THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF STATISTICS

TECHNICAL REPORT # 224

UNIFORM ASYMPTOTICS FOR S- AND MM-REGRESSION ESTIMATORS

MAREK OMELKA AND MATIAS SALIBIAN-BARRERA

JULY 2006

Uniform asymptotics for S- and MM-regression estimators

Marek Omelka* and Matías Salibián-Barrera†

September 27, 2006

Abstract

In this paper we find verifiable regularity conditions to ensure that S-estimators of scale and regression and MM-estimators of regression are uniformly consistent and uniformly asymptotically normally distributed over contamination neighbourhoods. Moreover, we show how to calculate the size of these neighbourhoods. In particular, we find that, for MM-estimators computed with Tukey's family of bisquare score functions, there is a trade-off between the size of these neighbourhoods and both the breakdown point of the S-estimators and the leverage of the contamination that is allowed in the neighbourhood. These results extend previous work of Salibian-Barrera and Zamar for the location-scale model.

*Department of Probability and Mathematical Statistics, Charles University – Research was supported by the Charles University Grant GAUK 341/2005 and by the Czech Republic Grant GAČR 201/05/H007.

†Department of Statistics, The University of British Columbia – Research supported by the National Science and Engineering Research Council of Canada

1 Introduction

Many robust estimators have been proposed in the literature since Huber's seminal paper (Huber, 1964). Unfortunately, interest on robust inference methods seems to have fallen behind. Generally, robust inference relies on the asymptotic distribution of the estimator of interest. However, many asymptotic results available in the robustness literature require regularity conditions that are difficult to verify in practice, or that may not be valid precisely when robust methods are more appropriate. Typical regularity conditions under which asymptotic properties of robust estimators have been studied include: symmetric errors (see for example Bickel, 1975; Maronna and Yohai, 1981; Huber, 1981; Simpson *et. al*, 1992; Simpson and Yohai, 1998); known error scale (e.g. Huber, 1964; Markatou and Hettmansperger, 1990); or conditions that involve the expected value of the estimating equations under the unknown distribution of the data (e.g. Huber, 1981).

These types of regularity conditions limit the applicability of these results in practice. In particular, results derived under the assumption of symmetric errors only apply to uncontaminated data or to the rather uncommon situation of outliers that are symmetrically distributed. Furthermore, symmetry conditions are in general very hard to verify and when the distribution of the errors is not symmetric the asymptotic properties of location or regression robust estimators are affected in a non-trivial way by the choice of the scale estimate (Carroll, 1978, 1979; Rocke and Downs, 1981; Salibian-Barrera, 2000; Croux *et al.* 2003). By proceeding as if the formulae obtained under symmetry assumptions held we may be underestimating the variability of the estimate. This in turn may lead to lower than nominal confidence levels for intervals and to wrong sample size calculations.

Since according to the robustness model one does not know the actual distribution of the data, it is desirable to have robust estimators with asymptotic properties that hold uniformly over some set of plausible distributions. We refer the interested reader to Davies (1993, 1998) for a detailed discussion on this topic. Hampel (1971) showed that under certain regularity conditions, M-location estimators have uniform asymptotic properties over Prokhorov neighbourhoods. These are the first asymptotic results for robust estimators that hold uniformly over a set of distribution functions. Unfortunately the size of the neighbourhood where these uniform asymptotics hold is unknown. Huber (1967, 1981) shows that when the scale is known, M-location estimates are asymptotically normal and the approximation is uniform on the set of symmetric distribution functions with no mass on the points where the estimating equation is not differentiable. Uniform convergence of simultaneous estimators of location and scale with Hubers Proposal 2 has been studied by Clarke (1980). In Clarke (1986) it is shown that Hubers Proposal 2 estimates with nonsmooth estimating equations fall in the framework of Hampel (1971).

Davies (1998) constructed M-location estimates with simultaneous scale estimate (Huber Proposal II) that are locally asymptotically normal. Davies' results are "locally uniform", that is, for each distribution function there exists a neighbourhood of distributions where the convergence holds uniformly. Unfortunately, the size of these neighbourhoods is unknown. Moreover, the regularity conditions needed for Davies's construction include that the score functions used in the estimating equations are strictly monotone. Martin and Zamar (1993) showed that Huber's Proposal 2 estimates have larger asymptotic bias than the estimates that use a fixed ad-hoc scale. Furthermore, Berrendero and Zamar (1999) showed that strictly monotone score functions negatively affect the estimate's breakdown rate. From a practical point of view this means that Huber Proposal II estimates with strictly monotone score functions may have considerably larger asymptotic bias than other robust estimators.

Clarke (2000) showed that certain M-location estimates [including the simultaneous location and scale estimation proposed in Heathcote and Silvapulle (1981)] are continuous over full Prokhorov neighbourhoods of the parametric model. It follows that these estimates have uniform asymptotic behaviour over Prokhorov neighbourhoods. Unfortunately, as in Hampel (1971) and Davies (1998), the size of these neighbourhoods is unknown.

Recently, Salibian-Barrera and Zamar (2004) proved that M-estimators of location computed with an S-estimator of scale (Rousseeuw and Yohai, 1984) are consistent and asymptotically normally distributed uniformly over contamination neighbourhoods. Unlike previous results, the size of the neighbourhoods where these uniform asymptotic results hold can be calculated. In Salibian-Barrera and Zamar (2004) it is shown that there is a trade-off between the breakdown point of the S-estimator and the size of the neighbourhood where uniform asymptotics hold. For contamination neighbourhoods of the form $\mathcal{H}_{\varepsilon_0}$ in (2.2) these range from $\epsilon_0 = 0.11$ for estimators with 50% breakdown point to $\epsilon_0 = 0.25$ for estimators with 25% breakdown point. See also Berrendero and Zamar (2006) for similar results for M-estimators of location computed with a generalized S-estimator of scale (Croux *et al.* 1994).

In this paper we extend the results of Salibian-Barrera and Zamar (2004) to the linear regression model. More specifically, we obtain sufficient conditions to ensure uniform asymptotic properties for S-estimators (Rousseeuw and Yohai, 1984) and MM-estimators (Yohai, 1987). Moreover, we compute lower bounds for the size of the neighbourhoods where these results hold. For the regression model we found that the trade-off between the size of the contamination neighbourhoods where uniform asymptotic results hold now involves not only the the breakdown point of the estimators but also the maximum leverage of outliers that are present in the neighbourhood. Non surprisingly the size of these neighbourhoods is smaller than those found by Salibian-Barrera and Zamar (2004) for the location-scale model.

The rest of the paper is organized as follows. Section 2 contains some basic definitions. The uniform consistency of the S-scale estimator is discussed in Section 3, while Section 4 deals with the uniform consistency of the S-regression estimator. Their uniform asymptotic distributions are derived in Section 5. MM-estimators are considered studied in Section 6 while Section 7 contains some concluding remarks. Finally, all proofs are given in Section 8.

2 Definitions

Consider the usual regression model with random carriers where we observe i.i.d. random vectors $(Y_i, \mathbf{X}_i) \in \mathbb{R}^{p+1}$, $i = 1, \dots n$, where $Y_i \in \mathbb{R}$ and $\mathbf{X}_i \in \mathbb{R}^p$ satisfy

$$Y_i = \boldsymbol{\theta}_0^\top \mathbf{X}_i + u_i, \quad i = 1, \dots, n\,, \tag{2.1}$$

u_i are random errors independent from the covariates $\mathbf{X}_i$, and $\boldsymbol{\theta}_0 \in \mathbb{R}^p$ is the parameter of interest. Let $G_0(\mathbf{x})$ and $F_0(u)$ be the distribution of the carriers $\mathbf{X}_i$ and the errors u_i, respectively. Then the distribution of $(Y_i, \mathbf{X}_i)$ is given by $H_0(y, \mathbf{x}) = G_0(\mathbf{x})\, F_0(y - \boldsymbol{\theta}_0^\top \mathbf{x})$. We are concerned with the case where a certain proportion of the observations may not follow model (2.1) above. Thus, we will only assume that the distribution H of the observed data belongs to a contamination neighbourhood of H_0 of size ϵ_0. More precisely, we will assume that $H \in \mathcal{H}_{\varepsilon_0}$ where

$$\mathcal{H}_{\varepsilon_0} = \{H \in \mathcal{D} : H = (1-\varepsilon)H_0 + \varepsilon H^*, \varepsilon \in [0, \varepsilon_0]\} \ , \tag{2.2}$$

$\mathcal{D}$ is the set of all distribution functions, H^* is arbitrary and $\varepsilon_0 < 0.5$.

In what follows we will focus on S- and MM-regression estimators (see Rousseeuw and Yohai (1984) and Yohai (1987) respectively). MM-estimates are based on two loss functions ρ_0 and ρ_1, which determine the breakdown point and the efficiency of the estimate, respectively. More precisely, let $\hat{\sigma}_n$ be an scale S-estimate. That is, $\hat{\sigma}_n$ satisfies:

$$\hat{\sigma}_n = \min_{\boldsymbol{\theta} \in \mathbb{R}^p} \hat{\sigma}_n(\boldsymbol{\theta}) \,, \tag{2.3}$$

where $\hat{\sigma}_n(\boldsymbol{\theta})$ is implicitly defined by the equation

$$\frac{1}{n} \sum_{i=1}^{n} \rho_0 \left(\frac{y_i - \mathbf{x}_i' \boldsymbol{\theta}}{\hat{\sigma}_n(\boldsymbol{\theta})} \right) = b \,, \tag{2.4}$$

where,

A.1 ρ_0 is even, continuous, nondecreasing on $[0, \infty)$, $\rho_0(0) = 0$ and $\sup_{u \in \mathbb{R}} \rho_0(u) = 1$.

To ensure consistency of $\hat{\sigma}_n$ under the central model, we choose $b = E_{F_0}[\rho_0(u_1)]$. Moreover, in what follows we will assume that:

A.2 $\varepsilon_0 < E_{F_0}[\rho_0(u_1)] = b < 1 - \varepsilon_0 \quad ,$

where ε_0 is the size of the contamination neighbourhood (2.2). The breakdown point of $\hat{\sigma}_n$ and of $\tilde{\boldsymbol{\theta}}$ is given by $\min(b, 1-b)$.

For future reference, let $\tilde{\boldsymbol{\theta}}_n$ be the S-regression estimator, i.e.

$$\tilde{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} \hat{\sigma}_n(\boldsymbol{\theta}) \,. \tag{2.5}$$

Let $\rho_1 : \mathbb{R} \to \mathbb{R}_+$ be such that $\rho_1(u) \leq \rho_0(u)$ for all $u \in \mathbb{R}$ and $\sup_u \rho_1(u) = \sup_u \rho_0(u)$. The MM-regression estimator $\hat{\boldsymbol{\theta}}_n$ is defined as any local minimum of $f : \mathbb{R}^p \to \mathbb{R}_+$ defined by

$$f(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \rho_1 \left(\frac{y_i - \mathbf{x}_i' \boldsymbol{\theta}}{\hat{\sigma}_n} \right) ,$$

such that $f(\hat{\boldsymbol{\theta}}_n) \leq f(\tilde{\boldsymbol{\theta}}_n)$. For technical reasons we will choose the global minimum of $f(\boldsymbol{\theta})$ as our MM-regression estimator. It is easy to see that $\hat{\boldsymbol{\theta}}_n$ satisfies the following equations:

$$\frac{1}{n} \sum_{i=1}^{n} \rho_1' \left(\frac{y_i - \mathbf{x}_i' \hat{\theta}_n}{\hat{\sigma}_n} \right) \mathbf{x}_i = \mathbf{0} \,, \tag{2.6}$$

and

$$\sum_{i=1}^{n} \rho_1 \left(\frac{y_i - \mathbf{x}_i' \hat{\boldsymbol{\theta}}_n}{\hat{\sigma}_n} \right) \leq \sum_{i=1}^{n} \rho_1 \left(\frac{y_i - \mathbf{x}_i' \tilde{\boldsymbol{\theta}}_n}{\hat{\sigma}_n} \right) .$$

Since these estimators are affine equivariant, without loss of generality, in what follows we will assume that $\boldsymbol{\theta}_0 = \mathbf{0}$.

When the observed data follow the central model (2.1) the consistency and asymptotic distribution of MM-estimates has been studied by Yohai (1987) for the case of random covariates, and by Salibian-Barrera (2006) for fixed designs. Consistency and asymptotic distribution of S-estimators has been studied by Rousseeuw and Yohai (1984), Davies (1990) and Salibian-Barrera (2006).

3 Uniform consistency of the S-scale estimator in linear regression

For each $\boldsymbol{\theta} \in \mathbb{R}^p$ let the M-scale functional $\sigma(\cdot, \boldsymbol{\theta}) : \mathcal{F} \mapsto \mathbb{R}_+$ be defined by

$$\mathsf{E}_H \left[\rho_0 \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} X_1}{\sigma(H, \boldsymbol{\theta})} \right) \right] = b \,, \tag{3.1}$$

where we assume that $\mathcal{H}_{\varepsilon_0} \subset \mathcal{F}$. The S-scale functional $\sigma(\cdot) : \mathcal{F} \mapsto \mathbb{R}_+$ is then defined as

$$\sigma(H) = \inf_{\boldsymbol{\theta} \in \mathbb{R}^p} \sigma(H, \boldsymbol{\theta}) \,, \tag{3.2}$$

and the corresponding S-regression functional $\tilde{\boldsymbol{\theta}}(\cdot) : \mathcal{F} \mapsto \mathbb{R}_+$ is given by

$$\tilde{\boldsymbol{\theta}}(H) = \arg \inf_{\boldsymbol{\theta} \in \mathbb{R}^p} \sigma(H, \boldsymbol{\theta}) \,. \tag{3.3}$$

Under certain regularity conditions (see references above) the S-estimators of regression and scale are consistent to the functionals $\tilde{\boldsymbol{\theta}}\,(H)$ and $\boldsymbol{\sigma}\,(H)$, respectively.

We need the following regularity condition for the covariates $\mathbf{X}$:

X.1 $P_{G_0}\left[\boldsymbol{\theta}^{\mathsf{T}}\mathbf{X}_1 = 0\right] = 0$ for every $\boldsymbol{\theta} \neq \mathbf{0}$.

Moreover, for each $s > 0$ and $\boldsymbol{\theta} \in \mathbb{R}^p$ let

$$h(s, \boldsymbol{\theta}) = \frac{\partial}{\partial s} \left\{ \mathsf{E}_{H_0}\, \rho_0 \left(\frac{Y_1 - \boldsymbol{\theta}^{\mathsf{T}}\mathbf{X}_1}{s} \right) \right\} \,, \tag{3.4}$$

and assume that:

B.1 the function $h(s, \boldsymbol{\theta})$ is continuous and $h(s, \boldsymbol{\theta}) < 0$ for all $s > 0$ and $\boldsymbol{\theta} \in \mathbb{R}^p$.

The main result of this section shows that, under these conditions, $\hat{\sigma}_n$ in (2.3) converges to $\sigma(H)$ uniformly on $\mathcal{H}_{\varepsilon_0}$. We need the following definition to make this statement precise:

Definition 1 – **Uniform consistency** – *We say that the sequence of estimates $\hat{\tau}_n$ is uniformly consistent to the functional $\tau\,(F)$ over the contamination neighbourhood $\mathcal{H}_{\epsilon_0}$ if for all $\delta > 0$*

$$\lim_{m \to \infty} \sup_{F \in \mathcal{H}_{\epsilon_0}} P_F \left[\sup_{n \geq m} |\hat{\tau}_n - \tau\,(F)| > \delta \right] = 0 \,,$$

where $\tau\,(F)$ is the a.s. limit of $\hat{\tau}_n$ for an i.i.d. sequence of observations with distribution function F. We will denote this type of convergence by $\hat{\tau}_n \xrightarrow{\epsilon_0} \tau$.

The following theorem extends the results of Martin and Zamar (1993) for the location-scale model to the linear regression case.

Theorem 1 *Let* $\rho_0 : \mathbb{R} \to \mathbb{R}_+$ *satisfy* **A.1** *and* **A.2** *and assume that* **B.1** *and* **X.1** *hold. Then* $\hat{\sigma}_n \xrightarrow{\epsilon_0} \sigma$ *over* $\mathcal{H}_{\varepsilon_0}$.

Remark 1 *It is easy to show that if* ρ_0 *belongs to Tukey's bi-square family* (4.2) *then* **A.1** *and* **A.2** *hold. If, in addition, the central distribution of the errors* F_0 *is absolutely continuous with even, unimodal and positive density over the real line, then* **B.1** *also holds. Furthermore, note that assumptions* **B.1** *and* **X.1** *only impose conditions on the central distribution of* $\mathcal{H}_{\varepsilon_0}$ *and not on the distribution of the observed (and potentially contaminated) data.*

4 Uniform consistency of the S-regression estimator

For each $H \in \mathcal{H}_{\varepsilon_0}$, let the S-regression functional $\tilde{\boldsymbol{\theta}}(H)$ be defined by

$$\tilde{\boldsymbol{\theta}}(H) = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} \sigma(H, \boldsymbol{\theta}) \,,$$

where $\sigma(H, \boldsymbol{\theta})$ is given in (3.1). To simplify the notation, for each $s > 0$ and $\boldsymbol{\theta} \in \mathbb{R}^p$ let

$$g(H, \boldsymbol{\theta}, s) = \mathsf{E}_H \, \rho_0 \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s} \right) . \tag{4.1}$$

For our main result in this section we need the following regularity conditions (also see Salibian-Barrera, 2000):

U.1 let $0 < s_1 < s_2 < \infty$ be given in Lemma 1, then the function $g(H, \boldsymbol{\theta}, s)$ is continuous in $s \in [s_1, s_2]$ uniformly in $\boldsymbol{\theta} \in \mathbb{R}^p$ and $H \in \mathcal{H}_{\varepsilon_0}$. That is: for any

$\tilde{\varepsilon} > 0$ there exists a $\delta = \delta(\tilde{\varepsilon}) > 0$ such that $|s' - s''| < \delta$ implies

$$\left| g(H, \boldsymbol{\theta}, s') - g(H, \boldsymbol{\theta}, s'') \right| < \tilde{\varepsilon}, \qquad \forall\, \boldsymbol{\theta} \in \mathbb{R}^p,\ \forall\, H \in \mathcal{H}_{\varepsilon_0}\,;$$

U.2 for each $H \in \mathcal{H}_{\varepsilon_0}$, the function $f_H(\boldsymbol{\theta}) = g(H, \boldsymbol{\theta}, \sigma(H))$ has a unique minimum $\tilde{\boldsymbol{\theta}}(H)$; and

U.3 for every $\delta > 0$, let $\tilde{\varepsilon}(\delta, H)$ be defined by the property

$$\inf_{\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}(H)\| > \delta} g(H, \boldsymbol{\theta}, \sigma(H)) \geq g(H, \tilde{\boldsymbol{\theta}}(H), \sigma(H)) + \tilde{\varepsilon}(\delta, H),$$

where $\tilde{\boldsymbol{\theta}}(H)$ is the global minimum of $g(H, \boldsymbol{\theta}, \sigma(H))$, then $\tilde{\varepsilon}(\delta, H)$ satisfies

$$\tilde{\varepsilon}(\delta) = \inf_{H \in \mathcal{H}_{\varepsilon_0}} \tilde{\varepsilon}(\delta, H) > 0$$

Section 4.1 below discusses sufficient and verifiable conditions for U.1 to U.3 to hold. Moreover, we calculate lower bounds for ϵ_0 where U.1-3 hold. The following theorem shows that under these conditions, the S-estimator $\tilde{\boldsymbol{\theta}}_n$ in (2.5) is uniformly consistent over $\mathcal{H}_{\varepsilon_0}$.

Theorem 2 *If* **U.1** *to* **U.3** *and the assumptions of Theorem 1 hold, then* $\tilde{\boldsymbol{\theta}}_n \xrightarrow{\epsilon_0} \tilde{\boldsymbol{\theta}}$.

4.1 Verification of conditions U.1-3

In this section we discuss sufficient conditions for U.1 to U.3 to hold for S-regression estimates as in (2.5), when the loss function ρ_0 belongs to Tukey's bi-square family (Beaton and Tukey, 1974):

$$\rho^d(u) = \begin{cases} 3\,(u/d)^2 - 3\,(u/d)^4 + (u/d)^6 & \text{if } |u| \leq d\,, \\ 1 & \text{if } |u| > d\,, \end{cases} \tag{4.2}$$

where $d > 0$ is a user-chosen tuning constant. Recall that $g(H, \boldsymbol{\theta}, s) = E_H[\rho_0((Y - \boldsymbol{\theta}'\mathbf{X})/s)]$.

Condition U.1 - Using the mean value theorem we have that

$$\left|g(H,\boldsymbol{\theta},s') - g(H,\boldsymbol{\theta},s'')\right| \leq \left| - \mathsf{E}_H \left[\rho_0'\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s^*}\right)\left(\frac{Y_1 - \boldsymbol{\theta}'\mathbf{X}_1}{s^*}\right)\right]\frac{(s'-s'')}{s^*}\right| \\ \leq \left|\frac{s'-s''}{s_1}\right| \sup_x \left|x\,\rho_0'(x)\right|.$$

Since $\rho_0'(u) = 0$ for $|u| > d$ and bounded for $|u| \leq d$ we have $\sup_x |x\,\rho_0'(x)| < \infty$, and thus **U.1** holds.

Conditions U.2 & U.3 - Since these conditions are closely related we will consider them jointly. First, we show that we only need to verify them for $\boldsymbol{\theta}$ in a compact set. Specifically: in the proof of Theorem 1 we show that there exist $K < \infty$ and $\eta > 0$, such that for all $H \in \mathcal{H}_{\varepsilon_0}$ we have

$$\sigma(H) = \inf_{\|\boldsymbol{\theta}\| \leq K} \sigma(H,\boldsymbol{\theta}).$$

Moreover,

$$\inf_{\|\boldsymbol{\theta}\| > K} \mathsf{E}_H\,\rho_0\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{\sigma(H)}\right) \geq (1-\varepsilon_0)\,\mathsf{E}_{H_0}\,\rho_0\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{\sigma(H,\mathbf{0})}\right) > b + \eta.$$

It follows that the minimum of $g(H,\boldsymbol{\theta},\sigma(H))$ as a function of $\boldsymbol{\theta}$ is attained in the ball $\Theta_K = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \leq K\}$.

Hence, we see that to verify U.2 it is sufficient to show that, for all $H \in \mathcal{H}_{\varepsilon_0}$, the function $g(H,\boldsymbol{\theta},\sigma(H))$ is convex on Θ_K. Alternatively, it is sufficient to show that the matrix of second derivatives of $g(H,\boldsymbol{\theta},\sigma(H))$ is positive definite on Θ_K. Since $\sigma(H)$ is unknown but $\sigma(H) \in [s_1, s_2]$ (see Lemma 1), we will verify this condition for $\theta \in \Theta_K$ and $s \in [s_1, s_2]$. Differentiating with respect to $\boldsymbol{\theta}$ twice we obtain

$$s^2\left[\nabla^2_{\boldsymbol{\theta}} g(H, \boldsymbol{\theta}, s)\right] = \mathsf{E}_H\left[\rho_0''\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right)\mathbf{X}_1\mathbf{X}_1^\mathsf{T}\right]$$

$$= (1-\varepsilon)\,\mathsf{E}_{H_0}\left[\rho_0''\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right)\mathbf{X}_1\mathbf{X}_1^\mathsf{T}\right] + (1-\varepsilon)\,\mathsf{E}_{H^*}\left[\rho_0''\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right)\mathbf{X}_1\mathbf{X}_1^\mathsf{T}\right]$$

$$= (1-\varepsilon)\,\mathbf{A}(H_0, \boldsymbol{\theta}, s) + \varepsilon\,\mathbf{A}(H^*, \boldsymbol{\theta}, s)\,, \quad (4.3)$$

where $A(H, \boldsymbol{\theta}, s) = \mathsf{E}_H[\rho_0''((Y_1 - \boldsymbol{\theta}'\mathbf{X}_1)/s)\mathbf{X}_1\mathbf{X}_1']$. Thus, we need to show that (4.3) is positive definite for all $H \in \mathcal{H}_{\varepsilon_0}$, $\boldsymbol{\theta} \in \Theta_K$ and $s \in [s_1, s_2]$. It is enough to show that the minimum eigenvalue of this matrix is positive for all $H \in \mathcal{H}_{\varepsilon_0}$, $\boldsymbol{\theta} \in \Theta_K$ and $s \in [s_1, s_2]$. A sufficient condition is then that the smallest eigenvalue of the matrix $\mathbf{A}(H_0, \boldsymbol{\theta}, s)$ is greater than the largest eigenvalue of the matrix $\varepsilon\,\mathbf{A}(H^*, \boldsymbol{\theta}, s)/(1-\varepsilon)$. Equivalently, we need to verify that

$$\inf_{\|\boldsymbol{\theta}\| \le K}\ \inf_{s_1 \le s \le s_2} \lambda_{\text{MIN}}\left(\mathsf{E}_{H_0}\left[\rho_0''\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right)\mathbf{X}_1\mathbf{X}_1^\mathsf{T}\right]\right)$$

$$> \frac{(1-\varepsilon_0)}{\varepsilon_0}\ \sup_x \rho_0''(x)^-\ \sup_{H^*} \lambda_{\text{MAX}}\left(\mathsf{E}_{H^*}\left[\mathbf{X}_1\mathbf{X}_1^\mathsf{T}\right]\right)\,, \quad (4.4)$$

where $\lambda_{\text{MIN}}(\mathbf{A})$ and $\lambda_{\text{MAX}}(\mathbf{A})$ denote the minimum and maximum eigenvalue of the matrix $\mathbf{A}$, respectively. Therefore, condition (4.4) ensures that the functions $f(\boldsymbol{\theta}) = g(H, \boldsymbol{\theta}, s)$ are strictly convex on Θ_K and thus that they have a unique global minimum in this set, and condition U.2 holds. Furthermore, a Taylor expansion of $f(\boldsymbol{\theta}) = g(H, \boldsymbol{\theta}, \sigma(H))$ around $\boldsymbol{\theta}(H)$ shows that (4.4) is also sufficient for U.3.

Unfortunately, condition (4.4) is rather strong. In the rest of this section we show how it can be relaxed. The main idea is to note that we do not need to verify (4.4) for all combinations of $\boldsymbol{\theta} \in \Theta_K$ and $s \in [s_1, s_2]$. For each $\boldsymbol{\theta} \in \Theta_K$ let

$$\mathcal{A}(\boldsymbol{\theta}) = \{s_H(\boldsymbol{\theta}) : g(H, \boldsymbol{\theta}, s_H(\boldsymbol{\theta})) = b,\ H \in \mathcal{H}_{\varepsilon_0}\}.$$

By a standard argument we can show that $\mathcal{A}(\boldsymbol{\theta}) \subset [s_1(\boldsymbol{\theta}), s_2(\boldsymbol{\theta})]$, where $s_1(\boldsymbol{\theta})$ and $s_2(\boldsymbol{\theta})$ solve equations

$$(1-\varepsilon_0)\, \mathsf{E}_{H_0}\, \rho_0 \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s_1(\boldsymbol{\theta})} \right) = b, \quad \text{and} \quad (1-\varepsilon_0)\, \mathsf{E}_{H_0}\, \rho_0 \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s_2(\boldsymbol{\theta})} \right) = b-\varepsilon_0\,,$$

respectively. It is trivial but useful to notice that instead of $s_2(\boldsymbol{\theta})$ it is sufficient to take $s_2^*(\boldsymbol{\theta}) = \min\{s_2(\boldsymbol{\theta}), s_2(\mathbf{0})\}$. This implies that while checking condition (4.4) for a particular value of $\boldsymbol{\theta}$ we can consider only those values of $s \in [s_1(\boldsymbol{\theta}), s_2^*(\boldsymbol{\theta})]$. Furthermore, if $\boldsymbol{\theta}$ is a minimum of $g(H, \boldsymbol{\theta}, \sigma(H))$, then it satisfies

$$(1-\varepsilon)\, \mathsf{E}_{H_0} \left[\rho_0' \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s_H(\boldsymbol{\theta})} \right) \mathbf{X}_1 \right] + \varepsilon\, \mathsf{E}_{H^*} \left[\rho_0' \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s_H(\boldsymbol{\theta})} \right) \mathbf{X}_1 \right] = \mathbf{0}\,, \quad (4.5)$$

or, equivalently

$$(1-\varepsilon)\, \mathsf{E}_{H_0} \left[\rho_0' \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s_H(\boldsymbol{\theta})} \right) \mathbf{X}_1 \right] = -\varepsilon\, \mathsf{E}_{H^*} \left[\rho_0' \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s_H(\boldsymbol{\theta})} \right) \mathbf{X}_1 \right].$$

This gives us for all $s \in [s_1(\boldsymbol{\theta}), s_2(\boldsymbol{\theta})]$ the following coordinate-wise inequality

$$\left| \mathsf{E}_{H_0} \left[\rho_0' \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s} \right) \mathbf{X}_1 \right] \right| \leq \frac{\varepsilon_0}{1-\varepsilon_0} \sup_x |\rho_0'(x)|\; \mathsf{E}_{H^*} |\mathbf{X}_1|\,, \quad (4.6)$$

and it follows that equation (4.5) cannot be satisfied if any component of the left-hand side of (4.6) is greater than the corresponding coordinate on the right-hand side. Hence, (4.4) only needs to be verified for those $\boldsymbol{\theta}$ for which (4.6) holds.

We verified these conditions numerically for the case when the covariates in the central model are independent standard normal random variables. The model includes an intercept and one and two covariates ($p = 1$ and $p = 2$ respectively). We used functions $\rho_0(u)$ in Tukey's bisquare family (4.2) with tuning constants d that yield S-estimators with breakdown points 25, 30, 35, 40, 45 and 50%. We considered contamination distributions H^* that produce outliers with different degrees of leverage by setting $\mathsf{E}(X_1)$ to be 1, 2, ..., 5. In all cases we set $\mathrm{Var}(X_1) = 0.5$. In other words, we are allowing outliers with varying leverage degrees and that are, without loss of generality, concentrated along X_1. For each tuning constant d, and for each value of $\mathsf{E}(X_1)$ Table 1 contains lower bounds for the proportion $\epsilon^*(d)$ of this type of contaminations for which conditions U.2 and U.3 hold when $p = 1$ whereas Table 2 displays the corresponding values for the case $p = 2$. We see that the size of the sets of distributions where uniform asymptotic properties hold decreases both with the breakdown point of the estimator and with the leverage of the contamination that is allowed. For example, when $p = 1$, if $\mathcal{H}_{\varepsilon_0}$ only allows for contaminations that produce outliers inside the ball of radius 3, then the size of the sets (2.2) where uniform consistency holds ranges from $\epsilon_0 = 0.04$ for estimators with 25% breakdown point to $\epsilon_0 = 0.01$ for estimators with 50% breakdown point.

5 Uniform asymptotic normality of the S-estimators

The main result in this section shows that, under certain regularity conditions, $\hat{\sigma}_n$ and $\tilde{\boldsymbol{\theta}}_n$ are uniformly linearizable. It then follows that they are asymptotically normally distributed uniformly over the contamination neighbourhood. We first list the required regularity conditions and the main theorem. In Section 5.1 we show when these conditions hold.

In what follows we need the covariates to satisfy:

X.2 $\sup_{H\in\mathcal{H}_{\varepsilon_0}} \mathsf{E}_H \left\|\mathbf{X}_1\mathbf{X}_1^{\mathsf{T}}\right\| < \infty$, where $\|\mathbf{A}\| = \sum_i \sum_j |a_{ij}|$.

We will also assume that the function ρ_0 is twice differentiable and that:

N.1 the following functions are continuous at the point $(\tilde{\boldsymbol{\beta}}(H), \sigma(H))$, uniformly

BP	$\mathsf{E}(X_1)$	ϵ_0	BP	$\mathsf{E}(X_1)$	ϵ_0
25%	1	0.14	40%	1	0.08
	2	0.08		2	0.04
	3	0.04		3	0.02
	4	0.03		4	0.01
	5	0.02		5	0.01
30%	1	0.12	45%	1	0.06
	2	0.06		2	0.03
	3	0.04		3	0.02
	4	0.02		4	0.01
	5	0.01		5	0.00
35%	1	0.10	50%	1	0.05
	2	0.05		2	0.02
	3	0.03		3	0.01
	4	0.02		4	0.01
	5	0.01		5	0.00

Table 1: Lower bound for the size of the contamination neighbourhood where uniform asymptotic properties hold for S-estimators with breakdown points between 25% and 50%. The model includes an intercept and a single covariate ($p = 1$) with $\mathrm{Var}(X_1) = 0.5$.

BP	$\mathsf{E}(X_1)$	ϵ_0	BP	$\mathsf{E}(X_1)$	ϵ_0
25%	1	0.12	40%	1	0.07
	2	0.08		2	0.04
	3	0.05		3	0.02
	4	0.03		4	0.01
	5	0.02		5	0.01
30%	1	0.11	45%	1	0.06
	2	0.07		2	0.03
	3	0.04		3	0.02
	4	0.02		4	0.01
	5	0.01		5	0.00
35%	1	0.09	50%	1	0.05
	2	0.05		2	0.02
	3	0.03		3	0.01
	4	0.02		4	0.01
	5	0.01		5	0.00

Table 2: Lower bound for the size of the contamination neighbourhood where uniform asymptotic properties hold for S-estimators with breakdown points between 25% and 50%. The model includes an intercept and two covariates ($p = 2$) with $\mathsf{E}(X_2) = 0$ and $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 0.5$.

in $\mathcal{H}_{\varepsilon_0}$:

$$
\begin{aligned}
a_1^H(\boldsymbol{\theta}, s) &= \mathsf{E}_H\left[\rho_0'\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right)\frac{\mathbf{X}_1}{s}\right], \\
a_2^H(\boldsymbol{\theta}, s) &= \mathsf{E}_H\left[\rho_0'\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right)\frac{(Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1)}{s}\right], \\
a_3^H(\boldsymbol{\theta}, s) &= \mathsf{E}_H\left[\rho_0''\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{s^2}\right], \\
a_4^H(\boldsymbol{\theta}, s) &= \mathsf{E}_H\left[\rho_0''\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right)\frac{(Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1)\mathbf{X}_1^\mathsf{T}}{s^2}\right].
\end{aligned}
$$

In other words, for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $H \in \mathcal{H}_{\varepsilon_0}$ and $i = 1, 2, 3, 4$, we have $\left\| a_i^H(\boldsymbol{\theta}, s) - a_i^H(\tilde{\boldsymbol{\beta}}(H), \sigma(H)) \right\| < \varepsilon$, whenever $\|\boldsymbol{\theta} - \tilde{\beta}(H)\| < \delta$ and $|s - \sigma(H)| < \delta$.

Moreover, we will assume that ρ_0 is such that:

A.3 $\rho_0'(t)$, $\rho_0'(t)t$, $\rho_0''(t)$ and $\rho_0''(t)t$ can be uniformly approximated by finite linear combinations of indicator functions as in Definition 2 below.

Definition 2 *We say that the function* $f : \mathbb{R} \mapsto \mathbb{R}$ *can be uniformly approximated by finite linear combinations of indicator functions if for every* $\varepsilon > 0$ *there exist constants* $a_1, \ldots, a_k$ *and* $b_1, \ldots, b_k$ *such that*

$$
\sup_{u \in \mathbb{R}} \left| f(u) - \sum_{j=1}^{k} a_j \, \mathbb{I}\{u > b_j\} \right| < \varepsilon.
$$

Finally, let $\mathbf{C}_H = a_3^H(\tilde{\boldsymbol{\beta}}(H), \sigma(H)) \in \mathbb{R}^{p \times p}$ and assume that:

N.2 $\lambda_1 = \inf_{H \in \mathcal{H}_{\varepsilon_0}} \lambda_{MIN}(\mathbf{C}_H) > 0$.

We have the following

Theorem 3 *Assume that conditions* ***B.1****,* ***A.1-3****,* ***X.1-2****,* ***U.1-3****, and* ***N.1-2*** *are satisfied. Then*

$$\sqrt{n}(\hat{\sigma}_n - \sigma(H)) = \frac{1}{b_H \sqrt{n}} \sum_{i=1}^{n} [\rho(\tilde{u}_i(H)) - b] + Uo_P(1)\,, \tag{5.1}$$

and

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}(H)) = \mathbf{C}_H^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho'(\tilde{u}_i(H))\, \mathbf{X}_i - \mathbf{C}_H^{-1} \mathbf{d}_H \sqrt{n}[\hat{\sigma}_n - \sigma(H)] + Uo_P(1)\,, \tag{5.2}$$

where $\mathbf{b}_H = a_2^H(\tilde{\boldsymbol{\theta}}(H), \sigma(H))$, $\mathbf{d}_H = a_4^H(\tilde{\boldsymbol{\theta}}(H), \sigma(H))$, $\tilde{u}_i(H) = (Y_i - \tilde{\boldsymbol{\theta}}(H)^\top \mathbf{X}_i)/\sigma(H)$ *and* $Uo_P(1)$ *denotes a term that approaches zero in probability uniformly on* $H \in \mathcal{H}_{\varepsilon_0}$ *(see Definition 4 on page 28).*

Remark 2 *From* (5.2) *we see that the sequence* $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}(H))$ *is asymptotically normally distributed with covariance matrix* $\boldsymbol{\Sigma}_H$*, given by*

$$\begin{aligned}\boldsymbol{\Sigma}_H = \mathbf{C}_H^{-1}\, \mathsf{E}_H \left[\rho'^{\,2}(u_1(H)) \tfrac{\mathbf{X}_1 \mathbf{X}_1^\top}{\sigma(H)^2}\right] \mathbf{C}_H^{-1} + \mathbf{C}_H^{-1}\, \tfrac{\mathbf{d}_H}{b_H}\, \tfrac{\mathbf{d}_H^\top}{b_H}\, \mathbf{C}_H^{-1}\, \mathsf{E}_H \left[(\rho(u_1(H)) - b)^2\right] \\ - \mathbf{C}_H^{-1}\, \mathsf{E}_H \left[\rho'(\tilde{u}_1(H))\, (\rho(\tilde{u}_1(H)) - b)\, \tfrac{\mathbf{X}_1}{\sigma(H)}\right] \tfrac{\mathbf{d}_H^\top}{b_H}\, \mathbf{C}_H^{-1} \\ - \mathbf{C}_H^{-1} \tfrac{\mathbf{d}_H}{b_H}\, \mathsf{E}_H \left[\rho'(\tilde{u}_1(H))\, (\rho(\tilde{u}_1(H)) - b)\, \tfrac{\mathbf{X}_1^\top}{\sigma(H)}\right] \mathbf{C}_H^{-1}\,.\end{aligned} \tag{5.3}$$

Since S-regression estimators are particular cases of MM-regression estimators, it can easily be shown that the above formula coincides with that of Croux et al. (2003) for MM-regression estimators for the case of homoscedastic and independent errors.

The following corollary shows that the standardized sequence $\boldsymbol{\Sigma}_H^{-1/2}\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}(H))$ is, for sufficiently large n, uniformly close to a p-variate standard normal distribution.

Corollary 1 *Suppose that the conditions of Theorem 3 are satisfied and that*

$$\lim_{x\to\infty} \sup_{H\in\mathcal{H}_{\varepsilon_0}} \mathsf{E}_H \left[\|\rho'(\tilde{u}_1(H))\, \mathbf{X}_1\|^2\, \mathbb{I}\{\|\rho'(\tilde{u}_1(H))\, \mathbf{X}_1\| > x\}\right] = 0\,. \tag{5.4}$$

Let ν_n^H denote the distribution of the random vector $\mathbf{\Sigma}_H^{-1/2}\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}(H))$ and let μ denote a p-variate standard normal distribution. Then

$$\lim_{n\to\infty} \sup_{H\in\mathcal{H}_{\varepsilon_0}} d_P(\nu_n^H, \mu) = 0,$$

where d_P denotes the Prokhorov metric.

Remark 3 *Corollary 1 implies that*

$$\lim_{n\to\infty} \sup_{H\in\mathcal{H}_{\varepsilon_0}} \sup_{\mathbf{x}\in\mathbb{R}_p} \left| P_H \left[\mathbf{\Sigma}_H^{-1/2}\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}(H)) \le \mathbf{x} \right] - \Phi(\mathbf{x}) \right| = 0\,. \qquad (5.5)$$

This is proved in Section 8.

5.1 Sufficient conditions for the regularity conditions of this Section

Condition A.3 - It is easy to verify that **A.3** is satisfied for ρ_0 in Tukey's family (4.2).

Conditions N.1 & N.2 - If ρ_0 belongs to Tukey's family (4.2) then condition **N.1** is satisfied whenever **X.2** holds because all the functions $\rho_0'(x), x\,\rho_0'(x), \rho_0''(x), x\,\rho_0''(x)$ are continuous. Finally, note that condition **N.2** is weaker than (4.4).

6 MM-estimators

Let $\rho_1 : \mathbb{R} \to \mathbb{R}_+$ as in (2.6) and assume that it satisfies **A.1**. In what follows we need the MM-estimator $\hat{\boldsymbol{\theta}}_n$ to satisfy:

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\|\boldsymbol{\theta}\|\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^n \rho_1\left(\frac{Y_i - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\hat{\sigma}_n}\right)\,.$$

Similarly, define the corresponding function $\boldsymbol{\theta}(H) : \mathcal{F} \to \mathbb{R}^p$ by

$$\boldsymbol{\theta}(H) = \arg \min_{\|\boldsymbol{\theta}\| \in \mathbb{R}^p} \mathsf{E}\, \rho_1 \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{\sigma(H)} \right) ,$$

where $\sigma(H)$ is given by (3.2). The proof of the uniform strong consistency of the MM-estimator is very similar to that of the S-regression estimator. In particular, note that

$$\begin{aligned} & P_H \left[\left\| \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}(H) \right\| > \delta \right] \\ & \quad \leq P_H \left[\inf_{\|\boldsymbol{\theta} - \boldsymbol{\theta}(H)\| > \delta} \frac{1}{n} \sum_{i=1}^{n} \rho_1 \left(\frac{Y_i - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_i}{\hat{\sigma}_n} \right) \leq \frac{1}{n} \sum_{i=1}^{n} \rho_1 \left(\frac{Y_i - \hat{\boldsymbol{\theta}}_n^\mathsf{T} \mathbf{X}_i}{\hat{\sigma}_n} \right) \right] . \end{aligned} \tag{6.1}$$

But as we can easily show,

$$g(H_n, \hat{\beta}(H), \sigma(H_n)) \xrightarrow[n \to \infty]{\text{a.s.}} \mathsf{E}_H\, \rho_1 \left(\frac{Y_1 - \hat{\beta}(H)^\mathsf{T} \mathbf{X}_1}{\sigma(H)} \right) ,$$

uniformly over $\mathcal{H}_{\varepsilon_0}$. So we can replace $g(H_n, \hat{\beta}(H), \sigma(H_n))$ by the quantity

$$\mathsf{E}_H\, \rho_1 \left(\frac{Y_1 - \hat{\beta}(H)^\mathsf{T} \mathbf{X}_1}{\sigma(H)} \right) + \gamma\, ,$$

where $\gamma > 0$ is arbitrarily small, and then follow the proof of the consistency of the regression S-estimator. The proof of the uniform asymptotic normality of MM-estimator is completely analogous to the proof for the S-estimator.

7 Conclusion

In this paper we find verifiable regularity conditions to ensure that S- and MM-regression estimators are uniformly consistent and asymptotically normally distributed over contamination neighbourhoods. These regularity conditions only involve the central model of the contamination neighbourhood and the estimating equations of the estimators. Moreover, we show how to compute the size of the neighbourhoods where these uniform asymptotic results hold. We verify that S-scale, S-regression and MM-regression estimators satisfy these conditions when the central model is normal and the estimators are computed using Tukey's bisquare family of score functions. While Salibian-Barrera and Zamar (2004) showed that for the simpler location-scale model there is a trade-off between the size of the contamination neighbourhoods where these uniform asymptotic results hold and the breakdown point of the S-estimator, our calculations show that for the linear regression model the trade-off additionally involves the leverage of the contamination that is allowed in the neighbourhood. Our results extend those of Salibian-Barrera and Zamar (2004) to the linear regression model.

8 Proofs

Lemma 1 *Let $K > 0$ be given and suppose that assumptions **A.1-2** hold. Then, there exist $0 < s_1 < s_2 < \infty$ such that $s_1 \leq \sigma(H, \boldsymbol{\theta}) \leq s_2$ for all $\|\boldsymbol{\theta}\| < K$ and $H \in \mathcal{H}_{\varepsilon_0}$.*

Proof (a) Existence of s_2

$$\begin{aligned}\mathsf{E}_H\, \rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right) &\leq (1-\varepsilon_0)\,\mathsf{E}_{H_0}\,\rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right) + \varepsilon_0 \\ &\leq (1-\varepsilon_0)\,\mathsf{E}_{H_0}\,\rho\left(\frac{|u_1| + \|\boldsymbol{\theta}\|\,\|\mathbf{X}_1\|}{s}\right) + \varepsilon_0 \\ &\leq (1-\varepsilon_0)\,\mathsf{E}_{H_0}\,\rho\left(\frac{|u_1| + K\|\mathbf{X}_1\|}{s}\right) + \varepsilon_0 \xrightarrow[s\to\infty]{} \varepsilon_0 < b.\end{aligned}$$

So there exists s_2 ($s_2 < \infty$) such that for all $s > s_2$

$$\sup_{H\in\mathcal{H}_{\varepsilon_0}}\ \sup_{\|\boldsymbol{\theta}\|\leq K} \mathsf{E}_H\,\rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right) < b,$$

which implies that

$$\sup_{H \in \mathcal{H}_{\varepsilon_0}} \sup_{\|\boldsymbol{\theta}\| \le K} \sigma(H, \boldsymbol{\theta}) \le s_2 .$$

(b) Existence of s_1

Note that

$$\mathsf{E}_H \, \rho \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s} \right) \ge (1 - \varepsilon_0) \, \mathsf{E}_{H_0} \, \rho \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s} \right),$$

and define

$$f(s) = \inf_{\|\boldsymbol{\theta}\| \le K} \mathsf{E}_{H_0} \, \rho \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s} \right).$$

We will now show that $\lim_{s \to 0_+} f(s) = 1$. Let $\delta > 0$ be given. Since for each fixed $\boldsymbol{\theta} \in \{\|\boldsymbol{\theta}\| \le K\}$ we have

$$\lim_{s \to 0_+} \mathsf{E}_{H_0} \, \rho \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s} \right) = 1 \, ,$$

we can define $s(\boldsymbol{\theta}) > 0$ such that

$$\mathsf{E}_{H_0} \, \rho \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s} \right) > 1 - \delta, \qquad \forall s < s(\boldsymbol{\theta}).$$

Now we can proceed similarly as in the proof of Lemma 4.2 of Salibian-Barrera (2000) to find $s' > 0$ such that $f(s) > 1 - \delta$ for all $s < s'$. As $\delta > 0$ was arbitrary, we have $\lim_{s \to 0_+} f(s) = 1$. Thus, it follows that there exists $s_1 > 0$ such that

$$\inf_{H \in \mathcal{H}_{\varepsilon_0}} \inf_{\|\boldsymbol{\theta}\| \le K} \mathsf{E}_H \, \rho \left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T} \mathbf{X}_1}{s} \right) > b \, ,$$

for all $s < s_1$, which implies that

$$\inf_{H \in \mathcal{H}_{\varepsilon_0}} \inf_{\|\boldsymbol{\theta}\| \le K} \sigma(H, \boldsymbol{\theta}) \ge s_1 .$$

■

Lemma 2 *Suppose that assumptions* ***A.1-2*** *hold, and that the function* $h(s, \boldsymbol{\theta})$ *defined in* (3.4) *is continuous and that* $h(s, \boldsymbol{\theta}) < 0$ *for all* $s > 0$ *and* $\boldsymbol{\theta} \in \mathbb{R}^p$. *Let* $K > 0$ *be fixed. Then, for all* $\delta > 0$

$$\lim_{m\to\infty} \sup_{H\in\mathcal{H}_{\varepsilon_0}} P_H\left[\sup_{n\geq m}\sup_{\|\boldsymbol{\theta}\|\leq K} |\sigma(H_n, \boldsymbol{\theta}) - \sigma(H, \boldsymbol{\theta})| > \delta\right] = 0\,,$$

where $\sigma(H, \boldsymbol{\theta})$ *is defined in* (3.1) *and* H_n *is the empirical distribution function of the sample.*

Proof We have

$$\begin{aligned}
P_H&\left[\sup_{n\geq m}\sup_{\|\boldsymbol{\theta}\|\leq K} |\sigma(H_n, \boldsymbol{\theta}) - \sigma(H, \boldsymbol{\theta})| > \delta\right] \\
&\leq \sum_{n=m}^{\infty} P_H\left[\sup_{\|\boldsymbol{\theta}\|\leq K} |\sigma(H_n, \boldsymbol{\theta}) - \sigma(H, \boldsymbol{\theta})| > \delta\right] \\
&\leq \sum_{n=m}^{\infty} P_H\left[\sup_{\|\boldsymbol{\theta}\|\leq K} (\sigma(H_n, \boldsymbol{\theta}) - \sigma(H, \boldsymbol{\theta})) > \delta\right] \\
&\qquad + \sum_{n=m}^{\infty} P_H\left[\inf_{\|\boldsymbol{\theta}\|\leq K} (\sigma(H_n, \boldsymbol{\theta}) - \sigma(H, \boldsymbol{\theta})) < -\delta\right]. \qquad (8.1)
\end{aligned}$$

The event $[\sigma(H_n, \boldsymbol{\theta}) - \sigma(H, \boldsymbol{\theta}) > \delta]$ satisfies

$$\begin{aligned}
[\sigma(H_n, \boldsymbol{\theta}) - \sigma(H, \boldsymbol{\theta}) > \delta] &= [\sigma(H_n, \boldsymbol{\theta}) > \sigma(H, \boldsymbol{\theta}) + \delta] \\
&\subset \left[\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{Y_i - \boldsymbol{\theta}^{\mathsf{T}}\mathbf{X}_i}{\sigma(H, \boldsymbol{\theta}) + \delta}\right) > b\right] \\
= \left[\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{Y_i - \boldsymbol{\theta}^{\mathsf{T}}\mathbf{X}_i}{\sigma(H, \boldsymbol{\theta}) + \delta}\right) - \mathsf{E}_H\,\rho\left(\frac{Y_1 - \boldsymbol{\theta}^{\mathsf{T}}\mathbf{X}_1}{\sigma(H, \boldsymbol{\theta}) + \delta}\right)\right. &> \left. b - \mathsf{E}_H\,\rho\left(\frac{Y_1 - \boldsymbol{\theta}^{\mathsf{T}}\mathbf{X}_1}{\sigma(H, \boldsymbol{\theta}) + \delta}\right)\right].
\end{aligned}$$

Now let $m = \sup_{\|\boldsymbol{\theta}\|\leq K} \sup_{s_1\leq s\leq s_2+\delta} h(s, \boldsymbol{\theta})$, where s_1 and s_2 are given by Lemma 1. From the assumptions we know that $m < 0$. Hence, for every $\|\boldsymbol{\theta}\| \leq K$ and every

$H \in \mathcal{H}_{\varepsilon_0}$ the mean value theorem yields

$$
\begin{aligned}
b - \mathsf{E}_H\, \rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{\sigma(H,\boldsymbol{\theta}) + \delta}\right) &= \mathsf{E}_H\left[\rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{\sigma(H,\boldsymbol{\theta})}\right) - \rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{\sigma(H,\boldsymbol{\theta}) + \delta}\right)\right] \\
&\geq (1-\varepsilon_0)\, \mathsf{E}_{H_0}\left[\rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{\sigma(H,\boldsymbol{\theta})}\right) - \rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{\sigma(H,\boldsymbol{\theta}) + \delta}\right)\right] \\
&\geq (1-\varepsilon_0) \inf_{s\in[s_1,s_2]} \mathsf{E}_{H_0}\left[\rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right) - \rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s+\delta}\right)\right] \\
&\geq -(1-\varepsilon_0)\, m\, \delta > 0\,.
\end{aligned}
$$

Call $\Delta_0 = -(1-\varepsilon_0)\, m\, \delta$. We have

$$
\begin{aligned}
&P_H\left[\sup_{\|\boldsymbol{\theta}\|\leq K} \left(\sigma(H_n,\boldsymbol{\theta}) - \sigma(H,\boldsymbol{\theta})\right) > \delta\right] \\
&\leq P_H\left[\sup_{\|\boldsymbol{\theta}\|\leq K, s\in[s_1,s_2]} \left(\frac{1}{n}\sum_{i=1}^n \rho\left(\frac{Y_i - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{s+\delta}\right) - \mathsf{E}_H\, \rho\left(\frac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s+\delta}\right)\right) > \Delta_0\right] \\
&\leq \mathcal{P}(n) \exp\left\{-2\, n\, \Delta_0^2\right\}, \quad (8.2)
\end{aligned}
$$

where $\mathcal{P}(n)$ is a polynomial in n which depends only on the dimension of the vector $\mathbf{X}_1$. The last inequality can be justified as follows. The set of functions $\{(y + \boldsymbol{\theta}^\mathsf{T}\mathbf{x})/s \,:\, \boldsymbol{\theta} \in \mathbb{R}^p, s \in \mathbb{R}_+\}$ is, according to Lemma 2.6.15 (iii) of van der Vaart and Wellner (1996) [VW], a VC-class of functions. Using (ii) and (iv) of Lemma 2.6.18 of VW we see that the set of functions $\{|y + \boldsymbol{\theta}^\mathsf{T}\mathbf{x}|/s \,:\, \boldsymbol{\theta} \in \mathbb{R}^p, s \in \mathbb{R}_+\}$ is a VC-class as well. Finally, because the function ρ is monotone on $[0,\infty)$, the set of functions $\{\rho(|y + \boldsymbol{\theta}^\mathsf{T}\mathbf{x}|/s) : \boldsymbol{\theta} \in \mathbb{R}^p, s \in \mathbb{R}_+\}$, which is the same as $\{\rho((y + \boldsymbol{\theta}^\mathsf{T}\mathbf{x})/s) : \boldsymbol{\theta} \in \mathbb{R}^p, s \in \mathbb{R}_+\}$, is a VC-class (see Lemma 2.6.18 (viii) of VW). Moreover we are assuming that the function ρ takes its values in the interval $[0, 1]$. Thus, (8.2) follows from Theorem 2.14.9 of VW. Next, note that (8.2) implies that

$$
\sum_{n=m}^{\infty} P_H\left[\sup_{\|\boldsymbol{\theta}\|\leq K} \left(\sigma(H_n,\boldsymbol{\theta}) - \sigma(H,\boldsymbol{\theta})\right) > \delta\right] < \infty\,. \quad (8.3)
$$

A similar argument shows that

$$\sum_{n=m}^{\infty} P_H\left[\inf_{\|\boldsymbol{\theta}\|\le K}(\sigma(H_n,\boldsymbol{\theta})-\sigma(H,\boldsymbol{\theta}))<-\delta\right]<\infty\,. \tag{8.4}$$

Finally, (8.3), (8.4) and (8.1) prove the lemma. ■

Proof of theorem 1 Let $\delta>0$ be given and fix an arbitrary $\varepsilon>0$. Denote $S=\sup_{H\in\mathcal{H}_{\varepsilon_0}}\sigma(H,\mathbf{0})$. By Lemma 1 we know that $S<\infty$. As in the proof of Lemma 4.4. in Salibian-Barrera (2000), we can prove that

$$\lim_{K\to\infty}\mathsf{E}_{H_0}\left[\inf_{\|\boldsymbol{\theta}\|>K}\rho\left(\frac{Y_1-\boldsymbol{\theta}^{\mathsf{T}}\mathbf{X}_1}{S+\varepsilon}\right)\mathbb{I}(|Y_1|\le K)\right]=1.$$

It follows that there is a sufficiently large K such that

$$\eta=(1-\varepsilon_0)\,\mathsf{E}_{H_0}\left[\inf_{\|\boldsymbol{\theta}\|>K}\rho\left(\frac{Y_1-\boldsymbol{\theta}^{\mathsf{T}}\mathbf{X}_1}{S+\varepsilon}\right)\right]-b>0. \tag{8.5}$$

Moreover, for every $H\in\mathcal{H}_{\varepsilon_0}$ we have

$$\left[\inf_{\|\boldsymbol{\theta}\|>K}\mathsf{E}_H\,\rho\left(\frac{Y_1-\boldsymbol{\theta}^{\mathsf{T}}\mathbf{X}_1}{S+\varepsilon}\right)\right]\ge(1-\varepsilon_0)\,\mathsf{E}_{H_0}\left[\inf_{\|\boldsymbol{\theta}\|>K}\rho\left(\frac{Y_1-\boldsymbol{\theta}^{\mathsf{T}}\mathbf{X}_1}{S+\varepsilon}\right)\right]>b\,,$$

which implies that $\sigma(H,\boldsymbol{\theta})\ge S+\varepsilon\ge\sigma(H,\mathbf{0})+\varepsilon$ for every $\|\boldsymbol{\theta}\|>K$. Thus, we can conclude that

$$\sigma(H)=\inf_{\boldsymbol{\theta}\in\mathbb{R}^p}\sigma(H,\boldsymbol{\theta})=\inf_{\|\boldsymbol{\theta}\|\le K}\sigma(H,\boldsymbol{\theta}).$$

Next note that

$$\begin{aligned}&P_H\Big[|\hat{\sigma}_n-\sigma(H)|>\delta\Big]\\&\qquad\le P_H\left[\sup_{\|\boldsymbol{\theta}\|\le K}|\sigma(H_n,\boldsymbol{\theta})-\sigma(H,\boldsymbol{\theta})|>\delta\right]+P_H\left[\hat{\sigma}_n=\inf_{\|\boldsymbol{\theta}\|>K}\sigma(H_n,\boldsymbol{\theta})\right]\,.\end{aligned} \tag{8.6}$$

As in the proof of Lemma 2, we can show that the first term above goes to zero exponentially fast. For the second term we have

$$
\begin{aligned}
&P_H\left[\hat{\sigma}_n = \inf_{\|\boldsymbol{\theta}\|>K} \sigma(H_n,\boldsymbol{\theta})\right] \\
&\quad \le P_H\left[\inf_{\|\boldsymbol{\theta}\|>K} \sigma(H_n,\boldsymbol{\theta}) < \sigma(H_n,\mathbf{0})\right] = P_H\left[\inf_{\|\boldsymbol{\theta}\|>K} \sum_{i=1}^{n} \rho\left(\frac{Y_i-\boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\sigma(H_n,\mathbf{0})}\right) < b\right] \\
&\quad \le P_H\left[\inf_{\|\boldsymbol{\theta}\|>K} \sum_{i=1}^{n} \rho\left(\frac{Y_i-\boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\sigma(H_n,\mathbf{0})}\right) < b\,,\ |\sigma(H_n,\mathbf{0})-\sigma(H,\mathbf{0})| \le \varepsilon\right] \\
&\qquad + P_H\Big[|\sigma(H_n,\mathbf{0})-\sigma(H,\mathbf{0})| > \varepsilon\Big] \\
&\quad \le P_H\left[\inf_{\|\boldsymbol{\theta}\|>K} \sum_{i=1}^{n} \rho\left(\frac{Y_i-\boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\sigma(H,\mathbf{0})+\varepsilon}\right) < b\right] + P_H\Big[|\sigma(H_n,\mathbf{0})-\sigma(H,\mathbf{0})| > \varepsilon\Big].
\end{aligned}
$$

The second term in the last equation converges to zero exponentially fast (according to the proof Lemma 2), so we only need to look at the first term. Note that

$$
\begin{aligned}
&P_H\left[\inf_{\|\boldsymbol{\theta}\|>K} \sum_{i=1}^{n} \rho\left(\frac{Y_i-\boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\sigma(H,\mathbf{0})+\varepsilon}\right) < b\right] \\
&\quad \le P_H\Bigg[\frac{1}{n}\sum_{i=1}^{n} \inf_{\|\boldsymbol{\theta}\|>K} \rho\left(\frac{Y_i-\boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\sigma(H,\mathbf{0})+\varepsilon}\right) - \mathsf{E}_H \inf_{\|\boldsymbol{\theta}\|>K} \rho\left(\frac{Y_1-\boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{\sigma(H,\mathbf{0})+\varepsilon}\right) \\
&\qquad < b - \mathsf{E}_H \inf_{\|\boldsymbol{\theta}\|>K} \rho\left(\frac{Y_1-\boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{\sigma(H,\mathbf{0})+\varepsilon}\right)\Bigg] \\
&\quad \overset{(8.5)}{\le} P_H\left[\frac{1}{n}\sum_{i=1}^{n} \inf_{\|\boldsymbol{\theta}\|>K} \rho\left(\frac{Y_i-\boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\sigma(H,\mathbf{0})+\varepsilon}\right) - \mathsf{E}_H \inf_{\|\boldsymbol{\theta}\|>K} \rho\left(\frac{Y_1-\boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{\sigma(H,\mathbf{0})+\varepsilon}\right) < -\eta\right] \\
&\quad \le 2\exp(-2n\eta^2)\,, \qquad (8.7)
\end{aligned}
$$

where the last inequality follows from Hoeffding's inequality. We see that both probabilities in (8.6) go to zero exponentially fast and thus the same holds for $P_H\left[|\hat{\sigma}_n - \sigma(H)| > \delta\right]$. Now the theorem follows from the standard inequality

$$
P_H\left[\sup_{n\ge m} |\hat{\sigma}_n - \sigma(H)| > \delta\right] \le \sum_{n=m}^{\infty} P_H\left[|\hat{\sigma}_n - \sigma(H)| > \delta\right].
$$

■

Proof of theorem 2 Fix $\delta > 0$ and let $\tilde{\varepsilon} = \tilde{\varepsilon}(\delta) > 0$ be as in assumption **U.3**. Next find $\delta_0 > 0$ such that assumption **U.1** holds with $\frac{\tilde{\varepsilon}}{2}$. We have

$$
\begin{aligned}
P_H\left[|\tilde{\boldsymbol{\beta}}(H_n) - \tilde{\boldsymbol{\beta}}(H)| > \delta\right] &\leq P_H\left[\inf_{\|\boldsymbol{\theta}-\tilde{\boldsymbol{\beta}}(H)\|>\delta} \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{Y_i - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\hat{\sigma}_n}\right) \leq b\right] \\
\leq P_H\left[\inf_{\|\boldsymbol{\theta}-\tilde{\boldsymbol{\beta}}(H)\|>\delta} \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{Y_i - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\hat{\sigma}_n}\right) \leq b\ ,\ |\hat{\sigma}_n - \sigma(H)| \leq \delta\right] &+ P_H\left[|\hat{\sigma}_n - \sigma(H)| > \delta_0\right] \\
\leq P_H\left[\inf_{\|\boldsymbol{\theta}-\tilde{\boldsymbol{\beta}}(H)\|>\delta} \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{Y_i - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\sigma(H)+\delta_0}\right) \leq b\right] &+ P_H\left[|\hat{\sigma}_n - \sigma(H)| > \delta_0\right] \\
\leq P_H\Bigg[\inf_{\|\boldsymbol{\theta}-\tilde{\boldsymbol{\beta}}(H)\|>\delta} \left(\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{Y_i - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_i}{\sigma(H)+\delta_0}\right) - g(H,\boldsymbol{\theta},\sigma(H)+\delta_0)\right) \\
\leq b - \inf_{\|\boldsymbol{\theta}-\tilde{\boldsymbol{\beta}}(H)\|>\delta} g(H,\boldsymbol{\theta},\sigma(H)+\delta_0)\Bigg] &+ P_H\left[|\hat{\sigma}_n - \sigma(H)| > \delta_0\right] .
\end{aligned}
$$

Theorem 1 shows that the second term goes to zero exponentially fast. The empirical process approach used in the proof of Lemma 2 we can obtain a similar bound as in (8.2) if we show that

$$
\eta = b - \inf_{\|\boldsymbol{\theta}-\tilde{\boldsymbol{\beta}}(H)\|>\delta} g(H,\boldsymbol{\theta},\sigma(H)+\delta_0) < 0 .
$$

We have

$$
\begin{aligned}
&\inf_{\|\boldsymbol{\theta}-\tilde{\boldsymbol{\beta}}(H)\|>\delta} g(H,\boldsymbol{\theta},\sigma(H)+\delta_0) \\
\geq &\inf_{\|\boldsymbol{\theta}-\tilde{\boldsymbol{\beta}}(H)\|>\delta} \left[g(H,\boldsymbol{\theta},\sigma(H)+\delta_0) - g(H,\boldsymbol{\theta},\sigma(H)+\delta_0)\right] + \inf_{\|\boldsymbol{\theta}-\tilde{\boldsymbol{\beta}}(H)\|>\delta} g(H,\boldsymbol{\theta},\sigma(H)+\delta_0) \\
&\geq -\frac{\tilde{\varepsilon}}{2} + b + \tilde{\varepsilon} = b + \frac{\tilde{\varepsilon}}{2} ,
\end{aligned}
$$

where the last inequality follows from assumptions **U.1** and **U.3**. Thus $\eta \leq -\frac{\tilde{\varepsilon}}{2} < 0$ and this completes the proof. ■

In what follows we will need the following definitions.

Definition 3 - **Uniform big O in probability**: *Let* a_n, $n \geq 1$, *be a sequence of real numbers and let* X_n, $n \geq 1$, *be a sequence of random variables. We say that* $X_n = UO_P(a_n)$ *over the set of distribution functions* $\mathcal{H}_{\epsilon_0}$ *if*

$$\lim_{k\to\infty} \sup_{F\in\mathcal{H}_{\epsilon_0}} \lim_{n\to\infty} P_F\left[\left|\frac{X_n}{a_n}\right| > k \right] = 0.$$

Definition 4 - **Uniform small o in probability**: *Let* a_n, $n \geq 1$, *be a sequence of real numbers and let* X_n, $n \geq 1$, *be a sequence of random variables. We say that* $X_n = Uo_P(a_n)$ *over the set of distribution functions* $\mathcal{H}_{\epsilon_0}$ *if* $\forall\,\delta > 0$

$$\lim_{n\to\infty} \sup_{F\in\mathcal{H}_{\epsilon_0}} P_F\left[\left|\frac{X_n}{a_n}\right| > \delta \right] = 0.$$

With the above definitions we can show that these "uniform little o", "uniform big O" and "uniform asymptotic distribution" behave similarly to their "non-uniform" counterparts. This is made more precise in the following remark. In particular, if $a_n = Uo_P(1)$ and X_n is UAN then $X_n + a_n$ is UAN.

Remark 4 - **Properties of** $UO_p(1)$ **and** $Uo_p(1)$ - *In what follows* a_n, b_n *and* X_n, $n \in \mathbb{N}$ *denote sequences of random variables. It is easy to see that the following properties hold. Proofs of these results can be found in Salibian-Barrera (2000, Chapter 2).*

Property 1 - *if* $a_n = UO_P(1)$ *and* $b_n = Uo_P(1)$, *then* $a_n\, b_n = Uo_P(1)$;

Property 2 - *if* $a_n = UO_P(1)$ *and there exists* $b \neq 0$ *with* $b_n - b = Uo_P(1)$, *then* $a_n/\,b_n = a_n/\,b + Uo_P(1)$;

Property 3 - *if* $a_n = UO_P(1)$ *and there exists* b *with* $b_n - b = Uo_P(1)$, *then* $a_n\, b_n = a_n\, b + Uo_P(1)$;

Lemma 3 *Assume that **A.1-3**, **X.1** and **N.1** hold and let* $\{\varepsilon_n\}_{n\geq 1} \subset \mathbb{R}$ *and* $\{\boldsymbol{\delta}_n\}_{n\geq 1} \subset \mathbb{R}$ *be two sequences converging to zero. Then, uniformly over* $\mathcal{H}_{\varepsilon_0}$,

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i-(\tilde{\boldsymbol{\theta}}(H)+\boldsymbol{\delta}_n)^{\mathsf{T}}\mathbf{X}_i}{\sigma(H)+\varepsilon_n}\right)$$
$$=\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i-\tilde{\boldsymbol{\theta}}(H)^{\mathsf{T}}\mathbf{X}_i}{\sigma(H)}\right)-\left[\mathbf{a}_H+o(1)+Uo_P(1)\right]^{\mathsf{T}}\boldsymbol{\delta}_n-\left[b_H+o(1)+Uo_P(1)\right]\varepsilon_n\,, \tag{8.8}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\rho'\left(\frac{Y_i-(\tilde{\boldsymbol{\theta}}(H)+\boldsymbol{\delta}_n)^{\mathsf{T}}\mathbf{X}_i}{\sigma(H)+\varepsilon_n}\right)\mathbf{X}_i$$
$$=\frac{1}{n}\sum_{i=1}^{n}\rho'\left(\frac{Y_i-\tilde{\boldsymbol{\theta}}(H)^{\mathsf{T}}\mathbf{X}_i}{\sigma(H)}\right)\mathbf{X}_i-\left[\mathbf{C}_H+o(1)+Uo_P(1)\right]\boldsymbol{\delta}_n-\left[\mathbf{d}_H+o(1)+Uo_P(1)\right]\varepsilon_n\,. \tag{8.9}$$

Proof We will prove (8.8), the proof of expansion (8.9) being completely analogous. According to the mean value theorem there exists a point $(\boldsymbol{\theta}_n^*, \sigma_n^*)$ which lies in the interior of the line segment connecting the points $(\tilde{\boldsymbol{\theta}}(H)+\boldsymbol{\delta}_n, \sigma_H+\varepsilon_n)$ and $(\tilde{\boldsymbol{\theta}}(H), \sigma_H)$ such that

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i-(\tilde{\boldsymbol{\theta}}(H)+\boldsymbol{\delta}_n)^{\mathsf{T}}\mathbf{X}_i}{\sigma(H)+\varepsilon_n}\right)-\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i-\tilde{\boldsymbol{\theta}}(H)^{\mathsf{T}}X_i}{\sigma(H)}\right)$$
$$=-\left[\frac{1}{n}\sum_{i=1}^{n}\rho'\left(\frac{Y_i-(\boldsymbol{\theta}_n^*)^{\mathsf{T}}\mathbf{X}_i}{\sigma_n^*}\right)\frac{\mathbf{X}_i}{\sigma_n^*}\right]^{\mathsf{T}}\boldsymbol{\delta}_n-\left[\frac{1}{n}\sum_{i=1}^{n}\rho'\left(\frac{Y_i-(\boldsymbol{\theta}_n^*)^{\mathsf{T}}\mathbf{X}_i}{\sigma_n^*}\right)\frac{Y_i-(\boldsymbol{\theta}_n^*)^{\mathsf{T}}\mathbf{X}_i}{(\sigma_n^*)^2}\right]\varepsilon_n$$
$$=-\mathbf{a}_1^{H_n}(\boldsymbol{\theta}_n^*,\sigma_n^*)^{\mathsf{T}}\boldsymbol{\delta}_n-\mathbf{a}_2^{H_n}(\boldsymbol{\theta}_n^*,\sigma_n^*)\,\varepsilon_n\,.$$

We will first show that

$$\mathbf{a}_1^{H_n}(\boldsymbol{\theta}_n^*,\sigma_n^*)=\mathbf{a}_1^{H}(\boldsymbol{\theta}_n^*,\sigma_n^*)+Uo_p(1)\,, \tag{8.10}$$

and

$$\mathbf{a}_2^{H_n}(\boldsymbol{\theta}_n^*,\sigma_n^*)=\mathbf{a}_2^{H}(\boldsymbol{\theta}_n^*,\sigma_n^*)+Uo_p(1)\,. \tag{8.11}$$

It is obvious that it suffices to work component-wise. In what follows let X_i^j denote the j-th component of the vector $\mathbf{X}_i$. So fix $\varepsilon > 0$ (without loss of generality assume that $\varepsilon < 1$) and let $j \in \{1, \ldots, p\}$. Since the set of functions

$$\mathcal{G}_1 = \{g_{\boldsymbol{\theta},s}(y, \mathbf{x}) = \mathbb{I}\{y - \boldsymbol{\theta}^\mathsf{T}\mathbf{x} \; > s\}, \boldsymbol{\theta} \in \mathbb{R}^p, s \; \in \mathbb{R}_+\}$$

is a VC-class with envelope $G = 1$, which trivially satisfies

$$\sup_{H \in \mathcal{H}_{\varepsilon_0}} \mathsf{E}_H \, G < \infty, \tag{8.12}$$

there exists $C > 0$ and a polynomial $\mathcal{P}_n$ in n (with coefficients not depending on the distribution H or the sample size n) such that

$$P_H \left[\sup_{g \in \mathcal{G}_1} \left| \frac{1}{n} \sum_{i=1}^{n} g(Y_i, \mathbf{X}_i) - \mathsf{E}_H \, g(Y_1, \mathbf{X}_1) \right| > \varepsilon \right] \leq \mathcal{P}_n \exp(-\, C \, n \, \varepsilon). \tag{8.13}$$

As X_i^j is integrable (by assumption **X.2**), the set of functions

$$\mathcal{G}_2 = \{x^j \, g_{\boldsymbol{\theta},s}(y, \mathbf{x}) = x^j \, \mathbb{I}\{y - \boldsymbol{\theta}^\mathsf{T}\mathbf{x} \; > s\}, \boldsymbol{\theta} \in \mathbb{R}^p, \, s \in \mathbb{R}_+\} = x^j \, \mathcal{G}_1$$

is a VC-class as well with envelope $G = |X_i^j|$ (which is integrable) and so an inequality like (8.13) (with different $C > 0$ and $\mathcal{P}_n$) holds for the set of functions $\mathcal{G}_2$ as well.

Let

$$K = \sup_{H \in \mathcal{H}_{\varepsilon_0}} \max_{j=1,\ldots,p} \mathsf{E}_H \, |X_1^j| \; (< \infty) \, .$$

According to assumption **A.3** we can approximate $\rho'(u)$ with a function $\rho_k'(u) = \sum_{j=1}^{k} a_j \mathbb{I}\{u > b_j\}$ such that

$$\sup_{u \in \mathbb{R}} \left| \rho'(u) - \rho_k'(u) \right| < \frac{\varepsilon^2 \, s_1}{4K}.$$

Then, uniformly in $\mathcal{H}_{\varepsilon_0}$, we have

$$P_H\left[\sup_{\boldsymbol{\theta}\in\mathbb{R}^p, s\in[s_1,s_2]}\left|\frac{1}{n}\sum_{i=1}^n \rho'\left(\frac{Y_i-\boldsymbol{\theta}^\top\mathbf{X}_i}{s}\right)\frac{\mathbf{X}_i}{s} - \mathsf{E}\,\rho'\left(\frac{Y_1-\boldsymbol{\theta}^\top\mathbf{X}_1}{s}\right)\frac{\mathbf{X}_1}{s}\right| > \varepsilon\right]$$
$$\le P_H\left[\sup_{\boldsymbol{\theta}\in\mathbb{R}^p, s\in[s_1,s_2]}\left|\frac{1}{n}\sum_{i=1}^n \rho_k'\left(\frac{Y_i-\boldsymbol{\theta}^\top\mathbf{X}_i}{s}\right)\frac{\mathbf{X}_i}{s} - \mathsf{E}\,\rho_k'\left(\frac{Y_1-\boldsymbol{\theta}^\top\mathbf{X}_1}{s}\right)\frac{\mathbf{X}_1}{s}\right| > \frac{\varepsilon}{2}\right]$$
$$+ P_H\left[\frac{\varepsilon^2}{4K}\frac{1}{n}\sum_{i=1}^n |X_i^j| + \frac{\varepsilon^2}{4} > \frac{\varepsilon}{2}\right] = A_n + B_n.$$

Using the exponential inequality (8.13) we obtain

$$A_n \le \mathcal{P}_n \exp(-\,C\,n\,\frac{\varepsilon}{2})\,.$$

Finally, the Markov inequality yields

$$B_n \le P_H\left[\frac{\varepsilon^2}{4K}\frac{1}{n}\sum_{i=1}^n |X_i^j| > \frac{\varepsilon}{4}\right] \le \frac{4}{\varepsilon}\,\mathsf{E}_H\,\frac{\varepsilon^2}{4K}|X_1^j| \le \varepsilon\,.$$

So (8.10) is proved, and (8.11) can be proved in a similar way. To finish the proof of the lemma note that using assumption **N.1** we obtain

$$\mathsf{E}_H\,\rho'\left(\frac{Y_1-(\boldsymbol{\theta}_n^*)^\top\mathbf{X}_1}{\sigma_n^*}\right)\frac{\mathbf{X}_1}{\sigma_n^*} = \mathsf{E}_H\,\rho'\left(\frac{Y_1-\tilde{\boldsymbol{\theta}}(H)^\top\mathbf{X}_1}{\sigma(H)}\right)\frac{\mathbf{X}_1}{\sigma(H)} + o(1)\,,$$

and

$$\mathsf{E}_H\,\rho'\left(\frac{Y_1-(\boldsymbol{\theta}_n^*)^\top\mathbf{X}_1}{\sigma_n^*}\right)\frac{Y_1-(\boldsymbol{\theta}_n^*)^\top\mathbf{X}_1}{(\sigma_n^*)^2} = \mathsf{E}_H\,\rho'\left(\frac{Y_1-\tilde{\boldsymbol{\theta}}(H)^\top\mathbf{X}_1}{\sigma(H)}\right)\frac{Y_1-\tilde{\boldsymbol{\theta}}(H)^\top\mathbf{X}_1}{\sigma(H)^2}+o(1)\,,$$

because $\varepsilon_n \to 0$ and $\boldsymbol{\delta}_n \to \mathbf{0}$ implies $\boldsymbol{\theta}_n^* \to \tilde{\boldsymbol{\theta}}(H)$ and $\sigma_n^* \to \sigma(H)$. ■

Proof of theorem 3 Since the assumptions of Theorems 1 and 2 are satisfied, we have

$$\tilde{\boldsymbol{\theta}}(H_n) = \tilde{\boldsymbol{\theta}}(H) + U o_P(1)\,,$$

and

$$\hat{\sigma}_n = \sigma(H) + U o_P(1)\,.$$

Substituting $\boldsymbol{\delta}_n \mapsto \left[\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}(H)\right]$ and $\varepsilon_n \mapsto [\hat{\sigma}_n - \sigma(H)]$ into (8.8) and (8.9) and noting that from the defition of S-estimators we have $\mathbf{a}_H = \mathbf{0}$, we obtain

$$b = \frac{1}{n}\sum_{i=1}^{n} \rho(\tilde{u}_i(H)) - Uo_P(1)^{\mathsf{T}} \left(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}(H)\right) - \left[b_H + Uo_P(1)\right]\left(\hat{\sigma}_n - \sigma(H)\right), \quad (8.14)$$

and

$$\mathbf{0} = \frac{1}{n}\sum_{i=1}^{n} \rho'(\tilde{u}_i(H))\mathbf{X}_i - \left[\mathbf{C}_H + Uo_P(1)\right]\left(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}(H)\right) - \left[\mathbf{d}_H + Uo_P(1)\right]\left(\hat{\sigma}_n - \sigma(H)\right). \quad (8.15)$$

The idea of the proof is as follows. We first use (8.15) to find an expression for $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}(H))$. We then substitute this expression into (8.14) and arrive at (5.1). Finally, we substitute this result back into (8.15) to obtain (5.2). The only difficulty is to verify that all our steps hold uniformly on $\mathcal{H}_{\varepsilon_0}$.

First note that assumptions **A.3** and **X.2** imply

$$\sup_{H \in \mathcal{H}_{\varepsilon_0}} \max\{|b_H|, \|\mathbf{C}_H\|, \|\mathbf{d}_H\|\} < \infty. \quad (8.16)$$

(This can be also derived from assumption **N.1**). Furthermore, by **N.2** we know that the smallest eigenvalue of the matrix $\mathbf{C}_H$ is bounded away from zero uniformly in $\mathcal{H}_{\varepsilon_0}$. This together with equation (8.16) yields $\sup_{H\in\mathcal{H}_{\varepsilon_0}} \|\mathbf{C}_H^{-1}\| < \infty$ and this implies

$$\left[\mathbf{C}_H + Uo_P(1)\right]^{-1} = \mathbf{C}_H^{-1} + Uo_P(1).$$

Thus,

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}(H_n) - \tilde{\boldsymbol{\theta}}(H)) = \frac{\mathbf{C}_H^{-1} + Uo_p(1)}{\sqrt{n}} \sum_{i=1}^{n} \rho'(\tilde{u}_i(H))\mathbf{X}_i - \sqrt{n}(\hat{\sigma}_n - \sigma(H))\left[\mathbf{d}_H + Uo_P(1)\right]. \quad (8.17)$$

It is easy to see that $\sum_{i=1}^{n} \rho'(\tilde{u}_i(H))\mathbf{X}_i/\sqrt{n} = UO_p(1)$, and thus substituting (8.17) into equation (8.14) yields

$$\sqrt{n}(\hat{\sigma}_n - \sigma(H))\,[b_H + Uo_P(1)]$$

$$+\mathbf{C}_H^{-1}\left\{UO_P(1) + [\mathbf{d}_H + Uo_P(1)]\,\sqrt{n}(\hat{\sigma}_n - \sigma(H))\right\}Uo_P(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\rho(\tilde{u}_i(H)) - b]\,. \tag{8.18}$$

After some reorganization and using (8.16) we get

$$\sqrt{n}(\hat{\sigma}_n - \sigma(H))\,[b_H + Uo_P(1)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\rho(\tilde{u}_i(H)) - b] + Uo_P(1)\,.$$

Again, it is easy to verify that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\rho(\tilde{u}_i(H)) - b] = UO_p(1)\,,$$

and from the proof of Lemma 2 we know that $\inf_{H\in\mathcal{H}_{\varepsilon_0}} b_H > 0$, so we can divide both sides of (8.18) by $(b_H + Uo_p(1))$ to arrive at (5.1). Substituting the last expansion for $\sqrt{n}(\hat{\sigma}_n - \sigma(H))$ into (8.17) yields the second part of the theorem.

Remark 5 *Before we start the proof of Corollary 1 we need to verify*

$$\mathbf{\Sigma}_H^{-1/2} Uo_p(1) = Uo_p(1)\,. \tag{8.19}$$

Since it is easy to verify that all the quantities in formula (5.3) *are uniformly finite, it is sufficient to show that the smallest eigenvalue of the matrix* $\mathbf{\Sigma}_H$ *is bounded away from zero uniformly for* $H \in \mathcal{H}_{\varepsilon_0}$*. For each* $\boldsymbol{\theta} \in \mathbb{R}_p$*,* $s \in \mathbb{R}_+$ *and* $\mathbf{t} \in \mathbb{R}_p$ *let*

$$\mathbf{Z}(\boldsymbol{\theta}, s, \mathbf{t}) = \rho'\left(\tfrac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right)\tfrac{\mathbf{X}_1}{s} - \left(\rho\left(\tfrac{Y_1 - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}_1}{s}\right) - b\right)\mathbf{t}\,.$$

We need to show that there exists $\eta > 0$ *such that*

$$\inf_{H\in\mathcal{H}_{\varepsilon_0}}\inf_{\|\boldsymbol{\lambda}\|=1}\mathrm{var}_H\left\{\boldsymbol{\lambda}^\mathsf{T}\mathbf{Z}\left(\tilde{\boldsymbol{\theta}}(H), \sigma(H), \frac{\mathbf{d}_H}{b_H}\right)\right\} \geq \eta\,.$$

Note that we have dropped the matrix $\mathbf{C}_H$ *since we know that*

$$\sup_{H\in\mathcal{H}_{\varepsilon_0}} \{\|\mathbf{C}_H\|, \|\mathbf{C}_H^{-1}\|\} < \infty\,.$$

Consider

$$\begin{aligned}
\mathrm{var}_H\left\{\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}\left(\tilde{\boldsymbol{\theta}}(H),\sigma(H),\frac{\mathbf{d}_H}{b_H}\right)\right\} &= \mathsf{E}_H\left[\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}\left(\tilde{\boldsymbol{\theta}}(H),\sigma(H),\frac{\mathbf{d}_H}{b_H}\right)\right]^2\\
&\geq (1-\varepsilon_0)\,\mathsf{E}_{H_0}\left[\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}\left(\tilde{\boldsymbol{\theta}}(H),\sigma(H),\frac{\mathbf{d}_H}{b_H}\right)\right]^2\\
&\geq (1-\varepsilon_0)\inf_{\|\boldsymbol{\theta}\|\leq K_1}\ \inf_{s\in[s_1,s_2]}\ \inf_{\|\mathbf{t}\|\leq K_2}\ \mathsf{E}_{H_0}\left[\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}(\boldsymbol{\theta},s,\mathbf{t})\right]^2, \quad (8.20)
\end{aligned}$$

where K_1 *is taken from the proof of Theorem 2,* s_1, s_2 *are from Lemma 1 and* K_2 *is chosen such that* $\sup_{H\in\mathcal{H}_{\varepsilon_0}}\|\mathbf{d}_H/b_H\| \leq K_2 < \infty$*. Note that the lower bound in* (8.20) *holds uniformly over* $\mathcal{H}_{\varepsilon_0}$*. We will now assume that*

$$\inf_{\|\boldsymbol{\lambda}\|=1}\ \inf_{\|\boldsymbol{\theta}\|\leq K_1}\ \inf_{s\in[s_1,s_2]}\ \inf_{\|\mathbf{t}\|\leq K_2}\ \mathsf{E}_{H_0}\left[\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{Z}(\boldsymbol{\theta},s,\mathbf{t})\right]^2 = 0\,, \qquad (8.21)$$

and show that this leads to a contradiction. If (8.21) *holds, then for each* $n\in\mathbb{N}$ *there exists a foursome* $(\boldsymbol{\lambda}_n,\boldsymbol{\theta}_n,s_n,\mathbf{t}_n)$*, such that*

$$(\boldsymbol{\lambda}_n,\boldsymbol{\theta}_n,s_n,\mathbf{t}_n)\ \in\ \{(\boldsymbol{\lambda},\boldsymbol{\theta},s,\mathbf{t}) : \|\boldsymbol{\lambda}\|=1, \|\boldsymbol{\theta}\|\leq K_1, s\in[s_1,s_2], \|\mathbf{t}\|\leq K_2\} = \mathcal{K}\,,$$

and satisfies

$$\mathsf{E}_{H_0}\left[\boldsymbol{\lambda}_n^{\mathsf{T}}\mathbf{Z}(\boldsymbol{\theta}_n,s_n,\mathbf{t}_n)\right]^2 \leq \frac{1}{n}\,.$$

Since the set $\mathcal{K}$ *is compact, there exists a foursome* $(\boldsymbol{\lambda}_*,\boldsymbol{\theta}_*,s_*,\mathbf{t}_*)\in\mathcal{K}$*, for which*

$$\mathsf{E}_{H_0}\left[\boldsymbol{\lambda}_*^{\mathsf{T}}\mathbf{Z}(\boldsymbol{\theta}_*,s_*,\mathbf{t}_*)\right]^2 = 0.$$

But this further implies that the random variable $\boldsymbol{\lambda}_*^{\mathsf{T}}\mathbf{Z}(\boldsymbol{\theta}_*,s_*,\mathbf{t}_*)$ *equals zero almost surely* H_0*. However, this is not possible because, under the central model* H_0*,* u_1 *is independent from* $\mathbf{X}_1$*, which has a nonsingular distribution.*

Using (8.19) *we can rewrite expansion* (5.2) *as*

$$\mathbf{\Sigma}_H^{-1/2}\sqrt{n}(\tilde{\boldsymbol{\theta}}_n-\tilde{\boldsymbol{\theta}}(H))$$
$$=\mathbf{\Sigma}_H^{-1/2}\mathbf{C}_H^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\rho'(\tilde{u}_i(H))\,\mathbf{X}_i-\mathbf{\Sigma}_H^{-1/2}\mathbf{C}_H^{-1}\mathbf{d}_H\sqrt{n}[\hat{\sigma}_n-\sigma(H)]+Uo_P(1)\,. \tag{8.22}$$

Proof of Corollary 1 Let μ_n^H denote the distribution of the random vector $\sum_{i=1}^n \mathbf{Z}_i/\sqrt{n}$, where

$$\mathbf{Z}_i=\mathbf{\Sigma}_H^{-1/2}\left[\mathbf{C}_H^{-1}\rho'(\tilde{u}_i(H))\,\frac{\mathbf{X}_i}{\sigma(H)}-\mathbf{C}_H^{-1}\,\frac{\mathbf{d}_H}{b_H}\,(\rho(u_i(H))-b)\right]\,. \tag{8.23}$$

Since

$$d_P(\nu_n^H,\mu)\leq d_P(\nu_n^H,\mu_n^H)+d_P(\mu_n^H,\mu)\,, \tag{8.24}$$

we only need to show that both terms on the right-hand side of the equation (8.24) are sufficiently small. Fix $\varepsilon>0$ (without loss of generality we can assume $0<\varepsilon<1$). Then, by Definition 4 ($Uo_p(1)$) we can find n_0 such that for all $n\geq n_0$ the remainder term in (8.22) satisfies $\sup_{H\in\mathcal{H}_{\varepsilon_0}}P_H(\|Uo_p(1)\|\geq\varepsilon)<\varepsilon$. Then, for every Borel-measurable set B we have

$$\nu_n^H(B)=P_H\left[\mathbf{\Sigma}_H^{-1/2}\sqrt{n}(\tilde{\boldsymbol{\theta}}_n-\tilde{\boldsymbol{\theta}}(H))\in B\right]=P_H\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{Z}_i+Uo_p(1)\in B\right]$$
$$\leq P_H\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{Z}_i+Uo_p(1)\in B,\,\|Uo_p(1)\|<\varepsilon\right]+P_H\left[\|Uo_p(1)\|\geq\varepsilon\right]$$
$$\leq P_H\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{Z}_i\in B^{\varepsilon}\right]+\varepsilon=\mu_n^H(B^{\varepsilon})+\varepsilon, \tag{8.25}$$

which implies that

$$\sup_{H\in\mathcal{H}_{\varepsilon_0}}d_P(\nu_n^H,\mu_n^H)<\varepsilon\,. \tag{8.26}$$

To bound the second term in (8.24) we use Proposition A.5.2 of van der Vaart and Wellner (1996), which yields

$$\sup_{H\in\mathcal{H}_{\varepsilon_0}} d_P(\mu_n^H,\mu) \le 2\max\{\varepsilon^{-2}g(\varepsilon\sqrt{n}),\varepsilon\} + 4^{1/3}\, g(\varepsilon\sqrt{n})^{1/3} + (p\, g(0)\, \varepsilon)^{1/4}\, C\left(1+\left|\log\frac{\varepsilon\, g(0)}{p}\right|^{1/2}\right), \tag{8.27}$$

where C is a constant and

$$g(x) = \sup_{H\in\mathcal{H}_{\varepsilon_0}} \mathsf{E}_H\, \|\mathbf{Z}_1\|^2\, \mathbb{I}\{\|\mathbf{Z}_1\| > x\}\,.$$

Since

$$\sup_{H\in\mathcal{H}_{\varepsilon_0}} \|\mathbf{C}_H^{-1}\,\frac{\mathbf{d}_H}{b_H}\,(\rho(u_i(H)) - b)\| < \infty\,,$$

and

$$\sup_{H\in\mathcal{H}_{\varepsilon_0}} \|\mathbf{\Sigma}_H^{-1/2}\| < \infty\,,$$

assumption (5.4) implies $\lim_{n\to\infty} g(\varepsilon\sqrt{n}) = 0$ and so we can make (8.27) arbitrarily small by taking n sufficiently large and ε sufficiently small. This concludes the proof.

■

Proof of Remark 3 We will use the same notation as in Corollary 1. The uniform convergence in Prokhorov metric means that for all $\varepsilon > 0$ there exists n_0 such that for all $n > n_0$ and for all Borel-measurable sets B we have

$$\nu_n^H(B) \le \mu(B^\varepsilon) + \varepsilon \quad \text{and} \quad \mu(B) \le \nu_n^H(B^\varepsilon) + \varepsilon\,. \tag{8.28}$$

Define the set of one-sided intervals $\mathcal{I} = \{(-\infty, \mathbf{x}] : \mathbf{x}\in\mathbb{R}^p\}$. Clearly the elements of $\mathcal{I}$ are Borel measurable.

Recall that, without loss of generality, we are using the maximum norm. For each $\varepsilon \in \mathbb{R}$ define

$$I^{\varepsilon} = (-\infty, \mathbf{x}]^{\varepsilon} = (-\infty, \mathbf{x} + \varepsilon] \ \in \mathcal{I}\,.$$

Note that when $\varepsilon > 0$ this definition is consistent with the definition of I^{ε} used in the definition of Prokhorov metric. Now the 'Prokhorov bounds' (8.28) imply that for all $n > n_0$ and for all $I \in \mathcal{I}$

$$\nu_n^H(I) \le \mu(I^{\varepsilon}) + \varepsilon \le \mu(I) + \left(1 + \frac{p}{(2\,\pi)^{p/2}}\right)\,\varepsilon \le \mu(I) + 2\,\varepsilon\,,$$

where μ denotes a p-variate standard normal measure. Similarly we obtain

$$\nu_n^H(I) \ge \mu^H(I^{-\varepsilon}) - \varepsilon \ge \mu(I) - \left(1 + \frac{p}{(2\,\pi)^{p/2}}\right)\,\varepsilon \ge \mu(I) - 2\,\varepsilon\,.$$

Adding the last two equations shows that for any $\varepsilon > 0$ and for all sufficiently large n we have

$$\sup_{H \in \mathcal{H}_{\varepsilon_0}} \sup_{I \in \mathcal{I}} |\nu_n^H(I) - \mu(I)| \le 2\,\varepsilon\,,$$

which verifies (5.5).

References

[1] Beaton, A.E. and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, **16**, 147-185.

[2] Berrendero, J.R. and Zamar, R.H. (1999). Global robustness of location and dispersion estimates. *Statistics and Probability Letters*, **44**, 63-72.

[3] Berrendero, J.R. and Zamar, R.H. (2006). A note on the uniform asymptotic normality of location M-estimates. *Metrika*, **63**, 55-69.

[4] Bickel, P.J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association*, **70**, 428-434.

[5] Carroll, R. J. (1978). On almost sure expansion for M-estimates. *The Annals of Statistics*, **6**, 314-318.

[6] Carroll, R.J. (1979). On estimating variances of robust estimators when the errors are asymmetric. *Journal of the American Statistical Association*, **74**, 674-679.

[7] Clarke, B.R. (1980). Robust estimation; limit theorems and their applications. Unpublished PhD Thesis. Australian National University.

[8] Clarke, B.R. (1986). Nonsmooth analysis and Fréchet differentiability of M-functionals. *Probability Theory and Related Fields*, **73**, 197-209.

[9] Clarke, B.R. (2000). A remark on robustness and weak continuity of M-estimators. *Journal of the Australian Mathematical Society (Ser. A)*, **68**, 411-418.

[10] Croux, C., Dhaene, G., and Hoorelbeke, D. (2003). Robust Standard Errors for Robust Estimators. Research Report, Department of Applied Economics, K.U. Leuven, Leuven, Belgium.

[11] Croux, C., Rousseeuw, P.J. and Hossjer, O. (1994). Generalized S-estimators. *Journal of the American Statistical Association*, **89**, 1271-1281.

[12] Davies, L. (1990). The asymptotics of S-estimators in the linear regression. *Ann. Statist.*, 18:1651–1675.

[13] Davies, L. (1993). Aspects of robust linear regression. *The Annals of Statistics*, **21**, 1843-1899.

[14] Davies, P.L. (1998). On locally uniformly linearizable high breakdown location and scale functionals. *The Annals of Statistics*, **26**, 1103-1125.

[15] Hampel, F.R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, **42**, 1887-1896.

[16] Heathcote, C.R. and Silvapulle, M.J. (1981). Minimum mean squared estimation of location and scale parameters under misspecification of the model. *Biometrika*, **68**, 501-514.

[17] Huber, P.J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, **35**, 73-101.

[18] Huber, P.J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, California.

[19] Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.

[20] Markatou, M. and Hettmansperger, T.P. (1990). Robust bounded-influence tests in linear models. *Journal of the American Statistical Association*, **85**, 187-190.

[21] Maronna, R.A. and Yohai, V.J. (1981). Asymptotic behavior of general M-estimates for regression and scale with random carriers. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **58**, 7-20.

[22] Martin, R.D. and Zamar, R. (1993). Bias robust estimation of scale. *The Annals of Statistics*, **21**, 991-1017.

[23] Rocke, D.M. and Downs, G.W. (1981). Estimating the variances of robust estimators of location: influence curve, jackknife and bootstrap. *Communications in Statistics, Part B – Simulation and Computation*, **10**, 221-248.

[24] Rousseeuw, P.J. and Yohai, V.J. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series.* (J. Franke, W. Hardle and D. Martin, eds.). *Lecture Notes in Statistics*, **26** 256-272. Berlin: Springer-Verlag.

[25] Salibian-Barrera, M. (2000). Contributions to the Theory of Robust Inference. Unpublished Ph.D. Thesis. University of British Columbia, Department of Statistics, Vancouver, BC. Available on-line at `http://hajek.stat.ubc.ca/~matias/pubs.html` .

[26] Salibian-Barrera and Zamar. (2004). Uniform asymptotics for robust location estimates when the scale is unknown. *The Annals of Statistics*, **32**, 1434-1447.

[27] Salibián-Barrera, M. (2006). The asymptotics of MM-estimators for linear regression with fixed designs. *Metrika*, **63**, 283-294.

[28] Simpson, D.G., Ruppert, D. and Carroll, R.J. (1992). On one step GM-estimates and stability of inferences in linear regression. *Journal of the American Statistical Association*, **87**, 439-450.

[29] Simpson, D.G. and Yohai, V.J. (1998). Functional stability of one-step GM-estimators in approximately linear regression. *The Annals of Statistics*, **26**, 1147-1169.

[30] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer, New York.

[31] Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, **15**, 642-656.