ASYMPTOTIC EFFICIENCY OF SEMIPARAMETRIC TWO-STEP GMM

By

Xiaohong Chen, Jinyong Hahn and Zhipeng Liao

October 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1880

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Asymptotic Efficiency of Semiparametric Two-step GMM

Xiaohong Chen* Jinyong Hahn† Zhipeng Liao‡

Yale UCLA UCLA

First Version: April 2011; This version: June 2012

Abstract

In this note, we characterize the semiparametric efficiency bound for a class of semiparametric models in which the unknown nuisance functions are identified via nonparametric conditional moment restrictions with possibly non-nested or over-lapping conditioning sets, and the finite dimensional parameters are potentially 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 nonparametric procedures in the first step. Regardless of whether the efficiency bound has a closed form expression or not, we provide easy-to-compute sieve based optimal weight matrices that lead to asymptotically efficient two-step GMM estimators.

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 θ_o that is (possibly over-) identified by the unconditional

*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

moment restrictions

$$E\left[g\left(Z;\theta_o,h_{1,o}(\cdot),...,h_{L,o}(\cdot)\right)\right]=0, \tag{1.1}$$

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

$$E\left[\Delta_\ell(Z,h_{\ell,o}\left(X_\ell\right))\middle|\,X_\ell\right]=0 \text{ almost surely } X_\ell,\quad \ell=1,...,L, \tag{1.2}$$

where the conditioning variables X_ℓ, $\ell=1,...,L$, could be nested, overlapping or non-nested, and the unknown functions $h_{\ell,o}\left(\cdot\right)$, $\ell=1,...,L$, are distinct from each other. The moment functions $g\left(Z;\theta,h(\cdot)\right)$ and $\Delta_\ell(Z,h_\ell\left(X_\ell\right))$, $\ell=1,...,L$, could be pointwise non-smooth with respect to the parameters θ and $h=(h_1(\cdot),...,h_L(\cdot))$. This class of models has been widely used in applied work in economics, allowing for semiparametric quantile treatment effects, endogenous default, censoring, sample selection, data combination and many more.

Given the conditional moment restrictions (1.2), we can estimate $h_{\ell,o}$ by any nonparametric estimator $\widehat{h}_\ell$ for $\ell=1,...,L$, and then estimate θ_o in (1.1) by setting the sample analog $n^{-1}\sum_{i=1}^n g(Z_i;\theta,\widehat{h})$ of $E\left[g\left(Z;\theta,h_o\right)\right]$ as close to zero as possible, an intuitive strategy suggested in Andrews (1994), Newey (1994), Pakes and Olley (1995), Chen, Linton and van Keilegom (2003) and many 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.

We pose a natural question whether the "limited information" estimation strategy in fact exhausts all the information in model (1.1) and (1.2). For this purpose, we derive the semiparametric efficiency bound for θ_o when the unknown parameters (θ_o,h_o) are identified by the model (1.1) and (1.2). We allow the conditioning variables X_ℓ, $\ell=1,...,L$, to be different from each other or to have arbitrary overlaps. To the best of our knowledge, our paper is the first to derive efficiency bound for θ_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 θ_o is estimated in the second step by GMM using the unconditional moment (1.1) with an optimal weight matrix that reflect the noise in estimating the nuisance functions h_o, the resulting semiparametric two-step GMM estimators achieve the semiparametric efficiency bound for θ_o. To the best of our knowledge, there is no published work addressing whether or not the semiparametric two-step GMM estimation is efficient for θ_o satisfying the over-identifying moment restriction (1.1).

The semiparametric efficiency bound for θ_o may not have a closed form expression in general, and hence it may be difficult to compute a feasible optimal weight matrix based on any nonparamertic first step. When the nuisance functions are estimated via a simple sieve M procedure in the first step, we provide easy-to-compute optimal weight matrices that lead to asymptotically efficient two-step GMM estimators.

The rest of the note is organized as follows. Section 2 establishes the semiparametric efficiency bound for θ_o, and discusses some special cases. Readers who would like to avoid technical details can jump directly to Section 3, where the main result of Section 2 is rephrased in a more intuitive way and some of its practical implications are discussed. Section 4 provides computationally attractive sieve semiparametric efficient two-step GMM estimates of θ_o. Additional proofs and technical derivations are gathered in the Appendix.

2 Semiparametric Efficiency Bound

In this section, we derive the semiparametric efficiency bound for θ_o when the unknown parameters $\alpha_o = (\theta_o, h_o) \in \Theta \times \mathcal{H}$ are identified by the sets of moment restrictions (1.1) and (1.2). To be precise, let $F_o(\cdot)$ be the unknown true probability distribution of Z. For $\ell = 1, ..., L$ with a fixed finite L, let $F_{\ell,o}(\cdot|x_\ell)$ be the unknown true conditional probability distribution of Y_ℓ given $X_\ell = x_\ell$, where Y_ℓ does not include X_ℓ but could contain some X_j, $j \neq \ell$, that does not overlap with X_ℓ. In this paper, model (1.1) - (1.2) is a simplified presentation for the model (2.1) - (2.2)

$$\int g\left(z, \theta_o, h_{1,o}(\cdot), ..., h_{L,o}(\cdot)\right) dF_o(z) = 0, \tag{2.1}$$

$$\int \Delta_\ell\left(z_{-\ell}, x_\ell, h_{\ell,o}(x_\ell)\right) dF_{\ell,o}(z_{-\ell}|x_\ell) = 0 \text{ for almost all } x_\ell,\ \ell = 1, ..., L. \tag{2.2}$$

where $Z_{-\ell}$ denotes the components of Z not in the conditioning variable X_ℓ. We note that although the unknown functions $h_{\ell,o}(\cdot)$, $\ell = 1, ..., L$ enter the conditional moment restrictions (1.2) (i.e., (2.2)) through $h_{\ell,o}(X_\ell)$ only, they could enter the unconditional moment restrictions (1.1) (i.e., (2.1)) in a very flexible way. We assume that the infinite dimensional nuisance functions $h_o(\cdot) = (h_{1,o}(\cdot), ..., h_{L,o}(\cdot)) \in \mathcal{H} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_L$ are identified by the conditional moment restrictions (2.2), and that if $h_o(\cdot)$ were known, the finite dimensional parameter $\theta_o \in \Theta$ is (possibly) over identified by the unconditional moment restrictions (2.1).

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 in that X_ℓ being nested. 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 the unknown $h_o()$, which are discussed in Subsection 2.1 and in Section 4.

We now introduce some notation and definitions used in this paper. $E(\cdot)$ and $Var(\cdot)$ are computed with respect to the true unknown distribution F_o of Z. Let Θ be a compact set in $\mathcal{R}^{d_\theta}$ that contains an open ball centering at $\theta_o \in int(\Theta)$. For $\ell = 1, ..., L$, we assume that the nuisance function space $\mathcal{H}_\ell$ is a linear subspace of the space of square integrable functions with respect to X_ℓ. The moment functions $g(\cdot)$ and $\Delta_\ell(\cdot)$ are respectively $d_g \times 1$ and $d_\ell \times 1$ vector valued, with $d_g \geq d_\theta$ and $d_\ell = \dim(h_\ell(x_\ell))$ for $\ell = 1, ..., L$. Let $\frac{\partial E[g(Z,\theta,h)]}{\partial \theta'}$ be the $d_g \times d_\theta$ matrix valued ordinary (partial) derivative of the function $G(\theta, h) = E[g(Z, \theta, h)]$ with respect to θ. Let $\frac{\partial E[g(Z,\theta,h)]}{\partial h_\ell}[v_\ell]$ be the $d_g \times 1$ vector valued pathwise derivative of $G(\theta, h)$ with respect to h_ℓ in the direction $v_\ell \in \mathcal{H}_\ell - \{h_\ell\}$

$$\frac{\partial E[g(Z,\theta,h)]}{\partial h_\ell}[v_\ell] = \left.\frac{\partial E[g(Z,\theta,h_\ell + \tau v_\ell, h_{-\ell})]}{\partial \tau}\right|_{\tau=0} \tag{2.3}$$

where $h_{-\ell,o} = (h_{1,o}, ..., h_{\ell-1,o}, h_{\ell+1,o}, ..., h_{L,o})$. Let $m_\ell(X_\ell, h_\ell) = E[\Delta_\ell(Z, h_\ell)|\, X_\ell]$, and its $d_\ell \times 1$ vector valued parthwise derivative with respect to h_ℓ in the direction $v_\ell \in \mathcal{H}_\ell - \{h_{\ell,o}\}$ is given by

$$\frac{\partial m_\ell(X_\ell, h_{\ell,o})}{\partial h_\ell}[v_\ell] = \left.\frac{\partial m_\ell(X_\ell, h_{\ell,o} + \tau v_\ell)}{\partial \tau}\right|_{\tau=0}. \tag{2.4}$$

Let $\Sigma_\ell(X_\ell)$ be any positive definite symmetric matrix, such as $\Sigma_\ell(X_\ell) = I_\ell$ or $\mathrm{Var}(\Delta_\ell(Z, h_{\ell,o})|\, X_\ell)$. For any v_ℓ, $\widetilde{v}_\ell \in \mathcal{H}_\ell - \{h_{\ell,o}\}$, we define the following inner product

$$\langle v_\ell, \widetilde{v}_\ell \rangle_\ell = E\left[\left(\frac{\partial m_\ell(X_\ell, h_{\ell,o})}{\partial h_\ell}[v_\ell]\right)' \Sigma_\ell(X_\ell)^{-1}\left(\frac{\partial m_\ell(X_\ell, h_{\ell,o})}{\partial h_\ell}[\widetilde{v}_\ell]\right)\right]. \tag{2.5}$$

Let $\mathcal{V}_\ell$ be the Hilbert space generated by $\mathcal{H}_\ell - \{h_{\ell,o}\}$ under the inner product $\langle , \rangle_\ell$. In this paper, because any $h_\ell \in \mathcal{H}_\ell$ and $v_\ell \in \mathcal{V}_\ell$ are restricted to be measurable functions of X_ℓ, and because the conditional moment function $m_\ell(X_\ell, h_\ell)$ depends on h_ℓ only through $h_\ell(X_\ell)$, the pathwise derivative $\frac{\partial m_\ell(X_\ell,h_{\ell,o})}{\partial h_\ell}[v_\ell]$ takes a simple form $\frac{\partial m_\ell(X_\ell,h_{\ell,o}(X_\ell)+\tau v_\ell(X_\ell))}{\partial \tau}|_{\tau=0}$. To stress this fact, we let $\partial m_\ell(x_\ell, h_{\ell,o}(x_\ell))/\partial h_\ell'$ be a $d_\ell \times d_\ell$ matrix-valued (ordinary derivative) function such that

$$\frac{\partial m_\ell(X_\ell, h_{\ell,o}(X_\ell))}{\partial h_\ell'} v_\ell(X_\ell) = \frac{\partial m_\ell(X_\ell, h_{\ell,o})}{\partial h_\ell}[v_\ell] \quad \text{for all } v_\ell \in \mathcal{V}_\ell, \tag{2.6}$$

where $v_\ell(X_\ell)$ is a $d_\ell \times 1$ vector-valued function of X_ℓ. Then the inner product could be equivalently written as

$$\langle v_\ell, \widetilde{v}_\ell \rangle_\ell = E\left[v_\ell(X_\ell)' \left(\frac{\partial m_\ell(X_\ell, h_{\ell,o}(X_\ell))}{\partial h'_\ell} \right)' \Sigma_\ell(X_\ell)^{-1} \frac{\partial m_\ell(X_\ell, h_{\ell,o}(X_\ell))}{\partial h'_\ell} \widetilde{v}_\ell(X_\ell) \right]. \tag{2.7}$$

Finally, we say that $\frac{\partial E[g(Z,\theta_o,h_o)]}{\partial h_\ell}[\cdot]$ is a bounded (or regular) linear functional on $\mathcal{V}_\ell$ if $\frac{\partial E[g_j(Z,\theta_o,h_o)]}{\partial h_\ell}[\cdot]$ is a bounded linear functional on $\mathcal{V}_\ell$ for all $j = 1, ..., d_g$, i.e.,

$$\max_{1\le j\le d_g} \sup_{v\neq 0, v\in\mathcal{V}_\ell} \frac{\left| \frac{\partial E[g_j(Z,\theta_o,h_o)]}{\partial h_\ell}[v] \right|^2}{\langle v, v\rangle_\ell} < \infty.$$

We impose the following basic regularity condition

Condition 1 *(i) the data $\{Z_i\}_{i=1}^n$ is a random sample drawn from the unknown $F_o(\cdot)$; (ii) (θ_o, h_o) satisfies model (2.1) - (2.2), $\frac{\partial E[g(Z,\theta_o,h_o)]}{\partial \theta'}$ has full (column) rank d_θ; (iii) $\frac{\partial m_\ell(X_\ell,h_{\ell,o}(X_\ell))}{\partial h'_\ell}$ is invertible almost surely - X_ℓ for $\ell = 1, ..., L$; (iv) $\frac{\partial E[g(Z,\theta_o,h_o)]}{\partial h_\ell}[\cdot]$ is a bounded linear functional on $\mathcal{V}_\ell$ for $\ell = 1, ..., L$.*

Under Conditions 1(ii) and (iii), the unknown θ_o could be over identified by the unconditional moment restrictions (2.1) if h_o were known, but the unknown function h_o is "exactly" identified by the conditional moment restrictions (2.2).

Our main efficiency bound result is contained in the following theorem.

Theorem 1 *Let Condition 1 hold. If $Var(\rho(Z, \theta_o, h_o))$ is non-singular, then the semiparametric information bound for θ_o is*

$$\left(\frac{\partial E[g(Z, \theta_o, h_o)]}{\partial \theta'} \right)' [Var(\rho(Z, \theta_o, h_o))]^{-1} \left(\frac{\partial E[g(Z, \theta_o, h_o)]}{\partial \theta'} \right), \tag{2.8}$$

where

$$\rho(Z, \theta, h) = g(Z, \theta, h) - \sum_{\ell=1}^{L} \mathbf{v}^*_\ell(X_\ell)\Delta_\ell(Z, h_\ell(X_\ell)) \tag{2.9}$$

with $\mathbf{v}^*_\ell(\cdot)$ $(\ell = 1, ..., L)$ defined in equation (2.15).

Proof. Proof, along with discussion, is presented in Subsection 2.1. ■

This semiparametric efficiency bound result is very general. In addition to allow for non-overlapping or arbitrarily overlapped conditional moment restrictions, to allow for over identified GMM restrictions, it also allows for moment functions $g(Z, \theta, h)$ and $\Delta_\ell(Z, h_\ell(X_\ell))$, $\ell = 1, ..., L$ to be pointwise nonsmooth with respect to parameters.

2.1 Proof of Theorem 1

We first develop a semiparametric information bound under an extra zero derivative restriction (2.10).

Lemma 1 *Let Condition 1 hold and $Var\left(g\left(Z,\theta_o,h_o\right)\right)$ be non-singular. If for all $\ell=1,...,L$, the restriction*

$$\frac{\partial E\left[g\left(Z,\theta_o,h_o\right)\right]}{\partial h_\ell}[v_\ell]=0 \text{ for all } v_\ell\in\mathcal{H}_\ell-\{h_{\ell,o}\} \tag{2.10}$$

is satisfied, then the semiparametric information bound for θ_o is

$$\left(\frac{\partial E[g\left(Z,\theta_o,h_o\right)]}{\partial\theta'}\right)'\left(Var\left[g\left(Z,\theta_o,h_o\right)\right]\right)^{-1}\left(\frac{\partial E[g\left(Z,\theta_o,h_o\right)]}{\partial\theta'}\right). \tag{2.11}$$

Proof. Proof in Appendix. ■

Lemma 1 shows that when the effects of estimating unknown h_o on the moment conditions $E\left[g\left(Z,\theta_o,h_o\right)\right]=0$ are ruled out, the semiparametric efficiency bound of θ_o only relies on $E\left[g\left(Z,\theta_o,h_o\right)\right]=0$ with assuming h_o to be known.

We now argue that the implication of Lemma 1 is not limited to the case where the zero derivative condition (2.10) is satisfied. This is because we can always transform the model such that the moment condition $E\left[g\left(Z,\theta_o,h_o\right)\right]=0$ is equivalent to $E\left[\rho\left(Z,\theta_o,h_o\right)\right]=0$ under (1.2) and moreover

$$\frac{\partial E\left[\rho\left(Z,\theta_o,h_o\right)\right]}{\partial h_\ell}[v_\ell]=0 \text{ for all } v_\ell\in\mathcal{H}_\ell-\{h_{\ell,o}\},\ \ell=1,...,L, \tag{2.12}$$

where the pathwise derivative $\frac{\partial E[\rho(Z,\theta,h)]}{\partial h_\ell}[v_\ell]$ of $\rho\left(Z,\theta,h\right)$ is defined similarly to that in equation (2.3).

To prove Theorem 1, we present a systematic method of transforming the model (1.1) such that the zero derivative restriction (2.12) is always satisfied by the transformed moment $\rho\left(Z,\theta,h\right)$ defined in equation (2.9). By Condition 1(iv) and the Riesz representation theorem, we have: for each $j=1,...,d_g$, there is a unique $u^*_{\ell,j}\in\mathcal{V}_\ell$ such that

$$\frac{\partial E\left[g_j\left(Z,\theta_o,h_o\right)\right]}{\partial h_\ell}[v_\ell]=\left\langle u^*_{\ell,j},v_\ell\right\rangle_\ell=E\left[\left(\frac{\partial m_\ell(X_\ell,h_{\ell,o})}{\partial h_\ell}[u^*_{\ell,j}]\right)'\Sigma_\ell(X_\ell)^{-1}\left(\frac{\partial m_\ell(X_\ell,h_{\ell,o})}{\partial h_\ell}[v_\ell]\right)\right] \tag{2.13}$$

for all $v_\ell \in \mathcal{V}_\ell$. Let

$$\mathbf{v}_\ell^*(X_\ell) \equiv \begin{bmatrix} v_{\ell,1}^*(X_\ell)' \\ \vdots \\ v_{\ell,d_g}^*(X_\ell)' \end{bmatrix} = \begin{bmatrix} \left(\frac{\partial m_\ell(X_\ell,h_{\ell,o})}{\partial h_\ell}[u_{\ell,1}^*]\right)' \Sigma_\ell^{-1}(X_\ell) \\ \vdots \\ \left(\frac{\partial m_\ell(X_\ell,h_{\ell,o})}{\partial h_\ell}[u_{\ell,d_g}^*]\right)' \Sigma_\ell^{-1}(X_\ell) \end{bmatrix}, \tag{2.14}$$

which is a $d_g \times d_\ell$ matrix valued function. Equations (2.13) - (2.14) imply that $\mathbf{v}_\ell^*(\cdot)$ ($\ell = 1, ..., L$) can be equivalently defined as solution to

$$\frac{\partial E\left[g_j\left(Z,\theta_o,h_o\right)\right]}{\partial h_\ell}[v_\ell] = E\left[v_{\ell,j}^*(X_\ell)'\left(\frac{\partial m_\ell(X_\ell,h_{\ell,o})}{\partial h_\ell}[v_\ell]\right)\right] \quad \text{for all } v_\ell \in \mathcal{V}_\ell \tag{2.15}$$

for each $j = 1, ..., d_g$. By equation (2.9),

$$\rho\left(Z,\theta,h\right) = g\left(Z,\theta,h\right) - \sum_{\ell=1}^{L} \mathbf{v}_\ell^*(X_\ell)\Delta_\ell(Z,h_\ell\left(X_\ell\right)).$$

By construction we have

$$\frac{\partial E[\rho\left(Z,\theta_o,h_o\right)]}{\partial\theta'} = \frac{\partial E[g\left(Z,\theta_o,h_o\right)]}{\partial\theta'}. \tag{2.16}$$

Because v_ℓ is restricted to be a function of X_ℓ, we have for each $j = 1, ..., d_g$,

$$\begin{aligned} \frac{\partial E\left[v_{\ell,j}^*\left(X_\ell\right)'\Delta_\ell(Z,h_{\ell,o})\right]}{\partial h_\ell}[v_\ell] &= \left.\frac{\partial E\left[v_{\ell,j}^*\left(X_\ell\right)'\Delta_\ell(Z,h_{\ell,o}\left(X_\ell\right)+\tau v_\ell\left(X_\ell\right))\right]}{\partial\tau}\right|_{\tau=0} \\ &= \left.\frac{\partial E\left[v_{\ell,j}^*\left(X_\ell\right)' m_\ell(X_\ell,h_{\ell,o}\left(X_\ell\right)+\tau v_\ell\left(X_\ell\right))\right]}{\partial\tau}\right|_{\tau=0} \\ &= E\left[v_{\ell,j}^*\left(X_\ell\right)'\left(\frac{\partial m_\ell(X_\ell,h_{\ell,o})}{\partial h_\ell}[v_\ell]\right)\right], \end{aligned}$$

where the last equal sign holds under the assumption allowing for interchanging the expectation and differentiation. Therefore, for all $j = 1, ..., d_g$,

$$\begin{aligned} \frac{\partial E\left[\rho_j\left(Z,\theta_o,h_o\right)\right]}{\partial h_\ell}[v_\ell] &= \frac{\partial E\left[g_j\left(Z,\theta_o,h_o\right)\right]}{\partial h_\ell}[v_\ell] - E\left[v_{\ell,j}^*\left(X_\ell\right)'\left(\frac{\partial m_\ell(X_\ell,h_{\ell,o})}{\partial h_\ell}[v_\ell]\right)\right] \\ &= 0 \quad \text{for all } v_\ell \in \mathcal{V}_\ell \text{ by equation (2.15)}, \end{aligned}$$

which implies that

$$\frac{\partial E\left[\rho\left(Z,\theta_o,h_o\right)\right]}{\partial h_\ell}[v_\ell] = 0 \quad \text{for all } v_\ell \in \mathcal{V}_\ell,\ \ell = 1, ..., L. \tag{2.17}$$

Moreover under the conditional moment restrictions (1.2), the original unconditional moment condition $E\left[g\left(Z,\theta_o,h_o\right)\right]=0$ and the transformed moment condition $E\left[\rho\left(Z,\theta_o,h_o\right)\right]=0$ are equivalent, i.e.

$$E\left[\rho\left(Z,\theta_o,h_o\right)\right]=0\Leftrightarrow E\left[g\left(Z,\theta_o,h_o\right)\right]=0. \tag{2.18}$$

From equations (2.16), (2.17) and (2.18), Lemma 1 is applicable with the transformed moment $E\left[\rho\left(Z,\theta_o,h_o\right)\right]=0$ and hence Theorem 1 holds.

2.2 Special cases

The semiparametric efficiency bound stated in Theorem 1 depends on the functions $\mathbf{v}_\ell^*(\cdot)$ ($\ell=1,...,L$), which are characterized by equation (2.15) but may not have simple closed form expressions in general.

We now consider a special case where the functions $\mathbf{v}_\ell^*(\cdot)$ ($\ell=1,...,L$) and hence the efficiency bound could be solved more explicitly. In the following we let $\frac{\partial E[g(Z,\theta_o,h_o)|X_\ell]}{\partial h_\ell}[v_\ell]$ be the pathwise derivative of the function $E[g\left(Z,\theta_o,h_o\right)|X_\ell]$ with respective to h_ℓ in the direction $v_\ell\in\mathcal{V}_\ell$

$$\frac{\partial E[g\left(Z,\theta_o,h_o\right)|X_\ell]}{\partial h_\ell}[v_\ell]=\left.\frac{\partial E[g\left(Z,\theta_o,h_{\ell,o}+\tau v_\ell,h_{-\ell,o}\right)|X_\ell]}{\partial\tau}\right|_{\tau=0}.$$

Lemma 2 *Let all the conditions of Theorem 1 hold. If for all $\ell=1,...,L$ there is a $d_g\times d_\ell$ matrix valued square integrable function $D_\ell(X_\ell,\theta_o,h_o)$ of X_ℓ such that for all $v_\ell\in\mathcal{V}_\ell$,*

$$D_\ell(X_\ell,\theta_o,h_o)v_\ell(X_\ell)=\frac{\partial E[g\left(Z,\theta_o,h_o\right)|X_\ell]}{\partial h_\ell}[v_\ell]. \tag{2.19}$$

Then the conclusion of Theorem 1 holds with

$$\rho\left(Z,\theta,h\right)=g\left(Z,\theta,h\right)-\sum_{\ell=1}^{L}D_\ell(X_\ell,\theta_o,h_o)\left(\frac{\partial m_\ell(X_\ell,h_{\ell,o}\left(X_\ell\right))}{\partial h_\ell'}\right)^{-1}\Delta_\ell(Z,h_\ell\left(X_\ell\right)). \tag{2.20}$$

Proof. By equations (2.19) and (2.15), we have: for each $j=1,...,d_g$,

$$E\left\{\left[D_{\ell,j}(X_\ell,\theta_o,h_o)-v_{\ell,j}^*\left(X_\ell\right)'\left(\frac{\partial m_\ell(X_\ell,h_{\ell,o}\left(X_\ell\right))}{\partial h_\ell'}\right)\right]v_\ell(X_\ell)\right\}=0$$

for all $v_\ell\in\mathcal{V}_\ell$. Hence

$$D_{\ell,j}(X_\ell,\theta_o,h_o)=v_{\ell,j}^*\left(X_\ell\right)'\left(\frac{\partial m_\ell(X_\ell,h_{\ell,o}\left(X_\ell\right))}{\partial h_\ell'}\right)\text{ almost surely }X_\ell$$

By Condition 1(iii), we obtain

$$\mathbf{v}_\ell^*(X_\ell) = D_\ell(X_\ell, \theta_o, h_o) \left(\frac{\partial m_\ell(X_\ell, h_{\ell,o}(X_\ell))}{\partial h_\ell'} \right)^{-1} \text{ almost surely } X_\ell. \tag{2.21}$$

The conclusion now follows immediately from Theorem 1 under equations (2.9) and (2.21). ∎

If the unconditional moment restrictions (1.1) (i.e., 2.1) take the special form

$$E[g(Z, \theta_o, h_{1,o}(X_1), ...h_{L,o}(X_L))] = 0, \tag{2.22}$$

then equation (2.19) is trivially satisfied with

$$D_\ell(X_\ell, \theta_o, h_o) = \frac{\partial E[g(Z, \theta_o, h_{\ell,o}(X_\ell), h_{-\ell,o}(X_{-\ell}))|X_\ell]}{\partial h_\ell'}, \;\; \ell = 1, ..., L,$$

which could be viewed as an ordinary partial derivative defined similarly as that in equation (2.6). We next give two examples when the unconditional moment restrictions (1.1) is of the special form $E[g(Z, \theta_o, h_o(X))] = 0$ with $L = 1$.

Example 1 (Nonparametric Regression) *The unknown function h_o is identified by the conditional mean restriction: $E[Y - h_o(X)|X] = 0$. For this case, we have $\frac{\partial m(X, h_o(X))}{\partial h'} = -1$ and*

$$\rho(Z, \theta, h) = g(Z, \theta, h) + \frac{\partial E[g(Z, \theta_o, h_o(X))|X]}{\partial h'}(Y - h(X)).$$

Example 2 (Nonparametric Quantile Regression) *The unknown function h_o is identified by the conditional quantile restriction: $E[\tau - I\{Y \leq h_o(X)\}|X] = 0$. Denote $U = Y - h_o(X)$. Let $f_U(\cdot|X)$ be the conditional density of U given X. For this case, we have $\frac{\partial m(X, h_o(X))}{\partial h'} = -f_U(0|X)$ and*

$$\rho(Z, \theta, h) = g(Z, \theta, h) + \frac{\partial E[g(Z, \theta_o, h_o(X))|X]}{\partial h'} \frac{(\tau - I\{Y \leq h(X)\})}{f_U(0|X)}.$$

3 Implication and Discussion of Theorem 1

Suppose that h_o were known, then we would estimate θ_o in (1.1) by Hansen's (1982) optimally weighted GMM

$$\min_{\theta \in \Theta} \left[n^{-1/2} \sum_{i=1}^n g(Z_i, \theta, h_o) \right]' W_n \left[n^{-1/2} \sum_{i=1}^n g(Z_i, \theta, h_o) \right]$$

with an optimal weight matrix W_n such that its probability limit is the inverse of $Var\left[g\left(Z;\theta_o,h_o\right)\right]$. Because under the *i.i.d.* assumption $Var\left[g\left(Z;\theta_o,h_o\right)\right] = Avar\left(n^{-1/2}\sum_{i=1}^{n} g(Z_i,\theta_o,h_o)\right)$, the asymptotic variance of such an infeasible GMM estimator would be equal to the inverse of

$$\left(E\left[\frac{\partial g\left(Z,\theta_o,h_o\right)}{\partial\theta'}\right]\right)'\left(Avar\left(n^{-1/2}\sum_{i=1}^{n} g(Z_i,\theta_o,h_o)\right)\right)^{-1}\left(E\left[\frac{\partial g\left(Z,\theta_o,h_o\right)}{\partial\theta'}\right]\right).$$

Now h_o is in fact unknown, we may consider a feasible version of the preceding GMM estimator by using a weight matrix W_n such that its probability limit is the inverse of $Avar\left(n^{-1/2}\sum_{i=1}^{n} g(Z_i,\theta_o,\widehat{h})\right)$; the asymptotic variance of such a feasible GMM estimator would be the inverse of

$$\left(E\left[\frac{\partial g\left(Z,\theta_o,h_o\right)}{\partial\theta'}\right]\right)'\left(Avar\left(n^{-1/2}\sum_{i=1}^{n} g(Z_i,\theta_o,\widehat{h})\right)\right)^{-1}\left(E\left[\frac{\partial g\left(Z,\theta_o,h_o\right)}{\partial\theta'}\right]\right) \tag{3.1}$$

where $\widehat{h}$ is any consistent nonparametric estimator of h_o. This feasible GMM estimator was discussed by Newey (1994), Ackerberg, Chen, and Hahn (2012), among others. It is not obvious whether the feasible GMM estimator exploits all the information in model (1.1) and (1.2); for one thing, it does not use the (conditional) covariance of the moments between (1.1) and (1.2).

A practical implication of Theorem 1 is that (3.1) is indeed the semiparametric information bound for model (1.1) and (1.2), and therefore, the feasible GMM estimator discussed above is actually semiparametrically efficient. In order to understand this implication, we need to relate $Var\left(\rho\left(Z,\theta_o,h_o\right)\right)$ in the middle of (2.8) in Theorem 1 to the $Avar\left(n^{-1/2}\sum_{i=1}^{n} g(Z_i,\theta_o,\widehat{h})\right)$ in the middle of (3.1). For this purpose, we first use Ai and Chen's (2007) result that when h_o is estimated by a sieve minimum distance (SMD) estimator $\widehat{h}$, we have

$$Avar\left(n^{-1/2}\sum_{i=1}^{n} g(Z_i,\theta_o,\widehat{h})\right) = Var\left(\rho\left(Z,\theta_o,h_o\right)\right). \tag{3.2}$$

Next, we note that the asymptotic variance of $n^{-1/2}\sum_{i=1}^{n} g(Z_i,\theta_o,\widehat{h})$ is invariant to the choice of any consistent nonparametric estimator $\widehat{h}$ of h_o, which follows from Newey's (1994, Proposition 1) observation that the asymptotic variance of a semiparametric root-n consistent estimator is independent of the types of first step consistent nonparametric estimators. Such invariance result implies that the semiparametric efficiency bound of θ_o in model (1.1) and (1.2) can be equivalently written as the term in (3.1). It is clear that equation (3.2) provides one example of illustrating the general form (3.1) when h_o is estimated by a SMD estimator. Another example is provided in the next section where h_o is estimated by a sieve M estimator.

The general expression of the information bound of θ_o in (3.1) indicates that under suitable regularity conditions, the second step GMM estimator $\widehat{\theta}_n$ that solves

$$\min_{\theta\in\Theta}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n}g(Z_i,\theta,\widehat{h})\right]' W_n\left[n^{-\frac{1}{2}}\sum_{i=1}^{n}g(Z_i,\theta,\widehat{h})\right], \tag{3.3}$$

is semiparametric efficient as long as the weighting matrix W_n satisfies

$$W_n^{-1}\rightarrow_p Avar\left(n^{-1/2}\sum_{i=1}^{n}g(Z_i,\theta_o,\widehat{h})\right) \tag{3.4}$$

for any consistent nonparametric estimator $\widehat{h}$ of h_o. In most of the empirical applications, it is a natural exercise to choose a weight matrix W_n satisfying (3.4) such that the two-step GMM estimate $\widehat{\theta}_n$ given in (3.3) is expected to be "limited efficient", i.e. having smallest asymptotic variance among all feasible two-step GMM estimates of θ_o satisfying the unconditional moment restriction (1.1). As a pleasant surprise, Theorem 1 indicates that this natural procedure actually exhausts all the information in model (1.1) and (1.2) and hence is fully efficient.

From the above discussion, one only needs to take care of the effect of the first-step nuisance function estimation in the optimal weight matrix W_n to ensure that two-step GMM estimate $\widehat{\theta}_n$ is asymptotically efficient. Such an adjustment is automatically preformed when W_n is constructed to ensure the two-step GMM estimate achieves the limited efficiency. The simple, optimally weighted two-step GMM estimate (3.3) is not fully efficient in general, as illustrated in Hayashi and Sims (1983), Chamberlain (1992), and Ai and Chen (2009).

4 Sieve Semiparametric Two-step GMM Estimation

Under mild regularity conditions, Chen, Linton and van Keilegom (2003) show that any semiparametric two-step GMM estimator $\widehat{\theta}_n$ defined in (3.3) with an arbitrary positive definite weight matrix W_n has the following asymptotically linear representation

$$\sqrt{n}\left(\widehat{\theta}_n-\theta_o\right)=-\left(\Gamma_1'W\Gamma_1\right)^{-1}\Gamma_1'W\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g\left(Z_i,\theta_o,\widehat{h}_n\right)\right)+o_p\left(1\right),$$

where $\Gamma_1=\frac{\partial E[g(Z,\theta_o,h_o)]}{\partial\theta'}$, W is the probability limit of W_n and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g\left(Z_i,\theta_o,\widehat{h}_n\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g\left(Z_i,\theta_o,h_o\right)+\sum_{\ell=1}^{L}\sqrt{n}\frac{\partial E[g\left(Z,\theta_o,h_o\right)]}{\partial h_\ell}[\widehat{h}_{\ell,n}-h_{\ell,o}]+o_p\left(1\right).$$

Under their condition 2.2.6, i.e.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g\left(Z_i,\theta_o,h_o\right)+\sum_{\ell=1}^{L}\sqrt{n}\frac{\partial E[g\left(Z,\theta_o,h_o\right)]}{\partial h_\ell}[\widehat{h}_{\ell,n}-h_{\ell,o}]\rightarrow_d \mathcal{N}\left(0,V_N\right)$$

where $V_N = Avar\left(n^{-1/2}\sum_{i=1}^{n} g(Z_i,\theta_o,\widehat{h})\right)$ and $\mathcal{N}\left(A,B\right)$ denotes a Gaussian random vector with mean A and variance-covariance matrix B, Chen, Linton and van Keilegom (2003) deduce that

$$\sqrt{n}\left(\widehat{\theta}_n-\theta_o\right)\rightarrow_d \mathcal{N}\left(0,V_\theta\right) \quad \text{with} \quad V_\theta=\left(\Gamma_1' W\Gamma_1\right)^{-1}\left(\Gamma_1' WV_N W\Gamma_1\right)\left(\Gamma_1' W\Gamma_1\right)^{-1}.$$

If we could find a consistent estimator $\widehat{V}_N$ for V_N, then, with the optimal weight matrix $W_n = \widehat{V}_N^{-1} \rightarrow_p W = V_N^{-1}$, we immediately obtain a feasible semiparametric efficient two-step GMM estimator $\widehat{\theta}_n$ with an asymptotic variance given by $\left(\Gamma_1' V_N^{-1}\Gamma_1\right)^{-1}$.

In this section, we provide one feasible efficient estimator of θ_o for the model (1.1) and (1.2), where the unknown nuisance functions $h_{\ell,o}$, $\ell = 1,...,L$, are estimated by sieve M estimation in the first step.

For each $\ell = 1,...,L$, since the unknown true function $h_{\ell,o}\in\mathcal{H}_\ell$ is assumed to be "exactly" identified via the conditional moment restriction $E\left[\Delta_\ell(Z,h_{\ell,o}\left(X_\ell\right))\middle|\, X_\ell\right]=0$ in the sense that Condition 1(iii) holds, one can equivalently define $h_{\ell,o}$ as a solution to a population M estimation problem:

$$\sup_{h\in\mathcal{H}_\ell} E[\varphi_\ell(Z,h_\ell\left(X_\ell\right))],$$

where $\varphi_\ell(Z,h_\ell\left(X_\ell\right))$ is a non-negative measurable criterion function such that

$$\begin{aligned}E\left[\left.\frac{\partial\varphi_\ell(Z,h_{\ell,o})}{\partial h_\ell}\right| X_\ell\right][v_\ell] &= E\left[\left.\frac{\partial\varphi_\ell(Z,h_{\ell,o}\left(X_\ell\right))}{\partial h_\ell'}\right| X_\ell\right] v_\ell(X_\ell)\\ &= E\left[\Delta_\ell(Z,h_{\ell,o}\left(X_\ell\right))\middle|\, X_\ell\right]' v_\ell(X_\ell)=0 \text{ for all } v_\ell\in\mathcal{H}_\ell-\{h_{\ell,o}\}.\end{aligned}$$

In fact, one can typically choose a function $\varphi_\ell(Z,h_\ell\left(X_\ell\right))$ such that

$$\frac{\partial\varphi_\ell(Z,h_\ell\left(X_\ell\right))}{\partial h_\ell'}=\Delta_\ell(Z,h_\ell\left(X_\ell\right))' \quad a.s.-X_\ell \quad \text{for } h_\ell \text{ in a neighborhood of } h_{\ell,o}.$$

Under Condition 1(ii) and (iii), for any $h\in\mathcal{H}_\ell$ in a small neighborhood of $h_{\ell,o}$ with $h_\ell\neq h_{\ell,o}$, we also have:

$$
\begin{aligned}
& E[\varphi_\ell(Z, h_{\ell,o}(X_\ell)) - \varphi_\ell(Z, h_\ell(X_\ell))] \\
& \asymp E\left(-\frac{\partial m_\ell(X_\ell, h_{\ell,o})}{\partial h_\ell}[h_\ell - h_{\ell,o}, h_\ell - h_{\ell,o}]\right) \\
& = -E\left((h_\ell(X_\ell) - h_{\ell,o}(X_\ell))' \frac{\partial m_\ell(X_\ell, h_{\ell,o}(X_\ell))}{\partial h_\ell'}(h_\ell(X_\ell) - h_{\ell,o}(X_\ell))\right) \\
& = \langle h_\ell - h_{\ell,o}, h_\ell - h_{\ell,o}\rangle_\ell > 0,
\end{aligned}
$$

where the third equal sign holds by choosing $\Sigma_\ell(X_\ell) = -\frac{\partial m_\ell(X_\ell, h_{\ell,o}(X_\ell))}{\partial h_\ell'}$ in the definition of the inner product (2.7). We note that such a choice is valid under Condition 1(ii) and (iii) and by the definition of M estimation.

Therefore, for any $\ell = 1, ..., L$, it is natural to estimate $h_{\ell,o}$ by a sieve M estimator $\widehat{h}_{\ell,n}$ that solves

$$
\frac{1}{n}\sum_{i=1}^{n} \varphi_\ell(Z_i, \widehat{h}_{\ell,n}(X_{\ell,i})) \geq \sup_{h_\ell \in \mathcal{H}_{\ell,n}} \frac{1}{n}\sum_{i=1}^{n} \varphi_\ell(Z_i, h_\ell(X_{\ell,i})) - o_p(\frac{1}{n}) \tag{4.1}
$$

where $\mathcal{H}_{\ell,n}$ is a finite dimensional sieve space that becomes dense in the function parameter space $\mathcal{H}_\ell$ as sieve complexity grows with the sample size. In particular, since $h_{\ell,o}$ is only a nuisance function, we could use linear sieve $\mathcal{H}_{\ell,n}$ to simplify the computation. See, e.g., Chen (2007) for many examples of sieve M estimation.

By Condition 1(iv) and the Riesz representation theorem, we have: for each $j = 1, ..., d_g$, there is a unique $u^*_{\ell,j} \in \mathcal{V}_\ell$ such that

$$
\frac{\partial E\left[g_j(Z, \theta_o, h_o)\right]}{\partial h_\ell}[v_\ell] = \left\langle u^*_{\ell,j}, v_\ell\right\rangle_\ell = -E\left[u^*_{\ell,j}(X_\ell)' \frac{\partial m_\ell(X_\ell, h_{\ell,o}(X_\ell))}{\partial h_\ell'} v_\ell(X_\ell)\right] \tag{4.2}
$$

for all $v_\ell \in \mathcal{V}_\ell$. In fact, this $u^*_{\ell,j}$ is exactly the same Riesz representer in the semiparametric efficiency bound calculation equation (2.13) with $\Sigma_\ell(X_\ell) = -\frac{\partial m_\ell(X_\ell, h_{\ell,o}(X_\ell))}{\partial h_\ell'}$. Immediately we also have $v^*_{\ell,j} = -u^*_{\ell,j}$ in equation

We can apply any existing results (such as those in Chen (2007, Theorem 4.3) or Chen, Liao and Sun (2012)) on plug-in sieve M estimation of bounded linear functionals to obtain that for all $j=1,...,d_g$

$$\begin{aligned}&\sqrt{n}\frac{\partial E[g_j\left(Z,\theta_o,h_o\right)]}{\partial h_\ell}[\widehat{h}_{\ell,n}-h_{\ell,o}]=\sqrt{n}\left\langle u^*_{\ell,j},\widehat{h}_{\ell,n}-h_{\ell,o}\right\rangle_\ell\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\partial\varphi_\ell(Z_i,h_{\ell,o})}{\partial h_\ell}[u^*_{\ell,j}]+o_p\left(1\right)\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\Delta_\ell(Z_i,h_{\ell,o}\left(X_{\ell,i}\right))'u^*_{\ell,j}\left(X_{\ell,i}\right)+o_p\left(1\right)\\&=\frac{-1}{\sqrt{n}}\sum_{i=1}^n v^*_{\ell,j}\left(X_{\ell,i}\right)'\Delta_\ell(Z_i,h_{\ell,o}\left(X_{\ell,i}\right))+o_p\left(1\right).\end{aligned}$$

Therefore,

$$\sqrt{n}\frac{\partial E[g\left(Z,\theta_o,h_o\right)]}{\partial h_\ell}[\widehat{h}_{\ell,n}-h_{\ell,o}]=\frac{-1}{\sqrt{n}}\sum_{i=1}^n\mathbf{v}^*_\ell(X_{\ell,i})\Delta_\ell(Z_i,h_{\ell,o}\left(X_{\ell,i}\right))+o_p\left(1\right).$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n g\left(Z_i,\theta_o,\widehat{h}_n\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\rho\left(Z_i,\theta_o,h_o\right)+o_p\left(1\right)$$

with

$$\rho\left(Z,\theta_o,h_o\right)=g\left(Z,\theta_o,h_o\right)-\sum_{\ell=1}^L\mathbf{v}^*_\ell(X_\ell)\Delta_\ell(Z,h_{\ell,o}\left(X_\ell\right)).$$

Hence

$$V_N=Avar\left(n^{-1/2}\sum_{i=1}^n g(Z_i,\theta_o,\widehat{h})\right)=Var\left(\rho\left(Z,\theta_o,h_o\right)\right).$$

Unfortunately, the Riesz representer $u^*_{\ell,j}$ or $v^*_{\ell,j}$ may not have a closed form expression in general. Following Chen, Liao and Sun (2012), we can always compute a sieve Riesz representer $u^*_{\ell,j,n}\in\mathcal{H}_{\ell,n}$ such that

$$\frac{\partial E\left[g_j\left(Z,\theta_o,h_o\right)\right]}{\partial h_\ell}[v_\ell]=-E\left[u^*_{\ell,j,n}(X_\ell)'\frac{\partial m_\ell(X_\ell,h_{\ell,o}(X_\ell))}{\partial h'_\ell}v_\ell(X_\ell)\right]\quad\text{for all } v_\ell\in\mathcal{H}_{\ell,n},$$

which has a closed form solution, and satisfies $\left\|v^*_{\ell,j}-v^*_{\ell,j,n}\right\|_\ell\to0$ as $\dim(\mathcal{H}_{\ell,n})\to\infty$. See the Appendix for details. Moreover,

$$\begin{aligned}&\sqrt{n}\frac{\partial E[g_j\left(Z,\theta_o,h_o\right)]}{\partial h_\ell}[\widehat{h}_{\ell,n}-h_{\ell,o}]\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\partial\varphi_\ell(Z_i,h_{\ell,o})}{\partial h_\ell}[u^*_{\ell,j,n}]+o_p\left(1\right)=\frac{-1}{\sqrt{n}}\sum_{i=1}^n v^*_{\ell,j,n}\left(X_{\ell,i}\right)'\Delta_\ell(Z_i,h_{\ell,o}\left(X_{\ell,i}\right))+o_p\left(1\right).\end{aligned}$$

Denote

$$\rho_n(Z,\theta,h) \equiv \begin{bmatrix} \rho_{1,n}(Z,\theta,h) \\ \vdots \\ \rho_{d_g,n}(Z,\theta,h) \end{bmatrix} = \begin{bmatrix} g_1(Z,\theta,h) - \sum_{\ell=1}^{L} v^*_{\ell,1,n}(X_\ell)' \Delta_\ell(Z, h_\ell(X_\ell)) \\ \vdots \\ g_{d_g}(Z,\theta,h) - \sum_{\ell=1}^{L} v^*_{\ell,d_g,n}(X_\ell)' \Delta_\ell(Z, h_\ell(X_\ell)) \end{bmatrix}$$
$$= g(Z,\theta,h) - \sum_{\ell=1}^{L} \mathbf{v}^*_{\ell,n}(X_\ell)\Delta_\ell(Z, h_\ell(X_\ell)),$$

which, unlike $\rho(Z,\theta,h)$, has a known functional form, and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g\left(Z_i,\theta_o,\widehat{h}_n\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \rho_n(Z_i,\theta_o,h_o) + o_p(1).$$

The next proposition summaries the normality result:

Proposition 1 *Under some regularity conditions, the GMM estimator defined in (3.3) with* $p\lim_n W_n = W$ *satisfies*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_o\right) \to_d \mathcal{N}\left(0, (\Gamma_1' W \Gamma_1)^{-1}(\Gamma_1' W V_N W \Gamma_1)(\Gamma_1' W \Gamma_1)^{-1}\right),$$

$$V_N = \lim_{n\to\infty} E\left[n^{-1}\sum_{i=1}^{n} \rho_n(Z_i,\theta_o,h_o)\rho_n(Z_i,\theta_o,h_o)'\right]. \tag{4.3}$$

Proof. The claimed result follows directly from Theorem 2 of Chen, Linton and van Keilegom (2003), Theorem 4.3 of Chen (2007) and Theorem 3.1 of Chen, Liao and Sun (2012). ■

Remark 1 *When the unconditional moment function* $g(Z,\theta,h)$ *is continuously differentiable at* (θ_o,h_o)*, the asymptotic variance of the semiparametric efficient two-step GMM estimator* $\widehat{\theta}_n$ *can be consistently estimated by*

$$\left(\widehat{\Gamma}_{1,n}' \widehat{V}_{N,n}^{-1} \widehat{\Gamma}_{1,n}\right)^{-1},$$

with $\widehat{\Gamma}_{1,n} = n^{-1}\sum_{i=1}^{n} \frac{\partial g(Z_i,\widehat{\theta}_n,\widehat{h}_n)}{\partial\theta}$ *and*

$$\widehat{V}_{N,n} = n^{-1}\sum_{i=1}^{n}\left(\widehat{\rho}_n(Z_i,\widehat{\theta}_n,\widehat{h}_n)\right)\left(\widehat{\rho}_n(Z_i,\widehat{\theta}_n,\widehat{h}_n)\right)',$$

$$\widehat{\rho}_n(Z,\theta,h) = g(Z,\theta,h) - \sum_{\ell=1}^{L} \widehat{\mathbf{v}}^*_{\ell,n}(X_\ell)\Delta_\ell(Z, h_\ell(X_\ell)),$$

where $\widehat{\mathbf{v}}^*_{\ell,n}$ is a sieve estimator of $\mathbf{v}^*_{\ell,n}$ and is defined in (5.43) of Appendix 5.2.

Finally, when sieve M procedure is used to estimate unknown functions $h_{\ell,o}$ in the first step, we can apply the numerical equivalence results in Ackerberg, Chen, and Hahn (2012) to compute $\widehat{V}_{N,n}$ using standard software packages for parametric two-step GMM estimators.

5 Appendix

5.1 Proof of the Main Results in Section 2

Proof of Lemma 1. For the ease of notation and without loss of generality, we assume in this proof that $L = 2$. Let $f_o(z)$ to be the true density of Z with respect to a sigma finite dominating measure $\mu(z)$, and $f_o(z_{-j}|x_j)$ be the true conditional density of Z_{-j} given $X_j = x_j$ ($j = 1, 2$). Here, Z_{-j} denotes the components of Z not in the conditioning variable X_j, $j = 1, 2$. and $\mathcal{F}$ be a class of candidate density function of Z with $f_o \in \mathcal{F}$. Define a class of density functions $\mathcal{F}_\alpha$ that satisfy the conditional and unconditional moment conditions:

$$\begin{aligned}\mathcal{F}_\alpha = \Big\{ f \in \mathcal{F} : \int \Delta_1\left(z_{-1}, h_1(x_1)\right) f(z_{-1}|x_1) d\mu(z_{-1}) = 0, \\ \int \Delta_2\left(z_{-2}, h_2(x_2)\right) f(z_{-2}|x_2) d\mu(z_{-2}) = 0, \\ \int g\left(z, \theta, h_1, h_2\right) f(z) d\mu(z) = 0 \Big\}. \end{aligned} \tag{5.1}$$

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

$$\mathcal{F}_\alpha = \left\{ f\left(z |\, \theta, h_1, h_2, \eta\right) : \eta \in \mathcal{G} \right\} \tag{5.2}$$

for any $\alpha = (\theta, h_1, h_2) \in \Theta \times \mathcal{H}_1 \times \mathcal{H}_2$. Let $\mathcal{V}_\theta \times \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{V}_\eta$ denote the completion of $\Theta \times \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{G} - \{(\theta_o, h_{1,o}, h_{2,o}, \eta_o)\}$ where η_o satisfies

$$f\left(z |\, \theta_o, h_{1,o}, h_{2,o}, \eta_o\right) = f_o(z).$$

We will consider the parametric family $f\left(z |\, \theta_o + \tau_\theta \theta, h_{1,o} + \tau_1 v_1, h_{2,o} + \tau_2 v_2, \eta_o + \tau_\eta v_\eta\right)$. The scores in the direction of τ_θ, τ_1, τ_2, τ_η of this family are such that

$$\begin{aligned} s_\theta\left(Z\right) &= c_{\theta,1}\left(Z_{-1} |\, X_1\right) + d_{\theta,1}\left(X_1\right) \\ &= c_{\theta,2}\left(Z_{-2} |\, X_2\right) + d_{\theta,2}\left(X_2\right) \end{aligned}$$

$$\begin{aligned} s_{h_1}(Z)[v_1] &= c_{h_1,1}(Z_{-1}|X_1)[v_1] + d_{h_1,1}(X_1)[v_1] \\ &= c_{h_1,2}(Z_{-2}|X_2)[v_1] + d_{h_1,2}(X_2)[v_1] \\ s_{h_2}(Z)[v_2] &= c_{h_2,1}(Z_{-1}|X_1)[v_2] + d_{h_2,1}(X_1)[v_2] \\ &= c_{h_2,2}(Z_{-2}|X_2)[v_2] + d_{h_2,2}(X_2)[v_2] \end{aligned}$$

$$\begin{aligned} s_{\eta}(Z)[v_{\eta}] &= c_{\eta,1}(Z_{-1}|X_1)[v_{\eta}] + d_{\eta,1}(X_1)[v_{\eta}] \\ &= c_{\eta,2}(Z_{-2}|X_2)[v_{\eta}] + d_{\eta,2}(X_2)[v_{\eta}] \end{aligned}$$

with

$$E[c_{\theta,1}(Z_{-1}, X_1)[v_{\eta}]|X_1] = 0 \tag{5.3}$$
$$E[d_{\theta,1}(X_1)[v_{\eta}]] = 0 \tag{5.4}$$
$$E[c_{\theta,2}(Z_{-2}, X_2)[v_{\eta}]|X_2] = 0 \tag{5.5}$$
$$E[d_{\theta,2}(X_2)[v_{\eta}]] = 0 \tag{5.6}$$

$$E[c_{h_1,1}(Z_{-1}|X_1)[v_1]|X_1] = 0 \tag{5.7}$$
$$E[d_{h_1,1}(X_1)[v_1]] = 0 \tag{5.8}$$
$$E[c_{h_2,1}(Z_{-1}|X_1)[v_1]|X_1] = 0 \tag{5.9}$$
$$E[d_{h_2,1}(X_1)[v_1]] = 0 \tag{5.10}$$

$$E[c_{h_2,2}(Z_{-2}|X_2)[v_2]|X_2] = 0 \tag{5.11}$$
$$E[d_{h_2,2}(X_2)[v_2]] = 0 \tag{5.12}$$
$$E[c_{h_1,2}(Z_{-2}|X_2)[v_2]|X_2] = 0 \tag{5.13}$$
$$E[d_{h_1,2}(X_2)[v_2]] = 0 \tag{5.14}$$

and

$$E[c_{\eta,1}(Z_{-1}, X_1)[v_{\eta}]|X_1] = 0 \tag{5.15}$$
$$E[d_{\eta,1}(X_1)[v_{\eta}]] = 0 \tag{5.16}$$
$$E[c_{\eta,2}(Z_{-2}, X_2)[v_{\eta}]|X_2] = 0 \tag{5.17}$$
$$E[d_{\eta,2}(X_2)[v_{\eta}]] = 0 \tag{5.18}$$

Here, $c_{h_1}(Z_{-1}|X_1)[v_1]$ and $d_{h_1}(X_1)[v_1]$ denote the conditional score of Z_{-1} given X_1 and the marginal score of X_1, obtained by differentiating the log likelihood with respect to τ_1, for example. Blow, we will write $c_{h_1}(Z)[v_1] \equiv c_{h_1}(Z_{-1}|X_1)[v_1]$, e.g., for simplicity of notations.

Differentiating[1] the moment restrictions $E[\Delta_\ell(Z, h_{\ell,o}(X_\ell))|X_\ell] = 0$ and $E[g(Z, \alpha_o)] = 0$, we obtain the nonparametric tangent space $\mathcal{T}$ as the completion of the set consisting of $s_{h_1}(z)[v_1] + s_{h_2}(z)[v_2] + s_k(z)[v_k]$, where s's satisfy (5.3) - (5.18) as well as

$$E[\Delta_1(Z, h_{1,o})\, c_{\theta,1}(Z)|X_1] = 0 \tag{5.19}$$

$$\frac{\partial m_1(X_1, h_{1,o}(X_1))}{\partial h_1'} v_1(X_1) + E[\Delta_1(Z, h_{1,o})\, c_{h_1,1}(Z)[v_1]|X_1] = 0 \tag{5.20}$$

$$E[\Delta_1(Z, h_{1,o})\, c_{h_2,1}(Z)[v_2]|X_1] = 0 \tag{5.21}$$

$$E[\Delta_1(Z, h_{1,o}) c_{\eta,1}(Z)[v_\eta]|X_1] = 0 \tag{5.22}$$

$$E[\Delta_2(Z, h_{2,o}) c_{\theta,2}(Z)|X_2] = 0 \tag{5.23}$$

$$E[\Delta_2(Z, h_{2,o}) c_{h_1,2}(Z)[v_1]|X_2] = 0 \tag{5.24}$$

$$\frac{\partial m_2(X_2, h_{2,o}(X_2))}{\partial h_2'} v_2(X_2) + E[\Delta_2(Z, h_{2,o}) c_{h_2,2}(Z)[v_2]|X_2] = 0 \tag{5.25}$$

$$E[\Delta_2(Z, h_{2,o}) c_{\eta,2}(Z)[v_\eta]|X_2] = 0 \tag{5.26}$$

and

$$\frac{\partial E[g(Z, \theta_o, h_{1,o}, h_{2,o})]}{\partial \theta'} + E\left[g(Z, \theta_o, h_{1,o}, h_{2,o})\, s_\theta(Z)'\right] = 0 \tag{5.27}$$

$$E[g(Z, \theta_o, h_{1,o}, h_{2,o})\, s_{h_1}(Z)[v_1]] = 0 \tag{5.28}$$

$$E[g(Z, \theta_o, h_{1,o}, h_{2,o})\, s_{h_2}(Z)[v_2]] = 0 \tag{5.29}$$

$$E[g(Z, \theta_o, h_{1,o}, h_{2,o})\, s_\eta(Z)[v_\eta]] = 0 \tag{5.30}$$

for any $(v_{h_1}, v_{h_2}, v_\eta) \in \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{V}_\eta$. Note that (2.17) is used in (5.28) and (5.29).

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

$$S_\theta^*(Z) = -\left(\frac{\partial E[g(Z)]}{\partial \theta'}\right)' \left\{E[g(Z)\, g(Z)']\right\}^{-1} g(Z) \tag{5.31}$$

[1] We assume that the regularity condition as in Newey (1990, Definition A.1) is satisfied.

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

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

$$0 = E\left[\Delta_1(Z, h_{1,o})\left\{c_{\theta,1}(Z) - S_\theta^*(Z) - c_{h_1,1}(Z)[\Lambda_1^*] - c_{h_2,1}(Z)[\Lambda_2^*]\right\}\middle| X_1\right] \tag{5.32}$$

and

$$0 = E\left[\Delta_2(Z, h_{2,o})\left\{c_{\theta,2}(Z) - S_\theta^*(Z) - c_{h_1,2}(Z)[\Lambda_1^*] - c_{h_2,2}(Z)[\Lambda_2^*]\right\}\middle| X_2\right] \tag{5.33}$$

Letting $v_{h_1} = \Lambda_1^*(X_1)$ in (5.20) and $v_{h_2} = \Lambda_2^*(X_2)$ in (5.21), we get

$$\frac{\partial m_1(X_1, h_{1,o}(X_1))}{\partial h_1'}\Lambda_1^*(X_1) + E\left[\Delta_1(Z, h_{1,o})\, c_{h_1,1}(Z)[\Lambda_1^*]\middle| X_1\right] = 0 \tag{5.34}$$

and

$$E\left[\Delta_1(Z, h_{1,o})\, c_{h_2,1}(Z)[\Lambda_2^*]\middle| X_1\right] = 0. \tag{5.35}$$

Using (5.19) along with (5.31), (5.34) and (5.32), we write

$$0 = \left(\frac{\partial E[g(Z)]}{\partial \theta'}\right)'\left\{E[g(Z)\, g(Z)']\right\}^{-1} E[g(Z)\,\Delta_1(Z, h_{1,o})| X_1] + \frac{\partial m_1(X_1, h_{1,o}(X_1))}{\partial h_1'}\Lambda_1^*(X_1)$$

which can be solved for $\Lambda_1^*(X_1)$ as long as $\partial m_1(X_1, h_{1,o}(X_1)) / \partial h_1' \neq 0$ almost surely. Similarly, we can solve for $\Lambda_2^*(X_2)$ as long as $\partial m_2(X_2, h_{2,o}(X_2)) / \partial h_2' \neq 0$ almost surely.

Now let

$$W = s_\theta(Z) - S_\theta^*(Z) - s_{h_1}(Z)[\Lambda_1^*] - s_{h_2}(Z)[\Lambda_2^*]$$

We will show that W satisfies the properties (5.15)-(5.18), (5.22), (5.26), and (5.30) of the $s_\eta(Z)[v_\eta]$.

By construction, we have $E[W] = 0$. Taking

$$\begin{aligned}
\widetilde{d}_{\eta,1}(X_1)[v_\eta] &= E[W| X_1] \\
&= d_{\theta,1}(X_1) - d_{h_1,1}(X_1)[\Lambda_1^*] - d_{h_2,1}(X_1)[\Lambda_2^*] \\
&\quad + E\left[c_{\theta,1}(Z) - S_\theta^*(Z) - c_{h_1,1}(Z)[\Lambda_1^*] - c_{h_2,1}(Z)[\Lambda_2^*]\middle| X_1\right] \\
&= d_{\theta,1}(X_1) - d_{h_1,1}(X_1)[\Lambda_1^*] - d_{h_2,1}(X_1)[\Lambda_2^*] - E[S_\theta^*(Z)| X_1]
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{c}_{\eta,1}(z)[v_\eta] &= W - \widetilde{d}_{\eta,1}(X_1)[v_\eta] \\
&= c_{\theta,1}(X_1) - c_{h_1,1}(Z)[\Lambda_1^*] - c_{h_2,1}(Z)[\Lambda_2^*] - S_\theta^*(Z) + E[S_\theta^*(Z)| X_1]
\end{aligned}$$

we can see that properties (5.15) and (5.16) are satisfied for

$$W = \widetilde{c}_{\eta,1}\left(z\right)\left[v_\eta\right] + \widetilde{d}_{\eta,1}\left(X_1\right)\left[v_\eta\right].$$

With $\widetilde{c}_{\eta,2}\left(Z\right)\left[v_\eta\right]$ and $\widetilde{d}_{\eta,2}\left(X_2\right)\left[v_\eta\right]$ similarly defined, we can see that properties (5.17) and (5.18) are also satisfied.

Equations (5.32) implies that

$$\begin{aligned} E\left[\Delta_1(Z,h_{1,o})\widetilde{c}_{\eta,1}\left(z\right)\left[v_\eta\right]\middle| X_1\right] &= E\left[\Delta_1(Z,h_{1,o})\left\{c_{\theta,1}\left(Z\right) - S_\theta^*\left(Z\right) - c_{h_1,1}\left(Z\right)\left[\Lambda_1^*\right] - c_{h_2,1}\left(Z\right)\left[\Lambda_2^*\right]\right\}\middle| X_1\right] \\ &\quad + E\left[\Delta_1(Z,h_{1,o})\middle| X_1\right]E\left[S_\theta^*(Z)\middle| X_1\right] \\ &= 0. \end{aligned}$$

which implies that the property (5.22) is satisfied by W. Likewise, (5.26) are satisfied by W.

Using (5.27)-(5.29), we obtain

$$\begin{aligned} E\left[Wg\left(Z\right)'\right] &= E\left[s_\theta\left(Z\right)g\left(Z\right)'\right] - E\left[S_\theta^*\left(Z\right)g\left(Z\right)'\right] \\ &= -\left(\frac{\partial E[g\left(Z\right)]}{\partial\theta'}\right)' + \left(\frac{\partial E[g\left(Z\right)]}{\partial\theta'}\right)'\left\{E[g\left(Z\right)g\left(Z\right)']\right\}^{-1}\left\{E[g\left(Z\right)g\left(Z\right)']\right\} \\ &= 0. \end{aligned} \tag{5.36}$$

which shows that the property (5.30) is satisfied.

These observations lead us to conclude that

$$s_{h_1}\left(Z\right)\left[\Lambda_1^*\right] + s_{h_2}\left(Z\right)\left[\Lambda_2^*\right] + W \in \mathcal{T}. \tag{5.37}$$

Because $S_\theta^*\left(Z\right)$ is proportional to $g\left(Z\right)$, we can deduce from (5.28)-(5.30) that $S_\theta^*(Z) \perp \mathcal{T}$. Along with (5.37), this implies that $S_\theta^*(Z)$ is the residual of the projection of s_θ on $\mathcal{T}$. Thus the semiparametric information bound of θ_o is

$$E[S_\theta^*(Z)S_\theta^*(Z)'] = \left(\frac{\partial E[g\left(Z\right)]}{\partial\theta'}\right)'\left\{E[g\left(Z\right)g\left(Z\right)']\right\}^{-1}\left(\frac{\partial E[g\left(Z\right)]}{\partial\theta'}\right). \tag{5.38}$$

■

5.2 Sieve Riesz representation of bounded linear functionals

It may be difficult to compute the Riesz representer $u_{\ell,j}^*$ $(j = 1, ..., d_g)$ on the infinite dimensional Hilbert space $\mathcal{V}_\ell$. But we can always explicitly compute a Riesz representer $u_{\ell,j,n}^*$ on the finite dimensional Hilbert space $\mathcal{V}_{\ell,n}$ generated by the completion of $\mathcal{H}_{\ell,n} - \{h_{\ell,o,n}\}$ where $h_{\ell,o,n} \in \mathcal{H}_{\ell,n}$ and one can show that $\left\| u^*_{\ell,j} - u^*_{\ell,j,n} \right\|_\ell \to 0$ as $\dim(\mathcal{H}_{\ell,n}) \to \infty$ (see, e.g. Chen, Liao and Sun, 2012).

Formally, as $\frac{\partial E[g_j(Z,\theta_o,h_o)]}{\partial h_\ell}[\cdot]$ is a bounded linear functional, by Riesz representation Theorem, there exists a $u^*_{\ell,j,n} \in \mathcal{V}_{\ell,n}$ such that

$$\frac{\partial E\left[g_j\left(Z,\theta_o,h_o\right)\right]}{\partial h_\ell}[v] = \langle v, u^*_{\ell,j,n}\rangle_\ell \text{ for all } v \in \mathcal{V}_{\ell,n}, \text{ and} \tag{5.39}$$

$$\left\| u^*_{\ell,j,n} \right\|^2_\ell = \sup_{v\in\mathcal{V}_{\ell,n}, v\neq 0} \frac{\left| \frac{\partial E[g_j(Z,\theta_o,h_o)]}{\partial h_\ell}[v] \right|^2}{\left\| v \right\|^2_\ell} < \infty. \tag{5.40}$$

To simplify notation, we assume that $h_{\ell,o}$ is a scalar-valued function that can be approximated by a linear sieve. In particular, we let $P_{K_n}(\cdot) = [p_1(\cdot), ..., p_{K_n}(\cdot)]'$ be a $K_n \times 1$ vector denoting the sieve basis functions for $\mathcal{H}_{\ell,n}$ and $\mathcal{V}_{\ell,n}$. Let $\Gamma_{\ell,j}(\alpha_o)[v] = \frac{\partial E[g_j(Z,\theta_o,h_o)]}{\partial h_\ell}[v]$. Using the fact that $v(\cdot) = P_K(\cdot)'\beta_K$ for any $v \in \mathcal{V}_{\ell,n}$, we deduce that

$$\Gamma_{\ell,j}(\alpha_o)[u^*_{\ell,j,n}] = \{\Gamma_{\ell,j}(\alpha_o)[P_K]\}' \left\{ E\left[-P_K\left(X_\ell\right) \frac{\partial m_\ell(X_\ell, h_{\ell,o}\left(X_\ell\right))}{\partial h'_\ell} P_K\left(X_\ell\right)' \right] \right\}^{-1} \Gamma_{\ell,j}(\alpha_o)[P_K] \tag{5.41}$$

where $\Gamma_{\ell,j}(\alpha_o)[P_K] = (\Gamma_{\ell,j}(\alpha_o)[p_1(X_\ell)], ..., \Gamma_{\ell,j}(\alpha_o)[p_K(X_\ell)])'$. From the expression in (5.41), we obtain

$$u^*_{\ell,j,n}(\cdot) = P_K(\cdot)' \left\{ E\left[-P_K\left(X_\ell\right) \frac{\partial m_\ell(X_\ell, h_{\ell,o}\left(X_\ell\right))}{\partial h'_\ell} P_K\left(X_\ell\right)' \right] \right\}^{-1} \Gamma_{\ell,j}(\alpha_o)[P_K] \tag{5.42}$$

By the definition of Riesz representer $u^*_{\ell,j,n}(\cdot)$, we can define an empirical Riesz representer $\widehat{u}^*_{\ell,j,n}(\cdot)$ in the following

$$\widehat{u}^*_{\ell,j,n}(\cdot) = P_K(\cdot)' \left[-\frac{1}{n}\sum_{i=1}^n P_K\left(X_{\ell,i}\right) \frac{\partial \Delta_\ell(Z_i, \widehat{h}_{\ell,n}\left(X_{\ell,i}\right))}{\partial h'_\ell} P_K\left(X_{\ell,i}\right)' \right]^{-1} \widehat{\Gamma}_{\ell,j}(\widehat{\alpha}_n)[P_K], \tag{5.43}$$

where $\frac{\partial \Delta_\ell(Z,h_\ell(X_\ell))}{\partial h'_\ell}$ satisfies $E\left[\frac{\partial \Delta_\ell(Z,h_\ell(X_\ell))}{\partial h'_\ell} \middle| X_\ell \right] = \frac{\partial m_\ell(X_\ell,h_\ell(X_\ell))}{\partial h'_\ell}$ for any $h_\ell\left(X_\ell\right)$ in the local neighborhood of $h_{\ell,o}\left(X_\ell\right)$, and

$$\widehat{\Gamma}_{\ell,j}(\widehat{\alpha}_n)[P_K()]' \equiv \left(\frac{1}{n}\sum_{i=1}^n \frac{\partial g_j\left(Z_i,\widehat{\alpha}_n\right)}{\partial h_\ell}\left[p_1()\right], ..., \frac{1}{n}\sum_{i=1}^n \frac{\partial g_j\left(Z_i,\widehat{\alpha}_n\right)}{\partial h_\ell}\left[p_K()\right] \right).$$

References

[1] Ackerberg, D., X. Chen, and J. Hahn (2012): "A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators," *Review of Economics and Statistics* 94, 481-498.

[2] Ai, C. and X. Chen (2007): "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables," *Journal of Econometrics* 141, 5-43.

[3] Ai, C. and X. Chen (2009): "Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions Containing Unknown Functions," *Cowles Foundation Discussion Paper No 1731*, forthcoming in *Journal of Econometrics*.

[4] Andrews, D. (1994): "Asymptotics for Semi-parametric Econometric Models via Stochastic Equicontinuity," *Econometrica* 62, 43-72.

[5] Chamberlain, G. (1992): "Comment: Sequential Moment Restrictions in Panel Data," *Journal of Business and Economic Statistics* 10, 20-26.

[6] Chen, X. (2007) "Large Sample Sieve Estimation of Semi-Nonparametric Models", in J.J. Heckman and E.E. Leamer (eds.), *The Handbook of Econometrics*, vol. 6B. Amsterdam: North-Holland.

[7] Chen, X., Z. Liao and Y. Sun (2012) "Sieve Inference on Semi-nonparametric Time Series Models," *Cowles Foundation Discussion Paper 1849*.

[8] Chen, X., O. Linton and I. van Keilegom (2003) "Estimation of Semiparametric Models when the Criterion Function is not Smooth," *Econometrica*, 71, 1591-1608.

[9] Hansen, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[10] Hayashi, F., and C. Sims, (1983): "Nearly Efficient Estimation of Time Series Models with Predetermined, But Not Exogenous, Instruments," *Econometrica*, 51, 783-798.

[11] Newey, W.K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics* 5, 99 – 135.

[12] Newey, W.K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica* 62, 1349 – 1382.

[13] Pakes, A. and G. Olley (1995): "A Limit Theorem for a Smooth Class of Semiparametric Estimators," *Journal of Econometrics* 65, 295–332.