An alternative method for constructing confidence intervals from M-estimates in linear models

Marek Omelka

Abstract: We will investigate an alternative way how to construct a confidence interval based on M-estimator for a single parameter in a linear model. We will compare this confidence interval with a traditional (Wald type) confidence interval theoretically as well as by the means of a Monte-Carlo experiment.

MSC 2000: 62F25, 62F35
Key words: M-estimator, linear regression, confidence interval

1 Introduction

Suppose that our observations $\mathbf{Y} = (Y_1, \ldots, Y_n)^\mathsf{T}$ follow the linear model

$$Y_i = \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + e_i = \boldsymbol{\beta}^\mathsf{T}\mathbf{x}_i + e_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\mathsf{T}$ is a vector of unknown parameters, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\mathsf{T}$, for $i = 1, \ldots, n$ are known constants and $e_1, \ldots, e_n$ are independent, identically distributed random variables with a cumulative distribution function (cdf) F. In the following we will assume that the model includes intercept, that is $x_{i1} = 1$ for $i = 1, \ldots, n$. The studentized M-estimator $\hat{\boldsymbol{\beta}}_M$ is usually defined as a solution of the system of equations

$$\sum_{i=1}^{n} \mathbf{x}_i \, \psi\big(\tfrac{Y_i - b^\mathsf{T}\mathbf{x}_i}{S_n}\big) = \mathbf{0} \tag{2}$$

where $\psi : \mathbb{R} \mapsto \mathbb{R}$ is a (bounded) monotone or redescending function and S_n is an appropriate estimate of scale.

If we omit resampling procedures, there are generally two main ways how to construct a confidence interval (CI) for β_l ($1 \leq l \leq p$) with the help of the knowledge of asymptotic distribution or $\sqrt{n}(\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})$. The first (and the most common) way directly exploits first order asymptotic linearity of the M-estimator (see Jurečková and Sen [4])

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}) = \frac{\mathbf{V}_n^{-1}}{\gamma_1 \sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \, \psi\big(\tfrac{e_i}{S}\big) + \tfrac{\gamma_{1e}}{\gamma_1} \sqrt{n}\big(\tfrac{S_n}{S} - 1\big)\, \mathbf{u}_1 + o_p(1), \tag{3}$$

where $\mathbf{V}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\mathsf{T}$, $\mathbf{u}_1 = (1, 0, \ldots, 0)^\mathsf{T}$, $\gamma_1 = \mathrm{E}\,\frac{1}{S}\psi'(\frac{e_1}{S})$, $\gamma_{1e} = \mathrm{E}\,\frac{e_1}{S}\psi'(\frac{e_1}{S})$ and by $S = S(F)$ we understand the theoretical value of the scale estimator S_n.

Research was supported by the Czech Republic Grant GAČR 201/05/H007 and by the Charles University Grant GAUK 341/2005.

Let us denote $\{\omega_{ij}\}_{i=1,\dots,p}^{j=1,\dots,p}$ the elements of the matrix $\mathbf{V}_n^{-1}$. Then we immediately see from (3) that the random variable $\sqrt{n}(\hat{b}_l - \beta_l)$ has asymptotically zero mean normal distribution with variance $\frac{\sigma_\psi^2 \omega_{ll}}{\gamma_1^2}$, where $\sigma_\psi^2 = \mathrm{E}\,\psi(\frac{e_1}{S})^2$. The (Wald type) confidence interval for the component β_l can be constructed as

$$D_n^I = \left[\hat{b}_l - \tfrac{z_\alpha}{\sqrt{n}} \tfrac{\hat{\sigma}_\psi \sqrt{\omega_{ll}}}{\hat{\gamma}_1},\ \hat{b}_l + \tfrac{z_\alpha}{\sqrt{n}} \tfrac{\hat{\sigma}_\psi \sqrt{\omega_{ll}}}{\hat{\gamma}_1}\right], \tag{4}$$

where $\hat{\sigma}_\psi$ and $\hat{\gamma}_1$ are estimates of the unknown quantities σ_ψ and γ_1, and $z_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$, with Φ^{-1} being the inverse cdf of the standard normal distribution. We will call this **type I confidence interval.** If we put $r_i = Y_i - \hat{\boldsymbol{\beta}}_M^\top \mathbf{x}_i$ for the residuals of the fit with the M-estimator $\hat{\boldsymbol{\beta}}_M$, the most simple estimates of σ_ψ and γ_1 are

$$\hat{\sigma}_\psi^2 = \frac{1}{n-p} \sum_{i=1}^n \psi^2(\tfrac{r_i}{S_n}), \qquad \hat{\gamma}_1 = \frac{1}{n} \sum_{i=1}^n \tfrac{1}{S_n}\, \psi'(\tfrac{r_i}{S_n}). \tag{5}$$

Sometimes we may find the confidence interval (4) inconvenient especially for two reasons. Firstly, we need to estimate two unknown quantities (σ_ψ and γ_1). Secondly, we may be doubtful whether the symmetry of the confidence interval (4) does not affect the good coverage properties, especially in the case of asymmetric errors in the model (1).

Boss in [1] proposed another method for the construction of confidence intervals from M-estimates. He considered the case of location parameter and suggested the confidence interval $[\hat{\theta}_n^-, \hat{\theta}_n^+]$, where

$$\hat{\theta}_n^- = \sup \left\{ t : \tfrac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\tfrac{X_i - t}{S_n}) \geq \hat{\sigma}_\psi\, z_\alpha \right\} \tag{6}$$

$$\hat{\theta}_n^+ = \inf \left\{ t : \tfrac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\tfrac{X_i - t}{S_n}) \leq -\hat{\sigma}_\psi\, z_\alpha \right\}. \tag{7}$$

We will call it **type II confidence interval**. It is apparent from the definition that this method can work only for **monotone** ψ in general. Although, at the cost of some further difficulties, the method can be generalized to include redescending ψ-functions as well, we will assume that ψ is monotone in the following. On the other hand, the advantage of this approach is that we do not need to estimate the functional γ_1. Boss in [1] showed that this method is asymptotically correct and that the length of the confidence interval multiplied by $\sqrt{n}$ converges in probability to the same constant as for the type I confidence interval. By the means of simulation he also demonstrated that his proposed method has sometimes a slightly better coverage properties then the type I method. Some partial explanation of this phenomenon can be found in Lloyd [5].

In the following we will modify the type II method for the multiple linear regression. Similarly as in [1] we will show that the length of the confidence interval for a single component (multiplied by $\sqrt{n}$) has the same probability limit as for the type I method, but the asymptotic distribution of this length (properly standardized) is generally different. The proofs and some further discussion will be included in Omelka [6].

2 Type II confidence interval

2.1 Construction

For the simplicity of notation we will construct the confidence interval for the last component of vector $\boldsymbol{\beta}$ – parameter β_p. The general case would follow by relabeling of the indices. To simplify the subsequent formulas we will make use of the following notations. Let $\mathbf{z}_i$ stands for the vector $\mathbf{x}_i$ without the last component, that is $\mathbf{z}_i = (x_{ip}, \dots, x_{i\,p-1})^\mathsf{T}$ and similarly $\hat{\mathbf{b}}_z = (\hat{b}_1, \dots, \hat{b}_{p-1})^\mathsf{T}$. Finally put $T_{np}^2 = \frac{1}{n}\sum_{i=1}^n x_{ip}^2$.

Now define the function $M_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n x_{ip}\,\psi(\frac{Y_i - \hat{\mathbf{b}}_z^\mathsf{T}\mathbf{z}_i - t\,x_{ip}}{u})$. The (type II) confidence interval for the parameter β_p is $D_n^{II} = [\hat{b}_p^-, \hat{b}_p^+]$, where

$$\hat{b}_p^- = \sup\left\{t : M_n(t) \ge T_{np}^2\,\sqrt{\omega_{pp}}\,\hat{\sigma}_\psi\,z_\alpha\right\} \tag{8}$$

$$\hat{b}_p^+ = \inf\left\{t : M_n(t) \le -T_{np}^2\,\sqrt{\omega_{pp}}\,\hat{\sigma}_\psi\,z_\alpha\right\}. \tag{9}$$

2.2 Assumptions

Before we proceed, we need to formulate some assumptions about the function ψ, the distribution function of the errors F, the scale estimator S_n and the design points $\mathbf{x}_1, \dots, \mathbf{x}_n$.

S.1 The function $h(t) = \mathrm{E}\,\rho(\frac{e_1 - t}{S})$ has the unique minimum at $t = 0$, where $\rho' = \psi$.

S.2 ψ is absolutely continuous with a derivative ψ' such that $\mathrm{E}\,\psi'\left(\frac{e_1}{S}\right)^2 < \infty$.

S.3 The function $\psi'\left(\frac{e_1+t}{Se^u}\right)$ is continuous in the quadratic mean, that is

$$\lim_{(t,u)\to(0,0)} \mathrm{E}\left[\psi'\left(\tfrac{e_1+t}{Se^u}\right) - \psi'\left(\tfrac{e_1}{S}\right)\right]^2 = 0.$$

S.4 The function $\lambda(t,u) = \mathrm{E}\,\psi(\frac{e_1+t}{Se^u})$ is twice differentiable and the partial derivatives are continuous and bounded in a neighborhood of the point $(0,0)$

S.5 $\mathrm{E}\,\psi^4(\frac{e_1}{S}) < \infty$

S.6 The function $\lambda^{(2)}(t,u) = \mathrm{E}\,\psi^2(\frac{e_1+t}{Se^u})$ is continuously differentiable in a neighborhood of the point $(0,0)$. Let us denote

$$\gamma_{01} = \mathrm{E}\,\tfrac{1}{S}\,\psi\left(\tfrac{e_1}{S}\right)\,\psi'\left(\tfrac{e_1}{S}\right), \qquad \gamma_{01e} = \mathrm{E}\,\tfrac{e_1}{S}\,\psi\left(\tfrac{e_1}{S}\right)\,\psi'\left(\tfrac{e_1}{S}\right). \tag{10}$$

Scale The scale estimator S_n is $\sqrt{n}$-consistent, that is

$$\sqrt{n}(\tfrac{S_n}{S} - 1) = O_p(1) \tag{11}$$

The following conditions are for the fixed design. When we have the model with random covariates, all these assumptions are needed to hold in probability.

X.1 There exists a positive definite matrix $\mathbf{V}$ such that

$$\mathbf{V} = \lim_{n\to\infty} \mathbf{V}_n = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \, \mathbf{x}_i^{\mathsf{T}}.$$

X.2

$$\lim_{n\to\infty} \max_{1\le i\le n} \frac{|\mathbf{x}_i|_2^2}{\sqrt{n}} = 0, \qquad \text{where } |\cdot|_2 \text{ stands for Euclidean norm}$$

X.3

$$\frac{1}{n} \sum_{i=1}^{n} |\mathbf{x}_i|_2^4 = O(1)$$

With the help of the assumptions **S.1-2, Scale** and **X.1** we can show (see [4]) that there exists a root $\hat{\boldsymbol{\beta}}_M$ of the system (2) such that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}) = O_p(1). \tag{12}$$

This fact can be further used to find the asymptotic representation (3) of the M-estimator $\hat{\boldsymbol{\beta}}_M$.

With the help of the $\sqrt{n}$-root consistency of $\hat{\boldsymbol{\beta}}_M$ and S_n and the assumption **X.2** it is not difficult to justify the usage of $\hat{\sigma}^2_\psi$ from (5) as the estimator of σ^2_ψ and to show that

$$\hat{\sigma}^2_\psi = \sigma^2_\psi + o_p(1).$$

Finally, as many of our further expressions simplifies considerably in a symmetric setup, we will state this assumption explicitly for the sake of future reference.

Sym The function ψ is antisymmetric and the distribution function of the errors F is symmetric, that is

$$\psi(-x) = -\psi(x), \qquad \text{and} \qquad F(x) = 1 - F(-x), \qquad x \in \mathbb{R}.$$

2.3 Asymptotic properties of type II conf. interval

Next, to simplify the notation put

$$a_F = \frac{\sqrt{\omega_{pp}}\, \sigma_\psi \, z_\alpha}{\gamma_1}. \tag{13}$$

Now, it is quite straightforward to prove the following theorem.

Theorem 2.1. *If the conditions* ***S.1-2, Scale*** *and* ***X.1-2*** *then the confidence interval* D_n^{II} *defined by* (8) *and* (9) *satisfies:*

1. $$P(D_n^{II} \ni \beta_p) \xrightarrow[n\to\infty]{} 1-\alpha$$

2. $$\sqrt{n}(\hat{b}_p^{+} - \hat{b}_p^{-}) = 2\,a_F + o_p(1)$$

Before we will proceed with a finer analysis of the length of the confidence interval, we need to find the asymptotic distribution of the random variable $\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)$. With the help of the assumption **S.6** we can derive the representation

$$\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi) = \tfrac{1}{2\sigma_\psi\sqrt{n}} \sum_{i=1}^{n} \left[\psi^2(\tfrac{e_i}{S}) - \sigma_\psi^2\right] - \tfrac{2\gamma_{01}}{2\sigma_\psi\gamma_1\sqrt{n}} \sum_{i=1}^{n} \psi(\tfrac{e_i}{S}) - \tfrac{\gamma_{01e}}{\sigma_\psi}\sqrt{n}(\tfrac{S_n}{S} - 1) + o_p(1). \quad (14)$$

Notice that if the distribution of the errors satisfies the symmetry condition **Sym**, then $\gamma_{01} = 0$ and the second term on the right hand side of (14) vanishes.

Now we are ready to formulate the main result about the asymptotic distribution of the standardized length of the type II confidence interval, which is defined as

$$L_n^{II} = \frac{\sqrt{n}\left[\sqrt{n}(\hat{b}_p^{+} - \hat{b}_p^{-}) - 2\,a_F\right]}{2\,a_F}.$$

Theorem 2.2. *If the conditions* ***X.1-3, S.1-6*** *and* ***Scale*** *are satisfied, then the random variable* L_n^{II} *admits the first order asymptotic representation*

$$L_n^{II} = -\tfrac{1}{\gamma_1\sqrt{n}} \sum_{i=1}^{n} \tfrac{x_{ip}^2}{T_{np}^2}\left[\psi'(\tfrac{e_i}{S}) - \gamma_1\right] + \tfrac{\gamma_1+\gamma_{2e}}{\gamma_1}\sqrt{n}\left(\tfrac{S_n}{S} - 1\right) + \frac{\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)}{\sigma_\psi} + \tfrac{\gamma_2}{\gamma_1}\sqrt{n}(\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})^\mathsf{T} \sum_{i=1}^{n} \tfrac{x_{ip}^2\,\mathbf{x}_i}{n\,T_{np}^2} + o_p(1), \quad (15)$$

where γ_2 *is the derivative of the function* $d(t) = \mathrm{E}\,\frac{1}{S}\,\psi'(\frac{e_1+t}{S})$ *at the point zero.*

The proof of this theorem is rather technical. The crucial point is the investigation of the centered process $\{M_n(\mathbf{t},u) - \mathrm{E}\,M_n(\mathbf{t},u), |\mathbf{t}|_2 \le C, |u| \le C\}$, where C is an arbitrarily large but fixed constant and

$$M_n(\mathbf{t},u) = \sum_{i=1}^{n} x_{ip}\left[\psi\left(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\right) - \psi(e_i/S)\right].$$

It turns out that this process may be approximated with a simple process linear in its parameters $\mathbf{t}$, u. These results are sometimes called as the second order asymptotic linearity.

From the expansion (15) we see that the variance of the standardized length L_n is rather complicated. But if the symmetry condition **Sym** is met then both of the functionals γ_2 and γ_{01} are zero and using the representation (14) we get

$$L_n^{II} = -\frac{1}{\gamma_1\sqrt{n}}\sum_{i=1}^{n}\tfrac{x_{ip}^2}{T_{np}^2}\left[\psi'(\tfrac{e_i}{S})-\gamma_1\right]+\left(\tfrac{\gamma_1+\gamma_{2e}}{\gamma_1}-\tfrac{\gamma_{01e}}{\sigma_\psi^2}\right)\sqrt{n}\left(\tfrac{S_n}{S}-1\right)$$
$$+\frac{1}{2\,\sigma_\psi^2\sqrt{n}}\sum_{i=1}^{n}\left[\psi(\tfrac{e_i}{S})^2-\sigma_\psi^2\right]+o_p(1).$$

3 Comparison with the type I confidence interval

Recall that the confidence interval of type I for β_p is

$$D_n^I = [\hat{b}'^{-}_p, \hat{b}'^{+}_p] = \left[\hat{b}_p - \tfrac{z_\alpha}{\sqrt{n}}\tfrac{\hat{\sigma}_\psi\sqrt{\omega_{pp}}}{\hat{\gamma}_1},\ \hat{b}_p + \tfrac{z_\alpha}{\sqrt{n}}\tfrac{\hat{\sigma}_\psi\sqrt{\omega_{pp}}}{\hat{\gamma}_1}\right].$$

If the estimators $\hat{\gamma}_1$ and $\hat{\sigma}_\psi$ are (weakly) consistent estimators of γ_1 and σ_ψ, we can easily see that $P(D_n^I \ni \beta_p) \to 1-\alpha$ and $\sqrt{n}[\hat{b}'^{+}_p - \hat{b}'^{-}_p] = \frac{2z_\alpha\sigma_\psi\sqrt{\omega_{pp}}}{\gamma_1} + o_p(1)$. Moreover it is not difficult to show that the standardized length of the type I CI satisfies

$$L_n^I = \frac{\sqrt{n}\left[\sqrt{n}(\hat{b}'^{+}_p - \hat{b}'^{-}_p) - 2\,a_F\right]}{2\,a_F} = -\tfrac{1}{\gamma_1\sqrt{n}}\sum_{i=1}^{n}\left[\psi'(\tfrac{e_i}{S})-\gamma_1\right]$$
$$+\tfrac{\gamma_1+\gamma_{2e}}{\gamma_1}\sqrt{n}\left(\tfrac{S_n}{S}-1\right)+\frac{\sqrt{n}(\hat{\sigma}_\psi-\sigma_\psi)}{\sigma_\psi}+\tfrac{\gamma_2}{\gamma_1^2\sqrt{n}}\sum_{i=1}^{n}\psi(\tfrac{e_i}{S})+o_p(1). \quad (16)$$

In comparison with the type II CI (15) we see that for the symmetric situation ($\gamma_2 = 0$), the main difference is in the first term $-\frac{1}{\gamma_1\sqrt{n}}\sum_{i=1}^n[\psi'(\frac{e_i}{S})-\gamma_1]$ for the type I CI vs. $\frac{1}{\gamma_1\sqrt{n}}\sum_{i=1}^n\frac{x_{ip}^2}{T_{np}^2}[\psi'(\frac{e_i}{S})-\gamma_1]$ for the type II CI. As the Cauchy-Schwartz inequality tells us that $\sum_{i=1}^n\frac{x_{ip}^4}{n} \geq \left[\sum_{i=1}^n\frac{x_{ip}^2}{n}\right]^2$ we can deduce, that the type I CI will be usually more stable (the length has a smaller asymptotic variance). The reason for this is that the type II CI implicitly uses $\hat{\gamma}_1' = \frac{1}{n\,S_n\,T_{np}^2}\sum_{i=1}^n x_{ip}^2\,\psi'(\frac{r_i}{S})$ as the estimator of γ_1. This can be seen from the formal expansion of $M_n(\hat{b}_p^+)$ or $M_n(\hat{b}_p^-)$ around the point $\hat{b}_p$. But the estimator $\hat{\gamma}_1'$ is generally (if the errors are iid) more variable then the simple estimator $\hat{\gamma}_1 = \frac{1}{n}\sum_{i=1}^n\psi'(r_i)$. On the other hand, our experience is that the estimator $\hat{\gamma}_1'$ is usually able to prevent the total failure of the CI in the case of heteroskedasticity. Some partial numerical evidence for these findings are to be found in the last section.

Finally note that if we use $\hat{\gamma}_1'$ as the estimator of γ_1 for the type I CI, then both L_n^I and L_n^{II} have the same asymptotic expansion and we get that $n(\hat{b}_p^+ - \hat{b}_p^-) = n(\hat{b'}_p^+ - \hat{b'}_p^-) + o_p(1)$. But even in this case it does not generally hold that $n(\hat{b}_p^+ - \hat{b'}_p^+) = o_p(1)$ or $n(\hat{b}_p^- - \hat{b'}_p^-) = o_p(1)$. This can be seen from a further more delicate analysis which shows that

$$\hat{b}_p^- = \hat{b}_p - \tfrac{\sqrt{\omega_{pp}}\,\hat{\sigma}_\psi\, z_\alpha}{\hat{\gamma}_1\sqrt{n}} + \tfrac{1}{n}\,\tfrac{\gamma_2\,\omega_{pp}\,\sigma_\psi^2\, z_\alpha^2}{\gamma_1^3\, T_{np}^2} \sum_{i=1}^n \tfrac{x_{ip}^3}{n} + o_p(\tfrac{1}{n}). \tag{17}$$

If we compare (17) with the lower bound of the type I confidence interval

$$\hat{b'}_p^- = \hat{b}_p - \tfrac{\sqrt{\omega_{pp}}\,\hat{\sigma}_\psi\, z_\alpha}{\hat{\gamma}_1\sqrt{n}},$$

we see that $\sqrt{n}(\hat{b'}_p^- - \hat{b}_p^-) = o_P(1)$ but not $n(\hat{b'}_p^- - \hat{b}_p^-) = o_P(1)$ unless $\gamma_2 = 0$ or $\sum_{i=1}^n \frac{x_{ip}^3}{n} = 0$. As an analogy of (17) holds for $\hat{b}_p^+$ as well, we see that the confidence interval D_n^{II} is asymptotically shifted a little bit to the right or left according to the sign of the quantity $\frac{\gamma_2}{\gamma_1^3 n}\sum_{i=1}^n x_{ip}^3$.

4 Monte Carlo Experiment

In our small simulation, we considered the simple linear model with one covariate and an intercept, that is $Y_i = \beta_1 + \beta_2\, x_i$. We compared the confidence intervals for β_2 based based on the three different methods of estimation of γ_1, and σ_ψ. For simplicity of notation, we assumed that the covariate is centered, that is $\sum_{i=1}^n x_i = 0$.

1. (**hom**) Estimates for homoskedasticity

$$\hat{\gamma}_1 = \frac{1}{n}\sum_{i=1}^n \psi'(\tfrac{r_i}{S_n}), \qquad \hat{\sigma}_\psi^2 = \frac{1+\kappa}{n-p}\sum_{i=1}^n \psi^2(\tfrac{r_i}{S_n}),$$

   where $\kappa = \frac{2}{\hat{\gamma}_1^2 n^2}\sum_{i=1}^n [\psi'(\frac{r_i}{S_n}) - \hat{\gamma}_1]^2$ is the finite sample correction suggested by Huber [3]

2. (**het**) Estimates more robust against heteroskedasticity ($T_n^2 = \sum_{i=1}^n \frac{x_i^2}{n}$)

$$\hat{\gamma}_1 = \frac{1}{nT_n^2}\sum_{i=1}^n x_i^2\psi'(\tfrac{r_i}{S_n}), \qquad \hat{\sigma}_\psi^2 = \frac{1}{(n-p)T_n^2}\sum_{i=1}^n x_i^2\psi^2(\tfrac{r_i}{S_n}),$$

3. (**het2**) It is well known (e. g. in Hansen [2]) that if we can neglect the influence of the estimation of scale then in the presence of heteroskedasticity, the variance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})$ is consistently estimated by $S_n^2\,\left(\boldsymbol{\Gamma}_n^{-1}\,\boldsymbol{\Sigma}_n\,\boldsymbol{\Gamma}_n^{-1}\right)$, where

$$\boldsymbol{\Gamma}_n = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \psi'(\frac{r_i}{S_n}) & \frac{1}{n}\sum_{i=1}^n x_i\,\psi'(\frac{r_i}{S_n}) \\ \frac{1}{n}\sum_{i=1}^n x_i\,\psi'(\frac{r_i}{S_n}) & \frac{1}{n}\sum_{i=1}^n x_i^2\,\psi'(\frac{r_i}{S_n}) \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_n = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \psi(\frac{r_i}{S_n}) & \frac{1}{n}\sum_{i=1}^n x_i\,\psi(\frac{r_i}{S_n}) \\ \frac{1}{n}\sum_{i=1}^n x_i\,\psi(\frac{r_i}{S_n}) & \frac{1}{n}\sum_{i=1}^n x_i^2\,\psi(\frac{r_i}{S_n}) \end{bmatrix}.$$

Let us denote the standard deviation of $\sqrt{n}(\hat{b}_2 - \beta_2)$ based on this estimates by s_b. Then the type I CI is simply $[\hat{b}_2 \mp s_b\,u_{1-\alpha/2}]$ and the type II confidence interval is defined by (8) and (9) with the quantity $T_{np}^2\,\sqrt{\omega_{pp}}\,\hat{\sigma}_\psi\,z_\alpha$ replaced by $g_{22}\,s_b\,u_{1-\alpha/2}$, where g_{22} is the 'last' element of the matrix $\boldsymbol{\Gamma}_n$.

In the following, we took the sample size $n = 20$ and prescribed $\alpha = 0.05$. As the distribution of the errors we chose the contaminated normal model given by the distribution function $F_{kont}(y) = 0.85\,\Phi(y) + 0.15\,\Phi(y-3)$, where $\Phi(y)$ is a distribution function of a standard normal random variable. But we got very similar results for errors following other type of distributions including the standard normal distribution, t-distribution, logistic and cauchy distribution. Because we do not want our results to be dependent on the particular design, we decided to work with a random covariate (correlation model). The covariate was always centered before a calculation of a M-estimator.

As the M-estimator we chose the Huber estimate with the psi function $\psi(x) = \min\{\max\{x, -k\}, k\}$, where the tuning constant k was set to 1.345. For estimation of scale we use the MAD (mean absolute deviation) computed from the residuals from the preliminary l_1 (least absolute deviation) fit. As l_1-regression has the exact fit property, that is p points are fitted exactly with zero residuals, these residuals were left out before computation of MAD.

To be more conservative we also replaced the quantiles of standard normal distribution with the quantiles of t-distribution with $(n-1)$ degrees of freedom in the formulas of CI's . By ℓ_n we denote the length of a particular CI. The number of random samples in all situations is 250 000.

Table 1 contains the result for the covariate X with uniform distribution on $[-1, 1]$. In the first line we see the actual coverage of the confidence interval, in the second line the median of the length of the CI (multiplied by $\sqrt{n}$) and finally the interquantile range of the lengths of conf. intervals (multiplied by n). Table 1 tells us that all the proposed methods work in this situation satisfactorily. The winner is the type I-**hom** CI. The coverage is OK, the length is shorter then the type II-**hom** CI and the stability (in terms of the variability of the length) is the best.

Table 2 contains the result for the similar situation as Table 1, but we have changed the distribution of the covariate X to exponential (with the density $f(x) =$

	I-hom	II-**hom**	I-**het**	II-**het**	I-**het2**	II-**het2**
Coverage	0.947	0.953	0.930	0.937	0.936	0.943
$\sqrt{n}\,\text{med}(\ell_n)$	10.38	10.62	9.79	10.09	10.00	10.33
$n\,\text{IQR}(\ell_n)$	12.03	12.74	13.78	14.64	14.24	15.19

Table 1: $n = 20$, $X \sim \text{U}[-1, 1]$ and $e \sim F_{kont}$

$e^{-x}\mathbb{I}(x > 0)$), which is strongly asymmetric. We have also added two rows concerning the one-sided coverage properties. By Covarage Lower (or Coverage Upper) we mean $\text{P}(\hat{\beta}_2^- < \beta_2)$ (or $\text{P}(\hat{\beta}_2^+ < \beta_2)$). At first we notice that the confidence intervals designed for heteroskedasticity are not doing a very good job. Some further simulations show that the actual coverage for this method does not approach 0.94 for the sample less than 200. The second view at Table 2 tells us that the actual coverage for the type **I-hom** CI is a little bit under 0.95. It is interesting that this is caused by the poor lower coverage.

	I-hom	II-**hom**	I-**het**	II-**het**	I-**het2**	II-**het2**
Coverage	0.943	0.956	0.884	0.898	0.892	0.906
Coverage Lower	0.961	0.974	0.942	0.952	0.944	0.955
Coverage Upper	0.982	0.982	0.942	0.946	0.948	0.951
$\sqrt{n}\,\text{med}(\ell_n)$	6.40	6.87	5.36	5.79	5.47	5.93
$n\,\text{IQR}(\ell_n)$	10.88	11.88	12.31	14.27	12.51	14.55

Table 2: $n = 20$, $X \sim \text{Exp}(1)$ and $e \sim F_{kont}$

Now we turn our attention to the models where heteroskedasticity is present. The covariate X will be once more uniformly distributed on $[-1, 1]$ and the errors of the i-th observation will be e_i multiplied by the absolute value of X_i. The results are to be found in Table 3. We see that if we are not aware of heteroskedasticity, the resulting conf. interval is too short with poor actual coverage. It is worth noting that for the type II-**hom** CI the situation is not so catastrophic. Notice that type II-**het2** CI almost achieves prescribed size. With increasing the sample size the actual coverage of the type **I-hom** CI is about 0.8, for the type **I-hom** CI about 0.9 and the coverage for all the other confidence intervals converges quite rapidly to 0.95.

Summary of the results

Let us summarize our empirical findings for the linear model with one explanatory variable.

CI's of type II are generally larger, more variable but with higher coverage than the type I conf. intervals. The higher coverage property is worth considering espe-

	I-hom	**II-hom**	**I-het**	**II-het**	**I-het2**	**II-het2**
Coverage	0.829	0.900	0.911	0.931	0.924	0.944
$\sqrt{n}\,\mathrm{med}(\ell_n)$	4.84	5.86	6.30	6.83	6.60	7.23
$n\,\mathrm{IQR}(\ell_n)$	6.29	8.20	10.17	11.14	10.86	12.11

Table 3: $n = 20$, $X \sim \mathrm{U}[-1, 1]$ and $e/|X| \sim F_{kont}$

cially in models in which errors or a covariate is heavily asymmetric. In such models type II methods are preferable for the construction of one-sided conf. intervals. At the same time the type II-**hom** CI usually prevents the total failure of the CI in the case of heteroskedasticity.

The CI's which take possible heteroskedasticity into considerations need in asymmetric models more than one hundred observations to be trustworthy.

If the number of observations is sufficiently large (> 200) and we would like to be robust against heteroskedasticity, the simple type II-**het** CI is worth considering.

It is behind the scope of this paper, but it would be desirable to verify to which extent these empirical findings hold for a multiple regression with two or more explanatory variables.

References

[1] D. D. Boss. A new method for constructing approximate confidence intervals from M-estimates. *J. Amer. Statist. Assoc.*, 75:142–145, 1980.

[2] L. Hansen. Large sample properties for generalized method of moments estimators *Econometrica*, 50:1029–1054, 1982.

[3] P. J. Huber. Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.*, 1:799–821, 1973.

[4] J. Jurečková and P. K. Sen. *Robust Statistical Procedures: Asymptotics and Interrelations*. Wiley, New York, 1996.

[5] C. J. Lloyd. Approximate pivots from M-estimators. *Statistica Sinica*, 4:701–714, 1994.

[6] M. Omelka, *Asymptotic properties of estimators and tests*. PhD thesis, Charles University in Prague. In Preparation, 2006.

Marek Omelka: Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Statistics, Sokolovská 83, Prague, 186 75, Czech Republic, omelka@karlin.mff.cuni.cz