Econometrica Supplementary Material

SUPPLEMENT TO "LOCAL IDENTIFICATION OF NONPARAMETRIC AND SEMIPARAMETRIC MODELS"

(*Econometrica*, Vol. 82, No. 2, March 2014, 785–809)

BY XIAOHONG CHEN, VICTOR CHERNOZHUKOV, SOKBAE LEE, AND WHITNEY K. NEWEY

THIS SUPPLEMENTAL MATERIAL GIVES PROOFS for the results of Sections 4 and 5 of the paper as well as some additional results and discussions.

S1. DISCUSSION AND AN EXAMPLE FOR LEMMA 3

The first item we consider is discussion and examples of the genericity result in Lemma 3. Below, we provide examples for Lemma 3 for the case $\mathcal{A} = \mathcal{B} = L^2[0, 1]$ that highlight the range of algorithms permitted by conditions 1 and 2 preceding Lemma 3, including cases where various restrictions on m' are imposed: boundedness, compactness, weak positivity, and density restrictions. Genericity arguments use the idea of randomization, and are often employed in economic theory, functional analysis, and probability theory; see, for example, Anderson and Zame (2000), Marcus and Pisier (1981), Ledoux and Talagrand (2011). Andrews (2011) previously used a related notion of genericity, called prevalence within bounded sets, to argue that rich classes of operators induced by densities in nonparametric IV are L^2-complete. Though inspired in part by Andrews (2011), Lemma 3 of the main article (Chen, Chernozhukov, Lee, and Newey (2014)) uses a somewhat different notion of genericity than prevalence.[1] We also note that while this construction implies identification with probability 1, it does not regulate in any way the strength of identification, and hence has no bearing on the choice of an inferential method.

It is useful to give explicit examples of the randomization algorithms obeying conditions 1 and 2 listed in Section 3. Suppose $\mathcal{A} = \mathcal{B} = L^2[0, 1]$, and that m' is an integral operator

$$m'\delta = \int K(\cdot, t)\delta(t)\,dt.$$

[1] Informally speaking, prevalence requires that it should be possible to construct a randomization device such that *all* finite dimensional distributions for λ_j's are absolutely continuous, that is, the distribution of $(\lambda_{j_1}, \dots, \lambda_{j_k})$ needs to be continuous with respect to the Lebesgue measure on $\mathbb{R}^k$, for any $(j_1, \dots, j_k) \subset \{1, 2, \dots, N\}$, and any $k \in \{1, 2, \dots\}$. The notion that we use requires *only* that the *one-dimensional* marginal distributions for λ_j are absolutely continuous for any j. The distinction is actually important to cover cases, where perfect dependence between some λ_j's may be required to maintain conditions imposed on the operator, such as, for example, the kernel of the operator being a conditional density.

 DOI: 10.3982/ECTA9988

The kernel K of this operator is generated as follows. The nature performs step 1 by selecting two, possibly different, orthonormal bases $\{\phi_j\}$ and $\{\varphi_j\}$ in $L^2[0,1]$. The nature performs step 2 by first selecting a bounded sequence $0 < \sigma_j < \sigma$ for $j = 0, 1, \dots,$ sampling u_j as i.i.d. $U[-1,1]$, and then setting $\lambda_j = u_j \sigma_j$. Finally, for some scalar $\kappa > 0$, it sets

$$K = \kappa \left(\sum_{j=0}^{\infty} \lambda_j \phi_j \varphi_j \right).$$

The operator defined in this way is well-defined over $\mathcal{A}$ and is bounded, but it need not be compact. If compactness is required, we impose $\sum_{j=1}^{\infty} \sigma_j^2 < \infty$ in the construction. If $K \geq 0$ is required, we can impose $\phi_0 = 1$, $\varphi_0 = 1$, $|\varphi_j| \leq c$, and $|\phi_j| \leq c$, for all j, where $c > 1$ is a constant, and $\sum_{j=0}^{\infty} \sigma_j < \infty$, and define instead λ_0 as $c \sum_{j=1}^{\infty} \lambda_j + |u_0| \sigma_0$. If in addition to positivity, $\int K(z,t)\,dt = 1$ is required, for example if $K(z,t) = f(t|z)$ is a conditional density, then we select $\kappa > 0$ so that $\kappa \lambda_0 = 1$. This algorithm for generating m' trivially obeys conditions 1 and 2 stated above Lemma 3. Furthermore, u_j need not be i.i.d. Take the extreme, opposite example, and set $u_j = U[-1,1]$ for all j, that is, u_j's are perfectly dependent. The resulting algorithm for generating m' still trivially obeys conditions 1 and 2. The latter point—of allowing perfect dependence—is useful for highlighting the differences with the approach and various examples given in Andrews (2011); other than that, our point is the same.

An important example where dependence matters is the case with normal instrumental regression, where the endogenous variable X conditional on the instrument $Z = z$ follows a normal distribution with mean ρz (and variance normalized to 1). Here we let $\mathcal{A} = \mathcal{B} = L^2(\mathbb{R})$ equipped with standard normal density as a measure. In this case, m' is an integral operator

$$m'\delta = \int K(\cdot, t)\delta(t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt,$$

similarly to what we had above, where $K(t,z)$ has the following well-known representation:

$$K(t,z) = \sum_{j=0}^{\infty} \rho^j \phi_j(t) \varphi_j(z),$$

where $(\phi_j)_{j=0}^{\infty}$ and $(\varphi_j)_{j=0}^{\infty}$ are the orthonormal (Hermite) polynomials. Hence, if the nature draws ρ from an absolutely continuous density on $(-1,1)$, then the full rank condition holds with probability 1. Note that the generalized Fourier coefficients $(\rho^j)_{j=0}^{\infty}$ exhibit perfect dependence here. To see that our randomization algorithm permits this, let the nature draw ρ as specified above and draw λ_0 as an independent from ρ random variable with support $(0,\infty)$

having an absolutely continuous distribution. Then nature sets $\lambda_j = \rho^j \lambda_0$, $\kappa = 1/\lambda_0$, and

$$K(t, z) = \kappa \left(\sum_{j=0}^{\infty} \lambda_j \phi_j(t) \varphi_j(z) \right).$$

S2. PROOFS FOR SECTION 4

The next part of the supplemental material is proofs for the general semiparametric model considered in Section 4. Before proving these results, we give two useful intermediate results.

S2.1. *Useful Results on Projections on Linear Subspaces*

Let $\mathrm{Proj}(b|\overline{\mathcal{M}})$ denote the orthogonal projection of an element b of a Hilbert space on a closed linear subset $\overline{\mathcal{M}}$ of that space.

LEMMA S1: *If* (a) $\overline{\mathcal{M}}$ *is a closed linear subspace of a Hilbert space* $\mathcal{H}$; (b) $b_j \in \mathcal{H}$ $(j = 1, \dots, p)$; (c) *the* $p \times p$ *matrix* Π *with* $\Pi_{jk} = \langle b_j - \mathrm{Proj}(b_j|\overline{\mathcal{M}}), b_k - \mathrm{Proj}(b_k|\overline{\mathcal{M}}) \rangle$ *is nonsingular*, *then*, *for* $b = (b_1, \dots, b_p)^T$, *there exists* $\varepsilon > 0$ *such that*, *for all* $a \in \mathbb{R}^p$ *and* $\zeta \in \overline{\mathcal{M}}$,

$$\|b^T a + \zeta\| \geq \varepsilon\big(|a| + \|\zeta\|\big).$$

PROOF: Let $\bar{b}_j = \mathrm{Proj}(b_j|\overline{\mathcal{M}})$, $\tilde{b}_j = b_j - \bar{b}_j$, $\bar{b} = (\bar{b}_1, \dots, \bar{b}_p)^T$, and $\tilde{b} = (\tilde{b}_1, \dots, \tilde{b}_p)^T$. Note that, for $\varepsilon_1 = \sqrt{\lambda_{\min}(\Pi)/2}$,

$$\begin{aligned}
&\|b^T a + \zeta\| \\
&\quad = \sqrt{\|\tilde{b}^T a + \zeta + \bar{b}^T a\|^2} = \sqrt{\|\tilde{b}^T a\|^2 + \|\zeta + \bar{b}^T a\|^2} \\
&\quad \geq \big(\|\tilde{b}^T a\| + \|\zeta + \bar{b}^T a\|\big)/\sqrt{2} = \big(\sqrt{a^T \Pi a} + \|\zeta + \bar{b}^T a\|\big)/\sqrt{2} \\
&\quad \geq \varepsilon_1 |a| + \|\zeta + \bar{b}^T a\|/\sqrt{2}.
\end{aligned}$$

Also note that, for any $C^* \geq \sqrt{\sum_j \|\bar{b}_j\|^2}$, it follows by the triangle and Cauchy–Schwarz inequalities that

$$\|\bar{b}^T a\| \leq \sum_j \|\bar{b}_j\| |a_j| \leq C^* |a|.$$

Choose C^* big enough that $\varepsilon_1/\sqrt{2}C^* \le 1$. Then by the triangle inequality,

$$\begin{aligned}\|\zeta + \bar{b}^T a\|/\sqrt{2} &\ge (\varepsilon_1/\sqrt{2}C^*)\|\zeta + \bar{b}^T a\|/\sqrt{2} = \varepsilon_1\|\zeta + \bar{b}^T a\|/2C^* \\ &\ge \varepsilon_1(\|\zeta\| - \|\bar{b}^T a\|)/2C^* \ge \varepsilon_1(\|\zeta\| - C^*|a|)/2C^* \\ &= (\varepsilon_1/2C^*)\|\zeta\| - \varepsilon_1|a|/2.\end{aligned}$$

Then combining the inequalities, for $\varepsilon = \min\{\varepsilon_1/2, \varepsilon_1/2C^*\}$,

$$\begin{aligned}\|b^T a + \zeta\| &\ge \varepsilon_1|a| + (\varepsilon_1/2C^*)\|\zeta\| - \varepsilon_1|a|/2 \\ &= (\varepsilon_1/2)|a| + (\varepsilon_1/2C^*)\|\zeta\| \ge \varepsilon(|a| + \|\zeta\|).\end{aligned}$$

Q.E.D.

LEMMA S2: *If Assumption* 4 *is satisfied, then there is an* $\varepsilon > 0$ *such that, for all* $(\beta, g) \in \mathbb{R}^p \times \mathcal{N}_g'$,

$$\varepsilon(|\beta - \beta_0| + \|m_g'(g - g_0)\|_{\mathcal{B}}) \le \|m'(\alpha - \alpha_0)\|_{\mathcal{B}}.$$

PROOF: Apply Lemma S1 with $\mathcal{H}$ being the Hilbert space $\mathcal{B}$ described in Section 4, $\overline{\mathcal{M}}$ in Lemma S1 being the closed linear span of $\mathcal{M} = \{m_g'(g - g_0) : g \in \mathcal{N}_g'\}$, $b_j = m_\beta' e_j$ for the jth unit vector e_j, and $a = \beta - \beta_0$. Then, for all $(\beta, g) \in \mathbb{R}^p \times \mathcal{N}_g'$, we have

$$\begin{aligned}&m'(\alpha - \alpha_0) = b^T a + \zeta, \quad b^T a = m_\beta'(\beta - \beta_0), \\ &\zeta = m_g'(g - g_0) \in \overline{\mathcal{M}}.\end{aligned}$$

The conclusion then follows from the conclusion of Lemma S1. *Q.E.D.*

We next give the proofs of Theorems 7 and 8.

S2.2. *Proof of Theorem 7*

Since Assumption 4 is satisfied, the conclusion of Lemma S2 holds. Let ε be from the conclusion of Lemma S2. Also let $\mathcal{N}_g = \mathcal{N}_g' \cap \mathcal{N}_g^\beta$ for $\mathcal{N}_g'$ from Assumption 4 and $\mathcal{N}_g^\beta$ from Assumption 5. In addition, let B be from Assumption 5 with

$$\begin{aligned}&\sup_{g \in \mathcal{N}_g^\beta} \mathrm{E}\Big[\sup_{\beta \in B}\big|\partial \mathrm{E}[\rho(Y, X, \beta, g)|W]/\partial\beta - \partial \mathrm{E}[\rho(Y, X, \beta_0, g_0)|W]/\partial\beta\big|^2\Big] \\ &\quad < \varepsilon^2.\end{aligned}$$

Then by $m(\beta_0, g)$ linear in g and expanding each element of $m(\beta, g)(W) = \mathrm{E}[\rho(Y, X, \beta, g)|W]$ in β, it follows that, for each $(\beta, g) \in B \times \mathcal{N}_g$, if $\beta \neq \beta_0$,

$$\begin{aligned}\left\|m(\alpha) - m'(\alpha - \alpha_0)\right\|_{\mathcal{B}} &= \left\|m(\beta, g) - m(\beta_0, g) - m'_\beta(\beta - \beta_0)\right\|_{\mathcal{B}} \\ &= \left\|\left[\partial m(\tilde{\beta}, g)/\partial\beta - m'_\beta\right](\beta - \beta_0)\right\|_{\mathcal{B}} \\ &\leq \left\|m'_\beta(\tilde{\beta}, g) - m'_\beta\right\|_{\mathcal{B}}|\beta - \beta_0| \\ &< \varepsilon|\beta - \beta_0| \leq \varepsilon\left(|\beta - \beta_0| + \left\|m'_g(g - g_0)\right\|_{\mathcal{B}}\right) \\ &\leq \left\|m'(\alpha - \alpha_0)\right\|_{\mathcal{B}},\end{aligned}$$

where $\tilde{\beta}$ is a mean value depending on W that actually differs from row to row of

$$m'_\beta(\tilde{\beta}, g) = \partial\mathrm{E}\left[\rho(Y, X, \tilde{\beta}, g)|W\right]/\partial\beta.$$

Thus, $\|m(\alpha) - m'(\alpha - \alpha_0)\|_{\mathcal{B}} < \|m'(\alpha - \alpha_0)\|_{\mathcal{B}}$, implying $m(\alpha) \neq 0$, giving the first conclusion.

To show the second conclusion, consider $(\beta, g) \in \mathcal{N}$. If $\beta \neq \beta_0$, then it follows as above that $m(\alpha) \neq 0$. If $\beta = \beta_0$ and $g \neq g_0$, then, by linearity in g, we have $\|m(\alpha) - m'(\alpha - \alpha_0)\|_{\mathcal{B}} = 0$, while $\|m'(\alpha - \alpha_0)\|_{\mathcal{B}} = \|m'_g(g - g_0)\|_{\mathcal{B}} > 0$, so $m(\alpha) \neq 0$ follows as in the proof of Theorem 1. *Q.E.D.*

S2.3. *Proof of Theorem 8*

Since Assumption 4 is satisfied, the conclusion of Lemma S2 holds. Let ε be from the conclusion of Lemma S2. Define B as in the proof of Theorem 7. By Assumption 2, for $g \in \mathcal{N}''_g$, $\|m(\beta_0, g) - m'_g(g - g_0)\|_{\mathcal{B}} \leq L\|g - g_0\|^r_{\mathcal{A}}$. Then, similarly to the proof of Theorem 7 for all $\alpha \in \mathcal{N}$ with $\alpha \neq \alpha_0$,

$$\begin{aligned}&\left\|m(\alpha) - m'(\alpha - \alpha_0)\right\|_{\mathcal{B}} \\ &\quad\leq \left\|m(\beta, g) - m(\beta_0, g) - m'_\beta(\beta - \beta_0)\right\|_{\mathcal{B}} \\ &\qquad+ \left\|m(\beta_0, g) - m'_g(g - g_0)\right\|_{\mathcal{B}} \\ &\quad< \varepsilon|\beta - \beta_0| + L\|g - g_0\|^r_{\mathcal{A}} \leq \varepsilon|\beta - \beta_0| + \varepsilon\left\|m'(g - g_0)\right\|_{\mathcal{B}} \\ &\quad\leq \left\|m'(\alpha - \alpha_0)\right\|_{\mathcal{B}}.\end{aligned}$$

The conclusion follows as in the conclusion of Theorem 2. *Q.E.D.*

S3. A SINGLE INDEX IV EXAMPLE

We next turn to the single index IV model as an example of the results for local identification of semiparametric models. We focus on how imposing a single index structure can lower the need for instrumental variables. This example is motivated by econometric applications that have too many covariates for fully nonparametric estimation to be practical, that is, they suffer from the curse of dimensionality. A single index model can help with this problem because it only has one nonparametric dimension. Here, we consider a single index model with endogeneity, given by

$$Y = g_0\big(X_1 + X_2^T\beta_0\big) + U, \quad \mathrm{E}[U|W] = 0, \tag{S3.1}$$

where β_0 is a vector of unknown parameters, $g_0(\cdot)$ is an unknown function, and W are instrumental variables. Here, the nonparametric part is just one dimensional rather than having the same dimension as X. This model is nonlinear in Euclidean parameters, and so is an example where our results apply. Our results add to the literature on dimension reduction with endogeneity, by showing how identification of an index model requires fewer instrumental variables than a fully nonparametric IV model. We could generalize the results to multiple indices, but focus on a single index for simplicity.

The location and scale of the parametric part are not identified separately from g_0, and hence, we normalize the constant to zero and the coefficient of X_1 to 1. Here,

$$m(\alpha)(W) = \mathrm{E}\big[Y - g\big(X_1 + X_2^T\beta\big)|W\big].$$

Let $V = X_1 + X_2^T\beta_0$ and, for differentiable $g_0(V)$, let

$$m'_\beta = -\mathrm{E}\big[g_0'(V)X_2^T|W\big].$$

Let ζ_j^* denote the projection of $m'_\beta e_j = -\mathrm{E}[g_0'(V)X_{2j}|W]$ on the mean squared closure of the set $\{\mathrm{E}[h(V)|W] : \mathrm{E}[h(V)^2] < \infty\}$ and Π the matrix with $\Pi_{jk} = \mathrm{E}[(m'_\beta e_j - \zeta_j^*)(m'_\beta e_k - \zeta_k^*)]$.

THEOREM S3: *Consider the model of equation* (S3.1). *If* (a) $g_0(V)$ *is continuously differentiable with bounded derivative* $g_0'(V)$ *satisfying* $|g_0'(\tilde{V}) - g_0'(V)| \le C_g|\tilde{V} - V|$ *for some* $C_g > 0$, (b) $\mathrm{E}[|X_2|^4] < \infty$, *and* (c) Π *is nonsingular*, *then there is a neighborhood* B *of* β_0 *and* $\delta > 0$ *such that*, *for*

$$\mathcal{N}_g^\delta = \Big\{g : g(v) \text{ is continuously differentiable and } \sup_v\big|g'(v) - g_0'(v)\big| \le \delta\Big\},$$

β_0 *is locally identified for* $\mathcal{N} = B \times \mathcal{N}_g^\delta$. *Furthermore*, *if there is* $\mathcal{N}_g'$ *such that* $\mathrm{E}[g(V) - g_0(V)|W]$ *is bounded complete on the set* $\{g(V) - g_0(V) : g \in \mathcal{N}_g'\}$, *then* (β_0, g_0) *is locally identified for* $\mathcal{N} = B \times (\mathcal{N}_g^\delta \cap \mathcal{N}_g')$.

PROOF: The proof will proceed by verifying the conditions of Theorem 7. Note that Assumption 4 is satisfied. We now check Assumption 5. Note that, for any $\delta > 0$ and $g \in \mathcal{N}_g^\delta$, $g(X_1 + X_2^T\beta)$ is continuously differentiable in β with $\partial g(X_1 + X_2^T\beta)/\partial\beta = g'(X_1 + X_2^T\beta)X_2$. Also, for Δ a $p \times 1$ vector and $\bar{B}$ a neighborhood of zero, it follows by boundedness of g_0' and the specification of $\mathcal{N}_g^\delta$ that, for some $C > 0$,

$$\mathrm{E}\Big[\sup_{\Delta\in\bar{B}}\big|g'\big(X_1 + X_2^T(\beta+\Delta)\big)X_2\big|\big|W\Big] \le C\mathrm{E}\big[|X_2||W\big] < \infty \quad \text{a.s.}$$

Therefore, by the dominated convergence theorem, $m(\alpha)(W) = \mathrm{E}[Y - g(X_1 + X_2^T\beta)|W]$ is continuously differentiable in β a.s. with

$$\partial m(\alpha)(W)/\partial\beta = -\mathrm{E}\big[g'\big(X_1 + X_2^T\beta\big)X_2|W\big].$$

Next, consider any $\varepsilon > 0$ and let B and δ satisfy

$$B = \big\{\beta : |\beta - \beta_0|^2 < \varepsilon^2/4C_g^2\mathrm{E}\big[|X_2|^4\big]\big\} \quad \text{and} \quad \delta^2 < \varepsilon^2/4\mathrm{E}\big[|X_2|^2\big].$$

Then, for $g \in \mathcal{N}_g^\delta$, we have, for $v(X,\beta) = X_1 + X_2^T\beta$,

$$\begin{aligned}
&\mathrm{E}\Big[\sup_{\beta\in B}\big|\partial m(\alpha)(W)/\partial\beta - m_\beta'(W)\big|^2\Big] \\
&\quad = \mathrm{E}\Big[\sup_{\beta\in B}\big|\mathrm{E}\big[\big\{g'\big(v(X,\beta)\big) - g_0'(V)\big\}X_2|W\big]\big|^2\Big] \\
&\quad \le \mathrm{E}\Big[|X_2|^2\sup_{\beta\in B}\big|g'\big(v(X,\beta)\big) - g_0'(V)\big|^2\Big] \\
&\quad \le 2\mathrm{E}\Big[|X_2|^2\sup_{\beta\in B}\big|g'\big(v(X,\beta)\big) - g_0'\big(v(X,\beta)\big)\big|^2\Big] \\
&\qquad + 2\mathrm{E}\Big[|X_2|^2\sup_{\beta\in B}\big|g_0'\big(v(X,\beta)\big) - g_0'(V)\big|^2\Big] \\
&\quad \le 2\delta^2\mathrm{E}\big[|X_2|^2\big] + 2C_g^2\mathrm{E}\big[|X_2|^4\big]\sup_{\beta\in B}|\beta - \beta_0|^2 < \varepsilon^2.
\end{aligned}$$

Thus, Assumption 5 is satisfied, so the first conclusion follows by the first conclusion of Theorem 7. Also, $m_g'(g - g_0) = \mathrm{E}[g(V) - g_0(V)|W]$ the rank condition for m_g' follows by the last bounded completeness on $\mathcal{N}_g'$, so that the final conclusion follows by the final conclusion of Theorem 7. *Q.E.D.*

Since this model includes as a special case the linear simultaneous equations model, the usual rank and order conditions are still necessary for Π to be nonsingular for all possible models, and hence are necessary for identification.

Relative to the linear nonparametric IV model in Newey and Powell (2003), the index structure lowers the requirements for identification by requiring that $m_g' h = -\mathrm{E}[h(V)|W]$ be complete on $\mathcal{N}_g'$, rather than completeness of the conditional expectation of functions of X given W. For example, it may be possible to identify β_0 and g_0 with only two instrumental variables, one of which is used to identify g_0 and functions of the other being used to identify β_0.

To further explain, we can give more primitive conditions for nonsingularity of Π. The following result gives a necessary condition for Π to be nonzero (and hence nonsingular) as well as a sufficient condition for nonsingularity of Π.

THEOREM S4: *Consider the model of* (S3.1). *If Π is nonsingular, then the conditional distribution of W given V is not complete. Also, if there is a measurable function $T(W)$ such that the conditional distribution of V given W depends only on $T(W)$ and, for every $p \times 1$ vector $\lambda \neq 0$, $\mathrm{E}[g_0'(V)\lambda^T X_2|W]$ is not measurable with respect to $T(W)$, then Π is nonsingular.*

PROOF: Suppose first that the conditional distribution of W given V is complete. Note that by the projection definition, for all $h(V)$ with finite mean squared, we have

$$\begin{aligned} 0 &= \mathrm{E}\big[\big\{-\mathrm{E}\big[g_0'(V)X_{2j}|W\big] - \zeta_j^*(W)\big\}\mathrm{E}\big[h(V)|W\big]\big] \\ &= \mathrm{E}\big[\big\{-\mathrm{E}\big[g_0'(V)X_{2j}|W\big] - \zeta_j^*(W)\big\}h(V)\big]. \end{aligned}$$

Therefore,

$$\mathrm{E}\big[-\mathrm{E}\big[g_0'(V)X_{2j}|W\big] - \zeta_j^*(W)|V\big] = 0.$$

Completeness of the conditional distribution of W given V then implies that $-\mathrm{E}[g_0'(V)X_{2j}|W] - \zeta_j^*(W) = 0$, and hence $\Pi_{jj} = 0$. Since this is true for each j, we have $\Pi = 0$, Π is singular.

Next, consider the second hypothesis and $\lambda \neq 0$. Let $\zeta_\lambda^*(W)$ denote the projection of $-\mathrm{E}[g_0'(V)\lambda^T X_2|W]$ on $\overline{\mathcal{M}}$. Since $\mathrm{E}[h(V)|W] = \mathrm{E}[h(V)|T(W)]$, it follows that $\zeta_\lambda^*(W)$ is measurable with respect to (i.e., is a function of) $T(W)$. Since $\mathrm{E}[g_0'(V)\lambda^T X_2|W]$ is not measurable with respect to $T(W)$, we have $-\mathrm{E}[g_0'(V)\lambda^T X_2|W] - \zeta_\lambda^*(W) \neq 0$, so that

$$\lambda^T \Pi \lambda = \mathrm{E}\big[\big\{-\mathrm{E}\big[g_0'(V)\lambda^T X_2|W\big] - \zeta_\lambda^*(W)\big\}^2\big] > 0.$$

Since this is true for all $\lambda \neq 0$, it follows that Π is positive definite, and hence nonsingular. *Q.E.D.*

To explain the conditions of this result, note that if there is only one variable in W, then the completeness condition (of W given V) can hold and hence Π

can be singular. If there is more than one variable in W, then generally completeness (of W given V) will not hold, because completeness would be like identifying a function of more than one variable (i.e., W) with one instrument (i.e., V). If W and V are joint Gaussian and V and W are correlated, then completeness holds (and hence Π is singular) when W is one dimensional but not otherwise. In this sense, having more than one instrument in W is a necessary condition for nonsingularity of Π. Intuitively, one instrument is needed for identification of the one-dimensional function $g_0(V)$, so that more than one instrument is needed for identification of β.

The sufficient condition for nonsingularity of Π is stronger than noncompleteness. It is essentially an exclusion restriction, where $\mathrm{E}[g_0'(V)X_2|W]$ depends on W in a different way than the conditional distribution of V depends on W. This condition can be shown to hold if W and V are Gaussian, W is two-dimensional, and $\mathrm{E}[g_0'(V)X_2|W]$ depends on all of W.

S4. PROOFS FOR SECTION 5

This section provides proofs for the results for identification of the CCAPM model in Section 5. We also give two supplementary results. One is a global identification result (Theorem S5) based on boundedness of g and the other is a functional version of the Perron–Frobenius theorem that is used to prove Theorem 11.

S4.1. *Proof of Theorem 9*

The proof will proceed by verifying the conditions of Theorem 7 for $\rho(Y,\beta,g)$ from Section 5 of Chen et al. (2014). We first check the first part of Assumption 4. Note that the mapping $m':\mathbb{R}^2\times\mathcal{G}\longrightarrow\mathcal{B}$ is given by $m'(\alpha-\alpha_0)=m'_\beta(\beta-\beta_0)+m'_g(g-g_0)$, where

$$m'_\beta(\beta-\beta_0)=\mathrm{E}\big[A_t g_0(c_{t+1})X_t^T|W_t\big](\beta-\beta_0)\quad\text{and}$$
$$m'_g(g-g_0)=\mathrm{E}\big[A_t\big\{g(c_{t+1})-g_0(c_{t+1})\big\}|W_t\big]-\big\{g(c_t)-g_0(c_t)\big\}e.$$

Therefore, the mapping m' is obviously linear. Note that $\mathrm{E}[D_t^2|W_t]$ and $\mathrm{E}[D_t|W_t]$ exist with probability 1 by $\mathrm{E}[D_t^2]<\infty$ and that $|A_t|^2\le CD_t^2$. Then by the Cauchy–Schwarz inequality, for any $h\in\mathcal{G}$, we have, by $D_t\ge 1$, $\mathrm{E}[D_t^2|W_t]\ge 1$:

$$\begin{aligned}&\big\|\mathrm{E}\big[A_t h(c_{t+1})|W_t\big]-h(c_t)e\big\|_{\mathcal{B}}^2\\&\quad\le C\mathrm{E}\big[\mathrm{E}\big[A_t^T h(c_{t+1})|W_t\big]\mathrm{E}\big[A_t h(c_{t+1})|W_t\big]+h(c_t)^2\big]\\&\quad\le C\mathrm{E}\big[\mathrm{E}\big[D_t^2|W_t\big]\mathrm{E}\big[h(c_{t+1})^2|W_t\big]\big]+C\mathrm{E}\big[\mathrm{E}\big[D_{t-1}^2|W_{t-1}\big]h(c_t)^2\big]\\&\quad\le C\|h\|_{\mathcal{G}}^2.\end{aligned}$$

Thus $m'_g : \mathcal{G} \longrightarrow \mathcal{B}$ is bounded. Also, noting that $|m'_\beta| \leq \mathrm{E}[D_t g_0(c_{t+1})|W_t]$, the Cauchy–Schwarz inequality gives

$$\mathrm{E}\big[\big|m'_\beta(W)\big|^2\big] \leq \mathrm{E}\big[\mathrm{E}\big[D_t^2|W_t\big]\mathrm{E}\big[g_0(c_{t+1})^2|W_t\big]\big] \leq \|g_0\|_{\mathcal{G}}^2 < \infty,$$

and hence $m'_\beta : \mathbb{R}^2 \longrightarrow \mathcal{B}$ is bounded. Therefore, the first part of Assumption 4 is satisfied with $\mathcal{N}'_g = \mathcal{G}$.

Turning now to Assumption 5, let $H_t(\beta, g) = \delta R_{t+1} c_{t+1}^{-\gamma} g(c_{t+1})$ and $B = [\delta_0 - \Delta, \delta_0 + \Delta] \times [\gamma_0 - \Delta, \gamma_0 + \Delta]$. Note that $H_t(\beta, g)$ is twice continuously differentiable in β and, by construction of D_t, that

$$\sup_{\beta \in B} \left| \frac{\partial H_t(\beta, g)}{\partial \beta} \right| \leq D_t g(c_{t+1}),$$

$$\sup_{\beta \in B} \left| \frac{\partial^2 H_t(\beta, g)}{\partial \beta_j \, \partial \beta} \right| \leq D_t g(c_{t+1}) \quad (j = 1, 2).$$

Therefore, by standard results, $\mathrm{E}[\rho(Y_t, \beta, g)|W_t] = \mathrm{E}[H_t(\beta, g)|W_t] - g(c_t)$ is twice continuously differentiable in β on B, $\partial \mathrm{E}[\rho(Y_t, \beta, g)|W_t]/\partial \beta = \mathrm{E}[\partial H_t(\beta, g)/\partial \beta|W_t]$. We also have

$$\begin{aligned}
&\big|\mathrm{E}\big[\partial H_t(\beta, g)/\partial \beta - \partial H_t(\beta, g_0)/\partial \beta|W_t\big]\big|^2 \\
&\quad \leq \mathrm{E}\big[D_t^2|W_t\big]\mathrm{E}\big[\big|g(c_{t+1}) - g_0(c_{t+1})\big|^2|W_t\big], \\
&\big|\mathrm{E}\big[\partial H_t(\beta, g_0)/\partial \beta - \partial H_t(\beta_0, g_0)/\partial \beta|W_t\big]\big|^2 \\
&\quad \leq \mathrm{E}\big[D_t^2|W_t\big]\mathrm{E}\big[g_0(c_{t+1})^2|W_t\big]|\beta - \beta_0|^2.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\left| \frac{\partial \mathrm{E}[\rho(Y, \beta, g)|W]}{\partial \beta} - \frac{\partial \mathrm{E}[\rho(Y, \beta_0, g_0)|W]}{\partial \beta} \right|^2 \\
&\quad = \big|\mathrm{E}\big[\partial H_t(\beta, g)/\partial \beta - \partial H_t(\beta_0, g_0)/\partial \beta|W_t\big]\big|^2 \\
&\quad \leq 2\mathrm{E}\big[D_t^2|W_t\big]\big\{\mathrm{E}\big[\big|g(c_{t+1}) - g_0(c_{t+1})\big|^2|W_t\big] \\
&\qquad + \mathrm{E}\big[g_0(c_{t+1})^2|W_t\big]|\beta - \beta_0|^2\big\}.
\end{aligned}$$

Note that by iterated expectations,

$$\begin{aligned}
&\mathrm{E}\big[\mathrm{E}\big[D_t^2|W_t\big]\mathrm{E}\big[\big|g(c_{t+1}) - g_0(c_{t+1})\big|^2|W_t\big]\big] = \|g - g_0\|_{\mathcal{G}}^2, \\
&\mathrm{E}\big[\mathrm{E}\big[D_t^2|W_t\big]\mathrm{E}\big[g_0(c_{t+1})^2|W_t\big]\big] = \|g_0\|_{\mathcal{G}}^2.
\end{aligned}$$

Consider any $\varepsilon > 0$. Let

$$\mathcal{N}_g^\beta = \{g : \|g - g_0\|_{\mathcal{G}} \le \varepsilon/2\} \quad \text{and}$$
$$\tilde{B} = B \cap \{\beta : |\beta - \beta_0| < \varepsilon/(2\|g_0\|_{\mathcal{G}})\}.$$

Then, for $g \in \mathcal{N}_g^\beta$, we have

$$\mathrm{E}\Big[\sup_{\beta\in\tilde{B}}\big|\partial m(\alpha)(W)/\partial\beta - m'_\beta(W)\big|^2\Big]$$
$$\le 2\|g - g_0\|_{\mathcal{G}}^2 + 2\|g_0\|_{\mathcal{G}}^2 \sup_{\beta\in\tilde{B}}|\beta - \beta_0|^2 < \varepsilon^2.$$

Therefore, Assumption 5 holds with B there equal to $\tilde{B}$ here. The conclusion then follows from Theorem 7. *Q.E.D.*

S4.2. *Proof of Theorem 10*

Let $\bar{a}(c_{t+1}, \tilde{W}_t) = \mathrm{E}[A_{tj}|c_{t+1}, \tilde{W}_t]$ and $\bar{d}(c_{t+1}) = \mathrm{E}[\mathrm{E}[D_t^2|\tilde{W}_t]|c_{t+1}]$. Let $\tilde{\mathcal{B}} = \{b(\tilde{W}_t) : \mathrm{E}[b(\tilde{W}_t)^2] < \infty\}$ and the operator $L : \mathcal{G} \longrightarrow \tilde{\mathcal{B}}$ be given by

$$Lg = \mathrm{E}\big[A_{tj}g(c_{t+1})|\tilde{W}_t\big] = \int \bar{a}(c, \tilde{W}_t)g(c)\frac{f_{c,\tilde{W}}(c, \tilde{W}_t)}{f_{\tilde{W}}(\tilde{W}_t)}\,dc$$
$$= \int g(c)K(c, \tilde{W}_t)f_c(c)\bar{d}(c)\,dc,$$
$$K(c, \tilde{W}_t) = \frac{\bar{a}(c, \tilde{W}_t)f_{c,\tilde{W}}(c, \tilde{W}_t)}{f_{\tilde{W}}(\tilde{W}_t)f_c(c)\bar{d}(c)}.$$

Note that $\bar{d}(c) \ge 1$ by $D_t^2 \ge 1$. Therefore,

$$\int K(c, w)^2\bar{d}(c)f_c(c)f_{\tilde{W}}(w)\,dc\,dw$$
$$= \int \frac{\bar{a}(c, w)^2 f_{c,\tilde{W}}(c, w)}{f_{\tilde{W}}(w)f_c(c)\bar{d}(c)}f_{c,\tilde{W}}(c, w)\,dc\,dw$$
$$\le \int \frac{\bar{a}(c, w)^2 f_{c,\tilde{W}}(c, w)}{f_{\tilde{W}}(w)f_c(c)}f_{c,\tilde{W}}(c, w)\,dc\,dw$$
$$= \mathrm{E}\left[\mathrm{E}[A_{tj}|c_{t+1}, \tilde{W}_t]^2\frac{f_{c,\tilde{W}}(c_{t+1}, \tilde{W}_t)}{f_c(c_{t+1})f_{\tilde{W}}(\tilde{W}_t)}\right]$$
$$\le \mathrm{E}\big[A_{tj}^2 f_c(c_{t+1})^{-1}f_{\tilde{W}}(\tilde{W}_t)^{-1}f_{c,\tilde{W}}(c_{t+1}, \tilde{W}_t)\big] < \infty.$$

It therefore follows by standard results that L is Hilbert–Schmidt and thus compact. Furthermore, it follows exactly as in the proof of Theorem 3.2 of Kress (1999), that

$$\tilde{\mathcal{M}} = \{\mathrm{E}[A_{tj}g(c_{t+1})|\tilde{W}_t] - g(c_t) : g \in \mathcal{G}\}$$

is closed.

Next, let $b = (b_1, b_2)^T$ be a constant vector and $\Delta(c) = b_1/\delta_0 - b_2 \ln(c)$. Suppose $b^T \Pi b = 0$. Then, by the definition of Π, there is $g_k \in \mathcal{G}$ such that

$$\mathrm{E}[A_t g_k(c_{t+1})|\tilde{W}_t] - g_k(c_t)e \longrightarrow \mathrm{E}[A_t g_0(c_{t+1})\Delta(c_{t+1})|\tilde{W}_t]$$

in mean squared as $k \longrightarrow \infty$. It follows that, for any j,

$$\mathrm{E}[A_{tj} g_k(c_{t+1})|\tilde{W}_t] - g_k(c_t) \longrightarrow \mathrm{E}[A_{tj} g_0(c_{t+1})\Delta(c_{t+1})|\tilde{W}_t]$$

in mean squared. By $\tilde{\mathcal{M}}$ a closed set, there exists $g^*(c)$ such that

$$\text{(S4.1)} \qquad \mathrm{E}[A_{tj} g_0(c_{t+1})\Delta(c_{t+1})|\tilde{W}_t] = \mathrm{E}[A_{tj} g^*(c_{t+1})|\tilde{W}_t] - g^*(c_t).$$

If $g^*(c_{t+1}) = 0$, then $\mathrm{E}[A_{tj} g_0(c_{t+1})\Delta(c_{t+1})|\tilde{W}_t] = 0$, and by completeness of $\mathrm{E}[A_{tj} h(c_{t+1}, c_t)|\tilde{W}_t]$, it follows that $g_0(c_{t+1})\Delta(c_{t+1}) = 0$. Then, by $\Pr(g_0(c_{t+1}) \neq 0) = 1$, we have $\Delta(c_{t+1}) = 0$.

Next, suppose $\Pr(g^*(c_t) \neq 0) > 0$. Then $\Pr(\min\{|g^*(c_t)|, g_0(c_t)\} > 0) > 0$, so for small enough $\varepsilon > 0$ and $\mathcal{C} = \{c_t : \min\{|g^*(c_t)|, g_0(c_t)\} \geq \varepsilon\}$, we have $\Pr(\mathcal{C}) > 0$. Let $1_t^\varepsilon = 1(c_t \in \mathcal{C})$. Then, multiplying through equation (S4.1) by $1_t^\varepsilon / g^*(c_t)$ and subtracting the conditional expectation on the right-hand side gives

$$\mathrm{E}\left[A_{tj} 1_t^\varepsilon \frac{g_0(c_{t+1})\Delta(c_{t+1}) - g^*(c_{t+1})}{-g^*(c_t)} \Big| \tilde{W}_t\right] = 1_t^\varepsilon.$$

By the moment condition from Section 5 of Chen et al. (2014), we also have

$$\mathrm{E}\left[A_{tj} 1_t^\varepsilon \left\{\frac{g_0(c_{t+1})}{g_0(c_t)}\right\} \Big| \tilde{W}_t\right] = 1_t^\varepsilon.$$

By the completeness condition in part (a), it then follows that

$$1_t^\varepsilon \frac{g_0(c_{t+1})\Delta(c_{t+1}) - g^*(c_{t+1})}{-g^*(c_t)} = 1_t^\varepsilon \frac{g_0(c_{t+1})}{g_0(c_t)}.$$

Multiplying, dividing, and subtracting gives

$$1_t^\varepsilon \left[\frac{g_0(c_{t+1})\Delta(c_{t+1}) - g^*(c_{t+1})}{-g_0(c_{t+1})} - \frac{g^*(c_t)}{g_0(c_t)}\right] = 0.$$

Then, by the additive separability condition in part (a) of the conditions, it follows that $g_0(c_{t+1})\Delta(c_{t+1}) - g^*(c_{t+1}) = Cg_0(c_{t+1})$ for some nonzero constant C. Then by equation (S4.1) and the second kind equation for g_0, we have

$$\begin{aligned} g^*(c_t) &= -\mathrm{E}\big[A_{tj}\big\{g_0(c_{t+1})\Delta(c_{t+1}) - g^*(c_{t+1})\big\}|\tilde{W}_t\big] \\ &= -C\mathrm{E}\big[A_{tj}g_0(c_{t+1})|\tilde{W}_t\big] = -Cg_0(c_t). \end{aligned}$$

Then by equation (S4.1),

$$\begin{aligned} \mathrm{E}\big[A_{tj}g_0(c_{t+1})\Delta(c_{t+1})|\tilde{W}_t\big] &= \mathrm{E}\big[A_{tj}g^*(c_{t+1})|\tilde{W}_t\big] - g^*(c_t) \\ &= -C\big\{\mathrm{E}\big[A_{tj}g_0(c_{t+1})|\tilde{W}_t\big] - g_0(c_t)\big\} = 0. \end{aligned}$$

By the completeness condition in part (a) of the conditions, it follows that $g_0(c_{t+1})\Delta(c_{t+1}) = 0$, so $\Delta(c_{t+1}) = 0$ follows by $\Pr(g_0(c_t) = 0) = 0$. Therefore, we find that $b^T\Pi b = 0$ implies $\Delta(c_{t+1}) = 0$. But we know that, for $b \neq 0$, it is the case that $\Delta(c_{t+1}) \neq 0$. Therefore, $b \neq 0$ implies $b^T\Pi b > 0$, that is, Π is nonsingular.

Next, under condition (b) of Theorem 10, if $\mathrm{E}[A_{tj}g(c_{t+1})|\tilde{W}_t] = g(c_t)$ for $g \in \mathcal{G}_{\bar{c}}$, it follows that, for $\bar{c}$ as given there,

$$\mathrm{E}\left[A_{tj}\frac{g(c_{t+1})}{g(\bar{c})}\Big|w(Z_t), c_t = \bar{c}\right] = 1 = \mathrm{E}\left[A_{tj}\frac{g_0(c_{t+1})}{g_0(\bar{c})}\Big|w(Z_t), c_t = \bar{c}\right].$$

Then, by the completeness condition in part (b) of the hypotheses, it follows that $g(c_{t+1})/g(\bar{c}) = g_0(c_{t+1})/g_0(\bar{c})$, that is,

$$g(c_{t+1}) = g_0(c_{t+1})g(\bar{c})/g_0(\bar{c}),$$

so g is equal to g_0 up to scale. This also implies that g_0 is the unique solution to $\mathrm{E}[A_t g(c_{t+1})|W_t] = g(c_t)$ up to scale. *Q.E.D.*

S4.3. *Proof of Theorem 11*

Note that $K(c, s) = r(c, s)s^{-\gamma_0}f(s, c)/[f(s)f(c)] > 0$ almost everywhere by $r(c, s) > 0$ and $f(s, c) > 0$ almost everywhere. Therefore, the conclusion follows from Lemma S6 with $f(s)\,ds = d\mu(s)$. *Q.E.D.*

S4.4. *Completeness and Global Identification in the CCAPM*

In this subsection, we give a result showing that a certain completeness condition is sufficient for global identification of the CCAPM model we consider, when $g(c)$ is bounded and bounded away from zero.

THEOREM S5: *Consider model of Section* 5 *of Chen et al*. (2014). *If* $(R_{t,j}, c_t)$ *is strictly stationary*, c_t *is continuously distributed with support* $[0, \infty)$, $g_0(c) \geq 0$ *is bounded and bounded away from zero*, $\mathrm{E}[|R_{t+1,j} c_t^{-\gamma_0}|] < \infty$, *and there is* $\bar{c}$ *such that* $\mathrm{E}[R_{t+1,j} h(c_{t+1})|w(Z_t), \bar{c}] = 0$ *and* $\mathrm{E}[|R_{t+1,j} h(c_{t+1})|] < \infty$ *implies* $h(c_{t+1}) = 0$, *then* $(\delta_0, \gamma_0, g_0)$ *is identified* (g_0 *up to scale*) *among all* (δ, γ, g) *with* $g(c) \geq 0$, $g(c)$ *bounded and bounded away from zero, and* $\mathrm{E}[|R_{t+1,j} c_t^{-\gamma}|] < \infty$.

PROOF: Consider any two solutions (β_0, g_0) and (β_1, g_1) to the moment conditions satisfying the conditions of Theorem S5. Then, by iterated expectations,

$$\mathrm{E}\left[R_{t+1,j} \delta_0 c_{t+1}^{-\gamma_0} \frac{g_0(c_{t+1})}{g_0(\bar{c})} \Big| w(Z_t), \bar{c}\right]$$

$$= 1 = \mathrm{E}\left[R_{t+1,j} \delta_1 c_{t+1}^{-\gamma_1} \frac{g_1(c_{t+1})}{g_1(\bar{c})} \Big| w(Z_t), \bar{c}\right].$$

By completeness with $h(c_{t+1}) = \delta_0 c_{t+1}^{-\gamma_0} g_0(c_{t+1})/g_0(\bar{c}) - \delta_1 c_{t+1}^{-\gamma_1} g_1(c_{t+1})/g_1(\bar{c})$, it follows by multiplying and dividing that

$$c_{t+1}^{\gamma_1 - \gamma_0} = \frac{g_1(c_{t+1})}{g_0(c_{t+1})} \left[\frac{\delta_1 g_0(\bar{c})}{\delta_0 g_1(\bar{c})}\right].$$

Since the object on the right is bounded and bounded away from zero and the support of c_{t+1} is $I = [0, \infty)$, it follows that $\gamma_0 = \gamma_1$. Then we have

$$g_0(c_{t+1}) = g_1(c_{t+1}) \left[\frac{\delta_1 g_0(\bar{c})}{\delta_0 g_1(\bar{c})}\right] \quad \text{a.e. in } I^2,$$

so that there is a constant $D > 0$ such that $g_0(c_{t+1}) = D g_1(c_{t+1})$ a.e. in I. We can also assume that $g_0(\bar{c}) = D g_1(\bar{c})$ since c_t is continuously distributed. Substituting then gives $D = (\delta_1/\delta_0) D$, implying $\delta_1 = \delta_0$. *Q.E.D.*

Previously, Chen and Ludvigson (2009) showed global identification of $(\delta_0, \gamma_0, g_0)$ under different conditions. In their results, $\mathrm{E}[R_{t+1,j} h(c_{t+1}, c_t)| w(Z_t), c_t]$ was assumed to be complete, which is similar to condition (a) in Theorem 10 and is stronger than completeness at $c_t = \bar{c}$, but $g(c)$ is not assumed to be bounded or bounded away from zero on $[0, \infty)$.

S4.5. *A Functional Perron–Frobenius Theorem*

The following result and its proof rely in part on the fundamental results of Krein and Rutman (1950), specifically their Theorem 6.1 and example β'. Krein and Rutman (1950) is one of many extensions of the Perron–Frobenius theory of positive matrices to the case of operators leaving invariant a cone in a Banach space.

Let I be a Borel subset of $\mathbb{R}^m$ and μ be a σ-finite measure with support I. Consider the space $L^2(\mu)$, equipped with the standard norm $\|\cdot\|$. We consider the following conditions on the kernel K:

1. $K(s,t)$ is a nonnegative, measurable kernel such that

$$\int\int K^2(s,t)\,d\mu(t)\,d\mu(s)<\infty.$$

2. $K(s,t)=0$ on a set of points (t,s) of measure zero under $\mu\times\mu$.

Consider an integral operator L from $L^2(\mu)$ to $L^2(\mu)$ defined by

$$L\varphi:=\int K(\cdot,t)\varphi(t)\,d\mu(t),$$

and its adjoint operator

$$L^*\psi:=\int K(t,\cdot)\psi(t)\,d\mu(t).$$

It is known that these operators are compact under condition 1. The lemma given below shows that, under these assumptions, we have existence and global uniqueness of the positive eigenpair (ρ,φ) such that $L\varphi=\rho\varphi$, in the sense that is stated below. This lemma extends example β' outlined in Krein and Rutman (1950) that looked at the complex Hilbert space $L^2[a,b]$, $0<a<b<\infty$, an extension which we were not able to track easily in the literature, so we simply derived it; we also provided an additional step (3), not given in the outline, to fully verify uniqueness. Note that we removed the complex analysis-based arguments, since they are not needed here.

LEMMA S6: *Under conditions* 1 *and* 2, *there exists a unique eigenpair* (ρ,φ), *consisting of an eigenvalue* ρ *and eigenfunction* φ *such that* $L\varphi=\rho\varphi$ *and* $\rho>0$, $\|\varphi\|=1$, $\varphi\geq 0$; *moreover*, $\varphi>0$ *μ-a.e.*

PROOF: The proof is divided in five steps.

(1) Let C^o be the cone of nonnegative functions in $A=L^2(\mu)$. In the proof, we shall use the following result on the existence of nonnegative eigenpair from Krein and Rutman (1950, Theorem 6.1).

> Consider a cone C^o in a Banach space A such that the closure of the linear hull of C^o is A. Consider a linear, compact operator $L:A\mapsto A$ such that $LC^o\subset C^o$, and that has one point of spectrum different from zero. Then it has a positive eigenvalue ρ, not less in modulus than every other eigenvalue, and to this eigenvalue there corresponds at least one eigenvector $\varphi\in C^o$ of the operator L $(L\varphi=\rho\varphi)$ and at least one eigenvector $\psi\neq 0$ of the dual operator L^* $(L^*\psi=\rho\psi)$.

The theorem requires that the closure of the linear hull of the cone is A. This is true in our case for $A = L^2(\mu)$ and the cone C^o of the nonnegative functions in A, since $C^o - C^o$ is dense in A. Moreover, since

$$\sigma_2 = \int K(s,t)K(t,s)\,d\mu(s)\,d\mu(t) > 0,$$

which is equal to sum of squared eigenvalues of L, the spectrum of L must have at least one point different from zero. Therefore, application of the theorem quoted above implies that there exist $\rho > 0$ and φ and ψ s.t. μ-a.e.

$$\varphi(s) = \rho^{-1}\int K(s,t)\varphi(t)\,d\mu(t), \quad \varphi \geq 0, \|\varphi\| = 1, \rho > 0; \tag{S4.2}$$

$$\psi(s) = \rho^{-1}\int K(t,s)\psi(t)\,d\mu(t), \quad \|\psi\| = 1. \tag{S4.3}$$

(2) We would like to prove that any eigenvalue $\rho > 0$ associated to a nonnegative eigenfunction $\varphi \geq 0$ must be a simple eigenvalue, that is, φ is the only eigenfunction in $L^2(\mu)$ associated with ρ. For this purpose, we shall use the following standard fact on linear compact operators, for example stated in Krein and Rutman (1950) and specialized to our context: An eigenvalue ρ of L is simple if and only if the equations $L\varphi = \rho\varphi$ and $L^*\psi = \rho\psi$ have no solutions orthogonal to each other, that is, satisfying $\varphi \neq 0$, $\psi \neq 0$, $\int \psi(s)\varphi(s)\,d\mu(s) = 0$. So for this purpose, we will show in steps (4) and (5) below that ψ is of constant sign μ-a.e. and φ and ψ only vanish on a set of measure 0 under μ. Since $\varphi \geq 0$, this implies

$$\int \psi(s)\varphi(s)\,d\mu(s) \neq 0,$$

and we conclude from the quoted fact that ρ is a simple eigenvalue.

(3) To assert the uniqueness of the nonnegative eigenpair (ρ, φ) (meaning that $L\varphi = \rho\varphi$, $\rho > 0$, $\varphi \geq 0$, $\|\varphi\| = 1$), suppose to the contrary that there is another nonnegative eigenpair (r, ζ). Then r is also an eigenvalue of L^* by the Fredholm theorem (Kress (1999, Theorem 4.14)), which implies, by definition of the eigenvalue, that there exists a dual eigenfunction $\eta \neq 0$ such that $L^*\eta = r\eta$ and $\|\eta\| = 1$.

By step (4) below, we must have $\zeta > 0$, $\varphi > 0$ μ-a.e. Hence, by step (5), the dual eigenfunctions η and ψ are non-vanishing and of constant sign μ-a.e., which implies $\int \eta(s)\varphi(s)\,d\mu(s) \neq 0$. Therefore, $r = \rho$ follows from the equality

$$\begin{aligned} r\int \eta(s)\varphi(s)\,d\mu(s) &= \int\int K(t,s)\eta(t)\,d\mu(t)\varphi(s)\,d\mu(s) \\ &= \rho\int \eta(t)\varphi(t)\,d\mu(t). \end{aligned}$$

(4) Let us prove that any eigenfunction $\varphi \geq 0$ of L associated with an eigenvalue $\rho > 0$ must be μ-a.e. positive. Let S denote the set of zeros of φ. Evidently, $\mu(S) < \mu(I)$. If $s \in S$, then

$$\int K(s, t)\varphi(t)\, d\mu(t) = 0.$$

Therefore, $K(s, t)$ vanishes almost everywhere on $(s, t) \in S \times (I \setminus S)$. However, the set of zeroes of $K(s, t)$ is of measure zero under $\mu \times \mu$, so $\mu(S) \times \mu(I \setminus S) = 0$, implying $\mu(S) = 0$.

(5) Here we show that any eigen-triple (ρ, φ, ψ) solving (S4.3) and (S4.2) obeys

$$\text{(S4.4)} \qquad \operatorname{sign}\big(\psi(s)\big) = 1 \quad \mu\text{-a.e.} \quad \text{or} \quad \operatorname{sign}\big(\psi(s)\big) = -1 \quad \mu\text{-a.e.}$$

From equation (S4.3), it follows that μ-a.e.

$$\big|\psi(s)\big| \leq \rho^{-1} \int K(t, s)\big|\psi(t)\big|\, d\mu(t).$$

Multiplying both sides by $\varphi(s)$, integrating, and applying (S4.2) yields

$$\begin{aligned}\int \big|\psi(s)\big|\varphi(s)\, d\mu(s) &\leq \rho^{-1} \int \int K(t, s)\varphi(s)\big|\psi(t)\big|\, d\mu(t)\, d\mu(s) \\ &= \int \big|\psi(t)\big|\varphi(t)\, d\mu(t).\end{aligned}$$

It follows that μ-a.e.

$$\big|\psi(s)\big| = \rho^{-1} \int K(t, s)\big|\psi(t)\big|\, d\mu(t),$$

that is, $|\psi|$ is an eigenfunction of L^*.

Next, equation $|\psi(s)| = \psi(s)\operatorname{sign}(\psi(s))$ implies that μ-a.e.

$$\rho^{-1} \int K(t, s)\big|\psi(t)\big|\, d\mu(t) = \rho^{-1} \int K(t, s)\psi(t)\, d\mu(t) \operatorname{sign}\big(\psi(s)\big).$$

It follows that, for a.e. (t, s) under $\mu \times \mu$,

$$\big|\psi(t)\big| = \psi(t)\operatorname{sign}\big(\psi(s)\big).$$

By the positivity condition on K, $|\psi| > 0$ μ-a.e. by the same reasoning as given in step (4). Thus, (S4.4) follows. *Q.E.D.*

S5. TANGENTIAL CONE CONDITIONS

In this section, we discuss some inequalities that are related to identification of α_0. These inequalities are related to the discussion of Assumptions 1 and 2 in Section 2.3 of Chen et al. (2014). Throughout this section, we maintain that $m(\alpha_0) = 0$. Define

$$\mathcal{N} = \{\alpha : m(\alpha) \neq 0\}, \quad \mathcal{N}' = \{\alpha : m'(\alpha - \alpha_0) \neq 0\},$$
$$\mathcal{N}'_\eta = \{\alpha : \|m(\alpha) - m'(\alpha - \alpha_0)\|_{\mathcal{B}} \leq \eta \|m'(\alpha - \alpha_0)\|_{\mathcal{B}}\}, \quad \eta > 0,$$
$$\mathcal{N}_\eta = \{\alpha : \|m(\alpha) - m'(\alpha - \alpha_0)\|_{\mathcal{B}} \leq \eta \|m(\alpha)\|_{\mathcal{B}}\}, \quad \eta > 0.$$

Here, $\mathcal{N}$ can be interpreted as the identified set and $\mathcal{N}'$ as the set where the rank condition holds. The set $\mathcal{N}'_\eta$ is a set on which an inequality version of equation (2.2) in Chen et al. (2014) holds. The inequality used to define $\mathcal{N}_\eta$ is similar to the tangential cone condition from the literature on computation in nonlinear ill-posed inverse problems, for example, Hanke, Neubauer, and Scherzer (1995) and Dunker, Florens, Hohage, Johannes, and Mammen (2014).

The following result gives some relations among these sets:

LEMMA S7: *For any* $\eta > 0$,

$$\mathcal{N}_\eta \cap \mathcal{N}' \subset \mathcal{N}, \quad \mathcal{N}'_\eta \cap \mathcal{N} \subset \mathcal{N}'.$$

If $0 < \eta < 1$, *then*

$$\mathcal{N}_\eta \cap \mathcal{N} \subset \mathcal{N}', \quad \mathcal{N}'_\eta \cap \mathcal{N}' \subset \mathcal{N}.$$

PROOF: Note that $\alpha \in \mathcal{N}_\eta$ and the triangle inequality gives

$$-\|m(\alpha)\|_{\mathcal{B}} + \|m'(\alpha - \alpha_0)\|_{\mathcal{B}} \leq \eta \|m(\alpha)\|_{\mathcal{B}},$$

so that $\|m(\alpha)\|_{\mathcal{B}} \geq (1+\eta)^{-1}\|m'(\alpha - \alpha_0)\|_{\mathcal{B}}$. Therefore, if $\alpha \in \mathcal{N}_\eta \cap \mathcal{N}'$, we have $\|m(\alpha)\|_{\mathcal{B}} > 0$, that is, $\alpha \in \mathcal{N}$, giving the first conclusion. Also, if $\alpha \in \mathcal{N}'_\eta$, we have

$$-\|m'(\alpha - \alpha_0)\|_{\mathcal{B}} + \|m(\alpha)\|_{\mathcal{B}} \leq \eta \|m'(\alpha - \alpha_0)\|_{\mathcal{B}},$$

so that $\|m'(\alpha - \alpha_0)\|_{\mathcal{B}} \geq (1+\eta)^{-1}\|m(\alpha)\|_{\mathcal{B}}$. Therefore, if $\alpha \in \mathcal{N}'_\eta \cap \mathcal{N}$, we have $\|m'(\alpha - \alpha_0)\|_{\mathcal{B}} > 0$, giving the second conclusion.

Next, for $0 < \eta < 1$ and $\alpha \in \mathcal{N}_\eta$, we have

$$\|m(\alpha)\|_{\mathcal{B}} - \|m'(\alpha - \alpha_0)\|_{\mathcal{B}} \leq \eta \|m(\alpha)\|_{\mathcal{B}},$$

so that $\|m'(\alpha - \alpha_0)\| \geq (1 - \eta)\|m(\alpha)\|_{\mathcal{B}}$. Therefore, if $\alpha \in \mathcal{N}_\eta \cap \mathcal{N}$, we have $\|m'(\alpha - \alpha_0)\|_{\mathcal{B}} > 0$, giving the third conclusion. Similarly, for $0 < \eta < 1$ and $\alpha \in \mathcal{N}'_\eta$, we have

$$\left\|m'(\alpha - \alpha_0)\right\|_{\mathcal{B}} - \left\|m(\alpha)\right\|_{\mathcal{B}} \leq \eta\left\|m'(\alpha - \alpha_0)\right\|_{\mathcal{B}},$$

so that $\|m(\alpha)\|_{\mathcal{B}} \geq (1 - \eta)\|m'(\alpha - \alpha_0)\|_{\mathcal{B}}$. Therefore, if $\alpha \in \mathcal{N}'_\eta \cap \mathcal{N}'$, we have $\|m(\alpha)\|_{\mathcal{B}} > 0$, giving the fourth conclusion. *Q.E.D.*

The first conclusion shows that when the tangential cone condition is satisfied, the set on which the rank condition holds is a subset of the identified set. The second condition is less interesting, but does show that the rank condition is necessary for identification when $\alpha \in \mathcal{N}'_\eta$. The third conclusion shows that the rank condition is also necessary for identification under the tangential cone condition for $0 < \eta < 1$. The last conclusion shows that when $\alpha \in \mathcal{N}'_\eta$ with $0 < \eta < 1$, the rank condition is sufficient for identification.

When the side condition that $\alpha \in \mathcal{N}_\eta$ or $\alpha \in \mathcal{N}'_\eta$ is imposed for $0 < \eta < 1$, the rank condition is necessary and sufficient for identification.

COROLLARY S8: *If* $0 < \eta < 1$, *then*

$$\mathcal{N}_\eta \cap \mathcal{N}' = \mathcal{N}_\eta \cap \mathcal{N}, \quad \mathcal{N}'_\eta \cap \mathcal{N}' = \mathcal{N}'_\eta \cap \mathcal{N}.$$

PROOF: By intersecting both sides of the first conclusion of Lemma S7 with $\mathcal{N}_\eta$, we find that $\mathcal{N}_\eta \cap \mathcal{N}' \subset \mathcal{N}_\eta \cap \mathcal{N}$. For $\eta < 1$, it follows similarly from the third conclusion of Lemma S7 that $\mathcal{N}_\eta \cap \mathcal{N} \subset \mathcal{N}_\eta \cap \mathcal{N}'$, implying $\mathcal{N}_\eta \cap \mathcal{N}' = \mathcal{N}_\eta \cap \mathcal{N}$, the first conclusion. The second conclusion follows similarly. *Q.E.D.*

The equalities in the conclusion of this result show that the rank condition (i.e., $\alpha \in \mathcal{N}'$) is necessary and sufficient for identification (i.e., $\alpha \in \mathcal{N}$) under either of the side conditions that

$$\alpha \in \mathcal{N}'_\eta \quad \text{or} \quad \alpha \in \mathcal{N}_\eta, \quad 0 < \eta < 1.$$

In parametric models, Rothenberg (1971) showed that when the Jacobian has constant rank in a neighborhood of the true parameter, the rank condition is necessary and sufficient for local identification. These conditions fill an analogous role here, in the sense that when α is restricted to either set, the rank condition is necessary and sufficient for identification.

The sets $\mathcal{N}_\eta$ and $\mathcal{N}'_\eta$ are related to each other in the way shown in the following result.

LEMMA S9: *If* $0 < \eta < 1$, *then* $\mathcal{N}_\eta \subset \mathcal{N}'_{\eta/(1-\eta)}$ *and* $\mathcal{N}'_\eta \subset \mathcal{N}_{\eta/(1-\eta)}$.

PROOF: By the triangle inequality,

$$\|m'(\alpha-\alpha_0)\|_{\mathcal{B}} \le \|m(\alpha)-m'(\alpha-\alpha_0)\|_{\mathcal{B}} + \|m(\alpha)\|_{\mathcal{B}},$$
$$\|m(\alpha)\|_{\mathcal{B}} \le \|m(\alpha)-m'(\alpha-\alpha_0)\|_{\mathcal{B}} + \|m'(\alpha-\alpha_0)\|_{\mathcal{B}}.$$

Therefore, for $\alpha \in \mathcal{N}_\eta$,

$$\|m(\alpha)-m'(\alpha-\alpha_0)\|_{\mathcal{B}} \le \eta\|m(\alpha)-m'(\alpha-\alpha_0)\|_{\mathcal{B}} + \eta\|m'(\alpha-\alpha_0)\|_{\mathcal{B}}.$$

Subtracting $\eta\|m(\alpha)-m'(\alpha-\alpha_0)\|_{\mathcal{B}}$ from both sides and dividing by $1-\eta$ gives $\alpha \in \mathcal{N}'_{\eta/(1-\eta)}$. The second conclusion follows similarly. *Q.E.D.*

REFERENCES

ANDERSON, R., AND W. R. ZAME (2000): "Genericity With Infinitely Many Parameters," *Advances in Theoretical Economics*, 1, 1–62. [1]

ANDREWS, D. W. K. (2011): "Examples of L2-Complete and Boundedly-Complete Distributions," Cowles Foundation Discussion Paper 1801, Yale University. [1,2]

CHEN, X., AND S. LUDVIGSON (2009): "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Journal of Applied Econometrics*, 24, 1057–1093. [14]

CHEN, X., V. CHERNOZHUKOV, S. LEE, AND W. K. NEWEY (2014): "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, 82, 785–809. [1,9,12,14,18]

DUNKER, F., J. P. FLORENS, T. HOHAGE, J. JOHANNES, AND E. MAMMEN (2014): "Iterative Estimation of Solutions to Noisy Nonlinear Operator Equations in Nonparametric Instrumental Regression," *Journal of Econometrics*, 178, 444–455. [18]

HANKE, M., A. NEUBAUER, AND O. SCHERZER (1995): "A Convergence Analysis of Landweber Iteration for Nonlinear Ill-Posed Problems," *Numer. Math.*, 72, 21–37. [18]

KREIN, M. G., AND M. A. RUTMAN (1950): *Linear Operators Leaving Invariant a Cone in a Banach Space*. Amer. Math. Soc. Translation 1950, Vol. 26. Providence: Amer. Math. Soc. [14-16]

KRESS, R. (1999): *Linear Integral Equations* (Second Ed.). New York: Springer-Verlag. [12,16]

LEDOUX, M., AND M. TALAGRAND (2011): *Probability in Banach Spaces. Isoperimetry and Processes*. Berlin: Springer-Verlag. [1]

MARCUS, M. B., AND G. PISIER (1981): *Random Fourier Series With Applications to Harmonic Analysis*. Princeton: Princeton University Press. [1]

NEWEY, W. K., AND J. L. POWELL (2003): "Instrumental Variables Estimation for Nonparametric Models," *Econometrica*, 71, 1565–1578. [8]

ROTHENBERG, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591. [19]

Dept. of Economics, Yale University, Box 208281, New Haven, CT 06520, U.S.A.; xiaohong.chen@yale.edu,

Dept. of Economics, MIT, Cambridge, MA 02142, U.S.A.; vchern@mit.edu,

Dept. of Economics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 151-742, Respublic of Korea; sokbae@snu.ac.kr,

and

Dept. of Economics, MIT, Cambridge, MA 02142, U.S.A.; wnewey@mit.edu.

Manuscript received April, 2011; final revision received September, 2013.