A Discontinuity Test for Identification in Nonlinear Models with Endogeneity

Carolina Caetano, Christoph Rothe, and Nese Yildiz*

PRELIMINARY AND INCOMPLETE

Abstract

In this paper, we consider a triangular system of equations with a potentially endogenous variable whose distribution has a mass point at the lower boundary of its support, but is otherwise continuous. We show that, together with a weak continuity condition on the structural function, this setup yields a testable implication of the set of assumptions that is commonly used in this class of models to achieve identification of various structural quantities through a control variable approach.

JEL Classification: C12, C14, C31, C36, C52

Keywords: *Nonseparable models, triangular systems, control variables, specification testing, identification.*

*This Version: November 23, 2013. Caetano: Department of Economics, University of Rochester, 238 Harkness Hall, P.O. Box: 270156, Rochester, NY 14627, Email: carol.caetano@rochester.edu. Rothe: Department of Economics, Columbia University, 420 W 118th St., New York, NY 10027, Email: cr2690@columbia.edu. Yildiz: Department of Economics, University of Rochester, 238 Harkness Hall, P.O. Box: 270156, Rochester, NY 14627, Email: nese.yildiz@rochester.edu.

1. Introduction

Econometric models with nonseparable unobservables have recently received considerable attention in econometrics (e.g. Matzkin, 2003; Chesher, 2003; Imbens and Newey, 2009; Blundell and Matzkin, 2010). These models are important because they can accommodate general forms of unobserved heterogeneity, including heterogeneous responses to policy interventions among observationally identical individuals. Both economic theory and empirical evidence strongly suggest that such general forms of unobserved heterogeneity are a common feature of economic data (Heckman, 2001). When explanatory variables are endogenous, however, the nonseparability of unobserved heterogeneity complicates identification of structural quantities relative to standard models with additively separable disturbances.

There exist various ways to achieve identification in nonseparable models with endogeneity. If the model has a triangular structure, one possibility is to use a control variable approach. This requires certain conditions on the primitives of the model to ensure that there exists a so-called control variable that can be identified from observable quantities, and that, when conditioned on, makes covariates and disturbances independent. Control variables can be used to identify a wide range of structural parameters, and have been studied extensively in the literature. One common concern about this method is that the conditions necessary for its validity could be overly restrictive. For example, Imbens and Newey (2009) prove the existence of a control variable in a general triangular model under the assumptions that instruments and disturbances are independent, and that a reduced form for the endogenous variable is strictly monotonic in a scalar disturbance. These conditions are not innocuous, and imply substantial restrictions on the underlying economic model. It would thus be important to test their validity in specific empirical applications.

In general, the conditions that are necessary to justify a control variable approach have no testable implications beyond continuity of the conditional distribution of the endogenous variable given the instruments. The validity of the control variable approach can thus not be rejected using data alone without imposing additional restrictions on the model. One such restriction is assuming that the outcome equation is additively separable in the unobservables as in (Newey, Powell, and Vella, 1999). Under this assumption conditional expectation of the outcome given the endogenous covariate and the control variable will be additively separable in the control variable, which is a testable implication. In this paper, we argue that the conditions that are necessary to justify a control variable approach are also testable if the setup exhibits a different structure. Specifically, we study a nonseparable model where the distribution of the endogenous covariate has a mass point at the lower (or upper) boundary of its support, is otherwise continuously distributed, and exerts a continuous effect on the outcome variable of interest. Our main contribution is to show that in such a setting it is possible to test the validity of the control variable approach to identification by checking whether a certain identified function is continuous at one particular point. To the best of our knowledge, our paper is the first to propose a test for this type of hypothesis.

Our test is related to the one presented in Caetano (2012), which exploits the idea that endogeneity of a covariate with the above-mentioned properties can be detected through a discontinuity in the conditional expectation of the outcome variable given the covariate at the boundary. The main motivation for the approach in the present paper is that if the true causal effect of the covariate is continuous, then additional conditioning on a control variable should remove this discontinuity. If the discontinuity remains, however, one has to conclude that the control variable approach is invalid, and that thus some of the assumptions that were made to justify it have to be violated.

This basic idea is not directly feasible, because in our model the control variable is only identified in the subpopulation whose realization of the endogenous variable differs from the mass point. While this is suffices to identify the conditional expectation of the outcome variable given the endogenous variable and the control variable in any neighborhood of the mass point, it is not enough to learn the value of that function *at* the mass point. Therefore it is not directly possible to check for the presence of a discontinuity. We address this issue by deriving a necessary condition for our main testable implication, which is feasible to verify in applications.

While the main contribution of this paper is to point out the existence of testable implications of assumptions commonly made to justify a control variable approach, the question how these implications can be tested in applications is of course also of interest. Our paper therefore proposes a test statistic based on the strategy we just laid out. The computation of our test statistic is straightforward, as it involves only bivariate nonparametric regression and a simple numerical integration step. The test statistic is asymptotically normal and converges at the one-dimensional nonparametric rate under standard conditions. We give an explicit formula for its asymptotic variance, which can be estimated to obtain critical values. Deriving the theoretical properties of our test statistic is a non-standard problem as it involves running nonparametric regressions on estimated data points (the control variable is unobserved, and has to be estimated from the data). To account for this two-stage structure, we use recent results in Mammen, Rothe, and Schienle (2012, 2013) on generated covariates in nonparametric models.

Requiring the endogenous variable to be bounded from below (or above) and to have a mass point at the lower (or upper) boundary of its support obviously restricts the number of empirical applications in which our testing procedure can be applied. Yet there are many variables with such a property that appear frequently as potentially endogenous covariates in empirical applications: weekly hours of work have to be non-negative, and a sizable fraction of the population does not to work; hourly wages cannot be lower than the local minimum wage, and a sizable fraction of the population earns exactly the legal minimum; the amount of consumption of some product also has to be non-negative, and a sizable fraction of the population might not consume this product at all. Many other examples could easily be given. Our test should therefore be useful for a wide range of empirical settings.

Our paper contributes to an extensive literature on identification in nonlinear models with endogeneity. Control variable methods for non- and semi-parametric triangular models are studied by Newey, Powell, and Vella (1999), Blundell and Powell (2003, 2004), Imbens (2007), Imbens and Newey (2009), Rothe (2009) and Kasy (2011), among others. Kasy (2013) considers identification in triangular systems under monotonicity restrictions on the instrument. Instrumental variable (IV) approaches to identification in nonparametric models with additive disturbances are studied in Newey and Powell (2003), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), or Darolles, Fan, Florens, and Renault (2011). Chernozhukov and Hansen (2005) and Chernozhukov, Imbens, and Newey (2007) consider IV methods in nonseparable models, but with restrictions on the dimension of the disturbances. Canay, Santos, and Shaikh (2012) show that the completeness condition, which is necessary for nonparametric IV approaches, is generally not testable.

The remainder of the paper is structured as follows. In Section 2, we formally introduce our model and review the control variable approach to identification. In Section 3, we explain the testing problem and describe our testing approach. In Section 4, we derive the test statistic and study its asymptotic properties. Numerical properties are studied in Section 5, and Section 6 concludes. All proofs are collected in the appendix.

2. Model and Identification

2.1. Model. In this paper, we consider a triangular system of simultaneous equations as the data generating process. For simplicity, the system only contains a single potentially endogenous covariate, but it is allowed to contain an arbitrary number of additional exogenous ones. The main difference between our setup relative to standard existing frameworks is that we consider the case of an endogenous variable whose distribution has a mass point at the lower bound of its support. Such a model could be adequate in settings where natural or legal restrictions lead to corner solutions in the individuals' optimization problem that determines the value of the endogenous variable. Examples for such variables include weekly hours of work (which have to be non-negative), hourly wages (which have to exceed the minimum wage), or the amount of consumption of a particular good (which has to be non-negative). Specifically, our model is given by

$$Y = m_1(X, W, U), \tag{2.1}$$

$$X = \max\{c, m_2(Z, W, V)\}, \tag{2.2}$$

where Y is the outcome of interest, X is a scalar and potentially endogenous covariate, W is a d_W-dimensional vector of additional exogenous covariates, U and V denote unobserved heterogeneity, and c is a known constant.

Our model reduces to the one studied in Imbens and Newey (2009) for $c \leq \inf\{x : x \in \text{supp}(X)\}$, since then the maximum operator in (2.2) would obviously become redundant. In this paper, however, we assume that c is such that

$$0 < \Pr(m_2(Z, W, V) \leq c|Z, W) < 1 \text{ with probability } 1, \tag{2.3}$$

so that the conditional distribution of X given (Z, W) has an actual mass point at c, but is not degenerate. We also assume that the structural function m_1 satisfies a weak continuity property in its first argument:

$$\lim_{x \downarrow c} \mathbb{E}(m_1(x, w, U)|V = v) = \mathbb{E}(m_1(c, w, U)|V = v) \text{ for all } (w, v) \in \text{supp}(W, V) \tag{2.4}$$

A sufficient condition for m_1 having this property is that the mapping $x \mapsto m_1(x, W, U)$ is right-continuous at $x = c$ with probability 1. This seems to be a reasonable assumption for many, although certainly not all, empirical settings. For the remainder of this paper, we take the threshold

c to be zero. This can be done without loss of generality by subtracting c from both X and $m_2(Z, W, V)$.

2.2. Identification. Imbens and Newey (2009) have shown that in the model (2.1)–(2.2) with $c \leq \inf\{x : x \in \mathrm{supp}(X)\}$ a large class of interesting structural parameters can be identified through control variable arguments. The idea behind this type of identification approach is to impose additional conditions on the primitives of the model that ensure that there exists a so-called control variable that can be identified from observable quantities, and that, when conditioned on, makes covariates and disturbances independent. The purpose of this subsection is simply to show that an analogous result applies to our model (2.1)–(2.4) under the same conditions.

Assumption 1. *The model in* (2.1)–(2.4) *satisfies the following restrictions:*

(i) (W, Z) *and* (U, V) *are stochastically independent.*

(ii) V *is scalar and continuously distributed, and without loss of generality its distribution is normalized such that* $V \sim U[0, 1]$.

(iii) The function $v \mapsto m_2(Z, W, v)$ *is strictly monotone, normalized increasing, with probability 1.*

Part (i) of Assumption 1 requires full independence between the instruments and the unobserved heterogeneity, whereas parts (ii)–(iii) imply that individuals with the same realization of the vector (X, Z, W) are also identical in terms of the unobserved heterogeneity V (as long as $X > 0$), and would thus respond identically to exogenous variation in the instruments. The following proposition shows that Assumption 1, which is arguably strong, implies that

$$V^* = F_{X|ZW}(X; Z, W)$$

is a valid control variable in the subpopulation with $X > 0$.

Proposition 1. *Under Assumption 1 it holds that* $U \perp (X, W)|V^*$ *in the subpopulation with* $X > 0$.

Proof. It follows from routine calculations as in Imbens and Newey (2009) that under Assumption 1 we have that $U \perp (X, W)|V$, and that $V = V^*$ in the subpopulation with $X > 0$ (whereas we can only conclude that $0 \leq V \leq V^*$ in the subpopulation with $X = 0$). Taken together, this implies the statement of the proposition. $\square$

Identification of various structural parameters of interest requires an additional support condition of the control variable V^*.

Assumption 2. *The support supp*$(V^*|X = x, W = w)$ *of* V^* *conditional on* $X = x$ *and* $W = w$ *is equal to* $[0, 1]$ *for all* $(x, w) \in$ *supp*$(X, W|X > 0)$.

Assumption 2 is a generalization of the usual rank condition in traditional IV models. This condition would be satisfied if the function $z \mapsto m_2(z, w, V)$ exhibits a sufficient amount of variation over the support of (Z, W). It thus requires the instruments to be strong enough to induce a large amount of variation in the endogenous variable. While Assumption 2 is thus restrictive, it is also directly testable as it only involves observable quantities.

To illustrate how Assumptions 1 and 2 can be used to identify structural objects of interest in our model (2.1)–(2.4), consider the case of the Average Structural Function (ASF), defined by Blundell and Powell (2003) as

$$a(x, w) = \mathbb{E}(m_1(x, w, U)),$$

which describes the average outcome of an individual whose covariates (X, W) are exogenously fixed at (x, w). The following proposition shows that the ASF can be identified for $x > 0$ by traditional arguments, and for $x = 0$ through the continuity property (2.4). Similar arguments apply to other structural parameters that are commonly studied in the literature.

Proposition 2. *Suppose that Assumptions 1 and 2 hold. Then*

$$a(x, w) = \begin{cases} \int_0^1 \mathbb{E}(Y|X = x, W = w, V^* = v)dv & \text{if } x > 0, \\ \lim_{x \downarrow 0} \int_0^1 \mathbb{E}(Y|X = x, W = w, V^* = v)dv & \text{if } x = 0, \end{cases}$$

and thus the ASF $a(x, w)$ *is identified for all* $(x, w) \in$ *supp*(X, W).

Proof. This result is a minor extension of Theorem 1 in Imbens and Newey (2009). From Proposition 1, it follows that $\mathbb{E}(Y|X = x, W = w, V^* = v) = \mathbb{E}(m_1(x, w, U)|V = v)$ for all $x > 0$ and all v. Identification of the average structural function for $x > 0$ then follows because $a(x, w) := \int_0^1 \mathbb{E}(m_1(x, w, U)|V = v)dv$, which is well defined because of Assumption 2. The continuity condition (2.4) then implies identification of the average structural function at $x = 0$, because $a(0, w) = \int_0^1 \lim_{x \downarrow 0} \mathbb{E}(m_1(x, w, U)|V = v)dv = \lim_{x \downarrow 0} a(x, w)$. □

3. Testing the Validity of a Control Variable Approach

As argued above, the assumptions that are necessary to justify a control variable approach are not innocuous, and can be restrictive in a wide range of settings. When identification arguments rely on a control variable approach, it is thus important to have a procedure that allows researchers to investigate empirically whether these conditions holds or not in their particular application. This section describes such a procedure. We first carefully formulate our testing problem, and show that without the mass point in the distribution of the endogenous variable there would be no test with non-trivial power for this problem. We then describe our approach and how it exploits the mass point, and discuss the types of alternatives it is potentially able to detect.

3.1. Testing Problem. Our aim in this paper is to test the assumptions that justify the validity of the control variable approach to identification of structural parameters. In that context, it is important to carefully state the main hypothesis being tested and the set of maintained assumptions. In this paper, we maintain the assumption that the data generating process is of the form given in (2.1)–(2.4). We also maintain a weakened version of Assumption 2, namely that

$$\operatorname{supp}(V^*|X=x,W=w)=(0,1) \text{ for all } (x,w)\in\operatorname{supp}(X,W|\delta>X>0) \text{ and some } \delta>0. \tag{3.1}$$

Our justification for maintaining this condition is that, similarly to Assumption 2, it only involves observable quantities, and can thus be directly verified by practitioners in the context of an empirical application. Given our maintained assumptions, the central condition for the validity of the control variable approach is Assumption 1. Our main testing problem is thus given by the following pair of hypotheses:

$$\mathbb{H}_0\text{: Assumption 1 holds} \quad \text{vs.} \quad \mathbb{H}_1\text{: Assumption 1 is violated.} \tag{3.2}$$

While the testing problem in (3.2) is important in empirical applications, it is not obvious a priori that this pair of hypotheses is actually testable. Indeed, the following proposition formally shows that Assumption 1 would have no testable implications in the model considered by Imbens and Newey (2009) beyond continuity of the conditional distribution of X given (Z,W).

Proposition 3. *Suppose that the data are generated according to the model* (2.1)–(2.2) *with* $c\leq\inf\{x:x\in \mathit{supp}(X)\}$*, and that* X *is continuously distributed given* (Z,W)*. Then Assumption 1 has no further testable implications.*

Proof. Put $\widetilde{m}_1(x, w, u) = Q_{Y|XW}(u|x, w)$ and $\widetilde{m}_2(x, w, v) = Q_{X|ZW}(v|x, w)$, where $Q_{A|B}(\tau|b)$ denotes the conditional τ-quantile of A given $B = b$. Also let $(\widetilde{U}, \widetilde{V})$ be a bivariate random vector drawn independently of the data from the distribution of $(F_{Y|XW}(Y|X, W), F_{X|ZW}(Y|X, W))$. Then the distribution of the random vector $(\widetilde{Y}, \widetilde{X}, W, Z)$ generated by the triangular system

$$\widetilde{Y} = \widetilde{m}_1(\widetilde{X}, W, \widetilde{U}) \text{ and } \widetilde{X} = \widetilde{m}_2(Z, W, \widetilde{V})$$

is identical to the distribution of (Y, X, W, Z). The above system is thus observationally equivalent to (2.1)–(2.2) with $c \leq \inf\{x : x \in \text{supp}(X)\}$, but its components clearly satisfy Assumption 1 up to differences in notation. $\square$

3.2. Testing Approach. The result in Proposition 3 shows that our conditions (2.3)–(2.4) are of central importance for deriving a testable implication of Assumption 1 beyond continuity properties of the distribution of X. Indeed, we now show that due to the specific structure of our model Assumption 1 implies a continuity condition on an identified function that can be exploited for testing purposes. To motivate our approach, define the function

$$\bar{\Delta}(w, v) = \lim_{x \downarrow 0} \mu(x, w, v) - \mu(0, w, v),$$

where

$$\mu(x, w, v) = \mathbb{E}(Y|X = x, W = w, V = v)$$

is the conditional expectation function of Y given (X, W, V). Using the structure of the model, the latter function can clearly be written as

$$\mu(x, w, v) = \begin{cases} \mathbb{E}(m_1(x, w, U)|m_2(Z, w, v) = x, W = w, V = v) & \text{if } x > 0, \\ \mathbb{E}(m_1(x, w, U)|m_2(Z, w, v) \leq x, W = w, V = v) & \text{if } x = 0. \end{cases}$$

Since the conditioning sets in the two conditional expectations on the right-hand side of the previous equation differ, we would generally expect the function $\mu(x, w, v)$ to be discontinuous at $x = 0$ for at least some (and potentially all) values $(w, v) \in \text{supp}(W) \times (0, 1)$, and thus that $\bar{\Delta}(w, v)$ is non-zero over some set with positive probability. Under the null hypothesis, however, this discontinuity vanishes. To see this, note that if Assumption 1 holds we have that $U \perp (Z, W)|V$, and thus the function $\mu(x, w, v)$ simplifies to

$$\mu(x, w, v) = \mathbb{E}(m_1(x, w, U)|V = v).$$

It then follows from condition (2.4) that $\mu(x,w,v)$ is right-continuous at $x=0$, and thus that

$$\Pr(\bar{\Delta}(W,V)=0)=1 \tag{3.3}$$

under the null hypothesis.

The result in (3.3) does not immediately yield a testable implication because the function $\bar{\Delta}$ is not identified even under the null hypothesis. Recall from the proof of Proposition 1 that under $\mathbb{H}_0$ the relationship $V^*=V$ holds in the subpopulation with $X>0$ only, whereas in the subpopulation with $X=0$ realizations of V^* are only an upper bound on realizations of V. This means that while we are able to learn the function $\lim_{x\downarrow 0}\mu(x,w,v)$ under the null hypothesis, the data is not informative about $\mu(0,w,v)$, and thus does not identify $\bar{\Delta}$.

To address this problem, we derive a necessary condition for the null hypothesis in (3.2) which only involves quantities that are identified from the data. This condition can then be used as the basis for a feasible testing procedure. We define the function

$$\Delta(w)=\int \lim_{x\downarrow 0}\mu^*(x,w,v)\Gamma(dv,w)-\mathbb{E}(Y|X=0,W=w), \tag{3.4}$$

where

$$\mu^*(x,w,v)=\mathbb{E}(Y|X=x,W=w,V^*=v), \tag{3.5}$$

$$\Gamma(v,w)=\frac{F_{U[0,1]}(v)-\Pr(V^*\leq v|X>0,W=w)\Pr(X>0|W=w)}{\Pr(X=0|W=w)}, \tag{3.6}$$

and $F_{U[0,1]}$ is the CDF of the uniform distribution on the unit interval. The definition of $\Delta(w)$ only involves observed or identified quantities, which together with the maintained assumption (3.1) implies that the function itself is well-defined and identified from the data under both the null hypothesis and the alternative. The following theorem shows that it is necessary for the null hypothesis to be true that $\Delta(w)$ is equal to zero over the support of W (up to a set of measure zero).

Theorem 1 (Main Testable Implication)**.** *Under the null hypothesis,* $\Pr(\Delta(W)=0)=1$.

Proof. The main insight is that under $\mathbb{H}_0$ we can identify the conditional CDF $F_{V|XW}(v,0,w)$ of V given W in the subpopulation with $X=0$, even though we cannot identify V itself in this subpopulation. To the see this, recall that under the null the unconditional CDF of V is known to be $F_{U[0,1]}$. It thus follows from the law of total probability and exogeneity of W that

$$
\begin{aligned}
F_{U[0,1]}(v) = \Pr(V \leq v|X > 0, W = w) \Pr(X > 0|W = w) \\
+ \Pr(V \leq v|X = 0, W = w) \Pr(X = 0|W = w).
\end{aligned}
$$

The terms $\Pr(X > 0|W = w)$ and $\Pr(X = 0|W = w)$ are clearly identified from the data, and under the null hypothesis we also have that $\Pr(V \leq v|X > 0, W = w) = \Pr(V^* \leq v|X > 0, W = w)$ is identified, because $V = V^*$ in the subpopulation with $X > 0$. Rearranging terms, this shows that $F_{V|XW}(v, 0, w) := \Pr(V \leq v|X = 0, W = w)$ is equal to $\Gamma(v, w)$ under the null hypothesis. By the Law of Iterated Expectations, and using that $\mu^*(x, w, v) = \mu(x, w, v)$ if $x > 0$, we then also find that under the null

$$
\begin{aligned}
\Delta(w) &= \int \left(\lim_{x \downarrow 0} \mu(x, w, v) - \mu(x, w, v) \right) F_{V|XW}(dv, 0, w) \\
&= \int \bar{\Delta}(w, v) F_{V|XW}(dv, 0, w).
\end{aligned}
$$

The statement of the theorem then follows from (3.3). □

A natural way to exploit the result in Theorem 1 for creating a feasible testing procedure is to construct a sample analogue $\widehat{\Delta}(w)$ of $\Delta(w)$, and to reject the null hypothesis when the realization of $\widehat{\Delta}(\cdot)$ is "too large" in some suitable norm. We formally describe such an approach in Section 4.

3.3. Detectable Alternatives. The discussion preceding Theorem 1 suggests that the condition that $\Pr(\Delta(W) = 0) = 1$ is generally only necessary but not sufficient for the null hypothesis to hold. It is therefore possible that there exists a class of fixed data generating processes that belong to the alternative, in the sense that Assumption 1 is violated, but for which nevertheless the conclusion of Theorem 1 holds. Naturally, our test would not be able to detect such an alternative. Instead, it can only be expected to have power against the class of global alternatives that imply that $\Pr(\Delta(W) \neq 0) > 0$. It is thus instructive to gain some understanding about which kind of deviations from the null hypothesis cannot be detected by our methodology before describing the specific form of the test statistic and the decision rule.

One group of alternatives that a test based on Theorem 1 is unable to detect are those where Assumption 1 is violated, but the the joint distribution of (Y, X, W, V^*) is such that the integral on the right-hand-side of equation (3.4) is incidentally equal to $\mathbb{E}(Y|W = w, X = 0)$. We are not aware of any interpretable restrictions on the primitives of the model under which this would be the case, but it is clear that a joint distribution with this property can be constructed in principle. While we therefore cannot rule out the existence of such alternatives, we consider this case as pathological, and would argue that it is unlikely to be encountered in empirical applications.

A second, economically more meaningful group of alternatives that a test based on Theorem 1 cannot detect are those where the control variable is only locally valid for a subgroup of the population with X close to the threshold. To see this, suppose that Assumption 1 is violated such that we no longer have that $U \perp (X, W)|V^*$ in the subpopulation with $X > 0$, but only in the subpopulation with $0 < X < \delta$ for some small $\delta > 0$. Then the control variable approach would be invalid, and e.g. no longer identify the ASF $a(x, w)$ for values $x > \delta$. However, it is easy to see that the conclusion of Theorem 1 would continue to hold in this case, which means that our test would have no power. This general behavior is to be expected, as our test explicitly only exploits potential discontinuities at the threshold. We do not consider this as a weakness of approach, but as an indication of the difficulty to derive *any* testable implication from a condition like Assumption 1.

4. Implementing the Testing Approach in Practice

In this section, we describe a feasible approach to testing our null hypothesis in an empirical application. For simplicity, we focus on the case that the distribution of the exogenous covariates W is discrete with support $\{w_1, \ldots, w_K\}$, but extensions to settings with continuous covariates are conceptually straightforward. The main idea behind the construction of our test statistic is to take a sample analogue $\widehat{\Delta}(w)$ of $\Delta(w)$, and to reject the null hypothesis if the vector

$$\widehat{\Delta} := (\widehat{\Delta}(w_1), \ldots, \widehat{\Delta}(w_K))$$

is "too large" in some suitable norm. We construct such a sample analogue as

$$\widehat{\Delta}(w) = \int \widehat{\mu}^+(w, v)\widehat{\Gamma}(dv, w) - \widehat{\mu}(w), \tag{4.1}$$

where $\widehat{\mu}^+(w, v)$ and $\widehat{\mu}(w)$ are nonparametric estimates of $\lim_{x\downarrow 0} \mathbb{E}(Y|X = x, W = w, V^* = v)$ and $\mathbb{E}(Y|X = 0, W = w)$, respectively, and $\widehat{\Gamma}$ is a nonparametric estimate of the function Γ. The construction of these estimates is described below. In Theorem 2, we then show that the vector

$\widehat{\Delta}$ is asymptotically normal under standard regularity conditions, and characterize the asymptotic variance. Since the latter can be difficult to estimate, we then consider a simple nonparametric bootstrap procedure to obtain asymptotically valid p-values for our test.

4.1. Estimation of Δ. We start by describing the construction of the various components that make up the estimate $\widehat{\Delta}$ of $\Delta := (\Delta(w_1), \ldots, \Delta(w_K))$. The data are given by an i.i.d. sample $\{(Y_i, X_i, Z_i, W_i)\}_{i=1}^n$ of size n from the distribution of (Y, X, Z, W). First, we estimate the conditional distribution function of $F_{X|Z,W}$ of X given (Z, W) by local linear estimation (Fan and Gijbels, 1996):

$$\widehat{F}_{X|Z,W}(x, z, w) = e_{1,d_z}^\top \underset{(a_1, a_2^\top)}{\operatorname{argmin}} \sum_{i=1}^n \Big(\mathbb{I}\{X_i \leq x\} - a_1 - a_2^\top(Z_i - z)\Big)^2 K_h(Z_i - z)\mathbb{I}\{W_i = w\}.$$

Here $K_g(z) = \prod_{j=1}^{d_z} \mathcal{K}(z_j/g)/g$ is a d_z-dimensional product kernel built from the univariate kernel function $\mathcal{K}$, g is a one-dimensional bandwidth that tends to zero as the sample size n tends to infinity, and $e_{1,d_z} = (1, 0, \ldots, 0)^\top$ denotes the first unit $(d_z + 1)$-vector. In a second step, we then use this estimated CDF to define estimates $\{\widehat{V}_i^*\}_{i=1}^n$ of the realizations of the unobserved but identified random variable $V^* = F_{X|Z,W}(X; Z, W)$ as

$$\widehat{V}_i^* = \widehat{F}_{X|Z,W}(X_i, Z_i, W_i) \text{ for } i = 1, \ldots n. \tag{4.2}$$

Third, we estimate function $\Gamma(v, w)$ defined in (3.6) by

$$\widehat{\Gamma}(v, w) = \frac{F_{U[0,1]}(v) - \sum_{i=1}^n \mathbb{I}\{\widehat{V}_i^* \leq v, X_i > 0, W_i = w\}/\sum_{i=1}^n \mathbb{I}\{W_i = w\}}{\widehat{\pi}(w)},$$

where

$$\widehat{\pi}(w) = \frac{\sum_{i=1}^n \mathbb{I}\{X_i = 0, W_i = w\}}{\sum_{i=1}^n \mathbb{I}\{W_i = w\}}$$

is the natural estimate of the conditional probability $P(X = 0|W = w)$. Fourth, we define the estimate $\widehat{\mu}^+(v, w)$ of the function $\lim_{x\downarrow 0} \mathbb{E}(Y|X = x, W = w, V^* = v)$ as

$$\widehat{\mu}^+(v, w) = e_{1,2}^\top \underset{(a_1, a_2^\top)}{\operatorname{argmin}} \sum_{i=1}^n \Big(Y_i - a_1 - a_2^\top(X_i, \widehat{V}_i^* - v)\Big)^2 K_h(X_i, \widehat{V}_i^* - v)\mathbb{I}\{X_i > 0, W_i = w\},$$

where $K_h(x, v) = \mathcal{K}(x/h)\mathcal{K}(v/h)/h^2$ is a bivariate product kernel built from the univariate kernel function $\mathcal{K}$, h is a one-dimensional bandwidth that tends to zero as the sample size n tends to infinity, and $e_{1,2} = (1,0,0)^\top$. Finally, we define the estimate $\widehat{\mu}(w)$ of $\mathbb{E}(Y|X=0,W=w)$ as a sample average of the observed outcomes Y_i among those observations with $(X_i, W_i) = (0, w)$:

$$\widehat{\mu}(w) = \frac{\sum_{i=1}^n Y_i \mathbb{I}\{X_i = 0, W_i = w\}}{\sum_{i=1}^n \mathbb{I}\{X_i = 0, W_i = w\}}.$$

For every $w \in \{w_1, \ldots, w_K\}$, the statistic $\widehat{\Delta}(w)$ is then constructed as described in (4.1). Note that because of the particular structure of the estimate $\widehat{\Gamma}$, the expression given there simplifies to

$$\widehat{\Delta}(w) = \frac{1}{\widehat{\pi}(w)} \left(\int_0^1 \widehat{\mu}^+(v, w) dv - \frac{1}{n} \sum_{i=1}^n \widehat{\mu}^+(\widehat{V}_i^*, w) \mathbb{I}\{X_i > 0, W_i = w\} \right) - \widehat{\mu}(w).$$

The computation of $\widehat{\Delta}(w)$ is thus straightforward, as it only involves calculating sample averages and a one-dimensional numerical integration problem.

4.2. Asymptotic Theory. Deriving the theoretical properties of $\widehat{\Delta}$, and thus of T_n, is a non-standard problem because its construction involves a nonparametric regression on the estimated data points $\{\widehat{V}_i^*\}_{i=1}^n$. We address this issue by using recent results in Mammen, Rothe, and Schienle (2012, 2013) on nonparametric regression with generated covariates. Making use of these results requires the following assumption, which is largely similar to conditions that are commonly imposed in the context of local linear estimation.

Assumption 3. *We assume the following properties for the data distribution, the bandwidths, and kernel function $\mathcal{K}$.*

(i) For every $w \in supp(W)$, the random vector Z is continuously distributed conditional on $W = w$ with support $S_{Z|w} = supp(Z|W = w) \subset \mathbb{R}^{d_Z}$. The corresponding conditional density function $f_{Z|w}(\cdot)$ is continuously differentiable, bounded, and bounded away from zero on $S_{Z|w}$ for every $w \in supp(W)$.

(ii) The conditional CDF $F_{X|Z,W}(x, z, w)$ of X given (Z, W) is twice continuously differentiable with respect to its second argument on $S_{Z|w}$ for every $w \in supp(W)$.

(iii) For every $w \in supp(W)$, the random vector (X, V^*) is continuously distributed conditional on $X > 0$ and $W = w$ with support $S_{XV^*|X>0,w} = supp(X, V^*|X > 0, W = w)$. The corresponding conditional density function $f_{XV^*|X>0,w}(\cdot)$ is continuously differentiable, bounded, and bounded away from zero on the compact set $S_{\delta,w} = \{(x, v) : (x, v) \in S_{XV^*|X>0,w}$ and $x \leq \delta\}$ with $\delta > 0$ as in (3.1) and every $w \in supp(W)$.

(iv) *The conditional expectation function* $\mathbb{E}(Y|X = x, W = w, V^* = v)$ *is twice continuously differentiable in (x,v) on* $S_{\delta,w}$ *for every* $w \in supp(W)$.

(v) *There exist a constant* $\lambda > 0$ *and some constant* $l > 0$ *small enough such that the residuals* $\varepsilon = Y - \mathbb{E}(Y|X, W, V^*)$ *satisfy the inequality* $\mathbb{E}(\exp(l|\varepsilon|\mathbb{I}\{X > 0\})|X, W, V^*) \leq \lambda$.

(vi) *The kernel function* $\mathcal{K}$ *is twice continuously differentiable and satisfies the following conditions:* $\int \mathcal{K}(u)du = 1$, $\int u\mathcal{K}(u)du = 0$, *and* $\mathcal{K}(u) = 0$ *for values of* u *not contained in some compact interval, say* $[-1, 1]$.

(vii) *The bandwidths* g *and* h *satisfy the following conditions as* $n \to \infty$*: (a)* $nh^5 \to 0$*, (b)* $nh^3/\log(n) \to \infty$ *, (c)* $nhg^4 \to 0$ *and (d)* $h^2(ng^{d_z}/\log(n) + g^{-4}) \to \infty$.

As stated above, Assumption 3 collects conditions that are very common in the literature on nonparametric regression. Parts (i) and (iii) ensures that the estimates $\widehat{F}_{X|ZW}(x, z, W)$ and $\widehat{\mu}+(v, w)$ are stable over their respective range of evaluation. Parts (ii) and (iv) are smoothness conditions used to control the magnitude of certain bias terms. Assuming subexponential tails of ε conditional on (X, W, V^*) in the subpopulation with $X > 0$ in part (v) is necessary to apply certain results from Mammen, Rothe, and Schienle (2012, 2013) in our proofs. Part (vi) describes a standard kernel function with compact support. At the expense of technically more involved arguments, this part could be relaxed to also allow for certain kernels with unbounded support. In particular, the Gaussian kernel would be allowed. Finally, part (vii) collects a number of restrictions on the bandwidths that are partly standard, and partly sufficient for certain "high-level" conditions in Mammen, Rothe, and Schienle (2012, 2013).

Assumption 3 allows us to derive the limiting distribution of the random vector $\widehat{\Delta}$. To state the result, we define $f^+_{V^*|XW}(v, 0, w) = \lim_{x\downarrow 0} f_{V^*|XW}(v, x, w)$ and $f^+_{XW}(0, w) = \lim_{x\downarrow 0} f_{XW}(x, w)$, let $\gamma(v, w) = \partial\Gamma(v, w)/\partial v$, and put

$$\sigma^2_+(w) = \lim_{x\downarrow 0} \mathrm{Var}\left(\varepsilon \cdot \frac{\gamma(V^*, W)}{f^+_{V^*|X,W}(V^*, 0, W)}\middle| X = x, W = w\right).$$

where $\varepsilon = Y - \mathbb{E}(Y|X, W, V^*)$ as in Assumption 3(v). Note that under the null hypothesis the function $\sigma^2_+(w)$ can be expressed in the following, somewhat more intuitive form:

$$\sigma^2_+(w) = \lim_{x\downarrow 0} \mathrm{Var}\left(\varepsilon \cdot \frac{f_{V|XW}(V, 0, W)}{f^+_{V|X,W}(V, 0, W)}\middle| X = x, W = w\right).$$

Also, for $j \in \{0, 1, 2\}$ we define the constants

$$\kappa_j = \int_0^\infty x^j \mathcal{K}(x)dx \quad \text{and} \quad \lambda_j = \int_0^\infty x^j \mathcal{K}(x)^2 dx,$$

which depend on the kernel function $\mathcal{K}$ only, put

$$C = \frac{\kappa_2^2\lambda_0 - 2\kappa_1\kappa_2\lambda_1 + \kappa_1^2\lambda_2}{(\kappa_2\kappa_0 - \kappa_1^2)^2},$$

and finally let

$$\rho^2(w) = C \cdot \frac{\sigma_+^2(w)}{f_{XW}^+(0, w)}.$$

With this notation, we obtain the following result.

Theorem 2. *Suppose that Assumption 3 holds. Then*

$$\sqrt{nh}\left(\widehat{\Delta} - \Delta\right) \xrightarrow{d} N(0, \operatorname{diag}(\rho^2(w_1), \ldots, \rho^2(w_K))).$$

Proof. See the appendix. □

The theorem shows that in large samples the distribution of the vector $\widehat{\Delta}$ approaches a multivariate normal distribution with mean Δ and diagonal variance matrix $\operatorname{diag}(\rho^2(w_1), \ldots, \rho^2(w_K))/(nh)$. Note that the rate of convergence is the same as that of a standard one-dimensional kernel smoother.

4.3. Test Statistic and Critical Values. Given the result in Theorem 2, a natural test statistic for the testing problem in (3.2) is given by

$$T_n = nh \sum_{k=1}^{K} \left(\frac{\widehat{\Delta}(w_k)}{\widehat{\rho}(w_k)}\right)^2, \tag{4.3}$$

where $\widehat{\rho}^2(w)$ is some consistent estimate of $\rho^2(w)$ for all $w \in \{w_1, \ldots, w_K\}$. Such a statistic asymptotically follows a χ^2-distribution with K degrees of freedom under the null hypothesis and the conditions of Theorem 2. The testing decision is thus to reject $\mathbb{H}_0$ at the nominal level $\alpha \in (0, 1)$ if

$$T_n > \chi_K^2(1 - \alpha),$$

where $\chi_K^2(\tau)$ denotes the τ-quantile of the χ^2-distribution with K degrees of freedom. The following result formally shows the validity of such an approach.

Theorem 3. *Suppose that Assumption 3 holds, and that* $\widehat{\rho}^2(w) \xrightarrow{p} \rho^2(w)$ *for all* $w \in supp(W)$*. Then the following statements hold.*

(i) Under the null hypothesis, i.e. if $\Delta(w) \equiv 0$,

$$\lim_{n\to\infty} \Pr(T_n > \chi^2_K(1-\alpha)) = \alpha.$$

(ii) Under any fixed alternative that implies $\Delta(w) \neq 0$ *for some* $w \in supp(W)$,

$$\lim_{n\to\infty} \Pr(T_n > \chi^2_K(1-\alpha)) = 1.$$

(iii) Under any local alternative that implies $\Delta(w) = \delta(w)/\sqrt{nh}$ *for all* $w \in supp(W)$,

$$\lim_{n\to\infty} \Pr(T_n > \chi^2_K(1-\alpha)) = 1 - \Xi_{\Gamma,K}(\chi^2_K(1-\alpha))$$

where $\Xi_{\theta,K}$ *is the CDF of the noncentral* χ^2 *distribution with* K *degrees of freedom and noncentrality parameter* $\theta = \sum_{k=1}^{K}(\delta(w_k)/\rho(w_k))^2$.

Proof. Follows from Theorem 2 using straightforward arguments. □

Since Theorem 3 holds for any consistent estimator $\widehat{\rho}^2(w)$ of $\rho^2(w)$, the only remaining issue for applying our test is to find such an estimator that is feasible to compute in the context of an empirical application. One possible approach would be to develop a direct sample-analogue estimator using boundary-corrected nonparametric estimates of the various components of $\rho^2(w)$. Such an estimator is explicitly described in the appendix. However, the approach is unattractive in practice because it requires additional smoothing parameters and turned out to be very sensitive to their choice in our simulation experiments. We therefore recommend the use of a nonparametric bootstrap variance estimator. Such a procedure is computationally expensive, but straightforward from a practical point of view. The estimator is obtained as follows. Let $\{(Y_i^*, X_i^*, Z_i^*, W_i^*)\}_{i=1}^n$ be a bootstrap sample drawn with replacement from the observed data $\{(Y_i, X_i, Z_i, W_i)\}_{i=1}^n$, and let $\widehat{\Delta}^*(w)$ be an estimate of $\Delta(w)$ computed exactly as described in Section 4.1 but using the bootstrap sample. Then

$$\widehat{\rho}^2(w) = \mathbb{E}^*((\widehat{\Delta}^*(w) - \widehat{\Delta}(w))^2),$$

where $\mathbb{E}^*$ denotes the expectation with respect to bootstrap sampling.

5. Numerical Evidence

[TO BE COMPLETED]

6. Concluding Remarks

In this paper, we have proposed a test for the validity of a control variable approach to identification in triangular nonseparable models with endogeneity. To the best of our knowledge, this is the first test of this type of hypothesis. Our test requires a particular data structure, namely that the endogenous covariate has a mass point at the lower (or upper) boundary of its support, and is otherwise continuously distributed. While this is certainly restrictive, we argue that our test is still useful for a wide range of empirical applications. We propose a test statistic that is simple to compute, and derive its asymptotic properties.

A. Mathematical Appendix

A.1. Proof of Theorem 2. The result in Theorem 2 follows directly from the three axillary results in Lemma 1–3 below. To simplify the notation, we assume that the additional exogenous covariates W are absent from the model, and can thus be dropped from the notation in the following. That is, we have $\widehat{\Delta}(w) \equiv \widehat{\Delta}$, $\Delta(w) \equiv \Delta$, etc. The first of these three findings gives a bound on the uniform rate of consistency of the estimated function $\widehat{\Gamma}(\cdot)$.

Lemma 1. *Suppose that the conditions of Theorem 1 hold. Then*

$$\sup_{v \in supp(V|X=0)} |\widehat{\Gamma}(v) - \Gamma(v)| = O_P(n^{-1/2}) + O(g^2).$$

Proof. Up to terms that are clearly $O_P(n^{-1/2})$, the estimate $\widehat{\Gamma}(\cdot)$ is equal to a continuous and deterministic transformation of empirical distribution function of the estimates $\{\widehat{V}_i^*\}_{i=1}^n$ in the subset of the sample with $X_i > 0$. The result then follows from arguments analogous to those in Akritas and Van Keilegom (2001). $\square$

To state our next result, we introduce an infeasible estimator of the function $\mu^+(v) = \lim_{x \downarrow 0} \mathbb{E}(Y|X = x, V^* = v)$ that uses the actual realizations of $V_i^* = F_{X|Z}(X_i, Z_i)$ instead of the corresponding estimated values $\widehat{V}_i^*$. The corresponding estimator is denoted by $\widetilde{\mu}^+_{Y|X,V}(0, v)$. We also define an infeasible version of our test statistic which uses the population values $\Gamma(\cdot)$ and $\mathbb{E}(Y|X = 0)$ instead of their estimates, and replaces $\widehat{\mu}^+(v)$ with its infeasible version $\widetilde{\mu}^+(v)$:

$$\widetilde{\Delta} = \int \widetilde{\mu}^+(v) d\Gamma(v) - \mathbb{E}(Y|X = 0).$$

The following lemma derives the asymptotic properties of the infeasible test statistic $\widetilde{\Delta}$.

Lemma 2. *Suppose that the conditions of Theorem 2 hold. Then*

$$\sqrt{nh}(\widetilde{\Delta} - \Delta) \xrightarrow{d} N\left(0, C \cdot \frac{\sigma_+^2(0)}{f_X^+(0)}\right).$$

as $n \to \infty$.

Proof. To show this result, we first introduce the additional notation that:

$$L_i(v) = (1, X_j/h, (V_i^* - v)/h)^\top \cdot \mathbb{I}\{X_i > 0\},$$
$$M_n(v) = \frac{1}{n}\sum_{i=1}^n L_i(v)L_i(v)^\top K_h(X_i, V_i^* - v),$$
$$N_n(v) = \mathbb{E}(L_i(v)L_i(v)^\top K_h(X_i, V_i^* - v)).$$

With this notation, the local linear estimator $\widetilde{\mu}^+(v)$ can be written as

$$\widetilde{\mu}^+(v) = \frac{1}{n}\sum_{i=1}^n e_1^\top M_n(v)^{-1} L_i(v) K_h(X_i, V_i^* - v) Y_i.$$

It also follows from straightforward calculations that the term $N_n(v)$ satisfies

$$\begin{aligned} N_n(v) &= A \lim_{x \downarrow 0} f_{V^*X}(v, x) + o(1) \\ &= A f_{V^*|X}^+(v, 0) f_X^+(0) + o(1) \end{aligned}$$

uniformly in v, where the matrix A is given by

$$A = \begin{pmatrix} \kappa_0 & \kappa_1 & 0 \\ \kappa_1 & \kappa_2 & 0 \\ 0 & 0 & \kappa_2^* \end{pmatrix} \quad \text{and} \quad \kappa_2^* = \int_{-\infty}^{\infty} x^2 \mathcal{K}(x) dx.$$

Note that the structure of A follows from the assumption that the kernel function $\mathcal{K}$ is a symmetric density function. We now introduce a particular stochastic expansion for this estimator, which follows from standard results in e.g. Masry (1996). Writing

$$S_n(v) = \frac{1}{n}\sum_{i=1}^n e_1^\top N_n(v)^{-1} L_i(v) K_h(X_i, V_i^* - v)\varepsilon_i$$

with $\varepsilon_i = Y_i - \mathbb{E}(Y_i|X_i, V_i^*)$, we have that

$$\widetilde{\mu}^+(v) = \mu^+(v) + S_n(v) + O(h^2) + O_P\left(\frac{\log(n)}{nh^2}\right)$$

uniformly over $v \in supp(V|X = 0)$. Using standard change-of-variables arguments, we find that

$$\int S_n(v) d\Gamma(v) = \frac{1}{n}\sum_{i=1}^n e_1^\top N_n(V_i^*)^{-1} L_i^* K_h(X_i) \Gamma(V_i^*)\varepsilon_i + O(h^2)$$

with $L_i^* = (1, X_i/h, 0)^\top \cdot \mathbb{I}\{X_i > 0\}$. The first term on the right-hand-side of the last equation is a sample average of n independent random variables, and clearly has mean zero. On the other hand, its variance is equal to

$$
\begin{aligned}
&n^{-1}\mathbb{E}((e_1^\top N_n(V_i^*)^{-1}L_i^*)^2 K_h(X_i)^2 \Gamma(V_i^*)^2 \varepsilon_i^2) \\
&\quad = \frac{1}{nhf_X^+(0)^2}\int_0^\infty (e_1^\top A^{-1}(1,x,0)^\top)^2 \mathcal{K}(x)^2 \mathbb{E}\left(\frac{\Gamma(V^*)^2}{f_{V|X}^+(V^*,0)^2}\cdot \varepsilon^2 \middle| X = xh\right) f_X(xh)dx + o\left(\frac{1}{nh}\right) \\
&\quad = \frac{1}{nhf_X^+(0)^2}\int_0^\infty (e_1^\top A^{-1}(1,x,0)^\top)^2 \mathcal{K}(x)^2 dx \cdot \lim_{x\downarrow 0}\mathbb{E}\left(\frac{\Gamma(V^*)^2}{f_{V|X}^+(V^*,0)^2}\cdot \varepsilon^2 \middle| X = xh\right) \cdot \lim_{x\downarrow 0} f_X(x) + o\left(\frac{1}{nh}\right) \\
&\quad = \frac{1}{nhf_X^+(0)} \cdot C \cdot \sigma_+^2(0) + o\left(\frac{1}{nh}\right).
\end{aligned}
$$

The statement of the lemma then follows from an application of Ljapunov's Central Limit Theorem. □

As the final step of our proof of Theorem 2, the following lemma shows that $\widetilde{\Delta}$ and $\widehat{\Delta}$ have the same first order asymptotic properties.

Lemma 3. *Suppose that the conditions of Theorem 1 hold. Then*

$$\widetilde{\Delta} - \widehat{\Delta} = o_P((nh)^{-1/2})$$

as $n \to \infty$.

Proof. First, using that $\widehat{\mu}(0) = \mathbb{E}(Y|X=0) + O_P(n^{-1/2})$ and Lemma 1, we find that

$$\widehat{\Delta} = \int \widehat{\mu}^+(v)d\Gamma(v) - \mathbb{E}(Y|X=0) + O_P(n^{-1/2}),$$

since $\widehat{\mu}_{Y|X,V}^+(0,v)$ is easily seen to be a consistent estimate of a bounded function under the conditions of the lemma. Similarly, we have that

$$\widetilde{\Delta} = \int \widetilde{\mu}^+(v)d\Gamma(v) - \mathbb{E}(Y|X=0) + O_P(n^{-1/2}),$$

It thus only remains to be shown that

$$\int \widehat{\mu}^+(v)d\Gamma(v) = \int \widetilde{\mu}^+(v)d\Gamma(v) + o_P((nh)^{-1/2}).$$

We use recent results on nonparametric regression with generated covariates obtained by Mammen, Rothe, and Schienle (2012, 2013) to show this statement. For convenience, we repeat the following notation, which was already introduced in the proof of Lemma 2:

$$
\begin{aligned}
L_i(v) &= (1, X_i/h, (V_i^* - v)/h)^\top \cdot \mathbb{I}\{X_i > 0\}, \\
M_n(v) &= \frac{1}{n}\sum_{i=1}^n L_i(v)L_i(v)^\top K_h(X_i, V_i^* - v), \\
N_n(v) &= \mathbb{E}(L_i(v)L_i(v)^\top K_h(X_i, V_i^* - v)).
\end{aligned}
$$

It then follows from an application of Theorem 1 in Mammen, Rothe, and Schienle (2013) that

$$\int \widehat{\mu}^{+}(v) - \widetilde{\mu}^{+}(v) - \varphi_n(v; \widehat{F}_{X|Z}) d\Gamma(v) = o_P((nh)^{-1/2})$$

under the conditions of the lemma, where for any conformable function Λ

$$\begin{aligned}\varphi_n(v; \Lambda) = &-(\partial m^{+}(v)/\partial v) e_1^\top N_n(v)^{-1} \mathbb{E}(L_i(v) K_h(X_i, V_i^* - v)(\Lambda(X_i, Z_i) - F_{X|Z}(X_i, Z_i))) \\ &+ e_1^\top N_n(v)^{-1} \mathbb{E}(L_i(v) K_h'(X_i, V_i^* - v)(\Lambda(X_i, Z_i) - F_{X|Z}(X_i, Z_i)) \Psi(X_i, Z_i))\end{aligned}$$

with $\Psi(X_i, Z_i) = \mathbb{E}(Y_i|X_i, Z_i) - \mathbb{E}(Y_i|X_i, V_i^*)$, and $K_h'(x, v) = \partial K_h(x, v)/\partial v$. We remark that the two expectations in the previous equation are both taken with respect to the distribution of (X_i, Z_i), so that the term $\varphi_n(v; \widehat{F}_{X|Z})$ remains a random variable due to its dependence on the estimate $\widehat{F}_{X|Z}$. Also note that under the null hypothesis the second summand in the formula for φ_n vanishes, because under correct specification our model implies that $\Psi(X_i, Z_i) = 0$, and thus the "index bias" term in Mammen, Rothe, and Schienle (2013) is equal to zero. Next, it follows from the same arguments as in the proof of Theorem 4 in Mammen, Rothe, and Schienle (2013) that

$$\int \varphi_n(v; \widehat{F}_{X|Z}) d\Gamma(v) = O_P(n^{-1/2}) + O(h^2) + O(g^2) + O_P\left(\frac{\log n}{ng}\right).$$

This completes our proof. $\square$

B. An Alternative Estimate of the Asymptotic Variance

In this section, we describe a plug-in estimator of $\rho^2(w)$ that uses kernel-based nonparametric smoothers to estimate the various density and conditional expectation functions involved in the definition of the asymptotic variance of $\widehat{\Delta}$. Some technical complications arise from the fact that many of these functions need to be evaluated at or close to the limits of their support. This is a problem for standard kernel estimators, which are well-known to be inconsistent at the boundary, and highly biased in its vicinity. Since we only require a consistent estimate of $\rho^2(w)$, and not one that converges with a particular rate, we adopt a simple solution to this problem and introduce a multiplicative correction term into all estimators of density functions. More elaborate procedures could to used to achieve better rates of convergence, but those are not necessary for our main results. The boundary correction terms are of the form

$$s_b(v) = \bar{\mathcal{K}}(\min\{v, 1-v\}/b)^{-1} \text{ with } \bar{\mathcal{K}}(t) = \int_{-\infty}^{t} \mathcal{K}(u) du \tag{B.1}$$

for any $b \in \mathbb{R}$ and $v \in (0, 1)$. We then estimate the function $\gamma(v, w) = \partial \Gamma(v, w)/\partial v$ by the sample analogue

$$\widehat{\gamma}(v, w) = (1 - \widehat{g}(v, w))/\widehat{p}(w),$$

where

$$\widehat{g}(v,w) = s_{b_1}(v) \cdot \frac{\sum_{i=1}^n K_{b_1}(\widehat{V}_i^* - v)\mathbb{I}\{X_i > 0, W_i = w\}}{\sum_{i=1}^n \mathbb{I}\{W_i = w\}}.$$

Here b_1 is a one-dimensional bandwidth that tends to zero as n tends to infinity. By including the boundary correction term $s_{b_1}(v)$ into the definition of $\widehat{g}(v,w)$, we achieve that the estimator $\widehat{\gamma}$ is uniformly consistent under weak regularity conditions. We also define

$$\widehat{f}^+_{V^*|X,W}(v;x,w) = s_{b_2}(v) \cdot \frac{\sum_{i=1}^n K_{b_2}(\widehat{V}_i^* - v, X_i - x)\mathbb{I}\{X_i > 0, W_i = w\}}{\sum_{i=1}^n K_{b_2}(X_i - x)\mathbb{I}\{X_i > 0, W_i = w\}} \text{ and}$$

$$\widehat{f}^+_{XW}(0,w) = \frac{2}{n}\sum_{i=1}^n K_{b_1}(X_i)\mathbb{I}\{X_i > 0, W_i = w\},$$

where b_2 is another one-dimensional bandwidth that tends to zero as n tends to infinity. Again, consistency of these two estimates for the corresponding population counterparts is achieved by including a boundary correction term for the first estimator, and multiplication by two for the second estimator. The estimate the term $\sigma^2_+(w)$ is given by

$$\widehat{\sigma}^2_+(w) = e_{1,1}^\top \operatorname*{argmin}_{(a_1,a_2)} \sum_{i=1}^n \left(\widehat{\eta}_i - a_1 - a_2(X_i - x)\right)^2 K_{b_1}(X_i - x)\mathbb{I}\{X_i > 0, W_i = w\}.$$

where

$$\widehat{\eta}_i = (Y_i - \widehat{\mu}(X_i, W_i, \widehat{V}_i^*)) \cdot \frac{\widehat{\gamma}(\widehat{V}_i^*, W_i)}{\widehat{f}_{V^*|XW}(\widehat{V}_i^*, 0, W_i)},$$

and

$$\widehat{\mu}(x,w,v) = e_{1,2}^\top \operatorname*{argmin}_{(a_1,a_2^\top)} \sum_{i=1}^n \left(Y_i - a_1 - a_2^\top(X_i - x, \widehat{V}_i^* - v)\right)^2 K_h(X_i - x, \widehat{V}_i^* - v)\mathbb{I}\{X_i > 0, W_i = w\},$$

which is similar in structure to the estimate $\widehat{\mu}^+(w,v)$ defined above. Our final estimator of $\rho^2(w)$ then given by

$$\widehat{\rho}^2(w) = C \cdot \frac{\widehat{\sigma}^2_+(w)}{\widehat{f}^+_{XW}(0,w)}.$$

The constant C depends on the kernel function and can be computed numerically. For example, $C \approx 1.78581$ for the the Gaussian kernel that we use in our simulations below.

Theorem 4. *Suppose that Assumption 3 holds, and that $b_j \to 0$, $nb_j \to \infty$ and $(ng^{d_z}/\log(n) + g^{-4})/b_j^2 \to 0$ as $n \to \infty$ for $j = 1, 2$. Then $\widehat{\rho}^2(w) \xrightarrow{p} \rho^2(w)$ for all $w \in supp(W)$.*

Proof. This result follows from straightforward arguments. Let

$$g(v,w) = \partial \Pr(V^* \leq v, X > 0 | W = w)/\partial v$$

be the population counterpart of $\widehat{g}(v,w)$, and

$$\widetilde{g}(v,w) = s_{b_1}(v) \cdot \frac{\sum_{i=1}^n K_{b_1}(V_i^* - v)\mathbb{I}\{X_i > 0, W_i = w\}}{\sum_{i=1}^n \mathbb{I}\{W_i = w\}}.$$

be an infeasible estimator of $g(v,w)$ that uses the the true $V_i^* = F_{X|ZW}(X,Z,W)$ instead of the estimates $\widehat{V}_i^* = \widehat{F}_{X|ZW}(X,Z,W)$. From a simple Taylor expansion, it follows that

$$\sup_{v,w} |\widehat{g}(v,w) - \widetilde{g}(v,w)| = O_P\left(\max_{i=1,\ldots,n} |\widehat{V}_i^* - V_i^*|/b_1\right) = o_p(1)$$

since $\max_{i=1,\ldots,n} |\widehat{V}_i^* - V_i^*| = O_P((ng^{d_z}/\log(n))^{1/2}) + O(g^{-2})$. Moreover, standard results from kernel density estimation imply that

$$\sup_{v,w} |\widetilde{g}(v,w) - g(v,w)| = o_p(1).$$

Similar arguments can be used to show that $\widehat{f}^+_{V^*|XW}(v;0,w)$ and $\widehat{f}^+_{XW}(0,w)$ are uniformly consistent estimates of $\lim_{x\downarrow 0} f_{V^*|XW}(v;x,w)$ and $\lim_{x\downarrow 0} f_{XW}(x,w)$, respectively. Consistency of $\widehat{\sigma}^2_+(0)$ for $\sigma^2_+(0)$ then follows from the linearity of the local linear smoothing operator. This completes our proof. □

References

Akritas, M. G., and I. Van Keilegom (2001): "Non-parametric Estimation of the Residual Distribution," *Scandinavian Journal of Statistics*, 28(3), 549–567.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75(6), 1613–1669.

Blundell, R., and R. L. Matzkin (2010): "Conditions for the existence of control functions in nonseparable simultaneous equations models," Discussion paper, cemmap working paper.

Blundell, R., and J. Powell (2003): "Endogeneity in nonparametric and semiparametric regression models," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, vol. 2, pp. 655–679.

Blundell, R., and J. Powell (2004): "Endogeneity in semiparametric binary response models," *The Review of Economic Studies*, 71(3), 655–679.

Caetano, C. (2012): "A Discontinuity Test of Endogeneity," *Working Paper.*

Canay, I. A., A. Santos, and A. M. Shaikh (2012): "On the Testability of Identification in some Nonparametric Models with Endogeneity," Discussion paper, Working paper.

Chernozhukov, V., and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73(1), 245–261.

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007): "Instrumental variable estimation of nonseparable models," *Journal of Econometrics*, 139(1), 4–14.

Chesher, A. (2003): "Identification in nonseparable models," *Econometrica*, 71(5), 1405–1441.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011): "Nonparametric instrumental regression," *Econometrica*, 79(5), 1541–1565.

Fan, J., and I. Gijbels (1996): *Local polynomial modelling and its applications*. CRC Press.

Hall, P., and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *The Annals of Statistics*, 33(6), 2904–2929.

Heckman, J. J. (2001): "Micro data, heterogeneity, and the evaluation of public policy," *Journal of Political Economy*, 109(4), 673–748.

Imbens, G., and W. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77(5), 1481–1512.

Imbens, G. W. (2007): "Nonadditive models with endogenous regressors," in *Advances in Economics and Econometrics*, ed. by R. Blundell, W. Newey, and T. Persson, vol. 43.

Kasy, M. (2011): "Identification in Triangular Systems using Control Functions," *Econometric Theory*, 27, 663–671.

——— (2013): "Identification in General Triangular Systems," *Working Paper*.

Mammen, E., C. Rothe, and M. Schienle (2012): "Nonparametric Regression with Nonparametrically Generated Covariates," *Annals of Statistics*.

——— (2013): "Semiparametric Estimation with Generated Covariates," *Working Paper*.

Masry, E. (1996): "Multivariate local polynomial regression for time series: uniform strong consistency and rates," *Journal of Time Series Analysis*, 17(6), 571–599.

Matzkin, R. L. (2003): "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71(5), 1339–1375.

Newey, W., J. Powell, and F. Vella (1999): "Nonparametric estimation of triangular simultaneous equations models," *Econometrica*, 67(3), 565–603.

NEWEY, W. K., AND J. L. POWELL (2003): "Instrumental variable estimation of nonparametric models," *Econometrica*, 71(5), 1565–1578.

ROTHE, C. (2009): "Semiparametric estimation of binary response models with endogenous regressors," *Journal of Econometrics*, 153(1), 51–64.