Ann Inst Stat Math (2009) 61:543–578
DOI 10.1007/s10463-007-0159-8

Consistency of the instrumental weighted variables

Jan Ámos Víšek

Received: 7 June 2006 / Revised: 16 April 2007 / Published online: 24 November 2007

Abstract A robust version of method of Instrumental Variables accommodating the idea of an implicit weighting the residuals is proposed and its properties studied. Firstly, it is shown that all solutions of the corresponding normal equations are bounded in probability. Then the weak consistency of them is proved. The algorithm, evaluating the estimate, is described and results of small MC study discussed.

Keywords Robustness · Instrumental variables · Implicit weighting · Consistency of estimate by instrumental weighted variables

1 Introduction of basic 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 r.v.'s are defined on a basic probability space $(\Omega, \mathcal{A}, P)$. The linear regression model given as

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

Research was supported by grant of GA ČR number 402/06/0408.

J. Á. Víšek
Department of Stochastic Informatics, Institute of Information Theory & Automation, Academy of Sciences of the Czech Republic, Praha 8, Czech Republic

J. Á. Víšek (✉)
Department of Macroeconomics and Econometrics, Faculty of Social Sciences, Charles University, Smetanovo nábřeží 6, 110 01 Prague 1, Czech Republic
e-mail: visek@mbox.fsv.cuni.cz

will be considered (all vectors throughout the paper will be considered to be the column ones). We shall assume that:

C1 The sequence $\left\{(X_i^{'}, e_i)^{'}\right\}_{i=1}^{\infty}$ is sequence of independent and identically distributed $(p+1)$-dimensional random vectors (i.i.d. r.v.'s) with absolutely continuous distribution function $F_{X,e}(x,v)$. Moreover, $\mathbb{E}\left\{(X_1^{'}, e)^{'} \cdot (X_1^{'}, e)\right\}$ is positive definite matrix and the density $f_{e|X}(v|X_1 = x)$ is uniformly in x bounded by a positive constant U_e.

Remark 1 Let us notice that we have not assumed that the explanatory variables X_i's and the error terms e_i's are not correlated. If the model (1) contains the intercept, we have $X_{i1} = 1, i = 1, 2, \ldots, n$.

The *error term* is in econometric texts called *disturbance*. We will use mostly the former and only in the case of mentioning some economic applications we employ the later one.

In what follows $F_X(x)$ and $F_e(r)$ will denote the corresponding marginals of $F_{X,e}(x,r)$. Finally, let us recall that the (*Ordinary*) *Least Squares* (OLS) are the most frequently used estimator of regression coefficients.

Definition 1 The estimator of the regression coefficient given as

$$\hat{\beta}^{(\mathrm{OLS},n)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} \left(Y_i - X_i^{'}\beta\right)^2 = \underset{\beta \in R^p}{\arg\min} \left\{(Y - X\beta)^{'} (Y - X\beta)\right\}$$

(where $X = (X_1, X_2, \ldots, X_n)^{'}$ is the design matrix and $Y = (Y_1, Y_2, \ldots, Y_n)^{'}$ is response vector) is called the (*Ordinary*) *Least Squares*.

Sometimes, there are reasons, why the observations are to have different influence on the value of the estimator of regression coefficients. Then the classical statistics and econometrics advise to utilize the *Weighted Least Squares* (WLS) given as follows.

Definition 2 Let $U_n : \{1, 2, \ldots, n\} \to [0, 1]$ and denote $U_n(i) = w_i$. Moreover, let $W = \mathrm{diag}\{w_1, w_2, \ldots, w_n\}$ be diagonal matrix of weights and $w = (w_1, w_2, \ldots, w_n)'$ the vector of weights. Then the solution of the extremal problem

$$\hat{\beta}^{(\mathrm{WLS},n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w_i \left(Y_i - X_i^{'}\beta\right)^2$$

$$= \underset{\beta \in R^p}{\arg\min} \left\{(Y - X\beta)^{'} W (Y - X\beta)\right\} = \left(X^{'}WX\right)^{-1} X^{'}WY. \tag{2}$$

is called the *Weighted Least Squares*.

Remark 2 The mapping U_n represents some *external rule* which is establish *prior* to evaluating $\hat{\beta}^{(\mathrm{WLS},n,w)}$. One of rules, sometimes (or frequently?) used, is that one based on the diagonal elements of the *hat matrix* $X\left(X'X\right)^{-1}X'$, see Chatterjee and Hadi (1988).

2 Recalling reasons for instrumental variables

It is well known that in the case when the orthogonality condition $\mathbb{E}\{X_i e_i\} = 0$ is broken, the ordinary least squares are not consistent. The best known example of the situation, when the orthogonality condition fails, is the model assuming that the explanatory variables are measured with random error. Assume that

$$Y_i = V_i' \beta^0 + u_i, \quad i = 1, 2, \ldots, n \tag{3}$$

with $\mathbb{E}u_i = 0$ and $\mathbb{E}u_i^2 = \sigma^2 \in (0, \infty)$ and that we observe $\tilde{V}_i = V_i + \eta_i$, assuming usually that $\mathbb{E}\eta_i = 0$, $\mathbb{E}\eta_i \cdot \eta_i' = \Sigma_\eta$ with Σ_η nonsingular and $\mathbb{E}\eta_i \cdot u_i = 0$. Then, substituting $\tilde{V}_i = V_i + \eta_i$ into (3), we obtain

$$Y_i = \left(\tilde{V}_i - \eta_i\right)' \beta^0 + u_i = \tilde{V}_i' \beta^0 - \eta_i' \beta^0 + u_i = \tilde{V}_i' \beta^0 + w_i, \tag{4}$$

where $w_i = -\eta_i' \beta^0 + u_i$. But then

$$\mathbb{E}\left(\tilde{V}_i \cdot w_i\right) = \mathbb{E}\left[(V_i + \eta_i) \cdot \left(-\eta_i' \beta^0 + u_i\right)\right] = -\Sigma_\eta \beta^0.$$

Then $\beta^0 \neq 0$ implies that $\Sigma_\eta \beta^0 \neq 0$ and then due to the fact that

$$\hat{\beta}^{(\mathrm{OLS},n)} = \left(\tilde{V}' \tilde{V}\right)^{-1} \tilde{V}' Y = \left(\frac{1}{n} \tilde{V}' \tilde{V}\right)^{-1} \frac{1}{n} \tilde{V}' Y = \beta^0 + \left(\frac{1}{n} \tilde{V}' \tilde{V}\right)^{-1} \frac{1}{n} \tilde{V}' w, \tag{5}$$

the OLS-estimator of regression coefficients of model (3) is inconsistent. Another example considers the lagged response variable as explanatory one, see Judge (1985) or Víšek (1998a).

The problem is treated, *in econometrics,* by means of the *Method of Instrumental Variables*. Another possibility how to solve the problem is to find so called the *Total Least Squares*, see e.g. Van Huffel (2004).

Definition 3 For any sequence of p-dimensional random vectors $\{Z_i\}_{i=1}^{\infty}$ the solution(s) of the (vector) equation

$$\sum_{i=1}^{n} Z_i \left(Y_i - X_i' \beta\right) = 0 \tag{6}$$

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

Remark 3 The elements of the sequence $\{Z_i\}_{i=1}^{\infty}$ are usually called *instruments*. In the case that the model (1) contains intercept, without loss of generality we may assume that $Z_{i1} = 1$ and $\mathbb{E}Z_{ij} = 0$, $j = 2, 3, \ldots, p$ and $i = 1, 2, \ldots$. We do not lose generality at first, due to the fact that $Z_{i1} = 1$ represents constants and hence they cannot be correlated with the error terms (in fact we have then $Z_{i1} = X_{i1}$). Secondly, what concerns the assumption that $\mathbb{E}Z_{ij} = 0$, $j = 2, 3, \ldots, p$, if it would not be fulfilled, we can "move" $\mathbb{E}Z_{ij}$ into the intercept of the original model (1).

Sometimes (see e.g. Judge 1985) $\hat{\beta}^{(IV,n)}$ is defined as a solution of the extremal problem

$$\hat{\beta}^{(IV,n)} = \underset{\beta \in R^p}{\arg\min} \left\{ (Y - X\beta)' ZZ' (Y - X\beta) \right\}$$

where $Z = (Z_1, Z_2, \ldots, Z_n)'$ is the matrix of instruments, X is the design matrix and Y is the response vector. Similarly as in the case of the (*Ordinary*) *Least Squares*, sometimes we have reasons for employing the classical *Weighted Instrumental Variables*

$$\hat{\beta}^{(\mathrm{WIV},n,W)} = \underset{\beta \in R^p}{\arg\min} \left\{ (Y - X\beta)' WZZ'W (Y - X\beta) \right\} = \left(Z'WX \right)^{-1} Z'WY \quad (7)$$

where W is a diagonal matrix of weights. Let us stress that the weights are again assigned to the observation a priori, usually according to an external (heuristic, frequently geometric) rule.

For the heuristics which show the reasons for defining $\hat{\beta}^{(\mathrm{IV},n)}$ in just described way see Bowden and Turkington (1984), Judge (1985), Manski and Pepper (2000), and Stock and Trebbi (2003). In nineties the method became a standard tool in many case studies of dynamic regression model since the correlation of explanatory variables and disturbances frequently appeared (in economic data). Many papers considering possibilities how to select the instruments for explanatory variables brought applicable results, see e.g. Arellano and Bond (1991), Arellano and Bover (1995), Erickson (2001), Hahn and Hausman (2002), Heckman (1996), and Sargan (1988) [for examples of implementation see: for SAS—Der and Everitt (2002), for R and S-PLUS—Fox (2002)].

As (6) is an analogy of the *normal equations* for the *Ordinary Least Squares*, $\hat{\beta}^{(\mathrm{IV},n)}$ is not robust with respect to the outliers and/or leverage points. Hence we are going to define its robustified version. We shall use the idea of *implicit weighting the squared residuals* which was firstly employed in the method of the *Least Weighted Squares*, see Víšek (2000c).

3 Why the implicit weighting of residuals

Prior to continuing, we need to enlarge a bit the notations. For any $\beta \in R^p$, define the ith residual as $r_i(\beta) = Y_i - X_i'\beta$ and $r^2_{(h)}(\beta)$ the hth order statistic among the squared residuals, i.e. we have

$$r^2_{(1)}(\beta) \le r^2_{(2)}(\beta) \le \cdots \le r^2_{(n)}(\beta). \quad (8)$$

Without loss of generality we may assume that $\beta^0 = 0$ (otherwise write $\beta - \beta^0$ instead of β).

Víšek (1992, 1996, 2002c) revealed that for the M-estimator with discontinuous ψ-function, *the deletion of even one observation may cause very large change of the estimate*. Víšek (2000b) conjectured and Víšek (2006d) established the same result for the *Least Trimmed Squares* (LTS). Similarly, it appeared that robust, especially *the high breakdown point estimators can be very sensitive to a very small change of data*. It started with the paper by Hettmansperger and Sheather (1992) showing by a case study that *Least Median of Squares* estimator (LMS) (Rousseeuw 1984) changes a lot its value when small change data is made. Their result was due to a bad algorithm, they used, and Víšek (1994) corrected the result employing the algorithm by Boček and Lachout (1995). However the phenomenon really exists, for the theoretical explanation see Víšek (1996b, 2000a). Both these unpleasant consequences of (high) robustness have one denominator, namely that the estimators do relay to much on a group of observations, they have selected (considering these observations to be "clean" or "proper", as you want), while the others are assumed to be contamination, i.e. they are deleted from the data. A remedy can be to weight down the observations which seem to be suspicious, i.e. to depress their influence on the value of the estimator smoothly. It led to a proposal of the *Least Weighted Squares* (LWS) in the form (Víšek 2000c; see also Víšek 2002a,b):

Definition 4 Let $w : [0, 1] \to [0, 1]$ is a weight function. Then the solution of the extremal problem

$$\hat{\beta}^{(\mathrm{LWS},n,w)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta) \tag{9}$$

will be called the *Least Weighted Squares*.

Remark 4 Let us mention that the Definition 2 recalled the classical *Weighted Least Squares*. Just defined *Least Weighted Squares* $\hat{\beta}^{(\mathrm{LWS},n,w)}$ differ from the *Weighted Least Squares* $\hat{\beta}^{(\mathrm{WLS},n,w)}$ by the *implicit* assigning the weights which may lead to the improvement in efficiency of estimation. It happens in the case when the leverage points, i.e. observations having the vector of the explanatory variables far away from the other data, are present among the data and they were generated by model in question. There can be also leverage points which represent contamination of data and they can (seriously) damage the estimation. We are able, e. g. by the hat matrix $X\left(X'X\right)^{-1}X'$, (usually) recognize the presence of leverage points among the data but it is not so simple to decide whether they are "in model" or whether they are contamination, see again Chatterjee and Hadi (1988).

As the *Least Trimmed Squares* and the *Least Median of Squares* are special cases of the *Least Weighted Squares*, it is straightforward that LWS can adapt to various situations. It hints that by "tailoring" the weight function to the character of data, we can create the estimator which is "appropriately robust" but avoiding the problems we have discussed a few lines earlier. Moreover, when we put some lower bound on values on the weight function, we facilitate the use of the estimator also for the *panel data* where we cannot afford to delete any observation completely - since otherwise we disturb the correlation structure of data. In addition, avoiding the discontinuous weight function we get rid of the high subsample sensitivity while keeping all plausible (robust) properties for finite sizes of data sets. That is why in what follows we shall assume that the weight function has following properties:

C2 Weight function $w : [0, 1] \to [0, 1]$ is absolutely continuous and nonincreasing, with the derivative $w'(\alpha)$ bounded from below by $-L$ ($L > 0$), $w(0) = 1$.

Please see also Čížek (2002) where the estimator is called the *Smoothed Least Trimmed Squares*. Although this name indicates that for a *special case* of weight function, we obtain the *Least Trimmed Squares* (LTS) as a *special case* of the *Least Weighted Squares*, it may however obscure the fact that LWS are able to control subsample sensitivity (see Víšek 1996, 2000c, 2002c). The same is true about the behaviour of LTS versus LWS with respect to a small shift of an observation (see Víšek 1996b, 2000a). The last but not least, as we have already mentioned, LWS can be used for panel data processing, while LTS can not because the deletion of (even only) one observation from panel data may destruct the correlation structure of the error terms and/or of explanatory variables.

For any $i \in \{1, 2, \ldots, n\}$ and any $\beta \in R^p$, let us define the *random rank of the ith residual* as

$$\pi(\beta, i) = j \in \{1, 2, \ldots, n\} \;\Leftrightarrow\; r_i^2(\beta) = r_{(j)}^2(\beta) \tag{10}$$

[the definition is an analogy of rank which is used in nonparametric statistics, see e.g. Hájek and Šidák (1967)]. Then we have

$$\hat{\beta}^{(\mathrm{LWS},n,w)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{\pi(\beta, i) - 1}{n}\right) r_i^2(\beta). \tag{11}$$

Now, we are going to show that (11) (and hence also (9)) has always a solution. In order to see it, let us denote for any $n \in \mathcal{N}$ by $\mathcal{P}_n$ be the set of all permutations of the indices $\{1, 2, \ldots, n\}$ and denote π_i the ith coordinate of the vector $\pi \in \mathcal{P}_n$. (The following considerations do not represent an algorithm for the evaluation of LWS. The algorithm will be discussed later directly for the proposed *Instrumental Weighted Variables*.) Let us consider following steps:

1. For any $\beta \in R^p$ and arbitrary $\pi \in \mathcal{P}_n$ put $S(\beta, \pi) = \sum_{i=1}^{n} w\left(\frac{\pi_i - 1}{n}\right) r_i^2(\beta)$.
2. Recalling that we have defined $\pi(\beta, i)$ in (10) ($i = 1, 2, \ldots, n$), for any $\beta \in R^p$ put $\pi(\beta) = (\pi(\beta, 1), \pi(\beta, 2), \ldots, \pi(\beta, n))' \in \mathcal{P}_n$. As $\pi(\beta) \in \mathcal{P}_n$ we have

$$\min_{\beta \in R^p} \min_{\pi \in \mathcal{P}_n} \sum_{i=1}^{n} w\left(\frac{\pi_i - 1}{n}\right) r_i^2(\beta) \le \min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{\pi(\beta, i) - 1}{n}\right) r_i^2(\beta),$$

i.e.

$$\min_{\beta \in R^p} \min_{\pi \in \mathcal{P}_n} S(\beta, \pi) \leq \min_{\beta \in R^p} S(\beta, \pi(\beta)). \tag{12}$$

3. Fix $\tilde{\beta} \in R^p$ and notice that according to the definition in the step 1 and due to (10) we have

$$S\left(\tilde{\beta}, \pi(\tilde{\beta})\right) = \sum_{i=1}^{n} w\left(\frac{\pi(\tilde{\beta}, i) - 1}{n}\right) r_i^2(\tilde{\beta}) = \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\tilde{\beta}). \tag{13}$$

But it means that the smallest residual obtains the largest weight, the second smallest residuals obtains the second largest weight, etc. Finally, any sum, in which the weights are prescribed to residuals in any other way, can't be smaller. Hence for any $\beta \in R^p$ and $\pi \in \mathcal{P}_n$, we have

$$S(\beta, \pi(\beta)) \leq S(\beta, \pi). \tag{14}$$

4. (12) and (14) yield

$$\min_{\beta \in R^p} \min_{\pi \in \mathcal{P}_n} S(\beta, \pi) = \min_{\beta \in R^p} S(\beta, \pi(\beta)). \tag{15}$$

5. Fix $\omega_0 \in \Omega$, $\pi \in \mathcal{P}_n$, and evaluate the (classical) *Weighted Least Squares*, please see Definition 2), with the mapping $U_n(i) = U_n^{(\pi)}(i) = w\left(\frac{\pi_i - 1}{n}\right)$, i.e. with the weight matrix $W(\pi) = \text{diag}\left\{w\left(\frac{\pi_1-1}{n}\right), w\left(\frac{\pi_2-1}{n}\right), \ldots, \; w\left(\frac{\pi_n-1}{n}\right)\right\}$. In this case $U_n(i) = U_n^{(\pi)}(i)$, $i = 1, 2, \ldots, n$, is uniquely given by π and we shall write in what follows $\hat{\beta}^{(\text{WLS},\, n,\, \pi)}$ instead of $\hat{\beta}^{(\text{WLS},\, n,\, U_n^{(\pi)})}$.

$$\hat{\beta}^{(\text{WLS},\, n,\, \pi)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{\pi_i - 1}{n}\right)\left(Y_i - X_i'\beta\right)^2 = \left(X'W(\pi)X\right)^{-1} X'W(\pi)Y$$

where $Y = (Y_1, Y_2, \ldots, Y_n)'$ and $X = (X_1, X_2, \ldots, X_n)'$. Then we have for any $\beta \in R^p$,

$$S\left(\hat{\beta}^{(\text{WLS},\, n,\, \pi)}, \pi\right) \leq S(\beta, \pi). \tag{16}$$

6. Repeat it for all $\pi \in \mathcal{P}_n$ and for our $\omega_0 \in \Omega$ (we have fixed in step 5) define $\pi(\omega_0)$ by

$$\pi(\omega_0) = \arg\min_{\pi \in \mathcal{P}_n} S\left(\hat{\beta}^{(\text{WLS},\, n,\, \pi)}, \pi\right).$$

7. Then for any $\pi \in \mathcal{P}_n$,

$$S\left(\hat{\beta}^{(\mathrm{WLS},\ n,\ \pi(\omega_0))}, \pi(\omega_0)\right) \leq S\left(\hat{\beta}^{(\mathrm{WLS},\ n,\ \pi)}, \pi\right). \tag{17}$$

8. Due to (17) and then due to (16), for any $\tilde{\pi} \in \mathcal{P}_n$ and any $\tilde{\beta} \in R^p$,

$$S\left(\hat{\beta}^{(\mathrm{WLS},\ n,\ \pi(\omega_0))}, \pi(\omega_0)\right) \leq S\left(\hat{\beta}^{(\mathrm{WLS},\ n,\ \tilde{\pi})}, \tilde{\pi}\right) \leq S\left(\tilde{\beta}, \tilde{\pi}\right),$$

   i.e., due to the fact that $\tilde{\pi} \in P_n$ and $\tilde{\beta} \in R^p$ were arbitrary,

$$S\left(\hat{\beta}^{(\mathrm{WLS},\ n,\ \pi(\omega_0))}, \pi(\omega_0)\right) = \min_{\beta \in R^p} \min_{\pi \in \mathcal{P}_n} S(\beta, \pi). \tag{18}$$

   Finally, due to (15) and then due to (13),

$$S\left(\hat{\beta}^{(\mathrm{WLS},\ n,\ \pi(\omega_0))}, \pi(\omega_0)\right) = \min_{\beta \in R^p} S(\beta, \pi(\beta)) = \min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta)$$

   and hence, due to definition of $\hat{\beta}^{(\mathrm{LWS},\ n,\ w)}(\omega_0)$ (see (9)), we have $\hat{\beta}^{(\mathrm{WLS},\ n,\ \pi(\omega_0))}(\omega_0) = \hat{\beta}^{(\mathrm{LWS},\ n,\ w)}(\omega_0)$.
9. Repeating steps 1–8 for all ω's, we conclude the proof of existence of solution of (11).

As a byproduct of the previous considerations we have found that the *Least Weighted Squares* estimator $\hat{\beta}^{(\mathrm{LWS},\ n,\ w)}(\omega_0)$ is, at fixed $\omega_0 \in \Omega$, equal to the (classical) *Weighted Least Squares* estimator $\hat{\beta}^{(\mathrm{WLS},\ n,\ \pi(\omega_0))}(\omega_0)$ with the weights $w(\pi(\omega_0)) = \left(w(\frac{\pi_1(\omega_0)-1}{n}), w(\frac{\pi_2(\omega_0)-1}{n}), \ldots, w(\frac{\pi_n(\omega)-1}{n})\right)'$. On the other hand, the *Weighted Least Squares* estimator $\hat{\beta}^{(\mathrm{WLS},\ n,\ \pi(\omega_0))}(\omega_0)$ is (one of) the solution(s) of normal equations

$$\sum_{i=1}^{n} w_i X_i \left(Y_i - X_i'\beta\right) = 0$$

with $w_i = w\left(\frac{\pi_i(\omega_0)-1}{n}\right)$. So, considering successively all $\omega \in \Omega$, we verify that $\hat{\beta}^{(\mathrm{LWS},n,w)}$ is one of solutions of *normal equations*

$$\mathbb{N}E_{Y,X,n}(\beta) = \sum_{i=1}^{n} w\left(\frac{\pi(\beta,i)-1}{n}\right) X_i \left(Y_i - X_i'\beta\right) = 0. \tag{19}$$

[An alternative way is to show that $\frac{\partial \pi(\beta,i)}{\partial \beta} = 0$, see Víšek (2006b).]

4 Instrumental weighted variables

As we have already recalled the estimator obtained by means of the method of *Instrumental Variable* is not robust. On the other hand, the inconsistency of the *Least Squares* when *the orthogonality condition* is broken, as it was explained in Introduction), takes place generally also for the *Least Weighted Squares*. That is why we define an estimator which will be an analogy of the estimator obtained by the method of *Instrumental Variables* but which will weight down the residuals of those observations which seem to be atypical.

Definition 5 For any sequence of p-dimensional random vectors $\{Z_i\}_{i=1}^{\infty}$ the solution(s) of the (vector) equation

$$\mathbb{N}E_{Y,Z,n}(\beta) = \sum_{i=1}^{n} w\left(\frac{\pi(\beta,i)-1}{n}\right) Z_i \left(Y_i - X_i^{'}\beta\right) = 0 \tag{20}$$

will be called the *Instrumental Weighted Variables* estimator (IWV) and denoted by $\hat{\beta}^{(\mathrm{IWV},n,w)}$.

Remark 5 Similarly as in the case of the *Least Weighted Squares* and the classical *Weighted Least Squares*, we shall use in the text which follows both the *Instrumental Weighted Variables* and the (classical) *Weighted Instrumental Variables*, given for some external rule $U_n : \{1, 2, \ldots, n\} \rightarrow [0, 1]$ and the corresponding diagonal matrix $W = \mathrm{diag}\{w_1, w_2, \ldots, w_n\}$ with $w_i = U_n(i)$ and the vector of weights $w = (w_1, w_2, \ldots, w_n)'$ as

$$\hat{\beta}^{(\mathrm{WIV},n,w)} = \left(Z'WX\right)^{-1}\left(Z'WY\right).$$

5 Algorithm for the instrumental weighted variables

We have already learnt that the algorithm for evaluating (a tight approximation to) the robust estimator play an important role for reasonability of any further considerations. We have mentioned the algorithm for the LMS by Boček and Lachout (1995) based on simplex method. Similarly, the algorithm for LTS was discussed and successfully tested in Víšek (1996b, 2000a). Modifying this algorithm so that we evaluate the *Weighted Least Squares* (2) instead of the *Ordinary Least Squares* (5) (at one step of the algorithm) appeared to be reliable algorithm for the *Least Weighted Squares*. Finally, an analogous modification of this algorithm, but now evaluating the *Weighted Instrumental Variables* (7) instead of the *Ordinary Least Squares* (5) can be used for *Instrumental Weighted Variables*. We are going to describe it in details (we shall follow the main steps of Víšek (2006c)). Nevertheless, prior to the explanation of the algorithm, step by step, let us say a few words generally. They allow to keep the below given explanation *reasonably simple and transparent*.

The algorithm consists of two cycles, outer and inner. Both of them need some stopping rule. Let us start with the stopping rule for the inner (the reason is that the stopping rule for outer will be connected with the definition of the stopping rule for the inner cycle).

The stopping rule for the inner cycle
At the moment when we reach, by an iterative process (performed just by the inner cycle), the minimum of the functional $S\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right)$ (see (21)), we stop the cycle. In other words, when the value of the functional $S\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right)$ in two successive steps of the inner cycle is the same, we stop the repetitions of the inner cycle and start a new repetition of the outer cycle. It means that for each repetition of the outer cycle we reach some value of the functional $S\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,W)}\right)$, say $S\left(\hat{\beta}_{(\mathrm{final})}^{(\mathrm{WIV},n,W)}\right)$. Evidently, there is a regression model which corresponds to $S\left(\hat{\beta}_{(\mathrm{final})}^{(\mathrm{WIV},n,W)}\right)$. If the value $S\left(\hat{\beta}_{(\mathrm{final})}^{(\mathrm{WIV},n,W)}\right)$ is *the smallest one among the values, we have reached up to this moment*, we denote the corresponding model the ***best***. Of course, it may happen that the model which was denoted as the ***best***, may lose this "characteristic" at the end of some next repetition of the outer cycle and another model attains this "characteristic". It may also happen that in the repetitions of the outer cycle we repeatedly reach this minimal value and, also the corresponding ***best*** regression model is repeatedly found.

The stopping rule for the outer cycle
Either the number of repetitions of outer cycle reached an a priori given (usually large) number of repetitions (see below, in the stage A, the "*maximal number of repetitions*, say $k_{\max}$"). Or an a priori given number of the same models denoted at given moment as the ***best*** is attained.

If the former branch of the stopping rule was applied, we may expect that there is no reasonable model for data in question. The reason is the fact that the algorithm found plenty (say several hundreds or thousands) different models for our data. If the latter branch of the stopping rule took place, it indicates that (hopefully) there can be some structure in data. Really, if we obtain at the end of outer cycle several times (say 20 times) the same regression model, say $\mathcal{M}$ (which corresponds to the minimum of the functional $S\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right)$ reached during the whole process of repeating the outer cycle) and the total number of repetitions of outer cycle is reasonable (say several hundreds), we may expect that the model $\mathcal{M}$ is acceptable for our data.

Now, let us explain the algorithm *step by step*. We assume that we have at hand data, i.e. the vector of response variable $Y = (Y_1, Y_2, \ldots, Y_n)'$ and matrices of explanatory and of instrumental variables

$$X = \begin{bmatrix} X_{11}, & \cdots, & X_{1p} \\ X_{21}, & \cdots, & X_{2p} \\ \vdots & & \vdots \\ X_{n1}, & \cdots, & X_{np} \end{bmatrix}, \quad Z = \begin{bmatrix} Z_{11}, & \cdots, & Z_{1p} \\ Z_{21}, & \cdots, & Z_{2p} \\ \vdots & & \vdots \\ Z_{n1}, & \cdots, & Z_{np} \end{bmatrix},$$

respectively. The instrumental variables are selected so that they are as much as possible of the same quality and character as the explanatory variables, however they are not correlated with the error terms (disturbances) of the regression model in question. Finally, prior to starting the description of the algorithm, let us recall the notion "points in general position", proposed by Rousseeuw and Leroy (1987) (Chap. 3, paragraph 4). We utilize a bit weaker definition than Rousseeuw and Leroy used, because it is sufficient to our purposes.

Definition 6 A k-tuple of points in the k dimensional Euclidean space R^k is said to be in general position, if they uniquely determine $k-1$ dimensional plane.

Notice that e.g. three points in R^3, if falling on line, don't determine uniquely two-dimensional plane.

Remark 6 Let us realize that in our framework (of the regression model (1)), the minimal number of points in general position is equal to p. Assume, we have selected p points, i.e. $(Y_i, X_{i1}, X_{i2}, \ldots, X_{ip})'$, $i = 1, 2, \ldots, p$. In the case when the model contains intercept, i.e. $X_{i1} = 1$ for $i = 1, 2, \ldots, p$, we take take into account for establishing $p-1$ dimensional plane going through selected observations just $(Y_i, X_{i2}, X_{i3}, \ldots, X_{ip})'$, $i = 1, 2, \ldots, p$. So, we have p points in R^p.

In the case when model does not contain intercept we consider points $(Y_i, X_{i1}, X_{i2}, \ldots, X_{ip})'$, $i = 1, 2, \ldots, p$ and point $(0, 0, 0, \ldots, 0)'$ because employing model without intercept implies that the regression plane goes through the origin (after all, intercept is not estimated and hence any estimated model contains origin).

A. Select some *maximal number of repetitions of the outer cycle*, say $k_{\max}$, *minimal number of the best models* (as the "best model" was described a few lines above), say $b_{\min}$, put $k = 0$, $b = 0$ and $S_{\text{total}} = \infty$.
B. Select randomly p observations $\left(Y_{i_j}, X_{i_j1}, X_{i_j2}, \ldots, X_{i_jp}\right)'$, $j = 1, 2, \ldots, p$. If they are in *general position* evaluate the (regression) plane going through them, otherwise repeat selection of observations. It gives an initial estimate of regression coefficients. Let us denote it by $\hat\beta_{\text{initial}}$. Evaluate for all observations the squared residuals $r_i^2(\hat\beta_{\text{initial}}) = \left(Y_i - X_i'\hat\beta_{\text{initial}}\right)^2$, $i = 1, 2, \ldots, n$, establish the order statistics of them $r_{(i)}^2(\hat\beta_{\text{initial}})$'s, see (8), and the ranks $\pi(\hat\beta_{\text{initial}}, i)$, see (10). Further, define the diagonal matrix

$$W\left(\hat\beta_{\text{initial}}\right) = \text{diag}\left\{w_1^*, w_2^*, \ldots, w_n^*\right\} \quad \text{with} \quad w_i^* = \left(\frac{\pi(\hat\beta_{\text{initial}}, i) - 1}{n}\right)$$

and evaluate

$$S\left(\hat\beta_{\text{initial}}\right) = \left(Y - X\hat\beta_{\text{initial}}\right)' W\left(\hat\beta_{\text{initial}}\right) Z' Z W\left(\hat\beta_{\text{initial}}\right)\left(Y - X\hat\beta_{\text{initial}}\right).$$

Then put $t = 1$ and $S_{\min,k} = S\left(\hat\beta_{\text{initial}}\right)$. Finally, evaluate

$$\hat\beta_{(1)}^{(\text{WIV},n,W)} = \left(Z'W\left(\hat\beta_{\text{initial}}\right)X\right)^{-1}\left(Z'W\left(\hat\beta_{\text{initial}}\right)Y\right).$$

C. Evaluate for all observations the squared residuals $r_i^2(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}) = \left(Y_i - X_i'\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right)^2$, $i = 1, 2, \ldots, n$, establish the order statistics of them $r_{(i)}^2(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)})$'s, see again (8) and the ranks $\pi(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}, i)$, see once again (10). Finally, define the diagonal matrix

$$W\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right) = \mathrm{diag}\left\{w_1^*, w_2^*, \ldots, w_n^*\right\} \quad \text{with } w_i^* = \left(\frac{\pi(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}, i) - 1}{n}\right)$$

and evaluate

$$S\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right) = \left(Y - X\hat{\beta}_{(t)}^{(\mathrm{WIV},n,W)}\right)' W\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right) Z \times Z' W\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right)\left(Y - X\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right). \tag{21}$$

D. If $S\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right) < S_{\min,k}$, put $S_{\min,k} = S\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,W)}\right)$. Otherwise go to F.

E. Evaluate the *Weighted Instrumental Variables*

$$\hat{\beta}_{(t+1)}^{(\mathrm{WIV},n,W)} = \left(Z'W\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right)X\right)^{-1}\left(Z'W\left(\hat{\beta}_{(t)}^{(\mathrm{WIV},n,w)}\right)Y\right),$$

put $t = t + 1$ and go to C.

F. If $S_{\min,k} = S_{\mathrm{total}}$, put $b = b + 1$ (i.e. in just finished inner cycle again the regression model which is at this moment considered as the "best model" up to this moment—as described in previous—was attained).

G. If $S_{\mathrm{total}} > S_{\min,k}$, put $S_{\mathrm{total}} = S_{\min,k}$ and $b = 1$. If $k = k_{\max}$, go to H, otherwise put $k = k+1$. If the number of already estimated models, for which the functional (21) is equal to S_{total} reached $b_{\min}$ (i.e. $b = b_{\min}$), go to H. Otherwise go to B.

H. Return as the estimate by means of the *Instrumental Weighted Variables* $\hat{\beta}^{(\mathrm{IWV},n,w)}$ (see (20)) the estimate of regression coefficients which corresponds to S_{total}.

6 Simulation study

We are going to present and briefly comment results of small simulation study. As the understanding of the simulation study is crucial for attaining a trust to the described algorithm, we try to explain each step very carefully. Three experiments were performed. First of all let us explain what is common for them.

6.1 Common steps of the first and second experiment:

S1 The regression model

$$Y_n = \beta_1 \cdot X_{n1} + \beta_2 \cdot X_{n2} + \beta_3 \cdot X_{n3} + \varepsilon_n, \quad n = 1, 2, \ldots, 50, \tag{22}$$

was considered. After having generated data $\left\{Y_n^*, [X_n^*]', [Z_n^*]'\right\}_{n=1}^{50}$ - details are described below, the estimates by means of the *Ordinary Least Squares*, the *Least Weighted Squares* and the *Instrumental Weighted Variables* were applied on them.

S2 All experiments were ten times repeated. The results were collected in tables below. The results of each repetition create one column in each table of one triplet of tables (more details will be given in the separate explanation for the first, the second and the third experiment).

S3 Each repetition of given experiment contains 100 samples, each sample consists of 50 observations. Each sample was generated as follows.

S4 A finite sequence $\{T_n\}_{n=1}^{52}$ of 3-dimensional random vectors normally distributed with zero mean and unit covariance matrix was generated.

S5 Then, the autoregressive sequence $\{V_n\}_{n=1}^{51}$ was defined by

$$V_n = 0.5 \cdot T_{n+1} + 0.5 \cdot T_n.$$

S6 The sequences of explanatory and instrumental variables, $\{X_n\}_{n=1}^{50}$ and $\{Z_n\}_{n=1}^{50}$, were constructed

$$X_n = V_{n+1} \quad \text{and} \quad Z_n = V_n.$$

Notice please that for any $j, k \in \{1, 2, 3\}$,

$$\begin{aligned}\operatorname{cov}\left(X_{nj}, Z_{nj}\right) &= \operatorname{cov}\left(V_{n+1,j}, V_{nj}\right)\\ &= \operatorname{cov}\left(0.5 \cdot T_{n+2,j} + 0.5 \cdot T_{n+1,j}\ ,\ 0.5 \cdot T_{n+1,j} + 0.5 \cdot T_{nj}\right)\\ &= 0.25\end{aligned}$$

and

$$\operatorname{var}\left(X_{nj}\right) = \operatorname{var}\left(Z_{nj}\right) = 0.5.$$

On the other hand

$$\operatorname{cov}\left(X_{nj}, Z_{nk}\right) = \operatorname{cov}\left(0.5 \cdot T_{n+2,j} + 0.5 \cdot T_{n+1,j}, 0.5 \cdot T_{n+1,k} + 0.5 \cdot T_{nk}\right) = 0.$$

Finally,

$$\operatorname{corr}\left(X_n, Z_n\right) = \begin{bmatrix} 0.5, & 0, & 0 \\ 0, & 0.5, & 0 \\ 0, & 0, & 0.5 \end{bmatrix} \tag{23}$$

i.e. the instrumental variables are correlated with the explanatory ones.

S7 The error terms $\left\{\varepsilon_n^{(\ell)}\right\}_{n=1}^{50}$, $\ell = 1, 2, 3$, were created by

$$\varepsilon_n^{(\ell)} = (-1)^{\ell+1} \sum_{k=1}^{3} T_{n+2,k}$$

(index $\ell = 1, 2, 3$ is for the first, the second and the third experiment, respectively). Notice please that again $\operatorname{cov}\left(X_{nj}, \varepsilon_n^{(\ell)}\right) = (-1)^{\ell+1} 0.5, \quad j = 1, 2, 3,$ $\ell = 1, 2, 3$ and $\operatorname{var}\left(\varepsilon_n^{(\ell)}\right) = 3, \quad \ell = 1, 2, 3$ and hence

$$\operatorname{corr}\left(X_n, \varepsilon_n^{(\ell)}\right) = \operatorname{corr}\left(0.5 \cdot T_{n+2} + 0.5 \cdot T_{n+1}, (-1)^{\ell+1} \sum_{k=1}^{3} T_{n+2,k}\right)$$
$$= \begin{bmatrix} (-1)^{\ell+1} \frac{0.5}{\sqrt{1.5}} \\ (-1)^{\ell+1} \frac{0.5}{\sqrt{1.5}} \\ (-1)^{\ell+1} \frac{0.5}{\sqrt{1.5}} \end{bmatrix}$$

for $\ell = 1, 2, 3$. It indicates that the explanatory variables are correlated with the error terms. On the other hand

$$\operatorname{cov}\left(Z_{nj}, \varepsilon_n^{(\ell)}\right) = 0, \quad j = 1, 2, 3, \quad \ell = 1, 2, 3,$$

i.e. the instrumental variables are not correlated with the error terms.

Now, we are going to describe the special features of the *first experiment*.

S8 The values of response variables Y_n's were calculated as

$$Y_n = 7 \cdot X_{n1} - 3 \cdot X_{n2} - 5 \cdot X_{n3} + \varepsilon_n^{(1)}, \quad n = 1, 2, \ldots, 50.$$

Then for $k = 1, 2, \ldots, 5$, we put $Y_k^* = 5 \cdot Y_k$ and $Y_k^* = Y_k$ for $6 \le k \le 50$, $X_n^* = X_n, Z_n^* = Z_n, \quad n = 1, 2, \ldots, 50$. It means that the first five response variables were "converted" into outliers, or in other words, a contamination of data (on the level of 10% of observations having damaged response variable) was perform.

S9 Data $\left\{\left(Y_n^*, [X_n^*]', [Z_n^*]'\right)'\right\}_{n=1}^{50}$ were taken into account. Then the estimates of regression coefficients estimated by means of the *Ordinary Least Squares*, by the *Least Weighted Squares* and by the *Istrumental Weighted Variables* evaluated. It was done for each of 100 repetitions (each repetition produced data $\left\{\left(Y_n^*, [X_n^*]', [Z_n^*]'\right)'\right\}_{n=1}^{50}$). Let us denote the results $\hat{\beta}_{(k)}^{(LS,50)}$, $\hat{\beta}_{(k)}^{(LWS,50,w)}$ and $\hat{\beta}_{(k)}^{(IWV,50,w)}, \quad k = 1, 2, \ldots, 100$.

S10 The mean values were calculated

$$\hat{\beta}_{(mean)}^{(LS,50)} = \frac{1}{100} \sum_{k=1}^{100} \hat{\beta}_{(k)}^{(LS,50)}, \quad \hat{\beta}_{(mean)}^{(LWS,50,w)} = \frac{1}{100} \sum_{k=1}^{100} \hat{\beta}_{(k)}^{(LWS,50,w)},$$
$$\hat{\beta}_{(mean)}^{(IWV,50,w)} = \frac{1}{100} \sum_{k=1}^{100} \hat{\beta}_{(k)}^{(IWV,50,w)}.$$

These (empirical) means are presented in the next triplet of Tables 1, 2 and 3, in the columns denoted (at the second row of Tables 1, 2 and 3) by 1.

S11 The whole procedure, starting with **S1** up to **S10**, was 10 times repeated and values collected in 1, 2 and 3. Each repetition gave results in one column of 1, 2 and 3, i.e. the results of first repetition are in the second columns of 1, 2 and 3, the results of second repetition are in the third columns of 1, 2 and 3, etc.

The *second experiment*:

S′8 The values of response variables Y_n's were calculated as

$$Y_n = 2.4 \cdot X_{n1} - 3.1 \cdot X_{n2} + 2.8 \cdot X_{n3} + \varepsilon_n^{(1)}, \quad n = 1, 2, \ldots, 50.$$

Then we put $Y_n^* = Y_n$ for $1 \leq n \leq 50$ and $X_n^* = X_n$ and $Z_n^* = Z_n$ for $1 \leq n \leq 45$. Finally, for $n = 46, 47, \ldots, 50$ we put $X_n^* = X_n + 5$ and $Z_n^* = Z_n + 5$. Then we took into account the data $\left\{ \left(Y_n^*, [X_n^*]', [Z_n^*]' \right)' \right\}_{n=1}^{50}$. It means that the last five explanatory as well as instrumental variables were "converted" into leverage points. In other words, a contamination of data (on the level of 10% of data having wrong explanatory as well as instrumental variables) was performed.

S′9, S′10, S′11 The steps **S′9, S′10, S′11** coincide with **S9, S10** and **S11**.

Table 1 The first experiment: $\beta_1 = 7$, $\beta_2 = -3$, $\beta_3 = -5$

	1	2	3	4	5	6	7	8	9	10
Ordinary Least Squares										
$\hat{\beta}_1$	7.996	8.000	8.014	8.027	8.019	8.001	8.003	7.974	8.014	8.027
$\hat{\beta}_2$	−2.022	−1.999	−1.976	−1.998	−1.999	−2.001	−2.001	−2.002	−1.976	−1.998
$\hat{\beta}_3$	−3.971	−3.986	−4.026	−4.031	−4.017	−4.003	−3.995	−3.983	−4.026	−4.03
Least Weighted Squares										
$\hat{\beta}_1$	8.019	7.994	7.980	8.009	8.04	8.008	8.015	7.963	7.980	8.010
$\hat{\beta}_2$	−2.021	−1.998	−2.000	−2.011	−2.026	−2.007	−1.998	−1.976	−2.000	−2.011
$\hat{\beta}_3$	−3.968	−3.978	−4.025	−4.038	−4.002	−4.018	−3.985	−4.013	−4.025	−4.038
Instrumental Weighted Variables										
$\hat{\beta}_1$	6.817	6.735	6.868	7.099	6.871	7.095	7.474	6.688	6.868	7.0993
$\hat{\beta}_2$	−3.790	−3.073	−3.208	−3.276	−3.255	−3.420	−3.915	−3.077	−3.208	−3.276
$\hat{\beta}_3$	−5.534	−4.785	−5.384	−5.144	−5.006	−5.139	−5.747	−5.260	−5.384	−5.144

Table 2 The second experiment: $\beta_1 = 2.4$, $\beta_2 = -3.1$, $\beta_3 = 2.8$

	1	2	3	4	5	6	7	8	9	10
Ordinary Least Squares										
$\hat\beta_1$	3.406	3.393	3.396	3.386	3.395	3.394	3.407	3.412	3.393	3.400
$\hat\beta_2$	−2.100	−2.111	−2.088	−2.105	−2.085	−2.104	−2.107	−2.103	−2.095	−2.096
$\hat\beta_3$	3.793	3.819	3.788	3.823	3.797	3.809	3.797	3.797	3.801	3.789
Least Weighted Squares										
$\hat\beta_1$	3.405	3.393	3.394	3.377	3.396	3.388	3.419	3.403	3.407	3.398
$\hat\beta_2$	−2.099	−2.109	−2.085	−2.102	−2.070	−2.101	−2.113	−2.101	−2.096	−2.098
$\hat\beta_3$	3.777	3.823	3.784	3.829	3.793	3.805	3.811	3.798	3.796	3.788
Instrumental Weighted Variables										
$\hat\beta_1$	2.446	2.296	2.160	2.352	2.221	2.289	2.227	2.311	2.316	2.343
$\hat\beta_2$	−3.261	−3.218	−3.222	−3.122	−3.125	−3.172	−3.254	−3.200	−3.102	−2.999
$\hat\beta_3$	2.892	2.832	2.748	2.896	2.632	2.603	2.797	2.677	2.688	2.742

Table 3 The third experiment: $\beta_1 = -1$, $\beta_2 = 4$, $\beta_3 = 2$

	1	2	3	4	5	6	7	8	9	10
Ordinary Least Squares										
$\hat\beta_1$	−0.025	0.025	−0.011	0.001	−0.015	0.011	0.006	−0.002	−0.006	−0.011
$\hat\beta_2$	5.013	4.979	4.996	5.006	5.006	4.985	4.986	5.017	5.001	5.027
$\hat\beta_3$	3.008	2.999	3.018	2.993	3.014	3.001	3.006	2.989	3.012	2.994
Least Weighted Squares										
$\hat\beta_1$	−0.012	0.023	−0.004	0.005	−0.019	0.006	0.013	0.001	−0.022	−0.006
$\hat\beta_2$	5.010	4.973	4.997	5.005	4.991	4.986	4.990	5.022	4.997	5.027
$\hat\beta_3$	3.007	3.007	3.026	3.000	3.016	2.998	3.008	2.981	3.007	2.995
Instrumental Weighted Variables										
$\hat\beta_1$	−1.151	−1.077	−1.176	−1.109	−0.966	−1.045	−1.025	−1.057	−1.034	−1.008
$\hat\beta_2$	3.961	3.894	3.921	3.933	3.895	3.833	3.768	3.789	3.801	3.920
$\hat\beta_3$	1.986	1.846	1.893	1.840	1.891	1.862	1.998	1.997	1.743	1.831

The *third experiment*:

S′8 The values of response variables Y_n's were calculated as

$$Y_n = -X_{n1} - 4 \cdot X_{n2} + 2 \cdot X_{n3} + \varepsilon_n^{(1)}, \quad n = 1, 2, \ldots, 50.$$

Then for $n = 1, 2, \ldots, 5$ we put $Y_n^* = 5 \cdot Y_n$ and $Y_n^* = Y_n$ for $6 \leq n \leq 50$. Moreover, for $n = 46, 47, \ldots, 50$ we put $X_n^* = 5 \cdot X_n$ and $Z_n^* = 5 \cdot Z_n$. Finally, $X_n^* = X_n$ and $Z_n^* = Z_n$ for $1 \leq n \leq 45$. Then we took into account the data $\left\{\left(Y_n^*, [X_n^*]', [Z_n^*]'\right)'\right\}_{n=1}^{50}$. It means that the first five response variables were again "converted" into outliers and the last five explanatory as well as instrumental variables were "converted" into leverage points. In other words, a contamination of data (on the level of 10% of observations having damaged response variable and another 10% of them having wrong explanatory as well as instrumental variables) was performed.

S′9, S′10, S′11 The steps **S′9**, **S′10**, **S′11** coincide with **S9, S10** and **S11**.

(All programs for evaluating all employed estimators as well as the "framework" for the simulation study are available from the author on request.)

6.2 Conclusions of simulation study

It is evident that the contamination 10% together with correlation between the regressors and the error terms destroyed the *Ordinary Least Squares* as well as the *Least Weighted Squares*. The situation under presence of outliers can be coped quite well by the *Instrumental Weighted Variables*. The performance of the *Instrumental Weighted Variables* under presence of leverage points is nearly of the same quality.

There are at least two things which may be of interest. Firstly, the estimation is satisfactorily good although the correlation between the explanatory and the instrumental variables is rather weak, see (23). In practice, the economic data often exhibit higher autocorrelation in the time series of explanatory variables and hence we have (frequently) at hand better instruments, see e. g. Víšek (2003b).

Secondly, the estimation by means of the *Ordinary Least Squares* and by the *Least Weighted Squares* was mainly destroyed by correlation between the explanatory variables and error terms, as it is indicated by a similar "bias" of the respective estimates. If the damage would be caused (mainly) by contamination, the bias would be much larger for the *Ordinary Least Squares* in comparison with the *Least Weighted Squares* [which are able to cope with the contamination of data in the case when there is no the correlation between explanatory variables and error terms, see Plát (2004b)]. The phenomenon can be presumably explained as follows: For the *Ordinary Least Squares* we have

$$\hat{\beta}^{(\mathrm{OLS},n)} = \left(X'X\right)^{-1} X'Y = \beta^0 + \left(\frac{1}{n}X'X\right)^{-1} \frac{1}{n}X'e,$$

compare with (5). A similar asymptotic (Bahadur) representation can be derived for $\hat{\beta}^{(\mathrm{LWS},n,w)}$, see Mašíček (2003) or Víšek (2002b). Then 50 observations already "activated" the *law of large numbers* and so $\left(\frac{1}{n}X'X\right)^{-1}$ and $\frac{1}{n}X'e$ are already near to $\mathbb{E}X_1X_1'$ and to $\mathbb{E}X_1e_1$, respectively, and hence the bias.

So, it seems that (a bit preliminary) conclusion may be that neglecting the correlation between regressors and error terms may be much more dangerous than the omission of the presence of contamination of data, especially when it is not of very large (high, if you want) level.

7 Consistency of the instrumental weighted variables

For any $\beta \in R^p$ the distribution of the absolute value of residual will be denoted $F_\beta(r)$, i.e.

$$F_\beta(r) = P(|Y_1 - X_1^{'}\beta| < r) = P(|e_1 - X_1^{'}\beta| < r) \tag{24}$$

(remember, we have assumed $\beta^0 = 0$). Similarly, for any $\beta \in R^p$ the empirical distribution of the absolute value of residual will be denoted $F_\beta^{(n)}(r)$. It means that, denoting the indicator of a set A by $I\{A\}$, we have

$$F_\beta^{(n)}(r) = \frac{1}{n}\sum_{j=1}^{n} I\left\{|r_j(\beta)| < r\right\} = \frac{1}{n}\sum_{j=1}^{n} I\left\{|e_j - X_j^{'}\beta| < r\right\}. \tag{25}$$

Realize now that denoting $|r_i(\beta)| = a_i(\beta)$, the order statistics $a_{(i)}(\beta)$'s and the order statistics of the squared residuals $r_{(i)}^2(\beta)$'s assign to given fix observation the same rank, i.e. the residual of given fix observation (say for $i = i_0$, for some $i_0 \in \{1, 2, \ldots, n\}$) is in the sequence

$$r_{(1)}^2(\beta) \le r_{(2)}^2(\beta) \le \cdots \le r_{(n)}^2(\beta) \tag{26}$$

and in the sequence

$$a_{(1)}(\beta) \le a_{(2)}(\beta) \le \cdots \le a_{(n)}(\beta) \tag{27}$$

on the same position. In other words, if the squared residual of the jth observation is the ℓth smallest among the squared residuals, also the absolute value of the jth residual is the ℓth smallest among the absolute values of residuals. Then looking for the empirical distribution function of the absolute values of residuals, we observe that the first "jump" (having the magnitude $\frac{1}{n}$) is at the smallest absolute value of residuals, i.e. at $a_{(1)}(\beta)$. But due to the sharp inequality in the definition (25) of the empirical distribution function (see (25)), it holds $F_\beta^{(n)}(a_{(1)}(\beta)) = 0$. Hence, at the ℓth "jump" at $a_{(\ell)}(\beta)$, we have $F_\beta^{(n)}(a_{(\ell)}(\beta)) = \frac{\ell-1}{n}$. Now, let us realize that $a_{(\pi(\beta,i))}(\beta) = |r_i(\beta)|$. It means that at the $\pi(\beta, i)$th "jump", we have

$$F_\beta^{(n)}(a_{(\pi(\beta,i))}(\beta)) = F_\beta^{(n)}(|r_i(\beta)|) = \frac{\pi(\beta, i) - 1}{n} \tag{28}$$

(for $\pi(\beta)$ see (10)) and so (20) can be written as

$$\sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) Z_i \left(Y_i - X_i^{'}\beta\right) = 0. \tag{29}$$

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

$$F_{\beta' ZX'\beta}(u) = P(\beta^{'}Z_1X_1^{'}\beta < u) \tag{30}$$

and similarly as in (24) and (25), the corresponding empirical distribution will be denoted $F^{(n)}_{\beta' ZX'\beta}(u)$, so that

$$F^{(n)}_{\beta' ZX'\beta}(u) = \frac{1}{n}\sum_{j=1}^{n} I\left\{\beta' Z_j X_j' \beta < u\right\} = \frac{1}{n}\sum_{j=1}^{n} I\left\{\omega \in \Omega \;:\; \beta' Z_j(\omega) X_j'(\omega)\beta < u\right\}. \tag{31}$$

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

$$\gamma_{\lambda,a} = \sup_{\|\beta\|=\lambda} F_{\beta' ZX'\beta}(a). \tag{32}$$

Notice please that due to the fact that the surface of ball $\{\beta \in R^p, \|\beta\| = \lambda\}$ is compact, there is $\beta_\lambda \in \{\beta \in R^p, \|\beta\| = \lambda\}$ so that

$$\gamma_{\lambda,a} = F_{\beta_\lambda' ZX'\beta_\lambda}(a). \tag{33}$$

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

$$\tau_\lambda = -\inf_{\|\beta\|\le\lambda} \beta' \mathbb{E}\left[Z_1 X_1' \cdot I\{\beta' Z_1 X_1' \beta < 0\}\right]\beta. \tag{34}$$

Notice please that $\tau_\lambda \ge 0$ 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

$$\tau_\lambda = -\tilde{\beta}' \mathbb{E}\left[Z_1 X_1' \cdot I\{\tilde{\beta}' Z_1 X_1' \tilde{\beta} < 0\}\right]\tilde{\beta}. \tag{35}$$

The classical regression analysis accepted the assumption that $\mathbb{E} Z_1 X_1'$ is regular and $\mathbb{E}\{e_1 | Z_1\} = 0$ (see e.g. Bowden and Turkington 1984 or Judge 1985) to be able to prove consistency of the estimator obtained by the method of *Instrumental Variables*. We need to assume similar ones. The following more or less academic considerations give us an inspiration. Transforming the variables so that we put $\tilde{X}_{11} = X_{11}$ and for any $j = 2, 3, \ldots, p$,

$$\tilde{X}_{1j} = X_{1j} - \sum_{k=1}^{j-1} \lambda_{jk} \tilde{X}_{1k}$$

where λ_{jk} are selected so that $\mathrm{cov}(\tilde{X}_{1j}, \tilde{X}_{1k}) = 0$ for $j \ne k$, we have the matrix $\mathbb{E}\tilde{X}_1 \tilde{X}_1'$ diagonal and the model for transformed data, namely $Y_i = \tilde{X}_i' \tilde{\beta} + u_i$ has the same "explanatory" abilities as (1). New explanatory variables $\left\{\tilde{X}_i\right\}_{i=1}^{\infty}$ would not allow presumably so direct (physical, biological, economic etc.) interpretation, nevertheless they have also at least one advantage, namely that overfitting the model does not imply automatically a decrease of efficiency of the etimates of regression coefficients, see Chatterjee and Hadi (1988).

Assuming that we shall look for a sequence of instrumental variables $\left\{\tilde{Z}_i\right\}_{i=1}^{\infty}$ for the sequence of transformed explanatory variables $\left\{\tilde{X}_i\right\}_{i=1}^{\infty}$. We would like to find it so that also $\mathbb{E}\tilde{Z}_1\tilde{X}_1^{'}$ is regular and diagonal. In other words, we would like to find the instrumental variables so that $\tilde{Z}_{1j}$ is correlated only with $\tilde{X}_{1j}$ (of course for all $j = 2, , 3 \ldots, p$). Assume that it is possible. Then we may assume that $\mathbb{E}\tilde{Z}_{1j}\tilde{X}_{1j} > 0$ (otherwise we take instead of $\tilde{Z}_{1j}$ the instrumental variable $-\tilde{Z}_{1j}$). Then however $\mathbb{E}\tilde{Z}_1\tilde{X}_1^{'}$ is positive definite. These (let us repeat academic) considerations can inspire us to made following assumptions about the instrumental variables:

C3 The instrumental variables $\{Z_i\}_{i=1}^{\infty}$ are independent and identically distributed with distribution function $F_Z(z)$. Moreover, they are independent from the sequence $\{e_i\}_{i=1}^{\infty}$. Further, the joint distribution function $F_{X,Z}(x,z)$ is absolutely continuous, $\mathbb{E}\left\{w(F_{\beta^0}(|e_1|))Z_1X_1^{'}\right\}$ as well as $\mathbb{E}Z_1Z_1^{'}$ are positive definite (one can compare C3 with Víšek (1998a) where we considered instrumental M-estimators and the discussion of assumptions for M-instrumental variables was given) and there is $q > 1$ so that $\mathbb{E}\left\{\|Z_1\| \cdot \|X_1\|\right\}^q < \infty$. Finally, there is $a > 0$, $b \in (0,1)$ and $\lambda > 0$ so that

$$a \cdot \left(b - \gamma_{\lambda,a}\right) \cdot w(b) > \tau_\lambda \tag{36}$$

for $\gamma_{\lambda,a}$ and τ_λ given by (32) and (34).

Remark 7 Let us briefly discuss assumptions we have made. Let us recall that the Least Squares ($\beta^{(\mathrm{LS},n)}$) are optimal only under normality of error terms. Here the optimality means that they reach the lower Rao–Cramer bound (of course, in multivariate Rao–Cramer lemma we consider the ordering of the covariance matrices in the sense of ordering the positive definite matrices).On the other hand, a small departure from normality may cause (and usually does) a large decrease of efficiency (see e.g. Fisher 1920, 1922). So, without the assumption of normality of the error terms $\hat{\beta}^{(\mathrm{LS},n)}$ is much worse, in fact they are the best unbiased estimator only in the class of linear unbiased estimators, for a discussion showing that restriction on linear estimators can be drastic see Hampel et al. (1986). Sometimes, however we may meet with the statement that we do not need necessarily the normality of error terms, just because $\hat{\beta}^{(\mathrm{LS},50)}$ is still (without normality) the best unbiased estimator in the class of linear unbiased estimators. And the restriction on the class of linear unbiased estimators is justified by a claim that we have to restrict ourselves on the class of linear estimators, as in the the class of linear unbiased estimators, the estimators are *scale-* and *regression-eqivariant*. Let us recall that having denoted $M(n, p)$ the set of all matrices of type $(n \times p)$ and recalling that the estimator $\hat{\beta}$ can be considered as a mapping

$$\hat{\beta}(Y, X): \; M(n, p+1) \rightarrow R^p,$$

the estimator $\hat{\beta}$ of β^0 is called *scale-equivariant*, if for any $c \in R^+$, $Y \in R^n$ and $X \in M(n, p)$ we have

$$\hat{\beta}(cY, X) = c\hat{\beta}(Y, X)$$

and *regression-equivariant* if for any $b \in R^p$, $Y \in R^n$ and $X \in M(n, p)$,

$$\hat{\beta}(Y + Xb, X) = \hat{\beta}(Y, X) + b.$$

But, there are a lot of nonlinear estimators which are scale- and regression-equivariant. In the regression framework, the estimators as the *Least Median of Squares*, the *Least Trimmed Squares* or the *Least Weighted Squares* can serve as examples [for an interesting discussion of this topic see again Hampel et al. (1986), and also Bickel (1975) or Jurečková and Sen (1993)].

Since LWS are also based on L_2-metric, we guess that they are approximately optimal for finite sample sizes under the (approximative) normality of error terms, for some hint consult Mašíček (2003). As the present proposal of robustified instrumental variables is based on the same metric (due to the normal equations (20)), we can expect that the estimate can be approximately optimal under (approximative) normality of the error terms. But then our assumptions seem to be quite acceptable.

The only assumption which deserve further discussion is the assumption (36). We are going to show that it is a restriction on the weight function w. Let us return to (32) (or to (33)). We have

$$\gamma_{\lambda,a} = F_{\beta_\lambda' Z X' \beta_\lambda}(a) = P\left(\beta_\lambda' Z_1 X_1' \beta_\lambda \le 0\right) + P\left(0 < \beta_\lambda' Z_1 X_1' \beta_\lambda \le a\right).$$

If we assume for a while $Z_j = X_j$, for any fix $\lambda \in R^+$ we have

$$\lim_{a \to 0^+} F_{\beta_\gamma' X X' \beta_\gamma}(a) = 0 \tag{37}$$

but generally, (if Z_j is not X_j) we have (again for fix $\lambda \in R^+$)

$$\lim_{a \to 0^+} F_{\beta_\gamma' Z X' \beta_\gamma}(a) = P\left(\beta_\lambda' Z_1 X_1' \beta_\lambda \le 0\right). \tag{38}$$

On the other hand, for any $a > 0$ we have

$$\gamma_{\lambda,a} < 1. \tag{39}$$

Now let us turn to τ_λ. As

$$\mathbb{E}\left|\beta' Z_1 X_1' \beta\right| \le \|\beta\|^2 \, \mathbb{E}\{\|Z_1\| \, \|X_1\|\} \le \|\beta\|^2 \, \mathbb{E}\{\|Z_1\| \, \|X_1\|\}^q < \infty,$$

we have

$$\limsup_{\|\beta\| \to 0} \left| \beta' \mathbb{E}\left[Z_1 X_1' I\{\beta' Z_1 X_1' \beta < 0\}\right] \beta \right| = 0. \tag{40}$$

In other words, τ_λ can be done arbitrary small (just selecting $\lambda \in R^+$ so that $\|\lambda\|$ is small). It says that if $w(b) \equiv 1$, there is $b \in (0,1) > \gamma_{\lambda,a}$ (even for any $a > 0$). It means that (37), (38), (39) and (40) indicate that (36) can be always fulfilled but we may have restricted possibility to depress the influence of "bad" observations.

In what follows there are defined some constants inside the proofs of lemmas. They are assumed to be defined only inside the corresponding proof. Now we can prove:

Lemma 1 *Let Conditions* **C1**, **C2** *and* **C3** *be fulfilled. Then for any* $\varepsilon > 0$ *and* $\delta > 0$ *there is* $\theta > \delta$ *and* $\Delta > 0$ *such that*

$$P\left(\left\{\omega \in \Omega : \inf_{\|\beta\| \geq \theta} -\frac{1}{n}\beta^{'}\mathbb{N}E_{Y,Z,n}(\beta) > \Delta\right\}\right) > 1 - \varepsilon.$$

In other words, any sequence $\left\{\hat{\beta}^{(\mathrm{IWV},n,w)}\right\}_{n=1}^{\infty}$ *of the solutions of the (sequence of) normal equations* $\mathbb{N}E_{Z,n}(\hat{\beta}^{(\mathrm{IWV},n,w)}) = 0$ *(see (19)) is bounded in probability.*

Proof The plan of the proof is simple: We shall show that for any positive ε there are positive κ and n_ε so that for any $n > n_\varepsilon$ with probability at least $1 - \varepsilon$, outside the ball of the diameter κ the expression $-\frac{1}{n}\beta^{'}\mathbb{N}E_{Y,Z,n}(\beta)$ is positive. The way how to demonstrate it is based on the idea to show that quadratic part of $-\frac{1}{n}\beta^{'}\mathbb{N}E_{Y,Z,n}(\beta)$ is positive and hence for enough large β it overcomes the linear one. In order to establish the positivity of quadratic part, we evaluate the number of terms in the corresponding sum which are negative and the number of terms which are positive and simultaneously having weight larger than a constant c (of course, there are some other positive terms, contribution of which will be neglected, since their weights are smaller than c). Since the mean of sum of the negative terms is bounded from below in probability, we estimate from below the value of quadratic term.

First of all, denote the set of all indices $i = 1, 2, \ldots, n$ by I_n, for b from Condition **C3** the set of indices for which $F_\beta^{(n)}(|r_i(\beta)|) \geq b$ by I_b and finally, for any $\beta \in R^p$ denote the set of indices for which $\beta^{'} Z_i X_i^{'} \beta < a$ by $I_a(\beta)$. Of course, the set of indices I_b also depends on β but due to the fact that we shall need only an upper estimate of number of elements of I_b which doesn't depend on β, we have omitted β in notations. Returning to (26) or (27), we easy verify that the empirical d.f. overcomes b at least at its $[nb] + 1$ jump, i.e. at least $[nb]$ of n observations are in I_b^C. Hence

$$\#I_b \leq n \cdot (1 - b) + 1 \tag{41}$$

where $\#A$ stays for the number of elements of the set A. Denote $\mathbb{E}\{|e_1| \cdot \|Z_1\|\} = \gamma^{(1)}$ and $\mathbb{E}\{\|X_1\| \cdot \|Z_1\|\} = \gamma^{(2)}$ and fix a positive ε. Further, let $\lambda > 0$ be that from **C3** and put (see (36))

$$\delta = \frac{a \cdot \left(b - \gamma_{\lambda,a}\right) \cdot w(b) - \tau_\lambda}{5}.$$

Recalling that we have assumed that $\beta^0 = 0$, we shall consider for $\beta \in R^p$

$$-\frac{1}{n}\beta^{'}\mathbb{N}E_{Y,Z,n}(\beta) = -\frac{1}{n}\sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right)\beta^{'}Z_i\left(e_i - X_i^{'}\beta\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right)\beta^{'}Z_iX_i^{'}\beta - \frac{1}{n}\sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right)e_iZ_i^{'}\beta. \quad (42)$$

Let us start with the first term in (42) and put $\tau^{(1)} = \delta/(2L\cdot\gamma^{(2)}\cdot\lambda^2)$, for L see **C2**. Due to Lemma 4 we can find $n_1 \in \mathcal{N}$ so that for any $n > n_1$ there is a set $B_n^{(1)}$ such that $P(B_n^{(1)}) > 1 - \varepsilon/5$ and for any $\omega \in B_n^{(1)}$,

$$\sup_{\beta\in R^p}\sup_{r\in R}\left|F_\beta^{(n)}(r) - F_\beta(r)\right| \le \tau^{(1)}.$$

Employing the law of large numbers, find $n_2 \in \mathcal{N}$ so that for any $n > n_2$ there is a set $B_n^{(2)}$ such that $P(B_n^{(2)}) > 1 - \varepsilon/5$ and for any $\omega \in B_n^{(2)}$

$$\frac{1}{n}\sum_{i=1}^{n}\|Z_i\|\cdot\|X_i\| < 2\gamma^{(2)}.$$

Since then for any $n >\ \max\{n_1, n_2\}$ and any $\omega \in B_n^{(1)} \cap B_n^{(2)}$ (of course $P\left(B_n^{(1)} \cap B_n^{(2)}\right)$ $> 1 - \frac{2\varepsilon}{5}$)

$$\frac{1}{n}\sup_{\beta\in R^p}\left\|\sum_{i=1}^{n}\left\{w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) - w\left(F_\beta(|r_i(\beta)|)\right)\right\}Z_iX_i^{'}\right\|$$

$$\le \frac{1}{n}L\cdot\tau^{(1)}\cdot\sum_{i=1}^{n}\|Z_i\|\cdot\|X_i\| \le L\cdot\tau^{(1)}\cdot 2\gamma^{(2)} = \frac{\delta}{\lambda^2},$$

we have for any $n > \max\{n_1, n_2\}$, any $\omega \in B_n^{(1)} \cap B_n^{(2)}$ and any $\beta \in R^p$,

$$\frac{1}{n}\sup_{\beta\in R^p}\left|\sum_{i=1}^{n}\left\{w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) - w\left(F_\beta(|r_i(\beta)|)\right)\right\}\beta^{'}Z_iX_i^{'}\beta\right| \le \frac{\delta\cdot\|\beta\|^2}{\lambda^2}. \quad (43)$$

Notice please that for any $\beta \in R^p$, for indices for which $F_\beta^{(n)}(|r_i(\beta)|) \le b$, we have $w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) \ge w(b)$. Now, let us consider for any $\beta \in R^p$,

$$\frac{1}{n}\sum_{i=1}^{n} w\left(F_\beta(|r_i(\beta)|)\right)\beta^{'}Z_iX_i^{'}\beta$$

$$= \frac{1}{n}\sum_{i=1}^{n} w\left(F_\beta(|r_i(\beta)|)\right)\beta^{'}Z_iX_i^{'}\beta\cdot I\{\beta^{'}Z_iX_i^{'}\beta < 0\}$$

$$+\frac{1}{n}\sum_{i=1}^{n} w\left(F_\beta(|r_i(\beta)|)\right)\beta^{'}Z_iX_i^{'}\beta\cdot I\{\beta^{'}Z_iX_i^{'}\beta \ge 0\}$$

$$
\begin{aligned}
&\geq \frac{1}{n}\sum_{i=1}^{n}\beta^{'}Z_iX_i^{'}\beta\cdot I\{\beta^{'}Z_iX_i^{'}\beta<0\}\\
&+\frac{1}{n}\sum_{I_n\setminus I_b}w(b)\beta^{'}Z_iX_i^{'}\beta\cdot I\{\beta^{'}Z_iX_i^{'}\beta\geq 0\}
\end{aligned}
\tag{44}
$$

where we have employed monotonicity of $w(r)$. Notice please that (44) holds for any $\beta \in R^p$. Utilizing Lemma 10 find such $n_3 \in \mathcal{N}$ that for all $n > n_3$ we have

$$
P\left(\left\{\omega\in\Omega\ :\ \inf_{\|\beta\|\leq\lambda}\frac{1}{n}\sum_{i=1}^{n}\beta^{'}Z_iX_i^{'}\beta\cdot I\{\beta^{'}Z_iX_i^{'}\beta<0\}>\tau_\lambda-\frac{\delta}{2}\right\}\right)>1-\frac{\varepsilon}{5}
\tag{45}
$$

and denote the corresponding set by $B_n^{(3)}$. Employing Lemma 5 find $n_4 \in \mathcal{N}$ so that for all $n > n_4$ we have

$$
P\left(\left\{\omega\in\Omega:\sup_{\beta\in R^p}\sup_{u\in R}\left|F^{(n)}_{\beta^{'}ZX^{'}\beta}(u)-F_{\beta^{'}ZX^{'}\beta}(u)\right|\leq\frac{\delta}{2\cdot a\cdot w(b)}\right\}\right)>1-\frac{\varepsilon}{5}
\tag{46}
$$

and denote the corresponding set by $B_n^{(4)}$. Recalling that, due to the fact how the empirical distribution function is defined, we have

$$
F^{(n)}_{\beta^{'}ZX^{'}\beta}(a)=\frac{\#\{i\ :\ \beta^{'}Z_iX_i^{'}\beta<a\}}{n}=\frac{\#I_a(\beta)}{n}
$$

(where again $\#A$ denotes the number of points of the set A), we conclude that (46) implies for any $n > n_4$ and $\omega \in B_n^{(4)}$,

$$
\#I_a(\beta)<\left(F_{\beta^{'}ZX^{'}\beta}(a)+\frac{\delta}{2\cdot a\cdot w(b)}\right)\cdot n\leq\left(\gamma_{\lambda,a}+\frac{\delta}{2\cdot a\cdot w(b)}\right)\cdot n
\tag{47}
$$

(for $\gamma_{\lambda,a}$ see (32)). Finally, find $n_5 \in \mathcal{N}$ so that for all $n > n_5$ we have

$$
\frac{a\cdot w(b)}{n}<\delta.
\tag{48}
$$

Consider $\omega \in B_n^* = B_n^{(3)} \cap B_n^{(4)}$ and $n > \max\{n_3, n_4, n_5\}$. Let us recall once again that for any $\beta \in R^p$, for indices for which $F_\beta^{(n)}(|r_i(\beta)|) \leq b$, we have $w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) \geq w(b)$. Hence, (41) and (47) imply that the number of indices for which $\beta^{'}Z_iX_i^{'}\beta \geq a$ and simultaneously $w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) \geq w(b)$ is at least

$$
n-n\cdot(1-b)-1-n\cdot\left(\gamma_{\lambda,a}+\frac{\delta}{2\cdot a\cdot w(b)}\right)=n\cdot\left(b-\gamma_{\lambda,a}-\frac{\delta}{2\cdot a\cdot w(b)}\right)-1.
$$

Now, taking into account (45) and (48) we have for any $n > \max\{n_3, n_4, n_5\}$, any $\omega \in B_n^* = B_n^{(3)} \cap B_n^{(4)}$ and any $\|\beta\| = \lambda$,

$$\frac{1}{n}\sum_{i=1}^{n} \beta^{'} Z_i X_i^{'} \beta \cdot I\{\beta^{'} Z_i X_i^{'} \beta < 0\} + \frac{1}{n}\sum_{I_n \setminus I_b} w(b)\beta^{'} Z_i X_i^{'} \beta \cdot I\{\beta^{'} Z_i X_i^{'} \beta \geq 0\}$$
$$\geq a \cdot \left(b - \gamma_{\lambda,a} - \frac{\delta}{2 \cdot a \cdot w(b)} - \frac{1}{n}\right) \cdot w(b) - \tau_\lambda - \frac{\delta}{2}$$
$$= a \cdot \left(b - \gamma_{\lambda,a} - \frac{1}{n}\right) \cdot w(b) - \tau_\lambda - \delta > 3\delta.$$

Consider now any $\beta \in R^p$, $\|\beta\| = \theta \geq \lambda$ and put $\tilde{\beta} = \theta^{-1} \cdot \lambda \cdot \beta$. Notice please that for any $\beta \in R^p$ for which $\beta' Z_i X_i' \beta < 0$, also $\tilde{\beta}' Z_i X_i' \tilde{\beta} < 0$ and similarly for the case when $\beta' Z_i X_i' \beta \geq 0$. Then $\left\|\tilde{\beta}\right\| = \lambda$ and hence, again for any $n > \max\{n_3, n_4, n_5\}$ and any $\omega \in B_n^* = B_n^{(3)} \cap B_n^{(4)}$ (due to (44))

$$\frac{1}{n}\sum_{i=1}^{n} w\left(F_\beta(|r_i(\beta)|)\right) \beta^{'} Z_i X_i^{'} \beta$$
$$\geq \frac{1}{n}\sum_{i=1}^{n} \beta^{'} Z_i X_i^{'} \beta \cdot I\{\beta^{'} Z_i X_i^{'} \beta < 0\} + \frac{1}{n}\sum_{I_n \setminus I_b} w(b)\beta^{'} Z_i X_i^{'} \beta \cdot I\{\beta^{'} Z_i X_i^{'} \beta \geq 0\}$$
$$= \left(\frac{\theta}{\lambda}\right)^2 \left\{\frac{1}{n}\sum_{i=1}^{n} \tilde{\beta}^{'} Z_i X_i^{'} \tilde{\beta} \cdot I\{\tilde{\beta}^{'} Z_i X_i^{'} \tilde{\beta} < 0\}\right.$$
$$\left. + \frac{1}{n}\sum_{I_n \setminus I_b} w(b)\tilde{\beta}^{'} Z_i X_i^{'} \tilde{\beta} \cdot I\{\tilde{\beta}^{'} Z_i X_i^{'} \tilde{\beta} \geq 0\}\right\} > 3\left(\frac{\|\beta\|}{\lambda}\right)^2 \delta. \tag{49}$$

Now, we shall consider the second term in (42). Recalling that we have denoted $\mathbb{E}\{|e_i| \cdot \|Z_1\|\} = \gamma^{(1)}$, we can find $n_6 \in \mathcal{N}$ so that for any $n > n_6$ there is $B_n^{(5)}$ so that $P(B_n^{(5)}) > 1 - \varepsilon/5$ and for any $\omega \in B_n^{(5)}$ we have

$$\frac{1}{n}\left|\sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) e_i Z_i^{'} \beta\right| \leq (\gamma^{(1)} + \delta)\|\beta\|. \tag{50}$$

Consider $n > \max\{n_1, n_2, n_3, n_4, n_5, n_6\}$ and $\omega \in B_n = \cap_{j=1}^{5} B_n^{(j)}$. Of course, $P(B_n) > 1 - \varepsilon$ and (42), (43), (49) and (50) imply that for any $\beta \in R^p$, $\|\beta\| \geq \lambda$

$$-\frac{1}{n}\beta^{'}\mathbb{N}E_{Y,Z,n}(\beta) \geq 2\left(\frac{\|\beta\|}{\lambda}\right)^2 \delta - (\gamma^{(1)} + \delta)\|\beta\|.$$

Then there is a $\kappa > 0$ such that for any $\beta \in R^p$, $\|\beta\| > \kappa$ with probability at least $1 - \varepsilon$ we have

$$-\frac{1}{n}\beta^{'}\mathbb{N}E_{Y,Z,n}(\beta) > \delta.$$

□

Remark 8 The fact that for any i and any $\omega \in \Omega$ the matrix $X_i X_i^{'}$ is positive semidefinite allows to prove the same assertion (i.e. that all solutions of the normal equations are bounded in probability) for the *Least Weighted Squares* in significantly simpler way, see Mašíček (2003).

Lemma 2 *Let Conditions* **C1***,* **C2** *and* **C3** *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*

$$P\left(\left\{\omega \in \Omega \ : \ \sup_{\|\beta\|\leq\zeta}\left|\frac{1}{n}\sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right)\beta^{'}Z_i\left(e_i - X_i^{'}\beta\right)\right.\right.\right.$$
$$\left.\left.\left. -\beta^{'}\mathbb{E}\left[w\left(F_{\beta}(|r_1(\beta)|)\right)Z_1\left(e_i - X_1^{'}\beta\right)\right]\right| < \delta\right\}\right) > 1 - \varepsilon.$$

Proof Denoting $\mathbb{E}\{|e_1| \cdot \|Z_1\|\} = \gamma^{(1)}$ and $\mathbb{E}\{\|X_1\| \cdot \|Z_1\|\} = \gamma^{(2)}$, let us fix a positive ε, $\delta \in (0, 1)$ and $\zeta > 0$. Recalling that we have assumed that $\beta^0 = 0$, we shall consider for $\beta \in R^p$, $\|\beta\| \leq \zeta$,

$$-\frac{1}{n}\beta^{'}\mathbb{N}E_{Y,Z,n}(\beta) = -\frac{1}{n}\sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right)\beta^{'}Z_i\left(e_i - X_i^{'}\beta\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right)\beta^{'}Z_iX_i^{'}\beta - \frac{1}{n}\sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right)e_iZ_i^{'}\beta. \tag{51}$$

Let us start with the first term in (51) and put $\tau^{(1)} = \delta/(16\gamma^{(2)}\zeta^2 \cdot L)$, for L see Condition **C2**. Due to Lemma 4 we can find $n_1 \in \mathcal{N}$ so that for any $n > n_1$ there is a set $B_n^{(1)}$ such that $P(B_n^{(1)}) > 1 - \varepsilon/8$ and for any $\omega \in B_n^{(1)}$,

$$\sup_{\beta\in R^p}\sup_{r\in R}\left|F_{\beta}^{(n)}(r) - F_{\beta}(r)\right| \leq \tau^{(1)}. \tag{52}$$

Employing the law of large numbers, find $n_2 > n_1$ so that for any $n > n_2$ there is a set $B_n^{(2)}$ such that $P(B_n^{(2)}) > 1 - \varepsilon/8$ and for any $\omega \in B_n^{(2)}$,

$$\frac{1}{n}\sum_{i=1}^{n}\|Z_i\| \cdot \|X_i\| < 2\gamma^{(2)}. \tag{53}$$

Since then for any $n > n_2$ and any $\omega \in B_n^{(1)} \cap B_n^{(2)}$ (of course $P\left(B_n^{(1)} \cap B_n^{(2)}\right) > 1 - \frac{\varepsilon}{4}$)

$$\frac{1}{n} \sup_{\|\beta\| \le \zeta} \left\| \sum_{i=1}^{n} \left\{ w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) - w\left(F_\beta(|r_i(\beta)|)\right) \right\} Z_i X_i' \right\|$$
$$\le \frac{1}{n} L \cdot \tau^{(1)} \cdot \sum_{i=1}^{n} \|Z_i\| \cdot \|X_i\| \le L \cdot \tau^{(1)} \cdot 2\gamma^{(2)} = \frac{\delta}{8\zeta^2},$$

we have for any $n > n_2$ and any $\omega \in B_n^{(1)} \cap B_n^{(2)}$,

$$\frac{1}{n} \sup_{\|\beta\| \le \zeta} \left| \sum_{i=1}^{n} \left\{ w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) - w\left(F_\beta(|r_i(\beta)|)\right) \right\} \beta' Z_i X_i' \beta \right| \le \frac{\delta}{8}. \tag{54}$$

Employ Lemma 3 and find for $\Delta = \frac{\delta}{16 \cdot L \cdot \gamma^{(2)} \zeta^2}$ such $\tau^{(2)} > 0$ that for

$$\mathcal{T}(\tau^{(2)}) = \left\{ \left\| \beta^{(1)} \right\| \le \zeta, \left\| \beta^{(2)} \right\| \le \zeta, \left\| \beta^{(1)} - \beta^{(2)} \right\| < \tau^{(2)} \right\} \tag{55}$$

we have

$$\sup_{(\beta^{(1)}, \beta^{(2)}) \in \mathcal{T}(\tau^{(2)})} \sup_{r \in R} \left| F_{\beta^{(1)}}(r) - F_{\beta^{(2)}}(r) \right| < \Delta.$$

Then for any $n > n_2$ and any $\omega \in B_n^{(1)} \cap B_n^{(2)}$,

$$\frac{1}{n} \sup_{(\beta^{(1)}, \beta^{(2)}) \in \mathcal{T}(\tau^{(2)})} \left| \sum_{i=1}^{n} \left\{ w\left(F_{\beta^{(2)}}(|r_i(\beta^{(2)})|)\right) \right. \right.$$
$$\left. \left. - w\left(F_{\beta^{(1)}}(|r_i(\beta^{(2)})|)\right) \right\} \left[\beta^{(1)}\right]' Z_i X_i' \beta^{(1)} \right|$$
$$\le L \cdot \Delta \cdot \zeta^2 \cdot \frac{1}{n} \sum_{i=1}^{n} \|Z_i\| \cdot \|X_i\| \le \frac{\delta}{8} \tag{56}$$

(notice that the in the previous inequality the subindices of the d.f.'s are $\beta^{(1)}$ and $\beta^{(2)}$ but the arguments are the same, namely $r_i(\beta^{(2)})$). Further denote $\gamma^{(3)} = \mathbb{E}\{ \|Z_1\| \cdot \|X_1\| \}^q$, $\gamma^{(4)} = \mathbb{E}\, \|X_1\|$ and applying the law of large numbers find $n_3 > n_2$ so that for any $n > n_3$ there is a set $B_n^{(3)}$ such that $P(B_n^{(3)}) > 1 - \varepsilon/8$ and for any $\omega \in B_n^{(3)}$ we have

$$\frac{1}{n} \sum_{i=1}^{n} \{\|Z_i\| \cdot \|X_i\|\}^q < 2\gamma^{(3)} \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^{n} \|X_i\| < 2\gamma^{(4)}.$$

Finally, let us recall that $w(r) \in [0,1]$, so that for any pair $r_1, r_2 \in R$ we have $|w(r_1) - w(r_2)| \leq 1$ and hence for any $q' > 1$,

$$|w(r_1) - w(r_2)|^{q'} \leq |w(r_1) - w(r_2)|. \tag{57}$$

Then select a $\tau^{(3)} \in \left(0, \min\left\{\tau^{(2)}, \delta \cdot \left(2^{q'} \cdot 2^{q} \cdot 8 \cdot U_e \cdot L \cdot \left[\gamma^{(3)}\right]^{\frac{q'}{q}} \cdot \gamma^{(4)} \cdot \zeta^{2q'}\right)^{-1}\right\}\right)$ (for U_e see **C1**) and put

$$\mathcal{T}(\tau^{(3)}) = \left\{\left\|\beta^{(1)}\right\| \leq \zeta, \left\|\beta^{(2)}\right\| \leq \zeta, \left\|\beta^{(1)} - \beta^{(2)}\right\| < \tau^{(3)}\right\}.$$

Employing Hőlder's inequality we arrive at

$$\begin{aligned}
&\sup_{(\beta^{(1)},\beta^{(2)})\in\mathcal{T}(\tau^{(3)})} \frac{1}{n}\left|\sum_{i=1}^{n}\left\{w\left(F_{\beta^{(1)}}(|r_i(\beta^{(2)})|)\right)\right.\right.\\
&\qquad\left.\left. -w\left(F_{\beta^{(1)}}(|r_i(\beta^{(1)})|)\right)\right\}\left[\beta^{(1)}\right]' Z_i X_i' \beta^{(1)}\right|\\
&\leq \sup_{(\beta^{(1)},\beta^{(2)})\in\mathcal{T}(\tau^{(3)})} \left\{\left[\frac{1}{n}\sum_{i=1}^{n}\left|w\left(F_{\beta^{(1)}}(|r_i(\beta^{(2)})|)\right) - w\left(F_{\beta^{(1)}}(|r_i(\beta^{(1)})|)\right)\right|^{q'}\right]^{\frac{1}{q'}}\right.\\
&\qquad\left.\times\left[\frac{1}{n}\sum_{i=1}^{n}\left(\left\|\beta^{(1)}\right\| \cdot \|Z_i\| \cdot \|X_i\| \cdot \left\|\beta^{(1)}\right\|\right)^{q}\right]^{\frac{1}{q}}\right\}\\
&\leq \sup_{(\beta^{(1)},\beta^{(2)})\in\mathcal{T}(\tau^{(3)})} \left\{\left[\frac{1}{n}\sum_{i=1}^{n}\left|w\left(F_{\beta^{(1)}}(|r_i(\beta^{(2)})|)\right) - w\left(F_{\beta^{(1)}}(|r_i(\beta^{(1)})|)\right)\right|\right]^{\frac{1}{q'}}\right.\\
&\qquad\left.\times\zeta^2\left[\frac{1}{n}\sum_{i=1}^{n}(\|Z_i\| \cdot \|X_i\|)^{q}\right]^{\frac{1}{q}}\right\}\\
&\leq \sup_{(\beta^{(1)},\beta^{(2)})\in\mathcal{T}(\tau^{(3)})} \left\{U_e^{\frac{1}{q'}} L^{\frac{1}{q'}}\left[\tau^{(3)}\right]^{\frac{1}{q'}}\left[\frac{1}{n}\sum_{i=1}^{n}\|X_i\|\right]^{\frac{1}{q'}}\right.\\
&\qquad\left.\times\zeta^2\left[\frac{1}{n}\sum_{i=1}^{n}(\|Z_i\|\,\|X_i\|)^{q}\right]^{\frac{1}{q}}\right\}\\
&\leq \zeta^2 \cdot U_e^{\frac{1}{q'}} \cdot L^{\frac{1}{q'}} \cdot \left[\tau^{(3)}\right]^{\frac{1}{q'}} \cdot \left[2\gamma^{(4)}\right]^{\frac{1}{q'}} \cdot \left[2\gamma^{(3)}\right]^{\frac{1}{q}} \leq \frac{\delta}{8}.
\end{aligned} \tag{58}$$

Finally, utilizing Lemma 8 find $\tau^{(4)} \in (0, \min\{\delta/8, \tau^{(3)}\})$ so that for any pair $\|\beta^{(1)}\| \le \zeta$, $\|\beta^{(2)}\| \le \zeta$, $\|\beta^{(1)} - \beta^{(2)}\| \le \tau^{(4)}$, we have

$$\Big|[\beta^{(1)}]\mathbb{E}\left[w\left(F_{\beta^{(1)}}(|r_1(\beta^{(1)})|)\right) Z_1 \left(e_i - X_1^{'}\beta^{(1)}\right)\right] \\ -[\beta^{(2)}]^{'}\mathbb{E}\left[w\left(F_{\beta^{(2)}}(|r_1(\beta^{(2)})|)\right) Z_1 \left(e_i - X_1^{'}\beta^{(2)}\right)\right]\Big| \le \frac{\delta}{8}. \tag{59}$$

Now find a minimal system of open balls of type $\mathcal{B}(\beta, \tau^{(4)})$ covering the p-dimensional ball with center at zero and radius ζ, i.e. $\mathcal{B}(\zeta) = \{\beta \in R^p : \|\beta\| \le \zeta\}$. Of course, due to the compactness of $\mathcal{B}(\zeta)$ the system has finite number of balls, say $K(\zeta)$, and denote this system by $\{\mathcal{B}(\beta^{(j)}, \tau^{(4)})\}_{j=1}^{K(\zeta)}$. Utilizing the law of large numbers find for any $j \in \{1, 2, \ldots, K(\zeta)\}$ some $n_j^* \in \mathcal{N}$ so that for all $n > n_j^*$ the set

$$B_{nj}^{(4)} = \left\{\omega \in \Omega : \frac{1}{n}\left\|\sum_{i=1}^{n}\left\{w\left(F_{\beta^{(j)}}(|r_i(\beta^{(j)})|)\right) X_i X_i' \\ -\mathbb{E}\left[w\left(F_{\beta^{(j)}}(|r_i(\beta^{(j)})|)\right) X_i X_i'\right]\right\}\right\| < \frac{\delta}{8\zeta^2}\right\} \tag{60}$$

has probability at least $1 - \frac{\varepsilon}{8K(\zeta)}$. Finally put $n_{\varepsilon,\delta,\zeta}^{(1)} = \max\left\{n_3, n_1^*, n_2^*, \ldots, n_{K(\zeta)}^*\right\}$ and $B_n = B_n^{(1)} \cap B_n^{(2)} \cap B_n^{(3)} \cap_{j=1}^{K(\zeta)} B_{nj}^{(4)}$. We have $P(B_n) > 1 - \frac{\varepsilon}{2}$. Since for any $n > n_{\varepsilon,\delta,\zeta}^{(1)}$ and any $\beta \in R^p$, $\|\beta\| \le \zeta$ there is $j \in \{1, 2, \ldots, K(\zeta)\}$ so that $\left\|\beta - \beta^{(j)}\right\| < \tau^{(4)}$, taking into account (54), (56), (58), (59) and (60) we have for for any $\omega \in B_n$

$$\sup_{\|\beta\| \le \zeta} \frac{1}{n}\left|\beta^{'}\sum_{i=1}^{n}\left\{w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) Z_i X_i^{'} - \mathbb{E}\left[w\left(F_\beta(|r_1(\beta)|)\right) Z_1 X_1^{'}\right]\right\}\beta\right| < \frac{\delta}{2}. \tag{61}$$

Now, we shall consider the second term in (51). Along similar lines as in the first part of the proof, we can find $n_{\varepsilon,\delta,\zeta}^{(2)} \in \mathcal{N}$ so that for any $n > n_{\varepsilon,\delta,\zeta}^{(2)}$ there is $C_n \subset \Omega$ so that $P(C_n) > 1 - \varepsilon/2$ and for any $\omega \in C_n$ we have

$$\sup_{\|\beta\| \le \zeta} \frac{1}{n}\left|\sum_{i=1}^{n}\left\{w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) e_i Z_i^{'}\beta - \mathbb{E}\left[w\left(F_\beta(|r_1(\beta)|)\right) e_1 Z_1^{'}\beta\right]\right\}\right| < \frac{\delta}{2}. \tag{62}$$

Taking into account (61) and (62), we conclude the proof. □

C4 The vector equation

$$\beta^{'}\mathbb{E}\left[w\left(F_\beta(|r_1(\beta)|)\right) Z_1 \left(e_1 - X_1^{'}\beta\right)\right] = 0 \tag{63}$$

in the variable $\beta \in R^p$ has unique solution $\beta^0 = 0$.

Theorem 1 *Let Conditions* **C1***,* **C2***,* **C3** *and* **C4** *be fulfilled. Then any sequence* $\left\{\hat{\beta}^{(\mathrm{IWV},n,w)}\right\}_{n=1}^{\infty}$ *of the solutions of normal equations* $\mathbb{N}E_{Z,n}(\hat{\beta}^{(\mathrm{IWV},n,w)}) = 0$ *is weakly consistent.*

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

$$P\left(\left\{\omega \in \Omega : \left\|\hat{\beta}^{(\mathrm{IWV},n,w)} - \beta^0\right\| < \delta\right\}\right) > 1 - \varepsilon. \tag{64}$$

So fix $\varepsilon_1 > 0$ and $\delta_1 > 0$. According to Lemma 1 there are $\Delta_1 > 0$ and $\theta_1 > \delta_1$ so that for ε_1 there is $n_{\Delta_1,\varepsilon_1} \in \mathcal{N}$ so that for any $n > n_{\Delta_1,\varepsilon_1}$,

$$P\left(\left\{\omega \in \Omega : \inf_{\|\beta\| \geq \theta_1} -\frac{1}{n}\beta^{'}\mathbb{N}E_{Y,Z,n}(\beta) > \Delta_1\right\}\right) > 1 - \frac{\varepsilon_1}{2}$$

(denote the corresponding set by B_n). It means that for all $n > n_{\Delta_1,\varepsilon_1}$ all solutions of the normal equations $\mathbb{N}E_{Y,Z,n}(\beta) = 0$ are inside the ball $\mathcal{B}(0,\theta_1)$ with probability at least $1 - \frac{\varepsilon_1}{2}$. Now, utilizing Lemma 2 we may find for ε_1, $\delta = \min\{\frac{\Delta_1}{2}, \delta_1\}$ 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 $\omega \in C_n$

$$\sup_{\|\beta\| \leq \theta_1} \left| \frac{1}{n}\sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right)\beta^{'}Z_i\left(e_i - X_i^{'}\beta\right) \right.$$
$$\left. -\beta^{'}\mathbb{E}\left[w\left(F_{\beta}(|r_1(\beta)|)\right)Z_1\left(e_i - X_1^{'}\beta\right)\right] \right| < \delta.$$

But it means that

$$\inf_{\|\beta\| = \theta_1} \left\{-\beta^{'}\mathbb{E}\left[w\left(F_{\beta}(|r_1(\beta)|)\right)Z_1\left(e_i - X_1^{'}\beta\right)\right]\right\} > \frac{\Delta_1}{2} > 0. \tag{65}$$

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

$$\tau_C = \inf_{\beta \in C} \left\{-\beta^{'}\mathbb{E}\left[w\left(F_{\beta}(|r_1(\beta)|)\right)Z_1\left(e_i - X_1^{'}\beta\right)\right]\right\}. \tag{66}$$

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

$$\lim_{k \to \infty} \beta_k^{'}\mathbb{E}\left[w\left(F_{\beta_k}(|r_1(\beta_k)|)\right)Z_1\left(e_i - X_1^{'}\beta_k\right)\right] = -\tau_C.$$

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 continuity of $\beta^{'}\mathbb{E}\left[w\left(F_\beta(|r_1(\beta)|)\right)Z_1\left(e_i - X_1^{'}\beta\right)\right]$ (see Lemma 8) we have

$$-\left[\beta^*\right]^{'}\mathbb{E}\left[w\left(F_{\beta^*}(|r_1(\beta^*)|)\right)Z_1\left(e_i - X_1^{'}\beta^*\right)\right] = \tau_C. \tag{67}$$

Then the continuity of $\beta^{'}\mathbb{E}\left[w\left(F_\beta(|r_1(\beta)|)\right)Z_1\left(e_i - X_1^{'}\beta\right)\right]$ together with Condition **C4** and (65) imply that $\tau_C > 0$ (otherwise there has to be a solution of (63) inside the compact C).

Now, utilizing Lemma 2 once again we may find for $\varepsilon_1, \delta_1, \theta_1$ 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 $\omega \in D_n$

$$\sup_{\|\beta\|\leq\theta_1}\left|\frac{1}{n}\sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right)\beta^{'}Z_i\left(e_i - X_i^{'}\beta\right)\right.$$
$$\left.-\beta^{'}\mathbb{E}\left[w\left(F_\beta(|r_1(\beta)|)\right)Z_1\left(e_i - X_1^{'}\beta\right)\right]\right| < \frac{\tau_C}{2}. \tag{68}$$

But (66) and (68) 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

$$\inf_{\|\beta\|>\delta_1} -\frac{1}{n}\beta^{'}\mathbb{N}E_{Y,Z,n}(\beta) > \frac{\tau_C}{2}. \tag{69}$$

Of course, $P\left(B_n \cap D_n\right) > 1 - \varepsilon_1$. But it means that all solutions of normal equations (63) are inside the ball of radius δ_1 with probability at least $1 - \varepsilon_1$, i.e. in other words, $\hat{\beta}^{(\mathrm{IWV},n,w)}$ is weakly consistent. □

8 Concluding remarks

We have added a small pebble (of mosaic) to equip the *Least Weighted Squares* by additional (or alternative, if you want) methods (similarly as the classical *(Ordinary) Least Squares* are equipped) to be able to build up the regression model in the situations when the basic assumptions are broken or when the "main" method is not suitable. We have discussed the situation when orthogonality condition is broken and hence the *(Ordinary) Least Squares* are biased. That is why we have proposed the robustified version of the classical instrumental variables. The other situation, e.g. discrete or limited response variable, will require also modifications of the *Least Weighted Variables*

The lack of such tools and of course the lack of easy available and reliable implementations of robust methods hamper a wide (or at least wider than the present) employment of robust methods. We have at present at hand already a reliable algorithm for the *Instrumental Weighted Variables* which is based on the same idea as the algorithm which for the *Least Trimmed Squares* was tested in Víšek (1996b, 2000a). The algorithm appeared to be reliable, we have referred about it on COMPSTAT 2006, Víšek (2006c). A paper with a sufficient number of case studies of its applications is under preparation. We can send on the request the code of algorithms (in MATLAB or MATHEMATICA) for TLS, LWS and IWV to anybody who would like to try to use it.

There are already available some other results for the *Least Weighted Squares*, see Kalina (2004), Mašíček (2004a,b), and Plát (2004a,b) which enlarge possibility of their applications. Some other results, similar to those established in Víšek (1998b, 2000d, 2002d, 2003a) for other type of robust estimators, are under progress.

So, we hope that the present result can help to improve a bit the situations when *"not using robust methods along with the classical ones we take a risk of obtaining misleading results of case studies under presence of even a slight contamination"*, see Hampel et al. (1986).

Acknowledgment We would like to express our gratitude to the anonymous referees for carefully reading the manuscript. In fact, a lot of improvements were made according to their recommendations. They read carefully even the revised version and advised corrections and/or improvements, especially of (discussion of) the Monte Carlo study.

Appendix

The appendix collects lemmas proofs of which are either simple "computation" on several lines or they are chains (sometimes long and boring) of routine statistical steps. Exception is the proof of Lemma 4 which was already published and the proofs of next two lemmas (Lemmas 5 and 6) which are "copies" of the proof of Lemma 4. Proofs (in details) are available from author on request.

Lemma 3 *Under Conditions* **C1** *the distribution function $F_\beta(r)$ is, uniformly with respect to $r \in R$, uniformly continuous in β, i.e. for any $\delta > 0$ there is $\varsigma \in (0, 1)$ so that for any pair $\beta^{(1)}$ and $\beta^{(2)}$ such that $\left\|\beta^{(1)} - \beta^{(2)}\right\| < \varsigma$ we have*

$$\sup_{r \in R} \left| F_{\beta^{(1)}}(r) - F_{\beta^{(2)}}(r) \right| \leq \delta.$$

Proof is just evaluation of $\sup_{r \in R} \left| F_{\beta^{(1)}}(r) - F_{\beta^{(2)}}(r) \right| \leq \delta$ which makes use the fact that

$$F_\beta(r) = P\left(\left| e_1 - X_1^{'}\beta \right| < r \right) = \int I\{ \left| s - x^{'}\beta \right| < r \} \mathrm{d}F_{X,e}(x, s).$$

□

Lemma 4 *Let Conditions* **C1** *hold and fix arbitrary $\varepsilon > 0$. Then there are $K < \infty$ and $n_\varepsilon \in \mathcal{N}$ so that for all $n > n_\varepsilon$*

$$P\left(\left\{ \omega \in \Omega : \sup_{v \in R^+} \sup_{\beta \in R^p} \sqrt{n} \left| F_\beta^{(n)}(v) - F_\beta(v) \right| < K \right\} \right) > 1 - \varepsilon. \tag{70}$$

For the *proof* of lemma, see Víšek (2006a).

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 (30)) and the corresponding empirical distribution by $F^{(n)}_{\beta' ZX'\beta}(u)$ (see (31)).

Lemma 5 *Let Condition* **C3** *hold and fix arbitrary* $\varepsilon > 0$. *Then there are* $K < \infty$ *and* $n_\varepsilon \in \mathcal{N}$ *so that for all* $n > n_\varepsilon$

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

Proof runs along the same lines as the proof of previous lemma. □

Lemma 6 *Let Condition* **C3** *hold and fix arbitrary* $\varepsilon > 0$. *Then there is* $K_\varepsilon < \infty$ *and* $n_\varepsilon \in \mathcal{N}$ *so that for all* $n > n_\varepsilon$

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

Proof runs again along the same lines as the proof of Lemma 4.

Lemma 7 *Let Condition* **C3** *hold and fix arbitrary* $\varepsilon > 0$ *and* $\zeta > 0$. *Then there is* $\Delta > 0$ *so that*

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

Proof is a chain of routine considerations employing the continuity of the probability measure.

Lemma 8 *Let Conditions* **C1**, **C2** *and* **C3** *hold. Then for any positive* ζ,

$$\beta^{'} \mathbb{E}\left[w\left(F_\beta(|r_1(\beta)|)\right) Z_1 \left(e_i - X_1^{'}\beta\right)\right]$$

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

Proof utilizes the assumption that the derivative of the weight function is bonded from below and that the ball $\mathcal{B} = \{\beta \in R^p : \|\beta\| \le \zeta\}$ is compact (for finite ζ).

Lemma 9 *Let Conditions* **C1**, **C2** *and* **C3** *hold. Then for any positive* ζ,

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

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

Proof runs along the same lines as the proof of previous lemma.

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

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

Lemma 10 *Let Conditions* **C1**, **C2** *and* **C3** *be fulfilled. Then for any* $\varepsilon > 0$, $\delta \in (0, 1)$ *and* $\zeta \ge 1$ *there is* $n_{\varepsilon,\delta,\zeta} \in \mathcal{N}$ *so that for any* $n > n_{\varepsilon,\delta,\zeta}$ *we have*

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

Proof in this case is long chain of steps utilizing law of large numbers, compactness of the ball $\{\beta \in R^p : \|\beta\| \le \zeta\}$ and Cauchy–Schwarz inequality.

Lemma 11 *Let Conditions* **C1** *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* $n > n_{\varepsilon,\delta}$,

$$P\left(\left\{\omega \in \Omega : \sup_{r \in R} \sup_{\|\beta^{(1)} - \beta^{(2)}\| < \zeta} \left| F_{\beta^{(1)}}^{(n)}(r) - F_{\beta^{(2)}}^{(n)}(r) \right| < \delta \right\}\right) > 1 - \varepsilon. \tag{71}$$

Proof is a straightforward application of Lemmas 3 and 4.

References

Arellano, M., Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies, 58*, 277–297.

Arellano, M., Bover, O. (1995). Another look at the instrumental variables estimation of error components models. *Journal of Econometrics, 68*(1), 29–52.

Bickel, P. J. (1975). One-step Huber estimates in the linear model. *Journal of American Statistical Association*, *70*, 428–433.

Boček, P., Lachout, P. (1995). Linear programming approach to LMS-estimation. *Memorial Volume of Computational Statistics and Data Analysis, 19*, 129–134.

Bowden, R. J., Turkington, D. A. (1984). *Instrumental Variables.* Cambridge: Cambridge University Press.

Breiman, L. (1968). *Probability*. London: Addison-Wesley.

Chatterjee, S., Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression.* New York: Wiley.

Čížek, P. (2002). Robust estimation with discrete explanatory variables. *COMPSTAT 2002, Berlin*, pp. 509–514.

Der, G., Everitt, B. S. (2002). *A handbook of statistical analyses using SAS.* Boca Raton: Chapman and Hall/CRC Press.

Erickson, T. (2001). Constructing instruments for regression with measurement error when no additional data are available: *Econometrica, Notes and Comments*, *69*, 221–222.

Fisher, R. A. (1920). 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 of Royal Astrophysical Society*, *80*, 758–770.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of Royal Society London Series*, *222*, 309–368.

Fox, J. (2002). *An R and S-PLUS companion to applied regression.* Thousand Oaks: SAGE Publications.

Glivenko, V. I. (1933). Sulla determinazione empirica delle leggi di probabilita. *Giornale Istituto Italiano Attuari*, *4*, 92.

Hahn, J., Hausman, J. (2002). A new specification test for the validity of instrumental variables. *Econometrica*, *70*, 163–189.

Hájek, J., Šidák, Z. (1967). *Theory of rank test.* New York: Academic Press.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A. (1986). *Robust statistics—The approach based on influence functions.* New York: Wiley.

Heckman, J. J. (1996). Randomization as instrumental variables. *The Review of Economics and Statistics*, *78*, 336–341.

Hettmansperger, T. P., Sheather, S. J. (1992). A cautionary note on the method of least median squares. *The American Statistician*, *46*, 79–83.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., Lee, T. C. (1985). *The theory and practice of econometrics* (2nd edn). New York: Wiley.

Jurečková, J., Sen, P. K. (1993). Regression rank scores scale statistics and studentization in linear models. In *Proceedings of the fifth Prague symposium on asymptotic statistics* (pp. 111–121). Heidelberg: Physica-Verlag/Springer.

Kalina, J. (2004). Durbin–Watson test for least weighted squares. In *Proceedings of COMPSTAT 2004* (pp. 1287–1294). Heidelberg: Physica-Verlag/Springer.

Manski, F. C., Pepper, J. V. (2000). Monotone instrumental variables: With application to the return to scholling. *Econometrica, 68*, 997–1010.

Mašíček, L. (2003). Diagnostika a sensitivita robustních odhadů.(Diagnostics and sensitivity of robust estimators, in Czech.) Disertační práce (PhD-disertation).

Mašíček, L. (2004a). Consistency of the least weighted squares estimator. In *Statistics for industry and technology* (pp. 183–194). Basel: Birkhấser Verlag.

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

Plát, P. (2004a). Modifikace Whiteova testu pro nejmenší vážené čtverce. (Modification of White's test for the least weighted squares, in Czech.) In J. Antoch, G. Dohnal (Eds.), *ROBUST 2004* (pp. 291–298).

Plát, P. (2004b). The least weighted squares estimator. *Proceedings of COMPSTAT 2004* (pp. 1653–1660). Heidelberg: Physica-Verlag/Springer.

Rousseeuw, P. J. (1984). Least median of square regression. *Journal of American Statistical Association*, *79*, 871–880.

Rousseeuw, P. J., Leroy, A. M. (1987). *Robust regression and outlier detection.* New York: Wiley.

Sargan, J. D. (1988). Testing for misspecification after estimating using instrumental variables. In Massouumi, E. (Ed.) *Contribution to econometrics: John Denis Sargan (Vol. 1). Cambridge: Cambridge University Press.*

Stock, J. H., Trebbi, F. (2003). Who invented instrumental variable regression? *Journal of Economic Perspectives*, *17*, 177–194.

Van Huffel, S. (2004). Total least squares and error-in-variables modelling: Bridging the gap between statistics, computational mathematics and enginnering. In J. Antoch (Ed.) *Proceedings in Computational Statistics*, *COMPSTAT 2004* (pp. 539–555). Heidelberg: Physica-Verlag/Springer.

Víšek, J. Á. (1992). Stability of regression model estimates with respect to subsamples. *Computational Statistics*, *7*, 183–203.

Víšek, J. Á. (1994). A cautionary note on the method of the Least Median of Squares reconsidered. In *Transactions of the twelth prague conference on information theory, statistical decision functions and random processes* (pp. 254–259).

Víšek, J. Á. (1996a). Sensitivity analysis of M-estimates. *Annals of the Institute of Statistical Mathematics*, *48*, 469–495.

Víšek, J. Á. (1996b). On high breakdown point estimation. *Computational Statistics*, *11*, 137–146.

Víšek, J. Á. (1998a). Robust instruments. In J. Antoch, G. Dohnal (Eds.), (*published by Union of Czech Mathematicians and Physicists*), *Robust'98* (pp. 195–224). Prague: MatFyz Press.

Víšek, J. Á. (1998b). Robust specification test. In M. Hušková, P. Lachout, J. Á. Víšek, Union of Czech Mathematicians and Physicists (Eds.), *Proceedings of Prague stochastics'98* (pp. 581–586). Prague: MatFyz Press.

Víšek, J. Á. (2000a). On the diversity of estimates. *Computational Statistics and Data Analysis*, *34*, 67–89.

Víšek, J. Á. (2000b). A new paradigm of point estimation. *Data analysis 2000/II, modern statistical methods—modelling, regression, classification and data mining* (pp. 195–230). Pardubice: Trilobyte.

Víšek, J. Á. (2000c): Regression with high breakdown point. In J. Antoch, G. Dohnal, Union of Czech Mathematicians and Physicists (Eds.), *Robust 2000* (pp. 324–356). Prague: MatFyz Press.

Víšek, J. Á. (2000d). Over- and underfitting the M-estimates. *Bulletin of the Czech Econometric Society*, *7*, 53–83.

Víšek, J. Á. (2000e). Character of the Czech economy in transition. In *Proceedings of the conference "The Czech society on the break of the third millennium"* (pp. 181–205). Karolinum: Publishing House of the Charles University. ISBN 80-7184-825-5.
Víšek, J. Á. (2002a). The least weighted squares I. The asymptotic linearity of normal equations. *Bulletin of the Czech Econometric Society*, *9*, 31–58.
Víšek, J. Á. (2002b). The least weighted squares II. Consistency and asymptotic normality. *Bulletin of the Czech Econometric Society*, *9*, 1–28.
Víšek, J. Á. (2002c). Sensitivity analysis of M-estimates of nonlinear regression model: Influence of data subsets. *Annals of the Institute of Statistical Mathematics, 54*, 261–290.
Víšek, J. Á. (2002d). White test for the least weigthed squares. In S. Klinke, P. Ahrend, L. Richter (Eds.), *COMPSTAT 2002, Proceedings of the conference CompStat 2002—Short communications and poster (CD).* Berlin: Springer.
Víšek, J. Á. (2003a). Durbin-Watson statistic in robust regression. *Probability and Mathematical Statistics*, *23*, 435–483.
Víšek, J. Á. (2003b). Development of the Czech export in nineties. In *Konsolidace vládnutí a podnikání v České republice a v Evropské unii I. Umění vládnout, ekonomika, politika*, 2003 (pp. 193–220). Prague: MatFyz Press. ISBN 80-86732-00-2.
Víšek, J. Á. (2006a). Kolmogorov–Smirnov statistics in multiple regression. In J. Antoch, G. Dohnal (Eds.), *Proceedings of the ROBUST 2006, JČMF and KPMS MFF UK* (pp. 367–374). Prague: MatFyz Press.
Víšek, J. Á. (2006b). The least trimmed squares. Part I—Consistency. Part II—$\sqrt{n}$-consistency. Part III—Asymptotic normality and Bahadur representation. *Kybernetika*, *42*, 1–36, 181–202, 203–224.
Víšek, J. Á. (2006c). Instrumental weighted variables—algorithm. In A. Rizzi, M. Vichi (Eds.), *Proceedings of the COMPSTAT 2006* (pp. 777–786). Heidelberg: Physica-Verlag/Springer.
Víšek, J. Á. (2006d). The least trimmed squares. Sensitivity study. In M. Hušková, M. Janžura (Eds.) *Proceedings of the Prague stochastics 2006* (pp. 728–738). Prague: MatFyz Press.