PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 21, Fasc. 2 (2001), pp. 467–492

SELECTING REGRESSION MODEL

BY

J. Á. VÍŠEK* (PRAGUE)

Abstract. A new tool for the identification of regression model is proposed and its properties are established. The key importance of the new tool is that it is able to solve still not very well-known problem of *diversity of estimates*, as described in Víšek [22] and [25]. Main idea of the proposal is as follows. Having evaluated an estimate of regression coefficients for given data, the data are partitioned into two disjoint subsets (e.g. by a geometric rule applied in the factor space). Then for each subset of corresponding residuals we evaluate the estimate of their density, e.g. the kernel one. If the estimate of regression model is "near to the true model", the density of disturbances is the same in the both subsets, and hence also the estimates of density of residuals are approximately equal each to other. Therefore, finally, the estimates of density are compared by means of the weighted Hellinger distance. It implies that the significant difference between the estimates of density indicates that the given estimate of the regression model is not near to the "true" model or, in other words, that it is not "adequate" for the data. In the case when we have at our disposal more estimates of the regression model, and especially when the estimates are considerably different (each from other), the test statistic may be also used for selecting the estimate of the regression model. We just accept the estimate with the smallest weighted Hellinger distance. The result of the paper is illustrated by two simple numerical examples demonstrating especially the sensitivity of the test statistic to the difference between the estimates of density.

AMS Subject Classification: 62 J 05, 62 J 20.

Key words and phrases: Weighted Hellinger distance, diagnostics and choice of model, diversity of (robust) estimates.

INTRODUCTION

The second half of the past century has brought into the statistics a lot of new methods. Let us recall e.g. an ample offer of robust methods. So, many statistical, especially classical problems can be treated by several methods.

* Research was supported by grant of GA UK number 255/20000/A EK/FSV.

Sometimes this offer of methods seems to be so wide that it is difficult to select, for given data, the best one.

Let us look on the situation more closely. However, prior to continuing, let us make two technical remarks.

Firstly, in the paper the problem of selecting the "best possible" estimate of the regression model for given data will be treated. Hence, let us restrict ourselves in this introductory text also on regression analysis.

Secondly, to avoid a misunderstanding hereafter the word *estimate* will be used exclusively for the value of the *estimator* at given data.

It is not difficult to learn that the conditions, under which the estimator of a regression model is asymptotically consistent and even normal, are rather weak (e.g. the conditions for a wide class of M-estimators may be found in Maronna and Yohai [13], Jurečková [9] or [10] or for LTS-estimator see Víšek [24], etc.). One may also easily find that having adopted some optimality criteria, like $\mathscr{B}$-robustness or $\mathscr{V}$-robustness, minimax bias etc., results proving the optimality of some classes of estimators are available (Hampel et al. [6], Martin et al. [14], Yohai and Zamar [27], etc.). Nevertheless, even if we restrict ourselves to some such classes, the range of methods which can be applied is still very wide. So, there is still a possibility to choose, and hence we may hope to be able to select the method not only according to some general principles but also adaptively to the data (Hogg [7]).

It is easy to verify (by a "numerical" experiments) that one may obtain for a one set of data several estimates of the model, in which the values of regression coefficients are considerably different (Víšek [17] or [25]). The difference may amount to the hundreds of percents of the value of one estimate with respect to the other estimate of the same coefficient. All the estimates of the model having been, of course, evaluated for the same data, i.e. for the same response variables and for the same set of regressors they have been only produced by different algorithms. Moreover, all of these estimates pass (typically without difficulties) some test as F-test or for the robust methods some τ-test; see Markatou et al. [12].

It may be of interest that the diversity of the estimates of the same regression model need not cause necessarily any problem what concerns the efficiency of estimation because the efficiency of the various estimators need not be considerably different (Víšek [19] or [23]). However, we may get in serious difficulties when explaining or interpreting a structure of data, and also the quality (or reliability) of prediction may be problematic. In Víšek [20] the example of data is presented for which the Least Median of Squares (LMS) and the Least Trimmed Squares (LTS) estimates are orthogonal each to other. An example of similar effect for real data was presented also in Víšek [20].

Of course, one may argue that when processing real data we may meet with such a situation rarely. It may be or need not be true but it does not help in the situation when we process one, unique sample of data. Then any considerations about the frequency of meeting such strange data are helpless. So we really stay in front of a serious problem **what to do when for the given data we obtain several, very different estimates of models.**

An advice (implicitly) given, e.g. in Rousseeuw and Leroy [15], Chapter 6, is to use a method with high breakdown point as a diagnostic tool, and then to select an M-estimate which is near to this one with high breakdown point. Another proposed possibility is to use a one-step M-estimate starting again from an estimate with high breakdown point (Hampel et al. [6], p. 330, Jurečková and Portnoy [11]). As follows from the example (which was already mentioned above) with LMS orthogonal to LTS (Víšek [20]), these advices are not helping too much. (For numerical examples of the diversity of one-step M-estimates, LTS and LMS estimates see also Boček and Víšek [3], [21] or [22].)

Of course, we usually select from the evaluated estimates of model according to some objectives or ideas commonly accepted in the branch of science the data came from. Sometimes however a statistically oriented tool may give a hint in a form of rejecting a hypothesis that the respective model has generated data.

In the present paper we propose a statistic which may be used to reject those estimates which are not "adequate" for given data. The statistic is based on the weighted Hellinger distance between the estimates of density of residuals in two halves of sample. These two halves of data may be created e.g. by a "natural" geometric rule. The statistic can be also used to select the estimate of the regression model which fits to the data in the best way. From the all evaluated estimates we just accept that one for which the weighted Hellinger distance reaches its minimum.

Let us give now some basic notation.

1. TEST STATISTICS

Let R be the real line and N the set of all positive integers. For any $n \in N$ let R^n denote the n-dimensional Euclidean space. We shall consider for all $i \in N$ a linear regression model with deterministic carriers

$$Y_i = X_i^T \cdot \beta^0 + e_i, \quad i = 1, 2, \ldots, \tag{1.1}$$

where Y_i is a response variable, $X_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$ is the i-th row (assumed as a column vector) of the design matrix (if the intercept is assumed, then we suppose $x_{i1} = 1$ for $i = 1, 2, \ldots$), $\beta^0 = (\beta_1^0, \beta_2^0, \ldots, \beta_p^0)^T$ is the vector of regression coefficients (unknown but fix), and $\{e_i\}_{i=1}^{\infty}$ is a sequence of independent and identically distributed random variables – distributed according to a distribution function F. The distribution function is assumed to be absolutely continuous with a finite supremum of its density f, say $\sup_{x \in R} f(x) < V < \infty$.

The matrix form of model (1.1) is then

$$Y = X\beta^0 + e.$$

Let w be a kernel, $w\colon R \to [0, \infty)$ and, for any $c \in [0, \infty)$, $y \in R$, $Y^* = (Y_l, Y_{l+1}, \ldots, Y_k)^T$, $1 \leqslant l \leqslant k \leqslant 2n$, and $\beta \in R^p$ define the kernel estimator of the density

$$\hat{f}_{l,k}(y, Y^*, \beta, c) = \frac{1}{(k-l)\,c} \sum_{i=l}^{k} w\big(c^{-1}(y - Y_i + X_i^T \beta)\big).$$

Assume that we have at our disposal $2n$ observations and put

$$Y^{(n,1)} = (Y_1, Y_2, \ldots, Y_n)^T, \qquad Y^{(n,2)} = (Y_{n+1}, Y_{n+2}, \ldots, Y_{2n})^T$$

and

$$Y^{(n)} = ([Y^{(n,1)}]^T, [Y^{(n,2)}]^T)^T.$$

Similarly for the corresponding design matrices $X^{(n,1)}$ and $X^{(n,2)}$. We shall consider for some sequence $\{c_n\}_{n=1}^{\infty} \searrow 0$ of the real numbers the kernel estimators

$$\hat{f}_{1,n}(y, Y^{(n,1)}, \beta, c_n), \qquad \hat{f}_{n+1,2n}(y, Y^{(n,2)}, \beta, c_n) \quad \text{and} \quad \hat{f}_{1,2n}(y, Y^{(n)}, \beta, c_n).$$

In what follows let us write briefly Y, $Y^{(1)}$ and $Y^{(2)}$ instead of $Y^{(n)}$, $Y^{(n,1)}$ and $Y^{(n,2)}$, respectively; similarly for the design matrix X and its partition. Throughout the paper the sequence $\{c_n\}_{n=1}^{\infty}$ will be fix, and hence we shall abbreviate the notation of the kernel estimators as $\hat{f}_n(y, Y^{(1)}, \beta)$ instead of full $\hat{f}_{1,n}(y, Y^{(n,1)}, \beta, c_n)$, $\hat{f}_n(y, Y^{(2)}, \beta)$ instead of $\hat{f}_{n+1,2n}(y, Y^{(n,2)}, \beta, c_n)$, etc. Finally, for a sequence of positive numbers $\{a_n\}_{n=1}^{\infty} \nearrow \infty$ and $\beta \in R^p$ define the statistics of the Hellinger type

$$H_n(Y, \beta) = n \int_{-a_n}^{a_n} [\hat{f}_n^{1/2}(y, Y^{(1)}, \beta) - \hat{f}_n^{1/2}(y, Y^{(2)}, \beta)]^2 \hat{f}_{2n}(y, Y, \beta)\, dy.$$

Evidently, $H_n(Y, \beta)$ is not precisely the Hellinger distance of respective estimators of density. We have included into the formula which determines $H_n(Y, \beta)$ a "weight function" in the form of $\hat{f}_{2n}(y, Y, \beta)$. We have done it to be able to cope with some technicalities. On the other hand, a heuristic justification of such a step is however rather straightforward. As the kernel estimates are "the most unreliable" at their tails. It is hence necessary to take the information about the density at the tails with a caution or, in other words, the information at the tails of the estimators is to be somewhat weighted down.

2. ASYMPTOTIC DISTRIBUTION OF TEST STATISTICS

The test which will be proposed below will be based on the following:

THEOREM 1. *Let the following conditions be fulfilled*:

(i) *The density f has a finite supremum and the bounded second derivative.*

(ii) *The sequences* $\{a_n\}_{n=1}^{\infty}$, $\{c_n\}_{n=1}^{\infty}$ *and* $\{d_n\}_{n=1}^{\infty}$ *are such that*
(a) $(-a_n, a_n) \subset \{y \in R\colon f(y) > d_n\}$,

$$a_{n+1} \geqslant a_n \quad \textit{and} \quad (-a_n, a_n) \nearrow \{y \in R\colon f(y) > 0\} \ \textit{as} \ n \to \infty,$$

(b) $c_n = o(1)$, $d_n c_n^{-1} = O(1)$ *as* $n \to \infty$,
(c) *there exists* $\tau \in (0, 1)$ *so that*

$$c_n^{-1/2} \int_{|y| > (1-\tau)a_n} f(y)\,dy = o(1) \quad \textit{as} \ n \to \infty,$$

(d) $n^{-1} c_n^{-4} a_n^2 = o(1)$ *and* $n^{-1} c_n^{-3} a_n d_n^{-1} = o(1)$ *as* $n \to \infty$.
(iii) *For the kernel* $w(t)$ *we have*

$$\lim_{|t| \to \infty} t^4 w(t) = 0, \quad \sup_{t \in R} w(t) < K < \infty \quad \textit{and} \quad \int_{-\infty}^{\infty} t w(t)\,dt = 0,$$

and there exists $\{L_n\}_{n=1}^{\infty}$ *such that, for* τ *as in* (ii) (c),

$$c_n^{-1/2} \int_{|t| > L_n} w^2(t)\,dt = o(1) \ \textit{as} \ n \to \infty \quad \textit{and} \quad \limsup_{n \to \infty} c_n L_n a_n^{-1} < \tau.$$

Then the asymptotic distribution of

$$\mathscr{H}_n(Y, \beta^0) = \Delta_n^{-1}\{H_n(Y, \beta^0) - m_n\}$$

is $N(0, 1)$, *where*

$$m_n = \frac{1}{2} c_n^{-1} \int_{-\infty}^{\infty} w^2(t)\,dt$$

and

$$\Delta_n^2 = \frac{1}{2} c_n^{-1} \int_{-\infty}^{\infty} f^2(y)\,dy \int_{-\infty}^{\infty} \Big\{\int_{-\infty}^{\infty} w(t) w(t+z)\,dt\Big\}^2 dz.$$

For the proof of Theorem 1 we shall prepare several assertions. However, prior to doing that let us discuss the assumptions of the theorem because it need not be clear immediately whether they can be fulfilled.

Remark 1. Firstly, to construct the sequences $\{a_n\}_{n=1}^{\infty}$, $\{c_n\}_{n=1}^{\infty}$ and $\{d_n\}_{n=1}^{\infty}$ in order to meet assumption (ii) of Theorem 1 we need to start from $\{d_n\}_{n=1}^{\infty}$. Then we may find $\{a_n\}_{n=1}^{\infty}$ (under some mild and acceptable assumption on the shape of f, e.g. that $\{y\colon f(y) > d_n\}$ is an interval) to fulfill (ii)(a). Now we need to find $\{c_n\}_{n=1}^{\infty}$ so that $n^{-1} c_n^{-4} a_n^2 = o(1)$. If it is not possible, we have to change $\{a_n\}_{n=1}^{\infty}$ to some $\{\tilde{a}_n\}_{n=1}^{\infty}$ so that (ii)(a) is fulfilled, $\tilde{a}_n \leqslant a_n$ and there is $\{c_n\}_{n=1}^{\infty}$ so that $n^{-1} c_n^{-4} \tilde{a}_n^2 = o(1)$. May be that we obtain a sequence $\{\tilde{a}_n\}_{n=1}^{\infty}$ such that $(-\tilde{a}_n, \tilde{a}_n\} \nearrow \{y \in R\colon f(y) > 0\}$ rather slowly. Nevertheless, we may always select $\tilde{c}_n \searrow 0$ so that $\tilde{c}_n \geqslant c_n$ and (ii)(b) and (ii)(c) are also fulfilled. If now the second requirement in (ii)(d) does not hold, we have to take some $\{d_n^*\}_{n=1}^{\infty}$

so that $d_n^* > d_n$ for all $n \in N$ to meet (ii)(d) (and naturally to keep $d_n^* \searrow 0$). This is also always possible. Now we have to return to (ii)(a) and may be that we will have to change the sequence $\{\tilde{a}_n\}_{n=1}^{\infty}$ to $\{a_n^*\}_{n=1}^{\infty}$ so that $a_n^* < \tilde{a}_n$ to fulfill (ii)(a). Similarly, it may happen that we will be forced to change $\{\tilde{c}_n\}_{n=1}^{\infty}$ so that $c_n^* > \tilde{c}_n$ for all $n \in N$ and that (ii)(b) and (ii)(c) will be fulfilled (this is possible for any $\{a_n^*\}_{n=1}^{\infty}$, a_n^* increasing to the upper bound of the support of $f(y)$). But then (ii)(d) keeps to hold also for $\{a_n^*\}_{n=1}^{\infty}$, $\{c_n^*\}_{n=1}^{\infty}$ and $\{d_n^*\}_{n=1}^{\infty}$. Now, we may select the kernel w (with sufficiently modest tails) so that the assumption (iii) is fulfilled. This means that the assumptions of Theorem 1 are not surely contradictory and, moreover, for the most of densities they can be easily fulfilled.

Secondly, it may seem that the second part of (ii)(b) together with the first part of (ii)(d) implies the second part of (ii)(d). Of course, this depends on the fact how the second part of the assumption (ii)(b) is understood. If we assume that it means that there is $n_0 \in N$ and $k > 0$, $K < \infty$ so that for all $n > n_0$ we have $k < d_n c_n^{-1} < K$, then really the second part of (ii)(d) is redundant. If, however, we assume that for all $n > n_0$ we have only $d_n c_n^{-1} < K$, then it may happen that $d_n c_n^{-1} \searrow 0$ as $n \to \infty$, and then even under $n^{-1} c_n^{-4} a_n^2 = o(1)$ we can have $n^{-1} c_n^{-3} a_n d_n^{-1} \nearrow \infty$ as $n \to \infty$. After all, the assumptions, except of the second part of (ii)(d), "allow" d_n's to converge to zero rather fast and it would cause that the second part of (ii)(d) can be violated. ∎

For the proof of the following assertion see Csörgő and Révész [5], Lemma 6.1.2.

Assertion 1 (Csörgő and Révész). *Let the kernel $w(t)$ be bounded with*

$$\lim_{|t| \to \infty} t^4 w(t) = 0 \quad \textit{and} \quad \int_{-\infty}^{\infty} t w(t)\, dt = 0.$$

Moreover, assume that the density $f(y)$ has a bounded second derivative on an interval $-\infty \leqslant A < B \leqslant \infty$. Then for any $\varepsilon > 0$

$$\sup_{A+\varepsilon \leqslant y \leqslant B-\varepsilon} |E\hat{f}_n(y, Y, \beta^0) - f(y)| = O(c_n^2).$$

Lemma 1. *Let the assumptions of Theorem 1 be fulfilled. Then*

$$n \int_{-a_n}^{a_n} [\hat{f}_n^{1/2}(y, Y^{(1)}, \beta^0) - \hat{f}_n^{1/2}(y, Y^{(2)}, \beta^0)]^2 [\hat{f}_{2n}(y, Y, \beta^0) - E\hat{f}_{2n}(y, Y, \beta^0)]\, dy = \mathcal{O}_p(n^{-1/2} c_n^{-2} a_n) \quad \textit{as } n \to \infty.$$

Proof. Making use of the Schwarz inequality we find an upper bound of the squared value on the left-hand side in the form

$$(2.1) \quad n^2 \int_{-a_n}^{a_n} [\hat{f}_n^{1/2}(y, Y^{(1)}, \beta^0) - \hat{f}_n^{1/2}(y, Y^{(2)}, \beta^0)]^4 E\hat{f}_{2n}(y, Y, \beta^0)\, dy \times \int_{-a_n}^{a_n} [\hat{f}_{2n}(y, Y, \beta^0) - E\hat{f}_{2n}(y, Y, \beta^0)]^2 E^{-1}\hat{f}_{2n}(y, Y, \beta^0)\, dy.$$

Now we easily verify that

$$(2.2)\qquad E\hat{f}_n(y,\, Y^{(1)},\, \beta^0) = \int_{-\infty}^{\infty} \left\{\frac{1}{nc_n}\sum_{i=1}^{n} w\big(c_n^{-1}(y-z)\big)\right\} f(z)\,dz$$

$$= \frac{1}{c_n}\int_{-\infty}^{\infty} w\big(c_n^{-1}(y-z)\big) f(z)\,dz = E\hat{f}_n(y,\, Y^{(2)},\, \beta^0) = E\hat{f}_{2n}(y,\, Y,\, \beta^0).$$

Consequently, the first factor of (2.1) may be written as

$$(2.3)\qquad n^2 \int_{-a_n}^{a_n} \{[\hat{f}_n^{1/2}(y,\, Y^{(1)},\, \beta^0) - E^{1/2}\hat{f}_n(y,\, Y^{(1)},\, \beta^0)]$$

$$-[\hat{f}_n^{1/2}(y,\, Y^{(2)},\, \beta^0) - E^{1/2}\hat{f}_n(y,\, Y^{(2)},\, \beta^0)]\}^4\, E\hat{f}_{2n}(y,\, Y,\, \beta^0)\,dy$$

$$\leqslant 8n^2 \Big\{ \int_{-a_n}^{a_n} [\hat{f}_n^{1/2}(y,\, Y^{(1)},\, \beta^0) - E^{1/2}\hat{f}_n(y,\, Y^{(1)},\, \beta^0)]^4\, E\hat{f}_n(y,\, Y^{(1)},\, \beta^0)\,dy$$

$$+ \int_{-a_n}^{a_n} [\hat{f}_n^{1/2}(y,\, Y^{(2)},\, \beta^0) - E^{1/2}\hat{f}_n(y,\, Y^{(2)},\, \beta^0)]^4\, E\hat{f}_n(y,\, Y^{(2)},\, \beta^0)\,dy\Big\}.$$

We shall use now the inequality $(a-b)^2 \leqslant b^{-2}(a^2-b^2)^2$ valid for $a \geqslant 0$ and $b > 0$. We obtain again an upper bound for both the terms in (2.3) in the form

$$8n^2 \int_{-a_n}^{a_n} E^{-1}\hat{f}_n(y,\, Y^{(j)},\, \beta^0)\,[\hat{f}_n(y,\, Y^{(j)},\, \beta^0) - E\hat{f}_n(y,\, Y^{(j)},\, \beta^0)]^4\,dy, \qquad j = 1, 2.$$

Now,

$$(2.4)\qquad P\left\{8n^2 \int_{-a_n}^{a_n} E^{-1}\hat{f}_n(y,\, Y^{(j)},\, \beta^0)\,[\hat{f}_n(y,\, Y^{(j)},\, \beta^0) - E\hat{f}_n(y,\, Y^{(j)},\, \beta^0)]^4\,dy \geqslant L\frac{a_n}{c_n^3}\right\}$$

$$\leqslant \frac{8n^2 c_n^3}{La_n} \int_{-a_n}^{a_n} E^{-1}\hat{f}_n(y,\, Y^{(j)},\, \beta^0)\, E[\hat{f}_n(y,\, Y^{(j)},\, \beta^0) - E\hat{f}_n(y,\, Y^{(j)},\, \beta^0)]^4\,dy$$

$$= \frac{8n^2 c_n^3}{La_n} \int_{-a_n}^{a_n} E^{-1}\hat{f}_n \left\{\frac{1}{n^3 c_n^4} E[w-Ew]^4 + \frac{6}{n^2 c_n^4}\{E[w-Ew]^2\}^2\right\} dy,$$

where $L \in R$ is arbitrary. Further (please, keep in mind that $w(t) \geqslant 0$)

$$(2.5)\quad E[w-Ew]^4 \leqslant E\{[w-Ew]^2[w-Ew]^2\} \leqslant 4[\sup_{x\in R} w(z)]^2 E[w-Ew]^2$$

$$\leqslant 4\sup_{z\in R}{}^2\, w(z)\, Ew^2 \leqslant \sup_{z\in R}{}^3\, w(z)\, Ew = 4c_n \sup_{z\in R}{}^3\, w(z)\, E\hat{f}_n,$$

and also

$$E[w-Ew]^2 \leqslant Ew^2 \leqslant \sup_{z\in R}{}^2\, w(z),$$

and so

$$\{E[w-Ew]^2\}^2 \leqslant c_n \sup_{z\in R}{}^3 w(z) E\hat{f}_n. \tag{2.6}$$

Consequently, taking into account (2.5) and (2.6) we conclude that (2.4) is proportional to L^{-1}, and hence the first factor of (2.1) has order $\mathcal{O}_p(c_n^{-3} a_n)$. Due to the independence of e_i's, for the second factor of (2.1) we have (again $L \in R$, arbitrary)

$$P\{\int_{-a_n}^{a_n} [\hat{f}_{2n}(y, Y, \beta^0) - E\hat{f}_{2n}(y, Y, \beta^0)]^2 E^{-1}\hat{f}_{2n}\, dy > Ln^{-1} c_n^{-1} a_n\}$$
$$\leqslant (4Lc_n^{-1} a_n)^{-1} \int_{-a_n}^{a_n} E^{-1}\hat{f}_{2n}\, E[w-Ew]^2\, dy \leqslant (2L)^{-1} \sup_{z\in Z} w(z).$$

Since L was arbitrary, we have the second factor of (2.1) of order $\mathcal{O}_p(n^{-1} c_n^{-1} a_n)$. Now, taking into account that we have considered the squared left-hand side of the expression given in the assertion of the lemma we complete the proof. ∎

LEMMA 2. *Let the assumptions of Theorem* 1 *be fulfilled. Then for all combinations of* $i, j = 1, 2$ *we have*

$$n \int_{-a_n}^{a_n} [\hat{f}_n^{1/2}(y, Y^{(i)}, \beta^0) - E^{1/2}\hat{f}_n(y, Y^{(i)}, \beta^0)]^2 [\hat{f}_n(y, Y^{(j)}, \beta^0) - E\hat{f}_n(y, Y^{(j)}, \beta^0)]$$
$$= \mathcal{O}_p(n^{-1/2} c_n^{-2} a_n).$$

Proof. As in the proof of Lemma 1 we obtain a boundary for the squared left-hand side in the form

$$n^2 \int_{-a_n}^{a_n} [\hat{f}_n^{1/2}(y, Y^{(i)}, \beta^0) - E^{1/2}\hat{f}_n(y, Y^{(i)}, \beta^0)]^4 E\hat{f}_n(y, Y^{(i)}, \beta^0)\, dy$$
$$\times \int_{-a_n}^{a_n} [\hat{f}_n(y, Y^{(j)}, \beta^0) - E\hat{f}_n(y, Y^{(j)}, \beta^0)]^2 E^{-1}\hat{f}_n(y, Y^{(j)}, \beta^0)\, dy$$

(keep in mind (2.2)). The rest of the proof uses the steps which were performed in the proof of the previous lemma.

LEMMA 3. *Let the assumptions of Theorem* 1 *be fulfilled. Then for* $i = 1, 2$ *we obtain*

$$n \int_{-a_n}^{a_n} [\hat{f}_n^{1/2}(y, Y^{(i)}, \beta^0) - E^{1/2}\hat{f}_n(y, Y^{(i)}, \beta^0)]^4\, dy = \mathcal{O}_p(n^{-1/2} c_n^{-3} a_n d_n^{-1}). \tag{2.7}$$

Proof. We shall use again the inequality $(a-b)^2 \leqslant b^{-2}(a^2-b^2)^2$, $a \geqslant 0$, $b > 0$. We infer then that the left-hand side of (2.7) is bounded by

$$n \int_{-a_n}^{a_n} E^{-2}\hat{f}_n(y, Y^{(i)}, \beta^0) [\hat{f}_n(y, Y^{(i)}, \beta^0) - E\hat{f}_n(y, Y^{(i)}, \beta^0)]^4\, dy.$$

Now, using Assertion 1 and the assumption (ii)(b) we may find $n_0 \in N$ so that for all $n \geqslant n_0$ we have

$$\sup_{y\in R} |E\hat{f}_n(y, Y^{(i)}, \beta^0) - f(y)| \leqslant \tfrac{1}{2} d_n,$$

and hence

$$\int_{-a_n}^{a_n} E^{-2}\hat{f}_n(y, Y^{(i)}, \beta^0)[\hat{f}_n(y, Y^{(i)}, \beta^0) - E\hat{f}_n(y, Y^{(i)}, \beta^0)]^4\, dy$$
$$\leqslant 2d_n^{-1} \int_{-a_n}^{a_n} E^{-1}\hat{f}_n(y, Y^{(i)}, \beta^0)[\hat{f}_n(y, Y^{(i)}, \beta^0) - E\hat{f}_n(y, Y^{(i)}, \beta^0)]^4\, dy.$$

Now

$$E[\hat{f}_n(y, Y^{(i)}, \beta^0) - E\hat{f}_n(y, Y^{(i)}, \beta^0)]^4 = c_n^{-4}\{n^{-3}E[w-Ew]^4 + n^{-2}E^2[w-Ew]^2\}.$$

Making use of the inequalities

$$E[w-Ew]^4 \leqslant 4[Ew^2 + E^2 w]\sup_{y\in R}{}^2 w(y) \leqslant 8c_n(\sup_{y\in R}{}^3 w(y))E\hat{f}_n(y, Y^{(i)}, \beta^0),$$

$$E^2[w-Ew]^2 \leqslant 4[Ew^2 + E^2 w]\sup_{y\in R}{}^2 w(y) \leqslant 8c_n(\sup_{y\in R}{}^3 w(y))E\hat{f}_n(y, Y^{(i)}, \beta^0)$$

and (ii)(a) of Theorem 1, we find (as in (2.4)) that

$$(2.8)\qquad 2\int_{-a_n}^{a_n} E^{-1}\hat{f}_n(y, Y^{(i)}, \beta^0)E[\hat{f}_n(y, Y^{(i)}, \beta^0) - E\hat{f}_n(y, Y^{(i)}, \beta^0)]^4\, dy$$
$$\leqslant 128n^{-2}a_n c_n^{-3}\sup_{y\in R}{}^3 w(y).$$

Now let the set $\mathscr{D}_{ni}$ be determined as

$$\{\omega\colon n\int_{-a_n}^{a_n} E^{-2}\hat{f}_n(y, Y^{(i)}, \beta^0)[\hat{f}_n(y, Y^{(i)}, \beta^0) - E\hat{f}_n(y, Y^{(i)}, \beta^0)]^4\, dy$$
$$\geqslant Kn^{-1/2}c_n^{-3}a_n d_n^{-1}\}$$

for a positive constant K. Taking into account (2.8), we have

$$P(\mathscr{D}_{ni}) \leqslant K^{-1}n^{3/2}c_n^3 a_n^{-1}d_n\int_{-a_n}^{a_n} E^{-2}\hat{f}_n(y, Y^{(i)}, \beta^0)$$
$$\times E[\hat{f}_n(y, Y^{(i)}, \beta^0) - E\hat{f}_n(y, Y^{(i)}, \beta^0)]^4\, dy$$
$$\leqslant K^{-1}n^{3/2}c_n^3 a_n^{-1}\int_{-a_n}^{a_n} E^{-1}\hat{f}_n(y, Y^{(i)}, \beta^0)E[\hat{f}_n(y, Y^{(i)}, \beta^0) - E\hat{f}_n(y, Y^{(i)}, \beta^0)]^4\, dy$$
$$\leqslant 128K^{-1}n^{-1/2}\sup_{y\in R}{}^3 w(y)$$

and the proof is complete. ∎

LEMMA 4. *Under the assumptions of Theorem* 1 *we have*

$$n \int_{-a_n}^{a_n} [\hat{f}_n^{1/2}(y, Y^{(1)}, \beta^0) - E^{1/2}\hat{f}_n(y, Y^{(1)}, \beta^0)]^2$$
$$\times [\hat{f}_n^{1/2}(y, Y^{(2)}, \beta^0) - E^{1/2}\hat{f}_n(y, Y^{(2)}, \beta^0)]^2\, dy = \mathcal{O}_p(n^{-1} c_n^{-2} a_n).$$

Proof. Let $L \in R$ be arbitrary. Using once again the inequality $(a-b)^2 < b^{-2}(a^2-b^2)$ ($a \geqslant 0$ and $b > 0$), we obtain

$$P\{n \int_{-a_n}^{a_n} [\hat{f}_n^{1/2}(y, Y^{(1)}, \beta^0) - E^{1/2}\hat{f}_n(y, Y^{(1)}, \beta^0)]^2$$
$$\times [\hat{f}_n^{1/2}(y, Y^{(2)}, \beta^0) - E^{1/2}\hat{f}_n(y, Y^{(2)}, \beta^0)]^2\, dy \geqslant L n^{-1} c_n^{-2} a_n\}$$
$$\leqslant \frac{n^2 c_n^2}{L a_n} \int_{-a_n}^{a_n} \Big\{ \frac{1}{E\hat{f}_n(y, Y^{(1)}, \beta^0)} \frac{1}{E\hat{f}_n(y, Y^{(2)}, \beta^0)}$$
$$\times E[\hat{f}_n(y, Y^{(1)}, \beta^0) - E\hat{f}_n(y, Y^{(1)}, \beta^0)]^2$$
$$\times E[\hat{f}_n(y, Y^{(2)}, \beta^0) - E\hat{f}_n(y, Y^{(2)}, \beta^0)]^2 \Big\} dy$$
$$= \frac{1}{L a_n c_n^2} \int_{-a_n}^{a_n} \frac{1}{E^2 \hat{f}_n} \{E[w - Ew]^2\}^2 \leqslant (\sup_{z \in R}{}^2 w(z)) \frac{2E^2 w}{L c_n^2 E^2 \hat{f}_n} \leqslant \frac{2}{L} \sup_{z \in R}{}^2 w(z).$$

Since L was arbitrary, the proof is complete. ∎

The proof of the next lemma will be based on Theorem 6.1.2 of Csörgő and Révész [5]. For convenience of the reader we will give this theorem here as an assertion without proof.

ASSERTION 2 (Csörgő and Révész). *Let* $G_n(x)$ $(-\infty \leqslant A < x < B \leqslant \infty;$ $n = 1, 2, \ldots)$ *be a sequence of Gaussian processes with*

$$EG_n(x) = 0, \qquad R_n(u, v) = EG_n(u)\, G_n(v)$$

and

$$E \int_A^B G_n^2(x)\, dx = \int_A^B EG_n^2(x)\, dx = \int_A^B R_n(x, x)\, dx = m_n < +\infty.$$

Assume that $R_n(u, v)$ $((u, v) \in (A, B)^2)$ *is continuous at any point* (u, u) $(A < u < B)$, *square integrable,*

$$\Delta_n^2 = \mathrm{Var} \int_A^B G_n^2(x)\, dx = 2 \int_Y^B \int_A^B R_n^2(u, v)\, du dv \to \infty \qquad (n \to \infty),$$

and

$$\frac{\int_A^B \left(\int_A^B R_n(u, v)\, h(v)\, dv\right)^2 du}{\int_A^B \int_A^B R_n^2(u, v)\, du dv} \to 0 \qquad (n \to \infty) \tag{2.9}$$

for any $h \in L^2(A, B)$. *Then*

$$\Delta_n^{-1}\Big(\int_A^B G_n^2(x)\,dx - m_n\Big) \xrightarrow{\mathscr{D}} N(0, 1).$$

LEMMA 5. *Let the assumptions of Theorem* 1 *be satisfied and put*

$$\tilde{H}_n(Y, \beta^0) = n \int_{-a_n}^{a_n} [\hat{f}_n(y, Y^{(1)}, \beta^0) - \hat{f}_n(y, Y^{(2)}, \beta^0)]^2\,dy. \tag{2.10}$$

Then the asymptotic distribution of

$$\tilde{\mathscr{H}}_n(Y, \beta^0) = \tilde{\Delta}_n^{-1}\{\tilde{H}_n(Y, \beta^0) - \tilde{m}_n\}$$

is $N(0, 1)$, *where* $\tilde{m}_n = 2c_n^{-1}\int_{-\infty}^{\infty} w^2(t)\,dt$ *and*

$$\tilde{\Delta}_n^2 = 8c_n^{-1}\int_{-\infty}^{\infty} f^2(y)\,dy \int_{-\infty}^{\infty}\Big\{\int_{-\infty}^{\infty} w(z)\,w(t+z)\,dz\Big\}^2 dt.$$

Proof. First of all, let us recall that for the sequence of independent and identically distributed random variables $\{e_i\}_{i=1}^{\infty}$ the empirical distribution function is determined as

$$F_n(x, \omega) = \frac{1}{n}\sum_{i=1}^{n} I_{\{e_i(\omega) \leqslant x\}}.$$

We shall write briefly $F_n(x)$. Denote by $F_n^{(j)}(x)$ the corresponding empirical distribution function of $Y_i^{(j)} - X_i^T\beta^0$ $(i = 1, 2, \ldots, n$ for $j = 1$; respectively, $i = n+1, n+2, \ldots, 2n$ for $j = 2)$. Put (for $t \in [0, 1]$)

$$B_n^{(j)}(t) = \sqrt{n}\,[F_n^{(j)}(\operatorname{inv} F(t)) - t]$$

and define $\varphi_n(y, z) = c_n^{-1} w\big(c_n^{-1}(y-z)\big)$. Then we have

$$\begin{aligned}\sqrt{n}\{\hat{f}_n(y, Y^{(1)}, \beta^0) - \hat{f}_n(y, Y^{(2)}, \beta^0)\} &= \sqrt{n}\{E_{F_n^{(1)}}\varphi_n(y, x) - E_{F_n^{(2)}}\varphi_n(y, x)\}\\ &\overset{\text{a.s.}}{=} \int_0^1 B_n^{(1)}(t)\,d_t\varphi_n\big(y, \operatorname{inv} F(t)\big) - \int_0^1 B_n^{(2)}(t)\,d_t\varphi_n\big(y, \operatorname{inv} F(t)\big)\end{aligned}$$

(see also Csörgő and Révész [5], p. 223). Let us denote by $G_n(y)$ the process $\sqrt{n}\{\hat{f}_n(y, Y^{(1)}, \beta^0) - \hat{f}_n(y, Y^{(2)}, \beta^0)\}$, $-\infty < y < \infty$. Then we have $E_F G_n(y) \equiv 0$ and

$$\begin{aligned}&R_n(u, v) = EG_n(u)\,G_n(v)\\ &= E\Big\{\int_0^1 [B_n^{(1)}(t) - B_n^{(2)}(t)]\,d_t\varphi_n\big(u, \operatorname{inv} F(t)\big)\int_0^1 [B_n^{(1)}(z) - B_n^{(2)}(z)]\,d_z\varphi_n\big(v, \operatorname{inv} F(z)\big)\Big\}\end{aligned}$$

$$= \int_0^1\int_0^1 E\{[B_n^{(1)}(t)-B_n^{(2)}(t)][B_n^{(1)}(z)-B_n^{(2)}(z)]\}\, d_t\varphi_n(u, \operatorname{inv} F(t))\, d_z\varphi_n(v, \operatorname{inv} F(z))$$

$$= \int_0^1\int_0^1 \{E[B_n^{(1)}(t)B_n^{(1)}(z)] - E[B_n^{(1)}(t)B_n^{(2)}(z)] - E[B_n^{(2)}(t)B_n^{(1)}(z)]$$

$$+ E[B_n^{(2)}(t)B_n^{(2)}(z)]\}\, d_t\varphi_n(u, \operatorname{inv} F(t))\, d_z\varphi_n(v, \operatorname{inv} F(z))$$

$$= \int_0^1\int_0^1 \{EB_n^{(1)}(t)B_n^{(1)}(z) + EB_n^{(2)}(t)B_n^{(2)}(z)\}\, d_t\varphi_n(u, \operatorname{inv} F(t))\, d_z\varphi_n(v, \operatorname{inv} F(z))$$

$$= 2\int_{-\infty}^{\infty} \varphi_n(u, x)\varphi_n(v, x) f(x)\, dx$$

(see again Csörgő and Révész [5], p. 227). Now let us recall that $\tilde{m}_n$ and $\tilde{\Delta}_n^2$ are given by

$$\tilde{m}_n = \int_{-\infty}^{\infty} R_n(t, t)\, dx \qquad \text{and} \qquad \tilde{\Delta}_n^2 = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R^2(t, s)\, dt ds$$

(see the previous assertion). We obtain

$$\tilde{m}_n = 2c_n^{-2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w^2(c_n^{-1}(t-x)) f(x)\, dx dt$$

$$= 2c_n^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w^2(z) f(t-c_n z)\, dt dz = 2c_n^{-1}\int_{-\infty}^{\infty} w^2(z)\, dz.$$

Similarly,

$$(2.11)\quad \tilde{\Delta}_n^2 = 8\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{c_n^{-1}\int_{-\infty}^{\infty} w(c_n^{-1}(t-x))\, w(c_n^{-1}(s-x)) f(x)\, dx\}^2\, dt ds$$

$$= 8\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{c_n^{-1}\int_{-\infty}^{\infty} w(z)\, w(c_n^{-1}(s-t)+z) f(t-c_n z)\, dz\}^2\, dt ds$$

$$= 8\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{c_n^{-1}\int_{-\infty}^{\infty} w(z)\, w(v+z) f(t-c_n z)\, dz\}^2\, dt\cdot c_n dv$$

$$= 8c_n^{-1}\int_{-\infty}^{\infty} f^2(r)\, dr \int_{-\infty}^{\infty}\{\int_{-\infty}^{\infty} w(z)\, w(v+z)\, dz\}^2\, dv.$$

To be able to use Assertion 2 we need to verify that $R_n(u, v)$ is continuous at every point (t, t), $t \in R$ (which is however evident due to continuity of kernel w), and also that the condition (2.9) holds. We have for any $h(t) \in L_2$

$$\int_{-\infty}^{\infty}\Big[\int_{-\infty}^{\infty}\{c_n^{-2}\int_{-\infty}^{\infty} w(c_n^{-1}(t-x))\, w(c_n^{-1}(s-x)) f(x)\, dx\}\, h(t)\, dt\Big]^2 ds$$

$$= \int_{-\infty}^{\infty} \Big[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(z)\, w(v+z)\, f(s-c_n v-c_n z)\, dz\, h(s-c_n v)\, dv \Big]^2 ds$$

$$\leqslant V \int_{-\infty}^{\infty} h^2(r)\, dr \Big[\int_{-\infty}^{\infty} w(z)\, w(v+z)\, dz dv \Big]^2 = O(1).$$

This represents the numerator of the ratio in (2.9). The denominator is equal to $\frac{1}{2}\tilde{\Delta}_n^2$ and (evidently, see (2.11)) converges to infinity. So the whole ratio converges to zero as $n \to \infty$, which means that the condition (2.9) is fulfilled. Hence using the Assertion 2 we infer that the statistic

$$\tilde{\Delta}_n^{-1} \{ n \int_{-\infty}^{\infty} [\hat{f}_n(y, Y^{(1)}, \beta^0) - \hat{f}_n(y, Y^{(2)}, \beta^0)]^2 dy - \tilde{m}_n \}$$

converges in distribution to $N(0, 1)$. To complete the proof it remains to show that

$$\tilde{\Delta}_n^{-1} n \{ \int_{|y| > a_n} [\hat{f}_n(y, Y^{(1)}, \beta^0) - \hat{f}_n(y, Y^{(2)}, \beta^0)]^2 dy \} = o_p(1).$$

In order to do this, let us write the left-hand side as

$$\tilde{\Delta}_n^{-1} n \{ \int_{|y| > a_n} [[\hat{f}_n(y, Y^{(1)}, \beta^0) - E\hat{f}_n(y, Y^{(1)}, \beta^0)]$$
$$- [\hat{f}_n(y, Y^{(2)}, \beta^0) - E\hat{f}_n(y, Y^{(2)}, \beta^0)]] \}^2 dy$$

$$\leqslant 2\tilde{\Delta}_n^{-1} n \int_{|y| > a_n} \{ [\hat{f}_n(y, Y^{(1)}, \beta^0) - E\hat{f}_n(y, Y^{(1)}, \beta^0)]^2$$
$$+ [\hat{f}_n(y, Y^{(2)}, \beta^0) - E\hat{f}_n(y, Y^{(2)}, \beta^0)]^2 \} dy.$$

Now for any $\delta > 0$

$$P \{ \tilde{\Delta}_n^{-1} n \int_{a_n}^{\infty} [\hat{f}_n(y, Y^{(1)}, \beta^0) - E\hat{f}_n(y, Y^{(1)}, \beta^0)]^2 dy > \delta \}$$

$$\leqslant \mathscr{K} \delta^{-1} c_n^{-3/2} \int_{a_n}^{\infty} \int_{-\infty}^{\infty} [w(c_n^{-1}(y-z)) - Ew(c_n^{-1}(y-e_1))]^2 f(z)\, dz dy$$

$$\leqslant \mathscr{K} \delta^{-1} c_n^{-3/2} \int_{a_n}^{\infty} \int_{-\infty}^{\infty} w^2(c_n^{-1}(y-z)) f(z)\, dz dy,$$

where $\mathscr{K}^{-2} = 8 \int_{-\infty}^{\infty} f(y)\, dy \int_{-\infty}^{\infty} \{ \int_{-\infty}^{\infty} w(t)\, w(v+t)\, dt \}\, dv$. Consequently, the upper bound for the studied probability may be written in the form

$$\mathscr{K} \delta^{-1} c_n^{-1/2} \int_{a_n}^{\infty} \{ \int_{-\infty}^{\infty} w^2(t) f(y - tc_n)\, dt \}\, dy$$

$$= \mathscr{K} \delta^{-1} c_n^{-1/2} \int_{a_n}^{\infty} \{ \int_{-L_n}^{L_n} w^2(t) f(y - tc_n)\, dt + \int_{|t| > L_n} w^2(t) f(y - tc_n)\, dt \}\, dy.$$

The first term may be (starting with some $n_0 \in N$) bounded by

$$\mathscr{K}\delta^{-1} c_n^{-1/2} \int_{-L_n}^{L_n} w^2(t)\,dt \int_{(1-\tau)a_n}^{\infty} f(y)\,dy = o(1) \quad \text{as } n \to \infty,$$

and the second one by

$$\mathscr{K}\delta^{-1} c_n^{-1/2} \int_{|t|>L_n} \{w^2(t) \int_{-\infty}^{\infty} f(y-tc_n)\,dy\}\,dt$$

$$= \mathscr{K}\delta^{-1} c_n^{-1/2} \int_{|t|>L_n} w^2(t)\,dt = o(1) \quad \text{as } n \to \infty.$$

That completes the proof. ∎

Proof of Theorem 1. Due to Lemma 1 we have

$$H_n(Y, \beta^0) = n \int_{-a_n}^{a_n} [\hat f_n^{1/2}(y, Y^{(1)}, \beta^0) - \hat f_n^{1/2}(y, Y^{(2)}, \beta^0)]^2 E\hat f_{2n}(y, Y, \beta^0)\,dy$$
$$+ \mathcal{O}_p(n^{-1/2} c_n^{-2} a_n).$$

Using the equality

$$\hat f_n^{1/2}(y, Y^{(i)}, \beta) - E^{1/2}\hat f_n(y, Y^{(i)}, \beta)$$
$$= \frac{1}{2}\{\hat f_n(y, Y^{(i)}, \beta) - E\hat f_n(y, Y^{(i)}, \beta) - [\hat f_n^{1/2}(y, Y^{(i)}, \beta) - E^{1/2}\hat f_n(y, Y^{(i)}, \beta)]^2\}$$
$$\times E^{-1/2}\hat f_n(y, Y^{(i)}, \beta)$$

and taking into account (2.2) we obtain

$$H_n(Y, \beta^0) = n \int_{-a_n}^{a_n} \Big\{\frac{1}{2}[\hat f_n(y, Y^{(1)}, \beta^0) - E\hat f_n(y, Y^{(1)}, \beta^0)$$
$$-[\hat f_n^{1/2}(y, Y^{(1)}, \beta^0) - E^{1/2}\hat f_n(y, Y^{(1)}, \beta^0)]^2] E^{-1/2}\hat f_n(y, Y^{(1)}, \beta^0)$$
$$-\frac{1}{2}[\hat f_n(y, Y^{(2)}, \beta^0) - E\hat f_n(y, Y^{(2)}, \beta^0) - [\hat f_n^{1/2}(y, Y^{(2)}, \beta^0) - E^{1/2}\hat f_n(y, Y^{(2)}, \beta^0)]^2]$$
$$\times E^{-1/2}\hat f_n(y, Y^{(2)}, \beta^0)\Big\}^2 E\hat f_{2n}(y, Y, \beta^0)\,dy + \mathcal{O}_p(n^{-1/2} c_n^{-2} a_n).$$

Now using Lemmas 2–4 we obtain

$$\text{(2.12)} \qquad H_n(Y, \beta^0) = \frac{n}{4} \int_{-a_n}^{a_n} [\hat f_n(y, Y^{(1)}, \beta^0) - \hat f_n(y, Y^{(2)}, \beta^0)]^2\,dy$$
$$+ \mathcal{O}_p(n^{-1/2} c_n^{-2} a_n d_n^{-1})$$

and the theorem follows from Lemma 5. ∎

Remark 2. It follows from (2.10) and (2.12) that the statistics $\tilde H_n(Y, \beta^0)$ and $H_n(Y, \beta^0)$ are asymptotically equivalent, and hence we could use for the purpose which was discussed in the Introduction the statistic $\tilde{\mathscr{H}}_n(Y, \beta^0)$ instead of $\mathscr{H}_n(Y, \beta^0)$. The reason why we have considered at first $\mathscr{H}_n(Y, \beta^0)$ was the fact that the Hellinger distance is a well-known statistic which was proved to be useful and applicable, and hence many statisticians became familiar with it. On the other hand, it is evident that it is much easier to treat $\tilde{\mathscr{H}}_n(Y, \beta^0)$ than $\mathscr{H}_n(Y, \beta^0)$.

To be able to construct a critical region for a test based on $\tilde{\mathscr{H}}_n(Y, \beta^0)$ one needs to estimate $\int_{-\infty}^{\infty} f^2(r)\,dr$.

To do this we shall use the following assertion. Since the assumptions are somewhat different from the assumptions of Theorem 1, we will give them precisely following Csörgő and Révész [5]. For the proof see Theorem 6.1.5 in [5].

ASSERTION 3. *Let f vanish outside a finite interval (C, D) and assume that it has a bounded second derivative on this interval. Moreover, let the kernel w vanish also out of a finite interval (A, B) with $A \leqslant C < D \leqslant B$ and $\operatorname{var}_{t\in(A,B)} w(t) \leqslant M$ together with $\int_{-\infty}^{\infty} tw(t)\,dt = 0$. Let us put*

$$\sigma^2 = \int_C^D f^2(u)\,du \int_A^B \Big\{\int_A^B w(z+v)\,w(z)\,dz\Big\}^2 dv .$$

Then

$$c_n^{-1/2}\,\sigma^{-1}\Big\{nc_n \int_A^B [\hat{f}_n(y, Z, \beta^0) - f(y)]^2\,dy - \int_A^B w^2(t)\,dt\Big\} \to N(0, 1)$$

in distribution, provided that $n^{-1} c_n^{-3/2} \log^2 n = o(1)$ and $nc_n^{9/2} = o(1)$.

Remark 3. We formulated Assertion 3 in the form as given in Csörgő and Révész [5]. As demonstrated in Theorem 1 we may avoid the assumption of the bounded support of the density f by a more complicated setup with increasing intervals $(-a_n, a_n)$. In both approaches, i.e. in the approach assuming the compact support of density as well as in the approach with a sequence of intervals (the length of which increases to infinity), we have to cope with the behaviour of the density and its estimator in the tail areas. The setup given in Theorem 1 is (a little) more general, however we have to pay for it by the smaller transparency of proofs and by a lot of technicalities in them. In spite of this disadvantage we have used it, just to show that the bounded support of density is not an inevitable assumption, and so that in Csörgő and Révész [5] the approach with compact support was used only to simplify the text. Of course, it also took into account the fact that when we apply such theoretical results for finite-size samples (especially for modest or small sizes) it is questionable to insist that data were "generated" by a distribution having the density with the unbounded support or not. That is why we will further use the setup which was preferred by Csörgő and Révész. ∎

THEOREM 2. *Let us assume that f has a bounded second derivative and vanishes outside a finite interval (C, D). Let also the kernel w vanish outside a finite interval (A, B), $A \leqslant C \leqslant 0 \leqslant D \leqslant B$, and $\mathrm{var}_{t\in(A,B)} w(t) \leqslant M$ together with $\int_{-\infty}^{\infty} tw(t)\,dt = 0$. Let us put*

$$\tilde{H}_n(Y, \beta^0) = n \int_A^B [\hat{f}_n(y, Y^{(1)}, \beta^0) - \hat{f}_n(y, Y^{(2)}, \beta^0)]^2 \hat{f}_{2n}(y, Y, \beta^0)\,dy.$$

Finally, let $n^{-1} c_n^{-3/2} \log^2 n = o(1)$ and $nc_n^{9/2} = o(1)$. Then

$$[\Delta_n^*]^{-1} \{\tilde{H}_n(Y, \beta^0) - m_n\}$$

has the asymptotic distribution $N(0, 1)$, where

$$[\Delta_n^*]^2 = \frac{1}{2} c_n^{-1} \int_A^B \hat{f}_n^2(r, Y, \beta^0)\,dr \int_A^B \Big\{\int_A^B w(z)\,w(v+z)\,dz\Big\}^2 dv$$

and where m_n is the same as in Theorem 1, i.e.

$$m_n = \frac{1}{2} c_n^{-1} \int_A^B w^2(z)\,dz.$$

Proof. The theorem follows from Lemma 5 and Assertion 3 (do not be confused that integration is taken only over the interval (A, B) although $\hat{f}_n(r, Y, \beta^0)$ does not necessarily vanish outside this interval; however, the integral over the complement to (A, B) is not asymptotically significant, see Assertion 3).

To bring the considerations which started in the Introduction to the end which would be applicable, we need to find a statistic which would depend only on the data and not on the unknown value β^0. This will be done in the next theorem.

THEOREM 3. *Let the assumptions of Theorem 2 be fulfilled. Moreover, let $\hat{\beta}$ be an $\sqrt{n}$-consistent estimator of β^0 and let $w'(z)$, $w''(z)$ and $f'(y)$ exist everywhere. Further, let*

$$(2.13) \qquad \int_{-\infty}^{\infty} |tw'(t)|\,dt < \infty, \qquad \sup_{t\in R} |w'(t)| < \infty, \qquad \sup_{t\in R} |w''(t)| < \infty \quad \text{and} \quad \sup_{y\in R} |f'(y)| < \infty.$$

Finally, let

$$(2.14) \qquad \frac{1}{n} \sum_{i=1}^{n} \|X_i\|^2 = \mathcal{O}(1)$$

and

$$(2.15) \qquad \int_C^D f^3(y)\,dy < \infty.$$

Then

$$\tilde{H}_n(Y, \hat{\beta}) = \tilde{H}_n(Y, \beta^0) + o_p(c_n).$$

Proof. Not to burden the paper (which is already full of technicalities) by a lot of steps which are similar to those from the previous text, we shall only indicate the idea of the proof. Let us write

$$\tilde{H}_n(Y, \hat{\beta}) = n\int_A^B [\hat{f}_n(y, Y^{(1)}, \hat{\beta}) - \hat{f}_n(y, Y^{(2)}, \hat{\beta})]^2 \hat{f}_{2n}(y, Y, \hat{\beta})\, dy$$

$$= n\int_A^B \{[\hat{f}_n(y, Y^{(1)}, \hat{\beta}) - \hat{f}_n(y, Y^{(1)}, \beta^0)] + [\hat{f}_n(y, Y^{(1)}, \beta^0) - \hat{f}_n(y, Y^{(2)}, \beta^0)]$$

$$+ [\hat{f}_n(y, Y^{(1)}, \hat{\beta}) - \hat{f}_n(y, Y^{(1)}, \beta^0)]\}^2 \hat{f}_{2n}(y, Y, \hat{\beta})\, dy$$

and consider at first

$$(2.16)\qquad n\int_A^B [\hat{f}_n(y, Y^{(1)}, \hat{\beta}) - \hat{f}_n(y, Y^{(1)}, \beta^0)]^2 \hat{f}_{2n}(y, Y, \hat{\beta})\, dy$$

$$= \frac{1}{2n^3 c_n^5}\int_A^B \Big[\sum_{i=1}^{n} \sqrt{n} X_i^T(\hat{\beta} - \beta^0) w'(c_n^{-1}\eta_i)\Big]^2 \sum_{j=1}^{2n} w\big(c_n^{-1}(y - Y_j + X_j^T\hat{\beta})\big)\, dy,$$

where

$$\eta_i \in (\min\{y - Y_i + X_i^T\hat{\beta},\, y - Y_i + X_i^T\beta^0\},\, \max\{y - Y_i + X_i^T\hat{\beta},\, y - Y_i + X_i^T\beta^0\}),$$

i.e.

$$\eta_i \in (\min\{y - e_i + X_i^T(\hat{\beta} - \beta^0),\, y - e_i\},\, \max\{y - e_i + X_i^T(\hat{\beta} - \beta^0),\, y - e_i\}).$$

To make an idea about the expression in (2.16), let us study

$$(2.17)\qquad \frac{1}{2n^3 c_n^5}\int_A^B \Big[\sum_{i=1}^{n} \sqrt{n} X_i^T(\hat{\beta} - \beta^0) w'(c_n^{-1}\eta_i)\Big]^2 \sum_{j=1}^{2n} w\big(c_n^{-1}(y - Y_j + X_j^T\beta^0)\big)\, dy$$

$$= \frac{1}{2n^3 c_n^5}\int_A^B \Big[\sum_{i=1}^{n} \sqrt{n} X_i^T(\hat{\beta} - \beta^0) w'(c_n^{-1}\eta_i)\Big]^2 \sum_{j=1}^{2n} w\big(c_n^{-1}(y - e_j)\big)\, dy.$$

Fix an $\varepsilon > 0$. Since $\hat{\beta}$ is $\sqrt{n}$-consistent (i.e. $\sqrt{n}(\hat{\beta} - \beta^0) = \mathcal{O}_p(1)$), there is a finite constant K_1 such that with probability at least $1 - \varepsilon$ we have $y - e_i - n^{-1/2}K_1 \leqslant \eta_i \leqslant y - e_i + n^{-1/2}K_1$. Now the expression in (2.17) can be rewritten as

$$(2.18)\qquad \frac{1}{2n^3 c_n^5}\int_A^B \Big[\sum_{i=1}^{n} \sqrt{n} X_i^T(\hat{\beta} - \beta^0)\big(w'(c_n^{-1}(y - e_i)) + w''(\xi_i)\tau_i\big)\Big]^2$$

$$\times \sum_{j=1}^{2n} w\big(c_n^{-1}(y - e_j)\big)\, dy,$$

where

$$\xi_i \in (\min\{y-e_i, \eta_i\}, \max\{y-e_i, \eta_i\})$$

and

$$|\tau_i| \leqslant n^{-1/2} K_1. \tag{2.19}$$

Since $w(z) \geqslant 0$, the expression in (2.18) is bounded by

$$\frac{1}{2n^3 c_n^5} \int_A^B 2 \{[\sqrt{n} \sum_{i=1}^n X_i^T (\hat\beta - \beta^0) w'(c_n^{-1}(y-e_i))]^2$$

$$+ [\sqrt{n} \sum_{i=1}^n X_i^T (\hat\beta - \beta^0) w''(\xi_i)\tau_i]^2\} \sum_{j=1}^{2n} w(c_n^{-1}(y-e_j))\, dy.$$

The assumption that $\sup_{z\in R} |w''(z)| < \infty$ together with (2.19) yield

$$\frac{1}{n^3 c_n^5} \int_A^B [\sqrt{n} \sum_{i=1}^n X_i^T (\hat\beta - \beta^0) w''(\xi_i)\tau_i]^2 \sum_{j=1}^{2n} w(c_n^{-1}(y-e_j))\, dy = \mathcal{O}(n^{-1} c_n^{-5}).$$

Therefore we have to cope with the expression

$$\frac{1}{n^3 c_n^5} \int_A^B \sum_{i=1}^n \|X_i\| w'(c_n^{-1}(y-e_i)) \sum_{k=1}^n \|X_k\| w'(c_n^{-1}(y-e_k)) \times \sum_{j=1}^{2n} w(c_n^{-1}(y-e_j))\, dy. \tag{2.20}$$

Now we shall use the fact that the expression

$$\frac{1}{n} \sum_{i=1}^n \|X_i\| w'(c_n^{-1}(y-e_i))$$

is approximately equal to

$$E w'(c_n^{-1}(y-e_1)) \frac{1}{n} \sum_{i=1}^n \|X_i\|.$$

So we need to study

$$\int_{-\infty}^{\infty} w'(c_n^{-1}(y-z)) f(z)\, dz \quad \text{and} \quad \int_{-\infty}^{\infty} [w'(c_n^{-1}(y-z))]^2 f(z)\, dz.$$

Using the transformation $c_n^{-1}(y-z) = u$ we arrive at

$$\int_{-\infty}^{\infty} w'(c_n^{-1}(y-z)) f(z)\, dz = c_n \int_{-\infty}^{\infty} w'(u) f(y - c_n u)\, du$$

$$= c_n \int_{-\infty}^{\infty} [w'(u) f(y) + c_n w'(u) f'(\kappa_i) u]\, du$$

for an appropriate $|\kappa_i| < c_n|u|$. But then by (2.13) we obtain

$$\int_{-\infty}^{\infty} w'(u) f(y)\,du = 0 \quad \text{and} \quad c_n^2 \int_{-\infty}^{\infty} w'(u) f'(\kappa_i) u\,du = \mathcal{O}(c_n^2). \tag{2.21}$$

Along similar lines we find that

$$\int_{-\infty}^{\infty} \left[w'\left(c_n^{-1}(y-z)\right)\right]^2 f(z)\,dz = \mathcal{O}(c_n).$$

Therefore, putting $\xi_i = c_n^{-1}\|X_i\|\left(w'\left(c_n^{-1}(y-e_i)\right) - Ew'\left(c_n^{-1}(y-e_i)\right)\right)$ and applying the strong law of large numbers (see e.g. Breiman [4], Theorem 3.27), we obtain

$$\frac{1}{nc_n}\sum_{i=1}^{n} \|X_i\|\left(w'\left(c_n^{-1}(y-e_i)\right) - Ew'\left(c_n^{-1}(y-e_i)\right)\right) = o_p(1).$$

Approximating the expressions

$$\frac{1}{nc_n^2}\sum_{i=1}^{n} \|X_i\|\, w'\left(c_n^{-1}(y-e_i)\right),$$

$$\frac{1}{nc_n^2}\sum_{k=1}^{n} \|X_k\|\, w'\left(c_n^{-1}(y-e_k)\right) \quad \text{and} \quad \frac{1}{nc_n}\sum_{j=1}^{2n} w\left(c_n^{-1}(y-e_j)\right)$$

in (2.20) successively by

$$Ew'\left(c_n^{-1}(y-e_1)\right)\frac{1}{n}\sum_{i=1}^{n} \|X_i\|,$$

$$Ew'\left(c_n^{-1}(y-e_1)\right)\frac{1}{n}\sum_{k=1}^{n} \|X_k\| \quad \text{and} \quad c_n Ew'\left(c_n^{-1}(y-e_1)\right),$$

respectively, and taking into account (2.14), (2.15) and (2.21), we find that (2.17) is $o_p(c_n)$. By similar considerations we obtain the same result for (2.16). A chain of analogous steps then completes the proof of the theorem.

Remark 4. The assumption that the support of F is finite represents naturally some (theoretical) restriction of generality. As mentioned in Remark 2, due to the fact that in applications we always have a finite number of observations, it may occur that it does not restrict applicability of the result. However, much more important for the practical purposes is the problem that is common to all the asymptotic results: "How will the asymptotics really work?" ∎

The answer to this question needs a reasonably large study, preferably on real data sets, taking into account some sufficiently rich family of methods for estimating the regression model. Therefore such a study has to be postponed to a special paper.

Nevertheless, to create at least a very first idea about a "sensitivity" of $\mathscr{H}_n(Y, \beta)$ with respect to changes of β (or, more precisely, about an "ability" to distinguish different β's) we give the following artificial Example 1. (Next Example 2 offers then a very first insight on the behaviour of statistic $\mathscr{H}_n(Y, \beta)$ on the real data.)

3. NUMERICAL EXAMPLES

Example 1. We have considered the linear model

$$Y = \beta_1 x_1 + \beta_2 x_2 + e \tag{3.1}$$

and we have generated data which are given in Table 1. The values of regressors (x_{i1}, x_{i2}) for the first half of data (i.e. for $i = 1, 2, \ldots, 40$) were generated so to be uniformly distributed in the triangle $[-1, -1]$, $[-1, 1]$, $[1, 1]$, while for the second half (i.e. for $i = 41, 42, \ldots, 80$) in the triangle $[-1, -1]$, $[1, -1]$, $[1, 1]$. Such a division of data seems to be quite realistic, since in real applications (see Example 2) we will divide data mostly according to the position in the factor space.

Values of $e_1, e_2, \ldots, e_{80}$ were generated by applying a polynomial approximation to the normal quantile function (see Abramowitz and Stegun [1]) on the uniformly distributed numbers and the resulting numbers were tested for normality (see Shapiro and Wilk [16]). Then we have assumed the true model (3.1) with $\beta_1 = 0$ and $\beta_2 = 0$. This means that $Y_i = e_i$, i.e. $e_1, e_2, \ldots, e_{80}$ coincide with values of response variables.

Table 1. Simulated data

The first half of data

Case	Y	X_1	X_2
1	−0.0278	0.7537	0.8400
2	0.9238	0.2399	0.2619
3	−0.0834	0.8962	0.9584
4	1.2208	−0.3673	0.6822
5	−0.1394	−0.8629	0.6864
6	−0.1958	−0.5276	0.3366
7	−0.2529	0.6635	0.8421
8	−0.3109	0.3726	0.8106
9	0.0278	−0.4063	0.9508
10	−0.3699	−0.2691	0.2209
11	0.7645	−0.9747	0.9088
12	0.3699	0.1885	0.6780
13	1.0129	0.8870	0.9873
14	0.1394	0.5002	0.9386
15	−0.4303	−0.3184	0.7459

The second half of data

Case	Y	X_1	X_2
41	−0.0556	0.2351	−0.6782
42	0.8819	−0.5857	−0.7695
43	0.2818	0.4168	−0.3343
44	0.6568	−0.4799	−0.6689
45	−0.1114	−0.6188	−0.8509
46	−0.1676	−0.3346	−0.5493
47	1.4204	0.6719	−0.7896
48	1.1640	−0.2910	−0.6352
49	−0.2243	−0.9843	−0.9899
50	−0.2818	0.5004	−0.3201
51	0.5891	0.8870	0.7417
52	0.1676	0.9318	0.6582
53	0.9674	−0.3656	−0.6346
54	−0.3403	−0.5189	−0.6075
55	0.0556	0.5101	−0.6666

Table 1. Simulated data (continued)

Case	Y	X_1	X_2	Case	Y	X_1	X_2
16	−0.4923	0.9421	0.9803	56	−0.3999	−0.4439	−0.4639
17	−0.5563	−0.0709	0.6035	57	0.2243	−0.4070	−0.4670
18	0.4923	0.2551	0.5740	58	−0.4611	0.8893	0.1933
19	−0.6226	−0.5090	0.8582	59	−0.5240	−0.9833	−0.9858
20	−0.6918	−0.2109	−0.1201	60	−0.5891	0.7260	−0.8686
21	−0.7645	−0.8251	0.9968	61	0.7277	−0.6695	−0.7386
22	−0.8415	−0.5962	0.2590	62	0.5240	−0.9731	−0.9980
23	−0.9238	0.0310	0.2470	63	−0.6568	0.1018	−0.0502
24	0.2529	0.6227	0.9019	64	0.3403	0.1052	−0.6904
25	0.5563	−0.4704	0.5341	65	0.4611	0.0643	−0.5707
26	1.3478	−0.5214	0.8752	66	−0.7277	0.3307	−0.7538
27	−1.0129	−0.6669	−0.2313	67	−0.8024	−0.6400	−0.9646
28	−1.1108	−0.4020	0.2044	68	−0.8819	0.0029	−0.7170
29	1.1108	−0.8547	0.3699	69	−0.9674	−0.4136	−0.5663
30	0.8415	0.5555	0.9758	70	0.8024	0.8188	−0.7828
31	−1.2208	−0.8713	0.6129	71	−1.0606	0.7319	0.1330
32	0.6918	−0.4807	−0.1217	72	−1.1640	−0.6340	−0.9350
33	0.3109	−0.9096	0.9114	73	0.1114	0.3593	−0.8698
34	−1.3478	−0.1972	0.1177	74	0.3999	0.3191	−0.2061
35	0.4303	−0.0936	−0.0247	75	−1.2817	0.5526	0.0260
36	0.0834	0.1495	0.4817	76	−1.4204	−0.7587	−0.8561
37	0.6226	0.5069	0.9061	77	0.0000	−0.0160	−0.7893
38	0.1958	−0.9474	0.7699	78	1.2817	−0.3150	−0.7289
39	1.5014	−0.2868	0.8067	79	1.0606	0.9344	−0.2196
40	−1.5014	−0.4667	0.2120	80	−1.5936	−0.4742	−0.8502

Now, let us explain Table 2. Take into account that for our case, i.e. when $\beta_1 = 0$ and $\beta_2 = 0$, the value $H_n(Y, \beta)$ in the first row represents the value of a weighted Hellinger distance of residuals in the "true" model and the values of statistics $H_{80}(H, \beta)$ and $\mathscr{H}_{80}(Y, \beta)$ hint that the both halves of data are really very similar. In the other rows of Table 2 the values of $H_n(Y, \beta)$ and $\mathscr{H}_n(Y, \beta)$ are given for 41 values of β equidistantly spread (as the vector) over one half of the unit circle (i.e. we may interpret the values e.g. in the second row of the left-hand half of Table 2 as follows: An estimator has given the estimate $\hat{\beta}_1 = 0$, $\hat{\beta}_2 = 1$, and using corresponding residuals

$$r_i = Y_i - (X_{i1} \cdot 0 + X_{i2} \cdot 1)$$

we have evaluated $H_{80}(Y, \hat{\beta})$. Since 5 percent quantile of the standard normal distribution is 1.645, our test rejects the hypothesis that $\hat{\beta}_1 = 0$, $\hat{\beta}_2 = 1$ may be a true model.

Table 2. Sensitivity of statistics

β_1	β_2	$H_{80}(Y, \beta)$	$\mathscr{H}_{80}(Y, \beta)$
0.0000	0.0000	0.0020	−0.3595
0.0000	1.0000	2.1870	2.6731
0.0785	0.9969	2.1490	2.6221
0.1564	0.9877	2.0548	2.4922
0.2334	0.9724	1.9023	2.2835
0.3090	0.9511	1.7320	2.0489
0.3827	0.9239	1.5763	1.8311
0.4540	0.8910	1.4052	1.5916
0.5225	0.8526	1.2581	1.3825
0.5878	0.8090	1.1397	1.2107
0.6494	0.7604	0.9888	0.9955
0.7071	0.7071	0.8106	0.7449
0.7604	0.6494	0.7049	0.5921
0.8090	0.5878	0.5979	0.4393
0.8526	0.5225	0.4818	0.2764
0.8910	0.4540	0.3496	0.0933
0.9239	0.3827	0.2386	−0.0599
0.9511	0.3090	0.1478	−0.1843
0.9724	0.2334	0.0811	−0.2750
0.9877	0.1564	0.0435	−0.3255
0.9969	0.0785	0.0278	−0.3461
1.0000	0.0000	0.0371	−0.3327

β_1	β_2	$H_{80}(Y, \beta)$	$\mathscr{H}_{80}(Y, \beta)$
0.0000	0.0000	0.0020	−0.3595
0.0000	1.0000	2.1870	2.6731
−0.0785	0.9969	2.2120	2.7017
−0.1564	0.9877	2.2552	2.7454
−0.2334	0.9724	2.0855	2.5147
−0.3090	0.9511	2.0170	2.4102
−0.3827	0.9239	1.8909	2.2289
−0.4540	0.8910	1.6813	1.9397
−0.5225	0.8526	1.5268	1.7187
−0.5878	0.8090	1.3007	1.4078
−0.6494	0.7604	1.0938	1.1238
−0.7071	0.7071	0.9078	0.8684
−0.7604	0.6494	0.7692	0.6755
−0.8090	0.5878	0.6202	0.4696
−0.8526	0.5225	0.4562	0.2435
−0.8910	0.4540	0.3334	0.0721
−0.9239	0.3827	0.2347	−0.0669
−0.9511	0.3090	0.1520	−0.1855
−0.9724	0.2334	0.0991	−0.2619
−0.9877	0.1564	0.0597	−0.3215
−0.9969	0.0785	0.0428	−0.3477
−1.0000	0.0000	0.0420	−0.3503

Notice that nearby the most sensitive direction represented by $\beta = (0, 1)$ (and it would be similarly for $\beta = (0, -1)$) we have rather a wide range (up to $\beta = (0.38, 0.92)$ on one side and similarly up to $\beta = (-0.45, 0.89)$ on the opposite side) in which the test rejects (on the 5 percent level) the hypothesis that β is the true value (i.e. that such an estimate of the model is acceptable). A small asymmetry in behaviour of the statistics (in the left and the right half of Table 2) is caused by the fact that the randomly generated regressors for the first 40 observations have fallen accidently into the region given by points $[-1, 0]$, $[-1, 1]$, $[1, 1]$, $[0, 0]$ rather than into the triangle $[-1, -1]$, $[1, -1]$, $[1, 1]$. A similar situation occurred for the second half of observations.

By this example it is clear that the method is sensitive with respect to division of points in the sample (according to position in the factor space). Therefore, to be able to verify whether the obtained estimate of the model is such that it generates in two (reasonably selected) halves of the sample approximately the same estimate of density of residuals we may need (for the more-dimensional factor space) more (orthogonal) divisions. However, it was clear from the very beginning that the statistic $\mathscr{H}_n(Y, \beta)$ will depend on division of the sample. Nevertheless, obtained results give a hope that $\mathscr{H}_n(Y, \beta)$ may work quite well.

EXAMPLE 2. For the second example we have used well-known "Star data" and gave them here only in the form of a figure.

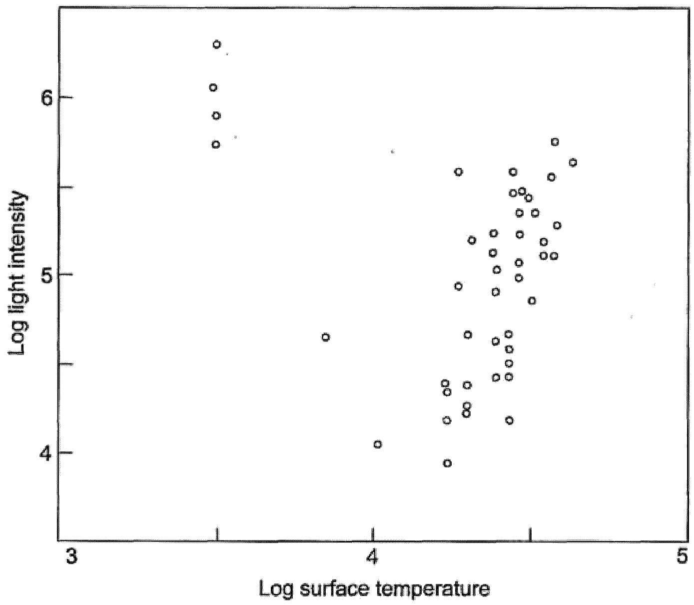

Fig. 1. Hertzprung–Russell diagram

The data describe the dependence of the light intensity of stars on their surface temperature, they may be found in Rousseeuw and Leroy [15] and were firstly published in Humphreys [8]. Since we shall construct the simple regression, we may easily verify whether a given estimate is reasonable.

We have applied on the data the Least Squares, then the Least Median of Squares (LMS) (see Rousseeuw and Leroy [15]), and finally twice again the Least Squares on restricted samples which were obtained deleting from the full data 4 (respectively, 6) points which had in the LMS estimate the largest residuals. (The LMS estimate has been evaluated by the software of Pavel Boček (see Boček and Lachout [2], see also discussion on the algorithms used for LMS in Víšek [22] or [25]); we are grateful for the possibility to use it.) The estimated models are briefly reported in Table 3, where $\hat{\beta}_1^0$ is an estimate for intercept, $\hat{\beta}_2^0$ is an estimate of the slope coefficient for the explaining variable "temperature", $\hat{\sigma}^2$ is an estimate of the scale, and finally R^2 stays for the coefficient of determination.

Then the data were divided into two halves (both containing the same number of observations) by a line which was orthogonal to the estimated regression line, and corresponding values of the statistics H_n and $\mathscr{H}_n$ for these models were gathered in Table 4.

Table 3. Results of regression analysis (LS — the Least Squares, LMS — the Least Median of Squares, LS-4 — the Least Squares with 4 points deleted, LS-6 — the Least Squares with 6 points deleted)

METHOD	LS	LMS	LS-4	LS-6
$\hat{\beta}_1^0$	6.79347	-12.74000	-4.05652	-8.50005
$\hat{\beta}_2^0$	-0.41330	4.00000	2.04666	3.04616
$\hat{\sigma}^2$	0.56463	0.19270	0.40580	0.34074
R^2	0.044273	0.87833	0.36656	0.55435

Table 4. Values of testing statistics (the abbreviations are the same as above)

METHOD	LS	LMS	LS-4	LS-6
$H_{47}(Y, \beta)$	1.5299	1.1084	1.2228	0.8262
$\mathscr{H}_{47}(Y, \beta)$	1.2932	0.8198	0.9254	0.5330

Although it is clear from Table 4 that the test will not reject the hypothesis that densities of residuals in the two subsets of data are the same for LS--estimate and LMS-estimate, the improvement of the model estimate when rejecting 6 points is evident. On the other hand, it is out of question that the test is (also) somewhat conservative one.

4. CONCLUSIONS

The previous two examples have demonstrated that the proposed statistic of the Hellinger type may help to select from all the evaluated estimates of the regression model such a model which fits best to the given data. The asymptotic normality of the test statistic allows then to construct also a critical region.

One may object that the critical region of the test is based on the asymptotic result which may (sometimes) start to work only for a large number of observations. Maybe that something like small-sample asymptotics for empirical processes could help, however an applicable theory is not still available.

As we have seen, the division of data into two (complementary) subsets is (theoretically) arbitrary. Somebody may consider it as a drawback, somebody as an advantage (to tailor the method to the topology of data). On the other hand, the division which was used in Example 2 seems to be quite natural. In the case of $p > 2$ we have to generalize the method of division as follows. At first, we project the data orthogonally into the estimated regression plane and determine the largest axis of the ellipsoid which is generated by the inverse matrix to the sample covariance matrix of these projected data. Secondly, we take from all the planes orthogonal to this axis such one which divides the data into two halves with the same numbers of points. In the case where some points are in the dividing plane we should add them randomly to one of the "halves".

The requirement that the (estimates of) density of residuals should be (very) similar in (reasonably selected) subsets of residuals is probably quite natural and transparent. The idea resembles, by its simplicity, e.g. the idea on which the normal plot is based. This is the strong point of the approach and it seems to be more than sufficient repay for the mentioned disadvantages. Of course, one can argue against, as well as for, the present choice of the density estimator and of the distance (i.e. against or for the kernel estimator and the weighted Hellinger distance). They were chosen to demonstrate that at least in a simple case the approach is tractable.

Naturally, there may appear a question why we did not look directly for the estimator defined as an argument minimizing the proposed test statistic $H_n(Y, \beta)$. It is true that we may study estimators based on the kernel estimators directly (Víšek [18]). However, the theory accompanying such estimators seems to be inevitably overcrowded by technicalities, and precise evaluation of the estimators of this type is nearly impossible. Therefore the efficiency of such a research is questionable. On the other hand, the evaluation of $H_n(Y, \beta)$ and $\mathscr{H}_n(Y, \beta)$ for several regression model estimates which we have at hand may be quite simple and quick (especially when we use some standard method of numerical integration – in the present paper the Romberg method was applied).

Acknowledgment. We would like to express our gratitude to the anonymous referee for the careful reading of the manuscript. In fact, part of the present form of Remark 1 is due to him/her.

REFERENCES

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Dover Publications, 1964.

[2] P. Boček and P. Lachout, *Linear programming approach to LMS-estimation*, in: *Memorial volume of "Comput. Statist. Data Anal." devoted to T. Havránek* 19 (1995), pp. 129–134.

[3] P. Boček and J. Á. Víšek, *On one-step M-estimates*, in: *Transactions of the Fifth Prague Symposium on Asymptotic Statistics*, M. Hušková and P. Mandl (Eds.), Physica Verlag, 1994, pp. 195–202.

[4] L. Breiman, *Probability*, Addison-Wesley, London 1968.

[5] M. Csörgő and P. Révész, *Strong Approximations in Probability and Statistics*, Akadémiai Kiadi, Budapest 1981.

[6] F. R. Hampel, P. J. Rousseeuw, E. M. Ronchetti and W. A. Stahel, *Robust Statistics – The Approach Based on Influence Functions*, Wiley, New York 1986.

[7] R. V. Hogg, *Adaptive robust procedures: A partial review and some suggestions for future applications and theory (with discussion)*, J. Amer. Statist. Assoc. 69 (1974), pp. 909–927.

[8] R. M. Humphreys, *Studies of luminous stars in nearby galaxies. I. Supergiants and* 0 *stars in the milky way*, Astrophys. J. Suppl. Ser. 38 (1978), pp. 309–350.

[9] J. Jurečková, *Consistency of M-estimators in linear model generated by nonmonotone and discontinuous ψ-functions*, Probab. Math. Statist. 10 (1988), pp. 1–10.

[10] J. Jurečková, *Consistency of M-estimators of vector parameters*, in: *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics*, Charles University, Prague 1988, pp. 305–312.

[11] J. Jurečková and S. Portnoy, *Asymptotics for one-step M-estimators in regression with application to combining efficiency and high breakdown point*, Comm. Statist. A 16 (1988), pp. 2187–2199.

[12] M. Markatou, W. A. Stahel and E. M. Ronchetti, *Robust M-type testing procedures for linear models*, in: *Directions in Robust Statistics and Diagnostics*, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 201–220.

[13] R. A. Maronna and V. J. Yohai, *Asymptotic behaviour of general M-estimates for regression and scale with random carriers*, Z. Wahrsch. verw. Gebiete 58 (1981), pp. 7–20.

[14] R. D. Martin, V. J. Yohai and R. H. Zamar, *Min-max bias robust regression*, Ann. Statist. 17 (1989), pp. 1608–1630.

[15] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, New York 1987.

[16] S. S. Shapiro and M. B. Wilk, *A comparative study of various tests for normality*, J. Amer. Statist. Assoc. 63 (1968), pp. 1343–1372.

[17] J. Á. Víšek, *Stability of regression model estimates with respect to subsamples*, Comput. Statist. 7 (1992), pp. 183–203.

[18] J. Á. Víšek, *Adaptive maximum-likelihood-like estimation in linear model*. Part I. *Consistency*, Kybernetika 28 (1992), pp. 357–382. Part II. *Asymptotic normality*, Kybernetika 28 (1992), pp. 454–471.

[19] J. Á. Víšek, *On the role of contamination level and the least favourable behaviour of gross-error sensitivity*, Probab. Math. Statist. 14 (1993), pp. 173–187.

[20] J. Á. Víšek, *A cautionary note on the method of Least Median of Squares reconsidered*, in: *Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Prague 1994, pp. 254–259.

[21] J. Á. Víšek, *High robustness and an illusion of truth*, in: *Transactions of ROBUST'94*, published by JČMF (Union of Czech Mathematicians), J. Antoch and G. Dohnal (Eds.), 1994, pp. 172–185.

[22] J. Á. Víšek, *On high breakdown point estimation*, Comput. Statist. 11 (1996), pp. 137–146.

[23] J. Á. Víšek, *Contamination level and sensitivity of robust tests*, in: *Handbook of Statistics*, Vol. 15, G. S. Maddala and C. R. Rao (Eds.), Elsevier Science B. V., Amsterdam 1997, pp. 633–642.

[24] J. Á. Víšek, *The least trimmed squares – random carriers*, Bulletin of the Czech Econometric Society 10 (1999), pp. 1–30.

[25] J. Á. Víšek, *On the diversity of estimates*, Comput. Statist. Data Anal. 34 (2000), pp. 67–89.

[26] J. Á. Víšek, *Sensitivity analysis of M-estimates of nonlinear model: influence of data subsets*, Ann. Inst. Statist. Math., to appear.

[27] V. J. Yohai and R. H. Zamar, *A minimax-bias property of the least α-quantile estimates*, Ann. Statist. 21 (1993), pp. 1824–1842.

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
Opletalova ulice 26,
CZ – 110 00 Prague 1, Czech Republic
E-mail: visek@mbox.fsv.cuni.cz

Received on 22.11.1999;
revised version on 25.6.2001