Econometric Theory, **31**, 2015, 1016–1053.
doi:10.1017/S0266466614000620

TESTING FOR TREATMENT DEPENDENCE OF EFFECTS OF A CONTINUOUS TREATMENT

XUN LU
Hong Kong University of Science and Technology

HALBERT WHITE
University of California, San Diego

This paper provides nonparametric tests for hypotheses about the effects of a *continuous* treatment variable in a nonseparable structural equation. Specifically, we consider local average responses and average partial effects and test whether these effects are identical across different levels of treatment, including testing whether they are zero for all treatment levels as a special case. Our tests are based on consistent procedures of Bierens (1982, *Journal of Econometrics* 20, 105–134; 1990, *Econometrica* 58, 1443–1458) and Stinchcombe and White (1998, *Econometric Theory* 14, 295–324). The tests are easy to implement and achieve $n^{-1/2}$ local power. Monte Carlo simulations show that the tests perform well in finite samples. We apply our tests to study the interest rate elasticities of loan demand in microfinance. We also extend our testing procedures to covariate-conditioned average effects and marginal effects.

1. INTRODUCTION

We consider a nonseparable structural relation

$$Y = r(X, U), \tag{1.1}$$

where Y is the response of interest, X is a continuous observable treatment variable or cause of interest, U denotes other causes that may be unobservable, and r is an unknown continuously differentiable function. Throughout the paper, random variables are denoted using uppercase letters and their realizations are denoted using lowercase letters. We consider two average effects of X on Y that

We are indebted to the editor and two anonymous referees for their helpful comments. We also thank Karim Chalak, Graham Elliott, Wolfgang Härdle, Stefan Hoderlein, Meng Huang, Hidehiko Ichimura, Ivana Komunjer, Andres Santos, Max Stinchcombe, Yixiao Sun, Zhijie Xiao, and the participants at the 3rd XMU-Humboldt Workshop at Xiamen University, the 2013 Asian meeting of the Econometric Society, and the seminars at Boston College, Hong Kong University of Science and Technology, Monash University, University of California, San Diego, University of New South Wales, University of Sydney, and Wayne State University for their comments and suggestions. Lu acknowledges the support from Hong Kong University of Science and Technology under grant number DAG11BM09.
Address correspondence to Xun Lu, Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; e-mail: xunlu@ust.hk.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

are often of interest: the *local average response (LAR)* and the *average partial effect (APE)* at treatment level x :

$$\text{LAR}: \beta_0(x) \equiv E[m_0(x, U)|X = x], \tag{1.2}$$

$$\text{APE}: \gamma_0(x) \equiv E[m_0(x, U)], \tag{1.3}$$

where $m_0(x, u)$ is the *marginal effect (ME)* of X on Y given x and u :

$$\text{ME}: m_0(x, u) \equiv \frac{\partial r(x, u)}{\partial x}. \tag{1.4}$$

LAR is analogous to the average treatment effect on the treated and APE is analogous to average treatment effect in the binary treatment effect literature (Florens, Heckman, Meghir, and Vytlacil, 2008); for detailed discussion of LAR and APE, see, e.g., Chamberlain (1984), Blundell and Powell (2003), Altonji and Matzkin (2005), Wooldridge (2005), Bester and Hansen (2009), and Schennach, White, and Chalak (2012).

Both effects are functions of the treatment variable (x). As yet, there are few procedures for testing the properties of these functions. As these properties are of direct policy relevance, our goal here is to provide practical new tests for specific features of these effects. We test whether these two effects depend on the level of treatment (x) ("treatment dependence," or simply "dependence") on a given interval. Let I_x be a prespecified interval on the support of X and $\mathbf{1}$ denote the indicator function. The interval I_x is determined by the researcher's specific interest. Formally, we test

$$\begin{aligned} &\textit{Hypothesis 1}: \ \mathbb{H}_{10}: \textit{ There exists an unknown constant } c_0 \\ &\qquad \textit{s.t. } \mathbf{1}[X \in I_x] \cdot [\beta_0(X) - c_0] = 0 \ a.s. \\ &\qquad \mathbb{H}_{1A}: \mathbb{H}_{10} \textit{ is false.} \end{aligned} \tag{1.5}$$

$$\begin{aligned} &\textit{Hypothesis 2}: \ \mathbb{H}_{20}: \textit{ There exists an unknown constant } c_0 \\ &\qquad \textit{s.t. } \mathbf{1}[X \in I_x] \cdot [\gamma_0(X) - c_0] = 0 \ a.s. \\ &\qquad \mathbb{H}_{2A}: \mathbb{H}_{20} \textit{ is false.} \end{aligned} \tag{1.6}$$

That is, we test the hypotheses that each of these two functions of x is constant in x. If these hypotheses are rejected, policymakers may want to allocate resources differentially, depending on the level of treatment x. This is similar to testing for nonlinearity in a parametric predictive regression model, but we consider this question in a nonparametric nonseparable *structural* setting.

As a special case, we also test whether these two effects are zero at all levels of x on a given interval ("zero treatment effect"). That is, we test

$$\begin{aligned} &\textit{Hypothesis 1}': \ \mathbb{H}'_{10}: \ \mathbf{1}[X \in I_x] \cdot \beta_0(X) = 0 \ a.s. \\ &\qquad \mathbb{H}'_{1A}: \mathbb{H}'_{10} \textit{ is false.} \end{aligned} \tag{1.7}$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

$$\textit{Hypothesis } 2' : \mathbb{H}'_{20} : \mathbf{1}[X \in I_x] \cdot \gamma_0(X) = 0 \ a.s.$$
$$\mathbb{H}'_{2A} : \mathbb{H}'_{20} \textit{ is false.} \tag{1.8}$$

$\mathbb{H}'_{10}$ and $\mathbb{H}'_{20}$ say that LAR and APE are zero for *all* values of x on the interval I_x, respectively. These are relevant hypotheses for policy evaluation, as policymakers are often interested in the *general* effectiveness of a policy instead of at a particular point. Moreover, compared to testing $\beta_0(x) = 0$ or $\gamma_0(x) = 0$ at a given x, testing over an interval has the advantage of achieving faster convergence (a parametric rate) as shown in Section 3 (for a similar argument, see Pagan and Ullah, 1999, p. 177).

The literature contains many instances of these types of hypotheses. For example, consider the relation between years of schooling and wages, i.e., let Y be hourly wages, X levels of education, and U other unobservable drivers of wage, such as ability. In this context, $\beta_0(x)$ is the average effect of education on wage for those who choose x years of schooling, while $\gamma_0(x)$ measures the average effect of education on wage at x for the whole population. One would be interested in whether these two effects are zeros globally on an interval and whether the effects vary at different levels of education. For example, Belzil and Hansen (2002) argue that "there is no obvious reason to presume that the local returns to schooling are independent of grade level. Indeed, heterogeneity in any component of schooling choices will lead to improper inference if the local returns are erroneously assumed to be constant". Take another example from Hoderlein and White (2012), where Y is the logarithm of calorie intake (demand), X is the logarithm of income, and U are the unobservable preferences of the household or measurement errors in calories. Then, $\beta_0(x)$ is the average income elasticity for those with income x, while $\gamma_0(x)$ measures the average income elasticity for the whole population. In this case, we would like to know whether these two elasticities are zero over an interval and whether they are constant across different income levels. These two types of hypotheses are similar to the treatment effect heterogeneity considered in Crump, Hotz, Imbens, and Mitnik (2008), who test whether the average effects of a binary treatment are zero or constant over different subpopulations defined by covariates. Here we consider a continuous treatment and test whether the average effects are constant or zero across different treatment levels.

Throughout, we impose a conditional form of exogeneity, namely that X and U are independent given covariates Z, as in Altonji and Matzkin (2005) and White and Chalak (2006). This is analogous to the unconfoundedness assumption in the treatment effect literature. Our tests are based on the integrated conditional moment type of test in the specification testing literature (Bierens, 1982, 1990; Bierens and Ploberger, 1997; Stinchcombe and White, 1998). The tests are simple to implement, yet achieve $n^{-1/2}$ local power. We show that the test statistics converge weakly to a nonstandard distribution, specifically an integral of a function of a Gaussian process. Critical values for these test statistics are straightforwardly obtained by subsampling or the bootstrap. Although, to the best of our knowledge, there exist no other tests that can be directly applied to test the hypotheses

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

discussed in this paper, in principle one can apply other types of consistent nonparametric tests, such as those based on L_2 distance (see, e.g., Härdle and Mammen, 1993). However, the L_2 distance test is only able to detect the local alternative that converges to the null at a rate slower than $n^{-1/2}$ and thus suffers from the "curse of dimensionality". Fan and Li (2000) provide a detailed comparison between these two types of tests.

In this paper, we also extend the tests of zero treatment effect and treatment dependence to covariate-conditioned average effects (White and Chalak, 2006) and marginal effects. The plan of the paper is as follows. In Section 2, we lay out the framework for analyzing the various effects and introduce the key conditional exogeneity assumption. In Section 3, we propose test statistics and discuss their asymptotic properties. Section 4 provides some Monte Carlo simulation evidence. In Section 5, we apply our tests to study the interest rate elasticities of demand for loans. In Section 6, we extend our tests to covariate-conditioned average effects and marginal effects. Section 7 contains concluding remarks. Mathematical proofs are relegated to the Appendix.

2. FRAMEWORK

We first introduce our data generating process.

Assumption A.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, on which we define random vectors $\{X, U\}$. X and U take values in $\mathbb{X}$ and $\mathbb{U}$ respectively, $\mathbb{X} \equiv [x_L, x_U]$ where x_L and $x_U \in \mathbb{R}$ $(x_L < x_U)$ and $\mathbb{U} \subset \mathbb{R}^{k_u}$, where k_u is a positive (possibly countably infinite) integer. Let Y be a random variable structurally generated as

$$Y = r(X, U),$$

where r is an unknown measurable function: $r : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$. Further, suppose that for each $u \in \mathbb{U}$, $r(\cdot, u) : \mathbb{X} \to \mathbb{R}$ is continuously differentiable with $E[m_0(x, U)] < \infty$ for all $x \in \mathbb{X}$, where

$$m_0(x, u) \equiv \frac{\partial r(x, u)}{\partial x}.$$

The boundedness of X is standard in nonparametric settings.

We follow Dawid (1979) and write $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$ to denote that $\mathcal{X}$ and $\mathcal{Y}$ are independent given $\mathcal{Z}$. We impose the key conditional exogeneity assumption next.

Assumption A.2. Let Z denote a random vector of finite dimension $k_z \in \mathbb{N}$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with compact support $\mathbb{Z} \subset \mathbb{R}^{k_z}$. Suppose that

$$X \perp U \mid Z,$$

and that X and U are not measurable with respect to the sigma-field generated by Z.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

A.2 is a common assumption for identifying effects (see, e.g., Altonji and Matzkin, 2005; White and Chalak, 2006; Hoderlein and Mammen, 2007, 2009; Hoderlein, 2011) and is closely related to the unconfoundedness assumption in the treatment effect literature. Variables Z are typically called control variables or covariates. Note that here Z can overlap with U, thus Z is not necessarily excluded from the structural equation r. A.2 is plausible in many different cases. First, one special case of A.2 is $X \perp U$, i.e., Z has null dimension. This is the case of strict exogeneity. When X is randomized as in an experiment, strict exogeneity holds. A second case in which A.2 is plausible is when Z is used to proxy for the unobservable causes U or for common causes of X and U. A classic example is to use IQ test scores to proxy unobservable ability to estimate the returns to education. It is plausible that conditional on the IQ test score, education and ability are independent, i.e., the IQ test score is a sufficient proxy for unobservable ability (Griliches, 1977; Dawid, 1979). A third case is one in which we can construct control variables Z from standard exogenous instrumental variables, provided certain strict monotonicity assumptions hold, as shown in Imbens and Newey (2009).[1]

Under the key conditional exogeneity assumption A.2, both LAR and APE are identified. We let $\mu(x,z) \equiv E(Y \mid X = x, Z = z)$ and $f(z \mid x)$ and $f(z)$ denote the conditional density of Z given X and the density of Z, respectively. Then LAR and APE are identified respectively as

$$\beta_0(x) \equiv E[m_0(x,U)|X = x] = \int \frac{\partial \mu(x,z)}{\partial x} \cdot f(z \mid x) \cdot dz, \text{ and}$$

$$\gamma_0(x) \equiv E[m_0(x,U)] = \int \frac{\partial \mu(x,z)}{\partial x} \cdot f(z) \cdot dz.$$

The key condition to achieving identification is Assumption A.2. To see this,

$$\begin{aligned}
\beta_0(x) &\equiv E\left[\frac{\partial r(x,U)}{\partial x}\middle| X = x\right] = \int E\left[\frac{\partial r(x,U)}{\partial x}\middle| X = x, Z = z\right] \cdot f(z|x) \cdot dz \\
&\overset{*}{=} \int E\left[\frac{\partial r(x,U)}{\partial x}\middle| Z = z\right] \cdot f(z|x) \cdot dz \\
&\overset{**}{=} \int \frac{\partial}{\partial x} E[r(x,U)|Z = z] \cdot f(z|x) \cdot dz \\
&\overset{*}{=} \int \frac{\partial}{\partial x} E[r(X,U)|Z = z, X = x] \cdot f(z|x) \cdot dz \\
&= \int \frac{\partial \mu(x,z)}{\partial x} \cdot f(z|x) \cdot dz,
\end{aligned}$$

where equalities $(*)$ follow by A.2 and equality $(**)$ follows by assuming that integral and derivative can be interchanged (see, e.g., Casella and Berger, 2002, Thm. 2.4.3, p. 70). The case for $\gamma_0(x)$ can also be derived similarly.

For the technical reason discussed below, we further require that the given interval I_x in $\mathbb{H}_{10}$, $\mathbb{H}_{20}$, $\mathbb{H}'_{10}$, and $\mathbb{H}'_{20}$ excludes the boundary points of the support

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

of X. That is,

$$I_x = [\underline{x},\ \bar{x}], \text{ such that } x_L + \epsilon \leq \underline{x} < \bar{x} \leq x_U - \epsilon,$$

where $\epsilon > 0$ and $\underline{x}$ and $\bar{x}$ are prespecified by the researcher's specific interests.

3. TEST STATISTICS

3.1. Introduction

Our tests are based on results in the specification testing literature (see, e.g., Bierens, 1982, 1990; Bierens and Ploberger, 1997; Stinchcombe and White, 1998). We first introduce a particular class of functions, the generically comprehensively revealing (GCR) functions.

DEFINITION 1 (GCR function, Stinchcombe and White, 1998). *We say $\varphi : \mathbb{R} \to \mathbb{R}$ is GCR if it is a positive bounded real analytic nonpolynomial function.*

Stinchcombe and White (1998) define a more general class of GCR functions, but this suffices here. GCR functions include many familiar functions such as the logistic CDF, normal CDF, and the normal density. Next, we restate a result of Stinchcombe and White (1998) that plays a key role in our test.

LEMMA 1. *Suppose φ is GCR. Let $\mathcal{X}$ be a random vector of dimension d where d is a finite integer and let $\mathcal{U}$ be a random scalar. Let $(\Lambda, \mathcal{G}, \nu)$ be a measure space where Λ is a compact subset of $\mathbb{R}^{d+1}$, $\mathcal{G}$ a Borel sigma-field on Λ, and ν is absolutely continuous with respect to Lebesgue measure. Let $\lambda \equiv (\lambda_0, \lambda_1) \in \Lambda$, where $\lambda_0 \in \mathbb{R}$ and $\lambda_1 \in \mathbb{R}^d$. Then,*

(a) *$E(\mathcal{U} | \mathcal{X}) = 0$ a.s. (almost surely) if and only if $E\left[\varphi\left(\lambda_0 + \mathcal{X}'\lambda_1\right) \cdot \mathcal{U}\right] = 0$ for ν-almost all $\lambda \in \Lambda$.*

(b) *Suppose $g : \mathbb{R}^d \to \mathbb{R}$ is a measurable function. Then $g(\mathcal{X}) = 0$ a.s. if and only if $E\left[\varphi\left(\lambda_0 + \mathcal{X}'\lambda_1\right) \cdot g(\mathcal{X})\right] = 0$ for ν-almost all $\lambda \in \Lambda$.*

In what follows, we use the notation of this lemma without further reference. We impose further assumptions on the density of (X, Y, Z). For this, we let $\Theta_{x,z}$ denote the joint support of X and Z and define

$$\sigma^2(x, z) \equiv var(Y \mid X = x, Z = z).$$

Assumption A.3.

(a) $\Theta_{x,z} \subset \mathbb{X} \times \mathbb{Z}$ is compact.

(b) The distribution function of (X, Y, Z) is absolutely continuous with respect to Lebesgue measure.

(c) $\mu(\cdot, z) : \mathbb{X}_z \to \mathbb{R}$ is continuously differentiable for all $z \in \mathbb{Z}$, where $\mathbb{X}_z$ is the conditional support of X given $Z = z$.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

(d) $\sup_{x,z\in\Theta_{x,z}} \sigma^2(x,z) < \infty$.

Assumption A.4.

(a) $(X_i, Y_i, Z_i), i = 1, 2, \ldots$ is a sequence of independent random vectors, identically distributed as (X, Y, Z).

(b) Realizations of X_i, Y_i, and Z_i are observed; elements of U_i can (but need not) be unobservable.

We let $f(x,y,z)$, $f(x,z)$, $f(z)$, and $f(x)$ denote the densities of (X,Y,Z), (X,Z), Z, and X respectively. Let $f_x(x,y)$ and $f_x(x,z)$ denote the derivatives of $f(x,y)$ and $f(x,z)$ with respect to x respectively. Throughout the paper, we assume that integral and derivative can be interchanged (see, e.g., Casella and Berger, 2002). We focus on testing $\mathbb{H}_{10}$ and $\mathbb{H}_{20}$, as testing $\mathbb{H}'_{10}$ and $\mathbb{H}'_{20}$ are special cases of them.

3.2. Testing $\mathbb{H}_{10}$ (Constant Effects for LAR)

To test $\mathbb{H}_{10}$, we first define a weight function, which we use to avoid boundary problems in deriving the asymptotic distribution of our test statistics.

DEFINITION 2. *Suppose* $w: \mathbb{R} \to \mathbb{R}_+$ *satisfies* (i) $w(x) > 0$ *if* $\underline{x} < x < \bar{x}$; (ii) $w(x) = 0$ *if* $x \le \underline{x}$ *or* $x \ge \bar{x}$; (iii) $\int w(x) \cdot dx = 1$; *and* (iv) w *is continuously differentiable on* $\mathbb{R}$. *Then we say* w *is a* boundary weight function.

For example, w could be given by

$$w(x) = \begin{cases} 30 \cdot (\bar{x} - \underline{x})^{-5} \cdot (x - \underline{x})^2 \cdot (x - \bar{x})^2 & \text{if } \underline{x} < x < \bar{x}, \\ 0 & \text{if } x \le \underline{x} \text{ or } x \ge \bar{x}. \end{cases} \tag{3.1}$$

In this example, $w(x)$ is similar to a quartic kernel.

Testing $\mathbb{H}_{10}$ is equivalent to testing

$$w(X) \cdot [\beta_0(X) - \bar{\beta}_0] = 0 \ a.s., \tag{3.2}$$

where

$$\bar{\beta}_0 = \int \beta_0(x) \cdot w(x) \cdot dx.$$

Further, based on Lemma 1, testing equation (3.2) is equivalent to testing

$$\mathcal{T}_{10}(\lambda) \equiv \int \varphi(\lambda_0 + \lambda_1 x) \cdot w(x) \cdot [\beta_0(x) - \bar{\beta}_0] \cdot dx = 0 \text{ for } \nu\text{-almost all } \lambda \in \Lambda,$$

where φ is GCR.[2] We summarize the result in the following lemma.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

LEMMA 2. *Let Assumption A.1 hold. Suppose* $f(x)$ *is bounded away from zero on* $[\underline{x}, \bar{x}]$. *Let* φ *be a GCR function and* w *a boundary weight function. Then,* $\mathbb{H}_{10}$ *holds if and only if* $\mathcal{T}_{10}(\lambda) = 0$ *for* ν*-almost all* $\lambda \in \Lambda$, *where, with* $d = 1$,

$$\mathcal{T}_{10}(\lambda) \equiv \int \psi(x, \lambda) \cdot \beta_0(x) \cdot dx,$$

$$\psi(x, \lambda) \equiv \varphi(\lambda_0 + x \cdot \lambda_1) \cdot w(x) - \left[\int \varphi(\lambda_0 + x \cdot \lambda_1) \cdot w(x) \cdot dx\right] \cdot w(x).$$

Remark 1. Similarly, we can show that $\mathbb{H}'_{10}$ holds if and only if $\mathcal{T}'_{10}(\lambda) = 0$ for ν-almost all $\lambda \in \Lambda$, where

$$\mathcal{T}'_{10}(\lambda) \equiv \int \psi'(x, \lambda) \cdot \beta_0(x) \cdot dx, \text{ and } \psi'(x, \lambda) \equiv \varphi(\lambda_0 + x \cdot \lambda_1) \cdot w(x).$$

Thus, testing $\mathbb{H}'_{10}$ is similar to testing $\mathbb{H}_{10}$; we just use different weight functions: $\psi(x, \lambda)$ for $\mathbb{H}_{10}$ and $\psi'(x, \lambda)$ for $\mathbb{H}'_{10}$. All the results below hold for $\mathbb{H}'_{10}$ if we simply replace $\psi(x, \lambda)$ with $\psi'(x, \lambda)$.

Under A.1 and A.2, it is easy to show that

$$\beta_0(x) = \int y \cdot \frac{f_x(x, y)}{f(x)} \cdot dy - \int \frac{\left(\int y \cdot f(x, y, z) \cdot dy\right) \cdot f_x(x, z)}{f(x, z) \cdot f(x)} \cdot dz.$$

Thus, Altonji and Matzkin (2005) propose the following estimator of $\beta_0(x)$:

$$\hat{\beta}(x) = \int y \cdot \frac{\hat{f}_x(x, y)}{\hat{f}(x)} \cdot dy - \int \frac{\left(\int y \cdot \hat{f}(x, y, z) \cdot dy\right) \cdot \tau\left(\hat{f}(x, z), t\right) \cdot \hat{f}_x(x, z)}{\hat{f}(x)} \cdot dz,$$

with trimming function $\tau : \mathbb{R} \times \mathbb{R}_{++} \to \mathbb{R}_+$:

$$\tau(x, t) = \begin{cases} 1/x & \text{if } x \geq 2t, \\ \frac{49(x-t)^3}{8t^4} - \frac{39(x-t)^4}{4t^5} + \frac{31(x-t)^5}{8t^6} & \text{if } t \leq x < 2t, \\ 0 & if\, x < t, \end{cases} \tag{3.3}$$

with $t \equiv t_n$; and $\hat{f}(x)$, $\hat{f}(x, z)$, $\hat{f}_x(x, y)$, $\hat{f}_x(x, z)$, and $\hat{f}(x, y, z)$ are kernel estimators of $f(x)$, $f(x, z)$, $f_x(x, y)$, $f_x(x, z)$, and $f(x, y, z)$ respectively with bandwidth $h \equiv h_n$.[3] For example,

$$\hat{f}(x, y, z) = \frac{1}{nh^{2+k_z}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_i - x}{h}\right) \mathbf{K}\left(\frac{Y_i - y}{h}\right) \mathbf{K}_z\left(\frac{Z_i - z}{h}\right),$$

where $\mathbf{K}(.)$ is a univariate kernel, and $\mathbf{K}_z(.)$ is a k_z-dimensioned product of univariate kernels:

$$\mathbf{K}_z\left(\frac{Z_i - z}{h}\right) = \prod_{j=1}^{k_z} \mathbf{K}\left(\frac{Z_{j,i} - z_j}{h}\right), \tag{3.4}$$

$Z_{j,i}$ and z_j denote the jth elements of Z_i and z respectively.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

An estimator of $\mathcal{T}_{10}(\lambda)$ is

$$\begin{aligned}\hat{\mathcal{T}}_{1n}(\lambda) &\equiv \int \psi(x,\lambda)\cdot\hat{\beta}(x)\cdot dx \\ &= \int \psi(x,\lambda)\left(\int y\cdot\frac{\hat{f}_x(x,y)}{\hat{f}(x)}\cdot dy\right)\cdot dx \\ &\quad - \int \psi(x,\lambda)\left[\int \frac{\left(\int y\cdot\hat{f}(x,y,z)\cdot dy\right)\cdot\tau\left(\hat{f}(x,z),t\right)\cdot\hat{f}_x(x,z)}{\hat{f}(x)}\cdot dz\right]\cdot dx.\end{aligned}$$

By construction, the weighting function $\psi(x,\lambda)$ is known and vanishes at the boundary points, which allows us to ignore the boundary term in the asymptotic distribution of $\hat{\mathcal{T}}_{1n}(\lambda)$. Otherwise, the asymptotic distribution of $\hat{\mathcal{T}}_{1n}(\lambda)$ will be dominated by the boundary points and $\sqrt{n}$-consistency of $\hat{\mathcal{T}}_{1n}(\lambda)$ is impossible in our case.[4] For a similar argument, see Newey and Stoker (1993, p. 1206).

We view $\hat{\mathcal{T}}_{1n}(\cdot)$ as a random element in the metric space $(\mathcal{C}(\Lambda), \|.\|_\infty)$, where $\mathcal{C}(\Lambda)$ is the class of continuous functions on Λ and $\|\cdot\|_\infty$ is the uniform norm (see, e.g., Billingsley, 1999; van der Vaart and Wellner, 2000). Our test statistic $\widehat{\mathcal{IT}}_{1n}$ is the integrated squared $\hat{\mathcal{T}}_{1n}(.)$:

$$\widehat{\mathcal{IT}}_{1n} \equiv \int_\Lambda \left[\hat{\mathcal{T}}_{1n}(\lambda)\right]^2 dv(\lambda).$$

Our test is similar to Bierens's (1990) and Bierens and Ploberger's (1997) integrated conditional moment type of test and the Cramer–von Mises test. To derive the asymptotic distribution of $\widehat{\mathcal{IT}}_{1n}$, we impose some further regularity conditions. These are similar to those of Newey (1994) and Altonji and Matzkin (2005).

Assumption B.1. (Kernel, similar to Assumption 3.3 in Altonji and Matzkin, 2005) The Kernel function $\mathbf{K}:\mathbb{R}\to\mathbb{R}$ is continuously differentiable of order $\Delta\geq 1$, with Δth order derivative that is Lipschitz and bounded. $\mathbf{K}$ vanishes outside a compact set. $\mathbf{K}$ is a kernel of order $s\geq 2$, so that $\int \mathbf{K}(u)\,du = 1$, $\int u^j\mathbf{K}(u)\,du = 0$, $j = 2,\ldots,s-1$, and $\int u^s\mathbf{K}(u)\,du < \infty$.

Assumption B.2. (Density, similar to Assumption 3.5 in Altonji and Matzkin, 2005) Let Δ and s be as in B.1. The density $f(x,z)$ is continuously differentiable up to order $q\geq\Delta+s$ with respect to both x and z and the derivatives of $f(x,z)$ of order q are Lipschitz and bounded. The density $f(x,y,z)$ and integral $\int y\cdot f(y,x,z)\cdot dy$ are continuously differentiable of order of q with respect to x and z; for some $p>2$, $E\left(|Y|^p \mid X=x, Z=z\right)$ is bounded for $(x,z)\in\Theta_{x,z}$; the density $f(x)$ is bounded away from 0 on $\left[\underline{x},\bar{x}\right]$; $\psi(\cdot,\lambda):\mathbb{X}\to\mathbb{R}$ is continuously differentiable of order $s+1$ for all $\lambda\in\Lambda$, and the derivative of $\psi(\cdot,\lambda)$ of order $s+1$ is Lipschitz and bounded.

Assumption B.3. (Bandwidth and trimming value, similar to Assumption 3.6 in Altonji and Matzkin, 2005) Let s and p be as in B.1 and B.2. As $n\to\infty$, $h_n\to 0$, $t_n\to 0$,

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

$$\sqrt{n}h_n^s \to 0,$$

$$t_n^{-3} \cdot \sqrt{n}\left(\sqrt{\ln(n)/\left(nh_n^{k_z+3}\right)} + h_n^s\right)^2 \to 0,$$

$$\sqrt{n}\left(\sqrt{\ln(n)/\left(nh_n^{k_z+3}\right)} + h_n^s\right) \cdot \left[\int\int \mathbf{1}[0 < f(x,z) < 2t_n] \cdot dz \cdot dx\right]^{\frac{1}{2}} \to 0,$$

and

$$\left(n^{1-\frac{2}{p}} h_n^{1+k_z} \Big/ \ln(n)\right) \to \infty.$$

Assumption B.4.

$$\int\int \left[w(x) \cdot f_x(x,z) \cdot f(x,z)^{-1}\right]^2 \cdot dx \cdot dz < \infty;$$

$$\int\int w(x) \cdot f(x,z)^{-1} \cdot dx \cdot dz < \infty;$$

and as $n \to \infty$,

$$\sqrt{n}\int \mathbf{1}[0 < f(x,z) < 2t_n] \cdot w(x) \cdot |\mu(x,z)| \cdot |f_x(x,z)| \cdot f(x)^{-1} \cdot dz \cdot dx \to 0.$$

THEOREM 3. *Suppose Assumptions A.1–A.4 and B.1–B.4 hold. Then*

(a) $\sqrt{n}\left[\hat{\mathcal{T}}_{1n}(\cdot) - \mathcal{T}_{10}(\cdot)\right]$ *converges weakly to* $\mathcal{T}_{1\infty}$ *in* $(\mathcal{C}(\Lambda), \|.\|_\infty)$, *where* $\mathcal{T}_{1\infty}(\cdot)$ *is a zero mean Gaussian process on* Λ *with covariance (kernel) function given by*

$$cov\left[\mathcal{T}_{1\infty}(\lambda), \mathcal{T}_{1\infty}(\lambda')\right] = cov\left[\kappa_1(X,Y,Z,\lambda), \kappa_1(X,Y,Z,\lambda')\right],$$

where

$$\kappa_1(x,y,z,\lambda) = \left[\frac{y-\mu(x,z)}{f(x)}\right] \cdot \frac{\partial}{\partial x}\psi(x,\lambda) + \frac{\partial}{\partial x}\left[\frac{y-\mu(x,z)}{f(x)}\right] \cdot \psi(x,\lambda) + \left[\frac{(y-\mu(x,z)) \cdot f_x(x,z)}{f(x,z) \cdot f(x)} + \frac{\beta_0(x)}{f(x)}\right] \cdot \psi(x,\lambda). \quad \textbf{(3.5)}$$

(b) *Under* $\mathbb{H}_{10}$, $\sqrt{n}\hat{\mathcal{T}}_{1n}(\cdot)$ *converges weakly to* $\mathcal{T}_{1\infty}$ *in* $(\mathcal{C}(\Lambda), \|\cdot\|_\infty)$ *and* $n \cdot \widehat{\mathcal{IT}}_{1n}$ *converges to* $\int_\Lambda [\mathcal{T}_{1\infty}(\lambda)]^2 d\nu(\lambda)$ *in distribution.*

We define a local alternative as:

$$\mathbb{H}_{1a} : Y = r(X,U) + \frac{\tilde{r}(X,U)}{\sqrt{n}},$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

such that there exists a constant c_0 such that

$$\Pr[\mathbf{1}[X \in I_x] \cdot (\beta_0(X) - c_0) = 0] = 1,$$

and for all $c \in \mathbb{R}$,

$$\Pr\left[\mathbf{1}[X \in I_x] \cdot \left(\tilde{\beta}_0(X) - c\right) = 0\right] < 1,$$

where

$$\tilde{\beta}_0(x) \equiv E\left[\frac{\tilde{r}(x, U)}{\partial x}\middle| X = x\right].$$

The next theorem gives the asymptotic distribution of $\sqrt{n}\hat{\mathcal{T}}_{1n}(\cdot)$ and $n \cdot \widehat{\mathcal{IT}}_{1n}$ under the local alternative $\mathbb{H}_{1a}$ and the global alternative $\mathbb{H}_{1A}$.

THEOREM 4. *Suppose Assumptions A.1–A.4 and B.1–B.4 hold. Then*

(a) *Under* $\mathbb{H}_{1a}$, $\sqrt{n}\hat{\mathcal{T}}_{1n}(\cdot)$ *converges weakly to* $\tilde{\mathcal{T}}_{1\infty}(\cdot) \equiv \mathcal{T}_{1\infty}(\cdot) + C_{10}(\cdot)$ *in* $(\mathcal{C}(\Lambda), \|.\|_\infty)$, *where*

$$C_{10}(\lambda) \equiv \int \psi(x, \lambda) \cdot \tilde{\beta}_0(x) \cdot dx,$$

and $C_{10}(\lambda) \neq 0$ *for* ν*-almost all* $\lambda \in \Lambda$. $n \cdot \widehat{\mathcal{IT}}_{1n}$ *converges weakly to* $\int [\mathcal{T}_{1\infty}(\lambda) + C_{10}(\lambda)]^2 \cdot d\nu(\lambda)$.

(b) *Under* $\mathbb{H}_{1A}$, *for any nonstochastic scalar sequence* $\{k_n\} = o\left(\sqrt{n}\right)$,

$$\Pr\left[\left|\sqrt{n}\hat{\mathcal{T}}_{1n}(\lambda)\right| > k_n\right] \to 1$$

for ν*-almost all* $\lambda \in \Lambda$; *and*

$$\Pr\left[n \cdot \widehat{\mathcal{IT}}_{1n} > k_n^2\right] \to 1.$$

Similar to Bierens and Ploberger (1997) and Chen and Fan (1999), one can show that $\int_\Lambda [\mathcal{T}_{1\infty}(\lambda)]^2 d\nu(\lambda)$ is an infinite sum of weighted χ_1^2 random variables with the weight depending on the true data generating process. Thus, the asymptotic distribution is nonstandard and difficult to tabulate. We can use either subsampling or bootstrap to simulate the asymptotic null distribution.

Specifically, for subsampling, we let $b \equiv b_n$ define a sequence of positive integers such that $b/n \to 0$ and $b \to \infty$ as $n \to \infty$. We simulate the null distribution in the following steps as in Politis, Romano, and Wolf (1999):

ALGORITHM 1 (Subsampling).

Step 1: Let B be a large integer. Randomly draw B subsamples $\left\{X^{*(1)}, Y^{*(1)}, Z^{*(1)}\right\}, \ldots, \left\{X^{*(B)}, Y^{*(B)}, Z^{*(B)}\right\}$ *from all possible subsamples of size b of the sample population* $\{X_i, Y_i, Z_i\}_{i=1}^n$.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

Step 2: *For* $k = 1, \ldots B$, *compute* $\hat{\mathcal{T}}_{1n}(\lambda)$ *using the subsample* $\left\{X^{*(k)}, Y^{*(k)}, Z^{*(k)}\right\}$ *and denote it as* $\hat{\mathcal{T}}^{*(k)}_{1n,b.}(\lambda)$.

Step 3: *The subsampling asymptotic null distribution is given by:*

$$F^{*}_{1n,b}(x) = \frac{1}{B}\sum_{k=1}^{B} \mathbf{1}\left\{ b\left[\int_{\Lambda} \left[\hat{\mathcal{T}}^{*(k)}_{1n,b.}(\lambda) - \hat{\mathcal{T}}_{1n}(\lambda) \right]^2 d\nu(\lambda) \right] \leq x \right\},$$

and the subsampling $100 \times (1-\alpha)\%$ *quantile is given by:*

$$c^{*}_{1n,b,1-\alpha} = \inf\left\{ x : F^{*}_{1n,b}(x) \geq 1-\alpha \right\}.$$

By the standard argument (see, e.g., Politis et al., 1999), we can show that subsampling works to obtain the null distribution. We may also apply the weighted bootstrap as in Hansen (1996). Specifically, we simulate the null distribution of $\hat{\mathcal{T}}_{1n}(\lambda)$ as $\hat{\mathcal{T}}^{*}_{1n}(\lambda)$:

$$\hat{\mathcal{T}}^{*}_{1n}(\lambda) \equiv \frac{1}{n}\sum_{i=1}^{n} \hat{\kappa}_1(X_i, Y_i, Z_i, \lambda) \cdot \epsilon^{*}_i,$$

where $\hat{\kappa}_1$ is an estimator of κ_1 in equation (3.5) and $\left\{\epsilon^{*}_i\right\}_{i=1}^{n}$ is i.i.d. (independent and identically distributed) standard normal random variables. It is straightforward to establish that $\hat{\mathcal{T}}^{*}_{1n}(\lambda)$ has the same limiting distribution as $\hat{\mathcal{T}}_{1n}(\lambda)$ under the null. Thus, the null distribution of $n\widehat{\mathcal{IT}}_{1n}$ can be simulated as

$$n\widehat{\mathcal{IT}}^{*}_{1n} \equiv n\int_{\Lambda} \left[\hat{\mathcal{T}}^{*}_{1n}(\lambda) \right]^2 d\nu(\lambda).$$

However, this approach requires to estimate κ_1 that is quite complicated. Also, the finite sample performance of this bootstrap may not be good, as it crucially depends on the normal approximation of $\hat{\mathcal{T}}_{1n}(\lambda)$. For these reasons, we do not follow this approach in our simulations.

We also expect that the nonparametric bootstrap works here. Nevertheless, providing a rigorous proof for the bootstrap is beyond the scope of this paper. In fact, the simulations of Section 4 show that subsampling and the nonparametric bootstrap do provide similar results.

3.3. Testing $\mathbb{H}_{20}$ (Constant Effects for APE)

To identify $\gamma_0(x)$, we impose one more assumption.

Assumption B.5. (Common Support) For all $x \in \mathbb{X}$, the support of Z conditional on $X = x$ equals the support of Z.

Under A.1, A2, and B.5, $\gamma_0(x)$ is identified as

$$\gamma_0(x) = \int \frac{\partial \mu(x,z)}{\partial x} \cdot f(z) \cdot dz.$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

B.5 is imposed to ensure that $\mu(x,z)$ is well defined for all z in the support of Z, so that we can identify $\gamma_0(x)$ by taking the integral over the whole support of Z. Note that B.5 is not required to identify $\beta_0(x)$. For detailed discussions of this assumption, see Blundell and Powell (2003, Sect. 2.2.4, p. 334), Florens et al. (2008, Sect. 3.2, p. 1199), and Imbens and Newey (2009, p. 1487).

Testing $\mathbb{H}_{20}$ is analogous to testing $\mathbb{H}_{10}$. The next lemma motivates our test of $\mathbb{H}_{20}$.

LEMMA 5. *Let Assumption A.1 hold. Suppose $f(x)$ is bounded away from zero on $[\underline{x}, \bar{x}]$. Then $\mathbb{H}_{20}$ holds if and only if $\mathcal{T}_{20}(\lambda) = 0$ for ν-almost all $\lambda \in \Lambda$, where*

$$\mathcal{T}_{20}(\lambda) \equiv \int \psi(x,\lambda) \cdot \gamma_0(x) \cdot dx,$$

with ψ as in Lemma 2.

Remark 2. Similarly, $\mathbb{H}'_{20}$ holds if and only if $\mathcal{T}'_{20}(\lambda) = 0$ for ν-almost all $\lambda \in \Lambda$, where

$$\mathcal{T}'_{20}(\lambda) \equiv \int \psi'(x,\lambda) \cdot \gamma_0(x) \cdot dx,$$

with ψ' as in Remark 1. In this subsection, we focus on testing $\mathbb{H}_{20}$. All the results hold for $\mathbb{H}'_{20}$ if we simply replace ψ with ψ'.

Under A.1, A.2, and B.5, integration by parts yields

$$\mathcal{T}_{20}(\lambda) = -\int\int \frac{\left(\int y \cdot f(x,y,z) \cdot dy\right) \cdot f(z)}{f(x,z)} \cdot \frac{\partial}{\partial x}[\psi(x,\lambda)] \cdot dz \cdot dx.$$

Then a sample analog of $\mathcal{T}_{20}(\lambda)$ is :

$$\hat{\mathcal{T}}_{2n}(\lambda) \equiv -\int\int \left(\int y \cdot \hat{f}(x,y,z) \cdot dy\right) \cdot \tau\left(\hat{f}(x,z), t\right) \cdot \hat{f}(z) \cdot \frac{\partial}{\partial x}[\psi(x,\lambda)] \cdot dz \cdot dx,$$

where $\hat{f}(x,y,z)$, $\hat{f}(z)$, and $\hat{f}(x,z)$ are kernel estimators of $f(x,y,z)$, $f(z)$, and $f(x,z)$ respectively, and τ is the trimming function (equation (3.3)). We also view $\hat{\mathcal{T}}_{2n}(\cdot)$ as a random element in the metric space $(\mathcal{C}(\Lambda), \|\cdot\|_\infty)$. Our test statistic $\widehat{\mathcal{IT}}_{2n}$ is the integrated squared $\hat{\mathcal{T}}_{2n}(\cdot)$:

$$\widehat{\mathcal{IT}}_{2n} \equiv \int_\Lambda \left[\hat{\mathcal{T}}_{2n}(\lambda)\right]^2 d\nu(\lambda).$$

We have results analogous to those for $\mathbb{H}_{10}$. We impose the same regularity conditions as in testing $\mathbb{H}_{10}$, except that we replace B.4 with B.6.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

Assumption B.6.

$$\int\int\left[f(z)\cdot f(x,z)^{-1}\cdot\frac{\partial}{\partial x}w(x)\right]^2\cdot dx\cdot dz<\infty;$$

$$\int\int f(x,z)^{-1}\cdot\left|\frac{\partial}{\partial x}w(x)\right|\cdot dx\cdot dz<\infty;$$

and as $n\to\infty$,

$$\sqrt{n}\int\int\mathbf{1}[f(x,z)<2t_n]\cdot|\mu(x,z)|\cdot f(z)\cdot\left|\frac{\partial}{\partial x}w(x)\right|\cdot dx\cdot dz\to 0.$$

THEOREM 6. *Suppose Assumptions A.1–A.4, B.1–B.3, B.5, and B.6 hold. Then*

(a) $\sqrt{n}\left[\hat{\mathcal{T}}_{2n}(\cdot)-\mathcal{T}_{20}(\cdot)\right]$ *converges weakly to* $\mathcal{T}_{2\infty}(\cdot)$ *in* $(\mathcal{C}(\Lambda),\|\cdot\|_\infty)$, *where* $\mathcal{T}_{2\infty}(\cdot)$ *is a zero mean Gaussian process on* Λ *with covariance (kernel) function given by*

$$cov\left[\mathcal{T}_{2\infty}(\lambda),\mathcal{T}_{2\infty}(\lambda')\right]=cov\left[\kappa_2(X,Y,Z,\lambda),\kappa_2(X,Y,Z,\lambda')\right],$$

where

$$\kappa_2(x,y,z,\lambda)\equiv-[y-\mu(x,z)]\cdot\frac{f(z)}{f(x,z)}\cdot\frac{\partial}{\partial x}\psi(x,\lambda)+\eta_2(z,\lambda),$$

$$\eta_2(z,\lambda)\equiv-\int\mu(x,z)\cdot\frac{\partial}{\partial x}\psi(x,\lambda)\cdot dx.$$

(b) *Under* $\mathbb{H}_{20}$, $\sqrt{n}\hat{\mathcal{T}}_{2n}(\cdot)$ *converges weakly to* $\mathcal{T}_{2\infty}(\cdot)$ in $(\mathcal{C}(\Lambda),\|\cdot\|_\infty)$ *and* $n\cdot\widehat{\mathcal{IT}}_{2n}$ *converges to* $\int_\Lambda[\mathcal{T}_{2\infty}(\lambda)]^2\,dv(\lambda)$ *in distribution.*

We define a local alternative as:

$$\mathbb{H}_{2a}:Y=r(X,U)+\frac{\tilde{r}(X,U)}{\sqrt{n}},$$

such that there exists a constant c_0 such that

$$\Pr[\mathbf{1}[X\in I_x]\cdot(\gamma_0(X)-c_0)=1]=1,$$

and for all $c\in\mathbb{R}$,

$$\Pr\left[\mathbf{1}[X\in I_x]\cdot(\tilde{\gamma}_0(X)-c)\right]<1,$$

where

$$\tilde{\gamma}_0(x)\equiv E\left[\frac{\partial\tilde{r}(x,U)}{\partial x}\right].$$

The next theorem gives the asymptotic behavior of $\sqrt{n}\hat{\mathcal{T}}_{2n}(\cdot)$ and $n\cdot\widehat{\mathcal{IT}}_{2n}$ under the local alternative $\mathbb{H}_{2a}$ and the global alternative $\mathbb{H}_{2A}$.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

THEOREM 7. *Suppose Assumptions A.1–A.4, B.1–B.3, B.5, and B.6 hold. Then (a) Under* $\mathbb{H}_{2a}$, $\sqrt{n}\hat{\mathcal{T}}_{2n}(\cdot)$ *converges weakly to* $\tilde{\mathcal{T}}_{2\infty}(\cdot) \equiv \mathcal{T}_{2\infty}(\cdot) + C_{20}(\cdot)$ *in* $(\mathcal{C}(\Lambda), \|\cdot\|_\infty)$, *where*

$$C_{20}(\lambda) = \int \psi(x, \lambda) \cdot \tilde{\gamma}_0(x) \cdot dx,$$

and $C_{20}(\lambda) \neq 0$ *for* ν*-almost all* $\lambda \in \Lambda$. $n \cdot \widehat{\mathcal{IT}}_{2n}$ *converges weakly to* $\int [\mathcal{T}_{2\infty}(\lambda) + C_{20}(\lambda)]^2 \cdot d\nu(\lambda)$. *(b) Under* $\mathbb{H}_{2A}$, *for any nonstochastic scalar sequence* $\{k_n\} = o(\sqrt{n})$,

$$\Pr\left[\left|\sqrt{n}\hat{\mathcal{T}}_{2n}(\lambda)\right| > k_n\right] \to 1$$

for ν*-almost all* $\lambda \in \Lambda$; *and*

$$\Pr\left[n \cdot \widehat{\mathcal{IT}}_{2n} > k_n^2\right] \to 1.$$

We also use subsampling or bootstrap to obtain the critical value.

ALGORITHM 2. *Identical to Algorithm 1, but replacing* $\hat{\mathcal{T}}_{1n}(\lambda)$, $\hat{\mathcal{T}}^{*(k)}_{1n,b.}(\lambda)$, $F^*_{1n,b}(x)$, *and* $c^*_{1n,b,1-\alpha}$ *with* $\hat{\mathcal{T}}_{2n}(\lambda)$, $\hat{\mathcal{T}}^{*(k)}_{2n,b}(\lambda)$, $F^*_{2n,b}(x)$, *and* $c^*_{2n,b,1-\alpha}$, *respectively.*

3.4. Implementing the Tests

To implement these tests, we must choose λ from a compact set Λ with measure ν and choose the bandwidth for kernel estimators. Providing formal analysis of these choices is beyond the scope of this paper. Nevertheless, we provide some practical guidance. The guidance is similar to that in Huang, Sun, and White (2014) and Lu and White (2014).

First, although as stated in Lemma 1, choosing λ from any compact set Λ delivers a consistent test, finite sample performance depends on the choice of Λ. As argued in White (2006) and Huang et al. (2013), we should choose λ to avoid $|\lambda_0 + \lambda_1 X|$ being too large or too small. For example, let φ be a normal density function. Then $\varphi(u)$ is close to zero if $|u|$ is too large and close to linear if $|u|$ is too small.

We choose ν to be the uniform probability measure, as Boning and Sowell (1999) show that this measure is optimal in the closely related specification testing literature. Specifically, as suggested in Huang et al. (2013), we choose λ in Lemma 1 as:

$$\lambda_0 \sim \text{uniform}\left(\lambda_0^* - 0.5,\ \lambda_0^* + 0.5\right) \text{ and } \lambda_1 \sim \text{uniform}\left(\lambda_1^* - 0.5,\ \lambda_1^* + 0.5\right), \quad \textbf{(3.6)}$$

where

$$\lambda_0^* = 1 - \frac{\bar{X}}{std(X)} \quad \text{and} \quad \lambda_1^* = \frac{\bar{X}}{std(X)},$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

with $\bar{X}$ the sample averages and $std(X)$ the sample standard deviations of X, respectively.

Second, we need to choose the bandwidth. We can use a simple "rule of thumb". Specifically, we can let $h = C \cdot [std(X), std(Z)] \cdot n^{\alpha}$, where C is a constant, $std(X)$ and $std(Z)$ are sample standard deviations of X and Z, respectively, and α is the convergence rate of the bandwidth. In the simulation where $k_z = 1$ and the order of kernel s is 6, we set $\alpha = -1/9$. We let $C = c \cdot 3.84$, where 3.84 is the constant associated with the sixth order quartic kernel and c takes a range of values close to 1. Alternatively, we can follow the method proposed by Newey (1994). The first step is to start at a value obtained by cross-validation (minimizing the asymptotic mean square error of the kernel estimator of the conditional mean function). The second step is to decrease the bandwidth until the test statistics do not change much.

Next, we have so far required that the conditioning variables Z be continuous. Nevertheless, in applications, Z may be discrete. Discrete Z does not cause any difficulty in testing $\mathbb{H}_{10}$, $\mathbb{H}_{20}$, $\mathbb{H}'_{10}$, or $\mathbb{H}'_{20}$. We can simply replace the integral with the summation in the test statistics. More sophisticated methods can also be applied, such as smoothing the discrete variables (see, e.g., Li and Racine, 2003; Huang, 2009).

Finally, numerical integration is required to compute the test statistics. There are many applicable methods. For low dimension cases, we apply Gaussian quadrature; we apply Monte Carlo integration in higher dimensions. Alternatively, one may replace the integral with summations over the data points. For example, we can write

$$\mathcal{T}_{10}(\lambda) \equiv \int \psi(x, \lambda) \cdot \beta_0(x) \cdot dx = E\left[\frac{\psi(X, \lambda) \cdot \beta_0(X)}{f(X)}\right],$$

which can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\psi(X_i, \lambda) \cdot \hat{\beta}(X_i)}{\hat{f}(X)}.$$

However, this approach introduces the estimator of the density in the denominator.

4. MONTE CARLO SIMULATIONS

In this section, we examine the finite sample performance of our test statistics. We first consider data generating process (DGP) 1:

$$\text{DGP } 1.0: Y = U_1,$$
$$\text{DGP } 1.1: Y = 1 + X + U_1,$$
$$\text{DGP } 1.2: Y = 1 + X + U_1 + U_2 \cdot \left[X + X^2 - cos(\pi X)\right],$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

where

$X = Z + Z^2 + \varepsilon_x$ with X truncated to $[-2, 2]$; $U_1 = Z + 0.5\varepsilon_1$; $U_2 = 1 + Z + 0.5\varepsilon_2$; $Z \sim N(0,1)$ truncated to $[-2, 2]$; $\varepsilon_x, \varepsilon_1, \varepsilon_2 \sim N(0,1)$.

We consider $\mathbb{H}_{10}$ (constant effects for LAR), $\mathbb{H}_{20}$ (constant effects for APE), $\mathbb{H}'_{10}$ (zero effects for LAR), and $\mathbb{H}'_{20}$ (zero effects for APE). By design, DGP 1.0 satisfies all four null hypotheses $\mathbb{H}_{10}$, $\mathbb{H}_{20}$, $\mathbb{H}'_{10}$, and $\mathbb{H}'_{20}$; DGP 1.1 satisfies the null hypotheses $\mathbb{H}_{10}$ and $\mathbb{H}_{20}$ and the alternatives $\mathbb{H}'_{1A}$ and $\mathbb{H}'_{2A}$; and DGP 1.2 satisfies all four alternatives $\mathbb{H}_{1A}$, $\mathbb{H}_{2A}$, $\mathbb{H}'_{1A}$, and $\mathbb{H}'_{2A}$, We use a sixth order quartic kernel for kernel estimators. The GCR function φ is the standard normal density as in Huang et al. (2013). The sample size is $n = 200$. For subsampling, we consider subsample sizes $b = \lfloor n^{0.9} \rfloor$ or $b = \lfloor n^{0.95} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part of $\cdot$. When $n = 200$, b equals 117 or 153. As a comparison, we also use the nonparametric bootstrap to simulate the null distributions. As discussed in Section 3.4, we use a simple "rule of thumb" for the bandwidth: $h = c \cdot 3.84 \cdot [std(X), std(Z)] \cdot n^{-1/9}$, where 3.84 is the constant associated with the sixth order quartic kernel. We choose a range of values of c close to 1 to examine whether performance is sensitive to these. Specifically, c equals 0.7, 0.8,..., and 1.2. The choice of λ follows equation (3.6). The number of random draws for subsampling and the bootstrap is $B = 200$. The number of Monte Carlo replications is $M = 400$.

The simulation results appear in Tables 1–4. The tests perform well for all four hypotheses. The simulated levels are all fairly close to the nominal level of 5%. For $\mathbb{H}_{10}$ and $\mathbb{H}_{20}$ in DGP 1.1, the levels are slightly oversized when the constant c is large. Test powers are excellent with the rejection frequencies being above 95% for all cases. In general, the performance in terms of level and power is not very sensitive to the constant c in the bandwidth choice and the subsample size b.

TABLE 1. Rejection frequencies ($\mathbb{H}_{10}$: constant effects for LAR, $n = 200$)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
DGP 1.0	Subsample $b = \lfloor n^{0.90} \rfloor$	0.065	0.063	0.058	0.050	0.065	0.060
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.060	0.060	0.055	0.050	0.045	0.048
	Bootstrap	0.043	0.053	0.050	0.038	0.038	0.038
DGP 1.1	Subsample $b = \lfloor n^{0.90} \rfloor$	0.055	0.078	0.078	0.073	0.083	0.088
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.045	0.063	0.075	0.070	0.078	0.078
	Bootstrap	0.045	0.065	0.060	0.068	0.073	0.075
DGP 1.2	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	0.998	0.998	0.995
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	0.998	0.995	0.995
	Bootstrap	0.998	0.998	1.000	0.998	0.993	0.995

Note: We let $\underline{x} = -1$ and $\bar{x} = 1$. The nominal level is 0.05. The bandwidth is chosen as $h = c \cdot 3.84 \cdot [std(X), std(Z)] \cdot n^{-1/9}$.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

TABLE 2. Rejection frequencies ($\mathbb{H}_{20}$: constant effects for APE, $n = 200$)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
DGP 1.0	Subsample $b = \lfloor n^{0.90} \rfloor$	0.038	0.033	0.028	0.025	0.033	0.040
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.025	0.025	0.025	0.023	0.025	0.043
	Bootstrap	0.023	0.030	0.033	0.020	0.020	0.025
DGP 1.1	Subsample $b = \lfloor n^{0.90} \rfloor$	0.030	0.043	0.030	0.060	0.073	0.080
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.035	0.028	0.020	0.045	0.070	0.088
	Bootstrap	0.035	0.038	0.035	0.055	0.063	0.075
DGP 1.2	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	0.998	1.000	0.998	0.988	0.970
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	0.995	0.998	0.998	0.993	0.968
	Bootstrap	1.000	1.000	0.998	0.993	0.985	0.960

Note: See the note to Table 1.

TABLE 3. Rejection frequencies ($\mathbb{H}'_{10}$: zero effects for LAR, $n = 200$)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
DGP 1.0	Subsample $b = \lfloor n^{0.90} \rfloor$	0.033	0.050	0.065	0.065	0.068	0.070
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.035	0.038	0.045	0.045	0.063	0.065
	Bootstrap	0.023	0.035	0.043	0.038	0.050	0.055
DGP 1.1	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000
DGP 1.2	Subsample $b = \lfloor n^{0.90} \rfloor$	0.998	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.998	1.000	1.000	1.000	1.000	1.000
	Bootstrap	0.998	1.000	1.000	1.000	1.000	1.000

Note: See the note to Table 1.

For this DGP, we also find that the performance of the nonparametric bootstrap is similar to that of subsampling.

Next, we consider DGP 2:

$$\text{DGP 2.0}: Y = U,$$
$$\text{DGP 2.1}: Y = X + U,$$
$$\text{DGP 2.2}: Y = X + U + U \cdot \left(X + X^2\right),$$

where

$$X = 0.5 \cdot Z + 2 \cdot \varepsilon_x \text{ with } X \text{ truncated to } [-2, 2];\ U = 0.5 \cdot Z + \sqrt{Z + 2} + 0.5 \cdot \varepsilon_1;$$
$$Z \sim N(0, 1) \text{ truncated by } [-2, 2],\ \varepsilon_x, \varepsilon_1 \sim N(0, 1).$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

TABLE 4. Rejection frequencies ($\mathbb{H}'_{20}$: zero effects for APE, $n = 200$)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
DGP 1.0	Subsample $b = \lfloor n^{0.90} \rfloor$	0.035	0.043	0.048	0.045	0.053	0.055
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.045	0.038	0.038	0.045	0.055	0.048
	Bootstrap	0.035	0.030	0.035	0.035	0.033	0.045
DGP 1.1	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000
DGP 1.2	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000

Note: See the note to Table 1.

Again, by design, DGP 2.0 satisfies $\mathbb{H}_{10}$, $\mathbb{H}_{20}$, $\mathbb{H}'_{10}$, and $\mathbb{H}'_{20}$; DGP 2.1 satisfies $\mathbb{H}_{10}$, $\mathbb{H}_{20}$, $\mathbb{H}'_{1A}$, and $\mathbb{H}'_{2A}$; and DGP 2.2 satisfies $\mathbb{H}_{1A}$, $\mathbb{H}_{2A}$, $\mathbb{H}'_{1A}$, and $\mathbb{H}'_{2A}$. For this DGP, test performance is also good. The results appear in Tables 5–8. The levels behave reasonably in general. But for $\mathbb{H}_{20}$ (Table 6), the level seems to be slightly undersized especially when the constant c is small. Compared with DGP 1, the powers against $\mathbb{H}_{10}$ and $\mathbb{H}_{20}$ are worse and more sensitive to the bandwidth, but they are in reasonable ranges; especially when the bandwidths are relatively large, the rejection frequencies are above 95%. We also note that in terms of the powers against $\mathbb{H}_{10}$ and $\mathbb{H}_{20}$, the nonparametric bootstrap slightly outperforms subsampling. The powers against $\mathbb{H}'_{10}$ and $\mathbb{H}'_{20}$ are excellent with all rejection frequencies being 1.

TABLE 5. Rejection frequencies ($\mathbb{H}_{10}$: constant effects for LAR, $n = 200$)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
DGP 2.0	Subsample $b = \lfloor n^{0.90} \rfloor$	0.040	0.073	0.068	0.063	0.075	0.078
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.043	0.060	0.063	0.060	0.065	0.068
	Bootstrap	0.030	0.053	0.058	0.050	0.055	0.045
DGP 2.1	Subsample $b = \lfloor n^{0.90} \rfloor$	0.038	0.068	0.063	0.058	0.058	0.058
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.050	0.078	0.070	0.063	0.078	0.053
	Bootstrap	0.040	0.068	0.070	0.050	0.063	0.043
DGP 2.2	Subsample $b = \lfloor n^{0.90} \rfloor$	0.765	0.810	0.910	0.973	0.998	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.795	0.825	0.920	0.975	1.000	1.000
	Bootstrap	0.800	0.830	0.933	0.988	1.000	1.000

Note: See the note to Table 1.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

TABLE 6. Rejection frequencies ($\mathbb{H}_{20}$: constant effects for APE, $n = 200$)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
DGP 2.0	Subsample $b = \lfloor n^{0.90} \rfloor$	0.015	0.030	0.030	0.035	0.038	0.043
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.015	0.025	0.020	0.030	0.030	0.038
	Bootstrap	0.013	0.023	0.020	0.020	0.028	0.035
DGP 2.1	Subsample $b = \lfloor n^{0.90} \rfloor$	0.013	0.020	0.035	0.038	0.040	0.040
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.018	0.020	0.035	0.030	0.043	0.035
	Bootstrap	0.023	0.025	0.040	0.040	0.035	0.033
DGP 2.2	Subsample $b = \lfloor n^{0.90} \rfloor$	0.625	0.690	0.800	0.943	0.993	0.998
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.655	0.720	0.845	0.968	0.995	1.000
	Bootstrap	0.668	0.728	0.888	0.968	0.998	1.000

Note: See the note to Table 1.

TABLE 7. Rejection frequencies ($\mathbb{H}'_{10}$: zero effects for LAR, $n = 200$)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
DGP 2.0	Subsample $b = \lfloor n^{0.90} \rfloor$	0.040	0.070	0.083	0.075	0.098	0.088
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.045	0.058	0.060	0.068	0.083	0.078
	Bootstrap	0.043	0.045	0.055	0.050	0.070	0.078
DGP 2.1	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000
DGP 2.2	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000

Note: See the note to Table 1.

TABLE 8. Rejection frequencies ($\mathbb{H}'_{20}$: zero effects for APE, $n = 200$)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
DGP 2.0	Subsample $b = \lfloor n^{0.90} \rfloor$	0.035	0.053	0.068	0.075	0.065	0.063
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.040	0.053	0.060	0.073	0.073	0.063
	Bootstrap	0.038	0.048	0.065	0.065	0.058	0.045
DGP 2.1	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000
DGP 2.2	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000

Note: See the note to Table 1.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

Finally, to further examine the power of the tests, we consider an alternative DGP similar to DGP 2.2:

$$\text{DGP } 2.2' : Y = X + U + 0.3 \cdot U \cdot \left(X + X^2\right).$$

DGP 2.2′ is identical to DGP 2.2, except that the coefficient on the term $U \cdot \left(X + X^2\right)$ is smaller. Thus, compared with DGP 2.2, for $\mathbb{H}_{10}$ and $\mathbb{H}_{20}$, DGP 2.2′ is an alternative that is closer to the null. The results for testing DGP 2.2′ appear in Tables 9–12. We first let the sample size $n = 200$ as before. Not surprisingly, the powers against $\mathbb{H}_{01}$ and $\mathbb{H}_{02}$ deteriorate compared with those in DGP 2.2. In fact, the rejection frequencies are below 30% when the bandwidths are small. However, if we increase the sample size to $n = 500$, the power values against $\mathbb{H}_{01}$ and $\mathbb{H}_{02}$ increase substantially and become comparable to those seen for DGP 2.2. The rejection frequencies can reach 90% for relatively large bandwidths. For $\mathbb{H}'_{01}$ and $\mathbb{H}'_{02}$, our tests perform well even with $n = 200$, with all rejection frequencies being 1.

TABLE 9. Rejection frequencies ($\mathbb{H}_{10}$: constant effects for LAR, DGP2.2′)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
$n = 200$	Subsample $b = \lfloor n^{0.90} \rfloor$	0.268	0.335	0.415	0.545	0.730	0.793
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.288	0.333	0.420	0.568	0.720	0.768
	Bootstrap	0.283	0.333	0.430	0.570	0.698	0.755
$n = 500$	Subsample $b = \lfloor n^{0.90} \rfloor$	0.678	0.688	0.730	0.810	0.885	0.965
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.688	0.703	0.743	0.820	0.900	0.965
	Bootstrap	0.698	0.708	0.748	0.830	0.898	0.970

Note: See the note to Table 1.

TABLE 10. Rejection frequencies ($\mathbb{H}_{20}$: constant effects for APE, DGP2.2′)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
$n = 200$	Subsample $b = \lfloor n^{0.90} \rfloor$	0.113	0.200	0.288	0.435	0.558	0.670
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.130	0.213	0.290	0.468	0.598	0.678
	Bootstrap	0.143	0.213	0.310	0.473	0.615	0.668
$n = 500$	Subsample $b = \lfloor n^{0.90} \rfloor$	0.540	0.598	0.683	0.775	0.870	0.950
	Subsample $b = \lfloor n^{0.95} \rfloor$	0.565	0.620	0.680	0.788	0.888	0.965
	Bootstrap	0.595	0.630	0.703	0.793	0.893	0.965

Note: See the note to Table 1.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

TABLE 11. Rejection frequencies ($\mathbb{H}'_{10}$: zero effects for LAR, DGP2.2′)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
$n = 200$	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000
$n = 500$	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000

Note: See the note to Table 1.

TABLE 12. Rejection frequencies ($\mathbb{H}'_{20}$: zero effects for APE, DGP2.2′)

	Bandwidth constant c	0.7	0.8	0.9	1.0	1.1	1.2
$n = 200$	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000
$n = 500$	Subsample $b = \lfloor n^{0.90} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Subsample $b = \lfloor n^{0.95} \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000
	Bootstrap	1.000	1.000	1.000	1.000	1.000	1.000

Note: See the note to Table 1.

5. AN ILLUSTRATIVE APPLICATION: INTEREST RATE ELASTICITIES FOR LOAN DEMAND

We apply our testing procedures to the dataset used in Karlan and Zinman (2008), where they estimate interest rate elasticities for loan demand in South Africa. Specifically, we let $Y = \ln(\textit{loan size})$ and $X = \ln(\textit{monthly interest rate})$. U represents other (unobserved) drivers of loan demand. Thus, the nonseparable structural equation we are interested in is:

$$\ln(\textit{loan size}) = r(\ln(\textit{monthly interest rate}), U).$$

We let Z be a client's preapproved risk category, i.e., Z is discrete, such that:

$$Z_i = \begin{cases} 1 \text{ if client } i \text{ is low risk} \\ 2 \text{ if client } i \text{ is medium risk} \\ 3 \text{ if client } i \text{ is high risk} \end{cases}.$$

An important feature of this dataset is that interest rates are *randomized*, stratified by the client's risk category (for details, see Karlan and Zinman, 2008, p. 1045). Thus, the conditional exogeneity assumption ($X \perp U \mid Z$) holds by the randomized experimental design. The sample size is 2,325. The numbers of low, medium, and high risk clients are 641, 459, and 1,225, respectively.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

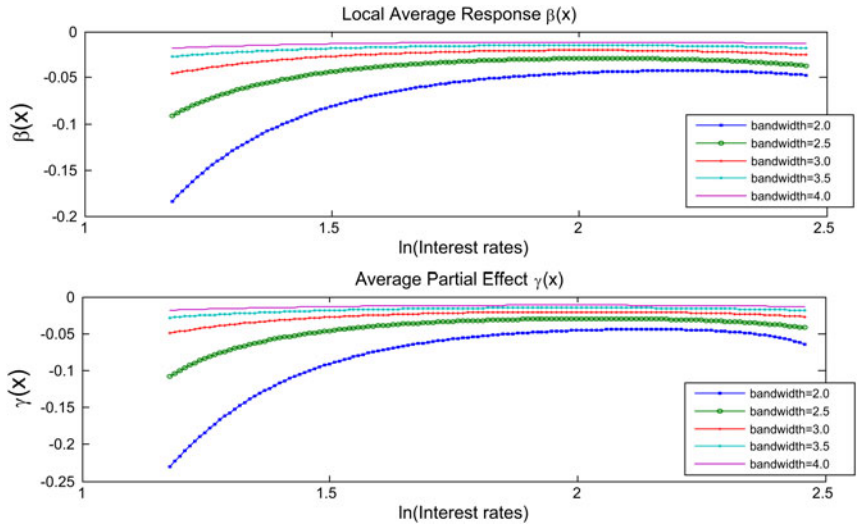


FIGURE 1. Interest rate elasticities for loan demand.

In this context, $\beta_0(x)$ measures the local average interest rate elasticities at level x for those who are assigned interest rate x. $\gamma_0(x)$ measures the average interest rate elasticities at level x over the whole population. We first estimate $\beta_0(x)$ and $\gamma_0(x)$ for different choices of bandwidth;[5] the results are shown in Figure 1. It appears that neither $\beta_0(x)$ nor $\gamma_0(x)$ varies much with x.

We test $\mathbb{H}_{10}$, $\mathbb{H}_{20}$, $\mathbb{H}'_{10}$, and $\mathbb{H}'_{20}$ formally using our new methods. In this application, $\mathbb{H}_{10}$ says the local elasticities ($\beta_0(x)$) do not depend on interest rates. $\mathbb{H}_{20}$ says that the average elasticities ($\gamma_0(x)$) are constant. $\mathbb{H}'_{10}$ means that on average, the interest rate elasticities of those who are assigned interest rate ($\beta_0(x)$) are zero. $\mathbb{H}'_{20}$ means that the average interest elasticities ($\gamma_0(x)$) are zero. It turns out that at the 5% significance level, we fail to reject both $\mathbb{H}_{10}$ and $\mathbb{H}_{20}$ but reject both $\mathbb{H}'_{10}$ and $\mathbb{H}'_{20}$. The results appear in Table 13.

In summary, we find that the two average effects of interest rates on loan demand are nonzero, but we do not have evidence showing that these two average elasticities depend on interest rates. This is in general consistent with Karlan and Zinman's (2008) finding that "loan size is far more responsive to changes in loan maturity than to changes in interest rates."

6. EXTENSIONS

Other effect measures, in addition to LAR and APE, may be of interest, such as the ME and the *covariate-conditioned average effect* (CAE) (see, e.g., White and Chalak, 2006):

$$\theta_0(x,z) \equiv E[m_0(x,U)|Z=z], \tag{6.1}$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

TABLE 13. Interest rate elasticities for loan demand

$\mathbb{H}_{10}$ (constant effects for LAR)									
Bandwidth (h)	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8
p-values	0.075	0.078	0.078	0.083	0.083	0.083	0.088	0.093	0.098
$\mathbb{H}_{20}$ (constant effects for APE)									
Bandwidth (h)	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8
p-values	0.123	0.138	0.145	0.170	0.183	0.200	0.208	0.218	0.223
$\mathbb{H}'_{10}$ (zero effects for LAR)									
Bandwidth (h)	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8
p-values	0.018	0.018	0.020	0.020	0.028	0.030	0.030	0.033	0.033
$\mathbb{H}'_{20}$ (zero effects for APE)									
Bandwidth (h)	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8
p-values	0.025	0.025	0.028	0.028	0.028	0.028	0.028	0.030	0.030

Note: We let $\underline{x} = \ln(3.25)$ and $\bar{x} = \ln(11.75)$. The bandwidth range is the same as in Figure 1 to give the estimates in a reasonable range. The asymptotic distribution is obtained by the nonparametric bootstrap with the number of random draws $B = 400$.

where $m_0(x, U)$ is the ME. CAE measures the average effects of X on Y for a subpopulation defined by covariates $(Z = z)$. This is similar to the conditional average treatment effect considered in Crump et al. (2008).

We can test for the treatment dependence of ME and CAE:

$$\textit{Hypothesis 3}: \ \mathbb{H}_{30}: \textit{There exists a measurable function } \theta_1 \textit{ s.t. } \theta_0(X, Z) = \theta_1(Z) \ a.s.$$
$$\mathbb{H}_{3A}: \mathbb{H}_{30} \textit{ is false.} \tag{6.2}$$
$$\textit{Hypothesis 4}: \ \mathbb{H}_{40}: \textit{There exists a measurable function } m_1 \textit{ s.t. } m_0(X, U) = m_1(U) \ a.s.$$
$$\mathbb{H}_{4A}: \mathbb{H}_{40} \textit{ is false.} \tag{6.3}$$

We can also test the hypotheses of zero treatment effects of CAE and ME as follows:

$$\textit{Hypothesis } 3': \mathbb{H}'_{30}: \theta_0(X, Z) = 0 \ a.s. \text{ versus } \mathbb{H}'_{3A}: \mathbb{H}'_{30} \textit{ is false.} \tag{6.4}$$
$$\textit{Hypothesis } 4': \mathbb{H}'_{40}: m_0(X, U) = 0 \ a.s. \text{ versus } \mathbb{H}'_{4A}: \mathbb{H}'_{40} \textit{ is false.} \tag{6.5}$$

Remark 3. Note that $\mathbb{H}_{40}$ implies $\mathbb{H}_{30}$; and $\mathbb{H}_{30}$ implies $\mathbb{H}_{20}$; $\mathbb{H}'_{40}$ implies $\mathbb{H}'_{30}$; and $\mathbb{H}'_{30}$ implies $\mathbb{H}'_{10}$ and $\mathbb{H}'_{20}$; The relationship suggests that the tests for LAR and APE discussed above are useful for testing the hypotheses about CAE and ME. For example, suppose that we reject $\mathbb{H}'_{10}$ (zero effects for LAR), then we must also reject $\mathbb{H}'_{30}$ (zero effects for CAE).

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

Under A.1 and A.2, $\theta_0(x,z)$ is identified as

$$\theta_0(x,z) = \frac{\partial \mu(x,z)}{\partial x}.$$

Thus, testing $\mathbb{H}_{30}$ and $\mathbb{H}'_{30}$ is equivalent to testing certain restrictions on the conditional expectation function $\mu(x,z)$ as shown in the next lemma.

LEMMA 8. *Suppose A.1, A.2, and A.3(c) hold. Then*

(a) $\mathbb{H}_{30}$ *is equivalent to* $\mathbb{H}''_{30}$:

$\mathbb{H}''_{30}$: *There exist measurable functions* μ_1 *and* μ_2 *s.t.*

$$\mu(X,Z) = \mu_1(Z) + \mu_2(Z)\cdot X \ a.s.$$

(b) $\mathbb{H}'_{30}$ *is equivalent to* $\mathbb{H}'''_{30}$:

$\mathbb{H}'''_{30}$: *There exists a measurable function* μ_1 *s.t.* $\mu(X,Z) = \mu_1(Z)$ *a.s.*

Under $\mathbb{H}''_{30}$, $\mu(X,Z)$ is a semiparametric varying coefficient process (see, e.g., Li and Racine, 2007, Sect. 9.2). Under $\mathbb{H}'''_{30}$, $\mu(X,Z)$ only depends on Z; this is conditional mean independence restriction. We can apply specification tests to test these restrictions on $\mu(X,Z)$. In particular, we can construct test statistics similar to Li, Hsiao, and Zinn (2003), who use series estimators to estimate conditional expectation functions under restrictions and propose an integrated conditional moment type of test. Though it is possible to use kernel estimators, using series estimators has the advantage of computational convenience, as it is easy to impose restrictions on μ with series estimators. Here we provide a brief discussion about how to test $\mathbb{H}''_{30}$. We can construct similar test statistics for $\mathbb{H}'''_{03}$. As shown in Lemma 1, testing $\mathbb{H}''_{30}$ is equivalent to testing:

$$\mathcal{T}''_{30}(\lambda) \equiv E\left[\varphi\left(\lambda_0 + X\lambda_1 + Z'\lambda_2\right)\cdot\varepsilon\right] = 0$$

for ν-almost all $\lambda \in \Lambda$, where $\lambda \equiv (\lambda_0, \lambda_1, \lambda_2)$, $\lambda_0, \lambda_1 \in \mathbb{R}$, $\lambda_2 \in \mathbb{R}^{k_z}$, $\varepsilon = Y - \mu(X,Z)$, and φ is GCR. We first estimate μ under the restriction using the series

$$\tilde{g}(x,z) = \sum_{k=0}^{K_0} \delta_{0,k}\cdot p_k(z) + \sum_{k=0}^{K_1} \delta_{1,k}\cdot p_k(z)\cdot x,$$

where $K_0 (\equiv K_{0n})$ and $K_1 (\equiv K_{1n})$ are positive integers increasing with n; $\{p_k(.)\}$ is some collection of basis functions; and $\{\delta_{0,k}\}$ and $\{\delta_{1,k}\}$ are corresponding coefficients. To simplify notation, we write

$$\tilde{g}(x,z) = \left[p^K(x,z)\right]'\delta^K,$$

where $K \equiv K_0 + K_1$, $p^K(x,z) \equiv \left[p_0(z), p_1(z), \ldots, p_{K_0}(z), p_0(z)\cdot x, p_1(z)\cdot x, \ldots, p_{K_1}(z)\cdot x,\right]'$, and $\delta^K \equiv \left[\delta_{0,0}, \delta_{0,1}, \ldots, \delta_{0,K_0}, \delta_{1,0}, \delta_{1,1}, \ldots, \delta_{1,K_1}\right]'$. We estimate δ^K using $\hat{\delta}^K$, the OLS estimator. Then we estimate ε_i by $\hat{\varepsilon}_i$:

$$\hat{\varepsilon}_i = Y_i - \hat{g}(X_i, Z_i),$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

where

$$\hat{g}(X_i, Z_i) = \left[p^K(X_i, Z_i)\right]' \hat{\delta}^K.$$

A sample analog of $\mathcal{T}''_{30}(\lambda)$ is $\hat{\mathcal{T}}_{3n}(\lambda)$:

$$\hat{\mathcal{T}}_{3n}(\lambda) \equiv n^{-1} \sum_{i=1}^{n} \varphi\left(\lambda_0 + X_i\lambda_1 + Z_i'\lambda_2\right) \cdot \hat{\varepsilon}_i.$$

Then the test statistic $\widehat{\mathcal{IT}}_{3n}$ is the integrated squared $\hat{\mathcal{T}}_{3n}(\lambda)$:

$$\widehat{\mathcal{IT}}_{3n} \equiv \int_{\Lambda} \left[\hat{\mathcal{T}}_{3n}(\lambda)\right]^2 dv(\lambda).$$

We can also show that the null distribution is nonstandard and we can obtain the critical values by using a residual-based wild bootstrap as in Li et al. (2003).

Next, we provide a method for testing $\mathbb{H}_{40}$ and $\mathbb{H}'_{40}$.

LEMMA 9. *Suppose A.1 holds. Then*

(a) $\mathbb{H}_{40}$ *is equivalent to* $\mathbb{H}''_{40}$:

$\mathbb{H}''_{40}$: There exist measurable functions r_1 and r_2 such that
$r(X, U) = r_1(U) + X \cdot r_2(U)$ *a.s.*

(b) $\mathbb{H}'_{40}$ *is equivalent to* $\mathbb{H}'''_{40}$:

$\mathbb{H}'''_{40}$: There exists a measurable function r_1 such that
$r(X, U) = r_1(U)$ *a.s.*

Under $\mathbb{H}''_{40}$, $r(X, U)$ has a random coefficient structure. It turns out that this implies many testable conditional moment restrictions, as shown in the next lemma.

LEMMA 10. *Suppose A.1 and A.2 hold. Let S be a positive integer and suppose* $E(Y^S|X = x, Z = z) < \infty$ *for all* $(x, z) \in \Theta_{x,z}$. *Then under* $\mathbb{H}''_{40}$,

$$E\left(Y^s \mid X = x, Z = z\right) = \sum_{j=0}^{s} x^j \cdot g_{s,j}(z), \quad s = 1, \ldots, S, \tag{6.6}$$

where $g_{s,0}, g_{s,1}, g_{s,2}, \ldots, g_{s,s} : \mathbb{R}^{k_z} \to \mathbb{R}$ *are measurable functions.*

We can test the restrictions in equation (6.6) as in testing for $\mathbb{H}''_{03}$ using series estimator to estimate $E(Y^s \mid X = x, Z = z)$ under the restriction and construct a test statistic similar to $\widehat{\mathcal{IT}}_{3n}$.

To test $\mathbb{H}'''_{40}$, one can simply test

$$X \perp Y \mid Z, \tag{6.7}$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

as $X \perp U \mid Z$ implies $X \perp r_1(U) \mid Z$. There are many conditional independence tests that can be applied directly (see, e.g., Huang et al., 2013 and the references therein).

We further note that equation (6.6) is not equivalent to $\mathbb{H}''_{40}$ and equation (6.7) is not equivalent to $\mathbb{H}'''_{40}$. Thus, testing equations (6.6) or (6.7) may not have power against all alternatives of $\mathbb{H}''_{40}$ or $\mathbb{H}'''_{40}$. In fact, we can show that given the relatively weak assumption imposed in this paper, it is impossible to design tests that are equivalent to $\mathbb{H}''_{40}$ or $\mathbb{H}'''_{40}$ (see Lu and White, 2014 for a specific example). Thus, these tests can be viewed as falsification tests.

So far, we have only considered different versions of average effects. We can also consider other effects, such as quantile effects and distributional effects. Following Imbens and Newey (2009, p. 1489), we define the CDF of the counterfactual Y when X is fixed at a particular value x as

$$G(y,x) \equiv \Pr[r(x,U) \leq y],$$

and the quantile structural function:

$$Q(\tau,x) \equiv G^{-1}(\tau,x).$$

Then the distributional effects and quantile effects are defined as

$$G'(y,x) \equiv \frac{\partial Q(y,x)}{\partial x} \text{ and } Q'(\tau,x) \equiv \frac{\partial Q(\tau,x)}{\partial x},$$

respectively. Under A.1 and A.2, both effects are identified. Specifically, it is easy to show under our assumptions that

$$G(y,x) = E[F(y|x,Z)],$$

where $F(y|x,z)$ is the conditional CDF of Y given $X = x$ and $Z = z$. The corresponding estimator of $G(y,x)$ proposed in Imbens and Newey (2009) is

$$\hat{G}(y,x) = \frac{1}{n}\sum_{i=1}^{n} \hat{F}(y|x,Z_i),$$

where $\hat{F}(y|x,Z_i)$ is a nonparametric estimator of $F(y|x,z)$, e.g., a kernel-based estimator. Then $Q(\tau,x)$ can be estimated by $\hat{G}^{-1}(\tau,x)$. Given the estimators of $G(y,x)$ and $Q(\tau,x)$, $G'(y,x)$ and $Q'(\tau,x)$ can be estimated by numerical derivatives (see, e.g., Hong, Mahajan, and Nekipelov, 2010). Then, we can test the hypothesis that the quantile effects or the distributional effects are zero or do not depend on the treatment level x using similar testing methods as discussed above. For example, to test whether $Q'(\tau,x) = c_\tau$ for all $x \in I_x$, where τ is given and c_τ is an unknown constant, we can construct the test statistic as

$$\widehat{\mathcal{IT}}_{\tau n} = \int_{\Lambda} \left[\hat{\mathcal{T}}_{\tau n}(\lambda)\right]^2 d\nu(\lambda),$$

where $\hat{\mathcal{T}}_{\tau n}(\lambda) = \int \psi(x,\lambda)\cdot\hat{Q}'(\tau,x)\cdot dx$ and $\hat{Q}'(\tau,x)$ is an estimator of $Q'(\tau,x)$.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

7. SUMMARY AND CONCLUDING REMARKS

In this paper, we propose a battery of nonparametric tests for zero treatment effects and treatment dependence of effects of a continuous treatment in nonseparable structural equations. Our tests are easy to implement, yet powerful: they detect local alternatives that converge to the null at the parametric rate $n^{-1/2}$. We illustrate with studies of interest rate elasticities for loan demand.

There are many interesting topics for further research. First, there are many other structural effect features that have significant policy implications, besides zero treatment effects and treatment dependence. Prominent examples are, for instance, monotonicity, and concavity. Second, if we reject the absence of treatment dependence, it is of interest to investigate where the effects are strongest and to analyze explicitly how to allocate limited resources given policymakers' objectives.

NOTES

1. However, as pointed out by an anonymous referee, in this case Z is the conditional CDF of X given the instrumental variable and has to be estimated. This leads to the problem of generated covariates. We conjecture that in this case, it is still possible to show that our test statistics achieve $n^{-1/2}$ local power. Then a similar subsampling procedure discussed can still be applied. A full analysis of this question is beyond the scope of this paper. For recent work on the semiparametric estimation with generated covariates, see, e.g., Mammen, Rothe, and Schienle (2013).

2. In principle, one can consider the alternative approach to test the L_2 distance between $\beta_0(X)$ and $\bar{\beta}_0$. That is, test

$$\int \left[\beta_0(x) - \bar{\beta}_0\right]^2 a(x)\, dx = 0,$$

where $a(.)$ is some weigh function. The similar approach is used in, e.g., Härdle and Mammen (1993) and Ait-Sahalia, Bickel, and Stoker (2001). The test proposed here has faster (parametric) convergence rates, while the L_2-distance test has typical nonparametric rates. However, the L_2-distance test may have better power to detect high frequency local alternatives (for a discussion, see, e.g., Fan and Li, 2000). Thus, our test and the L_2-distance test are complementary to each other. However, we are not aware of any L_2-distance test that can be directly applied to the hypotheses discussed in this paper.

3. For notational simplicity, we use the same bandwidth for X and Z here. In the simulations, we let the bandwidths be proportional to their sample standard deviations.

4. To better understand the boundary problem, let us consider a simple case where Z has null dimension. Then integration by parts yields

$$\mathcal{T}'_{10}(\lambda) = \psi'(x,\lambda)\,\mu(x)\big|_{\underline{x}}^{\bar{x}} - \int \mu(x)\,\frac{\partial}{\partial x}\psi'(x,\lambda)\cdot dx.$$

If $\psi'(x,\lambda)$ does not vanish at the boundary points, then we need to estimate $\mu(x)$ at the two boundary points $\bar{x}$ and $\underline{x}$, which cannot be estimated at the parametric rate.

5. The bandwidth range is chosen to ensure that the estimates are in a reasonable range. The simple "rule of thumb" discussed in Section 3.4 does not generate reasonable estimates in this application.

REFERENCES

Ait-Sahalia, Y., P. Bickel, & T.M. Stoker (2001) Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics* 105, 363–412.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

Altonji, J. & R. Matzkin (2005) Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica* 73, 1053–1102.

Andrews, D.W.K. (1994) Empirical process methods in econometrics. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2248–2296. Elsevier.

Belzil, C. & J. Hansen (2002) Unobserved ability and the return to schooling. *Econometrica* 70, 2075–2091.

Bester, C. & C. Hansen (2009) Identification of marginal effects in a nonparametric correlated random effects model. *Journal of Business and Economic Statistics* 27, 235–250.

Bierens, H. (1982) Consistent model specification tests. *Journal of Econometrics* 20, 105–134.

Bierens, H. (1990) A consistent conditional moment test of functional form. *Econometrica* 58, 1443–1458.

Bierens, H. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1151.

Billingsley, P. (1999) *Convergence of Probability Measures*. Wiley.

Blundell, R. & J. Powell (2003) Endogeneity in nonparametric and semiparametric regression models. In M. Dewatripont, L.P. Hansen, & S.J. Turnovsky (eds.), *Advances in Economics and Econometrics*, chap. 8, pp. 312–357. Cambridge University Press.

Boning, W.B. & F. Sowell (1999) Optimality for the integrated conditional moment test. *Econometric Theory* 15, 710–718.

Casella, G. & R.L. Berger (2002) *Statistical Inference*. Duxbury.

Chamberlain, G. (1984) Panel data. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, pp. 1247–1318. North-Holland.

Chen, X. & Y. Fan (1999) Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* 91, 373–401.

Crump, R., J. Hotz, G. Imbens, & O. Mitnik (2008) Nonparametric tests for treatment effect heterogeneity. *Review of Economics and Statistics* 90, 389–405.

Dawid, P. (1979) Conditional independence and statistical theory. *Journal of the Royal Statistical Society Series B* 41, 1–31.

Fan, Y. & Q. Li (2000) Consistent model specification tests: Kernel-based test versus Bierens' ICM tests. *Econometric Theory* 16, 1016–1041.

Florens, J., J. Heckman, C. Meghir, & E. Vytlacil (2008) Identification of treatment effects using control functions in models with continuous endogenous treatment and heterogeneous effects. *Econometrica* 76, 1191–1206.

Griliches, Z. (1977) Estimating the returns to schooling: Some econometric problems. *Econometrica* 45, 1–22.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

Hansen, B. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

Hoderlein, S. (2011) How many consumers are rational? *Journal of Econometrics* 164, 294–309.

Hoderlein, S. & E. Mammen (2007) Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* 75, 1513–1519.

Hoderlein, S. & E. Mammen (2009) Identification and estimation of local average derivatives in nonseparable models without monotonicity. *Econometrics Journal* 12, 1–25.

Hoderlein, S. & H. White (2012) Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics* 168, 300–314.

Hong, H., A. Mahajan, & D. Nekipelov (2010) Extremum Estimation and Numerical Derivatives. Discussion paper, UC Berkeley, Department of Economics.

Huang, M. (2009) Essays on testing conditional independence. Ph.D. dissertation, University of California, San Diego.

Huang, M., Y. Sun, & H. White (2013) A Flexible Test for Conditional Independence. Discussion paper, UCSD, Department of Economics.

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

Imbens, G. & W. Newey (2009) Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77, 1481–1512.
Karlan, D. & J. Zinman (2008) Credit elasticities in less-developed economies: Implications for microfinance. *American Economic Review* 98, 1040–1068.
Li, Q., C. Hsiao, & J. Zinn (2003) Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics* 112, 295–325.
Li, Q. & J. Racine (2003) Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis* 86, 266–292.
Li, Q. & J. Racine (2007) *Nonparametric Econometrics Theory and Practice*. Princeton University Press.
Lu, X. & H. White (2014) Testing for separability in structural equations. *Journal of Econometrics* 182, 14–26.
Mammen, E., C. Rothe, & M. Schienle (2013) Semiparametric Estimation with Generated Covariates. Discussion paper, Columbia University, Department of Economics.
Newey, W. (1994) Kernel estimation of partial means and a general variance estimator. *Econometric Theory* 10, 233–253.
Newey, W. & T. Stoker (1993) Efficiency of weighted average derivative estimators and index models. *Econometrica* 61, 1199–1223.
Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics*. Cambridge University Press.
Politis, D., J. Romano, & M. Wolf (1999) *Subsampling*. Springer-Verlag.
Schennach, S., H. White, & K. Chalak (2012) Estimating average marginal effects in nonseparable structural systems. *Journal of Econometrics* 166, 282–302.
Stinchcombe, M. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–324.
van der Vaart, A. & J. Wellner (2000) *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag.
White, H. (1994) *Estimation, Inference and Specification Analysis*. Econometric Society Monographs, No. 22. Cambridge University Press.
White, H. (2001) *Asymptotic Theory for Econometricians*. Academic Press.
White, H. (2006) Approximate nonlinear forecasting methods. In G. Elliott, C.W.J. Granger, & A. Timmermann (eds.), *Handbook of Economic Forecasting*, pp. 460–512. Elsevier.
White, H. & K. Chalak (2006) A Unified Framework for Defining and Identifying Causal Effects. Discussion paper, UCSD, Department of Economics.
White, H. & K. Chalak (2008) Identifying Structural Effects in Nonseparable Systems Using Covariates. Discussion paper, UCSD, Department of Economics.
Wooldridge, J. (2005) Unobserved heterogeneity and estimation of average partial effects. In D.W.K. Andrews & J. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 27–55. Cambridge University Press.

APPENDIX: Mathematical Proofs

Notations: In the proofs, we let $a \lesssim b$ denote that $a \leq Mb$ for some constant M that is universal in the context of the proofs. Further, we denote:

$$\begin{aligned}
&\Delta\hat{f}(x) \equiv \hat{f}(x) - f(x); && \Delta\hat{f}(z) \equiv \hat{f}(z) - f(z); \\
&\Delta\hat{f}_x(x,y) \equiv \hat{f}_x(x,y) - f_x(x,y); && \Delta\hat{f}(x,z) \equiv \hat{f}(x,z) - f(x,z); \\
&\Delta\hat{f}(x,y,z) \equiv \hat{f}(x,y,z) - f(x,y,z); && \Delta\left[\hat{f}(x)^{-1}\right] \equiv \hat{f}(x)^{-1} - f(x)^{-1}.
\end{aligned} \tag{A.1}$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

Proof of Lemma 1. For part (a), see Stinchcombe and White (1998). For part (b), since $E(\mathcal{U}|\mathcal{X})$ can be any measurable function of $\mathcal{X}$, we can find a random variable $\mathcal{U}^*$ such that $g(\mathcal{X}) = E(\mathcal{U}^*|\mathcal{X})$ for any given $g(\mathcal{X})$. Then the conclusion of part (b) directly follows from part (a). ■

Proof of Lemma 2. By Lemma 1, equation (3.2) is equivalent to

$$E\left[\varphi(\lambda_0+\lambda_1 X)\cdot f(X)^{-1}\cdot w(X)\cdot\left[\beta_0(X)-\bar{\beta}_0\right]\right]=0 \tag{A.2}$$

for ν-almost all $\lambda \in \Lambda$. Further, equation (A.2) can be simplified as:

$$\begin{aligned}
&E\left[\varphi(\lambda_0+\lambda_1 X)\cdot f(X)^{-1}\cdot w(X)\cdot\left[\beta_0(X)-\bar{\beta}_0\right]\right]\\
&\quad=\int\varphi(\lambda_0+\lambda_1 x)\cdot w(x)\cdot\left[\beta_0(x)-\bar{\beta}_0\right]\cdot dx\\
&\quad=\int\varphi(\lambda_0+\lambda_1 x)\cdot w(x)\cdot\beta_0(x)\cdot dx-\left[\int\beta_0(x)\cdot w(x)\cdot dx\right]\\
&\qquad\times\left[\int\varphi(\lambda_0+\lambda_1 x)\cdot w(x)\cdot dx\right]\\
&\quad=\int\psi(x,\lambda)\cdot\beta_0(x)\cdot dx\equiv\mathcal{T}_{10}(\lambda).
\end{aligned}$$

Thus the proof is complete. ■

Proof of Theorem 3. The proof of part (a) is in three steps. We first show (STEP 1) that

$$\begin{aligned}
\sqrt{n}\left[\hat{\mathcal{T}}_{1n}(\lambda)-\mathcal{T}_{10}(\lambda)\right]&=n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{\kappa_1(X_i,Y_i,Z_i,\lambda)-E\left[\kappa_1(X_i,Y_i,Z_i,\lambda)\right]\right\}\\
&\quad+\sqrt{n}I_1(\lambda)
\end{aligned} \tag{A.3}$$

and $\sup_{\lambda\in\Lambda}|\sqrt{n}I_1(\lambda)|=o_p(1)$. We then show (STEP 2) that for finite collections of $\lambda's$, $\sqrt{n}\left[\hat{\mathcal{T}}_{1n}(\lambda)-\mathcal{T}_{10}(\lambda)\right]$ is jointly asymptotically normal. Finally, we show (STEP 3) that $n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{\kappa_1(X_i,Y_i,Z_i,\lambda)-E\left[\kappa_1(X_i,Y_i,Z_i,\lambda)\right]\right\}$ is a tight process in Λ by showing it is stochastic equicontinuous. The key result used to prove this theorem is the uniform convergence rate of kernel estimators. Let $\bar{\Theta}_{x,z}$ denote the largest compact subset of $\Theta_{x,z}$ such that for all (x,z) in $\bar{\Theta}_{x,z}$, $\underline{x}\le x\le\bar{x}$. We define:

$$\begin{aligned}
A_n\equiv&\max_{k\in\{0,1\}}\sup_{\{x,z\}\in\bar{\Theta}_{x,z}}\left|\frac{\partial\hat{f}(x,z)}{\partial x^k}-\frac{\partial f(x,z)}{\partial x^k}\right|\\
&+\max_{k\in\{0,1\}}\sup_{\{x,z\}\in\bar{\Theta}_{x,z}}\left|\int y\cdot\frac{\partial\hat{f}(x,y,z)}{\partial x^k}\cdot dy-\int y\cdot\frac{\partial f(x,y,z)}{\partial x^k}\cdot dy\right|.
\end{aligned}$$

We can show (see Newey, 1994, Lem. B.3, p. 252 or Altonji and Matzkin, 2005, p. 1080) that

$$A_n=O_p\left(h^s+\sqrt{\frac{\ln(n)}{nh^{3+k_z}}}\right). \tag{A.4}$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

STEP 1: To show $\sup_{\lambda\in\Lambda}\left|\sqrt{n}I_1(\lambda)\right| = o_p(1)$, we decompose $I_1(\lambda)$ into many small terms and show that each term is uniformly $o_p(1)$. Specifically, after some tedious yet straightforward algebra, we can show $I_1(\lambda) = I_{11}(\lambda) + I_{12}(\lambda)$, where

$$
\begin{aligned}
I_{11}(\lambda) &\equiv \sum_{k=1}^{14} I_{11,k}(\lambda),\\
I_{11,1}(\lambda) &\equiv \int [\psi(x,\lambda)]\cdot\left[\Delta\left[\hat{f}(x)^{-1}\right] + f(x)^{-2}\cdot\Delta\hat{f}(x)\right]\cdot\left[\int y\cdot f_x(x,y)\cdot dy\right]\cdot dx\\
&\quad + \int [\psi(x,\lambda)]\cdot\Delta\left[\hat{f}(x)^{-1}\right]\cdot\left[\int y\cdot\Delta\hat{f}_x(x,y)\cdot dy\right]\cdot dx,\\
I_{11,k}(\lambda) &\equiv \int [\psi(x,\lambda)]\left[\int J_{11,k}(x,z)\cdot dz\right]\cdot dx,\ k=2,\ldots,12,\\
J_{11,2}(x,z) &\equiv \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right]\cdot\left[\int y\cdot\Delta\hat{f}(x,y,z)\cdot dy\right]\cdot\Delta\hat{f}_x(x,z)\cdot\Delta\left[\hat{f}(x)^{-1}\right],\\
J_{11,3}(x,z) &\equiv \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right]\cdot\left[\int y\cdot f(x,y,z)\cdot dy\right]\cdot\Delta\hat{f}_x(x,z)\cdot\Delta\left[\hat{f}(x)^{-1}\right],\\
J_{11,4}(x,z) &\equiv f(x,z)^{-1}\cdot\left[\int y\cdot\left(\hat{f}(x,y,z) - f(x,y,z)\right)\cdot dy\right]\cdot\Delta\hat{f}_x(x,z)\cdot\Delta\left[\hat{f}(x)^{-1}\right],\\
J_{11,5}(x,z) &\equiv f(x,z)^{-1}\cdot\left[\int y\cdot f(x,y,z)\cdot dy\right]\cdot\Delta\hat{f}_x(x,z)\cdot\Delta\left[\hat{f}(x)^{-1}\right],\\
J_{11,6}(x,z) &\equiv \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right]\cdot\left[\int y\cdot\Delta\hat{f}(x,y,z)\cdot dy\right]\cdot f_x(x,z)\cdot\Delta\left[\hat{f}(x)^{-1}\right],\\
J_{11,7}(x,z) &\equiv \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right]\cdot\left[\int y\cdot f(x,y,z)\cdot dy\right]\cdot f_x(x,z)\cdot\Delta\left[\hat{f}(x)^{-1}\right],\\
J_{11,8}(x,z) &\equiv f(x,z)^{-1}\cdot\left[\int y\cdot\Delta\hat{f}(x,y,z)\cdot dy\right]\cdot f_x(x,z)\cdot\Delta\left[\hat{f}(x)^{-1}\right],\\
J_{11,9}(x,z) &\equiv f(x,z)^{-1}\cdot\left[\int y\cdot f(x,y,z)\cdot dy\right]\cdot f_x(x,z)\cdot\left[\Delta\left[\hat{f}(x)^{-1}\right] + f(x)^{-2}\cdot\Delta\hat{f}(x)\right],\\
J_{11,10}(x,z) &\equiv \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right]\cdot\left[\int y\cdot\Delta\hat{f}(x,y,z)\cdot dy\right]\cdot\Delta\hat{f}_x(x,z)\cdot f(x)^{-1},\\
J_{11,11}(x,z) &\equiv \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right]\cdot\left[\int y\cdot f(x,y,z)\cdot dy\right]\cdot\Delta\hat{f}_x(x,z)\cdot f(x)^{-1},\\
J_{11,12}(x,z) &\equiv f(x,z)^{-1}\cdot\left[\int y\cdot\Delta\hat{f}(x,y,z)\cdot dy\right]\cdot\Delta\hat{f}_x(x,z)\cdot f(x)^{-1},\\
J_{11,13}(x,z) &\equiv \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right]\cdot\left[\int y\cdot\Delta\hat{f}(x,y,z)\cdot dy\right]\cdot f_x(x,z)\cdot f(x)^{-1},\\
J_{11,14}(x,z) &\equiv \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1} + f(x,z)^{-2}\cdot\Delta\hat{f}(x,z)\right]\\
&\quad\cdot\left[\int y\cdot f(x,y,z)\cdot dy\right]\cdot f_x(x,z)\cdot f(x)^{-1},\\
I_{12}(\lambda) &\equiv \int [\psi(x,\lambda)]\cdot f(x)^{-1}\cdot\left[\int y\cdot\Delta\hat{f}_x(x,y)\cdot dy\right]\cdot dx\\
&\quad - \int [\psi(x,\lambda)]\cdot f(x)^{-2}\cdot\Delta\hat{f}(x)\cdot\left[\int y\cdot f_x(x,y)\cdot dy\right]\cdot dx
\end{aligned}
$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

$$
\begin{aligned}
&+\int [\psi(x,\lambda)]\cdot\left[\int J_{12}(x,z)\cdot dz\right]\cdot dx\\
&-\frac{1}{n}\sum_{i=1}^{n}\{\kappa_1(X_i,Y_i,Z_i,\lambda)-E(\kappa_1(X_i,Y_i,Z_i,\lambda))\},\\
J_{12}(x,z)\equiv&-f(x,z)^{-1}\cdot\left[\int y\cdot f(x,y,z)\cdot dy\right]\cdot f_x(x,z)\cdot f(x)^{-2}\cdot\Delta\hat{f}(x)\\
&+f(x,z)^{-1}\cdot\left[\int y\cdot f(x,y,z)\cdot dy\right]\cdot\Delta\hat{f}_x(x,z)\cdot f(x)^{-1}\\
&+f(x,z)^{-1}\cdot\left[\int y\cdot\Delta\hat{f}(x,y,z)\cdot dy\right]\cdot f_x(x,z)\cdot f(x)^{-1}\\
&-f(x,z)^{-2}\cdot\Delta\hat{f}(x,z)\cdot\left[\int y\cdot f(x,y,z)\cdot dy\right]\cdot f_x(x,z)\cdot f(x)^{-1}.
\end{aligned}
$$

We verify that $\sup_{\lambda\in\Lambda}\left|\sqrt{n}I_{11}(\lambda)\right|=o_p(1)$ by showing $\sup_{\lambda\in\Lambda}\left|\sqrt{n}I_{11,k}(\lambda)\right|=o_p(1)$ for $k=1,\ldots,14$. For example,

$$
\begin{aligned}
&\sup_{\lambda\in\Lambda}\left|\sqrt{n}I_{11,1}(\lambda)\right|\\
&\overset{(1)}{\lesssim}\sqrt{n}A_n^2\sup_{\lambda\in\Lambda}\int|\psi(x,\lambda)|\cdot f(x)^{-3}\cdot\left|\int y\cdot f_x(x,y)\cdot dy\right|\cdot dx\\
&\quad+\sqrt{n}A_n^2\sup_{\lambda\in\Lambda}\int|\psi(x,\lambda)|\cdot f(x)^{-2}\cdot dx\\
&\overset{(2)}{\lesssim}\sqrt{n}A_n^2\int w(x)\cdot f(x)^{-3}\cdot\left|\int y\cdot f_x(x,y)\cdot dy\right|\cdot dx\\
&\quad+\sqrt{n}A_n^2\int w(x)\cdot f(x)^{-2}\cdot dx\overset{(3)}{=}o_p(1),
\end{aligned}
$$

where $\overset{(1)}{\lesssim}$ follows from Lemma A.1(a) and (b); $\overset{(2)}{\lesssim}$ follows from the boundedness of $\varphi(.)$; $\overset{(3)}{=}$ follows from equation (A.4) and Assumptions B.2 and B.3. Next,

$$
\begin{aligned}
&\sup_{\lambda\in\Lambda}\left|\sqrt{n}I_{11,2}(\lambda)\right|\\
&\overset{(1)}{\leq}\sqrt{n}A_n^2\sup_{\lambda\in\Lambda}\int|\psi(x,\lambda)|\cdot\left[\int\left|\tau\left(\hat{f}(x,z),t\right)-f(x,z)^{-1}\right|\left|\Delta\left[\hat{f}(x)^{-1}\right]\right|\cdot dz\right]\cdot dx\\
&\overset{(2)}{\lesssim}\sqrt{n}A_n^2\int w(x)\cdot\left[\int\left|\tau\left(\hat{f}(x,z),t\right)-f(x,z)^{-1}\right|\left|\Delta\left[\hat{f}(x)^{-1}\right]\right|\cdot dz\right]\cdot dx\\
&\overset{(3)}{\leq}\sqrt{n}A_n^3\int w(x)\cdot\left[\int\left|\tau\left(\hat{f}(x,z),t\right)-f(x,z)^{-1}\right|\cdot f(x)^{-2}\cdot dz\right]\cdot dx\\
&\overset{(4)}{\leq}\sqrt{n}A_n^3\int w(x)\cdot f(x)^{-2}\cdot\left[t^{-2}A_n+t^{-3}A_n^2+\int\mathbf{1}[f(x,z)<2t]\cdot f(x,z)^{-1}dz\right]\cdot dx\\
&\overset{(5)}{\lesssim}\sqrt{n}t^{-2}A_n^4+\sqrt{n}t^{-3}A_n^5+\sqrt{n}A_n^3\cdot\int\int w(x)\cdot f(x,z)^{-1}dx\cdot dz\overset{(6)}{=}o_p(1),
\end{aligned}
$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

where $\overset{(1)}{\leq}$ follows from the definition of A_n; $\overset{(2)}{\lesssim}$ follows from the boundedness of $\varphi(.)$; $\overset{(3)}{\leq}$ is by Lemma A.1(a); $\overset{(4)}{\leq}$ is by Lemma A.1(c); $\overset{(5)}{\lesssim}$ is by the boundedness of $f(x)^{-1}$ on $\left[\underline{x}, \bar{x}\right]$; $\overset{(6)}{=}$ follows from equation (A.4) and Assumptions B.3 and B.4. Similarly, we can show that $\sup_{\lambda\in\Lambda}\left|\sqrt{n}I_{11,k}(\lambda)\right| = o_p(1)$ for $k = 3, \ldots, 14$ and $\sup_{\lambda\in\Lambda}\left|\sqrt{n}I_{12}(\lambda)\right| = o_p(1)$.

STEP 2: This is a direct application of the Cramer–Wold device and the Lindeberg–Levy central limit theorem (see, e.g., White, 2001, p. 114).

STEP 3: We decompose $\kappa_1(X, Y, Z, \lambda)$ into three terms:

$$\kappa_1(X, Y, Z, \lambda) = \kappa_{11}(X, Y, Z, \lambda) + \kappa_{12}(X, Y, Z, \lambda) + \kappa_{13}(X, Y, Z, \lambda),$$

where

$$\kappa_{11}(X, Y, Z, \lambda) \equiv -\left[\frac{Y - \mu(X, Z)}{f(X)}\right] \cdot \frac{\partial}{\partial x}\psi(X, \lambda),$$

$$\kappa_{12}(X, Y, Z, \lambda) \equiv -\frac{\partial}{\partial x}\left[\frac{Y - \mu(X, Z)}{f(X)}\right] \cdot \psi(X, \lambda),$$

$$\kappa_{13}(X, Y, Z, \lambda) \equiv -\left[\frac{(Y - \mu(X, Z)) \cdot f_x(X, Z)}{f(X, Z) \cdot f(X)} + \frac{\beta_0(X)}{f(X)}\right] \cdot \psi(X, \lambda),$$

and show that each of them satisfies Ossiander's L^2 entropy condition. Using Theorems 4–6 in Andrews (1994, pp. 2276–2282), it is sufficient to show that there exist some finite constants $\delta > 0$ and $\phi > 0$ such that

$$E\left\{\sup_{\lambda:\|\lambda-\lambda'\|<\delta}\left|\kappa_{11}(X, Y, Z, \lambda) - \kappa_{11}\left(X, Y, Z, \lambda'\right)\right|^2\right\} \lesssim \delta^{\phi}, \tag{A.5}$$

$$E\left\{\sup_{\lambda:\|\lambda-\lambda'\|<\delta}\left|\kappa_{12}(X, Y, Z, \lambda) - \kappa_{12}\left(X, Y, Z, \lambda'\right)\right|^2\right\} \lesssim \delta^{\phi}, \tag{A.6}$$

$$E\left\{\sup_{\lambda:\|\lambda-\lambda'\|<\delta}\left|\kappa_{13}(X, Y, Z, \lambda) - \kappa_{13}\left(X, Y, Z, \lambda'\right)\right|^2\right\} \lesssim \delta^{\phi}. \tag{A.7}$$

To verify equation (A.5), we show that

$$E\left\{\left[\frac{Y - \mu(X, Z)}{f(X)}\right]^2 \sup_{\lambda:\|\lambda-\lambda'\|<\delta}\left|\frac{\partial}{\partial x}\psi(X, \lambda) - \frac{\partial}{\partial x}\psi\left(X, \lambda'\right)\right|^2\right\}$$

$$\overset{(1)}{\leq}\left[E\left[\left[\frac{Y - \mu(X, Z)}{f(X)}\right]^4\right]\right]^{\frac{1}{2}} \cdot E\left[\left[\sup_{\lambda:\|\lambda-\lambda'\|<\delta}\left|\frac{\partial}{\partial x}\psi(X, \lambda) - \frac{\partial}{\partial x}\psi\left(X, \lambda'\right)\right|^2\right]^2\right]^{\frac{1}{2}}$$

$$\overset{(2)}{=}\left[E\left[\left[\frac{Y - \mu(X, Z)}{f(X)}\right]^4\right]\right]^{\frac{1}{2}} \cdot E\left[\left[\sup_{\lambda:\|\lambda-\lambda'\|<\delta}\left|\begin{bmatrix}\frac{\partial^2}{\partial x\partial\lambda_0}\psi\left(X, \tilde{\lambda}_0, \lambda_1\right) \\ \frac{\partial^2}{\partial x\partial\lambda_1}\psi\left(X, \lambda_0, \tilde{\lambda}_1\right)\end{bmatrix}'\begin{bmatrix}\lambda_0 - \lambda_0' \\ \lambda_1 - \lambda_1'\end{bmatrix}\right|^2\right]^2\right]^{\frac{1}{2}}$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

$$\overset{(3)}{\lesssim}\left[E\left[\left[\frac{Y-\mu(X,Z)}{f(X)}\right]^4\right]\right]^{\frac{1}{2}}\cdot\sup_{\lambda:\|\lambda-\lambda'\|<\delta}\left|(\lambda_0-\lambda_0')+(\lambda_1-\lambda_1')\right|^2$$
$$\lesssim\left[E\left[\left[\frac{Y-\mu(X,Z)}{f(X)}\right]^4\right]\right]^{\frac{1}{2}}\cdot\delta^2\lesssim\delta^2,$$

where $\tilde{\lambda}_0$ is between λ_0 and λ_0' and $\tilde{\lambda}_1$ is between $\lambda_1$and λ_1'; $\overset{(1)}{\leq}$ follows from Cauchy–Schwarz inequality; $\overset{(2)}{=}$ follows from Taylor expansion; $\overset{(3)}{\lesssim}$ follows from the boundedness of $\frac{\partial^2}{\partial x\partial\lambda_0}\psi(.)$ and $\frac{\partial^2}{\partial x\partial\lambda_1}\psi(.)$. Similarly, we can also verify equations (A.6) and (A.7). The proof of part (a) is thus complete. Part (b) is a direct application of the continuous mapping theorem after imposing $\mathbb{H}_{10}$. ■

Proof of Theorem 4 . For part (a), following the proof of Theorem 3, we can show that

$$\sqrt{n}\left(\hat{\mathcal{T}}_{1n}(.)-\frac{C_{10}(.)}{\sqrt{n}}\right)\text{ converges to }\mathcal{T}_{1\infty}(.)\text{ in }\mathcal{C}(\Lambda,\|.\|_\infty)\text{ weakly.}$$

Thus $\sqrt{n}\hat{\mathcal{T}}_{1n}(.)$ converges to $\tilde{\mathcal{T}}_{1\infty}(.)=\mathcal{T}_{1\infty}(.)+C_{10}(.)$ weakly. By Lemma 1, $C_{10}(\lambda)\neq 0$ for ν-almost all λ. By the continuous mapping theorem, $n\cdot\widehat{\mathcal{IT}}_{1n}$ converges to $\int\left[\mathcal{T}_{1\infty}(\lambda)+C_{10}(\lambda)\right]^2\cdot d\nu(\lambda)$. The proof of part (a) is thus complete. For part (b), following the proof of Theorem 3, we can show that $\sqrt{n}\left[\hat{\mathcal{T}}_{1n}(\lambda)-\mathcal{T}_{10}(\lambda)\right]$ converges to $\mathcal{T}_{1\infty}(\lambda)$. Thus $\sqrt{n}\left[\hat{\mathcal{T}}_{1n}(\lambda)-\mathcal{T}_{10}(\lambda)\right]=O_p(1)$. By the continuous mapping theorem, $\widehat{\mathcal{IT}}_{1n}=\int_\Lambda\left[\mathcal{T}_{10}(\lambda)\right]^2\cdot d\nu(\lambda)+o_p(1)$. Then if $\mathcal{T}_{10}(\lambda)\neq 0$, $\hat{\mathcal{T}}_{1n}(\lambda)=\mathcal{T}_{10}(\lambda)+o_p(1)$ implies that $\Pr\left[\left|\sqrt{n}\hat{\mathcal{T}}_{1n}(\lambda)\right|>k_n\right]\to 1$ and $\widehat{\mathcal{IT}}_{1n}=\int\left[\mathcal{T}_{10}(\lambda)\right]^2\cdot d\nu(\lambda)+o_p(1)$ implies that $\Pr\left[n\cdot\widehat{\mathcal{IT}}_{1n}>k_n^2\right]\to 1$ by Theorem 8.13 in White (1994, p. 181). ■

Proof of Lemma 5. The proof is analogous to that of Lemma 2. ■

Proof of Theorem 6. For part (a), similar to the proof of Theorem 3, the proof is in three steps. First, we show (STEP 1) that

$$\sqrt{n}\left[\hat{\mathcal{T}}_{2n}(\lambda)-\mathcal{T}_{20}(\lambda)\right]=n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{\kappa_2(X_i,Y_i,Z_i,\lambda)-E\left[\kappa_2(X_i,Y_i,Z_i,\lambda)\right]\right\}+\sqrt{n}I_2(\lambda)$$

and $\sup_{\lambda\in\Lambda}|\sqrt{n}I_2(\lambda)|=o_p(1)$. Then, we show (STEP 2) that for finite collections of $\lambda's$, $\sqrt{n}\left[\hat{\mathcal{T}}_{2n}(\lambda)-\mathcal{T}_{20}(\lambda)\right]$ is jointly asymptotically normal. Finally, we show (STEP 3) that $n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{\kappa_2(X_i,Y_i,Z_i,\lambda)-E\left[\kappa_2(X_i,Y_i,Z_i,\lambda)\right]\right\}$ is a tight process in Λ by showing that it is stochastically equicontinuous. STEP 1: We can show that

$$\sqrt{n}I_2(\lambda)\equiv\sqrt{n}\left[\hat{\mathcal{T}}_{2n}(\lambda)-\mathcal{T}_{20}(\lambda)-n^{-1}\sum_{i=1}^{n}\left\{\kappa_2(X_i,Y_i,Z_i,\lambda)-E\left[\kappa_2(X_i,Y_i,Z_i,\lambda)\right]\right\}\right]$$
$$=\sqrt{n}I_{21}(\lambda)+\sqrt{n}I_{22}(\lambda),$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

where

$$I_{21}(\lambda) \equiv \sum_{k=1}^{5} I_{21,k}(\lambda); \quad I_{21,k} \equiv \int\int J_{21,k}(x,z) \cdot \frac{\partial}{\partial x}[\psi(x,\lambda)] \cdot dx \cdot dz, \quad k = 1,2,\ldots,5,$$

$$J_{21,1}(x,z) \equiv -\left[\int y \cdot f(y,x,z) \cdot dy\right] \cdot \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1} + f(x,z)^{-2} \cdot \Delta\hat{f}(x,z)\right] \cdot f(z),$$

$$J_{21,2}(x,z) \equiv -\left[\int y \cdot \Delta\hat{f}(y,x,z) \cdot dy\right] \cdot \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right] \cdot \Delta\hat{f}(z),$$

$$J_{21,3}(x,z) \equiv -\left[\int y \cdot f(y,x,z) \cdot dy\right] \cdot \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right] \cdot \Delta\hat{f}(z),$$

$$J_{21,4}(x,z) \equiv -f(x,z)^{-1} \cdot \left[\int y \cdot \Delta\hat{f}(y,x,z) \cdot dy\right] \cdot \Delta\hat{f}(z),$$

$$J_{21,5}(x,z) \equiv -f(z) \cdot \left[\int y \cdot \Delta\hat{f}(y,x,z) \cdot dy\right] \cdot \left[\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right],$$

$$\begin{aligned} I_{22}(\lambda) \equiv & \int\int J_{22}(x,z) \cdot \frac{\partial}{\partial x}[\psi(x,\lambda)] \cdot dz \cdot dx \\ & - n^{-1} \sum_{i=1}^{n} \{\kappa_2(X_i,Y_i,Z_i,\lambda) - E[\kappa_2(X_i,Y_i,Z_i,\lambda)]\}, \end{aligned}$$

$$\begin{aligned} J_{22}(x,z) \equiv & \mu(x,z) \cdot f(x,z)^{-1} \cdot f(z) \cdot \Delta\hat{f}(x,z) - \mu(x,z) \cdot \Delta\hat{f}(z) \\ & - \left[\int y \cdot \Delta\hat{f}(x,y,z) \cdot dy\right] \cdot f(x,z)^{-1} \cdot f(z). \end{aligned}$$

We can show that $\sup_{\lambda\in\Lambda}\left|\sqrt{n}I_{21,k}(\lambda)\right| = o_p(1)$ for $k = 1,\ldots,5$ and $\sup_{\lambda\in\Lambda}\left|\sqrt{n}I_{22}(\lambda)\right| = o_p(1)$. The details of the proof are similar to those of Theorem 3.

STEP 2: This is a direct application of the Cramer–Wold device and the Lindeberg–Levy central limit theorem (see, e.g., White, 2001, p. 114).

STEP 3: We decompose $\kappa_2(X,Y,Z,\lambda)$ into two terms:

$$\kappa_2(X,Y,Z,\lambda) = \kappa_{21}(X,Y,Z,\lambda) + \kappa_{22}(Z,\lambda),$$

where

$$\kappa_{21}(X,Y,Z,\lambda) \equiv -[Y - \mu(X,Z)] \cdot \frac{f(Z)}{f(X,Z)} \cdot \frac{\partial}{\partial x}[\psi(X,\lambda)],$$

$$\kappa_{22}(Z,\lambda) \equiv -\int \mu(x,Z) \cdot \frac{\partial}{\partial x}[\psi(x,\lambda)] \cdot dx.$$

We show that each of the two terms satisfies Ossiander's L^2 entropy condition. It is sufficient to show that there exist some finite constants $\delta > 0$ and $\phi > 0$ such that

$$E\left\{\sup_{\lambda:\|\lambda-\lambda'\|<\delta}\left|\kappa_{21}(X,Y,Z,\lambda) - \kappa_{21}\left(X,Y,Z,\lambda'\right)\right|^2\right\} \lesssim \delta^{\phi}, \tag{A.8}$$

$$E\left\{\sup_{\lambda:\|\lambda-\lambda'\|<\delta}\left|\kappa_{22}(Z,\lambda) - \kappa_{22}\left(Z,\lambda'\right)\right|^2\right\} \lesssim \delta^{\phi}. \tag{A.9}$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

The details of the proofs are similar to those of Theorem 3. The proof of part (a) is thus complete. Part (b) is a direct application of the continuous mapping theorem after imposing $\mathbb{H}_{20}$. ■

Proof of Theorem 7. The proof is analogous to that of Theorem 4. ■

Proof of Lemma 8. For part (a), it is immediate that $\mathbb{H}''_{30}$ implies $\mathbb{H}_{30}$. $\mathbb{H}_{30}$ says

$$\frac{\partial \mu(x,z)}{\partial x} = \theta_1(z) \quad \text{for almost all } (x,z). \tag{A.10}$$

Taking the integral with respect to x on both sides of equation (A.10) shows that $\mathbb{H}_{30}$ implies $\mathbb{H}''_{30}$. For part (b), it is immediate that $\mathbb{H}'''_{30}$ implies $\mathbb{H}'_{30}$. $\mathbb{H}'_{30}$ says that

$$\frac{\partial \mu(x,z)}{\partial x} = 0 \quad \text{for almost all } (x,z). \tag{A.11}$$

Taking the integral with respect to x on both sides of equation (A.11) shows that $\mathbb{H}'_{30}$ implies $\mathbb{H}'''_{30}$. ■

Proof of Lemma 9. The proof is analogous to that of Lemma 8. ■

Proof of Lemma 10. Under $\mathbb{H}''_{40}$,

$$\begin{aligned} E\left(Y^s \mid X=x, Z=z\right) &= E\left[\left[r_1(U)+X\cdot r_2(U)\right]^s \middle| X=x, Z=z\right] \\ &= E\left[\sum_{j=0}^{s}\binom{s}{j}\cdot\left[r_1(U)\right]^{s-j}\cdot\left[X\cdot r_2(U)\right]^j \middle| X=x, Z=z\right] \\ &= \sum_{j=0}^{s} x^j \cdot E\left[\binom{s}{j}\cdot\left[r_1(U)\right]^{s-j}\cdot\left[r_2(U)\right]^j \middle| Z=z\right] \\ &= \sum_{j=0}^{s} x^j \cdot g_{s,j}(z), \quad s=1,\ldots,S. \end{aligned}$$

Thus, the proof is complete. ■

LEMMA A1. *Suppose Assumptions A.1–A.4 and B.1–B.3 hold, then*

(a)

$$\left|\Delta\left[\hat{f}(x)^{-1}\right]\right| \lesssim f(x)^{-2} A_n;$$

(b)

$$\left|\Delta\left[\hat{f}(x)^{-1}\right] + f(x)^{-2}\cdot\Delta\hat{f}(x)\right| \lesssim f(x)^{-3} A_n^2;$$

(c)

$$\left|\tau\left(\hat{f}(x,z),t\right) - f(x,z)^{-1}\right| \le t^{-2}\cdot A_n + t^{-3}\cdot A_n^2 + \mathbf{1}[f(x,z) < 2t]\cdot f(x,z)^{-1}.$$

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press

Proof. For part (a),

$$\left|\Delta\left[\hat{f}(x)^{-1}\right]\right| = \left|\frac{1}{\hat{f}(x) f(x)}\right| \left|\hat{f}(x) - f(x)\right| \lesssim f(x)^{-2} A_n.$$

For part (b),

$$\begin{aligned}
&\left|\Delta\left[\hat{f}(x)^{-1}\right] + f(x)^{-2} \cdot \Delta\hat{f}(x)\right| \\
&\quad = \left|\frac{f(x) - \hat{f}(x)}{\hat{f}(x) f(x)} + f(x)^{-2}\left(\hat{f}(x) - f(x)\right)\right| \\
&\quad = \left|\frac{1}{\hat{f}(x) f(x)} - f(x)^{-2}\right| \left|\hat{f}(x) - f(x)\right| \\
&\quad = f(x)^{-2} \left|\frac{1}{\hat{f}(x)}\right| \left|\hat{f}(x) - f(x)\right|^2 \lesssim f(x)^{-3} A_n^2.
\end{aligned}$$

For part (c),

$$\begin{aligned}
&\left|\tau\left(\hat{f}(x,z), t\right) - f(x,z)^{-1}\right| \\
&\quad = \left|\left[\tau\left(\hat{f}(x,z), t\right) - \tau\left(f(x,z), t\right)\right] + \left[\tau\left(f(x,z), t\right) - f(x,z)^{-1}\right]\right| \\
&\quad = \left|\tau'\left(f(x,z), t\right) \cdot \left[\hat{f}(x,z) - f(x,z)\right] + \tau''\left(\tilde{f}(x,z), t\right) \cdot \left[\hat{f}(x,z) - f(x,z)\right]^2\right. \\
&\qquad \left. + \left[\tau\left(f(x,z), t\right) - f(x,z)^{-1}\right]\right| \\
&\quad \le t^{-2} \cdot \left|\hat{f}(x,z) - f(x,z)\right| + t^{-3} \cdot \left|\hat{f}(x,z) - f(x,z)\right|^2 + \mathbf{1}[f(x,z) < 2t] \cdot f(x,z)^{-1} \\
&\quad \le t^{-2} \cdot A_n + t^{-3} \cdot A_n^2 + \mathbf{1}[f(x,z) < 2t] \cdot f(x,z)^{-1}.
\end{aligned}$$

Thus the proof is complete. ■

https://doi.org/10.1017/S0266466614000620 Published online by Cambridge University Press