$\sqrt{n}$-CONSISTENCY OF THE INSTRUMENTAL WEIGHTED VARIABLES [1]

Jan Ámos Víšek

Department of Macroeconomics and Econometrics, Institute of Economic Studies,
Faculty of Social Siences, Charles University
Smetanovo nábřeží 6, 110 01 Praha 1, the Czech Republic, e-mail: visek@mbox.fsv.cuni.cz
&
Department of Stochastic Informatics, Institute of Information Theory and Automation,
Academy of Sciences, the Czech Republic

Abstract. *The definition of Instrumental Weighted Variables (IWV) (which is a robust version of the classical Instrumental Variables) and conditions for the weak consistency as given in Víšek (2009) are recalled. The reasons why the classical Instrumental Variables were introduced as well as the idea of weighting the order statistics of squared residuals (rather than directly the squared residual - firstly employed by the Least Weighted Squares, see Víšek (2000)) are also recalled. Then $\sqrt{n}$-consistency of all solutions of the corresponding normal equations is proved.*

Key words and phrases: Robustness, instrumental variables, weighting the order statistics of squared residuals, $\sqrt{n}$-consistency of estimate by instrumental weighted variables.

INTRODUCTION

The paper continues in studies of Víšek (2009). That it why we recall reasons for introducing the *Instrumental Weighted Variables* as well as for employing the idea of weighting the order statistics of squared residuals (see Víšek (2000)), only briefly. Nevertheless, we will do it in a way to make the paper self-contained.

Let $\mathcal{N}$ denote the set of all positive integers, R the real line and R^p the p-dimensional Euclidean space. All vectors will be assumed to be the column ones and throughout the paper, we assume that all random variables (r.v.'s) are defined on a basic probability space $(\Omega, \mathcal{A}, P)$. For a sequence of $(p+1)$-dimensional r. v.'s $\left\{(X_i', e_i)'\right\}_{i=1}^{\infty}$, any $n \in \mathcal{N}$ and $\beta^0 \in R^p$ the linear regression model given as

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

will be considered. Without loss of generality we may assume that $\beta^0 = 0$ (otherwise we should write in what follows $\beta - \beta^0$ instead of β; nevertheless, we will write fully $\beta - \beta^0$ instead of only β when formulating the asymptotic results because the analogous results are given in such a way, see e. g. Jurečková and Sen (1989), Kalina (2007) or Čížek (2008), among many others). We study the model with intercept, i.e. we assume that the first coordinate of explanatory variables X_i's is degenerated and equal to 1. The following conditions will be considered.

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

C1 *The sequence* $\left\{(X_i', e_i)'\right\}_{i=1}^{\infty}$ *is sequence of independent and identically distributed random variables (i.i.d. r.v.'s) with distribution function* $F_{X,e}(x,v) = F^{(1)}(x_1) \cdot F^{(2)}_{X,e}(x_2, v)$ *where* $F^{(1)}(x_1) : R^1 \to [0,1]$ *is d. f. degenerated at 1 and* $F^{(2)}_{X,e}(x_2, v)$ *is absolutely continuous. Moreover, the density* $f_{e|X}(v|X_1 = x)$ *is uniformly in* x *bounded by* U_e *and* $I\!E\left\{(X_1', e_1)' \cdot (X_1', e_1)\right\}$ *is positive definite matrix.*

$F_X(x)$ and $F_e(v)$ ($f_X(x)$ and $f_e(v)$) will stay for the marginals of $F^{(2)}_{X,e}(x_2, v)$ (and their densities, respectively). Finally, notice please that $f_e(v) = I\!E_x f_e(v|X_1 = x) \le I\!E_x U_e = U_e$.

The form of the conditions **C1** does not allow for dummy variable(s). However, it is clear (from what follows) that it is only matter of technicalities. On the other hand, accommodating **C1**, to allow the inclusion of dummies, would bring (unnecessarily) complicated notations.

ESTIMATING BY MEANS OF INSTRUMENTAL VARIABLES

The estimator of the regression coefficients β^0 of the model (1) which is probably the most frequently used, is the *(Ordinary) Least Squares* $\hat\beta^{(OLS,n)}$. On the other hand, due to the fact that

$$\hat\beta^{(OLS,n)} = \beta^0 + \left(\frac{1}{n}\sum_{k=1}^{n} X_k X_k'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i e_i \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i e_i = I\!E X_1 e_1 \quad \text{in probability,} \tag{2}$$

one easy verifies that the violation of orthogonality condition, i. e. when $I\!E\{e_i|X_i\} \neq 0$, implies bias and inconsistency of the *(Ordinary) Least Squares* (where, due to the almost sure convergence of $\frac{1}{n}\sum_{k=1}^{n} X_k X_k'$ to $I\!E(X_1 \cdot X_1')$, there is a set A, $P(A) = 1$ such that for any $\omega \in A$, there is $n_\omega \in \mathcal{N}$ so that for any $n > n_\omega$ we have $\frac{1}{n}\sum_{k=1}^{n} X_k X_k'$ positive definite. Moreover, (asymptotic) normality of the estimator is also violated.

For one of the best known example of the situation when the orthogonality condition fails - namely measurement of explanatory variables with a random error - see Judge et al. (1985) or Víšek (1998), and it was discussed in details in Víšek (2009). We are going to recall another famous example justifying employment of the method of instrumental variables, also given in Judge et al. (1985). So let us consider (with a bit of freedom from the rigor) the model with lagged explanatory variables. Assume the simplest one, with the geometric structure of coefficients, i. e.

$$Y_t = \gamma \sum_{j=1}^{\infty} \lambda^{j-1} x_{t-j+1} + e_t, \quad t = ..., -1, 0, 1, 2, ..., T \tag{3}$$

with a sequence of i.i.d. disturbances $\{e_t\}_{t=-\infty}^{\infty}$, $I\!E e_t = 0$ and $I\!E e_t^2 = \sigma^2 \in (0, \infty)$. Clearly, we are not able to estimate directly coefficients γ and λ, so writing model for $t-1$

$$Y_{t-1} = \gamma \sum_{j=1}^{\infty} \lambda^{j-1} x_{t-j} + e_{t-1},$$

multiplying it by λ and subtracting from (3), we obtain

$$Y_t = \lambda Y_{t-1} + \gamma x_t + e_t - \lambda e_{t-1} = \lambda Y_{t-1} + \gamma x_t + u_t. \tag{4}$$

Now, the "explanatory" variable Y_{t-1} is correlated with the error term u_t and then (2) indicates that $\hat{\beta}^{(OLS,n)}$ is inconsistent. Although just considered model does not fulfill **C1** (as the sequence $\{Y_{t-1}, x_t, u_t\}_{t=2}^{\infty}$ is not i.i.d.), it quite well illustrate that the situation when the orthogonality condition fails is not an academic one.

The classical econometrics solve such situations *usually* by means of the *Method of Instrumental Variables.*

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

$$\sum_{i=1}^{n} Z_i \left(Y_i - X_i' \beta \right) = 0 \tag{5}$$

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

Remark 1 *The elements of the sequence* $\{Z_i\}_{i=1}^{\infty}$ *are usually called instruments. Without loss of generality we may assume that* $Z_{i1} = 1$ *and* $I\!EZ_{ij} = 0, j = 2, 3, ..., p$ *and* $i = 1, 2, ...$ *We do not lose generality firstly, due to the fact that* $Z_{i1} = 1$ *represents constants and hence they cannot be correlated with disturbances (in fact we have then* $Z_{i1} = X_{i1}$*). Secondly, what concerns the assumption that* $I\!EZ_{ij} = 0, j = 2, 3, ..., p$*, if it would not be fulfilled, i. e. when* $I\!EZ_{ij} \neq 0$*, we can consider* $\tilde{Z}_{ij} = Z_{ij} - I\!EZ_{ij}$ *and change appropriately the intercept of the original model (1).*

The method became at the end of the last century more or less a standard tool in many case studies of panel data since the correlation of explanatory variables and disturbances frequently appeared, see Bowden and Turkington (1984), Judge et al. (1985), Manski and Pepper (2000), Stock and Trebbi (2003), to give some among many others. There are many papers exploring the best way of the selecting the instruments for explanatory variables which established useful, easy implemented results, see e.g. Arellano and Bond (1991), Arellano and Bover (1995) or Sargan (1988) (and for examples of implementation see for SAS - Der and Everitt (2002), for R and S-PLUS - Fox, J. (2002)).

There is also possibility to cope with the break of orthogonality condition by another approach, namely by the *Total Least Squares*, see e.g. Nievergelt (1994) or Van Huffel (2004) or Paige and Strakoš (2002). Advantage of this method is that we do not need to select instruments. On the other hand, the method need not have any solution or it can have several solutions. Moreover, the evaluation is much more complicated.

Last but not least, the interpretation of regression model is quite different from the point of view of economic applications (where the *disturbances* are assumed to include generally some (part of) explanatory variables, e. g. when the model (1) includes proxies to approximate explanatory variables which we cannot measure - *education*↔*schooling*) while in exact (especially natural) sciences the *error term* is (more or less) considered to be an error of measurement of response variable. As the philosophy behind the *Total Least Squares* corresponds with the interpretation of regression model in exact sciences, they are applied there, while the method of *Instrumental Variables* is employed mostly in the social sciences. On the other hand, in the social sciences we have at our disposal frequently "natural" *instruments*, e. g. lagged explanatory variables.

Let us enlarge a bit the notations. Let us denote for any $\beta \in R^p$ by $r_i(\beta) = Y_i - X_i'\beta$ the i-th residual and by $r^2_{(h)}(\beta)$ the h-th order statistic among the squared residuals. To be more explicit, we have

$$r^2_{(1)}(\beta) \le r^2_{(2)}(\beta) \le \ \dots \ \le r^2_{(n)}(\beta). \tag{6}$$

Then the *Least Weighted Squares* can be defined as follows (see Víšek (2000), see also (2002b, c))

$$\hat\beta^{(LWS,n,w)} = \mathop{\arg\min}_{\beta \in R^p} \sum_{i=1}^n w_i r^2_{(i)}(\beta) \tag{7}$$

where $w_i, i = 1, 2, ..., n$ are weights[2]. They are usually generated by a weight function with the following properties[3]:

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

Then put $w_i = w\left(\frac{i-1}{n}\right)$. Following Hájek and Šidák (1967) for any $i \in \{1, 2, ..., n\}$ let us denote by $\pi(\beta, i)$ the rank of the i-th squared residual. It means that $\pi(\beta, i) = j \in \{1, 2, ..., n\}$ iff $r^2_i(\beta) = r^2_{(j)}(\beta)$ (notice that $\pi(\beta, i)$ is r.v.). Then we have

$$\hat\beta^{(LWS,n,w)} = \mathop{\arg\min}_{\beta \in R^p} \sum_{i=1}^n w\left(\frac{\pi(\beta,i) - 1}{n}\right) r^2_i(\beta). \tag{8}$$

It is straightforward to show that the *Least Weighted Squares* are solution of *normal equations*

$$I\!N E_{X,n}(\beta) = \sum_{i=1}^n w\left(\frac{\pi(\beta,i) - 1}{n}\right) X_i \left(Y_i - X_i'\beta\right) = 0, \tag{9}$$

see Víšek (2009).

At the end of this paragraph let us recall that the *Least Weighted Squares* are the robust version of the *Ordinary Least Squares* which removes the sensitivity of OLS to influential observations. By flexibility of *weight function* it enable us to adapt the *level* and *character* of robustness of the method to the *level* and to the *character* of contamination of data, see Mašíček (2004). In this way we can obtain as the special cases of LWS the estimators as the *Least Median of Squares* or the *Least Trimmed Squares*, see Hampel et al. (1986) or Rousseeuw and Leroy (1987).

The sensitivity of OLS to the influential observations follows from the shape of normal equations

$$\sum_{i=1}^n X_i \left(Y_i - X_i'\beta\right) = 0. \tag{10}$$

High sensitivity to the outliers is due to the presence of residuals $r_i(\beta) = Y_i - X_i'\beta$ in (10) and similarly, sensitivity to the leverage points is implied by X_i's in (10). As the shape of the normal equations (5) is the same as the shape of normal equations (10), it is clear that the estimator by means of the *Instrumental Variables* is sensitive to influential points in an analogous way as OLS are.

[2]See also Čížek (2002) where the estimator is called the *Smoothed Least Trimmed Squares.*

[3]Compare Hájek and Šidák (1967).

On the other hand, the inconsistency of the *Ordinary Least Squares* in the case of failure of the orthogonality condition (as we recalled it 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 *Instrumental Variables* but which will weight down the residuals of those observations which seem to be atypical. For complex discussion of the situations when some observations are or seem to be clearly or more or less atypical see Hampel et al. (1986) or Rousseeuw and Leroy (1987).

INSTRUMENTAL WEIGHTED VARIABLES

Definition 2 *For any sequence of p-dimensional random vectors $\{Z_i\}_{i=1}^{\infty}$ the solution(s) of the (vector) equation*

$$I\!N E_{Z,n}(\beta) = \sum_{i=1}^{n} w\left(\frac{\pi(\beta,i)-1}{n}\right) Z_i \left(Y_i - X_i'\beta\right) = 0 \tag{11}$$

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

The form of normal equations (11) is not suitable for proving consistency and $\sqrt{n}$-consistency. So, let's make the following consideration.

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

$$F_{\beta}^{(n)}(v) = \frac{1}{n}\sum_{j=1}^{n} I\{|r_j(\beta)| < v\} = \frac{1}{n}\sum_{j=1}^{n} I\left\{|e_j - X_j'\beta| < v\right\}$$
$$= \frac{1}{n}\sum_{j=1}^{n} I\left\{\omega \in \Omega \,:\, |e_j(\omega) - X_j'(\omega)\beta| < v\right\}. \tag{12}$$

It is straightforward that then (for details see Víšek (2009))

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

and so (11) can be written as

$$\sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right) Z_i \left(Y_i - X_i'\beta\right) = 0. \tag{13}$$

CONSISTENCY OF THE INSTRUMENTAL WEIGHTED VARIABLES

We will need also the following notation. For any $\beta \in R^p$ the distribution of the product $\beta' Z_1 X_1' \beta$ will be denoted $F_{\beta' Z X' \beta}(u)$, i. e.

$$F_{\beta' Z X' \beta}(u) = P(\beta' Z_1 X_1' \beta < u) \tag{14}$$

and similarly as in previous, the corresponding empirical distribution will be denoted $F_{\beta' Z X' \beta}^{(n)}(u)$, so that

$$F_{\beta' Z X' \beta}^{(n)}(u) = \frac{1}{n}\sum_{j=1}^{n} I\left\{\beta' Z_j X_j' \beta < u\right\} = \frac{1}{n}\sum_{j=1}^{n} I\left\{\omega \in \Omega \,:\, \beta' Z_j X_j' \beta < u\right\}. \tag{15}$$

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

$$\gamma_{\zeta,a} = \sup_{\|\beta\|=\zeta} F_{\beta' Z X' \beta}(a). \tag{16}$$

Due to the fact that the surface of the ball $\{\beta \in R^p, \|\beta\| = \zeta\}$ is compact, there is $\beta_\gamma \in \{\beta \in R^p, \|\beta\| = \zeta\}$ so that

$$\gamma_{\zeta,a} = F_{\beta_\gamma' Z X \beta_\gamma}(a). \tag{17}$$

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

$$\tau_\zeta = - \inf_{\|\beta\| \leq \zeta} \beta' I\!\!E \left[Z_1 X_1' \cdot I\{\beta' Z_1 X_1' \beta < 0\} \right] \beta. \tag{18}$$

Notice please that $\tau_\zeta \geq 0$ (due to the presence of $I\{\beta' Z_1 X_1' \beta < 0\}$ in (18)) and that again due to the fact that the ball $\{\beta \in R^p, \|\beta\| \leq \zeta\}$ is compact, the infimum is finite, and hence there is a $\tilde{\beta} \in \{\beta \in R^p, \|\beta\| \leq \zeta\}$ so that

$$\tau_\zeta = -\tilde{\beta}' I\!\!E \left[Z_1 X_1' \cdot I\{\tilde{\beta}' Z_1 X_1' \tilde{\beta} < 0\} \right] \tilde{\beta}. \tag{19}$$

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, decomposing the joint distribution function $F_{X,Z}(x,z) = F^{(1)}(x_1, z_1) \cdot F^{(2)}(x_2, z_2)$ with $F^{(1)}(x_1, z_1) : R^2 \to [0,1]$ and $F^{(2)}(x_2, z_2) : R^{2(p-1)} \to [0,1]$, we have $F^{(2)}(x_2, z_2)$ absolutely continuous, $I\!\!E \left\{ w(F_{\beta^0}(|e_1|)) Z_1 X_1' \right\}$ as well as $I\!\!E Z_1 Z_1'$ are positive definite (one can compare C3 with Víšek (1998) where we considered instrumental M-estimators and the discussion of assumptions for M-instrumental variables was given) and there is $q > 1$ so that $I\!\!E \left\{ \|Z_1\| \cdot \|X_1\| \right\}^q < \infty$. Finally, there is $a > 0$, $b \in (0,1)$ and $\lambda > 0$ so that*

$$a \cdot (b - \gamma_{\lambda,a}) \cdot w(b) > \tau_\lambda \tag{20}$$

with $\gamma_{\lambda,a}$ and τ_λ given by (17) and (18).

Remark 2 *Let us briefly discuss assumptions we have made. Let us recall that the Ordinary Least Squares $\hat{\beta}^{(OLS,n)}$ are optimal only under normality of disturbances. Here the optimality means that the variance of $\hat{\beta}^{(OLS,n)}$ reaches the lower Rao-Cramer bound (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 a large decrease of efficiency (see e.g. Fisher (1920), (1922)). Without the assumption of normality of disturbances $\hat{\beta}^{(OLS,n)}$ is 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 we may meet with justification of the restriction on the class of linear unbiased estimators by asserting that the linear estimators are scale- and regression-equivariant*[4].

[4] 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

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

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

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

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)).

As the Least Weighted Squares can cope with contamination of data better than the Ordinary Least Squares, we may guess that they are approximately optimal under the approximative normality of disturbances, for some hint consult Mašíček (2003). As the present proposal of robustified instrumental variables is based on the same idea of weighting the order statistics of squared residuals as the Least Weighted Squares, we can expect that the estimate can be approximately optimal under approximative normality of disturbances. But then our assumptions seem to be quite acceptable.

Nevertheless, the assumption which deserve further discussion is the assumption (20). We are going to show that it is a restriction on the weight function w. Let us return to (17). 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

$$\lim_{a \to 0_+} F_{\beta_\gamma' X X' \beta_\gamma}(a) = 0 \qquad (21)$$

but for $\gamma_{\lambda,a}$ we have (again for fix $\lambda \in R^+$)

$$\lim_{a \to 0_+} F_{\beta_\gamma' Z X' \beta_\gamma}(a) = P\left(\beta_\lambda' Z_1 X_1' \beta_\lambda \le 0\right). \qquad (22)$$

So, we can have $\gamma_{\lambda,a} > 0$. On the other hand, for any $a > 0$ we have

$$\gamma_{\lambda,a} < 1. \qquad (23)$$

Now let us turn to τ_λ. As

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

we have

$$\limsup_{\|\beta\| \to 0} \left| \, \beta' I\!E\left[Z_1 X_1' I\{\beta' Z_1 X_1' \beta < 0\}\right] \beta\right| = 0. \qquad (24)$$

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 (21), (22), (23) and (24) indicate that (20) can be always fulfilled but we may have slightly restricted possibility to depress the influence of influential observations.

For consistency of $\hat\beta^{(LWS,n,w)}$ we need an *identification condition.* To be able to give it, we need: Let's denote by $F_\beta(v)$ the distribution of the absolute value of residual, i. e.

$$F_\beta(v) = P(|Y_1 - X_1'\beta| < v) = P\left(\left|e_1 - X_1'\left(\beta - \beta^0\right)\right| < v\right). \qquad (25)$$

---

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

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

C4 *The vector equation*

$$\beta' I\!\!E\left[w\left(F_\beta(|r_1(\beta)|)\right)Z_1\left(e_1 - X_1'\beta\right)\right] = 0 \tag{26}$$

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

Lemma 1 *Let the conditions* **C1**, **C2**, **C3** *and* **C4** *be fulfilled. Then any sequence $\left\{\hat\beta^{(IWV,n,w)}\right\}_{n=1}^{\infty}$ of the solutions of normal equations $I\!\!NE_{Z,n}(\hat\beta^{(IWV,n,w)}) = 0$ (see (11)) is weakly consistent.*

For the proof see Víšek (2009).

$\sqrt{n}$-CONSISTENCY OF THE INSTRUMENTAL WEIGHTED VARIABLES

We will need to enlarge the previous conditions.

NC1 *The density $f_{e|X}(r|X_1 = x)$ is uniformly with respect to x Lipschitz of the first order (with the corresponding constant equal to B_e). Moreover, $f_e'(r)$ exists and is bounded in absolute value by U_e'.*

NC2 *The derivative $w'(\alpha)$ of the weight function is Lipschitz of the first order (with the corresponding constant J_w).*

Lemma 2 *Let the conditions* **C1**, **C2**, **C3**, **C4**, **NC1** *and* **NC2** *be fulfilled. Then any sequence $\left\{\hat\beta^{(IWV,n,w)}\right\}_{n=1}^{\infty}$ of the solutions of normal equations (11) (or (13)) $I\!\!NE_{Z,n}(\hat\beta^{(IWV,n,w)}) = 0$ is $\sqrt{n}$-consistent.*

Proof: We have to prove that

$$\forall(\varepsilon > 0) \quad \exists(n_\varepsilon \in \mathcal{N}, K_\varepsilon < \infty) \quad \forall(n > n_\varepsilon) \qquad P\left(\left\{\omega \in \Omega \ : \ \sqrt{n}\left\|\hat\beta^{(IWV,n,w)} - \beta^0\right\| > K_\varepsilon\right\}\right) < \varepsilon.$$

We are going to give only a sketch of proof (to meet with space restriction). Moreover, the proof is a chain of nearly routine approximations employing standard tools of probability theory. That is why we give in details only a key step of proof[5].

Let us recall that $\hat\beta^{(IWV,n,w)}$ is given as solution of (13), i. e. as solution of the equation

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

Rewriting it, we obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) Z_i e_i = \frac{1}{n}\sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) Z_i X_i' \cdot \sqrt{n}\left(\beta - \beta^0\right). \tag{27}$$

Since $-L \le w'(v) \le 0$ (and recalling that we have denoted by $F_\beta(v)$ the distribution of the absolute value of residual), we have

$$\sup_{\beta\in R^p}\left|w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) - w\left(F_\beta(|r_i(\beta)|)\right)\right| \le L_w \cdot \sup_{v\in R^+}\ \sup_{\beta\in R^p}\left|F_\beta^{(n)}(v) - F_\beta(v)\right|.$$

[5]The paper with the proof containing all details in length is available on request from the present author.

Further

$$\frac{1}{\sqrt{n}} \sup_{\beta\in R^p} \left\| \sum_{i=1}^{n} \left[w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) - w\left(F_\beta(|r_i(\beta)|)\right)\right] Z_i e_i \right\|$$

$$\leq \frac{1}{\sqrt{n}} \sup_{\beta\in R^p} \sum_{i=1}^{n} \left| w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) - w\left(F_\beta(|r_i(\beta)|)\right)\right| \cdot \|Z_i\| \cdot |e_i|$$

$$\leq \sqrt{n} \cdot L_w \cdot \sup_{v\in R^+} \sup_{\beta\in R^p} \left| F_\beta^{(n)}(v) - F_\beta(v)\right| \cdot \frac{1}{n}\sum_{i=1}^{n} \|Z_i\| \cdot |e_i| \, .$$

As according to Lemma A.1 $\sqrt{n} \cdot \sup_{v\in R^+} \sup_{\beta\in R^p} \left| F_\beta^{(n)}(v) - F_\beta(v)\right|$ is $\mathcal{O}_p(1)$, we have

$$\frac{1}{\sqrt{n}} \sup_{\beta\in R^p} \left\| \sum_{i=1}^{n} \left[w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) - w\left(F_\beta(|r_i(\beta)|)\right)\right] Z_i e_i \right\| = \mathcal{O}_p(1)$$

as $n \to \infty$. Hence, denoting $X = (X_1, X_2, ..., X_n)'$, $Z = (Z_1, Z_2, ..., Z_n)'$ and $e = (e_1, e_2, ..., e_n)'$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) Z_i e_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(F_\beta(|r_i(\beta)|)\right) Z_i e_i + R_n^{(1)}(\beta, X, Z, e) \tag{28}$$

where

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

and $\mathcal{O}_p(1)$ is to be understood in the sense that

$$\forall(\varepsilon > 0) \;\; \exists(K_\varepsilon < \infty) \quad \inf_{n\in\mathcal{N}} P\left(\left\{\omega \in \Omega \; : \; \sup_{\beta\in R^p} \left\| R_n^{(1)}(\beta, X, Z, e)\right\| < K_\varepsilon\right\}\right) > 1 - \varepsilon. \tag{29}$$

(28) allows to substitute the left hand side in (27) by the right hand side of (28). Elaborating the same for the right hand side of (27), we arrive at

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

$$= \frac{1}{n} \sum_{i=1}^{n} \left[w\left(F_\beta(|r_i(\beta)|)\right) Z_i X_i' + R_n^{(2)}(\beta, X, Z, e)\right] \cdot \sqrt{n}\left(\beta - \beta^0\right). \tag{30}$$

The rest of proof is a (long) chain of approximations and estimations of upper boundaries of various expressions, employing tools of classical mathematical analysis (as simple as Taylor expansion, e. g.) and the means of mathematical statistics as laws of large numbers and the central limits theorem.

CONCLUDING REMARKS

Lemma 2 shows that $\hat{\beta}^{(IWV,n,w)}$ converges to β^0 in the rate $\frac{1}{\sqrt{n}}$. It means that, e. g., the length of the reliability intervals decreases as $\frac{1}{\sqrt{n}}$ with the increasing number of observations. As this is - under some natural regular conditions - the largest rate of convergence of estimators to the respective "true" value of parameters, Lemma 2 confirms that our estimator employs the information contained in data in an efficient way.

Appendix

Lemma A.1 *Let the conditions* **C1** *hold and fix arbitrary* $\varepsilon > 0$. *Then there is a constant* $K < \infty$ *and* $n_\varepsilon \in \mathcal{N}$ *so that for all* $n > n_\varepsilon$

$$P\left(\left\{\omega \in \Omega : \sup_{v \in R^+} \sup_{\beta \in R^p} \sqrt{n}\left|F_\beta^{(n)}(v) - F_\beta(v)\right| < K\right\}\right) > 1 - \varepsilon. \tag{A.31}$$

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

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

$$\lim_{n \to \infty} v_{ij}^{(n)} = q_{ij} \qquad \textit{in probability} \tag{A.32}$$

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

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

Then

$$\exists\ (\ \delta > 0) \quad \forall\ (H > 0)$$

so that

$$\limsup_{n \to \infty}\ P\left(\left\|\mathcal{V}^{(n)}\theta^{(n)}\right\| > H\right) > \delta.$$

Proof: Due to (A.32) the matrix $\mathcal{V}^{(n)}$ is regular in probability. Let then $0 < \lambda_{1n} < \lambda_{2n} < ... < \lambda_{pn}$ and $z_{1n}, z_{2n}, ..., z_{pn}$ be eigenvalues and corresponding eigenvectors (selected to be mutually orthogonal) of the matrix $[\mathcal{V}^{(n)}]^T\mathcal{V}^{(n)}$. Let us write $\theta^{(n)} = \sum_{j=1}^{p} a_{jn} z_{jn}$ (for an appropriate vector $a_n = (a_{1n}, a_{1n}, ..., a_{pn})^T$). Then we have

$$\left\|\mathcal{V}^{(n)}\theta^{(n)}\right\|^2 = \sum_{j=1}^{p} [a_{jn}]^2 \lambda_{jn} \|z_{jn}\|^2 \leq \lambda_{1n}\|\theta^{(n)}\|. \tag{A.33}$$

Moreover, denoting λ_1 the smallest eigenvalue of the matrix Q^TQ, we have $\lambda_{1n} \to \lambda_1$ in probability as $n \to \infty$. The assertion of the lemma then follows from (A.33), see also Víšek (1996) or (2002a). □

References

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

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

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

Čížek, P. (2008): Robust and efficient adaptive estimation of binary-choice regression models. *Journal of the American Statistical Association, 103(482), 687-696.*

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

Fox, J. (2002): *An R and S-PLUS Companion to Applied Regression.* Thousand Oaks: SAGE Publications.

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

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986): *Robust Statistics – The Approach Based on Influence Functions.* New York: J.Wiley *&* Sons.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, T. C. Lee (1985): *The Theory and Practice of Econometrics.* New York: J.Wiley *&* Sons(second edition).

Jurečková, J., P. K. Sen (1989): Uniform second order asymptotic linearity of M-statistics in linear models. *Statistics and Decisions 7, 263 - 276.*

Kalina, J. (2007): Autocorrelated errors of the least weighted squares. *Forum Statisticum Slovakum 3, Nr. 2, 136 - 141.*

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

Mašíček, L. (2004): Optimality of the least weighted squares estimator. *Kybernetika, vol. 40, n.6, 715-734.*

Nievergelt, Y. (1994): Total least squares - State-of-the-art regression in numerical analysis. *SIAM Review 36, 258 - 264.*

Paige, C. C., Strakoš, Z. (2002): Scaled total least squares fundamentals. *Numerische Mathematik 91, 117 - 146.*

Rousseeuw, P. J., A. M. Leroy (1987): *Robust Regression and Outlier Detection.* New York: J.Wiley *&* Sons.

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. *Proceedings in Computational Statistics, COMPSTAT 2004, ed. J. Antoch, Heidelberg: Physica-Verlag/Springer, 539 - 555.*

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

Víšek, J. Á. (1998): Robust instruments. *Robust'98 (eds. Jaromír Antoch & Gejza Dohnal, published by Union of Czech Mathematicians and Physicists), Prague: matfyzpress 195 - 224.*

Víšek, J. Á. (2000): Regression with high breakdown point. *Robust 2000 (eds. Jaromír Antoch & Gejza Dohnal, published by Union of Czech Mathematicians and Physicists), Prague: matfyzpress 324 - 356.*

Víšek, J. Á. (2002a): Sensitivity analysis of M-estimates of nonlinear regression model: Influence of data subsets. *Annals of the Institute of Statistical Mathematics, Vol. 54, No.2, 261 - 290, 2002.*

Víšek, J. Á. (2002b): The least weighted squares I. The asymptotic linearity of normal equations. *Bulletin of the Czech Econometric Society, Vol. 9, no. 15, 31 - 58.*

Víšek, J. Á. (2002c):The least weighted squares II. Consistency and asymptotic normality. *Bulletin of the Czech Econometric Society, Vol. 9, no. 16, 1 - 28.*

Víšek, J. Á. (2006): Kolmogorov-Smirnov statistics in multiple regression. *Proceedings of the ROBUST 2006, organized by CHarles University, Dept. of Probab. and Math. Statist. and the Union of the Czech Mathematicians and Physicists, eds. Jaromír Antoch & Gejza Dohnal, Prague: matfyzpress, 367-374.*

Víšek, J. Á. (2009): Consistency of the instrumental weighted variables. To appear in *Annals of the Institute of Statistical Mathematics, Vol.61, No.3 (September, 2009).*