Testing for causal effects in a generalized regression model with endogenous regressors[*]

Jason Abrevaya[†]

Jerry A. Hausman[‡]

Shakeeb Khan[§]

ABSTRACT

A unifying framework to test for causal effects in non-linear models is proposed. We consider a generalized linear-index regression model with endogenous regressors and no parametric assumptions on the error disturbances. To test the significance of the effect of an endogenous regressor, we propose a test statistic that is a kernel-weighted version of the rank correlation statistic (*tau*) of Kendall (1938). The semiparametric model encompasses previous cases considered in the literature (continuous endogenous regressors (Blundell and Powell (2003)) and a single binary endogenous regressor (Vytlacil and Yildiz (2007)), but the testing approach is the first to allow for (i) multiple discrete endogenous regressors, (ii) endogenous regressors that are neither discrete nor continuous (e.g., a censored variable), and (iii) an arbitrary "mix" of endogenous regressors (e.g., one binary regressor and one continuous regressor).

JEL Classification: C14, C25, C13.

Key Words: Endogeneity, causal effects, semiparametric estimation.

[†]Purdue University, Dept. of Economics, `abrevaya@purdue.edu`

[‡]Massachusetts Institute of Technology, Dept. of Economics, `jhausman@mit.edu`

[§]Duke University, Dept. of Economics, `shakeebk@econ.duke.edu`

1   Introduction

Endogenous regressors are frequently encountered in econometric models, and failure to correct for endogeneity can result in incorrect inference. With the availability of appropriate instruments, two-stage least squares (2SLS) yields consistent estimates in linear models without the need for making parametric assumptions on the error disturbances. Unfortunately, it is not theoretically appropriate to apply 2SLS to non-linear models, as the consistency of 2SLS depends critically upon the orthogonality conditions that arise in the linear-regression context.

Until recently, the standard approach for handling endogeneity in non-linear models has required parametric specification of the error disturbances (see, e.g., Smith and Blundell (1986), Rivers and Vuong (1988), or the treatment in Wooldridge (2002, Section 15.7)). A more recent literature in econometrics has developed methods that do not require parametric assumptions, which is more in line with the 2SLS approach in linear models. In the context of the model considered in this paper, existing approaches depend critically upon the form of the endogenous regressor(s).[1]

For continuous endogenous regressors, a "control-function approach" has been proposed by Blundell and Powell (2003, 2004) (see also Aradillas-Lopez, Honoré, and Powell (2005)). A linear model specifies a relationship between the continuous endogenous regressors and the full set of exogenous covariates (including the instruments). The first-stage estimation yields estimates of the residuals from this model, which are then plugged into a second-stage estimation procedure to appropriately "control" for the endogenous regressors.[2] The control-function approach, however, does not work if any of the endogenous regressors are non-continuous.

For a single binary endogenous regressor, Vytlacil and Yildiz (2007) establish conditions under which it is possible to identify the average treatment effect in non-linear models. Identification requires variation in exogenous regressors (including the instruments for the binary endogenous regressor) that has the same effect upon the outcome variable as a change in the binary endogenous regressor. Yildiz (2006) implements this identification strategy in

---

[1]Several papers have considered estimation in the presence of endogeneity under additional assumptions. These include Lewbel (2000), Hong and Tamer (2003), and Kan and Kao (2005).

[2]This idea was originally considered by Hausman (1978) in the linear-model context, where inclusion of residuals in the regression equation controls for endogeneity.

the context of a linear-index binary-choice model, where the outcome equation is

$$y_1 = 1(z_1 \beta_0 + \alpha_0 y_2 + \epsilon > 0)$$

for exogenous regressors z_1, a binary endogenous regressor y_2, and i.i.d. error disturbance ϵ. The reduced-form equation for y_2 is

$$y_2 = 1(z\delta_0 + \eta > 0)$$

for exogenous regressors z (which now includes instruments for y_2) and i.i.d. error disturbance η. Identification requires an extra support condition, specifically that for some $z\delta_0$ values (i.e., a positive-probability region), the conditional distribution $z_1 \beta_0$ has support wider than the parameter value α_0.[3]

In this paper, we consider the problem of testing the statistical significance of causal (or treatment) effects in a general non-linear setting. That is, rather than attempting to estimate the magnitude of causal effects, we seek to estimate the direction (or sign) of these effects. The focus upon the sign(s) of causal effects rather than the magnitude(s) turns out to have important implications for the generality of our proposed testing procedure. First, the testing procedure can handle endogenous regressors of arbitrary form, including continuous regressors as in Blundell and Powell (2003), a binary regressor as in Vytlacil and Yildiz (2007), or other types of regressors (e.g., a censored variable). Second, the approach extends easily to the case of multiple endogenous regressors; importantly, the set of endogenous regressors can include a "mix" of discrete and continuous variables. Third, the procedure can test the statistical significance of a causal effect even in cases in which the magnitude of the causal effect is not identified. For example, the extra support condition in Vytlacil and Yildiz (2007) and Yildiz (2006) is not required to identify the sign of the treatment effect and, therefore, is not needed for our testing procedure.

Even in situations in which it is possible to identify the average causal effect of an endogenous variable, the proposed test can be used as a preliminary stage of inference in empirical applications. The rank-based approach that we adopt (for both preliminary plug-in estimates and the test statistic itself) requires fewer smoothing and trimming parameters than either Blundell and Powell (2003) or Yildiz (2006). Moreover, in the binary case, due to the identification strategy required for estimation of the treatment-effect magnitude, it may be possible to find evidence of a statistically significant treatment effect even when it is difficult to precisely estimate the magnitude of the treatment effect.

---

[3]Alternatively, one can view this as a parameter restriction rather than a support condition. This restriction is substantive in the sense that identification of β_0 does no require unbounded support of $z_1 \beta_0$.

The outline of the paper is as follows. Section 2 introduces the generalized regression model, a model similar to Han (1987) but with the inclusion of an endogenous regressor. To complete the specification of the (triangular) model, a reduced-form model is utilized for the endogenous regressor. Focusing upon the case of a binary endogenous regressor, Section 3 introduces a three-step procedure for testing significance of the causal (or treatment) effect of the endogenous regressor. The first stage estimates the parameters of the reduced-form model and, thus, is not new. The second stage consistently estimates the coefficients (up-to-scale) for the non-endogenous regressors. The third stage computes the test statistic, which turns out to be a kernel-weighted version of the *tau* statistic of Kendall (1938). Section 4 presents the main asymptotic results for the second-stage estimator and third-stage test statistic, both of which are shown to be $\sqrt{n}$-consistent and asymptotically normal. Since the (scalar) test statistic is asymptotically normal, the test for statistical significance of the causal effect is simply a z-test. Section 5 discusses the natural extensions of the approach to non-binary endogenous regressors and to multiple endogenous regressors. Section 6 provides Monte Carlo simulations that focus upon the performance of the third-stage test statistic. Finally, Section 7 considers an empirical application based upon Angrist and Evans (1998), in which we test for a causal effect of fertility (specifically, having a third child) upon mothers' labor supply. The application highlights the feasibility of our approach (even with nearly 300,000 observations) and provides a comparison to the 2SLS approach often taken in binary-choice models with endogenous regressors.

2 The model

Let y_1 denote the dependent variable of interest, which is assumed to depend upon a vector of covariates z_1 and a single endogenous variable y_2. We consider the following (latent-variable) generalized regression model for y_1:

$$y_1^* = F(z_1\beta_0, y_2, \epsilon), \quad y_1 = D(y_1^*) \tag{2.1}$$

The model for the latent dependent variable y_1^* has a general linear-index form, where ϵ is the error disturbance (independent of z_1) and F is a possibly unknown function that is assumed to be strictly monotonic in its first and third arguments and weakly monotonic in its second argument. The observed dependent variable is y_1, where the function D is weakly increasing and non-degenerate. The model in (2.1) is similar to the *generalized regression model* of Han (1987), except for the inclusion of the endogenous variable y_2. This model

encompasses many non-linear microeconometric models of interest, including binary-choice models, ordered-choice models, censored-regression models, transformation (e.g., Box-Cox) models, and proportional hazards duration models.

Note that the endogenous variable y_2 enters separably in the model for y_1^*. This formulation includes the traditional additively separable case (i.e., $z_1\beta_0 + \alpha_0 y_2$) considered in Blundell and Powell (2003) and Yildiz (2006) but allows for other forms of separability.[4] In addition to consistently estimating β in the presence of y_2, researchers are also interested in determining whether the endogenous variable y_2 has an effect upon y_1 and, if so, the direction of this effect. More formally, in the context of the generalized regression model, the null hypothesis of no effect of y_2 upon y_1 is

$$H_0: \ F(v, y_2', \epsilon) = F(v, y_2'', \epsilon) \ \text{ for all } y_2', \ y_2'', \ v, \ \epsilon. \tag{2.2}$$

In contrast, a positive effect of y_2 upon y_1 is equivalent to

$$F(v, y_2', \epsilon) > F(v, y_2'', \epsilon) \ \text{ for all } y_2' > y_2'', \ v, \ \epsilon, \tag{2.3}$$

and a negative effect of y_2 upon y_1 is equivalent to

$$F(v, y_2', \epsilon) < F(v, y_2'', \epsilon) \ \text{ for all } y_2' < y_2'', \ v, \ \epsilon. \tag{2.4}$$

As is common in econometric practice, the three alternatives (2.2)–(2.4) rule out the case that y_2 may have a positive effect for some $z_1\beta_0$ values and a negative effect for other $z_1\beta_0$ values. For instance, in the traditional linear-index approach where z_1 and y_2 enter through the linear combination $z_1\beta_0 + \alpha_0 y_2$, the value of α_0 determines which of the above three cases is relevant ($\alpha_0 = 0$: no effect; $\alpha_0 > 0$: positive effect; and, $\alpha_0 < 0$: negative effect). In the presence of possibly non-monotonic effects of y_2 on y_1, it is straightforward to apply the testing component of this paper (i.e., testing H_0 above) to different regions of the covariate space. The proposed estimator of β_0 will be entirely unaffected by non-monotonicities in the effects of y_2 on y_1.

Turning to the model for the endogenous regressor, we first focus on the case of a binary endogenous regressor in order to simplify exposition. (The general treatment of a discrete or continuous endogenous regressor is considered in Section 5.) The binary endogenous variable y_2 is assumed to be determined by the following reduced-form model:

$$y_2 = 1[z\delta_0 + \eta > 0], \tag{2.5}$$

---

[4]Vytlacil and Yildiz (2007) also consider a weakly separable model with the added generality that z_1 enters non-parametrically (rather than through a linear index).

where $z \equiv (z_1, z_2)$ is the vector of "instruments" and η is an error disturbance independent of z. The z_2 subcomponent of z provides the exclusion restrictions in the model. Endogeneity of y_2 in (2.1) arises when ϵ and η are not independent of each other. Estimation of the model in (2.5) is standard. When dealing with a binary endogenous regressor, we will use the common terminology "treatment effect" rather then referring to the "causal effect of y_2 on y_1." Thus, for example, a positive treatment effect would correspond to the case of equation (2.3) where y_2 can take on only two values: $F(v, 1, \epsilon) > F(v, 0, \epsilon)$ for all v, ϵ.

The binary-choice model with a binary endogenous regressor is a special case of the model in (2.1). The linear-index form of this model, with an additively separable endogenous variable, is given by

$$y_1 = 1[z_1\beta_0 + \alpha_0 y_2 + \epsilon > 0]. \tag{2.6}$$

Parametric assumptions on the error disturbances (e.g., bivariate normality of (ϵ, η)) naturally lead to maximum likelihood estimation of $(\beta_0, \alpha_0, \delta_0)$, as described in Wooldridge (2002, Section 15.7.3) and implemented by Evans and Schwab (1995).[5] The semiparametric version of this model (i.e., the distribution of (ϵ, η) being left unspecified) has been considered by Yildiz (2006), whose estimation approach requires that all components of z be continuous.

3   Estimation and testing for a treatment effect

The testing approach consists of three stages. In the first stage, the reduced-form parameters δ_0 are estimated. In the second stage, the coefficients of the exogenous variables (β_0) in the structural model are estimated. Then, in the third stage, a treatment-effect test statistic is calculated. Each of the three stages is described in turn below.

Stage 1: Estimation of δ_0

When no distribution is assumed for η, several semiparametric binary-choice estimators exist for $\sqrt{n}$-consistent estimation of δ_0 up-to-scale (see Powell (1994) for a comprehensive review).[6] The following linear representation of the first-step estimator is assumed:

$$\hat{\delta} - \delta_0 = \frac{1}{n} \sum_{i=1}^{n} \psi_{\delta i} + o_p(n^{-1/2}), \tag{3.1}$$

---

[5]Another common estimation approach (see, e.g., Angrist and Evans (1998)) is to simply ignore the non-linearity in (2.6) and apply two-stage least squares to the system given by (2.6) and (2.5).

[6]With a parametric assumption on η, standard binary-choice MLE estimation (e.g., probit) would apply.

a representatation that exists for the available $\sqrt{n}$-consistent semiparametric estimators. Since the second stage of our estimation procedure utilizes rank-based procedures, we also focus our theoretical treatment of the first-stage estimator upon the use of a rank-based estimator (specifically, the maximum rank correlation (MRC) estimator of Han (1987)). We note, however, that any other $\sqrt{n}$-consistent estimator (parametric or semiparametric) of δ_0 could be used in the first stage; the empirical application of Section 7, for instance, uses a probit estimator in the first stage.

Stage 2: Estimation of β_0

The estimator of β_0 is based upon pairwise comparisons of the y_1 values. If (ϵ, η) is independent of z, note that the conditional distribution $\epsilon | y_2, z$ is given by

$$\Pr(\epsilon \le e \,|\, y_2, z) = \begin{cases} \Pr(\epsilon \le e \,|\, \eta \le -z\delta_0) & \text{if } y_2 = 0 \\ \Pr(\epsilon \le e \,|\, \eta > -z\delta_0) & \text{if } y_2 = 1 \end{cases} \tag{3.2}$$

If two observations (indexed i and j) have $y_{2i} = y_{2j}$ and $z_i\delta_0 = z_j\delta_0$, equation (3.2) implies that the conditional distributions $\epsilon_i | y_{i2}, z_i$ and $\epsilon_j | y_{j2}, z_j$ are identical. For such a pair of observations, the monotonicity of F with respect to the linear index $z_1\beta_0$ implies that

$$z_{1i}\beta_0 > z_{1j}\beta_0 \iff \tag{3.3}$$
$$\Pr(y_{1i} > y_{1j} \,|\, z_{1i}, z_{1j}, y_{2i} = y_{2j}, z_i\delta_0 = z_j\delta_0) > \Pr(y_{1i} < y_{1j} \,|\, z_{1i}, z_{1j}, y_{2i} = y_{2j}, z_i\delta_0 = z_j\delta_0).$$

Equation (3.3) forms the basis for the proposed estimator of β_0. Unfortunately, equation (3.3) can not be used directly for estimation since (i) δ_0 is unknown and (ii) having $z_i\delta_0 = z_j\delta_0$ might be a zero-probability event. Using the first-stage estimator $\hat{\delta}$ of δ_0,[7] note that equation (3.3) will be "approximately true" in large samples for a pair of observations with $y_{2i} = y_{2j}$ and $z_i\hat{\delta} \approx z_j\hat{\delta}$. This suggests the following kernel-weighted rank-based estimator of β_0:

$$\hat{\beta} \equiv \arg\max_{\beta \in \mathcal{B}} \frac{1}{n(n-1)} \sum_{i \ne j} 1[y_{2i} = y_{2j}] k_h(z_i\hat{\delta} - z_j\hat{\delta}) 1[y_{1i} > y_{1j}] 1[z_{1i}\beta > z_{1j}\beta], \tag{3.4}$$

where $k_h(u) \equiv h^{-1}k(u/h)$ for a kernel function $k(\cdot)$ and a bandwidth h that shrinks to zero as $n \to \infty$. The kernel weighting serves to place more weight on pairs of observations for

---

[7]Our method is not subject to the problems of the "forbidden regression" (in which fitted values are plugged in to a non-linear function prior to mimicking 2SLS). The first-stage plug-in estimator (of the reduced-form index) is used not as a regressor but rather as a matching mechanism. Matching also upon the value of the endogenous regressor ensures that there is no relationship between the structural error and the plug-in index.

which $z_i\hat{\delta}$ is close to $z_j\hat{\delta}$. Under appropriate assumptions, it is shown in Section 4 that $\hat{\beta}$ is a $\sqrt{n}$-consistent estimator of β_0.

Stage 3: Testing for a treatment effect

In order to test for the presence of a treatment effect, we propose a kernel-weighted version of Kendall's *tau* (or rank correlation) statistic (Kendall (1938)). To motivate this test statistic, we first substitute the reduced-form model (2.5) for the endogenous regressor into the structural model (2.1), which yields

$$y_1 = D(F(z_1\beta_0, 1(z\delta_0 + \eta > 0), \epsilon)). \tag{3.5}$$

For *fixed* $z_1\beta_0$, note that the sign of the rank correlation between y_1 and $z\delta_0$ will depend upon whether there is a positive treatment effect, a negative treatment effect, or no treatment effect. More precisely, for a pair of observations (indexed i and j) having $z_{1i}\beta_0 = z_{1j}\beta_0$, equation (3.5) implies

$$z_i\delta_0 > z_j\delta_0 \Longleftrightarrow \tag{3.6}$$
$$\Pr(y_{1i} > y_{1j} \mid z_i, z_j, z_{1i}\beta_0 = z_{1j}\beta_0) > \Pr(y_{1i} < y_{1j} \mid z_i, z_j, z_{1i}\beta_0 = z_{1j}\beta_0)$$

if there is a positive treatment effect (as in (2.3)), and

$$z_i\delta_0 > z_j\delta_0 \Longleftrightarrow \tag{3.7}$$
$$\Pr(y_{1i} > y_{1j} \mid z_i, z_j, z_{1i}\beta_0 = z_{1j}\beta_0) < \Pr(y_{1i} < y_{1j} \mid z_i, z_j, z_{1i}\beta_0 = z_{1j}\beta_0)$$

if there is a negative treatment effect (as in (2.4)). In the case of no treatment effect (as in (2.2)), it is trivially the case that

$$\Pr(y_{1i} > y_{1j} \mid z_i, z_j, z_{1i}\beta_0 = z_{1j}\beta_0) = \Pr(y_{1i} < y_{1j} \mid z_i, z_j, z_{1i}\beta_0 = z_{1j}\beta_0) \tag{3.8}$$

since y_{1i}^* and y_{1j}^* are identically distributed if $z_{1i}\beta_0 = z_{1j}\beta_0$.

Note that, unlike equation (3.3), these probability statements do *not* condition on y_2. In fact, the proposed treatment-effect test statistic below does not directly use the y_2 values. This approach is somewhat analogous to the second stage of two-stage least squares, where the endogenous regressors are not directly used in the regression; instead, their "fitted values" (projections onto the exogenous regressors) are included in the second-stage regression. In our context, the y_2 values play a role in estimation of δ_0 and β_0. Unlike two-stage least

squares, fitted values of y_2 are not used since linear projections are not appropriate in our general non-linear model.

To operationalize the empirical implications of the probability statements above, it is necessary to plug in the estimators $\hat{\delta}$ and $\hat{\beta}$ of δ_0 and β_0, respectively, and to place greater weight on pairs of observations for which $z_{1i}\hat{\beta} \approx z_{1j}\hat{\beta}$. This leads to the proposed treatment-effect test statistic, which is a kernel-weighted version of Kendall's tau:

$$\hat{\tau} \equiv \frac{\sum_{i \neq j} \hat{\omega}_{ij} sgn(y_{1i} - y_{1j}) sgn(z_i\hat{\delta} - z_j\hat{\delta})}{\sum_{i \neq j} \hat{\omega}_{ij}}, \tag{3.9}$$

where $sgn(v) = 1(v > 0) - 1(v < 0)$ and the (estimated) weights $\hat{\omega}_{ij}$ are defined as

$$\hat{\omega}_{ij} \equiv k_h(z_{1i}\hat{\beta} - z_{1j}\hat{\beta}). \tag{3.10}$$

Given asymptotically normal $\sqrt{n}$-consistent estimators $\hat{\delta}$ and $\hat{\beta}$, it is shown in Section 4 that $\hat{\tau}$ is also $\sqrt{n}$-consistent and asymptotically normal. The probability limit of $\hat{\tau}$ is

$$\tau_0 \equiv E[sgn(y_{1i} - y_{1j}) sgn(z_i\delta_0 - z_j\delta_0)|z_{1i}\beta_0 = z_{1i}\beta_0]. \tag{3.11}$$

Based upon (3.6)–(3.8), it is easy to show that $\tau_0 > 0$ for a positive treatment effect, $\tau_0 < 0$ for a negative treatment effect, and $\tau_0 = 0$ for no treatment effect. Therefore, it is straightforward to conduct a one-sided or two-sided z-test of $H_0 : \tau_0 = 0$ based upon $\hat{\tau}$ and its asymptotic standard error $se(\hat{\tau})$. This test for a treatment effect is consistent against the alternatives of a positive or negative treatment effect.[8]

In situations in which z_1 is empty (i.e., only y_2 enters the structural model for y_1) or z_1 has a single component, the test statistic defined in (3.9) simplifies somewhat. These two cases are considered separately:

Case 1: z_1 has no elements

In this case, one can re-write the structural latent-variable model in (2.1) as $y_1^* = F(y_2, \epsilon)$. Since the weights $\hat{\omega}_{ij}$ are identical for all i and j, it follows directly that

$$\hat{\tau} = \frac{1}{n(n-1)} \sum_{i \neq j} sgn(y_{1i} - y_{1j}) sgn(z_i\hat{\delta} - z_j\hat{\delta}), \tag{3.12}$$

---

[8]If the treatment effect is positive for some $z_1\beta_0$ and negative for some $z_1\beta_0$, it would be necessary to use local versions of $\hat{\tau}$ in order to construct a consistent test. See, for example, Ghosal, Sen, and van der Vaart (2000) and Abreveya and Jiang (2005), who develop consistent tests in similar U-statistic frameworks.

which is just the Kendall's tau statistic for rank correlation between y_1 and $z\hat{\delta}$. If, in addition, $z\,(=z_2)$ contains a single "instrument" (the just-identified case), then $\hat{\delta}$ would simply reflect the sign of z's coefficient within the reduced-form model (2.5).[9] If $\hat{\delta} > 0$, the test statistic would further simplify to be the rank correlation between y_1 and z:

$$\hat{\tau} = \frac{1}{n(n-1)} \sum_{i \neq j} sgn(y_{1i} - y_{1j}) sgn(z_i - z_j); \tag{3.13}$$

if $\hat{\delta} < 0$, the test statistic would be the negative of the rank correlation between y_1 and z. If z is a binary instrument, the test statistic in (3.13) can be thought of as a rank-based analogue to the "Wald estimator" commonly used in treatment-effect contexts.

Case 2: z_1 has one element

If z_1 has a single element, the second-stage estimation of β_0 is not necessary to construct the test statistic. The estimated weights $\hat{\omega}_{ij}$ simplify to $\hat{\omega}_{ij} = k_h(z_{1i} - z_{1j})$. For a continuous covariate z_1, the test statistic takes the same form as (3.9) with these simplified weights. For a discrete covariate z_1, since $h \to 0$ as $n \to \infty$, weights will only be placed on pairs of observations having $z_{1i} = z_{1j}$.[10] Therefore, for a discrete covariate z_1, the test statistic could be re-written as

$$\hat{\tau} = \frac{\sum_{i \neq j} 1(z_{1i} = z_{1j}) sgn(y_{1i} - y_{1j}) sgn(z_i \hat{\delta} - z_j \hat{\delta})}{\sum_{i \neq j} 1(z_{1i} = z_{1j})}. \tag{3.14}$$

4 Asymptotic properties

In this section, we outline the asymptotic theory for the three-stage test-statistic procedure. The first result concerns the asymptotic distribution for the second-stage estimator of β_0. Since β_0 is only identified up to scale, we will normalize its last component to 1 and denote its other components by θ_0 and the corresponding estimator by $\hat{\theta}$, where

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{i \neq j} 1[y_{2i} = y_{2j}] k_h(z_i'\hat{\delta} - z_j'\hat{\delta}) 1[y_{1i} > y_{1j}] 1[z_{1i}'\beta(\theta) > z_{1j}'\beta(\theta)] \tag{4.1}$$

[9]As is well known in the semiparametric literature, this sign can be estimated at faster than the $\sqrt{n}$ rate. As such, its estimation would not affect the asymptotic distribution of $\hat{\tau}$ as it does in the more general case where z has more than one element.

[10]We are making the fairly innocuous assumption that the discrete points of support for z_1 do not change as n gets larger. Also, if z_1 is discrete, z_2 would need to have a continuous element in order identify δ_0 in the semiparametric case.

We impose the following regularity conditions:

Assumption CPS (Parameter Space) θ_0 lies in the interior of Θ, a compact subset of $\mathbf{R}^{k-1}$.

Assumption FS The first stage estimator used to estimate δ_0 will be the maximum rank correlation estimator of Han (1987). Consequently, the same regularity conditions in that paper and Sherman (1993) will be assumed so we will have a linear representation as discussed in the previous section. We normalize one of the coefficients of δ_0 to 1, and assume the corresponding regressor is continuously distributed on its support.

Assumption K (Matching stage kernel function) The kernel function $k(\cdot)$ used in the second stage is assumed to have the following properties:

    K.1 $k(\cdot)$ is twice continuously differentiable, has compact support and integrates to 1.

    K.2 $k(\cdot)$ is symmetric about 0.

    K.3 $k(\cdot)$ is a pth order kernel, where p is an even integer:

$$\int u^l k(u)du = 0 \ \text{ for } l = 1, 2, ...p - 1$$
$$\int u^p k(u)du \neq 0$$

Assumption H (Matching stages bandwidth sequence) The bandwidth sequence h_n used in the second stages satisfies the conditions, $\sqrt{n}h_n^p \to 0$, $\sqrt{n}h_n^3 \to \infty$.

Assumption RD (Last Regressor Properties) $z_{1i}^{(k)}$ is continuously distributed, with positive density on the real line.

Assumption ED (Error Distribution) ϵ_i is distributed independently of the regressors z_{1i}, z_i, and is continuously distributed with positive density on the real line.

Assumption FR (Full Rank Condition) The support of z_{1i} does not lie in a proper linear subspace of R^k.

The following lemma, whose proof is provided in the Appendix, establishes the asymptotic properties of the second stage estimator of θ_0. Some additional notation is used in the statement of the lemma. The reduced-form linear index is denoted $\zeta_{\delta i} = z_i'\delta_0$ and $f_{\zeta\delta}()$ denotes its density function. F_{Z_1} denotes the distribution function of z_{1i}. Also, $\nabla_{\theta\theta}$ denotes the second-derivative operator.

Lemma 4.1 *If Assumptions CPS, FS, K, H, RD, ED, and FR hold, then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, V^{-1}\Omega V^{-1}) \tag{4.2}$$

or, alternatively, $\hat{\theta} - \theta_0$ has the linear representation

$$\hat{\theta} - \theta_0 = \frac{1}{n}\sum_{i=1}^{n}\psi_{\beta i} + o_p(n^{-1/2}) \tag{4.3}$$

with $V = \nabla_{\theta\theta}\mathcal{N}(\theta)\mid_{\theta=\theta_0}$ and $\Omega = E[\delta_{1i}\delta'_{1i}]$, and $\psi_{\beta i} = V^{-1}\delta_{1i}$, where

$$\mathcal{N}(\theta) = \int 1[z'_{1i}\beta(\theta) > z'_{1j}\beta(\theta)]\mathcal{H}(\zeta_j, \zeta_j)\mathcal{F}(z'_{1i}\beta_0, z'_{1j}\beta_0, \zeta_j, \zeta_j)dF_{Z_1,\zeta}(z_{1i}, \zeta_j)dF_{Z_1,\zeta}(z_{1j}, \zeta_j) \tag{4.4}$$

with $\zeta_i = z'_i\delta_0$, whose density function is denoted by f_ζ, and where

$$\mathcal{F}(z'_{1i}\beta_0, z'_{1j}\beta_0, \zeta_i, \zeta_j) = P(y_{1i} > y_{1j}|y_{2i} = y_{2j}, z_{1i}, z_{1j}, \zeta_i, \zeta_j) \tag{4.5}$$

$$\mathcal{H}(\zeta_i, \zeta_j) = P(y_{2i} = y_{2j}|\zeta_i, \zeta_j) \tag{4.6}$$

and the mean-zero vector δ_{1i} is given by

$$\delta_{1i} = \left(\int f_\zeta(\zeta_i)\mu_{31}(\zeta_i, \zeta_i, \beta_0)d\zeta_i\right)\psi_{\delta i} \tag{4.7}$$

where

$$\mu(t, \zeta, \beta) = \mathcal{H}(t, \zeta)\mathcal{M}(t, \zeta, \beta)f_\zeta(t) \tag{4.8}$$

with

$$\mathcal{M}(t, \zeta, \beta) = E[\mathcal{F}(z'_{1i}\beta_0, z'_{1j}\beta_0, \zeta_i, \zeta_j)1[z'_{1i}\beta > z'_{1j}\beta]z'_i \mid \zeta_i = t, \zeta_j = \zeta] \tag{4.9}$$

and $\mu_1(\cdot, \cdot, \cdot)$ denotes the partial derivative of $\mu(\cdot, \cdot, \cdot)$ with respect to its first argument and $\mu_{31}(\cdot, \cdot, \cdot)$ denotes the partial derivative of $\mu_1(\cdot, \cdot, \cdot)$ with respect to its third argument.

The asymptotic theory for the third-stage test statistic is based on the above conditions, now also assuming conditions K and H are valid for the third stage matching kernel, and the following additional smoothness condition:

Assumption S (Order of Smoothness of Density and Conditional Expectation Functions)

    S.1 Letting $\zeta_{\beta i}$ denote $z'_{1i}\beta_0$, and let $f_{\zeta\beta}(\cdot)$ denote its density function, we assume $f_{\zeta\beta}(\cdot)$ is p times continuously differentiable with derivatives that are bounded on the support of $\zeta_{\beta i}$.

S.2 The functions $G_{11}(\cdot)$ and $G_x(\cdot)$, defined as follows:

$$G_{11}(\cdot) = E[sgn(y_{1i} - y_{1j})f_{Z_k|Z_{-k}}(\Delta z'_{-kij}\delta_0^{(-k)})\Delta z'_{-kij}|\zeta_{\beta i} = \cdot, \zeta_{\beta j} = \cdot] \quad (4.10)$$

$$G_x(\cdot) = E[(sgn(y_{1i} - y_{1j})sgn(z'_i\delta_0 - z'_j\delta_0) - \tau_0)(z_{1i} - z_{1j})'|z_{1i} - z_{1j} = \cdot] \quad (4.11)$$

are all assumed to be all p times continuously differentiable with derivatives that are bounded on the support of $\zeta_{\beta i}$.

The following theorem, whose proof is also left to the Appendix, establishes the asymptotic distribution of the test statistic τ:

Theorem 4.1 *If Assumptions CPS, FS, K, H, RD, ED, FR, and S hold, then*

$$\sqrt{n}(\hat{\tau} - \tau_0) \Rightarrow N(0, V_2^{-2}\Omega_2) \quad (4.12)$$

with $V_2 = E[f_{\zeta\beta}(\zeta_{\beta i})]$ and $\Omega_2 = E[\delta_{2i}^2]$. The mean-zero random variable δ_{2i} is

$$\delta_{2i} = 2f_{\zeta\beta}(\zeta_{\beta i})G(y_{1i}, z_i, \zeta_{\beta i}) + E[G'_x(\zeta_{\beta i})f_{\zeta\beta}(\zeta_{\beta i})]\psi_{\beta i} + E[G_{11}(\zeta_{\beta i})f_{\zeta\beta}(\zeta_{\beta}i)]\psi_{\delta i}, \quad (4.13)$$

where $G'_x()$ denotes the derivative of G_x and $G()$ is given by

$$G(y_1, z, \zeta) = E[sgn(y_{1i} - y_1)sgn(z'_i\delta_0 - z'\delta_0)|\zeta_{\beta i} = \zeta]. \quad (4.14)$$

The (scalar) test statistic $\hat{\tau}$ is $\sqrt{n}$-consistent and asymptotically normal, which implies that testing the null hypothesis $H_0 : \tau_0 = 0$ is a simple z-test. Given $\hat{\tau}$ and an estimated asymptotic standard error $\hat{v}_{\hat{\tau}}$, one-sided or two-sided versions of this test can be implemented based upon the ratio $\hat{\tau}/\hat{v}_{\hat{\tau}}$. In order to compute the standard error $\hat{v}_{\hat{\tau}}$, we recommend the use of the bootstrap since the forms of the asymptotic variances in Lemma 4.1 and Theorem 4.1 are somewhat complicated. In the empirical application of Section 7, the bootstrap is used for inference purposes. Although the bootstrap has not formally been shown to be consistent in the specific context considered, there is no reason to expect failure of the bootstrap given that each stage of the testing procedure is $\sqrt{n}$-consistent. Recently, Subbotin (2006) has shown consistency of the bootstrap for the maximum rank correlation estimator (our first-stage estimator). It is worthy of future research to investigate whether the approach of Subbotin (2006) could be extended to kernel-weighted rank estimators (like $\hat{\beta}$ and $\hat{\tau}$).

5   Extensions

5.1   Non-binary endogenous regressors

The key to consistency of the second-stage estimator of β_0 is the ability to compare observation-pairs for which the distribution of ϵ (conditional on y_2 and z) is the same. In the case of a binary endogenous regressor considered above, the $1[y_{2i} = y_{2j}]k_h(z_i\hat{\delta} - z_j\hat{\delta})$ factor in the objective function in (3.4) serves to focus on such observation pairs. This same idea can be easily generalized to cases in which the endogenous regressor y_2 is non-binary. Specifically, we specify a reduced-form model for y_2 that is the generalized regression model of Han (1987):

$$y_2^* = F_2(z\delta_0, \eta), \ y_2 = D_2(y_2^*). \tag{5.1}$$

As before, we assume that (ϵ, η) are independent of z. The first-stage estimator of δ_0 would be the maximum rank correlation estimator of Han (1987) (or, as mentioned before, any other semiparametric linear-index estimator). In the case of a continuous y_2 variable, it is worth noting that the model in (5.1) does not require that the functional form of F_2 is known and, therefore, is more general than Blundell and Powell (2003).

The second-stage estimator of β_0 takes a form very similar to (3.4), except that weight is placed on observation-pairs having y_2 values close to each other (not necessarily equal):

$$\hat{\beta} \equiv \arg\max_{\beta\in\mathcal{B}} \frac{1}{n(n-1)} \sum_{i\neq j} k_h[y_{2i} - y_{2j}]k_h(z_i\hat{\delta} - z_j\hat{\delta})1[y_{1i} > y_{1j}]1[z_{1i}\beta > z_{1j}\beta]. \tag{5.2}$$

The third-stage test statistic ($\hat{\tau}$) of the previous section remains the same (once $\hat{\delta}$ and $\hat{\beta}$ have been computed).

5.2   Multiple endogenous regressors

To illustrate how our testing approach generalizes to multiple endogenous regressors, we first consider the case of two binary endogenous regressors. We then show that the basic idea easily extends to more general cases and can be used, for example, when there are a mix of discrete and continuous endogenous regressors.

5.2.1 Two binary endogenous regressors

Consider the following model, which extends the model of Section 2 to the case of two binary endogenous regressors y_{21} and y_{22}:

$$y_1^* = F(z_1\beta_0, y_{21}, y_{22}, \epsilon), \ y_1 = D(y_1^*) \tag{5.3}$$

$$y_{21} = 1[z\delta_{01} + \eta_1 > 0] \tag{5.4}$$

$$y_{22} = 1[z\delta_{02} + \eta_2 > 0] \tag{5.5}$$

The error disturbances $(\epsilon, \eta_1, \eta_2)$ are assumed to be independent of z. The reduced-form parameter vectors δ_{01} and δ_{02} can be consistently estimated with existing binary-choice-model estimators.

The second-stage estimator $\hat{\beta}$ would maximize the objective function

$$\frac{1}{n(n-1)} \sum_{i\neq j} 1[y_{21i} = y_{21j}]1[y_{22i} = y_{22j}]k_h(z_i\hat{\delta}_1 - z_j\hat{\delta}_1)k_h(z_i\hat{\delta}_2 - z_j\hat{\delta}_2)1[y_{1i} > y_{1j}]1[z_{1i}\beta > z_{1j}\beta].$$
$$\tag{5.6}$$

Observation pairs are considered in the objective function when their y_{21} and y_{22} values are identical and when their $z_i\hat{\delta}_1$ and $z_i\hat{\delta}_2$ index values are similar.

Given estimators for δ_{01}, δ_{02}, and β_0, we now consider the problem of testing the significance of causal effects. Combining the various equations of our model yields the following:

$$y_1 = D(F(z_1\beta, 1[z\delta_{01} + \eta_1 > 0], 1[z\delta_{02} + \eta_2 > 0], \epsilon)). \tag{5.7}$$

To test the effect of y_{21} (second argument) upon y_1, we want to fix $z_1\beta$ and $z\delta_{02}$ and examine the significance of the relationship between y_1 and $z\delta_{01}$. (Similarly, to test the effect of y_{22} upon y_1, we want to fix $z_1\beta$ and $z\delta_{01}$ and examine the significance of the relationship between y_1 and $z\delta_{02}$.) This idea can be operationalized with the following kernel-weighted rank-based test statistic:

$$\hat{\tau}_1 \equiv \frac{\sum_{i\neq j} \hat{\omega}_{ij,1} sgn(y_{1i} - y_{1j})sgn(z_i\hat{\delta}_1 - z_j\hat{\delta}_1)}{\sum_{i\neq j} \hat{\omega}_{ij,1}}, \tag{5.8}$$

where the (estimated) weights $\hat{\omega}_{ij,1}$ are defined as

$$\hat{\omega}_{ij,1} \equiv k_h(z_{1i}\hat{\beta} - z_{1j}\hat{\beta})k_h(z_i\hat{\delta}_2 - z_j\hat{\delta}_2). \tag{5.9}$$

The analogous test statistic (for testing the effect of y_{22} on y_1) would be as follows:

$$\hat{\tau}_2 \equiv \frac{\sum_{i\neq j} \hat{\omega}_{ij,2} sgn(y_{1i} - y_{1j})sgn(z_i\hat{\delta}_2 - z_j\hat{\delta}_2)}{\sum_{i\neq j} \hat{\omega}_{ij,2}}, \tag{5.10}$$

where the (estimated) weights $\hat{\omega}_{ij,2}$ are defined as

$$\hat{\omega}_{ij,2} \equiv k_h(z_{1i}\hat{\beta} - z_{1j}\hat{\beta})k_h(z_i\hat{\delta}_1 - z_j\hat{\delta}_1). \tag{5.11}$$

5.2.2 The general case

The endogenous regressors are denoted $y_{21}, y_{22}, \ldots, y_{2Q}$, where Q is the number of endogenous regressors. The $Q \times 1$ vector $\mathbf{y}_2$ is defined as $\mathbf{y}_2 = (y_{21}, y_{22}, \ldots, y_{2Q})'$. Each endogenous regressor y_{2q} (for $q = 1, \ldots, Q$) has a reduced-form generalized regression model as in (5.1):

$$y_{2q}^* = F_{2q}(z\delta_{0q}, \eta_q), \ y_{2q} = D_{2q}(y_{2q}^*). \tag{5.12}$$

The error disturbances $(\epsilon, \eta_1, \ldots, \eta_Q)$ are assumed to be independent of z. The functions F_{2q} and D_{2q} may differ over q, allowing for an arbitrary mix of discrete and continuous endogenous regressors.

To simplify notation somewhat, define $\Delta_0 \equiv (\delta_{01}, \ldots, \delta_{0Q})'$ to be the $Q \times \ell$ matrix containing all of the reduced-form coefficients (where ℓ is the dimension of z). Each of the δ_{0q} coefficient vectors can be estimated $\sqrt{n}$-consistently in a first stage using equation-by-equation semiparametric estimation (e.g., maximum rank correlation or some other linear-index estimator). The estimate of δ_{0q} (for $q = 1, \ldots, Q$) is denoted $\hat{\delta}_q$, and the $Q \times \ell$ matrix $\widehat{\Delta}$ is defined as $\widehat{\Delta} \equiv (\hat{\delta}_1, \ldots, \hat{\delta}_Q)'$.

For the second-stage estimator $\hat{\beta}$, we generalize the approach from Section 5.2.1 and focus upon observations pairs (i, j) for which $\mathbf{y}_{2i}$ is close to $\mathbf{y}_{2j}$ and $\widehat{\Delta}z_i$ is close to $\widehat{\Delta}z_j$. Specifically, the second-stage estimator $\hat{\beta}$ maximizes the objective function

$$\frac{1}{n(n-1)} \sum_{i \neq j} K_h(\mathbf{y}_{2i} - \mathbf{y}_{2j})K_h(\widehat{\Delta}z_i - \widehat{\Delta}z_j)1[y_{1i} > y_{1j}]1[z_{1i}\beta > z_{1j}\beta]. \tag{5.13}$$

where $K_h(\cdot)$ is a multivariate kernel function defined as $K_h(\mathbf{v}) \equiv \prod_{q=1}^{\dim(\mathbf{v})} k_h(v_q)$ for a vector $\mathbf{v}$.

To test the effect of y_{2q} upon y_1 (for any $q = 1, \ldots, Q$), we want to fix $z_1\beta$ and $z\delta_{0p}$ for all $p \neq q$ and examine the significance of the relationship between y_1 and $z\delta_{0q}$. Let $\widehat{\Delta}_{-q}$ denote the matrix $\widehat{\Delta}$ with the q-th row (i.e., $\hat{\delta}_q'$) removed, so that $\widehat{\Delta}_{-q}$ has dimension $(Q-1) \times \ell$. The test statistic associated with the q-th endogenous regressor is thus given by:

$$\hat{\tau}_q \equiv \frac{\sum_{i \neq j} \hat{\omega}_{ij,q} sgn(y_{1i} - y_{1j})sgn(z_i\hat{\delta}_q - z_j\hat{\delta}_q)}{\sum_{i \neq j} \hat{\omega}_{ij,q}}, \tag{5.14}$$

where the (estimated) weights $\hat{\omega}_{ij,q}$ are defined as

$$\hat{\omega}_{ij,q} \equiv k_h(z_{1i}\hat{\beta} - z_{1j}\hat{\beta})K_h(\widehat{\Delta}_{-q}z_i - \widehat{\Delta}_{-q}z_j). \tag{5.15}$$

The asymptotic theory for the general case is completely analogous to the results developed previously. The regularity conditions which change for the general case are conditions **H** and **S**. Condition **H** now becomes:

Assumption H' $\sqrt{n}h_n^p \to 0$, $\sqrt{n}h_n^{(m+2)} \to \infty$, where m denotes the number of matches, so $m = 2Q$ in the second stage and $m = Q$ in the third stage.

The smoothness condition in Assumption **S** now applies to each of the random variables being matched in the second and third stages.

6   Monte Carlo simulations

In this section, we consider evidence from Monte Carlo simulations which focus upon the performance of the third-stage test statistic proposed above. The following simple design, with a single (continuous) instrumental variable for the binary endogenous regressor, is considered:

$$y_{1i} = 1[\alpha_0 y_{2i} + \epsilon_i > 0] \tag{6.16}$$

$$y_{2i} = 1[z_i + \eta_i > 0] \tag{6.17}$$

where $z_i \sim N(0,1)$ and $\begin{pmatrix} \epsilon_i \\ \eta_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix} \right)$. Two parameters, α_0 (the coefficient on the binary endogenous variable) and ρ_0 (the correlation between η_i and ϵ_i), are chosen to vary over the simulation designs. In particular, the values $\rho_0 = 0, 0.25, 0.50, 0.75$ and $\alpha_0 = 0, 0.1, 0.2, 0.3$ are considered, yielding 16 different designs.

For each of the 16 designs, 1,000 simulations were conducted with a sample size of 500 observations ($n = 500$). Three different approaches to testing the significance of α (i.e., testing $H_0 : \alpha_0 = 0$) were considered: (i) a full MLE estimation strategy (with the test based upon the z-statistic of $\hat{\alpha}_{mle}$), (ii) a linear IV estimation strategy (with the test based upon the z-statistic of $\hat{\alpha}_{iv}$, and (iii) the third-stage test statistic $\hat{\tau}$ proposed above (which, in this case, is the Kendall's tau correlation between y_{1i} and z_i). Table 1 summarizes the results, with rejection rates reported for the 5% and 10% levels for the three approaches (labeled MLE, IV, and $\hat{\tau}$, respectively). The first four rows of the table correspond to $\alpha_0 = 0$ and, therefore, provide evidence on the size of the test. The rejection rates for the three tests are in line with the 5% and 10% levels, although at the highest level of correlation ($\rho_0 = 0.75$) between the error disturbances, the MLE approach exhibits some over-rejection. The remaining rows

of the table provide evidence on the power of the test (for $\alpha_0 = 0.1, 0.2, 0.3$). Overall, the power of the alternative approaches is remarkably similar across these designs. For $\rho_0 = 0.75$, the MLE approach does have higher rejection rates, but these are likely the result of the over-rejection phenomenon seen in the $\alpha_0 = 0/\rho_0 = 0.75$ design; for the other ρ_0 values, the MLE approach has rejection rates which are basically indistinguishable from the other two approaches.

For this simple Monte Carlo design, the semiparametric approach to testing for significance of the binary endogenous variable compares favorably with both an MLE approach and a linear IV approach. There seems to be no loss of power associated the rank-based test. The empirical application of the next section considers a more realistic situation with additional exogenous covariates.

7   Empirical application

In this section, we apply our estimation and testing methodology to an empirical application concerning the effects of fertility on female labor supply. In particular, we adopt the approach of Angrist and Evans (1998), who use the gender mix of a woman's first two children to instrument for the decision to have a third child. This instrumental-variable strategy allows one to identify the effect of having a third child upon the woman's labor-supply decision. The rationale for this strategy is that child gender is arguably randomly assigned and that, in the United States, families whose first two children are the same gender are significantly more likely to have a third child.

Using 1980 and 1990 Census data, Angrist and Evans (1998) find that married women whose first two children are the same gender are 5–8% more likely to have a third child; using the same-sex indicator as an instrument for having a third child, they find that having a third child lowers the probability of a married women working for pay by about 10–12%. Rather than using the 1980 and 1990 Census data, the sample for the current study is drawn from the 2000 Census data (5-percent public-use microdata sample (PUMS)) in order to see if any interesting changes have occurred in the relationship between fertility and labor supply. Starting from the household PUMS data, a mother was retained in the sample if all of the following criteria were satisfied: (i) mother has two or more children, (ii) mother is white, (iii) mother is a United States citizen, (iv) mother is married with spouse present in household, and (v) oldest child is 12 years of age or younger.[11] In addition, to eliminate

---

[11] The PUMS data contains information on children under the age of 18 that are living in the household.

any families that might have twin births (or higher-order multiple births), any family with same-aged first and second children or same-aged second and third children were dropped from the sample. The resulting sample consists of 293,771 observations.

Summary statistics for the variables to be used in the analysis are provided in Table 2. The table shows that 69.7% of mothers in the sample worked for pay during 1999 and 25.7% of mothers had a third child. The percentage of women working for pay represents a very slight increase over the comparable percentage from the 1990 Census data, and the percentage having a third child represents a decline from 1990. In the analysis, the outcome of interest (y_1) is whether the mother worked in 1999, the binary endogenous explanatory variable (y_2) is the presence of a third child, and the instrument is whether the mother's first two children were of the same gender.

Table 3 reports the first-stage regression results, i.e. regressing the have-third-child indicator upon the same-sex indicator variable and the other z variables. The linear probability estimates and the probit estimates indicate that mothers whose first two children are the same gender are 5.6–5.8 percentage points more likely to have a third child than mothers whose first two children are of different gender. These estimates are very similar, although slightly lower in magnitude, to those found by Angrist and Evans (1998, Table 5) for the earlier 1980 and 1990 samples. Table 4 reports the second-stage estimates for the (linear) two-stage least squares estimator, along with the OLS estimates for comparison. Again, the results are very similar to those found by Angrist and Evans (1998, Tables 7 and 8), with the OLS estimates of the had-third-child effect on labor supply larger in magnitude than the 2SLS estimates.

Table 5 considers the alternative tests for significance of the binary endogenous regressor, comparing the semiparametric $\hat{\tau}$ test proposed in this paper with the z-test based upon the 2SLS estimates. In order to examine the effect of additional covariates, testing results are reported starting from a model with no exogenous covariates and then adding covariates onb-by-one until the full set of three exogenous covariates are included. (Note that the 2SLS estimates and standard errors for the no-covariate and three-covariate models correspond to those presented in Table 4.) In the model with no exogenous covariates, the z-statistics associated with $\hat{\tau}$ and the 2SLS coefficient are extremely similar. This finding is very much in

Unlike earlier editions of the PUMS data, the 2000 edition does not contain a data item for the "total number of children ever born." Therefore, the last criterion is used in order to make it more likely that the oldest child in the household is actually the mother's first child. The cutoff could be lowered further to increase this certainty but at the expense of decreasing the sample size.

line with the Monte Carlo simulation evidence of the previous section. The 2SLS z-statistic for the larger models is basically unchanged from the no-covariate model, which is not too surprising given that the same-sex instrument is uncorrelated with the other exogenous covariates in the model. In contrast, the magnitude of the z-statistic for the semiparametric $\hat{\tau}$ method does decline. The addition of covariates to the model forces the semiparametric method to make comparisons based upon observation-pairs with similar first-stage (estimated) index values associated with these exogenous covariates. It is encouraging, however, that the z-statistic magnitude does not decline by much as the second and third covariates are added to the model. Table 5 highlights the inherent robustness-power tradeoff between the semiparametric and parametric methodologies. Although one might have worried that the tradeoff would be so drastic to render the semiparametric method useless in practice, the results indicate that this is not the case. Even in the model with three covariates, the $\hat{\tau}$ estimate provides strong statistical evidence ($z = -2.69$) that the endogenous third-child indicator variable has a causal effect upon mothers' labor supply. Importantly, this finding is not subject to the inherent misspecification of the linear probability model or any type of parametric assumption on the error disturbances.

8   Concluding Remarks

This paper proposes a new method for testing for the causal effects of endogenous variables in a generalized regression model. The model considered here allows for multiple continuously and/or discretely distributed endogenous variables, thereby offering a test for cases not previously considered in the literature. The proposed test statistic converges at the parametric rate to a limiting normal distribution under the null hypothesis of no causal effect. The simulation study in Section 6 indicates excellent finite-sample performance, and an application to testing the causal effect of fertility illustrates the usefulness of the proposed approach when compared to the standard 2SLS approach often implemented in empirical work.

A useful extension would be a localized version of the proposed procedure that would allow the sign of the causal effect(s) to vary over the support of the random variables in question.

References

Abrevaya, Jason and Wei Jiang, 2005, A nonparametric approach to measuring and testing curvature, Journal of Business and Economic Statistics 23: 1–19.

Angrist, Joshua D. and William N. Evans, 1998, Children and their parents' labor supply: evidence from exogenous variation in family size, American Economic Review 88: 450–477.

Aradillas-Lopez, Andres, Bo E. Honoré, and James L. Powell, 2005, Pairwise difference estimation with nonparametric control variables, mimeo.

Blundell, Richard W. and James L. Powell, 2003, Endogeneity in nonparametric and semi-parametric regression models, Review of Economic Studies 71: 655–679.

Blundell, Richard W. and James L. Powell, 2004, Endogeneity in semiparametric binary response models, in: M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, eds., Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. II (Cambridge, UK: Cambridge University Press).

Evans, William N. and Robert M. Schwab, 1995, Finishing high school and starting college: do Catholic schools make a difference? Quarterly Journal of Economics 110: 941–974.

Ghosal, Subhashis, Arusharka Sen, and Aad W. van der Vaart, 2000, Testing monotonicity of regression, Annals of Statistics 28: 1054–1082.

Han, Aaron K., 1987, Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator, Journal of Econometrics 35: 303–316.

Hausman, Jerry A., 1978, Specification tests in econometrics, Econometrica 46: 1251–1271.

Hong, Han and Elie Tamer, 2003, Inference in censored models with endogenous regressors, Econometrica 71: 905–932.

Kan, Kamhon and Chihwa Kao, 2005, Simulation-based two-step estimation with endogenous regressors, mimeo.

Kendall, Maurice G., 1938, A new measure of rank correlation, Biometrika 30: 81–93.

Khan, Shakeeb, 2001, Two stage rank estimation of quantile index models, Journal of Econometrics 100: 319–355.

Lewbel, Arthur, 2000, Semiparametric qualitative response model estimation with unknown heteroscedasticity and instrumental variables, Journal of Econometrics 97: 145–177.

Newey, Whitney K. and Daniel L. McFadden, 1994, Large sample estimation and hypothesis testing, in: Robert F. Engle and Daniel L. McFadden, eds., Handbook of Econometrics, Vol. IV (North-Holland, Amsterdam) 2111–2245.

Powell, James L., 1994, Estimation of semiparametric models, in: Robert F. Engle and Daniel L. McFadden, eds., Handbook of Econometrics, Vol. IV (North-Holland, Amsterdam) 2443–2521.

Powell, James L., James H. Stock, and Thomas M. Stoker, 1989, Semiparametric estimation of index coefficients, Econometrica 57: 474–523.

Serfling, Robert J., 1980, Approximation theorems of mathematical statistics, John Wiley & Sons.

Sherman, Robert P., 1993, The limiting distribution of the maximum rank correlation estimator, Econometrica 61: 123–137.

Sherman, Robert P., 1994, Maximal inequalities for degenerate U-processes with applications to optimization estimators, Annals of Statistics 22: 439–459.

Subbotin, Viktor, 2006, On the bootstrap of the maximum rank correlation estimator, mimeo.

Vytlacil, Edward and Nese Yildiz, 2007, Dummy endogenous variables in weakly separable models, Econometrica, forthcoming.

Wooldridge, Jeffrey M., 2002, Econometric analysis of cross section and panel data, MIT Press.

Yildiz, Nese, 2006, Estimation of binary choice models with linear index and dummy endogenous variables, mimeo.

A   Proof of Lemma 4.1

The proof strategy will be along the lines of Sherman(1994), and we deliberately aim to keep notation as similar as possible to that used in that paper. Specifically, let $G_n(\theta)$ and $\hat{G}_n(\theta)$ be defined as[12]

$$G_n(\theta) \ = \ \frac{1}{n(n-1)} \sum_{i \neq j} 1[y_{2i} = y_{2j}] k_h(z_i'\delta_0 - z_j'\delta_0) 1[y_{1i} > y_{1j}] 1[z_{1i}'\beta(\theta) > z_{1j}'\beta(\theta)] \quad (A.1)$$

$$\hat{G}_n(\theta) \ = \ \frac{1}{n(n-1)} \sum_{i \neq j} 1[y_{2i} = y_{2j}] k_h(z_i'\hat{\delta} - z_j'\hat{\delta}) 1[y_{1i} > y_{2j}] 1[z_{1i}'\beta(\theta) > z_{1j}'\beta(\theta)] \quad (A.2)$$

Similar to Sherman (1994), the proof strategy involves the following stages:

1. Establish consistency of the estimator.

2. Show the estimator converges at the parametric ($\sqrt{n}$) rate.

3. Establish asymptotic normality of the estimator.

For consistency, we apply Theorem 2.1 in Newey and McFadden (1994). Compactness follows from Assumption CPS. To show uniform convergence, we note that the estimated first-stage index converges uniformly to the true index, by Assumption FS, so we can replace estimated indexes with true values inside the objective function, and work with $G_n(\theta)$. Next, we note by Theorem 2 in Sherman (1994),

$$\sup_{\theta \in \Theta} (G_n(\theta) - E[G_n(\theta)]) = o_p(1) \tag{A.3}$$

by the Euclidean property of the indicator function $1[z_{1i}'\beta > z_{1j}'\beta]$, and the uniform (in n) boundedness of $E[G_n(\theta)]$. By a change of variables and Assumptions K,H,

$$\sup_{\theta \in \Theta} E[G_n(\theta)] - \mathcal{G}(\theta) \xrightarrow{p} 0 \tag{A.4}$$

where we will define $\mathcal{G}(\theta)$ as follows. First, define the indicator $\tilde{d}_{ij}$ as $I[y_{2i} = y_{2j}]$, and define

$$\mathcal{H}(z_i'\delta_0, z_j'\delta_0) = E[\tilde{d}_{ij} | z_i'\delta_0, z_j'\delta_0]$$

---

[12]Implicit in the proofs that follow, we are subtracting the function $I[z_{1i}'\beta_0 > z_{1j}'\beta_0]$ from $1[z_{1i}'\beta(\theta) > z_{1j}'\beta(\theta)]$ in each of the two objective functions, analogous to Sherman(1993). The terms subtracted do not affect the value of the estimator, and are omitted for notational convenience.

Furthermore, define

$$\mathcal{F}(z_{1i}'\beta_0, z_{1j}'\beta_0, z_i'\delta_0, z_j'\delta_0) = P(y_{1i} > y_{1j}|\tilde{d}_{ij} = 1, z_{1i}'\beta_0, z_{1j}'\beta_0, z_i'\delta_0, z_j'\delta_0)$$

So we can define

$$\mathcal{G}(\theta) = E[\mathcal{F}(z_{1i}'\beta_0, z_{1j}'\beta_0, z_i'\delta_0, z_i'\delta_0)\mathcal{H}(z_i'\delta_0, z_i'\delta_0)1[z_{1i}'\beta(\theta) > z_{1j}'\beta(\theta)]] \quad (A.5)$$

where the above expectation is taken with respect to z_i, z_j. This establishes uniform convergence of $\hat{G}_n(\theta)$ to $\mathcal{G}(\theta)$. $\mathcal{G}(\theta)$ is continuous by the smoothness assumptions on the regressor vector z_{1i}, and the index $z_i'\delta_0$ distribution. Finally, as a last condition to apply Theorem 2.1 in Newey and McFadden (1994), we need to show that $\mathcal{G}(\theta)$ is uniquely maximized at θ_0. This follows from the distributional assumption on ϵ_i (Assumption ED), the index distributional assumption (Assumption RD) and the full rank condition (Assumption FR). This establishes consistency.

With consistency established, the next two stages can be established along the lines of Sherman (1994) and Khan (2001). Turning to deriving asymptotic normality of the estimator, note that we can apply Theorem 2 of Sherman (1994), of which a sufficient condition will be to show that uniformly over $O_p(1/\sqrt{n})$ neighborhoods of θ_0,

$$\hat{G}_n(\theta) = \frac{1}{2}\theta'V\theta + \frac{1}{\sqrt{n}}\theta'W_n + o_p(n^{-1}) \quad (A.6)$$

where V is a negative definite matrix whose form will given below, and W_n is asymptotically normal, with mean 0 and variance Ω (whose form is given below).

To show (A.6), we will work with the following expansion:

$$\hat{G}_n(\theta) = G_n(\theta) + G_n'(\theta) + R_n \quad (A.7)$$

where

$$G_n'(\theta) = \frac{1}{n(n-1)}\sum_{i \neq j} 1[y_{2i} = y_{2j}]h_n^{-1}k_h'(z_i'\hat{\delta} - z_j'\hat{\delta})1[y_{1i} > y_{1j}]1[z_{1i}'\beta(\theta) > z_{1j}'\beta(\theta)](\Delta z_{ij}'\hat{\delta} - \Delta z_{ij}'\delta_0) \quad (A.8)$$

with $k_h'(\cdot)$ denoting the derivative of the function $k_h(\cdot)$. R_n in (A.7) denotes the remainder term in the expansion, whose asymptotic properties will be dealt with after we derive the asymptotic properties of $G_n(\theta)$ and $G_n'(\theta)$.

The following Lemma establishes a representation for $G_n(\theta)$:

Lemma A.1 *Under the conditions RD, ED, FR, uniformly over $O_p(n^{-1/2})$ neighborhoods of θ_0, we have*

$$G_n(\theta) = \frac{1}{2}\theta'V\theta + \frac{1}{\sqrt{n}}\theta'W_n + o_p(n^{-1}) \quad (A.9)$$

where V is negative definite and W_n is asymptotically normal with mean 0, and variance Ω.

Proof: We will first evaluate a representation for $E[G_n(\theta)]$. We do this because we will later work with the U-statistic representation theorems found in, e.g. Serfling (1978). Letting $\zeta_i = z_i'\delta_0$, we write $E[G_n(\theta)]$ as the integral:

$$\int k_h(\Delta\zeta_{ij})1[z_{1i}'\beta > z_{1j}'\beta]\mathcal{H}(\zeta_i,\zeta_j)\mathcal{F}(z_{1i}'\beta_0, z_{1j}'\beta_0, \zeta_i, \zeta_j)dF_{Z_1,\zeta}(z_{1i},\zeta_i)dF_{Z_1,\zeta}(z_{1j},\zeta_j) \tag{A.10}$$

Next, we do the change of variables $u = \frac{\Delta\zeta_{ij}}{h_n}$ and obtain the following integral

$$\int k(u)1[z_{1i}'\beta > z_{1j}'\beta]\mathcal{H}(\zeta_j + uh_n, \zeta_j)\mathcal{F}(z_{1i}'\beta_0, z_{1j}'\beta_0, \zeta_j + uh_n, \zeta_j)$$

$$dF_{Z_1,\zeta}(z_{1i},\zeta_j)dF_{Z_1,\zeta}(z_{1j}, \zeta_j + uh_n)du \tag{A.11}$$

Taking a second-order expansion inside the integral around $uh_n = 0$, the lead term is of the form:

$$\int k(u)1[z_{1i}'\beta > z_{1j}'\beta]\mathcal{H}(\zeta_j,\zeta_j)\mathcal{F}(z_{1i}'\beta_0, z_{1j}'\beta_0, \zeta_j, \zeta_j)dF_{Z_1,\zeta}(z_{1i},\zeta_j)dF_{Z_1,\zeta}(z_{1j},\zeta_j)du \tag{A.12}$$

Note the term $\mathcal{F}(z_{1i}'\beta_0, z_{1j}'\beta_0, \zeta_j, \zeta_j)$ controls for selection bias. Therefore, we can use the same arguments as in Sherman (1993) to conclude that the integral in (A.12) is of the form

$$\frac{1}{2}\theta'V\theta + o_p(n^{-1}) \tag{A.13}$$

for θ uniformly in $O_p(n^{-1/2})$ neighborhoods of θ_0. The first-order term in the second-order expansion is 0 since $\int uk(u)du = 0$. The second-order term can be bounded above by

$$\left(C\int 1[z_{1i}'\beta > z_{1j}'\beta]dF_{Z_1,\zeta}(z_{1i},\zeta_i)dF_{Z_1,\zeta}(z_{1j},\zeta_j)\right)h_n^2 \tag{A.14}$$

where C is a finite constant. This term is $O\left(\theta'h_n^2\right)$, which is $o(n^{-1})$ for θ in a $O(n^{-1/2})$ neighborhood of θ_0 by the assumptions on h_n which imply $\sqrt{n}h_n^2 \to 0$.

Next, we establish a representation for $E[G_n(\theta)|z_{1i}, y_{1i}, z_i]$. Using the same arguments as in the unconditional expectation, we conclude that $E[G_n(\theta)|z_{1i}, y_{1i}, z_i]$ is of the above form, now no longer integrating over the variables z_{1i}, y_{1i}, z_i.

Next, we derive a linear representation for (A.8). Regarding the term $\Delta z_{ij}'\hat{\delta}$, we will only derive the linear representation involving the component $\hat{\zeta}_i - \zeta_i$ as the term involving $\hat{\zeta}_j - \zeta_j$ can be dealt with similarly. The first step is to plug in a linear representation for the estimator $\hat{\delta}$ of δ_0:

$$\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k} h^{-1}k_h'((z_i - z_j)'\delta_0)1[y_{1i} > y_{1j}]z_i'\psi_{\delta k}1[z_{1i}'\beta > z_{1j}'\beta] \tag{A.15}$$

where $\psi_{\delta k}$ denotes the influence function in the linear representation of the first stage estimator $\hat{\delta}$, evaluated at the k-th observation. We have a centered third order U-process. Again,

we note the its unconditional mean is 0, as is its mean conditional on each of its first two arguments. Consequently, we derive a linear representation for its mean conditional on its third argument:

$$\frac{1}{n}\sum_{k=1}^{n}\left(\int \mathcal{H}(\zeta_i, \zeta_j)\mathcal{F}(z'_{1i}\beta_0, z'_{1j}\beta_0, \zeta_i, \zeta_j)h^{-1}k'_h(\zeta_i - \zeta_j)1[z'_{1i}\beta > z'_{1j}\beta]z'_i\times\right.$$

$$\left. dF_Z(z_i)dF_Z(z_j)\right)\psi_{\delta_k} \tag{A.16}$$

While the above integral is expressed with respect to z_i, z_j, it will prove convenient to express the integral in terms of ζ_i, ζ_j. We do so as follows:

$$\int \mathcal{H}(\zeta_i, \zeta_j)\mathcal{M}(\zeta_i, \zeta_j, \beta)h^{-1}k'_h(\zeta_i - \zeta_j)f_\zeta(\zeta_i)f_\zeta(\zeta_j)d\zeta_i d\zeta_j \tag{A.17}$$

where f_ζ denotes the density function of ζ, and

$$\mathcal{M}(\zeta_i, \zeta_j, \beta) = E[\mathcal{F}(z'_{1i}\beta_0, z'_{1j}\beta_0, \zeta_i, \zeta_j)1[z'_{1i}\beta > z'_{1j}\beta]z'_i|\zeta_i, \zeta_j] \tag{A.18}$$

Now we do a change of variables in (A.17) $v = (\zeta_i - \zeta_j)/h$, noting that under our assumptions we have $\int k'(v)dv = 0$ and $\int k'(v)vdv = -1$ so that the lead term in the expansion (inside the integral) around $vh_n = 0$ yields the integral

$$\left(\int \mu_1(\zeta_j, \zeta_j, \beta)f_\zeta(\zeta_j)d\zeta_j\right) \tag{A.19}$$

where

$$\mu(t, \zeta, \beta) = \mathcal{H}(t, \zeta)\mathcal{M}(t, \zeta, \beta)f_\zeta(t) \tag{A.20}$$

and $\mu_1(\cdot, \cdot, \cdot)$ denotes its partial derivative with respect to its first argument. The remaining terms in the expansion are negligible—i.e. $o_p(n^{-1})$ uniformly in $O_p(n^{-1/2})$ neighborhoods of β around β_0.

The next step is to expand the function in (A.19) around $\beta = \beta_0$. A second-order expansion of $\mu_1(\cdot, \cdot, \beta)$ around $\beta = \beta_0$, (inside the above integral) yields the term

$$\frac{1}{n}\sum_{k=1}^{n}\left(\int \mu_{31}(\zeta_j, \zeta_j, \beta_0)f_\zeta(\zeta_j)d\zeta_j\right)(\psi_{\delta k})'(\beta - \beta_0) + R_n \tag{A.21}$$

where $\mu_{31}(\cdot, \cdot, \cdot)$ denotes the partial derivative of $\mu_1(\cdot, \cdot, \cdot)$ with respect to its third argument, and the remainder term R_n is $o_p(n^{-1})$ uniformly in β in $O_p(n^{-1/2})$ neighborhoods of β_0. This concludes the linear representation of the term in the objective function involving $\hat{\zeta}_i - \zeta_i$. We note analogous arguments can be used to derive the analogous linear representation for the term involving $\hat{\zeta}_j - \zeta_j$ This concludes our proof that our estimator $\hat{\beta}$ of β_0 is root-n consistent and asymptotically normal. We will use these results in deriving the asymptotic properties of our test statistic.

B Proof of Theorem 4.1

The strategy is to derive a linear representation for $\hat{\tau} - \tau_0$, where

$$\tau_0 = E[sgn(y_{1i} - y_{1j})sgn(z_i'\delta_0 - z_j'\delta_0)|z_{1i}'\beta_0 - z_{1j}'\beta_0 = 0].$$

Note that

$$\hat{\tau} - \tau_0 = \left(\frac{1}{n(n-1)}\sum_{i\neq j}\hat{\omega}_{ij}\right)^{-1}\left(\frac{1}{n(n-1)}\sum_{i\neq j}\hat{\omega}_{ij}(sgn(y_{1i} - y_{1j})sgn(z_i'\hat{\delta} - z_j'\hat{\delta}) - \tau_0)\right) \quad \text{(B.1)}$$

from which we will derive the probability limit of the denominator term, a linear representation of the numerator term, and then apply Slutsky's theorem.

For the denominator term in (B.1), a mean-value expansion around the true index difference, the root-n consistency of $\hat{\beta}$, and a LLN for U-statistics implies:

$$\frac{1}{n(n-1)}\sum_{i\neq j}\hat{w}_{ij} \xrightarrow{p} E[f_{Z\beta}(\zeta_{\beta i})] \quad \text{(B.2)}$$

where $f_{Z\beta}(\cdot)$ denotes the density function of $\zeta_{\beta i} = z_{1i}'\beta_0$.

For the numerator term in (B.1), we first work with the term

$$\frac{1}{n(n-1)}\sum_{i\neq j}\omega_{ij}(sgn(y_{1i} - y_{1j})sgn(z_i'\delta_0 - z_j'\delta_0) - \tau_0) \quad \text{(B.3)}$$

to which we apply a U-statistic projection theorem. By a change of variables and the higher order properties of the kernel function inside ω_{ij}, the expectation of the term inside the double summation is $o_p(n^{-1/2})$; therefore, it remains to derive expressions for conditional expectations of the term inside the double summation conditional on its first and second arguments. Again, using a change of variables and the higher order properties of the kernel function inside the weighting function, we get the following expression for these conditional expectations:

$$\frac{1}{n}\sum_{i=1}^{n}2f_{Z\beta}(\zeta_{\beta i})(G(y_{1i}, z_i, \zeta_{\beta i}) - \tau_0) + o_p(n^{-1/2}) \quad \text{(B.4)}$$

where

$$G(y_1, z, \zeta_\beta) = E[sgn(y_{1i} - y_1)sgn(z_i'\delta_0 - z'\delta_0)|\zeta_{\beta i} = \zeta_\beta]. \quad \text{(B.5)}$$

Therefore, by the projection theorem (Powell et al. (1989)) (which is applicable due to the properties of the kernel function and bandwidth, which imply that the variance of the statistic is $o(n)$), (B.3) can be represented as (B.4).

Turning attention to the linear term in the expansion of $\hat{\omega}_{ij}$ around ω_{ij}

$$\frac{1}{n(n-1)} \sum_{i \neq j} h^{-1} k'_h(\zeta_i - \zeta_j)(z_{1i} - z_{1j})'(\hat{\beta} - \beta_0)(sgn(y_{1i} - y_{1j})sgn(z'_i\delta_0 - z'_j\delta_0) - \tau_0), \quad \text{(B.6)}$$

we can plug in the derived linear representation for $\hat{\beta} - \beta_0$, yielding a third order U-statistic plus a negligible remainder term. The U-statistic is of the form:

$$\frac{h_n^{-1}}{n(n-1)(n-2)} \sum_{i \neq j \neq k} k'_h(\zeta_{\beta i} - \zeta_{\beta j})(z_{1i} - z_{1j})'\psi_{\beta i}(sgn(y_{1i} - y_{1j})sgn(z'_i\delta_0 - z'_j\delta_0) - \tau_0). \quad \text{(B.7)}$$

We note the unconditional expectation of the above term is 0, as is the expectation conditional on each of its first two arguments. Using similar arguments as before, it follows that the expectation conditional on its third argument can be expressed as:

$$\frac{1}{n} \sum_{i=1}^{n} E[G'_x(\zeta_{\beta i}) f_{Z\beta}(\zeta_{\beta i})]\psi_{\beta i} + o_p(n^{-1/2}) \quad \text{(B.8)}$$

where

$$G_x(\cdot) = E[(sgn(y_{1i} - y_{1j})sgn(z'_i\delta_0 - z'_j\delta_0) - \tau_0)(z_{1i} - z_{1j})'|z_{1i} - z_{1j} = \cdot]$$

and $G'_x(\cdot)$ denotes the derivative of $G_x(\cdot)$ with respect to its argument.

A further term in the linear representation of the test statistic is

$$\frac{1}{n(n-1)} \sum_{i \neq j} \omega_{ij} sgn(y_{1i} - y_{1j})(sgn((z_i - z_j)'\hat{\delta}) - sgn((z_i - z_j)'\delta_0)) \quad \text{(B.9)}$$

Here we can effectively expand the above term with $\hat{\delta}$ around δ_0. To do so, since the sign function is not differentiable, we take the expectation of $sgn((z_i - z_j)'\delta)$ for any δ. Recall, as a normalization, we set the last component of $\delta = 1$ and assume its associated regressor was continuously distributed. Here we let $F_{Z_k|Z_{-k}}(\cdot)$ denote the cdf of $z_{ki} - z_{kj}$ conditional on z'_{-ki}, z'_{-kj} (where z_{ki} denotes the last component of z_i and z'_{-ki} denotes the other components of z_i). Consequently, we have:

$$E[sgn((z_i - z_j)'\delta)|z_{-ki}, z_{-kj}] = F_{Z_k|Z_{-k}}(\Delta z'_{-kij}\delta^{(-k)})$$

where Δz_{-kij} denotes the difference in the corresponding components of z_i and z_j, and $\delta^{(-k)}$ denotes the subvector of δ corresponding to z_{-ki}. We can expand this conditional expectation evaluated at δ around the conditional expectation evaluated at δ_0:

$$F_{Z_k|Z_{-k}}(\Delta z'_{-kij}\hat{\delta}^{(-k)}) = F_{Z_k|Z_{-k}}(\Delta z'_{-kij}\delta_0^{(-k)})$$

$$+f_{Z_k|Z_{-k}}(\Delta z'_{-kij}\delta_0^{(-k)})\Delta z'_{-kij}(\hat{\delta}-\delta_0)+O_p(\|\hat{\delta}-\delta_0\|^2)$$

where $f_{Z_k|Z_{-k}}(\cdot)$ denotes the density function of $z_{ki} - z_{kj}$ conditional on z'_{-ki}, z'_{-kj}. Note that since the *sgn* function is Euclidean and $\hat{\delta}-\delta_0$ is $O_p(n^{-1/2})$ (by using, e.g., Theorem 1 in Sherman (1994)), the remainder term that arises from replacing the difference in *sgn* functions with their expectations in (B.9) is $o_p(n^{-1/2})$.

Next, by applying the same arguments as before, involving plugging in a linear representation (this time for $(\hat{\delta}-\delta_0)$), and decomposing the resulting third order U-statistic, we get a linear representation. Specifically, let

$$G_{11}(\cdot) = E[sgn(y_{1i} - y_{1j})f_{Z_k|Z_{-k}}(\Delta z'_{-kij}\delta_0^{(-k)})\Delta z'_{-kij}|\zeta_{\beta i} = \cdot, \zeta_{\beta j} = \cdot].$$

Then, we may conclude that (B.9) has the following linear representation

$$\frac{1}{n}\sum_{i=1}^{n}E[G_{11}(\zeta_{\beta i})f_{Z\beta}(\zeta_{\beta i})]\psi_{\delta i} + o_p(n^{-1/2}). \tag{B.10}$$

Finally, we note that is easy to show that the remainder term, which involves the product of $\hat{\beta}-\beta_0$ and $\hat{\delta}-\delta_0$, is $o_p(n^{-1/2})$.

Therefore, collecting all our results we may conclude that

$$\hat{\tau} - \tau_0 = E[f_{Z\beta}(\zeta_{\beta i})]^{-1}\frac{1}{n}\sum_{i=1}^{n}\left(2f_{Z\beta}(\zeta_{\beta i})(G(y_{1i}, z_i, \zeta_{\beta i}) - \tau_0) + E[G'_x(\zeta_{\beta i})f_{Z\beta}(\zeta_{\beta i})]\psi_{\beta i} \right. \tag{B.11}$$

$$\left. + E[G_{11}(\zeta_{\beta i})f_{Z\beta}(\zeta_{\beta i})]\psi_{\delta i}\right) + o_p(n^{-1/2}) \tag{B.12}$$

which establishes the theorem.

Table 1: Monte Carlo simulation results for $n = 500$. Rejection rates (over 1,000 simulations) for tests at the 5% and 10% levels are reported. The three different z-tests are based upon MLE estimation, IV estimation, and rank correlation.

	ρ	α	5%-level rejections			10%-level rejections		
			MLE	IV	$\hat{\tau}$	MLE	IV	$\hat{\tau}$
Size of the test	0.00	0.0	0.064	0.052	0.056	0.110	0.095	0.105
	0.25	0.0	0.051	0.041	0.047	0.104	0.090	0.094
	0.50	0.0	0.059	0.051	0.051	0.116	0.092	0.096
	0.75	0.0	0.081	0.052	0.053	0.127	0.097	0.097
Power of the test	0.00	0.1	0.076	0.075	0.072	0.143	0.140	0.136
	0.25	0.1	0.090	0.087	0.097	0.160	0.157	0.156
	0.50	0.1	0.083	0.099	0.085	0.135	0.148	0.138
	0.75	0.1	0.129	0.119	0.118	0.205	0.185	0.182
	0.00	0.2	0.169	0.163	0.161	0.255	0.250	0.265
	0.25	0.2	0.187	0.193	0.189	0.286	0.288	0.295
	0.50	0.2	0.197	0.209	0.199	0.292	0.300	0.286
	0.75	0.2	0.269	0.262	0.245	0.393	0.364	0.353
	0.00	0.3	0.343	0.331	0.345	0.483	0.457	0.470
	0.25	0.3	0.315	0.318	0.326	0.434	0.439	0.437
	0.50	0.3	0.370	0.386	0.387	0.500	0.487	0.495
	0.75	0.3	0.470	0.454	0.430	0.588	0.548	0.556

Table 2: Summary statistics

Variable	Description	Mean (Stdev)
Worked in 1999	1 if worked for pay in 1999, 0 otherwise	0.697
Same-sex indicator	1 if first two children are the same gender, 0 otherwise	0.502
Had third child	1 if had third child, 0 otherwise	0.257
Age at first birth	Mother's age when first child was born	26.36 (5.03)
1st child's age	Age of first child in 2000	7.55 (3.03)
Education	Mother's education level (in years)	10.97 (2.19)

Table 3: First-stage regression results. The dependent variable is an indicator variable equal to one if the woman had a third child. Heteroskedasticity-robust standard errors are reported for the OLS estimates. Marginal effects (evaluated at the means of the explanatory variables) for the probit estimates are provided in brackets.

	OLS	OLS	Probit
Same-sex indicator	0.0564	0.0562	0.1865
	(0.0016)	(0.0015)	(0.0052)
			[0.0576]
Age at first birth		-0.0147	-0.0511
		(0.0002)	(0.0006)
			[-0.0158]
1st child's age		0.0318	0.1111
		(0.0002)	(0.0009)
			[0.0344]
Education		0.0104	0.0364
		(0.0004)	(0.0013)
			[0.0113]
# observations	293,771	293,771	293,771
R-squared	0.0042	0.0812	

Table 4: Second-stage regression results. The dependent variable is an indicator variable equal to one if the woman worked for pay in 1999. The 2SLS regressions use the same-sex indicator variable as an instrument for the had-third-child indicator variable. Heteroskedasticity-robust standard errors are reported.

	OLS	2SLS (Wald)	OLS	2SLS
Had third child	-0.1382	-0.1118	-0.1728	-0.1124
	(0.0020)	(0.0298)	(0.0020)	(0.0295)
Age at first birth			-0.0074	-0.0065
			(0.0002)	(0.0005)
1st child's age			0.0161	0.0142
			(0.0003)	(0.0010)
Education			0.0319	0.0313
			(0.0004)	(0.0005)
# observations	293,771	293,771	293,771	293,771
R-squared	0.0173	0.0167	0.0462	0.0432

Table 5: Testing significance of the binary endogenous regressor. The z-statistics for the semiparametric and 2SLS estimation approaches are reported for several different model specifications. The 2SLS standard errors are heteroskedasticity-robust.

Exogenous covariates	Semiparametric			2SLS		
in the model	$\hat{\tau}$	s.e.	z-stat	$\hat{\alpha}$	s.e.	z-stat
None	-0.00316	0.00085	-3.72	-0.1118	0.0298	-3.75
Education	-0.00299	0.00102	-2.94	-0.1103	0.0296	-3.72
Education, Mother's age at first birth	-0.00655	0.00229	-2.86	-0.1111	0.0296	-3.75
Education, Mother's age at first birth, Age of 1st child	-0.00695	0.00258	-2.69	-0.1124	0.0295	-3.80