PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 23, Fasc. 2 (2003), pp. 435–483

DURBIN–WATSON STATISTIC IN ROBUST REGRESSION

BY

JAN ÁMOS VÍŠEK* (PRAGUE)

Abstract. It is shown that the lower and upper critical values of the Durbin–Watson (D–W) statistic are asymptotically the same for the analysis based on M-estimators as for the classical least squares analysis. Moreover, the paper offers a possibility to make an idea when the asymptotics may start to work. Considering the B-robust optimal ψ-function, we demonstrate that the differences between the precise critical values of Durbin–Watson statistics evaluated for residuals corresponding to the M-estimate and critical values which were found by Durbin and Watson for the least squares analysis are rather small even for moderate sample size.

AMS Subject Classification: 62J20, 62F05.

Key words and phrases: Regression, diagnostics, M-estimators, robustified D–W statistics, critical values.

INTRODUCTION

Diagnostics are today an inseparable part of any data processing theory, see e.g. Atkinson (1985), Belsley et al. (1980), Geisser (1991), Greene (1993), Hauser (1997), Judge et al. (1985), McKean et al. (1991), Peters and Sibbertsen (2003), Schall and Dunne (1991), Tukey (1991), Urzua (1996) and many others. In fact, without an application of at least some basic diagnostic tools any processing of data loses its reliability and trustworthiness.

The econometrics offer diagnostics which are so rich that one has possibility to select for any problem really an efficient tool. Let us remind diagonal elements of a hat matrix (or an extended hat matrix) or many types of distances as Cook's, Mahalanobis' etc. which are employed to "discover" the influential points. Let us mention the condition indices for looking for collinearity, a lot of tests for heteroscedasticity as well as for the change points of model etc.; see Antoch and Vorlíčková (1992), Arslan (2003), Chatterjee and Hadi (1988), Cook and Weisberg (1983), Croux and Haesbroeck (1999), Hadi

* Research was supported by grant of GA CR number 402/03/0084.

(1992), (1994), Önder and Zaman (2003), Judge et al. (1985), Kmenta (1986), Markatou and He (1994) or Zvára (1989).

So it is not surprising that the robust statistics from very early days have also attempted to create appropriate diagnostics (Huber (1981), Welsh (1982), Hampel et al. (1986), Rousseeuw and Leroy (1987)), and the attention devoted to the topic is increasing (Huber (1991), Portnoy (1991), Stahel and Weisberg (1991), Willem et al. (2003)).

Nevertheless, an inspection of just mentioned references may indicate that there are mostly attempts to employ as diagnostic tools the robust procedures themselves. Much less attention was however paid to possibility to utilize the ideas of diagnostics of classical econometrics based on the least squares analysis (Chatterjee and Hadi (1988), Judge et al. (1985)) and to employ or to modify them for robust methods (Víšek (1992), (1996b, c)). Maybe that the roots of it are in an idea that robust procedures are "self-diagnosing". One can in fact trace out the spirit of it in the most papers devoted to the diagnostics in Stahel and Weisberg (1991). It is still believed that the application of a procedure with high breakdown point inevitably brings an indication of what is the "true" underlying model (of course, at a cost of some loss of efficiency). However, Hettmansperger and Sheather (1992) and Víšek (1994), (1996a, b), (2000a) show how misleading idea it is. Among others, in the referred papers the data are presented for which the least trimmed squares (LTS) and the least median of squares (LMS), both being methods with (possibly) 50% breakdown point, give mutually orthogonal estimates of regression models. Moreover, the model obtained for the same data by means of min-max bias estimation (see Martin et al. (1989)) appears to be completely different from the LTS as well as the LMS model and very "wild" what concerns residuals. But there are even more shocking results, showing that one can easily find such data that an arbitrarily small shift of one observation can imply a "rotation" of the hight breakdown point estimate of the regression model about 90 degrees; see Víšek (1996a) or (1997).

On the other hand, to be frank, the author believes that in the most situations some method with high breakdown point may serve very efficiently as diagnostic tools and the "final" estimate may be found by means of the ordinary least squares after deleting some observations, or by the weighted least squares, just weighting adequately down the influence of some points. The recent results of the author hint that one of the most promising methods (let us stress, we have in mind a linear regression) is probably the least weighted squares[1] method (see the discussion in Víšek (1999a) or (2001b)). And only in some exceptional

[1] Notice that we have said *the least weighted squares*, which differs from *the weighted least squares* in a similar manner to that *the least trimmed squares* are different from *the trimmed least squares*. In the both cases the latter methods, i.e. for *the trimmed least squares* and *the weighted least squares*, the trimming as well as the weighting are given explicitly, i.e. by an external explicit rule, while for *the least trimmed squares* and *the least weighted squares* the trimming and the weighting are given implicitly.

cases, when *a posteriori* diagnostics indicate that the estimate of model has e.g. rather "wild" structure of residuals, we should use several robust methods with high breakdown point to investigate all hidden substructures in data.

The present paper however will not discuss these tasks in detail (for more details see e.g. Víšek (1994), (1996a, b), (1997), (2000a) or (2001b)).

An instructive application of the robust method for diagnosing data can be found in Víšek (1999c) or (2003).

As one can easily learn, there are natural assumptions of the classical (regression) analysis which are accepted in the theoretical considerations of robust statistics and employed in applications without any (*a posteriori*) diagnostic verification. It is simply a consequence of the fact that corresponding tools are not yet available. For instance, the independence of the random fluctuations for different rows of the dynamic regression scheme for dynamic situation or the independence of explanatory variables and random fluctuations belong among such assumptions; see e.g. Huber (1981), Hampel et al. (1986), Boente and Fraiman (1991), Dollinger and Staudte (1991), Hettmansperger and Naranjo (1991), Lawrence (1991), Markatou et al. (1991) and many others. And it is well known that a break of the former assumption may cause (sometimes serious) decrease of efficiency while a failure of the latter causes a bias of the estimation of the least squares. And it is easy to guess, e.g. from the asymptotic representation of M-estimators, that in the case of robust analysis a failure of just mentioned assumptions may cause a similar damage of results as in the case of the least squares.

Of course, one can object that sometimes (frequently?) it is better to put up with a loss of efficiency than to make a correction (e.g. by means of the Prais–Winsten transformation, see Judge et al. (1985)) because if we wrongly assume e.g. an autoregressive structure of disturbances, we can even worsen the situation, see e.g. Mizon (1995).

There have already been known some results trying to establish diagnostics for robust methods. We have now at our disposal a robustified version of instrumental variables and the Hausman specification test (Víšek (1998a, b) and (2000b)), results of the sensitivity studies of some robust methods (Víšek (2000c) and (2001a)) and the White test for heteroscedasticity (Víšek (2002b)).

The present paper brings another result. It shows that under normality of disturbances the critical values of "lower" and "upper" Durbin–Watson statistics evaluated for residuals corresponding to an M-estimator are not dramatically different from those used in least squares analysis, and asymptotically they are even the same. For the special case of the optimal B-robust M-estimator of the regression coefficients for the frequently used values of the tuning constant they are even only slightly lower and higher than those given by Durbin and Watson for the least squares.

Of course, to have at all a possibility to compare results by Durbin and Watson with those for M-estimators, we had to assume normality of disturbances. Naturally, in the case when we decide to apply a robust method we assume that the normally distributed disturbances were contaminated by something, so that we should either delete some observations or weight them down to reach again normality. It may seem at a first glance rather restrictive or even senseless, since one can claim that the distribution of disturbances is something which is objectively given. Let us try to show that it need not be the case. Moreover, we shall explain why the normality is very important.

First of all, let us consider the classical ordinary least squares. Some monographs claim that even in the case when the normality of disturbances was rejected, the ordinary least squares are still a good estimator, since they are the best among all unbiased linear estimators. But the restriction on the class of linear estimators is drastic – for a nice discussion see Hampel et al. (1986). So in the case when the normality of disturbances is rejected, we should use some alternative method, e.g. the maximum likelihood estimator – nevertheless, practically all commercially available statistical packages use normal likelihood anyhow. However, there are only a few results about optimality of alternative methods at all. Moreover, the least squares are closely associated with the Euclidean geometry we are used to, and hence we would like to employ the least squares anyhow. That is why we try to reach the normality of disturbances by modifications of the model, e.g. by transformations of both response and explanatory variables. Of course, a philosophical point of view needs at least partially leave a "Renaissance" standpoint about objectivity of mathematical modelling, see e.g. Kuhn (1965), Popper (1972), Prigogine and Stengers (1977), (1984). In other words, we give up an idea of "discovering" (or "revealing") something like a "true" underlying model, being satisfied by establishing a reliable forecasting model.

Similarly, in the case of using the optimal B-robust M-estimator the normality is in fact frequently implicitly assumed since the corresponding ψ-function is (typically) a winsorised logarithmic derivative of normal density.

However, the assumption of normality, when considering D–W statistic, goes much deeper. It stems from the work of Anderson (1948) and of Durbin and Watson (1952) who constructed the test by means of the Neyman–Pearson lemma so that to be optimal under normality. It implies that in the case when the disturbances are not normally distributed and, as we have already mentioned, we should apply even in the classical regression analysis some other method than the least squares, we should also use another statistic (than D–W) for testing serial correlation (again, they are not offered in statistical packages).

Taking into account just discussed reasons for reaching normality in the case of employment of ordinary least squares together with the fact that in the case of robust methods the most complete results we have at hand for M-estimators, we may guess (of course, not prove) that it is better to apply some of them, assuming that we depressed an impact of contamination of data by weighting down (possibly to zero) the influence of some observations, reaching so an approximative normality. On the other hand, M-estimators are not generally scale- and regression-equivariant (see the last chapter of the paper) and have a bounded breakdown point, see Maronna and Yohai (1981). So, it may be preferable to use a method which is asymptotically equivalent to an M-estimator and simultaneously is "automatically" scale- and regression-equivariant with assignable breakdown point, e.g. the least weighted squares, see Víšek (2001b), (2002b).

The paper is organized as follows. In the next section we shall introduce notational framework for our considerations. Then we recall Durbin–Watson results, and in the following section we derive the basic tool for further study, namely an asymptotic representation of the M-estimators in the setup with random explanatory variables. The next step is the explanation of the modifications of the approach of Durbin and Watson which are necessary to be carried out for robust setup. At the end of the paper we shall give tables showing numerical approximations to the difference between the "lower" and "upper" Durbin–Watson statistics for the residuals from regression analysis based on the least squares estimator and on M-estimators.

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

$$Y_t = X_t^{\mathrm{T}} \beta^0 + e_t, \qquad t = 1, 2, \ldots, n, \tag{1}$$

where $Y = (Y_1, Y_2, \ldots, Y_n)^{\mathrm{T}}$ is the response variable, X_t^{T} is the t-th row of the random design matrix $X = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$, β^0 is the "true" vector of regression coefficients, and $e = (e_1, e_2, \ldots, e_n)^{\mathrm{T}}$ is the initial part of the sequence $\{e_t\}_{t=1}^{\infty}$ of independent identically distributed random variables. Sometimes we shall use the "matrix" notation for the regression model

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

Further, for any $\beta \in R^p$ let us define the t-th residual as

$$r_t(\beta) = Y_t - X_t^{\mathrm{T}} \beta \quad \text{and} \quad r^{(n)}(\beta) = \big(r_1(\beta), r_2(\beta), \ldots, r_n(\beta)\big)^{\mathrm{T}}. \tag{2}$$

RECALLING THE HISTORY

Durbin–Watson statistic which is today offered by many statistical packages (STASTICA, SPSS, TSP, RATS, to mention some among others) stems from the work by Anderson (1948). Assuming that the random fluctuations e_t's are governed by the multivariate normal distribution with covariance matrix either Ψ^{-1} (under hypothesis) or Φ^{-1} (under alternative), Anderson showed that in the cases in which the vector of regression coefficients is the eigenvector of the matrices Ψ and Φ, the statistic

$$z(\hat{\beta}^{(LS,n)}) = \frac{[r^{(n)}(\hat{\beta}^{(LS,n)})]^{\mathrm{T}}\ \Phi\ r^{(n)}(\hat{\beta}^{(LS,n)})}{[r^{(n)}(\hat{\beta}^{(LS,n)})]^{\mathrm{T}}\ \Psi\ r^{(n)}(\hat{\beta}^{(LS,n)})},$$

where $\hat{\beta}^{(LS,n)}$ is of course the least squares estimator, provides a test which is uniformly most powerful against a certain set of alternatives. The evident dependence of the statistics $z(\hat{\beta}^{(LS,n)})$ on the design matrix (which follows from $r(\hat{\beta}^{(LS,n)}) = [\mathscr{I}_n - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}]\,e$, where $\mathscr{I}_n$ is a unit matrix of type $n \times n$) and, consequently, a similar dependence of critical points on the design matrix, seemed to denounce the statistic to be a useless theoretical result. Durbin and Watson (1952) returned to the statistic for the special case when $\Psi = \mathscr{I}_n$, i.e. when we assume as the null hypothesis the independence of random fluctuations in the model (1). They slightly changed the notation used by Anderson and wrote the statistic as

$$z(\hat{\beta}^{(LS,n)}) = \frac{[r^{(n)}(\hat{\beta}^{(LS,n)})]^{\mathrm{T}}\ A\ r^{(n)}(\hat{\beta}^{(LS,n)})}{[r^{(n)}(\hat{\beta}^{(LS,n)})]^{\mathrm{T}}\ r^{(n)}(\hat{\beta}^{(LS,n)})} \tag{3}$$

and invented a spectacular "trick" which turned the statistic to be one of the most frequently used diagnostic tools. The "trick" is described in the following lemma:

LEMMA 1 (Durbin and Watson (1952)). *If r and e are $n \times 1$ vectors such that*

$$r = M \cdot e, \qquad \text{where } M = \mathscr{I}_n - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}},$$

and

$$z = \frac{r^{\mathrm{T}}Ar}{r^{\mathrm{T}}r}$$

(see (3)), *where A is a real symmetric matrix, then:*

(a) *There is an orthogonal transformation $e = H\xi$ such that*

$$z(\hat{\beta}^{(LS,n)}) = \frac{\sum_{t=1}^{n-p} \nu_t \xi_t^2}{\sum_{t=1}^{n-p} \xi_t^2}, \tag{4}$$

where $\nu_1, \nu_2, \ldots, \nu_{n-p}$ are eigenvalues of the matrix MA other than zeros.

(b) *If $n-p-s$ of columns of X are linear combinations of $n-p-s$ of the eigenvectors of A, then $n-p-s$ of ν's are equal to the eigenvalues corresponding to these eigenvectors; renumbering the remaining eigenvalues so that*

$$\nu_1 \leqslant \nu_2 \leqslant \ldots \leqslant \nu_s \quad \text{and} \quad \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_{s+p},$$

where λ's are eigenvalues of the matrix A, we have

$$\lambda_t \leqslant \nu_t \leqslant \lambda_{t+p}, \quad t = 1, 2, \ldots, s. \tag{5}$$

For the proof see Durbin and Watson (1952). Specifying then the alternative to the independence of e_t's as $AR(1)$ with $e_t = ae_{t-1} + v_t$, $t = 2, 3, \ldots, n$ (with $\{v_t\}_{t=1}^{\infty}$ being of course i.i.d. and $|a| < 1$), by a straightforward computation we obtain

$$A_a = \begin{bmatrix} 1 & -a & & & \\ -a & 1+a^2 & -a & & \\ & -a & \ddots & \ddots & \\ & & \ddots & \ddots & \\ & & & -a & 1 \end{bmatrix}, \tag{6}$$

where the blank space represents zeros. An "extreme" case, namely $a = 1$, gives then the traditional form of Durbin–Watson statistic:

$$\begin{aligned} z(\hat{\beta}^{(LS,n)}) &= \frac{\sum_{t=1}^{n-1} [r_t(\hat{\beta}) - r_{t+1}(\hat{\beta})]^2}{\sum_{t=1}^{n} r_t^2(\hat{\beta})} \\ &= 2 - \frac{2\sum_{t=1}^{n} r_t(\hat{\beta}) r_{t+1}(\hat{\beta}) + r_1^2(\hat{\beta}) + r_n^2(\hat{\beta})}{\sum_{t=1}^{n} r_t^2(\hat{\beta})} \approx 2 - \hat{\varrho}. \end{aligned} \tag{7}$$

For the second "extreme" case, namely $a = -1$, we obtain $z(\beta) \approx 2 + \hat{\varrho}$. It is clear that the statistics resulting from the both extreme cases are able to cope with positive ($a > 0$) as well as negative ($a < 0$) dependence since both contain an estimate of the correlation coefficient. Historical reasons established (7) as the usual form of Durbin–Watson statistic (see e.g. Kmenta (1986), Judge et al. (1985) or Zvára (1989)) but the consequence is that we use not only the critical values "implied" by

$$z_L(\hat{\beta}^{(LS,n)}) = \frac{\sum_{t=1}^{n-p} \lambda_t \xi_t^2}{\sum_{t=1}^{n-p} \xi_t^2} \quad \text{and} \quad z_U(\hat{\beta}^{(LS,n)}) = \frac{\sum_{t=1}^{n-p} \lambda_{t+p} \xi_t^2}{\sum_{t=1}^{n-p} \xi_t^2}, \tag{8}$$

say z_L and z_U, but also their "mirror reflections", i.e. we reject the hypothesis about independence when $z(\hat{\beta}^{(LS,n)}) < z_L$ or $z(\hat{\beta}^{(LS,n)}) > 4 - z_L$, and we do not reject this hypothesis if $z(\hat{\beta}^{(LS,n)}) \in (z_U, 4 - z_U)$. In the other cases the result of the test is (unfortunately) not decisive and the evaluation of the precise critical value for a given design matrix is advised, see Judge et al. (1985). These lower and upper bounds z_L and z_U were found by means of an approximation to the distribution function of $z_L(\hat{\beta}^{(LS,n)})$ and of $z_U(\hat{\beta}^{(LS,n)})$ based on the moments of these two statistics. The moments were established by using results of Pitman (1937) or von Neumann (1941), which reads:

ASSERTION 1. *Under the assumption of normality of random fluctuations in the regression model* (1), $z(\hat{\beta}^{(LS,n)})$ *and* $\sum_{t=1}^{n} \xi_t^2$ *are independent.*

Remark 1. Notice that $\sum_{t=1}^{n} \xi_t^2$ is the denominator of $z(\hat{\beta}^{(LS,n)})$.

COROLLARY 1. *Under the assumption of Assertion* 1 *we have*

$$E\{\sum_{t=1}^{n} v_t \xi_t^2\} = Ez(\hat{\beta}^{(LS,n)}) \cdot E\{\sum_{t=1}^{n} \xi_t^2\}.$$

(We have recalled this result because we shall need it later.)

Details about the Durbin–Watson contribution as well as a genesis of the matrix H (see Lemma 1) will be clear from the text generalizing their result (which will be presented in one of the next sections). As we have already indicated in the Introduction, to be able to study Durbin–Watson statistics for M-estimators, we shall need an asymptotic representation of these estimators for the framework we have introduced.

ASYMPTOTIC REPRESENTATION OF M-ESTIMATORS

We are going to establish an asymptotic representation of M-estimators. In order to achieve it, we employ the asymptotic linearity of the normal equation for M-estimator. The pioneering paper about the method was written by Portnoy (1983). The method was later used many times, see e.g. Jurečková and Malý (1995), Jurečková and Portnoy (1988), Jurečková and Sen (1989), Jurečková and Welsh (1990), Rubio and Víšek (1993), or Víšek (1998a, b). The method allows us (differently from the method used in Pollard (1991), Rao and Zhao (1992) or Yohai and Maronna (1979)) to derive also more complicated results, e.g. subsample sensitivity of the estimators; see Víšek (1996c) and (2002a).

Let us recall that the M-estimator, for an absolutely continuous (and, frequently but not necessarily, convex) function ϱ, is defined as

$$\hat{\beta}^{(M,n)} = \arg\min_{\beta \in R^p} \sum_{t=1}^{n} \varrho(Y_t - X_t^{\mathrm{T}} \beta) \tag{9}$$

and it is usually found as a solution of the equation

$$\sum_{t=1}^{n} X_t \psi(Y_t - X_t^{\mathrm{T}} \beta) = 0, \tag{10}$$

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

$$[X^{\mathrm{T}} X]^{-1/2} \sum_{t=1}^{n} X_t \psi(Y_t - X_t^{\mathrm{T}} \beta) = o_p(1) \tag{11}$$

instead of (10) allows us even to include the estimators generated by discontinuous ψ-functions (see Rao and Zhao (1992), compare also Jurečková and Welsh (1990)). Finally, studentizing the residuals (Huber (1964), (1981))

$$\hat{\beta}^{(M,n)} = \arg\min_{\beta \in R^p} \sum_{t=1}^{n} \varrho\left(\frac{Y_t - X_t^{\mathrm{T}} \beta}{\hat{\sigma}_{(n)}}\right), \tag{12}$$

where $\hat{\sigma}_{(n)}$ is a (preliminary) $\sqrt{n}$-consistent estimator of σ_{e_1} which is assumed to be scale-equivariant and regression-invariant, we obtain scale- and regression--equivariance of the estimator; see Bickel (1975) and Jurečková and Šen (1993).

As we shall see later, the asymptotic representations will be a key tool for considering the modification of D–W statistic. We shall find them using asymptotic linearity of "normal" equations for M-estimators.

Now, let us describe a range of ψ-functions for which we shall derive the promised results.

ASSUMPTIONS $\mathscr{A}$. The function ψ allows a decomposition

$$\psi = \psi_a + \psi_c + \psi_s, \tag{13}$$

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 derivative ψ_c' which is a step-function with a finite number of jump-points, and ψ_c is constant in a neighborhood of $-\infty$ and $+\infty$.

• ψ_s is a step-function with steps at points $r_1, r_2, \ldots, r_m$, i.e. there are $\alpha_0, \alpha_1, \ldots, \alpha_m$ so that $\psi_s(z) = \alpha_0$ for $z \in (-\infty, r_1)$, $\psi_s(z) = \alpha_l$ for $z \in (r_l, r_{l+1})$, $l = 1, 2, \ldots, m-1$, and $\psi_s(z) = \alpha_m$ for $z \in (r_m, \infty)$.

Moreover, $E\{\psi(e_1 \sigma_{e_1}^{-1})\} = 0$ and denote by $\psi'(z)$ the derivative of $\psi(z)$ where it exists. Finally, $\gamma = E\psi'(e_1 \sigma_{e_1}^{-1})$ is nonzero and finite.

Remark 2. The decomposition (13) is due to Jurečková (1988), see also Jurečková and Portnoy (1988). For the discussion that Assumptions $\mathscr{A}$ do not represent a considerable restriction see e.g. Hampel et al. (1986).

The conditions on explanatory variables and random fluctuations will be as follows.

ASSUMPTIONS $\mathscr{B}$. The sequences $\{X_t^{\mathrm{T}}\}_{t=1}^{\infty}$ and $\{e_t\}_{t=1}^{\infty}$ ($X_t \in R^p$, $e_t \in R$) are sequences of independent identically distributed random variables which are mutually independent. Moreover, $X_{t1} = 1$, $EX_{1j} = 0$, $EX_{1j}^2 = \sigma_{X_j}^2 \in (0, \infty)$ for $j = 2, 3, \ldots, p$ and $E|e_1| < \infty$. Finally, $E[X_1 X_1^{\mathrm{T}}] = Q$, $W = \sigma_{e_1}^{-1} E[e_1 \psi'(e_1 \sigma_{e_1}^{-1})]$ and $E[X_{1j}^2 X_{1k}^2]$ exist and are finite for $j, k = 2, 3, \ldots, p$, and put $Q_n = n^{-1} X^{\mathrm{T}} X$ (where X is the design matrix; see the text below (1)).

Remark 3. Without loss of generality we could assume that $W = 0$. In fact, it represents a shift of the derivative of ψ-function in the horizontal direction, i.e. $\tilde{\psi}'(z\sigma_{e_1}^{-1}) = \psi'(z\sigma_{e_1}^{-1} - d)$ for some $d \in R$. Shifting then the ψ-function in vertical direction, simply taking a modified function $\tilde{\tilde{\psi}}(z\sigma_{e_1}^{-1}) = \tilde{\psi}(z\sigma_{e_1}^{-1}) - E\tilde{\psi}(e_1\sigma_{e_1}^{-1})$, we may reach $E\{\tilde{\tilde{\psi}}(e_1\sigma_{e_1}^{-1})\} = 0$. The last modification does not change derivative $\tilde{\psi}'(z)$, and hence $W = 0$ will be kept. So this assumption is (nearly) of the same type as the assumption that the mean influence of the random fluctuations on the response variable is compensated, i.e. that $Ee_1 = 0$ (which we adopt in the least squares analysis) or $E\{\psi(e_1\sigma_{e_1}^{-1})\} = 0$ (which we assume for M-analysis).

Moreover, since the optimal B- and V-robust ψ-functions are of the shape

$$\psi(z) = \max\{-b, \min\{-f_{e_1}'(z)/f_{e_1}(z), b\}\},$$

where $f_{e_1}(z)$ is the density of random fluctuations, for symmetrically distributed random fluctuations we have $\psi(-z) = -\psi(z)$. 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 asymmetry of random fluctuations, we can consider $W = 0$.

Asymptotic linearity of M-statistics. At first we shall consider

$$S_n(\tau, \kappa) = \sum_{t=1}^{n} X_t\{\psi([e_t - n^{-1/2}X_t^{\mathrm{T}}\tau]\,\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa)) - \psi(e_t\sigma_{e_1}^{-1})\}$$

and we shall put for arbitrary $\Theta > 1$

$$\mathcal{T}_\Theta = \{\tau \in R^p, \kappa \in R; \max\{\|\tau\|, |\kappa|\} < \Theta\}.$$

In the proofs of the next theorems (which are postponed into Appendix I) some constants C_l'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 absolutely continuous derivative, i.e. $\psi = \psi_a$, and Assumptions $\mathcal{B}$ hold. Further, let*

$$E\{\psi(e_1\sigma_{e_1}^{-1})\} = 0 \quad \text{and} \quad 0 < E[e_1\psi'(e_1\sigma_{e_1}^{-1})]^2 < \infty. \tag{14}$$

Then for any $\Theta > 0$

$$\sup_{\mathcal{T}_\Theta}\left\|S_n(\tau, \kappa) + n^{1/2}\gamma Q_n\tau + n^{-1/2}W\sum_{t=1}^{n}X_t\kappa\right\| = \mathcal{O}_p(1) \quad \text{as } n \to \infty. \tag{15}$$

Remark 4. Let us observe 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 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.

It is nearly evident and it will be clear from the proof that the assertion of the theorem can use Q instead of Q_n. However, for our further purposes it will be more convenient to have this form of assertion.

Remark 5. Let us notice that the proof of the theorem is essentially based on the character of the processes $\{X_t X_t^T \tau\psi'(e_t)\}_{\tau\in\mathcal{T}_\Theta}$, $\{\kappa X_t e_t \psi'(e_t)\}_{\kappa\in\mathcal{T}_\Theta}$, etc. which are the products of some fixed sequences of random variables and of parameters of the processes. It allows us, roughly speaking, to treat the suprema[2] 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, $-\infty = r_0 < r_1 < \ldots < r_{k+1} = \infty$. Let Assumptions $\mathcal{B}$ hold and*

$$\max\{E\|X_1\|^3, E|e_1|\} < \infty \quad \textit{with } E\{\psi(e_1\sigma_{e_1}^{-1})\} = 0.$$

Finally, assume that $F_{e_1}(z)$ (the distribution of random fluctuations) has a bounded density $f_{e_1}(z)$. Then for any $\Theta > 0$

$$(16)\quad \sup_{\mathcal{T}_\Theta}\{\|S_n(\tau, \kappa)+n^{1/2}\gamma Q_n\tau+n^{-1/2}W\sum_{t=1}^{n} X_t\kappa\|\} = \mathcal{O}_p(1) \quad \textit{as } n\to\infty.$$

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 $-\infty = r_0 < r_1 < \ldots < r_k < r_{m+1} = \infty$. Let again Assumptions $\mathcal{B}$ hold and suppose that $F_{e_1}(z)$ has a bounded density $f_{e_1}(z)$ which is Lipschitz of the first order. Finally, put*

$$q = \sum_{s=1}^{m}\{(\alpha_s-\alpha_{s-1})E[X_1X_1^T f_{e_1}(r_s\sigma_{e_1})]\}$$

and

$$w = \sum_{s=1}^{m}\{r_s(\alpha_s-\alpha_{s-1})E[X_1 f_{e_1}(r_s\sigma_{e_1})]\}.$$

Then for any $\Theta > 0$

$$(17)\quad \sup_{\mathcal{T}_\Theta}\{\|n^{-1/2}S_n(\tau, \kappa)+q\tau+w\kappa\|\} = \mathcal{O}_p(n^{-1/4}) \quad \textit{as } n\to\infty.$$

CONSISTENCY, ASYMPTOTIC REPRESENTATION AND NORMALITY OF *M*-ESTIMATORS

THEOREM 4. *Let $\psi = \psi_a+\psi_c$ and the assumptions of Theorems* 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 $\mathcal{B}$) *be positive definite. Then the*

[2] The suprema over all possible values of the parameter of process, i.e. over $\tau\in\mathcal{T}_\Theta$ or $\kappa\in\mathcal{T}_\Theta$.

equation

$$(18) \qquad \sum_{t=1}^{n} X_t \psi\left(\frac{Y_t - X_t^{\mathrm{T}} \beta}{\hat{\sigma}_{(n)}}\right) = 0$$

has a $\sqrt{n}$-consistent solution, i.e. there is $\hat{\beta}^{(M,n)}$ for which (18) *is fulfilled and*

$$\sqrt{n}(\hat{\beta}^{(M,n)} - \beta^0) = \mathcal{O}_p(1).$$

Remark 6. Notice that the application of (15) and (16) and the assumption that Q is positive definite play the key role in the proof of the previous theorem. In what follows we shall use the same in a slightly more complicated situation.

Corollary 2. *Let $E\{X_1 X_1^{\mathrm{T}}\} = Q$ be a positive definite matrix and $\hat{\sigma}_{(n)}^2$ a $\sqrt{n}$-consistent estimate of $\sigma_{e_1}^2$. Then under the assumptions of Theorem* 4 *we have*

$$(19) \qquad \sqrt{n}(\hat{\beta}^{(M,n)} - \beta^0)$$

$$= n^{-1/2} \gamma^{-1} Q_n^{-1} \sum_{t=1}^{n} X_t \{\psi(e_t \sigma_{e_1}^{-1}) - W(\log \hat{\sigma}_{(n)} - \log \sigma_{e_1})\} + \mathcal{O}_p(n^{-1/2}).$$

Remark 7. Notice that the matrix Q has a block-structure of the type

$$\begin{bmatrix} 1 & 0 \\ 0 & Q^{(1)} \end{bmatrix}$$

since $E[X_{1j} X_{11}] = EX_{1j} = 0, j = 2, 3, \ldots, p$. It implies that

$$Q^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & [Q^{(1)}]^{-1} \end{bmatrix}.$$

Now having rewritten the right-hand side of (19) into the form

$$(20) \qquad n^{-1/2} \gamma^{-1} Q_n^{-1} \sum_{t=1}^{n} X_t \psi(e_t \sigma_{e_1}^{-1})$$

$$- n^{-1/2} \gamma^{-1} Q_n^{-1} \sum_{t=1}^{n} X_t W(\log \hat{\sigma}_{(n)} - \log \sigma_{e_1}) + o_p(1)$$

and taking into account that except of the first coordinate the vector $n^{-1/2} \sum_{t=1}^{n} X_t$ 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 (20) considerably affects only the first coordinate of $\sqrt{n}(\hat{\beta}^{(M,n)} - \beta^0)$. After all, the situation is not surprising because it is known that generally we cannot simultaneously very well estimate a location and scale (leaving aside that for this moment the question whether the scale is really a good indicator of a spread of random variable) without requiring at least some weak assumption(s) on the underlying distribution. On the other hand, it is also known that an unbiased and efficient estimator of the location parameter may be constructed e.g. in the case when we assume nothing more than the symmetry of the underlying distribution (see Beran (1978), Stone (1975) or Víšek (1991)) but the symmetry is substantial. But then, using e.g. the B-optimal M-estimator, we will have $W=0$.

COROLLARY 3. *Let $E\{X_1 X_1^{\mathrm{T}}\} = Q$ be a positive definite matrix, $W=0$ and let $\hat{\sigma}^2_{(n)}$ be a $\sqrt{n}$-consistent estimate of the variance $\sigma^2_{e_1}$. Then, under the assumptions of Theorem 4, $\sqrt{n}(\hat{\beta}^{(M,n)}-\beta^0)$ is asymptotically normal with zero mean and a covariance matrix*

$$C = \gamma^{-2} E\{\psi^2(e_1 \sigma_{e_1}^{-1})\}\,[Q^{-1}]^{\mathrm{T}}.$$

Now let us turn our attention to the case of discontinuous ψ-functions. We are going to define an approximation to the M-estimator for a discontinuous ψ-function and to show its consistency and asymptotic normality. We shall do it along similar lines to those in the case of continuous ψ-function, of course with necessary modifications. Let us consider the ψ-function defined in Theorem 3 and let K be a positive constant. Then for any $n \in N$ and $\vartheta > \frac{1}{2}$ define

$$\tilde{\psi}_n(z) = \tfrac{1}{2}\{\alpha_s + \alpha_{s-1} + (\alpha_s - \alpha_{s-1})\, n^{\vartheta} K^{-1}(z - r_s)\}$$
$$\text{for } |z - r_s| < K n^{-\vartheta} \text{ and } s \in \{1, 2, \ldots, m\}$$

and

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

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

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*

$$\sum_{t=1}^{n} X_t \tilde{\psi}_n\left(\frac{Y_t - X_t^{\mathrm{T}} \tilde{\beta}^{(n)}}{\hat{\sigma}_{(n)}}\right) = 0 \tag{21}$$

and

$$\sqrt{n}(\tilde{\beta}^{(n)} - \beta^0) = \mathcal{O}_p(1).$$

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

DEFINITION 1. Under the *approximate M-estimator for any discontinuous ψ-function* we shall understand that solution of equation (21) which was described in Theorem 5.

Remark 8. Notice that in some sense the construction which was presented a few lines above and which was a justification of the previous definition also gives an idea how to find the approximate M-estimator. Of course, in the case when (21) has more solutions we have to choose one of them, similarly to the case when we obtain several solutions of (10). The above-mentioned construction (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 the M-estimator generated by a discontinuous ψ-function need not be very simple because it is necessary to solve directly the corresponding extremal problem and not only an equation of type (10), see e.g. Koenker and Bassett (1978). From the above however it 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 formally (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 (hopefully for finite n and small K).

On the other hand, for fixed n and for $K \to 0$, the solutions of (21) converge to solutions of (10) (with studentized residuals, if such a solution exists). 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 (10) exists and that this solution is the solution of the corresponding extremal problem (9) (of course, if we have more solutions of (10), then one of them is a solution of (9); see Rubio and Víšek (1996). Then this solution is probably nearly the same as a solution of (21) for small K. After all, both estimators have in this case the same asymptotic representation.

Simple consequences of Theorem 5 (and of Definition 1) can be given as the following corollary.

COROLLARY 4. *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 the assumptions of Theorem* 5 *we have*

$$\sqrt{n}(\tilde{\beta}^{(M,n)} - \beta^0) = n^{-1/2} q^{-1} \sum_{t=1}^{n} X_t \{\psi(e_t \sigma_{e_1}^{-1}) - w(\log \hat{\sigma}_{(n)} - \log \sigma_{e_1})\} + o_p(1).$$

Moreover, if $w = 0$, $\sqrt{n}(\tilde{\beta}^{(M,n)} - \beta^0)$ *is asymptotically normal with zero mean and a covariance matrix* $C = q^{-1} E\{\psi^2(e_1 \sigma_{e_1}^{-1})\} Q [q^{-1}]^{\mathrm{T}}$.

Finishing preliminary considerations we have at hand sufficient tools for the study of Durbin–Watson statistics for M-estimation.

Modifying Durbin–Watson trick. In this section we shall show that the critical values of $z(\hat{\beta}^{(LS,n)})$ and $z(\hat{\beta}^{(M,n)})$ are asymptotically the same.

First of all we give a technical remark. In view of finiteness of Latin alphabet and also due to some historical reasons which established some habits in the notation in M-estimation as well as in studies devoted to Durbin–Watson statistics, some letters in the next text will denote different objects than in the previous text. We believe that a misinterpretation is (nearly) impossible.

We shall need some well-known assertions (the proof of which can be found e.g. in Judge et al. (1985) or in Zvára (1989)).

ASSERTION 2. *Let the matrix M be idempotent. Then its rank is equal to its trace.*

COROLLARY 5. *Let the design matrix X (of type $n \times p$) be of full rank. Then $M = \mathscr{I} - X(X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}}$ has rank equal to $n-p$.*

ASSERTION 3. *Let the matrix M be real and symmetric (of type $p \times p$). Then there exists an orthogonal matrix L such that $L^{\mathrm{T}} ML$ is the diagonal matrix $\Lambda = \operatorname{diag}\{\varsigma_1, \varsigma_2, \ldots, \varsigma_p\}$, where $\varsigma_1, \varsigma_2, \ldots, \varsigma_p$ are eigenvalues of the matrix M.*

COROLLARY 6. *If M is a projection matrix, then there exists an orthogonal matrix L such that $L^{\mathrm{T}} ML$ is a diagonal matrix the diagonal of which consists of ones and zeros (only).*

By analogy with (3) let us put for any $\beta \in R^p$ (see also (2))

$$z(\beta) = \frac{[r^{(n)}(\beta)]^{\mathrm{T}} A\, r^{(n)}(\beta)}{[r^{(n)}(\beta)]^{\mathrm{T}}\, r^{(n)}(\beta)}.$$

In what follows we are going to modify Durbin–Watson statistics for M-estimators. Due to the fact that there are not (generally) closed formulae for M-estimators, our considerations about them will be (inevitably) asymptotic. Hence instead of considering one fixed matrix A, we will need a sequence of matrices $\{A^{(n)}\}_{n=1}^{\infty}$. Moreover, we shall assume that all of these matrices are of type A_a (see (6)). To simplify the considerations in the following, let us assume that a is a fixed number $(a \in (-1, 1))$ for all $n \in N$. It will allow us to omit superindex a without danger of misunderstanding.

THEOREM 6. *Let Assumptions $\mathscr{A}$ and $\mathscr{B}$ hold and let the random fluctuations in the regression model* (1) *be normally and independently distributed. Then there are $z_L(\hat{\beta}^{(M,n)})$ and $z_U(\hat{\beta}^{(M,n)})$ such that*

$$z_L(\hat{\beta}^{(M,n)}) \leqslant z(\hat{\beta}^{(M,n)}) \leqslant z_U(\hat{\beta}^{(M,n)})$$

and the critical values of the statistics $z_L(\hat{\beta}^{(M,n)})$ and $z_U(\hat{\beta}^{(M,n)})$ are asymptotically equal to the critical values of $z_L(\hat{\beta}^{(LS,n)})$ and $z_U(\hat{\beta}^{(LS,n)})$, respectively.

In the rest of the paper we shall give an idea for which n of these asymptotics may start to work for a somewhat special, however frequently used setup. In order to do this, let us inspect the magnitude of the corresponding terms of (90) (below).

First of all, let us recall that $\tilde{M}$ is the projection matrix, and hence it is decomposable into the product $\tilde{M}_1 \cdot \tilde{M}_2 \cdot \ldots \cdot \tilde{M}_p$ of p matrices of type $\mathscr{I} - u \cdot u^{\mathrm{T}}$, where $u \in R^n$ (in fact, these vectors u's are the eigenvectors of

$\tilde{M}$ corresponding to nonzero eigenvalues of $\tilde{M}$). Then we may apply the part (b) of Durbin–Watson's lemma and we obtain

$$(22) \qquad \lambda_t \leqslant \nu_{n-p+t} \leqslant \lambda_{t+n-p}, \qquad t = 1, 2, \ldots, p.$$

First of all, let us give an idea about the moments of variables in (90) (for the definitions of the variables see (87)). We will assume that e_t's are independent and identically distributed according to a symmetric distribution function (around zero) and that ψ is a B-optimal function, i.e. that it is asymmetric ($\psi(-z) = -\psi(z)$). It implies that $Ee_1 = 0$ as well as $E\psi(e_1) = 0$. The evaluation of moments of $\tilde{\xi}_t$ is nearly the same as in Durbin and Watson (1952) and very similar to the evaluation of ζ_t which we shall start with, since it is the simplest. The evaluation for the other variables (ϑ and $\tilde{\vartheta}$) will be similar. Firstly, we will be interested in the second and the forth moments of ζ_t's and in covariances of ζ_t^2 and ζ_s^2 (for $t \neq s$) because they are relevant for the first and the second moment of z.

For the second moment of ζ_t we have (remember that $\tilde{H}$ is orthonormal)

$$(23) \quad E\zeta_t^2 = E\left\{\sum_{j=1}^{n} \tilde{h}_{tj}\left[e_j - \frac{\psi(e_j)}{\gamma}\right]\right\}\left\{\sum_{k=1}^{n} \tilde{h}_{tk}\left[e_k - \frac{\psi(e_k)}{\gamma}\right]\right\}$$
$$= \sum_{j=1}^{n} (\tilde{h}_{tj})^2 E\left[e_j - \frac{\psi(e_j)}{\gamma}\right]^2 = E\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2 \sum_{j=1}^{n} (\tilde{h}_{tj})^2 = E\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2,$$

and Table 1 shows that the values of $E[e_1 - \psi(e_1)/\gamma]^2$, and hence also the value of $E\zeta_t^2$ for the Huber ψ-function and the frequently used values of tuning constant are really rather small (the Huber ψ-function is given by $\psi(z) = \max\{-const, \min\{z, const\}\}$ and $const > 0$ is frequently called a *tuning constant*). (Let us recall that in our framework $E\xi_t^2 = 1$ and that moreover in the numerator of (90) we sum $n-p$ of ξ_t^2's while only p of ζ_t^2's.)

For $E\zeta_t^4$ the situation is slightly different from that one for $E\xi_t^4$; see Durbin and Watson (1952). Since $\sum_{j=1}^{n} \tilde{h}_{tj}^2 = 1$ (for $t = 1, 2, \ldots, n$), we have $\sum_{j=1}^{n} \tilde{h}_{tj}^4 = 1$, and so

$$E\zeta_t^4 \doteq E\left\{\sum_{j=1}^{n} \tilde{h}_{tj}\left[e_j - \frac{\psi(e_j)}{\gamma}\right]\right\}^4$$
$$= \sum_{j=1}^{n} \tilde{h}_{tj}^4 \left\{E\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^4 - 3E^2\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2\right\} + 3E^2\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2$$
$$\leqslant \max\left\{E\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^4, 3E^2\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2\right\}.$$

Consequently, for $\operatorname{var}(\zeta_t^2)$ we have

$$(24) \quad \operatorname{var}(\zeta_t^2) \leqslant \max\left\{E\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^4 - E^2\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2, 2E^2\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2\right\}.$$

Further we need to estimate an upper bound for $\operatorname{var}(\sum_{t=n-p+1}^{n} v_t \zeta_t^2)$. First of all, from (22) we have

$$(25)\qquad \sum_{t=n-p+1}^{n} \operatorname{var}(v_t \zeta_t^2)$$

$$\leqslant \sum_{t=n-p+1}^{n} \lambda_t^2 \max\left\{\boldsymbol{E}\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^4-\boldsymbol{E}^2\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2,\ 2\boldsymbol{E}^2\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\right\}.$$

Then for $t \neq s$ (keep in mind that $\sum_{j=1}^{n} \tilde{h}_{tj}\tilde{h}_{sj}=0$)

$$\begin{aligned}
\boldsymbol{E}\{\zeta_t^2\zeta_s^2\} &= \boldsymbol{E}\left\{\sum_{j=1}^{n}\tilde{h}_{tj}\left[e_j-\frac{\psi(e_j)}{\gamma}\right]\right\}^2\left\{\sum_{k=1}^{n}\tilde{h}_{sk}\left[e_k-\frac{\psi(e_k)}{\gamma}\right]\right\}^2\\
&= \sum_{j=1}^{n}\tilde{h}_{tj}^2\tilde{h}_{sj}^2\left\{\boldsymbol{E}\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^4-\boldsymbol{E}^2\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\right\}\\
&\quad+\sum_{j=1}^{n}\sum_{k=1}^{n}\tilde{h}_{tj}^2\tilde{h}_{sk}^2\boldsymbol{E}^2\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\\
&= \sum_{j=1}^{n}\tilde{h}_{tj}^2\tilde{h}_{sj}^2\left\{\boldsymbol{E}\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^4-\boldsymbol{E}^2\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\right\}+\boldsymbol{E}^2\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2.
\end{aligned}$$

Since the mean value of ζ_t^2 is given in (23) and does not depend on t (and s), we have

$$\operatorname{cov}(\zeta_t^2,\zeta_s^2)=\left\{\boldsymbol{E}\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^4-3\boldsymbol{E}^2\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\right\}\sum_{j=1}^{n}\tilde{h}_{tj}^2\tilde{h}_{sj}^2.$$

Now let us consider an extremal problem:

For $u, v \in R^n$, $\|u\|=1$, $\|v\|=1$, $u^{\mathrm{T}}v=0$, find

$$\max\left\{a_u^2\sum_{j=1}^{n}u_j^4+2a_u a_v\sum_{j=1}^{n}u_j^2v_j^2+a_v^2\sum_{j=1}^{n}v_j^4\right\}.$$

A straightforward computation for $n=2$ shows that the maximum is attained for

$$u_1=\frac{1}{\sqrt{2}},\qquad u_2=\frac{1}{\sqrt{2}},\qquad v_1=\frac{1}{\sqrt{2}},\qquad v_2=-\frac{1}{\sqrt{2}}$$

and is bounded by $\frac{1}{2}(a_u^2+a_v^2)$. One can easily verify that the same solution is valid for any n just putting $u_j=0$, $v_j=0$ for $j=3,4,\ldots,n$. (Any other "distribution of mass" among the coordinates, by the restriction that $\|u\|=1$ and

$\|v\| = 1$, decreases the maximum.) But then we have a little bit more general extremal problem:

For $u_1, u_2, \ldots, u_p \in R^n$, $u_s^{\mathrm{T}} u_t = \delta_{st}$, $s, t = 1, 2, \ldots, p$, $v_t \geqslant 0$, *find*

$$\max \sum_{s=1}^{p} \sum_{t=1}^{p} v_s v_t \sum_{j=1}^{n} u_{sj}^2 u_{tj}^2.$$

The solution of this problem gives an upper bound of the maximum equal to $\frac{1}{2}\sum_{t=1}^{p} v_t^2$, and hence

$$\sum_{t=n-p+1}^{n} \sum_{s=n-p+1}^{n} v_t v_s \sum_{j=1}^{n} \tilde{h}_{tj}^2 \tilde{h}_{sj}^2 \leqslant \frac{1}{2} \sum_{t=n-p+1}^{n} v_t^2.$$

This yields

$$(26) \quad \Big| \sum_{t=n-p+1}^{n} \sum_{s=n-p+1}^{n} v_t v_s \operatorname{cov}(\zeta_t^2, \zeta_s^2) \Big| \leqslant \frac{1}{2} \sum_{t=n-p+1}^{n} v_t^2 \left| E\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^4 - E^2\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2 \right|.$$

Table 1 again indicates that the values of $\operatorname{var}(\zeta_t^2)$ as well as the values of $\operatorname{cov}(\zeta_t^2, \zeta_s^2)$ are not very large, and hopefully also $\operatorname{var}\left(\sum_{t=n-p+1}^{n} v_t \zeta_t^2\right)$ will not be large. Similarly for variables from the last but one term of the numerator of (90):

$$(27) \quad E\,\tilde{\xi}_t \zeta_t = E\Big\{\sum_{j=1}^{n} \tilde{h}_{tj} e_j\Big\}\left\{\sum_{k=1}^{n} \tilde{h}_{tk}\left[e_k - \frac{\psi(e_k)}{\gamma}\right]\right\} = \sum_{j=1}^{n} (\tilde{h}_{tj})^2 E\left\{e_j\left[e_j - \frac{\psi(e_j)}{\gamma}\right]\right\}$$
$$= E\left\{e_1\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]\right\} \sum_{j=1}^{n} (\tilde{h}_{tj})^2 = \left\{E\, e_1\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]\right\}.$$

One can see that derivations of (23) and of (27) are very similar, and hence in the derivation of further expressions (which are analogical to (24) and (26)) the details will be omitted. Keeping in mind that we have assumed that $Ee_1^2 = 1$, we obtain

$$E(\tilde{\xi}_t \zeta_t)^2 \doteq E\left\{\sum_{j=1}^{n} \tilde{h}_{tj} e_j \sum_{k=1}^{n} \tilde{h}_{tk}\left[e_k - \frac{\psi(e_k)}{\gamma}\right]\right\}^2$$
$$\leqslant \max\left\{E\left\{e_1\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]\right\}^2, E\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2 + 2E^2\left\{e_1\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]\right\}\right\},$$

which means that

$$\operatorname{var}(\tilde{\xi}_t \zeta_t) \leqslant \max\left\{E\left\{e_1\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]\right\}^2 - E^2\left\{e_1\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]\right\},\right.$$
$$\left. E\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2 + E^2\left\{e_1\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]\right\}\right\}.$$

Of course, as above we obtain

$$(28)\qquad \sum_{t=n-p+1}^{n} \operatorname{var}(v_t \tilde{\xi}_t \zeta_t)$$
$$\leqslant \sum_{t=n-p+1}^{n} \lambda_t^2 \max\left\{E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2+E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\},\right.$$
$$\left.E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2+E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right\}.$$

And again, for $t \neq s$ (still keep in mind that $\sum_{j=1}^{n} \tilde{h}_{tj}\tilde{h}_{sj}=0$) we have

$$E\tilde{\xi}_t\zeta_t\tilde{\xi}_s\zeta_s = E\left\{\sum_{j=1}^{n}\tilde{h}_{tj}e_j\sum_{l=1}^{n}\tilde{h}_{tl}\left[e_l-\frac{\psi(e_l)}{\gamma}\right]\sum_{k=1}^{n}\tilde{h}_{sk}e_k\sum_{m=1}^{n}\tilde{h}_{sm}\left[e_m-\frac{\psi(e_m)}{\gamma}\right]\right\}$$
$$=\sum_{j=1}^{n}\tilde{h}_{tj}^2\tilde{h}_{sj}^2\left\{E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2-E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2-2E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right\}$$
$$+2E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}$$
$$\leqslant \max\left\{\frac{1}{2}\left[E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2-E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\right]\right.$$
$$\left.+E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\},\ 2E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right\}.$$

Consequently, along similar lines to those as above we obtain

$$(29)\qquad \left|\sum_{t=n-p+1}^{n}\sum_{s=n-p+1}^{n} v_t v_s \operatorname{cov}(\tilde{\xi}_t\zeta_t,\ \tilde{\xi}_s\zeta_s)\right|$$
$$\leqslant 4\sum_{t=n-p+1}^{n}\lambda_t^2\max\left\{\frac{1}{2}\left[E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2-E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\right],\right.$$
$$\left.E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right\}.$$

Using (28) and (29) we will be able later to estimate $\operatorname{var}(\sum_{t=n-p+1}^{n} v_t\tilde{\xi}_t\zeta_t)$. We need also to estimate (an upper bound of) covariance of two last terms in the numerator of (90). We can derive in a similar way to that as above

$$(30)\qquad \left|\sum_{t=n-p+1}^{n}\sum_{s=n-p+1}^{n} v_t v_s \operatorname{cov}(\tilde{\xi}_t\zeta_t,\ \zeta_s^2)\right|$$

$$\leqslant 4 \sum_{t=n-p+1}^{n} \lambda_t^2 \max\left\{\frac{1}{2}\left[E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2 - E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\right],\right.$$
$$\left. E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right\}.$$

This concludes the study of two last terms of the numerator of (90). To give an idea about moments of random variables from the second term of the numerator of the fraction in (90) let us put

$$\tilde{e} = A\cdot e = (e_1-e_2,\, 2e_2-e_1-e_3,\, \ldots,\, e_n-e_{n-1})^{\mathrm{T}}.$$

In what follows, let l_{tj} denote elements of the matrix L^{T} (notice that in order to simplify the next considerations we have denoted the (t,j)-th element of L^{T} by l_{tj} instead of l_{jt}). Then $\vartheta = L^{\mathrm{T}}\tilde{e}$ can be written as $\vartheta_t = \sum_{j=1}^{n} l_{tj}\tilde{e}_j$, and hence

$$(31)\qquad E\{\vartheta_t \tilde{\vartheta}_t\} = E\left\{\sum_{j=1}^{n} l_{tj}\tilde{e}_j \sum_{k=1}^{n} l_{tk}\left[e_k-\frac{\psi(e_k)}{\gamma}\right]\right\}$$
$$= 2\sum_{j=1}^{n} l_{tj}^2 E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\} - \sum_{j=1}^{n-1} l_{tj+1}l_{tj}E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}$$
$$-\sum_{j=2}^{n} l_{tj-1}l_{tj}E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}.$$

Using the Cauchy–Schwarz inequality we arrive at

$$\Big|\sum_{j=2}^{n} l_{tj-1}l_{tj}\Big| \leqslant \Big(\sum_{j=2}^{n} l_{tj-1}^2 \sum_{j=2}^{n} l_{tj}^2\Big)^{1/2} \leqslant 1,$$

and so from (31) we obtain

$$|E\vartheta_t\tilde{\vartheta}_t| \leqslant 4\left|E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right|.$$

(On the other hand, for the special case of Huber's ψ-function, $e_1-\psi(e_1)/\gamma$ is nearly zero for $|e_1|$ smaller than tuning constant, and hence $E\{e_1[e_1-\psi(e_1)/\gamma]\}\approx 0$; in fact, it is less than 10^{-8} for all the values of tuning constant. That was the reason why it was not included into Table 1.) Similarly we have

$$(32)\qquad E\{\vartheta_t\tilde{\vartheta}_t\}^2 = E\left\{\sum_{j=1}^{n} l_{tj}\tilde{e}_j \sum_{k=1}^{n} l_{tk}\left[e_k-\frac{\psi(e_k)}{\gamma}\right]\right\}^2$$
$$= E\left\{2\sum_{j=1}^{n} l_{tj}e_j \sum_{k=1}^{n} l_{tk}\left[e_k-\frac{\psi(e_k)}{\gamma}\right] - \sum_{j=2}^{n} l_{tj-1}e_j\sum_{k=1}^{n} l_{tk}\left[e_k-\frac{\psi(e_k)}{\gamma}\right]\right.$$
$$\left. -\sum_{j=1}^{n-1} l_{tj+1}e_j\sum_{k=1}^{n} l_{tk}\left[e_k-\frac{\psi(e_k)}{\gamma}\right]\right\}^2.$$

We shall give only examples of terms obtained from (32). Firstly we obtain the mean value of squares of the corresponding terms. The first one is

$$(33)\quad 4\left\{E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2\sum_{j=1}^{n} l_{tj}^4+Ee_1^2 E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\sum_{j=1}^{n} l_{tj}^2\sum_{k\neq j} l_{tk}^2 \right.$$
$$\left. +2E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\sum_{j=1}^{n} l_{tj}^2\sum_{k\neq j} l_{tk}^2\right\},$$

the second one equals

$$(34)\quad E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2\sum_{j=2}^{n} l_{tj-1}^2 l_{tj}^2+Ee_1^2 E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\sum_{j=2}^{n} l_{tj}^2\sum_{k\neq j} l_{tk}^2$$
$$+2E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\sum_{j=2}^{n} l_{tj-1}l_{tj}\sum_{k\neq j} l_{tk-1}l_{tk},$$

and the third is similar to the second one. Further we obtain "cross-terms". The term which is the mean value of the product of the first and the second term from (32) has the form

$$(35)\quad -4\left\{E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2\sum_{j=2}^{n} l_{tj-1}l_{tj}^3+Ee_1^2 E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\right.$$
$$\left.\times\sum_{j=2}^{n} l_{tj}l_{tj-1}\sum_{k\neq j} l_{tk}^2+E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\sum_{j=2}^{n} l_{tj}l_{tj-1}\sum_{k\neq j} l_{tk}^2\right\}.$$

The product of the first and the third term from (32) is similar while the product of the second and the third one has coefficient 2 instead of -4 but the rest is nearly the same as (35). Now let us recall that we have assumed that $Ee_1^2=1$ and that from the orthogonality of the matrix L it follows that $l_{tj}^2\leqslant 1$, $t,j=1,\ldots,n$, and hence

$$(36)\quad \left|\sum_{j=1}^{n-1} l_{tj+1}l_{tj}^3\right|\leqslant\sum_{j=1}^{n-1}|l_{tj+1}l_{tj}^3|\leqslant\sum_{j=1}^{n-1}|l_{tj+1}|\cdot|l_{tj}|\leqslant\Big(\sum_{j=1}^{n-1} l_{tj+1}^2\sum_{j=1}^{n} l_{tj}^2\Big)^{1/2}\leqslant 1.$$

Taking into account (32)–(36) (and similar relations valid for other terms) we conclude that

$$E\{\vartheta_t\tilde{\vartheta}_t\}^2\leqslant 16\left\{E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2+E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2+E^2\left\{e_1 E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right\},$$

and hence

$$(37)\quad \sum_{t=n-p+1}^{n}\operatorname{var}(\vartheta_t\tilde{\vartheta}_t)\leqslant 16p\left\{E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2\right.$$
$$\left.+E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2+E^2\left\{e_1 E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right\}.$$

In a similar way we may find that (for $t \neq s$)

$$E\{\vartheta_s \tilde{\vartheta}_s \vartheta_t \tilde{\vartheta}_t\} \leqslant 16\left\{E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2 + E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2 + E^2\left\{e_1 E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right\},$$

and hence again

$$\sum_{t=n-p+1}^{n} \sum_{s=n-p+1}^{n} \operatorname{cov}\{\vartheta_s \tilde{\vartheta}_s, \vartheta_t \tilde{\vartheta}_t) \tag{38}$$

$$\leqslant 16p^2\left\{E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2 + E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2 + E^2\left\{e_1 E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right\}.$$

Finally, we can find similar inequalities for the terms

$$E\{\tilde{\xi}_s^2 \vartheta_t \tilde{\vartheta}_t\},\ E\{\vartheta_s \tilde{\vartheta}_s \tilde{\xi}_t \zeta_t\},\ E\{\vartheta_t \tilde{\vartheta}_t \zeta_s^2\},\ E\{\tilde{\xi}_s^2 \tilde{\xi}_t \zeta_t\},\ E\{\tilde{\xi}_s^2 \zeta_t^2\}, \tag{39}$$

and inequalities for the sums of terms from (39) which would be analogical to (37) (or (38)). They would have however instead of coefficient 16 the coefficients

$$\sum_{t=p}^{n} \lambda_t \quad \text{(the first term)}, \tag{40}$$

$$4\sqrt{p} \sum_{t=n-p+1}^{n} \lambda_t \quad \text{(the second and the third ones)}, \tag{41}$$

$$2\Big(\sum_{t=p}^{n} \lambda_t^2\Big)^{1/2}\Big(\sum_{t=n-p+1}^{n} \lambda_t^2\Big)^{1/2} \quad \text{(the forth and the fifth ones)}. \tag{42}$$

Now taking into account (25), (26), (28)–(30), (37), (38) and (40)–(42) we obtain the term

$$\begin{aligned} &\sum_{t=n-p+1}^{n} \lambda_t^2 \left\{1.5E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^4 + 2E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2 \right. \\ &\left. -2.5E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2\right\} \\ &+\Big\{16p(1+p)+\sum_{t=p}^{n}\lambda_t+8\sqrt{p}\sum_{t=n-p+1}^{n}\lambda_t+8\Big(\sum_{t=p}^{n}\lambda_t^2\Big)^{1/2}\Big(\sum_{t=n-p+1}^{n}\lambda_t^2\Big)^{1/2}\Big\} \\ &\times\left\{E\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}^2 + E\left[e_1-\frac{\psi(e_1)}{\gamma}\right]^2 + E^2\left\{e_1\left[e_1-\frac{\psi(e_1)}{\gamma}\right]\right\}\right\}, \end{aligned} \tag{43}$$

which, together with the Durbin–Watson expression for $\operatorname{var}\left(\sum_{t=1}^{n-p} v_t \tilde{\xi}_t^2\right)$ (see (44) below), gives an upper bound for the variance of the numerator in (90). In a similar way we may find a lower bound for the numerator. Of course, to find similar bounds for the denominator is much easier. The results of these steps for the special case of Huber ψ-function will be given later in Table 2. Earlier, however, we give mean values of variables which enter the previous inequalities.

Table 1. Mean values of the random variables which are relevant for the estimate of bounds of the mean value and of variance for Durbin–Watson statistic (since $E|e_1[e_1 - \psi(e_1)/\gamma]| < 10^{-8}$ for all values of tuning constant, we have omitted it in the table),

$\mathscr{L}(e_1) = N(0, 1)$, ψ – Huber function, for γ see Assumptions $\mathscr{A}$.

Tuning constant	1.0	1.1	1.2	1.3	1.4	1.6
$E\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^2$	0.1073	0.0880	0.0718	0.05805	0.0466	0.0371
$E\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]^4$	0.0785	0.0664	0.0559	0.0467	0.0386	0.0315
$E\left\{e_1\left[e_1 - \frac{\psi(e_1)}{\gamma}\right]\right\}^2$	0.3779	0.3429	0.3077	0.2729	0.2392	0.2072

The considerations presented above lead to the conjecture that the behavior of Durbin–Watson statistics for $r(\hat{\beta}^{(M,n)})$ may be similar to that one for $\hat{\beta}^{(LS,n)}$. The values gathered in Table 1 give a hope that especially the mean values of the "upper" and the "lower" statistics z_L and z_U (see (8)) for $r(\hat{\beta}^{(M,n)})$ and $r(\hat{\beta}^{(LS,n)})$ will be nearly the same. The differences of variances (of the statistics evaluated for the least squares and for an M-estimator) can be expected to be somewhat larger. Table 2 shows lower and upper bounds for $Ez(\hat{\beta}^{(M,n)})$ and $\operatorname{var} z(\hat{\beta}^{(M,n)})$ obtained from (90) and evaluated so that we have assumed approximate independence of $z(\hat{\beta}^{(M,n)})$ and its denominator and evaluated moments of $z(\hat{\beta}^{(M,n)})$ in the same way as Durbin–Watson did that. In fact, for finding bounds of the mean value of $z(\hat{\beta}^{(M,n)})$ we have used the relation (6) from Durbin and Watson (1952), which reads

$$Ez(\hat{\beta}^{(LS,n)}) = \frac{1}{n-p} \sum_{t=1}^{n-p} v_t = \tilde{v} \text{ (say)},$$

together with the Durbin–Watson lemma (see (5) and (22)) and then we have carried out the corrections implied by (90) for $Ez(\hat{\beta}^{(M,n)})$. Similarly, when looking for the bounds of $\operatorname{var}\{z(\hat{\beta}^{(M,n)})\}$ we have employed (7) from Durbin and Watson (1952), i.e.

$$\operatorname{var} z(\hat{\beta}^{(LS,n)}) = \frac{2\sum_{t=1}^{n-p}(v_t - \bar{v})^2}{(n-p)(n-p+2)}, \tag{44}$$

(5), (22) and again "corrections" which follow from (90) and which were for the upper bound given (for the numerator) in (43). (Of course, Table 2 is given assuming the Huber ψ-function.)

Table 2. Lower and upper bounds of the mean value and of variance for Durbin–Watson statistics $z(\hat{\beta}^{(M,n)})$, i.e. for statistics evaluated for residuals corresponding to $\hat{\beta}^{(M,n)}$, $\mathscr{L}(e_1) = N(0, 1)$, ψ — Huber function with tuning constant c.

The rows of this table which start with the value of tuning constant c correspond to D–W statistic evaluated for the Huber M-estimator with this tuning constant. At the last row of every table the values of bounds of D–W statistic for the least squares are given for comparison with other rows.

$n = 20$

		$p = 1$		$p = 3$		$p = 5$	
		mean	variance	mean	variance	mean	variance
$c = 1.0$	lower	1.88	0.288	1.61	0.381	1.33	0.534
	upper	2.10	0.486	2.36	0.849	2.64	1.333
$c = 1.1$	lower	1.88	0.278	1.62	0.360	1.34	0.495
	upper	2.10	0.455	2.36	0.784	2.63	1.223
$c = 1.2$	lower	1.88	0.269	1.62	0.341	1.35	0.459
	upper	2.10	0.425	2.36	0.722	2.63	1.117
$c = 1.3$	lower	1.88	0.260	1.63	0.322	1.35	0.424
	upper	2.10	0.397	2.36	0.663	2.63	1.017
$c = 1.4$	lower	1.89	0.252	1.63	0.305	1.36	0.392
	upper	2.10	0.370	2.36	0.607	2.63	0.923
$c = 1.5$	lower	1.89	0.244	1.64	0.289	1.36	0.362
	upper	2.11	0.346	2.36	0.556	2.63	0.835
$z(\hat{\beta}^{(LS,n)})$	lower	1.89	0.157	1.65	0.101	1.38	0.048
	upper	2.11	0.200	2.35	0.249	2.62	0.313

$n = 40$

		$p=1$ mean	$p=1$ variance	$p=3$ mean	$p=3$ variance	$p=5$ mean	$p=5$ variance
$c = 1.0$	lower	1.94	0.132	1.82	0.149	1.69	0.171
	upper	2.05	0.201	2.17	0.285	2.30	0.376
$c = 1.1$	lower	1.94	0.129	1.82	0.144	1.69	0.163
	upper	2.05	0.190	2.17	0.266	2.30	0.349
$c = 1.2$	lower	1.94	0.125	1.82	0.138	1.70	0.155
	upper	2.05	0.179	2.17	0.248	2.29	0.323
$c = 1.3$	lower	1.94	0.122	1.83	0.133	1.70	0.148
	upper	2.05	0.169	2.17	0.231	2.29	0.298
$c = 1.4$	lower	1.95	0.119	1.83	0.129	1.70	0.141
	upper	2.05	0.160	2.17	0.215	2.29	0.274
$c = 1.5$	lower	1.95	0.116	1.83	0.124	1.71	0.135
	upper	2.05	0.151	2.17	0.200	2.29	0.253
$z(\hat{\beta}^{(LS,n)})$	lower	1.95	0.090	1.84	0.077	1.72	0.063
	upper	2.05	0.100	2.16	0.111	2.28	0.124

$n = 60$

		$p=1$ mean	$p=1$ variance	$p=3$ mean	$p=3$ variance	$p=5$ mean	$p=5$ variance
$c = 1.0$	lower	1.96	0.086	1.88	0.093	1.80	0.101
	upper	2.03	0.123	2.11	0.161	2.19	0.199
$c = 1.1$	lower	1.96	0.084	1.88	0.090	1.80	0.097
	upper	2.03	0.117	2.11	0.151	2.19	0.185
$c = 1.2$	lower	1.96	0.082	1.89	0.087	1.81	0.093
	upper	2.03	0.111	2.11	0.142	2.19	0.173
$c = 1.3$	lower	1.96	0.080	1.89	0.084	1.81	0.090
	upper	2.03	0.105	2.11	0.133	2.19	0.161
$c = 1.4$	lower	1.96	0.078	1.89	0.082	1.81	0.087
	upper	2.03	0.100	2.11	0.124	2.19	0.149
$c = 1.5$	lower	1.96	0.076	1.89	0.080	1.81	0.084
	upper	2.03	0.095	2.11	0.117	2.19	0.139
$z(\hat{\beta}^{(LS,n)})$	lower	1.97	0.062	1.89	0.057	1.82	0.051
	upper	2.03	0.067	2.11	0.071	2.18	0.077

USING SCALE- AND REGRESSION-EQUIVARIANT M-ESTIMATORS

As we have already mentioned the M-estimators defined by (9) or (10) are not scale- and regression-equivariant. To achieve it one can employ (12) with scale-equivariant and regression-invariant estimate of scale $\hat{\sigma}^2_{(n)}$, and consequently we may write the asymptotic expansion for $\hat{\beta}^{(M,n)}$ in the form

$$(45)\quad \hat{\beta}^{(M,n)} - \beta^0$$
$$= \gamma^{-1} Q_n^{-1} \left\{ \frac{1}{n} \sum_{t=1}^{n} X_t \psi\left(\frac{e_t}{\sigma_0}\right) - W\left(\log(\hat{\sigma}_{(n)}) - \log(\sigma_{e_1})\right) \frac{1}{n} \sum_{t=1}^{n} X_t \right\} + O_p(n^{-1})$$
$$\text{as } n \to \infty.$$

It means two things. At first the effect of the first term in the parentheses (in (45)) on the value of residuals is of order $1/\sqrt{n}$, since $(1/\sqrt{n}) \sum_{t=1}^{n} X_t \psi(e_t/\sigma_0)$ is bounded in probability as follows from CLT. Similarly, the effect of the second term is also of order $1/\sqrt{n}$ because $\sqrt{n}\left(\log(\hat{\sigma}_{(n)}) - \log(\sigma_{e_1})\right) = O_p(1)$. This means that the effect of the both terms (in the parentheses of (45)) on the value of residuals is of the same order $1/\sqrt{n}$. At the first glance, it may seem strange but it is necessary to realize that we have

$$r(\hat{\beta}^{(M,n)}) = Y - X\beta^0 - \gamma^{-1} X Q_n^{-1} \frac{1}{n} \sum_{t=1}^{n} X_t \left\{ \psi\left(\frac{e_t}{\sigma_0}\right) - W\left(\log(\hat{\sigma}_{(n)}) - \log(\sigma_{e_1})\right) \right\}$$
$$+ O_p(n^{-1})$$
$$= e - \gamma^{-1} X Q_n^{-1} \frac{1}{n} \sum_{t=1}^{n} X_t \left\{ \psi\left(\frac{e_t}{\sigma_0}\right) - W\left(\log(\hat{\sigma}_{(n)}) - \log(\sigma_{e_1})\right) \right\} + O_p(n^{-1}),$$

which indicates that $r(\hat{\beta}^{(M,n)})$ is equal to e plus some additional terms of order $1/\sqrt{n}$ etc.

Secondly, repeating the arguments of Remark 7, we should stress that due to the shape of the matrix Q^{-1} (and the fact that $Q_n^{-1} \to Q^{-1}$), the second term, i.e.

$$\gamma^{-1} X Q_n^{-1} \frac{1}{n} \sum_{t=1}^{n} X_t W\left(\log(\hat{\sigma}_{(n)}) - \log(\sigma_{e_1})\right),$$

has a considerable effect only on the intercept of $\hat{\beta}^{(M,n)}$. So, if we restrict ourselves to the Durbin–Watson statistic given by (7), it is clear that the modification of the asymptotic representation of $\hat{\beta}^{(M,n)}$, i.e. when we employ (45) instead of (74), does not affect the numerator of (7) because the effect of the modification in (45), namely the term containing the estimate of scale, on the residuals is the same for all of them, and hence it disappears. The effect of the modification in (45) on the denominator of (7) does not disappear, however, as we have shown above it is hopefully rather small.

CONCLUSION

The present paper shows that under the assumption of normality of random fluctuations in the regression model (1), Durbin–Watson statistic evaluated on the base of residuals from robust regression analysis (carried out by means of an M-estimator) has asymptotically the same critical values as Durbin–Watson statistic for the analysis based on the least squares. Moreover, it demonstrates that in the case where the Huber ψ-function is employed, the differences in the moments of $z_L(\hat{\beta}^{(LS,n)})$ and $z_L(\hat{\beta}^{(M,n)})$, and of $z_U(\hat{\beta}^{(LS,n)})$ and $z_U(\hat{\beta}^{(M,n)})$ (and hence also in their critical value) are even for moderate sample sizes as $n = 20$, 40 and 60 rather small (in the case of the mean, which plays of course a crucial role for the magnitude of the lower and the upper critical value, the differences are negligible).

One may object that we usually employ M-estimators just in the case when there is a suspicion that data are contaminated, and hence not normally distributed. As we have already discussed, in such a case we should weight down the influence of suspicious observations so that the random fluctuations are approximately normally distributed (and we should verify it *a posteriori*, i.e. after estimating the model). Consequently, we may hope that our results are still acceptable in the sense of approximately efficient forecast.

It also hints what the next step of the research should be, namely to study the behavior of Durbin–Watson statistic evaluated on the base of residuals obtained in regression analysis performed by the robust method with high breakdown point, e.g. the least trimmed squares $\hat{\beta}^{LTS}$ or the least weighted squares $\hat{\beta}^{LWS}$. Recent results describing the Bahadur representation of $\hat{\beta}^{LTS}$ as well as of $\hat{\beta}^{LWS}$ may allow us to do it (see Víšek (1999a)).

APPENDIX I. PROOFS OF THEOREMS AND LEMMAS

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

$$\exp(-n^{-1/2}\kappa) = 1 - n^{-1/2}\kappa + hn^{-1}\kappa^2, \tag{46}$$

where $h \in [\frac{1}{2}n^{-1/2}\exp(-n^{-1/2}\Theta), \frac{1}{2}n^{-1/2}\exp(n^{-1/2}\Theta)]$ and also

$$[e_t - n^{-1/2}X_t^{\mathrm{T}}\tau]\exp(-n^{-1/2}\kappa)$$
$$= e_t - n^{-1/2}(X_t^{\mathrm{T}}\tau + e_t\kappa) + n^{-1}X_t^{\mathrm{T}}\tau\kappa + he_t n^{-1}\kappa^2 - n^{-3/2}X_t^{\mathrm{T}}\tau h\cdot\kappa^2,$$

and finally

$$X_t\{\psi([e_t - n^{-1/2}X_t^{\mathrm{T}}\tau]\exp(-n^{-1/2}\kappa)) - \psi(e_t)\} \tag{47}$$
$$= -n^{-1/2}X_t(X_t^{\mathrm{T}}\tau + e_t u)\psi'(e_t)$$
$$+ n^{-1}X_t\{X_t^{\mathrm{T}}\tau\kappa + he_t - n^{-1/2}X_t^{\mathrm{T}}\tau\cdot h\cdot\kappa^2\}\psi'(e_t) + X_t R_{nt}(\tau, \kappa),$$

where the remainder term can be written in the form

$$R_{nt}(\tau, \kappa) = \{-n^{-1/2}(X_t^{\mathrm{T}}\tau + e_t\kappa) + n^{-1}\{X_t^{\mathrm{T}}\tau\kappa + he_t - n^{-1/2}X_t^{\mathrm{T}}\tau\cdot h\cdot\kappa^2\}\}$$
$$\times[\psi'(\xi_t^{(n)}) - \psi'(e_t)]$$

for some $\xi_t^{(n)}$ for which we have $|\xi_t^{(n)} - e_t| \leqslant n^{-1/2}|X_t^{\mathrm{T}}\tau + e_t\kappa|$. Now, for any fixed $j, k \in \{1, 2, \ldots, p\}$ the sequences

$$\{X_{tj}X_{tk}\psi'(e_t) - E[X_{tj}X_{tk}\psi'(e_t)]\}_{t=1}^{\infty} \quad \text{and} \quad \{X_{tj}[e_t\cdot\psi'(e_t) - E[e_t\cdot\psi'(e_t)]]\}_{t=1}^{\infty}$$

are sequences of independent identically distributed random variables with zero mean and finite (positive) variances. Hence the Lindeberg–Lévy form of the central limit theorem allows us to find an $n_0 \in N$ such that for any $\varepsilon > 0$ we may find $C_\varepsilon < \infty$ so that for any $n > n_0$

$$(48)\qquad P_F\{\max_{1\leqslant j,k\leqslant p}|n^{-1/2}\sum_{t=1}^{n}\{X_{tj}X_{tk}\psi'(e_t) - E[X_{tj}X_{tk}\psi'(e_t)]\}| > C_\varepsilon\} < \varepsilon$$

as well as

$$(49)\qquad P_F\{\max_{1\leqslant j\leqslant p}|n^{-1/2}\sum_{t=1}^{n}\{X_{tj}[e_t\cdot\psi'(e_t) - E[e_t\cdot\psi'(e_t)]]\}| > C_\varepsilon\} < \varepsilon.$$

Taking once again into account that for any $j = 1, 2, \ldots, p$

$$\{X_{tj}X_{tk}\psi'(e_t)\}_{t=1}^{\infty} \quad \text{and} \quad \{X_{tj}e_t\psi'(e_t)\}_{t=1}^{\infty}$$

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

$$(50)\qquad \begin{aligned} n^{-1}\max_{1\leqslant j,k\leqslant p}|\sum_{t=1}^{n}X_{tj}X_{tk}\psi'(e_t)| &= \mathcal{O}_p(1),\\ n^{-1}\max_{1\leqslant j\leqslant p}|\sum_{t=1}^{n}X_{tj}e_t\psi'(e_t)| &= \mathcal{O}_p(1)\end{aligned}$$

and

$$(51)\qquad n^{-3/2}\max_{1\leqslant j,k\leqslant p}|\sum_{t=1}^{n}X_{tj}X_{tk}\psi'(e_t)| = o_p(1).$$

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

$$\psi'(\xi_t^{(n)}) - \psi'(e_t) = \int_{e_t}^{\xi_t^{(n)}}\psi''(z)\,dz,$$

and hence for any n we have

$$|\psi'(\xi_t^{(n)}) - \psi'(e_t)| \leqslant n^{-1/2}|X_t^{\mathrm{T}}\tau|\,\psi''_{\sup}.$$

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

$$|X_t^{\mathrm{T}}\tau| \leqslant p^{-1/2}\,\|X_t\|\,\Theta,$$

and hence there is a constant C_1 such that

$$\sup_{\mathscr{T}_\Theta}\|X_t R_{nt}(\tau, \kappa)\| \leqslant n^{-1}\,\Theta^2\,C_1\,\psi''_{\sup}\,\|X_t\|^2\,\{\|X_t\|+|e_t|\}.$$

Now, it is again sufficient to take into account that $\{\|X_t\|^3\}_{t=1}^{\infty}$ and $\{\|X_t\|^2\,|e_t|\}_{t=1}^{\infty}$ are sequences of independent identically distributed random variables with finite mean values, and to apply Kolmogorov's law of large numbers onee again to find that

$$\sup_{\mathscr{T}_\Theta}\Big\|\sum_{t=1}^{n} X_t R_{nt}(\tau, \kappa)\Big\| = \mathcal{O}_p(1). \tag{52}$$

Now the theorem follows from (48)–(52). ∎

Proof of Theorem 2. Not loosing generality let us again assume that $\sigma_{e_1} = 1$ and notice that due to the character of the function $\psi(z)$ the assumptions (14) are fulfilled and the second derivative (where it exists) is equal to zero. Finally, let us put

$$g_t = X_t\,\{\psi\,([e_t - n^{-1/2}\,X_t^{\mathrm{T}}\tau]\,\sigma_{e_1}^{-1}\exp(-n^{-1/2}\,\kappa)) - \psi\,(e_t\,\sigma_{e_1}^{-1})\}$$

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 (47) in the case when there is an $s_0 \in \{1, 2, \ldots, k\}$ such that one of the following conditions holds:

$$(e_t - n^{-1/2}\,X_t^{\mathrm{T}}\tau)\exp(-n^{-1/2}\,\kappa) \leqslant r_{s_0} \leqslant e_t \qquad \text{or} \qquad e_t \leqslant r_{s_0} \leqslant (e_t - n^{-1/2}\,X_t^{\mathrm{T}}\tau)\exp(-n^{-1/2}\,\kappa). \tag{53}$$

So the idea of proof is to withdraw from $S_n(\tau, \kappa)$ the sum of all g_t's for which (53) takes place, then to show that the sum of the terms which were withdrawn from $S_n(\tau, \kappa)$ is small in probability, and finally to add to the "reduced" sum $S_n(\tau, \kappa)$ for all indices t which were in the previous step withdrawn appropriate terms (the sum of them will be shown to be also negligible in probability) to reach the assertion of the theorem.

In order to fulfill the just sketched plan, let us denote the event given by (53) by $B_n(t, \tau, \kappa)$ and its indicator by $I_{B_n(t,\tau,\kappa)}$. Since the left and the right inequality in (53) are successively equivalent to

$$r_{s_0} \leqslant e_t \leqslant r_{s_0}\exp(n^{-1/2}\kappa) + n^{-1/2}\,X_t^{\mathrm{T}}\tau \qquad \text{and} \qquad r_{s_0}\exp(-n^{-1/2}\,\kappa) + n^{-1/2}\,X_t^{\mathrm{T}}\tau \leqslant e_t \leqslant r_{s_0}$$

and due to the assumption on the upper bound of the density of random fluctuations, there is a constant C_1 such that the conditional probability of

$B_n(t, \tau, \kappa)$ for given X_t is bounded by $C_1 \{n^{-1/2} |X_t^T \tau| + r |\exp(n^{-1/2} \kappa) - 1|\}$ (starting from some n_0). Now

$$|\exp(n^{-1/2} \kappa) - 1| \leqslant \left|1 + n^{-1/2} \Theta + \frac{n^{-1} \Theta^2}{2!} + \frac{n^{-3/2} \Theta^3}{3!} + \ldots - 1\right|$$
$$\leqslant n^{-1/2} \Theta \left(1 + n^{-1/2} \Theta + \frac{n^{-1} \Theta^2}{2!} + \frac{n^{-3/2} \Theta^3}{3!} + \ldots\right)$$
$$= n^{-1/2} \Theta \exp(n^{-1/2} \Theta).$$

Then there is a constant C_2 such that for $\tau, \kappa \in \mathcal{T}_\Theta$ the corresponding probability is bounded by $n^{-1/2} C_2 \{\|X_t\| + 1\}$. Similar considerations lead to the existence of a constant C_3 such that

$$|\psi([e_t - n^{-1/2} X_t^T \tau] \exp(-n^{-1/2} \kappa)) - \psi(e_t)| \leqslant n^{-1/2} C_3 \{\|X_t\| + |e_t|\}.$$

Then

$$E \|g_t I_{B_n(t,\tau,\kappa)}\| \leqslant E \{\|X_t\| |\psi([e_t - n^{-1/2} X_t^T \tau] \exp(-n^{-1/2} \kappa)) - \psi(e_t)| I_{B_n(t,\tau,\kappa)}\}$$
$$\leqslant n^{-1/2} C_4 E \{\|X_t\| [\|X_t\| + |e_t|] I_{B_n(t,\tau,\kappa)}\} \leqslant n^{-1} C_5 E \{\|X_t\| (\|X_t\| + 1)(\|X_t\| + |e_t|)\}$$

for appropriate constants C_4 and C_5. Now we have

$$S_n(\tau, \kappa) = \sum_{t=1}^{n} g_t I_{B_n(t,\tau,\kappa)} + \sum_{t=1}^{n} g_t (1 - I_{B_n(t,\tau,\kappa)})$$

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

$$P\left(\left\|\sum_{t=1}^{n} g_t I_{B_n(t,\tau,\kappa)}\right\| > C_6\right) \leqslant C_6^{-1} C_5 E \{\|X_1\| (\|X_1\| + 1)(\|X_1\| + |e_1|)\}. \tag{54}$$

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

$$\sum_{t=1}^{n} n^{-1/2} X_t (X_t^T \tau + e_t \kappa) \psi'(e_t)(1 - I_{B_n(t,\tau,\kappa)}) \tag{55}$$

the sum

$$\sum_{t=1}^{n} n^{-1/2} X_t (X_t^T \tau + e_t \kappa) \psi'(e_t) I_{B_n(t,\tau,\kappa)}$$

to obtain the expression given in (16), namely $n^{1/2} Q_n \tau + n^{-1/2} W \sum_{t=1}^{n} X_t \kappa$. However, along similar lines to those we arrived at (54) we may find that

$$\sup_{\mathcal{T}_\Theta} \left\|\sum_{t=1}^{n} n^{-1/2} X_t (X_t^T \tau + e_t \kappa) \psi'(e_t) I_{B_n(t,\tau,\kappa)}\right\| = \mathcal{O}_p(1)$$

so that we can add this to (55). That completes the proof. ∎

Proof of Theorem 3. Without loss of generality let us assume that $m = 1$ (write r instead of r_1) and $\alpha_0 < \alpha_1$, and put $\delta = \alpha_1 - \alpha_0$. According to the assumptions, there is $C_1 < \infty$ such that $f_{e_1}(z) < C_1$. Let us write

$$\xi_t(n, \tau, \kappa) = \psi\left([e_t - n^{-1/2} X_t^{\mathrm{T}} \tau]\, \sigma_{e_1}^{-1} \exp(-n^{-1/2}\kappa)\right) - \psi(e_t \sigma_{e_1}^{-1})$$

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

$$e_t < r < [e_t - n^{-1/2} X_t^{\mathrm{T}} \tau] \exp(-n^{-1/2}\kappa) \leftrightarrow n^{-1/2} X_t^{\mathrm{T}} \tau + r\exp(-n^{-1/2}\kappa) < e_t < r \tag{56}$$

or

$$[e_t - n^{-1/2} X_t^{\mathrm{T}} \tau] \exp(n^{-1/2}\kappa) < r < e_t \leftrightarrow r < e_t < n^{-1/2} X_t^{\mathrm{T}} \tau + r\exp(n^{-1/2}\kappa). \tag{57}$$

Denote the events described in (56) and (57) successively by $B_t^{(k)}(n, \tau, \kappa)$, $k = 1, 2$. First of all, observe that (56) can take place when

$$n^{-1/2} X_t^{\mathrm{T}} \tau + r\exp(n^{-1/2}\kappa) < r \tag{58}$$

and similarly (57) can hold if

$$n^{-1/2} X_t^{\mathrm{T}} \tau + r\exp(n^{-1/2}\kappa) > r. \tag{59}$$

Fix an $l \in \{1, 2, \ldots, p\}$ and denote successively by $D_t^{(j)}(n, \tau, \kappa, l)$, $j = 1, 2, 3, 4$, the events

$$\{\omega \in \Omega\colon \{n^{-1/2} X_t^{\mathrm{T}} \tau + r\exp(n^{-1/2}\kappa) < r\} \cap \{X_{tl} \leqslant 0\}\},$$
$$\{\omega \in \Omega\colon \{n^{-1/2} X_t^{\mathrm{T}} \tau + r\exp(n^{-1/2}\kappa) < r\} \cap \{X_{tl} > 0\}\},$$
$$\{\omega \in \Omega\colon \{n^{-1/2} X_t^{\mathrm{T}} \tau + r\exp(n^{-1/2}\kappa) \geqslant r\} \cap \{X_{tl} \leqslant 0\}\},$$
$$\{\omega \in \Omega\colon \{n^{-1/2} X_t^{\mathrm{T}} \tau + r\exp(n^{-1/2}\kappa) \geqslant r\} \cap \{X_{tl} > 0\}\}.$$

Further, denote by $\pi_t^{(j,k)}(n, \tau, \kappa, l)$, $j = 1, 2, 3, 4$, $k = 1, 2$, the probabilities of the events $B_t^{(k)}(n, \tau, \kappa) \cap D_t^{(j)}(n, \tau, \kappa, l)$. For any $n \in N$, $j = 1, 2, 3, 4$ and $k = 1, 2$ we have

$$\begin{aligned}
\pi_t^{(j,k)}(n, \tau, \kappa, l) &= E\, I_{B_t^{(k)}(n,\tau,\kappa) \cap D_t^{(j)}(n,\tau,\kappa,l)} \\
&\leqslant \int_{-\infty}^{\infty} \Big\{ \int_{n^{-1/2} X_t^{\mathrm{T}}\tau + r\exp(n^{-1/2}\kappa)}^{r} f_{e_t}(z)\, dz \Big\} f_{X_t}(x)\, dx \\
&\leqslant C_1 \int_{-\infty}^{\infty} \Big\{ \int_{n^{-1/2} X_t^{\mathrm{T}}\tau + r\exp(n^{-1/2}\kappa)}^{r} dz \Big\} f_{X_t}(x)\, dx \\
&\leqslant n^{-1/2} \int_{-\infty}^{\infty} [X_t^{\mathrm{T}} \tau + r\kappa - h n^{-1/2} \kappa^2 r]\, f_{X_t}(x)\, dx,
\end{aligned} \tag{60}$$

and hence there is a constant C_2 such that

(61) $$\pi_t^{(j,k)}(n, \tau, \kappa, l) < n^{-1/2} C_2 \{E \|X_t\| + 1\}.$$

Of course, the lower and upper bounds of the integral in (60) should be interchanged if

$$r < n^{-1/2} [X_t^{\mathrm{T}} \tau + r \exp(n^{-1/2} \kappa)]$$

but (61) holds for any combination of j and k. Now, we shall study the sum

$$S_{nl}(\tau, \kappa) = \sum_{t=1}^{n} X_{tl} [\xi_t(n, \tau, \kappa) - E\,\xi_t(n, \tau, \kappa)].$$

Since $\bigcup_{j=1}^{4} D_t^{(j)}(n, \tau, \kappa, l) = \Omega$ a.s., we have

$$\xi_t(n, \tau, \kappa) = \sum_{j=1}^{4} [\xi_t(n, \tau, \kappa) I_{D_t^{(j)}(n,\tau,\kappa,l)}] \text{ a.s.,}$$

and hence

(62) $$E\xi_t(n, \tau, \kappa) = \sum_{j=1}^{4} E[\xi_t(n, \tau, \kappa) I_{D_t^{(j)}(n,\tau,\kappa,l)}].$$

Then

(63) $$\sum_{t=1}^{n} X_{tl} [\xi_t(n, \tau, \kappa) - E\xi_t(n, \tau, \kappa)]$$
$$= \sum_{t=1}^{n} \sum_{j=1}^{4} X_{tl} \left[\xi_t(n, \tau, \kappa) I_{D_t^{(j)}(n,\tau,\kappa,l)} - E[\xi_t(n, \tau, \kappa) I_{D_t^{(j)}(n,\tau,\kappa,l)}]\right].$$

Now consider $X_{tl} \left[\xi_t(n, \tau, \kappa) I_{D_t^{(1)}(n,\tau,\kappa,l)} - E[\xi_t(n, \tau, \kappa) I_{D_t^{(1)}(n,\tau,\kappa,l)}]\right]$. We easily find that (for δ see the second line of the proof of Theorem 3)

(64) $$X_{tl} \left[\xi_t(n, \tau, \kappa) I_{D_t^{(1)}(n,\tau,\kappa,l)} - E[\xi_t(n, \tau, \kappa) I_{D_t^{(1)}(n,\tau,\kappa,l)}]\right]$$
$$= \delta X_{tl} \left(1 - \pi_t^{(1,1)}(n, \tau, \kappa, l)\right) = -\delta |X_{tl}| \left(1 - \pi_t^{(1,1)}(n, \tau, \kappa, l)\right) > -\delta |X_{tl}|$$
$$\text{with probability } \pi_t^{(1,1)}(n, \tau, \kappa, l)$$

and

(65) $$X_{tl} \left[\xi_t(n, \tau, \kappa) I_{D_t^{(1)}(n,\tau,\kappa,l)} - E[\xi_t(n, \tau, \kappa) I_{D_t^{(1)}(n,\tau,\kappa,l)}]\right]$$
$$= \delta X_{tl} \pi_t^{(1,1)}(n, \tau, \kappa, l) = \delta |X_{tl}| \pi_t^{(1,1)}(n, \tau, \kappa, l) < \delta n^{-1/2} C_2 |X_{tl}| \cdot [E \|X_t\| + 1]$$
$$\text{with probability } 1 - \pi_t^{(1,1)}(n, \tau, \kappa, l).$$

Taking into account the expressions after the first sign of equality in (64) and in (65), and the corresponding probabilities, we immediately find that

$$E\left\{X_{tl} \left[\xi_t(n, \tau, \kappa) I_{D_t^{(1)}(n,\tau,\kappa,l)} - E[\xi_t(n, \tau, \kappa) I_{D_t^{(1)}(n,\tau,\kappa,l)}]\right]\right\} = 0.$$

(Notice that the last equality cannot be found as a consequence of taking the conditional mean value of the expression given in the square brackets). So, putting for any $n \in N$ and $t = 1, 2, \ldots, n$

$$a_{tl}(n, \tau, \kappa) = \delta |X_{tl}| \big(1 - \pi_t^{(1,1)}(n, \tau, \kappa, l)\big) \quad \text{and} \quad b_{tl}(n, \tau, \kappa) = \delta |X_{tl}| \, \pi_t^{(1,1)}(n, \tau, \kappa, l),$$

we may use Lemma A.2 (see Appendix II) and define $\mu_{tl}^{(1)}(n, \tau, \kappa)$, the time for Wiener process to exit the interval $\big(-a_{tl}(n, \tau, \kappa), b_{tl}(n, \tau, \kappa)\big)$. Then we obtain

$$X_{tl}\big[\xi_t(n, \tau, \kappa) I_{D_t^{(1)}(n,\tau,\kappa,l)} - E\,[\xi_t(n, \tau, \kappa) I_{D_t^{(1)}(n,\tau,\kappa,l)}]\big] \overset{\mathscr{D}}{=} \mathscr{W}\big(\mu_{tl}^{(1)}(n, \tau, \kappa)\big),$$

where $\overset{\mathscr{D}}{=}$ means equality in distribution. Similarly, for $j = 2, 3$ and 4 we find

$$X_{tl}\big[\xi_t(n, \tau, \kappa) I_{D_t^{(j)}(n,\tau,\kappa,l)} - E\,[\xi_t(n, \tau, \kappa) I_{D_t^{(j)}(n,\tau,\kappa,l)}]\big] \overset{\mathscr{D}}{=} \mathscr{W}\big(\mu_{tl}^{(j)}(n, \tau, \kappa)\big).$$

Finally, putting $\mu_{tl}(n, \tau, \kappa) = \sum_{j=1}^{4} \mu_{tl}^{(j)}(n, \tau, \kappa)$ and taking into account (63), we obtain

$$\begin{aligned} n^{-1/4}[S_{nl}(\tau, \kappa) - E\,S_{nl}(\tau, \kappa)] &= n^{-1/4} \sum_{t=1}^{n} X_{tl}[\xi_t(n, \tau, \kappa) - E\,\xi_t(n, \tau, \kappa)] \\ &\overset{\mathscr{D}}{=} n^{-1/4} \sum_{t=1}^{n} \sum_{j=1}^{4} \mathscr{W}\big(\mu_{tl}^{(j)}(n, \tau, \kappa)\big) \overset{\mathscr{D}}{=} n^{-1/4} \sum_{t=1}^{n} \mathscr{W}\big(\mu_{tl}(n, \tau, \kappa)\big) \\ &\overset{\mathscr{D}}{=} \mathscr{W}\big(n^{-1/2} \sum_{t=1}^{n} \mu_{tl}(n, \tau, \kappa)\big). \end{aligned}$$

Now, let us take into account inequalities which are given in (61), (64) and (65), and put $c_{tl} = \delta |X_{tl}|$ and $d_{tl} = n^{-1/2} \delta C_2 |X_{tl}| \cdot [E\,\|X_t\| + 1]$. Defining

(66) $\quad \tilde{\mu}_{tl}(n, \Theta)$ – the time for Wiener process to exit the interval $(-c_{tl}, d_{tl})$,

we obtain

$$\mu_{tl}(n, \tau, \kappa) \leqslant \tilde{\mu}_{tl}(n, \Theta),$$

and therefore

$$\text{(67)} \quad \begin{aligned} \sup_{\mathscr{T}_\Theta} n^{-1/4} |S_{nl}(\tau, \kappa) - E\,S_{nl}(\tau, \kappa)| &\overset{\mathscr{D}}{=} \sup_{\mathscr{T}_\Theta} \big|\mathscr{W}\big(n^{-1/2} \sum_{t=1}^{n} \mu_{tl}(n, \tau, \kappa)\big)\big| \\ &\leqslant \sup \big\{|\mathscr{W}(s)|:\ 0 \leqslant s \leqslant n^{-1/2} \sum_{t=1}^{n} \tilde{\mu}_{tl}(n, \Theta)\big\}. \end{aligned}$$

Moreover (see again Lemma A.2), we have from (66) for any $\tau, \kappa \in \mathscr{T}_\Theta$

$$E\tilde{\mu}_{tl}(n, \Theta) \leqslant 4n^{-1/2} C_2\, EX_{tl}^2 \cdot [E\,\|X_t\| + 1] \quad \text{for all } n \in N,$$

i.e.

$$n^{-1/2} \sum_{t=1}^{n} E\tilde{\mu}_{tl}(n, \Theta) \leqslant C_2\, EX_{1l}^2 \cdot [E\,\|X_1\| + 1].$$

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

$$P\{n^{-1/2} \sum_{t=1}^{n} \tilde{\mu}_{tl}(n, \Theta) > C_3\} < \varepsilon/2, \tag{68}$$

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

$$P\{\sup\{|\mathscr{W}(s)|:\ 0 \leqslant s \leqslant C_3\} > C_4\} < \varepsilon/2, \tag{69}$$

see e.g. Csörgö and Révész (1981). By (67)–(69), we get

$$P\{\sup_{\mathscr{T}_\Theta} n^{-1/4} |S_{nl}(\tau, \kappa) - ES_{nl}(\tau, \kappa)| > C_4\} < \varepsilon,$$

which means that also

$$\sup_{\mathscr{T}_\Theta} n^{-1/4} \|S_n(\tau, \kappa) - ES_n(\tau, \kappa)\|$$

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

$$n^{-1/4} \sup_{\mathscr{T}_\Theta} \|ES_n(\tau, \kappa) + n^{-1/2} \delta E \sum_{t=1}^{n} [X_t(X_t^{\mathrm{T}}\tau + \kappa r) f_{e_t}(r)]\| = \mathcal{O}(1).$$

We have already shown that $E\xi_t(n, \tau, \kappa) = \sum_{j=1}^{4} E[\xi_t(n, \tau, \kappa) I_{D_t^{(j)}(n,\tau,\kappa,l)}]$ (see (62)). On the other hand, we have

$$E\{X_t \xi_t(n, \tau, \kappa) I_{D^{(1)}(n,\tau,\kappa,l)}\} = E\{X_t E[\xi_t(n, \tau, \kappa) I_{D^{(1)}(n,\tau,\kappa,l)} \,|\, X_t]\}$$

$$= E\{X_t \int_{-\infty}^{\infty} \xi_t(n, \tau, \kappa) I_{D_t^{(j)}(n,\tau,\kappa,l)} f_{e_t}(z)\, dz\} = \delta E\{X_t \int_{n^{-1/2}X_t^{\mathrm{T}}\tau + r\exp(n^{-1/2}\kappa)}^{r} f_{e_t}(z)\, dz\}$$

$$= \delta E\{X_t \int_{n^{-1/2}X_t^{\mathrm{T}}\tau + r\exp(n^{-1/2}\kappa)}^{r} [f_{e_t}(r) + f_{e_t}(z) - f_{e_t}(r)]\, dz\}.$$

(Again, the upper and lower bounds of the integral should be interchanged if it is appropriate.) The last expression is equal to

$$-\delta E\{X_t[n^{-1/2} X_t^{\mathrm{T}} \tau + r(\exp(n^{-1/2}\kappa) - 1)] f_{e_t}(r)\} + ER_{tn}^*(\tau, \kappa),$$

where

$$R_{tn}^*(\tau, \kappa) = \delta\{X_t \int_{n^{-1/2}X_t^{\mathrm{T}}\tau + r\exp(n^{-1/2}\kappa)}^{r} [f_{e_t}(z) - f_{e_t}(r)]\, dz\}.$$

Moreover,

$$\exp(n^{-1/2}\kappa) - 1 = n^{-1/2}\kappa + \frac{n^{-1}\kappa^2}{2!} + \frac{n^{-3/2}\kappa^3}{3!} + \dots,$$

and hence we have

$$ES_n(\tau, \kappa)+n^{-1/2}\delta E\sum_{t=1}^{n}[X_t(X_t^{\mathrm{T}}\tau+\kappa r)f_{e_t}(r)]$$

$$=\sum_{t=1}^{n}\left\{ER_{tn\tau}^{*}-n^{-1/2}\delta rEX_t\left[\frac{n^{-1}\kappa^2}{2!}+\frac{n^{-3/2}\kappa^3}{3!}+\dots\right]\right\}.$$

Finally,

$$n^{-1/4}\sup_{\mathscr{T}_\Theta}\left\|ES_n(\tau, \kappa)+n^{-1/2}\delta E\sum_{t=1}^{n}[X_t(X_t^{\mathrm{T}}\tau+\kappa r)f_{e_t}(r)]\right\|$$

$$=n^{-1/4}\sup_{\mathscr{T}_\Theta}\left\|\sum_{t=1}^{n}ER_{tn}^{*}(\tau, \kappa)\right\|+\mathcal{O}(n^{-3/4})\leqslant n^{-1/4}\sup_{\mathscr{T}_\Theta}\sum_{t=1}^{n}\|ER_{tn}^{*}(\tau, \kappa)\|+\mathcal{O}(n^{-3/4}).$$

Recalling that $f_{e_1}(z)$ is Lipschitz, we have for $z\in[n^{-1/2}X_t^{\mathrm{T}}\kappa+r\exp(n^{-1/2}\kappa), r]$

$$|f_{e_1}(z)-f_{e_1}(r)|<n^{-1/2}C_5[\|X_1\|+1],$$

and hence

$$\|ER_{tn}^{*}(\tau, \kappa)\|\leqslant n^{-1/2}\delta C_5\left\|E\left\{\|X_t\|[\|X_t\|+1]\int_{n^{-1/2}X_t^{\mathrm{T}}\tau+r\exp(n^{-1/2}\kappa)}^{r}dz\right\}\right\|$$

$$\leqslant n^{-1}C_6E\{\|X_t\|\cdot[\|X_t\|+1]^2\}$$

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

$$n^{-1/4}\sup_{\mathscr{T}_\Theta}\sum_{t=1}^{n}\|ER_{tn}^{*}(\tau, \kappa)\|=\mathcal{O}(n^{-1/4}),$$

which completes the proof. ∎

Proof of Theorem 4. Firstly, let us realize that the assumption that Q is positive definite implies that the determinant of this matrix is positive (keep also in mind that the matrix Q is of type $p\times p$). Let us realize that determinant of matrix is a continuous function of the elements of the matrix. Secondly, the sequence of matrices $\{Q_n\}_{n=1}^{\infty}$ converges to the matrix Q in probability. Hence we can find for arbitrary positive numbers ε and δ an $n_0\in N$ so that for all $n>n_0$ the probability of the set on which the determinants of matrices Q_n differ from the determinant of matrix Q less than δ is at least $1-\varepsilon$. Assume that δ is smaller than the determinant of Q. Then, for the corresponding n and ω's, Q_n are also positive definite and regular.

Secondly, we are going to use Theorems 1 and 2 with properly selected Θ. First of all, let us fix $\varepsilon_1>0$. Due to the $\sqrt{n}$-consistency of $\hat{\sigma}_{(n)}^2(\omega)$ we may find $n_1\in N$, $n_1>n_0$, so that for all $n>n_1$ there is a set D_n such that $P(D_n^c)<\varepsilon_1$ and for any $\omega\in D_n$ we have $\sqrt{n}|\hat{\sigma}_{(n)}^2-\sigma_{e_1}^2|<L$ for some $L>0$. Then of course for any $\omega \in D_n$ also $\sqrt{n}\,|\hat{\sigma}_{(n)} - \sigma_{e_1}| < L^*$ for some $L^* > 0$. It means that we can write $\hat{\sigma}_{(n)}(\omega) = \sigma_{e_1} \exp(n^{-1/2}\kappa)$ for some $\kappa = \kappa(\omega)$, where $|\kappa(\omega)| < L^*$. Finally, select any $\Theta > L^*$. Now, using (15) and (16) we obtain

$$(70) \qquad n^{-1/2} \sum_{t=1}^{n} X_t \psi([e_t - n^{-1/2} X_t^{\mathrm{T}} \tau]\, \sigma_{e_1}^{-1} \exp(-n^{-1/2}\kappa))$$
$$= n^{-1/2} \sum_{t=1}^{n} X_t \psi(e_t \sigma_{e_1}^{-1}) - \gamma Q_n \tau - W E X_1 \kappa + o_p(1).$$

By Assumptions $\mathscr{B}$ and the assumptions on the functions ψ_a^- and ψ_c it is possible to verify that the assumptions of the Lindeberg–Lévy theorem are fulfilled for the sequence of random variables $\{X_t \psi(e_t \sigma_{e_1}^{-1})\}_{t=1}^{\infty}$ and from the assumption that $E\psi(e_1 \sigma_{e_1}^{-1}) = 0$ and X_t and e_t are mutually independent it follows that

$$n^{-1/2} \sum_{t=1}^{n} X_t \psi(e_t \sigma_{e_1}^{-1})$$

is bounded in probability (of course, independently of τ and κ). It means that for any $\varepsilon_2 > 0$ there is a constant $C_1 > 0$ and $n_2 \in N$, $n_2 > n_1$, such that for any $n > n_2$ we have

$$P(B_n) > 1 - \varepsilon_2 \qquad \text{for } B_n = \Big\{\omega \in \Omega \colon \Big\| n^{-1/2} \sum_{t=1}^{n} X_t \psi(e_t \sigma_{e_1}^{-1}) \Big\| < C_1 \Big\}.$$

Further, for any $\Delta > 0$, let $n_3 > n_2$ be selected so that for all $n > n_3$ there is a set C_n^{Δ} such that for all $\omega \in C_n^{\Delta}$ the term $o_p(1)$ in (70) is smaller than Δ and $P(C_n^{\Delta}) > 1 - \varepsilon_2$. Without loss of generality, let us assume that $\gamma > 0$. Taking into account the linearity in τ of

$$\tau^{\mathrm{T}} n^{-1/2} \sum_{t=1}^{n} X_t \psi(e_t \sigma_{e_1}^{-1}) \qquad \text{and} \qquad \tau^{\mathrm{T}} W E X_1 \kappa,$$

we may find for any $C_2 > 0$ and any fixed $\kappa \in (0, C_2)$ a constant $C_3 > 0$ so that for any $n > n_3$, $\omega \in B_n \cap C_n^{\Delta}$ and any $\tau \in R^p$, $\|\tau\| = C_3$, we have

$$-\tau^{\mathrm{T}} n^{-1/2} \sum_{t=1}^{n} X_t \psi([e_t - n^{-1/2} X_t^{\mathrm{T}} \tau]\, \sigma_{e_1}^{-1} \exp(-n^{-1/2}\kappa))$$
$$= -\tau^{\mathrm{T}} n^{-1/2} \sum_{t=1}^{n} X_t \psi(e_t \sigma_{e_1}^{-1}) + \gamma \tau^{\mathrm{T}} Q_n \tau + \tau^{\mathrm{T}} W E X_1 \kappa + o_p(1) > 0$$

(since we have assumed that the matrix Q is positive definite and $\gamma > 0$). Finally, applying Assertion A.4 (see Appendix II) for any $n > n_2$ and $\omega \in B_n \cap C_n^{\Delta} \cap D_n$ we find $\tau \in R^p$ such that $\|\tau\| \leqslant C_3$, $\tau = \tau(u, \omega)$ and

$$\sum_{t=1}^{n} X_t \psi\left(\frac{Y_t - X_t^{\mathrm{T}} \beta^0 + n^{-1/2} X_t^{\mathrm{T}} \tau}{\hat{\sigma}_{(n)}^2}\right) = 0.$$

Writing $\tau(u, \omega) = \sqrt{n}(\hat{\beta}(u, \omega) - \beta^0)$ and using once again $\sqrt{n}$-consistency of $\hat{\sigma}_{(n)}$, we complete the proof of the promised assertion. ∎

Proof of Corollary 2. Let us recall that by the previous theorem we get

$$\tau = \sqrt{n}(\hat{\beta}^{(M,n)} - \beta^0) = \mathcal{O}_p(1)$$

and by the assumptions of this corollary we also have

$$\kappa = \sqrt{n}(\log \hat{\sigma}_{(n)} - \log \sigma_{e_1}) = \mathcal{O}_p(1).$$

Moreover,

$$\sum_{t=1}^{n} X_t \psi([e_t - n^{-1/2} X_t^{\mathrm{T}} \tau]\, \sigma_{e_1}^{-1} \exp(-n^{-1/2}\kappa)) = \sum_{t=1}^{n} X_t \psi\left(\frac{Y_t - X_t^{\mathrm{T}} \hat{\beta}^{(M,n)}}{\hat{\sigma}_{(n)}}\right) = 0.$$

So using (15) and (16) for $\tau = \sqrt{n}(\hat{\beta}^{(M,n)} - \beta^0)$ and $\kappa = \sqrt{n}(\log \hat{\sigma}_{(n)} - \log \sigma_{e_1})$, we obtain

$$-n^{-1/2} \sum_{t=1}^{n} X_t \psi(e_t \sigma_{e_1}^{-1}) + \sqrt{n}\, \gamma Q_n (\hat{\beta}^{(M,n)} - \beta)$$
$$+ n^{-1/2}\, W \sum_{t=1}^{n} X_t (\log \hat{\sigma}_{(n)} - \log \sigma_{e_1}) = \mathcal{O}_p(n^{-1/2}),$$

which yields (19). ∎

Proof of Corollary 3 directly follows from the previous corollary. ∎

Proof of Theorem 5. Similarly to considerations in the proof of Theorem 3, without loss of generality let us assume that $m = 1$ (i.e. the function ψ has only one discontinuity), $r = 0$, $\alpha_2 = -\alpha_1 = \frac{1}{2}\delta$. Fixing in an appropriate way Θ (see the second paragraph of the proof of Theorem 4), let us consider for $\|\tau\| < \Theta$, $|\kappa| < \Theta$

$$\sum_{t=1}^{n} X_t \{\psi([e_t - n^{-1/2} X_t^{\mathrm{T}} \tau]\, \sigma_{e_1}^{-1} \exp(-n^{-1/2}\kappa))$$
$$- \tilde{\psi}([e_t - n^{-1/2} X_t^{\mathrm{T}} \tau]\, \sigma_{e_1}^{-1} \exp(-n^{-1/2}\kappa))\}.$$

Since $\psi(z) = \tilde{\psi}_n(z)$ for $|z| \geqslant Kn^{-\vartheta}$, the difference

$$\psi([e_t - n^{-1/2} X_t^{\mathrm{T}} \tau]\, \sigma_{e_1}^{-1} \exp(-n^{-1/2}\kappa)) - \tilde{\psi}_n([e_t - n^{-1/2} X_t^{\mathrm{T}} \tau]\, \sigma_{e_1}^{-1} \exp(-n^{-1/2}\kappa))$$

is nonzero only in the case when

$$|[e_t - n^{-1/2} X_t^{\mathrm{T}} \tau]\, \sigma_{e_1}^{-1} \exp(-n^{-1/2}\kappa)| < Kn^{-\vartheta},$$

i.e. when

$$(71)\qquad -Kn^{-\vartheta}\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa)+n^{-1/2}X_t^{\mathrm{T}}\tau < e_t < Kn^{-\vartheta}\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa)+n^{-1/2}X_t^{\mathrm{T}}\tau.$$

According to the assumption of Theorem 3 there is $J > 0$ such that $f_{e_t}(z) < J$. It means that probability of the event (71) is bounded by $2JKn^{-\vartheta}$ (notice that the presence of X_t in the boundaries of the interval given in (71) does not play any role). For the notational simplicity let I_t be the indicator of the set

$$\{\omega:\ \psi([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa)) \neq \tilde{\psi}_n([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa))\}.$$

Then

$$\begin{aligned} E\{X_t\{\psi([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa)) \\ -\tilde{\psi}_n([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa))\}\} \\ = E\{X_t\{\psi([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa)) \\ -\tilde{\psi}_n([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa))\}I_t\} \\ < E\|X_t\|\cdot\delta\cdot JKn^{-\vartheta}. \end{aligned}$$

Since $\vartheta > \frac{1}{2}$, it implies that for any $\varepsilon > 0$

$$\begin{aligned} P\Big(\Big\|n^{-1/2}\sum_{t=1}^{n}X_t\{\psi([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa)) \\ -\tilde{\psi}_n([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa))\}\Big\| > \varepsilon\Big) \\ \leqslant \varepsilon^{-1}n^{-1/2}\sum_{t=1}^{n}E\|X_t\|\cdot\delta\cdot JKn^{-\vartheta} = o(1). \end{aligned}$$

An analogical way leads to

$$(72)\qquad n^{-1/2}\sum_{t=1}^{n}X_t\{\psi(e_t\sigma_{e_1}^{-1})-\tilde{\psi}_n(e_t\sigma_{e_1}^{-1})\} = o_p(1).$$

Applying (17) we obtain

$$\begin{aligned} \sup_{\mathscr{T}_\theta}\Big\|n^{-1/2}\sum_{t=1}^{n}X_t\{\tilde{\psi}_n([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa))-\tilde{\psi}_n(e_t\sigma_{e_1}^{-1})\} \\ +q\tau+w\kappa\Big\| \\ \leqslant n^{-1/2}\sup_{\mathscr{T}_\theta}\Big\|\sum_{t=1}^{n}X_t\{\tilde{\psi}_n([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa)) \\ -\psi([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa))\}\Big\| \end{aligned}$$

$$+n^{-1/2}\sup_{\mathscr{T}_\theta}\Big\|\sum_{t=1}^{n} X_t\{\tilde{\psi}_n(e_t\sigma_{e_1}^{-1})-\psi(e_t\sigma_{e_1}^{-1})\}\Big\|+\sup_{\mathscr{T}_\theta}\|n^{-1/2}S_n(\tau,\kappa)+q\tau+w\kappa\|$$
$$= o_p(1) \quad \text{as } n\to\infty.$$

But it states that we may use the same idea which was used in the proof of Theorem 4, since we have

$$(73)\qquad n^{-1/2}\sum_{t=1}^{n} X_t\tilde{\psi}_n([e_t-n^{-1/2}X_t^{\mathrm{T}}\tau]\sigma_{e_1}^{-1}\exp(-n^{-1/2}\kappa))$$
$$=n^{-1/2}\sum_{t=1}^{n} X_t\tilde{\psi}_n(e_t\sigma_{e_1}^{-1})-q\tau-w\kappa+o_p(1)$$

(compare (70) and (73)). Since it is clear that for the function ψ (this is really ψ, not $\tilde{\psi}_n$) we can verify (as in the proof of Theorem 4) the assumptions of the Lindeberg–Lévy theorem, so that again for any $\varepsilon>0$ we find a constant C_1 and $n_0\in N$ such that for any $n>n_0$ we have

$$P(B_n)>1-\varepsilon \quad \text{for } B_n=\Big\{\omega\in\Omega\colon \Big\|n^{-1/2}\sum_{t=1}^{n} X_t\psi(e_t\sigma_{e_1}^{-1})\Big\|<C_1\Big\}.$$

Using (72) once again we can find an $n_1\geqslant n_0$ such that for any $n>n_1$ we have

$$P(\tilde{B}_n)>1-2\varepsilon \quad \text{for } \tilde{B}_n=\Big\{\omega\in\Omega\colon \Big\|n^{-1/2}\sum_{t=1}^{n} X_t\tilde{\psi}_n(e_t\sigma_{e_1}^{-1})\Big\|<C_1\Big\}.$$

Accomplishing similar modifications in the rest of the proof of Theorem 4 we complete the proof of Theorem 5. ∎

Proof of Theorem 6. For simplicity, let us assume at first that the corresponding ψ-function is continuous, $W=0$ and $\sigma_{e_1}=1$. The asymptotic representation of $\hat{\beta}^{(M,n)}$ may be written as (see (19))

$$(74)\qquad \hat{\beta}^{(M,n)}-\beta^0=\frac{1}{n}\gamma^{-1}Q_n^{-1}\sum_{t=1}^{n} X_t\psi(e_t)+\eta,$$

where η is a p-dimensional random vector of order $\mathcal{O}_p(n^{-1})$. Then

$$(75)\qquad r(\hat{\beta}^{(M,n)})=Y-X\cdot\hat{\beta}^{(M,n)}=X\beta^0+e-X\hat{\beta}^{(M,n)}=e-X(\hat{\beta}^{(M,n)}-\beta^0)$$
$$=e-\frac{1}{n}\gamma^{-1}XQ_n^{-1}\sum_{t=1}^{n} X_t\psi(e_t)+X\cdot\eta.$$

Let us recall what we have already known about the magnitude of the terms in (75) (under the hypothesis of independence). The vector of random fluctuations e has of course the coordinates of order $O_p(1)$. The second term

$$\frac{1}{n}\gamma^{-1}XQ_n^{-1}\sum_{t=1}^{n} X_t\psi(e_t)$$

has its coordinates of order $O_p(n^{-1/2})$, since the assumptions of the Lindeberg–Lévy theorem (see (14)) hold for the sequence $\{X_t\psi(e_t)\}_{t=1}^{\infty}$ of independent identically distributed random variables with zero mean and finite positive variance, and hence $n^{-1/2}\sum_{t=1}^{n} X_t\psi(e_t)$ is bounded in probability. Finally, coordinates of the term $X\eta$ are of order $O_p(n^{-1})$. Let us put

$$\kappa = e - \frac{1}{n}\gamma^{-1} X Q_n^{-1} \sum_{t=1}^{n} X_t \psi(e_t) \quad \text{and} \quad \tau = X\cdot\eta .$$

Then for any positive semidefinite matrix A we have

$$r^{\mathrm{T}}(\hat{\beta}^{(M,n)}) A r(\hat{\beta}^{(M,n)}) = \kappa^{\mathrm{T}} A \kappa + \kappa^{\mathrm{T}} A \tau + \tau^{\mathrm{T}} A \kappa + \tau^{\mathrm{T}} A \tau .$$

From what we have just stated on the order of coordinates of κ and τ, it is clear that the leading term $\kappa^{\mathrm{T}} A\kappa$ is about $n^{1/2}$ times greater than any other term in the last equality. That is why we shall take in the next considerations into account only the term $\kappa^{\mathrm{T}} A\kappa$ in the numerator of Durbin–Watson statistic evaluated for $r(\hat{\beta}^{(M,n)})$. Since the situation of the denominator is similar, we shall consider the statistics

$$\frac{\kappa^{\mathrm{T}} A \kappa}{\kappa^{\mathrm{T}} \kappa}$$

as an approximation to the precise value of $z(\hat{\beta}^{(M,n)})$. Since $\kappa^{\mathrm{T}}\kappa$ is not negligible in probability, we have

$$\left| z(\hat{\beta}^{(M,n)}) - \frac{\kappa^{\mathrm{T}} A \kappa}{\kappa^{\mathrm{T}} \kappa} \right| = O_p(n^{-1/2}).$$

Let us put

$$\Psi(e) = \big(\psi(e_1), \psi(e_2), \ldots, \psi(e_n)\big)^{\mathrm{T}}, \qquad \varphi = [\mathscr{I} - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}]\, e = M\cdot e$$

and

$$\phi = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\left[e - \frac{\Psi(e)}{\gamma}\right] = \tilde{M}\left[e - \frac{\Psi(e)}{\gamma}\right] \quad \text{(say)}. \tag{76}$$

Then

$$\kappa = \varphi + \phi \quad \text{and} \quad M\cdot\tilde{M} = 0, \tag{77}$$

so that

$$\varphi^{\mathrm{T}}\cdot\phi = 0,$$

which implies that

$$\frac{\kappa^{\mathrm{T}} A \kappa}{\kappa^{\mathrm{T}} \kappa} = \frac{\varphi^{\mathrm{T}} A \varphi + 2\varphi^{\mathrm{T}} A \phi + \phi^{\mathrm{T}} A \phi}{\varphi^{\mathrm{T}}\varphi + \phi^{\mathrm{T}}\phi} \tag{78}$$

$$= \frac{e^{\mathrm{T}} M^{\mathrm{T}} A M e + 2 e^{\mathrm{T}} M^{\mathrm{T}} A \tilde{M}[e - \Psi(e)/\gamma] + [e - \Psi(e)/\gamma]^{\mathrm{T}} \tilde{M}^{\mathrm{T}} A \tilde{M}[e - \Psi(e)/\gamma]}{e^{\mathrm{T}} M e + [e - \Psi(e)/\gamma]^{\mathrm{T}} \tilde{M}[e - \Psi(e)/\gamma]} .$$

Using Assertion 3 let us find an orthogonal matrix L (i.e. $L^T \cdot L = \mathscr{I}_n$) so that

$$L^T M L = D, \tag{79}$$

where D is the diagonal matrix and the diagonal elements are just the eigenvalues of the matrix M. (Notice that due to the fact that L is orthogonal, and hence regular, it has the left inverse matrix equal to the right inverse one, and so we have also $LL^T = \mathscr{I}_n$.) Let us recall that the matrix L is created from the eigenvectors of the matrix M and that M is the projection matrix (into the space $\mathscr{M}(M)$, the vector space generated by columns of the matrix M.) This implies that it is idempotent ($M \cdot M = M$), and hence D contains only ones and zeros. Since $\operatorname{rank}(D) = \operatorname{rank}(M) = n-p$, we may arrange the columns of L so that

$$L^T M L = \begin{bmatrix} \mathscr{I}_{n-p} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix},$$

where 0 stands for a zero matrix with appropriate numbers of rows and columns. Now we may write

$$\begin{aligned} L^T MAML &= L^T ML \cdot L^T AL \cdot L^T ML \\ &= \begin{bmatrix} \mathscr{I}_{n-p} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix} \cdot \begin{bmatrix} B_1 & \vdots & B_3 \\ \cdots & \cdots & \cdots \\ B_2 & \vdots & B_4 \end{bmatrix} \cdot \begin{bmatrix} \mathscr{I}_{n-p} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix}, \end{aligned} \tag{80}$$

where

$$\begin{bmatrix} B_1 & \vdots & B_3 \\ \cdots & \cdots & \cdots \\ B_2 & \vdots & B_4 \end{bmatrix}$$

is the appropriate partition of the real symmetric matrix $L^T AL$. Now returning to (79) we may notice that

$$L^T \tilde{M} L = L^T (\mathscr{I}_n - M) L = L^T L - L^T ML = \begin{bmatrix} 0 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \mathscr{I}_p \end{bmatrix}.$$

Similarly, returning to (80) we infer that for $\tilde{M} = X(X^T X)^{-1} X^T$ (see (76))

$$\begin{aligned} L^T \tilde{M} A \tilde{M} L &= L^T \tilde{M} L L^T A L L^T \tilde{M} L \\ &= \begin{bmatrix} 0 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \mathscr{I}_p \end{bmatrix} \begin{bmatrix} B_1 & \vdots & B_3 \\ \cdots & \cdots & \cdots \\ B_2 & \vdots & B_4 \end{bmatrix} \begin{bmatrix} 0 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \mathscr{I}_p \end{bmatrix}. \end{aligned}$$

Let N_1 and N_2 be orthogonal matrices diagonalizing B_1 and B_4, respectively, i.e.

$$N_1^{\mathrm{T}} B_1 N_1 = \begin{bmatrix} \nu_1 & & & \\ & \nu_2 & & \\ & & \ddots & \\ & & & \nu_{n-p} \end{bmatrix}$$

and

$$N_2^{\mathrm{T}} B_4 N_2 = \begin{bmatrix} \nu_{n-p+1} & & & \\ & \nu_{n-p+2} & & \\ & & \ddots & \\ & & & \nu_n \end{bmatrix},$$

where the blank spaces represent zeros, $\nu_1, \nu_2, \ldots, \nu_{n-p}$ are eigenvalues of B_1 (and of course also nonzero eigenvalues of MA) and $\nu_{n-p+1}, \nu_{n-p+2}, \ldots, \nu_n$ are eigenvalues of B_4 (which are equal to nonzero eigenvalues of $\tilde{M}A$). Then

$$\tilde{N} = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix}$$

is orthogonal and for $\tilde{H} = L \cdot \tilde{N}$ we have

$$\tilde{H}^{\mathrm{T}} M \tilde{H} = \tilde{N}^{\mathrm{T}} L^{\mathrm{T}} M L \tilde{N} = \tilde{N}^{\mathrm{T}} \begin{bmatrix} \mathscr{I}_{n-p} & 0 \\ 0 & 0 \end{bmatrix} \tilde{N} = \begin{bmatrix} \mathscr{I}_{n-p} & 0 \\ 0 & 0 \end{bmatrix} \tag{81}$$

and

$$\tilde{H}^{\mathrm{T}} \tilde{M} \tilde{H} = \tilde{N}^{\mathrm{T}} L^{\mathrm{T}} \tilde{M} L \tilde{N} = \tilde{N}^{\mathrm{T}} \begin{bmatrix} 0 & 0 \\ 0 & \mathscr{I}_p \end{bmatrix} \tilde{N} = \begin{bmatrix} 0 & 0 \\ 0 & \mathscr{I}_p \end{bmatrix}. \tag{82}$$

Further, we obtain

$$\tilde{H}^{\mathrm{T}} MAM \tilde{H} = \tilde{H}^{\mathrm{T}} M \tilde{H} \tilde{H}^{\mathrm{T}} A \tilde{H} \tilde{H}^{\mathrm{T}} M \tilde{H} = \left[\begin{array}{cccc:c} \nu_1 & & & & \\ & \nu_2 & & & \\ & & \ddots & & 0 \\ & & & \nu_{n-p} & \\ \hdashline & & 0 & & 0 \end{array}\right] \tag{83}$$

and

$$\tilde{H}^{\mathrm{T}} \tilde{M} A \tilde{M} \tilde{H} = \tilde{H}^{\mathrm{T}} \tilde{M} \tilde{H} \tilde{H}^{\mathrm{T}} A \tilde{H} \tilde{H}^{\mathrm{T}} \tilde{M} \tilde{H} \tag{84}$$

$$= \left[\begin{array}{c:cccc} 0 & & & 0 & \\ \hdashline & \nu_{n-p+1} & & & \\ & & \nu_{n-p+2} & & \\ 0 & & & \ddots & \\ & & & & \nu_n \end{array}\right].$$

Notice that by (83) we have $\operatorname{rank}(MA) = n-p$. Since due to the right part of (77) the "cross-term" in the denominator of (78) has disappeared, it seems quite natural to expect that similarly the "cross-term" in the numerator may disappear in view of the equality

$$(85) \qquad MA\tilde{M} = 0.$$

Unfortunately, (85) is not generally true, as the following example shows. Let

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \tilde{M} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Then

$$MA\tilde{M} = \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix}.$$

It causes that the analysis of the numerator of (78) will be somewhat more complicated. First of all, let us recall that $M = \mathscr{I}_n - \tilde{M}$, and hence

$$(86) \qquad MA\tilde{M} = A\tilde{M} - \tilde{M}A\tilde{M}.$$

Further, let us put

$$(87) \quad \xi = \tilde{H}^{\mathrm{T}}e, \quad \vartheta = L^{\mathrm{T}}Ae, \quad \tilde{\vartheta} = L^{\mathrm{T}}\left[e - \frac{\Psi(e)}{\gamma}\right], \quad \zeta = \tilde{H}^{\mathrm{T}}\left[\frac{e-\Psi(e)}{\gamma}\right],$$

and let us recall that $\tilde{H}$ is orthogonal, i.e. that we have also $\tilde{H}\xi = e$ and $\tilde{H}\zeta = e - \Psi(e)/\gamma$. Keeping in mind that X_i's are independent and identically distributed, notice that due to character of the matrices $L^{\mathrm{T}}A$, L^{T} and $\tilde{H}^{\mathrm{T}}$ all coordinates of the random vectors ϑ, $\tilde{\vartheta}$ and ζ are bounded in probability, and this holds uniformly with respect to n, i.e. with respect to the number of observations. Of course, we do not speak about the maximum of coordinates of these vectors, i.e. about e.g. $\max_{1 \leqslant t \leqslant n} |\vartheta_t|$ but we claim that we can find for all $\varepsilon > 0$ a constant $K > 0$ such that, for all $n \in N$ and $1 \leqslant t \leqslant n$, $P(|\vartheta_t| > K) < \varepsilon$ (it will be clear from the following that we shall need to know something about simultaneous behavior of the last p coordinates). Then we have

$$(88) \qquad e^{\mathrm{T}}A\tilde{M}\left[e - \frac{\Psi(e)}{\gamma}\right] = e^{\mathrm{T}}ALL^{\mathrm{T}}\tilde{M}LL^{\mathrm{T}}\left[e - \frac{\Psi(e)}{\gamma}\right] = \sum_{t=n-p+1}^{n} \vartheta_t \tilde{\vartheta}_t$$

and

$$(89) \qquad e^{\mathrm{T}}\tilde{M}A\tilde{M}\left[e - \frac{\Psi(e)}{\gamma}\right] = \sum_{t=n-p+1}^{n} \nu_t \xi_t \zeta_t.$$

Now using (78), (81)–(84), (86), (88) and (89), we obtain

$$(90)\quad z(\hat\beta^{(M,n)}) = \frac{\sum_{t=1}^{n-p} v_t \tilde\xi_t^2 + \sum_{t=n-p+1}^{n} \vartheta_t \tilde\vartheta_t + \sum_{t=n-p+1}^{n} v_t \tilde\xi_t \zeta_t + \sum_{t=n-p+1}^{n} v_t \zeta_t^2}{\sum_{t=1}^{n-p} \tilde\xi_t^2 + \sum_{t=n-p+1}^{n} \zeta_t^2}.$$

It is clear that generally $z(\hat\beta^{(M,n)})$ and the denominator of (90), i.e. $r^{\mathrm{T}}(\hat\beta^{(M,n)})\, r(\hat\beta^{(M,n)})$, are not independent, and hence it is not possible to use directly the approach for Durbin–Watson which was based on Pitman's or von Neumann's result (see Pitman (1937), von Neumann (1941)).

In the further analysis let us restrict ourselves on the matrix A_1 given in (6). Von Neumann (1941) has already shown (see e.g. Durbin and Watson (1952) or the original paper by von Neumann (1941)) that

$$(91)\qquad \lambda_t = 2\left\{1 - \cos\frac{\pi(t-1)}{n}\right\}, \qquad t = 1, 2, \ldots, n,$$

and so (see (5)) v_t's are bounded from above by 4 and from below by 0. Since $M = \mathscr{I} - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ has rank $n-p$ (as we have already mentioned ealier, this follows from (83)), there are $n-p$ of its eigenvalues which are different from zero. Then employing the Durbin–Watson lemma, we may see the following. Having a fixed sufficiently small positive number δ, there is an $n_0 \in N$ such that for all $n > n_0$ the value of the most of v_t's is greater than that δ. Taking into account the transformation which defines $\tilde\xi$, namely $\tilde\xi = \tilde H^{\mathrm{T}} e$, we conclude that $\tilde\xi$ has independent coordinates. Then using the Lindeberg–Lévy theorem, we may show that $n^{-1/2}\sum_{t=1}^{n-p} v_t \tilde\xi_t^2$ is asymptotically normally distributed, and hence $\sum_{t=1}^{n-p} v_t \tilde\xi_t^2$ (which is of course nonnegative) goes to infinity in probability. On the other hand, let us notice that all the sums in the numerator, except of the first one, contain for any n only p elements. It implies that these three last sums in the numerator of (90) are uniformly in n bounded in probability. In other words, they are asymptotically negligible with respect to the first one. A similar conclusion holds about the second sum in the denominator of $z(\hat\beta^{(M,n)})$. So we may conclude that asymptotically Pitman's and von Neumann's result holds, i.e. that $z(\hat\beta^{(M,n)})$ and its denominator are asymptotically indepedent. But it immediately implies the assertion of the theorem because then both types of statistics are the same, i.e. $z_L(\hat\beta^{(LS,n)})$ and $z_L(\hat\beta^{(M,n)})$ are asymptotically equal each to the other. Similarly we argue for the upper statistics.

The proof for a discontinuous ψ-function is analogical since the asymptotic representation of the $\hat\beta^{(M,n)} - \beta^0$ for the continuous and the discontinuous function are of the same character (see Remark 6 and Corollary 3). ∎

APPENDIX II. AUXILIARY ASSERTIONS

LEMMA A.2 (Štěpán (1987), p. 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 $E\xi = 0$. Moreover, let τ be the time for the Wiener process $W(s)$ to exit the interval $(-a, b)$. Then*

$$\xi \overset{\mathscr{D}}{=} W(\tau),$$

where $\overset{\mathscr{D}}{=}$ denotes the equality in distribution of the corresponding random variables. Moreover, $E\tau = a \cdot b = \operatorname{var} \xi$.

Remark A.9. Since the book by 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 (p. 277) of Breiman's book. (See also Theorem 13.6 on p. 276.)

We shall need however a somewhat generalized version of the previous lemma.

LEMMA A.3. *Let a and b be positive numbers. Further, let ξ be a random variable defined on a probability space $(\Omega, \mathscr{A}, P)$ 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 $E\xi = 0$. Finally, write $A = \{\omega \in \Omega \colon \xi(\omega) = 0\}$, for $\omega \in A$ put $\tau(\omega) = 0$, and for $\omega \in A^C$ let τ be the time for the Wiener process $W(s)$ to exit the interval $(-a, b)$. Then*

$$\xi \overset{\mathscr{D}}{=} W(\tau)$$

where $\overset{\mathscr{D}}{=}$ denotes the equality in distribution of the corresponding random variables. Moreover, $E\tau = a \cdot b \cdot (1-\pi_3) = \operatorname{var} \xi$.

Proof. Let us put $\tilde{\Omega} = A^C$, $\tilde{\mathscr{A}} = A^C \cap \mathscr{A}$, and $\tilde{P}(B) = \pi_3^{-1} P(B)$. Further, let $\tilde{\xi} = \xi$ for $\omega \in A^C$. 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 for $\omega \in A^C$ put $\tau(\omega) = \tilde{\tau}(\omega)$. According to Lemma A.1 we have $\tilde{\xi} \overset{\mathscr{D}}{=} W(\tilde{\tau})$ and $E\tilde{\tau} = a \cdot b = \operatorname{var} \tilde{\xi}$. Earlier than we shall continue, let us realize that $E\tilde{\tau} = \int_0^\infty z\tilde{f}(z)\,dz = a \cdot b$, where $\tilde{f}(z)$ is a density of distribution of $\tilde{\tau}$. Now, evidently, $\xi \overset{\mathscr{D}}{=} W(\tau)$ because on the set A we have $\xi = 0 = W(\tau)$ and

$$E\tau = \int_0^\infty zf(z)\,dz = 0 \cdot P(A) + \int_0^\infty z\tilde{f}(z)(1-\pi_3)\,dz = a \cdot b \cdot (1-\pi_3) = \operatorname{var} \xi. \blacksquare$$

ASSERTION A.4. *Let U be an open bounded set in R^p and assume that $Q(z) \colon \bar{U} \subset R^p \to R^p$ ($\bar{U}$ is the closure of U) is continuous and satisfies $(z-z_0)^T Q(z) \geqslant 0$ 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 p. 163.

REFERENCES

[1] T. W. Anderson, *On the theory of testing serial correlation*, Skandinavisk Aktuarietidskrift 31 (1948), pp. 88–116.

[2] J. Antoch and J. Á. Víšek, *Robust estimation in linear models and its computational aspects*, in: *Contributions to Statistics: Computational Aspects of Model Choice*, J. Antoch (Ed.), Springer, 1991, pp. 39–104.

[3] J. Antoch and D. Vorlíčková, *Vybrané metody statistické analýzy dat* (*Selected Methods of Statistical Analysis* – in Czech) Academia, Praha 1992.

[4] O. Arslan, *A simple test to identify good solution of redescending M-estimating equations for regression*, in: *Developments in Robust Statistics*, R. Dutter, P. Filzmoser, U. Gather and P. J. Rousseeuw (Eds.), Physica-Verlag, Springer-Verlag Company 2003, pp. 50–61.

[5] A. C. Atkinson, *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Claredon Press, Oxford 1985.

[6] D. A. Belsley, E. Kuh and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, New York 1980.

[7] R. Beran, *An efficient and robust adaptive estimator of location*, Ann. Statist. 6 (1978), pp. 292–313.

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

[9] G. Boente and R. Fraiman, *A functional approach to robust nonparametric regression*, in: *Directions in Robust Statistics and Diagnostics*, Part I, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 35–46.

[10] L. Breiman, *Probability*, Addison-Wesley Publishing Company, London 1968.

[11] S. Chatterjee and A. S. Hadi, *Sensitivity Analysis in Linear Regression*, Wiley, New York 1988.

[12] R. D. Cook and S. Weisberg, *Diagnostics of heteroscedasticity in regression*, Biometrika 70 (1983), pp. 1–10.

[13] C. Croux and G. Haesbroeck, *Influence function and efficiency of the minimum covariance determinant scatter matrix estimation*, J. Multivariate Anal. 71 (1999), pp. 161–190.

[14] M. Csörgö and P. Révész, *Strong Approximation in Probability and Statistics*, Akadémiai Kiadó, Budapest 1981.

[15] M. B. Dollinger and R. G. Staudte, *Efficiency of reweighted least squares iterates*, in: *Directions in Robust Statistics and Diagnostics*, Part I, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 61–66.

[16] N. R. Draper and H. Smith, *Applied Regression Analysis*, 3rd edition, Wiley, New York 1998.

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

[18] S. Geisser, *Diagnostics, divergence and perturbation analysis*, in: *Directions in Robust Statistics and Diagnostics*, Part I, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 89–100.

[19] W. H. Greene, *Econometric Analysis*, Macmillan Press, New York 1993.

[20] A. S. Hadi, *Identifying multiple outliers in multivariate data*, J. Roy. Statist. Soc., Ser. B, 54 (1992), pp. 761–777.

[21] A. S. Hadi, *A modification of a method for the detection of outliers in multivariate samples*, J. Roy. Statist. Soc., Ser. B, 56 (1994), pp. 393–396.

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

[23] M. A. Hauser, *Semiparametric and nonparametric testing for long memory: A Monte Carlo study*, Empirical Economics 22 (1997), pp. 247–271.

[24] T. P. Hettmansperger and J. D. Naranjo, *Some research directions in rank-based infer-*

ence, in: *Directions in Robust Statistics and Diagnostics*, Part I, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 113–120.

[25] T. P. Hettmansperger and S. J. Sheather, *A cautionary note on the method of least median squares*, Amer. Statist. 46 (1992), pp. 79–83.

[26] P. J. Huber, *Robust estimation of a location parameter*, Ann. Math. Statist. 35 (1964), pp. 73–101.

[27] P. J. Huber, *Robust Statistics*, Wiley, New York 1981.

[28] P. J. Huber, *Between robustness and diagnostics*, in: *Directions in Robust Statistics and Diagnostics*, Part I, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 121–131.

[29] G. G. Judge, W. E. Griffiths, R. C. Hill, H. Lütkepohl and T. Ch. Lee, *The Theory and Practice of Econometrics*, Wiley, New York 1985.

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

[31] J. Jurečková and M. Malý, *The asymptotics for studentized k-step M-estimators of location*, Sequential Anal. 14 (1995), pp. 229–245.

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

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

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

[35] J. Jurečková and A. H. Welsh, *Asymptotic relations between L- and M-estimators in the linear model*, Ann. Inst. Statist. Math. 42 (1990), pp. 671–698.

[36] J. Kmenta, *Elements of Econometrics*, Macmillan Publishing Company, New York 1986.

[37] R. Koenker and G. Bassett, *Regression quantiles*, Econometrica 46 (1978), pp. 33–50.

[38] T. S. Kuhn, *Structure of Scientific Revolution*, University of Chicago Press, Chicago, Phoenix Broks 159 (1965).

[39] A. J. Lawrence, *Local and deletion influence*, in: *Directions in Robust Statistics and Diagnostics*, Part I, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 141–158.

[40] C. C. Macduffee, *The Theory of Matrices*, Chelsea Publishing Company, 1946.

[41] G. S. Maddala, *Introduction to Econometrics*, Macmillan Press, New York 1988.

[42] M. Markatou and X. He, *Bounded influence and high breakdown point testing procedures in linear models*, J. Amer. Statist. Assoc. 89 (1994), pp. 543–549.

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

[44] R. A. Maronna and V. J. Yohai, *The breakdown point of simultaneous general M-estimates of regression and scale*, J. Amer. Statist. Assoc. 86 (1981), pp. 699–704.

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

[46] J. W. McKean, S. J. Sheather and T. P. Hettmansperger, *Regression diagnostics for rank-based method. II*, in: *Directions in Robust Statistics and Diagnostics*, Part II, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 21–32.

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

[48] J. von Neumann, *Distribution of the ratio of the mean-square successive difference to the variance*, Ann. Math. Statist. 12 (1941), pp. 367–395.

[49] A. Ö. Önder and A. Zaman, *A test of normality based on robust regression residuals*, in: *Developments in Robust Statistics*, R. Dutter, P. Filzmoser, U. Gather and P. J. Rousseeuw (Eds.), Physica-Verlag, Springer-Verlag Company, 2003, pp. 296–306.

[50] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York–London 1970.

[51] A. Peters and P. Sibbertsen, *Tests on fractional cointegration*, in: *Developments in Robust Statistics*, R. Dutter, P. Filzmoser, U. Gather and P. J. Rousseeuw (Eds.), Physica-Verlag, Springer-Verlag Company, 2003, pp. 307–316.

[52] E. J. G. Pitman, *The 'closest' estimates of statistical parameters*, Proc. Cambridge Phil. Soc. 33 (1937), pp. 212–222.

[53] D. Pollard, *Asymptotics for least absolute deviation regression estimator*, Econometric Theory 7 (1991), pp. 186–199.

[54] K. R. Popper, *Objective Knowledge*, Oxford Clarendon Press, 1972.

[55] S. Portnoy, *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 and D. Martin (Eds.), Springer, New York 1983, pp. 231–246.

[56] S. Portnoy, *Regression quantile diagnostics for multiple outliers*, in: *Directions in Robust Statistics and Diagnostics*, Part II, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 145–158.

[57] I. Prigogine and I. Stengers, *La Nouvelle Alliance*, SCIENTIA, 1977, issues 5–12.

[58] I. Prigogine and I. Stengers, *Order out of Chaos. Man's New Dialog with Nature*, Bantam Books, New York 1984.

[59] R. C. Rao and L. C. Zhao, *On the consistency of M-estimate in linear model obtained through an estimating equation*, Statist. Probab. Lett. 14 (1992), pp. 79–84.

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

[61] A. M. Rubio and J. Á. Víšek, *Asymptotic representation of constrained M-estimators*, in: *Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Prague, 1994* (1993), pp. 207–210.

[62] A. M. Rubio and J. Á. Víšek, *A note on asymptotic linearity of M-statistics in nonlinear models*, Kybernetika 32 (1996), pp. 353–374.

[63] R. Schall and T. T. Dunne, *Diagnostics for regression — ARMA time series*, in: *Directions in Robust Statistics and Diagnostics*, Part II, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 223–242.

[64] W. Stahel and S. Weisberg (Eds.), *Directions in Robust Statistics and Diagnostics*, Parts I and II, Springer, New York 1991.

[65] J. Štěpán, *Teorie Pravděpodobnosti* (*Theory of Probability* — in Czech), Academia, Praha 1987.

[66] C. Stone, *Adaptive maximum likelihood estimators of a location parameter*, Ann. Statist. 3 (1975), pp. 267–284.

[67] J. W. Tukey, *Graphical displays for alternate regression fits*, in: *Directions in Robust Statistics and Diagnostics*, Part II, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 145–158.

[68] C. M. Urzua, *On the correct use of omnibus tests for normality*, Econom. Lett. 53 (1996), pp. 247–251.

[69] J. Á. Víšek, *Adaptive estimation in linear regression model*, Kybernetika 28 (1991), *Part I. Consistency*, pp. 26–36, *Part II. Asymptotic normality*, pp. 100–119.

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

[71] 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*, Academia Prague–Kluwer Academic Press, Dordrecht 1994, pp. 254–259.

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

[73] J. Á. Víšek, *On the heuristics of statistical results*, in: *Proceedings of 'PROBASTAT' '94*, Tatra Mountains Publ. 7, Bratislava 1996b, pp. 349–357.

[74] J. Á. Víšek, *Sensitivity analysis on M-estimates*, Ann. Inst. Statist. Math. 48 (1996c), pp. 469–495.

[75] 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.

[76] J. Á. Víšek, *Robust instruments*, in: *Robust '98*, J. Antoch and G. Dohnal (Eds.), published by the Union of Czechoslovak Mathematicians and Physicists, 1998a, pp. 195–224.

[77] J. Á. Víšek, *Robust specification test*, in: *Proceedings of Prague Stochastics '98*, M. Hušková, P. Lachout and J. Á. Víšek (Eds.), published by the Union of Czechoslovak Mathematicians and Physicists 1998b, pp. 581–586.

[78] J. Á. Víšek, *The least trimmed squares — random carriers*, Bull. Czech Econometric Soc. 10 (1999a), pp. 1–30.

[79] J. Á. Víšek, *Robust estimation of regression model*, Bull. Czech Econometric Soc. 9 (1999b), pp. 57–79.

[80] J. Á. Víšek, *The robust regression and the experiences from its application on estimation of parameters in a dual economy*, in: *Proceedings of the Conference "Macromodels '99" organized by Wroclaw University, 4–6 December, 1999, in Rydzyna, Poland*, 1999c, pp. 424–445.

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

[82] J. Á. Víšek, *Robust instrumental variables and specification test*, in: *PRASTAN 2000, Proceedings of conference "Mathematical Statistics and Numerical Mathematics and Their Applications"*, 2000b, pp. 133–164.

[83] J. Á. Víšek, *Over- and underfitting the M-estimates*, Bull. Czech Econometric Soc. 7 (2000c), pp. 53–83.

[84] J. Á. Víšek, *Overfitting the least trimmed squares*, in: *PRASTAN 2001, Proceedings of conference "Mathematical Statistics and Numerical Mathematics and Their Applications"*, 2001a, pp. 160–166.

[85] J. Á. Víšek, *Regression with high breakdown point*, in: *ROBUST2000*, organized by the Union of the Czech Mathematicians and Physicists and the Czech Statistical Society, 2001b, pp. 324–356.

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

[87] J. Á. Víšek, *White test for the least weighted squares*, in: *COMPSTAT 2002, Berlin, Proceedings of the Conference CompStat 2002 — Short Communications and Poster (CD)*, S. Klinke, P. Ahrend and L. Richter (Eds.), 2002b.

[88] J. Á. Víšek, *Development of the Czech export in nineties*, in: *Consolidation of Governing and Business in the Czech Republic and in EU, 2003*, MatFyz Press, 2003, pp. 193–220.

[89] R. E. Welsh, *Influence function and regression diagnostic*, in: *Modern Data Analysis*, R. L. Launer and A. F. Sigel (Eds.), Academic Press, New York 1982.

[90] G. Willem, G. Pison, P. J. Rousseeuw and S. van Aelst, *A robust Hotelling test*, in: *Developments in Robust Statistics*, R. Dutter, P. Filzmoser, U. Gather and P. J. Rousseeuw (Eds.), Physica-Verlag, Springer-Verlag Company, 2003, pp. 417–431.

[91] V. J. Yohai and R. A. Maronna, *Asymptotic behaviour of M-estimators for the linear model*, Ann. Statist. 7 (1979), pp. 248–268.

[92] K. Zvára, *Regresní analýza (Regression Analysis* — in Czech), Academia, Praha 1989.

Department of Macroeconomics and Econometrics, Institute of Economic Studies
Faculty of Social Sciences, Charles University
and
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 5.7.2002;
revised version on 4.12.2003