cemmap
centre for microdata methods and practice

Semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions

Chunrong Ai
Xiaohong Chen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP28/09

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions Containing Unknown Functions*

Chunrong Ai[†] Xiaohong Chen[‡]

First draft: September 2004, Revised version: September 2009

Abstract

This paper computes the semiparametric efficiency bound for finite dimensional parameters identified by models of sequential moment restrictions containing unknown functions. Our results extend those of Chamberlain (1992b) and Ai and Chen (2003) for semiparametric conditional moment restriction models with identical information sets to the case of nested information sets, and those of Chamberlain (1992a) and Brown and Newey (1998) for models of sequential moment restrictions without unknown functions to cases with unknown functions of possibly endogenous variables. Our bound results are applicable to semiparametric panel data models and semiparametric two stage plug-in problems. As an example, we compute the efficiency bound for a weighted average derivative of a nonparametric instrumental variables (IV) regression, and find that the simple plug-in estimator is not efficient. Finally, we present

*Chen is grateful to Lars Peter Hansen for numerous insightful comments about efficiency of GMM and related models and methods. We are grateful to Jim Powell for his encouragement upon hearing the first presentation in 2005. We thank Demian Pouzo for the Monte Carlo study, and Jinyong Hahn, Alex Torgovitsky, the guest coeditor, three anonymous referees for suggestions that improve the exposition. We also thank D. Andrews, D. Bhattacharya, R. Blundell, J. Horowitz, W. Newey, E. Tamer, and participants of 2005 World Congress Econometric Society Meetings, 2007 LSE Conference in honor of Peter Robinson, 2007 CIRANO and CIREQ econometric conference on GMM in Montreal, econometric workshops at various universities for helpful comments. Ai's research is supported by 2003 summer grant from the Warrington Business School at University of Florida and Project 211 Phase III and Shanghai Leading Academic Discipline Project (Number: B803) at Shanghai University of Finance and Economics, China. Chen's research is supported by the National Science Foundation grants SES-0318091 and SES-0631613. All errors are the responsibility of the authors.

[†]Department of Economics, University of Florida, Gainesville, FL 32611, USA, and Shanghai University of Finance and Economics, China. Email: chunrong.ai@cba.ufl.edu.

[‡]Corresponding author. Xiaohong Chen, Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520, USA. Tel: 1 203 432 5852; fax: 1 203 432 6167. Email: xiaohong.chen@yale.edu.

an optimally weighted, orthogonalized, sieve minimum distance estimator that achieves the semiparametric efficiency bound.

JEL Classification: C14; C22
Keywords: Sequential moment models; Semiparametric efficiency bounds; Optimally weighted orthogonalized sieve minimum distance; Nonparametric IV regression; Weighted average derivatives; Partially linear quantile IV

1 Introduction

Since the publication of Hansen's (1982) seminal work on the Generalized Method of Moments (GMM), moment restriction models have become a popular and useful framework for analyzing economic data. See Hansen (2007) for an excellent review of the original GMM, its numerous extensions and a wide range of applications.

Motivated by an ever expanding variety of applications, one important branch of extensions focuses on semiparametric efficiency bounds and efficient estimation of more general moment restriction models.[1] In this paper, we contribute to this line of research by characterizing the semiparametric efficiency bound for finite dimensional parameters (θ_o) that are identified by models of sequential moment restrictions containing unknown functions:

$$E[\rho_t(Z;\theta_o,h_o(\cdot))|X^{(t)}]=0 \quad \text{for } t=1,...,T \quad \text{almost surely,} \tag{1}$$

where $Z=(Y',X')'$ denotes a multivariate random variable with support $\mathcal{Z}$ and $X\equiv X^{(T)}$, and $\rho_t(z;\theta,h(\cdot))$ denotes a $d_{\rho_t}\times 1$ vector of residual functions whose functional forms are known up to the unknown true parameter values (θ_o,h_o), with $h_o(\cdot)=(h_{o1}(\cdot),...,h_{oq}(\cdot))$ as the $q\times 1$ vector of real-valued measurable functions that may depend on endogenous variables Y and other unknown parameters. $E[\cdot|X^{(t)}]$ denotes the conditional expectation under the true (but unknown) conditional distribution function $F_{Z|X^{(t)}}$ for $t=1,...,T$. The sigma-field generated by the conditioning variable

[1]See, for example, Hansen (1985, 1993), Hansen, Heaton and Ogaki (1988), Chamberlain (1987, 1992a, 1992b), Newey (1993, 2004), Hahn (1997), Carrasco and Florens (2000), Ai and Chen (2003), Chernozhukov and Hong (2003), Donald, Imbens and Newey (2003), Kitamura, Tripathi and Ahn (2004), Newey and Smith (2004), Antoine, Bonnal and Renault (2007),to name only a few.

$X^{(t)}$, $\sigma\left(\left\{X^{(t)}\right\}\right)$, satisfies a nesting structure[2]

$$\{1\} \subseteq \sigma\left(\left\{X^{(1)}\right\}\right) \subset \sigma\left(\left\{X^{(2)}\right\}\right) \subset \cdots \subset \sigma\left(\left\{X^{(T)}\right\}\right).$$

When $X^{(1)}$ is the constant 1, the conditional expectation $E[\rho_1(Z;\theta,h(\cdot))|X^{(1)}]$ is simply the unconditional expectation $E[\rho_1(Z;\theta,h(\cdot))]$. Thus model (1) includes unconditional moment restrictions as a special case.

Model (1) includes many existing econometric models. First, it includes semiparametric and nonparametric panel data models where the information set expands over time. Second, it nests widely used semiparametric models that are estimated via two-stage plug-in procedures. For example, with $T=2$, $\theta=(\theta_1',\theta_2')'$, $X^{(1)}=1$ and $X=X^{(2)}$, model (1) becomes the following semiparametric "plug-in" problem:

$$E[\rho_1(Z;\theta_{o1},\theta_{o2},h_o(\cdot))] = 0 \text{ with } \dim(\rho_1)=\dim(\theta_1), \tag{2}$$

$$E[\rho_2(Z;\theta_{o2},h_o(\cdot))|X^{(2)}] = 0, \tag{3}$$

where the unknown parameter θ_{o2} and the unknown function $h_o(\cdot)$ can be estimated using the conditional moment restriction (3), and can then be plugged into the unconditional moment restriction (2) to compute the parameter θ_{o1}. An example of the plug-in problem is the estimation of a weighted average derivative of a possibly nonlinear nonparametric instrumental variables (IV) model:

$$\theta_o = E[W(Y_2)\nabla h_o(Y_2)], \quad E[\rho_2(Z;h_o(Y_2))|X^{(2)}]=0,$$

where $W()$ is a known positive weight function and $\nabla h_o()$ denotes the first derivative of h_o. Leading examples of the functional forms of $\rho_2(Z;h(Y_2))$ include $\rho_2(Z;h(Y_2))=Y_1-h(Y_2)$ for nonparametric mean IV regression and $\rho_2(Z;h(Y_2))=1\{Y_1\le h(Y_2)\}-0.5$ for nonparametric median IV regression. Note that our model (1) allows for more general plug-in problems where the unconditional moment restriction $E[\rho_1(Z;\theta_{o1},\theta_{o2},h_o(\cdot))]=0$ is overidentified for θ_{o1} (or $\dim(\rho_1)>\dim(\theta_1)$). Many semiparametric program evaluation models, semiparametric missing data models, choice-based sampling problems, some recent nonclassical measurement error models and some semiparametric control function models could also fit into framework (1).

[2] This sequential restriction on the sigma-fields allows us to gather all moment restrictions using the same set of conditioning variables into the same group. This grouping is convenient for our calculation of the efficiency bound.

When $T = 1$ and $X^{(1)} = X$, model (1) becomes the semiparametric conditional moment restriction model with the same conditioning information set: $E[\rho(Z;\theta_o, h_o(\cdot))|X] = 0$.[3] For this model, Chamberlain (1992b), Ai and Chen (2003) and Chen and Pouzo (2009) characterized the semiparametric efficiency bound of θ_o; and Ai and Chen (2003), Otsu (2008) and Chen and Pouzo (2009) considered efficient estimation of θ_o. However, these efficiency bound and efficient estimation results do not cover semiparametric models with sequential information sets like (1).

Without unknown functions $h_o(\cdot)$, model (1) becomes the one studied by Chamberlain (1992a), Hahn (1997) and Brown and Newey (1998). In particular, under the assumption that θ_o is identified by the model $E[\rho_t(Z;\theta_o)|X^{(t)}] = 0$, $t = 1, ..., T$, where $\{1\} \subseteq \sigma\left(\left\{X^{(1)}\right\}\right) \subset \cdots \subset \sigma\left(\left\{X^{(T)}\right\}\right)$, Chamberlain (1992a) established the semiparametric efficiency bound and Hahn (1997) obtained an efficient estimator of θ_o. Brown and Newey (1998) studied the semiparametric efficiency bound and suggested some efficient estimators of parameters that are defined as unconditional expectations; one of their models is $\theta_{o1} = E[g(Z, \theta_{o2})], E[\rho_2(Z;\theta_{o2})|X^{(2)}] = 0$ with known functional forms of g and ρ_2. In this paper we extend their results to the general model (1) that contains unknown functions $h_o(\cdot)$ which may depend on endogenous variables.

Ai and Chen (2007) studied consistent estimation of (θ_o, h_o) identified by the general model (1) with $T \geq 2$, and established the root-n asymptotic normality of their estimator of θ_o. But, to the best of our knowledge, there is no published work on the semiparametric efficiency bound or efficient estimation of θ_o for model (1), not even for the important special case of the semiparametric plug-in problem (2)-(3) when $h_o(\cdot)$ is an unknown function of endogenous variables. There are efficiency results for various special cases of the model (1) when the unknown function $h_o(\cdot)$ does not depend on endogenous variables. For example, Newey and Stoker (1993) computed the semiparametric efficiency bound and presented an efficient estimator for the weighted average derivative when the unknown $h_o(\cdot)$ is a measurable function of conditioning variables (say $X^{(2)}$ in terms of our notation) only. Our plug-in problem (2)-(3) extends their setup to allow for $h_o(\cdot)$ to be a function of endogenous variables. Since many economic structural models satisfy the sequential moment restrictions (1) with $h_o(\cdot)$ being an unknown function of Y, our extension is a useful one.

A key step in our calculation of the semiparametric efficiency bound for the general model (1) is to sequentially orthogonalize the residual functions $\rho_t(Z;\theta_o, h_o(\cdot))$ for $t = 1, ..., T$. The semiparametric efficiency bound is then obtained by optimally weighting the orthogonalized residual functions. When

[3] See Newey and Powell (2003), Chernozhukov, Imbens and Newey (2007), and Chen and Pouzo (2008a) for nonparametric estimation of this model when unknown $h_o(\cdot)$ depends on endogenous variables Y.

specialized to the models considered by Chamberlain (1987, 1992a, b), Brown and Newey (1998), Ai and Chen (2003) and Chen and Pouzo (2009), our semiparametric efficiency bound coincides with the bounds derived by these authors for their respective models. Although the efficiency bounds do not have explicit closed form expressions for general conditional moment models involving several unknown functions with different arguments, these bounds can be computed analytically for many specific sequential moment models containing only one unknown function. Our characterization of the efficiency bound is useful in evaluating and comparing different estimators for the general model (1). It also sheds lights on how to construct new estimators that attain the efficiency bound for θ_o. Specifically, it clearly suggests that any efficient estimation procedure has to account for both the correlation between the residual functions $\rho_t(Z;\theta_o,h_o(\cdot))$ as well as their conditional heteroskedasticity. For model (1) with nonsingular efficiency bound for θ_o, we provide an optimally weighted, orthogonalized, sieve minimum distance (SMD) estimator that is root-n asymptotically efficient for θ_o.

When applied to the semiparametric plug-in problem (2)-(3), our efficiency bound result reveals that one just needs to use the conditional moment model $E[\rho_2(Z;\theta_{o2},h_o(\cdot))|X^{(2)}] = 0$ alone to construct an efficient estimator of θ_{o2}. However, any efficient estimator of θ_{o1} has to account for the correlation between the two residuals $\rho_1(Z;\theta_{o1},\theta_{o2},h_o(\cdot))$ and $\rho_2(Z;\theta_{o2},h_o(\cdot))$ conditional on $X^{(2)}$. In particular, whenever $E[\rho_1(Z;\theta_{o1},\theta_{o2},h_o(\cdot))\rho_2(Z;\theta_{o2},h_o(\cdot))|X^{(2)}] \neq 0$, any simple plug-in estimator $\widehat{\theta}_1$, defined as a solution to $\frac{1}{n}\sum_{i=1}^{n}\rho_1(Z_i;\widehat{\theta}_1,\widehat{\theta}_2,\widehat{h}(\cdot)) = 0$, is not efficient for θ_{o1}, regardless of how one computes the first stage estimator $(\widehat{\theta}_2,\widehat{h}(\cdot))$. For example, to estimate the weighted average derivative $\theta_{o1} = E[W(Y_2)\nabla h_o(Y_2)]$ of a NPIV regression $E[Y_1 - h_o(Y_2)|X^{(2)}] = 0$, Ai and Chen (2007) presented a simple plug-in estimator $\widehat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n}W(Y_{2i})\nabla\widehat{h}(Y_{2i})$ where $\widehat{h}(\cdot)$ is a SMD estimator of h_o, and derived the root-n asymptotic normality of this estimator $\widehat{\theta}_1$. To compute a simple plug-in estimator of θ_{o1}, instead of using their SMD estimator of h_o, one could use other existing NPIV estimators, such as the estimators of Hall and Horowitz (2005), Darolles, Florens and Renault (2002), Blundell, Chen and Kristensen (2007). Unfortunately, since $E[W(Y_2)\nabla h_o(Y_2)\{Y_1 - h_o(Y_2)\}|X^{(2)}] \neq 0$, none of these simple plug-in estimators attain the semiparametric efficiency bound for θ_{o1}.

The rest of the paper is organized as follows: Section 2 first computes the semiparametric efficiency bound for θ_o in the general model (1), and then applies the bound result to the plug-in problem (2)-(3). Section 3 applies the efficiency bound result to several non-trivial examples, including estimating the weighted average derivative of $h_o(Y_2)$ in the NPIV regression $E[Y_1 - h_o(Y_2)|X^{(2)}] = 0$ or in the partially linear $\gamma-$quantile IV regression $E[1\{Y_1 \leq Y_2'\theta_{o2} + h_o(X_1)\}|X^{(2)}] = \gamma \in (0,1)$. Section 4 first discusses semiparametric efficient estimation of θ_o for the general model (1) and then presents a small Monte Carlo study to compare the inefficient simple plug-in estimator versus several efficient estimators of the average derivative of a NPIV regression. Section 5 concludes with suggestions of other efficient estimation procedures. All proofs are contained in the appendix.

2 Semiparametric Efficiency Bound

We begin by computing the semiparametric efficiency bound of θ_o. Our calculation of the bound closely follows the approach of Stein (1956), Newey (1990), Chamberlain (1992a, b) and others (see the appendix for details). We first sequentially orthogonalize the original sets of residual functions $\rho_t(Z;\theta_o,h_o(\cdot))$ for $t=1,...,T$. This procedure is called forward filtering by Hayashi and Sims (1983) in their study of linear time series rational expectation models, and has been used in Hansen, Heaton and Ogaki (1988), Chamberlain (1992a) and others in time series and panel data models without an unknown function $h_o(\cdot)$. In the following we denote $\Theta \subset \mathcal{R}^{d_\theta}$ as an open, finite-dimensional parameter space with $\theta_o \in \Theta$. Let $(\mathcal{H}, ||\cdot||_{\mathcal{H}})$ denote an infinite-dimensional metric space with $\mathcal{H}=\mathcal{H}_1\times\cdots\times\mathcal{H}_q$ and $h_o=(h_{o1},...,h_{oq})\in\mathcal{H}$, where h_{o1},..., h_{oq} are real-valued measurable functions. Let $\mathcal{A}=\Theta\times\mathcal{H}$ and $\alpha_o=(\theta_o',h_o)\in\mathcal{A}$. For any $\alpha=(\theta',h)\in\mathcal{A}$, we define

$$\begin{aligned}\varepsilon_T(Z,\alpha) &\equiv \rho_T(Z,\alpha),\\ \varepsilon_s(Z,\alpha) &\equiv \rho_s(Z,\alpha)-\sum_{t=s+1}^{T}\Gamma_{s,t}(X^{(t)})\varepsilon_t(Z,\alpha) \quad \text{for } s=T-1,...,1,\end{aligned}$$

where

$$\Gamma_{s,t}(X^{(t)})\equiv E[\rho_s(Z,\alpha_o)\varepsilon_t(Z,\alpha_o)'|X^{(t)}]\{\Sigma_{ot}(X^{(t)})\}^{-1} \quad \text{for } s<t$$

and

$$\Sigma_{ot}(X^{(t)})\equiv E[\varepsilon_t(Z,\alpha_o)\varepsilon_t(Z,\alpha_o)'|X^{(t)}].$$

For any $\alpha=(\theta',h)\in\mathcal{A}$, we denote

$$m_s(X^{(s)},\alpha)\equiv E\{\varepsilon_s(Z,\alpha)|X^{(s)}\} \quad \text{for } s=1,...,T.$$

We note that by construction

$$m_t(X^{(t)},\alpha_o)=E[\varepsilon_t(Z,\alpha_o)|X^{(t)}]=0 \quad \text{for } t=1,...,T,$$

and

$$E[\varepsilon_s(Z,\alpha_o)\varepsilon_t(Z,\alpha_o)'|X^{(t)}] = 0 \quad \text{for any } s < t.$$

This implies that

$$E\{\varepsilon_s(Z,\alpha_o)\varepsilon_t(Z,\alpha_o)v(X^{(s)})q(X^{(t)})\} = 0 \quad \text{for any } s \neq t \text{ and any measurable functions } v \text{ and } q.$$

Also, $\varepsilon_s(Z,\alpha) = \rho_s(Z,\alpha)$ (or $\Gamma_{s,t}(X^{(t)}) = 0$ for all $s < t$) if $\rho_s(Z,\alpha_o)$ is measurable with respect to the sigma field generated by $X^{(s+1)}$. Throughout the paper we assume the following conditions hold:

Assumption 1: (i) The data $\{Z_i = (Y_i', X_i')'\}_{i=1}^n$ is a random sample from an unknown distribution of Z on $\mathcal{Z}$; (ii) $\alpha_o = (\theta_o', h_o) \in \mathcal{A}$ satisfies model (1).

Let $\{\alpha(\tau) = (\theta(\tau), h(\tau)) : 0 \leq \tau \leq 1\} \subset \mathcal{A}$ be an arbitrarily smooth path in τ satisfying $\alpha(0) = \alpha_o$, $\alpha(1) = \alpha$, $\frac{d\theta(\tau)}{d\tau}|_{\tau=0} = \theta - \theta_o$, and $\frac{dh(\tau)}{d\tau}|_{\tau=0} = h - h_o \equiv \Delta h$.

Assumption 2: For $t = 1, ..., T$, $\frac{dm_t(X^{(t)},\alpha(\tau))}{d\tau}|_{\tau=0}$ is well-defined and has finite second moment.

For $t = 1, ..., T$ denote

$$\frac{dm_t(X^{(t)},\alpha_o)}{d\alpha}[\alpha - \alpha_o] = \frac{dE\{\varepsilon_t(Z,\alpha(\tau))|X^{(t)}\}}{d\tau}|_{\tau=0} = \frac{dm_t(X^{(t)},\alpha_o)}{d\theta'}(\theta - \theta_o) + \frac{dm_t(X^{(t)},\alpha_o)}{dh}[h - h_o]$$

where

$$\frac{dm_t(X^{(t)},\alpha_o)}{d\theta} = \frac{dE\{\varepsilon_t(Z,\theta,h_o)|X^{(t)}\}}{d\theta}|_{\theta=\theta_o}, \quad \frac{dm_t(X^{(t)},\alpha_o)}{dh}[h - h_o] = \frac{dE\{\varepsilon_t(Z,\theta_o,h(\tau))|X^{(t)}\}}{d\tau}|_{\tau=0}.$$

Assumption 3: For all $t = 1, ..., T$, (i) $\Sigma_{ot}(X^{(t)})$ is non-singular with probability one and (ii) $E\left\{\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta'}\right]' \Sigma_{ot}(X^{(t)})^{-1}\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta'}\right]\right\} < \infty$.

For any $h \in \mathcal{H}$, define a pseudo-metric $||h - h_o||$ as

$$||h - h_o||^2 \equiv \sum_{t=1}^{T} E\left\{\{\frac{dm_t(X^{(t)},\alpha_o)}{dh}[h - h_o]\}'\Sigma_{ot}(X^{(t)})^{-1}\{\frac{dm_t(X^{(t)},\alpha_o)}{dh}[h - h_o]\}\right\}. \tag{4}$$

Let $\overline{\mathcal{W}}$ denote the closed linear completion of $\mathcal{H} - \{h_o\}$ under $||\cdot||$. Denote $||x||_e^2 \equiv x'x$ for any vector

x. For each component θ^j (of θ), $j = 1, ..., d_\theta$, let $r_o^j \in \overline{\mathcal{W}}$ denote one solution to

$$\inf_{r^j \in \overline{\mathcal{W}}} \sum_{t=1}^{T} E\left\{ \left\| \{\Sigma_{ot}(X^{(t)})\}^{-\frac{1}{2}} \left[\frac{dm_t(X^{(t)}, \alpha_o)}{d\theta^j} - \frac{dm_t(X^{(t)}, \alpha_o)}{dh}[r^j] \right] \right\|_e^2 \right\}, \tag{5}$$

which solves: for all $r^j \in \overline{\mathcal{W}}$,

$$\sum_{t=1}^{T} E\left\{ \left[\frac{dm_t(X^{(t)}, \alpha_o)}{d\theta^j} - \frac{dm_t(X^{(t)}, \alpha_o)}{dh}[r_o^j] \right]' \Sigma_{ot}(X^{(t)})^{-1} \left[\frac{dm_t(X^{(t)}, \alpha_o)}{dh}[r^j] \right] \right\} = 0 \tag{6}$$

Let $r_o \equiv (r_o^1, ..., r_o^{d_\theta}) \in \prod_{j=1}^{d_\theta} \overline{\mathcal{W}}$, and $\frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o] \equiv (\frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o^1], ..., \frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o^{d_\theta}])$ be a $d_{\rho_t} \times d_\theta-$matrix valued measurable function of $X^{(t)}$. Denote

$$J_o \equiv \sum_{t=1}^{T} E\left\{ \left\| \{\Sigma_{ot}(X^{(t)})\}^{-\frac{1}{2}} \left[\frac{dm_t(X^{(t)}, \alpha_o)}{d\theta'} - \frac{dm_t(X^{(t)}, \alpha_o)}{dh}[r_o] \right] \right\|_e^2 \right\}.$$

Without further assumptions, the minimization problem (5) (or its normal equation (6)) may have several solutions in $\overline{\mathcal{W}}$. Nevertheless, the minimized criterion value is unique. Therefore J_o is always well-defined and unique. As shown in the appendix, J_o is actually the semiparametric Fisher information bound for θ_o.

Theorem 2.1 *Let Assumptions 1-3 and Assumption A in the appendix hold. (1) If J_o is singular, then θ_o can not be estimated at $\sqrt{n}-$rate. (2) If J_o is non-singular, then the semiparametric efficient variance bound for θ_o in model (1) is $\Omega_\theta^* = (J_o)^{-1}$.*

It is worth noting that Theorem 2.1 remains valid even if $\rho_t(Z; \alpha)$ is not pathwise differentiable but $m_t(X^{(t)}; \alpha)$ is pathwise differentiable in α_o. Thus our efficiency bound result applies to some nonsmooth problems such as a semi/nonparametric quantile regression with or without endogeneity.

Remark 2.1. (1). When specializing Theorem 2.1 to the case without unknown $h_o()$ in model (1):

$$E[\rho_t(Z; \theta_o)|X^{(t)}] = 0 \quad \text{for } t = 1, ..., T,$$

the semiparametric efficient variance bound for θ_o becomes $\Omega_\theta^* = (J_o)^{-1}$, where

$$J_o = \sum_{t=1}^{T} E\left\{ \left[\frac{dm_t(X^{(t)}, \theta_o)}{d\theta'} \right]' \Sigma_{ot}(X^{(t)})^{-1} \left[\frac{dm_t(X^{(t)}, \theta_o)}{d\theta'} \right] \right\},$$

which is the bound obtained in Chamberlain (1992a). If we further restrict the result to the original conditional moment restriction model: $E[\rho_1(Z;\theta_o)|X] = 0$ (i.e., the case of $T = 1$ and $X^{(T)} = X$), the semiparametric efficient variance bound for θ_o becomes

$$\left(E\left\{ \left[\frac{dm_1(X,\theta_o)}{d\theta'}\right]' \Sigma_{o1}(X)^{-1} \left[\frac{dm_1(X,\theta_o)}{d\theta'}\right] \right\} \right)^{-1} \quad \text{with } m_1(X,\theta) = E\{\rho_1(Z;\theta)|X\},$$

which is the bound derived in Chamberlain (1987). For the unconditional moment restriction model (i.e., $T = 1$ and $X^1 = 1$): $E[\rho_1(Z;\theta_o)] = 0$, the semiparametric efficient variance bound for θ_o becomes

$$\left(E\left\{ \left[\frac{dm_1(\theta_o)}{d\theta'}\right]' \Sigma_{o1}^{-1} \left[\frac{dm_1(\theta_o)}{d\theta'}\right] \right\} \right)^{-1} \quad \text{with } m_1(\theta) = E\{\rho_1(Z;\theta)\},$$

which is the bound obtained in Hansen (1982) (specialized to i.i.d. data).

(2). When specializing Theorem 2.1 to the case of $T = 1$ and $X^{(T)} = X$ in model (1):

$$E[\rho_1(Z;\theta_o,h_o())|X] = 0,$$

the semiparametric efficient variance bound for θ_o becomes $\Omega_\theta^* = (J_o)^{-1}$, where

$$J_o = E\left\{ \left[\frac{dm_1(X,\alpha_o)}{d\theta'} - \frac{dm_1(X,\alpha_o)}{dh}[r_o]\right]' \Sigma_{o1}(X)^{-1} \left[\frac{dm_1(X,\alpha_o)}{d\theta'} - \frac{dm_1(X,\alpha_o)}{dh}[r_o]\right] \right\}$$

with $m_1(X,\alpha) = E[\rho_1(Z;\theta,h())|X]$ and $\Sigma_{o1}(X) = Var\{\rho_1(Z;\theta_o,h_o())|X\}$. This recovers the semiparametric efficiency bound obtained by Chamberlain (1992b) and Ai and Chen (2003) for pathwise differentiable $\rho_1(Z;\theta,h())$ in α_o, and by Chen and Pouzo (2009) for possible non-pathwise differentiable $\rho_1(Z;\theta,h())$ in α_o.

2.1 Plug-in problem

We now apply Theorem 2.1 to the plug-in problem. We shall replace Assumption 1 by

Assumption 1s: (i) Assumption 1(i) holds; (ii) $\alpha_o = (\theta_o', h_o(\cdot))$ satisfies model (2)-(3) and $\frac{dE\{\rho_1(Z;\alpha_o)\}}{d\theta_1}$ has full rank $d_{\theta_1} \equiv \dim(\theta_1)$.

We first present the semiparametric efficiency bound of θ_{o2} for model (2)-(3). Recall that for this model $m_2(X^{(2)},\alpha) = E[\rho_2(Z,\alpha)|X^{(2)}]$ and $\Sigma_{o2}(X^{(2)}) = E[\rho_2(Z,\alpha_o)\rho_2(Z,\alpha_o)'|X^{(2)}]$. For each component θ_2^j (of θ_2), $j = 1, ..., d_{\theta_2}$, let $w_{o2}^j \in \overline{\mathcal{W}}$ denote one solution to

$$\inf_{w^j \in \overline{\mathcal{W}}} E\left\{\left\|\{\Sigma_{o2}(X^{(2)})\}^{-\frac{1}{2}}\left[\frac{dm_2(X^{(2)},\alpha_o)}{d\theta_2^j} - \frac{dm_2(X^{(2)},\alpha_o)}{dh}[w^j]\right]\right\|_e^2\right\}, \tag{7}$$

which solves for all $w^j \in \overline{\mathcal{W}}$:

$$E\left\{\left[\frac{dm_2(X^{(2)},\alpha_o)}{d\theta_2^j} - \frac{dm_2(X^{(2)},\alpha_o)}{dh}[w_{o2}^j]\right]' \Sigma_{o2}(X^{(2)})^{-1}\left[\frac{dm_2(X^{(2)},\alpha_o)}{dh}[w^j]\right]\right\} = 0.$$

Let $w_{o2} \equiv (w_{o2}^1, ..., w_{o2}^{d_{\theta_2}}) \in \prod_{j=1}^{d_{\theta_2}} \overline{\mathcal{W}}$, and $\frac{dm_2(X^{(2)},\alpha_o)}{dh}[w_{o2}] \equiv (\frac{dm_2(X^{(2)},\alpha_o)}{dh}[w_{o2}^1], ..., \frac{dm_2(X^{(2)},\alpha_o)}{dh}[w_{o2}^{d_{\theta_2}}])$ be a $d_{\rho_2} \times d_{\theta_2}-$matrix valued measurable function of $X^{(2)}$. Denote

$$J_{o\theta_2} \equiv E\left\{\left\|\{\Sigma_{o2}(X^{(2)})\}^{-\frac{1}{2}}\left[\frac{dm_2(X^{(2)},\alpha_o)}{d\theta_2'} - \frac{dm_2(X^{(2)},\alpha_o)}{dh}[w_{o2}]\right]\right\|_e^2\right\}.$$

As shown in the appendix, $J_{o\theta_2}$ is the semiparametric Fisher information bound for θ_{o2} in model (2)-(3).

Theorem 2.2 *Let Assumptions 1s, 2, 3 and assumption A in the appendix hold. (1) If $J_{o\theta_2}$ is singular, then θ_{o2} can not be estimated at $\sqrt{n}-$rate. (2) If $J_{o\theta_2}$ is non-singular, then the semiparametric efficient variance bound for θ_{o2} in model (2)-(3) is $\Omega_{\theta_2}^* = (J_{o\theta_2})^{-1}$.*

The next result provides the efficiency bound of θ_{o1}. Recall that for model (2)-(3), $\varepsilon_1(Z,\alpha) = \rho_1(Z;\alpha) - \Gamma_{1,2}(X^{(2)})\rho_2(Z;\alpha)$ where $\Gamma_{1,2}(X^{(2)}) = E[\rho_1(Z;\alpha_o)\rho_2(Z;\alpha_o)'|X^{(2)}]\{\Sigma_{o2}(X^{(2)})\}^{-1}$. When $X^{(1)} = 1$ we denote $m_1(\alpha) = E[\varepsilon_1(Z,\alpha)]$, $\frac{dm_1(\alpha_o)}{d\theta} = \frac{dE\{\varepsilon_1(Z,\theta,h_o)\}}{d\theta}|_{\theta=\theta_o}$, $\frac{dm_1(\alpha_o)}{dh}[h-h_o] = \frac{dE\{\varepsilon_1(Z,\theta_o,h(\tau))\}}{d\tau}|_{\tau=0}$ and $\Sigma_{o1} = E[\varepsilon_1(Z,\alpha_o)\varepsilon_1(Z,\alpha_o)']$. For each component θ_1^k (of θ_1), $k = 1, ..., d_{\theta_1}$, let $r_{o1}^k \in \overline{\mathcal{W}}$ denote one solution to

$$\inf_{r^k \in \overline{\mathcal{W}}} E\left\{\left\|\{\Sigma_{o1}\}^{-\frac{1}{2}}\left[\frac{dm_1(\alpha_o)}{d\theta_1^k} - \frac{dm_1(\alpha_o)}{dh}[r^k]\right]\right\|_e^2 + \left\|\{\Sigma_{o2}(X^{(2)})\}^{-\frac{1}{2}}\frac{dm_2(X^{(2)},\alpha_o)}{dh}[r^k]\right\|_e^2\right\}. \tag{8}$$

Let $r_{o\theta_1} \equiv (r_{o1}^1, ..., r_{o1}^{d_{\theta_1}}) \in \prod_{k=1}^{d_{\theta_1}} \overline{\mathcal{W}}$, and $\frac{dm_1(\alpha_o)}{dh}[r_{o\theta_1}] \equiv (\frac{dm_1(\alpha_o)}{dh}[r_{o1}^1], ..., \frac{dm_1(\alpha_o)}{dh}[r_{o1}^{d_{\theta_1}}])$ be a $d_{\rho_1} \times$

$d_{\theta_1}-$matrix of constants. Denote

$$E[S^*_{\theta_1} S^{*\prime}_{\theta_1}] \equiv E\left\{ \left\| \{\Sigma_{o1}\}^{-\frac{1}{2}} \left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{dh}[r_{o\theta_1}] \right] \right\|_e^2 + \left\| \{\Sigma_{o2}(X^{(2)})\}^{-\frac{1}{2}} \frac{dm_2(X^{(2)},\alpha_o)}{dh}[r_{o\theta_1}] \right\|_e^2 \right\}.$$

Similarly we define

$$J_{o\theta_1} \equiv \inf_{(b,w)\in\prod_{j=1}^{d_{\theta_1}}(\mathcal{R}^{d_{\theta_2}}\times\overline{\mathcal{W}})} E\left\{ \begin{array}{c} \left\| \{\Sigma_{o1}\}^{-\frac{1}{2}} \left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{d\theta_2'} b - \frac{dm_1(\alpha_o)}{dh}[w] \right] \right\|_e^2 \\ + \left\| \{\Sigma_{o2}(X^{(2)})\}^{-\frac{1}{2}} \left[\frac{dm_2(X^{(2)},\alpha_o)}{d\theta_2'} b + \frac{dm_2(X^{(2)},\alpha_o)}{dh}[w] \right] \right\|_e^2 \end{array} \right\}.$$

Theorem 2.3 *Let Assumptions 1s, 2, 3 and Assumption A in the appendix hold. (1) If $J_{o\theta_1}$ is singular, then θ_{o1} can not be estimated at $\sqrt{n}-$rate. (2) If $J_{o\theta_1}$ is non-singular, then the semiparametric efficient variance bound for θ_{o1} in model (2)-(3) is $\Omega^*_{\theta_1} = (J_{o\theta_1})^{-1}$. (3) If both $J_{o\theta_1}$ and $J_{o\theta_2}$ are non-singular, then another expression for the semiparametric efficient variance bound for θ_{o1} in model (2)-(3) is $\Omega^*_{\theta_1} = \left(E[S^*_{\theta_1} S^{*\prime}_{\theta_1}]\right)^{-1} + a^*\Omega^*_{\theta_2} a^{*\prime}$, with*

$$a^* = \left[\frac{dm_1(\alpha_o)}{d\theta_1'}\right]^{-1} \left[\frac{dm_1(\alpha_o)}{d\theta_2'} - \frac{dm_1(\alpha_o)}{dh}[w_{o2}]\right] \tag{9}$$

where $w_{o2} \equiv (w^1_{o2}, ..., w^{d_{\theta_2}}_{o2}) \in \prod_{j=1}^{d_{\theta_2}} \overline{\mathcal{W}}$ solves (7).

Remark 2.2: When applying Theorems 2.2 and 2.3 to the special case

$$\theta_{o1} = E[g(Z;\theta_{o2})], \quad E[\rho_2(Z;\theta_{o2})|X^{(2)}] = 0,$$

where the functional forms of $g()$ and $\rho_2()$ are known up to an unknown θ_{o2}, we obtain

$$\Omega^*_{\theta_2} = \left(E\left\{ \left[\frac{dE[\rho_2(Z;\theta_{o2})|X^{(2)}]}{d\theta_2'}\right]' \Sigma_{o2}(X^{(2)})^{-1} \left[\frac{dE[\rho_2(Z;\theta_{o2})|X^{(2)}]}{d\theta_2'}\right] \right\} \right)^{-1},$$

and

$$\varepsilon_1(Z,\theta) = g(Z;\theta_2) - E[g(Z;\theta_{o2})\rho_2(Z;\theta_{o2})'|X^{(2)}]\{\Sigma_{o2}(X^{(2)})\}^{-1}\rho_2(Z;\theta_2) - \theta_1,$$
$$\Omega^*_{\theta_1} = E[\varepsilon_1(Z,\theta_o)\varepsilon_1(Z,\theta_o)'] + \frac{dE\{\varepsilon_1(Z,\theta_o)\}}{d\theta_2'}\Omega^*_{\theta_2}\frac{dE\{\varepsilon_1(Z,\theta_o)\}'}{d\theta_2}.$$

We note that the semiparametric efficiency bound for $\theta_{o1} = E[g(Z;\theta_{o2})]$ coincides with that derived in Brown and Newey (1998).

3 Examples

For the general model (1) involving several unknown functions $h_{oj}(\cdot)$, $j = 1, ..., q$ with different arguments, our efficiency bound is characterized in a variational form,[4] however it can be solved explicitly for many examples of model (1) involving only one unknown function ($q = 1$). In this section we present several such examples.

Example 3.1: *Weighted average derivatives in a partially linear IV regression*:

$$\theta_{o1} = E[W(X_1)\{\nabla^s h_o(X_1)\}], \quad E[Y_1 - Y_2'\theta_{o2} - h_o(X_1)|X^{(2)}] = 0, \quad X_1 \subset X^{(2)},$$

where $X^{(1)} = 1$, $W(X_1)$ is a known scalar positive weight function, and Y_1 is a scalar. For this example, we have $\varepsilon_2(Z,\alpha) = \rho_2(Z,\alpha) = Y_1 - Y_2'\theta_2 - h(X_1)$, $\Sigma_{o2}(X^{(2)}) = E[\{Y_1 - Y_2'\theta_{o2} - h_o(X_1)\}^2|X^{(2)}]$, $m_2(X^{(2)},\alpha) = E[Y_1 - Y_2'\theta_2|X^{(2)}] - h(X_1)$, $\frac{dm_2(X^{(2)},\alpha_o)}{d\theta_2'} = -E[Y_2'|X^{(2)}]$ and $\frac{dm_2(X^{(2)},\alpha_o)}{dh}[r] = -r(X_1)$. Denote

$$w_{o2}(X_1) = \left(E[\Sigma_{o2}(X^{(2)})^{-1}|X_1]\right)^{-1} E\left[\{\Sigma_{o2}(X^{(2)})^{-1}E[Y_2'|X^{(2)}]\}|X_1\right].$$

We impose the following condition:

Condition 3.1: (i) $0 < \Sigma_{o2}(X^{(2)}) = E[\{Y_1 - Y_2'\theta_{o2} - h_o(X_1)\}^2|X^{(2)}] < \infty$, $0 < E[\Sigma_{o2}(X^{(2)})^{-1}|X_1] < \infty$; (ii) $E\left\{E[Y_2|X^{(2)}]\{\Sigma_{o2}(X^{(2)})\}^{-1}E[Y_2'|X^{(2)}]\right\} < \infty$, $E\left\{w_{o2}(X_1)'\{\Sigma_{o2}(X^{(2)})\}^{-1}w_{o2}(X_1)\right\} < \infty$; (iii) $E[Y_2|X^{(2)}]$ is not a measurable function of X_1.

Under Condition 3.1(i)(ii)(iii), the semiparametric Fisher information bound for θ_{o2}, as given by

$$J_{o\theta_2} = \inf_w E\left\{\left[-E[Y_2'|X^{(2)}] + w(X_1)\right]' \{\Sigma_{o2}(X^{(2)})\}^{-1}\left[-E[Y_2'|X^{(2)}] + w(X_1)\right]\right\},$$

is nonsingular, and has a unique minimizer $w_{o2}(X_1)$. Applying Theorem 2.2, we immediately obtain:

Proposition 3.1 (1) *Under Condition 3.1(i)(ii)(iii), $J_{o\theta_2}$ is non-singular, and the semiparametric*

[4] This is not a defect associated with our method of deriving efficiency bounds, but is due to the complexity of the model involving multiple unknown functions of different (and possibly endogenous) arguments. Even for the special case of the conditional moment model of common conditioning set: $E[\rho(Z;\theta_o, h_{o1}(X_1), ..., h_{oq}(X_q))|X] = 0$ with pointwise smooth $\rho()$ and $\{X_1, ..., X_q\} \subseteq \{X\}$, Chamberlain (1992b) points out that, when $q > 1$, the bound in an explicit form is no longer available and only characterizes his bound in a variational form.

efficient variance bound for θ_{o2} of Example 3.1 model is: $\Omega^*_{\theta_2} = (J_{o\theta_2})^{-1} =$

$$\left(E\left\{\left\|\{\Sigma_{o2}(X^{(2)})\}^{-\frac{1}{2}}\left[E[Y_2'|X^{(2)}] - \left(E[\Sigma_{o2}(X^{(2)})^{-1}|X_1]\right)^{-1} E\left[\{\Sigma_{o2}(X^{(2)})^{-1}E[Y_2'|X^{(2)}]\}|X_1\right]\right]\right\|_e^2\right\}\right)^{-1}.$$

Next since $\rho_1(Z,\alpha) = W(X_1)\nabla^s h(X_1) - \theta_1$ and $X_1 \subset X^{(2)}$, we have $E\{\rho_1(Z,\alpha_o)\rho_2(Z,\alpha_o)|X^{(2)}\} = 0$. Thus $\Gamma_{1,2}(X^{(2)}) = 0$, $\varepsilon_1(Z,\alpha) = \rho_1(Z,\alpha)$, $\Sigma_{o1} = Var\{W(X_1)\nabla^s h_o(X_1) - \theta_{o1}\}$, $m_1(\alpha) = E[W(X_1)\nabla^s h(X_1) - \theta_1]$, $\frac{dm_1(\alpha_o)}{d\theta_1'} = -I_{\theta_1}$, $\frac{dm_1(\alpha_o)}{d\theta_2'} = 0$ and $\frac{dm_1(\alpha_o)}{dh}[w_{o2}] = E[W(X_1)\nabla^s w_{o2}(X_1)]$. Hence $a^* = E[W(X_1)\nabla^s w_{o2}(X_1)]$.

Suppose that X_1 has a probability density $f_1(X_1)$ such that $W(x_1)f_1(x_1)$ goes to zero smoothly at the boundary of the support of X_1. Denote $l_s(X_1) \equiv \frac{\nabla^s[W(X_1)f_1(X_1)]}{f_1(X_1)}$. We impose the following condition:

Condition 3.1: (iv) $[W(x_1)f_1(x_1)]$ is $s-$times continuously differentiable and is zero on the boundary of the support of X_1; (v) $\Sigma_{o1} = Var\left(\{W(X_1)\nabla^s h_o(X_1) - \theta_{o1}\}\right)$ and $E[l_s(X_1)l_s(X_1)'(E[\Sigma_{o2}(X^{(2)})^{-1}|X_1])^{-1}]$ are finite, positive definite; (vi) $E[W(X_1)\nabla^s w_{o2}(X_1)] = (-1)^s E\{l_s(X_1)w_{o2}(X_1)\}$ exists.

We next compute one solution $r_{o\theta_1} \equiv (r^1_{o1}, ..., r^{d_{\theta_1}}_{o1})$ to

$$\begin{aligned}
& E[S^*_{\theta_1}S^{*\prime}_{\theta_1}] \\
= & \inf_r \left([I_{\theta_1} + E\{W(X_1)\nabla^s r(X_1)\}]' \Sigma_{o1}^{-1}[I_{\theta_1} + E\{W(X_1)\nabla^s r(X_1)\}] + E\left\{r(X_1)'\Sigma_{o2}(X^{(2)})^{-1}r(X_1)\right\}\right) \\
= & [I_{\theta_1} + E\{W(X_1)\nabla^s r_{o\theta_1}(X_1)\}]' \Sigma_{o1}^{-1}.
\end{aligned}$$

Applying integration-by-parts, $E\{W(X_1)\nabla^s r(X_1)\} = (-1)^s E\{l_s(X_1)r(X_1)\}$. Then by calculus variation, any solution $r_{o\theta_1}$ should satisfy

$$[I_{\theta_1} + (-1)^s E\{l_s(X_1)r_{o\theta_1}(X_1)\}]' \Sigma_{o1}^{-1} E\{(-1)^s l_s(X_1)r(X_1)\} = -E[r_{o\theta_1}(X_1)'\Sigma_{o2}(X^{(2)})^{-1}r(X_1)]$$

for all square measurable functions $r(X_1)$ such that all the expectations in the above equation are defined. This implies that $r_{o\theta_1}(X_1)$ solves

$$[I_{\theta_1} + (-1)^s E\{l_s(X_1)r_{o\theta_1}(X_1)\}]' \Sigma_{o1}^{-1}(-1)^s l_s(X_1) + r_{o\theta_1}(X_1)'E[\Sigma_{o2}(X^{(2)})^{-1}|X_1] = 0,$$

and thus

$$\{E[S^*_{\theta_1}S^{*\prime}_{\theta_1}]\}^{-1} = \Sigma_{o1} + E[l_s(X_1)l_s(X_1)'(E[\Sigma_{o2}(X^{(2)})^{-1}|X_1])^{-1}].$$

Applying Theorem 2.3, we immediately obtain:

Proposition 3.1 (2) *Under Condition 3.1(i)(ii)(iii)(iv)(v), $J_{o\theta_1}$ is non-singular, and the semiparametric efficient variance bound for θ_{o1} of Example 3.1 model is*:

$$\Omega^*_{\theta_1} = \{E[S^*_{\theta_1} S^{*\prime}_{\theta_1}]\}^{-1} + E[W(X_1)\nabla^s w_{o2}(X_1)]\Omega^*_{\theta_2} E[W(X_1)\nabla^s w_{o2}(X_1)]'.$$

Remark 3.1: When we specialize Example 3.1 to the case

$$\theta_{o1} = E[W(X_1)\nabla^s h_o(X_1)], \quad E[Y_1 - h_o(X_1)|X_1] = 0,$$

we have $X^{(2)} = X_1$, and the semiparametric efficient variance bound of θ_{o1} becomes

$$\Omega^*_{\theta_1} = \{E[S^*_{\theta_1} S^{*\prime}_{\theta_1}]\}^{-1} = \Sigma_{o1} + E[l_s(X_1)l_s(X_1)'\Sigma_{o2}(X_1)],$$

which coincides with the efficient variance bound of the weighted average first-derivative ($s = 1$) parameter in Newey and Stoker (1993, p.1205, equation (3.8)).

Example 3.2: *Weighted average derivatives in a partially linear quantile IV regression*:

$$\theta_{o1} = E[W(X_1)\{\nabla^s h_o(X_1)\}], \quad E[1\{Y_1 \leq Y_2'\theta_{o2} + h_o(X_1)\}|X^{(2)}] = \gamma \in (0,1), \quad X_1 \subset X^{(2)}.$$

For this example, we have $\varepsilon_2(Z,\alpha) = \rho_2(Z,\alpha) = 1\{Y_1 \leq Y_2'\theta_2 + h(X_1)\} - \gamma$, $\Sigma_{o2}(X^{(2)}) = \gamma(1-\gamma)$ and $m_2(X^{(2)},\alpha) = E[F_{Y_1|Y_2,X^{(2)}}(Y_2'\theta_2 + h(X_1))|X^{(2)}]$. Denote $U = Y_1 - Y_2'\theta_{o2} - h_o(X_1)$ and

$$w_{o2}(X_1) = \frac{E\{E[f_{U|Y_2,X^{(2)}}(0)Y_2'|X^{(2)}]E[f_{U|Y_2,X^{(2)}}(0)|X^{(2)}]|X_1\}}{E\{(E[f_{U|Y_2,X^{(2)}}(0)|X^{(2)}])^2|X_1\}}.$$

We impose the following condition:

Condition 3.2: (i) $E\left(E[f_{U|Y_2,X^{(2)}}(0)Y_2|X^{(2)}]E[f_{U|Y_2,X^{(2)}}(0)Y_2'|X^{(2)}]\right) < \infty$, $E\{w_{o2}(X_1)'w_{o2}(X_1)\} < \infty$; (ii) if $E[f_{U|Y_2,X^{(2)}}(0)[Y_2' - w_{o2}(X_1)]|X^{(2)}] \times a = 0$ a.s. then $a = 0$.

Then

$$J_{o\theta_2} = \inf_w E\left\{E[f_{U|Y_2,X^{(2)}}(0)[Y_2' - w(X_1)]|X^{(2)}]'E[f_{U|Y_2,X^{(2)}}(0)[Y_2' - w(X_1)]|X^{(2)}]\right\}[\gamma(1-\gamma)]^{-1},$$

is nonsingular, and has a unique minimizer $w_{o2}(X_1)$. Applying Theorem 2.2, we immediately obtain that the efficient variance bound for θ_{o2} is $\Omega^*_{\theta_2} = (J_{o\theta_2})^{-1}$.

Condition 3.2: (iii) $[W(x_1)f_1(x_1)]$ is $s-$times continuously differentiable and is zero on the boundary of the support of X_1; (iv) $E[W(X_1)\nabla^s w_{o2}(X_1)] = (-1)^s E\{l_s(X_1)w_{o2}(X_1)\}$ exists; (v) $\Sigma_{o1} = Var\left(\{W(X_1)\nabla^s h_o(X_1) - \theta_{o1}\}\right)$ and $E[l_s(X_1)l_s(X_1)'(E[\{E[f_{U|Y_2,X^{(2)}}(0)|X^{(2)}]\}^2|X_1])^{-1}]$ are finite, positive definite.

Under Condition 3.2,

$$
\begin{aligned}
& E[S^*_{\theta_1}S^{*\prime}_{\theta_1}] \\
= & \inf_r \begin{pmatrix} [I_{\theta_1} + E\{W(X_1)\nabla^s r(X_1)\}]' \, \Sigma_{o1}^{-1} \, [I_{\theta_1} + E\{W(X_1)\nabla^s r(X_1)\}] \\ +[\gamma(1-\gamma)]^{-1} E\left\{r(X_1)'\{E[f_{U|Y_2,X^{(2)}}(0)|X^{(2)}]\}^2 r(X_1)\right\} \end{pmatrix} \\
= & [I_{\theta_1} + E\{W(X_1)\nabla^s r_{o\theta_1}(X_1)\}]' \, \Sigma_{o1}^{-1}.
\end{aligned}
$$

is nonsingular, and has a unique minimizer $r_{o\theta_1}(X_1)$. Applying Theorem 2.3, the efficient variance bound for θ_{o1} is

$$
\Omega^*_{\theta_1} = \{E[S^*_{\theta_1}S^{*\prime}_{\theta_1}]\}^{-1} + E[W(X_1)\nabla^s w_{o2}(X_1)]\Omega^*_{\theta_2} E[W(X_1)\nabla^s w_{o2}(X_1)]',
$$

where now

$$
\{E[S^*_{\theta_1}S^{*\prime}_{\theta_1}]\}^{-1} = \Sigma_{o1} + \gamma(1-\gamma)E[l_s(X_1)l_s(X_1)'(E[\{E[f_{U|Y_2,X^{(2)}}(0)|X^{(2)}]\}^2|X_1])^{-1}].
$$

Example 3.3: *Weighted average derivatives of a nonparametric IV regression*

$$
\theta_o = E\left\{W(Y_2)\nabla^s h_o(Y_2)\right\}, \quad E\{Y_1 - h_o(Y_2)|X\} = 0,
$$

where Y_1, Y_2 and X are scalars, $X^{(1)} = 1$, $X^{(2)} = X$, and $W(Y_2)$ is a known positive weight function. For this example, we have $\varepsilon_2(Z,\alpha) = \rho_2(Z,\alpha) = Y_1 - h(Y_2)$ and $m_2(X^{(2)},\alpha) = E[Y_1 - h(Y_2)|X]$. Also, $\rho_1(Z,\alpha) = W(Y_2)\nabla^s h(Y_2) - \theta$, $\Gamma_{1,2}(X) = E[W(Y_2)\nabla^s h_o(Y_2)\{Y_1 - h_o(Y_2)\}|X]\{\Sigma_{o2}(X)\}^{-1}$ and $m_1(\alpha) = E[\{W(Y_2)\nabla^s h(Y_2) - \theta\} - \Gamma_{1,2}(X)\{Y_1 - h(Y_2)\}]$.

Let $f(X,Y_2)$ denote the joint density of (X,Y_2), and $f_2(Y_2)$ denote the marginal density of Y_2. Let K be the conditional expectation operator of Y_2 given X (i.e., $Kh \equiv E[h(Y_2)|X]$ for any measurable function h with $E\{[h(Y_2)]^2\} < \infty$), and K^* be the adjoint of K (i.e., $K^*g \equiv E[g(X)|Y_2]$ for any measurable function g with $E\{[g(X)]^2\} < \infty$). Denote $\|h\|_{Y_2} \equiv \sqrt{E\{h(Y_2)\}^2}$ and $l_s(Y_2) \equiv \frac{\nabla^s[W(Y_2)f_2(Y_2)]}{f_2(Y_2)}$.

We impose the following condition:

Condition 3.3: (i) $[W(Y_2)f_2(Y_2)]$ is $s-$times continuously differentiable and is zero on the boundary of the support of Y_2; (ii) $Kh \equiv E[h(Y_2)|X] = 0$ implies $h \equiv 0$;

(iii) $0 < \Sigma_{o1} = Var\left(\{W(Y_2)\nabla^s h_o(Y_2) - \theta_o\} - \Gamma_{1,2}(X)\{Y_1 - h_o(Y_2)\}\right) < \infty$; (iv) $0 < \Sigma_{o2}(X) = E\{[Y_1 - h_o(Y_2)]^2|X\} < \infty$; (v) $E\left(\{[K^*\Sigma_{o2}^{-1}K]^{-1/2}[(-1)^s l_s + K^*\Gamma_{1,2}](Y_2)\}^2\right) < \infty$.

Then:

$$\begin{aligned} J_o &= \inf_{r \in \overline{\mathcal{W}}} E \left\{ \begin{array}{c} [1 + E\{W(Y_2)\nabla^s r(Y_2) + \Gamma_{1,2}(X) r(Y_2)\}]^2 \{\Sigma_{o1}\}^{-1} \\ +\{\{E[r(Y_2)|X]\}^2 \Sigma_{o2}^{-1}(X)\} \end{array} \right\} \\ &= [1 + E\{[(-1)^s l_s + \Gamma_{1,2}(X)] r_o(Y_2)\}] \{\Sigma_{o1}\}^{-1}, \end{aligned}$$

Condition 3.3(iv)(v) can be replaced by the following sufficient assumptions:

Condition 3.3s: (i) K is compact with a singular-value system $\{\mu_j, q_j(Y_2), p_j(X)\}_{j=1}^{\infty}$ (i.e., $1 = \mu_1 \geq \mu_j \geq \mu_{j+1} \searrow 0$, $K^*Kq_j = \mu_j^2 q_j$ and $KK^*p_j = \mu_j^2 p_j$ for all $j \geq 1$); (ii) $\Sigma_{o2}(X) = \Sigma_{o2}$ a positive finite constant; (iii) $E\{[\Gamma_{1,2}(X)]^2\} < \infty$, $\sum_{j=1}^{\infty} \left(\frac{E\{l_s(Y_2)q_j(Y_2)\}}{\mu_j}\right)^2 < \infty$.

Applying Theorem 2.1 (the verification is very similar to that of example 2.2 in Ai and Chen (2007); hence we omit it), we obtain:

Proposition 3.3 *Under Condition 3.3, the semiparametric efficient variance bound for θ_o of Example 3.3 model is: $\Omega_\theta^* = (J_o)^{-1}$, with*

$$J_o = \frac{\{\Sigma_{o1}\}^{-1}}{1 + \left\|[K^*\Sigma_{o2}^{-1}K]^{-1/2}\{[(-1)^s l_s + K^*\Gamma_{1,2}]\}\right\|_{Y_2}^2 \Sigma_{o1}^{-1}} > 0.$$

Under Condition 3.3(i)(ii)(iii) and condition 3.3s, we have:

$$\Omega_\theta^* = \Sigma_{o1} + \Sigma_{o2} \sum_{j=1}^{\infty} \left\{ (-1)^s \frac{E\{l_s(Y_2)q_j(Y_2)\}}{\mu_j} + E\{\Gamma_{1,2}(X)p_j(X)\} \right\}^2 < \infty.$$

Remark 3.2: Notice that under condition 3.3(v), solutions $r_o(Y_2)$ are not unique, but the Fisher information bound J_o is unique and non-singular. If we strengthen condition 3.3(v) to condition 3.3(v)':

$$E\left(\{[K^*\Sigma_{o2}^{-1}K]^{-1}[(-1)^s l_s + K^*\Gamma_{1,2}](Y_2)\}^2\right) < \infty$$

then we obtain a unique solution $r_o(Y_2)$:

$$r_o(Y_2) = -\frac{[K^*\Sigma_{o2}^{-1}K]^{-1}[(-1)^s l_s + K^*\Gamma_{1,2}](Y_2)\Sigma_{o1}^{-1}}{1 + \left\|[K^*\Sigma_{o2}^{-1}K]^{-1/2}\{[(-1)^s l_s + K^*\Gamma_{1,2}]\}\right\|_{Y_2}^2 \Sigma_{o1}^{-1}}.$$

Remark 3.3: Condition 3.3(v) or Condition 3.3s(iii) imposes smoothness restrictions on $l_s(Y_2)$ relative to the smoothness of the operator K^*K. If Condition 3.3(v) is not satisfied, or under conditional homoskedastic error (i.e., Condition 3.3s(ii)), if Condition 3.3s(iii) is not satisfied, then V_o will be singular, and θ_o can not be estimated at the $\sqrt{n}-$rate. Ai and Chen (2007) discuss Condition 3.3s(iii), and point out that Condition 3.3s(iii) can still be satisfied even when the singular values μ_j of the conditional expectation operator K decay to zero exponentially fast. For example, consider the special case of $\theta_o = E[\nabla h_o(Y_2)]$ and assume that the conditional density of Y_2 given X is normal. For this case $l_s(Y_2) = \nabla\{\log f_2(Y_2)\}$ and $\mu_j \asymp \exp(-j)$, Condition 3.3s(iii) is satisfied.

Remark 3.4: When we specialize this example 3.3 to the case of no endogeneity (i.e., $Y_2 = X$):

$$\theta_o = E[W(Y_2)\nabla^s h_o(Y_2)], \quad E[Y_1 - h_o(Y_2)|Y_2] = 0,$$

we have $\Gamma_{1,2}(X) = 0$ and $K = K^* = identity$, and the semiparametrically efficient variance bound of θ_o becomes

$$\Omega_\theta^* = Var\left(\{W(X)\nabla^s h_o(X) - \theta_o\}\right) + E[l_s(X)l_s(X)'\Sigma_{o2}(X)].$$

4 Optimally Weighted Orthogonalized SMD Estimation

Our efficiency bound characterization for model (1) suggests that any efficient estimator has to account for both the correlation between the residual functions $\rho_t(Z;\theta_o,h_o(\cdot))$ as well as their conditional heteroskedasticity. In this section we propose an optimally weighted SMD procedure that is based on orthogonalized moment conditions. This procedure automatically leads to semiparametric efficiency regardless of whether or not the efficiency bound can be solved for analytically. We discuss alternative efficient procedures in the concluding section.

Let $\alpha = (\theta', h) \in \mathcal{A} = \Theta \times \mathcal{H}$. Recall that $\alpha_o = (\theta_o', h_o) \in \mathcal{A}$ is the unique solution to

$$\inf_{\alpha\in\Theta\times\mathcal{H}} \sum_{t=1}^{T} E\left\{m_t(X^{(t)},\alpha)'\Sigma_{ot}(X^{(t)})^{-1}m_t(X^{(t)},\alpha)\right\}, \tag{10}$$

where $m_t(X^{(t)}, \alpha) \equiv E\{\varepsilon_t(Z, \alpha)|X^{(t)}\}$, $\Sigma_{ot}(X^{(t)}) \equiv E[\varepsilon_t(Z, \alpha_o)\varepsilon_t(Z, \alpha_o)'|X^{(t)}]$, and $\varepsilon_t(Z, \alpha)$, $t = 1, ..., T$ are the sequentially orthogonalized residual functions that were introduced in section 2. Then, the optimally weighted orthogonalized SMD estimator $\widetilde{\alpha}_n = (\widetilde{\theta}'_n, \widetilde{h}_n) \in \Theta \times \mathcal{H}_{k(n)}$ is the solution to the following minimization problem:

$$(\widetilde{\theta}'_n, \widetilde{h}_n) = \arg \min_{\alpha \in \Theta \times \mathcal{H}_{k(n)}} \sum_{t=1}^{T} \frac{1}{n} \sum_{i=1}^{n} \widehat{m}_t(X_i^{(t)}, \alpha)' \{\widehat{\Sigma}_{ot}(X_i^{(t)})\}^{-1} \widehat{m}_t(X_i^{(t)}, \alpha), \tag{11}$$

where $\{\mathcal{H}_{k(n)} : k(n) = 1, 2, ...\}$ is a sequence of approximation spaces (sieves) whose union becomes dense in the infinite dimensional parameter space $\mathcal{H}$ as $k(n) \to \infty$, and $\widehat{m}_t(X^{(t)}, \alpha)$, $\widehat{\Sigma}_{ot}(X^{(t)})$ are consistent nonparametric estimators of $m_t(X^{(t)}, \alpha)$, $\Sigma_{ot}(X^{(t)})$ respectively. Note that the sample criterion function (11) corrects both the unknown correlation among the original sets of residual functions $\rho_t(Z; \theta, h(\cdot))$ for $t = 1, ..., T$, and the unknown heteroskedasticity.

In most applications, the sieve spaces $\mathcal{H}_{k(n)}$ are compact sets of series approximations truncated to a finite number of terms. Familiar series approximations include splines, power series, Fourier series, Hermite polynomials, wavelets; see, e.g., Chen (2007) for a review. The orthogonalized conditional means $m_t(X^{(t)}, \alpha)$ and the conditional variances $\Sigma_{ot}(X^{(t)})$ can be consistently estimated by many nonparametric regression methods such as kernel, local linear regression and series least squares (LS) regression. See, e.g., Andrews (1991), Newey (1997), Ai and Chen (2003) and Chen and Pouzo (2008b) for series LS regression estimators of $\widehat{m}_t(X^{(t)}, \alpha)$ and $\widehat{\Sigma}_{ot}(X^{(t)})$.

Suppose that the semiparametric efficiency bound for θ_o is nonsingular. Then, by proofs very similar to those in Ai and Chen (2003) for smooth $\rho_t()$ and those in Chen and Pouzo (2009) for nonsmooth $\rho_t()$, we can establish that $\sqrt{n}(\widetilde{\theta}_n - \theta_o) \Rightarrow \mathcal{N}(0, \Omega^*_\theta)$, where Ω^*_θ is the efficient variance bound derived in Theorem 2.1. If the semiparametric efficiency bound for θ_o is singular, then $\widetilde{\theta}_n$ converges to θ_o at a slower than root-n rate. See Chen and Pouzo (2008b) for details.

4.1 Efficient estimation for the Plug-in problem

For the semiparametric plug-in problem (2)-(3), the optimally weighted orthogonalized SMD estimator $(\widetilde{\theta}'_n, \widetilde{h}_n)$ given in (11) becomes the solution to:

$$\min_{\alpha \in \Theta \times \mathcal{H}_{k(n)}} \left\{ \left\| \widehat{\Sigma}_{o1}^{-\frac{1}{2}} \frac{1}{n} \sum_{i=1}^{n} \widehat{\varepsilon}_1(Z_i, \theta_1, \theta_2, h) \right\|_e^2 + \frac{1}{n} \sum_{i=1}^{n} \left\| \widehat{\Sigma}_{o2}(X_i^{(2)})^{-\frac{1}{2}} \widehat{E}[\rho_2(Z, \theta_2, h)|X_i^{(2)}] \right\|_e^2 \right\}, \tag{12}$$

which achieves the efficiency bound for $\theta_o' = (\theta_{o1}', \theta_{o2}')'$ in the model (2)-(3).

According to Theorems 2.2 and 2.3, when θ_{o1} is exactly identified by unconditional moment (2), the joint efficient estimator $(\widetilde{\theta}_n', \widetilde{h}_n)$ given in (12) is numerically equivalent to the following two stage efficient estimator:

First stage, estimate θ_{o2} efficiently by applying the optimally weighted SMD procedure to the original conditional moment restriction model (3):

$$(\widetilde{\theta}_{2n}', \widetilde{h}_n) = \arg \min_{(\theta_2', h) \in \Theta_2 \times \mathcal{H}_{k(n)}} \sum_{i=1}^{n} \widehat{E}[\rho_2(Z, \theta_2, h)|X_i^{(2)}]' \{\widehat{\Sigma}_{o2}(X_i^{(2)})\}^{-1} \widehat{E}[\rho_2(Z, \theta_2, h)|X_i^{(2)}], \tag{13}$$

where $\widehat{E}[\rho_2(Z, \theta_2, h)|X^{(2)}]$ and $\widehat{\Sigma}_{o2}(X^{(2)})$ are consistent nonparametric estimators of $E[\rho_2(Z, \theta_2, h)|X^{(2)}]$ and $\Sigma_{o2}(X^{(2)})$ respectively. By the results of Ai and Chen (2003) for smooth $\rho_2()$ and of Chen and Pouzo (2009) for nonsmooth $\rho_2()$, one immediately obtains $\sqrt{n}(\widetilde{\theta}_{2n} - \theta_{o2}) \Rightarrow \mathcal{N}(0, \Omega_{\theta_2}^*)$ hence the semiparametric efficiency of $\widetilde{\theta}_{2n}$.

Second stage, estimate θ_{o1} efficiently by plugging the first stage optimally weighted SMD estimator $(\widetilde{\theta}_{2n}', \widetilde{h}_n)$ into a consistently estimated, orthogonalized residual function $\varepsilon_1(Z, \theta_1, \theta_2, h) \equiv \rho_1(Z, \theta_1, \theta_2, h) - \Gamma_{1,2}(X^{(2)})\rho_2(Z, \theta_2, h)$:

$$\widetilde{\theta}_{1n} \text{ solves } \quad \frac{1}{n} \sum_{i=1}^{n} \widehat{\varepsilon}_1(Z_i, \theta_1, \widetilde{\theta}_{2n}, \widetilde{h}_n) = 0, \tag{14}$$

where

$$\widehat{\varepsilon}_1(Z, \theta_1, \theta_2, h) \equiv \rho_1(Z, \theta_1, \theta_2, h) - \widehat{\Gamma}_{1,2}(X^{(2)})\rho_2(Z, \theta_2, h),$$

and $\widehat{\Gamma}_{1,2}(X^{(2)})$ is some consistent nonparametric estimator of

$$\Gamma_{1,2}(X^{(2)}) \equiv E[\rho_1(Z, \theta_{o1}, \theta_{o2}, h_o)\rho_2(Z, \theta_{o2}, h_o)'|X^{(2)}]\{\Sigma_{o2}(X^{(2)})\}^{-1}.$$

For example, $\widehat{\Gamma}_{1,2}(X^{(2)})$ could be

$$\widehat{\Gamma}_{1,2}(X^{(2)}) = \widehat{E}[\rho_1(Z, \widehat{\theta}_{1n}, \widetilde{\theta}_{2n}, \widetilde{h}_n)\rho_2(Z, \widetilde{\theta}_{2n}, \widetilde{h}_n)'|X^{(2)}] \left(\widehat{\Sigma}_{o2}(X^{(2)})\right)^{-1},$$

where $\widehat{\Sigma}_{o2}(X^{(2)})$ is a consistent nonparametric estimator of $\Sigma_{o2}(X^{(2)})$, $\widehat{\theta}_{1n}$ is a consistent estimator of θ_{o1} (say a solution to $\frac{1}{n}\sum_{i=1}^{n} \rho_1(Z_i, \theta_1, \widetilde{\theta}_{2n}, \widetilde{h}_n) = 0$) and $\widehat{E}[\rho_1(Z, \widehat{\theta}_{1n}, \widetilde{\theta}_{2n}, \widetilde{h}_n)\rho_2(Z, \widetilde{\theta}_{2n}, \widetilde{h}_n)'|X^{(2)}]$ is a consistent nonparametric estimator of $E[\rho_1(Z, \theta_{o1}, \theta_{o2}, h_o)\rho_2(Z, \theta_{o2}, h_o)'|X^{(2)}]$, such as a series LS estimator:

$$
\begin{aligned}
& \widehat{E}[\rho_1(Z,\widehat{\theta}_{1n},\widetilde{\theta}_{2n},\widetilde{h}_n)\rho_2(Z,\widetilde{\theta}_{2n},\widetilde{h}_n)'|X^{(2)}] \\
= \; & p_2^{k_{2,n}}(X^{(2)})'(P_2'P_2)^{-1}\sum_{i=1}^{n} p_2^{k_{2,n}}(X_i^{(2)})\rho_1(Z_i,\widehat{\theta}_{1n},\widetilde{\theta}_{2n},\widetilde{h}_n)\rho_2(Z_i,\widetilde{\theta}_{2n},\widetilde{h}_n)',
\end{aligned}
$$

where $p_2^{k_{2,n}}(X^{(2)}) = (p_{2,1}(X^{(2)}),\ldots,p_{2,k_{2,n}}(X^{(2)}))'$ is a series basis that can approximate any square integrable function of $X^{(2)}$ well as $k_{2,n}\to\infty$. Instead of using $\widehat{\Sigma}_{o2}(X^{(2)})$ to compute $\widehat{\Gamma}_{1,2}(X^{(2)})$ one could use the following consistent series LS estimator of $\Sigma_{o2}(X^{(2)})$:

$$
\widetilde{\Sigma}_{o2}(X^{(2)}) = p_2^{k_{2,n}}(X^{(2)})'(P_2'P_2)^{-1}\sum_{i=1}^{n} p_2^{k_{2,n}}(X_i^{(2)})\rho_2(Z_i,\widetilde{\theta}_{2n},\widetilde{h}_n)\rho_2(Z_i,\widetilde{\theta}_{2n},\widetilde{h}_n)'.
$$

The semiparametric efficiency of $\widetilde{\theta}_{1n}$ and $\sqrt{n}(\widetilde{\theta}_{1n}-\theta_{o1}) \Rightarrow \mathcal{N}(0,\Omega^*_{\theta_1})$ can be established using proofs similar to those of Ai and Chen (2007) for smooth $\rho()$ and of Chen and Pouzo (2009) for nonsmooth $\rho()$.

Remark 4.1: Theorems 2.2 and 2.3 suggest many alternative asymptotically efficient estimators of θ_{o1}. In fact, one can use any efficient criterion based on the conditional moment restriction model (3) to construct an asymptotically efficient estimator $(\widetilde{\theta}'_{2n},\widetilde{h}_n)$ in the first stage. Then, in the second stage one can estimate θ_{o1} efficiently by plugging $(\widetilde{\theta}'_{2n},\widetilde{h}_n)$ into the sample moment based on any consistently estimated orthogonalized residual function $\varepsilon_1(Z,\theta_1,\theta_2,h)$. For example, a simple efficient estimator $\widetilde{\widetilde{\theta}}_{1n}$ of θ_{o1} can be computed as

$$
\widetilde{\widetilde{\theta}}_{1n} \text{ solves } \quad \frac{1}{n}\sum_{i=1}^{n}\widetilde{\varepsilon}_1(Z_i,\theta_1,\widetilde{\theta}_{2n},\widetilde{h}_n) = 0, \tag{15}
$$

where

$$
\widetilde{\varepsilon}_1(Z_i,\theta_1,\theta_2,h) \equiv \rho_1(Z_i,\theta_1,\theta_2,h) - \widetilde{\Gamma}_{1,2}(X_i^{(2)},\theta_1)\rho_2(Z_i,\theta_2,h),
$$

and $\widetilde{\Gamma}_{1,2}(X^{(2)},\theta_1)$ is some consistent nonparametric estimator of

$$
\Gamma_{1,2}(X^{(2)},\theta_1) \equiv E[\rho_1(Z,\theta_1,\theta_{o2},h_o)\rho_2(Z,\theta_{o2},h_o)'|X^{(2)}]\{\Sigma_{o2}(X^{(2)})\}^{-1}.
$$

For example, $\widetilde{\Gamma}_{1,2}(X^{(2)},\theta_1)$ could be

$$\widetilde{\Gamma}_{1,2}(X^{(2)},\theta_1)=\widehat{E}[\rho_1(Z,\theta_1,\widetilde{\theta}_{2n},\widetilde{h}_n)\rho_2(Z,\widetilde{\theta}_{2n},\widetilde{h}_n)'|X^{(2)}]\left(\widetilde{\Sigma}_{o2}(X^{(2)})\right)^{-1},$$

while $\widehat{E}[\rho_1(Z,\theta_1,\widetilde{\theta}_{2n},\widetilde{h}_n)\rho_2(Z,\widetilde{\theta}_{2n},\widetilde{h}_n)'|X^{(2)}]$ and $\widetilde{\Sigma}_{o2}(X^{(2)})$ could be series LS, kernel, or local linear regression estimators of $E[\rho_1(Z,\theta_1,\theta_{o2},h_o)\rho_2(Z,\theta_{o2},h_o)'|X^{(2)}]$ and $\Sigma_{o2}(X^{(2)})$ respectively.

4.2 Example: weighted average derivative of a NPIV model

For Example 3.3 $E[Y_1-h_o(Y_2)|X]=0$ (the NPIV model), it is known that the nonparametric estimation of h_o is a difficult ill-posed inverse problem; see, e.g., Newey and Powell (2003) and Carrasco, Florens and Renault (2007) for a detailed review. Let $\widehat{h}_n$ be any consistent estimator of h_o in the NPIV model, such as the estimators of Hall and Horowitz (2005), Darolles, Florens and Renault (2002) or Blundell, Chen and Kristensen (2007). An inefficient simple plug-in estimator of $\theta_o=E[W(Y_2)\nabla^s h(Y_2)]$ is:

$$\widehat{\theta}_n=\frac{1}{n}\sum_{i=1}^{n}W(Y_{2i})\nabla^s\widehat{h}_n(Y_{2i}).$$

Ai and Chen (2007) obtained root-n asymptotic normality of the inefficient simple plug-in estimator $\widehat{\theta}_n$ when $\widehat{h}_n$ is the original SMD estimator proposed in Newey and Powell (2003) and Ai and Chen (2003):

$$\widehat{h}_n=\min_{h\in\mathcal{H}_{k(n)}}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{E}[Y_1-h(Y_2)|X_i]\right)^2, \tag{16}$$

with

$$\widehat{E}[Y_1-h(Y_2)|X_i]=p_2^{k_{2,n}}(X_i)'(P_2'P_2)^{-1}\sum_{j=1}^{n}p_2^{k_{2,n}}(X_j)\{Y_{1j}-h(Y_{2j})\},$$

where the conditional mean function $m_2(X,h)=E[Y_1-h(Y_2)|X]$ is approximated by the series basis functions $p_2^{k_{2,n}}(X)=(p_{2,1}(X),\ldots,p_{2,k_{2,n}}(X))'$. The sieve space $\mathcal{H}_{k(n)}$ is a finite dimensional linear space generated by some spline basis functions $q^{k_{h,n}}(Y_2)=(q_1(Y_2),\ldots,q_{k_{h,n}}(Y_2))'$, with $k_{2,n}\geq k_{h,n}$. The identity weighted SMD estimator $\widehat{h}_n$ is a simple two stage least squares estimator of regressing Y_{1i} on $q^{k_{h,n}}(Y_{2i})$ with $p_2^{k_{2,n}}(X_i)$ as instruments.

We present two semiparametric efficient estimators of $\theta_o=E\{W(Y_2)\nabla^s h_o(Y_2)\}$. Both estimators can be computed in closed-form based on an orthogonalized (or transformed) residual function and an "efficient" first stage NPIV estimator:

$$\widetilde{h}_n = \min_{h \in \mathcal{H}_{k(n)}} \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{E}[Y_1 - h(Y_2)|X_i] \right)^2 \left(\widehat{\Sigma}_{o2}(X_i) \right)^{-1}, \tag{17}$$

$$\widehat{\Sigma}_{o2}(X_i) = p_2^{k_{2,n}}(X_i)'(P_2'P_2)^{-1} \sum_{i=1}^{n} p_2^{k_{2,n}}(X_i)\{Y_{1i} - \widehat{h}_n(Y_{2i})\}^2.$$

Efficient estimator 1: $\widetilde{\theta}_n$ solves (14), that is,

$$\widetilde{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} \left(W(Y_{2i}) \nabla^s \widetilde{h}_n(Y_{2i}) - \widehat{\Gamma}_{1,2}(X_i)\{Y_{1i} - \widetilde{h}_n(Y_{2i})\} \right), \tag{18}$$

with

$$\widehat{\Gamma}_{1,2}(X_i) = p_2^{k_{2n}}(X_i)'(P_2'P_2)^{-1} \sum_{i=1}^{n} p_2^{k_{2n}}(X_i)[W(Y_{2i}) \nabla^s \widehat{h}_n(Y_{2i}) - \widehat{\theta}_n]\{Y_{1i} - \widehat{h}_n(Y_{2i})\} \left(\widehat{\Sigma}_{o2}(X_i) \right)^{-1}.$$

Efficient estimator 2: $\widetilde{\widetilde{\theta}}_n$ solves (15), that is,

$$\frac{1}{n} \sum_{i=1}^{n} \left(W(Y_{2i}) \nabla^s \widetilde{h}_n(Y_{2i}) - \theta - \widetilde{\Gamma}_{12}(X_i, \theta)\{Y_{1i} - \widetilde{h}_n(Y_{2i})\} \right) = 0, \tag{19}$$

with

$$\widetilde{\Gamma}_{12}(X_i, \theta) = p_2^{k_{2,n}}(X_i)'(P_2'P_2)^{-1} \sum_{i=1}^{n} p_2^{k_{2,n}}(X_i)[W(Y_{2i}) \nabla^s \widetilde{h}_n(Y_{2i}) - \theta]\{Y_{1i} - \widetilde{h}_n(Y_{2i})\} \left(\widetilde{\Sigma}_{o2}(X_i) \right)^{-1},$$

$$\widetilde{\Sigma}_{o2}(X_i) = p_2^{k_{2,n}}(X_i)'(P_2'P_2)^{-1} \sum_{i=1}^{n} p_2^{k_{2,n}}(X_i)\{Y_{1i} - \widetilde{h}_n(Y_{2i})\}^2.$$

4.2.1 A Small Monte Carlo Study

We assess the finite sample performance of our estimators in a small simulation study. The parameter of interest is $\theta_o = E[\nabla h_o(Y_2)]$, and the model from which we simulate a random sample $\{(Y_{1i}, Y_{2i}, X_i)'\}_{i=1}^n$ is given by

$$Y_{1i} = h_o(Y_{2i}) + \sqrt{2} U_i, \quad U_i = \sqrt{\omega} \frac{E[h_o(Y_2)|X_i] - h_o(Y_{2i})}{\sigma_o} + \sqrt{1-\omega}\sqrt{\gamma}\varepsilon_i, \tag{20}$$

where $\sigma_o = \sqrt{Var\left(E\left[h_o\left(Y_2\right)|X\right] - h_o\left(Y_2\right)\right)}$ and $\varepsilon \sim \frac{1}{1+\exp\{-\varepsilon/a\}}$ with a chosen such that $Var(\varepsilon) = 1$. Following Blundell, Chen and Kristensen (2007), we generate our Monte Carlo (MC) experiment from the 1995 British Family Expenditure Survey (FES) data set using the subsample of families with no children. In particular, Y_2 is log-total expenditure (the endogenous regressor), $\widetilde{X}$ is log-gross earnings and $X = \Phi(\widetilde{X})$ is the instrument. We simulate $(Y_{2i}, \widetilde{X}_i)$ jointly from a bivariate Gaussian density f, with first and second moments estimated from the FES data set. Denote $B_r\left(x\right) = \frac{1}{(r-1)!}\sum_{j=0}^{r}\left(-1\right)^j \begin{pmatrix} r \\ j \end{pmatrix} \left(\max\left\{x-j,0\right\}\right)^{r-1}$ as the B-spline of order $r \geq 1$. We set the true function $h_o\left(y_2\right) = B_6\left(y_2 - c\right)$ where c is the highest integer less than the minimum value of the Y_2 series in the FES data set. We then draw ε_i independently from $\varepsilon \sim \frac{1}{1+\exp\{-\varepsilon/a\}}$, and generate U_i and then Y_{1i} according to model (20). We consider two cases: a "mid-endogeneity" case where $\sqrt{1-\omega} = 0.5$, and a "high-endogeneity" case where $\sqrt{1-\omega} = 0.001$. We let $\gamma = 0.1$ for the "mid-endogeneity" case, and for the "high-endogeneity" case we choose γ in such a way that the unconditional variance of U remains the same as for the "mid-endogeneity" case.

In this simulation study, we consider three different sample sizes: $n = 250,\ 500,\ 1500$. For each sample size we compute 4 different estimators: (a) $\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}\nabla\widehat{h}_n(Y_{2i})$, the inefficient simple plug-in estimator; (b) $\widetilde{\widehat{\theta}}_n = \frac{1}{n}\sum_{i=1}^{n}\left(\nabla\widehat{h}_n(Y_{2i}) - \widehat{\Gamma}_{1,2}(X_i)\{Y_{1i} - \widehat{h}_n(Y_{2i})\}\right)$, a modified plug-in estimator based on $\widehat{h}$; (c) $\widetilde{\widetilde{\theta}}_n$ given in (19), an efficient estimator based on $\widetilde{h}$; (d) $\widetilde{\theta}_n^s$ is an iterative solution to (12), similar to continuously updated optimally weighted SMD. This estimator is computationally costly; hence is only computed for $n = 250$. To compute the two inefficient estimators $\widehat{\theta}_n$ and $\widetilde{\widehat{\theta}}_n$ we used the SMD estimator $\widehat{h}$ given in (16), where $\widehat{E}[Y_1 - h(Y_2)|X_i]$ is a series LS estimator (say cosine polynomials basis with 10 coefficients) for $E[Y_1 - h(Y_2)|X_i]$, and $\mathcal{H}_{k(n)}$ is a fifth-order B-spline basis $\left(1, B_5\left(\cdot-2\right), B_5\left(\cdot-3\right), B_5\left(\cdot-4\right)\right)$. The MC results reported below actually correspond to the SMD estimator $\widehat{h}$ of Blundell, Chen and Kristensen (2007), which includes the L^2-norm of the second derivative of h as a smoothness penalty with a small tuning parameter $\lambda_n = 0.075$. To compute the efficient estimator $\widetilde{\widetilde{\theta}}_n$ given in (19) we used the SMD estimator $\widetilde{h}$ given in (17), where $\widehat{\Sigma}_{o2}(X_i)$ is a Gaussian kernel (with bandwidth $n^{-1/5}$) estimator of $\Sigma_{o2}(X)$, with $\widehat{h}$ as the initial consistent estimator of h_o. To compute estimators (b), (c) and (d), we also need to compute a consistent estimate of the correlation correction term $\Gamma_{1,2}(X)$. We used a series LS estimator with polynomial splines of order 2 and 2 equally spaced knots to estimate $\Gamma_{1,2}(X)$.

For each sample size we perform 1000 Monte Carlo (MC) repetitions. Tables 1 and 2 present the results for the different estimators of θ_o for the "high-endogeneity" ($\sqrt{1-\omega} = 0.001$) case and the "mid-endogeneity" ($\sqrt{1-\omega} = 0.5$) case respectively. In each table, the MC Bias is computed against the sample mean of the derivative of the true function. E_{MC} stands for the mean of the MC sample, Var_{MC} and MSE_{MC} are defined analogously. In the last column, $Var_{MC}/Var_{MC,a}$ is the ratio of the MC variance of each estimator, (b), (c) and (d), over the simple plug-in estimator (a).

Table 1: Monte Carlo Results for "high-endogeneity" case.

		$BIAS^2_{MC} \times 10^3$	$Var_{MC} \times 10^3$	$MSE_{MC} \times 10^3$	$\frac{Var_{MC}}{Var_{MC,a}}$
$n = 250$	$\widehat{\theta}_n(a)$	1.7600	86.832	88.596	-
	$\widetilde{\widehat{\theta}}_n(b)$	1.5900	84.707	86.292	0.9755
	$\widetilde{\widetilde{\theta}}_n(c)$	0.0330	68.441	68.473	0.7882
	$\widetilde{\theta}^s_n(d)$	0.0780	66.322	66.330	0.7638
$n = 500$	$\widehat{\theta}_n(a)$	0.3220	47.360	47.682	-
	$\widetilde{\widehat{\theta}}_n(b)$	0.3500	46.222	46.573	0.9759
	$\widetilde{\widetilde{\theta}}_n(c)$	1.8210	38.232	40.067	0.8072
$n = 1500$	$\widehat{\theta}_n(a)$	0.4650	10.512	10.938	-
	$\widetilde{\widehat{\theta}}_n(b)$	0.4028	10.327	10.752	0.9799
	$\widetilde{\widetilde{\theta}}_n(c)$	1.4327	9.512	10.918	0.8943

A brief summary of MC results: First, the MC variances of all the estimators (a), (b), (c) and (d) decrease approximately linearly as the sample size increases. The QQ plots, which are not reported here for length considerations, indicate that all the four estimators are root-n asymptotically normal. Second, the efficient estimators, (c) and (d), have lower MC variances than the inefficient simple plug-in estimator (a). Third, the MC variance gap between the estimators (or the finite sample efficiency gain) is bigger for the "high-endogeneity" case.[5] Lastly, the variance gap decreases as the sample size n increases. All of these findings are consistent with our theoretical results.

[5]The particular magnitude of the finite sample efficiency gain directly depends on the value of γ (the exogenous noise level). In MC studies that are not reported here, we discover that smaller γ values lead to smaller biases and smaller variances in the efficient estimators; hence bigger MC variance gaps.

Table 2: Monte Carlo Results for “mid-endogeneity” case.

		$BIAS^2_{MC} \times 10^3$	$Var_{MC} \times 10^3$	$MSE_{MC} \times 10^3$	$\frac{Var_{MC}}{Var_{MC,a}}$
$n = 250$	$\widehat{\theta}_n(a)$	0.3420	79.246	79.590	-
	$\widetilde{\widehat{\theta}}_n(b)$	0.2751	77.164	77.433	0.9716
	$\widetilde{\widetilde{\theta}}_n(c)$	0.1810	70.005	70.186	0.8717
	$\widetilde{\widetilde{\theta}}^s_n(d)$	0.3110	67.700	68.011	0.8530
$n = 500$	$\widehat{\theta}_n(a)$	0.2010	37.503	37.704	-
	$\widetilde{\widehat{\theta}}_n(b)$	0.1860	36.951	37.137	0.9853
	$\widetilde{\widetilde{\theta}}_n(c)$	0.2200	33.794	33.974	0.9064
$n = 1500$	$\widehat{\theta}_n(a)$	0.0200	9.2042	9.2244	-
	$\widetilde{\widehat{\theta}}_n(b)$	0.0201	9.1928	9.2132	0.9987
	$\widetilde{\widetilde{\theta}}_n(c)$	0.1062	8.6932	8.7996	0.9444

5 Conclusion

In this paper we computed the semiparametric efficiency bound for finite dimensional parameters of sequential moment restriction models (1) containing unknown functions that may depend on endogenous variables. The results extend those of Chamberlain (1992b), Ai and Chen (2003) and Chen and Pouzo (2009) to the case of semi/nonparametric conditional moment restriction with nested information sets. The results also extend those of Chamberlain (1992a), and Brown and Newey (1998) to the case of sequential moment restrictions involving unknown functions. Our characterization of the efficiency bound is useful in evaluating and comparing several competing estimators that are typically proposed for a particular semiparametric econometric model. Although we can only characterize the efficiency bounds for conditional moment models involving several unknown functions when they depend on different arguments, these bounds can be computed analytically for many specific models containing only one unknown function. In terms of semiparametric efficiency bound calculation, our approach carries over to allow for T to increase to infinity. However, any efficient estimation method would face the “curse-of-dimensionality” when T is very large.

We present an optimally weighted, orthogonalized SMD estimation procedure for (θ_o, h_o) identified by the sequential moment restriction model (1). When the semiparametric efficiency bound for θ_o is non-singular, we note that this estimator is root-n asymptotically normal and efficient for θ_o. There are many alternative procedures that can also achieve the semiparametric efficiency bound for θ_o in model (1). For instance, one could extend the constrained sieve MLE approach of Gallant and Tauchen (1989), Gallant, Hansen and Tauchen (1990) and Ai (2007) to estimate θ_o efficiently. Notice that (θ_o, h_o) in model (1) is the unique solution to

$$E[\rho_t(Z;\theta_o, h_o(\cdot)) \otimes p_t^{k_{t,n}}(X^{(t)})] = 0 \quad \text{for } t = 1, ..., T, \tag{21}$$

where $p_t^{k_{t,n}}(X^{(t)}) = (p_{t,1}(X^{(t)}), \ldots, p_{t,k_{t,n}}(X^{(t)}))'$ is a series of basis functions that can approximate any square integrable function of $X^{(t)}$ arbitrarily well as $k_{t,n} \to \infty$. One could also estimate θ_o in (21) by extending the GMM with increasing number of unconditional moments of Hahn (1997), or the continuum GMM of Carrasco and Florens (2000), or the empirical likelihood with increasing number of unconditional moments of Donald, Imbens and Newey (2003). We shall investigate these alternative efficient procedures in another paper.

References

[1] Ai, C. (2007): "A Nonparametric Maximum Likelihood Estimation of Conditional Moment Restrictions Models", *International Economic Review*, 48(4), 1093 - 1118.

[2] Ai, C. and X. Chen (2003): "Efficient Estimation of Conditional Moment Restrictions Models Containing Unknown Functions," *Econometrica,* 71, 1795-1843.

[3] 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.

[4] Andrews, D. (1991) "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models", *Econometrica,* 59, 307-345.

[5] Antoine, B., H. Bonnal and E. Renault (2007) "On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood", *Journal of Econometrics*, 138, 461-487.

[6] Begun, J., W. Hall, W. Huang, and J. Wellner (1983): "Information and Asymptotic Efficiency in Parametric-Nonparametric Models," Annals of Statistics, 11, 432–452.

[7] Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1993): Efficient and Adaptive Inference in Semiparametric Models. Johns Hopkins University Press, Baltimore.

[8] Blundell, R., X. Chen and D. Kristensen (2007): "Semi-nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613-1670.

[9] Brown, B. and W. Newey (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66, 453-464.

[10] Carrasco, M. and J. P. Florens. 2000. Generalization of GMM to a Continuum of Moment Conditions. Econometric Theory 20:797-834.

[11] Carrasco, M., J. P. Florens and E. Renault 2007. Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization. Handbook of Econometrics, Vol. 6B, eds. J.J. Heckman and E.E. Leamer. North-Holland, Amsterdam.

[12] Chamberlain, G. 1987. Asymptotic Effciency in Estimation with Conditional Moment Restrictions. Journal of Econometrics 34:305-334.

[13] Chamberlain, G. (1992a) "Comment: sequential moment restrictions in panel data," *Journal of Business and Economic Statistics* 10, 20-26.

[14] Chamberlain, G. (1992b) "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567-596.

[15] Chen, X., 2007. Large Sample Sieve Estimation of Semi-Nonparametric Models, Chp. 76 in Handbook of Econometrics, Vol. 6B, eds. J.J. Heckman and E.E. Leamer. North-Holland, Amsterdam.

[16] Chen, X. and D. Pouzo (2009): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *Journal of Econometrics* 152, 46-60.

[17] Chen, X. and D. Pouzo, 2008a. Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments. Cowles Foundation Discussion Paper 1650.

[18] Chen, X. and D. Pouzo, 2008b. Efficient Estimation of Nonparametric Quantile IV Weighted Average Derivative. Mimeo, Yale University and New York University.

[19] Chernozhukov, V. and H. Hong. 2003. An MCMC Approach to Classical Estimation. Journal of Econometrics 115:293-346.

[20] Chernozhukov, V., G. Imbens, and W. Newey. 2007. Instrumental Variable Estimation of Non-separable Models. Journal of Econometrics 139, 4-14.

[21] Darolles, S., J.-P. Florens and E. Renault (2002): "Nonparametric Instrumental Regression," mimeo, GREMAQ, University of Toulouse.

[22] Donald, S., G. Imbens and W. Newey (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, 117, 55–93.

[23] Gallant, A. R., L. P. Hansen, and G. Tauchen (1990): "Using Conditional Moments of Asset Payoffs to infer the Volatility of Intertemporal Marginal Rates of Subsitution," Journal of Econometrics, 45(2), 141–179.

[24] Gallant, A. R., and G. Tauchen (1989): "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," Econometrica, 57(5), 1091–1120.

[25] Hahn, J. (1997) "Efficient estimation of panel data models with sequential moment restrictions," *Journal of Econometrics* 79, 1-21.

[26] Hall, P. and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," Ann of Statistics, 33, 2904-2929.

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

[28] Hansen, L.P. (1985): "A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators," *Journal of Econometrics*, 30, 203-238.

[29] Hansen, L.P. (1993): "Semiparametric efficiency bounds for linear time-series models," in *Models, Methods, and Applications of Econometrics*: *Essays in Honor of A. R. Bergstrom*, edited by PCB Phillips, Cambridge, MA: Blackwell.

[30] Hansen, L.P. (2007). Generalized Method of Moments, entry for the New Palgrave Dictionary of Economics, 2nd edition: edited by S. Durlauf and L. Blume. New York: Elsevier.

[31] Hansen, L.P., J. Heaton, and M. Ogaki (1988): "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," Journal of the American Statistical Association, 83, 863–871.

[32] Hayashi, F., and C. Sims, 1983: "Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments. Econometrica, 51, 783–798.

[33] Kitamura, Y., G. Tripathi and H. Ahn (2004) "Empirical Likelihood-based Inference in Conditional Moment Restriction Models", *Econometrica*, 72, 1667-1714.

[34] Newey, W. 1990. "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99-135.

[35] Newey, W. 1993. Efficient Estimation of Models with Conditional Moment Restrictions. In Handbook of Statistics, vol. 11, edited by G. S. Maddala, C. R. Rao, and H. D. Vinod. Amsterdam: North Holland.

[36] Newey, W., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics, 79, 147-168.

[37] Newey, W., 2004. Efficient Semiparametric Estimation via Moment Restrictions, *Econometrica,* 72, 1877-1897.

[38] Newey, W. and J. Powell (2003) "Instrumental Variable Estimation of Nonparametric Models", *Econometrica,* 71, 1565-1578.

[39] Newey, W. and R. Smith (2004) "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators", *Econometrica*, 72, 219-256.

[40] Newey, W. and T. Stoker, 1993, Efficiency of Weighted Average Derivative Estimators and Index Models, Econometrica, 61, 1199-1223.

[41] Otsu, T. (2007) "Sieve Conditional Empirical Likelihood Estimation of Semiparametric Models", forthcoming in *Econometric Theory.*

[42] Stein, C. (1956): "Efficient Nonparametric Testing and Estimation," in Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability, vol. 1, pp. 187–196, Berkeley University of California Press.

[43] Van der Vaart, A., 1991. On Differentiable Functionals. Annals of Statistics, 19, 178-204.

Mathematical Appendix

We follow the approach of Stein (1956), Begun, Huang and Wellner (1983), Bickel, et al. (1993), Newey (1990), Chamberlain (1992a, b), and van der Vaart (1991) on semiparametric efficiency bound calculation. Recall the following notation for any $\alpha \in \mathcal{A}$:

$$\varepsilon_T(Z,\alpha) \equiv \rho_T(Z,\alpha),\ \varepsilon_s(Z,\alpha) \equiv \rho_s(Z,\alpha) - \sum_{t=s+1}^{T} \Gamma_{s,t}(X^{(t)})\varepsilon_t(Z,\alpha) \text{ for } s = T-1, ..., 1,$$

$\Gamma_{s,t}(X^{(t)}) \equiv E[\rho_s(Z,\alpha_o)\varepsilon_t(Z,\alpha_o)'|X^{(t)}]\{\Sigma_{ot}(X^{(t)})\}^{-1}$ for $s < t$ and $\Sigma_{ot}(X^{(t)}) \equiv E[\varepsilon_t(Z,\alpha_o)\varepsilon_t(Z,\alpha_o)'|X^{(t)}]$. Also, recall that

$$E\{\varepsilon_s(Z,\alpha_o)\varepsilon_t(Z,\alpha_o)v(X^{(s)})q(X^{(t)})\} = 0 \tag{22}$$

holds for any $s \neq t$ and for any measurable functions v and q. Finally we denote $m_t(X^{(t)},\alpha) \equiv E[\varepsilon_t(Z,\alpha)|X^{(t)}]$ for $t = 1, ..., T$.

Denote $d_z = \dim(Z)$. Let $p_o(\cdot)$ be the true probability density of $Z = (Y', X')'$ with respect to a sigma-finite measure μ on $\mathcal{Z} \subset \mathcal{R}^{d_z}$ that satisfies model (1), which is equivalent to the following model:

$$E[\varepsilon_t(Z,\alpha_o)|X^{(t)}] = 0 \quad \text{for } t = 1, ..., T \text{ and } \alpha_o \in \mathcal{A}, \tag{23}$$

where E denotes expectation taken with respect to the true density function $p_o(z)$. Let E_p denote expectation taken with respect to arbitrary density function $p(z)$. For arbitrary $\alpha \in \mathcal{A}$, the conditional moment restrictions

$$E_p[\varepsilon_t(Z;\theta,h)|X^{(t)}] = 0 \quad \text{for } t = 1, ..., T \tag{24}$$

do not uniquely determine $p(z)$. For any $\alpha \in \mathcal{A}$, let $\mathcal{F}_\alpha$ denote all probability density functions that satisfy (24):

$$\mathcal{F}_\alpha = \left\{ p(\cdot) : \int_{z\in\mathcal{Z}} p(z)d\mu(z) = 1,\ p(\cdot) \geq 0,\ E_p[\varepsilon_t(Z;\theta,h)|X^{(t)}] = 0 \text{ for } t = 1, ..., T \right\}.$$

For any $p(\cdot) \in \mathcal{F}_\alpha$, we can always write $p(z) \equiv f(z|\alpha, g)$ with $p_o(z) \equiv f(z|\alpha_o, g_o)$, where p is of a known functional form up to unknown parameters α and g, with g being an unknown measurable function of z (see Ai (2007) for an example). $g(z)$ can be viewed as the remainder of the probability density function $p(z) \equiv f(z|\alpha, g)$ that is not determined by (24), and is unrestricted except for satisfying $f(\cdot|\alpha, g) \in \mathcal{F}_\alpha$ and $f(\cdot|\alpha_o, g_o) \in \mathcal{F}_{\alpha_o}$. Let $\mathcal{G}$ denote a class of real valued measurable functions of Z satisfying (i) for each $\alpha \in \mathcal{A}$, $\{f(z|\alpha, g) : g \in \mathcal{G}\} = \mathcal{F}_\alpha$; (ii) there is a g_o in the interior of $\mathcal{G}$ such that $p_o(z) \equiv f(z|\alpha_o, g_o)$.

The following condition shall be imposed throughout the paper.

Assumption A: Let $\{(\alpha(\tau), g(\tau)) : \tau \in [0,1]\}$ denote a family of parametric specifications in the parameter space $\mathcal{A} \times \mathcal{G}$ satisfying: **(1)** $f(\cdot|\alpha(\tau), g(\tau)) \in \mathcal{F}_{\alpha(\tau)}$; $f(Z|\alpha(0), g(0)) = f(Z|\alpha_o, g_o) \equiv p_o(Z)$ holds with probability one; $\tau \to \sqrt{f(\cdot|\alpha(\tau), g(\tau))}$ is mean-square differentiable. **(2)** For all $j = 1, ..., T$, with probability one, $\varepsilon_j(Z, \alpha(\tau))$ is continuous at $\alpha(\tau)$ in a small neighborhood of $\tau = 0$, $\frac{dE[\varepsilon_j(Z,\alpha(\tau))|X^{(j)}]}{d\theta}|_{\tau=0}$ and $\frac{dE[\varepsilon_j(Z,\theta_o,h(\tau))|X^{(j)}]}{d\tau}|_{\tau=0}$ exist and have finite second moments, and $E[|\varepsilon_j(Z, \alpha_o)|^2|X^{(j)}]$ is bounded.

Proof. (**Theorem 2.1**) Let $\{(\alpha(\tau), g(\tau)) : \tau \in [0,1]\}$ be any parametric path in $\mathcal{A} \times \mathcal{G}$ satisfying assumption A. Denote the log-likelihood function (of one observation) of a parametric submodel by $\ell(z, \alpha(\tau), g(\tau)) = \log f(z|\alpha(\tau), g(\tau))$. Under assumption A, we can write the pathwise derivative of $\ell(z, \alpha(\tau), g(\tau))$ at (α_o, g_o) as

$$\begin{aligned}
\nabla \ell(z, \alpha_o, g_o) &= \lim_{\tau \to 0} \frac{\ell(z, \alpha(\tau), g(\tau)) - \ell(z, \alpha_o, g_o)}{\tau} \\
&= \ell_\theta(z, \alpha_o, g_o) \frac{d\theta(\tau)}{d\tau}|_{\tau=0} + \ell_h(z, \alpha_o, g_o) \frac{dh(\tau)}{d\tau}|_{\tau=0} + \ell_g(z, \alpha_o, g_o) \frac{dg(\tau)}{d\tau}|_{\tau=0} \\
&= \ell_\theta(z, \alpha_o, g_o)\Delta\theta + \ell_h(z, \alpha_o, g_o)\,[\Delta h] + \ell_g(z, \alpha_o, g_o)\Delta g,
\end{aligned}$$

where the second and the third term on the right-hand side denote the pathwise derivative with respect to h and g respectively. To simplify notation we denote $\ell_\theta(z) \equiv \ell_\theta(z, \alpha_o, g_o)$, $\ell_h(z)[\Delta h] \equiv \ell_h(z, \alpha_o, g_o)[\frac{dh(\tau)}{d\tau}|_{\tau=0}]$ and $\ell_g(z)\Delta g(z) \equiv \ell_g(z, \alpha_o, g_o)\frac{dg(\tau)}{d\tau}|_{\tau=0}$. Notice that any $p(z) = f(z|\alpha(\tau), g(\tau)) \in \mathcal{F}_{\alpha(\tau)}$ satisfies restrictions (24). By differentiating both sides of (24), we obtain:

$$\frac{dE[\varepsilon_j(Z, \alpha_o)|X^{(j)}]}{d\theta} + E[\varepsilon_j(Z, \alpha_o)\ell_\theta(Z)|X^{(j)}] = 0 \text{ for } j = 1, ..., T; \tag{25}$$

$$\frac{dE[\varepsilon_j(Z, \theta_o, h(\tau))|X^{(j)}]}{d\tau}|_{\tau=0} + E[\varepsilon_j(Z, \alpha_o)\ell_h(Z)\,[\Delta h]\,|X^{(j)}] = 0 \text{ for } j = 1, ..., T; \tag{26}$$

$$E[\varepsilon_j(Z, \alpha_o)\ell_g(Z)\Delta g(Z)|X^{(j)}] = 0 \text{ for } j = 1, ..., T. \tag{27}$$

Denote

$$\mathbb{T}_h = \left\{ a_h(\cdot) = \ell_h(\cdot)[\Delta h] : E[a_h(Z)] = 0, E[\{a_h(Z)\}^2] < \infty, \Delta h \in \mathcal{H} - \{h_o\},\ (26) \text{ holds} \right\}.$$

$$\mathbb{T}_g = \left\{ a_g(\cdot) = \ell_g(\cdot)\Delta g(\cdot) : E[a_g(Z)] = 0, E[\{a_g(Z)\}^2] < \infty, \text{ (27) holds} \right\}.$$

Let $\overline{\mathbb{T}}_h$ and $\overline{\mathbb{T}}_g$ respectively denote the closed linear completions of $\mathbb{T}_h$ and $\mathbb{T}_g$ under the mean squared norm $\|v(\cdot)\|_2^2 = E\{v(Z)^2\}$. Then $\overline{\mathbb{T}}_h$ and $\overline{\mathbb{T}}_g$ are the tangent spaces for the nonparametric parameters h and g respectively. Denote $\overline{\mathbb{T}} = \overline{\mathbb{T}}_h + \overline{\mathbb{T}}_g$. Let $\text{Proj}(\cdot|\overline{\mathbb{T}})$ denote the population least square projection of $\cdot$ onto the space $\overline{\mathbb{T}}$. The semiparametric efficient score of θ_o is given by $S_\theta^* \equiv \ell_\theta(Z) - \text{proj}(\ell_\theta(Z)|\overline{\mathbb{T}})$ (see, e.g., Bickel et al. 1993). Denote $J_o \equiv E[S_\theta^* S_\theta^{*\prime}]$ as the semiparametric Fisher information bound. If J_o is non-singular, then the semiparametric efficient variance bound for θ_o is $\Omega_\theta^* \equiv (J_o)^{-1} \equiv (E[S_\theta^* S_\theta^{*\prime}])^{-1}$.

To compute the least squares projections, note that, for each component θ^k (of θ), $k = 1, ..., d_\theta$, the projection $\text{proj}(\ell_{\theta^k}(Z)|\overline{\mathbb{T}})$ solves the following minimization problem:

$$\begin{aligned}
& E\left(\left[\ell_{\theta^k}(Z) - \text{proj}(\ell_{\theta^k}(Z)|\overline{\mathbb{T}})\right]^2\right) \equiv E\left(\left[\ell_{\theta^k}(Z) - a_h^*(Z) - a_g^*(Z)\right]^2\right) \\
= & \min_{a_h \in \overline{\mathbb{T}}_h, a_g \in \overline{\mathbb{T}}_g} E\left([\ell_{\theta^k}(Z) - a_h(Z) - a_g(Z)]^2\right) \\
= & \min_{a_h \in \overline{\mathbb{T}}_h} \left\{ \min_{a_g \in \overline{\mathbb{T}}_g} E\left([\ell_{\theta^k}(Z) - a_h(Z) - a_g(Z)]^2\right) \right\} \\
= & \min_{a_h \in \overline{\mathbb{T}}_h} \left\{ E\left(\left[\ell_{\theta^k}(Z) - a_h(Z) - \text{proj}(\ell_{\theta^k}(Z) - a_h(Z)|\overline{\mathbb{T}}_g)\right]^2\right) \right\},
\end{aligned}$$

where $a_h^* \in \overline{\mathbb{T}}_h, a_g^* \equiv \text{proj}(\ell_{\theta^k}(Z) - a_h(Z)|\overline{\mathbb{T}}_g) \in \overline{\mathbb{T}}_g$ denote a pair of solutions.

For any $a_h = \ell_h(\cdot)[\Delta h] \in \mathbb{T}_h$ (with $\Delta h \in \mathcal{H} - \{h_o\}$), to compute a solution $\text{proj}(\ell_{\theta^k}(Z) - a_h(Z)|\overline{\mathbb{T}}_g)$ to the problem $\min_{a_g \in \overline{\mathbb{T}}_g} E\left[\ell_{\theta^k}(Z) - a_h(Z) - a_g(Z)\right]^2$, we write the Lagrangian expression as

$$E\left\{ [\ell_{\theta^k}(Z) - a_h(Z) - a_g(Z)]^2 + 2\sum_{j=1}^{T} \lambda_j(X^{(j)})' \varepsilon_j(Z, \alpha_o) a_g(Z) \right\}, \tag{28}$$

where $\lambda_j(X^{(j)})$ is the Lagrangian multiplier for the constraint (27). Applying calculus of variation, any solution $\overline{a}_g$ to the unconstrained minimization problem (28) should satisfy the following first order condition:

$$\ell_{\theta^k}(Z) - a_h(Z) - \overline{a}_g(Z) = \sum_{j=1}^{T} \lambda_j(X^{(j)})' \varepsilon_j(Z, \alpha_o), \tag{29}$$

$$E[\varepsilon_j(Z, \alpha_o) \times \overline{a}_g(Z)|X^{(j)}] = 0 \text{ for } j = 1, 2, ..., T.$$

Note that $\overline{a}_g(Z)$ given by equation (29) satisfies $E\{\overline{a}_g(Z)\} = 0$. Under constraints (25) and (26), definition of $\varepsilon_j(Z,\alpha_o)$, $j=1,2,...,T$ and relation (22), it is straightforward to show that

$$\lambda_t(X^{(t)})' = -\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta^k} - \frac{dm_t(X^{(t)},\alpha_o)}{dh}[\Delta h]\right]' \Sigma_{ot}(X^{(t)})^{-1}, \ t=1,2,...,T$$

solves (29). Hence

$$\overline{a}_g(Z) = \ell_{\theta^k}(Z) - a_h(Z) - \sum_{t=1}^{T} \lambda_t(X^{(t)})'\varepsilon_t(Z,\alpha_o)$$

solves the unconstrained minimization problem (28). Moreover, because $E\{[\ell_{\theta^k}(Z)]^2\}<\infty$, $E\{[a_h(Z)]^2\}<\infty$, $E\left\{\left\|\lambda_t(X^{(t)})'\varepsilon_t(Z,\alpha_o)\right\|_e^2\right\}<\infty$, we have $E\{[\overline{a}_g(Z)]^2\}<\infty$ and $\overline{a}_g \in \overline{\mathbb{T}}_g$. Thus

$$\mathrm{proj}(\ell_{\theta^k}(Z) - a_h(Z)|\overline{\mathbb{T}}_g) = \overline{a}_g(Z) = \ell_{\theta^k}(Z) - a_h(Z) - \sum_{t=1}^{T} \lambda_t(X^{(t)})'\varepsilon_t(Z,\alpha_o).$$

That is, for any $a_h = \ell_h(\cdot)[\Delta h] \in \mathbb{T}_h$ (with $\Delta h \in \mathcal{H} - \{h_o\}$), we have:

$$\begin{aligned} & \ell_{\theta^k}(Z) - a_h(Z) - \mathrm{proj}(\ell_{\theta^k}(Z) - a_h(Z)|\overline{\mathbb{T}}_g) \\ = & -\sum_{t=1}^{T} \left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta^k} - \frac{dm_t(X^{(t)},\alpha_o)}{dh}[\Delta h]\right]' \Sigma_{ot}(X^{(t)})^{-1}\varepsilon_t(Z,\alpha_o). \end{aligned} \tag{30}$$

Recall that the space $\overline{\mathcal{W}}$ is the closed linear completion of $\mathcal{H}-\{h_o\}$ under the pseudo-norm $||\cdot||$

$$\|\Delta h\|^2 \equiv E\left[\sum_{t=1}^{T}\left[\frac{dm_t(X^{(t)},\alpha_o)}{dh}[\Delta h]\right]' \Sigma_{ot}(X^{(t)})^{-1}\left[\frac{dm_t(X^{(t)},\alpha_o)}{dh}[\Delta h]\right]\right] < \infty.$$

For any direction $\Delta h \in \mathcal{H}-\{h_o\}$ with $a_h = \ell_h(\cdot)[\Delta h] \in \mathbb{T}_h$, given assumptions 2, 3 and A and relation (30), we have $\|\Delta h\|^2 < \infty$ hence this direction Δh belongs to $\overline{\mathcal{W}}$. Conversely, for any $\Delta\widetilde{h} \in \mathcal{H}-\{h_o\}$ with $\left\|\Delta\widetilde{h}\right\|^2 < \infty$ (so that $\Delta\widetilde{h} \in \overline{\mathcal{W}}$), define

$$\widetilde{a}_h(z) = -\sum_{t=1}^{T}\left[\frac{dm_t(X^{(t)},\alpha_o)}{dh}[\Delta\widetilde{h}]\right]' \Sigma_{ot}(X^{(t)})^{-1}\varepsilon_t(z,\alpha_o).$$

It is obvious that $E[\widetilde{a}_h(Z)]=0$. By definition of $\overline{\mathcal{W}}$ and $\Delta\widetilde{h} \in \overline{\mathcal{W}}$ we have $E[\{\widetilde{a}_h(Z)\}^2] = \left\|\Delta\widetilde{h}\right\|^2 < \infty$.

Also relation (22) implies that

$$\frac{dm_j(X^{(j)},\alpha_o)}{dh}[\Delta\widetilde{h}] + E[\widetilde{a}_h(Z)\varepsilon_j(Z,\alpha_o)'|X^{(j)}] = 0 \text{ for } j = 1,...,T,$$

thus $\widetilde{a}_h(\cdot) \in \mathbb{T}_h$.

By definitions of $\mathbb{T}_h$, $\overline{\mathbb{T}}_h$ and $a_h^*()$, there exists a sequence $\{w_{h,j} \in \mathcal{H} - \{h_o\}, j = 1,2,...\}$ such that $a_{h,j}(\cdot) = \ell_h(\cdot)[w_{h,j}] \in \mathbb{T}_h$ converges to $a_h^*(\cdot) \in \overline{\mathbb{T}}_h$ under the mean-squared norm. Since the projection is a bounded linear functional, $a_{g,j}(Z) \equiv$proj$(\ell_{\theta^k}(Z) - a_{h,j}(Z)|\overline{\mathbb{T}}_g)$ converges to $a_g^*(Z) \in \overline{\mathbb{T}}_g$ under the mean-squared norm. By definition of $\overline{\mathcal{W}}$, such a sequence $\{w_{h,j} \in \mathcal{H} - \{h_o\}, j = 1,2,...\}$ belongs to $\overline{\mathcal{W}}$. By relation (30), we have:

$$\begin{aligned} \infty &> E\left[\ell_{\theta^k}(Z) - a_{h,j}(Z) - \text{proj}(\ell_{\theta^k}(Z) - a_{h,j}(Z)|\overline{\mathbb{T}}_g)\right]^2 \\ &= E\left[\sum_{t=1}^{T}\left\|\{\Sigma_{ot}(X^{(t)})\}^{-\frac{1}{2}}\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta^k} - \frac{dm_t(X^{(t)},\alpha_o)}{dh}[w_{h,j}]\right]\right\|_e^2\right] \\ &\geq E\left[\sum_{t=1}^{T}\left\|\{\Sigma_{ot}(X^{(t)})\}^{-\frac{1}{2}}\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta^k} - \frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o^k]\right]\right\|_e^2\right], \end{aligned} \tag{31}$$

where the last inequality is due to the fact that r_o^k is a solution to

$$\inf_{w\in\overline{\mathcal{W}}} E\left[\sum_{t=1}^{T}\left\|\{\Sigma_{ot}(X^{(t)})\}^{-\frac{1}{2}}\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta^k} - \frac{dm_t(X^{(t)},\alpha_o)}{dh}[w]\right]\right\|_e^2\right].$$

Taking limit as $j \to \infty$ in both sides of inequality (31), we obtain:

$$E\left(\left[\ell_{\theta^k}(Z) - a_h^*(Z) - a_g^*(Z)\right]^2\right) \geq E\left[\sum_{t=1}^{T}\left\|\{\Sigma_{ot}(X^{(t)})\}^{-\frac{1}{2}}\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta^k} - \frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o^k]\right]\right\|_e^2\right]. \tag{32}$$

On the other hand, by definitions of $\overline{\mathcal{W}}$ and r_o^k, there exists a subsequence $\{\widetilde{w}_{h,j} \in \mathcal{H} - \{h_o\}, j = 1,2,...\}$ in $\overline{\mathcal{W}}$ such that

$$\infty > E\left[\sum_{t=1}^{T}\left\|\{\Sigma_{ot}(X^{(t)})\}^{-\frac{1}{2}}\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta^k} - \frac{dm_t(X^{(t)},\alpha_o)}{dh}[\widetilde{w}_{h,j}]\right]\right\|_e^2\right]$$

converges to

$$E\left[\sum_{t=1}^{T}\left\|\{\Sigma_{ot}(X^{(t)})\}^{-\frac{1}{2}}\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta^k}-\frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o^k]\right]\right\|_e^2\right] \quad \text{as } j\to\infty.$$

Let

$$\widetilde{a}_{h,j}(z)=-\sum_{t=1}^{T}\left[\frac{dm_t(X^{(t)},\alpha_o)}{dh}[\widetilde{w}_{h,j}]\right]'\Sigma_{ot}(X^{(t)})^{-1}\varepsilon_t(z,\alpha_o).$$

Then $\widetilde{a}_{h,j}(\cdot)\in\mathbb{T}_h$ and hence $\widetilde{a}_{g,j}(z)\equiv$proj$(\ell_{\theta^k}(Z)-\widetilde{a}_{h,j}(Z)|\overline{\mathbb{T}}_g)\in\overline{\mathbb{T}}_g$, we have

$$\begin{aligned}&E\left[\sum_{t=1}^{T}\left\|\{\Sigma_{ot}(X^{(t)})\}^{-\frac{1}{2}}\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta^k}-\frac{dm_t(X^{(t)},\alpha_o)}{dh}[\widetilde{w}_{h,j}]\right]\right\|_e^2\right]\\ =\;& E\left[\ell_{\theta^k}(Z)-\widetilde{a}_{h,j}(Z)-\text{proj}(\ell_{\theta^k}(Z)-\widetilde{a}_{h,j}(Z)|\overline{\mathbb{T}}_g)\right]^2\geq E\left[\ell_{\theta^k}(Z)-a_h^*(Z)-a_g^*(Z)\right]^2. \qquad (33)\end{aligned}$$

Taking limit (as $j\to\infty$), and combining with the previous inequality (32), we obtain

$$\begin{aligned}&E\left(\left[\ell_{\theta^k}(Z)-\text{proj}(\ell_{\theta^k}(Z)|\overline{\mathbb{T}})\right]^2\right)\equiv E\left(\left[\ell_{\theta^k}(Z)-a_h^*(Z)-a_g^*(Z)\right]^2\right)\\ =\;& E\left[\sum_{t=1}^{T}\left\|\{\Sigma_{ot}(X^{(t)})\}^{-\frac{1}{2}}\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta^k}-\frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o^k]\right]\right\|_e^2\right].\end{aligned}$$

Denote $r_o=(r_o^1,...,r_o^{d_\theta})$ and

$$\frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o]=\left(\frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o^1],\frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o^2],...,\frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o^{d_\theta}]\right).$$

Then, an efficient score S_θ^* for θ_o is

$$\begin{aligned}S_\theta^* &\equiv \ell_\theta(Z)-\text{proj}(\ell_\theta(Z)|\overline{\mathbb{T}})\\ &= -\sum_{t=1}^{T}\left[\frac{dm_t(X^{(t)},\alpha_o)}{d\theta'}-\frac{dm_t(X^{(t)},\alpha_o)}{dh}[r_o]\right]'\Sigma_{ot}(X^{(t)})^{-1}\varepsilon_t(Z,\alpha_o).\end{aligned}$$

The semiparametric Fisher information bound for θ_o is $J_o\equiv E[S_\theta^*S_\theta^{*\prime}]$, and if J_o is non-singular, then the semiparametric efficient variance bound for θ_o is $\Omega_\theta^*\equiv(J_o)^{-1}\equiv(E[S_\theta^*S_\theta^{*\prime}])^{-1}$. ■

When we partition θ into $(\theta_1',\theta_2')'$ with $d_{\theta_i}=\dim(\theta_i)$ and $d_\theta=d_{\theta_1}+d_{\theta_2}$ for $i=1,2$, we let $r_{o\theta_1}\equiv(r_o^1,...,r_o^{d_{\theta_1}})\in\prod_{j=1}^{d_{\theta_1}}\overline{\mathcal{W}}$, $r_{o\theta_2}\equiv(r_o^{d_{\theta_1}+1},...,r_o^{d_\theta})\in\prod_{j=1}^{d_{\theta_2}}\overline{\mathcal{W}}$, $\frac{dm_1(X,\alpha_o)}{dh}[r_{o\theta_1}]\equiv(\frac{dm_1(X,\alpha_o)}{dh}[r_o^1],...,\frac{dm_1(X,\alpha_o)}{dh}[r_o^{d_{\theta_1}}])$

be $d_{\rho_1} \times d_{\theta_1}$, and $\frac{dm_2(X_2,\alpha_o)}{dh}[r_{o\theta_1}] \equiv (\frac{dm_2(X_2,\alpha_o)}{dh}[r_o^1], ..., \frac{dm_2(X_2,\alpha_o)}{dh}[r_o^{d_{\theta_1}}])$ be $d_{\rho_2} \times d_{\theta_1}$. Define $\frac{dm_1(X,\alpha_o)}{dh}[r_{o\theta_2}]$ and $\frac{dm_2(X^{(2)},\alpha_o)}{dh}[r_{o\theta_2}]$ accordingly. We have $S_\theta^* = (S_{\theta_1}^{*\prime}, S_{\theta_2}^{*\prime})'$. Then the semiparametric efficient variance bound of θ_{o1}, denoted as $\Omega_{\theta_1}^*$, is simply the inverse of the covariance matrix of the least squares projection residual of $S_{\theta_1}^*$ on $S_{\theta_2}^*$:

$$\Omega_{\theta_1}^* \equiv \left(E\{(S_{\theta_1}^* - b^{*\prime}S_{\theta_2}^*)(S_{\theta_1}^* - b^{*\prime}S_{\theta_2}^*)'\}\right)^{-1} = \left(E[S_{\theta_1}^* S_{\theta_1}^{*\prime}]\right)^{-1} + a^*\Omega_{\theta_2}^* a^{*\prime} \tag{34}$$

with $b^* \equiv \{E[S_{\theta_2}^* S_{\theta_2}^{*\prime}]\}^{-1}E[S_{\theta_2}^* S_{\theta_1}^{*\prime}]$ and $a^* \equiv \{E[S_{\theta_1}^* S_{\theta_1}^{*\prime}]\}^{-1}E[S_{\theta_1}^* S_{\theta_2}^{*\prime}]$. Then the semiparametric efficient variance bound of θ_{o2} is:

$$\Omega_{\theta_2}^* \equiv \left(E\{(S_{\theta_2}^* - a^{*\prime}S_{\theta_1}^*)(S_{\theta_2}^* - a^{*\prime}S_{\theta_1}^*)'\}\right)^{-1}. \tag{35}$$

Proof. (**Theorem 2.2**) Since the semiparametric efficient variance bound of θ_{o2} is $\Omega_{\theta_2}^*$ given by (35). We now show that $\Omega_{\theta_2}^* = (J_{o\theta_2})^{-1}$ for the plug-in model (2)-(3).

First we notice that for the plug-in model (2)-(3), the general expression of $S_\theta^* = (S_{\theta_1}^{*\prime}, S_{\theta_2}^{*\prime})'$ will take the following form: with $T = 2$, $X = X^{(2)}$, $\varepsilon = \varepsilon_1(Z, \alpha_o)$,

$$\begin{aligned} S_{\theta_2}^* &= \ell_{\theta_2}(Z) - \mathrm{proj}(\ell_{\theta_2}(Z)|\overline{\mathbb{T}}) \\ &= -\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} - \frac{dm_2(X,\alpha_o)}{dh}[r_{o\theta_2}]\right]' \Sigma_{o2}(X)^{-1}\rho_2(Z,\alpha_o) \\ &\quad -\left[\frac{dm_1(\alpha_o)}{d\theta_2'} - \frac{dm_1(\alpha_o)}{dh}[r_{o\theta_2}]\right]' \left(E\{\varepsilon\varepsilon'\}\right)^{-1}\varepsilon, \end{aligned}$$

$$\begin{aligned} S_{\theta_1}^* &= \ell_{\theta_1}(Z) - \mathrm{proj}(\ell_{\theta_1}(Z)|\overline{\mathbb{T}}) \\ &= \left[\frac{dm_2(X,\alpha_o)}{dh}[r_{o\theta_1}]\right]' \Sigma_{o2}(X)^{-1}\rho_2(Z,\alpha_o) \\ &\quad -\left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{dh}[r_{o\theta_1}]\right]' \left(E\{\varepsilon\varepsilon'\}\right)^{-1}\varepsilon, \end{aligned}$$

where $r_{o\theta_2}$ is such that, for all $r_{\theta_2} \in \prod_{j=1}^{d_{\theta_2}} \overline{\mathcal{W}}$

$$E\left\{ \begin{array}{c} \left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} - \frac{dm_2(X,\alpha_o)}{dh}[r_{o\theta_2}]\right]' \Sigma_{o2}(X)^{-1} \left[\frac{dm_2(X,\alpha_o)}{dh}[r_{\theta_2}]\right] + \\ \left[\frac{dm_1(\alpha_o)}{d\theta_2'} - \frac{dm_1(\alpha_o)}{dh}[r_{o\theta_2}]\right]' (E[\varepsilon\varepsilon'])^{-1} \left[\frac{dm_1(\alpha_o)}{dh}[r_{\theta_2}]\right] \end{array} \right\} = 0. \tag{36}$$

and $r_{o\theta_1}$ is such that, for all $r_{\theta_1} \in \prod_{j=1}^{d_{\theta_1}} \overline{\mathcal{W}}$

$$E\left\{ \begin{array}{c} -\left[\frac{dm_2(X,\alpha_o)}{dh}[r_{o\theta_1}]\right]' \Sigma_{o2}(X)^{-1} \left[\frac{dm_2(X,\alpha_o)}{dh}[r_{\theta_1}]\right] + \\ \left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{dh}[r_{o\theta_1}]\right]' (E[\varepsilon\varepsilon'])^{-1} \left[\frac{dm_1(\alpha_o)}{dh}[r_{\theta_1}]\right] \end{array} \right\} = 0. \tag{37}$$

Denote $w_2^* \equiv r_{o\theta_2} - r_{o\theta_1}a^* \in \prod_{j=1}^{d_{\theta_2}} \overline{\mathcal{W}}$ and compute "(36) $-$ $a^{*\prime}\times$(37)". Then we obtain: for all $r \in \prod_{j=1}^{d_{\theta_2}} \overline{\mathcal{W}}$,

$$E\left\{ \begin{array}{c} \left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} - \frac{dm_2(X,\alpha_o)}{dh}[w_2^*]\right]' \Sigma_{o2}(X)^{-1} \left[\frac{dm_2(X,\alpha_o)}{dh}[r]\right] + \\ \left[\frac{dm_1(\alpha_o)}{d\theta_2'} - \frac{dm_1(\alpha_o)}{d\theta_1'}a^* - \frac{dm_1(\alpha_o)}{dh}[w_2^*]\right]' (E[\varepsilon\varepsilon'])^{-1} \left[\frac{dm_1(\alpha_o)}{dh}[r]\right] \end{array} \right\} = 0.$$

Next by definition of a^*, we have for any non-zero $d_{\theta_1} \times d_{\theta_2}-$matrix a,

$$\begin{aligned} 0 &= E\left[\left(S_{\theta_2}^* - a^{*\prime}S_{\theta_1}^*\right) S_{\theta_1}^{*\prime}a\right] \\ &= -\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} - \frac{dm_2(X,\alpha_o)}{dh}[w_2^*]\right]' \Sigma_{o2}(X)^{-1} \left[\frac{dm_2(X,\alpha_o)}{dh}[r_{o\theta_1}a]\right] \\ &\quad + \left[\frac{dm_1(\alpha_o)}{d\theta_2'} - \frac{dm_1(\alpha_o)}{d\theta_1'}a^* - \frac{dm_1(\alpha_o)}{dh}[w_2^*]\right]' (E\{\varepsilon\varepsilon'\})^{-1} \left[\frac{dm_1(\alpha_o)}{d\theta_1'}a - \frac{dm_1(\alpha_o)}{dh}[r_{o\theta_1}a]\right]. \end{aligned}$$

Hence

$$0 = \left[\frac{dm_1(\alpha_o)}{d\theta_2'} - \frac{dm_1(\alpha_o)}{d\theta_1'}a^* - \frac{dm_1(\alpha_o)}{dh}[w_2^*]\right]' (E\{\varepsilon\varepsilon'\})^{-1} \left[\frac{dm_1(\alpha_o)}{d\theta_1'}\right].$$

Using the condition that $\frac{dm_1(\alpha_o)}{d\theta_1'}$ is invertible, we obtain that a^* satisfies (9) and

$$\left[\frac{dm_1(\alpha_o)}{d\theta_2'} - \frac{dm_1(\alpha_o)}{d\theta_1'}a^* - \frac{dm_1(\alpha_o)}{dh}[w_2^*]\right] = 0.$$

Thus

$$E\left\{\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'}-\frac{dm_2(X,\alpha_o)}{dh}[w_2^*]\right]'\Sigma_{o2}(X)^{-1}\left[\frac{dm_2(X,\alpha_o)}{dh}[r]\right]\right\}=0\quad\text{for all } r\in\prod_{j=1}^{d_{\theta_2}}\overline{\mathcal{W}}.$$

Hence $w_2^*\in\prod_{j=1}^{d_{\theta_2}}\overline{\mathcal{W}}$ solves

$$\inf_{w\in\prod_{j=1}^{d_{\theta_2}}\overline{\mathcal{W}}}E\left\{\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'}-\frac{dm_2(X,\alpha_o)}{dh}[w]\right]'\Sigma_{o2}(X)^{-1}\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'}-\frac{dm_2(X,\alpha_o)}{dh}[w]\right]\right\}.$$

Recall that $w_{o2}\in\prod_{j=1}^{d_\theta}\overline{\mathcal{W}}$ defined in (7) also solves the above minimization problem. Under Assumptions 1s, 2 and 3, we have $w_{o2}=w_2^*\equiv r_{o\theta_2}-r_{o\theta_1}a^*$,

$$J_{o\theta_2}=E\left\{\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'}-\frac{dm_2(X,\alpha_o)}{dh}[w_{o2}]\right]'\Sigma_{o2}(X)^{-1}\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'}-\frac{dm_2(X,\alpha_o)}{dh}[w_{o2}]\right]\right\},$$

and

$$S_{\theta_2}^*-a^{*\prime}S_{\theta_1}^*=\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'}-\frac{dm_2(X,\alpha_o)}{dh}[w_{o2}]\right]'\Sigma_{o2}(X)^{-1}\rho_2(Z,\alpha_o),$$

Hence $\Omega_{\theta_2}^*\equiv\left(E\{(S_{\theta_2}^*-a^{*\prime}S_{\theta_1}^*)(S_{\theta_2}^*-a^{*\prime}S_{\theta_1}^*)'\}\right)^{-1}=(J_{o\theta_2})^{-1}$. ■

Proof. (**Theorem 2.3**) Since the semiparametric efficient variance bound of θ_{o1} is $\Omega_{\theta_1}^*$ given by (34). We now show that $\Omega_{\theta_1}^*=(J_{o\theta_1})^{-1}$ for the plug-in model (2)-(3). Recall that $b^*\equiv\{E[S_{\theta_2}^*S_{\theta_2}^{*\prime}]\}^{-1}E[S_{\theta_2}^*S_{\theta_1}^{*\prime}]$. Denote $w_1^*\equiv r_{o\theta_1}-r_{o\theta_2}b^*\in\prod_{j=1}^{d_{\theta_1}}\overline{\mathcal{W}}$ and compute "(37) $-$ $b^{*\prime}\times$(36)". Then we obtain: for all $r\in\prod_{j=1}^{d_{\theta_1}}\overline{\mathcal{W}}$,

$$E\left\{\begin{array}{c}\left[-\frac{dm_2(X,\alpha_o)}{d\theta_2'}b^*-\frac{dm_2(X,\alpha_o)}{dh}[w_1^*]\right]'\Sigma_{o2}(X)^{-1}\left[\frac{dm_2(X,\alpha_o)}{dh}[r]\right]+\\ \left[\frac{dm_1(\alpha_o)}{d\theta_1'}-\frac{dm_1(\alpha_o)}{d\theta_2'}b^*-\frac{dm_1(\alpha_o)}{dh}[w_1^*]\right]'(E[\varepsilon\varepsilon'])^{-1}\left[\frac{dm_1(\alpha_o)}{dh}[r]\right]\end{array}\right\}=0. \tag{38}$$

Also $w_1^*\equiv r_{o\theta_1}-r_{o\theta_2}b^*$ implies:

$$\begin{aligned}&S_{\theta_1}^*-b^{*\prime}S_{\theta_2}^*\\ =\;&\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'}b^*+\frac{dm_2(X,\alpha_o)}{dh}[w_2^*]\right]'\Sigma_{o2}(X)^{-1}\rho_2(Z,\alpha_o)\\ &-\left[\frac{dm_1(\alpha_o)}{d\theta_1'}-\frac{dm_1(\alpha_o)}{d\theta_2'}b^*-\frac{dm_1(\alpha_o)}{dh}[w_1^*]\right]'(E[\varepsilon\varepsilon'])^{-1}\varepsilon.\end{aligned}$$

By definition of b^* we have for any non-zero $d_{\theta_2} \times d_{\theta_1}-$matrix,

$$\begin{array}{rcl} 0 & = & E\{(S^*_{\theta_1} - b^{*\prime} S^*_{\theta_2}) S^{*\prime}_{\theta_1} b\} \\ & = & E\left\{ \begin{array}{c} -\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} b^* + \frac{dm_2(X,\alpha_o)}{dh}[w_1^*]\right]' \Sigma_{o2}(X)^{-1} \left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} b - \frac{dm_2(X,\alpha_o)}{dh}[r_{o\theta_2} b]\right] + \\ \left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{d\theta_2'} b^* - \frac{dm_1(\alpha_o)}{dh}[w_1^*]\right]' E[\varepsilon\varepsilon']^{-1} \left[\frac{dm_1(\alpha_o)}{d\theta_2'} b - \frac{dm_1(\alpha_o)}{dh}[r_{o\theta_2} b]\right] \end{array} \right\}. \end{array}$$

This and equation (38) imply that

$$E\left\{ \begin{array}{c} -\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} b^* + \frac{dm_2(X,\alpha_o)}{dh}[w_1^*]\right]' \Sigma_{o2}(X)^{-1} \left[\frac{dm_2(X,\alpha_o)}{d\theta_2'}\right] + \\ \left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{d\theta_2'} b^* - \frac{dm_1(\alpha_o)}{dh}[w_1^*]\right]' E[\varepsilon\varepsilon']^{-1} \left[\frac{dm_1(\alpha_o)}{d\theta_2'}\right] \end{array} \right\} = 0. \tag{39}$$

Equations (38) and (39) imply that

$$E\left\{ \begin{array}{c} -\left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} b^* + \frac{dm_2(X,\alpha_o)}{dh}[w_1^*]\right]' \Sigma_{o2}(X)^{-1} \left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} b^* + \frac{dm_2(X,\alpha_o)}{dh}[w_1^*]\right] + \\ \left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{d\theta_2'} b^* - \frac{dm_1(\alpha_o)}{dh}[w_1^*]\right]' E[\varepsilon\varepsilon']^{-1} \left[\frac{dm_1(\alpha_o)}{d\theta_2'} b^* + \frac{dm_1(\alpha_o)}{dh}[w_1^*]\right] \end{array} \right\} = 0,$$

hence

$$\begin{array}{rl} & E\{(S^*_{\theta_1} - b^{*\prime} S^*_{\theta_2})(S^*_{\theta_1} - b^{*\prime} S^*_{\theta_2})'\} \\ = & E\left\{ \begin{array}{c} \left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} b^* + \frac{dm_2(X,\alpha_o)}{dh}[w_1^*]\right]' \Sigma_{o2}(X)^{-1} \left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} b^* + \frac{dm_2(X,\alpha_o)}{dh}[w_1^*]\right] \\ + \left\| \{E(\varepsilon\varepsilon')\}^{-\frac{1}{2}} \left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{d\theta_2'} b^* - \frac{dm_1(\alpha_o)}{dh}[w_1^*]\right] \right\|_e^2 \end{array} \right\} \\ = & \left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{d\theta_2'} b^* - \frac{dm_1(\alpha_o)}{dh}[w_1^*]\right]' E[\varepsilon\varepsilon']^{-1} \frac{dm_1(\alpha_o)}{d\theta_1'}. \end{array}$$

Equations (38) and (39) also imply that $(b^*, w_1^*) \in \prod_{j=1}^{d_{\theta_1}} (\mathcal{R}^{d_{\theta_2}} \times \overline{\mathcal{W}})$ solves

$$J_{o\theta_1} \equiv \inf_{(b,r) \in \prod_{j=1}^{d_{\theta_1}} (\mathcal{R}^{d_{\theta_2}} \times \overline{\mathcal{W}})} E\left\{ \begin{array}{c} \left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} b + \frac{dm_2(X,\alpha_o)}{dh}[r]\right]' \Sigma_{o2}(X)^{-1} \left[\frac{dm_2(X,\alpha_o)}{d\theta_2'} b + \frac{dm_2(X,\alpha_o)}{dh}[r]\right] \\ + \left\| \{E(\varepsilon\varepsilon')\}^{-\frac{1}{2}} \left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{d\theta_2'} b - \frac{dm_1(\alpha_o)}{dh}[r]\right] \right\|_e^2 \end{array} \right\}.$$

Thus with $b^* \equiv \{E[S^*_{\theta_2} S^{*\prime}_{\theta_2}]\}^{-1} E[S^*_{\theta_2} S^{*\prime}_{\theta_1}]$ and $w_1^* \equiv r_{o\theta_1} - r_{o\theta_2} b^*$, we have:

$$\begin{aligned} J_{o\theta_1} &= \left[\frac{dm_1(\alpha_o)}{d\theta_1'} - \frac{dm_1(\alpha_o)}{d\theta_2'} b^* - \frac{dm_1(\alpha_o)}{dh}[w_1^*]\right]' E[\varepsilon\varepsilon']^{-1} \frac{dm_1(\alpha_o)}{d\theta_1'} \\ &= E\{(S^*_{\theta_1} - b^{*\prime} S^*_{\theta_2})(S^*_{\theta_1} - b^{*\prime} S^*_{\theta_2})'\}. \end{aligned}$$

■