Identification of Partial Effects with Endogenous Controls

Kaicheng Chen* Kyoo il Kim†

Dec 25, 2025

Abstract

While control variables are often assumed to be exogenous, it is common to encounter endogenous controls in practice. It brings a dilemma: without controls, the treatment of interest may be endogenous; with controls, the endogeneity of controls may pollute the identification. The problem may not be solved with an instrumental variable when it is only conditionally valid. For these scenarios, we provide identification of partial effects of the treatment under a rank condition called measurable separability. For parametric models, our approach amounts to employ a nonparametric estimator of a nesting model. For nonparametric models, our results show that endogenous controls are generally innocuous except for the case of "bad control". We further propose a test for the endogenous control. Simulation results exemplify this prevalent issue and demonstrate the performance of the proposed methods in finite sample.

Keywords: endogenous control, average partial effects, local average response, nonseparable model.

JEL Classification Code: C14, C31

1 Introduction

In models with endogenous treatments, researchers often leverage the conditional independence or instrumental variables (IV) for identification of treatment effects while rather casu-

*School of Economics, Shanghai University of Finance and Economics. Email: chenkaicheng@sufe.edu.cn
†Department of Economics, Michigan State University. Email: kyookim@msu.edu

ally assuming the control variables are exogenous. In reality, however, empirical researchers often end up with control variables that may be subject to additional endogeneity concerns while finding instruments for every endogenous control is challenging or impossible. In this paper, we will demonstrate how endogeneity in control affects the identification of partial effects the treatments and, when it does, how can we get around the issue in certain scenarios.

1.1 Illustration in the simplest case.

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

$$Y = \tau W + \xi X + \varepsilon, \;\; E[\varepsilon|W, X] = E[\varepsilon|X], \tag{1}$$

where W is a treatment variable of interest, the control variable X is a determinant of the outcome Y and likely of W, too, and ε is an unobserved determinant. While it is assumed that ε is mean independent of W conditional on X, ε may not be mean independent of X. For example, if the function form of X is misspecified and some nonlinear effects of X ends up being part of ε, then the treatment is indeed conditionally independent while ε is dependent on X. Generally, this is more than a specification issue, although, as we will see in a moment, in many cases this can be mitigated by a flexible function form.

There are two consequences of the model above: First, without considering X, the dependence between W and X can cause bias in the OLS estimation. Second, with the presence of X, the dependence between X and ε can also pollute the identification of the average partial effect (APE) τ even if W is conditionally independent of ε. In either case, τ is not identified by the linear projection parameters, so the OLS estimator is inconsistent and biased. The bias in the first case can be interpreted as an omitted variable bias. To see the bias in the second case, let $D = (W, X)$ and $\theta = (\tau, \xi)'$, the linear projection parameters are defined as follows:

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

Therefore, without further restriction on $E[\varepsilon|X]$, OLS does not produce a consistent estimate for θ, or τ in particular. In a worse scenario, X could be an outcome of W, in which X is referred to as "bad control" in Angrist and Pischke (2009). With the presence of bad control, the conditional independence is not likely to hold (Lechner, 2008), and it is also noted that the bad control can cause problems for identification even if treatments are randomly assigned

(Wooldridge, 2005).

However, as is shown below, τ can still be identified as long as X is not solely a function of W. More formally, this extra condition is referred to as the measurable separability, as first introduced in Florens et al. (1990):

Definition 1. *W and X are measurably separated if, any function of W almost surely equal to a function of X must be almost surely equal to a constant.*

At its essence, this assumption ensures that we can vary the value of W while holding X at a particular value. Note that this still allows the distribution of X to depend on W, and vice versa. Although it does not rule out the bad control directly[1], it rules out situations where conditional independence is not likely to hold. Under this condition, for a continuous treatment W, τ is nonparametrically identified as follows:

$$\int_{\mathcal{W}\times\mathcal{X}} \partial_w E[Y|W=w, X=x]dF(w,x) = \tau + \int_{\mathcal{W}\times\mathcal{X}} \partial_w (x\xi + E[\varepsilon|X=x])dF(w,x) = \tau,$$

where the last result holds due to the measurable separability of W and X. To see why this condition is necessary, suppose the measurable separability does not hold, e.g., $X = f(W)$ almost surely and they are not constants, then conditioning on $W = w, X = x$ necessitates $W = w, X = f(w)$. In that case, we would have $\int_{\mathcal{W}\times\mathcal{X}} \partial_w(f(w)\xi + E[\varepsilon|X = f(w)])dF(w,x) \neq 0$.

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

$$\int_{\mathcal{X}} \left(E[Y|W=1, X=x] - E[Y|W=0, X=x]\right) dF(x) = \tau.$$

Example 1. *As a concrete example, consider the linear regression model relating district-level average test scores (avgscore) to district-level educational expenditure per student (expend) and average family income (avginc) from Wooldridge, 2019, Chapter 3:*

$$avgscore = \alpha + \tau \cdot expend + \xi \cdot avginc + \varepsilon.$$

Suppose we are interested in the partial effect of expend, τ. Since avginc is relevant for

[1] For example, $X = W + e$ where $e \perp\!\!\!\perp W$, then we can vary W while fixing X but in terms of potential outcome $X(w) \neq X$.

expend in the district level and avginc can also affect avgscore through other channels, e.g. private tutoring, including avginc as a control variable is sensible. However, avginc may also be correlated with other unobserved determinants that affect both avginc and avgscore, then the endogeneity of avginc can pollute the identification of (τ, ξ) and OLS fails to produce consistent estimates. However, τ is nonparametrically identified as long as: (1) expend is independent of ε conditional on avginc and (2) expend and avginc are measurably separated. Given those two conditions and the identification results above, a consistent estimate for τ is available through usual nonparametric estimators.

1.2 Does an IV solve the problem?

When the IV is truly exogenous and affects the outcome only through the treatment, the answer is yes. However, in practice, control variables are often included to either increase efficiency or to strengthen the validity of the IV. In these cases, again, the endogeneity in those controls can cause problems for identification. To illustrate the problem, let's consider the linear model again with an excludable IV:

$$\begin{aligned} Y &= \tau W + \xi X + \varepsilon,\ \ E[\varepsilon|W, X] \neq E[\varepsilon|X] \\ W &= \pi_Z Z + \pi_X X + \eta. \end{aligned}$$

In this model, the IV Z is needed because W is not conditionally independent of ε even after conditioning on X. However, without controlling X, the IV Z may not be valid if Z affects Y through X too. Therefore, again, the endogeneity of X brings a dilemma, and neither τ or ξ is identified by usual IV or *2SLS* projection. Nevertheless, τ can be nonparametrically identified through a control function approach[2]. If Z is independent of (ε, η) conditional on X, then W is independent of ε conditional on X and η. Accordingly, given the identification of η and the measurable separability between (X, η) and W[3], τ is identified:

$$\begin{aligned} &\int_{\mathcal{W}\times\mathcal{X}\times\mathcal{E}} \partial_w E[Y|W = w, X = x, \eta = e] dF(w, x, e) \\ =&\tau + \int_{\mathcal{W}\times\mathcal{X}\times\mathcal{E}} \partial_w E[\varepsilon|X = x, \eta = e] dF(w, x, e) = \tau. \end{aligned} \tag{2}$$

[2] It is not the only way of identification. In this linear triangular model, τ is also identified by $E[\partial_Z E[Y|X, Z]]/E[\partial_Z E[W|X, Z]] = \tau$.

[3] Because $W = Z\pi_Z + X\pi_X + \eta$ with $\pi_Z \neq 0$, the measurable separability between Z and X also implies the measurable separability between (X, η) and W in this case.

Example 2. *Consider a linear triangular model relating the individual wage with school attendance. In an influential paper that studies the causal impact of compulsory school attendance on earnings, Angrist and Krueger (1991) use quarter of birth (qbirth) as an instrument for educational attainment (totaledu) in wage equations, based on the observation that school-entry requirement and the compulsory schooling laws compel students born in the end of the year to attend school longer than students born in other months. However, this approach is also subject to some critiques that qbirth may not be truly exogenous because, for example, the family income level could affect conception planning*[4]*, which in turn affects of birth. Therefore, research may consider including the parents' income (parinc) as a control. The heuristic model can be specified as follows:*

$$\begin{aligned} log(wage) &= \alpha + \tau \cdot totaledu + \xi \cdot parinc + \varepsilon, \\ totaledu &= \pi_0 + \pi_1 \cdot qbirth + \pi_2 \cdot parinc + \eta. \end{aligned}$$

However, the way parents' income affects children's wage varies a lot and can be further correlated with other socio-economics determinants, which makes it an potential endogenous control. In that case, IV or 2SLS does not yield consistent estimates for (τ) *with or without including parinc linearly, but* τ *can be identified nonparametrically as shown in 2.*

1.3 More General Scenarios

Given the simple illustration of the endogenous control problem above, it seems that non-parametric methods are more immune to the endogenous control problem. This is first argued in Frölich (2008), but many questions remain: First, to what extent the nonparametric methods are immune to the endogenous control. For example, some extra regularity conditions may be needed such as the aforementioned measurable separability, and it is important to know the implication of such restrictions. Furthermore, we would like to ask how general can we define a class of models where the nonparametric methods are valid under endogenous controls. If we can define such an admissible class of models, then the next question is should researchers always use nonparametric methods? It is well-known that nonparametric methods may not be efficient, as a cost of being robust to specification, and, in this case, to endogenous control. Ideally, we would like to have a testing procedure that informs us about the potential endogenous controls in a general class of models.

[4]For example, high-income households may have more control and flexibility on conception timing to deliberately avoid late birth for better schooling planning.

As the first contribution, we study a class of parametric models that are nested in a nonseparable and nonparametric model. We show that under the usual conditional independence condition, the identification of the average response parameters is possible with the extra measurable separability condition in Section 2.1. Since average response of a nonseparable model coincides with the average partial effects of any nested models, researchers can instead resort to the nonseparable model to avoid the endogenous bias caused by the controls. An analogous result for nonseparable triangular model with IVs is also provided in Section 2.2.

Based on these results, we further propose a test for endogenous control in Section 3. For linear models, the test follows from the general idea of Hausman's specification test. For more general setups, the test is based on an asymptotic linear representation of the difference of two estimators. Particularly, we focus on the series estimation of average derivatives and inference using bootstrap/subsampling methods.

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. In a recent paper by Andrews et al. (2025), the issue of endogenous controls is attributed to misspecification and their method of "strong exclusion" amounts to projecting out from the instrument Z a conditional mean function of Z given X, which is conceptually and econometrically equivalent to including more flexible functional forms of X as the control function.

The rest of the paper is outlined as follows: The main identification results are given in Section **?**; Section 3 proposes the endogenous control test. Section 4 presents a simulation study to illustrates the endogenous bias of controls and the finite sample performance of the proposed methods. Section 5 concludes the paper with empirical recommendations.

2 Nonseparable Models with Endogenous Controls

To investigate the impact of endogenous control variables in a more general setting, we consider a class of models as follows:

$$Y = m(W, X, \varepsilon) = h(W, X, \varepsilon; \theta) \tag{3}$$

where m is nonparametric but can be parameterized by h. Suppose h is known to researcher up to the unknown parameters θ. A similar nonparametric and nonseparable model $Y = m(W, \varepsilon)$ has been studied by Altonji and Matzkin (2005) except that they consider X as some excludable instruments that do not enter the outcome equation. With similar notations, we refer to the parameters of interest as average response (AR), local AR (LAR), conditional LAR (CLAR), and they are defined as follows: For a continuous treatment W

$$\begin{aligned}
&\text{CLAR}: \beta(w, x) = \int \partial_w m(w, x, \epsilon) f_{\varepsilon|W=w, X=x}(\epsilon) d\epsilon, \\
&\text{LAR}: \quad \beta(w) = \int \beta(w, x) f_{X|W=w} dx, \\
&\text{AR}: \qquad \beta = \int \beta(w) f_W(w) dw
\end{aligned}$$

If Y is a binary outcome, then we can write

$$\begin{aligned}
m(W, X, \varepsilon) =& 1\{m^*(W, X, \varepsilon) > 0\}, \\
h(w, X, \varepsilon; \theta) =& 1\{h^*(w, X, \varepsilon; \theta^*) > 0\},
\end{aligned}$$

where m^*, h^*, and θ^* are implicitly defined. Similar to Altonji and Matzkin (2005), we partition ε as (u, v) and implicitly define $u^* = u^*(W, X, v)$ as a solution of $m^*(W, X, u^*, v) = 0$. Suppose that $m^*(w, x, u, v)$ has at least one root and that $m^*(w, x, u, v)$ is strictly monotonic in u, then $u^*(W, X, v)$ is uniquely defined. Additionally, suppose $m^*(w, x, u, v)$ is continuously differentiable in w and u and that $\partial_u m^*(w, x, u^*, v) \neq 0$, then by implicit function theorem $u^*(w, x, v)$ is continuously differentiable in w. Then, we can define the binary response parameters as

$$\begin{aligned}
&\text{BCLAR}: \beta^*(w,x) = \int -\partial_w u^*(w,x,v) f_{u,v|W=w,X=x}(u^*(w,x,v),v)dv,\\
&\text{BLAR}: \quad \beta^*(w) = \int \beta^*(w,x) f_{X|W=w}dx,\\
&\text{BAR}: \qquad \beta^* = \int \beta^*(w) f_W(w)dw
\end{aligned}$$

For a binary treatment, we can define a set of parameters analogously by replacing the derivatives above with the differences. For our main results, we will focus on the continuous treatment case to simplify the exposition.

Example 3. *Consider a parametric nonseparable model* $Y = h(W, X, \varepsilon; \theta) = 1\{\tau W + \xi X + \varepsilon > 0\}$ *where* $\theta = (\tau, \xi)$ *and* ε *is Gaussian. Suppose we can partition* ε *as* $\varepsilon = u + v$*, and* v *is also Gaussian. Note that* $-u^*(W, X, v) = \tau W + \xi X + v$*, then* $-\partial_w u^*(w, x, v) = \tau$ *and*

$$\begin{aligned}
\beta(w,x) =& \tau \int f_{u,v|W=w,X=x}(-(\tau w + \xi x + v), v)dv\\
=&\tau \int f_{\varepsilon,v|W=w,X=x}(-(\tau w + \xi x), v)dv = \tau f_{\varepsilon|W=w,X=x}(-(\tau w + \xi x))\\
=&\frac{\tau}{\sigma(w,x)}\phi\left(\frac{-(\tau w + \xi x) - \mu(w,x)}{\sigma(w,x)}\right)
\end{aligned}$$

where the second equality uses the fact that $\varepsilon = u + v$ *and the change-of-variables;* $\mu(W, X)$ *and* $\sigma(W, X)^2$ *are the conditional mean and variance since we don't assume* ε *is independent of* (W, X)*. If* $\varepsilon | W, X \sim N(0, 1)$*, then* $\beta(w, x) = \tau\phi(\tau w + \xi x)$*, which reduces to the usual average partial effect of a probit model.*

2.1 Identification under Conditional Independence

We first consider the scenario where ε is conditionally independent of W given X but the dependence between X and ε is unrestricted.

Assumption 1. $\varepsilon \perp\!\!\!\perp W \mid X$*.*

Assumption 2 is a key condition for identifying the response parameters. Importantly, it does not rule out X being endogenous. We emphasize that Assumption 2 itself may not be sufficient for identification due to the endogeneity of X: e.g., if W is some function of

X only, which is dependent on ε, then the response parameters would not be identified by Assumption 2, and, in which case, even Assumption 2 itself is not likely to hold. In other words, we need some extra restriction to avoid such extreme cases. The next theorem shows that the measurable separability condition between the treatment and the control suffices for identification of the response parameters of the treatment.

Theorem 1. *Suppose W and X are measurably separated and Assumption 2 holds.*

1. *If Y is continuous, suppose $m(w,x,e)$ is differentiable in w and there exists an integrable dominating function of $\partial_W m(W,X,\varepsilon)$ when it exists. Then, $\beta(w,x) = \partial_w E[Y|W = w, X = x]$, $\beta(w) = E[\partial_w E[Y|W = w, X]|W = w]$, and $\beta = E[\partial_W E[Y|W,X]]$.*

2. *If Y is binary, suppose $m^*(w,x,u,v)$ has at least one root and is strictly monotonic in u. Additionally, suppose $m^*(w,x,u,v)$ is continuously differentiable in w and u and $\partial_u m^*(w,x,u^*,e) \neq 0$, then $\beta^*(w,x) = \partial_w E[Y|W = w, X = x]$, $\beta^*(w) = E[\partial_w E[Y|W = w, X]|W = w]$, and $\beta^* = E[\partial_W E[Y|W,X]]$.*

Proof. By Assumptions 2 and the measurable separability between W and X, we have, for continuous W,

$$\partial_w f_{\varepsilon|W=w,X=x}(\epsilon) = \partial_w f_{\varepsilon|X=x}(\epsilon) = 0.$$

Then, applying Leibniz integral rule, under the regularity conditions given in the theorem, and the chain rule gives

$$\begin{aligned}
&\partial_w E[Y|W = w, X = x] = \partial_w \int m(w,x,\epsilon) f_{\varepsilon|W=w,X=x}(\epsilon) d\epsilon \\
=&\int \partial_w m(w,x,\epsilon) f_{\varepsilon|W=w,X=x}(\epsilon) + \int m(w,x,\epsilon) \partial_w f_{\varepsilon|W=w,X=x}(\epsilon) d\epsilon \\
=&E[\partial_w m(w,x,\varepsilon)|W = w, X = x] = \beta(w,x)
\end{aligned}$$

The other two statements for the continuous-Y case simply follow from the definition of the conditional expectation.

When Y is binary, we can write

$$\begin{aligned}
&E[Y|W = x, X = x] = \int 1\{m^*(w,x,e) > 0\} f_{\varepsilon|W=w,X=x}(e) de \\
=&\int_{1\{m^*(w,x,e)>0\}} f_{e|W=w,X=x}(e) de = \int \left[\int_{u>u^*(w,x,v)} f_{\varepsilon|W=w,X=x}(u,v) du \right] dv.
\end{aligned}$$

By Lebniz rule, chain rule, Assumption 2, and the measurable separability between W and X, we have

$$\partial_w E[Y|W=x, X=x] = \int -\partial_w u^*(w,x,v) f_{\varepsilon|W=w,X=x}(u^*(w,x,v),v)dv = \beta^*(w,x).$$

The other two statements for the binary-Y case follow from the definition of the conditional expectation, too.

$\square$

Theorem 1 show that even if ε and X are potentially dependent, the response parameters associated with the treatment can still be identified by the usual conditional expectations as long as the conditional independence and the measurable separability condition hold. This is a positive result: it justifies the prevalent use of potentially endogenous control variables because the extra measurable separability condition is not hard to justify in practice in common situations. Meanwhile, we have seen in previous examples that common linear parametric models are not immune to the endogeneity of control, so this result also encourages the use of fully flexible models in these scenarios. In the next example, we also show that a nonlinear parametric model is not immune to endogeneity of controls, while resorting to a nesting nonseparable model can get around this problem.

Example 3 continued. *For the binary model $Y = 1\{\tau W + \xi X > \varepsilon\}$, suppose $\varepsilon|W,X \sim N(\mu(X), \Sigma(X))$, i.e. Assumption 2 holds for this model. By Theorem 1, β^* is identified nonparametrically, which is also the common average partial effect of the treatment on the response probability, i.e. $E[\partial_W P(Y=1|W,X)] = E\left[\frac{\tau}{\Sigma(X)}\phi\left(\frac{\tau W + \xi X - \mu(X)}{\Sigma(X)}\right)\right]$. However, ignoring the endogeneity of the controls and employing a standard probit approach would produce an inconsistent estimate of the average partial effect.*

By its definition, measurable separability simply requires that W and X are not functions of each other only. It is intuitive and the reasoning could be straightforward in practice for most economics variables. For our purpose, the condition is already a very mild rank restriction. For further discussion, readers are referred to Florens et al. (2008), where they also provide primitive conditions on the data generating process under which measurable separability between two random variables are guaranteed.

2.2 Identification with IV

When the conditional independence condition is not plausible, it is common to consider an excluded IV for identification. In applications, additional control variables are often included to make the exogeneity condition of the IV more likely to hold. Although control variables are explicitly or implicitly assumed to be exogenous typically, we caution that they may be endogenous in practice 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) while explicitly allowing for endogenous controls.

We consider model 3 again except now we don't impose the conditional independence between W and ε. Instead, suppose there exists an excludable variable independent of (ε, η), at least conditionally:

Assumption 2. $W = q(Z, X, \eta),\ Z \perp\!\!\!\perp (\varepsilon, \eta) \mid X$.

Note that we don't impose exogeneity condition of X with respect to either ε or η, yet X is allowed in both outcome and the reduced form equations in a nonseparable way. The requirement for IV is relaxed, and Z may well be dependent on X, which motivates the inclusion of X in the model.

From a control-function perspective, we are looking for a control variable V such that

$$\varepsilon \perp\!\!\!\perp W \mid (X, V) \tag{4}$$

and both X and V are measurably separated from W, then we can identify the response parameters by Theorem 1. In the next proposition, we show that under Assumption 2, η can serve such a purpose. If $q(Z, X, e)$ is one-to-one in e almost surely, then we might recover the information of η from W given Z and X. However, this is possible only when both Z and X are independent of η, which is not the case here since Z is only conditionally independent and X is allowed to depend on η. However, as we show in the next proposition, it turns out that we don't need to recover the information of η completely, and it suffices to construct V such that the sigma algebra generated by (X, η) is the same as that by (X, V).

Assumption 3. *(i) $q(Z, X, e)$ is one-to-one in e almost surely; (ii) The conditional CDF $F_{\eta|X}(e)$ is continuous and strictly increasing in $e \in supp(\eta)$ almost surely.*

Proposition 1. *Suppose Assumption 2 holds for the nonseparable model (3). Then,*

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

(ii) If, additionally, Assumptions 3 holds, then $F_{W|Z,X}(W) = F_{\eta|X}(\eta)$, and condition (4) is satisfied with $V = F_{W|Z,X}(W)$.

Proof. For statement (i), let l be any continuous and bounded real function. Using the independence of Z and ε conditional on X, we first obtain the conditional mean independence as an intermediate result:

$$E[l(W)|\varepsilon, \eta, X] = \int l(q(z, X, \eta))dF_{Z|\varepsilon,\eta,X}(z) = \int l(q(z, X, \eta))dF_{Z|\eta,X}(z) = E[l(W)|\eta, X].$$

Then, we can check the conditional independence of W 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

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

Consider statement (ii). Under Assumption 3(i), there exists an inverse function $q^{-1}(Z, X, w) = e$. Then, we have

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

where the last equality follows from the independence of Z and η conditional on X. Note that $\eta = q^{-1}(Z, X, W)$ a.s., so we have $V = F_{W|Z,X}(W) = F_{\eta|X}(\eta)$. Let $Q_{\eta|X}(u) = \inf\{e \in \mathbb{R} : F_{\eta|X}(e) \geq u\}$ be the conditional quantile function of η given X. Under Assumption 3(ii), $F_{\eta|X}(e)$ is a one-to-one function of e a.s., so we have $Q_{\eta|X}(F_{\eta|X}(\eta)) = \eta$. We note that $\sigma(F_{\eta|X}(\eta), X) \subset \sigma(\eta, X)$ since $F_{\eta|X}(\eta)$ is a function of η and X. Because $(\eta, X) = (Q_{\eta|X}(F_{\eta|X}(\eta)), X)$, we also have $\sigma(\eta, X) \subset \sigma(F_{\eta|X}(\eta), X)$. By setting $V = F_{\eta|X}(\eta)$, it follows that

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

which implies condition (4). $\square$

By Proposition 1, $(X, V) = (X, F_{W|Z,X})$ together delivers the conditional independence condition. Then, the identification of response parameters associated with the treatment can be obtained by Theorem 1[5].

3 A Test for Endogenous Controls

In this section, we consider a test for the endogenous control in a large class of models, as an implication of the results presented in Section 2. Consider the parametric model $Y = h(W, X, \varepsilon; \theta)$ as in 3, which is nested by the nonparameteric and nonseparable model $Y = m(W, X, \varepsilon)$. To focus on the main idea, we limit our attention to an i.i.d sample of size n and the case of the conditional independence in Assumption 2 with the conditioning set observable to the researcher[6].

Suppose we are interested in the average response of the treatment: $\beta := E[\partial_W h(W, X, \varepsilon; \theta)]$. For the linear model (1) in the introduction, we find $\beta = \tau$. For the nonlinear binary model in Example 3, we have $\beta = E\left[\frac{\tau}{\Sigma(X)}\phi\left(\frac{\tau W + \xi X - \mu(X)}{\Sigma(X)}\right)\right]$. For both cases, under the conditional independence assumption, β coincide with the average partial effects, as guaranteed by Theorem 1, and both are non-parametrically identified. However, we also note that neither is parametrically identified, and so the corresponding parametric estimators assuming exogeneity of the controls would be inconsistent. Meanwhile, when these models are correctly specified without endogenous controls, we know that the parametric estimators corresponding to these specification can be efficient (attaining the Cramer-Rao lower bound) under certain conditions. This observation naturally leads to a Hausman-type test on the endogenous control.

Formally, we consider the null and the alternative hypotheses as follows:

$$
\begin{aligned}
H_0 &: X \text{ is } \text{ exogenous}, \\
H_1 &: X \text{ is } \text{ endogenous}.
\end{aligned}
$$

The null and alternative hypotheses are specified loosely to accommodate a large class of models. The meaning of the null and the alternative varies over the specification of $h(W, X, \varepsilon; \theta)$. For example, in the linear model, $E[X\varepsilon] = 0$ is necessary and sufficient for the

[5]Although V does not enter the outcome equation $m(W, X, \varepsilon)$, we can always rewrite $\tilde{m}(W, X, V, \varepsilon) = m(W, X, \varepsilon)$ and treat $\tilde{m}$ as the outcome equation in Theorem 1.

[6]Note that the analysis can be extended to the case of Section 2.2 where the conditioning variable V needs to be estimated, but that would complicates the asymptotic analysis and deviates from the main idea of the test.

null while $X \perp\!\!\!\perp \varepsilon$ is not necessary. However, in the endogenous probit model of Example 3, $X \perp\!\!\!\perp \varepsilon$ is necessary and sufficient for the null while $E[X\varepsilon] = 0$ is not sufficient. To make the null and alternative hypothesis concrete, we impose the following assumptions. We denote the distribution of (Y, W, X) as F and the empirical distribution as F_n.

Assumption 4. *The parametric estimator $\hat{\beta}$ under consideration corresponding to model 3 has the following properties:*

1. *Under the null, β is parametrically identified as a functional $\beta = \Gamma_p(F)$, and the parametric estimator $\hat{\beta} = \Gamma_p(F_n)$ is asymptotically linear with an influence function $\varphi(W, X)$:*

$$\hat{\beta} - \beta = \Gamma_p(F_n) - \Gamma_p(F) = \int \varphi(W, X) d(F_n - F) + o_P(n^{-1/2}).$$

2. *$\hat{\beta}$ is asymptotically biased with bias $B > \varepsilon$ for some constant $\varepsilon > 0$, such that $\hat{\beta} - \beta - B = \Gamma_p(F_n) - \Gamma_p(F) - B = \int \varphi(W, X) d(F_n - F) + o_P(n^{-1/2})$.*

Under the linear model 1, Assumption 4 is easily verified. Note that there could be multiple estimators corresponding to a parametric model that satisfy these conditions, and the test is with respect to any one of those under consideration.

For concreteness, we consider a series estimator. Suppose (W, X) is d-dimensional. For some given K, we define a power series $\{\tilde{g}_k\}_{1 \leq k \leq K}$ to approximate $g(W, X) := E[Y|W, X]$:

$$\tilde{g} := \sum_{k=1}^{K} \tilde{g}_k' \pi_k = G'\Pi,$$
$$\tilde{\beta} := E[\partial_W [G(W, X)'\Pi]]$$

where $G = (\tilde{g}_1', ..., \tilde{g}_K')'$; for each k,

$$\tilde{g}_k = \text{vec} \left\{ \prod_{j=1}^{d} q_j^{\lambda_j} : \sum_{j=1}^{d} \lambda_j = k, (q_1, ..., q_d) = (l_1(W), l(X_1), ..., l(X_{d-1})) \right\},$$

and $l = (l_1, ..., l_K)$ is some transformation function differentiable to all orders with bounded derivatives and has $\det(\partial_z l(z))$ bounded away from zero; $\Pi = (\pi_1', ..., \pi_K)'$ are the corresponding linear projection parameters. The series estimator of β is defined as

$$\hat{\hat{\beta}} := \frac{1}{n}\sum_{i=1}^{n} \partial_W \left[G(W_i, X_i)'\hat{\Pi} \right],$$

where $\hat{\Pi}$ are the least square estimators of Π. $\hat{\hat{\beta}}$ is an average derivative estimator studied in Example 3 of Newey (1994): it is shown that, with growing K and other regularity condition given in Theorem 7.2 of the same paper, $\hat{\hat{\beta}}$ is asymptotically linear with an influence function ϕ:

$$\sqrt{n}(\hat{\hat{\beta}} - E[\partial_W g(W,X)]) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi(W_i, X_i) + o_P(1), \tag{5}$$
$$\phi(W,X) = \partial_W g(W,X) - \beta + \partial_W \log f_{W,X}(W,X)(Y - g(W,X)),$$

where $f_{W,X}(W,X)$ is the joint density function of (W,X).

When Assumption 2 holds, we have $\beta = E[\partial_W g(W,X)]$ under both the null and the alternative. Therefore, under the null, $\tilde{\beta} - \hat{\beta}$ is asymptotically normal with the mean-zero influence function $\psi(W,X) = \phi(W,X) - \varphi(W,D)$:

$$\sqrt{n}\left(\tilde{\beta} - \hat{\beta}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(W_i, X_i) + o_P(1),$$

and the asymptotic variance $V = \lim_{n\to\infty} Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(W_i, D_i)\right)$. When the parametric estimator $\hat{\beta}$ is efficient under the null, we can appeal to the classic result of Hausman (1978) to express the V as the difference of the asymptotic variances of $\tilde{\beta}$ and $\hat{\beta}$. In general, however, assuming efficiency of the parametric estimator is not desirable. Moreover, $\varphi(W,X)$ depends on the specific parametric estimator and $\phi(W,X)$ involves estimation of a joint density function. Alternatively, we consider bootstrap the non-studentized statistic $\sqrt{n}\left(\tilde{\beta} - \hat{\beta}\right)$. Let $\hat{\theta} = \hat{\hat{\beta}} - \hat{\beta}$. Let J_n^* denote the bootstrap distribution of $\sqrt{n}\hat{\theta}$. We consider the bootstrap confidence region

$$\mathcal{B}_n(\alpha) = \{\theta \in \Theta : \hat{\theta} - n^{-1/2}\inf\{t : J_n^*(t) \geq \alpha/2\} \leq \theta \leq \hat{\theta} + n^{-1/2}\inf\{t : J_n^*(t) \geq 1 - \alpha/2\}\}.$$

The following result shows that this bootstrap confidence region is asymptotically valid.

Proposition 2. *Suppose Assumptions 4 and conditions of Theorem 1 hold, and 5 is satisfied.*

Additionally, (a) F has a finite support, (b) $0 < Var(\psi(W, X)) < \infty$, (c) $E_n[\psi^2] \to E[\psi^2]$, (d) $\limsup_n n^{1/2}\|F_n - F\|_\infty < \infty$, then we have, as $n \to \infty$:

1. *Under the null,*
$$P\{0 \in \mathcal{B}_n(\alpha)\} \to 1 - \alpha.$$

2. *Under the alternative,*
$$P\{0 \notin \mathcal{B}_n(\alpha)\} \to 1.$$

Proof. By Theorem 1, we have under both the null and the alternative,

$$\beta = E[\partial_W h(W, X, \varepsilon; \theta)] = E[\partial_W E[Y|W, X]] = E[\partial_W g(W, X)].$$

Then, under the null and Assumption 4 as well as condition 5, we can write

$$\begin{aligned}\hat{\hat{\beta}} - \hat{\beta} =& \hat{\hat{\beta}} - E[\partial_W g(W, X)] + \beta - \hat{\beta} \\ =& \int \phi(W, X) d(F_n - F) + \int \varphi(W, X) d(F - F_n) + o_P\left(n^{-1/2}\right) \\ =& \int \psi(W, X) d(F_n - F) + o_P\left(n^{-1/2}\right)\end{aligned}$$

Under the extra condition listed in the statement, we can apply Theorem 1.6.3 of Politis et al. (1999), which concludes the first statement.

Under the alternative, we have

$$\hat{\hat{\beta}} - \hat{\beta} - B = \int \psi(W, X) d(F_n - F) + o_P\left(n^{-1/2}\right)$$

Let $\mathcal{B}_n(\alpha) - B = \{\theta - B : \theta \in \mathcal{B}_n(\alpha)\}$. Then, applying Theorem 1.6.3 of Politis et al. (1999) to the reentered difference $\hat{\hat{\beta}} - \hat{\beta} - B$ gives,

$$P\{0 \in \mathcal{B}_n(\alpha) - B\} \to 1.$$

As $n \to \infty$, $n^{-1/2} \inf\{t : J_n^*(t) \geq 1 - \alpha/2\} - n^{-1/2} \inf\{t : J_n^*(t) \geq \alpha/2 \to 0$ because $J_n^*(t)$ is an asymptotical normal distribution by Theorem 1.6.3 of Politis et al. (1999). Since $B > \varepsilon > 0$, we have $P\{0 \notin \mathcal{B}_n(\alpha)\} \to 1$, which completes the proof. $\square$

4 Simulation

In this section, we use Monte Carlo simulations to demonstrate in finite sample: (1) the severe bias due to the endogenous control, (2) how our constructive identification results are immune to the endogeneity of controls, and (3) coverage and power of the proposed test of the endogenous control. We consider two data generating processes (DGPs) that corresponds to the two scenarios covered in Sections 2.1 and 2.2.

4.1 Endogenous control bias and the remedy

The first DGP covers the scenario where both the treatment and the controls are endogenous while the treatment is conditionally independent of the unobserved determinants:

$$\begin{aligned}\text{DGP(1): } Y = \tau W + \xi_1 X_1 + \xi_2 X_2 + U, \quad W = \frac{1}{2}\exp(a) + N(0,1)\\ X_1 = a + \frac{1}{2}\exp(p), \quad X_2 = p, \quad U = \frac{1}{2}\exp(b) + \frac{1}{2}\exp(q) + N(0,1),\end{aligned}$$

where $\tau = \pi_1 = \pi_2 = 1$; (a, b) and (p, q) are independent of each other, and each are jointly normal with mean zero, variance one, and covariance $\rho = 0.75$; $N(0,1)$ denotes a random draw from a standard normal distribution, independent of (a, b) and (c, d). We observe that (1) $X = (X_1, X_2)'$ are relevant for both Y and W; (2) W and X are dependent on U; (3) Conditional on X, W is independent of U; and (4) W and X are measurably separated. As a result, linear projection parameters does not identify τ. Meanwhile, τ is also the average response of W, and it is nonparametrically identified by Theorem 1.

The second DGP covers the scenario where the IV is only conditionally valid and the control is endogenous:

$$\begin{aligned}\text{DGP(2): } Y = \tau W + \pi_1 X_1 + \pi_2 X_2 + U, \quad W = X_1 + X_2 + Z + \eta\\ X_1 = a + \frac{1}{2}\exp(p), \quad X_2 = p, \quad Z = \frac{1}{2}\exp(a) + N(0,1)\\ \eta = \frac{1}{2}\exp(\xi), \quad U = \frac{1}{2}\exp(b) + \frac{1}{2}\exp(q) + \frac{1}{2}\exp(\zeta) + N(0,1),\end{aligned}$$

where (ξ, ζ) are also jointly normal with mean zero, variance one, and covariance $\rho = 0.75$. We observe that (i) W is not conditionally independent given X; (ii) Z is a valid IV only when conditional on X, but X is endogenous; (iii) Z and X are measurably separated. (iv) The measurable separability between Z and X here also implies the measurable separability between W and (X, η). As a result, τ is not identified by the usual IV projection, while τ is again, as an average response, nonparametrically identified: constructive identification is given by (2) or footnote 2 when the outcome equation is linear, and Theorem 1 gives a more general approach when the outcome equation is nonparametric and nonseparable.

Table 1: Simulation for DGP(1)

ρ	Methods	OLS w.o. X	OLS w. X	Series
0.75	Bias	0.732	0.156	0.013
	SD	0.091	0.080	0.048
	MSE	0.545	0.031	0.002
0.5	Bias	0.588	0.088	0.004
	SD	0.083	0.063	0.054
	MSE	0.352	0.012	0.003
0	Bias	0.383	0.000	0.000
	SD	0.065	0.045	0.058
	MSE	0.151	0.002	0.003

Note: Simulation results are based on $10,000$ replications and random samples of size $n = 1000$. Series regression uses 3rd-order polynomials.

Table 1 compares the estimates of τ using (i) OLS without control, (ii) OLS with control, (iii) the nonparametric estimations through the third-order polynomial series regression. The results are clear: while conventional methods that assume the exogeneity of controls are severely biased, the nonparametric methods perform much better in terms of bias.

Table 2 compares the following IV without control, IV with control, nonparametric approach as in footnote 2 using series regression, two-step control function approach in Theorem 1 using series regression for the second step. The results are as expected by theory. The IV-based methods that implicitly impose the exogeneity of the controls fail to produce consistent estimates of the partial effects of interest, and the alternative identification methods paired with usual nonparametric estimators perform well in finite sample.

4.2 Test of endogenous controls

In this section, we will focus on DGP(1) and demonstrate the finite sample performance of the proposed test in terms of the empirical coverage and power.

In the setting of DGP (1), the covariance parameter ρ controls the degree of endogeneity

Table 2: Simulation Results for DGP(2)

ρ	Methods	IV w.o. X	IV w. X	Series 1	Series 2
0.75	Bias	0.528	0.156	0.013	0.020
	SD	0.057	0.083	0.059	0.063
	MSE	0.282	0.031	0.004	0.004
0.5	Bias	0.424	0.089	0.005	0.001
	SD	0.052	0.068	0.064	0.069
	MSE	0.182	0.012	0.004	0.005
0	Bias	0.277	0.009	0.005	0.004
	SD	0.044	0.052	0.068	0.073
	MSE	0.079	0.003	0.005	0.005

Note: Simulation results are based on $10,000$ replications and random samples of size $n = 1000$. The conditional CDF is estimated based on a grid of 10 quantiles between 0.05 and 0.95 using the empirical support of W. Series 1 corresponds to the constructive identification in footnote 2 and Series 2 corresponds to the approach due to Theorem 1. Both series approaches use the 3rd-order polynomials.

in controls. As a result, the null and alternative of the test can be translated as

$$H_0 : \rho = 0,$$
$$H_1 : \rho \neq 0.$$

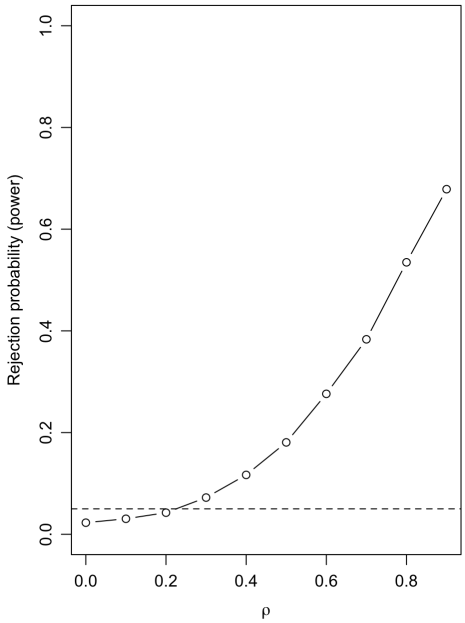
Therefore, we can simulate the coverage probability under the null $\rho = 0$, and we can choose a grid for $\rho > 0$ to simulate the power of the test.

Table 3 reports the empirical coverage under different sample sizes and number of Bootstrap replications. We find that (1) the coverage gets closer to the nominal rate as the sample size increases, and test tends to be conservative when the sample size is small; (2) The increment of Bootstrap replications from 500 to 1000 does not matter much.

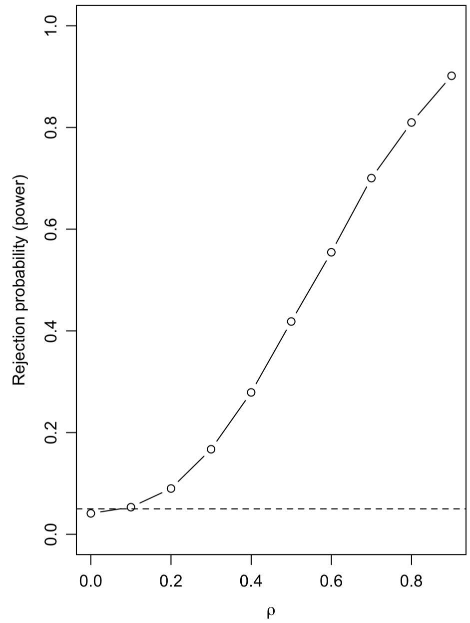
We also provide the empirical power of the test along a sequence of alternatives characterized by ρ. We find the test is slightly conservative for small values of ρ and when the sample size is small.

5 Conclusion

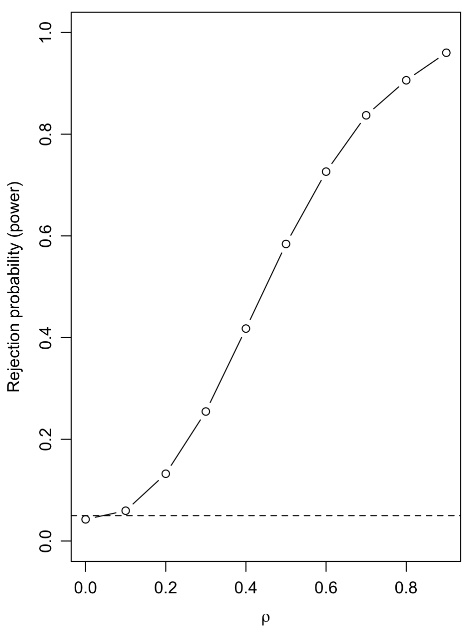
This note addresses a critical, prevalent, yet often overlooked problem in empirical research: the endogeneity of control variables. Building on the insightful observation and discussion in



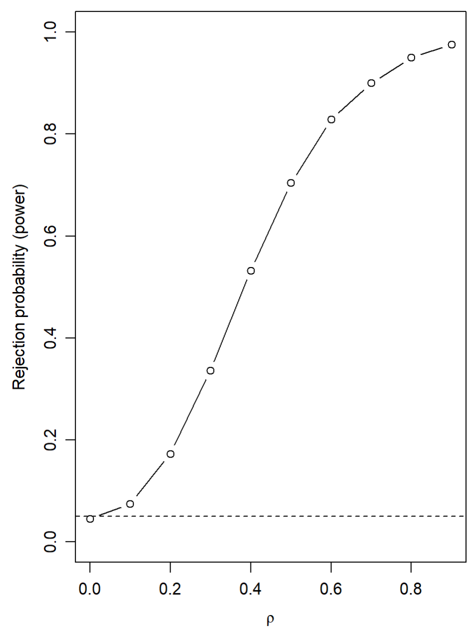
(a) Power with $n = 500$, $B = 500$



(b) Power with $n = 1000$, $B = 500$



(c) Power with $n = 500$, $B = 1000$



(d) Power with $n = 1000$, $B = 1000$

Figure 1: Power of the test with different sample sizes and bootstrap replications.

Note: Simulation results are based on $10,000$ replications. The test is based on the difference between the OLS estimator with controls and the series estimator with the 3rd-order polynomials.

Table 3: Coverage probabilities with a nominal rate 0.95

Sample size	1000		500		100	
Bootstrap reps	1000	500	1000	500	1000	500
Coverage(%)	96.6	95.9	97.3	97.7	99.9	99.9

Note: Simulation results are based on $10,000$ replications and random samples of size $n = 1000$. The series approach uses the 3rd-order polynomials.

Frölich (2008) that nonparametric estimation can help with the endogenous control problem, we provide constructive identification results for marginal effects in a simple linear model with or without the presence of IVs, and extend the results to a general class of nonseparable models.

Our results not only provide solutions for identifying marginal effects of the treatment in the presence of endogenous controls but also have important implications: Because the additional measurable separability condition barely adds any practical restrictions, endogenous controls are generally innocuous for nonparametric models. Furthermore, based on the identification results, we propose a test for endogenous controls.

For empirical studies, our results invite researcher to conduct a robustness check using nonparametric methods and the proposed test. In general, nonparametric approaches are more robust, not only due to specification concern but also in light of endogenous controls.

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.

Andrews, I., Barahona, N., Gentzkow, M., Rambachan, A., and Shapiro, J. M. (2025). Structural estimation under misspecification: Theory and implications for practice. *The Quarterly Journal of Economics*, 140(3):1801–1855.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.

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*. 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.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, pages 1251–1271.

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.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society*, pages 1349–1382.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer.

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

Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach*. Cengage Learning.