Computing semiparametric efficiency bounds in linear models with nonparametric regressors

Andrés Aradillas-López[*†]

September 14, 2019

Abstract

We use the computational method proposed by Severini and Tripathi (2001) to obtain semiparametric efficiency bounds in linear models with nonparametric regressors in the form of conditional expectations. Examples include social-interaction models. Explicit efficiency bounds for these models, with the degree of generality assumed here, had not been described before.
JEL Codes: C14.

1 Introduction

This note computes the semiparametric efficiency bound for linear models that include nonparametric regressors –in this case, conditional expectations. Notable examples include social-interactions models. Our derivation uses the method of "representers" proposed by Severini and Tripathi (2001). The bound described here is new in the literature given the level of generality assumed.

2 Computing efficiency bounds using representers: an outline

Here we will outline the approach we will use to compute the bounds. Our discussion follows Section 2 in Severini and Tripathi (2001) (henceforth ST).

Notational conventions

We will let $\mathbb{S}(z)$ denote the support of a random variable z. λ will denote the Lebesgue measure and $L^2(S,\lambda)$, the set of all real-valued functions on S that are square integrable with respect to

[*]email: aaradill@psu.edu. Pennsylvania State University. Department of Economics, 518 Kern Graduate Building, University Park, PA 16802

[†]The author is grateful to an anonymous referee for the useful suggestions and comments.

Lebesgue measure. For a random variable z, we will let $L^2(S, \lambda_z)$ denote the set of all functions defined on S which are square integrable with respect to the probability distribution of z.

Let $z_1, \dots, z_n$ be $d \times 1$ iid random vectors with Lebesgue density $p_0(z)$. Assume for simplicity that p_0 has full support[1] on $\mathbb{R}^d$ and let us express[2] $p_0(z) = \tau_0^2(z)$, with $\tau_0 \in \Gamma$ and Γ is a subset of the unit ball in $L^2(\mathbb{R}^d; \lambda)$. Assume for now that τ_0 is an unknown function and therefore an infinite-dimensional parameter. In the models we will study below, τ_0 itself will be a functional of other parameters, both finite and infinite-dimensional. Working with $\tau_0 = \sqrt{p_0}$ has the advantage that $\tau_0 \in L^2(\mathbb{R}^d; \lambda)$ while this may not be the case for p_0 itself.

Denote the parameter of interest as $\rho(\tau_0) \in \mathbb{R}$, where ρ is a pathwise differentiable functional and let $\nabla\rho$ denote the pathwise derivative of ρ. Ultimately, our focus will be a finite-dimensional parameter vector θ_0, in which case $\rho(\tau_0) = c'\theta_0$, where c is an arbitrary vector[3] . The objective is to obtain efficiency bounds for regular estimators of $\rho(\tau_0)$. Regular estimators are defined in Newey (1990, page 102). In essence, they require that the asymptotic distribution of the estimator be stable in a neighborhood of the true model (i.e, in a neighborhood of τ_0).

The method described in ST for computing efficiency bounds is based on the intuition provided by Stein (1956), who introduced the notion of efficiency bounds by noting that the problem of estimating a real-valued parameter with nonparametric components is at least as difficult (to first order of approximation) as any one-dimensional subproblem contained in it. Fix some $t_0 > 0$ and let $t \mapsto \tau_t$ denote a curve from $[0, t_0]$ on to Γ that passes through τ_0 at $t = 0$ (i.e, $\tau_t|_{t=0} = \tau_0$). Let $\dot{\tau}$ denote the slope of τ_t at $t = 0$. $\dot{\tau}$ is an element of the vector space $L^2(\mathbb{R}^d; \lambda)$ which is tangent[4] to Γ at τ_0. Let $T(\Gamma, \tau_0)$ denote the tangent cone that consists of all $\dot{\tau}$'s that are tangent to Γ at τ_0. Finally, let $\overline{lin\ \mathrm{T}(\Gamma, \tau_0)}$ denote the smallest closed (in the $L^2(\mathbb{R}^d; \lambda)$ norm) linear space containing $T(\Gamma, \tau_0)$.

Let $\ell_z(t) = \log \tau_t^2(z)$. The score and the Fisher information for estimating $t = 0$ are given, respectively by

$$S_0(z) = \left.\frac{d\ell_z(t)}{dt}\right|_{t=0} = \frac{2\dot{\tau}(z)}{\tau_0(z)} \quad \text{and} \quad i_F = \int_{\mathbb{R}^d} S_0^2(z)\tau_0^2(z)dz = 4\int_{\mathbb{R}^d} \dot{\tau}^2(z)dz.$$

[1]As the discussion that follows will illustrate, bounded support or point-masses can be readily incorporated into the analysis.

[2]While defining $\tau_0^2(z) = p_0(z)$ yields two solutions: $\tau_0(z) = \pm\sqrt{p_0(z)}$, we specifically define $\tau_0(z) = \sqrt{p_0(z)}$.

[3]This shows that focusing on the case where $\rho(\tau_0)$ is a scalar can be done without loss of generality in the case that will preoccupy us.

[4]Suppose M is a subset of a normed vector space $(X, \|\cdot\|_X)$. Take a point $x_0 \in M$. We say that a vector $\dot{x} \in X$ is tangent to M at x_0 if there exists $t_0 > 0$ and a mapping $t \mapsto r_t$ into X satisfying $\|r_t\| = o(t)$ as $t \downarrow 0$, such that $x_t \equiv x_0 + t\dot{x} + r_t \in M \ \forall\ t \in [0, t_0]$. The curve $t \mapsto x_t$ passes through x_0 at $t = 0$ and $\dot{x}$ is the slope of this curve at $t = 0$.

ST equip $\overline{lin\ \mathrm{T}(\Gamma,\tau_0)}$ with the Fisher-information inner product $\langle\cdot,\cdot\rangle_F$ defined as

$$\langle\dot\tau_1,\dot\tau_2\rangle_F = 4\int_{\mathbb{R}^d}\dot\tau_1(z)\dot\tau_2(z)dz \quad \forall\ \dot\tau_1,\dot\tau_1 \in \overline{lin\ \mathrm{T}(\Gamma,\tau_0)}.$$

We will use $avar(\widehat{A}_n)$ to denote the asymptotic variance of $\widehat{A}_n$. Let $\widehat{t}_n$ be any regular, $\sqrt{n}-$consistent estimator of $t = 0$ in the subproblem given by τ_t. The information inequality implies that $avar\left\{\sqrt{n}\cdot\widehat{t}_n\right\} \geq 1/i_F = \|\dot\tau\|_F^{-2}$. Next, since τ_t is ultimately a device to compute efficiency bounds, we should focus on subproblems that are informative about our parameter of interest $\rho(\tau_0)$. To this end, normalize ρ and reparameterize τ_t so that $\rho(\tau_t) = t$ for $t \in [0,t_0]$. Thus, estimating $t = 0$ will be equivalent to estimating $\rho(\tau_0)$. It follows that, for all the subproblems of interest, $avar\left\{\sqrt{n}\left[\rho(\tau_{\widehat{t}_n}) - \rho(\tau_0)\right]\right\} = avar\left\{\sqrt{n}\cdot\widehat{t}_n\right\} \geq \|\dot\tau\|_F^{-2}$. Next, by definition, $\nabla\rho$ is a continuous linear functional[5] and, for the suproblems we are interested in, it satisfies $\nabla\rho(\dot\tau) = 1$ (implying that $\dot\tau \neq 0$). Refer to such $\dot\tau$'s as *feasible.*

Thus, in searching for the lower bound (*l.b.*), we would look to maximize $\|\dot\tau\|_F^{-2}$ over those $\dot\tau$'s in $\overline{lin\ \mathrm{T}(\Gamma,\tau_0)}$ that satisfy $\dot\tau \neq 0$ and $\nabla\rho(\dot\tau) = 1$. That is,

$$l.b. = sup\left\{\|\dot\tau\|_F^{-2} : \dot\tau \in \overline{lin\ \mathrm{T}(\Gamma,\tau_0)},\ \dot\tau \neq 0,\ \nabla\rho(\dot\tau) = 1\right\}.$$

Suppose $\nabla\rho(\dot\tau)$ is a nonzero constant (a property shared by all feasible $\dot\tau$'s). Then, $\widetilde{\tau} \equiv \dot\tau/\nabla\rho(\dot\tau) \in \overline{lin\ \mathrm{T}(\Gamma,\tau_0)}$. In our search for *l.b.* we can focus on such $\widetilde{\tau}$'s. Since $\nabla\rho$ is a linear functional, we have $\nabla\rho(\widetilde{\tau}) = 1$ and therefore $\widetilde{\tau}$ is feasible. Furthermore, linearity of $\nabla\rho$ implies that

$$\|\widetilde{\tau}\|_F^{-1} = \left\|\frac{\dot\tau}{\nabla\rho(\dot\tau)}\right\|_F^{-1} = \frac{|\nabla\rho(\dot\tau)|}{\|\dot\tau\|_F} = \left|\nabla\rho\left(\frac{\dot\tau}{\|\dot\tau\|_F}\right)\right|$$

Obviously, we have $\left\|\frac{\dot\tau}{\|\dot\tau\|_F}\right\|_F = 1$. Therefore, going back to the notation of $\dot\tau$ instead of $\widetilde{\tau}$, the lower bound *l.b.* can be re-expressed as

$$l.b. = sup\left\{|\nabla\rho(\dot\tau)|^2 : \dot\tau \in \overline{lin\ \mathrm{T}(\Gamma,\tau_0)},\ \dot\tau \neq 0,\ \|\dot\tau\|_F = 1\right\}.$$

Since $\nabla\rho$ is a continuous linear functional on the tangent space $\overline{lin\ \mathrm{T}(\Gamma,\tau_0)}$ equipped with $\|\cdot\|_F$, its norm (see Luenberger (1969, Section 5.2)) is given by

$$\|\nabla\rho\|_* = sup\left\{|\nabla\rho(\dot\tau)| : \dot\tau \in \overline{lin\ \mathrm{T}(\Gamma,\tau_0)},\ \dot\tau \neq 0,\ \|\dot\tau\|_F = 1\right\}.$$

Therefore, $l.b. = \|\nabla\rho\|_*^2$. The key insight in ST is that the problem of computing *l.b* can be

[5]Let M, $\dot x$ and x_t be as defined in footnote 4. A functional $\rho : M \to \mathbb{R}$ is said to be pathwise differentiable at x_0 if, for any x_t, there exists a continuous linear functional $\nabla : X \to \mathbb{R}$ such that $\left|\frac{\rho(x_t)-\rho(x_0)}{t} - \nabla\rho(\dot x)\right| \to 0$ as $t \downarrow 0$.

solved by invoking the Riesz-Fréchet Theorem (R-F Theorem henceforth) which states[6] that, since $\left(\overline{lin\ \mathrm{T}(\Gamma,\tau_0)}, \langle\cdot,\cdot\rangle_F\right)$ is a Hilbert space and $\nabla\rho$ is a continuous, linear functional defined in it, there exists a *unique* $\tau^* \in \overline{lin\ \mathrm{T}(\Gamma,\tau_0)}$ such that

$$\nabla\rho(\dot{\tau}) = \langle\tau^*, \dot{\tau}\rangle_F \ \ \forall\ \dot{\tau} \in \overline{lin\ \mathrm{T}(\Gamma,\tau_0)} \quad \text{and} \quad \|\nabla\rho\|_* = \|\tau^*\|_F\,. \tag{R-F}$$

τ^* is called the *representer* of the linear functional $\nabla\rho$. Thus, computing *l.b.* is done in two steps:

Step 1: Find the representer τ^* by solving the condition (R-F).

Step 2: Compute $l.b. = \|\tau^*\|_F^2$

ST illustrate how this computational method can be used to recover the efficiency bound in a number of well-known econometric models (partially linear model, models with unconditional and conditional moment restrictions, the binary choice model and density-weighted average derivatives) whose bounds were derived previously by a variety of ad-hoc approaches.

3 A linear econometric model with conditional expectations as regressors

Consider the following model,

$$y = x'\beta_0 + E[s|z]'\gamma_0 + \varepsilon \equiv x'\beta_0 + \mu(z)'\gamma_0 + \varepsilon, \tag{1}$$

where $x \in \mathbb{R}^{d_x}$, $s \in \mathbb{R}^{d_s}$ and $z \in \mathbb{R}^{d_z}$. Denote $\omega \equiv (x', s', z')' \in \mathbb{R}^{d_\omega}$ and $v \equiv (x'\mu(z)')' \in \mathbb{R}^d$, where $d \equiv d_x + d_s$. Model (1) can therefore be written as $y = v'\theta_0 + \varepsilon$. We observe $(y, \omega')'$ but not ε and we treat $\mu(z)$ as a nonparametric regressor. The parameter vector of interest is $\theta_0 \equiv (\beta_0', \gamma_0')' \in \mathbb{R}^d$. Our goal is to characterize the efficiency bound for $\sqrt{n}$-consistent, regular estimators of θ_0 when the regressor μ is nonparametrically specified and the distribution of $\varepsilon|\omega$ is unknown but assumed to satisfy some qualitative conditions which we will describe below.

Example: A social-interactions model

The model described in (1) can encompass examples of social interaction models of the type studied, for example, in Manski (1993), Manski (1995, Section 7.2) and Brock and Durlauf (2001, Section 2.6). Consider a population of agents whose choice of y is given by

$$y = x'\beta_0 + \delta_0'\widehat{E}[u|z] + \alpha_0\widehat{E}[y|z] + \varepsilon, \quad \text{with} \quad \alpha_0 \neq 1$$

[6] See Luenberger (1969, Section 5.3, Theorem 2) or Young (1988, Theorem 6.8).

The operator $\widehat{E}$ denotes agents' subjective expectations. In the social-interactions literature, α_0 measures an endogenous "social effect" while δ_0 measures "contextual effects" and z describes "reference" characteristics. Suppose beliefs are not observed in the data but we assume that agents use *rational expectations* in their construction. This implies that subjective expectations are consistent with the true data generating process. Assuming $E[\varepsilon|z]=0$ and solving for $E[y|z]$, we have

$$y = x'\beta_0 + \frac{\alpha_0\beta_0' E[x|z]}{1-\alpha_0} + \frac{\delta_0' E[u|z]}{1-\alpha_0} + \varepsilon \equiv x'\beta_0 + E[s|z]'\gamma_0 + \varepsilon,$$

where $s \equiv (x', u')'$ and $\gamma_0 = (\alpha_0\beta_0'/(1-\alpha_0),\, \delta_0'/(1-\alpha_0))'$.

4 Semiparametric efficiency bound

We describe our maintained assumptions next.

Assumption 1 *The support of v is not contained in any proper linear subspace of $\mathbb{R}^d$. Denote $Pr(\varepsilon \le \epsilon|\omega) = Pr(\varepsilon \le \epsilon|\omega) \equiv G_0(\epsilon|\omega)$, with corresponding conditional density given by $g_0^2(\epsilon|\omega)$. This is unknown but assumed to satisfy $E[\Upsilon(\varepsilon)|\omega]=0$ for a known function $\Upsilon \in \mathbb{R}^\ell$. Furthermore, $E[\Upsilon(\varepsilon)\Upsilon(\varepsilon)'|\omega]$ is invertible w.p.1. Define*

$$\begin{aligned}\mathcal{G} =\Bigg\{ & g \in L^2(\mathbb{R}\times\mathbb{R}^d;\lambda\times\lambda_\omega):\ g^2(\epsilon|\omega)>0,\ g(\epsilon|\omega) \text{ is bounded, continuous and differentiable w.p.1},\\ & g'(\cdot|\omega) \equiv \frac{dg(\cdot|\omega)}{d\varepsilon}\text{:}\ \ 0 < \int_{\mathbb{R}} \big[g(\epsilon|\omega) + \epsilon g'(\epsilon|\omega)\big]^2 d\epsilon < \infty,\quad \int_{\mathbb{R}} g^2(\epsilon|\omega)d\epsilon = 1,\quad \int_{\mathbb{R}} \big\|\Upsilon(\epsilon)\big\|^2 g^2(\epsilon|\omega)d\epsilon < \infty,\\ & \text{and} \quad \int_{\mathbb{R}} \Upsilon(\epsilon) g^2(\epsilon|\omega)d\epsilon = 0 \ \text{w.p.1.}\Bigg\}.\end{aligned}$$

Then, $g_0 \in \mathcal{G}$. The characterization of $\mathcal{G}$ is meant to ensure that $\lim_{|\epsilon|\to\infty} g_0^2(\epsilon|\omega) = 0$ w.p.1, and, in particular, $\int_{-\infty}^{\infty} g_0'(\varepsilon|\omega)g_0(\varepsilon|\omega)d\varepsilon = 0$ w.p.1.
Let $h_0^2(\cdot|s,z)$ denote the conditional density of x given s,z. And let $f_0^2(\cdot|z)$ and $m_0^2(\cdot)$ denote the conditional density of s given z and the marginal density of z, respectively. Define

$$\mathcal{H} =\Bigg\{ h \in L^2(\mathbb{R}^{d_x}\times\mathbb{R}^{d_s+d_z};\lambda\times\lambda_{s,z}):\ h^2(x|s,z)>0,\ \int_{\mathbb{R}^{d_x}} h^2(x|s,z)ds = 1 \ \text{w.p.1.}\Bigg\};$$

$$\begin{aligned}\mathcal{F} =&\Bigg\{ f \in L^2(\mathbb{R}^{d_s}\times\mathbb{R}^{d_z};\lambda\times\lambda_z):\ f^2(s|z)>0,\ \int_{\mathbb{R}^{d_s}} f^2(s|z)ds = 1 \ \text{w.p.1.}\Bigg\};\\ \mathcal{M} =&\Bigg\{ m \in L^2(\mathbb{R}^{d_z};\lambda):\ m^2(z)>0,\ \int_{\mathbb{R}^{d_z}} m^2(z)dz = 1\Bigg\}.\end{aligned}$$

Then, $f_0 \in \mathcal{F}$, $h_0 \in \mathcal{H}$ and $m_0 \in \mathcal{M}$.

The unknown parameters of the model are $\tau_0 = (\theta_0, g_0, f_0, m_0)$. The nonparametric regressors μ are functionals of f_0. The assumption that $E[\Upsilon(\varepsilon)|\omega] = 0$ for a known Υ allows us to incorporate multiple cases of interest. For example,

- Mean-independence: $E[\varepsilon|\omega] = 0$, by letting $\Upsilon(\varepsilon) = \epsilon$.
- Quantile-independence: $Pr(\varepsilon \leq 0|\omega) = \kappa$ for a known κ, by letting $\Upsilon(\varepsilon) = \mathbb{1}\{\varepsilon \leq 0\} - \kappa$.
- Mean and quantile-independence, by letting $\Upsilon(\varepsilon) = (\varepsilon,\ \mathbb{1}\{\varepsilon \leq 0\} - \kappa)'$.

Assumption 1 focuses, for simplicity, on the case where all the components in ω (and in particular, z) are continuously distributed. The steps in the proof of our main result will show how to extend this to cases where these regressors have point masses.

Remark 1 *Rilstone (1993) describes efficiency bounds in a linear model with nonparametric regressors under the assumption that (a) these can be approximated arbitrarily well with a series function, and (b) the additive shock ε is independent of all the other explanatory variables and is Normally distributed with known variance. Our setting is much more general.*

Using the same arguments as Lemmas B.1 and B.2 in ST, the tangent spaces for $\mathcal{G}$, $\mathcal{H}$, $\mathcal{F}$ and $\mathcal{M}$ can be shown to be as follows,

$$\begin{aligned}
\overline{lin\ T(\mathcal{G}, g_0)} &= \left\{ \dot{g} \in L^2(\mathbb{R} \times \mathbb{R}^d; \lambda \times \lambda_\omega)\colon\ \int_{\mathbb{R}} \dot{g}(\epsilon|\omega) g_0(\epsilon|\omega) d\epsilon = 0,\ \int_{\mathbb{R}} \Upsilon(\epsilon) \dot{g}(\epsilon|\omega) g_0(\epsilon|\omega) d\epsilon = 0 \ \text{ w.p.1.} \right\} \\
\overline{lin\ T(\mathcal{H}, h_0)} &= \left\{ \dot{h} \in L^2(\mathbb{R}^{d_x} \times \mathbb{R}^{d_s+d_z}; \lambda \times \lambda_{s,z})\colon\ \int_{\mathbb{R}^{d_x}} \dot{h}(x|s,z) h_0(x|s,z) dx = 0 \ \text{ w.p.1.} \right\} \\
\overline{lin\ T(\mathcal{F}, f_0)} &= \left\{ \dot{f} \in L^2(\mathbb{R}^{d_s} \times \mathbb{R}^{d_z}; \lambda \times \lambda_z)\colon\ \int_{\mathbb{R}^{d_s}} \dot{f}(s|z) f_0(s|z) ds = 0 \ \text{ w.p.1.} \right\} \\
\overline{lin\ T(\mathcal{M}, m_0)} &= \left\{ \dot{m} \in L^2(\mathbb{R}^{d_z}; \lambda)\colon\ \int_{\mathbb{R}^{d_z}} \dot{m}(z) m_0(z) dz = 0 \right\}.
\end{aligned} \tag{2}$$

Let $\dot{\tau} = (\dot{\theta}, \dot{g}, \dot{h}, \dot{f}, \dot{m})$. This vector belongs to the product tangent space

$$\dot{\mathscr{T}} = \mathbb{R}^d \times \overline{lin\ T(\mathcal{G}, g_0)} \times \overline{lin\ T(\mathcal{H}, h_0)} \times \overline{lin\ T(\mathcal{F}, f_0)} \times \overline{lin\ T(\mathcal{M}, m_0)}.$$

Proposition 1 *Let*

$$\eta(\omega, \varepsilon) = 2\frac{g_0'(\varepsilon|\omega)}{g_0(\varepsilon|\omega)}, \quad \mathcal{C}(\omega) = E\left[\Upsilon(\varepsilon)\Upsilon(\varepsilon)'|\omega\right]^{-1} E\left[\Upsilon(\varepsilon)\eta(\omega, \varepsilon)|\omega\right] \quad \textit{and} \quad P_\Upsilon(\omega, \varepsilon) = \mathcal{C}(\omega)'\Upsilon(\varepsilon).$$

$P_\Upsilon(\omega, \varepsilon)$ is the orthogonal projection, conditional on ω, of $\eta(\omega, \varepsilon)$ onto $col(\Upsilon(\varepsilon))$ (the column space of $\Upsilon(\varepsilon)$). Let $r(\omega, \varepsilon) \equiv \eta(\omega, \varepsilon) - P_\Upsilon(\omega, \varepsilon)$ denote the residual of this projection. Let

$$M(z) = \left(I_{d_s} + E[\eta(\omega, \varepsilon)^2|z] \cdot Var[s|z]\gamma_0\gamma_0'\right)^{-1} Var[s|z]\gamma_0\gamma_0',$$

where I_{d_s} is the $d_s \times d_s$ identity matrix. Let

$$\Phi^*(\omega,\varepsilon) = -\left(v - \frac{E[v \cdot P_\Upsilon(\omega,\varepsilon)^2|z] \cdot tr(M(z))}{1 - E[r(\omega,\varepsilon)^2|z] \cdot tr(M(z))}\right) \cdot P_\Upsilon(\omega,\varepsilon) \tag{3}$$

and $\Omega(z) = E\left[\Phi^*(\omega,\varepsilon)\eta(\omega,\varepsilon)|z\right]$. Let

$$\Sigma_\theta^* = E\left[\Phi^*(\omega,\varepsilon)\Phi^*(\omega,\varepsilon)'\right] + E\left[\Omega(z)\gamma_0' Var[s|z]\gamma_0\Omega(z)'\right] \tag{4}$$

If Σ_θ^* is invertible, the semiparametric efficiency bound for $\sqrt{n}$-consistent, regular estimators of θ_0 in Model 1 under Assumption 1 is well-defined and is equal to Σ_θ^{*-1} .

Example: Normally distributed shocks with mean-independence

Suppose $E[\varepsilon|\omega] = 0$ (i.e, $\Upsilon(\varepsilon) = \varepsilon$). In this case, if $\varepsilon|\omega \sim \mathcal{N}(0,\sigma^2(\omega))$ we have $\eta(\omega,\varepsilon) = -\frac{\varepsilon}{\sigma^2(\omega)}$ and therefore $P_\Upsilon(\omega,\varepsilon) = -\frac{\varepsilon}{\sigma^2(\omega)}$ and $r(\omega,\varepsilon) = 0$. Finally, Φ^* and Ω in (3) become

$$\begin{aligned}
\Phi^*(\omega,\varepsilon) &= \left(v - E\left[\frac{v}{\sigma^2(\omega)}\middle|z\right] \cdot tr(M(z))\right) \cdot \frac{\varepsilon}{\sigma^2(\omega)}, \\
\Omega(z) &= E\left[\frac{v}{\sigma^2(\omega)}\middle|z\right] \cdot \left(tr(M(z)) \cdot E\left[\frac{1}{\sigma^2(\omega)}\middle|Z\right] - 1\right),
\end{aligned}$$

with $M(z) = \left(I_{d_s} + E\left[\frac{1}{\sigma^2(\omega)}\middle|z\right] \cdot Var[s|z]\gamma_0\gamma_0'\right)^{-1} Var[s|z]\gamma_0\gamma_0'$.

Proof of Proposition 1

We have

$$\begin{aligned}
S_0 &= \frac{d}{dt}\left[\log p_t^2(y|\omega) + \log h_t^2(x|s,z) + \log f_t^2(s|z) + \log m_t^2(z)\right]\Big|_{t=0} \\
&= 2\left[\frac{\dot{g}(\varepsilon|\omega) - g_0'(\varepsilon|\omega) \cdot (v'\dot{\theta} + \gamma_0'\dot{\mu}(z))}{g_0(\varepsilon|\omega)}\right] + 2\frac{\dot{h}(x|s,z)}{h_0(x|s,z)} + 2\frac{\dot{f}(s|z)}{f_0(s|z)} + 2\frac{\dot{m}(z)}{m_0(z)}
\end{aligned}$$

with $\dot{\mu}(z) = 2\int s\dot{f}(s|z)f_0(s|z)ds$. Since $\lim_{|\epsilon|\to\infty} g_0^2(\epsilon|\omega) = 0$ and $\int \dot{g}(\varepsilon|\omega)g_0(\varepsilon|\omega)d\varepsilon = 0$ w.p.1, using iterated expectations we have

$$
\begin{aligned}
E[S_0^2] = 4E&\left[\left(\frac{\dot{g}(\varepsilon|\omega) - g_0'(\varepsilon|\omega)\cdot(v'\dot{\theta} + \gamma_0'\dot{\mu}(z))}{g_0(\varepsilon|\omega)}\right)^2\right] + 4E\left[\left(\frac{\dot{h}(x|s,z)}{h_0(x|s,z)}\right)^2\right] + 4E\left[\left(\frac{\dot{f}(s|z)}{f_0(s|z)}\right)^2\right] \\
&+ 4\int \dot{m}(z)^2 dz, \quad \text{and therefore,} \\
\langle \dot{\tau}_1, \dot{\tau}_2 \rangle_F = 4E&\left[\left(\frac{\dot{g}_1(\varepsilon|\omega) - g_0'(\varepsilon|\omega)\cdot(v'\dot{\theta}_1 + \gamma_0'\dot{\mu}_1(z))}{g_0(\varepsilon|\omega)}\right)\left(\frac{\dot{g}_2(\varepsilon|\omega) - g_0'(\varepsilon|\omega)\cdot(v'\dot{\theta}_2 + \gamma_0'\dot{\mu}_2(z))}{g_0(\varepsilon|\omega)}\right)\right] \\
&+ 4E\left[\left(\frac{\dot{h}_1(x|s,z)}{h_0(x|s,z)}\right)\left(\frac{\dot{h}_2(x|s,z)}{h_0(x|s,z)}\right)\right] + 4E\left[\left(\frac{\dot{f}_1(s|z)}{f_0(s|z)}\right)\left(\frac{\dot{f}_2(s|z)}{f_0(s|z)}\right)\right] + 4\int \dot{m}_1(z)\dot{m}_2(z)dz
\end{aligned}
$$

We are interested in the efficiency bound for $\rho(\tau) = c'\theta_0$ for an arbitrary c. The R-F condition becomes

$$
\begin{aligned}
c'\dot{\theta} = 4E&\left[\left(\frac{g^*(\varepsilon|\omega) - g_0'(\varepsilon|\omega)\cdot(v'\theta^* + \gamma_0'\mu^*(z))}{g_0(\varepsilon|\omega)}\right)\left(\frac{\dot{g}(\varepsilon|\omega) - g_0'(\varepsilon|\omega)\cdot(v'\dot{\theta} + \gamma_0'\dot{\mu}(z))}{g_0(\varepsilon|\omega)}\right)\right] \\
&+ 4E\left[\left(\frac{h^*(x|s,z)}{h_0(x|s,z)}\right)\left(\frac{\dot{h}(x|s,z)}{h_0(x|s,z)}\right)\right] + 4E\left[\left(\frac{f^*(s|z)}{f_0(s|z)}\right)\left(\frac{\dot{f}(s|z)}{f_0(s|z)}\right)\right] + 4\int m^*(z)\dot{m}(z)dz \\
&\forall\ \dot{\tau} \in \dot{\mathscr{T}},
\end{aligned}
$$

where $\underbrace{\mu^*(z)}_{d_s\times 1} = 2\int sf^*(s|z)f_0(s|z)ds$. Firstly, we will set $m^* = 0$ and $h^* = 0$, as these two representers will prove to be ancillary to our problem. Secondly, it will be convenient to express $f^*(s|z) = \theta^{*\prime}\underbrace{t^*(s|z)}_{d\times 1}$ and $g^*(\varepsilon|\omega) = \theta^{*\prime}\underbrace{\lambda^*(\varepsilon|\omega)}_{d\times 1}$, where $t^* \in \overline{lin\ T(\mathcal{F}.f_0)}$ and $\lambda^* \in \overline{lin\ T(\mathcal{G}, g_0)}$ element-wise. From here, we have $\mu^*(z) = \delta^*(z)\theta^*$, where $\underbrace{\delta^*(z)}_{d_s\times d} \equiv 2\int st^*(s|z)'f_0(s|z)ds$. Next, let $\underbrace{\Phi^*(\omega,\varepsilon)}_{d\times 1} \equiv 2\left(\frac{\lambda^*(\varepsilon|\omega) - g_0'(\varepsilon|\omega)\cdot(v + \delta^*(z)'\gamma_0)}{g_0(\varepsilon|\omega)}\right)$ and $\Omega(z) = E\left[\Phi^*(\omega,\varepsilon)\cdot\eta(\omega,\varepsilon)|z\right]$. The R-F condition becomes

$$
\begin{aligned}
c'\dot{\theta} = &\underbrace{2\theta^{*\prime}E\left[\int\left\{2t^*(s|z) - \Omega(z)\gamma_0' s f_0(s|z)\right\}\dot{f}(s|z)ds\right]}_{(5A)} + \underbrace{2\theta^{*\prime}E\left[\frac{\Phi^*(\omega,\varepsilon)}{g_0(\varepsilon|\omega)}\cdot\dot{g}(\varepsilon|\omega)\right]}_{(5B)} \\
&\underbrace{-\theta^{*\prime}E\left[\Phi^*(\omega,\varepsilon)\eta(\omega,\varepsilon)v'\right]\dot{\theta}}_{(5C)} \quad \forall\ \dot{\tau} \in \dot{\mathscr{T}}
\end{aligned} \tag{5}
$$

To solve (5), we will first find the representers t^* and λ^* that make both (5A) and (5B) equal to zero. We will choose

$$t^*(s|z) = \frac{1}{2}\Omega(z)\gamma_0'(s - E[s|z])f_0(s|z). \tag{6}$$

Since $\int f_0(s|z)\dot{f}(s|z)ds = 0$ w.p.1, it is straightforward to verify that this choice for the representer t^* makes (5A) equal to zero. Furthermore, $\int t^*(s|z)f_0(s|z) = \frac{1}{2}\Omega(z)\gamma_0' \int (s - E[s|z])f_0^2(s|z)ds = 0$ w.p.1, and therefore $t^* \in \overline{lin\ T(\mathcal{F}, f_0)}$ element-wise (see (2)). From here,

$$\begin{aligned}
\delta^*(z) &= Var[s|z]\gamma_0\Omega(z)' \\
&= Var[s|z]\gamma_0 \cdot \left(2E\left[\left.\frac{\lambda^*(\varepsilon|\omega)' \cdot \eta(\omega,\varepsilon)}{g_0(\varepsilon|\omega)}\right| z\right] - E\left[v' \cdot \eta(\omega,\varepsilon)^2 |z\right]\right) - Var[s|z]\gamma_0\gamma_0' \cdot E\left[\eta(\omega,\varepsilon)^2|z\right]\delta^*(z).
\end{aligned}$$

Therefore,

$$\delta^*(z) = A(z) \cdot \left(2E\left[\left.\frac{\lambda^*(\varepsilon|\omega)' \cdot \eta(\omega,\varepsilon)}{g_0(\varepsilon|\omega)}\right| z\right] - E\left[v' \cdot \eta(\omega,\varepsilon)^2|z\right]\right), \tag{7}$$

where $\underbrace{A(z)}_{d_s\times 1} \equiv \left(I_{d_s} + Var[s|z]\gamma_0\gamma_0' \cdot E\left[\eta(\omega,\varepsilon)^2|z\right]\right)^{-1} \cdot Var[s|z]\gamma_0$. Next, we will characterize the representer λ^* that makes (5B) equal to zero. Consider

$$\lambda^*(\varepsilon|\omega) = \frac{1}{2}\big(v + \delta^*(z)'\gamma_0\big) \cdot r(\omega,\varepsilon) \cdot g_0(\varepsilon|\omega), \tag{8}$$

where $r(\omega,\varepsilon)$ is as described in the statement of Proposition 1. With this choice, we have

$$\begin{aligned}
E\left[\frac{\Phi^*(\omega,\varepsilon)}{g_0(\varepsilon|\omega)} \cdot \dot{g}(\varepsilon|\omega)\right] &= E\left[\left(\frac{\lambda^*(\varepsilon|\omega) - g_0'(\varepsilon|\omega)\left(v + \delta^*(z)'\gamma_0\right)}{g_0^2(\varepsilon|\omega)}\right) \cdot \dot{g}(\varepsilon|\omega)\right] \\
= -\frac{1}{2}E\left[\frac{\mathcal{C}(\omega)'\Upsilon(\varepsilon)}{g_0(\varepsilon|\omega)} \cdot \dot{g}(\varepsilon|\omega)\right] &= -\frac{1}{2}E\left[\mathcal{C}(\omega)' \int \Upsilon(\varepsilon)g_0(\varepsilon|\omega)\dot{g}(\varepsilon|\omega)d\varepsilon\right] = 0,
\end{aligned}$$

where the last equality follows because $\dot{g} \in \overline{lin\ T(\mathcal{G}, g_0)}$ (see (2)). Therefore, the choice for λ^* in (8) makes (5B) equal to zero. But we need to verify that $\lambda^* \in \overline{lin\ T(\mathcal{G}, g_0)}$ element-wise. $\lambda^*(\varepsilon|\omega)$ consists of d components: $\lambda_\ell^*(\varepsilon|\omega)$ for $\ell = 1,\ldots,d$. From (8), each can be written as $\lambda_\ell^*(\varepsilon|\omega) = \xi_\ell(\omega) \cdot r(\omega,\varepsilon) \cdot g_0(\varepsilon|\omega)$. We need to show that each $\lambda_\ell^* \in \overline{lin\ T(\mathcal{G}, g_0)}$. That is: (a) $\int \lambda_\ell^*(\varepsilon|\omega)g_0(\varepsilon|\omega)d\varepsilon = 0$ and (b) $\int \Upsilon(\varepsilon)\lambda_\ell^*(\varepsilon|\omega)g_0(\varepsilon|\omega)d\varepsilon = 0$ w.p.1. We have

$$\begin{aligned}
\int \lambda_\ell^*(\varepsilon|\omega)g_0(\varepsilon|\omega)d\varepsilon &= \xi_\ell(\omega)\int r(\omega,\varepsilon) \cdot g_0(\varepsilon|\omega) \cdot g_0(\varepsilon|\omega)d\varepsilon \\
&= \xi_\ell(\omega)\int \left(g_0'(\varepsilon|\omega) - \mathcal{C}(\omega)'\Upsilon(\varepsilon) \cdot g_0(\varepsilon|\omega)\right) \cdot g_0(\varepsilon|\omega)d\varepsilon \\
&= \xi_\ell(\omega)\int g_0'(\varepsilon|\omega)g_0(\varepsilon|\omega)d\varepsilon - \xi_\ell(\omega)\mathcal{C}(\omega)'E\left[\Upsilon(\varepsilon)|\omega\right] = 0 \ \text{ w.p.1},
\end{aligned}$$

since $\int g_0'(\varepsilon|\omega)g_0(\varepsilon|\omega)d\varepsilon = 0$ and $E\left[\Upsilon(\varepsilon)|\omega\right] = 0$ w.p.1. Next, note that $\int \Upsilon(\varepsilon)\lambda_\ell^*(\varepsilon|\omega)g_0(\varepsilon|\omega)d\varepsilon = \xi_\ell(\omega)\int \Upsilon(\varepsilon)\cdot r(\omega,\varepsilon)\cdot g_0^2(\varepsilon|\omega)d\varepsilon = \xi_\ell(\omega)\cdot E\left[\Upsilon(\varepsilon)r(\omega,\varepsilon)|\omega\right] = 0$ w.p.1 since, conditional on ω, $\Upsilon(\varepsilon)\perp r(\omega,\varepsilon)$ w.p.1. Therefore, $\lambda^* \in \overline{lin\ T(\mathcal{G},g_0)}$ element-wise. Thus, with the representers in (6) and (8), the R-F condition in (5) becomes

$$c'\dot{\theta} = \theta^{*\prime}\underbrace{\left\{E\left[\Phi^*(\omega,\varepsilon)\Phi^*(\omega,\varepsilon)'\right] + E\left[\Omega(z)\gamma_0' Var[s|z]\gamma_0\Omega(z)'\right]\right\}}_{\equiv\Sigma_\theta^*}\dot{\theta}$$

From here, the R-F condition is satisfied by choosing $\theta^* = \Sigma_\theta^{*-1}c$ and by the R-F Theorem, the efficiency bound is $l.b = \langle\tau^*,\tau^*\rangle_F = \theta^{*\prime}\Big\{E\left[\Phi^*(\omega,\varepsilon)\Phi^*(\omega,\varepsilon)'\right] + E\left[\Omega(z)\gamma_0' Var[s|z]\gamma_0\Omega(z)'\right]\Big\}\theta^* = c'\Sigma_\theta^{*-1}\Sigma_\theta^*\Sigma_\theta^{*-1}c = c'\Sigma_\theta^{*-1}c$. The final step is to simplify Φ^*. Combining (7) and (8) and simplifying, we obtain a closed-form expression: $\delta^*(z) = -\frac{A(z)\cdot E\left[v' P_\Upsilon(\omega,\varepsilon)^2|z\right]}{1-A(z)'\gamma_0 E[r(\omega,\varepsilon)^2|z]}$, and

$$\Phi^*(\omega,\varepsilon) = -\left(v - \frac{E\left[v\cdot P_\Upsilon(\omega,\varepsilon)^2|z\right]A(z)'\gamma_0}{1-E\left[r(\omega,\varepsilon)^2|z\right]A(z)'\gamma_0}\right)\cdot P_\Upsilon(\omega,\varepsilon).$$

Finally, since $A(z)'\gamma_0$ is a scalar, using the properties of traces, $A(z)'\gamma_0 = tr\left(A(z)'\gamma_0\right) = tr\left(A(z)\gamma_0'\right)$. Therefore,

$$A(z)'\gamma_0 = tr\Bigg(\underbrace{\left(I_{d_s} + Var[s|z]\gamma_0\gamma_0'\cdot E\left[\eta(\omega,\varepsilon)^2|z\right]\right)^{-1}Var[s|z]\gamma_0}_{=A(z)}\gamma_0'\Bigg) \equiv tr(M(z)),$$

where $M(z) \equiv \left(I_{d_s} + E[\eta(\omega,\varepsilon)^2|z]\cdot Var[s|z]\gamma_0\gamma_0'\right)^{-1}Var[s|z]\gamma_0\gamma_0'$. This concludes the proof of the proposition. ■

Extensions and directions for future work

There are multiple avenues to extend model (1). A particularly interesting one would be to allow the conditioning variable z in $\mu(z) \equiv E[s|z]$ to be a nonparametric regressor itself. The approach of representers in the tangent space employed here also has the potential to be used to compute efficiency bounds in this case, and the steps of our proof can provide a guidance. While in the case examined here we have $\dot{\mu}(z) = 2\int s\dot{f}(s|z)f_0(s|z)ds$, if z itself is a nonparametric regressor we would have $\dot{\mu}(z) = 2\int s\left[\dot{f}(s|z) + \nabla_z f_0(s|z)'\dot{z}\right]f_0(s|z)ds$. The tangent space for $\dot{z}$ would depend on the specific structure of z. For example, if $z = E[u_1|u_2] = \int u_1\nu_0^2(u_1|u_2)du_1$, we would have $\dot{z} = 2\int u_1\dot{\nu}(u_1|u_2)\nu_0(u_1|u_2)du_1$, and $\dot{\nu}$ would belong to a tangent space with the same type of properties described in our analysis (see equation (2)).

The model studied in Li and Wooldridge (2002) fits the general description of the extension outlined above. They focus on partially linear models of the type studied in Robinson (1988), described as $y = x'\beta_0 + m(\eta) + \varepsilon$ (with $m(\cdot)$ an unknown function) in cases where η is of the form $\eta = s - E[s|z]$. Assuming that $E[\varepsilon|x, z] = 0$, identification comes from the transformation $y - E[y|\eta] = (x - E[x|\eta])'\beta_0 + \varepsilon$. Assuming that (y, x, s, z) are observable, the tangent-space representer approach used here has the potential to derive the efficiency bound for $\sqrt{n}-$consistent, regular estimators of β_0 in this type of model. Even though it is not a special case of our analysis, the steps of our proof can provide a roadmap to derive the bounds once the tangent spaces are modified appropriately as we outlined above. We leave the details for future research.

References

Brock, W. and S. Durlauf (2001). Interactions-based models. In J. Heckman and E. Leamer (Eds.), *The Handbook of Econometrics*, Volume 5, pp. 3297–3380. North-Holland.

Li, Q. and J. Wooldridge (2002). Semiparametric estimation of partially linear models for dependent data with generated regressors. *Econometric Theory 18*, 625–645.

Luenberger, D. (1969). *Optimization by vector space methods.* Wiley Inter-Science.

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem. *Review of Economic Studies 60*, 531–542.

Manski, C. F. (1995). *Identification Problems in the Social Sciences.* Cambridge, MA: Harvard University Press.

Newey, W. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics 5*, 99–135.

Rilstone, P. (1993). Calculating the (local) semiparametric efficiency bounds for the generated regressors problem. *Journal of Econometrics 56*, 357–370.

Robinson, P. (1988). Root-n-consistent semiparametric regression. *Econometrica 56*, 931–954.

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

Stein, C. (1956). Efficient nonparametric testing and estimation. In M. Armstrong and P. Rob (Eds.), *Berkeley symposium on mathematical statistics and probability*, Volume 1, pp. 187–195. University of California Press.

Young, N. (1988). *An introduction to Hilbert space.* Cambridge.