Identification of Partial Effects with Endogenous Controls

Kaicheng Chen* Kyoo il Kim†

February 2026

Abstract

While control variables are often assumed to be exogenous, it is common to encounter endogenous controls in practice. This creates a dilemma: without controls, the treatment of interest may be endogenous; with controls, the endogeneity of the controls may contaminate identification. The problem may not be solved with an instrumental variable when it is only conditionally valid. For these scenarios, we establish identification of partial effects of the treatment under a rank condition called measurable separability. For parametric models, our approach amounts to employing a nonparametric estimator of the nonseparable model that nests the parametric specification. For nonparametric models, our results show that endogenous controls are generally innocuous under standard identification conditions, except in the case of "bad controls". We further propose a test for endogenous controls. Simulation results and an empirical application demonstrate this prevalent issue and provide practical implications of our methods.

Keywords: endogenous control, average response, nonseparable model.

JEL Classification Code: C14, C31

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

1 Introduction

We study a critical and prevalent, yet often overlooked, issue in empirical research: the endogeneity of control variables. In models with endogenous treatments, researchers often leverage a conditional independence assumption (CIA) or instrumental variables (IV) to identify treatment effects, while rather casually assuming the control variables are exogenous.

In applications, additional control variables are often included either because they are relevant for both the treatment variable and the outcome, or because the IV is more likely to be valid when controls are included. However, empirical researchers often end up with control variables that are subject to additional endogeneity concerns, while finding instruments for every endogenous control can be challenging or impossible.

In this paper, we demonstrate how endogeneity in control variables affects the identification of partial effects of the treatment and, when it does, how the issue can be addressed in certain scenarios. To fix the idea, suppose we are given a parametric model

$$Y = h(W, X, \varepsilon; \theta) \tag{1}$$

where h is known up to unknown parameters θ. W is the treatment/policy variable of interest and X is a vector of control variables. Throughout the paper, W takes scalar values for notational convenience. For now, suppose Y and W are continuous random variables, and $h(w, x, \epsilon; \theta)$ is differentiable with respect to w. Suppose a researcher is interested in the average response with respect to W, defined as

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

Throughout the paper, we consider the parametric model as correctly specified in the sense that there exists θ such that 1 holds almost surely[1]. However, we remain agnostic about the dependence between the X and ε. The goal is to establish identification of β under CIA

$$\varepsilon \perp\!\!\!\perp W | \widetilde{X}$$

where $\widetilde{X}$ is a collection of control variables such that $X \subseteq \widetilde{X}$ and $\widetilde{X}$ may contain extra variables excludable from the outcome equation (1).

[1] θ needs not to be unique since we are only interested in identifying β.

As we illustrate in the examples below, CIA alone does not suffice for identification of β without further restrictions, for example, when $\widetilde{X}$ is endogenous. Instead of imposing the exogeneity of $\widetilde{X}$, we establish identification through an intuitive rank condition associated with W and $\widetilde{X}$, effectively allowing controls to be endogenous.

Notation. Throughout the paper, we let $F(A, B, C)$, $F(A, B)$, and $F(A)$ denote the distribution functions of (A, B, C), (A, B), and A; $F_{A|B,C}$ denote the distribution functions of A conditional on B, C. We denote the density functions f similarly. Whenever a density is invoked, we assume it exists and is well-defined. $A \perp\!\!\!\perp B \mid C$ denotes the independence of A and B conditional on C. The supports of the random elements A are denoted by their calligraphic letters $\mathcal{A}$. We use P to denote the probability measure and P^* to denote the bootstrap probability measure conditional on the observed data.

1.1 Illustration in a Linear Model

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

$$Y = \tau W + X'\xi + \varepsilon, \;\; E[\varepsilon|W, X] = E[\varepsilon|X], \tag{2}$$

In this model, the average response with respect to W is given by $\beta = \tau$. 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 functional form of X is misspecified and some nonlinear effects of X end up being part of ε, then $E[\varepsilon|X] \neq 0$. Generally, this is more than a specification issue, although, as we show below, in many cases this can be mitigated by a flexible functional form.

There are two consequences of the dependency. First, without considering X, it may cause omitted variable bias. Second, with the presence of X, the dependence between X and ε can also pollute the identification of 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. To see the bias in the second case, let $D = (W, X')'$ and $\theta = (\tau, \xi')'$, the linear projection parameters are defined as follows:

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

Therefore, without further restriction on $E[\varepsilon|X]$, θ, or τ in particular, are not identified

and OLS would produce a biased and inconsistent estimator. 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, CIA 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 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[2], it rules out situations where CIA is not likely to hold.

The idea is as follows. Due to (1), there exists a nonparametric and nonseparable nesting model $m(W, X, \varepsilon)$ such that

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

It follows that the average response under the parametric specification h equals that under the nesting nonparametric model m:

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

It is shown later that, under CI given $\widetilde{X}$ as well as the measurable separability between W and $\widetilde{X}$,

$$E[\partial_W m(W, X, \varepsilon)] = E[\partial_W E(Y|W, \widetilde{X})] \tag{4}$$

In this case, by taking $\widetilde{X} = X$,

$$\tau = \beta = E[\partial_W m(W, X, \varepsilon)] = E[\partial_W E(Y|W, X)].$$

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

In this linear case, it is straightforward to examine $E[\partial_W E(Y|W,X)]$ directly. Under the same CI and the measurable separability as above,

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

where the last equality holds due to the measurable separability between 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, the last equality does not hold anymore.

Note that the differentiability with respect to W is not essential. Given the conditional mean independence and the measurable separability of W and X, the identification result extends to nonparametric regression models such as

$$Y = f(W) + h(X) + \varepsilon,\ E[\varepsilon|W,X] = E[\varepsilon|X].$$

From the conditional mean function

$$E[Y|W=w,X=x] = f(w) + h(x) + E[\varepsilon|W=w,X=x] = f(w) + h(x) + E[\varepsilon|X=x],$$

$f(w)$ is identified due to the measurable separability, even if $h(x)$ is not identified (i.e. $E[\varepsilon|X=x] \neq 0$). This result does not require differentiability of $E[Y|W=w,X=x]$ or $f(w)$.

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 CI, 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

expend at 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.

1.2 Does an IV Approach Solve the Problem?

When the instrument is genuinely exogenous and affects the outcome solely through the treatment, the answer is yes. However, in practice, control variables are often included to either increase precision or to make the IV assumptions more plausible. In these cases, the endogeneity of these controls can again jeopardize identification. To illustrate, let's consider the linear model with an excludable IV:

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

In this model, the average response with respect to W is also τ. However, differing from the previous example, W is not conditionally independent of ε even after conditioning on X. Meanwhile, without controlling X, Z may not be a valid IV if Z affects Y through X too. Therefore, endogeneity in X creates the same dilemma, and neither τ nor ξ is identified by the usual IV or 2SLS projection.

Nevertheless, τ can be nonparametrically identified through a control function approach[3]: If Z is independent of (ε, η) conditional on X, then W is independent of ε conditional on X and η. Now, by taking $\widetilde{X} = \{X, \eta\}$ and assuming $\widetilde{X}$ is measurably separated from W[4], the general result (4) implies

$$\tau = E[\partial_W E(Y|W, X, \eta)].$$

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

[4] 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.

This can also be obtained by examining $E[\partial_W E(Y|W, X, \eta)]$ directly:

$$\begin{aligned}&\int_{\mathcal{W}\times\mathcal{X}\times\mathcal{A}} \partial_w E[Y|W = w, X = x, \eta = a]dF(w, x, a)\\ =&\tau + \int_{\mathcal{W}\times\mathcal{X}\times\mathcal{A}} \partial_w E[\varepsilon|X = x, \eta = a]dF(w, x, a) = \tau. \end{aligned} \tag{5}$$

Example 2. Consider a linear triangular model relating the individual wage to 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 subject to critiques such that *qbirth* may not be truly exogenous because, for example, the family income level could affect conception planning[5], which in turn affects birth. Therefore, researchers 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 wages varies and can be further correlated with other socioeconomic determinants, which makes it a 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 (5).

1.3 More General Settings

As illustrated in the examples above, nonparametric methods appear to be more robust to endogenous controls. This is first argued in Frölich (2008), but many questions remain: First, to what extent are the nonparametric methods immune to the endogenous control? For example, some extra regularity conditions may be needed, such as the aforementioned measurable separability. Furthermore, we ask how broad the class of models is in which nonparametric methods remain valid in the presence of endogenous controls. If we can define

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

such an admissible class of models, then the next question is, should researchers always use nonparametric methods? It is well-known that nonparametric methods can be less efficient, at the cost of being robust to specification, and, in this case, to endogenous control. Ideally, we would like a testing procedure that detects potentially endogenous controls in a broad class of models.

In this paper, we study a broad class of parametric models that are nested in a nonseparable and nonparametric model, with controls that are potentially endogenous. We show that under CIA and the measurable separability condition, the identification of the average response associated with the treatment/policy variable is available for the nesting model even with endogenous controls. Since the average response of the nested parametric models is equivalent to the average response of the nesting nonseparable and nonparametric model, researchers can instead resort to the latter one to avoid the bias caused by endogenous controls. Note that the CIA can be either achieved by conditioning on the observable control variables or through the control function approach when there exists excludable exogenous variables.

Based on these results, we further propose a test for endogenous controls. For linear models, our test reduces to a Hausman-style specification test. For more general setups, the test follows the same principle by checking agreement between a restricted estimate and a more robust estimate of the same partial effect. Particularly, we focus on the average derivative estimator through the series approximation and inference based on bootstrap. We examine the size and the power of the test both theoretically and through simulation.

In the simulation study, we present some examples of data generating processes that feature endogeneity in controls. By contrasting our approach and methods that does not take into account the endogenous controls, we highlight the practical importance of this issue and document the finite-sample performance of our procedure. To further illustrate our methods and draw practical implications, we revisit a classic study of the import-competition impact on labor market outcomes by Autor et al. (2013). While adding control variables can alleviate the bias from the concern of IV validity,it may not address—and can even mask—bias arising from endogeneity in the controls themselves. Our approach and test together provide a simple robustness check for detecting this additional source of bias.

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) assesses 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 of the average response under CIA and the measurable separability are given in Section 2. The endogenous control test is proposed in Section 3. Section 4 presents the simulation study that compares our methods and those not robust to endogenous controls, and examines the finite sample performance of endogenous control test. Section 5 provides the empirical application to illustrate the practical implications of our methods. Section 6 concludes the paper with empirical recommendations.

2 Nonseparable Models with Endogenous Controls

To investigate the impact of endogenous control variables in a general setting, we consider a broad class of parametric models as (1). Let $m(W, X)$ as a nesting nonseparable and nonparametric model,

$$Y = m(W, X, \varepsilon), \tag{6}$$

A similar nonparametric and nonseparable model has been studied by Altonji and Matzkin (2005) except that they consider X as the excluded instruments and X do not enter the outcome equation.[6] For a continuous treatment W and a continuous outcome Y, the average response β associated with W can be defined as

$$\beta = \int \beta(w, x) dF(w, x),$$
$$\text{where} \quad \beta(w, x) = \int \partial_w m(w, x, \epsilon) f_{\varepsilon|W=w,X=x}(\epsilon) d\epsilon$$

[6]Our main motivation for model (6), which differs from Altonji and Matzkin (2005), is that empirical researchers often seek to include X as control variables in the outcome equation.

If Y is a binary outcome, we can write

$$m(W, X, \varepsilon) = 1\{m^*(W, X, \varepsilon) > 0\},$$

where m^* is implicitly defined. Following 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, for fixed (w, x, v), $m^*(w, x, u, v)$ has at least one root in u 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 differentiable in w. Then, we can define the average response β^* associated with a binary outcome as

$$\beta^* = \int \beta^*(w, x) dF(w, x),$$
$$\text{where} \quad \beta^*(w, x) = \int -\partial_w u^*(w, x, v) f_{u,v|W=w,X=x}(u^*(w, x, v), v) dv.$$

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

Example 3. Consider a parametric nonseparable model $Y = h(W, X, \varepsilon; \theta) = 1\{\tau W + X'\xi + \varepsilon > 0\}$ where $\theta = (\tau, \xi)$ and $\varepsilon | W, X \sim N(\mu(W, X), \sigma(W, X)^2)$; $\mu(W, X)$ and $\sigma(W, X)^2$ are the conditional mean and variance since we don't assume ε is independent of (W, X). Suppose we can partition ε as $\varepsilon = u + v$, and v is also Gaussian. Note that $-u^*(W, X, v) = \tau W + X'\xi + v$, then $-\partial_w u^*(w, x, v) = \tau$ and

$$\begin{aligned}
&\beta^*(w, x) = \tau \int f_{u,v|W=w,X=x}(-(\tau w + x'\xi + v), v) dv \\
=&\tau \int f_{\varepsilon,v|W=w,X=x}(-(\tau w + x'\xi), v) dv = \tau f_{\varepsilon|W=w,X=x}(-(\tau w + x'\xi)) \\
=&\frac{\tau}{\sigma(w, x)} \phi\left(\frac{\tau w + x'\xi - \mu(w, x)}{\sigma(w, x)}\right) \qquad (7)
\end{aligned}$$

where the second equality uses the fact that $\varepsilon = u + v$ and the change-of-variables. If $\varepsilon | W, X \sim N(0, 1)$, then $\beta^*(w, x) = \tau\phi(\tau w + x'\xi)$, and $\beta^* = \tau E[\phi(\tau W + X'\xi)]$ reduces to the usual average partial effect (APE) of W in a probit model.

2.1 Identification under CIA

Let $\widetilde{X}$ denote a collection of control variables such that $X \subseteq \widetilde{X}$. We consider the scenario where ε is conditionally independent of W given $\widetilde{X}$, but the dependence between X and ε is unrestricted.

Assumption 1. $\varepsilon \perp\!\!\!\perp W \mid \widetilde{X}$.

Assumption 1 is a key condition for identifying the average response. Importantly, it does not rule out $X \subseteq \widetilde{X}$, or X in particular, being endogenous. We emphasize that Assumption 1 itself may not be sufficient for identification due to the endogeneity of X: for example, if W is some function of X only, which is dependent on ε, then the average response would not be identified by Assumption 1, and, in which case, even Assumption 1 itself is not likely to hold. Thus, an additional restriction is needed to rule out such extreme cases. The next theorem shows that the measurable separability condition between W and $\widetilde{X}$ suffices for the identification of the average response.

Theorem 1. *Suppose W and $\widetilde{X}$ are measurably separated and Assumption 1 holds.*

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

2. *If Y is binary, suppose $m^*(w, x, u, v)$ has at least one root and is strictly increasing in u; $m^*(w, x, u, v)$ is continuously differentiable in w and u, and $\partial_u m^*(w, x, u^*, e) \neq 0$. Then, $\beta^* = E\left[\partial_W E\left[Y|W, \widetilde{X}\right]\right]$.*

A proof is provided in Appendix. Theorem 1 shows that even if ε and X are potentially dependent, the average response associated with the treatment can still be identified as long as CIA and the measurable separability condition hold. By definition, measurable separability simply requires that W and X are not exclusively determined by each other, which is 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 is guaranteed.

Theorem 1 is a positive result: since the measurable separability condition is a very mild rank condition that holds in most settings, it justifies the prevalent use of potentially endogenous control variables. Meanwhile, we have seen in previous examples that parametric models are not immune to the endogeneity of control, so this result encourages the use of

fully flexible models in these scenarios.

Example 3 continued. For the binary model $Y = 1\{\tau W + X'\xi + \varepsilon > 0\}$, suppose $\varepsilon|W, X \sim N(\mu(X), \sigma(X)^2)$, i.e. CIA holds for this model with $\widetilde{X} = X$. Following the calculation as (7), the true average response is given by

$$\beta^* = E[\beta^*(W, X)] = E\left[\frac{\tau}{\sigma(X)}\phi\left(\frac{\tau W + X'\xi - \mu(X)}{\sigma(X)}\right)\right].$$

By Theorem 1, β^* is identified as $\beta^* = E[\partial_W P(Y = 1|W, X)]$ as long as W is measurably separated from X. Indeed, $E[\partial_W P(Y = 1|W, X)] = E\left[\frac{\tau}{\sigma(X)}\phi\left(\frac{\tau W + X'\xi - \mu(X)}{\sigma(X)}\right)\right]$ under CIA. However, ignoring the endogeneity of the controls with standard probit approach would impose $P(Y = 1|W, X) = \Phi(\tau W + X'\xi)$ in constructing the likelihood function, which would fail to identify the true average response.

2.2 Identification with IV

When the CIA is not plausible given observable controls X, it is common to consider exogenous and excludable IVs for identification. In applications, additional control variables are often included to make the exogeneity condition of IVs more likely to hold. Although control variables are explicitly or implicitly assumed to be exogenous, we caution that they may be endogenous in practice, while finding IVs for all endogenous controls is not possible. In this section, by utilizing excludable exogenous variables, we construct an extra control variable V following the approach by Imbens and Newey (2009), such that CIA is satisfied given $\widetilde{X} = \{X, V\}$. Thus, combining with the results in the previous section, the average response of the treatment can be identified with using IVs while while explicitly allowing for endogenous controls.

The researcher is given with the parametric model (1), nested in (6), and the average response parameter β is defined the same way, except now the CIA is not available. Instead, suppose there exists exogenous and excludable variables IVs Z such that:

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

Note that we don't impose the 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, the goal is to find a control variable V such that

$$\varepsilon \perp\!\!\!\perp W \mid \widetilde{X} = \{X, V\}, \tag{8}$$

and $\widetilde{X}$ is measurably separated from W. Then, we can identify the average response 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 $F_{\eta|X}(a)$ is continuous and strictly increasing in $a \in supp(\eta)$ almost surely.

Theorem 2. *Suppose Assumption 2 holds for the nonseparable model* (6)*. 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 (8) is satisfied with $V = F_{W|Z,X}(W)$.

A proof can be found in Appendix.

By Proposition 2, CIA is satisfied with $\widetilde{X} = \{X, V\}$. Therefore, the identification of the average response associated with the treatment can be obtained by Theorem 1.[7]

3 A Test for Endogenous Controls

In this section, we consider a test for the endogenous control for a large class of models, as an implication of the results presented in Section 2. Consider the parametric model $Y = h(W, X, \varepsilon; \theta)$, nested by the nonparametric and nonseparable model $Y = m(W, X, \varepsilon)$. To focus on the main idea, we limit our attention to the case where $\widetilde{X} = X$ are observable, and we observe an i.i.d sample of size n.[8].

[7]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.

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

Regardless of the continuous or binary outcomes, the average response is universally defined for parametric models and is equivalent to the average response defined using the nesting nonparametric model. In the linear model (2), $\beta = \tau$. In the binary model of Example 3, $\beta^* = E\left[\frac{\tau}{\sigma(W,X)}\phi\left(\frac{\tau W + X'\xi - \mu(W,X)}{\sigma(W,X)}\right)\right]$. Under CIA and the measurable separability, both are identified by Theorem 1. However, 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, the parametric estimators corresponding to these specifications 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:

$$H_0 : X \text{ is exogenous,}$$
$$H_1 : X \text{ is endogenous.}$$

The null and alternative hypotheses are specified loosely to accommodate a broad 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 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. We will make the hypotheses concrete in a moment when we introduce the assumptions for the estimators under consideration.

To represent both binary and continuous outcomes, we denote the average response as $\beta_0 := E[\partial_W h(W, X, \varepsilon; \theta)]$ for both cases for the remainder of this section. Let $\hat{\beta}_p$ denote the parametric estimator under model 1, but treating X as exogenous, i.e. imposing H_0. Due to the identification results from last section, we consider an average derivative estimator $\hat{\beta}_{np}$: let $g(W, X) := E[Y|W, X]$, and

$$\hat{\beta}_{np} = \frac{1}{n}\sum_{i=1}^{n} \partial_W \hat{g}(W_i, X_i) \tag{9}$$

where $\hat{g}$ is some nonparametric estimator for g.

We define the limit of these two estimators as β_p and β_{np}, respectively. With an appropriate choice of $\hat{g}$ and sufficient regularity conditions, the average derivative estimator is consistent for $\beta_{np} = E[\partial_W g(W, X)]$ (e.g., see Example 3 of Newey, 1994). Under the

conditions of Theorem 1, $\beta_{np} = \beta_0$, regardless of H_0 or H_1. Meanwhile, under the same conditions, the limit β_p may change across data generating processes because the parametric model is correctly specified up to the exogeneity restrictions on the controls.

If, under H_0, $\beta_p = \beta_0$, and both estimators are asymptotically linear with some influence functions φ and ϕ, then we can define $\psi = \varphi - \phi$, and write

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

which, under common regularity conditions, is asymptotically normal with the asymptotic variance $V = \lim_{n\to\infty} Var\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i, X_i)\right)$. 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, since $\varphi(W, X)$ and $\phi(W, X)$ depends specific parametric estimator, and $\phi(W, X)$ may depend on extra infinite-dimensional nuisance estimation, the analytical formula for V is tricky. Alternatively, we can resort to a bootstrap approach, which is a common and general approach for inference that involves semiparametric estimation; see, for example, Chen et al. (2003).

Therefore, with $\theta_0 = \beta_p - \beta_{np}$ and $\hat{\theta} = \hat{\beta}_p - \hat{\beta}_{np}$, we consider bootstrap the non-studentized statistic $\sqrt{n}\left(\hat{\theta} - \theta_0\right)$. Let J_n^* denote the empirical bootstrap CDF of $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$ where $\hat{\theta}^*$ is the bootstrap estimator for $\hat{\theta}$. We consider the bootstrap confidence region

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

A brief algorithm for implementation of the test is given as follows.

Algorithm 1.

- Step 1: Based on the parametric model (1), obtain $\hat{\beta}_p$ while treating X as exogeneous. Calculate $\hat{\beta}_{np}$ as (9) with some nonparametric estimator $\hat{g}$ for $E[Y|W, X]$. Let $\hat{\theta} = \hat{\beta}_p - \hat{\beta}_{np}$.
- Step 2: Resample independently and with replacement from P_n with equal probability $1/n$ for each observation. Obtain the bootstrap sample $\{Y_i^*, W_i^*, X_i^*\}_{i=1}^n$.
- Step 3: Obtain $\hat{\beta}_p^*$, $\hat{g}^*$, $\hat{\beta}_{np}^* = \frac{1}{n} \sum_{i=1}^{n} \partial_W \hat{g}^*(W_i^*, X_i^*)$, and $\hat{\theta}^* = \hat{\beta}_p^* - \hat{\beta}_{np}^*$ using $\{Y_i^*, W_i^*, X_i^*\}_{i=1}^n$.
- Step 4: Repeat Steps 2 - 3 for a sufficiently large B times and denote each bootstrap realization using subscript b. Obtain the empirical bootstrap CDF $J^*(t) =$

$\frac{1}{B}\sum_{b=1}^{B} 1\{\sqrt{n}(\hat{\theta}_b^* - \hat{\theta}) \leq t\}$. Obtain $\mathcal{B}_n(\alpha)$. Reject the null if $0 \notin \mathcal{B}_n(\alpha)$.

We introduce a set of high-level conditions in terms of $\hat{\beta}_p$, $\hat{\beta}_{np}$, as well as their limits. Let P^* denote the bootstrap probability measure conditional on the sample $\{Y_i, W_i, X_i\}_{i=1}^n$.

Assumption 4. (i) Under H_0, $\beta_p = \beta_0$; Under H_1, $\beta_p = \beta_0 + B_n$ for some real sequence B_n with $\sqrt{n}|B_n| \to \infty$.

(ii) There exists $\varphi(W, X)$ with $E[\varphi(W, X)] = 0$,

$$\hat{\beta}_p - \beta_p = \frac{1}{n}\sum_{i=1}^{n} \varphi(W_i, X_i) + o_P(n^{-1/2}). \tag{10}$$

$$\hat{\beta}_p^* - \hat{\beta}_p = \frac{1}{n}\sum_{i=1}^{n} \varphi(W_i^*, X_i^*) - \varphi(W_i, X_i) + o_{P^*}(n^{-1/2}). \tag{11}$$

(iii) There exists $\phi(W, X)$ with $E[\phi(W, X)] = 0$,

$$\hat{\beta}_{np} - \beta_{np} = \frac{1}{n}\sum_{i=1}^{n} \phi(W_i, X_i) + o_P(n^{-1/2}). \tag{12}$$

$$\hat{\beta}_{np}^* - \hat{\beta}_{np} = \frac{1}{n}\sum_{i=1}^{n} \phi(W_i^*, X_i^*) - \phi(W_i, X_i) + o_{P^*}(n^{-1/2}). \tag{13}$$

(iv) For $\psi = \varphi - \phi$, $0 < \mathrm{Var}(\psi(W, X)) < \infty$.

Assumption 4 (i) operationalizes the null and alternative hypotheses. For example, under the linear model (2), $\beta_0 = \tau$ and β_p is the first element of the linear projection parameter γ_{LP}, thus B_n is the first element of $E[DD']^{-1}E[DE[\varepsilon|X]]$ where $D = (W, X')'$. The upper bound in Assumption 4(iv) is a common finite variance condition for applying the central limit theorem. The lower bound excludes the case where the parametric estimator and the average derivative estimator agrees.

Assumption 4 (ii) and (iii) impose asymptotic linear representations for the estimators and their bootstrap counterparts. For parametric estimators, asymptotic linear representation is very common. For example, it holds for M-estimators with twice continuously differentiable objective functions. For some estimators with non-smooth objective functions, e.g. quantile estimators, there are also existing results established under extra regularity conditions. The asymptotic linear representation for the average derivative estimator follows from the general results for semiparametric estimators. Newey (1994) provides a general

approach for deriving the influence function for semiparametric estimators, and Example 3 of the same paper exemplifies the average derivative estimator using series approximation for the conditional expectation. See the next section for a concrete implementation of this estimator for simulation.

The asymptotic linear representations for bootstrap estimators expect slightly stronger regularity conditions. To see how this representation can be justified, we consider the process of the recentered scores $S_n(g) := \frac{1}{\sqrt{n}}\sum_{i=1}^n (\partial_W g(W_i, X_i) - E[\partial_W g(W_i, X_i)])$ and the bootstrap process $S_n^*(g) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\partial_W g(W_i^*, X_i^*) - \partial_W g(W_i, X_i))$. Also, define the pathwise derivative of $S_n(g)$ in the direction of $\tilde{g} - g$, evaluated at g, as $\Gamma(g)[\tilde{g} - g]$. In a semiparametric GMM framework, Theorem B of Chen et al. (2003) shows that under a slightly strengthen set of sufficient conditions for asymptotic normality of the GMM estimator, (13) can be obtained by

$$\|\hat{g}^* - \hat{g}\|_\infty = o_{P^*}(n^{-1/4}), \tag{14}$$

$$S_n^*(\hat{g}) + \Gamma(\hat{g})[\hat{g}^* - \hat{g}] = \sum_{i=1}^n (\phi(W_i^*, X_i^*) - \phi(W_i, X_i)) + o_{P^*}(1). \tag{15}$$

Condition (14) is also assumed in Chen et al. (2003) and holds for, for example, series estimators(Newey, 1997; Chen and Christensen, 2015) and kernel regression estimators(Hall, 1991). For Condition (14), we further denote $M_n(g) = \frac{1}{\sqrt{n}}\sum_{i=1}^n g(W_i, X_i)$ and $M_n(g)^* = \frac{1}{\sqrt{n}}\sum_{i=1}^n g(W_i^*, X_i^*)$. By adding and subtracting $M_n^*(\hat{g}^*) - M_n^*(\hat{g})$,

$$S_n^*(\hat{g}) + \Gamma(\hat{g})[\hat{g}^* - \hat{g}] = M_n^*(\hat{g}^*) - M_n(\hat{g}) + \Gamma(\hat{g})[\hat{g}^* - \hat{g}] - (M_n^*(\hat{g}^*) - M_n^*(\hat{g}))$$

Under regularity conditions for $\Gamma(\hat{g})[\hat{g}^* - \hat{g}]$ on P^*, we can apply results of Newey (1994) (Lemma 5.1) to obtain (15).

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

Theorem 3. *Under Assumptions 4 and conditions of Theorem 1, as $n \to \infty$, $P\{0 \in \mathcal{B}_n(\alpha)|H_0\} \to 1 - \alpha$ and $P\{0 \notin \mathcal{B}_n(\alpha)|H_1\} \to 1$.*

A proof can be found in Appendix. Theorem 3 formalizes the Hausman-type principle for the endogenous control testing. Unlike a specification test, the rejection here can occur even when the parametric model is correctly specified.

4 Simulation

In this section, we use Monte Carlo simulations to demonstrate in finite samples: (1) severe bias can be caused by the endogenous control, (2) how our approach are robust to the endogeneity of controls, and (3) empirical coverages and statistical powers of the proposed test of endogenous controls. We consider two data generating processes (DGPs) that correspond to the two scenarios covered in Sections 2.1 and 2.2.

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

$$\tilde{g} := \sum_{k=1}^{K} \tilde{g}_k' \pi_k = G'\Pi,$$

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

$$\tilde{g}_k = \text{vec}\left\{\prod_{j=1}^{d} q_j^{\lambda_j} : \sum_{j=1}^{d} \lambda_j = k, (q_1, ..., q_d) = \Big(l_1(W), l_2(\widetilde{X}_1), ..., l_d(\widetilde{X}_{d-1})\Big)\right\},$$

and $l = (l_1, ..., l_d)$ 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 associated average derivative estimator is defined as

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

where $\hat{\Pi}$ are the (penalized) least square estimators of Π. For our implementation in this section, we fix $K = 3$ and set $l(x) = x$.

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:

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

where $\tau = 1$; (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 (p, q). 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 do 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:

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

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 (5) or footnote 3 when the outcome equation is linear, and Theorem 2 gives a more general approach when the outcome equation is nonparametric and nonseparable.

For DGP (1), Table 1 compares the estimates of τ using (i) OLS without control, (ii) OLS with control, and (iii) $\hat{\beta}_{np}$ through the third-order polynomial series regression with $\widetilde{X} = X$. We compare these three approaches across three endogeneity levels denoted by ρ, with $\rho = 0$ corresponding to the exogenous control case and $\rho > 0$ for endogenous controls. The results are clear: while conventional methods that assume the exogeneity of controls are severely biased, the nonparametric methods are quite robust to the endogenous controls across the endogeneity levels.

Table 2 reports simulation results under DGP (2). The comparison is among the IV estimator without control, the IV estimator with control, nonparametric approach due to footnote 3 using the series approach, control function approach due to Theorem 2 with $\widetilde{X} = X, V$, where the conditional CDF V is estimated by the series approach, too, with 10

Table 1: Simulation for DGP(1)

ρ	Methods	OLS w.o. X	OLS w. X	$\hat{\beta}_{np}$
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

Notes: Simulation results are based on $10,000$ replications and random samples of size $n = 1000$.

quantiles grids between 0.05 and 0.95. The results also support our theories: The IV-based methods that implicitly impose the exogeneity of the controls fail to produce consistent estimates of the average response, and the proposed approaches are robust to endogenous controls with much less bias.

Table 2: Simulation Results for DGP(2)

ρ	Methods	IV w.o. X	IV w. X	Footnote 3	$\hat{\beta}_{np}$
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

Notes: 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.

4.2 Test for Endogenous Controls

In this section, we will focus on DGP(1) and demonstrate the finite sample performance of the proposed test in terms of empirical coverages and statistical powers. The implementation

of the test is based on Algorithm 1 from last section.

In the setting of DGP (1), the covariance parameter ρ controls the degree of endogeneity in controls. As a result, the null and alternative hypotheses 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 the coverage gets closer to the nominal rate as the sample size increases, and the test tends to be conservative when the sample size is small.

Figure 1 displays the empirical powers of the test along a sequence of alternatives characterized by ρ. Overall, the pattern is as expected by the theory and the finite sample performance in term of power is as desired. For small values of ρ and when the sample size is small, the test is slightly conservative.

Table 3: Coverage probabilities with a nominal rate 0.95

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

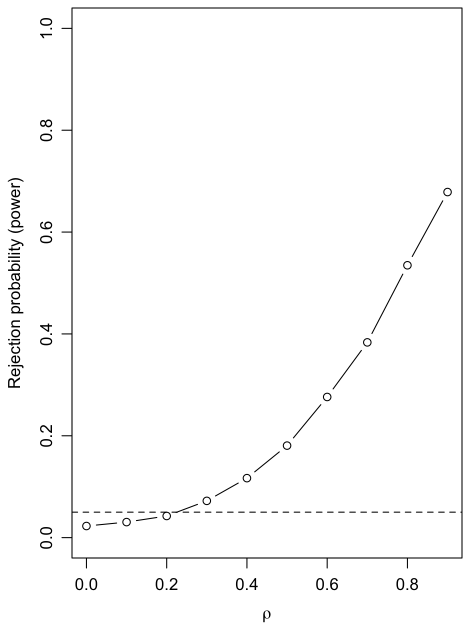
Notes: Note: Simulation results are based on $10,000$ replications. The series approach uses the 3rd-order polynomials.

5 Application: The China Syndrome Revisit

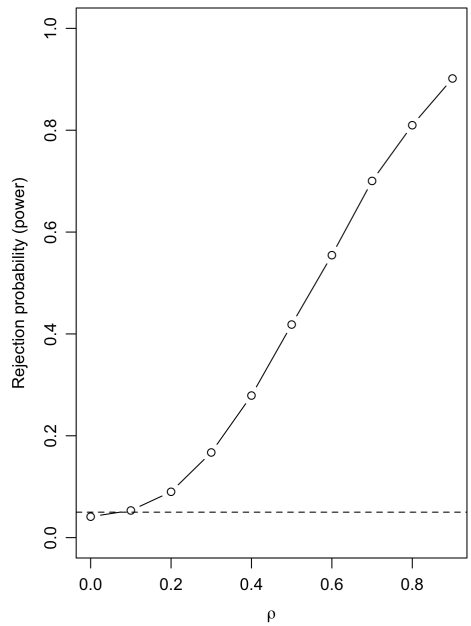
In this section, we revisit an empirical study of the import competition effect on the US labor market. During the late 20th and early 21st century, the world has witnessed a drastic surge of imports from the Chinese market. Meanwhile, there was a downturn of import-competing manufacturing in certain regions of the US, and it came with a higher unemployment rate and wage inequality in these regions. Therefore, a natural question arises: was the disruption of the local labor market mainly caused by the import competition, or was it rather a result of an overall economic transition in the US?

In a classic paper by Autor et al. (2013), the authors answer this question by exploiting the import-competition exposure variation across regional markets in the US. The idea is

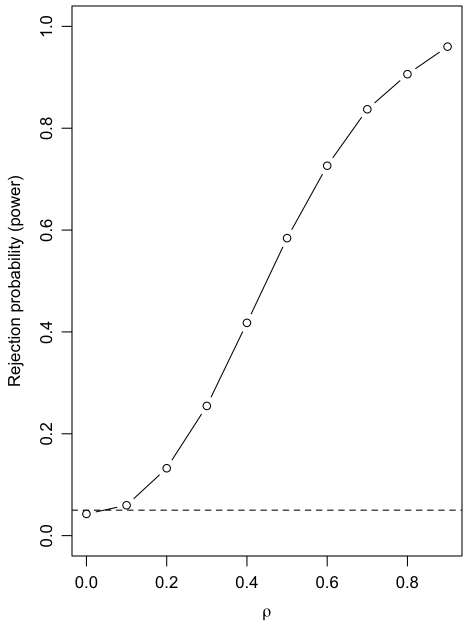
Figure 1: Power of the test with different sample sizes and bootstrap replications



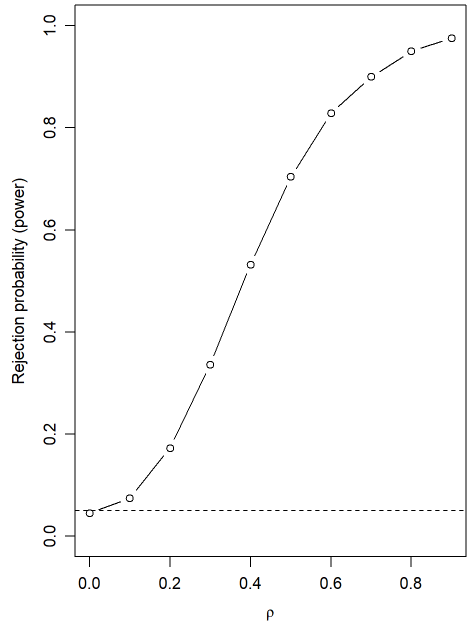
(a) Power with $n = 500$, $B = 500$



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



(c) Power with $n = 1500$, $B = 500$



(d) Power with $n = 2000$, $B = 500$

Notes: 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.

that regions with more initial specialization in labor-intensive industries are more exposed to the Chinese import competition. They measure the *share* of those industries at the start of the observation period and multiply it by the growth in US imports from China (*shift*), which produces a measure of import competition varied by regions. By taking commuting zones as analysis units, they are able to estimate effects on various labor market outcomes with a reasonable sample size.

There are two potential sources of endogeneity in this setup. Firstly, the import growth from China could be driven by unobserved local market transitions, such as industrial reallocation that causes a supply shortage. In that case, the growth in US imports from China could be driven by unobserved shocks that also move local outcomes. Secondly, the start-of-period measure of share in labor-intensive industries may also be related to other regional characteristics, such as educational attainment, which also determines the local labor market outcomes. To deal with the first type of endogeneity, they employ a measure of import growth in other high-income countries as an exogenous *shift* and multiply it by the same *share* variable to obtain the IV. They further assume that the growth in imports from China is not driven by demand shocks that are shared by the US and other high-income countries, so as to ensure the validity of the IV. For the second type of endogeneity, they take two approaches: (1) They first-difference both the treatment and the outcomes so that they are in terms of per capita changes. In a way, this is similar to removing the fixed effects in a linear panel model. (2) They add other start-of-period measures of regional market conditions and demographics to make the exogeneity of the *share* variables more plausible.

The results in Table 3 of Autor et al. (2013) show that adding more controls reduces the size of the estimated (negative) impact due to Chinese imports. This can be regarded as correcting the bias due to the second type of endogeneity. However, adding more controls may not fully capture the unobserved determinants that jeopardize the exogeneity of the *share* variables, and doing so may introduce extra endogeneity from the control variables. To make the exogeneity condition of the IV more likely to hold while avoiding the extra bias due to endogenous controls, we propose to implement the nonparametric control function approach in Section 2.2. We also illustrate how the endogenous control test works and its practical implications.

Our approach is also based on the exogeneity of the IV *shift* variables[9]. Thus, we do not

[9]Suppose the growth in imports from China is partially driven by positive demand shocks too, then the estimates of import-competition impact would look smaller because the positive demand would also generate positive outcomes in the US markets.

tend to evaluate or improve upon the choice of IV in the original study; instead, we intend to make this approach more robust by allowing more controls to achieve conditional validity of IV while relaxing the exogeneity requirement of the control variables.

To focus on our main concern, readers are referred to Autor et al. (2013) for a detailed construction of the data. The main outcome of interest, ΔL, is the regional decade-change (% pts) in the share of manufacturing employment; the explanatory variable of focus, ΔI_US, is a measure of changes in Chinese import exposure per worker in each region; and the IV, I_O, is the same measure of changes in exposure but with Chinese imports to the US replaced by those to other high-income countries. Two periods of decade-long first-difference cross-sectional data are stacked together as a panel. Five sets of time-invariant control variables, X, are augmented. The baseline model is given as follows:

$$\Delta L_{it} = \gamma_t + \tau \Delta I_US_{it} + X_i + e_{it}$$

where γ_t is a time fixed effect and e_{it} is the stochastic error term. Under the IV conditions and regularity conditions, τ is exactly identified by the linear IV approach, given that X_i is uncorrelated to e_{it}.

To implement our approach as in Section 2.2, we use the third-order Hermite polynomial series for approximating the nonparametric functions. Specifically, we estimate $F_{\Delta I_US|\Delta I_O,X}$ at a grid of values, taken from the unconditional quantiles of ΔI_US, using the third-order Hermite polynomial basis functions of $(\Delta I_O, X)$. Due to a large number of basis functions as the dimension of control increases, we add a ridge penalty to the least square estimation of the finite series regression, which is also a common practice in nonparametric series IV for better computation properties as suggested by Newey and Powell (2003). The time fixed effects and dummy variables in X are included linearly and not penalized. Using the ridge estimates, we obtain the fitted conditional CDF $\widehat{V}$ for each observation using linear interpolation. The conditional expectation $E(Y|W, V)$ is also approximated by the third-order Hermite polynomials with plugged-in $\widehat{V}$, and then the average derivative estimator is obtained from the least square estimator of the series regression with ridge penalty. Both ridge penalty levels are chosen by cross-validation. For inference, we obtain confidence intervals through cluster bootstrap that resamples with replacement at the state level.

Table 4 displays our results. The first IV section of the table replicates the results from Table 3 in Autor et al. (2013), except that the standard errors and confidence intervals are replaced by the bootstrap versions to match those used for other estimators. ADE denotes

Table 4: IV, ADE (Ridge-penalized), and Endogenous Control Test

	IV			ADE			IV − ADE		
Spec	$\hat{\beta}_p$	SE	95% CI	$\hat{\beta}_{np}$	SE	95% CI	$\hat{\beta}_p - \hat{\beta}_{np}$	SE	95% CI
(1)	−0.746	0.070	[−0.883, −0.613]	−1.036	0.144	[−1.289, −0.738]	0.290	0.111	[0.072, 0.510]
(2)	−0.610	0.102	[−0.793, −0.399]	−0.498	0.082	[−0.679, −0.361]	−0.113	0.124	[−0.315, 0.108]
(3)	−0.538	0.094	[−0.687, −0.336]	−0.420	0.069	[−0.559, −0.288]	−0.118	0.115	[−0.262, 0.111]
(4)	−0.508	0.082	[−0.648, −0.362]	−0.182	0.083	[−0.424, −0.109]	−0.326	0.111	[−0.518, −0.118]
(5)	−0.562	0.085	[−0.725, −0.405]	−0.126	0.075	[−0.285, −0.008]	−0.436	0.130	[−0.668, −0.184]
(6)	−0.596	0.095	[−0.759, −0.419]	−0.037	0.068	[−0.198, 0.066]	−0.559	0.127	[−0.780, −0.277]

Notes: The ridge penalty level is chosen by the 5-fold cross-validation with the minimum-square-error rule. The bootstrap confidence interval for $\hat{\theta}$ is defined as $\mathcal{B}_n(\alpha)$ in Section 3, with $\alpha = 0.05$; it is similarly defined for the other two estimators.

the average derivative estimator with plugged-in $\widehat{V}$. The last section IV − ADE gives the difference of these two estimators and is used for the test of endogenous controls. Six specifications in terms of controls are as follows: (1) No control; (2) start-of-period employment; (3) Census-division specific fixed effects + (1); (4) start-of-period shares of college attainment, immigrants, female employment + (2); (5) Routine-intensive occupations, offshorability index + (2); (6) All controls mentioned above.

First, looking at the IV estimates, we find that as more controls are included, the estimated effects get smaller: the magnitude drops from 0.746 (no control), to 0.610 (one control), 0.538 (census division dummies), 0.508 (demographics), 0.562 (industrial types), and 0.596 (all controls), all with a negative sign. The estimates of our approach display a similar pattern: the absolute sizes of the estimates reduce as more controls are included. As we discussed above, while the inclusion of more controls may solve the first type of bias, it may further introduce the second type, and our results provide alternative estimates robust to endogenous controls. If we believe more control variables help guard against the first type of bias, then the results in (6) suggest that after correcting the bias from the second type, the causal impact of Chinese import competition on the US market is not as large as what's found in the baseline results.

The standard errors of the nonparametric method are reasonably small, and some are even smaller than the linear IV estimates, which is attributed to the ridge regularization. However, the ridge penalty also introduces shrinkage bias on the slope estimators of the series regression, which in turn may cause shrinkage bias in the average derivative estimator. Therefore, we further conduct a robustness check by only penalizing the conditional CDF estimation and leaving the second-step average derivative estimator unpenalized. The results

are displayed in Table 5. We find that the estimates without penalization are indeed larger than their counterparts in Table 4, but the pattern remains: although the standard errors increase, we still find our approach generates smaller estimates of the average response except for the first case, where no controls are included.

Table 5: IV, ADE (un-penalized), and Endogenous Control Test

	IV			ADE			IV − ADE		
Spec	$\hat{\beta}$	SE	95% CI	$\hat{\hat{\beta}}$	SE	95% CI	$\hat{\theta}$	SE	95% CI
(1)	−0.746	0.070	[−0.883, −0.613]	−1.317	0.200	[−1.682, −0.901]	0.571	0.178	[0.237, 0.917]
(2)	−0.610	0.102	[−0.787, −0.409]	−0.602	0.182	[−0.972, −0.251]	−0.009	0.212	[−0.346, 0.432]
(3)	−0.538	0.107	[−0.728, −0.329]	−0.521	0.173	[−0.990, −0.294]	−0.017	0.199	[−0.328, 0.470]
(4)	−0.508	0.084	[−0.673, −0.351]	−0.311	0.207	[−0.947, −0.078]	−0.198	0.227	[−0.467, 0.427]
(5)	−0.562	0.091	[−0.723, −0.378]	−0.138	0.212	[−0.650, 0.151]	−0.424	0.236	[−0.788, 0.136]
(6)	−0.596	0.101	[−0.787, −0.401]	−0.286	0.217	[−0.914, −0.067]	−0.311	0.242	[−0.586, 0.349]

Notes: he ridge penalty level (for the conditional CDF estimation) is chosen by the 5-fold cross-validation with the minimum-square-error rule. The bootstrap confidence interval for $\hat{\theta}$ is defined as $\mathcal{B}_n(\alpha)$ in Section 3, with $\alpha = 0.05$; it is similarly defined for the other two estimators.

Lastly, the endogenous control test is conducted through estimates $\hat{\beta}_p - \hat{\beta}_{np}$ and the bootstrap 95% confidence interval from the last sections of Tables 4 and 5. Since row (1) does not include a control variable, our test does not apply there. Examining rows (2) - (6) in Table 4, we find rejections of no endogenous control in the last three rows, suggesting the ADEs are preferable in specifications (4) - (6). Due to larger standard errors, the test results in Table 5 are not very informative, given that none of the tests are rejected.

6 Conclusion

We address 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 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 we introduce imposes minimal practical restrictions, endogenous controls are generally innocuous in a nonparametric model. Furthermore, based on the identification results, we propose a test for endogenous controls.

For empirical studies, our results invite researchers to conduct robustness checks using nonparametric methods and the proposed test. In general, nonparametric approaches are more robust, not only with respect to specification concerns but also in light of endogenous controls.

Appendix

Proof of Theorem 1. By Assumptions 1 and the measurable separability between W and $\widetilde{X}$, $\partial_w f_{\varepsilon|W=w,\widetilde{X}=\tilde{x}}(\epsilon) = \partial_w f_{\varepsilon|\widetilde{X}=\tilde{x}}(\epsilon) = 0$. Then, applying Leibniz integral rule under the existence of an integrable dominating function and the chain rule gives

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

Case 1: If $\widetilde{X} = X$, then $\partial_w E[Y|W=w,\widetilde{X}=\tilde{x}] = \beta(w,x)$, and so $\beta = \int \beta(w,x) dF(w,x) = E\left[\partial_W E\left[Y|W,\widetilde{X}\right]\right]$. Case 2: If $\{\widetilde{X}\backslash X\} \neq \emptyset$, by integrating out $\{\widetilde{X}\backslash X\}$,

$$E[\partial_w E[Y|W=w,\widetilde{X}=\tilde{x}]|W=w,X=x] = E[\partial_w m(w,x,\varepsilon)|W=w,X=x] = \beta(w,x)$$

Then, by integrating out W and X, $\beta = \int \beta(w,x) dF(w,x) = E\left[\partial_W E\left[Y|W,\widetilde{X}\right]\right]$.

When Y is binary, given ε is partition as (u,v),

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

where the last equality follows because $m^*(w,x,u,v)$ strictly increases in u and $u^*(w,x,v)$ is

the unique solution of $m^*(w,x,u,v)=0$. By implicit function theorem under the conditions in the statement, $u^*(w,x,v)$ is differentiable in w. Then, by Leibniz rule, chain rule, CIA, and the measurable separability between W and $\widetilde{X}$, we have

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

Then, similarly, by considering Case (1) $\widetilde{X}=X$ and Case (2) $\{\widetilde{X}\backslash X\}\neq\emptyset$ as above, we have $\beta^*(w,x)=\partial_w E[Y|W=w,\widetilde{X}=\tilde{x}]$ under Case (1) and $\beta^*(w,x)=E[\partial_w E[Y|W=w,\widetilde{X}=\tilde{x}]|W=w,X=x]$ under Case (2). Therefore, by integrating out W and X, we conclude that $\beta^*=\int\beta^*(w,x)dF(w,x)=E\left[\partial_W E\left[Y|W,\widetilde{X}\right]\right]$, which completes the proof.

□

Proof of Theorem 2. 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 CIA 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 $Q_{\eta|X}(u)=\inf\{a\in\mathbb{R}:F_{\eta|X}(a)\geq u\}$ be the conditional quantile function of η given X. Under Assumption 3(ii), $F_{\eta|X}(a)$ is a one-to-one function of a 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 (8). □

Proof of Theorem 3. Under the conditions of Theorem 1, regardless of whether X are endogenous or exogenous, $\beta_0 = E[\partial_W g(W, X)] = \beta_{np}$. By Assumption 4 (i), $\theta_0 = 0$ under the null and $\theta_0 = B_n$ under the alternative. Then, under the null, by (10) and (12) as well as Assumption 4 (iv),

$$\begin{aligned} \sqrt{n}\hat{\theta} =& \sqrt{n}\left(\hat{\beta}_p - \beta_p\right) + \sqrt{n}\left(\beta_{np} - \hat{\beta}_{np}\right) \\ =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(W_i, X_i) + o_P(1) \xrightarrow{d} N(0, V) \end{aligned}$$

where the last equality follows from Lindeberg–Lévy CLT and $V = \mathrm{Var}(\psi(W, X))$.

Consider the bootstrap root $\sqrt{n}\left(\hat{\theta}^* - \hat{\theta}\right)$. By (11) and (13),

$$\sqrt{n}\left(\hat{\theta}^* - \hat{\theta}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(W_i^*, X_i^*) - \psi(W_i, X_i) + o_{P^*}(n^{-1/2}) \xrightarrow{d^*} N(0, V)$$

where the weak convergence conditional on the sample $\{Y_i, W_i, X_i\}_{i=1}^n$ follows from Theorem 10.4 in Hansen (2022). By Lemma 23.3 in Vaart (1998), with the standard errors replaced by constant 1, $P\{0 \in \mathcal{B}_n(\alpha)|H_0\} \to 1 - \alpha$.

Under the alternative, we have

$$\sqrt{n}\left(\hat{\theta} - B_n\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(W_i, X_i) + o_P(1) \xrightarrow{d} N(0, V)$$

Let $\mathcal{B}_n(\alpha) - B_n = \{\theta - B_n : \theta \in \mathcal{B}_n(\alpha)\}$. Then, by Lemma 23.3 in Vaart (1998) again, $P\{0 \in \mathcal{B}_n(\alpha) - B_n|H_1\} \to 1 - \alpha$. It follows that $P\{B_n \in \mathcal{B}_n(\alpha)|H_1\} \to 1 - \alpha$. Due to asymptotic normality of $\sqrt{n}\left(\hat{\theta}^* - \hat{\theta}\right)$, the confidence region shrinks at the rate $n^{-1/2}$ as $n \to \infty$. Since $\sqrt{n}|B_n| \to \infty$, $|B_n|$ dominates the (half) length of the confidence interval, thus $P\{0 \notin \mathcal{B}_n(\alpha)|H_1\} \to 1$, which completes the proof. □

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.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Cconomic Review*, 103(6):2121–2168.

Chen, X. and Christensen, T. M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics*, 188(2):447–465.

Chen, X., Linton, O., and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71(5):1591–1608.

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.

Hall, P. (1991). On convergence rates of suprema. *Probability theory and related fields*, 89(4):447–455.

Hansen, B. (2022). *Econometrics*. Princeton University Press.

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.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147–168.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Vaart, A. W. v. d. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

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.