Identification of Partial Effects with Bad/Endogenous Controls

Kaicheng Chen[*] Kyoo il Kim[†]

Feb 28, 2025

Abstract

Exogeneity of the treatment needed for identification are often achieved by conditioning. While control variables are explicitly or implicitly assumed to be exogenous, it is common to encounter endogenous controls in practice. It brings a dilemma: without controlling, the treatment may be endogenous; with controlling, the endogeneity of controls may pollute the identification. The problem is not solved with an instrumental variable when it is only conditionally valid and controls are endogenous. We provide an alternative identification for both cases under an extra measurable separability condition between the treatment and the controls. Noticeably, this condition permits the controls to be influenced by the treatment, effectively allowing for bad controls to some extent. The theoretical results apply to a wide class of models including linear, nonlinear, and non-separable models. Monte Carlo simulations exemplify this prevalent issue and demonstrate the performance of the proposed methods in finite sample.

Keywords: endogenous control, bad control, partial effects, local average response, non-separable model, triangular model.

JEL Classification Code: C14, C31

[*]Department of Economics, Michigan State University. Email: chenka19@msu.edu
[†]Department of Economics, Michigan State University. Email: kyookim@msu.edu

1 Introduction

In models with endogenous treatment, to obtain consistent estimates of treatment effects, researchers commonly impose conditional (mean) independence or use instrumental variables (IV) for the treatment while rather casually assuming other observable control variables are exogenous. In reality, however, empirical researchers often end up with control variables that may be subject to additional endogeneity concern while finding instruments for every endogenous control is challenging or impossible. In this note, we will demonstrate that if the objects of interest are limited to parameters associated with the treatment, then we can get around the endogeneity issue of the control variables in certain settings.

To illustrate the problem, let's consider the following linear model:

$$Y = D\tau + X\beta + \varepsilon, \;\; E[\varepsilon|D, X] = E[\varepsilon|X]$$

where D is a treatment variable of interest, X is another observable determinant of the outcome Y, and ε is an unobserved determinant. A common scenario where the endogenous control problem arises is that the treatment D is dependent on X while, at the same time, X is dependent onε, too. As a result, (1) without considering X in the specification, the dependence between D and X can cause bias in the OLS estimation; (2) with the presence of X, the dependence between X and ε can also pollute the estimation of the partial effects τ even if D is conditionally independent of ε. In either case (1) or (2), τ is not identified by the linear projection parameters, so the OLS estimator is biased. In case (1), it is simply an omitted variable bias. To see the bias in case (2), let $W = (D, X)$ and $\theta = (\tau', \beta')'$, the linear projection parameters are defined as follows:

$$\gamma_{LP} \equiv E[W'W]^{-1}E[W'Y] = \theta + E[W'W]^{-1}E\left[W'E[\varepsilon|X]\right].$$

Therefore, without further restriction on $E[\varepsilon|X]$, OLS does not produce consistent estimate for θ, or τ in particular. In a worse scenario, X may itself be an outcome of D, in which X is described as a bad control as in Angrist and Pischke (2009). It is well-known that a bad control can cause problem for identification and the problem is present even if we start with a randomly assigned treatment (see Wooldridge, 2005; Lechner, 2008).

However, as is shown below, τ can still be identified even when X is affected by D, as long as X is not sorely a function of D. More formally, this extra condition is referred to as the measurable separability, as first introduced in Florens et al. (1990) and will be defined formally later. At its essence, this assumption ensures that we can vary the value of $D = d$ while holding $X = x$ at a particular value of x. Note that this still allows the distribution of X to depend on D, and vice versa. In the case of a continuous random variable D, τ is nonparametrically identified as follows:

$$\partial_d E[Y|D = d, X = x] = \tau + \partial_d(x\beta + E[\varepsilon|D = d, X = x]) = \tau + \partial_d(x\beta + E[\varepsilon|X = x]) = \tau,$$

where the last result holds due to the measurable separability of D and X. To see this, suppose the measurable separability does not hold. For example, $X = f(D)$ almost surely and they are not constants, then conditioning on $D = d, X = x$ necessitates $D = d, X = x = f(d)$. In that case, we would have

$$\partial_d E[Y|D = d, X = x] = \tau + \partial_d(f(d)\beta + E[\varepsilon|X = f(d)]) \neq \tau.$$

In the case of a binary D, measurable separability between D and X allows for conditioning on $D = 1$ and $D = 0$ at different realized values of X. Combining with the conditional independence condition, the identification of τ is achieved as follows:

$$\begin{aligned} E[Y|D = 1, X = x] - E[Y|D = 0, X = x] &= \tau + E[\varepsilon|D = 1, X = x] - E[\varepsilon|D = 0, X = x] \\ &= \tau + E[\varepsilon|X = x] - E[\varepsilon|X = x] = \tau. \end{aligned}$$

Does the problem go away with an excludable instrumental variable? When the instrumental variable is truly exogenous and affect the outcome only through the treatment, the answer is yes because no control is needed. However, in practice, control variables are commonly considered in models with instrumental variables, sometimes due to the concern that the excludability condition is valid only after controlling certain observable variables. In those cases, again, the endogeneity in those controls can cause problem for identification. To illustrate the problem, let's consider the same outcome equation with an excludable instrumental variable in a linear triangular model:

$$\begin{aligned} Y &= D\tau + X\beta + \varepsilon, \\ D &= Z\pi_Z + X\pi_X + \eta. \end{aligned}$$

In this model, the instrumental variable Z is needed because D is not conditionally independent of ε even after conditioning on X and the nonparametric identification method introduced above is not valid anymore. However, without controlling X, the instrument Z itself may not suffice for identification because Z may affect Y through X too. Again, the endogeneity of X brings a dilemma: (1) without controlling X, Z is not a valid IV; (2) with the presence of X, neither τ or β is identified by usual *IV* or *2SLS* projection without further restriction on $E[\varepsilon|X]$.

Nevertheless, τ is nonparametrically identified given (i) the measurable separability between Z and X and (ii) $\pi_Z \neq 0$:

$$\begin{aligned}
E[Y|X=x, Z=z] &= (z\pi_Z + x\pi_X + E[\eta|X=x])\,\tau + x\beta + E[\varepsilon|X=x], \\
E[D|X=x, Z=z] &= z\pi_Z \ + x\pi_X + E[\eta|X=x], \\
\frac{\partial_z E[Y|X=x, Z=z]}{\partial_z E[D|X=x, Z=z]} &= \tau. \qquad (1)
\end{aligned}$$

Again, at its essence, the extra measurable separability condition ensures that we can vary the value of $Z = z$ while holding $X = x$. Alternatively, τ can also be nonparametrically identified through a control function approach. In this case, due to $D = Z\pi_Z + X\pi_X + \eta$ with $\pi_Z \neq 0$, the measurable separability between Z and X also implies the measurable separability between (X, η) and D. It will be shown later that given the exogeneity of Z, D is independent of ε conditional on (η, X). Therefore, τ can also be identified as follows:

$$\partial_d E[Y|D=d, X=x, \eta=e] = \partial_d \left[d\tau + E[\varepsilon|X=x, \eta=e]\right] = \tau. \qquad (2)$$

The heuristic exposition above focuses on linear models. When the true data generating process is nonlinear in the parameters or nonparametric, it is not clear whether the same idea is still applicable. Especially, when the unobserved determinant is not separable from other observable covariates, it is not clear whether the dependence between the controls and the unobserved determinant could largely change the exposition. Ideally, we would like our approach to be applicable to a large class of data generating processes. Therefore, we derive the main results under a nonparametric and nonseparable models, considering cases with or without instrumental variables.

The basic idea on how nonparametric estimation helps alleviate the bias due to endogenous controls is introduced in Frölich (2008). However, our work first provides a formal identification result, which not only justifies the use of nonparametric methods in the presence of endogenous control but also delineates the boundary of this method through the measurable separability condition.

The issue of endogenous controls is prevalent in empirical research but is not well studied in econometrics literature. One exception outside our setting is regarding the regression discontinuity (RD) design where Kim (2013) finds that endogenous control variables yield asymptotic bias in the RD estimator while the inclusion of these relevant controls may offset this bias and improve some higher-order properties of the estimator. Diegert et al. (2022) assess the omitted variable bias when the controls are potentially correlated with the omitted variables in a sensitivity analysis framework.

For the rest of the paper, it is outlined as follows: In Sections 2, we establish the main identification results for nonseparable models under conditional independence. In Section 3, we consider methods based on an instrumental variable that is only conditionally exogenous. In Section 4, Monte Carlo simulation demonstrate the issue of endogenous control and the performance of proposed methods in finite sample. Section 6 concludes the note with recommendations for empirical practice.

2 Nonseparable Models with Conditional Independence

Identification results for a nonseparable model with an endogenous treatment, $Y = m(D, \varepsilon)$, is given in Altonji and Matzkin (2005), assuming there exists some vector X such that conditional on X, the treatment variable D is independent of the stochastic error ε. However, in many empirical applications with either nonparametric, semiparametric, or parametric models, the vector of control variables usually appear in the model of the outcome Y.

The question is, would the endogeneity of X be a problem when we are interested in identifying, for example, LAR and ATE of D on Y given the conditional independence assumption? In this section, we show the answer is positive. To focus on our main point, for convenience, we assume all relevant (conditional) probability density functions are well defined below and also throughout the paper.

Consider a nonseparable nonparametric model as follows:

$$Y = m(D, X, \varepsilon) \tag{3}$$

We are interested in identifying conditional LAR ($CLAR$) and unconditional LAR, denoted by $\beta(d, x)$ and $\beta(d)$, respectively. For now, the focus is on a continuous outcome Y, but the results can be extended to binary choice models, as shown in Altonji and Matzkin (2005). Assume $m(.)$ is differentiable w.r.t its first argument and D is a continuous treatment, then

$\beta(d,x)$ and $\beta(d)$ are defined as:

$$\beta(d,x) = \int \frac{\partial m(d,x,\epsilon)}{\partial d} f_{\varepsilon|D=d,X=x}(\epsilon)d\epsilon,$$
$$\beta(d) = \int\int \frac{\partial m(d,x,\epsilon)}{\partial d} f_{X,\varepsilon|D=d}(x,\epsilon)dxd\epsilon,$$

where $f_{\varepsilon|D=d,X=x}(\epsilon)$ and $f_{X,\varepsilon|D=d}(x,\epsilon)$ denote relevant conditional density functions. If D is a binary random variable (or if we are interested in discrete change in D), we can define $CLAR$ and LAR as follows:

$$\tilde{\beta}(d,x) = \int (m(1,x,\epsilon) - m(0,x,\epsilon))\, f_{\varepsilon|D=d,X=x}(\epsilon)d\epsilon,$$
$$\tilde{\beta}(d) = \int\int (m(1,x,\epsilon) - m(0,x,\epsilon))\, f_{X,\varepsilon|D=d}(x,\epsilon)dxd\epsilon.$$

Assumption 1. $f_{\varepsilon|D,X}(\epsilon) = f_{\varepsilon|X}(\epsilon)$ *for all* $\epsilon \in \mathbb{R}$.

Assumption 1 is the conditional independence assumption which is also imposed in Altonji and Matzkin (2005). We note that Assumption 1 does not rule out X being endogenous (i.e. being not independent with ε).

If D is a continuous random variable, we can represent D by some function h as

$$D = h(X,U), \tag{4}$$

where X is independent of a continuous error term U and $h(X,u)$ is strictly monotonic in u almost surely (see Matzkin (2003)). To identify $CLAR$ and LAR in the continuous treatment case while allowing for endogeneity in X, we need an extra rank condition:

Assumption 2. *D and X are measurably separated, that is, any function of D almost surely equal to a function of X must be almost surely equal to a constant.*

To see why it is a type of rank condition, consider a case the condition is violated at some point in the interior of the support of (D,X), i.e. $l(D) = q(X)$ for some measurable functions $l(.)$ and $q(.)$. Then, $l(h(X,U)) = q(X)$. Differentiating both sides with respect to U, we have $\frac{\partial l}{\partial h}\frac{\partial h}{\partial U} = \frac{\partial q}{\partial U} = 0$. Given that the measurable separability fails, we have $\frac{\partial l}{\partial h} \neq 0$ and so $\frac{\partial h}{\partial U} = 0$. Therefore, Assumption 2 requires U to affect D. Following Theorem 3 in Florens et al. (2008), we give primitive conditions for Assumption 2 as follows:

Assumption 2′. *(i) D is determined by (4) where X is continuously distributed and independent of U, and $h(x,u)$ is continuous in x. (ii) Given any fixed x, the support of the distribution of $h(x,U)$ contains an open interval.*

In the Appendix, we give a lemma that follows from Theorem 3 in Florens et al. (2008) under which the conditions in Assumption 2′ are sufficient for Assumption 2 to hold. Note that Assumption 2′(i) implicitly restricts the treatment D to be continuous, so it is not proper to impose this restriction for a binary treatment D. As we see in Example 1.1, the identification of LAR is possible without the measurable separability restriction when D is binary, and we will show in Theorem 1 that this is the same case here. Assumption 2′(ii) requires U to be continuously distributed and that $h(x,U)$ is a continuous monotonic function of U for any fixed x.

We now give identification results of $CLAR$ and LAR for both continuous treatment and binary treatment cases in the following theorem:

Theorem 1. *Consider the model defined in (3) and (4).*

(i) For a continuous random variable D, suppose that Assumption 1 and 2' hold and $E\left[\left|\frac{\partial m(d,x,\varepsilon)}{\partial d}\right| |D=d, X=x\right] < \infty$, then LAR and $CLAR$ are identified for all $d \in Supp(D)$ and $x \in Supp(X|D=d)$ as follows:

$$\beta(d,x) = \frac{\partial E[Y|D=d, X=x]}{\partial d},$$
$$\beta(d) = \int \frac{\partial E[Y|D=d, X=x]}{\partial d} f_{X|D=d}(x)dx.$$

(ii) For a binary random variable D, suppose Assumption 1 holds and for all $d \in Supp(D)$ and $x \in Supp(X|D=d)$, $E\left[|m(1,\varepsilon) - m(0,\varepsilon)|\,|D=d, X=x\right] < \infty$, then LAR and $CLAR$ are identified for all $d \in Supp(D)$ and $x \in Supp(X|D=d)$ as follows:

$$\tilde{\beta}(d,x) = E[Y|D=1, X=x] - E[Y|D=0, X=x],$$
$$\tilde{\beta}(d) = \int \left(E[Y|D=1, X=x] - E[Y|D=0, X=x]\right) f_{X|D=d}(x)dx.$$

The proof of Theorem 1 is given in Appendix.

3 Nonseparable Triangular Models

As an alternative to the conditional independence assumption, another useful identifying restriction to solve the endogeneity problem of the treatment is the excluded IV by which we can construct a *control* variable that controls for the endogeneity from the treatment equation. In applications, other observable control variables are included to make the exogeneity condition of the IV more likely to hold. Commonly, these observable controls are assumed to be exogenous and included in both the outcome equation and the reduced form equation. We caution that these control variables may be endogenous, too, while finding IV for all endogenous controls is not possible. In this section, we study a nonseparable triangular model similar to the one in Imbens and Newey (2009) where we explicitly include potentially endogenous control variables in the model and provide identification results on LAR and treatment effects.

Consider the nonseparable triangular model as follows:

$$Y = g(D, X, \varepsilon), \tag{5}$$

$$D = q(Z, X, \eta) \tag{6}$$

where D is a continuously distributed random variable and endogenous to the stochastic error. X is a vector of observable control variables potentially endogenous to unobservable determinants of Y. Z is an exogenous variable excluded from the outcome equation (5) and is independent of (ε, η):

Assumption 3. $Z \perp\!\!\!\perp (\varepsilon, \eta) \mid X$.

This is different from Imbens and Newey (2009) in that the endogenous variables now have been separated into two vectors D and X, and we are only interested in identifying parameters associated with D. Note that X is allowed to be correlated with both D and Z, which also motivates the inclusion of X in the model as it makes Assumption 3 more likely to hold.

If there exists a control variable V such that

$$f_{\varepsilon|D,X,V}(\epsilon) = f_{\varepsilon|X,V}(\epsilon),\ \forall \epsilon \in \mathbb{R}, \tag{7}$$

and both X and V are measurably separated of D, then we can apply the similar approach from Section 2 to identify $CLAR$ and LAR, which in this case are defined as:

$$\begin{aligned}\beta(d,x) &= \int \frac{\partial g(d,x,\epsilon)}{\partial d} f_{\varepsilon|D=d,X=x}(\epsilon)d\epsilon \\ &= \int\int \frac{\partial g(d,x,\epsilon)}{\partial d} f_{\varepsilon|D=d,X=x,V=v}(\epsilon) f_{V|D=d,X=x}(v)dvd\epsilon, \\ \beta(d) &= \int\int \frac{\partial g(d,x,\epsilon)}{\partial d} f_{X,\varepsilon|D=d}(x,\epsilon)dxd\epsilon \\ &= \int\int\int \frac{\partial g(d,x,\epsilon)}{\partial d} f_{\varepsilon|D=d,X=x,V=v}(\epsilon) f_{V|D=d,X=x}(v) f_{X|D=d}(x)dxdvd\epsilon.\end{aligned}$$

Under the condition (7), measurably separability, and appropriate regularity conditions for the derivative to pass through the expectation, we have

$$\begin{aligned}\frac{\partial E[Y|D=d,X=x,V=v]}{\partial d} &= \frac{\partial \int g(d,x,\epsilon) f_{\varepsilon|D=d,X=x,V=v}(\epsilon)d\epsilon}{\partial d} \\ &= \int \frac{\partial g(d,x,\epsilon)}{\partial d} f_{\varepsilon|D=d,X=x,V=v}(\epsilon)d\epsilon.\end{aligned}$$

Then, $CLAR$ and LAR are identified for all $d \in Supp(D)$ and $x \in Supp(X|D=d)$ as follows:

$$\begin{aligned}&\int \frac{\partial E[Y|D=d,X=x,V=v]}{\partial d} f_{V|D=d,X=x}(v)dv \\ &\qquad = \int\int \frac{\partial g(d,x,\epsilon)}{\partial d} f_{\varepsilon|D=d,X=x,V=v}(\epsilon) f_{V|D=d,X=x}(v)dvd\epsilon = \beta(d,x), \\ &\int\int \frac{\partial E[Y|D=d,X=x,V=v]}{\partial d} f_{V|D=d,X=x}(v) f_{X|D=d}(x)dvdx \\ &\qquad = \int\int\int \frac{\partial g(d,x,\epsilon)}{\partial d} f_{\varepsilon|D=d,X=x,V=v}(\epsilon) f_{V|D=d,X=x}(v) f_{X|D=d}(x)dxdvd\epsilon = \beta(d).\end{aligned}$$

We construct such control variable V satisfying condition (7) in a fashion similar to Imbens and Newey (2009): $V = F_{D|Z,X}(D)$, i.e. the conditional CDF of D given (Z,X). The following assumption is essential for the construction of V and ensuring that the information contained in V is the same as that in η.

Assumption 4. *(i) $q(z,x,e)$ is strictly monotonic in e for any fixed (z,x); (ii) η is continuously distributed with its CDF $F_\eta(e)$ strictly increasing in the support of η.*

Assumption 4(i) allows the inverse function of $q(z,x,e)$ with respect to e to exist. Assumption 4(ii) implies that $F_\eta(e)$ is a one-to-one function of e.

We further discuss measurable separability conditions. First note that if η is independent of (Z, X) in (6), we can fix η and see that D and X are measurably separated (i) if Z and X are measurably separated, which would hold under similar sufficient conditions as Assumption $2'$, and (ii) if $q(z, \cdot, \cdot)$ is continuous in z. Again, this essentially means that given fixed values of $(X, \eta) = (x, e)$ we can vary $Z = z$, so does $D = d$ because $d = q(z, x, e)$. We also provides primitive conditions for the measurable separability between D and η:

Assumption 5. *(i) D is determined by (6) where η is continuously distributed and independent of (Z, X), and $q(z, x, e)$ is continuous in e. (ii) For any fixed e, the support of the distribution of $q(Z, X, e)$ contains an open interval.*

It is a counterpart of Assumption $2'$, it implies the measurable separability of D and η by Lemma 1 in Appendix. The difference is that Assumption 5(ii) does impose some restriction on the X and Z: Assumption 5(i) requires X to be independent of unobservable determinants of D and Assumption 5(ii) requires (Z, X) contains a continuous element and q is continuous in that element for any fixed e.

In the next theorem, we show that the the constructed control variable V satisfies condition (7) and is measurable separated of D.

Theorem 2. *Suppose Assumption 3 holds for the nonseparable model in (5) and (6). Then,*

(i) D is independent of ε conditional on (η, X).

(ii) If, additionally, Assumptions 4 and 5 holds, then condition (7) is satisfied with $V = F_\eta(\eta) = F_{D|Z,X}(D)$ and D is measurably separated of V.

The proof of Theorem 2 is given in Appendix.

4 Simulation

In this section, we use Monte Carlo simulations to demonstrate the bias due to the endogenous control, and how the proposed methods perform in finite sample. Two DGPs considered are as follows: (1) the first DGP covers the scenario where the control is endogenous and treatment is conditionally independent of the unobserved determinants as in Section 2; (2) the second DGP covers the scenario where the IV is conditionally valid but the control is endogenous as in Section 3.

First, consider DGP(1):

$$\begin{aligned}\text{DGP}(1): \ & Y = \beta_0 + \beta_1 D + \beta_2 X + U \\ & D = a^2 + N(0,1) \\ & X = a \\ & U = b^2 + N(0,1),\end{aligned}$$

where (a, b) are jointly normal with mean zero, variance one, and covariance 0.75; $N(0,1)$ denotes a random draw from a normal distribution, independent of a and b. We observe that (i) X is relevant for both Y and D; (ii) U affects both D and X; (iii) Conditional on X, D is independent of U; and (iv) D and X are measurably separated. As a result, linear projection parameters does not identify β_1, which is the local average response and is our parameter of interest. As is shown in Section 2, β_1 is nonparametrically identified.

Table 1: Simulation for DGP(1)

Methods	Bias	SD
OLS w.o. control	0.374	0.055
OLS w. control	0.374	0.044
Local linear regression	0.114	0.444
Series regression	0.001	0.048

Note: Simulation results are based on 1000 replications and random samples of size n = 1000. Series regression uses 3rd-order polynomials.

In Table 4, we compare the estimates of β_1 using (I) OLS without control, (II) OLS with control, (III) the nonparametric estimations through the local linear regression, and (IV) the third-order polynomial series regression. The results are clear: while conventional methods that assume the exogeneity of controls are severely biased, both the nonparametric methods perform much better in terms of the bias. The local linear kernel method is less efficient compared to the series approximation methods. Note that we can rewrite $X = (D - N(0,1))^{1/2}$, which is a function of the treatment D and so is subject to the critics of bad control. However, we see that the proposed methods still works in this case.

Next, consider DGP(2):

$$\begin{aligned}
\text{DGP(2)}:\ & Y = \beta_0 + \beta_1 D + \beta_2 X + U \\
& D = X + Z + \eta \\
& X = a \\
& Z = a^2 + N(0,1) \\
& \eta = c \\
& U = b^2 + d^2 + N(0,1),
\end{aligned}$$

where (c, d) are also jointly normal with mean zero, variance one, and covariance 0.75, and (c, d) are independent of a, b. We observe that (i) D is not conditionally independent given X; (ii) Z is not a valid IV except when it is conditional on X, but X is endogenous; (iii) Z and X are measurably separated, so the local average response parameter β_1 is nonparametrically identified as 1; (iv) The measurable separability between Z and X here also implies the measurable separability between D and (X, η), so β_1 can also be nonparametrically identified as 2 and we can use a two-step control function approach. Note that the model is linear, a special case of the nonseparable model in Section 3, so we can also resort to Theorem 2 for identification.

Table 1: Simulation for DGP(1)

Methods	(I)	(II)	(III)	(IV)
Bias	0.374	0.374	0.114	0.001
SD	0.055	0.044	0.444	0.048

Note: Simulation results are based on 1000 replications and random samples of size $n = 1000$. Series regression uses 3rd-order polynomials.

Table 2 compares the following 6 methods, corresponding to their column numbers in the same table: (I) IV without control; (II) IV with control; (III) nonparametric approach as 1 using series regression; (IV) nonparametric approach as 1 using local linear regression; (V) two-step control function approach as 2 using series regression; (VI) nonparametric approach using Theorem 2. Although (6) is the most general approach that allows for nonseparable models, it is very computationally costly because we need to estimate the conditional CDF $F_{D|Z,X}(D_i)$ for each realized D_i in the sample. In practice, method (6) is implemented as follows: (a) For some $i = 1, ..., n$, we generate the indicator variable $1\{D < D_i\}$. (b) Then we use local linear kernel methods to regress $1\{D < D_i\}$ on X and Z for each d and obtain the the fitted values of $1\{D < D_i\}$ only for the i-th observation. After, repeat steps (a)-(b) for each $i = 1, ..., n$, we obtain the estimated $F_{D|Z,X}(D_i)$ for each i. Finally, we nonparametrically regress Y on D, X, and the estimated $F_{D|Z,X}(D_i)$ using polynomial series regression. The results of this comparison are as expected by theory. While the IV-based methods (I) and (II) that implicitly imposes the exogeneity of the controls fail to produce consistent estimates of the partial effects of interest, the alternative identification methods paired with usual nonparametric estimators in (III)-(VI) perform well in finite sample.

Table 2: Simulation for DGP(2)

Methods	(I)	(II)	(III)	(IV)	(V)	(VI)
Bias	0.374	0.375	0.139	0.001	0.001	0.001
SD	0.057	0.052	0.161	0.068	0.100	0.100

Note: Simulation results are based on 1000 replications and random samples of size $n = 1000$.

5 Conclusion

This note addresses a critical, prevalent, yet often overlooked problem in empirical research: the endogeneity of control variables. Building on the insightful observation and discussion in Frölich (2008) that nonparametric estimation can help with the endogenous control problem, we provide formal identification results in a simple linear model with or without the presence of instrumental variables, and extend the results to a general class of nonseparable models focusing on identifying local average responses.

For empirical practice, this note provides a more flexible framework of dealing with endogenous control. Following primitive conditions we provide in this note, researchers could evaluate if the inclusion of potentially endogenous control variables is desirable and if the estimation and inference can be made robust to the potential endogeneity in the control. Estimation based on our identification results is also standard in common empirical settings.

Appendix

We first restate Theorem 3 Florens et al. (2008) as Lemma 1 below, which gives primitive conditions for measurable separability.

Lemma 1. *Suppose D is determined by $D = h(Z, V)$, where V is continuously distributed and independent of Z, and $h(z, v)$ is continuous in v. Further, for any fixed v, the support of the distribution of $h(Z, v)$ contains an open interval. Then, D and V are measurably separated.*

Proof of Theorem 1. First, note that Assumption $2'$ implies Assumption 2 using Lemma 1 with $Z = U$ and $V = X$. For continuous D, Assumptions 1 and 2 implies that

$$\frac{\partial f(\epsilon|D = d, X = x)}{\partial d} = \frac{\partial f(\epsilon|X = x)}{\partial d} = 0.$$

Then, we have

$$\frac{\partial E[Y|D = d, X = x]}{\partial d} = \frac{\partial \int m(d, x, \epsilon) f_{\varepsilon|D=d,X=x}(\epsilon) d\epsilon}{\partial d} = \int \frac{\partial m(d, x, \epsilon)}{\partial d} f_{\varepsilon|D=d,X=x}(\epsilon) d\epsilon,$$

where the last equality follows from the Leibniz integral rule and the chain rule. Therefore, $\beta(d, x)$ is identified by $\frac{\partial E[Y|D=d,X=x]}{\partial d}$ for all $d \in Supp(D)$ and $x \in Supp(X|D = d)$.

Furthermore, taking integrals on both sides with respect to X given $D = d$ gives

$$\int \frac{\partial E[Y|D = d, X = x]}{\partial d} f_{X|D=d}(x) dx = \int \int \frac{\partial m(d, x, \epsilon)}{\partial d} f_{X,\varepsilon|D=d}(x, \epsilon) dx d\epsilon.$$

So, $\beta(d)$ is identified by $\int \frac{\partial E[Y|D=d,X=x]}{\partial d} f_{X|D=d}(x) dx$.

In the case of binary D, Assumptions 1 implies that $f_{\varepsilon|D=1,X=x}(\epsilon) d\epsilon = f_{\varepsilon|D=0,X=x}(\epsilon) d\epsilon$, so we have

$$\begin{aligned}
&E[Y|D = 1, X = x] - E[Y|D = 0, X = x] \\
=& \int m(1, x, \epsilon) f_{\varepsilon|D=1,X=x}(\epsilon) d\epsilon - \int m(0, x, \epsilon) f_{\varepsilon|D=0,X=x}(\epsilon) d\epsilon \\
=& \int (m(1, x, \epsilon) - m(0, x, \epsilon)) f_{\varepsilon|D=d,X=x}(\epsilon) d\epsilon.
\end{aligned}$$

So, $\tilde{\beta}(d, x)$ is identified for all $d \in Supp(D)$ and $x \in Supp(X|D = d)$ and taking integral on both sides with respect to X given $D = d$ gives $\tilde{\beta}(d)$

$$
\begin{aligned}
&\int \left(E[Y|D=1, X=x] - E[Y|D=0, X=x]\right) f_{X|D=d}(x)dx \\
=&\int\int \left(m(1,x,\epsilon) - m(0,x,\epsilon)\right) f_{X,\varepsilon|D=d}(x,\epsilon)d\epsilon dx
\end{aligned}
$$

$\square$

Proof of Theorem 2. The proof of statement (i) and part of statement (ii) follows closely the proof of Theorem 1 in Imbens and Newey (2009). For statement (i), let l be any continuous and bounded real function. Due to the independence of Z and (ε, η) conditional on X, we first obtain the conditional mean independence as an intermediate result:

$$
\begin{aligned}
E[l(D)|\varepsilon, \eta, X] &= E[l(q(Z, X, \eta))|\varepsilon, \eta, X] \\
&= \int l(q(z, X, \eta))dF_{Z|\varepsilon,\eta,X}(z) \\
&= \int l(q(z, X, \eta))dF_{Z|X}(z) = E[l(D)|\eta, X].
\end{aligned}
$$

Then, we can check the conditional independence of D and ε given (η, X) by a conditional version of Theorem 2.1.12 of Durrett (2019). Let $a(\cdot)$ and $b(\cdot)$ be any continuous and bounded real functions, then

$$
\begin{aligned}
E[a(D)b(\varepsilon)|\eta, X] &= E[E[a(D)b(\varepsilon)|\varepsilon, \eta, X]|\eta, X] \\
&= E[E[a(D)|\varepsilon, \eta, X]b(\varepsilon)|\eta, X] \\
&= E[E[a(D)|\eta, X]b(\varepsilon)|\eta, X] \\
&= E[a(D)|\eta, X]E[b(\varepsilon)|\eta, X].
\end{aligned}
$$

Consider statement (ii). The measurable separability between D and η is implied by Assumption 5 using Lemma 1 with $Z = (Z, X)$ and $V = \eta$. So it suffices to show that the sigma-algebra generated by V is the same as that of η. By strict monotonicity of $q(z, x, e)$

in e for any fixed (z, x), there exists an inverse function $q^{-1}(z, x, d) = e$. Then, we have

$$\begin{aligned}F_{D|Z=z,X=x}(d) &= Pr(D \leq d|Z = z, X = x) = Pr(q(z, x, \eta) \leq d|Z = z, X = x) \\ &= Pr(\eta \leq q^{-1}(z, x, d)|Z = z, X = x) = Pr(\eta \leq q^{-1}(z, x, d)) \\ &= F_\eta(q^{-1}(z, x, d)).\end{aligned}$$

where the second to the last equality follows from the the independence of (Z, X) and η under Assumption 5. Note that $\eta = q^{-1}(Z, X, D)$ a.s., so we have $V = F_{D|Z,X}(D) = F_\eta(\eta)$. Under Assumption 4, $F_\eta(e)$ is a one-to-one function of e, which implies the sigma-algebra generated by $F_\eta(\eta)$ is the same as that of η.

Furthermore, combining with the independence of η and X under Assumption 5, we have

$$\begin{aligned}E[a(D)b(\varepsilon)|V, X] &= E[a(D)b(\varepsilon)|\eta, X] \\ &= E[a(D)|\eta, X]E[b(\varepsilon)|\eta, X] \\ &= E[a(D)|V, X]E[b(\varepsilon)|V, X]\end{aligned}$$

which implies condition (7). $\square$

References

Altonji, J. G. and Matzkin, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, 73(4):1053–1102.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Diegert, P., Masten, M. A., and Poirier, A. (2022). Assessing omitted variable bias when the controls are endogenous. *arXiv preprint arXiv:2206.02303*.

Durrett, R. (2019). *Probability: theory and examples*, volume 49. Cambridge university press.

Florens, J.-P., Heckman, J. J., Meghir, C., and Vytlacil, E. (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica*, 76(5):1191–1206.

Florens, J. P., Mouchart, M., and Rolin, J.-M. (1990). *Elements of Bayesian statistics*. CRC Press.

Frölich, M. (2008). Parametric and nonparametric regression in the presence of endogenous control variables. *International Statistical Review*, 76(2):214–227.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512.

Kim, K. (2013). Regression discontinuity design with endogenous covariates. *Journal of Economic Theory and Econometrics*, 24(4):320–337.

Lechner, M. (2008). A note on the common support problem in applied evaluation studies. *Annales d'Économie et de Statistique*, pages 217–235.

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

Wooldridge, J. M. (2005). Violating ignorability of treatment by controlling for too many factors. *Econometric Theory*, 21(5):1026–1028.