EFFICIENCY BOUNDS FOR ESTIMATING LINEAR FUNCTIONALS OF NONPARAMETRIC REGRESSION MODELS WITH ENDOGENOUS REGRESSORS

THOMAS A. SEVERINI AND GAUTAM TRIPATHI

ABSTRACT. Let $Y = \mu^*(X) + \varepsilon$, where $\mu^*$ is unknown and $\mathbb{E}[\varepsilon|X] \neq 0$ with positive probability but there exist instrumental variables $W$ such that $\mathbb{E}[\varepsilon|W] = 0$ w.p.1. It is well known that such nonparametric regression models are generally “ill-posed” in the sense that the map from the data to $\mu^*$ is not continuous. In this paper, we derive the efficiency bounds for estimating certain linear functionals of $\mu^*$ without assuming $\mu^*$ itself to be identified.

1. Introduction

Models containing unknown functions are commonly used in econometrics and statistics. For instance, consider the model for an observed random vector $(Y, X)$ given by $Y = \mu^*(X) + \varepsilon$, where $\mu^*$ is an unknown function and $\varepsilon$ is an unobserved random variable. If $\mathbb{E}[\varepsilon|X] = 0$, then $\mu^*(x) = \mathbb{E}[Y|X = x]$ and nonparametric regression methods can be used for inference about $\mu^*$. Now suppose that the condition $\mathbb{E}[\varepsilon|X] = 0$ is not satisfied. This typically occurs whenever some components of $X$ are determined endogenously. In this case, $\mu^*$ is no longer a conditional expectation. Nonetheless, estimation of $\mu^*$ may still be possible provided there exists a random vector $W$ such that $\mathbb{E}[\varepsilon|W] = 0$. Unfortunately, the results of [Ai and Chen (2003), Hall and Horowitz (2005), Darolles, Florens and Renault (2006), Severini and Tripathi (2006), and Blundell, Chen, and Kristensen (2007)] show that estimators of $\mu^*$ can have very poor rates of convergence because such models are “ill-posed” under general conditions. Thus, even relatively large sample sizes may not be of much help in accurately estimating $\mu^*$.

In contrast, it may be possible to accurately estimate certain features of $\mu^*$, such as linear functionals of the form $\mathbb{E}[\psi(X)\mu^*(X)]$ and $\int \psi(x)\mu^*(x) \, dx$, where $\psi$ is a known function. In particular, it may be possible to estimate such a linear functional at the usual parametric rate of convergence, even when $\mu^*$ itself is not identified.

Economists are often interested in estimating linear functionals of unknown functions. For instance, [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...

The main objective of this paper is to derive the efficiency bounds for estimating linear functionals of $\mu^*$ when it is not a conditional expectation without assuming $\mu^*$ to be identified. There are at least two reasons why such efficiency bounds are important. One is that efficiency bounds can be used to recognize, and in some cases help construct, an asymptotically efficient estimator of a linear functional. That is, if an estimator has asymptotic variance equal to the efficiency bound, then it is asymptotically efficient. A second use of the efficiency bounds derived in this paper is in understanding nonparametric regression models with endogenous regressors. Efficiency bounds for linear functionals allow us to measure the relative difficulty in estimating different features of the function $\mu^*$ thus telling us what may be learned from the data about $\mu^*$. For instance, we are able to characterize a condition that is necessary for $n^{1/2}$-estimability of these functionals when they are identified. This is particularly important in the present context since estimation of $\mu^*$ itself is generally quite difficult.

Estimation of functionals of $\mu^*$ has been considered by Ai and Chen (2005, 2007, 2008) and Darolles, Florens, and Renault (2006). Ai and Chen (2005, 2008) derive the efficiency bound for estimating functionals of $\mu^*$ when $\mu^*$ is identified. Ai and Chen (2007, Example 2.2) consider estimating a weighted average derivative of $\mu^*$ and show that their estimator is $n^{1/2}$-consistent and asymptotically normal. So the contribution of our paper is to derive the efficiency bound for estimating functionals of $\mu^*$ that remains valid even when $\mu^*$ is not identified and the proof is different from that of Ai and Chen. A discussion on the $n^{1/2}$-rate of convergence of inner products can be found in Darolles, Florens, and Renault (2006), cf. their Section 4.3 (p. 31–35). The results in this paper complement those obtained earlier by Ai and Chen and Darolles, Florens, and Renault.

The outline of the paper is as follows. The model under consideration is described in detail in Section 2. Section 3 contains a discussion of identification and ill-posedness in this model and the relationship between ill-posedness and $n^{1/2}$-estimability is considered in Section 4. The efficiency bound for a functional of the unknown function is presented in Section 5. Proofs are in the appendices.

2. THE MODEL

Consider the nonparametric regression model

$$Y = \mu^*(X) + \varepsilon, \quad (2.1)$$

where $X$ is a vector of regressors some or all of which are endogenous so that $\mathbb{E}[\varepsilon | X] \neq 0$ with positive probability. The functional form of $\mu^*$ 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 $X$. We assume that $\varepsilon$ satisfies the conditional moment restriction $\mathbb{E}[\varepsilon|W] = 0$ w.p.1, where $W$ denotes a vector of instrumental variables (IV’s); conditions under which a $\mu^*$ satisfying \eqref{eq:2.1} is uniquely defined are described below. Since $W$ does not coincide with $X$ (though they may have some elements in common because exogenous regressors are valid instruments), $\mu^*$ cannot be a conditional expectation; if all regressors are exogenous, i.e., $W = X$, then of course $\mu^* = \mathbb{E}[Y|X]$. The observed data consists of iid copies of $(Y, X, W)$.

Identification, i.e., uniqueness, of a $\mu^*$ satisfying \eqref{eq:2.1} is equivalent to the completeness of the conditional distribution of $X$ given $W$ (cf. Newey and Powell, 2003, p. 1567), a condition that may be well nigh impossible to check if $\text{Law}(X|W)$ is unknown — as is maintained in this paper. Moreover, even if $\mu^*$ happens to be uniquely defined, the equation defining it can still be ill-posed in the sense that the function mapping the data to $\mu^*$ may not be continuous (cf. Lemma 2.4 of Severini and Tripathi (2006) for additional properties of this mapping). Note that if $\mu^*$ is not identified then it is the equation defining the “identifiable part” of $\mu^*$ that can be ill-posed (cf. Section 3).

As mentioned in the introduction, a consequence of ill-posedness is that estimators of $\mu^*$, or its identified part, can have very poor rates of convergence. Hence, it makes good statistical sense to study functionals of $\mu^*$ that are estimable at parametric, i.e., $n^{1/2}$-rates. In this paper, we take a step in this direction by deriving the efficiency bounds for estimating linear functionals of the form $\mathbb{E}[\psi(X)\mu^*(X)]$ and $\int_{\text{supp}(X)} \psi(x)\mu^*(x) \, dx$, where $\psi$ is a known weight function and $\text{supp}(X)$ denotes the support of $X$, without assuming that $\mu^*$ is identified. Note that these functionals, being subfeatures of $\mu^*$, may be identifiable even when $\mu^*$ itself is not identified (cf. Section 3).

The results we obtain are most cleanly characterized in terms of operators on Hilbert spaces. With this in mind, the following notation is used throughout the paper. $L_2(Y, X, W)$ is the set of real valued functions of $(Y, X, W)$ that are square integrable with respect to the joint distribution of $Y$ and the distinct coordinates of $X$ and $W$. For $f \in L_2(Y, X, W)$ we write $\mathbb{E}[f|W]$ as $P_{L_2(W)} f$, where $P_A$ denotes orthogonal projection onto $A \subset L_2(Y, X, W)$ using the inner product $\langle f_1, f_2 \rangle_{L_2(Y, X, W)} := \mathbb{E}[f_1 f_2]$. In fact, we take advantage of this notation and continue to write $\mathbb{E}[f|W]$ as $P_{L_2(W)} f$ even if $f \in L_1(Y, X, W) \setminus L_2(Y, X, W)$; any ambiguity can be resolved by just thinking of $P_{L_2(W)}$ as $\mathbb{E}[\cdot|W]$. 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 $\text{cl}(A)$.

For $\int_{\text{supp}(X)} \psi(x)\mu^*(x) \, dx$ 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. Henceforth, these are written simply as $\int \psi \mu^*$ and $\mathbb{E}[\psi \mu^*]$. This
and other instances of functional notation, where arguments taken by functions are suppressed, should not cause any confusion. Furthermore, for the remainder of the paper, we focus attention on the expectation functional \( \mathbb{E}[\psi \mu^*] \); results for \( \int \psi \mu^* \) follow mutatis mutandis.

3. Identification and ill-posedness

In this section we briefly describe what we mean by the identifiable part of \( \mu^* \) and the sense in which the equation defining it can be ill-posed. Cf. [Kress (1999)] for the definition of ill-posed linear equations. Some recent papers that discuss identification conditions and results for ill-posed econometric models include Ai and Chen (2003), Newey and Powell (2003), Hall and Horowitz (2005), Darolles, Florens and Renault (2006), and Blundell, Chen, and Kristensen (2007). Severini and Tripathi (2006) have more on underidentification and ill-posedness in a general setting, but they use different notation.

Since \( P_{L_2(W)} \varepsilon = 0 \) by assumption, (2.1) holds if and only if \( T \mu^* = P_{L_2(W)} Y \). But a \( \mu^* \) satisfying this linear equation may not be uniquely defined. So, noting that \( \mu^* = P_{N(T)\perp} \mu^* + P_{N(T)} \mu^* \), write \( T \mu^* = P_{L_2(W)} Y \) as

\[
(T|_{N(T)\perp}) P_{N(T)\perp} \mu^* = P_{L_2(W)} Y,
\]

where \( T|_{N(T)\perp} \) is the restriction of \( T \) to \( N(T)\perp \). Since \( N(T|_{N(T)\perp}) = \{0\} \), it makes sense to call \( P_{N(T)\perp} \mu^* \) the identifiable-part of \( \mu^* \). Of course, if \( \mu^* \) is identified to begin with, i.e., \( N(T) = \{0\} \), then \( P_{N(T)\perp} \mu^* = \mu^* \) and there is no distinction between \( \mu^* \) and its identifiable-part.

Although \( T|_{N(T)\perp} \) is invertible on its range, the inverse may not be continuous. In fact, since \( (T|_{N(T)\perp})^{-1} \) is a closed map from \( \mathcal{R}(T) \) to \( N(T)\perp \), by the closed graph theorem it follows that \( (T|_{N(T)\perp})^{-1} \) is continuous if and only if \( \mathcal{R}(T) \) is closed. Therefore, following the definition of ill-posed linear equations given in [Kress (1999), p. 266], lack of closedness of \( \mathcal{R}(T) \) is equivalent to the ill-posedness of (3.1).

Functionals of \( \mu^* \) are identifiable under conditions weaker than those required for identification of \( \mu^* \) itself. In particular, even when \( \mu^* \) is not identified, i.e., \( N(T) \neq \{0\} \),

**Lemma 3.1.** \( \mathbb{E}[\psi \mu^*] \) is identified if and only if \( \psi \in N(T)\perp \).

If \( \mu^* \) happens to be identified, then \( \mathbb{E}[\psi \mu^*] \) is uniquely defined for every \( \psi \in L_2(X) \) because then \( N(T)\perp = L_2(X) \). The identification condition \( \psi \in N(T)\perp \) ensures that every \( \mu \in P_{N(T)\perp} \mu^* + N(T) \) yields the same expectation functional, namely, \( \mathbb{E}[\psi \mu] = \mathbb{E}[\psi P_{N(T)\perp} \mu^*] = \mathbb{E}[\psi \mu^*] \). Therefore, without loss of generality, we henceforth focus on \( \mathbb{E}[\psi P_{N(T)\perp} \mu^*] \) as the parameter of interest.

4. Ill-posedness and \( n^{1/2} \)-estimability

As mentioned earlier, ill-posedness of (3.1) can lead to very poor rates of convergence for estimators of \( P_{N(T)\perp} \mu^* \). In fact, convergence can be so slow that \( n^{1/2} \)-estimability of
$\mathbb{E}[\psi P_{N(T)}] \mu^*$ may not be possible for certain well behaved $\psi$. [An obvious exception is $\psi := 1$, for which $\mathbb{E}[\psi P_{N(T)}] = \mathbb{E}Y$ is $n^{1/2}$-estimable irrespective of the correlation between $X$ and $\varepsilon$.] The aim of this section is to characterize the $\psi$’s for which the corresponding expectation functionals are not $n^{1/2}$-estimable and make precise the connection between ill-posedness of (3.1) and $n^{1/2}$-estimability of $\mathbb{E}[\psi P_{N(T)}] \mu^*$.

To motivate these results, we begin with a simple but revealing example.

**Example 4.1.** Let $X$ and $W$ be jointly Gaussian with mean zero and variance $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, where the correlation $\rho \in (-1, 1) \setminus \{0\}$. Also, let $\phi$ be the standard normal density, $H_j(x) := (-1)^j \phi^{(j)}(x)/\phi(x)$ the $j$th Hermite polynomial and $h_j := H_j/\sqrt{j!}$ its normalized version. Since Gaussian distributions with varying means are complete, $T$ and $T'$ are both injective. Injectiveness of $T$ implies that $\mu^*$ is identified. [In fact, as shown later, the equation defining $\mu^*$, namely $T\mu^* = P_{L_2(W)}Y$, is also ill-posed since $\mathcal{R}(T)$ is not closed.] Hence, $\mathbb{E}[\psi \mu^*]$ is identified for every $\psi \in L_2(X)$. The reproducing property of Hermite polynomials (cf. Severini and Tripathi 2006, Example 2.4) implies that $T$ is Hilbert-Schmidt with singular system $(\rho^j, h_j(X), h_j(W))_{j \in \{0\} \cup \mathbb{N}}$. Its singular value decomposition (cf. Carrasco, Florens, and Renault 2007, Theorem 2.41) can be used to show that $T^{-1}b = \sum_{j=0}^{\infty} \rho^{-j} \langle b, h_j \rangle_{L_2(W)} h_j$ whenever $b \in \mathcal{R}(T)$. Therefore, since $\mu^* = T^{-1}P_{L_2(W)}Y$, in this example

$$
\mu^*(X) = \sum_{j=0}^{\infty} \rho^{-j} \mathbb{E}[Y h_j(W)] h_j(X). \quad (4.1)
$$

We now show that the sample analog of the expectation functional $\mathbb{E}[\psi \mu^*]$, although identified for each $\psi \in L_2(X)$, is not $n^{1/2}$-consistent for at least one well behaved $\psi$. Of course, this in itself does not prove that certain expectation functionals may not be $n^{1/2}$-estimable. But the fact that the sample analog of an expectation — probably its most obvious consistent estimator — can fail to converge at the $n^{1/2}$-rate is nonetheless very suggestive.

First, it is clear from (4.1) that $\mathbb{E}[\psi_K \mu^*]$ is $n^{1/2}$-estimable whenever $\psi_K$ is a polynomial of degree $K \in [0, \infty)$. This is because $\langle X^K, h_j \rangle_{L_2(X)} = 0$ for $j > K$, implying that $\mathbb{E}[\psi_K \mu^*]$ consists of a finite number of summands each of which is $n^{1/2}$-estimable.

Next, let $\psi_d := \mathbbm{1}_{(-\infty, d]}$, where $d < \infty$ is a known constant. The following argument reveals that $\theta_d^\ast := \mathbb{E}[\psi_d \mu^*]$ cannot be estimated at $n^{1/2}$-rate by its sample analog. Begin by observing that

$$
\int_{-\infty}^{d} H_j(x) \phi(x) \, dx = \begin{cases} \Phi(d) & \text{if } j = 0 \\ -H_{j-1}(d) \phi(d) & \text{if } j \geq 1. \end{cases} \quad (4.2)
$$

It is then straightforward to show that

$$
\theta_d^\ast \overset{(1.3)}{=} \mathbb{E}[Y (\Phi(d) - \phi(d) \sum_{j=0}^{\infty} h_{j+1}(W)/\rho^{j+1} \sqrt{j+1} h_j(d))] =: \mathbb{E}[Y Q_d(W)].
$$
Now consider the estimator \( \hat{\theta}_d := \sum_{j=1}^n Y_j Q_d(W_j)/n \), where we have assumed that \( \rho \) is known to keep things simple. Clearly, \( \hat{\theta}_d \) is consistent for \( \theta_d^* \). Moreover, assuming that \( \text{var}[Y|W] \) is bounded away from zero,

\[
\text{var}[n^{1/2}\hat{\theta}_d] = \text{var}[YQ_d(W)] \geq \mathbb{E}\text{var}[YQ_d(W)|W] \\
\geq \inf_{w \in \text{supp}(W)} \text{var}[Y|W = w] \mathbb{E}[Q_d^2(W)].
\]

But, by the orthonormality of Hermite polynomials,

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

Moreover, since \( (j+1)\rho^{2j} < 1 \) for all sufficiently large \( j \), there exists a positive integer \( N \) such that, for \( \tilde{H}_j(x/\sqrt{2}) := 2^{j/2}H_j(x) \), we have

\[
\sum_{j=0}^\infty \frac{H_j^2(d)}{(j+1)!\rho^{2j}} \geq \sum_{j=N}^\infty \frac{H_j^2(d)}{j!} = \sum_{j=N}^\infty \frac{\tilde{H}_j^2(d/\sqrt{2})}{j!2^j} = \infty,
\] (4.3)

where the last equality is because \( \sum_{j=0}^\infty \tilde{H}_j^2(x)r^j/(j!2^j) < \infty \) for \( x \in \mathbb{R} \) if and only if \( |r| < 1 \); cf. the second proof of (6.1.13) in \cite{Andrews1999}. Therefore, it follows that \( \text{var}[n^{1/2}\hat{\theta}_d] = \infty \). In other words, \( \hat{\theta}_d \) is not \( n^{1/2} \)-consistent. Incidentally, it is easy to see that \( \text{var}[\hat{\theta}_d] \) goes to zero slower than \( 1/n \) even when \( Q_d \) 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 \( \lim_{n \to \infty} m_n = \infty \).

Although it may not be obvious, the fundamental feature that distinguishes \( \psi_K \) from \( \psi_d \) is that the former lies in \( \mathcal{R}(T) \) whereas the latter does not. This makes sense because, as we show next, elements of \( \mathcal{R}(T') \) have to be smooth in a certain sense.

Singular value decompositions of \( T \) and \( T' \) can be used to show that

\[
\mathcal{R}(T) = \{ b \in L_2(W) : \sum_{j=0}^\infty \langle b, h_j \rangle_{L_2(W)}^2 \rho^{-2j} < \infty \} \subset L_2(W)
\]

\[
\mathcal{R}(T') = \{ a \in L_2(X) : \sum_{j=0}^\infty \langle a, h_j \rangle_{L_2(X)}^2 \rho^{-2j} < \infty \} \subset L_2(X).
\] (4.4)

Note that denseness is a consequence of duality, i.e., \( \mathcal{N}(T)^\perp = \text{cl}(\mathcal{R}(T')) \), plus injectivity of \( T \) and \( T' \). Since \( \mathcal{R}(T) \) and \( \mathcal{R}(T') \) are dense albeit proper subspaces of \( L_2(W) \) and \( L_2(X) \), they cannot be closed. In particular, non-closedness of \( \mathcal{R}(T) \) implies that the equation defining \( \mu^* \) is ill-posed.

It is clear from (4.4) that the Fourier coefficients of elements of \( \mathcal{R}(T) \) and \( \mathcal{R}(T') \) have to go to zero sufficiently fast. In fact, elements of \( \mathcal{R}(T) \) and \( \mathcal{R}(T') \) are infinitely differentiable in mean-square with each derivative being square-integrable. To see this, let \( b \in \mathcal{R}(T) \) and
\[ b^{(k)} \] denote its kth derivative. Since \( H_j^{(1)} = jH_{j-1} \), it is straightforward to show that \( b^{(k)} = \sum_{j=k}^{\infty} b^{(k)} L_{\Delta(W)}(j) h_{j-k} \), where \( (j) := j(j-1) \ldots (j-k+1) \). Furthermore, \( \| b^{(k)} \|^2_{L_{\Delta(W)}} = \sum_{j=k}^{\infty} b^{(k)} L_{\Delta(W)}(j) \) for each \( k \) since \( \lim_{j \to \infty} \rho^{2j} \rho^{2j} = 0 \). Same holds for \( \mathcal{R}(T') \) as well.

It only remains to verify that \( \psi \notin \mathcal{R}(T') \) — note that \( \psi \notin \mathcal{R}(T') \) is obvious because \( \langle X^k, h_j \rangle_{\mathcal{L}_{\Delta(X)}} = 0 \) for \( j > K \). But this is immediate since

\[
\sum_{j=0}^{\infty} \langle \psi_d, h_j \rangle_{\mathcal{L}_{\Delta(X)}}^2 \rho^{-2j} \phi^2(d) + \frac{\phi^2(d)}{\rho^2} \sum_{j=0}^{\infty} \frac{H_j^2(d)}{(j+1)!} \rho^{2j} \geq \infty.
\]

Therefore, \( \psi \notin \mathcal{R}(T') \). \( \square \)

Example 4.1 suggests that a \( \psi \) that is not sufficiently smooth relative to \( T' \), in the sense that \( \psi \notin \mathcal{R}(T') \), will lead to an expectation functional that is not \( n^{1/2}\)-estimable. For such a \( \psi \), as the proof of Lemma 4.1 reveals, the parameter of interest \( \mathbb{E}[\psi P_{N(T)\perp} \mu^*] \) is not a differentiable function of the distribution of \( (Y, X, W) \). By a well-known result, cf. van der Vaart (1991, p. 185; 1998, Section 25.5) and Newey (1994, p. 1353), it then follows that \( \mathbb{E}[\psi P_{N(T)\perp} \mu^*] \) cannot be estimated at \( n^{1/2}\)-rate.

Let \( \tilde{\varepsilon} : = Y - P_{N(T)\perp} \mu^* \), and \( \Omega : = P_{\mathcal{L}_2(W)} \varepsilon^2 \) be the scedastic function. The next assumption bounds \( \Omega \) and \( \tilde{\Omega} := 1/\Omega \) away from zero and infinity.

**Assumption 4.1.** \( 0 < \inf_{w \in \text{supp}(W)} \Omega(w) \leq \sup_{w \in \text{supp}(W)} \Omega(w) < \infty \).

Under this assumption we can show the main result of this section.

**Lemma 4.1.** Let Assumption 4.1 hold and \( T' \) be compact. Then, the condition \( \psi \in \mathcal{R}(T') \) is necessary for \( \mathbb{E}[\psi P_{N(T)\perp} \mu^*] \) to be \( n^{1/2}\)-estimable.

Recall that \( \mathbb{E}[\psi P_{N(T)\perp} \mu^*] \) is identified if and only if \( \psi \in N(T)^{\perp} = \text{cl}(\mathcal{R}(T')) \). Thus \( \psi \in \mathcal{R}(T') \) seems like a “natural” requirement for \( n^{1/2}\)-estimability of \( \mathbb{E}[\psi P_{N(T)\perp} \mu^*] \) because it strengthens the identification condition.

The assumption that \( T' \) (equivalently, \( T \)) is compact is not very restrictive, at least from an applied point of view, since compact operators are frequently encountered in applied work (it is well known that compact operators are precisely those that can be approximated arbitrarily well by finite rank operators, i.e., matrices). Moreover, we only use compactness of \( T' \) to show Lemma 4.1; it is not needed to prove our efficiency bound results. Note that conditional expectation operators can be shown to be compact under weak conditions on the joint density; cf. Bickel, Klassen, Ritov, and Wellner (1993, p. 440) and Kress (1999, Theorem 2.21).

Finally, here’s the connection between ill-posedness of (3.1) and \( n^{1/2}\)-estimability of \( \mathbb{E}[\psi P_{N(T)\perp} \mu^*] \). Recall that \( \mathcal{R}(T') \) is closed if and only if \( \mathcal{R}(T) \) is closed (van der Vaart 1991, p. 184). Therefore, \( \text{cl}(\mathcal{R}(T')) \setminus \mathcal{R}(T') \) is empty \( \iff \mathcal{R}(T') \) is closed \( \iff \mathcal{R}(T) \) is closed \( \iff \).
(3.1) is well-posed. Thus, ill-posedness of (3.1) implies the existence of at least one expectation functional of $P_{N(T)^{\perp}} \mu^*$ that is identified but (by Lemma 4.1) not $n^{1/2}$-estimable; Example 4.1 provides a nice illustration.

**Remarks.** (i) Ritov and Bickel (1990, p. 936) have a result that looks similar to Lemma 4.1. They define a class $P$ of large dimensional parametric models and show that if the true model lies in $\text{cl}(P) \setminus P$ then it cannot be consistently estimated. Darolles et al. also impose a similar condition (cf. their Theorem 4.3) although they do not show that their condition is necessary for $n^{1/2}$-consistency.

(ii) If $\psi \in \mathcal{R}(T')$, then the efficiency bound for estimating $\mathbb{E}[\psi P_{N(T)^{\perp}} \mu^*]$ is finite (cf. Theorem 5.1). The condition $\psi \in \mathcal{R}(T')$ plus some smoothness on $\mu^*$ and the joint distribution of $(Y, X, W)$ as in Ritov and Bickel may thus also be sufficient for $n^{1/2}$-estimability.

(iii) If there are no endogenous regressors, i.e., $W = X$, then $\psi \in \mathcal{R}(T')$ holds automatically because $T'$ then is just the identity.

5. **The efficiency bound**

Following the discussion in Section 3, let $\theta^* := \mathbb{E}[\psi P_{N(T)^{\perp}} \mu^*]$ denote the parameter of interest. In this section we determine the efficiency bound for estimating $\theta^*$ when

**Assumption 5.1.** $\psi \in \mathcal{R}(T')$, i.e., there exists $\delta^* \in L^2(W)$, not necessarily uniquely defined, such that $T' \delta^* = \psi$.

For maximum generality, the bound is derived under minimal assumptions on $\mu^*$. In particular, $\mu^*$ is allowed to be underidentified, i.e., $\mathcal{N}(T) \neq \{0\}$, and the equation defining $P_{N(T)^{\perp}} \mu^*$ is allowed to be ill-posed, i.e., $\mathcal{R}(T)$ is not assumed to be closed. Subsequent results simplify accordingly if $\mu^*$ is identified or (3.1) is well-posed.

To facilitate presentation, we express $\theta^*$ as the solution to a moment condition; i.e., we obtain the efficiency bound for estimating $\theta^*$ in the model

$$\mathbb{E} g(X, \theta^*, \mu^*) = 0,$$ (5.1)

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

The efficiency bound for estimating $\theta^*$ is the squared-length of an orthogonal projection onto the tangent space of nonparametric score functions. The tangent space here is $\text{cl}(\mathcal{M}) + L_{2,0}(W)$, where $\mathcal{M} := \{f \in L^2(W)^{\perp} : P_{L^2(W)}(\tilde{\varepsilon} f) \in \mathcal{R}(T)\}$ and $L_{2,0}(W) := \{f \in L^2(W) : \mathbb{E} f = 0\}$ describe how (3.1) restricts the scores of $\text{Law}(Y, X | W)$ and $\text{Law}(W)$. In the appendix we show that $\text{cl}(\mathcal{M}) = \{f \in L^2(W)^{\perp} : P_{L^2(W)}(\tilde{\varepsilon} f) \in \text{cl}(\mathcal{R}(T))\}$. If there are no endogenous regressors, then $\text{cl}(\mathcal{M}) = L^2(W)^{\perp} = L^2(X)^{\perp}$ because $T$ then is the identity map. Therefore, the size of $\text{cl}(\mathcal{M})$ is a measure of the information contained in (3.1), namely, smaller $\text{cl}(\mathcal{M})$ means more information; cf. Examples 5.1 and 5.2.
Theorem 5.1. Let Assumptions 4.1 and 5.1 hold. Then, the efficiency bound for estimating \( \theta^* \) is finite and is given by

\[
\mathbb{E}[P_{\text{cl}(\tilde{X})+L_{2,0}(W)}(\tilde{\varepsilon} P_{\text{cl}(\tilde{X})}) \delta^* + g)]^2, \tag{5.2}
\]

where \( \delta^* \in L_2(W) \) satisfies \( T^T \delta^* = \psi \).

From the discussion in Section 3, recall that \( P_{\mathcal{N}(T)} \mu^* \) is uniquely defined for every \( \mu^* \) that satisfies (2.1). Similarly, \( P_{\text{cl}(\mathcal{R}(T))} \delta^* \) is uniquely defined for every \( \delta^* \) satisfying \( T^T \delta^* = \psi \) because \( \text{cl}(\mathcal{R}(T)) = \mathcal{N}(T') \). Therefore, the tangent space and the efficient influence function \( P_{\text{cl}(\tilde{X})+L_{2,0}(W)}(\tilde{\varepsilon} P_{\text{cl}(\tilde{X})}) \delta^* + g) \) are invariant to choice of \( \mu^* \) and \( \delta^* \) in the sense that each \( (\mu^*, \delta^*) \in L_2(X) \times L_2(W) \) satisfying \( T \mu^* = P_{L_2(W)} Y \) and \( T^T \delta^* = \psi \) leads to the same efficiency bound. Hence, the bound derived above is robust to underidentification of \( \mu^* \) and \( \delta^* \). Similarly, since \( \mathcal{R}(T) \) enters (5.2) only via \( \text{cl}(\mathcal{R}(T)) \), the same bound holds whether (3.1) is ill-posed or not. Finiteness of the efficiency bound suggests that \( n^{1/2} \)-estimation may be possible.

If there are no endogenous regressors, then \( \delta^* = \psi \), and \( \mu^* = P_{L_2(X)} Y \); consequently, the efficiency bound reduces to \( \text{var}[\psi Y] \). This makes sense because if \( W = X \), then \( \mathbb{E}[\psi \mu^*] = \mathbb{E}[\psi Y] \); it also matches the result obtained earlier by Chamberlain (1992, p. 572).

Example 5.1 (Efficiency bound for estimating \( \mathbb{E} Y \)). Suppose \( \psi = 1 \). Then \( \theta^* = \mathbb{E} Y \) irrespective of whether \( \mu^* \) is identified or not. Therefore, by Theorem 5.1 and the fact that \( \text{cl}(\tilde{M}) + L_{2,0}(W) \) is closed, the efficiency bound for estimating \( \mathbb{E} Y \) is given by

\[
\mathbb{E}[P_{\text{cl}(\tilde{M})+L_{2,0}(W)}(Y - \mathbb{E} Y)]^2 = \text{var} Y - \mathbb{E}[P_{\tilde{M} \cap L_{2,0}(W)}(Y - \mathbb{E} Y)]^2.
\]

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

The following corollary of Theorem 5.1 is immediate.

Corollary 5.1. If \( \mu^* \) is identified, then (5.2) can be written as

\[
\mathbb{E}[P_{\text{cl}(\tilde{M})+L_{2,0}(W)}(\varepsilon P_{\text{cl}(\tilde{X})}) \delta^* + g)]^2,
\]

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

Ai and Chen (2005, 2009) give an expression for efficiency bounds of functionals of \( \mu^* \) for the case in which \( \mu^* \) is identified, using an approach different than the one used here. The approach used by Ai and Chen is based on the residuals from projecting the score with respect to the parameter of interest onto the space spanned by the nuisance parameter score; the approach used here is based on finding the norm of the representer of the pathwise derivative of the functional, viewed as a linear functional on the tangent space. One consequence of the approach used here is that the efficiency bound can be given explicitly, rather than as the solution to a variational problem.
Although Theorem 5.1 and Corollary 5.1 provide precise variational characterizations of the efficiency bound for estimating \( \theta^* \), 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 \( P_{\text{cl}(\hat{M})+L_{2,0}(W)} \) is available. Fortunately, an explicit expression for orthogonal projections onto \( \text{cl}(\hat{M}) + L_{2,0}(W) \) can be obtained by using Lemma 5.1, a result that may be of independent interest.

Let \((T'\hat{\Omega}T)^+\) denote the Moore-Penrose inverse of \( T'\hat{\Omega}T : L_2(X) \to L_2(X) \), cf. Engl, Hanke, and Neubauer (2000, Section 2.1), and let \( I \) be the identity operator (the domain of \( I \) will be clear from the context). Keep in mind that \( \mathcal{D}((T'\hat{\Omega}T)^+) = \mathcal{R}(T'\hat{\Omega}T) + \mathcal{R}(T'\hat{\Omega}T)^\perp \).

**Lemma 5.1.** Let Assumption 4.1 hold and \( f \in L_2(Y,X,W) \) be such that \( T'\hat{\Omega}P_{L_2(W)}(\tilde{\varepsilon} f) \) lies in the domain of \((T'\hat{\Omega}T)^+\). Then,

\[
P_{\text{cl}(\hat{M})} f = f - P_{L_2(W)} f - \tilde{\varepsilon}(I - \hat{\Omega}T(T'\hat{\Omega}T)^+T')\hat{\Omega}P_{L_2(W)}(\tilde{\varepsilon} f).
\]

Since \( \hat{M} \perp L_2(W) \) and \( P_{L_{2,0}(W)} f = P_{L_2(W)} f - \mathbb{E} f \),

\[
P_{\text{cl}(\hat{M})+L_{2,0}(W)} f = P_{\text{cl}(\hat{M})} f + P_{L_{2,0}(W)} f = P_{\text{cl}(\hat{M})} f + P_{L_2(W)} f - \mathbb{E} f.
\]

Therefore, an immediate corollary of Lemma 5.1 is that

\[
P_{\text{cl}(\hat{M})+L_{2,0}(W)} f = f - \mathbb{E} f - \tilde{\varepsilon}(I - \hat{\Omega}T(T'\hat{\Omega}T)^+T')\hat{\Omega}P_{L_2(W)}(\tilde{\varepsilon} f). \tag{5.3}
\]

We use (5.3) to derive a closed form for the efficiency bound in Theorem 5.1.

**Corollary 5.2.** Let \( \psi \in \mathcal{D}((T'\hat{\Omega}T)^+) \cap \mathcal{N}(T)^\perp \) and \( T'\hat{\Omega}P_{L_2(W)}(\tilde{\varepsilon} g) \in \mathcal{D}((T'\hat{\Omega}T)^+) \). Then, under the assumptions maintained in Theorem 5.1, (5.2) can be written as

\[
\mathbb{E}[\varepsilon \hat{\Omega}T(T'\hat{\Omega}T)^+ \psi + g - \varepsilon(I - \hat{\Omega}T(T'\hat{\Omega}T)^+T')\hat{\Omega}P_{L_2(W)}(\tilde{\varepsilon} g)]^2. \tag{5.4}
\]

If \( \mu^* \) is identified, then the closed form of the bound can be obtained by replacing \( \tilde{\varepsilon} \) with \( \varepsilon \) and \((T'\hat{\Omega}T)^+ \) with \((T'\hat{\Omega}T)^{-1}\) because \( \mathcal{N}(T'\hat{\Omega}T) = \mathcal{N}(T) \); i.e., if \( \mu^* \) is identified, then (5.4) can be written as

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

The non-variational characterization leads to some additional insight behind the form of the bound. To see this, assume that \( \mu^* \) is identified. Then, from Corollary 5.2, the efficient influence function for estimating \( \theta^* \) is given by

\[
[g - \varepsilon \hat{\Omega}P_{L_2(W)}(\varepsilon g)] + \varepsilon \hat{\Omega}T(T'\hat{\Omega}T)^{-1}(\psi + T'\hat{\Omega}P_{L_2(W)}(\varepsilon g)). \tag{5.5}
\]

A look at the proof of Theorem 5.1 reveals that the efficiency bound for estimating \( \theta^* \) when \( \mu^* \) is fully known is given by \( \mathbb{E}[g - \varepsilon \hat{\Omega}P_{L_2(W)}(\varepsilon g)]^2 \). The first term of (5.5), which has a very intuitive control variate interpretation, thus represents the contribution of \( P_{L_2(W)} \varepsilon = 0 \) if \( \mu^* \) 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
\[ E[\mathcal{U}P_{L_2(W)}(\varepsilon g)]^2 + \mathbb{E}[\mathcal{U}(T'\mathcal{U})^{-1}(\psi + T'\mathcal{U}P_{L_2(W)}(\varepsilon g))]^2. \]

Therefore, the efficiency bound for estimating \( \theta^* \) when \( \mu^* \) is known equals the efficiency bound
for estimating \( \theta^* \) when \( \mu^* \) is unknown if and only if
\[ T'(T'\mathcal{U})^{-1}(\psi + T'\mathcal{U}P_{L_2(W)}(\varepsilon g)) = 0. \]
But since this is a very restrictive condition, e.g., it may not hold even when \( W = X \), adaptive
(meaning invariance with respect to knowledge of \( \mu^* \) or lack thereof) estimation of \( \theta^* \) appears
for all practical purposes to be impossible.

Finally, we describe the efficiency bound for estimating \( \int \psi \mu^* \). The proofs of Corol-
laries 5.3–5.4 are very similar to those of Theorem 5.1 and Corollary 5.2 and are therefore
omitted.

**Corollary 5.3.** Let Assumption 4.1 hold and assume there exists \( \delta^* \in L_2(W) \) such that
\( T'\delta^* = \psi/h \), where \( h \) is the unknown density of \( X \). Then, the efficiency bound
for estimating \( \int \psi \mu^* \) is given by \( \mathbb{E}[\mathcal{U}P_{cl(M)}(\varepsilon P_{cl(X(T))}\delta^*)]^2 \). The bound when \( \mu^* \) is identified is obtained by replacing \( \varepsilon \) with \( \bar{\varepsilon} \).

In case of no endogeneity the above bound reduces to \( \mathbb{E}[\psi h] \), a result obtained earlier
by [Severini and Tripathi] (2001, Section 7). As before, Lemma 5.1 can be used to derive a
closed form expression for the bound.

**Corollary 5.4.** Let \( \psi/h \in \mathcal{D}(T'\mathcal{U}) \cap \mathcal{N}(T)^\perp \). Then, under the assumptions maintained
in Corollary 5.3 the efficiency bound obtained there can be written as \( \mathbb{E}[\mathcal{U}(T'\mathcal{U})^+(\psi/h)]^2 \). The bound when \( \mu^* \) is identified is obtained by replacing \( \varepsilon \) with \( \bar{\varepsilon} \) and \( T'\mathcal{U} \) with \( T'\mathcal{U}^{-1} \).

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

**Example 5.2** (Efficiency bound for probabilities). Let the vector \( Z \) contain \( Y \) and the distinct
components of \( X \) and \( W \). Then, modifying the proof of Theorem 5.1 it can be shown that the
efficiency bound for estimating \( p := \Pr(Z \in A) \), where \( A \) is a known Borel set, is given by
\[ \mathbb{E}[\mathcal{U}P_{cl(M)+L_2(W)}(\mathbb{1}(Z \in A) - p)]^2 = p(1 - p) - \mathbb{E}[\mathcal{U}(T'\mathcal{U})^+(\mathbb{1}(Z \in A) - p)]^2. \]
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 \( p \).

6. Conclusion

We derive a necessary condition for \( n^{1/2} \)-estimability as well as the efficiency bounds for
estimating \( \mathbb{E}[\psi \mu^*] \) and \( \int \psi \mu^* \) when \( \mu^* \) is underidentified and the model defining it is ill-posed.
Appendix A. Proofs

Proof of Lemma 3.1. Suppose \( \theta^* := \mathbb{E}[\psi \mu^*] \) is identified. This means that \( \mu^* \) and \( \mu^* + f \), where \( f \in \mathcal{N}(T) \) is arbitrary, both yield the same value of \( \theta^* \). Hence, \( \psi \in \mathcal{N}(T)^\perp \). Conversely, assume \( \psi \in \mathcal{N}(T)^\perp \) and let \( \theta_i^* := \mathbb{E}[\psi \mu_i^*] \), where \( \mu_1^* \) and \( \mu_2^* \) both satisfy (2.1), i.e., \( T \mu_i^* = P_{L_2(W)} Y \) for \( i = 1, 2 \). Then, \( \theta_1^* - \theta_2^* = \mathbb{E}[\psi (\mu_1^* - \mu_2^*)] = 0 \) since \( \mu_1^* - \mu_2^* \in \mathcal{N}(T) \). Hence, \( \theta^* \) is identified. \( \square \)

The proof of Lemma 4.1 appears after the proof of Theorem 5.1 because it uses notation introduced in the latter.

Proof of Theorem 5.1. Let \( v_0^2 \) be the conditional density of \( (Y, X)|W \) with respect to a dominating measure \( \lambda \) and \( b_0^2 \) the marginal density of \( W \) with respect to a dominating measure \( \gamma \). Let \( v_t \) be a real-valued function on an interval \( I_0 \ni 0 \) such that \( v_t|_{t=0} = v_0 \) and \( \int_{\text{supp}(Y,X)} v_t^2(y, x|w) \, d\lambda = 1 \) for all \( (t, w) \in I_0 \times \text{supp}(W) \). Similarly, \( b_t \) is a curve through \( b_0 \) satisfying \( \int_{\text{supp}(W)} b_t^2(w) \, d\gamma = 1 \) for all \( t \in I_0 \). Using \( \hat{\tau} = (\hat{v}, \hat{b}) \) to denote the tangent vector to \( (v_t, b_t) \) at \( t = 0 \), we have

\[
\hat{v} \in \hat{\mathcal{V}} := \{ S_v \in L_2(Y, X, W) : P_{L_2(W)} S_v = 0 \}
\]

\[
\hat{b} \in L_{2,0}(W) := \{ S_b \in L_2(W) : \mathbb{E}S_b = 0 \},
\]

where \( S_v := 2\hat{v}/v_0 \) and \( S_b := 2\hat{b}/b_0 \) are the score functions corresponding to \( \hat{v} \) and \( \hat{b} \), respectively. Since \( \hat{\mathcal{V}} = L_2(W)^\perp \), it is clear that \( \hat{\mathcal{V}} \perp L_{2,0}(W) \).

Let \( \kappa_t \) be a curve from \( I_0 \) into \( \mathcal{N}(T)^\perp \), passing through \( P_{\mathcal{N}(T)^\perp \mu^*} \) at \( t = 0 \), such that \( \mathbb{E}_t[Y - \kappa_t|W = w] = 0 \) for all \( (t, w) \in I_0 \times \text{supp}(W) \), where \( \mathbb{E}_t \) denotes conditional expectation under the submodel \( v_t^2 \). Hence, differentiating with respect to \( t \) and evaluating at \( t = 0 \), for some \( \check{\kappa} \in \mathcal{N}(T)^\perp \),

\[
T \check{\kappa} = P_{L_2(W)}(\check{\mathcal{E}} S_v).
\]

(A.1)

Since (A.1) further restricts \( \hat{\mathcal{V}} \), the conditional scores lie in

\[
\hat{\mathcal{M}} := \{ f \in L_2(W)^\perp : P_{L_2(W)}(\check{\mathcal{E}} f) \in \mathcal{R}(T) \}.
\]

(A.2)

Therefore, the tangent space of score functions relevant for our problem is \( \hat{\mathcal{F}} := \text{cl}(\hat{\mathcal{M}}) + L_{2,0}(W) \). As shown in Lemma B.1, an appealing expression for \( \text{cl}(\hat{\mathcal{M}}) \) can be obtained under the assumption that the scedastic function is bounded; namely, \( \text{cl}(\hat{\mathcal{M}}) = \{ f \in L_2(W)^\perp : P_{L_2(W)}(\check{\mathcal{E}} f) \in \text{cl}(\mathcal{R}(T)) \} \). Since \( \text{cl}(\hat{\mathcal{M}}) \) and \( L_{2,0}(W) \) are closed linear subspaces of \( L_2(Y, X, W) \) and \( \hat{\mathcal{M}} \perp L_{2,0}(W) \), the tangent space \( \hat{\mathcal{F}} \) 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 (5.1), the parameter of interest \( \theta^* \) 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 \( \int_{\text{supp}(Y,X,W)} g(x, \eta(v_t, b_t), \kappa_t) v_t^2(y, x|w) b_t^2(w) \, d\lambda \, d\gamma = 0 \) for all \( t \in I_0 \). Differentiating with respect to \( t \) and evaluating at \( t = 0 \), we obtain that
\[
\nabla \eta(\hat{\tau}) = \mathbb{E}[\psi \hat{\kappa}] + \mathbb{E}[g S_\hat{v}] + \mathbb{E}[g S_\hat{b}],
\]
(A.3)
where \( \nabla \eta \) is the derivative of \( \eta \) along one-dimensional paths through \((v_0, b_0)\). Next, we write \( \mathbb{E}[\psi \hat{\kappa}] \) in terms of the tangent vectors so that \( \nabla \eta \) can be expressed as a linear functional on \( \hat{\mathcal{J}} \). Observing that \( \hat{\kappa} \overset{\Delta}{=} T^*P_{L_2(W)}(\hat{\epsilon}S_\hat{v}) \), we have \( \mathbb{E}[\psi \hat{\kappa}] = J_{\psi, \hat{\epsilon}}(S_\hat{v}) \), where, for \( f \in \hat{\mathcal{M}} \),
\[
J_{\psi, \hat{\epsilon}}(f) := \mathbb{E}[\psi T^*P_{L_2(W)}(\hat{\epsilon}f)] = \langle \psi, T^*P_{L_2(W)}(\hat{\epsilon}f) \rangle_{L_2(X)}.
\]
Therefore, we can rewrite (A.3) as
\[
\nabla \eta(\hat{\tau}) = J_{\psi, \hat{\epsilon}}(S_\hat{v}) + \mathbb{E}[g S_\hat{v}] + \mathbb{E}[g S_\hat{b}].
\]
(A.4)
But \( \psi \in \mathcal{R}(T') \), i.e., \( T'\delta^* = \psi \) for \( \delta^* \in L_2(W) \), by Assumption 5.1 Therefore, for \( f \in \hat{\mathcal{M}} \),
\[
J_{\psi, \hat{\epsilon}}(f) = \langle T'^* \psi, P_{L_2(W)}(\hat{\epsilon}f) \rangle_{L_2(W)} \quad \text{(by Lemma B.3(i))}
\]
\[
= \langle P_{\mathcal{R}(T)} \delta^*, P_{L_2(W)}(\hat{\epsilon}f) \rangle_{L_2(W)} \quad \text{(by Lemma B.3(ii))}
\]
\[
= \langle \hat{\epsilon}P_{\mathcal{R}(T)} \delta^*, f \rangle_{L_2(Y,X,W)},
\]
implying, by Assumption 4.1, that \( J_{\psi, \hat{\epsilon}} \) is bounded on \( \hat{\mathcal{M}} \). Hence, it can be uniquely extended to a linear functional that is bounded on \( \text{cl}(\hat{\mathcal{M}}) \).

Consequently, \( \nabla \eta(\hat{\tau}) = \langle \hat{\epsilon}P_{\mathcal{R}(T)} \delta^* + g, S_\hat{v} \rangle_{L_2(Y,X,W)} + \langle g, S_\hat{b} \rangle_{L_2(X,W)} \) and \( \nabla \eta \) is bounded on \( \hat{\mathcal{J}} \). The expression for \( \nabla \eta \) further simplifies to
\[
\nabla \eta(\hat{\tau}) = \langle P_{\text{cl}(\hat{\mathcal{M}}) + L_{2,0}(W)}(\hat{\epsilon}P_{\mathcal{R}(T)} \delta^* + g), S_\hat{v} + S_\hat{b} \rangle_{L_2(Y,X,W)}
\]
upon noting that \( \hat{\epsilon}P_{\mathcal{R}(T)} \delta^* \in L_2(W)^+ \) and \( S_\hat{v} + S_\hat{b} \in \text{cl}(\hat{\mathcal{M}}) + L_{2,0}(W) \). Following Severini and Tripathi (2001), the efficiency bound for estimating \( \eta(v_0, b_0) \) is given by \( \| \nabla \eta \|^2 \), the squared operator norm of its derivative, where \( \| \nabla \eta \| := \sup \{ |\nabla \eta(\hat{\tau})| : \hat{\tau} \in \hat{\mathcal{J}} \setminus \{0\} \} \). Therefore,
\[
\| \nabla \eta \|^2 = \mathbb{E}[P_{\text{cl}(\hat{\mathcal{M}}) + L_{2,0}(W)}(\hat{\epsilon}P_{\mathcal{R}(T)} \delta^* + g)]^2 < \infty.
\]
\[
\square
\]

**Proof of Lemma 4.1.** To show that \( \mathbb{E}[\psi P_{\mathcal{R}(T')}^+ \mu^*] \) is not \( n^{1/2} \)-estimable, it is enough to demonstrate that \( \nabla \eta \) is unbounded on the tangent space. Since this implies that \( \eta \) is not a differentiable functional of \((v_0, b_0)\), the parameter \( \mathbb{E}[\psi P_{\mathcal{R}(T')}^+ \mu^*] \) cannot be estimated at \( n^{1/2} \)-rate; cf. van der Vaart (1991, p. 185; 1998, Section 25.5) and Newey (1994, p. 1353).

Begin by observing that \( f \mapsto \mathbb{E}[\psi T^*P_{L_2(W)}(\hat{\epsilon}f)] \) is well defined on \( \hat{\mathcal{M}} \) for each \( \psi \in L_2(X) \) because \( f \in \hat{\mathcal{M}} \) implies that \( P_{L_2(W)}(\hat{\epsilon}f) \in \mathcal{R}(T) \subset D(T^+) \). The domain of this linear functional can be enlarged to \( \text{cl}(\hat{\mathcal{M}}) \) if we have additional information about \( \psi \). To see this, assume that \( \psi \in \text{cl}(\mathcal{R}(T')) \) and let \((\lambda_j, a_j, b_j)_{j \in \mathbb{N}} \) denote the singular system for \( T \) and \( T' \), where \((a_j)\) and \((b_j)\) are orthonormal bases for \( \mathcal{N}(T)^+ (= \text{cl}(\mathcal{R}(T'))) \) and \( \text{cl}(\mathcal{R}(T')) \), respectively, and \((\lambda_j)\) are the
nonzero singular values. Then, since \( \psi = \sum_{j=1}^{\infty} F_{a_j}(\psi)a_j \), where \( F_{a_j}(\psi) := \langle \psi, a_j \rangle_{L^2(X)} \) is the \( j \)-th Fourier coefficient of \( \psi \) with respect to \( a_j \),

\[
\mathbb{E}[\psi^T P_{L^2(W)}(\bar{\varepsilon} f)] = \lim_{k \to \infty} \left( \sum_{j=1}^{k} F_{a_j}(\psi)a_j, T^+P_{L^2(W)}(\bar{\varepsilon} f) \right)_{L^2(X)}
\]

\[
= \lim_{k \to \infty} \left( T^+ \sum_{j=1}^{k} F_{a_j}(\psi)a_j, P_{L^2(W)}(\bar{\varepsilon} f) \right)_{L^2(W)}
\]

\[
= \lim_{k \to \infty} \left( P_{\text{cl}(\mathbb{R}(T))}(\sum_{j=1}^{k} \frac{F_{a_j}(\psi)}{\lambda_j} b_j), P_{L^2(W)}(\bar{\varepsilon} f) \right)_{L^2(W)}
\]

by Lemma \( \text{B.3(ii)} \). Therefore, since \((b_j) \subset \text{cl}(\mathbb{R}(T))\),

\[
\mathbb{E}[\psi^T P_{L^2(W)}(\bar{\varepsilon} f)] = \sum_{j=1}^{\infty} \lambda_j^{-1} F_{a_j}(\psi) F_{b_j}(P_{L^2(W)}(\bar{\varepsilon} f)) =: K_\psi(f).
\]

Since \( F_{b_j}(P_{L^2(W)}(\bar{\varepsilon} f)) \) is well defined for \( P_{L^2(W)}(\bar{\varepsilon} f) \in \text{cl}(\mathbb{R}(T)) \), it follows by Lemma \( \text{B.1} \) that the linear functional \( f \mapsto K_\psi(f) \) is well defined on \( \text{cl}(\mathbb{M}) \). Hence, the linear functional \( (S_\psi, S_\nu) \mapsto K_\psi(S_\psi) + \mathbb{E}[gS_\psi] + \mathbb{E}[gS_\nu] \) is well defined on \( \text{cl}(\mathbb{M}) + L^2_0(W) \) and represents the extension of \( \nabla \eta \) in (A.4) to the tangent space. Therefore, any \( \psi \in \text{cl}(\mathbb{R}(T')) \) which makes \( K_\psi \) unbounded on \( \text{cl}(\mathbb{M}) \) will also make \( \nabla \eta \) unbounded on the tangent space. Consequently, by the argument described at the beginning of the proof, for such a \( \psi \) the corresponding expectation functional \( \mathbb{E}[\psi P_{\mathbb{N}(T')} + \mu^*] \) will not be \( n^{1/2} \)-estimable.

Now, let \( f \in \text{cl}(\mathbb{M}) \). By Lemma \( \text{B.1} \) this is equivalent to assuming that \( P_{L^2(W)}(\bar{\varepsilon} f) \in \text{cl}(\mathbb{R}(T)) \). Since \( F_{b_j}(P_{L^2(W)}(\bar{\varepsilon} f)) := (P_{L^2(W)}(\bar{\varepsilon} f), b_j)_{L^2(W)} \), by Cauchy-Schwarz and Bessel

\[
|K_\psi(f)|^2 \leq \sum_{j=1}^{\infty} \lambda_j^{-2} F_{a_j}^2(\psi) \|P_{L^2(W)}(\bar{\varepsilon} f)\|_{L^2(Y,X,W)}^2
\]

\[
\lesssim \sum_{j=1}^{\infty} \lambda_j^{-2} F_{a_j}^2(\psi) \|f\|_{L^2(Y,X,W)}^2,
\]

where the second inequality holds by Assumption \( \text{A.1} \) and the \( \lesssim \) symbol signifies that the left hand side is bounded from above by a positive constant times the right hand side. Hence, since \( \mathbb{R}(T') = \{ a \in L^2(X) : \sum_{j=0}^{\infty} \lambda_j^{-2} (a, a_j)_{L^2(X)}^2 < \infty \} \) by the singular value decomposition of \( T' \), it follows that \( K_\psi \) is bounded on \( \text{cl}(\mathbb{M}) \) whenever \( \psi \in \mathbb{R}(T') \). Therefore, \( K_\psi \) can be unbounded on \( \text{cl}(\mathbb{M}) \) only if \( \psi \in \text{cl}(\mathbb{R}(T')) \setminus \mathbb{R}(T') \).
We now show that $\psi \in \mathcal{R}(T')$ is necessary for $K_\psi$ to be bounded on $\text{cl}(\mathcal{M})$. So let $\psi_0 \in \text{cl}(\mathcal{R}(T')) \setminus \mathcal{R}(T')$ (remember that $\sum_{j=1}^{\infty} \lambda_j^{-2} F_{a_j}^2(\psi_0) = \infty$). For each $r \in \mathbb{N}$,

$$d_r := \varepsilon \delta \sum_{i=1}^{r} \lambda_i^{-1} F_{a_i}(\psi_0) b_i \in L_2(Y, X, W),$$

$P_{L_2(W)} d_r = 0$, and $P_{L_2(W)}(\tilde{\varepsilon} d_r) = \sum_{i=1}^{r} \lambda_i^{-1} F_{a_i}(\psi_0) b_i \in \text{cl}(\mathcal{R}(T))$. Hence, $d_r \in \text{cl}(\mathcal{M})$ for each $r$ by Lemma [B.1]. Furthermore,

$$F_{b_j}(P_{L_2(W)}(\tilde{\varepsilon} d_r)) = \begin{cases} \lambda_j^{-1} F_{a_j}(\psi_0) & \text{for } j \leq r \\ 0 & \text{otherwise,} \end{cases}$$

implying that $K_{\psi_0}(d_r) = \|d_r\|^2_{L_2(Y, X, W)}$. Since $\#\{j : F_{a_j}(\psi_0) \neq 0\} = \infty$, because otherwise $\psi_0 \notin \mathcal{R}(T')$, assume without loss of generality that $\|d_r\|_{L_2(Y, X, W)} > 0$ for each $r$. Then, $f_r := d_r/\|d_r\|_{L_2(Y, X, W)}$ lies on the unit sphere in $\text{cl}(\mathcal{M})$ and $K_{\psi_0}(f_r) = \|d_r\|_{L_2(Y, X, W)}$. It follows that $(f_r)$ is a sequence of unit vectors in $\text{cl}(\mathcal{M})$ such that

$$\lim_{r \to \infty} K_{\psi_0}(f_r) = \left(\sum_{j=1}^{\infty} \lambda_j^{-2} F_{a_j}^2(\psi_0)\right)^{1/2} = \infty.$$ 

Therefore, $K_{\psi_0}$ is unbounded on $\text{cl}(\mathcal{M})$. 

**Proof of Lemma 5.1** Let $\pi^* := f - P_{L_2(W)} f - \tilde{\varepsilon}(I - \delta T(T'\delta T)^+ T') \delta P_{L_2(W)}(\tilde{\varepsilon} f)$. Clearly, $\pi^* \in L_2(W)^+$ because $P_{L_2(W)} \tilde{\varepsilon} = 0$. Furthermore, since $P_{L_2(W)}(\tilde{\varepsilon}^2) := \Omega$,

$$P_{L_2(W)}(\tilde{\varepsilon} \pi^*) = T(T'\delta T)^+T'\delta P_{L_2(W)}(\tilde{\varepsilon} f) \in \mathcal{R}(T).$$

Hence, $\pi^* \in \hat{\mathcal{M}} \subset \text{cl}(\mathcal{M})$. Next, let $R_T := I - \delta T(T'\delta T)^+ T'$. Then, for every $\hat{m} \in \hat{\mathcal{M}}$,

$$\langle f - \pi^*, \hat{m} \rangle_{L_2(Y, X, W)} = \langle \tilde{\varepsilon} R_T \delta P_{L_2(W)}(\tilde{\varepsilon} f), \hat{m} \rangle_{L_2(Y, X, W)}$$

$$= \langle P_{L_2(W)}(\tilde{\varepsilon} \hat{m}), R_T \delta P_{L_2(W)}(\tilde{\varepsilon} f) \rangle_{L_2(W)} \quad \text{(by iterated expectations)}$$

$$= \langle R_T' P_{L_2(W)}(\tilde{\varepsilon} \hat{m}), \delta P_{L_2(W)}(\tilde{\varepsilon} f) \rangle_{L_2(W)} = 0$$

because from [B.3] we know that $\hat{m} \in \hat{\mathcal{M}}$ implies $R_T' P_{L_2(W)}(\tilde{\varepsilon} \hat{m}) = 0$. Therefore, $f - \pi^* \perp \hat{\mathcal{M}}$; thus $f - \pi^* \perp \text{cl}(\mathcal{M})$ by continuity of the inner product. 

**Proof of Corollary 5.2** Since $\mathbb{E}[\tilde{\varepsilon} P_{\text{cl}(\mathcal{R}(T))} \delta^* + g] = 0$,

$$P_{\text{cl}(\mathcal{M}) + L_2,0(W)}(\tilde{\varepsilon} P_{\text{cl}(\mathcal{R}(T))} \delta^* + g) \xrightarrow{[B.3]} \tilde{\varepsilon} P_{\text{cl}(\mathcal{R}(T))} \delta^* + g$$

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

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

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

\[ T'P_{\text{cl}(R(T))} = T'(I - P_{R(T)^+}) = T'(I - P_{\mathcal{N}(T')}) = T'. \]

Therefore, \( T'P_{\text{cl}(R(T))}\delta^* = T\delta^* = \psi \) by Assumption \[5.1\] and the desired result follows. \( \square \)

**Appendix B. Some Useful Results**

**Lemma B.1.** Let Assumption \[4.1\] hold. Then,

\[ \text{cl}(\hat{M}) = \{ f \in L_2(W)^+ : P_{L_2(W)}(\tilde{\epsilon}f) \in \text{cl}(R(T)) \}, \]  

where \( \hat{M} \) was defined in \[A.2\].

**Proof of Lemma B.1** Let \( f \in \text{cl}(\hat{M}) \). Then, there exists a sequence \((f_k)_{k \in \mathbb{N}} \subset \hat{M} \) such that \( \lim_{k \to \infty} f_k = f \). Thus, for each \( k \), \( f_k \in L_2(W)^+ \) and \( P_{L_2(W)}(\tilde{\epsilon}f_k) = Ta_k \) for some \( a_k \in L_2(X) \). But, by Cauchy-Schwarz and Assumption \[4.1\]

\[ \| P_{L_2(W)}(\tilde{\epsilon}f) \|_{L_2(W)} \leq \| \Omega_{1/2}(P_{L_2(W)}f^2)^{1/2} \|_{L_2(W)} \lesssim \| f \|_{L_2(Y,X,W)}; \]

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

\[ \lim_{k \to \infty} Ta_k = P_{L_2(W)}(\tilde{\epsilon}f) \implies P_{L_2(W)}(\tilde{\epsilon}f) \in \text{cl}(R(T)). \]

Since \( f \in L_2(W)^+ \), because \( f_k \) converges in \( L_2(W)^+ \) and the latter is closed, it follows that “\( \subset \)” holds. To show the reverse inclusion, let \( m \) belong to the right hand side of \[B.1\]. Then, for every \( \epsilon > 0 \), there exists a \( b_\epsilon \in R(T) \) such that

\[ \| b_\epsilon - P_{L_2(W)}(\tilde{\epsilon}m) \|_{L_2(W)} < \epsilon. \]  

(B.2)

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

\[ P_{L_2(W)}(\tilde{\epsilon}\hat{m}_\epsilon) = P_{L_2(W)}(\tilde{\epsilon}m) + P_{L_2(W)}(\tilde{\epsilon}^2\sigma(b_\epsilon - P_{L_2(W)}(\tilde{\epsilon}m))) \]

\[ = b_\epsilon \in R(T), \]

it follows that \( \hat{m}_\epsilon \in \hat{M} \). Finally, by Assumption \[4.1\] and \[B.2\],

\[ \| \hat{m}_\epsilon - m \|_{L_2(Y,X,W)} \lesssim \| b_\epsilon - P_{L_2(W)}(\tilde{\epsilon}m) \|_{L_2(W)} \lesssim \epsilon. \]

Therefore, \( \hat{m}_\epsilon \in \hat{M} \) is arbitrarily close to \( m \). Hence, \( m \in \text{cl}(\hat{M}) \). \( \square \)

**Lemma B.2.** Let Assumption \[4.1\] hold. Then,

\[ \hat{M} = \{ \hat{m} \in L_2(W)^+ : T'\sigma P_{L_2(W)}(\tilde{\epsilon}\hat{m}) \in R(T'\sigma T) \]

\[ \& (I - T(T'\sigma T)^{+'T'\sigma})P_{L_2(W)}(\tilde{\epsilon}\hat{m}) = 0 \}. \]  

(B.3)
Proof of Lemma \[B.2\] Recalling the definition of \( \hat{M} \) from (A.2), let \( \hat{m} \in \hat{M} \). Then, \( \hat{m} \in L_2(W)^\perp \) and there exists \( a \in L_2(X) \) such that

\[
P_{L_2(W)}(\hat{m}) = Ta.
\] (B.4)

Now (B.4) implies that

\[
T' \delta P_{L_2(W)}(\hat{m}) = T' \delta Ta \in \mathcal{R}(T' \delta T).
\] (B.5)

Hence,

\[
a = (T' \delta T)^+ T' \delta P_{L_2(W)}(\hat{m}) \in \mathcal{R}((T' \delta T)^+).
\] (B.6)

But \( \mathcal{R}((T' \delta T)^+) = \mathcal{N}(T' \delta T)^\perp = \text{cl}(\mathcal{R}(T' \delta T)) \) since \( T' \delta T \) is bounded by Assumption 4.1.

Thus, by (B.6),

\[
a \in \text{cl}(\mathcal{R}(T' \delta T)).
\] (B.7)

Hence, \( T' \delta Ta \in \mathcal{D}((T' \delta T)^+) \). Thus, by (B.5), Lemma B.3(ii), and (B.7),

\[
(T' \delta T)^+ T' \delta P_{L_2(W)}(\hat{m}) = (T' \delta T)^+ T' \delta Ta = P_{\text{cl}(\mathcal{R}(T' \delta T))}a = a,
\]

implies that

\[
T(T' \delta T)^+ T' \delta P_{L_2(W)}(\hat{m}) = Ta.
\] (B.8)

Therefore, \( (I - T(T' \delta T)^+ T' \delta)P_{L_2(W)}(\hat{m}) = 0 \) upon subtracting (B.8) from (B.4). In other words, we have shown that “\( \subset \)” holds. The reverse inclusion is straightforward. Let \( R_T := I - \delta T(T' \delta T)^+ T' \) and \( \hat{m} \) be an arbitrary element in the right hand side of (B.3). Then, \( \hat{m} \in L_2(W)^\perp \) and satisfies \( R_T P_{L_2(W)}(\hat{m}) = 0 \) which is equivalent to

\[
P_{L_2(W)}(\hat{m}) = T(T' \delta T)^+ T' \delta P_{L_2(W)}(\hat{m}) \in \mathcal{R}(T).
\]

Hence, “\( \supset \)” also holds.

It is well known, cf., for instance, \cite{Luenberger}, Proposition 1, p. 165, that the generalized inverse and adjoint operations commute for closed-range operators. The following result, which looks very familiar although we have not been able to find it in the literature, shows that something similar also holds for operators whose range may not be closed. We use this result in the paper to derive an expression for the adjoint of \( T^+ \) without assuming that \( T^+ \) is bounded (cf. the proof of Theorem 5.1). It is important to allow \( T^+ \) to be unbounded because its boundedness is equivalent to \( \mathcal{R}(T) \) being closed (cf. \cite{Engl} Proposition 2.4, p. 34) and, consequently, the nonparametric regression model for \( \mu^* \) being well-posed.

Lemma B.3. Let \( A \) and \( B \) be Hilbert spaces and \( Q : A \rightarrow B \) a bounded linear operator whose range is not closed. Also, let \( Q^\ast \) denote the adjoint of \( Q \) and \( a \in \mathcal{R}(Q') \). Then, \( i) a \in \mathcal{D}(Q^\ast) \) and \( ii) Q^\ast a = P_{\text{cl}(\mathcal{R}(Q))}b \), where \( b \in B \) is such that \( Q'b = a \). Consequently, \( Q^\ast a = Q'^\ast a \) whenever \( a \in \mathcal{R}(Q') \).
Proof of Lemma B.3} Since \( R(Q) \) is not closed, the Moore-Penrose inverse \( Q^+ : R(Q) + R(Q)^\perp \rightarrow N(Q)^\perp \) is unbounded. Moreover, \( D(Q^+) \) is a dense subspace of \( B \) and \( R(Q^+) \subset A \). Hence, by Kreyszig (1978, Definition 10.1-2), the operator \( Q^+ : D(Q^+) \rightarrow B \) is such that
\[
D(Q^+) = \{ a \in A : \exists b^* \in B \text{ s.t. } \langle Q^+ f, a \rangle_A = \langle f, b^* \rangle_B \forall f \in D(Q^+) \}
\]
and \( Q^+ a := b^* \). Let \( Q|_{N(Q)^\perp} \) denote the restriction of \( Q \) to \( N(Q)^\perp \). To verify that \( a \) lies in the domain of \( Q^+ \) observe that, for \( f \in R(Q) + R(Q)^\perp \),
\[
\langle Q^+ f, a \rangle_A = \langle (Q|_{N(Q)^\perp})^{-1} P_{cl(\mathbb{R}(Q))} f, Q'|b \rangle_A = \langle P_{cl(\mathbb{R}(Q))} f, b \rangle_B,
\]
implying that \( \langle Q^+ f, a \rangle_A = \langle f, P_{cl(\mathbb{R}(Q))} b \rangle_B \). Furthermore, \( \|P_{cl(\mathbb{R}(Q))} b\|_B \leq \|b\|_B < \infty \) since \( b \in B \). Therefore, \( a \in D(Q^+) \) and \( Q^+ a = P_{cl(\mathbb{R}(Q))} b \). Finally, let \( a_0 \in R(Q') \). Hence, \( a_0 = Q' b_0 \) for some \( b_0 \in B \) and \( Q^+ a_0 = P_{cl(\mathbb{R}(Q))} b_0 \) by (i) and (ii). Since \( Q^+ : R(Q') + R(Q')^\perp \rightarrow N(Q')^\perp \), it is clear that \( a_0 \in D(Q'^+) \). It follows that
\[
Q'^+ a_0 = Q'^+ Q' b_0 = P_{N(Q')^\perp} b_0 = P_{cl(\mathbb{R}(Q))} b_0 = Q^+ a_0. \tag*{\Box}
\]

Acknowledgements

We thank the co-editors and two anonymous referees for comments that greatly improved this paper. We also thank Gary Chamberlain, Enno Mammen, Whitney Newey, and participants at several seminars for helpful suggestions and conversations. The first author is grateful for financial support from the NSF.

References


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

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