COUNTERFACTUAL MAPPING AND INDIVIDUAL TREATMENT EFFECTS IN NONSEPARABLE MODELS WITH DISCRETE ENDOGENEITY*

QUANG VUONG† AND HAIQING XU‡

ABSTRACT. 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, and the instrumental variable can take binary values. We distinguish the cases where the model includes or does not include a selection equation for the binary endogenous regressor. First, we establish point identification of the structural function when it 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 with its counterfactual. This then identifies every individual treatment effect. Second, we characterize all the testable restrictions on observables imposed by the model with or without the selection equation. Lastly, 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

Date: Wednesday 20th August, 2014.
*A previous version of this paper was circulated under the title "Identification of nonseparable models with a binary endogenous regressor". We thank Jason Abrevaya, Federico Bugni, Karim Chalak, Xiaohong Chen, Victor Chernozhukov, Denis Chetverikov, Stephen Donald, Jinyong Hahn, Stefan Hoderlein, Shakeeb Khan, Brendan Kline, Qi Li, Robert Lieli, Matthew Masten, Isabelle Perrigne, Geert Ridder, Xiaoxia Shi, Robin Sickles, Elie Tamer, Edward Vytlacil and Nese Yildiz for useful comments. We also thanks seminar participants at Duke, Rice, NYU, UCLA, 2013 Bilkent University Annual Summer Workshop, 23rd Annual Meeting of Midwest Econometrics Group and 2014 Cowles Summer Conference.
†(corresponding author) Department of Economics, New York University, 19 W. 4th Street, 6FL, New York, NY, 10012, qvuong@nyu.edu.
‡Department of Economics, University of Texas at Austin, BRB 3.160, Mailcode C3100, Austin, TX, 78712, h.xu@austin.utexas.edu.

1. INTRODUCTION AND RELATED LITERATURE

The primary aim of this paper is to establish nonparametric identification of individual treatment effects in a nonseparable model with a binary endogenous regressor. We focus on the case where the instrumental variable has limited variations, in particular when it takes only binary values. We introduce a counterfactual mapping, which links each realized outcome with its counterfactual to identify individual treatment effects. As a by–product, we also identify the structural function. The key idea is to exploit information from the "complier group" in the presence of endogeneity following Imbens and Angrist (1994). Our matching approach generalizes Vytlacil and Yildiz (2007)'s results to the fully nonseparable case.

The secondary objective of this paper is to investigate the identification power as well as the restrictions on observations provided by the selection equation, such as its weak monotonicity. In particular, we derive an easily verifiable high level condition that is sufficient for identification of the structural function and show that the weak monotonicity of the selection equation is sufficient for such a condition. Moreover, we characterize all the testable restrictions on observables imposed by the model with or without the selection equation.

Nonparametric identification of nonseparable models has become central for understanding the source of identification power in structural models, especially in models with discrete endogenous variables, see, e.g., Chesher (2005), Heckman and Vytlacil (2005) and the references therein. Theoretic models that admit structural relationships usually just provide qualitative properties, e.g., monotonicity between economic variables of interest (see e.g. Milgrom and Shannon, 1994; Athey, 2001, 2002; Reny, 2011). For the related empirical analysis, introducing additional parametric restrictions on the functional form, especially on the form of heterogeneity, without careful justifications may lead to spurious identification; see Heckman and Robb (1986) and Imbens and Wooldridge (2009).

The literature on nonseparable models and treatment effects is vast, see e.g. Imbens (2007), Imbens and Wooldridge (2009) for surveys. Our paper also considers the identification of the structural functions in nonseparable models with discrete endogenous regressors. Identification of such functions was studied in a setting with continuous endogenous regressors; see Chesher (2003), Matzkin (2008), D'Haultfœuille and Février (2011), Torgovitsky (2011), among others. For the binary endogenous variable case, Chernozhukov and Hansen (2005), Chen, Chernozhukov, Lee, and Newey (2013) and Chernozhukov and Hansen (2013) establish identification of the structural function without requiring a selection equation. Further, Chesher (2005) establishes partial identification of the structural function at some conditional quantile of the error term under local 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.

The idea behind our identification strategy differs from the above literature. It is based on identifying the counterfactual mapping that relates each individual outcome to its counterfactual under the monotonicity of the outcome equation. This then identifies the structural function. Similar to Imbens and Angrist (1994), we consider an exogenous change in the instrumental variable, which affects the distribution of outcomes through a subpopulation called the "complier" group. We can then identify the conditional probability distribution of the (potential) outcome for this subpopulation at each value of the binary endogenous variable. From these two conditional distributions, we can identify constructively the counterfactual mapping relating the quantiles of one distribution to the other for the whole population.

Our counterfactual mapping generalizes Vytlacil and Yildiz (2007), who match conditional expectations of two distributions — rather than quantiles of two distributions — to identify the average treatment effects (ATE) in a weakly nonseparable triangular model. The idea of matching several 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, and exploited by D'Haultfœuille and Février (2011) and Torgovitsky (2011) in a triangular model with continuous endogenous regressors.

Key among our identifying conditions is the (strong or weak) monotonicity of the structural function and the selection equation. Without any monotonicity restriction, Manski (1990) derives sharp bounds for the average treatment effect (ATE) with and without instrumental variables. Using an instrumental variable and the weak monotonicity assumption on the selection equation, Imbens and Angrist (1994) establish point identification of the local average treatment effect (LATE). Alternatively, in a similar setting, Heckman and Vytlacil (1999, 2005) develop the marginal treatment effect (MTE) and establish its identification by using local variations in instrumental variables. In this paper, we impose a strong/weak monotonicity assumption on the outcome equation, which allows us to constructively identify the counterfactual mapping, as well as the individual treatment effects.

In our setting, we allow the instrumental variable to take only binary values, a case where identification at infinity obviously fails; see, e.g., Heckman (1990). When the outcome variable is continuously distributed, we assume that the outcome equation is strictly monotone in the error term. Under this assumption, our identification strategy only requires a binary–valued instrumental variable. On the other hand, when the outcome variable has a mass probability, then the strict monotonicity condition is no longer proper given that the error term is usually assumed to be continuously distributed. We then replace the strong monotonicity by the weak monotonicity of the outcome equation. As a consequence, our rank condition requires more variations in the instrumental variable, though a finite support is still allowed.

Our method is related to the instrumental variable approach developed in Chernozhukov and Hansen (2005) and generalized by Chen, Chernozhukov, Lee, and Newey (2013) and Chernozhukov and Hansen (2013). The instrumental variable approach does not require a selection equation. Identification then relies on a full rank condition of an equation system. In our setting, we exploit the identification power from the monotonicity of some identified functions to deliver a weaker sufficient condition for identification in a constructive way. Moreover, we characterize all the testable restrictions on observables imposed by the model with and without the selection equation, thereby allowing researchers to assess the model from data.

The structure of the paper is organized as follows. We introduce our benchmark model in Section 2 and its identification is established in Section 3. Section 4 extends our identification mechanism to the case where there is no selection equation. Section 5 characterizes the restrictions imposed on data by the model with or without the selection equation. Section 6 generalizes our method to the case where the distribution of the outcome variable has some mass points. Section 7 concludes.

2. The Benchmark Model

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

$$Y = h(D, X, \epsilon), \tag{1}$$

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

where Y is the outcome variable, $D \in \{0, 1\}$ is a binary endogenous variable, $X \in \mathbb{R}^{d_X}$ is a vector of observed exogenous covariates and $Z \in \mathbb{R}^{d_Z}$ are instrumental variables for the binary endogenous variable D. The error terms ϵ and η are assumed to be scalar valued disturbances.[1] The functions h and m are unknown structural relationships, where h is a function of interest.

[1] Note that η being scalar is not essential and can be relaxed by the monotonicity assumption in Imbens and Angrist (1994). When $\epsilon \in \mathbb{R}^k$ with $k \geq 2$, there always exists a (known) bijection from $\mathbb{R}^k$ to $\mathbb{R}$ that depends on neither d nor x. With more structure such as when (1) is $Y = h(D, X, \epsilon_D)$, we can extend our results when $\{\epsilon_0, \epsilon_1\}$ are identically distributed conditional on η as in Chernozhukov and Hansen (2005, A4-(b)). Specifically, $h(d, x, \cdot)$ is identified under our assumptions below.

Following standard convention, we refer to (1) and (2) as the outcome equation and the selection equation, respectively. Note that (2) covers the general setting where $D = g(X, Z, \eta)$ under additional standard assumptions. Specifically, suppose that g is non–increasing 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 $g(x, z, \eta) = 1\{m(x, z) - \eta \geq 0\}$ for all (x, z). For a detailed discussion, see e.g. Vytlacil (2002, 2006).

A running example of this model is the return to education (see e.g. Chesher, 2005): let Y, D and (X, Z) be 'earnings', 'schooling' and 'demographics', respectively. Moreover, let ϵ be job related ability and η be education related talent. Intuitively, these two latent variables should be correlated to each other, which accounts for the endogeneity problem in the triangular system. The difference between demographics X and Z is that Z affects the education level of an individual, but not the wage outcome. For instance, Z could be some institutional features of the education system. Another example drawn from the auction literature arises when $h(D, X, \cdot)$ is the equilibrium bidding strategy with risk averse bidders, D indicates the level of competition (the number of bidders), X are characteristics of the auctioned object and ϵ is bidder's private value, see e.g. Guerre, Perrigne, and Vuong (2000, 2009). This example is especially relevant as strict monotonicity and nonseparability in the private value ϵ arises from auction theory.

The key to our identification strategy is to match $h(1, x, \cdot)$ with $h(0, x, \cdot)$ under weak conditions, i.e. $h(1, x, \cdot) = \phi_x(h(0, x, \cdot))$, where the function ϕ_x can be identified in a constructive manner. We call ϕ_x the counterfactual mapping because we can find the counterfactual outcome $h(1, x, e)$ from $h(0, x, e)$ for any $e \in \mathscr{S}_{h(0,x,\epsilon)}$. Then the individual treatment effect can be defined by

$$h(1, X, \epsilon) - h(0, X, \epsilon) = D(Y - \phi_X^{-1}(Y)) + (1 - D)(\phi_X(Y) - Y), \qquad (3)$$

which is identified as soon as ϕ_x is identified for all $x \in \mathscr{S}_X$. Moreover, given ϕ_x and any (d, x), we can also identify $h(d, x, \cdot)$ as the quantile function of $F_{h(d,x,\epsilon)}$ following Matzkin (2003).

We make the following assumptions on the benchmark model.

Assumption A. *Equation (1) holds where (i) h is continuous and strictly increasing in ϵ, with ϵ distributed as $U[0,1]$, and (ii) (X, Z) is independent of ϵ.*

Assumption B. *Equation (2) holds where (i) (X, Z) is independent of (ϵ, η), and (ii) the distribution of (ϵ, η) is absolutely continuous with respect to the Lebesgue measure with a rectangular support.*

In assumption A, the continuity and strict monotonicity of h follow e.g. Matzkin (1999, 2003), Chesher (2003) and Chernozhukov and Hansen (2005). This assumption rules out mass points in the distribution of Y. In Section 5, we generalize our results by allowing h to be flat inside the support of ϵ. The uniform distribution of ϵ on $[0,1]$ is a normalization which is standard in the literature (e.g., Chesher, 2003; Chernozhukov and Hansen, 2005).

Assumption B specifies the selection equation. Condition (i) strengthens assumption A–(ii) by including the disturbance term η in the independence condition. It is worth emphasizing that we require the instrumental variable Z to be fully independent of (ϵ, η), which is stronger than the local independence restriction imposed by Chesher (2005). Condition (ii) on the joint distribution of (ϵ, η) is standard and weak. It rules out the degenerate case.[2] Moreover, if the objects of interest are the individual treatment effects but not the structural function h, then the independence assumption could be relaxed to $Z \perp (\epsilon, \eta) | X$. Under such a weaker condition, we can still identify the counterfactual mapping ϕ_x as our identification argument is conditional on $X = x$.

Under assumption B, the function m is identified up to the marginal distribution function $F_\eta(\cdot)$, i.e., $m(x, z) = F_\eta^{-1}(\mathbb{E}(D|X = x, Z = z))$ for all $(x, z) \in \mathscr{S}_{XZ}$. Moreover,

[2] There are some overlaps in assumptions A and B. To simplify exposition, we keep them as is.

for any $x \in \mathscr{S}_X$ and $t \in \mathbb{R}$, we have by assumptions A and B

$$\mathbb{P}(h(1,x,\epsilon) \le t; \eta \le m(x,z)) = \mathbb{P}(Y \le t; D = 1|X = x, Z = z);$$
$$\mathbb{P}(h(1,x,\epsilon) \le t; \eta > m(x,z)) = \mathbb{P}(Y \le t; D = 0|X = x, Z = z);$$

In contrast, neither $\mathbb{P}(h(1,x,\epsilon) \le t; \eta > m(x,z))$ nor $\mathbb{P}(h(0,x,\epsilon) \le t; \eta \le m(x,z))$ is identified yet. Next, suppose there are $z_1, z_2 \in \mathscr{S}_{Z|X=x}$ with $p(x,z_1) < p(x,z_2)$, where $p(x,z) = \mathbb{E}(D = 1|X = x, Z = z)$ is the propensity score. Then, the conditional distribution $F_{h(d,x,\epsilon)|m(x,z_1)<\eta \le m(x,z_2)}$ can be identified as follows: For $d = 0, 1$,

$$\begin{aligned}
&\mathbb{P}\big(h(d,x,\epsilon) \le t | m(x,z_1) < \eta \le m(x,z_2)\big) \\
&= \frac{\mathbb{P}\big[h(d,x,\epsilon) \le t; m(x,z_1) < \eta \le m(x,z_2)\big]}{\mathbb{P}\big[m(x,z_1) < \eta \le m(x,z_2)\big]} \\
&= \frac{\mathbb{P}(Y \le t; D = d|X = x, Z = z_2) - \mathbb{P}(Y \le t; D = d|X = x, Z = z_1)}{\mathbb{P}(D = d|X = x, Z = z_2) - \mathbb{P}(D = d|X = x, Z = z_1)}. \qquad (4)
\end{aligned}$$

As a matter of fact, the above calculation follows Imbens and Angrist (1994) and Vytlacil and Yildiz (2007) who derive similar expressions for the expectations, while we consider the full conditional distributions. Equation (4) is used in the next section.

3. IDENTIFICATION

In this section, we develop a simple and constructive approach for the identification of ϕ_x and h under assumptions A and B. In particular, our method involves a weak rank condition: the instrumental variable Z takes at least two values such that the propensity score has a non–degenerate support, i.e., $p(x,z_1) \neq p(x,z_2)$ for some $z_1, z_2 \in \mathscr{S}_{Z|X=x}$.

We first provide a lemma that shows that identification of $\phi_x(\cdot)$ is sufficient for identifying $h(d,x,\cdot)$ under assumption A. Fix $x \in \mathscr{S}_X$. For each $y \in \mathscr{S}_{h(0,x,\epsilon)}$, let $\phi_x(y) \equiv h\big(1, x, h^{-1}(0,x,y)\big)$, where $h^{-1}(0,x,\cdot)$ is the inverse function of $h(0,x,\cdot)$. Under assumption A, the counterfactual mapping ϕ_x is well–defined, continuous and strictly increasing from $\mathscr{S}_{h(0,x,\epsilon)}$ onto $\mathscr{S}_{h(1,x,\epsilon)}$. Clearly, $h(1,x,\cdot) = \phi_x(h(0,x,\cdot))$ on $[0,1]$.

Lemma 1. *Suppose assumption A holds. Then, for any* $x \in \mathscr{S}_X$, $h(0,x,\cdot)$ *and* $h(1,x,\cdot)$ *are identified on* $[0,1]$ *if and only if* $\phi_x(\cdot)$ *is identified on* $\mathscr{S}_{h(0,x,\epsilon)}$.

Proof. See Appendix A.1 ☐

Lemma 1 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 $X = x$,

$$h(1,x,\epsilon) = YD + \phi_x(Y)(1-D), \quad h(0,x,\epsilon) = \phi_x^{-1}(Y)D + Y(1-D). \tag{5}$$

Then, the identification of ϕ_x provides the distributions of $h(0,x,\epsilon)$ and $h(1,x,\epsilon)$, which further identify $h(0,x,\cdot)$ and $h(1,x,\cdot)$ under assumption A, respectively.

To identify the counterfactual mapping $\phi_x(\cdot)$, we now introduce a rank condition.

Assumption C (Rank Condition)**.** *For any* $x \in \mathscr{S}_X$, $\mathscr{S}_{p(X,Z)|X=x}$ *is not a singleton.*

When $\mathscr{S}_{Z|X} = \{z_1, z_2\}$, the Rank Condition is satisfied if $p(x,z_1) \neq p(x,z_2)$ for any $x \in \mathscr{S}_X$. Though weak, this rank condition can be relaxed further as is discussed at the end of this section.

For a generic random variable W, let Q_W be the quantile function of its distribution, i.e. $Q_W(\tau) = \inf\{q \in \mathbb{R} : \mathbb{P}(W \leq q) \geq \tau\}$ for any $\tau \in [0,1]$. Further, for any x, d and a generic subset $A \subseteq \mathscr{S}_\eta$, let $Q_{h(d,x,\epsilon)|\eta\in A}$ be the conditional quantile function of $h(d,x,\epsilon)$ given $\eta \in A$, i.e. $Q_{h(d,x,\epsilon)|\eta\in A}(\tau) = \inf\{q : \mathbb{P}[h(d,x,\epsilon) \leq q | \eta \in A] \geq \tau\}$.

Theorem 1. *Suppose assumptions A to C hold. For every* $x \in \mathscr{S}_X$, *let* $z_1, z_2 \in \mathscr{S}_{Z|X=x}$ *such that* $p(x,z_1) < p(x,z_2)$. *Then the support of* $h(d,x,\epsilon)$ *for* $d = 0,1$ *is identified by* $\mathscr{S}_{h(d,x,\epsilon)} = \mathscr{S}_{Y|D=d,X=x}$. *Moreover, the counterfactual mapping* $\phi_x(\cdot)$ *is identified on* $\mathscr{S}_{h(0,x,\epsilon)}$ *by*

$$\phi_x(\cdot) = Q_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}\big(F_{h(0,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\cdot)\big), \tag{6}$$

where $Q_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\cdot)$ *and* $F_{h(0,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\cdot)$ *are identified by (4).*

Proof. See Appendix A.2 ☐

It should be noted that Theorem 1 actually holds pointwise in x. On the other hand, the requirement that the rank condition holds for any $x \in \mathscr{S}_X$ in assumption C is useful to identify every individual treatment effect in the whole population by (3). Moreover, using the counterfactual mapping $\phi_x(\cdot)$ and (5), we can identify other policy effects of interest, e.g.,

$$\begin{aligned}
\text{ATE} &\equiv \mathbb{E}\big[h(1,X,\epsilon) - h(0,X,\epsilon)\big] = \mathbb{E}\Big\{D\big[Y - \phi_X^{-1}(Y)\big] + (1-D)\big[\phi_X(Y) - Y\big]\Big\};\\
\text{ATT} &\equiv \mathbb{E}\big[h(1,X,\epsilon) - h(0,X,\epsilon)\big|D=1\big] = \mathbb{E}\big[Y - \phi_X^{-1}(Y)\big|D=1\big];\\
\text{QTE} &\equiv Q_{h(1,X,\epsilon)}(\cdot) - Q_{h(0,X,\epsilon)}(\cdot) = Q_{D\cdot Y + (1-D)\cdot\phi_X(Y)}(\cdot) - Q_{(1-D)\cdot Y + D\cdot\phi_X^{-1}(Y)}(\cdot).
\end{aligned}$$

where ATT and QTE respectively refer to the average treatment effect on the treated and the quantile treatment effect, see, e.g., Heckman and Robb (1986). It is worth emphasizing that Theorem 1 still holds under the weaker conditional independence assumption $Z \perp (\epsilon, \eta)|X$, used in the literature to identify the preceding policy effects.

In particular, if one imposes additivity on h, i.e. $h(D,X,\epsilon) = h^*(D,X) + \epsilon$, then $\phi_x(y) = h^*(1,x) - h^*(0,x) + y$. By Theorem 1, we can show that for any $\tau \in [0,1]$,

$$h^*(1,x) - h^*(0,x) = Q_{h(1,x,\epsilon)|m(x,z_1)<\eta\le m(x,z_2)}(\tau) - Q_{h(0,x,\epsilon)|m(x,z_1)<\eta\le m(x,z_2)}(\tau).$$

Thus, by integrating out τ, we obtain

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

Note that the RHS has the same expression as the (conditional) LATE in Imbens and Angrist (1994). This is not surprising, since there is no unobserved heterogeneity in individual treatment effects as $h(1,x,\epsilon) - h(0,x,\epsilon) = h^*(1,x) - h^*(0,x)$. Thus, individual treatment effects are the same as (conditional) LATE or conditional ATE, which is defined as $\mathbb{E}[h(1,X,\epsilon) - h(0,X,\epsilon)|X = x]$.[3] Further, additive separability of the structural function h in ϵ can be tested by testing the slope of $\phi_x(\cdot)$ equals 1.

By Lemma 1 and Theorem 1, the identification of h follows.

Corollary 1. *Suppose that assumptions A to C hold. For every $x \in \mathscr{S}_X$, $d = 0,1$ and $\tau \in [0,1]$, $h(d,x,\tau)$ is identified as the τ–th quantile of the distribution of $F_{h(d,x,\epsilon)}$, where*

$$F_{h(1,x,\epsilon)}(t) = \mathbb{P}\big[YD + \phi_x(Y)(1-D) \le t|X = x\big];$$
$$F_{h(0,x,\epsilon)}(t) = \mathbb{P}\big[Y(1-D) + \phi_x^{-1}(Y)D \le t|X = x\big].$$

It is worthnoting that the copula of (ϵ,η) is also identified on $[0,1] \times \mathscr{S}_{p(X,Z)}$ by

$$C(\tau,p) = \mathbb{P}(Y \le h(1,x,\tau); D = 1|X = x, p(X,Z) = p).$$

This is because, conditional on $X = x$ and $p(X,Z) = p$, the event $\{Y \le h(1,x,\tau); D = 1\}$ is equivalent to $\big\{\epsilon \le \tau; \eta \le Q_\eta(p)\big\}$. In particular, the copula is identified for all $\tau \in [0,1]$ but only for $p \in \mathscr{S}_{p(X,Z)}$ which is in general a subset of $[0,1]$ when the propensity score has limited variations.

As mentioned above, we can further relax the rank condition in Corollary 1 by allowing some x with a singleton support $\mathscr{S}_{p(X,Z)|X=x}$. To see this, let $(x,z) \in \mathscr{S}_{XZ}$ and $\mathscr{S}_{p(X,Z)|X=x} = \{p(x,z)\}$. Suppose that there exists $(\tilde{x},\tilde{z}) \in \mathscr{S}_{XZ}$ such that assumption C is satisfied for $\tilde{x}$ with $p(\tilde{x},\tilde{z}) = p(x,z)$. By Corollary 1, $h(d,\tilde{x},\cdot)$ is identified for $d = 0,1$. Moreover, for any $\tau \in [0,1]$,

$$\begin{aligned}\mathbb{P}(Y \le h(1,x,\tau)|D = 1, X = x, Z = z) &= \mathbb{P}(\epsilon \le \tau|\eta \le m(x,z)) \\ &= \mathbb{P}(\epsilon \le \tau|\eta \le m(\tilde{x},\tilde{z})) = \mathbb{P}\big[Y \le h(1,\tilde{x},\tau)|D = 1, X = \tilde{x}, Z = \tilde{z}\big],\end{aligned}$$

where the second step comes from $\mathbb{1}(\eta \le m(\tilde{x},\tilde{z})) \overset{a.s.}{=} \mathbb{1}(\eta \le m(x,z))$ since $p(\tilde{x},\tilde{z}) = p(x,z)$. Thus $h(1,x,\tau) = Q_{Y|D=1,X=x,Z=z}\big\{F_{Y|D=1,X=\tilde{x},Z=\tilde{z}}\,(h(1,\tilde{x},\tau))\,\big\}$. A similar result holds for $h(0,x,\tau)$.

[3] From such an intuition, (2) is not even needed for this simple expression of $h^*(1,x) - h^*(0,x)$.

4. Extending the identification argument

In this section, we investigate the identification power arising from the selection equation, as well as from variations in the instrumental variable Z. In particular, we drop 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, Theorem 2). Throughout, we maintain assumption A and for simplicity, we let Z be a binary variable.

Under assumption A, we have $\mathbb{P}\big[Y \leq h(D,X,\tau)|X = x, Z = z\big] = \tau$ for all $z \in \mathscr{S}_{Z|X=x}$ 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 for all $z \in \mathscr{S}_{Z|X=x}$,

$$\mathbb{P}\big[Y \leq h(1,x,\tau); D = 1|X = x, Z = z\big] + \mathbb{P}\big[Y \leq h(0,x,\tau); D = 0|X = x, Z = z\big] = \tau.$$

Suppose $z_1, z_2 \in \mathscr{S}_{Z|X=x}$ with $p(x,z_1) < p(x,z_2)$. Thus, we obtain

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

i.e.,

$$\Delta_0(h(0,x,\tau),x,z_1,z_2) = \Delta_1(h(1,x,\tau),x,z_1,z_2), \tag{7}$$

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

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

By definition, $\Delta_d(\cdot,x,z_1,z_2)$ is identified for $d = 0,1$. Let $y = h(0,x,\tau)$ in (7) so that $h(1,x,\tau) = \phi_x(y)$. Since τ is arbitrary, then

$$\Delta_0(y,x,z_1,z_2) = \Delta_1(\phi_x(y),x,z_1,z_2), \ \ \forall y \in \mathscr{S}_{h(0,x,\epsilon)}. \tag{8}$$

Our general identification result is based on (8) and exploits the strict monotonicity of $\phi_x(\cdot)$. Recall that under assumption B, $\Delta_d(y, x, z_1, z_2) = \mathbb{P}[h(d, x, \epsilon) \le y; m(x, z_1) < \eta \le m(x, z_2)]$, which is continuous and strictly increasing in $y \in \mathscr{S}_{h(d,x,\epsilon)}$, thereby identifying $\phi_x(\cdot) = \Delta_1^{-1}(\Delta_0(\cdot, x, z_1, z_2), x, z_1, z_2)$ on $\mathscr{S}_{h(0,x,\epsilon)}$. This suggests that identification of the counterfactual mapping ϕ_x can be achieved under weaker conditions than assumption B.

Definition 1 (Piecewise Monotone)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ and $S \subseteq \mathbb{R}$. We say that g is piecewise weakly (strictly) monotone on the support S if S can be partitioned into a countable number of non–overlapping intervals such that g is weakly (strictly) monotone in every interval.*

Lemma 2. *Suppose assumption A holds. Given $x \in \mathscr{S}_X$, suppose $z_1, z_2 \in \mathscr{S}_{Z|X=x}$ with $p(x, z_1) < p(x, z_2)$. Then, $\Delta_1(\cdot, x, z_1, z_2)$ is piecewise weakly (strictly) monotone on $\mathscr{S}_{h(1,x,\epsilon)}$ if and only if $\Delta_0(\cdot, x, z_1, z_2)$ is piecewise weakly (strictly) monotone on $\mathscr{S}_{h(0,x,\epsilon)}$.*

Proof. See Appendix A.3 □

Assumption D. *Suppose $\Delta_d(\cdot, x, z_1, z_2)$ is piecewise weakly monotone on $\mathscr{S}_{h(d,x,\epsilon)}$, $d = 0, 1$.*

Assumption D is weak. In particular, $\Delta_d(\cdot, x, z_1, z_2)$ may be discontinuous. On the other hand, if $\Delta_d(\cdot, x, z_1, z_2)$ is continuously differentiable as in Chernozhukov and Hansen (2005, Theorem 2-i), then assumption D holds.

Lemma 3. *Let $x \in \mathscr{S}_X$ such that $z_1, z_2 \in \mathscr{S}_{Z|X=x}$ with $p(x, z_1) < p(x, z_2)$. Suppose (i) $\phi_x(\cdot) : \mathscr{S}_{h(0,x,\epsilon)} \to \mathscr{S}_{h(1,x,\epsilon)}$ is continuous and strictly increasing, and (ii) equation (8) and assumption D hold where $\Delta_d(\cdot, x, z_1, z_2)$ and $\mathscr{S}_{h(d,x,\epsilon)}$ are known for $d = 0, 1$. Then, $\phi_x(\cdot)$ is identified if and only if $\Delta_d(\cdot, x, z_1, z_2)$ is piecewise strictly monotone on $\mathscr{S}_{h(d,x,\epsilon)}$ for some d.*

Proof. See Appendix A.4 □

Lemma 3 is useful when identification is based on (8) only. In this case, under the maintained assumption D, Lemma 3 provides a necessary and sufficient condition for the identification of $\phi_x(\cdot)$.

We now extend Theorem 1 and Corollary 1.

Theorem 2. *Suppose that assumptions A and C hold. Let $x \in \mathscr{S}_X$ such that $z_1, z_2 \in \mathscr{S}_{Z|X=x}$ with $p(x, z_1) < p(x, z_2)$. Suppose $\Delta_d(\cdot, x, z_1, z_2)$ is piecewise strictly monotone on $\mathscr{S}_{Y|X=x,D=d}$ for $d = 0, 1$. Then the support of $h(d, x, \epsilon)$ is identified by $\mathscr{S}_{h(d,x,\epsilon)} = \mathscr{S}_{Y|D=d,X=x}$. Moreover, $\phi_x(\cdot)$ and $h(d, x, \cdot)$ are identified on the supports $\mathscr{S}_{h(0,x,\epsilon)}$ and $[0, 1]$, respectively.*

Proof. See Appendix A.5. □

5. CHARACTERIZATION OF MODEL RESTRICTIONS

In empirical applications, an important question is whether to adopt the nonseparable model (1) with or without the selection equation (2). As shown in Section 3, the selection equation provides a simple and constructive identification result, but can introduce additional restrictions on the data. In this section, we characterize all the restrictions on observables imposed by the model with or without the selection equation. These restrictions are useful for developing model selection and model specification tests.

Formally, we denote these two models by

$$\mathcal{M}_0 \equiv \left\{ [h, F_{\epsilon D|XZ}] : \text{assumption A holds} \right\};$$
$$\mathcal{M}_1 \equiv \left\{ [h, m, F_{\epsilon\eta|XZ}] : \text{assumptions A and B hold} \right\}.$$

To simplify, hereafter we assume $\mathscr{S}_{XZ} = \mathscr{S}_X \times \{z_1, z_2\}$. Moreover, $p(x, z_1) < p(x, z_2)$ for all $x \in \mathscr{S}_X$. We say that a conditional distribution $F_{YD|XZ}$ of observables is rationalized by model $\mathcal{M}$ if and only if there exists a structure in $\mathcal{M}$ that generates $F_{YD|XZ}$.

Theorem 3. *A conditional distribution $F_{YD|XZ}$ can be rationalized by $\mathcal{M}_0$ if and only if*

(i) *$F_{Y|DXZ}$ is a continuous conditional CDF;*
(ii) *for each $x \in \mathscr{S}_X$, there exists a continuous and strictly increasing mapping $g_x : \mathbb{R} \to \mathbb{R}$ such that g_x maps $\mathscr{S}_{(1-D)Y+Dg_x^{-1}(Y)|X=x}$ onto $\mathscr{S}_{DY+(1-D)g_x(Y)|X=x}$ and*

$$\Delta_0(\cdot, x, z_1, z_2) = \Delta_1(g_x(\cdot), x, z_1, z_2). \tag{9}$$

Proof. See Appendix A.6 □

In Theorem 3, the key restriction is the existence of a solution to (9), which may not be unique.

We now turn to $\mathcal{M}_1$. Let $\mathscr{C}$ be a collection of differentiable and strictly increasing copula functions with support $(0,1]^2$.

Theorem 4. *A conditional distribution $F_{YD|XZ}$ rationalized by $\mathcal{M}_0$ can also be rationalized by $\mathcal{M}_1$ if and only if*

(i) *for any $x \in \mathscr{S}_X$ and $d \in \{0,1\}$, $\mathscr{S}_{Y|D=d,X=x}$ is an interval of $\mathbb{R}$, and $\Delta_d(\cdot, x, z_1, z_2)$ is continuous strictly increasing on $\mathscr{S}_{Y|D=d,X=x}$ and flat elsewhere;*

(ii) *there exists a copula function $C \in \mathscr{C}$ such that for any $(x,z) \in \mathscr{S}_{XZ}$ and $y \in \mathbb{R}$,*

$$\mathbb{P}\big[Y \le y; D = 1|X = x, Z = z\big] = C(\tau_x(y), p(x,z)),$$

where $\tau_x(\cdot) = \mathbb{P}(DY + (1-D)\Delta_1^{-1}(\Delta_0(Y, x, z_1, z_2), x, z_1, z_2) \le \cdot|X = x)$.

Proof. See Appendix A.7 □

Condition (i) strengthens Theorem 3–(ii): By Lemma 3, the fact that $\Delta_d(\cdot, x, z_1, z_2)$ is continuous, strictly increasing on $\mathscr{S}_{Y|D=d,X=x}$ and flat elsewhere implies that (9) has a unique solution $g_x(\cdot) = \phi_x(\cdot)$. Because (ϵ, η) is continuously distributed with a rectangular support, $\mathscr{S}_{Y|D=d,X=x}$ is an interval under assumption A. Condition (ii) comes from the identification of the copula of (ϵ, η) on $[0,1] \times \mathscr{S}_{p(X,Z)}$ under the additional assumption B. Moreover, condition (ii) implies that $\mathbb{P}\big[Y \le y; D = 1|X = x, Z = z\big] = \mathbb{P}\big[Y \le y; D = 1|p(X,Z) = p(x,z)\big]$ and the latter is strictly increasing in $p(x,z)$. Further, $\mathbb{P}\big[Y \le y; D = 1|X = x, Z = z_2\big] - \mathbb{P}\big[Y \le y; D = 1|X = x, Z = z_1\big]$ is strictly increasing in y, see also Mourifie and Wan (2014).[4]

[4]See also Vuong and Xu (2014) for a characterization of all the restrictions associated with two other models: $\{[h, F_{\epsilon D|XZ}]$: ϵ is continuously distributed on an interval and assumption A–(ii) holds$\}$ and $\{[h, m, F_{\epsilon\eta|XZ}]$: assumption B holds$\}$.

6. GENERALIZATION

This section provides another extension of Theorem 1. Specifically, we maintain assumption B but relax the continuity and strict monotonicity assumption of h so that our method applies to the case where the outcome variable has a probability mass in its distribution such as when the outcome variable is censored or binary (see e.g. Wooldridge, 2013).

Assumption E. *Equation (1) holds where (i) h is left–continuous and weakly increasing in* ϵ*, with* ϵ *distributed as* $U[0,1]$*, and (ii)* (X, Z) *is independent of* ϵ*.*

The left–continuity of h is a normalization for the identification of the structural function at its discontinuous points. For simplicity, the following rank condition is introduced for the identification of $h(1, x, \cdot)$. The identification of $h(0, x, \cdot)$ can be obtained similarly.

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

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

Condition (i) of assumption F requires that there exist $z_1, z_2 \in \mathscr{S}_{Z|X=x}$ and $\tilde{z}_1, \tilde{z}_2 \in \mathscr{S}_{Z|X=\tilde{x}}$ 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: For any $\tau_1, \tau_2 \in (0, 1)$,

$$F_{h(0,\tilde{x},\epsilon)|m(\tilde{x},\tilde{z}_1)<\eta\leq m(\tilde{x},\tilde{z}_2)}(\tau_1) = F_{h(0,\tilde{x},\epsilon)|m(\tilde{x},\tilde{z}_1)<\eta\leq m(\tilde{x},\tilde{z}_2)}(\tau_2)$$
$$\Longrightarrow F_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\tau_1) = F_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\tau_2).$$

Thus, we can verify whether any given $\tilde{x} \in \mathscr{S}_X$ satisfies assumption F. When $h(d, x, \cdot)$ is strictly monotone in ϵ, by setting $\tilde{x} = x$, assumption F reduces to assumption C.

Fix x and let $\tilde{x}$ satisfy assumption F. 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 $y \in \mathscr{S}_{h(0,\tilde{x},\epsilon)}$.[5] If $\tilde{x} = x$, then $\phi_{x,\tilde{x}}(\cdot)$ reduces to the counterfactual mapping $\phi_x(\cdot)$ in Section 3. Let $(z_1, z_2) \in \mathscr{S}_{Z|X=x}$ and $(\tilde{z}_1, \tilde{z}_2)\mathscr{S}_{Z|X=\tilde{x}}$ 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 Theorem 1 and Corollary 1.

Theorem 5. *Suppose assumptions B and E hold. Given x and $\tilde{x}$ satisfying assumption F, then $\phi_{x,\tilde{x}}(\cdot)$ is identified by*

$$\phi_{x,\tilde{x}}(\cdot) = Q_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}\big(F_{h(0,\tilde{x},\epsilon)|m(\tilde{x},\tilde{z}_1)<\eta\leq m(\tilde{x},\tilde{z}_2)}(\cdot)\big). \tag{10}$$

and for any $\tau \in [0,1]$, $h(1, x, \tau)$ is identified as the τ–th quantile of the distribution

$$F_{h(1,x,\epsilon)}(\cdot) = \mathbb{P}(Y \leq \cdot; D = 1|X = x) + \mathbb{P}\big[\phi_{x,\tilde{x}}(Y) \leq \cdot; D = 0|X = \tilde{x}\big].$$

Proof. See Appendix A.8 □

Similarly to assumption C, assumption F can be weaken. To illustrate Theorem 5, we discuss the generalized rank condition (assumption F) 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 Vytlacil and Yildiz (2007).

Example 1. *Suppose (X, Z) is independent of (ϵ, η) and the distribution of (ϵ, η) has a non–degenerate rectangular support with $\epsilon \sim U(0,1)$. Moreover, let*

$$Y = h^*(D, X, \epsilon)\mathbb{1}\big(h^*(D, X, \epsilon) \geq 0\big),$$
$$D = \mathbb{1}\big(m(X, Z) \geq \eta\big),$$

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

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

Fix $x \in \mathscr{S}_X$. For $d = 0, 1$, let τ_{dx} solve $h^*(d, x, \tau_{dx}) = 0$. W.l.o.g., let $\tau_{dx} \in (0, 1)$. Thus, assumption F is satisfied if there exists an $\tilde{x} \in \mathscr{S}_X$ 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 $z_1, z_2 \in \mathscr{S}_{Z|X=x}$, $\tilde{z}_1, \tilde{z}_2 \in \mathscr{S}_{Z|X=\tilde{x}}$.

Example 2. *Let Y and D denote a binary outcome variable and a binary endogenous regressor, respectively. Consider*

$$Y = \mathbb{1}\big(h^*(D, X) \geq \epsilon\big),$$
$$D = \mathbb{1}\big(m(X, Z) \geq \eta\big).$$

To identify $h^*(1, x)$ for some x, assumption F requires that there exists $\tilde{x} \in \mathscr{S}_X$ 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 $z_1, z_2 \in \mathscr{S}_{Z|X=x}$, $\tilde{z}_1, \tilde{z}_2 \in \mathscr{S}_{Z|X=\tilde{x}}$. Thus, assumption F reduces exactly to the support condition in Vytlacil and Yildiz (2007).

7. Conclusion

This paper establishes nonparametric identification of the counterfactual mapping and individual treatment effects 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. Moreover, we consider two extensions: One without a selection equation and the other with weak monotonicity in the outcome equation.

As indicated in Section 3, the counterfactual mapping and individual treatment effects can be identified under the weaker independence assumption $Z \perp (\epsilon, \eta) | X$ and the rank condition (assumption C). Some policy effects such as ATE can also be identified. It is important to note that such results do not require exogeneity of X. On the other hand, identification of the structural function h, which is necessary for some counterfactuals in empirical IO, requires the full independence $(X, Z) \perp (\epsilon, \eta)$ in assumption B.

To conclude, one needs to develop a nonparametric estimation method for the structural function h. To simplify ideas, we consider the benchmark model in Section 3 where X is discrete and Z is binary. We further assume $p(x,z_1) \neq p(x,z_2)$ for all $(x,z_1),(x,z_2) \in \mathscr{S}_{XZ}$. A method will be first to estimate the counterfactual mapping $\phi_x(\cdot)$ and its inverse by

$$\hat{\phi}_x(\cdot) = \hat{Q}_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\hat{F}_{h(0,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\cdot)),$$
$$\hat{\phi}_x^{-1}(\cdot) = \hat{Q}_{h(0,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\hat{F}_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\cdot)),$$

where $\hat{Q}_{h(d,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\cdot)$ is the quantile function of

$$\hat{F}_{h(d,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(\cdot) = \frac{\sum_{i=1}^n \mathbb{1}(Y_i \leq \cdot; D_i = d; X_i = x)[\mathbb{1}(Z_i = z_2) - \mathbb{1}(Z_i = z_1)]}{\sum_{i=1}^n \mathbb{1}(D_i = d; X_i = x)[\mathbb{1}(Z_i = z_2) - \mathbb{1}(Z_i = z_1)]}.$$

In the second step, one estimates the structural function for any $\tau \in (0,1)$ by

$$\hat{h}(1,x,\tau) = \hat{Q}_{YD+\phi_x(Y)(1-D)|X=x}(\tau), \;\; \hat{h}(0,x,\tau) = \hat{Q}_{Y(1-D)+\phi_x^{-1}(Y)D|X=x}(\tau),$$

which are respectively the τ–th quantiles of

$$\hat{F}_{YD+\phi_x(Y)(1-D)|X=x}(\cdot) = \frac{\sum_{i=1}^n \mathbb{1}(YD + \hat{\phi}_x(Y)(1-D) \leq \cdot; X_i = x)}{\sum_{i=1}^n \mathbb{1}(X_i = x)},$$
$$\hat{F}_{Y(1-D)+\phi_x^{-1}(Y)D|X=x}(\cdot) = \frac{\sum_{i=1}^n \mathbb{1}(Y(1-D) + \hat{\phi}_x^{-1}(Y)D \leq \cdot; X_i = x)}{\sum_{i=1}^n \mathbb{1}(X_i = x)}.$$

The above estimation procedure can be easily implemented.

Excluding possibly boundaries, the above suggests that the estimators of the counterfactual mapping $\phi_x(\cdot)$ and the structural function h are $\sqrt{n}$–consistent. Their precise asymptotic distributions can be derived using the functional delta method and the composition lemma in Van Der Vaart and Wellner (1996, Lemma 3.9.27). If Z is continuous, we conjecture that $\phi_x(\cdot)$ and h are still estimated at $\sqrt{n}$–rate. This is because we can average out Z and obtain $\mathbb{E}F_{h(0,x,\epsilon)|m(x,Z_1)<\eta\leq m(x,Z_2)}(\cdot) = \mathbb{E}F_{h(1,x,\epsilon)|m(x,Z_1)<\eta\leq m(x,Z_2)}(\phi_x(\cdot))$ where Z_1 and Z_2 are two different observations. On the other hand, if X is continuous, the estimation of ϕ_x and h will no longer be $\sqrt{n}$–consistent. An important question is then the optimal rate for estimating these functions. Moreover, we should note that even in the benchmark model, there might exist some values $\tilde{x} \neq x$ satisfying assumption F–(i). In this case, one has other counterfactual mapping $\phi_{x,\tilde{x}}$ that can be used to identify h. A second question is to use this property to improve efficiency.

References

ATHEY, S. (2001): "Single crossing properties and the existence of pure strategy equilibria in games of incomplete information," *Econometrica*, 69(4), 861–889.

——— (2002): "Monotone comparative statics under uncertainty," *The Quarterly Journal of Economics*, 117(1), 187–223.

ATHEY, S., AND G. W. IMBENS (2006): "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 74(2), 431–497.

CHEN, X., V. CHERNOZHUKOV, S. LEE, AND W. NEWEY (2013): "Local identification of nonparametric and semiparametric models," *Econometrica, Forthcoming*.

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV model of quantile treatment effects," *Econometrica*, 73(1), 245–261.

——— (2013): "Quantile Models with Endogeneity," *Annual Review of Economics*, 5, 57–81.

CHESHER, A. (2003): "Identification in nonseparable models," *Econometrica*, 71(5), 1405–1441.

——— (2005): "Nonparametric identification under discrete variation," *Econometrica*, 73(5), 1525–1550.

D'HAULTFŒUILLE, X., AND P. FÉVRIER (2011): "Identification of nonseparable models with endogeneity and discrete instruments," Discussion paper, working paper.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal nonparametric estimation of first–price auctions," *Econometrica*, 68(3), 525–574.

——— (2009): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica*, 77(4), 1193–1227.

HECKMAN, J. (1990): "Varieties of selection bias," *The American Economic Review*, 80(2), 313–318.

HECKMAN, J. J., AND R. ROBB (1986): *Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes*. Springer.

HECKMAN, J. J., AND E. VYTLACIL (2005): "Structural equations, treatment Effects, and econometric policy evaluation1," *Econometrica*, 73(3), 669–738.

HECKMAN, J. J., AND E. J. VYTLACIL (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.

IMBENS, G. W. (2007): "Nonadditive models with endogenous regressors," *Econometric Society Monographs*, 43, 17.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62(2), 467–475.

IMBENS, G. W., AND J. M. WOOLDRIDGE (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5–86.

JUN, S. J., J. PINKSE, AND H. XU (2011): "Tighter bounds in triangular systems," *Journal of Econometrics*, 161(2), 122–128.

MANSKI, C. F. (1990): "Nonparametric bounds on treatment effects," *The American Economic Review*, 80(2), 319–323.

MATZKIN, R. L. (1999): "Nonparametric Estimation of Nonadditive Random Functions," Discussion paper, Northwestern University.

——— (2003): "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71(5), 1339–1375.

——— (2008): "Identification in nonparametric simultaneous equations models," *Econometrica*, 76(5), 945–978.

MILGROM, P., AND C. SHANNON (1994): "Monotone comparative statics," *Econometrica: Journal of the Econometric Society*, pp. 157–180.

MOURIFIE, I., AND Y. WAN (2014): "Testing LATE Assumptions," *Available at SSRN 2429664*.

RENY, P. J. (2011): "On the existence of monotone pure–strategy equilibria in Bayesian games," *Econometrica*, 79(2), 499–553.

TORGOVITSKY, A. (2011): "Identification of nonseparable models with general instruments," Discussion paper, Yale University.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence*. Springer.

VUONG, Q., AND H. XU (2014): "Characterization of Model Restrictions in a Triangular Model," Discussion paper, Working Paper.

VYTLACIL, E. (2002): "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 70(1), 331–341.

——— (2006): "A note on additive separability and latent index models of binary choice: representation results," *Oxford Bulletin of Economics and Statistics*, 68(4), 515–518.

VYTLACIL, E., AND N. YILDIZ (2007): "Dummy endogenous variables in weakly separable models," *Econometrica*, 75(3), 757–779.

WOOLDRIDGE, J. (2013): "Control Function Methods in Applied Econometrics," Discussion paper, Michigan State University.

Appendix A. Proofs

A.1. **Proof of Lemma 1.**

Proof. The only if part is straightforward by the definition of $\phi_x(\cdot)$ and it suffices to show the if part. Suppose $\phi_x(\cdot)$ is identified. Fix x. By definition, $h(1,x,\cdot) = \phi_x(h(0,x,\cdot))$ on $[0,1]$. Then, conditional on $X = x$, we have $h(1,x,\epsilon) = YD + \phi_x(Y)(1-D)$ and $h(0,x,\epsilon) = Y(1-D) + \phi_x^{-1}(Y)D$. Therefore we can identify $F_{h(d,x,\epsilon)}$ as follows: for all $t \in \mathbb{R}$,

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

Further, by assumption A, we have $h(d,x,\tau) = Q_{h(d,x,\epsilon)}(\tau)$ for all $\tau \in [0,1]$. □

A.2. **Proof of Theorem 1.**

Proof. Fix $x \in \mathscr{S}_X$ and $z_1, z_2 \in \mathscr{S}_{Z|X=x}$ with $p(x,z_1) < p(x,z_2)$. To show the first part, w.l.o.g. let $d = 1$. By assumption B, $\mathscr{S}_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)} = \mathscr{S}_{h(1,x,\epsilon)}$. Moreover, $\mathscr{S}_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)} \subseteq \mathscr{S}_{h(1,x,\epsilon)|\eta\leq m(x,z_2)} = \mathscr{S}_{Y|D=1,X=x,Z=z_2} \subseteq \mathscr{S}_{Y|D=1,X=x} \subseteq \mathscr{S}_{h(1,x,\epsilon)}$. Hence, all these supports are equal.

For the second part, because h is strictly monotone in ϵ, we have for any $\tau \in (0,1)$

$$F_{h(0,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(h(0,x,\tau)) = F_{\epsilon|m(x,z_1)<\eta\leq m(x,z_2)}(\tau) = F_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(h(1,x,\tau)).$$

Because $F_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}$ is continuous and strictly increasing at $h(1,x,\tau)$, we have

$$h(1,x,\tau) = Q_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}\Big(F_{h(0,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(h(0,x,\tau))\Big).$$

Let $y = h(0,x,\tau)$. Then, $\tau = h^{-1}(0,x,y)$ and the above equation becomes

$$\phi_x(y) = Q_{h(1,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}\Big(F_{h(0,x,\epsilon)|m(x,z_1)<\eta\leq m(x,z_2)}(y)\Big). \quad \square$$

A.3. **Proof of Lemma 2.**

Proof. For the if part, suppose $\Delta_0(\cdot, x, z_1, z_2)$ is piecewise weakly monotone on $\mathscr{S}_{h(0,x,\epsilon)}$. Then, by definition, $\mathscr{S}_{h(0,x,\epsilon)}$ can be partitioned into a sequence of non–overlapping intervals $\{I_j : j = 1, \cdots, J\}$, where $J \in \mathbb{N} \cup \{+\infty\}$, such that $\Delta_0(\cdot, x, z_1, z_2)$ is weakly monotone on every interval.

By assumption A, $\phi_x(\cdot)$ is continuous and strictly increasing. Then, $\mathscr{S}_{h(1,x,\epsilon)}$ can be partitioned into a sequence of non–overlapped intervals $\{\phi_x(I_j) : j = 1, \cdots, J\}$. Moreover, by (8), we have

$$\Delta_1(y, x, z_1, z_2) = \Delta_0\big(\phi_x^{-1}(y), x, z_1, z_2\big), \ \ \forall y \in \mathscr{S}_{h(1,x,\epsilon)}.$$

Clearly, $\Delta_1(\cdot, x, z_1, z_2)$ is weakly monotone in every interval $\phi_x(I_j)$. The only if part can be shown similarly. □

A.4. **Proof of Lemma 3.**

Proof. We first show the if part. W.l.o.g., let $\Delta_0(\cdot, x, z_1, z_2)$ be piecewise strictly monotone on $\mathscr{S}_{h(0,x,\epsilon)}$, which can be partitioned into a sequence of non–overlapping intervals $\{I_j : j = 1, \cdots, J\}$. On each interval I_j, due to the monotonicity, $\Delta_0(\cdot, x, z_1, z_2)$ has at most a countable number of discontinuity points. Hence, $\mathscr{S}_{h(0,x,\epsilon)}$ can be further partitioned into a countable number of non–overlapped intervals such that $\Delta_0(\cdot, x, z_1, z_2)$ is continuous and strictly monotone on every interval. For such a sequence of intervals, we further merge adjacent intervals if $\Delta_0(\cdot, x, z_1, z_2)$ is continuous and strictly monotone on the union of them.

Thus, we partition $\mathscr{S}_{h(0,x,\epsilon)}$ into a countable number of non–overlapped intervals $\{I'_j : j = 1, \cdots, J'\}$, where $J' \in \mathbb{N} \cup \{+\infty\}$, such that $\Delta_0(\cdot, x, z_1, z_2)$ is continuous, strictly monotone on every interval, and discontinuous or achieve a local extreme value at each endpoints of these intervals. By the proof of Lemma 2, (8) and the fact that $\phi_x(\cdot) : \mathscr{S}_{h(0,x,\epsilon)} \to \mathscr{S}_{h(1,x,\epsilon)}$ is continuous and strictly increasing implies that $\mathscr{S}_{h(1,x,\epsilon)}$ can be partitioned into the same number of non–overlapped intervals $\{\phi_x(I'_j) : j = 1, \cdots, J'\}$, such that $\Delta_1(\cdot, x, z_1, z_2)$ is continuous, strictly monotone on every interval, and discontinuous or achieve a local extreme value at each endpoints of these intervals.

Moreover, for each $y \in I'_j$, we solve $\phi_x(y)$ by (8) as

$$\phi_x(y) = \Delta_{1,j}^{-1}(\Delta_{0,j}(y, x, z_1, z_2), x, z_1, z_2),$$

where $\Delta_{0,j}(\cdot, x, z_1, z_2)$ and $\Delta_{1,j}(\cdot, x, z_1, z_2)$ are the projections of $\Delta_0(\cdot, x, z_1, z_2)$ and $\Delta_1(\cdot, x, z_1, z_2)$ on the support I'_j and $\phi_x(I_j)$, respectively.

For the only if part, w.l.o.g., suppose $\Delta_0(\cdot, x, z_1, z_2)$ is constant on a non–degenerate interval $I \subseteq \mathscr{S}_{h(0,x,\epsilon)}$. By the proof of Lemma 2, $\Delta_1(\cdot, x, z_1, z_2)$ is also constant on $\phi_x(I)$. It suffices to construct a continuous and strictly increasing function $\tilde{\phi}_x \neq \phi_x$ such that (8) holds for $\tilde{\phi}_x$. Let $\tilde{\phi}_x(y) = \phi_x(y)$ for all $y \notin I$ and $\tilde{\phi}_x(y) = \phi_x(g(y))$ for all $y \in I$, where g is an arbitrary continuous, strictly increasing mapping from I onto I. Clearly, there are plenty of choices for the function g. Moreover, $\tilde{\phi}_x \neq \phi_x$ if $g(t)$ is not an identity mapping. By construction, $\Delta_1(\tilde{\phi}_x(y), x, z_1, z_2) = \Delta_1(\phi_x(y), x, z_1, z_2)$ holds for all $y \in I$ because $\Delta_1(\cdot, x, z_1, z_2)$ is constant on $\phi_x(I)$. This equation also holds for all $y \notin I$ by the definition of $\tilde{\phi}_x$. Then (8) holds for $\tilde{\phi}_x$. □

A.5. **Proof of Theorem 2.**

Proof. We first show the identification of $\mathscr{S}_{h(d,x,\epsilon)}$. Clearly, $\mathscr{S}_{Y|X=x,D=d} \subseteq \mathscr{S}_{h(d,x,\epsilon)}$. Suppose w.l.o.g. $\mathscr{S}_{Y|X=x,D=0} \subsetneq \mathscr{S}_{h(0,x,\epsilon)}$. Therefore, there exists an interval I_ϵ in $\mathscr{S}_\epsilon$ such that $\mathbb{P}(\epsilon \in I_\epsilon; D = 0|X = x, Z = z) = 0$ and $\mathbb{P}(\epsilon \in I_\epsilon; D = 1|X = x, Z = z) = \mathbb{P}(I_\epsilon)$ for all $z \in \mathscr{S}_{Z|X=x}$. In other words, conditional on $X = x$ and $Z = z$ for any $z \in \mathscr{S}_{Z|X=x}$, all individual with $\epsilon \in I_\epsilon$ choose $D = 1$ almost surely. Thus, $\mathscr{S}_{h(1,x,\epsilon)|\epsilon\in I_\epsilon} \subseteq \mathscr{S}_{Y|X=x,D=1}$ and the latter support is identified. For $e_1 < e_2 \in I_\epsilon$, note that

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

Hence, for $z_1, z_2 \in \mathscr{S}_{Z|X=x}$ with $p(x,z_1) < p(x,z_2)$, we have

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

which contradicts with the piecewise strict monotonicity of $\Delta_1(\cdot, x, z_1, z_2)$ on $\mathscr{S}_{Y|X=x,D=1}$. Therefore, we have $\mathscr{S}_{Y|X=x,D=d} = \mathscr{S}_{h(d,x,\epsilon)}$.

Thus, the identification of $\phi_x(\cdot)$ and $h(d,x,\cdot)$ follows directly from Lemmas 1 and 3. □

A.6. **Proof of Theorem 3.**

Proof. For the only if part, let $g_x(\cdot) = \phi_x(\cdot)$ on the support $\mathscr{S}_{h(0,x,\epsilon)}$. The result is straightforward by Section 4, For the if part, we prove by constructing a structure $S = [h, F_{\epsilon D|XZ}]$ to rationalize the given distribution $F_{YD|XZ}$. Fix an arbitrary x. Let $H_x(y) \equiv \mathbb{P}\big(Y \cdot (1-D) + g_x^{-1}(Y) \cdot D \leq y | X = x\big)$ and $h(0, x, \tau)$ be its τ–th quantile. Moreover, let $h(1, x, \tau) = g_x(h(0, x, \tau))$. Clearly, condition (i) and the fact that g_x is continuous and strictly increasing on the support ensure $h(d, x, \cdot)$ are continuous and strictly increasing on $[0, 1]$. Further let $\mathbb{P}_S(\epsilon \leq \tau; D = d | X = x, Z = z) = \mathbb{P}(Y \leq h(d, x, \tau); D = d | X = x, Z = z)$ for all $z \in \mathscr{S}_{Z|X=x}$ and $(\tau, d) \in [0, 1] \times \{0, 1\}$, where $\mathbb{P}_S$ denote the probability measure under the constructed structure S. We now show that $(X, Z) \perp \epsilon$ and $\epsilon \sim U[0, 1]$.

By (9), we have

$$\begin{aligned}
&\mathbb{P}\big[Y \leq g_x(y); D = 1 | X = x, Z = z_1\big] + \mathbb{P}\big[Y \leq y; D = 0 | X = x, Z = z_1\big] \\
&\qquad = \mathbb{P}\big[Y \leq g_x(y); D = 1 | X = x, Z = z_2\big] + \mathbb{P}\big[Y \leq y; D = 0 | X = x, Z = z_2\big] \\
&\qquad\qquad = \mathbb{P}\big[Y \leq g_x(y); D = 1 | X = x\big] + \mathbb{P}\big[Y \leq y; D = 0 | X = x\big] = H_x(y).
\end{aligned}$$

where the second equality is because $\mathscr{S}_{Z|X=x} = \{z_1, z_2\}$.

Hence, for any $\tau \in [0, 1]$ and $i = 1, 2$,

$$\begin{aligned}
\mathbb{P}_S(\epsilon \leq \tau | X = x, Z = z_i) &= \mathbb{P}\big(Y \leq h(0, x, \tau); D = 0 | X = x, Z = z_i\big) \\
&\quad + \mathbb{P}\big(Y \leq h(1, x, \tau); D = 1 | X = x, Z = z_i\big) = H_x(h(0, x, \tau)) = \tau,
\end{aligned}$$

where the last step is by the construction of $h(0, x, \cdot)$ and the strict monotonicity of $H_x(\cdot)$ on the support $\mathscr{S}_{(1-D)Y + Dg_x^{-1}(Y)|X=x}$. Therefore, $(X, Z) \perp \epsilon$ and $\epsilon \sim U[0, 1]$.

Now it suffices to show that the constructed structure $S = [h, F_{\epsilon D|XZ}]$ generates the given distribution $F_{YD|XZ}$. This is true because for any $(y, d, x, z) \in \mathscr{S}_{YDXZ}$, we have

$$\begin{aligned}
\mathbb{P}_S(Y \leq y; D = d | X = x, Z = z) &= \mathbb{P}_S(\epsilon \leq h^{-1}(d, x, y); D = d | X = x, Z = z) \\
&= \mathbb{P}(Y \leq y; D = d | X = x, Z = z).
\end{aligned}$$

The last step comes from the construction of $F_{\epsilon D|XZ}$.

A.7. **Proof of Theorem 4.**

Proof. By Section 4, the only if part is trivial.

For the if part, suppose condition (i) and (ii) hold for a structure $S_0 = [h, F_{\eta D|XZ}]$ satisfying assumption A. It suffices to construct an observational equivalent structural $S_1 \in \mathcal{M}_1$. Fix an arbitrary $x \in \mathscr{S}_X$. By assumption A, $\epsilon \sim U[0,1]$. By Theorem 2 the structural function $h(d, x, \cdot)$, is identified for $d = 0, 1$. Hence, it suffices to construct m and $F_{\epsilon\eta}$. Let $\eta \sim U[0,1]$ and $m(x,z) = p(x,z)$. Lastly, let C be the copula of the joint distribution $F_{\epsilon\eta}$. Now it suffices to show observational equivalence of the constructed structure $S_1 = [h, m, F_{\epsilon\eta}]$.

Fix $(x,z) \in \mathscr{S}_{XZ}$. We first look at the case where $0 < p(x,z) < 1$. Given the identification of $\phi_x(\cdot)$ by condition (i), we have $\tau_x(y) = h^{-1}(1, x, y)$ in the condition (ii), which further implies that for any $\tau \in (0,1)$,

$$\mathbb{P}(Y \leq h(1, x, \tau); D = 1 | X = x, Z = z) = C(\tau, p(x,z)).$$

Because C is a copula function with non degenerate support $[0,1]^2$, then $\mathscr{S}_{Y|D=1,X=x,Z=z} = \mathscr{S}_{h(1,x,\epsilon)}$. By assumption A, we have

$$\mathbb{P}(Y \leq h(0, x, \tau); D = 0 | X = x, Z = z) + \mathbb{P}(Y \leq h(1, x, \tau); D = 1 | X = x, Z = z) = \tau.$$

Therefore,

$$\begin{aligned} &\mathbb{P}(Y \leq h(0, x, \tau); D = 0 | X = x, Z = z) \\ &\qquad = \tau - \mathbb{P}(Y \leq h(1, x, \tau); D = 1 | X = x, Z = z) = \tau - C(\tau, p(x,z)), \end{aligned} \tag{11}$$

which implies $\mathscr{S}_{Y|D=0,X=x,Z=z} = \mathscr{S}_{h(0,x,\epsilon)}$ by a similar argument as above.

We now show observational equivalence, i.e., for $d = 0, 1$, any $z \in \mathscr{S}_{Z|X=x}$ and $y \in \mathbb{R}$,

$$\mathbb{P}_{S_1}(Y \leq y; D = d | X = x, Z = z) = \mathbb{P}(Y \leq y; D = d | X = x, Z = z).$$

This is true because for any $y \in \mathscr{S}_{h(1,x,\epsilon)}$,

$$\mathbb{P}_{S_1}(Y \leq y; D = 1 | X = x, Z = z) = C(h^{-1}(1, x, y), p(x,z)) = \mathbb{P}(Y \leq y; D = 1 | X = x, Z = z).$$

Moreover, or any $y \in \mathscr{S}_{h(0,x,\epsilon)}$,

$$\begin{aligned}\mathbb{P}_{S_1}(Y \leq y; D=0|X=x, Z=z) &= \mathbb{P}_{S_1}(\epsilon \leq h^{-1}(0,x,y); \eta > p(x,z)) \\ &= h^{-1}(0,x,y) - C(h^{-1}(0,x,y), p(x,z)) = \mathbb{P}(Y \leq y; D=0|X=x, Z=z)\end{aligned}$$

where the last step comes from (11).

When $p(x,z) = 1$, $m(x,z) = p(x,z) = 1$. Note that $\mathbb{P}_{S_1}(Y \leq y; D = 0|X = x, Z = z) = \mathbb{P}(Y \leq y; D = 0|X = x, Z = z) = 0$ for all y. Moreover, we have $\mathscr{S}_{h(1,x,\epsilon)} = \mathscr{S}_{Y|D=1,X=x,Z=z}$. Then, it suffices to show $\mathbb{P}_{S_1}(Y \leq y; D = 1|X = x, Z = z) = \mathbb{P}(Y \leq y; D = 1|X = x, Z = z)$ for all $y \in \mathscr{S}_{h(1,x,\epsilon)}$. This is true because

$$\mathbb{P}_{S_1}(Y \leq h(1,x,\tau); D=1|X=x, Z=z) = \mathbb{P}_{S_1}(\epsilon \leq \tau; \eta \leq 1) = \tau.$$

Moreover, $\mathbb{P}\big[Y \leq h(1,x,\tau); D = 1|X = x, Z = z\big] = C(\tau, 1) = \tau$. A similar argument holds for the case with $p(x,z) = 0$.

A.8. **Proof of Theorem 5.**

Proof. Our proofs take two steps: First, we will show that the constructed mapping defined by (10) satisfies

$$h(1,x,\tau) = \phi_{x,\tilde{x}}\big(h(0,\tilde{x},\tau)\big), \ \ \forall \tau \in [0,1];$$

Second, we will show that the distribution of $h(d,x,\epsilon)$ can be identified, from which we identify function $h(d,x,\cdot)$.

Fix $x \in \mathscr{S}_X$ and $\tilde{x} \in \mathscr{S}_X$ satisfying assumption F. Let further $z_1, z_2 \in \mathscr{S}_{Z|X=x}$ and $\tilde{z}_1, \tilde{z}_2 \in \mathscr{S}_{Z|X=\tilde{x}}$ such that $p(x,z_1) = p(\tilde{x},\tilde{z}_1) < p(x,z_2) = p(\tilde{x},\tilde{z}_2)$. By definition $\phi_{x,\tilde{x}}$ is well defined, weakly increasing and left–continuous.

We now show that $h(1,x,\cdot) = \phi_{x,\tilde{x}}(h(0,\tilde{x},\cdot))$. For any $\tau \in [0,1]$, $\psi(0,\tilde{x},\tau) = \sup\{e : h(0,\tilde{x},e) = h(0,\tilde{x},\tau)\}$. By assumption E, we have

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

Therefore,

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

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

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

which implies

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

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

$$\begin{aligned}& \mathbb{P}\left[h(1,x,\epsilon) \leq y\middle|m(x,z_1) < \eta \leq m(x,z_2)\right]\\ = \;& \mathbb{P}\left[h(1,x,\epsilon) \leq h\big(1,x,\psi(0,\tilde{x},\tau)\big)\middle|m(x,z_1) < \eta \leq m(x,z_2)\right]\\ - \;& \mathbb{P}\left[y < h(1,x,\epsilon) \leq h(1,x,\psi(0,\tilde{x},\tau))\middle|m(x,z_1) < \eta \leq m(x,z_2)\right]\\ < \;& F_{\epsilon|m(x,z_1)<\eta\leq m(x,z_2)}(\psi(0,\tilde{x},\tau)),\end{aligned}$$

where the last inequality comes from the fact that by assumption E, for any $y < h(1,x,\psi(0,\tilde{x},\tau)$,

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

Thus, we have that

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

which gives us that $\phi_{x,\tilde{x}}(h(0,\tilde{x},\tau)) = h\left(1,x,\psi(0,\tilde{x},\tau)\right)$. By the definition of $\psi(0,\tilde{x},\tau)$ and condition (ii), there is $h\left(1,x,\psi(0,\tilde{x},\tau)\right) = h\left(1,x,\tau\right)$. Because τ is arbitrary in $(0,1]$, then $h\left(1,x,\cdot\right) = \phi_{x,\tilde{x}}\big(h(0,\tilde{x},\cdot)\big)$ on $(0,1]$.

For the second step, it is straightforward that the distribution of $h(1,x,\epsilon)$ is identified as

$$F_{h(1,x,\epsilon)}(y) = \mathbb{P}(Y \leq y; D = 1|X = x) + \mathbb{P}\left[\phi_{x,\tilde{x}}(Y) \leq y; D = 0|X = \tilde{x}\right], \ \ \forall y \in \mathbb{R}.$$

Now we show the identification of $h(1,x,\cdot)$ from $F_{h(1,x,\epsilon)}$.

By definition, $Q_{h(1,x,\epsilon)}(\tau) = \inf\ \{y \in \mathbb{R} : \mathbb{P}\left[h(1,x,\epsilon) \leq y\right] \geq \tau\}$. Because of the weakly monotonicity of h in ϵ, we have $h(1,x,u) \leq h\left(1,x,\tau\right)$ for all $u \leq \tau$. Therefore, we have that

$$\mathbb{P}\left[h(1,x,\epsilon) \leq h\left(1,x,\tau\right)\right] \geq \tau.$$

It follows that $Q_{h(1,x,\epsilon)}(\tau) \leq h\left(1,x,\tau\right)$.

Moreover, fix arbitrary $y < h\left(1,x,\tau\right)$. Then

$$\mathbb{P}\left[h(1,x,\epsilon) \leq y\right] = \mathbb{P}\left[h(1,x,\epsilon) \leq y; \epsilon \leq \tau\right] = \mathbb{P}\left[\epsilon \leq \tau\right] - \mathbb{P}\left[\epsilon \leq \tau; h(1,x,\epsilon) > y\right] < \tau.$$

where the first equality is because $h(1,x,\epsilon) \leq y$ implies $\epsilon \leq \tau$, and the inequality comes from the fact that $\mathbb{P}\left[\epsilon \leq \tau; h(1,x,\epsilon) > y\right] > 0$ under assumption E. Thus, $Q_{h(1,x,\epsilon)}(\tau) > y$ for all $y < h(1,x,\tau)$. Hence, $Q_{h(1,x,\epsilon)}(\tau) = h\left(1,x,\tau\right)$. □