Charles University in Prague, Faculty of Mathematics and Physics

Department of Probability and Mathematical Statistics

DOCTORAL THESIS

Second order properties of some M-estimators and R-estimators

Marek Omelka

June 2006

Advisor: Prof. RNDr. Jana Jurečková, DrSc.

Univerzita Karlova v Praze, Matematicko-fyzikální fakulta
Katedra pravděpodobnosti a matematické statistiky

Asymptotické vlastnosti druhého řádu některých M-odhadů a R-odhadů

Disertační práce

Marek Omelka

Obor: m4 - pravděpodobnost a matematická statistika
Školitel: Prof. RNDr. Jana Jurečková, DrSc.

Acknowledgements

First of all, I would like to express my thanks to my advisor Prof. Jana Jurečková for her encouragement, guidance and support.

I owe much more than I can express here to my parents for their love, care and patience. I could always felt that they believed in me.

Many people helped me in different ways. In particular, I would like to thank David Kraus, who was always very willing to help me with computer staff, my teaching responsibilities, and many other things, and who listened very patiently to my psalms about this thesis.

Financial support was provided by the Charles University Grant GAUK 341/2005 and by the Czech Republic Grant GAČR 201/05/H007. Finally, I am very thankful to Prof. Rudy Beran, who enabled me to spend October 2005 at the University of Davis. This helped me a lot to start to write the thesis.

Prague, June 2006

I am very grateful to both of my referees – Doc. Petr Lachout and Doc. Jan Picek, whose remarks and comments led to a considerable improvement of the text.

Prague, September 2006

Contents

Chapter 1

Introduction

Statistical inferences could not be based solely upon the observations. To be able to answer the questions of scientific interest, statisticians have to assume, that there is a random mechanism producing data. This random mechanism is usually called a model. Obviously, the more we know about the model, the more efficiently we can handle data. There exists huge literature about optimal procedures for models, which are specified in a great detail.

Nevertheless, the research initiated by Peter J. Huber showed that even very tiny departures from model assumptions may have a dramatic effect on optimality of procedures. This initiated a highly dynamic growth of research literature on statistical methods, which are not so sensitive to the departures from a model, but remain efficient if this model holds. Roughly speaking, we can distinguish robust methods and nonparametric methods, which present very broad and lively parts of mathematical statistics.

Although the range of statistical methods is very diverse these days, the basic problem remains estimation in location and regression problems. Robust and nonparametric statistics offer us basically three families of estimators – M-estimators, L-estimators and R-estimators.

The aim of this thesis is to give some further insights into the asymptotic properties of the M-estimators (of location and regression) as well as an R-estimator based on Wilcoxon scores, which belongs to the most popular R-estimators.

1.1 M-estimators

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

$$Y_i = \beta_1 \, x_{i1} + \ldots + \beta_p \, x_{ip} + e_i = \boldsymbol{\beta}^\mathsf{T} \mathbf{x}_i + e_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

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 rows of a known matrix $\mathbf{X}_n$, and $e_1, \ldots, e_n$ are independent, identically distributed random variables with an unknown cumulative distribution function (cdf) F.

Then given an absolutely continuous function ρ with a derivative ψ, we define a fixed scale (studentized) M-estimator $\hat{\boldsymbol{\beta}}_n$ of the parameter $\boldsymbol{\beta}$ as the solution of the minimization

$$\rho\left(Y_i - \mathbf{t}^\mathsf{T}\mathbf{x}_i\right) := \min, \qquad \left(\text{or } \rho\left(\tfrac{Y_i - \mathbf{t}^\mathsf{T}\mathbf{x}_i}{S_n}\right) := \min\right),$$

where S_n is an appropriate scale estimator.

If the function $\psi = \rho'$ is continuous, then the estimator $\hat{\boldsymbol{\beta}}_n$ is a solution of the system of equations

$$\sum_{i=1}^{n} \mathbf{x}_i \, \psi(Y_i - b^\mathsf{T}\mathbf{x}_i) = \mathbf{0} \qquad \left(\text{or } \sum_{i=1}^{n} \mathbf{x}_i \, \psi(\tfrac{Y_i - b^\mathsf{T}\mathbf{x}_i}{S_n}) = \mathbf{0}\right). \tag{1.2}$$

As the defining equation (1.2) gives us more flexibility in tuning properties of M-estimators by a choice of a function ψ, $\hat{\boldsymbol{\beta}}_n$ is usually defined as a carefully chosen root of (1.2). For simplicity, we will focus on an M-estimator with a fixed scale for this moment.

As the M-estimator is defined implicitly, it is not obvious how to make statistical inference based on it. A very elegant approach to the investigation of asymptotic properties of $\hat{\boldsymbol{\beta}}_n$ (consistency, asymptotic normality) is based on the 'uniform linearity results'. This approach studies the (vector) process

$$\mathbf{T}_n(\mathbf{t}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \left[\psi(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - \psi(e_i)\right] + \gamma_1 \sqrt{n}\, \mathbf{V}_n \mathbf{t}, \qquad \mathbf{t} \in T = \{\mathbf{s} \in \mathbb{R}^p : |\mathbf{s}|_2 \le M\}, \tag{1.3}$$

where $\mathbf{V}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\mathsf{T}$, $\gamma_1 = \mathsf{E}\,\psi'(e_1)$, M is an arbitrarily large but fixed constant, and $|\cdot|_2$ stands for the euclidean norm.

It is well known that under some mild conditions $\sup_{\mathbf{t}\in T} |\mathbf{T}_n(\mathbf{t})| = o_p(1)$. This result, sometimes called the first order asymptotic linearity (FOAL), is the main tool in proving the first order asymptotic representation of the estimator $\hat{\boldsymbol{\beta}}_n$, that is

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \, \psi(e_i) + \mathbf{R}_n, \tag{1.4}$$

where the remainder term $\mathbf{R}_n$ is of order $o_p(1)$. It turns out that to investigate the remainder term $\mathbf{R}_n$ more carefully, a more delicate analysis of the process $\mathbf{T}_n$ defined in (1.3) is needed. Jurečková and Sen (1989b) proved that if the function ψ and the distribution of errors F are sufficiently smooth, then $\sup_{\mathbf{t}\in T} |T_n(\mathbf{t})| = O_p(\frac{1}{\sqrt{n}})$, which further gives us that the remainder term $\mathbf{R}_n$ is of the same order. The asymptotic distribution of the random variable $\sqrt{n}\, R_n$ was studied by Boos (1977) for the special case of a location model and by Jurečková and Sen (1990) for a general M-estimator of a scalar parameter. Arcones and Mason (1997) generalize these results to multivariate M-estimators.

Our thesis extends these results in the following way. If ψ and the underlying distributions of the errors are sufficiently smooth, then we find a simple process $\mathbf{P}_n$ (with a limiting gaussian distribution) linear in the parameter $\mathbf{t}$ such that

$$\sup_{\mathbf{t}\in T} |\sqrt{n}\; \mathbf{T}_n(\mathbf{t}) - \mathbf{P}_n(\mathbf{t})| = o_p(1).$$

This gives us not only the asymptotic distribution of the remainder term $n^{1/2}\mathbf{R}_n$ of (1.4), but we even derive a two-term von Mises expansion for the M-estimator $\hat{\boldsymbol{\beta}}_n$. This expansion enables us to compare the M-estimator with its one-step approximations or with some other estimators which are first order equivalent to the chosen M-estimator. Our results include studentized M-estimators as well.

The situation is qualitatively different if ψ is a step function. In this case we are only able to find the asymptotic distribution of the remainder term $n^{1/4}\mathbf{R}_n$. This extends the results of Jurečková and Sen (1989a), who dealt with the (unstudentized) location case, as well as the results of Knight (1997), who consider the special case of quantile regression.

1.2 R-estimators

Consider the linear regression model (1.1). Let $R_i(\mathbf{b})$ be the rank of $Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i$ among $Y_1 - \mathbf{b}^\mathsf{T}\mathbf{x}_1, \ldots, Y_n - \mathbf{b}^\mathsf{T}\mathbf{x}_n$ and $\bar{\mathbf{x}}_n = (\bar{x}_{n1}, \ldots, \bar{x}_{np})^\mathsf{T}$ be the vector of the column means of the design matrix $\mathbf{X}$. Further let $a_n(i), i = 1, \ldots, n$ be a nondecreasing set of scores, satisfying

$$\sum_{i=1}^{n} a_n(i) = 0.$$

These scores are usually generated as $a_n(i) = \phi(\frac{i}{n+1})$, or $a_n(i) = \mathsf{E}\, \phi(U_{n:i})$, where ϕ is a nondecreasing function defined on $(0,1)$ and $U_{n:i}$ is the i-th order statistic from a sample of n independent random variables uniformly distributed on $(0,1)$.

The R-estimator $\hat{\boldsymbol{\beta}}_n$ is usually defined as the argument of the minimum of the 'Jaeckel' measure of dispersion (Jaeckel (1972)), that is

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\mathbf{b}\in\mathbb{R}_p} D_n(\mathbf{b}), \quad \text{where} \quad D_n(\mathbf{b}) = \sum_{i=1}^{n} (Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i) a_n\left(R_i(\mathbf{b})\right). \tag{1.5}$$

Jaeckel (1972) showed that the function $D_n(\mathbf{b})$ is nonnegative, continuous and convex in $\mathbf{b} \in \mathbb{R}_p$. The convexity ensures that $D_n(\mathbf{b})$ is differentiable in $\mathbf{b}$ almost everywhere with the derivative

$$\frac{\partial D_n(\mathbf{b})}{\partial \mathbf{b}} = -\sum_{i=1}^{n} \mathbf{x}_i\, a_n(R_i(\mathbf{b})).$$

Thus $\boldsymbol{\beta}$ may be defined as the solution of the following minimization (see Jurečková (1971))

$$\sum_{j=1}^{p} |S_{nj}(\mathbf{b})| := \min, \qquad \text{where} \quad S_{nj}(\mathbf{b}) = \frac{1}{n^{3/2}} \sum_{i=1}^{n} x_{ij}\, a_n(R_i(\mathbf{b})). \tag{1.6}$$

It can be shown that both definitions are asymptotically equivalent (Jaeckel (1972), or Jurečková and Sen (1996)). For our purposes it will be more convenient to work with the definition (1.5).

In fact, we will only consider the case of Wilcoxon scores, that is $a_n(i) = \frac{i}{n+1} - \frac{1}{2}$. The reason is that these scores give the resulting R-estimator a relatively simple structure. On the other hand this 'Wilcoxon type' R-estimators belong to the most widely used R-estimators, see e.g. McKean (2004) or Terpstra and McKean (2004).

From any of the definitions of R-estimators we see that, informally speaking, a regression estimator 'almost' solves the following system of equations

$$\mathbf{S}_n(\mathbf{b}) = (S_{n1}(\mathbf{b}), \ldots, S_{np}(\mathbf{b}))^\mathsf{T} = \frac{1}{n^{3/2}} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}_n) R_i(\mathbf{b}) \stackrel{!}{=} \mathbf{0}. \tag{1.7}$$

Let c_{in}, $1 \leq i \leq n$, $n = 1, 2, \ldots$ be a triangular array of constants satisfying some conditions which will be specified later. In view of the 'normal equations' (1.7), and similarly to M-estimation, it is not surprising that to investigate the asymptotic properties of $\hat{\boldsymbol{\beta}}_n$, it turns out to be useful to study the asymptotic behaviour of the processes

$$\tilde{S}_n(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^{n} c_{in}\, R_i^{'}(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^{n} c_{in} \sum_{j=1}^{n} \mathbb{I}\{e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} \geq e_j - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_j}{\sqrt{n}}\}, \tag{1.8}$$

$$T_n(\mathbf{t}) = \tilde{S}_n(\mathbf{t}) - \tilde{S}_n(\mathbf{0}), \tag{1.9}$$

$$\bar{T}_n(\mathbf{t}) = T_n(\mathbf{t}) - \mathsf{E}\, T_n(\mathbf{t}), \tag{1.10}$$

where $\mathbf{t} = (t_1, \ldots, t_p)^\mathsf{T}$ and $R_i^{'}(\mathbf{t})$ stands for the rank of $e_i - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}$ among $e_1 - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_1}{\sqrt{n}}, \ldots, e_n - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_n}{\sqrt{n}}$. Similarly to the previous section, we will index these processes by the set $T = \{\mathbf{s} \in \mathbb{R}^p : |\mathbf{s}|_2 \leq M\}$, where $|\cdot|_2$ stands for the euclidean norm, and M is an arbitrarily large but fixed constant.

The process $S_{nj}(\mathbf{b})$ of the equation (1.6) and the process $\tilde{S}_n(\mathbf{t})$ of (1.8) are connected through the equation

$$S_{nj}(\mathbf{b}) = \tfrac{1}{\sqrt{n}}\, \tilde{S}_n(\sqrt{n}(\mathbf{b} - \boldsymbol{\beta})), \qquad \text{where} \qquad c_{in} = x_{ij} - \bar{x}_j.$$

The standard result, which is usually called the first order asymptotic uniform linearity, states that

$$\sup_{\mathbf{t} \in T} \frac{1}{\sqrt{n}} \left| T_n(\mathbf{t}) - \tfrac{\gamma\, \mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^{n} c_i\, \mathbf{x}_i \right| = o_p(1), \qquad \text{where} \qquad \gamma = \mathsf{E}\; f(e_1) = \int f^2(x)\, dx. \tag{1.11}$$

The results of this type proved to be very useful as they made possible an elegant asymptotic approach to statistical inference based on R-estimators. The research in this area was initiated by Jurečková (1971). For an overview of results see Jurečková and Sen (1996) and Koul (2002).

For the case of a simple regression (a one-dimensional parameter β) and Wilcoxon scores Jurečková (1973) showed that if we leave out the scaling factor $1/\sqrt{n}$ in (1.11), we obtain a stochastic process which converges weakly to a linear process. This result was further generalized for the Wilcoxon signed-rank statistics by Antille (1976) and for some other types of score functions by Hušková (1980), Puri and Wu (1985) and Kersting (1987).

In our thesis, we generalize the work of Jurečková (1973) to the case of a multi-dimensional parameter $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\mathsf{T}$. Our approach can be modified for a Wilcoxon-signed rank statistic (Section 3.4) and some other estimators (Section 3.5).

1.3 Thesis outline

The rest of this thesis is organized as follows.

Chapter 2 and Chapter 3 are of technical character. In Chapter 2 we are dealing with the M-processes of type (1.3) and with the 'studentized' processes of the form

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

where $T = \{(\mathbf{t}, u) : |\mathbf{t}|_2 \leq M, |u| \leq M\}$ $(\subset \mathbb{R}^{p+1})$.

Chapter 3 is dealing with the processes of the form (1.8) and (1.10).

The next two chapters, that is Chapter 4 and Chapter 5 are based on the technical results of the previous chapters.

In the first part of Chapter 4 we derive a two-term von Mises expansion for an M-estimator based on a 'smooth' function ψ and for an R-estimator based on Wilcoxon scores. Next we use these results to compare an M-estimator with its one-step approximation. This generalizes the work of Jurečková and Sen (1990) as well as it provides a partial theoretical background for the work of Welsh and Ronchetti (2002). For an M-estimator generated by an 'unsmooth' function ψ, we find the asymptotic distribution of the remainder term in the first order asymptotic representation (1.4).

In the second part of Chapter 4 we propose an alternative way of constructing a confidence interval for a single regression parameter and we investigate its properties. We compare this alternative procedure with the 'traditional' (Wald type) approach. This extends the results of Boos (1980), who proposed this alternative way of constructing confidence intervals for a location problem. In the case of R-estimators we generalize some results of Jurečková (1973).

Chapter 5 is dealing with a sequential problem of a confidence interval of a fixed size. The results of this chapter for M-estimators extend the work of Jurečková and Sen (1981a) and Jurečková and Sen (1981b), where the linear model (1.1) with one explanatory variable (without intercept or studentization) was treated. The results for R-estimators generalize the work of Jurečková (1978) and Hušková (1980) for a special case of Wilcoxon scores.

In Chapter 6 we briefly review the results obtained in this thesis and discuss some further possible extensions of this work.

The appendix in Chapter 7 contains most of the auxiliary results used in the proofs. We present the proofs for those results that could not be found in the literature.

Finally, following References we add a list of symbols and regularity conditions which are used throughout the text.

Chapter 2

SOAL of M-processes

In this chapter we generalize the results of Jurečková and Sen (1989b), Jurečková and Sen (1990), and Knight (1997). The basic building stone of our proofs is 2.11.11 Theorem of van der Vaart and Wellner (1996). The idea of using this theorem for our purposes originated from Knight (1997), who used this theorem to find the second order asymptotic distribution of L_1 regression estimators. For the sake of future reference, we derived several modifications of this theorem, which are to be found in Appendix.

The chapter is divided into two sections depending on the smoothness of the function ψ. Primarily we distinguish two cases:

- ψ is an absolutely continuous function;
- ψ is a step function.

To shorten the terminology, by 'smooth-ψ' we will mean that ψ is absolutely continuous and by 'unsmooth-ψ' we will mean that ψ is a step function.

As the studentization of M-estimators brings in new technical and theoretical difficulties, we will usually treat fixed scale estimators and studentized estimators separately.

2.1 An absolutely continuous ψ

2.1.1 Fixed scale

To motivate the following investigation, let us recall that we are considering the linear model

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

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.

Notations

Let $\mathbf{t} = (t_1, \ldots, t_p)^\mathsf{T}$. We will be interested in the asymptotic behaviour of the processes

$$M_n(\mathbf{t}) = \sum_{i=1}^{n} c_i \left[\psi(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - \psi(e_i) \right], \tag{2.1}$$

$$\bar{M}_n(\mathbf{t}) = M_n(\mathbf{t}) - \mathsf{E}\; M_n(\mathbf{t}), \tag{2.2}$$

with $\mathbf{t} \in T = \{\mathbf{s} \in \mathbb{R}^p : |\mathbf{s}|_2 \leq M\}$, where $|\cdot|_2$ stands for the euclidean norm, and M is an arbitrarily large but fixed constant.

Assumptions

First we formulate assumptions on the design $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and the constants $c_1, \ldots, c_n$. As it is sometimes convenient to allow to vary these quantities with n, we formulate the conditions for the triangular arrays $\mathbf{x}_{1n} \ldots, \mathbf{x}_{nn}$ and $c_{1n}, \ldots, c_{nn}$.

X.1

$$\frac{1}{n}\sum_{i=1}^{n} c_{in}^2 = O(1), \qquad \lim_{n\to\infty} \frac{\max_{1\leq i\leq n} |c_{in}|}{\sqrt{n}} = 0.$$

X.2

$$\frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_{in}|_2^2 = O(1), \qquad \lim_{n\to\infty} \frac{\max_{1\leq i\leq n} |\mathbf{x}_{in}|_2}{\sqrt{n}} = 0.$$

X.3

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \frac{|c_{in}|\,|\mathbf{x}_{in}|_2}{\sqrt{n}} = 0.$$

X.4

$$B_n^2 = \frac{1}{n}\sum_{i=1}^{n} c_{in}^2\, |\mathbf{x}_{in}|_2^2 = O(1).$$

The conditions **X.1-3** are analogous to the conditions in Jurečková (1973). The last condition **X.4** is for our convenience. If $B_n^2 = O(1)$ was not satisfied, we would work with the process $M_n'(\mathbf{t}) = \frac{M_n(\mathbf{t})}{B_n}$. To simplify the notation, we will write $\mathbf{x}_i$ instead of $\mathbf{x}_{in}$ and c_i instead of c_{in}.

Remark 1. Put $\varepsilon_n = \max_{1\leq i\leq n} \frac{M|\mathbf{x}_i|_2}{\sqrt{n}}$. We will often use this simple observation:

$$\max_{1\leq i\leq n} \sup_{\mathbf{t}\in T} \left| \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} \right| \leq \varepsilon_n = \max_{1\leq i\leq n} \frac{M\,|\mathbf{x}_i|_2}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathbf{X.2}} 0. \tag{2.3}$$

Later, we will substitute x_{ij} $(j = 1, \ldots, p)$ for c_i to find the second order asymptotic distribution of the regression M-estimator $\hat{\boldsymbol{\beta}}_n$. Taking $c_{in} = |\mathbf{x}_{in}|_2$ we get the following requirements on the design:

XX.1
$$\frac{1}{n}\sum_{i=1}^{n}|\mathbf{x}_{in}|_2^4 = O(1), \qquad \text{and} \qquad \lim_{n\to\infty}\frac{\max_{1\le i\le n}|\mathbf{x}_{in}|_2}{\sqrt{n}} = 0. \tag{2.4}$$

In the next, we will simultaneously impose conditions on the distribution function F and the function ψ. The abbreviation **SmFx** stands for 'a **Sm**ooth ψ and a **Fix**ed scale'.

SmFx.1 ψ is absolutely continuous with a derivative ψ' such that $\mathsf{E}\ \psi'(e_1)^2 < \infty$.

SmFx.2 The function $\psi'(e_1 + t)$ is continuous in the quadratic mean at the point 0, that is

$$\lim_{t\to 0}\mathsf{E}\,[\psi'(e_1+t) - \psi'(e_1)]^2 = 0.$$

SmFx.3 There exists a continuous second derivative of the function $\lambda(t) = \mathsf{E}\ \psi(e_1 + t)$ at the point 0.

As we will see later, the first two conditions are used to study the asymptotic behaviour of the process $\bar{M}_n$. The third condition is necessary to approximate $\mathsf{E}\ M_n$ (the mean function of M_n).

If the function ψ is twice differentiable, ψ' and ψ'' are both bounded continuous, then it is easy to verify that conditions **SmFx.2-3** are met. Notice that in this case we do not need to make any assumptions about the distribution of the errors.

An important class of functions which do not possess a smooth second derivative are piecewise linear functions. We will need the following assumptions about this class of functions.

A.1 ψ is a continuous piecewise linear function with the derivative

$$\psi'(x) = \alpha_j, \quad \text{for } r_j < x \le r_{j+1},\ j = 1,\ldots,k,$$

where $\alpha_0, \alpha_1, \ldots, \alpha_k$ are real numbers, (usually $\alpha_0 = \alpha_k = 0$) and $-\infty = r_0 < r_1 < \ldots < r_k < r_{k+1} = \infty$.

A.2 The cumulative distribution function F is continuous at the points $r_1, \ldots, r_k$.

A.3 The cumulative distribution function F is absolutely continuous with a derivative which is continuous at the points $r_1, \ldots, r_k$.

The condition **A.1** trivially implies **SmFx.1**. Analogously **A.2** ensures **SmFx.2** and **A.3** ensures **SmFx.3**.

Remark 2. Probably the most famous ψ functions are (in alphabetical order)

- Andrews' sine wave $\psi(x) = \sin\left(\frac{\pi}{k}\,x\right)\,\mathbb{I}\{|x| \le k\}$

	SmFx.2	**SmFx.3**
Andrew	F is continuous at $\pm k$	f exists and is continuous at $\pm k$
Hampel	F is continuous at $\pm a$, $\pm b, \pm c$	f exists and is continuous at $\pm a$, $\pm b, \pm c$
Huber	F is continuous at $\pm k$	f exists and is continuous at $\pm k$
Tukey	OK	F is continuous at $\pm k$
$\psi(x) = \frac{x}{1+x^2}$	OK	OK

Table 1: Conditions on the underlying distribution for different ψ functions

- Hampels' function

$$\psi(x) = \begin{cases} x, & 0 \le x \le a \\ a, & a \le x \le b \\ a\left(\frac{c-x}{c-b}\right), & b \le x \le c \\ 0 & x > c \end{cases}$$

  and $\psi(x) = -\psi(-x)$ for $x < 0$.

- Hubers' function $\psi(x) = \max\{\min\{x, k\}, -k\}$

- Tukeys' biweight $\psi(x) = \frac{x}{k}\left(1 - \frac{x^2}{k^2}\right)^2 \mathbb{I}\{|x| \le k\}$

Let f stand for the derivative of the cumulative distribution function F. Table 1 present the requirements on the underlying distribution of the errors so as the conditions **SmFx.2** and **SmFx.3** are met.

Many of the following results (in particular for studentized M-estimators) simplify significantly under some symmetric assumptions. For the sake of future reference, we state this assumption explicitly.

Sym The distribution of the errors is symmetric and the ψ-function is antisymmetric, that is $F(x) = 1 - F(-x)$ and $\psi(x) = -\psi(-x)$ for all $x \in \mathbb{R}$.

2.1.2 Theorems

For a function $g : T \mapsto \mathbb{R}$ set

$$\|g\|_T = \sup_{\mathbf{s} \in T} |g(\mathbf{s})|.$$

Theorem 2.1. *Under conditions **X.1-4** and **SmFx.1-2** it holds that*

$$\sup_{\mathbf{t} \in T} \left| \bar{M}_n(\mathbf{t}) + \frac{\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^{n} c_i\, \mathbf{x}_i [\psi'(e_i) - \gamma_1] \right| = o_p(1), \tag{2.5}$$

where $\gamma_1 = \mathsf{E}\ \psi'(e_1)$.

Specially, if we put

$$\mathbf{B}_n = \frac{1}{n}\sum_{i=1}^{n} c_i^2\, \mathbf{x}_i\mathbf{x}_i^{\mathsf{T}},$$

and the matrix $\mathbf{B}_n$ *is positive definite for all* n *large enough, then the process* $\bar{M}_n'(\mathbf{t}) = \bar{M}_n(\mathbf{B}_n^{-1/2}\mathbf{t})$ *indexed by the set* T *converges in distribution to a centered Gaussian process* $\{Y(\mathbf{t}), \mathbf{t}\in T\}$ *with the covariance function* $\operatorname{cov}(Y(\mathbf{t}), Y(\mathbf{s})) = \sigma^2\, \mathbf{t}^{\mathsf{T}}\mathbf{s}$*, where* $\sigma^2 = \operatorname{var}\{\psi'(e_1)\}$*.*

Remark 3. One may wonder, whether the quantity on the left-hand side of (2.5) is measurable, as we are taking supremum over an uncountable set. But all the the processes involved in the supremum are continuous in the parameter $\mathbf{t}$ and the index set T is compact, which ensures the measurability of the supremum in (2.5).

Proof. Let us denote

$$Z_{ni}(\mathbf{t}) = c_i\left[\psi(e_i - \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}) - \psi(e_i) + \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}\,\psi'(e_1)\right] \qquad \text{and} \qquad Z_n(\mathbf{t}) = \sum_{i=1}^{n} Z_{ni}(\mathbf{t}).$$

Then our theorem states that

$$\left\|\bar{Z}_n\right\|_T = \left\|Z_n - \mathsf{E}\ Z_n\right\|_T = o_p(1). \tag{2.6}$$

To prove (2.6), we need to check the assumptions of Corollary 7.13. As $Z_{ni}(\mathbf{0}) = 0$ for $i = 1, \ldots, n$, it is sufficient to verify (7.5).

First, we notice that for arbitrary $\mathbf{t}, \mathbf{s} \in T$

$$|Z_{ni}(\mathbf{t}) - Z_{ni}(\mathbf{s})| \le |c_i| \left| \int_{\frac{\mathbf{s}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}}^{\frac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}} \left[\psi'(e_i - v) - \psi'(e_i)\right] dv \right|.$$

Put $\varepsilon_n = \max_{1\le i\le n} \frac{M|\mathbf{x}_i|_2}{\sqrt{n}}$ and calculate

$$\begin{aligned}
\sum_{i=1}^{n} \mathsf{E} \sup_{|\mathbf{t}-\mathbf{s}|_2<\varepsilon} [Z_{ni}(\mathbf{t}) - Z_{ni}(\mathbf{s})]^2 &\le \sum_{i=1}^{n} |c_i|^2\, \mathsf{E} \sup_{|\mathbf{t}-\mathbf{s}|_2<\varepsilon} \left| \int_{\frac{\mathbf{s}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}}^{\frac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}} [\psi'(e_i - v) - \psi'(e_i)] dv \right|^2 \\
&\le \sum_{i=1}^{n} |c_i|^2\, \mathsf{E} \sup_{|\mathbf{t}-\mathbf{s}|_2<\varepsilon} \left| \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}} - \tfrac{\mathbf{s}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}} \right| \int_{\frac{\mathbf{s}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}}^{\frac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}} [\psi'(e_i - v) - \psi'(e_i)]^2 dv \\
&\le \sum_{i=1}^{n} |c_i|^2\, \mathsf{E}\ \tfrac{\varepsilon|\mathbf{x}_i|_2}{\sqrt{n}} \int_{-\frac{M|\mathbf{x}_i|_2}{\sqrt{n}}}^{-\frac{M|\mathbf{x}_i|_2}{\sqrt{n}}} [\psi'(e_i - v) - \psi'(e_i)]^2 dv \\
&\le \frac{2\,\varepsilon\, M}{n} \sum_{i=1}^{n} |c_i|^2 |\mathbf{x}_i|_2^2 \sup_{|v|\le\varepsilon_n} \mathsf{E}\ \left[\psi'(e_1 - v) - \psi'(e_1)\right]^2 \le C\,\varepsilon\, r_n,
\end{aligned}$$

where $r_n = \sup_{|v|\le\varepsilon_n} \mathsf{E}\ [\psi'(e_1 - v) - \psi'(e_1)]$. By the assumptions of the theorem we can take C large enough, so that the last inequality holds uniformly in n and $r_n = o(1)$, which verifies (7.5) and proves the first part of the theorem.

To show the second part of the theorem, about the process $\bar{M}_n'(\mathbf{t}) = \bar{M}_n(\mathbf{B}_n^{-1/2}\mathbf{t})$, we utilize the asymptotic expansion (2.5), which gives us that uniformly in T

$$\bar{M}_n'(\mathbf{t}) = \frac{\mathbf{B}_n^{-1/2}\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n c_i\, \mathbf{x}_i \left[\psi'(e_i) - \gamma_1\right] + o_p(1).$$

Now it is quite easy to verify the convergence of the process on the right-hand side by Theorem 7.12 (condition **X.3** is utilized here). □

Remark 4. Let us replace the condition **B.3** by an assumption that there exists $\delta > 0$ such that the set of random variables

$$\mathcal{G} = \left\{|\psi'(e_1 - s)|^2,\ |s| < \delta\right\} \tag{2.7}$$

is uniformly integrable (Definition 7.16). Then we can still prove the tightness of the process $\bar{M}_n$, which yields $\|\bar{M}_n\|_T = O_p(1)$.

The proof is based directly on Theorem 7.12. Put $\varepsilon_n = \max_{1\le i\le n} \frac{M|\mathbf{x}_i|_2}{\sqrt{n}}$ and define the metric ρ on T as $\rho(\mathbf{t},\mathbf{s}) = C\sqrt{|\mathbf{t}-\mathbf{s}|_2}$, where C is a constant which will be specified later. Denote $B(\varepsilon)$ ($\subset T$) a ρ-ball of radius ε. Then

$$\begin{aligned}
\sum_{i=1}^n \mathsf{E} \sup_{\mathbf{t},\mathbf{s}\in B(\varepsilon)} \left[M_{ni}(\mathbf{t}) - M_{ni}(\mathbf{s})\right]^2 &\le \sum_{i=1}^n |c_i|^2\, \mathsf{E} \sup_{\mathbf{t},\mathbf{s}\in B(\varepsilon)} \left| \int_{\frac{\mathbf{s}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}}^{\frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}} \psi'(e_i - v)\, dv \right|^2 \\
&\le \sum_{i=1}^n |c_i|^2\, \mathsf{E} \sup_{\mathbf{t},\mathbf{s}\in B(\varepsilon)} \left| \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} - \frac{\mathbf{s}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} \right| \int_{\frac{\mathbf{s}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}}^{\frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}} \left[\psi'(e_i - v)\right]^2 dv \\
&\le \frac{M\,\varepsilon^2}{C\,n} \sum_{i=1}^n |c_i|^2 |\mathbf{x}_i|_2^2 \sup_{|v|\le\varepsilon_n} \mathsf{E}\, \left[\psi'(e_1 - v)\right]^2.
\end{aligned}$$

We immediately see that if we take C large enough then our assumptions **X.3-4** and (2.7) imply the metric $\rho(\mathbf{t},\mathbf{s}) = C\sqrt{|\mathbf{t}-\mathbf{s}|_2}$ to satisfy both (7.3) and (7.4) of Theorem 7.12.

Analogously

$$\begin{aligned}
\sum_{i=1}^n \mathsf{E}\, \|M_{ni}\|_T\, \mathbb{I}_{\{\|M_{ni}\|_T > \eta\}} \le \frac{M^2}{\eta\, n} \sum_{i=1}^n |c_i|^2 |\mathbf{x}_i|^2\, \mathsf{E} \int_{-M}^{M} \left[\psi'\left(e_1 - \tfrac{v\,|\mathbf{x}_i|_2}{\sqrt{n}}\right)\right]^2 dv \\
\mathbb{I}\left\{ \tfrac{M^2 |c_i|^2 |\mathbf{x}_i|^2}{n} \int_{-M}^{M} \left[\psi'\left(e_1 - \tfrac{v\,|\mathbf{x}_i|_2}{\sqrt{n}}\right)\right]^2 dv > \eta^2 \right\} \xrightarrow[n\to\infty]{} 0,
\end{aligned}$$

which implies that the assumption (7.2) is satisfied as well.

Now we would like to approximate the expectation of the process $M_n(\mathbf{t})$. Let γ_2 stand for the second derivative of the function $\lambda(t) = \mathsf{E}\ \psi(e_1 + t)$ at the point 0. That is $\gamma_2 = \sum_{j=0}^k \alpha_j \left[f(r_{j+1}) - f(r_j)\right]$ for a piecewise linear ψ and $\gamma_2 = \mathsf{E}\ \psi''(e_1)$ for a sufficiently smooth ψ.[1]

[1] This is not correct in a strict mathematical sense as the smoothness of ψ alone does not allow us to interchange differentiation and integration. But as we write down the formulae for $\gamma_1, \gamma_2, \ldots$ only to give the reader an intuition behind this quantities, we will not discuss this problem in this thesis.

Lemma 2.2. *Under the conditions* ***X.1-4*** *and* ***SmFx.1-3*** *it holds uniformly in* $\mathbf{t} \in T$

$$\mathsf{E}\ M_n(\mathbf{t}) = -\tfrac{\gamma_1 \mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n c_i\, \mathbf{x}_i + \tfrac{\gamma_2}{2}\, \mathbf{t}^\mathsf{T} W_n \mathbf{t} + o(1),$$

where $W_n = \frac{1}{n} \sum_{i=1}^n c_i\, \mathbf{x}_i \mathbf{x}_i^\mathsf{T}$.

Proof. By the assumption **SmFx.3** the function $\lambda(t) = \mathsf{E}\ \lambda(e_1 + t)$ has a second derivative. This enables us to calculate

$$\begin{aligned}
R_n(\mathbf{t}) &= \mathsf{E}\ M_n(\mathbf{t}) + \tfrac{\gamma_1 \mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n c_i\, \mathbf{x}_i - \tfrac{\gamma_2}{2}\, \mathbf{t}^\mathsf{T} W_n \mathbf{t} \\
&= \sum_{i=1}^n c_i\, \mathsf{E} \left[\psi(e_1 - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - \psi(e_1) + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\, \psi'(e_1) - \tfrac{\gamma_2}{2} (\tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})^2\right] \\
&= \sum_{i=1}^n c_i \left[\lambda(-\tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - \lambda(0) + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\, \lambda'(0) - \tfrac{1}{2} (\tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})^2 \lambda''(0)\right]
\end{aligned}$$

$$= \sum_{i=1}^n c_i \int_0^{\frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}} \left[-\lambda'(-v) + \lambda'(0) - v\, \lambda''(0)\right] dv = \sum_{i=1}^n c_i \int_0^{\frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}} \int_0^v \left[\lambda''(-w) - \lambda''(0)\right] dw\, dv.$$

Let ε_n be given by (2.3). We get

$$\|R_n\|_T \le \frac{M^2}{n} \sum_{i=1}^n |c_i|\, |\mathbf{x}_i|_2^2 \sup_{|t| \le \varepsilon_n} |\lambda''(t) - \lambda''(0)|.$$

But the last quantity converges to zero as the second derivative of the function $\lambda(t)$ is continuous at zero and as the Cauchy-Schwartz inequality implies us

$$\frac{1}{n} \sum_{i=1}^n |c_i| |\mathbf{x}_i|_2^2 \le \left[\frac{1}{n} \sum_{i=1}^n |c_i|^2 |\mathbf{x}_i|_2^2\right]^{1/2} \left[\frac{1}{n} \sum_{i=1}^n |\mathbf{x}_i|_2^2\right]^{1/2} \overset{\mathbf{X.2,\ X.4}}{=} O(1).$$

The lemma is proved. □

Combining Theorem 2.1 and Lemma 2.2 yields the asymptotic representation for the process $\{M_n(\mathbf{t}),\, \mathbf{t} \in T\}$.

Corollary 2.3. *Under conditions* ***X.1-4*** *and* ***SmFx.1-3*** *it holds uniformly in* $\mathbf{t} \in T$

$$M_n(\mathbf{t}) + \tfrac{\gamma_1 \mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n c_i\, \mathbf{x}_i = -\tfrac{\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n c_i\, \mathbf{x}_i [\psi'(e_i) - \gamma_1] + \tfrac{\gamma_2}{2}\, \mathbf{t}^\mathsf{T} W_n \mathbf{t} + o_p(1). \tag{2.8}$$

Remark 5. The symmetry condition **Sym** implies $\gamma_2 = 0$ and the term $\frac{\gamma_2}{2}\, \mathbf{t}^\mathsf{T} W_n \mathbf{t}$ in the expansion (2.8) vanishes.

2.1.3 Studentized M-processes

As M-estimators are generally not scale invariant, in practice they are usually studentized. Let S_n be an estimator of a scale, which converges to in probability to S.

To investigate properties of studentized estimators, we need to study the studentized M-processes and consider

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

Put $\bar{M}_n(\mathbf{t}, u) = M_n(\mathbf{t}, u) - \mathsf{E}\ M_n(\mathbf{t}, u)$ and take $T = \{(\mathbf{s}, u) : |\mathbf{s}|_2 \leq M, |u| \leq M\}$ $(\subset \mathbb{R}^{p+1})$ as the index set with the metric $\rho((\mathbf{t}, u), (\mathbf{s}, v)) = |\mathbf{t} - \mathbf{s}|_2 + |u - v|$.

Assumptions

In the following, the abbreviation **SmSt** stands for **Sm**ooth **St**udentized.

SmSt.1 ψ is absolutely continuous with a derivative ψ' such that $\mathsf{E}\ \psi'\left(\frac{e_1}{S}\right)^2 < \infty$.

SmSt.2 the function $\psi'\left(\frac{e_1+t}{Se^u}\right)$ is continuous at the point $(0, 0)$ in the quadratic mean, that is

$$\lim_{(t,u)\to(0,0)} \mathsf{E}\ \left[\psi'\left(\tfrac{e_1+t}{Se^u}\right) - \psi'\left(\tfrac{e_1}{S}\right)\right]^2 = 0. \tag{2.9}$$

SmSt.3 There exist second partial derivatives of the function $\lambda(t, u) = \mathsf{E}\ \psi(\frac{e_1+t}{Se^u})$ in a neighbourhood of the point $(0, 0)$, which are continuous at the point $(0, 0)$.

By Lemma 7.17, the equation (2.9) is certainly satisfied, if ψ' is continuous.

If we suppose the function ψ to be twice differentiable then the condition **SmSt.3** is met provided the following functions

$$\begin{aligned}
a(t, u) &= \mathsf{E}\ \psi''(\tfrac{e_1+t}{Se^u}), \\
b(t, u) &= \mathsf{E}\ \left[\left(\tfrac{e_1+t}{Se^u}\right)^2\ \psi''(\tfrac{e_1+t}{Se^u})\right], \\
c(t, u) &= \mathsf{E}\ \left[\tfrac{e_1+t}{Se^u}\ \psi''(\tfrac{e_1+t}{Se^u})\right], \\
d(t, u) &= \mathsf{E}\ \left[\tfrac{e_1+t}{Se^u}\ \psi'(\tfrac{e_1+t}{Se^u})\right]
\end{aligned}$$

are bounded and continuous in the neighbourhood of the point $(0, 0)$ and we can interchange the derivative and the expectation. This condition is certainly satisfied if ψ is constant for all $|x| \geq K$ (K is sufficiently large) and if ψ'' is uniformly continuous.

If ψ is a piecewise linear function, then the conditions **SmSt.1-3** are satisfied if we simply replace the points $r_1, \ldots, r_k$ in the conditions **A.2** and **A.3** by the points $S\, r_1, \ldots, S\, r_k$.

Remark 6. Recall that in Remark 2 in Table 1, we gave a list of assumptions which ensures the conditions **SmFx.2-3** to hold for some of the most famous ψ functions. It is easy to verify that we can construct an analogous table for the conditions **SmSt.2-3**. The only thing we should do is to replace the points a, b, c, k with $S\,a, S\,b, S\,c, S\,k$.

Before we proceed, it is useful to introduce some notation. In the following, the partial derivatives of the functions $\lambda(t,u) = \mathsf{E}\ \psi(\frac{e_1+t}{Se^u})$ and $\delta(t,u) = \mathsf{E}\ \frac{e_1}{S}\,\psi'(\frac{e_1+t}{Se^u})$ will be indicated by lower subscripts. Put

$$\begin{aligned} \gamma_1 &= \lambda_t(0,0)\ \left(= \tfrac{1}{S}\,\mathsf{E}\ \psi'\left(\tfrac{e_1}{S}\right)\right), \qquad \gamma_{1e} = -\lambda_u(0,0)\ \left(= \mathsf{E}\ \tfrac{e_1}{S}\,\psi'\left(\tfrac{e_1}{S}\right)\right) \\ \gamma_2 &= \lambda_{tt}(0,0)\ \left(= \tfrac{1}{S^2}\,\mathsf{E}\ \psi''\left(\tfrac{e_1}{S}\right)\right), \qquad \gamma_{2e} = \delta_t(0,0)\ \left(= \mathsf{E}\ \tfrac{e_1}{S^2}\,\psi''\left(\tfrac{e_1}{S}\right)\right), \\ \gamma_{2ee} &= -\delta_u(0,0)\ \left(= \mathsf{E}\ \left(\tfrac{e_1}{S}\right)^2\psi''\left(\tfrac{e_1}{S}\right)\right). \end{aligned} \tag{2.10}$$

The formulae in the brackets are for the case of ψ sufficiently smooth (and integrable). We omit the formulae for the case of a piecewise continuous ψ as they are rather complicated in general case. By the assumptions **SmSt.1-3** all these quantities are finite. Notice that

$$\lambda_{tu}(0,0) = \gamma_1 + \gamma_{2e} \qquad \text{and} \qquad \lambda_{uu}(0,0) = \gamma_{1e} + \gamma_{2ee}.$$

Theorem 2.4. *Let the conditions **X.1-4** and **SmSt.1-2** be satisfied, then*

$$\sup_{(\mathbf{t},u)\in T}\left|\bar{M}_n(\mathbf{t},u) + \tfrac{\mathbf{t}^\mathsf{T}}{\sqrt{n}}\sum_{i=1}^n c_i\,\mathbf{x}_i\left[\tfrac{1}{S}\,\psi'(e_i/S) - \gamma_1\right] + \tfrac{u}{\sqrt{n}}\sum_{i=1}^n c_i\left[\tfrac{e_i}{S}\,\psi'(e_i/S) - \gamma_{1e}\right]\right| = o_P(1).$$

Proof. Let us denote

$$Z_{ni}(\mathbf{t},u) = c_i\left[\psi\left(e^{-u\,n^{-1/2}}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\right) - \psi(e_i/S) + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{S\sqrt{n}}\,\psi'(e_i/S) + \tfrac{u}{\sqrt{n}}\tfrac{e_i}{S}\,\psi'(e_i/S)\right].$$

Then after a little algebra we get

$$\begin{aligned} |Z_{ni}(\mathbf{t},u) - Z_{ni}(\mathbf{s},v)| \le \sum_{i=1}^n |c_i|\left|\int_{\frac{\mathbf{s}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}}^{\frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}}\left[\tfrac{1}{Se^{u/\sqrt{n}}}\,\psi'\left(\tfrac{e_i-r}{Se^{u/\sqrt{n}}}\right) - \tfrac{1}{S}\,\psi'\left(\tfrac{e_i}{Se^{u/\sqrt{n}}}\right)\right]dr\right| \\ + \sum_{i=1}^n |c_i|\left|\int_{v/\sqrt{n}}^{u/\sqrt{n}}\left[\tfrac{e_i}{Se^w}\,\psi'\left(\tfrac{e_i}{Se^w}\right) - \tfrac{e_i}{S}\,\psi'\left(\tfrac{e_i}{S}\right)\right]dw\right|. \end{aligned}$$

The rest of the proof is analogous to the proof of Theorem 2.1. □

The following lemma approximate the expectation of the process $M_n(\mathbf{t},u)$.

Lemma 2.5. *Suppose that the conditions **X.1-4** and **SmSt.1-3** are satisfied and let us denote $W_n = \frac{1}{n}\sum_{i=1}^n c_i\,\mathbf{x}_i\mathbf{x}_i^\mathsf{T}$. Then uniformly in $(\mathbf{t},u)\in T$*

$$\begin{aligned} \mathsf{E}\ M_n(\mathbf{t},u) = -\tfrac{\gamma_1\,\mathbf{t}^\mathsf{T}}{\sqrt{n}}\sum_{i=1}^n c_i\,\mathbf{x}_i - \tfrac{\gamma_{1e}\,u}{\sqrt{n}}\sum_{i=1}^n c_i \\ + \tfrac{\gamma_2}{2}\,\mathbf{t}^\mathsf{T}W_n\mathbf{t} + \tfrac{(\gamma_{2e}+\gamma_1)u\,\mathbf{t}^\mathsf{T}}{n}\sum_{i=1}^n c_i\,\mathbf{x}_i + \tfrac{(\gamma_{2ee}+\gamma_{1e})\,u^2}{2n}\sum_{i=1}^n c_i + o(1). \end{aligned} \tag{2.11}$$

Proof. The proof of this lemma is akin to the proof of Lemma 2.2. $\square$

Remark 7. If $\sum_{i=1}^{n} c_i = 0$, then the second and the fifth term in the expansion (2.11) disappear. Even simpler situation occurs if the symmetry condition **Sym** is satisfied. In this case $\gamma_2 = \gamma_{1e} = \gamma_{2ee} = 0$, which yields a more 'friendly' expansion

$$\mathsf{E}\ M_n(\mathbf{t}, u) = -\tfrac{\gamma_1\,\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^{n} c_i\,\mathbf{x}_i + \tfrac{(\gamma_{2e}+\gamma_1)\,u\,\mathbf{t}^\mathsf{T}}{n} \sum_{i=1}^{n} c_i\,\mathbf{x}_i + o(1). \tag{2.12}$$

Combining Theorem 2.4 and Lemma 2.5 gives us the asymptotic representation of the process M_n.

Corollary 2.6. *Under conditions **X.1-4** and **SmSt.1-3** it holds uniformly in* $(\mathbf{t}, u) \in T$

$$\begin{aligned} M_n(\mathbf{t}, u) &+ \tfrac{\gamma_1\,\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^{n} c_i\,\mathbf{x}_i + \tfrac{\gamma_{1e}\,u}{\sqrt{n}} \sum_{i=1}^{n} c_i \\ &= -\tfrac{\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^{n} c_i\,\mathbf{x}_i \left[\tfrac{1}{S}\,\psi'(e_i/S) - \gamma_1\right] - \tfrac{u}{\sqrt{n}} \sum_{i=1}^{n} c_i \left[\tfrac{e_i}{S}\,\psi'(e_i/S) - \gamma_{1e}\right] \\ &\quad + \tfrac{\gamma_2}{2}\,\mathbf{t}^\mathsf{T} W_n \mathbf{t} + \tfrac{(\gamma_{2e}+\gamma_1)\,u\,\mathbf{t}^\mathsf{T}}{n} \sum_{i=1}^{n} c_i\,\mathbf{x}_i + \tfrac{(\gamma_{2ee}+\gamma_{1e})\,u^2}{2n} \sum_{i=1}^{n} c_i + o_p(1). \end{aligned} \tag{2.13}$$

2.2 A step function ψ

In this section we suppose that the function ψ is a step-function, that is

$$\psi(x) = \alpha_j \qquad \text{for} \quad q_{j-1} < x \le q_j, \quad j = 1, \dots, m, \tag{2.14}$$

where $\alpha_0, \alpha_1, \dots, \alpha_m$ are real numbers (not all equal) and $-\infty = q_0 < q_1 < \dots < q_m = \infty$, where m is a positive integer.

For our purposes it is sometimes more convenient to rewrite the function ψ in a form

$$\psi(x) = \sum_{j=1}^{m} \beta_j\,\mathbb{I}\{x \le q_j\}, \tag{2.15}$$

where $\beta_j = \alpha_j - \alpha_{j+1}$ for $j = 1, \dots, m-1$ and $\beta_m = \alpha_m$.

2.2.1 Fixed scale

In the following, we will be interested in the processes

$$M_n(\mathbf{t}) = \frac{1}{n^{1/4}} \sum_{i=1}^{n} c_i \left[\psi(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - \psi(e_i)\right], \tag{2.16}$$

$$\bar{M}_n(\mathbf{t}) = M_n(\mathbf{t}) - \mathsf{E}\ M_n(\mathbf{t}), \tag{2.17}$$

with $\mathbf{t} \in T = \{\mathbf{s} \in \mathbb{R}^p : |\mathbf{s}|_2 \leq M\}$, where $|\cdot|_2$ stands for the euclidean norm, and M is an arbitrary large but fixed constant.

Notice that in comparison to the case of a smooth ψ we scale the process M_n with $n^{-1/4}$. As we will see later, this is because for small t the variance $\mathrm{var}\{\psi(e_1 - t) - \psi(e_1)\}$ is not of order t^2 (as in the case of a smooth ψ) but only of order $|t|$.

Remark 8. To make the following results more precise, we need to make a few technical comments. First, we look at the process M_n ($\bar{M}_n$) as a mapping from the underlying probability space into $\ell^\infty(T)$ (the space of bounded functions on T). It is well known (see e.g. Billingsley (1968)), that the empirical distribution function viewed as a mapping into the space of bounded functions is not Borel measurable. That is why we cannot expect our processes to be measurable. In what follows, by the weak convergence of such processes we will mean the (star) weak convergence in the sense of Hoffmann-Jørgensen (see Appendix Section 7.3).

Assumptions

We will need a slightly modified assumptions on the design $\mathbf{x}_{1n}, \ldots, \mathbf{x}_{nn}$ and the constants $c_{1n}, \ldots, c_{nn}$.

X'.1

$$\frac{1}{n}\sum_{i=1}^{n} c_{in}^2 = O(1), \qquad \lim_{n\to\infty} \frac{\max_{1\leq i\leq n} |c_{in}|}{n^{1/4}} = 0.$$

X'.2

$$\lim_{n\to\infty} \frac{\max_{1\leq i\leq n} |\mathbf{x}_{in}|_2}{\sqrt{n}} = 0.$$

X'.3

$$\frac{1}{n}\sum_{i=1}^{n} |c_{in}|^2\, |\mathbf{x}_{in}|_2 = O(1).$$

X'.4 There exists a $\delta > \frac{1}{2}$ such that

$$B_n^2 = \frac{1}{n}\sum_{i=1}^{n} |c_{in}|\, |\mathbf{x}_{in}|_2^{1+\delta} = O(1).$$

Notice that in comparison with the conditions **X.1-4** for a smooth function ψ we have to strengthen the convergence $\frac{\max_{1\leq i\leq n}|c_{in}|}{\sqrt{n}} \to 0$ to $\frac{\max_{1\leq i\leq n}|c_{in}|}{n^{1/4}} \to 0$. On the other hand we do not need to assume so much about the design points $\mathbf{x}_{1n}, \ldots, \mathbf{x}_{nn}$. Finally if we put $c_{in} = \mathbf{x}_{in}$, the conditions **X'.1-4** turns out to be

XX'.1

$$\frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_{in}|_2^3 = O(1), \qquad \lim_{n\to\infty} \frac{\max_{1\leq i\leq n} |\mathbf{x}_{in}|_2^2}{n^{1/2}} = 0.$$

Thus in comparison with the condition **XX.1** we only need the third moment of the norm of the rows of the matrix $\mathbf{X}$ to be finite. To simplify the notation we will write shortly c_i and $\mathbf{x}_i$.

As the function ψ is not continuous, we need to impose the following smoothness conditions on the cdf F in neighbourhoods of the points of discontinuity of ψ.

Step.1 F has a continuous derivative in a neighbourhood of the points $q_1, \ldots, q_m$.

Step.2 For every $j \in \{1, \ldots, m\}$ there exist $\delta_j > 0$, $\nu_j > \frac{1}{2}$, and $C_j < \infty$ such that for every $|t| < \delta_j$

$$|f(q_j + t) - f(q_j)| \leq C_j \, |t|^{\nu_j}.$$

Similarly to the case of a smooth ψ, we need the first condition to prove the asymptotic tightness of the process $\bar{M}_n$ and the second one to approximate $\mathsf{E}\, M_n$.

Remark 9. As there is only a finite number of discontinuities of ψ, the condition **Step.2** is certainly satisfied uniformly in $j \in \{1, \ldots, m\}$ for ν, δ, C given by

$$\nu = \min_{1 \leq j \leq m} \nu_j, \qquad \delta = \min_{1 \leq j \leq m} \delta_j \wedge 1, \qquad C = \max_{1 \leq j \leq m} C_j.$$

2.2.2 Theorems

Theorem 2.7. *Provided the conditions **X'.1-3** and **Step.1** are satisfied, then the centered process $\bar{M}_n = M_n - \mathsf{E}\, M_n$ is asymptotically tight.*[2]

Moreover, if there exists a real function $r : T \times T \to \mathbb{R}$ such that for every $\mathbf{t}$, $\mathbf{s} \in T$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} |c_i|^2 \min(|\mathbf{t}^\mathsf{T} \mathbf{x}_i|, |\mathbf{s}^\mathsf{T} \mathbf{x}_i|) \, \mathbb{I}\{\mathbf{t}^\mathsf{T} \mathbf{x}_i \mathbf{x}_i^\mathsf{T} \mathbf{s} > 0\} = r(\mathbf{t}, \mathbf{s}), \tag{2.18}$$

then the process $\bar{M}_n$ converges in distribution to a centered gaussian process Z with the covariance function

$$\operatorname{cov}\{Z(\mathbf{t}), Z(\mathbf{s})\} = r(\mathbf{t}, \mathbf{s}) \left(\sum_{j=1}^{m} \alpha_j^2 \left[f(q_j) - f(q_{j-1}) \right] \right).$$

Notice that in comparison with Theorem 2.1 we are not able to find an approximation up to a remainder term of order $o_p(1)$, but only a limit process.

Proof. As the function ψ is a linear combination of jumps, without lost of generality it suffices to consider $\psi(x) = \mathbb{I}\{x \leq q\}$.

Put

$$M_{ni}(\mathbf{t}) = c_i \left[\psi(e_i - \tfrac{\mathbf{t}^\mathsf{T} \mathbf{x}_i}{\sqrt{n}}) - \psi(e_i) \right] = c_i \left[\mathbb{I}\{e_i \leq q + \tfrac{\mathbf{t}^\mathsf{T} \mathbf{x}_i}{\sqrt{n}}\} - \mathbb{I}\{e_i \leq q\} \right].$$

[2] See Definition 7.9 of Appendix

To prove the theorem, we will verify the assumptions of Theorem 7.12 of Appendix.

Note that for arbitrary $\mathbf{t}, \mathbf{s} \in T$

$$|M_{ni}(\mathbf{t}) - M_{ni}(\mathbf{s})| \leq |c_i| \, \mathbb{I}\{e_i \text{ lies between } q + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} \text{ and } q + \tfrac{\mathbf{s}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\}.$$

By assumption **Step.1** there exists $\delta > 0$ such that the quantity $K = \sup_{|v| \leq \delta} f(q+v)$ is finite. Put $\varepsilon_n = \max_{1 \leq i \leq n} \frac{M|\mathbf{x}_i|_2}{\sqrt{n}}$. The convergence $\varepsilon_n \to 0$ implies that for all sufficiently large n it holds $\sup_{|v| \leq \varepsilon_n} f(q+v) \leq K$.

Let us define the metric ρ on T as $\rho(\mathbf{t}, \mathbf{s}) = C\sqrt{|\mathbf{t} - \mathbf{s}|_2}$, where C is a constant which will be specified later. Denote $B(\varepsilon)$ $(\subset T)$ a ρ-ball of radius ε. Further, for $i = 1, \ldots, n$ define

$$l_i = \inf_{\mathbf{w} \in B(\varepsilon)} \{\tfrac{\mathbf{w}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\}, \qquad u_i = \sup_{\mathbf{w} \in B(\varepsilon)} \{\tfrac{\mathbf{w}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\}.$$

Obviously

$$u_i - l_i \leq \frac{|\mathbf{x}_i|_2}{\sqrt{n}} \sup_{\mathbf{t}, \mathbf{s} \in B(\varepsilon)} |\mathbf{t} - \mathbf{s}|_2 \leq \frac{4\,\varepsilon^2 |\mathbf{x}_i|_2}{C^2\sqrt{n}}.$$

Thus we can bound

$$\sum_{i=1}^n \mathsf{E}^* \sup_{\mathbf{t}, \mathbf{s} \in B(\varepsilon)} [M_{ni}(\mathbf{t}) - M_{ni}(\mathbf{s})]^2$$
$$\leq \frac{1}{\sqrt{n}} \sum_{i=1}^n |c_i|^2 \, [F(q+u_i) - F(q+l_i)] \leq \frac{4\,\varepsilon^2 K}{C^2 n} \sum_{i=1}^n |c_i|^2 |\mathbf{x}_i|_2.$$

By the assumptions of the theorem we immediately see that our metric ρ satisfies the equation (7.3) as well as (7.4) provided we take the constant C in the definition of the metric ρ large enough.

To verify the condition (7.2), calculate

$$\sum_{i=1}^n \mathsf{E}^* \, \|M_{ni}\|_T \, \mathbb{I}_{\{\|M_{ni}\|_T > \eta\}} \leq \sum_{i=1}^n \mathsf{E}^* \, \|M_{ni}\|_T \, \mathbb{I}\left\{\tfrac{|c_i|}{n^{1/4}} \, \mathbb{I}\{|e_i - q| \leq \tfrac{M|\mathbf{x}_i|_2}{\sqrt{n}}\} > \eta\right\}. \qquad (2.19)$$

By assumption **X'.1** $\max_{1 \leq i \leq n} |c_i| = o(n^{-1/4})$, which implies that the right hand side of (2.19) diminishes for all sufficiently large n. This proves the asymptotic tightness of the process $\bar{M}_n$.

To prove the second part of the theorem, it remains to show that the process $\bar{M}_n$ converges marginally in distribution to the process Z. By Cramér-Wald device it would be enough to verify that for any $\mathbf{t}_1, \ldots, \mathbf{t}_k \in T$, $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$, $k \geq 1$ the random variable $\sum_{j=1}^k \lambda_j \, \bar{M}_n(\mathbf{t}_j)$ converges in distribution to the random variable $\sum_{j=1}^k \lambda_j \, Z(\mathbf{t}_j)$. But this will follow immediately by the Feller-Lindeberg theorem (Theorem 7.4) if we prove the pointwise convergence of the covariance function of the process $\bar{M}_n$, that is

$$\lim_{n \to \infty} \operatorname{cov}\{M_n(\mathbf{t}), M_n(\mathbf{s})\} = r(\mathbf{t}, \mathbf{s}) \, f(q), \quad \text{for all } \mathbf{t}, \mathbf{s} \in T.$$

It is straightforward to calculate

$$
\begin{aligned}
\operatorname{cov}\{M_n(\mathbf{t}), M_n(\mathbf{s})\} &= \sum_{i=1}^n \operatorname{cov}\{M_{ni}(\mathbf{t}), M_{ni}(\mathbf{s})\} \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^n |c_i|^2 \; \mathsf{E}\, \mathbb{I}\{e_i \text{ lies between } q \text{ and } q + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\} &\mathbb{I}\{e_i \text{ lies between } q \text{ and } q + \tfrac{\mathbf{s}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\} \\
- \frac{1}{\sqrt{n}} \sum_{i=1}^n |c_i|^2 \; \mathsf{E}\, \mathbb{I}\{e_i \text{ lies between } q \text{ and } q + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\} \; &\mathsf{E}\, \mathbb{I}\{e_i \text{ lies between } q \text{ and } q + \tfrac{\mathbf{s}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\} \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^n |c_i|^2 \left[F\left(q + \max\left\{0, \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} \wedge \tfrac{\mathbf{s}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\right\}\right) \right. &- \left. F\left(q + \min\left\{0, \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} \vee \tfrac{\mathbf{s}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\right\}\right)\right] + O(\tfrac{1}{\sqrt{n}}) \\
&= \frac{1}{n} \sum_{i=1}^n |c_i|^2 f(q) \min(|\mathbf{t}^\mathsf{T}\mathbf{x}_i|, |\mathbf{s}^\mathsf{T}\mathbf{x}_i|)\, \mathbb{I}\{\mathbf{t}^\mathsf{T}\mathbf{x}_i\mathbf{x}_i^\mathsf{T}\mathbf{s} > 0\} + o(1) \xrightarrow[n\to\infty]{(2.18)} r(\mathbf{t},\mathbf{s})\, f(q),
\end{aligned}
$$

which proves the theorem. $\square$

Now we approximate $\mathsf{E}\; M_n(\mathbf{t})$.

Lemma 2.8. *Under the conditions* ***X'.1-4*** *and* ***Step.1-2*** *it holds*

$$
\mathsf{E}\; M_n(\mathbf{t}) = \tfrac{\mathbf{t}^\mathsf{T}}{n^{3/4}} \sum_{i=1}^n c_i\, \mathbf{x}_i \sum_{j=1}^m \beta_j f(q_j) + o(1) = \tfrac{\gamma_1 \mathbf{t}^\mathsf{T}}{n^{3/4}} \sum_{i=1}^n c_i\, \mathbf{x}_i + o(1)
$$

uniformly in $\mathbf{t} \in T$.

Proof. Without loss of generality we can only consider the case $\psi(x) = \mathbb{I}\{x \le q\}$. By the assumption **Step.2** there exists $\nu > \frac{1}{2}$ such that

$$
\begin{aligned}
\left| \mathsf{E}\; M_n(\mathbf{t}) - \tfrac{\mathbf{t}^\mathsf{T}}{n^{3/4}} \sum_{i=1}^n c_i\, \mathbf{x}_i\, f(q) \right| &= \left| \frac{1}{n^{1/4}} \sum_{i=1}^n c_i [F(q + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - F(q)] - \tfrac{\mathbf{t}^\mathsf{T}}{n^{3/4}} \sum_{i=1}^n c_i\, \mathbf{x}_i\, f(q) \right| \\
= \left| \frac{1}{n^{1/4}} \sum_{i=1}^n c_i \int_0^{\frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}} [f(q+v) - f(q)]\, dv \right| &\overset{\textbf{Step.2}}{\le} \frac{C}{n^{1/4}} \sum_{i=1}^n |c_i| \int_0^{\left|\frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\right|} v^\nu dv \\
\le \frac{C}{(\nu+1) n^{1/4}} \sum_{i=1}^n |c_i| \left(\tfrac{M|\mathbf{x}_i|_2}{\sqrt{n}}\right)^{\nu+1} &= \frac{C M^{\nu+1}}{(\nu+1) n^{3/4+\nu/2}} \sum_{i=1}^n |c_i|\, |\mathbf{x}_i|^{\nu+1}.
\end{aligned}
$$

Notice that the last quantity does not depend on $\mathbf{t}$ and it converges to zero by the assumption **X'.4**. $\square$

Combining Theorem 2.7 and Lemma 2.8 yields the asymptotic distribution for the process $\{M_n(\mathbf{t}),\, \mathbf{t} \in T\}$.

Corollary 2.9. *Under conditions* ***X'.1-4****,* ***Step.1-2*** *and* (2.18) *the process*

$$
M_n(\mathbf{t}) - \tfrac{\gamma_1 \mathbf{t}^\mathsf{T}}{n^{3/4}} \sum_{i=1}^n c_i\, \mathbf{x}_i \sum_{j=1}^m \beta_j f(q_j), \qquad \mathbf{t} \in T
$$

converges in distribution to a centered gaussian process $\{Z(\mathbf{t}), \mathbf{t} \in T\}$ *with the covariance function*

$$\operatorname{cov}\{Z(\mathbf{t}), Z(\mathbf{s})\} = r(\mathbf{t}, \mathbf{s}) \left(\sum_{j=1}^{m} \alpha_j^2 \left[f(q_j) - f(q_{j-1}) \right] \right),$$

where the function $r(\mathbf{t}, \mathbf{s})$ *is given in* (2.18).

2.2.3 Studentized M-processes

Unlike in the case of a smooth ψ the studentization of the M-estimator based on a step ψ brings no new conceptual problems. Denote

$$M_n(\mathbf{t}, u) = \frac{1}{n^{1/4}} \sum_{i=1}^{n} c_i \left[\psi\left(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S \right) - \psi(e_i/S) \right], \qquad |\mathbf{t}|_2 \leq M,\ |u| \leq M,$$

and put $\bar{M}_n(\mathbf{t}, u) = M_n(\mathbf{t}, u) - \mathsf{E}\, M_n(\mathbf{t}, u)$. The index set is $T = \{(\mathbf{s}, u) : |\mathbf{s}|_2 \leq M, |u| \leq M\}$ $(\subset \mathbb{R}^{p+1})$ with the metric $\rho((\mathbf{t}, u), (\mathbf{s}, v)) = \sqrt{|\mathbf{t} - \mathbf{s}|_2} + \sqrt{|u - v|}$.

To prove the asymptotic results about the process M_n, we can very closely follow the proofs of the previous section. That is why the following theorem only summarizes the results. As we will see, only the covariance structure of the limiting process is more complicated.

But before we state the theorem, we note that speaking about conditions **Step.1-2** in the studentized case, we require the smoothness assumptions on the cdf F hold in a neighbourhood of the points $S\,q_1, \ldots, S\,q_m$ (instead of $q_1, \ldots, q_m$).

Theorem 2.10. *If the conditions* ***X'.1-3*** *and* ***Step.1*** *hold, then the process* $\bar{M}_n = M_n - \mathsf{E}\, M_n$ *is asymptotically tight.*

Moreover, if there exists a function r *defined on* $T \times T$ *such that for every* $(\mathbf{t}, u)$, $(\mathbf{s}, v) \in T$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{m} \beta_j^2 \sum_{i=1}^{n} |c_i|^2 \min(|Sq_j u + \mathbf{t}^\mathsf{T}\mathbf{x}_i|, |Sq_j v + \mathbf{s}^\mathsf{T}\mathbf{x}_i|)\, \mathbb{I}\{\mathbf{t}^\mathsf{T}\mathbf{x}_i\mathbf{x}_i^\mathsf{T}\mathbf{s} > 0\} = r\left((\mathbf{t}, u), (\mathbf{s}, v)\right),$$

then the process $\bar{M}_n$ *converges in distribution to a centered gaussian process* Z *with the covariance function*

$$\operatorname{cov}\{Z(\mathbf{t}, u), Z(\mathbf{s}, v)\} = r\left((\mathbf{t}, u), (\mathbf{s}, v)\right).$$

Finally, if in addition the conditions ***X'.4*** *and* ***Step.2*** *are satisfied, then we can replace the mean value* $\mathsf{E}\, M_n(\mathbf{t}, u)$ *by*

$$\tfrac{\mathbf{t}^\mathsf{T}}{n^{3/4}} \sum_{i=1}^{n} c_i\, \mathbf{x}_i \sum_{j=1}^{m} \beta_j f(S\,q_j) + \tfrac{u}{n^{3/4}} \sum_{i=1}^{n} c_i \sum_{j=1}^{m} S\,q_j\,\beta_j\, f(S\,q_j) = \tfrac{\gamma_1 \mathbf{t}^\mathsf{T}}{n^{3/4}} \sum_{i=1}^{n} c_i\, \mathbf{x}_i + \tfrac{\gamma_{1e}\, u}{n^{3/4}} \sum_{i=1}^{n} c_i$$

uniformly in T.

2.3 ψ is a sum of a continuous function and a step function

Sometimes, $\psi = \psi_c + \psi_s$, where ψ_c stands for a continuous function and ψ_s for a step function (e.g. skipped mean is generated with $\psi(x) = x\,\mathbb{I}\{|x < k|\}$). The simplest way how to handle this situation is to assume that the continuous part ψ_c is absolutely continuous and then to combine the results of the previous sections. But we have already seen that the order of the convergence of M-processes is slower for step functions. That is why the continuous part ψ_c only influences the expectation $\mathsf{E}\ M_n$ but not the limit process. In view of this, it would be possible to weaken the assumptions on ψ_c from the first section of this chapter. For instance, we do not need the existence of a derivative of ψ_c. It is certainly sufficient to assume the existence of $\delta_0 > 0$ and $\eta > \frac{3}{4}$ such that for any x

$$|\psi(x + \delta) - \psi(x)| \leq h(x)\,\delta^{\eta}, \qquad \text{for any } |\delta| < \delta_0,$$

where $\mathsf{E}\ h(e_1)^2 < \infty$. In this case we can even weaken the assumptions about the design $\mathbf{x}_1, \ldots, \mathbf{x}_n$ to some sort of conditions similar to **X'.1-4** (with $1 + \delta$ replaced by $2\,\eta$).

But we will not investigate this situation as we believe that with the presented results one can easily handle these special situations ad hoc.

Chapter 3

SOAL of the R-processes with Wilcoxon scores

To motivate the following technical results, recall the linear regression model (1.1). Let c_{in}, $\mathbf{x}_{in}$, $1 \leq i \leq n$, $n = 1, 2, \ldots$ be triangular arrays of constants satisfying some conditions which will be specified later. In this section we study the asymptotic behaviour of the processes

$$\tilde{S}_n(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n} c_{in}\, R_i^{'}(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n} c_{in} \sum_{j=1}^{n} \mathbb{I}\{e_i - \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_{in}}{\sqrt{n}} \geq e_j - \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_{jn}}{\sqrt{n}}\}, \tag{3.1}$$

$$T_n(\mathbf{t}) = \tilde{S}_n(\mathbf{t}) - \tilde{S}_n(\mathbf{0}), \tag{3.2}$$

$$\bar{T}_n(\mathbf{t}) = T_n(\mathbf{t}) - \mathsf{E}\; T_n(\mathbf{t}), \tag{3.3}$$

where $\mathbf{t} = (t_1, \ldots, t_p)^{\mathsf{T}}$ and $R_i^{'}(\mathbf{t})$ stands for the rank of $e_i - \frac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_{in}}{\sqrt{n}}$ among $e_1 - \frac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_{1n}}{\sqrt{n}}, \ldots, e_n - \frac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_{nn}}{\sqrt{n}}$. We will index these processes by the set $T = \{\mathbf{t} \in \mathbb{R}^p : |\mathbf{t}|_2 \leq M\}$, where $|\cdot|_2$ stands for the euclidean norm, and M is an arbitrarily large but fixed constant.

3.1 Assumptions, theorems

In this section we specify the conditions on the distribution of the errors F and on the constants $\mathbf{x}_{1n}, \ldots, \mathbf{x}_{nn}$ and $c_{1n}, \ldots, c_{nn}$ and formulate the results. The proofs of this results are to be found in the following sections.

3.1.1 Assumptions

W.1 F is absolutely continuous with a derivative f such that $\mathsf{E}\,[f(e_1)]^2 < \infty$.

W.2 The function $f(e_1 + s)$ is continuous in the quadratic mean at the point zero, that is

$$\lim_{s\to 0} \mathsf{E}\,[f(e_1 + s) - f(e_1)]^2 = 0.$$

W.3

$$\lim_{\Delta \to 0} \frac{1}{\Delta^2} \int_{-\infty}^{+\infty} \int_{-\Delta}^{+\Delta} [f(z+y) - f(y)]^2 dz\, dy = 0.$$

It follows from the proof of Lemma 2 of Antille (1976) that the condition **W.2** is met if **W.1** holds and the density f is continuous except for a finite number of jump discontinuities.

Further, according to Antille (1976), the condition **W.3** is satisfied in these two important cases

(*i*). f is such that $|f(x+t) - f(x)| \le |t|^\alpha h(x)$, with $\alpha > \frac{1}{2}$ and $h(x) \in L_2(-\infty, +\infty)$,

(*ii*). f is absolutely continuous and $f'(x) \in L_2(-\infty, +\infty)$.

Note that the second condition is satisfied if there exists a finite Fisher information of the density f.

Remark 10. It is easy to show that the condition **W.3** is met for many of the standard distributions with smooth densities like normal, lognormal, t-, χ^2- (with degrees of freedom greater or equal 2), cauchy, and logistic distribution. But it is not satisfied for an exponential distribution.

In addition to the conditions **X.1-4** of the previous chapter (with c_i replaced by c_{in}), we need:

X.5

$$\sum_{i=1}^{n} c_{in} = 0.$$

To simplify the notation, in the following we will write $c_i, \mathbf{x}_i$ instead of $c_{in}, \mathbf{x}_{in}$.

3.1.2 Theorems

Put $\gamma = \mathsf{E}\ f(e_1) = \int f^2(x)\, dx$.

Theorem 3.1. *Under conditions **X.1-5** and **W.1-2** the process* $\{\bar{T}_n(\mathbf{t}), \mathbf{t} \in T\}$ *satisfies uniformly in* $\mathbf{t} \in T$

$$\sup_{\mathbf{t} \in T} \left| \bar{T}_n(\mathbf{t}) + \frac{\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^{n} \left[c_i\, \mathbf{x}_i + \frac{1}{n} \sum_{j=1}^{n} c_j\, \mathbf{x}_j \right] (f(e_i) - \gamma) \right| = o_p(1). \tag{3.4}$$

Specially, if we put

$$\mathbf{A}_n^2 = \frac{1}{n} \sum_{i=1}^{n} c_i^2\, \mathbf{x}_i \mathbf{x}_i^\mathsf{T} + 3 \left(\frac{1}{n} \sum_{i=1}^{n} c_i\, \mathbf{x}_i \right) \left(\frac{1}{n} \sum_{i=1}^{n} c_i\, \mathbf{x}_i^\mathsf{T} \right),$$

and the matrix $\mathbf{A}_n^2$ *is regular for n large enough, then the process* $\bar{T}'_n(\mathbf{t}) = \bar{T}_n(\mathbf{A}_n^{-1}\mathbf{t})$ *converges in distribution to a centered gaussian process* $\{Y(\mathbf{t}), \mathbf{t} \in T\}$ *with the covariance function given by* $\operatorname{cov}(Y(\mathbf{t}), Y(\mathbf{s})) = \sigma^2\, \mathbf{t}^\mathsf{T}\mathbf{s}$, *where* $\sigma^2 = \int f^3(x) dx \ - \ \gamma^2$.

As the proof of this theorem is rather lengthy, we will postpone it and formulate the other results of this chapter.

Lemma 3.2. *Suppose that the density of the errors f satisfies the assumptions* ***W.1-3*** *and the conditions* ***X.1-5*** *are satisfied as well. Then uniformly in* $\mathbf{t} \in T$

$$\mathsf{E}\ T_n(\mathbf{t}) = -\tfrac{\gamma\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n c_i\,\mathbf{x}_i + o(1).$$

Combining Theorem 3.1 and Lemma 3.2 yields the following corollary.

Corollary 3.3. *Under conditions* ***X.1-5*** *and* ***W.1-3*** *it holds uniformly in* $\mathbf{t} \in T$

$$\tilde{S}_n(\mathbf{t}) - \tilde{S}_n(\mathbf{0}) + \tfrac{\gamma\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n c_i\,\mathbf{x}_i = -\tfrac{\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n \left(c_i\,\mathbf{x}_i + \tfrac{1}{n} \sum_{j=1}^n c_j\,\mathbf{x}_j \right) (f(e_i) - \gamma) + o_p(1). \tag{3.5}$$

3.2 Proof of Theorem 3.1

As a first step, we approximate the process $\{\bar{T}_n(\mathbf{t}), \mathbf{t} \in T\}$ by the 'Hájek projection' (see Hájek (1968) or Serfling (1980))

$$P_n = \sum_{i=1}^n \mathsf{E}\ \left[\bar{T}_n | \, e_i\right] - (n-1)\,\mathsf{E}\ \bar{T}_n = \sum_{i=1}^n \mathsf{E}\ \left[\bar{T}_n | \, e_i\right].$$

In the following, we will show that the projection P_n (leading term) has the asymptotic representation (3.4) and the remainder term $R_n = \bar{T}_n - P_n$ is asymptotically negligible, that is $\|R_n\|_T = o_p(1)$.

3.2.1 The convergence of the process P_n

Calculating the projection of the process $\bar{T}_n$, we find out that $P_n(\mathbf{t}) = V_n(\mathbf{t}) - \mathsf{E}\ V_n(\mathbf{t})$, where

$$V_n(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (c_i - c_j) \left[F(e_i - \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}) - F(e_i) \right] \overset{\text{Say}}{=} \sum_{i=1}^n V_{ni}(\mathbf{t}). \tag{3.6}$$

For $i = 1, \ldots, n$ define

$$\begin{aligned} Z_{ni}(\mathbf{t}) &= V_{ni}(\mathbf{t}) - \tfrac{\mathbf{t}^\mathsf{T}}{n^{3/2}} \sum_{j=1}^n (c_i - c_j)\,(\mathbf{x}_i - \mathbf{x}_j) f(e_i) \\ &= \frac{1}{n} \sum_{j=1}^n (c_i - c_j) \left[F(e_i - \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}) - F(e_i) + \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}} f(e_i) \right] \end{aligned}$$

and put $Z_n = \sum_{i=1}^n Z_{ni}$.

As the condition **X.5** gives us

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(c_i-c_j)(\mathbf{x}_i-\mathbf{x}_j)\,[f(e_i)-\gamma]=\sum_{i=1}^{n}\left[c_i\,\mathbf{x}_i+\tfrac{1}{n}\sum_{j=1}^{n}c_j\,\mathbf{x}_j\right](f(e_i)-\gamma),$$

it suffices to prove

$$\|\bar{Z}_n\|_T=o_p(1). \tag{3.7}$$

Similarly to the proof of Theorem 2.1, we verify that the assumptions of Corollary 7.13 are satisfied.

For convenience of notation put $\mathbf{y}_{ij}(n)=\frac{\mathbf{x}_i-\mathbf{x}_j}{\sqrt{n}}$, $\varepsilon_n=\max_{1\le i\le n}\frac{2M|\mathbf{x}_i|_2}{\sqrt{n}}$, and $r_n=\sup_{|s|\le\varepsilon_n}\mathsf{E}\,[f(e_1-s)-f(e_1)]^2$. Notice that the condition **W.2** implies $r_n\to 0$. With the help of Cauchy-Schwartz inequality we can bound

$$\begin{aligned}
\mathsf{E}\sup_{|\mathbf{t}-\mathbf{s}|_2<\varepsilon}|Z_{ni}(\mathbf{t})-Z_{ni}(\mathbf{s})|^2 &\le \mathsf{E}\left[\frac{1}{n}\sum_{j=1}^{n}|c_i-c_j|\sup_{|\mathbf{t}-\mathbf{s}|_2<\varepsilon}\int_{\mathbf{s}^{\mathsf{T}}\mathbf{y}_{ij}^{n}}^{\mathbf{t}^{\mathsf{T}}\mathbf{y}_{ij}^{n}}|f(e_i)-f(e_i-v)|\,dv\right]^2\\
&\le \frac{1}{n^2}\sum_{j=1}^{n}|c_i-c_j|^2\sum_{j=1}^{n}\mathsf{E}\sup_{|\mathbf{t}-\mathbf{s}|_2<\varepsilon}\left[\int_{\mathbf{s}^{\mathsf{T}}\mathbf{y}_{ij}^{n}}^{\mathbf{t}^{\mathsf{T}}\mathbf{y}_{ij}^{n}}|f(e_i)-f(e_i-v)|\,dv\right]^2\\
&\le \frac{1}{n^2}\sum_{j=1}^{n}|c_i-c_j|^2\sum_{j=1}^{n}\mathsf{E}\sup_{|\mathbf{t}-\mathbf{s}|_2<\varepsilon}|\mathbf{t}^{\mathsf{T}}\mathbf{y}_{ij}^{n}-\mathbf{s}^{\mathsf{T}}\mathbf{y}_{ij}^{n}|\int_{-\frac{M(|\mathbf{x}_i|_2+|\mathbf{x}_j|_2)}{\sqrt{n}}}^{+\frac{M(|\mathbf{x}_i|_2+|\mathbf{x}_j|_2)}{\sqrt{n}}}[f(e_i)-f(e_i-v)]^2\,dv\\
&\le \frac{1}{n^2}\sum_{j=1}^{n}|c_i-c_j|^2\sum_{j=1}^{n}\frac{\varepsilon\,|\mathbf{x}_i-\mathbf{x}_j|_2}{\sqrt{n}}\,\frac{2\,M(|\mathbf{x}_i|_2+|\mathbf{x}_j|_2)}{\sqrt{n}}\sup_{|v|\le\varepsilon_n}\mathsf{E}\,[f(e_i)-f(e_i-v)]^2\,dv\\
&\le \frac{2\,M\,\varepsilon}{n^3}\sum_{j=1}^{n}|c_i-c_j|^2\sum_{j=1}^{n}(|\mathbf{x}_i|_2+|\mathbf{x}_j|_2)^2\,r_n\\
&\le \frac{8\,M\,\varepsilon\,r_n}{n^3}\sum_{j=1}^{n}\left(|c_i|^2+|c_j|^2\right)\sum_{j=1}^{n}\left(|\mathbf{x}_i|_2^2+|\mathbf{x}_j|_2^2\right).
\end{aligned}$$

Thus by the assumptions of the theorem we can find a sufficiently large constant C such that for all $n\in\mathbb{N}$

$$\sum_{i=1}^{n}\mathsf{E}\sup_{|\mathbf{t}-\mathbf{s}|_2<\varepsilon}|Z_{ni}(\mathbf{t})-Z_{ni}(\mathbf{s})|^2\le C\,\varepsilon\,r_n=o(1),$$

which verifies the condition (7.5) of Corollary 7.13. The condition (7.6) trivially holds for $\mathbf{t}_0=\mathbf{0}$.

3.2.2 Asymptotic negligibility of R_n

In the following, we show that

$$\|R_n\|_T=\sup_{\mathbf{t}\in T}|R_n(\mathbf{t})|=o_p(1). \tag{3.8}$$

To prove (3.8), we adapt the theory of U-processes with the kernel of degree two introduced in Nolan and Pollard (1987). For convenience we denote this reference as NP. Let us recall that $R_n = \bar{T}_n - P_n$. With the help of (3.6) we can write $R_n(\mathbf{t}) = U_n(\mathbf{t}) - \mathsf{E}\, U_n(\mathbf{t})$, where

$$\begin{aligned} U_n(\mathbf{t}) &= \frac{1}{n}\sum_{i=1}^{n} c_i \sum_{j=1}^{n}\Big[\mathbb{I}\{e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} \geq e_j - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_j}{\sqrt{n}}\} - \mathbb{I}\{e_i \geq e_j\}\Big] \\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big[(c_i - c_j)(F(e_i - \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}) - F(e_i))\Big] = \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(e_i, e_j, \tfrac{\mathbf{t}}{\sqrt{n}}), \end{aligned} \tag{3.9}$$

with

$$g_{ij}(u, v, \mathbf{w}) = \tfrac{c_i}{n}\Big[\mathbb{I}\{u - \mathbf{w}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j) \geq v\} - \mathbb{I}\{u \geq v\}\Big] - \tfrac{c_i - c_j}{n}\Big[F(u - \mathbf{w}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)) - F(u)\Big].$$

We divide the proof into several steps, which are common in the theory of empirical processes. But before we proceed, we spend a few words about measurability. As the first part of the process U_n is a sum of indicators (depending on $\mathbf{t}$), we cannot hope that U_n, viewed as a mapping from Ω into $\ell^\infty(T)$, is measurable. But let us denote $S = \{\mathbf{s} : |\mathbf{s}|_2 \leq M+1,\ \mathbf{s} \in \mathbb{Q}_p\}$, where $\mathbb{Q}$ is the set of rational numbers. Then obviously

$$\sup_{\mathbf{t}\in T} |U_n(\mathbf{t}) - \mathsf{E}\, U_n(\mathbf{t})| \leq \sup_{\mathbf{s}\in S} |U_n(\mathbf{s}) - \mathsf{E}\, U_n(\mathbf{s})|,$$

where the quantity on the right hand side is a supremum over a countable set and thus measurable. In the following, let T stand for S. Then we do not need to worry about measurability, which is particularly important in Lemma 3.5, where we use Fubini Theorem.

Symmetrization

The first step is the symmetrization of the process $R_n(\mathbf{t})$. Let $e_1', \ldots, e_n'$ be independent copies of $e_1, \ldots, e_n$. Denote

$$\begin{aligned} U_n^{'}(\mathbf{t}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(e_i', e_j, \tfrac{t}{\sqrt{n}}), &\qquad R_n^{'}(\mathbf{t}) &= U_n^{'}(\mathbf{t}) - \mathsf{E}\, U_n^{'}(\mathbf{t}), \\ U_n^{\,'}(\mathbf{t}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(e_i, e_j', \tfrac{t}{\sqrt{n}}), &\qquad R_n^{\,'}(\mathbf{t}) &= U_n^{\,'}(\mathbf{t}) - \mathsf{E}\, U_n^{\,'}(\mathbf{t}), \\ U_n^{''}(\mathbf{t}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(e_i', e_j', \tfrac{t}{\sqrt{n}}), &\qquad R_n^{''}(\mathbf{t}) &= U_n^{''}(\mathbf{t}) - \mathsf{E}\, U_n^{''}(\mathbf{t}). \end{aligned}$$

With the help of these processes we define the symmetrized process

$$\begin{aligned} R_n^{sym}(\mathbf{t}) &= R_n(\mathbf{t}) - R_n^{'}(\mathbf{t}) - R_n^{\,'}(\mathbf{t}) + R_n^{''}(\mathbf{t}) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\Big[g_{ij}(e_i, e_j, \tfrac{\mathbf{t}}{\sqrt{n}}) - g_{ij}(e_i', e_j, \tfrac{\mathbf{t}}{\sqrt{n}}) - g_{ij}(e_i, e_j', \tfrac{\mathbf{t}}{\sqrt{n}}) + g_{ij}(e_i', e_j', \tfrac{\mathbf{t}}{\sqrt{n}})\Big] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}^{sym}(e_i, e_j, e_i', e_j', \tfrac{\mathbf{t}}{\sqrt{n}}), \end{aligned} \tag{3.10}$$

where

$$g_{ij}^{sym}(e_i, e_j, e_i', e_j', \tfrac{\mathbf{t}}{\sqrt{n}}) = g_{ij}(e_i, e_j, \tfrac{\mathbf{t}}{\sqrt{n}}) - g_{ij}(e_i', e_j, \tfrac{\mathbf{t}}{\sqrt{n}}) - g_{ij}(e_i, e_j', \tfrac{\mathbf{t}}{\sqrt{n}}) + g_{ij}(e_i', e_j', \tfrac{\mathbf{t}}{\sqrt{n}}).$$

Let $\sigma_1, \ldots, \sigma_n$ be Rademacher random variables, that is they are independent, identically distributed with $\mathrm{P}[\sigma_1 = 1] = \mathrm{P}[\sigma_1 = -1] = \frac{1}{2}$. Suppose further that $\sigma_1, \ldots, \sigma_n$ are independent of $(e_1, \ldots, e_n, e_1', \ldots, e_n')$. Then the process R_n^{sym} has the same distribution as the process

$$R_n^{\sigma}(\mathbf{t}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_i\, \sigma_j\, g_{ij}^{sym}(e_i, e_j, e_i', e_j', \tfrac{\mathbf{t}}{\sqrt{n}}).$$

Let us introduce the process

$$R_n^{\circ}(\mathbf{t}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_i\, \sigma_j\, g_{ij}(e_i, e_j, \tfrac{\mathbf{t}}{\sqrt{n}}).$$

It holds

$$\mathsf{E}\, \|R_n\|_T \le \mathsf{E}\, \|R_n^{sym}\|_T = \mathsf{E}\, \|R_n^{\sigma}\|_T \le 4\,\mathsf{E}\, \|R_n^{\circ}\|_T, \tag{3.11}$$

where the first inequality is a complete analogy of Lemma 1 in NP (the important thing is that the process $R_n(\mathbf{t})$ is degenerated in the sense that its projection is a zero process) and the second inequality follows simply by a triangular inequality. One more application of a triangular inequality yields

$$\begin{aligned}
\mathsf{E}\, \|R_n^{\circ}\|_T \le \mathsf{E} \sup_{\mathbf{t}\in T} &\left| \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_i\, \sigma_j\, \tfrac{c_i}{n} \left[\mathbb{I}\{e_i - \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}} \ge e_j - \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_j}{\sqrt{n}}\} - \mathbb{I}\{e_i \ge e_j\} \right] \right| \\
+ \mathsf{E} \sup_{\mathbf{t}\in T} &\left| \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_i\, \sigma_j\, \tfrac{c_i - c_j}{n} \left[F(e_i - \tfrac{\mathbf{t}^{\mathsf{T}}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}) - F(e_i) \right] \right| \\
&\overset{\text{Say}}{=} \mathsf{E}\, \|R_{n1}^{\circ}\|_T + \mathsf{E}\, \|R_{n2}^{\circ}\|_T
\end{aligned}$$

In the sequel, we show that $\mathsf{E}\, \|R_{n1}^{\circ}\|_T = o_P(1)$. The proof for the process R_{n2}° would be completely analogous.

Exponential inequality

The second step is an exponential inequality. Denote E_{σ} the operator of the expected value induced by the random variables $\sigma_1, \ldots, \sigma_n$ (we condition on the realizations of $e_1, \ldots, e_n$).

Lemma 3.4. *For each real square matrix $\mathbf{A} = [a_{ij}]$ with $\sum_{i=1}^{n}\sum_{j\ne i}^{n} a_{ij}^2 \le \frac{1}{4\pi^2}$ it holds*

$$\mathsf{E}_{\sigma} \exp\left(\sum_{i=1}^{n} \sum_{j=1, j\ne i}^{n} \sigma_i \sigma_j\, a_{ij} \right) \le \exp\left(\frac{\pi^2}{2} \sum_{i=1}^{n} \sum_{j=1, j\ne i}^{n} a_{ij}^2 \right).$$

Proof. This lemma follows easily by Lemma 3 of NP, where the lemma is proved for symmetric matrices.

$$\mathsf{E}_\sigma \exp\left(\sum_{i=1}^n \sum_{j=1,j\neq i}^n \sigma_i\,\sigma_j\,a_{ij}\right) = \mathsf{E}_\sigma \exp\left(\sum_{i=1}^n \sum_{j=1,j\neq i}^n \sigma_i\,\sigma_j \frac{a_{ij}+a_{ji}}{2}\right)$$
$$\overset{\text{Lemma 3 of NP}}{\leq} \exp\left(\sum_{i=1}^n \sum_{j=1,j\neq i}^n \frac{\pi^2}{2}\frac{(a_{ij}+a_{ji})^2}{4}\right) \leq \exp\left(\frac{\pi^2}{2}\sum_{i=1}^n \sum_{j=1,j\neq i}^n a_{ij}^2\right). \tag{3.12}$$

□

Chaining and the maximal inequality

In the third, step we make use of the technique known as chaining. Let the index set T is equipped with a pseudometric d. Then the covering number $N(\varepsilon, T, d)$ is the minimal number of balls of radius ε needed to cover the set T.

Suppose that the random variables $e_1, e_2, \ldots$ are defined on a common probability space Ω and for $1 \leq i, j \leq n$ put

$$f_{ij}(\mathbf{t}) = \tfrac{c_i}{n}\left[\mathbb{I}\{e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} \geq e_j - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_j}{\sqrt{n}}\} - \mathbb{I}\{e_i \geq e_j\}\right].$$

In fact, the f_{ij} are random functions depending on the realization of $\omega \in \Omega$. Let us define the random pseudometric d_ω on T as

$$d_\omega(\mathbf{t}, \mathbf{s}) = \left(\sum_{i=1}^n \sum_{j=1}^n [f_{ij}(\mathbf{t}) - f_{ij}(\mathbf{s})]^2\right)^{1/2}.$$

To make use of Lemma 7.11 of Appendix (which can be found e.g. in NP), we need to find a uniform upper bound for the (random) covering numbers $N(\varepsilon, T, d_\omega)$. For this purpose, we use the technique of pseudodimension introduced in Pollard (1990). For future convenience we denote this reference EP. Put

$$h_{ij}(\omega, \mathbf{s}) = \mathbb{I}\{e_i(\omega) - e_j(\omega) \geq \mathbf{s}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)\} - \mathbb{I}\{e_i(\omega) - e_j(\omega) \geq 0\}.$$

Using Lemma 4.4 of EP, we can deduce that the subset of the euclidean space $\mathbb{R}^{n(n-1)}$

$$\mathcal{H}_{n\omega} = \{(h_{ij}(\omega, \mathbf{s}),\ 1 \leq i \neq j \leq n),\ \mathbf{s} \in \mathbb{R}^p\}$$

has for all $\omega \in \Omega$ uniformly bounded pseudodimension. Now set

$$\boldsymbol{\alpha} = (\alpha_{ij}, 1 \leq i \neq j \leq n) = \left(\tfrac{|c_i|}{n}, 1 \leq i \neq j \leq n\right)$$

and let $\boldsymbol{\alpha} \odot \mathbf{h}$ stand for the pointwise product in $\mathbb{R}^{n(n-1)}$ with the k^{th} coordinate $\alpha_k h_k$. Then obviously

$$N(\varepsilon, T, d_\omega) \leq N(\varepsilon, \boldsymbol{\alpha} \odot \mathcal{H}_{n\omega}, |\cdot|_2). \tag{3.13}$$

Because $|h_{ij}(\omega, \mathbf{s})| \leq 1$, we can take the vector $\mathbf{H} = (1, 1, \ldots, 1)$ as the envelope for $\mathcal{H}_{n\omega}$. Notice that uniformly in ω and n

$$|\boldsymbol{\alpha} \odot \mathbf{H}|_2^2 = \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{c_i^2}{n^2} \leq \frac{1}{n} \sum_{i=1}^{n} c_i^2.$$

For simplicity, but without lost of generality, we will suppose that $\frac{1}{n}\sum_{i=1}^{n} c_i^2 \leq 1$. Now the Corollary 4.10 of EP guarantees the existence of universal constants A a W such that for any $\omega \in \Omega$, any $\varepsilon\,(0 < \varepsilon \leq 1)$ as well as any $n \in \mathbb{N}$

$$N(\varepsilon, \boldsymbol{\alpha} \odot \mathcal{H}_{n\omega}, |\cdot|_2) \leq A \left(\frac{1}{\varepsilon}\right)^{W}.$$

Combining this inequality with the inequality (3.13) yields $N(\varepsilon, T, d_\omega) \leq A(\frac{1}{\varepsilon})^W$. Thus we can bound the covering integral for $s < 1$ uniformly in ω and n by

$$\begin{aligned} J_{n\omega}(s) = \int_0^s \log(N(\varepsilon, T, d_\omega))\, d\varepsilon &\leq \int_0^s \log(A) - W\log(\varepsilon)\, d\varepsilon \\ &\leq s\log(A) + W \int_0^s \tfrac{1}{\sqrt{\varepsilon}}\, d\varepsilon \leq s\log(A) + 2\,W\sqrt{s}. \end{aligned} \tag{3.14}$$

Now we are ready to formulate the analogy of Theorem 6 of NP.

Lemma 3.5. *There exists a constant C such that for all $n \in \mathbb{N}$*

$$\mathsf{E}\, \|R_n^\circ\|_T \leq C\, \mathsf{E}\left[\theta_n + \sqrt{\theta_n}\right], \tag{3.15}$$

where $\theta_n = \frac{1}{4}\sup_{\mathbf{t}\in T} d_\omega(\mathbf{t}, \mathbf{0})$.

Proof. Set $\Psi(x) = \frac{1}{2}\exp(\frac{x}{2\pi} - \frac{1}{8})$. For a fixed ω we verify that the process $R_n^\circ(\mathbf{t})$ meets the conditions of Lemma 7.11. The only unobvious condition is *(ii)*. With the help of Lemma 3.4 we can calculate

$$\begin{aligned} &\mathsf{E}_\sigma \exp\left(\frac{|R_n^\circ(\mathbf{t}) - R_n^\circ(\mathbf{s})|}{2\pi\, d_\omega(\mathbf{t}, \mathbf{s})}\right) \\ &\qquad \leq \mathsf{E}_\sigma \exp\left(\frac{R_n^\circ(\mathbf{t}) - R_n^\circ(\mathbf{s})}{2\pi\, d_\omega(\mathbf{t}, \mathbf{s})}\right) + \mathsf{E}_\sigma \exp\left(\frac{-R_n^\circ(\mathbf{t}) + R_n^\circ(\mathbf{s})}{2\pi\, d_\omega(\mathbf{t}, \mathbf{s})}\right) \\ &\qquad \overset{\text{Lemma 3.4}}{\leq} 2\exp\left(\frac{\pi^2}{2}\, \frac{\sum_{i=1}^{n}\sum_{j\neq i}\left[f_{ij}(\mathbf{t}) - f_{ij}(\mathbf{s})\right]^2}{4\pi^2\, d_\omega(\mathbf{t}, \mathbf{s})^2}\right) = 2\exp(\tfrac{1}{8}). \end{aligned}$$

This yields the desired exponential inequality $\mathsf{E}_\sigma \Psi\left(\frac{|R_n^\circ(\mathbf{t}) - R_n^\circ(\mathbf{s})|}{d_\omega(\mathbf{t},\mathbf{s})}\right) \leq 1$. The choice $f_0 = \mathbf{0}$ in Lemma 7.11 implies

$$\begin{aligned} \mathsf{E}_\sigma \|R_n^\circ\|_T &\leq 8\int_0^{\theta_n} \psi^{-1}(N(\varepsilon, T, d_\omega))\, d\varepsilon \\ &\leq 8\int_0^{\theta_n} \frac{\pi}{4} + 2\pi\log\left[2N(\varepsilon, T, d_\omega)\right] d\varepsilon \leq C_1\left[\theta_n + J_{n\omega}(\theta_n)\right] \overset{(3.14)}{\leq} C_2\left[\theta_n + \sqrt{\theta_n}\right], \end{aligned}$$

where C_1 and C_2 are sufficiently large constants. Averaging out over the $\omega \in \Omega$ gives us the inequality (3.15). ∎

Let us denote $\varepsilon_n = \frac{2(M+1)\max_{1\le i\le n}|\mathbf{x}_i|_2}{\sqrt{n}}$. The condition **X.2** implies $\varepsilon_n = o(1)$. The simple bound

$$[f_{ij}(\mathbf{t})]^2 \le \frac{c_i^2}{n^2}\,\mathbb{I}\{|e_i - e_j| \le \varepsilon_n\},$$

yields

$$\mathsf{E}\ \theta_n^2 \le \sum_{i=1}^n \frac{c_i^2}{n}\,\mathsf{E}\ \mathbb{I}\{|e_1 - e_2| \le \varepsilon_n\} \xrightarrow[n\to\infty]{\textbf{X.1-2},\ \textbf{W.2}} 0.$$

But the last inequality implies that both $\mathsf{E}\ \theta_n$ and $\mathsf{E}\ \sqrt{\theta_n}$ converge to zero, which with the help of Lemma 3.5 implies $\mathsf{E}\ \|R_n^\circ\|_T \to 0$.

Finally using Markov's inequality (Lemma 7.1) and the inequality (3.11) gives us

$$P\left\{\|R_n\|_T > \varepsilon\right\} \le \tfrac{1}{\varepsilon}\,\mathsf{E}\ \|R_n\|_T \le \tfrac{4}{\varepsilon}\,\mathsf{E}\ \|R_n^\circ\|_T \xrightarrow[n\to\infty]{} 0. \tag{3.16}$$

3.3 Proof of Lemma 3.2

Directly by the definition of T_n we see that

$$\mathsf{E}\ T_n(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^n c_i \sum_{j=1}^n \int_{-\infty}^{+\infty} [F(y - \tfrac{\mathbf{t}^\mathsf{T}}{\sqrt{n}}(\mathbf{x}_i - \mathbf{x}_j)) - F(y)]\,dF(y).$$

Along the lines given in Antille (1976) we can calculate

$$\begin{aligned} D_n(\mathbf{t}) &= \frac{1}{n}\sum_{i=1}^n c_i \sum_{j=1}^n \int_{-\infty}^{+\infty} [F(y - \tfrac{\mathbf{t}^\mathsf{T}}{\sqrt{n}}(\mathbf{x}_i - \mathbf{x}_j)) - F(y)]\,dF(y) + \tfrac{\gamma\,\mathbf{t}^\mathsf{T}}{\sqrt{n}}\sum_{i=1}^n c_i\,\mathbf{x}_i \\ &= \frac{1}{n}\sum_{i=1}^n c_i \sum_{j=1}^n \int_{-\infty}^{+\infty}\int_0^{\frac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}}} -f(y-v)f(y) + f(y)^2 dv\,dy \\ &= \frac{1}{2n}\sum_{i=1}^n c_i \sum_{j=1}^n \int_{-\infty}^{+\infty}\int_0^{\frac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}}} [f(y-v) - f(y)]^2\,dv\,dy. \end{aligned} \tag{3.17}$$

With the help of the conditions **W.3** and **X.1-4** we get that for arbitrary $\varepsilon > 0$ and for all sufficiently large n

$$|D_n(\mathbf{t})| \le \frac{\varepsilon}{2n}\left|\sum_{i=1}^n |c_i| \sum_{j=1}^n \frac{|\mathbf{t}|_2^2}{n}|\mathbf{x}_i - \mathbf{x}_j|_2^2\right| \le \frac{\varepsilon M^2}{2n}\left|\sum_{i=1}^n |c_i|\left(|\mathbf{x}_i|_2^2 + \frac{1}{n}\sum_{j=1}^n |\mathbf{x}_j|_2^2\right)\right| = \varepsilon\,O(1)$$

uniformly in T, which proves the lemma.

3.4 Wilcoxon signed rank statistics

With the help of the techniques presented in the previous sections we are able to prove analogous results for a signed rank statistic. This will generalize the results of Antille (1976).

Recall that we are considering the linear regression model (1.1). Let $R_i^+(\mathbf{b})$ stand for the rank of $|Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i|$ among $|Y_1 - \mathbf{b}^\mathsf{T}\mathbf{x}_1|, \ldots, |Y_n - \mathbf{b}^\mathsf{T}\mathbf{x}_n|$. Then the signed rank estimator $\hat{\boldsymbol{\beta}}_n^+$ (based on Wilcoxon scores) of $\boldsymbol{\beta}$ is usually defined as

$$\hat{\boldsymbol{\beta}}_n^+ = \arg\min_{\mathbf{b}\in\mathbb{R}_p} D_n^+(\mathbf{b}), \quad \text{where} \quad D_n^+(\mathbf{b}) = \sum_{i=1}^n |Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i|\, R_i^+(\mathbf{b}). \tag{3.18}$$

Alternatively, we may define the estimator $\hat{\boldsymbol{\beta}}_n^+$ as the solution of the following minimization

$$\sum_{j=1}^p |S_{nj}^+(\mathbf{b})| := \min, \quad \text{where} \quad S_{nj}^+(\mathbf{b}) = \frac{1}{n^{3/2}} \sum_{i=1}^n (x_{ij} - \bar{x}_{nj})\,\mathrm{sign}(Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i)\, R_i^+(\mathbf{b}). \tag{3.19}$$

To explore the asymptotic properties of this estimator, it turns out to be useful to study the asymptotic behaviour of the processes

$$\tilde{S}_n(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^n c_i\,\mathrm{sign}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})\, R_i'(\mathbf{t}), \tag{3.20}$$

$$T_n(\mathbf{t}) = \tilde{S}_n^+(\mathbf{t}) - \tilde{S}_n^+(\mathbf{0}), \tag{3.21}$$

$$\bar{T}_n(\mathbf{t}) = T_n(\mathbf{t}) - \mathsf{E}\, T_n(\mathbf{t}), \tag{3.22}$$

where $\mathbf{t} = (t_1, \ldots, t_p)^\mathsf{T}$, and $R_i'(\mathbf{t})$ stands for the rank of the random variable $|e_i - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}|$ among $|e_1 - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_1}{\sqrt{n}}|, \ldots, |e_n - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_n}{\sqrt{n}}|$. As usual, we will index these processes by the set $T = \{\mathbf{s} \in \mathbb{R}^p : |\mathbf{s}|_2 \leq M\}$, where $|\cdot|_2$ stands for the euclidean norm, and M is an arbitrary large but fixed constant.

In addition to the assumptions **W.1-3**, we need the symmetry of the distribution of the errors.

W.4 The density of the distribution of the errors is symmetric, that is $f(x) = f(-x)$, for any $x \in \mathbb{R}$.

Similarly to Antille (1976), we can show that $T_n(\mathbf{t}) = T_{n1}(\mathbf{t}) + T_{n2}(\mathbf{t})$, where

$$T_{n1}(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^n \sum_{j=1, j\neq i}^n c_i \Big[\mathbb{I}\{|e_j - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_j}{\sqrt{n}}| < e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\} - \mathbb{I}\{|e_j| < e_i\} \\ + \mathbb{I}\{|e_j| < -e_i\} - \mathbb{I}\{|e_j - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_j}{\sqrt{n}}| < -e_i + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\}\Big]$$

and

$$T_{n2}(\mathbf{t}) = \frac{2}{n}\sum_{i=1}^n c_i \left[\mathbb{I}\{e_i > \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\} - \mathbb{I}\{e_i > 0\}\right].$$

Put $\varepsilon_n = \max_{1\le i\le n} \frac{M|\mathbf{x}_i|_2}{\sqrt{n}}$ and estimate

$$\mathsf{E}\ \|T_{n2}\|_T \le \frac{2}{n}\sum_{i=1}^{n} |c_i|\,|F(0) - F(-\varepsilon_n)| \xrightarrow[n\to\infty]{} 0.$$

Thus Markov's inequality (Lemma 7.1) gives us $\|T_{n2}\|_T = o_P(1)$. That is why we can only concentrate on the process T_{n1}. Along the lines of the second section of this chapter we approximate $\bar{T}_{n1} = T_{n1} - \mathsf{E}\ T_{n1}$ by its projection $P_n = \sum_{i=1}^{n} \mathsf{E}\left[\bar{T}_{n1}|e_i\right]$. It is a matter of simple but tedious algebraic manipulations to show that $P_n = V_n - \mathsf{E}\ V_n$, where

$$\begin{aligned} V_n(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} c_i \left[F\left(e_i + \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_j - \mathbf{x}_i)}{\sqrt{n}}\right) - F(e_i) - F\left(-e_i + \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_j + \mathbf{x}_i)}{\sqrt{n}}\right) + F(-e_i)\right] \\ + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} c_j \left[F\left(\left|e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\right| - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_j}{\sqrt{n}}\right) - F\left(\left|e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\right| + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_j}{\sqrt{n}}\right)\right]. \end{aligned}$$

Analogously to Section 3.2.1 we can prove that

$$\sup_{\mathbf{t}\in T}\left|P_n(\mathbf{t}) + \tfrac{2\,\mathbf{t}^\mathsf{T}}{\sqrt{n}}\sum_{i=1}^{n}\left[c_i\,\mathbf{x}_i + \tfrac{1}{n}\sum_{j=1}^{n} c_j\,\mathbf{x}_j\right](f(e_i) - \gamma)\right| = o_p(1). \tag{3.23}$$

Further, by the same technique as in Section 3.2.2 we can show the asymptotic negligibility of the remainder term $R_n = \bar{T}_{n1} - P_n$. Finally, repeating the idea of the proof of Lemma 3.2 gives us that uniformly in $\mathbf{t} \in T$

$$\mathsf{E}\ T_n(\mathbf{t}) = -\tfrac{2\,\gamma\,\mathbf{t}^\mathsf{T}}{\sqrt{n}}\sum_{i=1}^{n} c_i\,\mathbf{x}_i + o(1).$$

Combining the previous results, we get the second order asymptotic linearity of the Wilcoxon signed-rank statistics.

Theorem 3.6. *Under conditions **X.1-5** and **W.1-4** it holds uniformly in* $\mathbf{t} \in T$

$$\tilde{S}_n^+(\mathbf{t}) - \tilde{S}_n^+(\mathbf{0}) + \tfrac{2\,\gamma\,\mathbf{t}^\mathsf{T}}{\sqrt{n}}\sum_{i=1}^{n} c_i\,\mathbf{x}_i = -\tfrac{2\,\mathbf{t}^\mathsf{T}}{\sqrt{n}}\sum_{i=1}^{n}\left(c_i\,\mathbf{x}_i + \tfrac{1}{n}\sum_{j=1}^{n} c_j\,\mathbf{x}_j\right)(f(e_i) - \gamma) + o_p(1).$$

3.5 Further generalizations

Although R-estimators defined by (1.7) are robust against outlying Y-values, they remain sensitive to observations with outlying $\mathbf{x}$-values, or high leverage points. Sievers (1983) proposed to define the R-estimator as the minimum of the weighted loss function

$$D_n(\mathbf{b}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,|Y_i - Y_j - b^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)|, \tag{3.24}$$

where $\{w_{ij}\}$ are appropriate weights. As

$$\sum_{i=1}^{n}\sum_{j=1}^{n}|Y_i - Y_j - b^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)| = \sum_{i=1}^{n}(Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i)(R_i(\mathbf{b}) - \tfrac{n+1}{2}),$$

we see that the new weighted estimator coincides with the R-estimator based on the Wilcoxon scores for $w_{ij} \equiv 1$. Naranjo and Hettmansperger (1994) made some proposals how to choose the weights w_{ij} to achieve the robustness in the $\mathbf{x}$-space. Comparing the gradient of the loss function (3.24)

$$\mathbf{S}_n^w(\mathbf{b}) = \frac{\partial D_n(\mathbf{b})}{\partial \mathbf{b}} = -\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(\mathbf{x}_i - \mathbf{x}_j)\ \mathrm{sign}(Y_i - Y_j - \mathbf{b}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)) \tag{3.25}$$

with the process from (3.1), we see that the only difference is in the absence or presence of the weights w_{ij}. Thus, provided we impose some mild regularity conditions on the weights w_{ij}, we can use the methods of previous sections to explore asymptotic properties of $\mathbf{S}_n^w(\mathbf{b})$ of (3.25).

Generally, we believe that the technique presented in this thesis may be also useful for the processes in the form of weighted U-statistics of degree two, that is

$$T_n(\mathbf{t}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\, g(Y_i, Y_j, \mathbf{X}, \mathbf{t}), \qquad \mathbf{t} \in T,$$

where w_{ij} are weights and $\mathbf{X}$ may represent a design matrix.

Another estimator whose exploration leads us to processes of the above type is the generalized S-estimators (GS-estimators). This estimator, proposed by Croux et al. (1994), is defined as

$$\hat{\mathbf{b}} = \arg\min_{\mathbf{b}} s_n(\mathbf{b}),$$

with

$$s_n(\mathbf{b}) = \sup\left\{ s > 0, \binom{n}{2}^{-1} \sum_{i=1}^{n} \rho\left(\tfrac{Y_i - Y_j - b^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)}{s}\right) \geq k \right\},$$

where ρ is a loss function and k a tuning constant.

Chapter 4

Applications of SOAL

In the first part of this Section we derive a two-term von Mises expansions for M-estimators based on a smooth ψ function and for an R-estimator based on Wilcoxon scores. For M-estimators based on a step function ψ, we find the second order distribution.

In the second part we propose an alternative procedure of constructing a confidence interval (CI) for a single component of the vector parameter $\boldsymbol{\beta}$ in a linear model (1.1). We explore the asymptotic properties of this CI and compare it with the standard Wald type CI.

4.1 SO asymptotic representations

In what follows, we heavily utilize the expansions of Chapter 2.

4.1.1 M-estimators – an absolutely continuous ψ

First order results

The work on the second order asymptotic representations of M-estimators was initiated by Jurečková and Sen (1990). Their main interest was in comparison of an M-estimator defined as the root of the equation

$$\sum_{i=1}^{n} \psi(X_i - t) = 0$$

and its one-step approximations (one step of a Newton-Raphson iterative procedure starting from a consistent estimator). In the following, we will generalize their results to a regression estimator, which is defined as a root of the system of equations

$$\sum_{i=1}^{n} \mathbf{x}_i \, \psi(Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i) = \mathbf{0} \qquad \left(\text{or } \sum_{i=1}^{n} \mathbf{x}_i \, \psi\big(\tfrac{Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i}{S_n}\big) = \mathbf{0} \right). \tag{4.1}$$

But before we proceed, we recall some first order asymptotic results. For each of the following processes we assume that the corresponding conditions from the previous chapters hold. Of course, the first order results require weaker conditions, but as we are interested primarily in the second order properties, it is sufficient to use these stronger requirements. Moreover, we need some consistency conditions so that the parameter $\boldsymbol{\beta}$ is identifiable and consistent.

GenFx.1 (**GenSt.1**) The function $h(t) = \mathsf{E}\ \rho(e_1 - t)$ (or $h(t) = \mathsf{E}\ \rho(\frac{e_1-t}{S})$) has a unique minimum at $t = 0$.

XX.2 There exists a limit $(p \times p)$ matrix $\mathbf{V}$

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

which is positive definite.

We say that a (fixed scale) vector M-process

$$M_n(\mathbf{t}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i\, \psi(e_i - \tfrac{\mathbf{b}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}})$$

satisfies the first order asymptotic linearity (FOAL) if for an arbitrarily large but fixed constant M

$$\sup_{|\mathbf{t}|_2 \le M} \| M_n(\mathbf{t}) - M_n(\mathbf{0}) + \gamma_1 \mathbf{V}_n \mathbf{t}\, \| = o_p(1), \tag{4.2}$$

where $\|\cdot\|$ stands for the maximum norm. Once we have this FOAL result, we can use the technique of the proof of Theorem 5.5.1 of Jurečková and Sen (1996) to show that there exists a root $\hat{\boldsymbol{\beta}}_n$ of the system (4.1) such that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = O_p(1). \tag{4.3}$$

Now inserting $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$ in the FOAL of M-process (4.2) gives us the first order asymptotic representation (FOAR) for a M-estimator $\hat{\boldsymbol{\beta}}_n$

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \tfrac{\mathbf{V}_n^{-1}}{\gamma_1 \sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i\, \psi(e_i) + o_p(1). \tag{4.4}$$

Similarly we can handle studentized M-estimators. For simplicity of the following results, but with only a minor loss of generality, we will assume that the model includes an intercept, that is $x_{i1} = 1$ for $i = 1, \dots, n$. Moreover, we need to assume that the scale estimator is $\sqrt{n}$-consistent, i.e.

$$\sqrt{n}(\tfrac{S_n}{S} - 1) = O_p(1). \tag{4.5}$$

Then the FOAL linearity result gives us

$$\sup_{\|\mathbf{t}\| \le M} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \left[\psi\left((e_i - \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}})/S_n \right) - \psi\left(e_i/S\right) \right] + \gamma_1 \mathbf{V}_n \mathbf{t}\ + \gamma_{1e} \mathbf{v}_1 \sqrt{n}(\tfrac{S_n}{S} - 1) \right\| = o_p(1), \tag{4.6}$$

where $\mathbf{v}_1$ is the first column of the matrix $\mathbf{V}_n$ and γ_1, γ_{1e} are defined in (2.10) of Section 2.1.3.

Analogously to the case of a fixed scale M-estimator we can derive the first order asymptotic representation (FOAR)

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \frac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i\, \psi\left(\tfrac{e_i}{S}\right) - \tfrac{\gamma_{1e}}{\gamma_1}\, \sqrt{n}(\tfrac{S_n}{S} - 1)\, \mathbf{u}_1 + o_p(1), \tag{4.7}$$

where $\mathbf{u}_1 = (1, 0, \dots, 0)^{\mathsf{T}} \in \mathbb{R}_p$. Notice that only the asymptotic distribution of the intercept is influenced by the asymptotic distribution of the scale estimator.

Before we proceed, let us spend a few words about the choice of the scale estimator S_n. It is natural to require that this estimator is regression invariant and scale equivariant, that is

$$S_n(c\,(\mathbf{Y} + \mathbf{X}\mathbf{b})) = c\, S_n(\mathbf{Y}), \quad \text{for} \quad \mathbf{b} \in \mathbb{R}_p \quad \text{and} \quad c > 0.$$

Some possible choices of S_n are discussed in Welsh (1986) and Jurečková and Sen (1996) (see also Section 7.4 of Appendix).

Second order results

We are ready to derive the second order asymptotic expansion (a two-term von Mises expansion) of a regression M-estimator for an absolutely continuous ψ-function.

First, we restate Corollary 2.3, for the vector case by replacing the constants c_i by $\mathbf{x}_i$. Put

$$\mathbf{M}_n(\mathbf{t}) = \left(\sum_{i=1}^{n} x_{i1} \left[\psi(e_i - \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}) - \psi(e_i)\right], \dots, \sum_{i=1}^{n} x_{ip} \left[\psi(e_i - \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}) - \psi(e_i)\right] \right)^{\mathsf{T}}$$
$$= \sum_{i=1}^{n} \mathbf{x}_i \left[\psi(e_i - \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}) - \psi(e_i)\right]$$

and let $\mathbf{W}_n$ stand for the $(p^2 \times p)$ matrix written in the block form as $[\mathbf{W}_{n1}, \dots, \mathbf{W}_{np}]^{\mathsf{T}}$, where $\mathbf{W}_{nl} = \frac{1}{n} \sum_{i=1}^{n} x_{li}\, \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}}$ for $l = 1, \dots, p$. Further, by the symbol $\mathbf{t}^{\mathsf{T}}\mathbf{W}_n\mathbf{t}$ we will mean the vector $(\mathbf{t}^{\mathsf{T}}\mathbf{W}_{n1}\mathbf{t}, \dots, \mathbf{t}^{\mathsf{T}}\mathbf{W}_{np}\mathbf{t})^{\mathsf{T}}$.

Corollary 4.1. *Under conditions **XX.1** and **SmFx.1-3** it holds uniformly in* $\mathbf{t} \in T$

$$\mathbf{M}_n(\mathbf{t}) + \gamma_1 \sqrt{n}\, \mathbf{t}^{\mathsf{T}}\mathbf{V}_n = -\tfrac{\mathbf{t}^{\mathsf{T}}}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i\, \mathbf{x}_i^{\mathsf{T}} [\psi'(e_i) - \gamma_1] + \tfrac{\gamma_2}{2}\, \mathbf{t}^{\mathsf{T}}\mathbf{W}_n\mathbf{t} + o_p(1). \tag{4.8}$$

Suppose that $\hat{\boldsymbol{\beta}}_n$ is a $\sqrt{n}$-consistent estimator which satisfies the first order asymptotic representation (4.8). Thus we can insert $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$ in (4.8). With the help of (4.4) and after some algebra we get

$$\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i) \\
&\qquad = -\tfrac{1}{\sqrt{n}} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \mathbf{x}_i^{\mathsf{T}} [\psi'(e_i) - \gamma_1] \right\} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i) \right\} \\
&\qquad + \tfrac{\gamma_2}{2\,\gamma_1}\, \tfrac{\mathbf{V}_n^{-1}}{\sqrt{n}} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i) \right\}^{\mathsf{T}} \mathbf{W}_n \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i) \right\} + o_p(\tfrac{1}{\sqrt{n}}). \quad (4.9)
\end{aligned}$$

The last expansion simplifies if the symmetry condition **Sym** is satisfied. This condition implies $\gamma_2 = 0$, which further gives us

$$\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \frac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i) \\
&\qquad = -\tfrac{1}{\sqrt{n}} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \mathbf{x}_i^{\mathsf{T}} [\psi'(e_i) - \gamma_1] \right\} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i) \right\} \overset{\text{Say}}{=} \frac{1}{\sqrt{n}}\, \mathbf{A}_n \mathbf{d}_n, \quad (4.10)
\end{aligned}$$

where both quantities $\mathbf{A}_n$ $(\in \mathbb{R}_{p\times p})$ and $\mathbf{d}_n$ $(\in \mathbb{R}_p)$ are asymptotically multivariate normal. Moreover, as the symmetry condition **Sym** implies

$$\mathsf{E}\ \psi(e_1)[\psi'(e_1) - \gamma_1] = \mathsf{E}\ \psi(e_1)\psi'(e_1) = 0,$$

the quantities $\mathbf{A}_n$ and $\mathbf{d}_n$ are asymptotically independent.

Similarly, if the scale estimator is $\sqrt{n}$ consistent (see (4.5)) and conditions **SmSt.1-4** and **XX.1-2** are satisfied, then we can find the second order distribution representation for the studentized M-estimator by inserting $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$, $\sqrt{n}\log(\frac{S_n}{S})$ for u and $\mathbf{x}_i$ for c_i in (2.6). But as the resulting expression is rather awkward, we will assume the symmetry condition **Sym** to hold. Under the above assumptions we get

$$\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i/S) \\
&\qquad = -\tfrac{1}{\sqrt{n}} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \mathbf{x}_i^{\mathsf{T}} [\psi'(e_i/S) - \gamma_1] \right\} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i/S) \right\} \\
&\qquad\quad - \tfrac{1}{\sqrt{n}} \left\{ \sqrt{n}(\tfrac{S_n}{S} - 1)\, \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \left[\tfrac{e_i}{S}\, \psi'(e_i/S) \right] \right\} \\
&\qquad\quad + \tfrac{\gamma_{2e}+\gamma_1}{\gamma_1\sqrt{n}}\, \sqrt{n}(\tfrac{S_n}{S} - 1) \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i/S) \right\} + o_p(\tfrac{1}{\sqrt{n}}). \quad (4.11)
\end{aligned}$$

From the expansion (4.11) we see that although the first order asymptotic distribution of the studentized M-estimator of the slope does not depend on the asymptotic distribution of S_n, the second order distribution does even if the symmetry condition **Sym** holds.

Numerical illustration

To get an idea of how does the second order approximation of an M-estimator work, we performed a small numerical study. We worked with the simple linear regression model $Y_i = \beta_1 + \beta_2 x_i + e_i$. The design points $x_1, \ldots, x_n$ were generated from the uniform distribution $U[-1,1]$ and centered. The errors $e_1, \ldots, e_n$ were taken to be normally distributed. As an M-estimator we used the Huber estimator generated by the function $\psi = \max(\min(x,k), -k)$ with the tuning constant $k = 1.345$.

In the following, we compare the first order remainder term

$$\mathbf{Rem_1} = \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \tfrac{\mathbf{V}_n^{-1}}{\gamma_1 \sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \, \psi(e_i)$$

with the second order remainder term

$$\mathbf{Rem_2} = \mathbf{Rem_1} + \frac{1}{\sqrt{n}} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1 \sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \, \mathbf{x}_i^{\mathsf{T}} [\psi'(e_i) - \gamma_1] \right\} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1 \sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \, \psi(e_i) \right\}.$$

Notice that we do not need to calculate the second term on the right hand side of the expansion (4.9), as the symmetry condition **Sym** implies $\gamma_2 = 0$. By the above discussion we know that $\|\mathbf{Rem_1}\| = o_p(1)$ and $\|\sqrt{n}\, \mathbf{Rem_2}\| = o_p(1)$. Let us denote by R_1 and R_2 the second components (corresponding to β_2) of the quantities $\mathbf{Rem_1}$ and $\mathbf{Rem_2}$.

Table 1 shows estimated 10%, 20%,…,90%-quantiles for the sample sizes n = 20, 50, 100, 500, 1 000, and 5 000. The number of random samples was always 10 000. Comparing the columns with the quantities R_1 and R_2, we see that the two-term expansion really improves on the first order approximation of the quantity $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$. On the other hand it is worth noticing that the random variable $\sqrt{n}\, R_2$ converges to zero much more slowly than the quantity R_1. This is in an agreement with the results of Lachout and Paulauskas (2000), who studied the speed of convergence in the second order asymptotic results for M-estimators in the location case.

Some further experiments show that the convergence of $\sqrt{n}\, R_2$ is even slower if the errors are asymmetric or if we increase the number of explanatory variables. This indicates that there seems to be no point in deriving the third or even the fourth term in von Mises expansions.

Comparison with a one-step estimator

We can also use the asymptotic expansions from the previous chapters to find the two-term von Mises expansions for one-step M-estimators. This estimator, which can be viewed as an approximation to the M-estimator defined in (4.1), is constructed in the following way. Suppose $\hat{\boldsymbol{\beta}}_n^{(0)}$ to be an initial estimator of $\boldsymbol{\beta}$. We assume that this estimator is $\sqrt{n}$-consistent, that is $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(0)} - \boldsymbol{\beta}) = O_p(1)$.

Let us denote $r_i = Y_i - (\hat{\boldsymbol{\beta}}_n^{(0)})^{\mathsf{T}} \mathbf{x}_i$ (for $i = 1, \ldots, n$) the residuals from this initial fit. Then we define the one-step estimator $\hat{\boldsymbol{\beta}}_n^{(1)}$ as

$$\hat{\boldsymbol{\beta}}_n^{(1)} = \hat{\boldsymbol{\beta}}_n^{(0)} + \mathbf{H}_n^{-1} \mathbf{g}_n,$$

	n = 20			n = 50			n = 100		
q	R_1	R_2	$\sqrt{n}\,R_2$	R_1	R_2	$\sqrt{n}\,R_2$	R_1	R_2	$\sqrt{n}\,R_2$
0.1	-0.320	-0.225	-1.006	-0.210	-0.107	-0.760	-0.149	-0.062	-0.619
0.2	-0.195	-0.094	-0.422	-0.121	-0.050	-0.357	-0.087	-0.030	-0.297
0.3	-0.114	-0.042	-0.187	-0.068	-0.022	-0.159	-0.049	-0.014	-0.141
0.4	-0.050	-0.016	-0.072	-0.030	-0.008	-0.056	-0.022	-0.005	-0.049
0.5	-0.001	-0.001	-0.003	0.000	0.000	-0.001	0.000	0.000	0.000
0.6	0.053	0.014	0.061	0.029	0.007	0.047	0.021	0.005	0.050
0.7	0.111	0.040	0.179	0.066	0.021	0.148	0.049	0.015	0.149
0.8	0.188	0.090	0.403	0.120	0.048	0.338	0.088	0.031	0.313
0.9	0.315	0.217	0.969	0.204	0.101	0.711	0.154	0.064	0.641
	n = 500			n = 1 000			n = 5 000		
q	R_1	R_2	$\sqrt{n}\,R_2$	R_1	R_2	$\sqrt{n}\,R_2$	R_1	R_2	$\sqrt{n}\,R_2$
0.0	-0.066	-0.018	-0.402	-0.048	-0.011	-0.348	-0.022	-0.003	-0.238
0.1	-0.038	-0.009	-0.205	-0.028	-0.005	-0.173	-0.013	-0.002	-0.120
0.2	-0.021	-0.005	-0.103	-0.016	-0.003	-0.085	-0.007	-0.001	-0.057
0.3	-0.009	-0.002	-0.038	-0.007	-0.001	-0.033	-0.003	0.000	-0.020
0.4	0.001	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.001
0.5	0.010	0.002	0.035	0.007	0.001	0.029	0.003	0.000	0.024
0.6	0.022	0.004	0.094	0.015	0.003	0.081	0.007	0.001	0.061
0.7	0.037	0.009	0.190	0.027	0.005	0.169	0.012	0.002	0.122
0.8	0.064	0.017	0.378	0.048	0.011	0.345	0.022	0.003	0.235

Table 1: Comparison of the first and the second order remainder term for a fixed-scale M-estimator.

where

$$\mathbf{g}_n = \sum_{i=1}^{n} \mathbf{x}_i \, \psi\left(\tfrac{r_i}{S_n}\right)$$

and

$$\mathbf{H}_n = \tfrac{1}{S_n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}} \psi'\left(\tfrac{r_i}{S_n}\right) \qquad \text{or} \qquad \mathbf{H}_n = \tfrac{1}{S_n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \psi'\left(\tfrac{r_i}{S_n}\right) \right\}.$$

The first choice of the matrix $\mathbf{H}_n$ corresponds to the Newton-Raphson method and the second one to the scoring method. Welsh and Ronchetti (2002) discussed very carefully the asymptotic expansions for one-step estimators. They even formally derive the third order von Mises expansion in a more complex situation including different weights for observations Y_i and design points $\mathbf{x}_i$. In view of that, our results present a partial technical background for the formal derivations of Welsh and Ronchetti. We note that the second order von Mises expansions for the one-step estimators are rather complex in general, unless we assume the symmetry condition **Sym**.

An interesting question, initiated by Jurečková and Sen (1990), is the impact of the initial estimator on the distribution of the one-step estimator. The authors studied the case of M-estimators of location with fixed scale. One of their results states that if the symmetric condition **Sym** is satisfied and the initial estimator is $\sqrt{n}$-consistent, then the one-step estimator $\hat{\theta}_n^{(1)}$ is not only first order (FO) equivalent with the M-estimator $\hat{\theta}_n^M$ (that is $\sqrt{n}(\hat{\theta}_n^{(1)} - \hat{\theta}_n^M) = o_p(1)$) but also second order (SO) equivalent, that is

$$n(\hat{\theta}_n^{(1)} - \hat{\theta}_n^M) = o_p(1). \tag{4.12}$$

In what follows, we will suppose the scale estimator S_n is $\sqrt{n}$-consistent and the symmetry condition **Sym** is satisfied. Writing down two-term expansions for an M-estimator and a corresponding one-step M-estimator we can find out that the SO equivalence (4.12) holds for the studentized estimators of location as well. This is in a good agreement with the empirical findings of Andrews et al. (1972).

Some further straightforward but tedious algebra yields the generalization of the previous results to the regression settings. Suppose that the initial regression estimator $\hat{\boldsymbol{\beta}}_n^{(0)}$ is $\sqrt{n}$-consistent. Then the SO equivalence (4.12) is true for the (studentized) regression estimators too, provided we use the Newton-Raphson method. But the preceding statement is not generally true for the scoring method. For this method, even in a very simple case of fixed scale and the regression line going through the origin ($Y_i = \beta\, x_i + e_i$), we get

$$n(\hat{\beta}_n^{(1)} - \hat{\beta}_n^M) = \frac{1}{\gamma_1 V_n}\left[\sqrt{n}(\hat{\beta}_n^{(0)} - \hat{\beta}_n^M)\right]\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n (V_n - x_i^2)(\psi'(e_i) - \gamma_1)\right] + o_p(1),$$

where $V_n = \frac{1}{n}\sum_{i=1}^n x_i^2$. From the above expansion we see that $n(\hat{\beta}_n^{(1)} - \hat{\beta}_n^M) = O_p(1)$ unless the initial estimator $\hat{\beta}_n^{(0)}$ and M-estimator $\hat{\beta}_n^M$ are first order equivalent. This may happen, e.g. if we use an R-estimator which is FO equivalent with the target M-estimator (see Jurečková (1977)) as an initial estimator. As we have indicated by the simple example of a regression line going through the origin, when using the scoring we need two steps so that the resulting estimator is SO equivalent.

Remark 11. It should be stressed that the symmetry condition **Sym** is crucial for these results. The problem of the linear model with asymmetric distributions is that different estimators of the parameter $\boldsymbol{\beta}$ usually estimate different intercepts. If this happen to the initial estimator and the (target) M-estimator, then the one-step estimator is not even first order equivalent (see Simpson et al. (1992)).

4.1.2 M-estimators – a step function ψ

In this section we suppose that the function $\psi = \rho'$ is a nondecreasing step-function, that is

$$\psi(x) = \alpha_j \qquad \text{for} \quad q_j < x \leq q_{j+1}, \quad j = 0, 1, \ldots, m, \tag{4.13}$$

where $\alpha_0 < \alpha_1 < \ldots < \alpha_m$ are real numbers, $-\infty = q_0 < q_1 < \ldots < q_m < q_{m+1} = \infty$, and m being a positive integer.

Due to the discontinuity of the function ψ, there may not exist an exact root of the system of equations (4.1). That is why we define the M-estimator of $\boldsymbol{\beta}$ as a solution of the minimization problem

$$\sum_{i=1}^{n} \rho\left(\tfrac{Y_i - \mathbf{t}^\mathsf{T}\mathbf{x}_i}{S_n}\right) := \min, \qquad \text{where} \qquad \rho' = \psi. \tag{4.14}$$

As we suppose in this section that the ψ function is nondecreasing, the function ρ is convex, which implies that there exists a solution to the minimization problem (4.14).

Remark 12. Without the assumption of convexity of the function ρ several complications arise. First, it is rather complicated to prove $\sqrt{n}$-consistency of the estimator $\hat{\boldsymbol{\beta}}_n$. Second, it is nontrivial to show that

$$\tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi\left(\tfrac{Y_i - \hat{\boldsymbol{\beta}}_n^\mathsf{T}\mathbf{x}_i}{S_n}\right) = o_p(1) \quad \left(\text{or even } o_p\left(\tfrac{1}{n^{1/4}}\right)\right), \tag{4.15}$$

where the first equation is needed for the first order results and the equation to the quantity in the brackets is needed for the second order results. To the present knowledge of the author, the problems with the asymptotic behaviour of the regression M-estimators based on a discontinuous ψ function which is not monotone have not been completely solved. But we believe that the difficulties are only of technical character and they have not been solved yet, because these estimators are not in the centre of attention. Nevertheless, once we have shown the $\sqrt{n}$-consistency of the estimator $\hat{\boldsymbol{\beta}}_n$ and the equation (4.15), our further results hold.

Similarly to the case of a smooth ψ function, we would like use the FOAL of the M-process (4.6) to prove the FOAR for the M-estimator (4.7). But due to the discontinuity of the function ψ, it is not sufficient to have the $\sqrt{n}$-consistency of the estimates of regression parameters (and the scale estimator), but we need the equation (4.15) to hold as well.

To show that, we adopt the technique of Jurečková and Sen (1996) (see pp. 167 for details). Let us denote

$$G_j(t) = \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho\left((Y_i - \hat{\boldsymbol{\beta}}_n^\mathsf{T}\mathbf{x}_i - t\, x_{ij})/S_n\right),$$

and G_j^+ the right derivative of this function. As the function ψ is nondecreasing, the function

$$G_j^+(t) = \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{ij}\, \psi\left((Y_i - \hat{\boldsymbol{\beta}}_n^\mathsf{T}\mathbf{x}_i - t\, x_{ij})/S_n\right),$$

is nondecreasing in t. Using further the fact that the function G_j has its minimum in 0, we conclude that for every $\varepsilon > 0$ $|G_j^+(0)| \leq G_j^+(\varepsilon) - G_j^+(-\varepsilon)$. Letting $\varepsilon \searrow 0$ implies

$$|G_j^+(0)| \leq \tfrac{p}{\sqrt{n}} \left(\max_{0 \leq l \leq m} |\alpha_l| \max_{1 \leq i \leq n} |x_{ij}|\right) \overset{\mathbf{X}\underline{\mathbf{X}}\text{'}.\mathbf{1}}{=} o\left(\tfrac{1}{n^{1/4}}\right),$$

where the second inequality holds almost surely due to the continuity of the cdf of the errors F (assumption **Step.1**). This gives us the desired result

$$\tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \, \psi\left(\tfrac{Y_i - \hat{\boldsymbol{\beta}}_n^{\mathsf{T}} \mathbf{x}_i}{S_n}\right) = o(\tfrac{1}{n^{1/4}}) \qquad \text{almost surely.} \tag{4.16}$$

The next step is to show the $\sqrt{n}$-consistency of the estimator $\hat{\boldsymbol{\beta}}_n$. There are basically two techniques to use. We can either exploit the monotonicity of ψ or use some convexity arguments (see Jurečková and Sen (1996), pp. 167–169), or pp. 223–224). Once we have the $\sqrt{n}$-consistency and (4.16), we can insert $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$ in (4.6) and get the FOAR (4.7).

Second order results

To simplify the following formulae, we will only consider the fixed-scale M-estimators. Let us denote $\gamma_1 = \sum_{j=1}^{m} \beta_j f(q_j)$. Unlike the case of a smooth ψ-function, for which we have derived a two-term von Mises expansion, the best we can do here is to find the asymptotic distribution of the random variable

$$n^{1/4} \left(\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \frac{1}{\gamma_1 \sqrt{n}} \sum_{i=1}^{n} \mathbf{V}_n^{-1} \mathbf{x}_i \, \psi(e_i) \right).$$

Before we do that, we need some auxiliary results.

Lemma 4.2. *Suppose that the conditions **XX'.1**, **XX.2**, and **Step.1-2** are satisfied and there exists a (matrix) function $\mathbf{r} : T \times T \to \mathbb{R}_{p\times p}$ such that for every $\mathbf{t}$, $\mathbf{s} \in T$*

$$\lim_{n\to\infty} \mathbf{V}_n^{-1} \sum_{i=1}^{n} \tfrac{\mathbf{x}_i \mathbf{x}_i^{\mathsf{T}}}{n} \min(|\mathbf{t}^{\mathsf{T}} \mathbf{x}_i|, |\mathbf{s}^{\mathsf{T}} \mathbf{x}_i|) \, \mathbb{I}\{\mathbf{t}^{\mathsf{T}} \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}} \mathbf{s} > 0\} \, \mathbf{V}_n^{-1} = \mathbf{r}(\mathbf{t}, \mathbf{s}). \tag{4.17}$$

Then the vector process

$$\mathbf{Z}_n = (Z_n^1, \ldots, Z_n^p)^{\mathsf{T}} = \frac{1}{\gamma_1 \, n^{1/4}} \sum_{i=1}^{n} \mathbf{V}_n^{-1} \mathbf{x}_i \left[\psi(e_i - \tfrac{\mathbf{t}^{\mathsf{T}} \mathbf{x}_i}{\sqrt{n}}) - \psi(e_i) \right] - n^{1/4} \mathbf{t}, \qquad \mathbf{t} \in T \tag{4.18}$$

converges to a p-dimensional centered gaussian process $\mathbf{Z}$, with the covariance structure

$$\operatorname{cov}\{\mathbf{Z}(\mathbf{t}), \mathbf{Z}(\mathbf{s})\} = \mathbf{r}(\mathbf{t}, \mathbf{s}).$$

Proof. For simplicity of notation suppose that $p = 2$ and let us denote $X_n = Z_n^1$ and $Y_n = Z_n^2$. From Corollary 2.9 we know that both X_n and Y_n converge weakly to tight gaussian processes in the space $\ell^\infty(T)$ of bounded functions. By 1.3.8 Lemma of van der Vaart and Wellner (1996) (VW), both sequences $\{X_n, n \in \mathbb{N}\}$ and $\{Y_n, n \in \mathbb{N}\}$ are asymptotically tight. Further by 1.4.3 Lemma and 1.4.4 Lemma of VW, the sequence $\{(X_n, Y_n), n \in \mathbb{N}\}$ is asymptotically tight as well as asymptotically measurable. Now by Prohorov's theorem (1.3.9 Theorem of VW) there exists a subsequence $\{(X_{n_j}, Y_{n_j}), j \in \mathbb{N}\}$ that converges weakly to a tight Borel law, say $\mathbf{Z} = (X, Y)^\mathsf{T}$. It only remains to show that this limit is unique.

For this purpose, let us consider the collection $\mathcal{F}$ of all functions $f : \ell^\infty(T) \times \ell^\infty(T) \mapsto \mathbb{R}$ of the form

$$h(\mathbf{x}, \mathbf{y}) = f(x(t_1), \ldots, x(t_k))\, g(y(s_1), \ldots, y(s_l)),$$

where f and g are continuous and bounded real functions on $\mathbb{R}^k$ and $\mathbb{R}^l$ respectively, and $t_1, \ldots, t_k \in T$, $s_1, \ldots, s_l \in T$, $k \in \mathbb{N}$ and $l \in \mathbb{N}$. The collection $\mathcal{F}$ forms a vector lattice (vector space that is closed under taking positive parts), contains constant functions, and separates points of the space $\ell^\infty(T) \times \ell^\infty(T)$. By 1.3.12 Lemma of VW a Borel measure L on $\ell^\infty(T) \times \ell^\infty(T)$ is uniquely determined by the expectations $\{\int f\, dL : f \in \mathcal{F}\}$. That is why we can conclude that $\mathbf{Z}_n = (X_n, Y_n)^\mathsf{T}$ converges weakly to the process $\mathbf{Z} = (X, Y)^\mathsf{T}$, if for all $k, l \in \mathbb{N}$ and $t_1, \ldots, t_k, s_1, \ldots, s_l \in T$

$$(X_n(t_1), \ldots, X_n(t_k), Y_n(s_1), \ldots, Y_n(s_l)) \xrightarrow{w} (X(t_1), \ldots, X(t_k), Y(s_1), \ldots, Y(s_l)).$$

But this weak convergence follows follows easily by the assumptions of the lemma. $\square$

Theorem 4.3. *Suppose that the conditions of Lemma 4.2 are satisfied and the FOAR* (4.4) *holds for $\hat{\boldsymbol{\beta}}_n$. Then the random vector*

$$n^{1/4}\left(\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \frac{1}{\gamma_1\sqrt{n}}\sum_{i=1}^{n} \mathbf{V}_n^{-1}\mathbf{x}_i\, \psi(e_i)\right)$$

converges in distribution to the random variable $\mathbf{Z}(\mathbf{W})$, where the random vector $\mathbf{W}$ has the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ and $\mathbf{Z}$ is the limiting process of Lemma 4.2.

Proof. Let $\mathbf{Z}_n$ be defined by (4.18) and put $\mathbf{W}_n = \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$. Then following the proof of Lemma 4.2, we can show that the pair $(\mathbf{Z}_n, \mathbf{W}_n)$ converges in distribution to $(\mathbf{Z}, \mathbf{W})$ in the product space $[\prod_{i=1}^p \ell^\infty(T)] \times \mathbb{R}_p$.

Recall that $T = \{\mathbf{t} \in \mathbb{R}_p : |\mathbf{t}|_2 \leq M\}$, where M is an arbitrary large but fixed constant. Put $[\mathbf{W}_n]_M = \mathbf{W}_n\, \mathbb{I}\{|\mathbf{W}_n|_2 \leq M\}$. We claim that, $\mathbf{Z}_n([\mathbf{W}_n]_M)$ weakly converges to $\mathbf{Z}([\mathbf{W}]_M)$. This statement will follow from the continuous mapping theorem (see 1.3.6 Theorem of van der Vaart and Wellner (1996)), provided we show that the mapping $\psi : [\prod_{i=1}^p \ell^\infty(T)] \times T \mapsto \mathbb{R}_p$ defined by $\varphi(x, \phi) = x(\phi)$ is continuous on a subset $\mathbb{D}_0 \subset [\prod_{i=1}^p \ell^\infty(T)] \times T$ such that $\mathrm{P}\{(\mathbf{Z}, \mathbf{W}) \in \mathbb{D}_0\} = 1$.

Put $\mathbb{D}_0 = C^p(T) \times \mathbb{R}_p$, where $C^p(T)$ is a space of continuous (and therefore bounded) vector functions on T. As $\mathbf{Z}$ is a p-variate Gaussian process, then $\mathrm{P}\{(\mathbf{Z}, [\mathbf{W}]_M) \in \mathbb{D}_0\} = 1$. Further let $(x, \phi) \in \mathbb{D}_0$ and assume that x_n converges to x (in $\prod_{i=1}^p \ell^\infty(T)$) and ϕ_n to ϕ (in T). We can estimate

$$\begin{aligned}|\varphi_n - \varphi|_2 = |x_n(\phi_n) - x(\phi)|_2 &\leq |x_n(\phi_n) - x(\phi_n)|_2 + |x(\phi_n) - x(\phi)|_2 \\ &\leq \||x_n - x|_2\|_T + |x(\phi_n) - x(\phi)|_2 \xrightarrow[n\to\infty]{} 0,\end{aligned}$$

because x is a continuous function on T. Thus φ is continuous on $\mathbb{D}_0$.

Now, let $\varepsilon > 0$ be given. As $\hat{\boldsymbol{\beta}}_n$ satisfies the expansion (4.4), we can find $M > 0$ and n_0 such that for all $n \geq n_0$

$$\mathrm{P}\,\{[\mathbf{W}_n]_M \neq \mathbf{W}_n\} = \mathrm{P}\left\{|\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})|_2 > M\right\} < \varepsilon, \quad \text{and} \quad \mathrm{P}\,\{|\mathbf{W}|_2 > M\} < \varepsilon.$$

Further fix $\boldsymbol{\lambda} \in \mathbb{R}_p$ ($|\boldsymbol{\lambda}|_2 = 1$). Then for each $z \in \mathbb{R}$

$$\begin{aligned}\limsup_{n\to\infty} \mathrm{P}\left\{\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}_n(\mathbf{W}_n) \leq z\right\} &\leq \limsup_{n\to\infty} \mathrm{P}\left\{\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}_n([\mathbf{W}_n]_M) \leq z\right\} + \varepsilon \\ &= \mathrm{P}\left\{\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}([\mathbf{W}]_M) \leq z\right\} + \varepsilon \leq \mathrm{P}\left\{\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}(\mathbf{W}) \leq z\right\} + 2\,\varepsilon. \end{aligned} \tag{4.19}$$

Similarly

$$\liminf_{n\to\infty} \mathrm{P}\left\{\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}_n(\mathbf{W}_n) \leq z\right\} \geq \mathrm{P}\left\{\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}(\mathbf{W}) \leq z\right\} - 2\,\varepsilon. \tag{4.20}$$

As $\varepsilon > 0$ was arbitrary, the equations (4.19) and (4.20) together imply

$$\lim_{n\to\infty} \mathrm{P}\left\{\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}_n(\mathbf{W}_n) \leq z\right\} = \mathrm{P}\left\{\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}(\mathbf{W}) \leq z\right\}.$$

Finally, the Cramér-Wold device (Theorem 7.5 of Appendix) yields that $\mathbf{Z}_n(\mathbf{W}_n)$ converges in distribution to $\mathbf{Z}(\mathbf{W})$. But this is just the statement of the theorem. $\square$

4.1.3 R-estimators based on Wilcoxon scores

Let us remind that

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\mathbf{b}\in\mathbb{R}_p} D_n(\mathbf{b}), \quad \text{where} \quad D_n(\mathbf{b}) = \sum_{i=1}^{n}(Y_i - \mathbf{b}^{\mathsf{T}}\mathbf{x}_i)\left(\tfrac{R_i(\mathbf{b})}{n+1} - \tfrac{1}{2}\right), \tag{4.21}$$

and $R_i(\mathbf{b})$ is the rank of $Y_i - \mathbf{b}^{\mathsf{T}}\mathbf{x}_i$ among $Y_1 - \mathbf{b}^{\mathsf{T}}\mathbf{x}_1, \ldots, Y_n - \mathbf{b}^{\mathsf{T}}\mathbf{x}_n$.

Our aim is to use the asymptotic expansion of the previous chapter to derive the second order asymptotic representation for the estimator $\hat{\boldsymbol{\beta}}_n$. But before that, we need to recall some first order results. If conditions **W.1**, **X.2**, and **XX.2** are satisfied, then the estimator $\hat{\boldsymbol{\beta}}_n$ admits the following first order asymptotic representation (e.g. Ren (1994))

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}} \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})\,[F(e_i) - \tfrac{1}{2}] + o_p(1). \tag{4.22}$$

For simplicity of notation we will suppose in the following that

$$\sum_{i=1}^{n}\mathbf{x}_i = \mathbf{0}. \tag{4.23}$$

Otherwise we would work with $\mathbf{x}_i - \bar{\mathbf{x}}$.

Second order results

As a preliminary step, it is convenient to restate Corollary 3.3 for a vector process

$$\tilde{\mathbf{S}}_n(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\, R_i^{'}(\mathbf{t}), \qquad \mathbf{t} \in T,$$

where $R_i^{'}(\mathbf{t})$ stands for the rank of $e_i - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}$ among $e_1 - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_1}{\sqrt{n}}, \dots, e_n - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_n}{\sqrt{n}}$.

Corollary 4.4. *Under conditions **XX.1**, **XX.2**, and **W.1-3** it holds uniformly in* $\mathbf{t} \in T$

$$\tilde{\mathbf{S}}_n(\mathbf{t}) - \tilde{\mathbf{S}}_n(\mathbf{0}) + \gamma\sqrt{n}\,\mathbf{V}_n\mathbf{t} = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\mathbf{x}_i\,\mathbf{x}_i^\mathsf{T} + \mathbf{V}_n\Big)\left(f(e_i) - \gamma\right)\mathbf{t} + o_p(1). \tag{4.24}$$

Now we would like to insert $\mathbf{t} \to \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ into the equation (4.24). But before we do that, we need to verify that $\tilde{\mathbf{S}}_n(\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})) = \sqrt{n}\,\mathbf{S}_n(\hat{\boldsymbol{\beta}}_n)$ is sufficiently small (for the definition of $\mathbf{S}_n(\mathbf{b})$ see (1.7) or List of symbols on page 112).

It is well known (e.g. Hettmansperger (1984)) that we can rewrite the quantity $D_n(\mathbf{b})$ as

$$D_n(\mathbf{b}) = \sum_{i=1}^{n}(Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i)(R_i(\mathbf{b}) - \tfrac{n+1}{2}) = \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}|Y_i - Y_j - \mathbf{b}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)|. \tag{4.25}$$

Put $Z_{ij} = Y_i - Y_j$ and $\mathbf{d}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ and replace the indices i, j with a single index l. We get $D_n(\mathbf{b}) = 2\sum_{l=1}^{\binom{n}{2}}|Z_l - \mathbf{b}^\mathsf{T}\mathbf{d}_l|$. With the help of this representation of $D_n(\mathbf{b})$ we can proceed analogously to Section 4.1.2 (for details see Jurečková and Sen (1996), pp. 167) and show that for all $j = 1, \dots, p$

$$|S_{nj}(\hat{\boldsymbol{\beta}}_n)| = \left|\frac{1}{n^{3/2}}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)R_i(\hat{\boldsymbol{\beta}}_n)\right| \overset{\text{a.s}}{\leq} \frac{4\,p\max_{1\leq i\leq n}|\mathbf{x}_i|_2}{n^{3/2}} \overset{\mathbf{XX.1}}{=} o(\tfrac{1}{n}).$$

Thus $\sqrt{n}\,\mathbf{S}_n(\hat{\boldsymbol{\beta}}_n) = o(\frac{1}{\sqrt{n}})$ almost surely.

Remark 13. By a very similar argument we could show that the signed rank statistics satisfies $\sqrt{n}\,\mathbf{S}_n^{+}(\hat{\boldsymbol{\beta}}_n^{+}) = O(\frac{1}{\sqrt{n}})$ almost surely as well.

Now we are ready to insert $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$ in (4.24). After some reorganization we get

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}}\sum_{i=1}^{n}\mathbf{x}_i\,\tfrac{R_i(\boldsymbol{\beta})}{n} &= -\tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}}\sum_{i=1}^{n}\Big(\mathbf{x}_i\,\mathbf{x}_i^\mathsf{T} + \mathbf{V}_n\Big)\left(f(e_i) - \gamma\right)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + o_p(\tfrac{1}{\sqrt{n}}) \\ &\overset{(4.22)}{=} -\tfrac{1}{\sqrt{n}}\left\{\tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}}\sum_{i=1}^{n}\Big(\mathbf{x}_i\,\mathbf{x}_i^\mathsf{T} + \mathbf{V}_n\Big)\left(f(e_i) - \gamma\right)\right\}\left\{\tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}}\sum_{i=1}^{n}\mathbf{x}_i\,F(e_i)\right\} + o_p(\tfrac{1}{\sqrt{n}}).\end{aligned} \tag{4.26}$$

Specially, suppose that the l^{th}-column of the matrix $\mathbf{X}_n$ is orthogonal to the other columns of the matrix $\mathbf{X}_n$, that is $\sum_{i=1}^{n} x_{il}\,x_{ij} = 0$ for $l \neq j$. This implies that the l^{th}-column of the matrix $\mathbf{V}_n$ is given by $T_{nl}^2\,\mathbf{e_l}$, where $\mathbf{e_l} = (0, \dots, 0, 1, 0, \dots, 0)^\mathsf{T}$ is a vector of zeros with the only one nonzero element in the l^{th} coordinate and $T_{nl}^2 = \frac{1}{n}\sum_{i=1}^n x_{il}^2$. Now, taking only the l-th component of the vector equation (4.26) gives us the second order representation for the l-th component of $\hat{\boldsymbol{\beta}}_n$

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}_l - \beta_l) &- \frac{1}{\gamma\, T_{nl}^2 \sqrt{n}} \sum_{i=1}^n x_{il}\, \tfrac{R_i(\boldsymbol{\beta})}{n} \\
&= -\frac{1}{\sqrt{n}} \left\{ \tfrac{1}{\gamma T_{nl}^2 \sqrt{n}} \sum_{i=1}^n \left(\mathbf{x}_i\, x_{il} + \mathbf{e_l}\, T_{nl}^2\right)^{\mathsf{T}} (f(e_i) - \gamma) \right\} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \,[F(e_i) - \tfrac{1}{2}] \right\} + o_p(\tfrac{1}{\sqrt{n}}), \\
&\overset{\text{say}}{=} \tfrac{1}{\sqrt{n}}\, \mathbf{A}_n^{\mathsf{T}}\, \mathbf{B}_n + o_p(\tfrac{1}{\sqrt{n}}). \qquad (4.27)
\end{aligned}
$$

The random vectors $\mathbf{A}_n$, $\mathbf{B}_n$ have asymptotically multivariate normal distributions and their asymptotic covariance is

$$
\begin{aligned}
\lim_{n\to\infty} \operatorname{cov}\{\mathbf{A}_n, \mathbf{B}_n\} &= \lim_{n\to\infty} \tfrac{\mathbf{V}_n^{-1}}{\gamma^2 T_{nl}^2\, n} \sum_{i=1}^n \left[\mathbf{x}_i \mathbf{x}_i^{\mathsf{T}}\, x_{il} + \mathbf{x}_i \mathbf{e_l}^{\mathsf{T}} T_{nl}^2\right] \operatorname{cov}\{F(e_i), f(e_i)\} \\
&= \tfrac{\mathbf{V}^{-1}}{\gamma^2} \lim_{n\to\infty} \tfrac{1}{T_{nl}^2\, n} \sum_{i=1}^n \left[\mathbf{x}_i \mathbf{x}_i^{\mathsf{T}}\, x_{il}\right] \operatorname{cov}\{F(e_i), f(e_i)\}.
\end{aligned}
$$

Thus $\mathbf{A}_n$, $\mathbf{B}_n$ are asymptotically independent if $\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i^{\mathsf{T}} \mathbf{x}_i\, x_{il} \to 0$ or $\operatorname{cov}\{F(e_i), f(e_i)\} = 0$. The second condition is certainly satisfied if the distribution of the errors is symmetric.

Numerical illustration

To get an idea of how does the second order approximation of an R-estimator works we performed a small numerical study. We worked with a linear regression model with two explanatory variables $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$. The design points $\mathbf{x}_1 = (x_{11}, x_{12}), \ldots, \mathbf{x}_n = (x_{n1}, x_{n2})$ were generated from the uniform distribution $U[-1, 1]$ and centered in each coordinate. The errors $e_1, \ldots, e_n$ were taken to be normally distributed.

In the following, we focus on the coefficient β_1. Let R_1 be the first component of the vector

$$
\mathbf{Rem_1} = \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, F(e_i)
$$

and R_1^* be the first component of

$$
\mathbf{Rem_1^*} = \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \tfrac{R_i}{n}.
$$

We compare these first order remainder terms with the second order remainder term R_2 which is the first component of the vector

$$
\begin{aligned}
\mathbf{Rem_2} &= \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \tfrac{R_i(\boldsymbol{\beta})}{n} \\
&\quad + \tfrac{1}{\sqrt{n}} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}} \sum_{i=1}^n \left(\mathbf{x}_i\, \mathbf{x}_i^{\mathsf{T}} + \mathbf{V}_n\right) (f(e_i) - \gamma) \right\} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, F(e_i) \right\}. \qquad (4.28)
\end{aligned}
$$

From the theory presented above we know that $R_1 = o_p(1)$, as well as $R_1^* = o_p(1)$, and $\sqrt{n}\, R_2 = o_p(1)$.

Table 2 shows estimated 10%, 20%,…,90%-quantiles of the quantities R_1^*, R_1, R_2, and $\sqrt{n}\, R_2$ for the sample sizes n = 20, 100, 500, and 2 000. The number of random samples was always 10 000.

	n = 20				**n = 100**			
q	R_1^*	R_1	R_2	$\sqrt{n}\, R_2$	R_1^*	R_1	R_2	$\sqrt{n}\, R_2$
0.1	-0.638	-0.531	-0.514	-2.299	-0.261	-0.247	-0.171	-1.709
0.2	-0.401	-0.339	-0.330	-1.478	-0.153	-0.156	-0.109	-1.086
0.3	-0.242	-0.210	-0.198	-0.886	-0.090	-0.095	-0.065	-0.647
0.4	-0.117	-0.100	-0.094	-0.418	-0.044	-0.044	-0.031	-0.306
0.5	-0.005	0.000	-0.001	-0.006	-0.003	0.001	0.001	0.007
0.6	0.113	0.099	0.096	0.431	0.040	0.047	0.033	0.327
0.7	0.240	0.205	0.199	0.892	0.090	0.097	0.068	0.679
0.8	0.399	0.342	0.330	1.475	0.151	0.158	0.110	1.096
0.9	0.643	0.538	0.521	2.329	0.251	0.248	0.174	1.736
	n = 500				**n = 2 000**			
q	R_1^*	R_1	R_2	$\sqrt{n}\, R_2$	R_1^*	R_1	R_2	$\sqrt{n}\, R_2$
0.1	-0.095	-0.103	-0.051	-1.140	-0.044	-0.052	-0.020	-0.897
0.2	-0.055	-0.064	-0.032	-0.706	-0.026	-0.033	-0.013	-0.562
0.3	-0.031	-0.040	-0.019	-0.436	-0.015	-0.020	-0.008	-0.342
0.4	-0.014	-0.020	-0.009	-0.208	-0.006	-0.010	-0.004	-0.164
0.5	0.001	-0.001	0.000	0.008	0.000	0.000	0.000	-0.003
0.6	0.016	0.019	0.010	0.213	0.007	0.010	0.004	0.159
0.7	0.032	0.040	0.020	0.442	0.015	0.021	0.007	0.330
0.8	0.056	0.066	0.033	0.731	0.026	0.033	0.012	0.544
0.9	0.096	0.102	0.052	1.166	0.045	0.052	0.020	0.894

Table 2: Comparison of the first and second order remainder term for an R-estimator based on Wilcoxon scores

Comparing the columns with the quantities R_1^* and R_1, we see that while the representation $\frac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, F(e_i)$ approximates $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ better for small and moderate sample sizes, the approximation $\frac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \frac{R_i}{n}$, works better for large sample sizes.

Next, we see that a two-term expansion improves the first order approximation of the quantity $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ in particular for large n. But notice that the random variable $\sqrt{n}\, R_2$ converges to zero much more slowly than the quantity R_1 does. Comparing the quantities $\sqrt{n}\, R_2$ in Table 1 and Table 2 we see that the convergence of the term $\sqrt{n}\, R_2$ to zero is much slower for the R-estimator than for the M-estimator. This may be explained by a higher smoothness of the M-estimator (provided the function ψ and the underlying distribution of the errors are sufficiently smooth).

4.2 An alternative confidence interval

In this section we introduce an alternative way of constructing a confidence interval for a single component of a regression parameter $\boldsymbol{\beta}$ in the model (1.1). We derive asymptotic properties of this procedure and compare it with the standard Wald-type procedure.

4.2.1 Unstudentized M-estimators

Leaving out resampling procedures, the standard way for a construction of a confidence interval is a Wald-type method. This method directly exploits FOAL of the M-estimator. From (4.4) we immediately see that the random variable $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ is asymptotically normally distributed with zero mean and the variance $\frac{\sigma_\psi^2}{\gamma_1^2}\mathbf{V}_n^{-1}$, where $\sigma_\psi^2 = \mathsf{E}\,\psi(e_1)^2$. Let us denote $\{\omega_{ij}^n\}_{i=1,\ldots,p}^{j=1,\ldots,p}$ the elements of the matrix $\mathbf{V}_n^{-1}$ and $T_{nl}^2 = \frac{1}{n}\sum_{i=1}^n x_{il}^2$. Then the random variable $\sqrt{n}(\hat{b}_l - \beta_l)$ has asymptotically zero mean normal distribution with variance $\frac{\sigma_\psi^2\,\omega_{ll}^n}{\gamma_1^2}$. Thus we can construct the confidence interval for β_l as

$$D_n^I = \left[\hat{\beta}_l - \tfrac{z_\alpha}{\sqrt{n}}\tfrac{\hat{\sigma}_\psi\sqrt{\omega_{ll}^n}}{\hat{\gamma}_1},\ \hat{\beta}_l + \tfrac{z_\alpha}{\sqrt{n}}\tfrac{\hat{\sigma}_\psi\sqrt{\omega_{ll}^n}}{\hat{\gamma}_1}\right], \tag{4.29}$$

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 it a **type I confidence interval.** Putting $r_i = Y_i - \hat{\boldsymbol{\beta}}_n^{\mathsf{T}}\mathbf{x}_i$ for the residuals, the most simple estimates of σ_ψ and γ_1 are

$$\hat{\sigma}_\psi = \left[\frac{1}{n-1}\sum_{i=1}^n \psi^2(r_i)\right]^{1/2} \qquad \text{and} \qquad \hat{\gamma}_1 = \frac{1}{n}\sum_{i=1}^n \psi'(r_i). \tag{4.30}$$

Sometimes we may find the confidence interval (4.29) inconvenient particularly for two reasons. First, we need to estimate two unknown quantities (σ_ψ and γ_1). Second, we may be doubtful whether the symmetry of the confidence interval (4.29) does not affect coverage properties, especially in the presence of asymmetric distribution of errors or explanatory variables in the model (1.1).

Boos (1980) 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

$$\hat{\theta}_n^- = \sup\left\{t : \tfrac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i - t) \geq \hat{\sigma}_\psi\, z_\alpha\right\} \tag{4.31}$$

$$\hat{\theta}_n^+ = \inf\left\{t : \tfrac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i - t) \leq -\hat{\sigma}_\psi\, z_\alpha\right\}. \tag{4.32}$$

We will call it a **type II confidence interval**. It is apparent from the definition that this method can only work for **monotone** ψ in general. But we can easily modify the definition to include redescending ψ-functions as well. Suppose that $\hat{\theta}_n$ is the M-estimator. Then we define the confidence interval $[\hat{\theta}_n^-, \hat{\theta}_n^+] = [\hat{\theta}_n + \delta_n^-, \hat{\theta}_n + \delta_n^+]$, where

$$\begin{aligned}
\delta_n^- &= \sup\left\{t < 0 : \tfrac{1}{\sqrt{n}} \sum_{i=1}^n \psi(X_i - \hat{\theta}_n - t) \geq \hat{\sigma}_\psi \, z_\alpha \right\}, \\
\delta_n^+ &= \inf\left\{t > 0 : \tfrac{1}{\sqrt{n}} \sum_{i=1}^n \psi(X_i - \hat{\theta}_n - t) \leq -\hat{\sigma}_\psi \, z_\alpha \right\}.
\end{aligned}$$

The advantage of this approach is that we do not need to estimate the functional γ_1. Boos (1980) 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. He also demonstrated by the means of simulation that his proposed method sometimes has a slightly better coverage properties then the type I method. Some partial explanation of this phenomenon can be found in Lloyd (1994).

In the following, we will modify the type II method for a linear model (1.1). We will show that the length of the type II CI for a single component (multiplied by $\sqrt{n}$) has the same probability limit as the CI of type I, but asymptotic distributions of properly standardized lengths of CI's are in general different.

4.2.2 Construction of the confidence interval

For the simplicity of notation we will construct the confidence interval for the last component of $\boldsymbol{\beta}$ – parameter β_p. The general case would follow by relabeling of the indices. To simplify the subsequent formulae we will make use of the following notations. Let $\mathbf{z}_i$ stand for the vector $\mathbf{x}_i$ without the first component, that is $\mathbf{z}_i = (x_{i1}, \ldots, x_{i\,p-1})^\mathsf{T}$. Similarly $\mathbf{V}_n^z = \frac{1}{n}\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\mathsf{T}$ and $\hat{\boldsymbol{\beta}}_z = (\hat{\beta}_1, \ldots, \hat{\beta}_{p-1})^\mathsf{T}$. Further denote $\mathbf{d}_{x\,\mathbf{z}} = \frac{1}{n}\sum_{i=1}^n x_{ip}\,\mathbf{z}_i$ and $T_{np}^2 = \frac{1}{n}\sum_{i=1}^n x_{ip}^2$.

Finally put

$$M_n(t) = \tfrac{1}{\sqrt{n}} \sum_{i=1}^n x_{ip}\, \psi(r_i - t\, x_{ip}) \tag{4.33}$$

and notice that

$$M_n(\beta_p - \hat{\beta}_p) = \tfrac{1}{\sqrt{n}} \sum_{i=1}^n x_{ip}\, \psi(Y_i - \hat{\boldsymbol{\beta}}_z^\mathsf{T} \mathbf{z}_i - \beta_p\, x_{ip}).$$

Then the confidence interval for the parameter β_p is $D_n^{II} = [\hat{b}_p^-, \hat{b}_p^+] = [\hat{\beta}_p + \delta_n^-, \hat{\beta}_p + \delta_n^+]$, where

$$\delta_n^- = \sup\left\{t < 0 : M_n(t) \geq T_{np}^2 \sqrt{\omega_{pp}^n}\, \hat{\sigma}_\psi\, z_\alpha \right\}, \tag{4.34}$$

$$\delta_n^+ = \inf\left\{t > 0 : M_n(t) \leq -T_{np}^2 \sqrt{\omega_{pp}^n}\, \hat{\sigma}_\psi\, z_\alpha \right\}. \tag{4.35}$$

Remark 14. There exist no solution to either of the equations (4.34) or (4.35) if

$$\frac{\sup_t |\psi(t)|}{\sqrt{n}} \sum_{i=1}^n |x_{ip}| < T_{np}^2 \sqrt{\omega_{pp}^n}\, \hat{\sigma}_\psi\, z_\alpha. \tag{4.36}$$

This may happen if $\sum_{i=1}^n |x_{ip}|$ is 'too small' in comparison with $\sum_{i=1}^n x_{ip}^2$, that is if the second moment of the p-th column of the design matrix is 'too large'. To prevent this possibility at least partially, it is advisable to center the explanatory variables. If the linear model includes an intercept, this transformation does not affect the estimate of the slope coefficients. Moreover, as in practice we are usually interested mainly in the slope coefficients, this simple transformation does not cause any interpretation problems as well.

The second simple idea is whether the scale transformation $x'_{ip} = \frac{x_{ip}}{K}$ $(i = 1, \ldots, n)$ for an appropriate K would help. The answer is negative as we can easily see that $(T'_{np})^2 = \frac{T_{np}^2}{K^2}$ and $(\omega_{pp}^n)' = K\omega_{pp}^n$. That is why both sides of (4.36) for transformed variables are multiplied by the same factor $\frac{1}{K}$ and the problem is left unchanged.

Good news is that numerical experiments show that unless the behaviour of the p-th explanatory variable is 'very wild' and the sample size is very small ($n < 15$), the problematic situation (4.36) occurs very rarely. Nevertheless, we strongly recommend to compute both types of confidence intervals and compare the results. A huge difference is a good indication to look at our data more carefully.

In the next, we will use the following simple estimator of σ_ψ^2:

$$\hat{\sigma}_\psi^2 = \frac{1}{n}\sum_{i=1}^n \psi^2(r_i). \tag{4.37}$$

The next lemma justifies the usage of this estimator.

Lemma 4.5. *Let the estimator $\hat{\boldsymbol{\beta}}_n$ be $\sqrt{n}$-root consistent, the function ψ of bounded variation, and the condition **X.2** satisfied. Then $\hat{\sigma}_\psi^2 = \sigma_\psi^2 + o_p(1)$, that is, the estimator $\hat{\sigma}_\psi^2$ is weakly consistent.*

Proof. Let M be an arbitrarily large but fixed constant and denote

$$S_n(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^n \psi^2\left(e_i - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}\right), \quad \mathbf{t} \in T = \{\mathbf{s} \in \mathbb{R}_p : |\mathbf{s}|_2 \leq M\}.$$

In Lemma 5.6 we will prove that the quantity $\|S_n(\mathbf{t}) - \mathsf{E}\ \psi^2(e_1)\|_T$ converges to zero almost surely. Further, as the estimator $\hat{\boldsymbol{\beta}}_n$ is $\sqrt{n}$-root consistent, we can replace $\mathbf{t}$ with $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ and get the statement of the lemma. $\square$

Remark 15. Somebody may argue that some of the estimators

$$\tilde{\sigma}_\psi^2 = \frac{1}{n-p}\sum_{i=1}^n \psi^2(r_i), \quad \text{or} \quad \tilde{\sigma}_\psi^2 = \frac{1}{n-p}\sum_{i=1}^n [\psi(r_i) - \bar{\psi}]^2,$$

where $\bar{\psi} = \frac{1}{n}\sum_{i=1}^n \psi(r_i)$, would be more appropriate. But it is easy to see that all of these estimators are asymptotically equivalent in the sense: $\sqrt{n}(\hat{\sigma}_\psi^2 - \tilde{\sigma}_\psi^2) = o_p(1)$. To show this, it suffices to verify that $\bar{\psi}$ is 'small enough.' By the definition of the M-estimator $\bar{\psi} = 0$ if the linear model includes an intercept and the function ψ is continuous. If this is not the case, we can use $\sqrt{n}$-consistency of $\hat{\boldsymbol{\beta}}_n$, FOAL result 4.2 and the condition **GenFx.1**, which together give us

$$O_p(\tfrac{1}{\sqrt{n}}) = \frac{1}{n}\sum_{i=1}^{n}\psi(e_i) = \frac{1}{n}\sum_{i=1}^{n}\psi(r_i) + \tfrac{\gamma_1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{\mathsf{T}}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + o_p(\tfrac{1}{\sqrt{n}}) = \bar{\psi} + O_p(\tfrac{1}{\sqrt{n}}).$$

But this implies $\bar{\psi} = O_p(\frac{1}{\sqrt{n}})$, which further yields

$$\frac{1}{n-p}\sum_{i=1}^{n}[\psi(r_i) - \bar{\psi}]^2 = \frac{1}{n-p}\sum_{i=1}^{n}\psi^2(r_i) - \frac{n\,\bar{\psi}^{\,2}}{n-p} = \frac{1}{n-p}\sum_{i=1}^{n}\psi^2(r_i) + O_p(\tfrac{1}{n}).$$

We see, that we have even proved $\sqrt{n}(\hat{\sigma}_\psi^2 - \tilde{\sigma}_\psi^2) = O_p(\frac{1}{\sqrt{n}})$.

That is why the simplicity of notation is the main reason for working with the estimator (4.37).

Before we state the theorem about basic asymptotic properties of the confidence interval procedure, we need one more technical lemma for the case of a step function ψ.

Lemma 4.6. *Let ψ be an increasing step function* (4.13). *Then there exists $K < \infty$ such that almost surely it holds for all $n \in \mathbb{N}$*

$$|M_n(\delta_n^-) - T_{np}^2\,\sqrt{\omega_{pp}^n}\,\hat{\sigma}_\psi\,z_\alpha| \le \frac{K\max_{1\le i\le n}|x_{ip}|}{\sqrt{n}} \tag{4.38}$$

as well as

$$|M_n(\delta_n^+) + T_{np}^2\,\sqrt{\omega_{pp}^n}\,\hat{\sigma}_\psi\,z_\alpha| \le \frac{K\max_{1\le i\le n}|x_{ip}|}{\sqrt{n}}. \tag{4.39}$$

Proof. The proof is very similar to the considerations made in Section 4.1.2 to arrive at (4.16). We will only prove the first part of the lemma for δ_n^-, because the proof for δ_n^+ would be completely analogous.

Let us denote $c = T_{np}^2\,\sqrt{\omega_{pp}^n}\,\hat{\sigma}_\psi\,z_\alpha$ and

$$G_n(t) = M_n(\delta_n^- - t) - c = \tfrac{1}{\sqrt{n}}\sum_{i=1}^{n} x_{ip}\,\psi(r_i - \delta_n^- x_{ip} + t\,x_{ip}) - c.$$

By the definition of δ_n^- (4.34) and the monotonicity of the function ψ we see that for every $\varepsilon > 0$ it holds $G_n(\varepsilon) \ge 0$ and $G_n(-\varepsilon) \le 0$. This gives us the inequality $|G_n(0)| \le G_n(\varepsilon) - G_n(-\varepsilon)$.

Letting $\varepsilon \searrow 0$, we easily find $K > 0$ such that for every $n \in \mathbb{N}$

$$|G_n^+(0)| \le \frac{K\max_{1\le i\le n}|x_{ip}|}{\sqrt{n}} \qquad \text{a.s.,}$$

which proves the lemma. $\square$

Remark 16. If ψ is a step function, the monotonicity is an important technical assumption. But as we have discussed in Remark 12, we believe that the difficulties for step functions are only of technical character. Once there are solved, we justify the usage of type II confidence intervals for these M-estimators as well. Some preliminary numerical experiments show that the type II CI may be useful, for instance if we use the (skipped mean) ψ-function given by

$$\psi(x) = \begin{cases} x, & |x| \leq k \\ 0, & |x| > k. \end{cases}$$

As to construct type I confidence interval, we need to estimate the functional γ_1, given by $\gamma_1 = F(k) - F(-k) - k(f(k) + f(-k))$. But this estimation may be rather difficult particularly in not very large samples (< 100), when we do not have enough information to estimate the density at the points k and $-k$.

The condition **X'.2** implies that the quantities on the right-hand sides of (4.38) and (4.39) in Lemma 4.6 are of order $o(1)$ almost surely. This is sufficient for the first order results. For the second order results we require the stronger condition **XX'.1**, which yields order $o(n^{-1/4})$ almost surely.

Suppose the condition **XX.2** to hold and denote $\mathbf{V}^{-1} = [\omega_{ij}]_{i,j=1}^{p}$. For the sake of simplicity of notation put

$$a_F^n = \frac{\sigma_\psi\, z_\alpha \sqrt{\omega_{pp}^n}}{\gamma_1}, \qquad \text{and} \qquad a_F = \lim_{n\to\infty} a_F^n = \frac{\sigma_\psi\, z_\alpha \sqrt{\omega_{pp}}}{\gamma_1}. \tag{4.40}$$

Theorem 4.7. *Suppose that the FOAL of the M-process* (4.2) *as well as the condition **X'.2** hold and $\hat{\boldsymbol{\beta}}_n$ admits FOAR* (4.4), *then*

(i).

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

(ii).

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

Proof. Proof of (i)
Note that

$$\mathrm{P}\,(D_n^{II} \not\ni \beta_p) = \mathrm{P}\,(\hat{b}_p^- > \beta_p) + \mathrm{P}\,(\hat{b}_p^+ < \beta_p).$$

With the help of FOAL (4.2) and FOAR (4.4) we get

$$\begin{aligned} M_n(\beta_p - \hat{\beta}_p) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{ip}\, \psi(Y_i - \hat{\boldsymbol{\beta}}_z^{\mathsf{T}} \mathbf{z}_i - \beta_p\, x_{ip}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{ip}\, \psi(e_i) - \gamma_1 \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z)^{\mathsf{T}} \mathbf{d}_{x\mathbf{z}} + o_p(1) \end{aligned} \tag{4.41}$$

and

$$o(1) = M_n(0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{ip}\, \psi(e_i) - \gamma_1 \sqrt{n}(\hat{b}_p - \beta_p)\, T_{np}^2 - \gamma_1 \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z)^{\mathsf{T}} \mathbf{d}_{x\,\mathbf{z}} + o_p(1). \quad (4.42)$$

Comparing (4.41) and (4.42) yields

$$M_n(\beta_p - \hat{\beta}_p) = \gamma_1 \sqrt{n}(\hat{\beta}_p - \beta_p)\, T_{np}^2 + o_p(1).$$

To finish the proof of (i), it suffices to realize that by Lemma 4.5 $\hat{\sigma}_\psi \to \sigma_\psi$ in probability, from which it follows

$$\mathrm{P}\,(\hat{b}_p^- > \beta_p) = \mathrm{P}\,(\delta_n^- > \beta_p - \hat{\beta}_p) = \mathrm{P}\,\Big(M_n(\beta_p - \hat{\beta}_p) > T_{np}^2 \sqrt{\omega_{pp}^n}\, \hat{\sigma}_\psi\, z_\alpha\Big) \xrightarrow[n\to\infty]{} \tfrac{\alpha}{2}.$$

Analogously we can prove

$$\mathrm{P}\,(\hat{b}_p^+ < \beta_p) \xrightarrow[n\to\infty]{} \tfrac{\alpha}{2}.$$

<u>Proof of (ii)</u>

First, we need to check that

$$\sqrt{n}(\hat{b}_p^- - \beta_p) = O_p(1) \qquad \text{and} \qquad \sqrt{n}(\hat{b}_p^+ - \beta_p) = O_p(1). \quad (4.43)$$

Analogously to (4.41) we can calculate

$$\begin{aligned} \mathrm{P}\,(\sqrt{n}(\hat{b}_p^- - \beta_p) > t) &= \mathrm{P}\,(\hat{b}_p^- > \beta_p + \tfrac{t}{\sqrt{n}}) = \mathrm{P}\,(\delta_n^- > \beta_p - \hat{\beta}_p + \tfrac{t}{\sqrt{n}}) \\ &= \mathrm{P}\,\Big(\gamma_1 T_{np}^2 \sqrt{n}(\beta_p - \hat{\beta}_p) < T_{np}^2 \sqrt{\omega_{pp}^n}\, \hat{\sigma}_\psi\, z_\alpha - t\,\gamma_1\, T_{np}^2 + o_p(1)\Big). \end{aligned}$$

As the random variable $\sqrt{n}(\hat{b}_p - \beta_p)$ is asymptotically normal, we can make the last probability arbitrarily small for all sufficiently large $n \in \mathbb{N}$ by taking t large enough, which implies (4.43) for $\hat{b}_p^-$. Similarly we can prove (4.43) for $\hat{b}_p^+$.

(4.43) enables us to insert $\mathbf{b} \to \sqrt{n}\left((\hat{\boldsymbol{\beta}}_z, \hat{b}_p^+)^{\mathsf{T}} - \boldsymbol{\beta}\right)$ as well as $\mathbf{b} \to \sqrt{n}\left((\hat{\boldsymbol{\beta}}_z, \hat{b}_p^-)^{\mathsf{T}} - \boldsymbol{\beta}\right)$ in the asymptotic linearity result (4.2) with $\mathbf{x}_i = x_{ip}$ and get

$$\begin{aligned} M_n(\hat{b}_p^+ - \hat{\beta}_p) - \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n x_{ip}\, \psi(e_i) + \gamma_1 T_{np}^2 \sqrt{n}(\hat{b}_p^+ - \beta_p) + \gamma_1 \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z)^{\mathsf{T}} \mathbf{d}_{x\,\mathbf{z}} &= o_p(1), \\ M_n(\hat{b}_p^- - \hat{\beta}_p) - \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n x_{ip}\, \psi(e_i) + \gamma_1 T_{np}^2 \sqrt{n}(\hat{b}_p^- - \beta_p) + \gamma_1 \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z)^{\mathsf{T}} \mathbf{d}_{x\,\mathbf{z}} &= o_p(1). \end{aligned}$$

Subtracting these two equations and Lemma 4.5 (and Lemma 4.6 in the case of a step-ψ) yield

$$\begin{aligned} \gamma_1 T_{np}^2 \sqrt{n}(b_p^+ - b_p^-) &= M_n(\hat{b}_p^- - \hat{\beta}_p) - M_n(\hat{b}_p^+ - \hat{\beta}_p) + o(1) \\ &= M_n(\delta_n^-) - M_n(\delta_n^+) + o(1) = 2\, T_{np}^2 \sqrt{\omega_{pp}^n}\, \sigma_\psi\, z_\alpha + o_p(1), \end{aligned}$$

which yields the statement of the theorem. □

In the following, we will assume that the function ψ is **absolutely continuous**. The case of a step function is treated separately in Section 4.2.4.

Before we proceed with a finer analysis of the length of the confidence interval, we need to find the asymptotic distribution of the random variable $\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)$. For this purpose, we need to impose some further conditions on the function ψ and the distribution of the errors.

SmFx.4 There exists a $\delta > 0$ such that $\sup_{|t|<\delta} \mathsf{E}\ \psi^4(e_1 + t) < \infty$.

SmFx.5 The function $\lambda^{(2)}(t) = \mathsf{E}\ \psi^2(e_1 + t)$ is continuously differentiable in a neighbourhood of the point zero.

As the function ψ is continuous, the condition **SmFx.4** together with Lemma 7.17 imply the continuity of the function $\psi^2(e_1 + t)$ in the quadratic mean at the point zero, that is

$$\lim_{t\to 0} \mathsf{E}\ \left[\psi^2(e_1 + t) - \psi^2(e_1)\right]^2 = 0. \tag{4.44}$$

Let us denote $\gamma_{01} = \frac{\partial}{\partial t}\left(\mathsf{E}\ \psi^2(e_1 + t)\right)_{t=0}$. Usually we can interchange the derivative and integral and get

$$\gamma_{01} = 2\ \mathsf{E}\ \psi(e_1)\,\psi'(e_1).$$

Lemma 4.8. *Suppose that the conditions* ***XX.1-2**, **SmFx.1-5** (or **Step.1-2**), and **GenFx.1*** *are satisfied. Then*

$$\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi) = \frac{1}{2\,\sigma_\psi\sqrt{n}} \sum_{i=1}^{n} \left[(\psi^2(e_i) - \sigma_\psi^2) - b_i\,\psi(e_i)\right] + o_p(1), \tag{4.45}$$

where $b_i = \frac{\gamma_{01}\mathbf{x}_i^\mathsf{T}\mathbf{V}_n^{-1}}{\gamma_1} \sum_{j=1}^n \frac{\mathbf{x}_j}{n}$.

Proof. Let us define the processes

$$Z_n(\mathbf{t}) = \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} [\psi^2(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - \psi^2(e_i)], \qquad \bar{Z}_n(\mathbf{t}) = Z_n(\mathbf{t}) - \mathsf{E}\ Z_n(\mathbf{t}),$$

where $\mathbf{t} \in T = \{\mathbf{s} \in \mathbb{R}_p : |\mathbf{s}|_2 \le M\}$ and M is an arbitrarily large but fixed constant.

Then with the help of conditions **XX.1-2** and **SmFx.4** we can easily verify the assumptions of Corollary 7.13, which gives us $\sup_{|\mathbf{t}|_2\le M} |\bar{Z}_n(\mathbf{t})| = o_p(1)$. Let us only note that (4.44) would be utilized here.

As the next step we can use the assumption **SmFx.5** to show that we can replace $\mathsf{E}\ Z_n(\mathbf{t})$ by $-\gamma_{01}\mathbf{t}^\mathsf{T}\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$. Combining this two results gives us

$$\tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} [\psi^2(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - \psi^2(e_i)] + \gamma_{01}\mathbf{t}^\mathsf{T} \sum_{i=1}^{n} \tfrac{\mathbf{x}_i}{n} = o_p(1).$$

Because the estimate $\hat{\boldsymbol{\beta}}_n$ satisfies the first order representation (4.4), we can substitute $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$ in the last equation and get

$$\tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi^2(r_i) = \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi^2(e_i) - \tfrac{\gamma_{01}}{\gamma_1 \sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i^{\mathsf{T}} \mathbf{V}_n^{-1} \psi(e_i) \sum_{j=1}^{n} \tfrac{\mathbf{x}_j}{n} + o_p(1),$$

which after some reorganization implies

$$\sqrt{n}(\hat{\sigma}_\psi^2 - \sigma_\psi^2) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big[(\psi^2(e_i) - \sigma_\psi^2) - b_i\, \psi(e_i)\big] + o_p(1).$$

The representation (4.45) now follows easily by a (Delta-)Theorem 7.3. $\square$

Fortunately, the awkward representation (4.45) simplifies considerable in two important cased. First, if the linear model (1.1) includes an intercept, that is $x_{i1} = 1$ for $i = 1, \ldots, n$, then $\mathbf{V}_n^{-1} \sum_{j=1}^{n} \frac{\mathbf{x}_j}{n} = \mathbf{u}_1$, where $\mathbf{u}_1 = (1, 0, \ldots, 0)^{\mathsf{T}} \in \mathbb{R}_p$. This further implies

$$b_i = \tfrac{\gamma_{01} \mathbf{x}_i^{\mathsf{T}} \mathbf{V}_n^{-1}}{\gamma_1} \sum_{j=1}^{n} \tfrac{\mathbf{x}_j}{n} = \tfrac{\gamma_{01}}{\gamma_1}, \qquad i = 1, \ldots, n.$$

Second, if the distribution of the errors satisfies the symmetry condition **Sym**, then $\gamma_{01} = 0$, which implies $b_i = 0$ for all $i = 1, \ldots, n$.

Theorem 4.9. *Suppose that the conditions **XX.1-2**, **SmFx.1-5**, and **GenFx.1** are satisfied. Put*

$$L_n^{II} = \frac{\sqrt{n}[\sqrt{n}(\hat{b}_n^+ - \hat{b}_n^-) - 2\, a_F^n]}{2\, a_F}. \tag{4.46}$$

Then the random variable L_n is asymptotically normal with mean zero and the variance which can be deduced from the following asymptotic representation

$$L_n^{II} = -\tfrac{1}{\gamma_1 \sqrt{n}} \sum_{i=1}^{n} \tfrac{x_{ip}^2}{T_{np}^2} \left[\psi'(e_i) - \gamma_1\right] + \frac{\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)}{\sigma_\psi} + \tfrac{\gamma_2}{\gamma_1} \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^{\mathsf{T}} \sum_{i=1}^{n} \tfrac{x_{ip}^2\, \mathbf{x}_i}{n\, T_{np}^2} + o_p(1). \tag{4.47}$$

As we see from (4.47), the formula for the asymptotic variance of L_n (the standardized length of CI) is rather complicated in general. But if the symmetry condition **Sym** is met, then both of the functionals γ_2 and γ_{01} are zero and using the representation (4.45) we get

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

Proof. Substitute $\sqrt{n}(\hat{\boldsymbol{\beta}}_z^{\mathsf{T}} - \boldsymbol{\beta}_z^{\mathsf{T}}, \hat{b}_p^+ - \beta_p)$ for $\mathbf{t}^{\mathsf{T}}$ in the second order asymptotic expansion of the M-process (2.8) (with $c_i = x_{ip}$ for $i = 1, \dots, n$) and get

$$
\begin{aligned}
&-\sqrt{n}\, T_{np}^2 \sqrt{\omega_{pp}^n}\, z_\alpha\, \hat{\sigma}_\psi + \gamma_1\, n(\hat{b}_p^+ - \beta_p) T_{np}^2 + \gamma_1\, n(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z)^\mathsf{T} \mathbf{d}_{x\,\mathbf{z}} \\
&\quad = -\sqrt{n}(\hat{b}_p^+ - \beta_p) \tfrac{1}{\sqrt{n}} \sum_{i=1}^n x_{ip}^2\, [\psi'(e_i) - \gamma_1] - \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z)^\mathsf{T} \tfrac{1}{\sqrt{n}} \sum_{i=1}^n x_{ip}\, \mathbf{z}_i\, [\psi'(e_i) - \gamma_1] \\
&\qquad + \frac{\gamma_2}{2} \left\{ \frac{1}{n} \sum_{i=1}^n x_{ip}^3 \left[\sqrt{n}(\hat{b}_p^+ - \beta_p) \right]^2 + \frac{2}{n} \sum_{i=1}^n x_{ip}^2\, \mathbf{z}_i^\mathsf{T} \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z) \sqrt{n}(\hat{b}_p^+ - \beta_p) \right. \\
&\qquad\qquad \left. + \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z)^\mathsf{T} \frac{1}{n} \sum_{i=1}^n x_{ip}\, \mathbf{z}_i \mathbf{z}_i^\mathsf{T} \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z) \right\} + o_p(1). \qquad (4.48)
\end{aligned}
$$

Similarly substituting $\sqrt{n}(\hat{\boldsymbol{\beta}}_z^\mathsf{T} - \boldsymbol{\beta}_z^\mathsf{T}, \hat{b}_p^- - \beta_p)$ for $\mathbf{t}^\mathsf{T}$ in (2.8) yields

$$
\begin{aligned}
&\sqrt{n}\, T_{np}^2 \sqrt{\omega_{pp}^n}\, z_\alpha\, \hat{\sigma}_\psi + \gamma_1\, n(\hat{b}_p^- - \beta_p) T_{np}^2 + \gamma_1\, n(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z)^\mathsf{T} \mathbf{d}_{x\,\mathbf{z}} \\
&\quad = -\sqrt{n}(\hat{b}_p^- - \beta_p) \tfrac{1}{\sqrt{n}} \sum_{i=1}^n x_{ip}^2 [\psi'(e_i) - \gamma_1] - \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z)^\mathsf{T} \tfrac{1}{\sqrt{n}} \sum_{i=1}^n x_{ip}\, \mathbf{z}_i [\psi'(e_i) - \gamma_1] \\
&\qquad + \frac{\gamma_2}{2} \left\{ \frac{1}{n} \sum_{i=1}^n x_{ip}^3 \left[\sqrt{n}(\hat{b}_p^- - \beta_p) \right]^2 + \frac{2}{n} \sum_{i=1}^n x_{ip}^2\, \mathbf{z}_i^\mathsf{T} \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z) \sqrt{n}(\hat{b}_p^- - \beta_p) \right. \\
&\qquad\qquad \left. + \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z)^\mathsf{T} \frac{1}{n} \sum_{i=1}^n x_{ip}\, \mathbf{z}_i \mathbf{z}_i^\mathsf{T} \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z) \right\} + o_p(1). \qquad (4.49)
\end{aligned}
$$

Subtracting (4.48) from (4.49) gives us

$$
\begin{aligned}
&\gamma_1\, n(\hat{b}_p^+ - \hat{b}_p^-)\, T_{np}^2 - 2\, T_{np}^2 \sqrt{n} \sqrt{\omega_{pp}^n}\, z_\alpha\, \hat{\sigma}_\psi = -\sqrt{n}(\hat{b}_p^+ - \hat{b}_p^-) \tfrac{1}{\sqrt{n}} \sum_{i=1}^n x_{ip}^2\, [\psi'(e_i) - \gamma_1] \\
&\qquad + \frac{\gamma_2}{2} \left\{ \frac{1}{n} \sum_{i=1}^n x_{ip}^3 \left\{ \left[\sqrt{n}(\hat{b}_p^+ - \beta_p) \right]^2 - \left[\sqrt{n}(\hat{b}_p^- - \beta_p) \right]^2 \right\} \right. \\
&\qquad\qquad \left. + \frac{2}{n} \sum_{i=1}^n x_{ip}^2\, \mathbf{z}_i^\mathsf{T} \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z) \sqrt{n}(\hat{b}_p^+ - \hat{b}_p^-) \right\} + o_p(1), \qquad (4.50)
\end{aligned}
$$

which further implies

$$
\begin{aligned}
L_n^{II} &= -\tfrac{1}{T_{np}^2 \sqrt{n}} \sum_{i=1}^n x_{ip}^2\, [\psi'(e_i) - \gamma_1] + \frac{\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)}{2\, \sigma_\psi} \\
&\quad + \frac{\gamma_2}{2\, T_{np}^2} \left\{ \frac{1}{n} \sum_{i=1}^n x_{ip}^3 \sqrt{n}(\hat{b}_p^+ + \hat{b}_p^- - 2\, \beta_p) + \frac{2}{n} \sum_{i=1}^n x_{ip}^2\, \mathbf{z}_i^\mathsf{T} \sqrt{n}(\hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z) \right\} + o_p(1). \qquad (4.51)
\end{aligned}
$$

To prove the statement of the theorem, it only remains to show that the third term on the right-hand side of (4.51) is asymptotically equivalent (up to a term of order $o_p(1)$) to the third term in (4.47). But this is an immediate consequence of

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

where the last equality follows by $\sqrt{n}(\hat{b}_p^+ + \hat{b}_p^- - 2\,\hat{\beta}_p) = o_p(1)$, whose proof is analogous to the proof of the second part of Theorem 4.7. $\square$

Remark 17. As by the condition **XX.2** $a_F = a_F^n + o(1)$, we can replace the denominator of L_n^{II} in (4.47) by a_F^n.

Comparison with the type I confidence interval

Recall that the confidence interval of type I for β_p is

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

If the estimators $\hat{\gamma}_1$ and $\hat{\sigma}_\psi$ are (weakly) consistent estimators of γ_1 and σ_ψ, it is pretty straightforward to show that $P(D_n^I \ni \beta_p) \to 1-\alpha$ and $\sqrt{n}[\hat{b'}_p^+ - \hat{b'}_p^-] = 2\,a_F + o_p(1)$.

Suppose that we use $\hat{\gamma}_1 = \frac{1}{n}\sum_{i=1}^n \psi'(r_i)$ as the estimator of the functional γ_1. Then it is not difficult to find the expansion for the standardized length of the confidence interval

$$L_n^I = -\tfrac{1}{\gamma_1\sqrt{n}} \sum_{i=1}^n [\psi'(e_i) - \gamma_1] + \frac{\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)}{\sigma_\psi} + \tfrac{\gamma_2}{\gamma_1}\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^\mathsf{T} \sum_{i=1}^n \tfrac{\mathbf{x}_i}{n} + o_p(1). \tag{4.52}$$

If we compare (4.52) with the expansion of L_n^{II} (4.47) for type II confidence interval and realize that $(T_{np}^2)^2 \leq \frac{1}{n}\sum_{i=1}^n x_{ip}^4$, we immediately see that the type I is more stable in the sense that its standardized length has a smaller asymptotic variance. This is not so surprising if we realize that the type II confidence interval 'implicitly' uses $\hat{\gamma}_1' = \frac{1}{n\,T_{np}^2}\sum_{i=1}^n x_{ip}^2\,\psi'(r_i)$ as the estimator of γ_1 (this may be seen from the expansion of $M_n(\hat{b}_p^+ - \hat{\beta}_p)$ or $M_n(\hat{b}_p^- - \hat{\beta}_p)$ around the point zero derived in the proof Theorem 4.7). But this estimator is more variable then the simple estimator $\hat{\gamma}_1 = \frac{1}{n}\sum_{i=1}^n \psi'(r_i)$. On the other hand the results presented in Omelka (2006) indicate that the estimator $\hat{\gamma}_1'$ often prevents the worst in the case of heteroscedasticity.

Let us suppose for this moment that we use $\hat{\gamma}_1 = \frac{1}{n\,T_{np}^2}\sum_{i=1}^n x_{ip}^2\,\psi'(r_i)$ as the estimator of γ_1 in the type I confidence procedure. Then, after some algebra, we can find that the standardized length of the confidence interval satisfies (4.47). This implies that the lengths of the confidence intervals of type I and type II are asymptotically equivalent, that is

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

But it is not generally true that $n\,(\hat{b'}_p^- - \hat{b}_p^-) = o_p(1)$ or $n\,(\hat{b'}_p^+ - \hat{b}_p^+) = o_p(1)$. To see this, consider the expansion of the function $M_n(t)$ (defined in (4.33)) around the point 0 for t of order $O_p(\frac{1}{\sqrt{n}})$

$$
\begin{aligned}
M_n(t) = \tfrac{1}{\sqrt{n}} \sum_{i=1}^n x_{ip}\, \psi(r_i - t\, x_{ip}) \\
&= \tfrac{1}{\sqrt{n}} \sum_{i=1}^n x_{ip}\, \psi(r_i) - \tfrac{t}{\sqrt{n}} \sum_{i=1}^n x_{ip}^2 \psi'(r_i) + \tfrac{\gamma_2\, t^2}{\sqrt{n}} \sum_{i=1}^n x_{ip}^3 + o_p(\tfrac{1}{\sqrt{n}}) \\
&= -\sqrt{n}\, t\, \hat{\gamma}_1 T_{np}^2 + \tfrac{\gamma_2\, t^2}{\sqrt{n}} \sum_{i=1}^n x_{ip}^3 + o_p(\tfrac{1}{\sqrt{n}}).
\end{aligned} \tag{4.53}
$$

As during the proof of Theorem 4.7 we have shown $\sqrt{n}(\hat{b}_p^- - \beta_p) = O_p(1)$, which further implies $(\hat{b}_p^- - \hat{\beta}_p) = O_p(\frac{1}{\sqrt{n}})$, we can substitute $\delta_n^- = (\hat{b}_p^- - \hat{\beta}_p)$ for t in the expansion (4.53) and get

$$
T_{np}^2 \sqrt{\omega_{pp}^n}\, \hat{\sigma}_\psi\, z_\alpha \overset{(4.34)}{=} M_n(\delta_n^-) \overset{(4.53)}{=} -\sqrt{n}\,(\hat{b}_p^- - \hat{\beta}_p)\, \hat{\gamma}_1' T_{np}^2 + O_p(\tfrac{1}{\sqrt{n}}).
$$

This further yields $\hat{b}_p^- = \hat{\beta}_p - \frac{\sqrt{\omega_{pp}^n}\, \hat{\sigma}_\psi\, z_\alpha}{\hat{\gamma}_1' \sqrt{n}} + o_p(\frac{1}{\sqrt{n}})$. Now we can insert $\hat{b}_p^-$ for t in the equation (4.53) once more and after some algebra get

$$
\hat{b}_p^- = \hat{\beta}_p - \frac{\sqrt{\omega_{pp}^n}\, \hat{\sigma}_\psi\, z_\alpha}{\hat{\gamma}_1 \sqrt{n}} + \frac{\gamma_2\, \omega_{pp}\, \sigma_\psi^2\, z_\alpha^2}{\gamma_1^3\, T_{np}^2\, n} \sum_{i=1}^n \frac{x_{ip}^3}{n} + o_p(\tfrac{1}{n}). \tag{4.54}
$$

Comparing (4.54) with the lower bound of the type I confidence interval

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

we see that $n(\hat{b'}_p^- - \hat{b}_p^-) = O_p(1)$, 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 (4.54) holds for $\hat{b}_p^+$ as well, we conclude that the confidence interval D_n^{II} is asymptotically shifted 'a little' to the right or left depending on the sign of the quantity $\frac{\gamma_2}{n} \sum_{i=1}^n x_{ip}^3$.

4.2.3 Studentized M-estimators

As in practice we usually do not know the scale, we prefer to use studentized M-estimators. The difficulty with the inference for these estimators is that the asymptotic distribution of the estimator of the location parameter depends on the asymptotic distribution of the scale estimator (unless the underlying distribution is symmetric). That is why the confidence interval constructed by Boos (1980) for the parameter of location $[\hat{\theta}_n^-, \hat{\theta}_n^+]$ given by

$$
\hat{\theta}_n^- = \sup\left\{ t : \tfrac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\tfrac{X_i - t}{S_n}) \geq \hat{\sigma}_\psi\, z_\alpha \right\} \tag{4.55}
$$

$$
\hat{\theta}_n^+ = \inf\left\{ t : \tfrac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\tfrac{X_i - t}{S_n}) \leq -\hat{\sigma}_\psi\, z_\alpha \right\}, \tag{4.56}
$$

where $\hat{\sigma}_\psi^2$ is an estimate of $\mathsf{E}\,\psi^2(\frac{X_1-\theta}{S})$, is not generally asymptotically correct, that is $\lim_{n\to\infty} P(\theta \in [\hat{\theta}_n^-, \hat{\theta}_n^+]) \neq 1-\alpha$. Let us note that the type I confidence interval is facing the same problem.

Fortunately, if an intercept is included in the linear regression model (1.1), it turns out that the asymptotic distribution of the slope parameters does not depend on the asymptotic distribution of the scale estimate (see the FOAR of $\hat{\boldsymbol{\beta}}_n$ in (4.7)). This enables us to construct an asymptotically correct confidence interval for a single component of the 'slope' part of the vector parameter $\boldsymbol{\beta}$.

In this section we will suppose the linear model model (1.1) includes an intercept, that is $x_{i1} = 1$, for $i = 1, \ldots, n$. Further, let S_n be an estimator of scale.

Put

$$M_n(t) = \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{ip}\, \psi(\tfrac{r_i - t\, x_{ip}}{S_n}).$$

Then the (type II) confidence interval for the parameter β_p is given by $D_n^{II} = [\hat{b}_p^-, \hat{b}_p^+] = [\hat{\beta}_p + \delta_n^-, \hat{\beta}_p + \delta_n^+]$, where

$$\begin{aligned} \delta_n^- &= \sup\left\{t < 0 : M_n(t) \geq T_{np}^2 \sqrt{\omega_{pp}^n}\, \hat{\sigma}_\psi\, z_\alpha\right\} &\quad (4.57)\\ \delta_n^+ &= \inf\left\{t > 0 : M_n(t) \leq -T_{np}^2 \sqrt{\omega_{pp}^n}\, \hat{\sigma}_\psi\, z_\alpha\right\} &\quad (4.58)\end{aligned}$$

with $\hat{\sigma}_\psi^2 = \frac{1}{n}\sum_{i=1}^n \psi^2(\frac{r_i}{S_n})$.

The following theorem only restates the results of Theorem 4.7 and Theorem 4.9. We note that we will use the symbols γ_1 and γ_{2e} in a way defined in (2.10) (Section 2.1.3) and the quantity a_F is defined by (4.40).

Theorem 4.10. *If the conditions* ***XX.1-2, SmSt.1-3, GenSt.1****, and* $\sqrt{n}(\frac{S_n}{S} - 1) = O_p(1)$ *hold, then the confidence interval* D_n^{II} *defined by* (4.57) *and* (4.58) *satisfies:*

(i).

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

(ii).

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

(iii). Further suppose that $\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi) = O_p(1)$ *and put*

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

Then the random variable L_n *admits the first order asymptotic representation*

$$\begin{aligned} L_n^{II} = &-\tfrac{1}{\gamma_1\sqrt{n}} \sum_{i=1}^{n} \tfrac{x_{ip}^2}{T_{np}^2}\,[\psi'(e_i/S) - \gamma_1] + \tfrac{\gamma_1+\gamma_{2e}}{\gamma_1}\sqrt{n}\left(\tfrac{S_n}{S} - 1\right) \\ &+ \frac{\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)}{\sigma_\psi} + \tfrac{\gamma_2}{\gamma_1}\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^{\mathsf{T}} \sum_{i=1}^{n} \tfrac{x_{ip}^2\,\mathbf{x}_i}{n\,T_{np}^2} + o_p(1). \qquad (4.59)\end{aligned}$$

For this moment we postpone the discussion of the assumption $\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi) = O_p(1)$ to Section 5.2.2 of the next chapter. In that section we derive that the first order representation of the quantity $\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)$ is given by (5.24). To prove that expansion, we will need conditions **SmSt.4-5**, which are analogous to the conditions **SmFix.4-5**.

Proof. The proof of the theorem is completely analogous (only a little more complicated) to the proofs of Theorem 4.7 and Theorem 4.9. The only difference is that instead of FOAL of M-process with fixed scale, we use FOAL of studentized M-process to prove statements (i) and (ii). To prove the last statement we use Corollary 2.6 instead of Corollary 2.3. □

Remark 18. As the statements (i) and (ii) of Theorem 4.10 are 'only' first order results, they could be proved under weaker assumptions.

Numerical evidence

Some partial comparison of the finite sample performance of the confidence intervals of type I and type II are to be found in Omelka (2006). We can summarize the results as follows.

Confidence intervals (CI's) of type II are generally larger and more variable than the type I CI's. Moreover, in comparison with CI of type I, CI's of type II are conservative, that is their coverage is usually larger than the nominal value. On the other hand this higher coverage property is worth considering in models in which errors or explanatory variables are asymmetric. In such models the two-sided CI's of type I often have a slightly lower coverage than theirs nominal values and the one-sided CI's may be completely misleading. Moreover, the type II confidence intervals does not usually fail completely in the case of heteroscedasticity.

4.2.4 M-estimators based on discontinuous ψ function

Suppose that the function ψ is a step function given by (2.14) or equivalently by (2.15). A type II confidence intervals is an interesting alternative because a construction of a type I confidence interval usually requires density estimation.

Theorem 4.7 (or the first two statements of Theorem 4.10 for the studentized case) covers the first order results about asymptotic coverage and asymptotic length of the confidence interval.

But it is a more delicate task to find the limiting distribution of the properly standardized length of the confidence interval D_n^{II}.

For every $\mathbf{b} \in B = \{\mathbf{s} : |\mathbf{s}|_2 \leq M, \mathbf{s} \in \mathbb{R}_p\}$ we define the processes

$$\begin{aligned} M_n^{\mathbf{b}}(t) &= \frac{1}{n^{1/4}} \sum_{i=1}^{n} x_{ip} \left[\psi(e_i - \tfrac{\mathbf{b}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} - \tfrac{t\,x_{ip}}{\sqrt{n}}) - \psi(e_i - \tfrac{\mathbf{b}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) \right], \\ \bar{M}_n^{\mathbf{b}}(t) &= M_n^{\mathbf{b}}(t) - \mathsf{E}\, M_n^{\mathbf{b}}(t), \end{aligned}$$

indexed by the set $T = \{t,\ \ |t| \le M\}$.

It is a rather standard use of Theorem 7.12 to show that for every fixed $\mathbf{b} \in B$ the process $\bar{M}_n^{\mathbf{b}}(t)$ converges weakly to a centered gaussian process $\{W(t),\, |t| \le M\}$ with a covariance function

$$\operatorname{cov}(W(s), W(t)) \begin{cases} T_p^3\, \gamma_{01}\, |s| \wedge |t|, & t\, s > 0 \\ 0, & t\, s \le 0, \end{cases} \tag{4.60}$$

where $T_p^3 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n |x_{ip}|^3$ and $\gamma_{01} = \frac{\partial}{\partial t}(\mathsf{E}\ \psi^2(e_1 + t))_{t=0} = \sum_{j=1}^m \alpha_j^2\, [f(q_j) - f(q_{j-1})]$. In the following, we would like to show that this convergence is uniform in $\mathbf{b} \in B$.

By a partition of T we will mean a decomposition of T into finitely many disjoint subsets $T_1, \ldots, T_N$ such that $T = \bigcup_{j=1}^N T_j$. Choose from each partitioning set T_j a fixed element and denote it by t_j. Finally, define the map $\pi : T \to \{t_1, \ldots, t_N\}$ as $\pi(t) = t_j$ if $t \in T_j$.

Lemma 4.11. *Let the conditions **X'.1-3** and **Step.1-2** be satisfied. Then for every $\varepsilon > 0$ there exists a finite partition $T = \bigcup_{j=1}^N T_j$ and a map $\pi : T \to \{t_1, \ldots, t_N\}$ such that for all sufficiently large $n \in \mathbb{N}$*

$$\sup_{|\mathbf{b}|_2 \le M} \mathsf{E}^*\, \|\bar{M}_n^{\mathbf{b}}(t) - \bar{M}_n^{\mathbf{b}}(\pi(t))\|_T < \varepsilon. \tag{4.61}$$

Proof. Similarly to the proof of Theorem 2.7 let us define the metric ρ on T as $\rho(t,s) = C\sqrt{|t-s|_2}$, where C is a (large) constant. Notice that for this metric

$$N(\varepsilon, T, \rho) \le \left(\tfrac{2\, M\, C^2}{\varepsilon^2}\right) \wedge 1,$$

which implies condition (7.4). The other obvious but important fact is that the metric ρ does not depend on $\mathbf{b}$. Denote $B(\varepsilon)$ ($\subset T$) a ρ-ball of radius ε. Then it is easy to show that by taking C large enough for all sufficiently large $n \in \mathbb{N}$ it holds

$$\sup_{|\mathbf{b}|_2 \le M} \sum_{i=1}^n \mathsf{E}^* \sup_{t,s \in B(\varepsilon)} \Big[M_{ni}^{\mathbf{b}}(t) - M_{ni}^{\mathbf{b}}(s)\Big]^2 \le \varepsilon^2, \tag{4.62}$$

where $M_{ni}^{\mathbf{b}}(t) = \frac{x_{ip}}{n^{1/4}} \Big[\psi(e_i - \frac{\mathbf{b}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} - \frac{t\, x_{ip}}{\sqrt{n}}) - \psi(e_i - \frac{\mathbf{b}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})\Big]$.

By the assumptions of the lemma and by the previous inequality (4.62) we can for every $q \in \mathbb{N}$ construct a partition $T = \bigcup_{j=1}^{N_q} T_j^q$ such that

$$\sum_{q=1}^{\infty} 2^{-q} \sqrt{\log N_q} < \infty$$

and

$$\sup_{|\mathbf{b}|_2 \le M} \sum_{i=1}^n \mathsf{E}^* \sup_{t,s \in T_j^q} \Big[M_{ni}^{\mathbf{b}}(t) - M_{ni}^{\mathbf{b}}(s)\Big]^2 < \frac{1}{2^q}, \quad \text{for } j = 1, \ldots, N_q.$$

We can argue similarly to the proof of Theorem 2.5.6 of van der Vaart and Wellner (1996) to show that without loss of generality we can choose the sequence of partitions (for $q = 1, 2, \ldots$) as successive refinements.

Now we can follow step by step the proof of Theorem 7.12 (2.11.11 Theorem of van der Vaart and Wellner (1996)) and show that there exists a sufficiently large $q_\circ$ such that for all $q \geq q_\circ$ and for all n large enough

$$\mathsf{E}^* \|\bar{M}_n^{\mathbf{b}}(t) - \bar{M}_n^{\mathbf{b}}(\Pi_q(t))\|_T < \varepsilon.$$

As the construction of the partitions does not depend on $\mathbf{b}$, we can take q_0 sufficiently large so that the last inequality holds uniformly in $\mathbf{b}\,(|\mathbf{b}|_2 \leq M)$, which proves the statement of the lemma. $\square$

In the following, we will make use of the bounded Lipschitz metric (see e.g. van der Vaart and Wellner (1996)). This metric is an important tool as it metrizes the weak convergence to a separable Borel limit.

Definition 4.12. Let Q_1, Q_2 be two probability measures on $\ell^\infty(T)$ and BL_1 be the set of all real functions f on T with $\|f\|_\infty \leq 1$ and $|f(x) - f(y)| \leq d(x,y)$, for every $x, y \in T$. Then the **bounded Lipschitz metric** of the measures Q_1 and Q_2 is defined as

$$d_{BL}(Q_1, Q_2) = \sup_{f \in BL_1} \left| \int f\, dQ_1 - \int f\, dQ_2 \right|.$$

Lemma 4.13. *The process* $\{M_n^{\mathbf{b}}(t),\, t \in T\}$ *converges weakly to a gaussian process* $\{W(t),\, t \in T\}$ (*with the covariance function specified in* (4.60)) *uniformly in* $\mathbf{b} \in B$, *that is*

$$\sup_{|\mathbf{b}|_2 \leq M} d_{BL}(M_n^{\mathbf{b}}(\cdot), W(\cdot)) \to 0, \quad \textit{as} \quad n \to \infty.$$

Proof. We will very closely follow the second part of the proof of Theorem 2.8.2 of van der Vaart and Wellner (1996).

Fix $\varepsilon > 0$. With the help of Lemma 4.11 and by the fact that the limiting process $W(\cdot)$ is gaussian, we can find a finite set $T_0 = \{t_1, \ldots, t_N\}$ ($T_0 \subset T$) and a mapping $\pi : T \to T_0$ such that for all sufficiently large n

$$\sup_{|\mathbf{b}|_2 \leq M} \mathsf{E}^* \|\bar{M}_n^{\mathbf{b}}(t) - \bar{M}_n^{\mathbf{b}}(\pi(t))\|_T < \varepsilon^2 \tag{4.63}$$

as well as

$$\mathsf{E}\, \|W(t) - W(\pi(t))\|_T < \varepsilon^2. \tag{4.64}$$

Now we claim that

$$\mathbf{Z}_n^{\mathbf{b}} = \left(\bar{M}_n^{\mathbf{b}}(t_1), \ldots, \bar{M}_n^{\mathbf{b}}(t_N)\right)^{\mathsf{T}} \xrightarrow[n\to\infty]{w} (W(t_1), \ldots, W(t_N))^{\mathsf{T}} \text{ uniformly in } \mathbf{b} \in B. \tag{4.65}$$

This may be justified as follows. Denote by μ_n the distribution of the random vector $\mathbf{Z}_n^{\mathbf{b}}$ and put

$$\mathbf{X}_i = (X_1, \ldots, X_N)^{\mathsf{T}} = n^{1/4} \left(\bar{M}_{ni}^{\mathbf{b}}(t_1), \ldots, \bar{M}_{ni}^{\mathbf{b}}(t_N)\right)^{\mathsf{T}},$$

where

$$M_{ni}^{\mathbf{b}}(t) = x_{ip}\left[\psi(e_i - \tfrac{\mathbf{b}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} - \tfrac{\mathbf{t}_1^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - \psi(e_i - \tfrac{\mathbf{b}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})\right] \quad \text{and} \quad \bar{M}_{ni}^{\mathbf{b}}(t) = M_{ni}^{\mathbf{b}}(t) - \mathsf{E}\; M_{ni}^{\mathbf{b}}(t).$$

Then μ_n coincides with the distribution of the vector $\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{X}_i$. We immediately see that $\mathsf{E}\;|\mathbf{X}_i|_2^3 < \infty$ and $\rho_3 = \frac{1}{n}\sum_{i=1}^n \mathsf{E}\;|\mathbf{X}_i|_2^3 = O(n^{1/4})$ uniformly in B. By Theorem 7.19 the Prohorov distance between measure μ_n and the Gaussian measure ν_n with mean zero and with the same covariance matrix as μ_n is of order smaller then $o(n^{-1/32})$ uniformly in B. Finally, as the covariance matrix of ν_n converges to the covariance matrix of the random vector $\{W(t),\, t \in T_0\}$ uniformly in B, we can use Theorem 7.20 (with the subsequent discussion) to conclude the proof of (4.65).

The uniform convergence in (4.65) further implies that

$$\sup_{\mathbf{b}\in B}\sup_{h\in \mathrm{BL}_1}\left|\mathsf{E}^*\, h(\bar{M}_n^{\mathbf{b}}(\pi(\cdot))) - \mathsf{E}\; h(W(\pi(\cdot)))\right| \to 0. \tag{4.66}$$

Next, since every $h \in \mathrm{BL}_1$ satisfies the inequality $|h(x) - h(y)| \leq 2 \wedge |x-y|$, we get that uniformly in B for every $\varepsilon > 0$

$$\begin{aligned}
\sup_{h\in \mathrm{BL}_1}\left|\mathsf{E}^*\, h(\bar{M}_n^{\mathbf{b}}(\cdot)) - \mathsf{E}\; h(\bar{M}_n^{\mathbf{b}}(\pi(\cdot)))\right| &\leq 2 \wedge \mathsf{E}^*\left\|\bar{M}_n^{\mathbf{b}}(\cdot)) - \bar{M}_n^{\mathbf{b}}(\pi(\cdot))\right\|_T \\
&\leq \varepsilon + 2\ \mathrm{P}^*\left\{\|\bar{M}_n^{\mathbf{b}}(t) - \bar{M}_n^{\mathbf{b}}(\pi(t))\|_T > \varepsilon\right\} \\
&\leq \varepsilon + \frac{2\ \mathsf{E}^*\,\|\bar{M}_n^{\mathbf{b}}(t) - \bar{M}_n^{\mathbf{b}}(\pi(t))\|_T}{\varepsilon} \overset{(4.63)}{\leq} 3\,\varepsilon.
\end{aligned} \tag{4.67}$$

As the limit process W is gaussian, we can obtain an analogous result for the process W.

Combining the previous results (4.64), (4.66) and (4.67) we get that for sufficiently large n uniformly in B

$$\begin{aligned}
&\sup_{h\in \mathrm{BL}_1}\left|\mathsf{E}^*\, h(\bar{M}_n^{\mathbf{b}}(\cdot)) - \mathsf{E}\; h(W(\cdot))\right| \\
&\quad\leq \sup_{h\in \mathrm{BL}_1}\left|\mathsf{E}^*\, h(\bar{M}_n^{\mathbf{b}}(\cdot)) - \mathsf{E}\; h(\bar{M}_n^{\mathbf{b}}(\pi(\cdot)))\right| + \sup_{h\in \mathrm{BL}_1}\left|\mathsf{E}\; h(\bar{M}_n^{\mathbf{b}}(\pi(\cdot))) - \mathsf{E}\; h(W(\pi(\cdot)))\right| \\
&\qquad\qquad + \sup_{h\in \mathrm{BL}_1}\left|\mathsf{E}\; h(W(\pi(\cdot))) - \mathsf{E}\; h(W(\cdot))\right| \\
&\qquad\leq \sup_{h\in \mathrm{BL}_1}\left|\mathsf{E}\; h(\bar{M}_n^{\mathbf{b}}(\pi(\cdot))) - \mathsf{E}\; h(W(\pi(\cdot)))\right| + 3\,\varepsilon + 3\,\varepsilon \xrightarrow[n\to\infty]{} 6\,\varepsilon,
\end{aligned}$$

which concludes the proof of the lemma. $\square$

Before we find the asymptotic distribution of the length of the confidence interval, notice that uniformly in $t \in T$ and $\mathbf{b} \in B$

$$\mathsf{E}\; M_n^{\mathbf{b}}(t) + t\,\gamma_1\, n^{1/4}\, T_{np}^2 = o(1). \tag{4.68}$$

Theorem 4.14. *Put*

$$L_n = \frac{n^{1/4}[\sqrt{n}(\hat{b}_n^+ - \hat{b}_n^-) - 2\, a_F^n]}{2\, a_F}, \tag{4.69}$$

then the random variable L_n is asymptotically normally distributed with mean zero and the variance

$$\sigma_n^2 = \frac{\gamma_{01}\, T_{np}^3}{2\,\gamma_1^2\,(T_{np}^2)^2\, a_F}, \qquad \textit{where} \quad T_{np}^3 = \frac{1}{n}\sum_{i=1}^n |x_{ip}|^3. \tag{4.70}$$

Proof. First, notice that Lemma 4.8 implies $\hat{\sigma}_\psi - \sigma_\psi = o_p(\frac{1}{n^{1/4}})$. That is why, we can replace the estimate $\hat{\sigma}_\psi$ in the definition of the confidence interval with the true value σ_ψ.

Further, with the help of Lemma 4.13, equation (4.68), and the $\sqrt{n}$-consistency of $\hat{\boldsymbol{\beta}}_n$ we get that the process

$$W_n(t) = \frac{1}{n^{1/4}}\sum_{i=1}^n x_{ip}\,\psi(e_i - \tfrac{\hat{\boldsymbol{\beta}}_n^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} - \tfrac{t\,x_{ip}}{\sqrt{n}}) - \frac{1}{n^{1/4}}\sum_{i=1}^n x_{ip}\,\psi(e_i - \tfrac{\hat{\boldsymbol{\beta}}_n^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) + \gamma_1\, T_{np}^2\, n^{1/4}\, t, \quad t \in T,$$

weakly converges to the process $W(t)$. As the random variables $\sqrt{n}(\hat{b}_p^+ - \hat{\beta}_p)$ and $\sqrt{n}(\hat{b}_p^- - \hat{\beta}_p)$ are in probability equivalent to a_F^n and $-a_F^n$ respectively, we can substitute them for t in the process W_n. This substitution and subtraction of the two resulting equations yield

$$n^{1/4}\gamma_1 T_{np}^2 \left[\sqrt{n}(\hat{b}_p^+ - \hat{b}_p^-) - 2\, a_F^n\right] \sim AN(0,\, 2\,\gamma_{01}\, T_{np}^3\, a_F^n).$$

But as the random variable L_n equals the quantity on the left-hand side divided by $2\,\gamma_1 T_{np}^2 a_F^n$, the theorem is proved.

□

The situation is very similar for the **studentized estimators**. In this case we need to study the process

$$\{M_n^{\mathbf{b},u}(t) = \frac{1}{n^{1/4}}\sum_{i=1}^n x_{ip}\left[\psi\left(e^{-u/\sqrt{n}}(e_i - \tfrac{\mathbf{b}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} - \tfrac{t\,x_{ip}}{\sqrt{n}})/S\right) - \psi\left(e^{-u/\sqrt{n}}(e_i - \tfrac{\mathbf{b}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\right)\right], \quad |t| \le M\},$$

where $|\mathbf{b}|_2 \le M$ and $|u| \le M$.

We can proceed along the lines of Lemma 4.11 and Lemma 4.13 and show that the process $\{M_n^{\mathbf{b},u}(t),\, |t| \le M\}$ converges uniformly in $\mathbf{b} \in B$ and $u\,(|u| \le M)$ to the gaussian process $\{W(\frac{t}{S}),\, |t| \le M\}$. Thus if $\sqrt{n}(\frac{S_n}{S} - 1) = O_p(1)$ and the model (1.1) includes an intercept, no new complications arise. We have to only incorporate the scale functional S into our formulae.

Numerical illustration

Just to get an idea, how does the confidence interval of type II work in practice, we performed a small Monte-Carlo experiment. We considered a simple linear model with one explanatory

	$\mathbf{n = 20}$		$\mathbf{n = 50}$		$\mathbf{n = 100}$		$\mathbf{n = 200}$	
	I	II	I	II	I	II	I	II
Coverage	0.989	0.955	0.982	0.946	0.974	0.944	0.968	0.947
$\sqrt{n}$ mean(length)	8.413	6.900	6.933	5.663	6.314	5.413	5.950	5.318
$\sqrt{n}$ median(length)	8.122	6.583	6.859	5.530	6.281	5.325	5.935	5.254
$n^{3/4}$ sd(length)	4.913	5.418	3.129	4.260	2.503	4.110	2.105	4.084
$n^{3/4}$ IQR(length)	4.612	5.124	3.071	4.197	2.500	4.082	2.101	4.070

Table 3: Comparison of type I and type II CI; the nominal coverage is 0.95.

variable (plus intercept) $Y_i = \beta_0 + \beta_1 x_i + e_i$ and the median regression estimator with the ψ function given by $\psi(x) = \mathbb{I}\{x > 0\} - \frac{1}{2}$. To construct the type I confidence interval, we need to estimate the functional $\gamma_1 = f(F^{-1}(0))$ (density of the errors evaluated at the median of the distribution). In our study we used the estimate originally suggested by Siddiqui (1960) (see p. 139 of Koenker (2005)).

We generated the errors from t-distribution with 3 degrees of freedom and the design points from the uniform distribution on the interval $(-1, 1)$. We set the nominal coverage to 0.95. We took the sample sizes $20, 50, 100,$ and 200 respectively. The number of random samples was 100 000.

Table 3 shows some of the results, which seem to be typical. We see that while the type I confidence interval is rather conservative, the actual coverage of the type II confidence interval is slightly less then the prescribed value. The fact that speaks for the type II CI is its average shorter length. On the other hand we may be rather nervous that the length of the CI of type II is much more variable than the length of the CI of type I.

We can also compare the quantities $\sqrt{n}\,$mean(length) and $n^{3/4}\,$sd(length) from the Table 3 with its asymptotic counterparts $2\,a_F = 5.33$ and $\sigma_n' = \sigma_n\, a_F = 4.33$, where the quantities a_F and σ_n are given by the equations (4.40) and (4.70). We see that for $n \geq 100$ the approximation of the mean length CI (multiplied by $\sqrt{n}$) of the confidence interval by $2\,a_F$ works very satisfactorily. On the other hand the asymptotic variance of the length of CI overestimates the true variance and it gives only a rough idea unless the sample size is very large. Some further simulations show that the approximation of variance needs the sample size to be in thousands to be trustworthy. Finally, we should note that the comparisons made in this paragraph are only exploratory and not mathematically correct. The problem is that Theorem 4.7 and Theorem 4.14 only speak about the convergence in probability and distribution, none of which implies the convergence of moments. The justification of comparisons of finite sample mean (or variance) with the mean (or variance) of the asymptotic distribution would require to show the uniform integrability of the sequence in question. But this is beyond the scope of this thesis.

Some further numerical experiments show that a type II CI is not very convenient for more than one explanatory variables, as its actual coverage is considerably smaller than the prescribed value unless the sample size is very large. Moreover, even if there is only one explanatory variable the undercoverage of the type II CI is considerable higher if the prescribed nominal value is less than 0.95.

We conclude that a type II CI is a good alternative to a type I CI for the models with one explanatory variable (and an intercept) in particular for small and moderate sample sizes. Finally, unless the sample size is extremely large, type II CI's usually work better for asymmetric errors too.

4.2.5 R-estimators based on Wilcoxon scores

The situation is analogous to the case of an absolutely continuous ψ. For the simplicity of notation, we will construct the confidence interval for the last coordinate of the vector $\boldsymbol{\beta}$. Let $\hat{\boldsymbol{\beta}}_n$ be the R-estimator and $r_1, \ldots, r_n$ the residuals, that is $r_i = Y_i - \hat{\boldsymbol{\beta}}_n^{\mathsf{T}} \mathbf{x}_i$ for $i = 1, \ldots, n$. We define $S_{np}(t) = \frac{1}{n^{3/2}} \sum_{i=1}^n x_{ip} R_i(t)$, where $R_i(t)$ is the rank of the random variable $r_i - t\, x_{ip}$ among $r_1 - t\, x_{1p}, \ldots, r_n - t\, x_{np}$.

Then the (type II) confidence interval for the parameter β_p can be constructed as $D_n^{II} = [\hat{b}_p^-, \hat{b}_p^+] = [\hat{\beta}_p + \delta_n^-, \hat{\beta}_p + \delta_n^+]$, where

$$\delta_n^- = \sup\left\{t < 0 : S_{np}(t) \geq \frac{T_{np}^2\, z_\alpha \sqrt{\omega_{pp}^n}}{\sqrt{12}}\right\}, \qquad \delta_n^+ = \inf\left\{t > 0 : S_{np}(t) \leq \frac{-T_{np}^2\, z_\alpha \sqrt{\omega_{pp}^n}}{\sqrt{12}}\right\}, \tag{4.71}$$

and $z_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$, with Φ^{-1} being the inverse cdf of the standard normal distribution. Notice that we do not need to estimate any unknown parameters.

The following theorem is an analogy to Theorems 4.7 and 4.9. In fact, it only restates the results of Section 5 of Jurečková (1973).

For the simplicity of notations we put

$$a_F^n = \frac{z_\alpha \sqrt{\omega_{pp}^n}}{\gamma \sqrt{12}} \qquad \text{and} \qquad a_F = \lim_{n\to\infty} a_F^n \overset{\textbf{XX.2}}{=} \frac{z_\alpha \sqrt{\omega_{pp}}}{\gamma \sqrt{12}},$$

where $\gamma = \mathsf{E}\, f(e_1) = \int f^2(x) dx$.

Theorem 4.15. *If the conditions **XX.1-2**, **X.5**, **W.1-3**, and the representation* (4.22) *hold, then the confidence interval D_n^{II} defined by* (4.71) *satisfies:*

(i).

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

(ii).

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

(iii). Put

$$L_n = \frac{\sqrt{n}[\sqrt{n}(\hat{b}_p^+ - \hat{b}_p^-) - 2\, a_F^n]}{2\, a_F}, \tag{4.72}$$

then the random variable L_n is asymptotically normal and admits the first order asymptotic representation

$$L_n = -\frac{1}{\gamma\sqrt{n}\,T_{np}^2}\sum_{i=1}^{n}\left[x_{ip}^2 + \sum_{j=1}^{n}\frac{x_{jp}^2}{n}\right][f(e_i)-\gamma] + o_p(1). \tag{4.73}$$

Proof. Completely analogous to the proof of Theorem 4.7 and Theorem 4.9. The first two statements are 'first' order results, which can be deduced from (4.22). In the proof of the last statement we utilize the asymptotic expansion derived in Corollary 3.3. □

Remark 19. The confidence interval of type I for the parameter β_p would be

$$D_n^I = [\hat{b'}_p^-, \hat{b'}_p^+] = \left[\hat{\beta}_p - \tfrac{z_\alpha}{\sqrt{n}}\tfrac{\sqrt{\omega_{pp}}}{\hat{\gamma}\sqrt{12}},\ \hat{\beta}_p + \tfrac{z_\alpha}{\sqrt{n}}\tfrac{\sqrt{\omega_{pp}}}{\hat{\gamma}\sqrt{12}}\right], \tag{4.74}$$

where $\hat{\gamma}$ is an estimate of $\gamma = \mathsf{E}\ f(e_1)$, with f being the density of the distribution of errors. As the density f is unknown, it is not so straightforward to estimate the functional γ. Some estimators of γ can be found in Hettmansperger (1984).

Numerical illustration

We performed some simulations to illustrate the results of Theorem 4.15. We considered a linear model $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$. We used the Meyer matrix of order 27×2 (see Stigler (1986), pp. 16–25) as the design matrix. Further, we normalized this matrix such that $\sum_{i=1}^{n} x_{ij} = 0$ and $\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 = 1$ for $j = 1, 2$. We were interested in type I (R I) and type II (R II) 95% confidence intervals for the parameter β_2.

The type I CI requires the estimation of the functional γ. In our simulation we used the estimate constructed as follows.

1. Denote
$$H_n(t) = \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{I}\{|r_i - r_j| \le t\}$$

the distribution function of the pairwise differences of the residuals $r_i = Y_i - \hat{\boldsymbol{\beta}}_n^{\mathsf{T}}\mathbf{x}_i$.

2. Put $\tau_n = \frac{H_n^{-1}(0.8)}{\sqrt{n}}$
3. Finally estimate γ by
$$\hat{\gamma} = \frac{\sqrt{3}\,H(\tau_n)}{\tau_n}\sqrt{\frac{n-p-1}{n}}.$$

Once we have estimated γ, we can define the type I CI by (4.74).

First, we were interested in small sample coverages and mean lengths of the type I and type II CI's. Some of the results, for the errors generated from standard normal distribution ($N(0,1)$), logistic distribution with the density $f(x) = \frac{e^x}{(1+e^x)^2}$ (logistic), and exponential distribution with $f(x) = e^{-x}\mathbb{I}\{x > 0\}$ (exp) are to be found in Table 4. The number of random samples was 100 000. The first row of this table gives us the estimated two-sided coverage probability, the second (the third) row estimates the one-sided coverage probability

$n = 27$	$N(0,1)$		logistic		exp	
	R I	R II	R I	R II	R I	R II
Coverage	0.955	0.961	0.956	0.961	0.954	0.958
Coverage L	0.977	0.980	0.978	0.981	0.968	0.977
Coverage U	0.978	0.981	0.977	0.980	0.986	0.981
mean(length)	0.870	0.866	1.487	1.496	0.604	0.661
$\sqrt{n}$ mean(length)	4.520	4.498	7.725	7.776	3.136	3.436
$\sqrt{n}$ median(length)	4.481	4.466	7.625	7.675	3.071	3.348
n sd(length)	4.463	4.057	8.256	7.836	3.884	4.527
n IQR(length)	4.441	4.041	8.145	7.711	3.791	4.374

Table 4: Actual coverage probabilities of the true value of the parameter β_2 for the sample size $n = 27$.

	n = 27		**n = 54**		**n = 108**		**n = 216**	
	R I	R II	R I	R II	R I	R II	R I	R II
Coverage	0.959	0.958	0.956	0.953	0.952	0.957	0.947	0.951
Coverage L	0.970	0.976	0.971	0.974	0.970	0.976	0.969	0.976
Coverage U	0.989	0.983	0.986	0.979	0.983	0.981	0.978 .976	
mean(length)	0.832	0.932	0.508	0.541	0.317	0.344	0.218	0.230
$\sqrt{n}$ mean(length)	4.323	4.840	3.737	3.973	3.297	3.572	3.204	3.382
$\sqrt{n}$ median(length)	4.174	4.601	3.677	3.877	3.272	3.534	3.190	3.361
$2\,a_F$	3.297	3.297	3.217	3.217	3.180	3.180	3.162	3.162
n sd(length)	6.209	8.088	5.060	6.242	4.755	5.324	4.628	4.994
n IQR(length)	5.896	7.340	4.956	5.910	4.690	5.221	4.528	4.843
n asd(length)	*	4.298	*	4.194	*	4.145	*	4.121

Table 5: Results on confidence intervals for β_2 for lognormal errors and different sample sizes.

$\mathrm{P}\,(\hat{b}_p^- < \beta_2)$ $(\mathrm{P}\,(\hat{b}_p^+ > \beta_2))$. The next two rows measures the mean and median length of the confidence intervals and the final two rows the variability of the length of the confidence intervals. By sd we mean standard deviation and by IQR interquantile range divided by $2\,\Phi^{-1}(\frac{3}{4})$ (so that IQR is consistent to σ if the underlying distribution is normal).

We see that for symmetric errors both methods are almost equivalent. We were surprised that type II method seems to work better for normal errors. Some further experiments show that the picture is usually following. The type II method is usually a little bit more conservative in terms of coverage probability and resulting CI's are on average larger and more variable than the type I CI's. Let us notice more thoroughly what happens if the errors are asymmetric. The last two columns of Table 4 as well as Table 5 indicate that although the two-sided type I CI's keep the nominal value very closely, the one-sided confidence intervals may be slightly misleading.

Second, we wanted to assess the statements (*ii*) and (*iii*) of Theorem 4.15. We chose the sample sizes n = 27, 54, 108 and 216 (we used appropriate multiples of Meyer matrix) and estimated the mean length of CI's (multiplied by $\sqrt{n}$) and standard deviation of this length (multiplied by n). We compared these empirical results with their theoretical counterparts – $2\,a_F$ and asymptotical standard deviation of L_n ($asd(L_n)$) based on the asymptotic expansion (4.73). The number of samples was in each case at least 10 000. Notice that we are able to calculate $asd(L_n)$ only for type II method. The calculation of this quantity for type I method would require a detailed study of the estimate $\hat{\gamma}$.

Table 5 contains the results for errors following lognormal distribution (with density given by the formula $f(x) = \frac{1}{x\sqrt{2\pi}} \exp\{-\frac{\log^2(x)}{2}\} \mathbb{I}\{x > 0\}$). Comparing the finite sample results with their asymptotic values (the seventh and the last column of the table), we see that to approximate the mean and in particular the variance of length of CI's with their asymptotic values is too optimistic, even in the situations with more than one hundred observations and only two explanatory variables. But to be fair, we chose one of the worst cases – with heavily asymmetric errors (lognormal). For symmetric (e.g. normal) errors, the asymptotic approximations work for $n > 100$ satisfactory, provided that the distribution of the columns of the explanatory variable is not heavily skewed.

Chapter 5

A bounded length confidence interval

In this chapter we explore some asymptotic properties of a bounded length confidence interval for a single parameter, which is based on a (studentized) M-estimator or an R-estimator generated with the Wilcoxon scores. The results for M-estimators generalize the work of Jurečková and Sen (1981a) and Jurečková and Sen (1981b).

5.1 Preliminaries

It is natural that sometimes we would like to estimate the parameter of interest with a prescribed precision. But as the sampling distribution of the estimate usually depends on some unknown (nuisance) parameters, which we mostly do not know in practice, we need to incorporate a sequential procedure.

One of such procedures is a bounded length confidence interval. We can generally describe it as follows. Suppose we are estimating a scalar parameter θ and for every fixed sample size n we are able to construct an asymptotically correct confidence interval D_n. Denote L_n the length of this interval. Now we prescribe the quantity d (the 'precision' of interval) and we sample unless the length of the interval is shorter than $2d$. More precisely, we denote the stopping variable

$$N_d = \inf\{n \geq n_0 : L_n \leq 2d\}, \tag{5.1}$$

where n_0 may be interpreted as an initial (or the smallest reasonable) sample size. For the resulting confidence interval D_{N_d} holds $L_{N_d} \leq 2d$.

This procedure gives rise to some natural questions.

(i). What is the actual coverage probability?

(ii). Can we describe or at least approximate the behaviour of the stopping variable N_d?

Concerning the first question, we would like to show that $\lim_{d\to 0_+} \mathrm{P}\,(D_{N_d} \ni \theta) = 1-\alpha$, which would justify our approach at least asymptotically. As some numerical experiments show that actual coverage for a fixed length d may be substantially smaller than nominal coverage, some authors propose to define the stopping variable as

$$N_d = \inf\{n \geq n_0 :\, L_n + r_n \leq 2d\},$$

where r_n is a penalty term (e.g. $r_n = \frac{1}{n}$). This penalty should prevent very early stopping, which results in taking too few observations. For simplicity, we will not consider the penalty in what follows. By the definition of the stopping variable N_d we will mean (5.1).

Let us turn to the question (ii). As we will see later, we are usually able to find a nonrandom quantity n_d (depending on d) such that $\frac{N_d}{n_d} \xrightarrow[d\to 0_+]{\mathrm{P}} 1$. Sometimes we can even show that the random variable $\frac{N_d}{n_d}$, properly standardized, is asymptotically normally distributed.

5.2 Definitions and Theorem

Suppose that the model (1.1) includes an intercept, that is $x_{i1} = 1$ for $i = 1, \dots, n$. For simplicity of notation we will be interested in β_p (the last component of a regression parameter $\boldsymbol{\beta}$).

In the following, we would like to explore the asymptotic properties of a bounded length confidence interval $D^{II}_{N_d}$, where D^{II}_n is defined with the help of (4.57) and (4.58). The stopping variable N_d is given by

$$N_d = \inf\{n \geq n_0 :\, \hat{b}^{+}_{p,n} - \hat{b}^{-}_{p,n} \leq 2d\}, \tag{5.2}$$

where the symbol n in the subscript indicates the number of observations used to construct the estimate. Provided condition **XX.2** holds, put $\mathbf{V}_n^{-1} = [\omega^n_{ij}]^p_{i,j=1}$ and $\mathbf{V}^{-1} = [\omega_{ij}]^p_{i,j=1}$. Define

$$n_d = \frac{z^2_\alpha\, \sigma^2_\psi\, \omega_{pp}}{\gamma^2_1\, d^2} = \frac{a^2_F}{d^2}, \qquad \text{where} \qquad a_F = \lim_{n\to\infty} \frac{z_\alpha\, \sigma_\psi \sqrt{\omega^n_{pp}}}{\gamma_1}. \tag{5.3}$$

Now we are ready to formulate the basic properties of this sequential procedure.

Theorem 5.1. *Under the assumptions **XX.1-2**, **SmSt.1-3** (or **Step.1-2**), and **GenSt.1** it holds:*

(i). *N_d is nonincreasing in d $(d > 0)$;*

(ii). $\mathrm{P}\,(N_d < \infty) = 1$ *for any $d > 0$;*

(iii). $\lim_{d\to 0_+} N_d = \infty$ *a.s.;*

(iv). $\frac{N_d}{n_d} \xrightarrow[d\to 0_+]{\mathrm{P}} 1.$

Proof. The proof of the statements is analogous to the proof of Theorem 3.1.1 of Jurečková (1978). We show it only for the sake of completeness.

Proof of (i). The monotonicity of N_d follows directly from the definition of N_d.

Proof of (ii). For any fixed $d > 0$ we can write

$$\mathrm{P}\left(N_d = \infty\right) = \mathrm{P}\left\{\bigcap_{n=1}^{\infty}[N_d > n]\right\} \leq \lim_{n\to\infty} \mathrm{P}\left(L_n > 2d\right) = 0,$$

where the last equality holds in view of Theorem 4.10.

Proof of (iii). $\lim_{d\to 0_+} N_d = \infty$ if and only if

$$\mathrm{P}\left\{\bigcup_{K=1}^{\infty}\bigcap_{d>0}\bigcup_{d'<d}[N_{d'} \leq K]\right\} = 0. \tag{5.4}$$

Due to the monotonicity of N_d, the left-hand side of (5.4) equals to

$$\mathrm{P}\left\{\bigcup_{K=1}^{\infty}\bigcap_{m=1}^{\infty}[N_{\frac{1}{m}} \leq K]\right\} = \mathrm{P}\left\{\bigcup_{K=1}^{\infty}\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{K}[L_n \leq \tfrac{2}{m}]\right\} = \mathrm{P}\left\{\bigcup_{K=1}^{\infty}\bigcup_{n=1}^{K}[L_n = 0]\right\} = 0$$

as $\mathrm{P}\left(L_n > 0\right) = 1$ for any $n \in \mathbb{N}$.

Proof of (iv). Fix $\varepsilon > 0$ and put $n_{d\varepsilon} = \lfloor n_d(1+\varepsilon)\rfloor$. Then for sufficiently small d

$$\begin{aligned}\mathrm{P}\left\{\tfrac{N_d}{n_d} > 1+\varepsilon\right\} &= \mathrm{P}\left\{L_n > 2d, \forall n_0 \leq n \leq n_{d,\varepsilon}\right\} \leq \mathrm{P}\left\{\sqrt{n_{d,\varepsilon}}\, L_{n_{d,\varepsilon}} > d\sqrt{n_{d,\varepsilon}}\right\} \\ &\overset{(5.3)}{\leq} \mathrm{P}\left\{\sqrt{n_{d,\varepsilon}}\, L_{n_{d,\varepsilon}} > \tfrac{(1+\frac{\varepsilon}{2})\,\sigma_\psi\, z_\alpha \sqrt{\omega_{pp}}}{\gamma_1}\right\} \xrightarrow[d\to 0_+]{} 0,\end{aligned}$$

where the convergence follows by the statement (ii) of Theorem 4.10.

Similarly we can show that $\mathrm{P}\left\{\frac{N_d}{n_d} < 1-\varepsilon\right\} \xrightarrow[d\to 0_+]{} 0.$

□

5.2.1 Asymptotic coverage of the sequential confidence interval

In this section we prove that the sequential confidence interval $D_{N_d}^{II}$ has an asymptotically correct coverage, that is

$$\lim_{d\to 0_+} \mathrm{P}\left(D_{N_d}^{II} \ni \beta_p\right) = 1-\alpha. \tag{5.5}$$

The idea behind the following steps is very simple. From the first statement of Theorem 4.10 we know that the confidence interval $D_n^{II} = [\hat{b}_{pn}^-, \hat{b}_{pn}^+]$ is asymptotically correct. Now we would like to replace the index n with a random stopping variable N_d. As we know that $\frac{N_d}{n_d} \xrightarrow[d\to 0_+]{\mathrm{P}} 1$, then by the results of Anscombe (1952) all we need is to show that the sequences of random variables $\sqrt{n}(\hat{b}_{pn}^- - \beta_p)$ and $\sqrt{n}(\hat{b}_{pn}^+ - \beta_p)$ are both uniformly continuous in probability.

Definition 5.2. We say that a sequence of random variables $\{Z_n\}$ is **uniformly continuous in probability** (ucp) if given any small positive ε and η, there exist $n_\circ \in \mathbb{N}$ and δ such that for any $n > n_\circ$

$$\mathrm{P}\,\{|Z_n - Z_m| < \varepsilon, \quad \forall m \in \mathbb{N} : \, |n-m| < \delta\, n\} > 1 - \eta.$$

In the following, the symbol $o_{upc}(1)$ will stand for a random variable R_n which is of order $o_p(1)$ and which is uniformly continuous in probability. Further in this chapter we will use $\hat{\boldsymbol{\beta}}_n$ instead of $\hat{\boldsymbol{\beta}}_n$ (or $\hat{\boldsymbol{\beta}}_n$) to emphasize that the estimator depends on the sample size.

As a first step, we would like to prove that the remainder term in FOAR of the random variable $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ (4.7) is not only $o_p(1)$, but $o_{upc}(1)$. For this purpose, we need several technical lemmas. Similarly to Section 2.1.3 put $T = \{(\mathbf{t}, u) : |\mathbf{t}|_2 \leq M, |u| \leq M\}$ $(\subset \mathbb{R}^{p+1})$.

Lemma 5.3. *Let the assumptions **X.1-2** and **SmSt.1-3** (or **Step.1-2**) be satisfied and define*

$$M_{nk}(\mathbf{t}, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{k} c_i \left[\psi\left(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\right) - \psi(e_i/S)\right], \qquad (\mathbf{t}, u) \in T.$$

Then for every $\varepsilon > 0$ and $\eta > 0$ there exists $n_\circ$ such that for all $n > n_\circ$

$$\mathrm{P}^*\left\{\max_{k=1,\dots,n} \left\| M_{nk}(\mathbf{t}, u) + \tfrac{\gamma_1 \mathbf{t}^\mathsf{T}}{S} \sum_{i=1}^{k} \tfrac{c_i\,\mathbf{x}_i}{n} + \tfrac{\gamma_{1e} u}{S} \sum_{i=1}^{k} \tfrac{c_i}{n} \right\|_T > \varepsilon \right\} < \eta.$$

Proof. For the simplicity of notation put $\bar{M}_{nk} = M_{nk} - \mathsf{E}\ M_{nk}$. From the proof of Theorem 2.1 (or Theorem 2.7 for the case of step ψ) we know that $\sum_{i=1}^{n} \mathsf{E}^*\, \|\bar{Z}_{ni}\|_T^2 = o(1)$, where

$$Z_{ni} = c_i \left[\psi\left(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\right) - \psi(e_i/S)\right].$$

Corollary 7.14 yields

$$\mathrm{P}^*\left\{\max_{k=1,\dots,n} \|\bar{M}_{nk}\|_T > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0. \tag{5.6}$$

Next, we claim that

$$\max_{k=1,\dots,n} \left\| \mathsf{E}\ M_{nk}(\mathbf{t}, u) - \tfrac{\gamma_1 \mathbf{t}^\mathsf{T}}{S} \sum_{i=1}^{k} \tfrac{c_i\,\mathbf{x}_i}{n} + \tfrac{\gamma_{1e} u}{S} \sum_{i=1}^{k} \tfrac{c_i}{n} \right\|_T = o(1), \tag{5.7}$$

which follows easily by the continuity of the first derivatives of the function $\lambda(s, v) = \mathsf{E}\ \psi(\frac{e_1 - s}{S e^v})$ (condition **SmSt.3** or **Step.2**) and by the assumptions **X.1-2**.

The lemma follows from (5.6) and (5.7). $\square$

Remark 20. Lemma 5.3 implies

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} c_i \left[\psi\left(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\right) - \psi(e_i/S)\right] + \tfrac{\gamma_1 \mathbf{t}^\mathsf{T}}{S} \sum_{i=1}^{n} \tfrac{c_i\,\mathbf{x}_i}{n} + \tfrac{\gamma_{1e} u}{S} \sum_{i=1}^{n} \tfrac{c_i}{n} \right\|_T = o_{ucp}(1). \tag{5.8}$$

Now we would like to substitute $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$ and $\sqrt{n}(\frac{S_n}{S} - 1)$ for u in (5.8). To keep the remaining term of order $o_{ucp}(1)$, we need a stronger condition than the $\sqrt{n}$-consistency of $\hat{\boldsymbol{\beta}}_n$ and S_n.

We will say that the sequence of random variables $\{Z_n\}$ satisfies the condition **SUB** ('sequential uniform boundedness') if for every $\eta > 0$ there exist $C\,(C < \infty)$, $\delta\,(\delta > 0)$ and $n_\circ \in \mathbb{N}$ such that for all $n > n_\circ$

$$\mathrm{P}\left\{\max_{k=n_{-\delta},\dots,n_\delta} |Z_k| > C\right\} < \eta, \tag{5.9}$$

where $n_{-\delta} = \lfloor n(1-\delta)\rfloor$ and $n_\delta = \lceil n(1+\delta)\rceil$.

Lemma 5.4. *Let the conditions **SmSt.1-3** and **XX.1-2** be satisfied. If the sequence* $\sqrt{n}(\frac{S_n}{S} - 1)$ *meets the condition **SUB**, then there exists a sequence* $\{\hat{\boldsymbol{\beta}}_n\}$ *of the solutions of the system of equations* (4.1) *such that the sequence* $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ *satisfies the condition **SUB**.*

Proof. Our proof will be only a slight adaptation of the proof of Theorem 5.5.1. of Jurečková and Sen (1996). Let us denote $\mathbf{V}_k = \frac{1}{k}\sum_{i=1}^k \mathbf{x}_i\mathbf{x}_i^\mathsf{T}$ and $\mathbf{v}_1(k) = \frac{1}{k}\sum_{i=1}^k \mathbf{x}_i$. As our model includes intercept, the vector $\mathbf{v}_1(k)$ is just the first column of the matrix $\mathbf{V}_k$. Further put

$$\mathbf{E}_{nk}(\mathbf{t}, S_n) = \tfrac{1}{\sqrt{n}}\sum_{i=1}^k \mathbf{x}_i\,\psi\left((e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S_n\right).$$

Because the sequence $\sqrt{n}(\frac{S_n}{S} - 1)$ meets the condition **SUB**, Lemma 5.3 implies that for an arbitrarily large but fixed constant M in the definition of T

$$\left\|\mathbf{E}_n(\mathbf{t}, S_n) - \mathbf{E}_n(\mathbf{0}, S) + \gamma_1\mathbf{V}_n\mathbf{t} + \gamma_{1e}\mathbf{v}_1(n)\sqrt{n}\,(\tfrac{S_n}{S} - 1)\right\|_T = o_{ucp}(1). \tag{5.10}$$

With the help of this equation we will show that for every $\eta > 0$ there exist $\delta > 0$, $C > 0$ and $n_\circ \in \mathbb{N}$ such that for all $n > n_\circ$

$$\mathrm{P}\left\{\max_{k=n_{-\delta},\dots,n_\delta}\ \sup_{\|\mathbf{t}\|=C} \mathbf{t}^\mathsf{T}\mathbf{E}_k(\mathbf{t}, S_k) \geq 0\right\} < \eta. \tag{5.11}$$

Provided (5.11) holds, we can use Theorem 6.3.4 of Ortega and Rheinboldt (1970) (see Theorem 7.15) to conclude that for all $k = n_{-\delta},\dots,n_\delta$ with probability exceeding $1-\eta$ the system of equations

$$\sum_{i=1}^k \mathbf{x}_i\,\psi\left((e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S_k\right) = 0$$

has a root T_k such that $\|T_k\| \leq C$. Defining $\hat{\boldsymbol{\beta}}_k = \boldsymbol{\beta} + \frac{T_k}{\sqrt{k}}$ gives us the sequence of the solutions of the system of equations (4.1) with the desired property (**SUB**).

Let us return to the proof of (5.11). With the help of (5.10) for every $\varepsilon > 0$ and $\eta > 0$ we can find $n_\circ$, $\delta' > 0$, and $M' > 0$ such that for all $n > n_\circ$, $\delta < \delta'$, and $M > M'$

$$\begin{aligned}
&\mathrm{P}\left\{\max_{k=n_{-\delta},\ldots,n_\delta} \sup_{\|\mathbf{t}\|=M} \mathbf{t}^\top \mathbf{E}_k(\mathbf{t}, S_k) \geq 0\right\} \\
&\quad \leq \mathrm{P}\left\{\max_{k=n_{-\delta},\ldots,n_\delta} \sup_{\|\mathbf{t}\|=M} \left[\mathbf{t}^\top \mathbf{E}_k(0, S) - \gamma_1 \mathbf{t}^\top \mathbf{V}_k \mathbf{t} - \gamma_{1e} \mathbf{v}_1(k)\sqrt{k}(\tfrac{S_k}{S} - 1)\right] \geq -\varepsilon\right\} + \tfrac{\eta}{2} \\
&\quad \leq \mathrm{P}\left\{\max_{k=n_{-\delta},\ldots,n_\delta} |M|\,|\mathbf{E}_k(0,S)|_2 + |\gamma_{1e}|\,|\mathbf{v}_1(k)|_2 \left|\sqrt{k}\,(\tfrac{S_k}{S} - 1)\right| \geq \gamma_1 M^2 \lambda_1(k) - \varepsilon\right\} + \tfrac{\eta}{2},
\end{aligned} \tag{5.12}$$

where $\lambda_1(k)$ is the smallest eigenvalue of the matrix $\mathbf{V}_k$. We see that we can make the last probability in (5.12) arbitrarily small by taking δ sufficiently small and $n_\circ$ and M sufficiently large. This completes the proof of the lemma. □

As the proof of Lemma 5.4 rests on the continuity of the function ψ, it has to be modified for a step function ψ.

Lemma 5.5. *Suppose that the function ψ is monotone. Then the conclusion of Lemma 5.4 holds if we replace the conditions* ***SmSt.1-3*** *and* ***XX.1*** *by* ***Step.1-2*** *and* ***XX.1'*** *respectively.*

Proof. As the proof is only a minor modification of the proof of Theorem 4.7.1 of Jurečková and Sen (1996), we only indicate the main steps.

Similarly to the proof of Lemma 5.4 put

$$\mathbf{E}_{nk}(\mathbf{t}, S_n) = \tfrac{1}{\sqrt{n}} \sum_{i=1}^{k} \mathbf{x}_i\, \psi\left((e_i - \tfrac{\mathbf{t}^\top \mathbf{x}_i}{\sqrt{n}})/S_n\right).$$

With the help of (4.16), it is sufficient to prove that given any $\varepsilon > 0$, there exist $M > 0$, $\eta > 0$, $\delta > 0$ and a positive integer n_0 such that for all $n > n_0$

$$\mathrm{P}^*\left\{\min_{k=n_{-\delta},\ldots,n_\delta} \inf_{\|\mathbf{t}\|\geq M} \|\mathbf{E}_{nk}(\mathbf{t}, S_n)\| < \eta\right\} < \varepsilon.$$

First, exploiting the uniform asymptotic linearity result (5.10), we can show analogously to the proof of Lemma 5.4 that there exists $M > 0$, $\delta > 0$, and $\eta > 0$ such that for all sufficiently large n it holds

$$\begin{aligned}
&\mathrm{P}^*\left\{\min_{k=n_{-\delta},\ldots,n_\delta} \inf_{\|\mathbf{t}\|=M} \|\mathbf{E}_{nk}(\mathbf{t}, S_k)\| < \eta\right\} \\
&\qquad\qquad \leq \mathrm{P}^*\left\{\max_{k=n_{-\delta},\ldots,n_\delta} \sup_{\|\mathbf{t}\|=M} \left[\mathbf{t}^\top \mathbf{E}_{nk}(\mathbf{t}, S_k)\right] > -M\eta\right\} < \varepsilon.
\end{aligned} \tag{5.13}$$

Second, we utilize that the function

$$G(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{s}^{\mathsf{T}} \mathbf{x}_i \, \psi \left((e_i - \tau \tfrac{\mathbf{s}^{\mathsf{T}} \mathbf{x}_i}{\sqrt{n}}) / S_n \right)$$

is nonincreasing for $\tau \geq 1$. This further gives us

$$\begin{aligned} \mathrm{P}^* & \left\{ \min_{k=n_{-\delta},\ldots,n_\delta} \inf_{\|\mathbf{t}\| \geq M} \|\mathbf{E}_{nk}(\mathbf{t}, S_n)\| < \eta \right\} \\ & \leq \mathrm{P}^* \left\{ \max_{k=n_{-\delta},\ldots,n_\delta} \sup_{\|\mathbf{t}\| \geq M} \left[\tfrac{M}{\|\mathbf{t}\|} \, \mathbf{t}^{\mathsf{T}} \mathbf{E}_{nk}(\mathbf{t}, S_k) \right] > -M\eta \right\} \\ & \leq \mathrm{P}^* \left\{ \max_{k=n_{-\delta},\ldots,n_\delta} \sup_{\|\mathbf{s}\| = M} \sup_{\tau \geq 1} \left[\mathbf{s}^{\mathsf{T}} \mathbf{E}_{nk}(\tau \, \mathbf{s}, S_k) \right] > -M\eta \right\} \\ & = \mathrm{P}^* \left\{ \max_{k=n_{-\delta},\ldots,n_\delta} \sup_{\|\mathbf{s}\| = M} \left[\mathbf{s}^{\mathsf{T}} \mathbf{E}_{nk}(\mathbf{s}, S_k) \right] > -M\eta \right\}. \end{aligned}$$

But we can make the last probability arbitrarily small by the first part of the proof. □

With the help of Lemma 5.4 (or Lemma 5.5) we are ready to substitute $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$ into the equation (5.10) and get the following refinement of the FOAR (4.7):

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \frac{\mathbf{V}_n^{-1}}{\gamma_1 \sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \, \psi \left(\tfrac{e_i}{S} \right) - \tfrac{\gamma_{1e}}{\gamma_1} \sqrt{n} \left(\tfrac{S_n}{S} - 1 \right) \mathbf{u}_1 + o_{ucp}(1). \tag{5.14}$$

The representation (5.14) together with the conditions **XX.1-2** immediately imply that for $l \geq 2$ the sequence $\sqrt{n}(\hat{\beta}_{ln} - \beta_l)$ is uniformly continuous in probability (ucp). Moreover, if the sequence $\sqrt{n}(\frac{S_n}{S} - 1)$ is ucp, then the sequence $\sqrt{n}(\hat{\beta}_{1n} - \beta_1)$ is ucp as well.

As the definition of the confidence interval (4.57) and (4.58) includes the estimate of σ_ψ, we need to show that $\hat{\sigma}_\psi - \sigma_\psi = o_{ucp}(1)$. In the following lemma we assume that the function ψ^2 is of bounded variation. From the simple inequality

$$|\psi^2(x) - \psi^2(y)| \leq 2 \sup_t |\psi(t)| \, |\psi(x) - \psi(y)|$$

we immediately see, that if the function ψ is of bounded variation, then the function ψ^2 is of bounded variation as well. Notice that all the most famous ψ-functions given in Remark 2 are of bounded variation.

Lemma 5.6. *Suppose that the function ψ is monotone or ψ^2 is of bounded variation and the condition **XX.1** is met. Further let the function $\lambda^{(2)}(t, u) = \mathsf{E} \, \psi^2(\frac{e_1 - t}{S e^u})$ be continuous in a neighbourhood of the point $(0, 0)$ and denote*

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

Then $\|M_n(\mathbf{t}, u) - \mathsf{E} \, \psi^2(e_1/S)\|_T \xrightarrow[n \to \infty]{a.s.} 0.$

In fact, Lemma 5.6 gives us more than we need. For our purposes it would suffice to show that $\|M_n(\mathbf{t},u) - \mathsf{E}\ \psi^2(e_1/S)\|_T = o_{ucp}(1)$. This would follow trivially by a finer analysis of the process M_n given in Lemma 5.10, but this lemma requires slightly stronger assumptions.

Proof. As ψ is monotone or ψ^2 of bounded variation, we can write $\psi^2(x) = \Phi_1(x) + \Phi_2(x)$, where $\Phi_1(x)$ is nonincreasing and $\Phi_2(x)$ is nondecreasing. Thus

$$M_n(\mathbf{t},u) = \frac{1}{n}\sum_{i=1}^{n}\Big[\Phi_1\Big(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\Big) - \Phi_1(e_i/S)\Big]$$
$$+\frac{1}{n}\sum_{i=1}^{n}\Big[\Phi_2\Big(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\Big) - \Phi_2(e_i/S)\Big] \overset{\text{Say}}{=} M_n^1(\mathbf{t},u) + M_n^2(\mathbf{t},u).$$

It will suffice to deal only with the process M_n^1 because the proof for the process M_n^2 would be completely analogous.

Let $\varepsilon > 0$ be given. By the assumptions of the lemma we can find $\delta > 0$ such that

$$\sup_{\max\{|t|,|u|\}\le\delta} \big|\,\mathsf{E}\ [\Phi_1(e^u(e_1-t)/S) - \Phi(e_1/S)]\,\big| < \tfrac{\varepsilon}{2}. \tag{5.15}$$

Let us bound

$$\big\|M_n^1(\mathbf{t},u) - \mathsf{E}\ \Phi_1(e_1/S)\big\|_T \le \Bigg\|\frac{1}{n}\sum_{i=1}^{n}\Big[\Phi_1\Big(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\Big) - \Phi_1(e_i/S)\Big]\Bigg\|_T$$
$$+\Bigg|\frac{1}{n}\sum_{i=1}^{n}[\Phi_1(e_i/S) - \mathsf{E}\ \Phi_1(e_1/S)]\Bigg| \overset{\text{Say}}{=} A_n + B_n.$$

As the term B_n converges to zero almost surely by the strong law of large numbers, it suffices to deal with A_n. As usual, denote $\varepsilon_n = \max_{1\le i\le n}\frac{M|\mathbf{x}_i|_2}{\sqrt{n}}$ and further put

$$\phi_n(t,u) = \frac{1}{n}\sum_{i=1}^{n}\Phi_1\left(\tfrac{e_i-t}{S\,e^u}\right) \qquad \text{and} \qquad \phi(t,u) = \mathsf{E}\ \Phi_1\left(\tfrac{e_1-t}{S\,e^u}\right).$$

Exploiting the monotonicity of the function Φ_1 once more, we get that for all sufficiently large n (such that $\varepsilon_n < \delta$)

$$\begin{aligned} A_n &\le \Big|\phi_n(-\varepsilon_n, -\tfrac{M}{\sqrt{n}}) - \phi_n(0,0)\Big| + \Big|\phi_n(\varepsilon_n, \tfrac{M}{\sqrt{n}}) - \phi_n(0,0)\Big| \\ &\le |\phi_n(-\delta,-\delta) - \phi_n(0,0)| + |\phi_n(\delta,\delta) - \phi_n(0,0)| \\ &\le |\phi_n(-\delta,-\delta) - \phi(-\delta,-\delta)| + |\phi(-\delta,-\delta) - \phi(0,0)| + |\phi_n(\delta,\delta) - \phi(\delta,\delta)| \\ &\quad + |\phi(\delta,\delta) - \phi(0,0)| + 2\,|\phi_n(0,0) - \phi(0,0)| \xrightarrow[n\to\infty]{\text{a.s.}} \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

where the convergence of the first, the third and the fifth term follows by the strong law of large numbers and the second and the fourth term are bounded by (5.15). Because we can take ε arbitrary small, the proof is completed. □

As the almost sure convergence is a stronger condition than ucp, then with the help of the just proved Lemma 5.6 and the fact that both sequences $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ and $\sqrt{n}(\frac{S_n}{S} - 1)$ satisfy the **SUB** condition, we immediately get

$$\hat{\sigma}_\psi^2 - \sigma_\psi^2 = \frac{1}{n}\sum_{i=1}^{n} \psi^2\big(\tfrac{Y_i - \hat{\boldsymbol{\beta}}_n^{\mathsf{T}} \mathbf{x}_i}{S_n}\big) - \sigma_\psi^2 = o_{ucp}(1), \tag{5.16}$$

which further implies $\hat{\sigma}_\psi - \sigma_\psi = o_{ucp}(1)$.

Lemma 5.7. *If the conditions of the previous lemmas are satisfied, then both sequences $\sqrt{n}(\hat{\beta}_{pn}^{-} - \beta_p)$ and $\sqrt{n}(\hat{\beta}_{pn}^{+} - \beta_p)$ satisfy the* ***SUB*** *condition.*

Proof. First, we notice that with the help of (5.8) (with c_i replaced by x_{ip}) we get that for $t = O(\frac{1}{\sqrt{n}})$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_{ip}\,\psi\big(\tfrac{r_i - t\,x_{ip}}{S_n}\big) = \gamma_1 T_{np}^2 \sqrt{n}\,t + o_{ucp}(1). \tag{5.17}$$

With the help of this equation we can calculate

$$\begin{aligned} \mathrm{P}\left\{\max_{k=n_{-\delta},\dots,n_\delta} \sqrt{k}(\hat{b}_{pk}^{-} - \hat{\beta}_{pk}) > C\right\} &= \mathrm{P}\left\{\max_{k=n_{-\delta},\dots,n_\delta} \delta_k^{-} > \tfrac{C}{\sqrt{k}}\right\} \\ = \mathrm{P}\left\{\max_{k=n_{-\delta},\dots,n_\delta} \frac{1}{\sqrt{k}}\sum_{i=1}^{k} x_{ip}\,\psi\left(\tfrac{r_i - (C/\sqrt{k})\,x_{ip}}{S_k}\right) \right. &> \left. T_{kp}^2\sqrt{\omega_{pp}(k)}\,\hat{\sigma}_\psi(k)\,z_\alpha\right\} \\ \leq \mathrm{P}\left\{\max_{k=n_{-\delta},\dots,n_\delta} -\gamma_1 C\, T_{kp}^2 + o_{ucp}(1) \right. &> \left. T_{kp}^2\sqrt{\omega_{pp}(k)}\,\hat{\sigma}_\psi(k)\,z_\alpha\right\}, \end{aligned}$$

which can be made arbitrarily small by taking C large enough. Thus, the random sequence $\sqrt{n}(\hat{b}_{pn}^{-} - \hat{\beta}_{pn})$ is **SUB**. By Lemma 5.4 (or Lemma 5.5) the sequence $\sqrt{n}(\hat{\beta}_{pn} - \beta_p)$ is **SUB** as well. Now the simple equality

$$\sqrt{n}(\hat{b}_{pn}^{-} - \beta_p) = \sqrt{n}(\hat{b}_{pn}^{-} - \hat{\beta}_{pk}) + \sqrt{n}(\hat{\beta}_{pn} - \beta_p)$$

yields that $\sqrt{n}(\hat{\beta}_{pn} - \beta_p)$ satisfies the **SUB** condition too.

Similarly we can show that the sequence $\sqrt{n}(\hat{\beta}_{pn}^{+} - \beta_p)$ satisfies **SUB** as well. $\square$

With the help of the just proved lemma we can justify the replacement of t in (5.17) by $\hat{b}_{pn}^{-}$, which, after a slight rearrangement, gives us

$$\begin{aligned} \sqrt{n}(\hat{b}_{pn}^{-} - \beta_p) &= \sqrt{n}(\hat{\beta}_{pn} - \beta_p) - \tfrac{1}{\gamma_1}\sqrt{\omega_{pp}^{n}}\,\hat{\sigma}_\psi\, z_\alpha + o_{ucp}(1) \\ &= \sqrt{n}(\hat{\beta}_{pn} - \beta_p) - \tfrac{1}{\gamma_1}\sqrt{\omega_{pp}}\,\sigma_\psi\, z_\alpha + o_{ucp}(1), \end{aligned} \tag{5.18}$$

where the last equation follows by (5.16) and the conditions **XX.2**. Finally as the sequence $\sqrt{n}(\hat{\beta}_{pn} - \beta_p)$ is ucp, we see from (5.18) that the sequence $\sqrt{n}(\hat{b}_{pn}^{-} - \beta_p)$ is ucp as well. Analogously we can prove that the sequence $\sqrt{n}(\hat{b}_{pn}^{+} - \beta_p)$ is ucp.

Now we are ready to summarize the partial results of this section.

Theorem 5.8. *Let the conditions **SmSt.1-3** and **XX.1-2** (or **Step.1-2**, **XX'.1**, and **XX.2**) be satisfied. Further suppose that the sequence $\sqrt{n}(\frac{S_n}{S}-1)$ is **SUB**. Then the sequential confidence interval $D^{II}_{N_d}$ has the asymptotic coverage $1-\alpha$ as $d \to 0_+$, that is* (5.5) *holds.*

5.2.2 Asymptotic distribution of the stopping variable N_d

Absolutely continuous ψ-function

For the simplicity of notation put $\ell_n = \hat{b}^+_{pn} - \hat{b}^-_{pn}$. Notice that by (5.3) $\sqrt{n_d} = \frac{a_F}{d}$, which gives us

$$\tfrac{a_F}{d}\left(\sqrt{\tfrac{N_d}{n_d}}-1\right) = \sqrt{n_d}\left(\tfrac{d\sqrt{N_d}}{a_F}-1\right). \tag{5.19}$$

From the following inequalities

$$\sqrt{n_d}\left(\frac{\ell_{N_d}\sqrt{N_d}}{2\,a_F}-1\right) \leq \sqrt{n_d}\left(\frac{d\,\sqrt{N_d}}{a_F}-1\right) \leq \sqrt{n_d}\left(\frac{\ell_{N_d-1}\,\sqrt{N_d}}{2\,a_F}-1\right) \tag{5.20}$$

and by the previously proved fact that $\frac{N_d}{n_d} \xrightarrow[d\to 0_+]{\mathrm{P}} 1$, we see that the problem of finding asymptotic distribution of the random variable $\frac{a_F}{d}\left(\sqrt{\frac{N_d}{n_d}}-1\right)$ is the same as the problem of finding the asymptotic distribution of $\sqrt{N_d}\left(\frac{\ell_{N_d}\sqrt{N_d}}{2\,a_F}-1\right)$. But the last quantity can be rewritten as

$$\sqrt{N_d}\left(\tfrac{\sqrt{N_d}\,\ell_{N_d}}{2a_F}-1\right) = \sqrt{N_d}\,\tfrac{\sqrt{N_d}\,\ell_{N_d}-2a_F^{N_d}}{2a_F} + \tfrac{\sqrt{N_d}(a_F^{N_d}-a_F)}{a_F} = L_{N_d} + \tfrac{\sqrt{N_d}(a_F^{N_d}-a_F)}{a_F}. \tag{5.21}$$

To take care about the second term in the last equation we will assume:

XX.3 There exists a $\Delta \in \mathbb{R}$ such that

$$\Delta = \lim_{n\to\infty} \sqrt{n}\left(\sqrt{\omega^n_{pp}} - \sqrt{\omega_{pp}}\right).$$

Remark 21. Notice that the assumption **XX.3** is tied down to the situation of a fixed design. If we consider a correlation model with random covariates, this assumption is untenable. In this case we would need to find the asymptotic distribution of $\sqrt{n}(\sqrt{\omega^n_{pp}} - \sqrt{\omega_{pp}})$ and to show that this sequence is ucp.

Let us turn our attention to L_{N_d} – the first term on the right-hand side of equation (5.21). The asymptotic behaviour of the random variable L_n was studied in Subsection 4.2.3. Theorem 4.10 states that L_n is asymptotically normal with zero mean and variance which follows by the asymptotic expansion (4.59). Thus it only remains to show that the sequence L_n is ucp. Using the above mentioned expansion (4.59), we immediately see that all we need is to show that

(*i*). the term $o_p(1)$ in (4.59) is ucp,

(ii). the sequence $\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)$ is ucp,

(iii). the sequence $\sqrt{n}(\frac{S_n}{S} - 1)$ is ucp.

In the following, we will be dealing with the items (i) and (ii). The item (iii) depends heavily on the choice of a scale estimator. The verification of this item for two simple scale estimators can be found in Appendix.

The following lemma strengthen the results of Corollary 2.6. Recall that $T = \{(\mathbf{t}, u), |\mathbf{t}|_2 \leq M, |u| \leq M\}$.

Lemma 5.9. *Let the assumptions of Corollary 2.6 be satisfied. Then the remainder term* $o_p(1)$ *in the expansion* (2.13) *is ucp.*

Proof. First, let us define

$$M_{nk}(\mathbf{t}, u) = \sum_{i=1}^{k} c_i \left[\psi\left(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\right) - \psi(e_i/S)\right]$$
$$- \tfrac{\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^{n} c_i\,\mathbf{x}_i \left[\tfrac{1}{S}\psi'(e_i/S) - \gamma_1\right] - \tfrac{u}{\sqrt{n}} \sum_{i=1}^{n} c_i \left[\tfrac{e_i}{S}\,\psi'(e_i/S) - \gamma_{1e}\right], \qquad (\mathbf{t}, u) \in T.$$

As from the proof of Theorem 2.1 we know that $\mathsf{E}\, \|M_n\|_T = o_p(1)$, we can use Corollary 7.14 to show that for every $\varepsilon > 0$ and $\eta > 0$ there exists $n_\circ$ such that for all $n > n_\circ$

$$\mathrm{P}\left\{\max_{k=1,\ldots,n} \|M_{nk} - \mathsf{E}\, M_{nk}\|_T > \varepsilon\right\} < \eta.$$

Second, by virtue of Lemma 2.5 we get that uniformly in $k = 1, \ldots, n$ and $(\mathbf{t}, u) \in T$

$$\mathsf{E}\, M_{nk}(\mathbf{t}, u) = -\tfrac{\gamma_1\,\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^{k} c_i\,\mathbf{x}_i - \tfrac{\gamma_{1e}\,u}{\sqrt{n}} \sum_{i=1}^{k} c_i$$
$$+ \tfrac{\gamma_2}{2}\,\mathbf{t}^\mathsf{T} W_k \mathbf{t} + \tfrac{(\gamma_{2e}+\gamma_1)u\,\mathbf{t}^\mathsf{T}}{n} \sum_{i=1}^{k} c_i\,\mathbf{x}_i + \tfrac{(\gamma_{2ee}+\gamma_{1e})\,u^2}{2n} \sum_{i=1}^{k} c_i + o(1). \quad (5.22)$$

Combining these facts yields the statement of the lemma. $\square$

To find the asymptotic distribution of the sequence $\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)$, we need to impose some further conditions on the function ψ and the distribution of the errors.

SmSt.4 There exists $\delta > 0$ such that $\sup_{|t|<\delta, |u|<\delta} \mathsf{E}\, \psi^4(\frac{e_1+t}{Se^u}) < \infty$.

SmSt.5 The function $\lambda^{(2)}(t, u) = \mathsf{E}\, \psi^2(\frac{e_1+t}{Se^u})$ is continuously differentiable in a neighbourhood of the point $(0, 0)$.

As the function ψ is continuous, the condition **SmSt.4** together with Lemma 7.17 imply the continuity of the function $\lambda^{(2)}(t,u) = \psi^2(\frac{e_1+t}{S\,e^v})$ in the quadratic mean at the point $(0,0)$, that is

$$\lim_{(t,u)\to(0,0)} \mathsf{E}\,\left[\psi^2(\tfrac{e_1+t}{S\,e^u}) - \psi^2(\tfrac{e_1}{S})\right]^2 = 0. \tag{5.23}$$

The condition **SmSt.5** is certainly satisfied if the functions

$$\begin{aligned} e(t,u) &= \mathsf{E}\,\psi'(\tfrac{e_1+t}{Se^u})\,\psi(\tfrac{e_1+t}{Se^u}),\\ f(t,u) &= \mathsf{E}\,e_1\,\psi'(\tfrac{e_1+t}{Se^u})\,\psi(\tfrac{e_1+t}{Se^u}) \end{aligned}$$

are bounded and continuous in a neighbourhood of the point $(0,0)$.

Remark 22. The conditions **Step.1-2** trivially imply the conditions **SmSt.4-5**.

Lemma 5.10. *Suppose that the conditions* ***SmSt.4-5*** *and* ***XX.1*** *are satisfied. Define*

$$M_{nk}(\mathbf{t},u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{k}\left[\psi^2\left(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\right) - \psi^2(e_i/S)\right], \quad \bar{M}_{nk} = M_{nk} - \mathsf{E}\,M_{nk}.$$

Then $\max_{k=1,\ldots,n}\|\bar{M}_{nk}\|_T = o_p(1)$.

Proof. The proof of this lemma is rather standard. For $i = 1,\ldots,n$ define the processes

$$Z_{ni}(\mathbf{t},u) = \tfrac{1}{\sqrt{n}}\left[\psi^2\left(e^{-n^{-1/2}u}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\right) - \psi^2(e_i/S)\right], \qquad (\mathbf{t},u)\in T.$$

Now we can proceed along the lines of the proof of Theorem 2.1 and show that this process meets the conditions of Corollary 7.14 (the continuity in the quadratic mean (5.23) of the function $\psi^2(\frac{e_1+t}{Se^u})$ is utilized here). ☐

Remark 23. Provided the condition **SmSt.4** is satisfied, it does not matter whether ψ is continuous or not.

Let us denote

$$\gamma_{01} = \lambda_t^{(2)}(0,0)\ \left(= \tfrac{2}{S}\,\mathsf{E}\,\psi\left(\tfrac{e_1}{S}\right)\,\psi'\left(\tfrac{e_1}{S}\right)\right), \qquad \gamma_{01e} = -\lambda_u^{(2)}(0,0)\ \left(= \mathsf{E}\,\tfrac{e_1}{S}\,\psi\left(\tfrac{e_1}{S}\right)\psi'\left(\tfrac{e_1}{S}\right)\right).$$

The assumption **SmSt.5** implies that all these quantities are finite. If the condition **Sym** holds, then $\gamma_{01} = 0$, but $\gamma_{01e} > 0$.

With the help of the assumption **SmSt.5** we can easily find out that uniformly in $k = 1,\ldots,n$ and uniformly in T

$$\mathsf{E}\,M_{nk}(\mathbf{t},u) = -\tfrac{2\gamma_{01}\,\mathbf{t}^\mathsf{T}}{n}\sum_{i=1}^{k}\mathbf{x}_i - 2\,\gamma_{01e}\,u + o(1).$$

Combining the last equation with Lemma 5.10 and inserting $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$ and $\sqrt{n}\log(\frac{S_n}{S})$ for u give us

$$\begin{aligned}\sqrt{n}(\hat{\sigma}_\psi^2 - \sigma_\psi^2) &= \tfrac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\psi^2(e_i/S) - \sigma_\psi^2\right] - \tfrac{2\gamma_{01}\sqrt{n}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})^{\mathsf{T}}}{n}\sum_{i=1}^{n}\mathbf{x}_i - 2\,\gamma_{01e}\,\sqrt{n}(\tfrac{S_n}{S}-1) + o_{ucp}(1)\\ &= \tfrac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\psi^2(e_i/S) - \sigma_\psi^2\right] - \tfrac{2\gamma_{01}}{\gamma_1\sqrt{n}}\sum_{i=1}^{n}\psi(e_i/S) - 2\,\gamma_{01e}\,\sqrt{n}(\tfrac{S_n}{S}-1) + o_{ucp}(1). \quad (5.24)\end{aligned}$$

We have utilized that $\sum_{i=1}^n \frac{\mathbf{x}_i}{n}$ is the first column of the matrix $\mathbf{V}_n$ and that $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ admits expansion (5.14).

The asymptotic representation of $\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)$ now follows by the 'delta-type' approximation

$$\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi) = \tfrac{\sqrt{n}}{2\,\sigma_\psi}(\hat{\sigma}_\psi^2 - \sigma_\psi^2) + o_{ucp}(1). \quad (5.25)$$

Expansion (5.24) implies that the sequence $\sqrt{n}(\hat{\sigma}_\psi - \sigma_\psi)$ is ucp if the sequence $\sqrt{n}(\frac{S_n}{S}-1)$ is ucp.

Theorem 5.11. *Let the conditions **SmSt.1-5** and **XX.1-3** be satisfied. Further suppose that the sequence $\sqrt{n}(\frac{S_n}{S}-1)$ is ucp. Then the stopping variable N_d admits the following expansion as $d \to 0_+$*

$$\begin{aligned}\tfrac{a_F}{d}\left(\sqrt{\tfrac{N_d}{n_d}} - 1\right) &= -\tfrac{1}{\gamma_1\sqrt{n_d}}\sum_{i=1}^{n_d}\tfrac{x_{ip}^2}{T_{n_d p}^2}\left[\psi'(e_i/S) - \gamma_1\right] + \tfrac{\gamma_1+\gamma_{2e}}{\gamma_1}\sqrt{n_d}\left(\tfrac{S_{n_d}}{S} - 1\right)\\ &\quad + \frac{\sqrt{n_d}(\hat{\sigma}_\psi - \sigma_\psi)}{\sigma_\psi} + \tfrac{\gamma_2}{\gamma_1}\sqrt{n_d}\,(\hat{\boldsymbol{\beta}}_{n_d} - \boldsymbol{\beta})^{\mathsf{T}}\sum_{i=1}^{n_d}\tfrac{x_{ip}^2\,\mathbf{x}_i}{n_d\,T_{n_d p}} + \tfrac{\Delta}{\sqrt{\omega_{pp}}} + o_p(1). \quad (5.26)\end{aligned}$$

Proof. The theorem follows by (4.59), (5.19), (5.22), (5.24), and (5.25). □

Numerical illustration

To illustrate the theoretical results, we performed a small numerical experiment. We considered a simple linear model with one explanatory variable $Y_i = \beta_0 + \beta_1 x_i + e_i$. We took the errors independent identically distributed following the contaminated normal distribution given by the cdf $F(x) = 0.9\,\Phi(x) + 0.1\,\Phi(x-2)$, where $\Phi(x)$ is a cdf of a standard normal random variable. We generated the explanatory variable as the random sample from the uniform distribution on the interval $(-1, 1)$.

We were interested in the behaviour of the bounded-width confidence interval for the slope (parameter β_1). We studied the actual coverage of the 95%-confidence intervals of type I and type II for (the half-length of a CI) d decreasing from 0.6 to 0.3, which corresponds to the increasing of the 'theoretical' sample size n_d from 43 to 173. The initial sample size was 20 and the number of repetitions of our experiments 20 000.

Figure 5.1 presents the results for type I CI and type II CI. The first picture shows the actual coverage of confidence intervals for different values of d and the second picture presents

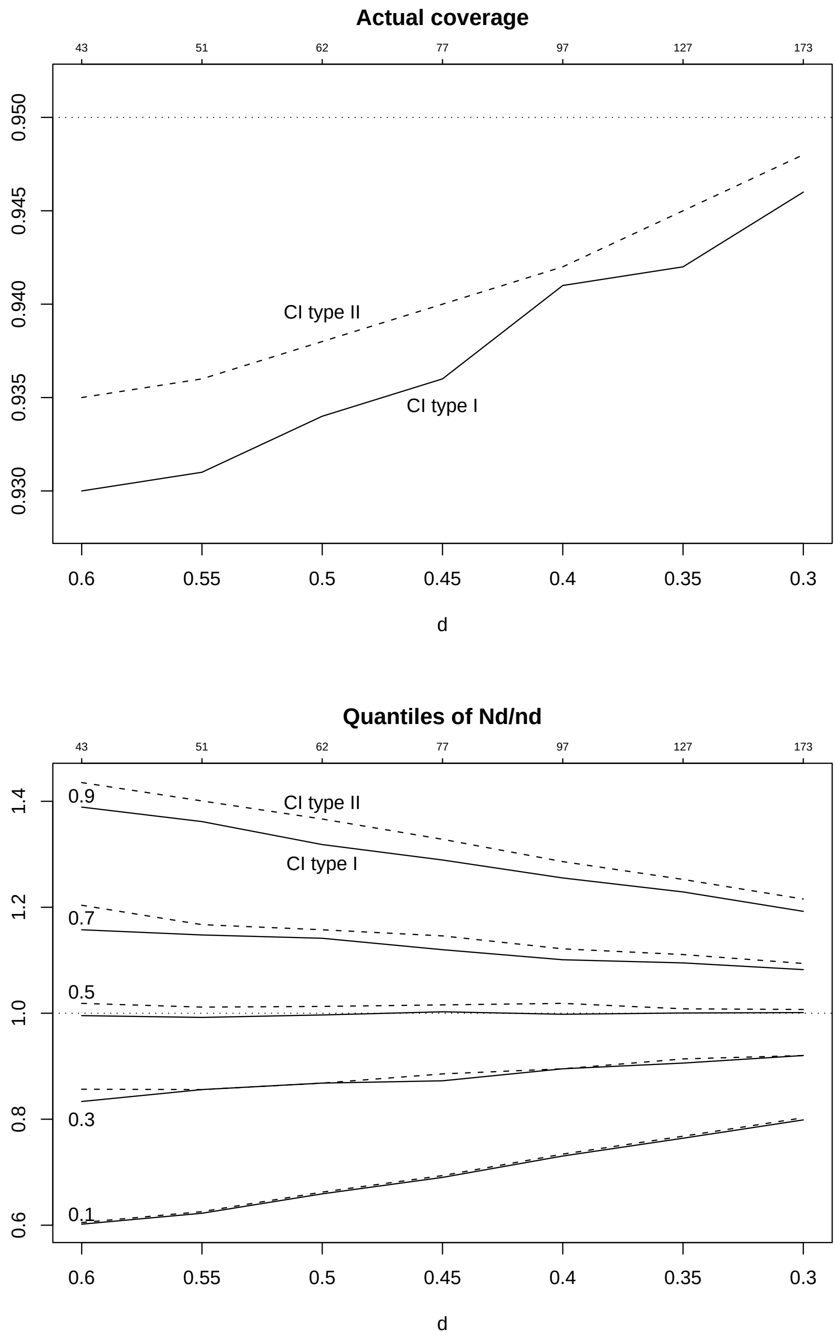

Figure 5.1: Actual coverage of CI's and the quantiles of the random variables N_d/n_d for CI of type I (solid line) and CI of type II (dashed).

the $10\%, 30\%, 50\%, 70\%, 90\%$-quantiles of the quantity N_d/n_d. From the first picture we see that both coverage probabilities converge to the prescribed nominal value 0.95 and that the actual coverage of type II CI is slightly closer to the target value. On the other hand, the second pictures indicates that we usually need more observations to stop sequential procedure for a type II CI. Moreover (in agreement with our theoretical results), the quantity N_d/n_d for type II CI is more variable than for the type I methods.

Unsmooth ψ-function

By the same trick as in (5.20) we can find out that the problem of deriving asymptotic distribution of the random variable $\sqrt{\frac{a_F}{d}}\left(\sqrt{\frac{N_d}{n_d}}-1\right)$ is the same as the problem of the asymptotic distribution of the random variable

$$L'_{N_d} = N_d^{1/4}\left(\frac{\sqrt{N_d}\,\ell_{N_d}}{2a_F}-1\right) = N_d^{1/4}\,\frac{\sqrt{N_d}\,\ell_{N_d}-2a_F}{2a_F}.$$

In the following, we will for simplicity assume that the condition **XX.3** holds, which implies $n^{1/4}(\sqrt{\omega_{pp}^n}-\sqrt{\omega_{pp}}=o(1))$. Thus

$$L'_{N_d} = N_d^{1/4}\,\frac{\sqrt{N_d}\,\ell_{N_d}-2a_F^{N_d}}{2a_F}+o(1) = L_{N_d}+o(1), \qquad \text{as } d\to 0_+.$$

The asymptotic behaviour of the random variable L_n was studied in Section 4.2.4. Theorem 4.14 states that L_n is asymptotically normal with a zero mean and a variance given by (4.70). That is why we only need to prove that L_n is ucp. As Lemma 5.10 implies

$$n^{1/4}(\hat\sigma_\psi-\sigma_\psi) = o_{ucp}(1), \tag{5.27}$$

it suffices to strengthen the result of Lemma 4.13. Before we do that, it is convenient to introduce some notation. Let us denote

$$B_k = \sum_{i=1}^{k}|x_{ip}|^3, \qquad B_n = \sum_{i=1}^{n}|x_{ip}|^3 = n\,T_{np}^3$$

and

$$n[s] = \max\{k:\ B_k \le s\,B_n\}, \qquad s\in[0,1]. \tag{5.28}$$

For simplicity we will suppose that there exists a finite and positive limit of the quantity $\frac{1}{n}\,B_n$. In the next, we will be dealing with a partial sum process ($\mathbf{b}\in B$)

$$M_n^{\mathbf{b}}(s,t) = \frac{1}{n^{1/4}}\sum_{i=1}^{n[s]} x_{ip}\left[\psi(e_i-\tfrac{\mathbf{b}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}-\tfrac{t\,x_{ip}}{\sqrt{n}})-\psi(e_i-\tfrac{\mathbf{b}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})\right], \qquad |t|\le M,\ 0\le s\le 1.$$

Put $\bar M_n^{\mathbf{b}} = M_n^{\mathbf{b}} - \mathsf{E}\,M_n^{\mathbf{b}}$.

Lemma 5.12. *If the conditions **X'.1-3** and **Step.1-2** holds, then the process $\bar M_n^{\mathbf{b}}(s,t)$ converges weakly to a zero mean gaussian process $W(s,t)$ uniformly in $\mathbf{b}\in B$. The limiting process has the covariance structure*

$$\operatorname{cov}\{W(s_1,t_1),\,W(s_2,t_2)\},=(s_1\wedge s_2)\,g(t_1,t_2),$$

where

$$g(t_1,t_2)\begin{cases} T_{np}^3\,\gamma_{01}\,|t_1|\wedge|t_2|, & t_1\cdot t_2>0,\\ 0, & t_1\cdot t_2\le 0.\end{cases}$$

Proof. Our proof will imitate the proof of Theorem 2.12.1 of van der Vaart and Wellner (1996).

We can look at the partial-sum process $\bar{M}_n^{\mathbf{b}}(s,t)$ as a process indexed by the index set $[0,1]\times T$, which we equip with the metric $\rho((s_1,t_1),(s_2,t_2)) = |s_1-s_2|+|t_1-t_2|$.

As it is easy to verify that the process $\bar{M}_n^{\mathbf{b}}(s,t)$ converges marginally, it suffices to show the asymptotic equicontinuity of the sequence $\bar{M}_n$. Thus, for every $\varepsilon>0$, $\eta>0$ there should exist $\delta>0$ and n_1 such that for all $n\ge n_1$

$$\mathrm{P}^*\left\{\sup_{|s_1-s_2|+|t_1-t_2|<\delta}|\bar{M}_n^{\mathbf{b}}(s_1,t_1)-\bar{M}_n^{\mathbf{b}}(s_2,t_2)|>\varepsilon\right\}<\eta. \tag{5.29}$$

Let us denote $T_\delta=\{(t_1,t_2):\,|t_1-t_2|<\delta,\ t_1,t_2\in T\}$. A triangular inequality yields

$$\begin{aligned}\sup_{|s_1-s_2|+|t_1-t_2|<\delta}&|\bar{M}_n^{\mathbf{b}}(s_1,t_1)-\bar{M}_n^{\mathbf{b}}(s_2,t_2)|\\ &\le \sup_{|s_1-s_2|<\delta}\|\bar{M}_n^{\mathbf{b}}(s_1,t)-\bar{M}_n^{\mathbf{b}}(s_2,t)\|_T+\sup_{0\le s\le 1}\|\bar{M}_n^{\mathbf{b}}(s,t_1)-\bar{M}_n^{\mathbf{b}}(s,t_2)\|_{T_\delta}.\end{aligned} \tag{5.30}$$

In the second term on the right-hand side the parameter s may be restricted to the points k/n with k ranging over $1,2,\dots,n$. Further by Ottaviani's inequality (Lemma 7.6) we get

$$\begin{aligned}\mathrm{P}^*\left\{\sup_{0\le s\le 1}\|\bar{M}_n^{\mathbf{b}}(s,t_1)-\bar{M}_n^{\mathbf{b}}(s,t_2)\|_{T_\delta}>\varepsilon\right\} &= \mathrm{P}^*\left\{\max_{k\le n}\|\bar{M}_{nk}^{\mathbf{b}}(t_1)-\bar{M}_{nk}^{\mathbf{b}}(t_2)\|_{T_\delta}>\varepsilon\right\}\\ &\le \frac{\mathrm{P}^*\left\{\|\bar{M}_n^{\mathbf{b}}(t_1)-\bar{M}_n^{\mathbf{b}}(t_2)\|_{T_\delta}>\varepsilon\right\}}{1-\max_{k\le n}\mathrm{P}^*\left\{\|\bar{M}_{nk}^{\mathbf{b}}(t_1)-\bar{M}_{nk}^{\mathbf{b}}(t_2)\|_{T_\delta}>\varepsilon\right\}}.\end{aligned} \tag{5.31}$$

By Lemma 4.13 the process $\bar{M}_n^{\mathbf{b}}(t)$ converges in distribution, which implies that it is asymptotically equicontinuous (see (7.1) of Appendix). Thus we can make the probability in the numerator arbitrarily small by taking δ small and n large enough.

To show that the denominator is bounded away from zero for all sufficiently large n, we can use a similar trick as in the proof of Corollary 7.14.

To prove the same for the first term on the right-hand side of (5.30), we estimate

$$\begin{aligned}\mathrm{P}^*&\left\{\sup_{|s_1-s_2|<\delta}\|\bar{M}_n^{\mathbf{b}}(s_1,t)-\bar{M}_n^{\mathbf{b}}(s_2,t)\|_T>3\,\varepsilon\right\}\\ &\le 3\,\mathrm{P}^*\left\{\max_{0\le j\,\delta\le 1}\sup_{j\,\delta\le s\le (j+1)\delta}\|\bar{M}_n^{\mathbf{b}}(s,t)-\bar{M}_n^{\mathbf{b}}(j\,\delta,t)\|_T>\varepsilon\right\}\\ &\le 3\sum_{j=1}^{\lceil 1/\delta\rceil}\mathrm{P}^*\left\{\sup_{j\,\delta\le s\le (j+1)\delta}\|\bar{M}_n^{\mathbf{b}}(s,t)-\bar{M}_n^{\mathbf{b}}(j\,\delta,t)\|_T>\varepsilon\right\}.\end{aligned} \tag{5.32}$$

With the help of Ottaviani's inequality we can bound each term in the last sum by

$$\mathrm{P}^*\left\{\sup_{j\,\delta\le s\le (j+1)\,\delta}\|\bar{M}_n^{\mathbf{b}}(s,t)-\bar{M}_n^{\mathbf{b}}(j\,\delta,t)\|_T>\varepsilon\right\}$$
$$\le\frac{\mathrm{P}^*\left\{\|\bar{M}_n^{\mathbf{b}}((j+1)\,\delta,t)-\bar{M}_n^{\mathbf{b}}(j\,\delta,t)\|_T>\varepsilon\right\}}{1-\max_{n[j\,\delta]\le k\le n[(j+1)\,\delta]}\mathrm{P}^*\left\{\|\bar{M}_{n\,n[(j+1)\,\delta]}^{\mathbf{b}}(t)-\bar{M}_{nk}^{\mathbf{b}}(t)\|_T>\varepsilon\right\}}. \tag{5.33}$$

Lemma 4.13 and the construction of $n(s)$ in (5.28) yields that the process

$$W_n(t)=\bar{M}_n^{\mathbf{b}}((j+1)\,\delta,t)-\bar{M}_n^{\mathbf{b}}(j\,\delta,t)$$

converges in distribution to the process $\sqrt{\delta}\,W$, where W is a zero mean gaussian process with the covariance function given by (4.60). By the portmanteau theorem (Theorem 7.8), the lim sup of the probability in the numerator of (5.33) is bounded by $\mathrm{P}\left\{\|W\|_T\ge\varepsilon/\sqrt{\delta}\right\}$. Since the norm $\|W\|_T$ has moments of all orders (Proposition A.2.3 of van der Vaart and Wellner (1996)), the latter probability converges to zero faster than any power of δ as $\delta\searrow 0$. We conclude that we are able to make the numerator of (5.30) smaller as δ^2 by choosing δ sufficiently small and n large enough. By an argument similar to the argument used in the investigation of the denominator in (5.31), but now also using the fact that $\mathrm{P}\left(\|W\|_T\ge\varepsilon\right)<1$, we can show that for every $\varepsilon>0$, the denominator in (5.33) remains bounded away from zero.

Thus we can make the probability in (5.32) (for sufficiently large n and small δ) arbitrarily small, which concludes the proof of the lemma for any fixed $\mathbf{b}\in B$.

A closer inspection of the proof shows that all the arguments hold uniformly in $\mathbf{b}\in B$, that is (5.29) holds uniformly in B. This enables us for every $\varepsilon>0$ and $\eta>0$ to find a finite subset S_0 of $[0,1]\times T$ and a mapping $\pi:[0,1]\times T\mapsto S_0$ such that

$$\sup_{\mathbf{b}\in B}\mathrm{P}^*\left\{\sup_{(s,t)\in[0,1]\times T}|\bar{M}_n^{\mathbf{b}}(s,t)-\bar{M}_n^{\mathbf{b}}(\pi(s,t))|>\varepsilon\right\}<\eta. \tag{5.34}$$

Now we can use the bounded Lipschitz metric (Definition 4.12) and finish the proof along the lines of Lemma 4.13. $\square$

Remark 24. For the case of studentized M-estimator we need to study the process

$$\{M_n^{\mathbf{b},u}(s,t)=\frac{1}{n^{1/4}}\sum_{i=1}^{n[s]}x_{ip}\left[\psi(e^{-u/\sqrt{n}}(e_i-\tfrac{\mathbf{b}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}-\tfrac{t\,x_{ip}}{\sqrt{n}})/S)-\psi(e^{-u/\sqrt{n}}(e_i-\tfrac{\mathbf{b}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}})/S)\right],$$
$$0\le s\le 1,\ |t|\le M\},$$

where $|\mathbf{b}|_2\le M$ and $|u|\le M$. We could proceed completely analogously to show that Lemma 5.12 holds for this 'studentized' process uniformly in u and $\mathbf{b}$ as well.

Theorem 5.13. *Let the conditions* ***XX'.1****,* ***XX.2-3****, and* ***Step.1-2*** *be satisfied. Further suppose that the sequence* $\sqrt{n}(\frac{S_n}{S}-1)$ *satisfies the* ***SUB*** *condition. Then for* $d \to 0_+$ *the random variable* $\sqrt{\frac{a_F}{d}}\left(\sqrt{\frac{N_d}{n_d}}-1\right)$ *is asymptotically normal with the zero mean and the variance given by*

$$\sigma^2 = \frac{\gamma_{01}\,\kappa}{2\,\gamma_1^2\,a_F}, \qquad \textit{where} \quad \kappa = \lim_{n\to\infty} \frac{T_{np}^3}{(T_{np}^2)^2}. \tag{5.35}$$

Proof. The uniform continuity in probability of L_n follows from (5.27), Lemma 5.12 (and its generalization to the studentized processes as indicated in Remark 24), and the assumption that $\sqrt{n}(\frac{S_n}{S}-1)$ satisfies the **SUB** condition.

Once we have proved that L_n is ucp, Theorem 4.14 gives us the asymptotic distribution of L_{N_d}. $\square$

5.3 Bounded length confidence interval for R-estimators based on Wilcoxon scores

In this section we discuss the asymptotic properties of a bounded length confidence interval based on an R-estimator generated with Wilcoxon scores. As the treatment of this situation will be very similar to the case of a smooth ψ-function, we will not go into great details.

Recall that the confidence interval (of type II) D_n^{II} is defined by (4.71). The theoretical counterpart of the random stopping time N_d defined in (5.2) is

$$n_d = \frac{z_\alpha^2\,\omega_{pp}}{12\,\gamma^2\,d^2} = \frac{a_F^2}{d^2}, \quad \text{where} \quad a_F = \frac{z_\alpha\,\sqrt{\omega_{pp}}}{\gamma\,\sqrt{12}}. \tag{5.36}$$

With the help of Theorem 4.15 we can follow the proof of Theorem 5.1 to show that under the conditions **W.1** and **XX.2** the statements of Theorem 5.1 hold for N_d as well.

5.3.1 Asymptotic coverage

Similarly to Section 5.2.1, we need to strengthen some first order results to prove that the confidence interval $D_{N_d}^{II}$ is asymptotically correct for $d \to 0_+$.

Put $\tilde{\mathbf{S}}_n(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i R_i'(\mathbf{t})$, where $R_i'(\mathbf{t})$ is the rank of the random variable $e_i - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}$ among $e_1 - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_1}{\sqrt{n}}, \dots, e_n - \frac{\mathbf{t}^\mathsf{T}\mathbf{x}_n}{\sqrt{n}}$. We would like to show that

$$\left\| \tfrac{1}{\sqrt{n}}\,\tilde{\mathbf{S}}_n(\mathbf{t}) - \tfrac{1}{\sqrt{n}}\,\tilde{\mathbf{S}}_n(\mathbf{0}) + \gamma \mathbf{V}_n \mathbf{t} \right\|_T = o_{ucp}(1). \tag{5.37}$$

To prevent some minor measurability difficulties, we can argue similarly as in Subsection 3.2.2, that instead of $T = \{\mathbf{t} \in \mathbb{R}_p : |\mathbf{t}|_2 \le M\}$ we can take $T = \{\mathbf{t} \in \mathbb{Q}_p : |\mathbf{t}|_2 \le M+1\}$, which makes the supremum on the left-hand side of the equation (5.37) a measurable random variable.

With the help of the linearity result (5.37) we will be able to proceed similarly as in Lemma 5.5 to show that $\sqrt{n}(\hat{\mathbf{b}}_n - \boldsymbol{\beta})$ satisfies the **SUB** condition. This will enable us to substitute $\sqrt{n}(\hat{\mathbf{b}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$ in (5.37) and get that the remainder term $o_p(1)$ in (4.22) of the asymptotic expansion for $\sqrt{n}(\hat{\mathbf{b}}_n - \boldsymbol{\beta})$ is $o_{ucp}(1)$ in fact. Combining this result with the uniform asymptotic linearity (5.37) immediately yields that the random variables $\sqrt{n}(\hat{b}^+_{np} - \beta_p)$ and $\sqrt{n}(\hat{b}^-_{np} - \beta_p)$ are both ucp, which would imply the asymptotic correctness of $D^{II}_{N_d}$.

As it is sufficient to show that the statement (5.37) holds for every single component, it is enough to deal with one-dimensional processes. In the following, we will make use of the notations introduced in Chapter 3 (we please the reader to recall the definitions of the processes $\tilde{S}_n$, T_n and $\bar{T}_n$ of (3.1)–(3.3)).

To show that the statement (5.37) holds, it is useful to decompose the process $\bar{T}_n$ into its projection and a remainder term, that is $\bar{T}_n = P_n + R_n$.

First, we will be dealing with the remainder term R_n. From Section 3.2.2 we know that $R_n = U_n - \mathsf{E}\, U_n$ with the process U_n defined by (3.9). Let us define $R_{nk} = U_{nk} - \mathsf{E}\, U_{nk}$, where

$$U_{nk}(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{k} c_i \sum_{j=1}^{k} \Big[\mathbb{I}\{e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}} \geq e_j - \tfrac{\mathbf{t}^\mathsf{T} x_j}{\sqrt{n}}\} - \mathbb{I}\{e_i \geq e_j\}\Big]$$
$$- \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k} \Big[(c_i - c_j)(F(e_i - \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}) - F(e_i))\Big]. \quad (5.38)$$

Lemma 5.14. *Under the conditions **X.1-5** and **W.1-3***

$$\max_{k\leq n} \|R_{nk}\|_T = o_p(1).$$

Proof. From the definition of the process R_n we see that for all $i = 1, \ldots, n$ and all $\mathbf{t} \in T$ it holds $\mathsf{E}\,[R_{nk}(\mathbf{t})|e_i] = 0$. Let $\mathcal{B}_k$ be the σ-field generated by $\{e_i, i \leq k\}$ and $\mathcal{B}_0$ be the trivial σ-field. Then, for every $n \geq 1$,

$$\{R_{nk}(\mathbf{t}),\ \mathbf{t} \in T,\ \mathcal{B}_k,\ k \leq n\}$$

is a martingale (process), which implies that $\{\|R_{nk}\|_T, \mathcal{B}_k, k \leq n\}$ is a nonnegative submartingale. Therefore, by Kolmogorov's inequality for submartingales (see Lemma 7.2), for every $\varepsilon > 0$

$$\mathrm{P}\left\{\max_{k\leq n} \|R_{nk}\|_T > \varepsilon\right\} \leq \tfrac{1}{\varepsilon}\,\mathsf{E}\,\|R_n\|_T.$$

But from (3.16) (Section 3.2.2) we know $\mathsf{E}\,\|R_n\|_T = o(1)$, which proves the lemma. □

Let us turn our attention to the leading term P_n. Recall that $P_n = V_n - \mathsf{E}\, V_n$, where V_n is defined in (3.6) (cf. Section 3.2.1). Similarly to the case of U_n put

$$V_{nk}(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k}(c_i - c_j)\Big[F(e_i - \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}) - F(e_i)\Big].$$

In the following we will need this technical lemma.

Lemma 5.15. *Let the conditions* ***X.1-5*** *and* ***W.1-3*** *be satisfied. Then for every* $\varepsilon > 0$ *and* $\eta > 0$ *there exist* $\delta > 0$ *and* n_0 *such that for every* $n > n_0$

$$\mathrm{P}\left\{\max_{n_{-\delta}\le k\le n_\delta}\left\|\tfrac{1}{\sqrt{n}}\,\bar V_{nk}-\tfrac{1}{\sqrt{n}}\,\bar V_n\right\|_T>\varepsilon\right\}<\eta,$$

where $n_{-\delta}=\lfloor n(1-\delta)\rfloor$ *and* $n_\delta=\lceil n(1+\delta)\rceil$.

Proof. Obviously it is sufficient to prove

$$\mathrm{P}\left\{\max_{n_{-\delta}\le k\le n}\left\|\tfrac{1}{\sqrt{n}}\,\bar V_{nk}-\tfrac{1}{\sqrt{n}}\,\bar V_n\right\|_T>\varepsilon\right\}<\eta.$$

For the simplicity of notation define

$$W_{ij}(\mathbf{t})=\frac{c_i-c_j}{n^{3/2}}\left[F(e_i-\tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}})-F(e_i)\right].$$

Note that $V_{nn}-V_{nk}=A_{nk}+B_{nk}$, where

$$A_{nk}=\sum_{i=k+1}^{n}\sum_{j=1}^{n}W_{ij},\qquad\text{and}\qquad B_{nk}=\sum_{i=1}^{k}\sum_{j=k+1}^{n}W_{ij}.$$

We can exploit the martingale structure of the process $\bar A_{nk}$ and with the help of Kolmogorov's inequality (Lemma 7.2) get

$$\mathrm{P}\left\{\max_{k\le n}\|\bar A_{nk}\|_T>\varepsilon\right\}\le\frac{1}{\varepsilon}\,\mathsf{E}\left\|\sum_{i=1}^{n}\sum_{j=1}^{n}\bar W_{ij}\right\|_T=\frac{1}{\varepsilon}\,\mathsf{E}\,\|\tfrac{1}{\sqrt{n}}\,\bar V_n\|_T.$$

But it is easy to verify the assumptions of Corollary 7.13 with $Z_{ni}=\sum_{j=1}^{n}\bar W_{ij}$, which yields $\frac{1}{\sqrt{n}}\,\mathsf{E}\,\|\bar V_n\|_T\to 0$.

Now turn our attention to the process B_{nk}. For $i=1,\dots,n$ and $j=1,\dots,n$ put

$$l_{ij}=\inf_{\mathbf{w}\in T}\{\tfrac{\mathbf{w}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}}\},\qquad u_{ij}=\sup_{\mathbf{w}\in T}\{\tfrac{\mathbf{w}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}}\}.$$

Then we can bound each term W_{ij} by

$$\|W_{ij}\|_T\le\frac{|c_i|+|c_j|}{n^{3/2}}\left[F(e_i+u_{ij})-F(e_i+l_{ij})\right]\overset{\text{Say}}{=}W_{ij}^{\circ},$$

which further gives us

$$\max_{n_{-\delta}\le k\le n}\|B_{nk}\|_T\le\sum_{i=1}^{n}\sum_{j=n_{-\delta}}^{n}W_{ij}^{\circ}\overset{\text{Say}}{=}W_n^{\circ}.$$

Notice that for sufficiently large n with the help of our assumptions

$$\begin{aligned}\mathsf{E}\,W_n^{\circ}&\le\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j=n_{-\delta}}^{n}(|c_i|+|c_j|)\,\mathsf{E}\,\left[F(e_i+u_{ij})-F(e_i+l_{ij})\right]\\&\le\frac{2\,CM}{n^2}\sum_{i=1}^{n}\sum_{j=n_{-\delta}}^{n}(|c_i|+|c_j|)\,(|\mathbf{x}_i|_2+|\mathbf{x}_j|_2)\le C'\delta,\end{aligned}$$

where C and C' are sufficiently large constants independent of n. Thus by taking δ small enough, we can make $\mathsf{E}\, W_n^\circ$ arbitrarily small. Now the statement of the lemma follows by the inequalities

$$\begin{aligned} \mathrm{P}\left\{\max_{n_{-\delta}\le k\le n}\left\|\frac{1}{\sqrt{n}}\,\bar{V}_{nk}-\frac{1}{\sqrt{n}}\,\bar{V}_n\right\|_T>\varepsilon\right\} \\ \le \mathrm{P}\left\{\max_{n_{-\delta}\le k\le n}\|\bar{A}_{nk}\|_T>\tfrac{\varepsilon}{2}\right\}+\mathrm{P}\left\{\max_{n_{-\delta}\le k\le n}\|\bar{B}_{nk}\|_T>\tfrac{\varepsilon}{2}\right\} \\ \le \frac{2}{\varepsilon}\,\mathsf{E}\,\|\tfrac{1}{\sqrt{n}}\,\bar{V}_n\|_T+\frac{4}{\varepsilon}\,\mathsf{E}\,\|W_n^\circ\|_T. \end{aligned}$$

□

Now, by a standard computation it is possible to show that uniformly in $k = n_{-\delta}, \ldots, n_\delta$

$$\mathsf{E}\,\tfrac{1}{\sqrt{n}}\,V_{nk}=\tfrac{2\gamma\,k\,\mathbf{t}^\mathsf{T}}{n^2}\sum_{i=1}^{k}(c_i-\bar{c}_k)\,\mathbf{x}_i+o(1). \tag{5.39}$$

Combining the last equation (5.39) together with the Lemma 5.14 and Lemma 5.15 gives us the strengthened uniform linearity result (5.37).

Theorem 5.16. *Let the conditions* ***W.1-2*** *and* ***XX.1-2*** *be satisfied. Then the sequential confidence interval* $D^{II}_{N_d}$ *has the asymptotic coverage* $1-\alpha$ *as* $d\to 0_+$*, that is* (5.5) *holds.*

5.3.2 Asymptotic distribution of the stopping variable N_d

We can use the same trick as in (5.20) to show that $\frac{a_F}{d}\left(\sqrt{\frac{N_d}{n_d}}-1\right)$ is asymptotically equivalent (as $d\to 0_+$) with the random variable

$$L'_{N_d}=\sqrt{N_d}\big(\tfrac{\sqrt{N_d}\,\ell_{N_d}}{2a_F}-1\big)=L_{N_d}+\tfrac{\sqrt{N_d}(a_F^{N_d}-a_F)}{a_F}.$$

As Theorem 4.15 gives us the asymptotic expansion for the random variable L_n (4.73), all we need is to prove that the term $o_p(1)$ in that expansion is $o_{ucp}(1)$ in fact.

Unfortunately, we have not succeeded to proceed without the following assumption:

W.5 There exist $\delta > 0$ and $C < \infty$ such that for all $|t| \le \delta$

$$\mathsf{E}\,|f(e_1-t)-f(e_1)|\le Ct.$$

Remark 25. The condition **W.5** is certainly satisfied if f has a bounded derivative.

If we recall the previous sections of this chapter, we see that it suffices to show that $\left\|\sum_{i=1}^n \bar{Z}_{ni}\right\|_T = o_{ucp}(1)$, where

$$Z_{ni}(\mathbf{t})=\frac{1}{n}\sum_{j=1}^{n}(c_i-c_j)\left[F(e_i-\tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}})-F(e_i)+\tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}}\,f(e_i)\right].$$

Similarly to the previous section define

$$V_{nk}(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k}(c_i - c_j)\left[F(e_i - \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}}) - F(e_i) + \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}}\, f(e_i)\right].$$

Lemma 5.17. *Let the conditions* ***X.1-5***, ***W.1-3*** *and* ***W.5*** *be satisfied. Then for every* $\varepsilon > 0$ *there exist* $\delta > 0$ *and* n_0 *such that for every* $n > n_0$

$$\mathrm{P}\left\{\max_{n_{-\delta}\le k\le n_\delta}\left\|\bar{V}_{nk} - \bar{V}_n\right\|_T > \varepsilon\right\} < \eta,$$

where $n_{-\delta} = \lfloor n(1-\delta)\rfloor$ *and* $n_\delta = \lceil n(1+\delta)\rceil$.

Proof. The proof is completely analogous to the proof of Lemma 5.15. First, we define

$$W_{ij}(\mathbf{t}) = \frac{c_i - c_j}{n}\left[F(e_i - \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}}) - F(e_i) + \tfrac{\mathbf{t}^\mathsf{T}(\mathbf{x}_i-\mathbf{x}_j)}{\sqrt{n}} f(e_i)\right].$$

and note that $V_{nn} - V_{nk} = A_{nk} + B_{nk}$, where

$$A_{nk} = \sum_{i=k+1}^{n}\sum_{j=1}^{n} W_{ij} \qquad \text{and} \qquad B_{nk} = \sum_{i=1}^{k}\sum_{j=k+1}^{n} W_{ij}.$$

As from the proof of Theorem 3.1 we know that $\mathsf{E}\,\|\bar{V}_n\|_T \xrightarrow[n\to\infty]{} 0$, we can use Kolmogorov's inequality for submartingales (Lemma 7.2) and get

$$\mathrm{P}\left\{\max_{k\le n}\|\bar{A}_{nk}\|_T > \varepsilon\right\} \le \frac{1}{\varepsilon}\,\mathsf{E}\left\|\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{W}_{ij}\right\|_T = \frac{1}{\varepsilon}\,\mathsf{E}\,\|\bar{V}_n\|_T \xrightarrow[n\to\infty]{} 0.$$

Now turn our attention to the process B_{nk}. For sufficiently large n we can bound

$$\mathsf{E}\,\|W_{ij}\|_T \le \frac{|c_i| + |c_j|}{n}\int_{-\frac{(M+1)(|\mathbf{x}_i|_2+|\mathbf{x}_j|_2)}{\sqrt{n}}}^{\frac{(M+1)(|\mathbf{x}_i|_2+|\mathbf{x}_j|_2)}{\sqrt{n}}} \mathsf{E}\,|f(e_i - v) - f(e_i)|\,dv$$

$$\overset{\textbf{W.5}}{\le} C(M+1)^2\,\frac{(|c_i| + |c_j|)(|\mathbf{x}_i|_2^2 + |\mathbf{x}_j|_2^2)}{n^2}.$$

Thus there exists $C' > 0$ such that

$$\mathsf{E}\max_{n_{-\delta}\le k\le n_\delta}\|B_{nk}\|_T \le \sum_{i=1}^{n}\sum_{j=n_{-\delta}}^{n}\mathsf{E}\,\|W_{ij}\|_T \le C'\delta.$$

Now we can finish the proof by the same argument as the proof of Lemma 5.15. ☐

We are ready to formulate the theorem about the asymptotic behaviour of N_d as $d \to 0_+$.

Theorem 5.18. *Let the conditions* ***W.1-3***, ***W.5***, *and* ***XX.1-3*** *be satisfied. Then the stopping variable* N_d *admits the following expansion as* $d \to 0_+$

$$\tfrac{a_F}{d}\left(\sqrt{\tfrac{N_d}{n_d}} - 1\right) = -\frac{1}{\gamma\,\sqrt{n_d}\,T^2_{n_d p}}\sum_{i=1}^{n_d}\left[x_{ip}^2 + \sum_{j=1}^{n_d}\tfrac{x_{jp}^2}{n_d}\right][f(e_i) - \gamma] + \tfrac{\Delta}{\omega_{pp}} + o_p(1).$$

Numerical illustration

To illustrate the theoretical results we performed a small numerical experiment. We considered a linear model with two explanatory variables, that is $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$. We took the errors to be independent identically distributed following the logistic distribution given by the cdf $F(x) = \frac{e^{-x}}{1+e^{-x}}$. The coordinates of the explanatory variable were independent and followed uniform distribution on the interval $(-1, 1)$.

We were interested in the behaviour of the bounded-width confidence interval for the slope parameter β_1, which is based on an R-estimator generated by Wilcoxon scores. We studied the actual coverage of the 95%-confidence intervals of type I and type II for (the half-length of a CI) d decreasing from 0.45 to 0.225, which corresponds to the 'theoretical' sample size n_d increasing from 43 to 171. We estimated the functional γ, needed to construct type I CI, by the same procedure as was used in Numerical illustration of Section 4.1.3. The number of repetitions of our experiments was 20 000.

The results of our small study are to be found in Figure 5.2. The first picture shows the actual coverage of confidence interval for different values of d and the second picture presents the $10\%, 30\%, 50\%, 70\%, 90\%,$ quantiles of the quantity N_d/n_d. From the first picture we see that while the coverage probability of type I CI converges to the prescribed nominal value 0.95 from below (starting at 0.938), the actual coverage of type II CI oscillates around the target value. On the other hand the second pictures clearly indicates that we usually need more observations to stop the sequential procedure based on the type II CI.

We conclude that the actual coverage of type II CI is closer to its nominal value at the cost of a larger number of observations needed. Some further numerical simulations show that the above results are quite typical.

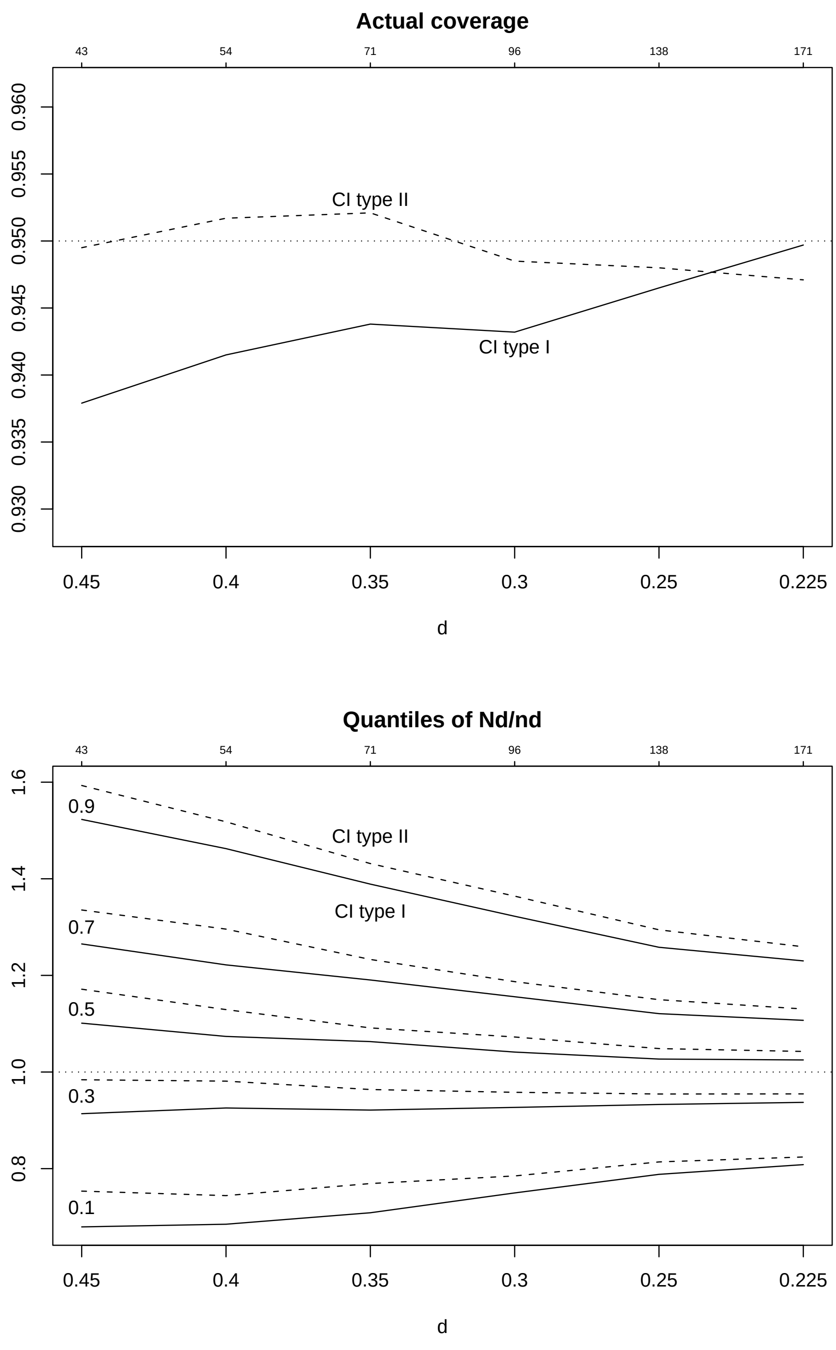

Figure 5.2: Actual coverage of CI's and the quantiles of the random variables N_d/n_d for CI of type I (solid line) and CI of type II (dashed).

Chapter 6

Conclusion

This chapter contains a short discussion of the results obtained in the thesis as well as some possible suggestions for future work.

M-estimators

M-estimators generated by an absolutely continuous ψ

We derived the two-term von Mises expansion for regression M-estimators. This enables us, e.g. to make a finer (second order) comparison of M-estimator with its one-step approximation. The possible further development would be to show that the remainder term R_n in the von Mises expansion is not only of order $o_p(\frac{1}{\sqrt{n}})$, but $\mathsf{E}\,\sqrt{n}R_n = o(1)$ or even $\mathsf{E}\,[\sqrt{n}R_n]^2 = o(1)$. This would justify the heuristic results of Huber (1973) about the approximation of finite-sample bias and variance of the regression M-estimators.

If ψ' and the distribution of the errors are sufficiently smooth, then with the machinery introduced in this thesis (together with some algebraic manipulation software) we can derive von Mises-expansion of the third or even higher order. But our numerical illustrations indicate that even the second order asymptotics needs a very large number of observations to kick in. That is why we do not see any practical importance for this higher order analysis.

M-estimators generated by a step function ψ

We derived the asymptotic distribution of the remainder term in the first order expansion. Unfortunately, as to the best of our knowledge the $\sqrt{n}$-consistency of these estimators have been generally treated only for a monotone ψ, our results are limited only to this special type of ψ functions.

R-estimators

We studied 'rank'-processes for a special case of Wilcoxon scores. This enabled us to derive a two-term von Mises expansion for R-estimators based on Wilcoxon scores. It would be interesting to show analogous results for the scores generated by a different (possibly unbounded) score function. But this would require to use some other techniques, as our approach applies only to a generalized U-statistics with a kernel of degree two. Nevertheless, we believe that our technique would be useful to prove some auxiliary results.

Our numerical illustrations indicates that the second order asymptotics 'works even slower' than in the case of M-estimators. This is probably explained by the fact, that R-estimators are based on ranks, so that we cannot hope them to be very smooth functionals.

Alternative confidence intervals

We proposed an alternative to a traditional Wald-type procedure to construct a confidence interval for a single regression parameter. We used the technical tools derived earlier in the thesis to explore asymptotic properties of the alternative method and compare it with the standard approach. We derived that these two methods are asymptotically equivalent in the sense of the asymptotic coverage and the length of the confidence interval (multiplied by $\sqrt{n}$). A finer analysis showed that Wald-type confidence intervals are more stable in the sense, that their lengths (properly standardized) have smaller variances. The degree of the difference in the variances of lengths depends on the ratio of the fourth to the second moment of the corresponding explanatory variable. This was confirmed with several numerical experiments, some of which are to be found in this thesis and some in Omelka (2006).

Our experiments show that in the case of M-estimators with a smooth ψ, our proposed confidence intervals have actual coverage larger than nominal coverage and they are on average longer than its standard competitors. On the other hand, the proposed confidence intervals usually work better in the presence of strong asymmetry, heteroscedasticity or if we are interested only in one-sided confidence intervals.

It is interesting that for M-estimators which are generated by a step function ψ the proposed confidence intervals are undersized, which is in opposite to the case of a smooth ψ. Our experience is that this lack of coverage is reasonably small if there is only one explanatory variable (plus an intercept).

Finally, if ψ is a sum of a smooth and a step function, our experiments indicate that unless the sample size is very large, the proposed procedure may be of interest, as it does not require density estimation. Unfortunately, as it has been said before, we miss $\sqrt{n}$-consistency results for such M-estimators.

The results for the R-estimator based on the Wilcoxon scores are similar to the results for the M-estimators based on a smooth ψ. The coverage of the proposed procedure is slightly oversized and the confidence intervals are on average longer than their standard competitors.

On the other hand, if we are dealing with a bounded length confidence interval problem, then unless the prescribed length of the confidence interval is very small, the coverage of our proposed procedure is usually closer to the nominal value. But this is at the cost of a larger number of observations, which the procedure usually requires.

Although it is not very common to criticize own proposed procedures, it is fair to end this short discussion with a few cautionary notes. Our experience showed that the performance of our suggested confidence interval depends heavily on a reasonable behaviour of the corresponding explanatory variable. That is why applying our confidence limits cannot be done automatically and the amount of care should be higher than for the Wald-type procedure. One of the strategies may be as follows. Compute both type of confidence intervals and if the results are comparable, we can choose our proposal. If there is a huge difference, then something wrong is happening and man should look carefully into data. But to justify this strategy, a larger simulation study is required.

Chapter 7

Appendix

7.1 Probability Inequalities

Lemma 7.1. *(Markov's inequality) For $a > 0$*

$$P(|X| \geq a) \leq \frac{\mathsf{E}\,|X|}{a}.$$

Lemma 7.2. *(Kolmogorov's inequality) Let $\{S_k,\, k = 1, \ldots, n\}$ be a nonnegative submartingale. Then for every $\varepsilon > 0$*

$$P\left\{\max_{k \leq n} S_k > \varepsilon\right\} \leq \tfrac{1}{\varepsilon}\,\mathsf{E}\,S_n.$$

7.2 Limit Theorems

We say that a sequence of random variables $\{X_n\}$ is asymptotically normal with 'mean' μ_n and 'variance' σ_n^2 if $\sigma_n > 0$ for all n sufficiently large and $\frac{X_n - \mu_n}{\sigma_n}$ converges in distribution to $N(0, 1)$. We write 'X_n is $AN(\mu_n, \sigma_n^2)$.'

The following theorem is sometimes called Delta-Theorem.

Theorem 7.3. *Suppose that X_n is $AN(\mu, \sigma_n^2)$, with $\sigma_n \to 0$. Let g be a real-valued function differentiable at $x = \mu$, with $g'(\mu) \neq 0$. Then $g(X_n)$ is $AN\left(g(\mu), [g'(\mu)]^2\sigma_n^2\right)$.*

Theorem 7.4. *Consider a triangular array (of row-wise independent r.v.'s) X_{nj}, $j \leq k_n$, $n \geq 1$, where $k_n \to \infty$ as $n \to \infty$. Let*

$$\mathsf{E}\,X_{nk} = 0, \qquad \operatorname{var}\{X_{nk}\} = \sigma_{nk}^2, \qquad k = 1, \ldots, k_n,\ n \geq 1.$$

Then $Z_n = \sum_{j=1}^{k_n} X_{nj}$ is $AN(0, 1)$ provided the Feller-Lindeberg condition holds, that is for any $\varepsilon > 0$

$$\frac{1}{\sigma_n^2}\sum_{k=1}^{k_n} \mathsf{E}\,X_{nk}^2\,\mathbb{I}\{|X_{nk}| > \varepsilon\,\sigma_n\} \xrightarrow[n\to\infty]{} 0,$$

where $\sigma_n^2 = \sum_{k=1}^{k_n} \sigma_{nk}^2$.

Theorem 7.5. *(Cramér-Wold device) Let* $\mathbf{T}_n = (T_{n1}, \ldots, T_{np})^\mathsf{T}$ *be a sequence of random vectors. Then* $\mathbf{T_n} \xrightarrow{\mathcal{D}} \mathbf{T}$ *if and only if for every fixed* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^\mathsf{T} \in \mathbb{R}_p$*,* $\boldsymbol{\lambda}^\mathsf{T}\mathbf{T}_n \xrightarrow{\mathcal{D}} \boldsymbol{\lambda}^\mathsf{T}\mathbf{T}$*.*

7.3 Outer probability, outer expectation, and star weak convergence

By the word *random element* we will mean an arbitrary map from the probability space to a metric space D. To be able to introduce the concept of weak convergence of random elements, we need to give some terminology (for more details see, e.g. the first part of the book van der Vaart and Wellner (1996)).

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be an arbitrary probability space and $T : \Omega \mapsto \mathbb{R}^* = [-\infty, +\infty]$ be an arbitrary map from this space to the extended real line. The *outer integral* of T with respect to P is defined as

$$\mathsf{E}^* T = \inf\left\{\mathsf{E}\, U : U \geq T, U \text{ is measurable and } \mathsf{E}\, U \text{ exists}\right\}.$$

The *outer probability* of an arbitrary subset B of Ω is

$$\mathrm{P}^*(B) = \inf\{\mathrm{P}(A) : B \subset A, A \in \mathcal{A}\}.$$

Similarly the *inner probability* is defined as

$$\mathrm{P}_*(B) = \sup\{\mathrm{P}(A) : A \subset B, A \in \mathcal{A}\}.$$

It turns out that these definitions are very natural. E.g. it can be easily shown that the inequalities of Section 7.1 hold even for random elements if we replace P and E with its 'outer versions' P^* and E^*. Before we define the (star) weak convergence, we give one very useful inequality.

Lemma 7.6. *(Ottaviani's inequality) Let* $X_1, \ldots, X_n$ *be independent stochastic processes indexed by an arbitrary set and* $S_k = X_1 + \ldots + X_k$ *the partial sum of these processes. Then for every* $\lambda, \mu > 0$

$$\mathrm{P}^*\left\{\max_{k \leq n} \|S_k\| > \lambda + \mu\right\} \leq \frac{\mathrm{P}^*\{\|S_n\| > \lambda\}}{1 - \max_{k \leq n} \mathrm{P}^*\{\|S_n - S_k\| > \mu\}}.$$

Definition 7.7. Let $\{X_n, n \in \mathbb{N}\}$ be arbitrary maps from the probability space $(\Omega, \mathcal{A}, P)$ into a metric space D and $X : (\Omega, \mathcal{A}, P) \mapsto D$ be Borel measurable. We say that the sequence X_n *converges weakly* to X if

$$\mathsf{E}^* f(X_n) \xrightarrow[n \to \infty]{} \mathsf{E}\, f(X), \text{ for every bounded and continuous function } f \text{ on } D.$$

Notice that if X_n are measurable, then Definition 7.7 coincides with the standard definition of weak convergence.

The concept of the weak convergence in the sense of Definition 7.7 proved to be very fruitful. We can reformulate many of the 'classic theorems' by using outer expectations and outer probabilities.

In the proof of Lemma 5.12 we make use of the following version of Portmanteau theorem.

Theorem 7.8. *Let $X_1, X_2, \ldots$ be a sequence of random elements and X be a random variable. Then the following statements are equivalent:*

(i). X_n converges weakly to X;

(ii). $\liminf_{n\to\infty} \mathrm{P}_*(X_n \in G) \geq \mathrm{P}(X \in G)$ for every open G;

(iii). $\limsup_{n\to\infty} \mathrm{P}^*(X_n \in F) \leq \mathrm{P}(X \in F)$ for every closed F.

The role of tightness in the 'classic theory' of weak convergence is played by asymptotic tightness, which is defined as follows.

Definition 7.9. We say that the sequence of random elements $\{X_n, n \in \mathbb{N}\}$ on a metric space (D, d) is *asymptotically tight* if for every $\varepsilon > 0$ there exists a compact set K such that

$$\liminf_{n\to\infty} \mathrm{P}_*(X_n \in K^\delta) \geq 1 - \varepsilon, \qquad \text{for every } \delta > 0,$$

where $K^\delta = \{y \in D : d(y, K) < \delta\}$.

It can be shown that for measurable random variables on separable and complete metric spaces the concept of asymptotic tightness and tightness coincide.

Theorem 7.10. *Let $\{X_n, n \in \mathbb{N}\}$ be random elements on $\ell^\infty(T)$. Then X_n converges weakly to a tight limit X if and only if the sequence $\{X_n\}$ is asymptotically tight and the marginals $(X_n(t_1), \ldots, X_n(t_k))$ converge weakly to the marginals $(X(t_1), \ldots, X(t_k))$ of X.*

The weak convergence implies the paths of the sequence of processes to be *asympotically uniformly ρ-continuous in probability.* More precisely, let ρ be a semimetric such that the space (T, ρ) is totally bounded and X_n converges weakly to X in $\ell^\infty(T)$ (the space of bounded functions). Then for every $\varepsilon, \eta > 0$, there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathrm{P}^*\left(\sup_{\rho(s,t)<\delta} |X_n(s) - X_n(t)| > \varepsilon\right) < \eta. \tag{7.1}$$

7.4 Clippings from empirical process theory

The following Lemma can be found in Nolan and Pollard (1987). We only change the notation so that it is consistent with the notation used in van der Vaart and Wellner (1996).

Let $(\mathcal{F}, \rho)$ be an index class equipped with the pseudometric ρ. Then the covering number $N(\varepsilon, \mathcal{F}, \rho)$ is the minimal number of balls of radius ε needed to cover the set $\mathcal{F}$.

Lemma 7.11. *Let Ψ be a convex, strictly increasing function on $[0, \infty)$ with $0 \leq \Psi(0) \leq 1$. Suppose that the stochastic process Z indexed by the class $(\mathcal{F}, \rho)$ satisfies:*

(i). if $\rho(f, g) = 0$, then $Z(f) = Z(g)$ almost surely;

(ii). if $\rho(f, g) > 0$, then $\mathsf{E}\, \Psi\left(\frac{Z(f)-Z(g)}{\rho(f,g)}\right) \leq 1$;

(iii). there exists a point $f_0 \in S$ for which $\sup_{f\in\mathcal{F}} \rho(f, f_0) < \infty$;

(iv). the sample paths of Z are continuous on $(\mathcal{F}, \rho)$.

Then

$$\mathsf{E} \sup_{f\in\mathcal{F}} |Z(f) - Z(f_0)| \leq 8 \int_0^\theta \Psi^{-1}(N(\varepsilon, \mathcal{F}, \rho))d\,\varepsilon,$$

where θ equals one quarter of the supremum in (iii).

The following theorem is a minor modification of 2.11.11 Theorem of van der Vaart and Wellner (1996).

Theorem 7.12. *For each n, let $Z_{n1}, \ldots, Z_{n\,m_n}$ be independent stochastic processes indexed by an arbitrary index set $\mathcal{F}$. Let*

$$\sum_{i=1}^{m_n} \mathsf{E}^* \, \|Z_{ni}\|_\mathcal{F} \, \mathbb{I}_{\{\|Z_{ni}\|_\mathcal{F} > \eta\}} \to 0, \qquad \textit{for every } \eta > 0, \tag{7.2}$$

and suppose there exists a semimetric ρ such that for every ρ-ball $B(\varepsilon) \subset \mathcal{F}$

$$\sum_{i=1}^{m_n} \mathsf{E}^* \sup_{f,g\in B(\varepsilon)} [Z_{ni}(f) - Z_{ni}(g)]^2 \leq \varepsilon^2, \qquad \textit{for every } \varepsilon > 0, \tag{7.3}$$

and

$$\int_0^\infty \sqrt{\log N(\varepsilon, \mathcal{F}, \rho)}\, d\varepsilon < \infty. \tag{7.4}$$

Then the sequence $\sum_{i=1}^{m_n}(Z_{ni} - \mathsf{E}\, Z_{ni})$ is asymptotically tight in $\ell^\infty(\mathcal{F})$. It converges in distribution provided it converges marginally.

In comparison with 2.11.11 Theorem of van der Vaart and Wellner (1996), the only difference is that we replaced the original conditions

$$\sum_{i=1}^{m_n} (Z_{ni}(f) - Z_{ni}(g)) \leq \rho^2(f, g), \qquad \text{for every } f, g \in \mathcal{F},$$

and

$$\sup_{t>0} \sum_{i=1}^{m_n} t^2 \, \mathrm{P}^* \left(\sup_{f,g \in B(\varepsilon)} |Z_{ni}(f) - Z_{ni}(g)| > t \right) \le \varepsilon^2$$

with a slightly stronger, but in our situation easily verifiable, condition (7.3).

For our purposes, it is convenient to formulate this simple corollary of Theorem 7.12.

Corollary 7.13. *For each n, let $Z_{n1}, \ldots, Z_{n\,m_n}$ be independent stochastic processes indexed by T – a bounded subset of the Euclidean space $\mathbb{R}^p$. Put $\bar{Z}_n = \sum_{i=1}^{m_n} (Z_{ni} - \mathsf{E}\, Z_{ni})$. Suppose there exists a quantity r_n of order $o(1)$ as $n \to \infty$, a constant $C > 0$, and $\mathbf{t}_0 \in T$ such that*

$$\sum_{i=1}^{m_n} \mathsf{E}^* \sup_{|\mathbf{s}-\mathbf{t}|_2 < \varepsilon} [Z_{ni}(\mathbf{s}) - Z_{ni}(\mathbf{t})]^2 \le C\, \varepsilon\, r_n, \qquad \textit{for every } \varepsilon > 0,\ n \in \mathbb{N}, \tag{7.5}$$

and

$$\sum_{i=1}^{m_n} \mathsf{E}^* |Z_{ni}(\mathbf{t}_0)|^2 \to 0, \qquad \textit{as } n \to \infty. \tag{7.6}$$

Then $\mathsf{E}^* \|\bar{Z}_n\|_T \xrightarrow[n\to\infty]{} 0.$

Proof. It is easy to verify the conditions of Theorem 7.12, with the semimetric $\rho(\mathbf{s}, \mathbf{t}) = \sqrt{|\mathbf{t} - \mathbf{s}|_2}$.

Let $\varepsilon > 0$ be given. By the assumptions of Theorem 7.12 we can for every $q \in \mathbb{N}$ construct a partition $T = \bigcup_{j=1}^{N_q} T_j^q$ (which does not depend on n) such that

$$\sum_{q=1}^{\infty} 2^{-q} \sqrt{\log N_q} < \infty$$

and

$$\sum_{i=1}^{n} \mathsf{E}^* \sup_{\mathbf{t},\mathbf{s} \in T_j^q} [Z_{ni}(\mathbf{t}) - Z_{ni}(\mathbf{s})]^2 < \frac{1}{2^q} \qquad \text{for every } n \in \mathbb{N}.$$

We can argue similarly to the proof of 2.5.6 Theorem of van der Vaart and Wellner (1996) to show that without loss of generality we can choose the sequence (for $q = 1, 2, \ldots$) of partitions as successive refinements.

Further choose an element $\mathbf{t}_{qj}$ from each partitioning set T_{qj} and define a mapping

$$\pi_q : T \to F_q = \{\mathbf{t}_{q1}, \ldots, \mathbf{t}_{qN_q}\} \tag{7.7}$$

such that $\pi_q(\mathbf{t}) = \mathbf{t}_{qj}$ if $\mathbf{t} \in T_j$.

Now we can follow step by step the proof of Theorem 7.12 and show that for every $\varepsilon > 0$ there exist sufficiently large $q_\circ$ and n_0 such that for all $n \ge n_\circ$

$$\mathsf{E}^* \|\bar{Z}_n(\mathbf{t}) - \bar{Z}_n(\pi_{q_\circ}(\mathbf{t}))\|_T < \varepsilon. \tag{7.8}$$

As the set F_q in (7.7) is finite, it follows by the assumption (7.5) that

$$\max_{\mathbf{t}\in F_q} \mathsf{E}\,|\bar{Z}_n(\mathbf{t})|^2 = \max_{\mathbf{t}\in F_q} \sum_{i=1}^{m_n} \operatorname{var}\{\bar{Z}_{ni}(\mathbf{t})\} \leq \sum_{i=1}^{m_n} \mathsf{E}^*\,\|Z_{ni}\|_T^2 \to 0, \tag{7.9}$$

which implies that $\max_{\mathbf{t}\in F_q} \mathsf{E}\,|\bar{Z}_n(\mathbf{t})| \to 0$.

Combining (7.8) and (7.9) gives the statement of the corollary. □

Corollary 7.14. *Let the conditions of Corollary 7.13 be satisfied and for $k = 1,\dots,m_n$ put $Z_n^k = \sum_{i=1}^k Z_{ni}$. Then*

$$\mathrm{P}^*\left\{\max_{1\leq k\leq m_n} \|Z_n^k - \mathsf{E}\,Z_n^k\|_T > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0.$$

Proof. By Ottaviani's inequality for every $\varepsilon > 0$

$$\mathrm{P}^*\left\{\max_{k=1,\dots,m_n} \|\bar{Z}_n^k\|_T > 2\,\varepsilon\right\} \leq \frac{\mathrm{P}^*\left\{\|\bar{Z}_n\|_T > \varepsilon\right\}}{1 - \max_{k=1,\dots,m_n} \mathrm{P}^*\left\{\|\bar{Z}_n^{m_n} - \bar{Z}_n^k\|_T > \varepsilon\right\}}. \tag{7.10}$$

By Markov's inequality we can bound the numerator in (7.10) by $\frac{1}{\varepsilon}\,\mathsf{E}^*\,\|\bar{Z}_n^{m_n}\|_T$, which converges to zero by Corollary 7.13. Now, if the processes Z_{ni} were measurable, we could exploit the fact that $Y_k = \|\bar{Z}_n^{m_n} - \bar{Z}_n^k\|_T$ is an inverse submartingale and estimate

$$\mathrm{P}\left\{\|\bar{Z}_n^{m_n} - \bar{Z}_n^k\|_T > \varepsilon\right\} \leq \frac{1}{\varepsilon}\,\mathsf{E}\,\|\bar{Z}_n^{m_n} - \bar{Z}_n^k\|_T \leq \tfrac{1}{\varepsilon}\,\mathsf{E}\,\|\bar{Z}_n^{m_n}\|_T \to 0.$$

Unfortunately to the presence knowledge of the author, the case of the unmeasurable Z_{ni} cannot be handle in such a simple way. As we need this type of considerations in the proof of Lemma 5.12, we will argue more generally than we would need only for the purpose of this lemma.

Let us for contradiction suppose that there exists $\varepsilon > 0$ such that

$$\limsup_{n\to\infty} \max_{k=1,\dots,k_n} \mathrm{P}^*\left\{\|\bar{Z}_n^{m_n} - \bar{Z}_n^k\|_T > \varepsilon\right\} = 1. \tag{7.11}$$

Thus for every $j \in \mathbb{N}$ we can find $n_j \geq j$ and k_j such that

$$\mathrm{P}^*\left\{\|\bar{Z}_{n_j}^{m_{n_j}} - \bar{Z}_{n_j}^{k_j}\|_T > \varepsilon\right\} > 1 - \tfrac{1}{j}.$$

Let us now consider the triangular array $Z_{n_j\,k_j+1}, Z_{n_j\,k_j+2}, \dots, Z_{n_j\,m_{n_j}}$, where $j = 1, 2, \dots$. As this array is chosen from the original array $Z_{n1}, \dots, Z_{n\,m_n}$, it surely satisfies the assumptions of Corollary 7.13, which further implies

$$\mathsf{E}^*\,\|\bar{Z}_{n_j}^{m_{n_j}} - \bar{Z}_{n_j}^{k_j}\|_T \xrightarrow[j\to\infty]{} 0.$$

This convergence together with Markov's inequality contradict the assumption (7.11). Thus we conclude that the denominator in (7.10) stays away bounded from zero, which finishes the proof of this lemma. □

Miscellaneous

The following theorem showed to be useful when proving the existence of a consistent root of the defining equations (4.1) for M-estimators. It can be found in Ortega and Rheinboldt (1970) as Theorem 6.3.4.

Theorem 7.15. *Let C be an open, bounded set in $\mathbb{R}_p$ and assume that the mapping $H : \overline{C} \subset \mathbb{R}_p \to \mathbb{R}_p$ is continuous and satisfies $(x - x_0)H(x) \geq 0$ for some $x_0 \in C$ and all $x \in \partial C$. Then the system of equations $H(x) = 0$ has a solution in C.*

Applying Theorem 7.15 to the function $H'(x) = -H(x)$, we see that the condition $(x - x_0)H(x) \geq 0$ (for some $x_0 \in C$ and all $x \in \partial C$) may be equivalently rewritten as $(x - x_0)H(x) \leq 0$.

Definition 7.16. A sequence of random variables $\{Y_n\}$ is uniformly integrable if

$$\lim_{c \to \infty} \sup_{n \in \mathbb{N}} \mathsf{E}\, |Y_n|\, \mathbb{I}_{\{|Y_n| > c\}} = 0.$$

Some conditions for the sequence of random variables to be uniformly integrable can be found, e.g. in Serfling (1980).

Lemma 7.17. *Suppose that the function $g(x)$ is continuous and let X be a random variable such that there exists a $\delta_\circ > 0$ for which*

$$\sup_{(t,s) \in U} \mathsf{E}\, g^2\left(\tfrac{X+t}{s}\right) < \infty, \quad \textit{where} \quad U = [-\delta_\circ, \delta_\circ] \times [1 - \delta_\circ, 1 + \delta_\circ].$$

Then the (random) function $g\left(\frac{X+t}{s}\right)$ is continuous in the quadratic mean at the point $(0,1)$, that is

$$\lim_{(t,s) \to (0,1)} \mathsf{E}\left[g\left(\tfrac{X+t}{s}\right) - g(X)\right]^2 = 0.$$

Proof. Let $\varepsilon > 0$ be given. First find $K > 0$ such that $\mathsf{E}\, g^2(X)\, \mathbb{I}\{|X| \geq K\} < \varepsilon$. Without loss of generality we can assume $\delta_\circ < \frac{1}{2}$. Then $|X| \geq 2K + 1$ implies $|\frac{X+t}{s}| > K$ for all $(t,s) \in U$. This yields

$$\sup_{(t,s) \in U} \mathsf{E}\, g^2\left(\tfrac{X+t}{s}\right) \mathbb{I}\{|X| \geq 2K + 1\} < \varepsilon.$$

As the function $(x,t,s) \to \frac{x+t}{s}$ is continuous on $A = [-K, K] \times [-\frac{1}{2}, \frac{1}{2}] \times [\frac{1}{2}, \frac{3}{2}]$ and the function g is continuous, the composition of these functions $h(x,t,s) = g(\frac{x+t}{s})$ is continuous on A. Because the set A is compact, the function $h(x,t,s)$ is even uniformly continuous on A. That is, there exists δ $(0 < \delta < \delta_\circ)$ such that $|t| < \delta$ together with $|s - 1| < \delta$ imply $|h(x,t,s) - h(x,0,1)| < \varepsilon$. This further gives us for all $(t,s) \in [-\delta, \delta] \times [1 - \delta, 1 + \delta]$

$$\begin{aligned}
&\mathsf{E}\left[g\left(\tfrac{X+t}{s}\right) - g(X)\right]^2 \\
&\quad = \mathsf{E}\left[g\left(\tfrac{X+t}{s}\right) - g(X)\right]^2 \mathbf{I}\{|X| > 2K+1\} + \mathsf{E}\left[g\left(\tfrac{X+t}{s}\right) - g(X)\right]^2 \mathbf{I}\{|X| \leq 2K+1\} \\
&\quad \leq 4\varepsilon + \varepsilon^2.
\end{aligned}$$

□

Some scale estimators in linear models

In this section we give some estimators of scale for studentization of regression M-estimators. The aim of the section is not to give comprehensive list of possible scale estimators, but to show that there exist simple estimators which are not only $\sqrt{n}$-consistent, but moreover, the sequence $\sqrt{n}(S_n - S)$ meets the **SUB** condition and it is usually ucp as well. In fact, we only generalize slightly the results of Welsh (1986), who found the first order representation for the interquantile range (IQR) and the median absolute deviation (MAD) computed from the residuals of a preliminary regression fit.

More precisely, suppose that the $\hat{\boldsymbol{\beta}}^{(0)}$ is an initial estimator of the parameter $\boldsymbol{\beta}$. Put $r_i = Y_i - (\hat{\boldsymbol{\beta}}^{(0)})^{\mathsf{T}}\mathbf{x}_i$ for the residuals from this preliminary fit and denote Q_n the interquantile range and S_n the median absolute deviation of these residuals. We immediately see that if the initial estimator $\hat{\boldsymbol{\beta}}^{(0)}$ is regression equivariant and scale invariant, then both scale estimators Q_n and S_n are regression invariant and scale equivariant.

In the following, we will suppose that the linear model (1.1) includes an intercept, the initial estimator is $\sqrt{n}$-consistent, that is

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}) = O_p(1), \tag{7.12}$$

and the condition **XX.2** is satisfied. Further, let us denote $\xi_q = F^{-1}(q)$ and $\hat{\xi}_{qn} = \hat{F}_n^{-1}(q)$, where $\hat{F}_n$ is the empirical distribution function of the residuals.

Interquantile range Q_n

Welsh (1986) showed that provided the distribution function F of the errors has a positive and continuous derivative at the points $\xi_{1/4} = F^{-1}(1/4)$ and $\xi_{3/4} = F^{-1}(3/4)$, then the estimator Q_n admits the expansion

$$\sqrt{n}(Q_n - Q(F)) =$$
$$\tfrac{1}{\sqrt{n}f(\xi_{3/4})}\sum_{i=1}^{n}\left[\tfrac{3}{4} - \mathbb{I}\{e_i \le \xi_{3/4}\}\right] - \tfrac{1}{\sqrt{n}f(\xi_{1/4})}\sum_{i=1}^{n}\left[\tfrac{1}{4} - \mathbb{I}\{e_i \le \xi_{1/4}\}\right] + o_p(1), \tag{7.13}$$

where $Q(F) = \xi_{3/4} - \xi_{1/4}$. Our aim is to strengthen this result and show that the remainder term in the expansion (7.13) is $o_{ucp}(1)$ in fact.

For this purpose we define the process

$$M_n(\mathbf{t}, s) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[q - \mathbb{I}\{e_i \le \xi_q + \tfrac{s}{n} + \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}\}\right], \qquad |\mathbf{t}|_2 \le M, |s| \le M,$$

where M is an arbitrarily large but fixed constant.

It is rather standard (see Section 5.2.1) to show that if F has a continuous derivative at the point ξ_q, then

$$\sup_{|\mathbf{t}|_2 \le M, |s| \le M}\left| M_n(\mathbf{t}, s) - M_n(\mathbf{0}, 0) + s\, f(\xi_q) + \tfrac{f(\xi_q)\,\mathbf{t}^{\mathsf{T}}}{n}\sum_{i=1}^{n}\mathbf{x}_i \right| = o_{ucp}(1). \tag{7.14}$$

Now we would like to show that the sequence $\sqrt{n}(\hat{\xi}_{qn} - \xi_q)$ meets the **SUB** condition. But to show that we need that the sequence $\sqrt{n}(\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta})$ meets the **SUB** condition as well. Thus it may seem that we are moving in a circle. But fortunately, there exist regression estimators which are scale invariant but which do not require to estimate the scale. Probably the most popular ones are the least squares estimator and least absolute deviation (LAD) estimator. To keep our procedure robust we will choose the LAD method. This method also fits into our frame, because the LAD-estimator can be viewed as an M-estimator with the psi-function $\psi(x) = \mathrm{sign}(x)$. Thus we can use the results of Section 5.2.1 to conclude that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{LAD} - \boldsymbol{\beta}) = \frac{\mathbf{V}_n^{-1}}{\sqrt{n} f(\xi_{1/2})} \sum_{i=1}^{n} \mathbf{x}_i \left[\tfrac{1}{2} - \mathbb{I}\{e_i \leq \xi_{1/2}\}\right] + o_{ucp}(1), \tag{7.15}$$

provided f exists and is continuous and positive at the point $\xi_{1/2}$. It is easy to verify that (7.15) is even a stronger result than the **SUB** condition for $\sqrt{n}(\hat{\boldsymbol{\beta}}_{LAD} - \boldsymbol{\beta})$. With the help of the linearity result (7.14) and the fact that $\sqrt{n}(\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta})$ meets the **SUB** condition for sufficiently large n it holds

$$\begin{aligned}
&\mathrm{P}\left\{\max_{n_{-\delta} \leq k \leq n_\delta} \sqrt{k}(\hat{\xi}_{qk} - \xi_q) > C\right\} \\
&\quad \leq \mathrm{P}\left\{\max_{n_{-\delta} \leq k \leq n_\delta} \tfrac{1}{\sqrt{k}} \sum_{i=1}^{k} \left[q - \mathbb{I}\left\{e_i \leq \xi_q + \tfrac{C}{\sqrt{k}} + \sqrt{k}(\hat{\boldsymbol{\beta}}_k^{(0)} - \boldsymbol{\beta}) \sum_{i=1}^{k} \tfrac{\mathbf{x}_i}{k}\right\}\right] > 0\right\} \\
&\leq \mathrm{P}\left\{\max_{n_{-\delta} \leq k \leq n_\delta} \tfrac{1}{\sqrt{k}} \sum_{i=1}^{k} [q - \mathbb{I}\{e_i \leq \xi_q\}] - f(\xi_q)\sqrt{k}(\hat{\boldsymbol{\beta}}_k^{(0)} - \boldsymbol{\beta}) \sum_{i=1}^{k} \tfrac{\mathbf{x}_i}{k} > f(\xi_q)\,\tfrac{C}{2} + o_{ucp}(1)\right\}.
\end{aligned} \tag{7.16}$$

From (7.16) we see that if $f(\xi_q) > 0$ we can make $\mathrm{P}\left\{\max_{n_{-\delta} \leq k \leq n_\delta} \sqrt{k}(\hat{\xi}_{qk} - \xi_q) > C\right\}$ arbitrarily small for all sufficiently large n by taking C large enough and δ small enough. Similarly we can estimate $\mathrm{P}\left\{\max_{n_{-\delta} \leq k \leq n_\delta} \sqrt{k}(\hat{\xi}_{qk} - \xi_q) < -C\right\}$.

Now we are ready to substitute $\sqrt{n}(\hat{\xi}_{qn} - \xi_q)$ for s and $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(0)} - \boldsymbol{\beta})$ for $\mathbf{t}$ in (7.14) and after some rearrangements we get

$$\sqrt{n}(\hat{\xi}_{qn} - \xi_q) = \frac{1}{f(\xi_q)\sqrt{n}} \sum_{i=1}^{n} [q - \mathbb{I}\{e_i \leq \xi_q\}] - \sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(0)} - \boldsymbol{\beta}) \sum_{i=1}^{n} \frac{\mathbf{x}_i}{n} + o_{ucp}(1). \tag{7.17}$$

Combining (7.17) for $q = \frac{3}{4}$ and for $q = \frac{1}{4}$ we arrive at (7.13) with the remainder term $o_{ucp}(1)$.

Median absolute deviation S_n

By the median absolute deviation we mean

$$S_n = \mathrm{med}\{|r_i - \hat{\xi}_{0.5\,n}|,\ i = 1, \ldots, n\}. \tag{7.18}$$

But as the scale estimator we usually use $S_n' = S_n/\Phi^{-1}(0.75)$, so as the estimator is consistent for the normal errors. For simplicity of notation we will be dealing with S_n defined by (7.18). Welsh (1986) showed that provided F has a continuous derivative at the points $\xi_{0.5} - S$, $\xi_{0.5}$, and $\xi_{0.5} + S$, and $f(\xi_{0.5}) > 0$, $f(\xi_{0.5} - S) + f(\xi_{0.5} + S) > 0$, then S_n admits the first order expansion

$$\sqrt{n}(S_n - S) = \tfrac{1}{g_1\sqrt{n}} \sum_{i=1}^{n} \left[\tfrac{1}{2} - \mathbb{I}\{\xi_{0.5} - S < e_i < \xi_{0.5} + S\}\right] - \tfrac{g_2}{g_1\, f(\xi_{0.5})\,\sqrt{n}} \sum_{i=1}^{n} \left[\tfrac{1}{2} - \mathbb{I}\{e_i < \xi_{0.5}\}\right] + o_p(1) \quad (7.19)$$

where $g_1 = (f(\xi_{0.5} + S) + f(\xi_{0.5} - S))$ and $g_2 = (f(\xi_{0.5} + S) - f(\xi_{0.5} - S))$. Our aim is to prove that the remainder term in (7.19) is $o_{ucp}(1)$. For this purpose it is useful to study the process

$$M_n(\mathbf{t}, s, r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\tfrac{1}{2} - \mathbb{I}\{\xi_{0.5} - \tfrac{r}{\sqrt{n}} - S - \tfrac{s}{\sqrt{n}} + \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}} < e_i < \xi_{0.5} - \tfrac{r}{\sqrt{n}} + S + \tfrac{s}{\sqrt{n}} + \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}\}\right],$$

where $|\mathbf{t}|_2 \le M, |s| \le M, |r| \le M$.

With the techniques of Section 5.2.1 it is easy to show

$$\sup_{\max\{|\mathbf{t}|_2, |s|, |r|\} \le C} \left| M_n(\mathbf{t}, s, r) - M_n(\mathbf{0}, 0, 0) - g_1\, s - g_2\, r + g_2\, \mathbf{t}^{\mathsf{T}} \sum_{i=1}^{n} \tfrac{\mathbf{x}_i}{n} \right| = o_{ucp}(1) \quad (7.20)$$

From the discussion of interquantile range Q_n we know that if the sequence $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(0)} - \boldsymbol{\beta})$ meets **SUB**, then $\sqrt{n}(\hat{\xi}_{0.5\,n} - \xi_{0.5})$ satisfies **SUB** too. Then analogously to (7.16) we can show that $\sqrt{n}(S_n - S)$ satisfies **SUB** as well. Now we are ready to substitute $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(0)} - \boldsymbol{\beta})$ for $\mathbf{t}$, $\sqrt{n}(S_n - S)$ for s and $\sqrt{n}(\hat{\xi}_{0.5\,n} - \xi_{0.5})$ for r in (7.20). With the help of (7.17) we get

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\tfrac{1}{2} - \mathbb{I}\{\hat{\xi}_{0.5\,n} - S_n < r_i < \hat{\xi}_{0.5\,n} + S_n\}\right] + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\tfrac{1}{2} - \mathbb{I}\{\xi_{0.5} - S < e_i < \xi_{0.5} + S\}\right] - g_1\sqrt{n}(S_n - S) + -\tfrac{g_2}{f(\xi_{0.5})\sqrt{n}} \sum_{i=1}^{n} \left[\tfrac{1}{2} - \mathbb{I}\{e_i \le \xi_{0.5}\}\right] = o_{ucp}(1).$$

It is not difficult to show that the first term on the left-hand side is of order $o(1)$. Then after some algebra we get exactly (7.19) with the remainder term $o_{ucp}(1)$.

Notice that, similarly to the interquantile range Q_n, we need the sequence $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(0)} - \boldsymbol{\beta})$ to meet **SUB** but we do not need it to be ucp. Further, the asymptotic distributions of Q_n and S_n do not depend on the asymptotic distribution of the preliminary regression estimator $\hat{\boldsymbol{\beta}}_n^{(0)}$. Welsh (1986) also pointed out, that if the distribution of the errors is symmetric, then $g_2 = 0$, $S = \xi_{3/4}$ and the expansions (7.13) and (7.19) for the estimators Q_n and S_n coincide (up to a multiplication by a factor 2).

7.5 Prohorov metric for probability measures

Definition 7.18. Let μ and ν be Borel measures on a metric space (S, d). Then the Prohorov distance between measures μ and ν is defined as

$$d_P(\mu, \nu) = \inf\{\varepsilon > 0 : \mu(B) \leq \nu(B^\varepsilon) + \varepsilon \text{ for all Borel sets } B\},$$

where $B^\varepsilon = \{x : \inf_{y \in B} d(x, y) \leq \varepsilon\}$.

Prohorov metric is very important because it metrizes the weak convergence. The following two useful theorems are to be found in Dehling (1983).

Theorem 7.19. *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be independent $\mathbb{R}_d$-valued random variables with $\mathsf{E}\,\mathbf{X}_i = 0$ and $\mathsf{E}\,|\mathbf{X}_i|_2^3 < \infty$. Denote the distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{X}_i$ by μ_n and let ν_n be the Gaussian measure with mean zero and same covariance as μ_n. Then:*

$$d_P(\mu_n, \nu_n) \leq c\, d^{1/4} \rho_3^{1/4} n^{-1/8} (1 + |\log(\rho_3\, n^{-1/2} d^{-1})|^{1/2})$$

where $\rho_3 = n^{-1} \sum_{i=1}^n \mathsf{E}\,|\mathbf{X}_i|_2^3$ and c is an absolute constant.

Theorem 7.20. *Let μ and ν be two Gaussian measures on $\mathbb{R}_d$ with mean zero and covariance functions $\mathbf{T}$ and $\mathbf{S}$. Then the following estimation for their Prohorov distance holds:*

$$d_P(\mu, \nu) \leq C\, \|\mathbf{T} - \mathbf{S}\|_1\, d^{1/6} (1 + |\log(\|\mathbf{T} - \mathbf{S}\|_1^{-1}\, d)|^{1/2})$$

where C is an absolute constant and $\|\mathbf{A}\|_1 = \sum_{i=1}^d |\lambda_i|$, with λ_i being the eigenvalues of the (symmetric) $d \times d$ matrix $\mathbf{A}$.

If we denote λ_{max} the largest eigenvalue of the matrix $\mathbf{A}$ then

$$\|\mathbf{A}\|_1 \leq d\, \lambda_{max} \leq d \left[\sum_{i=1}^d \sum_{j=1}^d |a_{ij}|^2, \right]^{1/2} \tag{7.21}$$

where the first inequality is obvious and the second one is the well known relation of the largest eigenvalue and the Frobenius norm (see e.g. Theorem 3.1.3 of Dennis and Schnabel (1996)). For our purposes it is suffices that inequality (7.21) implies that $\|\mathbf{A}\|_1$ converges to zero, provided all its elements converges to zero.

Bibliography

Andrews, D. F., Bickel, P., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). *Robust Estimates of Location.* Princeton University Press.

Anscombe, J. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.*, 48:51–58.

Antille, A. (1976). Asymptotic linearity of Wilcoxon signed-rank statistics. *Ann. Statist.*, 4:175–186.

Arcones, M. A. and Mason, D. M. (1997). A general approach to Bahadur-Kiefer representations for M-estimators. *Mathematical Methods of Statistics*, 6:267–292.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Boos, D. D. (1977). Comparison of L- and M-estimators using the second term of the von Mises expansion. Technical report, North Carolina State University, Raleigh, North Carolina.

Boos, D. D. (1980). A new method for constructing approximate confidence intervals from M-estimates. *J. Amer. Statist. Assoc.*, 75:142–145.

Croux, C., Rousseeuw, P. J., and Hossjer, O. (1994). Generalized S-estimators. *J. Amer. Statist. Assoc.*, 89:1271–1281.

Dehling, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 63:393–432.

Dennis, J. E. J. and Schnabel, R. B. (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Classics In Applied Mathematics. SIAM.

Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.*, 39:325–346.

Hettmansperger, T., P. (1984). *Statistical inference based on ranks.* Wiley, New York.

Huber, P. J. (1973). Robust regression: Asymptotics, Conjectures and Monte Carlo. *Ann. Statist.*, 1:799–821.

Hušková, M. (1980). On bounded length sequential confidence interval for parameter in regression model based on ranks. In *Colloq. Math. Soc. János Bolyai*, pages 435–463. North-Holland, Amsterdam.

Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.*, 43:1449–1458.

Jurečková, J. (1971). Nonparametric estimation of regression coefficients. *Ann. Math. Statist.*, 42:1328–1338.

Jurečková, J. (1973). Central limit theorem for Wilcoxon rank statistics process. *Ann. Statist.*, 1:1046–1060.

Jurečková, J. (1977). Asymptotic relations of M-estimates and R-estimates in linear regression model. *Ann. Statist.*, 5:464–472.

Jurečková, J. (1978). Bounded-length sequential confidence intervals for regression and location parameters. In *Proceedings of the Second Prague Symposium on Asymptotic Statistic*, pages 239–250.

Jurečková, J. and Sen, P. K. (1981a). Invariance principles for some stochastic processes relating to M-estimators and their role in sequential statistical inference. *Sankhya A*, 43:190–210.

Jurečková, J. and Sen, P. K. (1981b). Sequential M-procedures with discontinuous scores. *J. Statist. Plann. Inference*, 5:253–267.

Jurečková, J. and Sen, P. K. (1989a). A second-order asymptotic distributional representation of M-estimators with discontinuous score functions. *Statist. Dec.*, 7:263–276.

Jurečková, J. and Sen, P. K. (1989b). Uniform second order asymptotic linearity of M-statistics in linear models. *Statist. Dec.*, 7:263–276.

Jurečková, J. and Sen, P. K. (1990). Effect of the initial estimator on the asymptotic behavior of one-step M-estimator. *Annals of the Institute of Statistical Mathematics*, 2:345–357.

Jurečková, J. and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations.* Wiley, New York.

Kersting, G. (1987). Second-order linearity of the general signed-rank statistics. *Journal of Multivariate Analysis*, 21:274–295.

Knight, K. (1997). Asymptotics for L_1 regression estimators under general conditions. Unpublished manuscript.

Koenker, R. (2005). *Quantile Regression.* Cambridge University Press, New York.

Koul, H., L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models.* Springer, New York. Second Edition.

Lachout, P. and Paulauskas, V. (2000). On the second-order asymptotic distribution of M-estimators. *Statist. Dec.*, 18:231–257.

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

McKean, J., W. (2004). Robust analysis of linear models. *Statistical Science*, 19:562–570.

Naranjo, J. D. and Hettmansperger, T. P. (1994). Bounded influence rank regression. *J. Royal Statistical Society. Series B.*, 56:209–220.

Nolan, D. and Pollard, D. (1987). U-processes: Rates of convergence. *Ann. Statist.*, 15:780–799.

Omelka, M. (2006). An alternative method for constructing confidence intervals from M-estimators in linear models. In *Proceedings of Prague Stochastics 2006.*

Ortega, J. M. and Rheinboldt, W. (1970). *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York.

Pollard, D. (1990). *Empirical Processes: Theory and Applications.* Institute of Mathematical Statistics.

Puri, M. L. and Wu, T.-J. (1985). Gaussian approximation of signed rank statistics process. *J. Statist. Plann. Inference*, 11:277–312.

Ren, J. J. (1994). On Hadamard differentiability and its application to R-estimation in linear models. *Statistics & Decision*, 12:1–12.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Siddiqui, M. (1960). Distribution of quantiles from a bivariate population. *Journal of Research of the National Bureau of Standards*, 64:145–150.

Sievers, G. L. (1983). A weighted dispersion function for estimation in linear models. *Commun. Statist. Theory Meth.*, 12:1161–1179.

Simpson, D. G., Ruppert, R. J., and Carroll, R. J. (1992). On One-Step GM Estimates and Stability of Inferences in Linear Regression. *J. Amer. Statist. Assoc.*, 87:439–450.

Stigler, S. (1986). *The history of statistics. The measurement of Uncertainty before 1900.* The Belknap Press of Harvard Univ. Press, Cambridge, Massachussets.

Terpstra, J. T. and McKean, J. W. (2004). Rank-Based Analysis of Linear Models using R. Technical Report.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes With Applications to Statistics.* Springer, New York.

Welsh, A. H. (1986). Bahadur representations for robust scale estimators based on regression residuals. *Ann. Statist.*, 14:1246–1251.

Welsh, A. H. and Ronchetti, E. (2002). A journey in single steps: robust one-step M-estimation in linear regression. *J. Statist. Plann. Inference*, 103:287–310.

List of Symbols

$a \wedge b$	$a \wedge b = \min(a, b)$
$a \vee b$	$a \vee b = \max(a, b)$
$\hat{\boldsymbol{\beta}}_n$	M-estimator of the parameter $\boldsymbol{\beta}$
$\hat{\boldsymbol{\beta}}_n$	R-estimator of the parameter $\boldsymbol{\beta}$ based on Wilcoxon scores
$\hat{\boldsymbol{\beta}}_n^+$	signed R-estimator of the parameter $\boldsymbol{\beta}$ based on Wilcoxon scores
$\boldsymbol{\beta}_z$	$\boldsymbol{\beta}_z = (\beta_1, \dots, \beta_{p-1})^\mathsf{T}$
$\hat{\boldsymbol{\beta}}_z$	$\hat{\boldsymbol{\beta}}_z = (\hat{\beta}_1, \dots, \hat{\beta}_{p-1})^\mathsf{T}$
$\mathbf{d}_{x\,\mathbf{z}}$	$\mathbf{d}_{x\,\mathbf{z}} = \frac{1}{n}\sum_{i=1}^n x_{ip}\,\mathbf{z}_i$
D_n^I	type I confidence interval
D_n^{II}	type II confidence interval
$D_n(\mathbf{b})$	$D_n(\mathbf{b}) = \sum_{i=1}^n (Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i)(R_i(\mathbf{b}) - \frac{n+1}{2})$
f	density of the errors in the model (1.1)
F	cumulative distribution function of the errors in the model (1.1)
Φ	cumulative distribution function of $N(0, 1)$
$\ell^\infty(T)$	set of all uniformly bounded real functions on T
$n_{-\delta}$	$n_{-\delta} = \lfloor n(1-\delta) \rfloor$
n_δ	$n_\delta = \lceil n(1+\delta) \rceil$
$\mathcal{N}(\mu, \sigma^2)$	normal (Gaussian) distribution on $\mathbb{R}$
$\mathcal{N}_k(\mu, \Sigma)$	normal (Gaussian) distribution on $\mathbb{R}_k$
$\mathbb{Q}$	rational numbers
$\mathbb{R}$	real numbers
$\mathbb{R}^*$	extended real line – $[-\infty, +\infty]$
$\mathbb{R}_p$	p-dimensional euclidean space
$\mathbf{S}_n(\mathbf{b})$	$\mathbf{S}_n(\mathbf{b}) = \frac{1}{n^{3/2}}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}_n) R_i(\mathbf{b})$
$\mathbf{S}_n^+(\mathbf{b})$	$\mathbf{S}_n^+(\mathbf{b}) = \frac{1}{n^{3/2}}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}_n)\,\mathrm{sign}(Y_i - \mathbf{b}^\mathsf{T}\mathbf{x}_i)\, R_i^+(\mathbf{b})$
$\mathbf{u}_l$	the vector with the l-th component 1 and the other components 0
T_{np}^2	$T_{np}^2 = \frac{1}{n}\sum_{i=1}^n x_{ip}^2$
T_{np}^3	$T_{np}^3 = \frac{1}{n}\sum_{i=1}^n \lvert x_{ip} \rvert^3$
$\mathbf{v}_1$	the first column of the matrix $\mathbf{V}_n$
$\mathbf{V}_n$	the matrix $\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\mathsf{T}$

$\mathbf{V}_n^{\mathbf{z}}$	the matrix $\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{\mathsf{T}}$
ω_{ii}^n	the i-th diagonal element of the matrix $\Omega_n = \mathbf{V}_n^{-1}$
Ω	the limit of $\mathbf{V}_n^{-1}$, that is $\Omega = \lim_{n\to\infty}\mathbf{V}_n$
ω_{ii}	the i-th diagonal element of the matrix Ω
$\mathbf{x}_i$	the i-th row of the design matrix $\mathbf{X}$
$\mathbf{X}$	the design matrix, $\mathbf{X} = (x_{ij})_{i=1,\ldots,n}^{j=1,\ldots,p}$
$\mathbf{Y}$	the vector of observations, $\mathbf{Y} = (Y_1,\ldots,Y_n)^{\mathsf{T}}$
z_α	$z_\alpha = \Phi^{-1}(1-\frac{\alpha}{2})$, with Φ^{-1} being the inverse cdf of $N(0,1)$
$\mathbf{z}_i$	$\mathbf{z}_i = (x_{i1},\ldots,x_{i\,p-1})^{\mathsf{T}}$

List of symbols depending on the type of estimator

The partial derivatives are denoted by lower subscripts.

M-estimator with fixed scale

Let us denote $\lambda(t) = \mathsf{E}\ \psi(e_1+t)$ and $\lambda^{(2)}(t) = \mathsf{E}\ \psi^2(e_1+t)$. Then

$$\gamma_1 = \lambda'(0), \qquad \gamma_2 = \lambda''(0), \qquad \gamma_{01} = \lambda_t^{(2)}(0).$$

For ψ absolutely continuous (plus some integrability conditions)

$$\gamma_1 = \mathsf{E}\ \psi'(e_1), \qquad \text{and} \qquad \gamma_{01} = 2\ \mathsf{E}\ \psi(e_1)\,\psi'(e_1).$$

For ψ a step function (2.14) or equivalently (2.15),

$$\gamma_1 = \sum_{j=1}^{m}\beta_j\, f(q_j) \qquad \text{and} \qquad \gamma_{01} = \sum_{j=1}^{m}\alpha_j^2\,[f(q_j)-f(q_{j-1})].$$

Finally

$$a_F^n = \frac{z_\alpha\,\sigma_\psi\,\sqrt{\omega_{pp}^n}}{\gamma_1}, \qquad a_F = \frac{z_\alpha\,\sigma_\psi\,\sqrt{\omega_{pp}}}{\gamma_1}, \qquad \text{where} \quad \sigma_\psi^2 = \mathsf{E}\ \psi^2(e_1).$$

Studentized M-estimator

Let us denote $\lambda(t,u) = \mathsf{E}\ \psi(\frac{e_1+t}{Se^u})$ and $\delta(t,u) = \mathsf{E}\ \frac{e_1}{S}\,\psi'(\frac{e_1+t}{Se^u})$. Then

$$\gamma_1 = \lambda_t(0,0), \quad \gamma_{1e} = -\lambda_u(0,0), \quad \gamma_2 = \lambda_{tt}(0,0), \quad \gamma_{2e} = \delta_t(0,0), \quad \gamma_{2ee} = -\delta_u(0,0).$$

Let us denote $\lambda^{(2)}(t,u) = \mathsf{E}\ \psi^2(\frac{e_1+t}{Se^u})$. Then

$$\gamma_{01} = \lambda_t^{(2)}(0,0), \qquad \gamma_{01e} = -\lambda_u^{(2)}(0,0).$$

For ψ absolutely continuous (plus some integrability conditions)

$$\gamma_1 = \mathsf{E}\ \tfrac{1}{S}\,\psi'(\tfrac{e_1}{S}), \quad \gamma_{1e} = \mathsf{E}\ \tfrac{e_1}{S}\,\psi'(\tfrac{e_1}{S}), \quad \gamma_{01} = 2\ \mathsf{E}\ \psi(\tfrac{e_1}{S})\,\psi'(\tfrac{e_1}{S}), \quad \gamma_{01e} = 2\ \mathsf{E}\ \tfrac{e_1}{S}\,\psi(\tfrac{e_1}{S})\,\psi'(\tfrac{e_1}{S}).$$

For ψ a step function

$$\gamma_1 = \sum_{j=1}^{m} \beta_j\, f(S\,q_j), \qquad \gamma_{1e} = -\sum_{j=1}^{m} S\,q_j \beta_j\, f(S\,q_j), \qquad \gamma_{01} = \sum_{j=1}^{m} \alpha_j^2\,[f(q_j) - f(q_{j-1})].$$

Finally

$$a_F^n = \frac{z_\alpha\,\sigma_\psi\,\sqrt{\omega_{pp}^n}}{\gamma_1}, \qquad a_F = \frac{z_\alpha\,\sigma_\psi\,\sqrt{\omega_{pp}}}{\gamma_1}, \qquad \text{where} \quad \sigma_\psi^2 = \mathsf{E}\ \psi^2(\tfrac{e_1}{S}).$$

R-estimator with Wilcoxon scores

$$\gamma = \mathsf{E}\ f(e_1), \qquad a_F^n = \frac{z_\alpha\,\sqrt{\omega_{pp}^n}}{\gamma\,\sqrt{12}}, \qquad a_F = \frac{z_\alpha\,\sqrt{\omega_{pp}}}{\gamma\,\sqrt{12}}.$$

List of abbreviations

a.s.	almost surely
cdf	cumulative distribution function
CI	confidence interval
FOAL	first order asymptotic linearity
FOAR	first order asymptotic representation
SOAL	second order asymptotic linearity
SOAR	second order asymptotic representation
SUB	‘sequential uniformly bounded’ (see (5.9))
ucp	uniformly continuous in probability (see Definition 5.2)

List of conditions

Conditions on the design points

X.1

$$\frac{1}{n}\sum_{i=1}^{n} c_{in}^2 = O(1), \qquad \lim_{n\to\infty} \frac{\max_{1\le i\le n} |c_{in}|}{\sqrt{n}} = 0.$$

X.2

$$\frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_{in}|_2^2 = O(1), \qquad \lim_{n\to\infty} \frac{\max_{1\le i\le n} |\mathbf{x}_{in}|_2}{\sqrt{n}} = 0.$$

X.3

$$\lim_{n\to\infty} \max_{1\le i\le n} \frac{|c_{in}|\,|\mathbf{x}_{in}|_2}{\sqrt{n}} = 0.$$

X.4

$$B_n^2 = \frac{1}{n}\sum_{i=1}^{n} c_{in}^2\,|\mathbf{x}_{in}|_2^2 = O(1).$$

X.5

$$\sum_{i=1}^{n} c_{in} = 0.$$

XX.1

$$\frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_{in}|_2^4 = O(1), \qquad \lim_{n\to\infty} \frac{\max_{1\le i\le n} |\mathbf{x}_{in}|_2}{\sqrt{n}} = 0.$$

XX.2 There exists a limit $(p\times p)$ matrix $\mathbf{V}$

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

which is positive definite.

XX.3 There exists a $\Delta \in \mathbb{R}$ such that

$$\Delta = \lim_{n\to\infty} \sqrt{n}(\sqrt{\omega_{pp}^n} - \sqrt{\omega_{pp}}).$$

X'.1

$$\frac{1}{n}\sum_{i=1}^{n} c_{in}^2 = O(1), \qquad \lim_{n\to\infty} \frac{\max_{1\le i\le n} |c_{in}|}{n^{1/4}} = 0.$$

X'.2

$$\lim_{n\to\infty} \frac{\max_{1\le i\le n} |\mathbf{x}_{in}|_2}{\sqrt{n}} = 0.$$

X'.3

$$\frac{1}{n}\sum_{i=1}^{n} |c_{in}|^2 |\mathbf{x}_{in}|_2 = O(1).$$

X'.4 There exists a $\delta > \frac{1}{2}$

$$B_n^2 = \frac{1}{n}\sum_{i=1}^{n} |c_{in}|\,|\mathbf{x}_{in}|_2^{1+\delta} = O(1).$$

XX'.1

$$\frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_{in}|_2^3 = O(1), \qquad \lim_{n\to\infty} \frac{\max_{1\le i\le n} |\mathbf{x}_{in}|_2^2}{n^{1/2}} = 0.$$

Conditions on the ψ function and cdf F used in M-estimation

A.1 ψ is a continuous piecewise linear function with the derivative

$$\psi'(x) = \alpha_j, \quad \text{for } r_j < x \leq r_{j+1},\ j = 1, \ldots, k,$$

where $\alpha_0, \alpha_1, \ldots, \alpha_k$ are real numbers, (usually $\alpha_0 = \alpha_k = 0$) and $-\infty = r_0 < r_1 < \ldots < r_k < r_{k+1} = \infty$.

A.2 The cumulative distribution function F is continuous at the points $r_1, \ldots, r_k$.

A.3 The cumulative distribution function F is absolutely continuous with a derivative which is continuous at the points $r_1, \ldots, r_k$.

GenFx.1 (**GenSt.1**) The function $h(t) = \mathsf{E}\ \rho(e_1 - t)$ (or $h(t) = \mathsf{E}\ \rho(\frac{e_1 - t}{S})$) has a unique minimum at $t = 0$.

SmFx.1 ψ is absolutely continuous with a derivative ψ' such that $\mathsf{E}\ \psi'(e_1)^2 < \infty$.

SmFx.2 The function $\psi'(e_1 + t)$ is continuous in the quadratic mean at the point 0, that is

$$\lim_{t \to 0} \mathsf{E}\,[\psi'(e_1 + t) - \psi'(e_1)]^2 = 0.$$

SmFx.3 The second derivative of the function $\lambda(s) = \mathsf{E}\ \psi(e_1 + t)$ is finite and continuous at the point 0.

SmFx.4 There exists a $\delta > 0$ such that $\sup_{|t|<\delta} \mathsf{E}\ \psi^4(e_1 + t) < \infty$

SmFx.5 The function $\lambda^{(2)}(t)$ is continuously differentiable in a neighbourhood of the point zero.

SmSt.1 ψ is absolutely continuous with a derivative ψ' such that $\mathsf{E}\ \psi'\left(\frac{e_1}{S}\right)^2 < \infty$.

SmSt.2 There exists $\delta > 0$ such that

$$\lim_{t \to 0} \sup_{|u|<\delta} \mathsf{E}\,[\psi'\left(\tfrac{e_1+t}{Se^u}\right) - \psi'\left(\tfrac{e_1}{Se^u}\right)]^2 = 0$$

and

$$\lim_{u \to 0} \mathsf{E}\,[\psi'\left(\tfrac{e_1}{Se^u}\right) - \psi'\left(\tfrac{e_1}{S}\right)]^2 = 0.$$

SmSt.3 The function $\lambda(t, u) = \mathsf{E}\ \psi(\frac{e_1+t}{Se^u})$ is twice differentiable and the partial derivatives are continuous and bounded in a neighbourhood of the point $(0, 0)$.

SmSt.4 There exists $\delta > 0$ such that $\sup_{|t|<\delta, |u|<\delta} \mathsf{E}\ \psi^4(\frac{e_1+t}{Se^u}) < \infty$.

SmSt.5 The function $\lambda^{(2)}(t, u) = \mathsf{E}\ \psi^2(\frac{e_1+t}{Se^u})$ is continuously differentiable in a neighbourhood of the point $(0, 0)$.

Step.1 F has a continuous derivative in a neighbourhood of the points $q_1, \ldots, q_m$.

Step.2 For every $j \in \{1, \ldots, m\}$ there exists a $\delta_j > 0$, $\nu_j > \frac{1}{2}$ and a $C_j < \infty$ such that for every $|t| < \delta_j$

$$|f(q_j + t) - f(q_j)| \leq C_j \, |t|^{\nu_j}.$$

Sym The distribution of the errors is symmetric and the ψ-function is antisymmetric, that is $F(x) = 1 - F(-x)$ and $\psi(x) = -\psi(-x)$ for all $x \in \mathbb{R}$.

Conditions on the distribution function and density used in R-estimation

W.1 F is absolutely continuous with a derivative f such that $\mathsf{E}\,[f(e_1)]^2 < \infty$.

W.2 The function $f(e_1 + s)$ is continuous in the quadratic mean at the point zero, that is

$$\lim_{s \to 0} \mathsf{E}\,[f(e_1 + s) - f(e_1)]^2 = 0.$$

W.3

$$\lim_{\Delta \to 0} \frac{1}{\Delta^2} \int_{-\infty}^{+\infty} \int_{-\Delta}^{+\Delta} [f(z + y) - f(y)]^2 dz \, dy = 0.$$

W.4 The density of the distribution of the errors is symmetric, that is $f(x) = f(-x)$, for any $x \in \mathbb{R}$.

W.5 There exist $\delta > 0$ and $C < \infty$ such that for all $|t| \leq \delta$

$$\mathsf{E}\,|f(e_1 - t) - f(e_1)| \leq Ct.$$