Supplemental Appendix for "The Trade-Off Between Flexibility and Robustness in Instrumental Variables Analysis"

By Ben Deaner*

These appendices contain proofs of results in the main paper and other materials. Section B contains details regarding the practical illustration and an additional numerical exercise. Section C contains proofs of statements in the main text and appendices.

Additional Simulation Results and Details

B1. Calibration Details

The NPIV estimate that we use for h_0 in our numerical exercises is a penalized sieve 2SLS estimator based on Horowitz (2011) (we adapted code available with his paper when implementing the estimator). To define the estimator, let $\Phi(I_{i,a})$ and $\Psi(Z_{i,a})$ be vectors of technical regressors and instruments that are of the same length. In addition, let $Q_{i,a}$ be a vector of controls. We use the same vector of controls as Dahl & Lochner (2012) which includes polynomials of $P_{i,a-1}$ up to fifth powers, second differences of time-varying household characteristics, and constant household characteristics. Let $\mathcal{I}$ be the set of individual-age pairs in the dataset. We first partial out the controls from the technical instruments to get residualized technical instruments $\tilde{Z}_{i,a}$:

$$\tilde{Z}_{i,a} = \Psi(Z_{i,a}) - \left(\sum_{(i,a)\in\mathcal{I}} \Psi(Z_{i,a})Q'_{i,a}\right)\left(\sum_{(i,a)\in\mathcal{I}} Q_{i,a}Q'_{i,a}\right)^{-1} Q_{i,a}$$

Then, we perform linear IV taking as the outcome the second differenced test scores $\Delta Y_{i,a}$, and as the endogenous regressors, the second differenced technical regressors $\Delta\tilde{I}_{i,a} := \Phi(I_{i,a}) - \Phi(I_{i,a-2})$. Following Horowitz (2011) we regularize the inverse by adding the identify matrix $\mathbb{I}$ multiplied by a penalty parameter λ.

* Deaner: Department of Economics, University College London and CeMMAP (email: bendeaner@gmail.com)

The resulting coefficient estimates $\hat{\beta}$ are given below:

$$\Sigma =\Big(\frac{1}{|\mathcal{I}|}\sum_{(i,a)\in\mathcal{I}}\Delta\tilde{I}_{i,a}\tilde{Z}'_{i,a}\Big)\Big(\frac{1}{|\mathcal{I}|}\sum_{(i,a)\in\mathcal{I}}\Delta\tilde{I}_{i,a}\tilde{Z}'_{i,a}\Big)'$$
$$\hat{\beta} =(\Sigma+\lambda\mathbb{I})^{-1}\Big(\frac{1}{|\mathcal{I}|}\sum_{(i,a)\in\mathcal{I}}\Delta\tilde{I}_{i,a}\tilde{Z}'_{i,a}\Big)\frac{1}{|\mathcal{I}|}\sum_{(i,a)\in\mathcal{I}}\tilde{Z}_{i,a}\Delta Y_{i,a}$$

The resulting estimate of h_0 is given by $\hat{h}(x)=\Phi(x)'\hat{\beta}-\Phi(\bar{I})'\hat{\beta}$, where the second term normalizes the intercept so that $\hat{h}(\bar{I})=0$. In our implementation we take $\Phi(I_{i,a})$ and $\Psi(Z_{i,a})$ to contain polynomials up to cubic that are orthogonalized as in Horowitz (2011). For the penalty λ we use 10^{-15}.

We now discuss calibration of the distribution of $I_{i,a}$ given $Z_{i,a}$. We use the same procedure to calibrate the distribution of $I_{i,a-2}$ given $Z_{i,a}$ (simply replace $I_{i,a}$ with $I_{i,a-2}$ in the steps below). We first estimate the conditional mean of $I_{i,a}$ given $Z_{i,a}$ by Nadaraya-Watson. That is, we form an estimate of $E[I_{a,i}|Z_{i,a}=z]$ as follows:

$$\hat{E}[I_{a,i}|Z_{i,a}=z]=\frac{\sum_{(i,a)\in\mathcal{I}}\phi\big(\frac{|z-Z_{i,a}|}{w_Z}\big)I_{i,a}}{\sum_{(i,a)\in\mathcal{I}}\phi\big(\frac{|z-Z_{i,a}|}{w_Z}\big)}$$

To select bandwidths we use Silverman's rule, namely $(\frac{4}{5|\mathcal{I}|})^{1/(3+4)}\hat{\sigma}_Z$ for w_Z, where $\hat{\sigma}_Z$ is the standard deviation of $Z_{i,a}$. We estimate the residual variance by $\hat{\sigma}^2_I=\sum_{(i,a)\in\mathcal{I}}(I_{i,a}-\hat{E}[I_{a,i}|Z_{i,a}])^2$. The conditional distribution of $I_{i,a}$ given $Z_{i,a}=z$ is then normal with mean $\hat{E}[I_{a,i}|Z_{i,a}=z]$ and variance $\hat{\sigma}^2_I$. To be precise, we calibrate the density of $I_{i,a}$ given $Z_{i,a}$ as follows:

$$\hat{f}_{I_{i,a}|Z_{i,a}}(x|z)=\frac{1}{\sqrt{2\pi\hat{\sigma}^2_I}}\exp\Big(-\frac{(x-\hat{E}[I_{a,i}|Z_{i,a}=z])^2}{2\hat{\sigma}^2_I}\Big)$$

Statistical completeness of the calibrated distribution of $I_{a,i}$ given $Z_{i,a}$ then follows by Hu & Shiu (2018). For the bias exercise in this appendix and to evaluate the identified set for the slope of the best linear approximation to h_0 we also require the marginal density for $Z_{i,a}$. We use the following:

$$\hat{f}_{Z_{i,a}}(z)=\frac{1}{w_Z|\mathcal{I}|}\sum_{(i,a)\in\mathcal{I}}\phi\Big(\frac{|z-Z_{i,a}|}{w_Z}\Big)$$

Given an estimate $\hat{h}$ of h_0, and estimates $\hat{f}_{I_{i,a}|Z_{i,a}}$ and $\hat{f}_{I_{i,a-2}|Z_{i,a}}$ of the densities of $I_{i,a}$ and $I_{i,a-2}$ conditional on $Z_{i,a}$, we calibrate $E[\tilde{Y}_{i,a}|Z_{i,a}]$ and $E[\tilde{Y}_{i,a-2}|Z_{i,a}]$

as follows:

$$E[\tilde{Y}_{i,a}|Z_{i,a}=z]=\int \hat{h}(x)\hat{f}_{I_{i,a}|Z_{i,a}}(x|z)dx$$
$$E[\tilde{Y}_{i,a-2}|Z_{i,a}=z]=\int \hat{h}(x)\hat{f}_{I_{i,a-2}|Z_{i,a}}(x|z)dx$$

$E[\Delta\tilde{Y}_{i,a}|Z_{i,a}=z]$ is then the difference between these two objects. To evaluate the integrals we use numerical integration, discretizing the support into a grid of 1000 evenly spaced points between the largest and smallest values of $I_{i,a}$ in the data.

B2. Evaluating the Envelopes of the Identified Set

Our method for evaluating features of the identified set is based on Deaner (2019) who uses linear programming to estimate the identified set. First consider the case in which $\mathcal{H}$ contains parametric functions of the form $\Phi(\cdot)'\beta$ for some basis functions Φ. Then the envelopes of the identified set for $h_0(\cdot)-h_0(\bar{I})$ at some x, in the case of the moment condition in levels, are the maximum and minimum of $\Phi(x)'\beta$ subject to the constraint that for all z in the support of the instrument:

$$|E[\tilde{Y}_{i,a}|Z_{i,a}=z]-E[\Phi(I_{i,a})'\beta|Z_{i,a}=z]|\leq b$$

And for the moment condition in differences:

$$|E[\Delta\tilde{Y}_{i,a}|Z_{i,a}=z]-\big(E[\Phi(I_{i,a})|Z_{i,a}=z]-E[\Phi(I_{i,a-2})|Z_{i,a}=z]\big)'\beta|\leq b$$

We have already discussed evaluation of $E[\tilde{Y}_{i,a}|Z_{i,a}=z]$ and $E[\Delta\tilde{Y}_{i,a}|Z_{i,a}=z]$. In order to check the constraints at some value of z we must also evaluate $E[\Phi(I_{i,a})|Z_{i,a}=z]$ and $E[\Phi(I_{i,a-2})|Z_{i,a}=z]$. Again, this can be achieved by numerical integration of the integrals below:

$$E[\Phi(I_{i,a})|Z_{i,a}=z]=\int \Phi(x)\hat{f}_{I_{i,a}|Z_{i,a}}(x|z)dx$$
$$E[\Phi(I_{i,a-2})|Z_{i,a}=z]=\int \Phi(x)\hat{f}_{I_{i,a-2}|Z_{i,a}}(x|z)dx$$

Following Deaner (2019) we enforce the constraints only on a discrete grid (here 1000 evenly spaced points), and thus the problem can be solved by linear programming. If we wish to evaluate the set for the slope of the best linear approximation of h_0, we replace the objective with $E[(I_{i,a}-\bar{I})\Phi(I_{i,a})]'\beta/Var(I_{i,a})$ and we can again evaluate the expectation and variance by numerical integration using the densities $\hat{f}_{I_{i,a}|Z_{i,a}}$ and $\hat{f}_{Z_{i,a}}$.

For the bounds on second derivatives, we follow Deaner (2019) and alter the above by a) using a very rich set of basis functions Φ, we use cubic splines with eighteen evenly-spaced knot points, leading twenty basis functions, b) we add a constraint that $|\frac{\partial^2}{\partial x^2}\Phi(x)| \leq c$ over a dense grid of points in the support of $I_{i,a}$. We evaluate these second derivatives analytically.

B3. Bias Simulation Study

To supplement the empirical exercises in Section IV of the main text, we carry out a simulation exercise to evaluate the asymptotic bias in flexible 2SLS under misspecification. Consider vectors of basis functions Ψ and Φ. Using these basis functions, we may form endogenous regressors $\Phi(I_{i,a})$ and instruments $\Psi(Z_{i,a})$. Given the outcome $\tilde{Y}_{i,a}$, we can then estimate the structural function h_0 by 2SLS. The population analogue of this estimate, which we denote h°, is given below:

$$\pi = E[\Psi(Z_{a,i})\Psi(Z_{a,i})']^{-1}E[\Psi(Z_{a,i})\Phi(I_{a,i})']$$
$$h^\circ(x) = \Phi(x)'\big(\pi' E[\Psi(Z_{a,i})\Psi(Z_{a,i})']\pi\big)^{-1}\pi' E[\Psi(Z_{a,i})\tilde{Y}_{a,i}]$$

Alternatively, we may instead apply 2SLS only after second differencing outcomes and regressors. The population analogue of this estimate is denoted by $\tilde{h}$ and has the formula below, where $\Delta\Phi(I_a) := \Phi(I_a) - \Phi(I_{a-2})$.

$$\tilde{\pi} = E[\Psi(Z_{a,i})\Psi(Z_{a,i})']^{-1}E[\Psi(Z_{a,i})\Delta\Phi(I_{a,i})']$$
$$\tilde{h}(x) = \Phi(x)'\big(\tilde{\pi}' E[\Psi(Z_{a,i})\Psi(Z_{a,i})']\tilde{\pi}\big)^{-1}\tilde{\pi}' E[\Psi(Z_{a,i})\Delta\tilde{Y}_{a,i}]$$

In the main text we consider two, possibly misspecified, identifying moment restrictions. The misspecified restriction in levels and the restriction in second differences are given below.

$$E[\tilde{Y}_{i,a} - h_0(I_{i,a})|Z_{i,a}] = u_0(Z_{i,a}) \tag{B1}$$
$$E[\Delta\tilde{Y}_{i,a} - \Delta h_0(I_{i,a})|Z_{i,a}] = u_0(Z_{i,a}) \tag{B2}$$

Suppose that the first equation above holds. That is, the moment condition in levels is misspecified, with the misspecification captured in u_0. Then the asymptotic bias in the levels estimate can be decomposed into two parts:

$$\begin{aligned} h^\circ(x) - h_0(x) =& \Phi(x)'\big(\pi' E[\Psi(Z_{a,i})\Psi(Z_{a,i})']\pi\big)^{-1}\pi' E[\Psi(Z_{a,i})u_0(Z_{a,i})] \\ &+ \Phi(x)'\big(\pi' E[\Psi(Z_{a,i})\Psi(Z_{a,i})']\pi\big)^{-1}\pi' E[\Psi(Z_{a,i})h_0(I_{i,a})] - h_0(x) \end{aligned}$$

The term in the first line on the RHS above captures the bias due to the misspecification of the conditional moment restriction. That is, bias due to the fact that $u_0 \neq 0$ in (B1). The second line accounts for the approximation error. That is, the fact that there is no β_0 such that $h_0(x) = \Phi(x)'\beta_0$.

For the estimator in differences $\tilde{h}(x)$, an analogous decomposition applies when we impose (B2).

$$\begin{aligned}\tilde{h}(x)-h_0(x)=&\Phi(x)'\big(\tilde{\pi}'E[\Psi(Z_{a,i})\Psi(Z_{a,i})']\tilde{\pi}\big)^{-1}\tilde{\pi}'E[\Psi(Z_{a,i})u_0(Z_{a,i})]\\&+\Phi(x)'\big(\tilde{\pi}'E[\Psi(Z_{a,i})\Psi(Z_{a,i})']\tilde{\pi}\big)^{-1}\tilde{\pi}'E[\Psi(Z_{a,i})\Delta h_0(I_{i,a})]-h_0(x)\end{aligned}$$

Again, the second row captures the asymptotic bias due to approximation error. In this case, the first row captures bias due to misspecification of the conditional moment restriction in differences i.e., the bias that arises because $u_0 \neq 0$ in (B2).

We evaluate the bias due to approximation error and the bias due to misspecification of the moment condition under our calibrated DGP. Given our calibration of the joint distribution of $I_{i,a}$ and $Z_{i,a}$, and of $I_{i,a-2}$ and $Z_{i,a}$, and calibration of h_0 (which we take as the NPIV estimator in the main text) we can evaluate the approximation bias for both h° and $\tilde{h}$.

In order to evaluate the bias that results from misspecification of the relevant moment restriction, we must know u_0. Rather than keep u_0 fixed or find the u_0 that maximizes the magnitude of the bias, we draw many values of u_0 at random and evaluate the bias under each draw. Thus we obtain a distribution of this bias term and we can evaluate its quartiles. We constrain u_0 to be of the form $u_0(z) = \Psi(z)'\omega_0$, where ω_0 is a vector of coefficients and Ψ are the same basis functions that we use for our instruments. We draw the components of ω_0 independently from a standard normal distribution and then normalize the coefficients so that $\sup_{z\in S_Z}|u_0(z)| = b$. Throughout we take Ψ to be polynomial of degree four, that is $\Psi(z) = (1, z, z^2, z^3, z^4)'$. We consider $\Phi(x)$ to be powers up to degree K for $K = 1, 2, 3, 4$.

Figure B1 plots the results. Panels (a), (b), and (c) pertain to the estimates with second differencing $\tilde{h}$, and the remaining panels to the estimator in levels h°. From left to right, we plot subfigures for increasing values of K (we omit $K = 4$ in these figures to aid legibility). In all cases we subtract off the bias at $\bar{I}$, the mean of $I_{i,a}$. Thus we plot contributions to the bias in the causal effect of changing $I_{i,a}$ away from its mean $\bar{I}$. The approximation bias is given by the solid black curve. Dashed curves plot the top and bottom quartiles of the bias due to misspecification of the moment restriction for various choices of b. The dashed lines in red, blue, and black are upper and lower bias quartiles for $b = 0.01$, 0.02, and 0.03 respectively.

For $K = 3$ there is no approximation error as our calibrated h_0 is in fact cubic. Interestingly, the approximation error is not markedly lower for the quadratic approximation ($K = 2$) than for the linear approximation $K = 1$. Note that for sieve IV estimators, the bias due to approximation error is not guaranteed to decrease as the model becomes more flexible. In fact, in the NPIV literature researchers typically assume that a 'stability condition' holds (for example, Assumption 5.2(ii) in Chen & Pouzo (2012)) which ensures the bias due to approximation error vanishes.

In-line with the results in the main text, a more flexible specification is asso-

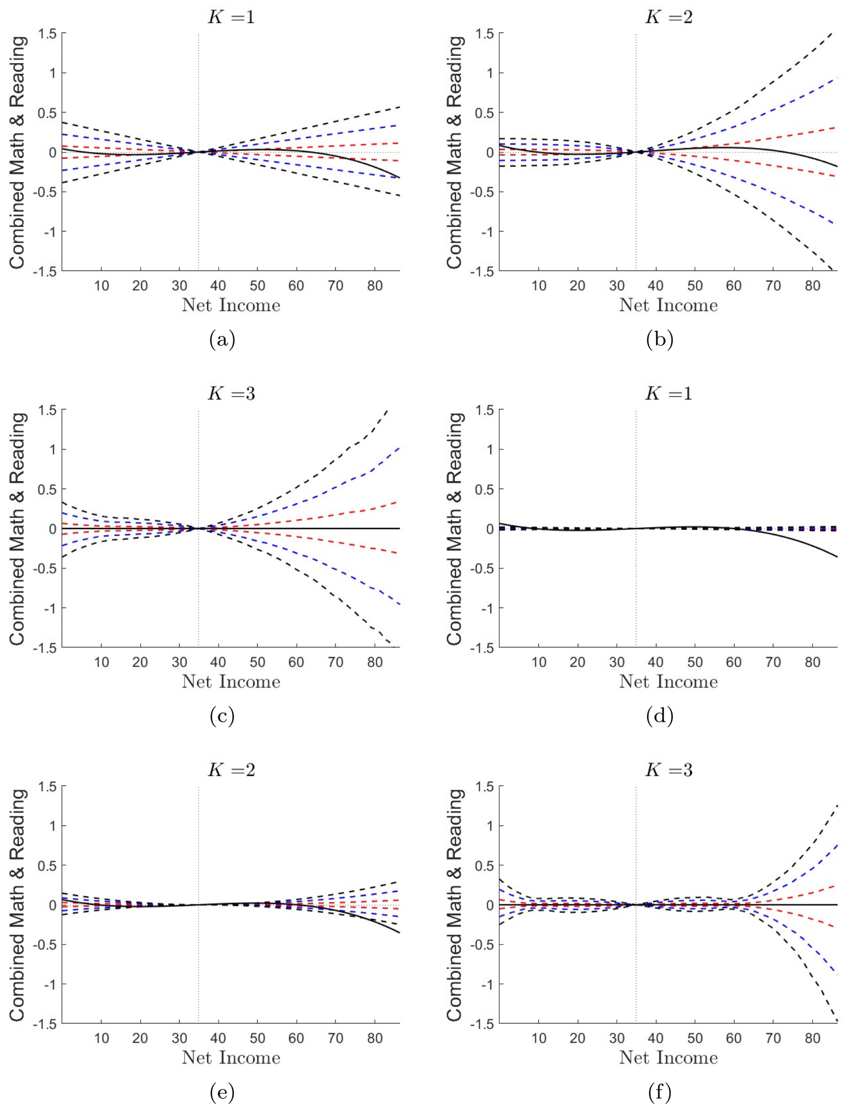

Figure B1. : Bias Quartiles (Polynomial Functional-Forms)

Bias contributions under the assumption that h_0 is polynomial of degree K. (a), (b), (c) are for $\tilde{h}$ and (d), (e), (f) for h°. Lower and upper dashed lines in red, blue, and black represent the upper and lower quartiles of the bias due to moment condition misspecification for $b = 0.01, 0.02, 0.03$ respectively. These quantiles are over 1000 random draws of u_0. The solid line is the bias due to approximation error.

ciated with greater maximum bias due to misspecification over the support of the treatment. This increase is small and difficult to discern in the differenced case between $K = 2$ and $K = 3$ but can be seen in Figure B2 (a). Figure B2 summarizes the trends in B1 and includes the $K = 4$ case. It plots the maximum distance over the support of the treatment between the upper and lower quartiles of the bias due to misspecification. Subfigure (a) corresponds to $\tilde{h}$ and the moment restriction in differences, (b) to h° and the moment condition in levels. Polynomials of degree $K = 1, ..., 4$ are given in green, red, blue, and black respectively. Note that the bias due to misspecification must scale linearly with b. Note that for $K = 1, 2, 3$ the bias due to misspecification of the moment condition is more pronounced for the estimator with differencing but this is reversed for $K = 4$.

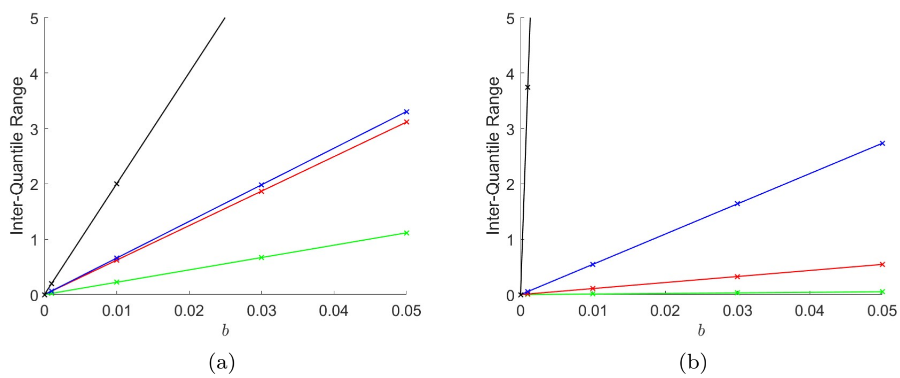

Figure B2. : Maximum Inter-Quantile Range

Maximum distances between the upper and lower 5% quantiles of the bias due to misspecification of the relevant moment restriction. (a) corresponds to the moment condition in differences, (b) in levels. Solid curves correspond to polynomials of degree $K = 1, ..., 4$ given in green, blue, red, and black respectively.

Appendix C: Proofs

Throughout the proofs it will be convenient to introduce some additional notation. We let $\mathbb{A}$ be the linear operator from $\mathcal{X}$ to $\mathcal{Z}$ defined by $\mathbb{A}[h](Z) = E[h(X)|Z]$. The operator norm of $\mathbb{A}$, $\|\mathbb{A}\|$ is defined by $\sup_{h \in \mathcal{X}: \|h\|_{\mathcal{X}}=1} \|A[h]\|_{\mathcal{X}}$. If $\mathbb{A}$ is injective we denote its inverse by $\mathbb{A}^{-1}$ so that $\mathbb{A}[h] = g$ if and only if $\mathbb{A}^{-1}[g] = h$. If $\mathcal{S}$ is a subset of $\mathcal{Z}$ then $\mathbb{A}^{-1}[\mathcal{S}]$ is the pre-image of $\mathcal{S}$ under $\mathbb{A}$ (i.e., the set of elements $h \in \mathcal{X}$ such that $\mathbb{A}[h] \in \mathcal{S}$). The closure of a set $\mathcal{S}$ is denoted $\bar{\mathcal{S}}$ and its interior by $int(\mathcal{S})$.

We let $\tilde{\mathcal{X}}$ be the subset of $\mathcal{X}$ that contains all $h \in \mathcal{X}$ such that $E[h(X)] = 0$. Similarly, we let $\tilde{\mathcal{Z}}$ be the subset of $\mathcal{Z}$ that contains all $g \in \mathcal{Z}$ with $E[g(Z)] = 0$. We equip $\tilde{\mathcal{X}}$ an $\tilde{\mathcal{Z}}$ with the norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Z}}$ respectively and thus define open and closed balls on these sets. Finally, we defined a function $g_0 \in \mathcal{Z}$ by $g_0(Z) = E[Y|Z]$ which is well-defined and finite under Assumption 2.iii.

C1. Proofs of Results in the Main Text

It is convenient to rewrite some of the objects in Section II in terms of the notation defined above. We can rewrite the identified set Θ_b as follows.

$$\Theta_b = [h \in \mathcal{X} : h \in \mathcal{H}, g_0 - \mathbb{A}[h] \in \mathcal{U}, \|g_0 - \mathbb{A}[h]\|_{\mathcal{Z}} \leq b]$$

Moreover, Assumption 1 states that $\mathbb{A}^{-1}[\mathcal{U}]$ contains a closed $\tilde{\mathcal{X}}$ ball of radius r_u centred at zero. Assumption 2.ii states that $\mathbb{A}$ is injective compact and infinite-dimensional. The pseudo-solution h^* (if it exists and is unique) is given by $h^* = \mathbb{A}^{-1}[g_0]$.

With Assumptions 1 and 2 and the identified set expressed in terms of $\mathbb{A}$ and g_0 as above, the proofs in this section apply for **any** linear operator $\mathbb{A}$, not just $h \mapsto E[h(X)|Z = \cdot]$. That is, the proofs continue to apply when we replace the condition $E[Y - h_0(X)|Z] = u_0(Z)$ with any equation of the form:

$$g_0 - \mathbb{A}[h_0] = u_0$$

Thus our main results apply for all linear conditional moment restriction models (but not necessarily non-linear moment restriction models). Of course, the specific choice of $\mathbb{A}$ then impacts the interpretation of Assumption 2.ii which must apply for this particular $\mathbb{A}$.

An important example appears in Section IV, in which the relevant operator and function g_0 are:

$$\mathbb{A}[h](z) = E[h(I_{i,a}) - h(I_{i,a-2})|Z_{i,a} = z], \; g_0(z) = E[\Delta Y_{i,a}|Z_{i,a} = z]$$

It is not difficult to show that compactness of the operator follows from compactness of $h \mapsto E[h(I_{i,a})|Z_{i,a} = \cdot]$ and $h \mapsto E[h(I_{i,a-2})|Z_{i,a} = \cdot]$. Injectivity of this operator requires that $\mathcal{X}$ incorporates some location normalization. For example, we can let $\mathcal{X}$ be the Banach space of bounded continuous functions h so that $E[h(I_{i,a})] = 0$ or $h(\bar{I}) = 0$. In this case, injectivity on $\mathcal{X}$ is equivalent to the statement that for any bounded continuous function h (which does not necessarily satisfy the location normalization), $E[h(I_{i,a}) - h(I_{i,a-2})|Z_{i,a} = \cdot] = 0$ if and only if h is constant. In Section IV we evaluate the identified set for $h_0(\cdot) - h_0(\bar{I})$ and so the normalization is of no consequence.

Lemma 1.1. *Suppose Assumptions 2.i and ii. hold. Then $\tilde{\mathcal{X}}$ is a Banach space with the norm $\|\cdot\|_{\mathcal{X}}$, and the restriction of $\mathbb{A}$ to $\tilde{\mathcal{X}}$ is a compact infinite-dimensional linear operator from $\tilde{\mathcal{X}}$ to $\mathcal{Z}$.*

Proof. First we show that $\tilde{\mathcal{X}}$ is a Banach space with the norm $\|\cdot\|_{\mathcal{X}}$, that is, a linear space that is complete with respect to $\|\cdot\|_{\mathcal{X}}$. Recall that $\tilde{\mathcal{X}}$ contains all elements of $\mathcal{X}$ that have mean zero. Let $h_1, h_2 \in \tilde{\mathcal{X}}$ and $\alpha, \beta \in \mathbb{R}$. Note that $E[h_1(X)] = E[h_2(X)] = 0$ and $h_1, h_2 \in \mathcal{X}$. Since $\mathcal{X}$ is a Banach space (by Assumption 2.i) and thus linear, $\alpha h_1 + \beta h_2 \in \mathcal{X}$, and by linearity of the mean $E[\alpha h_1(X) + \beta h_2(X)] = 0$. Thus $\alpha h_1 + \beta h_2 \in \tilde{\mathcal{X}}$ and so $\tilde{\mathcal{X}}$ is a linear space.

To show $\tilde{\mathcal{X}}$ is complete, let $\{h_n\}_{n=1}^{\infty}$ be a Cauchy sequence in $\tilde{\mathcal{X}}$. Since $\mathcal{X}$ is Banach and thus complete, this sequence converges (in the norm $\|\cdot\|_{\mathcal{X}}$) to an element $h_\infty \in \mathcal{X}$. By Assumption 2.i the mapping of a function in $\mathcal{X}$ to its mean is continuous, and so $h_n \to h_\infty$ implies $E[h_n(X)] \to E[h_\infty(X)]$ and since $E[h_n(X)] = 0$ for all n, we must have $E[h_\infty(X)] = 0$. Thus $h_\infty \in \tilde{\mathcal{X}}$ and so $\tilde{\mathcal{X}}$ is complete.

Finally we show that $\mathbb{A}$ is a compact operator from $\tilde{\mathcal{X}}$ to $\mathcal{Z}$. By definition a compact operator maps bounded sets into relatively compact sets. Assumption 2.ii states that $\mathbb{A}$ is a compact operator between $\mathcal{X}$ and $\mathcal{Z}$. Since any bounded set in $\tilde{\mathcal{X}}$ is also a bounded set in $\mathcal{X}$ we get that $\mathbb{A}$ is compact between $\tilde{\mathcal{X}}$ to $\mathcal{Z}$. □

Lemma 1.2. *Suppose Assumptions 1, 2.i, and 2.ii hold and in addition, there is an open $\tilde{\mathcal{X}}$-ball centred at an element $h^\circ \in \bar{\mathcal{S}}$ of radius r_s. Then for any $\delta, b > 0$ there exist functions $h_1, h_2 \in \mathcal{S}$, so that for $i = 1, 2$, $\mathbb{A}[h_i - h^\circ] \in \mathcal{U}$, $\|\mathbb{A}[h_i - h^\circ]\|_{\mathcal{Z}} \leq b$, $\|h_i - h^\circ\| \geq \min\{r_h, r_u\} - \delta$, and $\|h_1 - h_2\| \geq 2\min\{r_s, r_u\} - \delta$.*

Proof. From Lemma 1.1 $\tilde{\mathcal{X}}$ is a Banach space and $\mathbb{A}$ is a compact infinite-dimensional operator from $\tilde{\mathcal{X}}$ to $\mathcal{Z}$. Thus we can apply Theorem 15.4 (or 2.20) in Kress (2014) to get that $\mathbb{A}^{-1}$ is unbounded and so:

$$\text{(C1)} \qquad \sup_{h \in \tilde{\mathcal{X}}, \|\mathbb{A}[h]\|_{\mathcal{X}} = 1} \|h\|_{\mathcal{Z}} = \infty$$

Let $\|\mathbb{A}\|$ be the operator norm of $\mathbb{A}$, this must be finite because $\mathbb{A}$ is compact and therefore bounded (see Theorem 2.14 in Kress (2014)). By (C1) for any $0 < \delta < 2\min\{r_s, r_u, \frac{b}{\|\mathbb{A}\|}\}$ there exists an element $\tilde{h}$ of $\tilde{\mathcal{X}}$ so that $\|\tilde{h}\|_{\mathcal{X}} = \min\{r_s, r_u\} - \frac{\delta}{2}$ and $\|\mathbb{A}[\tilde{h}]\|_{\mathcal{Z}} \leq b - \frac{\delta}{2}\|\mathbb{A}\|$. By linearity of $\mathbb{A}$ and the elementary properties of norms we also have $\|-\tilde{h}\|_{\mathcal{X}} = \min\{r_s, r_u\} - \frac{\delta}{2}$ and $\|\mathbb{A}[-\tilde{h}]\|_{\mathcal{Z}} \leq b - \frac{\delta}{2}\|\mathbb{A}\|$.

Because $\|\tilde{h}\|_{\mathcal{X}} \leq r_s$ we have $h^\circ + \tilde{h} \in \bar{\mathcal{S}}$ and similarly $h^\circ - \tilde{h} \in \bar{\mathcal{S}}$. Since a set $\mathcal{S}$ must be dense in its closure $\bar{\mathcal{S}}$ there exists an $h_1 \in \mathcal{S}$ so that $\|h^\circ + \tilde{h} - h_1\|_{\mathcal{X}} \leq \frac{\delta}{2}$ in which case, by the triangle inequality and $\|\tilde{h}\|_{\mathcal{X}} = \min\{r_s, r_u\} - \frac{\delta}{2}$, we get $\|h_1 - h^\circ\|_{\mathcal{X}} \leq \min\{r_s, r_u\}$. Given Assumption 1, this inequality implies that $h_1 - h^\circ \in \mathbb{A}^{-1}[\mathcal{U}]$ and therefore $\mathbb{A}[h_1 - h^\circ] \in \mathcal{U}$. Moreover, by the definition of the operator norm, the triangle equality, and $\|\mathbb{A}[\tilde{h}]\|_{\mathcal{Z}} \leq b - \frac{\delta}{2}\|\mathbb{A}\|$, we get:

$$\begin{aligned}\|\mathbb{A}[h_1 - h^\circ]\|_{\mathcal{Z}} &\leq \|\mathbb{A}[\tilde{h}]\|_{\mathcal{Z}} + \|\mathbb{A}[h^\circ + \tilde{h} - h_1]\|_{\mathcal{Z}} \\ &\leq \|\mathbb{A}[\tilde{h}]\|_{\mathcal{Z}} + \|\mathbb{A}\|\|h^\circ + \tilde{h} - h_1\|_{\mathcal{X}} \leq b\end{aligned}$$

So in all, $h_1 \in \mathcal{S}$, $\mathbb{A}[h_1 - h^\circ] \in \mathcal{U}$, $\|\mathbb{A}[h^\circ - h_1]\|_{\mathcal{Z}} \leq b$, and $\|h^\circ + \tilde{h} - h_1\|_{\mathcal{X}} \leq \frac{\delta}{2}$. Applying the same reasoning with $\tilde{h}$ replaced by $-\tilde{h}$ we see that there exists an

$h_2 \in \mathcal{S}$ with $\mathbb{A}[h_2 - h^\circ] \in \mathcal{U}$, and $\|\mathbb{A}[h^\circ - h_2]\|_{\mathcal{Z}} \leq b$, and $\|h^\circ - \tilde{h} - h_2\|_{\mathcal{X}} \leq \frac{\delta}{2}$. Now that note that by the triangle inequality:

$$\begin{aligned}\|h_1 - h_2\|_{\mathcal{X}} &= \|2\tilde{h} + (h^\circ - \tilde{h} - h_2) - (h^\circ + \tilde{h} - h_1)\|_{\mathcal{X}} \\ &\geq 2\|\tilde{h}\|_{\mathcal{X}} - \|h^\circ + \tilde{h} - h_1\|_{\mathcal{X}} - \|h^\circ - \tilde{h} - h_2\|_{\mathcal{X}} \\ &\geq 2\min\{r_s, r_u\} - \delta\end{aligned}$$

□

Lemma 1.3. *Suppose Assumption 2.i and 2.ii hold. If $\mathcal{S}$ is a compact subset of $\mathcal{X}$ then:*

$$\lim_{b\to 0} \sup_{u\in\mathbb{A}[\mathcal{S}], \|u\|_{\mathcal{Z}}\leq b} \|\mathbb{A}^{-1}[u]\|_{\mathcal{X}} = 0$$

Proof. Since Assumption 2.ii holds, $\mathbb{A}$ is compact (and therefore continuous) and injective. Denote the restriction of $\mathbb{A}$ to $\mathcal{S}$ by $\mathbb{A}_{\mathcal{S}}$ and its inverse by $\mathbb{A}_{\mathcal{S}}^{-1}$. It is well-known that a continuous and injective function that maps from a compact set in a Banach space to another Banach space (in fact, any Hausdorff topological spaces, not necessarily Banach) has a continuous inverse. So by compactness of $\mathcal{S}$, $\mathbb{A}_{\mathcal{S}}^{-1}$ is continuous. Continuity of $\mathbb{A}_{\mathcal{S}}^{-1}$ implies:

$$\lim_{b\to 0} \sup_{u\in\mathbb{A}[\mathcal{S}], \|u\|_{\mathcal{Z}}\leq b} \|\mathbb{A}_{\mathcal{S}}^{-1}[u]\|_{\mathcal{X}} = 0$$

The final result follows because $\mathbb{A}_{\mathcal{S}}^{-1}[u]$ and $\mathbb{A}^{-1}[u]$ coincide for $u \in \mathbb{A}[\mathcal{S}]$. □

Lemma 1.4. *Suppose Assumptions 1 and 2 hold, and the set $\mathcal{S} \subset \mathcal{X}$ is such that $\mathcal{S}$ is absolutely convex and infinite dimensional. Let $h^\circ \in \mathcal{S}$ and suppose there exists $\alpha > 1$ so that $\alpha h^\circ \in \mathcal{S}$. Then:*

$$\lim_{b\to 0} \frac{\sup_{h\in\mathcal{S}, \mathbb{A}[h-h^\circ]\in\mathcal{U}:\, \|\mathbb{A}[h-h^\circ]\|_{\mathcal{Z}}\leq b} \|h - h^\circ\|_{\mathcal{X}}}{b} = \infty$$

Proof. Assume the contrary, then for some $b^* > 0$ there exists a finite scalar C so that if $h \in \mathcal{S}$, $\mathbb{A}[h - h^\circ] \in \mathcal{U}$, $\|\mathbb{A}[h - h^\circ]\|_{\mathcal{Z}} \leq b$, and $b \leq b^*$, then we must have $\|h - h^\circ\|_{\mathcal{X}} \leq Cb$. Therefore, if $h \in \mathcal{S}$, $\mathbb{A}[h - h^\circ] \in \mathcal{U}$, and $\|\mathbb{A}[h - h^\circ]\|_{\mathcal{Z}} \leq b^*$, we get $\|h - h^\circ\|_{\mathcal{X}} \leq C\|\mathbb{A}[h - h^\circ]\|_{\mathcal{Z}}$.

By definition of the operator norm of $\mathbb{A}$, $\|\mathbb{A}[h - h^\circ]\|_{\mathcal{Z}} \leq \|\mathbb{A}\|\|h - h^\circ\|_{\mathcal{X}}$. Assumption 2.ii implies that $\|\mathbb{A}\| > 0$. And so for any $h \in \mathcal{S}$ with $\mathbb{A}[h - h^\circ] \in \mathcal{U}$ we must have:

$$\|h - h^\circ\|_{\mathcal{X}} \leq \frac{1}{\|\mathbb{A}\|} b^* \implies \|h - h^\circ\|_{\mathcal{X}} \leq C\|\mathbb{A}[h - h^\circ]\|_{\mathcal{Z}}$$

By Assumption 1 there is an open $\tilde{\mathcal{X}}$-ball centred at zero in $\mathbb{A}^{-1}[\mathcal{U}]$ with radius

r_u. If $\|h - h^\circ\|_{\mathcal{X}} \leq r_u$ and $h - h^\circ \in \tilde{\mathcal{X}}$ it follows that $\mathbb{A}[h - h^\circ] \in \mathcal{U}$. Let $c^* = \min\{\frac{1}{\|\mathbb{A}\|} b^*, r_u\}$, we get that for any $h \in \mathcal{S}$ with $h - h^\circ \in \tilde{\mathcal{X}}$:

$$\text{(C2)} \qquad \|h - h^\circ\|_{\mathcal{X}} \leq c^* \implies \|h - h^\circ\|_{\mathcal{X}} \leq C\|\mathbb{A}[h - h^\circ]\|_{\mathcal{Z}}$$

We have $\alpha h^\circ \in \mathcal{S}$ for some $\alpha > 1$ and also that $\mathcal{S}$ is absolutely convex and hence convex. Note that $1/\alpha \in (0,1)$ so by convexity, for any $h \in \mathcal{S} \cap \tilde{\mathcal{X}}$, $h' := (1 - \frac{1}{\alpha})h + \frac{1}{\alpha}(\alpha h^\circ) \in \mathcal{S}$ in which case, using linearity of $\tilde{\mathcal{X}}$ (see Lemma 1.1), we have $(1 - \frac{1}{\alpha})h = h' - h^\circ \in \tilde{\mathcal{X}}$. Applying (C2) and then substituting $h' - h^\circ = (1 - \frac{1}{\alpha})h$ we get:

$$\|h\|_{\mathcal{X}} \leq \frac{\alpha}{\alpha - 1} c^* \implies \|h\|_{\mathcal{X}} \leq C\|\mathbb{A}[h]\|_{\mathcal{Z}}$$

Note the above holds for any such an $h \in \mathcal{S} \cap \tilde{\mathcal{X}}$.

Let R be the closed ball in $\tilde{\mathcal{X}}$ of radius $\frac{\alpha}{\alpha-1}c^*$. Let $\mathcal{C} = [\gamma h : h \in R \cap \mathcal{S}, \gamma \in \mathbb{R}]$. We have already shown that for any $h \in R \cap \mathcal{S}$, $\|h\|_{\mathcal{X}} \leq C\|\mathbb{A}[h]\|_{\mathcal{Z}}$. By linearity of $\mathbb{A}$ and properties of norms, for any $h \in \mathcal{C}$ we have that $\|h\|_{\mathcal{X}} \leq C\|\mathbb{A}[h]\|_{\mathcal{Z}}$.

Let $\bar{\mathcal{C}}$ be the closure of $\mathcal{C}$. By definition of the closure, for any $h \in \bar{\mathcal{C}}$ there is a sequence h_k in $\mathcal{C}$ so that $\|h - h_k\|_{\mathcal{X}} \to 0$. For all k, $\|h_k\|_{\mathcal{X}} \leq C\|\mathbb{A}[h_k]\|_{\mathcal{Z}}$, so by the triangle inequality and the definition of the operator norm:

$$\|h\|_{\mathcal{X}} \leq C\|\mathbb{A}[h]\|_{\mathcal{Z}} + (1 + C\|\mathbb{A}\|)\|h - h_k\|_{\mathcal{X}}$$

$\mathbb{A}$ is compact by Assumption 2.ii and thus continuous, so $\|\mathbb{A}\| < \infty$, and so since $\|h - h_k\|_{\mathcal{X}} \to 0$ we get $\|h\|_{\mathcal{X}} \leq C\|\mathbb{A}[h]\|_{\mathcal{Z}}$. Thus the inverse $\mathbb{A}^{-1}$ (which exists by Assumption 2.ii) is bounded on $\bar{\mathcal{C}}$.

Now, $R \cap \mathcal{S}$ is absolutely convex which implies $\mathcal{C}$ is a linear space and therefore so is $\bar{\mathcal{C}}$. Because $\mathcal{S}$ is infinite-dimensional and absolutely convex, $\mathcal{C}$ is infinite-dimensional and likewise $\bar{\mathcal{C}}$. It is well-known that a closed subset of a complete space is complete, $\bar{\mathcal{C}}$ is a closed subset of $\tilde{\mathcal{X}}$ by construction and $\tilde{\mathcal{X}}$ is a Banach space (see Lemma 1.1) and thus complete in the norm $\|\cdot\|_{\mathcal{X}}$. Thus $\bar{\mathcal{C}}$ is an infinite-dimensional, complete linear space, i.e., an infinite-dimensional Banach space. But the inverse of a compact injective operator on an infinite-dimensional Banach space cannot be bounded (see Theorem 15.4 in Kress (2014)), and so we have a contradiction. □

Proof of Theorem 1. **Part a.**

By Assumption 2.ii, the inverse $\mathbb{A}^{-1}$ exists. Since $\mathcal{H}$ is finite dimensional under Assumption 3.i, so too is its image under $\mathbb{A}$, denoted $\mathbb{A}[\mathcal{H}]$. The restriction of $\mathbb{A}^{-1}$ to $\mathbb{A}[\mathcal{H}]$ is then bounded (see e.g., Kress (2014) Theorem 2.6). Suppose the restriction of $\mathbb{A}^{-1}$ to $\mathbb{A}[\mathcal{H}]$ has norm c. Now, by the triangle inequality, for any

$h_1, h_2 \in \Theta_b$, $\|\mathbb{A}[h_1 - h_2]\|_{\mathcal{Z}} \leq 2b$, and since $\mathcal{H}$ is linear $h_1 - h_2 \in \mathcal{H}$. But then:

$$\|h_1 - h_2\|_{\mathcal{X}} = \|\mathbb{A}^{-1}\mathbb{A}[h_1 - h_2]\|_{\mathcal{X}} \leq c\|\mathbb{A}[h_1 - h_2]\|_{\mathcal{Z}} \leq 2cb$$

So we see $diam(\Theta_b) \leq 2cb$, and we are done.

Part b.

First we show that $\lim_{b\to 0} diam(\Theta_b) = 0$. Assumption 3.ii states that $\alpha h^* \in \mathcal{H}$ for some $\alpha > 1$ which, by absolute convexity of $\mathcal{H}$, implies $h^* \in \mathcal{H}$ and thus $\mathbb{A}[h^*] = g_0 \in \mathbb{A}[\mathcal{H}]$. Since $\mathbb{A}$ is injective by Assumption 2.ii:

$$\begin{aligned}
\sup_{h_1,h_2\in\Theta_b} \|h_1 - h_2\|_{\mathcal{X}} &\leq 2 \sup_{h\in\Theta_b} \|h - \mathbb{A}^{-1}[g_0]\|_{\mathcal{X}} \\
&\leq 2 \sup_{h\in\mathcal{H}:\, \|\mathbb{A}[h]-g_0\|_{\mathcal{Z}}\leq b} \|h - \mathbb{A}^{-1}[g_0]\|_{\mathcal{X}} \\
&\leq \sup_{u\in\mathbb{A}[\mathcal{H}]-g_0:\, \|u\|_{\mathcal{Z}}\leq 2b} \|\mathbb{A}^{-1}[u]\|_{\mathcal{X}}
\end{aligned}$$

Where the first inequality follows by the triangle inequality, the second because the set $[h \in \mathcal{H} : \|\mathbb{A}[h] - g_0\|_{\mathcal{Z}} \leq b]$ is a subset of Θ_b, and the final inequality by a reparametrization ($\mathbb{A}[\mathcal{H}] - g_0$ is defined so that $u \in \mathbb{A}[\mathcal{H}] - g_0$ if and only if $u + g_0 \in \mathbb{A}[\mathcal{H}]$). Finally, since $\mathcal{H}$ is compact by Assumption 3.ii, it follows that $\mathcal{H} - \mathbb{A}^{-1}[g_0]$ is compact. Thus we can apply Lemma 1.3 with $\mathcal{S} = \mathcal{H} - \mathbb{A}^{-1}[g_0]$ and we get the result.

Now we show that $\lim_{b\to 0} diam(\Theta_b)/b = \infty$. Since $\mathbb{A}^{-1}[g_0] \in \Theta_b$:

$$\begin{aligned}
\sup_{h_1,h_2\in\Theta_b} \|h_1 - h_2\|_{\mathcal{X}} &\geq \sup_{h\in\Theta_b} \|h - \mathbb{A}^{-1}[g_0]\|_{\mathcal{X}} \\
&= \sup_{h\in\mathcal{S},\mathbb{A}[h]-g_0\in\mathcal{U}:\, \|\mathbb{A}[h]-g_0\|_{\mathcal{Z}}\leq b} \|h - \mathbb{A}^{-1}[g_0]\|_{\mathcal{X}}
\end{aligned}$$

Applying Lemma 1.4 with $\mathcal{S} = \mathcal{H}$ and $h^\circ = \mathbb{A}^{-1}[g_0]$ then gives the result.

Part c.

By Assumption 3.iii, $\bar{\mathcal{H}}$ contains an open $\tilde{\mathcal{X}}$-ball of radius r_h centred at h^*. Applying Lemma 1.2 with $\mathcal{S} = \mathcal{H}$ and $h^\circ = h^*$ and using that $\mathbb{A}[h^*] = g_0$, we see that for any $\delta > 0$ there exists $h_1, h_2 \in \mathcal{H}$, with $\|h_1 - h_2\|_{\mathcal{X}} \geq 2\min\{r_h, r_u\} - 2\delta$ and for $i = 1, 2$, $\mathbb{A}[h_i] - g_0 \in \mathcal{U}$, and $\|g_0 - \mathbb{A}[h_i]\|_{\mathcal{Z}} \leq b$. Thus for $i = 1, 2$, $h_i \in \Theta_b$. Since we can make δ arbitrarily small, it follows that the diameter of Θ_b satisfies:

$$\sup_{h_1,h_2\in\Theta_b} \|h_1 - h_2\|_{\mathcal{X}} \geq 2\min\{r_h, r_u\}$$

$\square$

Proof of Corollary 1. By Assumption 2.ii, $0 < \|\mathbb{A}\| < \infty$. Fix some $b > 0$. Given $\mathcal{U} = \mathcal{Z}$, the identified set contains all functions $h \in \mathcal{H}$ with $\|\mathbb{A}[h] - g_0\|_{\mathcal{Z}} \leq b$. By Theorem 1 part c., for any $c < \infty$ and $0 < \epsilon < b$, there exist h_1 and h_2 in the identified set with $\mathcal{H} = \mathcal{X}$ so that $\|h_1 - h_2\|_{\mathcal{X}} \geq c$ and $\|\mathbb{A}[h_i] - g_0\|_{\mathcal{Z}} \leq b - \epsilon$ for $i = 1, 2$. Since $\inf_{h \in \mathcal{H}_k} \|h_i - h\|_{\mathcal{X}} \to 0$, there is some k^* so that for all $k \geq k^*$ there is a corresponding $h^*_{1,k}, h^*_{2,k} \in \mathcal{H}_k$ and $\|h_i - h^*_{i,k}\|_{\mathcal{X}} \leq \frac{\epsilon}{\|\mathbb{A}\|}$ for $i = 1, 2$. By the triangle inequality:

$$\|\mathbb{A}[h^*_{i,k}] - g_0\|_{\mathcal{Z}} \leq \|\mathbb{A}[h_i] - g_0\|_{\mathcal{Z}} + \|\mathbb{A}\|\|h_i - h^*_{i,k}\|_{\mathcal{X}} \leq b$$

So $h^*_{1,k}, h^*_{2,k} \in \Theta_{b,k}$, and again by the triangle inequality:

$$\|h^*_{1,k} - h^*_{2,k}\|_{\mathcal{X}} \geq \|h_1 - h_2\|_{\mathcal{X}} - \frac{2\epsilon}{\|\mathbb{A}\|} \geq c - \frac{2\epsilon}{\|\mathbb{A}\|}$$

And so for all $k \geq k^*$, $diam(\Theta_{b,k}) \geq c - \frac{2\epsilon}{\|\mathbb{A}\|}$. But since c can be made arbitrarily large and this will hold for some corresponding k^*, we have $diam(\Theta_{b,k}) \to \infty$. $\square$

Proof of Corollary 2. Consider some $h \in \Theta_{b,\eta}$. Then by definition there is an $h' \in \mathcal{H}$ so that $\|h - h'\|_{\mathcal{X}} \leq \eta$ and $\|g_0 - \mathbb{A}[h]\|_{\mathcal{Z}} \leq b$. Then by the triangle inequality, $\|g_0 - \mathbb{A}[h']\|_{\mathcal{Z}} \leq b + \|\mathbb{A}\|\eta$. As such, because $\mathcal{U} = \mathcal{Z}$, we see $h' \in \Theta_{b+\|\mathbb{A}\|\eta,0}$ and $\|h - h'\|_{\mathcal{X}} \leq \eta$. Since there is such an h' for any $h \in \Theta_{b,\eta}$, we have that for any pair $h_1, h_2 \in \Theta_{b,\eta}$ there is a pair $h'_1, h'_2 \in \Theta_{b+\|\mathbb{A}\|\eta,0}$ so that $\|h_i - h'_i\|_{\mathcal{X}} \leq \eta$ for $i = 1, 2$ and so by the triangle inequality:

$$\|h_1 - h_2\|_{\mathcal{X}} \leq \|h'_1 - h'_2\|_{\mathcal{X}} + 2\eta \leq diam(\Theta_{b+\|\mathbb{A}\|\eta,0}) + 2\eta$$

And since this holds for any $h_1, h_2 \in \Theta_{b,\eta}$, we see that

$$diam(\Theta_{b,\eta}) \leq diam(\Theta_{b+\|\mathbb{A}\|\eta,0}) + 2\eta.$$

Thus if $\mathcal{H}$ satisfies Assumption 3.i we have by Theorem 1 that for the same constant C in Theorem 1.a $diam(\Theta_{b,\eta}) \leq Cb + (C\|\mathbb{A}\| + 2)\eta$ and under either Assumption 3.i or 3.ii $\lim_{b,\eta \to 0} diam(\Theta_{b,\eta}) = 0$. $\square$

Proof of Theorem 2. We prove a more general result than the one stated in the paper. We let $\mathcal{X}$ and $\mathcal{Z}$ be Hilbert spaces (not necessarily L_2) with inner products $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{Z}}$ respectively. We show that for $b > 0$, $diam(\mathbb{L}(\Theta_b)) < \infty$ if and only if $\mathbb{L}[h] = \langle w, h \rangle_{\mathcal{X}}$ where $w = \mathbb{A}^*[\alpha]$, in which case $diam(\mathbb{L}(\Theta_b)) = 2b\|\alpha\|_{\mathcal{Z}}$. $\mathbb{A}^*$ is the adjoint of $\mathbb{A}$. Specializing this to the L_2 case in the main body of the paper, $\langle h_1, h_2 \rangle_{\mathcal{X}} = E[h_1(X)h_2(X)]$, $\langle g_1, g_2 \rangle_{\mathcal{Z}} = E[g_1(Z)g_2(Z)]$, and the adjoint of the operator $\mathbb{A}$, denoted $\mathbb{A}^*$ is given by $\mathbb{A}^*[g](X) = E[g(Z)|X]$ and so we recover the result in the paper.

First we prove the following:

$$diam(\mathbb{L}(\Theta_b)) = 2 \sup_{g \in \mathbb{A}[\mathcal{X}]:\, \|g\|_{\mathcal{Z}} \leq b} |\mathbb{L}\mathbb{A}^{-1}[g]| \tag{C3}$$

Given our $\mathcal{H}$ and $\mathcal{U}$, $\Theta_b = [h \in \mathcal{X} : \|\mathbb{A}[h] - g_0\|_{\mathcal{Z}} \leq b]$, so we have:

$$\begin{aligned} diam(\mathbb{L}(\Theta_b)) &= \sup_{h_1,h_2 \in \mathcal{X}:\, \|\mathbb{A}[h_i]-g_0\|_{\mathcal{Z}} \leq b,\, i=1,2} |\mathbb{L}[h_1] - \mathbb{L}[h_2]| \\ &= \sup_{g_1,g_2 \in \mathbb{A}[\mathcal{X}]:\, \|g_i\|_{\mathcal{Z}} \leq b,\, i=1,2} |\mathbb{L}\mathbb{A}^{-1}[g_1 - g_2]| \end{aligned}$$

Where the final line uses linearity of $\mathbb{L}$ and that $g_0 \in \mathbb{A}[\mathcal{X}]$. Restricting $g_2 = -g_1$ we get the lower bound $diam(\mathbb{L}(\Theta_b)) \geq 2\sup_{g \in \mathbb{A}[\mathcal{X}]:\, \|g\|_{\mathcal{Z}} \leq b} |\mathbb{L}\mathbb{A}^{-1}[g]|$, and by the triangle inequality we obtain:

$$\sup_{g_1,g_2 \in \mathbb{A}[\mathcal{X}]:\, \|g_i\|_{\mathcal{Z}} \leq b,\, i=1,2} |\mathbb{L}\mathbb{A}^{-1}[g_1 - g_2]| \leq 2 \sup_{g \in \mathbb{A}[\mathcal{X}]:\, \|g\|_{\mathcal{Z}} \leq b} |\mathbb{L}\mathbb{A}^{-1}[g]|$$

Noting that the LHS is equal to $diam(\mathbb{L}(\Theta_b))$ we get (C3). By linearity, from (C3) we see that $diam(\mathbb{L}(\Theta_b)) = b \times diam(\mathbb{L}(\Theta_1))$. Moreover, $\sup_{g \in \mathbb{A}[\mathcal{X}]:\, \|g\| \leq 1} |\mathbb{L}\mathbb{A}^{-1}[g]|$ is the operator norm of $\mathbb{L}\mathbb{A}^{-1}$. If this operator is unbounded, $diam(\mathbb{L}(\Theta_1)) = \infty$ and likewise for $diam(\mathbb{L}(\Theta_b))$ for $b > 0$.

Suppose instead that $\mathbb{L}\mathbb{A}^{-1}$ is bounded. We will apply the Reisz representation theorem (see Theorem 4.8 in Kress (2014)). First we need to deal with the technicality that $\mathbb{L}\mathbb{A}^{-1}$ is not defined on the whole of $\mathcal{Z}$, but rather on $\mathbb{A}[\mathcal{X}]$. Since $\mathbb{L}\mathbb{A}^{-1}$ is bounded we can apply the Hahn-Banach theorem which shows $\mathbb{L}\mathbb{A}^{-1}$ has a bounded linear extension (with the same operator norm) that is defined everywhere on $\mathcal{Z}$ and coincides with $\mathbb{L}\mathbb{A}^{-1}$ on $\mathbb{A}[\mathcal{X}]$. Applying the Reisz representation theorem to this extension, we see $\mathbb{L}\mathbb{A}^{-1}$ is bounded if and only if there exists an $\alpha \in \mathcal{Z}$ so that for all $g \in \mathbb{A}[\mathcal{X}]$ we have $\mathbb{L}\mathbb{A}^{-1}[g] = \langle \alpha, g \rangle_{\mathcal{Z}}$. In which case $\sup_{g \in \mathbb{A}[\mathcal{X}]:\, \|g\| \leq 1} |\mathbb{L}\mathbb{A}^{-1}[g]| = \|\alpha\|_{\mathcal{Z}}$. Thus $diam(\mathbb{L}(\Theta_b)) = 2b\|\alpha\|_{\mathcal{Z}}$. But since $g \in \mathbb{A}[\mathcal{X}]$, we have:

$$\mathbb{L}\mathbb{A}^{-1}[g] = \langle \alpha, \mathbb{A}\mathbb{A}^{-1}[g] \rangle_{\mathcal{Z}} = \langle \mathbb{A}^*[\alpha], \mathbb{A}^{-1}[g] \rangle_{\mathcal{Z}}$$

And so, for all $h \in \mathcal{X}$, $\mathbb{L}[h] = \langle \mathbb{A}^*[\alpha], h \rangle_{\mathcal{X}}$ in which case $\mathbb{L}[h] = \langle w, h \rangle_{\mathcal{X}}$, where $w = \mathbb{A}^*[\alpha]$. And so we are done. If $diam(\mathbb{L}(\Theta_b)) < \infty$ then $\mathbb{L}[h] = \langle w, h \rangle_{\mathcal{X}}$ where $w = \mathbb{A}^*[\alpha]$, in which case $diam(\mathbb{L}(\Theta_b)) = 2b\|\alpha\|_{\mathcal{Z}}$. □

C2. Proof of Results Stated in Appendix A

Proof of Proposition A.1. From Proposition A.2 (see proof below) it is enough to show the following are all zero:

$$\begin{aligned} \gamma_0(x,z) &= E[Y_{x,z} - Y_x] \\ \ell_0(z;Z) &= E[Y_{x_0,z}|Z] - E[Y_{x_0,z}] \\ \delta_0(x,z;X,Z) &= E[Y_{x,z} - Y_{x_0,z}|X,Z] - E[Y_{x,z} - Y_{x_0,z}] \end{aligned}$$

Under Assumption C.i $E[Y_{x,z}] = E[Y_x]$ and so $\gamma_0(x,z) = 0$. Under C.ii $E[Y_{x_0,z}|Z] = E[Y_{x_0,z}]$ and so $\ell_0(z;Z) = 0$. Under Assumption C.iii $E[Y_{x,z} - Y_{x_0,z}|X,Z] = E[Y_{x,z} - Y_{x_0,z}]$ and so $\delta_0(x,z;X,Z) = 0$. □

Proof of Proposition A.2. Fix some arbitrary treatment level x_0. Using $h_0(x) = E[Y_x]$ and adding and subtracting terms, we get:

$$\begin{aligned} Y_{x,z} - h_0(x) = &E[Y_{x_0} - Y_x] + Y_{x,z} - Y_{x_0,z} \\ &+ Y_{x_0,z} - E[Y_{x_0,z}] + E[Y_{x_0,z}] - E[Y_{x_0}] \end{aligned}$$

Setting x equal to X and z equal to Z in the above and taking expectations conditional on Z we get:

$$\begin{aligned} E[Y - h_0(X)|Z] =&E\big[E[Y_{x_0} - Y_x]|_{x=X}|Z] + E[Y_{X,Z} - Y_{x_0,Z}|Z] \\ &+E[Y_{x_0,Z}|Z] - E[Y_{x_0,z}]|_{z=Z} + E[Y_{x_0,z}]|_{z=Z} - E[Y_{x_0}] \\ =&E\big[E[Y_{x_0} - Y_x]|_{x=X}|Z] + E[Y_{X,Z} - Y_{x_0,Z}|Z] \\ &+\ell_0(Z;Z) + \gamma_0(x_0,Z) \end{aligned}$$

Now consider the term $E[Y_{X,Z} - Y_{x_0,Z}|Z]$ on the RHS of the final equality above. Applying the law of iterated expectations and then adding and subtracting terms, we get:

$$\begin{aligned} E[Y_{X,Z} - Y_{x_0,Z}|Z] =&E[E[Y_{X,Z} - Y_{x_0,Z}|X,Z]|Z] \\ =&E[E[Y_x - Y_{x_0}]|_{x=X}|Z] + E[\gamma_0(X,Z)|Z] \\ &-\gamma_0(x_0,Z) + E[\delta_0(X,Z;X,Z)\big|Z] \end{aligned}$$

Substituting this, we get:

$$E[Y - h_0(X)|Z] = E[\gamma_0(X,Z)|Z] + \ell_0(Z;Z) + E[\delta_0(X,Z;X,Z)\big|Z]$$

□

Proof of Proposition A.3. Consider some $h \in \tilde{\mathcal{X}}$ with $\sup_{x \in S_X} |h(x)| \leq a$. Define

$u = \mathbb{A}[h] \in \tilde{\mathcal{Z}}$. Note that for any $z_1, z_2 \in S_Z$:

$$\begin{aligned} |u(z_1) - u(z_2)| &= |\int h(x)\big(f_{X|Z}(x|z_1) - f_{X|Z}(x|z_2)\big)dx| \\ &\leq \sup_{x\in S_X} |h(x)| \int |f_{X|Z}(x|z_1) - f_{X|Z}(x|z_2)|dx \\ &\leq \sup_{x\in S_X} |h(x)| \sup_{x\in S_X} |f_{X|Z}(x|z_1) - f_{X|Z}(x|z_2)| \\ &\leq \sup_{x\in S_X} |h(x)|c\|z_1 - z_2\|_2 \\ &\leq ac\|z_1 - z_2\|_2 \end{aligned}$$

Where the second inequality follows by the Hölder inequality and the third by (**??**). From the above we get:

$$\sup_{z_1 \neq z_2 \in S_Z} \frac{|u(z_1) - u(z_2)|}{\|z_1 - z_2\|_2} \leq ac$$

In addition, note that for all $z \in S_Z$, we have $|u(z)| \leq \sup_{x\in S_X} |h(x)| \leq a$. So we see that if $a = \min\{\bar{c}, C/c\}$ then $\mathbb{A}[h] \in \mathcal{U}$. If $\mathcal{X}$ is equipped with the max norm then the set of functions $h \in \tilde{\mathcal{X}}$ with $\sup_{x\in S_X} |h(x)| \leq \min\{\bar{c}, C/c\}$ is precisely the closed ball around zero with radius $\min\{\bar{c}, C/c\}$ which contains the open ball with the same radius. So we see that $\mathbb{A}^{-1}[\mathcal{U}]$ contains this open ball with this radius. □

Proof of Proposition A.4. We will suppose **a.** holds and show **b.** cannot be true. Fix some $h_0 \in \mathcal{H}$. Let $\tilde{\Theta}_b$ be the identified set from assumptions $h_0 \in \mathcal{H}$ and $\|u_0\|_{\mathcal{Z}} \leq b$ when the joint distribution of observables (X, Z, Y) is pinned down by $(X, Z, \eta) \sim F_{XZ\eta}$ and $E[Y|Z] = E[h_0(X)|Z]$. Note that by construction, $\tilde{\Theta}_b$ is the set of functions $h \in \mathcal{H}$ so that $\|\mathbb{A}[h_0] - \mathbb{A}[h]\|_{\mathcal{Z}} \leq b$. By definition of the diameter and the triangle inequality, for any $\epsilon > 0$ there must be some $h \in \tilde{\Theta}_b$ so that $\|h - h_0\|_{\mathcal{X}} > \frac{1}{2}diam(\tilde{\Theta}_b) - \epsilon$. Fix such an h. Now suppose $g_0 = \mathbb{A}[h]$. Together with $(X, Z, \eta) \sim F_{XZ\eta}$ this pins down the joint distribution of observables F_{XZY}. Because h is in the identified set, we must have:

$$\|u_0\|_{\mathcal{Z}} = \|g_0 - \mathbb{A}[h_0]\|_{\mathcal{Z}} = \|\mathbb{A}[h_0] - \mathbb{A}[h]\|_{\mathcal{Z}} \leq b$$

Now, note that the joint distribution of observables, F_{XZY} is identical to the case in which $h_0 = h$ and $u_0 = 0$. Thus by the consistency properties of the estimator, it follows that $\hat{h} \to^p h$. And so:

$$\text{plim}\|\hat{h} - h_0\|_{\mathcal{X}} = \|h - h_0\|_{\mathcal{X}} > \frac{1}{2}diam(\tilde{\Theta}_b) - \epsilon$$

□

Proof of Proposition A.5. **Part a.** Pick some $0 < c < 1$. A compact infinite-dimensional operator cannot have a bounded inverse (see e.g., Theorem 15.4 in Kress (2014)) and so $\mathbb{A}^{-1}$ is unbounded. Thus for any $\epsilon > 0$ there exists a non-zero function $h \in \mathcal{X}$ so that $\|\mathbb{A}[h]\|_{\mathcal{Z}} \leq \epsilon \|h\|_{\mathcal{X}}$. Given our choice of Banach space this means there is $0 < E[h(X)^2] < \infty$ so that:

$$E\big[E[h(X)|Z]^2\big] \leq \frac{1}{4}c^2 E[h(X_i)^2]$$

Moreover, by the triangle inequality:

$$E\big[E[\beta'\Phi_k(X)|Z]^2\big]^{1/2} \leq E\big[E[h(X)|Z]^2\big]^{1/2} + E\big[E[h(X) - \beta'\Phi_k(X)|Z]^2\big]^{1/2}$$

And note that any random variable W with $E[W^2] < \infty$, $E\big[E[W|Z]^2\big] \leq E[W^2]$ and so:

$$E\big[E[h(X_i) - \beta'\Phi_k(X_i)|Z_i]^2\big] \leq E\big[\big(h(X_i) - \beta'\Phi_k(X_i)\big)^2\big]$$

So we get that:

$$E\big[E[\beta'\Phi_k(X)|Z]^2\big]^{1/2} \leq \frac{1}{2}cE[h(X)^2]^{1/2} + E\big[\big(h(X) - \beta'\Phi_k(X)\big)^2\big]^{1/2} \tag{C4}$$

Moreover, again by the triangle inequality:

$$E[h(X_i)^2]^{1/2} \leq E\big[\big(\beta'\Phi_k(X)\big)^2\big]^{1/2} + E\big[\big(h(X) - \beta'\Phi_k(X)\big)^2\big]^{1/2}$$

By assumption, for some k there is a β so that:

$$E\big[\big(h(X) - \beta'\Phi_k(X)\big)^2\big]^{1/2} \leq \frac{c}{2c+2}E[h(X)^2]^{1/2}$$

Since $\frac{c}{2c+2} < 1$, this β must be non-zero. Note that the choice of $\{\Psi_k\}_{k=1}^{\infty}$ has no bearing on the value of k that achieves the above. Combining the previous two equations we get that for such a β:

$$E[h(X_i)^2]^{1/2} \leq \frac{2c+2}{c+2}E\big[\big(\beta'\Phi_k(X)\big)^2\big]^{1/2} \tag{C5}$$

And so:

$$E\big[\big(h(X) - \beta'\Phi_k(X)\big)^2\big]^{1/2} \leq \frac{c}{c+2}E[\big(h(X)\big)^2]^{1/2} \tag{C6}$$

Substituting (C5) and (C6) into (C4) we get:

$$E\big[E[\beta'\Phi_k(X)|Z_i]^2\big]^{1/2} \leq cE\big[\big(\beta'\Phi_k(X)\big)^2\big]^{1/2}$$

Now, define q by the projection below:

$$q(Z) = \Psi_k(Z)' E[\Psi_k(Z)\Psi_k(Z)']^{-1} E[\Psi_k(Z)\Phi_k(X)]'\beta$$

By elementary properties of least-squares projections:

$$E[q(Z)^2] \leq E\big[E[\beta'\Phi_k(X)|Z]^2\big]$$

So we have $E[q(Z)^2] \leq c^2 E\big[\big(\beta'\Phi_k(X)\big)^2\big]$. Using the definition of q and π_k, we can re-write this as:

$$\beta'\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k\beta \leq c^2\beta' E[\Phi_k(X)\Phi_k(X)']\beta$$

Now, note that from $\Phi_k(X) = \pi_k'\Psi_k(Z) + V_k$ and the definition of π_k, we have:

$$E[\Phi_k(X)\Phi_k(X)'] = \pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k + E[V_k V_k']$$

Substituting the above into the RHS of the previous inequality, we get:

$$\beta'\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k\beta \leq c^2\beta'\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k\beta + c^2\beta' E[V_k V_k']\beta$$

Using $0 < c < 1$ the above implies $\beta'\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k\beta \leq \frac{c^2}{1-c^2}\beta' E[V_k V_k']\beta$. Reparametrizing $\tilde{\beta} = E[V_k V_k']^{1/2}\beta$ and dividing through by $\|\tilde{\beta}\|_2^2$, we get:

$$\frac{1}{\|\tilde{\beta}\|_2^2}\tilde{\beta}' E[V_k V_k']^{-1/2}\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k E[V_k V_k']^{-1/2}\tilde{\beta} \leq \frac{c^2}{1-c^2}$$

But by elementary properties of eigenvalues, for any non-zero vector β:

$$\begin{aligned}&\lambda_{\min}\big(E[V_k V_k']^{-1/2}\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k E[V_k V_k']^{-1/2}\big)\\ \leq &\frac{1}{\|\beta\|_2^2}\beta' E[V_k V_k']^{-1/2}\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k E[V_k V_k']^{-1/2}\beta\end{aligned}$$

Where $\lambda_{\min}(\cdot)$ returns the smallest eigenvalue. The LHS above is simply $t_{\min,k}$, so we conclude that $t_{\min,k} \leq \frac{c^2}{1-c^2}$. Since $0 < c < 1$ was set arbitrarily, we obtain the result.

Part b. Decomposing the 2SLS formula in the usual way, we get:

$$\beta_k^* - \beta_0 = (\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k)^{-1}\pi_k' E[\Psi_k(Z)u_0(Z)]$$

So we see that:

$$\begin{aligned}&E\big[\big(h_0(X) - \Phi_k(X)'\beta_k^*\big)^2\big]\\ =&E\big[\big(\Phi_k(X)'(\beta_0 - \beta_k^*)\big)^2\big]\\ =&E\big[\big(\Phi_k(X)'(\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k)^{-1}\pi_k' E[\Psi_k(Z)u_0(Z)]\big)^2\big]\end{aligned}$$

Substituting $u_0(Z) = \Psi_k(Z)'\pi_k E[V_k V_k']^{-1/2} v_0$ for some vector v_0 gives:

$$\begin{aligned}E\big[\big(h_0(X) - \Phi_k(X)'\beta_k^*\big)^2\big] =&E\big[\big(\Phi_k(X)' E[V_k V_k']^{-1/2} v_0\big)^2\big]\\ =&v_0' E[V_k V_k']^{-1/2} E[\Phi_k(X)\Phi_k(X)'] E[V_k V_k']^{-1/2} v_0\end{aligned}$$

As in part a., note that $E[\Phi_k(X)\Phi_k(X)'] = \pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k + E[V_k V_k']$, and so $E\big[\big(h_0(X) - \Phi_k(X)'\beta_k^*\big)^2\big] \geq \|v_0\|_2^2$. Moreover, we have

$$E[u_0(Z)^2] = v_0' E[V_k V_k']^{-1/2} \pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k E[V_k V_k']^{-1/2} v_0$$

But then, letting v_0 be the eigenvector of $E[V_k V_k']^{-1/2}\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k E[V_k V_k']^{-1/2}$ associated with its smallest eigenvalue (which is $t_{\min,k}$), we get:

$$\begin{aligned}E[u_0(Z)^2] &= \lambda_{min}\big(E[V_k V_k']^{-1/2}\pi_k' E[\Psi_k(Z)\Psi_k(Z)']\pi_k E[V_k V_k']^{-1/2}\big)\|v_0\|^2\\ &= t_{\min,k}\|v_0\|_2^2\end{aligned}$$

And so $E\big[\big(h_0(X) - \Phi_k(X)'\beta_k^*\big)^2\big] \geq \frac{E[u_0(Z)^2]}{t_{\min,k}}$ Since we can scale the eigenvector v_0 however we like, we can scale it so that $E[u_0(Z)^2] = b$. □

Proof of Proposition A.6. Given $E[u_0(Z)] = 0$, for any $h_1, h_2 \in \Theta_b$ we have $E[h_1(X)] = E[h_2(X)] = E[Y]$. Then by the triangle inequality we see:

$$\begin{aligned}|h_1(x) - h_2(x)| \leq&\big|\big(h_1(x) - h_1(\bar{x})\big) - \big(h_2(x) - h_2(\bar{x})\big)\big|\\ &+\big|E\big[\big(h_1(x) - h_1(\bar{x})\big) - \big(h_2(x) - h_2(\bar{x})\big)\big]\big|\\ \leq&\big|\big(h_1(x) - h_1(\bar{x})\big) - \big(h_2(x) - h_2(\bar{x})\big)\big|\\ &+E\big[\big|\big(h_1(x) - h_1(\bar{x})\big) - \big(h_2(x) - h_2(\bar{x})\big)\big|\big]\\ \leq&2\sup_{x\in S_X}\big|\big(h_1(x) - h_1(\bar{x})\big) - \big(h_2(x) - h_2(\bar{x})\big)\big|\end{aligned}$$

Taking the supremum over $x \in S_X$ and dividing by 2 gives the result. □

*

REFERENCES

Chen, Xiaohong, & Pouzo, Demian. 2012. Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals. *Econometrica*.

Dahl, Gordon B., & Lochner, Lance. 2012. The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. *American Economic Review*, **102**, 1927–1956.

Deaner, Ben. 2019. Nonparametric Instrumental Variables Estimation Under Misspecification. *arXiv e-prints*, Jan., arXiv:1901.01241.

Horowitz, Joel L. 2011. Applied nonparametric instrumental variables estimation. *Econometrica. Journal of the Econometric Society*, **79**(2), 347–394.

Hu, Yingyao, & Shiu, Ji-Liang. 2018. Nonparametric identification using instrumental variables: sufficient conditions for completeness. *Econometric Theory*, **34**(3), 659–693.

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