TESTS OF INDEPENDENCE IN SEPARABLE ECONOMETRIC MODELS

By

Donald J. Brown and Marten Wegkamp

January 2003

COWLES FOUNDATION DISCUSSION PAPER NO. 1395

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

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

TESTS OF INDEPENDENCE IN SEPARABLE ECONOMETRIC MODELS

DONALD J. BROWN AND MARTEN H. WEGKAMP

Abstract. A common stochastic restriction in econometric models separable in the latent variables is the assumption of stochastic independence between the unobserved and observed exogenous variables. Both simple and composite tests of this assumption are derived from properties of independence empirical processes and the consistency of these tests is established.

1. Introduction

Recently, Brown and Wegkamp (2002) proposed a family of extremum estimators for semiparametric econometric models separable in the latent variables W, where $W = \rho(X, Y, \theta)$, X a random vector of observed exogenous variables, Y a random vector of observed endogenous variables, W is drawn from a fixed but unknown distribution and θ is a vector of unknown parameters. An important special case is the implicit nonlinear simultaneous equations model, where a reduced form function $Y = \rho^{-1}(X, W, \theta)$ exists. Of course, in general $Y = \rho^{-1}(X, W, \theta)$ is non-additive in W, e.g., consider the random utility model proposed by Brown and Matzkin (1998), where the random utility function $V(Y, W, \theta) = U(Y, \theta) + W \cdot Y$. In this case the structural equations defined by $W = \rho(X, Y, \theta)$ are equivalent to the first order conditions of maximizing $V(Y, W, \theta)$ subject to the budget constraint $P \cdot Y = I$ (P and I stand for prices and income, respectively). The details can be found in Section 2 below.

The principal maintained assumption in Brown and Wegkamp (2002) is the stochastic independence between W and X. In this paper we propose tests of this assumption using the elements of empirical independence processes. We present both simple tests, i.e., the null hypothesis states that for a given θ_0, $\rho(X, Y, \theta_0)$ and W are independent, as well as composite tests where the null hypothesis is that there exists some $\theta_0 \in \Theta$, the set of possible parameter values, such that X and $\rho(X, Y, \theta_0)$ are independent.

Date: January 15, 2003.
Key words and phrases. Cramér–von Mises distance, empirical independence processes, random utility models, semiparametric econometric models, specification test of independence.

Here we extend the analysis of Brown and Wegkamp (2002) beyond the characterization of the independence of random vectors in terms of their distribution functions. In particular, we define a family of weighted minimum mean-square distance from independence estimators in terms of characteristic or moment generating functions. These estimates are computationally more tractable than the ones considered by Brown and Wegkamp (2002). We show asymptotic normality, and consistency of the bootstrap for our estimates and the consistency for the tests.

The paper is organized as follows. In Section 2 of this paper we present both the general econometric model and two examples which motivated this research. Properties of empirical independence processes are reviewed in Section 3. Asymptotic properties of our estimators are derived in Section 4, and Section 5 discusses tests of independence between the observed and unobserved exogenous variables.

2. The Econometric Model

In this paper we consider semiparametric econometric models, which are separable in the latent variables. In these models we have a triple $(X, Y, W) \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \times \mathbb{R}^{k_2}$ of random vectors, where X and W are stochastically independent. The exogenous variable $W = \rho(X, Y) \in \mathbb{R}^{k_2}$ is unobserved and drawn from a fixed but unknown distribution. In this paper we consider structural equations ρ of the parametric form $\rho(x, y) = \rho(x, y, \theta)$ for some $\theta \in \Theta \subseteq \mathbb{R}^p$.

In general, two random vectors $X \in \mathbb{R}^{k_1}$ and $W \in \mathbb{R}^{k_2}$ are independent if and only if

$$\mathbb{E} f(X) g(W) = \mathbb{E} f(X) \mathbb{E} g(W) \text{ for all } f \in \mathcal{F}_1, \ g \in \mathcal{F}_2, \tag{2.1}$$

where $\mathcal{F}_\ell$ $(\ell = 1, 2)$ are

$$\mathcal{F}_\ell = \left\{ 1_{(-\infty, t]}(\cdot), \ t \in \mathbb{R}^{k_\ell} \right\}. \tag{2.2}$$

Note that each $\mathcal{F}_\ell$ in (2.2) is a universal Donsker class, indexed by a set of finite dimensional parameters $(s, t) \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$ only. This situation has been considered in Brown and Wegkamp (2002). Indeed, there are other classes $\mathcal{F}_\ell$, for which (2.1) holds as well. For example, the classes

$$\mathcal{F}_\ell = \left\{ \exp(< t, \cdot >), \ t \in \mathbb{R}^{k_\ell} \right\}, \tag{2.3}$$

or the classes

$$\mathcal{F}_\ell = \left\{ \exp(i < t, \cdot >), \ t \in \mathbb{R}^{k_\ell} \right\} \text{ where } i = \sqrt{-1}, \tag{2.4}$$

or the classes of all $\mathcal{C}^\infty$ functions on $\mathbb{R}^{k_\ell}$. The first two sets of classes are Donsker, provided t ranges in a bounded subset. In (2.3) we compare the joint moment generating functions (m.g.f.'s) with the product of its marginal m.g.f.'s, and in (2.4) the comparison is based on characteristic functions. The class of all $\mathcal{C}^\infty$ functions is not finite dimensional, and therefore is uninteresting from a computational perspective. We note in passing that this formulation using expected values does not allow for comparison between the joint density of X and $\rho(X, Y, \theta)$, and the product of its marginal densities. In fact, our estimators can be viewed as moment estimators as (2.1) is a family, albeit infinite, of moment conditions.

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent copies of the pair (X, Y). Motivated by the equivalence (2.1), we compare the empirical version

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i)g(\rho(X_i, Y_i, \theta)) = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \cdot \frac{1}{n}\sum_{i=1}^{n} g(\rho(X_i, Y_i, \theta)),$$

for all $f \in \mathcal{F}_1$ and $g \in \mathcal{F}_2$. Letting $\mathbb{P}_n = n^{-1}\sum_{i=1}^{n}\delta_{X_i, Y_i}$ be the empirical measure based on the sample $(X_1, Y_1), \cdots, (X_n, Y_n)$, we can write the preceding display more compactly as

$$\mathbb{P}_n f(x)g(\rho(x, y, \theta)) = \mathbb{P}_n f(x)\mathbb{P}_n g(\rho(x, y, \theta)) \text{ for all } f \in \mathcal{F}_1,\ g \in \mathcal{F}_2.$$

Observe that this amounts to comparing the joint cumulative distribution functions (c.d.f.'s) with the product of the marginal c.d.f.'s.

In order to obtain a tractable large sample theory, we consider the statistics

$$\mathbb{M}_n(\theta; \mathbb{P}_n; \mu) \equiv \iint_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \left\{\mathbb{P}_n f_s(x)g_t(\rho(x, y, \theta)) - \mathbb{P}_n f_s(x)\mathbb{P}_n g_t(\rho(x, y, \theta))\right\}^2 d\mu(s, t),$$

where μ is a c.d.f. acting as a weight function. We require that μ has a strictly positive density. In this way, we guarantee that all values s and t, that is, all functions $f_\ell \in \mathcal{F}_\ell$, are taken into account. The heuristic idea is that the unique minimizer of $\mathbb{M}_n(\theta; \mathbb{P}_n; \mu)$ should be close to the unique minimizer of

$$M(\theta; P; \mu) \equiv \iint_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \left\{P f_s(x)g_t(\rho(x, y, \theta)) - P f_s(x)P g_t(\rho(x, y, \theta))\right\}^2 d\mu(s, t),$$

where P is the probability measure of the pair (X, Y). The unique minimizer of this criterion is denoted by $\theta_P = \theta(P; \mu)$. Observe that $M(\theta; P; \mu)$ is finite for all θ since μ is a distribution function, and that $M(\theta_P; P; \mu) = 0$ if and only if $\rho(X, Y, \theta_P)$ and X are independent. In this case $\theta(P; \mu)$ does not depend on μ and we say that the model is identified. We can interpret $M(\theta)$ as the Cramér-von Mises distance between the actual distribution of the pair $(X, \rho(X, Y, \theta))$ and the (product) distribution of (X, W_θ), where the marginals X and W_θ are independent and W_θ has the same distribution as $\rho(X, Y, \theta)$. Observe that

$$M(\widehat{\theta}_n) \quad = \quad M(\theta_P) + \frac{1}{2}(\widehat{\theta}_n - \theta_P)' M''(\theta_n)(\widehat{\theta}_n - \theta_P),$$

provided $M \in \mathcal{C}^2(\Theta)$, for some θ_n between θ_P and $\widehat{\theta}_n$. We can view the first term on the right as the approximation error due to the finite dimensional model, and the last term can be thought of as the estimation error, which has an asymptotic χ_p^2 distribution (cf. Theorem 4.1 below) under some regularity assumptions. For instance, suppose that $\rho(X, Y)$ and X are independent for some ρ which we approximate by some finite series

$$\rho(x, y) \cong \rho(x, y, \theta) \equiv \sum_{i=1}^{p} \theta_i \psi_i(x, y)$$

based on some finite dimensional basis $\psi_1, \cdots, \psi_p$.

We end this section with two examples of implicit nonlinear simultaneous equations models separable in the latent variables, which motivated our research. In both examples, we show that the econometric model is identified for the class of extremum estimators proposed in this paper and hence can be estimated by these methods.

Example 2.1 (A Random Utility Model of Consumer Demand). We consider a consumer with a random demand function $Y(P, I, W, \theta_0)$ derived from maximizing a random utility function $V(Y, W, \theta_0)$ subject to her budget constraint $P \cdot Y = I$. First, the consumer draws W from a fixed and known distribution. Then nature draws $X = (P, I)$, from a fixed but unknown distribution. The main model assumption is that W and X are stochastically independent. The consumer solves the following optimization problem:

$$\text{maximize } V(y, w, \theta_0) \text{ over } y \text{ such that } p \cdot y = I.$$

The econometrician knows $V(y, w, \theta)$ and Θ, the set of all possible values for the parameter θ, but does not know θ_0, the true value of θ. Nor does the econometrician observe W or know the distribution of W. The econometrician does observe $X = (P, I)$. The econometrician's problem is to estimate θ_0 and the distribution of W from a sequence of observations $Z_i = (X_i, Y_i)$ for $i = 1, 2, ..., n$. The structural equations for this model are simply the first-order conditions of the consumer's optimization problem. These conditions define an implicit nonlinear simultaneous equations model of the form $W = \rho(X, Y, \theta)$, where the reduced form function is the consumer's random demand function $Y(P, I, W, \theta_0)$ for the specification of $V(y, W, \theta)$ proposed by Brown–Matzkin

(1998), i.e., $V(y,W,\theta) = U(y,\theta) + W \cdot y$. They assume that for all $\theta \in \Theta$, $U(y,\theta)$ is a smooth monotone strictly concave utility function on the positive orthant of $\mathbb{R}^k$, i.e., $DU(y,\theta) > 0$ and $D^2U(y,\theta)$ is negative definite for all y in the positive orthant of $\mathbb{R}^k$.

Our examples are suggested by their model, where first we consider:

$$V(y,W,\theta) = \ln y_0 + \theta_1 \ln y_1 + \theta_2 \ln y_2 + W_1 y_1 + W_2 y_2,$$

where $\theta_1, \theta_2 \in (0,1)$ and y_0 is the numeraire good. Then the first-order conditions for this optimization problem can be written as $W = \rho(X,Y,\theta)$, where $X = (P_1, P_2, I)$, $Y = (Y_0, Y_1, Y_2)$ and $\theta = (\theta_1, \theta_2)$

(i) $W_1 = P_1(I - P_1Y_1 - P_2Y_2)^{-1} - \theta_1 y_1^{-1}$
(ii) $W_2 = P_2(I - P_1Y_1 - P_2Y_2)^{-1} - \theta_2 y_2^{-1}$

Equations (i) and (ii) can (in principle) be solved uniquely for the random demand functions $Y_1(X,W,\theta)$ and $Y_2(X,W,\theta)$. This verifies that there exists a unique reduced form $Y = \gamma(X,W,\theta)$ such that$W = \rho(X, \gamma(X,W,\theta),\theta)$. To verify that the matrix $\partial\rho/\partial y$ has full rank, we note that

$$\left(\frac{\partial \rho}{\partial y}\right) = \begin{bmatrix} \dfrac{\partial \rho_1}{\partial y_1} & \dfrac{\partial \rho_1}{\partial y_2} \\ \dfrac{\partial \rho_2}{\partial y_1} & \dfrac{\partial \rho_2}{\partial y_2} \end{bmatrix}$$

where

$$\begin{aligned} \frac{\partial \rho_1}{\partial y_1} &= \frac{p_1^2}{(I - p_1 y_1 - p_2 y_2)^2} + \frac{\theta_1}{y_1^2} \\ \frac{\partial \rho_1}{\partial y_2} &= \frac{\partial \rho_2}{\partial y_1} = \frac{p_1 p_2}{(I - p_1 y_1 - p_2 y_2)^2} \\ \frac{\partial \rho_2}{\partial y_2} &= \frac{p_2^2}{(I - p_1 y_1 - p_2 y_2)^2} + \frac{\theta_2}{y_2^2} \end{aligned}$$

and $\det(\partial\rho/\partial y) > 0$. Hence $\partial\rho/\partial y$ has (full) rank 2. This and the implicit function theorem imply that $\partial y(x,W,\theta)/\partial x$ can be computed from the structural equations $W = \rho(x,y,\theta)$. In fact,

$$\begin{aligned} \left(\frac{\partial y}{\partial x}\right) &= -\left[\frac{\partial \rho}{\partial y}\right]^{-1} \left[\frac{\partial \rho}{\partial x}\right] \\ \left[\frac{\partial \rho}{\partial x}\right] &= \begin{bmatrix} \dfrac{\partial \rho_1}{\partial p_1} & \dfrac{\partial \rho_1}{\partial p_2} & \dfrac{\partial \rho_1}{\partial I} \\ \dfrac{\partial \rho_2}{\partial p_1} & \dfrac{\partial \rho_2}{\partial p_2} & \dfrac{\partial \rho_2}{\partial I} \end{bmatrix} \end{aligned}$$

where

$$
\begin{aligned}
\frac{\partial \rho_1}{\partial p_1} &= \frac{1}{(I - p_1 y_1 - p_2 y_2)} + \frac{p_1 y_1}{(I - p_1 y_1 + p_2 y_2)^2} \\
\frac{\partial \rho_1}{\partial p_2} &= \frac{p_2 y_1}{(I - p_1 y_1 + p_2 y_2)^2} \\
\frac{\partial \rho_1}{\partial I} &= \frac{-p_1}{(I - p_1 y_1 + p_2 y_2)^2} \\
\frac{\partial \rho_2}{\partial p_1} &= \frac{1}{(I - p_1 y_1 + p_2 y_2)} + \frac{p_2 y_2}{(I - p_1 y_1 + p_2 y_2)^2} \\
\frac{\partial \rho_2}{\partial p_2} &= \frac{p_1 y_1}{(I - p_1 y_1 + p_2 y_2)^2} \\
\frac{\partial \rho_2}{\partial I} &= \frac{-p_2}{(I - p_1 y_1 + p_2 y_2)^2}
\end{aligned}
$$

We see that $\partial\rho/\partial x$ has rank 2 and is independent of θ and W and that $\partial\rho/\partial y$ is independent of W. Therefore,

$$\frac{\partial y(x, W, \theta_0)}{\partial x} = \frac{\partial y(x, W, \theta)}{\partial x} \text{ a.e.}$$

if and only if

$$\frac{\partial \rho(x, y, \theta_0)}{\partial y} = \frac{\partial \rho(x, y, \theta)}{\partial y} \text{ a.e.}$$

But

$$\text{(iii)} \qquad \frac{\partial \rho(x, y, \theta_0)}{\partial y} = \frac{\partial \rho(x, y, \theta)}{\partial y} \text{ a.e. if and only if } \theta = \theta_0,$$

where Brown and Matzkin (1998, Theorem 1') have showm that

$$\theta \neq \theta_0 \iff \frac{\partial y(x, w, \theta_0)}{\partial x} \neq \frac{\partial y(x, w, \theta)}{\partial x}$$

is necessay and sufficient for identification. It is important to notice that the structural equations for this model are nonlinear in *both* parameters and variables.

Example 2.2 (A Simple Pure Trade Model). In applied general equilibrium analysis, there are two methods for determining parameter values: calibration and econometric estimation. The latter method, although theoretically more appealing, suffers from a number of limitations. In particular, the random shocks to tastes and technology enter the model in an ad hoc fashion, i.e., in most cases they are simply added to reduced forms of the deterministic structural equations, such as demand or supply functions. In addition, given the nonlinear nature of the structural equations, assumptions of model identification are problematic. In fact, as pointed out by Mansur and Whalley (1984) in their survey article, these issues have not been successfully resolved even for simple textbook models of general equilibrium such as the pure trade model. Surprisingly, this is still the case [cf. Dawkins, Srinivasan and Whalley (2001, page 3664)].

In this example, we consider a simple pure trade model with two countries, where the tastes of each country is characterized by a random utility function, representing the distribution of tastes within the country. The analysis is partial equilibrium in that the random utility functions $V(y, y_0, w) = U(y) + y_0 + w \cdot y$ are quasi-linear with a random linear perturbation.

The assumption of quasi-linearity plays a number of roles in our analysis. Most importantly, this specification gives rise to monotone individual demand functions for fixed realizations of W, see Quah (2000) for discussion. If we posit a distribution economy where the income distribution is fixed, then monotonicity of individual demand implies monotonicity of aggregate demand, a sufficient condition for uniqueness of the equilibrium price vector, see Hildenbrand (1994, Theorem 1 in Appendix 1). This uniqueness of equilibrium price vectors is an essential ingredient in our proof of identification.

Let us denote the two countries as A and B and the aggregate endowment in the world as $(\varepsilon, \varepsilon_0)$. Then the countrywide endowments are $(\varepsilon_A, \varepsilon_{0A}) = \alpha_A(\varepsilon, \varepsilon_0)$ and $(\varepsilon_B, \varepsilon_{0B}) = \alpha_B(\varepsilon, \varepsilon_0)$, where $\alpha_A, \alpha_B > 0$ and $\alpha_A + \alpha_B = 1$. We now normalize prices (p, p_0) such that $(p, p_0) \cdot (\varepsilon, \varepsilon_0) = 1$.

The observable exogenous random variables are $(\varepsilon, \varepsilon_0)$. The unobservable exogenous random variables are W_A and W_B, the random shocks to tastes. The observable endogenous random variables are the equilibrium price vector (p, p_0) and the consumptions of country A, (y_A, y_{0A}). α_A and α_B are deterministic and fixed.

As noted earlier these assumptions are sufficient for uniqueness of the equilibrium price vector, conditional on the realizations of $W = (W_A, W_B)$ and $(\varepsilon, \varepsilon_0)$, but they limit our ability to identify each country's characteristics, i.e., $\langle U_A, f_{w_A} \rangle$ and $\langle U_B, f_{w_B} \rangle$ where f_{w_A} and f_{w_B} are the distributions of W_A and W_B, respectively, since $(\varepsilon_A, \varepsilon_{0A})$ and $(\varepsilon_B, \varepsilon_{0B})$ are dependent, i.e., linearly related. Hence we assume that $U_A = U_B$. That is, each country has the same quasi-linear location function, but the distribution of tastes about the location function in each country may differ.

The remaining assumptions follow those of Brown and Matzkin, except we assume that the quasi-linear utility functions under consideration are parameterized by a compact subset of $\mathbb{R}^L$, with nonempty interior, denoted Θ. All distributions have smooth densities and their supports are in the positive orthants of the relevant Euclidean spaces. The final identifying assumption is that $W = (W_A, W_B)$ is stochastically independent of $(\varepsilon, \varepsilon_0)$.

We now proceed to show that this model is identified, i.e., if $\theta \neq \widetilde{\theta}$ then the resulting distributions of data are unequal. The structural equations can be expressed in terms of each country's F.O.C.'s for utility maximization subject to their budget constraints. We use the market clearing conditions to express the F.O.C.'s for country B in terms of country A's consumptions.

Structural Equations:

$$w_A = p/p_0 - DU(y_A) \tag{2.5}$$

$$w_B = p/p_0 - DU(\varepsilon - y_A) \tag{2.6}$$

$$y_{0A} = (\alpha_A - p \cdot y_A)/p_0 \tag{2.7}$$

$$y_{0B} = (\alpha_B - p \cdot (\varepsilon - y_A))/p_0 \tag{2.8}$$

We can solve these equations in two steps, because of the assumption of quasi-linear utility functions. First, we solve (2.5) and (2.6) for $q = p/p_0$ and y_A. Then we substitute these values into the budget constraints, (2.7) and (2.8), to solve for y_{0A} and p_0. Hence the relevant structural equations for our estimation procedure are

$$w_A = q - DU(y_A) \tag{2.9}$$

$$w_B = q - DU(\varepsilon - y_A) \tag{2.10}$$

This is a system of $2k$ equations in $2k$ unknowns, q and y_A, with $2k$ unobserved random variables $W = (W_A, W_B)$. We write this system as $w = g(q, y_A, \varepsilon, \theta)$ where θ indexes U. The standard assumptions on U that it is smooth, strictly concave and monotone with interior optima on budget sets, together with our earlier assumptions, guarantee the existence of a unique smooth function $h(w, \varepsilon, \theta) = (q, y_A)$ such that $w \equiv g(h(w, \varepsilon, \theta), \varepsilon, \theta)$. As noted earlier, Brown and Matzkin (1998, Theorem 1') have shown the following necessary and sufficient condition for identification: $\forall \theta, \widetilde{\theta} \in \Theta$, $\theta \neq \widetilde{\theta}$ if and only if

$$\frac{\partial h(w, \varepsilon, \theta)}{\partial \varepsilon} \neq \frac{h(\widetilde{w}, \varepsilon, \widetilde{\theta})}{\partial \varepsilon} \text{ where } \widetilde{w} = g(h(w, \varepsilon, \theta), \varepsilon, \widetilde{\theta}).$$

Applying the implicit function theorem to the structural equations (2.9) and (2.10), we deduce that

$$\frac{\partial h(w,\varepsilon,\theta)}{\partial \varepsilon} = \begin{bmatrix} I & -D^2U(y_A) \\ I & D^2U(\varepsilon - y_A) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ D^2U(\varepsilon - y_A) \end{bmatrix}$$

Our assumptions of smoothness and strict concavity of $U(\cdot)$ guarantee that $D^2U(y_A)$ is negative definite, hence invertible. Moreover, these assumptions guarantee that $[D^2U(y_A) + D^2U(\varepsilon - y_A)]^{-1}$ exists. Let $R = D^2U(y_A)$ and $S = D^2U(\varepsilon - y_A)$, then

$$\begin{bmatrix} I & -R \\ I & S \end{bmatrix}^{-1} = \begin{bmatrix} S(R+S)^{-1} & R(R+S)^{-1} \\ -(R+S)^{-1} & (R+S)^{-1} \end{bmatrix}$$

and

$$\begin{aligned} \begin{bmatrix} I & -R \\ I & S \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ S \end{bmatrix} &= \begin{bmatrix} S(R+S)^{-1} & R(R+S)^{-1} \\ -(R+S)^{-1} & (R+S)^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ S \end{bmatrix} \\ &= \begin{bmatrix} R(R+S)^{-1}S \\ (R+S)^{-1}S \end{bmatrix}. \end{aligned}$$

Therefore $\partial h(w,\varepsilon,\theta)/\partial\varepsilon = \partial h(\widetilde{w},\varepsilon,\widetilde{\theta})/\partial\varepsilon$ if and only if

(i) $R(R+S)^{-1}S = \widetilde{R}(\widetilde{R}+\widetilde{S})^{-1}\widetilde{S}$ and

(ii) $(R+S)^{-1}S = (\widetilde{R}+\widetilde{S})^{-1}\widetilde{S}$.

(i) and (ii) imply $(R-\widetilde{R})(R+S)^{-1}S = 0$. Since $(R+S)^{-1}S$ is nonsingular we see that $R = \widetilde{R}$. Summarizing, we have proven that the structural equations $w_A = q - DU(y_A)$ and $w_B = q - DU(\varepsilon - y_A)$ are identified if and only if $\forall\ \theta, \widetilde{\theta} \in \Theta$, $\theta \neq \widetilde{\theta}$ implies that $\exists \overline{y}_A$ such that $D^2U(\overline{y}_A) \neq D^2\widetilde{U}(\overline{y}_A)$.

One obvious example of a family of utility functions with this property are Cobb–Douglas utility functions.

3. Independence Empirical Processes

Given the classes $\mathcal{F}_1$ and $\mathcal{F}_2$, we define $\mathcal{F} \equiv \mathcal{F}_1$ and

$$\mathcal{G} \equiv \{f(\rho(\cdot,\cdot,\theta)) :\ f \in \mathcal{F}_2,\ \theta \in \Theta\} = \{g_t(\rho(\cdot,\cdot,\theta)) :\ t \in \mathbb{R}^{k_2},\ \theta \in \Theta\}.$$

As before, we denote the joint probability measure of the pair (X,Y) by P, and the empirical measure based on the sample $(X_1,Y_1),\cdots,(X_n,Y_n)$ by $\mathbb{P}_n$. For any $f \in \mathcal{F}$ and $g \in \mathcal{G}$, set

$$\mathbb{D}_n(f,g) \equiv \mathbb{P}_n fg - \mathbb{P}_n f \mathbb{P}_n g$$

and

$$D(f,g) \equiv Pfg - PfPg,$$

so that

$$\mathbb{M}_n(\theta) = \iint\limits_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \mathbb{D}_n^2(f_s, g_{t,\theta})\, d\mu(s,t)$$

in the new notation. Finally, we define the independence empirical process $\mathbb{Z}_n$ indexed by $\mathcal{F}\times\mathcal{G}$ by

$$\mathbb{Z}_n(f,g) \equiv \sqrt{n}(\mathbb{D}_n - D)(f,g).$$

Observe that [cf. Van der Vaart and Wellner (1996, page 367)]

$$\begin{aligned} \mathbb{Z}_n(f,g) &= \sqrt{n}\{(\mathbb{P}_n - P)(fg) - (\mathbb{P}_n g)(\mathbb{P}_n - P)(f) - (Pf)(\mathbb{P}_n - P)(g)\} \\ &= \sqrt{n}(\mathbb{P}_n - P)((f-Pf)(g-Pg)) - \sqrt{n}(\mathbb{P}_n - P)(f)(\mathbb{P}_n - P)(g) \qquad (3.1) \end{aligned}$$

The minor difference with the original formulation of independence empirical processes due to Van der Vaart and Wellner (1996, Chapter 3.8) is that we consider the marginal distributions of $(X, \rho(X,Y,\theta))$ rather than (X,Y). The next result states sufficient conditions for weak convergence of the independence empirical process $\mathbb{Z}_n$ in $\ell^\infty(\mathcal{F}\times\mathcal{G})$. Let $\|P\|_{\mathcal{F}}$ be the sup-norm on $\ell^\infty(\mathcal{F})$ for any class $\mathcal{F}$, i.e. $\|P\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} P|f|$.

Theorem 3.1. *Let $\mathcal{F}, \mathcal{G}$ and $\mathcal{F}\times\mathcal{G}$ be P-Donsker classes, and assume that $\|P\|_{\mathcal{F}} < \infty$ and $\|P\|_{\mathcal{G}} < \infty$. Then $\mathbb{Z}_n$ converges weakly to a tight Gaussian process Z_P in $\ell^\infty(\mathcal{F}\times\mathcal{G})$.*

Proof. The first term on the right in (3.1) converges weakly as $\mathcal{F}\times\mathcal{G}$ is P-Donsker. The second term in this expression is asymptotically negligible, since $\mathcal{F}$ and $\mathcal{G}$ are P-Donsker. We invoke Slutsky's lemma to conclude the proof. □

We can also bootstrap the limiting distribution of $\mathbb{Z}_n$. Let $(X_1^*, Y_1^*), \cdots, (X_n^*, Y_n^*)$ be an i.i.d. sample from $\mathbb{P}_n$, and let $\mathbb{P}_n^*$ be the corresponding bootstrap empirical measure. Then we define the bootstrap counterpart of $\mathbb{Z}_n$ by

$$\mathbb{Z}_n^*(f,g) = \sqrt{n}(\mathbb{D}_n^* - \mathbb{D}_n)(f,g),$$

where $\mathbb{D}_n^*(f,g) = \sqrt{n}(\mathbb{P}_n^* fg - \mathbb{P}_n^* f \mathbb{P}_n^* g)$.

Theorem 3.2. *Let $\mathcal{F}, \mathcal{G}$ and $\mathcal{F}\times\mathcal{G}$ be P-Donsker classes, and assume that $\|P\|_{\mathcal{F}} < \infty$ and $\|P\|_{\mathcal{G}} < \infty$. Then $\mathbb{Z}_n^*$ converges weakly to a tight Gaussian process Z_P in $\ell^\infty(\mathcal{F}\times\mathcal{G})$, given P^∞-almost every sequence $(X_1,Y_1),(X_2,Y_2),\cdots$.*

Proof. We first note that

$$\mathbb{Z}_n^*(f,g) = \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)((f-\mathbb{P}_n f)(g - \mathbb{P}_n g)) - \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)(f)(\mathbb{P}_n^* - \mathbb{P}_n)(g)$$

and recall that $\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$ converges weakly [cf. Theorem 3.9.12 in Van der Vaart and Wellner (1996)]. An application of Slutsky's lemma concludes our proof. $\square$

4. Estimation of θ_P

4.1. **A general result.** Given P-Donsker classes $\mathcal{F} = \{f_s : \ s \in \mathbb{R}^{k_1}\}$ and $\mathcal{G} = \{g_{t,\theta} : \ t \in \mathbb{R}^{k_2}, \ \theta \in \theta\}$ and a c.d.f. μ, we can define

$$\mathbb{M}_n(\theta) = \iint\limits_{\mathbb{R}^{k_1} \times \mathbb{R}^{k_2}} \mathbb{D}_n^2(f_s, g_{t,\theta})\, d\mu(s,t)$$

and

$$M(\theta) = \iint\limits_{\mathbb{R}^{k_1} \times \mathbb{R}^{k_2}} D^2(f_s, g_{t,\theta})\, d\mu(s,t).$$

We propose to estimate $\theta_P = \theta(P;\mu)$ by $\widehat{\theta}_n = \theta(\mathbb{P}_n;\mu)$ which minimizes the random criterion function $\mathbb{M}_n$ over Θ. Then, provided M has a unique, well-separated minimum at an interior point θ_P of Θ, it follows immediately by the weak convergence of $\mathbb{Z}_n$ (cf. Theorem 3.1) and Theorem 5.9 in Van der Vaart (1998, page 46) that

$$\widehat{\theta}_n \in \arg\min \mathbb{M}_n(\theta) \to \arg\min M(\theta) = \theta_P,$$

in probability. We will now show the asymptotic normality of the standardized distribution $\sqrt{n}(\widehat{\theta}_n - \theta_P)$.

We impose the following set of assumptions:

(A1) M has a unique global, well-separated minimum at θ_P in the interior of Θ and $M(\theta;P) \in \mathcal{C}^2(\Theta)$ and $M''(\theta_P;P)$ is non-degenerate.

(A2) $D(f_s, g_{t,\theta})$ is differentiable with respect to θ for all s,t, and its derivative satisfies

$$\left|\dot{D}(s,t,\theta) - \dot{D}(s,t,\theta_P)\right| \le |\theta - \theta_P|\Delta(s,t)$$

for some $\Delta \in \mathcal{L}^2(\mu)$.

(A3) $\sup_{s,t} P|f_s g_{t,\theta} - f_s g_{t,\theta_P}|^2 \to 0$ as $\theta \to \theta_P$.

(A4) The map $\rho(\cdot,\cdot,\theta)$ is continuously differentiable in θ.

(A5) The classes $\mathcal{F}, \mathcal{G}$ and $\mathcal{F} \times \mathcal{G}$ are P-Donsker.

We have the following result:

Theorem 4.1. *Assume (A1) – (A5). Then,* $\sqrt{n}(\widehat{\theta}_n - \theta_P)$ *has a non-degenerate Gaussian limiting distribution.*

Proof. The result follows from Theorem 3.2 in Wegkamp (1999, page 48). We need to verify the following three conditions:

(i) $\widehat{\theta}_n \to \theta_P$ in probability.
(ii) M has a non-singular second derivative at θ_P.
(iii) $\sqrt{n}(\mathbb{Z}_n - Z)(\theta)$ is stochastically differentiable at θ_P.

As noted above, (i) follows from general theory. Condition (ii) is subsumed in (A1). It remains to establish (iii). Let the symbol $\rightsquigarrow$ denote weak convergence in general metric spaces. (A3) implies that

$$\mathbb{Z}_n(\theta) - \mathbb{Z}_n(\theta_P) \rightsquigarrow 0 \ \text{ as } \theta \rightsquigarrow \theta_P,\ n \to \infty.$$

Consequently, by the continuous mapping theorem

$$\int\int_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \left[\mathbb{Z}_n^2(f_s, g_{t,\theta}) - \mathbb{Z}_n^2(f_s, g_{t,\theta_P})\right] d\mu(s,t) \rightsquigarrow 0$$

as $\theta \rightsquigarrow \theta_P,\ n \to \infty$. (A2), (A3) and the continuous mapping theorem yield also that

$$\int\int_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \Big[D(f_s, g_{t,\theta})\mathbb{Z}_n(f_s, g_{t,\theta}) - D(f_s, g_{t,\theta_P})\mathbb{Z}_n(f_s, g_{t,\theta_P}) \\ -(\theta-\theta_P)'\dot{D}(s,t,\theta_P)\mathbb{Z}_n(f_s, g_{t,\theta_P})\Big] d\mu(s,t) \rightsquigarrow 0$$

as $\theta \rightsquigarrow \theta_P,\ n \to \infty$. Conclude that

$$\begin{aligned}
&\sqrt{n}(\mathbb{M}_n - M)(\theta) \\
&= \int\int_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \mathbb{Z}_n^2(f_s, g_{t,\theta})\, d\mu(s,t) + 2\int\int_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} D(f_s, g_{t,\theta})\mathbb{Z}_n(f_s, g_{t,\theta})\, d\mu(s,t) \\
&= \int\int_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \mathbb{Z}_n^2(f_s, g_{t,\theta_P})\, d\mu(s,t) + 2\int\int_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} D(f_s, g_{t,\theta_P})\mathbb{Z}_n(f_s, g_{t,\theta_P})\, d\mu(s,t) + \\
&\quad +2(\theta-\theta_P)'\int\int_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \dot{D}(s,t,\theta_P)\mathbb{Z}_n(f_s, g_{t,\theta_P})\, d\mu(s,t) + o_p(1 + \|\theta-\theta_P\|) \\
&= \sqrt{n}(\mathbb{M}_n - M)(\theta_P) + 2(\theta-\theta_P)'\int\int_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \dot{D}(s,t,\theta_P)\mathbb{Z}_n(f_s, g_{t,\theta_P})\, d\mu(s,t) \\
&\quad + o_p(1 + \|\theta-\theta_P\|),
\end{aligned}$$

which establishes (iii). □

In fact, the asymptotic linear expansion

$$\sqrt{n}(\widehat{\theta}_n - \theta_P) = -2\left[M''(\theta_P)\right]^{-1} \iint\limits_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \dot{D}(s,t,\theta_P)\mathbb{Z}_n(f_s, g_{t,\theta_P})\, d\mu(s,t) + o_p(1) \tag{4.3}$$

holds. This expression coincides with the one derived in Brown and Wegkamp (2002, page 2045).

In addition, the conditional distribution of the bootstrap estimators $\sqrt{n}(\widehat{\theta}^*_n - \widehat{\theta}_n)$ has the same limit in probability. Here $\widehat{\theta}^*_n$ is based on i.i.d. sampling from $\mathbb{P}_n$, see Section 3. The proof of this assertion follows from similar arguments as Theorem 4.1, see Brown and Wegkamp (2002, pages 2046 - 2048) and is for this reason omitted.

Theorem 4.2. *Assume (A1) – (A5). Then*

$$\sqrt{n}(\widehat{\theta}^*_n - \widehat{\theta}_n) - \sqrt{n}(\widehat{\theta} - \theta_P) \rightsquigarrow 0.$$

We apply the developed theory to the special cases where $\mathcal{F}$ and $\mathcal{G}$ are indicator functions of half-spaces $(-\infty, \cdot]$ or exponential functions $\exp(t'x)$.

4.2. **Estimators based on the distribution functions.** For every $s \in \mathbb{R}^{k_1}, t \in \mathbb{R}^{k_2}$ and $\theta \in \Theta$, define the empirical distribution functions

$$\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\{X_i \le s\},\ \mathbb{G}_{n\theta}(t) = \frac{1}{n}\sum_{i=1}^{n}\{\rho(X_i, Y_i, \theta) \le t\} \text{ and}$$

$$\mathbb{H}_{n\theta}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\{X_i \le s,\ \rho(X_i, Y_i, \theta) \le t\}.$$

The criterion function $\mathbb{M}_n$ becomes in this case

$$\mathbb{M}_n(\theta) \equiv M(\theta; \mathbb{P}_n; \mu) = \iint\limits_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \{\mathbb{F}_n(s)\mathbb{G}_{n\theta}(t) - \mathbb{H}_{n\theta}(s,t)\}^2\, d\mu(s,t).$$

This is essentially the empirical criterion proposed by Brown and Wegkamp (2002). We obtain its theoretical counterpart $M(\theta) = M(\theta; P; \mu)$ by replacing the empirical distributions $\mathbb{F}_n, \mathbb{G}_{n\theta}$ and $\mathbb{H}_{n\theta}$ by the population distributions.

Assumption (A3) is verified if $\rho(x, y, \theta)$ is Lipschitz in θ, see Brown and Wegkamp (2002, page 2043, proof of Lemma 3). Assumptions (A2), and (A4) follow from smoothness assumptions on $\rho(\cdot,\cdot,\theta)$ and P. For (A1), we refer to Brown and Wegkamp (2002, Theorem 3, page 2038). We now show how to verify (A5).

We define the sets

$$A_{\theta,t} = \left\{(x,y) \in \mathbb{R}^{k_1+k_2} : \rho(x,y,\theta) \leq t\right\}, \ t \in \mathbb{R}^{k_2}, \ \theta \in \Theta,$$

and the associated collection

$$\mathcal{A} = \left\{A_{\theta,t} : \theta \in \Theta, \ t \in \mathbb{R}^{k_2}\right\}.$$

Note that $\mathcal{G}$ corresponds to the indicators I_A of sets $A \in \mathcal{A}$, and $\mathcal{F}$ corresponds to corresponds to the indicators I_B of sets $B \in \mathcal{B} \equiv \left\{\{x \in \mathbb{R}^{k_1} : x \leq t\}, \ t \in \mathbb{R}^{k_1}\right\}$, which is universally Donsker. Condition (A5) becomes in this specific setting

(A5') The classes of sets $\mathcal{A}$, $\mathcal{A} \times \mathcal{B}$ are P-Donsker.

Sufficient conditions for $\mathcal{A}$ to be P-Donsker are either smoothness of $\rho(x,y,\theta)$ (with respect to x and y, not θ) or that ρ ranges over a finite dimensional vector space. See Brown and Wegkamp (2002) for a discussion.

Example 4.3. Let $\{\rho(\cdot,\cdot,\theta), \ \theta \in \Theta\}$ be a subset of a finite dimensional vector space. Then both $\mathcal{A}$ and $\mathcal{B}$ are VC-classes, and $\mathcal{A} \times \mathcal{B}$, the product of two VC-classes, is again VC, see Van der Vaart and Wellner (1996, page 147). Hence $\mathcal{A}, \mathcal{B}$ and $\mathcal{A} \times \mathcal{B}$ are universally Donsker.

Example 4.4. Let the support of (X,Y) be a bounded, convex subset of $\mathbb{R}^{k_1+k_2}$ with non-empty interior, and, for each θ, $\rho(x,y,\theta)$ have uniformly bounded (by K) partial derivatives through order $\beta = \lfloor \alpha \rfloor$, and the derivatives of order β satisfy a uniform Hölder condition of order $\alpha - \beta$, and with Lipschitz constant bounded by K. For a complete description of the space $C_K^\alpha[X \times \mathcal{Y}]$, we refer to Van der Vaart and Wellner (1996), page 154. If $\alpha > d$ and P has a bounded density, then $\mathcal{A}$ and $\mathcal{A} \times \mathcal{B}$ are P-Donsker. To see why, we first notice that $\mathcal{A} \times \mathcal{B}$ has constant envelope 1, and that

$$Q|fg - \tilde{f}\tilde{g}|^2 \leq 2Q|f - \tilde{f}|^2 + 2Q|g - \tilde{g}|^2,$$

and that $f_L \leq f \leq f_U$ and $g_L \leq g \leq g_U$ implies $f_L g_L \leq fg \leq f_U g_U$. Hence

$$\mathcal{N}_B(2\varepsilon, L^2(Q), \mathcal{F} \times \mathcal{G}) \leq \mathcal{N}_B(\varepsilon, L^2(Q), \mathcal{F})\mathcal{N}_B(\varepsilon, L^2(Q), \mathcal{G}),$$

where $\mathcal{N}_B(\varepsilon, L^2(Q), \mathcal{F})$ is the ε-bracketing number of the set $\mathcal{F}$ with respect to the $L^2(Q)$ norm. Since $\log \mathcal{N}_B(\varepsilon, L^2(Q), \mathcal{B}) \lesssim \log(1/\varepsilon)$, the bound on the bracketing numbers in Corollary 2.7.3 in Van der Vaart and Wellner (1996) on $\mathcal{A}$ implies that $\mathcal{A} \times \mathcal{B}$ is P-Donsker.

4.3. **Estimators based on the moment generating functions.** Assume that X and $\rho(X, Y, \theta)$ are bounded, so that in particular their m.g.f.'s exist. For every $s \in \mathbb{R}^{k_1}, t \in \mathbb{R}^{k_2}$ and $\theta \in \Theta$, define the empirical m.g.f.'s

$$\phi_n(s) = \frac{1}{n}\sum_{k=1}^{n} \exp(< s, X_k >),\ \psi_{n\theta}(t) = \frac{1}{n}\sum_{k=1}^{n} \exp\left\{< t, \rho(X_k, Y_k, \theta) >\right\}$$

and $$\zeta_{n\theta}(s,t) = \frac{1}{n}\sum_{k=1}^{n} \exp\left\{< s, X_k > + < t, \rho(X_k, Y_k, \theta) >\right\}.$$

Let $k = k_1 + k_2$, and $C_\varepsilon > 0$ be such that $\mu[-C_\varepsilon, C_\varepsilon]^k = 1 - \varepsilon$. In this case we take the random criterion function $\mathbb{M}_n$

$$\mathbb{M}_n(\theta) \equiv M(\theta; \mathbb{P}_n; \mu) = \int\int_{[-C_\varepsilon, +C_\varepsilon]^k} \left\{\phi_n(s)\psi_{n\theta}(t) - \zeta_{n\theta}(s,t)\right\}^2 \, d\mu(s,t).$$

This setting corresponds to

$$\mathcal{F}_\varepsilon = \{\exp(< t, x >),\ t \in [-C_\varepsilon, +C_\varepsilon]^{k_1}\}$$

and

$$\mathcal{G}_\varepsilon = \{\exp(< t, \rho(x, y, \theta) >),\ t \in [-C_\varepsilon, +C_\varepsilon]^{k_2},\ \theta \in \Theta\}.$$

Van de Geer (2000, Lemma 2.5) shows that the box $[-C_\varepsilon, C_\varepsilon]^{k_1}$ can be covered by $(4C_\varepsilon\delta^{-1} + 1)^{k_1}$ many δ-balls in $\mathbb{R}^{k_1}$. Since

$$\mathbb{P}_n \left|\exp(< s, X >) - \exp(< t, X >)\right|^2 \lesssim \mathbb{P}_n \|X\|^2 \|s - t\|^2,$$

it follows from the above covering number calculation that the uniform entropy condition (cf. Van der Vaart and Wellner (1996, page 127) is met, and consequently the class $\mathcal{F}_\varepsilon$ is P-Donsker. Restricting the integration over $[-C_\varepsilon, C_\varepsilon]^k$, which has μ-probability equal to $1 - \varepsilon$, forces the function M to be within ε of the original criterion function, since

$$\left| \int\int_{\mathbb{R}^{k_1} \times \mathbb{R}^{k_2}} D^2(s,t)\, d\mu(s,t) - \int\int_{[-C_\varepsilon, C_\varepsilon]^k} D^2(s,t)\, d\mu(s,t) \right| \leq \mu\left(\mathbb{R}^k \setminus [-C_\varepsilon, C_\varepsilon]^k\right) \leq \varepsilon.$$

Assumption (A1) will force the corresponding unique minimizers to be close as well. Notice that $\mathcal{F}_\varepsilon$ is not a Donsker class if we take $C_\varepsilon = +\infty$. $\mathcal{G}_\varepsilon$ will be a P-Donsker class if $\{\rho(\cdot, \cdot, \theta) :\ \theta \in \Theta\}$ has this property. This is a consequence of the fact that the Donsker property of a class is preserved under Lipschitz transformations, see Theorem 2.10.6 in Van der Vaart and Wellner (1996, page 192).

Assumptions (A2) and (A3) follow from (A4), smoothness of $\rho(\cdot,\cdot,\theta)$, and the smoothness of the exponential function. Again, for (A1) we refer to Brown and Wegkamp (2002, Theorem 3, page 2038).

5. Tests of independence

Our null hypothesis is that $\rho(X,Y)$ and X are independent for some specified structural equation $\rho(x,y) = \rho(x,y,\theta_0)$. Following the discussion in Van der Vaart and Wellner (Chapter 3.8, 1996), a reasonable test is based on the Kolmogorov-Smirnov type statistic

$$\mathbb{K}_n \equiv \sup_{s,t} \sqrt{n} \left| \mathbb{P}_n f_s(x) g_t(\rho(x,y)) - \mathbb{P}_n f_s(x) \mathbb{P}_n g_t(\rho(x,y)) \right|.$$

Provided $\mathcal{F} \times \mathcal{G}$, $\mathcal{F}$ and $\mathcal{G}$ are P-Donsker, the limiting distribution of $\mathbb{K}_n$ under the null is known and can be bootstrapped (see Van der Vaart and Wellner, 1996, pages 367 -369).

Alternatively, we propose tests based on the criteria $\mathbb{M}_n$ defined above. Given observations (X_i, Y_i) we can compute $(X_i, W_i) \equiv (X_i, \rho(X_i, Y_i))$. Next, we note that

$$\begin{aligned} \mathbb{Z}_n(f,g) &\equiv \sqrt{n}(\mathbb{D}_n - D)(f, g \circ \rho) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ f(X_i) g(W_i) - \mathbb{E} f(X_i) \mathbb{E} g(W_i) \} \\ &\qquad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ f(X_i) - \mathbb{E} f(X_i) \} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ g(W_i) - \mathbb{E} g(W_i) \} \end{aligned}$$

is the same independence empirical process discussed in Van der Vaart and Wellner (1996, Section 3.8). Theorem 3.8.1 in Van der Vaart and Wellner (1996, page 368) states that $\mathbb{Z}_n(f,g)$ converges weakly to a tight Gaussian process Z_H in $\mathcal{F} \times \mathcal{G}$. Consequently, under the null hypothesis

$$n\mathbb{M}_n = \int\int_{\mathbb{R}^{k_1} \times \mathbb{R}^{k_2}} \left\{ \mathbb{Z}_n(f_s, g_t) + \sqrt{n} D(f_s, g_t) \right\}^2 \, d\mu(s,t) \tag{5.1}$$

converges weakly to

$$\int\int_{\mathbb{R}^{k_1} \times \mathbb{R}^{k_2}} Z_P^2(f_s, g_t) \, d\mu(s,t) \tag{5.2}$$

by the continuous mapping theorem. However,

$$n\mathbb{M}_n \to +\infty \ \ \text{(in probability)}$$

under any alternative $P_{X,W}$ with

$$\int D^2(f_s, g_t)\, d\mu(s,t) > 0,$$

which, provided $\mathcal{F}$ and $\mathcal{G}$ are generating classes as in (2.2), (2.3) or (2.4), is equivalent with X and $W = \rho(X,Y)$ are dependent. This implies that the power of the test converges to one under each alternative, that is, the test is consistent.

In lieu of the normal limiting distribution (5.2), we can also rely on the following bootstrap approximation for the distribution of the test statistic under the null. Let P^X and P^W be the probability measures of X and W, respectively, with empirical counterparts denoted by $\mathbb{P}_n^X$ and $\mathbb{P}_n^W$, respectively. Under the null hypothesis, the joint distribution of (X,W) is the product measure $P^X \times P^W$, and a natural estimate for the joint distribution of (X,W) is $\mathbb{P}_n^X \times \mathbb{P}_n^W$. In order to imitate the independence structure under the null hypothesis, we sample from the product measure $\mathbb{P}_n^X \times \mathbb{P}_n^W$. Let $(X_1^*, W_1^*), \cdots, (X_n^*, W_n^*)$ be the resulting i.i.d. sample from $\mathbb{P}_n^X \times \mathbb{P}_n^W$, and let $\mathbb{Z}_n^*$ and $\mathbb{M}_n^*$ be the bootstrap counterparts of $\mathbb{Z}_n$ and $\mathbb{M}_n$, respectively, based on this sample. Van der Vaart and Wellner (1996, Theorem 3.8.3) show that $\mathbb{Z}_n^*(f,g)$ converges weakly to $Z_{P^X \times P^W}$ almost surely. Since this limit coincides with the limiting distribution of $n\mathbb{M}_n$ under the null hypothesis, $n\mathbb{M}_n^*$ can be used to approximate the finite sample distribution of $n\mathbb{M}_n$ in a consistent manner (under the null hypothesis). Note that this procedure is model based as the resampling is done from the estimated model under the null hypothesis.

In addition, we present a specification test where the composite null hypothesis is the existence of a $\theta_0 \in \Theta$ such that X and $\rho(X,Y,\theta_0)$ are independent. We base the test on the statistic $T_n \equiv n\mathbb{M}_n(\widehat{\theta})$, and we show that T_n equals in distribution approximately $n\mathbb{M}_n(\theta_0)$ plus some drift due to $\widehat{\theta}_n$. In general the limiting distribution depends on θ_0, but it can be bootstrapped.

Theorem 5.1. *Assume (A1) – (A5) and $M(\theta_0) = 0$. Then*

$$n\mathbb{M}_n(\widehat{\theta}_n) - \int \left[\mathbb{Z}_n(f_s, g_{t,\theta_0}) + \sqrt{n}(\hat{\theta}_n - \theta_0)'\dot{D}(s,t,\theta_0)\right]^2 d\mu(s,t) \rightsquigarrow 0, \tag{5.3}$$

and

$$\int \left[\mathbb{Z}_n(f_s, g_{t,\theta_0}) + \sqrt{n}(\hat{\theta}_n - \theta_0)'\dot{D}(s,t,\theta_0)\right]^2 d\mu(s,t) \tag{5.4}$$

is asymptotically tight. Moreover,

$$n\mathbb{M}_n^*(\theta^*) - 4n\mathbb{M}_n(\widehat{\theta}) \rightsquigarrow 0. \tag{5.5}$$

Proof. First, we note that $\mathbb{Z}_n(f_s, g_{t,\theta})$ is stochastically differentiable in θ for all s, t by Condition (A3). An application of the functional continuous mapping theorem yields that

$$\int\int_{\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}} \left[\mathbb{Z}_n(f_s, g_{t,\theta'}) - \mathbb{Z}_n(f_s, g_{t,\theta})\right]^2 \, d\mu(s,t) \rightsquigarrow 0, \text{ for } \theta' \rightsquigarrow \theta.$$

The stochastic equicontinuity, weak convergence of $\widehat{\theta}_n$ and (A2) yield the following expansion of $\mathbb{M}_n(\hat{\theta}_n)$:

$$\begin{aligned} n\mathbb{M}_n(\widehat{\theta}_n) &= \int n\mathbb{D}_n^2(f_s, g_{t,\widehat{\theta}_n})\, d\mu(s,t) \\ &= \int \left[\left\{\mathbb{Z}_n(f_s, g_{t,\widehat{\theta}_n}) - \mathbb{Z}_n(f_s, g_{t,\theta_0})\right\} + \mathbb{Z}_n(f_s, g_{t,\theta_0}) + \sqrt{n}D(f_s, g_{t,\widehat{\theta}_n})\right]^2 d\mu(s,t) \\ &= \int \left[\mathbb{Z}_n(f_s, g_{t,\theta_0}) + \sqrt{n}(\widehat{\theta}_n - \theta_0)'\dot{D}(s,t,\theta_0)\right]^2 d\mu(s,t) + o_p(1). \end{aligned}$$

Since $\widehat{\theta}_n$ is asymptotically linear [cf. (4.3)], i.e.,

$$\widehat{\theta}_n = \theta_0 + n^{-1}\sum_{i=1}^n \psi(X_i) + o_p(n^{-1/2})$$

for some $\psi \in \mathcal{L}^2(P^X)$. Hence the vector

$$\left(\mathbb{Z}_n(f_s, g_{t,\theta_0})\ ,\ \sqrt{n}(\widehat{\theta}_n - \theta_0)\right) = \left(\mathbb{Z}_n(f_s, g_{t,\theta_0})\ ,\ n^{-1/2}\sum_{i=1}^n \psi(X_i) + o_p(1)\right)$$

converges weakly to a tight limit. Claim (5.3) and (5.4) follow from the continuous mapping theorem.

We note that by Theorem 4.2

$$\sqrt{n}(\theta^* - \widehat{\theta}_n) - \sqrt{n}(\widehat{\theta}_n - \theta_0) \rightsquigarrow 0 \tag{5.6}$$

and by condition (A3)

$$\mathbb{Z}_n(\widehat{\theta}_n) - \mathbb{Z}_n(\theta_0) \rightsquigarrow 0 \text{ and } \mathbb{Z}_n^*(\theta_0) - \mathbb{Z}_n(\theta_0) \rightsquigarrow 0. \tag{5.7}$$

Also, Theorem 3.2 and condition (A2) imply that

$$D(f_s, g_{t,\theta^*}) = (\theta^* - \theta_0)'\dot{D}(s,t,\theta_0) + o(\|\theta^* - \theta_0\|)$$

in $\mathcal{L}^2(\mu)$, so that

$$\int \left[D(f_s, g_{t,\theta^*}) - (\theta^* - \widehat{\theta}_n)'\dot{D}(s,t,\theta_0) - (\widehat{\theta}_n - \theta_0)'\dot{D}(s,t,\theta_0)\right]^2 d\mu(s,t) = o(n^{-1}).$$

This in term implies that

$$\begin{aligned} n\mathbb{M}_n^*(\theta^*) &= \int \left[\sqrt{n}\mathbb{D}_n^*(f_s, g_{t,\theta^*})\right]^2 \, d\mu(s,t) \\ &= \int \left[\mathbb{Z}_n^*(f_s, g_{t,\theta^*}) + \mathbb{Z}_n(f_s, g_{t,\theta^*}) + \sqrt{n}D(f_s, g_{t,\theta^*})\right]^2 \, d\mu(s,t) \\ &= \int \Big[\mathbb{Z}_n^*(f_s, g_{t,\theta^*}) + \mathbb{Z}_n(f_s, g_{t,\theta^*}) + \sqrt{n}(\theta^* - \widehat{\theta}_n)'\dot{D}(s,t,\theta_0) + \\ &\qquad +\sqrt{n}(\widehat{\theta}_n - \theta_0)'\dot{D}(s,t,\theta_0)\Big]^2 \, d\mu(s,t). \end{aligned}$$

Invoke (5.6) and (5.7) to deduce the second claim (5.5). □

This result says that the distribution of $n\mathbb{M}_n(\theta^*)/4$ can be used to approximate the finite sample distribution of our test statistics T_n. Again, we note that the power of the test converges to one, as $n\mathbb{M}_n(\widehat{\theta}_n) \to +\infty$ under any alternative $P_{X,W}$ with $\int D^2(f_s, g_t)\, d\mu(s,t) > 0$, that is, $P^X P^W \neq P^{X,W}$.

References

[1] Dawkins, C., Srinivasan, T.N. and Whalley, J. (2001). Calibration. In Heckman and Leamer (Eds) *Handbook of Econometrics, Vol 5.* Amsterdam: Elsevier, North Holland.

[2] Brown, D.J. and Matzkin, R. (1998). Estimation of Nonparametric functions in simultaneous equations models, with an application to consumer demand. *Working paper, Yale University*

[3] Brown, D. and Wegkamp, M. (2002). Weighted Minimum Mean-Square Distance from Independence Estimation. *Econometrica*, 70(5), 2035 – 2051.

[4] Hildenbrand, W. (1994). *Market demand: theory and empirical evidence.* Princeton: Princeton University Press.

[5] Mansur, A.H. and Whalley, J. (1984). Numerical specifications of applied general equilibrium models: estimation, calibration and data. In H.E. Scart and J.B. Shoven (eds) *Applied general Equilibrium analysis.* Cambridge: Cambridge University Press.

[6] Quah, J. (2000). The monotonicity of individual and market demand. *Econometrica*, 68, 911 – 930.

[7] Vaart, A. van der (1998). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

[8] Vaart, A. van der and Wellner, J. (1996). *Weak Convergence and Empirical Processes.* New York: Springer-Verlag.

[9] Wegkamp, M. (1999) *Entropy Methods in Statistical Estimation.* CWI-tract 125, Amsterdam.

Department of Economics, Yale University, Box 208268, New Haven, CT 06520-8268
E-mail address: donald.brown@yale.edu

Department of Statistics, Yale University, Box 208290 New Haven, CT 06520-8290
E-mail address: marten.wegkamp@yale.edu