

Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity

QUANG VUONG

Department of Economics, New York University

HAIQING XU

Department of Economics, University of Texas at Austin

This paper establishes nonparametric identification of individual treatment effects in a nonseparable model with a binary endogenous regressor. The outcome variable may be continuous, discrete, or a mixture of both, while the instrumental variable can take binary values. First, we study the case where the model includes a selection equation for the binary endogenous regressor. We establish point identification of the individual treatment effects and the structural function when the latter is continuous and strictly monotone in the latent variable. The key to our results is the identification of a so-called counterfactual mapping that links each outcome of the dependent variable with its counterfactual. Second, we extend our identification argument when there is no selection equation. Last, we generalize our identification results to the case where the outcome variable has a probability mass in its distribution such as when the outcome variable is censored or binary.

KEYWORDS. Nonparametric identification, nonseparable models, discrete endogenous variable, counterfactual mapping, individual treatment effects.

JEL CLASSIFICATION. C14, C18, C30, C36, C50.

1. INTRODUCTION AND RELATED LITERATURE

The primary aim of this paper is to establish nonparametric identification of individual treatment effects (ITE) in a nonseparable model with a binary endogenous regressor.

Quang Vuong: quuong@nyu.edu

Haiqing Xu: h.xu@austin.utexas.edu

A previous version of this paper was circulated under the title “Identification of nonseparable models with a binary endogenous regressor.” We thank the editor as well as three referees for their comments, which have greatly improved the paper. We also thank Jason Abrevaya, Federico Bugni, Karim Chalak, Xiaohong Chen, Victor Chernozhukov, Andrew Chesher, Denis Chetverikov, Xavier D’Haultfoeuille, Stephen Donald, Junlong Feng, Jinyong Hahn, Shakeeb Khan, Brendan Kline, Qi Li, Robert Lieli, Matthew Masten, Isabelle Perrigne, Geert Ridder, Xiaoxia Shi, Robin Sickles, Maxwell Stinchcombe, Elie Tamer, Edward Vytlačil, Kaixi Wang, and Nese Yildiz as well as seminar participants at Duke, Rice, NYU, UCLA, Princeton, Rochester, Emory, Pittsburgh, 2013 Bilkent University Annual Summer Workshop, 23rd Annual Meeting of the Midwest Econometrics Group, 2014 Cowles Summer Conference, 2014 CEMMAP Conference, the 6th French Econometrics Conference, and the 5th Shanghai Econometrics Workshop. The first author gratefully acknowledges financial support from the National Science Foundation through Grant SES 1148149, while the second author thanks UT Austin for a 2013 Summer Research Fellowship.

We focus on the case where the instrumental variable has limited variations, namely when it takes only binary values. As a byproduct, we also identify the structural function linking the outcome variable to the binary endogenous regressor, other covariates, and the unobserved term. Key to our results is the strict/weak monotonicity in the outcome equation with respect to the unobserved term, which allows us to extrapolate information from the complier group to the whole population. Our secondary objective is to investigate the identification power provided by the selection equation. In particular, we derive an easily verifiable high level condition that is sufficient for identification and show that the monotonicity of the selection equation implies such a condition.

In program evaluations, nonseparable structures admit heterogeneous treatment effects that “vary across individuals that, measured by covariates, are identical” (Chesher (2003)), and therefore allow the researcher to study the distribution of policy effects, rather than the mean effects only. For instance, if one considers the return to education, individuals who have the same demographics might benefit differently from schooling due to some unobserved job-related ability. As is argued in, for example, Heckman, Smith, and Clements (1997), the conventional assumption of identical treatment effects across individuals (with the same value of covariates) is convenient but implausible. In the literature, however, it is emphasized that measuring the distribution of heterogeneous individual treatment effects is difficult. In this paper, we develop a counterfactual mapping approach that identifies heterogeneous individual effects as well as their distribution among the population. In particular, our approach does not require estimation of the structural function.

The idea behind our identification strategy differs from the literature. It is based on identifying a counterfactual mapping that relates each individual outcome to its counterfactual under the monotonicity of the outcome equation. This then identifies the ITE and the structural function. Following Imbens and Rubin (1997), an exogenous change in the instrumental variable provides information on the distribution of (potential) outcomes at each value of the binary endogenous variable for a subpopulation called the *complier* group. From these two conditional distributions, we can identify constructively the counterfactual mapping by relating the quantiles of one distribution to the other for the whole population. The idea of matching two quantile functions to exploit limited variations of instrumental variables was introduced by Athey and Imbens (2006) in the context of policy intervention analysis with repeated cross-sectional data in nonlinear difference-in-difference models. It was also used by Guerre, Perrigne, and Vuong (2009) in the empirical auction literature with an endogenous number of bidders, and exploited by D’Haultfœuille and Février (2015) and Torgovitsky (2015) in a triangular model with continuous endogenous regressors.

Monotonicity between economic variables are prevalent in theoretical models. See, for example, Milgrom and Shannon (1994), Athey (2001), and Reny (2011). In econometric modeling, monotonicity is widely used in nonseparable models with continuous endogenous regressors. See, for example, Chesher (2003) and Matzkin (2008). For the binary endogenous variable case, Chernozhukov and Hansen (2005, 2013) and Chen, Chernozhukov, Lee, and Newey (2014) establish identification of the structural function without requiring a selection equation. Furthermore, Chesher (2005) establishes partial

identification of the structural function at some conditional quantile of the error term under local independence conditions on the instrumental variable. Subsequently, Jun, Pinkse, and Xu (2011) tighten Chesher (2005)'s bounds by strengthening Chesher's local conditions to the full independence of the instrumental variable.

Without monotonicity in either the outcome equation or the selection equation, Manski (1990) derives sharp bounds for the average treatment effect (ATE) with and without instrumental variables. Using a binary instrumental variable and monotonicity in the selection equation, Imbens and Angrist (1994) establish point identification of the local average treatment effect (LATE). In a similar setting, Heckman and Vytlacil (1999) develop the marginal treatment effect (MTE) and establish its identification by using local variations in instrumental variables. In this paper, we show that monotonicity in the outcome equation provides identification power to extrapolate information from local treatment effects to population treatment effects. Furthermore, we relax strict monotonicity to weak monotonicity in the outcome equation, thereby generalizing Vytlacil and Yildiz (2007) to the fully nonseparable setting for an outcome variable with mass points in its distribution.

In our setting with strict monotonicity in the outcome equation, we allow the instrumental variable to take only binary values, a case where identification at infinity obviously fails (see, e.g., Chamberlain (1986), Heckman (1990)). With weak monotonicity, our rank condition requires more variations in the instrumental variable, though a finite support is still allowed. Our method is also related to the instrumental variable approach developed in Chernozhukov and Hansen (2005, 2013) and generalized by Chen et al. (2014). The instrumental variable approach does not require a selection equation. Identification then relies on a full rank condition of an equation system. In contrast, we exploit the identification power from the monotonicity of some identified functions to deliver a weaker sufficient condition for identification. In addition, our identification is constructive.

The paper is organized as follows. In Section 2, we introduce our benchmark model and present our main identification results. Section 3 extends our identification argument to the case where there is no selection equation. Section 4 generalizes our method to the case where the distribution of the outcome variable has some mass points. Section 5 concludes. Appendix A collects the proofs of our main results. Proofs of auxiliary lemmas are presented in Appendix B, available in a supplementary file on the journal website, <http://qeconomics.org/supp/579/supplement.pdf>. In Appendix B, we also study partial identification when a support condition fails, and we characterize the restrictions imposed on data by the model with or without the selection equation.

2. BENCHMARK MODEL AND MAIN RESULTS

In this section, we present the benchmark model with its assumptions. We provide some examples for illustration and we state our main identification results regarding individual treatment effects and the structural function.

2.1 Triangular model

We consider the nonseparable triangular system with a binary endogenous variable,

$$Y = h(D, X, \varepsilon), \quad (1)$$

$$D = \mathbb{1}[m(X, Z) - \eta \geq 0], \quad (2)$$

where Y is the outcome variable, $D \in \{0, 1\}$ is the binary endogenous variable, $X \in \mathbb{R}^{d_X}$ is a vector of observed covariates, and $Z \in \{z_1, z_2\}$ is a binary instrumental variable for the binary endogenous variable D .¹ The error terms ε and η are scalar valued disturbances.² The functions h and m are unknown structural relationships. We define the individual treatment effect (ITE) as

$$\text{ITE} \equiv h(1, X, \varepsilon) - h(0, X, \varepsilon).$$

See, for example, [Rubin \(1974\)](#) and [Heckman, Smith, and Clements \(1997\)](#).

Following standard convention, we refer to (1) and (2) as the outcome equation and the selection equation, respectively.³ For the identification of ITE in our benchmark model, we make the following assumptions.

ASSUMPTION A. Equation (1) holds where (i) h is continuous and strictly increasing in ε , (ii) Z is conditionally independent of ε given X , that is, $Z \perp \varepsilon | X$, and (iii) the conditional cumulative distribution function (c.d.f.) $F_{\varepsilon|X}(\cdot|x)$ is continuous on \mathbb{R} .

ASSUMPTION B. Equation (2) holds where (i) Z is conditionally independent of η given (X, ε) , that is, $Z \perp \eta | (X, \varepsilon)$, and (ii) the conditional c.d.f. $F_{\eta|X, \varepsilon}(\cdot|x, e)$ is continuous on \mathbb{R} .

In Assumption A(i), the continuity and strict monotonicity of h follow, for example, [Matzkin \(1999, 2003\)](#), [Chesher \(2003\)](#), and [Chernozhukov and Hansen \(2005\)](#). In Section 4, we relax this assumption by allowing h to be flat inside the support of ε . The essential restriction in Assumption A(i) is the so-called rank preservation/invariance condition; see, for example, [Heckman, Smith, and Clements \(1997\)](#) and [Chernozhukov and Hansen \(2005\)](#), among others. Specifically, for any given $\tau \in (0, 1)$, it requires that the relative rank/quantile of $h(1, x, \tau)$ in the distribution of $Y_{1x} \equiv h(1, x, \varepsilon)$ be the same as that of $h(0, x, \tau)$ in the distribution of $Y_{0x} \equiv h(0, x, \varepsilon)$. If the error term is additive, as is often assumed, that is, $Y = h^*(D, X) + \varepsilon$ for some real valued function h^* , then Assumption A(i) holds. Assumption A(ii) is standard. Assumptions A(ii) and B(i) are equivalent

¹An early use of this triangular model is the seminal paper by [Heckman \(1979\)](#) on sample selection in a parametric setting. Considering a binary valued instrument highlights the identification power of the instrumental variable. For the importance of having binary instruments in the treatment effect literature, see, for example, [Imbens and Wooldridge \(2009\)](#).

²Note that η being scalar is not essential and can be relaxed by the monotonicity assumption in [Imbens and Angrist \(1994\)](#). In the nonseparable model context, multidimensional ε was considered by [Hoderlein and Mammen \(2007\)](#) and [Kasy \(2014\)](#).

³Note that (2) covers the general setting where $D = g(X, Z, \eta)$ under standard assumptions. Specifically, suppose that g is nonincreasing and left-continuous in η . For each (x, z) , let $m(x, z) = \inf\{\eta \in \mathbb{R} : g(x, z, \eta) = 0\}$. It follows that $g(x, z, \eta) = \mathbb{1}\{m(x, z) - \eta \geq 0\}$ for all (x, z) . See [Vytlačil \(2002\)](#).

to the conditional independence of Z and (ε, η) given X , that is, $Z \perp (\varepsilon, \eta) | X$, which is a standard requirement for the instrumental variable Z .⁴ Assumptions A(iii) and B(ii) are weaker than the standard assumption that conditional on X , (ε, η) are absolutely continuous with respect to Lebesgue measure. Assumptions A(i) and A(iii) together rule out mass points in the distribution of Y .

Let $p(x, z) \equiv \mathbb{P}(D = 1 | X = x, Z = z)$ denote the propensity score for $z \in \{z_1, z_2\}$ and $x \in \mathcal{X}$.⁵ Following Imbens and Angrist (1994), we define the complier group under Assumption B as $C_x \equiv \{X = x, m(x, z_1) < \eta \leq m(x, z_2)\}$.⁶ Under Assumptions A(ii) and B(i), similar to Imbens and Rubin (1997), the conditional distribution of $Y_{dx} \equiv h(d, x, \varepsilon)$ given the complier group is identified on its support from the data, that is,

$$F_{Y_{dx}|C_x}(t) = \frac{\mathbb{P}(Y \leq t; D = d | X = x, Z = z_2) - \mathbb{P}(Y \leq t; D = d | X = x, Z = z_1)}{\mathbb{P}(D = d | X = x, Z = z_2) - \mathbb{P}(D = d | X = x, Z = z_1)} \tag{3}$$

for $d = 0, 1$ and all $t \in \mathbb{R}$. Let $\mathcal{S}_{Y_{dx}|C_x}$ be the support of Y_{dx} given C_x . To identify ITE, we make two additional assumptions.

ASSUMPTION C (Rank Condition). For every $x \in \mathcal{X}$, $p(x, z_1) \neq p(x, z_2)$.

ASSUMPTION D (Support Condition). For every $x \in \mathcal{X}$ and $d = 0, 1$, $\mathcal{S}_{Y_{dx}|C_x} = \mathcal{S}_{Y_{dx}|X=x}$.

Assumption C is a minimal rank condition as it requires the propensity score $p(x, z)$ to vary with z . In general, we have $\mathcal{S}_{Y_{dx}|C_x} \subseteq \mathcal{S}_{Y_{dx}|X=x}$ (see the proof of Lemma 1). Thus Assumption D requires that the complier group $C_x \equiv \{X = x, m(x, z_1) < \eta \leq m(x, z_2)\}$ be sufficiently large so that these two supports are equal. Note that given the monotonicity of h in Assumption A, the support condition in Assumption D is equivalent to $\mathcal{S}_{\varepsilon|C_x} = \mathcal{S}_{\varepsilon|X=x}$. Such a condition is minimal as it is indispensable for the identification of ITE (and ATE as well) in the whole population, even if the error term ε was directly observed in the data. Because the complier group depends on the values z_1 and z_2 , we say that Z is *effective* if Assumption D is satisfied. We discuss it further in Appendix B.3. It is, however, worth noting that a sufficient condition for Assumption D is that conditional on X , (ε, η) has a rectangular support as is often assumed for the triangular model (1) and (2).⁷

2.2 Examples

To illustrate Assumptions A–D, we provide two examples of nonseparable structures. These examples are discussed further in Section 2.4.

EXAMPLE 1 (Additive Error With Generalized Heterogeneity). A running example of the triangular model (1) and (2) is the return to education (see, e.g., Heckman (1979),

⁴This assumption is stronger than the local independence restriction imposed by Chesher (2005).

⁵Hereafter, for a generic random variable W with distribution F_w , we denote its support by \mathcal{S}_W , defined as the closure of the open set $\{w : F_w(w)$ is strictly increasing in a neighborhood of $w\}$.

⁶Without loss of generality (w.l.o.g.), we implicitly assume $m(x, z_1) < m(x, z_2)$.

⁷This follows because $\mathcal{S}_{\varepsilon|X=x, m(x, z_1) < \eta \leq m(x, z_2)} = \mathcal{S}_{\varepsilon|X=x, \eta = \eta_0} = \mathcal{S}_{\varepsilon|X=x}$ for any $\eta_0 \in \mathcal{S}_{\eta|X=x}$.

Chesher (2005)). Let Y , D , and (X, Z) be earnings, schooling, and demographics, respectively, where the schooling dummy D indicates whether or not an individual has graduated from high school. Moreover, let ε be job-related ability and let η be education-related talent. Intuitively, these two latent variables are correlated to each other, which accounts for the endogeneity problem. The difference between demographics X and Z is that Z affects the education level of an individual, but not earnings. For instance, Z could indicate whether an individual was born in the first quarter of a calendar year following Angrist and Krueger (1991). Let (1) be

$$Y = h^*(D, X) + \sigma(D, X) \times \varepsilon \equiv h(D, X, \varepsilon)$$

for some real valued function h^* and positive function σ . In particular, the “heterogeneity” σ depends on the endogenous binary variable D .

Assumption A(i) holds in this specification, which extends Heckman (1979). To be a valid instrument in the triangular model (1) and (2), Z needs to satisfy the exogeneity condition $Z \perp (\varepsilon, \eta) | X$, which is Assumptions A(ii) and B(i), as well as the relevance condition, which is Assumption C. Angrist and Krueger (1991) take the quarter-of-birth dummies as exogenous in their triangular model determining weekly wage and education level. They also provide some evidence that quarter of birth does affect education level, thereby suggesting that $\mathbb{P}(D = 1 | X = x, Z = z_1) \neq \mathbb{P}(D = 1 | X = x, Z = z_2)$ so that Assumption C is satisfied by the first-quarter-of-birth dummy Z . Last, Assumption D relates to the effectiveness of the instrumental variable. It says that the potential earnings Y_{dx} of individuals in the complier group with characteristics x have the same range of values as the potential earnings Y_{dx} of individuals in the population with characteristics x . Since $C_x = \{X = x, Z = z_1, D = 0, \eta \leq m(x, z_2)\} \cup \{X = x, Z = z_2, D = 1, m(x, z_1) < \eta\}$, the complier group is composed of those individuals with characteristics x who are either (i) born in the first quarter and drop out of high school but who would have completed high school if born in any other quarters, or (ii) born in any other quarter and complete high school but who would have dropped out of high school if born in the first quarter. There does not seem to be a reason why such individuals have different potential earnings with or without high school graduation than the rest of the population so that Assumption D holds. In any case, Assumption D is testable as discussed after Lemma 1. Moreover, if one believes that (ε, η) has a full support conditional on X , namely \mathbb{R}^2 as in the original Heckman (1979) specification, then Assumption D is automatically satisfied.

EXAMPLE 2 (First-Price Auction With Risk Aversion). Another example arises in a first price auction with risk averse bidders. Consider each bidder's equilibrium bid in the private value paradigm: In a symmetric equilibrium,

$$B_i = h(D, X, \varepsilon_i) \quad \forall i = 1, \dots, I,$$

where B_i is bidder i 's equilibrium bid, D indicates the level of competition (e.g., $D = \mathbb{1}\{I = I_h\}$, where $I \in \{I_h, I_\ell\}$), X are characteristics of the auctioned object such as its appraisal value, and ε_i is bidder i 's private value. See, for example, Guerre, Perrigne, and

Vuong (2000, 2009). From an empirical point of view, one expects that private values depend on the level of competition, as higher valued objects may attract more bidders. To account for endogenous participation, the level of competition D is modeled as

$$D = \mathbb{1}\{\eta \leq m(X, Z)\},$$

where Z affects entry but not bidders' valuations so that $Z \perp (\varepsilon_i, \eta) | X$. For instance, Z could be the reserved price, as shown in Bajari and Hortacsu (2003). Furthermore, it is assumed that (ε_i, η) has a rectangular support conditional on (X, Z) .

This example is especially relevant as strict monotonicity and nonseparability in the scalar private value ε (in Assumption A) arise from auction theory. Clearly, Assumptions B–D are also satisfied.

2.3 Main identification results

The key to our identification strategy is to match $Y_{0x} \equiv h(0, x, \varepsilon)$ with $Y_{1x} \equiv h(1, x, \varepsilon)$ through a mapping ϕ_x , that is, $Y_{1x} = \phi_x(Y_{0x})$. We call ϕ_x a counterfactual mapping because we can find the counterfactual outcome Y_{1x} from Y_{0x} using ϕ_x and vice versa. Note that ϕ_x is uniquely defined under Assumption A(i) by $\phi_x(y) \equiv h(1, x, h^{-1}(0, x, y))$ for each $y \in \mathcal{S}_{Y_{0x}|X=x}$, where $h^{-1}(0, x, \cdot)$ is the inverse function of $h(0, x, \cdot)$. Moreover, the counterfactual mapping ϕ_x is continuous and strictly increasing from $\mathcal{S}_{Y_{0x}|X=x}$ onto $\mathcal{S}_{Y_{1x}|X=x}$. Thus, the ITE satisfies

$$\text{ITE} = D \times (Y - \phi_x^{-1}(Y)) + (1 - D) \times (\phi_x(Y) - Y), \tag{4}$$

where $\phi_x^{-1}(\cdot)$ denotes the inverse of $\phi_x(\cdot)$. In particular, we recover ITE for every individual in the population as soon as $\phi_x(\cdot)$ is identified for all $x \in \mathcal{S}_X$.

For a generic random variable W , let Q_W be the quantile function of W , that is, $Q_W(\tau) = \inf\{w \in \mathbb{R} : \mathbb{P}(W \leq w) \geq \tau\}$ for any $\tau \in [0, 1]$. Let $\mathcal{S}_{Y_{0x}|X=x}^\circ$ be the interior of $\mathcal{S}_{Y_{0x}|X=x}$.

LEMMA 1. *Suppose that Assumptions A–D hold and, w.l.o.g., let $p(x, z_1) < p(x, z_2)$. Then we have $\mathcal{S}_{Y_{dx}|X=x} = \mathcal{S}_{Y|D=d, X=x}$ for $d = 0, 1$ and the counterfactual mapping $\phi_x(\cdot)$ is identified on $\mathcal{S}_{Y_{0x}|X=x}$ by the continuous extension of*

$$\phi_x(y) = Q_{Y_{1x}|C_x}(F_{Y_{0x}|C_x}(y)) \quad \forall y \in \mathcal{S}_{Y_{0x}|X=x}^\circ, \tag{5}$$

where $Q_{Y_{1x}|C_x}(\cdot)$ and $F_{Y_{0x}|C_x}(\cdot)$ are given by (3).

Lemma 1 establishes the identification of the counterfactual mapping ϕ_x on $\mathcal{S}_{Y_{0x}|X=x}^\circ$ as well as its support $\mathcal{S}_{Y_{0x}|X=x}$ in a constructive way. It relies on matching the quantiles of Y_{0x} and Y_{1x} in the complier group. Moreover, since $\mathcal{S}_{Y_{dx}|X=x} = \mathcal{S}_{Y|D=d, X=x}$, then Assumption D can be written as $\mathcal{S}_{Y_{dx}|C_x} = \mathcal{S}_{Y|D=d, X=x}$ for $d = 0, 1$. Because $F_{Y_{dx}|C_x}$ is identified by (3), it follows that Assumption D is testable.

The next theorem follows immediately from Lemma 1 and (4).

THEOREM 1. *Suppose that Assumptions A–D hold. Then, with probability 1, the ITE of every individual in the population can be recovered by Lemma 1 and (4).*

Given that we recover ITE, the distribution of ITE is identified. Hence, some important features of the distribution are also identified, such as

$$\begin{aligned} \text{ATE} &= \mathbb{E}\{D[Y - \phi_X^{-1}(Y)] + (1 - D)[\phi_X(Y) - Y]\}, \\ \text{ATT} &= \mathbb{E}[Y - \phi_X^{-1}(Y)|D = 1], \end{aligned}$$

where ATE and ATT are the average treatment effect (ATE) and the average treatment effect on the treated (ATT), respectively. See, for example, Heckman and Vytlačil (2007) and Imbens and Wooldridge (2009).⁸ It is also worth noting that identification of the counterfactual mapping ϕ_X and all such treatment effects does not require the exogeneity of X in either (1) or (2).

As a byproduct of Lemma 1, we can also identify the structural function $h(d, x, \cdot)$ for $d = 0, 1$ and $x \in \mathcal{S}_X$. We need to make an additional assumption, namely, the exogeneity of X and a distributional assumption on ε .

ASSUMPTION E. *We have that (i) ε is independent of X , that is, $\varepsilon \perp X$, and (ii) $\varepsilon \sim U[0, 1]$.*

Assumption E(i) is the exogeneity of X in (1), which is indispensable for the identification of the structural function h . On the other hand, exogeneity of X in the selection equation (2) is not required.⁹ In Assumption E(ii), the uniform distribution of ε on $[0, 1]$ is a normalization that is standard in nonseparable models. See, for example, Chesher (2003).¹⁰

We now provide a lemma that shows that identification of $\phi_X(\cdot)$ is necessary and sufficient for identifying $h(d, x, \cdot)$ under Assumptions A and E.

LEMMA 2. *Suppose that Assumptions A and E hold. Then, for any $x \in \mathcal{S}_X$, $h(0, x, \cdot)$ and $h(1, x, \cdot)$ are identified on $[0, 1]$ if and only if $\phi_X(\cdot)$ is identified on $\mathcal{S}_{Y_{0x}}$.*

Lemma 2 reduces the identification of $h(1, x, \cdot)$ and $h(0, x, \cdot)$ into the identification of one function, namely, the counterfactual mapping $\phi_X(\cdot)$. To see the if part, note that conditional on $X = x$,

$$\begin{aligned} Y_{1x} &\equiv h(1, x, \varepsilon) = YD + \phi_X(Y)(1 - D), \\ Y_{0x} &\equiv h(0, x, \varepsilon) = \phi_X^{-1}(Y)D + Y(1 - D). \end{aligned} \tag{6}$$

⁸The quantile treatment effect (QTE) considered by Chernozhukov and Hansen (2005) can also be identified given that we can recover Y_{dX} using ϕ_X .

⁹If one was interested in identifying m in (2), then $\eta \perp X$ would be useful as such exogeneity combined with $\eta \perp Z|X$, which follows from Assumptions A(ii) and B(i), would give $\eta \perp (X, Z)$.

¹⁰Matzkin (2003) proposes an alternative normalization on the structural function h , instead of normalizing the distribution of ε .

Thus, the identification of $\phi_x(\cdot)$ provides the marginal distributions of Y_{0x} and Y_{1x} by Assumption E(i), which further identify $h(0, x, \cdot)$ and $h(1, x, \cdot)$ by Assumption E(ii). Specifically, for $d = 0, 1$ we have $h(d, x, e) = Q_{Y_{dx}}(e)$ for every $e \in (0, 1)$.

By Lemmas 2 and 1, the identification of h follows immediately. Let $p(x) \equiv \mathbb{P}(D = 1 | X = x)$.

THEOREM 2. *Suppose that Assumptions A–E hold. For every $x \in \mathcal{S}_X$ and for $d = 0, 1$, $h(d, x, \cdot)$ is identified on $[0, 1]$. Specifically, $h(d, x, 0) = \underline{Y}_{dx}$, which is the lower bound of $\mathcal{S}_{Y|D=d, X=x}$, while $h(d, x, \tau)$ is the τ th quantile of the (unconditional) distribution $F_{Y_{dx}}$ of Y_{dx} , where*

$$\begin{aligned}
 F_{Y_{1x}}(t) &= [1 - p(x)] \times \mathbb{P}(Y \leq \phi_x^{-1}(t) | D = 0, X = x) \\
 &\quad + p(x) \times \mathbb{P}(Y \leq t | D = 1, X = x), \\
 F_{Y_{0x}}(t) &= [1 - p(x)] \times \mathbb{P}(Y \leq t | D = 0, X = x) \\
 &\quad + p(x) \times \mathbb{P}(Y \leq \phi_x(t) | D = 1, X = x)
 \end{aligned}$$

for $\tau \in (0, 1]$.

Theorem 2 shows that $h(1, x, \tau)$ is the τ -quantile of $F_{Y_{1x}}(\cdot)$, which is a weighted average of $F_{Y|D=0, X=x}(\phi_x^{-1}(\cdot))$ and $F_{Y|D=1, X=x}(\cdot)$ with respective weights $1 - p(x)$ and $p(x)$. Similarly, $h(0, x, \tau)$ is the τ -quantile of $F_{Y_{0x}}(\cdot)$, which is a weighted average of $F_{Y|D=0, X=x}(\cdot)$ and $F_{Y|D=1, X=x}(\phi_x(\cdot))$ with respective weights $1 - p(x)$ and $p(x)$.

Figure 1 illustrates our identification strategy behind Theorems 1 and 2 when there are no covariates X . It displays the conditional distributions of $Y_0 \equiv h(0, \varepsilon)$ and $Y_1 \equiv h(1, \varepsilon)$ given the complier group \mathcal{C} , respectively, in the top graph, and the conditional distributions of Y given $D = 0$ and $D = 1$, respectively, in the bottom graph. These four distributions are identified from the data. Identification of $h(0, e)$ for an arbitrary value $e \in [0, 1]$ is equivalent to identifying $e = h^{-1}(0, y_0)$ for arbitrary $y_0 \in \mathcal{S}_{Y_0} = \mathcal{S}_{Y|D=0} = \mathcal{S}_{Y_0|\mathcal{C}}$. Point A gives $F_{Y|D=0}(y_0)$ while point B gives $F_{Y_0|\mathcal{C}}(y_0)$. Point C gives $y_1 = \phi(y_0)$ while point D gives $F_{Y|D=1}(y_1)$. Thus, e obtains by a weighted average of $F_{Y|D=0}(y_0)$ and $F_{Y|D=1}(y_1)$, that is, $e = \mathbb{P}(D = 0) \times F_{Y|D=0}(y_0) + \mathbb{P}(D = 1) \times F_{Y|D=1}(y_1)$, where $\mathbb{P}(D = 0)$ and $\mathbb{P}(D = 1)$ come from the data. Hence, e is identified. Point F is the corresponding weighted average of A and E and gives $F_{Y_0}(y_0)$. Similarly, point G gives $F_{Y_1}(y_1)$. When y_0 varies, points F and G trace out the population distributions of Y_0 and Y_1 , respectively.

A special case of our results arises when one imposes additive separability of ε in h , that is, $h(D, X, \varepsilon) = h^*(D, X) + \varepsilon$. It follows that the counterfactual mapping takes a specific functional form. Namely, $\phi_x(y) = h^*(1, x) - h^*(0, x) + y$.¹¹ In this case, there is no unobserved heterogeneity in individual treatment effects as $h(1, x, \varepsilon) - h(0, x, \varepsilon) = h^*(1, x) - h^*(0, x)$, which does not depend on ε . Hence, conditional on X , individual

¹¹Because of the identification of ϕ_x , the additive separability of ε implies testable model restrictions.

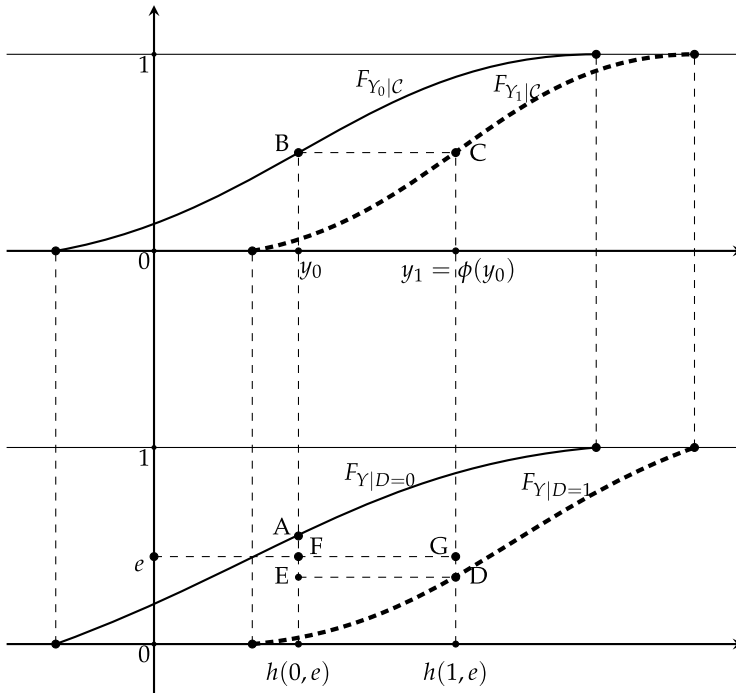


FIGURE 1. Identification illustration.

treatment effects are the same as LATE and ATE, that is,

$$\begin{aligned}
 h^*(1, x) - h^*(0, x) &= \mathbb{E}[h(1, x, \varepsilon) - h(0, x, \varepsilon)|C_x] \\
 &= \frac{\mathbb{E}(Y|X = x, Z = z_2) - \mathbb{E}(Y|X = x, Z = z_1)}{p(x, z_2) - p(x, z_1)}.
 \end{aligned}$$

See [Imbens and Angrist \(1994\)](#). In this case, Assumption D is not needed for extrapolating information from the complier group to the whole population. When Assumption D fails in the general triangular model (1) and (2), we provide some results on partial identification of the counterfactual mapping $\phi_x(\cdot)$ and the structural function $h(d, x, \cdot)$ in Appendix B.3.

2.4 Examples

To illustrate Lemma 1, and Theorems 1 and 2, we consider again the two examples of Section 2.2.

EXAMPLE 1 (Continued). Given Assumptions A–D, Theorem 1 establishes the identification of the ITE, that is, the potential gain from completing high school, for every individual in the population by observing each individual’s current earning Y , schooling status D , and demographics X . Specifically to apply (4), we need to know the counterfactual mapping ϕ_x and its inverse ϕ_x^{-1} . For instance, from Lemma 1 we can use a sample analog of (3) to estimate nonparametrically the distribution $F_{Y_{dx}|C_x}$ of potential outcomes in

the complier group for $d = 0, 1$.¹² Note that this requires neither exogeneity of X in the triangular model (1) and (2), nor a normalization of ε such as $\mathbb{E}(\varepsilon) = 0$.

On the other hand, to identify the structural functions $h^*(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ on their respective supports, Theorem 2 requires Assumption E, namely (i) exogeneity of X in (1) and (ii) a normalization of the distribution of ε . Under such assumptions, the general structural function $h(d, x, \cdot)$ is obtained as the quantile of the distribution of Y_{dx} . Again, we can use a sample analog of the expression given in Theorem 2 to estimate the latter. Regarding (i), Angrist and Krueger (1991) take several demographics as exogenous in their weekly wage equation. Regarding (ii), the normalization of the distribution of ε is tailored to the identification of the general structural function h in (1). In the present example we can exploit the special form of h , namely $h(D, X, \varepsilon) \equiv h^*(D, X) + \sigma(D, X) \times \varepsilon$, to obtain a weaker normalization than Assumption E(ii). For instance, assuming $\mathbb{E}(\varepsilon) = 0$ and $\text{Var}(\varepsilon) = 1$ suffices since $h^*(d, x)$ and $\sigma(d, x)$ are identified as the mean and standard deviation of Y_{dx} whose marginal distribution is identified on its full support $\mathcal{S}_{Y_{dx}}$ as a consequence of Lemma 1 and Assumption E(i).

Though we have not required exogeneity of X in the selection equation (2), such an exogeneity (or a weaker form of it) is useful for identifying the function m , which determines whether an individual completes high school.

EXAMPLE 2 (Continued). Given Assumptions A–D, Lemma 1 identifies the counterfactual mapping ϕ_x for every $x \in \mathcal{S}_X$. Consequently, for every individual with observed bid B , competition level D , and auction characteristics X , we can identify his/her counterfactual bid B_c , that is, what this individual would bid under the alternative competition level $1 - D$, from

$$B_c = D \times \phi_x^{-1}(B) + (1 - D) \times \phi_x(B).$$

This is useful to conduct some counterfactuals. As in the previous example, we can use Lemma 1 to estimate nonparametrically ϕ_x and its inverse ϕ_x^{-1} from sample analogs of (3) of the distributions $F_{Y_{dx}|\mathcal{C}_x}$ for $d = 0, 1$. Alternatively, we can use the extremum estimator developed in Feng, Vuong, and Xu (2016). In contrast to Guerre, Perrigne, and Vuong (2000, 2009), such a counterfactual bid does not require estimation of the inverse bidding strategy $h^{-1}(d, x, \cdot)$. It also does not require that X be exogenous, while allowing for competition level to be correlated with bidders' private values.

3. IDENTIFICATION WITHOUT SELECTION EQUATION

In this section, we drop the selection equation (2) with its Assumption B and provide a general sufficient condition for the identification of the counterfactual mapping ϕ_x as well as the structural function h . Such a condition is related to, but weaker than, the rank condition for global identification developed in Chernozhukov and Hansen (2005,

¹²See also Feng, Vuong, and Xu (2016), who consider an alternative extremum estimator of ITE and its density in the triangular model (1) and (2). In particular, they show that the counterfactual mapping as well as the ITEs can be estimated uniformly at the \sqrt{n} rate when the covariates X are discrete.

Theorem 2). Throughout, we maintain Assumptions **A**, **C**, and **E**. In Appendix B.5, we characterize the restrictions on the data imposed by the model with or without the selection equation that can guide a researcher about which model to use.

Under Assumptions **A** and **E**, we have $\mathbb{P}[Y \leq h(D, X, \tau)|X = x, Z = z] = \tau$ for all $z \in \mathcal{Z}_{|X=x}$ and $\tau \in (0, 1)$, which is called the main testable implication by **Chernozhukov and Hansen (2005, Theorem 1)**. Because D is binary, we have

$$\begin{aligned} & \mathbb{P}[Y \leq h(1, x, \tau); D = 1|X = x, Z = z] \\ & + \mathbb{P}[Y \leq h(0, x, \tau); D = 0|X = x, Z = z] = \tau \end{aligned} \quad (7)$$

for all $z \in \mathcal{Z}_{|X=x}$. W.l.o.g., let $p(x, z_1) < p(x, z_2)$. Thus, we obtain

$$\begin{aligned} & \mathbb{P}[Y \leq h(1, x, \tau); D = 1|X = x, Z = z_1] + \mathbb{P}[Y \leq h(0, x, \tau); D = 0|X = x, Z = z_1] \\ & = \mathbb{P}[Y \leq h(1, x, \tau); D = 1|X = x, Z = z_2] \\ & + \mathbb{P}[Y \leq h(0, x, \tau); D = 0|X = x, Z = z_2], \end{aligned}$$

that is,

$$\Delta_0(h(0, x, \tau), x, z_1, z_2) = \Delta_1(h(1, x, \tau), x, z_1, z_2), \quad (8)$$

where $\Delta_d(\cdot, x, z_1, z_2)$ is defined for $y \in \mathbb{R}$ as

$$\begin{aligned} & \Delta_0(y, x, z_1, z_2) \\ & \equiv \mathbb{P}[Y \leq y; D = 0|X = x, Z = z_1] - \mathbb{P}[Y \leq y; D = 0|X = x, Z = z_2], \\ & \Delta_1(y, x, z_1, z_2) \\ & \equiv \mathbb{P}[Y \leq y; D = 1|X = x, Z = z_2] - \mathbb{P}[Y \leq y; D = 1|X = x, Z = z_1]. \end{aligned}$$

Note that Δ_d is (up to sign) the numerator of the right-hand side of (3). By definition, $\Delta_d(\cdot, x, z_1, z_2)$ is identified on \mathbb{R} for $d = 0, 1$. Let $y = h(0, x, \tau)$ in (8). Thus, $h(1, x, \tau) = \phi_x(y)$. Since τ is arbitrary, then

$$\Delta_0(y, x, z_1, z_2) = \Delta_1(\phi_x(y), x, z_1, z_2) \quad \forall y \in \mathcal{Y}_{0x|X=x}. \quad (9)$$

Our general identification result is based on (9) and exploits the strict monotonicity of $\phi_x(\cdot)$.¹³ As a motivation, recall that under Assumptions **B** and **D**, $\Delta_d(y, x, z_1, z_2)$ is continuous and strictly increasing in $y \in \mathcal{Y}_{dx|X=x}^\circ$, thereby identifying the counterfactual mapping $\phi_x(\cdot) = \Delta_1^{-1}(\Delta_0(\cdot, x, z_1, z_2), x, z_1, z_2)$ on $\mathcal{Y}_{0x|X=x}$; see Lemma 1. This suggests that identification of the counterfactual mapping ϕ_x can be achieved under weaker conditions than Assumption **B**.

¹³In contrast, **Chernozhukov and Hansen (2005)** first write (7) for $Z = z_1$ and $Z = z_2$, and then find sufficient conditions for the resulting system of two (nonlinear) equations to have a unique local/global solution in the two unknowns $h(1, x, \tau)$ and $h(0, x, \tau)$.

DEFINITION 1 (Piecewise Monotone). Let $g : \mathbb{R} \rightarrow \mathbb{R}$ and $S \subseteq \mathbb{R}$. We say that g is piecewise weakly (strictly) monotone on S if S can be partitioned into a (finite or infinite) sequence of non-overlapping intervals with increasing left endpoints such that g is weakly (strictly) monotone in each interval.

The next lemma exploits the strict monotonicity of ϕ_x . Its proof is given in Appendix B.1.

LEMMA 3. Fix $x \in \mathcal{X}_X$. W.l.o.g, let $p(x, z_1) < p(x, z_2)$. Suppose (9) holds where $\phi_x(\cdot) : \mathcal{Y}_{0x|X=x} \rightarrow \mathcal{Y}_{1x|X=x}$ is continuous and strictly increasing. Then $\Delta_1(\cdot, x, z_1, z_2)$ is piecewise weakly (strictly) monotone on $\mathcal{Y}_{1x|X=x}$ if and only if $\Delta_0(\cdot, x, z_1, z_2)$ is piecewise weakly (strictly) monotone on $\mathcal{Y}_{0x|X=x}$.

We now provide a necessary and sufficient condition for identification of ϕ_x when $\Delta_d(\cdot, x, z_1, z_2)$ is piecewise weakly monotone on $\mathcal{Y}_{dx|X=x}$ for some d . See Appendix B.1 for its proof.

LEMMA 4. In addition to the assumptions of Lemma 3, suppose that $\Delta_d(\cdot, x, z_1, z_2)$ and $\mathcal{Y}_{dx|X=x}$ for $d = 0, 1$ are known. Suppose also that $\Delta_d(\cdot, x, z_1, z_2)$ is piecewise weakly monotone on $\mathcal{Y}_{dx|X=x}$ for some d . Then $\phi_x(\cdot)$ is identified on $\mathcal{Y}_{0x|X=x}$ if and only if $\Delta_d(\cdot, x, z_1, z_2)$ is piecewise strictly monotone on $\mathcal{Y}_{dx|X=x}$.

Lemma 4 is useful when identification is based on (9) only.

In view of Lemma 4, we make the following assumption.

ASSUMPTION F. We have that $\Delta_d(\cdot, x, z_1, z_2)$ is piecewise strictly monotone on $\mathcal{Y}_{dx|X=x}$ for some d .

Assumption F is weak. In particular, if $\Delta_d(\cdot, x, z_1, z_2)$ is continuously differentiable as in Chernozhukov and Hansen (2005, Theorem 2-i), and their full rank condition holds, then Assumption F holds.¹⁴ Moreover, by the proof of Theorem 3, Assumption F is equivalent to the piecewise strict monotonicity of $\Delta_d(\cdot, x, z_1, z_2)$ on $\mathcal{Y}_{Y|D=d, X=x}$, which is testable.

By replacing Assumptions B and D with the weaker Assumption F, we now extend Theorems 1 and 2. Note that the nonflatness of $\Delta_d(\cdot, x, z_1, z_2)$ on $\mathcal{Y}_{Y|D=d, X=x}$ for $d = 0, 1$ plays a role similar to Assumption D.

THEOREM 3. Suppose that Assumptions A, C, and F hold. Let $x \in \mathcal{X}_X$. W.l.o.g, let $p(x, z_1) < p(x, z_2)$. Then the support $\mathcal{Y}_{dx|X=x}$ is identified by $\mathcal{Y}_{Y|D=d, X=x}$, and $\phi_x(\cdot)$ is identified on $\mathcal{Y}_{0x|X=x}$, thereby recovering the ITE of individuals with $X = x$. In addition, suppose that Assumption E holds. Then $h(d, x, \cdot)$ is identified on $[0, 1]$ for $d = 0, 1$.

¹⁴More precisely, it can be shown that continuous differentiability and full rank on any closed rectangle \mathcal{L} contained in the support $\mathcal{Y}_{0x|X=x} \times \mathcal{Y}_{1x|X=x}$ imply that $\Delta_d(\cdot, x, z_1, z_2)$ for $d = 0, 1$ are piecewise strictly monotone on the projection of \mathcal{L} onto $\mathcal{Y}_{dx|X=x}$. Thus, Assumption F is satisfied. See Appendix B.1 for a proof. Refinements can be developed by allowing for more flexible sets \mathcal{L} .

To achieve point identification without the selection equation we replace Assumptions **B** and **D** with Assumption **F**. There are some “natural” selection rules that violate Assumption **B** but for which Assumption **F** still holds.¹⁵ For instance, [Gautier and Hoderlein \(2015\)](#) consider the selection mechanism

$$D = \mathbb{1}(\beta_0 + \beta_1 Z \geq 0),$$

where covariates X are suppressed, β_0, β_1 are random coefficients, and $(\beta_0, \beta_1, \varepsilon)$ are independent of Z . Clearly, Assumption **B** is not satisfied. Suppose in addition that $(\beta_0, \beta_1, \varepsilon)$ conforms to a joint normal distribution. For $d = 0, 1$, let r_d be the correlation coefficient between ε and $\beta_0 + \beta_1 \times d$. Then it can be shown that $\Delta_d(\cdot, z_1, z_2)$ for $d = 0, 1$ are piecewise strictly monotone if and only if $r_0 \neq r_1$.

4. IDENTIFICATION UNDER WEAK MONOTONICITY

This section provides another extension of Lemma 1 and Theorem 2 by generalizing our counterfactual mapping approach. Specifically, we relax the continuity and strict monotonicity assumption of h in Assumption **A** so that our method applies to an outcome variable with probability masses in its distribution. This is the case when the outcome variable is censored or binary, a challenging situation according to [Wooldridge \(2015\)](#). Specifically, we make the following assumption.

ASSUMPTION A'. Equation (1) holds where (i) h is left-continuous and weakly increasing in ε and (ii) Assumption **A(ii)** and (iii) holds.

The left-continuity of h is a normalization for the identification of the structural function at its discontinuity points. Throughout we maintain the selection equation with its Assumption **B**.

Because Assumption **A'(i)** relaxes Assumption **A(i)**, we need to strengthen other assumptions, namely, Assumptions **C–E**, to achieve point identification of h .

ASSUMPTION D'. We have that $\mathcal{S}_{(\varepsilon, \eta)|X}$ is a rectangle.

ASSUMPTION E'. (i) We have that (ε, η) is independent of X , that is, $(\varepsilon, \eta) \perp X$; (ii) Assumption **E(ii)** holds.

Assumption **D'** simplifies the exposition. It is slightly stronger than the support condition in Assumption **D** as noted there. Assumption **E'(i)** strengthens Assumption **E(i)** by requiring $\eta \perp X | \varepsilon$. In particular, it also requires exogeneity of X in the selection equation (2). Combining Assumptions **A'(ii)**, **B(i)**, and **E'(i)**, we have $(\varepsilon, \eta) \perp (X, Z)$. A key insight from [Vytlacil and Yildiz \(2007\)](#) is that Assumption **E'(i)** allows us to use variations in the exogenous covariates X to identify $h(d, x, \cdot)$ when the instrumental variable Z has sufficient variations (more than binary).

¹⁵We thank a referee for this point as well as for the following example.

ASSUMPTION C' (Generalized Rank Condition). *For any $x \in \mathcal{X}$, there exists $\tilde{x} \in \mathcal{X}$ such that (i) the set $\mathcal{S}_{p(X,Z)|X=x} \cap \mathcal{S}_{p(X,Z)|X=\tilde{x}}$ contains at least two different values, and (ii) for any pair $\tau_1, \tau_2 \in (0, 1)$,*

$$h(0, \tilde{x}, \tau_1) = h(0, \tilde{x}, \tau_2) \implies h(1, x, \tau_1) = h(1, x, \tau_2).$$

Assumption C' is the rank condition for the identification of $h(1, x, \cdot)$. The rank condition for the identification of $h(0, x, \cdot)$ is similar. Condition (i) requires that there exist $z_1, z_2 \in \mathcal{S}_{Z|X=x}$ and $\tilde{z}_1, \tilde{z}_2 \in \mathcal{S}_{Z|X=\tilde{x}}$ such that $p(x, z_1) = p(\tilde{x}, \tilde{z}_1) < p(x, z_2) = p(\tilde{x}, \tilde{z}_2)$. Condition (ii) is testable since it is equivalent to the following condition:¹⁶ For any $\tau_1, \tau_2 \in (0, 1)$,

$$\begin{aligned} \mathcal{Q}_{Y_{0\tilde{x}}|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(\tau_1) &= \mathcal{Q}_{Y_{0\tilde{x}}|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(\tau_2) \\ \implies \mathcal{Q}_{Y_{1x}|m(x, z_1) < \eta \leq m(x, z_2)}(\tau_1) &= \mathcal{Q}_{Y_{1x}|m(x, z_1) < \eta \leq m(x, z_2)}(\tau_2). \end{aligned}$$

Since $\mathcal{Q}_{Y_{dx}|m(x, z_1) < \eta \leq m(x, z_2)} = \mathcal{Q}_{Y_{dx}|X=x, m(x, z_1) < \eta \leq m(x, z_2)}$ is identified by the inverse of (3), we can verify whether there is $\tilde{x} \in \mathcal{X}$ satisfying Assumption C'. When $h(d, x, \cdot)$ is strictly monotone in ε , Assumption C' reduces to Assumption C by setting $\tilde{x} = x$.

Fix x and let \tilde{x} satisfy Assumption C'. We define a generalized counterfactual mapping $\phi_{x, \tilde{x}}(\cdot)$ as $\phi_{x, \tilde{x}}(y) = h(1, x, h^{-1}(0, \tilde{x}, y))$ for all $y \in \mathcal{S}_{Y_{0\tilde{x}}}$.¹⁷ If $\tilde{x} = x$, then $\phi_{x, \tilde{x}}(\cdot)$ reduces to the counterfactual mapping $\phi_x(\cdot)$ of Section 2. Let $(z_1, z_2) \in \mathcal{S}_{Z|X=x}$ and $(\tilde{z}_1, \tilde{z}_2) \in \mathcal{S}_{Z|X=\tilde{x}}$ with $p(x, z_1) = p(\tilde{x}, \tilde{z}_1) < p(x, z_2) = p(\tilde{x}, \tilde{z}_2)$. The next theorem generalizes Lemma 1 and Theorem 2.

THEOREM 4. *Suppose that Assumptions A', B, D', and E' hold. Fix $x \in \mathcal{X}$. Suppose also that Assumption C' holds with \tilde{x} . Then $\phi_{x, \tilde{x}}(\cdot)$ is identified by*

$$\phi_{x, \tilde{x}}(\cdot) = \mathcal{Q}_{Y_{1x}|m(x, z_1) < \eta \leq m(x, z_2)}(F_{Y_{0\tilde{x}}|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(\cdot)) \tag{10}$$

for all $y \in \mathcal{S}_{Y_{0\tilde{x}}|D=0, X=\tilde{x}}$. Moreover, for any $\tau \in (0, 1)$, $h(1, x, \tau)$ is identified as the τ th quantile of the distribution

$$F_{Y_{1x}}(\cdot) = \mathbb{P}(Y \leq \cdot; D = 1|X = x, Z = z) + \mathbb{P}[\phi_{x, \tilde{x}}(Y) \leq \cdot; D = 0|X = \tilde{x}, Z = \tilde{z}],$$

where $z \in \mathcal{S}_{Z|X=x}$ and $\tilde{z} \in \mathcal{S}_{Z|X=\tilde{x}}$ satisfy $p(x, z) = p(\tilde{x}, \tilde{z})$.

To illustrate Theorem 4, we discuss the generalized rank condition (Assumption C') with two examples: The first example is a fully nonseparable censored regression model while the second example is a weakly separable binary response model. The first example seems to be new, though special cases have been studied previously under some parametric and/or separability assumptions. The second example was studied by Vytlačil and Yildiz (2007).

¹⁶To see this, first note that $\mathcal{Q}_{h(d,x,\varepsilon)}(\tau_1) = \mathcal{Q}_{h(d,x,\varepsilon)}(\tau_2)$ is equivalent to $h(d, x, \tau_1) = h(d, x, \tau_2)$ under Assumption E'(ii). Moreover, $\mathcal{Q}_{h(d,x,\varepsilon)|A}(\tau) = h(d, x, \mathcal{Q}_{\varepsilon|A}(\tau))$ for any event A by the weak monotonicity of $h(d, x, \cdot)$. Therefore, $\mathcal{Q}_{h(d,x,\varepsilon)|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(\tau) = h(d, x, \mathcal{Q}_{\varepsilon|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(\tau)) = \mathcal{Q}_{h(d,x,\varepsilon)}(\mathcal{Q}_{\varepsilon|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(\tau))$.

¹⁷Due to the weak monotonicity of h in ε , we define $h^{-1}(0, \tilde{x}, y) = \inf_{\tau} \{ \tau : h(0, \tilde{x}, \tau) \geq y \}$.

EXAMPLE 3 (Fully Nonseparable Censored Regression Model). Consider the model

$$Y = h^*(D, X, \varepsilon) \times \mathbb{1}[h^*(D, X, \varepsilon) \geq 0],$$

$$D = \mathbb{1}[m(X, Z) \geq \eta],$$

where h^* is continuous and strictly increasing in ε . The structural unknowns are the triple $(h^*, m, F_{\varepsilon\eta})$.

Fix $x \in \mathcal{X}$. For $d = 0, 1$, let τ_{dx} solve $h^*(d, x, \tau_{dx}) = 0$, where $\tau_{dx} \in (0, 1)$. Thus, Assumption C' is satisfied if there exists an $\tilde{x} \in \mathcal{X}$ such that $\tau_{0\tilde{x}} \leq \tau_{1x}$ and $p(x, z_1) = p(\tilde{x}, \tilde{z}_1) < p(x, z_2) = p(\tilde{x}, \tilde{z}_2)$ for some $z_1, z_2 \in \mathcal{Z}|_{X=x}$, $\tilde{z}_1, \tilde{z}_2 \in \mathcal{Z}|_{X=\tilde{x}}$.

EXAMPLE 4 (Vytlačil and Yildiz (2007)). Let Y and D denote a binary outcome variable and a binary endogenous regressor, respectively. Consider

$$Y = \mathbb{1}(h^*(D, X) \geq \varepsilon),$$

$$D = \mathbb{1}(m(X, Z) \geq \eta).$$

To identify $h^*(1, x)$ for some value x , Assumption C' requires that there exists $\tilde{x} \in \mathcal{X}$ such that $h^*(1, x) = h^*(0, \tilde{x})$ and $p(x, z_1) = p(\tilde{x}, \tilde{z}_1) < p(x, z_2) = p(\tilde{x}, \tilde{z}_2)$ for some $z_1, z_2 \in \mathcal{Z}|_{X=x}$ and $\tilde{z}_1, \tilde{z}_2 \in \mathcal{Z}|_{X=\tilde{x}}$. Thus, Assumption C' represents the support condition in Vytlačil and Yildiz (2007).¹⁸

Motivated by Shaikh and Vytlačil (2011), Assumption C'(ii) can be relaxed, leading to partial identification. We can also drop the support condition Assumption D' to construct bounds under Assumption C'(i). Appendix B.4 studies such issues.

5. CONCLUSION

This paper establishes nonparametric identification of the counterfactual mapping, individual treatment effects, and structural function in a nonseparable model with a binary endogenous regressor. Our benchmark model assumes strict monotonicity in the outcome equation and weak monotonicity in the selection equation. Our counterfactual mapping then links each outcome with its counterfactual. We also consider two extensions: One without the selection equation and the other with weak monotonicity in the outcome equation.

APPENDIX A: PROOFS

PROOF OF LEMMA 1. Fix $x \in \mathcal{X}$ and $z_1, z_2 \in \mathcal{Z}|_{X=x}$ with $p(x, z_1) < p(x, z_2)$. First, we show that $\mathcal{S}_{Y_{dx}|C_x} \subseteq \mathcal{S}_{Y|D=d, X=x}$ for $d = 0, 1$. Because of Assumption A(i), this is equivalent to $\mathcal{S}_{\varepsilon|C_x} \subseteq \mathcal{S}_{\varepsilon|D=d, X=x}$. Consider, say, $d = 1$. We have $\mathcal{S}_{\varepsilon|C_x} = \mathcal{S}_{\varepsilon|C_x, Z=z_2}$ by $(\varepsilon, \eta) \perp Z|X$. But $\{m(x, z_1) < \eta \leq m(x, z_2), Z = z_2, X = x\} \subseteq \{\eta \leq m(x, z_2), Z = z_2, X = x\} = \{D = 1, Z = z_2, X = x\} \subseteq \{D = 1, X = x\}$, thereby implying $\mathcal{S}_{\varepsilon|C_x} \subseteq \mathcal{S}_{\varepsilon|D=1, X=x}$ as claimed. A similar reasoning applies to $d = 0$. Next, we have $\mathcal{S}_{Y_{dx}|C_x} \subseteq \mathcal{S}_{Y|D=d, X=x} =$

¹⁸Note that Vytlačil and Yildiz (2007) require X to be excluded from the selection equation.

$\mathcal{S}_{Y_{dx}|D=d,X=x} \subseteq \mathcal{S}_{Y_{dx}|X=x}$, where the equality follows from (1) and the definition of Y_{dx} . By Assumption D we obtain $\mathcal{S}_{Y_{dx}|C_x} = \mathcal{S}_{Y|D=d,X=x} = \mathcal{S}_{Y_{dx}|D=d,X=x} = \mathcal{S}_{Y_{dx}|X=x}$, which includes the desired result.

Because h is strictly monotone in ε , we have for any $\tau \in \mathcal{S}_{\varepsilon|C_x}^\circ$ (the interior of $\mathcal{S}_{\varepsilon|C_x}$),

$$F_{\varepsilon|C_x}(\tau) = F_{Y_{0x}|C_x}(h(0, x, \tau)) = F_{Y_{1x}|C_x}(h(1, x, \tau)).$$

Hence, $h(d, x, \tau) \in \mathcal{S}_{Y_{dx}|C_x}^\circ$ for $d = 0, 1$ so that $F_{Y_{dx}|C_x}$ is continuous and strictly increasing at $h(d, x, \tau)$. Thus

$$h(1, x, \tau) = Q_{Y_{1x}|C_x}(F_{Y_{0x}|C_x}(h(0, x, \tau))).$$

Let $y = h(0, x, \tau) \in \mathcal{S}_{Y_{0x}|C_x}^\circ = \mathcal{S}_{Y_{0x}|X=x}^\circ$. Then $\tau = h^{-1}(0, x, y)$ and the above equation becomes

$$\phi_x(y) = Q_{Y_{1x}|C_x}(F_{Y_{0x}|C_x}(y)),$$

showing that ϕ_x is identified on $\mathcal{S}_{Y_{0x}|X=x}^\circ$. Then ϕ_x is identified on $\mathcal{S}_{Y_{0x}|X=x}$ by its continuous extension. \square

PROOF OF LEMMA 2. The only if part is straightforward by the definition of $\phi_x(\cdot)$. Thus it suffices to show the if part. Suppose $\phi_x(\cdot)$ is identified on $\mathcal{S}_{Y_{0x}}$. Fix x . By definition, $h(1, x, \cdot) = \phi_x(h(0, x, \cdot))$ on $[0, 1]$. Then, conditional on $X = x$, we have $h(1, x, \varepsilon) = YD + \phi_x(Y)(1 - D)$ and $h(0, x, \varepsilon) = Y(1 - D) + \phi_x^{-1}(Y)D$. Thus, under Assumption E(i) we can identify $F_{h(d,x,\varepsilon)} = F_{h(d,x,\varepsilon)|X=x}$ as follows: For all $t \in \mathbb{R}$,

$$F_{h(1,x,\varepsilon)}(t) = \mathbb{P}(YD + \phi_x(Y)(1 - D) \leq t | X = x),$$

$$F_{h(0,x,\varepsilon)}(t) = \mathbb{P}(Y(1 - D) + \phi_x^{-1}(Y)D \leq t | X = x).$$

Furthermore, for all $\tau \in [0, 1]$, $F_{h(d,x,\varepsilon)}(h(d, x, \tau)) = F_\varepsilon(\tau) = \tau$ by Assumption E(ii). Thus, for all $\tau \in (0, 1)$ we have $h(d, x, \tau) \in \mathcal{S}_{Y_{dx}}^\circ$ and $h(d, x, \tau) = Q_{h(d,x,\varepsilon)}(\tau)$. That is, $h(d, x, \cdot)$ is identified on $(0, 1)$ and hence on $[0, 1]$ by continuity. In particular, $h(d, x, 0) = \underline{Y}_{dx}$, which is the lower bound of the support of $\mathcal{S}_{Y_{dx}|X=x} = \mathcal{S}_{Y|D=d,X=x}$ by Lemma 1, while $h(d, x, 1) = Q_{h(d,x,\varepsilon)}(1) = \overline{Y}_{dx}$ by left-continuity of the quantile function. \square

PROOF OF THEOREM 3. By Lemma 3 and Assumption F, $\Delta_d(\cdot, x, z_1, z_2)$ for $d = 0, 1$ are strictly increasing on $\mathcal{S}_{Y_{dx}|X=x}$. We now prove $\mathcal{S}_{Y_{dx}|X=x} = \mathcal{S}_{Y|X=x,D=d}$ by contradiction. Clearly, $\mathcal{S}_{Y|X=x,D=d} \subseteq \mathcal{S}_{Y_{dx}|X=x}$. Suppose w.l.o.g. $\mathcal{S}_{Y|X=x,D=0} \subsetneq \mathcal{S}_{Y_{0x}|X=x}$. Therefore, there exists an interval $I_{\varepsilon x}$ in $\mathcal{S}_{\varepsilon|X=x}$ such that $\mathbb{P}(\varepsilon \in I_{\varepsilon x}; D = 0 | X = x) = 0$. In other words, conditional on $X = x$, all individual with $\varepsilon \in I_{\varepsilon x}$ choose $D = 1$ almost surely. Thus, $\mathcal{S}_{Y_{1x}|X=x,\varepsilon \in I_{\varepsilon x}} \subseteq \mathcal{S}_{Y|X=x,D=1}$, where the latter support is identified. For $e_1 < e_2 \in I_{\varepsilon x}$, note that

$$\begin{aligned} & \mathbb{P}(Y \leq h(1, x, e_2); D = 1 | X = x, Z = z) - \mathbb{P}(Y \leq h(1, x, e_1); D = 1 | X = x, Z = z) \\ &= \mathbb{P}(\varepsilon \in (e_1, e_2] | X = x), \quad \forall z \in \mathcal{S}_{Z|X=x}, \end{aligned}$$

where we have used Assumption A(ii). Hence, for $z_1, z_2 \in \mathcal{S}_{Z|X=x}$ with $p(x, z_1) < p(x, z_2)$, we have

$$\begin{aligned} & \Delta_1(h(1, x, e_2), x, z_1, z_2) - \Delta_1(h(1, x, e_1), x, z_1, z_2) \\ &= \mathbb{P}(Y \leq h(1, x, e_2); D = 1 | X = x, Z = z_2) \\ & \quad - \mathbb{P}(Y \leq h(1, x, e_1); D = 1 | X = x, Z = z_2) \\ & \quad - \mathbb{P}(Y \leq h(1, x, e_2); D = 1 | X = x, Z = z_1) \\ & \quad + \mathbb{P}(Y \leq h(1, x, e_1); D = 1 | X = x, Z = z_1) = 0, \end{aligned}$$

which contradicts the piecewise strict monotonicity of $\Delta_1(\cdot, x, z_1, z_2)$ on $\mathcal{S}_{Y|X=x, D=1} \subseteq \mathcal{S}_{Y_{1x}|X=x}$. Therefore, we have $\mathcal{S}_{Y|X=x, D=0} = \mathcal{S}_{Y_{0x}|X=x}$. Similarly, we obtain $\mathcal{S}_{Y|X=x, D=1} = \mathcal{S}_{Y_{1x}|X=x}$. The identification of $\phi_x(\cdot)$ and $h(d, x, \cdot)$ follows directly from Lemmas 4 and 2, respectively. \square

PROOF OF THEOREM 4. Our proof is in two steps: First, we show that

$$h(1, x, \tau) = Q_{Y_{1x}|m(x, z_1) < \eta \leq m(x, z_2)}(F_{Y_{0\tilde{x}}|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(h(0, \tilde{x}, \tau))) \quad \forall \tau \in [0, 1].$$

Second, we show that the distribution of $h(1, x, \varepsilon)$ is identified, from which we identify the function $h(1, x, \cdot)$.

Fix $x \in \mathcal{S}_X$ and $\tilde{x} \in \mathcal{S}_X$ satisfying Assumption C'. Let further $z_1, z_2 \in \mathcal{S}_{Z|X=x}$ and $\tilde{z}_1, \tilde{z}_2 \in \mathcal{S}_{Z|X=\tilde{x}}$ be such that $p(x, z_1) = p(\tilde{x}, \tilde{z}_1) < p(x, z_2) = p(\tilde{x}, \tilde{z}_2)$. By definition, the right-hand side of (10) is weakly increasing and left-continuous. For any $\tau \in [0, 1]$, let $\psi(0, \tilde{x}, \tau) = \sup\{e : h(0, \tilde{x}, e) = h(0, \tilde{x}, \tau)\}$. Clearly, $\psi(0, \tilde{x}, \tau) \geq \tau$ by definition. Moreover, Assumption A' implies that

$$F_{Y_{0\tilde{x}}|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(h(0, \tilde{x}, \tau)) = F_{\varepsilon|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(\psi(0, \tilde{x}, \tau)).$$

Therefore,

$$\begin{aligned} & Q_{Y_{1x}|m(x, z_1) < \eta \leq m(x, z_2)}(F_{Y_{0\tilde{x}}|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(h(0, \tilde{x}, \tau))) \\ &= Q_{Y_{1x}|m(x, z_1) < \eta \leq m(x, z_2)}(F_{\varepsilon|m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(\psi(0, \tilde{x}, \tau))) \\ &= Q_{Y_{1x}|m(x, z_1) < \eta \leq m(x, z_2)}(F_{\varepsilon|m(x, z_1) < \eta \leq m(x, z_2)}(\psi(0, \tilde{x}, \tau))). \end{aligned}$$

The last step comes from the fact that $m(x, z_j) = m(\tilde{x}, \tilde{z}_j)$ for $j = 1, 2$. Note that

$$\begin{aligned} & \mathbb{P}[Y_{1x} \leq h(1, x, \psi(0, \tilde{x}, \tau)) | m(x, z_1) < \eta \leq m(x, z_2)] \\ & \geq \mathbb{P}[\varepsilon \leq \psi(0, \tilde{x}, \tau) | m(x, z_1) < \eta \leq m(x, z_2)] = F_{\varepsilon|m(x, z_1) < \eta \leq m(x, z_2)}(\psi(0, \tilde{x}, \tau)), \end{aligned}$$

which implies

$$Q_{Y_{1x}|m(x, z_1) < \eta \leq m(x, z_2)}(F_{\varepsilon|m(x, z_1) < \eta \leq m(x, z_2)}(\psi(0, \tilde{x}, \tau))) \leq h(1, x, \psi(0, \tilde{x}, \tau)).$$

Moreover, for any $y < h(1, x, \psi(0, \tilde{x}, \tau))$,

$$\begin{aligned} & \mathbb{P}[Y_{1x} \leq y | m(x, z_1) < \eta \leq m(x, z_2)] \\ &= \mathbb{P}[Y_{1x} \leq h(1, x, \psi(0, \tilde{x}, \tau)) | m(x, z_1) < \eta \leq m(x, z_2)] \\ &\quad - \mathbb{P}[y < Y_{1x} \leq h(1, x, \psi(0, \tilde{x}, \tau)) | m(x, z_1) < \eta \leq m(x, z_2)] \\ &< F_{\varepsilon | m(x, z_1) < \eta \leq m(x, z_2)}(\psi(0, \tilde{x}, \tau)), \end{aligned}$$

where the last inequality follows from

$$\mathbb{P}[y < h(1, x, \varepsilon) \leq h(1, x, \psi(0, \tilde{x}, \tau)) | m(x, z_1) < \eta \leq m(x, z_2)] > 0$$

for any $y < h(1, x, \psi(0, \tilde{x}, \tau))$ by Assumptions A' and D'. Thus, we have

$$Q_{Y_{1x} | m(x, z_1) < \eta \leq m(x, z_2)}(F_{\varepsilon | m(x, z_1) < \eta \leq m(x, z_2)}(\psi(0, \tilde{x}, \tau))) = h(1, x, \psi(0, \tilde{x}, \tau)),$$

which gives us

$$Q_{Y_{1x} | m(x, z_1) < \eta \leq m(x, z_2)}(F_{Y_{0\tilde{x}} | m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(h(0, \tilde{x}, \tau))) = h(1, x, \psi(0, \tilde{x}, \tau)).$$

By definition of $\psi(0, \tilde{x}, \tau)$ and Assumption C'(ii), there is $h(1, x, \psi(0, \tilde{x}, \tau)) = h(1, x, \tau)$ for all $\tau \in [0, 1]$. Hence, by letting $y = h(0, \tilde{x}, \tau) \in \mathcal{S}_{Y_{0\tilde{x}} | X=x}$, the above equation implies that

$$Q_{Y_{1x} | m(x, z_1) < \eta \leq m(x, z_2)}(F_{Y_{0\tilde{x}} | m(\tilde{x}, \tilde{z}_1) < \eta \leq m(\tilde{x}, \tilde{z}_2)}(y)) = h(1, x, h^{-1}(0, \tilde{x}, y)) \equiv \phi_{x, \tilde{x}}(y).$$

Note that $\mathcal{S}_{Y_{0\tilde{x}} | X=\tilde{x}} = \mathcal{S}_{Y | D=0, X=\tilde{x}}$ under Assumption D' since $(\varepsilon, \eta) \perp (X, Z)$ under Assumptions A'(ii), B(i), and E'(i).

For the second step, note that $\phi_{x, \tilde{x}}(h(0, \tilde{x}, \cdot)) = h(1, x, \cdot)$ under Assumptions A'(i) and C'(ii). Therefore, for any $y \in \mathbb{R}$,

$$\begin{aligned} F_{Y_{1x}}(y) &\equiv \mathbb{P}[h(1, x, \varepsilon) \leq y] \\ &= \mathbb{P}[h(1, x, \varepsilon) \leq y; m(x, z_1) \geq \eta] + \mathbb{P}[h(1, x, \varepsilon) \leq y; m(\tilde{x}, \tilde{z}_1) < \eta] \\ &= \mathbb{P}(Y \leq y; D = 1 | X = x, Z = z_1) + \mathbb{P}[\phi_{x, \tilde{x}}(Y) \leq y; D = 0 | X = \tilde{x}, Z = \tilde{z}_1]. \end{aligned}$$

Now we show the identification of $h(1, x, \cdot)$ from $F_{Y_{1x}}$. Because of the weak monotonicity of h in ε and Assumption E'(ii), we have

$$\tau = \mathbb{P}(\varepsilon \leq \tau) \leq \mathbb{P}[Y_{1x} \leq h(1, x, \tau)].$$

It follows that $Q_{Y_{1x}}(\tau) \leq h(1, x, \tau)$. Moreover, fix arbitrary $y < h(1, x, \tau)$. Then

$$\mathbb{P}(Y_{1x} \leq y) = \mathbb{P}(Y_{1x} \leq y; \varepsilon \leq \tau) = \mathbb{P}(\varepsilon \leq \tau) - \mathbb{P}(\varepsilon \leq \tau; Y_{1x} > y) < \tau,$$

where the first equality is because $Y_{1x} \equiv h(1, x, \varepsilon) \leq y$ implies $\varepsilon \leq \tau$, while the inequality comes from the fact that $\mathbb{P}(\varepsilon \leq \tau; Y_{1x} > y) > 0$ under Assumptions A'(i) and E'(ii). Thus, $Q_{Y_{1x}}(\tau) > y$ for all $y < h(1, x, \tau)$. Hence, $Q_{Y_{1x}}(\tau) = h(1, x, \tau)$. \square

REFERENCES

- Angrist, J. D. and A. B. Krueger (1991), "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics*, 106 (4), 979–1014. [594, 599]
- Athey, S. (2001), "Single crossing properties and the existence of pure strategy equilibria in games of incomplete information." *Econometrica*, 69 (4), 861–889. [590]
- Athey, S. and G. W. Imbens (2006), "Identification and inference in nonlinear difference-in-differences models." *Econometrica*, 74 (2), 431–497. [590]
- Bajari, P. and A. Hortacsu (2003), "The winner's curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions." *RAND Journal of Economics*, 34, 329–355. [595]
- Chamberlain, G. (1986), "Asymptotic efficiency in semi-parametric models with censoring." *Journal of Econometrics*, 32 (2), 189–218. [591]
- Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014), "Local identification of non-parametric and semiparametric models." *Econometrica*, 82 (2), 785–809. [590, 591]
- Chernozhukov, V. and C. Hansen (2005), "An IV model of quantile treatment effects." *Econometrica*, 73 (1), 245–261. [590, 591, 592, 596, 599, 600, 601]
- Chernozhukov, V. and C. Hansen (2013), "Quantile models with endogeneity." *Annual Review of Economics*, 5, 57–81. [590, 591]
- Chesher, A. (2003), "Identification in nonseparable models." *Econometrica*, 71 (5), 1405–1441. [590, 592, 596]
- Chesher, A. (2005), "Nonparametric identification under discrete variation." *Econometrica*, 73 (5), 1525–1550. [590, 591, 593, 594]
- D'Haultfœuille, X. and P. Février (2015), "Identification of nonseparable triangular models with discrete instruments." *Econometrica*, 83 (3), 1199–1210. [590]
- Feng, Q., Q. Vuong, and H. Xu (2016), "Nonparametric estimation of heterogeneous individual treatment effects with endogenous treatments." Preprint. Available at arXiv:1610.08899. [599]
- Gautier, E. and S. Hoderlein (2015), "A triangular treatment effect model with random coefficients in the selection equation." Preprint. Available at arXiv:1109.0362. [602]
- Guerre, E., I. Perrigne, and Q. Vuong (2000), "Optimal nonparametric estimation of first-price auctions." *Econometrica*, 68 (3), 525–574. [594, 595, 599]
- Guerre, E., I. Perrigne, and Q. Vuong (2009), "Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions." *Econometrica*, 77 (4), 1193–1227. [590, 595, 599]
- Heckman, J. (1990), "Varieties of selection bias." *American Economic Review*, 80 (2), 313–318. [591]

- Heckman, J. J. (1979), "Sample selection bias as a specification error." *Econometrica*, 47, 153–161. [592, 593, 594]
- Heckman, J. J., J. Smith, and N. Clements (1997), "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts." *The Review of Economic Studies*, 64 (4), 487–535. [590, 592]
- Heckman, J. J. and E. J. Vytlačil (1999), "Local instrumental variables and latent variable models for identifying and bounding treatment effects." *Proceedings of the National Academy of Sciences*, 96 (8), 4730–4734. [591]
- Heckman, J. J. and E. J. Vytlačil (2007), "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation." *Handbook of econometrics*, 6, 4779–4874. [596]
- Hoderlein, S. and E. Mammen (2007), "Identification of marginal effects in nonseparable models without monotonicity." *Econometrica*, 75 (5), 1513–1518. [592]
- Imbens, G. W. and J. D. Angrist (1994), "Identification and estimation of local average treatment effects." *Econometrica*, 62 (2), 467–475. [591, 592, 593, 598]
- Imbens, G. W. and D. B. Rubin (1997), "Estimating outcome distributions for compliers in instrumental variables models." *The Review of Economic Studies*, 64 (4), 555–574. [590, 593]
- Imbens, G. W. and J. M. Wooldridge (2009), "Recent developments in the econometrics of program evaluation." *Journal of Economic Literature*, 47 (1), 5–86. [592, 596]
- Jun, S. J., J. Pinkse, and H. Xu (2011), "Tighter bounds in triangular systems." *Journal of Econometrics*, 161 (2), 122–128. [591]
- Kasy, M. (2014), "Instrumental variables with unrestricted heterogeneity and continuous treatment." *The Review of Economic Studies*, 81 (4), 1614–1636. [592]
- Manski, C. F. (1990), "Nonparametric bounds on treatment effects." *American Economic Review*, 80 (2), 319–323. [591]
- Matzkin, R. L. (1999), "Nonparametric estimation of nonadditive random functions." Discussion paper, Northwestern University. [592]
- Matzkin, R. L. (2003), "Nonparametric estimation of nonadditive random functions." *Econometrica*, 71 (5), 1339–1375. [592, 596]
- Matzkin, R. L. (2008), "Identification in nonparametric simultaneous equations models." *Econometrica*, 76 (5), 945–978. [590]
- Milgrom, P. and C. Shannon (1994), "Monotone comparative statics." *Econometrica*, 62, 157–180. [590]
- Reny, P. J. (2011), "On the existence of monotone pure-strategy equilibria in Bayesian games." *Econometrica*, 79 (2), 499–553. [590]

- Rubin, D. B. (1974), “Estimating causal effects of treatments in randomized and nonrandomized studies.” *Journal of Educational Psychology*, 66 (5), 688. [592]
- Shaikh, A. M. and E. J. Vytlacil (2011), “Partial identification in triangular systems of equations with binary dependent variables.” *Econometrica*, 79 (3), 949–955. [604]
- Torgovitsky, A. (2015), “Identification of nonseparable models using instruments with small support.” *Econometrica*, 83 (3), 1185–1197. [590]
- Vytlacil, E. (2002), “Independence, monotonicity, and latent index models: An equivalence result.” *Econometrica*, 70 (1), 331–341. [592]
- Vytlacil, E. and N. Yildiz (2007), “Dummy endogenous variables in weakly separable models.” *Econometrica*, 75 (3), 757–779. [591, 602, 603, 604]
- Wooldridge, J. M. (2015), “Control function methods in applied econometrics.” *Journal of Human Resources*, 50 (2), 420–445. [602]

Co-editor Rosa L. Matzkin handled this manuscript.

Manuscript received 19 May, 2015; final version accepted 9 October, 2016; available online 25 October, 2016.