EXAMPLES OF L^2-COMPLETE AND BOUNDEDLY-COMPLETE DISTRIBUTIONS

By

Donald W. K. Andrews

May 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1801

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Examples of L^2-Complete and Boundedly-Complete Distributions

Donald W. K. Andrews*

Cowles Foundation for Research in Economics

Yale University

November 2010

Revised: May 2011

*The author gratefully acknowledges the research support of the National Science Foundation via grant number SES-0751517. The author thanks Xiaohong Chen, Jean-Pierre Florens, Phil Haile, Yingyao Hu, Whitney Newey, Alex Torgovitsky, and Daniel Wilhelm for very helpful comments.

Abstract

Completeness and bounded-completeness conditions are used increasingly in econometrics to obtain nonparametric identification in a variety of models from nonparametric instrumental variable regression to non-classical measurement error models. However, distributions that are known to be complete or boundedly complete are somewhat scarce.

In this paper, we consider an L^2-completeness condition that lies between completeness and bounded completeness. We construct broad (nonparametric) classes of distributions that are L^2-complete and boundedly complete. The distributions can have any marginal distributions and a wide range of strengths of dependence. Examples of L^2-incomplete distributions also are provided.

Keywords: Bivariate distribution, bounded completeness, canonical correlation, completeness, identification, measurement error, nonparametric instrumental variable regression.

JEL Classification Numbers: C14.

1 Introduction

Lehmann and Scheffé (1950, 1955) introduce the concept of completeness and use it to determine estimators with minimal risk in classes of unbiased estimators and to characterize tests that are similar. More recently, completeness and the weaker concept of bounded completeness have been used in the econometrics literature to obtain global and local identification conditions for a variety of nonparametric and semiparametric models. See the references below. In consequence, it is important to have available a broad array of distributions that are known to satisfy or fail these conditions.

A number of papers in the literature provide sufficient conditions for completeness and bounded completeness. Newey and Powell (2003) give a rank condition for completeness of distributions with finite support. Lehmann (1986) and Newey and Powell (2003) give sufficient conditions for parametric families in the exponential family, with the leading case being multivariate normal distributions. Ghosh and Singh (1966), Isenbeck and Rüschendorf (1992), and Mattner (1992) give conditions for location and scale families. Hu and Shiu (2011) provide some additional results.

d'Haultfoeuille (2011) provides sufficient conditions for bounded completeness for random vectors X and Z that satisfy $\mu(X) = \nu(Z) + \varepsilon$, where Z and ε are independent, $\nu(Z)$ is absolutely continuous with respect to (wrt) Lebesgue measure with full support R^{d_X}, and ε is absolutely continuous wrt Lebesgue measure with nowhere vanishing characteristic function. These are quite useful results but they do not allow for unbounded regression functions in the nonparametric instrumental variables (IV) regression model or non-classical measurement error in measurement error models, and the full support condition can be restrictive. d'Haultfoeuille (2011) also provides some sufficient conditions for completeness, but these conditions are quite restrictive.

In addition, several other papers in the literature provide examples of distributions that are boundedly complete, but not complete. These include Hoeffding (1977), Bar-Lev and Plachky (1989), and Mattner (1993). The boundedly complete distributions in these papers are restrictive and are not very suitable for typical econometric applications.

In this paper, we provide additional examples of distributions that satisfy completeness-type conditions and others that fail them. We consider the concept of L^2-completeness. This concept, or at least very closely related concepts, have been used by others, e.g., Florens, Mouchart, and Rolin (1990, Ch. 5), Isenbeck and Rüschendorf (1992), Mattner (1992, 1996), San Martin and Mouchart (2006), and Severini and Tripathi

(2006). Completeness and bounded completeness can be viewed as L^1-completeness and L^∞-completeness, respectively, so L^2-completeness lies between the two. It allows for unbounded regression functions in the nonparametric IV regression model and related semiparametric models, which are ruled out when the bounded completeness condition is used. The joint distribution of two random vectors X and Z is L^2-complete wrt X if and only if every non-constant square-integrable function of X is correlated with some square-integrable function of Z.

We construct distributions of (X, Z) that are L^2-complete or L^2-incomplete wrt X by starting with (i) any marginal distributions F_X and F_Z, respectively, (ii) two arbitrary sets of bounded orthonormal functions in $L^2(F_X)$ and $L^2(F_Z)$, and (iii) a sequence of constants $\{\tau_j : j \geq 1\}$. The constructed bivariate density is

$$k_\tau(x, z) = 1 + \sum_{j=1}^{r_Z} \tau_j x^{(j)}(x) z^{(j)}(z), \tag{1.1}$$

where the density is wrt the product of the marginal distributions $F_X \times F_Z$, $\{x^{(j)} : j = 0, ..., r_X\}$ is an orthonormal basis of $L^2(F_X)$ consisting of bounded functions, $x^{(0)}(x) = 1$ $\forall x$, and $\{z^{(j)} : j = 1, ..., r_Z\}$ is a set of bounded orthonormal functions in $L^2(F_Z)$. Under a condition on $\{\tau_j\}$, $k_\tau(x, z)$ is a proper density—it integrates to one and is non-negative. The resulting bivariate distribution is L^2-complete wrt X if $r_Z = r_X$ and $\tau_j \neq 0$ for all $j = 1, ..., r_Z$. Hence, one can construct easily a broad array of bivariate distributions that are L^2-complete and also a broad array that are L^2-incomplete. The method of construction employs the method used in a simple example of Lancaster (1958), which does not consider completeness.

If X and Z are absolutely continuous wrt Lebesgue measure, then the bivariate density k_τ wrt to the product of the marginals F_X and F_Z can be converted easily into a standard bivariate density wrt Lebesgue measure on $R^{d_X + d_Z}$, where d_X and d_Z denote the dimensions of X and Z, respectively.

Starting with Darolles, Florens, and Renault (2000), it is common in the nonparametric IV regression literature, to obtain identification as follows. Given the conditional distribution of X and Z, one defines the conditional expectation operator, say T, one obtains the singular value decomposition (SVD) of T using standard operator results, e.g., see Kress (1999, Sec. 15.4), and one assumes that the eigenvalues of T are all non-zero. The SVD yields a density of the form in (1.1).

The L^2-completeness results of this paper give a converse to this procedure. Starting with orthonormal functions $\{x^{(j)}\}$ and $\{z^{(j)}\}$ and constants $\{\tau_j\}$, one can define a function $k_\tau(x, z)$ as in (1.1). But, such a function is not necessarily a density because the orthonormal functions $\{x^{(j)}\}$ and $\{z^{(j)}\}$ take on positive and negative values and, hence, $k_\tau(x, z)$ can be negative. The contribution of this paper is to provide a simple set of sufficient conditions to guarantee that $k_\tau(x, z)$ is a proper density. The conditions given are sufficiently weak that one can construct a broad (i.e., nonparametric) class of distributions that are L^2-complete. In a certain sense, the distributions that are L^2-complete wrt X are generic in the class of distributions that are constructed. (The sense considered follows the concepts of shyness and prevalence introduced in Christensen (1974), Hunt, Sauer, and Yorke (1992), and Anderson and Zame (2001).) Nevertheless, one also can construct many L^2-incomplete distributions.

We now briefly discuss the use of completeness conditions in the econometrics literature. Completeness, L^2-completeness, and bounded completeness conditions can be used to obtain global or local identification in a variety of models. These models include: (i) the nonparametric IV regression model, see Newey and Powell (2003), Darolles, Florens, and Renault (2000), Hall and Horowitz (2005), and references in Horowitz (2010), (ii) semiparametric IV models, see Ai and Chen (2003), Blundell, Chen, and Kristensen (2007), and Chen and Pouzo (2009a), (iii) nonparametric IV quantile models, see Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007), Horowitz and Lee (2007), Chen and Pouzo (2009b), and Chen, Chernozhukov, Lee, and Newey (2010), (iv) measurement error models, see Bissantz, Hohage, Munk, and Ruymgaart (2007), Hu and Schennach (2008), Carroll, Chen, and Hu (2009), An and Hu (2009), Song (2011), and Wilhelm (2011), (v) demand models, see Berry and Haile (2009a, 2010), (vi) dynamic optimization models, see Hu and Shum (2009), (vii) generalized regression models with group effects, see Berry and Haile (2009a), (viii) asset pricing models, see Chen and Ludvigson (2009), and (ix) missing data models, see Sasaki (2011).

The remainder of the paper is organized as follows. Section 2 discusses the L^2-completeness condition. Section 3 gives classes of bivariate distributions of random variables and vectors that are L^2-complete and others that are L^2-incomplete. Section 4 provides proofs.

2 L^2-Completeness

In this section, we define the concept of L^2-completeness, which is very closely related to the well-known concepts of completeness and bounded completeness, see Lehmann (1986, p. 173). In consequence, L^2-completeness can be used to give conditions for nonparametric identification in a variety of models. L^2-completeness is not original to this paper.[1]

Let X and Z be random elements that take values in complete separable metric spaces $\mathcal{X}$ and $\mathcal{Z}$, respectively, and are defined on the same probability space. In applications, X and Z typically are random variables or vectors, possibly of different dimensions, but they could be stochastic processes. We say that a bivariate distribution F_{XZ} of random elements X and Z is L^2-*complete* wrt X if $\forall h \in L^2(F_X)$,

$$E(h(X)|Z) = 0 \text{ a.s.}[F_Z] \text{ implies that } h(X) = 0 \text{ a.s.}[F_X], \tag{2.1}$$

where the expectation is taken under F_{XZ}.[2,3,4] In words, L^2-*completeness* means that if $h \in L^2(F_X)$ has conditional mean zero given Z, then h equals zero a.s. In contrast to the well-known conditions of completeness and bounded completeness, the family of functions h considered here is $L^2(F_X)$, rather than $L^1(F_X)$ or $L^\infty(F_X)$, respectively. Although L^2-completeness imposes a second moment condition, rather than the weaker first moment condition imposed by a completeness condition, it is still useful in most applications for which unbounded functions h are of interest.

[1] See the Introduction for references to its use in the literature.

[2] The L^2-completeness of a bivariate distribution F_{XZ} wrt X depends on F_{XZ} through the conditional distribution of X given Z, the marginal distribution of X (because $L^2(F_X)$ enters the definition), and the support of Z (because a.s.$[F_Z]$ enters the definition). The marginal distribution of Z only affects L^2-completeness through its support.

[3] An L^2 version of Oosterhoff and Schriever's (1987) definition of $\mathcal{P}^*$ completeness (which is an L^1 definition) is almost the same as the definition used here.

[4] One can give a closely-related definition of L^2-completeness that is more akin to the definition of completeness given in Lehmann and Scheffé (1950). One can define a family of distributions $\mathcal{F} \in \{F_{X,\theta} : \theta \in \Theta\}$ of the random vector X to be L^2-*complete* if $\forall h \in L^2(F_X)$,

$$Eh(X) = 0 \ \forall F_{X,\theta} \in \mathcal{F} \text{ implies that } h(X) = 0 \text{ a.s.}[F_{X,\theta}] \ \forall \theta \in \Theta, \tag{2.2}$$

where the expectation of X is taken under $F_{X,\theta}$. Here, θ is a fixed parameter and its parameter space is Θ, which play the role of z and the support of F_Z, respectively, in (2.1), and $F_{X,\theta}$ is the distribution of X, which plays the role of the conditional distribution of X given $Z = z$ in (2.1).

For the purposes of identification in nonparametric models, the definition of L^2-completeness of the bivariate distribution F_{XZ} wrt X given in (2.1) is more convenient than the definition given in (2.2).

For example, consider the nonparametric IV regression model with regressor X, IV Z, and regression function $h(X)$. The use of $L^2(F_X)$-completeness wrt X to identify h, rather than completeness, imposes the restriction $h \in L^2(F_X)$, rather than $h \in L^1(F_X)$, but allows for a much broader class of joint distributions of X and Z. Unlike the bounded completeness condition, $L^2(F_X)$-completeness does not require that h is bounded, which can be restrictive.

The definition above can be weakened to *bounded completeness* of the bivariate distribution F_{XZ} wrt X by requiring the function h in the definition to be a bounded function. Obviously, L^2-completeness of F_{XZ} wrt X implies bounded completeness of F_{XZ} wrt X.

We say that a random variable is *non-constant* if its distribution is not a point mass distribution.

A simple and intuitive characterization of L^2-completeness is the following result, which is a slightly different statement of Lemma 2.1 of Severini and Tripathi (2006):

Proposition 1. *F_{XZ} is L^2-complete wrt X if and only if every non-constant rv $\lambda(X) \in L^2(F_X)$ is correlated with some rv $\phi(Z) \in L^2(F_Z)$.*[5]

3 Examples of $\mathbf{L^2}$-Complete Distributions

In this section, we construct bivariate distributions $F_{XZ,\tau}$ that are L^2-complete and others that are L^2-incomplete. The marginal distributions can be any distributions F_X and F_Z of interest.

3.1 Bivariate Distributions $\mathbf{F_{XZ,\tau}}$ of Random Elements X and Z

Given any marginal distributions F_X and F_Z, we construct a distribution $F_{XZ,\tau}$ by specifying its density $k_\tau(x,z)$ wrt the product of its marginal distributions $F_X \times F_Z$. To do so, we use the following assumptions.

[5]By definition, $h \in L^2(F_X)$ means $\int h^2(x)dF_X(x) < \infty$. For convenience, but with some abuse of notation, we let $h(X) \in L^2(F_X)$ mean that the random variable $h(X)$ satisfies $Eh^2(X) < \infty$ when $X \sim F_X$. Thus, $h \in L^2(F_X)$ and $h(X) \in L^2(F_X)$ are equivalent. This notation also is used for functions $\phi(Z)$ of $Z \sim F_Z$.

Assumption 1. $\{x^{(j)} : 0 \leq j \leq r_X\}$ is an orthonormal basis of $L^2(F_X)$ with $x^{(0)}(x) = 1$ $\forall x \in \mathcal{X}$ and $\{z^{(j)} : 0 \leq j \leq r_Z\}$ is a set of orthonormal functions in $L^2(F_Z)$ with $z^{(0)}(z) = 1$ $\forall z \in \mathcal{Z}$, where $0 \leq r_X, r_Z \leq \infty$.

Assumption 2. The functions $\{x^{(j)}\}$ and $\{z^{(j)}\}$ are bounded in absolute value on the supports of F_X and F_Z, respectively, with bounds $\{B_{X,j} : 0 \leq j \leq r_X\}$ and $\{B_{Z,j} : 0 \leq j \leq r_Z\}$.

Assumption 3. $\{\tau_j : j = 1, ..., r_Z\}$ is a sequence of constants that satisfies $\sum_{j=1}^{r_Z} |\tau_j| B_{X,j} B_{Z,j} \leq 1$, where $0 \leq r_Z \leq r_X$.

Assumptions 1 and 2 hold for a wide variety of functions. Some examples are given below. However, Assumption 2 does rule out some orthonormal functions, such as the Hermite polynomials on R or R^d, which appear in an orthonormal expansion of the bivariate normal distribution, see Lancaster (1957). Note that r_X is finite only if the support of F_X is finite. When $r_Z = \infty$, Assumption 3 holds for sequences that converge to zero arbitrarily quickly, as well as those that converge as slowly as $|\tau_j|^{-(1+\delta)}$ for any $\delta > 0$ when $\sup_{j \geq 1}(B_{X,j} B_{Z,j}) < \infty$.

Define the density $k_\tau(x, z)$ by

$$k_\tau(x, z) = 1 + \sum_{j=1}^{r_Z} \tau_j x^{(j)}(x) z^{(j)}(z). \tag{3.1}$$

Theorem 1 below shows that $k_\tau(x, z)$ is a density function wrt $F_X \times F_Z$ for any choice of functions $\{x^{(j)}\}$ and $\{z^{(j)}\}$ and any constants $\{\tau_j\}$ that satisfy Assumptions 1-3. In particular, Assumption 1 guarantees that $k_\tau(x, z)$ integrates to one and Assumptions 2 and 3 ensure that $k_\tau(x, z)$ is non-negative. Let $F_{XZ,\tau}$ denote the bivariate distribution of X and Z that corresponds to the density $k_\tau(x, z)$ and the marginal distributions F_X and F_Z.

Example 1. Now we illustrate functions $\{x^{(j)}\}$ and $\{z^{(j)}\}$ that satisfy Assumptions 1 and 2 in the case of absolutely continuous random vectors X and Z with dimensions d_X and d_Z, respectively. Consider bounded functions $\{u^{(j)} : j = 0, 1, ..., r_X\}$ and $\{v^{(j)} : j = 0, 1, ..., r_Z\}$ on $[0, 1]^{d_X}$ and $[0, 1]^{d_Z}$, respectively, that are orthonormal wrt Lebesgue measure, have $u^{(0)}(x) = v^{(0)}(z) = 1$ $\forall x \in [0, 1]^{d_X}$, $z \in [0, 1]^{d_Z}$, and for which $\{u^{(j)} : j = 0, 1, ..., r_X\}$ is a basis of the set of L^2 functions (wrt Lebesgue measure) on $[0, 1]^{d_X}$. Products of trigonometric functions (each scaled to lie in $[0, 1]^{d_m}$, rather than $[0, 2\pi]^{d_m}$, for $m = X, Z$) provide one example. Products of shifted Legendre polynomials provide another example. In fact, any countably dense sets of bounded functions on $[0,1]^{d_X}$ and $[0,1]^{d_Z}$ that are orthonormalized, e.g., by the Gramm-Schmidt process, yield other examples. The type of orthonormal functions for X, e.g., trigonometric functions, can be different from the orthonormal functions for Z, e.g., shifted Legendre polynomials. Shuffling the orders of the functions $\{u^{(j)}\}$ and $\{v^{(j)}\}$ (so that different $u^{(j)}$ functions match up with different $v^{(j)}$ functions) provides additional examples.

Then, the functions

$$\{x^{(j)} = u^{(j)} \circ F_X : j = 0, 1, ...\} \text{ and } \{z^{(j)} = v^{(j)} \circ F_Z : j = 0, 1, ..., r_Z\} \tag{3.2}$$

satisfy Assumptions 1 and 2 with $r_X = \infty$.[6] $\square$

Example 2. Suppose F_X and F_Z are uniform on $[0,1]^{d_X}$ and $[0,1]^{d_Z}$, respectively. Then, the density $k_\tau(x,z)$ of $F_{XZ,\tau}$ is a copula density on $[0,1]^{d_X+d_Z}$, which we denote by $c_\tau(x,z)$. Using the functions $\{u^{(j)}\}$ and $\{v^{(j)}\}$ defined in Example 1, the following function is a copula density provided Assumption 3 holds:

$$c_\tau(x,z) = 1 + \sum_{j=1}^{r_Z} \tau_j u^{(j)}(x) v^{(j)}(z). \tag{3.3}$$

$\square$

Example 3. As above, suppose F_X and F_Z are absolutely continuous with densities f_X and f_Z wrt Lebesgue measure on $[0,1]^{d_X}$ and $[0,1]^{d_Z}$, respectively. Then, the following functions, k_τ and f_τ, are proper densities wrt $F_X \times F_Z$ and wrt Lebesgue measure on $R^{d_X+d_Z}$, respectively, of a bivariate distribution $F_{XZ,\tau}$:

$$\begin{aligned} k_\tau(x,z) &= c_\tau(F_X(x), F_Z(z)) \text{ and} \\ f_\tau(x,z) &= c_\tau(F_X(x), F_Z(z)) f_X(x) f_Z(z) \end{aligned} \tag{3.4}$$

provided Assumption 3 holds. Given any copula density c_τ as in (3.3) one obtains bivariate distributions $F_{XZ,\tau}$ with any absolutely continuous marginal distributions F_X and F_Z that are desired. $\square$

[6]By definition, $(u^{(j)} \circ F_X)(x) = u^{(j)}(F_X(x))$ for $x \in R^{d_X}$ and likewise for $v^{(j)} \circ F_Z$.

3.2 $\mathbf{L^2}$-Completeness of $\mathbf{F_{XZ,\tau}}$

The L^2-completeness of $F_{XZ,\tau}$ wrt X depends on whether the following Assumption holds or not.

Assumption 4. (i) $\tau_j \neq 0 \; \forall j = 1, ..., r_Z$ and (ii) $r_Z = r_X$.

If Assumption 4 holds, then every basis function $x^{(j)}$ for $j = 0, ..., r_X$ enters the density $k_\tau(x, z)$ with a non-zero coefficient. Given this, one can show that every non-constant function $\lambda(X) \in L^2(F_X)$ is correlated with some function $\phi(Z) \in L^2(F_Z)$, see Theorem 1 below.

If the support of X is finite, then Assumption 4(ii) requires that the number of points in the support of Z is greater than or equal to the number in the support of X. If the support of X is infinite, then Assumption 4 requires $r_Z = \infty$ and $\tau_j \neq 0 \; \forall j \geq 1$. But, Assumption 4 does not require that $\{z^{(j)} : 0 \leq j \leq r_Z\}$ is an orthonormal basis of $L^2(F_Z)$. For example, $\{z^{(j)} : 0 \leq j \leq r_Z\}$ could consist of the odd-numbered terms of some orthonormal basis of $L^2(F_Z)$.

The following Theorem is the main result of the paper.

Theorem 1. *Suppose Assumptions* 1-3 *hold. Then,*

(a) $k_\tau(x, z)$ *is a proper bivariate density function wrt* $F_X \times F_Z$ *and*

(b) *the bivariate distribution* $F_{XZ,\tau}$ *defined by the density* k_τ *is* L^2*-complete wrt* X *if and only if Assumption* 4 *holds.*

Comments. 1. Given any marginal distributions F_X and F_Z, consider the class of bivariate distributions with densities k_τ (wrt $F_X \times F_Z$) of the form in (3.1) that is generated by a fixed choice of orthonormal functions $\{x^{(j)}\}$ and $\{z^{(j)}\}$ that satisfy Assumptions 1 and 2 and all sequences of constants $\{\tau_j : j = 1, ..., r_Z\}$ that satisfy Assumptions 3 and 4. This is a nonparametric (i.e., infinite-dimensional) class of L^2-complete distributions wrt X when $r_Z = \infty$. Re-ordering the orthonormal functions $\{x^{(j)}\}$ and $\{z^{(j)}\}$ leads to additional nonparametric classes of L^2-complete distributions. Different orthonormal functions $\{x^{(j)}\}$ and $\{z^{(j)}\}$ lead to additional nonparametric classes of L^2-complete distributions. Taking unions of the preceding classes of L^2-complete distributions over different marginal distributions leads to larger nonparametric classes of L^2-complete distributions.

2. The question naturally arises: How many bivariate distributions F_{XZ} can be written in the form of $F_{XZ,\tau}$? Results of Lancaster (1958, 1963) for bivariate distributions and results of Darolles, Florens, and Renault (2000) based on the singular value decomposition of the conditional expectation operator (see Kress (1998, Sec. 15.4)) show that the answer is that many are of this form. Let $F_{XZ} << F_X \times F_Z$ denote that F_{XZ} is absolutely continuous wrt $F_X \times F_Z$. Consider the following assumption:

Assumption A. (i) $F_{XZ} << F_X \times F_Z$ and (ii) $k \in L^2(F_X \times F_Z)$, where k is the Radon-Nykodym derivative of F_{XZ} wrt $F_X \times F_Z$.

Assumption A(i) rules out joint distributions of (X, Z) for which X is a deterministic function of Z and vice versa. In econometric applications of completeness or L^2-completeness, this usually is not restrictive. Note that $F_{XZ,\tau}$ and k_τ satisfy Assumption A under Assumptions 1-3.[7] The references immediately above show that any bivariate distribution F_{XZ} that satisfies Assumption A has a density k wrt $F_X \times F_Z$ of the form in (3.1) and Assumption 1 holds.

Theorem 1(a) is a partial converse to these results. Theorem 1(a) says that given suitable orthonormal functions and some conditions on the constants $\{\tau_j\}$ one obtains a proper bivariate distribution.

3. Assumptions 2 and 3 in Theorem 1 can be replaced by the more general, but less easily verified, condition:

Assumption 2*. (i) $k_\tau(x, z) \geq 0$ a.s.$[F_X \times F_Z]$ and (ii) $k_\tau \in L^2(F_X \times F_Z)$.

Several bivariate distributions in the literature have been shown to satisfy Assumptions 1 and 2*, but not 2 and 3, including the bivariate normal, gamma, Poisson, binomial, hypergeometric, and negative binomial, see Campbell (1934), Aitken and Gonin (1935), Kibble (1941), Eagleson (1964), and Hamdan and Al-Bayyani (1971).[8] In all cases, Assumption 4 holds with $r_X = r_Z = \infty$, so the distributions are L^2-complete wrt both X and Z by Theorem 1.[9]

4. Using the canoncial correlation representation of Lancaster (1958, 1963), it can be shown that when (X, Z) has density k_τ, as in (3.1), then $x^{(1)}(X)$ and $z^{(1)}(Z)$ are the mean-zero variance-one functions of X and Z, respectively, that maximize the correlation

[7] This holds because $k_\tau^2(x, z) = (1+\sum_{j=1}^{r_Z} \tau_j x^{(j)} z^{(j)})^2 \leq (1+\sum_{j=1}^{r_Z} |\tau_j| B_{X,j} B_{Z,j})^2 \leq 4$ by Assumption 3.

[8] I thank Daniel Wilhelm for references.

[9] One cannot use the bivariate density expansions just listed to obtain nonparametric classes of L^2-complete distributions just by perturbing the coefficients $\{\tau_j : j \geq 1\}$ in these expansions. The reason is that the resulting functions are not necessarily non-negative. Note that the basis functions $\{x^{(j)} : j \geq 1\}$ and $\{z^{(j)} : j \geq 1\}$ necessarily take negative values because they integrate to zero wrt F_X and F_Z, respectively.

between X and Z. In addition, by direct calculation, $\tau_1 = Corr(x^{(1)}(X), z^{(1)}(Z))$. Furthermore, for $j = 2, ..., r_Z$, $x^{(j)}(X)$ and $z^{(j)}(Z)$ are the mean-zero variance-one functions of X and Z that maximize the correlation between X and Z subject to being uncorrelated with $\{x^{(j)}(X) : m = 1, ..., j-1\}$ and $\{z^{(m)}(Z) : m = 1, ..., j-1\}$, respectively. Also, $\tau_j = Corr(x^{(j)}(X), z^{(j)}(Z))$.

5. It is sometimes of interest to view X as a function of Z and some unobservable V. Suppose (X, Z) have a joint df $F_{XZ,\tau}$ as in Theorem 1, X is a scalar random variable, and Z is a random vector (or a random element). Then, one can generate X via the equation

$$X = h(Z, V), \tag{3.5}$$

where Z and V are independent random variables, $Z \sim F_Z$, $V \sim F_V$ for any distribution F_V that is absolutely continuous wrt Lebesgue measure on R, and $h(z, v)$ is a suitably chosen function.[10] Hence, if $F_{XZ,\tau}$ satisfies Assumption 4, then (X, Z) generated as in (3.5) are L^2-complete wrt X but otherwise are not.

6. Suppose $r_X = r_Z = \infty$. Consider the space of ℓ^p sequences for some $1 \leq p \leq \infty$ that satisfy Assumption 3:

$$C_B = \left\{ \{\tau_j : j \geq 1\} \in \ell^p : \sum_{j=1}^{\infty} |\tau_j| B_{X,j} B_{Z,j} \leq 1 \right\}. \tag{3.6}$$

One can ask: for sequences in C_B, how generic is the property $\tau_j \neq 0 \ \forall j \geq 1$? That is, how generic is the completeness property specified by Assumption 4? For finite-dimensional spaces, a property often is said to be generic if the set of points for which the property fails has Lebesgue measure zero. In infinite-dimensional spaces, such as C_B, the concept of genericity is more complicated. Topological notions of genericity often are too weak, see Anderson and Zame (2001) and Stinchcombe (2002). Measure-theoretic notions are more useful. Christensen (1974) and Hunt, Sauer, and Yorke (1992) (independently) develop a measure-theoretic notion of "genericity" for vector

[10] Let $F_{X|Z,\tau}(x|z)$ denote the conditional distribution of X given $Z = z$ under $F_{XZ,\tau}$. Let $h^*(z, u) = F^{-1}_{X|Z,\tau}(u|z)$ for $u \in [0, 1]$, where $F^{-1}_{X|Z,\tau}(u|z)$ is the u-th quantile of $F_{X|Z,\tau}(\cdot|z)$. Let $h(z, v) = h^*(z, F_V(v))$. Let $X = h(Z, V)$. We claim that $(X, Z) \sim F_{XZ,\tau}$. Under the assumptions, $U = F_V(V) \sim U[0, 1]$. We have $X = h^*(Z, U) = F^{-1}_{X|Z,\tau}(U|Z)$. The conditional distribution of X given $Z = z$ is the distribution of $F^{-1}_{X|Z,\tau}(U|z)$. But, this is the conditional distribution $F_{X|Z,\tau}(\cdot|z)$ as desired, because for any distribution F, $\widetilde{X} = F^{-1}(U) \sim F$. This holds whether or not $F_{X|Z,\tau}(\cdot|z)$ is a continuous conditional distribution.

spaces that the latter authors call *prevalence.* A set is prevalent if its complement is *shy.* The shyness of a set is a natural extension to an infinite-dimensional vector space of a set having Lebesgue measure zero in a finite-dimensional space.

The set C_B is not a vector space. In fact, it is a shy set in the vector space ℓ^p. Thus, the definition of shyness and prevalence in Christensen (1974) and Hunt, Sauer, and Yorke (1992) cannot be applied here. However, the same issue that the space of interest is not a vector space also arises in various areas of economic theory where it is natural to ask whether a property is generic. In consequence, Anderson and Zame (2001) have extended the concept of shyness and prevalence to convex subsets of vector spaces. The set C_B is convex and hence their definition is applicable here.

Their definition is as follows. Let X be a topological vector space and let $C \subset X$ be a subset that is completely metrizable in the relative topology induced from X. Let $c \in C$. A set $E \subset C$ which is universally measurable in X is said to be *shy in C at c* if for each $\delta > 0$ and each neighborhood W of 0 in X, there is a regular Borel probability measure μ on X with compact support such that $supp(\mu) \subset [\delta(C-c)+c] \cap (W+c)$ and $\mu(E+x) = 0\ \forall x \in X$. By definition, E is *shy in C* if it is shy in C at c for all $c \in C$. An arbitrary subset $F \subset C$ is *shy in C* if it is contained in a shy universally measurable set. A subset $S \subset C$ is *prevalent in C* if its complement $C \backslash S$ is shy in C. Anderson and Zame (2001, p. 12) show that if E is shy at some $c \in C$ then it is shy at every $c \in C$ and hence is shy at C.

See Hunt, Sauer, and Yorke (1992) and Anderson and Zame (2001) for discussions of why the concept of shyness is a suitable extension to infinite-dimensional spaces of a set (in a finite-dimensional space) having Lebesgue measure zero. The key is that a set E in R^k is shy if and only if it has Lebesgue measure zero. See Hunt, Sauer, and Yorke (1992, p. 219).

We have the following genericity result for Assumption 4.

Lemma 1. Suppose $r_X = r_Z$. The set of sequences $S = \{\{\tau_j\} \in C_B : \tau_j \neq 0\ \forall j \geq 1\}$ is a prevalent subset of C_B.

7. Because the definition of prevalence is somewhat complicated, we give an alternative genericity result here. Consider the space of sequences $S_\tau = \{\{\tau_j : j \geq 1\} : |\tau_j| \leq DB_{X,j}^{-1} B_{Z,j}^{-1} j^{-1-\delta}\ \forall j \geq 1\}$ for some $\delta > 0$ and $D = (\sum_{j=1}^{\infty} j^{-1-\delta})^{-1}$. Such sequences all satisfy Assumption 3. How generic is the property $\tau_j \neq 0\ \forall j \geq 1$? If one considers a property to be generic if the μ-measure of the set for which the property fails is zero for some measure μ, then the property $\tau_j \neq 0\ \forall j \geq 1$ is generic for any measure μ on

S_τ (coupled with some σ-field $\mathcal{F}_{S_\tau}$) for which the induced measure on any set of finite subsequences is absolutely continuous wrt Lebesgue measure.[11]

8. Given any marginal distributions F_X and F_Z, consider the class of bivariate distributions with densities k_τ (wrt $F_X \times F_Z$) of the form in (3.1) that is generated by a fixed choice of orthonormal functions $\{x^{(j)}\}$ and $\{z^{(j)}\}$ that satisfy Assumptions 1 and 2 and all sequences of constants $\{\tau_j : j = 1, ..., r_Z\}$ that satisfy Assumption 3 for some fixed constants $\{B_{X,j} : 1 \leq j \leq r_X\}$ and $\{B_{Z,j} : 1 \leq j \leq r_Z\}$. The set of incomplete distributions in this class (i.e., those that fail Assumption 4) is a dense subset (under the $L^2(F_X \times F_Z)$ metric).

See Santos (2009, Lemma 2.1) for a related L^∞-denseness result for incomplete distributions (roughly speaking) in the class of distributions with compact support and smooth density functions wrt Lebesgue measure.

4 Proofs

4.1 Proof of Propositon 1

The proof of Proposition 1 uses the following Lemma.

Lemma 2. For any non-constant $\lambda(X) \in L^2(F_X)$ with $E\lambda(X) = 0$,

$$E(\lambda(X)|Z) = 0 \text{ a.s.}[F_Z] \text{ iff } Corr(\lambda(X), \phi(Z)) = 0 \text{ for all non-constant } \phi(Z) \in L^2(F_Z).$$

Proof of Lemma 2. Let $\sigma_\lambda^2 = Var(\lambda(X)) > 0$ and $\sigma_\phi^2 = Var(\phi(Z)) > 0$. We have

$$Corr(\lambda(X), \phi(Z)) = E\lambda(X)\phi(Z)/(\sigma_\lambda\sigma_\phi) = E[E(\lambda(X)|Z)\phi(Z)]/(\sigma_\lambda\sigma_\phi), \tag{4.1}$$

where the first equality uses $E\lambda(X) = 0$ and the second holds by iterated expectations.

If $E(\lambda(X)|Z) = 0$ a.s.$[F_Z]$, then the right-hand side of (4.1) equals zero, which establishes the "only if" statement of the Lemma.

[11] That is, the condition on μ is that $\mu_{\{j_1,...,j_K\}}$ is absolutely continuous wrt to Lebesgue measure on R^K for any non-redundant finite positive integers $\{j_1, ..., j_K\}$, where $\mu_{\{j_1,...,j_K\}}$ is the measure defined by $\mu_{\{j_1,...,j_K\}}(\{\{\tau_{j_k} : k = 1, ..., K\} : |\tau_{j_k}| \leq CB_{X,j_k}^{-1} B_{Z,j_k}^{-1} j_k^{-1-\delta} \text{ for } k = 1, ..., K\}) = \mu(\{\{\tau_j : j \geq 1\} : |\tau_{j_k}| \leq CB_{X,j_k}^{-1} B_{Z,j_k}^{-1} j_k^{-1-\delta} \text{ for } k = 1, ..., K\}).)$

To prove the "if" statement, take $\phi(Z) = E(\lambda(X)|Z)$ in (4.1) to obtain

$$Corr(\lambda(X), \phi(Z)) = E[E(\lambda(X)|Z)]^2/(\sigma_\lambda \sigma_\phi). \tag{4.2}$$

Then, $Corr(\lambda(X), \phi(Z)) = 0$ implies $E(\lambda(X)|Z) = 0$ a.s.$[F_Z]$ and the proof is complete. $\square$

Proof of Proposition 1. The following are equivalent:

1. Every non-constant rv $\lambda(X) \in L^2(F_X)$ is correlated with some rv $\phi(Z) \in L^2(F_Z)$.
2. Every mean zero, non-constant rv $\lambda(X) \in L^2(F_X)$ is correlated with some rv $\phi(Z) \in L^2(F_Z)$.
3. For every mean zero, non-constant rv $\lambda(X) \in L^2(F_X)$, $E(\lambda(X)|Z) = 0$ a.s.$[F_Z]$ fails to hold.
4. For every mean zero rv $\lambda(X) \in L^2(F_X)$, if $\lambda(X) = 0$ a.s.$[F_X]$ fails to hold, then $E(\lambda(X)|Z) = 0$ a.s.$[F_Z]$ fails to hold.
5. If $h(X) \in L^2(F_X)$ and $E(h(X)|Z) = 0$ a.s.$[F_Z]$, then $h(X) = 0$ a.s.$[F_X]$.
6. F_{XZ} is L^2-complete wrt X.

The equivalences of 1 and 2, 3 and 4, and 4 and 5 are straightforward. The equivalence of 2 and 3 holds by Lemma 2. The equivalence of 5 and 6 holds by the definition of L^2-completeness. $\square$

4.2 Proof of Theorem 1

First, we provide some useful expressions for $h \in L^2(F_X)$ and $E(h(X)|Z = z)$ when $(X, Z) \sim F_{XZ,\tau}$. These results are used in the proof of Theorem 1. Define the inner products $\langle \cdot, \cdot \rangle_{F_X}$ and $\langle \cdot, \cdot \rangle_{F_Z}$ by

$$\langle h_1, h_2 \rangle_{F_X} = \int h_1(x) h_2(x) dF_X(x) \text{ and } \langle m_1, m_2 \rangle_{F_Z} = \int m_1(z) m_2(z) dF_Z(z) \tag{4.3}$$

for $h_1, h_2 \in L^2(F_X)$ and $m_1, m_2 \in L^2(F_Z)$. Note that $\langle h_1, x^{(j)} \rangle_{F_X} = Cov(h_1(X), x^{(j)}(X))$ and $\langle m_1, z^{(j)} \rangle_{F_X} = Cov(m_1(Z), z^{(j)}(Z))$ for any functions $x^{(j)}$ and $z^{(j)}$ as in Assumption 1 and (3.1) for $j = 1, ..., r_Z$.[12]

[12] This holds because $x^{(0)}(x) = z^{(0)}(z) = 1$ $\forall x \in \mathcal{X}$, $\forall z \in \mathcal{Z}$ implies that $E_{F_X} x^{(j)}(X) = 0$ and $E_{F_Z} z^{(j)}(Z) = 0$ $\forall j \geq 1$.

For $h \in L^2(F_X)$, let

$$h_Z(z) = E(h(X)|Z = z). \tag{4.4}$$

Define $\tau_0 = 1$. Let $S_{X,\tau}$ denote the linear subspace of $L^2(F_X)$ that is generated by the functions $\{x^{(j)}1(\tau_j \neq 0) : j = 0, ..., r_Z\}$. Let $S_{X,\tau}^{\perp}$ denote the orthogonal complement of $S_{X,\tau}$ in $L^2(F_X)$.

Lemma 3. *Suppose $F_{XZ,\tau}$ satisfies Assumptions* 1-3. *Then,*
(a) for $h \in L^2(F_X)$,

$$h_Z(z) = \sum_{j=0}^{r_Z} \tau_j \langle h, x^{(j)} \rangle_{F_X} z^{(j)}(z) \text{ a.s. } [F_Z],$$

(b) for $h \in L^2(F_X)$ *and* $j = 0, ..., r_Z$ *with* $\tau_j \neq 0$,

$$\langle h, x^{(j)} \rangle_{F_X} = \tau_j^{-1} \langle h_Z, z^{(j)} \rangle_{F_Z},$$

(c) *if* $F_{XZ,\tau}$ *satisfies Assumption* 4, *then for* $h \in L^2(F_X)$,

$$h(x) = \sum_{j=0}^{r_Z} \tau_j^{-1} \langle h_Z, z^{(j)} \rangle_{F_Z} x^{(j)}(x) \text{ a.s.}[F_X], \textit{ and}$$

(d) *for* $h \in S_{X,\tau}^{\perp}$, $h_Z(z) = 0$ a.s.$[F_Z]$.

Comment. Lemma 3(a) provides an expression for the conditional mean of a function in terms of the function itself and the orthogonal functions and constants $\{\tau_j\}$ of $F_{XZ,\tau}$. Lemma 3(b) provides an expression for certain weighted averages of a function h in terms of its conditional mean h_Z and the orthogonal functions and constants $\{\tau_j\}$ of $F_{XZ,\tau}$. Lemma 3(c) provides an expression for a function h in terms of its conditional mean h_Z and the orthogonal functions and constants $\{\tau_j\}$ of $F_{XZ,\tau}$ that holds when Assumption 4 holds. Lemma 3(d) shows that the conditional mean given Z of a function in $S_{X,\tau}^{\perp}$ is zero.

Proof of Theorem 1. We have $k_\tau \in L^2(F_X \times F_Z)$ because $k_\tau^2(x,z) = (1 + \sum_{j=1}^{r_Z} \tau_j x^{(j)}(x) z^{(j)}(z))^2 \leq (1 + \sum_{j=1}^{r_Z} |\tau_j| B_{X,j} B_{Z,j})^2 \leq 4 \; \forall x \in \mathcal{X}, z \in \mathcal{Z}$ by Assumption 3. Let $\langle \cdot, \cdot \rangle_{F_X \times F_Z}$ denote the inner product on $L^2(F_X \times F_Z)$.

Now we apply the Parseval-Bessel equality, e.g., see Dudley (1989, Thm. 5.4.4), to show that the density k_τ integrates to one: $\langle k_\tau, 1 \rangle_{F_X \times F_Z} = \int \int k_\tau(x,z) dF_X(x) dF_Z(z) =$

1. The Parseval-Bessel equality says: If $\{e_\alpha\}$ is an orthonormal set in a Hilbert space H over the real numbers, $x \in H$, $y \in H$, $x = \sum_{\alpha \in I} x_\alpha e_\alpha$ for scalars $\{x_\alpha\}$, and $y = \sum_{\alpha \in I} y_\alpha e_\alpha$ for scalars $\{y_\alpha\}$, then $\langle x, y \rangle = \sum_{\alpha \in I} x_\alpha y_\alpha$, where $\langle \cdot, \cdot \rangle$ is the inner product on H.

We apply this result with (i) $H = L^2(F_X \times F_Z)$, (ii) $\{e_\alpha\}$ equal to the functions $\{x^{(j)} z^{(j)} : j = 0, ..., r_Z\}$, which are orthonormal in $L^2(F_X \times F_Z)$, (iii) $x = k_\tau = \sum_{j=0}^{r_Z} \tau_j x^{(j)} z^{(j)}$ $(= 1 + \sum_{j=1}^{r_Z} \tau_j x^{(j)} z^{(j)})$, and (iv) $y = 1 = \sum_{j=0}^{r_Z} \tau_j^* x^{(j)} z^{(j)}$, where $\tau_0^* = 1$ and $\tau_j^* = 0$ for $j = 1, ..., r_Z$. This yields

$$\int \int k_\tau(x, z) dF_X(x) dF_Z(z) = \langle k_\tau, 1 \rangle_{F_X \times F_Z} = \sum_{j=0}^{r_Z} \tau_j \tau_j^* = 1. \tag{4.5}$$

Next, we have

$$k_\tau(x, z) = 1 + \sum_{j=1}^{r_Z} \tau_j x^{(j)}(x) z^{(j)}(z) \geq 1 - \sum_{j=1}^{r_Z} \tau_j B_{X,j} B_{Z,j} \geq 0 \tag{4.6}$$

for all x and z in the supports of F_X and F_Z, respectively, using Assumption 3. Because $k_\tau(x, z)$ integrates to one and is non-negative on the support $F_X \times F_Z$, it is a proper density function wrt $F_X \times F_Z$, which proves part (a).

Now, we prove one direction of the if and only if result of Theorem 1(b). Suppose Assumption 4 holds. If $h_Z(z) = 0$ a.s.$[F_Z]$, then $\langle h_Z, z^{(j)} \rangle_{F_Z} = 0$ for $j = 0, ..., r_Z$. This and Lemma 3(c) (which applies because Assumption 4 holds) yield $h(x) = 0$ a.s.$[F_X]$, which establishes that $F_{XZ,\tau}$ is L^2-complete.

Next, we prove the other direction of the if and only if result of Theorem 1(b). Suppose Assumption 4 does not hold. Then, $\tau_j = 0$ for some $j = 1, ..., r_Z$ or $r_Z < r_X$. This implies that $\dim(S_{X,\tau}^\perp) > 0$ and the orthonormal basis $\{x^{(j)} : j = r_Z + 1, ..., r_X$ if $r_Z < r_X$ or $j \leq r_Z$ & $\tau_j = 0\}$ of $S_{X,\tau}^\perp$ has at least one element. Let $x_*^{(1)}$ denote any element in this basis. We show that the function $x_*^{(1)} \in S_{X,\tau}^\perp \subset L^2(F_X)$ satisfies (i) $E(x_*^{(1)}(X)|Z = z) = 0$ a.s.$[F_Z]$ and (ii) $x_*^{(1)}(x) = 0$ a.s.$[F_X]$ does not hold, which implies that $F_{XZ,\tau}$ is not L^2-complete wrt X. Property (i) holds by Lemma 3(d). Property (ii) holds because $||x_*^{(1)}||^2 = \int [x_*^{(1)}]^2 dF_X = 1$ by orthonormality. $\square$

Proof of Lemma 3. First, we establish Lemma 3(a) and 3(b). Let S_Z denote the linear subspace of $L^2(F_Z)$ generated by the orthonormal functions $\{z^{(j)} : 0 \leq j \leq r_Z\}$. Let $S_Z^\perp$ denote the orthogonal complement to S_Z in $L^2(F_Z)$. Let $\{z_*^{(j)} : j = 1, ..., r_{Z*}\}$

be an orthonormal basis for $S_Z^{\perp}$, where $0 \leq r_{Z*} \leq \infty$. If $\{z^{(j)} : j \leq r_Z\}$ is a basis of $L^2(F_X)$, then $r_{Z*} = 0$ and $\{z_*^{(j)}\}$ is the empty set. By construction, $\{z^{(j)}\} \cup \{z_*^{(j)}\}$ is an orthonormal basis of $L^2(F_Z)$.

By definition, $k_\tau(x, z)$ is the density of $F_{XZ,\tau}$ wrt $F_X \times F_Z$. The density of F_X wrt F_X is the constant function 1. Hence, $k_\tau(x, z)$ also is the conditional density of $F_{XZ,\tau}$ wrt $F_X \times F_Z$. This yields

$$h_Z(z) = \int h(x) k_\tau(x, z) dF_X(x) \text{ a.s.}[F_Z]. \tag{4.7}$$

The second equality of the following equation holds by (4.7): for $m = 0, ..., r_Z$,

$$\begin{aligned} \langle h_Z, z^{(m)} \rangle_{F_Z} &= \int h_Z(z) z^{(m)}(z) dF_Z(z) \\ &= \int \int h(x) k_\tau(x, z) dF_X(x) z^{(m)}(z) dF_Z(z) \\ &= \langle h z^{(m)}, k_\tau \rangle_{F_X \times F_Z}, \end{aligned} \tag{4.8}$$

where $\langle \cdot, \cdot \rangle_{F_X \times F_Z}$ denotes the $L^2(F_X \times F_Z)$ inner product. Equation (4.8) also holds with $z_*^{(m)}$ in place of $z^{(m)}$ for $m = 1, ..., r_{Z*}$.

Now we apply the Parseval-Bessel equality, see the proof of Theorem 1 above, to the right-hand side of (4.8). For each $m = 0, ..., r_Z$, we apply the Parseval-Bessel equality with (i) $H = L^2(F_X \times F_Z)$, (ii) $\{e_\alpha\}$ equal to the functions $\{x^{(j)} z^{(\ell)} : 0 \leq j \leq r_X, 0 \leq \ell \leq r_Z\} \cup \{x^{(j)} z_*^{(\ell)} : 0 \leq j \leq r_X, 1 \leq \ell \leq r_{Z*}\}$, which are orthonormal in $L^2(F_X \times F_Z)$,

$$\text{(iii) } x = h z^{(m)} = \sum_{j=0}^{r_X} \langle h, x^{(j)} \rangle_{F_X} x^{(j)} z^{(m)}, \tag{4.9}$$

where the second equality holds a.s.$[F_X]$ because $\{x^{(j)} : 0 \leq j \leq r_X\}$ is an orthonormal basis of $L^2(F_X)$, and

$$\text{(iv) } y = k_\tau = 1 + \sum_{j=1}^{r_Z} \tau_j x^{(j)} z^{(j)}, \tag{4.10}$$

where the second equality holds by definition, see (3.1). We have $k_\tau \in L^2(F_X \times F_Z)$ because $k_\tau^2(x, z) = (1 + \sum_{j=1}^{r_Z} \tau_j x^{(j)} z^{(j)})^2 \leq (1 + \sum_{j=1}^{r_Z} |\tau_j| B_{X,j} B_{Z,j})^2 \leq 4$ by Assumption 3. In addition, $h z^{(m)} \in L^2(F_X \times F_Z)$ because $h \in L^2(F_X)$ and $z^{(m)} \in L^2(F_Z)$. By the

Parseval-Bessel equality, we have

$$\langle hz^{(m)}, k_\tau \rangle_{F_X \times F_Z} = \tau_m \langle h, x^{(m)} \rangle_{F_X} \text{ for } m = 0, ..., r_Z, \tag{4.11}$$

because $x^{(m)} z^{(m)}$ is the only orthonormal basis function of $L^2(F_X \times F_Z)$ that $\{x^{(j)} z^{(m)} : 0 \leq j \leq r_X\}$ and $\{x^{(j)} z^{(j)} : 0 \leq j \leq r_X\}$ have in common.

By the same argument as in (4.9)-(4.11) but with $z_*^{(m)}$ in place of $z^{(m)}$, we obtain

$$\langle hz_*^{(m)}, k_\tau \rangle_{F_X \times F_Z} = 0 \text{ for } m = 1, ..., r_{Z*} \tag{4.12}$$

because $\{x^{(j)} z_*^{(m)} : 0 \leq j \leq r_X\}$ and $\{x^{(j)} z^{(j)} : 0 \leq j \leq r_X\}$ have no functions in common.

Equations (4.8), (4.11), and (4.12) combine to give

$$\begin{aligned} \langle h_Z, z^{(m)} \rangle_{F_Z} &= \tau_m \langle h, x^{(m)} \rangle_{F_X} \text{ for } m = 0, ..., r_Z \text{ and} \\ \langle h_Z, z_*^{(m)} \rangle_{F_Z} &= 0 \text{ for } m = 1, ..., r_{Z*}. \end{aligned} \tag{4.13}$$

Because $\{z^{(m)} : 0 \leq m \leq r_Z\} \cup \{z_*^{(m)} : 1 \leq m \leq r_{Z*}\}$ is an orthonormal basis of $L^2(F_Z)$, this yields the result of Lemma 3(a). It also gives the result of Lemma 3(b).

To prove Lemma 3(c), suppose Assumption 4 holds. Then, $r_Z = r_X$ and $\{x^{(j)} : 0 \leq j \leq r_Z\}$ is an orthonormal basis of $L^2(F_X)$. In consequence,

$$h(x) = \sum_{j=0}^{r_Z} \langle h, x^{(j)} \rangle_{F_X} x^{(j)}(x) = \sum_{j=0}^{r_Z} \tau_j^{-1} \langle h_Z, z^{(j)} \rangle_{F_Z} x^{(j)}(x), \tag{4.14}$$

where both equalities hold a.s.$[F_X]$, the first equality holds by the definition of an orthonormal basis, and the second equality holds by Lemma 3(b) and Assumption 4(i).

Lemma 3(d) follows from Lemma 3(a) because $h \in S_{X,\tau}^{\perp}$ implies that $\langle h, x^{(j)} \rangle_{F_X} = 0$ for those $x^{(j)}$ for which $\tau_j \neq 0$ for $j = 0, ..., r_Z$. $\square$

4.3 Proof of Lemma 1

Proof of Lemma 1. The set C_B is a closed convex subset of ℓ^p. Hence, it is completely metrizable in the relative topology induced from X. We show that the (universally measurable) set E, defined by

$$E = \{\{\tau_j\} \in C_B : \tau_j = 0 \text{ for some } j \geq 1\}, \tag{4.15}$$

is a shy subset of C_B. By Facts 0 and 3 in Anderson and Zame (2001, p. 12), it suffices to show that the set $E(1)$ is shy in C at $c = 0$, where

$$E(1) = \{\{\tau_j\} \in C_B : \tau_1 = 0\} \tag{4.16}$$

(because E is a countable union of sets of the form $E(k) = \{\{\tau_j\} \in C_B : \tau_k = 0\}$ and a set being shy at some $c \in C$ implies that it is shy at all $c \in C$). Given $\delta > 0$ and a neighborhood W of 0, define $\mu_{\delta,W}$ by

$$\begin{aligned} \mu_{\delta,W}(A) &= \lambda_{Leb}(A_1)/\lambda_{Leb}(C_{B,1}) \text{ for } A \subset C_B, \text{ where} \\ A_1 &= \{\tau_1 : \{\tau_j\} \in A \cap [\delta C_B] \cap W\}, \\ C_{B,1} &= \{\tau_1 : \{\tau_j\} \in C_B \cap [\delta C_B] \cap W\}, \end{aligned} \tag{4.17}$$

and λ_{Leb} denotes Lebesgue measure on R. The support of $\mu_{\delta,W}$ is in $[\delta C_B] \cap W$. Because δC_B is compact, the support of $\mu_{\delta,W}$ is in a compact set, as required. Note that $\lambda_{Leb}(C_{B,1}) > 0$, so $\mu_{\delta,W}$ is well defined.

Given the definition of $\mu_{\delta,W}$, we have

$$\mu_{\delta,W}(E(1)) = \lambda_{Leb}(E_1(1))/\lambda_{Leb}(C_{B,1}) = 0, \tag{4.18}$$

where $E_1(1) = \{\tau_1 : \{\tau_j\} \in E(1)\} = \{0\}$. Similarly, for all $x \in \ell^p$,

$$\mu_{\delta,W}(E(1) + x) = \lambda_{Leb}((E(1) + x)_1)/\lambda_{Leb}(C_{B,1}) = 0. \tag{4.19}$$

This holds because $E(1) + x = \{\{\tau_j\} \in C_B + x : \tau_1 = x_1\}$, where x_1 is the first element in the sequence x, $(E(1) + x)_1 = \{\tau_1 : \{\tau_j\} \in (E(1) + x) \cap [\delta C_B] \cap W\}$, and the latter set equals $\{x_1\}$ or ϕ. (The set $(E(1) + x)_1$ could be the null set ϕ because $(E(1) + x) \cap [\delta C_B] \cap W$ could be empty. In contrast, $E(1) \cap [\delta C_B] \cap W$ contains 0 and hence is not empty.) By (4.19), $E(1)$ is a shy set at $c = 0$ and the proof is complete. $\square$

References

AI, C. AND X. CHEN (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica,* 71, 1795-1843.

AITKEN, A. C. AND H. T. GONIN (1935): "Fourfold Sampling with and without Replacement," *Proceedings of the Royal Society of Edinburgh,* 55, 114-125.

AN, Y. AND Y. HU (2009): "Well-Posedness of Measurement Error Models for Self-reported Data," Department of Economics Working Paper No. 556, Johns Hopkins University.

ANDERSON, R. M. AND W. R. ZAME (2001): "Genericity with Infinitely Many Parameters," *Advances in Theoretical Economics,* 1, 1-62.

BERRY, S. T. AND P. A. HAILE (2009a): "Identification of a Heterogeneous Generalized Regression Model with Group Effects," Cowles Foundation Discussion Paper No. 1732, Yale University.

——— (2009b): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Cowles Foundation Discussion Paper No. 1718, Yale University.

——— (2010): "Identification in Differentiated Products Markets Using Market Level Data," Cowles Foundation Discussion Paper No. 1744, Yale University.

BAR-LEV, S AND D. PLACHKY (1989): "Boundedly Complete Families Which Are Not Complete," *Metrika,* 36, 331-336.

BISSANTZ, N., T. HOHAGE, A. MUNK, AND F. RUYMGAART (2007): "Convergence Rates of General Regularization Methods for Statistical Inverse Problems and Applications," *SIAM Journal of Numerical Analysis,* 45, 2610-2636.

BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica,* 75, 1613-1669.

CAMPBELL, J. T. (1934): "The Poisson Correlation Function," *Proceedings of the Edinburgh Mathematical Society,* Ser. 2, 4, 18-26.

Carroll, R. J., X. Chen, and Y. Hu (2009): "Identification and Estimation of Nonlinear Models Using Two Samples with Nonclassical Measurement Errors," *Journal of Nonparametric Statistics,* 22, 379-399.

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.

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

——— (2009b): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals," Cowles Foundation Discussion Paper No. 1650R, Yale University.

Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2010): "Identification in Semiparametric and Nonparametric Conditional Moment Models," unpublished manuscript, Cowles Foundation, Yale University.

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.

Christensen, J. P. R. (1974): *Topology and Borel Structure.* Volume 10 in North Holland Mathematical Studies. Amsterdam: North Holland.

Darolles, S., J.-P. Florens, and E. Renault (2000): "Nonparametric Instrumental Regression," unpublished manuscript, GREMAQ, University of Toulouse.

d'Haultfoeuille, X. (2011): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory,* 27, forthcoming.

Dudley, R. M. (1989): *Real Analysis and Probability.* Pacific Grove, CA: Wadsworth and Brooks/Cole.

Eagleson, G. K. (1964): "Polynomial Expansions of Bivariate Distributions," *Annals of Mathematical Statistics,* 35, 1208-1214.

Florens, J.-P., M. Mouchart, and J. M. Rolin (1990): *Elements of Bayesian Statistics.* New York: Marcel Dekker.

Ghosh, J. K. and R. Singh (1966): "Unbiased Estimation of Location and Scale Parameters," *Annals of Mathematical Statistics,* 37, 1671-1675.

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

Hamdan, M. A. and H. A. Al-Bayyani (1971): "Canonical Expansion of the Compound Correlated Bivariate Poisson Distribution," *Journal of the American Statistical Association,* 66, 390-393.

Hoeffding, W. (1977): "Some Incomplete and Boundedly Complete Families of Distributions," *Annals of Statistics,* 5, 278-291.

Horowitz, J. L. (2010): "Applied Nonparametric Instrumental Variables Estimation," unpublished manuscript, Department of Economics, Northwestern University.

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

Hu, Y. and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica,* 76, 195-216.

Hu, Y. and J. L. Shui (2011): "Nonparametric Identification Using Instrumental Variables," unpublished manuscript, Department of Economics, John Hopkins University.

Hu, Y. and M. Shum (2009): "Nonparametric Identification of Dynamic Models with Unobserved State Variables," Department of Economics Working Paper No. 543, Johns Hopkins University.

Hunt, B. R., T. Sauer, and J. A. Yorke (1992): "Prevalence: A Translation Invariant 'Almost Every' on Infinite-Dimensional Spaces," *Bulletin* (*New Series*) *of the American Mathematical Society,* 27, 217-238.

Isenbeck, M. and L. Ruschendorf (1992): "Completeness in Location Families," *Probability and Mathematical Statistics,* 13, 321-343.

Kibble, F. W. (1941): "A Two-variate Gamma Type Distribution," *Sankhya,* 5, 137-150

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

Lancaster, H. O. (1957): "Some Properties of the Bivariate Normal Distribution Considered in the Form of a Contingency Table," *Biometrika,* 44, 289-292.

——— (1958): "The Structure of Bivariate Distributions," *Annals of Mathematical Statistics,* 29, 719-736.

——— (1963): "Correlations and Canonical Forms of Bivariate Distributions," *Annals of Mathematical Statistics,* 34, 532-538.

Lehmann, E. L. (1986): *Testing Statistical Hypotheses,* 2nd ed. New York: Wiley.

Lehmann, E. L. and H. Scheffe (1950): "Completeness, Similar Regions, and Unbiased Estimation," *Sankhya,* 10, 305-340.

——— (1955): "Completeness, Similar Regions, and Unbiased Estimation," *Sankhya,* 15, 219-236.

Mattner, L. (1992): "Completeness of Location Families, Translated Moments, and Uniqueness of Charges" *Probability Theory and Related Fields,* 92, 137-149.

——— (1993): "Some Incomplete But Bounded Complete Location Families," *Annals of Statistics,* 21, 2158-2162.

——— (1996): "Complete Order Statistics in Parametric Models," *Annals of Statistics,* 24, 1265-1282.

Newey, W. K. and J. Powell (2003): "Instrumental Variables Estimation of Nonparametric Models," *Econometrica,* 71, 1565-1578.

Oosterhoff, J. and B. F. Schriever (1987): "A Note on Complete Families of Distributions," *Statistica Neerlandica,* 41, 183-189.

San Martin, E. and M. Mouchart (2006): "On Joint Completeness: Sampling and Bayesian Versions, and Their Connections," Discussion Paper No. 0625, Institut de Statistique, Université Catholique de Louvain.

SANTOS, A. (2009): "Inference in Nonparametric Instrumental Variables with Partial Identification," unpublished manuscript, Department of Economics, UCSD.

SASAKI, Y. (2011): "Identification and Estimation of a Nonseparable Dynamic Panel Model with Attrition Hazards," unpublished manuscript, Department of Economics, Brown University.

SEVERINI, T. A. AND G. TRIPATHI (2006): "Some Identification Issues in Nonparametric Linear Models with Endogenous Regressors," *Econometric Theory,* 22, 258-278.

SONG, S. (2011): "Semiparametric Estimation of Models with Conditional Moment Restrictions in the Presence of Nonclassical Measurement Errors," unpublished manuscript, Department of Economics, University of Wisconsin, Milwaukee.

STINCHCOMBE, M. (2002): "Some Genericity Analyses in Nonparametric Statistics," unpublished manuscript, Department of Economics, University of Texas, Austin.

WILHELM, D. (2011): "Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error," unpublished manuscript, Booth School of Business, University of Chicago.