widetilde{\varphi}_1(t, f) = -1 - F^{-1}(t)\, \frac{f'(F^{-1}(t))}{f(F^{-1}(t))}. \tag{4.7}$$

Particularly, $\widetilde{\varphi}_1(t) = (\Phi^{-1}(t))^2 - 1$ is locally optimal for normal and $\widetilde{\varphi}_1(t) = (2u-1)\log(u/(1-u)) - 1$ for logistic model errors.

Having chosen φ, we calculate the rank scores

$$\widehat{\mathbf{b}}_n = (\hat{b}_{n,1}, \dots, \hat{b}_{n,n})^\top, \quad \hat{b}_{n,i} = -\int_0^1 (\varphi(t) - \bar{\varphi}) d\hat{a}_{n,i}(t), \ i = 1, \dots, n \tag{4.8}$$

and consider the rank statistic

$$\mathbf{S}_n = n^{-1/2} \sum_{i=1}^n \mathbf{z}_i \hat{b}_{n,i}. \tag{4.9}$$

The proposed test criterion for $\mathbf{H}_1$ is

$$\mathcal{T}_n^2 = \frac{1}{A^2(\varphi)}\, \mathbf{S}_n^\top \widehat{\mathbf{D}}_n^{-1} \mathbf{S}_n, \quad \text{where} \tag{4.10}$$

$$\widehat{\mathbf{D}}_n = n^{-1}(\mathbf{Z}_n - \widehat{\mathbf{Z}}_n)^\top (\mathbf{Z}_n - \widehat{\mathbf{Z}}_n)$$

and $\widehat{\mathbf{Z}}_n = \mathbf{X}_n^*(\mathbf{X}_n^{*\top}\mathbf{X}_n^*)^{-1}\mathbf{X}_n^{*\top}\mathbf{Z}_n$ is the projection of $\mathbf{Z}_n$ on the space spanned by the columns of $\mathbf{X}_n^*$.

Remark. *Notice that* $\mathbf{X}_n^{*\top}\widehat{\mathbf{b}}_n = \mathbf{0}$*, due to the constraints in (4.6); hence*

$$\sum_{i=1}^{n} \widehat{\mathbf{z}}_i \hat{b}_{n,i} = \mathbf{0},$$

where $\widehat{\mathbf{z}}_i$ *is the i-th row of* $\widehat{\mathbf{Z}}_n$*,* $i = 1, \dots, n$*. It implies that* $\mathbf{S}_n$ *automatically reduces to*

$$\mathbf{S}_n = n^{-1/2} \sum_{i=1}^{n} (\mathbf{z}_i - \widehat{\mathbf{z}}_i)\hat{b}_{n,i}$$

and only the orthogonal complement $\mathbf{Z}_n - \widehat{\mathbf{Z}}_n$ *to* $\mathbf{X}_n^*$ *plays a role in testing. The situation* $\mathbf{Z}_n = \mathbf{X}_n^*$ *needs an estimation of the unknown regression parameter; see Gutenbrunner (1994), Koenker and Zhao (1994), Gutenbrunner et al. (1995) and Koenker (1997) for some attempts.*

The asymptotics for $\mathcal{T}_n^2$ is proven under the following conditions on matrices $\widehat{\mathbf{D}}_n$ and $\mathbf{X}_n^*$. Assume that there exist positive definite matrices $\widehat{\mathbf{D}}$, $\mathbf{M}$ such that

$$\lim_{n\to\infty} \widehat{\mathbf{D}}_n = \widehat{\mathbf{D}}, \ \lim_{n\to\infty} n^{-1}\mathbf{X}_n^{*\top}\mathbf{X}_n^* = \mathbf{M}. \tag{4.11}$$

In addition, let regressors $\mathbf{x}_i^* = (1, \mathbf{x}_i^\top)^\top$ satisfy:

$$\max_{1\le i\le n} \|\mathbf{x}_i^*\| = o(n^{\frac{1}{4}-\eta}) \ \text{ for any } \ \eta > 0, \ \text{ as } \ n \to \infty. \tag{4.12}$$

Remark. *In (1.10) we defined a characteristic* $\gamma(\varphi, f)$*. In this chapter* $\boldsymbol{\gamma}$ *stands for heteroscedastic parameter. To avoid confusion we will change a notation for* $\gamma(\varphi, f)$*; instead we denote*

$$\tau(\varphi, f) = \gamma(\varphi, f) = \int_0^1 \varphi(t)\widetilde{\varphi}(t, f)dt, \quad \widetilde{\varphi}(t, f) = -\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}.$$

We will also define

$$\tau_1(\varphi, f) = \int_0^1 \varphi(t)\widetilde{\varphi}_1(t, f)dt$$

with $\widetilde{\varphi}_1(t, f)$ *defined in (4.7).*

The behavior of the test criterion under $\mathbf{H}_1$ and under the local alternative is summarized in the following theorem.

Theorem 4.1. *Assume that F satisfies conditions (F.1)–(F.2) stated in Section 1.3, and has finite Fisher information with respect to the location and scale (1.6) and(4.4). Let matrices* $\widehat{\mathbf{D}}_n$ *and* $\mathbf{X}_n^*$ *satisfy (4.11) and (4.12). Then in model (4.1) the test statistic* $\mathcal{T}_n^2$ *has under* $\mathbf{H}_1$ *asymptotically* χ^2 *distribution with q degrees of freedom.*

Under the local alternative

$$\mathbf{K}_{1,n}: \ \boldsymbol{\gamma} = n^{-1/2}\boldsymbol{\gamma}^*, \quad \mathbf{0} \neq \boldsymbol{\gamma}^* \in \mathbb{R}^q \text{ fixed}$$

$\mathcal{T}_n^2$ *has asymptotically* χ^2 *with q degrees of freedom and noncentrality parameter*

$$\eta^2 = \frac{\tau_1^2(\varphi, f)}{A^2(\varphi)} \cdot \boldsymbol{\gamma}^{*\top}\widehat{\mathbf{D}}\boldsymbol{\gamma}^*. \tag{4.13}$$

Proof. Following the lines of Gutenbrunner et al. (1993), the linear rank statistic (4.9) admits, under (F.1)–(F.2), (4.12) and under validity of $\mathbf{K}_{1,n}$, the asymptotic representation

$$\mathbf{S}_n = n^{-1/2}(\mathbf{Z}_n - \widehat{\mathbf{Z}}_n)^\top \varphi\Big(F(\exp\{n^{-1/2}\mathbf{z}_i^\top \boldsymbol{\gamma}^*\}U_i)\Big) + o_p(1), \quad \text{as} \quad n \to \infty. \tag{4.14}$$

This implies the asymptotic normality and the covariance structure of $\mathbf{S}_n$ under $\mathbf{H}_1$ as well as under the contiguous alternative $\mathbf{K}_{1,n}$. This, in turn, gives the asymptotic distribution of the test criterion under hypothesis $\mathbf{H}_1$ and under local alternative $\mathbf{K}_{1,n}$. □

Remark. *Notice that both the asymptotic null distribution and the power of the test coincide with those of the corresponding rank test, used under known* $\boldsymbol{\beta}$.

4.3 Rank tests for regression under local heteroscedasticity

Let us proceed to testing hypothesis

$$\mathbf{H}_2 : \boldsymbol{\beta} = \mathbf{0} \quad \text{against} \quad \mathbf{K}_2 : \boldsymbol{\beta} \neq \mathbf{0}.$$

We speak on the local heteroscedasticity, when

$$\boldsymbol{\gamma} = n^{-1/2}\boldsymbol{\gamma}^*, \quad \mathbf{0} \neq \boldsymbol{\gamma}^* \in \mathbb{R}^q \text{ fixed.} \tag{4.15}$$

Hence we have the following model

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \exp\{n^{-1/2}\mathbf{z}_i^\top \boldsymbol{\gamma}^*\}U_i, \quad i = 1, \dots, n.$$

We intend to use standard rank test for regression based on the test statistic T_n^2 defined in (1.4); we simply ignore the local heteroscedasticity. The behavior of T_n^2 in the absence of heteroscedasticity was discussed in Section 1.1.1. It was shown that T_n^2 is distribution-free and under some regularity conditions under $\mathbf{H}_2$ has asymptotically χ^2 distribution with p degrees of freedom. However, the problem of our interest is how an eventual heteroscedasticity affects the validity of the test.

Denote

$$\begin{aligned}
\mathbf{Q}_n &= \frac{1}{n}\sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\top, \quad \text{with} \quad \overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i, \\
\mathbf{U}_n &= \frac{1}{n}\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top, \\
\mathbf{C}_n &= \frac{1}{n}\sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}})\mathbf{z}_i^\top,
\end{aligned}$$

Assume that these exist a positive definite matrix $\mathbf{U}$ and a matrix $\mathbf{C}$, such that as $n \to \infty$

$$\mathbf{U}_n \to \mathbf{U}, \tag{4.16}$$

$$\mathbf{C}_n \to \mathbf{C}, \tag{4.17}$$

$$\max_{1\le i\le n} \|\mathbf{x}_i\| = o(n^{1/2}), \tag{4.18}$$

$$\max_{1\le i\le n} \|\mathbf{z}_i\| = o(n^{1/2}), \tag{4.19}$$

$$\frac{1}{n} \max_{i=1,\dots,n} \mathbf{z}_i^\top \mathbf{U}_n^{-1} \mathbf{z}_i \to 0. \tag{4.20}$$

Theorem 4.2. *Let conditions (1.7) – (1.8) and (4.16) – (4.18) in model (4.3) be satisfied. Assume that f has finite Fisher information with respect to the location and scale. Then the test statistic T_n^2 has under $\mathbf{H}_2$ and under the local heteroscedasticity (4.15) asymptotically χ^2 distribution with p degrees of freedom with noncentrality parameter*

$$\eta^2 = \boldsymbol{\gamma}^{*\top} \mathbf{C}^\top \mathbf{Q}^{-1} \mathbf{C} \boldsymbol{\gamma}^* \frac{\tau_1^2(\varphi, f)}{A^2(\varphi)}. \tag{4.21}$$

Proof. The test T_n^2 is invariant with respect to the location parameter β_0, hence we may further on without any loos of generality assume that $\beta_0 = 0$. Under $\mathbf{H}_2$ and (4.15) Y_i has distribution function

$$P(Y_i \le y) = F(y \exp\{-n^{1/2} \mathbf{z}_i^\top \boldsymbol{\gamma}^*\})$$

and density

$$\exp\{-n^{1/2} \mathbf{z}_i^\top \boldsymbol{\gamma}^*\} f(y \exp\{-n^{1/2} \mathbf{z}_i^\top \boldsymbol{\gamma}^*\})$$

Hence the distribution of $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ under $\mathbf{H}_2$ has density

$$q_{n,\gamma^*}(y_1, \dots, y_n) = \prod_{i=1}^n \exp\{-n^{1/2} \mathbf{z}_i^\top \boldsymbol{\gamma}^*\} f\left(y_i \exp\{-n^{1/2} \mathbf{z}_i^\top \boldsymbol{\gamma}^*\}\right). \tag{4.22}$$

The sequence of densities $\{q_{n,\gamma^*}\}$ is contiguous to $\{q_{n,0}\}$ corresponding to $\boldsymbol{\gamma} = \mathbf{0}$ (see Hájek and Šidák (1967, Chapter VI)). It implies that the asymptotic distribution of $\mathbf{S}_n$ under $\mathbf{H}_2$ and under (4.15) is normal $\mathcal{N}_p\Big(\boldsymbol{\mu}_{\gamma^*}, \mathbf{Q}A^2(\varphi)\Big)$, where

$$\boldsymbol{\mu}_{\gamma^*} = \mathbf{C}\ \boldsymbol{\gamma}^* \int_0^1 \varphi(u) \widetilde{\varphi}_1(u, f) du = \mathbf{C}\ \boldsymbol{\gamma}^* \tau_1(\varphi, f).$$

Hence, the criterion $T_n^2 = A^{-2}(\varphi) \mathbf{S}_n^\top \mathbf{Q}_n^{-1} \mathbf{S}_n$ has under $\mathbf{H}_2$ and under local heteroscedasticity asymptotically noncentral $\chi^2(p)$ distribution with the parameter of noncentrality η^2. □

Remark. *Particularly, the noncentrality parameter vanishes if f is symmetric and φ is skew-symmetric, i.e.*

$$f(x) = f(-x), \quad x \in \mathbb{R} \quad \textit{and} \quad \varphi(u) = -\varphi(1-u),\ 0 < u < 1.$$

Hence, because a skew-symmetric score generating function is our possible choice, the local heteroscedasticity then does not affect the asymptotic null distribution of the rank test under symmetric parent distribution.

The noncentrality parameter also vanishes if $(\mathbf{x}_i - \overline{\mathbf{x}})$ *and* $\mathbf{z}_i$ *are asymptotically orthogonal, i.e.*

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})\mathbf{z}_i^\top \to \mathbf{0}, \quad \textit{as} \quad n \to \infty. \tag{4.23}$$

Consider a sequence of local alternatives

$$\mathbf{K}_{2,n}: \ \boldsymbol{\beta} = n^{-1/2}\boldsymbol{\beta}^*, \quad \mathbf{0} \neq \boldsymbol{\beta}^* \in \mathbb{R}^p \text{ fixed.} \tag{4.24}$$

Again, as in proof of Theorem 4.2 without loos of generality assume that $\beta_0 = 0$. The distribution function and density of Y_i under $\mathbf{H}_2$ and under the local heteroscedasticity are

$$\begin{aligned} P_{n,i}(y) &= F\left(y \exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right), \\ p_{n,i}(y) &= \exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\ f\left(y \exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right), \quad i = 1, \dots, n. \end{aligned} \tag{4.25}$$

The distribution function and density of Y_i under the local regression alternative $\mathbf{K}_{2,n}$ in the presence of the local heteroscedasticity are

$$Q_{n,i}(y) = F\left((y - n^{-1/2}\mathbf{x}_i^\top\boldsymbol{\beta}^*)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right),$$

$$q_{n,i}(y) = \exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\ f\left((y - n^{-1/2}\mathbf{x}_i^\top\boldsymbol{\beta}^*)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right), \quad i = 1, \dots, n.$$

Using the result of Oosterhoff and van Zwet (1979) stated in Lemma 3.2, we are able to prove the contiguity of the sequence $\{\prod_{i=1}^n Q_{n,i}\}_{n=1}^\infty$ with respect to the sequence $\{\prod_{i=1}^n P_{n,i}\}_{n=1}^\infty$.

Lemma 4.1. *Let conditions (1.7) – (1.8) and (4.16) – (4.20) in model (4.3) be satisfied. Assume that* f *has finite Fisher information with respect to the location and scale. Then the sequence of distributions* $\{\prod_{i=1}^n Q_{n,i}\}_{n=1}^\infty$ *is contiguous with respect to the sequence* $\{\prod_{i=1}^n P_{n,i}\}_{n=1}^\infty$ *for every fixed* $\boldsymbol{\beta}^*$, $\|\boldsymbol{\beta}^*\| < \infty$ *and* $\boldsymbol{\gamma}^*$, $\|\boldsymbol{\gamma}^*\| < \infty$.

Proof. Thanks to Lemma 3.2 we need to verify (3.7) and (3.8). Actually

$$\begin{aligned} \sum_{i=1}^{n} H^2(P_{n,i}, Q_{n,i}) &= \sum_{i=1}^{n} \int \exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\} \Big\{ \left[f\left(y \exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)\right]^{\frac{1}{2}} \\ &\quad - \left[f\left((y - n^{-1/2}\mathbf{x}_i^\top\boldsymbol{\beta}^*)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)\right]^{\frac{1}{2}} \Big\}^2 dy \\ &= \sum_{i=1}^{n} \exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\} \int \Big\{ \left[f\left(y\ \exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)\right]^{\frac{1}{2}} \\ &\quad - \left[f\left((y - n^{-1/2}\mathbf{x}_i^\top\boldsymbol{\beta}^*)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)\right]^{\frac{1}{2}} \Big\}^2 dy. \end{aligned}$$

If $n^{-1/2}\mathbf{x}_i^\top \boldsymbol{\beta}^* = \eta_i > 0$, then

$$\begin{aligned}
&\left\{\left[f\left(y\ \exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)\right]^{\frac{1}{2}} - \left[f\left((y-n^{-1/2}\mathbf{x}_i^\top\boldsymbol{\beta}^*)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)\right]^{\frac{1}{2}}\right\}^2 \\
&\leq \left\{\frac{1}{2}\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\int_0^{\eta_i}\frac{\left|f'\left((y-t)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)\right|}{f^{\frac{1}{2}}\left((y-t)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)}\,dt\right\}^2 \\
&\leq \left(\frac{1}{2}\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)^2\ \eta_i\int_0^{\eta_i}\frac{\left[f'\left((y-t)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)\right]^2}{f\left((y-t)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\right)}\,dt
\end{aligned}$$

similarly for $n^{-1/2}\mathbf{x}_i^\top\boldsymbol{\beta}^* = -\eta_i < 0$. Hence,

$$\begin{aligned}
\sum_{i=1}^n H^2(P_{n,i}, Q_{n,i}) &\leq \frac{1}{4}\sum_{i=1}^n \eta_i^2\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}I(f) \\
&= \frac{1}{4n}I(f)\sum_{i=1}^n(\mathbf{x}_i^\top\boldsymbol{\beta}^*)^2\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}
\end{aligned}$$

what confirms (3.7). It remains to prove (3.8). Actually, for any $c_n \to \infty$

$$\begin{aligned}
&\lim_{n\to\infty}\sum_{i=1}^n Q_{n,i}\left\{\frac{q_{n,i}(Y_i)}{p_{n,i}(Y_i)} \geq c_n\right\} \\
=\ &\lim_{n\to\infty}\sum_{i=1}^n\int_{-\infty}^{\infty}\mathbb{I}\left[\frac{f\Big((y-n^{-1/2}\mathbf{x}_i^\top\boldsymbol{\beta}^*)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\Big)}{f\Big(y\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\Big)} \geq c_n\right]\cdot \\
&\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}f\Big((y-n^{-1/2}\mathbf{x}_i^\top\boldsymbol{\beta}^*)\exp\{-n^{-1/2}\mathbf{z}_i^\top\boldsymbol{\gamma}^*\}\Big)dy \\
=\ &\lim_{n\to\infty}\sum_{i=1}^n\int_{-\infty}^{\infty}\mathbb{I}\left[\frac{f(t)}{f(t+n^{-1/2}\mathbf{x}_i^\top\boldsymbol{\beta}^*\exp\{-n^{-1/2}\mathbf{z}_i^\top\gamma^*\})} \geq c_n\right]f(t)dt \\
=\ &\lim_{n\to\infty}\sum_{i=1}^n\int_{-\infty}^{\infty}\mathbb{I}\left[\frac{f(t)}{f(t+n^{-1/2}\widetilde{\mathbf{x}}_i^\top\boldsymbol{\beta}^*)} \geq c_n\right]f(t)dt = 0,
\end{aligned}$$

applying the fact that $\{\prod_{i=1}^n \widetilde{Q}_{n,i}\}$ is contiguous with respect to $\{\prod_{i=1}^n \widetilde{P}_{n,i}\}$ with

$$\widetilde{q}_{n,i}(y) = f(y - n^{-1/2}\widetilde{\mathbf{x}}_i^\top\boldsymbol{\beta}^*), \quad \widetilde{p}_{n,i}(y) = f(y), \ \ i = 1,\ldots,n$$

for $\widetilde{\mathbf{x}}_i = \mathbf{x}_i(1+o(1)),\ i = 1,\ldots,n.$ □

Theorem 4.3. *Let conditions (1.7) – (1.8) and (4.16) – (4.20) in model (4.3) be satisfied. Assume that f has finite Fisher information with respect to the location and scale. Then the test statistic T_n^2 has under local alternative (4.24) and under the local heteroscedasticity (4.15) asymptotically χ^2 distribution with p degrees of freedom and noncentrality parameter*

$$\begin{aligned}
&\widetilde{\eta}^2 = A^{-2}(\varphi)\cdot &(4.26)\\
&\left\{\boldsymbol{\gamma}^{*\top}\mathbf{C}^\top\mathbf{Q}^{-1}\mathbf{C}\boldsymbol{\gamma}^*\tau_1^2(\varphi,f) + \boldsymbol{\beta}^{*\top}\mathbf{Q}\boldsymbol{\beta}^*\tau^2(\varphi,f) + 2\boldsymbol{\gamma}^{*\top}\mathbf{C}^\top\boldsymbol{\beta}^*\tau_1(\varphi,f)\tau(\varphi,f)\right\}.
\end{aligned}$$

Proof. Lemma 4.1 and Hájek and Šidák (1967, Chapter VI) imply that under the local regression alternative, in presence of the local heteroscedasticity, $\mathbf{S}_n$ has asymptotically normal distribution $\mathcal{N}_p\Big(\boldsymbol{\mu}_{\gamma^*,\beta^*}, \mathbf{Q}A^2(\varphi)\Big)$, where

$$\boldsymbol{\mu}_{\gamma^*,\beta^*} = \mathbf{C}\ \boldsymbol{\gamma}^*\tau_1(\varphi, f) + \mathbf{Q}\ \boldsymbol{\beta}^*\tau(\varphi, f). \tag{4.27}$$

Hence, the criterion T_n^2 has under the local regression alternative, in presence of the local heteroscedasticity, asymptotically noncentral χ_p^2 distribution with the noncentrality parameter (4.26). □

Hence if φ is skew-symmetric and f symmetric or if (4.23) is satisfied, then the local heteroscedasticity affects neither the critical region nor the asymptotic power of the test T_n^2. In general case we may get a formula for the asymptotic relative efficiency this test in presence of the local heteroscedasticity with respect to the test in the homoscedastic case. It is given by

$$ARE = \frac{\boldsymbol{\beta}^{*\top}\mathbf{Q}\boldsymbol{\beta}^*\tau^2(\varphi, f)}{\boldsymbol{\gamma}^{*\top}\mathbf{C}^\top\mathbf{Q}^{-1}\mathbf{C}\boldsymbol{\gamma}^*\tau_1^2(\varphi, f) + \boldsymbol{\beta}^{*\top}\mathbf{Q}\boldsymbol{\beta}^*\tau^2(\varphi, f) + 2\boldsymbol{\gamma}^{*\top}\mathbf{C}^\top\boldsymbol{\beta}^*\tau_1(\varphi, f)\tau(\varphi, f)}.$$

However, if we are not sure that the heteroscedasticity is only local, we must replace the unknown $\boldsymbol{\gamma}$ by an appropriate estimate and use the aligned rank test for testing $\mathbf{H}_2$, see Akritas and Albers (1993). We may also use suitable regression rank score test to avoid an estimation of nuisance $\boldsymbol{\gamma}$.

4.4 RRS tests for regression under heteroscedasticity

Recall that we test the hypothesis

$$\mathbf{H}_2:\ \boldsymbol{\beta} = \mathbf{0} \quad \text{against} \quad \mathbf{K}_2:\ \boldsymbol{\beta} \neq \mathbf{0}$$

in model (4.3). Without estimating the unknown $\boldsymbol{\gamma}$, we are able to construct the tests for $\mathbf{H}_2$ only under symmetric F. The symmetrization of the form $Y_i^* = Y_i - Y_i'$, where Y_i' is an independent copy of Y_i, would not bring a convenient extension to asymmetric F, because it would eliminate even the regression, the object of study. When symmetric F cannot be assumed, the unknown $\boldsymbol{\gamma}$ should be replaced with an estimate and some aligned test should be used.

Remark. *Without loss of generality we will assume that there is no intercept in model (4.3), i.e.* $\beta_0 = 0$. *Otherwise one would include it in the parameter* $\widetilde{\boldsymbol{\beta}} = (\beta_0, \boldsymbol{\beta}^\top)^\top$.

Under validity of $\mathbf{H}_2$, we can write the following identity:

$$|Y_i| = \exp\{\mathbf{z}_i^\top\boldsymbol{\gamma}\}\ |U_i|,\ i = 1, \ldots, n. \tag{4.28}$$

Consider the statement (4.28) as an hypothesis $\mathbf{H}_2'$, asserting that $Y_1, \ldots, Y_n$ satisfy (4.28). We shall verify the hypothesis $\mathbf{H}_2'$ instead of $\mathbf{H}_2$. If $\mathbf{H}_2'$ is not true, then $\mathbf{H}_2$ is not true, either. For this purpose, put $W_i = \ln|Y_i|,\ V_i = \ln|U_i|,\ i = 1,\dots,n$ (notice that $Y_i \neq 0,\ U_i \neq 0$ with probability 1, $i = 1,\dots,n$), and rewrite (4.28) in the form

$$W_i = \mathbf{z}_i^\top \boldsymbol{\gamma} + V_i,\ i = 1,\dots,n. \tag{4.29}$$

Then $V_1,\dots,V_n$ are i.i.d. random variables with distribution function $G(v)$ and density $g(v)$, satisfying the following relations:

$$\begin{aligned}
G(v) &= P(\ln|U_i| \le v) = 2F(\mathrm{e}^v) - 1,\ v \in \mathbb{R} \\
G^{-1}(\alpha) &= \ln\Big(F^{-1}\Big(\frac{\alpha+1}{2}\Big)\Big),\ 0 < \alpha < 1 \\
g(v) &= 2\,\mathrm{e}^v f(\mathrm{e}^v),\ v \in \mathbb{R} \\
g(G^{-1}(\alpha)) &= 2F^{-1}\Big(\frac{\alpha+1}{2}\Big) f\Big(F^{-1}\Big(\frac{\alpha+1}{2}\Big)\Big),\ 0 < \alpha < 1 \\
-\frac{g'(v)}{g(v)} &= -1 - \mathrm{e}^v \frac{f'(\mathrm{e}^v)}{f(\mathrm{e}^v)},\ v \in \mathbb{R} \\
-\frac{g'(G^{-1}(\alpha))}{g(G^{-1}(\alpha))} &= -1 - F^{-1}\Big(\frac{\alpha+1}{2}\Big) \cdot \frac{f'(F^{-1}(\frac{\alpha+1}{2}))}{f(F^{-1}(\frac{\alpha+1}{2}))},\ 0 < \alpha < 1.
\end{aligned} \tag{4.30}$$

The following lemma shows that G has positive and finite Fisher information with respect to the shift location, which coincides with Fisher information for F with respect to the scale:

Lemma 4.2. *Let F have an absolutely continuous symmetric density with finite Fisher information, then distribution function G in (4.30) has an absolutely continuous density g and finite Fisher information with respect to the location and $I(g) = I_1(f)$.*

Proof. Indeed,

$$\begin{aligned}
I(g) &= \int_{-\infty}^{\infty} \Big(1 + \mathrm{e}^v \frac{f'(\mathrm{e}^v)}{f(\mathrm{e}^v)}\Big)^2 2\,\mathrm{e}^v (f(\mathrm{e}^v) dv \\
&= \int_0^{\infty} \Big(1 + y\frac{f'(y)}{f(y)}\Big)^2 2f(y)dy = I_1(f) < \infty.
\end{aligned}$$

□

If F has an absolutely continuous symmetric density $f > 0$ with nondegenerate tails, finite Fisher information with respect to the location and scale, and satisfies conditions (F.1)–(F.2) in Section 1.3, then G has an absolutely continuous density g satisfying (4.30) with finite Fisher information for location. Moreover, (4.30) implies that

$$\Big|-\frac{g'(v)}{g(v)}\Big| = \Big|-1 - \mathrm{e}^v \frac{f'(\mathrm{e}^v)}{f(\mathrm{e}^v)}\Big| \le \mathrm{e}^{2v} - 1 \quad \text{for} \quad |v| > K_1. \tag{4.31}$$

Then G fulfills conditions (F.1)–(F.2) in Section 4.2, hence the regression rank scores for model (4.29) with error distribution G have a sense.

We replace testing $\mathbf{H}_2$ with testing $\mathbf{H}_2'$ corresponding to a class of distribution functions G, related to a symmetric F according to (4.30). Because of the symmetry of F, we shall use the signed rank test of $\mathbf{H}_2'$, using the regression rank scores corresponding to the model (4.29). We shall assume that, as $n \to \infty$,

(Z.1) $z_{i,1} = 1,\ i = 1, \dots, n$ and $\max_{1\le i\le n} \|\mathbf{z}_i\| = \mathcal{O}(1)$ as $n \to \infty$.

(Z.2) $\mathbf{U}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i^\top \to \mathbf{U}$, where $\mathbf{U}$ is a positive definite matrix.

(Z.3) $\widetilde{\mathbf{Q}}_n(\boldsymbol{\gamma}) = n^{-1}\sum_{i=1}^n \mathrm{e}^{-\mathbf{z}_i^\top\gamma}(\mathbf{x}_i - \widehat{\mathbf{x}}_i)(\mathbf{x}_i - \widehat{\mathbf{x}}_i)^\top \to \widetilde{\mathbf{Q}}(\boldsymbol{\gamma}), \quad \forall \boldsymbol{\gamma} \in \mathbb{R}^q$, where $\widetilde{\mathbf{Q}}(\boldsymbol{\gamma})$ is a positive definite matrix. Here $\widehat{\mathbf{x}}_i^\top$ is the i-th row of $\widehat{\mathbf{X}}_n = \mathbf{Z}_n(\mathbf{Z}_n^\top\mathbf{Z}_n)^{-1}\mathbf{Z}_n^\top\mathbf{X}_n$.

For simplicity, denote $\widetilde{\mathbf{Q}}_n(\mathbf{0}) \equiv \widetilde{\mathbf{Q}}_n$ and $\widetilde{\mathbf{Q}}(\mathbf{0}) \equiv \widetilde{\mathbf{Q}}$, respectively. Let $\widehat{\mathbf{a}}_n(\alpha) = (\hat{a}_{n,1}(\alpha), \dots, \hat{a}_{n,n}(\alpha))^\top$ be the α-regression rank scores for the model (4.29), i.e. the optimal solution of the linear programming problem

$$\widehat{\mathbf{a}}_n(\alpha) = \arg\max\{\mathbf{W}_n^\top\mathbf{a} \mid \mathbf{Z}_n^\top\mathbf{a} = (1-\alpha)\mathbf{Z}_n^\top\mathbf{1}_n, \mathbf{a} \in [0,1]^n\}, \quad 0 < \alpha < 1, \tag{4.32}$$

where $\mathbf{W}_n = (W_1, \dots, W_n)^\top$. Then $\widehat{\mathbf{a}}_n(\alpha)$ is invariant to the transformations

$$W_i \mapsto W_i + \mathbf{z}_i^\top\boldsymbol{\gamma}, \quad \boldsymbol{\gamma} \in \mathbb{R}^q. \tag{4.33}$$

Let $\varphi : (0,1) \mapsto \mathbb{R}$ be a nonconstant, nondecreasing, square-integrable, skew-symmetric score function, whose derivative $\varphi'(u)$ exists and satisfies for all $0 < u < \alpha_0,\ 1-\alpha_0 < u < 1$ (for some $\alpha_0 > 0$):

$$|\varphi'(u)| \le c(u(1-u))^{-1-\delta} \quad \text{for some } \delta > 0. \tag{4.34}$$

Remark. *The class of functions satisfying (4.34) covers the Wilcoxon- and van der Waerden-type tests, among others.*

Regarding the symmetry of F, we use the signed regression rank scores test of $\mathbf{H}_2$ with the statistic

$$\begin{aligned}
\widetilde{\mathbf{S}}_n &= n^{-1/2}\sum_{i=1}^n (\mathbf{x}_i - \widehat{\mathbf{x}}_i)\ \text{sign}\ Y_i\ \hat{b}_{n,i}^+,\\
\widehat{\mathbf{b}}_n^+ &= (\hat{b}_{n,1}^+, \dots, \hat{b}_{n,n}^+)^\top, \quad \hat{b}_{n,i}^+ = -\int_0^1 \varphi^+(u)\ d\hat{a}_{n,i}(u),\ i = 1, \dots, n.\\
\varphi^+(u) &= \varphi\Big(\frac{u+1}{2}\Big),\ 0 < u < 1
\end{aligned} \tag{4.35}$$

Remark. *Notice that* sign Y_i *and* $\hat{b}_{n,i}^+$ *are independent under* $\mathbf{H}_2,\ i = 1, \dots, n$, *due to the symmetry of F. Moreover,* W_i *in (4.32) can be replaced with i.i.d.* $V_i,\ i = 1, \dots, n$, *free of* $\boldsymbol{\gamma}$, *due to the invariance of regression rank scores.*

Finally, we propose the following test criterion for $\mathbf{H}_2$:

$$\widetilde{\mathcal{T}}_n^2 = \frac{1}{A^2(\varphi)}\widetilde{\mathbf{S}}_n^\top\widetilde{\mathbf{Q}}_n^{-1}\widetilde{\mathbf{S}}_n. \tag{4.36}$$

The asymptotic behavior of $\widetilde{\mathcal{T}}_n^2$ is stated in the following theorem.

Theorem 4.4. *Assume that F is symmetric, satisfies conditions (F.1)–(F.2) and has finite Fisher information with respect to the location and scale. The regressors in model (4.1) let satisfy conditions (Z.1)–(Z.3). Then in model (4.1) $\widetilde{\mathcal{T}}_n^2$ has under $\mathbf{H}_2$ asymptotically χ^2 distribution with p degrees of freedom and under the local alternative (4.24) the asymptotic distribution of $\widetilde{\mathcal{T}}_n^2$ is noncentral χ^2 with p degrees of freedom and noncentrality parameter*

$$\tilde{\eta}^2(\boldsymbol{\gamma}) = \frac{\tau_1^2(\varphi, f)}{A^2(\varphi)}\, \boldsymbol{\beta}^{*\top} \widetilde{\mathbf{Q}}^\top(\boldsymbol{\gamma}) \widetilde{\mathbf{Q}}^{-1} \widetilde{\mathbf{Q}}(\boldsymbol{\gamma}) \boldsymbol{\beta}^*. \tag{4.37}$$

Proof. Indeed, it follows from Gutenbrunner et al. (1993), Hušková (1970) and Puri and Sen (1985), that under $\mathbf{H}_2'$, the statistic $\widetilde{\mathbf{S}}_n$ admits the asymptotic representation

$$\begin{aligned} \widetilde{\mathbf{S}}_n &= n^{-1/2} \sum_{i=1}^n (\mathbf{x}_i - \widehat{\mathbf{x}}_i) \text{ sign } Y_i\ \varphi^+(2F(|\,\mathrm{e}^{-\mathbf{z}_i^\top \gamma} Y_i|) - 1) + o_p(1) \\ &= n^{-1/2} \sum_{i=1}^n (\mathbf{x}_i - \widehat{\mathbf{x}}_i) \varphi(F(\mathrm{e}^{-\mathbf{z}_i^\top \gamma} Y_i)) + o_p(1) \\ &= n^{-1/2} \sum_{i=1}^n (\mathbf{x}_i - \widehat{\mathbf{x}}_i) \varphi(F(U_i)) + o_p(1) \end{aligned} \tag{4.38}$$

as $n \to \infty$, (notice that $\mathrm{e}^{-\mathbf{z}_i^\top \gamma} Y_i = U_i,\ i = 1, \ldots, n$). This further implies that the asymptotic null distribution of $\widetilde{\mathbf{S}}_n$ under $\mathbf{H}_2'$ is normal $\mathcal{N}_p\Big(\mathbf{0}, A^2(\varphi)\widetilde{\mathbf{Q}}\ \Big)$.

Consider the local alternative $\mathbf{K}_{2,n}$ for $\mathbf{H}_2'$ defined in (4.24), which is contiguous to $\mathbf{H}_2'$ for every fixed $\boldsymbol{\gamma}$ (see Hájek and Šidák (1967, Chapter VI)).

The distribution function of Y_i under $\mathbf{K}_{2,n}$ and under a fixed $\boldsymbol{\gamma} \in \mathbb{R}^q$ is $F\left((y - n^{-1/2}\mathbf{x}_i^\top \boldsymbol{\beta}^*) \exp\{-\mathbf{z}_i^\top \boldsymbol{\gamma}\}\right)$. The representation (4.38) is true also under $\mathbf{K}_{2,n}$ and under any fixed $\boldsymbol{\gamma}$. Regarding the invariance (4.33), it can be expressed as

$$\widetilde{\mathbf{S}}_n = n^{-1/2} \sum_{i=1}^n (\mathbf{x}_i - \widehat{\mathbf{x}}_i) \varphi\Big(F(U_i + n^{-1/2} \mathbf{x}_i^\top \boldsymbol{\beta}^* \, \mathrm{e}^{-\mathbf{z}_i^\top \gamma})\Big) + o_p(1). \tag{4.39}$$

Hence $\widetilde{\mathbf{S}}_n$ is asymptotically normal under $\mathbf{K}_{2,n}$ and under a fixed $\boldsymbol{\gamma} \in \mathbb{R}^q$:

$$\mathcal{N}_p\Big(\tau_1(\varphi, f) \widetilde{\mathbf{Q}}(\boldsymbol{\gamma}) \boldsymbol{\beta}^*, A^2(\varphi) \widetilde{\mathbf{Q}}\Big). \tag{4.40}$$

Hence, the asymptotic distribution of $\widetilde{\mathcal{T}}_n^2$ under $\mathbf{H}_2'$ is central χ^2 with p degrees of freedom, while under $\mathbf{K}_{2,n}$ and fixed $\boldsymbol{\gamma}$ it is noncentral χ^2 with p degrees of freedom and noncentrality parameter (4.37). □

The behavior of test (4.36) is in accord with the signed rank tests of $\mathbf{H}_2$ under $\boldsymbol{\gamma}$ known; such tests were studied by Hušková (1970) and Puri and Sen (1985). The numerical study in the following section illustrates the finite sample behavior of all proposed tests from previous sections, for more details and other simulations see Jurečková and Navrátil (2012) and Jurečková and Navrátil (2014).

$\gamma \setminus U_i$	$\mathcal{N}\left(0, \frac{3}{2}\right)$			$Log\left(0, \frac{\sqrt{2}\pi}{3}\right)$			$Lap\left(0, \frac{\sqrt{3}}{2}\right)$			$t(6)$		
0	5.16	3.99	3.17	4.77	3.62	2.86	4.78	3.65	2.87	5.53	4.21	3.39
0.1	17.79	20.69	18.32	16.54	16.86	14.09	13.02	12.48	10.33	16.23	16.10	13.57
−0.1	18.61	20.90	18.21	15.91	16.05	13.63	12.85	11.88	10.11	15.93	15.33	12.87
0.2	53.77	63.88	59.91	48.16	52.36	47.95	36.76	38.84	34.16	46.98	50.28	45.50
−0.2	53.98	63.44	59.69	47.95	52.87	48.11	36.67	38.99	34.53	47.32	51.04	45.52
0.3	84.71	91.54	89.97	80.27	85.51	81.91	67.32	71.68	67.08	79.08	83.78	79.80
−0.3	84.43	91.48	89.78	81.17	85.99	82.38	67.80	72.14	67.10	79.25	83.29	79.17

Table 4.1: Percentage of rejections of hypothesis $\mathbf{H}_1 : \gamma = 0$ for various model errors U_i and for three choices of the score function φ (absolute value of Wilcoxon scores for location, Wilcoxon scores for scale and van der Waerden scores for scale); $n = 100$.

4.5 Numerical illustration

4.5.1 Test of $\mathbf{H}_1 : \boldsymbol{\gamma} = \mathbf{0}$

Consider the model of regression line with a possible heteroscedasticity,

$$Y_i = \beta_0 + \beta x_i + \exp\{z_i \gamma\} U_i, \quad i = 1, \dots, n$$

and the problem of testing the homoscedasticity $\mathbf{H}_1 : \gamma = 0$ against two-sided alternative $\gamma \neq 0$, considering β_0 and β as nuisance parameters. The nuisance parameters do not affect power of the tests (as it was shown in Section 4.2 and confirmed in the simulation study); thus these parameters will be further considered as fixed: $\beta_0 = 2$ and $\beta = 1$.

First we compare the regression rank score test (4.10) for three different choices of score function φ:

$\varphi^{(1)}(t) = \lvert t - \frac{1}{2} \rvert$	absolute value of Wilcoxon scores for location,
$\varphi^{(2)}(t) = -1 + (2t-1)\log\frac{t}{1-t}$	Wilcoxon scores for scale,
$\varphi^{(3)}(t) = (\Phi^{-1}(t))^2 - 1$	van der Waerden scores for scale.

The regressors x_i and z_i were once generated from independent samples of sizes $n = 100$ from uniform $(-2, 10)$ distribution, eventually the z_i from standard normal distribution and then considered fixed. The model errors U_i were generated from normal, logistic, Laplace and t-distribution with 6 degrees of freedom, respectively, always with 0 mean and variance 3/2 The empirical powers of the tests were computed as a percentage of rejections of $\mathbf{H}_1$ among 10 000 replications, at significance level $\alpha = 0.05$. The results are summarized in Table 4.1; they show a good performance for all three scores.

Next we shall compare the regression rank score tests with their classical rank test counterparts; we consider Wilcoxon scores with score function $\varphi^{(2)}(t)$. The

$\gamma \setminus U_i$	$\mathcal{N}\left(0, \frac{3}{2}\right)$		$Log\left(0, \frac{\sqrt{2}\pi}{3}\right)$		$Lap\left(0, \frac{\sqrt{3}}{2}\right)$		$t(6)$	
0	2.84	3.93	2.91	4.14	2.72	4.14	3.00	3.97
0.1	17.96	20.62	13.74	16.97	10.70	12.85	13.49	15.81
−0.1	18.02	20.72	13.90	16.51	10.62	12.69	13.80	16.13
0.2	60.40	62.96	49.99	52.62	36.50	39.8	48.37	50.74
−0.2	60.60	63.27	50.53	53.30	36.72	40.13	48.63	51.70
0.3	91.23	91.60	84.73	85.44	69.81	71.74	82.96	84.14
−0.3	91.98	92.56	84.94	85.65	69.76	71.57	82.72	83.78

Table 4.2: Percentage of rejections of hypothesis $\mathbf{H}_1 : \gamma = 0$ for various errors U_i, for classical rank test for the scale alternative and regression rank score test (in this order), both with Wilcoxon scores for scale; $n = 100$.

value of β is fixed as $\beta = 1$ in the rank test, for the purpose of comparison, though β is typically unknown. Anyway, the theoretical as well as simulation studies indicate that the regression rank score test (with parameter β unspecified) achieves the same asymptotic power as the classical rank test for the scale with β known.

The regressors x_i and z_i were once generated from independent samples of sizes $n = 100$ from the standard normal distribution and then considered fixed. The model errors U_i were generated from normal, logistic, Laplace and t-distributions with 6 degrees of freedom, respectively, all with 0 mean a variance 3/2. The empirical powers of the tests were again computed as a percentage of rejections of $\mathbf{H}_1$ among 10 000 replications, at significance level $\alpha = 0.05$. The results are summarized in Table 4.2.

4.5.2 Test of $\mathbf{H}_2 : \boldsymbol{\beta} = \mathbf{0}$

Again, consider the model of regression line with a possible heteroscedasticity,

$$Y_i = \beta x_i + \exp\{\gamma_0 + \gamma_1 z_i\} U_i, \quad i = 1, \ldots, n$$

and the problem of testing $\mathbf{H}_2 : \beta = 0$ against two-sided alternative $\beta \neq 0$, considering $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)^\top$ as a nuisance parameter.

We start with a comparison of the signed regression rank score test (4.36) for three choices of score function φ:

$\varphi^{(1)}(t) = t - \frac{1}{2}$ Wilcoxon scores,
$\varphi^{(2)}(t) = \Phi^{-1}(t)$ van der Waerden scores,
$\varphi^{(3)}(t) = \text{sign}(t - \frac{1}{2})$ sign scores.

The regressors x_i and z_i were once generated from independent samples of sizes $n = 100$ from uniform $(-2, 10)$ distribution and then considered fixed. The model errors U_i were generated from normal, logistic, Laplace and t-distribution with 6 degrees of freedom, respectively, always with 0 mean and variance 3/2.

The empirical powers of tests were computed as a percentage of rejections of $\mathbf{H}_2$ among 10 000 replications, at $\alpha = 0.05$. Table 4.3 compares empirical powers of tests under a fixed value of nuisance parameter $\boldsymbol{\gamma} = (0,\ 0.1)^T$, while Table 4.5 compares empirical powers of tests under fixed values $\beta = 0.05$ and $\gamma_0 = 0$, and various γ_1. Table 4.4 illustrates the situation where the regressors in the linear regression and in the nuisance heteroscedasticity coincide, i.e. $z_i = x_i$; the rest of the simulation setup remains the same as in Table 4.3.

$\beta \setminus U_i$	$\mathcal{N}\left(0, \frac{3}{2}\right)$			$Log\left(0, \frac{\sqrt{2}\pi}{3}\right)$			$Lap\left(0, \frac{\sqrt{3}}{2}\right)$			$t(6)$		
0	5.21	4.95	5.33	5.21	4.96	5.01	5.21	4.97	5.01	5.18	5.24	5.45
0.05	15.85	15.51	12.38	16.65	15.95	13.77	20.86	18.18	20.74	18.29	17.22	14.58
−0.05	15.67	15.80	12.37	17.20	16.38	14.25	20.82	18.34	21.55	17.54	16.72	14.27
0.1	46.17	46.30	32.11	49.30	47.44	38.37	57.82	52.48	51.57	51.40	49.09	40.53
−0.1	45.20	46.02	31.68	49.31	47.88	38.48	58.52	53.42	52.21	51.27	49.13	39.74
0.2	93.06	93.84	77.77	94.65	94.18	82.58	96.52	95.24	87.67	95.06	94.74	84.23
−0.2	93.32	94.06	78.02	94.83	94.37	83.02	96.75	95.34	88.67	95.30	94.40	84.56

Table 4.3: Percentage of rejections of hypothesis $\mathbf{H}_2 : \beta = 0$ for various model errors U_i and for three choices of score function φ (Wilcoxon, van der Waerden, sign scores), under fixed $\boldsymbol{\gamma} = (0,\ 0.1)^\top$; $n = 100$.

$\beta \setminus U_i$	$\mathcal{N}\left(0, \frac{3}{2}\right)$			$Log\left(0, \frac{\sqrt{2}\pi}{3}\right)$			$Lap\left(0, \frac{\sqrt{3}}{2}\right)$			$t(6)$		
0	4.77	4.60	4.61	4.71	4.32	5.03	4.72	4.32	5.03	4.84	4.48	5.26
0.05	7.90	7.80	7.02	8.50	8.22	7.66	9.72	8.94	10.25	8.80	8.27	7.88
−0.05	8.02	7.89	7.07	8.43	8.06	7.83	9.56	8.77	10.45	8.54	8.36	7.69
0.1	17.58	17.16	13.65	19.42	17.78	16.14	23.18	20.35	23.71	19.81	18.43	16.73
−0.1	17.17	16.76	13.61	18.48	17.43	15.71	22.42	19.74	23.78	19.85	18.63	16.60
0.2	43.80	42.77	35.63	46.32	43.88	40.50	49.88	45.71	49.29	46.32	43.26	41.54
−0.2	43.72	42.80	35.30	45.61	43.27	40.78	49.23	45.23	49.45	46.68	44.14	41.92

Table 4.4: Percentage of rejections of hypothesis $\mathbf{H}_2 : \beta = 0$ for various model errors U_i and for three choices of score function φ (Wilcoxon, van der Waerden, sign scores), under fixed $\boldsymbol{\gamma} = (0,\ 0.1)^\top$; $z_i = x_i$, $n = 100$.

The powers depend on nuisance parameter $\boldsymbol{\gamma}$, in accordance with (4.37). The powers increase for γ_1 negative and decrease for γ_1 positive, due to influence of z_i on the variance. Similarly as in the previous subsection, we observe that the van der Waerden scores achieve approximately the same power as Wilcoxon scores, even for normal model errors. The performance of the sign scores is the best for the Laplace model errors, but with approximately the same power as Wilcoxon. Apart from their simple form, the Wilcoxon scores generally achieve the best results for all model errors.

$\gamma_1 \setminus U_i$	$\mathcal{N}\left(0, \frac{3}{2}\right)$			$Log\left(0, \frac{\sqrt{2}\pi}{3}\right)$			$Lap\left(0, \frac{\sqrt{3}}{2}\right)$			$t(6)$		
−0.3	97.07	97.68	83.82	97.76	98.00	86.11	98.65	98.41	90.42	97.71	97.93	86.72
−0.2	85.60	87.02	66.75	88.70	88.14	72.43	91.68	89.65	80.52	89.53	88.75	73.92
−0.1	57.03	54.66	38.04	57.43	55.54	44.54	66.22	61.11	58.28	59.60	57.27	46.91
0	26.32	26.45	19.27	29.12	27.65	22.88	36.00	31.53	34.54	30.40	28.90	24.77
0.1	15.62	15.85	11.90	17.02	16.43	13.85	20.89	18.47	21.40	18.08	16.83	14.55
0.2	11.62	11.18	9.63	12.21	11.90	10.54	14.63	13.22	15.10	13.13	12.25	11.18
0.3	9.50	9.49	8.08	9.86	9.53	8.70	11.67	10.46	11.93	10.73	10.23	9.14

Table 4.5: Percentage of rejections of hypothesis $\mathbf{H}_2$: $\beta = 0$ for various model errors U_i and for three choices of score function φ (Wilcoxon scores, van der Waerden scores, sign scores), under fixed $\beta = 0.05$ and $\gamma_0 = 0$, and for various γ_1; $n = 100$.

Next we shall compare the regression rank score tests with their classical rank test counterparts: namely with the rank tests for regression ignoring the heteroscedasticity, and the rank tests for regression considering the nuisance parameter $\boldsymbol{\gamma}$ to be known. We consider Wilcoxon scores generated by score function $\varphi^{(1)}(t)$, and model errors U_i generated from standard normal distribution. The remaining design is the same as in the previous cases.

$\gamma_1 \setminus \beta$	0			0.01			0.02			0.03		
−0.3	5.26	4.79	5.21	18.64	85.42	34.86	53.40	99.96	73.65	81.48	100	90.90
−0.2	4.85	4.82	4.99	10.58	24.38	14.54	29.22	73.75	40.15	54.57	96.39	67.32
−0.1	5.19	4.82	5.26	7.23	8.70	7.94	16.42	22.26	17.91	31.32	42.39	33.51
0	4.55	4.55	4.55	6.13	6.13	6.34	9.75	9.75	9.95	16.74	16.74	16.36
0.1	4.84	5.07	5.31	5.71	5.40	5.73	7.03	8.16	7.52	9.77	11.83	10.64
0.2	4.70	4.69	4.91	5.35	5.56	5.48	5.88	8.57	6.64	6.97	12.57	9.01
0.3	4.63	5.06	5.02	5.02	5.75	5.27	4.99	9.13	5.72	5.56	13.09	6.79

Table 4.6: Percentage of rejections of hypothesis $\mathbf{H}_2 : \beta = 0$ for classical rank test ignoring the heteroscedasticity, classical rank test considering $\boldsymbol{\gamma}$ to be known, and regression rank score test with Wilcoxon scores (in this order) for normal model errors and for fixed $\gamma_0 = 0$; $n = 100$.

For small values of parameters $\boldsymbol{\gamma}$ and β, the regression rank score test and the classical rank test, considering $\boldsymbol{\gamma}$ to be known, have very similar powers. The power of regression rank score test is a bit smaller for larger values $\boldsymbol{\gamma}$ and β. The classical rank test ignoring heteroscedasticity has probability of the error of the first kind equal to prescribed α, but its power is smaller than the power of proposed signed regression rank score test.

4.5.3 Application to real data

Consider the data from Ezekiel and Fox (1959), giving the speed of $n = 63$ cars and the distances taken to stop.

Denote Y_i the distance (ft) taken to stop and z_i the speed (mph) of the i-th car, $i = 1, \ldots, n$. Due to the physical nature of the situation, the relationship between Y_i and z_i is quadratic, in addition zero speed means zero stopping distance, hence there is no intercept in our model:

$$Y_i = \beta z_i^2 + e^{\gamma z_i} U_i, \; i = 1, \ldots, n. \tag{4.41}$$

Figure 4.1 illustrates the relation between the data, when parameter β is estimated by an R-estimator generated by Wilcoxon scores. We also plot a graph of fitted values and speed in Figure 4.1; that indicates that there might be a heteroscedasticity in our data.

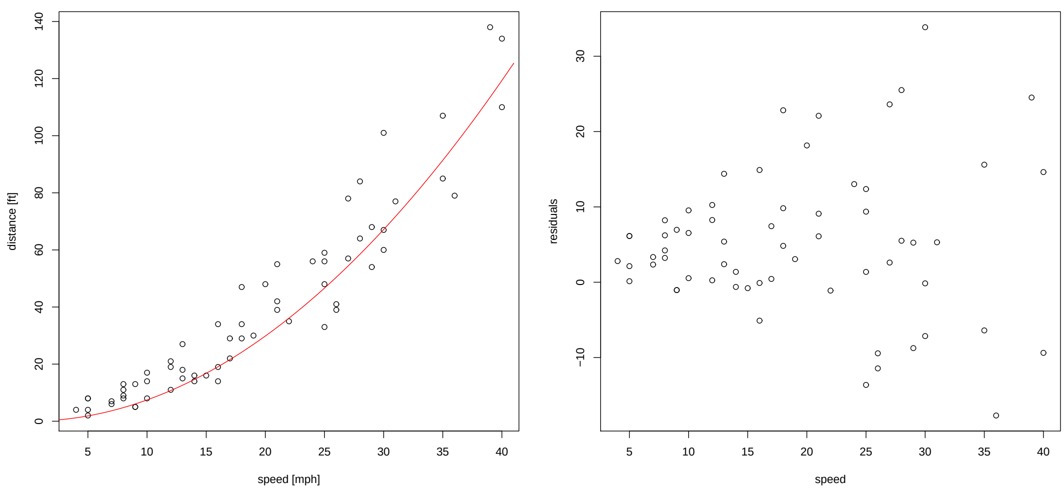

Figure 4.1: Data and their quadratic relationship; relationship between residuals and speed.

Let us start with test of homoscedasticity $\mathbf{H}_1 : \gamma = 0$ against $\gamma > 0$. Compute test statistic (4.10) with Wilcoxon scores for scale, i.e. generated by score function $\varphi(t) = -1 + (2t - 1) \log \frac{t}{1-t}$. The corresponding p-value is 0.043; hence, at usual level of significance $\alpha = 0.05$ we reject $\mathbf{H}_1$ and heteroscedasticity has to be admitted.

On the other hand, consider the test of significance of regression with nuisance heteroscedasticity, $\mathbf{H}_2 : \beta = 0$ against $\beta > 0$, with the test statistic (4.36) with Wilcoxon scores generated by score function $\varphi(t) = t - \frac{1}{2}$. It leads to the p-value 0.0085, hence the regression is really present in model (4.41).

5. Rank tests for location parameter

Unlike the previous chapters, where we dealt with linear regression model, in the following chapters we will discuss a bit simpler model with no regressors, namely location model.

Let $X_1, \ldots, X_n$ be i.i.d. random variables with an unknown continuous distribution function $F(x - \Delta)$ and continuous density $f(x - \Delta)$, where f is supposed to be symmetric around zero, i.e. $f(y) = f(-y)$ for all y, and has finite Fisher information with respect to the location $I(f)$. We want to test the hypothesis

$$\mathbf{H}_0 : \ \Delta = 0, \quad \text{against} \quad \mathbf{K}_0 : \ \Delta > 0.$$

One may rewrite the model as

$$X_i = \Delta + e_i, \quad i = 1, \ldots, n, \tag{5.1}$$

where e_i are i.i.d. random variables with an unknown symmetric density f with finite Fisher information $I(f)$.

Remark. *According to the terminology used by Hájek and Šidák (1967) we will call the following tests as rank tests of symmetry, although they actually do not test symmetry of underlying distribution, but they only assume it.*

5.1 Rank tests of symmetry

Again, we choose a nondecreasing, nonconstant, square integrable score function $\varphi : (0, 1) \mapsto \mathbb{R}$, consider

$$\varphi^+(u) = \varphi\left(\frac{u+1}{2}\right), \quad 0 < u < 1$$

and define approximate scores based on φ^+ as

$$a_n^+(i) = \varphi^+\left(\frac{i}{n+1}\right).$$

Let R_i^+ be the rank of $|X_i|$ among $|X_1|, \ldots, |X_n|$ and finally consider test statistic

$$S_n^+ = n^{-1/2} \sum_{i=1}^{n} a_n^+(R_i^+) \operatorname{sign}(X_i). \tag{5.2}$$

For deriving the exact distribution of S_n^+ under null hypothesis denote N_+ number of positive components of $(X_1, \ldots, X_n)$. Then the distribution of N_+ under $\mathbf{H_0}$ is binomial $Bi(n, 1/2)$. Further denote $\overline{R}_1^+ < \ldots < \overline{R}_{N_+}^+$ the ordered ranks corresponding to $(X_1, \ldots, X_n)$ which signs are positive. Then it holds

$$P(\overline{R}_1^+ = r_1, \ldots, \overline{R}_{N_+}^+ = r_{N_+}, N_+ = n_+ | \mathbf{H}_0) = \left(\frac{1}{2}\right)^n$$

for any n_+-tuple $(r_1, \ldots, r_{n_+})$, $1 \leq r_1 < \ldots < r_{n_+} \leq n$. Hence we can get the critical region of the test of the size α. For each $0 \leq m \leq n$ and for each of $\binom{n}{m}$ possible choices of positive signs of n-tuple $(r_1 = 1, \ldots, r_n = n)$ we calculate the value of the statistic S_n^+ and order these values in the increasing magnitude. The critical region is formed by $k = \lfloor \alpha \cdot 2^n \rfloor$ largest sums and the combination leads to (k+1)-st largest value can be possibly randomized.

Theorem 5.1. *Assume that f is symmetric with $I(f) < \infty$. Then in model (5.1) under $\mathbf{H}_0$ test statistic S_n^+ has asymptotically as $n \to \infty$ normal distribution $\mathcal{N}(0, A^2(\varphi^+))$.*

Proof. See Hájek et al. (1999, Theorem 1, Section 6.1.7). □

Consider a sequence of local alternatives

$$\mathbf{K}_{0,n}: \ \Delta = n^{-1/2}\Delta^*, \quad 0 < \Delta^* \text{ fixed.}$$

Theorem 5.2. *Assume that f is symmetric with $I(f) < \infty$. Then in model (5.1) under $\mathbf{K}_{0,n}$ test statistic S_n^+ has asymptotically as $n \to \infty$ normal distribution $\mathcal{N}(\mu, A^2(\varphi^+))$, where*

$$\mu = \Delta^* \gamma(\varphi^+, f).$$

Proof. See Hájek et al. (1999, Theorem 1, Section 7.2.5). □

From the previous theorem we can obtain asymptotic power of this test:

$$\beta = \beta(\Delta^*) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\Delta^* \gamma(\varphi^+, f)}{A(\varphi^+)}\right), \tag{5.3}$$

where $\Phi(x)$ denotes the distribution function of standard normal distribution.

This procedure is mainly used for testing homogeneity in two populations. For example we want to compare two treatments — we divide experimental objects into n homogeneous pairs (to exclude effects due to the inhomogeneity of the data) and apply the new treatment to one unit of the pair while the other one is control.

We may restate the previous problem as: $(X_1^{(T)}, X_1^{(C)}), \ldots, (X_n^{(T)}, X_n^{(C)})$ being i.i.d. two-dimensional random vectors with the distribution function $F(x, y)$; $X_i^{(T)}$ being treatment observations and $X_i^{(C)}$ control observations. Our aim is to test the hypothesis that the distribution function $F(x, y)$ is symmetric around the straight line $y = x$, i.e.

$$\mathbf{H}: \quad F(x, y) = F(y, x), \quad \forall x, y \in \mathbb{R},$$

against the alternative that the distribution of random vector $(X^{(T)}, X^{(C)})$ is shifted toward positive half-plane $y > x$.

Now, introduce new random variables $X_i = X_i^{(C)} - X_i^{(T)}$ and $G(z)$ its distribution function. Suppose that these exists a continuous density $g(z)$. The problem of testing $\mathbf{H}$ is then equivalent to stating that the distribution G is symmetric around 0, against the alternative, that the distribution is symmetric around $\Delta > 0$.

5.2 Rank tests of symmetry with additive measurement errors

Now, suppose that we observe instead of X_i random variables $Z_i = X_i + W_i$, $i = 1, \ldots, n$, where W_i are i.i.d. random variables independent with $X_1, \ldots, X_n$ with an unknown continuous density g symmetric around 0. We have

$$\begin{aligned} X_i &= \Delta + e_i, \\ Z_i &= X_i + W_i. \end{aligned}$$

Hence we may rewrite it as

$$Z_i = \Delta + e_i^*, \quad i = 1, \ldots, n, \tag{5.4}$$

where $e_i^* = e_i + W_i$ are i.i.d. random variables with density h.

Lemma 5.1. *Density of e_i^* is*

$$h(x) = \int_{-\infty}^{\infty} f(x - v)g(v)dv, \tag{5.5}$$

it is symmetric and has finite Fisher information with respect to the location.

Proof. Formula for density h follows from the convolution theorem and finiteness of Fisher information from Lemma 1.2.

Thanks to symmetry of f and g we have

$$\begin{aligned} h(-x) &= \int_{-\infty}^{\infty} f(-x - v)g(v)dv = \int_{-\infty}^{\infty} f(x + v)g(-v)dv \\ &= -\int_{\infty}^{-\infty} f(x - u)g(u)du = h(x), \quad \forall x \in \mathbb{R}. \end{aligned}$$

□

Denote $S_{Z,n}^+$ test statistic (5.2) based on observed values Z_i. The presence of measurement errors may change both ranks R_i^+ and signs of X_i, fortunately this change does not affect hypothetical distribution of $S_{Z,n}^+$. Actually, under $\mathbf{H}_0$ we have

$$Z_i = e_i^*$$

that has a continuous and symmetric density. Hence the exact distribution of $S_{Z,n}^+$ under $\mathbf{H}_0$ is the same as S_n^+. We can also formulate analogies of Theorems 5.1 and 5.2. Their proofs are immediate corollaries of Theorem 5.1, resp. 5.2 and Lemma 5.1.

Theorem 5.3. *Assume that f and g are symmetric and $I(f) < \infty$. Then in model (5.4) under $\mathbf{H}_0$ test statistic $S_{Z,n}^+$ has asymptotically as $n \to \infty$ normal distribution $\mathcal{N}(0, A^2(\varphi^+))$.*

$S_{Z,n}^+$ has under $\mathbf{K}_{0,n}$ asymptotically normal distribution $\mathcal{N}(\mu_Z, A^2(\varphi^+))$, where

$$\mu_Z = \Delta^* \gamma(\varphi^+, h).$$

Definition. *Let T_1 and T_2 be two tests for $\boldsymbol{H}$ against $\boldsymbol{K}$ such that T_1 has asymptotically $\mathcal{N}(0,\sigma_1^2)$ distribution under $\boldsymbol{H}$ and $\mathcal{N}(\mu_1,\sigma_1^2)$ distribution under $\boldsymbol{K}$ and T_2 has asymptotically $\mathcal{N}(0,\sigma_2^2)$ distribution under $\boldsymbol{H}$ and $\mathcal{N}(\mu_2,\sigma_2^2)$ distribution under $\boldsymbol{K}$. Then the number*

$$ARE(T_1,T_2)=\left(\frac{\mu_1\sigma_2}{\mu_2\sigma_1}\right)^2$$

will be called asymptotic relative efficiency of T_1-test relative to T_2-test.

Hence for our case we get

$$ARE(S^+_{Z,n},S^+_n)=\left(\frac{\mu_Z}{\mu}\right)^2=\left(\frac{\gamma(\varphi^+,h)}{\gamma(\varphi^+,f)}\right)^2. \tag{5.6}$$

According to the Cauchy-Schwarz inequality

$$ARE(S^+_{Z,n},S^+_n)\le\frac{I(h)A^2(\varphi^+)}{\gamma^2(\varphi^+,f)}.$$

If the test S^+_n with score function φ^+ is asymptotically optimal for f, i.e. if $\varphi^+(u)=\tilde{\varphi}^+(u,f)$, then

$$ARE(S^+_{Z,n},S^+_n)\le\frac{I(h)}{I(f)}\le 1.$$

We can look easily that the case of testing effects of two treatments with measurement errors is a special case of the previous one. Suppose that instead of $X_i^{(T)}$ we observe $\widetilde{X}_i^{(T)}=X_i^{(T)}+A_i$, where A_i are i.i.d. random variables independent with $X_1^{(T)},\dots,X_n^{(T)}$ with a continuous density symmetric around 0. And analogously instead of $X_i^{(C)}$ we observe $\widetilde{X}_i^{(C)}=X_i^{(C)}+B_i$, where B_i are i.i.d. random variables independent with $X_1^{(C)},\dots,X_n^{(C)}$ and $A_1,\dots,A_n$ with a continuous density symmetric around 0. Denote $\widetilde{X}_i=\widetilde{X}_i^{(T)}-\widetilde{X}_i^{(C)}=X_i^{(T)}-X_i^{(C)}+(A_i-B_i)$ for $i=1,\dots,n$. We can see that we are in the same situation as at the beginning of the section: we test the symmetry of r.v. $X_i=X_i^{(T)}-X_i^{(C)}$ with the presence of measurement errors $W_i=A_i-B_i$. Note that this case contains also the situation when both measurement errors A_i,B_i have distribution (not necessarily the same) symmetric about the same point $\delta>0$, or generally when the distribution of W_i is symmetric.

The following example comes from Lehmann (1975), but the original data are from Ury and Forrester (1970).

Example. *In a study of the comparative tensile strength of tape-closed and sutured wounds, the following results were obtained on 10 rats, 40 days after incisions made on their backs had been closed by suture or by surgical tape. The results are summarized in Table 5.1.*
We test the hypothesis of no difference between tape-closed and sutured wounds against the alternative that the tape-closed wounds are stronger. We illustrate it on the Wilcoxon scores $\varphi(u)=u-1/2$. Because the sample size is small

Rat	1	2	3	4	5	6	7	8	9	10
Tape	659	984	397	574	447	479	676	761	647	577
Suture	452	587	460	787	351	277	234	516	577	513
Difference	207	397	−63	−213	96	202	442	245	70	64

Table 5.1: Comparative tensile strength (lb. per sq. in.) of tape-closed and sutured wounds of rats.

Error	0	$\mathcal{N}$(0,100)	$\mathcal{N}$(0,400)	$\mathcal{U}$(−30,30)	$\mathcal{C}$(0,5)	$\mathcal{C}$(0,20)
p-value	0.0244	0.0271	0.0285	0.0283	0.0346	0.0564

Table 5.2: Effect of the measurement errors on p-value of Wilcoxon signed-rank test.

$(n = 10)$, *we use the exact distribution of* S_n^+ *and because we want to know exact p-value of this test, we use the non-randomized test. The p-value of this test is* $25/2^{10} \doteq 0.0244$, *that means that at the level of significance* $\alpha = 0.05$ *we reject the hypothesis of no difference between this two treatments.*

Now, suppose that we observe the differences of both treatments with a measurement error (it can arise from both measurement simultaneously). We produce 10 000 *replications, every time we contaminate the original data with random errors and compute respective p-value. Finally we estimate the p-value as the mean of p-values of these* 10 000 *replications. The results are summarized in Table* 5.2.

We can see that measurement errors increase the p-value and even errors with large variance can mask the effect of new treatment, so that we would not reject the hypothesis $\mathbf{H}_0$.

5.3 Rank tests of symmetry with additive shifted measurement errors

In the previous section we assumed that the measurement errors are symmetric around 0 – errors are random and not systematic. That assumption was crucial for construction of the test. Now, let the measurement errors $\widetilde{W}_i$ be symmetric around $\Delta_0 \neq 0$ with density $\widetilde{g}(x) = g(x - \Delta_0)$. Again, we consider model

$$\begin{aligned} X_i &= \Delta + e_i, \\ \widetilde{Z}_i &= X_i + \widetilde{W}_i. \end{aligned}$$

With the aid of the notation from the previous section, we can write

$$\widetilde{Z}_i = \Delta + \Delta_0 + e_i^*, \quad i = 1, \dots, n, \tag{5.7}$$

where $e_i^* = e_i + W_i$ are i.i.d. random variables with density h given in (5.5).

Denote $S^+_{\widetilde{Z},n}$ test statistic (5.2) based on observed values $\widetilde{Z}_i$. Again, the presence of measurement errors may change both ranks R_i^+ and signs of X_i, unfortunately in this case this change does affect hypothetical distribution of $S^+_{\widetilde{Z},n}$. Actually, under $\mathbf{H}_0$ we have

$$\widetilde{Z}_i = \Delta_0 + e_i^*.$$

Distribution of $\widetilde{Z}_i$ is then symmetric around Δ_0 that is generally unknown. Hence we are unable to do any statistical inference about the parameters. We are only able to describe the situation where the systematic errors are close to zero, in that sense that they are only local:

$$\Delta_0 = n^{-1/2}\Delta_0^*. \tag{5.8}$$

Theorem 5.4. *Assume that f and g are symmetric and $I(f) < \infty$. Then in model (5.7) the test based on statistic $S^+_{Z,n}$ under (5.8) achieves asymptotically error of the first kind*

$$\alpha^* = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\Delta_0^*\gamma(\varphi^+, h)}{A(\varphi^+)}\right). \tag{5.9}$$

Proof. Combining results from Theorem 5.3 for $\mathbf{K}_{0,n} : \Delta_0 = n^{-1/2}\Delta_0^*$ and (5.3) we get the expression (5.9) for the asymptotic error of the first kind for the test $S^+_{\widetilde{Z},n}$. □

From the previous formula we can see that if $\Delta 0 < 0$, then $\alpha^* < \alpha$ (the real value of the error of the first kind is lower than prescribed) and if $\Delta_0 > 0$, then $\alpha^* > \alpha$ (the real value is greater than prescribed).

Remark. *If we consider the opposite alternative $\mathbf{K}_0' : \Delta < 0$, the situation is symmetric and for the both-sided alternative $\mathbf{K}_0'' : \Delta \neq 0$ the real value is always greater than prescribed.*

5.4 Rank tests of symmetry with additive asymmetric measurement errors

The situation with asymmetric measurement errors is more complicated than the previous cases. Hence we will first discuss the model without measurement errors (5.1). Recall that it is

$$X_i = \Delta + e_i, \quad i = 1, \ldots, n,$$

where here e_i are i.i.d. random variables with an unknown asymmetric density f with finite Fisher information $I(f)$ and median 0.

The symmetry assumption is crucial for rank tests of symmetry, see the following lemma.

Lemma 5.2. *Let X be a symmetric random variable with density f, then* $\operatorname{sign} X$ *and $|X|$ are independent.*

Proof. Thanks to symmetry of X

$$\begin{aligned} P(\operatorname{sign} X = 1) &= P(\operatorname{sign} X = -1) = \frac{1}{2}, \\ P(-x < X < 0) &= P(0 < X < x), \quad \forall x > 0. \end{aligned}$$

Then for any $x > 0$

$$\begin{aligned} P(\operatorname{sign} X = 1, |X| < x) &= P(0 < X < x) = \frac{1}{2} P(-x < X < x) \\ &= P(\operatorname{sign} X = 1) P(|X| < x). \end{aligned}$$

and analogously

$$P(\operatorname{sign} X = -1, |X| < x) = P(\operatorname{sign} X = -1) P(|X| < x), \quad \forall x > 0$$

that completes the proof. □

Hence if the symmetry assumption is not satisfied Lemma 5.2 will not be valid and the test statistic S_n^+ from (5.2) will not have required properties. Only exception it might be the choice $\varphi(u) = \operatorname{sign}(2u - 1)$, so called *sign test*. Then

$$\varphi^+(u) = \varphi\left(\frac{u+1}{2}\right) = \operatorname{sign}(u) = 1, \quad 0 < u < 1.$$

Hence corresponding test statistic S_n^+ is

$$S_n^+ = n^{-1/2} \sum_{i=1}^{n} \operatorname{sign}(X_i).$$

Anyway, more known is an equivalent test statistic

$$T_n = \sum_{i=1}^{n} \mathbb{I}\{X_i > 0\}$$

number of positive components of $(X_1, \ldots, X_n)$. Actually

$$S_n^+ = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{n} \mathbb{I}\{X_i > 0\} - \sum_{i=1}^{n} \mathbb{I}\{X_i < 0\} \right] = \frac{2T_n}{\sqrt{n}} - \sqrt{n},$$

where we used

$$\sum_{i=1}^{n} \mathbb{I}\{X_i < 0\} + \sum_{i=1}^{n} \mathbb{I}\{X_i > 0\} = n.$$

According to Theorem 5.1 S_n^+ has asymptotically standard normal distribution $\mathcal{N}(0, 1)$, because $A^2(\varphi) = 1$. In general this test will work for all distributions with 0 median (and finite Fisher information).

Anyway, return back to our problem. The problem here is that if we add independent asymmetric measurement error (with 0 median) to symmetric random variable, it is not assured that median of the resultant distribution will be 0, see the following example.

Example. *Let X and Y be independent random variables, X with uniform distribution $\mathcal{U}(-1/2, 1/2)$ and Y with shifted exponential distribution with density*

$$g(y) = e^{-(y+\log 2)}\mathbb{I}\left\{y > -\log 2\right\}.$$

Denote $Z = X + Y$ their convolution. Apparently X is symmetric around 0 while Y is asymmetric but with median 0.

According to well-known formula for density of Z we have

$$\begin{aligned} h(z) &= \left(1 - e^{-z-\frac{1}{2}-\log 2}\right)\mathbb{I}\left\{-\log 2 - \frac{1}{2} < z < -\log 2 + \frac{1}{2}\right\} \\ &+ \left(e^{-z+\frac{1}{2}-\log 2} - e^{-z-\frac{1}{2}-\log 2}\right)\mathbb{I}\left\{z > -\log 2 + \frac{1}{2}\right\}. \end{aligned}$$

Thus, Z is not symmetric and its median is approximately 0.042.

Hence, even if we used the sign test and ignored measurement errors it would be biased. We do not put any assumption of the knowledge of the distribution of errors, that is why, it is difficult to express mathematically the influence of measurement errors. Anyway, our simulations indicate that measurement errors may change error of the first kind of the corresponding test. If we ignore the measurement errors and use the critical value for the test without measurement errors, the error of the first kind will be less (greater) than α (here, it depends on the shape of distribution of measurement errors).

5.5 Simulations

We have made a simulation study to show how the previous tests perform in finite sample situation ($n = 20$). We tested the hypothesis $\Delta = 0$ against the alternative $\Delta > 0$. We have also compared power of t-test and considered rank tests for various choices of the score function φ:

$(T1)$	sign test	$\varphi(u) = \text{sign}\left(u - \frac{1}{2}\right)$	$A^2(\varphi^+) = 1$
$(T2)$	Wilcoxon test	$\varphi(u) = u - \frac{1}{2}$	$A^2(\varphi^+) = \frac{1}{12}$
$(T3)$	normal scores test	$\varphi(u) = \Phi^{-1}(u)$	$A^2(\varphi^+) = 1$

In Tables 5.3 – 5.6 power of the tests is estimated based on percentage of rejections (based on 10 000 replications) for various distributions of the random variables X_i as well as for various distributions of measurement errors V_i. In Table 5.7 actual error of the first kind is estimated the same way for normal distributed variables X_i for various shifted measurement errors. The level of significance was $\alpha = 0.05$.

The simulation study indicates the influence of measurement errors on the power of used tests. For distributions with short tails t-test arises similar results as rank tests, but for heavy-tailed distributions the power of t-test is much lower than the power of rank tests. The power of tests decreases with increasing variance of measurement errors and it also depends on the choice of score function φ. Wilcoxon scores arise the best results for short-tailed distributions, but sign test arises the best results for heavy-tailed distributions. Normal scores do not arise

Test \ W_i	0	$\mathcal{N}(0, 0.5)$	$\mathcal{U}(-2, 2)$	$Log(0, 0.5)$	$Lap(0, 0.5)$	$\mathcal{C}(0, 0.5)$	$t(2)$
t	70.13	54.03	40.84	48.50	54.72	22.02	26.56
T1	58.49	45.50	31.25	41.43	47.24	36.73	30.56
T2	67.88	52.69	38.27	47.08	53.28	36.27	31.14
T3	64.59	48.74	35.64	43.06	49.20	29.41	26.23

Table 5.3: Percentage of rejections based on 10 000 replications for various distributions of measurement errors W_i and various tests, sample size $n = 20$ and the original variables X_i with $\mathcal{N}(0.5, 1)$.

Test \ W_i	0	$\mathcal{N}(0, 0.5)$	$\mathcal{U}(-2, 2)$	$Log(0, 0.5)$	$Lap(0, 0.5)$	$\mathcal{C}(0, 0.5)$	$t(2)$
t	44.62	38.19	30.79	34.71	37.76	18.04	21.96
T1	51.32	39.14	26.96	35.54	40.99	32.25	26.98
T2	51.92	41.39	30.86	37.37	41.72	29.50	26.21
T3	44.34	35.48	27.11	32.05	35.90	23.63	21.40

Table 5.4: Percentage of rejections based on 10 000 replications for various distributions of measurement errors W_i and various tests, sample size $n = 20$ and the original variables X_i with $t(3) + 0.5$.

Test \ W_i	0	$\mathcal{N}(0, 0.5)$	$\mathcal{U}(-2, 2)$	$Log(0, 0.5)$	$Lap(0, 0.5)$	$\mathcal{C}(0, 0.5)$	$t(2)$
t	34.15	30.36	26.68	28.93	30.34	16.30	20.85
T1	32.95	28.60	24.10	27.10	28.74	24.27	22.12
T2	34.94	30.58	26.19	28.68	30.82	23.35	22.15
T3	30.41	26.60	23.03	24.92	26.26	19.13	19.05

Table 5.5: Percentage of rejections based on 10 000 replications for various distributions of measurement errors W_i and various tests, sample size $n = 20$ and the original variables X_i with $Log(0.5, 1)$.

Test \ W_i	0	$\mathcal{N}(0, 0.5)$	$\mathcal{U}(-2, 2)$	$Log(0, 0.5)$	$Lap(0, 0.5)$	$\mathcal{C}(0, 0.5)$	$t(2)$
t	49.28	40.64	33.26	37.59	40.98	18.77	28.04
T1	59.88	41.65	27.76	37.96	44.77	34.75	30.10
T2	56.24	43.09	32.74	39.80	44.31	30.43	30.13
T3	48.31	37.25	29.11	34.14	38.15	24.50	25.29

Table 5.6: Percentage of rejections based on 10 000 replications for various distributions of measurement errors W_i and various tests, sample size $n = 20$ and the original variables X_i with $Lap(0.5, 1)$.

Test\\W_i	0	$\mathcal{C}(0,1)$	$t(2)$	$Log(0.2,1)$	$Log(-0.4,1)$	$Log(0.4,1)$	$\mathcal{U}(-1.8,2.2)$	$Lap(0.2,1)$
t	5.16	3.17	4.95	10.93	21.74	1.93	13.96	13.18
T1	5.99	5.76	5.84	11.52	20.67	2.53	12.98	13.74
T2	5.02	4.98	5.33	11.10	21.66	1.88	13.56	13.19
T3	4.06	3.94	4.23	9.06	18.21	1.37	11.37	10.75

Table 5.7: Percentage of rejections based on 10 000 replications for various distributions of measurement errors W_i and various tests, sample size $n = 20$ and the original variables X_i with $\mathcal{N}(0,1)$.

good results nor for the normal distribution. It is caused by the small sample size (convergence for normal scores is very slow), for large n it has similar power as t-test for short-tailed distributions, but it is more robust for heavy-tailed distributions.

6. R-estimates of location parameter

In this chapter we will stay at the location model and we will try to estimate the location parameter Δ.

6.1 R-estimates in model without measurement errors

Again, let us start with the model without measurement errors. Remind that we study model (5.1):

$$X_i = \Delta + e_i, \quad i = 1, \dots, n,$$

where e_i are i.i.d. random variables with an unknown symmetric density f with finite Fisher information $I(f)$. The problem of estimation of location parameter Δ is in fact dual to the testing problem, hence there will be a lot of similarities with the previous chapter.

Analogously as in Section 5.1 we choose a nondecreasing, nonconstant, square integrable score function $\varphi : (0,1) \mapsto \mathbb{R}$, consider

$$\varphi^+(u) = \varphi\left(\frac{u+1}{2}\right), \quad 0 < u < 1$$

and define approximate scores based on φ^+ as

$$a_n^+(i) = \varphi^+\left(\frac{i}{n+1}\right).$$

For $t \in \mathbb{R}$ define $R_i^+(t)$ the rank of $|X_i - t|$ among $|X_1 - t|, \dots, |X_n - t|$ and consider

$$S_n^+(t) = n^{-1/2} \sum_{i=1}^{n} a_n^+(R_i^+(t)) \operatorname{sign}(X_i - t). \tag{6.1}$$

Remark. *Note that for $t = 0$ (6.1) reduces to (5.2) used for testing the hypothesis $\Delta = 0$.*

As an estimator of Δ it is proposed the value of t which solves the equation $S_n^+(t) = 0$. Since $S_n^+(t)$ is discontinuous, such an equation may have no solution; then we define the R-estimator of Δ as

$$\hat{\Delta}_{n,(X)}^{(R)} = \hat{\Delta}_n^{(R)} = \frac{1}{2}\left[\sup\{t : S_n^+(t) > 0\} + \inf\{t : S_n^+(t) < 0\}\right].$$

Remark. *In general, the value of $\hat{\Delta}_n^{(R)}$ cannot be expressed by a formula, that's why we have to use some numerical method for finding $\hat{\Delta}_n^{(R)}$, but for special choices of the scores a_n^+ we can get the accurate expression. If $a_n^+(i) = 1$ for all $i = 1, \dots, n$, then $\hat{\Delta}_n^{(R)} = \hat{\Delta}_n^{(med)} = \operatorname{med}\{X_1, \dots, X_n\}$ and if $a_n^+(i) = \frac{i}{n+1}$ we have $\hat{\Delta}_n^{(R)} = \hat{\Delta}_n^{(H-L)} = \operatorname{med}\left\{\frac{X_i + X_j}{2}, 1 \le i \le j \le n\right\}$ (Hodges–Lehmann estimator).*

Theorem 6.1. *Assume that f is symmetric with $I(f) < \infty$. Then*

$$\sqrt{n}(\hat{\Delta}_n^{(R)} - \Delta) \xrightarrow{d} \mathcal{N}\left(0, \frac{A^2(\varphi^+)}{\gamma^2(\varphi^+, f)}\right).$$

Proof. The proof follows from asymptotic representation for $\hat{\Delta}_n^{(R)}$ (see Jurečková and Sen (1996, page 244)):

$$\sqrt{n}(\hat{\Delta}_n^{(R)} - \Delta) = \frac{1}{\sqrt{n}\gamma(\varphi^+, f)} \sum_{i=1}^{n} \varphi^+(F(X_i - \Delta)) + o_p(1), \quad n \to \infty.$$

□

With the aid of Theorem 6.1 we can also express the asymptotic confidence interval for Δ with (asymptotic) confidence level $(1 - \alpha)$:

$$\left(\hat{\Delta}_n^{(R)} - n^{-\frac{1}{2}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{A(\varphi^+)}{\gamma(\varphi^+, f)}, \hat{\Delta}_n^{(R)} + n^{-\frac{1}{2}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{A(\varphi^+)}{\gamma(\varphi^+, f)}\right). \quad (6.2)$$

However, this confidence interval has one disadvantage – we do not know the density f and for practical computations it is necessary to estimate it, respectively one has to estimate $\gamma(\varphi^+, f)$.

Anyway, it also exists another approach based on the knowledge of distribution of $S_n^+(t)$. Let $C_{n,\alpha}$ be the smallest value for which the following inequality holds

$$P(|S_n^+(0)| \leq C_{n,\alpha} | \Delta = 0) \geq 1 - \alpha.$$

We can compute $C_{n,\alpha}$ from the exact distribution of $S_n^+(0) = S_n^+$, or we can use the asymptotic normal approximation mentioned in Section 5.1, i.e. we use

$$\widetilde{C}_{n,\alpha} = \sqrt{n}A(\varphi^+)\Phi^{-1}(1 - \alpha/2). \quad (6.3)$$

And finally define the confidence interval $\left(\hat{\Delta}_{L,n}^{(R)}, \hat{\Delta}_{U,n}^{(R)}\right)$ such that

$$\hat{\Delta}_{L,n}^{(R)} = \frac{1}{2}\left[\sup\{t : S_n^+(t) > \widetilde{C}_{n,\alpha}\} + \inf\{t : S_n^+(t) < \widetilde{C}_{n,\alpha}\}\right],$$

$$\hat{\Delta}_{U,n}^{(R)} = \frac{1}{2}\left[\sup\{t : S_n^+(t) > -\widetilde{C}_{n,\alpha}\} + \inf\{t : S_n^+(t) < -\widetilde{C}_{n,\alpha}\}\right]. \quad (6.4)$$

The previous approach might be summarized in the following theorem.

Theorem 6.2. *Assume that f is symmetric with $I(f) < \infty$. Then confidence interval $\left(\hat{\Delta}_{L,n}^{(R)}, \hat{\Delta}_{U,n}^{(R)}\right)$ for Δ has asymptotic coverage probability $(1 - \alpha)$.*

Remark. *Jurečková and Sen (1996) proved a connection between confidence interval (6.4) and the previous one (6.2), namely for $n \to \infty$*

$$\left(\hat{\Delta}_{L,n}^{(R)}, \hat{\Delta}_{U,n}^{(R)}\right) = \left(\hat{\Delta}_n^{(R)} - n^{-1/2}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{A(\varphi^+)}{\gamma(\varphi^+, f)} + o\left(n^{-1/2}\right),\right.$$
$$\left.\hat{\Delta}_n^{(R)} + n^{-1/2}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{A(\varphi^+)}{\gamma(\varphi^+, f)} + o\left(n^{-1/2}\right)\right).$$

In this case we do not need to estimate the variance $A^2(\varphi^+)/\gamma^2(\varphi^+, f)$, nor to use the point estimator $\hat{\Delta}_n^{(R)}$. Disadvantage of this procedure is that the solutions for $\hat{\Delta}_{L,n}^{(R)}$ and $\hat{\Delta}_{U,n}^{(R)}$ cannot be expresses and must be computed iteratively.

6.2 R-estimates of location parameter under additive measurement errors

Now, suppose that we observe instead of X_i random variables $Z_i = X_i + W_i$, $i = 1, \dots, n$, where W_i are i.i.d. random variables independent with $X_1, \dots, X_n$ with an unknown continuous density $g(v)$ symmetric around 0. We still want to estimate the parameter Δ.

Recall that $h(w) = \int_{-\infty}^{\infty} f(w - v)g(v)dv$ is density of Z_i and define

$$\widetilde{S}_n^+(t) = n^{-1/2} \sum_{i=1}^{n} a_n^+(\widetilde{R}_i^+(t)) \operatorname{sign}(Z_i - t),$$

where $\widetilde{R}_i^+(t)$ are the ranks of $|Z_i - t|$ among $|Z_1 - t|, \dots, |Z_n - t|$ and an estimator of Δ is

$$\hat{\Delta}_{n,(Z)}^{(R)} = \frac{1}{2}\left[\sup\{t : \widetilde{S}_n^+(t) > 0\} + \inf\{t : \widetilde{S}_n^+(t) < 0\}\right]. \tag{6.5}$$

Theorem 6.3. *Assume that f and g are symmetric and $I(f) < \infty$. Then*

$$\sqrt{n}(\hat{\Delta}_{n,(Z)}^{(R)} - \Delta) \xrightarrow{d} \mathcal{N}\left(0, \frac{A^2(\varphi^+)}{\gamma^2(\varphi^+, h)}\right).$$

Proof. Lemma 5.1 assures symmetry and finiteness of Fisher information of h. The proof then follows from Theorem 6.1. □

It also means that if we use the same setup for confidence interval as for model without measurement errors, the resulting confidence interval for Δ in measurement error model will remain valid. Hence define the confidence interval $\left(\hat{\Delta}_{L,n}^{(Z,R)}, \hat{\Delta}_{U,n}^{(Z,R)}\right)$ as

$$\hat{\Delta}_{L,n}^{(Z,R)} = \frac{1}{2}\left[\sup\{t : \widetilde{S}_n^+(t) > \widetilde{C}_{n,\alpha}\} + \inf\{t : \widetilde{S}_n^+(t) < \widetilde{C}_{n,\alpha}\}\right],$$

$$\hat{\Delta}_{U,n}^{(Z,R)} = \frac{1}{2}\left[\sup\{t : \widetilde{S}_n^+(t) > -\widetilde{C}_{n,\alpha}\} + \inf\{t : \widetilde{S}_n^+(t) < -\widetilde{C}_{n,\alpha}\}\right],$$

where $\widetilde{C}_{n,\alpha}$ was defined in (6.3).

Theorem 6.4. *Assume that f and g are symmetric and $I(f) < \infty$. Then confidence interval $\left(\hat{\Delta}_{L,n}^{(Z,R)}, \hat{\Delta}_{U,n}^{(Z,R)}\right)$ for Δ has asymptotic coverage probability* $(1-\alpha)$.

As well as for the tests, for comparing two estimates one may define asymptotic relative efficiency.

Definition. *Let $\hat{\Delta}_1$ and $\hat{\Delta}_2$ be two estimates of parameter Δ such that $\hat{\Delta}_1$ has asymptotically $\mathcal{N}(\Delta, \sigma_1^2)$ distribution and $\hat{\Delta}_2$ has asymptotically $\mathcal{N}(\Delta, \sigma_2^2)$.*

Then the number

$$ARE(\hat{\Delta}_1, \hat{\Delta}_2) = \frac{\sigma_2^2}{\sigma_1^2}$$

will be called asymptotic relative efficiency of $\hat{\Delta}_1$ relative to $\hat{\Delta}_2$.

The ARE $\hat{\Delta}^{(R)}_{n,(W)}$ relative to $\hat{\Delta}^{(R)}_{n,(X)}$ (R-estimate in model with measurement errors relative to R-estimate in model without measurement errors) is

$$ARE\left(\hat{\Delta}^{(R)}_{n,(W)}, \hat{\Delta}^{(R)}_{n,(X)}\right) = \frac{\gamma^2(\varphi^+, h)}{\gamma^2(\varphi^+, f)} = \left(\frac{\int_{\frac{1}{2}}^{1} \varphi(u)\widetilde{\varphi}(u,h)du}{\int_{\frac{1}{2}}^{1} \varphi(u)\widetilde{\varphi}(u,f)du}\right)^2. \tag{6.6}$$

Remark. *Since estimation and testing are dual problems, it is not very surprising that formula for asymptotic relative efficiency of two tests in model with and without measurement errors (5.6) coincides with the formula for asymptotic relative efficiency of two estimates in model with and without measurement errors (6.6).*

6.3 R-estimates of location parameter under additive shifted measurement errors

Now, as in Section 5.3 let the measurement errors $\widetilde{W}_i$ be symmetric around $\Delta_0 \neq 0$ with density $\widetilde{g}(x) = g(x - \Delta_0)$. Again, we may write

$$\widetilde{Z}_i = \Delta + \Delta_0 + e_i^*, \quad i = 1, \dots, n,$$

where $e_i^* = e_i + W_i$ are i.i.d. random variables with symmetric density h given in (5.5).

Because of the form of $\widetilde{Z}_i$ without any knowledge about Δ_0 it is impossible to estimate parameter Δ. If we try to consider the R-estimate given by the equation (6.5), we have to realize that the estimate $\hat{\Delta}^{(R)}_{n,(\widetilde{Z})}$ does not estimate parameter Δ, but $\Delta + \Delta_0$, hence $\hat{\Delta}^{(R)}_{n,(\widetilde{Z})}$ is not consistent estimate of Δ. This result is stated in the following theorem.

Theorem 6.5. *Assume that f and g are symmetric and $I(f) < \infty$. Then*

$$\sqrt{n}(\hat{\Delta}^{(R)}_{n,(\widetilde{Z})} - \Delta) \xrightarrow{d} \mathcal{N}\left(\Delta_0, \frac{A^2(\varphi^+)}{\gamma^2(\varphi^+, h)}\right).$$

Remark (Asymmetric measurement errors)**.** *The symmetry assumption was very important in this chapter due to identifiability of parameter Δ. Using the same arguments as in Section 5.4 we conclude that if the measurement errors are asymmetric, corresponding R-estimator will be biased. The bias depends on the shape of underlying distribution of measurement errors. This result was further supported in the simulation study.*

6.4 Simulations

We have also made a simulation study to show how the estimates perform for finite sample situation ($n = 20$). We compared mean ($\hat{\Delta}_n^{(LSE)}$), median ($\hat{\Delta}_n^{(med)}$),

Est.\ W_i	0	$\mathcal{N}(0,1)$	$\mathcal{U}(-3,3)$	$\mathcal{C}(0,1)$	$Lap(0,1)$	$t(2)$
$\hat{\Delta}_n^{(LSE)}$	0.004 (0.152)	0.000 (0.201)	0.001 (0.300)	1.041 (4287)	0.000 (0.250)	0.014 (1.061)
$\hat{\Delta}_n^{(med)}$	0.004 (0.224)	0.000 (0.294)	0.004 (0.491)	0.003 (0.518)	0.002 (0.335)	0.005 (0.386)
$\hat{\Delta}_n^{(H-L)}$	0.006 (0.161)	0.000 (0.212)	0.002 (0.332)	0.005 (0.518)	0.001 (0.252)	0.006 (0.312)
$\hat{\Delta}_n^{(\Phi)}$	0.005 (0.156)	0.000 (0.206)	0.001 (0.311)	0.008 (0.630)	0.000 (0.251)	0.007 (0.326)

Table 6.1: Estimates of $\Delta = 0$ and their variance based on 10 000 replications for various distributions of measurement errors W_i, sample size $n = 20$ and the original variables X_i with $\mathcal{N}(0, 3)$.

Est.\ W_i	0	$\mathcal{U}(-3,3)$	$\mathcal{C}(0,1)$	$Lap(0,1)$	$t(2)$
$\hat{\Delta}_n^{(LSE)}$	0.005 (0.145)	0.000 (0.292)	3.483 (77861)	-0.002 (0.243)	0.019 (1.104)
$\hat{\Delta}_n^{(med)}$	0.005 (0.090)	0.002 (0.443)	-0.003 (0.355)	0.003 (0.209)	0.006 (0.252)
$\hat{\Delta}_n^{(H-L)}$	0.004 (0.082)	-0.002 (0.277)	0.001 (0.417)	-0.002 (0.181)	0.005 (0.232)
$\hat{\Delta}_n^{(\Phi)}$	0.005 (0.091)	-0.003 (0.262)	0.002 (0.532)	-0.003 (0.194)	0.005 (0.260)

Table 6.2: Estimates of $\Delta = 0$ and their variance based on 10 000 replications for various distributions of measurement errors W_i, sample size $n = 20$ and the original variables X_i with $t(3)$.

Hodges–Lehmann estimator ($\hat{\Delta}_n^{(H-L)}$) and R-estimate based on the score function $\varphi(u) = \Phi^{-1}(u)$ (denote it $\hat{\Delta}_n^{(\Phi)}$). The last one has not been mentioned yet, because it cannot be expressed by some formula, but it must be computed iteratively (we used Newton's method).

In Tables 6.1 – 6.5 the estimates of parameter $\Delta = 0$ based on 10 000 replications are computed for various distributions of the random variables X_i as well as for various distributions of measurement errors W_i. For each estimator the estimate of its variance based on performed simulations is added (in parentheses).

The simulation study indicates that all considered estimators estimate parameter Δ approximately the same except from the mean in heavy-tailed distributions. If the original distribution X_i or the distribution of errors W_i is heavy-tailed, then mean fails. Much more interesting is the comparison of variances of

Est.\ W_i	0	$\mathcal{N}(0,1)$	$\mathcal{U}(-3,3)$	$\mathcal{C}(0,1)$	$Lap(0,1)$	$t(2)$
$\hat{\Delta}_n^{(LSE)}$	0.006 (0.156)	0.003 (0.200)	0.005 (0.308)	2.170 (57258)	0.005 (0.257)	0.014 (1.104)
$\hat{\Delta}_n^{(med)}$	0.007 (0.404)	0.005 (0.397)	0.008 (0.485)	0.007 (0.653)	0.005 (0.453)	0.001 (0.488)
$\hat{\Delta}_n^{(H-L)}$	0.007 (0.189)	0.003 (0.234)	0.006 (0.350)	0.004 (0.555)	0.007 (0.281)	0.007 (0.330)
$\hat{\Delta}_n^{(\Phi)}$	0.006 (0.139)	0.003 (0.205)	0.005 (0.318)	0.005 (0.657)	0.007 (0.259)	0.009 (0.331)

Table 6.3: Estimates of $\Delta = 0$ and their variance based on 10 000 replications for various distributions of measurement errors W_i, sample size $n = 20$ and the original variables X_i with $\mathcal{U}(-3, 3)$.

Est.\ W_i	0	$\mathcal{N}(0,1)$	$\mathcal{U}(-3,3)$	$\mathcal{C}(0,1)$	$Lap(0,1)$	$t(2)$
$\hat{\Delta}_n^{(LSE)}$	0.007 (0.417)	0.003 (0.448)	0.003 (0.569)	2.168 (57260)	0.003 (0.518)	0.014 (1.116)
$\hat{\Delta}_n^{(med)}$	0.006 (0.280)	0.005 (0.403)	0.009 (0.699)	0.008 (0.732)	0.006 (0.475)	−0.001 (0.518)
$\hat{\Delta}_n^{(H-L)}$	0.009 (0.305)	0.003 (0.366)	0.008 (0.526)	0.001 (0.816)	0.008 (0.432)	0.007 (0.495)
$\hat{\Delta}_n^{(\Phi)}$	0.009 (0.343)	0.004 (0.394)	0.007 (0.536)	0.001 (0.997)	0.007 (0.461)	0.010 (0.548)

Table 6.4: Estimates of $\Delta = 0$ and their variance based on 10 000 replications for various distributions of measurement errors W_i, sample size $n = 20$ and the original variables X_i with $Lap(0, 2)$.

Est.\ W_i	0	$\mathcal{N}(0,1)$	$\mathcal{U}(-3,3)$	$\mathcal{C}(0,1)$	$Lap(0,1)$	$t(2)$
$\hat{\Delta}_n^{(LSE)}$	0.005 (0.172)	0.002 (0.214)	0.004 (0.324)	2.168 (57258)	0.003 (0.273)	0.014 (0.880)
$\hat{\Delta}_n^{(med)}$	0.005 (0.199)	0.005 (0.276)	0.007 (0.506)	0.006 (0.512)	0.005 (0.328)	0.001 (0.363)
$\hat{\Delta}_n^{(H-L)}$	0.006 (0.161)	0.003 (0.212)	0.006 (0.347)	0.002 (0.537)	0.006 (0.261)	0.007 (0.309)
$\hat{\Delta}_n^{(\Phi)}$	0.006 (0.165)	0.002 (0.212)	0.005 (0.330)	0.003 (0.662)	0.006 (0.265)	0.009 (0.332)

Table 6.5: Estimates of $\Delta = 0$ and their variance based on 10 000 replications for various distributions of measurement errors W_i, sample size $n = 20$ and the original variables X_i with $Log(0, 1)$.

R-estimates. It depends on the choice of score function φ and the distributions of measurement errors and it increases with increasing variance of measurement errors. Hodges–Lehmann estimator and R-estimator based on normal scores arise the best accuracy for short-tailed distributions, median arises the best accuracy for heavy-tailed distributions. But in general we can say that Hodges–Lehmann estimator arises the best accuracy among other estimates for any measurement errors. More simulation results might be found in Navrátil and Saleh (2011) and Navrátil (2012).

Conclusion

In a broader sense all statistical problems involve measurement errors. In this thesis measurement errors occur when we cannot observe some variable of our interest exactly. Most often that might be caused by instrument or sampling error. Application of these models is wide in many branches from physics and chemistry to social science.

We studied the behavior of rank procedures in measurement error models. Although both measurement error models and rank tests and estimates are relatively old and well-known, their combination (application of rank procedures to measurement error models) is only a couple of years old. Hence most of the results are original, generalizations of those in Jurečková et al. (2010) and other applications of rank procedures in various measurement error models.

The main goal of the thesis was to investigate if classical rank tests and estimates stay valid and applicable when there are some measurement errors present and if not how to modify these procedures to be still able to do some statistical inference.

First, we proposed a new rank test for the slope parameter in regression model based on minimum distance estimator and an aligned rank test for an intercept. We also investigated the bias of R-estimator in measurement error model; the results correspond to those for least squares estimate, but R-estimates are more robust to departures from normal model. Besides measurement errors we also dealt with the problem of heteroscedastic model errors. We proposed regression rank score tests of heteroscedasticity with nuisance regression and tests of regression with nuisance heteroscedasticity without estimation of nuisance parameters. Finally, in location model tests and estimates of shift parameter were studied for various measurement errors.

Anyway, the idea of using rank tests in measurement error model is quite new. Hence there is still a lot of open problems that remain to be solved (and proven), some of them have been already mentioned in the thesis, some are still waiting to be found. This thesis should have demonstrated that it is reasonable to use rank tests due to their simplicity and only a weak set of assumptions for their validity together with the robustness and high efficiency characteristics, relative to the parametric methods. All the results were derived theoretically and then demonstrated numerically with examples or simulations.

List of Tables

List of Figures

List of Symbols

$\uparrow$	...	convergence from the left
$\downarrow$	...	convergence from the right
$\xrightarrow{d}$	...	convergence in distribution
$\xrightarrow{p}$	...	convergence in probability
$(\cdot)'$	...	derivative
$\square$	...	end of the proof
$\equiv$	...	equivalence
$\Vert \cdot \Vert$	...	Euclidean norm
$\lfloor \cdot \rfloor$	...	floor function
$\langle \cdot \rangle$	...	inner product
$\mathbf{1}_n$	...	n-dimensional vector of ones
$\nabla(\cdot)$	...	subgradient
$(\cdot)^\top$	...	transposition
$\Phi(\cdot)$	...	distribution function of $\mathcal{N}(0,1)$
$\Gamma(\cdot)$	...	Gamma function
$\Phi^{-1}(\cdot)$	...	quantile function of $\mathcal{N}(0,1)$
$\arg\max(\cdot)$	...	argument of the maximum
$\arg\min(\cdot)$	...	argument of the minimum
a.e.	...	almost everywhere
a.s.	...	almost surely
$\mathcal{C}[0,1]$	...	space of continuous functions
$\mathcal{D}[0,1]$	...	space of right continuous functions with left-hand limits
$\mathbb{E}(\cdot)$	...	expectation
$f(x_-)$	...	limit from left
$F^{-1}(\cdot)$	...	quantile function
$\mathbb{I}(\cdot)$	...	indicator function
i.i.d.	...	independent and identically distributed
LSE	...	least squares estimate
$\max(\cdot)$	...	maximum
$\mathrm{med}(\cdot)$	...	median
$\min(\cdot)$	...	minimum
$\mathbb{N}$	...	natural numbers
$o_p(1)$	...	negligibility in probability
$O_p(1)$	...	bounded in probability
$\mathbb{R}$	...	real numbers
$\mathbb{R}^p$	...	p-dimensional Euclidean space
$\mathbb{R}^p_+$	...	p-dimensional vector of nonnegative numbers
RRS	...	regression rank scores
$\mathrm{sign}(\cdot)$	...	signum
$\mathrm{var}(\cdot)$	...	variance

List of Distributions

Symbol	Distribution	Density
$Bi(n,p)$	Binomial distribution	$\binom{n}{k}p^k(1-p)^{n-k},\ k=0,\dots,n$
$\chi^2(n)$	Chi-squared distribution	$\frac{x^{n/2-1}}{2^{n/2}\Gamma(\frac{n}{2})}\exp\{-\frac{x}{2}\},\ x\geq 0$
$\mathcal{C}(a,b)$	Cauchy distribution	$\frac{1}{\pi b(1+(\frac{x-a}{b})^2)},\ x\in\mathbb{R}$
$Lap(a,b)$	Laplace distribution	$\frac{1}{2b}\exp\{-\frac{\lvert x-a\rvert}{b}\},\ x\in\mathbb{R}$
$Log(a,b)$	Logistic distribution	$\frac{\exp\{-(x-a)/b\}}{b(1+\exp\{-(x-a)/b\})^2},\ x\in\mathbb{R}$
$\mathcal{N}(\mu,\sigma^2)$	Normal distribution	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(x-\mu)^2}{\sigma^2}\},\ x\in\mathbb{R}$
$\mathcal{P}(\alpha,x_m)$	Pareto distribution	$\frac{\alpha x_m^\alpha}{x^{\alpha+1}},\ x\geq x_m$
$\mathcal{U}(a,b)$	Uniform distribution	$\frac{1}{b-a},\ x\in(a,b)$
$t(n)$	t-distribution	$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\ x\in\mathbb{R}$
$\mathcal{N}_2(\boldsymbol{\mu},\boldsymbol{\Sigma})$	2D-normal distribution	$\frac{\lvert\boldsymbol{\Sigma}\rvert^{-1/2}}{2\pi}\exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\},\ \mathbf{x}\in\mathbb{R}^2$

Bibliography

Adcock, R. J. Note on the method of least squares. *The Analyst*, **4**:183–184, 1877.

Adcock, R. J. A problem in least squares. *The Analyst*, **5**:53–54, 1878.

Akritas, M. G., Albers, W. Aligned rank tests for the linear model with heteroscedastic errors. *J. Statist. Planning and Inference*, **37**:23–41, 1993.

Ali, M. M., Giaccotto, C. A study of several new and existing tests for heteroscedasticity in the general linear model. *J. of Econometrics*, **26**:355–373, 1984.

Anderson, T., Darling, D. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.*, **23**:193–212, 1952.

Arbuthnot, J. An argument for divine providence taken from the constant regularity observed in the births of both sexes. *Philos. Trans.*, **27**:186–190, 1710.

Barker, F., Soh, Y. C., Evans, R. J. Properties of the geometric mean functional relationship. *Biometrics*, **44**:279–282, 1988.

Bartlett, M. S. Fitting a straight line when both variables are subject to error. *Biometrics*, **5**:207–212, 1949.

Bickel, P. J. Using residuals robustly i: Tests for heteroscedasticity, non-linearity. *Ann. Statist.*, **6**:266–291, 1978.

Breusch, T. S., Pagan, A. R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, **47**:1287–1294, 1979.

Buonaccorsi, J. P. *Measurement Error Models, Methods and Applications*. Boca Raton: Chapman & Hall/CRC, 2010.

Carroll, R. J., Ruppert, D. On robust tests for heteroscedasticity. *Ann. Statist.*, **9**:206–210, 1981.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. *Measurement error in nonlinear models: a modern perspective*. Boca Raton: Chapman & Hall/CRC, 2006.

Cheng, C. L., Van Ness, J. W. *Statistical Regression with Measurement Error*. Kendall's Library of Statistics 6. Arnold, London, 1999.

Csörgő, M., Révész, L. Strong approximations of the quantile process. *Ann. Statist.*, **6**:882–894, 1978.

DeGracie, J. S., Fuller, W. A. Estimation of the slope and analysis of covariance when the concomitant variable is measured with error. *Journal of the American statistical association*, **67**:930–937, 1972.

Deming, W. E. The application of least squares. *Philos. Mag. Ser.*, **7**:146–158, 1931.

Dette, H., Munk, A. Testing heteroscedasticity in nonparametric regression. *J. R. Statist. Soc. B*, **60**:693–708, 1998.

Dixon, S. L., McKean, J. W. Rank-based analysis of the heteroscedastic linear model. *J. Amer. Statist. Assoc.*, **91**:699–712, 1996.

Drion, E. F. Estimation of the parameters of a straight line and of the variances of the variables, if they are both subject to error. *Indagationes Math.*, **13**:256–260, 1951.

Durbin, J. Errors in variables. *Rev. Inst. Internat. Statist.*, **22**:23–32, 1954.

Ezekiel, M., Fox, K. A. *Methods of correlation and regression analysis.* John Wiley & Sons, New York, 1959.

Fleiss, J. L. *Statistical methods for rates and proportions.* John Wiley & Sons, New York, 1981.

Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Assoc.*, **32**:675–701, 1937.

Fuller, W. A. *Measurement error models.* John Wiley & Sons, New York, 1987.

Geary, R. C. Inherent relations between random variables. *Proc. R. Irish. Acad. Sect. A.*, **47**:1541–1546, 1942.

Golub, G. H., Van Loan, C. F. An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, **17**:883–893, 1980.

Gutenbrunner, C. Tests for heteroscedasticity based on regression quantiles and regression rank scores. In P. Mandl, M. Hušková, editors, *Asymptotic Statistics: Proceedings of the 5th Prague Symposium.* Physica-Verlag, Heidelberg, 1994.

Gutenbrunner, C., Jurečková, J. Regression rank scores and regression quantiles. *The Annals of Statistics*, **20**:305–330, 1992.

Gutenbrunner, C., Jurečková, J., Koenker, R. Regression rank test for heteroscedasticity. Technical report, Charles University in Prague, 1995.

Gutenbrunner, C., Jurečková, J., Koenker, R., Portnoy, S. Tests of linear hypotheses based on regression rank scores. *Journal of Nonparametric Statistics*, **2**:307–333, 1993.

Hájek, J., Šidák, Z. *Theory of rank tests.* Academia, Prague, 1967.

Hájek, J., Šidák, Z., Sen, P. K. *Theory of rank tests, Second Edition.* Academic Press, New York, 1999.

Heiler, S., Willers, R. Asymptotic normality of r-estimates in the linear model. *Statistics*, **19**:173–184, 1988.

Hodges, J. L., Lehmann, E. L. Estimation of location based on rank tests. *Ann. Math. Statist.*, **34**:598–611, 1963.

Hotteling, H., Pabst, M. R. Rank correlation and tests of significance involving no assumptions of normality. *Ann. Math. Statist.*, **7**:29–43, 1936.

Hušková, M. Asymptotic distribution of simple linear rank statstics for testing symmetry. *Zeitschr. Wahrscheinlichkeitsth. verw. Geb.*, **12**:308–322, 1970.

Jaeckel, L. A. Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.*, **43**:1449–1459, 1972.

Jurečková, J. Asymptotic linearity of a rank statistic in regression parameter. *Ann. Math. Statist.*, **40**:1889–1900, 1969.

Jurečková, J. Nonparametric estimate of regression coefficients. *Ann. Math. Statist.*, **42**:1328–1338, 1971.

Jurečková, J., Kalina, J., Picek, J., Saleh, A. K. M. E. Rank tests of linear hypothesis with measurement errors both in regressors and responses. *Preprint of Department of Probability and Mathematical Statistics, Charles University in Prague*, **66**, 2009.

Jurečková, J., Koul, H. L., Navrátil, R., Picek, J. Behavior of r-estimators under measurement errors. *Bernoulli*, submitted.

Jurečková, J., Navrátil, R. Rank tests under uncertainty: Regression and local heteroscedasticity. *Synergies of Soft Computing and Statistics for Intelligent Data Analysis Advances in Intelligent Systems and Computing*, **190**:255–261, 2012.

Jurečková, J., Navrátil, R. Rank tests in heteroscedastic linear model with nuisance parameters. *Metrika*, **77**:433–450, 2014.

Jurečková, J., Picek, J., Saleh, A. K. M. E. Rank tests and regression rank score tests in measurement error models. *Computational Statistics and Data Analysis*, **54**:3108–3120, 2010.

Jurečková, J., Sen, P. K. *Robust Statistical Procedures: Asymptotics and Interrelations.* John Wiley & Sons, New York, 1996.

Kendall, M. G. A new measure of rank correlation. *Biometrika*, **30**:81–93, 1938.

Koenker, R. A note on studentizing a test for heteroscedasticity. *J. Econometrics*, **17**:107–112, 1981.

Koenker, R. Rank tests for linear models. *Handbook of Statistics*, **15**:175–199, 1997.

Koenker, R., Basset, G. Regression quantiles. *Econometrica*, **46**:33–50, 1978.

Koenker, R., Bassett, G. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, **50**:43–61, 1982.

Koenker, R., Zhao, Q. *l*-estimation for linear heteroscedastic models. *Nonparametric Statistics*, **3**:223–235, 1994.

Koul, H. L. Asymptotic behavior of the wilcoxon type confidence regions for the multiple linear regression. *Ann. Math. Statist.*, **40**:1950–1979, 1969.

Koul, H. L. *Weighted empirical processes in dynamic nonlinear models*. Springer, New York, 2002.

Kummel, C. H. Reduction of observed equations which contain more than one observed quantity. *The Analyst*, **6**:97–105, 1879.

Lehmann, E. L. *Nonparametrics. Statistical methods based on ranks*. Holden-Day, 1975.

Lewbel, A. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *J. Business and Econ. Statist.*, **30**:67–80, 2012.

Lin, J.-G., Qu, X.-Y. A consistent test for heteroscedasticity in semi-parametric regression with nonparametric variance function based on the kernel method. *Statistics*, **46**:565–576, 2012.

Lindley, D. V. Regression lines and the linear functional relationship. *Suppl. J. Roy. Statist. Soc.*, **9**:218–244, 1947.

Lyon, J. D., Tsai, C.-L. A comparison of tests for heteroscedasticity. *The Statistician*, **45**:337–349, 1996.

Navrátil, R. Rank tests of symmetry in measurement errors models. In J. Šafránková, J. Pavlů, editors, *WDS'10 Proceedings of Contributed Papers: Part I - Mathematics and Computer Sciences*, pages 177–182. Matfzypress, Prague, 2010.

Navrátil, R. Rank tests and r-estimates in location model with measurement errors. In J. Zelinka, I. Horová, editors, *Proceedings of Workshop of the Jaroslav Hájek Center and Financial Mathematics in Practise I. Book of short papers*, pages 37–42. Masarykova Univerzita, Brno, 2012.

Navrátil, R., Saleh, A. K. M. E. Rank tests of symmetry and r-estimation of location parameter under measurement errors. *Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica*, **50**:95–102, 2011.

Nummi, T., Möttönen, J. Estimation and prediction for low degree polynomial models under measurement errors with an application to forest harvesters. *Appl. Statist.*, **53**:495–505, 2004.

Oosterhoff, J., van Zwet, W. R. A note on contiguity and hellinger distance. In J. Jurečková, editor, *Contribution to Statistics, Jaroslav Hájek Memorial Volume*, pages 157–166. Academia, Prague and Reidel, Dordrecht, 1979.

Pakes, A. On the asymptotic bias of the wald-type estimators of a straight line when both variables are subject to error. *Int. Econ. Rev.*, **23**:491–497, 1982.

Parzen, E. Nonparametric statistical data modeling. *J. Amer. Statist. Assoc.*, **74**:105–121, 1979.

Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos. Mag.*, **2**:559–572, 1901.

Picek, J. *Statistical Procedures Based on Regression Rank Scores*. Ph.D. thesis, Charles University in Prague, 1996.

Pitman, E. J. G. *Lecture notes on nonparametric statistics*. Columbia University, New York, 1948.

Pollard, D. Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, **7**:186–199, 1991.

Puri, M. L., Sen, P. K. *Nonparametric methods in multivariate analysis*. John Wiley & Sons, New York, 1971.

Puri, M. L., Sen, P. K. *Nonparametric methods in general linear models*. John Wiley & Sons, New York, 1985.

Saleh, A. K. M. E., Picek, J., Kalina, J. R-estimation of the parameters of a multiple regression model with measurement errors. *Metrika*, **75**:311–328, 2012.

Scott, E. L. Note on consistent estimates of the linear structural relation between two variables. *Anal. Math. Stat.*, **21**:284–288, 1950.

Sen, P. K., Saleh, A. K. M. E. The theil-sen estimator in a measurement error perspective. *IMS Collections: Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková*, **7**:224–234, 2010.

Smirnov, N. V. Sur la distribution de ω^2 (criterium de m. r. v. mises). *C. R. Akad. Sci. Paris*, **202**:449–452, 1936.

Teissier, G. La relation d'allometrie sa signification statistique et biologique. *Biometrics*, **4**:14–53, 1948.

Theil, H. A rank-invariant method for linear and polynomial regression analysis. *Indag. Math.*, **12**:85–91, 1950.

Tolmatz, L. On the distribution of the square integral of the brownian bridge. *The Annals of Probability*, **30**:253–269, 2002.

Tolmatz, L. Addenda: On the distribution of the square integral of the brownian bridge. *The Annals of Probability*, **31**:530–532, 2003.

Ury, H. K., Forrester, J. C. More on scale dependence of wilcoxon signed-rank test. *The American Statistician*, **24**:25–26, 1970.

Wald, A. The fitting of straight lines if both variables are subject to error. *Ann. Math. Statistics*, **11**:285–300, 1940.

Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics*, **1**:80–83, 1945.