Econometric Theory, **25**, 2009, 847–855. Printed in the United States of America.
doi:10.1017/S0266466609090732

EFFICIENCY BOUNDS FOR SEMIPARAMETRIC ESTIMATION OF INVERSE CONDITIONAL-DENSITY-WEIGHTED FUNCTIONS

DAVID T. JACHO-CHÁVEZ
Indiana University

Consider the unconditional moment restriction $\mathrm{E}[\mathbf{m}(\mathbf{y}, v, \mathbf{w}; \pi_0)/f_{V|\mathbf{W}}(v|\mathbf{w}) - \mathbf{s}(\mathbf{w}; \pi_0)] = 0$, where $\mathbf{m}(\cdot)$ and $\mathbf{s}(\cdot)$ are known vector-valued functions of data $(\mathbf{y}^\top, v, \mathbf{w}^\top)^\top$. The smallest asymptotic variance that $\sqrt{n}$-consistent regular estimators of π_0 can have is calculated when $f_{V|\mathbf{W}}(\cdot)$ is only known to be a bounded, continuous, nonzero conditional density function. Our results show that "plug-in" kernel-based estimators of π_0 constructed from this type of moment restriction, such as Lewbel (1998, *Econometrica* 66, 105–121) and Lewbel (2007, *Journal of Econometrics* 141, 777–806), are semiparametric efficient.

1. MOTIVATION AND INTRODUCTION

This paper determines the semiparametric efficiency bound for a class of $\sqrt{n}$-semiparametric regular estimators for latent variable models introduced by Lewbel (1998). The calculation of their smallest asymptotic variance is important, because they quantify the efficiency loss that can result from a semiparametric, rather than a parametric, approach.

These estimators are defined by an unconditional moment restriction sharing the following common structure:

$$\mathrm{E}\left[\mathbf{g}\left(\mathbf{y}_i, v_i, \mathbf{w}_i; \pi_0, f_{V|\mathbf{W}}\right)\right] = \mathbf{0}, \quad \text{with} \tag{1.1}$$

$$\mathbf{g}\left(\mathbf{y}_i, v_i, \mathbf{w}_i; \pi_0, f_{V|\mathbf{W}}\right) = \frac{\mathbf{m}(\mathbf{y}_i, v_i, \mathbf{w}_i; \pi_0)}{f_{V|\mathbf{W}}(v_i|\mathbf{w}_i)} - \mathbf{s}(\mathbf{w}_i; \pi_0), \tag{1.2}$$

where $\mathbf{y}_i$ is a dependent variable, $\left(v_i, \mathbf{w}_i^\top\right)^\top$ is a vector of regressors, $\top$ represents transpose, i indexes observations, $\mathbf{m}(\cdot)$ and $\mathbf{s}(\cdot)$ are known vector-valued functions, π_0 is a vector of unknown constant coefficients to be estimated, and $f_{V|\mathbf{W}}(\cdot)$ is the unknown conditional probability density of v_i given $\mathbf{w}_i$

We thank the co-editor Richard Smith, two anonymous referees, Francesco Bravo, Juan Carlos Escanciano, Javier Hidalgo, Kim Huynh, Oliver Linton, and Pravin Trivedi for many helpful comments, corrections, and suggestions. The usual disclaimers apply. Address correspondence to David T. Jacho-Chávez, Department of Economics, Indiana University, Wylie Hall 251, 100 South Woodlawn Avenue, Bloomington, IN 47405, USA; e-mail: djachoch@indiana.edu.

© 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S0266466609090732 Published online by Cambridge University Press

(which is assumed throughout the paper to be continuous and nonzero on its compact support). The parameter space in (1.1)–(1.2) contains two parts: the finite-dimensional parameter π_0 and the unknown function $f_{V|\mathbf{W}}(v|\mathbf{w})$. General theory for calculating the minimum asymptotic variance for $\sqrt{n}$-consistent regular estimators of π_0 in (1.1) is readily available in Chamberlain (1992) and Ai and Chen (2003). Whereas their results did not impose any particular structure on the unknown infinite-dimensional parameter $f_{V|\mathbf{W}}(\cdot)$, here it is a conditional density function.[1] This extra piece of information regarding the underlying distribution will generally alter the (semiparametric) efficiency bound in (1.1)–(1.2) (see, e.g., Chamberlain, 1987).

The underlying framework is described in the next section, which also states the main theorem and discusses its implications. The resulting bounds are then applied to "plug-in" kernel-based estimators in Lewbel (1998) and Lewbel (2007), demonstrating their asymptotic efficiency. These results seem to be new in the literature. The proof and technical details are confined to the Appendix.

2. EFFICIENCY RESULT AND DISCUSSION

The framework is as follows: For an independent and identically distributed sample $i = 1, \ldots, n$, the information on one unit i is contained in a $d \times 1$ vector $\mathbf{u} = \left(\mathbf{y}^\top, v, \mathbf{w}^\top\right)^\top \in (S_{\mathbf{Y}} \times S_V \times S_{\mathbf{W}}) \equiv S_{\mathbf{U}} \subset \Re^d$ with unknown joint density $f(\mathbf{y}, v, \mathbf{w})$ with respect to a dominating measure $\mu = \mu_{\mathbf{Y}} \times \lambda \times \mu_{\mathbf{W}} = \left(\otimes_{l=1}^{d_{\mathbf{Y}}} \mu_{Y_l}\right) \times \lambda \times \left(\otimes_{l=1}^{d_{\mathbf{W}}} \mu_{W_l}\right)$, where $d = d_{\mathbf{Y}} + d_V + d_{\mathbf{W}}$ and λ represents Lebesgue measure. Because $\mathbf{y}$ and $\mathbf{w}$ can have discrete components, μ_{Y_i}'s and μ_{W_i}'s need not be Lebesgue measures. Furthermore, $f(\cdot)$ can be decomposed into conditional probability densities with respect to measures $\mu_{\mathbf{Y}}$, λ, and $\mu_{\mathbf{W}}$; i.e., if $\mathbf{g}$: $S_{\mathbf{U}} \to \Re^q$ is an integrable function, then

$$\int \mathbf{g}(\mathbf{u}) f(\mathbf{u}) \mu(d\mathbf{u}) = \int\int\int \mathbf{g}(\mathbf{u}) f_{\mathbf{Y}|V,\mathbf{W}}(\mathbf{y}|v,\mathbf{w}) \mu_{\mathbf{Y}}(d\mathbf{y}) \times f_{V|\mathbf{W}}(v|\mathbf{w}) dv f_{\mathbf{W}}(\mathbf{w}) \mu_{\mathbf{W}}(d\mathbf{w}),$$

where $f_{\mathbf{Y}|V,\mathbf{W}}$, $f_{V|\mathbf{W}}$, and $f_{\mathbf{W}}$ are the implied densities with respect to the conditional measures $\mu_{\mathbf{Y}}$, λ, and $\mu_{\mathbf{W}}$, respectively. For a given open set $\Pi \subset \Re^p$, define functions $\mathbf{m}$: $S_{\mathbf{U}} \times \Pi \to \Re^q$ and $\mathbf{s}$: $S_{\mathbf{W}} \times \Pi \to \Re^q$ such that for each $\pi \in \Pi$, $\mathbf{m}(\cdot;\pi)$: $S_{\mathbf{U}} \to \Re^q$ and $\mathbf{s}(\cdot;\pi)$: $S_{\mathbf{w}} \to \Re^q$ are measurable and for each $\mathbf{u} \in S_{\mathbf{U}}$, $\nabla_\pi \mathbf{m}(\mathbf{u};\pi)$ and $\nabla_\pi \mathbf{s}(\mathbf{w};\pi)$ are continuous mappings on Π. The restrictions on $f_{\mathbf{Y}|V,\mathbf{W}}(\cdot)$, $f_{V|\mathbf{W}}(\cdot)$, and $f_{\mathbf{W}}(\cdot)$ also involve the unconditional moment (1.1), where π_0 is some point in Π and $f_{V|\mathbf{W}}(\cdot)$ is the true conditional density function of V given $\mathbf{W}$. These assumptions and notation will be used in the proof of the following theorem.

THEOREM 2.1. *Assume that the true distribution of* $\mathbf{u}$ *satisfies (1.1)–(1.2) for a unique value* $\pi_0 \in \Pi$ *and that the matrix* $\mathrm{E}[\mathbf{g}\left(\mathbf{u};\pi_0, f_{V|\mathbf{W}}\right)\mathbf{g}\left(\mathbf{u};\pi_0, f_{V|\mathbf{W}}\right)^\top$

https://doi.org/10.1017/S0266466609090732 Published online by Cambridge University Press

$|v, \mathbf{w}]$ *is finite and nonsingular with probability one; then the semiparametric efficiency bound for regular estimators of* π_0 *is given by*

$$\left(\mathbf{M}_0^\top \widetilde{\Omega}_0^{-1} \mathbf{M}_0\right)^{-1}, \tag{2.1}$$

where

$$\mathbf{M}_0 = \mathrm{E}\left[\nabla_\pi \mathbf{g}\left(\mathbf{u}; \pi_0, f_{V|\mathbf{W}}\right)\right],$$

$$\widetilde{\Omega}_0 = \mathrm{E}\left[\widetilde{\mathbf{g}}\left(\mathbf{u}; \pi_0, f_{V|\mathbf{W}}\right) \widetilde{\mathbf{g}}\left(\mathbf{u}; \pi_0, f_{V|\mathbf{W}}\right)^\top\right],$$

$$\mathbf{g}\left(\mathbf{u}; \pi_0, f_{V|\mathbf{W}}\right) = \frac{\mathbf{m}(\mathbf{y}, v, \mathbf{w}; \pi_0)}{f_{V|\mathbf{W}}(v|\mathbf{w})} - \mathbf{s}(\mathbf{w}; \pi_0),$$

$$\widetilde{\mathbf{g}}\left(\mathbf{u}; \pi_0, f_{V|\mathbf{W}}\right) = \mathbf{g} - \mathrm{E}(\mathbf{g}|v, \mathbf{w}) + \mathrm{E}(\mathbf{g}|\mathbf{w}).$$

Proof. See the Appendix.

This bound corresponds to the covariance matrix of the influence function $-\left(\mathbf{M}_0^\top \widetilde{\Omega}_0^{-1} \mathbf{M}_0\right)^{-1} \mathbf{M}_0 \times \widetilde{\Omega}_0^{-1} \widetilde{\mathbf{g}}\left(\mathbf{u}; \pi_0, f_{V|\mathbf{W}}\right)$, where $\widetilde{\mathbf{g}}(\cdot)$ can be thought as $\mathbf{g}(\cdot)$ "net" of effects from the first-stage estimation of $f_{V|\mathbf{W}}(\cdot)$. Its form resembles that of standard generalized method of moments (GMM) estimators. Discussion of some relevant implications of this theorem concludes this paper.

Remark 1. Consider the semiparametric latent variable specification, $y = L\left(\alpha_0 v + \mathbf{x}^\top \beta_0 + \varepsilon\right)$, where $L(\cdot)$ represents a known or unknown function and ε is an unobserved error such that conditional on $\mathbf{w} = (\mathbf{x}^\top, \mathbf{z}^\top)^\top$, its distribution is not affected by v, and $\mathrm{E}[\mathbf{z}\varepsilon] = 0$. In the case when the support of $\mathbf{x}^\top \beta_0 + \varepsilon$ is contained in the support of v, Lewbel (1998) showed that (1.1)–(1.2) holds with

$$\mathbf{m}(y, v, \mathbf{x}, \mathbf{z}; \alpha_0, \beta_0) = \left[\alpha_0 v \theta(y), \mathbf{z}\alpha_0^2 v \theta(y)\right]^\top,$$

$$\mathbf{s}(\mathbf{x}, \mathbf{z}; \alpha_0, \beta_0) = \left[\kappa_1, \kappa_2 \mathbf{z} - \kappa_1 \mathbf{z}\mathbf{x}^\top \beta_0\right]^\top,$$

where $\pi_0 = (\alpha_0, \beta_0^\top)^\top$ and θ is a known real-valued function, with $\kappa_1 = \int_{-\infty}^{\infty} \theta[L(a)]da$ and $\kappa_2 = \int_{-\infty}^{\infty} a\theta[L(a)]da$. A direct application of Theorem 2.1 shows that the Lewbel (1998) estimator is semiparametric efficient. Detailed calculations can be found in Jacho-Chávez (2006).

Under similar conditions, with $\pi_0 = \left(1, \beta_0^\top\right)^\top$ and $L(a) = \mathbb{I}(a > 0)$, ($\mathbb{I}(\cdot)$ is the indicator function), Lewbel (2000) shows that moment restriction (1.1)–(1.2) holds with $\mathbf{m}(y, v, \mathbf{x}, \mathbf{z}; \beta_0) = \mathbf{z}(y - \mathbb{I}(v > 0))$ and $\mathbf{s}(\mathbf{x}, \mathbf{z}; \beta_0) = \mathbf{z}\mathbf{x}^\top \beta_0$. Magnac and Maurin (2007) establish that the Lewbel (2000) estimator is semiparametric efficient, using techniques similar to ours.

Remark 2. Let $\left(y, d, v, \mathbf{x}^\top, \mathbf{z}^\top\right)$ be a random vector from a data generating process where $y = y^* \times d$, with $d = \mathbb{I}(0 \leq v + M^* \leq c)$, for some known $c < \infty$,

https://doi.org/10.1017/S0266466609090732 Published online by Cambridge University Press

and unobserved random scalars y^* and M^*. Furthermore, for some finitely parameterized (by $\theta \in \Theta$) known vector-value function $\Psi(\cdot)$, the unconditional moment $\mathrm{E}\left[\Psi\left(y^*, v, \mathbf{x}^\top, \mathbf{z}^\top; \theta_0\right)\right] = 0$ holds. Let $\mathbf{y} = (y, d)^\top$ and $\mathbf{w} = \left(\mathbf{x}^\top, \mathbf{z}^\top\right)^\top$; then Lewbel (2007) shows that (1.1)–(1.2) holds with

$$\mathbf{m}(y, d, v, \mathbf{x}, \mathbf{z}; \theta_0) = \Psi(y, v, \mathbf{x}, \mathbf{z}; \theta_0)\, d,$$

$$\mathbf{s}(\mathbf{x}, \mathbf{z}; \theta_0) = \mathbf{0},$$

for the same unknown $\theta_0 \in \Theta$, which is assumed to be unique. The support of v, conditional on $(\mathbf{x}^\top, \mathbf{z}^\top)^\top$, is again assumed to contain the support of $-M^*$ and $c - M^*$. Furthermore, v does not affect the distribution of M^*. Another direct application of Theorem 2.1 shows that "plug-in" kernel-based estimators proposed in Lewbel (2007) are semiparametric efficient (for calculations and examples, see Jacho-Chávez, 2006).

Remark 3. Support conditions on v in Remarks 1 and 2 not only ensure the identifiability of parameters of interest, but that they are not needed in the derivation of the semiparametric efficiency bounds.

NOTE

1. We are grateful to an anonymous referee for raising this point.

REFERENCES

Ai, Chunrong, & Xiaohong Chen (2003) Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71, 1795–1843.

Chamberlain, G. (1987) Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34, 305–334.

Chamberlain, G. (1992) Efficiency bounds for semiparametric regression. *Econometrica* 60, 567–596.

Hansen, L.P. (1982) Large sample properties of generalized methods of moments estimators. *Econometrica* 50, 1029–1054.

Jacho-Chávez, D.T. (2006) Identification, estimation and efficiency of nonparametric and semiparametric models in microeconometrics. Ph.D. Dissertation, London School of Economics and Political Science.

Lewbel, A. (1998) Semiparametric latent variable model estimation with endogenous or mismeasured regressors. *Econometrica* 66, 105–121.

Lewbel, A. (2000) Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics* 97, 145–177.

Lewbel, A. (2007) Endogenous selection or treatment model estimation. *Journal of Econometrics* 141, 777–806.

Magnac, T. & E. Maurin (2007) Identification and information in monotone binary models. *Journal of Econometrics* 127, 76–104.

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

Rothenberg, T.J. (1971) Identification in parametric models. *Econometrica* 39, 577–591.

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

https://doi.org/10.1017/S0266466609090732 Published online by Cambridge University Press

APPENDIX A: Main Proofs

We use results of Severini and Tripathi (2001), who, using some standard Hilbert-space theory, present an alternative approach to computing efficiency bounds in semiparametric models. This approach simplifies the bounds' derivation in comparison to the popular procedure described in Newey (1990). Severini and Tripathi (2001) demonstrate that both methodologies are equivalent. We begin by introducing some notation and present the main results of Severini and Tripathi (2001) in the context of deriving the efficiency bound for estimating π_0.

Notation. In what follows, $L^2(S_B;\mu_B)$ represents the set of all real-valued functions on S_B that are square integrable with respect to the μ_B measure. The term λ represents the Lebesgue measure; $L^2(S_B;B)$ also denotes the set of all real-valued functions on S_B that are square integrable with respect to the probability distribution of B. We are also going to use the symbol E_B to denote integrals with respect to the distribution of B, and we will use $\mathrm{E}[\cdot|b]$ as conditional expectation given $B=b$.

Proof of Theorem 2.1. The random vector $\left(\mathbf{Y}^\top, V, \mathbf{W}^\top\right)^\top$ has unknown density function $f(\mathbf{y},\upsilon,\mathbf{w})$ (with respect to a dominating measure $\mu=\mu_{\mathbf{Y}}\times\lambda\times\mu_{\mathbf{W}}$) that is rewritten as

$$f(\mathbf{y},\upsilon,\mathbf{w}) = f_{\mathbf{Y}|V,\mathbf{W}}(\mathbf{y}|\upsilon,\mathbf{w})\, f_{V|\mathbf{W}}(\upsilon|\mathbf{w})\, f_{\mathbf{W}}(\mathbf{w})$$

$$= \phi^2_{0;\mathbf{Y}}(\mathbf{y}|\upsilon,\mathbf{w})\,\psi^2_0(\upsilon|\mathbf{w})\,\phi^2_{0;\mathbf{W}}(\mathbf{w}),$$

where $\phi_{0;\mathbf{Y}}\in\Phi_Y$, $\psi_0\in\Psi$, $\phi_{0;\mathbf{W}}\in\Phi_{\mathbf{W}}$, and

$$\Phi_{\mathbf{Y}} = \left\{ \phi_{\mathbf{Y}} \in S_{\mathbf{Y}}\times S_V\times S_{\mathbf{W}}\to\Re, \qquad \phi^2_{\mathbf{Y}}(\mathbf{y}|\upsilon,\mathbf{w})\geq 0, \quad \text{bounded}, \right.$$

$$\left. \int_{S_{\mathbf{Y}}} \phi^2_{\mathbf{Y}}(\mathbf{y}|\upsilon,\mathbf{w})\,\mu_{\mathbf{Y}}(d\mathbf{y}) = 1 \right\},$$

$$\Psi = \left\{ \psi\in L^2(S_V;\lambda), \qquad \psi^2(\upsilon|\mathbf{w})>0, \right.$$

$$\psi^2(\upsilon|\mathbf{w}) \quad \text{is bounded, and absolutely continuous}$$

$$\left. \text{Radon–Nikodym conditional density such that } \int_{S_V}\psi^2(\upsilon|\mathbf{w})\,d\upsilon = 1 \right\},$$

$$\Phi_{\mathbf{W}} = \left\{ \phi_{\mathbf{W}}\in L^2\left(S_{\mathbf{W}};\mu_{\mathbf{W}}\right), \phi^2_{\mathbf{W}}(\mathbf{w})>0, \quad \text{bounded}, \right.$$

$$\left. \int_{S_{\mathbf{W}}}\phi^2_{\mathbf{W}}(\mathbf{w})\,\mu_{\mathbf{W}}(d\mathbf{w}) = 1 \right\}.$$

https://doi.org/10.1017/S0266466609090732 Published online by Cambridge University Press

The assumption that matrix $\mathrm{E}[\mathbf{g}(\mathbf{u};\pi_0,\psi_0^2)\mathbf{g}(\mathbf{u};\pi_0,\psi_0^2)^\top|\upsilon,\mathbf{w}]$ exists and is not singular implies that none of the components of $\widetilde{\mathbf{m}}(\mathbf{u};\pi_0,\psi_0^2)-\mathbf{s}(\mathbf{w};\pi_0)$ are redundant, i.e., linearly dependent, where $\widetilde{\mathbf{m}}(\mathbf{u};\pi_0,\psi_0^2)=\mathbf{m}(\mathbf{u};\pi_0)/\psi_0^2$.

The objective is to calculate the efficiency bound for estimating $\pi_0\in\Pi\subset\Re^p$, our parameter of interest. To simplify this problem, for any arbitrary $\mathbf{c}\in\Re^p$ we look at estimating the real-valued function $\rho\left(\phi_{0;\mathbf{Y}},\psi_0,\phi_{0;\mathbf{W}}\right)=\mathbf{c}^\top\pi_0$ as our "structural" parameter of interest. As $\mathbf{c}$ is arbitrary, it is then straightforward to deduce the efficiency bound for estimators of π_0.

As described in Section 2 in Severini and Tripathi (2001) we begin by parameterizing $\xi_0\equiv(\phi_{0;\mathbf{Y}},\ \psi_0,\ \phi_{0;\mathbf{W}})$ and π_0 as a one-dimensional subproblem. For some $t_0>0$ let $t\longmapsto\left(\phi_{t;\mathbf{Y}},\psi_t,\phi_{t;\mathbf{W}},\pi_t\right)$ be a curve from $[0,t_0]$ into $\mathcal{A}\times\Pi$ that passes through (ξ_0,π_0) at $t=0$ and let $\mathcal{A}\equiv\Phi_{\mathbf{Y}}\times\Psi\times\Phi_{\mathbf{W}}$. In this context, a vector $\dot{\xi}\equiv(\dot{\phi}_{\mathbf{Y}},\dot{\psi},\dot{\phi}_{\mathbf{W}})$ is said to be tangent to $\mathcal{A}$ at ξ_0 if it is the slope of $\xi_t\equiv\left(\phi_{t;\mathbf{Y}},\psi_t,\phi_{t;\mathbf{W}}\right)$ at $t=0$, i.e., $\lim_{t\to0}||t^{-1}(\xi_t-\xi_0)-\dot{\xi}||=0$. Let the tangent space to $\mathcal{A}$ at the true value $(\phi_{0;\mathbf{Y}},\ \psi_0,\ \phi_{0;\mathbf{W}})$ be denoted by $\overline{\text{lin}\,T\left(\mathcal{A},\left(\phi_{0;Y},\psi_0,\phi_{0;\mathbf{W}}\right)\right)}$. This tangent space is the smallest linear space that is closed in the L^2-norm and that contains all $\dot{\xi}\in L^2(S_{\mathbf{U}};\mu_{\mathbf{Y}}\times\lambda\times\mu_{\mathbf{W}})$ that are tangent to $\mathcal{A}$ at $(\phi_{0;\mathbf{Y}},\ \psi_0,\ \phi_{0;\mathbf{W}})$.

As shown in Severini and Tripathi (2001), the tangent space is the product of $\overline{\text{lin}\,T\left(\Phi_{\mathbf{Y}},\phi_{0;\mathbf{Y}}\right)}$, $\overline{\text{lin}\,T\left(\Psi,\psi_0\right)}$, and $\overline{\text{lin}\,T\left(\Phi_{\mathbf{W}},\phi_{0;\mathbf{W}}\right)}$, where

$$\overline{\text{lin}\,T\left(\Phi_{\mathbf{Y}},\phi_{0;\mathbf{Y}}\right)}=\left\{\dot{\phi}_{\mathbf{Y}}\in L^2\left(S_{\mathbf{U}};\mu_Y\times\upsilon\times\mathbf{w}\right),\right.$$
$$\left.\int_{S_{\mathbf{Y}}}\dot{\phi}_{\mathbf{Y}}\phi_{0;\mathbf{Y}}\left(\mathbf{y}|\upsilon,\mathbf{w}\right)\mu_Y\left(dy\right)=0\quad\text{for almost all }\left(\upsilon,\mathbf{w}^\top\right)\right\}$$

$$\overline{\text{lin}\,T\left(\Phi_{\mathbf{W}},\phi_{0;\mathbf{W}}\right)}=\left\{\dot{\phi}_{\mathbf{W}}\in L^2\left(S_{\mathbf{W}};\mu_{\mathbf{W}}\right),\int_{S_{\mathbf{W}}}\dot{\phi}_{\mathbf{W}}\phi_{0;\mathbf{W}}\left(\mathbf{w}\right)\mu_{\mathbf{W}}\left(d\mathbf{w}\right)=0\right\}.$$

(In fact, they showed this for μ_Y and $\mu_{\mathbf{W}}$ Lebesgue measures. These sets can be made more specific in a given latent variable model at the expense of additional notation and complexity without altering the final result.) We should notice that we do not need to define explicitly the set $\overline{\text{lin}\,T\left(\Psi,\psi_0\right)}$ because ψ_0^* will soon be shown to be zero, which is always an element of the tangent space, whatever it may be. Severini and Tripathi (2001) also provide the framework in which we define, for any $(\dot{\phi}_{\mathbf{Y}},\dot{\psi},\dot{\phi}_{\mathbf{W}})$ and $(\phi'_{\mathbf{Y}},\psi',\phi'_{\mathbf{W}})$ elements of the tangent space, the Fisher information inner product $\langle\cdot,\cdot\rangle_F$ and the corresponding norm $\|\cdot\|_F$ as

$$\langle(\dot{\phi}_{\mathbf{Y}},\dot{\psi},\dot{\phi}_{\mathbf{W}}),(\phi'_{\mathbf{Y}},\psi',\phi'_{\mathbf{W}})\rangle_F=4\int_{S_{\mathbf{Y}}}\mathrm{E}_{V,\mathbf{W}}(\dot{\phi}_{\mathbf{Y}}\phi'_{\mathbf{Y}})\mu_{\mathbf{Y}}\left(dy\right)+4\mathrm{E}_{\mathbf{W}}\left(\int_{S_V}\dot{\psi}\,\psi'dv\right)$$
$$+4\int_{S_{\mathbf{W}}}\dot{\phi}_{\mathbf{W}}\phi'_{\mathbf{W}}\mu_{\mathbf{W}}\left(d\mathbf{w}\right)\qquad\textbf{(A.1)}$$

$$\|\left(\dot{\phi}_{\mathbf{Y}},\dot{\psi},\dot{\phi}_{\mathbf{W}}\right)\|_F^2=\langle(\dot{\phi}_{\mathbf{Y}},\dot{\psi},\dot{\phi}_{\mathbf{W}}),(\dot{\phi}_{\mathbf{Y}},\dot{\psi},\dot{\phi}_{\mathbf{W}})\rangle_F.$$

https://doi.org/10.1017/S0266466609090732 Published online by Cambridge University Press

Because the tangent space is a closed subspace of $L^2\left(S_{\mathbf{U}};\mu_Y\times\upsilon\times\mathbf{w}\right)$, the tangent space with this inner product, $\overline{(\text{lin}\,T(\mathcal{A},(\phi_{0;Y},\psi_0,\phi_{0;\mathbf{W}})),\langle\cdot,\cdot\rangle)}$, is a Hilbert space.

In the semiparametric model at hand, not every $(\dot{\phi}_{\mathbf{Y}},\dot{\psi},\dot{\phi}_{\mathbf{W}})$ in the tangent space may be used to calculate $\langle\cdot,\cdot\rangle_F$. In particular, after differentiating the unconditional moment conditions,

$$\mathrm{E}\left[\widetilde{\mathbf{m}}\left(\mathbf{u};\pi_t,\psi_t^2\right)-\mathbf{s}\left(\mathbf{w};\pi_t\right)\right]=\mathrm{E}\left[\mathbf{g}\left(\mathbf{u};\pi_0,\psi_0^2\right)\right],$$

with respect to t, we can notice that only those $(\dot{\phi}_{\mathbf{Y}},\dot{\psi},\dot{\phi}_{\mathbf{W}})$ may be used that also satisfy

$$\begin{aligned}
\mathbf{0}=&\int_{S_{\mathbf{Y}}\times S_V\times S_{\mathbf{W}}}\nabla_\pi\left[\mathbf{m}\left(\mathbf{y},\upsilon,\mathbf{w};\pi_0\right)-\mathbf{s}\left(\mathbf{w};\pi_0\right)\psi_0^2\right]\phi_{0;\mathbf{Y}}^2\phi_{0;\mathbf{W}}^2\mu_{\mathbf{Y}}\left(d\mathbf{y}\right)d\upsilon\mu_{\mathbf{W}}\left(d\mathbf{w}\right)\dot{\pi}\\
&+2\int_{S_{\mathbf{Y}}\times S_V\times S_{\mathbf{W}}}\Big\{\mathbf{m}\left(\mathbf{y},\upsilon,\mathbf{w};\pi_0\right)\left[\phi_{0;\mathbf{Y}}\dot{\phi}_{\mathbf{Y}}\phi_{0;\mathbf{W}}^2+\phi_{0;\mathbf{Y}}^2\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}\right]\\
&\qquad-\mathbf{s}\left(\mathbf{w};\pi_0\right)\left[\phi_{0;\mathbf{Y}}\dot{\phi}_{\mathbf{Y}}\psi_0^2\phi_{0;\mathbf{W}}^2+\phi_{0;\mathbf{Y}}^2\psi_0\dot{\psi}_0\phi_{0;\mathbf{W}}^2\right.\\
&\qquad\left.+\phi_{0;\mathbf{Y}}^2\psi_0^2\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}\right]\Big\}\mu_{\mathbf{Y}}\left(d\mathbf{y}\right)d\upsilon\mu_{\mathbf{W}}\left(d\mathbf{w}\right),\\
\mathbf{0}=&\int_{S_{\mathbf{Y}}\times S_V\times S_{\mathbf{W}}}\Big\{\nabla_\pi\left[\mathbf{m}\left(\mathbf{y},\upsilon,\mathbf{w};\pi_0\right)/\psi_0^2-\mathbf{s}\left(\mathbf{w};\pi_0\right)\right]\phi_{0;\mathbf{Y}}^2\psi_0^2\phi_{0;\mathbf{W}}^2\dot{\pi}\\
&\qquad+2\mathbf{m}\left(\mathbf{y},\upsilon,\mathbf{w};\pi_0\right)\left[\phi_{0;\mathbf{Y}}\dot{\phi}_{\mathbf{Y}}\phi_{0;\mathbf{W}}^2+\phi_{0;\mathbf{Y}}^2\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}\right]\Big\}\\
&\qquad\times\mu_{\mathbf{Y}}\left(d\mathbf{y}\right)d\upsilon\mu_{\mathbf{W}}\left(d\mathbf{w}\right)\\
&-2\int_{S_{\mathbf{W}}}\mathbf{s}\left(\mathbf{w};\pi_0\right)\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}\mu_{\mathbf{W}}\left(d\mathbf{w}\right),\\
\mathbf{0}=&\left[\mathrm{E}\nabla_\pi\mathbf{g}\left(\mathbf{y},\upsilon,\mathbf{w};\pi_0,\psi_0^2\right)\right]\dot{\pi}-2\int_{S_{\mathbf{W}}}\mathbf{s}\left(\mathbf{w};\pi_0\right)\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}\mu_{\mathbf{W}}\left(d\mathbf{w}\right)\\
&+2\int_{S_{\mathbf{Y}}\times S_V\times S_{\mathbf{W}}}\widetilde{\mathbf{m}}\left(\mathbf{y},\upsilon,\mathbf{w};\pi_0,\psi_0^2\right)\psi_0^2\left[\phi_{0;\mathbf{Y}}\dot{\phi}_{\mathbf{Y}}\phi_{0;\mathbf{W}}^2\right.\\
&\left.+\phi_{0;\mathbf{Y}}^2\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}\right]\mu_{\mathbf{Y}}\left(d\mathbf{y}\right)d\upsilon\mu_{\mathbf{W}}\left(d\mathbf{w}\right).
\end{aligned}$$

The second equality follows from the fact that for any $(\dot{\phi}_Y,\dot{\psi},\dot{\phi}_{\mathbf{W}})$ in the tangent space, $\int_{S_Y}\phi_{0;Y}\dot{\phi}_Y\mu_Y\left(dy\right)=0$ for almost all $\mathbf{w}$ and $\int_{S_V}\psi_0\dot{\psi}d\upsilon=0$. Because $q>p$, i.e., π_0 may be overidentified, the preceding equation will generally have a nonunique solution in $\dot{\pi}$. A sufficient condition for π_0 to be locally identified (see Rothenberg, 1971) is that we find a nonstochastic full rank $q\times q$ matrix $\mathbf{G}$ such that the matrix $\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}$ exists, where $\mathbf{M}_0=\mathrm{E}\left[\nabla_\pi\mathbf{g}\left(\mathbf{u};\pi_0,\psi_0^2\right)\right]$,

https://doi.org/10.1017/S0266466609090732 Published online by Cambridge University Press

$$\dot{\pi} = -2 \int_{S_Y \times S_V \times S_{\mathbf{W}}} \left(\mathbf{M}_0^\top \mathbf{G}\mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{G}\tilde{\mathbf{m}}\left(y, v, \mathbf{w}; \pi_0, \psi_0^2\right)$$
$$\times \psi_0^2[\phi_{0;Y}\dot{\phi}_Y\phi_{0;\mathbf{W}}^2 + \phi_{0;Y}^2\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}]$$
$$\times \mu_Y(dy)\, dv\, \mu_{\mathbf{W}}(d\mathbf{w}) + 2\int_{S_{\mathbf{W}}} \left(\mathbf{M}_0^\top \mathbf{G}\mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{G}\mathbf{s}(\mathbf{w}; \pi_0)\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}\mu_{\mathbf{W}}(d\mathbf{w}).$$

The tangent vector $(\dot{\phi}_{\mathbf{Y}}, \dot{\psi}, \dot{\phi}_{\mathbf{W}})$ used to calculate $\dot{\pi}$ also has to satisfy $\nabla\rho(\dot{\phi}_{\mathbf{Y}}, \dot{\psi}, \dot{\phi}_{\mathbf{W}}) = \mathbf{c}^\top\dot{\pi}$. Therefore $\nabla\rho$ is a bounded linear functional on the product tangent space. Furthermore, ρ is also pathwise differentiable. Hence Severini and Tripathi (2001) demonstrate that the Riesz-Fréchet theorem implies that for all $(\dot{\phi}_{\mathbf{Y}}, \dot{\psi}, \dot{\phi}_{\mathbf{W}})$ in the tangent space there exists a unique $(\phi_{\mathbf{Y}}^*, \psi^*, \phi_{\mathbf{W}}^*) \in \overline{\text{lin}}\, T(\mathcal{A}, (\phi_{0;Y}, \psi_0, \phi_{0;\mathbf{W}}))$ such that

$$\nabla\rho(\dot{\phi}_{\mathbf{Y}}, \dot{\psi}, \dot{\phi}_{\mathbf{W}}) = \langle(\phi_{\mathbf{Y}}^*, \psi^*, \phi_{\mathbf{W}}^*), (\dot{\phi}_{\mathbf{Y}}, \dot{\psi}, \dot{\phi}_{\mathbf{W}})\rangle_F.$$

The tangent vector $(\phi_{\mathbf{Y}}^*, \psi^*, \phi_{\mathbf{W}}^*)$ is called the representer of the tangent space, and Severini and Tripathi (2001) show that the lower bound for the asymptotic variance of a $\sqrt{n}$-consistent regular estimator of $\rho(\phi_{0;\mathbf{Y}}, \psi_0, \phi_{0;\mathbf{W}})$ is given by $\|(\phi_{\mathbf{Y}}^*, \psi^*, \phi_{\mathbf{W}}^*)\|_F^2$.

In our case, notice that $\nabla\rho(\dot{\phi}_{\mathbf{Y}}, \dot{\psi}, \dot{\phi}_{\mathbf{W}})$ can be rewritten as

$$\nabla\rho(\dot{\phi}_{\mathbf{Y}}, \dot{\psi}, \dot{\phi}_{\mathbf{W}})$$
$$= -2\int_{S_{\mathbf{Y}} \times S_V \times S_{\mathbf{W}}} \left\{\mathbf{c}^\top\left(\mathbf{M}_0^\top \mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top \mathbf{G}\tilde{\mathbf{m}}\left(\mathbf{y}, v, \mathbf{w}; \pi_0, \psi_0^2\right)\right.$$
$$\left.\times\psi_0^2[\phi_{0;\mathbf{Y}}\dot{\phi}_{\mathbf{Y}}\phi_{0;\mathbf{W}}^2 + \phi_{0;\mathbf{Y}}^2\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}]\right\}\mu_{\mathbf{Y}}(d\mathbf{y})\, dv\, \mu_{\mathbf{W}}(d\mathbf{w})$$
$$+ 2\int_{S_{\mathbf{W}}} \mathbf{c}^\top\left(\mathbf{M}_0^\top \mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top \mathbf{G}\mathbf{s}(\mathbf{w}; \pi_0)\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}\mu_{\mathbf{W}}(d\mathbf{w})$$
$$= -2\int_{S_{\mathbf{Y}}} \left\{\int_{S_V \times S_{\mathbf{W}}} \mathbf{c}^\top\left(\mathbf{M}_0^\top \mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top \mathbf{G}\tilde{\mathbf{m}}\left(\mathbf{y}, v, \mathbf{w}; \pi_0, \psi_0^2\right)\right.$$
$$\left.\times\psi_0^2\phi_{0;\mathbf{W}}^2 dv\, \mu_{\mathbf{W}}(d\mathbf{w})\right\}\phi_{0;\mathbf{Y}}\dot{\phi}_{\mathbf{Y}}\mu_{\mathbf{Y}}(d\mathbf{y})$$
$$-2\int_{S_{\mathbf{W}}} \left\{\int_{S_{\mathbf{Y}} \times S_V} \mathbf{c}^\top\left(\mathbf{M}_0^\top \mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top \mathbf{G}\left[\tilde{\mathbf{m}}\left(\mathbf{y}, v, \mathbf{w}; \pi_0, \psi_0^2\right)\right.\right.$$
$$\left.\left.\times\phi_{0;\mathbf{Y}}^2\psi_0^2\mu_{\mathbf{Y}}(d\mathbf{y})\, dv - \mathbf{s}(\mathbf{w}; \pi_0)\right]\right\}\phi_{0;\mathbf{W}}\dot{\phi}_{\mathbf{W}}\mu_{\mathbf{W}}(d\mathbf{w}).$$

Comparing the last expression with (A.1), and using the fact that $\int_{S_{\mathbf{Y}}}\phi_{0;\mathbf{Y}}\dot{\phi}_{\mathbf{Y}}\mu_{\mathbf{Y}}(d\mathbf{y}) = 0$ for almost all $(v, \mathbf{w}^\top)^\top$ for any $\dot{\phi}_{\mathbf{Y}}$ in the tangent space, it is possible to deduce $(\phi_{\mathbf{Y}}^*, \psi^*, \phi_{\mathbf{W}}^*)$ as

https://doi.org/10.1017/S0266466609090732 Published online by Cambridge University Press

$$\phi^*_{\mathbf{Y}}(\mathbf{y}|\upsilon,\mathbf{w}) = -\frac{1}{2}\mathbf{c}^\top\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top\mathbf{G}\left[\tilde{\mathbf{m}} - \mathrm{E}\left(\tilde{\mathbf{m}}\middle|\upsilon,\mathbf{w}\right)\right]\phi_{0;\mathbf{Y}},$$

$$\psi^*(\upsilon|\mathbf{w}) = 0,$$

$$\phi^*_{\mathbf{W}}(\mathbf{w}) = -\frac{1}{2}\mathbf{c}^\top\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top\mathbf{G}\left[\mathrm{E}\left(\tilde{\mathbf{m}}\middle|\mathbf{w}\right) - \mathbf{s}\right]\phi_{0;\mathbf{W}}.$$

It is straightforward to verify that the proposed $\left(\phi^*_{\mathbf{Y}}, \psi^*, \phi^*_{\mathbf{W}}\right)$ belongs to the tangent space, i.e., $\int_{S_{\mathbf{Y}}}\phi_{0;\mathbf{Y}}\phi^*_{\mathbf{Y}}\mu_Y(dy) = 0$ and $\int_{S_{\mathbf{W}}}\phi_{0;\mathbf{W}}\phi^*_{\mathbf{W}}\mu_{\mathbf{W}}(d\mathbf{w}) = 0$.

Therefore, as shown in Severini and Tripathi (2001), the efficiency bound is

$$\begin{aligned}
&\left\|\left(\phi^*_{\mathbf{Y}}, \psi^*, \phi^*_{\mathbf{W}}\right)\right\|_F^2 \\
&\quad= \mathbf{c}^\top\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top\mathbf{G}\mathrm{E}\left[\left(\tilde{\mathbf{m}} - \mathrm{E}\left(\tilde{\mathbf{m}}\middle|\upsilon,\mathbf{w}\right)\right)\left(\tilde{\mathbf{m}} - \mathrm{E}\left(\tilde{\mathbf{m}}\middle|\upsilon,\mathbf{w}\right)\right)^\top\right] \\
&\qquad\times\mathbf{G}^\top\mathbf{M}_0\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{c} + \mathbf{c}^\top\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top\mathbf{G}\mathrm{E} \\
&\qquad\times\left[\left(\mathrm{E}\left(\tilde{\mathbf{m}}\middle|\mathbf{w}\right) - \mathbf{s}\right)\left(\mathrm{E}\left(\tilde{\mathbf{m}}\middle|\mathbf{w}\right) - \mathbf{s}\right)^\top\right]\mathbf{G}^\top\mathbf{M}_0\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{c} \\
&\quad= \mathbf{c}^\top\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top\mathbf{G}\mathrm{E}\left[\left(\tilde{\mathbf{m}} - \mathbf{s} - \mathrm{E}\left(\tilde{\mathbf{m}}\middle|\upsilon,\mathbf{w}\right) + \mathrm{E}\left(\tilde{\mathbf{m}}\middle|\mathbf{w}\right)\right)\right. \\
&\qquad\left.\times\left(\tilde{\mathbf{m}} - \mathbf{s} - \mathrm{E}\left(\tilde{\mathbf{m}}\middle|\upsilon,\mathbf{w}\right) + \mathrm{E}\left(\tilde{\mathbf{m}}\middle|\mathbf{w}\right)\right)^\top\right]\mathbf{G}^\top\mathbf{M}_0\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{c} \\
&\quad= \mathbf{c}^\top\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top\mathbf{G}\mathrm{E}\left[\tilde{\mathbf{g}}(\mathbf{y},\upsilon,\mathbf{w};\pi_0)\tilde{\mathbf{g}}^\top(\mathbf{y},\upsilon,\mathbf{w};\pi_0)\right] \\
&\qquad\times\mathbf{G}^\top\mathbf{M}_0\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{c} \\
&\quad= \mathbf{c}^\top\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top\mathbf{G}\tilde{\Omega}_0\mathbf{G}^\top\mathbf{M}_0\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{c}.
\end{aligned}$$

However, this result is not satisfactory, because the lower bound depends upon $\mathbf{G}$, the auxiliary matrix used to solve for $\dot{\pi}$. Because by assumption $\tilde{\Omega}_0$ is nonsingular, and using the Hansen (1982) trick, it is possible to show that

$$\begin{aligned}
\text{l.b.}\left(\mathbf{c}^\top\hat{\pi}_n\right) &= \mathbf{c}^\top\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{M}_0^\top\mathbf{G}\tilde{\Omega}_0\mathbf{G}^\top\mathbf{M}_0\left(\mathbf{M}_0^\top\mathbf{G}\mathbf{M}_0\right)^{-1}\mathbf{c} \\
&\geq \mathbf{c}^\top\left(\mathbf{M}_0^\top\tilde{\Omega}_0^{-1}\mathbf{M}_0\right)^{-1}\mathbf{c}.
\end{aligned}$$

Therefore the efficiency bound for regular estimators of $\mathbf{c}^\top\pi_0$ is given by $\mathbf{c}^\top\left(\mathbf{M}_0^\top\tilde{\Omega}_0^{-1}\mathbf{M}_0\right)^{-1}\mathbf{c}$. Because $\mathbf{c} \in \Re^p$ was arbitrary, it can be concluded that the efficiency bound for regular estimators of π_0 is $\left(\mathbf{M}_0^\top\tilde{\Omega}_0^{-1}\mathbf{M}_0\right)^{-1}$. ■

https://doi.org/10.1017/S0266466609090732 Published online by Cambridge University Press