LOCAL IDENTIFICATION OF NONPARAMETRIC AND SEMIPARAMETRIC MODELS

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

Abstract. In parametric models a sufficient condition for local identification is that the vector of moment conditions is differentiable at the true parameter with full rank derivative matrix. We show that there are corresponding sufficient conditions for nonparametric models. A nonparametric rank condition and differentiability of the moment conditions with respect to a certain norm imply local identification. It turns out these conditions are slightly stronger than needed and are hard to check, so we provide weaker and more primitive conditions. We extend the results to semiparametric models. We illustrate the sufficient conditions with endogenous quantile and single index examples. We also consider a semiparametric habit-based, consumption capital asset pricing model. There we find the rank condition is implied by an integral equation of the second kind having a one-dimensional null space.

JEL Classification: C12, C13, C23
Keywords: Identification, Local Identification, Nonparametric Models, Asset Pricing.

Date: 2009. Revised April 2011.

The NSF and the ERC provided financial support for this paper. Helpful comments were provided by D. Andrews, D. Brown, K. Evdikimov, L. Hansen, S. Hoderlein, H. Ichimura, T. Komarova, O. Linton, A. Santos and participants in seminars at June 2010 Cowles Conference, June 2010 Beijing Conference, June 2010 LSE Conference, and December 2010 EC2 conference.

1. Introduction

There are many important models that give rise to conditional moment restrictions. These restrictions often take the form

$$E[\rho(Y,X,\alpha_0)|W]=0,$$

where $\rho(Y,X,\alpha)$ has a known functional form but α_0 is unknown. Parametric models of this form are well known from the work of Hansen (1982), Chamberlain (1987), and others. Nonparametric versions are motivated by the desire to relax functional form restrictions. Identification and estimation of linear nonparametric conditional moment models have been studied by Newey and Powell (1988, 2003), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), Darolles, Fan, Florens, and Renault (2011), and others.

The purpose of this paper is to derive identification conditions for α_0 when ρ may be nonlinear in α. Models with nonlinear ρ are important. They include models with conditional quantile restrictions, as discussed in Chernozhukov and Hansen (2005) and Chernozhukov, Imbens, and Newey (2007). Allowing ρ to be nonlinear in α is also important for economic structural models and for semiparametric models, as further discussed below. In this paper we focus on conditions for local identification of these models. It should be possible to extend these results to provide global identification conditions by linking the local conditions with global conditions.

In parametric models there are easily interpretable rank conditions for local identification; see Fisher (1966) and Rothenberg (1971). A sufficient condition for local identification from solving a set of equations is that the equations are differentiable at the true value with full rank derivative matrix. We show a nonparametric analog of this result. If a nonparametric rank condition holds and the equations are differentiable at the true value with respect to a certain norm then the unknown function is locally identified. However, the conditions of this result are sensitive to the choice of norm for the derivative and are not primitive. For these reasons we add Hilbert space structure that leads to more primitive sufficient conditions. We also consider semiparametric models, providing conditions for identification of a vector of real parameters. These conditions are based on "partialling out" the nonparametric part and allow for identification of the parametric part even when the nonparametric part is not identified.

The usefulness of these conditions is illustrated by three examples. One example gives primitive conditions for local identification of the nonparametric endogenous quantile models of Chernozhukov and Hansen (2005) and Chernozhukov, Imbens, and Newey (2007). Another gives conditions for local identification of a semiparametric index model with endogeneity. There we give conditions for identification of parametric components when nonparametric components are not identified. The third example give conditions for local identification of a semiparametric consumption capital asset pricing model with habit formation.

In relation to previous literature, the nonparametric rank condition is a local version of identification conditions for linear conditional moment restriction models that were considered in Newey and Powell (1988, 2003). Chernozhukov, Imbens, and Newey (2007) also suggested the nonparametric rank condition and differentiability as sufficient conditions for local identification but did not use the right norm in defining differentiability. Florens and Sbai (2010) recently gave local identification conditions for games but their conditions do not seem to apply to the kind of conditional moment restrictions that arise in instrumental variable settings and are a primary subject of this paper. Also, the models we consider belong to the difficult class of nonlinear ill-posed inverse problems, that have not received much treatment in the mathematics literature.

Section 2 presents a general nonparametric local identification result and relates it to sufficient conditions for identification in parametric models. Section 3 gives more easily interpretable conditions for local identification and applies these to the endogenous quantile model. Section 4 provides conditions for identification in semiparametric models and applies these to the endogenous index model. Section 5 discusses the asset pricing example and Section 6 briefly concludes.

2. Nonparametric Models

To help explain the nonparametric results and give them context we give a brief description of sufficient conditions for local identification in parametric models. Let α be a $p\times 1$ vector of parameters and $m(\alpha)$ a $J\times 1$ vector of moment conditions with $m(\alpha_0)=0$ for true value α_0. Also let $|\cdot|$ denote the Euclidean norm in either $\Re^p$ or $\Re^J$ depending on the context. We say that α_0 is locally identified if there is a neighborhood of α_0 such that $m(\alpha)\neq 0$ for all $\alpha\neq\alpha_0$ in the neighborhood. Let m' denote the derivative of $m(\alpha)$ at α_0 when it exists. Sufficient conditions for local identification can be stated as follows:

If $m(\alpha)$ is differentiable at α_0 and $rank(m')=p$ then α_0 is locally identified.

This result follows from two observations: 1) By $rank(m')=p$, for $h\in\Re^p$ the Euclidean norm $|h|$ is equivalent to the norm $|m'h|$; 2) By m' being the derivative at α_0 there is a neighborhood of α_0 such that for all $\alpha\neq\alpha_0$ in that neighborhood

$$\frac{|m(\alpha)-m'(\alpha-\alpha_0)|}{|m'(\alpha-\alpha_0)|}=\frac{|m(\alpha)-m(\alpha_0)-m'(\alpha-\alpha_0)|}{|m'(\alpha-\alpha_0)|}<1. \tag{2.1}$$

This inequality implies $m(\alpha)\neq 0$.

To extend these observations to provide sufficient conditions for local identification of nonparametric models we will let α denote a function with true value α_0 and $m(\alpha)$ a function of the object of interest. The true value of the object of interest satisfies

$$m(\alpha_0)=0,$$

where we will be precise about the meaning of the equality in the discussion to follow. Conditional moment restrictions are an important example where $\rho(Y,X,\alpha)$ is a finite dimensional residual vector depending on an unknown function α and $m(\alpha)=E[\rho(Y,X,\alpha)|W]$.

To be precise we impose some mathematical structure. Assume that $\alpha\in\mathcal{A}$, a Banach space with norm $\|\cdot\|_{\mathcal{A}}$. Let $\mathcal{B}$ be a Banach space with a norm $\|\cdot\|_{\mathcal{B}}$ and assume that m maps $\mathcal{A}$ into $\mathcal{B}$, i.e. $m:\mathcal{A}\mapsto\mathcal{B}$. The restrictions of the model are that $\|m(\alpha_0)\|_{\mathcal{B}}=0$.

Definition: *α_0 is locally identified for $\mathcal{N}\subseteq\mathcal{A}$, with $\alpha_0\in\mathcal{N}$, if for all $\alpha\in\mathcal{N}$, $\|m(\alpha)\|_{\mathcal{B}}=0\Rightarrow\|\alpha-\alpha_0\|_{\mathcal{A}}=0$.*

This local identification concept is more general than the one introduced by Chernozhukov, Imbens and Newey (2007). Note that local identification is defined relative to a set $\mathcal{N}$. Often there will be $\epsilon > 0$ such that $\mathcal{N}$ is a subset of an open ball $\{\alpha : \|\alpha - \alpha_0\|_{\mathcal{A}} < \epsilon\}$. The set $\mathcal{N}$ may be strictly smaller than an open ball due to other restrictions being imposed on $\mathcal{N}$. For example, one could restrict $\mathcal{N}$ to be a bounded set in a Sobolev space. Or one could restrict $\mathcal{N}$ to only include α that are bounded functions. This restriction is useful for local identification in conditional moment models as further discussed below.

To formulate a nonparametric rank condition we will use a nonparametric version of the derivative. We will be specific below about what we require of this derivative but for now we just specify it to be a linear mapping $m' : \mathcal{A} \mapsto \mathcal{B}$. Under the conditions we give, m' will be a Gâteaux derivative at α_0, that can be calculated as

$$m'h = \frac{\partial}{\partial t} m(\alpha_0 + th)|_{t=0} \tag{2.2}$$

for $h \in \mathcal{A}$ and t a scalar. The result of this calculation can be used as a candidate for checking the conditions given below.

The following condition is a nonparametric rank condition.

Assumption 1 (Rank): *There is a continuous linear mapping $m' : \mathcal{A} \mapsto \mathcal{B}$ and a set $\mathcal{N}'$ containing α_0 such that for all $\alpha \in \mathcal{N}'$,*

$$\|m'(\alpha - \alpha_0)\|_{\mathcal{B}} = 0 \Rightarrow \|\alpha - \alpha_0\|_{\mathcal{A}} = 0. \tag{2.3}$$

This condition means that on $\mathcal{N}'$ the only α with $m'(\alpha - \alpha_0) = 0$ is $\alpha = \alpha_0$. In other words, on the domain $\{\alpha - \alpha_0 : \alpha \in \mathcal{N}'\}$ the null space of the linear operator m' is 0. If α were finite dimensional this condition would be equivalent to a full rank derivative matrix (as long as $\mathcal{N}'$ is open and nonempty). This motivates our interpretation of Assumption 1 as being like the rank condition for local identification in parametric models.

A similar condition is used to characterize identification in linear conditional moment models. For example, consider the linear conditional moment restriction where $Y = \alpha_0(X) + U$ and $E[U|W] = 0$. Let $\rho(Y, X, \alpha) = Y - \alpha(X)$. Here $m(\alpha) = E[Y - \alpha(X)|W]$ so that equation (2.2) is satisfied with $m'h = -E[h(X)|W]$. In this case Assumption

1 requires that $E[\alpha(X)-\alpha_0(X)|W] \neq 0$ for any $\alpha \in \mathcal{N}'$ with $\alpha-\alpha_0 \neq 0$. This is the completeness condition discussed in Newey and Powell (2003) with α restricted to $\mathcal{N}'$.

Similarly to local identification, the rank condition is defined in terms of a set $\mathcal{N}'$. In general there is a trade-off between different sets $\mathcal{N}'$. With smaller $\mathcal{N}'$ it is easier to verify rank but the identification result is weaker. For example, in the linear conditional moment model we could let $\mathcal{N}' = \{\alpha \in \mathcal{A} : \|\alpha-\alpha_0\|_{\mathcal{A}} < \infty\}$, where $\|h\|_{\mathcal{A}} = \{E[h(X)^2]\}^{1/2}$, and X and W are continuous random variables. Then Assumption 1 requires completeness of the conditional distribution of X given W. Sufficient conditions for completeness can be found in Newey and Powell (2003), Chernozhukov, Imbens, and Newey (2007), and Andrews (2011). If we consider the same mean square norms for $\|\cdot\|_{\mathcal{A}}$ and $\|\cdot\|_{\mathcal{B}}$ but restrict $\alpha-\alpha_0$ to be a bounded function of X, then Assumption 1 requires that the conditional distribution of X given W be bounded complete, which is weaker than completeness. See, for example, Mattner (1993), Chernozhukov and Hansen (2005), Blundell, Chen and Kristensen (2007), D'Haultfoeuille (2010), and Andrews (2011) for discussions of completeness and bounded completeness.

As for parametric models, the rank condition and differentiability will imply local identification. We base differentiability on the following definition.

Definition: *The map $m(\alpha)$ is differentiable on $\mathcal{N}''$ at α_0 for the norm $\|\cdot\|$ if for all $\delta > 0$ there is $\varepsilon > 0$ such that for all $\alpha \in \mathcal{N}''$ with $0 < \|\alpha-\alpha_0\| < \varepsilon$,*

$$\frac{\|m(\alpha)-m(\alpha_0)-m'(\alpha-\alpha_0)\|_{\mathcal{B}}}{\|\alpha-\alpha_0\|} < \delta.$$

This condition is the same as Frechet differentiability only we do not require that the domain of $m(\alpha)$ be a Banach space with norm $\|\cdot\|$, i.e. we do not require that all Cauchy sequences converge in the metric implied by $\|\cdot\|$. This condition does depend on $\|\cdot\|$, which is important, because different norms are not equivalent in nonparametric models. The rank condition and differentiability for the norm $\|h\| = \|m'h\|_{\mathcal{B}}$ are sufficient for local identification in nonparametric models.

Theorem 1: *If Assumption 1 is satisfied and* $m(\alpha)$ *is differentiable on* $\mathcal{N}'$ *at* α_0 *for the norm* $\|m'h\|_{\mathcal{B}}$ *then there is* $\varepsilon > 0$ *such that* α_0 *is locally identified for* $\mathcal{N} = \mathcal{N}' \cap \{\alpha : \|m'(\alpha - \alpha_0)\|_{\mathcal{B}} < \varepsilon\}$.

Differentiability is actually a stronger assumption than is needed for local identification result. Intuitively, it is sufficient that an inequality analogous to equation (2.1) be satisfied. For this reason we also consider identification when we just impose that inequality.

Assumption 2: (Derivative) *There is a set* $\mathcal{N}''$ *containing* α_0 *such that for all* $\alpha \in \mathcal{N}''$ *with* $\alpha \neq \alpha_0$,

$$\frac{\|m(\alpha) - m(\alpha_0) - m'(\alpha - \alpha_0)\|_{\mathcal{B}}}{\|m'(\alpha - \alpha_0)\|_{\mathcal{B}}} < 1.$$

The rank and derivative conditions are sufficient for local identification.

Theorem 2: *If Assumptions 1 and 2 are satisfied then* α_0 *is locally identified for* $\mathcal{N}$=$\mathcal{N}' \cap \mathcal{N}''$.

In linear conditional moment restriction models Assumption 2 will automatically be satisfied and $m(\alpha)$ will be differentiable for any norm. That is because in the linear case

$$m(\alpha) - m(\alpha_0) - m'(\alpha - \alpha_0) = 0.$$

Therefore Theorem 2 includes previous identification results for linear conditional moment restrictions as a special case.

It is important to note that Theorems 1 and 2 just provide sufficient, and not necessary, conditions for local identification. In particular, Assumption 1 may not be needed for identification in nonlinear models, although its absence may affect the attainable convergence rate of estimators, as occurs in parametric models, see Sargan (1983).

3. Local Identification in Hilbert Spaces

In Hilbert spaces we can give more primitive conditions for local identification of nonlinear models. This will be based on a lower bound for the rank norm $\|m'(\alpha - \alpha_0)\|_{\mathcal{B}}^2$.

Assumption 3: $(\mathcal{A}, \|\cdot\|_{\mathcal{A}})$ *and* $(B, \|\cdot\|_{\mathcal{B}})$ *are separable Hilbert spaces and either* a) *there is a set* $\mathcal{N}'$, *an orthonormal basis* $\{\phi_1, \phi_2, ...\}$, *and a positive, nonincreasing sequence*

$(\mu_1, \mu_2, ...)$ *such that for all* $\alpha \in \mathcal{N}'$,

$$\|m'(\alpha - \alpha_0)\|_{\mathcal{B}}^2 \geq \sum_{j=1}^{\infty} \mu_j^2 \langle \alpha - \alpha_0, \phi_j \rangle^2;$$

or b) $\mathcal{N}' = \mathcal{A}$ *and* m' *is a compact linear operator with positive singular values* $(\mu_1, \mu_2, ...)$.

The hypothesis in b) that m' is a compact operator is not very strong; e.g. see Kress (1999). It implies that there is an orthonormal basis $\{\phi_j : j = 1, \dots\}$ for $\mathcal{A}$ with

$$\|m'(\alpha - \alpha_0)\|_{\mathcal{B}}^2 = \sum_{j=1}^{\infty} \mu_j^2 \langle \alpha - \alpha_0, \phi_j \rangle^2,$$

where μ_j^2 are the eigenvalues and ϕ_j the eigenfunctions of the self-adjoint operator $m'^* m'$, so that condition a) is satisfied when $\mu_j > 0$ for all j. The assumption that the singular values are all positive is quite strong and implies the rank condition for $\mathcal{N}' = \mathcal{A}$. In the linear conditional moment restriction model this condition implies (L^2-) completeness of the conditional expectation $E[\cdot|W]$. Part a) differs from part b) by imposing a lower bound on $\|m'(\alpha - \alpha_0)\|_{\mathcal{B}}^2$ only over a subset $\mathcal{N}'$ of $\mathcal{A}$ and by allowing the basis $\{\phi_j\}$ to be different from the eigenfunction basis of the operator $m'^* m'$. In principle this allows us to impose restrictions on $\alpha - \alpha_0$, like boundedness and smoothness, which help the rank condition to hold. Assumption 3 a) is a natural extension of the reverse link condition in Chen and Reiß (2010), that is used to establish the rate of convergence for the linear nonparametric instrumental variables (NPIV) problem. It has been used in Chen and Pouzo (2008) for the convergence rates of their estimators of functions identified by nonlinear nonparametric conditional moment restrictions. Here we demonstrate that Assumption 3 a) is useful also for local identification.

It is difficult to show that $m(\alpha)$ is differentiable for the norm $\|m'h\|_{\mathcal{B}}$ even when it is easy to show differentiability for $\|h\|_{\mathcal{A}}$. It also seems often impossible to make $\|m'h\|_{\mathcal{B}}$ equivalent to $\|h\|_{\mathcal{A}}$ by restricting h. For these reasons we follow a different approach where we strengthen the assumption of differentiability of $m(\alpha)$ for the norm $\|h\|_{\mathcal{A}}$ and forge a link between the norms $\|h\|_{\mathcal{A}}$ and $\|m'h\|_{\mathcal{B}}$ using Assumption 3. The following condition strengthens differentiability of $m(\alpha)$ for the norm $\|h\|_{\mathcal{A}}$.

Assumption 4: *There are constants $L \geq 0$ and $r > 1$ and a set $\mathcal{N}''$ such that for all $\alpha \in \mathcal{N}''$,*

$$\|m(\alpha) - m(\alpha_0) - m'(\alpha - \alpha_0)\|_{\mathcal{B}} \leq L \|\alpha - \alpha_0\|_{\mathcal{A}}^r .$$

This condition is like Holder continuity of the derivative and $L = 0$ corresponds to the case that $m(\alpha)$ is linear in α. Let $\langle \cdot, \cdot \rangle$ denote the inner product on $\mathcal{A}$ and for any $q > 1$ define

$$\|h\|_q = \left[\sum_{j=1}^{\infty} \mu_j^{-2/(q-1)} \langle h, \phi_j \rangle^2 \right]^{1/2} .$$

The following is an identification result based on Theorem 1.

Theorem 3: *If Assumptions 3 and 4 (with $L > 0$) are satisfied then for any $C > 0$ and any q with $1 < q < r$ there is $\epsilon > 0$ such that α_0 is locally identified for*

$$\mathcal{N} = \mathcal{N}' \cap \mathcal{N}'' \cap \{\alpha : \|m'(\alpha - \alpha_0)\|_{\mathcal{B}} < \epsilon, \|\alpha - \alpha_0\|_q \leq C\}.$$

We can also base a result on the inequality of Assumption 2 rather than differentiability.

Theorem 4: *If Assumptions 3 and 4 (with $L > 0$) are satisfied then α_0 is locally identified on*

$$\mathcal{N} = \mathcal{N}' \cap \mathcal{N}'' \cap \{\alpha : \|\alpha - \alpha_0\|_r < L^{-1/(r-1)}\}.$$

These results can be explained in a straightforward way. In conditional moment restriction models the operator m' often will not have a continuous inverse, i.e. there will be an ill-posed inverse problem. Under Assumption 3 that corresponds to $\mu_j \to 0$ as $j \to \infty$. A consequence of this is that the norms $\|m'h\|_{\mathcal{B}}$ and $\|h\|_{\mathcal{A}}$ are not equivalent. However, $\|h\|_{\mathcal{A}}$ is generally a natural norm to use in the remainder of Assumption 4, as is illustrated in the quantile example below. Therefore, to obtain sufficient conditions for local identification it is useful to forge a link between the norms $\|m'h\|_{\mathcal{B}}$ and $\|h\|_{\mathcal{A}}$. The bounds on $\|h\|_q$ or $\|h\|_r$ allow us to forge such a link. Implicitly these bounds restrict the higher-order Fourier coefficients of h to go to zero at certain rates, so that smallness of $\|m'h\|_{\mathcal{B}}$ implies that $\|h\|_{\mathcal{A}}$ is small. In this way the link leads to a small remainder in the derivative expansion, which in turn leads to Assumption 2.

The bounds on $\|\alpha - \alpha_0\|_q$ in Theorems 3 and 4 require that the Fourier coefficients $\langle \alpha - \alpha_0, \phi_j \rangle$ of the deviations $\alpha - \alpha_0$ vanish faster as j grows than $\mu_j^{1/(q-1)}$. This bound is a "source condition" under Assumption 3 b) and is similar to conditions used by Florens, Johannes and Van Bellegem (2010) and others. Under Assumption 3 a) it is similar to norms in generalized Hilbert scales, for example, see Engl, Hanke, and Neubauer (1996) and Chen and Reiß (2010). Theorems 3 and 4 also remain valid if we impose uniform bounds on the size of Fourier coefficients, corresponding to a hyperrectangle instead of an ellipsoid.

To illustrate the usefulness of the results, we consider an endogenous quantile example where $0 < \tau < 1$ is a scalar and

$$\rho(Y, X, \alpha) = 1(Y \leq \alpha(X)) - \tau$$

Here we have

$$m(\alpha) = E[1(Y \leq \alpha(X))|W] - \tau.$$

Let $f_Y(y|X, W)$ denote the conditional density of Y given X and W.

Proposition 5: *If $f_Y(y|X, W)$ is continuously differentiable with $|df_Y(y|X, W)/dy| \leq L_1$, X has conditional pdf $f_X(x|W)$ given W and marginal pdf $f(x)$ satisfying $f_X(x|W) \leq L_2 f(x)$, and $m'h = E[f_Y(\alpha_0(X)|X, W)h(X)|W]$ satisfies Assumption 3, then α_0 is locally identified for*

$$\mathcal{N} = \mathcal{N}' \cap \{\alpha : \alpha : \|\alpha - \alpha_0\|_2 < (L_1 L_2)^{-1}\}.$$

This result gives a precise link between a neighborhood on which α_0 is locally identified and the bounds L_1 and L_2. Also, here the neighborhood is defined in terms of $\|\alpha - \alpha_0\|_2$ which is a strong norm. This result corrects Theorem 3.2 of Chernozhukov, Imbens, and Newey (2007) and has more primitive conditions than the global identification characterization of Chernozhukov and Hansen (2005). Horowitz and Lee (2007) impose analogous bounds on a strong norm in their paper on convergence rates of nonparametric endogenous quantile estimators.

4. Semiparametric Models

In this section, we consider local identification in possibly nonlinear semiparametric models, where α can be decomposed into a $p \times 1$ dimensional parameter vector β and nonparametric component g, so that $\alpha = (\beta, g)$. Let $|\cdot|$ denote the Euclidean norm for β and $\mathcal{G}$ the parameter space for g, where we assume that $\mathcal{G}$ is a Banach space with norm $\|\cdot\|_{\mathcal{G}}$, such as a Hilbert space. We focus here on the model

$$E[\rho(Y, X, \beta_0, g_0)|W] = 0,$$

where $\rho(y, x, \beta, g)$ is a $J \times 1$ vector of residuals. Here $m(\alpha) = E[\rho(Y, X, \beta, g)|W]$ will be considered as an element of the Hilbert space $\mathcal{B}$ of $J \times 1$ random vectors with inner product

$$\langle a, b\rangle = E[a(W)^T b(W)].$$

The differential $m'(\alpha - \alpha_0)$ can be expressed as

$$m'(\alpha - \alpha_0) = m'_\beta(\beta - \beta_0) + m'_g(g - g_0),$$

where m'_β is the derivative of $m(\beta, g_0) = E[\rho(Y, X, \beta, g_0)|W]$ with respect to β at β_0 and m'_g is the Gateaux derivative of $m(\beta_0, g)$ with respect to g at g_0. To give conditions for local identification of β_0 in the presence of the nonparametric component g it is helpful to partial out g. Let $\bar{\mathcal{M}}$ be the closure of the linear span $\mathcal{M}$ of $m'_g(g - g_0)$ for $g \in \mathcal{N}'_g$ where $\mathcal{N}'_g$ will be specified below. In general $\bar{\mathcal{M}} \neq \mathcal{M}$ because the linear operator m'_g does not have closed range (due to the ill-posed inverse problem). For the j^{th} unit vector e_j let

$$\zeta_j^* = \arg\min_{\zeta \in \bar{\mathcal{M}}} E[\{m'_\beta(W)e_j - \zeta(W)\}^T\{m'_\beta(W)e_j - \zeta(W)\}],$$

which exists by standard Hilbert space results, and satisfies

$$E[\{m'_\beta(W)e_j - \zeta_j^*\}^T m'_g(g - g_0)] = 0 \text{ for all } g \in \mathcal{N}'_g.$$

Define Π to be the $p \times p$ matrix with

$$\Pi_{jk} := E\left[\left\{m'_\beta(W)e_j - \zeta_j^*(W)\right\}^T \left\{m'_\beta(W)e_k - \zeta_k^*(W)\right\}\right], (j, k = 1, ..., p).$$

The following condition is important for local identification of β_0.

ASSUMPTION 5: *The mapping* $m' : \Re^p \times \mathcal{N}'_g \longrightarrow \mathcal{B}$ *is continuous and linear and* Π *is nonsingular.*

This assumption is similar to those imposed in Ai and Chen (2003, assumption 4.1(i)) and Chen and Pouzo (2009, assumption 2.10), who used it for establishing the $n^{-1/2}$-normality of the sieve minimum distance estimator for the parametric part. Nonsingularity of Π can be shown to be equivalent to finiteness of the semiparametric variance bound for β_0, when $E[\rho(Y, X, \alpha_0)\rho(Y, X, \alpha_0)^T|W]$ is bounded with smallest eigenvalue bounded away from zero; see, e.g., Chamberlain (1992). In the local identification analysis considered here it leads to local identification of β_0 without identification of g when $m(\beta_0, g)$ is linear in g. It allows us to separate conditions for identification of β_0 from conditions for identification of g, via the following result:

LEMMA 6: *If Assumption 5 is satisfied then there is* $\varepsilon > 0$ *such that for all* $(\beta, g) \in \Re^p \times \mathcal{N}'_g$,

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

An implication of Lemma 6 is that if Assumption 5 is satisfied then Assumption 1 for m'_g will imply Assumption 1 for m'. In this way Assumption 5 is a critical condition that allows us to specify conditions for local identification of β_0. One other condition is also useful for this purpose.

ASSUMPTION 6: *For every* $\varepsilon > 0$ *there is a neighborhood* B *of* β_0 *and a set* $\mathcal{N}'''_g$ *such that for all* $g \in \mathcal{N}'''_g$ *with probability one* $E[\rho(Y, X, \beta, g)|W]$ *is continuously differentiable in* β *on* B *and*

$$\sup_{g \in \mathcal{N}'''_g} \sqrt{E[\sup_{\beta \in B} |\partial E[\rho(Y, X, \beta, g)|W]/\partial \beta - \partial E[\rho(Y, X, \beta_0, g_0)|W]/\partial \beta|^2]} < \varepsilon.$$

It turns out that Assumptions 5 and 6 will be sufficient for local identification of β_0 when $m(\beta_0, g)$ is linear in g, i.e. for $m(\beta, g) = 0$ to imply $\beta = \beta_0$ when (β, g) is in some neighborhood of (β_0, g_0). This works because Assumption 5 removes the effect of unknown g on local identification of β_0 by partialling out g.

THEOREM 7: *If Assumptions 5 and 6 are satisfied and* $m(\beta_0, g)$ *is linear in* g *then there is a neighborhood* B *of* β_0 *and a set* $\mathcal{N}_g$ *containing* g_0 *such that* β_0 *is locally identified* *for $\mathcal{N} = B \times \mathcal{N}_g$. If, in addition, Assumption 1 is satisfied for m'_g and $\mathcal{N}'_g$ replacing m' and $\mathcal{N}'$ then $\alpha_0 = (\beta_0, g_0)$ is locally identified for $\mathcal{N} = B \times (\mathcal{N}_g \cap \mathcal{N}'_g)$.*

This result is more general than Florens, Johannes, and Van Bellegem (2008) and Santos (2010) in allowing for nonlinearities in β.

For semiparametric models that are nonlinear in g we can give local identification results based on differentiability of $m(\beta_0, g)$ with respect to g or on the more primitive conditions of Section 3. For brevity we will focus on a result based on Theorem 4.

THEOREM 8: *If Assumptions 3 and 4 are satisfied with $\alpha = g$, $m(\alpha) = m(\beta_0, g)$, $m' = m'_g$, $\mathcal{N}' = \mathcal{N}'_g, \mathcal{N}'' = \mathcal{N}''_g$ and Assumptions 5 and 6 are satisfied then there is a neighborhood B of β_0 and $\delta > 0$ such that $\alpha_0 = (\beta_0, g_0)$ is locally identified for $\mathcal{N} = B \times \mathcal{N}_g$, where*

$$\mathcal{N}_g = \mathcal{N}'_g \cap \mathcal{N}''_g \cap \mathcal{N}'''_g \cap \{g : \|g - g_0\|_r < \delta\}.$$

An interesting and potentially important example is a single index model with endogeneity. This model is given by

$$Y = g_0(X_1 + X_2^T \beta_0) + U, \quad E[U|W] = 0, \tag{4.4}$$

where β_0 is a vector of unknown parameters, $g_0(\cdot)$ is an unknown function, and W are instrumental variables. The scale of the parametric part is not identified separately, and hence, we normalize the coefficient of X_1 to 1. Here

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

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

$$m'_\beta = -E[g'_0(V) X_2|W].$$

Let ζ_j^* denote the projection of $m'_\beta e_j = -E[g'_0(V) X_{2j}|W]$ on the mean-square closure of the set $\{E[h(V)|W] : E[h(V)^2] < \infty\}$ and Π the matrix with $\Pi_{jk} = E[(m'_\beta e_j - \zeta_j^*)(m'_\beta e_k - \zeta_k^*)]$.

PROPOSITION 9: *Consider the model of (4.4). 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) $E[|X_2|^4] < \infty$, and c) Π is nonsingular then there is a neighborhood B of β_0 and $\delta > 0$*

such that for

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

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

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 model in Newey and Powell (1988, 2003) the index structure lowers the requirements for identification by requiring that $m_g' h = -E[h(V)|W]$ be complete on $\mathcal{N}_g'$ rather than requiring completeness of $E[r(X)|W]$. 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 nonlinear 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.

Proposition 9A: *Consider the model of (4.4). 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* $\lambda \neq 0$, $E[g_0'(V)\lambda' X_2|W]$ *is not measurable with respect to* $T(W)$ *then* Π *is nonsingular.*

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 $E[g_0'(V)X_2|W]$ 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 $E[g_0'(V)X_2|W]$ depends on all of W.

5. Semiparametric CCAPM Models

Consumption capital asset pricing models (CCAPM) provide interesting examples of nonparametric and semiparametric moment restrictions, see Gallant and Tauchen (1989), Hansen, Heaton, Lee, and Roussanov (2007), Chen and Ludvigson (2009), and others. In this section, we illustrate our results by applying them to identification of a particular semiparametric specification. These results could easily be extended to other specifications. Newey and Powell (1988), Chen and Ludvigson (2009), Lewbel and Linton (2010), and Escanciano and Hoderlein (2010) have analyzed nonparametric marginal utility specifications.

To describe the model let C_t denote consumption level at time t and $c_t \equiv C_t/C_{t-1}$ be consumption growth. Suppose that the marginal utility of consumption at time t is given by

$$MU_t = C_t^{-\gamma_0} g_0(C_t/C_{t-1}) = C_t^{-\gamma_0} g_0(c_t),$$

where $g_0(c)$ is an unknown positive function. For this model the intertemporal marginal rate of substitution is

$$\delta_0 MU_{t+1}/MU_t = \delta_0 c_{t+1}^{-\gamma_0} g_0(c_{t+1})/g_0(c_t),$$

where $0 < \delta_0 < 1$ is the rate of time preference. Let $R_{t+1} = (R_{t+1,1}, ..., R_{t+1,J})^T$ be a $J \times 1$ vector of gross asset returns. A semiparametric CCAPM equation is then given by

$$E[R_{t+1}\delta_0 c_{t+1}^{-\gamma_0}\{g_0(c_{t+1})/g_0(c_t)\}|W_t] = e, \tag{5.5}$$

where $W_t \equiv (Z_t, c_t)$ is a vector of variables observed by the agent at time t, and e is a $J\times 1$ vector of ones. This corresponds to an external habit formation model with only one lag as considered by Chen and Ludvigson (2009). We focus here on consumption growth $c_t = C_t/C_{t-1}$ to circumvent the potential nonstationarity of the level of consumption, as has long been done in this literature, e.g. Hansen and Singleton (1982).

As discussed in the previous Section, local identification of $\beta_0 = (\delta_0, \gamma_0)^T$ and g_0 will follow from nonsingularity of a matrix and from identification of the nonparametric part at β_0. Identification of β_0 is straightforward while nonparametric identification is interesting, so we focus first on the nonparametric part. We consider two approaches, based on an integral equation of the first and second kind respectively. While our results are specific to the semiparametric model of equation (5.5), both approaches are applicable to a broad class of semiparametric consumption based asset pricing models, such as models with durable good consumption, housing, etc..

5.1. **Identification via integral equation of first-kind.** Let $h(c_{t+1}, c_t) = g(c_{t+1})/g(c_t)$. If g_0 is known to be bounded and bounded away from zero then it is sufficient for identification of h that at least one of the "adjusted" conditional expectation operators

$$E_j^*[h(c_{t+1}, c_t)|W_t] = \frac{E[R_{t+1,j}c_{t+1}^{-\gamma_0}h(c_{t+1}, c_t)|W_t]}{E[R_{t+1,j}c_{t+1}^{-\gamma_0}|W_t]}$$

be boundedly complete. Since identifying $h_0(c_{t+1}, c_t)$ identifies g_0 only up to scale we also normalize g_0 to satisfy $E[g_0(c_t)^2] = 1$. Let $\bar{\mathcal{G}}$ denote the set of positive functions g that are bounded, bounded away from zero, and satisfy $E[g(c_t)^2] = 1$.

Assumption 7a: *$E_j^*[\cdot|W_t]$ is boundedly complete for some j and $g_0 \in \bar{\mathcal{G}}$.*

An alternative scale normalization is also interesting. If $g_0(c^*) = 1$ for some c^*, then g_0 is identified by $g_0(c_{t+1}) = h_0(c_{t+1}, c^*)$. We could directly impose this scale normalization in equation (5.5) and then $g_0(c_{t+1})$ is identified when at least one of the "return adjusted" conditional expectation operators

$$E_j^*[g(c_{t+1})|Z_t, c_t = c^*] = \frac{E[R_{t+1,j}c_{t+1}^{-\gamma_0}g(c_{t+1})|Z_t, c_t = c^*]}{E[R_{t+1,j}c_{t+1}^{-\gamma_0}|Z_t, c_t = c^*]}$$

is boundedly complete. This identification condition is consistent with existing findings that c_t is a "weak instrument" and that one needs other more powerful instruments Z_t for strong identification and reliable estimation of CCAPM; see, e.g., Stock and Wright (2000). In fact, Chen and Ludvigson (2009) find that all the empirical results of their semiparametric habit formation CCAPM remain virtually unchanged when c_t is dropped from the conditioning set $W_t = (Z_t, c_t)$.

5.2. **Identification via integral equation of second-kind.** Multiplying $g_0(c_t)$ through CCAPM equation (5.5) we see that identification of $g_0(c)$ (up to scale) just requires a unique solution (up to scale) of

$$E[\delta_0 R_{t+1} c_{t+1}^{-\gamma_0} g(c_{t+1})|W_t] - g(c_t)e = 0. \tag{5.6}$$

This is a vector homogenous linear integral equation of the second kind. It will identify $g_0(c)$ (up to scale) if and only if the intersection of its null space $\mathcal{N}$ with $\bar{\mathcal{G}}$ is a singleton. A one-dimensional null space $\mathcal{N}$ is thus sufficient for identification of g_0, since the $E[g_0(c_t)^2] = 1$ normalization will reduce that to a singleton. This condition is analogous to the well known rank condition for identification in parametric simultaneous equations models, which requires a one-dimensional null space for the restriction matrix multiplied by the matrix of structural coefficients (see Fisher, 1966, Theorem 2.3.1).

Assumption 7b: *$\mathcal{N}$ is one dimensional and $g_0 \in \bar{\mathcal{G}}$.*

The following reasoning suggests that one-dimensional $\mathcal{N}$ is a weak condition that is generic. Note first that $\mathcal{N} = \cap_{j=1}^{J} \mathcal{N}_j$ where $\mathcal{N}_j$ is the null space of the operator $E[\delta_0 R_{t+1,j} c_{t+1}^{-\gamma_0} g(c_{t+1})|W_t] - g(c_t)$. Furthermore, for any finite valued, measurable function $T(Z_t)$, by iterated expectations it follows that

$$\mathcal{N}_j \subseteq \mathcal{N}_j^T = \{g : E[\delta_0 R_{t+1,j} c_{t+1}^{-\gamma_0} g(c_{t+1})|T(Z_t), c_t] - g(c_t) = 0\}.$$

If $E[\delta_0 R_{t+1,j} c_{t+1}^{-\gamma_0} g(c_{t+1})|T(Z_t) = \tau, c_t]$ is a compact operator then $\mathcal{N}_j^T$ is finite dimensional (see e.g. Kress, 1999, Chapter 3). Therefore, $\mathcal{N}_j \subseteq \cap_T \mathcal{N}_j^T$ has finite dimension, where the intersection is over all measurable functions T. Furthermore, if Z_t is continuously distributed then there will be an infinite number of distinct such T. Generically the intersection of an infinite number of finite dimensional spaces is the linear space that is common to each, which is just constant multiples of $g_0(c)$, so that $\mathcal{N}_j$ is one dimensional. It follows that generically $\mathcal{N}$ will be one-dimensional, and hence $g_0(c)$ identified (up to scale).

Many overidentifying restrictions may be present in this model. The argument given for generic identification holds if $J = 1$ and Z_t consists of one continuous variable. Larger J and more instrumental variables in Z_t constitute overidentifying restrictions.

The identification condition in Assumption 7B is interesting because it shows that Assumption 1 need not reduce to completeness of a conditional expectation. Instead, the rank condition holds if an integral equation of the second kind has a one-dimensional null space. Lewbel and Linton (2010) and Escanciano and Hoderlein (2010) also consider identification of nonparametric marginal utility of consumption, $MU_t = \alpha_0(C_t)$, using an integral equation of the second kind, but their formulations and conditions are different from ours.

Imposing the scale normalization $g_0(c^*) = 1$ gives another view of identification from an integral equation of the second kind. With that normalization (5.6) becomes the integral equation of the first kind discussed in the previous subsection, namely

$$E[\delta_0 R_{t+1} c_{t+1}^{-\gamma_0} g_0(c_{t+1})|W_t^*] - e = 0 \quad \text{with } W_t^* \equiv (Z_t, c_t = c^*).$$

Turning to the identification of parametric component $\beta_0 = (\delta_0, \gamma_0)^T$, let

$$m_{\beta 1}(W_t) = E[R_{t+1} c_{t+1}^{-\gamma_0} g_0(c_{t+1})|W_t],\ m_{\beta 2}(W_t) = -\delta_0 E[R_{t+1} \ln(c_{t+1}) c_{t+1}^{-\gamma_0} g_0(c_{t+1})|W_t].$$

Let $\bar{\mathcal{M}}$ be the mean square closure of the linear span of

$$\{E[R_{t+1}\delta_0 c_{t+1}^{-\gamma_0} g(c_{t+1})|W_t] - g(c_t)e : E[g(c_t)^2] < \infty\}.$$

Define

$$\zeta_j^* = \arg\min_{\zeta \in \bar{\mathcal{M}}} E[\{m_{\beta j}(W_t) - \zeta(W_t)\}^T \{m_{\beta j}(W_t) - \zeta(W_t)\}],$$

and let matrix Π be a 2×2 symmetric matrix with

$$\Pi_{jk} = E[\{m_{\beta j}(W_t) - \zeta_j^*(W_t)\}^T \{m_{\beta k}(W_t) - \zeta_j^*(W_t)\}], (j, k = 1, 2).$$

Nonsingularity of Π leads to local identification of $\beta_0 = (\delta_0, \gamma_0)^T$.

The matrix Π appears to be nonsingular quite generally as long as W_t includes other variables Z_t in addition to c_t. Similarly to the index example the instrument c_t is used in identifying g_0 so that addition instruments will be useful for identifying β. It should be possible to formulate necessary and sufficient conditions similar to those for the index model but for brevity we leave this to future work.

To help Assumption 6 be satisfied it is useful to impose a dominance condition. For any $\Delta > 0$ define

$$D_t = (1 + |R_{t+1}|)[2 + |\ln(c_{t+1})|] \sup_{\gamma \in [\gamma_0 - \Delta, \gamma_0 + \Delta]} c_{t+1}^{-\gamma}.$$

We can now give a local identification result for this model. Let $\bar{\mathcal{G}}$ denote the set of functions $g(c)$ that are bounded and bounded away from zero.

PROPOSITION 10: *If Π is nonsingular, $g_0(\cdot) \in \bar{\mathcal{G}}$, $0 < \delta_0 < 1$, and $E[D_t^2] < \infty$ then there is a neighborhood B of β_0 and $\varepsilon > 0$ such that for*

$$\mathcal{N}_g^{\varepsilon} = \{g : E[E[D_t^2|W_t]|g(c_{t+1}) - g_0(c_{t+1})|^2] < \varepsilon, g \in \bar{\mathcal{G}}\},$$

β_0 is locally identified for $\mathcal{N} = B \times \mathcal{N}_g^{\varepsilon}$. Furthermore, if Assumption 7A or 7B is satisfied then (β_0, g_0) is locally identified for $\mathcal{N} = B \times \mathcal{N}_g^{\varepsilon}$.

6. Conclusion

We provide sufficient conditions for local identification for a general class of semi-parametric and nonparametric conditional moment restriction models. We find that the choice of norms and neighborhoods are important for local identification of nonparametric models. We provide new examples to illustrate the usefulness of our identification results.

7. Appendix

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

LEMMA A1: *If a) $\mathcal{M}$ is a closed linear subset of Hilbert space $\mathcal{H}$;b) $b_j \in \mathcal{H}$ ($j = 1, \dots, p$), c) the $p \times p$ matrix Π with $\Pi_{jk} = \langle b_j - \mathrm{Proj}(b_j|\mathcal{M}), b_k - \mathrm{Proj}(b_k|\mathcal{M})\rangle$ is non-singular then for $b = (b_1, \dots, b_p)$ there exists $\varepsilon > 0$ such that for all $a \in \Re^p$ and $\zeta \in M$,*

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

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

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

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

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

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

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

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

$$\begin{aligned}
\left\|b^T a + \zeta\right\| &\geq \varepsilon_1|a| + (\varepsilon_1/2C^*)\|\zeta\| - \varepsilon_1|a|/2 \\
&= (\varepsilon_1/2)|a| + (\varepsilon_1/2C^*)\|\zeta\| \geq \varepsilon\left(|a| + \|\zeta\|\right).Q.E.D.
\end{aligned}$$

Proof of Theorem 1: Choosing $\delta = 1$ in the definition of differentiability, it follows that there is an $\varepsilon > 0$ such that for all $\alpha \in \mathcal{N}'$ with $0 < \|m'(\alpha - \alpha_0)\|_{\mathcal{B}} < \varepsilon$,

$$\frac{\|m(\alpha) - m'(\alpha - \alpha_0)\|_{\mathcal{B}}}{\|m'(\alpha - \alpha_0)\|_{\mathcal{B}}} = \frac{\|m(\alpha) - m(\alpha_0) - m'(\alpha - \alpha_0)\|_{\mathcal{B}}}{\|m'(\alpha - \alpha_0)\|_{\mathcal{B}}} < 1.$$

This can only be the case if $m(\alpha) \neq 0$. Therefore, $m(\alpha) \neq 0$ for all $\alpha \in \mathcal{N}$ with $\alpha \neq \alpha_0$ *Q.E.D.*

Proof of Theorem 2: For $\alpha \in \mathcal{N}$ it follows by Assumptions 1 and 2 that

$$\frac{\|m(\alpha) - m'(\alpha-\alpha_0)\|_{\mathcal{B}}}{\|m'(\alpha-\alpha_0)\|_{\mathcal{B}}} = \frac{\|m(\alpha) - m(\alpha_0) - m'(\alpha-\alpha_0)\|_{\mathcal{B}}}{\|m'(\alpha-\alpha_0)\|_{\mathcal{B}}} < 1.$$

The conclusion then follows as in the proof of Theorem 1. *Q.E.D.*

Proof of Theorem 3: Assumption 3 implies Assumption 1 and

$$\|m'h\|_{\mathcal{B}}^2 \geq \sum_{j=1}^{\infty} \mu_j^2 \left\langle h, \phi_j \right\rangle^2,$$

where $\langle h, \phi_j\rangle$ are the Fourier coefficients satisfying $h = \sum_{j=1}^{\infty} \langle h, \phi_j\rangle \phi_j$ and the inequality is an equality under Assumption 3 b). Also, by the Holder inequality, for any $q > 1$ and $a_j = |\langle h, \phi_j\rangle|$,

$$\begin{aligned}
(\sum_j a_j^2)^{1/2} &= \left(\sum_j \mu_j^{-2/q} a_j^{2-2/q} \mu_j^{2/q} a_j^{2/q}\right)^{1/2} \\
&\leq \left(\sum_j \mu_j^{-2/(q-1)} a_j^2\right)^{(q-1)/2q} \left(\sum_j \mu_j^2 a_j^2\right)^{1/2q}.
\end{aligned}$$

Therefore we have

$$\|h\|_{\mathcal{A}} \leq \left(\|h\|_q\right)^{1-1/q} \|m'h\|_{\mathcal{B}}^{1/q}. \tag{7.7}$$

Let $\mathcal{N}''' = \mathcal{N}'' \cap \{\|\alpha - \alpha_0\|_q \leq C\}$. Then, by Assumption 4, for $\alpha \in \mathcal{N}'''$ we have

$$\begin{aligned}
\|m(\alpha) - m(\alpha_0) - m'(\alpha-\alpha_0)\|_{\mathcal{B}} &\leq L\|\alpha-\alpha_0\|_q^{r(1-1/q)} \|m'(\alpha-\alpha_0)\|_{\mathcal{B}}^{r/q} \\
&\leq LC^{r(1-1/q)} \|m'(\alpha-\alpha_0)\|_{\mathcal{B}}^{r/q}.
\end{aligned}$$

Since $r/q > 1$ it follows that $m(\alpha)$ is differentiable on $\mathcal{N}'''$ at α_0 for the norm $\|m'h\|_{\mathcal{B}}$, so the conclusion follows by Theorem 1. Q.E.D.

Proof of Theorem 4: By Assumption 4 and equation (7.7) with $q = r$ it follows that for $\alpha \in \mathcal{N}$ with $\alpha \neq \alpha_0$,

$$\begin{aligned}
\|m(\alpha) - m'(\alpha-\alpha_0)\|_{\mathcal{B}} &\leq L\|\alpha-\alpha_0\|_r^{r-1} \|m'(\alpha-\alpha_0)\|_{\mathcal{B}} \\
&< \|m'(\alpha-\alpha_0)\|_{\mathcal{B}},
\end{aligned}$$

implying $m(\alpha) \neq 0$ by Theorem 2. *Q.E.D.*

Proof of Proposition 5: Let $F(y|X,W) = \Pr(Y \leq y|X,W)$, $m(\alpha) = E[1(Y \leq \alpha(X))|W] - \tau$, and $m'h = E[f_Y(\alpha_0(X)|X,W)h(X)|W]$, so that by iterated expectations,

$$m(\alpha) = E[F(\alpha(X)|X,W)|W] - \tau.$$

Then by a mean value expansion, and by $f_Y(y|X,W)$ continuously differentiable

$$\begin{aligned} & |F(\alpha(X)|X,W) - F(\alpha_0(X)|X,W) - f_Y(\alpha_0(X)|X,W)(\alpha(X) - \alpha_0(X))| \\ = \;& |[f_Y(\bar{\alpha}(X)|X,W) - f_Y(\alpha_0(X)|X,W)][\alpha(X) - \alpha_0(X)]| \\ \leq \;& L_1 [\alpha(X) - \alpha_0(X)]^2 . \end{aligned}$$

Then for $L_1 L_2 = L$

$$\begin{aligned} |m(\alpha) - m(\alpha_0) - m'(\alpha - \alpha_0)| &\leq L_1 E\left[\{\alpha(X) - \alpha_0(X)\}^2 |W\right] \\ &\leq LE[\{\alpha(X) - \alpha_0(X)\}^2] = L \|\alpha - \alpha_0\|_{\mathcal{A}}^2 . \end{aligned}$$

Therefore,

$$\|m(\alpha) - m(\alpha_0) - m'(\alpha - \alpha_0)\|_{\mathcal{B}} \leq L \|\alpha - \alpha_0\|_{\mathcal{A}}^2 ,$$

so that Assumption 4 is satisfied with $r = 2$ and $\mathcal{N}'' = \mathcal{A}$. The conclusion then follows from Theorem 4. *Q.E.D.*

Proof of Lemma 6: Apply Lemma A1 with $\mathcal{B}$ there equal to $\mathcal{B}$ from the text, $\bar{\mathcal{M}}$ in Lemma A1 equal to 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 j^{th} unit vector e_j, and $a = \beta - \beta_0$. Then for all $(\beta, g) \in \Re^p \times \mathcal{N}'_g$ we have

$$m'(\alpha - \alpha_0) = b'a + \zeta, b'a = m'_\beta(\beta - \beta_0), \zeta = m'_g(g - g_0) \in \bar{\mathcal{M}}.$$

The conclusion then follows by the conclusion of Lemma A1. *Q.E.D.*

Proof of Theorem 7: Let ε be from the conclusion of Lemma 6 and let B and $\mathcal{N}_g = \mathcal{N}_g'''$ be as in Assumption 6 with

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

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

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

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

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

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, suppose $\beta = \beta_0$ and $g \in \mathcal{N}_g \cap \mathcal{N}'_g$ with $g \neq g_0$. Then

$$\|m'(\alpha - \alpha_0)\|_{\mathcal{B}} = \left\| m'_g(g - g_0) \right\|_{\mathcal{B}} > 0,$$

while $\|m(\alpha) - m'(\alpha - \alpha_0)\|_{\mathcal{B}} = 0$, so $m(\alpha) \neq 0$ follows as in the proof of Theorem 1. *Q.E.D.*

Proof of Theorem 8: Let ε be from the conclusion of Lemma 6. Then similarly to the proof of Theorem 7, for all $g \in \mathcal{N}''_g \cap \mathcal{N}'''_g$.

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

where the last inequality is strict if $\beta \neq \beta_0$ and L is from Assumption 4. Choose $\delta = (\varepsilon/L)^{1/(r-1)}$. Then for $g \in \mathcal{N}'_g$ it follows as in the proof of Theorem 4 that for $g \neq g_0$ with $\|g - g_0\|_r < \delta$

$$L \, \|g - g_0\|^r_{\mathcal{A}} \leq L \, \|g - g_0\|^{r-1}_r \left\| m'_g(g - g_0) \right\|_{\mathcal{B}} < \varepsilon \left\| m'_g(g - g_0) \right\|_{\mathcal{B}} .$$

Combining this inequality with the previous one, it then follows from Lemma 6 that for $\alpha \neq \alpha_0$, implying either $\beta \neq \beta_0$ or $g \neq g_0$,

$$\|m(\alpha) - m'(\alpha - \alpha_0)\|_{\mathcal{B}} < \varepsilon(|\beta - \beta_0| + \left\|m'_g(g - g_0)\right\|_{\mathcal{B}} \leq \|m'(\alpha - \alpha_0)\|_{\mathcal{B}}.$$

The conclusion then follows by Theorem 2. *Q.E.D.*

Proof of Proposition 9: The proof will proceed by verifying the conditions of Theorem 7. Note that Assumption 5 is satisfied. We now check Assumption 6. 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

$$E[\sup_{\Delta\in\bar{B}} \left|g'(X_1 + X_2^T(\beta + \Delta))X_2\right| |W] \leq CE[|X_2| |W] < \infty \text{ a.s.}$$

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

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

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

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

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

$$\begin{aligned}
& E[\sup_{\beta\in B} \left|\partial m(\alpha)(W)/\partial\beta - m'_\beta(W)\right|^2] \\
= \;& E[\sup_{\beta\in B} |E[\{g'(v(X,\beta)) - g_0'(V)\}X_2|W]|^2] \leq E[|X_2|^2 \sup_{\beta\in B} |g'(v(X,\beta)) - g_0'(V)|^2] \\
\leq \;& 2E[|X_2|^2 \sup_{\beta\in B} |g'(v(X,\beta)) - g_0'(v(X,\beta))|^2] + 2E[|X_2|^2 \sup_{\beta\in B} |g_0'(v(X,\beta)) - g_0'(V)|^2] \\
\leq \;& 2\delta^2 E[|X_2|^2] + 2C_g^2 E[|X_2|^4] \sup_{\beta\in B} |\beta - \beta_0|^2 < \varepsilon^2.
\end{aligned}$$

Thus Assumption 6 is satisfied. The other conditions of Theorem 7 are assumed to be satisfied so the conclusion follows from Theorem 7. *Q.E.D.*

Proof of Proposition 9A: Suppose first that the conditional distribution of W given V is complete. Note that by the projection definition of for all $h(V)$ with finite mean-square we have

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

Therefore,

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

Completeness of the conditional distribution of W given V then implies that $-E[g_0'(V)X_{2j}|W] - \zeta_j^*(W) = 0$, 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 $-E[g_0'(V)\lambda' X_2|W]$ on $\bar{\mathcal{M}}$. Since $E[h(V)|W] = E[h(V)|T(W)]$ it follows that $\zeta_\lambda^*(W)$ is measurable with respect to $T(W)$. Since $E[g_0'(V)\lambda' X_2|W]$ is not measurable with respect to $T(W)$, we have $-E[g_0'(V)\lambda' X_2|W] - \zeta_\lambda^*(W) \neq 0$, so that

$$\lambda'\Pi\lambda = E[\{-E[g_0'(V)\lambda' X_2|W] - \zeta_\lambda^*(W)\}^2] > 0.$$

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

Proof of Proposition 10: The proof will proceed by verifying the conditions of Theorem 7 for the linear in g version of the moment condition from eq. (5.6). For bounded h let

$$m_g' h = E[\delta_0 R_{t+1} c_{t+1}^{-\gamma_0} h(c_{t+1})|W_t] - h(c_t)e$$

and $m'(\alpha - \alpha_0) = m_\beta(W)'(\beta - \beta_0) + m_g'(g - g_0)$. Let $\mathcal{A}$ be the set of functions $g(\cdot)$ with norm

$$\|g\|_{\mathcal{A}} = \sqrt{E[\{E[D_t^2|W_t] + 1\}g(c_t)^2]}.$$

Note that $\delta_0^2 |R_{t+1}|^2 c_{t+1}^{-2\gamma_0} \leq CD_t^2$. Then by the Cauchy-Schwartz inequality we have

$$\begin{aligned} & \left\| E[\delta_0 R_{t+1} c_{t+1}^{-\gamma_0} h(c_{t+1})|W_t] - h(c_t)e \right\|_{\mathcal{B}}^2 \\ \leq\ & CE[E[\delta_0 R_{t+1}' c_{t+1}^{-\gamma_0} h(c_{t+1})|W_t]E[\delta_0 R_{t+1} c_{t+1}^{-\gamma_0} h(c_{t+1})|W_t] + h(c_t)^2] \\ \leq\ & CE[E[D_t^2|W_t]E[h(c_{t+1})^2|W_t] + CE[h(c_t)^2] \leq C\|h\|_{\mathcal{A}}^2. \end{aligned}$$

It also follows similarly that $\|m_\beta(W)\|_{\mathcal{B}} < \infty$. Therefore $m'(\alpha - \alpha_0)$ is continuous so that Assumption 5 is satisfied.

We now check Assumption 6. Let $\beta = (\delta, \gamma)$ and for bounded g let $H_{t+1}(\beta, g) = \delta R_{t+1} c_{t+1}^{-\gamma} g(c_{t+1})$. Note that H_{t+1} is twice continuously differentiable in β and that there is a neighborhood B of β_0 such that

$$\begin{aligned}
\sup_{\beta \in B} \left| \frac{\partial H_{t+1}(\beta, g)}{\partial \beta} \right| &\leq D_t g(c_{t+1}), \\
|E[\partial H_{t+1}(\beta, g)/\partial\beta - \partial H_{t+1}(\beta, g_0)/\partial\beta | W_t]|^2 &\leq E[D_t^2|W_t] E[|g(c_{t+1}) - g_0(c_{t+1})|^2 | W_t], \\
|E[\partial H_{t+1}(\beta, g_0)/\partial\beta - \partial H_{t+1}(\beta_0, g_0)/\partial\beta | W_t]|^2 &\leq E[D_t^2|W_t] E[g_0(c_{t+1})^2 | W_t] \, |\beta - \beta_0|^2 .
\end{aligned}$$

By $E[D_t^2] < \infty$ we have $E[D_t|W_t]$ exists a.s. implying that $E[H_{t+1}(\beta, g)|W_t]$ is continuously differentiable on B with probability one with

$$\frac{\partial E[H_{t+1}(\beta, g)|W_t]}{\partial \beta} = E[\partial H_{t+1}(\beta, g)/\partial\beta | W_t].$$

By $g_0(c_{t+1})$ bounded we also have

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

Note that by iterated expectations,

$$E[E[D_t^2|W_t]E[|g(c_{t+1}) - g_0(c_{t+1})|^2|W_t]] = E[E[D_t^2|W_t]|g(c_{t+1}) - g_0(c_{t+1})|^2].$$

Then choosing $\mathcal{N}_g$ and B so that

$$E[E[D_t^2|W_t]|g(c_{t+1}) - g_0(c_{t+1})|^2] < \varepsilon/4, |\beta - \beta_0|^2 < \varepsilon/4E[D_t^2],$$

we have

$$\begin{aligned}
& E[\sup_{\beta \in B} \left| \partial m(\alpha)(W)/\partial\beta - m'_\beta(W) \right|^2] \\
\leq & \ 2E[E[D_t^2|W_t]E[|g(c_{t+1}) - g_0(c_{t+1})|^2|W_t]] + 2E[E[D_t^2|W_t]]\varepsilon/4E[D_t^2] < \varepsilon.
\end{aligned}$$

so $E[E[D_t^2|W_t]|g(c_{t+1}) - g_0(c_{t+1})|^2] < \varepsilon/4$ and B, choosing ε small enough in the the conditions of Proposition 10 it follows that Assumption 6 is satisfied. The first conclusion then follows by the first conclusion of Theorem 7.

To show the second conclusion it suffices to show that m'_g satisfies the rank condition under Assumption 7A or 7B. Consider first Assumption 7A. Consider a $g \in \bar{\mathcal{G}}$ with

$m_g'(g-g_0)=0$. Then $m_g'g=m_g'g_0=0$. Divide $m_g'g=0$ by $g(c_t)$ and $m_g'g_0=0$ by $g_0(c_t)$ to obtain

$$E[\delta_0 R_{t+1}c_{t+1}^{-\gamma_0}g(c_{t+1})/g(c_t)|W_t]=e=E[\delta_0 R_{t+1}c_{t+1}^{-\gamma_0}g_0(c_{t+1})/g_0(c_t)|W_t].$$

Since $0<\delta_0<1$ and $E[R_{t+1,j}c_{t+1}^{-\gamma_0}|W_t]$ is positive random variable,

$$E_j^*[\{g(c_{t+1})/g(c_t)\}\,|W_t]=\frac{E[R_{t+1,j}c_{t+1}^{-\gamma_0}g(c_{t+1})/g(c_t)|W_t]}{E[R_{t+1,j}c_{t+1}^{-\gamma_0}|W_t]}=\frac{E[R_{t+1,j}c_{t+1}^{-\gamma_0}g_0(c_{t+1})/g_0(c_t)|W_t]}{E[R_{t+1,j}c_{t+1}^{-\gamma_0}|W_t]}.$$

By Assumption 7A (bounded completeness for some j), it follows that $g(c_{t+1})/g(c_t)=g_0(c_{t+1})/g_0(c_t)$ almost surely. Square both sides and integrate both sides with respect the distribution of c_{t+1} and to obtain $g(c_t)^{-2}=g_0(c_t)^{-2}$. Since $g(c_t)>0$ it follows that $g(c_t)=g_0(c_t)$.

Consider next Assumption 7B. Then for $m_g'(g-g_0)=0$ we have $m_g'g=0$ so $g(c_t)=Kg_0(c_t)$ by m_g' having a one-dimensional null space containing g_0. Squaring and integrating both sides with respect the distribution of c_t gives $K^2=1$. Since $g(c_t)$ is restricted to be positive it follows that $K=1$. *Q.E.D.*

8. References

Ai, C. and X. Chen (2003) "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica* 71, 1795-1844.

Andrews, D.W.K. (2011) "Examples of L2-Complete and Boundedly-Complete Distributions," Yale Working Paper.

Blundell, R., X. Chen, and D. Kristensen (2007), "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica* 75 1613-1669.

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.

Chen, X. and S. C. Ludvigson (2009) "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Journal of Applied Econometrics* 24, 1057–1093.

Chen, X., and Pouzo, D., (2008) "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals," *Cowles Foundation Discussion Paper*, No. 1650R.

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

Chen, X. and M. Reiß (2010) "On Rate Optimality for Ill-Posed Inverse Problems in Econometrics," *Econometric Theory*, forthcoming.

Chernozhukov, V., and C. Hansen (2005) "An IV Model of Quantile Treatment Effects. *Econometrica* 73, 245–261.

Chernozhukov, V., G.W. Imbens, and W.K. Newey (2007) "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics* 139, 4–14.

Darolles, S, Y. Fan, J.-P. Florens, and E. Renault (2011) "Nonparametric Instrumental Regression," *Econometrica,* forthcoming.

Engl, H. W., M. Hanke, and A. Neubauer (1996) *Regularization of Inverse Problems.* Dordrecht: Kluwer.

Escanciano, J.C. and S. Hoderlein (2010) "Nonparametric Identification of Euler Equations," working paper.

Fisher, F. (1966) *The Identification Problem in Econometrics,*New York: McGraw–Hill Book Co., 1966; Huntington: Robert Krieger Publishing Co., 1976.

Florens, J.-P., J. Johannes, and S. Van Bellegem (2008) "Instrumental Regression in Partially Linear Models," Working Paper, Toulouse School of Economics (GREMAQ).

Florens, J.-P., J. Johannes, and S. Van Bellegem (2010) "Identification and Estimation by Penalization in Nonparametric Instrumental Regression," *Econometric Theory,* forthcoming.

Florens, J.-P., and E. Sbai (2010) "Local Identification in Empirical Games of Incomplete Information," *Econometric Theory* 26, 1638 -1662.

Gallant, A.R. and G. Tauchen (1989) "Semiparametric Estimation of Conditional Constrained Heterogenous Processes: Asset Pricing Applications," *Econometrica* 57, 1091-1120.

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

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

Hansen, L.P. and K.J. Singleton. (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50, 1269-1286.

Hansen L.P., J. Heaton, J. Lee, N. Roussanov (2007) "Intertemporal Substitution and Risk Aversion," In: Heckman,J.J. and E.E. Leamer, eds., *Handbook of Econometrics,* Elsevier, Volume 6, 3967–4056.

D'Haultfoeuille, X., (2010) "On the Completeness Condition in Nonparametric instrumental regression," *Econometric Theory*, forthcoming.

Horowitz, J. L., and S. Lee (2007) "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica*, 75, 1191–1208.

Kress, R. (1999) *Linear Integral Equations,* New York: Springer-Verlag.

Lewbel, A. and O. Linton (2010) "Nonparametric Euler Equation Identification and Estimation," working paper.

Mattner, L., (1993) "Some Incomplete But Boundedly Complete Location Families,". *Annals of Statistics* 21, 2158–2162.

Newey, W.K. and J.L. Powell (1988) "Instrumental Variables Estimation for Nonparametric Models," Working Paper Version of Published Paper, MIT.

Newey, W.K. and J.L. Powell (2003) "Instrumental Variables Estimation for Nonparametric Models," *Econometrica* 71, 1565–1578.

Rothenberg, T.J., (1971) "Identification in Parametric Models," *Econometrica* 39, 577–591.

Santos, A. (2010) "Instrumental Variables Methods for Recovering Continuous Linear Functionals," UC San Diego working paper.

Sargan, D. (1983), "Identification and Lack of Identification," *Econometrica* 51, 1605-1634.

Stock, J. and J. Wright (2000), "GMM with Weak Identification," *Econometrica* 51, 1055-1096.

Department of Economics, Yale University

Department of Economics, Massachusetts Institute of Technology

Department of Economics, Seoul National University

Department of Economics, Massachusetts Institute of Technology