Acta Universitatis Carolinae. Mathematica et Physica

Jan Ámos Víšek

Weak $\sqrt{n}$-consistency of the least weighted squares under heteroscedasticity

Acta Universitatis Carolinae. Mathematica et Physica, Vol. 52 (2011), No. 1, 81--92

Persistent URL: http://dml.cz/dmlcz/143670

Weak $\sqrt{n}$-Consistency of the Least Weighted Squares under Heteroscedasticity

JAN ÁMOS VÍŠEK

Praha

Received April 26, 2010
Revised June 20, 2010

Weak $\sqrt{n}$-consistency of the *Least Weighted Squares* estimator of the coefficients of regression model is proved generally under the heteroscedasticity of error terms. The assumptions required for the weak $\sqrt{n}$-consistency are briefly discussed. The roots of the heteroscedasticity are also critically considered.

1. Introduction

The paper continues in studies of [21] where the reasons for introducing the *Least Weighted Squares* (*LWS*) were already discussed in details, see also [14].

Let $\mathscr{N}$ denote the set of all positive integers, $\mathscr{R}$ the real line and $\mathscr{R}^p$ the p-dimensional Euclidean space. The linear regression model given as

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

will be considered. For any $\beta \in \mathscr{R}^p$ $r_i(\beta) = Y_i - X_i'\beta$ denotes the i-th residual and $r_{(h)}^2(\beta)$ stays for the h-th order statistic among the squared residuals, i.e. we have $r_{(1)}^2(\beta) \le r_{(2)}^2(\beta) \le \ldots \le r_{(n)}^2(\beta)$. We shall assume:

Conditions $\mathscr{C}1$ *The sequence* $\left\{\left(X_i', e_i\right)'\right\}_{i=1}^{\infty}$ *is sequence of independent* $(p+1)$*-dimensional random variables (r.v.'s) distributed according to a distribution functions (d.f.)*

Faculty of Social Sciences, Charles University, Institute of Economic Studies, Opletalova 26, 110 01 Praha 1

We would like to thank an anonymous referee for careful reading the draft of paper which led to significant improvement of it. Research was supported by grant of GA ČR 402/09/0557.

Key words and phrases. Robustness, implicit weighting, weak $\sqrt{n}$-consistency of estimate by the least weighted squares, heteroscedasticity

E-mail address: visek@fsv.cuni.cz

$F_{X,e_i}(x,r) = F_X(x)\cdot F_{e_i}(r)$ where $F_{e_i}(r) = F_e(r\sigma_i^{-1})$ with $\mathbb{E}e_i = 0$, $\mathrm{var}\,(e_i) = \sigma_i^2$ and $0 < \liminf_{i\to\infty} \sigma_i \leq \limsup_{i\to\infty} \sigma_i < \infty$. Moreover, $F_e(r)$ is absolutely continuous with the density $f_e(r)$ bounded by U_e. Finally, there is $q > 1$ so that $\mathbb{E}\,\|X_1\|^{2q} < \infty$ (as $F_X(x)$ does not depend on i, the sequence $\{X_i\}_{i=1}^{\infty}$ is sequence of independent and identically distributed (i.i.d.) r.v.'s).

Prior to continuing, let us briefly discuss Conditions $\mathscr{C}1$. Such discussion will reflect also the reasons, pros and cons for the present studies of weak $\sqrt{n}$-consistency of LWS under heteroscedasticity.

First of all, let us recall that e_i's are called either *error terms* or *disturbances*. The former name is used mainly in the exact sciences. In this context the regression model describes a mutual relation between a set of explanatory variables on one side and a response variable on the other. As the set of explanatory variables is nearly always exhaustively determined by the merits of problem in question, the error terms represent errors of the measurement of response variable. Then of course, the heteroscedasticity of the error terms is mostly due to the dependence of the variances of e_i's on the explanatory variables. After all, one of the tests of heteroscedasticity is based on testing independence of the error terms and the explanatory variables, see [22]. If White's test rejects the hypotheses of homoscedasticity, we should assume that also the *Orthogonality Condition* is broken, i. e. that $\mathbb{E}\{e_i|X_i\} \neq 0$. Then the method of *Instrumental Variables* (see [11]) or its robustified version the *instrumental weighted variables* (see [18, 19]) is to be used.

In contrast, in the social sciences we typically meet with the situations when we may assume that some explanatory variables were not (or even could not be) included into the model, see e.g. [2, 3, 10, 23, 24]. So, we may assume that some explanatory variables are (as latent variables) in the disturbances (there are attempts to cope with the situation, e.g. *fix and random effects* represent one such a proposal, see again [7] or [24]). Then the heteroscedaticity of disturbances may be caused just by these "hidden" part of a set of all possible explanatory variables. This situation is (indirectly) reflected by proposals of the tests of heteroscedasticity which are based on a specification of models of the heteroscedasticity. These models usually assume that the variances of e_i's can be given as $var(e_i) = h(\alpha' \cdot Z_i)$ where α is a vector of coefficients and Z_i's can be (statistically) independent from X_i's (for a whole family of such tests and a discussion see e.g. [7] or [12]).

The applications of the regression analysis in social sciences deserves (maybe) also other remark. The framework for regression model treats the explanatory variables either as a sequence of deterministic vectors (see e.g. [26]) or as a sequence of independent and identically distributed, say p-dimensional, random variables (see e.g. [5, 24], among many others). The former approach reflects the situation when it seems (evidently) strange to treat explanatory variables as random (mainly in the situations when we can assign the values to them – e.g. when performing some experiment). Nevertheless, even then – to be able to study asymptotic properties – we have to assume something like "pseudorandom" character of these vectors, see e.g. [16] or again [26]. The later approach takes into account the fact that in some other cases we just collect data (observations) and so the explanatory variable can be assumed to be a realization of a random sequence. Of course, even then there can be a discussion whether we can assume that they are independent and identically distributed. To allow at least for heteroscedasticity is just the goal of this paper. In the case of panel data (when we look in fact for a cointegration model of time series) we can express our ideas about a correlation structure by some model of Box-Jenkins type, see e.g. [1, 7, 24, 25]. But after all, in such a situation we usually help ourselves by transforming the time series in question to the independence by e.g. Cochrane-Orcutt or Prais-Winsten transformation (see [4] or [9]). The last step however requires serious care, as we can worsen the situation rather than improve, if we do not "guess" a proper model for correlation structure (for a nice example see [8]).

Last but not least, a proof of consistency of LWS under the heteroscedasticity open a possibility to establish a robustified version of a test of specificity [6] for LWS and so to put a decision of when apply the least weighted squares and when the instrumental weighted variables on a sound basis.

Remark 1 *Notice that under Conditions* $\mathscr{C}1$ *there are constants* $0 < s_\sigma \leq S_\sigma < \infty$ *so that* $s_\sigma \leq \sigma_i \leq S_\sigma$ *for all i's. Moreover, as the density of* e_i *is given as* $f_e(r \cdot \sigma_i^{-1}) \cdot \cdot \sigma_i^{-1}$, *there is a constant* $f_\sigma < \infty$ *such that*

$$\sup_{i \in \mathscr{N}} \sup_{r \in \mathscr{R}} f_{e_i}(r) < f_\sigma.$$

The assumption that the d.f. $F_e(r)$ *is continuous is not only a technical assumption. Possibility that the error terms in regression model are discrete r.v.'s implies problems with treating response variable and it requires special considerations - see chapters on logit or probit models or limited response variables e.g.* [7]. *Absolute continuity is then a technical assumption. Without the density (even bounded density) we have to assume that* $F_e(r)$ *is Lipschitz and it would bring a more complicated form of all what follows.*

Conditions $\mathscr{C}2$ *The weight function* $w(u)$ *is continuous, nonincreasing,* $w : [0,1] \to \to [0,1]$ *with* $w(0) = 1$. *Moreover, w is Lipschitz in absolute value, i.e. there is* L_w *such that for any pair* $u_1, u_2 \in [0,1]$ *we have* $|w(u_1) - w(u_2)| \leq L_w \cdot |u_1 - u_2|$.

Remark 2 *If the weight function w is not continuous, as e.g. for LTS, the techniques of proofs of asymptotic properties have to be more complicated (see* [17]*) and on the other hand, due to discontinuity of w, the estimator may be much more sensitive to (even very small) change of one observation, for details see* [13, 15] *or* [16].

Further, let e be a r.v. distributed according to $F_e(v)$ and for any $\beta \in \mathscr{R}^p$ denote $F_\beta(v) = P(|e - X_1'\left(\beta - \beta^0\right)| < v)$ and $r(\beta) = e - X_1'\left(\beta - \beta^0\right)$.

Conditions $\mathscr{C}3$ *There is the only solution of*

$$\mathbb{E}\left[w\left(F_\beta(|r(\beta)|)\right)\beta' X_1 \left[e - X_1'\left(\beta - \beta^0\right)\right]\right] = 0 \tag{2}$$

namely β^0. *Moreover*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |1 - \sigma_i| = 0.$$

Definition 1 *Let* $w : [0,1] \to [0,1]$ *be a nonincreasing function with* $w(0) = 1$. *Then the solution of the extremal problem*

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta\in\mathscr{R}^p}{arg\,min} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta) \tag{3}$$

will be called the Least Weighted Squares.

For any $\beta \in \mathscr{R}^p$ the empirical d. f. of the absolute values of residuals will be denoted $F_\beta^{(n)}(r)$. It means that, denoting the indicator of a set A by $I\{A\}$, we have

$$F_\beta^{(n)}(r) = \frac{1}{n}\sum_{i=1}^{n} I\{|r_i(\beta)| < r\} = \frac{1}{n}\sum_{i=1}^{n} I\left\{\left|Y_i - X_i'\beta\right| < r\right\}. \tag{4}$$

Taking into account that empirical d. f. has the jumps (of a magnitude $\frac{1}{n}$) just at the order statistics, we obtain

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta\in\mathscr{R}^p}{\arg\min} \sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) r_i^2(\beta) \tag{5}$$

and, then, it is straightforward that $\hat{\beta}^{(LWS,n,w)}$ is given as one of solutions of the *normal equations* (for details see [21]

$$\mathbb{N}E_{Y,X}^{(n)}(\beta) = \sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) X_i \left(Y_i - X_i'\beta\right) = 0. \tag{6}$$

One of key steps in the considerations about the weak $\sqrt{n}$-consistency of $\hat{\beta}^{(LWS,n,w)}$ is an approximation of the empirical d. f. $F_\beta^{(n)}(r)$ (see (4)) by the theoretical one $\overline{F}_{n,\beta}(r)$ where

$$\overline{F}_{n,\beta}(v) = \frac{1}{n}\sum_{i=1}^{n} F_{\beta,i}(v) = \frac{1}{n}\sum_{i=1}^{n} P\left(\left|Y_i - X_i'\beta\right| < v\right). \tag{7}$$

Theorem 1 *Let Conditions* $\mathscr{C}1$, $\mathscr{C}2$ *and* $\mathscr{C}3$ *be fulfilled. Then any sequence* $\left\{\hat{\beta}^{(LWS,n,w)}\right\}_{n=1}^{\infty}$ *of solutions of the normal equations (6)* $\mathbb{N}E_{Y,X}^{(n)}(\hat{\beta}^{(LWS,n,w)}) = 0$ *is consistent.*

For the proof see [21].

2. Weak $\sqrt{n}$-consistency of the least weighted squares

Conditions $\mathscr{NC}1$ *The derivative* $f_e'(r)$ *exists and is bounded in absolute value by* B_e. *The derivative* $w'(\alpha)$ *exists and is Lipschitz of the first order (with the corresponding constant* J_w*). Moreover, for any* $i \in \mathscr{N}$

$$\mathbb{E}\left[w'(\overline{F}_{n,\beta^0}(|e_i|))\left(f_e(|e_i|) - f_e(-|e_i|)\right)\cdot e_i\right] = 0.$$

Finally, for any $j, k, \ell = 1, 2, ..., p$ $\mathbb{E}\left|X_{1j} \cdot X_{1k} \cdot X_{1\ell}\right| < \infty$ *(as* $F_X(x)$ *does not depend on i, the sequence* $\{X_i\}_{i=1}^{\infty}$ *is sequence of independent and identically distributed p-dimensional r.v.'s).*

Theorem 2 *Let Conditions* $\mathscr{C}1$, $\mathscr{C}2$, $\mathscr{C}3$ *and* $\mathscr{N}\mathscr{C}1$ *hold. Then any sequence* $\left\{\hat{\beta}^{(LWS,n,w)}\right\}_{n=1}^{\infty}$ *of solutions of the normal equations (6)* $\mathbb{N}E_{Y,X}^{(n)}(\hat{\beta}^{(LWS,n,w)}) = 0$ *is weakly* $\sqrt{n}$*-consistent, i.e.* $\forall(\varepsilon > 0)\ \exists(K_\varepsilon < \infty)\ \forall(n \in \mathscr{N})$

$$P\left(\left\{\omega \in \Omega :\ \sqrt{n}\left\|\hat{\beta}^{(LWS,n,w)} - \beta^0)\right\| < K_\varepsilon\right\}\right) > 1 - \varepsilon.$$

Proof: Let us recall that $\hat{\beta}^{(LWS,n,w)}$ is one of the solutions of (see (6))

$$\sum_{i=1}^{n} w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) X_i\,(Y_i - X_i'\beta) = 0. \tag{8}$$

For any solution of (8), we have (write $w_{r_i(\beta)}$ instead of $w\left(F_\beta^{(n)}(|r_i(\beta)|)\right)$)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{r_i(\beta)} X_i e_i = \frac{1}{n}\sum_{i=1}^{n} w_{r_i(\beta)} X_i X_i' \cdot \sqrt{n}\left(\beta - \beta^0\right). \tag{9}$$

The idea of the proof of Theorem 2 is then as follows. We show that (9) can be rewritten as

$$L_n = Q_n \cdot (1 + q_n) \cdot \sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right) \tag{10}$$

where $L_n = \mathscr{O}_p(1)$, $Q_n \to Q$ in probability, Q being a regular matrix, and $q_n = o_p(\hat{\beta}^{(LWS,n,w)} - \beta^0)$. Then assuming that $(1 + q_n) \cdot \sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)$ is not $\mathscr{O}_p(1)$ and employing the Lemma 6, we prove that also L_n cannot be $\mathscr{O}_p(1)$, which is a contradiction.

Since w is Lipschitz, employing Lemma 1 (see Appendix)

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

$$\leq \sqrt{n} \cdot L_w \cdot \sup_{v \in \mathscr{R}^+}\sup_{\beta \in \mathscr{R}^p}\left|F_\beta^{(n)}(v) - \overline{F}_{n,\beta}(v)\right| \cdot \frac{1}{n}\sum_{i=1}^{n}\|X_i\| \cdot |e_i| = O_p(1)$$

as $n \to \infty$. Hence

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

where we have $\forall(\varepsilon > 0)\ \exists(K_\varepsilon < \infty)\ \forall(n \in \mathscr{N})$

$$P\left(\left\{\omega \in \Omega\ :\ \sup_{\beta \in \mathscr{R}^p}\left\|R_{X,e}^{(1,n)}(\beta)\right\| < K_\varepsilon\right\}\right) > 1 - \varepsilon.$$

In a similar way we derive

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

where we have $\forall(\varepsilon > 0, \delta > 0)\ \exists(n_{\varepsilon,\delta} \in \mathcal{N})\ \forall(n > n_{\varepsilon,\delta})$

$$P\left(\left\{\omega \in \Omega : \sup_{\beta \in \mathscr{R}^p} \left\|R_{X,e}^{(2,n)}(\beta)\right\|_M < \delta\right\}\right) > 1 - \varepsilon$$

(where $\|A\|_M = \max_{i,j=1,2,\ldots,k} |a_{ij}|$).Utilizing these approximations in (9) we get

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) X_i e_i + R_{X,e}^{(1,n)}(\beta) \tag{11}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) X_i X_i' + R_{X,e}^{(2,n)}(\beta)\right] \cdot \sqrt{n}\left(\beta - \beta^0\right). \tag{12}$$

Now employing successively Lemma 2, 3, 4 and 5 (see Appendix), we find that (11) can be written as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Bigg\{ w\left(\overline{F}_{n,\beta^0}(|e_i|)\right) + w'(\overline{F}_{n,\beta^0}(|e_i|)) \times \tag{13}$$

$$\times \left\{\left(f_e(|e_i|) - f_e(-|e_i|)\right) \cdot (X_i' - E_X X_1')\right\} \left[\beta - \beta^0\right] \Bigg\} X_i e_i + R_{X,e}^{(3,n)}(\beta) \tag{14}$$

with $\forall(\varepsilon > 0, \delta > 0)\ \exists(K_\varepsilon < \infty)\ \forall(n \in \mathcal{N})$

$$P\left(\left\{\omega \in \Omega \ : \ \sup_{\beta \in \mathscr{R}^p} \left\|R_{X,e}^{(3,n)}(\beta)\right\| < K_\varepsilon\right\}\right) > 1 - \varepsilon \tag{15}$$

and (12) turns into

$$\frac{1}{n} \sum_{i=1}^{n} w\left(\overline{F}_{n,\beta^0}(|e_i|)\right) X_i X_i' \cdot \sqrt{n}\left(\beta - \beta^0\right) + R_{X,e}^{(4,n)}(\beta) \tag{16}$$

with $\forall(\varepsilon > 0, \delta > 0)\ \exists(n_{\varepsilon,\delta} \in \mathcal{N})\ \forall(n \in \mathcal{N}, n > n\varepsilon, \delta)$

$$P\left(\left\{\omega \in \Omega : \sup_{\beta \in \mathscr{R}^p} \left\|R_{X,e}^{(4,n)}(\beta)\right\|_M < \delta\right\}\right) > 1 - \varepsilon. \tag{17}$$

Utilizing CLT and taking into account (16) we find that (13) and (14) are $\mathcal{O}_p(1)$. Similarly, due to the law of large numbers $\left\|\frac{1}{n} \sum_{i=1}^{n} w\left(\overline{F}_{n,\beta^0}(|e_i|)\right) \cdot \ \cdot X_i X_i'\right\|$ is $\mathcal{O}_p(1)$. Plugging in for β the estimate $\hat{\beta}^{(LWS,T,w)}$ and employing Lemma 6 together with Theorem 1 we conclude the proof.

3. Conclusions

We have shown the weak $\sqrt{n}$-consistency of the *least weighted squares* under the heteroscedasticity of disturbances. Of course, the *least weighted squares* are a special case of the *instrumental weighted variables* (see [19]) when we assume as instruments just the explanatory variables. Nevertheless, as one can see in [19], we cannot rid of some requirements on the instruments which however need not be automatically fulfilled by the explanatory variables. On the other hand, as it was already recalled in the introduction, we need the consistency of $\hat{\beta}^{(LWS,n,w)}$ for a possibility to study an asymptotics of a robustified version of the test of specificity. Then we have no choice but to prove the consistency of the *least weighted squares* under the heteroscedasticity directly. It is clear that to be able to propose a robustified version of the test of specificity (employing $\hat{\beta}^{(LWS,n,w)}$) and to study in a reasonable way its asymptotics, we will need also the consistency of the *instrumental weighted variables* under the heteroscedasticity (this problem is still under research).

4. Appendix

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

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

For the proof see [20].

Lemma 2 *Let Conditions $\mathscr{C}1$, $\mathscr{C}2$, $\mathscr{C}3$ and $\mathscr{N}\mathscr{C}1$ hold. Then, as $n \to \infty$,*

$$\sup_{\beta\in\mathscr{R}^p} \sup_{v\in\mathscr{R}^+} \sup_{i\in\mathscr{N}} \left|\overline{F}_{n,\beta}(v) - \overline{F}_{n,\beta^0}(v) - \left[f_e(-v\cdot\sigma_i^{-1}) \right.\right.$$
$$\left.\left. + f_e(v\cdot\sigma_i^{-1})\right]\mathbb{E}_X X_1' \cdot \left[\beta - \beta^0\right]\right| \cdot \left\|\beta - \beta^0\right\|^{-2} = \mathscr{O}(1)$$

and

$$\sup_{\beta\in\mathscr{R}^p} \sup_{v\in\mathscr{R}^+} \left|\overline{F}_{n,\beta}(v) - \overline{F}_{n,\beta^0}(v)\right| \cdot \left\|\beta - \beta^0\right\|^{-1} = \mathscr{O}(1). \tag{18}$$

Proof: Putting $a = \left[v - x'\left(\beta - \beta^0\right)\right]\sigma_i$ and $b = v\cdot\sigma_i$, the proof is based on evaluating $\overline{F}_{n,\beta}(v) - \overline{F}_{n,\beta^0}(v)$ as

$$\frac{1}{n}\sum_{i=1}^n \left\{\int_{-\infty}^{\infty}\left[\int_{-a}^{a} f_e(r\sigma_i^{-1})\sigma_i^{-1}\mathrm{d}r\right]\mathrm{d}F_X(x) - \int_{-b}^{b} f_e(r\sigma_i^{-1})\sigma_i^{-1}\mathrm{d}r\right\}.$$

Lemma 3 *Let Conditions $\mathscr{C}1$, $\mathscr{C}2$, $\mathscr{C}3$ and $\mathscr{N}\mathscr{C}1$ hold. Then, as $n \to \infty$,*

$$\sup_{\beta\in\mathscr{R}^p} \sup_{v\in\mathscr{R}^+} \left\|\frac{1}{n}\sum_{i=1}^n \left\{\left[w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) - w\left(\overline{F}_{n,\beta^0}(|r_i(\beta)|)\right)\right]\right.\right.$$
$$\left.\left. - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|))\cdot\left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right]\right\} X_i e_i\right\|$$
$$= \mathscr{O}_p(\|\beta - \beta^0\|^2)$$

and for any $\ell, k = 1, 2, \ldots, p$

$$\sup_{\beta \in \mathscr{R}^p} \sup_{v \in \mathscr{R}^+} \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ \left[w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) - w\left(\overline{F}_{n,\beta^0}(|r_i(\beta)|)\right) \right] \right. \right.$$
$$\left. \left. - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|)) \cdot \left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|) \right] \right\} X_{i\ell} X_{ik} \right|$$
$$= \mathscr{O}_p(\|\beta - \beta^0\|^2).$$

Remark 3 *Notice please that (21) says that when we substitute*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) X_i e_i$$

by

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ w\left(\overline{F}_{n,\beta^0}(|r_i(\beta)|)\right) - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|)) \right.$$
$$\left. \times \left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|) \right] \right\} X_i e_i,$$

the rest (in norm) is of order $\sqrt{n} \left\| \beta - \beta^0 \right\| \cdot \mathscr{O}_p(\left\| \beta - \beta^0 \right\|)$. Plugging, then the estimate $\hat{\beta}^{(LWS,n,w)}$ instead of β, we obtain the rest of order $\sqrt{n} \left\| \hat{\beta}^{(LWS,n,w)} - \beta^0 \right\| \cdot o_p(1)$ (due to weak consistency of $\hat{\beta}^{(LWS,n,w)}$). So the rest can be "shifted" to the right hand side of (10), into $q_n \cdot \sqrt{n} \left(\hat{\beta}^{(LWS,n,w)} - \beta^0 \right)$.

Similarly, when we substitute

$$\frac{1}{n} \sum_{i=1}^{n} w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) X_{i\ell} X_{ik}$$

by

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ w\left(\overline{F}_{n,\beta^0}(|r_i(\beta)|)\right) - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|)) \right.$$
$$\left. \times \left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|) \right] \right\} X_{i\ell} X_{ik},$$

the rest (in absolute value) is of order $\mathscr{O}_p(\left\| \beta - \beta^0 \right\|^2)$. When we plug moreover instead of β the estimate $\hat{\beta}^{(LWS,n,w)}$, we obtain the corresponding rest of order $\sqrt{n} \left\| \hat{\beta}^{(LWS,n,w)} - \beta^0 \right\| \cdot o_p(n^{-\frac{1}{2}})$ (due to the weak consistency of $\hat{\beta}^{(LWS,n,w)}$). So the rest can be again "shifted" into $\sqrt{n} \left(\hat{\beta}^{(LWS,n,w)} - \beta^0 \right) \cdot q_n$ of (10).

The proof of Lemma 3 is typical one and moreover, it is not very long. That is why we give it in full details.

Consider the first assertion of the lemma. We have for any $n \in \mathcal{N}$ and any $i = 1, 2, \ldots, n$

$$
\begin{aligned}
& w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) - w\left(\overline{F}_{n,\beta^0}(|r_i(\beta)|)\right) \\
& = w'(\xi_i)\left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right] \\
& = \left[w'(\xi_i) - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|)\right] \cdot \left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right] \\
& + w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|) \cdot \left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right]
\end{aligned} \tag{19}
$$

where $\xi_i \in \left[\overline{F}_{n,\beta}(|r_i(\beta)|), \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right]_{ord}$. Moreover, using J_w from Remark 1 (and recalling that w is monotone function)

$$
\begin{aligned}
& \left|w'(\xi_i) - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|)\right| \cdot \left|\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right| \\
& \leq J_w \cdot \left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right]^2 \\
& \leq J_w \cdot \sup_{v \in R^+} \left[\overline{F}_{n,\beta}(v) - \overline{F}_{n,\beta^0}(v)\right]^2 = \mathcal{O}(\left\|\beta - \beta^0\right\|^2)
\end{aligned} \tag{20}
$$

where the last equality is due to (18). Finally, we have for any $n \in \mathcal{N}$

$$
\begin{aligned}
& \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[w'(\xi_i) - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|))\right] \right. \\
& \left. \times \left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right] \cdot X_i e_i \right\| \\
& = \sqrt{n} \left\|\beta - \beta^0\right\| \cdot \mathcal{O}_p(\left\|\beta - \beta^0\right\|) \frac{1}{n} \sum_{i=1}^{n} \|X_i\| \cdot |e_i| \, .
\end{aligned}
$$

As $\frac{1}{n} \sum_{i=1}^{n} \|X_i\| \cdot |e_i|$ is finite in probability, we conclude that

$$
\begin{aligned}
& \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[w'(\xi_i) - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|))\right] \right. \\
& \left. \times \left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right] \cdot X_i e_i \right\| \\
& \leq \sqrt{n} \left\|\beta - \beta^0\right\| \cdot \mathcal{O}_p(\left\|\beta - \beta^0\right\|).
\end{aligned} \tag{21}
$$

Now, let us prove the second assertion of the lemma, i. e.

$$
\begin{aligned}
& \sup_{\beta \in \mathcal{R}^p} \sup_{v \in \mathcal{R}^+} \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ \left[w\left(\overline{F}_{n,\beta}(|r_i(\beta)|)\right) - w\left(\overline{F}_{n,\beta^0}(|r_i(\beta)|)\right)\right] \right. \right. \\
& \left. \left. - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|)) \cdot \left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right] \right\} X_{i\ell} X_{ik} \right| \\
& = \mathcal{O}_p(\|\beta - \beta^0\|^2)
\end{aligned}
$$

for any $\ell, k = 1, 2, \ldots, p$. Employing (20) once again we arrive at (for any $n \in \mathscr{N}$)

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left[w'(\xi_i) - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|))\right]\right.$$
$$\left.\times\left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right]\cdot X_{i\ell}X_{ik}\right|$$
$$\leq J_w \cdot \sup_{v\in R^+}\left[\overline{F}_{n,\beta}(v) - \overline{F}_{n,\beta^0}(v)\right]^2 \frac{1}{n}\sum_{i=1}^{n}|X_{i\ell}X_{ik}|$$
$$= \mathscr{O}_p(\left\|\beta - \beta^0\right\|^2)\frac{1}{n}\sum_{i=1}^{n}|X_{i\ell}X_{ik}|\,.$$

As $\frac{1}{n}\sum_{i=1}^{n}|X_{i\ell}X_{ik}|$ is finite in probability, we conclude that

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left[w'(\xi_i) - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|))\right]\right.$$
$$\left.\times\left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right]\cdot X_{i\ell}X_{ik}\right|$$
$$= \mathscr{O}_p(\left\|\beta - \beta^0\right\|^2). \tag{22}$$

□

Lemma 4 *Let Conditions $\mathscr{C}1'$, $\mathscr{C}2$, $\mathscr{C}3$ and $\mathscr{N}\mathscr{C}1$ hold. Then, as $n \to \infty$,*

$$\sup_{\beta\in\mathscr{R}^p}\frac{1}{n}\sum_{i=1}^{n}\left|w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|)) - w'(\overline{F}_{n,\beta^0}(|r_i(\beta^0)|))\right|\times$$
$$\times\left|\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta^0)|)\right|\cdot\left\|X_i e_i\right\| = \mathscr{O}_p(\left\|\beta - \beta^0\right\|^2)$$

and for any $\ell, k = 1, 2, \ldots, p$

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left[w'(\xi_i) - w'(\overline{F}_{n,\beta^0}(|r_i(\beta)|))\right]\right.$$
$$\left.\times\left[\overline{F}_{n,\beta}(|r_i(\beta)|) - \overline{F}_{n,\beta^0}(|r_i(\beta)|)\right]\cdot X_{i\ell}\cdot X_{ik}\right| = \mathscr{O}_p(\left\|\beta - \beta^0\right\|^2).$$

The proof consists of a chain of straightforward steps of finding the upper bound of respective expressions.

Lemma 5 *Let Conditions $\mathscr{C}1$, $\mathscr{C}2$, $\mathscr{C}3$ and $\mathscr{N}\mathscr{C}1$ hold. Then, as $n \to \infty$,*

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\beta - \beta^0\right\|^{-2}\cdot\left\|w\left(\overline{F}_{n,\beta^0}(|r_i(\beta)|)\right) - w\left(\overline{F}_{n,\beta^0}(|r_i(\beta^0)|)\right)\right.$$
$$\left.- \, w'(\overline{F}_{n,\beta^0}(|e_i|))\left[\left(f_e(e_i) - f_e(-e_i)\right)\cdot X_i'\left(\beta - \beta^0\right)\right]\cdot X_i e_i\right\| = \mathscr{O}(1).$$

The proof again consists of a chain of straightforward steps employing CLT and the law of large numbers.

Lemma 6 *Let for some $p \in \mathcal{N}$, $\left\{V^{(n)}\right\}_{n=1}^{\infty}$, $V^{(n)} = \left\{v_{ij}^{(n)}\right\}_{i=1,2,\ldots,p}^{j=1,2,\ldots,p}$ be a sequence of $(p \times p)$ matrixes such that for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, p$*

$$\lim_{n\to\infty} v_{ij}^{(n)} = q_{ij} \qquad \text{in probability} \tag{23}$$

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

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

Then

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

so that

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

For the proof see [13] or [15].

References

[1] Anděl, J., Zichová, J.: *A method foe estimating parameter in nonnegative MA(1) models.* Comm. Statist. Theory Methods **31** (2002), 2101–2111.

[2] Berndt, E.: The Practice of Econometrics. Addison-Wesley, Reading (1990).

[3] Bollen, K. A.: Structural Equations with Latent Variables. New York: Wiley (1989).

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

[5] Greene, W. H.: Econometric Analysis. Macmillam Press, New York (1993).

[6] Hausman, J.: *Specification test in econometrics.* Econometrica **46** (1978), 1251–1271.

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

[8] Mizon, G. E.: *A simple message for autocorrelation correctors: Don't.* J. Econometrics **69** (1995), 267–288.

[9] Prais, S. J., Winsten, C. B.: *Trend estimators and serial correlation.* Cowless Commission Discussion Paper No 383, Chicago (1954).

[10] Raudenbush, S. W., Bryk, A. S.: Hierarchical Linear Models: Applications and Data Analysis Methods (2nd Edition). Sage, Thousand Oaks (2002).

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

[12] Szroeter, J.: *A class of parametric tests of heteroscedasticity in linear econometric models.* Econometrica **46** (1978), 1311–1328.

[13] Víšek, J. Á.: *Sensitivity analysis of M-estimates.* Ann. Inst. Statist. Math. **48** (1996), 469–495.

[14] Víšek, J. Á.: *Regression with high breakdown point.* In: Robust 2000, J. Antoch & G. Dohnal, (eds) published by Union of Czech Mathematicians and Physicists), Matfyzpress, Prague (2000), 324–356.

[15] Víšek, J. Á.: *Sensitivity analysis of M-estimates of nonlinear regression model: Influence of data subsets.* Ann. Inst. Statist. Math. **54** (2) (2002), 261–290.

[16] Víšek, J. Á.: *The least trimmed squares*. Sensitivity study. In: Proceedings of the Prague Stochastics 2006, M. Hušková & M. Janžura, (eds) Matfyzpress, Prague (2006a), 728–738.

[17] Víšek, J. Á.: *The least trimmed squares. Part I-Consistency.* Kybernetika **42** (2006b), 1–36.

[18] Víšek, J. Á.: *Instrumental weighted variables.* Austrian Journal of Statistics **35** (2& 3) (2006c), 379–387.

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

[20] Víšek, J. Á.: *Empirical distribution function under heteroscedasticity.* To appear in Statistics (2010a).

[21] Víšek, J. Á.: *Consistency of the least weighted squares under heteroscedasticity.* Submitted to Kybernetika (2010b).

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

[23] Wiley, D. E.: *The identification problem for structural equation models with unmeasured variables.* In: Structural Equation Models in the Social Sciences, A. S. Goldberger, O. D. Duncan, (eds) 69–83. Seminar Press, New York (1973).

[24] Wooldridge, J. M.: Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA (2001).

[25] Zichová, J.: *On a method of estimating parameters in non-negative ARMA models.* Kybernetika **32** (1996), 409 – 424.

[26] Zvára, K.: Regresní analýza (Regression Analysis – in Czech). Academia, Prague (1989).