ROBUST'98, 195 – 224

ROBUST INSTRUMENTS

Jan Ámos VÍŠEK[1]

IES FSV UK, Prague

Abstract. It is not well-known even among statisticians that in the case when the explanatory variables and random fluctuations in regression model are not independent the least squares estimate is generally biased and always inconsistent. The present paper explains the reason why it is so and offers examples of such situations. Then the method of instrumental variables which is able to cope with such situations, is recalled.

On the other hand, it would be naive to assume that the situation when the assumption of independence of explanatory variables and random fluctuations does not hold is disjoint with the situation when the data are contaminated or "at least" not normally distributed. In fact, much more realistic is to expect the opposite, i. e. that not only the assumption of independence of regressors and fluctuations is broken but also assumption of normality of fluctuations will be doubtful. At such situations it is reasonable to use a robust method but such which is able to cope with just mentioned dependence.

And this is the topic the paper is focussed on, namely to study robust version of the method of instrumental variables. Paper offers a proposal of corresponding estimator and proves its $\sqrt{n}$-consistency, Bahadur representation and asymptotic normality (under broken assumption of independence)[2].

Резюме: Не очень знакомо даже между статистиками, что когда существует взаимная зависимость между независимими переменными и случайными флуктуациями в регресснной модели, тогда оценка наименших квадратов регресионных коэффициентов смещенная и тоже потеряет консистенцию. Статья показывает почему это так. Потом напоминает прикладов таких ситуаций метод инструменталйных переменных.

С другой стороны было бы нереалистически предполагать, что ситуации, когда независимые переменные и случайные флуктуации взаимно зависимы, непересекаются со ситуациями когда наблюдения контаминованы или по крайней мере нет нормально распределены. Более реалистическое предполагать что нет только одно предположение, но оба предположения неверны. В таком случае лучше место оценки наименьших квадратов использовать робастную оценку, но такую, которая консистентная даже при зависимости между флуктуациями и независимими переменными (о несмещенности нельзя говорить, так как у большинства робастных оценок нельзя добиться смещенности).

Єтой проблеме посвещена эта статья, что значит, что она рассматривает робастную версию метода инструментальных переменных. Статья вносит предложение отвечающей оценки и доказывает ее $\sqrt{n}$-консистенцию, выводит асымптотическую репресентацию оценки и ее асымптотическое распределение (при нарушении предположения о независимости между флуктуациями и независимими переменными).

[1]Research was supported by grant of GA ČR A 2075803.

[2]It is clear that under the assumption of independence the results hold, too.

1. INTRODUCTION AND NOTATION

Let N denote the set of all positive integers, R the real line and R^p the p dimensional Euclidean space. We shall consider for any $n \in N$ the linear regression model

$$Y_i = X_i^{\mathrm{T}}\beta^0 + e_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where $Y = (Y_1, Y_2, \ldots, Y_n)^{\mathrm{T}}$ is the response variable, $\{X_i^{\mathrm{T}}, e_i\}_{i=1}^{\infty}$ ($X_i \in R^p, e_i \in R$) is a sequence of independent identically distributed random vectors and β^0 is the "true" vector of regression coefficients. Let us mention (we shall need it later) that the data $\{Y_i, X_i^{\mathrm{T}}, e_i\}$ for one i are denoted as one case. The upper index "$^{\mathrm{T}}$" indicates transposition. (As implicitly follows from this notation, we shall assume all vectors to be column ones.) Finally, let us denote for any $n \in N$ by $X = (X_1, X_2, ..., X_n)^{\mathrm{T}}$ the design matrix and by $e = (e_1, e_2, ..., e_n)^{\mathrm{T}}$ the vector of random fluctuations. Then we can rewrite (1) into, sometimes more convenient, form

$$Y = X\beta^0 + e. \tag{2}$$

We have omitted an indication of the dimension of matrix and of vectors which would presumably unnecessarily burden the notation. Let us notice that in the case that the intercept is included in the model the first coordinates of all vectors X_i's are assumed to be equal to 1. In other words, the explanatory vectors X_i's which are assumed to be random, have degenerated first coordinate. There are of course, except of special cases, well-known reasons for inclusion of the intercept into the model, see Víšek (1997 b). And even when we would leave them aside, we should realize that in the case when we decide not to include intercept into the model we implicitly assume in some sense an absolute character of data and in fact simultaneously abandon otherwise natural requirement of scale- and regression-equivariance of the estimator of the regression coefficients. That is why we shall assume in the rest of paper that the intercept is included into the model. Then we can without any restriction of generality assume:

Assumptions $\mathcal{A}$.. The sequence $\{X_i^{\mathrm{T}}, e_i\}_{i=1}^{\infty}$ ($X_i \in R^p, e_i \in R$) is the sequence of independent identically distributed random variables with $X_{i1} = 1$, $\mathbb{E}X_{1j} = 0$, $\mathbb{E}X_{1j}^2 = \sigma_X^2 \in (0, \infty)$ and $\mathbb{E}X_{1j}^4 < \infty$) for $j = 2, 3, ..., p$.

Notice that in the Assumptions $\mathcal{A}$ there is no specification of the probabilistic characteristics of e_i's. We shall specify them later. Let us recall that the (ordinary) least squares estimator of β^0 is given by

$$\hat{\beta}^{(LS,n)} = \operatorname*{arg\,min}_{\beta \in R^p} \sum_{t=1}^{n} (Y_i - X_i^{\mathrm{T}}\beta)^2 = \operatorname*{arg\,min}_{\beta \in R^p} (Y - X\beta)^{\mathrm{T}}(Y - X\beta)$$

which yields

$$\hat{\beta}^{(LS,n)} = (X^{\mathrm{T}}X)^{-1}\,X^{\mathrm{T}}Y \tag{3}$$

where we have assumed that X is of full rank.

It[3]is of course commonly assumed that a reasonable proposal of estimator is such for which we are able to prove some convenient properties, as e. g. unbiasedness, consistency, optimality in some class(es) of estimators, asymptotic normality. Recently, let us say in the last three decades, we are pleased if also some robustness characteristics are proved and for the estimators which are defined implicitly we surely appreciate if also feasibility of evaluation of the estimator is demonstrated.

The last requirement may seem at the first glance as something superfluous or at least something which is not necessary at the age of powerful computers. A few sentences devoted to the problem may clarify the situation.

Firstly, let us realize and keep in mind that in fact we nearly never compute the *precise* value of the estimator but only better or worse approximation to it. Of course, with exception of the least squares or some ML-estimates but they are usually those estimators which are easily corrupted by contamination.

Secondly, let us take also into account that the robust estimators are frequently applied on relatively small samples of data. Hence their asymptotic properties "do not yet work". It is clear, that their "small-sample" properties may be quite different from the asymptotic ones. And the experiences confirm that, see e. g. Víšek (1997 a). As we have already said, by contrast with the least squares estimator we have not at hand a close formula for evaluation of the estimator and hence we are building our ideas about the "small-sample" properties of it nearly exclusively on numerical studies and experiences.

Third, in the case, when the evaluation of the approximation to the precise value of the estimator is too complicated or there is not at all a reasonable algorithm for solving corresponding extremal problem, a "naive" algorithm, as e. g. a simple resampling, may give false results and consequently it may lead to misleading conclusions, see Hettmansperger, Sheather (1992) and Víšek (1992). Hence feasibility of evaluation of a good approximation to the estimator is crucial inevitable requirement.

It is well-known that to prove the unbiasedness of the $\hat{\beta}^{(LS,n)}$ we (usually) accept the

- *orthogonality condition*, i.e. $\mathbb{E}(e|X) = 0$ (for any $n \in N$).

Similarly to be able to show the optimality of $\hat{\beta}^{(LS,n)}$ in the class of all linear unbiased estimators we adopt the

- *sphericality condition*, i.e. $\mathbb{E}(e\cdot e^{\mathrm{T}}|X) = \sigma_{e_1}^2\mathcal{I}$ (again for any $n \in N$)[4] with $\sigma_{e_1}^2 \in (0,\infty)$,

(see e. g. Hausman (1978)), although it is clear that it is not necessary assumption (see Drygas (1976)). Under the assumption of independence and identical distribution of all vectors $(X_i^{\mathrm{T}}, e_i)^{\mathrm{T}}$, which we have adopted a few lines above, the orthogonality and sphericality conditions reduce of course to $\mathbb{E}(e_1|X_1) = 0$ and to $\mathbb{E}(e_1^2|X_1) = \sigma_{e_1}^2,\ \sigma_{e_1}^2 \in (0,\infty)$, respectively.

[3]Less important parts of text and proofs were given in smaller type of letters to safe the space.

[4]"$\mathcal{I}$" denote the unit matrix.

The sphericality condition is, in the case of deterministic regressors, divided into two conditions and denoted as the assumption of homoscedasticity and absence of correlation among random fluctuations for different cases (Judge et al. (1985)).

But the unbiasedness of the estimator without its consistency would be rather weak certificate of a good behaviour because the estimator could be even asymptotically far from the "true" value of parameters. To guarantee the consistency we usually assume that

$$\mathbb{E}\left\{X_1 \cdot X_1^{\mathrm{T}}\right\} = Q \tag{4}$$

is a regular matrix, see again Dhrygas (1976). This assumption, together with the previous ones, is also sufficient for asymptotic normality.

Since the restriction on the class of linear estimators (which we have considered in connection with the sphericality condition) is (rather) drastic, we should assume and a posteriori verify normality of the random fluctuations to reach the optimality of $\hat{\beta}^{(LS,n)}$ in the class of all unbiased estimators. This is not usually underline even in good monographs that if we find by the least squares such an estimate of model which is well determined, in which the studentized values of the estimates of regression coefficients are significant, for which Durbin-Watson statistic is near to 2, etc., nevertheless in which the residuals are not normally distributed, the model is (nearly) worthless. In this case the estimate of coefficients is the best one only in the class of linear estimators and hence there can be, and frequently is available, much better nonlinear estimator. The sense, in which it is *much better*, is usually related to the sum of squares of residuals of substantial subsample of data. Moreover, when the random fluctuations are not normally distributed, there are easy justifiable doubts whether any sun of squared residuals, for all or for some part of observations is appropriate criterion for quality of the model estimate. E. g. sum of absolute deviation or something else may be preferable.

Substituting from (2) into (3) we obtain

$$\hat{\beta}^{(LS,n)} = \beta^0 + (X^{\mathrm{T}}X)^{-1}\,X^{\mathrm{T}}e = \beta^0 + (\frac{1}{n}X^{\mathrm{T}}X)^{-1}\,\frac{1}{n}X^{\mathrm{T}}e \tag{5}$$

which confirms that orthogonality condition implies unbiasedness. Nevertheless, for consistency it is sufficient - of course together with (4) - only $\mathbb{E}\{X_1 \cdot e_1\} = 0$ which is somewhat weaker than the orthogonality condition. Similarly, for optimality of the least squares estimator in the class of all linear unbiased estimators we need sphericality condition while for the asymptotic normality we need to assume that e. g. that $\mathbb{E}\{X_1 X_1^T e_1^2\}$ is a regular matrix (which is again slightly weaker than sphericality condition).

There are however situations at which the orthogonality condition apparently fails. Probably the most famous example of such a situation is the

situation when the explanatory variable is measured with a random error. Let us look somewhat more closely on the model since it will be useful to keep the example in mind when we shall need later to accept some new assumptions. In the simplest version the model reads

$$Y_i = \alpha + \beta\chi_i + e_i \tag{6}$$

and

$$X_i = \chi_i + v_i, \text{ for } i = 1, 2, \ldots, n \tag{7}$$

where random fluctuations in (6) as well as in (7) are i.i.d. with zero mean and finite positive variance, and they are mutually independent. In words, we assume that response variable Y is proportional to an explanatory variable χ but this variable is measured with an error v so that we have at hand values of the variable X. Of course, we shall assume existence of moments which we shall need bellow. A modification for a model with multidimensional explanatory variable is straightforward.

It may seem strange, in the case of random explanatory variable, to include explicitly in the model a random error of the measurement. But it will be clear in a moment that in the case when the values of random explanatory variable are measured without error (which means that they are measured with error which is negligible with respect to the variance of the random fluctuations e_i's), the ordinary least squares estimator is unbiased while in the case of measurement with an error it is generally biased.

The famous example of this type of model is Friedman's permanent income hypothesis (Friedman (1957)), although nowadays there are some doubts whether this hypothesis does hold.

Substituting from (7) into (6), we arrive at

$$Y_i = \alpha + \beta(X_i - v_i) + e_i = \alpha + \beta X_i + e_i - \beta v_i = \alpha + \beta X_i + u_i$$

with

$$I\!E\{u_i \cdot X_i\} = I\!E\{(e_i - \beta v_i)(\chi_i + v_i)\} = -\beta\sigma_v^2 \tag{8}$$

where σ_v^2 is the variance of the fluctuations v_i's. It indicates that orthogonality condition does not hold. On the other hand,

$$I\!E\left\{u_i^2 | X_i\right\} = I\!E\left\{(e_i - \beta v_i)^2 | X_i\right\} = const$$

where *const* does not depend on i. Finally

$$I\!E\{u_i \cdot u_j\} = I\!E\{(e_i - \beta v_i)(e_j - \beta v_j)\} = 0.$$

Another model which is frequently recalled in this context, is the infinite distributed lag model, e. g. with geometric structure of coefficients, i. e. the

model (again in the simplest version)

$$Y_i = \alpha + \beta \sum_{j=1}^{\infty} \lambda^{j-1} X_{i-j+1} + e_i, \ i = 1, 2, ... \tag{9}$$

where random fluctuations e_i's are i.i.d. with $I\!E e_i = 0$, $I\!E e_i^2 = \sigma_{e_1}^2 \in (0, \infty)$ and for further purposes we shall assume also that $I\!E e_i^3$ exist. Again, a modification for a model with multidimensional explanatory variable is straightforward. It is easy to see that in this form of model we are not able to estimate its parameters. So multiplying the model for $i-1$, i. e.

$$Y_{i-1} = \alpha + \beta \sum_{j=1}^{\infty} \lambda^{j-1} X_{i-j} + e_{i-1} \tag{10}$$

by λ and subtracting this from (9), we obtain for $i = 2, 3, ...$

$$Y_i = (1-\lambda)\alpha + \lambda Y_{i-1} + \beta X_i + e_i - \lambda e_{i-1} = \kappa + \lambda Y_{i-1} + \beta X_i + u_i, \tag{11}$$

with

$$I\!E \{u_i | Y_{i-1} = y_{i-1}\} = I\!E \{(e_i - \lambda e_{i-1}) \mid Y_{i-1} = y_{i-1}\} \neq 0$$

since

$$I\!E \{Y_{i-1} I\!E \{u_i | Y_{i-1} = y_{i-1}\}\} = I\!E \{u_i \cdot Y_{i-1}\} = -\lambda \sigma_{e_1}^2 \tag{12}$$

as follows from (11). Similarly as above

$$I\!E \left\{u_i^2 | Y_{i-1}\right\} = I\!E \left\{(e_i - \lambda e_{i-1})^2 | Y_{i-1}\right\} = const$$

but

$$I\!E\{u_i \cdot u_{i-1}\} = -\lambda \sigma_{e_1}^2 \text{ while } I\!E\{u_i \cdot u_j\} = 0 \text{ for } |i-j| > 1. \tag{13}$$

In both models orthogonality condition fails and moreover in the second model also the sphericality condition is broken, and the random fluctuation are of moving average character. Of course it is possible in such a case to use generalized least squares estimator or estimated generalized least squares estimator since the covariance matrix of random fluctuation (or at least its structure) is known. Nevertheless, (5) together with (8) as well as (12) implies the inconsistency of the ordinary least squares.

We can continue by the model with random regression coefficients (Hildreth, Houck (1968)) or with the simultaneous equations model (e. g. Kmenta (1986)), etc.

So we may conclude: *There are situations when we need to cope with the problem that the orthogonality condition and possibly also the sphericality condition do not hold. On the other hand, as it was indicated and as it is well known, we are able by some transformations of data, sometimes of course not very simple, to cope with heteroscedasticity or with dependence between random fluctuations for different observations. Moreover, it is clear that*

even in the situation when the sphericality condition fails the least squares estimator is still unbiased and consistent provided the orthogonality condition holds. Of course we lose some amount of efficiency.

However, if the orthogonality condition fails the least squares estimator and presumably at least some of robust estimators[5] *are biased and loose their consistency. So it seems that the problems caused by a failure of the orthogonality condition are more acute than the problems which yields the failure of sphericality condition.*

Of course, there are naturally circumstances under which it is not trivial to decide whether the orthogonality condition failed or not. Let us recall that the statistics offers Hausman's specification test, as a tool for decision (see Hausman (1978) or Greene (1993), see also Wu (1973) or Ramsey (1974)), and the instrumental variable estimator, as a tool for estimating β^0 (see e. g. Bowden, Turkington (1984) or Judge et al. (1985)).

Earlier than we shall continue let us recall the heuristics which led to the *instrumental variable estimator* (since we shall need it later). Let us imagine for a while that the least squares estimator was derived in the following, somewhat intuitive way. Multiplying (2) from the left hand side by $\frac{1}{n}X^{\mathrm{T}}$, we obtain

$$\frac{1}{n}X^{\mathrm{T}}Y = \frac{1}{n}X^{\mathrm{T}}X\beta^0 + \frac{1}{n}X^{\mathrm{T}}e.$$

It implies that under the orthogonality condition the expressions

$$\frac{1}{n}X^{\mathrm{T}}Y \qquad \text{and} \qquad \frac{1}{n}X^{\mathrm{T}}X\beta^0$$

have the same limit in probability. It can be considered as a justification for a proposal of an estimator

$$\hat{\beta} = \left(\frac{1}{n}X^{\mathrm{T}}X\right)^{-1}\frac{1}{n}X^{\mathrm{T}}Y = \left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}Y = \hat{\beta}^{(LS,n)}$$

and of an investigation of its properties. So assuming that there is a sequence of p-dimensional random variables $\{Z_i^{\mathrm{T}}\}_{i=1}^{\infty}$ such that

$$\lim_{n\to\infty}\frac{1}{n}Z^{\mathrm{T}}e = 0 \text{ in probability} \tag{14}$$

where of course $Z = (Z_1, Z_2, ..., Z_n)^{\mathrm{T}}$, we can multiply (2) from left by $\frac{1}{n}Z^{\mathrm{T}}$ and we obtain

$$\frac{1}{n}Z^{\mathrm{T}}Y = \frac{1}{n}Z^{\mathrm{T}}X\beta^0 + \frac{1}{n}Z^{\mathrm{T}}e.$$

It means, that the expressions

$$\frac{1}{n}Z^{\mathrm{T}}Y \qquad \text{and} \qquad \frac{1}{n}Z^{\mathrm{T}}X\beta^0 \tag{15}$$

[5] It is possible to judge it e. g. from their asymptotic representations.

have again the same limit in probability. In the analogy to the least squares estimator we can study an estimator

$$\hat{\beta}^{IV} = \left(\frac{1}{n} Z^{\mathrm{T}} X\right)^{-1} \frac{1}{n} Z^{\mathrm{T}} Y = \left(Z^{\mathrm{T}} X\right)^{-1} Z^{\mathrm{T}} Y. \tag{16}$$

This estimator is usually denoted as the *instrumental variable estimator.*

Let us note that in some monographs we can find seemingly more complicated heuristics which should lead to a more general form of the instrumental variable estimator. However, this approach yields more general estimator only in the case when $Z^{\mathrm{T}}X$ is not regular matrix (see Víšek (1997 b)).

Substituting now from (2) into (16) we obtain

$$\hat{\beta}^{IV} = \left(Z^{\mathrm{T}} X\right)^{-1} Z^{\mathrm{T}} \left(X^{\mathrm{T}}\beta^0 + e\right) = \beta^0 + \left(\frac{1}{n} Z^{\mathrm{T}} X\right)^{-1} \frac{1}{n} Z^{\mathrm{T}} e, \tag{17}$$

which indicates consistency of $\hat{\beta}^{IV}$ and hints how to trace out the conditions for unbiasedness. Notice that the consistency of this estimator holds independently of the fact whether the orthogonality condition holds for X_i's or not. In other words, $\hat{\beta}^{IV}$ is consistent both under the hypothesis as well as under the alternative.

On the other hand, more or less frequently we find ourselves in the situation when we feel that except of failure of the orthogonality condition the assumption of normality is also "more or less" broken, e. g. by a contamination of data. And in fact we may expect that the failure of the assumption of normality of random fluctuations take place much more frequently than the failure of the orthogonality condition. We are aware of it due to the fact that more and more studies of real data have already demonstrated that the assumption that the random fluctuations in model are distributed according to a distribution which is well approximated by the normal one is an illusion, the illusion leading sometimes to fatal error. It may be of interest that already Sir R. A. Fisher knew it, see Fisher (1922).

Let us mention at this point that those statisticians who insist on nearly exclusive use of the classical statistical methods (like the least squares or maximum likelihood etc.) still object that they met very rarely with suspicious data. But it is just due to the fact that they use only the classical methods which are not able even to indicate that the data may be damaged. And it may happen even in the case when data are heavily contaminated.

The present author may offer a dozen sample of data for which results of the least squares analysis seemed to be well or even excellent and also the application of diagnostic tools (as hat matrix, Durbin-Watson test, Hausman test, White test, test for normality, etc.) gave satisfactory results. Nevertheless, these results were misleading due to presence of several observations which had (completely) different character than the rest of data. Deleting these observations and applying least squares leads to a model with considerably smaller variance of random fluctuations than had the model estimated for the initial data. That can be assumed to be a justification that this model is more adequate for bulk of

data. However to discover such "dirty" observations is usually nearly impossible without the application of an estimator with reasonably high breakdown point.

So we may conclude that there are inevitably situations when we would like to use (and even should use) a robust method but we cannot guarantee that the random fluctuations are not correlated with the explanatory variables, so that the consistency of the estimation can be seriously damaged[6]. We shall show that a straightforward generalization of the instrumental variable estimator can help.

In order to achieve this let us recall that the M-estimator, for an absolutely continuous (and frequently, but not necessarily, convex) function ρ, is defined as

$$\hat{\beta}^{(M,n)} = \underset{\beta \in R^p}{\arg\min} \sum_{t=1}^{n} \rho(Y_i - X_i^{\mathrm{T}}\beta) \tag{18}$$

and it is usually found as a solution of

$$\sum_{t=1}^{n} X_i \, \psi(Y_i - X_i^{\mathrm{T}}\beta) = 0 \tag{19}$$

where ψ is the derivative of ρ (due to the assumption that ρ is absolutely continuous, ψ exists almost everywhere). Asking for

$$[X^{\mathrm{T}}X]^{-\frac{1}{2}} \sum_{t=1}^{n} X_i \, \psi(Y_i - X_i^{\mathrm{T}}\beta) = o_p(1) \tag{20}$$

instead of (19) allows even to include the estimators with discontinuous ψ-functions (see Rao, Zhao (1992), compare also Jurečková and Welsh (1990)). Let us notice that the M-estimators defined in this way (i. e. either by (18) or (19) or (20)) are not scale equivariant, while e. g. the least squares estimator is scale- and regression-equivariant. Hence it is preferable to studentize the residuals and to define the M-estimators as

$$\hat{\beta}^{(M,n)} = \underset{\beta \in R^p}{\arg\min} \sum_{t=1}^{n} \rho\left(\frac{Y_i - X_i^{\mathrm{T}}\beta}{\hat{\sigma}_{(n)}}\right) \tag{21}$$

where $\hat{\sigma}_{(n)}$ is a (preliminary) $\sqrt{n}$-consistent estimator of σ_{e_1} which is assumed to be scale-equivariant, i. e. for any $c \in R^+, Y \in R^n$ and any matrix X of the type $n \times p$

$$\hat{\sigma}^2_{(n)}(cY, X) = c^2 \hat{\sigma}^2_{(n)}(Y, X)$$

and regression-invariant, i. e. for any $b \in R^p, Y \in R^n$ and any matrix X of the type $n \times p$

$$\hat{\sigma}^2_{(n)}(Y + Xb, X) = \hat{\sigma}^2_{(n)}(Y, X).$$

[6]There are nearly no robust estimator which is unbiased, so we take care typically only of consistency.

Surprisingly, there are only two papers which explicitly stressed that not asking for the regression-invariance of the preliminary scale estimator[7], we would lose both the scale- as well as regression-equivariance of the estimator of β^0, see Bickel (1975) and Jurečková, Sen (1993). Let us recall that the estimator $\hat{\beta}$ of β^0 is scale-equivariant, if for any $c \in R^+, Y \in R^n$ and for any matrix X of type $n \times p$ we have

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

and regression-equivariant if for any $b \in R^p, Y \in R^n$ and for any matrix X of type $n \times p$

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

Let us also recall at this point that there is of course another possibility of studentization, namely simultaneous estimations of β^0 and $\sigma^2_{\varepsilon_1}$, see e. g. Huber (1981) or Hampel et al. (1986). However it proved to be much worse, from the computational standpoint, than iterative procedure with a preliminary scale estimator, since the steps of an algorithm which looks for the simultaneous minimum may lead to an oscillating sequence of values for both parameters.

On the other hand, for the approach with a preliminary scale estimator, the question of existence and computational feasibility of such scale estimator arises immediately. So let us add that an example of such estimators can be found in Jurečková, Sen (1993). An idea based on geometry (or topology, if you want) of observations and using eigenvectors and eigenvalues of covariance matrix of all variables, yields also a suitable preliminary scale estimator, see Víšek (1998 a).

Another possibility is to use a preliminary estimator of β^0 which is consistent (or even $\sqrt{n}$-consistent), scale- and regression-equivariant and evaluate a robust scale estimator, e. g. median absolute deviation of residuals. Of course, we should use as the preliminary estimator preferably an estimator of β^0 with high breakdown point, as e. g. the least trimmed squares (see Víšek (1996 a)). The latter possibility is however applicable only under hypothesis of independence (we shall return to this problem, namely possibility to use a method with high breakdown point, at the end of paper).

We are going now to give assumptions on ψ under which we shall derive the promised results.

Assumptions $\mathcal{B}$.. The function ψ allows a decomposition

$$\psi = \psi_a + \psi_c + \psi_s \tag{22}$$

where:

- *ψ_a is absolutely continuous with absolutely continuous derivative ψ'_a. Denote by ψ''_a the second derivative (where it exists) and*

$$\psi''_{\sup} = \sup\{|\psi''_a(z)| : z \in R\} < \infty.$$

- *ψ_c is continuous with ψ'_c a step-function with a finite number of jump-points, and ψ_c is constant in a neighborhood of $-\infty$ and $+\infty$.*

[7]There was of course a common agreement that the estimator has to be scale-equivariant to be appropriate for studentization.

- *ψ_s is a monotone step-function with steps at points $r_1, r_2, \ldots, r_h$, i. e. there are $\alpha_0, \alpha_1, \ldots, \alpha_h$ so that $\psi_s(z) = \alpha_0$ for $z \in (-\infty, r_1)$, $\psi_s(z) = \alpha_\ell$ for $z \in (r_\ell, r_{\ell+1})$, $\ell = 1, 2, \ldots, h-1$ and $\psi_s(z) = \alpha_h$ for $z \in (r_h, \infty)$.*

Moreover, $\mathbb{E}\{\psi(e_1\sigma_{e_1}^{-1})\} = 0$.

REMARK 1. *It is easy to see that all ψ-functions which are nowadays employed in robust statistics allow the decomposition (22) and most of them also fulfill other assumptions. In fact, it is a consequence of results describing the shape of the optimal B- and V-robust ψ-functions, see Hampel et al. (1986). Roughly speaking they are obtained as $\max\{-b, \min\{-f'_{e_1}/f_{e_1},\ b\}\}$ where f is density of random fluctuations. Due to the decomposition (22) we can prove corresponding assertions successively for special type of ψ-functions.*

The plan of the paper is as follows: First of all we shall prove asymptotic linearity of instrumental M-statistics and define instrumental M-estimators. Then we prove $\sqrt{n}$-consistency of the instrumental M-estimators, find their asymptotic representation and finally also establish asymptotic normality. All results are valid both under orthogonality condition as well as without it.

We shall assume that $\{Z_i\}_{i=1}^{\infty}$ is a sequence of instrumental variables. Since these variables will serve as a substitute of the explanatory variables in the sense which was explained above, we shall assume about them nearly the same what we have assumed about X_i's, namely:

Assumptions $\mathcal{C}$.. The sequence of instrumental variables $\{Z_i\}_{i=1}^{\infty}$, $Z_i \in R^p$, is the sequence of i.i.d. random variables, independent from the sequence $\{e_i\}_{i=1}^{\infty}$, $Z_{i1} = 1$, $\mathbb{E}Z_1 = (1, 0, 0, ..., 0)^{\mathrm{T}}$ $\mathbb{E}Z_{1j}^2 \in (0, \infty)$ for $j = 2, 3, ..., p$. Moreover, $\mathbb{E}\left[Z_1 X_1^{\mathrm{T}} \psi'(e_1\sigma_{e_1}^{-1})\right] = Q$, $\Gamma = \sigma_{e_1}^{-1}\mathbb{E}[e_1\psi'(e_1\sigma_{e_1}^{-1})]$ and $\mathbb{E}\left\{Z_{1j}^2 X_{1k}^2 \left[\psi'(e_1\sigma_{e_1}^{-1})\right]^2\right\}$ exist and are finite for $j, k = 2, 3, ..., p$.

REMARK 2. *Without loss of generality we could assume that $\Gamma = 0$. In fact, it represents a shift of the derivative of ψ-function in the horizontal direction, i. e. assuming $\bar{\psi}(z) = \psi(z + a)$ for an appropriate a. Shifting then the ψ-function in vertical direction, simply taking a modified function $\bar{\bar{\psi}} = \psi(z) - \mathbb{E}\psi(e_1)$, we may reach $\mathbb{E}\{\psi(e_1\sigma_{e_1}^{-1})\} = 0$. The last modification does not change derivative $\psi'(z)$ and hence $\Gamma = 0$ will be kept. So this assumption is (nearly) of the same type as assumption that the mean influence of the random fluctuations on the response variable is compensated, i. e. that $\mathbb{E}e_1 = 0$ (which we adopt in the least square analysis) or $\mathbb{E}\{\psi(e_1\sigma_{e_1}^{-1})\} = 0$ (which we assume for M-analysis).*

Moreover, it was already recalled the optimal B- and V-robust ψ-function are of a shape $\max\{-b, \min\{f'_{e_1}/f_{e_1}, b\}\}$ where f_{e_1} is the density of random

fluctuations, and hence we have for symmetrically distributed random fluctuation $\psi(-z) = \psi(z)$. *Although we do not take every time the optimal* ψ*-function, we usually employ symmetric ones. It implies that in the case when we have no reasons to assume assymetry of random fluctuations, we can consider* $\Gamma = 0$.

2. ASYMPTOTIC LINEARITY OF INSTRUMENTAL M-STATISTICS

At first we shall consider

$$S_n(t,u) = \sum_{i=1}^{n} Z_i \left\{ \psi(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t \right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}) - \psi(e_i \sigma_{e_1}^{-1}) \right\}$$

and we shall put for $M > 1$

$$\mathcal{T}_M = \{ t \in R^p,\, u \in R \,;\, \max\{\|t\|,\, |u|\} < M \}\,.$$

In the proofs of the next theorems some constants C_i's will be defined. These definitions will be assumed valid only within the respective proof.

THEOREM 1. *Let* ψ *be an absolutely continuous function with the absolutely continuous derivative, i. e.* $\psi = \psi_a$, *and the Assumptions* $\mathcal{C}$ *hold. Let us have*

$$\max\left\{ I\!E\left[\|Z_i\| \cdot \|X_1\|^2 \right],\, I\!E\left[\|Z_i\| \|X_1\| |e_1| \right] \right\} < \infty. \tag{23}$$

Finally, let

$$I\!E\{\psi(e_1 \sigma_{e_1}^{-1})\} = 0 \text{ and } 0 < I\!E[e_1 \psi'(e_1 \sigma_{e_1}^{-1})]^2 < \infty. \tag{24}$$

Then for any $M > 0$

$$\sup_{\mathcal{T}_M} \left\| S_n(t,u) + n^{\frac{1}{2}} Q t + n^{-\frac{1}{2}} \Gamma \sum_{i=1}^{n} Z_i u \right\| = \mathcal{O}_p(1) \quad \text{as } n \to \infty. \tag{25}$$

REMARK 3. *Let us again observe, similarly as in Remark 1, that for the* ψ*-functions which are constant in a neighborhood of* $\pm\infty$ *the assumptions of THEOREM 1 hold since* ψ' *as well as* ψ'' *is equal to zero in this neighborhood. As it was already recalled, all the optimal B- and V-robust estimators are generated by such functions (Hampel et al. (1986)). Hence the assumptions of the theorem do not represent a considerable restriction.*

Proof of theorem. Without loss of generality let $\sigma_{e_1} = 1$. First of all let us write for $u \in R,\ |u| < M$

$$e^{-n^{-\frac{1}{2}} u} = 1 - n^{-\frac{1}{2}} u + h\, n^{-1} u^2$$

where $h \in \left[\frac{1}{2} n^{-\frac{1}{2}} e^{-n^{-\frac{1}{2}} M},\, \frac{1}{2} n^{-\frac{1}{2}} e^{n^{-\frac{1}{2}} M} \right]$ and also

$$\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t \right] e^{-n^{-\frac{1}{2}} u}$$

$$= e_i - n^{-\frac{1}{2}} \left(X_i^{\mathrm{T}} t + e_i u \right) + n^{-1} X_i^{\mathrm{T}} t u + h\, e_i\, n^{-1} u^2 - n^{-\frac{3}{2}} X_i^{\mathrm{T}} t\, h \cdot u^2,$$

and finally

$$Z_i\left\{\psi\left(\left[e_i - n^{-\frac{1}{2}}X_i^{\mathrm{T}}t\right]e^{-n^{-\frac{1}{2}}u}\right) - \psi(e_i)\right\} = -n^{-\frac{1}{2}}Z_i(X_i^{\mathrm{T}}t + e_iu)\,\psi'(e_i)$$

$$+n^{-1}Z_i\left\{X_i^{\mathrm{T}}\,t\,u + h\,e_i - n^{-\frac{1}{2}}X_i^{\mathrm{T}}\,t\cdot h\cdot u^2\right\}\psi'(e_i) + Z_iR_{ni}(t) \tag{26}$$

where the remainder term can be written in a form

$$R_{ni}(t) = \left\{-n^{-\frac{1}{2}}(X_i^{\mathrm{T}}t + e_iu) + n^{-1}\left\{X_i^{\mathrm{T}}\,t\,u + h\,e_i\right.\right.$$
$$\left.\left.-n^{-\frac{1}{2}}X_i^{\mathrm{T}}\,t\cdot h\cdot u^2\right\}\right\}\left[\psi'(\xi_i^{(n)}) - \psi'(e_i)\right]$$

for some $\xi_i^{(n)}$ for which, starting from some n_0, we have $|\xi_i^{(n)} - e_i| \le 2n^{-\frac{1}{2}}|X_i^{\mathrm{T}}t + e_iu|$. Now, for any fix j, $k \in \{1, 2, \ldots, p\}$ the sequences

$$\left\{Z_{ij}X_{ik}\psi'(e_i) - I\!E\left[Z_{ij}X_{ik}\psi'(e_i)\right]\right\}_{i=1}^{\infty}$$

and

$$\left\{Z_{ij}\left[e_i\cdot\psi'(e_i) - I\!E[e_i\cdot\psi'(e_i)]\right]\right\}_{i=1}^{\infty}$$

are sequences of independent identically distributed random variables with zero mean and finite (positive) variances and hence Lindeberg-Lévy version of central limit theorem allows to find an $n_1 \in N$ so that for any $\varepsilon > 0$ we may find $C_\varepsilon < \infty$ so that for any $n > n_1$

$$P_F\left\{\max_{1\le j,\,k\le p}\left|n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{Z_{ij}X_{ik}\psi'(e_i) - I\!E\left[Z_{ij}X_{ik}\psi'(e_i)\right]\right\}\right| > C_\varepsilon\right\} < \varepsilon \tag{27}$$

as well as

$$P_F\left\{\max_{1\le j\le p}\left|n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{Z_{ij}\left[e_i\cdot\psi'(e_i) - I\!E[e_i\cdot\psi'(e_i)]\right]\right\}\right| > C_\varepsilon\right\} < \varepsilon. \tag{28}$$

Taking once again into account that

$$\left\{Z_{ij}X_{ik}\psi'(e_i)\right\}_{i=1}^{\infty} \text{ and } \left\{Z_{ij}e_i\psi'(e_i)\right\}_{i=1}^{\infty}$$

are sequences of independent identically distributed random variables with finite mean and applying Kolmogorov's law of large numbers we again find that

$$n^{-1}\max_{1\le j,\,k\le p}\left|\sum_{i=1}^{n}Z_{ij}X_{ik}\psi'(e_i)\right| = \mathcal{O}_p(1) \quad \text{as } n\to\infty, \tag{29}$$

$$n^{-1}\max_{1\le j\le p}\left|\sum_{i=1}^{n}Z_{ij}e_i\psi'(e_i)\right| = \mathcal{O}_p(1) \quad \text{as } n\to\infty \tag{30}$$

and

$$n^{-\frac{3}{2}}\max_{1\le j,\,k\le p}\left|\sum_{i=1}^{n}Z_{ij}X_{ik}\psi'(e_i)\right| = o_p(1) \quad \text{as } n\to\infty. \tag{31}$$

Due to the fact that ψ' is absolutely continuous we may write

$$\psi'(\xi_i^{(n)}) - \psi'(e_i) = \int_{e_i}^{\xi_i^{(n)}}\psi''(z)\mathrm{d}z$$

and hence we have for any n

$$|\psi^{'}(\xi_i^{(n)}) - \psi^{'}(e_i)| \leq n^{-\frac{1}{2}} |X_i^T t|\, \psi''_{\sup}.$$

On the other hand, applying Hlder's inequality we obtain

$$\left| X_i^{\mathrm{T}} t \right| \leq p^{-\frac{1}{2}} \, \|X_i\| \, M$$

and hence there is a constant C_1 such that

$$\sup_{\mathcal{T}_M} \|Z_i R_{ni}(t)\| \leq n^{-1} M^2 C_1 \psi''_{\sup} \, \|Z_i\| \, \|X_i\| \, \{\|X_i\| + |e_i|\}\,.$$

Now, it is again sufficient to take into account that the sequences

$$\left\{ \|Z_i\| \, \|X_i\|^2 \right\}_{i=1}^{\infty}$$

and

$$\{\|Z_i\| \, \|X_i\| \, |e_i|\}_{i=1}^{\infty}$$

are sequences of i. i. d. random variables with finite mean values, and to apply Kolmogorov's law of large numbers once again to find that

$$\sup_{\mathcal{T}_M} \|Z_i R_{ni}(t)\| = \mathcal{O}_p(1). \tag{32}$$

Now, the proof follows from (27), (28), (29), (30), (31) and (32). □

REMARK 4. *Let us notice that the proof of the theorem is essentially based on the character of the processes* $\left\{ Z_i X_i^{\mathrm{T}} t \psi'(e_i) \right\}_{t \in \mathcal{T}_M}$, $\left\{ Z_i X_i^{\mathrm{T}} t \right\}_{t \in \mathcal{T}_M}$, $\{u Z_i \; e_i \psi'(e_i)\}_{u \in \mathcal{T}_M}$, *etc. which are the products of some fixed sequences of random variables and of parameters of the processes. It allows, roughly speaking, to treat the suprema of the processes as the products of these sequences of random variables and of suprema of parameters.*

THEOREM 2. *Let* $\psi'(z) = \alpha_s$, *for* $z \in (r_s, r_{s+1})$, $s = 0, 1, \ldots, k$ *where* $0 = \alpha_0, \alpha_1, \ldots, \alpha_k = 0$ *are real numbers,* $-\infty = r_0 < r_1 < \cdots < r_k < r_{k+1} = \infty$. *Let the Assumptions* $\mathcal{C}$ *hold and* $I\!E\{\psi(e_1 \sigma_{e_1}^{-1})\} = 0$. *Moreover,*

$$\max\left\{ I\!E\left[\|Z_i\| \|X_1\|^2 \right], \, I\!E\left[\|Z_i\| \|X_1\| \cdot |e_1| \right] \right\} < \infty$$

and $I\!E\{\psi(e_1 \sigma_{e_1}^{-1})\} = 0$. *Finally, let* $F_{e_1}(v|X_1)$ *(the conditional distribution of random fluctuations given* X_1*) have a bounded density* $f_{e_1}(v|X_1)$ *and let us assume that the bound may be found independent on* X_1. *Then for any* $M > 0$

$$\sup_{\mathcal{T}_M} \left\{ \left\| S_n(t,u) + n^{\frac{1}{2}} Q t + n^{-\frac{1}{2}} \Gamma \sum_{i=1}^{n} Z_i u \right\| \right\} = \mathcal{O}_p(1) \quad \text{as } n \to \infty. \tag{33}$$

Proof. Not loosing generality let us again assume $\sigma_{e_1} = 1$ and notice that due to the character of the function $\psi(z)$ the second part of assumption (24) is fulfilled. Finally, let us denote $\eta_i = Z_i \left\{ \psi(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t \right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}) - \; \psi(e_i \sigma_{e_1}^{-1}) \right\}$ and $r = \max\{|r_1|, \, |r_k|\}$.

The problem induced by the fact that the derivative ψ' is a step-function is that we cannot use the relation (26) in the case when for given i there is an $s_0 \in \{1, 2, ..., k\}$ such that either

$$(e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t) e^{-n^{-\frac{1}{2}} u} \leq r_{s_0} \leq e_i \tag{34}$$

or

$$e_i \le r_{s_0} \le (e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t) e^{-n^{-\frac{1}{2}} u}. \tag{35}$$

Hence the idea of proof is to withdraw from $S_n(t,u)$ the sum of all η_i's for which either (34) or (35) case takes place, then to show that the sum of the terms which were withdrawn from $S_n(t,u)$ is small in probability and finally to add to the "reduced" $S_n(t,u)$ for all indeces i which were in the previous step withdrawn appropriate terms (sum of them will be shown to be also negligable in probability) to reach assertion of theorem.

In order to fulfill just sketched plan, let us denote the event given by (34) and (35) by $B_n(i,t,u)$ and its indicator by $I_{B_n(i,t,u)}$. Since (34) and (35) are successively equivalent to

$$r_{s_0} \le e_i \le r_{s_0} e^{n^{-\frac{1}{2}} u} + n^{-\frac{1}{2}} X_i^{\mathrm{T}} t$$

and

$$r_{s_0} e^{-n^{-\frac{1}{2}} u} + n^{-\frac{1}{2}} X_i^{\mathrm{T}} t \le e_i \le r_{s_0}$$

and due to the assumption about the upper bound of the density of random fluctuations, there is a constant C_1 such that the conditional probability of $B_n(i,t,u)$ for given X_i is bounded by $C_1 \left\{ n^{-\frac{1}{2}} \left| X_i^{\mathrm{T}} t \right| + r \cdot |e^{n^{-\frac{1}{2}} u} - 1| \right\}$. Now

$$|e^{n^{-\frac{1}{2}} u} - 1| \le |1 + n^{-\frac{1}{2}} M + \frac{n^{-1} M^2}{2!} + \frac{n^{-\frac{3}{2}} M^3}{3!} + ... - 1|$$

$$\le n^{-\frac{1}{2}} M \left(1 + n^{-\frac{1}{2}} M + \frac{n^{-1} M^2}{2!} + \frac{n^{-\frac{3}{2}} M^3}{3!} + ... \right) = n^{-\frac{1}{2}} M e^{n^{-\frac{1}{2}} M}.$$

Then there is a constant C_2 such that for $t, u \in \mathcal{T}_M$ corresponding probability is bounded by $n^{-\frac{1}{2}} C_2 \{\|X_i\| + 1\}$. Similar consideration lead to the existence of a constant C_3 such that

$$|\psi(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t \right] e^{-n^{-\frac{1}{2}} u}) - \psi(e_i)| \le n^{-\frac{1}{2}} C_3 \{\|X_i\| + |e_i|\}.$$

Then

$$I\!E \|\eta_i I_{B_n(i,t,u)}\|$$

$$\le n^{-1} C_4 I\!E \left\{ \|Z_i\| I\!E \left\{ \left[\psi(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t \right] e^{-n^{-\frac{1}{2}} u}) - \psi(e_i) \right] I_{B_n(i,t,u)} \middle| X_i \right\} \right\}$$

$$\le n^{-1} C_4 I\!E \left\{ \|Z_i\| I\!E \left\{ [\|X_i\| + |e_i|] I_{B_n(i,t,u)} | X_i \right\} \right\}$$

$$\le n^{-1} C_5 I\!E \left\{ \|Z_i\| (\|X_i\| + 1) (\|X_i\| + |e_i|) \right\}$$

for appropriate constants C_4 and C_5. Now we have

$$S_n(t,u) = \sum_{i=1}^n \eta_i I_{B_n(i,t,u)} + \sum_{i=1}^n \eta_i (1 - I_{B_n(i,t,u)})$$

and using Chebyshev's inequality for the nonnegative random variable, we obtain

$$P\left(\| \sum_{i=1}^n \eta_i I_{B_n(i,t,u)} \| > C_6 \right) \le C_6^{-1} C_5 I\!E \left\{ \|Z_1\| (\|X_1\| + 1) (\|X_1\| + |e_1|) \right\}$$

On the other hand, $\sum_{i=1}^n \eta_i (1 - I_{B_n(i,t,u)})$ can be treated in the same way as the $S_n(t,u)$ in the proof of Theorem 1. To finish the proof we need to add to

$$\sum_{i=1}^n n^{-\frac{1}{2}} Z_i (X_i^{\mathrm{T}} t + e_i u) \psi'(e_i) (1 - I_{B_n(i,t,u)}) \tag{36}$$

the sum

$$\sum_{i=1}^{n} n^{-\frac{1}{2}} Z_i(X_i^{\mathrm{T}} t + e_i u)\, \psi'(e_i) I_{B_n(i,t,u)}$$

to reach the assertion of present theorem, since we can then show (in the same way as it was done in the proof of Theorem 1) that

$$\sum_{i=1}^{n} n^{-\frac{1}{2}} Z_i(X_i^{\mathrm{T}} t + e_i u)\, \psi'(e_i) - \left[n^{\frac{1}{2}} Q t + n^{-\frac{1}{2}} \Gamma \sum_{i=1}^{n} Z_i u \right] = \mathcal{O}_p(1) \quad \text{as } n \to \infty.$$

However, along similar lines as above we may find that

$$\sup_{\mathcal{T}_M} \left\| \sum_{i=1}^{n} n^{-\frac{1}{2}} Z_i(X_i^{\mathrm{T}} t + e_i u)\, \psi'(e_i) I_{B_n(i,t,u)} \right\| = \mathcal{O}_p(1)$$

so that we can add this to (36). That concludes the proof. □

THEOREM 3. *Let $\psi(z) = \alpha_s$ for $z \in (r_s, r_{s+1})$, $s = 0, 1, \ldots, m$ where $\alpha_0, \alpha_1, \ldots, \alpha_m$ are real numbers and $-\infty = r_0 < r_1 < \cdots < r_k < r_{m+1} = \infty$. Let again the Assumptions $\mathcal{C}$ hold. Moreover, let $F_{e_1}(v|X_1)$ have for all values of X_1 bounded density $f_{e_1}(v|X_1)$ which is Lipschitz of the first order and let us assume that the bound as well as Lipschitz constant may be found independent of X_1. Assume also that the joint distribution of X_1 and e_1 is continuous. Finally put*

$$q = \sum_{s=1}^{m} \left\{ (\alpha_s - \alpha_{s-1})\, I\!E \left[Z_1 X_1^{\mathrm{T}} f_{e_1}(\sigma_{e_1} r_s | X_1) \right] \right\}$$

and

$$\gamma = \sum_{s=1}^{m} \left\{ r_s (\alpha_s - \alpha_{s-1})\, I\!E \left[Z_1 f_{e_1}(\sigma_{e_1} r_s | X_1) \right] \right\}.$$

Then for any $M > 0$

$$\sup_{\mathcal{T}_M} \left\{ \left\| n^{-\frac{1}{2}} S_n(t,u) + qt + \gamma u \right\| \right\} = \mathcal{O}_p\left(n^{-\frac{1}{4}} \right) \quad \text{as } n \to \infty. \tag{37}$$

Proof. Without loss of generality let us assume that $m = 1$ (write r instead of r_1) and $\alpha_0 < \alpha_1$, and denote $\tau = \alpha_1 - \alpha_0$. Then according to the assumptions, there is $C_1 < \infty$ such that we have $f_{e_1}(v|X_1) < C_1$. Let us denote

$$\xi_i(n,t,u) = \psi(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t \right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}) - \psi(e_i \sigma_{e_1}^{-1})$$

and assume that $\sigma_{e_1} = 1$. It is clear that $\xi_i(n,t,u) \neq 0$ only if either

$$e_i < r < [e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t] e^{-n^{-\frac{1}{2}} u} \quad \leftrightarrow \quad n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u} < e_i < r \tag{38}$$

or

$$[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t] e^{n^{-\frac{1}{2}} u} < r < e_i \quad \leftrightarrow \quad r < e_i < n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u}. \tag{39}$$

Denote the events described in (38) and (39) successively by $B_i^{(k)}(n,t,u)$, $k = 1, 2$. First of all, please observe that (38) can take place when

$$n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u} < r \tag{40}$$

and similarly (39) can hold if

$$n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u} > r. \tag{41}$$

Fix an $\ell \in \{1, 2, ..., p\}$ and denote successively by $D_i^{(j)}(n, t, u, \ell),\ j = 1, 2, 3, 4$ the events

$$\left\{\omega \in \Omega : \{n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u} < r\} \cap \{Z_{i\ell} \le 0\}\right\},$$

$$\left\{\omega \in \Omega : \{n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u} < r\} \cap \{Z_{i\ell} > 0\}\right\},$$

$$\left\{\omega \in \Omega : \{n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u} \ge r\} \cap \{Z_{i\ell} \le 0\}\right\}$$

and

$$\left\{\omega \in \Omega : \{n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u} \ge r\} \cap \{Z_{i\ell} > 0\}\right\}.$$

Further denote by $\pi_i^{(j,k)}(n, t, u, \ell)$ $j = 1, 2, 3, 4,\ k = 1, 2$ the conditional probabilities of the events $B_i^{(k)}(n, t, u) \cap D_i^{(j)}(n, t, u, \ell)$ given the event $D_i^{(j)}(n, t, u, \ell)$. For any $n \in N$, $j = 1, 2, 3, 4$ and $k = 1, 2$ we have

$$\begin{aligned}
\pi_i^{(j,k)}(n, t, u, \ell) &= I\!E \left\{ I_{B_i^{(k)}(n,t,u) \cap D_i^{(j)}(n,t,u,\ell)} | D_i^{(j)}(n, t, u, \ell) \right\} \\
&= I\!E \left\{ I\!E \left[I_{B_i^{(k)}(n,t,u) \cap D_i^{(j)}(n,t,u,\ell)} | X_i, D_i^{(j)}(n, t, u, \ell) \right] \right\} \\
&\le \int_{-\infty}^{\infty} \left\{ \int_{n^{-\frac{1}{2}} \left[X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u} \right]}^{r} f_{e_i}(v | X_i = x) \mathrm{d}v \right\} f_{X_i}(x) \mathrm{d}x \qquad (42) \\
&\le C_1 \int_{-\infty}^{\infty} \left\{ \int_{n^{-\frac{1}{2}} \left[X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u} \right]}^{r} \mathrm{d}v \right\} f_{X_i}(x) \mathrm{d}x
\end{aligned}$$

and hence there is a constant C_2 such that

$$\pi_i^{(j,k)}(n, t, u, \ell) < n^{-\frac{1}{2}} C_2 \left\{ I\!E \|X_i\| + 1 \right\}. \tag{43}$$

Of course, lower and upper bound in (42) should be interchanged if $r < n^{-\frac{1}{2}} \left[X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u} \right]$ but (43) holds for any combination of j and k. Now, we shall study the sum

$$S_{n\ell}(t, u) = \sum_{i=1}^{n} Z_{i\ell} \left[\xi_i(n, t, u) - I\!E\, \xi_i(n, t, u) \right].$$

Since $\cup_{j=1}^{4} D_i^{(j)}(n, t, u, \ell) = \Omega$ a.s., we have

$$\xi_i(n, t, u) = \sum_{j=1}^{4} \left[\xi_i(n, t, u) I_{D_i^{(j)}(n,t,u,\ell)} \right] \text{ a.s.},$$

and hence $I\!E \xi_i(n, t, u) = \sum_{j=1}^{4} I\!E \left[\xi_i(n, t, u) I_{D_i^{(j)}(n,t,u,\ell)} \right]$. Then

$$\sum_{i=1}^{n} Z_{i\ell} \left[\xi_i(n, t, u) - I\!E\, \xi_i(n, t, u) \right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{4} Z_{i\ell}\left[\xi_i(n,t,u) I_{D_i^{(j)}(n,t,u,\ell)} - I\!E\,[\xi_i(n,t,u) I_{D_i^{(j)}(n,t,u,\ell)}]\right]. \tag{44}$$

Now consider $Z_{i\ell}\left[\xi_i(n,t,u) I_{D_i^{(1)}(n,t,u,\ell)} - I\!E\,[\xi_i(n,t,u) I_{D_i^{(1)}(n,t,u,\ell)}]\right]$. We easy find that, under condition that $I_{D_i^{(1)}(n,t,u,\ell)} = 1$, we have

$$Z_{i\ell}\left[\xi_i(n,t,u) I_{D_i^{(1)}(n,t,u,\ell)} - I\!E\,[\xi_i(n,t,u) I_{D_i^{(1)}(n,t,u,\ell)}]\right]$$

$$= \tau Z_{i\ell}(1-\pi_i^{(1,1)}(n,t,u,\ell)) = -\tau|Z_{i\ell}|(1-\pi_i^{(1,1)}(n,t,u,\ell)) > -\tau|Z_{i\ell}| \tag{45}$$

with probability $\pi_i^{(1,1)}(n,t,u,\ell)$

and

$$= -\tau Z_{i\ell}\pi_i^{(1,1)}(n,t,u,\ell) = \tau|Z_{i\ell}|\pi_i^{(1,1)}(n,t,u,\ell) < \tau n^{-\frac{1}{2}} C_2 |Z_{i\ell}|\cdot[I\!E\|X_i\|+1] \tag{46}$$

with probability $1-\pi_i^{(1,1)}(n,t,u,\ell)$.

Taking into account the expressions staying after the first sign of equality in (45) and in (46), and corresponding probabilities, we find that

$$I\!E\left\{ Z_{i\ell}\left[\xi_i(n,t,u) I_{D_i^{(1)}(n,t,u,\ell)} - I\!E\,[\xi_i(n,t,u) I_{D_i^{(1)}(n,t,u,\ell)}]\right]\right\} = 0$$

Notice please that for $I_{D_i^{(1)}(n,t,u,\ell)} = 0$ we have

$$Z_{i\ell}\left[\xi_i(n,t,u) I_{D_i^{(1)}(n,t,u,\ell)} - I\!E\,[\xi_i(n,t,u) I_{D_i^{(1)}(n,t,u,\ell)}]\right] = 0.$$

So, putting for any $n \in N$ and $i = 1,2,...,n$ $a_{i\ell}(n,t,u) = \tau|Z_{i\ell}|\,\pi_i^{(1,1)}(n,t,u,\ell)$ and $b_{i\ell}(n,t,u) = \tau|Z_{i\ell}|(1-\pi_i^{(1,1)}(n,t,u,\ell))$, and utilizing Lemma A.2, define for $I_{D_i^{(1)}(n,t,u,\ell)} = 1$ $\mu_{i\ell}^{(1)}(n,t,u)$ the time for Wiener process to exit the interval $(-a_{i\ell}(n,t,u), b_{i\ell}(n,t,u))$ and for $I_{D_i^{(1)}(n,t,u,\ell)} = 0$ $\mu_{i\ell}^{(1)}(n,t,u) = 0$. Then we obtain

$$Z_{i\ell}\left[\xi_i(n,t,u) I_{D_i^{(1)}(n,t,u,\ell)} - I\!E\,[\xi_i(n,t,u) I_{D_i^{(1)}(n,t,u,\ell)}]\right] =_{\mathcal{D}} W(\mu_{i\ell}^{(1)}(n,t,u))$$

where "$=_{\mathcal{D}}$" denotes equality in distribution. Similarly we find for $j = 2, 3$ and 4

$$Z_{i\ell}\left[\xi_i(n,t,u) I_{D_i^{(j)}(n,t,u,\ell)} - I\!E\,[\xi_i(n,t,u) I_{D_i^{(j)}(n,t,u,\ell)}]\right] =_{\mathcal{D}} W(\mu_{i\ell}^{(j)}(n,t,u)).$$

Finally, putting $\mu_{i\ell}(n,t,u) = \sum_{j=1}^{4}\mu_{i\ell}^{(j)}(n,t,u)$ and taking into account (44), we obtain

$$n^{-\frac{1}{4}}[S_{n\ell}(t,u) - I\!E\,S_{n\ell}(t,u)] = n^{-\frac{1}{4}}\sum_{i=1}^{n} Z_{i\ell}\,[\xi_i(n,t,u) - I\!E\,\xi_i(n,t,u)]$$

$$=_{\mathcal{D}}\ n^{-\frac{1}{4}}\sum_{i=1}^{n}\sum_{j=1}^{4} W(\mu_{i\ell}^{(j)}(n,t,u)) =_{\mathcal{D}}\ n^{-\frac{1}{4}}\sum_{i=1}^{n} W(\mu_{i\ell}(n,t,u))$$

$$=_{\mathcal{D}} W(n^{-\frac{1}{2}}\sum_{i=1}^{n}\mu_{i\ell}(n,t,u)).$$

Now, let us take into account inequalities which are given in (43), (45) and (46), and put $c_{i\ell} = \tau|Z_{i\ell}|$ and $d_{i\ell} = n^{-\frac{1}{2}}\tau C_2|Z_{i\ell}|\cdot[\|X_i\|+1]$. Defining

$\kappa_{i\ell}(n,M)$ the time for Wiener process to exit the interval $(-c_{i\ell}, d_{i\ell})$, (47)

we obtain

$$\mu_{i\ell}(n,t,u) \le \kappa_{i\ell}(n,M).$$

So, we arrive at

$$\sup_{\mathcal{T}_M} n^{-\frac{1}{4}} \left|S_{n\ell}(t,u) - I\!E\ S_{n\ell}(t,u)\right| =_{\mathcal{D}} \sup_{\mathcal{T}_M} \left| W(n^{-\frac{1}{2}} \sum_{i=1}^{n} \mu_{i\ell}(n,t,u)) \right|$$

$$\le \sup \left\{ |W(s)| \ :\ 0 \le s \le n^{-\frac{1}{2}} \sum_{i=1}^{n} \kappa_{i\ell}(n,M) \right\}. \tag{48}$$

Moreover, see again Lemma A.2, we have from (47) for any $t, u \in \mathcal{T}_M$

$$I\!E\kappa_{i,\ell}(n,C) \le 4n^{-\frac{1}{2}} C_2 I\!E Z_{i\ell}^2 \cdot [I\!E\|X_i\| + 1]$$

for all $n \in N$, i. e.

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} I\!E\kappa_{i,\ell}(n,C) \le C_2 I\!E Z_{1\ell}^2 \cdot [I\!E\|X_1\| + 1].$$

It means that for any positive ε there is a constant C_3 and $n_\varepsilon \in N$ so that for any $n > n_\varepsilon$

$$P\left\{ n^{-\frac{1}{2}} \sum_{i=1}^{n} \kappa_i(n,C) > C_3 \right\} < \frac{\varepsilon}{2} \tag{49}$$

and then there is also $C_4 \in (0,\infty)$ such that

$$P\left\{\sup\{|W(s)| \ :\ 0 \le s \le C_3\} > C_4\right\} < \frac{\varepsilon}{2}, \tag{50}$$

see e. g. Csörgö, Révész (1981). Taking into account (48), (49) and (50), we arrive at

$$P\left\{ \sup_{\mathcal{T}_M} n^{-\frac{1}{4}} \left|S_{n\ell}(t,u) - I\!E\ S_{n\ell}(t,u)\right| > C_4 \right\} < \varepsilon$$

and it means that also

$$\sup_{\mathcal{T}_M} n^{-\frac{1}{4}} \left\|S_n(t,u) - I\!E\ S_n(t,u)\right\|$$

is bounded in probability. We shall conclude the proof if we show that

$$n^{-\frac{1}{4}} \sup_{\mathcal{T}_M} \left\| I\!E\ S_n(t,u) + n^{-\frac{1}{2}} \tau I\!E \sum_{i=1}^{n} \left[Z_i \left(X_i^{\mathrm{T}} t + ur \right) f_{e_i}(r|X_i) \right] \right\| = \mathcal{O}(1)$$

as $n \to \infty$. We have already shown that $I\!E\xi_i(n,t,u)$
$= \sum_{j=1}^{4} I\!E \left[\xi_i(n,t,u) I_{D_i^{(j)}(n,t,u,\ell)} \right]$. On the other hand, we have

$$I\!E\ \left\{ Z_i \xi_i(n,t,u) I_{D^{(1)}(n,t,u,\ell)} \right\} = I\!E\ \left\{ Z_i I\!E \left[\xi_i(n,t,u) I_{D^{(1)}(n,t,u,\ell)} | X_i, Z_i \right] \right\}$$

$$= I\!E \left\{ Z_i \int_{-\infty}^{\infty} \xi_i(n,t,u) I_{D_i^{(j)}(n,t,u,\ell)} f_{e_i}(v|X_i,Z_i)\ \mathrm{d}v \right\}$$

$$= \tau I\!E \left\{ Z_i \int_{n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u}}^{r} f_{e_i}(v|X_i,Z_i) \mathrm{d}v \right\}$$

$$= \tau I\!E \left\{ Z_i \int_{n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u}}^{r} \left[f_{e_i}(r|X_i,Z_i) + f_{e_i}(v|X_i,Z_i) - f_{e_i}(r|X_i,Z_i) \right] \mathrm{d}v \right\}.$$

(Again the upper and lower bound of the integral should be interchanged, if it is appropriate.) Since Z_i and e_i are independent, the last expression is equal to

$$\tau I\!E \left\{ Z_i \int_{n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u}}^{r} \left[f_{e_i}(r|X_i) + f_{e_i}(v|X_i) - f_{e_i}(r|X_i) \right] \mathrm{d}v \right\}$$

$$= -\tau I\!E \left\{ Z_i \left[n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r(e^{n^{-\frac{1}{2}} u} - 1) \right] f_{e_i}(r|X_1) \right\} + I\!E R^*_{int}$$

where

$$R^*_{int} = \tau \left\{ Z_i \int_{n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u}}^{r} \left[f_{e_i}(v|X_i) - f_{e_i}(r|X_i) \right] \mathrm{d}v \right\}.$$

Moreover

$$e^{n^{-\frac{1}{2}} u} - 1 = n^{-\frac{1}{2}} u + \frac{n^{-1} u^2}{2!} + \frac{n^{-\frac{1}{2}} u^3}{3!} + ...,$$

and hence we have

$$I\!E\ S_n(t,u) + n^{-\frac{1}{2}} \tau I\!E \sum_{i=1}^{n} \left[Z_i \left(X_i^{\mathrm{T}} t + ur \right) f_{e_i}(r|X_i) \right]$$

$$= \sum_{i=1}^{n} \left\{ I\!E\ R^*_{int} - \tau r I\!E Z_i \left[\frac{n^{-1} u^2}{2!} + \frac{n^{-\frac{3}{2}} u^3}{3!} + ... \right] f_{e_i}(r|X_1) \right\}.$$

Finally,

$$n^{-\frac{1}{4}} \sup_{\mathcal{T}_M} \left\| I\!E\ S_n(t,u) + n^{-\frac{1}{2}} \tau I\!E \sum_{i=1}^{n} \left[Z_i \left(X_i^{\mathrm{T}} t + ur \right) f_{e_i}(r|X_i) \right] \right\|$$

$$= n^{-\frac{1}{4}} \sup_{\mathcal{T}_M} \left\| \sum_{i=1}^{n} I\!E\ R^*_{int} \right\| + \mathcal{O}(n^{-\frac{1}{4}}) \leq n^{-\frac{1}{4}} \sup_{\mathcal{T}_M} \sum_{i=1}^{n} \| I\!E\ R^*_{int} \| + \mathcal{O}(n^{-\frac{1}{4}})$$

as $n \to \infty$. Recalling that $f_{e_1}(v|X_1)$ is Lipschitz and $|r - v| \leq n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u}$, we have for $t, u \in \mathcal{T}_M$

$$|f_{e_1}(v|X_1) - f(r|X_1)| < n^{-\frac{1}{2}} C_5 \left[\|X_1\| + 1 \right]$$

and hence for $t, u \in \mathcal{T}_M$

$$\| I\!E R^*_{int} \| \leq n^{-\frac{1}{2}} \tau C_5 \left\| I\!E \left\{ \|Z_i\| \left[\|X_i\| + 1 \right] \cdot \int_{n^{-\frac{1}{2}} X_i^{\mathrm{T}} t + r e^{n^{-\frac{1}{2}} u}}^{r} \mathrm{d}v \right\} \right\|$$

$$\leq n^{-1} C_6 I\!E \left\{ \|Z_i\| \cdot \left[\|X_i\| + 1 \right]^2 \right\}$$

for some constant C_6 (again the upper and the lower bound of the integral should be interchanged, if it is appropriate). So we have

$$n^{-\frac{1}{4}} \sup_{\mathcal{T}_M} \sum_{i=1}^{n} \| I\!E\ R^*_{int} \| = \mathcal{O}(n^{-\frac{1}{4}})$$

which concludes the proof. □

3. CONSISTENCY AND ASYMPTOTIC NORMALIRY OF INSTRUMENTAL M-ESTIMATES

THEOREM 4. *Let $\psi = \psi_a + \psi_c$ and the assumptions of Theorem 1 and 2 be fulfilled. Moreover, let $\hat{\sigma}_{(n)}$ be a $\sqrt{n}$-consistent estimator of scale of random fluctuations. Finally, let Q (see Assumptions $\mathcal{C}$) be positive definite matrix. Then the equation*

$$\sum_{t=1}^{n} Z_i \, \psi\left(\frac{Y_i - X_i^{\mathrm{T}}\beta}{\hat{\sigma}_{(n)}}\right) = 0 \tag{51}$$

has a $\sqrt{n}$-consistent solution, i. e. there is $\hat{\beta}^{(IM,n)}$ for which (51) is fulfilled [8] *and*

$$\sqrt{n}\left(\hat{\beta}^{(IM,n)} - \beta^0\right) = \mathcal{O}_p(1) \quad \textit{as } n \to \infty.$$

Proof. Using (25) and (33) we arrive at

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \psi(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u})$$

$$= n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \psi(e_i \sigma_{e_1}^{-1}) - Qt - \Gamma I\!E Z_1 u + o_p(1) \quad \text{as } n \to \infty. \tag{52}$$

Due to the assumptions on the functions ψ_a and ψ_c it is possible to verify that the assumptions of Lindeberg-Lévy theorem are fulfilled for the sequence of random variables

$$\left\{Z_i \psi(e_i \sigma_{e_1}^{-1})\right\}_{i=1}^{\infty}$$

and due to the fact that we have assumed that $I\!E\left\{\psi(e_1 \sigma_{e_1}^{-1})\right\} = 0$ and Z_i and e_i are mutually independent,

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \psi(e_i \sigma_{e_1}^{-1})$$

is bounded in probability (of course, independently on t and u). It means that for any $\varepsilon > 0$ there is a constant $C_1 > 0$ and $n_0 \in N$ so that for any $n > n_0$ we have for

$$B_n = \left\{\omega \in \Omega \ : \ \left\| n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \psi(e_i \sigma_{e_1}^{-1}) \right\| < C_1 \right\}$$

$P(B_n) > 1 - \varepsilon$. Further for any $\Delta > 0$, let $n_1 > n_0$ be selected so that for all $n > n_1$ there is a set C_n^{Δ} such that for all $\omega \in C_n^{\Delta}$ the term "$o_p(1)$" in (52) is smaller than Δ and $P(C_n^{\Delta}) > 1 - \varepsilon$. Taking into account the linearity in t of

$$t^T n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \psi(e_i \sigma_{e_1}^{-1})$$

and of

$$t^T \Gamma I\!E Z_1 u,$$

[8] The upper index 'IM' means *instrumental M-estimate*

we may find for any $C_2 > 0$ and any fixed $u \in (0, C_2)$ a constant $C_3 > 0$ so that for any $n > n_1$, $\omega \in B_n \cap C_n^{\Delta}$ and any $t \in R^p$, $\|t\| = C_3$, we have

$$-t^T n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \psi(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u})$$

$$= -t^T n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \psi(e_i \sigma_{e_1}^{-1}) + t^T Q t + t^T \Gamma I\!E Z_1 u + o_p(1) \geq 0$$

since we have assumed the matrix Q to be positive definite. Now due to the $\sqrt{n}$-consistency of $\hat{\sigma}^2_{(n)}(\omega)$ we may find $n_2 > n_1$ so that for all $n > n_2$ there is a set D_n such that $P(D_n^c) < \varepsilon$ and for any $\omega \in D_n$ $\hat{\sigma}^2_{(n)}(\omega) = \sigma_{e_1} \cdot \exp^{n^{-\frac{1}{2}} u}$ for some $|u| < M$ (of course, the consistency would suffice in this case). Finally applying Assertion A.1 we find for any $n > n_2$ and $\omega \in B_n \cap C_n^{\Delta} \cap D_n$ such $t \in R^p$ that $\|t_0\| \leq C_3$, $t_0 = t_0(u, \omega)$ and

$$\sum_{i=1}^{n} Z_i \psi(\frac{Y_i - X_i^{\mathrm{T}} \beta^0 + n^{-\frac{1}{2}} X_i^{\mathrm{T}} t_0)}{\hat{\sigma}^2_{(n)}}) = 0.$$

Writing

$$t_0(u, \omega) = \sqrt{n} \left(\hat{\beta}(u, \omega) - \beta^0\right)$$

and taking into account finiteness of $\hat{\sigma}_{(n)}$, we conclude the proof of the promised assertion. □

REMARK 5. *Notice please that the key role play in the previous proof the application of (25) and (33) and the assumption that Q is positive definite. In what follows we shall use the same in a somewhat more complicated situation.*

COROLLARY 1. *Let $I\!E\left\{Z_1 X_1^{\mathrm{T}} \psi'(e_1 \sigma_{e_1}^{-1})\right\} = Q$ be a positive definite matrix and $\hat{\sigma}_{(n)}$ a $\sqrt{n}$-consistent estimate of σ_{e_1}. Then under assumptions of Theorem 4 we have*

$$\sqrt{n} \left(\hat{\beta}^{(IM,n)} - \beta^0\right) = n^{-\frac{1}{2}} Q^{-1} \sum_{i=1}^{n} Z_i \left\{\psi(e_i \sigma_{e_1}^{-1}) - \Gamma(\log\hat{\sigma}_{(n)} - \log\sigma_{e_1})\right\}$$

$$+\mathcal{O}_p(n^{-\frac{1}{2}}) \quad \textit{as } n \to \infty. \tag{53}$$

Proof. Let us recall that due to previous theorem $t = \sqrt{n} \left(\hat{\beta}^{(IM,n)} - \beta^0\right) = \mathcal{O}_p(1)$ and due to the assumptions of this corollary we also have $u = \sqrt{n}(\log\hat{\sigma}_{(n)} - \log\sigma_{e_1}) = \mathcal{O}_p(1)$. Moreover

$$\sum_{i=1}^{n} Z_i \psi(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}) = \sum_{i=1}^{n} Z_i \psi(\frac{Y_i - X_i^{\mathrm{T}} \hat{\beta}^{(IM,n)}}{\hat{\sigma}_{(n)}}) = 0.$$

So using (25) and (33) we obtain

$$-n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \psi(e_i \sigma_{e_1}^{-1}) + \sqrt{n} Q(\hat{\beta}^{(IM,n)} - \beta)$$

$$+n^{-\frac{1}{2}} \Gamma \sum_{i=1}^{n} Z_i (\log\hat{\sigma}_{(n)} - \log\sigma_{e_1}) = \mathcal{O}_p(n^{-\frac{1}{2}}) \quad \text{as } n \to \infty$$

which yields (53).

REMARK 6. *Notice please that even in the case when the orthogonality condition is broken, the matrix Q has a block-structure of the type*

$$\begin{bmatrix} \nu & \zeta^{\mathrm{T}} \\ 0 & Q_{(1)} \end{bmatrix}$$

where $\nu = I\!E\psi'(e_1\sigma_{e_1}^{-1}$ *and* $zeta_{j-1} = I\!E\left[Z_{11}X_{1j}\psi'(e_1\sigma_{e_1}^{-1})\right]$, $j = 2, 3, ...p$. *The elements in the first column starting with thw second row are zero, since* $I\!E\left[Z_{1j}X_{11}\psi(e_1\sigma_{e_1}^{-1})\right] = I\!E\left[Z_{1j}\psi(e_1\sigma_{e_1}^{-1})\right] = 0$ $j = 2, 3, ...p$. *It implies that*

$$Q^{-1} = \begin{bmatrix} d^{\mathrm{T}} \\ 0 \quad Q_{(1)}^{-1} \end{bmatrix}$$

where $d^{\mathrm{T}} = -a^{\mathrm{T}} \cdot Q_{(1)}^{-1}$. *Now having rewritten (53) into the form*

$$\sqrt{n}\left(\hat{\beta}^{(IM,n)} - \beta^0\right) = n^{-\frac{1}{2}}Q^{-1}\sum_{i=1}^{n} Z_i\psi(e_i\sigma_{e_1}^{-1})$$

$$-n^{-\frac{1}{2}}Q^{-1}\sum_{i=1}^{n} Z_i\Gamma(\log\hat{\sigma}_{(n)} - \log\sigma_{e_1}) + \mathcal{O}_p(n^{-\frac{1}{2}}) \quad \textit{as } n \to \infty$$

and taking into account that except of the first coordinate the vector $n^{-\frac{1}{2}}\sum_{i=1}^{n} Z_i$ *is bounded in probability while* $\log\hat{\sigma}_{(n)} - \log\sigma_{e_1}$ *converges to zero, we find that the last but one term in (53) really affects only the first coordinate of* $\sqrt{n}\left(\hat{\beta}^{(IM,n)} - \beta^0\right)$.

On the other hand, to avoid completely this influence, we need $\Gamma = 0$. *As we have already shown, it takes place e. g. in the case when the distribution of random fluctuations is symmetric and we employ optimal B-robust M-estimator. After all, the situation is not surprising because it is known that an unbiased and efficient estimator of the location parameter may be constructed in the case when we assume nothing more then symmetry of distribution of underlying probabilistic model, see Beran (1978), Stone (1975) or Víšek (1991), but the symmetry is substantial.*

COROLLARY 2. *Let* $I\!E\left\{Z_1X_1^{\mathrm{T}}\psi'(e_1\sigma_{e_1}^{-1})\right\} = Q$ *be a positive definite matrix,* $\Gamma = 0$ *and let* $\hat{\sigma}_{(n)}$ *be a* $\sqrt{n}$*-consistent estimate of the variance* σ_{e_1}. *Then under assumptions of Theorem 4* $\sqrt{n}\left(\hat{\beta}^{(IM,n)} - \beta^0\right)$ *is asymptotically normal with zero mean and a covariance matrix*

$$C = Q^{-1}I\!E\left\{\psi^2(e_1\sigma_{e_1}^{-1})Z_1Z_1^{\mathrm{T}}\right\}[Q^{-1}]^T.$$

Proof. directly follows from the previous corollary. □

Let us turn now our attention to the case of discontinuous ψ-functions. We are going to define instrumental M-estimator for discontinuous ψ-function and to show its consistency and asymptotic normality. We shall do it along similar lines as in the case of continuous ψ-function, of course with necessary

modifications. Let us consider the ψ-function which we treated in Theorem 3 and let K be a positive constant. Then for any $n \in N$ and $\vartheta > \frac{1}{2}$ define

$$\tilde{\psi}_n(v) = \frac{1}{2}\left\{\alpha_s + \alpha_{s-1} + (\alpha_s - \alpha_{s-1})\, n^{\vartheta} K^{-1}\,(v - r_s)\right\}$$

$$\text{for } |v - r_s| < K n^{-\vartheta} \text{ and } s \in \{1, 2, ..., m\}$$

and

$$\tilde{\psi}_n(v) = \psi(v) \text{ elsewhere.}$$

(Since ϑ will be assumed to be fix, we have omitted it in the notation for $\tilde{\psi}_n(v)$.)

THEOREM 5. *Let the assumptions of Theorem 3 hold and q (see Theorem 3) be positive definite. Then there is a $\tilde{\beta}^{(n)}$ such that*

$$\sum_{i=1}^{n} Z_i \tilde{\psi}_n\left(\frac{Y_i - X_i^{\mathrm{T}} \tilde{\beta}^{(n)}}{\hat{\sigma}_{(n)}}\right) = 0 \tag{54}$$

and

$$\sqrt{n}\left(\tilde{\beta}^{(n)} - \beta^0\right) = \mathcal{O}_p(1) \quad \textit{as } n \to \infty.$$

Proof. Similarly as in the proof of Theorem 3, without loss of generality let us assume that $m = 1$ (i. e. function ψ has only one discontinuity), $r = 0$, $\alpha_2 - \alpha_1 = \tau$. First of all, let us consider for $\|t\| < M, |u| < M$

$$\sum_{i=1}^{n} Z_i \left\{\psi\left(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right) - \tilde{\psi}\left(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right)\right\}.$$

Since $\psi(v) = \tilde{\psi}_n(v)$ for $|v| \geq K n^{-\vartheta}$, the difference

$$\psi\left(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right) - \tilde{\psi}_n\left(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right)$$

is nonzero only in the case when

$$\left|\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right| < K n^{-\vartheta},$$

i. e. when

$$-K n^{-\vartheta} \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u} + n^{-\frac{1}{2}} X_i^{\mathrm{T}} t < e_i < K n^{-\vartheta} \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u} + n^{-\frac{1}{2}} X_i^{\mathrm{T}} t. \tag{55}$$

According to assumption of Theorem 3 there is $J > 0$ such that $f_{e_i}(v|X_i) < J$. It means that probability of the event (55), independently on X_i, is bounded by $2JKn^{-\vartheta}$. For the notational simplicity let I_i be the indicator of the set

$$\left\{\omega : \psi\left(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right) \neq \tilde{\psi}_n\left(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right)\right\}.$$

Then

$$I\!E\left\{Z_i\left\{\psi\left(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right) - \tilde{\psi}_n\left(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right)\right\}\right\}$$

$$= I\!E\left\{Z_i\left\{\psi\left(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right)\right.\right.$$

$$\left.\left. - \tilde{\psi}_n\left(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}\right)\right\} I_i\right\} < I\!E\|Z_i\| \cdot \tau \cdot JKn^{-\vartheta}.$$

It implies that for any $\varepsilon > 0$

$$P\left(\left\| n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \left\{ \psi(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}) \right.\right.\right.$$
$$\left.\left.\left. -\tilde{\psi}_n(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}) \right\} \right\| > \varepsilon \right)$$
$$\leq \varepsilon^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} I\!E \|Z_i\| \cdot \tau \cdot JK n^{-\vartheta} = o(1) \quad \text{as } n \to \infty.$$

Analogical way leads to

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \left\{ \psi(e_i \sigma_{e_1}^{-1}) - \tilde{\psi}_n(e_i \sigma_{e_1}^{-1}) \right\} = o_p(1) \quad \text{as } n \to \infty. \tag{56}$$

Applying (37) we obtain

$$\sup_{\mathcal{T}_M} \left\| n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \left\{ \tilde{\psi}_n(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}) - \tilde{\psi}_n(e_i \sigma_{e_1}^{-1}) \right\} + qt + \gamma u \right\|$$
$$\leq n^{-\frac{1}{2}} \sup_{\mathcal{T}_M} \left\| \sum_{i=1}^{n} Z_i \left\{ \tilde{\psi}_n(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}) \right.\right.$$
$$\left.\left. -\psi(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u}) \right\} \right\|$$
$$+ n^{-\frac{1}{2}} \sup_{\mathcal{T}_M} \left\| \sum_{i=1}^{n} Z_i \left\{ \tilde{\psi}_n(e_i \sigma_{e_1}^{-1}) - \psi(e_i \sigma_{e_1}^{-1}) \right\} \right\|$$
$$+ \sup_{\mathcal{T}_M} \left\| n^{-\frac{1}{2}} S_n(t,u) + qt + \gamma u \right\| = o_p(1) \quad \text{as } n \to \infty.$$

Taking into account (56) once again we can easy find that

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \tilde{\psi}_n(\left[e_i - n^{-\frac{1}{2}} X_i^{\mathrm{T}} t\right] \sigma_{e_1}^{-1} e^{-n^{-\frac{1}{2}} u})$$
$$= n^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i \psi_n(e_i \sigma_{e_1}^{-1}) - qt - \gamma u + o_p(1) \quad \text{as } n \to \infty \tag{57}$$

where q and γ correspond to ψ. Since the right hand sides of (52) and (57) are of the same character, (57) in fact indicates that we may use the same idea which was used in the proof of Theorem 4. □

Theorem 5 allows us to give definition of instrumental M-estimator for discontinuous ψ-function.

DEFINITION 1. *Under instrumental M-estimator $\hat{\beta}^{(IM,K,n)}$ for any discontinuous ψ-function we shall understand that solution of equation (54) which was described in Theorem 5.*

REMARK 7. *Please notice that in some sense the construction which was presented a few lines above and which was a justification of Definition 1, also gives an idea how to find the instrumental M-estimator. Of course, in the case when (54) has more solutions we have to choose one of them, similarly as in the case when we obtain several solutions of (21). Mentioned*

construction (and namely the proof of the previous theorem) however indicates even more. It is known (and after all, it is clear without any special knowledge) that to find M-estimator generated by a discontinuous ψ-function need not be very simple because it is sometimes (or even ususally) necessary to solve directly corresponding extremal problem and not only an equation of the type (21) which moreover need not have any solution, see e. g. Koenker and Bassett (1978). From previous text however follows that we may find an approximation to M-estimator generated by a ρ-function with discontinuous ψ-function in a way which is used for finding M-estimators generated by continuous ψ-functions (see e. g. Antoch and Víšek (1991)), simply considering a "continuous modification" $\tilde{\psi}_n$ of the function ψ. At this moment we are not able to show generally (after all, this is not the goal of the paper) that this approximation to the M-estimator has asymptotically the same properties as the "precise" M-estimator.

On the other hand, for fixed n and for $K \to 0$, the solutions of (54) converge to solutions of (19) or (20). It supports a hope that for small K the statistical properties of the approximation to "precise" M-estimator will be similar to the properties of that "precise" M-estimator. Moreover, for some ρ with discontinuous ψ we may guarantee that a solution of (19) exists and gives solution of the corresponding extremal problem (18) (of course, if we have more solutions of (19), then one of them is solution of (18)), see Rubio and Víšek (1996). Then, as the next corrolary shows, this solution has the same asymptotic representation as "precise" M-estimator.

Similarly as above simple consequences of Theorem 5 (and of Definition 1) can be given as corollary.

COROLLARY 3. *Let q (see Theorem 3) be a positive definite matrix and $\hat{\sigma}_{(n)}$ a $\sqrt{n}$-consistent estimate of σ_{e_1}. Then under assumptions of Theorem 5 we have*

$$\sqrt{n}\left(\hat{\beta}^{(IM,K,n)} - \beta^0\right) = n^{-\frac{1}{2}} q^{-1} \sum_{i=1}^{n} Z_i \left\{\psi(e_i \sigma_{e_1}^{-1}) - \gamma(\log\hat{\sigma}_{(n)} - \log\sigma_{e_1})\right\} + o_p(1) \quad \text{as } n \to \infty.$$

Moreover, if $\gamma = 0$ $\sqrt{n}\left(\hat{\beta}^{(IM,n)} - \beta^0\right)$ is asymptotically normal with zero mean and a covariance matrix $C = q^{-1} I\!E \left\{\psi^2(e_1 \sigma_{e_1}^{-1}) Z_1 Z_1^{\mathrm{T}}\right\} [q^{-1}]^T$.

4. CONCLUSIONS

The results of paper demonstrated that the instrumental M-estimates can be treated in a similar way as "ordinary" M-estimates and of course, they have similar behaviour as M-estimates. It means that they are able to cope with contamination of data in the same way as M estimators. In other words, they easily overcome an influence of outliers, however somewhat more

attention is necessary to pay to the situations when the diagonal elements of hat matrix indicates that there are some leverage points. Then of course instrumental M-estimates with (strongly) redescending ψ-function should be used. By contrast with "ordinary" M-estimates they are able to cope with the break of orthogonality condition, i. e. they give consistent estimate of regression model also in the case when random fluctuations are correlated with some explanatory variable(s).

Of course, similarly as in the least squares analysis we should have a tool indicating whether the instrumental M-estimate should be used or the "ordinary" M-estimate may suffices. Naturally, the latter estimator is easier to apply and the result is more efficient then the result of the former one in the case that this is appropriate, i. e. when fluctuations and regressors are not correlated. We are aware that the next step is to be generalization of a specification test, e. g. Hausman test. The forthcoming paper submitted to the Prague Stochastics'98 will bring it, see Víšek (1998 b).

The paper brought also something more at two points. Firstly, it showed that the idea of application of Ortega and Rheinboldt's result in order to achieve $\sqrt{n}$-consistency of M-estimators [9] can be also used for discontinuous ψ-functions. So it simplifies and unifies the theory. Unfortunately the same is not yet true about the proofs of asymptotic linearity of (instrumental) M-statistics, where the case of discontinuous ψ-functions still requires Skorohod embedding into Wiener process (the idea is due to Portnoy (1983), compare also Jurečková and Sen (1989)).

Secondly, we have seen in the paper that an approximation to the M-estimators with discontinuous ψ-functions can be found by the same way as the estimators with the continuous ψ-functions. It may considerably simplify their evaluation because the evaluation of M-estimates with continuous ψ-function can be based on the algorithms used for evaluation of the least squares. Such algorithms have been thoroughly studied, carefully implemented and innumerably used and hence they are quick and reliable. Naturally, since the result is asymptotic and moreover it includes a free parameter (the constant K), it requires some numerical study to make an idea about a real possibility to use this way.

5. Appendix

LEMMA A.1. *(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*

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

[9] The idea was used e. g. in Rubio and Víšek (1996) to prove $\sqrt{n}$-consistency of M-estimators with continuous ψ-functions and is due to Jana Jurečková and Stephan Portnoy (personal discussion).

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

REMARK 8. *Since the book of Štěpán (1987) is in Czech language we refer also to Breiman (1968) where however this simple 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 Breinman's book. (See also Theorem 13.6 on the page 276.)*

We shall need however somewhat generalized version of previous lemma.

LEMMA A.2. *Let a and b be positive numbers. Further let ξ be a random variable defined on a probability space $(\Omega, \mathcal{A}, P)$ such that $P(\xi = -a) = \pi_1$, $P(\xi = b) = \pi_2$ and $P(\xi = 0) = \pi_3$, $\pi_j \in (0,1)$ for $j = 1,2,3$ and $\pi_1 + \pi_2 + \pi_3 = 1$. Moreover let $I\!E\xi = 0$. Finally, denoting $A = \{\omega \in \Omega : \xi(\omega) = 0$, put for $\omega \in A$ $\tau(\omega) = 0$ and for $\omega \in A^{\mathrm{C}}$ 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, $I\!E\tau = a \cdot b \cdot (1 - \pi_3) = var\,\xi$.

Proof. Let us put $\tilde{\Omega} = A^{\mathrm{C}}$, $\tilde{\mathcal{A}} = A^{\mathrm{C}} \cap \mathcal{A}$ and $\tilde{P}(B) = \pi_3^{-1} P(B)$. Further let $\tilde{\xi} = \xi$ for $\omega \in A^{\mathrm{C}}$. Then $\tilde{P}(\tilde{\xi} = -a) = \pi_3^{-1} \cdot \pi_1$ and $\tilde{P}(\tilde{\xi} = b) = \pi_3^{-1} \cdot \pi_2$. Let finally $\tilde{\tau}$ be the time for the Wiener process $W(s)$ to exit the interval $(-a, b)$ and put for $\omega \in A^{\mathrm{C}}$ $\tau(\omega) = \tilde{\tau}(\omega)$. According to Lemma A.1 we have $\tilde{\xi} =_{\mathcal{D}} W(\tilde{\tau})$ and $I\!E\tilde{\tau} = a \cdot b = var\,\tilde{\xi}$. Earlier than we shall continue, let us realize that $I\!E\tilde{\tau} = \int_0^\infty z\tilde{f}(z)\mathrm{d}z = a \cdot b$ where $\tilde{f}(z)$ is a density of distribution of $\tilde{\tau}$. Now, evidently $\xi =_{\mathcal{D}} W(\tau)$ because on the set A we have $\xi = 0 = W(\tau)$ and $I\!E\tau = \int_0^\infty z f(z)\mathrm{d}z = 0 \cdot P(A) + \int_0^\infty z\tilde{f}(z)(1 - \pi_3)\mathrm{d}z = a \cdot b \cdot (1 - \pi_3) = var\,\xi$. □

ASSERTION 1. *Let U be an open, bounded set in R^p and assume that $Q(z) : \bar{U} \subset R^p \to R^p$ ($\bar{U}$ is the closure of U) is continuous and satisfies $(z - z_0)^T Q(z) \geq 0$ for some $z_0 \in U$ and all $z \in \bar{U} \setminus U$. Then the equation $Q(z) = 0$ has a solution in $\bar{U}$.*

For the proof see Ortega and Rheinboldt (1970) , assertion 6.3.4 on the page 163.

References

[1] Antoch J., Víšek J. Á. (1991): Robust estimation in linear models and its computational aspects. *Contributions to Statistics: Computational Aspects of Model Choice, Springer Verlag, (1992), ed. J. Antoch, 39 - 104.*

[2] Bickel P. J. (1975): One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc. 70, 428–433.*

[3] Bowden R. J., Turkington D. A. (1984): *Instrumental Variables.* Cambridge: Cambridge University Press.

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

[5] Drygas H. (1976): Weak and strong consistency of the least squares estimates in regression models. *Wahrsch. Verw. Gebiete 34, 119 - 127.*

[6] Fisher R. A. (1922): On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A 222, pp. 309–368.*

[7] Friedman M. (1957): *A Theory of the Consumption Function.* Princeton: Princeton University Press.

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

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

[10] Hausman J. (1978): Specification test in econometrics. *Econometrica, 4, 1978, 1251 - 1271.*

[11] Hettmansperger T. P., Sheather S. J. (1992): A Cautionary Note on the Method of Least Median Squares. *The American Statistician 46, 79–83.*
[12] Hildreth C., Houck J. P. (1968): Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association 63, 584 –595.*
[13] Huber P. J.(1981): *Robust Statistics.* New York: J.Wiley *&* Sons.
[14] Judge G. G., Griffiths W. E., Hill R. C., Lütkepohl H., Lee T. Ch (1985): *The Theory and Practice of Econometrics.* New York: J.Wiley *&* Sons (second edition).
[15] Jurečková J. and Procházka B. (1994): Regression quantiles and trimmed least squares estimator in nonlinear regression model. *Nonparametric Statistics, vol. 3., pp. 201 - 222.*
[16] Jurečková J., Sen P. K. (1989): Uniform second order asymptotic linearity of M-statistics in linear models. *Statistics and Decisions 7, 263 - 276*
[17] Jurečková J. and Sen P. K. (1993): Regression rank scores scale statistics and studentization in linear models. *Proceedings of the Fifth Prague Symposium on Asymptotic Statistics, Physica Verlag, 111-121.*
[18] Jurečková J., Welsh A. H. (1990): Asymptotic relations between L- and M-estimators in the linear model. *Annals of the Institute of Statistical Mathematics, 42, 671–698.*
[19] Kmenta J. (1986): *Elements of Econometrics.* New York: Macmillan Publishing Company.
[20] Koenker R., Bassett G. (1978): Regression quantiles. *Econometrica, 46, 33-50.*
[21] Ortega J. M., Rheinboldt W. C. (1970): *Iterative Solution of Nonlinear Equations in Several Variables*, New York and London: Academic Press.
[22] Portnoy S. (1983): Tightness of the sequence of empiric c.d.f. processes defined from regression fractiles. In *Robust and Nonlinear Time-Series Analysis (J. Franke, W. Härdle, D. Martin eds.), 231 - 246. Springer-Verlag, New York, 1983.*
[23] Ramsey J. B. (1974): Classical model selection through specification error tests. *Frontiers of Econometrics.* Ed. P. Zarembka., New York: Academic Press.
[24] Rao R. C., Zhao LC. (1992): On the consistency of M-estimate in linear model obtained through an estimating equation. *Statistics & Probability Letters 14, 79 - 84.*
[25] Rubio A. M., Víšek J. Á. (1996): A note on asymptotic linearity of M-statistics in non-linear models. *Kybernetika, 32(1996), 353-374.*
[26] Štěpán J. (1987): *Teorie pravděpodobnosti (Probability Theory).* Prague: Academia.
[27] Víšek J. Á. (1992): A cautionary note on the method of Least Median of Squares reconsidered, *Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Prague, 1994, 254 - 259.*
[28] Víšek J.Á.(1996 a): On high breakdown point estimation. *Computational Statistics (1996) 11:137-146, Berlin.*
[29] Víšek, J. Á. (1996 b): Sensitivity analysis M-estimates. *Annals of the Institute of Statistical Mathematics, 48(1996), 469-495.*
[30] Víšek J.Á.(1997 a): On the diversity of estimates. Submitted to *Computational Statistics and Data Analysis.*
[31] Víšek J.Á.(1997 b): *Ekonomerie I* (Econom trics, in czech) Carolinum, publishing house of Charles university.
[32] Víšek J.Á.(1998 a): A simple scale-invariant and regression-equivariant estimator of scale. *Preprint*
[33] Víšek J.Á.(1998 b): Robust specification test. Submitted to the proceedings of the Prague Asymptotics'98.
[34] Wu D. (1973): Alternative tests of independence between stochastic regressors and disturbances. *Econometrica, 41, 733 - 750.*