On the Asymptotic Behavior of One-step Estimates in Heteroscedastic Regression Models

Ana Bianco * ‡ Graciela Boente † ‡

Abstract

In this paper, the asymptotic distribution of one–step Newton– Raphson estimates is established for a regression model with random carriers and heteroscedastic errors under mild conditions. We also include, the robust estimates defined as the solution of an implicit equation, such as the MM–estimates.

1 Introduction.

This paper will deal with heteroscedastic regression models where the variance function has a given parametric form, i.e., the model can be written as:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \sigma G\left(\mathbf{x}_i, \boldsymbol{\lambda}, \boldsymbol{\beta}\right) , \tag{1}$$

where, as usual, $(\mathbf{x}_i, y_i)$, $1 \leq i \leq n$, $\mathbf{x}_i \in I\!R^p$, are i.i.d. random vectors, with ε_i and $\mathbf{x}_i$ independent and $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and σ are unknown parameters and $G(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \exp\{\boldsymbol{\lambda}'\, h(\mathbf{x}, \boldsymbol{\beta})\}$.

Some of the most common models for the function G are $G(\mathbf{x}_i, \boldsymbol{\lambda}, \boldsymbol{\beta}) = (1 + |\mathbf{x}_i'\boldsymbol{\beta}|)^{\boldsymbol{\lambda}}$ which has been introduced by Box and Hill (1974), $G(\mathbf{x}_i, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \exp\{\boldsymbol{\lambda}|\mathbf{x}_i'\boldsymbol{\beta}|\}$ considered by Bickel (1978) and $G(\mathbf{x}_i, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \exp\{\boldsymbol{\lambda}'\, h(\mathbf{x}_i)\}$. All these models have in common that the ratio $\left[\frac{\partial}{\partial \boldsymbol{\lambda}}\, G(\mathbf{x}_i, \boldsymbol{\lambda}, \boldsymbol{\beta})\right] / G(\mathbf{x}_i, \boldsymbol{\lambda}, \boldsymbol{\beta})$ does not depend on $\boldsymbol{\lambda}$.

In order to obtain bounded influence estimates, Giltinan, Carroll and Ruppert (1986) generalized homoscedastic GM–estimates to heteroscedastic regression models by considering both Mallows–type and Krasker–Welsch optimal weights. Although, these estimates fail to have high breakdown point when the dimension of the carriers increases. One–step estimates were considered for location–scale models by Bickel (1975), Davies (1992) and Lopuhaä (1992), among others. They have been adapted to homoscedastic regression models with fixed carriers by Simpson, Ruppert

*Universidad de Buenos Aires, Argentina.

†Universidad de Buenos Aires and CONICET Argentina.

‡This research was partially supported by Grants TX–49 from the Universidad de Buenos Aires, PIP #4186 from the CONICET and PICT # 03-00000-00576 from ANPCYT at Buenos Aires, Argentina.

and Carroll (1992). To solve the problem of low breakdown point, Bianco, Boente and di Rienzo (2000) considered a one–step version of GM–estimates based on a high breakdown point estimates of the regression parameters. As in the homoscedastic setting, these estimates inherit the breakdown point of the initial estimate. Moreover, in this paper we will show that using Newton–Raphson estimates one can reach final high–breakdown point root–n estimates with the same asymptotic distribution as the related GM–estimates, even if the initial regression estimates have a lower convergence rate. In Section 2.3, consistency results will be derived while in Section 2.4, the asymptotic distribution of the Newton–Raphson estimates when the initial regression estimates have order n^{τ} with $\tau \in (1/4, 1/2]$ is considered. Theorem 1 states the asymptotic normality of the estimators requiring only consistency to the matrix which estimates the scale of the carriers and a uniform second order condition to the variance function $h(\mathbf{x}, \boldsymbol{\beta})$ and to its derivative $\frac{\partial}{\partial \boldsymbol{\beta}} h(\mathbf{x}, \boldsymbol{\beta})$.

The asymptotic results given in Section 2.4, also include estimates, as MM–type estimates, defined through an implicit equation modified to take into account heteroscedasticity. Their asymptotic behavior is derived in Theorem 2, when the initial estimates are consistent. Remark 6 comments on the order of convergence of the reweighted estimate.

Some technical Lemmas are stated in Section 2.2 and their proofs may be found in the Appendix where the notion of uniform-entropy is described.

2 Main Results

2.1 Definitions and Assumptions.

High breakdown point estimates should be considered in heteroscedastic regression models since GM–estimates, as their relatives in the homoscedastic case, have a breakdown point which decreases with the dimension of the carriers.

Consistent estimates can be obtained by ignoring heteroscedasticity and by estimating $\boldsymbol{\beta}$ through a high breakdown point estimate for homoscedastic models such as the LMS, the S–estimates, the MM–estimates, the $\tau-$estimates or the P–estimates proposed by Rousseeuw (1984), Rousseeuw and Yohai (1984), Yohai (1987), Yohai and Zamar (1988) and Maronna and Yohai (1991), respectively. The disadvantage of such estimates is that they are not efficient under heteroscedasticity.

Denote $\boldsymbol{\beta}_{\mathrm{H}} = \boldsymbol{\beta}_{\mathrm{H},n}$ a high breakdown point estimate of $\boldsymbol{\beta}$, computed as if the regression model was homoscedastic, and $\widehat{\sigma}$ the related scale estimate. Besides, let $\mathbf{S}_{\mathrm{H}}$ be such that $\mathbf{W}_n = \mathbf{S}_n^{-1}\mathbf{S}_n^{-1t}$ is an estimate of the scale matrix of the carriers $\{\mathbf{x}_i\}$ with high breakdown point. Possible choices are the minimum volume estimate (Rousseeuw and van Zomeren (1990)) or the Donoho (1982)–Stahel (1981) estimate. Finally, let $\boldsymbol{\lambda}_{\mathrm{H}}$ also be a high breakdown point estimate of $\boldsymbol{\lambda}$ and denote $\sigma_{\mathrm{H}} =$

$\kappa^{-1}\text{med}\left(|y_i - \mathbf{x}_i'\boldsymbol{\beta}_{\rm H}|/G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)\right)$ with κ a standardizing constant (for normally distributed errors the usual choice is 0.6745).

Consider any score function χ, as those usually used for the scale parameter in robust estimation, and w_3 a weight function. When $\lambda \in I\!R$ and $G(\mathbf{x}, \lambda, \boldsymbol{\beta}) = \exp\{\lambda h(\mathbf{x}, \boldsymbol{\beta})\}$, one can define $\lambda_{\rm H}$ as the solution of

$$\sum_{i=1}^{n} \chi\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}_{\rm H}}{\widehat{\sigma} G\left(\mathbf{x}_i, \lambda, \boldsymbol{\beta}_{\rm H}\right)}\right) w_3\left(h\left(\mathbf{x}_i, \boldsymbol{\beta}_{\rm H}\right)\right) h\left(\mathbf{x}_i, \boldsymbol{\beta}_{\rm H}\right) = 0 \;,$$

which has the asymptotic breakdown point stated in Bianco, Boente and di Rienzo (2000).

The one–step Newton–Raphson estimate is defined as

$$\boldsymbol{\beta}_{\mathcal{N}} = \boldsymbol{\beta}_{\rm H} + \sigma_{\rm H}\mathbf{A}_n^{-1}\mathbf{g}_n \;, \tag{2}$$

with $\mathbf{A}_n$ and $\mathbf{g}_n$ given by

$$\begin{aligned} \mathbf{A}_n &= \sum_{i=1}^{n} \Psi_1'\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}_{\rm H}}{\sigma_{\rm H} G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) w_2\left(\frac{\mathbf{S}_{\rm H}\mathbf{x}_i}{G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) \frac{\mathbf{x}_i\, \mathbf{x}_i'}{G^2\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)} \\ \mathbf{g}_n &= \sum_{i=1}^{n} \Psi_1\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}_{\rm H}}{\sigma_{\rm H} G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) w_2\left(\frac{\mathbf{S}_{\rm H}\mathbf{x}_i}{G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) \frac{\mathbf{x}_i}{G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)} \;. \end{aligned}$$

On the other hand, a reweighted estimate can also be defined as

$$\boldsymbol{\beta}_{\mathcal{R}} = \boldsymbol{\beta}_{\rm H} + \sigma_{\rm H}\mathbf{B}_n^{-1}\mathbf{g}_n \;, \tag{3}$$

where

$$\mathbf{B}_n = \sum_{i=1}^{n} w_1\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}_{\rm H}}{\sigma_{\rm H} G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) w_2\left(\frac{\mathbf{S}_{\rm H}\mathbf{x}_i}{G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) \frac{\mathbf{x}_i\, \mathbf{x}_i'}{G^2\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)} \;,$$

with $w_1(t) = \Psi_1(t)/t$ and $\mathbf{g}_n$ is defined above.

In Bianco, Boente and di Rienzo (2000) it was shown that both estimates have a breakdown which is at least the minimum between the breakdown points of the initial estimates. Therefore, consistent and high breakdown point estimates can be obtained through this procedure. The Newton–Rahphson estimate reaches a root–n order of convergence even if the initial estimate has a lower order. However, as in the homocedastic case (see He and Portnoy (1992)), the reweighted estimate does not improve the order of the initial high breakdown point used in the procedure.

In order to include other estimates, we will also study the asymptotic distribution of the estimates $\widehat{\boldsymbol{\beta}}$ defined as solution of

$$\sum_{i=1}^{n} \Psi_1\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma_{\rm H} G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) w_2\left(\frac{\mathbf{S}_{\rm H}\mathbf{x}_i}{G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) \frac{\mathbf{x}_i}{G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)} = 0 \;, \tag{4}$$

where $\sigma_{\rm H}$, $\boldsymbol{\lambda}_{\rm H}$, $\boldsymbol{\beta}_{\rm H}$ and $\mathbf{S}_{\rm H}$ are defined as in (2). This approach follows the MM–approach given in Yohai (1987).

Let $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \sigma_0, \lambda_0)'$ be such that $\boldsymbol{\theta}_{\rm H} = (\boldsymbol{\beta}_{\rm H}, \sigma_{\rm H}, \lambda_{\rm H})' \xrightarrow{p} \boldsymbol{\theta}_0$ and $\mathbf{S}_0$ be such that $\mathbf{S}_{\rm H} \xrightarrow{p} \mathbf{S}_0$, where $\xrightarrow{p}$ stands for convergence in probability. For the sake of simplicity, we will assume throughout this paper $\lambda \in I\!R$, $G(\mathbf{x}, \lambda, \boldsymbol{\beta}) = \exp\{\lambda h(\mathbf{x}, \boldsymbol{\beta})\}$ and $\mathbf{S}_0 = \mathbf{I}$.

The consistency and the asymptotic distribution of $\boldsymbol{\beta}_{\mathcal{N}}$ and $\widehat{\boldsymbol{\beta}}$ will be derived under the following set of assumptions:

A1. $P_{\mathbf{x}}$, the conditional distribution of $y - \mathbf{x}'\boldsymbol{\beta}_0$ given $\mathbf{x}$, is symmetric around 0 for all $\mathbf{x}$.

A2. Ψ_1 is an odd, bounded and continuous function,

A3. Ψ_1 is twice continuously differentiable with bounded derivatives Ψ_1' and Ψ_1'', such that $\eta_1(t) = t\Psi_1'(t)$ is bounded.

A4. $w_2(\mathbf{x}) = \Psi_2(|\mathbf{x}|)\,|\mathbf{x}|^{-1} > 0$ with Ψ_2 a bounded and continuously differentiable function with derivative Ψ_2'.

A5. The function $\eta_2(t) = t\Psi_2'(t)$ is bounded.

A6. Ψ_2 is twice continuously differentiable with second derivative Ψ_2'' such that the function $\eta_3(t) = t^2\Psi_2''(t)$ is bounded.

A7. There exists $0 < \delta_0 < 1$ such that, for any $K > 0$ the function

$$\alpha(\mathbf{x}) = \sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\delta_0} |h(\mathbf{x}, \boldsymbol{\beta})|$$

is bounded in $\{|\mathbf{x}| \le K\}$.

A8. The function $h(\mathbf{x}, \boldsymbol{\beta})$ is equicontinuous as a function of $\boldsymbol{\beta}$, for all $|\boldsymbol{\beta} - \boldsymbol{\beta}_0| \le \delta_0$, in any compact set $\{|\mathbf{x}| \le K\}$, i.e., given $\epsilon > 0$ there exists $\delta > 0$ such that

$$|h(\mathbf{x}, \boldsymbol{\beta}) - h(\mathbf{x}, \widetilde{\boldsymbol{\beta}})| < \epsilon$$

for $|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}| < \delta$, $|\boldsymbol{\beta} - \boldsymbol{\beta}_0| \le \delta_0$, $|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0| \le \delta_0$ and $|\mathbf{x}| \le K$.

A9. There exists $\delta_0 > 0$ such that

$$E\left[\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\le\delta_0} \frac{|\mathbf{x}|}{G(\mathbf{x}, \lambda, \boldsymbol{\beta})}\right] < \infty\,.$$

A10. There exists $\delta_0 > 0$ such that

$$E\left[\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\delta_0\,,\,|\lambda-\lambda_0|\le\delta_0} \Psi_2\left(\frac{|\mathbf{x}|}{G(\mathbf{x}, \lambda, \boldsymbol{\beta})}\right) \frac{|\mathbf{x}|^2}{G^2(\mathbf{x}, \lambda, \boldsymbol{\beta})}\right] < \infty\,.$$

A11. The matrix

$$\mathbf{A} = E\left(\Psi_1'\left(\frac{y-\mathbf{x}'\boldsymbol{\beta}_0}{\sigma_0\, G\left(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0\right)}\right) w_2\left(\frac{\mathbf{x}}{G\left(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0\right)}\right)\frac{\mathbf{x}\,\mathbf{x}'}{G^2\left(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0\right)}\right)$$

is non–singular.

A12. The function $h(\mathbf{x},\boldsymbol{\beta})$ is continuously differentiable as a function of $\boldsymbol{\beta}$, for each fixed $\mathbf{x}$ and $E\left(\gamma^2(\mathbf{x})\right)<\infty$ where

$$\gamma(\mathbf{x}) = \sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\delta_0} |h(\mathbf{x},\boldsymbol{\beta})| + |\frac{\partial}{\partial\boldsymbol{\beta}}h(\mathbf{x},\boldsymbol{\beta})|\ .$$

Remark 1. **A2** to **A4** are standard conditions on the score functions in regression models. Condition **A7** is obviously fulfilled if $h(\mathbf{x},\boldsymbol{\beta})$ is a continuous function as it is the case for the variance functions described in the Introduction. This continuity assumption implies **A8.**
Assumptions **A9** to **A11** are moment conditions necessary to ensure the asymptotic order and asymptotic normal distribution of the estimates. However, **A10** is necessary only if the initial estimates have order of convergence $\tau<\frac{1}{2}$.

Remark 2. **A4** and **A5** imply that there exists $c>0$ such that

$$|w_2(\mathbf{x})-w_2(\mathbf{z})| \le c\frac{|\mathbf{z}-\mathbf{x}|}{[\min(|\mathbf{x}|,|\mathbf{z}|)]^2}\ . \tag{5}$$

2.2 Technical Lemmas.

In this section, we state some technical Lemmas whose results are necessary to derive the asymptotic distribution of the Newton–Raphson estimates. Their proofs are given in the Appendix.

For the sake of simplicity, we will begin by fixing some notation. Let us denote

$$\begin{aligned}
r(\mathbf{x},y,\boldsymbol{\theta}) &= \frac{y-\mathbf{x}'\boldsymbol{\beta}}{\sigma\, G\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)}\\
r_1(\mathbf{x},y,\boldsymbol{\theta}) &= \frac{y-\mathbf{x}'\boldsymbol{\beta}_0}{\sigma\, G\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)}\\
\mathbf{z}\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right) &= \frac{\mathbf{x}}{G\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)}\\
\mathbf{H}\left(\mathbf{x},y,\boldsymbol{\theta}\right) &= \varphi\left(r(\mathbf{x},y,\boldsymbol{\theta})\right) w_2\left(\mathbf{z}\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)\right)\mathbf{z}\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)\\
\mathbf{H}_1\left(\mathbf{x},y,\boldsymbol{\theta}\right) &= \mathbf{H}\left(\mathbf{x},y,\boldsymbol{\theta}\right)\ \mathbf{H}\left(\mathbf{x},y,\boldsymbol{\theta}\right)'\\
\mathbf{H}_2\left(\mathbf{x},y,\boldsymbol{\theta}\right) &= \mathbf{H}\left(\mathbf{x},y,\boldsymbol{\theta}\right)\ \mathbf{H}\left(\mathbf{x},y,\boldsymbol{\theta}_0\right)'\\
\mathbf{H}_3\left(\mathbf{x},y,\boldsymbol{\theta},\mathbf{S}\right) &= \varphi\left(r(\mathbf{x},y,\boldsymbol{\theta})\right) w_2\left(\mathbf{S}\,\mathbf{z}\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)\right)\mathbf{z}\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)\ \mathbf{z}\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)'\\
\mathbf{H}_4\left(\mathbf{x},y,\boldsymbol{\theta},\mathbf{S}\right) &= \varphi\left(r_1\left(\mathbf{x},y,\boldsymbol{\theta}\right)\right)\ w_2\left(\mathbf{S}\,\mathbf{z}\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)\right)\ \mathbf{z}\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)\\
\mathbf{H}_5\left(\mathbf{x},y,\boldsymbol{\theta}\right) &= \mathbf{H}_4\left(\mathbf{x},y,\boldsymbol{\theta},\mathbf{I}\right)\ ,
\end{aligned}$$

where $\varphi(t)$ is a bounded function. Later on, the function $\varphi(t)$ will be taken as $\Psi_1(t)$, $\Psi_1'(t)$ or $t\Psi_1'(t)$.

For any matrix $\mathbf{B}$, $|\mathbf{B}|$ denotes $\left[\sum_{k,l}\left(\mathbf{B}^{kl}\right)^2\right]^{\frac{1}{2}}$ where $\mathbf{B}^{kl}$ stands for the $(k,l)-$th coordinate of the matrix $\mathbf{B}$. For any symmetric and positive definite matrix $\mathbf{B}$, $\|\mathbf{B}\|$ denotes the maximum eigenvalue of the matrix $\mathbf{B}$. Both norms are equivalent, that is, there exist constants c_p and C_p depending only on the dimension, such that $c_p\|\mathbf{B}\| \leq |\mathbf{B}| \leq C_p\|\mathbf{B}\|$. However, we distinguish between them in order to simplify the proofs.

In what follows, denote $\mathcal{V}$ and $\mathcal{S}$ neighborhoods of $\boldsymbol{\theta}_0$ and $\mathbf{S}_0 = \mathbf{I}$ respectively, such that for any $\boldsymbol{\theta} \in \mathcal{V}$ and $\mathbf{S} \in \mathcal{S}$ we have that $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \delta_0$, where δ_0 is given in **A7,** and $C^{-1} < \sigma < C$, $|\boldsymbol{\beta}| < C$, $|\lambda| < C$ and $\max\left(\|\mathbf{S}^{-1}\|, \|\mathbf{S}\|\right) \leq C$ for some positive constant C.

Lemma 1. *Under* **A4, A7** *and* **A8,** *if φ is a bounded and continuous function, we have that,*

$$a)\quad \lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0} E\left(\mathbf{H}_j\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right) = \mathbf{A}_j \quad j = 1, 2$$

$$b)\quad \sup_{\boldsymbol{\theta}\in\mathcal{V}} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbf{H}_j\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}\right) - E\left(\mathbf{H}_j\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right)\right| \xrightarrow{p} 0\,,$$

which entails that for any weakly consistent estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{H}_j\left(\mathbf{x}_i, y_i, \widehat{\boldsymbol{\theta}}\right) \xrightarrow{p} \mathbf{A}_j\,,$$

where $\mathbf{A}_j = E\left(\mathbf{H}_j\left(\mathbf{x}, y, \boldsymbol{\theta}_0\right)\right)$ for $j = 1, 2$.

Lemma 2. *Under* **A4, A5** *and* **A7** *to* **A9,** *if φ is a bounded and continuous function, we have that,*

$$a)\quad \lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0, \mathbf{S}\to\mathbf{I}} E\left(\mathbf{H}_3\left(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{S}\right)\right) = \mathbf{A}_3$$

$$b)\quad \sup_{\boldsymbol{\theta}\in\mathcal{V}, \mathbf{S}\in\mathcal{S}} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbf{H}_3\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}, \mathbf{S}\right) - E\left(\mathbf{H}_3\left(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{S}\right)\right)\right| \xrightarrow{p} 0\,,$$

which entails that, for any weakly consistent estimates $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ and $\widehat{\mathbf{S}}$ of the scatter matrix of the carriers

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{H}_3\left(\mathbf{x}_i, y_i, \widehat{\boldsymbol{\theta}}, \widehat{\mathbf{S}}\right) \xrightarrow{p} \mathbf{A}_3\,,$$

where $\mathbf{A}_3 = E\left(\mathbf{H}_3\left(\mathbf{x}, y, \boldsymbol{\theta}_0, \mathbf{I}\right)\right)$.

Remark 3. The conclusion of Lemma 1 also holds for the functions $\mathbf{H}\left(\mathbf{x}, y, \boldsymbol{\theta}\right)$ and $\mathbf{H}_3^*\left(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{S}\right)$, where

$$\mathbf{H}_3^*\left(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{S}\right) = \varphi\left(r(\mathbf{x}, y, \boldsymbol{\theta})\right) w_2\left(\mathbf{S}\,\mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)\right) \mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)\ .$$

Moreover, Lemmas 1 and 2 still hold when $r(\mathbf{x}, y, \boldsymbol{\theta})$ is replaced by either

$$\begin{aligned} r_1(\mathbf{x}, y, \boldsymbol{\theta}) &= \frac{y - \mathbf{x}'\boldsymbol{\beta}_0}{\sigma\, G\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)} \\ &\text{or} \\ r_2(\mathbf{x}, y, \lambda, \boldsymbol{\beta}) &= \frac{y - \mathbf{x}'\boldsymbol{\beta}_0}{\sigma_0\, G\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)}\,. \end{aligned}$$

Furthermore, if we consider a weakly consistent estimate, $\boldsymbol{\xi}_n$, of $\boldsymbol{\beta}_0$ and

$$r_3(\mathbf{x}, y, \boldsymbol{\theta}, \boldsymbol{\xi}) = \frac{y - \mathbf{x}'\boldsymbol{\xi}}{\sigma\, G\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)}\,,$$

we also get that

$$\frac{1}{n}\sum_{i=1}^{n} \widetilde{\mathbf{H}}_3\left(\mathbf{x}_i, y_i, \widehat{\boldsymbol{\theta}}, \widehat{\mathbf{S}}, \boldsymbol{\xi}_n\right) \xrightarrow{p} \mathbf{A}_3\,,$$

where

$$\widetilde{\mathbf{H}}_3\left(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{S}, \boldsymbol{\xi}\right) = \varphi\left(r_3(\mathbf{x}, y, \boldsymbol{\theta}, \boldsymbol{\xi})\right)\, w_2\left(\mathbf{S}\,\mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)\right) \mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)\, \mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)'\,.$$

The Maximal Inequality given in Kim and Pollard (1990) for manageable classes of functions and stated in van der Vaart and Wellner (1996) (Theorem 3.14.1, page 239) provides a useful tool in order to get convergence rates under mild conditions. From this inequality, we will obtain the following two Lemmas using the measurability and entropy conditions described in the Appendix.

Lemma 3. *Under* **A1, A4, A5, A8** *and* **A12,** *if in addition, φ is an odd, continuously differentiable and bounded function with derivative φ' such that $\eta(t) = t\varphi'(t)$ is bounded, we have that for any weakly consistent estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$*

$$J_n\left(\widehat{\boldsymbol{\theta}}\right) \xrightarrow{p} 0\,, \tag{6}$$

where

$$J_n(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbf{H}_5\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}\right) - \mathbf{H}_5\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}_0\right)\right)\,.$$

Lemma 4. *Under* **A1, A4, A5, A8** *and* **A12** *if, in addition, φ is an odd and bounded function we have that for any weakly consistent estimate $(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{S}})$ of $(\boldsymbol{\theta}_0, \mathbf{I})$*

$$J_n\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{S}}\right) \xrightarrow{p} 0\,, \tag{7}$$

with

$$J_n(\boldsymbol{\theta}, \mathbf{S}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbf{H}_4\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}, \mathbf{S}\right) - \mathbf{H}_4\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}, \mathbf{I}\right)\right)\,.$$

2.3 Consistency Results

Proposition 1. *Under* **A1** *to* **A8** *and* **A10**, *we have that for any initial consistent sequence of estimates* $(\boldsymbol{\beta}_{\rm H}, \sigma_{\rm H}, \lambda_{\rm H}, \mathbf{S}_{\rm H})$

$$\boldsymbol{\beta}_{\mathcal{N}} \xrightarrow{p} \boldsymbol{\beta}_0 .$$

PROOF. Using Lemma 2 with $\varphi(t) = \Psi_1'(t)$, we obtain that $\frac{\mathbf{A}_n}{n} \xrightarrow{p} \mathbf{A}$. On the other hand, using Remark 3 with $\varphi(t) = \Psi_1(t)$ we get that $\frac{\mathbf{g}_n}{n} \xrightarrow{p} \mathbf{g}$, where

$$\mathbf{g} = E\left(\Psi_1\left(\frac{y - \mathbf{x}'\boldsymbol{\beta}_0}{\sigma_0\, G\left(\mathbf{x}, \lambda_0, \boldsymbol{\beta}_0\right)}\right) w_2\left(\frac{\mathbf{x}}{G\left(\mathbf{x}, \lambda_0, \boldsymbol{\beta}_0\right)}\right) \frac{\mathbf{x}}{G\left(\mathbf{x}, \lambda_0, \boldsymbol{\beta}_0\right)}\right) . \tag{8}$$

From **A1** and **A2** we get that $\mathbf{g} = \mathbf{0}$, which together with **A10** and the consistency of $\boldsymbol{\beta}_{\rm H}$ and $\sigma_{\rm H}$ entail the desired result.□

Proposition 2. *Let* $\widehat{\boldsymbol{\beta}}$ *be the solution of*

$$\sum_{i=1}^n \Psi_1\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma_{\rm H} G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) w_2\left(\frac{\mathbf{S}_{\rm H}\mathbf{x}_i}{G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) \frac{\mathbf{x}_i}{G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)} = 0 ,$$

where $(\boldsymbol{\beta}_{\rm H}, \sigma_{\rm H}, \lambda_{\rm H}, \mathbf{S}_{\rm H})$ *is any initial consistent sequence of estimates.*

Under **A1** *to* **A8** *and* **A10**, *we have that*

$$\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0 .$$

PROOF. For any $\boldsymbol{\beta}^* \in I\!R^p$ denote

$$\begin{aligned}
\mathbf{L}_n\left(\mathbf{S}, \boldsymbol{\theta}, \boldsymbol{\beta}^*\right) &= \frac{\sigma}{n}\sum_{i=1}^n \Psi_1\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}^*}{\sigma\, G\left(\mathbf{x}_i, \lambda, \boldsymbol{\beta}\right)}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})\right)\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta}) \\
\mathbf{L}_n^{(1)}\left(\mathbf{S}, \boldsymbol{\theta}, \boldsymbol{\beta}^*\right) &= \frac{1}{n}\sum_{i=1}^n \Psi_1'\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}^*}{\sigma\, G\left(\mathbf{x}_i, \lambda, \boldsymbol{\beta}\right)}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})\right)\, \mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})\, \mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})' .
\end{aligned}$$

Using a first order Taylor's expansion around $\boldsymbol{\beta}_0$, we get

$$\mathbf{L}_n\left(\mathbf{S}, \boldsymbol{\theta}, \widehat{\boldsymbol{\beta}}\right) = \mathbf{L}_n\left(\mathbf{S}, \boldsymbol{\theta}, \boldsymbol{\beta}_0\right) + \mathbf{L}_n^{(1)}\left(\mathbf{S}, \boldsymbol{\theta}, \boldsymbol{\xi}\right)(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}})$$

with $\boldsymbol{\xi}$ an intermediate point between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. This implies that

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \left(\frac{\widetilde{\mathbf{A}}_n}{n}\right)^{-1} \mathbf{L}_n\left(\mathbf{S}_{\rm H}, \boldsymbol{\theta}_{\rm H}, \boldsymbol{\beta}_0\right) , \tag{9}$$

where $\widetilde{\mathbf{A}}_n = \mathbf{L}_n^{(1)}\left(\mathbf{S}_{\rm H}, \boldsymbol{\theta}_{\rm H}, \boldsymbol{\xi}\right)$. Using **A8**, the boundness of Ψ_1' and of Ψ_2, we obtain that $\frac{\widetilde{\mathbf{A}}_n}{n}$ is bounded in probability. On the other hand, using Remark 3 with $\varphi(t) = \Psi_1'(t)$ and $\varphi(t) = \Psi_1(t)$, we obtain $\mathbf{L}_n\left(\mathbf{S}_{\rm H}, \boldsymbol{\theta}_{\rm H}, \boldsymbol{\beta}_0\right) \xrightarrow{p} \sigma\mathbf{g}$, where $\mathbf{g}$ is defined in (8) and equals $\mathbf{0}$. Therefore, as in Proposition 1, from **A1**, **A2**, **A10** and the consistency of $\boldsymbol{\beta}_{\rm H}$ and $\sigma_{\rm H}$ we obtain the desired result. □

2.4 Asymptotic Distribution

In this section we will assume that the initial estimates $\boldsymbol{\beta}_{\rm H}$ have rate of convergence n^{τ} with $\tau \in (1/4, 1/2]$. Theorem 1 is derived under **A12,** by requiring only consistency to $\lambda_{\rm H}$, $\sigma_{\rm H}$ and $\mathbf{S}_{\rm H}$. On the other hand, Theorem 3 does not need assumption **A12,** but requires, to the estimates $\mathbf{S}_{\rm H}$, an order of convergence n^{ν}, with $\nu \in (1/4, 1/2]$. Using similar arguments to those given in Simpson, Ruppert and Carroll (1992), we will obtain the asymptotic distribution of one–step Newton–Raphson estimates in the following Theorem.

Theorem 1. *Under* **A1** *to* **A5** *and* **A7** *to* **A12** , *we have that*

$$\sqrt{n}\,(\boldsymbol{\beta}_{\mathcal{N}} - \boldsymbol{\beta}_0) \xrightarrow{D} N(0, \boldsymbol{\Sigma})\,,$$

where $\boldsymbol{\Sigma} = \sigma_0^2 \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ *with* $\mathbf{A}$ *defined in* **A11** *and*

$$\mathbf{B} = E\left(\Psi_1^2\left(r(\mathbf{x}, y, \boldsymbol{\theta}_0)\right)\, w_2^2\left(\mathbf{z}(\mathbf{x}, \lambda_0, \boldsymbol{\beta}_0)\right) \mathbf{z}(\mathbf{x}, \lambda_0, \boldsymbol{\beta}_0)\, \mathbf{z}(\mathbf{x}, \lambda_0, \boldsymbol{\beta}_0)'\right)\,,$$

for any initial consistent estimates $(\boldsymbol{\beta}_{\rm H}, \sigma_{\rm H}, \lambda_{\rm H}, \mathbf{S}_{\rm H})$ *such that* $\boldsymbol{\beta}_{\rm H}$ *has order* n^{τ} *with* $\tau \in (\frac{1}{4}, \frac{1}{2}]$.

Proof. Denote

$$\mathbf{L}_n\left(\mathbf{S}, \boldsymbol{\theta}, \boldsymbol{\beta}^*\right) = \frac{\sigma}{n}\sum_{i=1}^{n} \Psi_1\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}^*}{\sigma\, G\left(\mathbf{x}_i, \lambda, \boldsymbol{\beta}\right)}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})\right) \mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})\,.$$

Using a second order Taylor's expansion around $\boldsymbol{\beta}_{\rm H}$, we get

$$\begin{aligned}
\mathbf{L}_n\left(\mathbf{S}, \boldsymbol{\theta}, \boldsymbol{\beta}_0\right) &= \frac{\sigma}{n}\sum_{i=1}^{n} \Psi_1\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}_{\rm H}}{\sigma\, G\left(\mathbf{x}_i, \lambda, \boldsymbol{\beta}\right)}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})\right) \mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta}) \\
&+ \frac{1}{n}\sum_{i=1}^{n} \Psi_1'\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}_{\rm H}}{\sigma\, G\left(\mathbf{x}_i, \lambda, \boldsymbol{\beta}\right)}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})\right) \mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})'\left(\boldsymbol{\beta}_{\rm H} - \boldsymbol{\beta}_0\right) \\
&+ \frac{1}{2\,n\,\sigma}\sum_{i=1}^{n} \Psi_1''\left(\frac{y_i - \mathbf{x}_i'\widetilde{\boldsymbol{\beta}}}{\sigma\, G\left(\mathbf{x}_i, \lambda, \boldsymbol{\beta}\right)}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})\right) \mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta}) \times \\
&\quad \times \mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})'\left(\boldsymbol{\beta}_{\rm H} - \boldsymbol{\beta}_0\right)\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})'\left(\boldsymbol{\beta}_{\rm H} - \boldsymbol{\beta}_0\right)\,,
\end{aligned}$$

with $\widetilde{\boldsymbol{\beta}}$ an intermediate point between $\boldsymbol{\beta}_{\rm H}$ and $\boldsymbol{\beta}_0$. This implies that

$$\mathbf{L}_n\left(\mathbf{S}_{\rm H}, \boldsymbol{\theta}_{\rm H}, \boldsymbol{\beta}_0\right) = \mathbf{g}_n \frac{\sigma_{\rm H}}{n} + \frac{\mathbf{A}_n}{n}\left(\boldsymbol{\beta}_{\rm H} - \boldsymbol{\beta}_0\right) + \mathbf{R}_n\,, \tag{10}$$

where

$$\begin{aligned}
\mathbf{R}_n &= \frac{1}{2\,\sigma_{\rm H}}\frac{1}{n}\sum_{i=1}^{n} \Psi_1''\left(\frac{y_i - \mathbf{x}_i'\widetilde{\boldsymbol{\beta}}}{\sigma_{\rm H}\, G\left(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H}\right)}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})\right)\, \mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H}) \times \\
&\quad \times \left\{\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})'\left(\boldsymbol{\beta}_{\rm H} - \boldsymbol{\beta}_0\right)\right\}^2\,.
\end{aligned}$$

From (10) and the definition of $\boldsymbol{\beta}_{\mathcal{N}}$, we get

$$\mathbf{L}_n\left(\mathbf{S}_{\rm H}, \boldsymbol{\theta}_{\rm H}, \boldsymbol{\beta}_0\right) = \frac{\mathbf{A}_n}{n}\left(\boldsymbol{\beta}_{\mathcal{N}} - \boldsymbol{\beta}_0\right) + \mathbf{R}_n\,. \tag{11}$$

Using Lemma 2 with $\varphi(t) = \Psi_1'(t)$, we obtain that

$$\frac{\mathbf{A}_n}{n} \xrightarrow{p} \mathbf{A}\,.$$

On the other hand, since we can write

$$\begin{aligned}\mathbf{L}_n\left(\mathbf{S}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\beta}_0\right) - \mathbf{L}_n\left(\mathbf{I}, \boldsymbol{\theta}_0, \boldsymbol{\beta}_0\right) &= \mathbf{L}_n\left(\mathbf{S}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\beta}_0\right) - \mathbf{L}_n\left(\mathbf{I}, \boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\beta}_0\right) \\ &+ \mathbf{L}_n\left(\mathbf{I}, \boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\beta}_0\right) - \mathbf{L}_n\left(\mathbf{I}, \boldsymbol{\theta}_0, \boldsymbol{\beta}_0\right)\end{aligned}$$

from Lemmas 3 and 4 and the consistency of $\mathbf{S}_{\mathrm{H}}$ and $\boldsymbol{\theta}_{\mathrm{H}}$, we have that

$$\sqrt{n}\left(\mathbf{L}_n\left(\mathbf{S}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\beta}_0\right) - \mathbf{L}_n\left(\mathbf{I}, \boldsymbol{\theta}_0, \boldsymbol{\beta}_0\right)\right) \xrightarrow{p} 0\,.$$

Thus,

$$\frac{\mathbf{A}_n}{n}\sqrt{n}\left(\boldsymbol{\beta}_{\mathcal{N}} - \boldsymbol{\beta}_0\right) + \sqrt{n}\mathbf{R}_n$$

has the same asymptotic distribution as $\mathbf{Z}_n = \sqrt{n}\,\mathbf{L}_n\left(\mathbf{I}, \boldsymbol{\theta}_0, \boldsymbol{\beta}_0\right)$. Since $\mathbf{Z}_n$ is asymptotically normally distributed with zero mean and covariance matrix $\sigma_0^2\mathbf{B}$, it only remains to prove that

$$\sqrt{n}\,\mathbf{R}_n \xrightarrow{p} 0\,. \tag{12}$$

It is easy to see that

$$|\mathbf{R}_n| \le \frac{\|\mathbf{S}_{\mathrm{H}}^{-1}\|}{2\sigma_{\mathrm{H}}}\|\Psi_1''\|_\infty \frac{1}{n}\sum_{i=1}^n \Psi_2\left(|\mathbf{S}_{\mathrm{H}}\,\mathbf{z}\left(\mathbf{x}_i, \lambda_{\mathrm{H}}, \boldsymbol{\beta}_{\mathrm{H}}\right)|\right)|\mathbf{z}\left(\mathbf{x}_i, \lambda_{\mathrm{H}}, \boldsymbol{\beta}_{\mathrm{H}}\right)|^2|\boldsymbol{\beta}_{\mathrm{H}} - \boldsymbol{\beta}_0|^2\,.$$

Thus, from **A10** we obtain that $\mathbf{R}_n = O_p\left(|\boldsymbol{\beta}_{\mathrm{H}} - \boldsymbol{\beta}_0|^2\right)$, which implies (12) since $\tau \in (\frac{1}{4}, \frac{1}{2}]$. □

Remark 4. When $\tau = 1/2$, one can only require to the score function Ψ_1 continuous first differentiability, by using a first order Taylor's expansion. Besides, **A10** will not be necessary.

Theorem 2. *Let $\widehat{\boldsymbol{\beta}}$ be the solution of*

$$\sum_{i=1}^n \Psi_1\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma_{\mathrm{H}}G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\mathrm{H}}, \boldsymbol{\beta}_{\mathrm{H}}\right)}\right) w_2\left(\frac{\mathbf{S}_{\mathrm{H}}\mathbf{x}_i}{G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\mathrm{H}}, \boldsymbol{\beta}_{\mathrm{H}}\right)}\right)\frac{\mathbf{x}_i}{G\left(\mathbf{x}_i, \boldsymbol{\lambda}_{\mathrm{H}}, \boldsymbol{\beta}_{\mathrm{H}}\right)} = 0\,, \tag{13}$$

where $(\boldsymbol{\beta}_{\mathrm{H}}, \sigma_{\mathrm{H}}, \lambda_{\mathrm{H}}, \mathbf{S}_{\mathrm{H}})$ is any initial consistent sequence of estimates.

Under **A1** *to* **A5** *and* **A7** *to* **A12** , *we have that*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{D} N(0, \boldsymbol{\Sigma})\,,$$

where $\boldsymbol{\Sigma} = \sigma_0^2\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ with $\mathbf{A}$ defined in **A11** *and*

$$\mathbf{B} = E\left(\Psi_1^2\left(r(\mathbf{x}, y, \boldsymbol{\theta}_0)\right)\, w_2^2\left(\mathbf{z}(\mathbf{x}, \lambda_0, \boldsymbol{\beta}_0)\right)\mathbf{z}(\mathbf{x}, \lambda_0, \boldsymbol{\beta}_0)\,\mathbf{z}(\mathbf{x}, \lambda_0, \boldsymbol{\beta}_0)'\right)\,.$$

PROOF. As in Theorem 1, denote

$$\mathbf{L}_n(\mathbf{S},\boldsymbol{\theta},\boldsymbol{\beta}^*) = \frac{\sigma}{n}\sum_{i=1}^{n}\Psi_1\left(\frac{y_i-\mathbf{x}_i'\boldsymbol{\beta}^*}{\sigma\, G(\mathbf{x}_i,\lambda,\boldsymbol{\beta})}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i,\lambda,\boldsymbol{\beta})\right)\mathbf{z}(\mathbf{x}_i,\lambda,\boldsymbol{\beta})\,.$$

Using a first order Taylor's expansion around $\boldsymbol{\beta}_0$, we get

$$\begin{aligned}\mathbf{L}_n\left(\mathbf{S},\boldsymbol{\theta},\widehat{\boldsymbol{\beta}}\right) &= \mathbf{L}_n(\mathbf{S},\boldsymbol{\theta},\boldsymbol{\beta}_0) + \\ &+ \frac{1}{n}\sum_{i=1}^{n}\Psi_1'\left(\frac{y_i-\mathbf{x}_i'\boldsymbol{\xi}}{\sigma\, G(\mathbf{x}_i,\lambda,\boldsymbol{\beta})}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i,\lambda,\boldsymbol{\beta})\right)\mathbf{z}(\mathbf{x}_i,\lambda,\boldsymbol{\beta})\mathbf{z}(\mathbf{x}_i,\lambda,\boldsymbol{\beta})'(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})\,,\end{aligned}$$

with $\boldsymbol{\xi}$ an intermediate point between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. This implies that

$$(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = \left(\frac{\widetilde{\mathbf{A}}_n}{n}\right)^{-1}\mathbf{L}_n(\mathbf{S}_{\mathrm{H}},\boldsymbol{\theta}_{\mathrm{H}},\boldsymbol{\beta}_0)\ , \tag{14}$$

where

$$\widetilde{\mathbf{A}}_n = \sum_{i=1}^{n}\Psi_1'\left(\frac{y_i-\mathbf{x}_i'\boldsymbol{\xi}}{\sigma_{\mathrm{H}}G(\mathbf{x}_i,\boldsymbol{\lambda}_{\mathrm{H}},\boldsymbol{\beta}_{\mathrm{H}})}\right) w_2\left(\frac{\mathbf{S}_{\mathrm{H}}\mathbf{x}_i}{G(\mathbf{x}_i,\boldsymbol{\lambda}_{\mathrm{H}},\boldsymbol{\beta}_{\mathrm{H}})}\right)\frac{\mathbf{x}_i\,\mathbf{x}_i'}{G^2(\mathbf{x}_i,\boldsymbol{\lambda}_{\mathrm{H}},\boldsymbol{\beta}_{\mathrm{H}})}\,.$$

Using Remark 3 with $\varphi(t)=\Psi_1'(t)$, we obtain that

$$\frac{\widetilde{\mathbf{A}}_n}{n}\xrightarrow{p}\mathbf{A}\,.$$

Since as in Theorem 1, we have that

$$\sqrt{n}\left(\mathbf{L}_n(\mathbf{S}_{\mathrm{H}},\boldsymbol{\theta}_{\mathrm{H}},\boldsymbol{\beta}_0)-\mathbf{L}_n(\mathbf{I},\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)\right)\xrightarrow{p}0\,,$$

$\sqrt{n}\,(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ has the same asymptotic distribution as $\left(\frac{\widetilde{\mathbf{A}}_n}{n}\right)^{-1}\mathbf{Z}_n$, where $\mathbf{Z}_n = \sqrt{n}\,\mathbf{L}_n(\mathbf{I},\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)$. Finally, the desired result follows from the fact that $\mathbf{Z}_n$ is asymptotically normally distributed with zero mean and covariance matrix $\sigma_0^2\mathbf{B}$. □

The following Theorem requires an entropy condition, as given in van der Vaart and Wellner (1996, page 127) and described in the Appendix. Its proof is also given in the Appendix.

Theorem 3. *Assume that* **A1** *to* **A11** *hold. If in addition,*

a) *the estimates* $\mathbf{S}_{\mathrm{H}}$ *have order of convergence* n^{ν} *with* $\nu\in(1/4,1/2]$,

and

b) *the class of functions*

$$\begin{aligned}\mathcal{F} &= \{f_{\boldsymbol{\theta}}(\mathbf{x},y) = \Psi_1(r_1(\mathbf{x},y,\boldsymbol{\theta}))\ \Psi_2(|\mathbf{z}(\mathbf{x},\lambda,\boldsymbol{\beta})|) - \\ &- \Psi_1(r_1(\mathbf{x},y,\boldsymbol{\theta}_0))\ \Psi_2(|\mathbf{z}(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0)|)\ ,\ \boldsymbol{\theta}\in\mathcal{V}\}\end{aligned}$$

with enveloppe $F = 2\,\|\Psi_1\|_\infty\,\|\Psi_2\|_\infty$, *has finite uniform–entropy,*

c) *the class of functions*

$$\begin{aligned}\mathcal{G} &= \{g_{\boldsymbol{\theta}}(\mathbf{x},y) = \Psi_1\left(r_1\left(\mathbf{x},y,\boldsymbol{\theta}\right)\right)\,\eta_2\left(\left|\mathbf{z}\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)\right|\right) - \\ &- \Psi_1\left(r_1\left(\mathbf{x},y,\boldsymbol{\theta}_0\right)\right)\,\eta_2\left(\left|\mathbf{z}\left(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0\right)\right|\right)\,,\ \boldsymbol{\theta}\in\mathcal{V}\}\end{aligned}$$

with enveloppe $G = 2\,\|\Psi_1\|_\infty\,\|\eta_2\|_\infty$, *has finite uniform–entropy,*

we have that

$$\sqrt{n}\left(\boldsymbol{\beta}_{\mathcal{N}} - \boldsymbol{\beta}_0\right) \xrightarrow{D} N(0,\boldsymbol{\Sigma})\,,$$

where $\boldsymbol{\Sigma} = \sigma_0^2\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ *with* $\mathbf{A}$ *defined in* **A11** *and*

$$\mathbf{B} = E\left(\Psi_1^2\left(r(\mathbf{x},y,\boldsymbol{\theta}_0)\right)\,w_2^2\left(\mathbf{z}(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0)\right)\mathbf{z}(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0)\,\mathbf{z}(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0)'\right)\,,$$

for any initial consistent sequence of estimates $(\boldsymbol{\beta}_{\mathrm{H}},\sigma_{\mathrm{H}},\lambda_{\mathrm{H}},\mathbf{S}_{\mathrm{H}})$ *such that* $\boldsymbol{\beta}_{\mathrm{H}}$ *has order* n^τ *with* $\tau\in(\frac{1}{4},\frac{1}{2}]$.

An analogous result can be obtained for the solution $\widehat{\boldsymbol{\beta}}$ of (13).

The following Proposition, whose proof is given in the Appendix, gives a condition under which the class $\mathcal{F}$ defined in Theorem 3 has finite uniform–entropy.

Proposition 3. *Assume that there exists a monotone function* $m(t)$ *such that the class of functions* $\{m\left(h(\mathbf{x},\boldsymbol{\beta})\right) : |\boldsymbol{\beta}-\boldsymbol{\beta}_0| < \delta_0\}$ *has finite dimension and that* Ψ_1 *and* Ψ_2 *are bounded and bounded variation functions, then the class of functions defined by*

$$\begin{aligned}\mathcal{F} &= \{f_{\boldsymbol{\theta}}(\mathbf{x},y) = \Psi_1\left(r_1\left(\mathbf{x},y,\boldsymbol{\theta}\right)\right)\,\Psi_2\left(\left|\mathbf{z}\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)\right|\right) - \\ &- \Psi_1\left(r_1\left(\mathbf{x},y,\boldsymbol{\theta}_0\right)\right)\,\Psi_2\left(\left|\mathbf{z}\left(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0\right)\right|\right)\,,\ \boldsymbol{\theta}\in\mathcal{V}\}\end{aligned}$$

has finite uniform–entropy.

Remark 5. Note that, for instance, when $h(\mathbf{x},\boldsymbol{\beta}) = h(\mathbf{x})$, Theorem 3 and Proposition 3 entail that the one–step Newton–Raphson estimate has the same asymptotic distribution as the related GM–estimate, without requiring any moment condition to the variance function $h(\mathbf{x})$, if Ψ_1, Ψ_2 and Ψ_2' are bounded variation functions. This latter condition is fulfilled for most score functions used in robust estimation. Proposition 3 also includes the variance functions $G(\mathbf{x},\lambda,\boldsymbol{\beta}) = (1+|\mathbf{x}'\boldsymbol{\beta}|)^\lambda$ or $G(\mathbf{x},\lambda,\boldsymbol{\beta}) = \exp\{\lambda|\mathbf{x}'\boldsymbol{\beta}|\}$ which were introduced by Box and Hill (1974) and by Bickel (1978), respectively.

Remark 6. From the definition of the reweighted estimate given in (3) and using (10), one can obtain that

$$\boldsymbol{\beta}_{\mathcal{R}} - \boldsymbol{\beta}_0 = \left(\mathbf{I} - \left(\frac{\mathbf{B}_n}{n}\right)^{-1}\frac{\mathbf{A}_n}{n}\right)\left(\boldsymbol{\beta}_{\mathrm{H}} - \boldsymbol{\beta}_0\right) + \left(\frac{\mathbf{B}_n}{n}\right)^{-1}\mathbf{L}_n\left(\mathbf{S}_{\mathrm{H}},\boldsymbol{\theta}_{\mathrm{H}},\boldsymbol{\beta}_0\right) - \left(\frac{\mathbf{B}_n}{n}\right)^{-1}\mathbf{R}_n\,.$$

Similar arguments to those used in Theorems 1 and 3, entail that the reweighted estimate has the same order of convergence as the initial high breakdown point estimate, if the matrix

$$\mathbf{B} = E\left(w_1\left(\frac{y-\mathbf{x}'\boldsymbol{\beta}_0}{\sigma_0\, G\left(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0\right)}\right) w_2\left(\frac{\mathbf{x}}{G\left(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0\right)}\right)\frac{\mathbf{x}\,\mathbf{x}'}{G^2\left(\mathbf{x},\lambda_0,\boldsymbol{\beta}_0\right)}\right)$$

is non–singular and if $\mathbf{B}^{-1}\mathbf{A}$ is not the identity matrix, where $\mathbf{A}$ is defined in assumption **A11** .

3 Appendix

3.1 Proofs of Lemmas 1 to 4

Proof of Lemma 1. We will begin by proving (a). The boundness of φ and Ψ_2 entails that $|\mathbf{H}_j\left(\mathbf{x},y,\boldsymbol{\theta}\right)-\mathbf{H}_j\left(\mathbf{x},y,\boldsymbol{\theta}_0\right)|$ is bounded. Thus, from **A8,** the continuity of φ and Ψ_2 and the Dominated Convergence Theorem, we have that

$$\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0} |E\left(\mathbf{H}_j\left(\mathbf{x},y,\boldsymbol{\theta}\right)\right)-E\left(\mathbf{H}_j\left(\mathbf{x},y,\boldsymbol{\theta}_0\right)\right)|=0$$

concluding (a).

(b) From Theorem 3, Chapter 2 of Pollard (1984) it will be enough to show that for each $\eta>0$ there exists a finite class $\mathcal{H}_\eta$ such that, for each $\boldsymbol{\theta}\in\mathcal{V}$ there exist functions $\mathbf{H}_{\eta,\mathrm{L}}$ and $\mathbf{H}_{\eta,\mathrm{U}}$ in $\mathcal{H}_\eta$ such that

$$\mathbf{H}^{kl}_{\eta,\mathrm{L}}(\mathbf{x},y)\le\mathbf{H}^{kl}(\mathbf{x},y,\boldsymbol{\theta})\le\mathbf{H}^{kl}_{\eta,\mathrm{U}}(\mathbf{x},y) \tag{15}$$

and

$$E\left(\mathbf{H}^{kl}_{\eta,\mathrm{U}}(\mathbf{x},y)-\mathbf{H}^{kl}_{\eta,\mathrm{L}}(\mathbf{x},y)\right)\le\eta\,, \tag{16}$$

where $\mathbf{H}$ denotes either $\mathbf{H}_1$ or $\mathbf{H}_2$.

Given $K\in I\!N$, denote $K_1=[\;K\left(|\boldsymbol{\beta}_0|+\sigma_0\,C_1\right)\;]\;+1$, where $[\;t\;]$ stands for the integer part of the real number t, $C_1=\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta_0,|\mathbf{x}|\le K} G\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)$ and define the sets

$$\mathcal{A}_K=\{(\mathbf{x},y)\,:\,|\mathbf{x}|\le K\,,\,|r\left(\mathbf{x},y,\boldsymbol{\theta}_0\right)|\le K\}\quad\text{and}\quad\mathcal{B}_K=\{|\mathbf{x}|\le K\,,\,|y|\le K_1\}\,.$$

Note that $\mathcal{A}_K\subset\mathcal{B}_K$.

Let $K\in I\!N$, be such that

$$P\left(\mathcal{A}_K\right)>1-\eta_1\,, \tag{17}$$

where $\eta_1=\eta/\left(5\,M^2\right)$ with $M=\|\varphi\|_\infty\|\Psi_2\|_\infty$.

From **A7** , $C_2=\inf_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta_0,|\mathbf{x}|\le K} G\left(\mathbf{x},\lambda,\boldsymbol{\beta}\right)$ is positive. Hence, for $(\mathbf{x},y)$ in $\mathcal{B}_K$ and $\boldsymbol{\theta}$ in $\mathcal{V}$ we have that $|r(\mathbf{x},y,\boldsymbol{\theta})|\le B_1$ and $|\mathbf{z}(\mathbf{x},\lambda,\boldsymbol{\beta})|\le B_2$, where $B_1=(K_1+K\,C)\,C/C_2$ and $B_2=K/C_2$.

Since the functions $\varphi(t)$ and $w_2(\mathbf{z})\mathbf{z}$ are continuous, they are uniformly continuous in $\mathcal{C}_K = \{|t| \le B_1\,,\, |\mathbf{z}| \le B_2\}$. Therefore, there exists $\delta > 0$ such that

$$|\varphi^2(t)w_2^2(\mathbf{z})\,\mathbf{z}\,\mathbf{z}' - \varphi^2(u)w_2^2(\mathbf{v})\,\mathbf{v}\,\mathbf{v}'| \le \frac{\eta}{10} \tag{18}$$

and

$$|\varphi(t)w_2(\mathbf{z})\,\mathbf{z}\; - \varphi(u)w_2(\mathbf{v})\,\mathbf{v}\,| \le \frac{\eta}{10\,M} \tag{19}$$

for $|t-u| < \delta$, $|\mathbf{z}-\mathbf{v}| < \delta$ and $(t, \mathbf{z})$ and $(u, \mathbf{v})$ in $\mathcal{C}_K$.

From **A7** and **A8,** in $\mathcal{B}_K$, we have that $r(\mathbf{x}, y, \boldsymbol{\theta})$ and $\mathbf{z}(\mathbf{x}, \lambda, \boldsymbol{\beta})$ are equicontinuous functions of $\boldsymbol{\theta}$, for $\boldsymbol{\theta}$ in $\mathcal{V}$, i.e., there exists $\delta_1 > 0$ such that for $|\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}| < \delta_1$, $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ in $\mathcal{V}$, $(\mathbf{x}, y)$ in $\mathcal{B}_K$, we have

$$|r(\mathbf{x}, y, \boldsymbol{\theta}) - r(\mathbf{x}, y, \widetilde{\boldsymbol{\theta}})| < \delta \tag{20}$$

and

$$|\mathbf{z}(\mathbf{x}, \lambda, \boldsymbol{\beta}) - \mathbf{z}(\mathbf{x}, \widetilde{\lambda}, \widetilde{\boldsymbol{\beta}})| < \delta\;. \tag{21}$$

Let $(\mathcal{V}_i)_{1 \le i \le N}$ be a finite collection of balls centered at points $\boldsymbol{\theta}_i \in \mathcal{V}$ with radius smaller than δ_1 such that $\mathcal{V} \subset \cup_{i=1}^N \mathcal{V}_i$.

Given $\boldsymbol{\theta} \in \mathcal{V}$, let i be such that $\boldsymbol{\theta} \in \mathcal{V}_i$. Define

$$\begin{aligned} \mathbf{H}^{kl}_{\eta,\mathrm{L}}(\mathbf{x}, y) &= \mathbf{H}^{kl}(\mathbf{x}, y, \boldsymbol{\theta}_i) - D(\mathbf{x}, y) \\ \mathbf{H}^{kl}_{\eta,\mathrm{U}}(\mathbf{x}, y) &= \mathbf{H}^{kl}(\mathbf{x}, y, \boldsymbol{\theta}_i) + D(\mathbf{x}, y)\,, \end{aligned}$$

where

$$D(\mathbf{x}, y) = \frac{\eta}{10} + 2\,M^2\,\mathrm{I}_{\mathcal{A}_K^c}(\mathbf{x}, y)\;.$$

For the sake of simplicity, we have omitted the subscript i in the functions $\mathbf{H}^{kl}_{\eta,\mathrm{L}}$ and $\mathbf{H}^{kl}_{\eta,\mathrm{U}}$.

Using (18) to (21) and the fact that $|\boldsymbol{\theta} - \boldsymbol{\theta}_i| < \delta_1$, it is easy to see that

$$|\mathbf{H}^{kl}(\mathbf{x}, y, \boldsymbol{\theta}) - \mathbf{H}^{kl}(\mathbf{x}, y, \boldsymbol{\theta}_i)| \le D(\mathbf{x}, y)\,.$$

Therefore, $\mathbf{H}^{kl}_{\eta,\mathrm{L}}(\mathbf{x}, y)$ and $\mathbf{H}^{kl}_{\eta,\mathrm{U}}(\mathbf{x}, y)$ satisfy (15).

It remains to show (16). Since

$$E\left(\mathbf{H}^{kl}_{\eta,\mathrm{U}}(\mathbf{x}, y) - \mathbf{H}^{kl}_{\eta,\mathrm{L}}(\mathbf{x}, y)\right) = 2E\left(D(\mathbf{x}, y)\right) = \frac{\eta}{5} + 4\,M^2\,\left(1 - P\left(\mathcal{A}_K\right)\right)\,,$$

(16) follows from inequality (17). □

Proof of Lemma 2. We will just point the differences with the proof of Lemma 1.

(a) Note that from Remark 2 there exists $c > 0$ such that

$$|w_2(\mathbf{x}) - w_2(\mathbf{z})| \le c|\mathbf{z} - \mathbf{x}|/[\min(|\mathbf{x}|, |\mathbf{z}|)]^2.$$

Then, for any $\boldsymbol{\theta} \in \mathcal{V}$, $\mathbf{V} \in \mathcal{S}$ and $\mathbf{W} \in \mathcal{S}$, we have that

$$|\mathbf{H}_3(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{W}) - \mathbf{H}_3(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{V})| \le c\, C^2\, \|\varphi\|_\infty \|\mathbf{W} - \mathbf{V}\| \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta_0} |\mathbf{z}(\mathbf{x}, \lambda, \boldsymbol{\beta})| \,.$$

On the other hand, since

$$|\mathbf{H}_3(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{S}) - \mathbf{H}_3(\mathbf{x}, y, \boldsymbol{\theta}_0, \mathbf{S})| \le 2\, C\, \|\varphi\|_\infty\, \|\Psi_2\|_\infty \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta_0} |\mathbf{z}(\mathbf{x}, \lambda, \boldsymbol{\beta})| \,,$$

using **A8, A9,** the continuity of φ and Ψ_2 and the Dominated Convergence Theorem, after some algebra we obtain (a).

(b) As in Lemma 1, it will be enough to show that for each $\eta > 0$ there exists a finite class $\mathcal{H}_\eta$ such that, for each $\boldsymbol{\theta} \in \mathcal{V}$ and $\mathbf{S} \in \mathcal{S}$ there exist functions $\mathbf{H}_{\eta,\mathrm{L}}$ and $\mathbf{H}_{\eta,\mathrm{U}}$ in $\mathcal{H}_\eta$ such that

$$\mathbf{H}^{kl}_{\eta,\mathrm{L}}(\mathbf{x}, y) \le \mathbf{H}^{kl}_3(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{S}) \le \mathbf{H}^{kl}_{\eta,\mathrm{U}}(\mathbf{x}, y) \tag{22}$$

and

$$E\left(\mathbf{H}^{kl}_{\eta,\mathrm{U}}(\mathbf{x}, y) - \mathbf{H}^{kl}_{\eta,\mathrm{L}}(\mathbf{x}, y)\right) \le \eta \,. \tag{23}$$

Let $K \in I\!N$ be such that

$$E\left(\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta_0} |\mathbf{z}(\mathbf{x}, \lambda, \boldsymbol{\beta})|\, \mathrm{I}_{\mathcal{A}^c_K}(\mathbf{x}, y)\right) < \eta_1 \tag{24}$$

where $\mathcal{A}_K$ is as in Lemma 1 and $\eta_1 = \eta/(5\, C\, M)$ with $M = \|\varphi\|_\infty\, \|\Psi_2\|_\infty$.

From the continuity of $T(\mathbf{z}, \mathbf{S}) = \mathbf{S}\mathbf{z}$, we have that $w_2(\mathbf{S}\mathbf{z})\mathbf{z}$ is a continuous function of $(\mathbf{z}, \mathbf{S})$ for any nonsingular $\mathbf{S}$. Thus, as in Lemma 1, we have that $\varphi(t)$ and $w_2(\mathbf{S}\mathbf{z})\mathbf{z}$ are uniformly continuous in $\mathcal{C}_K \times \{\max\left(\|\mathbf{S}^{-1}\|, \|\mathbf{S}\|\right) \le C\}$ with $\mathcal{C}_K = \{|t| \le B_1\,, |\mathbf{z}| \le B_2\}$, where as above $B_1 = (K_1 + K\, C)\, C/C_2$ and $B_2 = K/C_2$. Then, there exists $\delta > 0$ such that

$$|\varphi(t) w_2(\mathbf{S}\mathbf{z})\, \mathbf{z}\, \mathbf{z}' - \varphi(u) w_2(\mathbf{W}\mathbf{v})\, \mathbf{v}\, \mathbf{v}'| \le \frac{\eta}{10} \,, \tag{25}$$

for $|t - u| < \delta$, $|\mathbf{z} - \mathbf{v}| < \delta$, $\|\mathbf{S} - \mathbf{W}\| < \delta$ and $(t, \mathbf{z}, \mathbf{S})$ and $(u, \mathbf{v}, \mathbf{W})$ in $\mathcal{C}_K \times \{\max\left(\|\mathbf{S}^{-1}\|, \|\mathbf{S}\|\right) \le C\}$.

Let $(\mathcal{S}_j)_{1\le j\le N_1}$ be a finite collection of balls centered at points $\mathbf{S}_j \in \mathcal{S}$ with radius smaller than δ such that $\mathcal{S} \subset \bigcup_{j=1}^{N_1} \mathcal{S}_j$.

Since $r(\mathbf{x}, y, \boldsymbol{\theta})$ and $\mathbf{z}(\mathbf{x}, \lambda, \boldsymbol{\beta})$ are equicontinuous functions of $\boldsymbol{\theta}$ in $\mathcal{B}_K = \{|\mathbf{x}| \le K\,, |y| \le K_1\}$ with K_1 defined as in Lemma 1 for $\boldsymbol{\theta}$ in $\mathcal{V}$, we obtain that there exists $\delta_1 > 0$ and a finite collection of balls $(\mathcal{V}_i)_{1\le i\le N_2}$ centered at points $\boldsymbol{\theta}_i \in \mathcal{V}$ with radius smaller than δ_1 such that $\mathcal{V} \subset \bigcup_{i=1}^{N_2} \mathcal{V}_i$ and

$$|r(\mathbf{x}, y, \boldsymbol{\theta}) - r(\mathbf{x}, y, \widetilde{\boldsymbol{\theta}})| < \delta \tag{26}$$

and

$$|\mathbf{z}(\mathbf{x},\lambda,\boldsymbol{\beta}) - \mathbf{z}(\mathbf{x},\tilde{\lambda},\tilde{\boldsymbol{\beta}})| < \delta \,, \tag{27}$$

for $|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}| < \delta_1$, $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$ in $\mathcal{V}$, $(\mathbf{x}, y)$ in $\mathcal{B}_K$.

Given $\boldsymbol{\theta} \in \mathcal{V}$ and $\mathbf{S} \in \mathcal{S}$, let (i,j) be such that $\boldsymbol{\theta} \in \mathcal{V}_i$ and $\mathbf{S} \in \mathcal{S}_j$. Define

$$\begin{aligned} \mathbf{H}^{kl}_{\eta,\mathrm{L}}(\mathbf{x},y) &= \mathbf{H}^{kl}_3(\mathbf{x},y,\boldsymbol{\theta}_i,\mathbf{S}_j) - D(\mathbf{x},y) \\ \mathbf{H}^{kl}_{\eta,\mathrm{U}}(\mathbf{x},y) &= \mathbf{H}^{kl}_3(\mathbf{x},y,\boldsymbol{\theta}_i,\mathbf{S}_j) + D(\mathbf{x},y)\,, \end{aligned}$$

where

$$D(\mathbf{x},y) = \frac{\eta}{10} + 2\,M\,C\,\mathrm{I}_{\mathcal{A}_K^c}(\mathbf{x},y) \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta_0} |\mathbf{z}(\mathbf{x},\lambda,\boldsymbol{\beta})|\,.$$

Using (25) to (27) and the fact that $|\boldsymbol{\theta} - \boldsymbol{\theta}_i| < \delta_1$ and that $\|\mathbf{S} - \mathbf{S}_j\| < \delta$, it is easy to see that

$$|\mathbf{H}^{kl}_3(\mathbf{x},y,\boldsymbol{\theta},\mathbf{S}) - \mathbf{H}^{kl}_3(\mathbf{x},y,\boldsymbol{\theta}_i,\mathbf{S}_j)| \le D(\mathbf{x},y)\,.$$

Therefore, $\mathbf{H}^{kl}_{\eta,\mathrm{L}}(\mathbf{x},y)$ and $\mathbf{H}^{kl}_{\eta,\mathrm{U}}(\mathbf{x},y)$ satisfy (22).

It remains to show (23). Since

$$\begin{aligned} E\left(\mathbf{H}^{kl}_{\eta,\mathrm{U}}(\mathbf{x},y) - \mathbf{H}^{kl}_{\eta,\mathrm{L}}(\mathbf{x},y)\right) &= 2E\left(D(\mathbf{x},y)\right) \\ &= \frac{\eta}{5} + 4\,M\,C\,E\left(\mathrm{I}_{\mathcal{A}_K^c}(\mathbf{x},y) \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta_0} |\mathbf{z}(\mathbf{x},\lambda,\boldsymbol{\beta})|\right) \end{aligned}$$

using (24), (23) follows. □

The notion of $I\!P-$ measurability of a class of functions $\mathcal{F}$ can be found in van der Vaart and Wellner (1996, page 110) and it is needed in order to guarantee the measurability of a supremum over $\mathcal{F}$. In particular, if the class of functions $\mathcal{F}$ contains a countable subset $\mathcal{G}$ such that for every $f \in \mathcal{F}$ there exists a sequence g_m in $\mathcal{G}$ such that $g_m(\mathbf{x}) \to f(\mathbf{x})$ for every $\mathbf{x}$, then, $\mathcal{F}$ is $I\!P-$measurable for every probability measure $I\!P$.

Let $\mathcal{F}$ be a class of functions with enveloppe F and Q a probability measure. Remind that given two functions f and g, the bracket $[f,g]$ is the set of all functions l with $f \le l \le g$ and an $\epsilon-$bracket is a bracket $[f,g]$ with $\|g - f\|_{\mathbb{Q},2} < \epsilon$, where $\|f\|_{\mathbb{Q},2} = \left(E_{\mathbb{Q}}(f^2)\right)^{\frac{1}{2}}$. Denote $N_{[\,]}\left(\epsilon,\mathcal{F},L^2(Q)\right)$ the bracketing number, more precisely, the minimum number of $\epsilon-$brackets needed to cover $\mathcal{F}$. As above, the upper and lower bounds need not to belong to the class $\mathcal{F}$, but they should have finite $\|.\|_{\mathbb{Q},2}$ norms.

Define the bracketing integral

$$J_{[\,]}(\delta,\mathcal{F}) = \int_0^\delta \sqrt{1 + \log\left(N_{[\,]}\left(\epsilon\,\|F\|_{\mathrm{IP},2},\mathcal{F},L^2(I\!P)\right)\right)}\,d\epsilon\,.$$

The function $J_{[\,]}$ is increasing, $J_{[\,]}(0,\mathcal{F}) = 0$ and $J_{[\,]}(1,\mathcal{F}) < \infty$ and $J_{[\,]}(\delta,\mathcal{F}) \to 0$ as $\delta \to 0$ for classes of functions $\mathcal{F}$ which satisfies the bracketing entropy condition, i. e.,

$$\int_0^\infty \sqrt{\log\left(N_{[\,]}\left(\epsilon \|F\|_{\mathbb{P},2}, \mathcal{F}, L^2(I\!P)\right)\right)}\, d\epsilon < \infty\,. \tag{28}$$

In particular, classes of monotone functions and classes of functions which are Lipschitz in a parameter satisfy (28) if, for instance, the parameter set is bounded and has a finite covering number (see van der Vaart and Wellner (1996) page 164).

Maximal Inequality for bracketing numbers. *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be i.i.d. random vectors with common distribution $I\!P$. Let $\mathcal{F}$ be a $I\!P-$measurable class of functions with an enveloppe F such that $\|F\|_{\mathbb{P},2} = \left[E_{\mathbb{P}}(F^2)\right]^{\frac{1}{2}} < \infty$. For a given $\delta > 0$, set*

$$a(\delta) = \frac{\delta \|F\|_{\mathbb{P},2}}{\sqrt{1 + \log\left(N_{[\,]}\left(\delta \|F\|_{\mathbb{P},2}, \mathcal{F}, L^2(I\!P)\right)\right)}}\,.$$

Then, if $\|f\|_{\mathbb{P},2} < \delta \|F\|_{\mathbb{P},2}$, for every $f \in \mathcal{F}$, there exists a constant D_2 not depending on n, such that

$$\begin{aligned} \|\sup_{f\in\mathcal{F}} |T_n f|\,\|_{\mathbb{P},1} &\leq D_2\, J_{[\,]}(\delta,\mathcal{F})\, \|F\|_{\mathbb{P},2} + \sqrt{n}\, E_{\mathbb{P}}\left(F\, I_{\{F > \sqrt{n}\, a(\delta)\}}\right) \\ &\leq D_2\, J_{[\,]}(1,\mathcal{F})\, \|F\|_{\mathbb{P},2}\,, \end{aligned}$$

where

$$T_n f = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \left(f(\mathbf{X}_i) - E\left(f(\mathbf{X}_1)\right)\right)\right)\,.$$

Proof of Lemma 3. In order to obtain (6) we will again use the maximal inequality for bracketing numbers. Let us show that for all $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ in $\mathcal{V}$, there exists a constant C_1 such that

$$|\mathbf{H}_5\left(\mathbf{x}, y, \boldsymbol{\theta}\right) - \mathbf{H}_5\left(\mathbf{x}, y, \widetilde{\boldsymbol{\theta}}\right)| \leq C_1 \max(\gamma(\mathbf{x}), 1)\, |\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}|\,. \tag{29}$$

Using that

$$\begin{aligned} \frac{\partial}{\partial\lambda}\varphi\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right) &= -\eta\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right) h(\mathbf{x}, \boldsymbol{\beta})\,, \\ \frac{\partial}{\partial\boldsymbol{\beta}}\varphi\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right) &= -\lambda\, \eta\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right) \frac{\partial}{\partial\boldsymbol{\beta}} h(\mathbf{x}, \boldsymbol{\beta}) \end{aligned}$$

and

$$\frac{\partial}{\partial\sigma}\varphi\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right) = -\frac{1}{\sigma}\,\eta\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right)\,,$$

we obtain that

$$\begin{aligned} |\varphi\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right) - \varphi\left(r_1\left(\mathbf{x}, y, \widetilde{\boldsymbol{\theta}}\right)\right)| &\leq |\eta\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}^*\right)\right)|\, |h(\mathbf{x}, \boldsymbol{\theta}^*)|\, |\lambda - \widetilde{\lambda}| + \\ + \ |\lambda^*|\, |\eta\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}^*\right)\right)|\, |\frac{\partial}{\partial\boldsymbol{\beta}} h(\mathbf{x}, \boldsymbol{\beta}^*)|\, |\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}| &+ \frac{1}{\sigma^*}\, |\eta\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}^*\right)\right)|\, |\sigma - \widetilde{\sigma}|\,, \end{aligned} \tag{30}$$

where $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^*, \sigma^*, \lambda^*)'$ is an intermediate point between $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$.

From (30), the boundness of η implies that

$$|\varphi\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right) - \varphi\left(r_1\left(\mathbf{x}, y, \tilde{\boldsymbol{\theta}}\right)\right)| \leq M_1 \max\{|\gamma(\mathbf{x})|, 1\}\, |\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}|\,, \tag{31}$$

where $M_1 = ||\eta||_\infty(2C + 1)$.

Analogously, we obtain

$$\begin{aligned} |\Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)|\right) - \Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \tilde{\lambda}, \tilde{\boldsymbol{\beta}}\right)|\right)| \ &\leq\ |\eta_2\left(|\mathbf{z}\left(\mathbf{x}, \lambda^*, \boldsymbol{\beta}^*\right)|\right)|\,|h(\mathbf{x}, \boldsymbol{\beta}^*)|\,|\lambda - \tilde{\lambda}| + \\ &\quad + |\eta_2\left(|\mathbf{z}\left(\mathbf{x}, \lambda^*, \boldsymbol{\beta}^*\right)|\right)|\,|\lambda^*|\,|\frac{\partial}{\partial \boldsymbol{\beta}} h(\mathbf{x}, \boldsymbol{\beta}^*)|\,|\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}| \\ &\leq\ M_2\, \gamma(\mathbf{x})\, |\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}|\,, \end{aligned} \tag{32}$$

where $\eta_2(t) = t\Psi_2'(t)$ is bounded by **A5** and $M_2 = ||\eta_2||_\infty(C + 1)$.

Thus, from (31) and (32) we get

$$\begin{aligned} |\mathbf{H}_5\left(\mathbf{x}, y, \boldsymbol{\theta}\right) - \mathbf{H}_5\left(\mathbf{x}, y, \tilde{\boldsymbol{\theta}}\right)| \ &\leq\ |\varphi\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right) - \varphi\left(r_1\left(\mathbf{x}, y, \tilde{\boldsymbol{\theta}}\right)\right)|\,|\Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)|\right)| + \\ &\quad + |\varphi\left(r_1\left(\mathbf{x}, y, \tilde{\boldsymbol{\theta}}\right)\right)|\,|\Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)|\right) - \Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \tilde{\lambda}, \tilde{\boldsymbol{\beta}}\right)|\right)| \\ &\leq\ ||\Psi_2||_\infty\, |\varphi\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right) - \varphi\left(r_1\left(\mathbf{x}, y, \tilde{\boldsymbol{\theta}}\right)\right)| + \\ &\quad + ||\varphi||_\infty |\Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)|\right) - \Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \tilde{\lambda}, \tilde{\boldsymbol{\beta}}\right)|\right)| \\ &\leq\ C_1 \max\{\gamma(\mathbf{x}), 1\}|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}|\,, \end{aligned}$$

where $C_1 = M_1||\Psi_2||_\infty + M_2||\varphi||_\infty$ which entails (29).

Fix $1 \leq k \leq p$. Let $\mathcal{F} = \{f_{\boldsymbol{\theta}}(\mathbf{x}, y) = \mathbf{H}_5^k\left(\mathbf{x}, y, \boldsymbol{\theta}\right) - \mathbf{H}_5^k\left(\mathbf{x}, y, \boldsymbol{\theta}_0\right),\ \boldsymbol{\theta} \in \mathcal{V}\}$, where $\mathbf{H}_5^k$ denotes the $k-$th coordinate of $\mathbf{H}_5$. Note that from **A1,** $E\left(f_{\boldsymbol{\theta}}(\mathbf{x}, y)\right) = 0$.

A natural enveloppe for $\mathcal{F}$ is $2\,\|\varphi\|_\infty\,\|\Psi_2\|_\infty$. However, since $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \delta_0 < 1$, from (29) we can take $F(\mathbf{x}) = C_1 \max\{\gamma(\mathbf{x}), 1\}$.

Since $\mathcal{V}$ is separable and **A4, A8** and the continuity of φ entail that $\mathbf{H}_5\left(\mathbf{x}, y, \boldsymbol{\theta}\right)$ is a continuous function of $\boldsymbol{\theta}$, we have that the class $\mathcal{F}$ is $I\!P-$measurable for every $I\!P$. From (29), and using Theorem 3.7.11 of van der Vaart and Wellner (1996), we conclude that $N_{[\,]}\left(2\,\epsilon\,\|F\|_{I\!P,2}, \mathcal{F}, L^2(I\!P)\right) \leq N(\epsilon, \mathcal{V}, |\ .\ |)$ and so $\mathcal{F}$ satisfies the bracketing entropy condition given in (28).

In order to obtain (6), it will be enough to show that

$$\lim_{\delta\to 0}\lim_{n\to\infty} E\left(\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} |J_n^k(\boldsymbol{\theta})|\right) = 0\,, \tag{33}$$

since, using that $\hat{\boldsymbol{\theta}}$ is a consistent estimate of $\boldsymbol{\theta}_0$, for any $\epsilon > 0$ and $\delta > 0$, we have that for n large enough

$$\begin{aligned} P\left(|J_n^k(\hat{\boldsymbol{\theta}})| > \epsilon\right) \ &\leq\ P\left(|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| > \delta\right) + P\left(|J_n^k(\hat{\boldsymbol{\theta}})| > \epsilon\,,\, |\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < \delta\right) \\ &\leq\ \delta + P\left(\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} |J_n^k(\boldsymbol{\theta})| > \epsilon\right) \end{aligned}$$

$$\leq \delta + \frac{1}{\epsilon} E\left(\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} |J_n^k(\boldsymbol{\theta})|\right).$$

Given any $\delta > 0$, we will apply the maximal inequality for bracketing numbers to the subclass $\mathcal{F}_\delta = \{f_{\boldsymbol{\theta}}(\mathbf{x}, y),\ \boldsymbol{\theta} \in \mathcal{V} \text{ and } |\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \delta\}$. Thus, we get

$$\begin{aligned} E\left(\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} |J_n^k(\boldsymbol{\theta})|\right) &\leq D_2\, J_{[\,]}\,(\delta, \mathcal{F})\, \left[E\left(F^2(\mathbf{x})\right)\right]^{\frac{1}{2}} + \sqrt{n}\, E\left(F(\mathbf{x})\, \mathrm{I}_{\{F(\mathbf{x}) > \sqrt{n}\, a(\delta)\}}\right) \\ &\leq D_2\, J_{[\,]}\,(\delta, \mathcal{F})\, \left[E\left(F^2(\mathbf{x})\right)\right]^{\frac{1}{2}} + \frac{1}{a(\delta)} E\left(F^2(\mathbf{x})\, \mathrm{I}_{\{F(\mathbf{x}) > \sqrt{n}\, a(\delta)\}}\right). \quad (34) \end{aligned}$$

Using that $E\left(F^2(\mathbf{x})\right) < \infty$, we obtain that the second term of the inequality (34) converges to 0 as $n \to \infty$ and so, we have

$$\lim_{n\to\infty} E\left(\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} |J_n^k(\boldsymbol{\theta})|\right) \leq D_2\, J_{[\,]}\,(\delta, \mathcal{F})\, \left[E\left(F^2(\mathbf{x})\right)\right]^{\frac{1}{2}}.$$

Now, (33) follows from the fact that $J_{[\,]}\,(\delta, \mathcal{F}) \to 0$ as $\delta \to 0$. □

Proof of Lemma 4. In order to obtain (7) we will use the maximal inequality for bracketing numbers.

As above, let $\mathcal{V}$ be a neighborhood of $\boldsymbol{\theta}_0$ and $\mathcal{S}$ a neighborhood of $\mathbf{I}$ such that for any $(\boldsymbol{\theta}, \mathbf{S})$ in $\mathcal{V} \times \mathcal{S}$ we have that $\|\mathbf{S} - \mathbf{I}\| < \delta_0$ and $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \delta_0$, with δ_0 given in **A7** and $|\boldsymbol{\beta}| < C$, $|\lambda| < C$, $C^{-1} < \sigma < C$ and $\max\left(\|\mathbf{S}^{-1}\|, \|\mathbf{S}\|\right) \leq C$ for some positive constant C.

As in Lemma 2, using (5) it is easy to show that

$$\begin{aligned} |\mathbf{H}_4\,(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{W}) - \mathbf{H}_4\,(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{V})\,| &\leq c\, C^2\, \|\varphi\|_\infty \|\mathbf{W} - \mathbf{V}\| \\ &\leq c\, C^2\, \|\varphi\|_\infty \|(\boldsymbol{\theta}, \mathbf{W}) - (\boldsymbol{\theta}, \mathbf{V})\|, \quad (35) \end{aligned}$$

where $\|(\boldsymbol{\theta}, \mathbf{W})\| = \max\{|\boldsymbol{\theta}|, \|\mathbf{W}\|\}$.

As in the proof of Lemma 3 it is easy to show that for all $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ in $\mathcal{V}$, $\mathbf{W} \in \mathcal{S}$ there exists a constant C_1 such that

$$|\mathbf{H}_4\,(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{W}) - \mathbf{H}_4\left(\mathbf{x}, y, \widetilde{\boldsymbol{\theta}}, \mathbf{W}\right)| \leq C_1 \max(\gamma(\mathbf{x}), 1)\, |\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}|\,. \quad (36)$$

and so (35) and (36) entail that for some constant C

$$|\mathbf{H}_4\,(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{W}) - \mathbf{H}_4\left(\mathbf{x}, y, \widetilde{\boldsymbol{\theta}}, \mathbf{V}\right)| \leq C\, \left(1 + \max(\gamma(\mathbf{x}), 1)\right) \|(\boldsymbol{\theta}, \mathbf{W}) - (\widetilde{\boldsymbol{\theta}}, \mathbf{V})\|, \quad (37)$$

Fix $1 \leq k \leq p$. Let $\mathcal{F} = \{f_{(\boldsymbol{\theta}, \mathbf{S})}(\mathbf{x}, y) = \mathbf{H}_4^k\,(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{S}) - \mathbf{H}_4^k\,(\mathbf{x}, y, \boldsymbol{\theta}, \mathbf{I}),\ (\boldsymbol{\theta}, \mathbf{S}) \in \mathcal{V} \times \mathcal{S}\}$. Note that **A1** and the oddness of φ entail $E\left(f_{(\boldsymbol{\theta}, \mathbf{S})}(\mathbf{x}, y)\right) = 0$ and so, $E\left(J_n^k(\boldsymbol{\theta}, \mathbf{S})\right) = 0$.

A natural enveloppe for $\mathcal{F}$ is $2\,C\,\|\varphi\|_\infty\,\|\Psi_2\|_\infty$. However, since $\|(\boldsymbol{\theta},\mathbf{S})-(\boldsymbol{\theta}_0,\mathbf{I})\|<\delta_0<1$, (37) entails that we can take $F(\mathbf{x})=C(1+\max(\gamma(\mathbf{x}),1))$.

Since $\mathcal{V}\times\mathcal{S}$ is separable and **A4, A8** and the continuity of φ entail that $\mathbf{H}_4\,(\mathbf{x},y,\boldsymbol{\theta},\mathbf{S})$ is a continuous function of $(\boldsymbol{\theta},\mathbf{S})$, we have that, for every $I\!P$, the class $\mathcal{F}$ is $I\!P-$measurable. From (37) and **A12,** and using Theorem 3.7.11 of van der Vaart and Wellner (1996), we conclude that $N_{[\,]}\left(2\,\epsilon\,\|F\|_{\mathbb{P},2},\mathcal{F},L^2(I\!P)\right)\le N(\epsilon,\mathcal{V}\times\mathcal{S},|\;.\;|)$ and so $\mathcal{F}$ satisfies the bracketing entropy condition given in (28).

In order to obtain (7), it will be enough to show that

$$\lim_{\delta\to 0}\lim_{n\to\infty} E\left(\sup_{|(\boldsymbol{\theta},\mathbf{S})-(\boldsymbol{\theta}_0,\mathbf{I})|<\delta}|J_n^k(\boldsymbol{\theta},\mathbf{S})|\right)=0\,, \tag{38}$$

since, using that $(\widehat{\boldsymbol{\theta}},\widehat{\mathbf{S}})$ is a consistent estimate of $(\boldsymbol{\theta}_0,\mathbf{I})$, for any $\epsilon>0$ and $\delta>0$, we have that for n large enough

$$\begin{aligned}
P\left(|J_n^k(\widehat{\boldsymbol{\theta}},\widehat{\mathbf{S}})|>\epsilon\right) &\le \delta+P\left(\sup_{\|(\boldsymbol{\theta},\mathbf{S})-(\boldsymbol{\theta}_0,\mathbf{I})\|<\delta}|J_n^k(\boldsymbol{\theta},\mathbf{S})|>\epsilon\right)\\
&\le \delta+\frac{1}{\epsilon}E\left(\sup_{\|(\boldsymbol{\theta},\mathbf{S})-(\boldsymbol{\theta}_0,\mathbf{I})\|<\delta}|J_n^k(\boldsymbol{\theta},\mathbf{S})|\right).
\end{aligned}$$

Given any $\delta>0$, we will apply the maximal inequality for bracketing numbers to the subclass $\mathcal{F}_\delta=\{f_{\boldsymbol{\theta},\mathbf{S}}(\mathbf{x},y),\ (\boldsymbol{\theta},\mathbf{S})\in\mathcal{V}\times\mathcal{S}\text{ and }\|(\boldsymbol{\theta},\mathbf{S})-(\boldsymbol{\theta}_0,\mathbf{I})\|<\delta\}$. Thus, we get

$$\begin{aligned}
E\left(\sup_{\|(\boldsymbol{\theta},\mathbf{S})-(\boldsymbol{\theta}_0,\mathbf{I})\|<\delta}|J_n^k(\boldsymbol{\theta},\mathbf{S})|\right) &\le D_2\,J_{[\,]}\,(\delta,\mathcal{F})\,\left[E\left(F^2(\mathbf{x})\right)\right]^{\frac{1}{2}}+\\
&+\sqrt{n}\,E\left(F(\mathbf{x})\,\mathrm{I}_{\{F(\mathbf{x})>\sqrt{n}\,a(\delta)\}}\right)\\
&\le D_2\,J_{[\,]}\,(\delta,\mathcal{F})\,C_1\,,
\end{aligned} \tag{39}$$

for n large enough. And so, we have

$$\lim_{n\to\infty} E\left(\sup_{|(\boldsymbol{\theta},\mathbf{S})-(\boldsymbol{\theta}_0,\mathbf{I})|<\delta}|J_n^k(\boldsymbol{\theta},\mathbf{S})|\right)\le C_1\,D_2\,J_{[\,]}\,(\delta,\mathcal{F})\,.$$

Now, (38) follows from the fact that $J_{[\,]}\,(\delta,\mathcal{F})\to 0$ as $\delta\to 0$. □

3.2 Proof of Theorem 3

Before proving Theorem 3, we recall some definitions which can be found, for instance, in van der Vaart and Wellner (1996). Let $\mathcal{F}$ be a class of functions with enveloppe F and $\mathcal{Q}$ a probability measure. Denote $N\left(\epsilon,\mathcal{F},L^2(\mathcal{Q})\right)$ the covering number, i.e., the minimum number of balls $\mathcal{B}(\epsilon,g)=\{h\,:\,\|h-g\|_{\mathbb{Q},2}<\epsilon\}$ of radius

ϵ needed to cover $\mathcal{F}$, where $\|f\|_{\mathbb{Q},2} = \left(E_{\mathbb{Q}}(f^2)\right)^{\frac{1}{2}}$. The centers of the balls need not to belong to the class $\mathcal{F}$, but they should have finite $\|.\|_{\mathbb{Q},2}$ norms.

Define the integral

$$J(\delta,\mathcal{F}) = \sup_{\mathbb{Q}} \int_0^\delta \sqrt{1+\log\left(N\left(\epsilon\,\|F\|_{\mathbb{Q},2},\mathcal{F},L^2(\mathbb{Q})\right)\right)}\,d\epsilon\,,$$

where the supremum is taken over all discrete probability measures with $\|F\|_{\mathbb{Q},2} > 0$. The function J is increasing, $J(0,\mathcal{F}) = 0$ and $J(1,\mathcal{F}) < \infty$ and $J(\delta,\mathcal{F}) \to 0$ as $\delta \to 0$ for classes of functions $\mathcal{F}$ which satisfies the uniform– entropy condition, i. e.,

$$\int_0^\infty \sup_{\mathbb{Q}} \sqrt{\log\left(N\left(\epsilon\,\|F\|_{\mathbb{Q},2},\mathcal{F},L^2(\mathbb{Q})\right)\right)}\,d\epsilon < \infty\,. \tag{40}$$

In particular, Vapnis-Cervonenkis classes of functions satisfy (40).

Maximal Inequality for covering numbers. *Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be i.i.d. random vectors with common distribution $I\!P$. Let $\mathcal{F}$ be a $I\!P-$measurable class of functions with an enveloppe F such that $E_{I\!P}(F^2) < \infty$. Suppose that 0 is in $\mathcal{F}$. Then, there exists a constant $D_1 = D_1(q)$ not depending on n, such that*

$$\begin{aligned}\|\sup_{f\in\mathcal{F}}|T_nf|\,\|_{I\!P,q} &\leq D_1\,\|J\left(\delta_n,\mathcal{F}\right)\left[\frac{1}{n}\sum_{i=1}^n F^2\left(X_i\right)\right]^{\frac{1}{2}}\|_{I\!P,q}\\ &\leq D_1\,J(1,\mathcal{F})\,\|F\|_{I\!P,\max(2,q)}\,,\end{aligned}$$

where $\|Y\|_{I\!P,q} = \left[E_{I\!P}(Y^q)\right]^{\frac{1}{q}}$,

$$T_nf = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n\left(f(\mathbf{X}_i) - E\left(f(\mathbf{X}_1)\right)\right)\right)$$

and

$$\delta_n^2 = \frac{\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n f^2(\mathbf{X}_i)}{\frac{1}{n}\sum_{i=1}^n F^2(\mathbf{X}_i)}\,.$$

It is worthwile noting that the same function J can still be used independently of the subclass $\mathcal{F}_0$ if the same enveloppe F is used for $\mathcal{F}_0$. More precisely, if $\mathcal{F}_0 \subset \mathcal{F}$ and the enveloppe F is used for $\mathcal{F}_0$, then $J(\delta,\mathcal{F}_0) \leq J(\delta,\mathcal{F})$ and thus one still has

$$\begin{aligned}\|\sup_{f\in\mathcal{F}_0}|T_nf|\,\|_{I\!P,q} &\leq D_1\,\|J\left(\delta_{n,0},\mathcal{F}\right)\left[\frac{1}{n}\sum_{i=1}^n F^2\left(X_i\right)\right]^{\frac{1}{2}}\|_{I\!P,q}\\ &\leq D_1\,J(1,\mathcal{F})\,\|F\|_{I\!P,\max(2,q)}\,,\end{aligned}$$

with

$$\delta_{n,0}^2 = \frac{\sup_{f\in\mathcal{F}_0} \frac{1}{n}\sum_{i=1}^n f^2(\mathbf{X}_i)}{\frac{1}{n}\sum_{i=1}^n F^2(\mathbf{X}_i)}\,.$$

Lemma 5. *Under* **A1, A4, A7** *and* **A8,** *if in addition,*

a) φ *is an odd, continuous and bounded function,*

and, in addition,

b) *for each* $1 \le k \le p$, *(40) holds for the class of functions*

$$\mathcal{F}_k = \{f_{\boldsymbol{\theta}}(\mathbf{x}, y) = \mathbf{H}_5^k(\mathbf{x}, y, \boldsymbol{\theta}) - \mathbf{H}_5^k(\mathbf{x}, y, \boldsymbol{\theta}_0)\,,\ \boldsymbol{\theta} \in \mathcal{V}\}$$

with enveloppe $F = 2\,\|\varphi\|_\infty\,\|\Psi_2\|_\infty$,

we have that for any weakly consistent estimate $\widehat{\boldsymbol{\theta}}$ *of* $\boldsymbol{\theta}_0$

$$J_n\left(\widehat{\boldsymbol{\theta}}\right) \xrightarrow{p} 0\,, \tag{41}$$

where

$$J_n(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\mathbf{H}_5(\mathbf{x}_i, y_i, \boldsymbol{\theta}) - \mathbf{H}_5(\mathbf{x}_i, y_i, \boldsymbol{\theta}_0)\right)\,.$$

Proof. Fix $1 \le k \le p$. From **A1,** $E\left(f_{\boldsymbol{\theta}}(\mathbf{x}, y)\right) = 0$. Since $\mathcal{V}$ is separable, **A4, A8** and the continuity of φ entail that $\mathbf{H}_5(\mathbf{x}, y, \boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, we have that the class $\mathcal{F}_k$ is $IP-$measurable for every IP.

As in Lemma 3, in order to obtain (41), it will be enough to show that

$$\lim_{\delta\to 0}\lim_{n\to\infty} E\left(\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} |J_n^k(\boldsymbol{\theta})|\right) = 0\,, \tag{42}$$

Given any $\delta > 0$, we will apply, component–wise, the maximal inequality for covering numbers to the subclass $\mathcal{F}_\delta$ of $\mathcal{F}_k$ defined as $\mathcal{F}_\delta = \{f_{\boldsymbol{\theta}}(\mathbf{x}, y) \in \mathcal{F}_k, \boldsymbol{\theta} \in \mathcal{V}$ and $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \delta\}$. Thus, we get

$$\begin{aligned} E\left(\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} |J_n^k(\boldsymbol{\theta})|\right) &\le D_1\, E\left(J(\delta_n, \mathcal{F}_k)\left[\frac{1}{n}\sum_{i=1}^n F^2(\mathbf{x}_i, y_i)\right]\right) \\ &\le D\, E\left(J(\delta_n, \mathcal{F}_k)\right)\,, \end{aligned} \tag{43}$$

where $D = D_1\, M$, $M = 2\,\|\varphi\|_\infty\,\|\Psi_2\|_\infty$ and

$$\delta_n^2 = \frac{1}{M^2}\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta}\frac{1}{n}\sum_{i=1}^n |f_{\boldsymbol{\theta}}(\mathbf{x}_i, y_i)|^2\,.$$

From Lemma 1, we have that

$$\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} |\frac{1}{n}\sum_{i=1}^{n} |f_{\boldsymbol{\theta}}(\mathbf{x}_i, y_i)|^2 - E\left(|f_{\boldsymbol{\theta}}(\mathbf{x}, y)|^2\right)| \xrightarrow{p} 0\,.$$

Therefore, since $\lim_{\delta\to 0^+} J(\delta, \mathcal{F}_k) = 0$ it will be enough to show that

$$\lim_{\delta\to 0} \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\delta} E\left(|f_{\boldsymbol{\theta}}(\mathbf{x}, y)|^2\right) = 0\,,$$

which is equivalent to show that

$$\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0} E\left(|f_{\boldsymbol{\theta}}(\mathbf{x}, y)|^2\right) = 0 \tag{44}$$

Using **A8,** the continuity of φ and w_2, the fact that $|f_{\boldsymbol{\theta}}(\mathbf{x}, y)| \le M$ and the Dominated Convergence Theorem, we obtain (44), concluding the proof. □

Proof of Theorem 3. As in the proof of Theorem 1, denoting

$$\mathbf{L}_n\left(\mathbf{S}, \boldsymbol{\theta}, \boldsymbol{\beta}^*\right) = \frac{\sigma}{n}\sum_{i=1}^{n}\Psi_1\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}^*}{\sigma\, G\left(\mathbf{x}_i, \lambda, \boldsymbol{\beta}\right)}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})\right)\mathbf{z}(\mathbf{x}_i, \lambda, \boldsymbol{\beta})$$

and using a second order Taylor's expansion around $\boldsymbol{\beta}_{\mathrm{H}}$, we get

$$\mathbf{L}_n\left(\mathbf{S}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\beta}_0\right) = \frac{\mathbf{A}_n}{n}\left(\boldsymbol{\beta}_{\mathcal{N}} - \boldsymbol{\beta}_0\right) + \mathbf{R}_n\,, \tag{45}$$

where

$$\begin{aligned}\mathbf{R}_n &= \frac{1}{2\,\sigma_{\mathrm{H}}}\frac{1}{n}\sum_{i=1}^{n}\Psi_1''\left(\frac{y_i - \mathbf{x}_i'\widetilde{\boldsymbol{\beta}}}{\sigma_{\mathrm{H}}\, G\left(\mathbf{x}_i, \lambda_{\mathrm{H}}, \boldsymbol{\beta}_{\mathrm{H}}\right)}\right) w_2\left(\mathbf{S}\,\mathbf{z}(\mathbf{x}_i, \lambda_{\mathrm{H}}, \boldsymbol{\beta}_{\mathrm{H}})\right) \times \\ &\quad \times \mathbf{z}(\mathbf{x}_i, \lambda_{\mathrm{H}}, \boldsymbol{\beta}_{\mathrm{H}})\left\{\mathbf{z}(\mathbf{x}_i, \lambda_{\mathrm{H}}, \boldsymbol{\beta}_{\mathrm{H}})'\left(\boldsymbol{\beta}_{\mathrm{H}} - \boldsymbol{\beta}_0\right)\right\}^2\,.\end{aligned}$$

Using Lemma 2 with $\varphi(t) = \Psi_1'(t)$, we obtain that

$$\frac{\mathbf{A}_n}{n} \xrightarrow{p} \mathbf{A}\,.$$

On the other hand, we have that

$$\mathbf{L}_n\left(\mathbf{S}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\beta}_0\right) - \mathbf{L}_n\left(\mathbf{I}, \boldsymbol{\theta}_0, \boldsymbol{\beta}_0\right) = \mathbf{U}_{1n} + \mathbf{U}_{2n}\,,$$

where

$$\begin{aligned}\mathbf{U}_{1n} &= \mathbf{L}_n\left(\mathbf{S}_{\mathrm{H}}, \boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\beta}_0\right) - \mathbf{L}_n\left(\mathbf{I}, \boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\beta}_0\right) \\ \mathbf{U}_{2n} &= \mathbf{L}_n\left(\mathbf{I}, \boldsymbol{\theta}_{\mathrm{H}}, \boldsymbol{\beta}_0\right) - \mathbf{L}_n\left(\mathbf{I}, \boldsymbol{\theta}_0, \boldsymbol{\beta}_0\right)\,.\end{aligned}$$

From Lemma 5 and the consistency of $\mathbf{S}_{\mathrm{H}}$ and $\boldsymbol{\theta}_{\mathrm{H}}$, we have that

$$\sqrt{n}\mathbf{U}_{2n} \xrightarrow{p} 0\,.$$

Thus, if we had

$$\sqrt{n}\mathbf{U}_{1n} \xrightarrow{p} 0\,, \tag{46}$$

we would obtain that

$$\frac{\mathbf{A}_n}{n}\sqrt{n}\,(\boldsymbol{\beta}_{\mathcal{N}} - \boldsymbol{\beta}_0) + \sqrt{n}\mathbf{R}_n$$

has the same asymptotic distribution as $\mathbf{Z}_n = \sqrt{n}\,\mathbf{L}_n\,(\mathbf{I}, \boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$. Since, as in Theorem 1, $\mathbf{Z}_n$ is asymptotically normally distributed with zero mean and covariance matrix $\sigma_0^2\mathbf{B}$ and

$$\sqrt{n}\,\mathbf{R}_n \xrightarrow{p} 0\,,$$

the proof would be concluded.

Let us show (46). Consider the function defined by $g(\mathbf{W}) = \mathbf{L}_n\,(\mathbf{W}, \boldsymbol{\theta}_{\rm H}, \boldsymbol{\beta}_0)$. By using a second order Taylor's expansion of $g(\mathbf{S}_{\rm H})$ around $\mathbf{I}$ we get that $g(\mathbf{S}_{\rm H}) = g(\mathbf{I}) + \mathbf{T}_{1n} + \mathbf{T}_{2n}$, where $\mathbf{T}_{1n}$ and $\mathbf{T}_{2n}$ are the first and second order term respectively. Therefore, we have that

$$\begin{aligned}\mathbf{T}_{1n} &= \frac{2}{n}\sum_{i=1}^{n}\Psi_1\left(r_1\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}_{\rm H}\right)\right)\,\varphi_1\left(|\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})|^2\right)\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H}) \times \\ &\times \;\{\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})'\,(\mathbf{S}_{\rm H} - \mathbf{I})\,\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})\}\end{aligned}$$

and

$$\begin{aligned}|\mathbf{T}_{2n}| &\leq 2\,p^4\,C_p^2\,\|\mathbf{S}_{\rm H} - \mathbf{I}\|^2\frac{1}{n}\sum_{i=1}^{n}|\Psi_1\left(r_1\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}_{\rm H}\right)\right)|\,|\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})| \times \\ &\quad \times\Big\{|\varphi_2\left(|\tilde{\mathbf{S}}\,\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})|^2\right)|\;\|\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})\,\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})'\tilde{\mathbf{S}}\|^2 \\ &\quad + \varphi_1\left(|\tilde{\mathbf{S}}\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})|^2\right)\|\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})'\|\Big\}\;,\end{aligned}$$

where

$$\begin{aligned}\varphi_1(t) &= \frac{t^{-1}}{2}\left(\Psi_2'(t^{\frac{1}{2}}) - t^{-\frac{1}{2}}\,\Psi_2(t^{\frac{1}{2}})\right) \\ \varphi_2(t) &= \frac{t^{-2}}{2}\left(\frac{1}{2}t^{\frac{1}{2}}\Psi_2''(t^{\frac{1}{2}}) - \frac{3}{2}\Psi_2'(t^{\frac{1}{2}}) + \frac{3}{2}t^{-\frac{1}{2}}\,\Psi_2(t^{\frac{1}{2}})\right)\;.\end{aligned}$$

Note that $|\mathbf{T}_{1n}| \leq 2p^2|\mathbf{S}_{\rm H} - \mathbf{I}|\,|\mathbf{V}_n(\boldsymbol{\theta}_{\rm H})|$ where

$$\begin{aligned}\mathbf{V}_n(\boldsymbol{\theta}_{\rm H}) &= \frac{1}{n}\sum_{i=1}^{n}\Psi_1\left(r_1\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}_{\rm H}\right)\right)\,\varphi_1\left(|\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})|^2\right) \times \\ &\quad \times(\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})') \otimes \mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H}) \\ &= \mathbf{V}_{1n}(\boldsymbol{\theta}_{\rm H}) + \mathbf{V}_{2n}(\boldsymbol{\theta}_{\rm H})\,,\end{aligned}$$

with

$$\begin{aligned}\mathbf{V}_{1n}(\boldsymbol{\theta}_{\rm H}) &= \frac{1}{2n}\sum_{i=1}^{n}\Psi_1\left(r_1\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}_{\rm H}\right)\right)\eta_2\left(|\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})|\right)\left(\frac{\mathbf{x}_i}{|\mathbf{x}_i|}\frac{\mathbf{x}_i'}{|\mathbf{x}_i|}\right) \otimes \frac{\mathbf{x}_i}{|\mathbf{x}_i|} \\ \mathbf{V}_{2n}(\boldsymbol{\theta}_{\rm H}) &= \frac{1}{2n}\sum_{i=1}^{n}\Psi_1\left(r_1\left(\mathbf{x}_i, y_i, \boldsymbol{\theta}_{\rm H}\right)\right)\Psi_2\left(|\mathbf{z}(\mathbf{x}_i, \lambda_{\rm H}, \boldsymbol{\beta}_{\rm H})|\right)\left(\frac{\mathbf{x}_i}{|\mathbf{x}_i|}\frac{\mathbf{x}_i'}{|\mathbf{x}_i|}\right) \otimes \frac{\mathbf{x}_i}{|\mathbf{x}_i|}\;.\end{aligned}$$

Using that the covering number for the family $\mathcal{H}_{k\ell} = \left\{ f_{\boldsymbol{\theta}}(\mathbf{x}, y) \left[\left(\frac{\mathbf{x}}{|\mathbf{x}|} \frac{\mathbf{x}'}{|\mathbf{x}|} \right) \otimes \frac{\mathbf{x}}{|\mathbf{x}|} \right]^{k\ell} \right\}$ can be bounded by the covering number of the family $\mathcal{F}$, similar arguments to those used in Lemma 5 allow us to conclude that $\sqrt{n}\left(\mathbf{V}_{2n}(\boldsymbol{\theta}_{\mathrm{H}}) - \mathbf{V}_{2n}(\boldsymbol{\theta}_0)\right) \xrightarrow{p} 0$.

Since assumption (c) implies that the family $\mathcal{G}_{k\ell} = \left\{ g_{\boldsymbol{\theta}}(\mathbf{x}, y) \left[\left(\frac{\mathbf{x}}{|\mathbf{x}|} \frac{\mathbf{x}'}{|\mathbf{x}|} \right) \otimes \frac{\mathbf{x}}{|\mathbf{x}|} \right]^{k\ell} \right\}$ has finite entropy and η_2 is a continuous bounded function, as in Lemma 5 we can get that $\sqrt{n}\left(\mathbf{V}_{1n}(\boldsymbol{\theta}_{\mathrm{H}}) - \mathbf{V}_{1n}(\boldsymbol{\theta}_0)\right) \xrightarrow{p} 0$. Therefore,

$$\sqrt{n}\left(\mathbf{V}_{n}(\boldsymbol{\theta}_{\mathrm{H}}) - \mathbf{V}_{n}(\boldsymbol{\theta}_0)\right) \xrightarrow{p} 0\,, \tag{47}$$

which entails that $\sqrt{n}\mathbf{V}_n(\boldsymbol{\theta}_{\mathrm{H}})$ is bounded in probability. On the other hand, denoting $c_\varphi = \frac{1}{2}\left(\frac{1}{2}\|\eta_3\|_\infty + \frac{3}{2}\|\eta_2\|_\infty + \frac{3}{2}\|\Psi_2\|_\infty\right)$, assumption **A6** and **A4** imply that

$$|\varphi_2(|z|^2)|z|^5| \;\leq\; \frac{1}{2}\left(\frac{1}{2}|\eta_3(|z|)| + \frac{3}{2}|\eta_2(|z|)| + \frac{3}{2}|\Psi_2(|z|)|\right) \leq c_\varphi \tag{48}$$

Denote $C_1 = c_\varphi + \|\eta_2\|_\infty + \|\Psi_2\|_\infty$. Since

$$\begin{aligned} \sqrt{n}|\mathbf{U}_{1n}| &= \sqrt{n}|\left(g(\mathbf{S}_{\mathrm{H}}) - g(\mathbf{I})\right)| \\ &\leq 2p^2|\mathbf{S}_{\mathrm{H}} - \mathbf{I}|\,|\sqrt{n}\mathbf{V}_n(\boldsymbol{\theta}_{\mathrm{H}})| + 6p^4 C_p^2 C^7 \|\Psi_1\|_\infty C_1 \sqrt{n}\|\mathbf{S}_{\mathrm{H}} - \mathbf{I}\|^2 \end{aligned}$$

from (47), (48) and the assumption (a) made we obtain (46) □

PROOF OF PROPOSITION 3. Note that since

$$f_{\boldsymbol{\theta}}(\mathbf{x}, y) = \Psi_1\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right)\, \Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)|\right) - \Psi_1\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}_0\right)\right)\, \Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \lambda_0, \boldsymbol{\beta}_0\right)|\right)\,,$$

it will be enough to show that $\widetilde{\mathcal{F}} = \{f_{\boldsymbol{\theta}}(\mathbf{x}, y) = \Psi_1\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right)\, \Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)|\right)$, $\boldsymbol{\theta} \in \mathcal{V}\}$ has finite uniform–entropy.

Since for any class of functions $\mathcal{F}$ such that $\mathcal{F} = \{f = f_1 + f_2 \,:\, f_i \in \mathcal{F}_i\,,\ i = 1, 2\}$, we have that

$$N\left(\epsilon, \mathcal{F}, L^2(\mathbb{Q})\right) \leq N\left(\frac{\epsilon}{2}, \mathcal{F}_1, L^2(\mathbb{Q})\right)\ \boldsymbol{.}\ N\left(\frac{\epsilon}{2}, \mathcal{F}_2, L^2(\mathbb{Q})\right)\,,$$

we only need to obtain the result when Ψ_1 and Ψ_2 are bounded increasing functions.

On the other hand, $\widetilde{\mathcal{F}} \subset \widetilde{\mathcal{F}}_1\ \boldsymbol{.}\ \widetilde{\mathcal{F}}_2$ where

$$\begin{aligned} \widetilde{\mathcal{F}}_1 &= \{f_{1,\boldsymbol{\theta}}(\mathbf{x}, y) = \Psi_1\left(r_1\left(\mathbf{x}, y, \boldsymbol{\theta}\right)\right)\,,\ \boldsymbol{\theta} \in \mathcal{V}\} \\ \widetilde{\mathcal{F}}_2 &= \{f_{2,\boldsymbol{\theta}}(\mathbf{x}, y) = \Psi_2\left(|\mathbf{z}\left(\mathbf{x}, \lambda, \boldsymbol{\beta}\right)|\right)\,,\ \boldsymbol{\theta} \in \mathcal{V}\} \end{aligned}$$

and so $N\left(\epsilon, \widetilde{\mathcal{F}}, L^2(\mathbb{Q})\right) \leq N\left(\epsilon, \widetilde{\mathcal{F}}_1\ \boldsymbol{.}\ \widetilde{\mathcal{F}}_2, L^2(\mathbb{Q})\right)$. According to Corollary 2.6.12 and Lemmas 2.6.13 and 2.6.20 of van der Vaart and Wellner (1996), the desired conclusion can be derived from the fact that the classes of functions $\widetilde{\mathcal{F}}_1$ and $\widetilde{\mathcal{F}}_2$ are bounded VC–major classes of functions. Since $\widetilde{\mathcal{F}}_1$ and $\widetilde{\mathcal{F}}_2$ can be bounded by $\|\Psi_1\|_\infty$ and $\|\Psi_2\|_\infty$, respectively, and any finite dimensional vector space of measurable functions is a VC–major class, the result follows easily by applying Lemma 2.6.19 of van der Vaart and Wellner (1996). □

4 References

Bianco, A. M., Boente, G. and di Rienzo J. (2000). Some results of GM–based estimators in heteroscedastic regression models. *J. Statist. Inf. and Planning* **89**, 215-242.

Bickel, P. (1975). One–Step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70**, 428-434.

Bickel, P. (1978). Using residuals robustly I. Test for heteroscedasticity, nonlinearity. *Ann. Statist.* **6**, 266-291.

Box, G. and Hill, W. (1974). Correcting inhomogeinity of variance with power transformation weighting. *Technometrics* **16**, 385-389.

Carroll, R.J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *Ann. Statist.* **10**, 429-441.

Davies, L. (1992). An efficient Fréchet differentiable high breakdown multivariate location and dispersion estimator. *J. Mult. Anal.* **40**, 311-327.

Donoho, D. (1982). *Breakdown Properties of Multivariate Location Estimators,* Ph.D. Qualifying paper, Department of Statistics, Harvard University.

Giltinan, D.M. , Carroll, R.J. and Ruppert, D. (1986). Some new estimation methods for weighted regression when there are possible outliers. *Technometrics* **28**, 219-230.

He, X. and Portnoy, S. (1992). Reweighted LS estimators converge at the same rate as the initial estimator. *Ann. Statist.* **20**, 2161-2167.

Kim, J. and Pollard, D. (1990). Cube root asymptotics .*Ann. Statist.* **18**, 191-219.

Lopuhaä, H. P. (1992). Highly efficient estimators of multivariate location with breakdown point. *Ann. Statist.* **20**, 398-413.

Maronna, R. and Yohai, V. (1991). Recent results on bias-robust regression estimates. In *Directions in Robust Statistics and Diagnostics*, Part I, W. Stahel and S. Weisberg (eds.). New York: Springer-Verlag, p. 221-232.

Pollard, D. (1984). *Convergence of Stochastic Processes.* New York: Springer.

Rousseeuw, P.J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79**, 871-880.

Rousseeuw, P. and van Zomeren, B. (1990). Unmasking multivariate outliers and leverage points. *J. Amer. Statist. Assoc.* **85**, 633-639.

Rousseeuw, P. and Yohai, V. (1984). Robust regression by means of S-estimates. In *Robust and nonlinear time series*, J. Franke, W. Härdle and R. Martin (eds.). Lecture Notes in Statistics N° 26, 256-272.

Stahel, W. (1981). *Robust estimation: Infinitesimal optimality and covariance ma-*

trix estimators, Thesis (in German), ETH, Zurich, 1981.

Simpson, D., Ruppert, D. and Carroll, R.J. (1992). On one-step GM-estimates and stability of inferences in linear regressions. *J. Amer. Statist. Assoc.* **87**, 439-450.

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

Yohai, V. (1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.* **15**, 642-656.

Yohai, V. and Zamar, R. (1988). High breakdown estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.* **83**, 406-413.