Supplemental material for: Projection Estimators for Generalized Linear Models

Andrea Bergesio and Víctor Yohai

1 Proof of Theorem 1

Put $\lambda^* = \mathbf{C}'^{-1}\lambda$, then

$$S(\mathcal{L}(\lambda^{*\prime}\mathbf{C}\mathbf{x}, H)) = S(\mathcal{L}(\lambda'\mathbf{x}, H))$$

and

$$T_0(\mathcal{L}((y, \lambda^{*\prime}\mathbf{C}\mathbf{x}, \beta'\mathbf{x}), H)) = T_0(\mathcal{L}((y, \lambda'\mathbf{x}, \beta'\mathbf{x}), H)).$$

Therefore

$$A(\mathbf{C}'^{-1}\beta,\mathcal{L}((y, \mathbf{C}\mathbf{x}), H)) = A(\beta,H),$$

and then the Theorem holds.

2 Proof of Theorem 2

According to (8) we have

$$A(\beta_0, H_0) = 0. \tag{36}$$

On the other hand take $\beta\neq\beta_0$ and put $\lambda_0 = (\beta_0-\beta)/||\beta_0-\beta||$ and $\alpha_0 = ||\beta-\beta_0||$. Then $\beta_0'\mathbf{x}=\alpha_0\lambda_0'\mathbf{x}+\beta'\mathbf{x}$ and this implies that $(y, \lambda_0'\mathbf{x},\beta'\mathbf{x})$ follows model (6)-(7). Then (8) implies that

$$T_0(\mathcal{L}((y, \lambda_0'\mathbf{x}, \beta'\mathbf{x}),H_0) = \alpha_0 = ||\beta - \beta_0|| > 0, \tag{37}$$

and since $S(\mathcal{L}(\lambda'\mathbf{x}, G_0)) > 0$, we obtain

$$A(\beta,H_0) > 0. \tag{38}$$

Then from (36),(38) and (14) we get that $\mathbf{T}(H_0) =\beta_0$.

3 Proof of Theorem 3

Since both sides of (18) are affine invariant, we can assume that $V(G_0)$ is the $p \times p$ identity matrix.

Take β such that

$$||\beta - \beta_0|| > \left(\mathbf{1}+\frac{d^+(G_0, \varepsilon)}{d^-(G_0, \varepsilon)}\right) \sup_{||\lambda||=1} B(T_0, \lambda, \beta_0,G_0, \varepsilon). \tag{39}$$

Let $H \in \mathcal{V}(H_0, \varepsilon)$. We have to show that

$$\mathbf{T}(H) \neq \beta. \tag{40}$$

We have

$$\begin{aligned} A(\beta_0, H) &\leq \sup_{||\lambda||=1} \sup_{H \in \mathcal{V}(H_0,\varepsilon)} |T_0(\mathcal{L}((y, \lambda'\mathbf{x}, \beta_0'\mathbf{x}),H)|d^+(G_0, \varepsilon) \\ &\leq \sup_{||\lambda||=1} \sup_{M \in \mathcal{V}(M_{\lambda,\beta_0},\varepsilon)} |T_0(M)|d^+(G_0, \varepsilon) \\ &\leq \sup_{||\lambda||=1} B(T_0, \lambda, \beta_0, G_0, \varepsilon) d^+(G_0, \varepsilon). \end{aligned} \tag{41}$$

Put $\lambda_0 = (\beta_0 - \beta)/||\beta_0 - \beta||$ and $\alpha_0 = ||\beta - \beta_0||$. Then using (9)-(10) we obtain

$$\begin{aligned} T_0(\mathcal{L}((y, \lambda_0'\mathbf{x}, \beta'\mathbf{x}),H)) &= ||\beta - \beta_0|| T_0(\mathcal{L}((y, (\beta_0 - \beta)'\mathbf{x}, \beta'\mathbf{x}),H)) \\ &= ||\beta - \beta_0|| T_0(\mathcal{L}((y, (\beta_0 - \beta)'\mathbf{x}, (\beta_0 - (\beta_0 - \beta))'\mathbf{x}),H)) \\ &= ||\beta - \beta_0|| (T_0(\mathcal{L}((y, (\beta_0 - \beta)'\mathbf{x}, \beta_0'\mathbf{x}),H)) + 1) \\ &= T_0(\mathcal{L}((y, \lambda_0'\mathbf{x}, \beta_0'\mathbf{x}),H)) + ||\beta - \beta_0||, \end{aligned}$$

and therefore by (39) we get

$$\begin{aligned} |T_0(\mathcal{L}((y, \lambda'\mathbf{x}, \beta'\mathbf{x}),H))| &\geq ||\beta - \beta_0|| - \sup_{||\lambda||=1} B(T_0, \lambda, \beta_0, G_0, \varepsilon) \\ &> \left(\mathbf{1} + \frac{d^+(G_0, \varepsilon)}{d^-(G_0, \varepsilon)}\right) \sup_{||\lambda||=1} B(T_0, \lambda, \beta_0, G_0, \varepsilon) \\ &- \sup_{||\lambda||=1} B(T_0, \lambda, \beta_0, G_0, \varepsilon). \end{aligned} \tag{42}$$

Then, we obtain

$$A(\beta, H) > d^+(G_0, \varepsilon) \sup_{||\lambda||=1} B(T_0, \lambda, \beta_0, G_0, \varepsilon). \tag{43}$$

Inequalities (41) and (43) imply (40).

4 Proof of Theorem 4

Let $\mathbf{U}_n = n^{1/2}(\mathbf{T}(H_n) - \beta_0)$, $k_n = ||\mathbf{U}_n||$, $\lambda_n = \mathbf{U}_n/||k_n||$. We have to prove that k_n is bounded in probability. We can also write

$$\mathbf{T}(H_n) = \beta_0 + k_n \lambda_n / n^{1/2}. \tag{44}$$

By (19) we have

$$n^{1/2} T_0 \left(\mathcal{L}\left(\left(y, \lambda_n'\mathbf{x}, \beta_0'\mathbf{x}\right), H_n\right)\right) = O_p(1).$$

Then by (20), (21) and the definition of $\mathbf{T}$ we obtain

$$n^{1/2}T_0\left(\mathcal{L}\left((y,\lambda_n'\mathbf{x},\mathbf{T}(H_n)'\mathbf{x}),H_n\right)\right)=O_p(1),$$

and then using (44) we can write

$$V_n=n^{1/2}T_0\left(\mathcal{L}\left(\left(y,\lambda_n'\mathbf{x},\left(\beta_0+\frac{k_n}{n^{1/2}}\lambda_n\right)'\mathbf{x}\right),H_n\right)\right)=O_p(1). \tag{45}$$

However, because of (10) we have

$$V_n=n^{1/2}T_0\left(\mathcal{L}\left(\left(y,\lambda_n'\mathbf{x},\beta_0'\mathbf{x}\right),H_n\right)\right)-k_n,$$

and then

$$k_n=n^{1/2}T_0(\mathcal{L}((y,\lambda_n'\mathbf{x},\beta_0'\mathbf{x}),H_n)-V_n.$$

Then because of (19) and (45), k_n is bounded in probability.

5 Proof of Theorem 5

In order to simplify the proof of Theorem 5, in this Section we omit $S(L)$ in the definition of the initial estimating functionals given in (22). The proof when the scale is included is essentially the same but it requires to take care of some cumbersome technical details.

5.1 Uniform consistency of the initial estimator

In Lemma 2 we prove a uniform convergence property of the initial estimator T_0 given by (22). The following Lemma is required to prove Lemma 2.

Lemma 1. Assume P1-P4, then

(i) Let $r(\theta_1,\theta_2)=E_{\theta_2}\left[\eta(y-\delta(\theta_1))\right]$. Then $\theta_1<\theta_2$ implies $r(\theta_1,\theta_2)>0$ and $\theta_1>\theta_2$ implies $r(\theta_1,\theta_2)<0$.

(ii) The function $\delta(\theta)$ is strictly increasing.

(iii) $\delta(\theta)$ is continuous.

(iv) The function $\delta(\theta)$ is continuously differentiable.

(v) For all $\alpha>0$ we have

$$E_{H_0}\left[\eta(y-\delta(g^{-1}(\alpha\lambda'\mathbf{x}+\beta_0'\mathbf{x})))\kappa(\lambda'\mathbf{x})\right]<0 \tag{46}$$

and

$$E_{H_0}\left[\eta(y-\delta(g^{-1}(-\alpha\lambda'\mathbf{x}+\beta_0'\mathbf{x})))\kappa(\lambda'\mathbf{x})\right]>0. \tag{47}$$

Proof. Assume first $\theta_1<\theta_2$. As in Lemma 3 of Bianco, García Ben and Yohai (2005), we can prove that $F_{\theta_1}(y)>F_{\theta_2}(y)$ for all y. Let $\theta_1<\theta_2$, then $r(\theta_1,\theta_2)>0$ follows from the following two facts (a) $r(\theta_1,\theta_1)=0$ and (b) there exists $c>0$ such that $P_{\theta_1}(|y-\delta(\theta_1)|\leq c)>0$ and $\eta(u)$ is strictly increasing for

$|u| \leq c$. The proof in the case $\theta_1 > \theta_2$ is similar. This proves (i), and (ii) is a direct consequence of (i).

Parts (iii) and (iv) are proved here only for the case of a continuous family of distributions. The proofs for the discrete case are similar. To prove the continuity of $\delta(\theta)$ is enough to show the left and right continuity.We only show the right continuity of $\delta(\theta)$ and the proof of left continuity is completely similar. Let θ_n be a non-increasing sequence such that $\theta_n \to \theta$, then $\delta(\theta_n)$ is a non-increasing sequence such that $\delta(\theta_n) \to a$. Then to prove that $\delta(\theta)$ is right continuous is enough to show that $\delta(\theta) = a$. According to the definition of $\delta(\theta)$ we have

$$E_{\theta_n}\left[\eta(y - \delta(\theta_n))\right] = 0,$$

or equivalently

$$\int_D \eta(y - \delta(\theta_n)) f(y, \theta_n) dy = 0. \tag{48}$$

Since $f(y, \theta_n)$ is the density of an exponential family and η is bounded, we can find an integrable function $f^*(y)$ such that

$$\sup_n |\eta(y - \delta(\theta_n))| \, f(y, \theta_n) \leq f^*(y)$$

and therefore by the Dominated Convergence Theorem we can take limits in (48) inside the integral obtaining

$$\int_D \eta(y - a) f(y, \theta) dy = 0 \tag{49}$$

proving that $\delta(\theta) = a$.

To prove (iv) note that according to (23) we have

$$E_{\theta+h}\left[\eta(y - \delta(\theta + h)\right] = \int_D \eta(y - \delta(\theta + h)) f(y, \theta + h) dy = 0. \tag{50}$$

Applying the Mean Value Theorem (MVT) we get

$$\begin{aligned}
&\int_D \eta(y - \delta(\theta + h)) f(y, \theta + h) dy \\
&= \int_D \eta(y - \delta(\theta)) f(y, \theta + h) dy \\
&- (\delta(\theta + h) - \delta(\theta)) \int_D \eta'(y - \delta^*(y, h)) f(y, \theta + h) dy,
\end{aligned}$$

where $|\delta^*(y, h) - \delta(\theta)| \leq |\delta(\theta + h)) - \delta(\theta)|$. Therefore

$$\delta(\theta + h) - \delta(\theta) = \frac{\int_D \eta(y - \delta(\theta)) f(y, \theta + h) dy}{\int_D \eta'(y - \delta^*(y, h)) f(y, \theta + h) dy}. \tag{51}$$

Using the MVT theorem again, (23) and part (i) of this lemma we get

$$
\begin{aligned}
0 < R(\theta,\theta+h) &= \int_D \eta(y-\delta(\theta))f(y,\theta+h)dy \\
&= \int_D \eta(y-\delta(\theta))f(y,\theta)dy + h\int_D \eta(y-\delta(\theta))\left.\frac{\partial f(y,\theta)}{\partial\theta}\right|_{\theta=\theta^*(y,h)} dy \\
&= h\int_D \eta(y-\delta(\theta))\left.\frac{\partial f(y,\theta)}{\partial\theta}\right|_{\theta=\theta^*(y,h)} dy, \qquad (52)
\end{aligned}
$$

where $|\theta^*(y,h)-\theta|\leq h$. Then from (51) and (52) we obtain

$$
\frac{\delta(\theta+h)-\delta(\theta)}{h} = \frac{\int_D \eta(y-\delta(\theta))\left.\frac{\partial f(y,\theta)}{\partial\theta}\right|_{\theta=\theta^*(y,h)} dy}{\int_D \eta'(y-\delta^*(y,h))f(y,\theta+h)dy}. \qquad (53)
$$

Note that P1 implies

$$
\int_D \eta'(y-\delta(\theta))f(y,\theta)dy > 0
$$

and then from (53) we derive

$$
\delta'(\theta) = \lim_{h\to 0}\frac{\delta(\theta+h))-\delta(\theta)}{h} = \frac{\int_D \eta(y-\delta(\theta))\frac{\partial f(y,\theta)}{\partial\theta}dy}{\int_D \eta'(y-\delta(\theta))f(y,\theta)dy}
$$

which is a continuous function. This proves part (iii).

Equation (46) can be written as

$$
E_{G_0}\left[r(g^{-1}(\alpha\lambda'\mathbf{x}+\beta_0'\mathbf{x}),g^{-1}(\beta_0'\mathbf{x}))\kappa(\lambda'\mathbf{x})\right] < 0. \qquad (54)
$$

Note that Lemma 1 (i), P2 and P4 imply that

$$
r(g^{-1}(\alpha\lambda'\mathbf{x}+\beta_0'\mathbf{x}),g^{-1}(\beta_0'\mathbf{x}))\kappa(\lambda\mathbf{x}) \leq 0
$$

and

$$
r(g^{-1}(\alpha\lambda'\mathbf{x}+\beta_0'\mathbf{x}),g^{-1(}\beta_0'\mathbf{x}))\kappa(\lambda\mathbf{x}) < 0
$$

when $|\lambda'\mathbf{x}|>0$. Then (54) follows from P3. The proof of (47) is similar.

Lemma 2. *Assume P1-P3. Consider a random sample $(y_1,\mathbf{x}_1),\ldots,(y_n,\mathbf{x}_n)$ from model (1)-(3), where $y|\mathbf{x}$ is F_θ with $g(\theta)=\beta_0'\mathbf{x}_i$. For any $\lambda\in R^p$ let $\widehat{\alpha}_n(\lambda)$ be the estimating functional T_0 given in (22) applied to $(y_i,\lambda_i'\mathbf{x},\beta_0'\mathbf{x}_i)$, $1\leq i\leq n$. Then*

$$
\sup_{||\lambda||=\mathbf{1}}|\widehat{\alpha}_n(\lambda)| \longrightarrow 0. \text{ a.s..}
$$

Proof. Note that $\widehat{\alpha}_n(\lambda)$ is the value of α satisfying

$$
\sum_{i=1}^{b}\eta(y_i-\delta(g^{-1}(\alpha\lambda'\mathbf{x_i}+\beta_0'\mathbf{x_i})))\kappa(\lambda'\mathbf{x_i}) = 0.
$$

Then, to prove the Lemma it suffices to show that for all $\alpha_0 > 0$, there exists $c > 0$ such that

$$\text{limsup}_{n\to\infty} \sup_{\alpha>\alpha_0, ||\lambda||=1} \frac{1}{n}\sum_{i=1}^{n} \eta(y_i - \delta(g^{-1}(\alpha\lambda'\mathbf{x_i} + \beta_0'\mathbf{x_i})))\kappa(\lambda'\mathbf{x_i}) \leq -c \quad (55)$$

and

$$\text{liminf}_{n\to\infty} \inf_{\alpha<-\alpha_0, ||\lambda||=1} \frac{1}{n}\sum_{i=1}^{n} \eta(y_i - \delta(g^{-1}(\alpha\lambda'\mathbf{x_i} + \beta_0'\mathbf{x_i})))\kappa(\lambda'\mathbf{x_o}) \geq c. \quad (56)$$

Since the proofs of (55) and (56) are similar we only prove (55).

Note that P1, P2,P4 and part (ii) of Lemma 1 imply that

$$\eta(y_i - \delta(g^{-1}(\alpha\lambda'\mathbf{x} + \beta_0'\mathbf{x})))\kappa(\lambda'\mathbf{x})$$

is non-increasing in α and therefore we have

$$\begin{aligned}
&\sup_{\alpha>\alpha_0, ||\lambda||=1} \frac{1}{n}\sum_{i=1}^{n} \eta(y_i - \delta(g^{-1}(\alpha\lambda'\mathbf{x}_i + \beta_0'\mathbf{x}_i)))\kappa(\lambda'\mathbf{x_i}) \\
&= \sup_{|||\lambda||=1} \frac{1}{n}\sum_{i=1}^{n} \eta(y_i - \delta(g^{-1}(\alpha_0\lambda'\mathbf{x}_i + \beta_0'\mathbf{x}_i)))\kappa(\lambda'\mathbf{x}_i). \quad (57)
\end{aligned}$$

Then in order to prove (55) it is enough to show that

$$\text{limsup}_{n\to\infty} \sup_{|||\lambda||=1} \frac{1}{n}\sum_{i=1}^{n} \eta(y_i - \delta(g^{-1}(\alpha_0\lambda'\mathbf{x}_i + \beta_0'\mathbf{x}_i)))\kappa(\lambda'\mathbf{x}_i) \leq -c. \quad (58)$$

From Lemma 3 of Muler and Yohai (2002) we get that

$$\begin{aligned}
&\text{limsup}_{n\to\infty} \sup_{||\lambda||=1} \frac{1}{n}\sum_{i=1}^{n} \eta(y_i - \delta(g^{-1}(\alpha_0\lambda'\mathbf{x}_i + \beta_0'\mathbf{x}_i)))\kappa(\lambda'\mathbf{x}_i) \\
&\leq \sup_{||\lambda||=1} E_{H_0}\left[\eta(y_i - \delta(g^{-1}(\alpha_0\lambda'\mathbf{x}_i + \beta_0'\mathbf{x}_i)))\kappa(\lambda'\mathbf{x}_i)\right].
\end{aligned}$$

The Dominated Convergence Theorem implies that

$$a(\lambda) = E_{H_0}\left[\eta(y - \delta(g^{-1}(\alpha_0\lambda'\mathbf{x} + \beta_0'\mathbf{x})))\kappa(\lambda'\mathbf{x})\right]$$

is a continuous function of λ. Since by part (iv) of Lemma 1 we have $a(\lambda) < 0$, (58) follows.

5.2 Consistency order of the initial estimator

Given a compact metric space S with a metric d, we denote by $C(S)$ the space of the real continuous functions defined on S and $C(S)^*$, the space of linear and continuous functionals defined on $C(S)$. We denote by $N_d(S,u)$ the smallest number of spheres of radio equal or less than μ that cover S.

The following Theorem is proved in Jain and Marcus (1975).

Theorem 10. *Let S be a compact metric space with a metric d. Let X be a random variable with values in $C(S)$ the space of all continuous functions on S satisfying*

$$E\left[f(X)\right] = 0 \text{ for all } f \in C(S)^* \tag{59}$$

and such that

$$\sup_{\lambda \in S} E[X^2(\lambda)] < \infty. \tag{60}$$

Assume also that there exists a non-negative random variable M such that $E[M^2] < \infty$ and a metric d^* in S, which is continuous with respect to d such that given $s,t \in S$

$$|X(s) - X(t)| \leq M d^*(s,t). \tag{61}$$

Assume also

$$\sum_{n=1}^{\infty} 2^{-n} \sqrt{\ln N_{d^*}(S, 2^{-n})} < \infty. \tag{62}$$

Then if $X_1, \ldots, X_n$ are i.i.d. random variables with the distribution of X, we have

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} X_i$$

converges in distribution to a Gaussian process in $C(S)$.

Lemma 3. *Assume P1-P4 and let $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$ be a random sample from model (1)-(3). Then*

$$\sqrt{n} \sup_{\|\lambda\|=1} \left| \frac{1}{n} \sum_{i=1}^{n} \eta(y_i - \delta(g^{-1}(\beta_0' \mathbf{x}_i))) \kappa(\lambda' \mathbf{x}_i) \right| \tag{63}$$

is bounded in probability.

Proof. Let $C_p = \{\lambda \in R^p : ||\lambda|| = 1\}$. Note that

$$X_i(\lambda) = \eta(y_i - \delta(g^{-1}(\beta_0' \mathbf{x}))) \kappa(\lambda' \mathbf{x}_i), 1 \leq i \leq n,$$

is a set of i.i.d. random variables with values in $C(C_p)$, where C_p is endowed with the Euclidean metric. Then, it is easy to verify that taking $d^* = d$, conditions (59)-(62) hold. Then by Theorem 6 we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta(y_i - \delta(g^{-1}(\beta_0' \mathbf{x}_i))) \kappa(\lambda' \mathbf{x}_i) \to^D J(\lambda),$$

where $J(\lambda)$ is a Gaussian process in C_p. Then

$$\sqrt{n} \sup_{\|\lambda\|=1} \left| \frac{1}{n} \sum_{i=1}^{n} \eta(y_i - \delta(g^{-1}(\beta_0' \mathbf{x}_i))) \kappa(\lambda' \mathbf{x}_i) \right| \to^D \max_{||\lambda||=1} J(\lambda),$$

and this proves the Lemma.

Proof of Theorem 5. Let $\widehat{\alpha}_n(\lambda)$ be the solution of

$$\frac{1}{n}\sum_{i=1}^{n}\eta(y_i-\delta^*(\widehat{\alpha}_n(\lambda)\lambda'\mathbf{x}_i+\beta_0'\mathbf{x}_i))\kappa(\lambda'\mathbf{x}_i)=0,$$

where $\delta^*(u)=\delta(g^{-1}(u))$,Then we have to show that

$$\sqrt{n}\sup_{\lambda\in C_p}|\widehat{\alpha}_n(\lambda)|=O_p(1).$$

A Taylor expansion yields

$$A_n(\lambda)-B_n(\lambda)\widehat{\alpha}_n(\lambda)=0, \tag{64}$$

where

$$A_n(\lambda)=\frac{1}{n}\sum_{i=1}^{n}\eta(y_i-\delta^*(\beta_0'\mathbf{x}_i))\kappa(\lambda'\mathbf{x}_i),$$

$$B_n(\lambda)=\frac{1}{n}\sum_{i=1}^{n}\eta'(y_i-\delta^*(\alpha_n^*(\lambda)\lambda'\mathbf{x}_i+\beta_0'\mathbf{x}_i))\delta^{*\prime}(\alpha_n^*(\lambda)\lambda'\mathbf{x}_i+\beta_0'\mathbf{x}_i)\kappa(\lambda'\mathbf{x}_i)(\lambda'\mathbf{x}_i)$$

and $|\alpha_n^*(\lambda)|\leq|\alpha_n(\lambda)|$. Then by Lemma 3, it is enough to show that

$$\operatorname{liminf}_{n\to\infty}\inf_{||\lambda||=1}B_n(\lambda)>0.$$

Put $b(u)=\min(\kappa(u)u,1)$. Since by Lemma 2 we have $\sup_{||\lambda||=1}|\alpha_n^*(\lambda)|\to 0$ a.s. and by P1, P2, P4 and Lemma 1 $\eta'(y_i-\delta^*(\beta_0'\mathbf{x}_i))\delta^{*\prime}(\beta_0'\mathbf{x}_i)b(\lambda'\mathbf{x}_i)\geq 0$, Lemma 4.2 of Yohai (1985) implies that

$$\operatorname{liminf}_{n\to\infty}\inf_{||\lambda||=1}B_n(\lambda)\geq\inf_{||\lambda||=1}E_{H_0}\left[\eta'(y-\delta^*(\beta_0'\mathbf{x}))\delta^{*\prime}(\beta_0'\mathbf{x})b(\lambda'\mathbf{x})\right]\geq 0.$$

Moreover as $F(\lambda)=E_{H_0}\left[\eta'(y-\delta^*(\beta_0'\mathbf{x}))\delta^{*\prime}(\beta_0'\mathbf{x})b(\lambda'\mathbf{x})\right]$ is a continuous function of λ, it is enough to prove that for all λ with $||\lambda||=1$

$$F(\lambda)>0. \tag{65}$$

Note that P1 implies that $e(\theta)=E_\theta\left[\eta'(y-\delta(\theta))\right]>0$ for all θ. We also have

$$F(\lambda)=E_{G_0}\left[e(g^{-1}(v_1))\delta^{*\prime}(v_1)b(v_{\mathbf{2}})\right],$$

where $v_1=\beta_0'\mathbf{x}$, $v_2=\lambda'\mathbf{x}$. Put $\mathbf{v}=(v_1,v_2)$, then P3 implies that there are intervals I_1 and I_2 such that $\mathbf{v}$ has a positive density in $I_1\times I_2$. Since $\delta^*(\theta)$ is strictly increasing we can also assume that $\delta^{*\prime}(v)>0$ for $v\in I_1$. Then (65) holds.

6 Proof of Theorem 6

Note that $Q(0,\lambda,s) = 0$ and since $Q(\alpha,\lambda,s)$ is non-increasing in α, we have $\alpha_1(\lambda) \leq 0$ and $\alpha_2(\lambda) \geq 0$. Let $H = (1-\varepsilon)H_0 + \varepsilon\widetilde{H}$, then

$$(1-\varepsilon)Q(T_0(\mathcal{L}((y,\lambda'\mathbf{x},\beta_0'\mathbf{x}),H)),\lambda,S(\mathcal{L}(\lambda'\mathbf{x},H)))+$$

$$\varepsilon\ E_{\widetilde{H}}\left[\psi(y-\delta^*(T_0(\mathcal{L}((y,\lambda'\mathbf{x},\beta_0'\mathbf{x}),H))))\kappa\left(\frac{\lambda'\mathbf{x}}{S(\mathcal{L}(\lambda'\mathbf{x},H))}\right)\right] = \mathbf{0}$$

and this implies that

$$\begin{aligned}&|Q(T_0(\mathcal{L}((y,\lambda'\mathbf{x},\beta_0'\mathbf{x}),H)),\lambda,S(\mathcal{L}(\lambda'\mathbf{x},H)))|\\ &\leq \frac{\varepsilon c_1 c_2}{1-\varepsilon}\end{aligned}$$

for all $H \in \mathcal{V}(\varepsilon, H_0)$. Since $|Q(\alpha,\lambda,s)|$ in non-increasing in s we get

$$-e(\varepsilon) \leq Q(T(\mathcal{L}((y,\lambda'\mathbf{x},\beta_0'\mathbf{x}),H),\lambda,c^+(G_0,\varepsilon,\lambda)) \leq e(\varepsilon)$$

for all $H \in \mathcal{V}(\varepsilon, H_0)$. Using that $Q(\alpha,\lambda,s)$ is monotone non-decreasing in α we obtain

$$\alpha_1(\varepsilon,\lambda) \leq T(\mathcal{L}((y,\lambda'\mathbf{x},\beta_0'\mathbf{x}),H)) \leq \alpha_2(\varepsilon,\lambda)$$

for all $H \in \mathcal{V}(\varepsilon, H_0)$.This proves the Theorem.

7 Proof of Theorem 7

Suppose that

$$\varepsilon < \min(\varepsilon^{+*}(S,G_0),\varepsilon^{-*}(S,G_0)) \tag{66}$$

and

$$\varepsilon > \varepsilon^*_\infty(\mathbf{T},\beta_0,G_0). \tag{67}$$

Then to prove (i), it is enough to show

$$\varepsilon \geq \vartheta_1(G_0) \tag{68}$$

and to prove (ii) we have to show

$$\varepsilon \geq \vartheta_2(H_0). \tag{69}$$

Since ε satisfies (66) and (67), according to (18) and (17), there exists a sequence of distributions $H_n = (1-\varepsilon)H_0 + \varepsilon\widetilde{H}_n$ of $(y,\mathbf{x})$, and a sequence $\lambda_n \in R^p$ satisfying $||\lambda_n|| = 1$ such that $d_n = T_0(\mathcal{L}((y,\lambda_n'\mathbf{x},\beta_0'\mathbf{x}),H_n))$ converges either to $-\infty$ or to ∞. Without loss of generality we can also assume that $\lambda_n \to \lambda_0$. Then we should have

$$(1-\varepsilon)E_{H_0}\left[\eta(y-\delta(g^{-1}(d_n\lambda_n'\mathbf{x}+\beta_0'\mathbf{x})))\text{sign}\,(\lambda_n'\mathbf{x})\right]$$

$$+\varepsilon E_{\widetilde{H}_n}\left[\eta(y-\delta(g^{-1}(d_n\lambda_n'\mathbf{x}+\beta_0'\mathbf{x})))\text{sign}\,(\lambda_n'\mathbf{x})\right]=0,$$

Then if we put $c_1=\sup\eta$ we get

$$\begin{aligned}\frac{1-\varepsilon}{\varepsilon}&=\frac{\left|E_{\widetilde{H}_n}\left[\eta(y-\delta(g^{-1}(d_n\lambda_n'\mathbf{x}+\beta_0'\mathbf{x})))\ \text{sign}\,(\lambda_n'\mathbf{x})\right]\right|}{\left|E_{H_0}\left[\eta(y-\delta(g^{-1}(d_n\lambda_n'\mathbf{x}+\beta_0'\mathbf{x})))\text{sign}\,(\lambda_n'\mathbf{x})\right]\right|}\\&\leq\frac{c_1}{\left|E_{H_0}\left[\eta(y-\delta(g^{-1}(d_n\lambda_n'\mathbf{x}+\beta_0'\mathbf{x})))\text{sign}\,(\lambda_n'\mathbf{x})\right]\right|}.\end{aligned}\tag{70}$$

In case A, using the same arguments as in Lemma 3 of Bianco, García Ben and Yohai (2005), it can be shown that $m(\theta)\to\infty$ implies $y\to^P\infty$ and $m(\theta)\to-\infty$ implies $y\to^p-\infty$. This implies that $\delta(g^{-1}(\phi))\to\infty$ when $\phi\to\infty$ and $\delta(g^{-1}(\phi))\to-\infty$ when $\phi\to-\infty$. As a consequence of this and P4 we have that $\lambda_0'\mathbf{x}\neq0$ implies

$$\lim_{n\to\infty}\eta(y-\delta(g^{-1}(d_n\lambda_n'\mathbf{x}+\beta_0'\mathbf{x})))\text{sign}\,(\lambda_n'\mathbf{x})=\begin{cases}-c_1 & \text{if} \quad d_n\to\infty\\ c_1 & \text{if}\quad d_n\to-\infty,\end{cases}$$

and therefore

$$\begin{aligned}\lim_{n\to\infty}\left|E_{H_0}\left[\eta(y-\delta(g^{-1}(d_n\lambda_n'\mathbf{x}+\beta_0'\mathbf{x})))\text{sign}\,(\lambda_n'\mathbf{x})\right]\right|&\geq c_1P_{G_0}(\lambda_0'\mathbf{x}\neq0)\\&\geq c_1\xi_1(G_0).\end{aligned}$$

From (70) we get

$$\frac{(1-\varepsilon)}{\varepsilon}\leq\xi_1(G_0)$$

and this implies $\varepsilon\geq\xi_1(G_0)/(1+\xi_1(G_0))=\vartheta_1(G_0)$ proving (68). This proves (i).

In case B, using once more arguments similar to those in Lemma 3 of Bianco et al. (2005), we can prove that $m(\theta)\to\infty$ implies that $y\to^P\infty$ and $m(\theta)\to-\infty$ implies $y\to^p0$. Then $\delta(g^{-1}(\phi))\to\infty$ when $\phi\to\infty$ and $\delta(g^{-1}(\phi))\to0$ when $\phi\to-\infty$ and therefore $\lambda_0'\mathbf{x}\neq0$ implies

$$\begin{aligned}&\lim_{n\to\infty}\eta(y-\delta(g^{-1}(d_n\lambda_n'\mathbf{x}+\beta_0'\mathbf{x})))\text{sign}\,(\lambda_n'\mathbf{x})\\&=\begin{cases}-c_1I(\lambda_0'\mathbf{x}>0)-c_1I(\lambda_0'\mathbf{x}<0,y>0) & \text{if}\quad d_n\to\infty\\ c_1I(\lambda_0'\mathbf{x}<0)+c_1I(\lambda_0'\mathbf{x}>0,y>0) & \text{if}\quad d_n\to-\infty,\end{cases}\end{aligned}$$

and then

$$\begin{aligned}\lim_{n\to\infty}\left|E_{H_0}\left[\eta(y-\delta(g^{-1}(d_n\lambda_n'\mathbf{x}+\beta_0'\mathbf{x})))\text{sign}\,(\lambda_n'\mathbf{x})\right]\right|&\geq c_1P_{H_0}(\lambda'\mathbf{x}\neq0,y\neq0)\\&\geq c_1\xi_2(H_0).\end{aligned}$$

Then (69) follows from (70) and the last equation. This proves (ii).

8 Proof of Theorem 8

The following Lemma is required to prove Theorem 8.

Lemma 4. *Let* $e(\mathbf{z},\beta) : R^p \times R^k \rightarrow R^k$ *be a continuously differentiable function in* β *and let* $\mathbf{z}_1,...,\mathbf{z}_n$ *be i.i.d random vectors of dimension p. Consider a sequence of estimators* $\widehat{\beta}_n$ *such* $n^{1/2}(\widehat{\beta}_n - \beta_0) = O_p(1)$*. Suppose also that there exists* $\zeta > 0$ *such that*

$$\sup_{||\beta-\beta_0||\leq\zeta} |e(\mathbf{z},\beta)|$$

has finite expectation. *Then we have:*

(a) If $E\left[e(\mathbf{z},\beta_0)\right] = 0$*, then*

$$\frac{1}{n^{1/2}}\sum_{i=1}^{n} e(\mathbf{z}_i,\widehat{\beta}_n) = O_p(1).$$

(b) Also assume that

$$E\left[\left.\frac{\partial e(\mathbf{z},\beta)}{\partial\beta}\right|_{\beta=\beta_0}\right] = \mathbf{0}, \tag{71}$$

then

$$\frac{1}{n^{1/2}}\sum_{i=1}^{n} e(\mathbf{z}_i,\widehat{\beta}_n) - \frac{1}{n^{1/2}}\sum_{i=1}^{n} e(\mathbf{z}_i,\beta_0) = o_p(1).$$

Proof. Using the Mean Value Theorem (MVT) we have

$$\frac{1}{n^{1/2}}\sum_{i=1}^{n} e(\mathbf{z}_i,\widehat{\beta}_n) - \frac{1}{n^{1/2}}\sum_{i=1}^{n} e(\mathbf{z}_i,\beta_0) = n^{1/2}(\widehat{\beta}_n - \beta_0)\frac{1}{n}\sum_{i=1}^{n} \left.\frac{\partial e(\mathbf{z}_i,\beta)}{\partial\beta}\right|_{\beta=\beta_n^*}, \tag{72}$$

where $\beta_n^* \rightarrow^p \beta_0$.

By Lemma 4.2 of Yohai (1985) we have

$$\lim_{n\rightarrow\infty} \frac{1}{n}\sum_{i=1}^{n} \left.\frac{\partial e(\mathbf{z}_i,\beta)}{\partial\beta}\right|_{\beta=\beta_n^*} = E\left[\left.\frac{\partial e(\mathbf{z}_i,\beta)}{\partial\beta}\right|_{\beta=\beta_0}\right]. \tag{73}$$

By the Central Limit Theorem (CLT) $n^{-1/2}\sum e(\mathbf{x}_i,\beta_0)$ is bounded in probability. Then part (a) follows from (72) and (73).

Part (b) follows from (72), (73) and (71).

Proof of Theorem 8. Put

$$\widehat{\beta}_{2n} = \widehat{\beta}_{0n} - \mathbf{A}_0^{-1}\frac{1}{n}\sum_{i=1}^{n} w(d(\mathbf{x}_i,\widehat{\mu}_n,\widehat{\Sigma}_n))\psi(y_i,\widehat{\beta}_{0n}'\mathbf{x}_i,s_n^*)\mathbf{x}_i.$$

We start proving that

$$n^{1/2}(\widehat{\beta}_{1n} - \widehat{\beta}_{2n}) \rightarrow 0. \tag{74}$$

Since

$$\widehat{\beta}_{1n}=\widehat{\beta}_{0n}-\widehat{\mathbf{A}}_n^{-1}\frac{1}{n}\sum_{i=1}^{n}w(d(\mathbf{x}_i,\widehat{\mu}_n,\widehat{\mathbf{\Sigma}}_n))\psi(y_i,\widehat{\beta}_{0n}'\mathbf{x}_i,s_n^*)\mathbf{x}_i,$$

we have that

$$\begin{aligned}n^{1/2}(\widehat{\beta}_{1n}-\widehat{\beta}_{2n})&=(\mathbf{A}_0^{-1}-\widehat{\mathbf{A}}_n^{-1})\frac{1}{n^{1/2}}\sum_{i=1}^{n}w(d(\mathbf{x}_i,\widehat{\mu}_n,\widehat{\Sigma}_n))\psi(y_i,\widehat{\beta}_{0n}'\mathbf{x}_i,s_n^*)\mathbf{x}_i\\&=(\mathbf{A}_0^{-1}-\widehat{\mathbf{A}}_n^{-1})\frac{1}{n^{1/2}}\sum_{i=1}^{n}e(\mathbf{z}_i,\widehat{\beta}_{0n},\widehat{\mu}_n\widehat{\mathbf{\Sigma}}_n),\end{aligned}\tag{75}$$

where $\mathbf{z}_i=(y_i,\mathbf{x}_i)$, and given $\mathbf{z}=(y,\mathbf{x})$ we define

$$e(\mathbf{z},\beta,\mu,\mathbf{\Sigma},s)=w(d(\mathbf{x},\mu,\Sigma))\psi(y,\beta'\mathbf{x},s)\mathbf{x}.$$

Note that A2 implies $E\left[\psi(y,\beta_0'\mathbf{x},\mathbf{s})|\mathbf{x}\right]=0$ and then

$$\begin{aligned}E\left[e(\mathbf{z},\beta_0,\mu_0,\mathbf{\Sigma}_0,s_0^*)\right]&=E\left[E\left[e(\mathbf{z},\beta_0,\mu_0,\mathbf{\Sigma}_0,s_0^*)|\mathbf{x}\right]\right]\\&=E\left[E\left[\psi[y,\beta_0'\mathbf{x},s_0^*)|\mathbf{x}\right]w(d(\mathbf{x},\mu_0,\Sigma_0))\mathbf{x}\right]\\&=0.\end{aligned}$$

Then, by part (a) of Lemma 4 (a) we have

$$\frac{1}{n^{1/2}}\sum_{i=1}^{n}e(\mathbf{z}_i,\widehat{\beta}_{0n},\widehat{\mathbf{\Sigma}}_n,s_n^*)\ =O_p(1)$$

and therefore (74) follows from A5 and (75).

Note that by the Multivariate Central Limit Theorem

$$D_n=-n^{-1/2}\mathbf{A}_0^{-1}\sum_{i=1}^{n}w(d(\mathbf{x}_i,\mu_0,\Sigma_0))\psi(y_i,\beta_0'\mathbf{x}_i,s_0^*)\mathbf{x}_i\ \to^d N(0,V_0).$$

Then according to (74) in order to prove the Theorem it is enough to show

$$n^{1/2}(\widehat{\beta}_{2n}-\beta_0)-D_n=o_p(1).\tag{76}$$

We can write

$$\begin{aligned}n^{1/2}(\widehat{\beta}_{2n}-\beta_0)-D_n&=n^{1/2}(\widehat{\beta}_{0n}-\beta_0)\\&+\mathbf{A}_0^{-1}\frac{1}{n^{1/2}}\sum_{i=1}^{n}w(d(\mathbf{x}_i,\mu_0,\Sigma_0,s_0^*))\psi(y_i,\mathbf{x}_i'\beta_0,s_0^*)\mathbf{x}_i\\&-\mathbf{A}_0^{-1}\frac{1}{n^{1/2}}\sum_{i=1}^{n}\psi(y_i,\mathbf{x}_i'\widehat{\beta}_{0n},s_n^*)w(d(\mathbf{x}_i,\widehat{\mu}_n,\widehat{\Sigma}_n))\mathbf{x}_i\\&=n^{1/2}(\sum_{i=1}^{n}(e^*(\mathbf{x}_i,y_i,\widehat{\beta}_{0n},\widehat{\Sigma}_n,s_n^*)-e^*(y_i,\mathbf{x}_i,\beta_0,\Sigma_0,s_0^*)),\end{aligned}$$

where

$$e^{*}(\mathbf{x},y,\beta,\mu,\Sigma,s)=\beta-\psi(y,\mathbf{x}'\beta,s)w(d(\mathbf{x},\mu,\Sigma))\mathbf{A}_0^{-1}\mathbf{x}.$$

Note that

$$\frac{\partial e^{*}(y,\mathbf{x},\beta,\mu,\Sigma,s)}{\partial\Sigma}=-\psi(y,\mathbf{x}'\beta_0,s)\frac{\partial w(d(\mathbf{x},\mu_0,\Sigma))}{\partial\Sigma}\mathbf{A}_0^{-1}\mathbf{x},$$

and therefore putting $\theta^{*}=g^{-1}(\mathbf{x}'\beta_0\ )$we get

$$E\left[\psi(y,\mathbf{x}'\beta_0,s)|\mathbf{x}\right]=E_{\theta^{*}}\left[\psi(y,g(\theta^{*}),s)\right]=0.$$

This implies

$$\begin{aligned}E\left[\left.\frac{\partial e^{*}(y,\mathbf{x},\beta_0,\mu_0,\Sigma,s_0^{*})}{\partial\Sigma}\right|\mathbf{x}\right]&=-E\left[\psi(y,\mathbf{x}'\beta_0,s_0^{*})|\mathbf{x}\right]\frac{\partial w(d(\mathbf{x},\mu_0,\Sigma))}{\partial\Sigma}\mathbf{A}_0^{-1}\mathbf{x}\\&=\mathbf{0}\end{aligned}$$

and then for all $\mathbf{\Sigma}$

$$\begin{aligned}E\left[\frac{\partial e^{*}(y,\mathbf{x},\beta_0,\mu_0,\Sigma,s_0^{*})}{\partial\Sigma}\right]&=E\left[E\left[\left.\frac{\partial e^{*}(y,\mathbf{x},\beta_0,\mu_0,\Sigma,s_0^{*})}{\partial\Sigma}\right|\mathbf{x}\right]\right]\\&=\mathbf{0}.\end{aligned}$$

Similarly we can prove that for all μ

$$E\left[\frac{\partial e^{*}(y,\mathbf{x}'\beta_0,\mu,\Sigma_0,s)}{\partial\mu}\right]=0.$$

We can write

$$\frac{\partial e^{*}(\mathbf{z},\beta,\mu_0,\Sigma_0,s_0^{*})}{\partial\beta}=I-\mathbf{A}_0^{-1}w(d(\mathbf{x},\mu_0,\Sigma_0)\psi_{\theta}(y,\beta'\mathbf{x},s_0^{*})\ \mathbf{x}\mathbf{x}'$$

and therefore by (31) we get

$$E\left[\left.\frac{\partial\ e^{*}(y,\mathbf{x},\beta,\mu_0,\Sigma_0,s_0^{*})}{\partial\beta}\right|_{\beta=\beta_0}\right]=I-\mathbf{A}_0^{-1}\mathbf{A}_0=0.$$

Since

$$E\left[\psi(y,\beta_0^{'}\mathbf{x},s)\right]=0\text{ for all }s,$$

we get that

$$E\left[\left.\frac{\partial\psi(y,\beta_0'\mathbf{x},s)}{\partial s}\right|\mathbf{x}\right]=0$$

and this implies

$$E\left[\left.\frac{\partial\ e^{*}(y,\mathbf{x},\beta_0,\mu_0,\Sigma_0,s)}{\partial s}\right|_{s=s_0^{*}}\right]=0.$$

Then (76) follows from part (b) of Lemma 4.

9 Proof of Theorem 9

Take

$$\varepsilon < \min(\varepsilon^*(\mathbf{T}_0, H_0), \varepsilon^*(\mathbf{T}_\mu, H_0), \varepsilon^*(\mathbf{T}_{\mathbf{\Sigma}}, H_0), \varepsilon^{-*}(\mathbf{T}_A, H_0)). \tag{77}$$

To prove (a) we have to show that $||\mathbf{T}_1(H)||$ remains bounded for $H \in \mathcal{V}(H_0, \varepsilon)$. Note that the functional $\mathbf{T}_1(H)$ is given by

$$\mathbf{T}_1(H) = \mathbf{T}_0(H) - \mathbf{T}_{\mathbf{A}}(H)^{-1} E_H \left[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)) \psi(y, \mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{x} \right]. \tag{78}$$

We can write

$$\begin{aligned}
&E_H[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\psi(y, \mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{x}] \\
&= \mathbf{T}_\Sigma(H)^{1/2} E_H[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\psi(y, \mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{T}_\Sigma(H)^{-1/2}(\mathbf{x} - \mathbf{T}_\mu(H))] \\
&+ E_H[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\psi(y, \mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{T}_\mu(H)].
\end{aligned} \tag{79}$$

Let $K_1 = \sup \psi$, $K_2 = \sup w$ and $K_3 = \sup w(d)d$,

$$K_4 = \max_{H \in \mathcal{V}(H_0,\varepsilon)} \gamma_p(\mathbf{T}_{\mathbf{\Sigma}}(H)),$$

$$K_5 = \max_{H \in \mathcal{V}(H_0,\varepsilon)} ||\mathbf{T}_\mu(H)||.$$

By (77) we have $K_4 < \infty$ and $K_5 < \infty$. Then for $H \in \mathcal{V}(H_0, \varepsilon)$ we have

$$\begin{aligned}
&||\mathbf{T}_\Sigma(H)^{1/2} E_H[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\psi(y, \mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{T}_\Sigma(H)^{-1/2}(\mathbf{x} - \mathbf{T}_\mu(H))]|| \\
&\leq K_4 K_1 E_H[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))||\mathbf{T}_\Sigma(H)^{-1/2}(\mathbf{x} - \mathbf{T}_\mu(H))||] \\
&\leq K_4 K_1 K_2^{1/2} E_H[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))^{1/2} d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H))^{1/2}] \\
&\leq K_4 K_1 K_2^{1/2} K_3^{1/2}.
\end{aligned} \tag{80}$$

We also have

$$\begin{aligned}
||E_H[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\psi(y, \mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{T}_\mu(H)]|| &\leq K_1 K_2 ||\mathbf{T}_\mu(H)|| \\
&\leq K_1 K_2 K_5.
\end{aligned} \tag{81}$$

From (79), (80) and (81) we conclude that

$$||E_H[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\psi(y, \mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{x}]||$$

remains bounded when $H \in \mathcal{V}(H_0, \varepsilon)$. Since $\varepsilon < \varepsilon^{-*}(\mathbf{T}_{\mathbf{A}}, H_0)$, the eigenvalues of $\mathbf{T}_{\mathbf{A}}(H)^{-1}$ remain also bounded. Therefore using that $\varepsilon < \varepsilon^*(\mathbf{T}_1, H_0)$ and (78), we obtain that $||\mathbf{T}_1(H)||$ remains bounded too. This proves part (a) of the Theorem.

Now we prove part (b). The functional $\mathbf{T}_{\mathbf{A}}$ is defined as

$$\mathbf{T}_{\mathbf{A}}(H) = E_H[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\tau^*(\mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{x}\mathbf{x}'],$$

where $\tau^*(\phi, s) = \tau(g^{-1}(\phi), s)$.Since by A10 $\tau^*(\phi, s) < 0$, $\mathbf{A}(H)$ is semidefinite negative. Then, it is enough to show that if

$$\varepsilon < \min\left(\varepsilon^*(\mathbf{T}_0, H_0),\ \varepsilon^*(\mathbf{T}_{S^*}, H_0), \frac{\xi_1(G_0) - 0.5}{\xi_1(G_0)}\right), \tag{82}$$

then

$$\inf_{H\in\mathcal{V}(H_0,\varepsilon)} \inf_{||c||=1} -\mathbf{c}'\mathbf{T}_{\mathbf{A}}(H)\mathbf{c} > \mathbf{0}. \tag{83}$$

Given $H = (1-\varepsilon)H_0 + \varepsilon H^*$, we have that

$$\begin{aligned} -\mathbf{T}_{\mathbf{A}}(H) = &-(1-\varepsilon)E_{H_0}[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\tau^*(\mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{x}\mathbf{x}'] \\ &- \varepsilon E_{H^*}[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\tau^*(\mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{x}\mathbf{x}'], \end{aligned}$$

and since

$$-\varepsilon E_{H^*}\left[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\tau^*(\mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{x}\mathbf{x}'\right]$$

is semidefinite positive, in order to prove (83) it will be enough to show that

$$\inf_{H\in\mathcal{V}(H_0,\varepsilon)} \inf_{||c||=1} -\mathbf{c}'\mathbf{A}^*(H)\mathbf{c} > \mathbf{0},$$

where

$$\mathbf{A}^*(H) = E_{G_0}\left[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\tau^*(\mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))\mathbf{x}\mathbf{x}'\right].$$

Since $\varepsilon < (\xi_1(G_0) - 0.5)/\xi_1(G_0)$ we have

$$\begin{aligned} \gamma_0 &= \frac{1}{3}\left(\xi_1(G_0) - \frac{0.5}{1-\varepsilon}\right) \\ &= \frac{-\varepsilon\xi_1(G_0) + \xi_1(G_0)) - 0.5}{3(1-\varepsilon)} \\ &= -\frac{\varepsilon - [(\xi_1(G_0) - 0.5)/\xi_1(G_0)]}{3(1-\varepsilon)} \\ &> 0 \end{aligned} \tag{84}$$

and from the definition of $\xi_1(G_0)$ we can prove that there exists $\gamma_1 > 0$ such that

$$\sup_{||\mathbf{c}||=\mathbf{1}} P_{G_0}(|c'\mathbf{x}| \leq \gamma_1) < 1 - \xi_1(G_0) + \gamma_0. \tag{85}$$

Assumptions A8 and A9 imply

$$P_{G_0}(w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H))) < w_0) \leq \frac{0.5}{1-\varepsilon}. \tag{86}$$

Let K_6 be such that

$$P_{G_0}(||\mathbf{x}|| > K_6) < \gamma_0. \tag{87}$$

By (82) we can find $k_0 > 0$ and $K_7 < \infty$ such that

$$\inf_{H\in\mathcal{V}(H_0,\varepsilon)} T_{S^*}(H) > k_0, \sup_{H\in\mathcal{V}(H_0,\varepsilon)} T_{S^*}(H) < K_7$$

and $K_8 < \infty$ such that

$$\sup_{H\in\mathcal{V}(H_0,\varepsilon)} ||\mathbf{T}_1(H)|| < K_8.$$

Since $\tau^*(\phi, s)$ is continuous and by A10 $\tau^*(\phi, s) < 0$ for all (ϕ, s) with $s > 0$ we have

$$b_2 = \inf_{||\phi||\leq K_6 K_8, k_0\leq s\leq K_7} -\tau^*(\phi, s) > 0.$$

Put

$$D_{\mathbf{c}} = \{|\mathbf{c}'\mathbf{x}| \leq\gamma_1\} \cup \{||\mathbf{x}|| \geq K_6\} \cup \{w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H))) < w_0\}.$$

Then, from (85), (86), (87) and (84) we obtain

$$\begin{aligned} P_{G_0}(D_{\mathbf{c}}) &< (1 - \xi_1(G_0) + \gamma_0) + \gamma_0 + \frac{0.5}{1-\varepsilon} \\ &= 1 - \xi_1(G_0) + 2\gamma_0 + \xi_1(G_0) - 3\gamma_0 \\ &= 1 - \gamma_0 \end{aligned}$$

and if we denote by $D_{\mathbf{c}}^C$ the complement of D_c we have

$$P_{G_0}(D_{\mathbf{c}}^C) > \gamma_0.$$

Take $\mathbf{c} \in R^p$ with $||\mathbf{c}|| = \mathbf{1}$. Since in $D_{\mathbf{c}}^C$ we have $w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H))) \geq w_0$, $\tau^*(\mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H)) > b_2$ and $|\mathbf{c}'\mathbf{x}| > \gamma_1$ we get

$$\begin{aligned} &\inf_{H\in\mathcal{V}(H_0,\varepsilon)} -\mathbf{c}'\mathbf{A}^*(H)\mathbf{c} \\ &\geq \inf_{H\in\mathcal{V}(H_0,\varepsilon)} -E_{H_0}\left[w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\tau^*(\mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))(\mathbf{c}'\mathbf{x})^2\right] \\ &\geq \inf_{H\in\mathcal{V}(H_0,\varepsilon)} -E_{H_0}\left[I(D_{\mathbf{c}}^C)w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H)))\tau^*(\mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H))(\mathbf{c}'\mathbf{x})^2\right] \\ &\geq \gamma_1^2 w_0 b_2 P_{G_0}(D_{\mathbf{c}}^C) \\ &\geq \gamma_1^2 w_0 b_2 \gamma_0, \end{aligned}$$

and then (83) follows. This proves part (b) of the Theorem.

10 Derivation of the CUMGES estimators

A *general M-estimator* (GM-estimator) of α for the auxiliary model (6)-(7) is defined by

$$\sum_{i=1}^n \psi(y, w, z, \alpha) = 0,$$

where $\psi : R^4 \to R$. The GM-estimator corresponding to ψ is *conditionally unbiased* (CUGM-estimator) if ψ satisfies

$$E_\alpha(\psi(y,w,z,\alpha)|w,z) = 0.$$

Assume $g(\theta) = m(\theta)$. The maximum likelihood estimator for the auxiliary model has score function

$$\psi_0(y,w,z,\alpha) = (y - q(g^{-`1}(\alpha w + z)))w,$$

where $q(\theta) = E_\theta(y)$ and then it is a CUGM-estimator.

Let $\delta_k(\theta)$ and $\psi_k(y,w,z,\alpha)$ be defined by

$$E_\theta\left(\psi_k(y - \delta_k(\theta)\right) = 0$$

and

$$\begin{aligned}\psi_k(y,w,z,\alpha) &= \eta_k^H\left(\left(y - \delta_{\frac{k}{|w|}}(g^{-1}(\alpha w + z))\right)w\right)\\ &= \text{sign}(w)|w|\eta^H_{\frac{k}{|w|}}\left(y - \delta_{\frac{k}{|w|}}\left(g^{-1}(\alpha w + z)\right)\right). \end{aligned} \tag{88}$$

Then if $\theta_{w,z} = g^{-1}(\alpha w + z)$, we have

$$\begin{aligned}E_\alpha(\psi_k(y,w,z,\alpha|w,z)) &= \text{sign}(w)|w|E_{\theta_{w,z}}\left[\eta^H_{\frac{k}{|w|}}\left(y - \delta_{\frac{k}{|w|}}(\theta_{w,z})\right)\right]\\ &= 0.\end{aligned}$$

Then the GM-estimator of α based on $\psi_k(y,w,z,\alpha)$ is conditionally unbiased. Let $B(\alpha)$ be defined by

$$B_k(\alpha) = \left|E\left[\frac{\partial\psi_k(y,w,z,\alpha)}{\partial\alpha}\right]\right|.$$

Then as in Künsch et al. (1989) it can be proved that for the auxiliary model the GM-estimator based on $\psi_k(y,w,z,\alpha)$ is the conditional unbiased GM-estimator with smallest variance subject to the constrain that the GES is smaller or equal than $k/B_k(\alpha)$.

To obtain the minimum GES it seems natural to take $k \to 0$. Then in case that the median of F_θ is well defined we have $\lim_{k\to 0}\delta_k(\theta) = \delta_0(\theta) =\text{median}_\theta(y)$ and from (88) we get

$$\begin{aligned}&\lim_{k\to 0}\frac{1}{k}\psi_k(y,w,z,\alpha)\\ &= \text{sign}(w)\lim_{k\to 0}\frac{1}{\frac{k}{|w|}}\eta^H_{\frac{k}{|w|}}\left(y - \delta_{\frac{k}{|w|}}(g^{-1}(\alpha w + z))\right)\text{sign}(y - \delta_0(g^{-1}(\alpha w + z)))\text{sign}(w)\\ &= \text{sign}(w)\lim_{u\to 0}\frac{1}{u}\eta_u^H(y - \delta_u(g^{-1}(\alpha w + z)))\\ . &= \text{sign}(y - \delta_0(g^{-1}(\alpha w + z)))\text{sign}(w).\end{aligned}$$

Therefore the GM-estimator with score function

$$\psi_0(y, w, z, \alpha) = \text{sign}(y - \delta_0(g^{-1}(\alpha w + z)))\text{sign}(w)$$

is a CUMGES estimator. A rigorous proof that the GES is minimized when $k \to 0$ requires arguments similar to those used in Section 3.8.6 of Maronna et al. (2006).

When y takes only non-negative integer values the median of F_θ is not in general well defined. In this case we will show that

$$\lim_{k\to 0} \frac{1}{k}\psi_k(y, w, z, \alpha) = \frac{1}{0.5}\eta_{0.5}^H(y - \delta_{0.5}(g^{-1}(\alpha w + z)))\text{sign}(w), \tag{89}$$

and therefore the M-estimator with score function

$$\psi_0(y, w, z, \alpha) = \eta_{0.5}^H(y - \delta_{0.5}(g^{-1}(\alpha w + z)))\text{sign}(w)$$

is the CUMGES estimator. We first prove that for $0 < k < 0.5$.

$$\frac{1}{k}\eta_k^H(y - \delta_k(\theta)) = \frac{1}{0.5}\eta_{0.5}^H(y - \delta_{0.5}(\theta)). \tag{90}$$

Since η_k^H $(u)$and $\delta_k(\theta)$ are continuous in k,to prove (90) it is enough to show that given $k_2 < k_1 < 0.5$

$$\frac{1}{k_2}\eta_{k_2}^H(y - \delta_{k_2}(\theta)) = \frac{1}{k_1}\ \eta_{k_1}^H(y - \delta_{k_1}(\theta)). \tag{91}$$

Note that $k_1 < 0.5$ implies that there exists at most one non-negative integer y such that $|y - \delta_{k_1}(\theta)| < k_1$. Assume first that there exists one integer $y_0(\theta)$ satisfying this condition. In this case we can write

$$\eta_{k_1}(y - \delta_{k_1}(\theta)) = \begin{cases} k_1 & \text{if} \quad y > y_0(\theta) \\ -k_1 & \text{if} \quad y < y_0(\theta) \\ y_0(\theta) - \delta_{k_1}(\theta) & \text{if} \quad y = y_0(\theta). \end{cases} \tag{92}$$

Now we prove prove that $\delta_{k_2}(\theta)$ is given by δ^* defined by

$$y_0(\theta) - \delta^* = (k_2/k_1)(y_0(\theta) - \delta_{k_1}(\theta)). \tag{93}$$

Note that (93) implies that $|y_0(\theta) - \delta^*| < k_2$ and therefore

$$\eta_{k_2}(y - \delta^*) = \begin{cases} k_2 & \text{if} \quad y > y_0(\theta) \\ -k_2 & \text{if} \quad y < y_0(\theta) \\ \frac{k_2}{k1}(y_0(\theta) - \delta_{k_1}(\theta)) & \text{if} \quad y = y_0(\theta). \end{cases} \tag{94}$$

From (92) and (94) we get that

$$\eta_{k_2}(y - \delta^*) = (k_2/k_1)\eta_{k_1}(y - \delta_{k_1}(\theta)). \tag{95}$$

and then

$$E_\theta(\eta_{k_2}(y - \delta^*)) = (k_2/k_1)E_\theta(\eta_{k_1}(y - \delta_{k_1}(\theta))) = 0.$$

This implies that $\delta_{k_2}(\theta)=\delta^*$ and by (95) we get (91).

Consider now the case that $|y-\delta_{k_1}(\theta)|\geq k_1$ for all non-negative integer y. Then

$$\eta_{k_1}(y-\delta_{k_1}(\theta))=\begin{cases} k_1 & \text{if} \quad y>\delta_{k_1}(\theta) \\ -k_1 & \text{if} \quad y<\delta_{k_1}(\theta), \end{cases} \tag{96}$$

and since for all y we have $|y-\delta_{k_1}(\theta)|\geq k_1>k_2$ we get

$$\eta_{k_2}(y-\delta_{k_1}(\theta))=\begin{cases} k_2 & \text{if} \quad y>\delta_{k_1}(\theta) \\ -k_2 & \text{if} \quad y<\delta_{k_1}(\theta). \end{cases} \tag{97}$$

Then

$$\eta_{k_2}(y-\delta_{k_1}(\theta))=\frac{k_2}{k_1}\eta_{k_1}(y-\delta_{k_1}(\theta))$$

and

$$E(\eta_{k_2}(y-\delta_{k_1}(\theta))=\frac{k_2}{k_1}E_\theta(\eta_{k_1}(y-\delta_{k_1}(\theta)))=0.$$

This implies that $\delta_{k_2}(\theta)=\delta_{k_1}(\theta)$ and (91) holds too.

From (88) and (90) we have that

$$\begin{aligned}
\lim_{k\to 0}\frac{1}{k}\psi_k(y,w,z,\alpha) &= \operatorname{sign}(w)\lim_{k\to 0}\frac{1}{\frac{k}{|w|}}\eta^H_{\frac{k}{|w|}}\left(y-\delta_{\frac{k}{|w|}}(g^{-1}(\alpha w+z))\right) \\
&= \operatorname{sign}(w)\lim_{u\to 0}\frac{1}{u}\eta^H_u(y-\delta_u(g^{-1}(\alpha w+z))) \\
&= \frac{1}{0.5}\operatorname{sign}(w)\eta^H_{0.5}(y-\delta_{0.5}(g^{-1}(\alpha w+z))).
\end{aligned}$$

This proves (89).

11 Numerical Algorithm

In this Section we give an efficient algorithm to approximately compute P-estimators for the GLM. The algorithm, based on subsampling, is similar to the one given in Maronna and Yohai (1993) for projection estimators in the linear model. Suppose that we have a set $B=\{\beta^{(1)},\ldots,\beta^{(N)}\}$ of candidates for β_0. For any $\beta^{(i)}$ we define the following set of candidates for λ

$$\lambda^{(j)}=\frac{\beta^{(j)}-\beta^{(i)}}{||\beta^{(j)}-\beta^{(i)}||},\ j\neq i.$$

Then we define

$$A_n(\beta^{(i)})=\max_j S(\mathcal{L}(\lambda^{(j)\prime}\mathbf{x},H_n))\ |T_0(\mathcal{L}((y,\lambda^{(j)\prime}\mathbf{x},\beta^{(i)\prime}\mathbf{x}),H_n))|. \tag{98}$$

Then the approximate projection estimator is $\widehat{\beta}_n=\beta^{(i_0)}$, where

$$A_n(\beta^{(i_0)})=\min_{1\leq i\leq N}A_n(\beta^{(i)}).$$

It is not necessary to compute $A_n(\beta^{(i)})$ for all i. Suppose that we have already computed $A_n(\beta^{(1)}),\ldots,A_n(\beta^{(i)})$ and let

$$a_i=\min(A_n(\beta^{(1)}),\ldots,A_n(\beta^{(i)})).$$

If while computing $A_n(\beta^{(i+1)})$ we find j such that

$$S(\mathcal{L}(\lambda^{(j)\prime}\mathbf{x},H_n))\ |T_0(\mathcal{L}((y,\lambda^{(j)\prime}\mathbf{x},\beta^{(i+1)\prime}\mathbf{x}),H_n))>a_i,$$

we would know that $A_n(\beta^{(i+1)})>a_i$ and therefore we can stop computing $A_n(\beta^{(i+1)})$.

Then, it only remains to explain how to obtain the candidates $\beta^{(i)}, 1\leq i\leq N$. These candidates are obtained by generating at random N subsamples of size h from the original sample of size n, where h is small. Let $\widehat{\beta}_i$, $1\leq i\leq N$, be the value of ML estimator computed with the i-th subsample. In the case that y is Bernoulli, the directions $\widehat{\beta}_i/||\widehat{\beta}_i||$ of these estimates are not far from the true direction $\beta_0/||\beta_0||$ however the norms $||\widehat{\beta}_i||$ are in general much larger than $||\beta_0||$ especially if the number of overlaps between zeros and ones is small. For this reason we replace $\widehat{\beta}_i$ by $\widehat{\beta}_i^*=\alpha\widehat{\beta}_i$ and the value of α is estimated using the CUMGES estimator. Then the estimate of α is obtained by solving

$$\sum_{i=1}^{n}\frac{y_i-g^{-1}\left(\alpha\widehat{\beta}^{\prime}\mathbf{x}_i\right)}{\max\left(g^{-1}\left(\alpha\widehat{\beta}^{\prime}\mathbf{x}_i\right),1-g^{-1}\left(\alpha\widehat{\beta}^{\prime}\mathbf{x}_i\right)\right)}\operatorname{sign}\left(\widehat{\beta}^{\prime}\mathbf{x}_i\right)=0. \tag{99}$$

where g is the logistic function. Since this function is continuous and monotone in α, to solve this equation we can apply any algorithm for a nonlinear equation with one unknown parameter, e.g., binary search or the Brent 's (1973) algorithm.

When the zeros and ones of one sample are separated, the MLE is not defined. However in this case we can compute a value a $\widehat{\beta}_i$ separating ones and zeros. Even if $\widehat{\beta}_i$ is far from β_0, for some subsamples $\widehat{\beta}_i/||\widehat{\beta}_i||$ would be a reasonable estimate of $\beta_0/||\beta_0||$. Then again we replace $\beta_i$by $\widehat{\beta}_i^*=\alpha\widehat{\beta}_i$,where α is obtained by solving (99).

The number N of candidates can be calculated so that probability of obtaining at least one subsample of size h without outliers be smaller or equal that a number p_0 close to one, e.g. $p_0=0.99$. Then if ε is the expected fraction of outlier contamination, N is given by

$$N=\frac{\log(1-p_0)}{\log(1-(1-\varepsilon)^h)}.$$

The size h of the subamples should be small to increase the probability that the subsamples do not contain outliers. In our Monte Carlo study we took $h=2p$ for the logistic model and $h=p$ for Poisson regression.

REFERENCES

Bianco, A., García Ben, M. and Yohai, V. J (2005), "Robust estimation for linear regression with asymmetric errors", *Canadian Journal of Statistics*, **33** 1-18.

Brent, R. P. (1973), "*Algorithms for Minimization without Derivatives* ", Prentice-Hall, Englewood Cliffs.

Jain, N. and Marcus, M. (1975), "Central Limit Theorems for C(S)-valued Random Variables," *Journal of Functional Analysis*, **19**, 216–231.

Künsch, H., Stefanski, L. and Carroll, R. (1989), "Conditionally Unbiased Bounded-Influence Estimation in General Regression Models, with Applications to Generalized Linear Models," *Journal of the American Statistical Association,* 84, 460–466.

Muler, N. and Yohai, V.J. (2002), "Robust Estimates for ARCH Processes", *Journal of Time Series Analysis*, **23**. 341-375.

Yohai, V.J. (1985), "High Breakdown Point and High Efficiency Robust Estimates for Regression ". Technical report No. 66. Department of Statistics, University of Washington. Available at http://www.stat.washington.edu/research/reports/1985/tr066.pdf.