Online Supplementary Material for "Semiparametric Identification and Fisher Information"*

Juan Carlos Escanciano†
Universidad Carlos III de Madrid

February 21th, 2021

1 Supplemental Appendix

I will extensively use basic results from operator theory and Hilbert spaces in this online Supplemental Material. See Carrasco, Florens and Renault (2007) for an excellent review of these results. This Appendix is organized as follows. Section 1.1 establishes sufficient conditions for local irregular identification in models linear in nuisance parameters. Section 1.2 characterizes identification of linear continuous functionals of nuisance parameters in semiparametric models. Section 1.3 establishes sufficient conditions for identification in general nonlinear models.

1.1 Models Linear in Nuisance Parameters

Define the nuisance score operator

$$\dot{l}_{\eta(\theta)}b_\eta = \frac{f_{\theta,\eta_0+b_\eta} - f_{\theta,\eta_0}}{f_{\theta_0,\eta_0}}, \tag{1}$$

and the (negative) approximated score for θ as

$$s_\theta = \frac{f_{\theta_0,\eta_0} - f_{\theta,\eta_0}}{f_{\theta_0,\eta_0}}.$$

I drop the dependence on θ_0 and denote $\dot{l}_\eta \equiv \dot{l}_{\eta(\theta_0)}$. Define the (negative) approximated efficient score $\tilde{s}_\theta := s_\theta - \Pi_{\overline{\mathcal{R}(\dot{l}_{\eta(\theta)})}} s_\theta$, and the approximated Fisher Information

$$G(\theta) = ||\tilde{s}_\theta||^2.$$

*First version: September 20th, 2016.

†Department of Economics, Universidad Carlos III de Madrid, email: jescanci@eco.uc3m.es. Research funded by the Spanish Programa de Generación de Conocimiento, reference number PGC2018-096732-B-I00.

Let Ψ be the class of measurable functions $\psi : [0, \infty) \longrightarrow [0, \infty)$ that are increasing, right continuous at 0 and with $\psi(0) = 0$. Then, consider the following assumption.

Assumption D: (i) The map $\dot{l}_{\eta(\theta)} : T(\eta_0) \subseteq \mathcal{H} \mapsto L_2$ is linear for each θ in a neighborhood of θ_0 (ii) there exists a positive constant C such that $G(\theta) > C\psi(|\theta - \theta_0|^2)$ in a neighborhood of θ_0, where $\psi \in \Psi$.

Assumption D(i) holds for many models of interest. Assumption D(ii) follows from conditions on the derivative of $G(\theta)$ at θ_0. For example, if $G(\theta)$ is differentiable at θ_0 with full rank derivative at θ_0, then Assumption D(ii) holds with $\psi(\epsilon) = \epsilon$. This corresponds to the case of regular local identification. A necessary condition for Assumption D(ii) is that $\mathcal{N}(\dot{l}^*_{\eta(\theta)}) \neq 0$, since otherwise $G(\theta) = 0$.

Theorem 1.1 *Let Assumption D hold. Then, θ is locally identified at θ_0.*

Proof of Theorem 1.1: Write

$$\begin{aligned} \frac{f_{\theta,\eta} - f_{\theta_0,\eta_0}}{f_{\theta_0,\eta_0}} &= \frac{f_{\theta,\eta} - f_{\theta,\eta_0}}{f_{\theta_0,\eta_0}} - \frac{f_{\theta_0,\eta_0} - f_{\theta,\eta_0}}{f_{\theta_0,\eta_0}} \\ &= \dot{l}_{\eta(\theta)} b_\eta - s_\theta. \end{aligned}$$

Note that by standard least squares theory for all $b_\eta \in T(\eta_0)$, and all θ in a neighborhood of θ_0,

$$\begin{aligned} ||\dot{l}_{\eta(\theta)} b_\eta - s_\theta||^2 &\geq ||\Pi_{\overline{\mathcal{R}(\dot{l}_{\eta(\theta)})}} s_\theta - s_\theta||^2 \\ &> C\psi(|\theta - \theta_0|^2). \end{aligned}$$

This inequality implies local identification. ■

1.2 Functionals of Nuisance Parameters in Semiparametric Models

Let $\chi : \mathcal{H} \mapsto \mathbb{R}$ be a linear continuous functional, and let $r_\chi \in T(\eta_0) \subset \mathcal{H}$ be such that for all $b_\eta \in T(\eta_0)$,

$$\chi(b_\eta) = \langle b_\eta, r_\chi \rangle_{\mathcal{H}}.$$

To give a general result, I allow for θ to be infinite-dimensional, and ask the question: When does lack of identification of one parameter, here θ, have no effect, at least locally, on identification on another parameter $\chi(\eta)$?

A similar characterization to that of Proposition **??** is obtained for $\phi(\lambda) = \chi(\eta)$, allowing for singular information for both θ and the functional $\phi(\lambda) = \chi(\eta)$. Define the operator

$$A_{\eta\theta} = \left(\dot{l}^*_\eta \dot{l}_\eta\right)^- \dot{l}^*_\eta \dot{l}_\theta,$$

where B^- denotes the generalized Moore-Penrose inverse of B.

Proposition 1.1 *For the functional $\phi(\lambda) = \chi(\eta) \in \mathbb{R}$: (i) if $\mathcal{R}(\dot{l}_\theta) \cap \mathcal{R}(\dot{l}_\eta) = \{0\}$, then $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ holds iff $r_\chi \in \overline{\mathcal{R}(\dot{l}^*_\eta)}$; (ii) if $\mathcal{R}(\dot{l}_\theta) \cap \mathcal{R}(\dot{l}_\eta) \neq \{0\}$, then $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ holds if $r_\chi \in \overline{\mathcal{R}(\dot{l}^*_\eta)} \cap \mathcal{N}(A^*_{\eta\theta})$.*

Proof of Proposition 1.1: Note that for the functional $\phi(\lambda) = \chi(\eta)$, where $\chi : H \mapsto \mathbb{R}$ is a linear continuous functional with

$$\chi(b_\eta) = \langle b_\eta, r_\chi \rangle_H,$$

it holds that $\mathcal{N}(\dot{\phi}) = \{(b_\theta, b_\eta) : \langle b_\eta, r_\chi \rangle_H = 0\}$. Therefore, by the proof of Proposition **??** (which is also valid for infinite-dimensional θ, with $\tilde{I}_\theta$ interpreted as an operator), $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ iff $b'_\theta \tilde{I}_\theta b_\theta = 0$ and $\Pi_{\overline{\mathcal{R}(\dot{l}_\eta)}} \dot{l}'_\theta b_\theta = -\dot{l}_\eta b_\eta$ implies $\langle b_\eta, r_\chi \rangle_H = 0$. If $\tilde{I}_\theta$ is positive definite, then $(b_\theta, b_\eta) \in \mathcal{N}(S)$ iff $b_\theta = 0$ and $0 = \dot{l}_\eta b_\eta$. Therefore, $(b_\theta, b_\eta) \in \mathcal{N}(\dot{\phi})$ iff $\mathcal{N}(\dot{l}_\eta) \subset \mathcal{N}(\chi)$, which is equivalent to $r_\chi \in \overline{\mathcal{R}(\dot{l}^*_\eta)}$. If $\tilde{I}_\theta$ is semi-positive definite, there are two cases (i) $\mathcal{R}(\dot{l}_\theta) \cap \mathcal{R}(\dot{l}_\eta) \neq \{0\}$ and (ii) $\mathcal{R}(\dot{l}_\theta) \subset \overline{\mathcal{R}(\dot{l}_\eta)} \setminus \mathcal{R}(\dot{l}_\eta)$. In case (i), $\dot{l}_\theta b_\theta = -\dot{l}_\eta b_\eta$, and for all such b_η it must hold that $\langle b_\eta, r_\chi \rangle_H = 0$. All the solutions of $\dot{l}_\theta b_\theta = -\dot{l}_\eta b_\eta$ can be written as $b_\eta = \mathcal{N}(\dot{l}_\eta) - A_{\eta\theta} b_\theta$. Thus, the orthogonality $\langle b_\eta, r_\chi \rangle_H = 0$ holds if $r_\chi \in \overline{\mathcal{R}(\dot{l}^*_\eta)} \cap \mathcal{N}(A^*_{\eta\theta})$. In case (ii) $0 = \dot{l}_\eta b_\eta$ must imply that $(b_\theta, b_\eta) \in \mathcal{N}(\dot{\phi})$, which holds if $\mathcal{N}(\dot{l}_\eta) \subset \mathcal{N}(\chi)$ or equivalently $r_\chi \in \overline{\mathcal{R}(\dot{l}^*_\eta)}$. Therefore, if $\mathcal{R}(\dot{l}_\theta) \cap \mathcal{R}(\dot{l}_\eta) = \{0\}$ ($\tilde{I}_\theta$ is positive definite or case (ii) above) then $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ holds iff $r_\chi \in \overline{\mathcal{R}(\dot{l}^*_\eta)}$; (ii) if $\mathcal{R}(\dot{l}_\theta) \cap \mathcal{R}(\dot{l}_\eta) \neq \{0\}$ (case (i) above) then $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ holds if $\overline{\mathcal{R}(\dot{l}^*_\eta)} \cap \mathcal{N}(A^*_{\eta\theta})$.
■

Remark 1.1 *The conditions for local identification of $\chi(\eta_0)$ depend on whether θ_0 is locally identified or not. The case (ii) corresponds to the situation of local unidentification of θ_0, and it is shown that despite this lack of local identification of θ_0, $\chi(\eta_0)$ might still be locally identified. To interpret the result, one can think of $r_\chi \in \overline{\mathcal{R}(\dot{l}^*_\eta)}$ as the identification condition for $\chi(\eta_0)$ that would be needed if θ_0 was known. If θ_0 is not known, but is identified, one can treat it as known for the purpose of identifying $\chi(\eta_0)$. However, if θ_0 is not identified, an additional condition must be met to avoid the lack of identification of θ_0 to spread out to $\chi(\eta_0)$. Technically, this condition is that for all $b = (b_\theta, b_\eta)$ such that $\dot{l}_\theta b_\theta = -\dot{l}_\eta b_\eta$ (these b's are directions that lead to zero nonparametric information), it must hold that $\langle b_\eta, r_\chi \rangle_H = 0$. Under $r_\chi \in \overline{\mathcal{R}(\dot{l}^*_\eta)}$, a simple condition for this orthogonality is $r_\chi \in \mathcal{N}(A^*_{\eta\theta})$.*

Remark 1.2 *In both cases $r_\chi \in \overline{\mathcal{R}(\dot{l}^*_\eta)} \setminus \mathcal{R}(\dot{l}^*_\eta)$ corresponds to the case of zero information for $\phi(\lambda) = \chi(\eta)$ at $\phi(\lambda_0) = \chi(\eta_0)$. Regular identification of $\chi(\eta)$ in case (ii) requires that for all r^*_χ that solve $r_\chi = \dot{l}^*_\eta r^*_\chi$ it holds that $r^*_\chi \in \mathcal{N}(\dot{l}^*_\theta)$. Under this condition, lack of identification of θ_0 does not affect regular identification of $\chi(\eta_0)$.*

Van der Vaart (1991) has shown that positive information of $\chi(\eta_0)$ is equivalent to $r_\chi \in \mathcal{R}(\dot{l}^*_\eta)$ when θ_0 is locally regularly identified and η_0 is identified. Proposition 1.1 characterizes local regular and irregular identification of $\chi(\eta_0)$, allowing for θ_0 to be locally regular or irregularly identified, or even unidentified. The results of Proposition 1.1 are applied to measures of risk aversion in Example 3 on the Euler equation.

1.3 General Nonlinear Models

The following modulus of continuity is shown to be useful for the study of identification

$$\varpi(\epsilon) = \sup_{\lambda \in \mathcal{B}_\delta(\lambda_0): ||\left(f_\lambda - f_{\lambda_0}\right) f_{\lambda_0}^{-1}|| \leq \epsilon} |\phi(\lambda) - \phi(\lambda_0)| \,. \tag{2}$$

I drop the dependence of $\varpi(\epsilon)$ on δ for simplicity of notation. Lemma 1.3 below shows that $\varpi(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$ is sufficient for local identification of $\phi(\lambda_0)$. A related modulus of continuity was introduced in Donoho and Liu (1987) for the purpose of obtaining bounds on the optimal rate of convergence for functionals of a density (they assume identification and use the Hellinger metric). Using $|| \left(f_\lambda - f_{\lambda_0}\right) f_{\lambda_0}^{-1}||$ is convenient because we can exploit simultaneously the linearity of certain models and the Hilbert space structure.

Lemma *If there exists $\delta > 0$ such that $\varpi(\epsilon) \to 0$ as $\epsilon \to 0$, then $\phi(\lambda_0)$ is locally identified.*
Proof of Lemma 1.3: Suppose that $\phi(\lambda_0)$ is not locally identified. Then, for all $\delta > 0$, we can find a $\lambda^* \in \Lambda_\delta(\lambda_0)$ such that $\|(f_{\lambda^*} - f_{\lambda_0}) / f_{\lambda_0}\| = 0$ and $\phi(\lambda^*) \neq \phi(\lambda_0)$, and therefore, for all $\epsilon > 0$,

$$\varpi(\epsilon) \geq |\phi(\lambda^*) - \phi(\lambda_0)| > 0,$$

showing that $\varpi(\epsilon)$ does not converge to zero as $\epsilon \to 0$. ■

The following result provides a general local identification result. Recall Ψ is the class of measurable functions $\psi : [0, \infty) \longrightarrow [0, \infty)$ that are increasing, right continuous at 0 and with $\psi(0) = 0$.

Assumption N: *For all $\varepsilon > 0$, there exists $\delta > 0$, $\psi_1, \psi_2 \in \Psi$, and a continuous linear operator $S : T(\lambda_0) \subseteq H \mapsto L_2$, such that for all $\lambda = (\theta, \eta) \in \mathcal{B}_\delta(\lambda_0)$,*

(i)

$$\|(f_\lambda - f_{\lambda_0}) / f_{\lambda_0} - S(\lambda - \lambda_0)\| < \varepsilon \psi_1 \left(\|\lambda - \lambda_0\|_{\mathbf{H}}\right);$$

(ii)

$$|\phi(\lambda) - \phi(\lambda_0)| \leq \psi_2 \left(\|\lambda - \lambda_0\|_{\mathbf{H}}\right); \text{ and}$$

(iii)

$$\inf_{\lambda \in \mathcal{B}_\delta(\lambda_0)} \frac{||S(\lambda - \lambda_0)||}{\psi_1 \left(\|\lambda - \lambda_0\|_{\mathbf{H}}\right)} > 0.$$

Assumption N(i) and N(ii) are mild smoothness conditions that often hold in applications. Condition N(iii) is a positive nonparametric generalized information condition. Then, I have the following result.

Theorem 1.2 *Let Assumption N hold. Then, $\phi(\lambda)$ is locally identified at $\phi(\lambda_0)$.*

Proof of Theorem 1.2: Assumptions N(i-ii) imply that if $||\left(f_\lambda - f_{\lambda_0}\right) f_{\lambda_0}^{-1}|| \le \epsilon$ then we can find a positive constant C and $0 < \varepsilon < C$ such that for all $\lambda = (\theta, \eta) \in \mathcal{B}_\delta(\lambda_0)$,

$$C\psi_1\left(\|\lambda - \lambda_0\|_{\mathbf{H}}\right) \le \|S(\lambda - \lambda_0)\| \le \varepsilon\psi_1\left(\|\lambda - \lambda_0\|_{\mathbf{H}}\right) + \epsilon,$$

which in turn implies

$$\psi_1\left(\|\lambda - \lambda_0\|_{\mathbf{H}}\right) \le \frac{\epsilon}{C - \varepsilon}.$$

Hence, by Assumption N(ii)

$$\begin{aligned} \varpi(\epsilon) &= \sup_{\lambda \in \mathcal{B}_\delta(\lambda_0): ||\left(f_\lambda - f_{\lambda_0}\right) f_{\lambda_0}^{-1}|| \le \epsilon} \left|\phi(\lambda) - \phi(\lambda_0)\right|, \\ &\le \sup_{\lambda \in \mathcal{B}_\delta(\lambda_0): \psi_1\left(\|\lambda - \lambda_0\|_{\mathbf{H}}\right) \le \frac{\epsilon}{C - \varepsilon}} \psi_2\left(\|\lambda - \lambda_0\|_{\mathbf{H}}\right) \\ &\le \psi_2\left(\psi_1^{-1}\left(\frac{\epsilon}{C - \varepsilon}\right)\right) \\ &\to 0 \text{ as } \epsilon \to 0. \end{aligned}$$

Thus, the Theorem follows from Lemma 1.3. ■

1.3.1 A Counterexample

I provide a counterexample, building on that given in Chen et al. (2014, pg. 791), that shows that regular identification is not equivalent to $I_\phi > 0$ in general (and hence to Van der Vaart's (1991) differentiability condition). Let $\lambda = (\lambda_1, \lambda_2, ...)$ be a sequence of real numbers. Let $(p_1, p_2, ...)$ be probabilities, $p_j > 0$, $\sum_{j=1}^{\infty} p_j = 1$. Let $f(x)$ be a twice continuously differentiable function of a scalar x that is bounded with bounded second derivative. Suppose $f(x) = 0$ if and only if $x \in \{0, 1\}$ and $\partial f(0)/\partial x = 1$. Let $m(\lambda) = (f(\lambda_1), f(\lambda_2), ...)$ also be a sequence with $\|m(\lambda)\|^2 = \sum_{j=1}^{\infty} p_j f^2(\lambda_j) < \infty$. Then, for $\|\lambda\|_\Lambda = \left(\sum_{j=1}^{\infty} p_j \lambda_j^4\right)^{1/4}$ the mapping m is Frechet differentiable at $\lambda_0 = 0$ with derivative $Sb = b$, but $\lambda_0 = 0$ is not locally identified (Chen et al. 2014).

Consider the nonlinear functional

$$\phi(\lambda) = \sum_{j=1}^{\infty} f(\lambda_j) p_j.$$

This functional has a derivative at $\lambda_0 = 0$ given by

$$\dot{\phi}(b) = \sum_{j=1}^{\infty} b_j p_j,$$

and by Cauchy-Schwarz

$$\begin{aligned} \left|\dot{\phi}(b)\right|^2 &\leq \left(\sum_{j=1}^{\infty} b_j^2 p_j\right) \\ &= \|Sb\|^2 . \end{aligned}$$

Hence, $I_\phi \geq 1 > 0$. However, the functional is not identified, since $\phi(\alpha^k) = 0 = \phi(0)$, where $\alpha^k = (0, .., 0, 1, 1, 1...)$ has zeros in the first k positions and a one everywhere else.

References

Carrasco, M., J. P. Florens, and E. Renault (2007): "Linear Inverse Problem in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, vol. 6, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: North-Holland, 5633–5751.

Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2014), "Identification in Semiparametric and Nonparametric Conditional Moment Models," *Econometrica*, 82, 785-809.

Donoho, D.L. and Liu, R.C. (1987): "Geometrizing Rates of Convergence, I." Technical Report 137, Dept. Statistics, Univ. California, Berkeley.

van der Vaart, A. W. (1991): "On Differentiable Functionals," *The Annals of Statistics*, 19, 178-204.