ASYMPTOTICS FOR REDESCENDING M-ESTIMATORS IN LINEAR MODELS WITH INCREASING DIMENSION

Instituto de Cálculo, Universidad de Buenos Aires - CONICET

Supplementary Material

This supplement contains the proofs of all the results stated in the paper, and a discussion of one of the main assumptions needed to prove the results.

S1 Assumption X3

We begin by discussing assumption X3, needed to prove the consistency of the estimators.

Suppose that $\mathbf{X}_i$, $i = 1, \ldots, n$ are independent and identically distributed random vectors in $\mathbb{R}^p$ such that there exist η_1, η_2 with $0 < \eta_1, \eta_2 < 1$ such that, for all n, $\sup_{\|\boldsymbol{\theta}\|=1} \mathbb{P}\left(|\mathbf{X}^T\boldsymbol{\theta}| < \eta_1\right) < 1 - \eta_2$. We will show that in this case, X3 holds in probability for $\mathbf{X}_i$. Note that $\sup_{\|\boldsymbol{\theta}\|=1} \mathbb{P}\left(|\mathbf{X}^T\boldsymbol{\theta}| < \eta_1\right) < 1 - \eta_2$ holds, for some $0 < \eta_1, \eta_2 < 1$ and all n, for example, if $\mathbf{X}_i \sim N_p(\mathbf{0}, \mathbf{M}_n)$ and there exists some $\kappa > 0$ such that the smallest eigenvalue of $\mathbf{M}_n$ is bounded below by κ for all n. It is easy to show, using maximal inequalities such as those of Theorem 1, that if $p/n \to 0$,

$$\sup_{\|\boldsymbol{\theta}\|=1} \left| \frac{1}{n} \sum_{i=1}^{n} I\left\{ |\mathbf{X}_i^T \boldsymbol{\theta}| < \eta_1 \right\} - \mathbb{P}\left(|\mathbf{X}^T \boldsymbol{\theta}| < \eta_1 \right) \right| \xrightarrow{P} 0.$$

Hence, with arbitrarily high probability, for large enough n,

$$\sup_{\|\boldsymbol{\theta}\|=1} \frac{1}{n} \sum_{i=1}^{n} I\left\{ |\mathbf{X}_i^T \boldsymbol{\theta}| < \eta_1 \right\} < \sup_{\|\boldsymbol{\theta}\|=1} \mathbb{P}\left(|\mathbf{X}^T \boldsymbol{\theta}| < \eta_1 \right) + \eta_2/2 < 1 - \eta_2/2.$$

In this case, for any α such that $1 - \eta_2/2 < \alpha < 1$, for large enough n it follows that for all $\boldsymbol{\theta}$ with $\|\boldsymbol{\theta}\| = 1$ and all subsets $\mathcal{A}$ of $\{1, \ldots, n\}$ with $\#\mathcal{A} = [n\alpha]$ there exists $i \in \mathcal{A}$ such that $|\mathbf{X}_i^T \boldsymbol{\theta}| \geq \eta_1$, which implies $\lambda_n(\alpha) \geq \eta_1$. See also Examples 1, 2 and 3 of Davies [1990].

S2 Proofs

Proof of Lemma 1. Let $0 < \alpha < 1$ be such that $\liminf \lambda_n(\alpha) > 0$. Note that for all $\eta > 0$, $\#\{i : \|\mathbf{x}_i\| \geq \eta\} \leq nMp\eta^{-2}$. Take $\eta = \sqrt{2Mp(1-\alpha)^{-1}}$. Let $\alpha_1 = (1/2)(1+\alpha)$, $\eta_1 = \sqrt{2M(1-\alpha)^{-1}}$ and $\mathcal{A} = \{i : \|\mathbf{x}_i\| < \eta_1 \sqrt{p}\}$. Then $\#\mathcal{A} \geq n\alpha_1$, with $0 < \alpha < \alpha_1 < 1$. Take $\boldsymbol{\theta}^*$ with $\|\boldsymbol{\theta}^*\| = 1$ such that

$$\sum_{i \in \mathcal{A}} |\mathbf{x}_i^T \boldsymbol{\theta}^*|^2 = \min_{\|\boldsymbol{\theta}\|=1} \sum_{i \in \mathcal{A}} |\mathbf{x}_i^T \boldsymbol{\theta}|^2.$$

Let $\mathcal{G}$ be the set of $i \in \mathcal{A}$ giving rise to the smallest $[n\alpha]$ values of $|\mathbf{x}_i^T \boldsymbol{\theta}^*|$. Then, by definition of $\lambda_n(\alpha)$, $\lambda_n(\alpha) \leq \max_{i \in \mathcal{G}} |\mathbf{x}_i^T \boldsymbol{\theta}^*|$. Hence, $\lambda_n(\alpha) \leq$

$|\mathbf{x}_i^T \boldsymbol{\theta}^*|$ for all $i \in \mathcal{A} \setminus \mathcal{G}$. Thus

$$\min_{\|\boldsymbol{\theta}\|=1} \frac{1}{n} \sum_{i\in\mathcal{A}} |\mathbf{x}_i^T \boldsymbol{\theta}|^2 = \frac{1}{n} \sum_{i\in\mathcal{A}} |\mathbf{x}_i^T \boldsymbol{\theta}^*|^2 \geq \frac{1}{n} \sum_{i\in\mathcal{A}\setminus\mathcal{G}} |\mathbf{x}_i^T \boldsymbol{\theta}^*|^2 \geq \frac{(n\alpha_1 - [n\alpha])\lambda_n(\alpha)^2}{n}$$

$$\geq (\alpha_1 - \alpha)\lambda_n(\alpha)^2.$$

The lemma is proven. □

We recall some of the notation used in the statement of Theorem 1 and further introduce additional notation. Let $\varepsilon > 0$. Let $\mathcal{H}$ be a class of functions defined on $\mathbb{R}^d$ and let $\|.\|$ be a pseudo-norm on $\mathcal{H}$.

- The capacity number of $\mathcal{H}$, $D(\varepsilon, \mathcal{H}, \|.\|)$, is the largest N such that there exists $h_1, \ldots, h_N$ in $\mathcal{H}$ with $\|h_i - h_j\| > \varepsilon$ for all $i \neq j$. The capacity number is also called the packing number in the literature.

- The covering number of $\mathcal{H}$, $N(\varepsilon, \mathcal{H}, \|.\|)$, is the minimal number of open balls of radius ε needed to cover $\mathcal{H}$.

- Given two functions h, g a bracket $[h, g]$ is the set of all functions f such that $h \leq f \leq g$. An ε-bracket is a bracket $[h, g]$ such that $\|h - g\| < \varepsilon$. $N_{[\,]}(\varepsilon, \mathcal{H}, \|.\|)$ is the bracketing number of $\mathcal{H}$, that is, the minimum number of ε-brackets needed to cover $\mathcal{H}$.

- Given a metric space (T, d), the covering number of T, $N(\varepsilon, T, d)$, is the minimal number of open balls of radius ε needed to cover T.

It is easy to show that $D(\varepsilon, \mathcal{H}, \|.\|) \leq N(\varepsilon/2, \mathcal{H}, \|.\|) \leq N_{[\,]}(\varepsilon, \mathcal{H}, \|.\|)$. Given Q, a probability measure on $\mathbb{R}^d$ with finite support, let $\|.\|_{2,Q}$ be the $L^2(Q)$ pseudo-norm.

Proof of Theorem 1. We prove (ii). Let $\mathbb{P}_n$ be the empirical probability measure that places mass $1/n$ at each of the points $(\mathbf{v}_i, \mathbf{z}_i)$ $i = 1, \ldots, n$. Let $\|.\|_{2,n}$ be the $L^2(\mathbb{P}_n)$ pseudo-norm. Let $\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_n$ be i.i.d. random vectors independent of and with the same distribution as $\mathbf{v}_1, \ldots, \mathbf{v}_n$. With a slight abuse of notation denote $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and let $E_{\mathbf{v}}$ be the expectation conditional on $\mathbf{v}$. It follows that for all $i = 1, \ldots, n$, $h(\mathbf{v}_i, \mathbf{z}_i) = E_{\mathbf{v}} h(\mathbf{v}_i, \mathbf{z}_i)$ and $Eh(\mathbf{v}_i, \mathbf{z}_i) = Eh(\tilde{\mathbf{v}}_i, \mathbf{z}_i) = E_{\mathbf{v}} h(\tilde{\mathbf{v}}_i, \mathbf{z}_i)$. Then, for all $h \in \mathcal{H}$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - Eh(\mathbf{v}_i, \mathbf{z}_i)\right) = E_{\mathbf{v}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - h(\tilde{\mathbf{v}}_i, \mathbf{z}_i)\right).$$

By Jensen's inequality

$$\left| E_{\mathbf{v}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - h(\tilde{\mathbf{v}}_i, \mathbf{z}_i)\right) \right|^2 \leq E_{\mathbf{v}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - h(\tilde{\mathbf{v}}_i, \mathbf{z}_i)\right) \right|^2.$$

Hence

$$E \sup_{h \in \mathcal{H}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - Eh(\mathbf{v}_i, \mathbf{z}_i)\right) \right|^2 \leq E \sup_{h \in \mathcal{H}} E_{\mathbf{v}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - h(\tilde{\mathbf{v}}_i, \mathbf{z}_i)\right) \right|^2$$
$$\leq E E_{\mathbf{v}} \sup_{h \in \mathcal{H}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - h(\tilde{\mathbf{v}}_i, \mathbf{z}_i)\right) \right|^2 = E \sup_{h \in \mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - h(\tilde{\mathbf{v}}_i, \mathbf{z}_i)\right) \right|^2.$$

Let $g_1, \ldots, g_n$ be i.i.d random variables, independent of $\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_n$ and of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ such that $g_i \sim N(0, 1)$. Define $w_i = g_i/|g_i|$ for $i = 1, \ldots, n$.

Then $w_1, \dots, w_n$ are independent of $\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_n$ and of $\mathbf{v}_1, \dots, \mathbf{v}_n$. Note that $\mathbb{P}(w_i = 1) = \mathbb{P}(w_i = -1) = 1/2$ and that w_i is independent of $|g_i|$. Let $\boldsymbol{w} = (w_1, \dots, w_n)$. By the symmetry between $\tilde{\mathbf{v}}_i$ and $\mathbf{v}_i$ we have that

$$E \sup_{h \in \mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - h(\tilde{\mathbf{v}}_i, \mathbf{z}_i) \right) \right|^2 = E \sup_{h \in \mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^{n} w_i \left(h(\mathbf{v}_i, \mathbf{z}_i) - h(\tilde{\mathbf{v}}_i, \mathbf{z}_i) \right) \right|^2 .$$

Now

$$\sup_{h \in \mathcal{H}} \left| \sum_{i=1}^{n} w_i \left(h(\mathbf{v}_i, \mathbf{z}_i) - h(\tilde{\mathbf{v}}_i, \mathbf{z}_i) \right) \right| \leq \sup_{h \in \mathcal{H}} \left| \sum_{i=1}^{n} w_i h(\mathbf{v}_i, \mathbf{z}_i) \right| + \sup_{h \in \mathcal{H}} \left| \sum_{i=1}^{n} w_i h(\tilde{\mathbf{v}}_i, \mathbf{z}_i) \right| .$$

Hence

$$E \sup_{h \in \mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^{n} w_i \left(h(\mathbf{v}_i, \mathbf{z}_i) - h(\tilde{\mathbf{v}}_i, \mathbf{z}_i) \right) \right|^2 \leq 4E \sup_{h \in \mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^{n} w_i h(\mathbf{v}_i, \mathbf{z}_i) \right|^2 .$$

Let δ be the expectation of $|g_1|$ and let $E_{\mathbf{v},\boldsymbol{w}}$ be the expectation conditional on $\mathbf{v}$ and $\boldsymbol{w}$. Then for all $i = 1, \dots, n$, $E_{\mathbf{v},\boldsymbol{w}} w_i h(\mathbf{v}_i, \mathbf{z}_i) = w_i h(\mathbf{v}_i, \mathbf{z}_i)$ and

$E_{\mathbf{v},\boldsymbol{w}}|g_i| = \delta$. Hence, applying Jensen's inequality

$$\begin{aligned}
E \sup_{h\in\mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^{n} w_i h(\mathbf{v}_i, \mathbf{z}_i) \right|^2 &= E \sup_{h\in\mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^{n} w_i h(\mathbf{v}_i, \mathbf{z}_i) E_{\mathbf{v},\boldsymbol{w}} |g_i|/\delta \right|^2 \\
&= E \sup_{h\in\mathcal{H}} \frac{1}{n} \left| E_{\mathbf{v},\boldsymbol{w}} \sum_{i=1}^{n} w_i h(\mathbf{v}_i, \mathbf{z}_i) |g_i|/\delta \right|^2 \\
&= E \sup_{h\in\mathcal{H}} \frac{1}{n} \left| E_{\mathbf{v},\boldsymbol{w}} \sum_{i=1}^{n} g_i h(\mathbf{v}_i, \mathbf{z}_i)/\delta \right|^2 \\
&\leq E \sup_{h\in\mathcal{H}} E_{\mathbf{v},\boldsymbol{w}} \frac{1}{n} \left| \sum_{i=1}^{n} g_i h(\mathbf{v}_i, \mathbf{z}_i)/\delta \right|^2 \\
&\leq E E_{\mathbf{v},\boldsymbol{w}} \sup_{h\in\mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^{n} g_i h(\mathbf{v}_i, \mathbf{z}_i)/\delta \right|^2 \\
&= \delta^{-2} E \sup_{h\in\mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^{n} g_i h(\mathbf{v}_i, \mathbf{z}_i) \right|^2 .
\end{aligned}$$

In summary, we have shown that

$$E \sup_{h\in\mathcal{H}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - E h(\mathbf{v}_i, \mathbf{z}_i) \right) \right|^2 \leq 4\delta^{-2} E \sup_{h\in\mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^{n} g_i h(\mathbf{v}_i, \mathbf{z}_i) \right|^2 . \tag{S2.1}$$

Define for $h \in \mathcal{H}$, $Z_n(h, \mathbf{v}) = (1/\sqrt{n}) \sum_{i=1}^{n} g_i h(\mathbf{v}_i, \mathbf{z}_i)$. Then (S2.1) can be written as

$$E \sup_{h\in\mathcal{H}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\mathbf{v}_i, \mathbf{z}_i) - E h(\mathbf{v}, \mathbf{z}_i) \right) \right|^2 \leq 4\delta^{-2} E \sup_{h\in\mathcal{H}} \left| Z_n(h, \mathbf{v}) \right|^2 . \tag{S2.2}$$

Note that, conditionally on the $\mathbf{v}_i$, Z_n is a zero-mean Gaussian process with increments bounded by the $L^2(\mathbb{P}_n)$ pseudo-norm: for all $h_1, h_2 \in \mathcal{H}$

$$
\begin{aligned}
& E_{\mathbf{v}} \left| Z_n(h_1, \mathbf{v}) - Z_n(h_2, \mathbf{v}) \right|^2 \\
&= \frac{1}{n} E_{\mathbf{v}} \sum_{i,j} (h_1(\mathbf{v}_i, \mathbf{z}_i) - h_2(\mathbf{v}_i, \mathbf{z}_i))(h_1(\mathbf{v}_j, \mathbf{z}_j) - h_2(\mathbf{v}_j, \mathbf{z}_j)) g_i g_j \\
&= \frac{1}{n} \sum_{i,j} (h_1(\mathbf{v}_i, \mathbf{z}_i) - h_2(\mathbf{v}_i, \mathbf{z}_i))(h_1(\mathbf{v}_j, \mathbf{z}_j) - h_2(\mathbf{v}_j, \mathbf{z}_j)) E_{\mathbf{v}} g_i g_j \\
&= \frac{1}{n} \sum_{i=1}^{n} (h_1(\mathbf{v}_i, \mathbf{z}_i) - h_2(\mathbf{v}_i, \mathbf{z}_i))^2 = \|h_1 - h_2\|_{2,n}^2.
\end{aligned}
$$

Also, for fixed $\mathbf{v}$, Z_n has continuous sample paths in the $L^2(\mathbb{P}_n)$ pseudo-norm: if $\|h - h_k\|_{2,n} \to 0$ when $k \to \infty$ then $h_k(\mathbf{v}_i, \mathbf{z}_i) \to h(\mathbf{v}_i, \mathbf{z}_i)$ for all $i = 1, \ldots, n$ and hence $Z_n(h_k, \mathbf{v}) \to Z_n(h, \mathbf{v})$ for each realization of the g_i. Therefore, we can apply Theorem 3.3 of Pollard [1989]: there exists an universal constant $M > 0$ such that

$$
\left(E_{\mathbf{v}} \sup_{h \in \mathcal{H}} \left| Z_n(h, \mathbf{v}) \right|^2 \right)^{1/2} \leq M \int_0^{\Delta(\mathbf{v})} \left(\log D(x, \mathcal{H}, \|.\|_{2,n}) \right)^{1/2} dx, \qquad \text{(S2.3)}
$$

where $\Delta(\mathbf{v}) = \sup_{h \in \mathcal{H}} \|h\|_{2,n}$.

If $\|H\|_{2,n} > 0$, since by assumption $D(\varepsilon \|H\|_{2,n}, \mathcal{H}, \|.\|_{2,n}) \leq D(\varepsilon)$ for all $0 < \varepsilon < 1$, we have that

$$
\int_0^1 \left(\log D(\varepsilon \|H\|_{2,n}, \mathcal{H}, \|.\|_{2,n}) \right)^{1/2} d\varepsilon \leq \int_0^1 \left(\log D(\varepsilon) \right)^{1/2} d\varepsilon < \infty.
$$

Also, since $\Delta(\mathbf{v})/\|H\|_{2,n} \leq 1$

$$\int_0^{\Delta(\mathbf{v})/\|H\|_{2,n}} \left(\log D(\varepsilon\|H\|_{2,n}, \mathcal{H}, \|.\|_{2,n})\right)^{1/2} d\varepsilon$$
$$\leq \int_0^1 \left(\log D(\varepsilon\|H\|_{2,n}, \mathcal{H}, \|.\|_{2,n})\right)^{1/2} d\varepsilon$$

The change of variables $x = \varepsilon\|H\|_{2,n}$ gives

$$\|H\|_{2,n} \int_0^{\Delta(\mathbf{v})/\|H\|_{2,n}} \left(\log D(\varepsilon\|H\|_{2,n}, \mathcal{H}, \|.\|_{2,n})\right)^{1/2} d\varepsilon$$
$$= \int_0^{\Delta(\mathbf{v})} \left(\log D(x, \mathcal{H}, \|.\|_{2,n})\right)^{1/2} dx.$$

Then

$$\left(E_{\mathbf{v}} \sup_{h\in\mathcal{H}} |Z_n(h,\mathbf{v})|^2\right)^{1/2} \leq M\|H\|_{2,n} \int_0^1 \left(\log D(\varepsilon)\right)^{1/2} d\varepsilon. \tag{S2.4}$$

On the other hand, if $\|H\|_{2,n} = 0$ then $\Delta(\mathbf{v}) = 0$ and this implies that the right hand side of (S2.3) is zero. In this case (S2.4) holds trivially.

We have thus shown that

$$E \sup_{h\in\mathcal{H}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(h(\mathbf{v}_i,\mathbf{z}_i) - Eh(\mathbf{v}_i,\mathbf{z}_i)\right) \right|^2$$
$$\leq 4\delta^{-2} E \sup_{h\in\mathcal{H}} |Z_n(h,\mathbf{v})|^2$$
$$= 4\delta^{-2} E E_{\mathbf{v}} \sup_{h\in\mathcal{H}} |Z_n(h,\mathbf{v})|^2$$
$$\leq 4\delta^{-2} M^2 E\|H\|_{2,n}^2 \left(\int_0^1 \left(\log D(\varepsilon)\right)^{1/2} d\varepsilon\right)^2$$
$$= 4\delta^{-2} M^2 \frac{1}{n} \sum_{i=1}^n EH^2(\mathbf{v}_i,\mathbf{z}_i) \left(\int_0^1 \left(\log D(\varepsilon)\right)^{1/2} d\varepsilon\right)^2,$$

which is what we wanted to prove. Part (i) can be proved by substituting $L^1(\mathbb{P}_n)$ norms by $L^2(\mathbb{P}_n)$ in the arguments leading to (S2.2) and then applying Theorem 3.2 of Pollard [1989]. $\square$

Proof of Lemma 2. We will apply the maximal inequalities of Theorem 1 to $\mathcal{H} \cup \{0\}$.

Let $\mathcal{L} = \left\{ l_{s,\boldsymbol{\theta}}(t, \mathbf{x}) = (t - \mathbf{x}^T \boldsymbol{\theta})/s : \boldsymbol{\theta} \in \mathbb{R}^p,\ s > 0 \right\}$. Then $\mathcal{L}$ is a subset of the vector space of all linear functions in $p+1$ variables. This vector space has dimension $p+1$. It follows from Lemma 2.6.15 of van der vaart and Wellner [1996] that $\mathcal{L}$ has VC-index at most $p+3$.

Note that $\rho = m^1 + m^2$, where $m^1(t) = \rho(t) I\{t \geq 0\}$ and $m^2(t) = \rho(t) I\{t < 0\}$. Note that m^1 is non-decreasing and m^2 is non-increasing. By Lemma 9.9 (viii) of Kosorok [2008], $m^1 \circ \mathcal{L}$ and $m^2 \circ \mathcal{L}$ have VC-index at most $p+3$. $m^1 \circ \mathcal{L}$ and $m^2 \circ \mathcal{L}$ have a constant envelope equal to 1.

Let Q be a probability measure on $\mathbb{R}^{p+1}$ with finite support. Fix $0 < \varepsilon < 1$. By Theorem 2.6.7 from van der vaart and Wellner [1996], for some universal constant C_0 we have that for $i = 1, 2$

$$N(\varepsilon, m^i(\mathcal{L}), \|.\|_{2,Q}) \leq C_0 (p+3)(16e)^{p+3} \varepsilon^{-2(p+2)}.$$

Note that $m^1 \circ \mathcal{L} + m^2 \circ \mathcal{L}$ has constant envelope equal to 2. It is easy to show that

$$N(2\varepsilon, m^1 \circ \mathcal{L} + m^2 \circ \mathcal{L}, \|.\|_{2,Q}) \leq N(\varepsilon/2, m^1 \circ \mathcal{L}, \|.\|_{2,Q}) N(\varepsilon/2, m^2 \circ \mathcal{L}, \|.\|_{2,Q})$$
$$\leq (C_0(p+3)(16e)^{p+3} (\varepsilon/2)^{-2(p+2)})^2.$$

Note that $\mathcal{H}$ has envelope $H(t, \mathbf{x}) = 2$ and that $\mathcal{H} \subset m^1 \circ \mathcal{L} + m^2 \circ \mathcal{L}$. Hence

$$N(\varepsilon \|H\|_{2,Q}, \mathcal{H}, \|.\|_{2,Q}) \leq (C_0(p+3)(16e)^{p+3} (\varepsilon/2)^{-2(p+2)})^2.$$

Furthermore $\mathcal{H} \cup \{0\}$ also has envelope H. We can assume without loss of generality that $C_0 > 1$. Hence,

$$N(\varepsilon \|H\|_{2,Q}, \mathcal{H} \cup \{0\}, \|.\|_{2,Q}) \leq N(\varepsilon \|H\|_{2,Q}, \mathcal{H}, \|.\|_{2,Q}) + 1$$
$$\leq 2(C_0(p+3)(16e)^{p+3} (\varepsilon/2)^{-2(p+2)})^2$$

implies that

$$D(\varepsilon \|H\|_{2,Q}, \mathcal{H} \cup \{0\}, \|.\|_{2,Q}) \leq N((\varepsilon/2) \|H\|_{2,Q}, \mathcal{H} \cup \{0\}, \|.\|_{2,Q}) \leq D(\varepsilon)$$

where $D(\varepsilon) = 2(C_0(p+3)(16e)^{p+3} (\varepsilon/4)^{-2(p+2)})^2$.

It follows from Theorem 1(i) that for some fixed $C_1 > 0$

$$E \sup_{h \in \mathcal{H}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (h(u_i, \mathbf{x}_i) - Eh(u, \mathbf{x}_i)) \right| \leq E \sup_{h \in \mathcal{H} \cup \{0\}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (h(u_i, \mathbf{x}_i) - Eh(u, \mathbf{x}_i)) \right|$$
$$\leq C_1 \int_0^1 (\log D(\varepsilon))^{1/2}.$$

Note that $\log D(\varepsilon) \leq C_2 p (1 - \log \varepsilon)$ for some fixed $C_2 > 0$. Hence

$$E \sup_{h \in \mathcal{H}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (h(u_i, \mathbf{x}_i) - Eh(u, \mathbf{x}_i)) \right| \leq \sqrt{p} C_1 \sqrt{C_2} \int_0^1 (1 - \log \varepsilon)^{1/2} \, d\varepsilon = \sqrt{p} C_3$$

where $C_3 > 0$ is fixed. Finally, the last part of the lemma follows from applying Markov's inequality. □

Proof of Lemma 3. This follows from Theorem 3 of Davies [1990], replacing any appeals in the proof of that theorem to Lemma 2 of Davies [1990] by appeals to Lemma 2. □

The following lemma is similar to Lemma 1 of Davies [1990]. However, unlike our result, Lemma 1 of Davies [1990] requires the loss function to satisfy $\rho_1(t) = 1$ for all sufficiently large t, which excludes some interesting loss functions, such as Welsh's. For $v, s \in \mathbb{R}$, let $R(v,s) = E\rho_1\left((u-v)/s\right)$.

Lemma 4. *Assume R1 and F0 hold. Then*

(i) $R : \mathbb{R} \times \mathbb{R}_+ \to [0,1]$ *is continuous.*

(ii) $R(0,s) \leq R(v,s)$ *for* $v \in \mathbb{R}$*,* $s > 0$*.*

(iii) $R(0,s) < \inf_{|v|\geq\eta} R(v,s)$ *for all* $\eta > 0$ *and* $s > 0$*.*

Proof of Lemma 4. We first prove (i). Let $\{(v_k, s_k)_k\} \subset \mathbb{R} \times \mathbb{R}_+$ be a sequence converging to (v,s) with $s > 0$, we will show that $R(v_k, s_k) \to R(v,s)$. Since ρ_1 is continuous we have that, for each fixed value of t, $\rho_1\left((t-v_k)/s_k\right) \to \rho_1\left((t-v)/s\right)$. In particular, the sequence of functions $g_k(t) = \rho_1\left((t-v_k)/s_k\right)$ converges F_0-almost surely to $g(t) = \rho_1\left((t-v)/s\right)$.

Moreover, by assumption, $|g_k(t)| \leq 1$ for all t. Hence, the Bounded Convergence Theorem implies that $Eg_k(u) = R(v_k, s_k) \rightarrow Eg(u) = R(v, s)$. We have thus shown that R is continuous.

Next we prove (ii). This is roughly Lemma 3.1 of Yohai [1985]. Note that for any $v \neq 0$, the distribution function R_v of $|u - v|$ satisfies: $R_v(t) \leq R_0(t)$ for all $t > 0$ and there exists $\delta > 0$ such that $R_v(t) < R_0(t)$ for $0 < t \leq \delta$. Since $\rho_1(t/s)$ is non decreasing in $|t|$ and strictly increasing in a neighbourhood of 0, it follows that for all s, $R(v, s)$ has a unique minimum at $v = 0$.

Now we prove (iii). Suppose for some $\eta, s > 0$, $R(0, s) \geq \inf_{|v| \geq \eta} R(v, s)$. Note that by R1 and F0, $R(0, s) < 1$. Take v_n with $|v_n| \geq \eta$ such that $R(v_n, s) \rightarrow \inf_{|v| \geq \eta} R(v, s)$. Note that if for some subsequence v_{n_k}, $|v_{n_k}| \rightarrow \infty$, then by the Bounded Convergence Theorem $R(v_{n_k}, s) \rightarrow 1$ and hence $R(0, s) \geq 1$, leading to a contradiction. Hence v_n must be bounded. We can assume, eventually passing to a subsequence, that $v_n \rightarrow v^*$, with $|v^*| \geq \eta$. Hence $R(v^*, s) = \inf_{|v| \geq \eta} R(v, s) \leq R(0, s)$. But by (ii), $R(v, s)$ has a unique minimum at $v = 0$. Hence (iii) follows. $\square$

Proof of Theorem 2. Fix $0 < \alpha < 1$. Note that by definition of $\hat{\boldsymbol{\beta}}$

$$\frac{1}{n} \sum_{i=1}^{n} \rho_1 \left(\frac{u_i - \mathbf{x}_i^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)}{\hat{\sigma}_n} \right) \leq \frac{1}{n} \sum_{i=1}^{n} \rho_1 \left(\frac{u_i}{\hat{\sigma}_n} \right).$$

By Lemma 2 we have that

$$\sup_{\boldsymbol{\theta}\in\mathbb{R}^p, 0<s<2s_0} \frac{1}{n}\left|\sum_{i=1}^{n}\left(\rho_1\left(\frac{u_i-\mathbf{x}_i^T\boldsymbol{\theta}}{s}\right)-R(\mathbf{x}_i^T\boldsymbol{\theta},s)\right)\right| \xrightarrow{P} 0. \tag{S2.5}$$

Since by assumption $\hat{\sigma}_n \xrightarrow{P} s_0$, Lemma 4 (i) implies that the right hand side of the last inequality converges in probability to

$$b^* = E\rho_1\left(\frac{u}{s_0}\right). \tag{S2.6}$$

By Lemma 4 (ii), $R(0,s) \leq R(v,s)$ for all $v \in \mathbb{R}$, $s \in \mathbb{R}$. Then

$$R(0,\hat{\sigma}_n) \leq \frac{1}{n}\sum_{i=1}^{n} R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0),\hat{\sigma}_n). \tag{S2.7}$$

By Lemma 4 (i), $R(0,s)$ is a continuous function of s. Since by assumption $\hat{\sigma}_n \xrightarrow{P} s_0$

$$R(0,\hat{\sigma}_n) \xrightarrow{P} R(0,s_0) = b^*. \tag{S2.8}$$

Then, it follows from (S2.5), (S2.6), (S2.7) and (S2.8) that

$$\frac{1}{n}\sum_{i=1}^{n}\rho_1\left(\frac{u_i-\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)}{\hat{\sigma}_n}\right) \xrightarrow{P} b^*$$

and

$$\frac{1}{n}\sum_{i=1}^{n} R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0),\hat{\sigma}_n) \xrightarrow{P} b^*. \tag{S2.9}$$

By (S2.9), given $\delta > 0$, with arbitrarily high probability, for large enough n we have that

$$\frac{1}{n}\sum_{i=1}^{n} R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0),\hat{\sigma}_n) \leq b^* + \delta. \tag{S2.10}$$

Let $\varepsilon > 0$, we will show that with arbitrarily high probability, for large enough n, $\lambda_n(\alpha)\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\| \leq \varepsilon$. Let $A = \{i : |\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)| \geq \varepsilon\}$ and $N = \#A$. Then

$$\frac{1}{n}\sum_{i=1}^{n} R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0), \hat{\sigma}_n) = \frac{1}{n}\sum_{i\in A} R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0), \hat{\sigma}_n) + \frac{1}{n}\sum_{i\in A^c} R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0), \hat{\sigma}_n).$$

Note that

$$\frac{1}{n}\sum_{i\in A^c} R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0), \hat{\sigma}_n) \geq \frac{n-N}{n} R(0, \hat{\sigma}_n). \tag{S2.11}$$

Also, if $|\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)| \geq \varepsilon$ then $R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0), \hat{\sigma}_n) \geq \inf_{|v|\geq\varepsilon} R(v, \hat{\sigma}_n)$ and hence

$$R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0), \hat{\sigma}_n) \geq R(0, \hat{\sigma}_n) + \left(\inf_{|v|\geq\varepsilon} R(v, \hat{\sigma}_n) - R(0, \hat{\sigma}_n)\right).$$

We will show that with arbitrarily high probability, for large enough n and $i \in A$

$$R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0), \hat{\sigma}_n) \geq R(0, \hat{\sigma}_n) + \kappa, \tag{S2.12}$$

for some $\kappa = \kappa(\varepsilon) > 0$. First, we will show that

$$\sup_v |R(v, \hat{\sigma}_n) - R(v, s_0)| \xrightarrow{P} 0 \tag{S2.13}$$

Fix $u, v \in \mathbb{R}$. Let $\phi_1(t) = \psi_1(t)t$. By R1, ϕ_1 is bounded. Applying the Mean Value Theorem we get that, for some $\hat{\sigma}_n^*$ such that $|\hat{\sigma}_n^* - s_0| \leq |\hat{\sigma}_n - s_0|$

$$\begin{aligned}\left|\rho_1\left(\frac{u-v}{\hat{\sigma}_n}\right) - \rho_1\left(\frac{u-v}{s_0}\right)\right| &\leq \left|\psi_1\left(\frac{u-v}{\hat{\sigma}_n^*}\right)\left(\frac{u-v}{\hat{\sigma}_n^*}\right)\right|\left|\frac{\hat{\sigma}_n - s_0}{\hat{\sigma}_n^*}\right| \\ &\leq \|\phi_1\|_\infty \left|\frac{\hat{\sigma}_n - s_0}{\hat{\sigma}_n^*}\right|. \end{aligned} \tag{S2.14}$$

Fix some $\eta > 0$. Since $\hat{\sigma}_n \xrightarrow{P} s_0$, with arbitrarily high probability, for large enough n, the right hand side of (S2.14) is smaller than η for all u, v. (S2.13) is proven.

By Lemma 4 (iii), $\inf_{|v|\geq\varepsilon} R(v, s_0) > R(0, s_0)$. Let

$$\eta_1 = (\inf_{|v|\geq\varepsilon} R(v, s_0) - R(0, s_0))/4.$$

Fix $\eta_2 > 0$. Take n_0 such that for all $n \geq n_0$, $\sup_v |R(v, \hat{\sigma}_n) - R(v, s_0)| < \eta_1/2$ with probability greater than $1 - \eta_2$. For each $n_1 \geq n_0$, take v_{n_1} with $|v_{n_1}| \geq \varepsilon$ such that $\inf_{|v|\geq\varepsilon} R(v, \hat{\sigma}_{n_1}) \geq R(v_{n_1}, \hat{\sigma}_{n_1}) - \eta_1/2$. Note that v_{n_1} is random. It follows that with probability greater than $1 - \eta_2$, for all $n_1 \geq n_0$

$$\begin{aligned}\inf_{|v|\geq\varepsilon} R(v, s_0) - \inf_{|v|\geq\varepsilon} R(v, \hat{\sigma}_{n_1}) &\leq R(v_{n_1}, s_0) - R(v_{n_1}, \hat{\sigma}_{n_1}) + \eta_1/2 \\ &\leq \sup_v |R(v, \hat{\sigma}_{n_1}) - R(v, s_0)| + \eta_1/2 < \eta_1.\end{aligned}$$

Since $R(0, \hat{\sigma}_n) \xrightarrow{P} R(0, s_0)$, with arbitrarily high probability, for large enough n,

$$\begin{aligned}\inf_{|v|\geq\varepsilon} R(v, \hat{\sigma}_n) - R(0, \hat{\sigma}_n) &= \inf_{|v|\geq\varepsilon} R(v, \hat{\sigma}_n) - \inf_{|v|\geq\varepsilon} R(v, s_0) \\ &\quad + \inf_{|v|\geq\varepsilon} R(v, s_0) - R(0, s_0) \\ &\quad + R(0, s_0) - R(0, \hat{\sigma}_n) \\ &\geq 2\eta_1.\end{aligned}$$

We have proven (S2.12) for $\kappa(\varepsilon) = \left(\inf_{|v|\geq\varepsilon} R(v, s_0) - R(0, s_0)\right)/2$. Hence with arbitrarily high probability, for large enough n

$$\frac{1}{n}\sum_{i\in A} R(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0),\hat{\sigma}_n) \geq \frac{N}{n}(R(0,\hat{\sigma}_n)+\kappa)$$

and thus by (S2.10), (S2.11) and (S2.12) with arbitrarily high probability, for large n, we have that if $N \geq (1-\alpha)n$ then $R(0,\hat{\sigma}_n) \leq b^* + \delta - (1-\alpha)\kappa$. In summary, we have shown that

$$\{N \geq (1-\alpha)n\} \subseteq \{R(0,\hat{\sigma}_n) \leq b^* + \delta - (1-\alpha)\kappa\} \cup A_n, \qquad \text{(S2.15)}$$

where $\mathbb{P}(A_n) \to 0$. For any given ε, we can find a sufficiently small δ such that $\delta - (1-\alpha)\kappa < 0$. Then by (S2.8) and (S2.15), $\mathbb{P}(N \geq (1-\alpha)n) \to 0$. Hence, with arbitrarily high probability, for sufficiently large n, $n\alpha < n - N$. In this case, there must exist $\mathcal{A} \subset \{1, ..., n\}$ with $\#\mathcal{A} = [n\alpha]$ such that $\left|\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\right| < \varepsilon$ for all $i \in \mathcal{A}$ and this implies that

$$\lambda_n(\alpha)\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\| \leq \min_{\mathcal{A}\subset\{1,\ldots,n\},\#\mathcal{A}=[n\alpha]} \max_{i\in\mathcal{A}} \left|\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\right| \leq \varepsilon,$$

which is what we wanted to prove. □

The following lemmas are needed in the proof of Theorem 3.

Lemma 5. *Assume R2, F0 and X1 a) hold. Let* $0 < a < c$. *For* $\mathbf{x} \in \mathbb{R}^p$, *consider the class of functions* $\mathcal{H} = \{h_s(u,\mathbf{x}) = \psi_1(u/s)\,\mathbf{x} : s \in [a,c]\}$. *Then, for some fixed constant* $A > 0$ *that depends only on* a, c, ψ_1 *and the* *constant that appears in X1 a),*

$$E \sup_{h \in \mathcal{H}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(u_i, \mathbf{x}_i) \right\| \leq A\sqrt{p}.$$

Proof. Let $\mathcal{G} = \{g_s(t,x) = \psi_1(t/s)\, x : s \in [a,c]\}$. Fix $1 \leq j \leq p$. Note that $Eg(u, x_{i,j}) = 0$ for $g \in \mathcal{G}$ and $i = 1, \ldots, n$. Note also that $\mathcal{G}$ has envelope $G(t,x) = \|\psi_1\|_\infty |x|$ and that $(1/n) \sum_{i=1}^n EG^2(u_i, x_{i,j}) = (\|\psi_1\|_\infty^2/n) \sum_{i=1}^n x_{i,j}^2 < \infty$. Let Q be a probability measure on $\mathbb{R}^2$ with finite support such that $\|G\|_{2,Q} > 0$. This implies that $\|x\|_{2,Q} > 0$

Let $\chi_1(t) = t\psi_1'(t)$. By R2, χ_1 is bounded. Also, if $s_1, s_2 \in [a,c]$, then by the Mean Value Theorem $|g_{s_1}(t,x) - g_{s_2}(t,x)| \leq (\|\chi_1\|_\infty |x|\, |s_1 - s_2|)/a$. Then, by Theorem 2.7.11 of van der vaart and Wellner [1996], for all $\varepsilon > 0$ the bracketing number of $\mathcal{G}$ satisfies

$$N_{[\,]}(2\varepsilon \|\chi_1\|_\infty \frac{1}{a} \| |x| \|_{2,Q}, \mathcal{G}, \|.\|_{2,Q}) \leq N(\varepsilon, [a,c], |\,.\,|). \tag{S2.16}$$

Note that for some constant C_1 that depends only on a and c, for all $\varepsilon > 0$

$$N(\varepsilon, [a,c], |\,.\,|) \leq \frac{C_1}{\varepsilon} + 1. \tag{S2.17}$$

Fix $0 < \varepsilon < 1$. It follows from (S2.16) and (S2.17) that

$$
\begin{aligned}
N(\varepsilon \|G\|_{2,Q}, \mathcal{G}, \|.\|_{2,Q}) &= N(\varepsilon \|\psi_1\|_\infty \||x|\|_{2,Q}, \mathcal{G}, \|.\|_{2,Q}) \\
&\leq N_{[\,]}(2\varepsilon \|\psi_1\|_\infty \||x|\|_{2,Q}, \mathcal{G}, \|.\|_{2,Q}) \\
&\leq N(\frac{a\varepsilon \|\psi_1\|_\infty}{\|\chi_1\|_\infty}, [a,b], |\,.\,|) \\
&\leq \frac{C_1 \|\chi_1\|_\infty}{a\varepsilon \|\psi_1\|_\infty} + 1 = \frac{C_2}{\varepsilon} + 1.
\end{aligned}
$$

Note that $\mathcal{G} \cup \{0\}$ has envelope G, $Eg(u, x_{i,j}) = 0$ for $g \in \mathcal{G} \cup \{0\}$ and $i = 1, \ldots, n$, and that

$$
N(\varepsilon \|G\|_{2,Q}, \mathcal{G} \cup \{0\}, \|.\|_{2,Q}) \leq N(\varepsilon \|G\|_{2,Q}, \mathcal{G}, \|.\|_{2,Q}) + 1 \leq \frac{C_2}{\varepsilon} + 2.
$$

Thus

$$
D(\varepsilon \|G\|_{2,Q}, \mathcal{G} \cup \{0\}, \|.\|_{2,Q}) \leq N(\varepsilon \|G\|_{2,Q}/2, \mathcal{G} \cup \{0\}, \|.\|_{2,Q}) \leq \frac{2C_2}{\varepsilon} + 2.
$$

Let $D(\varepsilon) = 2C_2/\varepsilon + 2$. Then by Theorem (1)(ii), for some fixed $C_3 > 0$

$$
\begin{aligned}
E \sup_{g \in \mathcal{G}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(u_i, x_{i,j}) \right|^2 &\leq E \sup_{g \in \mathcal{G} \cup \{0\}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(u_i, x_{i,j}) \right|^2 \\
&\leq C_3 \left(\frac{1}{n} \sum_{i=1}^{n} x_{i,j}^2 \right) \left(\int_0^1 \left(\log\left(\frac{2C_2}{\varepsilon} + 2 \right) \right)^{1/2} d\varepsilon \right)^2 .
\end{aligned} \tag{S2.18}
$$

Note that (S2.18) holds for all $1 \leq j \leq p$. Then, by (S2.18) and X1 a), for some fixed $C_4 > 0$

$$E \sup_{h\in\mathcal{H}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(u_i, \mathbf{x}_i) \right\|^2 = E \sup_{s\in[a,c]} \sum_{j=1}^{p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_1\left(\frac{u_i}{s}\right) x_{i,j} \right|^2$$
$$\leq \sum_{j=1}^{p} E \sup_{s\in[a,c]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_1\left(\frac{u_i}{s}\right) x_{i,j} \right|^2$$
$$\leq \sum_{j=1}^{p} C_3 \left(\frac{1}{n}\sum_{i=1}^{n} x_{i,j}^2\right) \left(\int_0^1 \left(\log\left(\frac{2C_2}{\varepsilon}+2\right)\right)^{1/2} d\varepsilon\right)^2 \leq C_4 p.$$

The result now follows from applying Jensen's inequality. □

Lemma 6. *Assume R2 and F0 hold. Then* $E\psi_1'(u/s_0) > 0$.

Proof of Lemma 6. Let $[-c, c]$ be the (possibly infinite) interval where $\rho_1(t) < 1$. In this case, $\psi_1 = \rho_1'$ is an odd function such that $\psi_1(t) \geq 0$ for $t \geq 0$ and $\psi_1(t) > 0$ if $t \in (0, c)$. Note that by partial integration we have that

$$E\psi_1'(u/s_0) = -\, s_0 \int \psi_1\left(\frac{u}{s_0}\right) f_0'(u)du = -\,2\, s_0 \int_0^\infty \psi_1\left(\frac{u}{s_0}\right) f_0'(u)du$$

where the last equality follows from the fact that ψ_1 and f_0' are odd functions. Hence, $E\psi_1'(u/s_0) = 2\, s_0(I_1 + I_2)$ where

$$\begin{aligned} I_1 &= \int_0^{c\, s_0} \psi_1\left(\frac{u}{s_0}\right)(-f_0'(u))du \\ I_2 &= \int_{c\, s_0}^{\infty} \psi_1\left(\frac{u}{s_0}\right)(-f_0'(u))du. \end{aligned}$$

Note that F0 entails that $-f_0'(u) \geq 0$ for any $u > 0$ and furthermore, in some neighbourhood of 0, $-f_0'(u) > 0$, which together with the fact that

$\psi_1(u/s_0) \geq 0$ for $u \geq 0$ and $\psi_1(u/s_0) > 0$ if $u \in (0, c\, s_0)$ implies that $I_1 > 0$ and $I_2 \geq 0$, leading to $E\psi_1'(u/s_0) > 0$.

$\square$

The following lemma extends Lemma 3.1 of Portnoy [1984] to M-estimators defined using a scale to standardize the residuals. It is needed to obtain the rate of consistency of the estimators. Define

$$K_i(t) = \inf_{|v| \leq |t|} \psi_1^{'} \left(\frac{u_i - v}{s_0} \right)$$

and

$$K_i^n(t) = \inf_{|v| \leq |t|} \psi_1^{'} \left(\frac{u_i - v}{\hat{\sigma}_n} \right).$$

Lemma 7. *Assume R2, F0, X1, X2, X4 and X5 hold and* $(p \log n)/n \to 0$. *Then there exists* $a^* > 0$ *and* $\delta > 0$ *such that*

$$\mathbb{P} \left(\inf_{\|\mathbf{z}\|=1, \|\boldsymbol{\beta}\| \leq \delta} \sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{z})^2 K_i^n(\mathbf{x}_i^T \boldsymbol{\beta}) \geq a^* n \right) \to 1. \tag{S2.19}$$

Proof. Note that for any $\delta > 0$

$$\begin{aligned} \inf_{\|\mathbf{z}\|=1, \|\boldsymbol{\beta}\| \leq \delta} \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{z})^2 K_i^n(\mathbf{x}_i^T \boldsymbol{\beta}) \geq & \inf_{\|\mathbf{z}\|=1, \|\boldsymbol{\beta}\| \leq \delta} \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{z})^2 K_i(\mathbf{x}_i^T \boldsymbol{\beta}) \\ & + \inf_{\|\mathbf{z}\|=1, \|\boldsymbol{\beta}\| \leq \delta} \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{z})^2 \left(K_i^n(\mathbf{x}_i^T \boldsymbol{\beta}) - K_i(\mathbf{x}_i^T \boldsymbol{\beta}) \right). \end{aligned}$$

By Lemma 6, $E\psi_1'(u/s_0) > 0$. Hence, by Lemma 3.1 of Portnoy [1984], (S2.19) holds when K_i^n is replaced by K_i. Hence, for some $a^* > 0$ and

$\delta > 0$, for sufficiently large n, with arbitrarily high probability

$$\inf_{\|\mathbf{z}\|=1, \|\boldsymbol{\beta}\| \leq \delta} \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{z})^2 K_i(\mathbf{x}_i^T \boldsymbol{\beta}) \geq a^*.$$

We will show that

$$\sup_{\|\mathbf{z}\|=1, \|\boldsymbol{\beta}\| \leq \delta} \left| \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{z})^2 \left(K_i^n(\mathbf{x}_i^T \boldsymbol{\beta}) - K_i(\mathbf{x}_i^T \boldsymbol{\beta}) \right) \right| \xrightarrow{P} 0.$$

Fix $i \leq n$, $\mathbf{z}$ with $\|\mathbf{z}\| = 1$, and $\boldsymbol{\beta}$ with $\|\boldsymbol{\beta}\| \leq \delta$. We will bound $\left| K_i^n(\mathbf{x}_i^T \boldsymbol{\beta}) - K_i(\mathbf{x}_i^T \boldsymbol{\beta}) \right|$. Assume $K_i^n(\mathbf{x}_i^T \boldsymbol{\beta}) \geq K_i(\mathbf{x}_i^T \boldsymbol{\beta})$. By R2, $K_i(\mathbf{x}_i^T \boldsymbol{\beta}) = \psi_1'\left((u_i - v_i^*)/s_0\right)$ for some v_i^* with $|v_i^*| \leq |\mathbf{x}_i^T \boldsymbol{\beta}|$. Then

$$\begin{aligned}
\left| K_i^n(\mathbf{x}_i^T \boldsymbol{\beta}) - K_i(\mathbf{x}_i^T \boldsymbol{\beta}) \right| &= K_i^n(\mathbf{x}_i^T \boldsymbol{\beta}) - K_i(\mathbf{x}_i^T \boldsymbol{\beta}) \\
&= K_i^n(\mathbf{x}_i^T \boldsymbol{\beta}) - \psi_1'\left(\frac{u_i - v_i^*}{s_0}\right) \\
&\leq \left| \psi_1'\left(\frac{u_i - v_i^*}{\hat{\sigma}_n}\right) - \psi_1'\left(\frac{u_i - v_i^*}{s_0}\right) \right|.
\end{aligned}$$

Note that by R2, $\varsigma_1(t) = \psi_1''(t)t$ is bounded. Applying the Mean Value Theorem we get that

$$\begin{aligned}
\left| \psi_1'\left(\frac{u_i - v_i^*}{\hat{\sigma}_n}\right) - \psi_1'\left(\frac{u_i - v_i^*}{s_0}\right) \right| &= \left| \psi_1''\left(\frac{u_i - v_i^*}{s_{i,n}^*}\right) \left(\frac{u_i - v_i^*}{s_{i,n}^*}\right) \right| \left| \frac{\hat{\sigma}_n - s_0}{s_{i,n}^*} \right| \\
&\leq \|\varsigma_1\|_\infty \left| \frac{\hat{\sigma}_n - s_0}{s_{i,n}^*} \right|.
\end{aligned}$$

where $s_{i,n}^*$ is such that $|s_{i,n}^* - s_0| \leq |\hat{\sigma}_n - s_0|$. Note that $s_{i,n}^*$ may depend on $\boldsymbol{\beta}$, say $s_{i,n}^* = s_{i,n}^*(\boldsymbol{\beta})$. The same type of argument can be used to show that an analogous bound holds when $K_i^n(\mathbf{x}_i^T \boldsymbol{\beta}) \leq K_i(\mathbf{x}_i^T \boldsymbol{\beta})$.

Note that since $\hat{\sigma}_n \xrightarrow{P} s_0$, we have that $\sup_{\|\boldsymbol{\beta}\|\leq\delta} \max_{i\leq n} |s^*_{i,n}(\boldsymbol{\beta}) - s_0| \leq |\hat{\sigma}_n - s_0| \xrightarrow{P} 0$. Then, using assumption X2

$$\begin{aligned}
&\sup_{\|\mathbf{z}\|=1,\|\boldsymbol{\beta}\|\leq\delta} \left| \frac{1}{n}\sum_{i=1}^{n} (\mathbf{x}_i^T\mathbf{z})^2 \left(K_i^n(\mathbf{x}_i^T\boldsymbol{\beta}) - K_i(\mathbf{x}_i^T\boldsymbol{\beta})\right)\right| \\
&\leq \sup_{\|\mathbf{z}\|=1,\|\boldsymbol{\beta}\|\leq\delta} \frac{1}{n}\sum_{i=1}^{n} (\mathbf{x}_i^T\mathbf{z})^2 \left|K_i^n(\mathbf{x}_i^T\boldsymbol{\beta}) - K_i(\mathbf{x}_i^T\boldsymbol{\beta})\right| \\
&\leq \sup_{\|\mathbf{z}\|=1,\|\boldsymbol{\beta}\|\leq\delta} \|\varsigma_1\|_\infty \max_{i\leq n} \frac{1}{\left|s^*_{i,n}(\boldsymbol{\beta})\right|} |\hat{\sigma}_n - s_0| \frac{1}{n}\sum_{i=1}^{n} (\mathbf{x}_i^T\mathbf{z})^2 \\
&\leq \left(\|\varsigma_1\|_\infty |\hat{\sigma}_n - s_0| \sup_n \gamma_{2,n}\right) \sup_{\|\boldsymbol{\beta}\|\leq\delta} \max_{i\leq n} \frac{1}{|s^*_{i,n}(\boldsymbol{\beta})|} \xrightarrow{P} 0.
\end{aligned}$$

The conclusion of the lemma now follows easily. □

Proof of Theorem 3. A first order Taylor expansion shows that there exists a sequence of numbers ζ_i, $1 \leq i \leq n$, satisfying $0 \leq \zeta_i \leq 1$, such that

$$\begin{aligned}
&0 \geq \frac{1}{n}(L_n(\hat{\boldsymbol{\beta}}) - L_n(\boldsymbol{\beta}_0)) = -\frac{1}{n\hat{\sigma}_n}\sum_{i=1}^{n} \psi_1\left(\frac{u_i}{\hat{\sigma}_n}\right) \mathbf{x}_i^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&+ \frac{1}{2}\frac{1}{\hat{\sigma}_n^2}\frac{1}{n}\sum_{i=1}^{n} \psi_1'\left(\frac{u_i - \zeta_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{x}_i}{\hat{\sigma}_n}\right)(\mathbf{x}_i^{\mathrm{T}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0))^2 \\
&= A_n + B_n.
\end{aligned}$$

Since by assumption $\hat{\sigma}_n \xrightarrow{P} s_0$, with arbitrarily high probability, for large enough n, we have that $\hat{\sigma}_n \in [s_0/2, 2s_0]$, and hence

$$\begin{aligned}
|A_n| &= \frac{1}{\hat{\sigma}_n}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \left| \frac{1}{n}\sum_{i=1}^{n} \psi_1\left(\frac{u_i}{\hat{\sigma}_n}\right) \mathbf{x}_i^T \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)}{\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|} \right| \\
&= \frac{1}{\hat{\sigma}_n}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \left| \left(\frac{1}{n}\sum_{i=1}^{n} \psi_1\left(\frac{u_i}{\hat{\sigma}_n}\right) \mathbf{x}_i^T\right) \left(\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)}{\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|}\right) \right| \\
&\leq \frac{1}{\hat{\sigma}_n}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \left\| \frac{1}{n}\sum_{i=1}^{n} \psi_1\left(\frac{u_i}{\hat{\sigma}_n}\right) \mathbf{x}_i \right\| \\
&\leq \frac{1}{\hat{\sigma}_n}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \frac{1}{\sqrt{n}} \sup_{s \in [s_0/2, 2s_0]} \left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_1\left(\frac{u_i}{s}\right) \mathbf{x}_i \right\|.
\end{aligned}$$

By Lemma 5,

$$E \sup_{s \in [s_0/2, 2s_0]} \left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_1\left(\frac{u_i}{s}\right) \mathbf{x}_i \right\| \leq C\sqrt{p},$$

for a fixed constant C. Hence, by Markov's inequality

$$\sup_{s \in [s_0/2, 2s_0]} \left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_1\left(\frac{u_i}{s}\right) \mathbf{x}_i \right\| = O_P(\sqrt{p}).$$

Also, since by assumption $\hat{\sigma}_n \xrightarrow{P} s_0 > 0$, we have that $1/\hat{\sigma}_n = O_P(1)$. We have thus shown that

$$A_n = \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| O_P\left(\sqrt{\frac{p}{n}}\right). \tag{S2.20}$$

Let δ and a^* be as in Lemma 7. Since $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \xrightarrow{P} 0$, for sufficiently large n, with arbitrarily high probability we have that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| < \delta$ and hence

$$
\begin{aligned}
B_n &\geq \frac{1}{2\,\hat{\sigma}_n^2}\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0))^2 \inf_{|v|\leq|\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)|}\psi_1'\left(\frac{u_i-v}{\hat{\sigma}_n}\right)\\
&= \frac{1}{2\,\hat{\sigma}_n^2}\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0))^2 K_i^n(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0))\\
&= \frac{1}{2\,\hat{\sigma}_n^2}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^T\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)}{\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|}\right)^2 K_i^n(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0))\\
&\geq \frac{1}{2\,\hat{\sigma}_n^2}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2 \inf_{\|\mathbf{z}\|=1,\|\boldsymbol{\beta}\|\leq\delta}\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i^T\mathbf{z})^2 K_i^n(\mathbf{x}_i^T\boldsymbol{\beta}).
\end{aligned}
$$

By Lemma 7, for sufficiently large n, with arbitrarily high probability

$$
\frac{1}{2\,\hat{\sigma}_n^2}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2 \inf_{\|\mathbf{z}\|=1,\|\boldsymbol{\beta}\|\leq\delta}\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i^T\mathbf{z})^2 K_i^n(\mathbf{x}_i^T\boldsymbol{\beta}) \geq \frac{a^*}{2\,\hat{\sigma}_n^2}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2. \tag{S2.21}
$$

Hence, it follows from (S2.20) and (S2.21) that with arbitrarily high probability, for large enough n and some positive constants M_1 and M_2, $0 \geq A_n + B_n \geq -M_1\sqrt{p/n}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\| + M_2\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2$. Then, $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\| \leq (M_1/M_2)\sqrt{p/n}$, which proves the theorem. □

The following lemma is needed in the proof of Theorem 4.

Lemma 8. *Assume R2, F0, X1, X2, X3 and X6 hold. Let* $\mathbf{a}_n \in \mathbb{R}^p$, $\|\mathbf{a}_n\| = 1$. *Let* $r_n^2 = \mathbf{a}_n^T\boldsymbol{\Sigma}_n^{-1}\mathbf{a}_n$. *Then*

a)

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\psi_1\left(\frac{u_i}{\hat{\sigma}_n}\right)-\psi_1\left(\frac{u_i}{s_0}\right)\right)\left(\mathbf{a}_n^T\boldsymbol{\Sigma}_n^{-1}\mathbf{x}_i\right) \xrightarrow{P} 0.
$$

b)

$$\frac{1}{r_n\sqrt{n}}\sum_{i=1}^{n}\psi_1\left(\frac{u_i}{s_0}\right)\mathbf{a}_n^T\boldsymbol{\Sigma}_n^{-1}\mathbf{x}_i \xrightarrow{d} N\left(0, E\psi_1^2\left(\frac{u}{s_0}\right)\right).$$

Proof. We first prove a). For $t \in [0,1]$ let

$$F_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_1\left(\frac{u_i}{0.5s_0 + ts_0}\right)\mathbf{a}_n^T\boldsymbol{\Sigma}_n^{-1}\mathbf{x}_i.$$

Since by assumption $\hat{\sigma}_n \xrightarrow{P} s_0$, it suffices to show that $(F_n)_n$ is a tight sequence in $C[0,1]$. By Theorem 12.3 of Billingsley [1968], it suffices to show that

(i) $F_n(0)$ is tight

(ii) There exists $\delta \geq 0$, $\alpha > 1$ and a nondecreasing, continuous function f on $[0,1]$, such that for any $0 \leq t_1 \leq t_2 \leq 1$ and any $\lambda > 0$ we have

$$\mathbb{P}\left(|F_n(t_2) - F_n(t_1)| \geq \lambda\right) \leq \frac{1}{\lambda^\delta}\left(f(t_2) - f(t_1)\right)^\alpha \text{ for all } n.$$

We first prove (i). Let $h_n^2 = E\psi_1^2\left(u/(0.5s_0)\right)\mathbf{a}_n^T\boldsymbol{\Sigma}_n^{-1}\mathbf{a}_n$. By X1, X3 and Lemma 1, $\inf_n \gamma_{1,n} > 0$. This together with X2 implies that h_n and $1/h_n$ are bounded. Note that since ψ_1 is odd and the errors have a symmetric distribution, $E\psi_1\left(u/(0.5s_0)\right) = 0$. Also,

$$\sum_{i=1}^{n}E\left(\frac{1}{\sqrt{n}}\psi_1\left(\frac{u}{0.5s_0}\right)\mathbf{a}_n^T\boldsymbol{\Sigma}_n^{-1}\mathbf{x}_i\right)^2 = h_n^2.$$

Note that by X6 $\max_{i\leq n}(\mathbf{a}_n^T\mathbf{\Sigma}_n^{-1}\mathbf{x}_i)/(\sqrt{n}h_n)\to 0$. Then for any fixed $\varepsilon>0$,

$$\sum_{i=1}^{n} E\left(\frac{1}{\sqrt{n}h_n}\psi_1\left(\frac{u}{0.5s_0}\right)\mathbf{a}_n^T\mathbf{\Sigma}_n^{-1}\mathbf{x}_i\right)^2 I\left\{\left|\psi_1\left(\frac{u}{0.5s_0}\right)(\mathbf{a}_n^T\mathbf{\Sigma}_n^{-1}\mathbf{x}_i)/(\sqrt{n}h_n)\right|>\varepsilon\right\} \to 0.$$

Hence, by the Lindberg-Feller Theorem, $F_n(0)/h_n \xrightarrow{d} N(0,1)$ and (i) follows. Note that roughly the same argument proves b).

Now, we prove (ii). By Tchebyshev's inequality, it suffices to show that there exists $M>0$ such that for all t_1,t_2 in $[0,1]$, $E(F_n(t_1)-F_n(t_2))^2\leq M(t_2-t_1)^2$ for all n. Let

$$\Delta_i(t_1,t_2)=\psi_1\left(\frac{u_i}{0.5s_0+t_1s_0}\right)-\psi_1\left(\frac{u_i}{0.5s_0+t_2s_0}\right).$$

Note that $E\Delta_i(t_1,t_2)=0$ for all t_1,t_2 and i. Using the independence of $u_1,\ldots,u_n$, we get

$$\begin{aligned}E(F_n(t_1)-F_n(t_2))^2 &= \frac{1}{n}\sum_{i,j}E\Delta_i(t_1,t_2)\Delta_j(t_1,t_2)(\mathbf{a}_n^T\mathbf{\Sigma}_n^{-1}\mathbf{x}_i)(\mathbf{a}_n^T\mathbf{\Sigma}_n^{-1}\mathbf{x}_j)\\ &= E\Delta_1(t_1,t_2)^2\frac{1}{n}\sum_{i=1}^{n}(\mathbf{a}_n^T\mathbf{\Sigma}_n^{-1}\mathbf{x}_i)^2\\ &= E\left(\psi_1\left(\frac{u}{0.5s_0+t_1s_0}\right)-\psi_1\left(\frac{u}{0.5s_0+t_2s_0}\right)\right)^2\mathbf{a}_n^T\mathbf{\Sigma}_n^{-1}\mathbf{a}_n.\end{aligned} \tag{S2.22}$$

Let $\chi_1(t)=\psi_1'(t)t$. By R2 χ_1 is bounded. Applying the Mean Value

Theorem we get that

$$\left|\psi_1\left(\frac{u}{0.5s_0+t_1s_0}\right)-\psi_1\left(\frac{u}{0.5s_0+t_2s_0}\right)\right|\leq 2\|\chi_1\|_\infty\left|(t_1-t_2)\right|.$$

Hence, for some fixed constant $C>0$

$$E\left(\psi_1\left(\frac{u}{0.5s_0+t_1s_0}\right)-\psi_1\left(\frac{u}{0.5s_0+t_2s_0}\right)\right)^2\leq C(t_2-t_1)^2.$$

Since $\inf_n\gamma_{1,n}>0$, from (S2.22) it follows that (ii) holds and thus the lemma is proven. $\square$

The following lemma is needed in the proof of Theorem 4. Its proof is very similar to that of Lemma 5 and for this reason it is ommitted.

Lemma 9. *Assume R2, F0 and X1 a) hold. Let $0<a<c$. Then for some fixed constant $A>0$ that depends only on a,c, ψ_1' and the constant that appears in X1 a),*

$$E\sup_{s\in[a,c]}\left\|(\frac{1}{\sqrt{n}})\sum_{i=1}^n\left(\psi_1'\left(\frac{u_i}{s}\right)-E\psi_1'\left(\frac{u}{s}\right)\right)\mathbf{x}_i\mathbf{x}_i^T\right\|_F\leq A\sqrt{p}\max_{i\leq n}\|\mathbf{x}_i\|,$$

where $\|.\|_F$ is the Frobenius norm.

Proof of Theorem 4. From the definition of $\hat{\boldsymbol{\beta}}$, (2.4), it follows that

$$\mathbf{0}_p=\frac{1}{\sqrt{n}}\frac{-1}{\hat{\sigma}_n}\sum_{i=1}^n\psi_1\left(\frac{y_i-\mathbf{x}_i^T\hat{\boldsymbol{\beta}}}{\hat{\sigma}_n}\right)\mathbf{x}_i.$$

Then the Mean Value Theorem gives

$$\mathbf{0}_p=\frac{1}{\sqrt{n}}\frac{-1}{\hat{\sigma}_n}\sum_{i=1}^n\psi_1\left(\frac{u_i}{\hat{\sigma}_n}\right)\mathbf{x}_i+\frac{1}{\hat{\sigma}_n^2}\mathbf{W}_n\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0),$$

where

$$\mathbf{W}_n = \frac{1}{n}\sum_{i=1}^{n} \psi_1' \left(\frac{u_i - \zeta_i \mathbf{x}_i^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)}{\hat{\sigma}_n} \right) \mathbf{x}_i \mathbf{x}_i^T$$

and $0 \le \zeta_i \le 1$. Let

$$\mathbf{W}_n^1 = \frac{1}{n}\sum_{i=1}^{n} \psi_1' \left(\frac{u_i}{\hat{\sigma}_n} \right) \mathbf{x}_i \mathbf{x}_i^T, \;\; \mathbf{W}_n^2 = E\psi_1' \left(\frac{u}{\hat{\sigma}_n} \right) \boldsymbol{\Sigma}_n,$$

where the expectation in $E\psi_1'\left(u/\hat{\sigma}_n\right)$ is taken only with respect to u. Then

$$\begin{aligned}
\sqrt{n}\mathbf{a}_n^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \frac{\hat{\sigma}_n}{E\psi_1'\left(u/\hat{\sigma}_n\right)} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_1 \left(\frac{u_i}{\hat{\sigma}_n} \right) \mathbf{a}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{x}_i \\
&- \frac{1}{E\psi_1'\left(u/\hat{\sigma}_n\right)} \mathbf{a}_n^T \boldsymbol{\Sigma}_n^{-1} (\mathbf{W}_n - \mathbf{W}_n^1) \sqrt{n} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&- \frac{1}{E\psi_1'\left(u/\hat{\sigma}_n\right)} \mathbf{a}_n^T \boldsymbol{\Sigma}_n^{-1} (\mathbf{W}_n^1 - \mathbf{W}_n^2) \sqrt{n} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&= \frac{\hat{\sigma}_n}{E\psi_1'\left(u/\hat{\sigma}_n\right)} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_1 \left(\frac{u_i}{\hat{\sigma}_n} \right) \mathbf{a}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{x}_i \\
&+ A_n + B_n.
\end{aligned}$$

We will show that $A_n + B_n = o_P(1)$. Note that by R2 and the Bounded Convergence Theorem, $E\psi_1'\left(u/\hat{\sigma}_n\right) \xrightarrow{P} E\psi_1'\left(u/s_0\right)$.

For a matrix $\mathbf{W}$ let $\|\mathbf{W}\|$ be its spectral norm and let $\|\mathbf{W}\|_F$ be its Frobenius norm. Recall that for any $\mathbf{W}$, $\|\mathbf{W}\| \le \|\mathbf{W}\|_F$. We will show that $\|\mathbf{W}_n - \mathbf{W}_n^1\| = o_P(1/\sqrt{p})$ and $\|\mathbf{W}_n^1 - \mathbf{W}_n^2\| = o_P(1/\sqrt{p})$. Take $\boldsymbol{\theta} \in \mathbb{R}^p$ with

$\|\boldsymbol{\theta}\| = 1$. Then, applying the Mean Value Theorem, we get

$$\begin{aligned}
|\boldsymbol{\theta}^T(\mathbf{W}_n - \mathbf{W}_n^1)\boldsymbol{\theta}| &\leq \frac{1}{n}\sum_{i=1}^{n}\left|\psi_1'\left(\frac{u_i}{\hat{\sigma}_n}\right) - \psi_1'\left(\frac{u_i - \zeta_i \mathbf{x}_i^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)}{\hat{\sigma}_n}\right)\right|(\boldsymbol{\theta}^T\mathbf{x}_i)^2 \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\frac{\|\psi_1''\|_\infty}{\hat{\sigma}_n}|\mathbf{x}_i^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)|(\boldsymbol{\theta}^T\mathbf{x}_i)^2 \\
&\leq \frac{\|\psi_1''\|_\infty}{\hat{\sigma}_n}\max_{i\leq n}\|\mathbf{x}_i\|\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|\sup_n \gamma_{2,n}.
\end{aligned}$$

Since $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_P(\sqrt{p/n})$, taking supremum over $\boldsymbol{\theta}$, from X6 it follows that $\|\mathbf{W}_n - \mathbf{W}_n^1\| = o_P\left(1/\sqrt{p}\right)$ and hence we have that $A_n = o_P(1)$. By Lemma 9 and X6,

$$\|\mathbf{W}_n^1 - \mathbf{W}_n^2\|_F = O_P\left(\sqrt{p/n}\max_{i\leq n}\|\mathbf{x}_i\|\right) = o_P\left(\frac{1}{\sqrt{p}}\right)$$

and hence we have that $B_n = o_P(1)$.

We have thus shown that $A_n + B_n = o_P(1)$ and so it follows that

$$\sqrt{n}\mathbf{a}_n^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \frac{\hat{\sigma}_n}{E\psi_1'\left(u/\hat{\sigma}_n\right)}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_1\left(\frac{u_i}{\hat{\sigma}_n}\right)\mathbf{a}_n^T\boldsymbol{\Sigma}_n^{-1}\mathbf{x}_i + o_P(1).$$

Note that since by assumptions X1, X2 and X3, $\inf_n \gamma_{1,n} > 0$ and $\sup_n \gamma_{2,n} < \infty$, r_n and $1/r_n$ are bounded. By Lemma 8

$$r_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_1\left(\frac{u_i}{\hat{\sigma}_n}\right)\mathbf{a}_n^T\boldsymbol{\Sigma}_n^{-1}\mathbf{x}_i \xrightarrow{d} N(0, a(\psi_1)).$$

The theorem now follows from Slutzky's Theorem. □

Bibliography

P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

L. Davies. The asymptotics of S-estimators in the linear regression model. *Ann. Statist.*, 18(4):1651–1675, 1990.

M. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer, 2008.

D. Pollard. Asymptotics via empirical processes. *Statist. Sci.*, 4(4):341–354, 1989.

S. Portnoy. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. consistency. *Ann. Statist.*, 12(4):1298–1309, 1984.

A. W. van der vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag New York, New York, 1996.

V. J. Yohai. High Breakdown Point and High Efficiency Robust Estimates for Regression. Technical Report 66, University of Washington, 1985. Available at `http://www.stat.washington.edu/research/reports/1985/tr066.pdf`.