White estimator of covariance matrix for instrumental weighted variables

Proceedings of the COMPSTAT 2008, 355 - 362, ISBN 978-3-7908-2083-6, Physica-Verlag, Springer Company, ed. Paula Brito, Heidelberg 2008.

Jan Ámos Víšek[1]

Faculty of Social Sciences, Charles University
& Inst. of Information Theory and Automation, Academy of Sciences,
Smetanovo nábřeží 6, 110 01 Prague, the Czech Republic *visek@mbox.fsv.cuni.cz*

Abstract. Under heteroscedasticity of disturbances the significances of explanatory variables in a linear regression model have to be established employing the *White estimator of covariance matrix* of the *(Ordinary) Least Squares* estimator of regression coefficients. When the orthogonality condition is broken the *Instrumental Variables* (in econometrics, sociology, etc.) or the *Total Least Squares* (in natural sciences) are used to preserve unbiasedness of estimation. If moreover, data are contaminated a robust version of instrumental variables called the *Instrumental Weighted Variables* is to be used to cope both with the break of orthogonality condition as well as with contamination. Significance of explanatory variables (and of instruments) is to be examined by a robust version of White estimator of covariance matrix.

Keywords: Robustness, heteroscedasticity, Instrumental Weighted Variables, White estimator

1 Introduction of basic framework

The set of all positive integers will be denoted by N and p-dimensional Euclidean space by R^p.Let us consider the linear regression model

$$Y_i = X_i'\beta^0 + e_i, \quad i = 1, 2, ..., n. \tag{1}$$

We shall assume that:

C1 *The sequence $\{(V_i', e_i)'\}_{i=1}^{\infty}$ is sequence of independent p-dimensional random variables. There is an absolutely continuous d.f., say $F_{V,e}(v,r)$ (denote density $f_{V,e}(v,r)$), so that the d.f.'s $F_{V,e_i}(v,r) = F_{V,e}(v, \sigma_i \cdot r)$ and $\mathbb{E}e_i = 0$ for all $i \in N$. The marginal d.f.'s $F_V(v)$ of vectors V_i's are the same for all $i \in N$ and have a bounded support, i.e. putting $M = \sup\{\|v\| \;:\; f_V(v) > 0\}$ we have $M < \infty$. Moreover, the existence of second moments is assumed, the density $f_{V,e}(v,r)$ is bounded, say by B, and $\sup_{i \in N} \sigma_i < \infty$. Finally, consider the sequence $\{(X_i', e_i)'\}_{i=1}^{\infty}$ where $X_{i1} = 1$ and $X_{ij} = V_{i,j-1}$, $j = 2, 3, ..., p-1$ for all $i \in N$.*

Notice please that we assume that the error terms e_i's can be correlated with explanatory variables V_i's. Moreover, error terms are assumed generally heteroscedastic. Finally, as $f_{V,e_i}(v,r) = \sigma_i \cdot f_{v,e}(v, \sigma_i \cdot r)$, we have $f_{V,e_i}(v,r) < \sup_{i \in N} \sigma_i \cdot B$. For any $\beta \in R^p$ $r_i(\beta) = Y_i - X_i'\beta$ denotes the i-th residual and $r^2_{(h)}(\beta)$ the h-th order statistic among the squared residuals, i.e. we have

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

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

2 Why Instrumental Weighted Variables?

The violation of orthogonality condition $I\!E\{e_i | X_i\} = 0$ implies that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i e_i \ne 0 \quad \text{in probability} \tag{3}$$

and hence also inconsistency of

$$\hat{\beta}^{(OLS,n)} = \beta^0 + \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} X_i e_i. \tag{4}$$

The most frequently given examples of failure of the condition of orthogonality are the measurement of explanatory variables with a random error or the (dynamic) regression model with lagged response in the role of explanatory variable (Judge et al. (1985) or Víšek (1998)). Econometricians offer as a remedy the method of the *Instrumental Variables* which defines the estimator as (any) solution of the normal equations

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

where the sequence $\{Z_i\}_{i=1}^{\infty}$ is a sequence of i.i.d. instruments for explanatory variables X_i's given as follows: Let $\{U_i\}_{i=1}^{\infty}$ be a sequence of $p-1$-dimensional i.i.d. r.v.'s such that $I\!EU_1 \cdot e_1 = 0$, so that putting $Z_{i1} = 1$ and $Z_{ij} = U_{i,j-1}$ for all $i \in N$ the orthogonality condition $I\!EZ_1 e_1 = 0$ holds. The analogy of (4)

$$\hat{\beta}^{(IV,n)} = \beta^0 + \left(\frac{1}{n} \sum_{i=1}^{n} Z_i X_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} Z_i e_i. \tag{6}$$

hints that the estimator evaluated by means of method of the *Instrumental Variables* is consistent provided (e. g.)

$$I\!EZ_1 X_1' = Q \;\text{ is regular} \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Z_i e_i = 0 \quad \text{in probability} \tag{7}$$

In 1992 Hettmansperger and Sheather showed that the *Least Median of Squares (LMS)* (Rousseeuw (1984)) can be considerably sensitive to some very small changes of data. It appeared later that their result was due to bad algorithm for *LMS* (Víšek (1994)). Nevertheless, evaluating the *Least Trimmed Squares (LTS)* (Hampel (1986)) by total search for data used by Hettmansperger and Sheather (1992) (and hence reaching the exact value of the estimator) revealed that the problem exists for *LTS*. Academic examples in Víšek (1996b) and (2000a) indicated the reason for it (for any robust estimator with high *breakdown point*) and Víšek (1992), (1996a), (2000b) and (2002c) brought the theoretical justification of the fact that the discontinuous objective functions can cause (extremely) high sensitivity of robust estimators to some changes of data. That was an inspiration for defining the *Least Weighted Squares (LWS)* (Víšek (2000c), see also (2002a, b))

$$\hat{\beta}^{(LWS,n,w)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta)$$
$$= \arg\min_{\beta \in R^p} \sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i^2(\beta)|)\right) r_i^2(\beta) \tag{8}$$

where

$$F_\beta^{(n)}(v) = \frac{1}{n} \sum_{j=1}^{n} I\left\{|r_j(\beta)| < v\right\} = \frac{1}{n} \sum_{j=1}^{n} I\left\{|e_j - X_j'\beta| < v\right\} \tag{9}$$

is the empirical distribution function of the absolute values of residuals and w is a weight function fulfilling:

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

It is only a technicality to show that $\hat{\beta}^{(LWS,n,w)}$ has to be a solution of

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

Then again, if

$$w\left(F_\beta^{(n)}(|e_1|)\right) X_1 e_1 \neq 0,$$

$\hat{\beta}^{(LWS,n,w)}$ is inconsistent. The remedy is straightforward, given by *normal equations*

$$\sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) Z_i \left(Y_i - X_i'\beta\right) = 0 \tag{11}$$

where again the sequence $\{Z_i\}_{i=1}^{\infty}$ is a sequence of i.i.d. instruments for X_i's (see text below the equation (5) and Víšek (2004)).

In the case of "classical" *Instrumental Variables* (6) and (7) indicated that we don't need any "qualitative relation" between explanatory variables and instruments (although in practise it is not so - if there are not "natural" instrument, e.g. lagged values, the method can work poorly). However for robust version of the method we need some assumption about the mutual behaviour of X_i's and Z_i's. Let's recall that we assume heteroscedasticity of the error terms (see **C1**) and define a "mean" d.f.

$$\overline{F}_{n,\beta}(v) = \frac{1}{n}\sum_{i=1}^{n} P\left(|Y_i - X_i'\beta| < v\right). \tag{12}$$

(a possibility to approximate the empirical distribution $F_\beta^{(n)}(v)$ - see (9) - by $\overline{F}_{n,\beta}(v)$ uniformly in $v \in R$ as well as in $\beta \in R^p$ opened in fact the way for results given below, see Víšek (2008d)). Further define

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

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

$$\gamma_{\lambda,a} = \sup_{\|\beta\|=\lambda} F_{\beta' Z X' \beta}(a). \tag{13}$$

Finally, for any $\lambda \in R^+$ let us denote

$$\tau_\lambda = -\inf_{\|\beta\| \le \lambda} \beta' I\!E\left[Z_1 X_1' \cdot I\{\beta' Z_1 X_1' \beta < 0\}\right]\beta. \tag{14}$$

C3 *The $p-1$-dimensional r.v.'s $\{U_i\}_{i=1}^\infty$ are independent and identically distributed with distribution function $F_U(u)$. Moreover, they are independent from the sequence $\{e_i\}_{i=1}^\infty$, the joint distribution function $F_{V,U}(v,u)$ is absolutely continuous, $I\!E Z_1 Z_1'$ is positive definite and there is $q > 1$ so that $I\!E\left\{\|Z_1\| \cdot \|X_1\|\right\}^q < \infty$. Further, there is $n_0 \in N$ so that for all $n > n_0$ $I\!E\left\{\frac{1}{n}\sum_{i=1}^n \left[w(\overline{F}_{n,\beta}(|e_i|)) Z_i X_i'\right]\right\}$ is regular. Finally, there is $a > 0$, $b \in (0,1)$ and $\lambda > 0$ so that*

$$a \cdot (b - \gamma_{\lambda,a}) \cdot w(b) > \tau_\lambda \tag{15}$$

For discussion of **C3** see Víšek (2008a).

C4 *There is $n_0 \in N$ so that for all $n > n_0$ the vector equation*

$$\beta' I\!E\left\{\frac{1}{n}\sum_{i=1}^{n}\left[w(\overline{F}_{n,\beta}(|r_i(\beta)|)) Z_i\left(e_i - X_i'\beta\right)\right]\right\} = 0 \tag{16}$$

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

Lemma 1. *Let 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 (11) is weakly consistent.*

Proof is given in Víšek (2008a) where also a simulation study demonstrates that the algorithm, firstly presented in Víšek (2006a), works very well. Result in Víšek (2006b) opened way to prove $\sqrt{n}$-consistency and to find an asymptotic representation of $\hat\beta^{(IWV,n,w)}$ under following conditions (denote by $f_{e|V}(r|V_1 = x)$ the conditional density corresponding to the d.f. $F_{V,e}(v,r)$):

NC1 *The density $f_{e|V}(r|V_1 = x)$ 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) are* $\sqrt{n}$*-consistent.*

For the proof see Víšek (2008b).

Denote by $g(r)$ the density of the d.f. $G(r) = P(e_1^2 < r)$ (notice that under **C1** density $g(r)$ always exists). Moreover, for any $\alpha \in (0,1)$ denote by u_α^2 the upper α-quantile of d.f. G, i.e. we have $P(e_1^2 > u_\alpha^2) = \alpha$.

AC1 *For any* $\alpha \in (0,1)$ *there is* $\delta(\alpha) > 0$ *so that*

$$\inf_{r \in (0, u_\alpha^2 + \delta(\alpha))} g(r) > L_{g,\alpha} > 0 \quad \text{and} \quad \inf_{|r| \in (0, \sqrt{u_\alpha^2 + \delta(\alpha)})} f(r) > L_{f,\alpha} > 0. \tag{17}$$

Similarly as above (see text under **C1**) the condition **AC1** implies in fact that (17) holds for all densities $g_{e_i}(r)$ and $f_{e_i}(r)$, i.e. for all $i \in N$.

AC2 *There is* $q > 1$ *so that* $\sup_{i \in N} I\!\!E \left|e_i\right|^{2q} < \infty$.

Lemma 3. *Let the conditions* **C1***,* **C2***,* **C3***,* **C4***,* **NC1***,* **NC2***,* **AC1** *and* **AC2** *hold. Then*

$$\sqrt{n}\left(\hat{\beta}^{(IWV,n,w)} - \beta^0\right) =$$

$$\left[\frac{1}{n}\sum_{i=1}^{n} w\left(\overline{F}_{n,\beta}(^0|e_i|)\right) \cdot Z_i X_i'\right]^{-1} \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w\left(\overline{F}_{n,\beta^0}(|e_i|)\right) \cdot Z_i e_i + o_p(1) \tag{18}$$

as $n \to \infty$.

Having at hand the algorithm for the IWV and applying it on data, one needs a test for homoscedasticity of error terms as disregarding heteroscedasticity my lead to poor identification of regression model, frequently wrongly assuming some insignificant explanatory variables as significant. Such a test was for IWV established in Víšek (2007). When the test rejects the homoscedasticity, we need estimators of variances of the estimates of regression coefficient "robust" against heteroscedasticity. Following Halbert White (1980) and employing (4), we may prove:

Lemma 4. *Let the conditions* **C1***,* **C2***,* **C3***,* **C4***,* **NC1***,* **NC2***,* **AC1** *and* **AC2** *hold. Then*

$$\left[\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right]^{-1} \left[\sum_{i=1}^{n} r_i^2(\hat{\beta}^{(IWV,n,w)}) Z_i X_i'\right] \left[\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right]^{-1}$$

is weakly consistent estimator of covariance matrix of $\hat{\beta}^{(IWV,n,w)}$.

References

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.

Hettmansperger, T. P., S. J. Sheather (1992): A Cautionary Note on the Method of Least Median Squares. *The American Statistician 46, 79–83.*

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)

Rousseeuw, P.J. (1984): Least median of square regression. *Journal of Amer. Statist. Association 79, 871-880.*

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

Víšek, J. Á. (1994): A cautionary note on the method of the Least Median of Squares reconsidered, *Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, 1994, 254 - 259.*

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

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

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

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

Víšek, J. Á. (2000b): A new paradigm of point estimation. Proceedings of *Data Analysis 2000/II, Modern Statistical Methods - Modelling, Regression, Classification and Data Mining*, ISBN 80-238-6590-0, 195 - 230.

Víšek, J. Á. (2000c): Regression with high breakdown point. *Robust 2000 (eds. Jaromír Antoch & Gejza Dohnal, published by Union of Czechoslovak Mathematicians and Physicists), 2001, ISBN 80-7015-792-5, 324 - 356.*

Víšek, J. Á. (2002a): 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. (2002b):The least weighted squares II. Consistency and asymptotic normality. *Bulletin of the Czech Econometric Society, Vol. 9, no. 16, 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 (2002), 261 - 290.*

Víšek, J. Á. (2004): Robustifying instrumental variables. *Proceedings of COMPSTAT'2004. Physica-Verlag/Springer. ISBN 3-7908-1554-3. 1947 - 1954.*

Víšek, J. Á. (2006a): Instrumental Weighted Variables- algorithm. *Proceedings of the COMPSTAT 2006, Rome (28.8.- 1.9.2006), eds. A. Rizzi & M. Vichi, Physica-Verlag (Springer Company) Heidelberg 2006, ISBN-10 3-7908-1708-2 ISBN-13 978-3–7908-1708-2, 777-786.*

Víšek, J. Á. (2006b): Kolmogorov-Smirnov statistics in multiple regression. *Proceedings of the ROBUST 2006, eds. Jaromír Antoch & Gejza Dohnal, ISBN 80-7015-073-4, 367-374.*

Víšek, J. Á. (2007): White test for the instrumental weighted variables. *Submited to Bulletin of the Czech Econometrc Society*, presented on ICORS 2006.

Víšek, J. Á. (2008a): Consistency of the instrumental weighted variables. To appear in the *Annals of the Institute of Statistical Mathematics.*

Víšek, J. Á. (2008b): $\sqrt{n}$-consistency of the instrumental weighted variables. Submited to the *Annals of the Institute of Statistical Mathematics.*

Víšek, J. Á. (2008c): Asymptotic representation of the instrumental weighted variables. *Preprint*

Víšek, J. Á. (2008d): Empirical distribution function under heteroscedasticity. Submitted to the *Statistics.*

White, H. (1980): A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica, 48, 817 - 838.*