Identification of Nonseparable Models with General Instruments

Alexander Torgovitsky*

April 21, 2011†

Abstract

I consider nonparametric identification of a nonseparable model with a continuous endogenous variable (treatment), a scalar unobservable and an excluded instrumental variable. If the first-stage relationship between the instrument and the treatment is strictly monotone in unobservables then many kinds of relevant instruments can be used to identify the levels of the outcome equation. In particular, binary instruments, such as the intent to treat, can be used. This contrasts sharply with related work on nonparametric identification for nonseparable models, which has often required continuous instruments, sometimes with large support. The key insight is that strict

*Ph.D. candidate, Department of Economics, Yale University, a-torgovitsky@northwestern.edu. I thank my advisors, Donald Andrews, Xiaohong Chen and Edward Vytlacil, who have been gracious with their advice, support and feedback. I have benefited from discussions with Joseph Altonji, Lanier Benkard, Steven Berry, Kirill Evdokimov, Jinyong Hahn, Philip Haile, James Heckman, Keisuke Hirano, Yuichi Kitamura, Sokbae (Simon) Lee, Oliver Linton, Geert Ridder, Jean-Marc Robin and Elie Tamer. I thank participants of the 2010 Cowles Foundation Summer Conference, the 2010 World Congress of the Econometric Society, the 2010 New York Area Econometrics Colloquium and of seminars at Yale, Chicago, Northwestern, Columbia, Michigan, UCLA and USC. While writing this paper I was generously supported by a Carl Arvid Anderson fellowship from the Cowles Foundation.

†This paper is based on a portion of my job market paper, which was circulated under the title "Identification and Estimation of Nonparametric Quantile Regressions with Endogeneity" and dated November 12, 2010.

monotonicity (rank invariance) in the first-stage relationship imposes a restriction on the copula function between the treatment and unobservables. I develop several examples of economic models in which strict monotonicity arises naturally from a decision problem, functional form considerations or informal arguments.

JEL classification: C14; C20; C51

Keywords: Nonseparable models, nonparametric identification, instrumental variables, selection, unobserved heterogeneity, rank invariance, copulas, quantile regression

1 Introduction

The classical linear model, $Y = X\beta + U$, with $\beta \in \mathbb{R}$ and U unobserved, is separable because it assumes that the effect of X on Y is β, which is deterministic. The nonparametric model, $Y = m(X) + U$, where m is an unknown function, is also separable because the effect of X is $\nabla_x m(X)$, which is deterministic after conditioning on observables. This paper is about the nonseparable model $Y = m(X, U)$. In this model, the effect of X on Y is $\nabla_x m(X, U)$, which is still stochastic after conditioning on observables because U is unobserved. Nonseparable models can capture generalized unobserved heterogeneity in the effect of X, whereas separable models cannot.

Separable models are frequently used in applications even though they are rarely justified by economic theory or empirical evidence (Heckman, 2001). One reason for this is that nonseparable models present challenging identification problems when X is endogenous (statistically dependent with U) and only a single cross-section of data is available.[1] Standard instrumental variables techniques cannot be applied directly (Heckman and Vytlacil, 2005; Heckman et al., 2006). Modified instrumental variables arguments usually require continuous instruments. Chesher (2003) uses a continuously distributed instrument to identify derivatives under local conditions. To identify levels,

[1]This paper only considers cross-sections. There is a recent body of work that uses the additional dimension of a panel to nonparametrically identify a nonseparable model. The most closely related to this paper are Altonji and Matzkin (2005), Athey and Imbens (2006) and Evdokimov (2010).

a continuous instrument often must have large support, which means that it has positive density over the entire real line (e.g. Heckman, 1990; Imbens and Newey, 2009). However, many commonly available instruments are discrete, such as exogenous policy shifts, natural experiments or the intent to treat in a randomized controlled experiment with partial compliance. Among continuous instruments, few will satisfy the large support assumption. Given that plausibly exogenous instruments are hard to find, these are important limitations.

Other approaches to identifying nonseparable models have come at the cost of economic interpretability. An extreme example is Heckman's (1979) selection model, which doesn't require any instruments. Its identification rests on parametric assumptions about the error distributions. In more recent work, Florens et al. (2008) introduce an instrument and impose a flexible polynomial structure on the outcome equation. As the authors point out, this type of structure can be difficult to generate from economic primitives. Chernozhukov and Hansen (2005), Chernozhukov et al. (2007) and Chen et al. (2011) identify nonseparable instrumental variables models by using nonlinear versions of the completeness condition employed by Newey and Powell (2003).[2] However, the economic content of these completeness conditions has not yet been determined.

In this paper, I present conditions under which m is nonparametrically identified in the nonseparable model $Y = m(X, U)$, when X and U may be arbitrarily dependent and only a cross-section of data is available. I consider the case where Y and X are continuously distributed and m is strictly increasing in U, which is scalar. Identification is obtained by utilizing an excluded instrument that is related to X through a first-stage equation that is strictly monotone in unobservables. This condition has economic content as a rank invariance assumption and allows for the use of discrete-valued instruments. As I show, it can also be formulated as an assumption about the

[2] Chernozhukov and Hansen (2005) use a completeness condition only when the treatment is continuous. Their main results are for discrete treatments.

copula function of (X,U) conditional on the instrument. In Section 2, I discuss the model and its interpretation. In Section 3, I present the identification result and describe the intuition behind it. Section 4 provides examples based on recent empirical work where rank invariance is a reasonable assumption. Section 5 concludes. In Torgovitsky (2011a), I use the identification result to construct a minimum distance from independence estimator and I establish its asymptotic properties.

2 Model

Suppose that a scalar response variable Y is determined by the relationship

$$Y = m^*(X, W, U), \tag{1}$$

where X is a scalar explanatory variable (treatment), W are covariates, U is a scalar unobservable and m^* is an unknown function. I allow for X to be endogenous, i.e. potentially dependent with U, even conditional on W. This situation arises frequently in economics. For example, if Y is an agent's log wage, X is their investment in education and W are observable socioeconomic variables, selection on latent ability suggests that $U \not\perp\!\!\!\perp X|W$. I focus on identifying the distribution of the counterfactual random variable $Y_x \equiv m^*(x, W, U)$, conditional on $W = w$. This distribution describes the impact on Y of exogenously setting $X = x$ for the population subgroup determined by $W = w$. Suppose that there exists a scalar instrument Z that is excluded from (1) and that the following assumptions about (1) hold.[3]

M.C. (Continuity) *For every* $(w, z) \in \mathcal{WZ} \equiv \text{supp}(W, Z)$, $(X, U)|(W, Z) = (w, z)$ *and* $(X, U)|W = w$ *are continuously distributed with connected support and have*

[3] The results throughout this paper extend to cases where X and/or Z are vector-valued. Most of the underlying intuition remains unchanged, but there are some additional details and the notation becomes much more complicated. To keep the exposition clear, I treat these extensions separately in Torgovitsky (2011b).

everywhere differentiable distribution functions.[4]

M.SI. (Scalar heterogeneity) *$m^*(x, w, \cdot)$ is strictly increasing for every $(x, w) \in \mathcal{X}\mathcal{W} \equiv \text{supp}(X, W)$.*

M.C and M.SI imply that the distribution of $Y_x|W$ is continuous and can be related to that of $U|W$ through their quantile functions as $Q_{Y_x|W}(t|w) = m^*(x, w, Q_{U|W}(t|w))$.[5] A normalization is needed to separate the scale of m^* from that of the random variable $U|W$. This was formally shown to be necessary for identification by Matzkin (2003) who considered several different normalizations that arise naturally in econometric models. The normalizations from that work are all applicable to this paper. For concreteness, I focus on the following, which endows m^* with a particularly straightforward interpretation.[6]

N.QR. (Quantile regression normalization) *$U|W \sim Unif[0, 1]$.*

Under N.QR, $Q_{Y_x|W}(u \mid w) = m^*(x, w, u)$ is interpretable as the u^{th} quantile treatment response (QTR) of Y to setting $X = x$ for the population subgroup corresponding to $W = w$. Similarly, quantile treatment effects (QTEs) can be formed for any pair x, x' as $Q_{Y_{x'}|W}(u|w) - Q_{Y_x|W}(u|w) = m^*(x', w, u) - m^*(x, w, u)$. These effects are exogenous (or causal) with respect to X, but not necessarily with respect to the covariates, W. QTEs describe the horizontal distance between two counterfactual distributions and have attracted considerable interest among both theoretical and applied researchers interested in the distributional effects of treatments.[7]

[4]Only almost everywhere differentiability is guaranteed by absolute continuity. This additional technical condition is convenient in the proofs, but not necessary.

[5]For any scalar random variable Y and random element W, $Q_Y(t) \equiv \inf\{y : F_Y(y) \geq t\}$ and $Q_{Y|W}(t|w) \equiv \inf\{y : F_{Y|W}(y \mid w) \geq t\}$. For any scalar– or vector– valued random variable Y, I use the notation $F_{Y|W}(y \mid w) \equiv \mathbb{P}[Y \leq y \mid W = w]$ for $w \in \text{supp}\, W$.

[6]I demonstrate the identification result in Section 3 in a way that is invariant to the normalization used. Other normalizations are provided in Appendix A.

[7]In addition to the previously cited papers, see, e.g., Abadie et al. (2002), Bitler et al. (2006), Firpo (2007), Djebbari and Smith (2008) and Chernozhukov et al. (2009).

For identifying QTEs, the assumption that U is scalar is not, by itself, restrictive. Su et al. (2010) show that if U were a high-dimensional unobservable, then (1) would be observationally equivalent to an equation with a scalar unobservable. This result implies that it is not possible to identify the separate impacts of a high-dimensional U in (1). Indeed, authors who have studied models with high-dimensional unobservables have ultimately identified quantities that integrate over them, e.g. Blundell and Powell (2003), Hoderlein and Mammen (2007) and Imbens and Newey (2009). As Chesher (2007) points out, this is representative of a fundamental trade-off in identification analysis between the number of observable and distinct unobservable quantities. If m and U are desired to have a specific structural interpretation different from that given by N.QR, then requiring U to be scalar can be restrictive.

Given a scalar U, the additional strict monotonicity assumption of M.SI is a further restriction. It has the interpretation of *rank invariance*, meaning that $F_{Y_x|W}(Y_x \mid W) = F_{U|W}(U \mid W) = F_{Y_{x'}|W}(Y_{x'} \mid W)$ for all x, x'. Rank invariance is a concept that was originally introduced by Doksum (1974) and has been recently revisited by Heckman et al. (1997) and Chernozhukov and Hansen (2005). It can be interpreted as positing an underlying proneness or ranking of agents for Y, conditional on W, that is not affected by counterfactual manipulations of X. For example, if rank invariance holds, then relatively high earning (Y_x) 35-year-old white males (W) with a high school education ($X = x$) would also be relatively high earning ($Y_{x'}$) if they had a college education ($X = x'$).[8]

The identification issues in this model arise because the counterfactual distribution $Y_x|W$ is only observed conditional on $X = x$. When X is exogenous, i.e. $U \perp\!\!\!\perp X|W$, the distribution of $Y_x|W$ is equal to the observable distribution of $Y|W, X = x$, so $Q_{Y|WX}(t \mid w, x) = Q_{Y_x|W}(t \mid w) = m^*(x, w, Q_{U|W}(t \mid w))$ is immediately identified from

[8]Chernozhukov and Hansen (2005) also introduce a slightly weaker alternative to rank invariance that they call rank similarity. This allows the ranks to deviate from a common ranking as long as the deviations are exogenous. Assumption M.SI can also be replaced by a rank similarity condition without affecting the results.

the data. A normalization such as N.QR then provides identification of m^*. When $U \not\perp\!\!\!\perp X|W$, $m^*(x, w, Q_{U|W}(t \mid w))$ is not identified because the distribution of $Y_x|W$ differs in an unknown way from that of $Y|W, X = x$. I will show that the excluded instrument, Z, can be used to identify the function m^* on $\mathcal{XWU} \equiv \text{supp}(X, W, U)$ if it satisfies the following assumptions.

Z.FS. ***(First stage equation)*** *There exists a function g such that $X = g(W, Z, V)$, where V is an unobserved scalar and*

- ***Z.FS.EX.*** ***(Exogenous instrument)*** $(V, U) \perp\!\!\!\perp Z|W$.
- ***Z.FS.SI.*** ***(Scalar heterogeneity)*** *$g(w, z, \cdot)$ is strictly increasing for every $(w, z) \in \mathcal{WZ}$.*

Z.R. ***(Relevance)*** *Let $\mathcal{X}_w$ and $\mathcal{Z}_w$ denote $\text{supp}\, X|W = w$ and $\text{supp}\, Z|W = w$ for $w \in \mathcal{W} \equiv \text{supp}\, W$. Suppose that either Z.R.D or Z.R.C holds for each $w \in \mathcal{W}$.*

- ***Z.R.D.*** ***(Discrete)*** *$\mathcal{Z}_w$ is finite, $\mathcal{X}_w$ is bounded either from above or below (or both) and, except for finitely many $x \in \mathcal{X}_w$, there exist $\overline{z}_w(x), \underline{z}_w(x) \in \mathcal{Z}_w$ such that $1 > F_{X|WZ}(x \mid w, \overline{z}_w(x)) > F_{X|WZ}(x \mid w, \underline{z}_w(x)) > 0$.*[9]
- ***Z.R.C.*** ***(Continuous)*** *$Z|W = w$ is continuously distributed and for almost every $x \in \mathcal{X}_w$ there exists a $z'_w(x) \in \mathcal{Z}_w$ such that $F_{X|WZ}(x \mid w, \cdot)$ is differentiable at $z'_w(x)$ with $\nabla_z F_{X|WZ}(x \mid w, z'_w(x)) \neq 0$.*

The exogeneity condition of Z.FS.EX is a standard assumption. Full independence, rather than mean independence, is required because of the nonseparability of (1). The instrument only needs to be exogenous conditional on covariates. This can be an important distinction when considering the validity of an instrument that is not randomly assigned. Assumption Z.FS.SI has the same rank invariance interpretation as M.SI, only it concerns the effect of the instrument on the treatment. That is, $F_{X_z|W}(X_z \mid W) = F_{V|W}(V \mid W) = F_{X_{z'}|W}(X_{z'} \mid W)$ for any z, z', where $X_z \equiv g(W, z, V)$

[9] The bound(s) on $\mathcal{X}_w$ can vary with w.

is a counterfactual realization of X. The common ranking is $F_{V|W}(V \mid W)$, which can change with covariates. Conversely, if the instrument is exogenous and rank invariance holds, then a g and V satisfying Z.FS will always exist.

Proposition 1. *Let $\mathcal{Z} \equiv \operatorname{supp} Z$. Given M.C, Z.FS is satisfied if and only if there exists a collection of random variables $\{X_z\}_{z\in\mathcal{Z}}$ such that $X = \sum_{z\in\mathcal{Z}} \mathbb{1}[Z = z] X_z$ with $X_z|W = w$ continuously distributed for every $(w, z) \in \mathcal{WZ}$, $(X_z, U) \perp\!\!\!\perp Z|W$ for every $z \in \mathcal{Z}$ and $F_{X_z|W}(X_z \mid W) = F_{X_{z'}|W}(X_{z'} \mid W)$ for every $z, z' \in \mathcal{Z}$.*

In Section 4, I show how Z.FS can be justified in specific applications, either by specifying g and V directly or by appealing to the equivalent rank invariance interpretation.

Assumption Z.R is a relevance condition for the instrument.[10] Both the discrete and continuous versions of Z.R require Z to have an impact on X at almost every point in its support. This is a natural requirement for an instrument in a nonparametric model, because there is no functional form to use for extrapolation. It may only hold on some proper subset of the observed values of X and Z. For example, if X is income and Z has no effect on those with large incomes, then Z.R may still be satisfied in an analysis that considers only agents with low incomes.[11] If the researcher is willing to assume that $m^* \in \{m_\theta : \theta \in \Theta \subseteq \mathbb{R}^{d_\theta}\}$ is an element of a finite-dimensional parameter space, the analysis can be restricted to a determining set $\mathcal{XW}_\Theta$ that is rich enough that $m_\theta(x, w, Q_{U|W}(\cdot \mid w)) = m_{\theta^*}(x, w, Q_{U|W}(\cdot \mid w))$ for all $(x, w) \in \mathcal{XW}_\Theta$ implies $\theta = \theta^*$ for any $\theta, \theta^* \in \Theta$. Extrapolation outside of $\mathcal{XW}_\Theta$ is then possible, but it is based on nonparametric identification achieved inside $\mathcal{XW}_\Theta$. As Chesher (2005, 2007) shows, linear instrumental variables models can be misleading in this regard because correlations provide no information about the set of treatment intensities at which the instrument is relevant.

[10] This is often called a rank condition. I avoid that terminology because it could create confusion with rank invariance, which is unrelated.

[11] However, a first-stage relationship that satisfies Z.FS for all incomes may or may not still satisfy Z.FS.EX and/or Z.FS.SI when restricted to only low incomes. Truncating the support of X to satisfy Z.R necessitates reconsidering the appropriateness of Z.FS for the smaller support.

Assumption Z.R.D contrasts sharply with the existing literature on nonseparable instrumental variables models for continuous treatments, which requires a continuously distributed instrument.[12] This is important in practice. Many commonly used instruments assume only two values, such as the intent to treat in a randomized controlled experiment with partial compliance or an exogenous policy shift in a natural experiment. Assumption Z.R.C applies to continuous instruments, but does not require the familiar large support assumption, i.e. $\text{supp}\, Z = \mathbb{R}$, that Imbens and Newey (2009) use for point identification of the QTR.[13] Their model is the same as the one in this paper except, crucially, they do not assume M.SI. As discussed, this imposes rank invariance on the outcome equation. The trade-off between the identification result of Imbens and Newey (2009) and the one in this paper is between large support for the instrument and strict monotonicity in the outcome equation.

Another notable aspect of both Z.R.D and Z.R.C is that, except when $Z \in \{0, 1\}$ is binary, they do not generally require the common support assumption $\text{supp}(X, Z)|W = \text{supp}\, X|W \times \text{supp}\, Z|W$.[14] This significantly complicates the proof of identification in Section 3, but it is important in practice. For example, if different types of subsidies, Z, are offered to agents with different incomes, X, then the common support assumption does not hold.

It is possible to state Z.R.D and Z.R.C in terms of the first-stage equation in Z.FS. While this is perhaps more primitive, the equivalent conditions are less intuitive. The given statement of Z.R is attractive because it can be easily verified by examining the data. Moreover, Proposition 1 shows that an analyst who believes that rank invariance is an appropriate assumption does not need to specify a first-stage equation at all. On the other hand, examining some simple sufficient conditions for Z.R in terms of g and

[12] For example, see Chesher (2003), Florens et al. (2008) and Imbens and Newey (2009).

[13] Imbens and Newey (2009) also provide sharp set identification results when the support of Z is not large enough to obtain point identification.

[14] When Z is binary, common support corresponds to the intuitively appealing requirement that at every $X = x$ there are non-trivial treatment ($Z = 1$) and control ($Z = 0$) groups.

V helps to clarify just how weak this assumption is. For Z.R.C to hold in the case of no covariates, it suffices that $\nabla_z g(z, v)$ exists and is non-zero almost everywhere. Similarly, Z.R.D will hold if, e.g., $\operatorname{supp} V = (0, \infty)$ and there are $\overline{z}, \underline{z} \in \mathcal{Z}$ such that $g(\overline{z}, \cdot)$ and $g(\underline{z}, \cdot)$ have the same range with $g(\overline{z}, v) \neq g(\underline{z}, v)$ for almost every v.

3 Identification

The first-stage relationship described by Z.FS has an equivalent characterization that is helpful in the identification analysis.

Proposition 2. *If M.C holds then Z.FS is equivalent to*

Z.RE. (Conditional rank exogeneity) $(R, U) \perp\!\!\!\perp Z | W$ *where* $R \equiv F_{X|WZ}(X \mid W, Z)$ *is called the* conditional rank *of* $X|W$*.*

Let $\mathcal{M}$ denote the collection of all outcome functions that satisfy M.SI and are everywhere continuous in x and u on $\mathcal{XWU}$. Identification under Z.R.C will require differentiability as well. Let $\mathcal{M}^d \subseteq \mathcal{M}$ denote those outcome functions that are also everywhere differentiable in x and u on the interior of $\mathcal{XWU}$. For any $m \in \mathcal{M}$, let $U^m \equiv m^{-1}(X, W, Y)$, where m^{-1} is the inverse of m with respect to its last argument and define $\overline{m}(x, w, t) \equiv m(x, w, Q_{U^m|W}(t \mid w))$. The following theorem shows that if the model is correctly specified, i.e. (1) holds for some $m^* \in \mathcal{M}$, then $\overline{m}^*(x, w, t) = m^*(x, w, Q_{U|W}(t \mid w))$ is identified for every $(x, w) \in \mathcal{XW}$ and every $t \in [0, 1]$. This demonstrates identification of m^* up to a normalization on the scale of m^* or U, such as N.QR in Section 2 or N.S, N.H in Appendix A.

Theorem 1. *Suppose that* (1) *holds,* $m^* \in \mathcal{M}$ *and that M.C, Z.R.D and Z.FS are satisfied. Then for any* $m \in \mathcal{M}$*,* $(R, U^m) \perp\!\!\!\perp Z | W$ *if and only if* $\overline{m}(x, w, t) = \overline{m}^*(x, w, t)$ *for every* $(x, w) \in \mathcal{XW}$ *and every* $t \in [0, 1]$*. The same statement is true for* $m \in \mathcal{M}^d$ *if* $m^* \in \mathcal{M}^d$ *and Z.R.C holds instead of Z.R.D.*

Theorem 1 can be used to construct a minimum distance from independence estimator for m^*.[15] In Torgovitsky (2011a), I consider

$$\widehat{m} = \underset{m \in \mathcal{M}_N}{\arg\min} \int \left[\widehat{F}_{RU^m Z}(r,u,z) - \widehat{F}_{RU^m}(r,u)\,\widehat{F}_Z(z)\right]^2 d\mu(r,u,z), \tag{2}$$

where $\mathcal{M}_N \subseteq \mathcal{M}$ is a collection of normalized outcome functions, μ is a measure over the support of (R, U^m, Z) and covariates have been suppressed for simplicity. The integrand is composed of empirical distribution functions that are constructed from the pseudo-sample $(\widehat{R}_i, U_i^m, Z_i)$, where $\widehat{R}_i$ is an estimate of $F_{X|Z}(X_i \mid Z_i)$ and $U_i^m = m^{-1}(X_i, Y_i)$. Theorem 1 shows that in the absence of sampling error, $\widehat{m} = m^*$ is the unique minimizer of (2).

The proof of Theorem 1 is facilitated by some elementary copula theory. Sklar (1959) showed that the distribution function for any continuously distributed random vector can be uniquely expressed as a combination of its constituent marginal distributions and a joint distribution, called a copula function, that admits unit uniform marginals. His result is the following.

Theorem 2 (Sklar's Theorem). *Let $(X,U) \in \mathbb{R}^2$ be a random vector with joint distribution F_{XU}. There exists a probability distribution function C such that for all $(x,u) \in \overline{\mathbb{R}}^2$,*

$$F_{XU}(x,u) = C\left(F_X(x), F_U(u)\right).$$

The function C is called the copula function of (X,U) and has support $[0,1]^2$. If X and U are continuously distributed then C is unique. If Z is a random element such that $X|Z = z$ and $U|Z = z$ are continuously distributed for $z \in \operatorname{supp} Z$ then there exists a

[15]Econometric applications of minimum distance from independence estimation include Manski (1983), Brown and Wegkamp (2002) and Komunjer and Santos (2010).

unique conditional copula function $C(\cdot,\cdot;z)$ *such that*

$$F_{XU|Z}(x,u \mid z) = C\left(F_{X|Z}\left(x \mid z\right), F_{U|Z}(u \mid z); z\right)$$

for all $(x,u) \in \overline{\mathbb{R}}^2$.

Proof. See, e.g., Nelsen (2006) for the traditional result and Patton (2006) for the extension to the conditional case. *Q.E.D.*

Sklar's Theorem shows that the reason that two random vectors with the same marginal distributions can have different joint distributions is precisely because they can have different copulas. For a given set of marginals, the copula therefore fully characterizes the dependence structure of a random vector.

The analysis in Theorem 1 involves the copulas for the collection of random vectors $\{(X,U)|W,Z=z\}$ indexed by z. The next proposition shows that under Z.RE (or Z.FS, by Proposition 2), this collection is actually just a singleton.

Proposition 3. *Given M.C, Z.RE holds if and only if both*

Z.EX. ***(Marginal exogeneity)*** $U \perp\!\!\!\perp Z|W$.

Z.CI. ***(Copula invariance)*** *The copula function for* $(X,U)|(W,Z)=(w,z)$ *is equal to the copula function for* $(X,U)|(W,Z)=(w,z')$ *for every* $(w,z),(w,z') \in \mathcal{WZ}$.

This result means that the first-stage formulation of Z.FS is a nonparametric restriction on the dependence structure of $(X,U)|W,Z$. The loose interpretation is that whatever the underlying source of the dependence between X and U is, it is not affected by exogenous manipulations of the instrument. This is consistent with the rank invariance interpretation given in Section 2. In the first-stage specification of Z.FS, the source of the dependence between X and U is V. The random variable $F_{V|W}(V \mid W)$ corresponds to an agent's ranking for X. As noted before, Z.FS implies that this ranking is invariant to exogenous manipulations of Z.

Given the related literature on nonseparable models, it may seem surprising that Theorem 1 holds for a discrete instrument that satisfies Z.FS and Z.R.D, especially one that has only the two-point (binary) support $\{0, 1\}$. An intuitive requirement for identification of levels in a nonparametric model is that the assumptions are sufficient to provide a comparison of the exogenous effects of x_a and x_b on Y for any x_a, x_b. That this is the case here can be demonstrated through a chaining argument that uses Z.RE. For simplicity, assume that there are no covariates. The conditional rank always satisfies $R \perp\!\!\!\perp Z$,[16] so Z.RE is equivalent to the combination of $U \perp\!\!\!\perp Z$ and $U \perp\!\!\!\perp Z|R$. This can be interpreted as saying that changes in Z are unconditionally exogenous and remain exogenous after including the conditional rank of X in the information set. Assumptions M.C and Z.R.D imply that conditional on $R = r \in (0, 1)$, an exogenous shift of Z from z to z' corresponds to a unique shift between two distinct realizations of X, say x and x'. Thus, given the exclusion of Z from (1) and $U \perp\!\!\!\perp Z$, the difference in the distributions of $Y|(R, Z) = (r, z)$ and $Y|(R, Z) = (r, z')$ is entirely attributable to the associated exogenous shift from $m^*(x, \cdot)$ to $m^*(x', \cdot)$. This difference is observable because R is observable.

When Z is binary, only one such exogenous shift is possible for any given rank, r. However, M.C and Z.R.D allow shifts at different ranks to be chained together so that any two arbitrary points can be compared by repeatedly shifting Z. For example, to compare the direct effect of x_a to that of x_b, first find the r_a^0 such that $x_a = Q_{X|Z}\left(r_a^0 \,\middle|\, 0\right)$. Such an r_a^0 exists by M.C and Z.R.D. Now shift $Z = 0$ to $Z = 1$ while holding $R = r_a^0$. This represents an exogenous shift from x_a to $x_a^1 = Q_{X|Z}\left(r_a^0 \,\middle|\, 1\right)$. Assumption Z.R.D ensures that $x_a^1 \neq x_a$, so this comparison is not trivial. Next, find the r_a^1 such that $x_a^1 = Q_{X|Z}\left(r_a^1 \,\middle|\, 0\right)$. Repeat the process by finding an $x_a^2 = Q_{X|Z}\left(r_a^1 \,\middle|\, 1\right)$, which gives an exogenous shift from x_a^1 to x_a^2. Notice that now x_a has been exogenously compared with x_a^2 through their mutual comparisons with x_a^1. The proof of Theorem 1 shows that this sequence x_a, x_a^1, x_a^2 can be continued indefinitely

[16]This is because $\mathbb{P}\left[R \leq r \mid Z = z\right] = \mathbb{P}\left[X \leq Q_{X|Z}(r \mid z) \,\middle|\, Z = z\right] = r$ for any $r \in [0, 1]$, $z \in \operatorname{supp} Z$.

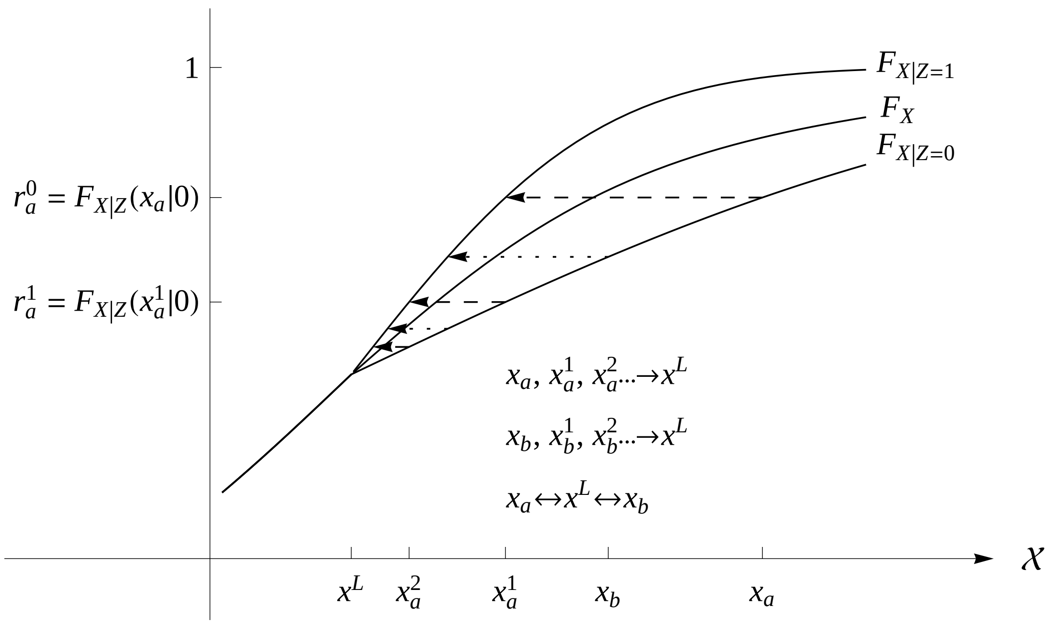

Figure 1: **Chaining together multiple exogenous comparisons when Z is binary.** The dashed arrows represent exogenous shifts starting from x_a while the dotted arrows represent exogenous shifts starting from x_b. The relevance condition, Z.R.D, is only satisfied here for an analysis with $\mathcal{X} = [x^L, \infty)$. Further extrapolating assumptions would be needed to identify $\overline{m}$ in a model including treatment levels $x < x^L$.

and that M.C allows x_a to be compared with a unique limiting point, x^L. Moreover, an analogous sequence started at x_b has the same limit. Thus x_a and x_b can be exogenously compared through their mutual comparisons with x^L. This is the intuition behind Theorem 1. Figure 1 depicts this argument graphically.

4 Examples

Example 1 (**A structural model of the returns to schooling**). [17] Suppose that Y is lifetime earnings, X is a measure of investment in schooling and W is a set of socioeconomic and family background controls. The analyst is interested in the effect of schooling investment in the educational production function $Y = m^*(X, W, U)$. The classic endogeneity problem in this situation is that $U \not\perp\!\!\!\perp X | W$ because U captures, among other things, latent traits such as ability, which are likely to be dependent with both education decisions and earnings.

Suppose that agents choose X by maximizing expected lifetime earnings net of costs,

$$X = g(W, Z, V) = \arg\max_x \mathbb{E}\left[m^*(x, W, U)|W, V\right] - c(x, W, Z),$$

where c is the educational cost function and V is a scalar signal of U that is observed by the agent. The instrument Z is an exogenous cost shifter that is excluded from the production function. For example, Card (1995) uses an indicator for residence in a county with a four year college as a cost shifter and argues that Z is exogenous, conditional on W, i.e. Z satisfies Z.FS.EX. As discussed in Imbens and Newey (2009), $g(w, z, \cdot)$ is strictly increasing under the following assumptions.

1. m^* is strictly increasing in x.

2. m^* is twice continuously differentiable.

[17] This example is due to Imbens and Newey (2009). I have modified it slightly to include covariates.

3. There are diminishing returns to schooling, $\nabla_x^2 m^* < 0$, and the returns to schooling increase in ability, $\nabla_{x,u}^2 m^* > 0$.
4. Costs increase in education at an increasing rate, $\nabla_x c, \nabla_x^2 c > 0$.
5. V and U are affiliated random variables, conditional on W.

Theorem 1 shows that m^* is identified as long as Z.R holds. ∎

Example 2 (**The returns to schooling in a natural experiment**). Duflo (2001) estimates the returns to schooling using an instrument derived from a natural experiment in Indonesia. The experiment was the result of a 1970's era government campaign to construct primary schools. In this example, Y is hourly wage, X is schooling obtained and W contains characteristics of the region of birth. The instrument, Z, is an interaction term between year of birth and the number of schools planned for construction in the agent's region of birth. Duflo shows that this instrument is relevant and argues convincingly that it is exogenous and can be excluded from (1). Her analysis suggests that Z.R.D and Z.FS.EX are reasonable assumptions. Assumption Z.FS.SI requires rank invariance in schooling. An agent who would have obtained a large amount of schooling, relative to his peers, had he been of an age and in a region where many schools were built, would also have obtained a relatively large amount of schooling if he were of an age and/or region in which few schools were built. In other words, Z.FS.SI assumes that agents possess an underlying proclivity for education that is not affected by their year of birth or the intensity of the school building program in their birth region. Under this assumption, Theorem 1 shows that a nonseparable outcome function that is continuous and satisfies M.SI is nonparametrically identified. ∎

Example 3 (**The effect of class size with a first-stage equation**). Let Y be a measure of schooling outcomes (e.g., standardized test scores), X be the average class size of a school and W be a set of observable controls such as school characteristics and socioeconomic variables. The unobservable U aggregates the litany of other factors involved in determining outcomes, including parental involvement and unobserved family background characteristics. The assumption that $U \perp\!\!\!\perp X$ is unlikely to hold, even conditional on W, because families that value education more highly are more likely to select into schools on the basis of the prevailing class size.

Hoxby (2000) uses an instrument that captures the exogenous fluctuations in the number of enrolled students caused by changes in the timing of births around the calendar year. Suppose that $\bar{s}(W, V)$ represents the number of students that would be enrolled if the timing of births were non-varying, where V is some random element that may be arbitrarily dependent with U. Letting $Z \approx 1 > 0$ represent proportional exogenous fluctuations in enrollment, the actual number of enrolled students is $s(W, Z, V) = Z\bar{s}(W, V)$. Similarly, suppose that $\bar{c}(W, V)$ denotes the number of classes that the school would maintain in a baseline year and let the actual number of classes be given by $c(W, Z, V) = d(W, Z)\bar{c}(W, V)$, where $d > 0$ and $d(W, 1) = 1$. Assuming that classes are split into equal sizes,

$$X = \frac{s(W, Z, V)}{c(W, Z, V)} = \frac{Z}{d(W, Z)} \frac{\bar{s}(W, V)}{\bar{c}(W, V)} \equiv h_1(W, Z)h_2(W, V) \equiv h_1(W, Z)\widetilde{V},$$

where $h_1 > 0$. Thus $X = g(W, Z, \widetilde{V}) = h_1(W, Z)\widetilde{V}$ is strictly increasing in $\widetilde{V}$, so Z.FS.SI is satisfied. Under the assumption that the fluctuations are indeed exogenous, i.e. $(V, U) \perp\!\!\!\perp Z|W$, then also $(\widetilde{V}, U) \perp\!\!\!\perp Z|W$, so Z.FS.EX holds. ■

Example 4 (The effect of class size with a traditional instrument). Consider the same inference problem as in Example 3. Feinstein and Symons (1999) use geographic indicator variables for instruments. Variation in these indicators corresponds to different local authorities (a unit of local government in England) which have different policies on class size. The authors cite work on the determinants of migration to argue that geographic location at the local authority level is exogenous to schooling outcomes after conditioning on measures of social class, parents' education and parental interest. That is, Z.FS.EX holds when W is a set of controls containing these variables.

To assess the plausibility of rank invariance, Z.FS.SI, consider a school that is located in local authority A. Suppose that this school has relatively small class size compared to other schools in A with the same socioeconomic makeup. Rank invariance implies that if this school were actually located in local authority B, then it would also have a small class size relative to comparable schools in B. In other words, whatever unobservable factor it is (say, pushy parents) that makes the school have a relatively small class size in local authority A is intrinsic to the school and is not due to the region. It is important to notice that the absolute class size of the school can be different from local authority A to local authority B. In fact, it must be for Z.R.D to hold. ■

5 Conclusion

In this paper I have shown that a general nonseparable outcome equation can be nonparametrically identified under a scalar heterogeneity restriction on the first-stage equation. An unusual aspect of the identification result is that it allows for both discrete and continuous instruments that do not have large support. Because discrete instruments are widely used in applications and credibly exogenous instruments of all types are rare, this is important in practice. In particular, this result applies to randomized controlled experiments with partial compliance. In that context, the implication is that easily interpretable, nonparametric assumptions about the dimension of heterogeneity enable one to extrapolate outside the context of the experiment, while still allowing for general unobserved heterogeneity.

A Normalizations

Matzkin (2003) showed that a functional normalization is needed to separate the scale of m^* from that of U in outcome equation (1) with M.C and M.SI. In particular, m^* is identified up to a strictly increasing transformation of its unobservable component. Normalization N.QR amounts to choosing this transformation to be $Q_{U|W}$ with $U|W \sim \text{Unif}[0,1]$. It is then easy to see that m^* is identified if $\overline{m}^*$ is: $\overline{m}^*(x,w,t) = m^*(x,w,Q_{U|W}(t \mid w)) = m^*(x,w,t)$. In this case, $m^*(x,w,u)$ has the quantile treatment effect interpretation discussed in Section 2. What follows are two other normalizations that were suggested by Matzkin (2003). See that work for a lucid discussion of their interpretations.

Proposition 4. *If $\overline{m}^*(x,w,t)$ is identified for every $(x,w) \in \mathcal{XW}$ and $t \in [0,1]$ then $m^*(x,w,u)$ is constructively identified for every $(x,w,u) \in \mathcal{XWU}$ under N.S or N.H.*

N.S. ***(Scale)*** *There is a known $\overline{x}$ such that for all $(\overline{x},w,u) \in \mathcal{XWU}$, $m^*(\overline{x},w,u) = u$. Then $m^*(x,w,u) = \overline{m}^*(x,w,(\overline{m}^*)^{-1}(\overline{x},w,u))$ for all $(x,w,u) \in \mathcal{XWU}$.*

N.H. ***(Homogeneity of degree one)*** *For a known $\overline{x}$, a known $\overline{u} > 0$, a known $\alpha > 0$ and all w such that $(\overline{x},w,\overline{u}) \in \mathcal{XWU}$, $m^*(\overline{x},w,\lambda\overline{u}) = \lambda\alpha$ for all $\lambda > 0$. In this case, $m^*(x,w,u) = \overline{m}^*\left(x,w,(\overline{m}^*)^{-1}(\overline{x},w,(u/\overline{u})\alpha)\right)$ for $(x,w,u) \in \mathcal{XWU}$.*

Proof. See Matzkin (2003).

B Proofs

Notation. To compress notation I write an event like $[X = x, W = w]$ as simply $[x,w]$.

Proof of Proposition 1. Suppose Z.FS holds and take $X_z \equiv g(W,z,V)$ for $z \in \mathcal{Z}$. Then $X = g(W,Z,V) = \sum_{z\in\mathcal{Z}} \mathbb{1}[Z = z]\, X_z$. For every $(w,z) \in \mathcal{WZ}$, $X_z|W = w$ is given by $g(w,z,V)|(W,Z) = (w,z)$, which is equal to $X|(W,Z) = (w,z)$, by Z.FS. This random variable is continuously distributed under M.C. By Z.FS.EX, $(X_z,U) = (g(W,z,V),U) \perp\!\!\!\perp Z|W$, and Z.FS.SI implies that $F_{X_z|W}(X_z \mid W) = F_{V|W}(V \mid W) = F_{X_{z'}|W}(X_{z'} \mid W)$ for all $z,z' \in \mathcal{Z}$, as pointed out in the text.

On the other hand, suppose that the stated counterfactual representation holds. Take $V \equiv F_{X_z|W}(X_z \mid W)$, which does not depend on $z \in \mathcal{Z}$ by the rank invariance assumption. Then $(V, U) \perp\!\!\!\perp Z|W$, because $(X_z, U) \perp\!\!\!\perp Z|W$, so Z.FS.EX holds. Take $g(w, z, \cdot) \equiv Q_{X_z|W}(\cdot \mid w)$ for $(w, z) \in \mathcal{WZ}$. Then $g(w, z, \cdot)$ is strictly increasing for every $(w, z) \in \mathcal{WZ}$ because $X_z|W = w$ is continuously distributed by assumption. Also, because $V = F_{X_z|W}(X_z \mid W)$ for every $z \in \mathcal{Z}$,

$$
\begin{aligned}
g(W, Z, V) &= \sum_{z \in \mathcal{Z}} \mathbb{1}[Z = z]\, g(W, z, V) \\
&= \sum_{z \in \mathcal{Z}} \mathbb{1}[Z = z]\, Q_{X_z|W}(V \mid W) \\
&= \sum_{z \in \mathcal{Z}} \mathbb{1}[Z = z]\, Q_{X_z|W}(F_{X_z|W}(X_z \mid W) \mid W) \\
&= \sum_{z \in \mathcal{Z}} \mathbb{1}[Z = z]\, X_z = X.
\end{aligned}
$$

Q.E.D.

Proof of Proposition 2. Suppose that a first-stage equation satisfying Z.FS exists. By Z.FS.SI and Z.FS.EX,

$$
Q_{X|WZ}(r \mid w, z) = g(w, z, Q_{V|WZ}(r \mid w, z)) = g(w, z, Q_{V|W}(r \mid w)) \tag{3}
$$

for any $r \in [0, 1]$ and any $(w, z) \in \mathcal{WZ}$. Then for any $r \in [0, 1]$, $u \in \mathbb{R}$ and $(w, z) \in \mathcal{WZ}$,

$$
\begin{aligned}
\mathbb{P}[R \leq r, U \leq u \mid w, z] &= \mathbb{P}\left[X \leq Q_{X|WZ}(r \mid w, z), U \leq u \,\middle|\, w, z\right] \\
&= \mathbb{P}\left[g(w, z, V) \leq g(w, z, Q_{V|W}(r \mid w)), U \leq u \,\middle|\, w, z\right] \\
&= \mathbb{P}\left[V \leq Q_{V|W}(r \mid w), U \leq u \,\middle|\, w, z\right] \\
&= \mathbb{P}\left[V \leq Q_{V|W}(r \mid w), U \leq u \,\middle|\, w\right]
\end{aligned} \tag{4}
$$

where the first equality is by M.C, the second is by (3), the third is because $g(w, z, \cdot)$ is strictly increasing by Z.FS.SI and the fourth is from Z.FS.EX. Because (w, z) is arbitrary, (4) shows that $\mathbb{P}[R \leq r, U \leq u \mid w, z]$ does not depend on z for any r, u and hence that $(R,U) \perp\!\!\!\perp Z|W$, which is Z.RE.

Conversely, suppose that Z.RE holds. Let $V \equiv R$ and take $g(w,z,\cdot) \equiv Q_{X|WZ}(\cdot|w,z)$ for $(w,z) \in \mathcal{WZ}$. Then $g(W,Z,V) = Q_{X|WZ}(R \mid W,Z) = X$. By M.C, $g(w,z,\cdot)$ is strictly increasing for $(w,z) \in \mathcal{WZ}$, so Z.FS.SI is satisfied. In addition, Z.FS.EX holds because $(V,U) \equiv (R,U) \perp\!\!\!\perp Z|W$ by hypothesis.

Q.E.D.

Proof of Proposition 3. Let $r \in [0,1]$, $u \in \mathbb{R}$ and $(w,z),(w,z') \in \mathcal{WZ}$. Note that Z.RE immediately implies Z.EX. Moreover, if Z.EX holds then by M.C and Sklar's Theorem there exists a unique conditional copula $C(\cdot,\cdot;(w,z))$ such that

$$\begin{aligned}
\mathbb{P}\left[R \le r, U \le u \mid w,z\right] &= \mathbb{P}\left[X \le Q_{X|WZ}(r \mid w,z), U \le u \,\middle|\, w,z\right] \\
&= C(F_{X|WZ}(Q_{X|WZ}(r \mid w,z) \mid w,z), F_{U|WZ}(u \mid w,z);(w,z)) \\
&= C(r, F_{U|WZ}(u \mid w,z);(w,z)) \\
&= C(r, F_{U|W}(u \mid w);(w,z)). \qquad (5)
\end{aligned}$$

By assumption, $(R,U) \perp\!\!\!\perp Z|W$, so $\mathbb{P}\left[R \le r, U \le u \mid w,z\right] = \mathbb{P}\left[R \le r, U \le u \mid w,z'\right]$. Repeating the derivation in (5) with the conditional copula for $(W,Z) = (w,z')$ shows that Z.RE implies $C(\cdot, F_{U|W}(\cdot \mid w);(w,z)) = C(\cdot, F_{U|W}(\cdot \mid w);(w,z'))$. Given M.C, this is equivalent to $C(\cdot,\cdot;(w,z)) = C(\cdot,\cdot;(w,z'))$, which is Z.CI. Conversely, if Z.EX and Z.CI hold then (5) shows that Z.RE is satisfied. *Q.E.D.*

Proof of Theorem 1. For ease of exposition, I suppress the conditioning on covariates W in the proof.

Recall the definition of $U^m \equiv m^{-1}(X,Y)$, where the inverse is in the second argument of $m \in \mathcal{M}$. If $\overline{m} = \overline{m}^*$ then by M.C, $m = m^*$. By (1), Z.FS and Proposition 2, $(R,U^m) = (R,U)$ and Z are independent.

Now suppose that $(R,U^m) \perp\!\!\!\perp Z$ for some $m \in \mathcal{M}$. The continuity and strict monotonicity of $m, m^* \in \mathcal{M}$ imply that the random vector $(X,U^m) = (X, m^{-1}(X,Y)) =$

$\big(X, m^{-1}[X, m^*(X,U)]\big)$ satisfies M.C if (X,U) does.[18] Proposition 3 shows that Z.EX and Z.CI are satisfied for both (X, U^m, Z) and (X, U, Z). Let $C(\cdot,\cdot)$ and $C^*(\cdot,\cdot)$ denote the common (for all realizations of Z) conditional copula functions for $(X, U^m)|Z$ and $(X,U)|Z$, which are unique by Sklar's Theorem. Let $\mathcal{XZ} \equiv \text{supp}(X,Z)$ and $\mathcal{XZ}^\circ \equiv \{(x,z) \in \mathcal{XZ} : F_{X|Z}(x \mid z) \in (0,1)\}$. Then for all $u \in \mathbb{R}$ and $(x,z) \in \mathcal{XZ}^\circ$, the common copula of $(X,U)|Z$ satisfies

$$\begin{aligned} F_{U|XZ}(u \mid x, z) &= \frac{\nabla_x F_{XU|Z}(x, u \mid z)}{f_{X|Z}(x \mid z)} \\ &= \frac{\nabla_x C^*(F_{X|Z}(x \mid z), F_{U|Z}(u \mid z))}{f_{X|Z}(x \mid z)} = C_r^*(F_{X|Z}(x \mid z), F_U(u)), \end{aligned} \tag{6}$$

where M.C guarantees that $C_r^*(r,t) \equiv \nabla_r C^*(r,t)$ exists, is continuous in both arguments and is strictly increasing in t for every $r \in (0,1)$.[19] The analogous relationship holds for C,

$$F_{U^m|XZ}(u \mid x, z) = C_r(F_{X|Z}(x \mid z), F_{U^m}(u)). \tag{7}$$

[18]The connectedness of the support of $(X, U^m)|Z$ follows from the continuity of m and m^*.

[19]This can be seen by rewriting the conditional version of Sklar's Theorem (under Z.CI) as $C^*(r,t) = F_{XU|Z}\big(Q_{X|Z}(r \mid z), Q_{U|Z}(t \mid z) \bigm| z\big)$. The assumption that $(X,U)|Z = z$ is absolutely continuous with connected support ensures that C^* is almost everywhere differentiable. M.C strengthens this slightly to everywhere differentiability, which is convenient in the following analysis, but not necessary. Strict monotonicity of $C_r^*(r,\cdot)$ for $r \in (0,1)$ also follows from M.C.

It follows that for any $(x, z) \in \mathcal{XZ}^\circ$ and any $t \in [0, 1]$,

$$
\begin{aligned}
C_r(F_{X|Z}(x \mid z), t) &= \mathbb{P}\left[U^m \le Q_{U^m}(t) \mid x, z\right] \\
&= \mathbb{P}\left[m^{-1}(x, Y) \le Q_{U^m}(t) \,\middle|\, x, z\right] \\
&= \mathbb{P}\left[Y \le \overline{m}(x, t) \mid x, z\right] \\
&= \mathbb{P}\left[m^*(x, U) \le \overline{m}(x, t) \mid x, z\right] \\
&= \mathbb{P}\left[F_U(U) \le (\overline{m}^*)^{-1}(x, \overline{m}(x, t)) \,\middle|\, x, z\right] \\
&= C_r^*\left(F_{X|Z}(x \mid z), (\overline{m}^*)^{-1}\left[x, \overline{m}(x, t)\right]\right). \qquad (8)
\end{aligned}
$$

Here the first equality uses (7), the second uses the definition of U^m with $m \in \mathcal{M}$, the third uses the definition of $\overline{m}$, the fourth is by (1), the fifth uses $m^* \in \mathcal{M}$ with the definition of $(\overline{m}^*)^{-1}(x, y) = F_U((m^*)^{-1}(x, y))$ and the last equality is by (6).

Inverting (8) gives

$$
(\overline{m}^*)^{-1}\left[x, \overline{m}(x, t)\right] = (C_r^*)^{-1}\left(F_{X|Z}(x \mid z), C_r\left[F_{X|Z}(x \mid z), t\right]\right) \equiv I(x, t), \qquad (9)
$$

which holds for all $(x, z) \in \mathcal{XZ}^\circ$ and all $t \in [0, 1]$. Equality (9) shows that $I(x, t)$ is everywhere continuous in both arguments if $m, m^* \in \mathcal{M}$ and everywhere differentiable in both arguments if $m, m^* \in \mathcal{M}^d$.[20] It also shows that $I(x, t)$ is not a function of $z \in \mathcal{Z} \equiv \operatorname{supp} Z$. The rest of the proof uses this observation and the force of Z.R to demonstrate that in fact $I(x, t) = t$ for almost every (a.e.) $x \in \mathcal{X} \equiv \operatorname{supp} X$. By inverting (9), one then obtains $\overline{m}^*(x, t) = \overline{m}(x, t)$ for a.e. $x \in \mathcal{X}$ and every $t \in [0, 1]$. This shows identification everywhere because $\overline{m}^*, \overline{m}$ are everywhere continuous.

The proof that $I(x, t) = t$ is composed of two steps. First, I show that $I(x, t) = J(t)$ is not a function of x. This step depends on whether Z.R.C or Z.R.D is maintained. Second, I show that $J(t) = t$.

[20] Continuity could easily be deduced from the preceding discussion as well. Note that $\overline{m}$ inherits the differentiability of m because of the continuity and connectedness in M.C which, as mentioned, holds for (X, U^m). That $(\overline{m}^*)^{-1}(x, \cdot)$ is differentiable also depends on the strict monotonicity of $\overline{m}^*(x, \cdot)$.

Step 1 under Z.R.C: Let $D^*(r,s) \equiv (C_r^*)^{-1}(r,s)$. As noted, if $m, m^* \in \mathcal{M}^d$ then $I(x,t)$ is differentiable. If $F_{X|Z}(x \mid \cdot)$ is differentiable as well then $I(x,t)$ can be differentiated with respect to z to give

$$\begin{aligned}
0 &= \Big[\nabla_r D^*\left(F_{X|Z}(x \mid z), C_r\left[F_{X|Z}(x \mid z), t\right]\right) \\
&\quad + \nabla_t D^*\left(F_{X|Z}(x \mid z), C_r\left[F_{X|Z}(x \mid z), t\right]\right)\nabla_r C_r(F_{X|Z}(x \mid z), t)\Big]\nabla_z F_{X|Z}(x \mid z) \\
&\equiv A(x,z,t)\nabla_z F_{X|Z}(x \mid z),
\end{aligned}$$

which holds for any $(x,z) \in \mathcal{XZ}^\circ$ and $t \in (0,1)$. By Z.R.C, for a.e. $x \in \mathcal{X}$, there exists a $z'(x) \in \mathcal{Z}$ with $(x, z'(x)) \in \mathcal{XZ}^\circ$ for which $\nabla_z F_{X|Z}(x \,|\, z'(x))$ exists and is non-zero.[21] This implies $A(x, z'(x), t) = 0$. On the other hand, differentiating (9) with respect to x gives

$$\nabla_x I(x,t) = A(x,z,t) f_{X|Z}(x \mid z) = A(x, z'(x), t) f_{X|Z}(x \mid z'(x)) = 0,$$

where the second equality is because $I(x,t)$ is does not depend on z. This shows that $I(x,t) = J(t)$ is not a function of x a.e. on $\mathcal{X}$.

Step 1 under Z.R.D: Let $\mathcal{Z}(x) \equiv \{z \in \mathcal{Z} : F_{X|Z}(x \mid z) \in (0,1)\}$ and

$$\mathcal{X}^L \equiv \{x \in \mathcal{X} : F_{X|Z}(x \mid \overline{z}) = F_{X|Z}(x \mid \underline{z}) \;\; \forall \overline{z}, \underline{z} \in \mathcal{Z}(x)\} \cup \{x \in \mathcal{X} : |\mathcal{Z}(x)| < 2\}.$$

These are the elements of $\mathcal{X}$ at which the condition in Z.R.D does not hold, either because the instrument is locally irrelevant (the first set) or because there are fewer than two non-trivial comparison groups (the second). The statement of Z.R.D is that $\mathcal{X}^L$ is finite. Also from Z.R.D, $\mathcal{X}$ is bounded below or above, or both. Assume that $\mathcal{X}$ is bounded below—the argument when $\mathcal{X}$ is bounded above is analogous. Let $\{\xi_k\}_{k=0}^K$ be the unique elements of $\mathcal{X}^L \cup \{Q_X(1)\}$ ordered to be increasing, where $Q_X(1)$ may

[21] Note that the set of $x \in \mathcal{X}$ for which there exists a $z'(x) \in \mathcal{Z}$ such that $\nabla_z F_{X|Z}(x \mid z'(x)) \neq 0$ but yet $F_{X|Z}(x \mid z'(x)) = 0$ or 1 is negligible by M.C.

be $+\infty$. The first element is $\xi_0 = Q_X(0) > -\infty$, because $\mathcal{X}$ is bounded below and $\mathcal{Z}(Q_X(0))$ is the empty set. Let $\mathcal{X}_k \equiv [\xi_k, \xi_{k+1}]$ for $0 \leq k \leq K-1$. Then $\mathcal{X} = \bigcup_{k=0}^{K-1} \mathcal{X}_k$ and $\mathcal{X}_k \subseteq \mathcal{X}$ for every k because $\mathcal{X}$ is connected. Let $\mathcal{X}^\circ$ denote the interior of $\mathcal{X}$ and $\mathcal{X}_k^\circ = (\xi_k, \xi_{k+1})$ the interior of $\mathcal{X}_k$. The strategy is to show that $I(x,t)$ is not a function of x on $\mathcal{X}_0$ and then extend this to $\mathcal{X}_1, \mathcal{X}_2, \ldots$ and hence to $\mathcal{X}$, by induction.

Define the mappings

$$F^* : \mathcal{X}^\circ \to (0,1) : F^*(x) = \min_{z \in \mathcal{Z}(x)} F_{X|Z}(x \mid z)$$

$$\text{and} \qquad \pi : \mathcal{X}^\circ \to \mathcal{X}^\circ : \pi(x) = \min_{z \in \mathcal{Z}} Q_{X|Z}(F^*(x) \mid z),$$

where the minima are attained because $\mathcal{Z}(x)$ is non-empty for $x \in \mathcal{X}^\circ$ and $\mathcal{Z}$ has finitely many elements by Z.R.D. The mapping π is weakly decreasing because for any $z \in \mathcal{Z}(x)$,

$$\pi(x) \leq Q_{X|Z}(F^*(x) \mid z) \leq Q_{X|Z}(F_{X|Z}(x \mid z) \mid z) = x.$$

However, F^* and hence π may be discontinuous. Figure 2 illustrates these definitions.

Now take any $x^0 \in \mathcal{X}_0^\circ$ and consider the recursive decreasing sequence $\{x^n\}_{n=0}^\infty$ formed as $x^{n+1} = \pi(x^n)$ for $n \geq 0$.[22] This sequence has a limit $x^L \in \mathcal{X}_0$. In fact, $x^L \in \mathcal{X}^L$. Otherwise, there would exist $\overline{z}, \underline{z} \in \mathcal{Z}(x^L)$ such that $F_{X|Z}(x^L \mid \overline{z}) > F_{X|Z}(x^L \mid \underline{z})$.

[22] For this construction to make sense, the range of π should be $\mathcal{X}^\circ$, to match its domain. This is the case because $F^*(x) \in (0,1)$ by definition of $\mathcal{Z}(x)$.

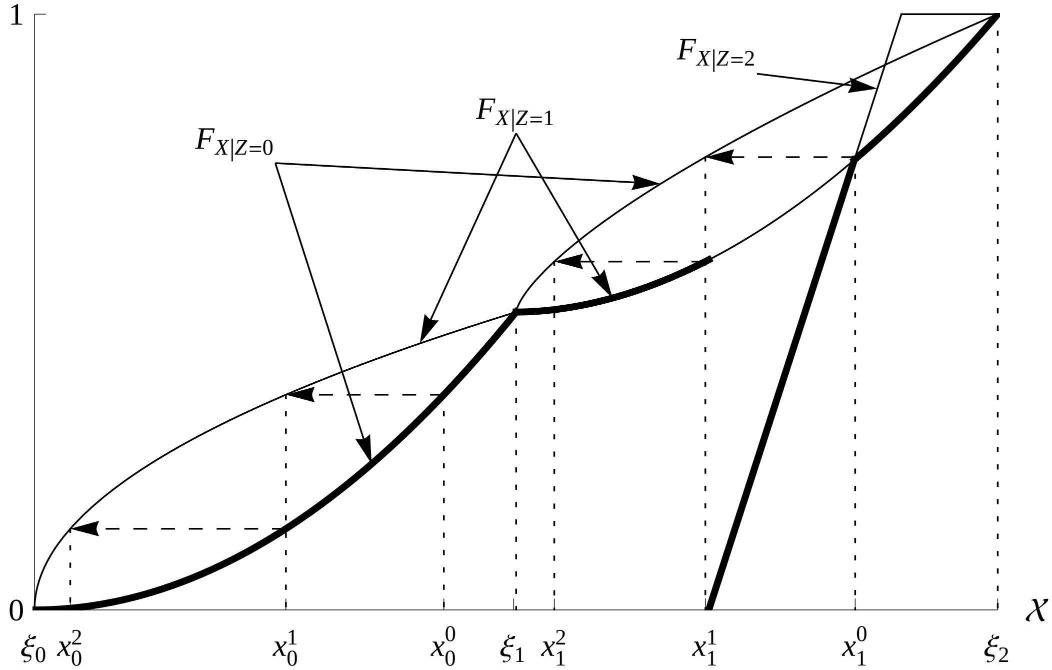

Figure 2: **Definitions for Step 1 under Z.R.D:** Here $\mathcal{Z} = \{0, 1, 2\}$ and the associated conditional distribution functions are as indicated. Note that the instrument is locally irrelevant at ξ_1 (where $F_{X|Z}(\cdot \mid 0)$ and $F_{X|Z}(\cdot \mid 1)$ cross) and that the support of $X|Z = 2$ is a proper subset of $\mathcal{X}$. Both of these features are allowed for under Z.R.D and can be important in practice. The bold line indicates F^*. The mapping π is indicated by the horizontal dashed arrows with $x_M^2 = \pi(x_M^1) = \pi(\pi(x_M^0))$ for $M = 0, 1$. In this diagram, $\mathcal{X}^L = \{\xi_0, \xi_1\}$, $\xi_2 = Q_X(1)$ is finite (although this is not required), $\mathcal{X}_0 = [\xi_0, \xi_1]$ and $\mathcal{X}_1 = [\xi_1, \xi_2]$. The sequence $\{x_1^n\}_{n=0}^{\infty}$ is converging to ξ_1. If x_1^0 were a bit larger, the sequence would instead converge to ξ_0 because of the discontinuity in F^*.

By M.C, $\underline{z} \in \mathcal{Z}(x^n)$ for sufficiently large n as well, so this would imply

$$
\begin{aligned}
F_{X|Z}(x^L \mid \overline{z}) = F_{X|Z}(\lim_{n\to\infty} x^{n+1} \mid \overline{z}) &= F_{X|Z}(\lim_{n\to\infty} \pi(x^n) \mid \overline{z}) \\
&= \lim_{n\to\infty} F_{X|Z}(\pi(x^n) \mid \overline{z}) \\
&= \lim_{n\to\infty} F_{X|Z}\left(\min_{z\in\mathcal{Z}} Q_{X|Z}(F^*(x^n) \mid z) \,\middle|\, \overline{z} \right) \\
&\leq \lim_{n\to\infty} F_{X|Z}\left(Q_{X|Z}(F^*(x^n) \mid \overline{z}) \,\middle|\, \overline{z} \right) \qquad (10) \\
&= \lim_{n\to\infty} F^*(x^n) \leq \lim_{n\to\infty} F_{X|Z}(x^n \mid \underline{z}) = F_{X|Z}(x^L \mid \underline{z}),
\end{aligned}
$$

which is a contradiction since $F_{X|Z}(x^L \mid \overline{z}) > F_{X|Z}(x^L \mid \underline{z})$. Because $\mathcal{X}_0 \cap \mathcal{X}^L = \{\xi_0\}$, it must be the case that $x^L = \xi_0$.

Return to (9) and recall that $I(x,t)$ does not depend on z. For $x \in \mathcal{X}^\circ$, let $z_F^*(x) \in \arg\min_{z\in\mathcal{Z}(x)} F_{X|Z}(x \mid z)$ and $z_Q^*(x) \in \arg\min_{z\in\mathcal{Z}} Q_{X|Z}(F^*(x) \mid z)$. Then $(\pi(x), z_Q^*(x))$, $(x, z_F^*(x)) \in \mathcal{XZ}^\circ$, because

$$
F_{X|Z}(\pi(x) \mid z_Q^*(x)) = F^*(x) = F_{X|Z}(x \mid z_F^*(x)) \in (0,1). \qquad (11)
$$

From (9) and (11) it follows that

$$
\begin{aligned}
I(\pi(x), t) &= (C_r^*)^{-1} \left(F_{X|Z}(\pi(x) \mid z_Q^*(x)), C_r \left[F_{X|Z}(\pi(x) \mid z_Q^*(x)), t \right] \right) \\
&= (C_r^*)^{-1} \left(F^*(x), C_r \left[F^*(x), t \right] \right) \\
&= (C_r^*)^{-1} \left(F_{X|Z}(x \mid z_F^*(x)), C_r \left[F_{X|Z}(x \mid z_F^*(x)), t \right] \right) = I(x,t), \qquad (12)
\end{aligned}
$$

where the first and last equalities hold because $I(x,t)$ does not depend on z. Equation (12) shows that $I(x^n, t) = I(x^0, t)$ for all $n \geq 0$ and all $t \in [0,1]$. As mentioned, $I(x,t)$ is continuous in x and $\mathcal{X}$ is bounded below by $\xi_0 > -\infty$, so $I(\xi_0, t) \equiv \lim_{x \searrow \xi_0} I(x,t)$ exists. It follows that $I(x^0, t) = \lim_{n\to\infty} I(x^n, t) = I(\lim_{n\to\infty} x^n, t) = I(x^L, t) = I(\xi_0, t) = J(t)$. Because $x^0 \in \mathcal{X}_0^\circ$ was arbitrary, this shows that $I(x,t) = J(t)$ for every

$x \in \mathcal{X}_0^\circ$ and every $t \in [0,1]$. By continuity, conclude that $I(x,t) = J(t)$ on the closure, $\mathcal{X}_0$, as well.

Now proceed inductively. Suppose that $I(x,t) = J(t)$ for all $x \in \bigcup_{k=0}^{M-1} \mathcal{X}_k$ with $1 \leq M \leq K-1$ and take an arbitrary $x_M^0 \in \mathcal{X}_M^\circ$. Construct the same sequence $\{x_M^n\}_{n=0}^\infty$ with $x_M^{n+1} = \pi(x_M^n)$ for $n \geq 0$. Then $\{x_M^n\}_{n=0}^\infty$ is a decreasing sequence and by the argument in (10), it has a limit point $x_M^L \in \{x : x \leq x_M^0\} \cap \mathcal{X}^L = \{\xi_k\}_{k=0}^M \subset \bigcup_{k=0}^{M-1} \mathcal{X}_k$. Hence $I(x_M^0, t) = I(x_M^L, t) = J(t)$. Because $x_M^0 \in \mathcal{X}_M^\circ$ was arbitrary, conclude that $I(x,t) = J(t)$ for all $x \in \mathcal{X}_M^\circ$ and hence for all $x \in \mathcal{X}_M$ by continuity. By induction, $I(x,t) = J(t)$ for any $x \in \mathcal{X}_M$ and every M. Therefore $I(x,t) = J(t)$ is not a function of x on $\mathcal{X} = \bigcup_{k=0}^{K-1} \mathcal{X}_k$.

Step 2: Copulas have unit-uniform marginal distributions, so for any $z \in \mathcal{Z}$, a change of variables provides

$$t = C(1,t) = \int_0^1 C_r(r,t)\,dr = \int_{-\infty}^{\infty} C_r(F_{X|Z}(x \mid z), t) f_{X|Z}(x \mid z)\,dx. \tag{13}$$

From (9) and the finding in Step 1 that $I(x,t) = J(t)$ does not depend on $x \in \mathcal{X}$ a.e., the right hand side of (13) is equal to

$$\int_{-\infty}^{\infty} C_r^*(F_{X|Z}(x \mid z), J(t)) f_{X|Z}(x \mid z)\,dx = \int_0^1 C_r^*(r, J(t))\,dr = C^*(1, J(t)) = J(t).$$

As noted in the paragraph following (9), $I(x,t) = J(t) = t$ for a.e. $x \in \mathcal{X}$ and every $t \in [0,1]$ implies that $\overline{m}(x,t) = \overline{m}^*(x,t)$ for every $x \in \mathcal{X}$ and every $t \in [0,1]$. *Q.E.D.*

References

Abadie, A., J. Angrist, and G. Imbens (2002): “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings,” *Econometrica*, 70, 91–117. 5

Altonji, J. G. and R. L. Matzkin (2005): “Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors,” *Econometrica*, 73, 1053–1102. 2

Athey, S. and G. W. Imbens (2006): “Identification and Inference in Nonlinear Difference-in-Differences Models,” *Econometrica*, 74, 431–497. 2

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2006): "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *The American Economic Review*, 96, 988–1012. 5

Blundell, R. and J. L. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, Cambridge: Cambridge University Press, vol. II, 312–357. 6

Brown, D. J. and M. H. Wegkamp (2002): "Weighted Minimum Mean-Square Distance from Independence Estimation," *Econometrica*, 70, 2035–2051. 11

Card, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. by L. N. Christofides, K. E. Grant, and R. Swidinsky, Toronto: University of Toronto Press, 201–222. 15

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2011): "Local Identification of Nonparametric and Semiparametric Models," Working paper, Yale University. 3

Chernozhukov, V., I. Fernández-Val, and B. Melly (2009): "Inference on counterfactual distributions," Working paper CWP09/09, cemmap. 5

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261. 3, 6

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007): "Instrumental variable estimation of nonseparable models," *Journal of Econometrics*, 139, 4–14. 3

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441. 2, 9

——— (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 1525–1550. 8

——— (2007): "Instrumental values," *Journal of Econometrics*, 139, 15–34. 6, 8

Djebbari, H. and J. Smith (2008): "Heterogeneous impacts in PROGRESA," *Journal of Econometrics*, 145, 64–80. 5

Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *The Annals of Statistics*, 2, 267–277. 6

Duflo, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *The American Economic Review*, 91, 795–813. 16

Evdokimov, K. (2010): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," Working paper, Princeton University. 2

Feinstein, L. and J. Symons (1999): "Attainment in secondary school," *Oxford Economic Papers*, 51, 300–321. 17

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259–276. 5

Florens, J. P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008): "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 76, 1191–1206. 3, 9

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161. 3

——— (1990): "Varieties of Selection Bias," *The American Economic Review*, 80, 313–318. 3

——— (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *The Journal of Political Economy*, 109, 673–748. 2

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, 487–535. 6

Heckman, J. J., S. Urzua, and E. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88, 389–432. 2

Heckman, J. J. and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738. 2

Hoderlein, S. and E. Mammen (2007): "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica*, 75, 1513–1518. 6

Hoxby, C. M. (2000): "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *The Quarterly Journal of Economics*, 115, 1239–1285. 17

Imbens, G. W. and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512. 3, 6, 9, 15

Komunjer, I. and A. Santos (2010): "Semi-parametric estimation of non-separable models: a minimum distance from independence approach," *Econometrics Journal*, 13, S28–S55. 11

Manski, C. F. (1983): "Closest Empirical Distribution Estimation," *Econometrica*, 51, 305–319. 11

Matzkin, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375. 5, 18, 19

Nelsen, R. (2006): *An introduction to copulas*, New York: Springer. 12

Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578. 3

Patton, A. J. (2006): "Modelling Asymmetric Exchange Rate Dependence," *International Economic Review*, 47, 527–556. 12

Sklar, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8, 1. 11

Su, L., S. Hoderlein, and H. White (2010): "Testing Monotonicity in Unobservables with Panel Data," Working paper, Singapore Management University. 6

Torgovitsky, A. (2011a): “Minimum Distance from Independence Estimation of a Nonseparable Instrumental Variables Model,” Working paper, Yale University. 4, 11

——— (2011b): “Relevance Conditions for Nonseparable Instrumental Variables Models,” Working paper, Yale University. 4