CONSISTENCY OF INSTRUMENTAL WEIGHTED VARIABLES

J. Á. Víšek[1]

[1] Inst. of Economic Studies, Fac. of Social Sciences, Charles University, Smetanovo nábřeží 6, 110 01 Prague, the Czech republic

Keywords: Robust regression, break of orthogonality condition, instrumental variables, implicitly weighted residuals, reliability of algorithm

1 Recalling the Least Weighted Squares

Considering the linear regression model

$$Y_t = X_t'\beta^0 + e_t = \sum_{j=1}^{p} X_{tj}\beta_j^0 + e_t, \quad t = 1, 2, ..., T,$$

denote for any $\beta \in R^p$ the t-th residual by $r_t(\beta) = Y_t - X_t'\beta$ and by $r^2_{(h)}(\beta)$ the h-th order statistic among the squared residuals. For a weight function $w : [0,1] \to [0,1]$ define the *Least Weighted Squares* (LWS) (Víšek (2000 b))

$$\hat{\beta}^{(LWS,T,w)} = \arg\min_{\beta \in R^p} \sum_{t=1}^{T} w\left(\frac{t-1}{n}\right) r^2_{(t)}(\beta). \tag{1}$$

(Notice please that the weights are assigned to the order statistics, not to observations, i. e. in an implicit way, hence the order of words in the name of estimator.) Having denoted the empirical distribution function of the absolute values of residuals as

$$F_\beta^{(T)}(r) = \frac{1}{T}\sum_{t=1}^{T} I\left\{|r_t(\beta)| < r\right\} = \frac{1}{T}\sum_{t=1}^{n} I\left\{|e_t - X_t'\beta| < r\right\},$$

we can show that $\hat{\beta}^{(LWS,T,w)}$ is one of solution of the *normal equations*

$$\sum_{t=1}^{T} w\left(F_\beta^{(T)}(|r_t(\beta)|)\right) X_t\left(Y_t - X_t'\beta\right) = 0. \tag{2}$$

All solutions of (2) fulfill usual requirements on robust point estimators (see Hampel et al. (1986)) as the *consistency* (see Mašíček (2003)) and the *asymptotic normality*, they have *controllable break-down point* and *subsample sensitivity* (in contrast to M-estimators, see Víšek (1996 a), (2002)). They preserve the *scale-* and *regression-equivariance* of the *Ordinary Least Squares* (again contrary to M-estimators, see Bickel (1975) or Jurečková and Sen (1993)). Last but not least, there is reliable, implemented algorithm for evaluating a tight approximation to the solution of extremal problem (1) (the algorithm is similar to that one tested for the *Least Trimmed Squares* in Víšek (1996 b), (2000 a)).

2 Reasons for instrumental variable

When the orthogonality condition $I\!E\left\{e_t|X_t\right\} = 0$ is broken, the ordinary least squares are not consistent. Well-known example of the situation when the orthogonality condition fails, is the model assuming that the explanatory variables are measured with random error or the model with lagged response (and/or explanatory) variable, see e. g. Judge et al. (1985). The problem is usually solved by considering, for a sequence of *Iinstrumental variable* $\{Z_t\}_{t=1}^{\infty}$, the solution of the *normal equations*

$$\sum_{i=1}^{T} Z_t\left(Y_t - X_t'\beta\right) = 0.$$

In nineties the *Method of Instrumental Variable* became a standard tool in many case studies of dynamic regression model since the correlation of explanatory variables and disturbances frequently appeared. Moreover, many papers considering possibilities how to select the instruments for explanatory variables brought applicable results (including also easy available implementations), see e.g. Arellano, Bond (1991), Arellano, Bover (1995) or Sargan (1988).

3 Instrumental weighted variable

Of course, similarly as the *Ordinary Least Squares*, the classical *Instrumental Variables* are vulnerable to the influential points. In analogy with (2) we can define the *Instrumental Weighted Variables* estimator by

$$\sum_{t=1}^{T} w\left(F_{\beta}^{(T)}(|r_t(\beta)|)\right) Z_t\,(Y_t - X_t'\beta) = 0. \tag{3}$$

Let us consider assumptions (compare with Víšek (1998)):

C1 *The sequence* $\{(X_t', e_t)'\}_{t=1}^{\infty} \subset R^{p+1}$ *is sequence of independent and identically distributed random variables with absolutely continuous d. f.* $F_{X,e}(x,r)$*. Moreover, the existence of second moments of* (X,e) *is assumed and the density* $f_{e|X}(r|X=x)$ *is uniformly in* x *bounded.*

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

C3 *The instrumental variables* $\{Z_t\}_{t=1}^{\infty} \subset R^p$ *are independent and identically distributed with d. f.* $F_Z(z)$*. Moreover, they are independent from the sequence* $\{e_t\}_{t=1}^{\infty}$*. Finally,* $\mathbb{E}\left\{w(F_{\beta^0}(|e|))Z_1X_1'\right\}$ *as well as* $\mathbb{E}Z_1Z_1'$ *are positive definite,* $\mathbb{E}\,\|Z_1\|\cdot\|X_1\|^2 < \infty$ *and for any* $\beta \in R^p$ *and any* $v \in R$

$$\beta'\left\{\int [w(P(-r + x'\beta < e_1 < r + x'\beta)) - w(P(-r < e_1 < r))]\, zx'\mathrm{d}F(x,z,r)\right\}\beta \geq 0.$$

Then any sequence $\left\{\hat{\beta}^{(IWV,T,w)}\right\}_{T=1}^{\infty}$ of the solutions of the *normal equations* (3) is *weakly consistent.*

Numerical examples of the Instrumental Weighted Variables will be also presented.

References

Arellano, M., S. Bond (1991): Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. of Econ. Studies (1991) 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.*

Bickel, P. J. (1975): One-step Huber estimates in the linear model. *JASA 70, 428–433.*

Bowden, R. J., D. A. Turkington (1984): *Instrumental Variables.* Cambridge Univ. 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 (1993): Regression rank scores scale statistics and studentization in linear models. *Proc. of the 5th Prague Symposium on Asymptotic Statistics, Physica Verlag, 111-121.*

Mašíček, L. (2003): Consistency of the least weighted squares estimator. *Statistics for Industry and Technology, 2004 Birkhǻser Verlag Basel, 183 - 194.*

Sargan, J.D. (1988): Testing for misspecification after estimating using instrumental variables. In *Contribution to Econometrics: J. D. Sargan, Vol. 1. Cambridge Univ.*

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

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

Víšek, J. Á. (1998): Robust instruments. *Robust'98 (eds. J. Antoch & G. Dohnal, Union of Czechoslovak Mathematicians and Physicists), 1998, 195 - 224.*

Víšek, J. Á. (2000 a): On the diversity of estimates. *CSDA 34, (2000), 67 - 89.*

Víšek, J. Á. (2000 b): Regression with high breakdown point. *Robust 2000 (eds. J. Antoch & G. Dohnal, Union of Czechoslovak Mathematicians and Physicists), 2001, 324 - 356.*

Víšek, J. Á. (2002): Sensitivity analysis of M-estimates of nonlinear regression model: Influence of data subsets. *Annals of the Institute of Statistical Mathematics, 54 (2002), 261 - 290.*