Nonparametric Evidence of Nonlinear Effects Using Instrumental Variables

Nicholas Y. Li and Benjamin Williams*
George Washington University

January 31, 2025

Preliminary and Incomplete. Do not cite without permission.

Abstract

Economic decisions take place at the margin. When treatment is binary, instrumental variables recover causal effects for "compliers" whose treatment decisions are marginal to the instrument (Imbens and Angrist, 1994). When treatment is a non-binary dosage, instruments must shift the same agents' treatment (e.g., price) more than once to show evidence of nonlinear responses (e.g., quantity demanded). If there are no "multi-marginal units," then any nonlinearity can be rationalized by the data, including linear structural effects. However, the existence of multi-marginal units is not sufficient to show that structural effects are nonlinear, even if structural effects are assumed to be monotone (e.g., demand slopes downward).

This paper provides (1) tests for the non-existence of multi-marginal units; (2) tests for monotone structural effects; (3) bounds for distributions of potential outcomes under monotonicity; and (4) tests for convex structural effects under monotonicity. Each contribution highlights the importance of instrument relevance in detecting structural nonlinearities, at odds with the practice of using narrow comparisons to support the exclusion restriction.

*Nicholas Li: nicholas.li@gwu.edu. Benjamin Williams: bdwilliams@gwu.edu. We thank Perry Sharaf, Mark Behringer, and Vibha Mital for excellent research assistance.

1 Introduction

[S]eparation of the observer from the phenomenon to be observed is no longer possible.

– Werner Heisenberg

I don't understand why economists are so obsessed with linear models.

– Anonymous Physicist

When direct inference of the causal effect of a treatment X on an outcome Y is infeasible, unethical, or unpersuasive, analysts often use an instrumental variable Z to change the dosage of X, holding other determinants ε constant. They often operationalize inference with instruments by specifying structural relationships between Y and X that are linear-in-parameters. On the one hand, the workhorse linear model is transparent—the eponymous two-stage least squares (2SLS) estimator is a ratio of two reduced-form slopes—and convenient—the parameter-of-interest directly measures the marginal effect of treatment. Despite its misspecification-robust interpretation as a weighted average derivative, the estimates give no indication of whether, for example, increasing prices yields the same demand effect regardless of the baseline price. In other scenarios, conclusions hinge specifically on structural nonlinearities (Bhattacharya, 2024).

Thus, analysts try to capture nonlinear responses in two ways: moment condition estimators (including 2SLS) (e.g., Newey and Powell, 2003; Horowitz, 2011) and control function estimators (Imbens and Newey, 2009). While they accommodate more flexibility in modeling average responses, both approaches place restrictions on effect heterogeneity, an increasingly prominent explanation for economic phenomena and important component of economic modeling. The former approach assumes no effect heterogeneity (Hahn and Ridder, 2011), and the latter imposes rank restrictions that do not permit Roy selection on gains (Kasy, 2011), for example. Empiricists are left with little guidance for documenting reduced-form evidence supporting or rejecting the presence of nonlinear effects under more minimal assumptions.

This paper argues that quantifying nonlinear effects without restrictions on heterogeneity is inherently fraught. Because unrestricted selection into higher dosages can rationalize many patterns in data, ruling out linear (or convex or concave or other nonlinear) effects requires highly predictive instruments that explain much of the overall variation in the data. However, (1) such instruments are at odds with the practice of making narrow comparisons to support the exclusion restriction assumption, and (2) if instruments explain much of the variation in treatment, then treatment cannot be "too endogenous," undermining the need for exogenous instruments. The paper argues this in two parts.

The first part generalizes "compliers" to more general treatment settings. Naturally, relevant instruments must shift agents' treatment: agents must be marginal to the instrument. In turn, to provide evidence that causal effects change with higher dosages, instruments must shift the *same* agents' treatment to multiple values. We call agents that are marginal at multiple values of the instrument "multi-marginal." However, if an agent is marginal at one value of the instrument, they may be inframarginal at other values. Therefore, instrument relevance is not sufficient to guarantee that any multi-marginal agents exist. We provide a test for the non-existence of multi-marginal agents. If there are no multi-marginal agents, then *any nonlinearity*, including linear effects, can be rationalized by the reduced-form relationships and heterogeneous selection into dosage.

The second part of the paper analyzes what instrumental variable analysis can say about structural effects' direction and convexity. Specifically, we provide tests for whether (1) structural effects are stochastically increasing (decreasing) and (2) whether they are stochastically increasing (decreasing) and convex (concave). In contrast to the test for stochastic direction, one may fail to reject both stochastic convexity and concavity. In other words, either interpretation is consistent with the data.

The starting point of our analysis is a non-separable, triangular model of potential outcomes that satisfies minimal exclusion and monotonicity assumptions, mirroring Imbens and Angrist (1994) and Angrist and Imbens (1995). We first analyze selection into dosage for general treatments. We partition the population into principle strata of inframarginal and (single- and multi-)marginal agents (Frangakis and Rubin, 2002). Unlike the case of binary treatment, the size of the marginal population (i.e., the analogous "complier" share) is not point identified. However, inframarginal agents' potential outcomes of X and Y do not change (Kitagawa, 2015), allowing us to place sharp bounds on agents who are inframarginal between different instrument values. Using this, we provide a test for a null hypothesis that there are no multi-marginal agents over all values of the instrument. In many empirical settings where instruments may be highly significant but explain little of the overall variation (i.e., large first-stage F-stat but low first-stage partial R^2), analysts may not be able to reject the null of no multi-marginals.

In other settings, analysts may have powerful instruments that reject the non-existence of multi-marginals. However, agents whose behavior is only shifted twice gives no information about higher-order nonlinearities. We can show that unless the instrument is perfectly predictive, there is always a point where higher-order nonlinearities are not identified.

Nevertheless, even if treatment is exogenous, many empiricists are not interested in higher order nonlinearities. Analysts approximate the conditional expectation function with OLS regression models and use quadratic terms to test for structural convexity. Since estimating such a model with instruments and 2SLS restricts effect heterogeneity, we ask what can be learned about structural convexity from instruments.

In the second part of the analysis, we study how instruments can test for stochastic direction and convexity of potential outcomes. Stochastic order (and convexity) are desirable properties because they are equivalent to qualitative conclusions being robust to increasing (and convex) transformations of the outcome variable. This part has three key elements. First, we provide necessary and sufficient conditions for stochastic order. If potential outcomes are stochastically ordered, then they can be rationalized by a model where effects are monotonic (e.g., demand slopes downward for everyone). Second, we derive sharp bounds on the marginal distributions of potential outcomes under monotonic effects.[1] Third, we provide necessary and sufficient conditions for potential outcomes to be stochastically increasing (decreasing) and convex (or concave). If the bounds in the marginal distributions of potential outcomes distributions are wide, then analysts will likely fail to reject both convexity and concavity. Finally, each of the three elements involve finding maximal lower sets, an open question in probability and qualitative multidimensional welfare analyses. We show that in two dimensions, this is equivalent to a very simple variational calculus problem, which is straightforward to compute.

Altogether, both sets of results imply that credibly identifying any nonlinearities requires much more than "relevant" instruments from a sampling error "strong first-stage" standpoint. Typically in structural analyses, there is little focus on the "first-stage" relationship: i.e., the first-stage relationship is generally not part of the "structural model." Instead, analysts focus on arguing that the instruments are exogenous and independent of the structural unobservables in the "second-stage" that they wish to hold constant. However, "sensitivity to the first-stage implies that the identified relation is not structural... (Haavelmo, 1944) calls this invariance property 'autonomy.'" (Hahn and Ridder, 2011) We show that structural relationships of interest are inherently sensitive to the first-stage relationship between the endogenous variable the instrument. Absent structural justifications for the first-stage specification, the data can be consistent with the opposite conclusions. Thus, for a fixed instrument set, the results raise questions of whether complicated structural models incorporating flexibility for potentially nonlinear responses (via e.g., moment conditions or sieve estimators) in fact yield more accurate counterfactual predictions than linear structural models that are both simpler and more transparent.

[1] If one does not assume that effects are monotonic, the marginal distributions are generally unbounded, to say nothing of the joint distributions.

This paper contributes to several distinct literatures studying causal inference and structural identification when treatment is non-binary. The first studies discrete treatments. In the case of binary instruments, Angrist and Imbens (1995) showed that with an ordered, multi-valued treatment, 2SLS identifies the average causal response (ACR), a weighted average of the causal effect of a one unit increase in treatment. A wide literature has emerged that addresses the limited interpretability of the ACR, but these generally come at the expense of stronger assumptions. Vohra and Goldin (2024) derive informative sharp bounds on a cumulative complier effect by restricting effect heterogeneity across different complier groups. Nibbering and Oosterveen (2024) derive a similar result with unordered treatment. Rose and Shem-Tov (2024) show that recoding treatment as a binary indicator produces an interpretable weighted average of treatment effects by excluding types of compliers in the population (see also Andresen and Huber, 2021). Chernozhukov et al., 2024 also derive results on identification with a binary instrument under copula restrictions. Multi-valued instruments generally give an opportunity to learn more about the structural functions, but existing approaches still restrict heterogeneity. In the case of multi-valued instruments, Heckman, Urzua, and Vytlacil (2006), Heckman and Vytlacil (2007), and Heckman, Urzua, and Vytlacil (2008) showed that identification of more interpretable economic parameters is possible by relating treatment and the instrument using a discrete choice model. Heckman and Pinto, 2018 and Lee and Salanié, 2018; Lee and Salanié, 2024 study identification under more general monotonicity conditions that do not require treatment values to be ordered. Rather than showing that assumptions are sufficient for identification of target parameters, Goff (2024) and Navjeevan, Pinto, and Santos (2023) take the opposite approach and study necessary conditions for point identification of target parameters. We show that under unrestricted heterogeneity and ordered instruments, the complier share between treatment values is generally not point identified.

The second studies continuous treatments. Many of the above approaches do not apply, but similarly, identification generally relies on restricting heterogeneity. The control function approach of Imbens and Newey (2009) restricts the dimension of heterogeneity in the selection equation. The nonparametric IV approach of Newey and Powell (2003) restricts the dimension of heterogeneity in the potential outcomes. Finally, Torgovitsky (2015), D'Haultfœuille and Février (2015), and Chernozhukov et al. (2024) derive results that combine restrictions on outcome heterogeneity with restrictions on selection to derive identification with discrete instruments.

Encompassing these literatures, we consider both continuous and discrete treatments. In contrast to these literatures, we focus on identification of nonlinear effects across ordered treatment values. We maintain the monotonicity condition of Angrist and Imbens (1995),[2], and we do not restrict the dimension of heterogeneity in either the potential outcomes or treatments. One of the reasons LATE is influential is because of what it implies about the reduced form. The reduced-form of credible research designs provide evidence of structural effects' direction without strong assumptions. In contrast, our analysis suggests that because credible reduced-forms can be consistent with the opposite nonlinear conclusion, that absent very powerful instruments, conclusions of nonlinear effects may hinge solely on restrictions on heterogeneity.

This paper also links two literatures. The paper extends Manski's (1997) analysis of what can be learned under sign restrictions to two dimensions, utilizing tools from the broad literature on stochastic order, spanning economics (Atkinson and Bourguignon, 1982) and probability (Shaked, 2007). In an instrumental variable setting where the instrument is excludable and the structural relationships are monotonic, the bivariate marginal distributions of potential outcomes are stochastically ordered. We use this to derive bounds on potential outcomes distributions and provide tests for structural concavity and convexity.

Relatedly, testing stochastic order for random vectors involves showing that for any lower set the probability of one being in the set is greater than or equal to the probability of the other being in the set. Practically, searching over all lower sets is computationally burdensome, so existing tests focus on searching over a subset of them (Crawford, 2005; Stengos and Thompson, 2012; McCaig and Yatchew, 2007). Moreover, the tests are infeasible when one wants to test whether uncountable random vectors are stochastically ordered (as is our case with continuous instruments). We show that finding the maximal lower set is a variational calculus problem and provide necessary and sufficient conditions for the maximal lower set.

This paper is organized as follows. Section 2 lays out the basic assumptions and notation. Section 3 analyzes marginals and multi-marginals. Section 4 provides bounds for potential outcomes under monotone effects and analyzes structural convexity and concavity. Proofs for Section 3 are in the Appendix, and proofs for Section 4 are in progress.

2 Setup and Notation

2.1 Standard Fare

We consider a situation where an analyst observes the joint distribution of (Y, X, Z), the outcome, the potentially endogenous treatment, and the instrument. Formally, we follow

[2] This assumption is not nested by the unordered monotonicity condition of (Heckman and Pinto, 2018)

Imbens (2007) and assume exclusion and monotonicity mirroring assumptions used in Imbens and Angrist (1994):

Assumption 1 (Triangular System and Exclusion Restriction)**.**

1. ***Triangular System**: We assume that random variables Y and X are unknown functions of Z, η, ε,*

$$\begin{aligned} Y &= g(X; \varepsilon) \\ X &= h(Z; \eta) \end{aligned}$$

where X, Y, Z are observed with real-valued realizations, ε and η are latent, potentially infinite dimensional nuissance parameters capturing unobservables to hold constant, and g and h are structural functions of interest.

2. ***Exclusion Restriction**: We assume that our instrument is independent of the unobservables,*

$$Z \perp (\varepsilon, \eta)$$

Assumption 2 (Weak Monotonicity in the Instrument)**.** *$h(z; \eta)$ is weakly increasing in z for all η.*

All our results apply for the case where $h(z; \eta)$ is decreasing in z by considering $-Z$ as the instrument. Note that our Assumption 2 is weaker than Kasy (2014) and equivalent to Imbens and Angrist's: not every increase in z necessarily corresponds to an increase in X.

To focus on notions of marginality, we depart from usual notation and index potential outcomes with z.

Definition 1 (Potential Outcomes)**.**

$$\begin{aligned} X^z &= h(z; \eta) \\ Y^z &= g(X^z; \varepsilon) \\ &= g(h(z; \eta); \varepsilon) \end{aligned}$$

We consider structural counterfactual predictions using those potential outcomes, omitting the notation for unobserved dimensions of heterogeneity. Namely, potential outcomes have a superscript, and observed outcomes do not. This notation instead puts stronger focus on the direct "intention-to-treat" causal effects of Z. When considering nonlinear effects of X, we will use g directly rather than the abbreviated potential outcomes notation.

The data generating process selectively reveals agents' structural potential outcomes as observed outcomes. In other words, random variables Y and X are functions of observed Z and unobserved ε and η. The goal of instrumental variables is generally to hold constant those unobservables to indirectly understand the causal effect of X on Y.

2.2 Notation

Throughout, we denote cumulative distribution functions (CDFs) with F and their densities by f, and we use subscripts to denote the corresponding random variable. We denote the quantile function or generalized inverse of the CDF of random variable W as $F_W^{-1}(u) = \inf_w \{w : F_W(w) \geq u\}$. Where possible, we will write functions of observed random variables on the left hand side of equations and functions of potential outcomes on the right hand side of equations.

3 Selection: Inframarginals, Marginals, and Multi-marginals

3.1 Inframarginals and their potential outcomes

Let $\mathcal{Z}$ denote the support of the instrument Z. Here we define the units that are inframarginal and marginal units on any arbitrary subset $\mathcal{Z}_0 \subseteq \mathcal{Z}$.

Definition 2 (Inframarginal units)**.** A unit is **inframarginal** on $\mathcal{Z}_0$ if X^z is a constant function of z over $\mathcal{Z}_0$. Let $N_{\mathcal{Z}_0}$ be an indicator for whether a unit is inframarginal on $\mathcal{Z}_0$.

There are two immediate implications of the exclusion restriction. First, if a unit is inframarginal on $\mathcal{Z}_0$, its intention-to-treat effect is also zero on the interval. Second, the conditional distribution of $(Y, X) \mid Z = z, N_{\mathcal{Z}_0} = 1$ is fixed for $z \in \mathcal{Z}_0$.

Lemma 1 (Y^z constant for inframarginals)**.** *If $N_{\mathcal{Z}_0} = 1$, then Y^z is a constant function of z over $\mathcal{Z}_0$.*

Lemma 2 (Fixed distribution of observed outcomes for inframarginals)**.** *The distribution function of realized outcomes for inframarginals is fixed. For any $z, z' \in \mathcal{Z}_0$, $F_{Y,X|Z,N_{\mathcal{Z}_0}}(y, x \mid z', 1) = F_{Y^z,X^z|N_{\mathcal{Z}_0}}(y, x \mid z, 1)$.*

Many of the subsequent results come from noting that the observed distributions are mixtures of marginals and inframarginals and applying Lemma 2.

3.2 Marginals and binary instruments

The objective of this section is to derive bounds on the share of inframarginals and their marginal counterparts in the simplest possible scenario where Z takes two values, $\mathcal{Z} = \{0, 1\}$. We later extend the results to Z with arbitrary support, including the common case where Z is continuously distributed.

Definition 3 (Marginal units)**.** A unit is **marginal** on $\mathcal{Z}_0$ if it is not inframarginal on $\mathcal{Z}_0$. Let $M_{\mathcal{Z}_0} := 1 - N_{\mathcal{Z}_0}$ be an indicator for whether a unit is marginal on $\mathcal{Z}_0$.

Because units are either inframarginal or marginal, observed distributions are mixtures of distributions for inframarginals and marginals. Throughout, we assume that the distribution functions of potential outcomes X^z, Y^z are absolutely continuous with respect to the Lebesgue measure.

Assumption 3. *The conditional CDF of potential outcomes $F_{X^z,Y^z}(x, y)$ is absolutely continuous with respect to the Lebesgue measure with corresponding density $f_{X^z,Y^z}(x, y)$.*

It follows that the conditional distribution functions of observed outcomes $F_{X,Y|Z}(x, y \mid z)$ are also absolutely continuous.

3.2.1 Sharp Bounds on Inframarginal Share

The Upper Bound The upper bound on inframarginals (and correspondingly, the lower bound on marginals) comes from attributing as much of the data as possible to inframarginals whose potential outcomes do not change. Denote this overlapping density $\mathsf{overlap}_{\mathcal{Z}_0}(x, y) := \min_{z\in\mathcal{Z}_0} \left\{ f_{X,Y|Z}(x, y \mid z) \right\}$.

Lemma 3 (Upper Bound on Inframarginal Share)**.** *Let $z_0 \in \mathcal{Z}_0$ and let the inframarginal share $\mathbf{E}[N_{\mathcal{Z}_0}] = p_{\mathcal{Z}_0}$. Then,*

$$p_{\mathcal{Z}_0} \leq \bar{p}_{\mathcal{Z}_0} := \int_{\mathbb{R}^2} \mathsf{overlap}_{\mathcal{Z}_0}(x, y)\, dxdy.$$

Proof. By Lemma 2, $f_{X^z,Y^z|N_{\mathcal{Z}_0}}(x, y \mid 1) = f_{X^{z_0},Y^{z_0}|N_{\mathcal{Z}_0}}(x, y \mid 1)\ \forall z \in \mathcal{Z}_0$, and every conditional distribution is a mixture of marginals and inframarginals, $f_{X,Y|Z}(x, y \mid z) = p_{\mathcal{Z}_0} f_{X^{z_0},Y^{z_0}|N_{\mathcal{Z}_0}}(x, y \mid 1) + (1 - p_{\mathcal{Z}_0})\, f_{X^z,Y^z|N_{\mathcal{Z}_0}}(x, y \mid 0), \forall z \in \mathcal{Z}_0$. Because both terms are non-negative, $\mathsf{overlap}_{\mathcal{Z}_0}(x, y) \geq p f_{X^{z_0},Y^{z_0}|N_{\mathcal{Z}_0}}(x, y \mid 1)$. The inequality follows by integrating both sides. □

Lemma 4 (Sharpness of Bounds with Binary Support)**.** *If the support of $\mathcal{Z}$ is binary, $\bar{p}_{\mathcal{Z}}$ is a sharp upper bound.*

The proof involves subtracting the overlapping density from each of the observed densities and assuming constant ranks among marginal agents. The lower bound on marginals is immediate.

Corollary (Sharp Lower Bound on Marginal Share)**.** *The marginal share* $\mathbf{E}[M_{\mathcal{Z}}] = 1 - p$ *has a sharp lower bound of* $1 - \bar{p}_{\mathcal{Z}}$.

We make three additional remarks. First, we assume that the random variables are absolutely continuous with respect to the Lebesgue measure because most empirical data admit densities, especially those where one would be interested in estimating structural nonlinearities. However, the inequalities hold for arbitrary probability measures, so one can construct a similar proof by instead defining the overlap function as the greatest lower bound of the two probability measures conditional on Z and defining the bound as the Lebesgue integral over the state space. Second, it is no surprise that data generated by a binary instrument can be rationalized by linear structural functions: each set of potential outcomes can be connected by a line. However, the key idea here is that one need only draw lines between mass "missing" at $Z = 0$ to "excess" mass at $Z = 1$. We apply this same idea when considering multi-valued Z. Third, the upper bound in Lemma 3 does not use monotonicity. Monotonicity is equivalent to first order stochastic dominance of $X \mid Z = 0$ by $X \mid Z = 1$, which are observable (or X^0 by X^1, which are not), and provides no other restrictions to refine the upper bounds. Monotonicity plays a role in the lower bound for inframarginals.

The Lower Bound Consider an irrelevant instrument such that $X \perp Z$. All agents must be inframarginal to the instrument. By opposite token, if the distribution of $X \mid Z = 1$ is simply a horizontal shift of the distribution of $X \mid Z = 0$, then all agents are marginal to the instrument under a rank equivalence condition. The lower bound on inframarginals comes from measuring the overlap in the distribution functions, and the proof is a nearly immediate consequence of Example 2 in Arnold, Molchanov, and Ziegel (2020).

Lemma 5 (Lower Bound on Inframarginal Share)**.** *The share of units that are inframarginal on* $\mathcal{Z}_0$, $\mathbf{E}[N_{\mathcal{Z}_0}] = p_{\mathcal{Z}_0}$, *is bounded from below by*

$$\begin{aligned} p_{\mathcal{Z}_0} \geq \underline{p}_{\mathcal{Z}_0} &:= \int_{\mathbb{R}} D(x) dF_X \\ &= \int_0^1 \tilde{D}(p) dp \end{aligned}$$

where $D(x) := \mathbf{1}\left[\left\{x : F_{X|Z}(x \mid z) = F_X(x) \forall z \in \mathcal{Z}_0\right\}\right]$ *is an indicator for* x *where the conditional distribution functions overlap so* $\mathbf{1}[X \leq x] \perp Z$. *Alternatively, one can write the bound* *in terms of* $\tilde{D}(p) := \mathbf{1}\left[\left\{p : F_{X|Z}^{-1}(p \mid z) = F_X^{-1}(p) \quad \forall z \in \mathcal{Z}_0\right\}\right]$, *an indicator for where the quantile functions overlap, i.e. if* $U = F_X(X)$, *the* p *such that* $\mathbf{1}[U \le p] \perp Z$ *and* $z_0 \in \mathcal{Z}_0$.

Proof. Fix $z_0 \in \mathcal{Z}_0$. By Arnold, Molchanov, and Ziegel (2020) Example 2, if $F_{X^z}(x) = F_{X^{z_0}}(x)$, then $\Pr[X^z = X^{z_0} \mid X^{z_0} = x] = 1$. Extending this argument, if $F_{X^z}(x)$ is some constant for some x and $\forall z \in \mathcal{Z}_0$, then $F_{X^z}(x) = F_{X^{z_0}}(x)$. Correspondingly, $\Pr[X^z = x \quad \forall z \mid X^{z_0} = x] = 1$ and $F_{X|Z}(x \mid z) = F_X(x) \quad \forall z \in \mathcal{Z}_0$. The inequality comes from integrating over the support of X with respect to the corresponding measure. □

Lemma 6 (Sharpness of Lower Bound with Binary Support)**.** *If the support of* $\mathcal{Z}$ *is binary,* $\underline{p}_{\mathcal{Z}}$ *is a sharp lower bound.*

The bound is achieved by assuming units have constant ranks.

Corollary (Sharp Upper Bound on Marginal Share)**.** *The marginal share* $\mathbf{E}[M_{\mathcal{Z}}] = 1 - p_{\mathcal{Z}}$ *has a sharp upper bound of* $1 - \underline{p}_{\mathcal{Z}}$.

Interestingly, these bounds use monotonicity but not exclusion. We summarize the bounds from Lemmas 4 and 6 in Theorem 1.

Theorem 1 (Sharp bounds on share of inframarginals and marginals)**.**

1. *The inframarginal share* $\mathbf{E}[N_{\mathcal{Z}}] = p_{\mathcal{Z}}$ *is sharply bounded by* $\underline{p}_{\mathcal{Z}} \le p_{\mathcal{Z}} \le \bar{p}_{\mathcal{Z}}$.
2. *The marginal share* $\mathbf{E}[M_{\mathcal{Z}}] = 1 - p_{\mathcal{Z}}$ *is sharply bounded by* $1 - \bar{p}_{\mathcal{Z}} \le 1 - p_{\mathcal{Z}} \le 1 - \underline{p}_{\mathcal{Z}}$

Special Case: Binary Treatment If treatment is binary so the support of X is $\{0, 1\}$, then the bounds coincide and the marginal share is point identified, corresponding to the familiar expression for the complier share.

Example 1 (Share of inframarginals (never and always takers) and marginal compliers with binary treatment)**.**

$$p_{\mathcal{Z}} = \bar{p}_{\mathcal{Z}} = \underline{p}_{\mathcal{Z}} = \mathbf{E}[X \mid Z = 1] - \mathbf{E}[X \mid Z = 0]$$

Proof. Modifying the proof of Lemma 3 slightly to accommodate discrete X,, note:

$$\min\left\{f_{XY|Z}(y \mid 0), f_{XY|Z}(y \mid 1)\right\} = f_{XY|Z}(y \mid 0)$$
$$\min\left\{f_{(1-X)Y|Z}(y \mid 0), f_{(1-X)Y|Z}(y \mid 1)\right\} = f_{(1-X)Y|Z}(y \mid 1),$$

corresponding to the densities of Y^1 and Y^0 for the always takers and the never takers, respectively (see e.g., Abadie, 2003; Kitagawa, 2015). Integrating over the support of Y, the upper bound on the inframarginal share is $\bar{p}_{\mathcal{Z}} = \mathbf{E}[X \mid Z = 0] + \mathbf{E}[1 - X \mid Z = 1]$.

For the other bound, we can compute the expression directly, noting that $F^{-1}_{X|Z}(p \mid z) = \mathbf{1}[p \geq \mathbf{E}[X \mid Z = z]]$, so $\underline{p}_{\mathcal{Z}} = \mathbf{E}[X \mid Z = 0] + 1 - \mathbf{E}[X \mid Z = 1]$. □

3.3 Multi-marginals and multi-valued instruments

When the instrument is binary, at most two potential outcomes are revealed to the analyst, so the data can be rationalized by heterogeneous (but linear) effects. To rule out linear effects, analysts must observe more than two potential outcomes. Instruments with larger support allow analysts to observe more than two potential outcomes.

However, just as binary instruments do not guarantee that agents see more than one potential outcome, larger support does not guarantee two (much less three) uncensored potential outcomes. In this section, we define agents that are marginal to multiple values of the instrument. If there are no such units, then the data can be rationalized by linear structural relationships. Paralleling Definition 3, we first define multi-marginals.

Definition 4 (Multi-marginal units)**.** A unit is **multi-marginal** on an interval $\mathcal{Z}_0 = [z_0, z_1] \subseteq \mathcal{Z}$ if there is some $z^* \in \mathcal{Z}_0$ such that the unit is marginal on the interval $[z_0, z^*]$ *and* the unit is marginal on interval $[z^*, z_1]$. Let $M^{multi}_{\mathcal{Z}_0}$ be an indicator for whether a unit is multi-marginal on $\mathcal{Z}_0$.

The definition has two immediate implications. First, if a unit is multi-marginal then $X(z) > X(z') > X(z'')$ for some $z > z' > z''$. Second, multi-marginal units on $\mathcal{Z}_0$ are also marginal on $\mathcal{Z}_0$.

3.3.1 Three-valued instruments

We can derive explicit bounds on the share of multi-marginal units in the simplest possible scenario accommodating multi-marginal units, where $\mathcal{Z}_0 = \mathcal{Z} = \{0, 1, 2\}$. The lower bound on multi-marginals comes from attributing as much of the data as possible to units that are inframarginal on $\mathcal{Z}_1 = \{0, 1\}$ or $\mathcal{Z}_2 = \{1, 2\}$.

Lemma 7 (Sharp Lower Bound on Multimarginal Share)**.** *If $\mathcal{Z} = \{0, 1, 2\}$ then the multi-marginal share $\boldsymbol{E}[M^{multi}_{\mathcal{Z}}] = p^{multi}_{\mathcal{Z}}$ has a sharp lower bound given by*

$$p^{multi}_{\mathcal{Z}} \geq \int \max\left(f_{X,Y|Z}(x, y \mid 1) - \mathsf{overlap}_{\mathcal{Z}_1}(x, y) - \mathsf{overlap}_{\mathcal{Z}_2}(x, y), 0\right) dxdy$$

where $\mathcal{Z}_1 = \{0,1\}$ *and* $\mathcal{Z}_2 = \{1,2\}$.

The upper bound comes from assuming rank invariance. Under rank invariance, everyone is multi-marginal except those where $F_{X|Z}(x|0) = F_{X|Z}(x|1)$ or $F_{X|Z}(x|1) = F_{X|Z}(x|2)$. This leads to the following result.

Lemma 8 (Sharp Upper Bound on Multimarginal Share)**.** *If* $\mathcal{Z} = \{0,1,2\}$ *then the multi-marginal share* $\boldsymbol{E}[M_{\mathcal{Z}}^{multi}] = p_{\mathcal{Z}}^{multi}$ *has a sharp upper bound given by*

$$p_{\mathcal{Z}}^{multi} \leq 1 - \left(\underline{p}_{\mathcal{Z}_0} + \underline{p}_{\mathcal{Z}_1} - \underline{p}_{\mathcal{Z}}\right)$$

where $\mathcal{Z}_1 = \{0,1\}$ *or* $\mathcal{Z}_2 = \{1,2\}$.

3.3.2 Continuous instruments

Testing for multi-marginals when Z is continuously distributed follows similar ideas from Sections XXYY. Between two instrument values, units are inframarginal if their potential outcomes do not change. The bounds on the population come from attributing all overlapping mass to inframarginals. In contrast, where mass is "missing," units must be marginal. If there are no multi-marginals, then all missing mass must come from the baseline density of potential outcomes.

For this section, we assume that Y, X and Z are continuously distributed.

Assumption 4. *The conditional density,* $f_{Y,X|Z}(y, x \mid z)$ *is absolutely continuous in* z *for all* x, y *with partial derivative* $\psi(y,x,z) := \frac{\partial}{\partial z} f_{Y,X|Z}(y, x \mid z)$*, which is continuous in* x, y *and* z *and absolutely integrable,* $\int_0^1 \int_{\mathcal{X}} \int_{\mathcal{Y}} |\psi(y,x,z)| dydxdz < \infty$

We further assume that $\mathcal{Z} = [z_0, z_1]$, which we normalize to the unit interval, $[0,1]$.

To operationalize the notions of excess and missing mass, let $\psi^+(y,x,z) := \max(\psi(y,x,z), 0)$ and $\psi^-(y,x,z) := \max(-\psi(y,x,z), 0)$ denote the positive and negative parts of the density derivative. We can then state a simple condition for whether the identified set for $p_{\mathcal{Z}}^{multi}$ contains 0.

Theorem 2 (Test for multimarginals)**.** *The identified set for* $p_{\mathcal{Z}}^{multi}$ *contains* 0 *if and only if* $\int_0^1 \psi^-(y,x,z)dz \leq f_{Y^0,X^0}(y,x)$ *for all* $x, y \in Supp\{X,Y\}$.

If the total amount of "missing mass" does not exceed the baseline density, then as in Section 3, the data can be rationalized by heterogeneous step functions in the first stage and linear structural equations with heterogeneous slopes and intercepts.

Rationalizing nonlinearities without multimarginals Theorem 2 shows that the density condition is both necessary and sufficient for the nonexistence of multimarginals. If there are no multimarginals, then following the sufficiency direction of the proof, any structural nonlinearity is rationalizable.

Corollary 1. *Let $d(x;\varepsilon)$ be continuous in x on $[\underline{x}, \bar{x}]$ where $\underline{x}$ is a lower bound for $x_0(\eta)$, and $\bar{x}$ is an upper bound for $x_1(\eta, Z^*)$. If $p_{\mathcal{Z}}^{multi} = 0$, then $d(x;\varepsilon)$ is in the identified set of $\frac{\partial g^{(\delta)}(x;\varepsilon)}{\partial x^\delta}$ for any $\delta \geq 2$.*

Since any structural nonlinearity is rationalizable, it immediately follows that any nonlinearity in the average structural function or quantile structural function among any subpopulation is also rationalizable.

Corollary 2. *Let d be defined as in Corollary 1. If $p_{\mathcal{Z}}^{multi} = 0$, then d is in the identified set of derivatives of order $\delta \geq 2$ of the conditional average structural function $\frac{\partial^{(\delta)}}{\partial x^\delta}\mathbf{E}_\varepsilon[g(x,\varepsilon)|W]$ and the conditional quantile structural function $\frac{\partial^{(\delta)}}{\partial x^\delta}\inf\{y : \mathbf{E}_\varepsilon[\mathbf{1}[g(x,\varepsilon) \leq yt]|W] \geq \tau\}$ for any quantile τ.*

4 *All I wanted to do was regress Y on X and X^2*: Testing direction and convexity with instruments

4.1 Direction: Testing two-dimensional stochastic order and monotonicity in structural effects

Definition 5. W^z stochastically dominates $W^{z'}$ for $z > z'$ if $\Pr[W^z \in L] - \Pr\left[W^{z'} \in L\right] \leq 0$ for any lower set L.

Corollary 3 (Shaked and Shantikumar Theorem 6.B.1)**.** *(X^z, Y^z) is increasing in z if and only if there exists a structural function $g(x;\varepsilon)$ that (1) is increasing in x for all ε and (2) rationalizes the data (i.e. $Y^z = g(X^z;\varepsilon)$).*

Let,

$$M(z) = \max_{L_z \in LS} \Pr[W^z \in L_z] - \Pr\left[W^{z'} \in L_z\right]$$
$$M(z) = \max_{L_z \in LS} \frac{\partial}{\partial z} \Pr[W^z \in L_z]$$

for discrete and continuous Z, respectively, and LS is the set of all lower sets in $\mathbb{R}^2$

Lemma 9. *W^z stochastically dominates $W^{z'}$ for $z > z'$ if and only if* $\max_z M(z) \leq 0$.

Assumption 5 (Monotone effects)**.** *g is increasing in x for all ε.*

4.2 Sharp bounds on potential outcomes distributions under monotone structural effects

Let

- $Q2(x_0, y_0) = \{(x, y) : x \leq x_0, y > y_0\}$ and $Q4(x_0, y_0) = \{(x, y) : x > x_0, y \leq y_0\}$
- $S_z^2(x, y) = \Pr[W^z \in Q2(x, y)]$ and $S_z^4(x, y) = \Pr[W^z \in Q4(x, y)]$

Fix x and y and for ease of notation denote $Q2(x, y) = Q2$ and $Q4(x, y) = Q4$

Lemma 10. *Under assumption 5,*

1. *If $g(x; \varepsilon) \leq y$, then $W^z \notin Q2$ for all z.*
2. *If $W^z \in Q4$ for any z, then $g(x; \varepsilon) \leq y$*

Corollary 4. *The CDF of potential outcomes are bounded above and below by*

1. $\Pr[g(x; \varepsilon) \leq y] \leq \Pr[\cap_{z \in \mathcal{Z}} W^z \notin Q2]$
2. $\Pr[g(x; \varepsilon) \leq y] \geq \Pr[\cup_{z \in \mathcal{Z}} W^z \in Q4]$

4.2.1 Binary instruments

Lemma 11. *The CDF of potential outcomes are sharply bounded above by*

$$\begin{aligned}\Pr[g(x; \varepsilon) \leq y] &\leq \Pr\left[W^0 \notin Q2 \cap W^1 \notin Q2\right] \\ &= 1 - \Pr\left[W^0 \in Q2\right] - \Pr\left[W^1 \in Q2\right] + \Pr\left[W^0, W^1 \in Q2\right]\end{aligned}$$

and sharply bounded below by

$$\begin{aligned}\Pr[g(x; \varepsilon) \leq y] &\geq \Pr\left[W^0 \in Q4 \cup W^1 \in Q4\right] \\ &= \Pr\left[W^0 \in Q4\right] + \Pr\left[W^1 \in Q4\right] - \Pr\left[W^0, W^1 \in Q4\right]\end{aligned}$$

Lemma 12. $\Pr[W^0, W^1 \in Q2]$ *is sharply bounded above by*

$$\Pr\left[W^0, W^1 \in Q2\right] \leq \Pr\left[W^1 \in Q2\right] - \max_{L \in LS}\left\{\Pr\left[W^1 \in Q2 \cap L\right] - \Pr\left[W^0 \in Q2 \cap L\right]\right\}$$

and $\Pr\left[W^0, W^1 \in Q4\right]$ *is sharply bounded above by*

$$\Pr\left[W^0, W^1 \in Q4\right] \leq \Pr\left[W^1 \in Q4\right] - \max_{L \in LS}\left\{\Pr\left[W^1 \in Q4 \cap L\right] - \Pr\left[W^0 \in Q4 \cap L\right]\right\}$$

Theorem 3. *The CDF of potential outcomes are sharply bounded above by*

$$\Pr\left[g\left(x;\varepsilon\right) \leq y\right] \leq 1 - \Pr\left[W^0 \in Q2\right] - \max_{L \in LS}\left\{\Pr\left[W^1 \in L \cap Q2\right] - \Pr\left[W^0 \in L \cap Q2\right]\right\}$$

and sharply bounded below by

$$\Pr\left[g\left(x;\varepsilon\right) \leq y\right] \geq \Pr\left[W^0 \in Q4\right] + \max_{L \in LS}\left\{\Pr\left[W^1 \in L \cap Q4\right] - \Pr\left[W^0 \in L \cap Q4\right]\right\}$$

4.2.2 Multi-valued/continuous instruments

Discrete instruments

Theorem 4. *The CDF of potential outcomes are sharply bounded above by*

$$\Pr\left[g\left(x;\varepsilon\right) \leq y\right] \leq 1 - \Pr\left[W^0 \in Q2\right] - \sum_{k=1}^{K} \max_{L_k \in LS}\left\{\Pr\left[W^k \in L_k \cap Q2\right] - \Pr\left[W^{k-1} \in L_k \cap Q2\right]\right\}$$

and sharply bounded below by

$$\Pr\left[g\left(x;\varepsilon\right) \leq y\right] \geq \Pr\left[W^0 \in Q4\right] + \sum_{k=1}^{K} \max_{L_k \in LS}\left\{\Pr\left[W^k \in L_k \cap Q4\right] - \Pr\left[W^{k-1} \in L_k \cap Q4\right]\right\}$$

Continuous instruments

Lemma 13. *The CDF of potential outcomes is sharply bounded above by*

$$\begin{aligned}\Pr\left[g\left(x;\varepsilon\right) \leq y\right] \leq & 1 - \Pr\left[W^0 \in Q2\right] \\ & - \int_0^1 \frac{\max_{L_z \in LS}\left\{\Pr\left[W^{z+dz} \in L_z \cap Q2\right] - \Pr\left[W^z \in L_z \cap Q2\right]\right\}}{dz} dz\end{aligned}$$

and sharply bounded below by

$$\begin{aligned}\Pr\left[g\left(x;\varepsilon\right) \geq y\right] \leq & \Pr\left[W^0 \in Q4\right] \\ & + \int \frac{\max_{L_z \in LS}\left\{\Pr\left[W^{z+dz} \in L_z \cap Q4\right] - \Pr\left[W^z \in L_z \cap Q4\right]\right\}}{dz} dz\end{aligned}$$

Theorem 5. *If regularity conditions, then the CDF of potential outcomes is sharply bounded above by*

$$\Pr\left[g\left(x;\varepsilon\right)\leq y\right]\leq 1-\Pr\left[W^{0}\in Q2\right]-\int_{0}^{1}\frac{d\Pr\left[W^{z}\in Q2\right]}{dz}dz$$

and sharply bounded below by

$$\Pr\left[g\left(x;\varepsilon\right)\leq y\right]\leq \Pr\left[W^{0}\in Q4\right]-\int_{0}^{1}\frac{d\Pr\left[W^{z}\in Q4\right]}{dz}dz$$

4.3 Finding maximal lower sets

Lemma 14. *Let* $S\left(x\right)=\left\{y\in\mathbb{R}:\left(x,y\right)\in L\right\}$, $T=\left\{x\in\mathbb{R}:S\left(x\right)\neq\emptyset\right\}$, $b^{L}\left(x\right)=\sup S\left(x\right)$, *and* $\tilde{L}=\left\{\left(x,y\right):x\in T,y<b^{L}\left(x\right)\right\}$. *Then,* $\Pr\left[W\in L\right]=\Pr\left[W\in\tilde{L}\right]=\Pr\left[Y<b^{L}\left(X\right)\right]$

Lemma 15.

$$M\left(z\right)=\min_{b(x;z)}\int_{-\infty}^{\infty}\int_{-\infty}^{b(x;z)}\left[f_{z_{j}}\left(x,y\right)-f_{z_{j+1}}\left(x,y\right)\right]dydx$$
$$M\left(z\right)=\min_{b(x;z)}\int_{-\infty}^{\infty}\int_{-\infty}^{b(x;z)}\frac{\partial f\left(x,y;z\right)}{\partial z}dydx$$

for discrete and continuous Z*, respectively.*

4.3.1 Binary instruments

First, consider the case of binary instruments where $\mathcal{Z}=\{0,1\}$. We can write $\Pr\left[W^{0}\in L\right]-\Pr\left[W^{1}\in L\right]=\Pr\left[Y^{0}\leq b^{L}\left(X^{0}\right)\right]-\Pr\left[Y^{1}\leq b^{L}\left(X^{1}\right)\right]$. Combining Lemmas 10–13,

Lemma 16. *If the distribution of* (X^{z},Y^{z}) *is absolutely continous for all* z*, then* $g(x;\varepsilon)$ *is increasing in* x *for all* ε *if and only if*

$$\min_{L\in LS}\Pr\left[W^{0}\in L\right]-\Pr\left[W^{1}\in L\right]=\min_{b(x)\in\mathcal{B}}\int_{-\infty}^{\infty}\int_{-\infty}^{b(x)}\left[f_{0}\left(x,y\right)-f_{1}\left(x,y\right)\right]dydx\geq 0$$

where $\mathcal{B}$ *is the set of decreasing functions in* $\mathbb{R}$.

Let $L(x)=\int_{-\infty}^{b(x)}\left[f_{0}\left(x,y\right)-f_{1}\left(x,y\right)\right]dy$. Frolov and Frolov show that this is a calculus of variations problem solved by an "upgraded" Lagrangian

$$\mathcal{L}\left(x,b,b';\chi,\zeta\right)=L(x)+\chi\left(x\right)\left(b'\left(x\right)+\zeta^{2}\left(x\right)\right).$$

Thus,

Lemma 17. *$b^*(x)$ is a local minimizer of $\int_{-\infty}^{\infty} \mathcal{L}(x,b,b';\chi,\zeta)dx$ if and only if*

1. *$b^*(x)$ is non-increasing,*
$$0 \geq \frac{db^*(x)}{dx}$$

2. *When $b^*(x)$ is decreasing, the bivariate densities are equal,*
$$0 = \frac{db^*(x)}{dx}\left[f_0\left(x, b^*\left(x\right)\right) - f_1\left(x, b^*\left(x\right)\right)\right]^2$$

3. *The bivariate densities are increasing in y at the optimum,*
$$\frac{\partial f_1}{\partial y}\left(x, b^*\left(x\right)\right) > \frac{\partial f_0}{\partial y}\left(x, b^*\left(x\right)\right)$$

Lemma 18. *b^* is a global minimizer of $\int_{-\infty}^{\infty} \mathcal{L}(x,b,b';\chi,\zeta)dx$ if it is a local minimizer and $\frac{\partial f_1}{\partial y}\left(x,y\right) > \frac{\partial f_0}{\partial y}\left(x,y\right) \quad \forall x,y$*

4.3.2 Multi-valued/continuous instruments

Theorem 6. *Under continuous Z, $g(x;\varepsilon)$ is increasing in x for all ε if $\int_{-\infty}^{\infty}\int_{-\infty}^{b^*(x;z^*)} \frac{\partial f(x,y;z)}{\partial z} dydx \geq 0$ where $b^*(x;z^*)$...*

References

Abadie, Alberto (2003). "Semiparametric instrumental variable estimation of treatment response models". In: *Journal of econometrics* 113.2, pp. 231–263.

Andresen, Martin E and Martin Huber (Feb. 2021). "Instrument-based estimation with binarised treatments: issues and tests for the exclusion restriction". In: *The Econometrics Journal* 24.3, pp. 536–558. ISSN: 1368-4221. DOI: 10.1093/ectj/utab002. eprint: https://academic.oup.com/ectj/article-pdf/24/3/536/40344597/utab002.pdf. URL: https://doi.org/10.1093/ectj/utab002.

Angrist, Joshua D and Guido W Imbens (1995). "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity". In: *Journal of the American Statistical Association* 90.430, pp. 431–442. ISSN: 01621459, 1537274X. URL: http://www.jstor.org/stable/2291054 (visited on 01/31/2025).

Arnold, Sebastian, Ilya Molchanov, and Johanna F Ziegel (2020). "Bivariate distributions with ordered marginals". In: *Journal of Multivariate Analysis* 177, p. 104585.

Atkinson, Anthony B and Francois Bourguignon (1982). "The comparison of multi-dimensioned distributions of economic status". In: *The Review of Economic Studies* 49.2, pp. 183–201.

Bhattacharya, Debopam (2024). "Nonparametric Approaches to Empirical Welfare Analysis". In: *Journal of Economic Literature* 62.2, pp. 554–593.

Chernozhukov, Victor et al. (2024). *Estimating Causal Effects of Discrete and Continuous Treatments with Binary Instruments*. arXiv: 2403.05850 [econ.EM]. URL: https://arxiv.org/abs/2403.05850.

Crawford, Ian (2005). "A nonparametric test of stochastic dominance in multivariate distributions". In: *Discussion Papers in Economics, DP* 12.05.

D'Haultfœuille, Xavier and Philippe Février (2015). "Identification of Nonseparable Triangular Models With Discrete Instruments". In: *Econometrica* 83.3, pp. 1199–1210. DOI: https://doi.org/10.3982/ECTA10038. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA10038. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA10038.

Frangakis, Constantine E and Donald B Rubin (2002). "Principal stratification in causal inference". In: *Biometrics* 58.1, pp. 21–29.

Goff, Leonard (2024). *When does IV identification not restrict outcomes?* arXiv: 2406.02835 [econ.EM]. URL: https://arxiv.org/abs/2406.02835.

Haavelmo, Trygve (1944). "The probability approach in econometrics". In: *Econometrica: Journal of the Econometric Society*, pp. iii–115.

Hahn, Jinyong and Geert Ridder (2011). "Conditional moment restrictions and triangular simultaneous equations". In: *Review of Economics and Statistics* 93.2, pp. 683–689.

Heckman, James J and Rodrigo Pinto (2018). "Unordered Monotonicity". In: *Econometrica* 86.1, pp. 1–35. DOI: https://doi.org/10.3982/ECTA13777. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA13777. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13777.

Heckman, James J, Sergio Urzua, and Edward Vytlacil (Aug. 2006). "Understanding Instrumental Variables in Models with Essential Heterogeneity". In: *The Review of Economics and Statistics* 88.3, pp. 389–432. ISSN: 0034-6535. DOI: 10.1162/rest.88.3.389. eprint: https://direct.mit.edu/rest/article-pdf/88/3/389/1614210/rest.88.3.389.pdf. URL: https://doi.org/10.1162/rest.88.3.389.

— (2008). "Instrumental Variables in Models with Multiple Outcomes: the General Unordered Case". In: *Annales d'Économie et de Statistique* 91/92, pp. 151–174. ISSN: 0769489X, 22726497. URL: http://www.jstor.org/stable/27917243 (visited on 01/31/2025).

Heckman, James J and Edward Vytlacil (2007). "Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments". In: *Handbook of econometrics* 6, pp. 4875–5143.

Horowitz, Joel L (2011). "Applied nonparametric instrumental variables estimation". In: *Econometrica* 79.2, pp. 347–394.

Imbens, Guido W (2007). "Nonadditive models with endogenous regressors". In: *Econometric Society Monographs* 43, p. 17.

Imbens, Guido W and Joshua D Angrist (1994). "Identification and Estimation of Local Average Treatment Effects". In: *Econometrica: Journal of the Econometric Society*, pp. 467–475.

Imbens, Guido W and Whitney K Newey (2009). "Identification and estimation of triangular simultaneous equations models without additivity". In: *Econometrica* 77.5, pp. 1481–1512.

Kasy, Maximilian (2011). "Identification in triangular systems using control functions". In: *Econometric Theory* 27.3, pp. 663–671.

— (2014). "Instrumental variables with unrestricted heterogeneity and continuous treatment". In: *The Review of Economic Studies* 81.4, pp. 1614–1636.

Kitagawa, Toru (2015). "A test for instrument validity". In: *Econometrica* 83.5, pp. 2043–2063.

Lee, Sokbae and Bernard Salanié (2018). "Identifying Effects of Multivalued Treatments". In: *Econometrica* 86.6, pp. 1939–1963. DOI: https://doi.org/10.3982/ECTA14269.

eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA14269. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA14269.

Lee, Sokbae and Bernard Salanié (2024). *Treatment Effects with Targeting Instruments*. arXiv: 2007.10432 [econ.EM]. URL: https://arxiv.org/abs/2007.10432.

Manski, Charles F (1997). "Monotone treatment response". In: *Econometrica: Journal of the Econometric Society*, pp. 1311–1334.

McCaig, Brian and Adonis Yatchew (2007). "International welfare comparisons and nonparametric testing of multivariate stochastic dominance". In: *Journal of Applied Econometrics* 22.5, pp. 951–969.

Navjeevan, Manu, Rodrigo Pinto, and Andres Santos (2023). *Identification and Estimation in a Class of Potential Outcomes Models*. arXiv: 2310.05311 [econ.EM]. URL: https://arxiv.org/abs/2310.05311.

Newey, Whitney K and James L Powell (2003). "Instrumental variable estimation of nonparametric models". In: *Econometrica* 71.5, pp. 1565–1578.

Nibbering, Didier and Matthijs Oosterveen (Aug. 2024). "Instrument-Based Estimation of Full Treatment Effects with Partial Compliers". In: *The Review of Economics and Statistics*, pp. 1–46. ISSN: 0034-6535. DOI: 10.1162/rest_a_01486. eprint: https://direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01486/2466981/rest_a_01486.pdf. URL: https://doi.org/10.1162/rest%5C_a%5C_01486.

Rose, Evan K and Yotam Shem-Tov (June 2024). "On Recoding Ordered Treatments as Binary Indicators". In: *The Review of Economics and Statistics*, pp. 1–32. ISSN: 0034-6535. DOI: 10.1162/rest_a_01462. eprint: https://direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01462/2383847/rest_a_01462.pdf. URL: https://doi.org/10.1162/rest%5C_a%5C_01462.

Shaked, Moshe (2007). *Stochastic orders*. Springer Science & Business Media.

Stengos, Thanasis and Brennan S Thompson (2012). "Testing for bivariate stochastic dominance using inequality restrictions". In: *Economics Letters* 115.1, pp. 60–62.

Torgovitsky, Alexander (2015). "Identification of Nonseparable Models Using Instruments With Small Support". In: *Econometrica* 83.3, pp. 1185–1197. DOI: https://doi.org/10.3982/ECTA9984. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA9984. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA9984.

Vohra, Vedant and Jacob Goldin (Oct. 2024). "Identifying the Cumulative Causal Effect of a Non-Binary Treatment from a Binary Instrument". In: *The Review of Economics and Statistics*, pp. 1–20. ISSN: 0034-6535. DOI: 10.1162/rest_a_01526. eprint: https://direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf. URL: https://doi.org/10.1162/rest%5C_a%5C_01526.

A Proofs

Lemma 4 (Sharpness of Bounds with Binary Support)**.** *If the support of $\mathcal{Z}$ is binary, $\bar{p}_{\mathcal{Z}}$ is a sharp upper bound.*

Proof of Lemma 4. To prove that the bound is sharp, we must show there always exists a DGP, a $g(x;\varepsilon)$ and an $h(z;\eta)$, that (1) satisfies exclusion and monotonicity; (2) rationalizes the data; and (3) achieves the bound. The bounds can be achieved by assuming three dimensions of heterogeneity. Define

1. $\tilde{f}(x,y) = \frac{\mathsf{overlap}_{\mathcal{Z}}(x,y)}{\bar{p}}$
2. $\hat{f}(x,y \mid z) = \frac{f_{X,Y|Z}(x,y|z) - \mathsf{overlap}_{\mathcal{Z}}(x,y)}{1-\bar{p}}$

and note that these are valid density functions, representing the regions of overlapping density and the observed conditional densities less the overlapping density. For the purpose of defining quantile functions associated with the densities, let (U_X, U_Y) be distributed according to $\tilde{f}$ and (V_{X^z}, V_{Y^z}) be distributed according to $\hat{f}$.

The DGP comes from rationalizing each of these densities with inframarginals and marginals, respectively. All variation for the inframarginals must come from heterogeneity, so constant structural functions suffice. The observed distributions for marginals can be rationalized using step and affine functions.

Let

1. $N_{\mathcal{Z}} \sim \text{Bernoulli}(\bar{p})$ denote whether agents are inframarginal;
2. $\eta \sim \text{Uniform}[0,1]$ and $\varepsilon \sim \text{Uniform}[0,1]$ be independent.

Finally, let

$$X(z;\eta,N_{\mathcal{Z}}) = \begin{cases} \tilde{x}(\eta) & N_{\mathcal{Z}} = 0 \\ \hat{x}_0(\eta)\,\mathbf{1}(z<1) + \hat{x}_1(\eta)\,\mathbf{1}(z \geq 1) & N_{\mathcal{Z}} = 1 \end{cases}$$

$$Y(x;\varepsilon,\eta,N_{\mathcal{Z}}) = \begin{cases} \tilde{y}(\varepsilon,\eta) & N_{\mathcal{Z}} = 0 \\ m(\varepsilon,\eta)(x - \hat{x}_0(\eta)) + \hat{y}_0(\varepsilon,\eta) & N_{\mathcal{Z}} = 1 \end{cases}$$

where

1. $\tilde{x}(\eta) = F^{-1}_{U_X}(\eta)$
2. $\tilde{y}(\varepsilon,\eta) = F^{-1}_{U_Y|U_X}(\varepsilon \mid \tilde{x}(\eta))$

3. $\hat{x}_z(\eta) = F^{-1}_{V_{X^z}}(\eta)$

4. $\hat{y}_z(\varepsilon, \eta) = F^{-1}_{V_{Y^z}|V_{X^z}}(\varepsilon \mid \hat{x}_z(\eta))$

5. $m(\varepsilon, \eta) = \frac{\hat{y}_1(\varepsilon,\eta) - \hat{y}_0(\varepsilon,\eta)}{\hat{x}_1(\eta) - \hat{x}_0(\eta)}$.

The distribution of $(X^z, Y^z) \mid Z = z$ has the same distribution as $(X, Y) \mid Z = z$. Monotonicity holds because the quantile function is weakly increasing. Exclusion holds by construction because the potential outcomes of Y are not a function of the realization of random variable Z. □

Lemma 6 (Sharpness of Lower Bound with Binary Support)**.** *If the support of $\mathcal{Z}$ is binary, $\underline{p}_{\mathcal{Z}}$ is a sharp lower bound.*

Proof of Lemma 6. The bound is achieved by assuming rank invariance in X. Let $\eta \sim$ Uniform$[0, 1]$ and $\varepsilon \sim$ Uniform$[0, 1]$ be independent. Let

$$
\begin{aligned}
X(z; \eta) &= x_0(\eta)\,\mathbf{1}(z < 1) + x_1(\eta)\,\mathbf{1}(z \geq 1) \\
Y(x; \varepsilon, \eta) &= m(\varepsilon, \eta)(x - x_0(\eta)) + y_0(\varepsilon, \eta)
\end{aligned}
$$

where

1. $x_z(\eta) = F^{-1}_{X|Z}(\eta \mid z)$
2. $y_z(\varepsilon, \eta) = F^{-1}_{Y|X,Z}(\varepsilon \mid x_z(\eta), z)$
3. $m(\varepsilon, \eta) = \frac{y_1(\varepsilon,\eta) - y_0(\varepsilon,\eta)}{x_1(\eta) - x_0(\eta)}$.

□

Lemma 7 (Sharp Lower Bound on Multimarginal Share)**.** *If $\mathcal{Z} = \{0, 1, 2\}$ then the multimarginal share $\boldsymbol{E}[M^{multi}_{\mathcal{Z}}] = p^{multi}_{\mathcal{Z}}$ has a sharp lower bound given by*

$$
p^{multi}_{\mathcal{Z}} \geq \int \max\left(f_{X,Y|Z}(x, y \mid 1) - \mathsf{overlap}_{\mathcal{Z}_1}(x, y) - \mathsf{overlap}_{\mathcal{Z}_2}(x, y), 0\right) dxdy
$$

where $\mathcal{Z}_1 = \{0, 1\}$ and $\mathcal{Z}_2 = \{1, 2\}$.

Lemma 8 (Sharp Upper Bound on Multimarginal Share)**.** *If $\mathcal{Z} = \{0, 1, 2\}$ then the multimarginal share $\boldsymbol{E}[M^{multi}_{\mathcal{Z}}] = p^{multi}_{\mathcal{Z}}$ has a sharp upper bound given by*

$$
p^{multi}_{\mathcal{Z}} \leq 1 - \left(\underline{p}_{\mathcal{Z}_0} + \underline{p}_{\mathcal{Z}_1} - \underline{p}_{\mathcal{Z}}\right)
$$

where $\mathcal{Z}_1 = \{0, 1\}$ or $\mathcal{Z}_2 = \{1, 2\}$.

Theorem 2 (Test for multimarginals)**.** *The identified set for* $p_{\mathcal{Z}}^{multi}$ *contains* 0 *if and only if* $\int_0^1 \psi^-(y,x,z)dz \leq f_{Y^0,X^0}(y,x)$ *for all* $x, y \in Supp\{X,Y\}$.

Proof of Theorem 2. ($\Rightarrow$): We first want to show that if $p_{\mathcal{Z}}^{multi} = 0$, then $\int_0^1 \psi^-(y,x,z)dz \leq f_{Y^0,X^0}(y,x)$ for all x, y. By way of contradiction, suppose $\int_{\mathcal{Z}} \psi^-(y^*,x^*,z)dz > f_{Y_0,X_0}(y^*,x^*)$ for some x^*, y^*.

We must first translate the inequality into probability terms. We do this by showing the inequality (1) holds in a neighborhood of x^*, y^*; (2) holds when integrating ψ over regions of $\mathcal{Z}$ where it is negative; and (3) holds when approximating the inequality using a finite collection of intervals.

First, by Assumption 4 and compactness of $\mathcal{Z}$, the inequality holds for a neighborhood around x^*, y^*. Let $\Delta(\delta) = \{(x,y) : |x - x^*| < \delta, |y - y^*| < \delta\}$ be a square region on the outcome surface with sides of length 2δ. Then for some $\bar{\delta} > 0$,

$$\int_{\Delta(\delta)} \int_{\mathcal{Z}} \psi^-(y,x,z)dzdxdy > \int_{\Delta(\delta)} f_{Y^0,X^0}(y,x)dxdy$$

for any $0 < \delta \leq \bar{\delta}$. Let $\varepsilon_0 = \int_{\Delta(\delta)} \int_{\mathcal{Z}} \psi^-(y,x,z)dzdxdy - \int_{\Delta(\delta)} f_{Y^0,X^0}(y,x)dxdy > 0$.

Second, we change the integral of ψ^- over $\mathcal{Z}$ to an integral of ψ over regions where it is negative. Formally, let $\mathcal{Z}^*(\delta) = \{z \in \mathcal{Z} : \sup_{\Delta(\delta)} \psi(y,x,z) \leq 0\}$ be the instrument values where the density is not increasing in z on the entire square $\Delta(\delta)$. Because ψ is continuous, ψ^- is continuous in y and x. Thus, $\forall \varepsilon > 0$, $\exists \delta$ small enough so that the density is not decreasing too much outside $\mathcal{Z}^*(\delta)$, i.e., $\psi^-(y,x,z) < \varepsilon$ for all $z \in \mathcal{Z} \backslash \mathcal{Z}^*(\delta)$. Correspondingly,

$$\begin{aligned}\int_{\Delta(\delta)} \int_{\mathcal{Z}} \psi^-(y,x,z)dzdxdy &= \int_{\Delta(\delta)} \left(\int_{\mathcal{Z}^*(\delta)} \psi^-(y,x,z)dzdxdy + \int_{\mathcal{Z} \backslash \mathcal{Z}^*(\delta)} \psi^-(y,x,z)dzdxdy \right) \\ &< \int_{\Delta(\delta)} \left(\int_{\mathcal{Z}^*(\delta)} \psi^-(y,x,z)dzdxdy + \varepsilon \right)\end{aligned}$$

So,

$$\begin{aligned}-\int_{\Delta(\delta)} \int_{\mathcal{Z}^*(\delta)} \psi(y,x,z)dzdxdy &= \int_{\Delta(\delta)} \int_{\mathcal{Z}^*(\delta)} \psi^-(y,x,z)dzdxdy \\ &> \int_{\Delta(\delta)} \int_{\mathcal{Z}} \psi^-(y,x,z)dzdxdy - 4\delta^2\varepsilon \\ &> \int_{\Delta(\delta)} f_{Y^0,X^0}(y,x)dxdy\end{aligned}$$

if ε is small enough, i.e. $4\delta^2\varepsilon < \varepsilon_0$. Let $\varepsilon_1 = -\int_{\Delta(\delta)} \int_{\mathcal{Z}^*(\delta)} \psi(y,x,z)dzdxdy - \int_{\Delta(\delta)} f_{Y^0,X^0}(y,x)dxdy >$

0.

Third, we apply Littlewood's first principle and approximate $\mathcal{Z}^*(\delta)$ with a finite number of open intervals, i.e. $\forall\varepsilon > 0$, $\exists S = \cup_k \{(z_k, z_{k+1})\}_k$ such that $\Pr[Z \in \mathcal{Z}^*(\delta)\Delta S] < \varepsilon$. Since $\psi(x, y, z)$ is continuous in z on $\mathcal{Z}$, it is also bounded between $[-M, M]$. So,

$$\begin{aligned}
-\int_{\Delta(\delta)}\int_{\mathcal{Z}^*(\delta)} \psi(y,x,z)dzdxdy = &-\int_{\Delta(\delta)}\int_{S} \psi(y,x,z)dzdxdy \\
&-\int_{\Delta(\delta)}\int_{\mathcal{Z}^*(\delta)\backslash S} \psi(y,x,z)dzdxdy \\
&+\int_{\Delta(\delta)}\int_{S\backslash Z^*(\delta)} \psi(y,x,z)dzdxdy \\
<&-\int_{\Delta(\delta)}\int_{S} \psi(y,x,z)dzdxdy + 4\delta^2 M\varepsilon
\end{aligned}$$

So,

$$\begin{aligned}
-\int_{\Delta(\delta)}\int_{S} \psi(y,x,z)dzdxdy &> -\int_{\Delta(\delta)}\int_{\mathcal{Z}^*(\delta)} \psi(y,x,z)dzdxdy - 4\delta^2 M\varepsilon \\
&> \int_{\Delta(\delta)} f_{Y^0,X^0}(y,x)dxdy
\end{aligned}$$

if ε is small enough, i.e. $4\delta^2 M\varepsilon < \varepsilon_1$. Integrating the density derivative over the box on the left-hand side forms a probability measure,

$$\begin{aligned}
\sum_k \int_{\Delta(\delta)}\int_{z_k}^{z_{k+1}} \psi(y,x,z)dzdxdy &= \sum_k \Pr[(Y,X) \in \Delta(\delta)\,|Z = z_k] - \Pr[(Y,X) \in \Delta(\delta)\,|Z = z_{k+1}] \\
&= \sum_k \Pr[(Y^{z_k}, X^{z_k}) \in \Delta(\delta)] - \Pr[(Y^{z_{k+1}}, X^{z_{k+1}}) \in \Delta(\delta)]
\end{aligned}$$

Thus, the contradiction hypothesis implies that the sum of changes in likelihood of potential outcomes being in the square exceeds the baseline likelihood of potential outcomes being in the square,

$$\Pr\left[\left(Y^0, X^0\right) \in \Delta(\delta)\right] < \sum_k \Pr[(Y^{z_k}, X^{z_k}) \in \Delta(\delta)] - \Pr[(Y^{z_{k+1}}, X^{z_{k+1}}) \in \Delta(\delta)].$$

We will show that this is inconsistent with there being almost no multimarginals.

Consider $z < z'$. Monotonicity implies that either (1) $X^z = X^{z'}$ or (2) $X^z < X^{z'}$. If $X^z = X^{z'}$ and $(X^z, Y^z) \in \Delta(\delta)$, then exclusion implies that $\left(X^{z'}, Y^{z'}\right) \in \Delta(\delta)$. The contrapositive of this implication is that if $\left(X^{z'}, Y^{z'}\right) \notin \Delta(\delta)$, then $X^z < X^{z'}$.

Thus,

$$\begin{aligned}\Pr\left[\left(X^{z_k},Y^{z_k}\right)\in\Delta\left(\delta\right)\right]&=\Pr\left[\left(X^{z_k},Y^{z_k}\right)\in\Delta\left(\delta\right),X^{z_k}=X^{z_{k+1}}\right]+\Pr\left[\left(X^{z_k},Y^{z_k}\right)\in\Delta\left(\delta\right),X^{z_k}<X^{z_{k+1}}\right]\\&\leq\Pr\left[\left(X^{z_{k+1}},Y^{z_{k+1}}\right)\in\Delta\left(\delta\right)\right]+\Pr\left[\left(X^{z_k},Y^{z_k}\right)\in\Delta\left(\delta\right),X^{z_k}<X^{z_{k+1}}\right]\end{aligned}$$

So rearranging and substituting yields,

$$\Pr\left[\left(Y^0,X^0\right)\in\Delta\left(\delta\right)\right]<\sum_k\Pr\left[\left(X^{z_k},Y^{z_k}\right)\in\Delta\left(\delta\right),X^{z_k}<X^{z_{k+1}}\right]$$

If the events, $\{(X^{z_k},Y^{z_k})\in\Delta(\delta),X^{z_k}<X^{z_{k+1}}\}$ are not disjoint then there is some k and k' such that $0<Pr\left[X^{z_k}<X^{z_{k+1}},X^{z_{k'}}<X^{z_{k'+1}}\right]<p_{\mathcal{Z}}^{multi}$. On the other hand, if they are disjoint then the sum of probabilities is the probability of the union of the disjoint events so that

$$\Pr\left[\left(Y^0,X^0\right)\in\Delta\left(\delta\right)\right]<\Pr\left[\left(X^z,Y^z\right)\in\Delta\left(\delta\right),X^z<X^{z'}\quad\text{for some } z<z'\right],$$

This then implies that $\Pr\left[\left(Y^0,X^0\right)\notin\Delta\left(\delta\right),\left(X^z,Y^z\right)\in\Delta\left(\delta\right),X^z<X^{z'}\quad\text{for some } z<z'\right]>0$, which implies that $p_{\mathcal{Z}}^{multi}=\Pr\left[X^0<X^z<X^{z'}\right]>0$.

($\Leftarrow$): Let $f_{Z^*}(z^*)=\int_{\mathcal{X}}\int_{\mathcal{Y}}\psi^-(x,y,z^*)dxdy$, and define a random variable Z^* distributed according to f_{Z^*} with a point mass at 3 (a number outside the support of $\mathcal{Z}$), so $\Pr\left[Z^*=3\right]=1-\int_{\mathcal{X}}\int_{\mathcal{Y}}\psi^-(x,y,z^*)dxdy$. (Note that $\int_{\mathcal{X}}\int_{\mathcal{Y}}\psi^-(x,y,z^*)dxdy=\int_{\mathcal{X}}\int_{\mathcal{Y}}\psi^+(x,y,z^*)dxdy$ because $\int_{\mathcal{X}}\int_{\mathcal{Y}}\psi^+(x,y,z^*)dxdy-\int_{\mathcal{X}}\int_{\mathcal{Y}}\psi^-(x,y,z^*)dxdy=\int_{\mathcal{X}}\int_{\mathcal{Y}}\psi(x,y,z^*)dxdy=\frac{\partial}{\partial z}\int_{\mathcal{X}}\int_{\mathcal{Y}}f\left(x,y|z\right)dxdy=0$.) Next, define

$$\varphi\left(x,y|z\right)=\frac{\psi^+\left(y,x,z\right)}{f_{Z^*}\left(z\right)}$$

And let $\left(U^X,U^Y\right)|Z^*=z$ be distributed according to $\varphi\left(y,x|z\right)$. Finally, let $\eta\sim\text{Uniform}\left[0,1\right]$ and $\varepsilon\sim\text{Uniform}\left[0,1\right]$ be independent. Our DGP is given by

$$\begin{aligned}X\left(z;\eta,Z^*\right)&=x_0\left(\eta\right)\mathbf{1}\left[z\leq Z^*\right]+x_1\left(\eta,Z^*\right)\mathbf{1}\left[z>Z^*\right]\\Y\left(x;\eta,\varepsilon,Z^*\right)&=\begin{cases}y_0\left(\varepsilon,\eta\right)&Z^*=3\\m\left(\varepsilon,\eta,Z^*\right)\left(x-x_0\left(\eta\right)\right)+y_0\left(\varepsilon,\eta,Z^*\right)&Z^*<3\end{cases}\end{aligned}$$

where

1. $x_0\left(\eta\right)=F_{X|Z}^{-1}\left(\eta|0\right)$
2. $x_1\left(\eta,Z^*\right)=F_{U^X|Z}^{-1}\left(\eta|Z^*\right)$

3. $y_0(\varepsilon,\eta) = F^{-1}_{Y|X,Z}(\varepsilon|x_0(\eta),0)$

4. $y_1(\varepsilon,\eta,Z^*) = F^{-1}_{U^Y|U^X,Z^*}(\varepsilon|x_1(\eta,Z^*),Z^*)$

5. $m(\varepsilon,\eta,Z^*) = \frac{y_1(\varepsilon,\eta,Z^*)-y_0(\varepsilon,\eta)}{x_1(\eta,Z^*)-x_0(\eta)}$.

Note: I think that the dgp above doesn't quite work. The following is a restatement of the one in my proof in a similar form to the one above:

$$X(z;\eta,Z^*) = x_0(\eta)\mathbf{1}[z \leq Z^*] + x_1(\eta,Z^*)\mathbf{1}[z > Z^*]$$
$$Y(x;\eta,\varepsilon,Z^*) = m(\varepsilon,\eta,Z^*)(x - x_0(\eta)) + y_0(\varepsilon,\eta)$$

where

1. $x_0(\eta) = F^{-1}_{X|Z}(\eta|0)$

2. $y_0(\varepsilon,\eta) = F^{-1}_{Y|X,Z}(\varepsilon|x_0(\eta),0)$

3. $Z^* \mid x_0(\eta), y_0(\varepsilon,\eta)$ drawn from distribution with density $\frac{\psi^-(x_0,y_0,z)}{f_{X_0,Y_0}(x_0,y_0)}$ for $z \in [0,1)$ and a point mass of $1 - \frac{\int_0^1 \psi^-(y_0,x_0,z)dz}{f_{X_0,Y_0}(x_0,y_0)}$ at 1 (or some value above one)

4. $x_1(\eta,Z^*) = F^{-1}_{U^X_+|Z^*}(\tilde{\eta}|Z^*)$ where $\left(U^X_+,U^Y_+\right)|Z^* = z$ is distributed according to $\varphi(y,x|z)$ and $\tilde{\eta} = F_{U^X_-|Z^*}(x_0(\eta) \mid Z^*)$ where $\left(U^X_-,U^Y_-\right)|Z^* = z$ is distributed according to $\tilde{\varphi}(y,x|z) := \frac{\psi^-(y,x,z)}{f_{Z^*}(z)}$

5. $y_1(\varepsilon,\eta,Z^*) = F^{-1}_{U^Y_+|U^X_+,Z^*}(\varepsilon|x_1(\eta,Z^*),Z^*)$

6. $m(\varepsilon,\eta,Z^*) = \frac{y_1(\varepsilon,\eta,Z^*)-y_0(\varepsilon,\eta)}{x_1(\eta,Z^*)-x_0(\eta)}$.

□

Corollary 1. *Let $d(x;\varepsilon)$ be continuous in x on $[\underline{x},\bar{x}]$ where $\underline{x}$ is a lower bound for $x_0(\eta)$, and $\bar{x}$ is an upper bound for $x_1(\eta,Z^*)$. If $p^{multi}_{\mathcal{Z}} = 0$, then $d(x;\varepsilon)$ is in the identified set of $\frac{\partial g^{(\delta)}(x;\varepsilon)}{\partial x^\delta}$ for any $\delta \geq 2$.*

Proof of Corollary 1. We prove the case where $\delta = 2$; the cases for $\delta > 2$ follow immediately. We must show that there exists a function g such that $\frac{\partial g^{(2)}(x;\varepsilon)}{\partial x^2} = p$ and $g(x_0(\eta),\varepsilon) = y_0(\varepsilon,\eta)$ and $g(x_1(\eta,Z^*)) = y_1(\varepsilon,\eta,Z^*)$. To simplify notation, we omit notation indexing heterogeneity, η,ε,Z^*.

Let $a \in [\underline{x},\bar{x}]$ and define $\tilde{d}(x,C) = \int_a^x d(x)\,dx + C$ and $\tilde{\tilde{d}}(x) = \int_a^x \tilde{d}(x)\,dx + D = \int_a^x \int_a^x d(x) + (x-a)C + D$. By the Fundamental Theorem of Calculus, $\tilde{d}$ and $\tilde{\tilde{d}}$ are continuous and $\tilde{\tilde{d}}' = \tilde{d}$ and $\tilde{\tilde{d}}'' = d$. Let $\tilde{y}_0 = \int_a^{x_0}\int_a^{x_0} d(x)\,dx$ and $\tilde{y}_1 = \int_a^{x_1}\int_a^{x_1} d(x)\,dx$. Then let $C = \frac{(y_1-y_0)-(\tilde{y}_1-\tilde{y}_0)}{x_1-x_0}$ and $D = y_0 - \tilde{y}_0 - (x_0 - a)C$. □

Corollary 2. *Let d be defined as in Corollary 1. If $p_{\mathcal{Z}}^{multi} = 0$, then d is in the identified set of derivatives of order $\delta \geq 2$ of the conditional average structural function* $\frac{\partial^{(\delta)}}{\partial x^{\delta}} \mathbf{E}_{\varepsilon} \left[g\left(x, \varepsilon\right) | W \right]$ *and the conditional quantile structural function* $\frac{\partial^{(\delta)}}{\partial x^{\delta}} \inf \left\{ y : \mathbf{E}_{\varepsilon} \left[\mathbf{1} \left[g\left(x, \varepsilon\right) \leq yt \right] | W \right] \geq \tau \right\}$ *for any quantile τ.*