MIT Open Access Articles

Testing for causal e ffects in a generalized regression model with endogenous regressors

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

Jason Abrevaya[†]

Jerry A. Hausman[‡]

Shakeeb Khan[§]

This version: November 2009

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 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.

[†]The University of Texas at Austin, Department of Economics, `abrevaya@eco.utexas.edu`

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

[§]Duke University, Department of Economics, `shakeebk@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., Heckman (1978), Smith and Blundell (1986), and Rivers and Vuong (1988)). A more recent literature in econometrics has developed methods that do not require parametric distributional 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) for many nonlinear models (see also Aradillas-Lopez, Honoré, and Powell (2007) and, without linear-index and separability restrictions, Imbens and Newey (forthcoming)). In Blundell and Powell (2003, 2004), a reduced-form 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 of Blundell and Powell (2003, 2004), however, requires the endogenous regressors to be continuously distributed. For the endogenous regressors, this restriction is necessary to identify the average structural function and its derivatives (i.e., the structural effects).[3]

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 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

---

[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]See also, for example in linear models, Telser (1964) and Dhrymes (1970).

[3]On the other hand, following their analysis, in certain cases one could recover the sign of the structural effect(s) without the support condition. However, there are relevant cases, such as a binary endogenous variable, for which their method is unable to recover even the sign of the effect.

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,[4] 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.[5]

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, 2004), 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.

The work proposed here is also related to a recent literature on bounding causal effects in models with a binary endogenous variable. See, for example, Bhattacharya, Shaikh, and Vytlacil (2005), Shaikh and Vytlacil (2005), and Chiburis (2008). These studies focus upon partial identification of causal-effect parameters (such as the average treatment effect). Chesher (2005) concludes that point identification of these parameters is generally not possible without additional assumptions. Therefore, it is natural to focus directly upon the sign (rather than the magnitude) of the causal effect. In addition, while the aforementioned studies imply identification of the sign in specific settings, this paper will focus upon such identification in a more general model that allows for additional covariates, potential non-linearities, and non-discrete endogenous regressors. The generality of our model is also in contrast to related studies in the biostatistics literature, where tests for the significance of a treatment effect with time-to-event data have been proposed (e.g., Mantel (1966), Peto and Peto (1972), Prentice (1978), Yang and Zhao (2007), and Yang and Prentice (2005)).

---

[4]This support condition pertains only to index and parameter values and not the unobserved error terms. Thus, this result is distinct from previous "identification at infinity" results.

[5]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 not require unbounded support of $z_1'\beta_0$. Our assumption RD later on will impose an unbounded support condition, but this is only for convenience. See Khan (2001) on how to relax this condition in the context of rank estimation, and note that such a condition is not required when alternative estimators are used for binary choice models (e.g. Ahn, Ichimura and Powell (2004)).

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 third stage of this procedure computes the statistic of interest, which turns out to be a kernel-weighted version of the *tau* statistic of Kendall (1938). Since the (scalar) statistic is $\sqrt{n}$-consistent and asymptotically normal, the proposed test for statistical significance of the causal effect is simply a z-test. Section 4 describes the causal-effect testing approach for the general version of the model, which allows for multiple endogenous regressors (with an arbitrary mix of discrete, continuous, and possible censored endogenous regressors). Section 5 provides a brief empirical illustration of the methodology, 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. In the interest of space, Monte Carlo simulations, additional details on the empirical application, and also some of the asymptotic proofs have been provided as on-line "Supplements" to this note.

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. (The general treatment of multiple endogenous regressors with a mix of continuous/discrete covariates is considered in Section 4.) 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 and F is a possibly unknown function that is assumed to be strictly monotonic (without loss of generality, strictly increasing) in its first and third arguments and weakly monotonic in its second argument. For example, strict monotonicity for the first argument is

$$v' > v'' \implies F(v', y, e) > F(v'', y, e) \text{ for all } y, e$$

and similarly for the third argument. The direction of the monotonicity with respect to the second argument is assumed to be invariant to the values of the first and third arguments. That is, either

$$y_2' > y_2'' \implies F(v, y_2', e) \geq F(v, y_2'', e) \text{ for all } v, e$$

or

$$y_2' > y_2'' \implies F(v, y_2', e) \leq F(v, y_2'', e) \text{ for all } v, e.$$

3

The observed dependent variable is y_1, where the function D is weakly increasing and non-degenerate (i.e., strictly increasing on some region of its argument). 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.[6] 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 (2004) and Yildiz (2006) but allows for other forms of separability.[7] 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

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

with strict inequality on some region of the support of y_2. Similarly, a negative effect of y_2 upon y_1 is equivalent to

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

with strict inequality on some region of the support of y_2. 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.[8] 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

---

[6]Hence, we impose the same monotonicity conditions as Han (1987), noting that strict monotonicity of the third argument ensures that the support of y_1 does not depend on β_0. Under further restrictions on the support of exogenous regressors (e.g., a component with positive density on the real line), the strict monotonicity condition on $F(\cdot, \cdot, \cdot)$ with respect to its first argument may be relaxed to weak monotonicity as long as non-degeneracy of $D(\cdot)$ is preserved. An alternative specification to (2.1), which explicitly separates the strictly and weakly monotonic relationships, is $y_1^* = F(z_1'\beta_0, \epsilon)$, $y_1 = D(y_1^*, y_2)$, where $F(\cdot, \cdot)$ is strictly monotonic in each of its arguments for all values of the other and $D(\cdot, \cdot)$ is weakly monotonic (but non-degenerate) in each of its arguments for all values of the other.

[7]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).

[8]There is also a tradition in the biostatistics literature to focus upon testing the null of no treatment effect; see, for example, Rosenbaum (2002) and the references cited in the Introduction.

straightforward to apply the testing component of this paper (i.e., testing H_0 above) to different regions of the covariate space.

Turning to the model for the endogenous regressor, we first focus on the case of a single binary endogenous regressor in order to simplify exposition. The binary endogenous variable y_2 is assumed to be determined by the following reduced-form model:

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

where $z \equiv (z_1, z_2)$ is the vector of "instruments" and η is an error disturbance. The z_2 subcomponent of z provides the exclusion restrictions in the model. z_2 will only required to be nondegenerate conditional on $z_1'\beta_0$, which is a particularly weak condition. We assume that (ϵ, η) is independent of z. 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 $(\beta_0, \alpha_0, \delta_0)$ (Heckman (1978)).[9] 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, the treatment-effect statistic is calculated. Each of the three stages is described in turn below.

Stage 1: Estimation of δ_0

---

[9]Another common estimation approach is to simply ignore the non-linearity in (2.6) and apply 2SLS to the system given by (2.6) and (2.5).

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).[10] 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.

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

$$\Pr(\epsilon \le e \,|\, y_2, z) = \begin{cases} \Pr(\epsilon \le e \,|\, \eta \le -z'\delta_0) & \text{if } y_2 = 0 \\ \Pr(\epsilon \le e \,|\, \eta > -z'\delta_0) & \text{if } y_2 = 1 \end{cases} \tag{3.1}$$

If two observations (indexed i and j) have $y_{2i} = y_{2j}$ and $z_i'\delta_0 = z_j'\delta_0$, equation (3.1) 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 strict monotonicity of F with respect to its first and third arguments implies that

$$z_{1i}'\beta_0 \ge z_{1j}'\beta_0 \iff \tag{3.2}$$
$$\Pr(y_{1i} > y_{1j} \,|\, z_{1i}, z_{1j}, y_{2i} = y_{2j}, z_i'\delta_0 = z_j'\delta_0) \ge \Pr(y_{1i} < y_{1j} \,|\, z_{1i}, z_{1j}, y_{2i} = y_{2j}, z_i'\delta_0 = z_j'\delta_0).$$

Equation (3.2) forms the basis for the proposed estimator of β_0. Unfortunately, equation (3.2) 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,[11] note that equation (3.2) 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:

$$\hat{\beta} \equiv \arg\max_{\beta \in \mathcal{B}} \frac{1}{n(n-1)} \sum_{i \ne j} 1[y_{2i} = y_{2j}] k_h(z_i'\hat{\delta} - z_j'\hat{\delta}) 1[y_{1i} > y_{1j}] 1[z_{1i}'\beta > z_{1j}'\beta], \tag{3.3}$$

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 which $z_i'\hat{\delta}$ is close to $z_j'\hat{\delta}$.[12] Under appropriate regularity assumptions, it can be shown that $\hat{\beta}$ is a $\sqrt{n}$-consistent estimator of β_0 (see Appendix).

---

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

[11]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.

[12]For a binary endogenous regressor, the weighting is analogous to propensity-score matching.

Stage 3: Testing for a treatment effect

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

$$y_1 = D(F(z_1'\beta_0, 1(z'\delta_0 + \eta > 0), \epsilon)). \tag{3.4}$$

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) with $z_{1i}'\beta_0 = z_{1j}'\beta_0$, (3.4) implies

$$z_i'\delta_0 \geq z_j'\delta_0 \iff \Pr(y_{1i} > y_{1j} \mid z_i, z_j, z_{1i}'\beta_0 = z_{1j}'\beta_0) \geq \Pr(y_{1i} < y_{1j} \mid z_i, z_j, z_{1i}'\beta_0 = z_{1j}'\beta_0) \tag{3.5}$$

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

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

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

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

since y_{1i}^* and y_{1j}^* are identically distributed if $z_{1i}'\beta_0 = z_{1j}'\beta_0$. The ability to find statistical evidence against the null of no treatment effect (equation (3.7)) will require that the inequality in (3.5) (or (3.6)) is strict in some region.

Unlike equation (3.2), these probability statements do *not* condition on y_2. In fact, the proposed statistic below does not directly use the y_2 values. This feature is somewhat analogous to the second stage of 2SLS, where endogenous regressors are not directly used in the regression but rather their "fitted values" (projections onto the exogenous regressors) are included. In our context, the y_2 values play a role in estimation of δ_0 and β_0. Unlike 2SLS, however, 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 statistic, which is a kernel-weighted version of Kendall's tau:[13]

$$\hat{\tau} \equiv \frac{\sum_{i \neq j} \hat{\omega}_{ij} sgn(y_{1i} - y_{1j}) sgn(z_i'\hat{\delta} - z_j'\hat{\delta})}{\sum_{i \neq j} \hat{\omega}_{ij}}, \tag{3.8}$$

---

[13]In the context of a binary endogenous regressor, note that the sign of $(z_i'\hat{\delta} - z_j'\hat{\delta})$ is simply the sign of the difference in propensity scores for the pair of observations.

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

$$\hat{\omega}_{ij} \equiv k_h(z_{1i}'\hat{\beta} - z_{1j}'\hat{\beta}). \tag{3.9}$$

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

$$\tau_0 \equiv E[sgn(y_{1i} - y_{1j})sgn(z_i'\delta_0 - z_j'\delta_0)|z_{1i}'\beta_0 = z_{1j}'\beta_0]. \tag{3.10}$$

Based upon (3.5)–(3.7), 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.

Remark 3.1 *The testing approach does not require the index structure of equations (2.1) and (2.5). A nonparametric version of the model in (2.1) would be of the form $y_1^* = F(z_1, y_2, \epsilon)$. Then, the statistic described in this section could match on all components of the z_1 vector, which would be attractive when some (or all) of its components are discrete. Moreover, the linear-index restriction in (2.5) is unnecessary; a non-parametric specification (e.g., a non-parametric propensity score in the binary-endogenous variable case) could be used instead.*

Remark 3.2 *A comparison of our proposed test procedure with the standard 2SLS approach is warranted. As a referee correctly pointed out, while the 2SLS coefficient on the endogenous variable might not get the right magnitude of any particular parameter of interest, it could be getting the right sign given the monotonicity conditions of the model. For the case of binary y_1 and y_2 and no exogenous regressors (empty z_1), Shaikh and Vylacil (2005) and Bhattacharya, Shaikh, and Vytlacil (2005) show that the probability limit of 2SLS identifies the correct sign of the average treatment effect when the outcome is assumed to be weakly monotonic in the treatment.[14] Unfortunately, there are reasons to question the validity of a 2SLS-based test in more general settings. Specifically, in a non-linear model (say, a probit model) with non-empty z_1 and no causal effect of y_2 on y_1, the probability limit of the 2SLS endogenous-variable coefficient will generally be non-zero; when this is the case, simply looking at the sign of the 2SLS coefficient introduces Type-I error whose probability converges to one.*

---

[14]Shaikh and Vytlacil (2005) generalize this result to the case with covariates (non-empty z_1) by conditioning upon the covariates. This approach can be interpreted as a conditional version of 2SLS, but is distinct from a standard 2SLS regression in which z_1 and y_2 are explicitly included as right-hand-side variables. It is an open question whether or not their results extend to situations with non-binary outcomes, non-binary endogenous regressors, and/or multiple endogenous regressors.

Remark 3.3 *Interestingly, even in the case when the sign of the treatment effect can vary across individuals (contrary to our maintained assumptions), τ_0 can represent an interesting parameter. For example, in the case without z_1, τ_0 is the rank correlation between y_1 and the treatment probability (propensity score).[15] This correlation has the usual advantages when compared to other measures, e.g. linear correlation, such as being more robust to outliers, which is especially important for the generalization to non-binary variables.*

Remark 3.4 *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 Abrevaya and Jiang (2005), who develop consistent tests in similar U-statistic frameworks.*

Remark 3.5 *No consideration has been given to the efficiency of the various estimators discussed above. As a referee pointed out, in principle, more efficient estimates of $(\hat{\delta}_0, \hat{\beta}_0, \hat{\tau}_0)$ could be obtained through the use of some joint estimation technique (e.g., a version of joint GMM).*

The (scalar) 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. The theoretical assumptions required for this result are provided in the Appendix, as is the formal statement of the asymptotic properties (Theorem 1); proofs are provided in the Supplement. 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 form of the asymptotic variances in Theorem 1 is somewhat complicated.[16] Furthermore, estimating the components of the analytical asymptotic variance matrix would require the choice of additional smoothing parameters.

4 The general case

This section presents the general version of the model for which the rank-based testing procedure can be applied. The model allows for multiple endogenous regressors. A given endogenous regressor

---

[15]More generally, τ_0 has an interpretation as a conditional rank correlation. In the case of a binary (non-binary) endogenous variable, τ_0 is the rank correlation between the outcome y_1 and the propensity score (reduced-form index $z'\delta_0$) conditional on having an identical structural index $z_1'\beta_0$.

[16]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 $\hat{\tau}$).

may be discrete, continuous, or even censored in some way. The endogenous regressors are denoted $y_{21}, y_{22}, \ldots, y_{2Q}$, where Q is the number of endogenous regressors. Let $Q_C \leq Q$ denote the number of non-discrete 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:

$$y_{2q}^* = F_{2q}(z'\delta_{0q}, \eta_q), \ y_{2q} = D_{2q}(y_{2q}^*). \tag{4.1}$$

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. Similar to the model for the generalized regression for y_1, we assume that (for $q = 1, \ldots, Q$) F_{2q} is strictly increasing in each of its two arguments and D_{2q} is weakly increasing and non-degenerate (i.e., strictly increasing on some region of its argument).

To simplify notation, 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 3 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

$$\frac{1}{n(n-1)} \sum_{i \neq j} K_h(\mathbf{y}_{2i} - \mathbf{y}_{2j}) K_h(\widehat{\Delta} z_i - \widehat{\Delta} z_j) 1[y_{1i} > y_{1j}] 1[z_{1i}'\beta > z_{1j}'\beta]. \tag{4.2}$$

where $K_h(\cdot)$ is a multivariate kernel function defined as $K_h(\mathbf{v}) \equiv \prod_{q=1}^{\dim(\mathbf{v})} k_{hq}(v_q)$ for a vector $\mathbf{v}$. Kernel-weighting functions $k_{hq}(\cdot)$ are used for non-discrete components of $\mathbf{y}_2$, whereas exact matching (i.e., $k_{hq}(v) = 1(v = 0)$) is used for discrete components of $\mathbf{y}_2$.

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 statistic associated with the q-th endogenous regressor is thus given by:

$$\hat{\tau}_q \equiv \frac{\sum_{i \neq j} \hat{\omega}_{ij,q} sgn(y_{1i} - y_{1j}) sgn(z_i'\hat{\delta}_q - z_j'\hat{\delta}_q)}{\sum_{i \neq j} \hat{\omega}_{ij,q}}, \tag{4.3}$$

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

$$\hat{\omega}_{ij,q} \equiv k_h(z_{1i}'\hat{\beta} - z_{1j}'\hat{\beta}) K_h(\widehat{\Delta}_{-q} z_i - \widehat{\Delta}_{-q} z_j). \tag{4.4}$$

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 on the

10

bandwidth sequence used in matching variables and the order of smoothness assumed on certain density and conditional expectation functions.

An example: two binary endogenous regressors

Consider the following model with two binary endogenous regressors y_{21} and y_{22}:

$$y_1^* = F(z_1'\beta_0, y_{21}, y_{22}, \epsilon), \ y_1 = D(y_1^*) \tag{4.5}$$

$$y_{21} = 1[z'\delta_{01} + \eta_1 > 0], \ y_{22} = 1[z'\delta_{02} + \eta_2 > 0] \tag{4.6}$$

Given estimators for δ_{01}, δ_{02}, and β_0, one tests the effect of y_{21} (second argument) upon y_1 by fixing $z_1'\beta$ and $z'\delta_{02}$ and examine the significance of the relationship between y_1 and $z'\delta_{01}$. This idea can be operationalized with the following kernel-weighted rank-based statistic:

$$\hat{\tau}_1 \equiv \frac{\sum_{i \neq j} \hat{\omega}_{ij,1} sgn(y_{1i} - y_{1j}) sgn(z_i'\hat{\delta}_1 - z_j'\hat{\delta}_1)}{\sum_{i \neq j} \hat{\omega}_{ij,1}}, \tag{4.7}$$

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

$$\hat{\omega}_{ij,1} \equiv k_h(z_{1i}'\hat{\beta} - z_{1j}'\hat{\beta}) k_h(z_i'\hat{\delta}_2 - z_j'\hat{\delta}_2). \tag{4.8}$$

An analogous statistic (for testing the effect of y_{22} on y_1) can easily be constructed.

5 Empirical illustration

In this section, we apply our 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.

The sample for the current study is drawn from the 2000 Census data (5-percent public-use microdata sample (PUMS)). 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. We consider specifications in which education, mother's age at first birth, and age of first child enter as exogenous covariates (z_1). More complete details on the empirical application, including construction of the sample, descriptive statistics, and first-stage results, are reported in the Supplement to this note.

The primary results of interest relate to the conclusions from the causal-effect significance tests, which are reported in Table 1. The table compares results obtained from the semiparametric

$\hat{\tau}$ statistic with those obtained from the z-statistic based upon 2SLS estimates. The bootstrap was used in order to compute standard errors for $\hat{\tau}$. 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 one-by-one until the full set of three exogenous covariates are included. In the model with no exogenous covariates, the z-statistics associated with $\hat{\tau}$ and the 2SLS coefficient are extremely similar. 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 1 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. In addition, this illustration highlights the feasibility of the semiparametric approach even for a very large sample (n close to 300,000 here).

6   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 statistic converges at the parametric rate to a limiting normal distribution under the null hypothesis of no causal effect. 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. In addition, it would be of interest to improve the efficiency of the $\hat{\tau}$ estimator.

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, 2007, Pairwise difference estimation with nonparametric control variables, International Economic Review 48: 1119–1158.

Bhattacharya, Jay, Azeem Shaikh, and Edward Vytlacil, 2005, Treatment effect bounds: an application to Swan-Ganz catheterization, NBER working paper 11263.

Blundell, Richard W. and James L. Powell, 2003, Endogeneity in nonparametric and semiparametric regression 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).

Blundell, Richard W. and James L. Powell, 2004, Endogeneity in semiparametric binary response models, Review of Economic Studies 71: 655–679.

Chesher, Andrew, 2005, Nonparametric identification under discrete variation, Econometrica 73: 1525–1550.

Chiburis, Richard C., 2008, Semiparametric bounds on treatment effects, mimeo.

Dhrymes, Phoebus J., 1970, Econometrics: Statistical Foundations and Applications, Springer-Verlag.

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.

Heckman, James, 1978, Dummy endogenous variables in a simultaneous equation system, Econometrica 46, 931–959.

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

Imbens, Guido W. and Whitney K. Newey, forthcoming, Identification and estimation of triangular simultaneous equations models without additivity, Econometrica.

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.

Mantel, N., 1966, Evaluations of survival data and two new rank order statistics arising in its consideration, Cancer Chemotherapy Reports 50: 163-170.

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.

Peto, R. and J. Peto, 1972, Asymptotically efficient rank invariant test procedures with discussion, Journal of the Royal Statistical Society Series A 135: 185-206.

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.

Prentice, R. L., 1978, Linear rank tests with right censored data, Biometrika 65: 167-179.

Rosenbaum, Paul R., 2002, Observational studies, Springer Series in Statistics.

Shaikh, Azeem M. and Edward Vytlacil, 2005, Threshold crossing models and bounds on treatment effects: a nonparametric analysis, NBER working paper T307.

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, U-processes in the analysis of a generalized semiparametric regression estimator, Econometric Theory 10: 372–395.

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

Telser, Lester G., 1964, Iterative estimation of a set of linear regression equations, Journal of the American Statistical Association 59: 845–862.

Vytlacil, Edward and Nese Yildiz, 2007, Dummy endogenous variables in weakly separable models, Econometrica 75: 757–779.

Yang, S. and R. L. Prentice, 2005, Semiparametric analysis of short term and long term hazard ratios with two sample survival data, Biometrika 92: 1-17.

Yang, S. and Y. Zhao, 2007, Testing treatment effect by combining weighted log-rank tests and using empirical likelihood, Statistics and probability letters 77: 1385–1393.

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

Appendix

In this appendix, we outline the asymptotic theory for the three-stage testing procedure. We state the main asymptotic-normality results and also explicitly state sufficient regularity conditions for these results. The proofs, which are somewhat standard given previous results in this literature, are provided in the Supplement.

The following linear representation of the first-step estimator is assumed:

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

where $\psi_{\delta i}$ is an influence-function term with zero mean and finite variance. This representation exists for the available $\sqrt{n}$-consistent semiparametric estimators. We do not specify a particular form for the influence-function term $\psi_{\delta i}$ since it will depend upon the particular estimator chosen.

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 normalize its last component to 1 and denote its other components by θ_0 and the corresponding estimator by $\hat{\theta}$, where

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

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 above. 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 stages kernel function) The kernel function $k(\cdot)$ used in the second stage and the third 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 stage and the third stage satisfies $\sqrt{n} h_n^p \to 0$ and $\sqrt{n} h_n^3 \to \infty$.

Assumption RD (Last regressor and index properties) $z_{1i}^{(k)}$ is continuously distributed with positive density on the real line conditional on $z_i'\delta_0$ and all other elements of z_{1i}. Moreover, $z_i'\delta_0$ is nondegenerate conditional on $z_{1i}'\beta_0$.

Assumption ED (Error distribution) (ϵ_i, η_i) is distributed independently of z_i and is continuously distributed with positive density on $\mathbf{R}^2$.

Assumption FR (Full rank condition) Conditional on $(z_i'\delta_0, y_{2i})$, the support of z_{1i} does not lie in a proper linear subspace of R^k.

The following lemma 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 1 *If Assumptions CPS, FS, K, H, RD, ED, and FR hold, then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, V^{-1}\Omega V^{-1}) \tag{A.3}$$

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

$$\hat{\theta} - \theta_0 = \frac{1}{n} \sum_{i=1}^{n} \psi_{\beta i} + o_p(n^{-1/2}) \tag{A.4}$$

15

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

$$\mathcal{N}(\theta) = \int 1[z'_{1i}\beta(\theta) > z'_{1j}\beta(\theta)]\mathcal{H}(\zeta_i, \zeta_j)\mathcal{F}(z'_{1i}\beta_0, z'_{1j}\beta_0, \zeta_i, \zeta_j)dF_{Z_1,\zeta}(z_{1i}, \zeta_i)dF_{Z_1,\zeta}(z_{1j}, \zeta_j) \tag{A.5}$$

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

$$\mathcal{F}(z'_{1i}\beta_0, z'_{1j}\beta_0, \zeta_i, \zeta_j) = P(y_{1i} > y_{1j}|y_{2i} = y_{2j}, z_{1i}, z_{1j}, \zeta_i, \zeta_j) \tag{A.6}$$

$$\mathcal{H}(\zeta_i, \zeta_j) = P(y_{2i} = y_{2j}|\zeta_i, \zeta_j) \tag{A.7}$$

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

$$\delta_{1i} = \left(\int f_\zeta(\zeta_i)\mu_{31}(\zeta_i, \zeta_i, \beta_0)d\zeta_i \right) \psi_{\delta i} \tag{A.8}$$

where

$$\mu(t, \zeta, \beta) = \mathcal{H}(t, \zeta)\mathcal{M}(t, \zeta, \beta)f_\zeta(t) \tag{A.9}$$

with

$$\mathcal{M}(t, \zeta, \beta) = E[\mathcal{F}(z'_{1i}\beta_0, z'_{1j}\beta_0, \zeta_i, \zeta_j)1[z'_{1i}\beta > z'_{1j}\beta]z'_i \mid \zeta_i = t, \zeta_j = \zeta] \tag{A.10}$$

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.

Although the particular expressions for V and Ω are quite involved, note that V represents the second derivative of the limit of the expectation of the maximand and Ω represents the variance of the limit of its projection.

The asymptotic theory for the third-stage 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:

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

$$G_x(\cdot) = E[(sgn(y_{1i} - y_{1j})sgn(z'_i\delta_0 - z'_j\delta_0) - \tau_0)(z_{1i} - z_{1j})'|z_{1i} - z_{1j} = \cdot] \tag{A.12}$$

where $f_{Z_k|Z_{-k}}()$ in (A.11) denotes the density function of the last component of $z_i - z_j$, conditional on its other components, and Δz_{-kij} denotes the difference for all the components of z_i except the last one, are all assumed to be all p times continuously differentiable with derivatives that are bounded on the support of $\zeta_{\beta i}$.

The main theorem establishes the asymptotic distribution of the statistic $\hat{\tau}$:

Theorem 1 *If Assumptions CPS, FS, K, H, RD, ED, FR, and S hold, then*

$$\sqrt{n}(\hat{\tau} - \tau_0) \Rightarrow N(0, V_2^{-2}\Omega_2) \tag{A.13}$$

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

$$\delta_{2i} = 2f_{\zeta\beta}(\zeta_{\beta i})G(y_{1i}, z_i, \zeta_{\beta i}) + E[G'_x(\zeta_{\beta i})f_{\zeta\beta}(\zeta_{\beta i})]\psi_{\beta i} + E[G_{11}(\zeta_{\beta i})f_{\zeta\beta}(\zeta_\beta i)]\psi_{\delta i}, \tag{A.14}$$

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

$$G(y_1, z, \zeta) = E[sgn(y_{1i} - y_1)sgn(z'_i\delta_0 - z'\delta_0)|\zeta_{\beta i} = \zeta]. \tag{A.15}$$

16

Table 1: Testing significance of the binary endogenous regressor (having a third child) in women's labor-force participation. 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 in the model	Semiparametric			2SLS		
	$\hat{\tau}$	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