Statistics, Volume 45, Issue 5, October 2011, pages 497-508, ISSN: 0233-1888, (článek byl také publikován elektronicky 18. 5. 2010 - viz
http://www.informaworld.com/smpp/title content=g769572243 db=all?stem=2#messages)

EMPIRICAL DISTRIBUTION FUNCTION UNDER HETEROSCEDASTICITY

Jan Ámos Víšek [1]

Abstract

Neglecting heteroscedasticity of error terms may imply a wrong identification of regression model - see Appendix. Employment of (heteroscedasticity resistent) White's estimator of covariance matrix of estimates of regression coefficients may lead to the correct decision about significance of individual explanatory variables under heteroscedasticity. However, White's estimator of covariance matrix was established for LS-regression analysis (in the case when error terms are normally distributed, LS- and ML-analysis coincide and hence then White's estimate of covariance matrix is available for ML-regression analysis, too). To establish White's-type estimate for another estimator of regression coefficients requires Bahadur representation of the estimator in question, under heteroscedasticity of error terms. The derivation of Bahadur representation for other (robust) estimators requires some tools. As the key one proved to be a tight approximation of the empirical distribution function of residuals by the theoretical distribution function of the error terms of the regression model. We need the approximation to be uniform in the argument of distribution function as well as in regression coefficients. The present paper offers this approximation for the situation when the error terms are heteroscedastic.

Key words: Regression, asymptotics of Kolmogorov-Smirnov statistics under heteroscedasticity, robustified White's estimate of covariance matrix.

AMS Classification: Primary 62J05, 62F35, Secondary 62F12

1. INTRODUCTION

The goal of this paper is to establish a generalized version of the famous Kolmogorov-Smirnov result that

$$\sup_{-\infty<x<\infty} \sqrt{n} \left| F^{(n)}(x) - F(x) \right| = \mathcal{O}_p(1) \tag{1}$$

(where $F^{(n)}(x)$ denotes the empirical distribution function), see Smirnov (1939), Doob (1949), Donsker (1952) or Kolmogorov (1950). Research on Kolmogorov-Smirnov statistic was very extensive in various fields (especially its form for multivariate distributions was considered) and many applications performed - see e. g. Fasano, Franceschini (1987), An Hong-Zhi, Cheby Bing (1991), Ustel et al. (1997), Drew et al. (2000) (they were supported by results on power of test - e. g. Jansen (2000)). Many modifications were proposed - see e. g. Mason, Schuenemeyer (1983), Khamis (1993) or Andrews (1997) and even the computational problems were studied - Glen and Premis (2000). Research touched also regression scheme, for the results on logistic model see Jing Qin, Biao Zhang (1997) and some results on testing validity of the shape of regression function (considering Kolmogorov-Smirnov (K-S) statistics as one possible measure of distance) can be found in Dette, Munk (1998). The results closest to our one were presented in Lee, Ching-Zong Wei (1999) where stochastic process based on residuals of the least squares analysis

[1]Research was supported by grant of GA ČR 402/09/0557.

was considered and supremum with respect of argument of distribution function (d. f.) (not with respect to regression coefficients) were analyzed. Regression scheme was also studied in Delgado, Mora(2000) and residual empirical process is employed for testing serial independence of error terms. Attention was also paid to the problem of large deviations of the K-S statistic - Inglot, Ledwina (1990). Later, some results of more general approach via empirical processes (requiring naturally stronger assumptions) started to appear - see e. g. Cabana, Cabana (1997), Vaart, Welner (2000) or Koul (2002). We are going to prove an analogous result to (1) for the regression framework for situation when the error terms are heteroscedastic and supremum is assumed over the argument of d. f. as well as over the regression coefficients, see Assertion 1 below. This result already serves as a tool for the study of the asymptotic behavior of robust estimator of regression coefficients (and the paper also explains how), see Víšek (2009d). It enables us also to establish an estimator of covariance matrix of robust estimators of regression coefficients. Why we need such an estimator of covariance matrix ?

It is easy to see that the estimate of regression coefficients by means of the least squares is under heteroscedasticity still unbiased and (strongly) consistent. From the practical point of view, for the finite sample (i. e. for given data), the picture may be however quite different. As it is demonstrated in the Appendix by a numerical example, neglecting heteroscedasticity may lead to the misleading estimation, in our example it leads to overidentified model. The remedy is to find reliable conclusions about the significance of the explanatory variables. To do it, we need an estimator of covariance matrix (of the estimator of regression coefficients) which is consistent under heteroscedasticity of error terms. Then the corresponding studentization (employing the square roots of diagonal elements of just mentioned matrix) will be appropriate.

For the (ordinary) least squares such estimator was proposed by Halbert White in 1980. To be able to establish an analogous estimator of covariance matrix (of the estimator of regression coefficients which were found by a robust estimator) we need an analogy of Halbert-White-estimator but for given robust estimator of regression coefficients. The key tool for this is just the main result of the present paper. Now, let us explain how the result of the paper is used in robust regression.

Robust regression - the least weighted squares. Several decades the robust regression is one of main topics in robust statistics. A lot of methods of robust identification of regression model was studied, even with a priori selected "level of robustness", selected according to an expected level of data contamination. Nevertheless, there is quite large difference between the conception of the classical regression analysis and the robust one. The former establishes, especially for the *Least Squares* (LS) or *Maximum Likelihood* (ML) a lot of diagnostic tools as Durbin-Watson statistic, (Durbin, Watson (1952)), White's test and White's estimator of covariance matrix of estimates of regression coefficients (White (1980)), Hausman test of specificity (Hausman (1978)), etc. There are also sensitivity studies describing the influence of deleting an observation(s) or deleting an explanatory variable(s) (Chatterjee, Hadi (1988)). Unfortunately, all of these supporting tools and results of sensitivity studies were derived mostly for the LS- or ML-regression analysis. As however reviewers pointed out, today we would be presumably able to derive the key assertion for more estimators, especially for those which are asymptotically normal. What is however true these tools suffer by lack of robustness.

Moreover, some modifications of the LS or ML as instrumental variables (Balestra, Nerlove (1966) or Brundy, Jorgenson (1971)) or modification of LS or ML for categorical and limited dependent variables, (Long (1997)) were established to enable us to identify the model under "nonstandard" situations. Nearly all these lack in robust statistics.

In other words, only a few accompanying means (as diagnostic tools, modifications of basic method or sensitivity studies) for given individual robust regression estimator were established. The present paper prepares the tool for starting to fulfill this gap for the *Least Weighted Squares* (LWS). The definition of LWS, together with a discussion of reasons for the definition, is given in Víšek (2001). Nevertheless, let us recall it briefly.

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

$$Y_i = g\left(X_i, \beta^0\right) + e_i, \quad i = 1, 2, ..., n \tag{2}$$

will be considered (we shall assume $g\left(X_i, \beta^0\right) = 0$ otherwise we would write everywhere $g\left(X_i, \beta\right) -$

$g\left(X_i, \beta^0\right)$). Further, let us denote (for any $\beta \in I\!R^p$) by $r^2_{(h:n)}(\beta)$ the h-th order statistic among the squared residuals $r^2_i(\beta) = (Y_i - g(X_i, \beta))^2$, $i = 1, 2, ..., n$, i.e. we assume

$$r^2_{(1:n)}(\beta) \le r^2_{(2:n)}(\beta) \le \ \dots \ \le r^2_{(n:n)}(\beta). \tag{3}$$

DEFINITION 1 *Let for any $n \in \mathcal{N}$ w_i, $i = 1, 2, ..., n$ be some fixed (non-random) weights. Then the solution of the extremal problem*

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{arg\,min} \sum_{i=1}^{n} w_i r^2_{(i:n)}(\beta) \tag{4}$$

will be called the Least Weighted Squares.

REMARK 1 *We usually generate the weights w_i's by a non-increasing (absolutely continuous) weight function $w(v) : [0, 1] \to [0, 1]$, putting $w_i = w\left(\frac{i-1}{n}\right)$. Then (4) is changed to*

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{arg\,min} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r^2_{(i:n)}(\beta). \tag{5}$$

Notice please that the fact that the weights are prescribed to order statistics of squared residuals rather than to squared residuals directly, leads to the robustness of the estimator.

Let us recall that LWS have some plausible features. Firstly, LWS are, similarly as the *Least Median of Squares* (LMS) and the *Least Trimmed Squares* (LTS) (for definition of both see Hampel et al. (1986)), scale- and regression-equivariant. The same is true for *Instrumental Weighted Variables* (IWV) or a robustified version of the *Total Least Squares* (TLS) (for the definition of former see again Víšek (2006a), for the latter - for the definition of "classical" *Total Least Squares* see e. g. Golub, Van Loan (1980), Paige, Starkoš (2002) or Van Huffel (2004), for a robustified version of TLS see Víšek (2009c)). It is an advantage in comparison with (some) other robust estimators of regression coefficients. E. g. in order to achieve *scale- and regression-equivariance* of M-estimators (of regression coefficients) we have to studentize the residuals by a scale estimator which is *scale-invariant and regression-equivariant*, see Bickel (1975). Such an estimator is not easy to evaluate, see e. g. Jurečková, Sen (1993) where the estimator is based on the regression scores. Of course, it is possible only under homoscedasticity of error terms. If they are heteroscedastic we should studentize the i-th residual by an estimate of the standard deviation of the i-th error term, which however can be estimated only in the case when the observations for the i-th case are repeated.

Secondly, LWS can be used for panel data (if the weight function is strictly positive), in contrast to the *Least Trimmed Squares* or the *Least Median of Squares* (on the other hand, the *Least Trimmed Squares* and the *Least Median of Squares* are special case of LWS). Of course, when LWS employs strictly positive weight function it has zero breakdown point. Nevertheless, for finite data one can reach even by positive weight function a very similar behaviour of the estimator as if it would be equal to zero on a part of the interval $[0, 1]$. So, that by "tailoring" the weight function to given data one can reach a compromise between efficiency of estimation and the "level" of robustness. Finally, they have a reliable algorithm for evaluating a tight approximation to the precise value of estimator. The algorithm is a simple generalization of the algorithm for the LTS, see Víšek (1996, 2000), or a straightforward simplification of the algorithm for the *Instrumental Weighted Variables*, see Víšek (2006a).

Nowadays there are already available some results of building up supporting tools for LWS, see Čížek (2001, 2002), Kalina (2004a, b) Mašíček (2003a, b, 2004), Plát (2004a, b), Víšek (2002, 2004, 2006a, 2007). But we still not have any results for the case when the error terms are heteroscedastic.

Now following Hájek, Šidák (1967) for any $i \in \{1, 2, ..., n\}$ let us define the rank of i-th residual by

$$\pi(\beta, i) = j \in \{1, 2, ..., n\} \qquad \Leftrightarrow \qquad r^2_i(\beta) = r^2_{(j)}(\beta) \tag{6}$$

(notice that $\pi(\beta, i)$ is r.v. since it depends on $X_i(\omega)$'s and $e_i(\omega)$'s). Then we have

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\frac{\pi(\beta, i) - 1}{n}\right) r^2_i(\beta). \tag{7}$$

Now, realize please, that having fixed $\beta \in R^p$ and denoting $|r_i(\beta)| = a_i(\beta)$, the order statistics $a_{(i)}(\beta)$'s and the order statistics of the squared residuals $r^2_{(i)}(\beta)$'s assign to given fix observation the same rank, i. e. if the squared residual of given fix observation is on the ℓ-th position (say) in the sequence

$$r^2_{(1)}(\beta) \leq r^2_{(2)}(\beta) \leq ... r^2_{(n)}(\beta), \tag{8}$$

then the absolute value of residual of the same observation is in the sequence

$$a_{(1)}(\beta) \leq a_{(2)}(\beta) \leq ... a_{(n)}(\beta) \tag{9}$$

also on the ℓ-th position. Further, let us denote for any $\beta \in I\!R^p$ and any $v \in I\!R$ the empirical d. f. of the absolute value of residuals $|r_i(\beta)| = |Y_i - g(X_i, \beta)|$ by $F_n^{(\beta)}(v)$, i. e.

$$F_\beta^{(n)}(v) = \frac{1}{n}\sum_{i=1}^n I\{|r_i(\beta)| < v\} = \frac{1}{n}\sum_{i=1}^n I\{|Y_i - g(X_i, \beta)| < v\} \tag{10}$$

where $I\{A\}$ denotes the indicator of the set A. Now, let us realize that the empirical d. f. $F_\beta^{(n)}(v)$ has the first jump at $a_{(1)}(\beta)$ (but due to the sharp inequality in (10), $F_\beta^{(n)}(a_{(1)}(\beta)) = 0$), the second jump at $a_{(2)}(\beta)$ (but $F_\beta^{(n)}(a_{(2)}(\beta)) = \frac{1}{n}$), the third at $a_{(3)}(\beta)$ (but $F_\beta^{(n)}(a_{(3)}(\beta)) = \frac{2}{n}$), etc. Hence it has the $\pi(\beta, i)$-th jump at $a_{(\pi(\beta,i))}(\beta)$, i. e.

$$F_\beta^{(n)}(a_{(\pi(\beta,i))}(\beta)) = F_\beta^{(n)}(|r_i(\beta)|) = \frac{\pi(\beta, i) - 1}{n}. \tag{11}$$

So, we can rewrite (7) as

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^n w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) r_i^2(\beta). \tag{12}$$

(12) indicates that we need an approximation to the empirical d. f. $F_\beta^{(n)}(v)$ which is uniform in $v \in R$ as well as in $\beta \in R^p$ (for the homoscedastic situation such approximation was studied in Víšek (2006b)). Then, having at hand a uniform in $v \in I\!R$ and $\beta \in I\!R^p$ approximation of empirical d. f. $F_\beta^{(n)}(v)$ by the "mean" d. f.

$$\overline{F}_{n,\beta}(v) = \frac{1}{n}\sum_{i=1}^n F_{i,\beta}(v) \tag{13}$$

where

$$F_{i,\beta}(v) = P(|Y_i - g(X_i, \beta)| < v) = P\left(\left|e_i - g(X_i, \beta) + g\left(X_i, \beta^0\right)\right| < v\right) \tag{14}$$

(remember that e_i's have different variances σ_i^2), we can perform technicalities leading to asymptotic representation of the estimator in question, and it, in turn, allows to "generalize" classical diagnostic tools for LWS (or other modifications of LS- or ML-estimators, as the *Instrumental Weighted Variables*, see Víšek (2009a)) even under heteroscedasticity. The same is true about the *Total Least Squares*, the research is under progress.

At the first glance, it seems that the same can be achieved by employing results on *Weighted empirical processes*, see e. g. Koul (2002). Leaving aside that present approach is much more transparent and applicable in wider range of situations - see discussion bellow devoted to studentization of M-estimators, it would be as using a sledgehammer to crack a nut.

Finally, throughout the paper we shall assume:

Conditions $\mathcal{C}$ *The sequence* $\{(X_i', e_i)'\}_{i=1}^\infty$ *is sequence of independent* $(p+1)$*-dimensional random variables with* $X_{i1} = 1$ *for all* $i = 1, 2, ...$ *(i. e. the model with intercept is considered). The random vectors* $(X_{i2}, X_{i3}, ..., X_{ip}, e_i)'$ *are distributed according to distribution functions* $\{F(x, v\sigma_i)\}_{i=1}^\infty, x \in I\!R^{p-1}, v \in I\!R$*, i. e.*

$$P(X_i < x, e_i < v) = F(x, v\sigma_i)$$

where $F(x, v)$ *is a parent d. f. . Moreover,* $I\!E(e_i|X_i) = 0$ *and* $\mathrm{var}(e_i|X_i) = \sigma_i^2$ *with* $0 < \sigma_i^2 < \infty$.

2. APPROXIMATION OF EMPIRICAL D. F. UNDER HETEROSCEDASTICITY

Let us consider a sequence of i.i.d. r. v.'s $\{V_n\}_{n=1}^{\infty}$ distributed according to d. f. $G(v)$. Denote by $G^{(n)}(v)$ empirical d. f., i. e.

$$G^{(n)}(v) = \frac{1}{n}\sum_{i=1}^{n} I\{V_i < v\}.$$

(let' repeat that by $I\{V_i < v\}$ we have denoted the indicator of the set $\{\omega \in \Omega : V_i(\omega) < v\}$). As we have already mentioned, Smirnov (1939), using results by Kolmogorov (1933) (see also Doob (1949), Donsker (1952), or Kolmogorov (1950)), established the asymptotic d. f. of the statistic

$$D^{(n)} = \sup_{-\infty < v < \infty} \sqrt{n}\left|G^{(n)}(v) - G(v)\right|. \tag{15}$$

Considering regression framework with i.i.d. error terms (i. e. assuming that $F_\beta(v) = P\left(|Y_n - X_n'\beta| < v\right)$ is the same for all $n \in \mathcal{N}$) and employing somewhat generalized steps of Smirnov, we have derived (Víšek (2006b)):

ASSERTION 1 *Let $B(t), 0 \le t \le 1$ be normalized Brownian motion. Then under* **Conditions** $\mathcal{C}$ *with $\sigma_i^2 = \sigma^2 \in (0,\infty)$, we have*

$$\sup_{v\in \mathbb{R}^+} \sup_{\beta \in \mathbb{R}^p} \sqrt{n}\left|F_\beta^{(n)}(v) - F_\beta(v)\right| \underset{\mathcal{D}}{\longrightarrow} \sup_{0\le t\le 1} |B(t) - t\cdot B(1)|. \tag{16}$$

(for $F_\beta^{(n)}(v)$ see (10)). Now, we are going to study $D^{(n)}$ without restriction that all variances of error terms are the same. Unfortunately, the technique which was used in previous papers cannot be used here. So we employed the Skorohod embedding (in the sense as it was used in Portnoy (1983) or Jurečková (1984)) for which we will need the following three assertions.

ASSERTION 2 *(Štěpán (1987), page 420, VII.2.8) Let a and b be positive numbers. Further let ξ be a random variable such that $P(\xi = -a) = \pi$ and $P(\xi = b) = 1 - \pi$ (for a $\pi \in (0,1)$) and $\mathbb{E}\xi = 0$. Moreover let τ be the time for the Wiener process $W(s)$ to exit the interval $(-a, b)$. Then*

$$\xi =_{\mathcal{D}} W(\tau)$$

where "$=_{\mathcal{D}}$" denotes the equality of distributions of the corresponding random variables. Moreover, $\mathbb{E}\tau = a\cdot b = var\,\xi$.

REMARK 2 *Since the book by Štěpán (1987) is in Czech language we refer also to Breiman (1968) where however this assertion is not isolated. Nevertheless, the assertion can be found directly in the first lines of the proof of Proposition 13.7 (page 277) of Breiman's book. (See also Theorem 13.6 on the page 276.) The next assertion can be found, in a bit modified form also in Breiman's book, Proposition 12.20 (page 258).*

ASSERTION 3 *(Štěpán (1987), page 409, VII.1.6) Let a and b be positive numbers. Then*

$$P\left(\max_{0\le t\le b} |W(t)| > a\right) \le 2\cdot P\left(|W(b)| > a\right).$$

DEFINITION 2 *Let S be a subset of a separable metric space. The stochastic process $V = (V(s),\ s \in S)$ is called separable if there is a countable dense subset $T \subset S$ (i.e. T is countable and dense in S) such that for any $(\omega, s) \in \Omega \times S$ there is a sequence such that*

$$s_n \in T, \quad \lim_{n\to\infty} s_n = s \quad \text{and} \quad \lim_{n\to\infty} V(\omega, s_n) = V(\omega, s).$$

ASSERTION 4 *(Štěpán (1987), page 85, I.10.4) Let $V = (V(s),\ s \in S)$ be a separable stochastic process defined on the probability space $(\Omega, \mathcal{A}, P)$. Moreover, let $G \subset S$ be open and denote by $k(G)$ the set of all finite subsets of G. Then for any close set $K \subset I\!R^p$ we have*

$$\{\omega \in \Omega : V(s) \in K, s \in G\} \in \mathcal{A}$$

and

$$P\left(\{\omega \in \Omega : V(s) \in K, s \in G\}\right) = \inf_{J \in k(G)} P\left(\{\omega \in \Omega : V(s) \in K, s \in J\}\right).$$

Proof: Since the book by Štěpán is in Czech language and the proof is short, we will give it. Let T be countable dense subset of S. Then we have

$$\{\omega \in \Omega : V(s) \in K, s \in G\} = \{\omega \in \Omega : V(s) \in K, s \in G \cap T\}$$

and

$$P\left(\{\omega \in \Omega : V(s) \in K, s \in G\}\right) \leq \inf_{J \in k(G)} P\left(\{\omega \in \Omega : V(s) \in K, s \in J\}\right)$$

$$\leq \inf_{J \in k(G \cap S)} P\left(\{\omega \in \Omega : V(s) \in K, s \in J\}\right) = P\left(\{\omega \in \Omega : V(s) \in K, s \in G \cap S\}\right)$$

$$= P\left(\{\omega \in \Omega : V(s) \in K, s \in G\}\right). \qquad \square$$

Now, we are going to give the main result of paper.

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

$$P\left(\left\{\omega \in \Omega : \sup_{v \in R^+} \sup_{\beta \in I\!R^p} \sqrt{n}\left|F_\beta^{(n)}(v) - \overline{F}_{n,\beta}(v)\right| < K_\varepsilon\right\}\right) > 1 - \varepsilon. \tag{17}$$

Proof: Fix $\varepsilon > 0$ and put $K_\varepsilon = \sqrt{\frac{8}{\varepsilon}} + 1$. Recalling that (see (10))

$$F_\beta^{(n)}(v) = \frac{1}{n}\sum_{i=1}^{n} I\left\{|Y_i - g\left(X_i, \beta\right)| < v\right\}, \tag{18}$$

let us put

$$b_i(v, \beta) = I\left\{\omega \in \Omega : |Y_i - g\left(X_i, \beta\right)| < v\right\} = I\left\{\omega \in \Omega : -v < Y_i - g\left(X_i, \beta\right) < v\right\}. \tag{19}$$

Further put

$$\xi_i(v, \beta) = b_i(v, \beta) - I\!E b_i(v, \beta) \tag{20}$$

and denote

$$p_i(v, \beta) = I\!E b_i(v, \beta) = P\left(b_i(v, \beta) = 1\right) = F_{i,\beta}(v) \tag{21}$$

(see (14)). Then $\{\xi_i(v, \beta)\}_{i=1}^{\infty}$, for any $v \in I\!R^+$ and any $\beta \in I\!R^p$, is a sequence of independently distributed r.v.'s. Finally, (18), (19) and (21) yield

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i(v, \beta) = F_\beta^{(n)}(v) - \overline{F}_{n,\beta}(v),$$

i. e.

$$\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n} \xi_i(v, \beta)\right| = \sqrt{n}\left|F_\beta^{(n)}(v) - \overline{F}_{n,\beta}(v)\right|.$$

Moreover

$$P\Big(\xi_i(v, \beta) = 1 - p_i(v, \beta)\Big) = p_i(v, \beta)$$

and

$$P\left(\xi_i(v,\beta) = -p_i(v,\beta)\right) = 1 - p_i(v,\beta).$$

Now, we are going to employ Assertion 2. We have already mentioned that $\{\xi_i(v,\beta)\}_{i=1}^{\infty}$ is a sequence of independently distributed r.v.'s. Let us denote by $\{W_i(s)\}_{i=1}^{\infty}$ the sequence of independent Wiener processes (we may assume e. g. that each of them is defined on "an own probability space", say $\{(\Omega_i, \mathcal{A}_i, P_i)\}_{i=1}^{\infty}$ and then consider the product space $(\Omega, \mathcal{A}, P)$ in the same way as it is done in the proof of Daniell-Kolmogorov theorem, see e. g. Tucker (1967)) and, following Portnoy (1983) or Jurečková (1984) or Jurečková, Sen (1989), let us define $\tau_i(v,\beta)$ to be the time for the Wiener process $W_i(s)$ to exit the interval $(-p_i(v,\beta), 1-p_i(v,\beta))$ (please keep in mind that $\tau_i(v,\beta)$ is r.v., i.e. $\tau_i(v,\beta) = \tau_i(v,\beta,\omega)$). Then $\xi_i(v,\beta) =_{\mathcal{D}} W_i(\tau_i(v,\beta))$ and hence for any $\beta \in I\!\!R^p$

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \xi_i(v,\beta) =_{\mathcal{D}} n^{-\frac{1}{2}} \sum_{i=1}^{n} W_i(\tau_i(v,\beta)) =_{\mathcal{D}} W_1\left(n^{-1} \sum_{i=1}^{n} \tau_i(v,\beta)\right) \tag{22}$$

where the last equality follows from the properties of the Wiener process. Further, let us define U_i to be the time for the Wiener process $W_i(s)$ to exit interval $(-1,1)$. Due to the fact that for all $i = 1, 2, ..., n$ for any $v \in I\!\!R^+$ and any $\beta \in I\!\!R^p$

$$p_i(v,\beta) \le 1 \quad \text{and} \quad 1 - p_i(v,\beta) \le 1, \qquad \text{i. e.} \qquad \left(-p_i(v,\beta)), 1 - p_i(v,\beta)\right) \subset (-1, 1)\,,$$

we conclude that for any $v \in I\!\!R^+$, any $\beta \in I\!\!R^p$ and any $\omega \in \Omega$

$$\tau_i(v,\beta) \le U_i$$

and hence (again for any $\omega \in \Omega$)

$$n^{-1} \sum_{i=1}^{n} \tau_i(v,\beta) \le n^{-1} \sum_{i=1}^{n} U_i. \tag{23}$$

Of course, $\{U_i\}_{i=1}^{\infty}$ is the sequence of i.i.d r.v.'s and due to Assertion 2 we have

$$I\!\!E U_i = 1,$$

so, employing the law of large numbers, we can find n_1 so that for all $n > n_1$ and for

$$B_n = \left\{\omega \in \Omega \;:\; n^{-1} \sum_{i=1}^{n} U_i \le 2\right\}$$

we have

$$P\left(B_n\right) \ge 1 - \frac{\varepsilon}{2}. \tag{24}$$

Let us consider $n > n_1$ and a fix $\omega_0 \in B_n$ and let us realize that for any $v \in I\!\!R^+$ and any $\beta \in I\!\!R^p$ the left hand side of (23), i. e. $n^{-1} \sum_{i=1}^{n} \tau_i(v,\beta) = n^{-1} \sum_{i=1}^{n} \tau_i(v,\beta,\omega_0)$, is not larger than $n^{-1} \sum_{i=1}^{n} U_i = n^{-1} \sum_{i=1}^{n} U_i(\omega_0) \in [0,2]$. So for our fix ω_0, we have

$$\left\{t \in I\!\!R : t = n^{-1} \sum_{i=1}^{n} \tau_i(v,\beta,\omega_0), v \in I\!\!R^+, \beta \in I\!\!R^p\right\} \subset \left\{t \in I\!\!R : 0 \le t \le n^{-1} \sum_{i=1}^{n} U_i(\omega_0)\right\}.$$

It means that

$$\sup_{v \in \boldsymbol{R}^+} \sup_{\beta \in \boldsymbol{R}^p} W\left(n^{-1} \sum_{i=1}^{n} \tau_i(v,\beta,\omega_0)\right) \le \sup_{0 \le t \le n^{-1} \sum_{i=1}^{n} U_i(\omega_0)} \left|W_1\left(t,\omega_0\right)\right|. \tag{25}$$

So, we arrived at: We have two processes which are equivalent in distribution, i. e. $\sum_{i=1}^n \xi_i(v,\beta,\omega) =_{\mathcal{D}} W_1\left(n^{-1}\sum_{i=1}^n \tau_i(v,\beta,\omega)\right)$ with the same index sets, $v \in I\!R, \beta \in I\!R^p$ (see (22)), both of them are separable. Then employing Assertion 4, we obtain

$$n^{-\frac{1}{2}} \sup_{v\in \boldsymbol{R}^+} \sup_{\beta \in \boldsymbol{R}^p} \left| \sum_{i=1}^n \xi_i(v,\beta,\omega_0) \right| =_{\mathcal{D}} \sup_{v\in \boldsymbol{R}^+} \sup_{\beta \in \boldsymbol{R}^p} \left| W_1\left(n^{-1} \sum_{i=1}^n \tau_i(v,\beta,\omega_0) \right) \right|$$

and due to (25)

$$n^{-\frac{1}{2}} \sup_{v\in \boldsymbol{R}^+} \sup_{\beta \in \boldsymbol{R}^p} \left| \sum_{i=1}^n \xi_i(v,\beta,\omega_0) \right| \le \sup_{0\le t\le n^{-1}\sum_{i=1}^n U_i(\omega_0)} \left| W_1\left(t,\omega_0\right) \right| .$$

In other words, for any $n > n_1$ and any $\omega \in B_n$

$$n^{-\frac{1}{2}} \sup_{v\in \boldsymbol{R}^+} \sup_{\beta \in \boldsymbol{R}^p} \left| \sum_{i=1}^n \xi_i(v,\beta) \right| \le \sup_{0\le t\le n^{-1}\sum_{i=1}^n U_i} \left| W_1\left(t\right) \right| . \tag{26}$$

Further, employing (26), we arrive at

$$\begin{aligned} & P\left(\left\{ \omega\in\Omega \ : \ n^{-\frac{1}{2}} \sup_{v\in \boldsymbol{R}^+} \sup_{\beta \in \boldsymbol{R}^p} \left| \sum_{i=1}^n \xi_i(v,\beta) \right| > K \right\} \right) \\ \le\ & P\left(\left\{ \omega\in\Omega \ : \ n^{-\frac{1}{2}} \sup_{v\in \boldsymbol{R}^+} \sup_{\beta \in \boldsymbol{R}^p} \left| \sum_{i=1}^n \xi_i(v,\beta) \right| > K \right\} \cap \left\{ \omega\in\Omega \ : \ n^{-1}\sum_{i=1}^n U_i > 2 \right\} \right) \\ +\ & P\left(\left\{ \omega\in\Omega \ : \ \sup_{0\le t\le n^{-1}\sum_{i=1}^n U_i} \left| W_1\left(t\right) \right| > K \right\} \cap \left\{ \omega\in\Omega \ : \ n^{-1}\sum_{i=1}^n U_i \le 2 \right\} \right) \\ \le\ & P\left(\left\{ \omega\in\Omega \ : \ n^{-1}\sum_{i=1}^n U_i > 2 \right\} \right) + P\left(\left\{ \omega\in\Omega \ : \ \sup_{0\le t\le 2} \left| W_1\left(t\right) \right| > K \right\} \right). \end{aligned} \tag{27}$$

Now, utilizing Assertion 3, we obtain

$$P\left(\sup_{0\le t\le 2} |W_1(t)| > K \right) \le 2\cdot P\left(|W_1(2)| > K \right). \tag{28}$$

Further, recalling the fact that $\text{var}\{W(2)\} = 2$ and using Chebyshev inequality, we arrive at

$$2\cdot P\left(\left| W_1(2) \right| > K \right) \le 4\cdot \frac{1}{K^2} = \frac{\varepsilon}{2}. \tag{29}$$

Finally, (24), (27), (28) and (29) imply

$$P\left(n^{-\frac{1}{2}} \sup_{v\in \boldsymbol{R}^+,\ \beta \in \boldsymbol{R}^p} \left| \sum_{i=1}^n \xi_i(v,\beta) \right| > K \right) \le \varepsilon$$

which concludes the proof. ■

REMARK 3 *It seems that the result of Lemma 1 can be generalized e. g. for the situation when the sequence $\{(X_i', e_i)'\}_{i=-\infty}^{\infty}$ is AR vector process, just applying Cochrane-Orcutt transformation (Cochrane, Orcutt (1949)). Similarly for other structure of dependence of r. v.'s which allows a transformation to "back" to independence we can achieve the same.*

Appendix

Neglecting heteroscedasticity - an example. We looked for a model of export from the Czech republic into EU for the period from 1993 to 2001. The Czech economy was divided into 61 industries and following (response and explanatory) variables were available:

X	-	export,	M	-	import,	PE	-	export prices,	PM	- import prices,
VA	-	value added,	K	-	capital,	L	-	labor,	DE	- debts,

FDI - foreigner direct investment, $GDPeu$ - gross domestic product in EU

After some research we arrived at the model ($t = 1994, 1995, ..., 2001$, $i = 1, 2, ..., 61$; p-values are given as subindices of the coefficients)

$$ln(X_{i,t}) = 9.6_{(.104)} + 0.83_{(.000)} \cdot ln(X_{i,t-1}) - 0.16_{(.007)} \cdot ln(PE_{i,t}) + 0.2_{(.001)} \cdot ln(PE_{i,t-1}) + 0.34_{(.000)} \cdot ln(VA_{i,t})$$
$$-0.23_{(.004)} \cdot ln(VA_{i,t-1}) - 0.63_{(.000)} \cdot ln(K_{i,t}/L_{i,t}) + 0.52_{(.001)} \cdot ln(K_{i,t-1}/L_{i,t-1}) + 0.3_{(.016)} \cdot ln(DE_{i,t}/VA_{i,t})$$
$$-0.3_{(.015)} \cdot ln(DE_{t-1}/VA_{t-1}) + 0.15_{(.009)} \cdot ln(FDI_{i,t}) - 0.15_{(.007)} \cdot ln(FDI_{i,t-1}) + 1.13_{(.045)} \cdot ln(GDPeu_{i,t})$$
$$-1.97_{(.002)} \cdot ln(GDPeu_{i,t-1}) \quad (30)$$

which seemed rather complicated. Nevertheless, it was well determined with approximately normally distributed residuals and, as the p-values indicate, all explanatory variables (except of intercept) were significant. Nevertheless, White's test of homoscedasticity gave value 244.1 with corresponding p-value equal to 0.0000. Clearly the error terms in model were heteroscedastic. So, for appropriate judgement about the significance of explanatory variables it was necessary to employ White's estimator of covariance matrix of the estimates of regression coefficients which is given as

$$\hat{\Sigma} = \frac{1}{8 \cdot 61} \sum_{t=1994}^{2001} \sum_{i=1}^{61} r_{i,t}^2(\hat{\beta}^{(OLS,n)}) \cdot X_{i,t} \cdot X_{i,t}^{\mathrm{T}}$$

where $r_{i,t}(\beta) = Y_{i,t} - X_{i,t}^{\mathrm{T}}\beta$, for details see White (1980). We obtained then (notice please the changes of the values of subindices of the coefficients)

$$ln(X_{i,t}) = 9.6_{(.200)} + 0.83_{(.000)} \cdot ln(X_{i,t-1}) - 0.16_{(.127)} \cdot ln(PE_{i,t}) + 0.2_{(.062)} \cdot ln(PE_{i,t-1}) + 0.34_{(.098)} \cdot ln(VA_{i,t})$$
$$-0.23_{(.235)} \cdot ln(VA_{i,t-1}) - 0.63_{(.016)} \cdot ln(K_{i,t}/L_{i,t}) + 0.52_{(.087)} \cdot ln(K_{i,t-1}/L_{i,t-1}) + 0.3_{(.312)} \cdot ln(DE_{i,t}/VA_{i,t})$$
$$-0.3_{(.302)} \cdot ln(DE_{t-1}/VA_{t-1}) + 0.15_{(.300)} \cdot ln(FDI_{i,t}) - 0.15_{(.222)} \cdot ln(FDI_{i,t-1}) + 1.13_{(.305)} \cdot ln(GDPeu_{i,t})$$
$$-1.97_{(.049)} \cdot ln(GDPeu_{i,t-1}). \quad (31)$$

Excluding successively insignificant explanatory variables, we finally arrived at the model

$$ln(X_{i,t}) = 9.6_{(.200)} + 0.80_{(.000)} \cdot ln(X_{i,t-1}) + 0.15_{(.000)} \cdot ln(VA_{i,t})$$
$$-0.21_{(.001)} \cdot ln(K_{i,t}/L_{i,t}) + 1.90_{(.016)} \cdot ln(GDPeu_{i,t}) - 2.54_{(.001)} \cdot ln(GDPeu_{i,t-1}) \quad (32)$$

which is much simpler than (30). Of course, neglecting heteroscedasticity leads to the estimates of regression coefficients which are still unbiased but which can have rather large variances and hence they are less reliable. Moreover, due to the wrong p-values, the model may include some insignificant explanatory variables which can "contest" with those really significant. The consequence may be that they in fact "divide" the influence on the response variable, so that (all) estimates of regression coefficients may be incorrect. In other words, when applying White's estimate of covariance matrix of the estimates of regression coefficients and correcting significance judgement, excluding successively the "most insignificant" explanatory variable, we may meet with situation when the estimates of other coefficients may dramatically change. So, the conclusion is that the heteroscedasticity of error terms should be taken into account seriously.

References

[1] An Hong-Zhi, Cheby Bing (1991): A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *International statistical review 59, 287 - 307.*

[2] Andrews, D. W. K. (1997): A conditional Kolmogorov test. *Econometrica 65, 1097 - 1128.*

[3] Balestra, P., M. Nerlove (1966): Pooling cross-section and time-series data in the estimation of a dynamic model. *Econometrica 34, 585 - 612.*

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

[5] Breiman, L. (1968): *Probability.* Addison-Wesley Publishing Company, London.

[6] Brundy, J., D. Jorgenson (1971): Consistent and efficient estimation of system of simultaneous equations by means of instrumental variables. *Review of Economics and Statistics 53, 207 - 224.*

[7] Cabana, A., E. M. Cabana (1997): Transformed empirical processes and modified Kolmogorov-Smirnov tests fo multivariate distribution. *Annals of Statistics 25, 2388 - 2409.*

[8] Chatterjee, S., A. S. Hadi (1988): *Sensitivity Analysis in Linear Regression.* J. Wiley *&* Sons, New York.

[9] Čížek, P. (2001): Robust estimation in nonlinear regression models - sensitivity analysis. *PhD. disertation, Center of Economic Research and Graduate Education, Charles University.*

[10] Čížek, P. (2002): Robust estimation with discrete explanatory variables. *COMPSTAT 2002, Berlin, eds. Klinke, S., Ahrend, P., Richter, L., Springer, 509 - 514.*

[11] Cochrane, D., G. H. Orcutt (1949): Application of least squares regression to relationhips containing autocorrelated error terms. *Journal of the American Statistical Association, 44, 32 - 61.*

[12] Delgado, M. A., Mora, J. (2000): A nonparametric test for serial independence of regression errors. *Biometrika 87, 228 - 234.*

[13] Dette, H., Munk, A. (1998): Validation of regression model. *The Annals of Statistics 26, 778 - 800.*

[14] Donsker, M. (1952): Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Annals of Mathematical Statistics 23, 277 - 281.*

[15] Doob, J. L. (1949): Heuristic approach to the Kolmogorov-Smirnov theorem. *Annals of Mathematical Statistics 20, 393 - 403.*

[16] Drew, J. H., A. G. Glen, L. M. Premis (2000): Computing the cumulative distribution function of the Kolmogorov-Smirnov statistic. *Computational Statistics and Data Analysis 34, 1 - 15.*

[17] Durbin, J., G. S. Watson (1952): Testing for serial correlation in least squares regression. I. *Biometrika 37, 409-428.*

[18] Fasano, G., A. Franceschini (1987): A multidimensional version of the Kolmogorov-Smirnov test. *Royal Astronomical Society, Monthly Notice 225, 155 - 170.*

[19] Golub, G.H., C.F. Van Loan (1980): An analysis of the total least squares problem. *Siam J. Numer. Anal., 17, 883 -893.*

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

[21] 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 *&* Son.

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

[23] Inglot, T., T. Ledwina (1990): On probabilities of excessive deviations for Kolmogorov-Smirnov, Cramer-von Mises andchi-square statistics. *Annals of Statistics 18, 1491 - 1495.*

[24] Jansen, A. (2000): Global power function of goodness of fit tests. *Annals of Statistics 28, 239 - 253.*

[25] Jing Qin, Biao Zhang (1997): A goodness-of-fit test for logistic regression model based on case-control data. *Biometrica 84, 609 - 618.*

[26] Jurečková, J. (1984): Regression quantiles and trimmed least squares estimator under a general design. *Kybernetika 20, 345–357.*

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

[28] Jurečková, J., P. K. Sen (1993): Regression rank scores scale statistics and studentization in linear models. *Proceedings of the Fifth Prague Symposium on Asymptotic Statistics, Physica Verlag, 111-121.*

[29] Kalina, J. (2004a): Autocorrelated disturbances of robust regression. *Proceedings of European Young Statisticians Meeting 2003 Ovronnaz, Switzerland, 65-72.*

[30] Kalina, J. (2004b): Durbin-Watson Test for Least Weighted Squares. *Proceedings of COMPSTAT 2004, Physica-Verlag/Springer, ed. Antoch, J., 1287 - 1294.*

[31] Khamis, H. J. (1993): A comparative study of the α-corrected Kolmogorov-Smirnov test. *Journal of Applied Statistics 20, 401 - 421.*

[32] Kolmogorov, A. (1933): Sulla determinazione empirica di una legge di distribuzione. *Giorn. dell'Instituto Italiano degli Attuari 4, 83 - 91.*

[33] Kolmogorov, A. (1950): *Foundations of Probability.* (English translation) Chelsea Publishing Co., New York.

[34] Koul, H. L. (2002): *Weighted Dynamical Processes in Dynamic Nonlinear Models.* Second edition, Springer, Lecture Notes in Statistics 166.

[35] Lee, S., Ching-Zong Wei (1999): On residual empirical process of stochastic regression models with applications to time series. *The Annals of Statistics 27, 237 - 261.*

[36] Long, J. S. (1997): *Regression Models for Categorical and Limited Dependent Variables.* SAGE publications, London.

[37] Mason, M. D., J. H. Schuenemeyer (1983): A modified Kolmogorov-Smirnov test sensitive to tail alternatives. *Annals of Statistics 11, 933 - 946.*

[38] Mašíček, L. (2003a): Diagnostika a sensitivita robustních modelů (Diagnostics and sensitivity of robust models), PhD. thesis, Faculty of Mathematics and Physics, Charles University.

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

[40] Mašíček, L. (2004): Behaviour of the Least Weighted Squares Estimator for Data with Correlated Regressors. *Proceedings of COMPSTAT 2004, Physica-Verlag/Springer, ed. Antoch, J., 1463 - 1470.*

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

[42] Paige, C.C., Z. Starkoš (2002): Scaled total least squ ares fundamentals. *Numer. Math. 91, 117 - 146.*

[43] Plát, P. (2004a): Modifikace Whiteova testu pro nejmen ven ctverce. (Modification of White's test for the least weighted squares, in Czech.) *ROBUST 2004, Union of the Czech Mathematicians and Physicists 57-552-04, ed. Antoch, J., Dohnal, G., 291 - 298.*

[44] Plát, P. (2004b): The Least Weighted Squares Estimator. *Proceedings of COMPSTAT 2004, Physica-Verlag/Springer, ed. Antoch, J., 1653 - 1660.*

[45] Portnoy, S. (1983): Tightness of the sequence of empiric c.d.f. processes defined from regression fractiles. In *Robust and Nonlinear Time-Series Analysis, eds. J. Franke, W. Hårdle, D. Martin, 231 - 246, Springer-Verlag, New York.*

[46] Smirnov, N. (1939): On the estimation of discrepancy between empirical curves of distribution for two independent samples. *Bulletin of Mathematics, University of Moscow 2, 3 - 14.*

[47] Štěpán, J. (1987): *Teorie pravděpodobnosti (Theory of Probability, in Czech).* Academia, Praha.

[48] Tucker, H. (1967): *A Graduate Course in Probability.* Academic Press, London.

[49] Ustel, A., D. Pena, R. Zamar (1997): A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics and probability letters 35, 251-259.*

[50] Vaart, A. W. van der, J. A. Welner (2000): *Weak Convergence and Empirical Processes.* Springer-Verlag, New York, second edition.

[51] 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, Physica-Verlag, Springer, 539 - 555.*

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

[53] Víšek, J.Á.(2000): On the diversity of estimates. *Computational Statistics and Data Analysis 34, 67 - 89.*

[54] Víšek, J. Á. (2001): Regression with high breakdown point. *ROBUST 2000, 324 - 356, ed. Antoch, J. & Dohnal, G., The Union of the Czech Mathematicians and Physicists and the Czech Statistical Society 2001.*

[55] Víšek, J. Á. (2002): White test for the least weigthed squares. *COMPSTAT 2002, Berlin, Proceedings of the Conference CompStat 2002 - Short Communications and Poster (CD), ISBN 3-00-009819-4, eds. Klinke, S., Ahrend, P., Richter, L.*

[56] Víšek, J. Á. (2004): Robustifying instrumental variables. *Proceedings of COMPSTAT'2004. Physica-Verlag/Springer, ed. Antoch, J., 1947 - 1954.*

[57] Víšek, J. Á. (2006a): Instrumental Weighted Variables- algorithm. *Proceedings of the COMPSTAT 2006, Rome (28.8.- 1.9.2006), eds. Rizzi, A. & Vichi, M., Physica-Verlag (Springer Company) Heidelberg 2006, 777-786.*

[58] Víšek, J. Á. (2006b): Kolmogorov-Smirnov statistics in multiple regression. *Proceedings of the ROBUST 2006, eds. Antoch, J. & Dohnal, G. The Union of the Czech Mathematicians and Physicists 57-554-06, 367-374.*

[59] Víšek, J. Á.(2009a): Consistency of the instrumental weighted variables. *Annals of the Institute of Statistical Mathematics, (2009) 61, 543 - 578.*

[60] Víšek, J. Á. (2009b): $\sqrt{n}$-consistency of the instrumental weighted variables. Asymptotic representation of the instrumental weighted variables. *Preprints.*

[61] Víšek, J. Á. (2009c): Robustifying total least squares. *Preprint* presented on the 2nd Workshop of the ERCIM Working Group on COMPUTING *&* STATISTICS, Limasol, October, 2009.

[62] Víšek, J. Á. (2009d): Consistency of the least weighted squares under heteroscedasticity. Submitted to the *Kybernetika.*

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

Affiliation:
Jan Ámos Víšek

Department of Macroeconomics and Econometrics, Institute of Economic Studies, Faculty of Social Sciences, Charles University

&

Department of Stochastic Informatics, Institute of Information Theory and Automation, Academy of Sciences of Czech Republic
Mailing address:
Opletalova ulice 26, CZ - 110 01 Prague 1, Czech Republic,
e-mail: visek@fsv.cuni.cz