

ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

*An International Society for the Advancement of Economic
Theory in its Relation to Statistics and Mathematics*

<http://www.econometricsociety.org/>

Econometrica, Vol. 83, No. 4 (July, 2015), 1581–1600

A TEST OF EXOGENEITY WITHOUT INSTRUMENTAL VARIABLES IN MODELS WITH BUNCHING

CAROLINA CAETANO

University of Rochester, Rochester, NY 14627, U.S.A.

The copyright to this Article is held by the Econometric Society. It may be downloaded, printed and reproduced only for educational or research purposes, including use in course packs. No downloading or copying may be done for any commercial purpose without the explicit permission of the Econometric Society. For such commercial purposes contact the Office of the Econometric Society (contact information may be found at the website <http://www.econometricsociety.org> or in the back cover of *Econometrica*). This statement must be included on all copies of this Article that are made available electronically or in any other format.

A TEST OF EXOGENEITY WITHOUT INSTRUMENTAL VARIABLES IN MODELS WITH BUNCHING

BY CAROLINA CAETANO¹

This paper presents a test of the exogeneity of a single explanatory variable in a multivariate model. It does not require the exogeneity of the other regressors or the existence of instrumental variables. The fundamental maintained assumption is that the model must be continuous in the explanatory variable of interest. This test has power when unobservable confounders are discontinuous with respect to the explanatory variable of interest, and it is particularly suitable for applications in which that variable has bunching points. An application of the test to the problem of estimating the effects of maternal smoking in birth weight shows evidence of remaining endogeneity, even after controlling for the most complete covariate specification in the literature.

KEYWORDS: Exogeneity test, discontinuity, nonparametric models, bunching, maternal smoking, birth weight.

1. INTRODUCTION

ENDOGENEITY IS ONE OF THE MOST STUDIED PROBLEMS IN ECONOMETRICS. Failure to address it generally results in biased estimates and therefore wrong conclusions. There are a number of techniques designed to identify effects in models with endogeneity, such as instrumental variables, panel data with fixed effects, and proxy variables. Unfortunately, such techniques are often difficult to apply to several important problems because they rely on the existence of specific data or natural phenomena. Because of this, a test that can detect endogeneity can be useful, especially when it does not require that a solution to the problem be immediately available.

The available tests of exogeneity in the literature are either specification tests (e.g., [Gourieroux, Monfort, Renault, and Trognon \(1987\)](#), [Bierens \(1990\)](#), and [Bierens and Ploberger \(1997\)](#)), or require the existence of instrumental variables (e.g., [Hausman \(1978\)](#) and [Blundell and Horowitz \(2007\)](#)). This paper presents a test of the exogeneity of a single explanatory variable in a multivariate model. The test does not require the existence of instrumental variables, and the other variables in the model may be endogenous. Therefore, this test may be useful for two reasons. First, the test can be useful in the earlier stages of the analysis, before the researcher makes the effort to find and implement any new identification strategy (which could require, e.g., a different data set or the existence of an instrumental variable) in order to deal with a

¹I am especially indebted to Jim Powell, Gregorio Caetano, and David Brillinger for their time and valuable input. This paper also benefited from discussions with Bin Chen and Nese Yildiz, as well as several other colleagues and seminar participants, in fact too many to name. Umair Khalil and Liquan Huang provided excellent research assistance. Douglas Almond, Ken Chay, and David Lee graciously provided the data used in this paper. I am thankful to all of them. All errors are my own.

possible endogeneity problem. Second, the test can be applied as an omitted variable test to provide guidance in choosing the appropriate model, even with an approach deliberately based on selection on observables.

The test is best explained in an applied example. Consider the problem of estimating the marginal effect of the amount a woman smokes during pregnancy on the baby's birth weight. The variable "average number of cigarettes per day" is naturally prone to endogeneity, given that there are many pre-existing selection factors associated with both smoking and birth weight. Examples of such factors include the mother's education level, marital status, age, etc. The question is whether the amount smoked is exogenous after controlling for the observable factors available in the data. If this is the case, then it is possible to identify the marginal effects of maternal smoking on birth weight.

The fundamental maintained assumption of the exogeneity test is that the structural function must be continuous in the explanatory variable of interest. In the smoking example, it means that the mother's smoking amount must have a continuous effect on the baby's birth weight. Suppose that this is indeed true, and then consider Figure 1(a), which illustrates the expected birth weight for each amount smoked (see Figure 7(a) for the empirical version of this figure). If the expected birth weight conditional on the amount smoked is discontinuous, this discontinuity cannot be due to smoking, since smoking has a continuous effect on birth weight. However, this discontinuity may be due to selection on observables, since the observable mother's characteristics may be discontinuous at zero cigarettes. This is indeed the case, as can be seen, for example, in Figures 4(a) to 4(f), 5(a), and 5(b). Next, consider Figure 1(b), which depicts the expected birth weight for each amount smoked for a subgroup of mothers who share the same observable characteristics. If the birth weight per amount smoked is still discontinuous, this discontinuity cannot be caused by smoking (by assumption) nor by selection on observables (because the observable mother's characteristics are held fixed). The only explanation

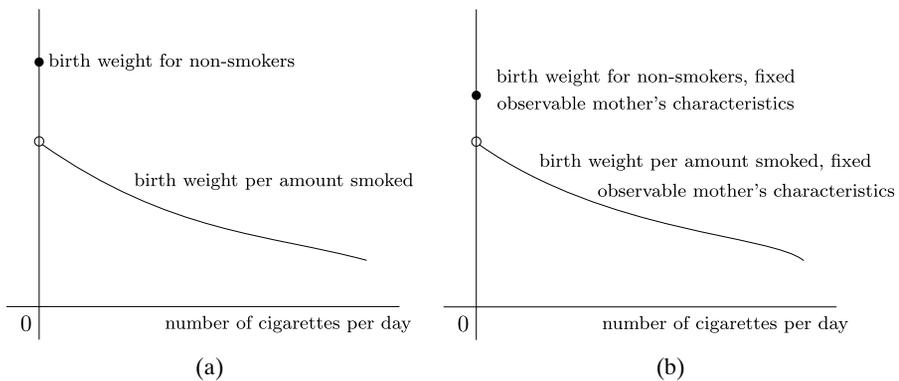


FIGURE 1.—Discontinuities in birth weight at zero cigarettes.

for this discontinuity is that there is at least one confounder that was not included in the structural equation, and thus smoking is endogenous.

The test consists of estimating the expected outcome variable conditional on all the observed variables, and assessing whether it is discontinuous in the variable whose exogeneity requires testing. If a discontinuity is found, then the variable is endogenous.

The smoking example evidences the source of the test's power. The test has power when at least one unobservable confounder is discontinuously distributed with respect to the endogenous variable. This condition is discussed in more depth in Section 2.2, mentioning several examples in which such discontinuities can be found. Typically this phenomenon can be argued when the variable of interest has a bunching point. This is the case in the maternal smoking example, where more than 80% of the observations do not smoke. It should be noted, however, that bunching points in the variable of interest are not a necessary requirement for the applicability of the test.

The paper is organized as follows. Section 2 presents the test idea, as well as discussions of the test's maintained assumptions (Section 2.1), power (Section 2.2), and implementation (Section 2.3). An application of the test to the problem of the estimation of the effects of maternal smoking on birth weight can be seen in Section 3. That section discusses the applicability of the test (Section 3.1), introduces a test statistic that is particularly suited for this application and can be useful for practitioners interested in implementing the test (Section 3.2), and presents the test results (Section 3.3). The details about the test statistic, including the asymptotic results and small sample behavior in simulations, can be found in the Supplemental Material (Caetano (2015)). Finally, Section 4 concludes.

2. A DISCONTINUITY TEST OF EXOGENEITY

This section presents the assumptions under which the test is built as well as the test idea. The following subsections discuss the assumptions themselves (Section 2.1), when such a test would have power (Section 2.2), and how it can be implemented in practice (Section 2.3).

Suppose that

$$Y = g(X, Z) + U,$$

where Y and X are scalar observable variables, Z is a vector of observable variables, and U is scalar and unobservable. This model specifies U as separable from X and Z only for exposition purposes. All the arguments and proofs can be derived analogously when U is nonseparable and may be a vector.

If $\mathbb{E}[U|X, Z] = 0$, g is identifiable in the entire support of the joint distribution of X and Z . This is the usual "exogeneity" condition commonly re-

quired in the literature (see, e.g., [Blundell and Horowitz \(2007\)](#)).² However, this condition is often too strong because it imposes the same level of mean independence between all the components of the vector (X, Z) and the unobservable U .

In order to identify partial effects, for example, quantities such as $\partial g(x, z)/\partial x$, $\mathbb{E}[\partial g(X, Z)/\partial x]$, or $\mathbb{E}[\partial g(X, Z)/\partial x|Z = z]$, it is not necessary that all the components of the vector (X, Z) be exogenous. In fact, a weaker condition is sufficient:³

$$(1) \quad \mathbb{E}[U|X, Z] = \mathbb{E}[U|Z] \quad \text{with probability equal to 1.}$$

In the remainder of this paper, if condition (1) is satisfied, then X is said to be exogenous. Otherwise X is said to be endogenous. Notice that condition (1) is agnostic about the exogeneity of Z , and thus g may not be identifiable even if this condition is satisfied.

This paper claims that it is possible to build a test of the exogeneity of X by estimating a quantity which is equal to zero when X is exogenous and different from zero when X is endogenous in an important set of cases. In order to build this quantity, consider the following assumption.

ASSUMPTION 2.1: *Suppose that $[0, \delta) \subset \text{supp}(X)$ ⁴ for some $\delta > 0$, and that $(x, z) \mapsto g(x, z)$ is continuous at $x = 0$ uniformly for all z .*

Assumption 2.1's implications are discussed in Section 2.1. Observe that if the objective is the identification of the average partial effect of X , then Assumption 2.1 is redundant. For the partial effect to be well defined, $\text{supp}(X)$ must be connected near zero, and g must be differentiable in the first coordinate and thus continuous.

Let $\Delta(X, Z) := \mathbb{E}[Y|X = 0, Z] - \mathbb{E}[Y|X, Z]$. Then the quantity in which the test of exogeneity is based is

$$\tau(z) := \lim_{x \downarrow 0} \Delta(x, z),$$

which denotes the right discontinuity of $\mathbb{E}[Y|X = x, Z = z]$ as $x \downarrow 0$.

If Assumption 2.1 holds, then $\tau(z) = \mathbb{E}[U|X = 0, Z = z] - \lim_{x \downarrow 0} \mathbb{E}[U|X = x, Z = z]$. It is immediate to conclude that if X is exogenous, then $\tau(z) = 0$. This quantity is attractive as a basis for a test of the exogeneity of X , provided it is identifiable. The following condition guarantees the identifiability of $\tau(z)$.

²Exogeneity is always defined with respect to the model. Hence, since the model examined here is separable, exogeneity is defined as mean independence. If the model were nonseparable, exogeneity would be defined as actual independence, i.e., $U \perp\!\!\!\perp X, Z$. See [Blundell and Powell \(2003\)](#).

³Along with regularity conditions that allow the order of the derivative and the expectation to be exchanged.

⁴The notation $\text{supp}(W)$ denotes the support of the distribution of W .

ASSUMPTION 2.2: $\text{supp}(Z) \subseteq \text{supp}(Z|X = 0) \cap \text{supp}(Z|0 < X < \epsilon)$ for all $\epsilon > 0$.

The pair of hypotheses of interest is thus

$$\mathbb{H}_0 : X \text{ is exogenous} \quad \text{vs.} \quad \mathbb{H}_1 : X \text{ is endogenous}$$

under the maintained Assumptions 2.1 and 2.2. The previous arguments validate the following theorem.

THEOREM 1: *Suppose Assumptions 2.1 and 2.2 hold. Then $\tau(z)$ is identifiable for all $z \in \text{supp}(Z)$ and, under \mathbb{H}_0 , $\tau(z) = 0$.*

REMARK 2.1—Interpreting the Discontinuity: One must exercise caution when giving an economic interpretation to the magnitude of $\tau(z)$. It is an incomplete measure of the bias of endogeneity at zero. To see this, denote $\mathbb{E}[U|X = x, Z = z] = m_1(x, z) + m_2(z)\mathbf{1}(x = 0)$, where m_1 is continuous in x for all z (w.l.o.g.), then $\tau(z) = m_2(z)$, but the bias of endogeneity at $x = 0$ is given by $m_1(0, z) + m_2(z)$. Therefore, $\tau(z)$ does not account for the part of the bias that is caused by the continuous variation of the unobservables, $m_1(x, z)$, be it at zero or anywhere else.

In fact, because the endogeneity at zero has the two components m_1 and m_2 , it may often be the case that the magnitude of the bias of endogeneity is larger at zero than at other points. One cannot claim this formally, since m_1 may vary freely, but it could often be the case. In the maternal smoking example, women who smoke 10 cigarettes may have different unobservables from those who smoke 9, but it is unlikely that they would be as different as the comparison between the women who smoked one cigarette and those who did not smoke at all. The group of nonsmokers is much more heterogeneous, including women who are far from indifference.

REMARK 2.2—Relation to the RDD: The technical similarities with the Regression Discontinuity Design methods may give rise to questions about the relation between the RDD and this test. The methods are not formally related, but this test can be seen as an inversion of sorts of the RDD. Assumption 2.1 is also a requirement in the RDD (it is an indirect consequence of Assumption A1 in Hahn, Todd, and Van der Klaauw (2001)). However, Assumption 2.2 is the exact opposite (also as an indirect requirement of Assumption A1 in Hahn, Todd, and Van der Klaauw (2001)).

2.1. Maintained Assumptions

The first requirement of Assumption 2.1 is that the distribution of X has a connected support in a neighborhood of zero. This can be a strong constraint

for applications mainly due to a lack of refined data. For example, since it is possible to smoke any fraction of a cigarette, the variable “average cigarettes per day” is likely continuous. However, surveys on maternal smoking behavior measure the average cigarettes per day in integers, which may be too coarse a measure. The concern is that if the expected birth weight varies sharply as the smoked amount approaches zero, the data point nearest to zero, “one cigarette per day,” is in fact too distant to be used to predict the right-limit.

The continuity requirement in Assumption 2.1 is common to most of the applied econometrics literature (often as an indirect consequence of linearity). However, in this paper it plays a crucial role, and it cannot be overlooked. For example, in the smoking example this test could not be implemented if the effect of smoking on birth weight was discontinuous at zero cigarettes. See more details in Section 3, especially Remark 3.1.

One example where the continuity condition in Assumption 2.1 likely does not hold is for certain levels of schooling in a typical Mincer equation (e.g., Card (1999)), where X is education and Y is wages. Potential employers may use a high school degree as a proxy for ability and thus offer discontinuously higher wages to workers with a high school degree (this discontinuous treatment effect is often labeled as the “sheepskin effect”). It should be noted, however, that in principle the continuity condition can be argued for all other levels of schooling that are not susceptible to sheepskin effects.

Assumption 2.2 can be directly verified in the data. Put in terms of the smoking example, it requires that, for each value of the mother’s characteristics (e.g., white, high school educated, unmarried, etc.), there exist both nonsmoker mothers and mothers that smoke infinitesimal amounts. This could be restrictive, but the test can be easily modified to substantially relax this condition (see Remark 2.3 in Section 2.3). It should also be mentioned that Assumption 2.2 can be dropped if some semiparametric restrictions are imposed on $\mathbb{E}[Y|X, Z]$ (e.g., if $\mathbb{E}[Y|X, Z] = \psi(X) + Z'\gamma$).

2.2. Detectable Alternatives

Under \mathbb{H}_1 , $\mathbb{E}[U|X = x, Z]$ varies with x with positive probability. However, a test based on $\tau(z)$ has nontrivial power only when the following condition holds.

ASSUMPTION 2.3: *Under \mathbb{H}_1 , $\mathbb{E}[U|X = x, Z = z]$ is right-discontinuous as $x \rightarrow 0$.*

The set of cases where the discontinuity test can be applied is thus determined by Assumptions 2.1, 2.2, and 2.3. It must be noted that the endogeneity is detected locally. If X is locally exogenous in a neighborhood of zero ($\mathbb{E}[U|X = x, Z] = \mathbb{E}[U|Z]$, for $x < \delta$ with probability equal to 1), then the test will have no power.

Assumption 2.3 is novel and grants further discussion. It can often be argued when the potentially endogenous variable has bunching points. Some of the most common causes of bunching are natural or law restrictions that generate corner solutions. Examples include wages in the problem of the estimation of Engel curves (bunching at the minimum wage), schooling in the problem of the estimation of Mincer equations (bunching at the minimum attendance laws' thresholds), and weekly hours of work in the problem of the estimation of the effects of hours worked by the mother on the child's test scores (bunching at zero hours of work). Number of cigarettes (as in the application in Section 3) is another example, as it cannot take nonnegative values.

To illustrate how bunching due to a corner solution generates discontinuities in the unobservables, consider the maternal smoking example. It is helpful to think of U as a specific variable, say an aggregate measure of mother's quality. As the smoking level X varies, so does the distribution of the mother's quality. In general, it would be expected that incremental changes in smoking would generate incremental changes in the distribution of mother's quality, and thus $\mathbb{E}[U|X = x]$ would change continuously as $x \downarrow 0$. However, the group of mothers that do not smoke is different, because it includes all the mothers that would have chosen $X < 0$ if it was possible. This group is likely to have a disproportional amount of higher quality mothers, and therefore the expected quality of the mothers will be discontinuously higher in comparison with that of the mothers that smoked marginal amounts.

This is indeed the case, as shown in Section 3: every observable indicator of the mother's quality in the application data (e.g., mother's education and alcohol consumption) is discontinuous at zero cigarettes. The hope is that the vector Z can predict the entire part of the mother's quality that is related to smoking. The remaining quality should be unrelated to smoking. If that is not the case, and thus X is still endogenous, then the test will have power if the remaining quality is actually discontinuous at zero. The following example proposes a model that explains how the discontinuities arise when bunching is generated by a lower boundary constraint, as in the case of maternal smoking.

EXAMPLE 1: Suppose that smoking X is determined by the equation $X = \max\{0, Z'\pi - Q\}$, where Q is a continuously distributed unobservable variable. One can understand Q as an index of the mother's unobservable quality (the quality of the mother after accounting for the observables Z). Then

$$\tau(z) = \mathbb{E}[U|Q \geq z'\pi, Z = z] - \lim_{x \downarrow 0} \mathbb{E}[U|Q = z'\pi - x, Z = z].$$

The test has nontrivial power if, when X is endogenous, $\tau(z) \neq 0$. This is always true, for example, if $(Q, U)|Z = z \sim \mathcal{N}((0, 0), (1, \rho(z); \rho(z), 1))$. In this case, $\tau(z) = \rho(z)(\lambda(-z'\pi) - z'\pi)$, where $\lambda(\cdot)$ is the inverse Mills ratio. If X is endogenous, $\rho(z) \neq 0$, and thus $\tau(z) \neq 0$.

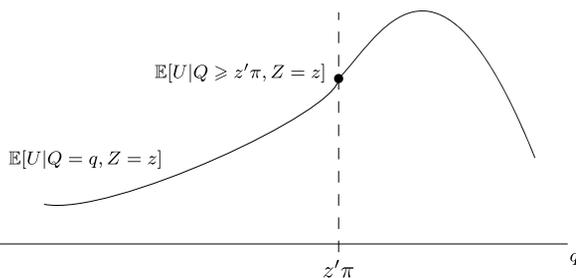


FIGURE 2.—Case (iii): pathological relationship between unobservables and mother's quality yields no testing power.

In general cases (not assuming normality), one can expect this test to have nontrivial power. However, there are three instances in which X may be endogenous and yet $\tau(z)$ may be equal to zero:

(i) (no censoring) If $\mathbb{P}(Q > z'\pi | Z = z) = 0$. In this instance, no mother has quality above $z'\pi$. Hence, the constraint that smoking cannot be negative is irrelevant, and there is effectively no censoring. In this case, $\tau(z) = \mathbb{E}[U|Q = z'\pi, Z = z] - \lim_{x \downarrow 0} \mathbb{E}[U|Q = z'\pi - x, Z = z]$. Hence, if $\mathbb{E}[U|Q = z'\pi - x, Z = z]$ is continuous in x at $x = 0$, then $\tau(z) = 0$ even if X is endogenous.

(ii) (local exogeneity) If $\mathbb{E}[U|Q = q, Z = z] = \mathbb{E}[U|Z = z]$ for all $q > z'\pi - \delta$ for some $\delta > 0$. In this instance, the mother's quality is not a confounder for all mothers with quality above $z'\pi - \delta$, and therefore X is locally exogenous in $[0, \delta)$. Hence, for x small enough, $\tau(z) = \mathbb{E}[U|Z = z] - \lim_{x \downarrow 0} \mathbb{E}[U|Z = z] = 0$ even if X is endogenous.

(iii) (pathological functional shape) If the shape of $\mathbb{E}[U|Q = q, Z = z]$ for $q > z'\pi$ is such that, although it varies with q , incidentally $\mathbb{E}[U|Q \geq z'\pi, Z = z] = \lim_{x \downarrow 0} \mathbb{E}[U|Q = z'\pi - x, Z = z]$, and thus $\tau(z) = 0$. This is a rather odd situation, which requires that the shape of the relationship between the mother's quality and birthweight be of a peculiar form. Figure 2 illustrates one such situation when $\mathbb{E}[U|Q = q, Z = z]$ is continuous in q . Notice that for this phenomenon to happen, $\mathbb{E}[U|Q = q, Z = z]$ cannot be monotonic when $q > z'\pi$. This means that for certain higher levels of the mother's quality, more quality would have a decreasing effect on birth weight.

Other examples of bunching can be found in the empirical literature in several topics. For instance, Madrian and Shea's (2001) study on 401(k) savings showed that people tend to bunch at default contract levels, and hence the test might be applied, for example, to the problem of estimating the effects of savings for retirement on other savings. Similarly, Saez (2010) showed that taxpayers bunch at kink points of the U.S. income tax schedule, and hence the test might be applied, for example, to the problem of estimating the effect of taxable income on tax deductions or to the problem of estimating Engel curves.

In both situations, to the extent that bunching occurs along any unobserved variable, that unobserved variable is likely to be discontinuous at the bunching point. If at least one such unobserved variable is not absorbed by the controls, then the test will have power.

Discontinuities in the distribution of the unobservables can also be found for reasons other than bunching. For example, suppose that one is studying the effects of hours worked per week on the amount of time spent weekly on childcare. One should be concerned with potential endogeneity due to variables such as, for example, stress or career ambition. Fortunately, there is a phenomenon that can be explored, which is the fact that workers with a minimum fixed workload, such as the typical “9 to 5” worker, are discontinuously more likely to report working 40 hours or more. These kinds of occupations are a very particular subset of all occupations, and therefore one should expect to see discontinuities at 40 in most variables which are related to the choice of occupation. For example, Figures 3(a) and 3(b), which use the March Current Population Survey (CPS) data from 2000 to 2009, show that workers who report working more than 40 hours per week are discontinuously more likely to be male and to work for the federal government compared with workers who report working strictly less than 40 hours. Analogously, the same patterns may be expected of unobservable variables related to the choice of occupation, such as stress and career ambition. These discontinuities are not driven by bunching (although it is true that there exists bunching of 25% of the labor force at 40 hours). They are driven by the sharp change in the distribution of industry, profession, and occupation at 40 hours.

This example is particularly instructive. Notice that the averages at the multiples of 5 (e.g., 20, 25, 30, 35, etc.) seem to follow a different process in comparison with the other points. There is an overrepresentation of males and federal employees among workers who report working multiples of 5 in comparison

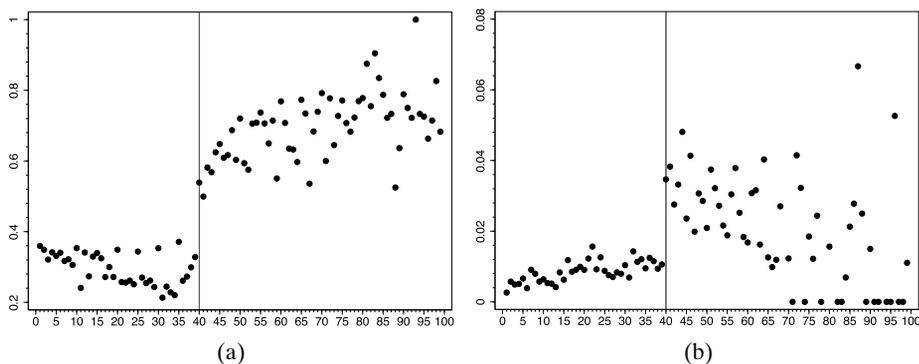


FIGURE 3.—Weekly hours of work: (a) percentage of males, (b) percentage of federal employees. Dots represent percentages of workers at each level of hours worked per week. Source: CPS 2000–2009.

with workers who report other workloads. This may be due to individuals' tendency to round, for example if males are more likely to round than females. More likely, this phenomenon is caused by the type of occupation. Some occupations require a fixed amount of hours every day of the week, which add up to a weekly workload which is a multiple of 5. Also, some occupations do not require that the worker report how many hours they work (e.g., artists, writers, journalists, professors, CEOs, many types of self-employment). When asked their weekly workload, such workers would have an imprecise recall and thus tend to round. Analogously to the figures above, there may be an overrepresentation at the multiples of 5 of all the characteristics that are more likely to be encountered among workers that have such occupations, including unobservable characteristics such as stress and career ambition. One could then test for the presence of such unobservable confounders after controlling for observables not only at 40 hours, but also at all the multiples of 5.

2.3. Implementing the Test

A natural way to exploit the result from Theorem 1 is to estimate $\tau(z)$ and to reject the null hypothesis when the realization of $\hat{\tau}(z)$ is "too large" in some suitable norm. For example, if $\mathbb{P}(X = 0) > 0$, $\mathbb{E}[Y|X = 0, Z = z]$ can be estimated with a local linear regression of Y onto Z at z using only observations such that $X = 0$, and $\lim_{x \downarrow 0} \mathbb{E}[Y|X = x, Z = z]$ can be estimated with a local linear regression of Y onto $X = 0$ and $Z = z$ using only observations such that $X > 0$. A good explanation of this approach for practitioners can be found on [Imbens and Lemieux \(2008, p. 625\)](#).

From a technical point of view, this approach presents no difficulties. Given the standard assumptions for the convergence of the multivariate local linear estimator (see [Masry \(1996\)](#)), it is immediate to show that the $\lim_{x \downarrow 0} \mathbb{E}[Y|X = x, Z = z]$ term will dominate the asymptotic variance because it has one more estimation dimension and thus converges more slowly. Therefore, under \mathbb{H}_0 , $\sqrt{nh^{d+1}}\hat{\tau}(z) \xrightarrow{d} \mathcal{N}(0, V)$, where V is the asymptotic variance of the estimator of $\lim_{x \downarrow 0} \mathbb{E}[Y|X = x, Z = z]$ as given in [Fan and Gijbels \(1992, p. 2011\)](#).

In practice, it is worthwhile to aggregate the discontinuities $\tau(z)$ for several values of z . The aggregation can increase the rate of convergence of the test statistic considerably (often up to that of a univariate nonparametric regression) and thus significantly increase the power of the test. Possible aggregations include $\mathbb{E}[\tau(Z)]$, $\mathbb{E}[|\tau(Z)|]$, and $\mathbb{E}[\tau(Z)^2]$. The last two have the desirable property that if the discontinuities have different signs for different values of Z , no power is lost because the positive discontinuities cancel out the negative ones. An alternative possibility is to build a test based on the quantity $\sup_{z \in \text{supp}(Z)} |\tau(z)|$.

Section 3 implements the test using an aggregation strategy that is different from the ones suggested above. Though it is not necessarily the most powerful alternative in general, it is well adjusted to the particular features of the data,

and it is remarkably simple to implement on standard packaged software. Details of this approach can be found in Section 3 and in the Supplemental Material.

REMARK 2.3—Relaxing the Support Assumption 2.2: If Assumption 2.2 is too restrictive in the particular application, the aggregation can be modified to include only the discontinuities over the set $\mathcal{A} = \bigcap_{\epsilon > 0} \{\text{supp}(Z|X = 0) \cap \text{supp}(Z|0 < X < \epsilon)\}$, for example $\mathbb{E}[\tau(Z)^2|Z \in \mathcal{A}]$, provided $\mathbb{P}(Z \in \mathcal{A}) > 0$.

3. AN APPLICATION TO THE EFFECTS OF MATERNAL SMOKING ON BIRTH WEIGHT

Investments before a person is born can substantially affect the person's adult life, with regard to both health (e.g., [Barker \(1990\)](#)) and educational and labor market outcomes (e.g., [Currie and Hyson \(1999\)](#), [Currie and Moretti \(2007\)](#), [Black, Devereux, and Salvanes \(2007\)](#), [Oreopoulos, Stabile, Walld, and Roos \(2008\)](#), and [Royer \(2009\)](#)). For example, [Black, Devereux, and Salvanes \(2007\)](#) found that a 10% increase in birth weight increases high school graduation by 1.2%, IQ (of men) by 1.2%, earnings by 0.9%, and height by 0.3% (see [Almond and Currie \(2011\)](#) for a recent survey on the topic).

Smoking while pregnant is viewed as the leading modifiable cause of low birth weight in the United States ([Almond, Chay, and Lee \(2005\)](#)). There is a very large literature of observational studies on the subject, of which [Almond, Chay, and Lee \(2005\)](#) seems to be the most exhaustive approach. The rest of this section applies the exogeneity test to the model in Section 2, where X represents the average cigarettes the mother smoked per day during pregnancy, Y is the weight of the child at birth, and Z is the entire vector of covariates used in the full specification in [Almond, Chay, and Lee \(2005\)](#) (the variables list can be seen in their footnote 36, p. 1064) and using the same data set as in that paper.

3.1. *Applicability of the Test*

The first concern for the applicability of the exogeneity test to this problem is whether the causal effect of smoking on birth weight is continuous at zero. An observationally equivalent concern is that the observed smoking variable may not be continuous enough. The average number of cigarettes smoked per day is measured (at best) in integers. The fear is that smoking 1 cigarette may have a very steep effect on birth weight in comparison with subsequent cigarettes so that what is perceived as a discontinuity is in fact a very large derivative. Though this is indeed a possibility, there is currently no data set that can verify whether this is the case. In any event, the results are so pronounced that the qualitative results of the test are likely to hold anyway. See Remark 3.1 for more details.

The second concern is whether Assumption 2.3 is satisfied. Figures 4(a) to 4(f), 5(a), and 5(b) provide heuristic evidence of discontinuities at $X = 0$ for several variables in the data set. The figures were cropped at $X = 40$, which

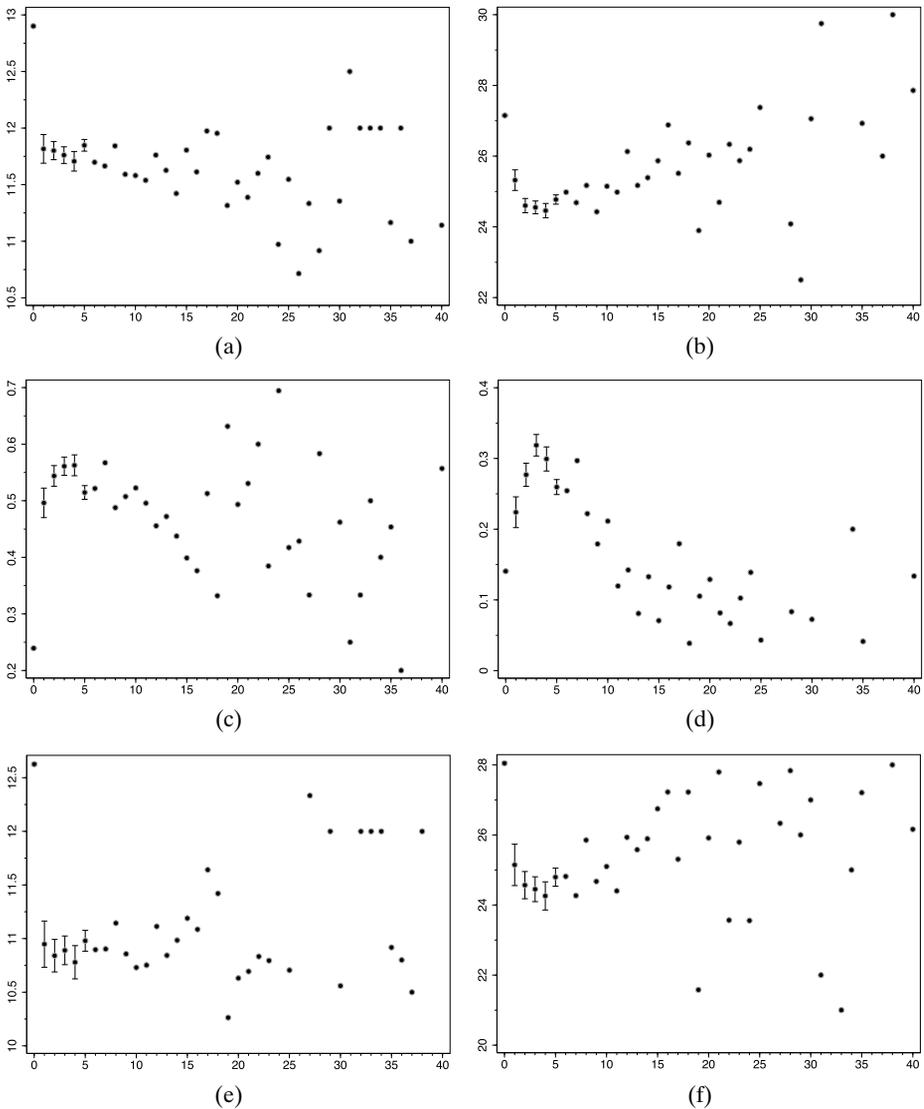


FIGURE 4.—(a)–(d) Mother’s demographic characteristics: (a) education (years), (b) age, (c) unmarried, (d) race: black. (e)–(f) Father’s demographic characteristics: (e) education (years), (f) age. Dots represent average values among pregnant mothers for each level of daily cigarette consumption. The vertical lines represent the 95% confidence interval of the mean.

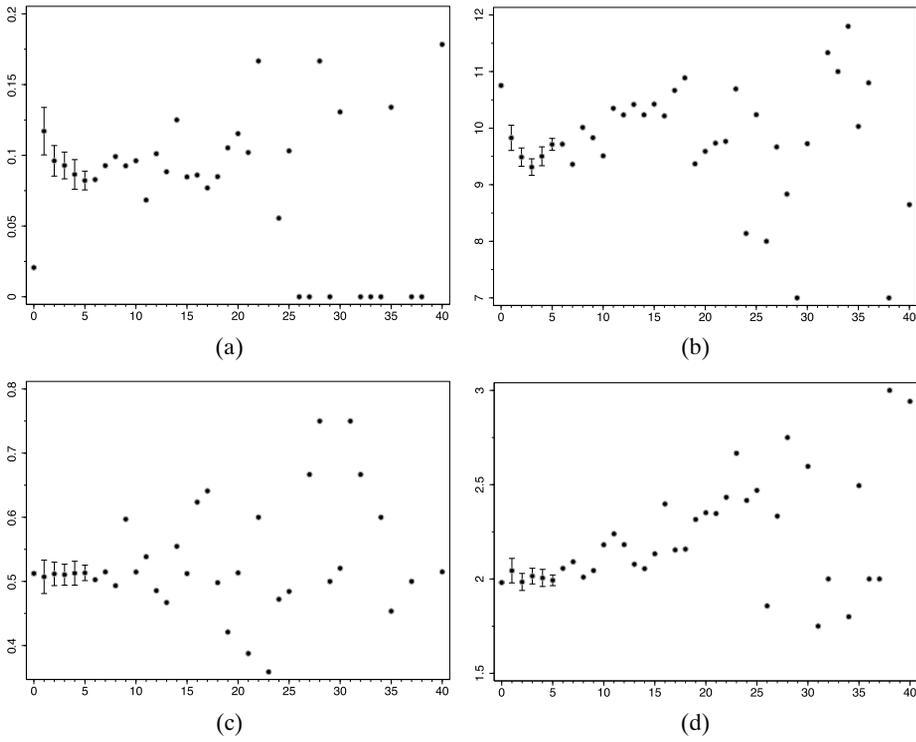


FIGURE 5.—Mother’s behavior variables: (a) consumed alcohol, (b) # of prenatal visits, (c) gender of newborn, (d) order of newborn (among live births). Dots represent average values among pregnant mothers for each level of daily cigarette consumption. The vertical lines represent the 95% confidence interval of the mean.

represents 99.95% of the entire population. The dots are estimates of $\mathbb{E}[Z|X]$, and the lines show the 95% confidence interval of the mean for some values of X . The confidence intervals become much larger for $X > 20$, due to small samples. Similar discontinuities can be found for most of the confounders in the data, and the hope is that the same patterns would also be found with the unobservable confounders. There are a few exceptions, as seen in Figures 5(c) and 5(d), where no discontinuity is found. However, it is not clear that these variables are confounders at all.

The discontinuities are not only present in the first moment of the distribution. To illustrate this point, Figures 6(a) to 6(d) show the (kernel density) distribution of a few variables for each level of cigarettes, from $X = 0$ to $X = 3$. These figures provide evidence that the distribution for $X = 0$ is not a simple mean shift of the distribution for X positive. Rather, the variance and often the shape change as well.

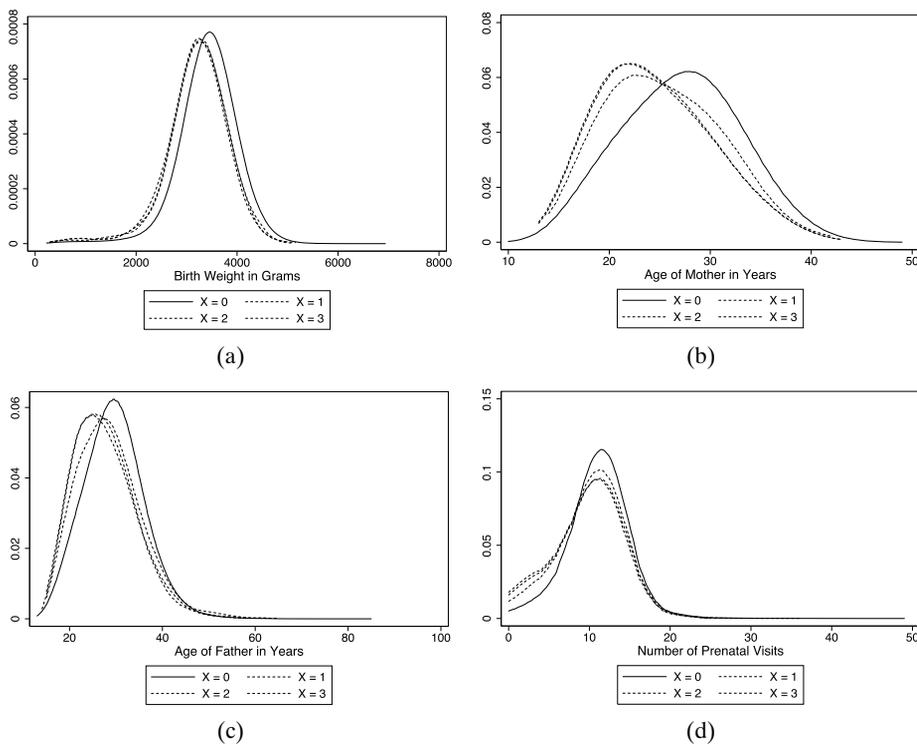


FIGURE 6.—Discontinuity in distributions: (a) birth weight, (b) mother’s age, (c) father’s age, (d) # of prenatal visits. Kernel density of the distribution conditional on the amount smoked X .

3.2. Building the Test Statistic

As discussed in Section 2.3, because of concerns of severe loss of efficiency due to high dimensionality, it is recommendable to aggregate the discontinuities. This application will use the following aggregation:

$$\theta = \lim_{x \downarrow 0} \mathbb{E}[\Delta(x, Z)|X = x] = \lim_{x \downarrow 0} \mathbb{E}[\mathbb{E}[Y|X = 0, Z] - Y|X = x].$$

The advantages of this particular aggregation over other (perhaps more efficient) choices discussed in Section 2.3 are practical. With over 30 covariates (more than 300 covariates if one counts all the interactions used in the full specification of Almond, Chay, and Lee (2005)), the multivariate nonparametric boundary term $\lim_{x \downarrow 0} \mathbb{E}[Y|X = x, Z]$ cannot be estimated with a multivariate local linear estimator using reasonably small bandwidths. However, because of the law of iterated expectations, the aggregation in θ eliminates this term. The remaining multivariate nonparametric term $\mathbb{E}[Y|X = 0, Z]$ is not a boundary quantity and thus can be estimated with other methods such

as series, or flexible OLS. Moreover, since 80% of the sample does not smoke, the estimation of $\mathbb{E}[Y|X = 0, Z]$ counts with over 400,000 observations. The only boundary quantity that has to be estimated is the limit of the outer expectation, which can be done with a univariate local linear regression without any practical difficulty.

The immediate concern with the aggregation in θ is that there could be a loss of power if the discontinuities $\tau(z)$ change signs for different values of z . In this application, this would become a problem if the mothers that do not smoke are often selected worse than the mothers that smoke marginal amounts. This seems unlikely. The practical advantages of this aggregation in this application far outweigh the risk of some loss of power due to this possibility.

Additionally, θ has a direct intuitive interpretation in relation to the model. If smoking is exogenous, then the observable mother characteristics Z are enough to account for the selection to different levels of smoking treatment. In this case, the mothers that did not smoke can be used to calculate the counterfactual effect if the mothers that smoked marginal amounts stopped smoking altogether. If the effect of smoking is continuous, there should be no effect, since the treatment difference is only marginal across both groups. A difference can be found only if the nonsmoking mothers cannot serve as counterfactual for the mothers that smoked marginal amounts, that is, they have systematic unobservable differences that affect birth weight.

The test is thus based on θ , which is estimated in two steps. The first step assumes that $\mathbb{E}[Y|X = 0, Z] = Z'\gamma$, where Z is the full covariate specification in Almond, Chay, and Lee (2005), and estimates γ with an OLS regression of Y onto Z using only observations such that $X = 0$. The second step is a local linear regression of $Z'\hat{\gamma} - Y$ onto X at $X = 0$ using only observations such that $X > 0$. The results in the next section use the Epanechnikov kernel.⁵ The test statistic is $T_n = \hat{\theta} / \sqrt{\hat{\Omega}/nh}$, where $\hat{\Omega}$ is the estimator of the asymptotic variance of $\sqrt{nh}(\hat{\theta} - \theta)$. T_n should be compared to the percentiles of the standard normal distribution.

The Supplemental Material provides more details about the estimators $\hat{\theta}$ and $\hat{\Omega}$. It also contains the results that establish the asymptotic behavior of T_n , as well as real-data Monte Carlo simulations that showcase its small sample behavior (although the sample in the application is anything but small). The test can be programmed directly in any packaged software as two regressions. A Stata code can be obtained in the author's website.

⁵Rectangular and triangular kernels, as well as local polynomial regressions of degrees 2 and 3 in place of the local linear estimator were also used with very similar results. The Monte Carlo study in the Supplemental Material suggests that the local linear estimator is more stable. This is likely a consequence of the fact that the estimation is done at the boundary, where Runge's phenomenon causes large variability in the results for the higher order polynomials when larger bandwidths are used.

3.3. Test Results

Table I shows the discontinuity test results for two outcome variables commonly studied in the medical literature. The first is the baby's birth weight in grams. The second is the probability of low birth weight (LBW) expressed in percent likelihood, which is defined as birth weight below 2500 grams. Specification I estimates the discontinuity in the outcome conditional on the amount smoked, without covariates, that is, $\lim_{x \downarrow 0} \mathbb{E}[\mathbb{E}[Y|X=0] - Y|X=x]$. For example, the values in the top row in Table I can be interpreted as the expected difference in birth weight between a nonsmoking mother and a mother who smokes a marginally positive amount. Of course, this difference can be explained by selection on observables. Specification II includes all the covariates. Hence, the numbers in the first row of Specification II in Table I are the values of $\hat{\theta}$.

The columns in Table I show the estimates for several bandwidths used in the second step of the estimator. For the smallest bandwidth, $h = 3$, the results for birth weight show that nonsmoking women have babies that are 203 grams heavier than those that smoke a marginally positive amount. This difference is at least partially explained by selection on observables. Indeed, the results of Specification II show that 66 grams are explained by the covariates. However, 137 grams remain unexplained and are thus evidence of endogeneity. The results show that the discontinuity is notably larger for $h = 3$ in comparison with the other bandwidths. However, $\hat{\theta}$ becomes even more significant as the bandwidth increases, and thus the qualitative point stands under any bandwidth.

TABLE I
TEST RESULTS^a

Variable	Specification	Bandwidth					
		3 5%	5 12%	7 22%	9 25%	11 60%	13 61%
Birth Weight	I	203 (34)	178 (18)	184 (12)	186 (11)	202 (7)	204 (7)
	II	137 (30)	99 (16)	75 (11)	74 (10)	85 (6)	87 (6)
LBW (%)	I	-4.7 (1.6)	-3.9 (0.9)	-4.3 (0.6)	-4.3 (0.5)	-4.4 (0.4)	4.4 (0.3)
	II	-2.8 (1.6)	-1.9 (0.8)	-1.5 (0.6)	-1.4 (0.5)	-1.5 (0.3)	-1.5 (0.3)

^aTable shows the values of $\hat{\theta}$ and the standard errors $\sqrt{\hat{\Omega}/nh}$ (between parentheses) for birth weight and the percent likelihood of low birth weight. Specification I uses no controls, and Specification II uses the most complete control specification in Almond, Chay, and Lee (2005). The numbers under the bandwidths represent the proportion of the smoking mothers weighted positively by the kernel (out of 93,263 observations).

The results for the probability of low birth weight show less evidence of endogeneity. The test was significant for all but the smallest bandwidth, but the test statistic was always considerably smaller than in the birth weight case. Since LBW is a threshold variable, its determination has to be equal to or less complex (in terms of the number of possible confounders) than the actual birth weight level. Thus, it is expected that the endogeneity will be the same or lower than for the birth weight process.

Figures 7(a) and 7(b) depict the main results from Table I. Intuitively, the difference between the solid dot representing $\hat{\mathbb{E}}[Y|X = 0]$ and the hollow dot cannot be due to smoking. However, it can be due to selection on observables. The \times point depicts the estimated expected birth weight of a nonsmoker with the same covariates as a woman that smokes just a marginally positive amount. The difference between this point and the solid dot $\hat{\mathbb{E}}[Y|X = 0]$ is due entirely to differences in the covariates of nonsmokers and those that smoke a marginally positive amount. This part of the discontinuity is due to selection on observables. The distance between the \times point and the hollow dot is the expected difference in birth weight between a nonsmoking mother who has the same covariates as a mother that smoked a marginally positive amount and a mother who did smoke a marginally positive amount. Since both mothers have the same covariates, the difference can only be explained by selection on unobservables.

Figure 7(b) has the same interpretation. In the case of the probability of low birth weight, selection on observables explains a comparatively larger proportion of the discontinuity. The part of the difference that is due to endogeneity (the difference between the \times point and the hollow dot) is significant for most choices

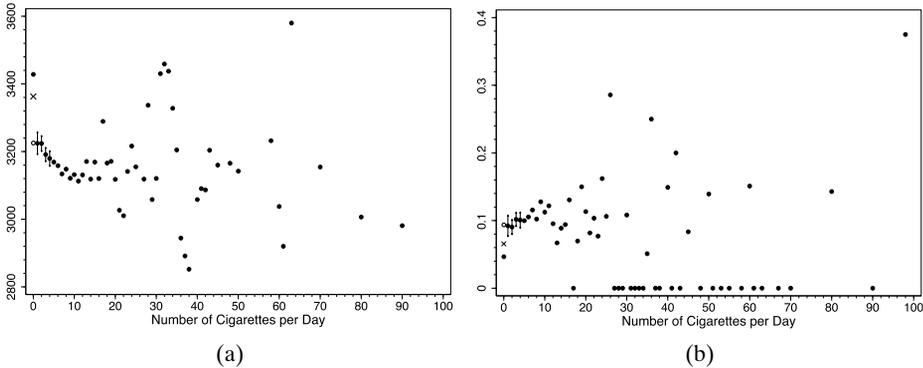


FIGURE 7.—(a) Birth weight, (b) probability of low birth weight. Dots represent average values among pregnant mothers for each level of daily cigarette consumption. The vertical lines represent the 95% confidence interval of the mean. High variability for higher amounts of smoking is just small sample variance. The hollow dot represents the estimate of $\lim_{x \downarrow 0} \mathbb{E}[Y|X = x]$. The “ \times ” point represents the estimate of $\lim_{x \downarrow 0} \mathbb{E}[\mathbb{E}[Y|X = 0, Z]|X = x]$. Their difference is $\hat{\theta}$. (Bandwidth is $h = 3$.)

of bandwidth, albeit the test statistic is considerably smaller than in the birth weight case.

REMARK 3.1: A natural concern with this application is that the causal effect of smoking could be discontinuous at zero. An observationally equivalent concern is that the observed smoking variable may be too coarse and smoking 1 cigarette may have a very steep effect on birth weight in comparison with subsequent cigarettes, so that what is perceived as a discontinuity is in fact a very large derivative. To entertain this possibility, suppose that the model is

$$Y = g(X) + Z'\gamma + U,$$

where $\mathbb{E}[U|X, Z] = 0$, $g(0) = 0$ (w.l.o.g.) and g may be discontinuous at zero. The effect of a marginal amount of smoking is thus $\lim_{x \downarrow 0} g(x) - g(0)$, and the next x cigarettes will have a combined effect of $g(x) - \lim_{x \downarrow 0} g(x)$.

This is a selection on observables model that attributes the discontinuity at zero to the causal effect of smoking. This model is identifiable. γ can be estimated by regressing Y onto Z using only observations such that $X = 0$ (because $\mathbb{E}(Y|X = 0, Z) = Z'\gamma$). Then, since $\mathbb{E}[Y - Z'\hat{\gamma}|X = x] = g(x)$, g is estimated by regressing $Y - Z'\hat{\gamma}$ onto X using a local linear estimator with $h = 3$ and the Epanechnikov kernel.

The estimates yield $\lim_{x \downarrow 0} g(x) - g(0) = -137$ grams, and $g(10) - \lim_{x \downarrow 0} g(x) = -67$ grams. This means that, if this model is correct, smoking just a minimal amount would have more than two times the effect of the next 10 cigarettes. It seems reasonable to presume instead that even if the smoking effects are discontinuous or very steep for small smoking amounts, at least part of the estimated effect is due to endogeneity.

REMARK 3.2: There could be a concern with measurement error on the cigarettes variable, namely that mothers can misreport how much they smoke. Indeed, in the data there is a large concentration of observations at 5 and 15 cigarettes, and a much higher concentration at 10 and 20 cigarettes. However, medical research that relates self-reported smoking levels and smoking biomarkers such as urine and blood cotinine levels shows that the measurement error is much smaller than intuition suggests. The most recent studies with the best sets of controls conclude that less than 1% of the population misreport the amount smoked (e.g., Yeager and Krosnick (2010), Pickett, Rathouz, Kasza, Wakschlag, and Wright (2005), and Caraballo, Giovino, Pechacek, and Mowery (2001)).

Even if it is true that there is measurement error in the reporting of cigarettes, the test detects endogeneity of any kind, including the kind generated by measurement error. Thus, although measurement error will affect the power of the test, it will not invalidate it. In fact, endogeneity combined with measurement error may yield a more powerful test, for example if the mothers that underreport are better selected than the average nonsmoker.

4. CONCLUSION

This paper develops a test of the exogeneity of a single variable, which does not require instrumental variables. In fact, it does not require the identifiability of the structural function even under the null hypothesis. Thus, this test can be used to validate a selection on observables approach.

The test depends on the continuity of the causal effect of the explanatory variable of interest. The power of the test derives from the discontinuity of the distribution of the unobservable confounders at a known point in the domain of the variable of interest. When the variable is endogenous, the discontinuities in the unobservables generate discontinuities in the expectation of the outcome variable conditional on the observables. The test consists of estimating such discontinuities, averaging them over a particular distribution of the covariates, and then testing for whether this average is equal to zero.

This paper also presents an application of the test to the problem of the estimation of the effect of maternal smoking on birth weight. A particular aggregation is suggested in the context of this application which has great practical advantages. The results suggest that there is substantial endogeneity remaining for the birth weight outcome, even after controlling for the most complete covariate specification in the literature. The evidence of endogeneity is considerably weaker for the probability of low birth weight outcome.

REFERENCES

- ALMOND, D., AND J. CURRIE (2011): "Killing Me Softly: The Fetal Origins Hypothesis," *Journal of Economic Perspectives*, 25 (3), 153–172. [1591]
- ALMOND, D., K. Y. CHAY, AND D. S. LEE (2005): "The Costs of Low Birth Weight," *Quarterly Journal of Economics*, 120 (3), 1031–1083. [1591,1594-1596]
- BARKER, D. J. (1990): "The Fetal and Infant Origins of Adult Disease," *BMJ: British Medical Journal*, 301 (6761), 1111. [1591]
- BIERENS, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58 (6), 1443–1458. [1581]
- BIERENS, H. J., AND W. PLOBERGER (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65 (5), 1129–1151. [1581]
- BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2007): "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," *Quarterly Journal of Economics*, 122 (1), 409–439. [1591]
- BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics: Theory and Applications. Eighth World Congress*, Vol. II, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky. Cambridge, U.K.: Cambridge University Press, 312–357. [1584]
- BLUNDELL, R. W., AND J. L. HOROWITZ (2007): "A Non-Parametric Test of Exogeneity," *Review of Economic Studies*, 74 (4), 1035–1058. [1581,1584]
- CAETANO, C. (2015): "Supplement to 'A Test of Exogeneity Without Instrumental Variables in Models With Bunching'," *Econometrica Supplemental Material*, 83, <http://dx.doi.org/10.3982/ECTA11231>. [1583]
- CARABALLO, R. S., G. A. GIOVINO, T. F. PECHACEK, AND P. D. MOWERY (2001): "Factors Associated With Discrepancies Between Self-Reports on Cigarette Smoking and Measured Serum Cotinine Levels Among Persons Aged 17 Years or Older: Third National Health and Nutrition Examination Survey, 1988–1994," *American Journal of Epidemiology*, 153 (8), 807–814. [1598]

- CARD, D. E. (1999): "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, Vol. 3. Amsterdam: Elsevier, 1801–1863. [1586]
- CURRIE, J., AND R. HYSON (1999): "Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birthweight," *American Economic Review*, 89 (2), 245–250. [1591]
- CURRIE, J., AND E. MORETTI (2007): "Biology as Destiny? Short- and Long-Run Determinants of Intergenerational Transmission of Birth Weight," *Journal of Labor Economics*, 25 (2), 231–264. [1591]
- FAN, J., AND I. GIJBELS (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, 20 (4), 2008–2036. [1590]
- GOURIEROUX, C., A. MONFORT, E. RENAULT, AND A. TROGNON (1987): "Generalised Residuals," *Journal of Econometrics*, 34 (1–2), 5–32. [1581]
- HAHN, J., P. TODD, AND W. VAN DER KLAUW (2001): "Identification and Estimation of Treatment Effects With a Regression-Discontinuity Design," *Econometrica*, 69 (1), 201–209. [1585]
- HAUSMAN, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46 (6), 1251–1271. [1581]
- IMBENS, G. W., AND T. LEMIEUX (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142 (2), 615–635. [1590]
- MADRIAN, B. C., AND D. F. SHEA (2001): "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116 (4), 1149–1187. [1588]
- MASRY, E. (1996): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17 (6), 571–599. [1590]
- OREOPOULOS, P., M. STABILE, R. WALLD, AND L. L. ROOS (2008): "Short-, Medium-, and Long-Term Consequences of Poor Infant Health: An Analysis Using Siblings and Twins," *Journal of Human Resources*, 43 (1), 88–138. [1591]
- PICKETT, K. E., P. J. RATHOUZ, K. KASZA, L. S. WAKSCHLAG, AND R. WRIGHT (2005): "Self-Reported Smoking, Cotinine Levels, and Patterns of Smoking in Pregnancy," *Paediatric and Perinatal Epidemiology*, 19 (5), 368–376. [1598]
- ROYER, H. (2009): "Separated at Girth: US Twin Estimates of the Effects of Birth Weight," *American Economic Journal: Applied Economics*, 1 (1), 49–85. [1591]
- SAEZ, E. (2010): "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2 (3), 180–212. [1588]
- YEAGER, D. S., AND J. A. KROSNICK (2010): "The Validity of Self-Reported Nicotine Product Use in the 2001–2008 National Health and Nutrition Examination Survey," *Medical Care*, 48 (12), 1128–1132. [1598]

Dept. of Economics, University of Rochester, 280 Hutchison Road, Box 270156, Rochester, NY 14627, U.S.A.; carol.caetano@rochester.edu.

Manuscript received November, 2012; final revision received February, 2015.