Asymptotic Efficiency of Semiparametric Two-step GMM

Xiaohong Chen* Jinyong Hahn† Zhipeng Liao‡
Yale UCLA UCLA

First Version: April 2011; This version: May 2012

Abstract

In this note, we characterize the semiparametric efficiency bound for a class of semi-parametric models in which the unknown nuisance functions are "exactly identified" via nonparametric conditional moment restrictions with possibly non-nested or over-lapping conditioning sets, and the finite dimensional parameters are "over-identified" via unconditional moment restrictions involving the nuisance functions. We discover a surprising result that semiparametric two-step optimally weighted GMM estimators achieve the efficiency bound, where the nuisance functions could be estimated via any consistent non-parametric procedures in the first step. When the nuisance functions are estimated via sieve M estimation in the first step, we provide explicit formula for the asymptotic variance of the second step GMM estimate, and present simple ways to compute standard errors.

JEL Classification: C14, C31, C32

Keywords: Overlapping Information Sets; Semiparametric Efficiency; Two-Step GMM

1 Introduction

In this note, we consider semiparametric efficiency bound and efficient estimation of a finite dimensional parameter of interest $\theta_0$ that is (over-) identified by the unconditional moment

---

*Department of Economics, Yale University, 30 Hillhouse, Box 208281, New Haven, CT 06520, USA. Email: xiaohong.chen@yale.edu
†Department of Economics, UC Los Angeles, 8383 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: hahn@econ.ucla.edu
‡Department of Economics, UC Los Angeles, 8379 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: zhipeng.liao@econ.ucla.edu
restrictions

\[ E[g(Z; \theta, h_o)] = 0, \]  
(1.1)

where the nuisance functions \( h_o = (h_{1,o}, \ldots, h_{L,o}) \) are identified by the conditional moment restrictions

\[ E[\Delta_l(Z_l, h_{l,o}(X_l)) | X_l] = 0, \quad l = 1, \ldots, L, \]  
(1.2)

where the conditioning variables \( X_l, l = 1, \ldots, L \), could be nested, overlapping or non-nested. Here, \( Z_l' = (Y_l', X_l') \) for \( l = 1, \ldots, L \), and \( Z \) contains all the distinct elements of \( Z_1, \ldots, Z_L \), and possibly \( Z_{L+1} = (Y'_{L+1}, X'_{L+1})' \) that could be empty. Finally, \( h_{l,o}(\cdot) \neq h_{j,o}(\cdot) \) for \( l \neq j \) and \( l, j \in \{1, \ldots, L\} \) for fixed finite \( L \geq 1 \).

Given the conditional moment restrictions in (1.2), we can estimate \( h_{l,o} \) by any nonparametric estimator \( \tilde{h}_l \), and then estimate \( \theta_o \) by setting the sample analog \( n^{-1} \sum_{i=1}^n g(Z_i; \theta, \tilde{h}) \) of \( E[g(Z; \theta, h_o)] \) as close to zero as possible, a strategy suggested in Andrews (1994), Newey (1994), Pakes and Olley (1995), Chen, Linton and van Keilegom (2003) and others. This is a ‘limited information’ inference in the sense that the information contained in moment conditions (1.1) and (1.2) are not simultaneously considered. If \( h_o \) were known, it is well-known that the optimal ‘limited information’ inference would take the form of GMM with the weight matrix equal to the inverse of \( E[g(Z; \theta, h_o) g(Z; \theta, h_o)'] \). Because the moment equation used in estimating \( \theta_o \) is actually based on \( n^{-1} \sum_{i=1}^n g(Z_i; \theta, \tilde{h}) \), we can speculate that the optimal ‘limited information’ procedure would now take the form of GMM, where the weight matrix converges in probability to the inverse of the asymptotic variance of \( n^{-1/2} \sum_{i=1}^n g(Z_i; \theta_o, \tilde{h}) \), which is the basis of proposal by Ackerberg, Chen, and Hahn (2012).

Two questions arise. Is this intuitive proposal in fact efficient in the ‘limited information’ sense? That is, we would like to ask if there is any other two-step estimate which is more efficient than the proposed two-step GMM estimate. If the answer is in the affirmative, the next question is whether the ‘limited information’ efficient estimator exhausts all the information contained in both moments (1.1) and (1.2). To the best of our knowledge, there is no published work addressing whether or not the semiparametric two-step GMM estimation is efficient for \( \theta_o \) satisfying the over-identifying moment restriction (1.1). We show that the answers to both questions are ‘Yes’ by showing that the asymptotic variance of the proposed two-step GMM estimator is equal to the inverse of the semiparametric information bound for models characterized by the moments (1.1) and (1.2).

Under the i.i.d. data assumption, we derive the semiparametric efficiency bound for \( \theta_o \) when the unknown parameters \( \alpha_o = (\theta_o, h_o) \) are identified by the set of moment restrictions (1.1) and (1.2). When the conditioning variables \( X_l, l = 1, \ldots, L \), are nested, our efficiency bound recovers those derived in Ai and Chen (2009). To the best of our knowledge, our paper...
is the first to derive efficiency bound for $\theta_o$ that could be over-identified by the unconditional moment restriction (1.1) when the sets of conditional moment restrictions (1.2) could be non-nested or overlapping. We then discover an intriguing result that, when the nuisance functions $h_o = (h_{1,o}, ..., h_{L,o})$ are estimated via any consistent nonparametric procedures in the first step, and when $\theta_o$ is estimated in the second step by GMM using the moment (1.1) with the optimal weight matrix that reflect the noise in estimating the nuisance functions $h_o$, the resulting semiparametric two-step GMM estimators achieve the efficiency bound for $\theta_o$. Our efficiency result indicates that the simple estimator proposed in subsection 5.3 of Ackerberg, Chen and Hahn (2012) is in fact efficient, which would probably surprise them.

The rest of the note is organized as follows. Section 2 derives the semiparametric efficiency bound for $\theta_o$. Section 3 gives the proof of the efficiency bound and provides several examples. Readers who are unfamiliar with the technical material presented in Sections 2 and 3 can jump directly to Section 4, where the main results of this paper are rephrased in a more heuristic way and some of its implications are discussed. Section 5 provides a feasible semiparametric efficient two-step GMM estimate of $\theta_o$ as an illustration of our main result. All the proofs and additional technical derivations are gathered in the Appendix.

## 2 Semiparametric Efficiency Bound

In this section, we derive the semiparametric efficiency bound for $\theta_o$ when the unknown parameters $\alpha_o = (\theta_o, h_o)$ are identified by the sets of moment restrictions (1.1) and (1.2). We assume that the infinite dimensional nuisance parameters $h_o = (h_{1,o}, ..., h_{L,o})$ are identified by the conditional moment restrictions (1.2), and that if $h_o$ were known, the finite dimensional parameter $\theta_o$ is (possibly) over identified by the unconditional moment restrictions (1.1).

Note that the conditioning variables $X_l$ in the conditional moment restrictions (1.2) can be over-lapped or totally different. All previous literatures on efficiency bound that we are aware of, including Chamberlain (1992) and Ai and Chen (2009), only allow for sequential moment restrictions. We make progress over the existing literature in this regard. Our new efficiency bound allows for arbitrary structure in the conditioning variables, and is derived using a new technique based on an orthogonality argument. The orthogonalization has an interesting relationship to adjustment of the influence function for estimation of nonparametric components, which is discussed in Section 3.

Our main result is contained in the following theorem.

**Theorem 1** Suppose that the data are i.i.d., the unknown parameters $\alpha_o = (\theta_o, h_o)$ are identified by model (1.1) and (1.2). If $\text{Var}(\rho(Z, \theta_o, h_o))$ is non-singular, then the information bound
for $\theta_o$ is

$$
\left( \frac{\partial E[g(Z, \alpha_o)]}{\partial \theta} \right) [\Var(\rho(Z, \theta, h_o))]^{-1} \left( \frac{\partial E[g(Z, \alpha_o)]}{\partial \theta} \right)
$$

(2.1)

where

$$
\rho(Z, \theta, h) = g(Z, \theta, h) + \sum_{l=1}^{L} \Delta_l(Z_l, h_l)v_l^*(X_l)
$$

(2.2)

and $v_l^* (\cdot)$ ($l = 1, ..., L$) are defined in (3.3) of Section 3.

**Proof.** Proof, along with discussion, is presented in Section 3. 

The semiparametric efficiency bound of $\theta_o$ depends on some general functions $v_l^*(\cdot)$ ($l = 1, ..., L$) which do not have explicit expressions unless in some special examples. In such scenarios, there exists some function $\frac{\partial g(Z, \theta, h_o)}{\partial h_l}$ such that

$$
\frac{\partial g \left[ Z, \theta, h_{l,o} + \tau \tilde{h}_l, h_{-l,o} \right]}{\partial \tau} \bigg|_{\tau = 0} = \frac{\partial g(Z, \theta, h_o)}{\partial h_l} \tilde{h}_l(X_l)
$$

(2.3)

for any $\tilde{h}_l$ in the local neighborhood of $h_{l,o}$. When the equation (2.3) holds, the general function $v_l^*(X_l)$ has the following closed form expression

$$
v_l^*(X_l) = E \left[ \frac{\partial g(Z, \theta, h_o)}{\partial h_l} \bigg| X_l \right] \left( \frac{\partial m_l(X_l, h_{l,o})}{\partial h_l} \right)^{-1}
$$

(2.4)

where $\frac{\partial m_l(X_l, h_{l,o})}{\partial h_l}$ is defined in (3.6) of Section 3. It is clear that under (2.2) and (2.4), the semiparametric efficiency bound of $\theta_o$ defined in (2.1) has closed form expression. More specific examples of (2.2) can be found in the next section.

### 3 Proof of Theorem 1

We first develop the information bound under some zero derivative restriction. As in (6.23) of Appendix, we define $\Gamma_{1+l}(\theta, h)[v_l]$ as the pathwise derivative of the function $G(\theta, h) \equiv E[g(Z, \theta, h)]$ with respective to $h_l$ in the direction $v_l$ ($l = 1, ..., L$). We will assume that

$$
\Gamma_{1+l}(\theta_o, h_o)[v_l] = \frac{\partial E[g(Z, \theta_o, h_o)]}{\partial h_l}[v_l] = 0 \text{ for any } v_l \in \mathcal{V}_l
$$

(3.1)

where $\mathcal{V}_l$ is the Hilbert space generated by $\mathcal{H}_l - \{h_{l,o}\}$.
Lemma 1 Suppose that the data are i.i.d., the unknown parameters \( \alpha_o = (\theta_o, h_o) \) are identified by (1.1) and (1.2), and the condition (3.1) is satisfied. Then the information bound for \( \theta_o \) is

\[
\left( E \left[ \frac{\partial g(Z, \alpha_o)}{\partial \theta} \right] \right) \left( E \left[ g(Z, \alpha_o) g(Z, \alpha_o)' \right] \right)^{-1} \left( E \left[ \frac{\partial g(Z, \alpha_o)}{\partial h} \right] \right). 
\]

(3.2)

Proof. Proof in Appendix. ■

Lemma 1 shows that when the effects of estimation of \( h_{l,o} \) on the moment conditions \( E[g(Z, \alpha_o)] = 0 \) are ruled out, the semiparametric efficiency bound of \( \theta_o \) only relies on \( E[g(Z, \alpha_o)] = 0 \) with assuming \( h_{l,o} \) to be known.

We now argue that the implication of Lemma 1 is not limited to the case where the zero derivative condition (3.1) is satisfied. It is because we can often transform the model such that the moment condition \( E[g(Z, \alpha_o)] = 0 \) is equivalent to \( E[\rho(Z, \alpha_o)] = 0 \) under (1.2) and moreover

\[
\frac{\partial E[\rho(Z, \theta_o, h_o)]}{\partial h_l}[v_l] = 0 \quad \text{for any } v_l \in \mathcal{V}_l,
\]

where the pathwise derivative of \( \rho(Z, \alpha) \) is defined similarly to (6.23). In many cases, the transformation can be found by inspection, as illustrated in the following examples. For the ease of notation and without loss of generality, we assume that the equation (2.3) is satisfied in these examples.

Example 1 (Nonparametric Regression) The unknown function \( h_o \) is identified by the following conditional mean restriction

\[
E[Y_1 - h_o(X_1) | X_1] = E[U_1 | X_1] = 0. 
\]

(3.3)

For this case, we have

\[
\rho(Z, \theta, h) = g(Z, \theta, h) + E \left[ \frac{\partial g(Z, \theta_o, h_o)}{\partial h} \right] X_1 (Y_1 - h(X_1)).
\]

Example 2 (Nonparametric Quantile Regression) The unknown function \( h_o \) is identified by the following conditional quantile restriction

\[
E[\tau - I\{U_1 \leq 0\} | X_1] = 0, 
\]

(3.4)

where \( U_1 = Y_1 - h_o(X_1) \) and \( \tau \in [0, 1] \). Let \( f_{u_1}(0 | X_1) \) denote the conditional density of \( U_1 \)

---

\footnote{Equation (2.3) is assumed here to provide explicit expression for the general function \( v^*_l(X_l) \) and to make the zero derivative condition (3.1) easier to understand in specific examples. It is remarkable that our main efficiency bound result (i.e., Theorem 1) does not depend on (2.3).}
given $X_1$. For this case, we have

$$
\rho(Z, \theta, h) = g(Z, \theta, h) + E \left[ \frac{\partial g(Z, \theta, h)}{\partial h} \bigg| X_1 \right] \frac{(\tau - I\{Y_1 \leq h(X_1)\})}{f_{u_1}(0 | X_1)}.
$$

**Example 3 (Nonparametric Density Estimation)** Let $f(z_1, h_o(x_1))$ denote the conditional density of $Y_1$ given $X_1$, where $h_o$ is identified by the following conditional moment condition

$$
E \left[ \frac{1}{f(Z_1, h_o(x_1))} f_{o,1} \bigg| X_1 \right] = 0
$$

(3.5)

where $f_{o,1} \equiv \partial f(Z_1, h_o(X_1))/\partial h_1$ satisfies

$$
\left. \frac{\partial f(Z_1, h_o + \tau(h - h_o))}{\partial \tau} \right|_{\tau=0} = f_{o,1}(h - h_o).
$$

Let $f_o \equiv f(Z_1, h_o)$. For this case, we have

$$
\rho(Z, \theta, h) = g(Z, \theta, h) + E \left[ \frac{\partial g(Z, \theta, h)}{\partial h} \bigg| X_1 \right] \frac{f_{o,1}/f_o}{E \left[ f_{o,1}^2/f_o^2 \bigg| X_1 \right]}.
$$

**Example 4 (Different Unknown Functions Estimated in First-step)** Suppose that $h_o = (h_{1,o}, h_{2,o}, h_{3,o})$ and the moment condition (1.1) has the following form

$$
E [g(Z, \theta, h_{1,o}, h_{2,o}, h_{3,o})] = 0
$$

where $h_{1,o}$, $h_{2,o}$ and $h_{3,o}$ are identified by the conditional moment conditions (3.3), (3.4) and (3.5) respectively. For this case, we have

$$
\rho(Z, \theta, h) = g(Z, \theta, h) + E \left[ \frac{\partial g(Z, \theta, h_o)}{\partial h_1} \bigg| X_1 \right] (Y_1 - h_1(X_1))
$$

$$
+ E \left[ \frac{\partial g(Z, \theta, h_o)}{\partial h_2} \bigg| X_2 \right] \frac{(\tau - I\{Y_2 \leq h_2(X_2)\})}{f_{u_2}(0 | X_2)}
$$

$$
+ E \left[ \frac{\partial g(Z, \theta, h_o)}{\partial h_3} \bigg| X_3 \right] \frac{f_{o,1}(Z_3)/f_o(Z_3)}{E \left[ f_{o,1}^2(Z_3)/f_o^2(Z_3) \bigg| X_3 \right]}
$$

where $h = (h_1, h_2, h_3)$.

In order to gain a similar interpretation for the model characterized by (1.1) and (1.2), we present a systematic method of transforming the model (1.2) and (1.1) such that the zero derivative restriction (3.1) is always satisfied. For this purpose, we define $\Sigma_t(X_i) = \text{Var}(\Delta_t(Z_l, h_{t,o}(X_i)) | X_i)$ and $m_t(X_i, h_i) = E[\Delta_t(Z_l, h_i(X_i)) | X_i]$. Let $\partial m_t(X_i, h_{t,o})/\partial h_l$ be the
function which satisfies

\[
\frac{\partial m_t(X_t, h_{t,o} + \tau h_t)}{\partial \tau} \bigg|_{\tau = 0} = \frac{\partial m_t(X_t, h_{t,o})}{\partial h_t} h_t(X_t) \tag{3.6}
\]

for any \( h_t \) in the local neighborhood of \( h_{t,o} \). For any \( v_t, \tilde{v}_t \in \mathcal{V}_t \), we define the following inner product

\[
\langle v_t, \tilde{v}_t \rangle_t = E \left[ \left( \frac{\partial m_t(X_t, h_{t,o})}{\partial h_t} v_t \right) \Sigma_{t,o}^{-1}(X_t) \left( \frac{\partial m_t(X_t, h_{t,o})}{\partial h_t} \tilde{v}_t \right) \right].
\]

Applying the Riesz representation theorem, we can find the \( u_t^* \) such that

\[
\frac{\partial E [g(Z, \theta, h_{t,o})]}{\partial h_t} [v_t] = \langle v_t, u_t^* \rangle_t = E \left[ \left( \frac{\partial m_t(X_t, h_{t,o})}{\partial h_t} v_t \right) \Sigma_{t,o}^{-1}(X_t) \left( \frac{\partial m_t(X_t, h_{t,o})}{\partial h_t} u_t^* \right) \right]. \tag{3.7}
\]

It follows that if we let

\[
\rho(Z, \theta, h) = g(Z, \theta, h) + \sum_{l=1}^{L} \Delta_l(Z_l, h_l) v_t^*(X_l) \tag{3.8}
\]

where

\[
v_t^*(X_t) = -\Sigma_{t,o}^{-1}(X_t) \left( \frac{\partial m_t(X_t, h_{t,o})}{\partial h_t} \right) u_t^*(X_t), \tag{3.9}
\]

then we have

\[
E \left[ \frac{\partial \rho(Z, \theta, h_{t,o})}{\partial \theta'} \right] = E \left[ \frac{\partial g(Z, \theta, h_{t,o})}{\partial \theta'} \right] \quad \text{and} \quad \frac{\partial E [\rho(Z, \theta, h_{t,o})]}{\partial h_t} [v_t] = 0 \tag{3.10}
\]

for any \( v_t \in \mathcal{V}_t \). Moreover under (1.2), the original moment condition \( E [g(Z, \alpha_o)] = 0 \) and the transformed moment condition \( E [\rho(Z, \alpha_o)] = 0 \) are equivalent, i.e.

\[
E [\rho(Z, \alpha_o)] = 0 \Leftrightarrow E [g(Z, \alpha_o)] = 0. \tag{3.11}
\]

From (3.10) and (3.11), we can use Lemma 1 to deduce that Theorem 1 holds.

### 4 Implication and Discussion of Theorem 1

As discussed in Ai and Chen (2007), when \( h_o \) is estimated by sieve minimum distance (SMD) estimator \( \tilde{h} \), there is

\[
E [\rho(Z, \alpha_o) \rho(Z, \alpha_o)'] = \text{Avar} \left( n^{-1/2} \sum_{i=1}^{n} g(Z_i, \theta_o, \tilde{h}) \right), \tag{4.1}
\]
which together with Theorem 1 implies that the information bound for $\theta_o$ in model (1.1) and (1.2) is equal to
\[
\left( E \left[ \frac{\partial g(Z, \theta_o, h_o)}{\partial \theta'} \right] \right)' \left( \text{Avar} \left( n^{-1/2} \sum_{i=1}^{n} g(Z_i, \theta_o, \tilde{h}) \right) \right)^{-1} \left( E \left[ \frac{\partial g(Z, \theta_o, h_o)}{\partial \theta'} \right] \right). \tag{4.2}
\]

On the other hand, the asymptotic variance of $n^{-1/2} \sum_{i=1}^{n} g(Z_i, \theta_o, \tilde{h})$ is known to be invariant to the choice of any nonparametric estimator $\tilde{h}$ of $h_o$, which follows from Newey’s (1994, Theorem 1) observation that the asymptotic variance of the semiparametric estimators is independent of the type of estimator. Such invariance result implies that the semiparametric efficiency bound of $\theta_o$ in model (1.1) and (1.2) can be equivalently written as
\[
\left( E \left[ \frac{\partial g(Z, \theta_o, h_o)}{\partial \theta'} \right] \right)' \left( \text{Avar} \left( n^{-1/2} \sum_{i=1}^{n} g(Z_i, \theta_o, \tilde{h}_n) \right) \right)^{-1} \left( E \left[ \frac{\partial g(Z, \theta_o, h_o)}{\partial \theta'} \right] \right) \tag{4.3}
\]
where $\tilde{h}_n$ denotes any consistent nonparametric estimator of $h_o$. It is clear that (4.1) provides one example of illustrating the general form (4.3) when $h_o$ is estimated by SMD estimator. Another example is provided in the next section where $h_o$ is estimated by sieve M estimator.

The general expression of the information bound of $\theta_o$ in (4.3) indicates that under suitable regularity conditions, the GMM estimator $\hat{\theta}_n$ that solves
\[
\min_{\theta \in \Theta} \left[ n^{-1/2} \sum_{i=1}^{n} g(Z_i, \theta, \tilde{h}_n) \right]' W_n \left[ n^{-1/2} \sum_{i=1}^{n} g(Z_i, \theta, \tilde{h}_n) \right], \tag{4.4}
\]
is efficient as long as the weighting matrix $W_n$ satisfies
\[
W_n^{-1} \rightarrow_p \text{Avar} \left( n^{-1/2} \sum_{i=1}^{n} g(Z_i, \theta_o, \tilde{h}_n) \right) \tag{4.5}
\]
where $\tilde{h}_n$ again denotes any consistent nonparametric estimator of $h_o$. In most of the empirical applications, it is a natural exercise to choose the weighting matrix $W_n$ which satisfies (4.5) such that the two-step GMM estimate $\hat{\theta}_n$ is expected to be limited efficient, i.e. efficient among all feasible two-step estimates. However, Theorem 1 actually indicates that the result two-step GMM estimate $\hat{\theta}_n$ exhausts all the information in (1.1) and (1.2) and thus is fully efficient.

From the above discussion, we see that one only need to take care of the first-step estimation effect in the weighting matrix $W_n$ to ensure that two-step GMM estimate is asymptotically efficient. Such adjustment is automatically preformed when $W_n$ is constructed to ensure the
two-step GMM estimate achieves the limited efficiency. Such simple, but surprising result is
due to the fact that the nuisance functions estimated in the first-step are exactly identified.
This is in sharp contrast with the scenario where the parameters estimated in the first-step are
over identified. In that case, optimally weighted two-step GMM estimates are not fully efficient
in general, unless some orthogonalization/filtering operations are performed on the second-step
moment functions, as illustrated in Hayashi and Sims (1983), Chamberlain (1992) and Ai and
Chen (2007).

5 Semiparametric Two-step GMM Estimation

In this section, we discuss implementation of a feasible efficient estimator of \( \theta_0 \) for the model
categorized by (1.1) and (1.2), where the unknown functions \( h_o \) are estimated by sieve M
estimation. We start by introducing some notations. Let \( G_n(\theta, h) = \frac{1}{n} \sum_{i=1}^{n} g(Z_i, \theta, h) \) and
\( G(\theta, h) = E[g(Z, \theta, h)] \). For any \((\theta, h) \in \Theta \times \mathcal{H}\), we denote the ordinary derivative of \( G(\theta, h) \)
with respect to \( \theta \) as \( \Gamma_1(\theta, h) \), which satisfies

\[
\Gamma_1(\theta, h)(\hat{\theta} - \theta) = \frac{\partial}{\partial \tau} \left[ G(\theta + \tau (\hat{\theta} - \theta), h(\cdot)) \right]_{\tau=0}
\]

for all \( \hat{\theta} \in \Theta \). Let \( \hat{h}_{l,n} \) \((l = 1, \ldots, L)\) be the first-step sieve M estimates that solve

\[
\frac{1}{n} \sum_{i=1}^{n} \varphi_l(Z_{l,i}, \hat{h}_{l,n}) \geq \sup_{h_l \in \mathcal{H}_{l,n}} \frac{1}{n} \sum_{i=1}^{n} \varphi_l(Z_{l,i}, h_l) - O_p(\varepsilon_n^2) \tag{5.1}
\]

where \( \varphi_l(\cdot, \cdot) \) is some measurable criterion function, \( \mathcal{H}_{l,n} \) is a finite dimensional sieve space
approximating the parameter spaces \( \mathcal{H}_l \) \((l = 1, \ldots, L)\), the term \( O_p(\varepsilon_n^2) \) denotes the maximization
error and \( \varepsilon_n^2 = o(n^{-1}) \) is the rate that this error converges to zero.

For any \( \alpha = (\theta, h) \in \Theta \times \mathcal{H} \) and \( v_l \in \otimes_{s=1}^{k} \mathcal{V}_s \), denote

\[
V_l(\alpha) [v] = g(Z_i, \theta, h) + \sum_{l \leq L} \Delta_l(Z_{l,i}, h_l) v_l, \tag{5.2}
\]

where \( v_l = (v_{l,1}, \ldots, v_{l,k})' \). The asymptotic variance of \( \hat{\theta}_n \) has an explicit form when the unknown
functions \( h_{l,o} \) \((l = 1, \ldots, L)\) are estimated by sieve M estimation, as illustrated in the following
proposition.

**Proposition 1** Under some regularity conditions, the GMM estimator defined in (4.4) satisfies
\[ \sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \to_d \mathcal{N} (0, V_\theta), \] where

\[ V_\theta = (\Gamma'_1 W T_1)^{-1} (\Gamma'_1 W V_N W T_1) (\Gamma'_1 W T_1)^{-1}, \] (5.3)

\[ \Gamma_1 = \Gamma_1 (\theta_0, h_0), \] \( W \) is the probability limit of \( W_n \) and

\[ V_N = \lim_{n \to \infty} E \left[ n^{-1} \sum_{i=1}^{n} \nu_i (\alpha_0) \nu_i' (\alpha_0) \right], \] (5.4)

where \( \nu_i \in \otimes_{s=1}^{k} V_s \) is defined in (6.25) of Appendix 6.2.

**Proof.** The claimed result follows directly from Theorem 2 of Chen, Linton and van Keilegom (2003) and Theorem 3.1 of Chen and Liao (2012).

From Theorem 2 of Chen, Linton and van Keilegom (2003), we know that

\[ V_N = \text{Avar} \left( n^{-1/2} \sum_{i=1}^{n} g(Z_i, \theta_0, \hat{h}_n) \right). \] (5.5)

Using the identity in (6.27), i.e. \( \nu_i = \nu_i^* \) a.e. \( (l = 1, ..., L) \) and the expression (5.2), we get

\[ V_N = \text{Avar} \left( n^{-1/2} \sum_{i=1}^{n} \rho (Z_i, \alpha_0) \right), \] (5.6)

where \( \rho (Z_i, \alpha_0) \) is defined in (2.2). When the data are i.i.d., from (5.5) and (5.6),

\[ \text{Avar} \left( n^{-1/2} \sum_{i=1}^{n} g(Z_i, \theta_0, \hat{h}_n) \right) = E \left[ \rho (Z, \alpha_0) \rho (Z, \alpha_0)' \right] = V_N. \] (5.7)

The first equality in (5.7) was predicted by Theorem 1 of Newey (1994) that the asymptotic variance of \( n^{-1/2} \sum_{i=1}^{n} g(Z_i, \theta_0, \hat{h}_n) \) is invariant to the choice of nonparametric estimator of \( h_0 \). Moreover, the second equality in (5.7) together with Theorem 1 implies that if \( W = V_N^{-1} \) in (5.3), then we immediately get the following efficiency result.

**Corollary 1** The two-step estimator \( \hat{\theta}_n \) defined in (4.4) is semiparametrically efficient for the model characterized by moment conditions (1.1) and (1.2).

**Remark 1** The asymptotic variance of the efficient estimator \( \hat{\theta}_n \) can be estimated by

\[ \hat{V}_{\theta,n} = \left( \hat{\Gamma}_{1,n}^{-1} \hat{V}_{N,n} \hat{\Gamma}_{1,n}^{-1} \right)^{-1} \] with
\[ \hat{V}_{N,n} = n^{-1} \sum_{i=1}^{n} \left( V_i(\hat{\theta}_n, \hat{h}_n) \right) \left( V_i(\hat{\theta}_n, \hat{h}_n) \right)' \text{ and } \hat{V}_{1,n} = n^{-1} \sum_{i=1}^{n} \frac{\partial g(Z_i, \hat{\theta}_n, \hat{h}_n)}{\partial \theta} \]

where \( \hat{e}_n^* \) is the estimator of \( e^* \) and is defined in (6.32) of Appendix 6.2.

6 Appendix

6.1 Proof of the Main Results in Section 3

Proof of Lemma 1. For the ease of notation and without loss of generality, we assume in this proof that \( L = 2 \) and that \( h_l \) are scalar-valued. Let \( f_o(z) \) to be the true density of \( Z \), \( f_o(z_j|x_j) \) be the true conditional density of \( Z_j \) given \( X_j \) (\( j = 1, 2 \)) and \( \mathcal{F} \) be a class of function on \( (Z, || \cdot ||) \) with \( f_0 \in \mathcal{F} \). Define a class of density functions \( \mathcal{F}_\alpha \) that satisfy the conditional moment conditions:

\[
\mathcal{F}_\alpha = \left\{ f \in \mathcal{F} : \int \Delta_1(z_1, h_1(x_1)) f(z|x_1) d\mu(z) = 0, \int \Delta_2(z_2, h_2(x_2)) f(z|x_2) d\mu(z) = 0, \int g(z, \theta, h_1, h_2) f(z) d\mu(z) = 0 \right\}. \tag{6.1}
\]

Let \( \mathcal{G} \) denote a class of real valued measurable function of \( Z \) such that

\[
\mathcal{F}_\alpha = \left\{ f(z|\theta, h_1, h_2, k) : k \in \mathcal{G} \right\} \tag{6.2}
\]

for any \( \alpha = (\theta, h_1, h_2) \in \Theta \times \mathcal{H}_1 \times \mathcal{H}_2 \). Let \( \mathcal{V}_\theta \times \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{V}_k \) denote the completion of \( \Theta \times \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{G} - \{ (\theta, h_1, h_2, k_o) \} \) where \( k_o \) satisfies

\[
f(z|\theta_o, h_{1, o}, h_{2, o}, k_o) = f_o(z).
\]

We will consider the parametric family \( f(z|\theta_o + \tau_\theta, h_{1, o} + \tau_1 v_1, h_{2, o} + \tau_2 v_2, k_o + \tau_k v_k) \). Inspection of the moment restrictions \( E[\Delta_l(Z_l, h_{l, o}(X_l))|X_l] = 0 \) and \( E[g(Z, \alpha_o)] = 0 \) indicates that the nonparametric tangent space \( \mathcal{T} \) is given by the completion of \( s_{h_1}(z) [v_1] + s_{h_2}(z) [v_2] + \)
\( s_k (z) [v_k] \), where \( s \)'s satisfy

\[
E[\Delta_1 (Z_1, h_{1, o}) s_{\theta} (Z) | X_1] = 0 \quad (6.3)
\]

\[
\frac{\partial m_1 (X_1, h_{1, o})}{\partial h_1} v_1 (X_1) + E[\Delta_1 (Z_1, h_{1, o}) s_{h_1} (Z) [v_1] | X_1] = 0 \quad (6.4)
\]

\[
E[\Delta_1 (Z_1, h_{1, o}) s_{h_2} (Z) [v_2] | X_1] = 0 \quad (6.5)
\]

\[
E[\Delta_1 (Z_1, h_{1, o}) s_k (Z) [v_k] | X_1] = 0 \quad (6.6)
\]

\[
E[\Delta_2 (Z_2, h_{2, o}) s_{\theta} (Z) | X_2] = 0 \quad (6.7)
\]

\[
E[\Delta_2 (Z_2, h_{2, o}) s_{h_1} (Z) [v_1] | X_2] = 0 \quad (6.8)
\]

\[
\frac{\partial m_2 (X_2, h_{2, o})}{\partial h_2} v_2 (X_2) + E[\Delta_2 (Z_2, h_{2, o}) s_{h_2} (Z) [v_2] | X_2] = 0 \quad (6.9)
\]

\[
E[\Delta_2 (Z_2, h_{2, o}) s_k (Z) [v_k] | X_2] = 0 \quad (6.10)
\]

and

\[
\frac{\partial E [g (Z, \theta, h_{1, o}, h_{2, o})]}{\partial \theta'} + E [g (Z, \theta, h_{1, o}, h_{2, o}) s_{\theta} (Z)'] = 0 \quad (6.11)
\]

\[
E [g (Z, \theta, h_{1, o}, h_{2, o}) s_{h_1} (Z) [v_1]] = 0 \quad (6.12)
\]

\[
E [g (Z, \theta, h_{1, o}, h_{2, o}) s_{h_2} (Z) [v_2]] = 0 \quad (6.13)
\]

\[
E [g (Z, \theta, h_{1, o}, h_{2, o}) s_k (Z) [v_k]] = 0 \quad (6.14)
\]

for any \( (v_{h_1}, v_{h_2}, v_k) \in \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{V}_k \). Note that \([3.10]\) is used in \([6.12]\) and \([6.13]\).

The residual of the projection of \( s_{\theta} \) on \( T \), \( s_{\theta} (Z) - \text{proj}[s_{\theta} (Z) | T] \) will give the semiparametric score \( S_{\theta}^* (Z) \) and the semiparametric efficiency bound of \( \theta_0 \) will be \( E[S_{\theta}^* (Z) S_{\theta}^* (Z)'] \). We show that the residual of the projection of \( s_{\theta} \) on \( T \) is equal to

\[
S_{\theta}^* (Z) = - \left( \frac{\partial E [g (Z)]}{\partial \theta'} \right)' \left\{ E [g (Z) g (Z)'] \right\}^{-1} g (Z) \quad (6.15)
\]

where \( g (Z) = g (Z, \theta, h_{1, o}, h_{2, o}) \).

We first solve for \( \Lambda_1^* (X_1) \) and \( \Lambda_2^* (X_2) \) for the equalities

\[
0 = E[\Delta_1 (Z_1, h_{1, o}) \{ s_{\theta} (Z) - S_{\theta}^* (Z) - s_{h_1} (Z) [\Lambda_1^*] - s_{h_2} (Z) [\Lambda_2^*] \} | X_1] \quad (6.16)
\]

and

\[
0 = E[\Delta_2 (Z_2, h_{2, o}) \{ s_{\theta} (Z) - S_{\theta}^* (Z) - s_{h_1} (Z) [\Lambda_1^*] - s_{h_2} (Z) [\Lambda_2^*] \} | X_2] \quad (6.17)
\]
Letting $v_{h_1} = \Lambda_1^*(X_1)$ in (6.4) and $v_{h_2} = \Lambda_2^*(X_2)$ in (6.5), we get
\[
\frac{\partial m_1(X_1, h_{1,1})}{\partial h_1} \Lambda_1^*(X_1) + E \left[ \Delta_1(Z_{1,1}, h_{1,1}) s_{h_1}(Z) [\Lambda_1^*]_1 | X_1 \right] = 0 \quad (6.18)
\]
and
\[
E \left[ \Delta_1(Z_{1,1}, h_{1,1}) s_{h_2}(Z) [\Lambda_2^*]_1 | X_1 \right] = 0. \quad (6.19)
\]
Using (6.3) along with (6.18) and (6.16), we write
\[
0 = \frac{\partial E\left[ g(Z) \right]}{\partial \theta} \{ E[g(Z) g(Z)] \}^{-1} E[g(Z) \Delta_1(Z_{1,1}, h_{1,1}) | X_1] + \frac{\partial m_1(X_1, h_{1,1})}{\partial h_1} \Lambda_1^*(X_1)
\]
which can be solved for $\Lambda_1^*(X_1)$ as long as $\partial m_1(X_1, h_{1,1})/\partial h_1 \neq 0$ almost surely. Similarly, we can solve for $\Lambda_2^*(X_2)$ as long as $\partial m_2(X_2, h_{2,1})/\partial h_2 \neq 0$ almost surely.

Now let
\[
W = s_{\theta}(Z) - S_{\theta}^*(Z) - s_{h_1}(Z) [\Lambda_1^*] - s_{h_2}(Z) [\Lambda_2^*]
\]
Using (6.11)-(6.14), we obtain
\[
E[W g(Z)] = E[s_{\theta}(Z) g(Z)] - E[S_{\theta}^*(Z) g(Z)]
\]
\[
= - \left( \frac{\partial E[g(Z)]}{\partial \theta} \right)' \left( \frac{\partial E[g(Z)]}{\partial \theta} \right)' \{ E[g(Z) g(Z)] \}^{-1} \{ E[g(Z) g(Z)] \}
\]
\[
= 0. \quad (6.20)
\]
Inspection of equations (6.16), (6.17), and (6.20) reveals that $W$ satisfies the property of $s_k(z) [v_k]$ listed above and thus
\[
s_{h_1}(Z) [\Lambda_1^*] + s_{h_2}(Z) [\Lambda_2^*] + W \in \mathcal{T}. \quad (6.21)
\]
Because $S_{\theta}^*(Z)$ is proportional to $g(Z)$, we can deduce from (6.12)-(6.14) that $S_{\theta}^*(Z) \perp \mathcal{T}$. Along with (6.21), this implies that $S_{\theta}^*(Z)$ is the residual of the projection of $s_{\theta}$ on $\mathcal{T}$. Thus the semiparametric efficiency bound of $\theta$ is
\[
E[S_{\theta}^*(Z)] = \left( \frac{\partial E[g(Z)]}{\partial \theta} \right)' \left( \frac{\partial E[g(Z)]}{\partial \theta} \right)' \{ E[g(Z) g(Z)] \}^{-1} \left( \frac{\partial E[g(Z)]}{\partial \theta} \right). \quad (6.22)
\]
6.2 Riesz Representor of Smooth Functionals

For any $(\theta, h) \in \Theta \times \mathcal{H}$, we say that $G(\theta, h)$ is pathwise differentiable at $h_l \in \mathcal{H}_l$ in the direction $[\tilde{h}_l - h_l]$, if $\{h_l + \tau(\tilde{h}_l - h_l) : \tau \in [0, 1]\} \subset \mathcal{H}_l$ and

$$
\Gamma_{1+l}(\theta, h)[\tilde{h}_l - h_l] \equiv \left. \frac{\partial G[\theta, h_1(\cdot) + \tau(\tilde{h}_l(\cdot) - h_l(\cdot)), h_{-l}(\cdot)]}{\partial \tau} \right|_{\tau=0} \tag{6.23}
$$

exits, where $h_{-l} = (h_1, ..., h_{l-1}, h_{l+1}, ..., h_L)$. Recall that $\mathcal{V}_l$ is the Hilbert space generated by $\mathcal{H}_l - \{h_{l,o}\}$. We define an inner products on $\mathcal{V}_l$:

$$
\langle v_1, v_2 \rangle_{m,l} = E \left\{ - \frac{\partial m_l(X_l, h_{l,o})}{\partial h_l} v_1(X_l) v_2(X_l) \right\} \text{ for any } v_1, v_2 \in \mathcal{V}_l \tag{6.24}
$$

which induces a semi-norm $\|\cdot\|_{m,l}$ on $\mathcal{V}_l$ ($l = 1, ..., L$).

Denote $\Gamma_{1+l,j}(\alpha_o)[\cdot] = \Gamma_{1+l,j}(\theta_o, h_o)[\cdot]$ for ($l = 1, ..., L$ and $j = 1, ..., k$). Suppose $\Gamma_{1+l,j}(\alpha_o)[\cdot]$ is a linear and bounded functional on $\mathcal{V}_l$, then by Riesz representation Theorem, there exists some $e_{l,j}^* \in \mathcal{V}_l$ such that

$$
\Gamma_{1+l,j}(\alpha_o)[v] = \langle v, e_{l,j}^* \rangle_{m,l} \text{ for all } v \in \mathcal{V}_l, \text{ and} \tag{6.25}
$$

$$
\Gamma_{1+l,j}(\alpha_o)[e_{l,j}^*] = \sup_{v \in \mathcal{V}_l, v \neq 0} \frac{|\Gamma_{1+l,j}(\alpha_o)[v]|^2}{\|v\|_{m,l}^2} < \infty. \tag{6.26}
$$

The function $e_{l,j}^*$ is called the Riesz representor, which is well defined and unique almost surely. As a result, we get $e_l^* = (e_{l,1}^*, ..., e_{l,k}^*)'$ for $l = 1, ..., L$.

Using the Riesz representations in (3.7) and (6.25), we have

$$
\langle e_{l,j}^* - v_{l,j}^*, u_{l,j}^* \rangle_l = \Gamma_{1+l,j}(\alpha_o)[e_{l,j}^* - v_{l,j}^*] = \langle e_{l,j}^* - v_{l,j}^*, e_{l,j}^* \rangle_{m,l}. \tag{6.27}
$$

From the expressions in (3.7) and (6.24), we can deduce that

$$
0 = E \left[ \frac{\partial m_l(X_l, h_{l,o})}{\partial h_l} (e_{l,j}^* - v_{l,j}^*) e_{l,j}^* \right] - E \left\{ \frac{\partial m_l(X_l, h_{l,o})}{\partial h_l} (e_{l,j}^* - v_{l,j}^*) v_{l,j}^* \right\}
$$

$$
= E \left[ \frac{\partial m_l(X_l, h_{l,o})}{\partial h_l} (e_{l,j}^* - v_{l,j}^*)^2 \right] = \|e_{l,j}^* - v_{l,j}^*\|_{m,l}^2
$$

which implies that

$$
e_{l,j}^* = v_{l,j}^* \text{ a.e. for any } l = 1, ..., L \text{ and } j = 1, ..., k. \tag{6.27}
$$
In some examples, it may be difficult to compute the Riesz representer $e_{i,j}^*(j = 1, ..., k)$ on the infinite dimensional Hilbert space $V_i$. One solution is to define a Riesz representer $e_{i,j,n}^*$ on the finite dimensional Hilbert space $V_{i,n}$ generated by $H_{i,n} = \{h_{i,n}\}$ where $h_{i,n} \in H_{i,n}$ and then use $e_{i,j,n}^*$ to approximate $e_{i,j}^*$. In the linear sieve M-estimation, $e_{i,j,n}^*$ is well defined and always has the explicit expression, which not only is useful for constructing estimate for $e_i^*$ in finite samples but also makes it to derive the asymptotic distribution for plug-in estimate of irregular functionals which are not root-n estimable (see, e.g., Chen and Liao (2012)).

Formally, as $\Gamma_{1+t,j}(\alpha_o)[\cdot]$ is a linear functional, by Riesz representation Theorem, there exists some $e_{i,j,n}^* \in V_{i,n}$ such that

$$
\Gamma_{1+t,j}(\alpha_o)[v] = \langle v, e_{i,j,n}^* \rangle_{m,l} \text{ for all } v \in V_{i,n}, \text{ and }
$$

$$
\Gamma_{1+t,j}(\alpha_o)[e_{i,j,n}^*] = \sup_{v \in V_{i,n}, v \neq 0} \frac{|\Gamma_{1+t,j}(\alpha_o)[v]|^2}{\|v\|^2_{m,l}} < \infty.
$$

Let $P_K(\cdot) = [p_1(\cdot), ..., p_K(\cdot)]'$ denote the basis functions in $H_{i,n}$. Using the fact that $v(\cdot) = P_K(\cdot)'\beta_v, K$ for any $v \in V_{i,n}$, we deduce that

$$
\Gamma_{1+t,j}(\alpha_o)[e_{i,j,n}^*] = \Gamma_{1+t,j}(\alpha_o)[P_K]' \left\{ E \left[ -\frac{\partial m_l(X_l, h_{l,o})}{\partial h_l} P_K(X_l) P_K(X_l)' \right] \right\}^{-1} \Gamma_{1+t,j}(\alpha_o)[P_K]
$$

where

$$
\Gamma_{1+t,j}(\alpha_o)[P_K] = (\Gamma_{1+t,j}(\alpha_o)[p_1(X_l)], ..., \Gamma_{1+t,j}(\alpha_o)[p_K(X_l)])'.
$$

From the expression in (6.30), we obtain

$$
e_{i,j,n}^*(\cdot) = P_K(\cdot)' \left\{ E \left[ -\frac{\partial m_l(X_l, h_{l,o})}{\partial h_l} P_K(X_l) P_K(X_l)' \right] \right\}^{-1} \Gamma_{1+t,j}(\alpha_o)[P_K]
$$

and $e_{i,n}^* = (e_{i,n,1}^*, ..., e_{i,n,L}^*)$ for $l = 1, ..., L$. By the definition of Riesz representer $e_{i,j,n}^*(\cdot)$, we can define an empirical Riesz representer $\hat{e}_{i,j,n}^*(\cdot)$ in the following

$$
\hat{e}_{i,j,n}^*(\cdot) = P_K(\cdot)' \left\{ -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \Delta_l(Z_{i,l}, \hat{h}_{l,n})}{\partial h_l} P_K(X_{i,l}) P_K(X_{i,l})' \right\}^{-1} \hat{\Gamma}_{1+t,j}(\hat{\alpha}_n)[P_K],
$$

where $\partial \Delta_l(Z_l, h_l)/\partial h_l$ satisfies

$$
E \left[ \frac{\partial \Delta_l(Z_l, h_l)}{\partial h_l} | X_l \right] = \partial m_l(X_l, h_l)/\partial h_l \text{ for any } h_l \text{ in the local neighborhood of } h_{o,l}
$$

and

$$
\hat{\Gamma}_{1+t,j}(\hat{\alpha}_n)[P_K()]' \equiv \left( \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g_j(Z_l, \hat{\alpha}_n)}{\partial h_l} [p_1()], ..., \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g_j(Z_l, \hat{\alpha}_n)}{\partial h_l} [p_K()] \right).
$$
The vector $\widehat{e}_{t,n}' = [\widehat{e}_{t,1,n} (\cdot), \ldots, \widehat{e}_{t,K,n} (\cdot)]'$ can be defined accordingly using (6.32).

References


