Submitted to Econometrica

Appendices for Nonparametric Instrumental Variables Estimation Under Misspecification: Proofs and Some Additional Results

November 16, 2019

APPENDICES FOR NONPARAMETRIC INSTRUMENTAL VARIABLES ESTIMATION UNDER MISSPECIFICATION: PROOFS AND SOME ADDITIONAL RESULTS

APPENDIX A: SUPPLEMENTARY MATERIAL

In this appendix we provide some additional results that supplement those in the main body of the paper. In Theorem A.1 we provide high-level conditions for the consistency of NPIV estimators of the form detailed in Section 1. In Theorem A.2 we show that under an arbitrarily small degree of misspecification the NPIV moment condition may have no solution.

A.1 Consistency of NPIV Estimators Under Correct Specification

Below we provide conditions for consistency of an NPIV estimator of the type described in Section 1. The conditions are likely too high-level for practical use but we include the result and proof in order to provide additional exposition for those unfamiliar with NPIV or the use of regularization in statistical inverse problems.

THEOREM A.1 *Suppose Assumption 1.1 holds and $g_0 = A[h_0]$. Let $\{Q_k\}_{k=1}^{\infty}$ be a sequence of bounded linear operators between $\mathcal{B}_Z$ and $\mathcal{B}_X$ so that for any $g \in R(A)$ (where $R(A) \subset \mathcal{B}_Z$ is the range of the operator A):*

$$\underset{k\to\infty}{lim} ||Q_k[g] - A^{-1}[g]||_{\mathcal{B}_X} = 0$$

Let $\{\hat{Q}_{n,k}\}_{k=1}^{\infty}$ be a sequence of estimators so that for each k:

$$||\hat{Q}_{n,k} - Q_k||_{op} \to^p 0$$

Let $\hat{g}_n$ be an estimator with:

$$||\hat{g}_n - g_0||_{\mathcal{B}_Z} \to^p 0$$

Let $\{K_n\}_{n=1}^{\infty}$ be a sequence of natural numbers so that $K_n \to \infty$. Define the estimator $\hat{h}_{n,K_n}$ by $\hat{h}_{n,K_n} = \hat{Q}_{n,K_n}[\hat{g}_n]$. If $E[Y - h_0(X)|Z] = 0$ and K_n grows sufficiently slowly with n then:

$$||\hat{h}_{n,K_n} - h_0||_{\mathcal{B}_X} \to^p 0$$

PROOF: By Assumption 1.1, $h_0 = A^{-1}[g_0]$. By the triangle inequality and the boundedness and linearity of Q_{K_n}:

$$\begin{aligned} ||\hat{h}_{n,K_n} - h_0||_{\mathcal{B}_X} \leq ||\hat{Q}_{n,K_n} - Q_{K_n}||_{op}||\hat{g}_n||_{\mathcal{B}_Z} + ||Q_{K_n}||_{op}||\hat{g}_n - g_0||_{\mathcal{B}_Z} \\ + ||Q_{K_n}[g_0] - A^{-1}[g_0]||_{\mathcal{B}_X} \end{aligned}$$

By assumption, if $K_n \to \infty$ then $||Q_{K_n}[g_0] - A^{-1}[g_0]||_{\mathcal{B}_X} \to 0$.

By assumption, for any fixed k, $||\hat{Q}_{n,k} - Q_k||_{op} \to^p 0$. So if K_n grows sufficiently slowly:

$$||\hat{Q}_{n,K_n} - Q_{K_n}||_{op} \to^p 0$$

Also by assumption $||\hat{g}_n - g_0||_{\mathcal{B}_Z} \to^p 0$ and so by the triangle inequality $||\hat{g}_n||_{\mathcal{B}_X} \to^p ||g_0||_{\mathcal{B}_X} < \infty$. Therefore, if K_n grows sufficiently slowly:

$$||\hat{Q}_{n,K_n} - Q_{K_n}||_{op}||\hat{g}_n||_{\mathcal{B}_Z} \to^p 0$$

Furthermore, because $||\hat{g}_n - g_0||_{\mathcal{B}_Z} \to^p 0$, if K_n and therefore $||Q_{K_n}||_{op}$ grows sufficiently slowly:

$$||Q_{K_n}||_{op}||\hat{g}_n - g_0||_{\mathcal{B}_Z} \to^p 0$$

Q.E.D.

A.2 Non-Existence of a Solution to the NPIV Moment Condition Under Misspecification

Below we show that even if the degree of misspecification is very small, the NPIV estimating equation may not have a solution. Provided that is, that

Assumptions 1.1 and 1.2 hold. This contrasts with the case of instrumental validity, under which the estimating equation must have a solution by construction. The possibility that a solution to the NPIV moment condition might not exist is noted in Darolles *et al.* (2011) without reference to misspecification. We capture the non-existence result formally in the Theorem A.2 below.

Note that a similar situation may arise in the parametric (that is, the finite-dimensional) case. In an over-identified linear instrumental variables model, if some instruments are not valid then there may be no parameter value that satisfies all of the moment conditions simultaneously. Indeed, the non-existence of a solution to the population moment conditions is testable and forms the basis of common tests of instrumental validity (for example the Sargan-Hansen test).

However, a necessary condition for over-identification in the linear instrumental variables case is that there be strictly more instruments than regressors. By contrast, Theorem A.2 applies even when there are as many regressors as there are instruments. Moreover, recent work by Chen & Santos (2018) suggests that the existence of a solution to the NPIV estimating equation may not be a testable hypothesis in some cases.

THEOREM A.2 *Let $h_0 \in int(\mathcal{H})$ and A be an infinite-dimensional linear operator from $\mathcal{B}_X$ to $\mathcal{B}_Z$ that satisfies Assumptions 1.1 and 1.2. Then for any $b > 0$ there exists a $u_0 \in \mathcal{B}_Z$ with $||u_0||_{\mathcal{B}_Z} \leq b$ so that $A[h] \neq A[h_0] + u_0 = g_0$ for all $h \in \mathcal{H}$.*

PROOF: The range of a compact, injective, infinite-dimensional linear operator between Banach spaces cannot be closed (Kress (2014) Chapter 15). Hence there must exist some $f \in \mathcal{B}_Z$ such that $f \notin R(A)$. Note that this implies $||f||_{\mathcal{B}_Z} \neq 0$ and let $u_0 = \frac{f}{||f||_{\mathcal{B}_Z}} b \in \mathcal{B}_Z$. Then $||u_0||_{\mathcal{B}_Z} \leq b$ and the linearity of A implies $A[h_0] + u_0 \notin R(A)$.

Q.E.D.

APPENDIX B: PROOFS

This appendix contains proofs of results in the main body of the paper and supporting lemmas.

Proofs For Section 2

To state the following lemma, we define the 'modulus of continuity' (see Chen & Pouzo (2015)). Let $\mathcal{H} \subseteq \mathcal{B}_X$, the modulus of continuity at some $h_0 \in \mathcal{H}$ for the subset $\mathcal{H}$ and for a given positive scalar b is given by:

$$\omega(b, h_0, \mathcal{H}) = \sup_{h \in \mathcal{H}:\, ||A[h]-A[h_0]||_{\mathcal{B}_Z} \leq b} ||h - h_0||_{\mathcal{B}_X}$$

LEMMA 2.1 *Fix $\mu_{XZ\eta}$ so that the conditional expectation operator A satisfies Assumptions 1.1 and 1.2.*

Let $\hat{h}_n$ be an NPIV estimator that is consistent under instrumental validity whenever $h_0 \in \mathcal{H}$. That is, for any $h_0 \in int(\mathcal{H})$, if $u_0 = 0$ then $\underset{n\to\infty}{plim}||\hat{h}_n - h_0||_{\mathcal{B}_X} = 0$. Then, $bias_{\hat{h}_n}(b) \geq \omega(b, h_0, \mathcal{H})$.

PROOF: Suppose $g_0 \in A[\mathcal{H}]$, if $u_0 = 0$ then $h_0 = A^{-1}[g_0]$ and so by consistency, for any $g_0 \in A[\mathcal{H}]$, $\underset{n\to\infty}{\text{plim}}||\hat{h}_n - A^{-1}[g_0]||_{\mathcal{B}_X} = 0$.

Now fix h_0. Pick some $u_0 \in A[\mathcal{H} - h_0]$ with $||u_0||_{\mathcal{B}_Z} \leq b$ and let $h = h_0 + A^{-1}[u_0]$. Then $h \in \mathcal{H}$, $||A[h_0] - A[h]||_{\mathcal{B}_Z} \leq b$ and $g_0 = A[h_0] + u_0 = A[h] \in A[\mathcal{H}]$, so $\underset{n\to\infty}{\text{plim}}||\hat{h}_n - h||_{\mathcal{B}_X} = 0$.

And so by the triangle inequality:

$$\underset{n\to\infty}{\text{plim}}||\hat{h}_n - h_0||_{\mathcal{B}_X} \geq ||h - h_0||_{\mathcal{B}_X}$$

Taking the supremum over all over all $u_0 \in A[\mathcal{H} - h_0]$ with $||u_0||_{\mathcal{B}_Z} \leq b$, or equivalently the supremum over all $h \in \mathcal{H}$ with $||A[h_0] - A[h]||_{\mathcal{B}_Z} \leq b$, we get:

$$\sup_{u_0 \in A[\mathcal{H}-h_0]:\, ||u_0||_{\mathcal{B}_Z} \leq b} \operatorname*{plim}_{n\to\infty} ||\hat{h}_n - h_0||_{\mathcal{B}_X} \geq \sup_{h\in\mathcal{H}:\, ||A[h]-A[h_0]||_{\mathcal{B}_Z}\leq b} ||h-h_0||_{\mathcal{B}_X}$$

Since $A[\mathcal{H} - h_0] \subseteq R(A)$:

$$\sup_{u_0 \in R(A):\, ||u_0||_{\mathcal{B}_Z} \leq b} \operatorname*{plim}_{n\to\infty} ||\hat{h}_n - h_0||_{\mathcal{B}_X} \geq \sup_{h\in\mathcal{H}:\, ||A[h]-A[h_0]||_{\mathcal{B}_Z}\leq b} ||h-h_0||_{\mathcal{B}_X}$$

Applying the definitions of the worst-case asymptotic bias and the modulus of continuity gives the result.

Q.E.D.

LEMMA 2.2 *Let A be an infinite-dimensional linear operator from $\mathcal{B}_X$ to $\mathcal{B}_Z$ that satisfies Assumptions 1.1 and 1.2.*

Suppose $h \in int(\mathcal{H})$. Let $r > 0$ be the radius of a closed ball in $int(\mathcal{H})$ centered at h (such a ball must exist). For any $b > 0$:

$$\omega(b, h, \mathcal{H}) \geq r$$

PROOF: By definition of the closed ball of radius r centered at h in $int(\mathcal{H})$, if $h' \in \mathcal{B}_X$ and $||h - h'||_{\mathcal{B}_X} \leq r$ then $h' \in \mathcal{H}$.

A^{-1} is an unbounded linear operator so for any $C > 0$ we can find $h' \in R(A)$ with $||h'||_{\mathcal{B}_X} = 1$ and $||A[h']||_{\mathcal{B}_Z} \leq C$.

Fix some $0 < C \leq \frac{b}{r}$ and fix some h' with $||h'||_{\mathcal{B}_X} = 1$ that satisfies the inequality above. Define h'' by $h'' = h + rh'$.

Note that $||h - h''||_{\mathcal{B}_X} = r$ and therefore $h'' \in \mathcal{H}$ and using linearity of A and the properties of norms:

$$||A[h] - A[h'']||_{\mathcal{B}_Z} = r||A[h']||_{\mathcal{B}_Z} \leq b$$

And so h'' satisfies $||A[h]-A[h'']||_{\mathcal{B}_Z} \leq b$, $||h-h''||_{\mathcal{B}_X} = r$ and since $h'' \in \mathcal{H}$:

$$\sup_{h'\in\mathcal{H}:\, ||A[h]-A[h']||_{\mathcal{B}_Z}\leq b} ||h'-h||_{\mathcal{B}_X} \geq r$$

Q.E.D.

PROOF OF THEOREM 2.1: The first two statements follow immediately from Lemmas 2.1 and 2.2. The final statement follows from the fact that the Banach space $\mathcal{B}_X$ must contain a closed ball centered at h_0 of arbitrarily large radius.

Q.E.D.

PROOF OF THEOREM 2.2: It is well-known that a continuous and injective function defined on a compact set has a continuous inverse. It is also well-known that a continuous linear operator on a finite-dimensional linear space has a continuous inverse. So under either of Assumptions 2.1.i and 2.1.ii $A_{\mathcal{H}}^{-1}$ is continuous. Since $A_{\mathcal{H}}^{-1}$ is linear it is also uniformly continuous.

P_Z is uniformly continuous by assumption, and the composition of two uniformly continuous operators is uniformly continuous, hence $A_{\mathcal{H}}^{-1}P_Z$ is uniformly continuous. By uniform continuity, for any $g \in R(A)$:

$$\lim_{b\to 0} \sup_{g_0 \in R(A):\, ||g_0 - g|| \leq b} ||A_{\mathcal{H}}^{-1}P_Z[g_0] - A_{\mathcal{H}}^{-1}P_Z[g]||_{\mathcal{B}_X} = 0$$

Set $g = A[h_0]$ in the above. Since $h_0 \in \mathcal{H}$ we have $A_{\mathcal{H}}^{-1}P_Z[A[h_0]] = h_0$. Since $R(A)$ is a linear space, for any $g_0 \in R(A)$ there exists $u_0 \in R(A)$ so that $g_0 = u_0 + A[h_0]$. So we get:

$$\lim_{b\to 0} \sup_{u_0 \in R(A):\, ||u_0|| \leq b} ||A_{\mathcal{H}}^{-1}P_Z[g_0] - h_0||_{\mathcal{B}_X} = 0 \tag{B.1}$$

The triangle inequality implies:

$$||\hat{h}_n - h_0||_{\mathcal{B}_X} \leq ||A_{\mathcal{H}}^{-1}P_Z[g_0] - h_0||_{\mathcal{B}_X} + ||\hat{h}_n - A_{\mathcal{H}}^{-1}P_Z[g_0]||_{\mathcal{B}_X}$$

Since $||\hat{h}_n - A_{\mathcal{H}}^{-1}P_Z[g_0]||_{\mathcal{B}_X} \to^p 0$ it follows that:

$$\operatorname*{plim}_{n\to\infty} ||\hat{h}_n - h_0||_{\mathcal{B}_X} = ||A_{\mathcal{H}}^{-1}P_Z[g_0] - h_0||_{\mathcal{B}_X}$$

Substituting the above into (B.1) and applying the definition of the worst-case asymptotic bias gives $\lim_{b\to 0} bias_{\hat{h}_n}(b) = 0$.

Q.E.D.

LEMMA 2.3 *Suppose $\mathcal{H}$ and h_0 satisfy Assumptions 2.1.i and 2.2 and A satisfies Assumptions 1.1 and 1.2. Then:*

$$\lim_{b\to 0} \frac{\omega(b, h_0, \mathcal{H})}{b} = \infty$$

PROOF: Assume on the contrary, then for some $b > 0$ there exists a scalar C so that for any $h \in \mathcal{H}$:

$$||A[h] - A[h_0]||_{\mathcal{B}_Z} \leq b \implies ||h - h_0||_{\mathcal{B}_X} \leq C||A[h] - A[h_0]||_{\mathcal{B}_Z}$$

By Assumption 2.2 $\mathcal{H}$ is convex, and symmetry of $\mathcal{H}$ implies $0 \in \mathcal{H}$. Further we have $\frac{1}{\alpha}h_0 \in \mathcal{H}$ for some $\alpha \in (0,1)$. So for any $h \in \mathcal{H}$ there exists $h' \in \mathcal{H}$ so that $(1-\alpha)h = h' - h_0 \in \mathcal{H}$. Therefore the above implies that for any $h \in \mathcal{H}$:

$$||A[h]||_{\mathcal{B}_Z} \leq \frac{1}{1-\alpha} b \implies ||h||_{\mathcal{B}_X} \leq C||A[h]||_{\mathcal{B}_Z}$$

Further, note that:

$$||A[h]||_{\mathcal{B}_Z} \leq ||A||_{op}||h||_{\mathcal{B}_X}$$

Assumption 1.1 implies that $||A||_{op} > 0$. And so for any $h \in \mathcal{H}$:

$$||h||_{\mathcal{B}_X} \leq \frac{1}{||A||_{op}(1-\alpha)} b \implies ||h||_{\mathcal{B}_X} \leq C||A[h]||_{\mathcal{B}_Z}$$

Let R be the closed ball in $\mathcal{B}_X$ of radius $\frac{1}{||A||_{op}(1-\alpha)}b$. The intersection of a closed set and a compact set is also compact and so $R \cap \mathcal{H}$ is compact. The intersection of two convex sets is also convex and the intersection of two symmetric sets is symmetric and so $R \cap \mathcal{H}$ is convex and symmetric.

Let $\tilde{\mathcal{H}}$ be the cone of $R \cap \mathcal{H}$, that is the set defined by:

$$\tilde{\mathcal{H}} = [\gamma h : h \in R \cap \mathcal{H}, \gamma \in \mathbb{R}_+]$$

We have already shown that for any $h \in R \cap \mathcal{H}$, $||h||_{\mathcal{B}_X} \leq C||A[h]||_{\mathcal{B}_Z}$. By linearity of A and properties of norms, for any $h \in \tilde{\mathcal{H}}$ we have $||h||_{\mathcal{B}_X} \leq C||A[h]||_{\mathcal{B}_Z}$. So A^{-1} is bounded on $A[\tilde{\mathcal{H}}]$.

Now, $R \cap \mathcal{H}$ is convex and symmetric which implies $\tilde{\mathcal{H}}$ is a linear space. Because $\mathcal{H}$ is infinite-dimensional, convex, and contains zero, the space $\tilde{\mathcal{H}}$ is infinite-dimensional. And because $R \cap \mathcal{H}$ is compact, $\tilde{\mathcal{H}}$ is a closed subset of $\mathcal{B}_X$ and therefore complete in the norm $|| \cdot ||_{\mathcal{B}_X}$. In other words, $\tilde{\mathcal{H}}$ is an infinite-dimensional Banach space with the norm $|| \cdot ||_{\mathcal{B}_X}$.

But the inverse of a compact injective operator on an infinite-dimensional Banach space cannot be bounded. By Assumptions 1.1 and 1.2 the operator A is compact and injective, and so we have a contradiction.

Q.E.D.

PROOF OF THEOREM 2.3: First a. It is well-known that a bounded linear operator on a finite-dimensional linear space has a bounded inverse, so $A_{\mathcal{H}}^{-1}$ is bounded and linear. P_Z is bounded and linear and the composition of two bounded linear operators is a bounded linear operator, and so there exists a constant C so that for any $g \in \mathcal{B}_Z$, $A_{\mathcal{H}}^{-1} P_Z[g] \leq C||g||_{\mathcal{B}_Z}$. And so by linearity:

$$\sup_{g_0 \in R(A):\, ||g_0 - g|| \leq b} ||A_{\mathcal{H}}^{-1} P_Z[g_0] - A_{\mathcal{H}}^{-1} P_Z[g]||_{\mathcal{B}_X} \leq Cb$$

Set $g = A[h_0]$, since $h_0 \in \mathcal{H}$ we have $A_{\mathcal{H}}^{-1} P_Z[A[h_0]] = h_0$. Since $R(A)$ is a linear space, for any $g_0 \in R(A)$ there exists $u_0 \in R(A)$ so that $g_0 = u_0 + A[h_0]$. So we get:

$$\sup_{u_0 \in R(A):\, ||u_0|| \leq b} ||A_{\mathcal{H}}^{-1} P_Z[g_0] - h_0||_{\mathcal{B}_X} \leq Cb$$

And hence:

$$\frac{\sup_{u_0 \in R(A):\, ||u_0|| \leq b} ||A_{\mathcal{H}}^{-1} P_Z[g_0] - h_0||_{\mathcal{B}_X}}{b} = C$$

By the triangle inequality:

$$||\hat{h}_n - h_0||_{\mathcal{B}_X} \leq ||A_{\mathcal{H}}^{-1} P_Z[g_0] - h_0||_{\mathcal{B}_X} + ||\hat{h}_n - A_{\mathcal{H}}^{-1} P_Z[g_0]||_{\mathcal{B}_X}$$

Since $||\hat{h}_n - A_{\mathcal{H}}^{-1} P_Z[g_0]||_{\mathcal{B}_X} \to^p 0$ it follows that:

$$\text{plim}_{n\to\infty} ||\hat{h}_n - h_0||_{\mathcal{B}_X} = ||A_{\mathcal{H}}^{-1} P_Z[g_0] - h_0||_{\mathcal{B}_X}$$

And so we have:

$$\frac{bias_{\hat{h}_n}(b)}{b} = \frac{\sup_{u_0 \in R(A):\, ||u_0|| \leq b} \text{plim}_{n\to\infty} ||\hat{h}_n - h_0||_{\mathcal{B}_X}}{b} \leq C$$

b. Follows immediately from Lemma 2.1 and Lemma 2.3.

Q.E.D.

PROOF OF THEOREM 2.4: First we prove the following:

$$bias_{\hat{\gamma}_n}(b) = \sup_{u_0 \in R(A):\, ||u_0||_{L_2(\mu_Z)} \leq b} |E[w(X)A^{-1}[u_0](X)]| \tag{B.2}$$

By consistency under instrumental validity, for any $g_0 \in R(A)$ we must have:

$$|\hat{\gamma}_n - E[w(X)A^{-1}[g_0](X)]| \to^p 0$$

If the above did not hold for some $g_0 \in R(A)$ then $\hat{\gamma}_n$ would be inconsistent when $u_0 = 0$ and $h_0 = A^{-1}[g_0]$.

Setting $g_0 = A[h_0] + u_0$ for $u_0 \in R(A)$ we have:

$$|E[w(X)A^{-1}[g_0](X)] - \gamma_0| = |E[w(X)A^{-1}[u_0](X)]|$$

And so:

$$\underset{n\to\infty}{\text{plim}} |\hat{\gamma}_n - \gamma_0| = |E[w(X)A^{-1}[u_0](X)]|$$

Applying the definition of the worst-case asymptotic bias then gives (B.2).

Now let us introduce some convenient notation. The $L_2(\mu_X)$ and $L_2(\mu_Z)$ inner products are denoted $\langle \cdot, \cdot \rangle_{L_2(\mu_X)}$ and $\langle \cdot, \cdot \rangle_{L_2(\mu_Z)}$. They are defined by:

$$\langle \delta_1, \delta_2 \rangle_{L_2(\mu_X)} = E[\delta_1(X)\delta_2(X)]$$

For $\delta_1, \delta_2 \in L_2(\mu_X)$. And:

$$\langle \delta_1, \delta_2 \rangle_{L_2(\mu_Z)} = E[\delta_1(Z)\delta_2(Z)]$$

For $\delta_1, \delta_2 \in L_2(\mu_Z)$.

The linear functional of interest γ_0 can then be written as $\gamma_0 = \langle w, h_0 \rangle_{L_2(\mu_X)}$.

Define the operator $A^* : L_2(\mu_Z) \to L_2(\mu_X)$ by:

$$A^*[g](X) = E[g(Z)|X]$$

Note that by iterated expectations, for any $g \in L_2(\mu_Z)$ and $h \in L_2(\mu_X)$:

$$E\Big[E[h(X)|Z]g(Z)\Big] = E\Big[h(X)E[g(Z)|X]\Big]$$

And so A^* is the adjoint of the operator A with respect to the $L_2(\mu_Z)$ and $L_2(\mu_X)$ inner products.

Let us now prove a. We will show that:

$$\sup_{\substack{g \in R(A) \\ ||g||_{L_2(\mu_Z)} \le b}} |E[w(X)A^{-1}[g](Z)]| = b \inf_{\substack{\alpha \in L_2(\mu_Z) \\ w(X) = E[\alpha(Z)|X]}} ||\alpha||_{L_2(\mu_Z)}$$

Suppose that for some α, $w(X) = E[\alpha(Z)|X]$, equivalently $w = A^*[\alpha]$. Then for any $g \in R(A)$:

$$\begin{aligned}
\langle w, A^{-1}[g] \rangle_{L_2(\mu_X)} &= \langle A^*[\alpha], A^{-1}[g] \rangle_{L_2(\mu_X)} \\
&= \langle \alpha, AA^{-1}[g] \rangle_{L_2(\mu_Z)} \\
&= \langle \alpha, g \rangle_{L_2(\mu_Z)}
\end{aligned}$$

Therefore, by Cauchy-Schwartz:

$$\begin{aligned}
|\langle w, A^{-1}[g] \rangle_{L_2(\mu_X)}| &= |\langle \alpha, g \rangle_{L_2(\mu_Z)}| \\
&\le ||g||_{L_2(\mu_Z)} ||\alpha||_{L_2(\mu_Z)}
\end{aligned}$$

Since the inequality above holds for any α with $w = A^*[\alpha]$:

$$|E[w(X)A^{-1}[g](X)]| \leq ||g||_{L_2(\mu_Z)} \inf_{\substack{\alpha \in L_2(\mu_Z) \\ w(X) = E[\alpha(Z)|X]}} ||\alpha||_{L_2(\mu_Z)}$$

Since the above holds for any $g \in L_2(\mu_Z)$ we have:

(B.3)
$$\sup_{||g||_{L_2(\mu_Z)} \leq b} |E[w(X)A^{-1}[g](Z)]| \leq b \inf_{\substack{\alpha \in L_2(\mu_Z) \\ w(X) = E[\alpha(Z)|X]}} ||\alpha||_{L_2(\mu_Z)}$$

Now, let $N(A^*)$ denote the null space of the operator A^*, that is:

$$N(A^*) = \Big[\delta \in L_2(\mu_Z) : \; E[\delta(Z)^2] > 0, E\big[E[\delta(Z)|X]^2\big] = 0\Big]$$

$N(A^*)$ is a closed linear subspace of $L_2(\mu_Z)$ and therefore (by, e.g., Theorem 13.3 in Kress (2014)) there exists an orthogonal projection operator P_N that maps from $L_2(\mu_Z)$ to $N(A^*)$.

Note that by definition of the null space, for any $\alpha \in L_2(\mu_Z)$, $A^*P_N[\alpha] = 0$. And so if $E[\alpha(Z)|X] = w(X)$ then $A^*\big[\alpha - P_N[\alpha]\big] = w$.

Further, for any $\alpha' \in L_2(\mu_Z)$ and $\alpha'' \in L_2(\mu_Z)$ with $A^*[\alpha'] = A^*[\alpha''] = w$ we have:

$$\alpha' - P_N[\alpha'] = \alpha'' - P_N[\alpha'']$$

And note that, by orthogonality of the projection $\alpha - P_N[\alpha] \in N(A^*)^\perp$ (where $N(A^*)^\perp$ denotes the orthogonal complement of $N(A^*)$), and so:

$$||\alpha - P_N[\alpha]||_{L_2(\mu_Z)} = ||\alpha||_{L_2(\mu_Z)} - ||P_N[\alpha]||_{L_2(\mu_Z)} \leq ||\alpha||_{L_2(\mu_Z)}$$

It follows that for any $\alpha \in L_2(\mu_Z)$ so that $E[\alpha(Z)|X] = w(X)$:

(B.4) $$||\alpha - P_N[\alpha]||_{L_2(\mu_Z)} = \inf_{\substack{\alpha \in L_2(\mu_Z) \\ w(X) = E[\alpha(Z)|X]}} ||\alpha||_{L_2(\mu_Z)}$$

Note that by Theorem 15.8 of Kress (2014) $N(A^*)^\perp$ is the closure of the range of A, that is $N(A^*)^\perp = \overline{R(A)}$. Therefore:

$$\alpha - P_N[\alpha] \in \overline{R(A)}$$

Let $0 < \epsilon < b$ be strictly positive scalars. Since $\overline{R(A)}$ is a linear space:

$$\frac{b-\epsilon}{||(\alpha - P_N[\alpha])||_{L_2(\mu_Z)}}(\alpha - P_N[\alpha]) \in \overline{R(A)}$$

By definition of the closure, there exists a $g \in R(A)$ so that:

$$||\frac{b-\epsilon}{||(\alpha - P_N[\alpha])||_{L_2(\mu_Z)}}(\alpha - P_N[\alpha]) - g||_{L_2(\mu_Z)} \leq \epsilon \tag{B.5}$$

Note that by the reverse triangle inequality the above implies that:

$$||g||_{L_2(\mu_Z)} \leq b$$

For such a function g, by the bilinearity of the inner product and the triangle inequality:

$$\begin{aligned} &|\langle \alpha - P_N[\alpha], g\rangle_{L_2(\mu_Z)}| \\ \geq &(b-\epsilon)||\alpha - P_N[\alpha]||_{L_2(\mu_Z)} \\ -&|\langle \alpha - P_N[\alpha], \frac{b-\epsilon}{||(\alpha - P_N[\alpha])||_{L_2(\mu_Z)}}(\alpha - P_N[\alpha]) - g\rangle_{L_2(\mu_Z)}| \end{aligned}$$

By Cauchy-Schwartz and (B.5):

$$\begin{aligned} &|\langle \alpha - P_N[\alpha], \frac{b-\epsilon}{||(\alpha - P_N[\alpha])||_{L_2(\mu_Z)}}(\alpha - P_N[\alpha]) - g\rangle_{L_2(\mu_Z)}| \\ \leq &\epsilon||\alpha - P_N[\alpha]||_{L_2(\mu_Z)} \end{aligned}$$

And so:

$$|\langle \alpha - P_N[\alpha], g\rangle_{L_2(\mu_Z)}| \geq (b-2\epsilon)||\alpha - P_N[\alpha]||_{L_2(\mu_Z)}$$

Recall that $g \in R(A)$ with $||g||_{L_2(\mu_Z)} \leq b$, and there exists such a g that satisfies the above for any value of $0 < \epsilon < b$. So we have that:

$$\sup_{\substack{g \in R(A) \\ ||g||_{L_2(\mu_Z)} \leq b}} |\langle \alpha - P_N[\alpha], g \rangle_{L_2(\mu_Z)}| \geq b ||\alpha - P_N[\alpha]||_{L_2(\mu_Z)}$$

Recall that $|\langle \alpha - P_N[\alpha], g \rangle_{L_2(\mu_Z)}| = |\langle w, A^{-1}[g] \rangle_{L_2(\mu_X)}|$ and by (B.4):

$$\sup_{\substack{g \in R(A) \\ ||g||_{L_2(\mu_Z)} \leq b}} |E[w(X)A^{-1}[g](Z)]| \geq b \inf_{\substack{\alpha \in L_2(\mu_Z) \\ w(X) = E[\alpha(Z)|X]}} ||\alpha||_{L_2(\mu_Z)}$$

Combining with (B.3) we can replace the inequality above with equality. Applying (B.2) then gives:

$$bias_{\hat{\gamma}_n}(b) = b \inf_{\substack{\alpha \in L_2(\mu_Z) \\ w(X) = E[\alpha(Z)|X]}} ||\alpha||_{L_2(\mu_Z)}$$

Let us now prove ii.

We will show that if the worst case asymptotic bias is finite for some choice of b then there must be an $\alpha \in L_2(\mu_Z)$ so that $w(X) = E[\alpha(Z)|X]$.

By (B.2):

$$bias_{\hat{\gamma}_n}(b) = \sup_{u_0 \in R(A):\, ||u_0||_{L_2(\mu_Z)} \leq b} |\langle w, A^{-1}[u_0] \rangle_{L_2(\mu_X)}|$$

Suppose that for some $\bar{b} > 0$ the worst-case asymptotic bias is finite. That is, there exists a scalar $c < \infty$ so that:

$$\sup_{u_0 \in R(A):\, ||u_0||_{L_2(\mu_Z)} \leq \bar{b}} |\langle w, A^{-1}[u_0] \rangle_{L_2(\mu_X)}| \leq c$$

By linearity of A^{-1} and bilinearity of the inner-product it follows that for all $b > 0$:

$$\sup_{u_0 \in R(A):\, ||u_0||_{L_2(\mu_Z)} \leq b} |\langle w, A^{-1}[u_0] \rangle_{L_2(\mu_X)}| \leq \frac{c}{\bar{b}} b$$

Note that by linearity:

$$\sup_{u_0 \in R(A):\, ||u_0||_{L_2(\mu_Z)} \leq b} |\langle w, A^{-1}[u_0]\rangle_{L_2(\mu_X)}|$$
$$= \sup_{u_0 \in R(A):\, ||u_0||_{L_2(\mu_Z)} \leq b} \langle w, A^{-1}[u_0]\rangle_{L_2(\mu_X)}$$

Define the function $D : R(A) \to \mathbb{R}$ by $D[g] = \langle w, A^{-1}[g]\rangle_{L_2(\mu_X)}$. Then we have:

$$\sup_{u_0 \in R(A):\, ||u_0||_{L_2(\mu_Z)} \leq b} |D[u_0]| \leq \frac{c}{\bar{b}} b$$

By the Hahn-Banach theorem we can extend D to a bounded linear function $\bar{D}$ defined on the whole space $L_2(\mu_Z)$ which then satisfies:

$$\sup_{u_0 \in L_2(\mu_Z):\, ||u_0||_{L_2(\mu_Z)} \leq b} |\bar{D}[u_0]| \leq \frac{c}{\bar{b}} b$$

And for each $u_0 \in R(A)$, $\bar{D}[u_0] = D[u_0]$.

Since $\bar{D}$ is a bounded linear functional defined on a Hilbert space, by the Riesz representation theorem for Hilbert spaces there must then exist an element $\alpha \in L_2(\mu_Z)$ so that for all $g \in L_2(\mu_Z)$, $\bar{D}[g] = \langle \alpha, g\rangle_{L_2(\mu_Z)}$. And so for any $u_0 \in R(A)$:

$$\begin{aligned} D[u_0] &= \langle \alpha, u_0\rangle_{L_2(\mu_Z)} \\ &= \langle A^*[\alpha], A^{-1}[u_0]\rangle_{L_2(\mu_X)} \\ &= \langle w, A^{-1}[u_0]\rangle_{L_2(\mu_X)} \end{aligned}$$

Where the final equality follows by the definition of D.

Since the equality above holds for all $u_0 \in R(A)$ then for all $h \in L_2(\mu_X)$ we have (using bilinearity of the inner product):

$$\langle A^*[\alpha] - w, h\rangle_{L_2(\mu_X)} = 0$$

But we can set $h = A^*[\alpha] - w$ and the above implies that the norm of $A^*[\alpha] - w$ equals zero and so we have that $A^*[\alpha] - w = 0$. Or equivalently $w(X) = E[\alpha(Z)|X]$.

Q.E.D.

Proofs For Section 3

Throughout we let $|\cdot|_\infty$ denote the essential supremum norm with respect to Z. That is, for any real valued function g defined on $\mathcal{Z}$ the support of Z:

$$|g|_\infty = \inf[y \in \mathbb{R} : P(|g(Z)| \geq y) = 0]$$

PROPOSITION 3.1 *If $\mathbb{T}$ is linear then for any x in the support of X, Θ_x is convex and so Θ_x is an interval. If, in addition, $\mathcal{H}$ is compact, then Θ_x is a closed interval.*

PROOF: The constraints a. and b. are clearly convex and therefore so is Θ. Suppose $\theta' \in \Theta_x$ and $\theta'' \in \Theta_x$, then there exists $h' \in \mathcal{B}_X$ and $h'' \in \mathcal{B}_X$ so that $h'(x) = \theta'$, $h''(x) = \theta''$ and $h' \in \Theta$ and $h'' \in \Theta$. So consider $h''' = \alpha h' + (1-\alpha)h''$ for some $\alpha \in [0,1]$. Because Θ is convex $h''' \in \Theta$ and so by definition $h'''(x) = \alpha\theta' + (1-\alpha)\theta'' \in \Theta_x$. Therefore Θ_x is convex.

Now suppose $\mathcal{H}$ is compact. For a given $x \in \mathcal{X}$, by definition of the supremum there exists a sequence $\{h_k\}_{k=1}^\infty$ in $\mathcal{H}$ so that each h_k satisfies condition **a.** and $h_k(x) \to \bar{\theta}(x)$.

By compactness of $\mathcal{H}$, there must exist a subsequence of $\{h_k\}_{k=1}^\infty$ that converges in the sup-norm to some element $h_\infty \in \mathcal{H}$, note then that continuity of the operator A implies:

$$|g_0 - A[h_\infty]|_\infty = \lim_{k\to\infty} |g_0 - A[h_k]|_\infty \leq b$$

And clearly:

$$h_\infty(x) = \lim_{k\to\infty} h_k(x) = \bar{\theta}(x)$$

So h_∞ achieves the supremum and satisfies conditions **a.** and **b.** Similar reasoning applies for the infimum.

Q.E.D.

In order to derive convergence rates for the estimators $\hat{\bar{\theta}}_n$ and $\hat{\underline{\theta}}_n$ we define intermediate functions $\bar{\theta}^*_n$ and $\underline{\theta}^*_n$, and functions $\bar{\theta}^\circ_n$ and $\underline{\theta}^\circ_n$. We then bound the error in each estimate by the sum of three terms:

$$|\bar{\theta}(x) - \hat{\bar{\theta}}_n(x)| \leq |\bar{\theta}(x) - \bar{\theta}^*_n(x)| + |\bar{\theta}^*_n(x) - \bar{\theta}^\circ_n(x)| + |\bar{\theta}^\circ_n(x) - \hat{\bar{\theta}}_n(x)|$$

And:

$$|\underline{\theta}(x) - \hat{\underline{\theta}}_n(x)| \leq |\underline{\theta}(x) - \underline{\theta}^*_n(x)| + |\underline{\theta}^*_n(x) - \underline{\theta}^\circ_n(x)| + |\underline{\theta}^\circ_n(x) - \hat{\underline{\theta}}_n(x)|$$

Below we restate the definitions of the estimators $\hat{\bar{\theta}}_n(x)$ and $\hat{\underline{\theta}}_n(x)$. Under Assumption 3.1 (which we assume holds in all the subsequent results) the maximum and minimum in the problems below are achieved, let $\hat{\bar{\beta}}_n(x)$ and $\hat{\underline{\beta}}_n(x)$ respectively denote (not necessarily unique) elements that achieve the maximum and minimum.

$$\hat{\bar{\theta}}_n(x) = \max_{\beta \in \mathbb{R}^{K_n}} \Phi_n(x)'\beta$$

subject to constraints A.1 and A.2

$$\hat{\underline{\theta}}_n(x) = \min_{\beta \in \mathbb{R}^{K_n}} \Phi_n(x)'\beta$$

subject to constraints A.1 and A.2

Where the constraints A.1 and A.2 are given below:

A.1 $|\hat{g}_n - \hat{\Pi}_n'\beta|_n \leq b$

A.2 $\mathbb{T}[\Phi_n'](x)\beta \leq c(x), \forall x \in \mathcal{X}_n$

Note that $\hat{\bar{\beta}}_n$ and $\hat{\underline{\beta}}_n$ must exist if Assumption 3.1 holds because the subset of $\mathbb{R}^{K_n}$ that satisfies A.1 and A.2 is clearly closed and under Assumption 3.1 it is also bounded and therefore it is compact.

Next we define $\bar{\theta}^*_n(x)$ and $\underline{\theta}^*_n(x)$, let $\bar{\beta}^*_n(x)$ and $\underline{\beta}^*_n(x)$ respectively denote elements that achieve the maximum and minimum.

$$\bar{\theta}^*_n(x) = \max_{\beta \in \mathbb{R}^{K_n}} \Phi_n(x)'\beta$$

subject to constraints B.1 and B.2

$$\underline{\theta}_n^*(x) = \min_{\beta \in \mathbb{R}^{K_n}} \Phi_n(x)'\beta$$

$$\text{subject to constraints B.1 and B.2}$$

Where constraints B.1 and B.2 are defined below:

B.1 $|g_0 - \Pi_n'\beta|_\infty \leq b$

B.2 $\mathbb{T}[\Phi_n'](x)\beta \leq c(x)$, $\forall x \in \mathcal{X}$

Finally we define $\bar{\theta}_n^\circ(x)$ and $\underline{\theta}_n^\circ(x)$, and let $\bar{\beta}_n^\circ(x)$ and $\underline{\beta}_n^\circ(x)$ respectively denote elements that achieve the maximum and minimum.

$$\bar{\theta}_n^\circ(x) = \max_{\beta \in \mathbb{R}^{K_n}} \Phi_n(x)'\beta$$

$$\text{subject to constraints C.1 and B.2}$$

$$\underline{\theta}_n^\circ(x) = \min_{\beta \in \mathbb{R}^{K_n}} \Phi_n(x)'\beta$$

$$\text{subject to constraints C.1 and B.2}$$

Where the constraint C.1 is given below:

C.1 $|\hat{g}_n - \hat{\Pi}_n'\beta|_\infty \leq b$

LEMMA 3.1 *Let $\mathbb{T}_1$ be a linear operator from $\mathcal{B}_X$ to $\mathcal{B}_Z$ and let $\mathbb{T}_2$ be a linear operator from $\mathcal{B}_X$ to $\mathcal{B}_X$. Suppose $\bar{y}$ solves:*

$$\bar{y} = \sup_{\beta \in \mathbb{R}^{K_n}} \Phi_n(x)'\beta$$

$$s.t.$$

$$|\mathbb{T}_1[\Phi_n'\beta]|_\infty \leq b_1$$

$$\mathbb{T}_2[\Phi_n'\beta](x) \leq b_2(x), \ \forall x \in \mathcal{X}$$

And $\underline{y}$ solves the corresponding minimization problem.

Suppose there exists a $\tilde{\beta} \in \mathbb{R}^{K_n}$, $\epsilon > 0$ so that:

$$|\mathbb{T}_1[\Phi_n'\tilde{\beta}]|_\infty \leq b_1 - \epsilon$$

$$\mathbb{T}_2[\Phi_n'\tilde{\beta}](x) \leq b_2(x) - \epsilon, \, \forall x \in \mathcal{X}$$

And for a given $\hat{\beta}$ *and* $r \geq 0$*:*

$$|\mathbb{T}_1[\Phi_n'\hat{\beta}]|_\infty \leq b_1 + r$$

$$\mathbb{T}_2[\Phi_n'\hat{\beta}](x) \leq b_2(x) + r, \, \forall x \in \mathcal{X}$$

Then:

$$\Phi_n(x)'\hat{\beta} - \bar{y} \leq \frac{r}{r+\epsilon}\Phi_n(x)'[\hat{\beta} - \tilde{\beta}]$$

$$\underline{y} - \Phi_n(x)'\hat{\beta} \leq \frac{r}{r+\epsilon}\Phi_n(x)'[\tilde{\beta} - \hat{\beta}]$$

PROOF: Let $\beta = (1 - \frac{r}{r+\epsilon})\hat{\beta} + \frac{r}{r+\epsilon}\tilde{\beta}$. Note that by the linearity of $\mathbb{T}_1$ and the triangle inequality:

$$\begin{aligned}
|\mathbb{T}_1[\Phi_n'\beta]|_\infty &\leq (1 - \frac{r}{r+\epsilon})|\mathbb{T}_1[\Phi_n'\hat{\beta}]|_\infty + \frac{r}{r+\epsilon}|\mathbb{T}_1[\Phi_n'\tilde{\beta}]|_\infty \\
&\leq (1 - \frac{r}{r+\epsilon})(b_1 + r) + \frac{r}{r+\epsilon}(b_1 - \epsilon) \\
&= b_1
\end{aligned}$$

And following analogous steps:

$$\mathbb{T}_2[\Phi_n'\beta](x) \leq b_2(x), \, \forall x \in \mathcal{X}$$

So β satisfies the constraints in the problems for $\bar{y}$ and $\underline{y}$, so we must have:

$$\underline{y} \leq \Phi_n(x)'\beta \leq \overline{y}$$

Substituting the definition of β into the above gives:

$$\underline{y} \leq \Phi_n(x)'\hat{\beta} - \frac{r}{r+\epsilon}\Phi_n(x)'[\hat{\beta} - \tilde{\beta}] \leq \overline{y}$$

And so we get:

$$\Phi_n(x)'\hat{\beta} - \bar{y} \leq \frac{r}{r+\epsilon}\Phi_n(x)'[\hat{\beta} - \tilde{\beta}]$$

$$\underline{y} - \Phi_n(x)'\hat{\beta} \leq \frac{r}{r+\epsilon}\Phi_n(x)'[\tilde{\beta} - \hat{\beta}]$$

Q.E.D.

Lemma 3.2 *Suppose Assumptions 3.1, 3.2 and 3.3 hold. Suppose that there is an $h \in \mathcal{H}$ with $|g_0 - A[h]|_\infty < b$. Then there is a sequence $\epsilon_n \to \epsilon > 0$ so that for n sufficiently large there exists $\tilde{\beta}_n \in \mathbb{R}^{K_n}$ so that:*

$$|g_0 - \Pi_n' \tilde{\beta}_n|_\infty \leq b - \epsilon_n$$
$$\mathbb{T}[\Phi_n'] \tilde{\beta}_n(x) \leq c(x) - \epsilon_n, \, \forall x \in \mathcal{X}$$

And with probability approaching 1*:*

$$|\hat{g}_n - \hat{\Pi}_n' \tilde{\beta}_n|_\infty \leq b - \epsilon_n$$

Proof: By presumption there exists some $h_* \in \mathcal{H}$ and $|g_0 - A[h_*]|_\infty < b$. Therefore there is an $\epsilon_* > 0$ so that (using the definition of $\mathcal{H}$):

$$|g_0 - A[h_*]|_\infty \leq b - \epsilon_*$$
$$\mathbb{T}[h_*](x) \leq c(x) \, \forall x \in \mathcal{X}$$

Therefore, using linearity of $\mathbb{T}$ (from Assumption 3.1) and A, and using the triangle inequality, for any $\alpha \in (0, 1)$:

$$|g_0 - A[\alpha h_*]|_\infty \leq b - \epsilon_* + (1 - \alpha)|A[h_*]|_\infty$$
$$\mathbb{T}[\alpha h_*](x) \leq \alpha c(x) \leq c(x) - (1 - \alpha)\underline{c}, \, \forall x \in \mathcal{X}$$

Where we have also used that c is bounded below by $\underline{c} > 0$ from Assumption 3.1.

So setting α sufficiently close to 1 we see that for some $\epsilon > 0$ and $\tilde{h} = \alpha h_*$:

$$|g_0 - A[\tilde{h}]|_\infty \leq b - \epsilon$$
$$\mathbb{T}[\tilde{h}](x) \leq c(x) - \epsilon, \, \forall x \in \mathcal{X}$$

For some sufficiently large constant $C > 1$ (that does not depend on n) define ϵ_n by $\epsilon_n = \epsilon - \kappa_n - C a_n$. κ_n and a_n the sequences in Assumptions 3.2 and 3.3. Note that $\epsilon_n \to \epsilon$.

By Assumption 3.3 there is a $\tilde{\beta}_n$ with:

$$\mathbb{T}[\Phi_n'](x)\tilde{\beta}_n \leq c(x) - \epsilon \leq c(x) - \epsilon_n, \, \forall x \in \mathcal{X}$$

And $|\tilde{h} - \Phi_n'\tilde{\beta}_n|_\infty \leq \kappa_n$.

Since A has operator norm of unity and $|g_0 - A[\tilde{h}]|_\infty \leq b - \epsilon$, by the triangle inequality the above implies that for n sufficiently large:

$$|g_0 - \Pi_n'\tilde{\beta}_n|_\infty \leq b - (\epsilon - \kappa_n) \leq b - \epsilon_n$$

Which is the first statement in the Lemma. By the triangle inequality:

$$\begin{aligned}
&\sup_{\beta \in \mathbb{R}^{K_n}:\, \Phi_n'\beta \in \mathcal{H}} \Big| |\hat{g}_n - \hat{\Pi}_n'\beta|_\infty - |g_0 - \Pi_n'\beta|_\infty \Big| \\
\leq &\sup_{\beta \in \mathbb{R}^{K_n}:\, \Phi_n'\beta \in \mathcal{H}} \Big| |(\hat{g}_n - \hat{\Pi}_n'\beta) - (g_0 - \Pi_n'\beta)|_\infty \Big| \\
\leq &|\hat{g}_n - g_0|_\infty + \sup_{\beta \in \mathbb{R}^{K_n}:\, \Phi_n'\beta \in \mathcal{H}} |(\hat{\Pi}_n - \Pi_n)'\beta|_\infty \\
= &O_p(a_n)
\end{aligned}$$

Where the final equality follows by Assumption 3.2. And so, since $\Phi_n'\tilde{\beta}_n \in \mathcal{H}$

$$\begin{aligned}
|\hat{g}_n - \hat{\Pi}_n'\tilde{\beta}_n|_\infty &\leq |g_0 - \Pi_n'\tilde{\beta}_n|_\infty + O_p(a_n) \\
&\leq b - (\epsilon - \kappa_n) + O_p(a_n)
\end{aligned}$$

And so with probability approaching 1, if C is sufficiently large:

$$|\hat{g}_n - \hat{\Pi}_n'\tilde{\beta}_n|_\infty \leq b - \epsilon + \kappa_n + Ca_n = b - \epsilon_n$$

Q.E.D.

PROPOSITION 3.2 *Suppose Assumptions 3.1, 3.2 and 3.3 hold and there is an $h \in \mathcal{H}$ with $|g_0 - A[h]|_\infty < b$. Then:*

$$|\bar{\theta}(x) - \bar{\theta}_n^*(x)| = O(\kappa_n) = o(1)$$

$$|\underline{\theta}(x) - \underline{\theta}_n^*(x)| = O(\kappa_n) = o(1)$$

PROOF: First note that the objective and constraints in the optimization problems that define $\bar{\theta}^*_n(x)$ and $\underline{\theta}^*_n(x)$ are identical to those for $\bar{\theta}(x)$ and $\underline{\theta}(x)$ respectively. However, the space of functions over which we optimize for $\bar{\theta}^*_n(x)$ and $\underline{\theta}^*_n(x)$ is restricted to linear combinations of the components of Φ_n. Therefore:

$$\bar{\theta}(x) - \bar{\theta}^*_n(x) \geq 0$$

$$\underline{\theta}^*_n(x) - \underline{\theta}(x) \geq 0$$

Note that Assumption 3.1 implies $\bar{\theta}(x)$ and $\underline{\theta}(x)$ are finite. For a given $\delta > 0$ let each $x \in \mathcal{X}$ let $\bar{h}_x$ and $\underline{h}_x$ be functions that satisfy constraints **a.** and **b.** of the infeasible problem with $\bar{h}_x(x) \geq \bar{\theta}(x) - \delta$ and $\underline{h}_x(x) \leq \underline{\theta}(x) + \delta$. Such functions must exist by definition of the supremum and infimum.

Let $h_x \in \{\bar{h}_x, \underline{h}_x\}$. Since h_x satisfies constraint **b.**, by Assumption 3.3 for n sufficiently large there must exist β_n so that:

$$\mathbb{T}[\Phi_n'](x)\beta_n \leq c(x), \, \forall x \in \mathcal{X}$$

And $|\Phi_n'\beta_n - h_x|_\infty \leq \kappa_n$.

In which case, because operator A has operator norm of unity:

$$|g_0 - \Pi_n'\beta_n|_\infty \leq b + \kappa_n$$

By presumption there is an $h \in \mathcal{H}$ with $|g_0 - A[h]|_\infty < b$ so by Lemma 3.2 there must exist $\tilde{\beta}_n \in \mathbb{R}^{K_n}$ and $\epsilon_n \to \epsilon > 0$ so that:

$$|g_0 - \Pi_n'\tilde{\beta}_n|_\infty \leq b - \epsilon_n$$
$$\mathbb{T}[\Phi_n'](x)\tilde{\beta}_n \leq c(x) - \epsilon_n, \, \forall x \in \mathcal{X}$$

Now, $\bar{\beta}^*_n(x)$ and $\underline{\beta}^*_n(x)$ are optima for the problem with constraints B.1 and B.2, so by Lemma 3.1:

$$\begin{aligned}\Phi_n(x)'\beta_n - \Phi_n(x)'\bar{\beta}^*_n(x) &\leq \frac{\kappa_n}{\kappa_n + \epsilon_n}[\Phi_n(x)'\beta_n - \Phi_n(x)'\tilde{\beta}_n] \\ &\leq \frac{\kappa_n}{\kappa_n + \epsilon_n} 2\bar{c}\end{aligned}$$

Where we have used that β_n and $\tilde{\beta}_n$ satisfy constraint B.2 and so $\Phi_n'\beta_n$ and $\Phi_n'\tilde{\beta}_n$ are uniformly bounded by $\bar{c}$ from Assumption 3.1.

And furthermore:

$$\begin{aligned}\Phi_n(x)'\underline{\beta}_n^*(x) - \Phi_n(x)'\beta_n &\leq \frac{\kappa_n}{\kappa_n + \epsilon_n}[\Phi_n(x)'\tilde{\beta}_n - \Phi_n(x)'\beta_n] \\ &\leq \frac{\kappa_n}{\kappa_n + \epsilon_n} 2\bar{c}\end{aligned}$$

So using $|\Phi_n'\beta_n - h_x|_\infty \leq \kappa_n$ we get (recalling $\kappa_n \geq 0$), in the case of $h_x = \bar{h}_x$:

$$\bar{h}_x(x) - \Phi_n(x)'\bar{\beta}_n^*(x) \leq \kappa_n + \frac{\kappa_n}{\kappa_n + \epsilon_n} 2\bar{c}$$

And if $h_x = \underline{h}_x$:

$$\Phi_n(x)'\underline{\beta}_n^*(x) - \underline{h}_x(x) \leq \kappa_n + \frac{\kappa_n}{\kappa_n + \epsilon_n} 2\bar{c}$$

Recall that $\bar{h}_x(x) \geq \bar{\theta}(x) - \delta$ and $\underline{h}_x(x) \leq \underline{\theta}(x) + \delta$ and that $\Phi_n(x)'\bar{\beta}_n^*(x) = \bar{\theta}_n^*(x)$ and $\Phi_n(x)'\underline{\beta}_n^*(x) = \underline{\theta}_n^*(x)$. Since (as we showed at the beginning of the proof) the expressions on the LHSs above are positive:

$$|\bar{\theta}(x) - \bar{\theta}_n^*(x)| \leq \kappa_n + \frac{\kappa_n}{\kappa_n + \epsilon_n} 2\bar{c} + \delta$$

Since the above holds for arbitrary $\delta > 0$ we have:

$$\begin{aligned}|\bar{\theta}(x) - \bar{\theta}_n^*(x)| &\leq \kappa_n + \frac{\kappa_n}{\kappa_n + \epsilon_n} 2\bar{c} \\ &= O(\kappa_n)\end{aligned}$$

And by similar reasoning:

$$\begin{aligned}|\underline{\theta}(x) - \underline{\theta}_n^*(x)| &\leq \kappa_n + \frac{\kappa_n}{\kappa_n + \epsilon_n} 2\bar{c} \\ &= O(\kappa_n)\end{aligned}$$

Since the $\kappa_n + \frac{\kappa_n}{\kappa_n+\epsilon_n} 2\bar{c}$ does not depend on x the inequalities and hence convergence applies uniformly over $x \in \mathcal{X}$. Finally, noting that $\kappa_n = o(1)$ we get the desired result.

Q.E.D.

Proposition 3.3 *Under Assumptions 3.1, 3.2 and 3.3, if there is an $h \in \mathcal{H}$ with $|g_0 - A[h]|_\infty < b$ then:*

$$|\bar{\theta}_n^* - \bar{\theta}_n^\circ|_\infty = O_p(a_n) = o_p(1)$$
$$|\underline{\theta}_n^* - \underline{\theta}_n^\circ|_\infty = O_p(a_n) = o_p(1)$$

Proof: Define $\hat{r}_n$ by:

$$\hat{r}_n = \sup_{\beta \in \mathbb{R}^{K_n} : \Phi_n' \beta \in \mathcal{H}} \Big| |\hat{g}_n - \hat{\Pi}_n' \beta|_\infty - |g_0 - \Pi_n' \beta|_\infty \Big|$$

By the triangle inequality:

$$\begin{aligned}
\hat{r}_n \leq & \sup_{\beta \in \mathbb{R}^{K_n} : \Phi_n' \beta \in \mathcal{H}} \Big| |(\hat{g}_n - \hat{\Pi}_n' \beta) - (g_0 - \Pi_n' \beta)|_\infty \Big| \\
\leq & |\hat{g}_n - g_0|_\infty + \sup_{\beta \in \mathbb{R}^{K_n} : \Phi_n' \beta \in \mathcal{H}} |(\hat{\Pi}_n - \Pi_n)' \beta|_\infty \\
= & O_p(a_n)
\end{aligned}$$

Where the final equality follows by Assumption 3.2.

Let $\beta_{n,x}^*$ equal either $\bar{\beta}_n^*(x)$ or $\underline{\beta}_n^*(x)$. By constraint B.1 and the triangle inequality we must have:

$$|\hat{g}_n - \hat{\Pi}_n' \beta_{n,x}^*|_\infty \leq b + \hat{r}_n$$

By Lemma 3.2 there is a sequence $\epsilon_n \to \epsilon$ with $\epsilon > 0$ so that with probability approaching 1 there exists $\tilde{\beta}_n \in \mathbb{R}^{K_n}$ so that the following three inequalities hold:

$$|g_0 - \Pi_n' \tilde{\beta}_n|_\infty \leq b - \epsilon_n$$
$$|\hat{g}_n - \hat{\Pi}_n' \tilde{\beta}_n|_\infty \leq b - \epsilon_n$$
$$\mathbb{T}[\Phi_n'](x) \tilde{\beta}_n \leq c(x) - \epsilon_n, \ \forall x \in \mathcal{X}$$

So applying Lemma 3.1 we get that with probability approaching 1 for all

$x \in \mathcal{X}$:

$$\Phi_n(x)'\bar{\beta}_n^*(x) - \Phi_n(x)'\bar{\beta}_n^\circ(x) \leq \frac{\hat{r}_n}{\hat{r}_n + \epsilon_n}[\Phi_n(x)'\bar{\beta}_n^*(x) - \Phi_n(x)'\tilde{\beta}_n] \\ \leq \frac{\hat{r}_n}{\hat{r}_n + \epsilon_n}2\bar{c}$$

And similarly:

$$\Phi_n(x)'\underline{\beta}_n^\circ(x) - \Phi_n(x)'\underline{\beta}_n^*(x) \leq \frac{\hat{r}_n}{\hat{r}_n + \epsilon_n}2\bar{c}$$

Conversely, let $\beta_{n,x}^\circ$ equal either $\bar{\beta}_n^\circ(x)$ or $\underline{\beta}_n^\circ(x)$. By constraint C.1 and the triangle inequality we must have:

$$|g_0 - \Pi_n'\beta_{n,x}^\circ|_\infty \leq b + \hat{r}_n$$

Similarly, if $\beta_{n,x}^* = \underline{\beta}_n^*(x)$:

$$\underline{\theta}_n^\circ(x) - \underline{\theta}_n^*(x) \leq \frac{\hat{r}_n}{\hat{r}_n + \epsilon_n}[\tilde{\theta}_n(x) - \underline{\theta}_n^*(x)] \\ \leq 2\bar{c}\frac{\hat{r}_n}{\hat{r}_n + \epsilon_n}$$

So applying Lemma 3.1 we get that with probability approaching 1 for all $x \in \mathcal{X}$:

$$\Phi_n(x)'\bar{\beta}_n^\circ(x) - \Phi_n(x)'\bar{\beta}_n^*(x) \leq \frac{\hat{r}_n}{\hat{r}_n + \epsilon_n}[\Phi_n(x)'\beta_{n,x}^\circ - \Phi_n(x)'\tilde{\beta}_n] \\ \leq \frac{\hat{r}_n}{\hat{r}_n + \epsilon_n}2\bar{c}$$

And:

$$\Phi_n(x)'\underline{\beta}_n^*(x) - \Phi_n(x)'\underline{\beta}_n^\circ(x) \leq \frac{\hat{r}_n}{\hat{r}_n + \epsilon_n}2\bar{c}$$

Note that the final LHSs of the inequalities above do not depend on x.

Combining, and noting that $\Phi_n(x)'\bar{\beta}_n^\circ(x) = \bar{\theta}_n^\circ$, $\Phi_n(x)'\underline{\beta}_n^\circ(x) = \underline{\theta}_n^\circ$, $\Phi_n(x)'\bar{\beta}_n^*(x) = \bar{\theta}_n^*$ and $\Phi_n(x)'\underline{\beta}_n^*(x) = \underline{\theta}_n^*$ we get with probability approaching 1:

$$|\bar{\theta}_n^\circ - \bar{\theta}_n^*|_\infty \leq \frac{\hat{r}_n}{\hat{r}_n + \epsilon_n}2\bar{c} = O_p(a_n)$$

And:

$$|\underline{\theta}_n^* - \underline{\theta}_n^\circ|_\infty \leq \frac{\hat{r}_n}{\hat{r}_n + \epsilon_n} 2\bar{c} = O_p(a_n)$$

Q.E.D.

PROPOSITION 3.4 *Under Assumptions 3.1 and 3.4 if there is an $h \in \mathcal{H}$ with $|g_0 - A[h]|_\infty < b$ then:*

$$|\bar{\theta}_n^\circ - \hat{\bar{\theta}}_n|_\infty = O_p(C_{K_n} G_n D_{2,n} + C_{K_n} \xi_{K_n} D_{1,n}) = o_p(1)$$
$$|\underline{\theta}_n^\circ - \hat{\underline{\theta}}_n|_\infty = O_p(C_{K_n} G_n D_{2,n} + C_{K_n} \xi_{K_n} D_{1,n}) = o_p(1)$$

PROOF: First note that constraints C.1 and B.2 are weaker than A.1 and A.2 and so:

$$\hat{\bar{\theta}}_n(x) - \bar{\theta}_n^\circ(x) \geq 0$$
$$\underline{\theta}_n^\circ(x) - \hat{\underline{\theta}}_n(x) \geq 0$$

By Assumption 3.4.i Φ_n is Lipschitz continuous with constant ξ_n, and so the function $\Phi_n'\beta$ is Lipschitz continuous with constant at most $||\beta||_2\xi_n$. Therefore:

$$|\Phi_n'\beta|_\infty - |\Phi_n'\beta|_n \leq ||\beta||_2 \xi_n D_{1,n}$$

It follows that:

$$|\Phi_n'\beta|_n \leq \bar{c} \implies |\Phi_n'\beta|_\infty \leq \bar{c} + ||\beta||_2 \xi_n D_{1,n}$$

By the definition of C_n:

$$|\Phi_n'\beta|_\infty \leq \bar{c} + ||\beta||_2 \xi_n D_{1,n} \implies ||\beta||_2 \leq C_n(\bar{c} + ||\beta||_2 \xi_n D_{1,n})$$

Assumption 3.1 implies that if β satisfies constraint A.2 then $|\Phi_n'\beta|_n \leq \bar{c}$. Then if n is sufficiently large so that $C_n\xi_n D_{1,n} < 1$ (recall Assumption 3.4.iii states $C_n\xi_n D_{1,n} \to 0$) we must have (for any β that satisfies constraint A.2):

$$||\beta||_2 \leq \frac{C_n}{1 - C_n \xi_n D_{1,n}} \bar{c}$$

Similarly, by Assumption 3.4.i, the function $\mathbb{T}[\Phi_n']$ is Lipschitz with constant at most ξ_n and c with some constant we will denote by ξ^c and so:

$$\begin{aligned}&\mathbb{T}[\Phi_n'](x)\beta \leq c(x),\, \forall x \in \mathcal{X}_n \\ &\qquad \implies \mathbb{T}[\Phi_n'](x)\beta \leq c(x) + (\xi^c + ||\beta||_2 \xi_n) D_{1,n},\, \forall x \in \mathcal{X}\end{aligned}$$

So for any β that satisfies constraint A.2 and sufficiently high n:

$$\mathbb{T}[\Phi_n'](x)\beta \leq c(x) + \frac{C_n \xi_n D_{1,n}}{1 - C_n \xi_n D_{1,n}} \bar{c} + \xi^c D_{1,n},\, \forall x \in \mathcal{X}$$

By Assumption 3.4.ii with probability approaching 1 $\hat{g}_n$ and $\hat{\Pi}_n$ are both Lipschitz continuous with Lipschitz constant G_n. Then $\hat{g}_n - \hat{\Pi}_n'\beta$ has Lipschitz constant at most $(1 + ||\beta||_2) G_n$ and so by similar reasoning to the above, for any β that satisfies constraints A.1 and A.3:

$$|\hat{g}_n - \hat{\Pi}_n'\beta|_\infty \leq b + \left[1 + \frac{C_n}{1 - C_n \xi_n D_{1,n}}\right] G_n D_{2,n} b$$

So define S_n by:

$$\begin{aligned}S_n &= \left[1 + \frac{C_n}{1 - C_n \xi_n D_{1,n}}\right] G_n D_{2,n} b + \frac{C_n \xi_n D_{1,n}}{1 - C_n \xi_n D_{1,n}} \bar{c} + \xi^c D_{1,n} \\ &= O_p\big(C_n (G_n D_{2,n} + \xi_n D_{1,n})\big)\end{aligned}$$

Then with probability approaching 1, any β that satisfies constraints A.1, A.2 and A.3 satisfies:

$$\begin{aligned}|\hat{g}_n - \hat{\Pi}_n'\beta|_\infty &\leq b + S_n \\ \max_{||\lambda||_1 = m} |D_\lambda \Phi_n'\beta|_\infty &\leq d + S_n \\ |\Phi_n'\beta|_\infty &\leq c + S_n\end{aligned}$$

And so, in particular the inequalities above are satisfied by $\hat{\bar{\beta}}_n(x)$ and $\hat{\underline{\beta}}_n(x)$ for all $x \in \mathcal{X}$.

By Lemma 3.2 there is a sequence $\epsilon_n \to \epsilon$ with $\epsilon > 0$ so that with probability approaching 1 there exists $\tilde{\beta}_n \in \mathbb{R}^{K_n}$ so that:

$$
\begin{aligned}
&|\hat{g}_n - \hat{\Pi}_n' \tilde{\beta}_n|_\infty \leq b - \epsilon_n \\
&\quad \mathbb{T}[\Phi_n'](x)\tilde{\beta}_n \leq c(x) - \epsilon_n, \, \forall x \in \mathcal{X}
\end{aligned}
$$

Recall $\bar{\beta}_n^\circ$ and $\underline{\beta}_n^\circ$ satisfy constraints C.1 and B.2. So by Lemma 3.1, with probability approaching 1 for all $x \in \mathcal{X}$:

$$
\begin{aligned}
\Phi_n(x)'\hat{\bar{\beta}}_n(x) - \Phi_n(x)'\bar{\beta}_n^\circ(x) &\leq \frac{S_n}{\epsilon_n + S_n}[\Phi_n(x)'\hat{\bar{\beta}}_n(x) - \Phi_n(x)'\tilde{\beta}_n] \\
&\leq \frac{S_n}{\epsilon_n + S_n} 2\bar{c}
\end{aligned}
$$

And similarly:

$$
\Phi_n(x)'\underline{\beta}_n^\circ(x) - \Phi_n(x)'\hat{\underline{\beta}}_n(x) \leq \frac{S_n}{\epsilon_n + S_n} 2\bar{c}
$$

Recall $\Phi_n(x)'\bar{\beta}_n^\circ(x) = \bar{\theta}_n^\circ(x)$, $\Phi_n(x)'\underline{\beta}_n^\circ(x) = \underline{\theta}_n^\circ(x)$, $\Phi_n(x)'\hat{\bar{\beta}}_n(x) = \hat{\bar{\theta}}_n(x)$ and $\Phi_n(x)'\hat{\underline{\beta}}_n(x) = \hat{\underline{\theta}}_n(x)$, and we already showed (at the beginning of the proof) the the LHSs in the inequalities above are negative. So we get that with probability approaching 1:

$$
\begin{aligned}
|\bar{\theta}_n^\circ - \hat{\bar{\theta}}_n|_\infty &\leq \frac{S_n}{\epsilon_n + S_n} 2\bar{c} \\
&= O_p(C_n G_n D_{2,n} + C_n \xi_n D_{1,n})
\end{aligned}
$$

And similarly:

$$
|\underline{\theta}_n^\circ - \hat{\underline{\theta}}_n|_\infty = O_p(C_n G_n D_{2,n} + C_n \xi_n D_{1,n})
$$

Finally Assumption 3.4.iii gives that $C_n G_n D_{2,n} + C_n \xi_n D_{1,n} = o(1)$ and so we get the result.

Q.E.D.

Proof of Theorem 3.1: Follows immediately from Propositions 3.2, 3.3 and 3.4.

Q.E.D.

Define $\hat{Q}_n$ by:

$$\hat{Q}_n = \frac{1}{n}\sum_{i=1}^{n}\Psi_n(Z_i)\Psi_n(Z_i)'$$

For each $h \in \mathcal{H}$, let $\hat{\gamma}[h]$ be the least squares estimator defined by:

$$\hat{\gamma}[h] = \hat{Q}_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\Psi_n(Z_i)h(X_i)$$

If $\hat{Q}_n$ is singular we take $\hat{\gamma}[h]$ to be zero (under Assumption 3.5.i this event happens with probability approaching zero).

LEMMA 3.3 *Suppose Assumptions 3.5 and 3.6 hold. Then:*

$$\begin{aligned}\sup_{h\in\mathcal{H}}|\Psi_n'\hat{\gamma}[h] - A[h]|_\infty &= O_p\left(\frac{\bar{\xi}_n}{\sqrt{n}}[1+\sqrt{log(\ell_n)}+\sqrt{L_n}R_n(s)] + R_n(s)\right)\\ &= o_p(1)\end{aligned}$$

PROOF: The proof follows some steps of Belloni *et al.* (2015) Lemma 4.2 with alterations to achieve uniformity over $\mathcal{H}$.

Recall the matrix $Q_n = E[\Psi_n(Z_i)\Psi_n(Z_i)']$. Assumption 3.5.i states that the eigenvalues of Q_n are uniformly bounded above and away from zero and so we can normalize:

$$E[\Psi_n(Z)\Psi_n(Z)'] = I$$

The assumptions then hold for the normalized Ψ_n with sequences $\bar{\xi}_n$ and ℓ_n (that satisfy Assumption 3.6) changed only by a constant positive factor not dependent on n. We maintain this normalization throughout.

Define γ by:

$$\gamma[h] = E[\Psi_n(Z_i)h(X_i)]$$

Let $\epsilon_i[h] = h(X_i) - E[h(X_i)|Z_i]$ and $r_i[h] = E[h(X_i)|Z_i] - \gamma[h]'\Psi_n(Z_i)$. Then we can decompose $h(X_i)$ as:

$$h(X_i) = \Psi_n(Z_i)'\gamma[h] + \epsilon_i[h] + r_i[h]$$

We bound $|\epsilon_i[h]|$. $\mathcal{H}$ contains functions bounded in the supremum norm by $\bar{c}$, so the magnitude of $\epsilon_i[h]$ is bounded uniformly over $\mathcal{H}$:

$$\begin{aligned} |\epsilon_i[h]| &\leq |h(X_i)| + |E[h(X_i)|Z_i]| \\ &\leq 2\bar{c} \end{aligned}$$

Where, as usual the equality is understood to hold with probability 1.

Note as well that:

$$\begin{aligned} |\epsilon_i[h_1] - \epsilon_i[h_2]| &\leq |h_1(X_i) - h_2(X_i)| + |E[h_1(X_i) - h_2(X_i)|Z_i]| \\ &\leq 2|h_1 - h_2|_\infty \end{aligned} \tag{B.6}$$

We also bound $|r_i[h]|$. To do so we first show that if $|h|_\infty < \infty$ then Assumption 3.5.ii implies $A[h]$ is smooth.

For some $m \leq s$ (where s is the smoothness in Assumption 3.5.ii) let $\{q_j\}_{j=1}^{\dim(Z)}$ be a sequence of positive integers with $\sum_{j=1}^{\dim(Z)} q_j = m$. Let D_q then be the partial derivative operator given by $D_q[f](z) = \frac{\partial^m}{\partial^{q_1}\partial^{q_2}\ldots\partial^{q_{\dim(Z)}}} f(z)$ for any sufficiently differentiable function $f : \mathcal{Z} \to \mathbb{R}$. Now, from Assumption 3.5.ii, $D_q[f_{X|Z}(x,\cdot)](z)$ is bounded uniformly over x and z by some constant $\bar{\ell}$. Then it follows by the dominated convergence theorem that:

$$\begin{aligned} |D_q A[h](z)| &= |D_q\Big[\int h(x) f_{X|Z}(x,z)dx\Big](z)| \\ &= |\int h(x) D_q\big[f_{X|Z}(x,\cdot)\big](z)dx| \\ &\leq |h|_\infty \bar{\ell} \end{aligned}$$

And for $m = s$, note that:

$$
\begin{aligned}
&|D_qA[h](z_1) - D_qA[h](z_2)| \\
=&|\int h(x)D_q\big[f_{X|Z}(x,\cdot)\big](z_1)dx - \int h(x)D_q\big[f_{X|Z}(x,\cdot)\big](z_2)dx| \\
\leq&|h|_\infty \sup_{x\in\mathcal{X}} |D_q\big[f_{X|Z}(x,\cdot)\big](z_1) - D_q\big[f_{X|Z}(x,\cdot)\big](z_2)| \\
\leq&|h|_\infty \bar{\ell}||z_1 - z_2||_2^s
\end{aligned}
$$

Where the last inequality again follows because Assumption 3.5.ii states $f_{X|Z}(x,\cdot)$ is of Hölder smoothness class $s$with constant $\bar{\ell}$ for all x. From the above we see that $A[h]$ is of Hölder smoothness class $s$with constant at most $|h|_\infty\bar{\ell}$. So we can apply Assumption 3.6.i and get that:

$$|r_i[h]| \leq |h|_\infty \bar{\ell} R_n(s)$$

Since $|h|_\infty \leq \bar{c}$ for all $h \in \mathcal{H}$ (from Assumption 3.6.iii), we get for any $h \in \mathcal{H}$:

$$|r_i[h]| \leq \bar{c}\bar{\ell} R_n(s)$$

Further, note that by linearity of r_i:

$$|r_i[h_1] - r_i[h_2]| = |r_i[h_1 - h_2]| \leq |h_1 - h_2|_\infty \bar{\ell} R_n(s) \tag{B.7}$$

Using Assumption 3.6.iv we can apply Rudelson's matrix LLN (Rudelson (1999), Belloni *et al.* (2015) Lemma 6.2) to get:

$$
\begin{aligned}
E||\hat{Q}_n - I||_{op} &= O\Big(\sqrt{\frac{\bar{\xi}_n^2 log(L_n)}{n}}\Big) \\
&= o(1)
\end{aligned}
$$

Which implies $||\hat{Q}_n^{-\frac{1}{2}}||_{op} = O_p(1)$, $||\hat{Q}_n^{\frac{1}{2}}||_{op} = O_p(1)$ and $||\hat{Q}_n^{-1}||_{op} = O_p(1)$. And also by Rudelson's LLN , $E||\hat{Q}_n^{\frac{1}{2}}||_{op} = O(1)$.

It follows that $\hat{Q}_n$ is non-singular with probability approaching 1 and so below we treat it as non-singular. Recall we define $\alpha_n : \mathcal{Z} \to \mathbb{R}^{L_n}$ by:

$$\alpha_n(z) = \frac{\Psi_n(z)}{||\Psi_n(z)||}$$

And note that:

$$\text{(B.8)} \quad \sqrt{n}\alpha_n(z)'(\hat{\gamma}[h] - \gamma[h]) = \alpha_n(z)'\hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_n(Z_i)\epsilon_i[h] + \alpha_n(z)'\hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_n(Z_i)r_i[h]$$

Let us bound the first term on the RHS above. Note that:

$$E\left[\alpha_n(z)'\hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_n(Z_i)\epsilon_i[h]\middle|Z_1, ..., Z_n\right] = 0$$

Let $(\eta_1, ..., \eta_n)$ be a sample of iid Rademachers, independent of the data. By the symmetrization inequality:

$$E\left[\sup_{z\in\mathcal{Z},h\in\mathcal{H}}|\alpha_n(z)'\hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_n(Z_i)\epsilon_i[h]|\middle|Z_1, ..., Z_n\right]$$

$$\text{(B.9)} \quad \leq 2E\left[E_\eta\left[\sup_{z\in\mathcal{Z},h\in\mathcal{H}}|\alpha_n(z)'\hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_n(Z_i)\eta_i\epsilon_i[h]|\right]\middle|Z_1, ..., Z_n\right]$$

Where the inner expectation on the RHS above is over the Rademachers with ϵ_i and Z_i for $i = 1, ..., n$ treated as fixed.

Define the set $\mathcal{T} \subseteq \mathbb{R}^n$ by:

$$\mathcal{T} = [t = (t_1, ..., t_n) \in \mathbb{R}^n : t_i = \alpha_n(z)'\hat{Q}_n^{-1}\Psi_n(Z_i)\epsilon_i[h], z \in \mathcal{Z}, h \in \mathcal{H}]$$

Define a norm $||\cdot||_{n,2}$ on $\mathbb{R}^n$ by $||t||_{n,2}^2 = \frac{1}{n}\sum_{i=1}^{n} t_i^2$. By Dudley (1967) there exists a universal constant D so that:

$$E_\eta\left[\sup_{z\in\mathcal{Z},h\in\mathcal{H}}|\alpha_n(z)'\hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_n(Z_i)(\eta_i\epsilon_i[h])|\right]$$

$$\leq D\int_0^\theta\sqrt{log\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta)}d\delta$$

Where $\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta)$ is the smallest number of radius-δ $||\cdot||_{n,2}$ balls needed to cover $\mathcal{T}$ and θ is the smallest upper bound on the $||\cdot||_{n,2}$-distance between any two points in $\mathcal{T}$. Using our bounds on $|\epsilon_i[h]|$:

$$\theta = 2 \sup_{t\in\mathcal{T}} ||t||_{n,2} \leq 4\bar{c}||\hat{Q}_n^{-\frac{1}{2}}||_{op}$$

Let $\tilde{A}_i[h] = \hat{Q}_n^{-1}\Psi_n(Z_i)\epsilon_i[h]$. Using the Lipschitz constant for α_n given in Assumption 3.6.ii, and using B.6 we get (for $h_1, h_2 \in \mathcal{H}$):

$$\left(\frac{1}{n}\sum_{i=1}^{n} |\alpha_n(z_1)'\tilde{A}_i[h_1] - \alpha_n(z_2)'\tilde{A}_i[h_2]|^2\right)^{\frac{1}{2}}$$

$$\text{(B.10)} \quad \leq \left((\bar{c}\ell_n)^2||z_1 - z_2||_2^2 + |h_1 - h_2|_\infty^2\right)^{\frac{1}{2}} 4\sqrt{2}||\hat{Q}_n^{-\frac{1}{2}}||_{op}$$

Define b_n by $b_n = 4||\hat{Q}_n^{-\frac{1}{2}}||_{op}$. Note that $||\hat{Q}_n^{-\frac{1}{2}}||_{op} = O_p(1)$ and so $b_n = O_p(1)$. Then from B.10:

$$\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta) \leq \mathcal{N}(\mathcal{Z}, ||\cdot||_2, \frac{\delta}{\bar{c}\ell_n b_n})\mathcal{N}(\mathcal{H}, |\cdot|_\infty, \frac{\delta}{b_n})$$

And so (using sub-additivity of the square root):

$$\int_0^\theta \sqrt{log\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta)}d\delta \leq \int_0^{\bar{c}b_n} \sqrt{log\mathcal{N}(\mathcal{Z}, ||\cdot||_2, \frac{\delta}{\bar{c}\ell_n b_n})}d\delta + \int_0^{\bar{c}b_n} \sqrt{log\mathcal{N}(\mathcal{H}, |\cdot|_\infty, \frac{\delta}{b_n})}d\delta$$

Making the substitution $u = \frac{\delta}{\bar{c}b_n}$ into the first integral and $u = \frac{\delta}{b_n}$ into the second:

$$\int_0^\theta \sqrt{log\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta)}d\delta \leq \bar{c}b_n \int_0^1 \sqrt{log\mathcal{N}(\mathcal{Z}, ||\cdot||_2, \frac{u}{\ell_n})}du + b_n \int_0^{\bar{c}} \sqrt{log\mathcal{N}(\mathcal{H}, |\cdot|_\infty, u)}du$$

Because the integrand is decreasing in u:

$$\int_0^{\bar{c}} \sqrt{log\mathcal{N}(\mathcal{H}, |\cdot|_\infty, u)}du \leq \max\{\bar{c}, 1\} \int_0^1 \sqrt{log\mathcal{N}(\mathcal{H}, |\cdot|_\infty, u)}du$$

By Assumption 3.6.iii the integral on the RHS above is finite. So let ω_1 denote the finite constant on the RHS above.

By Assumption 3.5.ii $\mathcal{Z}$ is bounded and it has dimension $\dim(Z) < \infty$ therefore for some constant ω_2:

$$\mathcal{N}(\mathcal{Z}, ||\cdot||_2, \delta) \leq \omega_2 \left(\frac{1}{\delta}\right)^{\dim(Z)}$$

So we have:

$$\int_0^\theta \sqrt{log\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta)}d\delta \leq \bar{c}b_n \int_0^1 \sqrt{log\omega_2 + \dim(Z)log\left(\frac{\ell_n}{u}\right)}du + b_n\omega_1$$

Using sub-additivity of the square root we get from the above:

$$\begin{aligned}&\int_0^\theta \sqrt{log\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta)}d\delta \\ \leq&\bar{c}b_n\sqrt{log\omega_2} + b_n\omega_1 \\ &+\bar{c}b_n\sqrt{\dim(Z)}\left(\sqrt{log(\ell_n)} + \int_0^1 \sqrt{log\left(\frac{1}{u}\right)}du\right) \\ =&O_p\left(1 + \sqrt{log(\ell_n)}\right)\end{aligned}$$

And so from B.9 and Markov's inequality:

$$\sup_{z\in\mathcal{Z}, h\in\mathcal{H}} |\alpha_n(z)'\hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi_n(Z_i)\epsilon_i[h]| = O_p\left(1 + \sqrt{log(\ell_n)}\right)$$

Now we bound the second term on the RHS in B.8. Define $\mathcal{S}^{L_n-1} = [\beta \in \mathbb{R}^{L_n} : ||\beta||_2 \leq 1]$. Note that:

$$\begin{aligned}&|\alpha_n(z)'\hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi_n(Z_i)r_i[h]| \\ \leq&||\hat{Q}_n^{-1}||_{op}||\frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi_n(Z_i)r_i[h]||_2 \\ =&||\hat{Q}_n^{-1}||_{op}\sup_{\beta\in\mathcal{S}^{L_n-1}} |\beta'\frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi_n(Z_i)r_i[h]|\end{aligned}$$

And we already have by the matrix LLN that $||\hat{Q}_n^{-1}||_{op} = O_p(1)$. Note as well that:

$$E\left[\beta'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_n(Z_i)r_i[h]\right] = 0$$

Again, let $(\eta_1, ..., \eta_n)$ be a sample of iid Rademachers, independent of the data. By the symmetrization inequality:

$$\begin{aligned}
&E\left[\sup_{h\in\mathcal{H},\beta\in\mathcal{S}^{L_n-1}}|\beta'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_n(Z_i)r_i[h]|\right] \\
\text{(B.11)}\quad &\leq 2E\left[E_\eta\Big[\sup_{h\in\mathcal{H},\beta\in\mathcal{S}^{L_n-1}}|\beta'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_n(Z_i)r_i[h]|\Big]\right]
\end{aligned}$$

Where the inner expectation on the RHS above is over the Rademachers with r_i and Z_i for $i = 1, ..., n$ treated as fixed. Define a new set $\mathcal{T} \subseteq \mathbb{R}^n$ by:

$$\mathcal{T} = [t = (t_1, ..., t_n) \in \mathbb{R}^n : \ t_i = \beta'\psi_{n,l}(Z_i)r_i[h], h \in \mathcal{H}, \beta \in \mathcal{S}^{L_n-1}]$$

Again, by Dudley (1967):

$$\begin{aligned}
&E_\eta\left[\sup_{h\in\mathcal{H},\beta\in\mathcal{S}^{L_n-1}}|\beta'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_n(Z_i)(\eta_i r_i[h])|\right] \\
&\leq D\int_0^\theta \sqrt{log\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta)}d\delta
\end{aligned}$$

Using our bound on $|r_i[h]|$:

$$\theta = 2\sup_{t\in\mathcal{T}}||t||_{n,2} \leq 2\bar{c}\bar{\ell}R_n(s)||\hat{Q}_n^{\frac{1}{2}}||_{op}$$

Using B.7 we get (for $h_1, h_2 \in \mathcal{H}$):

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{i=1}^{n}|\Psi_n(Z_i)r_i[h_1] - \Psi_n(Z_i)r_i[h_2]|^2\right)^{\frac{1}{2}} \\
&\leq \sqrt{2}\bar{\ell}R_n(s)||\hat{Q}_n^{\frac{1}{2}}||_{op}\left(\bar{c}^2||\beta_1 - \beta_2||_2^2 + |h_1 - h_2|_\infty^2\right)^{\frac{1}{2}}
\end{aligned}$$

Define c_n by $c_n = 2\bar{\ell}R_n(s)||\hat{Q}_n^{\frac{1}{2}}||_{op}$. Then from B.10:

$$\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta) \leq \mathcal{N}(\mathcal{S}^{L_n-1}, ||\cdot||_2, \frac{\delta}{\bar{c}c_n})\mathcal{N}(\mathcal{H}, |\cdot|_\infty, \frac{\delta}{b_n})$$

And so (using sub-additivity of the square root) and making substitutions, much as before:

$$\int_0^\theta \sqrt{log\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta)}d\delta \leq \bar{c}c_n \int_0^1 \sqrt{log\mathcal{N}(\mathcal{S}^{L_n-1}, ||\cdot||_2, u)}du + c_n \int_0^{\bar{c}} \sqrt{log\mathcal{N}(\mathcal{H}, |\cdot|_\infty, u)}du$$

We have already shown that that the second integral on the RHS above is bounded by a finite constant ω_1. The covering number of a unit ball in $\mathbb{R}^{L_n}$ satisfies for some universal constant $\omega_3 > 0$:

$$\mathcal{N}(\mathcal{S}^{L_n-1}, ||\cdot||_2, \delta) = \omega_3\left(\frac{1}{\delta}\right)^{L_n}$$

Substituting this and using sub-additivity of the square root we get:

$$\int_0^\theta \sqrt{log\mathcal{N}(\mathcal{T}, ||\cdot||_{n,2}, \delta)}d\delta \leq \bar{c}c_n\sqrt{log\omega_3} + c_n\omega_1 + \bar{c}c_n\sqrt{L_n}\int_0^1 \sqrt{log\left(\frac{1}{u}\right)}du$$

And so:

$$\begin{aligned}
&E\left[\sup_{h\in\mathcal{H},\beta\in\mathcal{S}^{L_n-1}} |\beta'\frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi_n(Z_i)r_i[h]|\right] \\
\leq& \bar{c}E[c_n]\sqrt{L_n}\int_0^1 \sqrt{log\left(\frac{1}{u}\right)}du + \bar{c}E[c_n]\sqrt{L_n}\int_0^1 \sqrt{log\left(\frac{1}{u}\right)}du \\
=& O\big(R_n(s)\sqrt{L_n}\big)
\end{aligned}$$

And so, by Markov's inequality the second term in the RHS of B.8 is

$O_p\big(R_n(s)\sqrt{L_n}\big)$. So in all:

$$\sup_{h\in\mathcal{H}}|\Psi_n'\hat{\gamma}[h]-A[h]|_\infty \le \frac{\bar{\xi}_n}{\sqrt{n}}\sup_{z\in\mathcal{Z},h\in\mathcal{H}}|\sqrt{n}\alpha_n(z)'(\hat{\gamma}[h]-\gamma[h])| + \sup_{h\in\mathcal{H}}|\Psi_n'\gamma[h]-A[h]|_\infty = O_p\left(\frac{\bar{\xi}_n}{\sqrt{n}}[1+\sqrt{log(\ell_n)}+\sqrt{L_n}R_n(s)]+R_n(s)\right)$$

Q.E.D.

THEOREM 3.2: From Lemma 3.3:

$$\sup_{h\in\mathcal{H}}|\Psi_n'\hat{\gamma}[h]-A[h]|_\infty = O_p\left(\frac{\bar{\xi}_n}{\sqrt{n}}[1+\sqrt{log(\ell_n)}+\sqrt{L_n}R_n(s)]+R_n(s)\right)$$

Note that if $\Phi_n'\beta\in\mathcal{H}$ then $\hat{\Pi}_n'\beta=\Psi_n'\hat{\gamma}[\Phi_n'\beta]$ and by definition $\Pi_n'\beta = A[\Phi_n'\beta]$, and so:

$$\sup_{\beta\in\mathbb{R}^{K_n}:\,\Phi_n'\beta\in\mathcal{H}}|(\hat{\Pi}_n-\Pi_n)'\beta|_\infty \le \sup_{h\in\mathcal{H}}|\Psi_n'\hat{\gamma}[h]-A[h]|_\infty$$

Applying the rates in Assumption 3.7.ii:

$$\frac{\bar{\xi}_n}{\sqrt{n}}[1+\sqrt{log(\ell_n)}+\sqrt{L_n}R_n(s)]+R_n = O\left(\frac{\sqrt{L_n}}{\sqrt{n}}\sqrt{log(L_n)}+L_n^{-s_0(s)/\dim(Z)}]\right)$$

Note that Assumption 3.5.i is identical to Assumption A.2 in Belloni *et al.* (2015), Assumption 3.6.i implies Assumption A.3 in Belloni *et al.* (2015) (with '$\ell_k c_k$' in their notation equal to $R_n(s)$), Assumption 3.7.i implies Assumption A.4 in Belloni *et al.* (2015) and Assumption 3.7.i and 3.7.ii imply Assumption A.5. Assumption A.1 in Belloni *et al.* (2015), that the data are iid, is assumed in our paper throughout. Therefore we can apply Belloni *et al.* (2015) Theorem 4.3 for $\hat{g}_n$, and under our other assumptions the rate simplifies to:

$$|\hat{g}_n-g_0|_\infty = O\left(\frac{\sqrt{L_n}}{\sqrt{n}}\sqrt{log(L_n)}+L_n^{-s_0(s)/\dim(Z)}]\right)$$

Then the triangle inequality gives the result.

Q.E.D.

THEOREM 3.3: We note two facts about B-spline basis functions on an interval. Firstly, if $K_n \geq 2$ then we can apply a linear transformation to Φ_n so that the first two entries are 1 and x. We will assume without loss of generality that Φ_n has been transformed in this way.

Secondly, because the basis functions are at least third order, the vector of functions Φ_n is at least twice continuously differentiable at all but a finite set of points in its domain. For any z at which the second derivatives are defined, let $\frac{\partial^2}{\partial x^2}\Phi_n(z)$ denote the vector of second derivatives of each component of Φ_n at z. We can define this function elsewhere by right-continuity, that is if the second derivatives are not defined for some z then let $\frac{\partial^2}{\partial x^2}\Phi_n(z) = \lim_{z' \downarrow z} \frac{\partial^2}{\partial x^2}\Phi_n(z')$. $\frac{\partial^2}{\partial x^2}\Phi_n$ is an invertible linear transformation of the vector of $(s_0-2)^{th}$-order B-spline basis functions with the same knot points as the original spline basis. It then follows from the approximation properties of splines that there is a sequence $\tilde{\kappa}_n = O(K_n^{-1})$ so that for any Lipschitz continuous function h defined on $[a,b]$ with Lipschitz constant L, there exists $\beta \in \mathbb{R}^{K_n}$ with:

$$|\frac{\partial^2}{\partial x^2}\Phi_n'\beta - h|_\infty \leq L\tilde{\kappa}_n$$

See for example DeVore & Lorentz (1993).

First we show that any $h \in \mathcal{H}$ is Lipschitz continuous with constant at most $C = \frac{2}{(b-a)} + c(b-a)$.

Consider some $h \in \mathcal{H}$, since h is twice differentiable its first derivative also exists and so for any $x_1, x_2 \in [a,b]$:

$$h(x_2) - h(x_1) = \int_0^1 (x_2 - x_1)\frac{\partial}{\partial x}h(x_1 + t(x_2 - x_1))dt$$

Because $|\frac{\partial^2}{\partial x^2}h|_\infty \leq c$ the above implies:

$$|h(x_2) - h(x_1)| + c|x_2 - x_1|^2 \geq |(x_2 - x_1)\frac{\partial}{\partial x}h(x_1)|$$

Substituting $x_1 = \frac{1}{2}(a+b)$ and $x_2 = b$ into the above we get:

$$|h(x_2) - h(x_1)| + c\frac{1}{4}(b-a)^2 \geq \frac{1}{2}(b-a)|\frac{\partial}{\partial x}h(x_1)|$$

Since $h(x) \in [0,1]$, $|h(x_1) - h(x_2)| \leq 1$ and so we get:

$$|\frac{\partial}{\partial x} h(x_1)| \leq \frac{2}{(b-a)} + c\frac{1}{2}(b-a)$$

And again, since $|\frac{\partial^2}{\partial x^2} h|_\infty \leq c$ we then have that for any $x \in [a,b]$:

$$|\frac{\partial}{\partial x} h(x)| \leq \frac{2}{(b-a)} + c(b-a)$$

Denote the RHS by C. We thus have that any $h \in \mathcal{H}$ is Lipschitz continuous with constant at most C.

Let the functional P extend a function $h \in \mathcal{H}$ to a function defined on $\mathbb{R}$ as follows:

$$P[h](x) = \begin{cases} h(x) & \text{if } x \in [a,b] \\ h(a) & \text{if } x < a \\ h(b) & \text{if } x > b \end{cases}$$

For $r \in [0, b-a]$ define the linear operator $M_r : \mathcal{B}_X \to \mathcal{B}_X$ by:

$$M_r[h](x) = \frac{\int_{x-r}^{x+r} P[h](y)dy}{2r}$$

For $x \in [a,b]$.

If $|h|_\infty < \infty$ then it is easy to see that:

$$|M_r[h]|_\infty \leq |h|_\infty$$

With a substitution we get:

$$M_r[h](x) = \frac{\int_{-r}^{r} P[h](y+x)dy}{2r}$$

For any $h \in \mathcal{H}$, $\frac{\partial^2}{\partial x^2} h$ is uniformly bounded by c. So we can use the dominated convergence theorem to get that for any $h \in \mathcal{H}$:

$$\begin{aligned} \frac{\partial^2}{\partial x^2} M_r[h](x) =& \frac{1}{2r} \int_{-r}^{r} \frac{\partial^2}{\partial x^2} P[h](y+x)dy \\ =& \frac{1}{2r} \int_{\max\{a,x-r\}}^{\min\{b,x+r\}} \frac{\partial^2}{\partial y^2} h(y)dy \end{aligned}$$

Where we have used that the second derivative of $P[h]$ is zero outside of $[a,b]$.

So for any $h \in \mathcal{H}$:

$$|\frac{\partial^2}{\partial x^2} M_r[h](x)| \leq c$$

Assuming without loss of generality that $x_1 \geq x_2$, we see for any $h \in \mathcal{H}$:

$$\begin{aligned}
&|\frac{\partial^2}{\partial x^2} M_r[h](x_1) - \frac{\partial^2}{\partial x^2} M_r[h](x_2)| \\
=&\frac{1}{2r}\left| \int_{\max\{x_1-r, x_2+r\}}^{x_1+r} \frac{\partial^2}{\partial y^2} P[h](y)dy - \int_{x_2-r}^{\min\{x_2+r, x_1-r\}} \frac{\partial^2}{\partial y^2} P[h](y)dy \right| \\
\leq&\frac{c}{r}|x_1 - x_2|
\end{aligned}$$

Where the last inequality follows because $\frac{\partial^2}{\partial x^2} P[h]$ is uniformly bounded by c for $h \in \mathcal{H}$. So we have established that for any $h \in \mathcal{H}$ the function $M_r[h]$ has second derivatives that are Lipschitz continuous with Lipschitz constant at most $\frac{c}{r}$.

And now note that, because $h \in \mathcal{H}$ is Lipschitz continuous with constant C for any $x \in [a,b]$:

$$|M_r[h](x) - h(x)| \leq rC$$

Now, recall the properties of B-splines discussed at the beginning of this proof. Because $\frac{\partial^2}{\partial x^2} M_r[h]$ is Lipschitz continuous with constant $\frac{c}{r}$, there is some $\beta \in \mathbb{R}^{K_n}$ with:

$$|\frac{\partial^2}{\partial x^2} \Phi_n' \beta - \frac{\partial^2}{\partial x^2} M_r[h]|_\infty \leq \frac{c}{r}\tilde{\kappa}_n$$

In which case $|\frac{\partial^2}{\partial x^2} \Phi_n' \beta|_\infty \leq c(1 + \frac{\tilde{\kappa}_n}{r})$. And so letting $\tilde{\beta} = (1 + \frac{\tilde{\kappa}_n}{r})^{-1}\beta$ we get $|\frac{\partial^2}{\partial x^2} \Phi_n' \tilde{\beta}|_\infty \leq c$, and from the triangle inequality:

$$\begin{aligned}
|\frac{\partial^2}{\partial x^2} \Phi_n' \tilde{\beta} - \frac{\partial^2}{\partial x^2} M_r[h]|_\infty \leq&\frac{c}{r}\tilde{\kappa}_n + \left(\frac{\tilde{\kappa}_n / r}{1 + (\tilde{\kappa}_n / r)}\right) |\frac{\partial^2}{\partial x^2} \Phi_n' \beta|_\infty \\
\leq&2\frac{c}{r}\tilde{\kappa}_n
\end{aligned}$$

Which implies:

$$|\frac{\partial}{\partial x}\Phi_n'\tilde{\beta} - \frac{\partial}{\partial x}M_r[h]|_\infty \leq \left|\frac{\partial}{\partial x}\Phi_n(\frac{a+b}{2})'\tilde{\beta} - \frac{\partial}{\partial x}M_r[h](\frac{a+b}{2})\right| + (b-a)\frac{c}{r}\tilde{\kappa}_n$$

Where again $\frac{\partial}{\partial x}\Phi_n$ is defined at points at which the derivative is undefined by right continuity. The inequality above then implies:

$$|\Phi_n'\tilde{\beta} - M_r[h]|_\infty \leq |\Phi_n(\frac{a+b}{2})'\tilde{\beta} - M_r[h](\frac{a+b}{2})| + \frac{b-a}{2}\left|\frac{\partial}{\partial x}\Phi_n(\frac{a+b}{2})'\tilde{\beta} - \frac{\partial}{\partial x}M_r[h](\frac{a+b}{2})\right| + \left(\frac{b-a}{2}\right)^2 2\frac{c}{r}\tilde{\kappa}_n \tag{B.12}$$

But the first two entries of $\Phi_n(x)$ are 1 and x, so let β^* be identical to $\tilde{\beta}$ aside from its first two entries β_1^* and β_2^* which are given by:

$$\beta_2^* = \frac{\partial}{\partial x}M_r[h](\frac{a+b}{2}) - \frac{\partial}{\partial x}\Phi_n(\frac{a+b}{2})'\tilde{\beta}$$

And:

$$\beta_1^* = M_r[h](\frac{a+b}{2}) - \Phi_n(\frac{a+b}{2})'\tilde{\beta} - (\frac{a+b}{2})\beta_2^*$$

Then $\frac{\partial^2}{\partial x^2}\Phi_n'\beta^* = \frac{\partial^2}{\partial x^2}\Phi_n'\tilde{\beta}$, and so $|\frac{\partial^2}{\partial x^2}\Phi_n'\beta^*|_\infty \leq c$. Moreover, repeating the same steps used to get (B.12) above we see:

$$|\Phi_n'\beta^* - M_r[h]|_\infty \leq \left(\frac{b-a}{2}\right)^2 2\frac{c}{r}\tilde{\kappa}_n$$

We already showed $|M_r[h](x) - h(x)| \leq rC$ and so

$$|\Phi_n'\beta^* - h|_\infty \leq \left(\frac{b-a}{2}\right)^2 2\frac{c}{r}\tilde{\kappa}_n + rC$$

And so setting $r = \sqrt{\tilde{\kappa}_n}$:

$$|\Phi_n'\beta^* - h|_\infty \leq \left[\left(\frac{b-a}{2}\right)^2 2c + C\right]\sqrt{\tilde{\kappa}_n}$$

Since we found such a β^* for any h the result follows.

Q.E.D.

REFERENCES

Belloni, Alexandre, Chernozhukov, Victor, Chetverikov, Denis, & Kato, Kengo. 2015. Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, **186**, 345–366.

Chen, Xiaohong, & Pouzo, Demian. 2015. Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models. *Econometrica*, **83**, 1013–1079.

Chen, Xiaohong, & Santos, Andres. 2018. Overidentification in Regular Models. *Econometrica*, **86**, 1771–1817.

Darolles, Serge, Fan, Yanqin, Florens, Jean-Pierre, & Renault, Eric. 2011. Nonparametric Instrumental Regression. *Econometrica*.

DeVore, Ronald A., & Lorentz, George G. 1993. *Constructive Approximation*. Springer-Verlag.

Dudley, R. M. 1967. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *Journal of Functional Analysis*, **1**, 290–330.

Kress, Rainer. 2014. *Linear Integral Equations*.

Rudelson, M. 1999. Random Vectors in the Isotropic Position. *Journal of Functional Analysis*, **164**, 60–72.