Score estimation in the single index model: supplement

Fadoua Balabdaoui

Centre de Recherche en Mathématiques de la Décision, Université Paris-Dauphine, Paris, France
e-mail: fadoua.balabdaoui@gmail.com

Piet Groeneboom

Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands.
e-mail: P.Groeneboom@tudelft.nl

Kim Hendrickx

Hasselt University, I-BioStat, Agoralaan, B3590 Diepenbeek, Belgium.
e-mail: kim.hendrickx@uhasselt.be

We give the proofs of the remaining results given in the main manuscript that were not considered in Section 10 of Balabdaoui, Groeneboom and Hendrickx (2017) together with additional technical lemmas needed for proving our main results.

1. Supplement A: Asymptotic normality of the efficient score estimator

In this section we prove (iii) of Theorem 5.1 on the asymptotic normality of the efficient score estimator $\tilde{\boldsymbol{\alpha}}_n$. The proofs of existence and consistency of $\tilde{\boldsymbol{\alpha}}_n$, given in (i) and (ii) of Theorem 5.1 follow the same lines as the corresponding proofs for the simple score estimator $\hat{\boldsymbol{\alpha}}_n$ given in Sections 10.2.1 and 10.2.1 and are omitted.

Proof of asymptotic normality: Let τ_i denote the sequence of jump points of the monotone LSE $\hat\psi_{n\boldsymbol{\alpha}}$. We introduce the piecewise constant function $\bar\rho_{n,\boldsymbol{\beta}}$ defined for $u \in [\tau_i, \tau_{i+1})$ as

$$\bar\rho_{n,\boldsymbol{\beta}}(u) = \begin{cases} \mathbb{E}[\boldsymbol{X}|\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{X} = \tau_i]\psi'_{\boldsymbol{\alpha}}(\tau_i) & \text{if } \psi_{\boldsymbol{\alpha}}(u) > \hat\psi_{n\boldsymbol{\alpha}}(\tau_i) \text{ for all } u \in (\tau_i, \tau_{i+1}), \\ \mathbb{E}[\boldsymbol{X}|\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{X} = s]\psi'_{\boldsymbol{\alpha}}(s) & \text{if } \psi_{\boldsymbol{\alpha}}(s) = \hat\psi_{n\boldsymbol{\alpha}}(s) \text{ for some } s \in (\tau_i, \tau_{i+1}), \\ \mathbb{E}[\boldsymbol{X}|\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{X} = \tau_{i+1}]\psi'_{\boldsymbol{\alpha}}(\tau_{i+1}) & \text{if } \psi_{\boldsymbol{\alpha}}(u) < \hat\psi_{n\boldsymbol{\alpha}}(\tau_i) \text{ for all } u \in (\tau_i, \tau_{i+1}). \end{cases}$$

We can write,

$$\begin{aligned}
&\xi_{nh}(\tilde{\boldsymbol{\beta}}_n) \\
&= \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \left\{\boldsymbol{x}\tilde\psi'_{nh,\boldsymbol{\alpha}}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\left\{y - \hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\} d\mathbb{P}_n(\boldsymbol{x}, y) \\
&+ \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \left\{\mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \bar\rho_{n,\tilde{\boldsymbol{\beta}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\left\{y - \hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\} d\mathbb{P}_n(\boldsymbol{x}, y) \\
&= J + JJ, \qquad (1.1)
\end{aligned}$$

using,

$$\int \bar\rho_{n,\tilde{\boldsymbol{\beta}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\left\{y - \hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\} d\mathbb{P}_n(\boldsymbol{x}, y) = \boldsymbol{0}.$$

imsart-generic ver. 2014/10/16 file: BGH_single_index_models_Supplement1.tex date: June 15, 2017

The term JJ can be written as

$$\begin{aligned}
JJ &= \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \left\{\mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \bar{\rho}_{n,\tilde{\boldsymbol{\beta}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\\
&\qquad\qquad \cdot\left\{y - \hat{\psi}_{n\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\,d(\mathbb{P}_n - P_0)(\boldsymbol{x},y)\\
&\quad + \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \left\{\mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \bar{\rho}_{n,\tilde{\boldsymbol{\beta}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\\
&\qquad\qquad \cdot\left\{y - \psi_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\,dP_0(\boldsymbol{x},y)\\
&\quad + \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \left\{\mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \bar{\rho}_{n,\tilde{\boldsymbol{\beta}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\\
&\qquad\qquad \cdot\left\{\psi_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \hat{\psi}_{n\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\,dP_0(\boldsymbol{x},y)\\
&= JJ_a + JJ_b + JJ_c, \qquad (1.2)
\end{aligned}$$

We first note that by Assumption A10, the functions $u \mapsto \psi'_{\boldsymbol{\alpha}}(u) := \psi'_{\mathbb{S}(\boldsymbol{\beta})}(u)$ are uniformly bounded and have a total variation that is uniformly bounded for all $\boldsymbol{\beta} \in \mathcal{C}$. This also implies, using Lemma 3.4, that the functions $u \mapsto \mathbb{E}\left(X_i|\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{X} = u\right)\psi'_{\boldsymbol{\alpha}}(u)$ have a bounded variation for all $\boldsymbol{\beta} \in \mathcal{C}$. Using the same arguments as those for term II_a defined in (10.26) in the proof of Theorem 4.1, it easily follows that,

$$JJ_a = o_p(n^{-1/2}).$$

We next consider the term JJ_b. By Lemma 3.6 we know that $\psi'_{\boldsymbol{\alpha}}$ stays away from zero for all $\mathbb{S}(\boldsymbol{\beta})$ in a neighborhood of $\mathbb{S}(\boldsymbol{\beta}_0)$. Using the same techniques as in Groeneboom and Jongbloed (2014), we can find a constant $K > 0$ such that for all $i = 1, \ldots, d$ and $u \in \mathcal{I}_{\boldsymbol{\alpha}}$,

$$\left|\mathbb{E}\left(X_i|\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{X} = u\right)\psi'_{\boldsymbol{\alpha}}(u) - \bar{\rho}_{ni,\boldsymbol{\beta}}(u)\right| \le K\left|\psi_{\boldsymbol{\alpha}}(u) - \hat{\psi}_{n\boldsymbol{\alpha}}(u)\right| \qquad (1.3)$$

where $\bar{\rho}_{ni,\boldsymbol{\beta}}$ denotes the ith component of $\rho_{n,\boldsymbol{\beta}}$. This implies that the difference $\mathbb{E}\left(X_i|\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{X} = u\right)\psi'_{\boldsymbol{\alpha}}(u) - \bar{\rho}_{ni,\boldsymbol{\beta}}(u)$ converges to zero for all $u \in \mathcal{I}_{\boldsymbol{\alpha}}$. Using Lemma 3.1 and a Taylor expansion of $\boldsymbol{\beta} \mapsto \psi_{\boldsymbol{\alpha}}\left(\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{x}\right)$ we get,

$$\begin{aligned}
\psi_{\boldsymbol{\alpha}}\left(\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{x}\right) &= \psi_0\left(\mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{x}\right) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T\left[\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)^T\left(\boldsymbol{x} - \mathbb{E}(\boldsymbol{X}|\mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{X} = \mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{x})\right)\psi'_0\left(\mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{x}\right)\right]\\
&\quad + o(\boldsymbol{\beta} - \boldsymbol{\beta}_0), \qquad (1.4)
\end{aligned}$$

such that

$$\begin{aligned}
JJ_b &= \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \left\{\mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \bar{\rho}_{n,\tilde{\boldsymbol{\beta}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\\
&\qquad \cdot\left\{\psi_0\left(\mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{x}\right) - \psi_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\,dP_0(\boldsymbol{x},y) = o_p\left(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right)
\end{aligned}$$

For the therm JJ_c, we get by an application of the Cauchy-Schwarz inequality together with the uniform boundedness of $\boldsymbol{J}_{\mathbb{S}}$, Proposition 3.2 and (1.3) that,

$$\begin{aligned}
JJ_c &\le \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T\left(\int \left\{\mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \bar{\rho}_{n,\tilde{\boldsymbol{\beta}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}^2 dP_0(\boldsymbol{x},y)\right)^{1/2}\\
&\qquad \cdot\int\left(\left\{\psi_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \hat{\psi}_{n\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}^2 dP_0(\boldsymbol{x},y)\right)^{1/2}\\
&\lesssim \int\left\{\psi_{\tilde{\boldsymbol{\alpha}}_n}\left(\tilde{\boldsymbol{\alpha}}_n^T\boldsymbol{x}\right) - \hat{\psi}_{n\tilde{\boldsymbol{\alpha}}_n}\left(\tilde{\boldsymbol{\alpha}}_n^T\boldsymbol{x}\right)\right\}^2 dG(\boldsymbol{x}) = O_p\left((\log n)^2 n^{-2/3}\right) = o_p(n^{-1/2}).
\end{aligned}$$

We conclude that (1.1) can be written as

$$
\begin{aligned}
&\xi_{nh}(\tilde{\boldsymbol\beta}_n)\\
&= \boldsymbol J_{\mathbb S}(\tilde{\boldsymbol\beta}_n)^T \int \left\{\boldsymbol x \tilde\psi'_{nh,\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right) - \mathbb E\left(\boldsymbol X|\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\psi'_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}\\
&\qquad\qquad\cdot\left\{y-\hat\psi_{n\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}d\mathbb P_n(\boldsymbol x,y)\\
&\qquad + o_p\left(n^{-1/2}+(\tilde{\boldsymbol\beta}_n-\boldsymbol\beta_0)\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \boldsymbol J_{\mathbb S}(\tilde{\boldsymbol\beta}_n)^T \int \left\{\boldsymbol x \tilde\psi'_{nh,\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right) - \mathbb E\left(\boldsymbol X|\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\psi'_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}\\
&\qquad\qquad\cdot\left\{y-\psi_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}d\mathbb P_n(\boldsymbol x,y)\\
&\quad + \boldsymbol J_{\mathbb S}(\tilde{\boldsymbol\beta}_n)^T \int \left\{\boldsymbol x \tilde\psi'_{nh,\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right) - \mathbb E\left(\boldsymbol X|\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\psi'_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}\\
&\qquad\qquad\cdot\left\{\psi_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)-\hat\psi_{n\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}d(\mathbb P_n-P_0)(\boldsymbol x,y)\\
&\quad + \boldsymbol J_{\mathbb S}(\tilde{\boldsymbol\beta}_n)^T \int \left\{\boldsymbol x \tilde\psi'_{nh,\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right) - \mathbb E\left(\boldsymbol X|\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\psi'_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}\\
&\qquad\qquad\cdot\left\{\psi_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)-\hat\psi_{n\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}dP_0(\boldsymbol x,y)\\
&\quad + o_p\left(n^{-1/2}+(\tilde{\boldsymbol\beta}_n-\boldsymbol\beta_0)\right)\\
&= J_a+J_b+J_c+o_p\left(n^{-1/2}+(\tilde{\boldsymbol\beta}_n-\boldsymbol\beta_0)\right). \qquad (1.5)
\end{aligned}
$$

We first consider the term J_b. By Assumption A10, Lemma 3.4 and Lemma 3.7 we get that the functions $u\mapsto \mathbb E\left(\boldsymbol X|\mathbb S(\boldsymbol\beta)^T\boldsymbol x=u\right)\psi'_{\tilde{\boldsymbol\alpha}_n}(u)$ and $u\mapsto\tilde\psi'_{nh,\tilde{\boldsymbol\alpha}_n}(u)$ have a uniformly bounded total variation for all $\boldsymbol\beta\in\mathcal C$. Using similar arguments as for the term I_b defined in (10.28) we get for $A>0$ and $\nu>0$ that

$$P(|J_b|\ge An^{-1/2})\le\nu,$$

for n large enough and we conclude that $J_b=o_p(n^{-1/2})$. For the term J_c we get,

$$
\begin{aligned}
J_c &= \boldsymbol J_{\mathbb S}(\tilde{\boldsymbol\beta}_n)^T \int \left\{\boldsymbol x - \mathbb E\left(\boldsymbol X|\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}\tilde\psi'_{nh,\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\\
&\qquad\qquad\cdot\left\{\psi_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)-\hat\psi_{n\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}dP_0(\boldsymbol x,y)\\
&\quad + \boldsymbol J_{\mathbb S}(\tilde{\boldsymbol\beta}_n)^T \int \left\{\tilde\psi'_{nh,\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)-\psi'_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}\mathbb E\left(\boldsymbol X|\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\\
&\qquad\qquad\cdot\left\{\psi_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)-\hat\psi_{n\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}dP_0(\boldsymbol x,y)\\
&= \boldsymbol J_{\mathbb S}(\tilde{\boldsymbol\beta}_n)^T \int \left\{\tilde\psi'_{nh,\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)-\psi'_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}\mathbb E\left(\boldsymbol X|\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\\
&\qquad\qquad\cdot\left\{\psi_{\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)-\hat\psi_{n\tilde{\boldsymbol\alpha}_n}\left(\mathbb S(\tilde{\boldsymbol\beta}_n)^T\boldsymbol x\right)\right\}dP_0(\boldsymbol x,y)
\end{aligned}
$$

Furthermore, let $H_{\boldsymbol\beta}$ be the distribution function of the random variable $\mathbb S(\boldsymbol\beta)^T\boldsymbol X$ and let $\mathbb E(\boldsymbol X|u)$ denote the conditional expectation of $\boldsymbol{X}$ given $\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{X} = u$, then

$$\begin{aligned}
&\int \left\{\tilde\psi'_{nh,\tilde{\boldsymbol{\alpha}}_n}(u) - \psi'_{\tilde{\boldsymbol{\alpha}}_n}(u)\right\} \mathbb{E}\left(\boldsymbol{X}|u\right) \left\{\psi_{\tilde{\boldsymbol{\alpha}}_n}(u) - \hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}(u)\right\} dH_{\tilde{\boldsymbol{\beta}}_n}(u)\\
&= \int \left\{\frac{1}{h}\int K\left(\{u-v\}/h\right) d\hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}(v) - \psi'_{\tilde{\boldsymbol{\alpha}}_n}(u)\right\} \mathbb{E}\left(\boldsymbol{X}|u\right) \left\{\psi_{\tilde{\boldsymbol{\alpha}}_n}(u) - \hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}(u)\right\} dH_{\tilde{\boldsymbol{\beta}}_n}(u)\\
&= \int \left(\frac{1}{h^2}\int K'\left(\{u-v\}/h\right) \left\{\hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}(v) - \psi_{\tilde{\boldsymbol{\alpha}}_n}(v)\right\} dv\right) \mathbb{E}\left(\boldsymbol{X}|u\right) \left\{\psi_{\tilde{\boldsymbol{\alpha}}_n}(u) - \hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}(u)\right\} dH_{\tilde{\boldsymbol{\beta}}_n}(u)\\
&\quad + \int \left(\frac{1}{h}\int K\left(\{u-v\}/h\right) \psi'_{\tilde{\boldsymbol{\alpha}}_n}(v)\, dv - \psi'_{\tilde{\boldsymbol{\alpha}}_n}(u)\right) \mathbb{E}\left(\boldsymbol{X}|u\right) \left\{\psi_{\tilde{\boldsymbol{\alpha}}_n}(u) - \hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}(u)\right\} dH_{\tilde{\boldsymbol{\beta}}_n}(u)
\end{aligned}$$

The last term on the right hand side is $O_p\left(n^{-2/7-1/3}\right) = o_p\left(n^{-1/2}\right)$. This follows by an application of the Cauchy-Schwarz inequality since

$$\left\{\int \left(\frac{1}{h}\int K\left(\{u-v\}/h\right) \psi'_{\tilde{\boldsymbol{\alpha}}_n}(v)\, dv - \psi'_{\tilde{\boldsymbol{\alpha}}_n}(u)\right)^2 dH_{\tilde{\boldsymbol{\beta}}_n}(u)\right\}^{1/2} = O_p\left(n^{-2/7}\right),$$

and

$$\left\{\int \left(\psi_{\tilde{\boldsymbol{\alpha}}_n}(u) - \hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}(u)\right)^2 dH_{\tilde{\boldsymbol{\beta}}_n}(u)\right\}^{1/2} = O_p\left(n^{-1/3}\right),$$

The first term on the right hand side is $O_p\left(n^{1/7-2/3}\right) = o_p\left(n^{-1/2}\right)$ using that for small h

$$\begin{aligned}
&\int \left(\frac{1}{h^2}\int K'\left(\{u-v\}/h\right) \left\{\hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}(v) - \psi_{\tilde{\boldsymbol{\alpha}}_n}(v)\right\} dv\right) \mathbb{E}\left(\boldsymbol{X}|u\right) \left\{\psi_{\tilde{\boldsymbol{\alpha}}_n}(u) - \hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}(u)\right\} dH_{\tilde{\boldsymbol{\beta}}_n}(u)\\
&\lesssim \frac{1}{h}\int \left(\psi_{\tilde{\boldsymbol{\alpha}}_n}(u) - \hat\psi_{n\tilde{\boldsymbol{\alpha}}_n}(u)\right)^2 dH_{\tilde{\boldsymbol{\beta}}_n}(u),
\end{aligned}$$

We conclude that (1.5) can be written as,

$$\begin{aligned}
\xi_{nh}(\tilde{\boldsymbol{\beta}}_n) &= \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \left\{\boldsymbol{x}\tilde\psi'_{nh,\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\\
&\qquad\qquad \cdot \left\{y - \psi_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\} d\mathbb{P}_n(\boldsymbol{x},y) + o_p\left(n^{-1/2} + (\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)\right)\\
&= \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \boldsymbol{x}\left\{\tilde\psi'_{nh,\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\left\{y - \psi_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\} d(\mathbb{P}_n - P_0)(\boldsymbol{x},y)\\
&\quad + \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \boldsymbol{x}\left\{\tilde\psi'_{nh,\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right) - \psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\left\{y - \psi_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\} dP_0(\boldsymbol{x},y)\\
&\quad + \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \left\{\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\left\{y - \psi_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\} d(\mathbb{P}_n - P_0)(\boldsymbol{x},y)\\
&\quad + \boldsymbol{J}_{\mathbb{S}}(\tilde{\boldsymbol{\beta}}_n)^T \int \left\{\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\}\psi'_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\left\{y - \psi_{\tilde{\boldsymbol{\alpha}}_n}\left(\mathbb{S}(\tilde{\boldsymbol{\beta}}_n)^T\boldsymbol{x}\right)\right\} dP_0(\boldsymbol{x},y)\\
&\quad + o_p\left(n^{-1/2} + (\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)\right)\\
&= JJJ_a + JJJ_b + JJJ_c + JJJ_d + o_p\left(n^{-1/2} + (\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)\right) \qquad (1.6)
\end{aligned}$$

We consider JJJ_a first and note that by Assumption A10 and Lemma 3.7, the functions $\psi'_{\boldsymbol{\alpha}}$ and $\tilde\psi'_{nh,\boldsymbol{\alpha}}$ have a uniformly bounded total variation. By an application of Lemma 3.5 we can write the difference $\tilde\psi'_{nh,\boldsymbol{\alpha}} - \psi'_{\boldsymbol{\alpha}}$ as the difference of two monotone functions, say $f_1, f_2 \in \mathcal{M}_{RC_1}$ for some constant $C_1 > 0$. This implies that the class of functions

$$\mathcal{F}_1 = \left\{f(\boldsymbol{x},y) := \{\tilde\psi'_{nh,\boldsymbol{\alpha}}\left(\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{x}\right) - \psi'_{\boldsymbol{\alpha}}\left(\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{x}\right)\}\{y - \psi_{\boldsymbol{\alpha}}\left(\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{x}\right)\}, (\boldsymbol{x},y,\boldsymbol{\beta}) \in \mathcal{X}\times\mathbb{R}\times\mathcal{C})\right\}$$

is contained in the class $\mathcal{H}_{RC_1 v}$ where $v \asymp h^{-1}\log n n^{-1/3}$ (See the proof of Lemma 3.7). By Lemma 2.4 and the fact that the order bracketing entropy of a class does not get altered after multiplication with the fixed and bounded function $\boldsymbol{x} \mapsto x_i$ we get that the class of functions involved with the term JJJ_a, say $\mathcal{F}_a$, satisfies

$$H_B\big(\epsilon, \mathcal{F}_a, \|\cdot\|_{B,P_0}\big) \lesssim \frac{1}{\epsilon} \qquad \text{and} \qquad \|f\|_{B,P_0} \lesssim v$$

Using again an application of Markov's inequality, together with Lemma 3.4.3 of van der Vaart and Wellner (1996) we conclude that for $A > 0$

$$P(|JJJ_a| > An^{-1/2}) \lesssim v^{1/2} = h^{-1/2}(\log n)^{1/2} n^{-1/6}$$

which can be made arbitrarily small for n large enough and $h \asymp n^{-1/7}$. We conclude that

$$JJJ_a = o_p(n^{-1/2})$$

Using similar arguments as for the term JJ_b defined in (1.2) we also get,

$$JJJ_b = o_p\left(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right).$$

The result of Theorem 5.1 follows by noting that, using the same techniques as for the term I_a in (10.31), we get

$$\begin{aligned} JJJ_c = &(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0))^T \int \left\{\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{x}\right)\right\} \psi_0'\left(\mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{x}\right)\left\{y - \psi_0\left(\mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{x}\right)\right\} d(\mathbb{P}_n - P_0)(\boldsymbol{x}, y) \\ &+ o_p(n^{-1/2}) + o_p(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \end{aligned}$$

and that by a Taylor expansion of $\boldsymbol{\beta} \mapsto \psi_{\boldsymbol{\alpha}}\left(\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{x}\right)$ we get,

$$\begin{aligned} JJJ_d = -\Bigg\{&(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0))^T\left(\int \left(\psi_0'\left(\mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{x}\right)\right)^2 \cdot \left\{\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{x}\right)\right\}\left\{\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{X}|\mathbb{S}(\boldsymbol{\beta}_0)^T\boldsymbol{x}\right)\right\}^T dP_0(\boldsymbol{x}, y)\right) \\ &\times J_{\mathbb{S}}(\boldsymbol{\beta}_0)\Bigg\}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \end{aligned}$$

The rest of the proof follows the same line as the proof of asymptotic normality of the simple score estimator defined in Theorem 4.1 and is omitted.

2. Supplement B: Entropy results

Lemma 2.1. *Fix $\epsilon > 0$, and consider $\mathcal{F}_1$ a class of functions defined on $\mathcal{X} \times \mathbb{R}$ bounded by some constant $A > 0$ and equipped by the L_2 norm $\|\cdot\|_{P_0}$ with respect to P_0. Also, let $\mathcal{F}_2$ be another class of continuous functions defined on a bounded set $\mathcal{C} \subset \mathbb{R}^{d-1}$ such that $\mathcal{F}_2$ is equipped by the supremum norm $\|\cdot\|_\infty$, and bounded by some constant $B > 0$. Moreover assume that $H_B(\epsilon, \mathcal{F}_1, \|\cdot\|_{P_0}) < \infty$ and $H_B(\epsilon, \mathcal{F}_2, \|\cdot\|_\infty) < \infty$. Consider*

$$\mathcal{F} = \mathcal{F}_1\mathcal{F}_2 = \Big\{f(\boldsymbol{x}) = f_{\boldsymbol{\beta}}(\boldsymbol{x}, y) = f_1(\boldsymbol{x}, y) f_2(\boldsymbol{\beta}) : (\boldsymbol{x}, y, \boldsymbol{\beta}) \in \mathcal{X} \times \mathbb{R} \times \mathcal{C}\Big\}.$$

Then there exists some constant $B > 0$ such that

$$H_B(\epsilon, \mathcal{F}, \|\cdot\|_{P_0}) \le H_B(B\epsilon, \mathcal{F}_1, \|\cdot\|_{P_0}) + H_B(B\epsilon, \mathcal{F}_2, \|\cdot\|_\infty).$$

Proof. Let $f = f_1 f_2 \in \mathcal{F}$ for some pair $(f_1, f_2) \in \mathcal{F}_1 \times \mathcal{F}_2$. For $\epsilon > 0$ consider the (f_1^L, f_1^U) and (f_2^L, f_2^U) ϵ-brackets with respect to $\|\cdot\|_{P_0}$ for f_1 and f_2. Note that since $\mathcal{F}_1$ and $\mathcal{F}_2$ are bounded by $M = \max(A, B)$ we can always assume that $-M \le f_i^L \le f_i^U \le M$ for $i \in \{1, 2\}$. As we deal with a product of two functions, construction of a bracket for f requires considering different sign cases for a given pair $(\boldsymbol{x}, \boldsymbol{\beta})$:

1. $0 \le f_1^L(\boldsymbol{x})$ and $0 \le f_2^L(\boldsymbol{\beta})$,
2. $0 \le f_1^L(\boldsymbol{x})$, $f_2^L(\boldsymbol{\beta}) < 0$ and $f_2^U(\boldsymbol{\beta}) \ge 0$,
3. $f_1^L(\boldsymbol{x}) \le 0$, $f_1^U(\boldsymbol{x}) \ge 0$ and $0 \le f_2^L(\boldsymbol{\beta})$,
4. $f^U(\boldsymbol{x}) \le 0$, $f^L(\boldsymbol{\beta}) \ge 0$,
5. $f^L(\boldsymbol{x}) \ge 0$, $f^U(\boldsymbol{\beta}) \ge 0$,
6. $f_1^L(\boldsymbol{x}) \le 0$, $f^U(\boldsymbol{x}) \ge 0$, $f_2^L(\boldsymbol{\beta}) \le 0$ and $f_2^U(\boldsymbol{\beta}) \ge 0$,
7. $f_1^L(\boldsymbol{x}) \le 0$, $f^U(\boldsymbol{x}) \ge 0$ and $f_2^U(\boldsymbol{\beta}) \le 0$,
8. $f^U(\boldsymbol{x}) \le 0$, $f^L(\boldsymbol{\beta}) \le 0$ and $f^U(\boldsymbol{\beta}) \ge 0$,
9. $f_1^U(\boldsymbol{x}) \le 0$ and $f_2^U(\boldsymbol{\beta}) \le 0$.

We can assume without loss of generality that each one these cases occur for all $x \in \mathcal{X}$ and $\boldsymbol{\beta} \in \mathcal{C}$ since the general case can be handled by considering the 9 different subsets of $\mathcal{X} \times \mathcal{C}$. In the proof, we will restrict ourselves to making the calculations explicit for cases 1 and 2 since the remaining cases can be handled very similarly. Then, $f_1^L f_2^L \le f \le f_1^U f_2^U$. Also, we have that

$$f_1^U f_2^U - f_1^L f_2^L = \Big(f_1^U - f_1^L\Big) f_2^U + f_1^L \Big(f_2^U - f_2^L\Big).$$

Recall that $M = \max(A, B)$. Then, it follows that

$$\begin{aligned} \int_{\mathcal{X}} \Big(f_1^U f_2^U - f_1^L f_2^L\Big)^2 dP_0 \quad &\le \quad 2M \left(\int_{\mathcal{X}} \Big(f_1^U - f_1^L\Big)^2 dP_0(\boldsymbol{x}) + \|f_2^U - f_2^L\|_\infty^2 \right) \\ &\le \quad 4M\epsilon^2. \end{aligned}$$

This in turn implies that $H_B(\epsilon, \mathcal{F}, \|\cdot\|_{P_0}) \le H_B(C\epsilon, \mathcal{F}_1, \|\cdot\|_{P_0}) + H_B(C\epsilon, \mathcal{F}_2, \|\cdot\|_\infty)$ with $C = (2M)^{-1}$. Now we consider case 2. It is not difficult to show that

$$f_2^L f_1^U \le f \le f_1^U f_2^U.$$

Hence,

$$\int_{\mathcal{X}} \Big(f_1^U f_2^U - f_2^L f_1^U\Big)^2 dP_0 \le A^2 \|f_2^U - f_2^L\|_\infty^2 \le A^2 \epsilon^2$$

and we can take $C = A^{-1}$. □

Lemma 2.2. *Let $\mathcal{F}$ be a class of functions satisfying $H_B(\epsilon, \mathcal{F}, \|\cdot\|_{P_0}) < \infty$ for every $\epsilon \in (0, \epsilon_0)$ for some given $\epsilon_0 > 0$. If $\mathcal{D} = \mathcal{F} - \mathcal{F}$ the class of all differences of elements of $\mathcal{F}$, then*

$$H_B(\epsilon, \mathcal{D}, \|\cdot\|_{P_0}) \le 2H_B(\epsilon/2, \mathcal{F}, \|\cdot\|_{P_0}).$$

Proof. Let $\epsilon \in (0, \epsilon_0)$ and $d = f_2 - f_1$ denote an element in $\mathcal{D}$ with $(f_1, f_2) \in \mathcal{F}^2$. Also, let (f_1^L, f_1^U) and (f_2^L, f_2^U) ϵ-brackets for f_1 and f_2. Define $d^L = f_2^L - f_1^U$ and $d^U = f_2^U - f_1^L$. It is clear that (d^L, d^U) is a bracket for d. Furthermore, we have that

$$\begin{aligned} &\int_{\mathcal{X}} \left(d^U(\boldsymbol{x}, y) - d^L(\boldsymbol{x}, y)\right)^2 dP_0(\boldsymbol{x}, y) \\ &\quad \le 2 \left\{ \int_{\mathcal{X}} \left(f_1^U(\boldsymbol{x}, y) - f_1^L(\boldsymbol{x}, y)\right)^2 dP_0(\boldsymbol{x}, y) + \int_{\mathcal{X}} \left(f_2^U(\boldsymbol{x}, y) - f_2^L(\boldsymbol{x}, y)\right)^2 dP_0(\boldsymbol{x}, y) \right\} \\ &\quad \le 4\epsilon^2. \end{aligned}$$

Thus,

$$\exp\Big(H_B(2\epsilon, \mathcal{D}, \|\cdot\|_{P_0})\Big) \le \exp\Big(H_B(\epsilon, \mathcal{F}, \|\cdot\|_{P_0})\Big)^2$$

which is equivalent to the statement of the lemma. □

Consider the class $\mathcal{G}_{RK}$ defined as

$$\mathcal{G}_{RK} = \Big\{ g : g(\boldsymbol{x}) = g_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \psi(\boldsymbol{\alpha}^T \boldsymbol{x}), \boldsymbol{x} \in \mathcal{X}, (\psi, \boldsymbol{\alpha}) \in \mathcal{M}_{RK} \times \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0) \Big\}. \tag{2.1}$$

where $\mathcal{M}_{RK}$ is the same class defined in (10.7).

Lemma 2.3. *There exists $A > 0$ such that for $\epsilon \in (0, K)$ we have that*

$$H_B(\epsilon, \mathcal{G}_{RK}, \| \cdot \|_{P_0}) \le \frac{AK}{\epsilon}.$$

Proof. See the proof of Lemma 4.9 in Balabdaoui, Durot and Jankowski (2016). □

Lemma 2.4. *For some constants $C > 0$ and $\delta > 0$ consider the class of functions*

$$\mathcal{D}_{RC\delta} = \Big\{ d : d = f_{1,\boldsymbol{\alpha}} - f_{2,\boldsymbol{\alpha}},\ (f_{1,\boldsymbol{\alpha}}, f_{2,\boldsymbol{\alpha}}) \in \mathcal{G}^2_{RC}, \|d(\alpha^T \cdot)\|_{P_0} \le \delta \text{ for all } \boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0) \Big\}.$$

Let $\mathcal{H}_{RCv}$ be a class of functions such that

$$\mathcal{H}_{RCv} = \Big\{ h : h(\boldsymbol{x}, y) = y d_1(\boldsymbol{\alpha}^T \boldsymbol{x}) - d_2(\boldsymbol{\alpha}^T \boldsymbol{x}),\ (\boldsymbol{x}, y, \boldsymbol{\alpha}) \in \mathcal{X} \times \mathbb{R} \times \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0), (d_1, d_2) \in \mathcal{D}^2_{RCv} \Big\} \tag{2.2}$$

where $C \ge K_0 \vee 1$. Then, for all $\epsilon \in (0, C)$ we have that

$$H_B\Big(\epsilon, \widetilde{\mathcal{H}}, \| \cdot \|_{B,P_0}\Big) \le H_B\Big(\epsilon \tilde{C}^{-1}, \mathcal{H}_{RCv}, \| \cdot \|_{P_0}\Big) \le \frac{\tilde{C} C}{\epsilon} \asymp \frac{1}{\epsilon}$$

$$\|\tilde{h}\|_{B,P_0} \lesssim \tilde{D}^{-1} v$$

where

$$A' = A\Big(2(a_0 M_0 + 1)\Big)^{-1/2}, \quad \tilde{D} = 16 M_0 C \quad \text{and} \quad \tilde{C} = \frac{1}{8M_0} \left(2a_0 + \frac{1}{2} e^{(2M_0)^{-1}} \right)^{1/2} \frac{1}{C} \tag{2.3}$$

with a_0, M_0 the same constants from Assumption A6, A the same constant in Lemma 2.3, and $\widetilde{\mathcal{H}} \stackrel{def}{=} \mathcal{H}_{RCv} \tilde{D}^{-1}$.

Proof. Consider (d_1^L, d_1^U) and (d_2^L, d_2^U) to be ϵ-brackets of the functions $\boldsymbol{x} \mapsto d_1(\boldsymbol{\alpha}^T \boldsymbol{x})$ and $\boldsymbol{x} \mapsto d_2(\boldsymbol{\alpha}^T \boldsymbol{x})$ and some $\boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)$. It follows from Lemma 4.9 of Balabdaoui, Durot and Jankowski (2016) and Lemma 2.2 that there exists some constant $A > 0$ such that

$$H_B\Big(\epsilon, \mathcal{D}_{RC}, \| \cdot \|_{P_0}\Big) \le \frac{AC}{\epsilon}.$$

Define now

$$h^L(\boldsymbol{x}, y) = \begin{cases} y d_1^L(\boldsymbol{x}) - d_2^U(\boldsymbol{x}), & \text{if } y \ge 0 \\ y d_1^U(\boldsymbol{x}) - d_2^U(\boldsymbol{x}), & \text{if } y < 0 \end{cases}$$

and

$$h^U(\boldsymbol{x}, y) = \begin{cases} y d_1^U(\boldsymbol{x}) - d_2^L(\boldsymbol{x}), & \text{if } y \ge 0 \\ y d_1^L(\boldsymbol{x}) - d_2^L(\boldsymbol{x}), & \text{if } y < 0. \end{cases}$$

Note first that (h^L, h^U) is a bracket for $h(\boldsymbol{x}, y) = yd_1(\boldsymbol{\alpha}^T\boldsymbol{x}) - d_2(\boldsymbol{\alpha}^T\boldsymbol{x})$. Next we compute the size of this bracket with respect to $\|\cdot\|_{P_0}$. We have that

$$\begin{aligned}\int_{\mathcal{X}\times\mathbb{R}}\Big(h^U(\boldsymbol{x},y)-h^L(\boldsymbol{x},y)\Big)^2 dP_0(\boldsymbol{x},y) &\le 2\left\{\int_{\mathcal{X}\times\mathbb{R}} y^2\Big(d_1^U(\boldsymbol{x})-d_1^L(\boldsymbol{x})\Big)^2 dP_0(\boldsymbol{x},y)+\int_{\mathcal{X}}\Big(d_2^U(\boldsymbol{x})-d_2^L\boldsymbol{x})\Big)^2 dG(\boldsymbol{x})\right\}\\ &= 2\Big\{2a_0\int_{\mathcal{X}}\Big(d_1^U(\boldsymbol{x})-d_1^L(\boldsymbol{x})\Big)^2 dG(\boldsymbol{x})+\int_{\mathcal{X}}\Big(d_2^U(\boldsymbol{x})-d_2^L\boldsymbol{x})\Big)^2 dG(\boldsymbol{x})\Big\}\\ &\le 2\big(2a_0+1\big)\epsilon^2\end{aligned}$$

where a_0 is the same constant of Assumption A6. It follows that

$$H_B\Big(\epsilon,\mathcal{H},\|\cdot\|_{P_0}\Big)\le\frac{\tilde{A}C}{\epsilon}$$

with $\tilde{A} = A\Big(2(2a_0+1)\Big)^{-1/2}$ and A is the same constant of Lemma 2.3. Let now $D > 0$ be some constant to be determined later. For a given $h \in \mathcal{H}_{RK^2v}$, we consider $\tilde{h} = D^{-1}h$ which admits $[D^{-1}h^L, D^{-1}h^U]$ as bracket. We will compute the size of this bracket with respect to the Bernstein norm. By definition of the latter we can write for any function h such that h^k is P_0 integrable that

$$\|h\|^2_{B,P_0} = 2\sum_{k=2}^{\infty}\frac{1}{k!}|h|^k dP_0.$$

Thus, using this and convexity of the function $x \mapsto |x|^k$ for all $k \ge 2$ it follows that

$$\begin{aligned}\|D^{-1}h^U-D^{-1}h^L\|^2_{B,P_0} &= 2\sum_{k=2}^{\infty}\frac{1}{k!D^k}\int_{\mathcal{X}\times\mathbb{R}}\Big|y\big(d_1^U(\boldsymbol{x})-d_1^L(\boldsymbol{x})\big)+d_2^U(\boldsymbol{x})-d_2^L(\boldsymbol{x})\Big|^k dP_0(\boldsymbol{x},y)\\ &\le 2\sum_{k=2}^{\infty}\frac{2^{k-1}}{k!D^k}\left\{\int_{\mathcal{X}\times\mathbb{R}}|y|^k\big(d_1^U(\boldsymbol{x})-d_1^L(\boldsymbol{x})\big)^k dP_0(\boldsymbol{x},y)+\int_{\mathcal{X}\times\mathbb{R}}\big(d_2^U(\boldsymbol{x})-d_2^L(\boldsymbol{x})\big)^k dP_0(\boldsymbol{x},y)\right\}.\end{aligned}$$

Using Assumption A7 and the fact that $|d_i^L| \le K^2$ and $|d_i^U| \le 2C$ for $i \in \{1, 2\}$ (an assumption that one can always make in constructing brackets for a bounded class) we can write

$$\begin{aligned}\|D^{-1}h^U-D^{-1}h^L\|^2_{B,P_0} &\le \sum_{k=2}^{\infty}\frac{1}{k!}\left(\frac{2}{D}\right)^k\left\{a_0M_0^{k-2}k!(4C)^{k-2}\int_{\mathcal{X}}\big(d_1^U(\boldsymbol{x})-d_1^L(\boldsymbol{x})\big)^2 dP_0(\boldsymbol{x},y)\right.\\ &\qquad\left.+(4C)^{k-2}\int_{\mathcal{X}}\big(d_1^U(\boldsymbol{x})-d_1^L(\boldsymbol{x})\big)^2 dP_0(\boldsymbol{x},y)\right\}\\ &= \left(\frac{2}{D}\right)^2\left\{a_0\sum_{k=2}^{\infty}\left(\frac{8M_0C}{D}\right)^{k-2}+\sum_{k=2}^{\infty}\frac{1}{k!}\left(\frac{8C}{D}\right)^{k-2}\right\}\epsilon^2\\ &\le \left(\frac{2}{D}\right)^2\left\{a_0\sum_{k=0}^{\infty}\left(\frac{8M_0C}{D}\right)^{k}+\frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{8C}{D}\right)^{k}\right\}\epsilon^2,\quad\text{using } k!\ge 2(k-2)!.\end{aligned}$$

Taking $D = \tilde{D} = 16M_0C$ yields

$$\|\tilde{D}^{-1}h^U-\tilde{D}^{-1}h^L\|^2_{B,P_0}\le\left(\frac{2}{\tilde{D}}\right)^2\left(2a_0+\frac{1}{2}e^{(2M_0)^{-1}}\right)\ \epsilon^2$$

which in turn implies that

$$\|\tilde{D}^{-1}h^U-\tilde{D}^{-1}h^L\|_{B,P_0}\le\frac{1}{8M_0}\left(2a_0+\frac{1}{2}e^{(2M_0)^{-1}}\right)^{1/2}\ \frac{1}{C}\ \epsilon.$$

This completes the proof of the first claim about the entropy bound of the class $\widetilde{\mathcal{H}}$ with $\tilde{D}$ defined as above. Now for a given element $\tilde{h} \in \widetilde{\mathcal{H}}$ we calculate

$$\begin{aligned}
\|\tilde{h}\|^2_{B,P_0} &= 2\sum_{k=2}^{\infty} \frac{1}{\tilde{D}^k}\frac{1}{k!}\int_{\mathcal{X}\times\mathbb{R}} \left|y d_1(\boldsymbol{\alpha}^T\boldsymbol{x}) - d_2(\boldsymbol{\alpha}^T\boldsymbol{x})\right|^k dP_0(\boldsymbol{x},y) \\
&\le 2\sum_{k=2}^{\infty} \frac{2^{k-1}}{\tilde{D}^k}\frac{1}{k!}\int_{\mathcal{X}\times\mathbb{R}} \left\{\left|y\right|^k \left|d_1(\boldsymbol{\alpha}^T\boldsymbol{x})\right|^k + \left|d_2(\boldsymbol{\alpha}^T\boldsymbol{x})\right|^k dP_0(\boldsymbol{x},y)\right\} \\
&\le 2\sum_{k=2}^{\infty} \frac{2^{k-1}}{\tilde{D}^k}\frac{1}{k!}(2C)^{k-2}\left\{a_0 M_0^{k-2} k! \int_{\mathcal{X}\times\mathbb{R}} \left|d_1(\boldsymbol{\alpha}^T\boldsymbol{x})\right|^2 dP_0(\boldsymbol{x},y) + \int_{\mathcal{X}\times\mathbb{R}} \left|d_2(\boldsymbol{\alpha}^T\boldsymbol{x})\right|^2 dP_0(\boldsymbol{x},y)\right\} \\
&\le \left(\frac{2}{\tilde{D}}\right)^2 \left\{a_0 \sum_{k=2}^{\infty}\left(\frac{8M_0C}{\tilde{D}}\right)^{k-2} + \sum_{k=2}^{\infty}\frac{1}{k!}\left(\frac{8C}{\tilde{D}}\right)^{k-2}\right\} v^2 \quad \text{using the definition of the class} \\
&\le \left(\frac{2}{\tilde{D}}\right)^2 \left(2a_0 + \frac{1}{2}e^{(2M_0)^{-1}}\right) v^2 \quad \text{using arguments as above,}
\end{aligned}$$

implying that

$$\|\tilde{h}\|_{B,P_0} \le 2\left(2a_0 + \frac{1}{2}e^{(2M_0)^{-1}}\right)^{1/2} \frac{1}{\tilde{D}}\, v \lesssim \tilde{D}^{-1} v$$

as claimed. □

Recall that $\mathcal{X}$ is the support of the covariates $X_i,\ i = 1,\dots,n$. Let us denote by $\mathcal{X}_j,\ j = 1,\dots,d$ the set of the j-th projection of $\boldsymbol{x} \in \mathcal{X}$. Also, consider some function s that $d-1$ times continuously differentiable on a convex and bounded set $\mathcal{C} \in \mathbb{R}^{d-1}$ with a nonempty interior such that there exists $M > 0$ satisfying

$$\max_{k.\le d-1}\sup_{\boldsymbol{\beta}\in\mathcal{C}} |D^k s(\boldsymbol{\beta})| \le M \tag{2.4}$$

where $k = (k_1,\dots,k_d)$ with k_j an integer $\in \{0,\dots,d-1\}$, $k. = \sum_{i=1}^{d-1} k_i$ and

$$D^k \equiv \frac{\partial^{k\cdot} s(\boldsymbol{\beta})}{\partial\beta_{k_1}\dots\partial\beta_{k_d}}.$$

Consider now the class

$$\mathcal{Q}_{jRC} = \left\{q_j(\boldsymbol{x},y) = s(\boldsymbol{\beta})x_j(y - \psi(\boldsymbol{\alpha}^T\boldsymbol{x})),\ (\boldsymbol{\alpha},\boldsymbol{\beta},\psi) \in \mathcal{B}(\alpha_0,\delta_0)\times\mathcal{C}\times\mathcal{M}_{RC} \text{ and } (x_j,y) \in \mathcal{X}_j \times \in \mathbb{R}\right\}. \tag{2.5}$$

Define

$$\widetilde{\mathcal{Q}}_{jRC} = \left\{\tilde{q}_j : \tilde{q}_j = q_j\tilde{D}^{-1}, q_j \in \mathcal{Q}_{RC}\right\},$$

where $\tilde{D} > 0$ is some appropriate constant.

Lemma 2.5. *Let $\epsilon \in (0,1)$ and $C \ge \max(1, 2M_0, Me^{-1/4}2^{-1/2}R^{-1}, 2a_0^{1/2}e^{-1/2})$. Then, there exist some constant $B_1 > 0$ and B_2 depending on a_0, M_0, and R such that*

$$H_B\left(\epsilon, \widetilde{\mathcal{Q}}_{jRC}, \|\cdot\|_{B,P_0}\right) \le \frac{B_1 C}{\epsilon}, \quad \|\tilde{q}_j\|_{B,P_0} \le B_2,$$

if $\tilde{D} = 8MRC$ where a_0 and M_0 are the same positive constants in Assumption A6, and M is from (2.4).

Proof. Fix $j \in \{1, \ldots, d\}$. The proof of this lemma uses similar techniques as in showing Lemma 2.4. Let (g^L, g^U) be ϵ-brackets for the class $\mathcal{G}_{RC}$. Using the result of Lemma 2.3 we know that there are at most $N \le \exp(AC/\epsilon)$ such brackets covering $\mathcal{G}_{RC}$ for some constant $A > 0$. Define

$$\Big(k^L(\boldsymbol{x}, y), k^U(\boldsymbol{x}, y)\Big) = \begin{cases} \Big(x_j(y - g^L(\boldsymbol{x})), x_j(y - g^U(\boldsymbol{x}))\Big), & \text{if } x_j \ge 0 \\ \Big(x_j(y - g^U(\boldsymbol{x})), x_j(y - g^L(\boldsymbol{x}))\Big), & \text{if } x_j < 0\ . \end{cases} \tag{2.6}$$

Then, the collection of all possible pairs (q^L, q^U) form brackets for the class of functions

$$\mathcal{K}_{jRC} \equiv \Big\{k_j(\boldsymbol{x}, y) = x_j(y - \psi(\boldsymbol{\alpha}^T\boldsymbol{x})),\ (\boldsymbol{\alpha}, \psi) \in \mathcal{B}(\alpha_0, \delta_0) \times \mathcal{M}_{RC} \text{ and } (x_j, \boldsymbol{x}, y) \in \mathcal{X}_j \times \mathcal{X} \times \mathbb{R}\Big\}.$$

Furthermore we have that

$$\begin{aligned} \|k^U - k^L\|_{P_0}^2 &= \int_{\mathcal{X}} x_j^2 \big(g^U(\boldsymbol{x}) - g^L(\boldsymbol{x})\big)^2 dG(\boldsymbol{x}) \\ &\le \|\boldsymbol{x}\|_2^2 \int_{\mathcal{X}} \big(g^U(\boldsymbol{x}) - g^L(\boldsymbol{x})\big)^2 dG(\boldsymbol{x}) \le R^2\epsilon^2. \end{aligned}$$

This implies that

$$H_B\Big(\epsilon, \mathcal{K}_{jRC}, \|\cdot\|_{P_0}\Big) \le \frac{ARC}{\epsilon}$$

where A is the same constant of Lemma 2.3. Furthermore, the assumption in (2.4) implies that the function s belongs to $C_{\tilde{M}}^{d-1}$ as defined in Section 2.7 in van der Vaart and Wellner (1996), with $\tilde{M} = 2M$. Using now Theorem 2.7.1 of van der Vaart and Wellner (1996) it follows that there exists some constant $B > 0$ such that

$$\log N\Big(\epsilon, C_{\tilde{M}}^{d-1}, \|\cdot\|_\infty\Big) \le B\left(\frac{1}{\epsilon}\right)^{d/(d-1)} \le \frac{B}{\epsilon}.$$

This also implies that

$$H_B\Big(\epsilon, C_{\tilde{M}}^{d-1}, \|\cdot\|_\infty\Big) = \log N\Big(\epsilon/2, C_{\tilde{M}}^{d-1}, \|\cdot\|_\infty\Big) \le \frac{2B}{\epsilon}.$$

Indeed, for an arbitrary $s \in C_{\tilde{M}}^{d-1}$ there exists $s_i, i \in \{1, \ldots, N\}$, with $N = N\big(\epsilon/2, C_{\tilde{M}}^{d-1}, \|\cdot\|_\infty\big)$, such that $\|s - s_i\|_\infty \le \epsilon/2$. The claim follows from noting that $(s_i - \epsilon/2, s_i + \epsilon/2)$ is an ϵ-bracket for $C_{\tilde{M}}^{d-1}$ with respect to $\|\cdot\|_\infty$. Using Lemma 2.1 it follows that there exists some constant $L > 0$ such that

$$\begin{aligned} H_B\Big(\epsilon, \mathcal{Q}_{jRC}, \|\cdot\|_{P_0}\Big) &\le L\left(\frac{1}{\epsilon} + \frac{C}{\epsilon}\right) \\ &\le \frac{2LC}{\epsilon} \end{aligned} \tag{2.7}$$

using that $C \ge 1$, $d - 1 \ge 1$ and $\epsilon \in (0, 1)$. Consider now a constant $D > 0$, and (q^L, q^U) and ϵ-bracket. From the proof of Lemma 2.1 we know that we can restrict attention to the case for example to case 1 assumed to occur for all $(\boldsymbol{x}, \boldsymbol{\beta}) \in \mathcal{X} \times \mathcal{C}$. In such that we have $q^L = s^L k^L$ and $q^U = s^U k^U$ where (s^L, s^U) is an ϵ-bracket for $C_{\tilde{M}}^1$equipped with $\|\cdot\|_\infty$, where the expression of (k^L, k^U) is given in (2.6). We can now write

$$\begin{aligned} \|D^{-1}q^U - D^{-1}q^L\|_{B,P_0}^2 &= 2\sum_{k=2}^\infty \frac{1}{k!}\frac{1}{D^k}\int_{\mathcal{X}\times\mathbb{R}} \big|s^U k^U - s^L k^L\big|^k dP_0 \\ &\le \sum_{k=2}^\infty \frac{2^k}{k!}\frac{1}{D^k}\int_{\mathcal{X}\times\mathbb{R}} \Big\{\big|s^U\big(k^U - k^L\big)\big|^k + \big|k^L\big(s^U - s^L\big)\big|^k\Big\} dP_0 \end{aligned}$$

with

$$\int_{\mathcal{X}\times\mathbb{R}} \left|s^U\left(k^U-k^L\right)\right|^k dP_0 \le M^k(2RC)^{k-2}\int_{\mathcal{X}\times\mathbb{R}} \left(k^U-k^L\right)^2 dP_0 = M^2(2MCR)^{k-2}\epsilon^2$$

where we used the fact that $|s| \le M$ by assumption of the lemma (implying that we can constructs brackets (s^L, s^U) satisfying the same property), and $k^U - k^U = x_j(g^U - g^L) \le 2RC$. Also, if we assume without loss of generality that $x_j \ge 0$ is satisfied for all $\boldsymbol{x} \in \mathcal{X}$ we have that

$$\begin{aligned}\int_{\mathcal{X}\times\mathbb{R}} \left|k^L\left(s^U-s^L\right)\right|^k dP_0 &\le (2M)^{k-2}\int_{\mathcal{X}\times\mathbb{R}} |x_j(y-g^L(\boldsymbol{x}))|^k dP_0(\boldsymbol{x},y)\times\epsilon^2\\ &\le (2M)^{k-2}R^k2^{k-1}\int_{\mathcal{X}\times\mathbb{R}}\left\{|y|^k+\left|g^L(\boldsymbol{x})\right|^k\right\}dP_0(\boldsymbol{x},y)\times\epsilon^2\\ &\le (2M)^{k-2}R^k2^{k-1}\left(a_0M_0^{k-2}k!+C^k\right)\epsilon^2.\end{aligned}$$

Putting these inequalities together and after some algebra we get

$$\begin{aligned}&\|D^{-1}q^U-D^{-1}q^L\|^2_{B,P_0}\\ &\le \left(\frac{1}{2}\left(\frac{2M}{D}\right)^2 e^{4MCR/D}+\left(\frac{2RC}{D}\right)^2 e^{8MCR/D}+2a_0\left(\frac{2R}{D}\right)^2\frac{1}{1-8MM_0R/D})\right)\epsilon^2.\end{aligned}$$

Now let us choose $\tilde{D} = D \ge \max(16MM_0R, 8MRC)$. In particular, we can assume that C is large enough so that $\max(16MM_0R, 8MRC) = 8MRC = \tilde{D}$ (or equivalently $C \ge 2M_0$). Then, $4MCR/\tilde{D} = 1/2$, $8MCR/\tilde{D} = 1/4$, and $8MM_0R/\tilde{D} = M_0/C \le 1/2$. Therefore,

$$\begin{aligned}\|\tilde{D}^{-1}q^U-\tilde{D}^{-1}q^L\|^2_{B,P_0} &\le \left(\frac{1}{2}\left(\frac{2M}{\tilde{D}}\right)^2 e^{1/2}+\left(\frac{2RC}{\tilde{D}}\right)^2 e+4a_0\left(\frac{2R}{\tilde{D}}\right)^2\right)\ \epsilon^2\\ &= \left(2M^2e^{1/2}+4R^2e\ C^2+16a_0R^2\right)\ \frac{1}{\tilde{D}^2}\ \epsilon^2\\ &\le \frac{\tilde{A}C^2}{\tilde{D}^2}\ \epsilon^2 = \frac{\tilde{A}}{64M^2R^2}\epsilon^2\end{aligned}$$

if C is large enough, where $\tilde{A} = 2M^2e^{1/2} + 4R^2e + 16a_0R^2$. It follows that we can find some constant $\tilde{L} > 0$ such that

$$\|\tilde{D}^{-1}q^U-\tilde{D}^{-1}q^L\|_{B,P_0} \le \tilde{L}\epsilon.$$

This in turn implies that

$$\begin{aligned}H_B\left(\tilde{L}\epsilon, \widetilde{Q}_{jRC}, \|\cdot\|_{B,P_0}\right) &\le H_B\left(\epsilon, Q_{jRC}, \|\cdot\|_{P_0}\right)\\ &\lesssim \frac{2MC}{\epsilon}\end{aligned}$$

using (2.7). Hence, we can find a constant $B_1 > 0$ such that

$$H_B\left(\epsilon, \widetilde{Q}_{jRC}, \|\cdot\|_{B,P_0}\right) \le \frac{B_1C}{\epsilon}.$$

Now we turn to computing an upper bound for $\|\tilde{q}_j\|_{B,P_0}$. We have

$$\begin{aligned}
\|\tilde{q}_j\|^2_{B,P_0} &= 2\sum_{k=2}^{\infty}\frac{1}{k!}D^{-k}\int_{\mathcal{X}\times\mathbb{R}}|s(\boldsymbol{\beta})|^k\big|x_j\big(y-\psi(\boldsymbol{\alpha}^T\boldsymbol{x})\big)\big|^k dP_0(\boldsymbol{x},y)\\
&\le \sum_{k=2}^{\infty}\frac{1}{k!}2^kD^{-k}(RM)^k\int_{\mathcal{X}\times\mathbb{R}}\Big\{|y|^k+|\psi(\boldsymbol{\alpha}^T\boldsymbol{x})|^k\Big\}dP_0(\boldsymbol{x},y)\\
&\le \sum_{k=2}^{\infty}\frac{1}{k!}2^kD^{-k}(RM)^k\Big(a_0M_0^{k-2}k!+C^k\Big)\\
&\le a_0\left(\frac{2MR}{D}\right)^2\sum_{k=2}^{\infty}\left(\frac{2RMM_0}{D}\right)^{k-2}+\frac{1}{2}\left(\frac{2MRC}{D}\right)^2\sum_{k=2}^{\infty}\frac{1}{(k-2)!}\left(\frac{2RMC}{D}\right)^{k-2}\\
&\le a_0\left(\frac{1}{2C}\right)^2\sum_{k=0}^{\infty}\left(\frac{1}{4}\right)^k+\frac{1}{2}\left(\frac{1}{2}\right)^2\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{1}{2}\right)^k\\
&\le a_0\left(\frac{1}{2C}\right)^2\frac{3}{4}+\frac{1}{2}\left(\frac{1}{2}\right)^2e^{1/2}\qquad\text{if } D=4MRC \text{ and } C\ge\max(1,2M_0).
\end{aligned}$$

The proof of the lemma is complete if we write $B_2=(3a_0/16+e^{1/2}/8)^{1/2}$.

□

In the next lemma, we consider a given a class of functions $\mathcal{F}$ which admits a bounded bracketing entropy with respect to $\|\cdot\|_{P_0}$ for $\epsilon\in(0,1]$. Suppose also that there exists $D>0$ such that $\|f\|_\infty\le D$ and $\delta>0$ such that $\|f\|_{P_0}\le\delta$ for all $f\in\mathcal{F}$. Then we can derive an upper bound for the bracketing entropy for the class

$$\widetilde{\mathcal{F}} = \Big\{\tilde{f}:\tilde{f}(\boldsymbol{x},y)=(4M_0D)^{-1}f(\boldsymbol{x})\Big(y-\lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x})\Big),(\boldsymbol{x},y)\in\mathcal{X}\times\mathbb{R}\text{ and } f\in\mathcal{F}\Big\} \tag{2.8}$$

with respect to the Bernstein norm. Here, M_0 is the same constant from Assumption A6 and $\tilde{D}$ is a positive constant that will be determined below.

Lemma 2.6. *Let $\mathcal{F}$ be a class of functions satisfying the conditions above. Then,*

$$H_B(\epsilon,\widetilde{\mathcal{F}},\|\cdot\|_{B,P_0})\le H_B(\epsilon\tilde{D}^{-1},\mathcal{F},\|\cdot\|_{P_0}),\quad\textit{and}\quad\|\tilde{f}\|_{B,P_0}\le\tilde{D}\delta$$

where

$$\tilde{D}=\left(\frac{a_0}{2M_0^2}+\frac{\lambda^2K_0^2}{8M_0^2}e^{\lambda K_0(2M_0)^{-1}}\right)^{1/2}D^{-1} \tag{2.9}$$

and a_0,M_0 are the same constants from Assumption A6.

Proof. Let (L,U) be an ϵ-bracket for $\mathcal{F}$ with respect to $\|\cdot\|_{P_0}$. Consider the class

$$\mathcal{F}'=\Big\{f':f'(\boldsymbol{x},y)=f(\boldsymbol{x})\Big(y-\lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x})\Big),(\boldsymbol{x},y)\in\mathcal{X}\times\mathbb{R}\text{ and } f\in\mathcal{F}\Big\}.$$

Then for $f'\in\mathcal{F}'$ we have

$$\begin{aligned}
&L(\boldsymbol{x})(y-\lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x}))\le f'(\boldsymbol{x},y)\le U(\boldsymbol{x})(y-\lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x})),\ \text{if } y-\lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x})\ge0\ \text{ or}\\
&U(\boldsymbol{x})(y-\lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x}))\le f'(\boldsymbol{x},y)\le L(\boldsymbol{x})(y-\lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x})),\ \text{if } y-\lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x})<0\ .
\end{aligned}$$

Let (L', U') denote the new bracket. Using the definition of the Bernstein norm, convexity of $x \mapsto x^k,\ k \geq 2$ and $\|\psi_0\|_\infty \leq K_0$ we have that

$$
\begin{aligned}
&\left\|(U'-L')(4M_0D)^{-1}\right\|^2_{B,P_0}\\
&= 2\sum_{k=2}^{\infty}\frac{(4M_0D)^{-k}}{k!}\int_{\mathcal{X}\times\mathbb{R}}(U(\boldsymbol{x})-L(\boldsymbol{x}))^k|y-\lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x})|^k dP_0(\boldsymbol{x},y)\\
&\leq 2\sum_{k=2}^{\infty}\frac{(4M_0D)^{-k}}{k!}\int_{\mathcal{X}\times\mathbb{R}}(U(\boldsymbol{x})-L(\boldsymbol{x}))^k 2^{k-1}\Big(|y|^k+\lambda^k|\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x})|^k\Big)dP_0(\boldsymbol{x},y)\\
&\leq \sum_{k=2}^{\infty}\frac{1}{2^kD^kM_0^kk!}\int_{\mathcal{X}\times\mathbb{R}}(U(\boldsymbol{x})-L(\boldsymbol{x}))^k\big(|y|^k+\lambda^kK_0^k\big)dP_0(\boldsymbol{x},y)\\
&\leq \sum_{k=2}^{\infty}\frac{1}{2^kD^kM_0^kk!}\int_{\mathcal{X}\times\mathbb{R}}(U(\boldsymbol{x})-L(\boldsymbol{x}))^k(a_0k!M_0^{k-2}+\lambda^kK_0^k)dP_0(\boldsymbol{x},y)\\
&\leq \frac{a_0}{4M_0^2D^2}\sum_{k=2}^{\infty}\frac{1}{2^{k-2}}\int_{\mathcal{X}\times\mathbb{R}}(U(\boldsymbol{x})-L(\boldsymbol{x}))^2g(\boldsymbol{x})dx+\frac{\lambda^2K_0^2}{4D^2M_0^2}\sum_{k=2}^{\infty}\left(\frac{\lambda K_0}{2M_0}\right)^{k-2}\frac{1}{k!}\int_{\mathcal{X}\times\mathbb{R}}(U(\boldsymbol{x})-L(\boldsymbol{x}))^2g(\boldsymbol{x})d\boldsymbol{x}\\
&\leq \frac{a_0}{2M_0^2D^2}\epsilon^2+\frac{\lambda^2K_0^2}{8D^2M_0^2}\sum_{k=2}^{\infty}\frac{1}{(k-2)!}\left(\frac{\lambda K_0}{2M_0}\right)^{k-2}\ \epsilon^2\\
&\leq \left(\frac{a_0}{2M_0^2D^2}+\frac{\lambda^2K_0^2}{8D^2M_0^2}e^{\lambda K_0(2M_0)^{-1}}\right)\epsilon^2=\tilde{D}^2\ \epsilon^2.
\end{aligned}
$$

This implies that

$$H_B\Big(\epsilon\tilde{D},\widetilde{\mathcal{F}},\|\cdot\|_{B,P_0}\Big)\leq H_B\Big(\epsilon,\mathcal{F},\|\cdot\|_{P_0}\Big)$$

or equivalently

$$H_B\Big(\epsilon,\mathcal{F}',\|\cdot\|_{B,P_0}\Big)\leq H_B\Big(\epsilon\tilde{D}^{-1},\mathcal{F},\|\cdot\|_{P_0}\Big).$$

Using similar calculations we can write

$$
\begin{aligned}
\|\tilde{f}\|^2_{B,\mathbb{P}} &= 2\sum_{k=2}^{\infty}\frac{1}{(4M_0D)^k}\frac{1}{k!}\int_{\mathcal{X}\times\mathbb{R}}|f(\boldsymbol{x})|^k|y-\lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x})|^k dP_(\boldsymbol{x},y)\\
&\leq \frac{1}{D^2}\sum_{k=2}^{\infty}\frac{1}{(2M_0)^k}\frac{1}{k!}\int_{\mathcal{X}\times\mathbb{R}}f(\boldsymbol{x})^2(a_0k!M_0^{k-2}+\lambda^kK_0^k)g(\boldsymbol{x})d\boldsymbol{x}\\
&\leq \left(\frac{a_0}{4M_0^2D^2}\sum_{k=2}^{\infty}\frac{1}{2^{k-2}}+\frac{\lambda^2K_0^2}{8D^2M_0^2}\sum_{k=2}^{\infty}\left(\frac{\lambda K_0}{2M_0}\right)^{k-2}\frac{1}{(k-2)!}\right)\int_{\mathcal{X}\times\mathbb{R}}f(\boldsymbol{x})^2g(\boldsymbol{x})d\boldsymbol{x}\\
&\leq \tilde{D}^2\delta^2
\end{aligned}
$$

which completes the proof.

□

In the next corollary, we consider the class

$$\mathcal{F}=\Big\{x\mapsto f_{\boldsymbol{\alpha}}(\boldsymbol{x})=E_{i,\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^T\boldsymbol{x})-E_{i,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x}),\ \boldsymbol{x}\in\mathcal{X},\boldsymbol{\alpha}\in\mathcal{B}(\boldsymbol{\alpha}_0,\delta)\Big\}$$

where $E_{i,\boldsymbol{\alpha}}(u)=\mathbb{E}\left\{X_i|\boldsymbol{\alpha}^T\boldsymbol{X}=u\right\}$ for $i\in\{1,\ldots,d\}$ and $\delta\in(0,\delta_0)$. Using the same arguments in the proof of Lemma 3.3 with $f(\boldsymbol{x})=x_i$ it follows that for all $x\in\mathcal{X}$ and $\boldsymbol{\alpha},\boldsymbol{\alpha}'\in\mathcal{B}(\boldsymbol{\alpha}_0,\delta)$

$$|f_{\boldsymbol{\alpha}'}(\boldsymbol{x})-f_{\boldsymbol{\alpha}}(\boldsymbol{x})|\leq M\|\boldsymbol{\alpha}'-\boldsymbol{\alpha}\|$$

for the same constant M of that lemma. Now, we can apply Theorem 2.7.11 of van der Vaart and Wellner (1996) to conclude that

$$N_B\big(2\epsilon, \mathcal{F}, \|\cdot\|_{P_0}\big) \le N\big(\epsilon, \mathcal{B}(\boldsymbol{\alpha}_0, \delta), \|\cdot\|\big)$$

where $N\big(\epsilon, \mathcal{B}(\boldsymbol{\alpha}_0, \delta), \|\cdot\|\big)$ is the ϵ-covering number for $\mathcal{B}(\boldsymbol{\alpha}_0, \delta)$ with respect to the norm $\|\cdot\|$ which is of order $(\delta/\epsilon)^d$ for $\epsilon \in (0, \delta)$. Hence, using the inequality $\log(\boldsymbol{x}) \le x$ for $x > 0$ we can find a constant $M' > 0$ depending on d such that

$$H_B\big(\epsilon, \mathcal{F}, \|\cdot\|_{P_0}\big) \le \frac{M'\delta}{\epsilon}.$$

Furthermore, there exists $\tilde{M} > 0$ such that $\|f\|_\infty \le \tilde{M}\delta$ and $\|f\|_{P_0} \le \tilde{M}\delta$.

Lemma 2.7. *Let $\mathcal{F}$ be the class of functions as above and consider the related class*

$$\mathcal{F}' = \Big\{f' : f'(\boldsymbol{x}, y) = f(\boldsymbol{x})\Big(y - \lambda\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x})\Big), (\boldsymbol{x}, y) \in \mathcal{X} \times \mathbb{R}, f \in \mathcal{F}\Big\}. \tag{2.10}$$

Then,

$$E[\|\mathbb{G}_n\|_{\mathcal{F}'}] \lesssim \delta.$$

Proof. Note that for any function $f' \in \mathcal{F}'$ and constant $C > 0$ we have that $\mathbb{G}_n(f'C^{-1}) = C^{-1}\mathbb{G}_n f'$ implying that $\|\mathbb{G}_n\|_{\mathcal{F}'} = 4M_0\tilde{M}\delta\|\mathbb{G}_n\|_{\widetilde{\mathcal{F}}}$, where

$$\widetilde{\mathcal{F}} = \Big\{\tilde{f} : \tilde{f}(\boldsymbol{x}, y) = (4M_0\tilde{M}\delta)^{-1} f'(\boldsymbol{x}, y), f' \in \mathcal{F}'\Big\}.$$

Note also that the constant $\tilde{D}$ in Lemma 2.6 is given by $\tilde{D} \asymp \delta^{-1}$, where $\tilde{D}$ depends on $\tilde{M}$, a_0, M_0 and K_0. Also, using the entropy calculations along with Lemma 2.6 we can show easily that

$$H_B(\epsilon, \widetilde{\mathcal{F}}, \|\cdot\|_{B,P_0}) \lesssim \frac{1}{\epsilon}$$

and that $\|\tilde{f}\|_{B,P_0} \lesssim 1$. Using Lemma 3.4.3 of van der Vaart and Wellner (1996) it follows that there exists some constant $B > 0$ such that

$$E[\|\mathbb{G}_n\|_{\widetilde{\mathcal{F}}}] \lesssim J_n\left(1 + \frac{J_n}{\sqrt{n}B^2}\right)$$

with $J_n = \int_0^B \sqrt{1 + B/\epsilon}\,d\epsilon$. Hence, $E[\|\mathbb{G}_n\|_{\widetilde{\mathcal{F}}}] \lesssim 1$ and $E[\|\mathbb{G}_n\|_{\mathcal{F}'}] \lesssim \delta$ as claimed. □

3. Supplement C: Auxiliary results

Proof of Lemma 4.1. We have:

$$(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0))^T \boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\Big\{(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0))^T \boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\Big\}^{-1}(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0))^T \boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0) = (\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0))^T \boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0).$$

In the parametrizations that we consider, the columns of $\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)$ are orthogonal to $\boldsymbol{\alpha}_0$. We can therefore extend the matrix $\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)$ with a last column $\boldsymbol{\alpha}_0$ to a square nonsingular matrix $\bar{\boldsymbol{J}}_{\mathbb{S}}(\boldsymbol{\beta}_0)$. This leads to the equality

$$\big(\bar{\boldsymbol{J}}_{\mathbb{S}}(\boldsymbol{\beta}_0)\big)^T \boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\Big\{(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0))^T \boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\Big\}^{-1}(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0))^T \boldsymbol{A}\bar{\boldsymbol{J}}_{\mathbb{S}}(\boldsymbol{\beta}_0) = \big(\bar{\boldsymbol{J}}_{\mathbb{S}}(\boldsymbol{\beta}_0)\big)^T \boldsymbol{A}\bar{\boldsymbol{J}}_{\mathbb{S}}(\boldsymbol{\beta}_0).$$

Multiplying on the left by $\left(\left(\bar{\boldsymbol{J}}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\right)^{-1}$ and on the right by $\bar{\boldsymbol{J}}_{\mathbb{S}}(\boldsymbol{\beta}_0)^{-1}$, we get:

$$\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}=\boldsymbol{A}. \tag{3.1}$$

This shows that $\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T$ is a generalized inverse of $\boldsymbol{A}$.

To complete the proof and show that it is indeed the Moore-Penrose inverse of $\boldsymbol{A}$, we first note that

$$\begin{aligned}&\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\\&=\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T.\end{aligned} \tag{3.2}$$

Furthermore,

$$\begin{aligned}&\left(\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\right)^T\\&=\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}^T\\&=\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A},\end{aligned}$$

where the last equality holds since $\boldsymbol{A}$ is symmetric, being a covariance matrix. We have to show that

$$\begin{aligned}&\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\\&=\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T.\end{aligned} \tag{3.3}$$

Multiplying on the left by $\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T$ and on the right by $\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)$, we get:

$$\begin{aligned}&\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\\&=\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\\&=\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0),\end{aligned}$$

and (3.3) follows by the orthogonality relation of the columns of $\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)$ with $\boldsymbol{\alpha}_0$ in the same way as before, replacing the matrix $\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)$ by $\bar{\boldsymbol{J}}_{\mathbb{S}}(\boldsymbol{\beta}_0)$ in the outer factors of the equality relation.

In a similar way we obtain:

$$\begin{aligned}&\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\right)^T\\&=\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}.\end{aligned} \tag{3.4}$$

Since the matrix $\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\boldsymbol{A}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T$ satisfies properties (3.1), (3.2),(3.3) and (3.4), the matrix satisfies the four properties which define the Moore-Penrose pseudo-inverse matrix of $\boldsymbol{A}$. This completes the proof of Lemma 4.1. □

Remark 3.1. The same proof holds for showing that the Moore-Penrose inverse $\tilde{\boldsymbol{A}}$ is given by

$$\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\left\{\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T\tilde{\boldsymbol{A}}\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right\}^{-1}\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)\right)^T.$$

Lemma 3.1 (Derivative $\boldsymbol{\alpha} \mapsto \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})$).

$$\frac{\partial}{\partial \alpha_j}\psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = \left(x_j - E(X_j|\boldsymbol{\alpha}^T\boldsymbol{X} = \boldsymbol{\alpha}_0^T\boldsymbol{x})\right)\psi_0'(\boldsymbol{\alpha}_0^T\boldsymbol{x})$$

and

$$\begin{aligned}\frac{\partial}{\partial \beta_j}\psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} &= \frac{\partial}{\partial \beta_j}\psi_{\mathbb{S}(\boldsymbol{\beta})}(\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{x})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\\ &= \left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)^T\right)_j\left(\boldsymbol{x} - E(\boldsymbol{X}|\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{X} = \mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{x})\right)\psi_0'\left(\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{x}\right),\end{aligned}$$

where $\left(\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)^T\right)_j$ *denotes the jth row of* $\boldsymbol{J}_{\mathbb{S}}(\boldsymbol{\beta}_0)^T$

Proof. We assume without loss of generality that the first component α_1 of $\boldsymbol{\alpha}$ is not equal to zero. Denote the conditional density of $(X_2,\dots,X_d)^T$ given $\boldsymbol{\alpha}^T\boldsymbol{X} = u$ by $h_{\boldsymbol{\alpha}}(\cdot|u)$ Using the change of variables $t_1 = \boldsymbol{\alpha}^T\boldsymbol{x}$, $t_j = x_j$ for $j = 1,\dots,d$, the function $\psi_{\boldsymbol{\alpha}}$ can be written as

$$\begin{aligned}\psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x}) &= \mathbb{E}[\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X} = \boldsymbol{\alpha}^T\boldsymbol{x}]\\ &= \int \psi_0\left(\frac{\alpha_{01}}{\alpha_1}(\boldsymbol{\alpha}^T\boldsymbol{x} - \alpha_2\tilde{x}_2 - \dots - \alpha_d\tilde{x}_d) + \sum_{j=2}^d \alpha_{0j}\tilde{x}_j\right) h_{\boldsymbol{\alpha}}(\tilde{x}_2,\dots,\tilde{x}_d|\boldsymbol{\alpha}^T\boldsymbol{x})\prod_{j=2}^d d\tilde{x}_j\end{aligned}$$

with partial derivatives w.r.t. α_j for $j = 2,\dots,d$ given by,

$$\begin{aligned}&\frac{\partial}{\partial \alpha_j}\psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x}) = \frac{\partial}{\partial \alpha_j}\mathbb{E}[\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X} = \boldsymbol{\alpha}^T\boldsymbol{x}]\\ &= \int \frac{\alpha_{01}}{\alpha_1}(x_j - \tilde{x}_j)\psi_0'\left(\frac{\alpha_{01}}{\alpha_1}(\boldsymbol{\alpha}^T\boldsymbol{x} - \alpha_2\tilde{x}_2 - \dots - \alpha_d\tilde{x}_d) + \sum_{j=2}^d \alpha_{0j}\tilde{x}_j\right) h_{\boldsymbol{\alpha}}(\tilde{x}_2,\dots,\tilde{x}_d|\boldsymbol{\alpha}^T\boldsymbol{x})\prod_{j=2}^d d\tilde{x}_j\\ &\quad + \int \psi_0\left(\frac{\alpha_{01}}{\alpha_1}(\boldsymbol{\alpha}^T\boldsymbol{x} - \alpha_2\tilde{x}_2 - \dots - \alpha_d\tilde{x}_d) + \sum_{j=2}^d \alpha_{0j}\tilde{x}_j\right)\frac{\partial}{\partial \alpha_j}h_{\boldsymbol{\alpha}}(\tilde{x}_2,\dots,\tilde{x}_d|\boldsymbol{\alpha}^T\boldsymbol{x})\prod_{j=2}^d d\tilde{x}_j\end{aligned}$$

which is at $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ equal to

$$\begin{aligned}\frac{\partial}{\partial a_j}\psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} &= \int (x_j - \tilde{x}_j)\psi_0^T\left(\boldsymbol{\alpha}_0^T\boldsymbol{x}\right) h_{\boldsymbol{\alpha}_0}(\tilde{x}_2,\dots,\tilde{x}_d|\boldsymbol{\alpha}_0^T\boldsymbol{x})\prod_{j=2}^d d\tilde{x}_j\\ &= \psi_0'(\boldsymbol{\alpha}_0^T\boldsymbol{x})\left\{x_j - \mathbb{E}(X_j|\boldsymbol{\alpha}_0^T\boldsymbol{X} = \boldsymbol{\alpha}_0^T\boldsymbol{x})\right\}.\end{aligned}$$

For the partial derivatives w.r.t. α_1 we have,

$$\begin{aligned}&\frac{\partial}{\partial \alpha_1}\psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\\ &\quad = \int\left\{\frac{\alpha_{01}}{\alpha_1}x_1 - \frac{\alpha_{01}}{\alpha_1^2}(\boldsymbol{\alpha}^T\boldsymbol{x} - \alpha_2\tilde{x}_2 - \dots - \alpha_d\tilde{x}_d)\right\}\psi_0'\left(\frac{\alpha_{01}}{\alpha_1}(\boldsymbol{\alpha}^T\boldsymbol{x} - \alpha_2\tilde{x}_2 - \dots - \alpha_d\tilde{x}_d) + \sum_{j=2}^d \alpha_{0j}\tilde{x}_j\right)\\ &\qquad\qquad h_{\boldsymbol{\alpha}}(\tilde{x}_2,\dots,\tilde{x}_d|\boldsymbol{\alpha}^T\boldsymbol{x})\prod_{j=2}^d d\tilde{x}_j\\ &+ \int \psi_0\left(\boldsymbol{\alpha}^T\boldsymbol{x} + (\alpha_{01} - \alpha_1)\frac{\boldsymbol{\alpha}^T\boldsymbol{x} - \alpha_2\tilde{x}_2 - \dots - \alpha_d\tilde{x}_d}{\alpha_1} + \sum_{j=2}^d(\alpha_{0j} - \alpha_j)\tilde{x}_j\right)\frac{\partial}{\partial \alpha_1}h(\tilde{x}_2,\dots,\tilde{x}_d|\boldsymbol{\alpha}^T\boldsymbol{x})\prod_{j=2}^d d\tilde{x}_j,\end{aligned}$$

and,

$$\frac{\partial}{\partial a_1}\psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = \psi_0'(\boldsymbol{\alpha}_0^T\boldsymbol{x})\left\{x_1 - E(X_1|\boldsymbol{\alpha}_0^T\boldsymbol{X} = \boldsymbol{\alpha}_0^T\boldsymbol{x})\right\}.$$

This proves the first result of Lemma 3.1. The proof for the second results follows similarly and is omitted. □

Lemma 3.2. *Let $\bar\phi$ be defined by*

$$\bar\phi(\boldsymbol{\alpha}) = \int \boldsymbol{x}\left\{y - \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\right\}\, dP_0(\boldsymbol{x}, y) = \int \boldsymbol{x}\left\{\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x}) - \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\right\}\, dG(\boldsymbol{x}), \tag{3.5}$$

then we have for each $\boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)$,

$$\bar\phi(\boldsymbol{\alpha}) = \mathbb{E}\left[\mathrm{Cov}\left[\boldsymbol{X}, \psi_0(\boldsymbol{\alpha}^T\boldsymbol{X} + (\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X}\right]\right].$$

Moreover,

$$\boldsymbol{\alpha}^T\bar\phi(\boldsymbol{\alpha}) = 0$$

and,

$$(\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\bar\phi(\boldsymbol{\alpha}) = \mathbb{E}\left[\mathrm{Cov}\left[(\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\boldsymbol{X}, \psi_0(\boldsymbol{\alpha}^T\boldsymbol{X} + (\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X}\right]\right] \geq 0,$$

and $\boldsymbol{\alpha}_0$ is the only value such that the above equation holds uniform in $\boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)$.

Proof. We have,

$$\begin{aligned}\bar\phi(\boldsymbol{\alpha}) &= \int \boldsymbol{x}\left\{y - \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\right\}\, dP_0(\boldsymbol{x}, y) = \int \boldsymbol{x}\left\{\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x}) - \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\right\}\, dG(\boldsymbol{x})\\ &= \int \boldsymbol{x}\left[\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x}) - \mathbb{E}\left\{\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X} = \boldsymbol{\alpha}^T\boldsymbol{x}\right\}\right]\, dG(\boldsymbol{x})\\ &= \mathbb{E}\left[\mathrm{Cov}\left[\boldsymbol{X}, \psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X}\right]\right], \end{aligned} \tag{3.6}$$

and

$$\boldsymbol{\alpha}^T\int \boldsymbol{x}\left[\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{x}) - \mathbb{E}\left\{\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X} = \boldsymbol{\alpha}^T\boldsymbol{x}\right\}\right]\, dG(\boldsymbol{x}) = \mathbb{E}\left[\mathrm{Cov}\left[\boldsymbol{\alpha}^T\boldsymbol{X}, \psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X}\right]\right] = \boldsymbol{0}.$$

We next note that,

$$\begin{aligned}(\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\bar\phi(\boldsymbol{\alpha}) &= \mathbb{E}\left[\mathrm{Cov}\left[(\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\boldsymbol{X}, \psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X}\right]\right]\\ &= \mathbb{E}\left[\mathrm{Cov}\left[(\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\boldsymbol{X}, \psi_0(\boldsymbol{\alpha}^T\boldsymbol{X} + (\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X}\right]\right],\end{aligned}$$

which is positive by the monotonicity of ψ_0. This can be seen as follows. Using Fubini's theorem, one can prove that for any random variables X and Y such that XY, X and Y are integrable, we have

$$\mathrm{Cov}\left\{X, Y\right\} = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = \int\left\{\mathbb{P}(X \geq s, Y \geq t) - \mathbb{P}(X \geq s)\mathbb{P}(Y \geq t)\right\} ds\, dt.$$

Denote $Z_1 = (\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\boldsymbol{X}$ and $Z_2 = \psi_0(u + (\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\boldsymbol{X}) = \psi_0(u + Z_1)$, then, using monotonicity of the function ψ_0, we have

$$\begin{aligned}\mathbb{P}(Z_1 \geq z_1, Z_2 \geq z_2) &= \mathbb{P}(Z_1 \geq \max\{z_1, \tilde z_2\}) \geq \mathbb{P}(Z_1 \geq \max\{z_1, \tilde z_2\})\mathbb{P}(Z_1 \geq \min\{z_1, \tilde z_2\})\\ &= \mathbb{P}(Z_1 \geq z_1)\mathbb{P}(Z_2 \geq z_2\}\end{aligned}$$

where

$$\tilde z_2 = \psi_0^{-1}(z_2) - u = \inf\{t \in \mathbb{R} : \psi_0(t) \geq z_2\} - u.$$

We conclude that,

$$\begin{aligned}&\mathrm{Cov}\left\{(\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T\boldsymbol{X}, \psi_0(\boldsymbol{\alpha}^T\boldsymbol{X} + (\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})'\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X} = u\right\}\\ &\qquad = \int\left\{\mathbb{P}(Z_1 \geq z_1, Z_2 \geq z_2) - \mathbb{P}(Z_1 \geq z_1)\mathbb{P}(Z_2 \geq z_2)\right\} ds\, dt \geq 0,\end{aligned}$$

and hence the first part of the Lemma follows. We next prove the uniqueness of the parameter $\boldsymbol{\alpha}_0$. We start by assuming that, on the contrary, there exists $\boldsymbol{\alpha}_1 \neq \boldsymbol{\alpha}_0$ in $\mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)$ such that

$$(\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T \bar{\phi}(\boldsymbol{\alpha}) \geq 0 \qquad \text{and } (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha})^T \bar{\phi}(\boldsymbol{\alpha}) \geq 0 \qquad \text{for all } \boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0),$$

and we consider the point $\boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)$ such that

$$|\alpha_j - \alpha_{j0}| = |\alpha_j - \alpha_{j1}| \qquad \text{for } j = 1, \dots, d.$$

For this point, we have,

$$(\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T \bar{\phi}(\boldsymbol{\alpha}) = -(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha})^T \bar{\phi}(\boldsymbol{\alpha}) \qquad \text{for all } \boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0),$$

which is not possible since both terms should be positive. This completes the proof of Lemma 3.2. □

Lemma 3.3. *Let $f : \mathcal{X} \to \mathbb{R}^k$, $k \leq d$ be a differentiable function on $\mathcal{X}$ such that there exists a constant $M > 0$ satisfying $\|f\|_\infty \leq M$. Then, under the assumptions A1 and A5 we can find a constant $\tilde{M} > 0$ such that for all $\boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)$ we have that*

$$\sup_{(\boldsymbol{x}, \mathcal{X}} \left| \mathbb{E}[f(\boldsymbol{X}) | \boldsymbol{\alpha}^T \boldsymbol{X} = \boldsymbol{\alpha}^T \boldsymbol{x}] - \mathbb{E}[f(\boldsymbol{X}) | \boldsymbol{\alpha}_0^T \boldsymbol{X} = \boldsymbol{\alpha}_0^T \boldsymbol{x}] \right| \leq M \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|.$$

Proof. We can assume without loss of generality that $\alpha_{0,1} \neq 0$ where $\alpha_{0,1}$ is the first component of $\boldsymbol{\alpha}_0$. At the cost of taking a smaller δ_0, we can further assume that $\tilde{\boldsymbol{\alpha}}_1 \neq 0$ for all $\boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)$. Consider the change of variables $t_1 = \boldsymbol{\alpha}^T \boldsymbol{X}$, $t_i = x_i$ for $i = 1, \dots, d$. Then, the density of $(\boldsymbol{\alpha}^T \boldsymbol{X}, X_2, \dots, X_d)$ is given by

$$g_{(\boldsymbol{\alpha}^T \boldsymbol{X}, X_2, \dots, X_d)}(t_1, \dots, t_d) = g\left(\frac{1}{\alpha_1}\Big(t_1 - \alpha_2 t_2 - \dots - \alpha_d t_d), t_2, \dots, t_d\right) \frac{1}{\alpha_1}.$$

Then, for $i = 2, \dots, d$, the conditional density $g_{(X_2, \dots, X_d) | \boldsymbol{\alpha}^T \boldsymbol{X} = u}(x_2, \dots, x_d)$ of the $(d-1)$-dimensional vector $(X_2, \dots, X_d)$ given that $\boldsymbol{\alpha}^T \boldsymbol{X} = u$ is equal to

$$\frac{g\left(\frac{u - \alpha_2 x_2 - \dots - \alpha_d x_d}{\alpha_1}, x_2, \dots, x_d\right)}{\int g\left(\frac{u - \alpha_2 t_2 - \dots - \alpha_d t_d}{\alpha_1}, t_2, \dots, t_d\right) \prod_{j=2}^d dt_j} \quad := \quad h_{\boldsymbol{\alpha}}(x_2, \dots, x_d | u) \tag{3.7}$$

where the domain of integration in the denominator is the set $\{(x_2, \dots, x_d) : (\boldsymbol{x}, \mathcal{X})\}$. Note that $X_1 = (\boldsymbol{\alpha}^T \boldsymbol{X} - \alpha_2 X_2 - \dots - \alpha_d X_d)/\alpha_1$. Thus, for $(\boldsymbol{x}, \mathcal{X})$ we have that

$$\begin{aligned}
&\mathbb{E}[f(\boldsymbol{X}) | \boldsymbol{\alpha}^T \boldsymbol{X} = \boldsymbol{\alpha}^T \boldsymbol{x}] = \mathbb{E}[f(X_1, X_2, \dots, X_d) | \boldsymbol{\alpha}^T \boldsymbol{X} = \boldsymbol{\alpha}^T \boldsymbol{x}] \\
&= \mathbb{E}\left[f\left(\frac{\boldsymbol{\alpha}^T \boldsymbol{X} - \alpha_2 X_2 - \dots - \alpha_d X_d}{\alpha_1}, X_2, \dots, X_d\right) \mid \boldsymbol{\alpha}^T \boldsymbol{X} = \boldsymbol{\alpha}^T \boldsymbol{x}\right] \\
&= \int f\left(\frac{\boldsymbol{\alpha}^T \boldsymbol{x} - \alpha_2 x_2 - \dots - \alpha_d x_d}{\alpha_1}, x_2, \dots, x_d\right) h_{\boldsymbol{\alpha}}(x_2, \dots, x_d | \boldsymbol{\alpha}^T \boldsymbol{x}) \prod_{j=2}^d dx_j.
\end{aligned}$$

Note now that function

$$\boldsymbol{\alpha} \mapsto h_{\boldsymbol{\alpha}}(x_2, \dots, x_d | \boldsymbol{\alpha}^T \boldsymbol{x}) = \frac{g\left(\frac{\boldsymbol{\alpha}^T \boldsymbol{x} - \alpha_2 x_2 - \dots - \alpha_d x_d}{\alpha_1}, x_2, \dots, x_d\right)}{\int g\left(\frac{\boldsymbol{\alpha}^T \boldsymbol{x} - \alpha_2 t_2 - \dots - \alpha_d t_d}{\alpha_1}, t_2, \dots, t_d\right) \prod_{j=2}^d dt_j}$$

is continuously differentiable on $\mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)$. This follows from assumptions A1 and A5 together with Lebesgue dominated convergence theorem which allows us to differentiate the density g under the integral sign. With some notation abuse we write $\partial h/\partial x_i$ for the i-th partial derivative of $\boldsymbol{\alpha} \mapsto h_{\boldsymbol{\alpha}}(x_2,\dots,x_d|\boldsymbol{\alpha}^T\boldsymbol{x})$. Straightforward calculations yield

$$\begin{aligned}\frac{\partial h_{\boldsymbol{\alpha}}}{\partial \alpha_1} &= g\left(\frac{\boldsymbol{\alpha}^T\boldsymbol{x}-\alpha_2x_2-\dots-\alpha_dx_d}{\alpha_1},x_2,\dots,x_d\right)\\ &\quad\times\frac{\int\sum_{i=2}^d(x_i-t_i)\frac{\partial g}{\partial x_1}\left(\frac{\boldsymbol{\alpha}^T\boldsymbol{x}-\alpha_2t_2-\dots-\alpha_dt_d}{\alpha_1},t_2,\dots,t_d\right)\prod_{j=2}^d dt_j}{\alpha_1^2\left(\int g\left(\frac{\boldsymbol{\alpha}^T\boldsymbol{x}-\alpha_2t_2-\dots-\alpha_dt_d}{\alpha_1},t_2,\dots,t_d\right)\prod_{j=2}^d dt_j\right)^2},\end{aligned}$$

and for $i=2,\dots,d$

$$\begin{aligned}\frac{\partial h_{\boldsymbol{\alpha}}}{\partial \alpha_i} &= -g\left(\frac{\boldsymbol{\alpha}^T\boldsymbol{x}-\alpha_2x_2-\dots-\alpha_dx_d}{\alpha_1},x_2,\dots,x_d\right)\\ &\quad\times\frac{\int(x_i-t_i)\frac{\partial g}{\partial x_i}\left(\frac{\boldsymbol{\alpha}^T\boldsymbol{x}-\alpha_2t_2-\dots-\alpha_dt_d}{\alpha_1},t_2,\dots,t_d\right)\prod_{j=2}^d dt_j}{\alpha_1\left(\int g\left(\frac{\boldsymbol{\alpha}^T\boldsymbol{x}-\alpha_2t_2-\dots-\alpha_dt_d}{\alpha_1},t_2,\dots,t_d\right)\prod_{j=2}^d dt_j\right)^2}.\end{aligned}$$

Assumptions A1 and A5 allow us to find a constant $D>0$ depending on R, $\underline{c}_0$, $\bar{c}_0$ and $\bar{c}_1$ such that

$$\left\|\frac{\partial h_{\boldsymbol{\alpha}}}{\partial \alpha_i}\right\|_\infty \le D,$$

for $i=1,\dots,d$. Consider now the function $\alpha \mapsto E[f(X)|\boldsymbol{\alpha}^T\boldsymbol{X}=\boldsymbol{\alpha}^T\boldsymbol{x}]$. Using the assumptions of the lemma and applying again Lebesgue dominated convergence theorem we conclude for $i\in\{1,\dots,d\}$ that we have

$$\begin{aligned}&\frac{\partial \mathbb{E}[f(X)|\boldsymbol{\alpha}^T\boldsymbol{X}=\boldsymbol{\alpha}^T\boldsymbol{x}]}{\partial\alpha_i}\\ &=\int f\left(\frac{\boldsymbol{\alpha}^T\boldsymbol{x}-\alpha_2x_2-\dots-\alpha_dx_d}{\alpha_1},x_2,\dots,x_d\right)\frac{\partial h_{\boldsymbol{\alpha}}(x_2,\dots,x_d|\boldsymbol{\alpha}^T\boldsymbol{x})}{\partial\alpha_i}\prod_{j=2}^d dx_j.\end{aligned}$$

Furthermore, we have that

$$\sup_{(\boldsymbol{x},\mathcal{X}}\left|\frac{\partial \mathbb{E}[f(\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X}=\boldsymbol{\alpha}^T\boldsymbol{x}]}{\partial\alpha_i}\right| \le MD\int\prod_{j=2}^d dx_j = M',$$

for all $i\in\{1,\dots,d\}$ and $(\boldsymbol{x},\mathcal{X})$ and $\boldsymbol{\alpha}\in\mathcal{B}(\boldsymbol{\alpha}_0,\delta)$. The results now follow using a first order Taylor expansion to obtain

$$\left|\mathbb{E}[f(\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X}=\boldsymbol{\alpha}^T\boldsymbol{x}]-\mathbb{E}[f(\boldsymbol{X})|\boldsymbol{\alpha}_0^T\boldsymbol{X}=\boldsymbol{\alpha}_0^T\boldsymbol{x}]\right| = \left|\sum_{i=1}^d\frac{\partial \mathbb{E}[f(\boldsymbol{X})|\widetilde{\boldsymbol{\alpha}}^T\boldsymbol{X}=\widetilde{\boldsymbol{\alpha}}^T\boldsymbol{x}]}{\partial\alpha_i}(\alpha_i-\alpha_{0,i})\right|$$

for some $\widetilde{\boldsymbol{\alpha}}\in\mathbb{R}^d$ such that $\|\widetilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|\le\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|$. Bounding the right side of the preceding display by $\tilde{M}\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|$ with $\tilde{M}=dM'$ gives the result.

□

Lemma 3.4. *Denote for $i\in\{1,\dots,d\}$ the ith component of the function $u\mapsto\mathbb{E}[\boldsymbol{X}|\boldsymbol{\alpha}^T\boldsymbol{X}=u]$ by $E_{i,\boldsymbol{\alpha}}$. Then $E_{i,\boldsymbol{\alpha}}$ has a total bounded variation. Furthermore, there exists a constant $B>0$ such that for all $\boldsymbol{\alpha}\in\mathcal{B}(\boldsymbol{\alpha}_0,\delta_0)$*

$$\|E_{i,\boldsymbol{\alpha}}\|_\infty \le B, \quad \textit{and} \quad \int_{\mathcal{I}_{\boldsymbol{\alpha}}}|E'_{i,\boldsymbol{\alpha}}(u)|du\le B.$$

where $\mathcal{I}_{\boldsymbol{\alpha}}=\{\boldsymbol{\alpha}^T\boldsymbol{x}:\boldsymbol{x}\in\mathcal{X}\}$.

Proof. Since $\mathcal{X} \subset \mathcal{B}(0,R)$, it is clear that $\|E_{i,\boldsymbol{\alpha}}\|_\infty \le R$. As above let us assume without loss of generality that the first component of $\boldsymbol{\alpha}_0$ is not equal to 0. At the cost of taking a smaller δ_0, we can further assume that $\tilde{\alpha}_1 \neq 0$ for all $\boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)$. We known that for $i = 2, \dots, d$

$$E_{i,\boldsymbol{\alpha}}(u) = \int x_i h_{\boldsymbol{\alpha}}(x_2, \dots, x_d | u) dx_2 \dots dx_d,$$

where integration is done over the set $\{(x_2, \dots, x_d) : (\boldsymbol{x}, \mathcal{X})\}$ and $u \in \mathcal{I}_{\boldsymbol{\alpha}} \subset (a_0 - \delta_0 R, b_0 + \delta_0 R)$ and where $h_{\boldsymbol{\alpha}}$ denotes conditional density of $(X_2, \dots, X_d)'$ given $\boldsymbol{\alpha}^T \boldsymbol{X} = u$, defined in (3.7). Using assumptions A1 and A5 along with the Lebesgue dominated convergence theorem we are allowed to write

$$E'_{i,\boldsymbol{\alpha}}(u) = \int x_i \frac{\partial}{\partial u} h_{\boldsymbol{\alpha}}(x_2, \dots, x_d | u) \; dx_2 \dots dx_d.$$

Straightforward calculations yield that

$$\begin{aligned}
&\frac{\partial}{\partial u} h_{\boldsymbol{\alpha}}(x_2, \dots, x_d | u) \\
&= \frac{\frac{\partial g}{\partial x_1}\left(\frac{u - \alpha_2 x_2 - \dots - \alpha_d x_d}{\alpha_1}, x_2, \dots, x_d\right)}{\alpha_1 \left(\int g\left(\frac{u - \alpha_2 t_2 - \dots - \alpha_d t_d}{\alpha_1}, t_2, \dots, t_d\right) \prod_{j=2}^d dt_j\right)} \\
&\quad - \frac{g\left(\frac{1}{\alpha_1}\left(u - \alpha_2 x_2 - \dots - \alpha_d x_d\right), x_2, \dots, x_d\right) \int \frac{\partial g}{\partial x_1}\left(\frac{u - \alpha_2 t_2 - \dots - \alpha_d t_d}{\alpha_1}, t_2, \dots, t_d\right) \prod_{j=2}^d dt_j}{\alpha_1 \left(\int g\left(\frac{u - \alpha_2 t_2 - \dots - \alpha_d t_d}{\alpha_1}, t_2, \dots, t_d\right) \prod_{j=2}^d dt_j\right)^2}.
\end{aligned}$$

Thus, we can find constant $C > 0$ depending only on $|\alpha_{0,1}|$, $\underline{c}_0$, $\underline{c}_1$, $\bar{c}_1$ and R such that $\int |E'_{i,\boldsymbol{\alpha}}(u)| du \le C$ for all $\alpha \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)$. Now $B = \max(R, C)$ gives the claimed inequalities. If $i = 1$, then

$$E_{1,\boldsymbol{\alpha}}(u) = \frac{1}{\alpha_1}\Big(u - \alpha_j \sum_{j=2}^d E_{j,\boldsymbol{\alpha}}(u)\Big), \quad \text{and} \quad e'_{1,\boldsymbol{\alpha}}(u) = \frac{1}{\alpha_1}\Big(1 - \alpha_j \sum_{j=2}^d e'_{j,\boldsymbol{\alpha}}(u)\Big).$$

for $u \in I_{\boldsymbol{\alpha}}$. We conclude again that the claimed inequalities are true at the cost of increasing the constant B obtained above. □

Lemma 3.5. *Let f be a function defined on some interval $[a,b]$ such that*

$$\|f\|_\infty \le M, \quad V(f, [a,b]) := \sup_{a = x_0 < x_1 \dots < x_n = b} \sum_{j=1}^n |f(x_j) - f(x_{j-1})| \le M$$

for some finite constant $M > 0$. Then, there exist two non-decreasing functions f_1 and f_2 on $[a,b]$ such that $\|f_1\|_\infty, \|f_2\|_\infty \le 2M$ and $f = f_2 - f_1$.

Proof. The fact that $f = f_2 - f_1$ with f_1 and f_2 non-decreasing on $[a,b]$ follows from the well-known Jordan's decomposition. Furthermore, we can take $f_1(\boldsymbol{x}) = V(f, [a, x])$ and $f_2(\boldsymbol{x}) = f(\boldsymbol{x}) - f_1(\boldsymbol{x})$for $(\boldsymbol{x}, [a,b]$. By assumption, $\|f_1\|_\infty \le M \le 2M$ and $\|f_2\| \le \|f\|_\infty + \|f_1\|_\infty \le 2M$. □

Lemma 3.6. *Under Assumptions A4-A5, we can find a constant $C > 0$ such that for all $\boldsymbol{\alpha}$ close enough to $\boldsymbol{\alpha}_0$ we have that*

$$\psi'_{\boldsymbol{\alpha}}(u) > C$$

for all $u \in \mathcal{I}_{\boldsymbol{\alpha}}$.

Proof. We assume again that $a_1 \neq 0$. By calculations similar to the calculations made in the proof of Lemma 3.1, we get

$$\psi_{\boldsymbol{\alpha}}(u) = \frac{\alpha_{01}}{\alpha_1}\int \psi_0'\left(\frac{\alpha_{01}}{\alpha_1}(u - \alpha_2\tilde{x}_2 - \ldots - \alpha_d\tilde{x}_d) + \sum_{j=2}^d \alpha_{0j}\tilde{x}_j\right) h_{\boldsymbol{\alpha}}(\tilde{x}_2,\ldots,\tilde{x}_d|u)\prod_{j=2}^d d\tilde{x}_j$$
$$+ \int \psi_0\left(\frac{\alpha_{01}}{\alpha_1}(u - \alpha_2\tilde{x}_2 - \ldots - \alpha_d\tilde{x}_d) + \sum_{j=2}^d \alpha_{0j}\tilde{x}_j\right) \frac{\partial}{\partial u} h(\tilde{x}_2,\ldots,\tilde{x}_d|u)\prod_{j=2}^d d\tilde{x}_j.$$

Now, a Taylor expansion of α_i in the neighborhood of $\alpha_{0,i}$ and using that $\alpha_{0,1}/\alpha_1 = 1 - \epsilon_1/\alpha_{0,1} + o(\epsilon_1)$ yields

$$\begin{aligned}
&\psi_0\left(\frac{\alpha_{0,1}}{\alpha_1}(u - \alpha_2 x_2 - \cdots - \alpha_d x_d) + \alpha_{0,2}x_2 + \cdots + \alpha_{0,d}x_d\right)\\
&= \psi_0\left(u - \frac{\epsilon_1}{\alpha_{0,1}}(u - \epsilon_2 x_2 - \ldots - \epsilon_d x_d) + o(\epsilon_1)\right)\\
&= \psi_0(u) - \frac{\epsilon_1}{\alpha_{0,1}}(u - \epsilon_2 x_2 - \ldots - \epsilon_d x_d)\psi_0'(u) + o(\epsilon_1)\\
&= \psi_0(u) - \frac{\epsilon_1}{\alpha_{0,1}} u\psi_0'(u) + o(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|).
\end{aligned}$$

Using the Lebesgue dominated convergence theorem and the fact that $h_{\boldsymbol{\alpha}}(\tilde{x}_2,\ldots,\tilde{x}_d|u)$ is a conditional density it follows that

$$\int \psi_0\left(\frac{\alpha_{01}}{\alpha_1}(u - \alpha_2\tilde{x}_2 - \ldots - \alpha_d\tilde{x}_d) + \sum_{j=2}^d \alpha_{0j}\tilde{x}_j\right) \frac{\partial}{\partial u} h(\tilde{x}_2,\ldots,\tilde{x}_d|u)\prod_{j=2}^d d\tilde{x}_j, = o(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0),$$

such that

$$\psi_{\boldsymbol{\alpha}}'(u) \geq C\left(1 - \frac{\epsilon_1}{\alpha_{0,1}}\right) + o(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \geq C > 0,$$

provided that $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|$ is small enough. □

Lemma 3.7. *If $h \asymp n^{-1/7}$, then there exists a constant $B > 0$ such that for all $\boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta)$*

$$\|\psi_{nh,\boldsymbol{\alpha}}'\|_\infty \leq B \qquad \textit{and} \qquad \int_{\mathcal{I}_{\boldsymbol{\alpha}}} |\psi_{nh,\boldsymbol{\alpha}}''(u)|du \leq B,$$

where $\mathcal{I}_{\boldsymbol{\alpha}} = \{\boldsymbol{\alpha}^T\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}\}$

Proof. Using integration by parts and Proposition 3.2, we have for all $u \in \mathcal{I}_{\boldsymbol{\alpha}}$

$$\begin{aligned}
\psi_{nh,\boldsymbol{\alpha}}'(u) &= \frac{1}{h}\int K\left(\frac{u-x}{h}\right) d\hat{\psi}_{n\boldsymbol{\alpha}}(x)\\
&= \frac{1}{h}\int K\left(\frac{u-x}{h}\right)\psi_{\boldsymbol{\alpha}}'(x)dx + \frac{1}{h^2}\int K'\left(\frac{u-x}{h}\right)(\hat{\psi}_{n\boldsymbol{\alpha}}(x) - \psi_{\boldsymbol{\alpha}}(x))dx\\
&= \frac{1}{h}\int K\left(\frac{u-x}{h}\right) d\psi_{\boldsymbol{\alpha}}(x) + \frac{1}{h}\int K'(w)(\hat{\psi}_{n\boldsymbol{\alpha}}(u+hw) - \psi_{\boldsymbol{\alpha}}(u+hw))dw\\
&= \psi_{\boldsymbol{\alpha}}'(u) + O(h^2) + O_p(h^{-1}\log n n^{-1/3}) = \psi_{\boldsymbol{\alpha}}'(u) + o_p(1).
\end{aligned}$$

This proves the first part of Lemma 3.7. For the second part, we get by a similar calculation that,

$$\begin{aligned}
\psi_{nh,\boldsymbol{\alpha}}''(u) &= \frac{1}{h}\int K\left(\frac{u-x}{h}\right)\psi_{\boldsymbol{\alpha}}''(x)dx + \frac{1}{h^2}\int K''(w)(\hat{\psi}_{n\boldsymbol{\alpha}}(u+hw) - \psi_{\boldsymbol{\alpha}}(u+hw))dw\\
&= \frac{1}{h}\int K\left(\frac{u-x}{h}\right)\psi_{\boldsymbol{\alpha}}''(x)dx + O_p(h^{-2}\log n n^{-1/3}).
\end{aligned}$$

Since $h^{-2}\log n n^{-1/3} = o(1)$ for $h \asymp n^{-1/7}$, the second part follows by Assumption A10. □

References

Balabdaoui, F., Durot, C. and Jankowski, H. (2016). Least squares estimation in the monotone single index model. *arXiv preprint arXiv:1610.06026*.

Balabdaoui, F., Groeneboom, P. and Hendrickx, K. (2017). Score estimation in the monotone single index model. working paper.

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge Univ. Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes. Springer Series in Statistics*. Springer-Verlag, New York. With applications to statistics. MR1385671 (97g:60035)