University of Connecticut

Department of Economics Working Paper Series

Efficiency Bounds for Estimating Linear Functionals of Nonparametric Regression Models with Endogenous Regressors

Thomas A. Severini
Northwestern University

Gautam Tripathi
University of Connecticut

Working Paper 2007-18

May 2007

Abstract

Consider a nonparametric regression model $Y = \mu^*(X) + \varepsilon$, where the explanatory variables X are endogenous and ε satisfies the conditional moment restriction $E[\varepsilon|W] = 0$ w.p.1 for instrumental variables W. It is well known that in these models the structural parameter μ^* is "ill-posed" in the sense that the function mapping the data to μ^* is not continuous. In this paper, we derive the efficiency bounds for estimating linear functionals $E[\psi(X)\mu^*(X)]$ and $\int_{supp(X)} \psi(x)\mu^*(x)dx$, where ψ is a known weight function and supp(X) the support of X, without assuming μ^* to be well-posed or even identified.

Journal of Economic Literature Classification: C14

Keywords: Efficiency bounds, Linear functionals, Nonparametric regression, Endogenous regressors

We thank Gary Chamberlain, Enno Mammen, Whitney Newey, and participants at several seminars for helpful suggestions and conversations. The first author also thanks the NSF for financial support.

1. Introduction

Models containing unknown functions, typically characterized as conditional expectations, are common in economics and economists are often interested in estimating linear functionals of these unknown functions; e.g., Stock (1989) estimates the contrast between functionals of $\mathbb{E}[Y|X]$ using before-and-after policy intervention data; letting Y denote the market demand and X the price, Newey and McFadden (1994) consider estimating $\int_a^b \mathbb{E}[Y|X=x]\,dx$, the approximate change in consumer surplus for a given price change; additional examples can be found in Brown and Newey (1998) and Ai and Chen (2005a, 2005b).

However, in models where variables are determined endogenously, unknown functions cannot always be interpreted as conditional expectations which complicates the problem of estimating their linear functionals. For instance, market demand functions are not identifiable as conditional expectations because prices are endogenous. Hence, simply integrating an estimator of the conditional expectation of equilibrium quantity given equilibrium price over a certain interval will not lead to a consistent estimator of the change in consumer surplus.

The main objective of this paper is to derive the efficiency bounds for estimating certain linear functionals of an unknown structural function when the latter is not itself a conditional expectation. To set up the problem, consider the nonparametric regression model

$$Y = \mu^*(X) + \varepsilon, \qquad \mathbb{E}[\varepsilon|W] = 0 \text{ w.p.1}, \tag{1.1}$$

where X is a vector of regressors some or all of which are endogenous and W denotes the vector of instrumental variables (IV's); since exogenous explanatory variables act as their own instruments, W and X can have elements in common. The functional form of μ^* is unknown; we only assume that it lies in $L_2(X)$, the set of real-valued functions of X that are square integrable with respect to the distribution of X. Endogeneity of regressors means that μ^* cannot be a conditional expectation function because W does not contain all of X; of course, if $W = X$ so that there are no endogenous regressors, then $\mu^*(X) = \mathbb{E}[Y|X]$.

Even if the structural parameter μ^* in (1.1) is identified, i.e., uniquely defined, it is said to be "ill-posed" because the function that maps the data to μ^* is not continuous; see Lemma 2.4 of Severini and Tripathi (2006), hereafter abbreviated as ST, for additional properties of this mapping. Although μ^* may be ill-posed and hence difficult to estimate, in this paper we focus on obtaining the efficiency bounds for estimating its functionals $\mathbb{E}[\psi(X)\mu^*(X)]$ and $\int_{\mathrm{supp}(X)} \psi(x)\mu^*(x)\,dx$, where ψ is a known weight function and supp(X) the support of X.[1]

In addition to the papers cited earlier, recent works on nonparametric IV estimation include Ai and Chen (2003), Blundell and Powell (2003), Newey and Powell (2003), Florens, Johannes, and van Bellegem (2005), Hall and Horowitz (2005), Darolles, Florens, and Renault

[1]For $\int_{\mathrm{supp}(X)} \psi(x)\mu^*(x)\,dx$ to make sense it is implicitly understood that X is continuously distributed; the expectation functional $\mathbb{E}[\psi(X)\mu^*(X)]$ is of course well defined even when some components of X are discrete.

(2006), and the references therein. Our main contribution to this literature is to derive variational and non-variational, i.e., closed form, expressions for the efficiency bounds for estimating $\mathbb{E}[\psi(X)\mu^*(X)]$ and $\int_{\text{supp}(X)} \psi(x)\mu^*(x)\,dx$ without assuming that μ^* is well-posed or even identified. Some useful examples are also developed to illustrate the insights resulting from the efficiency bound calculations; for instance, we are able to characterize a condition that is necessary for $n^{1/2}$-estimability of these functionals when they are identified.[2] We conclude the paper by motivating the conjecture that plug-in estimators of these functionals based on a suitable estimator of μ^* can be asymptotically efficient. To the best of our knowledge, the results obtained in this paper are new and cannot be found elsewhere in the literature.

For the remainder of the paper, $\mathbb{E}[\psi(X)\mu^*(X)]$ and $\int_{\text{supp}(X)} \psi(x)\mu^*(x)\,dx$ are written simply as $\mathbb{E}[\psi\mu^*]$ and $\int \psi\mu^*$, respectively. These and other instances of functional notation, where arguments taken by functions are suppressed, should not cause any confusion.

2. Efficiency bounds for linear functionals

The efficiency bounds we obtain are most cleanly characterized in terms of operators on Hilbert spaces. So let $L_2(Y, X, W)$ be the set of real valued functions of (Y, X, W) that are square integrable with respect to the joint distribution of (Y, X, W) (recall that $L_2(Y, X, W)$ is a Hilbert space with inner product $\langle f_1, f_2\rangle_{L_2(Y,X,W)} := \mathbb{E}[f_1 f_2]$) and $P_{\mathcal{A}} : L_2(Y, X, W) \to \mathcal{A}$ denote orthogonal projection onto $\mathcal{A} \subseteq L_2(Y, X, W)$; hence, we can write $\mathbb{E}[f(Y, X, W)|W]$ as $P_{L_2(W)}f$ with much economy of notation. Let T be the restriction of $P_{L_2(W)}$ to $L_2(X)$, i.e., $T : L_2(X) \to L_2(W)$ is the bounded linear operator given by $Ta = \mathbb{E}[a(X)|W]$; its adjoint $T' : L_2(W) \to L_2(X)$ is the bounded linear map $T'b = \mathbb{E}[b(W)|X]$. The domain, range, and null space of T are $\mathcal{D}(T)$, $\mathcal{R}(T)$, and $\mathcal{N}(T)$, respectively; the orthogonal complement of a set A is denoted by $A^\perp$ and its closure in the norm topology is $\text{cl}(A)$. Throughout the paper the letter c denotes a generic constant that may vary from case to case.

2.1. Ill-posedness and $n^{1/2}$-estimability.

We begin by considering efficient estimation of $\mathbb{E}[\psi\mu^*]$, where the weight function ψ satisfies the following condition.

Assumption 2.1. *$\psi \in \mathcal{R}(T')$, i.e., there exists a $\delta^* \in L_2(W)$ such that $T'\delta^* = \psi$.*

In Section 4 of their paper, ST show that $\mathbb{E}[\psi\mu^*]$ is identified, i.e., uniquely defined, if and only if $\psi \in \mathcal{N}(T)^\perp$. Since $\mathcal{N}(T)^\perp = \text{cl}(\mathcal{R}(T'))$, identification requires that $\psi \in \text{cl}(\mathcal{R}(T'))$. Thus, Assumption 2.1 strengthens the identification condition. As shown next in Lemma 2.1, this is necessary because expectation functionals corresponding to $\psi \in \text{cl}(\mathcal{R}(T')) \setminus \mathcal{R}(T')$ cannot

[2] While it is well known that evaluation functionals in large dimensional spaces can be identified but not $n^{1/2}$-estimable, the result that this can also happen with expectation functionals seems to be a new insight.

be estimated at the $n^{1/2}$-rate even though they are identified.[3] Notice that since Assumption 2.1 does not require T' to be injective, δ^* above may not be uniquely defined. Also, if there are no endogenous regressors then Assumption 2.1 holds because T' is then just the identity map.

Lemma 2.1. *$\psi \in \mathcal{R}(T')$ is necessary for $\mathbb{E}[\psi\mu^*]$ to be $n^{1/2}$-estimable.*

Since $\mathcal{R}(T')$ is closed if and only if $\mathcal{R}(T)$ is closed (van der Vaart, 1991, p. 184),

$$\mathrm{cl}(\mathcal{R}(T')) \setminus \mathcal{R}(T') = \varnothing \iff \mathcal{R}(T') \text{ closed} \iff \mathcal{R}(T) \text{ closed} \iff \mu^* \text{ well-posed},$$

where the equivalence between the closure of $\mathcal{R}(T)$ and well-posedness of μ^* follows from Lemma 2.4 of ST. Therefore, if μ^* is ill-posed or, equivalently, $\mathcal{R}(T)$ is not closed, then there exists at least one expectation functional of μ^* that is identified but not $n^{1/2}$-estimable; see Example 2.2 for a nice illustration. Of course, if μ^* is well-posed then every identified expectation functional of μ^* is $n^{1/2}$-estimable.

The following example shows how $\mathcal{R}(T)$ and $\mathcal{R}(T')$ look in a Gaussian setup.

Example 2.1. Let X and W be jointly normal with mean zero and variance $\left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]$, where $\rho \in (-1,1) \setminus \{0\}$. Furthermore, let ϕ be the standard normal density, $H_j(x) := (-1)^j \phi^{(j)}(x)/\phi(x)$ the jth Hermite polynomial, and $h_j := H_j/\sqrt{j!}$ its normalized version.[4] Using the reproducing property of Hermite polynomials, see Example 2.4 of ST, it is straightforward to show that T and T' are injective and compact, in fact, Hilbert-Schmidt, with singular system $\{(\rho^j, h_j(X), h_j(W)) : j \in \mathbb{N}\}$. Therefore, by Lemma B.1 and Corollary B.1,[5]

$$\begin{aligned} \mathcal{R}(T) &= \{b \in L_2(W) : \sum_{j=0}^{\infty} \langle b, h_j \rangle_{L_2(W)}^2 \rho^{-2j} < \infty\} \overset{\text{dense}}{\subsetneq} L_2(W) \\ \mathcal{R}(T') &= \{a \in L_2(X) : \sum_{j=0}^{\infty} \langle a, h_j \rangle_{L_2(X)}^2 \rho^{-2j} < \infty\} \overset{\text{dense}}{\subsetneq} L_2(X). \end{aligned} \tag{2.1}$$

Since $\mathcal{R}(T)$ and $\mathcal{R}(T')$ are dense albeit proper subspaces of $L_2(W)$ and $L_2(X)$, respectively, they cannot be closed. Moreover, their elements are infinitely differentiable with each derivative being square-integrable. To see this, let $b \in \mathcal{R}(T)$ and $b^{(k)}$ denote its kth derivative. Then, since $H_j^{(1)} = jH_{j-1}$, it is straightforward to show that $b^{(k)} = \sum_{j=k}^{\infty} \langle b, h_j \rangle_{L_2(W)} \sqrt{(j)_k} h_{j-k}$ for every $k \in \mathbb{N}$, where $(j)_k := j(j-1)\dots(j-k+1)$. Therefore, b is infinitely differentiable. Furthermore, $\|b^{(k)}\|_{L_2(W)}^2 = \sum_{j=k}^{\infty} \langle b, h_j \rangle_{L_2(W)}^2 (j)_k < \infty$ for each $k \in \mathbb{N}$ since $\lim_{j\to\infty} \rho^{2j}(j)_k = 0$; hence, each $b^{(k)}$ is square integrable. Same results hold for $\mathcal{R}(T')$ as well. □

[3] Ritov and Bickel (1990, p. 936) have a similar looking result. They define a class P of large dimensional parametric models and show that if the true model lies in $\mathrm{cl}(P) \setminus P$ then it cannot be consistently estimated.

[4] In contrast to the statistics literature, mathematicians seem to prefer $\tilde{H}_j(x) := (-1)^j e^{x^2}(d^j e^{-x^2}/dx^j)$ as the definition of Hermite polynomials; see, e.g., (6.1.3) of Andrews, Askey, and Roy (1999). It is easy to show that $H_j(x) = 2^{-j/2}\tilde{H}_j(x/\sqrt{2})$; this fact is used in Example 2.2.

[5] Denseness of $\mathcal{R}(T)$ and $\mathcal{R}(T')$ follows by injectivity of T and T' and the fact that $\mathcal{N}(T)^{\perp} = \mathrm{cl}(\mathcal{R}(T'))$.

Example 2.1 can be used to describe some weight functions that satisfy Assumption 2.1.

Example 2.2 (Example 2.1 contd.). Since $\langle X^k, h_j\rangle_{L_2(X)} = 0$ for $j > k$, all polynomials lie in $\mathcal{R}(T')$; consequently, $\mathbb{E}[X^k\mu^*]$ can be efficiently estimated for every $k < \infty$. In contrast, the indicator function $\mathbb{1}_{(-\infty,d]}$, where $d < \infty$ is a known constant, is not an element of $\mathcal{R}(T')$; hence, $\mathbb{E}[\mathbb{1}_{(-\infty,d]}\mu^*]$ is not $n^{1/2}$-estimable even though it is identified.[6] This suggests that in the presence of unknown functions of endogenous regressors, identifiability of finite dimensional parameters may not be sufficient to ensure their $n^{1/2}$-estimability. Since such problems do not arise if $W = X$, this example illustrates the importance of being careful about identification and ill-posedness when dealing with nonparametric IV models.

It only remains to show that $\mathbb{1}_{(-\infty,d]} \notin \mathcal{R}(T')$; although this follows from Example 2.1, we provide a direct verification because the same logic is also used in Example 2.3. So let $\psi_d := \mathbb{1}_{(-\infty,d]}$ and Φ be the standard normal cdf. Then, using the fact that

$$\int_{-\infty}^{d} H_j(x)\phi(x)\,dx = \begin{cases} \Phi(d) & \text{if } j = 0 \\ -H_{j-1}(d)\phi(d) & \text{if } j \geq 1, \end{cases} \tag{2.2}$$

it is easily verified that

$$\sum_{j=0}^{\infty} \langle \psi_d, h_j\rangle_{L_2(X)}^2 \rho^{-2j} = \Phi^2(d) + \frac{\phi^2(d)}{\rho^2}\sum_{j=0}^{\infty} \frac{H_j^2(d)}{(j+1)!\rho^{2j}}.$$

But since $(j+1)\rho^{2j} < 1$ for all sufficiently large j, there exists a positive integer N such that

$$\sum_{j=0}^{\infty} \frac{H_j^2(d)}{(j+1)!\rho^{2j}} \geq \sum_{j=N}^{\infty} \frac{H_j^2(d)}{j!} = \sum_{j=N}^{\infty} \frac{\tilde{H}_j^2(d/\sqrt{2})}{j!2^j}, \tag{2.3}$$

where the last equality follows upon recalling that $\tilde{H}_j(x/\sqrt{2}) = 2^{j/2}H_j(x)$. Therefore, since $\sum_{j=0}^{\infty} \tilde{H}_j^2(x) r^j/(j!2^j) < \infty$ for every $x \in \mathbb{R}$ if and only if $|r| < 1$, see, e.g., the second proof of (6.1.13) in Andrews et al., it follows that $\sum_{j=0}^{\infty}\langle \psi_d, h_j\rangle_{L_2(X)}^2\rho^{-2j} = \infty$; hence, $\psi_d \notin \mathcal{R}(T')$. $\square$

The next example provides some additional intuition behind why expectation functionals of the form $\mathbb{E}[\mathbb{1}_{(-\infty,d]}\mu^*]$ are not $n^{1/2}$-estimable in a Gaussian setting.

[6] Of course, indicator functions may lie in $\mathcal{R}(T')$ and their expectation functionals can be $n^{1/2}$-estimable if X and W are not jointly Gaussian.

Example 2.3 (Example 2.2 contd.)**.** Let $\theta_d^* := \mathbb{E}[\mathbb{1}_{(-\infty,d]}\mu^*]$ and from Example 2.4 of ST recall that $\mu^*(X) = \sum_{j=0}^{\infty} \rho^{-j}\mathbb{E}[Yh_j(W)]h_j(X)$. Using (2.2), it is then straightforward to show that

$$\begin{aligned}\theta_d^* &= \Phi(d)\mathbb{E}Y - \phi(d)\sum_{j=0}^{\infty}\frac{\mathbb{E}[Yh_{j+1}(W)]}{\rho^{j+1}\sqrt{j+1}}h_j(d)\\ &= \mathbb{E}[Y(\Phi(d) - \phi(d)\sum_{j=0}^{\infty}\frac{h_{j+1}(W)}{\rho^{j+1}\sqrt{j+1}}h_j(d))]\\ &:= \mathbb{E}[YQ_d(W)].\end{aligned}$$

Now consider the estimator $\hat\theta_d := \sum_{j=1}^{n} Y_jQ_d(W_j)/n$, where we have assumed that ρ is known to keep things simple. Clearly, $\hat\theta_d$ is consistent for θ_d^*. Moreover, assuming that $\mathrm{var}[Y|W]$ is bounded away from zero,

$$\mathrm{var}[n^{1/2}\hat\theta_d] = \mathrm{var}[Y_1Q_d(W_1)] \geq \mathbb{E}\,\mathrm{var}[Y_1Q_d(W_1)|W_1] \geq c\mathbb{E}[Q_d^2(W_1)].$$

But, by the orthonormality of Hermite polynomials,

$$\mathbb{E}[Q_d^2(W_1)] = \Phi^2(d) + \frac{\phi^2(d)}{\rho^2}\sum_{j=0}^{\infty}\frac{h_j^2(d)}{\rho^{2j}(j+1)} = \Phi^2(d) + \frac{\phi^2(d)}{\rho^2}\sum_{j=0}^{\infty}\frac{H_j^2(d)}{(j+1)!\rho^{2j}}.$$

Hence, by the same argument used to show that the RHS of (2.3) is unbounded when $\rho^2 < 1$, it follows that $\mathrm{var}[n^{1/2}\hat\theta_d] = \infty$. Therefore, $\hat\theta_d$ is not $n^{1/2}$-consistent.[7] □

2.2. **Efficiency bound.** We are now ready to determine the efficiency bound for estimating $\mathbb{E}[\psi\mu^*]$ when ψ satisfies Assumption 2.1. For maximum generality, the bound is derived under minimal assumptions on μ^*. In particular, μ^* is allowed to be underidentified, i.e., $\mathcal{N}(T) \neq \{0\}$, and ill-posed, i.e., $\mathcal{R}(T)$ is not assumed to be closed.

Since the μ^* appearing in (1.1) is not assumed to be identified, think of it as a fixed but arbitrary element of $L_2(X)$ satisfying the conditional moment restriction[8]

$$T\mu^* = P_{L_2(W)}Y \iff TP_{\mathcal{N}(T)^\perp}\mu^* = P_{L_2(W)}Y \tag{2.4}$$

because $T\mu^* = T(P_{\mathcal{N}(T)^\perp} + P_{\mathcal{N}(T)})\mu^* = TP_{\mathcal{N}(T)^\perp}\mu^*$. Moreover, since $\psi \in \mathcal{N}(T)^\perp$ by Assumption 2.1, for every $\mu \in P_{\mathcal{N}(T)^\perp}\mu^* + \mathcal{N}(T)$ we also have

$$\langle\psi,\mu\rangle_{L_2(X)} = \langle\psi, P_{\mathcal{N}(T)^\perp}\mu^*\rangle_{L_2(X)} = \langle\psi,\mu^*\rangle_{L_2(X)};$$

[7]This argument breaks down if $\rho^2 = 1$, i.e., no endogenous regressors, or $d = \infty$, i.e., $\theta_\infty^* = \mathbb{E}Y$; in both these cases $n^{1/2}$-consistent estimation of θ_d^* is possible. Excluding these two special cases, the result that the variance of $\hat\theta_d$ goes to zero at a rate slower than $1/n$ remains valid even if Q_d is replaced by its truncated version $Q_{d,m_n}(W) := \Phi(d) - \phi(d)\sum_{j=0}^{m_n}\rho^{-(j+1)}(j+1)^{-1/2}h_{j+1}(W)h_j(d)$, where m_n is any sequence of positive integers such that $\lim_{n\to\infty} m_n = \infty$.

[8]Following Section 3 of ST, $P_{\mathcal{N}(T)^\perp}\mu^*$ is the "identifiable part" of μ^*.

i.e., $\mathbb{E}[\psi\mu]$ is uniquely defined for every $\mu \in P_{\mathcal{N}(T)^\perp}\mu^* + \mathcal{N}(T)$. Hence, without loss of generality, let $\theta^* := \mathbb{E}[\psi P_{\mathcal{N}(T)^\perp}\mu^*]$ denote the parameter of interest. As shown later, each μ^* satisfying (2.4) leads to the same efficiency bound for estimating θ^*. Subsequent results simplify accordingly if μ^* is identified to begin with, i.e., $\mathcal{N}(T) = \{0\}$; see, e.g., Corollary 2.1 and 2.2.

To facilitate presentation, we express θ^* as the solution to a moment condition; namely, we obtain the efficiency bound for estimating θ^* in the model

$$\mathbb{E}\, g(X, \theta^*, \mu^*) = 0, \tag{2.5}$$

where $g(X, \theta^*, \mu^*) := \psi P_{\mathcal{N}(T)^\perp}\mu^* - \theta^*$; henceforth, $g := g(X, \theta^*, \mu^*)$ for notational convenience.

Let $\tilde{\varepsilon} := Y - P_{\mathcal{N}(T)^\perp}\mu^*$, so that $P_{L_2(W)}\tilde{\varepsilon} = 0$, and $\Omega := P_{L_2(W)}\tilde{\varepsilon}^2$ be the scedastic function. The next assumption bounds Ω away from zero and infinity.

Assumption 2.2. $0 < \inf_{w \in \operatorname{supp}(W)} \Omega(w) \le \sup_{w \in \operatorname{supp}(W)} \Omega(w) < \infty$.

From now on, we write $\mho := 1/\Omega$ instead of $1/\Omega$ to avoid confusing the reciprocal with an operator inverse.

As shown later, the variance bound for estimating θ^* is the squared-length of an orthogonal projection onto $\mathrm{cl}(\dot{\mathcal{M}}) + L_2(W)$, the tangent space of nonparametric score functions, where $\dot{\mathcal{M}} := \{f \in L_2(W)^\perp : P_{L_2(W)}(\tilde{\varepsilon} f) \in \mathcal{R}(T)\}$; in the appendix we show that

$$\mathrm{cl}(\dot{\mathcal{M}}) = \{f \in L_2(W)^\perp : P_{L_2(W)}(\tilde{\varepsilon} f) \in \mathrm{cl}(\mathcal{R}(T))\}.$$

If there are no endogenous regressors, then $\mathrm{cl}(\dot{\mathcal{M}}) = L_2(W)^\perp = L_2(X)^\perp$. Therefore, as made clear in Examples 2.4 and 2.5, the size of $\mathrm{cl}(\dot{\mathcal{M}})$ is a measure of the information contained in the conditional moment restriction (2.4); smaller $\mathrm{cl}(\dot{\mathcal{M}})$ means more information.

Theorem 2.1. *Let Assumptions* 2.1 *and* 2.2 *hold. Then, the efficiency bound for estimating* θ^* *is given by*

$$\mathbb{E}[P_{\mathrm{cl}(\dot{\mathcal{M}}) + L_2(W)}(\tilde{\varepsilon} P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g)]^2, \tag{2.6}$$

where $\delta^* \in L_2(W)$ *satisfies* $T'\delta^* = \psi$.

If $\mu_1^* \neq \mu_2^*$ satisfy (2.4), then

$$P_{\mathcal{N}(T)^\perp}\mu_1^* - P_{\mathcal{N}(T)^\perp}\mu_2^* = P_{\mathcal{N}(T)^\perp}(\mu_1^* - \mu_2^*) = 0$$

because $\mu_1^* - \mu_2^* \in \mathcal{N}(T)$; similarly, $P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*$ is uniquely defined for every δ^* satisfying $T'\delta^* = \psi$ because $\mathrm{cl}(\mathcal{R}(T)) = \mathcal{N}(T')^\perp$. Therefore, $\mathrm{cl}(\dot{\mathcal{M}})$ and the efficient influence function $P_{\mathrm{cl}(\dot{\mathcal{M}}) + L_2(W)}(\tilde{\varepsilon} P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g)$ are invariant to choice of μ^* and δ^*, implying that the above efficiency bound is robust to underidentification of μ^* and δ^*. Similarly, since $\mathcal{R}(T)$ enters (2.6) only via $\mathrm{cl}(\mathcal{R}(T))$, the same bound holds whether μ^* is ill-posed or not.

Example 2.4 (Efficiency bound for estimating $\mathbb{E}Y$)**.** Suppose $\psi = 1$. Then $\theta^* = \mathbb{E}Y$ irrespective of whether μ^* is identified or not. Therefore, by Theorem 2.1 and the fact that $\mathrm{cl}(\dot{\mathcal{M}}) + L_2(W)$ is closed, the efficiency bound for estimating $\mathbb{E}Y$ is given by

$$\mathbb{E}[P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(Y - \mathbb{E}Y)]^2 = \operatorname{var} Y - \mathbb{E}[P_{\dot{\mathcal{M}}^\perp \cap L_2(W)^\perp}(Y - \mathbb{E}Y)]^2.$$

Hence, the sample mean is asymptotically efficient if there are no endogenous regressors. $\square$

The following corollary of Theorem 2.1 is immediate.

Corollary 2.1. *If μ^* is identified, i.e., $\mathcal{N}(T) = \{0\}$, then* (2.6) *can be written as*

$$\mathbb{E}[P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(\varepsilon P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g)]^2,$$

where $\varepsilon = Y - \mu^*$, $g = \psi\mu^* - \theta^*$, and $\dot{\mathcal{M}} = \{f \in L_2(W)^\perp : P_{L_2(W)}(\varepsilon f) \in \mathcal{R}(T)\}$.

If there are no endogenous regressors, then $\delta^* = \psi$ and $\mu^* = P_{L_2(X)}Y$, and the efficiency bound for estimating θ^* reduces to $\operatorname{var}[\psi Y]$; see Chamberlain (1992, p. 572). This makes sense because if $W = X$, then $\mathbb{E}[\psi\mu^*] = \mathbb{E}[\psi Y]$; but the efficiency bound for estimating $\mathbb{E}[Z]$ when the random variable Z is fully observed is $\operatorname{var}[Z]$. Therefore, when there are no endogenous regressors the efficiency bound for estimating θ^* is given by $\operatorname{var}[\psi Y]$.

Although Theorem 2.1 and Corollary 2.1 provide precise variational characterizations of the efficiency bound for estimating θ^*, in practice it may not be easy to use these results to construct efficient estimators or to determine whether a proposed estimator is asymptotically efficient unless a closed form for $P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}$ is available. Fortunately, an explicit expression for orthogonal projections onto $\mathrm{cl}(\dot{\mathcal{M}}) + L_2(W)$ can be obtained by using Lemma 2.2, which may be of independent interest.

Henceforth, let $(T'\mho T)^+$ denote the Moore-Penrose generalized inverse of $T'\mho T$, see, e.g., Engl, Hanke, and Neubauer (2000, Section 2.1), and I be the identity operator; keep in mind that $\mathcal{D}((T'\mho T)^+) = \mathcal{R}(T'\mho T) + \mathcal{R}(T'\mho T)^\perp$.

Lemma 2.2. *Let Assumption* 2.2 *hold and $f \in L_2(Y, X, W)$ be such that $T'\mho P_{L_2(W)}(\tilde{\varepsilon} f)$ lies in the domain of $(T'\mho T)^+$. Then,*

$$P_{\mathrm{cl}(\dot{\mathcal{M}})}f = f - P_{L_2(W)}f - \tilde{\varepsilon}(I - \mho T(T'\mho T)^+T')\mho P_{L_2(W)}(\tilde{\varepsilon} f).$$

Since $\dot{\mathcal{M}} \perp L_2(W)$, it follows that $P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)} = P_{\mathrm{cl}(\dot{\mathcal{M}})} + P_{L_2(W)}$. Therefore, an immediate corollary of Lemma 2.2 is that

$$P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}f = f - \tilde{\varepsilon}(I - \mho T(T'\mho T)^+T')\mho P_{L_2(W)}(\tilde{\varepsilon} f). \tag{2.7}$$

Hence, we can use (2.7) to derive a closed form for the efficiency bound in Theorem 2.1.

Theorem 2.2. *Let $\psi \in \mathcal{D}((T'\mho T)^+) \cap \mathcal{N}(T)^\perp$ and $T'\mho P_{L_2(W)}(\tilde{\varepsilon} g) \in \mathcal{D}((T'\mho T)^+)$.*[9] *Then, under the assumptions maintained in Theorem* 2.1, (2.6) *can be written as*

$$\mathbb{E}[\tilde{\varepsilon}\mho T(T'\mho T)^+\psi + g - \tilde{\varepsilon}(I - \mho T(T'\mho T)^+T')\mho P_{L_2(W)}(\tilde{\varepsilon} g)]^2. \tag{2.8}$$

When μ^* is identified, the closed form of the bound can be obtained by replacing $\tilde{\varepsilon}$ with ε and $(T'\mho T)^+$ with $(T'\mho T)^{-1}$ (because $\mathcal{N}(T'\mho T) = \mathcal{N}(T)$); i.e.,

Corollary 2.2. *If μ^* is identified, then* (2.8) *can be written as*

$$\mathbb{E}[\varepsilon\mho T(T'\mho T)^{-1}\psi + g - \varepsilon(I - \mho T(T'\mho T)^{-1}T')\mho P_{L_2(W)}(\varepsilon g)]^2.$$

An interesting feature of the non-variational characterization is that it does not depend upon the "nuisance" parameter δ^*. Apart from this, it also leads to some additional insight behind the form of the bound. To see this, assume that μ^* is identified. Then, from Corollary 2.2, the efficient influence function for estimating θ^* is given by

$$[g - \varepsilon\mho P_{L_2(W)}(\varepsilon g)] + \varepsilon\mho T(T'\mho T)^{-1}(\psi + T'\mho P_{L_2(W)}(\varepsilon g)). \tag{2.9}$$

But a look at the proof of Theorem 2.1 reveals that the efficiency bound for estimating θ^* when μ^* is fully known is given by $\mathbb{E}[g - \varepsilon\mho P_{L_2(W)}(\varepsilon g)]^2$. Thus the first term of (2.9), which has a very intuitive control variate interpretation, represents the contribution of $P_{L_2(W)}\varepsilon = 0$ if μ^* is assumed known whereas the second term represents the penalty for not knowing its functional form. Since the two terms are orthogonal, the efficiency bound can also be written as

$$\mathbb{E}[g - \varepsilon\mho P_{L_2(W)}(\varepsilon g)]^2 + \mathbb{E}[\varepsilon\mho T(T'\mho T)^{-1}(\psi + T'\mho P_{L_2(W)}(\varepsilon g))]^2.$$

Therefore, the efficiency bound for estimating θ^* when μ^* is known will be equal to the efficiency bound for estimating θ^* when μ^* is unknown if and only if

$$T(T'\mho T)^{-1}(\psi + T'\mho P_{L_2(W)}(\varepsilon g)) = 0. \tag{2.10}$$

But since (2.10) is a very restrictive condition, e.g., it may not hold even when $W = X$, adaptive (meaning invariance with respect to knowledge of μ^* or lack thereof) estimation of θ^* appears for all practical purposes to be impossible.

Next, we obtain the efficiency bound for estimating $\int \psi\mu^*$. The proofs of Theorems 2.3–2.4 are very similar to those of Theorems 2.1–2.2 and are therefore omitted.

Theorem 2.3. *Let Assumption* 2.2 *hold and assume there exists a $\delta^* \in L_2(W)$ such that $T'\delta^* = \psi/h$, where h is the unknown Lebesgue density of X.*[10] *Then, the efficiency bound for estimating $\int \psi\mu^*$ is given by $\mathbb{E}[P_{\mathrm{cl}(\dot{\mathcal{M}})}(\tilde{\varepsilon} P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*)]^2$. The bound when μ^* is identified is obtained by replacing $\tilde{\varepsilon}$ with ε.*

---

[9] Since $\mathcal{D}((T'\mho T)^+) \cap \mathcal{N}(T)^\perp = \mathcal{R}(T'\mho T)$ by Lemma B.4, $T'\mho P_{L_2(W)}(\varepsilon g) \in \mathcal{R}(T'\mho T)$ when μ^* is identified.

[10] ST(Section 4) show that $\int \psi\mu^*$ is identified if and only if $\psi/h \in \mathcal{N}(T)^\perp$.

In case of no endogeneity the above bound reduces to $\mathbb{E}[\psi\varepsilon/h]^2$, a result obtained earlier by Severini and Tripathi (2001, Section 7). As before, Lemma 2.2 can be used to derive a closed form expression for the efficiency bound.

Theorem 2.4. *Let $\psi/h \in \mathcal{D}((T'\mho T)^+) \cap \mathcal{N}(T)^\perp$. Then, under the assumptions maintained in Theorem 2.3, the efficiency bound obtained there can be written as $\mathbb{E}[\tilde{\varepsilon}\mho T(T'\mho T)^+(\psi/h)]^2$. The bound when μ^* is identified is obtained by replacing $\tilde{\varepsilon}$ with ε and $(T'\mho T)^+$ with $(T'\mho T)^{-1}$.*

The methodology developed in this paper can be used to obtain efficiency bounds for other parameters of interest as well.

Example 2.5 (Efficiency bound for probabilities)**.** Let the vector Z contain Y and the distinct components of X and W. Then, modifying the proof of Theorem 2.1, it can be shown that the efficiency bound for estimating $p := \Pr(Z \in A)$, where A is a known region, is given by

$$\mathbb{E}[P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(\mathbb{1}(Z \in A) - p)]^2 = p[1-p] - \mathbb{E}[P_{\dot{\mathcal{M}}^\perp \cap L_2(W)^\perp}(\mathbb{1}(Z \in A) - p)]^2.$$

Hence, unless there are no endogenous regressors, the empirical measure $\sum_{j=1}^n \mathbb{1}(Z_j \in A)/n$ is not an efficient estimator of p. □

2.3. **Approximating the bound.** In this section we provide some justification to show that the efficiency bound in Theorem 2.1 is attainable; similar results have been obtained earlier by Wong (1986, Section 5.2) and Chamberlain (1987, Section 4.2). To do so, we assume the existence of a sequence of parametric models satisfying (2.4) and show that the corresponding efficiency bound for estimating θ^* approaches (2.6) as the models get richer. Since the bound is attainable in the parametric case, this suggests that our semiparametric efficiency bound is achievable as well. A similar argument also works for estimating $\int \psi\mu^*$.

So let $P_{\mathcal{N}(T)^\perp}\mu^*$ be embedded in a smooth parametric family $\mathcal{F}_p := \{m(\cdot,\beta) : \beta \in \mathbb{R}^p\}$, i.e., $(P_{\mathcal{N}(T)^\perp}\mu^*)(X) = m(X,\beta^*)$ for all sufficiently large p. Also, let $\mathcal{S}_p$ denote the set spanned by the coordinates of the parametric gradient $\nabla_\beta m(X,\beta^*)$ and define $T_p : \mathcal{S}_p \to L_2(W)$ to be the restriction of T to $\mathcal{S}_p$. Then we have the following result.

Lemma 2.3. *Let $P_{\mathcal{N}(T)^\perp}\mu^*$ be embedded in $\mathcal{F}_p$ and Assumptions 2.1 and 2.2 hold. Then, the efficiency bound for estimating θ^* is given by*

$$\mathbb{E}[P_{\dot{\mathcal{M}}_p+L_2(W)}(\tilde{\varepsilon}P_{\mathcal{R}(T_p)}\delta^* + g)]^2,$$

where $\dot{\mathcal{M}}_p := \{f \in L_2(W)^\perp : P_{L_2(W)}(\tilde{\varepsilon}f) \in \mathcal{R}(T_p)\}$ and $\delta^* \in L_2(W)$ satisfies $T'\delta^* = \psi$.

Next, we show that if the sequence of parametric families is dense in a certain sense then the parametric and semiparametric efficiency bounds can get arbitrarily close.

Lemma 2.4. *Let the parametric family $\mathcal{F}_p$ be nested so that $\mathcal{R}(T_p) \uparrow \mathrm{cl}(\mathcal{R}(T))$ as $p \to \infty$. Then, under the assumptions maintained in Lemma 2.3,*

$$\lim_{p\to\infty} \mathbb{E}[P_{\dot{\mathcal{M}}_p+L_2(W)}(\tilde{\varepsilon}P_{\mathcal{R}(T_p)}\delta^* + g)]^2 = \mathbb{E}[P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(\varepsilon P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g)]^2.$$

Since the $\mathcal{F}_p$'s are nested, the parametric efficiency bound is non-decreasing in p. Therefore, by Lemma 2.4, the efficiency bound under any particular parametric model can be no greater than the semiparametric efficiency bound.

3. Concluding Remarks

We conclude by giving some intuition as to why a plug-in estimator of θ^* based on a suitable estimator of μ^* can be asymptotically efficient;[11] similar justification holds for estimating $\int \psi\mu^*$ as well. Since efficient estimation of θ^* is particularly simple when $\psi = 1$ and T' is injective, for the remainder of this section we assume that either $\psi \neq 1$ or that $\mathcal{N}(T') \neq \{0\}$ to keep the estimation problem interesting.[12]

So let $\hat{\mu}$ denote a consistent estimator of μ^* (assumed to be identified) such that $\hat{\theta} := \sum_{j=1}^{n} \psi(X_j)\hat{\mu}(X_j)/n$ converges in probability to $\mathbb{E}[\psi\mu^*]$ as $n \to \infty$. Since every $\hat{\mu}$ may not lead to an efficient estimator of θ^*, the former will have to satisfy additional regularity conditions for $\hat{\theta}$ to be asymptotically efficient.[13]

The motivation behind the asymptotic efficiency of $\hat{\theta}$ comes from the fact that the pathwise derivative (see Newey (1994, p. 1351) for the definition) of $\mathbb{E}[\psi\mu^*]$, when the latter is regarded as a function of the true but unknown distribution of the data, is just $\varphi :=$

[11] Although an efficient estimator of θ^* can also be based on the efficient influence function by using the moment condition $\mathbb{E}[P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(\varepsilon P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g)] = 0$ and the closed form expression of the efficient influence function given in (2.9), this will be significantly more complicated than a plug-in estimator because apart from μ^* it requires the nonparametric estimation of additional functions and operators.

[12] If T' is injective, the efficiency bound for estimating θ^* is given by $\mathbb{E}[\tilde{\varepsilon}\delta^* + g]^2$; since $\mathcal{N}(T') = \{0\}$ implies that $\mathcal{R}(T)$ is dense in $L_2(W)$, this follows directly from Theorem 2.1 by setting $\mathrm{cl}(\mathcal{R}(T)) = L_2(W)$. If, in addition, $\psi = 1$, so that $\delta^* = 1$ is the unique solution to $T'\delta^* = 1$, the bound reduces to $\mathrm{var}[Y]$. Therefore, if $\psi = 1$ and T' is injective, θ^* can be efficiently estimated by $\sum_{j=1}^{n} Y_j/n$, irrespective of whether μ^* is identified or not. The intuition behind why $\mathrm{var}[Y]$, the efficiency bound for estimating $\mathbb{E}Y$ when $W = X$, is also the bound for estimating $\mathbb{E}Y$ when T' is injective, is straightforward: since the IV's enter the bound only through δ^* and $\delta^* = 1$ uniquely solves $T'\delta^* = 1$, endogeneity of regressors is not an issue for this case.

[13] Darolles et al. propose an estimator for θ^* and under the assumption that ε is homoscedastic with respect to W, show that their estimator, suitably centered and scaled, converges at rate $n^{1/2}$ to a standard Gaussian random variable. However, the value at which the estimator is centered depends on n and may not converge to θ^* at a rate faster than $n^{-1/2}$, so that their estimator may be asymptotically biased; see Theorem 4.3 and Corollary 4.1 of their paper. Thus, the efficiency bound given here, which is valid only for asymptotically unbiased estimators, does not necessarily apply to their estimator. However, it is interesting to note that the asymptotic variance of their estimator for the case $\psi = 1$, i.e., when $\theta^* = \mathbb{E}Y$, is $\mathrm{var}[\varepsilon]$, which does not match the efficiency bound for estimating $\mathbb{E}Y$ described in Example 2.4.

$P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(\varepsilon P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g)$, the efficient influence function for estimating θ^*; see (A.7) in the appendix for the proof. Consequently, as noted by Newey, the asymptotic variance of $\hat{\theta}$ is the variance of the pathwise derivative of its probability limit, i.e., $\mathbb{E}\varphi^2$.

Following Newey (1994, Section 5), sufficient "high level" conditions that $\hat{\mu}$ should satisfy so that $\hat{\theta}$ is asymptotically efficient can also be described. So let $\mathbb{P}$ be the probability measure generating the data and $\hat{\mathbb{P}}$ the empirical measure.

Lemma 3.1. *If $n^{1/2}\int \psi(\hat{\mu}-\mu^*)\,d\mathbb{P} = -n^{1/2}\int(g-\varphi)\,d\hat{\mathbb{P}}+o_p(1)$ and $n^{1/2}\int \psi(\hat{\mu}-\mu^*)\,(d\hat{\mathbb{P}}-d\mathbb{P}) = o_p(1)$, then $n^{1/2}(\hat{\theta}-\theta^*) = n^{1/2}\int \varphi\,d\hat{\mathbb{P}} + o_p(1)$.*

The first requirement on $\hat{\mu}$ determines the $n^{1/2}$-consistency and efficiency of $\hat{\theta}$ whereas the second is a stochastic equicontinuity condition; see Newey (1994, p. 1365–1366) for the intuition behind these assumptions. Although a $\hat{\mu}$ satisfying these conditions will lead to an efficient plug-in estimator of θ^*, we were unable to find such an estimator. In particular, it remains to be determined whether the estimators of μ^* proposed earlier in the literature, see Section 1 for the references, satisfy the requirements of Lemma 3.1 (especially the first one).

In conclusion, it seems reasonable to believe that plug-in estimation of $\mathbb{E}[\psi\mu^*]$ and $\int \psi\mu^*$ can be asymptotically efficient. Showing this explicitly, however, remains an open problem and a topic for future research.

Appendix A. Proofs

Proof of Lemma 2.1. Following the discussion in van der Vaart (1991, p. 185) and Newey (1994, p. 1353), we know that θ^* is not $n^{1/2}$-estimable if the derivative $\nabla\eta$ is unbounded.[14] But, by (A.6), $\nabla\eta$ is bounded on the tangent space $\dot{\mathcal{T}}$ if and only if the linear functional $J_{\psi,\tilde{\varepsilon}}$ defined in (A.4) is bounded on $\mathrm{cl}(\dot{\mathcal{M}})$. We now show that $J_{\psi,\tilde{\varepsilon}}$ is bounded if and only if $\psi \in \mathcal{R}(T')$. In fact, since sufficiency follows directly from (A.5), it only remains to show that

$$\psi \in \mathrm{cl}(\mathcal{R}(T')) \setminus \mathcal{R}(T') \implies J_{\psi,\tilde{\varepsilon}} \text{ is unbounded on } \mathrm{cl}(\dot{\mathcal{M}}).$$

We demonstrate necessity via the following example. First, assume that T and T' are both injective and Hilbert-Schmidt with singular system $\{(\lambda_j, a_j, b_j) : j \in \mathbb{N}\}$, where the λ_j's are singular values satisfying $\sum_{j=1}^{\infty}\lambda_j^2 < \infty$, and $\{a_j\}$ and $\{b_j\}$ are orthonormal bases for $L_2(X)$ and $L_2(W)$, respectively.[15] Next, let $\psi_0 := \sum_{j=1}^{\infty}\lambda_j a_j$. Clearly, $\psi_0 \notin \mathcal{R}(T')$ by Corollary B.1. However, $\psi_k := \sum_{j=1}^{k}\lambda_j a_j \in \mathcal{R}(T')$ for every $k \in \mathbb{N}$ since $T'b_j = \lambda_j a_j$. Therefore, since $\psi_k \to \psi_0$

[14] We are using notation introduced subsequently in the proof of Theorem 2.1.

[15] This holds, for instance, in the Gaussian framework of Example 2.1.

as $k \to \infty$ and $\mathcal{N}(T) = \{0\}$, we have that, for every $f \in \mathrm{cl}(\dot{\mathcal{M}})$,

$$\begin{aligned} J_{\psi_0,\varepsilon}(f) &= \lim_{k\to\infty} \langle \psi_k, T^+ P_{L_2(W)}(\varepsilon f)\rangle_{L_2(X)} = \lim_{k\to\infty} \langle T^{+\prime}\psi_k, P_{L_2(W)}(\varepsilon f)\rangle_{L_2(W)} \\ &= \lim_{k\to\infty} \langle P_{\mathrm{cl}(\mathcal{R}(T))}(\sum_{j=1}^{k} b_j), P_{L_2(W)}(\varepsilon f)\rangle_{L_2(W)} \qquad \text{(by Lemma B.3(ii))} \\ &= \sum_{j=1}^{\infty} \mathbb{E}[\varepsilon b_j f], \end{aligned}$$

since $\mathcal{R}(T)$ is dense in $L_2(W)$. Hence, $J_{\psi_0,\varepsilon}$ is unbounded on $\mathrm{cl}(\dot{\mathcal{M}})$; e.g., if $f_0 := \varepsilon\mho\sum_{j=1}^{\infty} b_j/j$, then it is easy to verify that $f_0 \in \mathrm{cl}(\dot{\mathcal{M}})$ but $J_{\psi_0,\varepsilon}(f_0) = \infty$. $\square$

Proof of Theorem 2.1. Let v_0^2 be the conditional density of $(Y, X)|W$ with respect to a product dominating measure $\lambda(dy, dx)$ and b_0^2 the density of W with respect to a dominating measure $\gamma(dw)$. Let v_t be a real-valued function on I_0, an interval containing zero, such that $v_t|_{t=0} = v_0$ and $\int_{\mathrm{supp}(Y,X)} v_t^2(y, x|w)\, d\lambda = 1$ for all $(t, w) \in I_0 \times \mathrm{supp}(W)$; similarly, b_t is a curve through b_0 satisfying $\int_{\mathrm{supp}(W)} b_t^2(w)\, d\gamma = 1$ for all $t \in I_0$. Using $\dot{\tau} = (\dot{v}, \dot{b})$ to denote the tangent vector to (v_t, b_t) at $t = 0$, we have

$$\begin{aligned} \dot{v} \in \dot{\mathcal{V}} &:= \{S_{\dot{v}} \in L_2(Y, X, W) : \mathbb{E}[S_{\dot{v}}(Y, X, W)|W] = 0 \text{ w.p.1}\} \\ \dot{b} \in \dot{\mathcal{B}} &:= \{S_{\dot{b}} \in L_2(W) : \mathbb{E}[S_{\dot{b}}(W)] = 0\}, \end{aligned}$$

where $S_{\dot{v}}(y, x, w) := 2\dot{v}(y, x|w)/v_0(y, x|w)$ and $S_{\dot{b}}(w) := 2\dot{b}(w)/b_0(w)$ are the score functions corresponding to $\dot{v}$ and $\dot{b}$, respectively. Since $\dot{\mathcal{V}} = L_2(W)^{\perp}$, it is clear that $\dot{\mathcal{V}} \perp \dot{\mathcal{B}}$.

Now let κ_t be a curve from I_0 into $\mathcal{N}(T)^{\perp}$, passing through $P_{\mathcal{N}(T)^{\perp}}\mu^*$ at $t = 0$, such that

$$\mathbb{E}_t[Y - \kappa_t|W = w] = 0 \quad \text{for all } (t, w) \in I_0 \times \mathrm{supp}(W),$$

where $\mathbb{E}_t$ denotes conditional expectation under the sub-model $v_t^2(y, x|w)$. Hence, differentiating with respect to t and evaluating at $t = 0$,

$$T\dot{\kappa} = P_{L_2(W)}(\tilde{\varepsilon} S_{\dot{v}}) \tag{A.1}$$

for some $\dot{\kappa} \in \mathcal{N}(T)^{\perp}$. Since (A.1) further restricts $\dot{\mathcal{V}}$, the tangent vectors are given by

$$\dot{\mathcal{M}} := \{f \in L_2(W)^{\perp} : P_{L_2(W)}(\tilde{\varepsilon} f) \in \mathcal{R}(T)\}. \tag{A.2}$$

Therefore, the tangent space of score functions relevant for our problem is $\dot{\mathcal{T}} := \mathrm{cl}(\dot{\mathcal{M}}) + \dot{\mathcal{B}}$. As shown in Lemma B.2, an appealing expression for $\mathrm{cl}(\dot{\mathcal{M}})$ can be obtained under the assumption that the scedastic function is bounded; namely,

$$\mathrm{cl}(\dot{\mathcal{M}}) = \{f \in L_2(W)^{\perp} : P_{L_2(W)}(\tilde{\varepsilon} f) \in \mathrm{cl}(\mathcal{R}(T))\}.$$

Note that since $\mathrm{cl}(\dot{\mathcal{M}})$ and $\dot{\mathcal{B}}$ are closed linear subspaces of $L_2(Y,X,W)$ and $\dot{\mathcal{M}} \perp \dot{\mathcal{B}}$, the tangent space $\dot{\mathcal{T}}$ is a Hilbert space with inner product $\langle \cdot,\cdot\rangle_{L_2(Y,X,W)} + \langle \cdot,\cdot\rangle_{L_2(W)}$.

Since, by (2.5), the parameter of interest θ^* is an implicitly defined function of v_0 and b_0, write it as $\eta(v_0,b_0)$ for some $\eta : L_2(Y,X,W)\times L_2(W) \to \mathbb{R}$. Suppose that $\eta(v_t,b_t)$ satisfies the moment condition

$$\int_{\mathrm{supp}(Y,X,W)} g(x,\eta(v_t,b_t),\kappa_t)v_t^2(y,x|w)b_t^2(w)\,d\lambda\,d\gamma = 0 \quad \text{for all } t\in I_0.$$

Differentiating with respect to t and evaluating at $t=0$, we obtain that

$$\nabla\eta(\dot{\tau}) = \mathbb{E}[\psi\dot{\kappa}] + \mathbb{E}[gS_{\dot{v}}] + \mathbb{E}[gS_{\dot{b}}], \tag{A.3}$$

where $\nabla\eta$ is the derivative of η along one-dimensional paths through (v_0,b_0).

Next, we write $\mathbb{E}[\psi\dot{\kappa}]$ in terms of the tangent vectors so that $\nabla\eta$ can be expressed as a linear functional on the tangent space $\dot{\mathcal{T}}$. So, noting that $\dot{\kappa} \overset{(A.1)}{=} T^+P_{L_2(W)}(\tilde{\varepsilon}S_{\dot{v}})$, where T^+ is the Moore-Penrose inverse of T, we have $\mathbb{E}[\psi\dot{\kappa}] = J_{\psi,\tilde{\varepsilon}}(S_{\dot{v}})$, where $J_{\psi,\tilde{\varepsilon}} : \mathrm{cl}(\dot{\mathcal{M}}) \to \mathbb{R}$ is given by

$$J_{\psi,\tilde{\varepsilon}}(f) := \mathbb{E}[\psi T^+ P_{L_2(W)}(\tilde{\varepsilon}f)]. \tag{A.4}$$

But $\psi \in \mathcal{R}(T')$ by Assumption 2.1; equivalently, $T'\delta^* = \psi$ for some $\delta^*\in L_2(W)$. Therefore,

$$\begin{aligned}
J_{\psi,\tilde{\varepsilon}}(f) &= \langle \psi, T^+P_{L_2(W)}(\tilde{\varepsilon}f)\rangle_{L_2(X)} \\
&= \langle T^{+\prime}\psi, P_{L_2(W)}(\tilde{\varepsilon}f)\rangle_{L_2(W)} && \text{(by Lemma B.3(i))} \\
&= \langle P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*, P_{L_2(W)}(\tilde{\varepsilon}f)\rangle_{L_2(W)} && \text{(by Lemma B.3(ii))} \\
&= \langle \tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*, f\rangle_{L_2(Y,X,W)}, && \text{(A.5)}
\end{aligned}$$

implying, by Assumption 2.2, that $J_{\psi,\tilde{\varepsilon}}$ is bounded on $\mathrm{cl}(\dot{\mathcal{M}})$. Consequently, by (A.3)–(A.5),

$$\begin{aligned}
\nabla\eta(\dot{\tau}) &= J_{\psi,\tilde{\varepsilon}}(S_{\dot{v}}) + \mathbb{E}[gS_{\dot{v}}] + \mathbb{E}[gS_{\dot{b}}] && \text{(A.6)} \\
&= \mathbb{E}[\tilde{\varepsilon}S_{\dot{v}}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*] + \mathbb{E}[gS_{\dot{v}}] + \mathbb{E}[gS_{\dot{b}}] \\
&= \langle \tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g, S_{\dot{v}}\rangle_{L_2(Y,X,W)} + \langle g, S_{\dot{b}}\rangle_{L_2(X,W)};
\end{aligned}$$

i.e., $\nabla\eta$ is bounded on $\dot{\mathcal{T}}$ or, equivalently, that η is a differentiable functional of (v_0,b_0). To further simplify the expression for $\nabla\eta$, notice that since $\tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* \in L_2(W)^\perp$,

$$\begin{aligned}
\nabla\eta(\dot{\tau}) &= \langle \tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g, S_{\dot{v}}\rangle_{L_2(Y,X,W)} + \langle \tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g, S_{\dot{b}}\rangle_{L_2(Y,X,W)} \\
&= \langle \tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g, P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(S_{\dot{v}} + S_{\dot{b}})\rangle_{L_2(Y,X,W)} \\
&= \langle P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(\tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g), S_{\dot{v}} + S_{\dot{b}}\rangle_{L_2(Y,X,W)}, && \text{(A.7)}
\end{aligned}$$

where the third equality is because $S_{\dot{v}} + S_{\dot{b}} \in \mathrm{cl}(\dot{\mathcal{M}}) + L_2(W)$.

Following Severini and Tripathi (2001), the efficiency bound for estimating $\eta(v_0, b_0)$ is given by $\|\nabla\eta\|^2$, the squared norm of its derivative, where $\|\nabla\eta\| := \sup_{\{\dot{\tau}\in\dot{\mathcal{T}}:\dot{\tau}\neq 0\}} |\nabla\eta(\dot{\tau})|$. Therefore, by (A.7) and Riesz-Fréchet, it is immediate that

$$\|\nabla\eta\|^2 = \mathbb{E}[P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(\tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g)]^2. \qquad \square$$

Proof of Lemma 2.2. Let $\pi^* := f - P_{L_2(W)}f - \tilde{\varepsilon}(I - \mho T(T'\mho T)^+T')\mho P_{L_2(W)}(\tilde{\varepsilon}f)$. Since $P_{L_2(W)}\tilde{\varepsilon} = 0$, it is clear that $\pi^* \in L_2(W)^\perp$. Moreover, since $P_{L_2(W)}(\tilde{\varepsilon}^2) = \Omega$ by definition,

$$\begin{aligned} P_{L_2(W)}(\tilde{\varepsilon}\pi^*) &= P_{L_2(W)}(\tilde{\varepsilon}f) - \Omega(I - \mho T(T'\mho T)^+T')\mho P_{L_2(W)}(\tilde{\varepsilon}f) \\ &= T(T'\mho T)^+T'\mho P_{L_2(W)}(\tilde{\varepsilon}f) \in \mathcal{R}(T). \end{aligned}$$

Hence, $\pi^* \in \dot{\mathcal{M}} \subseteq \mathrm{cl}(\dot{\mathcal{M}})$. Next, let $R_T := I - \mho T(T'\mho T)^+T'$. Then, for every $\dot{m} \in \dot{\mathcal{M}}$,

$$\begin{aligned} \langle f - \pi^*, \dot{m}\rangle_{L_2(Y,X,W)} &= \langle \tilde{\varepsilon}R_T\mho P_{L_2(W)}(\tilde{\varepsilon}f), \dot{m}\rangle_{L_2(Y,X,W)} \\ &= \langle P_{L_2(W)}(\tilde{\varepsilon}\dot{m}), R_T\mho P_{L_2(W)}(\tilde{\varepsilon}f)\rangle_{L_2(W)} \qquad \text{(by iterated expectations)} \\ &= \langle R_T' P_{L_2(W)}(\tilde{\varepsilon}\dot{m}), \mho P_{L_2(W)}(\tilde{\varepsilon}f)\rangle_{L_2(W)} = 0 \end{aligned}$$

because from (B.12) we know that $\dot{m} \in \dot{\mathcal{M}}$ implies $R_T' P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) = 0$. Therefore, $f - \pi^* \perp \dot{\mathcal{M}}$ and $f - \pi^* \perp \mathrm{cl}(\dot{\mathcal{M}})$ follows by continuity of the inner product. $\square$

Proof of Theorem 2.2. By (2.7),

$$P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(\tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*+g) = \tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*+g-\tilde{\varepsilon}(I-\mho T(T'\mho T)^+T')\mho P_{L_2(W)}(\tilde{\varepsilon}^2P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*+\tilde{\varepsilon}g).$$

Next, since $P_{L_2(W)}(\tilde{\varepsilon}^2 P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*) = \Omega P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*$,

$$\tilde{\varepsilon}(I - \mho T(T'\mho T)^+T')\mho P_{L_2(W)}(\tilde{\varepsilon}^2 P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*) = \tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* - \tilde{\varepsilon}\mho T(T'\mho T)^+T'P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*.$$

But note that

$$T'P_{\mathrm{cl}(\mathcal{R}(T))} = T'(I - P_{\mathcal{R}(T)^\perp}) = T'(I - P_{\mathcal{N}(T')}) = T'. \tag{A.8}$$

Therefore,

$$T'P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* \overset{\text{(A.8)}}{=} T'\delta^* \overset{\text{Ass. 2.1}}{=} \psi,$$

and we obtain that

$$P_{\mathrm{cl}(\dot{\mathcal{M}})+L_2(W)}(\tilde{\varepsilon}P_{\mathrm{cl}(\mathcal{R}(T))}\delta^* + g) = \tilde{\varepsilon}\mho T(T'\mho T)^+\psi + g - \tilde{\varepsilon}(I - \mho T(T'\mho T)^+T')\mho P_{L_2(W)}(\tilde{\varepsilon}g). \quad \square$$

Proof of Lemma 2.3. Since the basic idea is very similar to the proof of Theorem 2.1, we only describe the essential steps; notation and symbols not defined here have the same meaning as in the proof of Theorem 2.1. Let β_t be a smooth curve through β^* and $\dot{\beta} := d\beta_t/dt|_{t=0}$. Then, for the one-dimensional submodel $\mathbb{E}_t[Y - m(X, \beta_t)|W] = 0$ w.p.1,

$$T_p\dot{m} = P_{L_2(W)}(\tilde{\varepsilon}S_{\dot{v}}),$$

where $\dot{m} := \dot{\beta}'\nabla_\beta m(X,\beta^*)$. Therefore, the tangent space $\dot{\mathcal{T}}_p := \dot{\mathcal{M}}_p + \dot{\mathcal{B}}$, where $\dot{\mathcal{M}}_p$ is closed because $\mathcal{R}(T_p)$ is finite dimensional. Next, following the argument to (A.4), for every $\dot{\tau} \in \dot{\mathcal{T}}_p$,

$$\nabla\eta(\dot{\tau}) = \mathbb{E}[\psi T_p^+ P_{L_2(W)}(\tilde{\varepsilon} S_{\dot{v}})] + \langle g, S_{\dot{v}}\rangle_{L_2(Y,X,W)} + \langle g, S_{\dot{b}}\rangle_{L_2(X,W)}.$$

The same reasoning that led to (A.5), plus the fact that $T_p' = T'$, can be used to show that

$$\mathbb{E}[\psi T_p^+ P_{L_2(W)}(\tilde{\varepsilon} S_{\dot{v}})] = \langle \psi, T_p^+ P_{L_2(W)}(\tilde{\varepsilon} S_{\dot{v}})\rangle_{L_2(X)} = \langle \tilde{\varepsilon} P_{\mathcal{R}(T_p)}\delta^*, S_{\dot{v}}\rangle_{L_2(Y,X,W)}.$$

Hence,

$$\nabla\eta(\dot{\tau}) = \langle \tilde{\varepsilon} P_{\mathcal{R}(T_p)}\delta^* + g, S_{\dot{v}}\rangle_{L_2(Y,X,W)} + \langle g, S_{\dot{b}}\rangle_{L_2(X,W)},$$

which, as in (A.7), can be written as

$$\nabla\eta(\dot{\tau}) = \langle P_{\dot{\mathcal{M}}_p + L_2(W)}(\tilde{\varepsilon} P_{\mathcal{R}(T_p)}\delta^* + g), S_{\dot{v}} + S_{\dot{b}}\rangle_{L_2(Y,X,W)}.$$

Therefore, the efficiency bound is for estimating θ^* when $P_{\mathcal{N}(T)^\perp}\mu^*$ be embedded in the parametric family $\mathcal{F}_p$ is given by $\mathbb{E}[P_{\dot{\mathcal{M}}_p + L_2(W)}(\tilde{\varepsilon} P_{\mathcal{R}(T_p)}\delta^* + g)]^2$. $\square$

Proof of Lemma 2.4. Since $P_{\dot{\mathcal{M}}_p + L_2(W)}(\tilde{\varepsilon} P_{\mathcal{R}(T_p)}\delta^* + g) = P_{\dot{\mathcal{M}}_p}(\tilde{\varepsilon} P_{\mathcal{R}(T_p)}\delta^*) + P_{L_2(W)}g$, it suffices to show that

$$\|P_{\dot{\mathcal{M}}_p}(\tilde{\varepsilon} P_{\mathcal{R}(T_p)}\delta^*)\|_{L_2(Y,X,W)} \to \|P_{\mathrm{cl}(\dot{\mathcal{M}})}(\tilde{\varepsilon} P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*)\|_{L_2(Y,X,W)} \quad \text{as } p \to \infty.$$

Begin by observing that

$$P_{\dot{\mathcal{M}}_p}(\tilde{\varepsilon} P_{\mathcal{R}(T_p)}\delta^*) - P_{\mathrm{cl}(\dot{\mathcal{M}})}(\tilde{\varepsilon} P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*) = (P_{\dot{\mathcal{M}}_p} - P_{\mathrm{cl}(\dot{\mathcal{M}})})(\tilde{\varepsilon} P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*) + r_p,$$

where $r_p := P_{\dot{\mathcal{M}}_p}(\tilde{\varepsilon}(P_{\mathcal{R}(T_p)} - P_{\mathrm{cl}(\mathcal{R}(T))})\delta^*)$. Since projection operators are bounded with norm equal to one, by Assumption 2.2 we have that

$$\|r_p\|_{L_2(Y,X,W)} \leq \|\tilde{\varepsilon}(P_{\mathcal{R}(T_p)} - P_{\mathrm{cl}(\mathcal{R}(T))})\delta^*\|_{L_2(Y,X,W)} \leq c\|(P_{\mathcal{R}(T_p)} - P_{\mathrm{cl}(\mathcal{R}(T))})\delta^*\|_{L_2(Y,X,W)}.$$

But $P_{\mathcal{R}(T_p)}$ is a monotone sequence of projection operators since $\mathcal{R}(T_p) \uparrow \mathrm{cl}(\mathcal{R}(T))$ by assumption. Therefore, by Akhiezer and Glazman (1993, p. 68), the sequence of operators $P_{\mathcal{R}(T_p)}$ converges strongly to $P_{\mathrm{cl}(\mathcal{R}(T))}$, implying that

$$\|r_p\|_{L_2(Y,X,W)} \to 0 \quad \text{as } p \to \infty.$$

Since $P_{\dot{\mathcal{M}}_p}$ is also monotone, because $\mathcal{R}(T_p) \uparrow \mathrm{cl}(\mathcal{R}(T)) \implies \dot{\mathcal{M}}_p \uparrow \mathrm{cl}(\dot{\mathcal{M}})$, we also have

$$\|(P_{\dot{\mathcal{M}}_p} - P_{\mathrm{cl}(\dot{\mathcal{M}})})(\tilde{\varepsilon} P_{\mathrm{cl}(\mathcal{R}(T))}\delta^*)\|_{L_2(Y,X,W)} \to 0 \quad \text{as } p \to \infty.$$

The desired result follows. $\square$

Proof of Lemma 3.1. Follows immediately from the decomposition

$$\hat{\theta} - \theta^* = \int \psi(\hat{\mu} - \mu^*)\,(d\hat{\mathbb{P}} - d\mathbb{P}) + \int \psi(\hat{\mu} - \mu^*)\,d\mathbb{P} + \int (g - \varphi)\,d\hat{\mathbb{P}} + \int \varphi\,d\hat{\mathbb{P}}. \qquad \square$$

Appendix B. Some Useful Results

Lemma B.1. *Let A and B be Hilbert spaces and $K : A \to B$ a compact linear operator with singular system $\{(\lambda_j, a_j, b_j) : j \in \mathbb{N}\}$, where $\{\lambda_j\}$ is the set of non-zero singular values, $\{a_j\}$ an orthonormal basis for $\mathcal{N}(K)^\perp$, and $\{b_j\}$ an orthonormal basis for* $\mathrm{cl}(\mathcal{R}(K))$. *Then,* (i) $\mathcal{R}(K) = \{b \in B : \sum_{j=1}^\infty \langle b, b_j\rangle_B^2 \lambda_j^{-2} < \infty\}$; (ii) *if $\sum_{j=1}^\infty \lambda_j^2 < \infty$, i.e., if K is Hilbert-Schmidt, then $\mathcal{R}(K)$ is a proper subspace of B, i.e., K is not surjective.*

Proof of Lemma B.1. Let $b \in \mathcal{R}(K)$. Then $KP_{\mathcal{N}(K)^\perp} a = b$ for some $a \in A$. Thus, by the singular value decomposition of K, see, e.g., Kress (1999, Section 15.4),

$$\sum_{j=1}^\infty \lambda_j \langle P_{\mathcal{N}(K)^\perp} a, a_j\rangle_A b_j = \sum_{j=1}^\infty \langle b, b_j\rangle_B b_j \implies \langle P_{\mathcal{N}(K)^\perp} a, a_j\rangle_A = \langle b, b_j\rangle_B / \lambda_j.$$

Hence, $\sum_{j=1}^\infty \langle b, b_j\rangle_B^2 \lambda_j^{-2} < \infty$ since $P_{\mathcal{N}(K)^\perp} a \in A$; i.e., we have shown that

$$\mathcal{R}(K) \subseteq \{b \in B : \sum_{j=1}^\infty \langle b, b_j\rangle_B^2 \lambda_j^{-2} < \infty\}. \tag{B.1}$$

To show the reverse inclusion, let b belong to the RHS of (B.1) and $a := \sum_{j=1}^\infty \langle b, b_j\rangle_B \lambda_j^{-1} a_j$. Since $Ka = b$, it follows that $b \in \mathcal{R}(K)$; therefore, (i) holds. Finally, assume that $\sum_{j=1}^\infty \lambda_j^2 < \infty$. Then, $b := \sum_{j=1}^\infty \lambda_j b_j \in B \setminus \mathcal{R}(K)$ since $\sum_{j=1}^\infty \langle b, b_j\rangle_B^2 \lambda_j^{-2} = \infty$. Hence, $\mathcal{R}(K) \subsetneq B$. $\square$

Corollary B.1. *Let K be as in Lemma* B.1. *Then,* $\mathcal{R}(K') = \{a \in A : \sum_{j=1}^\infty \langle a, a_j\rangle_A^2 \lambda_j^{-2} < \infty\}$; *and if $\sum_{j=1}^\infty \lambda_j^2 < \infty$, then $\mathcal{R}(K') \subsetneq A$.*

Proof of Corollary B.1. Follow the proof of Lemma B.1 keeping in mind the singular value decomposition of K', i.e., $K'b = \sum_{j=1}^\infty \lambda_j \langle b, b_j\rangle_B a_j$ for $b \in \mathrm{cl}(\mathcal{R}(K))$. $\square$

Lemma B.2. *Let Assumption* 2.2 *hold. Then, recalling the definition of $\dot{\mathcal{M}}$ from* (A.2),

$$\mathrm{cl}(\dot{\mathcal{M}}) = \{f \in L_2(W)^\perp : P_{L_2(W)}(\tilde{\varepsilon} f) \in \mathrm{cl}(\mathcal{R}(T))\}.$$

Proof of Lemma B.2. Let $f \in \mathrm{cl}(\dot{\mathcal{M}})$. Then, there exists a sequence f_k in $\dot{\mathcal{M}}$ such that $f_k \to f$ as $k \to \infty$. Therefore, for every $k \in \mathbb{N}$, we have $f_k \in L_2(W)^\perp$ and $P_{L_2(W)}(\tilde{\varepsilon} f_k) = Ta_k$ for some $a_k \in L_2(X)$. But, by Cauchy-Schwarz and Assumption 2.2,

$$\|P_{L_2(W)}(\tilde{\varepsilon} f)\|_{L_2(W)} \le \|\Omega^{1/2}(P_{L_2(W)} f^2)^{1/2}\|_{L_2(W)} \le c\|f\|_{L_2(Y,X,W)};$$

i.e., $f \mapsto P_{L_2(W)}(\tilde{\varepsilon} f)$ is a bounded linear map from $L_2(Y, X, W) \to L_2(W)$. Hence,

$$\lim_{k\to\infty} Ta_k = P_{L_2(W)}(\tilde{\varepsilon} f) \implies P_{L_2(W)}(\tilde{\varepsilon} f) \in \mathrm{cl}(\mathcal{R}(T)).$$

Since $f \in L_2(W)^\perp$, because $L_2(W)^\perp$ is closed, it follows that

$$\mathrm{cl}(\dot{\mathcal{M}}) \subseteq \{f \in L_2(W)^\perp : P_{L_2(W)}(\tilde{\varepsilon} f) \in \mathrm{cl}(\mathcal{R}(T))\}. \tag{B.2}$$

To show the reverse inclusion, let m belong to the RHS of (B.2). Then, for every $\epsilon > 0$, there exists a $b_\epsilon \in \mathcal{R}(T)$ such that

$$\|b_\epsilon - P_{L_2(W)}(\tilde{\varepsilon} m)\|_{L_2(W)} < \epsilon. \tag{B.3}$$

Now let $\dot{m}_\epsilon := m + \tilde{\varepsilon}\mho(b_\epsilon - P_{L_2(W)}(\tilde{\varepsilon} m))$. Since $m \in L_2(W)^\perp$ and $b_\epsilon - P_{L_2(W)}(\tilde{\varepsilon} m) \in L_2(W)$, it is clear that $\dot{m}_\epsilon \in L_2(W)^\perp$. Therefore, since

$$\begin{aligned} P_{L_2(W)}(\tilde{\varepsilon}\dot{m}_\epsilon) &= P_{L_2(W)}(\tilde{\varepsilon} m) + P_{L_2(W)}(\tilde{\varepsilon}^2\mho(b_\epsilon - P_{L_2(W)}(\tilde{\varepsilon} m))) \\ &= P_{L_2(W)}(\tilde{\varepsilon} m) + P_{L_2(W)}(b_\epsilon - P_{L_2(W)}(\tilde{\varepsilon} m)) \\ &= b_\epsilon \in \mathcal{R}(T), \end{aligned}$$

it follows that $\dot{m}_\epsilon \in \dot{\mathcal{M}}$. Finally, by iterated expectations, Assumption 2.2, and (B.3),

$$\begin{aligned} \|\dot{m}_\epsilon - m\|_{L_2(Y,X,W)} &= \|\tilde{\varepsilon}\mho(b_\epsilon - P_{L_2(W)}(\tilde{\varepsilon} m))\|_{L_2(Y,X,W)} = \|\mho^{1/2}(b_\epsilon - P_{L_2(W)}(\tilde{\varepsilon} m))\|_{L_2(W)} \\ &\le c\|b_\epsilon - P_{L_2(W)}(\tilde{\varepsilon} m)\|_{L_2(W)} \\ &< c\epsilon. \end{aligned}$$

Therefore, $\dot{m}_\epsilon \in \dot{\mathcal{M}}$ is arbitrarily close to m; hence, $m \in \mathrm{cl}(\dot{\mathcal{M}})$. □

Lemma B.3. *Let $Q : A \to B$ be a bounded linear operator, where A and B are Hilbert spaces, $Q^{+\prime}$ denote the adjoint of Q^+, and $a \in \mathcal{R}(Q')$. Then,* (i) *$a \in \mathcal{D}(Q^{+\prime})$; and* (ii) *$Q^{+\prime}a = P_{\mathrm{cl}(\mathcal{R}(Q))}b$, where $b \in B$ is such that $Q'b = a$.*

Proof of Lemma B.3. Without loss of generality, assume that $\mathcal{R}(Q)$ is not closed; the result for the closed-range case follows by replacing $\mathrm{cl}(\mathcal{R}(Q))$ in (ii) with $\mathcal{R}(Q)$. Since $\mathcal{R}(Q)$ is not closed, the Moore-Penrose inverse $Q^+ : \mathcal{R}(Q) + \mathcal{R}(Q)^\perp \to \mathcal{N}(Q)^\perp$ is unbounded; moreover, $\mathcal{D}(Q^+)$ is a dense subspace of B and $\mathcal{R}(Q^+) \subseteq A$. Hence, by Kreyszig (1978, Definition 10.1-2), the operator $Q^{+\prime} : \mathcal{D}(Q^{+\prime}) \to B$ is such that

$$\mathcal{D}(Q^{+\prime}) = \{a \in A : \exists b^* \in B \text{ s.t. } \langle Q^+ f, a\rangle_A = \langle f, b^*\rangle_B \; \forall f \in \mathcal{D}(Q)^+\} \tag{B.4}$$

$$Q^{+\prime}a := b^*. \tag{B.5}$$

Let $Q|_{\mathcal{N}(Q)^\perp}$ denote the restriction of Q to $\mathcal{N}(Q)^\perp$. To verify that a lies in the domain of $Q^{+\prime}$, observe that for every $f \in \mathcal{R}(Q) + \mathcal{R}(Q)^\perp$,

$$\begin{aligned} \langle Q^+ f, a\rangle_A &= \langle (Q|_{\mathcal{N}(Q)^\perp})^{-1} P_{\mathrm{cl}(\mathcal{R}(Q))} f, a\rangle_A \\ &= \langle (Q|_{\mathcal{N}(Q)^\perp})^{-1} P_{\mathrm{cl}(\mathcal{R}(Q))} f, Q'b\rangle_A \\ &= \langle Q(Q|_{\mathcal{N}(Q)^\perp})^{-1} P_{\mathrm{cl}(\mathcal{R}(Q))} f, b\rangle_B \\ &= \langle P_{\mathrm{cl}(\mathcal{R}(Q))} f, b\rangle_B \\ &= \langle f, P_{\mathrm{cl}(\mathcal{R}(Q))} b\rangle_B. \end{aligned}$$

Furthermore, since $b \in B$,

$$\|P_{\mathrm{cl}(\mathcal{R}(Q))}b\|_B \leq \|b\|_B < \infty.$$

Therefore, by (B.4) and (B.5) it follows that $a \in \mathcal{D}(Q^{+\prime})$ and $Q^{+\prime}a = P_{\mathrm{cl}(\mathcal{R}(Q))}b$. $\square$

Lemma B.4. $(\mathcal{R}(T'\mho T) + \mathcal{R}(T'\mho T)^{\perp}) \cap \mathcal{N}(T)^{\perp} = \mathcal{R}(T'\mho T)$.

Proof of Lemma B.4. Let $a \in (\mathcal{R}(T'\mho T) + \mathcal{R}(T'\mho T)^{\perp}) \cap \mathcal{N}(T)^{\perp}$. Since $\mathcal{N}(T)^{\perp} = \mathrm{cl}(\mathcal{R}(T'))$, we have $a = f + g$, where $f \in \mathcal{R}(T'\mho T)$, $g \in \mathcal{R}(T'\mho T)^{\perp}$, and $f + g \in \mathrm{cl}(\mathcal{R}(T'))$. Hence,

$$g = f - \mathrm{cl}(\mathcal{R}(T')) \in \mathcal{R}(T'\mho T) - \mathrm{cl}(\mathcal{R}(T')) \subseteq \mathrm{cl}(\mathcal{R}(T')) \subseteq L_2(X). \tag{B.6}$$

Moreover,

$$\begin{aligned} g \in \mathcal{R}(T'\mho T)^{\perp} &\iff \langle g, T'\mho Ta\rangle_{L_2(X)} = 0 \text{ for every } a \in L_2(X) \\ &\overset{\text{(B.6)}}{\Longrightarrow} \langle g, T'\mho Tg\rangle_{L_2(X)} = 0 \\ &\iff \langle Tg, \mho Tg\rangle_{L_2(X)} = 0 \\ &\overset{\mho>0}{\iff} Tg = 0 \iff g \in \mathcal{N}(T). \end{aligned}$$

But $g \overset{\text{(B.6)}}{\in} \mathrm{cl}(\mathcal{R}(T')) = \mathcal{N}(T)^{\perp}$. Hence,

$$g = 0 \implies a = f \in \mathcal{R}(T'\mho T).$$

Therefore, $(\mathcal{R}(T'\mho T)+\mathcal{R}(T'\mho T)^{\perp})\cap\mathcal{N}(T)^{\perp} \subseteq \mathcal{R}(T'\mho T)$. To show the converse, let $b \in \mathcal{R}(T'\mho T)$. Then, for some $d \in L_2(X)$ and every $v \in \mathcal{N}(T)$,

$$\langle b, v\rangle_{L_2(X)} = \langle T'\mho Td, v\rangle_{L_2(X)} = \langle \mho Td, Tv\rangle_{L_2(X)} = 0 \implies b \perp \mathcal{N}(T).$$

Hence, $\mathcal{R}(T'\mho T) \subseteq \mathcal{N}(T)^{\perp}$, implying that

$$\mathcal{R}(T'\mho T) \subseteq (\mathcal{R}(T'\mho T) + \mathcal{R}(T'\mho T)^{\perp}) \cap \mathcal{N}(T)^{\perp}. \qquad \square$$

Lemma B.5. *Let Assumption* 2.2 *hold and recall the definition of* $\dot{\mathcal{M}}$ *from* (A.2)*. Then,*

$$\dot{\mathcal{M}} = \{\dot{m} \in L_2(W)^{\perp} : T'\mho P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) \in \mathcal{R}(T'\mho T) \text{ and } (I - T(T'\mho T)^{+\prime}T'\mho)P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) = 0\}.$$

Proof of Lemma B.5. Let $\dot{m} \in \dot{\mathcal{M}}$. Then, by (A.2), we know $\dot{m} \in L_2(W)^{\perp}$ and there exists an $a \in L_2(X)$ such that

$$P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) = Ta. \tag{B.7}$$

Since (B.7) implies that

$$T'\mho P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) = T'\mho Ta, \tag{B.8}$$

we get that $T'\mho P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) \in \mathcal{R}(T'\mho T)$. Hence, by (B.8),

$$a = (T'\mho T)^{+}T'\mho P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) \in \mathcal{R}((T'\mho T)^{+}). \tag{B.9}$$

But, since $T'\mho T$ is bounded by Assumption 2.2,

$$\mathcal{R}((T'\mho T)^{+}) = \mathcal{N}(T'\mho T)^{\perp} = \mathrm{cl}(\mathcal{R}(T'\mho T)).$$

Thus, by (B.9),

$$a \in \mathrm{cl}(\mathcal{R}(T'\mho T)). \tag{B.10}$$

Hence, $T'\mho Ta$ lies in the domain of $(T'\mho T)^{+'}$.[16] Therefore,

$$\begin{aligned}(T'\mho T)^{+'}T'\mho P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) &\overset{(B.8)}{=} (T'\mho T)^{+'}T'\mho Ta \\ &\overset{\text{Lemma B.3(ii)}}{=} P_{\mathrm{cl}(\mathcal{R}(T'\mho T))}a \\ &\overset{(B.10)}{=} a,\end{aligned}$$

implying that

$$T(T'\mho T)^{+'}T'\mho P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) = Ta. \tag{B.11}$$

Hence, subtracting (B.11) from (B.7),

$$(I - T(T'\mho T)^{+'}T'\mho)P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) = 0.$$

In other words, letting $R_T := I - \mho T(T'\mho T)^{+}T'$, we have shown that

$$\dot{\mathcal{M}} \subseteq \{\dot{m} \in L_2(W)^{\perp} : T'\mho P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) \in \mathcal{R}(T'\mho T) \text{ and } R_T' P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) = 0\}. \tag{B.12}$$

The reverse inclusion is straightforward. Let $\dot{m}$ be an arbitrary element in the RHS of (B.12). Then, $\dot{m}$ lies in $L_2(W)^{\perp}$ and satisfies

$$R_T' P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) = 0 \iff P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) = T(T'\mho T)^{+'}T'\mho P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) \in \mathcal{R}(T).$$

Hence, we have

$$\{\dot{m} \in L_2(W)^{\perp} : T'\mho P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) \in \mathcal{R}(T'\mho T) \text{ and } R_T' P_{L_2(W)}(\tilde{\varepsilon}\dot{m}) = 0\} \subseteq \dot{\mathcal{M}}. \tag{B.13}$$

[16] Note that $(T'\mho T)^{+}$, which maps $\mathcal{R}(T'\mho T) + \mathcal{R}(T'\mho T)^{\perp}$ onto $\mathcal{N}((T'\mho T))^{\perp} \subseteq L_2(X)$, is unbounded because $\mathcal{R}(T'\mho T)$ is not assumed to be closed. Therefore, referring to the proof of Lemma B.3, the domain of $(T'\mho T)^{+'}$ consists of all $q \in L_2(X)$ such that there exists a $q^* \in L_2(X)$ satisfying

$$\langle (T'\mho T)^{+}f, q\rangle_{L_2(X)} = \langle f, q^*\rangle_{L_2(X)} \quad \text{for every } f \in \mathcal{D}((T'\mho T)^{+}).$$

To verify that $T'\mho Ta$ lies in the domain of $(T'\mho T)^{+'}$ observe that, for every $f \in \mathcal{R}(T'\mho T) + \mathcal{R}(T'\mho T)^{\perp}$,

$$\begin{aligned}\langle (T'\mho T)^{+}f, T'\mho Ta\rangle_{L_2(X)} &= \langle (T'\mho T|_{\mathcal{N}(T'\mho T)^{\perp}})^{-1}P_{\mathrm{cl}(\mathcal{R}(T'\mho T))}f, T'\mho Ta\rangle_{L_2(X)} \\ &= \langle T'\mho T|_{\mathcal{N}(T'\mho T)^{\perp}}(T'\mho T|_{\mathcal{N}(T'\mho T)^{\perp}})^{-1}P_{\mathrm{cl}(\mathcal{R}(T'\mho T))}f, a\rangle_{L_2(X)} \\ &= \langle P_{\mathrm{cl}(\mathcal{R}(T'\mho T))}f, a\rangle_{L_2(X)} \\ &= \langle f, P_{\mathrm{cl}(\mathcal{R}(T'\mho T))}a\rangle_{L_2(X)} \\ &\overset{(B.10)}{=} \langle f, a\rangle_{L_2(X)}.\end{aligned}$$

Since we already know that $a \in L_2(X)$, it follows that $T'\mho Ta$ lies in the domain of $(T'\mho T)^{+'}$.

The desired result follows by (B.12) and (B.13). □

References

Ai, C. and X. Chen (2003): "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, 71, 1795–1843.

——— (2005a): "Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables," Manuscript.

——— (2005b): "On efficient sequential estimation of semi-nonparametric moment models," Manuscript.

Akhiezer, N. and Glazman (1993): *Theory of linear operators in Hilbert space*, vol. 1, Dover Publications.

Andrews, G. E., R. Askey, and R. Roy (1999): *Special Functions*, Cambridge University Press.

Blundell, R. and J. L. Powell (2003): "Endogeneity in nonparametric and semiparametric regression models," in *Advances in economics and econometrics: Theory and applications*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, Cambridge University Press, vol. 2, 312–357.

Brown, B. W. and W. K. Newey (1998): "Efficient semiparametric estimation of expectations," *Econometrica*, 66, 453–464.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

——— (1992): "Efficiency bounds for semiparametric regression," *Econometrica*, 60, 567–596.

Darolles, S., J.-P. Florens, and E. Renault (2006): "Nonparametric instrumental regression," Manuscript.

Engl, H. W., M. Hanke, and A. Neubauer (2000): *Regularization of inverse problems*, Kluwer.

Florens, J.-P., J. Johannes, and S. van Bellegem (2005): "Instrumental regression in partially linear models," Manuscript.

Hall, P. and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *Annals of Statisics*, 33, 2904–2929.

Kress, R. (1999): *Linear integral equations*, Springer Verlag, 2nd ed.

Kreyszig, E. (1978): *Introductory functional analysis with applications*, John Wiley and Sons.

Newey, W. K. (1994): "The asymptotic variance of semiparametric estimators," *Econometrica*, 62, 1349–1382.

Newey, W. K. and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics, vol. IV*, ed. by R. Engle and D. McFadden, Elsevier Science B.V., 2111–2245.

Newey, W. K. and J. L. Powell (2003): "Instrumental variables estimation of nonparametric models," *Econometrica*, 71, 1557–1569.

Ritov, Y. and P. J. Bickel (1990): "Achieving information bounds in non and semiparametric models," *Annals of Statistics*, 18, 925–938.

Severini, T. A. and G. Tripathi (2001): "A simplified approach to computing efficiency bounds in semiparametric models," *Journal of Econometrics*, 102, 23–66.

——— (2006): "Some identification issues in nonparametric linear models with endogenous regressors," *Econometric Theory*, 22, 258–278.

Stock, J. H. (1989): "Nonparametric policy analysis," *Journal of the American Statistical Association*, 84, 567–575.

van der Vaart, A. (1991): "On differentiable functionals," *Annals of Statistics*, 19, 178–204.

Wong, W. H. (1986): "Theory of partial likelihood," *Annals of Statistics*, 14, 88–123.

Department of Statistics, Northwestern University, Evanston, IL - 60201.
E-mail address: `severini@northwestern.edu`

Department of Economics, University of Connecticut, Storrs, CT - 06269.
E-mail address: `gautam.tripathi@uconn.edu`