Kybernetika

Jan Ámos Víšek

Instrumental weighted variables under heteroscedasticity. Part I – Consistency

Kybernetika, Vol. 53 (2017), No. 1, 1–25

Persistent URL: http://dml.cz/dmlcz/146704

KYBERNETIKA — VOLUME *53* (2017), NUMBER 1, PAGES 1–25

INSTRUMENTAL WEIGHTED VARIABLES UNDER HETEROSCEDASTICITY PART I – CONSISTENCY

Jan Ámos Víšek

The proof of consistency instrumental weighted variables, the robust version of the classical instrumental variables is given. It is proved that all solutions of the corresponding normal equations are contained, with high probability, in a ball, the radius of which can be selected – asymptotically – arbitrarily small. Then also $\sqrt{n}$-consistency is proved. An extended numerical study (the Part II of the paper) offers a picture of behavior of the estimator for finite samples under various types and levels of contamination as well as various extent of heteroscedasticity. The estimator in question is compared with two other estimators of the type of "robust instrumental variables" and the results indicate that our estimator gives comparatively good results and for some situations it is better.

The discussion on a way of selecting the weights is also offered. The conclusions show the resemblance of our estimator with the M-estimator with Hampel's ψ-function. The difference is that our estimator does not need the studentization of residuals (which is not a simple task) to be scale- and regression-equivariant while the M-estimator does. So the paper demonstrates that we can directly compute – moreover by a quick algorithm (reliable and reasonably quick even for tens of thousands of observations) - the scale- and the regression-equivariant estimate of regression coefficients.

Keywords: weighting order statistics of the squared residuals, consistency of the instrumental weighted variables, heteroscedasticity of disturbances, numerical study

Classification: 62J02, 62F35

1. INTRODUCTION OF BASIC FRAMEWORK

We are going to start with usual framework. Let $\mathcal{N}$ denote the set of all positive integers, R the real line and R^p the p-dimensional Euclidean space. We assume that all random variables (r.v.'s) are defined on a basic probability space $(\Omega, \mathcal{A}, P)$. The linear regression model

$$Y_i = X_i^{'}\beta^0 + e_i = \sum_{j=1}^{p} X_{ij}\beta_j^0 + e_i, \quad i = 1, 2, \ldots, n \tag{1}$$

DOI: 10.14736/kyb-2017-1-0001

will be considered (all vectors throughout the paper will be assumed to be the column ones)[1]. We will need also the matrix form of the model

$$Y = X\beta^0 + e \tag{2}$$

where $Y = (Y_1, Y_2, \ldots, Y_n)'$, $X = (X_1, X_2, \ldots, X_n)'$ and $e = (e_1, e_2, \ldots, e_n)'$. We shall further assume that:

Conditions $\mathcal{C}1$. The sequence $\left\{(X_i', e_i)'\right\}_{i=1}^{\infty}$ is sequence of independent $(p+1)$-dimensional random variables (r.v.'s) distributed according to distribution functions (d. f.) $F_{X,e_i}(x, r) = F_{X,e}(x, r\sigma_i^{-1})$ where $F_{X,e}(x, r)$ is a parent d. f. and $\sigma_i^2 = \text{var}\,(e_i)$. Further, $I\!E e_i = 0$ and

$$0 < \liminf_{i\to\infty} \sigma_i \le \limsup_{i\to\infty} \sigma_i < \infty.$$

Denote $F_{e|X}(r|X_1 = x)$ the conditional d.f. corresponding to the parent d.f. $F_{X,e}(x, r)$. Then, for all $x \in R^p$ $F_{e|X}(r|X_1 = x)$ is absolutely continuous with density $f_{e|X}(r|X_1 = x)$ bounded by U_e (which does not depend on x).

Remark 1.1. The assumption that the d.f. $F_{e|X}(r|X_1 = x)$ is continuous is not only technical assumption. Without the (bounded) density, we should assume that $F_{e|X}(r|X_1 = x)$ is Lipschitz and it would bring a more complicated form of all what follows. The absolute continuity is then a technical assumption.

In what follows $F_X(x)$ and $F_e(r)$ will denote the corresponding marginal d. f.'s of the *parent d. f.* $F_{X,e}(x, r)$. Then, assuming that e is a *"parent" r. v.* distributed according to *parent d. f.* $F_e(r)$, we have, e. g., $F_{e_i}(r) = P(e_i < r) = P(\sigma_i \cdot e < r) = P(e < \sigma_i^{-1} \cdot r) = F_e(\sigma_i^{-1} \cdot r)$, etc.. Conditions $\mathcal{C}1$ imply that the marginal d. f. $F_X(x)$ does not depend on i, the sequence $\{X_i\}_{i=1}^{\infty}$ is sequence of independent and identically distributed (i.i.d.) r.v.'s. Finally, let us recall that the *ordinary least squares* (*OLS*) are – when processing economic (and other social sciences) data – more and more substituted by *instrumental variables* just due to the fact that we meet more and more with situations when disturbances are correlated with explanatory variables. Then, however, it is nearly straightforward to assume that (at least) the variances of the individual disturbances depend on explanatory variables, i. e. we can expect the heteroscedasticity[2].

Remark 1.2. Let us realize that under Conditions $\mathcal{C}1$ there are $s \in R$ and $S \in R$ so that for any $i \in \mathcal{N}$ we have $0 < s \le \sigma_i \le S < \infty$. Further, $F_{e_i|X}(r|X_i = x) = F_{e|X}(\sigma_i^{-1} r|X_1 = x)$ and $f_{e_i|X}(r|X_i = x) = \sigma^{-1} \cdot f_{e|X}(\sigma_i^{-1} r|X_1 = x) \le s^{-1} \cdot U_e$.

[1]When we assume the regression model with intercept, we have $X_{i1} = 1$ for all i. In some papers the intercept is given in the notation explicitly, to stress that it plays – mostly from the application point of view but sometimes also from the theoretical one – a special role. We will not need to stress it except in one discussion on assumptions given below.

[2]The last sentence is not to be understood as a "justification" for studying the behavior of robustified versions of instrumental variables under heteroscedasticity because the results of such a study have only a limited direct importance for *practical applications* (as already Fisher's results [19] indicated and Mizon's example [42] recalled and stressed with even higher power) but it does be a justification from the point of view of famous Halmos' paper [24], i. e. it has an indirect consequence for applications, see also a discussion below.

Remark 1.3. The problems, resulting from ignoring the heteroscedasticity, were recognized very early, see e. g. [14, 15] or [27]. It is known that the efficient estimator (under the normality of disturbances) is $\hat{\beta}^{(GLS,n)} = \left(\sum_{i=1}^{n} \sigma_i^{-2} X_i X_i'\right)^{-1} \sum_{i=1}^{n} \sigma_i^{-2} X_i Y_i$ but the unknown σ_i's do not allow to employ it. The asymptotic efficiency can be reached utilizing estimators of individual variances (see e. g. [47, 49] or [75]). An estimator of covariance matrix resistant to heteroscedasticity was established in [74]. It opens way for proper studentization of the coordinates of $\hat{\beta}^{(OLS,n)}$ but we can compute the corresponding p-values only approximately because the studentized estimators have only approximately t-distribution, see e. g. [20, pp. 96–97] or [48][3].

2. INSTRUMENTAL VARIABLES AS ALTERNATIVE TO THE ORDINARY LEAST SQUARES

When the *orthogonality condition* $I\!E\left\{e_i | X_i = x\right\} = 0$ is broken, the *ordinary least squares* are not consistent (see (10) below where substitute X instead of Z). The best known example of such a situation is the model assuming that the explanatory variables are measured with random error (sometimes the model is called *error-in-variables model* – compare [31] and [55]) – for further discussion see [66]. A possibility how to solve the problem is to employ the *orthogonal regression* (sometimes also called the *total least squares*), see e. g. [43] or [55]. This approach is usually considered in natural sciences. The econometricians offer another remedy in the form of the *method of instrumental variables.* We can meet with various definitions. The first one, probably the most frequently given, copes with the situation by modifying the orthogonality condition.

Definition 2.1. For any sequence of p-dimensional random vectors $\{Z_i\}_{i=1}^{\infty}$ the solution(s) of the equation

$$\sum_{i=1}^{n} Z_i \left(Y_i - X_i^{'}\beta\right) = Z'\left(Y - X\beta\right) = 0 \tag{3}$$

will be called the *estimator obtained by means of the method of instrumental variables* (or *instrumental variables*, for short) and denoted by $\hat{\beta}^{(IV,n)}$.

Notice that (3) is an analogy of the *classical normal equations*

$$\sum_{i=1}^{n} X_i \left(Y_i - X_i^{'}\beta\right) = X'\left(Y - X\beta\right) = 0. \tag{4}$$

Generally the dimensionality of Z_i's can be $q \neq p$. If $q < p$, (3) implies immediately the underidentification of model, see [22] or [76] (for the case $q > p$ see discussion below).

The classical regression adopts the assumption (more or less technical one) that the

[3]Of course, the previous remark has to be again understood in the sense of already mentioned papers [19, 42] and [24]. In other words, the theoretical results have for the applications mostly a "vicarious" importance of keeping sound mathematical traditions in proposing a new method (and there is no substitute for it) but really reliable information about its behavior can be found (only) by well designed numerical study.

matrix $Z'X$ is regular (for discussion see e. g. [22, 33, 75] or [76], and also Remark 2.6 below). Then it follows from (3) that

$$\hat{\beta}^{(IV,n)} = (Z'X)^{-1} Z'Y. \tag{5}$$

The instrumental variable can be introduced also in an alternative way – over an extremal problem, see e. g. [10, 31] or [33].

Definition 2.2. Let D be a positive definite, symmetric matrix and put

$$\hat{\beta}^{(IV,n)} = \underset{\beta \in R^p}{\arg\min} \; (Y - X\beta)' ZDZ' (Y - X\beta). \tag{6}$$

Remark 2.3. Definition 2.2 allows for $q > p$. Moreover, assuming that Z has full rank and considering instead of Z the matrix $\tilde{Z} = Z(Z'Z)^{-1}Z'X$, we can show that solution of (6) coincides formally with (5), see [31]. Recently the situation when we have a large number of instruments was considered and a plausible solution can be established by borrowing an idea from the *partial least squares*, see e.,g. [28] and [29].

We can meet also with definition utilizing (in fact) two stage least squares, see again [10] or [31].

Definition 2.4. Let $q \geq p$ and Z be of full rank and put $P_Z = Z(Z'Z)^{-1}Z'$. Then define

$$\hat{\beta}^{(IV,n)} = \underset{\beta \in R^p}{\arg\min} \; (X'P_Z X)^{-1} X'P_Z Y. \tag{7}$$

Remark 2.5. Plugging $Z(Z'Z)^{-1}Z'$ instead of P_Z in (7), we again easy demonstrate that the solution of (7) formally coincides with (5). Definition 2.4 says that $\hat{\beta}^{(IV,n)}$ uses the endogenous part of X. Finally, we can consider the transformed data $\tilde{Y} = P_Z \cdot Y$ and $\tilde{X} = P_Z \cdot X$ and employ the fact that P_Z is idempotent. Then (7) reads

$$\hat{\beta}^{(IV,n)} = \underset{\beta \in R^p}{\arg\min} \; \left(\tilde{X}'\tilde{X}\right)^{-1} \tilde{X}'\tilde{Y}. \tag{8}$$

It says that $\hat{\beta}^{(IV,n)}$ can be considered to be the *Ordinary Least Squares* estimator for appropriately transformed data[4]. By the way, it also says that Z can be selected so that $Z'X$ is positive definite with preserving the values of $\hat{\beta}^{(IV,n)}$. In other words, it shows that without a loss of generality we can assume that (see also (12) in [10])

$$X_i = \Pi \cdot Z_i + \eta \cdot e_i + \xi_i \tag{9}$$

where Π (matrix of type $(p \times q)$) has full rank, $\eta \in R^p$, Z_i, e_i and ξ_i are mutually independent. In the ideal case we would be able to find such an instrument Z_i that it represents the whole exogenous information in X_i, i. e. we decompose X_i so that $\xi_i \equiv 0$. In fact, it hints how the various ways of defining the *instrumental variables* have the same roots (for details see [10] and references given there). Finally, it shows that without a significant restriction of generality we can assume that $q = p$.

[4]This explain why this approach is called – *two stage least squares*.

Remark 2.6. The classical theory (for simplicity) assumes that $\{Z_i\}_{i=1}^{\infty}$ is a sequence of independent and identically distributed random variables, which are not correlated with the sequence of disturbances $\{e_i\}_{i=1}^{\infty}$ (to simplify the notation in the next discussion, let us assume for a while the homoscedasticity of disturbances). Substituting (2) into (5), we obtain

$$\hat{\beta}^{(IV,n)} = \left(\frac{1}{n}Z'X\right)^{-1}\frac{1}{n}Z'(X\beta^0 + e) = \beta^0 + \left(\frac{1}{n}Z'X\right)^{-1}\frac{1}{n}Z'e. \tag{10}$$

It implies that if the *orthogonality condition* $I\!\!E\,(Z_1e_1) = 0$ holds, $\hat{\beta}^{(IV,n)}$ is unbiased and consistent, provided the matrix $I\!\!E\,(Z_1X_1')$ is regular.

Remark 2.7. Having assumed again the homoscedasticity and $I\!\!E\,(Z_1e_1) = 0$, then substituting (2) into (3) and finally computing the mean value of respective expression (remember that the sequences $\{Z_i\}_{i=1}^{\infty}$ and $\{X_i\}_{i=1}^{\infty}$ are sequences of identically distributed r. v.'s), we obtain

$$I\!\!E\left\{Z_1\left[e_1 - X_1'\left(\beta - \beta^0\right)\right]\right\} = 0 \tag{11}$$

which implies that

$$I\!\!E\left\{Z_1X_1'\right\}\left(\beta - \beta^0\right) = 0. \tag{12}$$

If $I\!\!EZ_1X_1'$ is positive definite (see (8)), (12) holds iff

$$\left(\beta - \beta^0\right)' I\!\!E\left\{Z_1X_1'\right\}\left(\beta - \beta^0\right) = 0 \tag{13}$$

and it is equivalent to

$$\left(\beta - \beta^0\right)' I\!\!E\left\{Z_1\left[e_1 - X_1'\left(\beta - \beta^0\right)\right]\right\} = 0 \tag{14}$$

(remeber that we assume that the orthogonality condition $I\!\!E\,(Z_1e_1) = 0$ holds). Notice that – due to the fact that the orthogonality condition does not contain β, it holds uniformly with respect to $\beta \in R^p$. When we robustify the classical methods we need some uniformity in this conditions, see e. g. [9, 10, 39] or (32) below.

There is a lot of papers discussing the heuristic reasons for defining the instrumental variables, the possibilities how to select the instruments and the problem of the implementation – for many references see [66]. But there was only a limited number of papers trying to robustify the method, see e. g. [1, 9, 10, 11, 34, 35, 36, 38, 39, 40, 50, 73] or [57] (and the references given there[5]). Moreover, even these papers, including [66], consider the situation when the disturbances are homoscedastic, except of [7] where the idea of the *generalized method of moments* (GMM) is employed, see [26]. Due to the fact that GMM does not need to specify the underlying d. f., the method covers also the heteroscedasticity – in a latent way because GMM is in fact studied in the i. i. d. framework. So, although [7] does not address directly (in formalism) the heteroscedasticity, it is a way how to cope with it. On the other hand, as the heteroscedasticity is not implemented in the model, it is not explicitly treated which could (and in fact

[5]By the way, in [10] one can find a brief list of the previous attempts of robustification of the instrumental variables. The spirit of paper then says why the authors gave preference to the robustification based on S-estimator, i. e. the reason why the smooth depression of the influence of suspicious points seems to be preferable way.

does) result in a bad efficiency of GMM, see e. g. Table 2 in [7]. Let's recall that the heteroscedasticity does not prevent the unbiasedness and the consistency of $\hat{\beta}^{(OLS,n)}$. However it is problem for its efficiency – for a discussion how to treat the problem – e. g. by estimating a model for heteroscedasticity – see [76].

The problem may be also with the zero-one objective function (employed in [7]), i. e. with trimming the observations. It implies (nearly) inevitably an instability of estimator with respect to *inliers*, even for estimators with moderate level of robustness, see [30] and also [56]. The problem of (possible) implausible effect of trimming away some points has its roots in the fact that it classifies – a bit omnipotently – the observations either as "clean" or as "contaminated" and then a small shift of an "inlier" can change its classification from "clean" to "contaminated" (or vice versa). Finally, it implies a switch of the estimate of model, see again [30] and [56]. So, it seems that the smooth depression of the influential observations may be preferable. We will return to the problem below.

The (technical) problems which the allowance of heteroscedasticity implies, one can learn by comparing the papers [61] and [69]. Nevertheless, much better insight into the problem we obtain when we compare the papers [65] and [71]. Both papers generalize the Kolmogorov–Smirnov result for the regression scheme. The former under homoscedasticity of disturbances, the latter under the heteroscedasticity, see Lemma 6.7 of this paper. The proof of former result mimics the steps of Kolmogorov's proof, see [48] or [13], verifying that they can be done in the regression framework. The proof of the latter result however requires the application of the Skorohod embedding into the Wiener process, see [46] or [32]. That is why it is long and technically complicated. The same unfortunately holds for some proofs from the present paper and [66].

The shape of the *normal equations* (3) defining $\hat{\beta}^{(IV,n)}$ (as being of the same shape as the shape of *normal equations* for $\hat{\beta}^{(OLS,n)}$) indicates that $\hat{\beta}^{(IV,n)}$ is not robust with respect to the outliers and/or leverage points. So a robustified version of $\hat{\beta}^{(IV,n)}$ is to be employed when the orthogonality condition is broken and simultaneously there is a suspicion of a contamination of data. An inspiration for one possible way how to do it may be taken from the *normal equations* for the *least weighted squares* $\hat{\beta}^{(LWS,n,w)}$ (see [59]) which are robustified version of the $\hat{\beta}^{(OLS,n)}$. So, let us briefly recall it.

3. ROBUSTIFYING THE INSTRUMENTAL VARIABLES

In what follows, for any $\beta \in R^p$ define the ith residual as $r_i(\beta) = Y_i - X_i^{'}\beta$ and by $r^2_{(h)}(\beta)$ denote the hth order statistic among the squared residuals, i. e. we have

$$r^2_{(1)}(\beta) \le r^2_{(2)}(\beta) \le \ \cdots \ \le r^2_{(n)}(\beta). \tag{15}$$

Rather general way how to robustify the *ordinary least squares* is to weight down the residuals of observations which seem to be suspicious. Nevertheless, it is known that when the weighting is done according to an external rule (typically based on some diagnostics of data), it need not be (and usually it doesn't be) optimal. Hence it is better to let the method itself to assign the weights "implicitly", in other words, to assign the weights to the order statistics of the squared residuals rather than directly to the squared residuals.

Definition 3.1. Let $w : [0,1] \to [0,1]$ be a weight function. Then the solution of the extremal problem

$$\hat{\beta}^{(LWS,n,w)} = \arg\min_{\beta \in R^p} \sum_{j=1}^{n} w\left(\frac{j-1}{n}\right) r_{(j)}^2(\beta) \tag{16}$$

will be called the *least weighted squares*, ([59], see also [60] and [61]).

Notice please that the *least median of squares* $\hat{\beta}^{(LMS,n,h)}$ (see [51]), the *least trimmed squares* $\hat{\beta}^{(LTS,n,h)}$ (see [25][6]) and the *ordinary least squares* $\hat{\beta}^{(OLS,n)}$ are special cases of the *least weighted squares*. Moreover, the possibility to accommodate the shape of the weight function w to the level and to the character of contamination[7] guarantee that $\hat{\beta}^{(LWS,n,w)}$ can adapt to various situations, i.e. to various level and/or characters of contamination. The trimming of observations can (more or less) cope with the level of contamination employing the *forward search*, see [2]. But it cannot accommodate to the character of contamination. The problem was in details discussed on the *Workshop on algorithm for outliers/regressors selection*, see [72]. After all, we shall demonstrate it in Part II of paper in the promised numerical study. Moreover, $\hat{\beta}^{(LWS,n,w)}$ can be easier used for panel data processing, easier than $\hat{\beta}^{(LMS,n,h)}$ and $\hat{\beta}^{(LTS,n,h)}$, see [5]. Last but not least, $\hat{\beta}^{(LWS,n,w)}$ is, similarly as $\hat{\beta}^{(OLS,n)}$, $\hat{\beta}^{(LMS,n,h)}$ and $\hat{\beta}^{(LTS,n,h)}$, *scale-* and *regression-equivariant*.

Conditions $\mathcal{C}2$. Weight function $w : [0,1] \to [0,1]$ is absolutely continuous and nonincreasing, with the derivative $w'(\alpha)$ bounded from below by $-L$ ($L > 0$), $w(0) = 1$.

Remark 3.2. In the simulations, results of which are presented in Part II, the weight function, borrowing the shape from the famous Tukey ρ-function, was used – see Figure 1 in Part II.

We will study scale- and regression-equivariant estimators and hence without loss of generality, we may assume in what follows in theoretical considerations that $\beta^0 = 0$. Then $r_i(\beta) = e_i - X_i^{'}\beta$.

Due to presence of order statistics in Definition 3.1, (16) is not convenient for proving consistency. That is why, following [23] for any $i \in \{1, 2, \ldots, n\}$ and any $\beta \in R^p$ let us define the *rank of the ith residual* as

$$\pi(\beta, i) = j \in \{1, 2, \ldots, n\} \qquad \Leftrightarrow \qquad r_i^2(\beta) = r_{(j)}^2(\beta). \tag{17}$$

Then we have

$$\hat{\beta}^{(LWS,n,w)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{\pi(\beta,i)-1}{n}\right) r_i^2(\beta). \tag{18}$$

[6]Although it became common to give this reference on $\hat{\beta}^{(LTS,n,h)}$, $\hat{\beta}^{(LTS,n,h)}$ was proposed in fact nearly simultaneously with $\hat{\beta}^{(LMS,n,h)}$, prior to the *S-etimator* $\hat{\beta}^{(S,n,\rho)}$, see [53].

[7]Of course, not to all characters, otherwise it would not be a scientific method, see [45].

Now, denoting the indicator of a set A by $I\{A\}$, for any $\beta \in R^p$ the empirical distribution function (e. d. f.) of the absolute value of residual will be considered in the form

$$F_\beta^{(n)}(r) = \frac{1}{n}\sum_{j=1}^n I\{|r_j(\beta)| < r\} = \frac{1}{n}\sum_{j=1}^n I\left\{|e_j - X_j^{'}\beta| < r\right\}. \tag{19}$$

Then, it is only a technicality to show that

$$\frac{\pi(\beta,i)-1}{n} = F_\beta^{(n)}(|r_i(\beta)|) \tag{20}$$

(the fact that $\pi(\beta,i)-1$ stays in the numerator of fraction in (20) – not $\pi(\beta,i)$ – is due to the sharp inequality in (19)) and hence $\hat\beta^{(LWS,n,h)}$ can be finally given as

$$\hat\beta^{(LWS,n,w)} = \arg\min_{\beta\in R^p} \sum_{i=1}^n w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) r_i^2(\beta). \tag{21}$$

It may seem, at the first glance, strange to consider e. d. f. when the observations are not identically distributed. However, as Lemma 6.7 shows the e. d. f. $F_\beta^{(n)}(r)$ can be *uniformly in* $r \in R$, *uniformly in* $\beta \in R^p$ and even *uniformly in* $\sigma^2 \in R^+$ approximated by the "mean" d. f.

$$\overline{F}_{n,\beta}(v) = \frac{1}{n}\sum_{i=1}^n F_{i,\beta}(v) \tag{22}$$

where (remember that e_i's have different variances σ_i^2)

$$F_{i,\beta}(r) = P\left(|Y_i - X_i'\beta| < r\right) = P\left(|e_i - X_i'\beta| < r\right). \tag{23}$$

It is nearly straightforward that the solution of the extremal problem (21) is one of solutions of the *normal equations*

$$I\!N E_{Y,X,n}^{(LWS)}(\beta) = \sum_{i=1}^n w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) X_i\left(Y_i - X_i^{'}\beta\right) = 0. \tag{24}$$

Remark 3.3. Notice that the robustification of $\hat\beta^{(OLS,n)}$ was achieved by including $w\left(F_\beta^{(n)}(|r_i(\beta)|)\right)$ into the *normal equations* (4), see (24). It gives immediately an inspiration how to robustify the classical *instrumental variables* (3). It simultaneously explains why we needed the similar modifications of the orthogonality condition and of the assumption that $I\!E\left[X_1 \cdot X_1'\right]$ is positive definite, e. g. including into the corresponding condition also $w\left(F_\beta^{(n)}(|r_i(\beta)|)\right)$.

Definition 3.4. For any sequence of p-dimensional random vectors $\{Z_i\}_{i=1}^\infty$ the solution(s) of the equation

$$I\!N E_{Y,X,Z,n}^{(IWV)}(\beta) = \sum_{i=1}^n w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) Z_i\left(Y_i - X_i^{'}\beta\right) = 0 \tag{25}$$

will be called the *estimator obtained by means of the method of instrumental weighted variables* (or *instrumental weighted variables*, for short) and denoted by $\hat\beta^{(IWV,n,w)}$, see [63, 64].

4. CONSISTENCY OF THE INSTRUMENTAL WEIGHTED VARIABLES

In what follows we shall denote the joint d. f. of explanatory variables and of instrumental variables by $F_{X,Z}(x,z)$ and of course the marginal d. f.'s by $F_X(x)$ and $F_Z(z)$. We will need also the following notation. For any $\beta \in R^p$ the distribution of the product $\beta^{'} Z X^{'} \beta$ will be denoted $F_{\beta' Z X' \beta}(u)$, i. e.

$$F_{\beta' Z X' \beta}(u) = P(\beta^{'} Z_1 X_1^{'} \beta < u) \tag{26}$$

and similarly as in (19), the corresponding empirical d. f. will be denoted $F^{(n)}_{\beta' Z X' \beta}(u)$, so that

$$F^{(n)}_{\beta' Z X' \beta}(u) = \frac{1}{n} \sum_{j=1}^{n} I\left\{\beta^{'} Z_j X_j^{'} \beta < u\right\}. \tag{27}$$

For any $\lambda \in R^+$ and any $a \in R$ put

$$\gamma_{\lambda,a} = \sup_{\|\beta\|=\lambda} F_{\beta' Z X' \beta}(a). \tag{28}$$

Notice please that due to the fact that the surface of ball $\{\beta \in R^p, \|\beta\| = \lambda\}$ is compact and taking into account Conditions $\mathcal{C}3$ (below), there is $\beta_\lambda \in \{\beta \in R^p, \|\beta\| = \lambda\}$ so that

$$\gamma_{\lambda,a} = F_{\beta_\lambda' Z X' \beta_\lambda}(a). \tag{29}$$

For any $\lambda \in R^+$ let us denote

$$\tau_\lambda = -\inf_{\|\beta\| \le \lambda} \beta^{'} I\!E \left[Z_1 X_1^{'} \cdot I\{\beta^{'} Z_1 X_1^{'} \beta < 0\}\right] \beta. \tag{30}$$

Notice also that due to the indicator $I\{\beta^{'} Z_1 X_1^{'} \beta < 0\}$ in (30) we have $\tau_\lambda \ge 0$ and that again due to the fact that the ball $\{\beta \in R^p, \|\beta\| \le \lambda\}$ is compact, the infimum is finite, since there is a $\tilde{\beta} \in \{\beta \in R^p, \|\beta\| \le \lambda\}$ so that

$$\tau_\lambda = -\tilde{\beta}^{'} I\!E \left[Z_1 X_1^{'} \cdot I\{\tilde{\beta}^{'} Z_1 X_1^{'} \tilde{\beta} < 0\}\right] \tilde{\beta}. \tag{31}$$

Conditions $\mathcal{C}3$. The instrumental variables $\{Z_i\}_{i=1}^{\infty}$ are independent and identically distributed with distribution function $F_Z(z)$. Further, the joint distribution function $F_{X,Z}(x,z)$ is absolutely continuous with a density $f_{X,Z}(x,z)$ bounded by $U_{ZX} < \infty$. Further for any $n \in \mathcal{N}$ we have $I\!E \sum_{i=1}^{n} \left\{ w(F_{\beta^0}(|e_i|)) \cdot e_i \cdot Z_i \right\} = 0$ and the matrices $I\!E Z_1 Z_1^{'}$ as well as $I\!E \sum_{i=1}^{n} \left\{ w(F_{\beta^0}(|e_i|)) Z_i X_i^{'} \right\}$ are positive definite. Moreover, there is $q > 1$ so that $I\!E \left\{ \|Z_1\| \cdot \|X_1\| \right\}^q < \infty$. Finally, there is $a > 0$, $b \in (0,1)$ and $\lambda > 0$ so that

$$a \cdot (b - \gamma_{\lambda,a}) \cdot w(b) > \tau_\lambda \tag{32}$$

for $\gamma_{\lambda,a}$ and τ_λ given by (28) and (30).

Remark 4.1. As $I\!E \sum_{i=1}^{n} \left\{ w(F_{\beta^0}(|e_i|)) \cdot e_i \cdot Z_i \right\} = \sum_{i=1}^{n} I\!E \left\{ w(F_{\beta^0}(|e_i|)) \cdot e_i \right\} \cdot I\!E Z_i$, we have $I\!E \sum_{i=1}^{n} \left\{ w(F_{\beta^0}(|e_i|)) \cdot e_i \cdot Z_i \right\} = 0$ if e (the "parent" r. v., see text below Remark 1.1) is symmetrically distributed and its mean value exists – just due to symmetry of $w(F_{\beta^0}(|v|))$ in v. The condition (32) "regulates" the mutual relations between X_i's and Z_i's. We have discussed it in Remark 2.6. The condition seems to be easier directly verifiable from data than (11) or (12). But it contains $\gamma_{\lambda,a}$ and τ_λ so that we have to rely also on some empirically obtained estimates of them. It may be also of interest to compare Conditions $\mathcal{C}3$ with the conditions in [57] where we considered instrumental M-estimators and the discussion of assumptions for M-instrumental variables was given.

Lemma 4.2. Let Conditions $\mathcal{C}1$, $\mathcal{C}2$ and $\mathcal{C}3$ be fulfilled. Then for any $\varepsilon > 0$ there is $\zeta > 0$ and $\delta > 0$ such that

$$P\left(\left\{\omega \in \Omega : \inf_{\|\beta\| \geq \zeta} -\frac{1}{n}\beta^{'} I\!N E_{Y,X,Z,n}(\beta) > \delta\right\}\right) > 1 - \varepsilon$$

(for $I\!N E_{Y,X,Z,n}(\beta)$ see (25)). In other words, any sequence $\left\{\hat{\beta}^{(IWV,n,w)}\right\}_{n=1}^{\infty}$ of the solutions of the sequence of normal equations $I\!N E_{Y,X,Z,n}(\hat{\beta}^{(IWV,n,w)}) = 0$ is bounded in probability.

The proof is formally nearly the same as the proof of Lemma 1 in [66]. The allowance for the heteroscedasticity of disturbances requires some *formally* straightforward modifications[8]. □

Lemma 4.3. Let Conditions $\mathcal{C}1$, $\mathcal{C}2$ and $\mathcal{C}3$ be fulfilled. Then for any $\varepsilon > 0$, $\delta \in (0, 1)$ and $\zeta > 0$ there is $n_{\varepsilon,\delta,\zeta} \in \mathcal{N}$ so that for any $n > n_{\varepsilon,\delta,\zeta}$ we have

$$P\left(\left\{\omega \in \Omega \ : \ \sup_{\|\beta\| \leq \zeta} \left| \frac{1}{n}\sum_{i=1}^{n}\left\{ w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right)\beta^{'} Z_i \left(e_i - X_i^{'}\beta\right) \right.\right.\right.\right.$$
$$\left.\left.\left.\left. -\beta^{'} I\!\!E\left[w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) Z_i\left(e_i - X_i^{'}\beta\right)\right]\ \right\} \right| < \delta\right\}\right) > 1-\varepsilon$$

(for $\overline{F}_{n,\beta}(.)$ see (22)).

The proof has formally similar structure as the proof of Lemma 2 in [66]. It is a bit more complicated because again instead of employing a limiting distribution we need to estimate differences of empirical d. f. from a sequence of the arithmetic means of underlying d. f.'s $\{\overline{F}_{n,\beta}(v)\}_{n=1}^{\infty}$, see (22). □

As we already recalled, see the end of Remark 2.6, the classical regression analysis accepted (under the framework of homoscedasticity of disturbances – hence the assumption is sufficient to make about one observation) the orthogonality condition $I\!\!E\{e_1 Z_1\} = 0$

[8]The fact that the modifications are relatively simple and straightforward is due to the fact that the complicated steps were made in [71] but the background of proof is different from the proof in [66]. The approximation of empirical d. f. is not by the underlying d. f. as the limit of the empirical d. f.'s but we employ the knowledge about convergence of the difference of the empirical d. f.'s and the arithmetic mean of the d. f. of individual disturbances.

and the assumption that $I\!E\left\{Z_1 X_1^{'}\right\}$ is a regular matrix[9] (see e. g. [4], [31, 76] or [75]) to be able to prove consistency of $\hat{\beta}^{(IV,n)}$. We have discussed it in Remark 2.6. So one way can be to modify (11) and/or (12) as follows:

Conditions $\mathcal{C}4$. For any $n \in \mathcal{N}$ the equation

$$\beta' \sum_{i=1}^{n} I\!E\left[w\left(\overline{F}_{n,\beta}(|r_i(\beta)\right) Z_i \left(e_i - X_i^{'}\beta\right)\right] = 0 \tag{33}$$

in the variable $\beta \in R^p$ has unique solution at $\beta = 0$, i. e. at $\beta = \beta^0$.

An alternative way would be to give conditions on the weight function and on the distributions of disturbances, of explanatory variables and of instruments which would imply uniqueness of solution of (33). We could take an inspiration from [10, 37] and mainly from [39] (because the roots of both [10] and [37] go back to [39]). We can learn that the conditions are assumed to be valid uniformly in (or independently of) β and then Remarks 2.6 and 3.3 indicate that they can read $I\!E[w(\overline{F}_{n,\beta}(|r_i(\beta)|)) \cdot Z_i \cdot e_i] = 0$ together with the assumption that $I\!E[w(\overline{F}_{n,\beta}(|r_i(\beta)|)) \cdot Z_i \cdot X_i'$] is positive definite for all β from some compact subset of R^p containing β^0.

It is clear that the last two conditions implies (33). We can however give less intuitive (and less "straightforward") conditions which leads to (33), see [9].

Conditions $\mathcal{C}4'$ The distributions F_Z, F_e and F_ξ are elliptically symmetric around the zero and the corresponding mean values and the covariance matrices $I\!E Z_1 Z_1'$ and $I\!E \xi_1 \xi_1'$ exist. Further, $I\!E Z_1 Z_1'$ is regular. Finally, (9) holds with Π of the full rank (for simplicity assume that $q = p$).

Remark 4.4. It is clear that in the framework, we have used throughout the paper, the Conditions $\mathcal{C}4'$ cannot be fulfilled if the intercept is present. In such a case the Conditions $\mathcal{C}1$ have to be given in a bit modified way:
The sequence $\{(V_i', e_i)'\}_{i=1}^{\infty}$ is sequence of independent p-dimensional random variables distributed according to the distribution functions (d. f.) $F_{V,e_i}(v,r) = F_{V,e}(v, \sigma_i \cdot r)$, $i \in \mathcal{N}$ where $F_{V,e}(v,r)$ is a parent d. f., $I\!E V_1 = 0$, the covariance matrix $I\!E\left\{V_1 V_1'\right\}$ is regular and $\sigma_i^2 = \mathrm{var}\,(e_i|V_i)$ with $0 < \sigma_i^2 < K < \infty$. ... Finally, consider the sequence $\{(X_i', e_i)'\}_{i=1}^{\infty}$ where $X_{i1} = 1$ and $X_{ij} = V_{i,j-1}$, $j = 2, 3, \ldots, p$ for all $i \in \mathcal{N}$. This sequence will be considered as the sequence of explanatory variables and of disturbances, see e. g. [68]. It is nearly straightforward that the proofs in the previous text can be carry out in this framework but they would be formally a bit more complicated, see again [68].

---

[9] It means that Z is of full rank and then $I\!E\left\{Z_1 X_1^{'}\right\} = I\!E \tilde{X}_1' \tilde{X}_1$, i. e. $I\!E\left\{Z_1 X_1^{'}\right\}$ is positive definite, see Remark 2.5.

Remark 4.5. The assumption that F_e is symmetric is widely used, especially in the regression framework[10]. The assumption on symmetry of explanatory variables and of the instruments is not so frequent but in fact from the theoretical point of view it is not substantially limiting the generality. When we are going to study the behavior of an estimator we can assume that prior to performing the considerations we can transform the data. Moreover, there are even studies hinting that the "spirit of statistics" implicitly assumes symmetry (while it is hidden by some latent transformation of "underlying raw data"), see e. g. [16] or [17]. For some further discussion on the assumption of elliptical symmetry see [11, 12, 39] and others, in fact it goes back (at least) to [3]. Already [25] or [52] hinted that the assumption can be relax at the cost of an increase intricacy of proofs.

Assertion 4.6. Let Conditions $\mathcal{C}1$ and $\mathcal{C}4'$ hold. Then for any $\beta \in R^p$ we have for all $i \in \mathcal{N}$

$$\mathbb{E}\left[w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) \cdot Z_i \cdot e_i\right] = 0 \tag{34}$$

and

$$\mathbb{E}\left[w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) \cdot Z_i \cdot X_i'\right] \tag{35}$$

is positive definite. Consequently, (33) holds.

P r o o f . Writing

$$r_i(\beta) = e_i - X_i\beta = e_i - (\Pi \cdot Z_i + \eta \cdot e_i + \xi_i)\,\beta$$

(see (9)) and recalling that the joint distribution $F_{Z,e,\xi}$ is the product of the marginals, we can calculate the corresponding integrals. We have

$$\mathbb{E}\left[w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) \cdot Z_i \cdot e_i\right]$$

$$= \int \left\{ \int \left[w\left(\overline{F}_{n,\beta}(|v - (\Pi \cdot z + \eta \cdot v + t)'\,\beta|)\right)\right] \mathrm{d}F_\xi(t) \right\} z \cdot v \cdot \mathrm{d}F_Z(z) \cdot \mathrm{d}F_e(v).$$

Then due to the fact that the argument of $\overline{F}_{n,\beta}$ is an absolute value of the expression $v - (\Pi \cdot z + \eta \cdot v + t)'\,\beta$ and due to the symmetry of F_ξ,

$$H_\beta(z, v) = \int \left[w\left(\overline{F}_{n,\beta}(|v - (\Pi \cdot z + \eta \cdot v + t)'\,\beta|)\right)\right] \mathrm{d}F_\xi(t)$$

is symmetric in z and v and moreover $H_\beta(z, v) \in [0, 1]$. From Conditions $\mathcal{C}4'$ we have $\mathbb{E}Z_1 = 0$ and $\mathbb{E}e_i = 0$ and since the function $H_\beta(z, v)$ weights down z and v symmetrically, we conclude the proof of (34). Nearly the same arguments can be used to verify (35). □

[10] One of very first employment was probably in [21] but see also [3, 31] or [33], and all after also [10].

Theorem 4.7. Let Conditions $\mathcal{C}1$, $\mathcal{C}2$, $\mathcal{C}3$ and $\mathcal{C}4$ or $\mathcal{C}4'$ be fulfilled. Then any sequence $\left\{\hat{\beta}^{(IWV,n,w)}\right\}_{n=1}^{\infty}$ of the solutions of normal equations (25) $I\!\!N E_{Y,X,Z,n}^{(IWV)}(\beta) = 0$ is weakly consistent.

P r o o f. To prove the consistency of $\{\hat{\beta}^{(IWV,n,w)}\}_{n=1}^{\infty}$, we have to show that for any $\varepsilon > 0$ and $\delta > 0$ there is $n_{\varepsilon,\delta} \in \mathcal{N}$ such that for all $n > n_{\varepsilon,\delta}$

$$P\left(\left\{\omega \in \Omega : \left\|\hat{\beta}^{(IWV,n,w)} - \beta^0\right\| < \delta\right\}\right) > 1 - \varepsilon. \tag{36}$$

So fix $\varepsilon_1 > 0$ and $\delta_1 > 0$. According to Lemma 4.2 there are $\delta_1 > 0$ and $\theta_1 > 0$ so that for ε_1 there is $n_{\delta_1,\varepsilon_1} \in \mathcal{N}$ so that for any $n > n_{\delta_1,\varepsilon_1}$

$$P\left(\left\{\omega \in \Omega : \inf_{\|\beta\| \geq \theta_1} -\frac{1}{n}\beta' I\!\!N E_{Y,X,Z,n}(\beta) > \delta_1\right\}\right) > 1 - \frac{\varepsilon_1}{2}$$

(denote the corresponding set by B_n). It means that for all $n > n_{\delta_1,\varepsilon_1}$ all solutions of the normal equations $I\!\!N E_{Y,X,Z,n}(\beta) = 0$ are inside the ball $\mathcal{B}(0, \theta_1)$ with probability at least $1 - \frac{\varepsilon_1}{2}$. If $\theta_1 \leq \delta$, we have finished the proof. Generally of course we can have $\theta_1 > \delta$.

Then, utilizing Lemma 4.3 we may find for ε_1, $\delta = \min\{\frac{\theta_1}{2}, \delta_1\}$ and θ_1 such $n_{\varepsilon_1,\delta,\theta_1} \in \mathcal{N}, n_{\varepsilon_1,\delta,\theta_1} \geq n_{\delta_1,\varepsilon_1}$ so that for any $n > n_{\varepsilon_1,\delta,\theta_1}$ there is a set C_n (with $P(C_n) > 1 - \frac{\varepsilon}{2}$) such that for any $\omega \in C_n$

$$\sup_{\|\beta\| \leq \theta_1} \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) \beta' Z_i \left(e_i - X_i'\beta\right) \right.\right.$$
$$\left.\left. - \beta' I\!\!E\left[w\left(F_\beta(|r_i(\beta)|)\right) Z_i \left(e_i - X_i'\beta\right)\right] \right\} \right| < \delta.$$

But it means that

$$\inf_{\|\beta\| = \theta_1} \left\{ -\beta' \frac{1}{n} \sum_{i=1}^{n} I\!\!E\left[w\left(F_\beta(|r_i(\beta)|)\right) Z_i \left(e_i - X_i'\beta\right)\right] \right\} > \frac{\delta_1}{2} > 0. \tag{37}$$

Further consider the compact set $C = \{\beta \in R^p : \delta_1 \leq \|\beta\| \leq \theta_1\}$ and find

$$\tau_C = \inf_{\beta \in C} \left\{ -\beta' \frac{1}{n} \sum_{i=1}^{n} I\!\!E\left[w\left(F_\beta(|r_i(\beta)|)\right) Z_i \left(e_i - X_i'\beta\right)\right] \right\}. \tag{38}$$

Then there is a $\{\beta_k\}_{k=1}^{\infty}$ such that

$$\lim_{k \to \infty} \beta_k' \frac{1}{n} \sum_{i=1}^{n} I\!\!E\left[w\left(F_{\beta_k}(|r_i(\beta_k)|)\right) Z_i \left(e_i - X_i'\beta_k\right)\right] = -\tau_C.$$

On the other hand, due to compactness of C there is a β^* and a subsequence $\left\{\beta_{k_j}\right\}_{j=1}^{\infty}$ such that

$$\lim_{j \to \infty} \beta_{k_j} = \beta^*$$

and due to the uniform continuity (uniform in $i \in \mathcal{N}$ and $\beta \in C$) of $\beta' I\!\!E \left[w \left(F_\beta(|r_i(\beta)|) \right) \times \right.$ $\left. \times Z_i \left(e_i - X_i' \beta \right) \right]$ (see Lemma 6.16) we have

$$- [\beta^*]' \frac{1}{n} \sum_{i=1}^{n} I\!\!E \left[w \left(F_{\beta^*}(|r_i(\beta^*)|) \right) Z_i \left(e_i - X_i' \beta^* \right) \right] = \tau_C. \tag{39}$$

Employing once again the uniform continuity (uniform in $i \in \mathcal{N}$ and $\beta \in C$) of $\beta' I\!\!E \left[w \left(F_\beta(|r_i(\beta)|) \right) Z_i \left(e_i - X_1' \beta \right) \right]$ together with Condition $\mathcal{C}4$ and (37) we find that $\tau_C > 0$, otherwise there has to be a solution of (33) inside the compact C which does not contain $\beta = 0$.

Now, utilizing Lemma 4.3 once again we may find for ε_1, δ_1, θ_1 and τ_C $n_{\varepsilon_1,\delta_1,\theta_1,\tau_C} \in \mathcal{N}, n_{\varepsilon_1,\delta_1,\theta_1,\tau_C} \geq n_{\varepsilon_1,\delta,\theta_1}$ so that for any $n > n_{\varepsilon_1,\delta_1,\theta_1,\tau_C}$ there is a set D_n (with $P(D_n) > 1 - \frac{\varepsilon}{2}$) such that for any $\omega \in D_n$

$$\sup_{\|\beta\| \leq \theta_1} \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ w \left(F_\beta^{(n)}(|r_i(\beta)|) \right) \beta' Z_i \left(e_i - X_i' \beta \right) \right. \right.$$
$$\left. \left. - \beta' I\!\!E \left[w \left(F_\beta(|r_i(\beta)|) \right) Z_i \left(e_i - X_i' \beta \right) \right] \right\} \right| < \frac{\tau_C}{2}. \tag{40}$$

But (38) and (40) imply that for any $n > n_{\varepsilon_1,\delta_1,\theta_1,\tau_C}$ and any $\omega \in B_n \cap D_n$ we have

$$\inf_{\|\beta\| > \delta_1} -\frac{1}{n} \beta' I\!\!N\!E_{Y,X,Z,n}(\beta) > \frac{\tau_C}{2}. \tag{41}$$

Of course, $P(B_n \cap D_n) > 1 - \varepsilon_1$. But it means that all solutions of normal equations (33) are inside the ball of radius δ_1 with probability at least $1 - \varepsilon_1$, i. e. in other words, $\hat\beta^{(IWV,n,w)}$ is weakly consistent. □

We will need to enlarge the previous conditions.

Conditions $\mathcal{NC}1$. The derivative $f_e'(r)$ exists and is bounded in absolute value by B_e. The derivative $w'(\alpha)$ exists and is Lipschitz of the first order (with the corresponding constant J_w). Moreover, for any $i \in \mathcal{N}$ $I\!\!E[w'(\overline{F}_{n,\beta^0}(|e_i|))(f_e(|e_i|) - f_e(-|e_i|)) \cdot e_i] = 0$. Finally, for any $\ell, k, j = 1, 2, \ldots, p$ $I\!\!E \left| V_{1\ell} \cdot V_{1k} \cdot V_{1j} \right| < \infty$ where for V_{1s} can be substituted either X_{1s} or Z_{1s} and the mean value is bounded for any combination of X's and Z's.

Theorem 4.8. Let the conditions $\mathcal{C}1$, $\mathcal{C}2$, $\mathcal{C}3$, $\mathcal{C}4$ (or $\mathcal{C}4'$) and $\mathcal{NC}1$ hold. Then any sequence $\left\{ \hat\beta^{(IWV,n,w)} \right\}_{n=1}^{\infty}$ of solutions of normal equations (25) $I\!\!N\!E_{Y,X,Z,n}^{(IWV)}(\beta) = 0$ is $\sqrt{n}$-consistent, i. e.

$$\forall (K > 0) \qquad \exists (n_K \in \mathcal{N}) \qquad \forall (n > n_K) \qquad \sqrt{n} \left\| \hat\beta^{(IWV,n,w)} - \beta^0 \right\| < K.$$

P r o o f. Recalling that $\hat{\beta}^{(IWV,n,w)}$ solves the normal equations (25), we obtain (we will write it in a "traditional" form with β^0 although we have assumed that $\beta^0 = 0$)

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right) Z_i e_i$$
$$= \frac{1}{n} \sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right) Z_i X_i' \cdot \sqrt{n}\left(\hat{\beta}^{(IWV,n,w)} - \beta^0\right). \tag{42}$$

Now, taking into account that the weight function w is Lipschitz and Lemma 6.7, we find that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right) Z_i e_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) Z_i e_i + R_n^{(1)}(\beta, X, Z, e)$$

and

$$\frac{1}{n} \sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right) Z_i X_i' = \frac{1}{n} \sum_{i=1}^{n} w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) Z_i X_i' + R_n^{(2)}(\beta, X, Z, e)$$

where

$$\sup_{\beta \in R^p} \left\| R_n^{(1)}(\beta, X, Z, e) \right\| = \mathcal{O}_p(1) \quad \text{and} \quad \sup_{\beta \in R^p} \left\| R_n^{(2)}(\beta, X, Z, e) \right\| = o_p(1).$$

By a chain of similar approximations[11] we finally show that (42) can be written as

$$L_n = Q_n \cdot (1 + q_n) \cdot \sqrt{n}\left(\hat{\beta}^{(IWV,n,w)} - \beta^0\right) \tag{43}$$

with $L_n = \mathcal{O}_p(1)$, $Q_n \to Q$ in probability, Q being a regular matrix, and $q_n = o_p(\hat{\beta}^{(IWV,n,w)} - \beta^0)$. Then assuming that $(1 + q_n) \cdot \sqrt{n}(\hat{\beta}^{(IWV,n,w)} - \beta^0)$ is not $\mathcal{O}_p(1)$ and employing the Lemma 6.20 we prove that also L_n cannot be $\mathcal{O}_p(1)$, which is a contradiction. □

5. CONCLUSIONS

The paper proves the consistency and $\sqrt{n}$-consistency of the *Instrumental Weighted Variables* under heteroscedasticity employing the generalization of Kolmogorov–Smirnov's result for the regression framework, see [13] and [71]. The present proofs, if given in the full length, would triple the length of paper but they are significantly simpler than proofs in [60] and [61]. The complexity of the proofs for robust estimators is the sacrifice we have to pay for possibility to employ them for data processing, see [39]. However the idea that the tax, we pay for the robustness of methods, is a loss of efficiency, is wrong. The speed of PCs and abilities of modern languages (the codes are written in) allow to tailor (the parameters of) the method (we have decided to utilize) just to the level and the character of contamination of the data (if any) – just by repeating the computations up to the moment of a break of stability of estimated models and so not to loose any information (see e. g.[2] or [62]).

Employing the results by Phillips and Solo [44] the results of the present paper can be

[11] It contains 19 steps, for details see [70].

generalized to the autoregressive framework which in turn would allow to utilize the lagged values as instruments.

A decade or two ago people tried to make an idea about behaviour of estimators by the *small sample asymptotics* (see e. g. [18]) however they had to check the results by numerical studies anyway. Nowadays the IT tools allow to study behavior of estimators for finite sample sizes directly. And we would be foolish not to employ this possibility. Nevertheless, let us stress that numerical studies cannot substitute theoretical verification of such basic properties as e. g. consistency – for a nice discussion we recall again the famous Halmos' paper [24].

An important question of course is how to select the weight function. Intuitively it seems acceptable to use some function which is equal to one on an interval $[0, h]$, then it is decreasing (smootly) to zero and finally, it is equal to zero on $[g, 1]$. In the simulation study (in the Part II of this paper) we have used in that decreasing part the function of Tukey's type. The optimality of function (i. e. the "free" parameters h, g and the constant c of Tukey's function) can be approximately established (even for real data) by something which is usually called the *forward search*, for a theory see [2], for an example with economic data, see e. g. [58]. The experiences (with simulated data) confirmed an intuitive idea that the interval $(g, 1)$ has to cover the bad leverage points.

On the other hand, the simulations revealed the fact that the decreasing part of the weight function w should be rather wide (i. e. $h << g$) in the case of high heteroscedasticity. It says in other words, that when we assign the weights to the order statistics of squared residuals, we simultaneously cope (partially) with heteroscedasticity, even in the sense of decreasing MSE of the estimator. There is also a theoretical result answering (at least partially) this problem, see [41].

Last but not least, the preparation of a simulation study makes us to start really think repeatedly about the framework of simulations. In the case of a new robust estimator we try to invent such a framework which can cause to the estimator and its competitors really considerable problems. Then it may happen that we realize that the cliché which became a statistical/econometric folklore can be wrong, may be due to the fact that we removed by the (up to now) proposed estimators the problem which the classical estimators have with outliers and/or leverage points, but some other type of atypical observations can appear to be dangerous. And they can cause problems to the estimators in question. We will meet with something like this in the Part II of paper.

6. APPENDIX

An efficient tool for proving the key lemma allowing a uniform approximation of the empirical distribution function is the Skorohod embedding (in the sense as it was used in [32] or [46]) for which we will need the following three assertions.

Assertion 6.1. (Štěpán [54], page 420, VII.2.8) Let a and b be positive numbers. Further let ξ be a random variable such that $P(\xi = -a) = \pi$ and $P(\xi = b) = 1 - \pi$ (for a $\pi \in (0, 1)$) and $I\!E\xi = 0$. Moreover let τ be the time for the Wiener process $W(s)$ to exit the interval $(-a, b)$. Then

$$\xi =_{\mathcal{D}} W(\tau)$$

where "$=_{\mathcal{D}}$" denotes the equality of distributions of the corresponding random variables. Moreover, $I\!\!E\tau = a \cdot b = var\,\xi$.

Remark 6.2. Since the book by Štěpán [54] is in Czech language we refer also to [6] where however this assertion is not isolated. Nevertheless, the assertion can be found directly in the first lines of the proof of Proposition 13.7 (page 277) of Breiman's book. (See also Theorem 13.6 on the page 276.) The next assertion can be found, in a bit modified form also in Breiman's book, Proposition 12.20 (page 258).

Assertion 6.3. (Štěpán [54], page 409, VII.1.6) Let a and b be positive numbers. Then

$$P\left(\max_{0\le t\le b} |W(t)| > a\right) \le 2 \cdot P\left(|W(b)| > a\right).$$

Definition 6.4. Let S be a subset of a separable metric space. The stochastic process $V = (V(s),\ s \in S)$ is called *separable* if there is a countable dense subset $T \subset S$ (i. e. T is countable and dense in S) such that for any $(\omega, s) \in \Omega \times S$ there is a sequence such that

$$s_n \in T, \quad \lim_{n\to\infty} s_n = s \quad \text{and} \quad \lim_{n\to\infty} V(\omega, s_n) = V(\omega, s).$$

Assertion 6.5. (Štěpán [54], page 85, I.10.4) Let $V = (V(s),\ s \in S)$ be a *separable* stochastic process defined on the probability space $(\Omega, \mathcal{A}, P)$. Moreover, let $G \subset S$ be open and denote by $k(G)$ the set of all finite subsets of G. Then for any closed set $K \subset I\!\!R^p$ we have

$$\{\omega \in \Omega : V(s) \in K, s \in G\} \in \mathcal{A}$$

and

$$P\left(\{\omega \in \Omega : V(s) \in K, s \in G\}\right) = \inf_{J\in k(G)} P\left(\{\omega \in \Omega : V(s) \in K, s \in J\}\right).$$

P r o o f. Since the book by Štěpán [54] is in Czech language and the proof is short, we will give it. Let T be countable dense subset of S. Then we have

$$\{\omega \in \Omega : V(s) \in K, s \in G\} = \{\omega \in \Omega : V(s) \in K, s \in G \cap T\}$$

and

$$\begin{aligned} P\left(\{\omega \in \Omega : V(s) \in K, s \in G\}\right) &\le \inf_{J\in k(G)} P\left(\{\omega \in \Omega : V(s) \in K, s \in J\}\right) \\ \le \inf_{J\in k(G\cap S)} P\left(\{\omega \in \Omega : V(s) \in K, s \in J\}\right) &= P\left(\{\omega \in \Omega : V(s) \in K, s \in G \cap S\}\right) \\ &= P\left(\{\omega \in \Omega : V(s) \in K, s \in G\}\right). \end{aligned}$$

□

Assertion 6.6. Kolmogorov law of large numbers (Breiman [6], Theorem 3.27). Let $\{V_i\}_{i=1}^{\infty}$ be a sequence of independent random variables, $I\!\!EV_i = 0$, $I\!\!EV_i^2 < \infty$. Let b_n converge to $+\infty$. If $\sum_{i=1}^{\infty} I\!\!EV_i^2 \cdot b_i^{-2} < \infty$, then

$$\frac{1}{b_n}\sum_{i=1}^{n} V_i \quad \underset{n\to\infty}{\longrightarrow} \quad 0 \qquad a.\,s.$$

Now, we are going to give the key lemma for proving consistency of $\hat{\beta}^{(IWV,n,w)}$.

Lemma 6.7. Let the Conditions $\mathcal{C}1$ hold. For any $\varepsilon > 0$ there is a constant K_ε and $n_\varepsilon \in \mathcal{N}$ so that for all $n > n_\varepsilon$

$$P\left(\left\{\omega \in \Omega : \sup_{v \in R^+} \sup_{\beta \in \boldsymbol{R}^p} \sqrt{n}\left|F_\beta^{(n)}(v) - \overline{F}_{n,\beta}(v)\right| < K_\varepsilon\right\}\right) > 1 - \varepsilon \tag{44}$$

(for $F_\beta^{(n)}(v)$ and $\overline{F}_{n,\beta}(v)$ see (19) and (22), respectively).

Proof. The proof can be found in [71]. □

Remark 6.8. It seems that the result of Lemma 6.7 can be generalized e. g. for the situation when the sequence $\{(X_i', e_i)'\}_{i=-\infty}^{\infty}$ is AR vector process. We just apply Cochrane–Orcutt transformation (see [8]) and we transform the problem into i. i. d. framework. Similarly for other structures of dependence of r. v.'s which allow a transformation "back" to independence we can achieve the same.

Lemma 6.9. Under Conditions $\mathcal{C}1$ the distribution function $F_{i,\beta}(r)$ is, uniformly with respect to $r \in R$ and $i \in \mathcal{N}$, uniformly continuous in $\beta \in R^p$, i. e. for any $\varepsilon > 0$ there is $\delta \in (0,1)$ so that for any pair $\beta^{(1)}$ and $\beta^{(2)}$ such that $\left\|\beta^{(1)} - \beta^{(2)}\right\| < \delta$ we have

$$\sup_{i \in \mathcal{N}} \sup_{r \in R} \left|F_{i,\beta^{(1)}}(r) - F_{i,\beta^{(2)}}(r)\right| \le \varepsilon$$

(for $F_{i,\beta}(r)$ see (23)).

Proof. Proof is a chain of finding some upper bounds of some integrals representing the empirical d. f.s. □

Corollary 6.10. Under Conditions $\mathcal{C}1$ the distribution function $\overline{F}_{n,\beta}(v)$ is, uniformly with respect to $v \in R$ and $n \in \mathcal{N}$, uniformly continuous in $\beta \in R^p$, i. e. for any $\varepsilon > 0$ there is $\delta \in (0,1)$ so that for any pair $\beta^{(1)}$ and $\beta^{(2)}$ such that $\left\|\beta^{(1)} - \beta^{(2)}\right\| < \delta$ we have

$$\sup_{n \in \mathcal{N}} \sup_{v \in R} \left|\overline{F}_{n,\beta^{(1)}}(v) - \overline{F}_{n,\beta^{(2)}}(v)\right| \le \varepsilon.$$

Lemma 6.11. Under Conditions $\mathcal{C}3$ the distribution function $F_{\beta' Z X' \beta}(u)$ is, uniformly with respect to $u \in R$, uniformly continuous in $\beta \in R^p$, i. e. for any $\varepsilon > 0$ there is $\delta \in (0,1)$ so that for any pair $\beta^{(1)}$ and $\beta^{(2)}$ such that $\left\|\beta^{(1)} - \beta^{(2)}\right\| < \delta$ we have

$$\sup_{u \in R} \left|F_{[\beta^{(1)}]' Z X' \beta^{(1)}}(u) - F_{[\beta^{(2)}]' Z X' \beta^{(2)}}(u)\right| \le \varepsilon.$$

Proof. Proof utilizes the Skorohod embedding into the Wiener process in a nearly the same way as Lemma 6.7. As the paper [71] is easily available we shall skip it. □

Lemma 6.12. Under Conditions $\mathcal{C}3$ the distribution function $F_{\beta' ZX'\beta}(u)$ is, uniformly with respect to $\beta \in R^p$, uniformly continuous in $u \in R$, i. e. for any $\varepsilon > 0$ there is $\delta \in (0,1)$ so that for any pair $u^{(1)}$ and $u^{(2)}$ such that $\left\|u^{(1)} - u^{(2)}\right\| < \delta$ we have

$$\sup_{\beta \in R^p} \left| F_{\beta' ZX'\beta}(u^{(1)}) - F_{\beta' ZX'\beta}(u^{(2)}) \right| \leq \varepsilon.$$

P r o o f. Proof runs again along similar lines as the proof of Lemma 6.7 and taking account again that paper [71] is available we skip the proof. □

We will have in what follows several such situations – the proofs are really very similar and the employment of the Skorohod embedding into the Wiener process is straightforward.

Let us recall that we have denoted for any $\beta \in R^p$ by $F_{\beta' ZX'\beta}(u)$ the distribution of the product $\beta^{'} ZX^{'} \beta$ (see (26)), i. e.

$$F_{\beta' ZX'\beta}(u) = P(\beta^{'} ZX^{'} \beta < u)$$

and the corresponding empirical distribution by $F^{(n)}_{\beta' ZX'\beta}(u)$ (see (27)), i. e.

$$F^{(n)}_{\beta' ZX'\beta}(u) = \frac{1}{n} \sum_{j=1}^{n} I\left\{\beta^{'} Z_j X_j^{'} \beta < u\right\}.$$

Lemma 6.13. Let Conditions $\mathcal{C}3$ hold and fix arbitrary $\varepsilon > 0$. Then there are $K < \infty$ and $n_\varepsilon \in \mathcal{N}$ so that for all $n > n_\varepsilon$

$$P\left(\left\{\omega \in \Omega : \sup_{\beta \in R^p} \sup_{u \in R} \sqrt{n} \left| F^{(n)}_{\beta' ZX'\beta}(u) - F_{\beta' ZX'\beta}(u) \right| \leq K \right\}\right) > 1 - \varepsilon.$$

P r o o f. Proof runs along the similar (nearly the same) lines as the proof of Lemma 6.7. □

Lemma 6.14. Let Conditions $\mathcal{C}3$ hold and fix arbitrary $\varepsilon > 0$. Then there is $K_\varepsilon < \infty$ and $n_\varepsilon \in \mathcal{N}$ so that for all $n > n_\varepsilon$

$$P\left(\left\{\omega \in \Omega : \sup_{\beta^{(1)}, \beta^{(2)} \in R^p} \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^{n} I\left\{ \left[\beta^{(1)}\right]^{'} Z_i X_i^{'} \beta^{(1)} < 0, \left[\beta^{(2)}\right]^{'} Z_i X_i^{'} \beta^{(2)} \geq 0 \right\} \right.\right.\right.$$
$$\left.\left.\left. - P\left(\left[\beta^{(1)}\right]^{'} Z_1 X_1^{'} \beta^{(1)} < 0, \left[\beta^{(2)}\right]^{'} Z_1 X_1^{'} \beta^{(2)} \geq 0 \right) \right| > K_\varepsilon \right\}\right) > 1 - \varepsilon.$$

P r o o f. Proof runs again along the similar lines as the proof of Lemma 6.7. □

Let us put

$$\mathcal{T}(\zeta, \Delta) = \left\{ \left\|\beta^{(1)}\right\| \leq \zeta, \left\|\beta^{(2)}\right\| \leq \zeta, \left\|\beta^{(1)} - \beta^{(2)}\right\| < \Delta \right\}. \tag{45}$$

Lemma 6.15. Let Conditions $\mathcal{C}3$ hold and fix arbitrary $\varepsilon > 0$ and $\zeta > 0$. Then there is $\Delta > 0$ so that

$$\sup_{(\beta^{(1)},\ \beta^{(2)}) \in \mathcal{T}(\zeta,\Delta)} P\left(\left[\beta^{(1)}\right]^{'} ZX^{'}\beta^{(1)} < 0, \left[\beta^{(2)}\right]^{'} ZX^{'}\beta^{(2)} \geq 0 \right) < \varepsilon.$$

P r o o f. Proof is again a chain of approximations of probabilities of some sets employing the continuity of probability. □

Lemma 6.16. Let Conditions $\mathcal{C}1$, $\mathcal{C}2$ and $\mathcal{C}3$ hold. Then for any positive ζ

$$\beta^{'} I\!\!E \left[w\left(F_\beta(|r_i(\beta)|)\right) Z_i \left(e_i - X_i^{'}\beta\right)\right] \tag{46}$$

is uniformly in $i \in \mathcal{N}$, uniformly continuous in β on $\mathcal{B} = \{\beta \in R^p : \|\beta\| \leq \zeta\}$, i. e. for any $\varepsilon > 0$ there is $\delta > 0$ so that for any pair of vectors $\beta^{(1)}, \beta^{(2)} \in R^p$, $\left\|\beta^{(1)} - \beta^{(2)}\right\| < \delta$ we have

$$\sup_{i\in\mathcal{N}} \Bigg| \left[\beta^{(1)}\right]^{'} I\!\!E \left[w\left(F_{\beta^{(1)}}(|r_i(\beta^{(1)})|)\right) Z_i \left(e_i - X_i^{'}\beta^{(1)}\right)\right]$$
$$- \left[\beta^{(2)}\right]^{'} I\!\!E \left[w\left(F_{\beta^{(2)}}(|r_i(\beta^{(2)})|)\right) Z_i \left(e_i - X_i^{'}\beta^{(2)}\right)\right] \Bigg| < \varepsilon.$$

P r o o f. Proof is full of technicalities utilizing a simple estimate of upper bounds of differences of the values of (46) for close pair of points in R^p. □

Lemma 6.17. Let Conditions $\mathcal{C}3$ hold. Then for any positive ζ

$$\beta^{'} I\!\!E \left[Z_1 X_1^{'} \cdot I\left\{\beta^{'} Z_1 X_1^{'}\beta < 0\right\}\right] \beta$$

is uniformly continuous in β on $\mathcal{B} = \{\beta \in R^p \ : \ \|\beta\| \leq \zeta\}$.

P r o o f. Proof is a rather long series of approximations of some differences of weight function employing mostly a basic differential calculus. □

Let us recall that for any $\zeta \in R^+$ we have denoted

$$\tau_\zeta = - \inf_{\|\beta\|\leq\zeta} \beta^{'} I\!\!E \left[Z_1 X_1^{'} \cdot I\{\beta^{'} Z_1 X_1^{'}\beta < 0\}\right] \beta.$$

Lemma 6.18. Let Conditions $\mathcal{C}3$ be fulfilled. Then for any $\varepsilon > 0$, $\delta \in (0,1)$ and $\zeta \geq 1$ there is $n_{\varepsilon,\delta,\zeta} \in \mathcal{N}$ so that for any $n > n_{\varepsilon,\delta,\zeta}$ we have

$$P\left(\left\{ \omega \in \Omega \ : \ \inf_{\|\beta\|\leq\zeta} \frac{1}{n} \sum_{i=1}^{n} \beta^{'} Z_i X_i^{'} \beta \cdot I\{\beta^{'} Z_i X_i^{'}\beta < 0\} > -\tau_\zeta - \delta \right\} \right) > 1 - \varepsilon.$$

P r o o f . Proof is a similar chain of approximations and estimations of upper bounds as the proof of Lemma 6.16. □

Lemma 6.19. Let Conditions $\mathcal{C}1$ hold. Then for any $\varepsilon > 0$ and $\delta \in (0,1)$ there is $\zeta > 0$ and $n_{\varepsilon,\delta} \in \mathcal{N}$ so that for all $n > n_{\varepsilon,\delta}$

$$P\left(\left\{\omega \in \Omega : \sup_{r \in R} \sup_{\left\|\beta^{(1)}-\beta^{(2)}\right\|<\zeta} \left|F^{(n)}_{\beta^{(1)}}(r) - F^{(n)}_{\beta^{(2)}}(r)\right| < \delta\right\}\right) > 1 - \varepsilon. \tag{47}$$

P r o o f . Proof follows nearly immediately from Corollary 6.10. □

Lemma 6.20. Let for some $p \in \mathcal{N}$, $\left\{\mathcal{V}^{(n)}\right\}_{n=1}^{\infty}$, $\mathcal{V}^{(n)} = \left\{v^{(n)}_{ij}\right\}_{i=1,2,\ldots,p}^{j=1,2,\ldots,p}$ be a sequence of $(p \times p)$ matrixes such that for $i = 1,2,\ldots,p$ and $j = 1,2,\ldots,p$

$$\lim_{n\to\infty} v^{(n)}_{ij} = q_{ij} \qquad \text{in probability} \tag{48}$$

where $Q = \{q_{ij}\}_{i=1,2,\ldots,p}^{j=1,2,\ldots,p}$ is a fixed nonrandom regular matrix. Moreover, let $\left\{\theta^{(n)}\right\}_{n=1}^{\infty}$ be a sequence of p–dimensional random vectors such that

$$\exists\ (\varepsilon > 0)\ \forall\ (K > 0)\ \limsup_{n\to\infty} P\left(\|\theta^{(n)}\| > K\right) > \varepsilon.$$

Then
$$\exists\ (\ \delta > 0)$$
so that
$$\forall\ (H > 0) \qquad \limsup_{n\to\infty}\ P\left(\left\|\mathcal{V}^{(n)}\theta^{(n)}\right\| > H\right) > \delta.$$

ACKNOWLEDGEMENT

We would like to thank two anonymous referees for valuable comments and recommendations how to improve the quality of explanation of ideas from the previous version of paper and also how to cover much better various items connected with the instrumental variables. The paper was written with the support of the Czech Science Foundation projects 13-01930S Robust methods for nonstandard situations, their diagnostics and implementations.

(Received July 28, 2014)

REFERENCES

[1] T. Amemiya: Two stage least absolute deviation estimators. Econometrica *50* (1982), 689–711. DOI:10.2307/1912608

[2] A. C. Atkinson, M. Riani, and A. Cerioli: Exploring Multivariate Data with the Forward Search. Springer Series in Statistics 2004, 31–88. DOI:10.1007/978-0-387-21840-3_2

[3] R. Beran: An efficient and robust adaptive estimator of location. Ann. Statist. *6* (1978), 292–313. DOI:10.1214/aos/1176344125

[4] R. J. Bowden and D. A. Turkington: Instrumental variables. Cambridge Univ. Press, Cambridge 1984. DOI:10.1017/ccol0521262410

[5] C. M. Bramati and C. Croux: Robust estimators for the fixed effects panel data model. The Econometr. J. *10* (2077), 521–540. DOI:10.1111/j.1368-423x.2007.00220.x

[6] L. Breiman: Probability. Addison–Wesley Publishing Company, London 1968.

[7] P. Čížek: Generalized method of trimmed moments. J. Statist. Planning Inference *171* (2009), 63–78. DOI:10.1016/j.jspi.2015.11.004

[8] D. Cochrane and G. H. Orcutt: Application of least squares regression to relationhips containing autocorrelated error terms. J. Amer. Statist. Assoc. *44* (1949), 32–61. DOI:10.1080/01621459.1949.10483290

[9] G. V. Cohen-Freue, H. Ortiz-Molina, and R. H. Zamar: A New Robust Instrumental Variables Estimator. Preprint 2008. University of British Columbia, Vancouvee, Canada (see http://www.stat.ubc.ca/∼ruben/website/cv/cohen-zamar.pdf).

[10] G. V. Cohen-Freue, H. Ortiz-Molina and R. H. Zamar: Natural robustification of the ordinary instrumental variables estimator. Biometrics *69* (2013), 641–650. DOI:10.1111/biom.12043

[11] C. Croux, S. Van Aelst, and C. Dehon: Bounded influence regression using high breakdown scatter matrices. Ann. Inst. Statist. Math. *55* (2013), 265–285. DOI:10.1007/bf02530499

[12] P. L. Davies: Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. Ann. Statist. *15* (1987), 1269–1292. DOI:10.1214/aos/1176350505

[13] M. Donsker (1952): Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. Ann. Math. Statist. *23* (1952), 277–281. DOI:10.1214/aoms/1177729445

[14] F. Eicker: Asymptotic normality and consistency of the least squares estimators for families of linear regressions. Ann. Math. Stat. *34* (1963), 447–456. DOI:10.1214/aoms/1177704156

[15] F. Eicker: Limit theorems for regression with unequal and dependent errors. In: Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability (L. Le Cam and J. Neyman, eds.), University of California Press, Berkeley 1967.

[16] Z. Fabián: Induced cores and their use in robust parametric estimation. Comm. Statist. – Theory and Methods *30* (2001), 537–555. DOI:10.1214/aoms/1177704156

[17] Z. Fabián: New measures of central tendency and variability of continuous distributions. Comm. Statist. – Theory and Methods *37* (2008), 159–174. DOI:10.1080/03610920701648987

[18] C. A. Field and E. M. Ronchetti (1990): Small Sample Asymptotics. Institute of Mathematical Statistics Monograph Series, Hayward 1990.

[19] R. A. Fisher: A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean squares error. Monthly Notes Royal Astronomical Society *80* (1920), 758–770. DOI:10.1093/mnras/80.8.758

[20] R. A. Fisher: Statistical Methods for Research Workers. Second edition (1928).

[21] F. Galton: Regression towards mediocrity in hereditary stature. J. Anthropol. Inst. *15* (1886), 246–263. DOI:10.2307/2841583

[22] W. H. Greene: Econometric Analysis. Macmillam Press, New York 1993.

[23] J. Hájek and Z. Šidák: Theory of Rank Test. Academic Press, New York 1967.

[24] P. R. Halmos: Applied mathematics is a bad mathematics. In: Mathematics Tomorrow (L. Steen, ed.), Springer Verlag, New York 1981, pp. 9–20. DOI:10.1007/978-1-4613-8127-3_2

[25] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel: Robust Statistics – The Approach Based on Influence Functions. J. Wiley and Sons, New York 1986. DOI:10.1002/9781118186435

[26] L. P. Hansen: Large sample properties of generalized method of moments estimators. Econometrica *50* (1982), 1029–1054. DOI:10.2307/1912775

[27] A. C. Harvey: Estimating regression models with multiplicative heteroscedasticity. Econometrica *44* (1976), 461–465. DOI:10.2307/1913974

[28] J. Hausman, W. Newey, T. Voutersen, J. Chao, and N. Swanson: Instrumental variable estimation with heteroscedasticity and many instruments. Quantitative Economics *3* (2012), 211–255. DOI:10.3982/qe89

[29] I. S. Helland: Partial Least Squares Regression and Statistical Models. Scand. J. Statist. *17* (1990), 97–114.

[30] T. P. Hettmansperger and S. J. Sheather: A cautionary note on the method of least median squares. Amer. Statist. *46* (1992), 79–83. DOI:10.2307/2684169

[31] G. G. Judge, W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T. C. Lee: The Theory and Practice of Econometrics. Second edition. J. Wiley and Sons, New York 1985.

[32] J. Jurečková: Regression quantiles and trimmed least squares estimator under a general design. Kybernetika *20* (1984), 345–357.

[33] J. Kmenta: Elements of Econometrics. Macmillan Publishing Company, New York 1986. DOI:10.3998/mpub.15701

[34] W. S. Krasker: Two-stage bounded-influence estimators for simultaneous equations models. J. Business Econom. Statist. *4* (1986), 437–444. DOI:10.2307/1391499

[35] W. S. Krasker and R. E. Welsch: Resistant estimation for simultaneous – equations models using weighted instrumental variables. Econometrica *53* (1985), 1475–1488. DOI:10.2307/1913223

[36] J. Krishnakumar and E. Ronchetti: Robust-estimators for simultaneous equations models. J. Econometr. *78* (1997), 295–314. DOI:10.1016/s0304-4076(97)80014-0

[37] H. P. Lopuhaa: On the relations between S-estimators and M-estimatros of multivariate location and covariance. Ann. Statist. *17* (1989), 1662–1683. DOI:10.1214/aos/1176347386

[38] R. A. Maronna and S. Morgenthaler: Robust regression through robust covariances. Comm. Statist. – Theory and Methods *15* (1986), 1347–1365. DOI:10.1080/03610928608829187

[39] R. A. Maronna and V. J. Yohai: Asymptotic behaviour of general M-estimates for regression and scale with random carriers. Zeitschrift fűr Wahrscheinlichkeitstheorie und verwandte Gebiete *58* (1981), 7–20. DOI:10.1007/bf00536192

[40] R. A. Maronna and V. J. Yohai: Robust estimation in simultaneous equations models. J. Statist. Planning Inference *57* (1997), 233–244. DOI:10.1016/s0378-3758(96)00046-8

[41] L. Mašíček: Optimality of the least weighted squares estimator. Kybernetika *40* (2004), 715–734.

[42] G. E. Mizon: A simple message for autocorrelation correctors: Don't. J. Econometr. *69* (1995), 267–288. DOI:10.1016/0304-4076(94)01671-l

[43] C. C. Paige and Z. Strakoš: Scaled total least squares fundamentals. Numer. Math. *91* (2002), 117–146. DOI:10.1007/s002110100314

[44] P. C. B. Phillips and V. Solo: Asymptotics for linear processes. Ann. Statist. *20* (1992), 971–1001. DOI:10.1214/aos/1176348666

[45] K. R. Popper: The Logic of Scientific Discovery. (Logik der Forscung, Springer, Vienna 1935). Hutchinson and co., New York 1952.

[46] S. Portnoy: Tightness of the sequence of empiric c.d.f. processes defined from regression fractiles. In: Robust and Nonlinear Time – Series Analysis (J. Franke, W. Hărdle, and D. Martin, eds.), Springer Verlag, New York 1983, pp. 231–246. DOI:10.1007/978-1-4615-7821-5_13

[47] R. C. Rao: Estimation of heteroscedastic variances in linear models. J. Amer. Statist. Assoc. *65* (1970), 161–172. DOI:10.2307/2283583

[48] R. C. Rao: Linear Statistical Inference and Its Applications. J. Wiley and Sons, New York 1973. DOI:10.1002/9780470316436

[49] P. M. Robinson: Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. Econometrica *55* (1987), 875–891. DOI:10.2307/1911033

[50] E. Ronchetti and F. Trojani: Robust inference with GMM estimators. J. Econometrics *101* (2001), 37–69. DOI:10.1016/s0304-4076(00)00073-7

[51] P. J. Rousseeuw: Least median of square regression. J. Amer. Statist. Assoc. *79* (1984), 871–880. DOI:10.1080/01621459.1984.10477105

[52] P. J. Rousseeuw and A. M. Leroy: Robust Regression and Outlier Detection. J. Wiley and Sons, New York 1987. DOI:10.1002/0471725382

[53] P. J. Rousseeuw and V. Yohai: Robust regressiom by means of S-estimators. In: Robust and Nonlinear Time Series Analysis (J. Franke, W. Hărdle, and R. D. Martin, eds.), Lecture Notes in Statistics *26* Springer Verlag, New York 1984, pp. 256–272. DOI:10.1007/978-1-4615-7821-5_15

[54] J. Štěpán: Teorie pravděpodobnosti (Probability Theory). Academia, Praha 1987.

[55] S. Van Huffel: Total least squares and error–in–variables modelling: Bridging the gap between statistics, computational mathematics and enginnering. In: Proc. Computational Statistics, COMPSTAT 2004 (J. Antoch, ed.), Physica Verlag/Springer, Heidelberg 2004, pp. 539–555. DOI:10.1007/978-3-7908-2656-2_44

[56] J. Á. Víšek: A cautionary note on the method of Least Median of Squares reconsidered. In: Trans. Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes (P. Lachout, ed.), Academy of Sciences of the Czech Republic, Praha 1994, pp. 254–259.

[57] J. Á. Víšek: Robust instruments. In: Robust'98 (J. Antoch and G. Dohnal, eds.), Union of Czech Mathematicians and Physicists, Matfyzpress, Praha 1998, pp. 195–224.

[58] J. Á. Víšek: The robust regression and the experiences from its application on estimation of parameters in a dual economy. In: Proc. Conference Macromodels'99, Wroclaw University 1999, pp. 424–445.

[59] J. Á. Víšek: Regression with high breakdown point. In: Robust 2000 (J. Antoch and G. Dohnal, eds.), Union of Czech Mathematicians and Physicists, Matfyzpress, Praha 2000, pp. 324–356.

[60] J. Á. Víšek: The least weighted squares I. The asymptotic linearity of normal equations. Bull. Czech Econometr. Soc. *9* (2002), 31–58.

[61] J. Á. Víšek: The least weighted squares II. Consistency and asymptotic normality. Bull. Czech Econometr. Soc. *9* (2002), 1–28.

[62] J. Á. Víšek: Development of the Czech export in nineties. In: Konsolidace vládnutí a podnikání v České republice a v Evropské unii I. Umění vládnout, ekonomika, politika, Matfyzpress, Praha 2003, pp. 193–220.

[63] J. Á. Víšek: Robustifying instrumental variables. In: Proc. COMPSTAT'2004 (J. Antoch, ed.), Physica Verlag/Springer, pp. 1947–1954. DOI:10.1007/978-3-7908-2656-2

[64] J. Á. Víšek: Instrumental weighted variables – algorithm. In: Proc. COMPSTAT 2006 (A. Rizzi and M. Vichi, eds.), Physica Verlag/Springer, Heidelberg 2006, pp. 777–786. DOI:10.1007/978-3-7908-1709-6

[65] J. Á. Víšek: Kolmogorov–Smirnov statistics in multiple regression. In: Proc. ROBUST 2006 (J. Antoch and G. Dohnal, eds.), pp. 367–374.

[66] J. Á. Víšek: Consistency of the instrumental weighted variables. Ann. Inst. Statist. Math. *61* (2009), 543–578.

[67] J. Á. Víšek: Robust error – term – scale estimate. In: IMS Collections. Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: Festschrift for Jana Jurečková, 2010, pp. 254–267. DOI:10.1007/s10463-007-0159-8

[68] J. Á. Víšek: Heteroscedasticity resistant robust covariance matrix estimator. Bull. Czech Econometric Society *17* (2010), 33–49.

[69] J. Á. Víšek: Consistency of the least weighted squares under heteroscedasticity. Kybernetika *47* (2011), 179–206.

[70] J. Á. Víšek: Weak $\sqrt{n}$ – consistency of the least weighted squares under heteroscedasticity. Acta Universitatis Carolinae, Mathematica et Physica *2* (2011), 51, 71–82.

[71] J. Á. Víšek: Empirical distribution function under heteroscedasticity. Statistics *45* (2011), 497–508. DOI:10.1080/02331881003768891

[72] J. Á. Víšek: Robustifying estimation of the model with fixed and random effects. Part I – Theoretical considerations. Part II – Numerical study. Workshop on Algorithm for Outliers/regressors Selection organized by Bent Nielsen, Nuffield College, Oxford 2013. Methodology and Computing in Applied Probability *17* (2014), 4, 999–1014. DOI:10.1007/s11009-014-9432-5

[73] R. Wagenvoort and R. Waldmann: On B-robust instrumental variable estimation of the linear model with panel data. J. Econometr. *106* (2002), 297–324. DOI:10.1016/s0304-4076(01)00102-6

[74] H. White: A heteroskedasticity – consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica *48* (1980), 817–838. DOI:10.2307/1912934

[75] J. M. Wooldridge: Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge 2001. (Second edition 2008.)

[76] J. M. Wooldridge: Introductory Econometrics. A Modern Approach. MIT Press, Cambridge 2006. (Second edition 2009.)

Jan Ámos Víšek, Department of Macroeconomics and Econometrics, Faculty of Social Sciences, Charles University, Opletalova 26, Praha 1, the Czech Republic.
e-mail: visek@fsv.cuni.cz