COMPLEMENTARITY AND IDENTIFICATION

Tᴀᴛᴇ Tᴡɪɴᴀᴍ
University of Washington

This paper examines the identification power of assumptions that formalize the notion of complementarity in the context of a nonparametric bounds analysis of treatment response. I extend the literature on partial identification via shape restrictions by exploiting cross-dimensional restrictions on treatment response when treatments are multidimensional; the assumption of *supermodularity* can strengthen bounds on average treatment effects in studies of policy complementarity. This restriction can be combined with a statistical independence assumption to derive improved bounds on treatment effect distributions, aiding in the evaluation of complex randomized controlled trials. Complementarities arising from treatment effect heterogeneity can be incorporated through *supermodular instrumental variables* to strengthen identification in studies with one or multiple treatments. An application examining the long-run impact of zoning on the evolution of urban spatial structure illustrates the value of the proposed identification methods.

1. INTRODUCTION

Complementarities arise naturally in many economic problems, often manifesting as policy interactions or treatment effect heterogeneity among observed subgroups of a population. This paper examines how assumptions that formalize the notion of complementarity can aid in the identification of treatment effects. The analysis employs a nonparametric bounds approach, where identification is driven by qualitative restrictions rooted in economic theory or empirical evidence rather than strong functional form or unconfoundedness assumptions. This approach will yield interval estimates of parameters of interest; however, informative bounds are often preferable to precise (but wrong) estimates obtained under incorrect assumptions. Partial identification tools have been fruitfully applied to a wide range of empirical problems.[1]

In particular, I explore the identification power yielded by assuming that individual treatment response functions exhibit *supermodularity* when treatments are multidimensional. This assumption allows one to construct more informative bounds in studies of policy complementarity, which are typically stymied by the absence of pseudo-experimental variation in the assignment of multiple

1154

treatments. Complementarities arising from interactions between treatment effects and observable covariates can be formalized as *supermodular instrumental variables* to improve bounds on average treatment effects. This novel instrumental variable approach is broadly applicable to studies with one or multiple treatments. Complementarity is frequently invoked in economics, but studies of its identification power have been limited to very specific contexts. This paper develops general results applicable to program evaluation in a wide range of empirical situations, and their value is illustrated by an empirical application on the long-run effects of zoning on urban spatial structure.

Typically, empirical studies seek to estimate the effect of a single treatment on one or more outcome variables. However, the effect of a treatment may vary substantially with the value of other (endogenously-determined) treatment variables. When policymakers have multiple tools at their disposal, understanding how different policies enhance or offset each other is crucial. If the positive impact of some policy intervention is substantially larger when combined with a second (costly) intervention, a measure of the magnitude of this difference is necessary for a proper cost-benefit analysis. The supermodularity and submodularity assumptions proposed here can aid in quantifying how policy impacts differ with the associated policy environment.

For example, unemployment relief is a multidimensional policy, involving a choice of both potential benefit duration and the wage replacement rate. Lalive, Van Ours, and Zweimüller (2006) show both theoretically and empirically that these two dimensions are complementary, with simultaneous increases in both the replacement rate and potential benefit duration leading to an increase in unemployment duration substantially larger than the sum of the effects measured individually for particular subgroups. The Lalive et al. study exploits variation in both dimensions of unemployment relief that has the characteristics of a natural experiment, but such opportunities are very rare. Pseudo-experimental variation along multiple policy dimensions is far less common than similar variation in individual policies. This has arguably led to the overwhelming focus on the effects of policies in isolation. The partial identification tools developed here, which are applicable in the absence of any unusual pseudo-experimental policy variation, should enhance the ability of researchers to measure treatment effect heterogeneity due to policy complementarities in a wide range of contexts.[2]

Relatedly, responses to a treatment may differ among subpopulations defined by observable covariates. Many recent experimental studies have discussed the importance of treatment effect heterogeneity between subgroups (Bitler, Gelbach, and Hoynes, 2006, 2008, 2014; Djebbari and Smith, 2008; Feller and Holmes, 2009). Qualitative information about such treatment effect heterogeneity leads naturally to supermodular instrumental variables, which can help narrow the bounds on average treatment effects in the same manner as a traditional instrumental variable or a monotone instrumental variable.[3] Supermodular instrumental variables can be applied in the case of a single treatment or multiple treatments, making them a potentially valuable addition to the range of identifying assumptions available to applied researchers.

While the bulk of the paper focuses on identification using non-experimental data, the assumptions developed in this paper can be applied in the evaluation of complex randomized controlled trials (RCTs) involving multiple treatments. The structural assumptions introduced can be used to obtain stronger bounds on the (partially identified) distribution of treatment effects. Since average treatment effects are identified in this context, the supermodularity or submodularity of average effects can be established, and this can be used to provide some justification for the stronger structural assumptions. Similarly, the validity of supermodular instrumental variable assumptions can be established and used to justify stronger *quantile supermodular instrumental variable* assumptions, which can also be applied in the case of a single treatment.

The literature on partial identification is extensive.[4] Many of the contributions of Charles Manski and coauthors are relevant to the results developed below and are reviewed as appropriate. The literature on complementarity and identification is relatively small. Molinari and Rosen (2008) connect supermodularity to identification in the context of game estimation. They show that the approach of Aradillas-Lopez and Tamer (2008) applies to games with supermodular payoff functions. Eeckhout and Kircher (2011) find that they cannot identify (using wage data alone) whether or not the technology of a firm is supermodular, i.e., whether or not more productive workers sort towards more productive jobs. Graham, Imbens, and Ridder (2014) analyze how reallocations of indivisible heterogeneous inputs across production units (leaving a potentially complementary input fixed) may affect average output. They discuss identification and estimation of the effects of a variety of correlated matching rules. Lazzati (2015) uses monotone comparative statics to partially identify treatment response in the presence of endogenous social interactions. The shape restrictions proposed have previously been used in the context of estimation to improve efficiency; Beresteanu (2005, 2007) considers the efficiency gains from imposing a variety of restrictions, including supermodularity and submodularity.

2. NOTATION AND SETUP

Individuals are drawn from a population I. The set I, the Borel σ-algebra of subsets of I denoted by $\mathcal{I}$, and the probability measure P together form a probability space $(I, \mathcal{I}, P)$. Every individual $i \in I$ is associated with a vector of covariates $x_i \in X$ and a vector of realized treatments $z_i \in T$, where T is the treatment set.[5] Since I focus on the identification of treatment effects in the presence of multiple treatments, the following assumption is imposed on the structure of the treatment space:

Assumption. The treatment space T is such that

- $T \subseteq \mathbb{R}^L$ with $L \in \mathbb{N}$,

- T is partially ordered under the product order,[6] and

- T is a nonempty lattice.

The lattice assumption means that, for any $t_1, t_2 \in T$, T contains the join (least upper bound) of t_1 and t_2, denoted by $t_1 \vee t_2$, as well as the meet (greatest lower bound) of t_1 and t_2, denoted by $t_1 \wedge t_2$.[7] Examples of lattices include $\mathbb{R}^2$, $\mathbb{Z} \times \mathbb{R}$, and $\{0, 1\}^n$ for $n \in \mathbb{N}$. An element t_1 of a lattice T is the top (bottom) of T if $t_2 \leq t_1$ $(t_1 \leq t_2)$ for all $t_2 \in T$; if t_1 is not the top or bottom, it is in the interior. If the top (or bottom) of a lattice exists, it is unique. A subset $S \subseteq T$ is a sublattice of T if, for any $t_1, t_2 \in S$, S contains the meet and join of t_1 and t_2 in T. The advantage of the lattice assumption is the notational clarity it provides when employing supermodularity and submodularity assumptions. In this paper, I restrict attention to discrete treatments, as these are most commonly encountered in practice. Dimensions of the treatment may be binary or multivalued (Cattaneo 2010).

Every individual i is associated with a (measurable) response function $y_i(\cdot)$: $T \to Y \in \mathbb{R}$ mapping treatments into outcomes $y_i(t)$.[8] $z_i \in T$ is the treatment that i actually receives, so $y_i(z_i)$ is individual i's realized outcome, $\{y_i(t)\}_{t \neq z_i}$ are individual i's counterfactual outcomes, and $\{y_i(t)\}_{t \in T}$ are individual i's potential outcomes.[9] Throughout, I assume that there exist $\underline{K}, \overline{K} \in \mathbb{R}$ such that $\underline{K} \leq y(t) \leq \overline{K}$ for all t; these are global bounds on response functions. In the absence of these global bounds, the results below will generally be uninformative. All well-defined expectations are assumed to exist.[10]

3. SHAPE RESTRICTIONS

In this section, I explore the identifying power of shape restrictions that formalize complementarity and substitutability, with an emphasis on the identification of average treatment effects. Shape restrictions proposed in the previous literature are reviewed before moving on to the novel restrictions proposed here. Using these assumptions, I derive bounds on average treatment effects for both simple and complex treatment spaces.

Manski (1989) introduced the no-assumption bounds on $E[y(t)]$. The no-assumption upper bound is the average of $E[y(t) \mid z = t]$ and the global upper bound $\overline{K}$, weighted, respectively, by $P(z = t)$ and $P(z \neq t)$; likewise for the lower bound. Since they are typically wide, research has focused on other credible assumptions that yield additional identifying power.

Manski (1997) studied the identification power of assumptions on the shape of individual response functions; in particular, he considered restricting response functions to be monotone, semimonotone, or concave-monotone. Semimonotone treatment response (SMTR), which I employ below, requires response functions to be weakly increasing in t. SMTR has the same identification power regardless of whether $T \subseteq \mathbb{R}$ or $T \subseteq \mathbb{R}^L$ for $L > 1$, except that in the latter case, it is possible that $t_1 \parallel t_2$. Bhattacharya, Shaikh, and Vytlacil (2008) derives bounds using SMTR without assuming a particular direction of monotonicity. Tsunao and Usui (2014) study the identification power of concave-monotone treatment response combined with monotone treatment selection (discussed in Section 4).

SMTR is a within-dimension restriction on the response functions. Additional identification power can be obtained from cross-dimension restrictions, where the marginal effect of a change in some dimensions of the treatment variable depends on the values of the other dimensions:

Assumption SPM (Supermodularity)**.** Response functions are *supermodular* on a sublattice $S \subseteq T$ if, for all $t_1, t_2 \in S$,

$$y(t_2) + y(t_1) \leq y(t_1 \vee t_2) + y(t_1 \wedge t_2). \tag{3.1}$$

Assumption SBM (Submodularity)**.** Response functions are *submodular* on a sublattice $S \subseteq T$ if, for all $t_1, t_2 \in S$,

$$y(t_2) + y(t_1) \geq y(t_1 \vee t_2) + y(t_1 \wedge t_2). \tag{3.2}$$

SPM is a formalization of the notion of complementarity. If two dimensions of a treatment $t = (t^1, t^2)$ are complementary, then the magnitude of the change in the response variable due to an increase in the first dimension t^1 is increasing with t^2. Thus, each component of the treatment amplifies the marginal effect of the other component. In the case of a linear model

$$y = \alpha + \beta t^1 + \delta t^2 + \gamma t^1 t^2, \tag{3.3}$$

supermodularity is equivalent to the sign restriction $\gamma \geq 0$. More generally, if y is a (sufficiently smooth) nonlinear function of the treatment, supermodularity can be interpreted as a nonnegativity restriction on the cross-partial derivative $\frac{\partial^2 y}{\partial t^1 \partial t^2}$. SBM is a formalization of substitutability, the case where elements of the treatment may mitigate each others effects. In the linear model above, submodularity is equivalent to the sign restriction $\gamma \leq 0$. If both supermodularity and submodularity hold, response functions are said to be *modular*. Since assumptions SPM and SBM can be applied on sublattices of T, it is possible to allow some dimensions of a treatment to be complements while those same dimensions are substitutes with other dimensions.

Neumark and Wascher (2011) provide an example of policy complementarity in a study on the interaction between the Earned Income Tax Credit (EITC) and the minimum wage. They find that a higher minimum wage enhances the positive effect of the EITC on the labor supply of single mothers; they find the opposite effect for childless individuals, suggesting a crowding-out effect. These findings suggest that assumptions SPM and SBM, respectively, for each subgroup, could be applied in other studies on how the effect of minimum wage changes are influenced by the EITC or similar programs. Another naturally multidimensional policy is zoning. Zoning laws typically regulate many aspects of the built environment; most broadly, they regulate both what types of uses are allowed (commercial, industrial, etc.) and how densely land can be developed (lot coverage of buildings, maximum height, etc.). The effects of specific zoning policies may vary with the overall policy bundle, which I explore further in the empirical application in Section 6.

The amount one can learn about $E[y(t_1)]$ or $E[y(t_1) - y(t_2)]$ from the data alone depends on $P(z = t_1)$ and, in the latter case, $P(z = t_2)$. If $P(z = t_1)$ is small, the data are practically uninformative about $E[y(t_1)]$.[11] Thus, the researcher faces a trade-off where richer treatment spaces (which entail a larger number of treatments) allow for more interesting questions but generally lead to less precise answers. Adding "nuisance" dimensions to the treatment space that allow for the application of additional SPM or SBM assumptions will generally not aid in the identification of treatment effects of interest.

In Propositions 1 and 2, I show how SPM and SBM can be used to compute bounds on the expectations of average treatment effects. In general, these bounds will improve upon the no-assumption bounds in the case of multidimensional treatments; with only a single treatment, SPM and SBM have no identifying power. The simplest nontrivial lattice treatment space is $T = \{(0,0), (1,0), (0,1), (1,1)\}$, which corresponds to two binary treatments. The following result shows the implications of supermodularity for identification on this simple treatment space:

PROPOSITION 1. *Assume that* $T = \{(0,0), (1,0), (0,1), (1,1)\}$. *Assume that SPM holds on* T. *Then, the bounds*

$$E[y(1,0) \mid z = (1,0)]P(z = (1,0)) + \underline{K}P(z \neq (1,0))$$
$$- E[y(0,0) \mid z = (0,0)]P(z = (0,0)) - \overline{K}P(z \neq (0,0))$$
$$\leq E[y(1,0) - y(0,0)]$$
$$\leq E[y(1,1) \mid z = (1,1)]P(z = (1,1))$$
$$+ E[y(1,0) \mid z = (1,0)]P(z = (1,0))$$
$$+ \overline{K}P(z \in \{(0,0), (0,1)\}) - E[y(0,1) \mid z = (0,1)]P(z = (0,1))$$
$$- E[y(0,0) \mid z = (0,0)]P(z = (0,0)) - \underline{K}P(z \in \{(1,0), (1,1)\}) \quad \textbf{(3.4)}$$

and

$$E[y(1,1) \mid z = (1,1)]P(z = (1,1)) + E[y(1,0) \mid z = (1,0)]P(z = (1,0))$$
$$+ \underline{K}P(z \in \{(0,0), (0,1)\}) - E[y(0,1) \mid z = (0,1)]P(z = (0,1))$$
$$- E[y(0,0) \mid z = (0,0)]P(z = (0,0)) - \overline{K}P(z \in \{(1,0), (1,1)\})$$
$$\leq E[y(1,1) - y(0,1)]$$
$$\leq E[y(1,1) \mid z = (1,1)]P(z = (1,1)) + \overline{K}P(z \neq (1,1))$$
$$- E[y(0,1) \mid z = (0,1)]P(z = (0,1)) - \underline{K}P(z \neq (0,1)) \quad \textbf{(3.5)}$$

are sharp.[12] *The no-assumption bounds remain sharp for* $E[y(1,1) - y(0,0)]$, $E[y(1,0) - y(0,1)]$, $E[y(0,1) - y(1,0)]$, *and each average potential outcome* $E[y(\cdot)]$ *defined on* T.

Proof of Proposition 1. First, I show that SPM does not improve upon the no-assumption bounds on potential outcomes. SPM implies that

$$y(1,0) + y(0,1) \leq y(1,1) + y(0,0).$$

For each i, exactly one of these outcomes is observed. The unobserved terms may take any value in $[\underline{K}, \overline{K}]$. Consider bounding the average potential outcome $E[y(1,0)]$. When $z \neq (1,0)$, there are three cases to consider. If $z = (1,1)$, then SPM implies

$$y(1,0) \leq \overline{K} \leq y(1,1) + \overline{K} - \underline{K}.$$

If $z = (0,1)$, then

$$y(1,0) \leq \overline{K} \leq \overline{K} + \overline{K} - y(0,1).$$

If $z = (0,0)$, then

$$y(1,0) \leq \overline{K} \leq \overline{K} + y(0,0) - \underline{K}.$$

All three of these inequalities are implied without SPM, so the assumption is nonbinding. Thus, it follows that

$$y(1,0) \in \begin{cases} \{y(1,0)\} & \text{if } z = (1,0) \\ [\underline{K}, \overline{K}] & \text{if } z \in \{(0,0), (0,1), (1,1)\}. \end{cases}$$

Taking expectations yields the no-assumption bounds. A similar argument applies to the other elements of T.

The SPM inequality does permit strengthened identification results for treatment effects. In the no-assumption case, if $z = (1,1)$ or $z = (0,1)$, then $y(1,0) - y(0,0) \in [\underline{K} - \overline{K}, \overline{K} - \underline{K}]$. Under SPM, the fact that we observe one of $\{y(1,1), y(0,1)\}$ allows us to further reduce this upper bound. Sharp bounds for the treatment effects $y(1,0) - y(0,0)$, $y(1,1) - y(0,1)$, and $y(1,1) - y(0,0)$ are given below:

$$y(1,0) - y(0,0) \in \begin{cases} [\underline{K} - y(z), \overline{K} - y(z)] & \text{if } z = (0,0) \\ [y(z) - \overline{K}, y(z) - \underline{K}] & \text{if } z = (1,0) \\ [\underline{K} - \overline{K}, \overline{K} - y(z)] & \text{if } z = (0,1) \\ [\underline{K} - \overline{K}, y(z) - \underline{K}] & \text{if } z = (1,1) \end{cases} \tag{3.6}$$

$$y(1,1) - y(0,1) \in \begin{cases} [\underline{K} - y(z), \overline{K} - \underline{K}] & \text{if } z = (0,0) \\ [y(z) - \overline{K}, \overline{K} - \underline{K}] & \text{if } z = (1,0) \\ [\underline{K} - y(z), \overline{K} - y(z)] & \text{if } z = (0,1) \\ [y(z) - \overline{K}, y(z) - \underline{K}] & \text{if } z = (1,1) \end{cases} \tag{3.7}$$

$$y(1,1) - y(0,0) \in \begin{cases} [\underline{K} - y(z), \overline{K} - y(z)] & \text{if } z = (0,0) \\ [\underline{K} - \overline{K}, \overline{K} - \underline{K}] & \text{if } z = (1,0) \\ [\underline{K} - \overline{K}, \overline{K} - \underline{K}] & \text{if } z = (0,1) \\ [y(z) - \overline{K}, y(z) - \underline{K}] & \text{if } z = (1,1) \end{cases} \tag{3.8}$$

Taking expectations in equations (3.6) and (3.7) yields the bounds in (3.4) and (3.5), respectively. Equation (3.8) shows that the no-assumption bounds remain

sharp for E[$y(1, 1) - y(0, 0)$]. A similar analysis establishes that the no-assumption bounds remain sharp for E[$y(0, 1) - y(1, 0)$] and E[$y(1, 0) - y(0, 1)$].   ∎

In Proposition 1, assumption SPM improves the upper bound on E[$y(1, 0) - y(0, 0)$] and the lower bound on E[$y(1, 1) - y(0, 1)$] by establishing a monotonicity relationship between the two treatment effects. Similar bounds for the treatment effects E[$y(0, 1) - y(0, 0)$] and E[$y(1, 1) - y(1, 0)$] can be obtained by permuting the order of the treatments and applying the result. As stated in the proposition, the no-assumption bounds on E[$y(1, 1) - y(0, 0)$], E[$y(1, 0) - y(0, 1)$], and E[$y(0, 1) - y(1, 0)$] remain sharp under SPM. In the special case where y is bounded between zero and one, SPM can establish that E[$y(1, 0) - y(0, 0)$] $\in [-1, 0]$ or E[$y(1, 1) - y(0, 1)$] $\in [0, 1]$ if the observed expectations in (3.4) and (3.5) take certain boundary values. In general, however, SPM is not sufficient to identify the sign of a treatment effect in the absence of other assumptions.

Sharp bounds can be derived on more general treatment spaces using the same approach. For any t_1 and t_2,

$$\mathrm{E}[\, y(t_1) - y(t_2)\,] = \sum_{t_3 \in T} \mathrm{E}[\, y(t_1) - y(t_2) \mid z = t_3\,] P(z = t_3).$$

Thus, bounds can be derived for E[$y(t_1) - y(t_2)$] by bounding E[$y(t_1) - y(t_2) \mid z = t_3$] for each t_3. Bounds on E[$y(t_1) - y(t_2) \mid z = t_3$] will differ depending on the data, the ordering of the three treatments, and whether SPM and/or SBM hold on any sublattices containing the treatments. To capture all of the information implied by the data and assumptions, one must examine every sublattice containing t_1, t_2, and t_3, determine which of SPM and SBM hold on the sublattice, and derive the implied bounds given the ordering of t_1, t_2, and t_3.

To this end, I formally define a collection of sets of treatments (indexed by t_1 and t_2) which will allow for a derivation of the bounds on E[$y(t_1) - y(t_2) \mid z = t_3$] by identifying which assumptions (if any) imply tightened bounds and the form those bounds will take. Let $\{S_\gamma\}_{\gamma \in \Gamma}$ be the collection of all sublattices of T such that, for every $\gamma \in \Gamma$, $|S_\gamma| = 4$ and S_γ is not a chain. These are the minimal sublattices on which SPM and SBM can have any implications for identification. Let Γ_{t_1, t_2} denote the set of γ such that $t_1, t_2 \in S_\gamma$, and let Γ_{t_1, t_2}^{SPM} refer to the subset of Γ_{t_1, t_2} for which SPM (but not SBM) holds on S_γ; likewise, let Γ_{t_1, t_2}^{SBM} refer to the subset where only SBM holds on S_γ. Let Γ_{t_1, t_2}^{MOD} refer to the subset where both SPM and SBM hold. Then, one can define the following sets:

$$\Lambda_1 = \big\{ t_3 \mid t_2 < t_1 < t_3 \text{ and either } \exists \gamma \in \Gamma_{t_1, t_2}^{MOD} \text{ s.t. } t_3 \in S_\gamma$$
$$\text{or } \exists \gamma \in \Gamma_{t_1, t_2}^{SPM}, \gamma' \in \Gamma_{t_1, t_2}^{SBM} \text{ s.t. } t_3 \in S_\gamma \cap S_{\gamma'} \big\},$$

$$\Lambda_2 = \big\{ t_3 \mid t_2 < t_1 < t_3, \nexists \gamma \in \Gamma_{t_1, t_2}^{MOD} \cup \Gamma_{t_1, t_2}^{SBM} \text{ s.t. } t_3 \in S_\gamma,$$
$$\text{and } \exists \gamma \in \Gamma_{t_1, t_2}^{SPM} \text{ s.t. } t_3 \in S_\gamma \big\},$$

$$\Lambda_3 = \big\{ t_3 \mid t_2 < t_1 < t_3, \nexists \gamma \in \Gamma_{t_1, t_2}^{MOD} \cup \Gamma_{t_1, t_2}^{SPM} \text{ s.t. } t_3 \in S_\gamma,$$
$$\text{and } \exists \gamma \in \Gamma_{t_1, t_2}^{SBM} \text{ s.t. } t_3 \in S_\gamma \big\},$$

$$\Lambda_4 = \big\{ t_3 \mid t_3 < t_2 < t_1 \text{ and either } \exists \gamma \in \Gamma_{t_1,t_2}^{MOD} \text{ s.t. } t_3 \in S_\gamma$$
$$\text{or } \exists \gamma \in \Gamma_{t_1,t_2}^{SPM}, \gamma' \in \Gamma_{t_1,t_2}^{SBM} \text{ s.t. } t_3 \in S_\gamma \cap S_{\gamma'} \big\},$$

$$\Lambda_5 = \big\{ t_3 \mid t_3 < t_2 < t_1, \nexists \gamma \in \Gamma_{t_1,t_2}^{MOD} \cup \Gamma_{t_1,t_2}^{SBM} \text{ s.t. } t_3 \in S_\gamma,$$
$$\text{and } \exists \gamma \in \Gamma_{t_1,t_2}^{SPM} \text{ s.t. } t_3 \in S_\gamma \big\},$$

$$\Lambda_6 = \big\{ t_3 \mid t_3 < t_2 < t_1, \nexists \gamma \in \Gamma_{t_1,t_2}^{MOD} \cup \Gamma_{t_1,t_2}^{SPM} \text{ s.t. } t_3 \in S_\gamma,$$
$$\text{and } \exists \gamma \in \Gamma_{t_1,t_2}^{SBM} \text{ s.t. } t_3 \in S_\gamma \big\},$$

$$\Lambda_7 = \big\{ t_3 \mid t_3 < t_1, t_3 \parallel t_2, \text{ and } \exists \gamma \in \Gamma_{t_1,t_2}^{MOD} \text{ s.t. } t_3 \in S_\gamma \big\},$$

$$\Lambda_8 = \big\{ t_3 \mid t_3 < t_1, t_3 \parallel t_2, \text{ and } \exists \gamma \in \Gamma_{t_1,t_2}^{SPM} \text{ s.t. } t_3 \in S_\gamma \big\},$$

$$\Lambda_9 = \big\{ t_3 \mid t_3 < t_1, t_3 \parallel t_2, \text{ and } \exists \gamma \in \Gamma_{t_1,t_2}^{SBM} \text{ s.t. } t_3 \in S_\gamma \big\},$$

$$\Lambda_{10} = \big\{ t_3 \mid t_2 < t_3, t_3 \parallel t_1, \text{ and } \exists \gamma \in \Gamma_{t_1,t_2}^{MOD} \text{ s.t. } t_3 \in S_\gamma \big\},$$

$$\Lambda_{11} = \big\{ t_3 \mid t_2 < t_3, t_3 \parallel t_1, \text{ and } \exists \gamma \in \Gamma_{t_1,t_2}^{SPM} \text{ s.t. } t_3 \in S_\gamma \big\},$$

$$\Lambda_{12} = \big\{ t_3 \mid t_2 < t_3, t_3 \parallel t_1, \text{ and } \exists \gamma \in \Gamma_{t_1,t_2}^{SBM} \text{ s.t. } t_3 \in S_\gamma \big\},$$

$$\Lambda_{13} = \left(\bigcup_{j=1}^{12} \Lambda_j \right)^c \setminus \{t_1, t_2\}. \tag{3.9}$$

These partition the set of sublattices of T containing t_1, t_2, and t_3 based on the ordering of the treatments and the assumptions imposed. These sets are mutually exclusive and exhaustive, so each t_3 will belong to exactly one. For an example of how the sets can be used, consider Λ_5. If $t_3 \in \Lambda_5$, then t_1, t_2, and t_3 belong to one or more sublattices on which SPM holds, but none on which SBM holds. The required ordering implies that

$$\underline{K} - y(t_3) \le y(t_1) - y(t_2)$$

and thus

$$y(t_1) - y(t_2) \in [\underline{K} - y(t_3), \overline{K} - \underline{K}],$$

which implies that

$$\underline{K} - \mathrm{E}[\, y(t_3) \mid z = t_3 \,] \le \mathrm{E}[\, y(t_1) - y(t_2) \mid z = t_3 \,] \le \overline{K} - \underline{K}.$$

The following result uses the above sets to derive sharp bounds on $\mathrm{E}[\, y(t_1) - y(t_2) \,]$ under arbitrary combinations of supermodularity and submodularity.

PROPOSITION 2. *Let $\{S_\gamma\}_{\gamma \in \Gamma}$ be the collection of all sublattices of T such that, for every $\gamma \in \Gamma$, S_γ is not a chain and $|S_\gamma| = 4$. Define $\Gamma^{SPM} \subseteq \Gamma$ to be the set of γ such that SPM holds on S_γ and SBM does not hold on S_γ iff $\gamma \in \Gamma^{SPM}$; likewise, define $\Gamma^{SBM} \subseteq \Gamma$ to be the set of γ such that SBM holds on S_γ and SPM does not hold on S_γ iff $\gamma \in \Gamma^{SBM}$. Define $\Gamma^{MOD} \subseteq \Gamma$ to be the set of γ such that SPM and SBM hold on S_γ iff $\gamma \in \Gamma^{MOD}$. Let $\Gamma_{t_1,t_2}^{SPM} \subseteq \Gamma^{SPM}$ be the set of γ such that $t_1, t_2 \in S_\gamma$ and $\gamma \in \Gamma^{SPM}$; likewise for Γ_{t_1,t_2}^{SBM} and Γ_{t_1,t_2}^{MOD}. Then, for $t_2 < t_1$,*

$[\mathrm{E}[\,y(t_1)\,|\,z=t_1\,]-\overline{K}]P(z=t_1)+[\underline{K}-\mathrm{E}[\,y(t_2)\,|\,z=t_2\,]]P(z=t_2)$

$\qquad+\displaystyle\sum_{t_3\in\bigcup_{j\in\{1,3,7,8\}}\Lambda_j}[\mathrm{E}[\,y(t_3)\,|\,z=t_3\,]-\overline{K}]P(z=t_3)$

$\qquad+\displaystyle\sum_{t_3\in\bigcup_{j\in\{4,5,10,12\}}\Lambda_j}[\underline{K}-\mathrm{E}[\,y(t_3)\,|\,z=t_3\,]]P(z=t_3)$

$\qquad+\displaystyle\sum_{t_3\in\bigcup_{j\in\{2,6,9,11,13\}}\Lambda_j}[\underline{K}-\overline{K}]P(z=t_3)$

$\quad\leq\mathrm{E}[\,y(t_1)-y(t_2)\,]$

$\quad\leq[\overline{K}-\mathrm{E}[\,y(t_2)\,|\,z=t_2\,]]P(z=t_2)+[\mathrm{E}[\,y(t_1)\,|\,z=t_1\,]-\underline{K}]P(z=t_1)$

$\qquad+\displaystyle\sum_{t_3\in\bigcup_{j\in\{1,2,7,9\}}\Lambda_j}[\mathrm{E}[\,y(t_3)\,|\,z=t_3\,]-\underline{K}]P(z=t_3)$

$\qquad+\displaystyle\sum_{t_3\in\bigcup_{j\in\{4,6,10,11\}}\Lambda_j}[\overline{K}-\mathrm{E}[\,y(t_3)\,|\,z=t_3\,]]P(z=t_3)$

$\qquad+\displaystyle\sum_{t_3\in\bigcup_{j\in\{3,5,8,12,13\}}\Lambda_j}[\overline{K}-\underline{K}]P(z=t_3).$ **(3.10)**

These bounds are sharp.

Proof of Proposition 2. By the Law of Iterated Expectations,

$$\mathrm{E}[\,y(t_1)-y(t_2)\,]=\sum_{t_3\in T}\mathrm{E}[\,y(t_1)-y(t_2)\,|\,z=t_3\,]P(z=t_3). \qquad \textbf{(3.11)}$$

Sharp bounds for the partially identified expectations on the right hand side of (3.11) will yield sharp bounds on $\mathrm{E}[\,y(t_1)-y(t_2)\,]$. I proceed by finding the sharp identification region for an arbitrary $y(t_1)-y(t_2)$ and every possible value that z may take. These can be averaged to find sharp bounds on $\mathrm{E}[\,y(t_1)-y(t_2)\,|\,z\,]$ for all z. SPM and SBM provide identifying power by establishing monotonicity relationships between treatment effects, so for each possible value of z, it must be determined which of the following inequalities are implied by the maintained assumptions:

$y(t_1)-y(t_2)\leq y(z)-\underline{K},$ **(3.12)**

$y(t_1)-y(t_2)\leq\overline{K}-y(z),$ **(3.13)**

$y(z)-\overline{K}\leq y(t_1)-y(t_2),$ **(3.14)**

$\underline{K}-y(z)\leq y(t_1)-y(t_2).$ **(3.15)**

I show that *1)* the union of these sets is $T\setminus\{t_1,t_2\}$, *2)* these sets are mutually exclusive, and *3)* they allow for a characterization of the identification region for $\mathrm{E}[\,y(t_1)-y(t_2)\,]$. $\bigcup_{j=1}^{12}\Lambda_j=T\setminus\{t_1,t_2\}$ follows from the definition of Λ_{13}. Due to the orderings required for membership, the sets $\Lambda_1\cup\Lambda_2\cup\Lambda_3$, $\Lambda_4\cup\Lambda_5\cup\Lambda_6$, $\Lambda_7\cup\Lambda_8\cup\Lambda_9$, and $\Lambda_{10}\cup\Lambda_{11}\cup\Lambda_{12}$ are mutually exclusive.

The sets Λ_1, Λ_2, and Λ_3 collect those sublattices where t_2 is the bottom and t_1 is in the interior. Λ_1, Λ_2, and Λ_3 are mutually exclusive because Λ_1 requires either the existence of a four-point sublattice containing t_1, t_2, and t_3 on which both SPM and SBM hold or the existence of at least two distinct sublattices containing t_1, t_2, and t_3 where SPM holds on one while SBM holds on the other. Λ_2 rules out both of these possibilities, only allowing for the existence of a sublattice containing t_1, t_2, and t_3 on which SPM holds; similarly, Λ_3 only allows for the existence of a sublattice containing t_1, t_2, and t_3 on which SBM holds. An identical argument establishes that Λ_4, Λ_5, and Λ_6 are mutually exclusive; these sets are defined identically to Λ_1, Λ_2, and Λ_3, respectively, with the modification that $t_3 < t_2 < t_1$ rather than $t_2 < t_1 < t_3$, so t_2 is in the interior and t_1 is the top of each sublattice.

The set construction is complicated in the above cases because the orderings $t_2 < t_1 < t_3$ and $t_3 < t_2 < t_1$ are compatible with multiple four-point sublattices containing t_1, t_2, and t_3. This results from the fact that there may be multiple t_4 such that $t_1 \vee t_4 = t_3$ and $t_1 \wedge t_4 = t_2$ (in the former case) and $t_2 \vee t_4 = t_1$ and $t_2 \wedge t_4 = t_3$ (in the latter case). When exactly two of t_1, t_2, and t_3 are incomparable, there can be at most one four-point sublattice containing t_1, t_2, and t_3, since the incomparable treatments define a unique meet and join. This simplifies the construction of the sets $\Lambda_7, \ldots, \Lambda_{12}$. By the above argument and the fact that

$$\Gamma_{t_1,t_2}^{MOD} \cap \Gamma_{t_1,t_2}^{SPM} = \Gamma_{t_1,t_2}^{MOD} \cap \Gamma_{t_1,t_2}^{SBM} = \Gamma_{t_1,t_2}^{SPM} \cap \Gamma_{t_1,t_2}^{SBM} = \emptyset,$$

the sets Λ_7, Λ_8, and Λ_9 are mutually exclusive. The same argument establishes that Λ_{10}, Λ_{11}, and Λ_{12} are mutually exclusive.

The sets $\Lambda_1, \ldots, \Lambda_{12}$ define every sublattice membership pattern for t_1 and t_2 for which SPM and SBM may have any implications; this follows from Proposition 1 and its straightforward extension to the case of SBM. I now outline the implications for the identification of $y(t_1) - y(t_2)$ when $z \in \Lambda_j$ for each $j \in \{1, \ldots, 12\}$.

If $z \in \Lambda_1$, then both SPM and SBM hold on one or more sublattices containing t_1, t_2, and z, and the required ordering implies that both (3.12) and (3.14) hold. Thus,

$$y(t_1) - y(t_2) \in [y(z) - \overline{K}, y(z) - \underline{K}]. \tag{3.16}$$

If $z \in \Lambda_2$, then t_1, t_2, and z belong to one or more sublattices on which only SPM holds. The required ordering implies that (3.15) holds and thus

$$y(t_1) - y(t_2) \in [\underline{K} - \overline{K}, y(z) - \underline{K}]. \tag{3.17}$$

If $z \in \Lambda_3$, then t_1, t_2, and z belong to one or more sublattices on which only SBM holds. The required ordering implies that (3.14) holds and thus

$$y(t_1) - y(t_2) \in [y(z) - \overline{K}, \overline{K} - \underline{K}]. \tag{3.18}$$

If $z \in \Lambda_4$, then both SPM and SBM hold on one or more sublattices containing t_1, t_2, and z, and the required ordering implies that both (3.13) and (3.15) hold. Thus,

$$y(t_1) - y(t_2) \in [\underline{K} - y(z), \overline{K} - y(z)]. \tag{3.19}$$

If $z \in \Lambda_5$, then t_1, t_2, and z belong to one or more sublattices on which only SPM holds. The required ordering implies that (3.15) holds and thus

$$y(t_1) - y(t_2) \in [\underline{K} - y(z), \overline{K} - \underline{K}]. \tag{3.20}$$

If $z \in \Lambda_6$, then t_1, t_2, and z belong to one or more sublattices on which only SBM holds. The required ordering implies that (3.13) holds and thus

$$y(t_1) - y(t_2) \in [\underline{K} - \overline{K}, \overline{K} - y(z)]. \tag{3.21}$$

If $z \in \Lambda_7$, then both SPM and SBM hold on the only sublattice containing t_1, t_2, and z, and the required ordering implies that both (3.12) and (3.14) hold. Thus,

$$y(t_1) - y(t_2) \in [y(z) - \overline{K}, y(z) - \underline{K}]. \tag{3.22}$$

If $z \in \Lambda_8$, then SPM holds on the only sublattice containing t_1, t_2, and z, and the required ordering implies that (3.14) holds and thus

$$y(t_1) - y(t_2) \in [y(z) - \overline{K}, \overline{K} - \underline{K}]. \tag{3.23}$$

If $z \in \Lambda_9$, then SBM holds on the only sublattice containing t_1, t_2, and z, and the required ordering implies that (3.12) holds and thus

$$y(t_1) - y(t_2) \in [\underline{K} - \overline{K}, y(z) - \underline{K}]. \tag{3.24}$$

If $z \in \Lambda_{10}$, then both SPM and SBM hold on the only sublattice containing t_1, t_2, and z, and the required ordering implies that both (3.15) and (3.13) hold. Thus,

$$y(t_1) - y(t_2) \in [\underline{K} - y(z), \overline{K} - y(z)]. \tag{3.25}$$

If $z \in \Lambda_{11}$, then SPM holds on the only sublattice containing t_1, t_2, and z, and the required ordering implies that (3.13) holds and thus

$$y(t_1) - y(t_2) \in [\underline{K} - \overline{K}, \overline{K} - y(z)]. \tag{3.26}$$

If $z \in \Lambda_{12}$, then SBM holds on the only sublattice containing t_1, t_2, and z, and the required ordering implies that (3.15) holds and thus

$$y(t_1) - y(t_2) \in [\underline{K} - y(z), \overline{K} - \underline{K}]. \tag{3.27}$$

The set Λ_{13} contains those t_3 that obey one of the orderings from $\Lambda_1, \ldots, \Lambda_{12}$ but do not belong to a sublattice on which SPM or SBM hold; thus, the no-assumption bounds $[\underline{K} - \overline{K}, \overline{K} - \underline{K}]$ will hold for these. It also contains t_3 such that $t_2 < t_3 < t_1$, and any four-point sublattice containing these treatments must have t_2 as the bottom and t_1 as the top; on these sublattices, SPM and SBM have no implications for identification. Additionally, Λ_{13} contains t_3 such that $t_3 \parallel t_1$ and $t_3 \parallel t_2$; SPM and SBM have no identifying power in these circumstances, as t_1, t_2, and t_3 do not share a four-point sublattice.

The focus on four-point sublattices is without loss of generality, since the implications of assumptions SPM and SBM only appear on four-point sublattices. SPM and SBM have no implications on chains, so sublattices that are chains can be ignored. Restricting attention to elements of $\{S_\gamma\}_{\gamma \in \Gamma_{t_1, t_2}} \subseteq \{S_\gamma\}_{\gamma \in \Gamma}$ is without loss of generality as well. This follows from the fact that SPM and SBM have no implications for potential outcomes under the maintained assumptions, and any

implications for the treatment effect $y(t_1) - y(t_2)$ from another treatment effect which are mediated by a third treatment effect are realized directly on a sublattice containing the treatments from the first two treatment effects. To see this concretely, suppose that $S_\gamma = \{t_2, t_1, t_3, t_4\}$ and $S_{\gamma'} = \{t_3, t_4, t_5, t_6\}$ where $t_1 \parallel t_3$, $t_4 \parallel t_5$, $t_2 = t_1 \wedge t_3$, $t_4 = t_1 \vee t_3$, $t_3 = t_4 \wedge t_5$, and $t_6 = t_4 \vee t_5$. Suppose that SPM holds on both S_γ and $S_{\gamma'}$. This implies

$$y(t_1) - y(t_2) \le y(t_4) - y(t_3) \le y(t_6) - y(t_5)$$
$$\implies y(t_1) - y(t_2) \le y(t_6) - y(t_5).$$

The fact that $\{t_2, t_1, t_5, t_6\} \in \{S_\gamma\}_{\gamma \in \Gamma_{t_1, t_2}^{SPM}}$ follows from Lemma 1 below and the definition of Γ_{t_2, t_1}^{SPM}:

LEMMA 1. *Assume that* $t_1 \parallel t_3$, $t_4 \parallel t_5$, $t_2 = t_1 \wedge t_3$, $t_4 = t_1 \vee t_3$, $t_3 = t_4 \wedge t_5$, *and* $t_6 = t_4 \vee t_5$. *Then,* $t_5 \wedge t_1 = t_2$ *and* $t_5 \vee t_1 = t_6$.

Proof. See appendix.   ∎

The argument in the SBM case simply reverses the inequalities.

The results from (3.16)–(3.27) are summarized below:

$$y(t_1) - y(t_2) \in \begin{cases} [\underline{K} - y(z), \overline{K} - y(z)] & \text{if } z \in \{t_2\} \cup \Lambda_4 \cup \Lambda_{10} \\ [y(z) - \overline{K}, y(z) - \underline{K}] & \text{if } z \in \{t_1\} \cup \Lambda_1 \cup \Lambda_7 \\ [\underline{K} - \overline{K}, y(z) - \underline{K}] & \text{if } z \in \Lambda_2 \cup \Lambda_9 \\ [\underline{K} - \overline{K}, \overline{K} - y(z)] & \text{if } z \in \Lambda_6 \cup \Lambda_{11} \\ [y(z) - \overline{K}, \overline{K} - \underline{K}] & \text{if } z \in \Lambda_3 \cup \Lambda_8 \\ [\underline{K} - y(z), \overline{K} - \underline{K}] & \text{if } z \in \Lambda_5 \cup \Lambda_{12} \\ [\underline{K} - \overline{K}, \overline{K} - \underline{K}] & \text{if } z \in \Lambda_{13} \end{cases} \qquad (3.28)$$

Since the sets $\{t_1\}, \{t_2\}, \Lambda_1, \ldots, \Lambda_{13}$ are mutually exclusive and exhaustive, (3.28) characterizes the sharp identification region for $y(t_1) - y(t_2)$ and each possible z. Averaging the bounds in (3.28) across i yields sharp bounds on $\mathrm{E}[y(t_1) - y(t_2)]$ via (3.11). These are given in (3.10).   ∎

Proposition 2 generalizes Proposition 1 by allowing for a much richer set of treatments. The treatment may have any finite number of dimensions, and each may be binary or multivalued. Some dimensions of the treatment may be complements while others are substitutes; the result allows for arbitrary combinations of SPM and SBM as appropriate. The complexity of the result is due to two factors. First, the treatment pair t_1, t_2 may belong to multiple sublattices. Second, the position of the treatment pair within a lattice, i.e., whether it includes the top and/or bottom of the sublattice, differs across sublattices. The position of the treatment pair within a sublattice combined with the assumptions that hold on the sublattice determine whether the upper and/or lower bound (or neither) are improved. While the result appears complicated, defining the Λ sets in practice seems to be fairly straightforward.

I have focused on bounding expectations of treatment effects using only super-modularity and submodularity assumptions. However, in applications these will often be paired with other assumptions, such as monotonicity. The next result modifies Proposition 2 by adding Manski's assumption of semimonotone treatment response:

PROPOSITION 3. *Assume the conditions of Proposition 2 and that SMTR holds on T. Then, for $t_2 < t_1$,*

$$0 \leq \mathrm{E}[\, y(t_1) - y(t_2)\,]$$
$$\leq [\overline{K} - \mathrm{E}[\, y(t_2) \mid z \leq t_2\,]]P(z \leq t_2) + [\mathrm{E}[\, y(t_1) \mid z \geq t_1\,] - \underline{K}]P(z \geq t_1)$$
$$+ \sum_{t_3 \in \bigcup_{j \in \{7,9\}} \Lambda_j} [\mathrm{E}[\, y(t_3) \mid z = t_3\,] - \underline{K}]P(z = t_3)$$
$$+ \sum_{t_3 \in \cup \bigcup_{j \in \{10,11\}} \Lambda_j} [\overline{K} - \mathrm{E}[\, y(t_3) \mid z = t_3\,]]P(z = t_3)$$
$$+ \sum_{t_3 \in \{t_3 \mid t_1 \npreceq t_3 \text{ and } t_3 \npreceq t_2\} \cap \bigcup_{j \in \{8,12,13\}} \Lambda_j} [\overline{K} - \underline{K}]P(z = t_3). \tag{3.29}$$

These bounds are sharp.

Proposition 3 shows that bounds computed under combinations of SMTR and SPM/SBM will generally be narrower than bounds computed under the assumptions separately. In particular, SPM and SBM may tighten the SMTR upper bound, and SMTR at least weakly improves the SPM/SBM lower bound and may tighten the upper bound as well. (B.1) shows that when $z \in \Lambda_7 \cup \Lambda_9$, SBM will improve the SMTR upper bound, and when $z \in \Lambda_{10} \cup \Lambda_{11}$, SPM will improve the SMTR upper bound. The SMTR lower bound at zero cannot be improved by using SPM or SBM. In the empirical application in Section 6, I show how the addition of SBM improves the SMTR bounds with real data.

4. INSTRUMENTAL VARIABLES

Traditional instrumental variable (IV) analysis of treatment response relies on the existence of a variable that is correlated with the treatment variable of interest but is mean-independent or independent of the distribution of response functions. Whether or not such independence assumptions are justified in a particular context is often the subject of vigorous debate. This has motivated researchers to find weaker and more credible forms of these assumptions that still retain some identification power. A leading example is the notion of a monotone instrumental variable (MIV): A variable x is an MIV if average potential outcomes conditional on x are monotone in x (Manski and Pepper, 2000, 2009). MIV and its generalizations impose restrictions on functionals of potential outcome distributions. Restrictions can also be imposed directly on functionals of treatment effect distributions:

Assumption SPMIV (Supermodular instrumental variable). x^k is a *supermodular instrumental variable* for $E[y(t_1) - y(t_2) \mid x^k, x^{-k}]$ with $t_2 \leq t_1$ if

$$x_1^k \leq x_2^k \implies E[y(t_1) - y(t_2) \mid x_1^k, x^{-k}] \leq E[y(t_1) - y(t_2) \mid x_2^k, x^{-k}] \qquad \textbf{(4.1)}$$

for all x^{-k}.[13]

SPMIV is an alternative formulation of complementarity where average treatment effects vary monotonically with an observed covariate x^k.[14] An advantage of these assumptions is that evidence for their validity may be provided by previous studies where strong identifying assumptions are credible due to controlled randomization or a natural experiment. This evidence can motivate the application of these assumptions in other contexts where similar identification strategies are not available. This contrasts with traditional IV assumptions, which tend to be highly context-specific.

The Djebbari and Smith (2008) study of the heterogeneous impacts of the PROGRESA conditional cash transfer program provides some examples of potential SPMIVs. PROGRESA provided payments to households conditional on regular school attendance by the household's children as well as visits to health centers. Djebbari and Smith find that the impact of this program on per capita consumption is substantially larger for poorer households and households in more "marginal" villages, i.e., villages with greater rates of illiteracy, more limited infrastructure, and a greater dependence on agricultural activities. Evaluations of cash transfer programs in other contexts could make use of this information by using household poverty or village marginality as SPMIVs.

Further examples are provided by the Bitler et al. (2014) study of the impact of the Connecticut Jobs First experiment. This program substantially lowered the marginal tax rate on earnings below the poverty line for families on relief, relative to the existing Aid to Families with Dependent Children (AFDC) program. In the Jobs First program, the entire benefit package is terminated once earnings rise above the poverty line; this is in contrast to the AFDC, where benefits decline linearly with earnings. Labor supply theory clearly suggests that the impact of this alternative budget scheme should boost earnings and employment much more for those who were previously out of work or whose earnings left them far below the poverty line. These hypotheses are strongly borne out by the data, suggesting that measures of pre-program earnings and employment could serve as SPMIVs in studies of similar programs which are not implemented experimentally.

Manski and Pepper (2000) considered a specific case of an MIV where the average potential outcomes are monotone in the realized treatment; they referred to this as the monotone treatment selection assumption. A similar assumption can be developed in this context:

Assumption SPMTS (Supermodular treatment selection). The *supermodular treatment selection* assumption holds for $E[y(t_1) - y(t_2)]$ with $t_2 \leq t_1$ if

$$E[y(t_1) - y(t_2) \mid z = t_1] \geq E[y(t_1) - y(t_2) \mid z = t_2]. \qquad \textbf{(4.2)}$$

The bounds under this assumption will be analogous to those under SPMIV derived below. Let $\underline{B}(t,x)$ and $\overline{B}(t,x)$ be defined as

$$\underline{B}(t,x) = \mathrm{E}[\, y(t)\mid z=t,x\,]P(z=t\mid x) + \underline{K}P(z\neq t\mid x) \quad \forall t\in T, x\in X$$

and

$$\overline{B}(t,x) = \mathrm{E}[\, y(t)\mid z=t,x\,]P(z=t\mid x) + \overline{K}P(z\neq t\mid x) \quad \forall t\in T, x\in X.$$

The following bounds can be derived using SPMIV:

PROPOSITION 4. *Assume that x^k is an SPMIV for $\mathrm{E}[\, y(t_1)-y(t_2)\mid x^k, x^{-k}\,]$ with $t_1, t_2 \in T$. Then, the bounds*

$$\sup_{x_2^k \leq x_1^k} \left\{ \underline{B}\big(t_1, x_2^k, x^{-k}\big) - \overline{B}\big(t_2, x_2^k, x^{-k}\big) \right\}$$

$$\leq \mathrm{E}\big[\, y(t_1) - y(t_2) \mid x_1^k, x^{-k}\,\big]$$

$$\leq \inf_{x_1^k \leq x_2^k} \left\{ \overline{B}\big(t_1, x_2^k, x^{-k}\big) - \underline{B}\big(t_2, x_2^k, x^{-k}\big) \right\} \tag{4.3}$$

are sharp.

As is the case for bounds derived under IV or MIV assumptions, inference is complicated by the sup and inf operators in equation (4.3) (Manski and Pepper, 2009). Analog estimators of the bounds in (4.3) are consistent but biased in finite samples; the estimated bounds will generally be too narrow. Fortunately, the methods developed by Chernozhukov, Lee, and Rosen (2013) can be applied to find bias-corrected estimates and associated confidence intervals. Chernozhukov et al. discuss in detail the special cases of estimating nonparametric bounds using instrumental variables and MIVs; the bounds in (4.3) are essentially identical for the purposes of estimation, so their results can be applied directly to my estimation problem. The theoretical extension allowing for multiple SPMIVs is straightforward, and presents no novel estimation challenges besides those associated with high-dimensional nonparametric conditioning.

Returning to assumption SPMIV: If the second inequality in (4.1) is reversed, x^k becomes a *submodular instrumental variable*. If x^k is a supermodular and submodular instrumental variable, i.e., average treatment effects are constant across different values of x^k, then x^k is a *modular instrumental variable*. While this may seem like a strong assumption, it is routinely employed in applied work that models treatment effects without allowing for interactions.

SPMIVs may also improve the bounds on functionals of potential outcome distributions, as the following corollary illustrates:

COROLLARY 1. *Assume that x^k is an SPMIV for $\mathrm{E}[\, y(t_1)-y(t_2)\mid x^k, x^{-k}\,]$ with $t_1, t_2 \in T$. Then, the bounds*

$$\max\left\{\underline{B}(t_1, x_1^k, x^{-k}), \sup_{x_2^k \le x_1^k}\left\{\underline{B}(t_1, x_2^k, x^{-k}) - \overline{B}(t_2, x_2^k, x^{-k})\right\} + \underline{B}(t_2, x_1^k, x^{-k})\right\}$$

$$\le \mathrm{E}\left[y(t_1) \mid x_1^k, x^{-k}\right]$$

$$\le \min\left\{\overline{B}(t_1, x_1^k, x^{-k}), \inf_{x_1^k \le x_2^k}\left\{\overline{B}(t_1, x_2^k, x^{-k}) - \underline{B}(t_2, x_2^k, x^{-k})\right\} + \overline{B}(t_2, x_1^k, x^{-k})\right\} \quad \textbf{(4.4)}$$

and

$$\max\left\{\underline{B}(t_2, x_1^k, x^{-k}), \underline{B}(t_1, x_1^k, x^{-k}) - \inf_{x_1^k \le x_2^k}\left\{\overline{B}(t_1, x_2^k, x^{-k}) - \underline{B}(t_2, x_2^k, x^{-k})\right\}\right\}$$

$$\le \mathrm{E}\left[y(t_2) \mid x_1^k, x^{-k}\right]$$

$$\le \min\left\{\overline{B}(t_2, x_1^k, x^{-k}), \overline{B}(t_1, x_1^k, x^{-k}) - \sup_{x_2^k \le x_1^k}\left\{\underline{B}(t_1, x_2^k, x^{-k}) - \overline{B}(t_2, x_2^k, x^{-k})\right\}\right\} \quad \textbf{(4.5)}$$

are sharp.

The proof of this result is straightforward, requiring only verification that these potential outcome bounds do not imply that the bounds in (4.3) can be tightened. As in the case of Proposition 4, the Chernozhukov et al. approach to inference can be applied here. It is natural to consider the application of this IV assumption in concert with the shape restrictions considered above. Proposition 5 below shows how SPMIV can be combined with SPM and SMTR:

PROPOSITION 5. *Assume the conditions of Proposition 2 and assume that* x^k *is an SPMIV for* $\mathrm{E}[y(t_1) - y(t_2) \mid x^k, x^{-k}]$. *Let* $\overline{\Delta B}(t_1, t_2, x^k, x^{-k})$ *and* $\underline{\Delta B}(t_1, t_2, x^k, x^{-k})$ *denote the conditional upper and lower bounds, respectively, for* $\mathrm{E}[y(t_1) - y(t_2) \mid x^k, x^{-k}]$ *from (3.10). Then, the bounds*

$$\sup_{x_2^k \le x_1^k}\left\{\underline{\Delta B}(t_1, t_2, x_2^k, x^{-k})\right\} \le \mathrm{E}\left[y(t_1) - y(t_2) \mid x_1^k, x^{-k}\right]$$

$$\le \inf_{x_1^k \le x_2^k}\left\{\overline{\Delta B}(t_1, t_2, x_2^k, x^{-k})\right\} \quad \textbf{(4.6)}$$

are sharp. Under the additional assumption of SMTR, these bounds are sharp when $\overline{\Delta B}(t_1, t_2, x^k, x^{-k})$ *and* $\underline{\Delta B}(t_1, t_2, x^k, x^{-k})$ *denote the conditional upper and lower bounds, respectively, from (3.29).*

This result, combining SPMIV with SMTR and the new shape restrictions proposed above, will be applied in the empirical exercise below.

5. INDEPENDENCE

Independence assumptions have been used to operationalize the belief that individuals' realized treatments are unrelated to any individual characteristics which may influence responses. This should be the case, for example, in a randomized controlled trial. I show how statistical independence can be combined with shape restrictions and instrumental variables assumptions to narrow the bounds on entire treatment effect distributions.

The familiar assumption of statistical independence of treatments and response functions is defined in my notation as follows:

Assumption SI (Statistical independence)**.** Potential outcomes are *statistically independent* of realized treatments if

$$P(y(t) \mid z) = P(y(t)) \quad \forall t \in T.$$

Assumption SI implies that the marginal distribution of $y(t)$, denoted F_t, is point identified for all $t \in T$ such that $P(z = t) > 0$. However, the distribution of $y(t_1) - y(t_2)$, whose cumulative distribution function is denoted by F_{t_1, t_2}, is only partially identified. Makarov (1982) was the first to derive pointwise sharp bounds on the distribution of the sum of two random variables with fixed marginal distributions. Frank, Nelsen, and Schweizer (1987) derived these bounds in a simpler manner and extended them to allow for other operations such as differences and products as well as more than two variables. However, as Kreinovich and Ferson (2006) show, these bounds are not sharp in the case of more than two variables. The following result, taken from Theorem 2 of Williamson and Downs (1990), gives the sharp bounds on the distribution of $y(t_1) - y(t_2)$ for any $t_1, t_2 \in T$:

$$\underline{F}_{t_1,t_2}(w) = \sup_{u+v=w} \left\{ \max \left\{ F_{t_1}(u) - F_{t_2}(-v), 0 \right\} \right\}$$

$$\leq F_{t_1,t_2}(w) \leq 1 + \inf_{u+v=w} \left\{ \min \left\{ F_{t_1}(u) - F_{t_2}(-v), 0 \right\} \right\} = \overline{F}_{t_1,t_2}(w)$$

Fan and Park (2010) discuss consistent nonparametric estimation of these bounds. This result can be applied to derive sharp bounds on the quantile function of $y(t_1) - y(t_2)$, $F_{t_1,t_2}^{-1}(q)$, where F_{t_1,t_2}^{-1} denotes the generalized inverse of the cdf F_{t_1,t_2}. Define the following functions:

$$\underline{F}_{t_1,t_2}^{-1}(q) = \begin{cases} \inf_{u \in [q,1]} [F_{t_1}^{-1}(u) - F_{t_2}^{-1}(u-q)] & \text{if } q \neq 0 \\ F_{t_1}^{-1}(0) - F_{t_2}^{-1}(1) & \text{if } q = 0 \end{cases},$$

$$\overline{F}_{t_1,t_2}^{-1}(q) = \begin{cases} \sup_{u \in [0,q]} [F_{t_1}^{-1}(u) - F_{t_2}^{-1}(1+u-q)] & \text{if } q \neq 1 \\ F_{t_1}^{-1}(1) - F_{t_2}^{-1}(0) & \text{if } q = 1 \end{cases}. \tag{5.1}$$

Then, $\overline{F}_{t_1,t_2}^{-1}(q) \leq F_{t_1,t_2}^{-1}(q) \leq \underline{F}_{t_1,t_2}^{-1}(q)$. SI can be combined with SPM to refine these bounds, as the following result shows:

PROPOSITION 6. *Assume that SI holds and that* $T = \{t_1 \wedge t_2, t_1, t_2, t_1 \vee t_2\}$ *with* $t_1 \wedge t_2 < t_1, t_2 < t_1 \vee t_2$. *Assume that SPM holds on* T. *Then, the bounds*

$$\max \left\{ \overline{F}_{t_1 \vee t_2, t_1}^{-1}(q), \overline{F}_{t_2, t_1 \wedge t_2}^{-1}(q) \right\} \leq F_{t_1 \vee t_2, t_1}^{-1}(q) \leq \underline{F}_{t_1 \vee t_2, t_1}^{-1}(q)$$

and

$$\overline{F}_{t_1, t_1 \wedge t_2}^{-1}(q) \leq F_{t_1, t_1 \wedge t_2}^{-1}(w) \leq \min \left\{ \underline{F}_{t_1, t_1 \wedge t_2}^{-1}(q), \underline{F}_{t_1 \vee t_2, t_2}^{-1}(q) \right\}$$

are sharp.

Similar results can be derived for SBM. These shape restrictions could be justified by theoretical arguments; alternatively, since average treatment effects are point-identified in this context, the supermodularity or submodularity of average effects could be used to provide some justification for stronger structural assumptions. Extending these results to general lattices is problematic due to the fact that sharp bounds on the distribution function of a sum of more than two variables are an open question. Nonetheless, it is straightforward to collect all possible stochastic dominance relations implied by the maintained assumptions, and bounds which contain the true value (but are not necessarily sharp) can be obtained in a manner similar to that in Proposition 6. Such bounds may be useful in policy evaluation.

A reformulation of the SPMIV assumption can also be applied in this setting:[15]

Assumption Q-SPMIV (Quantile supermodular instrumental variable). x^k is a *quantile supermodular instrumental variable* for $y(t_1) - y(t_2)$ if

$$x_1^k \leq x_2^k \implies F_{t_1,t_2}^{-1}(q \mid x_1^k, x^{-k}) \leq F_{t_1,t_2}^{-1}(q \mid x_2^k, x^{-k})$$

for all x^{-k}.

Giustinelli (2011) analyzes the returns to education in Italy using a similar restriction on the quantile function of potential outcomes. Blundell, Gosling, Ichimura, and Meghir (2007) impose monotonicity in a covariate on the conditional cdf of a potential outcome. Their bounds are simplified by the fact that the distribution of potential outcomes is partially observed, while the distribution of treatment effects is never observed, necessitating the use of the Williamson and Downs (1990) bounds. The following proposition computes the bounds derived under Q-SPMIV:

PROPOSITION 7. *Assume that SI holds. Assume that x^k is a Q-SPMIV for $y(t_1) - y(t_2)$ with $t_1, t_2 \in T$. Then, the bounds*

$$\sup_{x_2^k \leq x_1^k} \left\{ \overline{F}_{t_1,t_2}^{-1}(q \mid x_2^k, x^{-k}) \right\} \leq F_{t_1,t_2}^{-1}(q \mid x_1^k, x^{-k}) \leq \inf_{x_1^k \leq x_2^k} \left\{ \underline{F}_{t_1,t_2}^{-1}(q \mid x_2^k, x^{-k}) \right\}$$

are sharp.

Again, since conditional average treatment effects are point-identified, they can provide some evidence to support the validity of the stronger Q-SPMIV assumption.

6. EMPIRICAL ILLUSTRATION

To illustrate the use of the identification results developed in this paper, I re-analyze data from Shertzer, Twinam, and Walsh (2016b). That study examines the extent to which Chicago's first zoning ordinance, passed in 1923, influenced

the evolution of the spatial distribution of commercial, industrial, and residential activity in the city. Evidence of substantial treatment effect heterogeneity was found, motivating the use of SBM and SPMIV assumptions in the analysis below.

Chicago's 1923 zoning ordinance regulated land by restricting uses and density.[16] Here, I bound the effects of 1923 commercial zoning on the probability that a city block will contain any commercial activity in 2005. Zoning for each type of use (commercial, industrial, single/multi-family residential) was coupled with an allowed maximum density level. Focusing on four different allowed density levels and an indicator for whether or not a block received commercial zoning yields eight distinct treatments and four distinct commercial zoning treatment effects.

Formally, the outcome variable $y(\cdot)$ is an indicator equal to 1 iff city block i contains any commercial activity in 2005. y is a function of a treatment $t \in T = \{0, 1\} \times \{1, 2, 3, 4\}$. The first dimension of t is equal to 1 if the block received any commercial zoning in 1923 and 0 otherwise. The second dimension of t is equal to 1 if the block was zoned for the lowest density development (3 or fewer stories), 2 if it was zoned for higher densities suitable for apartment buildings (8–10 stories), 3 if it was zoned for high-rise commercial buildings, and 4 if it was zoned for the tallest commercial skyscrapers (this latter zoning was reserved for the central business district and immediately surrounding area).

Since numerous other factors, such as pre-zoning land use, property values, and demographics, shaped both the initial zoning ordinance as well as the future development of the city, the exogenous treatment selection (ETS) assumption is likely too strong. As an alternative, I employ mixtures of SBM, SMTR, and SPMIV in the empirical analysis. SBM implies that the commercial zoning treatment effect will be larger when paired with lower density zoning. This is motivated by the fact that areas zoned for lower densities will be more residential in character and contain a larger proportion of single-family homes (Shertzer et al., 2016b). It is well documented that residential property owners (especially single-family homeowners) generally oppose the encroachment of commercial uses and have substantial power to block such development (Fischel, 2001). It is likely that the early establishment of commercial activity through zoning will be a more important determinant of future commercial land use in areas also zoned for lower densities.[17] This assumption is also consistent with previous literature showing that mixed use areas are more likely to see conversion to completely nonresidential use than strictly residential use (McMillen and McDonald, 1991).

While the effect of commercial zoning is likely to be larger in low-density areas, it is generally the case that commercial uses appear more often in denser areas. Commercial uses can afford the higher rents that prevail in dense areas, benefit more from agglomeration economies, and are compatible with residential development through mixed-use structures. This, combined with the fact that

commercial zoning is unlikely to lower the likelihood of commercial development, suggests that SMTR is an appropriate assumption in this context. Table 1 shows how the upper and lower bounds on the average treatment effect of commercial zoning, $E[y(1, d) - y(0, d)]$, vary under different shape restrictions; bounds on the ATE are obtained for each density level $d \in \{1, 2, 3, 4\}$. Appendix Figure 1 depicts the results from Table 1 graphically.

The results in Table 1 show how the imposition of the shape restrictions SBM and SMTR can separately and jointly improve upon the no assumption upper and lower bounds on treatment effects. The no assumption bounds are very wide due to the large number of treatments. The imposition of SBM establishes monotonicity of the upper and lower bounds as expected; upper bounds on the commercial zoning effects are smaller in higher density areas, while lower bounds are higher in low density areas. SMTR restricts the lower bound to be zero but its effect on upper bounds is theoretically ambiguous. Upper bounds on the potential outcomes treated with commercial zoning are monotone increasing with density:

$$E[y(1, 1)] \leq E[y(1, 2)] \leq E[y(1, 3)] \leq E[y(1, 4)].$$

However, the same is true of lower bounds on potential outcomes of observations not treated with commercial zoning:

$$E[y(0, 1)] \leq E[y(0, 2)] \leq E[y(0, 3)] \leq E[y(0, 4)].$$

In this case, the SMTR upper bounds on treatment effects are increasing with density; however, with the addition of SBM, this is reversed, since SBM implies that

TABLE 1. Bounds on the effects of commercial zoning on future commercial use

Treatment effect	ETS	None		SBM		SMTR		SBM/SMTR	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
$E[y(1, 1) - y(0, 1)]$	0.579	−0.702	0.932	−0.271	0.932	0	0.754	0	0.754
	(0.557, 0.601)	−0.710	0.936	−0.277	0.936	0	0.761	0	0.761
$E[y(1, 2) - y(0, 2)]$	0.503	−0.664	0.858	−0.569	0.790	0	0.803	0	0.738
	(0.483, 0.524)	−0.672	0.863	−0.578	0.796	0	0.809	0	0.745
$E[y(1, 3) - y(0, 3)]$	0.311	−0.936	0.957	−0.905	0.747	0	0.923	0	0.732
	(0.269, 0.354)	−0.941	0.961	−0.910	0.754	0	0.927	0	0.739
$E[y(1, 4) - y(0, 4)]$	0.341	−0.969	0.982	−0.969	0.729	0	0.957	0	0.729
	(0.266, 0.417)	−0.972	0.984	−0.972	0.736	0	0.960	0	0.736
Observations	14,690	14,690	14,690	14,690	14,690	14,690	14,690	14,690	14,690

Bounds on the effect of 1923 commercial zoning on the probability of commercial use in 2005 under different shape restrictions. Commercial zoning effect is estimated separately for all four levels of 1923 density zoning. Upper and lower bounds are accompanied below by bootstrapped upper and lower 95% confidence intervals. Column 1 presents point estimates assuming exogenous treatment selection (ETS), along with 95% confidence intervals. Columns 2–3 present the no assumption bounds. Columns 4–5 add the assumption of submodularity (SBM). Columns 6–7 add the assumption of semimonotone treatment response (SMTR). Columns 8–9 combine SBM and SMTR.

the commercial zoning effect will be larger in areas also zoned for low densities. Under both SMTR and SBM, the results indicate that commercial zoning in 1923 increases the probability of commercial activity in 2005 by at most roughly 0.75 on the entire sample. While this still leaves a sizable range of possible values, it shows that historical zoning is not necessarily decisive, a fact which is not discernible from the uninformative no assumption bounds or the SMTR bounds in high density areas.

While SBM cannot establish the nonnegativity of the treatment effects on the whole sample, an analysis of the impact of commercial and low density zoning on the subsample of blocks which were not treated with commercial zoning is more informative. Table 2 shows that SBM alone establishes a lower bound of -0.068 on $E[y(1,1) - y(0,1)]$. This can be strengthened using the fraction of the block's population that are not at least 3rd generation native white as an SPMIV. The argument here is that native whites were a politically powerful group relative to blacks and recent immigrants, especially those from Eastern Europe. These latter groups would have less power to alter zoning ex post, and so one would expect the 1923 zoning effect to be larger if more of the population belongs to these politically marginalized groups.[18] Computing the average treatment effect conditional on 40% of the population being white using this SPMIV yields a lower bound of -0.01, so these two assumptions alone nearly identify the sign of the treatment effect.

Table 3 presents a further analysis of the upper bounds of the commercial zoning effect. The sample is restricted to blocks that fall into the bottom quartile of pre-zoning development; development is measured using a continuous index

TABLE 2. Bounds on the effect of commercial zoning on future commercial use: subsample not treated with commercial zoning

Treatment effect	Assumptions					
	None		SBM		SBM/SPMIV	
	Lower	Upper	Lower	Upper	Lower	Upper
$E[y(1,1) - y(0,1)]$	-0.509	0.991	-0.068	0.991	-0.01	0.992
	-0.521	0.994	-0.075	0.994	-0.0213	0.994
Observations	6,072	6,072	6,072	6,072	6,027	6,027
Sample	No comm. zoning		No comm. zoning		No comm. zoning	

Bounds on the effect of 1923 commercial zoning on the probability of commercial use in 2005 when paired with the most restrictive 1923 density zoning. Sample is restricted to blocks not treated with commercial zoning. No assumption and SBM upper and lower bounds are reported with bootstrapped upper and lower 95% confidence intervals. The average treatment effect in the final two columns is conditional on 40% of the population being white (3rd gen. and above) with the SPMIV assumption that conditional average treatment effects are monotone increasing in the percentage of the population that are non-3rd gen. white. SPMIV bounds are half-median unbiased estimates with associated 95% confidence intervals calculated using the Chernozhukov, Kim, Lee, and Rosen (2015). Stata implementation of the inference methods of Chernozhukov et al. (2013).

TABLE 3. Bounds on the effect of commercial zoning on future commercial use: undeveloped subsample

Treatment effect	ETS	None	SMTR	SBM/SMTR	SBM/SMTR/SPMIV
		Assumptions			
		Upper	Upper	Upper	Upper
$E[y(1,1) - y(0,1)]$	0.51 (0.469, 0.550)	0.846 0.862	0.806 0.825	0.806 0.825	0.661 0.688
$E[y(1,2) - y(0,2)]$	0.425 (0.352, 0.496)	0.921 0.931	0.918 0.928	0.769 0.786	0.628 0.656
Observations	3,669	3,669	3,669	3,669	3,669
Sample	Undeveloped	Undeveloped	Undeveloped	Undeveloped	Undeveloped

Bounds on the effect of 1923 commercial zoning on the probability of commercial use in 2005 when paired with the two most restrictive levels of 1923 density zoning. Estimation is on a subset of blocks that were largely undeveloped in 1923; since these areas exclusively received zoning for either density level 1 or 2, I only discuss those treatment effects. Average treatment effects are conditional on 40% of the population being white (3rd gen. and above). The bounds are computed under no assumptions, SBM/SMTR, and SBM/SMTR/SPMIV. No assumption and SBM/SMTR upper bounds are reported with upper 95% confidence intervals from a kernel-weighted local polynomial smoother. SPMIV bounds assume that conditional average treatment effects are monotone increasing in the percentage of the population that are non-3rd gen. white. SPMIV bounds are half-median unbiased estimates with associated 95% confidence intervals calculated using the Chernozhukov et al. (2015) Stata implementation of the inference methods of Chernozhukov et al. (2013).

which is a function of population density, building heights, commercial and manufacturing establishment density, and distance to the central business district and Lake Michigan. The blocks in this sample lie in the outlying areas of the city which were largely undeveloped in 1923, with many vacant lots, very low population density, and few business establishments.[19] This is the area where the effect of commercial zoning should be the largest, as these areas also received the most restrictive density zoning and the pattern of future development was essentially undetermined.

The results in Table 3 show that commercial zoning in 1923 increases the probability of commercial activity in 2005 by at most 0.66 in areas zoned for the lowest density and 0.63 in areas zoned for higher densities. While still compatible with a range of magnitudes, these results show that historical zoning was limited in its ability to direct future development, leaving a large role for market-driven rezonings to alter the initial plan.

7. CONCLUSION

In this paper, I contribute to the literature on the partial identification of treatment effects by developing and applying assumptions that formalize the notion of complementarity. I examine the identification power of these assumptions and discuss how they can be justified. The supermodularity and submodularity assumptions proposed can be used to narrow bounds on treatment effects in studies of policy complementarity, which have traditionally been stymied by a lack of

pseudo-experimental variation in multiple policies simultaneously. Proposition 1 shows how these shape restrictions can improve bounds on average treatment effects in the simple case of two binary treatments. Proposition 2 extends this result to a more general treatment set with an arbitrary finite number of (possibly multivalued) treatments and the possibility of complex combinations of supermodularity and submodularity. Proposition 3 combines the supermodularity and submodularity assumptions with the semimonotone treatment response assumption of Manski (1997).

Complementarity may also stem from differential treatment response among subpopulations defined by observed covariates. Subgroup heterogeneity in treatment effects is an increasingly widely recognized phenomenon, and can often be motivated directly from economic theory (see, e.g., Bitler et al., 2014). Proposition 4 and Corollary 1 show how qualitative information about treatment effect heterogeneity embodied in supermodular instrumental variables can be used to improve bounds on average treatment effects and average potential outcomes, respectively. Proposition 5 combines supermodular IVs with the new shape restrictions as well as semimonotone treatment response. Supermodular instrumental variables can be used in studies with one or many treatments, making them a versatile and potentially powerful addition to the arsenal of applied econometricians.

The assumptions developed here can be useful in the experimental context as well. Proposition 6 shows how supermodularity can be combined with an assumption of statistical independence between assigned treatments and responses to yield improved bounds on the cumulative distribution function of a treatment effect. These results can be applied to the evaluation of outcomes in complex (multitreatment) randomized controlled trials, which are increasingly prevalent in many fields, including development economics. Since average treatment effects are point-identified in this context, one can determine if average responses exhibit supermodularity or submodularity. This can provide evidence that individual response functions are supermodular or submodular. Similarly, the behavior of (point-identified) conditional average treatment effects can motivate the use of a quantile supermodular instrumental variable; Proposition 7 shows how this assumption can strengthen the bounds on the quantile function of a treatment effect distribution.

Bounds derived under the assumptions I propose here are of interest only to the extent that such assumptions are considered credible. Where might evidence for their validity come from? Arguments for policy complementarity may be provided by economic theory, as in Lalive et al. (2006), or they may come from multitreatment randomized controlled trials. Evidence on subgroup heterogeneity in treatment effects may be provided by previous studies where strong identifying assumptions are credible due to controlled randomization or a natural experiment. In such studies, conditional average treatment effects are point-identified, so the validity of the proposed assumptions can be established. This can motivate their use in other contexts where similar identification strategies are not available.

This distinguishes supermodular IV assumptions from traditional IV assumptions, since the latter tend to be context-specific.

The empirical illustration in Section 6 employs submodularity, semimonotone treatment response, and supermodular IVs to study the impact of historical zoning on the evolution of land use in Chicago. In some cases, I am able to essentially rule out the possibility of a negative commercial zoning effect using only submodularity and a supermodular IV. I am also able to show that the effect of commercial zoning was far from decisive even in the areas of the city that were undeveloped at the time of zoning, demonstrating the potentially sizable role of market-driven zoning revisions.

NOTES

1. Examples include: Giustinelli (2011) and Tsunao and Usui (2014) on the returns to education, Kreider and Pepper (2007) on disability and employment, Bhattacharya, Shaikh, and Vytlacil (2008, 2012) on the mortality effects of Swan–Ganz catheterization, Kreider and Hill (2009) on the effect of universal health insurance on medical expenditures, Pepper (2000) on the intergenerational transmission of welfare receipt, Manski and Nagin (1998) on sentencing and recidivism, and Gundersen, Kreider, and Pepper (2012) on the health effects of the National School Lunch Program.

2. The sensitivity of effects to the surrounding policy environment may partly explain the wide variation in estimates of treatment effects for similar policies in different contexts found in many literatures; see, for example, the discussion in Lalive et al. (2006) on the effects of unemployment benefit policies on re-employment rates.

3. See Manski and Pepper (2000) and Section 4.

4. See Manski (2003) for a comprehensive overview.

5. I reserve superscripts to refer to vector components.

6. If t_1 and t_2 are incomparable, I write $t_1 \parallel t_2$.

7. The meet and join operations depend on the particular order imposed on the lattice; for example, the join of $(2,0)$ and $(1,1)$ in $\mathbb{R}^2$ is equal to $(2,1)$ under the product order and $(2,0)$ under the lexicographic order.

8. I suppress i when referring to arbitrary response functions, covariates, or realized treatments.

9. The stable unit treatment value assumption (alternatively referred to as noninterference by Cox (1958) and individualistic treatment response by Manski (2013)) is maintained throughout the paper; this states that individuals' potential outcomes $\{y(t)\}_{t \in T}$ do not depend on other individuals' realized treatments (Rubin, 1978).

10. If an expectation $\mathrm{E}[\, y(t_1) \mid z = t_2 \,]$ is ill-defined because the event $z = t_2$ is off the support of z, I establish the convention that $\mathrm{E}[\, y(t_1) \mid z = t_2 \,] P(z = t_2) \equiv 0$.

11. In the case where one or more of the dimensions of the treatment are continuous, the data are necessarily uninformative about almost all of the treatments. This motivates my restriction to discretely-valued treatments.

12. There is no guarantee that these bounds will be nonempty; if an assumption implies that the bounds on the parameter of interest are empty, the assumption can be falsified by the data. This caveat applies to all the results that follow.

13. The weak inequality in (4.1) can be reversed, in which case x^k would be a *submodular instrumental variable* (SBMIV). If the inequality is replaced with equality, x^k becomes a *modular instrumental variable* (MODIV).

14. The SPM/SPMIV distinction is analogous to the MTR/MIV distinction; see Manski and Pepper (2009).

15. SPMIV itself is unhelpful, since conditional average treatment effects are point-identified.

16. For details on the ordinance, consult Shertzer et al. (2016b).

17. Shertzer et al. (2016b) provide direct evidence of the veracity of this assumption.

18. This is consistent with the historical record as well as the results of Shertzer, Twinam, and Walsh (2016a), which provide direct evidence of discrimination against blacks and recent immigrants in the drafting of the 1923 ordinance in Chicago. An alternative approach could involve including recent Irish immigrants with native whites, as the Irish became politically influential early in Chicago's history; experimenting with this alternate approach yielded similar results.

19. These blocks had a median of two residents and zero commercial/industrial uses per acre.

REFERENCES

Aradillas-Lopez, A. & E. Tamer (2008) The identification power of equilibrium in simple games. *Journal of Business & Economic Statistics* 26(3), 261–283.

Beresteanu, A. (2005) Nonparametric analysis of cost complementarities in the telecommunications industry. *RAND Journal of Economics* 36(4), 870–889.

Beresteanu, A. (2007) Nonparametric estimation of regression functions under restrictions on partial derivatives. Unpublished manuscript.

Bhattacharya, J., A.M. Shaikh, & E. Vytlacil (2008) Treatment effect bounds under monotonicity assumptions: An application to Swan–Ganz catheterization. *American Economic Review* 98(2), 351–356.

Bhattacharya, J., A.M. Shaikh, & E. Vytlacil (2012) Treatment effect bounds: An application to Swan–Ganz catheterization. *Journal of Econometrics* 168(2), 223–243.

Bitler, M.P., J.B. Gelbach, & H.W. Hoynes (2006) What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review* 96(4), 988–1012.

Bitler, M.P., J.B. Gelbach, & H.W. Hoynes (2008) Distributional impacts of the self-sufficiency project. *Journal of Public Economics* 92(3), 748–765.

Bitler, M.P., J.B. Gelbach, & H.W. Hoynes (2014) Can variation in subgroups' average treatment effects explain treatment effect heterogeneity? Evidence from a social experiment. NBER Working Paper 20142.

Blundell, R., A. Gosling, H. Ichimura, & C. Meghir (2007) Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica* 75(2), 323–363.

Cattaneo, M.D. (2010) Multi-valued treatment effects. In S.N. Durlauf & L.E. Blume (eds.), *New Palgrave Dictionary of Economics*, Palgrave Macmillan, pp. 855–857.

Chernozhukov, V., W. Kim, S. Lee, & A.M. Rosen (2015) Implementing intersection bounds in Stata. *Stata Journal* 15(1), 21–44.

Chernozhukov, V., S. Lee, & A.M. Rosen (2013) Intersection bounds: Estimation and inference. *Econometrica* 81(2), 667–737.

Cox, D.R. (1958) *Planning of Experiments*. Wiley.

Djebbari, H. & J. Smith (2008) Heterogeneous impacts in PROGRESA. *Journal of Econometrics* 145(1), 64–80.

Eeckhout, J. & P. Kircher (2011) Identifying sorting – In theory. *Review of Economic Studies* 78(3), 872–906.

Fan, Y. & S.S. Park (2010) Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory* 26, 931–951.

Feller, A. & C.C. Holmes (2009) Beyond toplines: Heterogeneous treatment effects in randomized experiments. Unpublished manuscript.

Fischel, W.A. (2001) *The Homevoter Hypothesis*. Harvard University Press.

Frank, M.J., R.B. Nelsen, & B. Schweizer (1987) Best-possible bounds for the distribution of a sum – A problem of Kolmogorov. *Probability Theory and Related Fields* 74(2), 199–211.

Giustinelli, P. (2011) Non-parametric bounds on quantiles under monotonicity assumptions: With an application to the Italian education returns. *Journal of Applied Econometrics* 26(5), 783–824.

Graham, B.S., G.W. Imbens, & G. Ridder (2014) Complementarity and aggregate implications of assortative matching: A nonparametric analysis. *Quantitative Economics* 5(1), 29–66.

Gundersen, C., B. Kreider, & J.V. Pepper (2012) The impact of the National School Lunch Program on child health: A nonparametric bounds analysis. *Journal of Econometrics* 166(1), 79–91.

Kreider, B. & S.C. Hill (2009) Partially identifying treatment effects with an application to covering the uninsured. *Journal of Human Resources* 44(2), 409–449.

Kreider, B. & J.V. Pepper (2007) Disability and employment: Reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association* 102(478), 432–441.

Kreinovich, V. & S. Ferson (2006) Computing best–possible bounds for the distribution of a sum of several variables is NP–hard. *International Journal of Approximate Reasoning* 41(3), 331–342.

Lalive, R., J. Van Ours, & J. Zweimüller (2006) How changes in financial incentives affect the duration of unemployment. *Review of Economic Studies* 73(4), 1009–1038.

Lazzati, N. (2015) Treatment response with social interactions: Partial identification via monotone comparative statics. *Quantitative Economics* 6(1), 49–83.

Makarov, G.D. (1982) Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability & its Applications* 26(4), 803–806.

Manski, C.F. (1989) Anatomy of the selection problem. *Journal of Human Resources* 24(3), 343–360.

Manski, C.F. (1997) Monotone treatment response. *Econometrica* 65(6), 1311–1334.

Manski, C.F. (2003) *Partial Identification of Probability Distributions*. Springer.

Manski, C.F. (2013) Identification of treatment response with social interactions. *Econometrics Journal* 16(1), S1–S23.

Manski, C.F. & D.S. Nagin (1998) Bounding disagreements about treatment effects: A case study of sentencing and recidivism. *Sociological Methodology* 28(1), 99–137.

Manski, C.F. & J.V. Pepper (2000) Monotone instrumental variables: With an application to the returns to schooling. *Econometrica* 68(4), 997–1010.

Manski, C.F. & J.V. Pepper (2009) More on monotone instrumental variables. *Econometrics Journal* 12, S200–S216.

McMillen, D.P. & J.F. McDonald (1991) A Markov chain model of zoning change. *Journal of Urban Economics* 30(2), 257–270.

Molinari, F. & A.M. Rosen (2008) The identification power of equilibrium in games: The supermodular case. *Journal of Business & Economic Statistics* 26(3), 297–302.

Neumark, D. & W. Wascher (2011) Does a higher minimum wage enhance the effectiveness of the Earned Income Tax Credit? *Industrial and Labor Relations Review* 64(4), 712–746.

Pepper, J.V. (2000) The intergenerational transmission of welfare receipt: A nonparametric bounds analysis. *Review of Economics and Statistics* 82(3), 472–488.

Rubin, D.B. (1978) Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6(1), 34–58.

Shertzer, A., T. Twinam, & R.P. Walsh (2016a) Race, ethnicity, and discriminatory zoning. *American Economic Journal: Applied Economics* 8(3), 1–31.

Shertzer, A., T. Twinam, & R.P. Walsh (2016b) Zoning and the economic geography of cities. Unpublished manuscript.

Tsunao, O. & E. Usui (2014) Concave–monotone treatment response and monotone treatment selection: With an application to the returns to schooling. *Quantitative Economics* 5(1), 175–194.

Williamson, R.C. & T. Downs (1990) Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning* 4(2), 89–158.
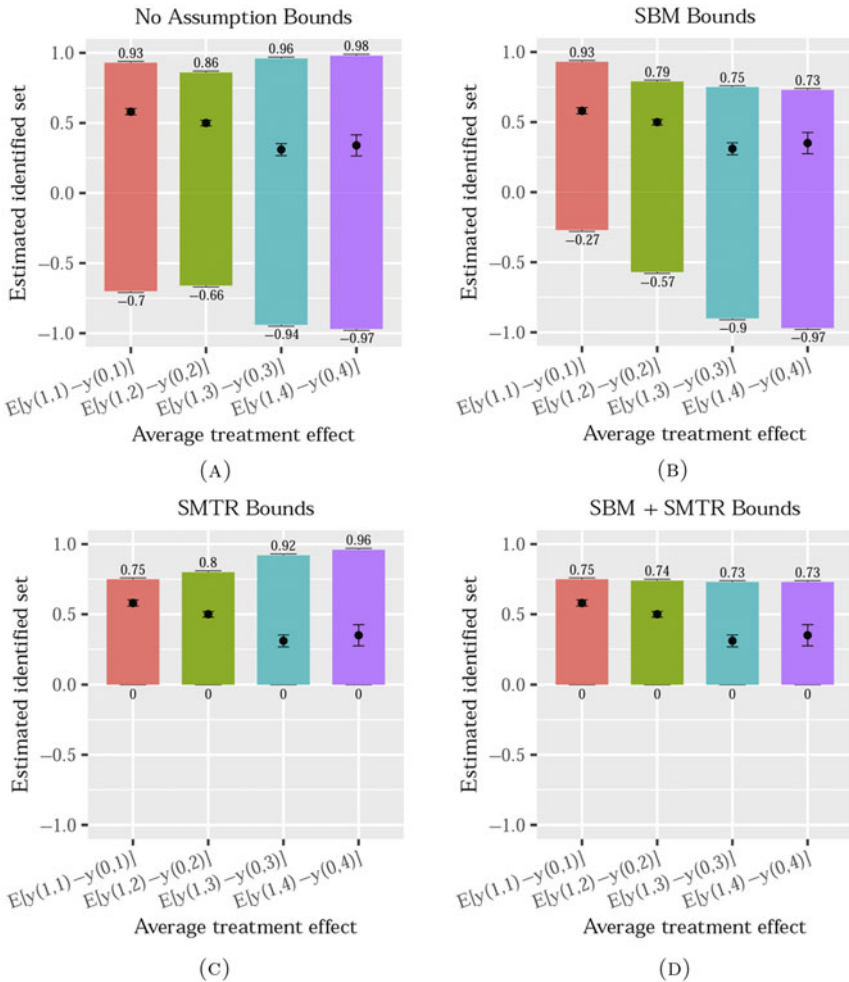
APPENDIX A: Figures

FIGURE 1. Bounds on the effects of commercial zoning on future commercial use.
Bounds on the effect of 1923 commercial zoning on the probability of commercial use in 2005 under a variety of shape restrictions. The commercial zoning effect is estimated separately for all four levels of 1923 density zoning; colors denote the level of density zoning. Upper and lower bounds are accompanied by upper and lower 95% confidence bounds, respectively. Confidence intervals estimated using a standard bootstrap. Point estimates from a naive OLS regression are included along with 95% confidence intervals. Panel A shows the no assumption bounds. Panel B adds the assumption of submodularity (SBM). Panel C adds the assumption of semimonotone treatment response (SMTR). Panel D combines SBM and SMTR.

APPENDIX B: Proofs

Proof of Lemma 1. I first show that $t_1 \wedge t_5 = t_2$:

$$t_5 \wedge t_4 = t_3$$
$$\implies t_1 \wedge (t_4 \wedge t_5) = t_1 \wedge t_3$$
$$\implies (t_1 \wedge t_4) \wedge t_5 = t_2$$
$$\implies t_1 \wedge t_5 = t_2,$$

where the last implication follows from the fact that $t_1 \leq t_4$. Now, I show that $t_1 \vee t_5 = t_6$:

$$t_4 \vee t_5 = t_6$$
$$\implies (t_1 \vee t_3) \vee t_5 = t_6$$
$$\implies t_1 \vee (t_3 \vee t_5) = t_6$$
$$\implies t_1 \vee t_5 = t_6,$$

where the last implication follows from the fact that $t_3 \leq t_5$. ∎

Proof of Proposition 3. This proof requires only a minor modification of the proof of Proposition 2. Following Manski (1997), SMTR implies $y(t_4) \leq y(t_3)$ and thus $y(t_3) - y(t_4) > 0$ for all $t_4 < t_3$, and this lower bound is sharp under SMTR alone. The lower bounds in (3.28) take the form $y(z) - \overline{K}$, $\underline{K} - y(z)$, and $\underline{K} - \overline{K}$; these are necessarily weakly negative regardless of the value of $y(z)$, so they do not improve upon the SMTR lower bound. Thus, a sharp lower bound for the individual treatment effect of interest is

$$y(t_1) - y(t_2) \geq 0.$$

Turning to the upper bounds, if $z \geq t_1$, then SMTR implies

$$y(t_1) - y(t_2) \leq y(z) - \underline{K}.$$

This is the same upper bound implied by SPM when $z \in \Lambda_1 \cup \Lambda_2$, so SMTR has no additional identification power in this case. If $z \geq t_1$ and $z \in \Lambda_3 \cup \Lambda_{13}$, SMTR generally improves upon the no-assumption upper bound $\overline{K} - \underline{K}$. If $z \leq t_2$, then SMTR implies

$$y(t_1) - y(t_2) \leq \overline{K} - y(z).$$

If $z \in \Lambda_4 \cup \Lambda_6$, SBM implies the same upper bound, so SMTR has no additional identifying power. If $z \leq t_2$ and $z \in \Lambda_5 \cup \Lambda_{13}$, SMTR improves upon the no-assumption upper bound. If $z \in \Lambda_{10} \cup \Lambda_{11}$, then SMTR has no implications for the upper bound on $y(t_1) - y(t_2)$, but SPM does via 3.28. If $z \in \Lambda_7 \cup \Lambda_9$, SMTR also has no implications for the upper bound on $y(t_1) - y(t_2)$, but SBM weakly improves upon it, again via 3.28. Combining these results yields

$$y(t_1) - y(t_2) \in \begin{cases} [0, \overline{K} - y(z)] & \text{if } z \in \{t_3 \mid t_3 \leq t_2\} \cup \Lambda_{10} \cup \Lambda_{11} \\ [0, y(z) - \underline{K}] & \text{if } z \in \{t_3 \mid t_1 \leq t_3\} \cup \Lambda_7 \cup \Lambda_9 \\ [0, \overline{K} - \underline{K}] & \text{if } z \in \{t_3 \mid t_1 \not\leq t_3 \text{ and } t_3 \not\leq t_2\} \cap (\Lambda_8 \cup \Lambda_{12} \cup \Lambda_{13}) \end{cases}.$$

$$\textbf{(B.1)}$$

Averaging over i yields the result in (3.29). ∎

Proof of Proposition 4. In the absence of other assumptions, the bounds

$$\underline{B}(t_1, x) \le \mathrm{E}[y(t_1) \mid x] \le \overline{B}(t_1, x)$$

and

$$\underline{B}(t_2, x) \le \mathrm{E}[y(t_2) \mid x] \le \overline{B}(t_2, x)$$

and thus

$$\underline{B}(t_1, x) - \overline{B}(t_2, x) \le \mathrm{E}[y(t_1) \mid x] - \mathrm{E}[y(t_2) \mid x] \le \overline{B}(t_1, x) - \underline{B}(t_1, x)$$

are sharp for all $x \in X$. The assumption that x^k is an SPMIV for $\mathrm{E}[y(t_1) - y(t_2) \mid x^k, x^{-k}]$ implies that

$$\underline{B}(t_1, x_2^k, x^{-k}) - \overline{B}(t_2, x_2^k, x^{-k}) \le \mathrm{E}[y(t_1) - y(t_2) \mid x_1^k, x^{-k}]$$

for all $x_2^k \le x_1^k$ and

$$\mathrm{E}[y(t_1) - y(t_2) \mid x_1^k, x^{-k}] \le \underline{B}(t_1, x_2^k, x^{-k}) - \overline{B}(t_2, x_2^k, x^{-k})$$

for all $x_1^k \le x_2^k$. The result follows.   ■

Proof of Proposition 1. Proposition 4 implies that the bounds

$$\sup_{x_2^k \le x_1^k} \{\underline{B}(t_1, x_2^k, x^{-k}) - \overline{B}(t_2, x_2^k, x^{-k})\}$$

$$\le \mathrm{E}[y(t_1) - y(t_2) \mid x_1^k, x^{-k}]$$

$$\le \inf_{x_1^k \le x_2^k} \{\overline{B}(t_1, x_2^k, x^{-k}) - \underline{B}(t_2, x_2^k, x^{-k})\}$$

are sharp. Thus, $\mathrm{E}[y(t_1) \mid x]$ and $\mathrm{E}[y(t_2) \mid x]$ must simultaneously satisfy the no-assumption bounds

$$\underline{B}(t_1, x) \le \mathrm{E}[y(t_1) \mid x] \le \overline{B}(t_1, x) \tag{B.2}$$

and

$$\underline{B}(t_2, x) \le \mathrm{E}[y(t_2) \mid x] \le \overline{B}(t_2, x) \tag{B.3}$$

as well as

$$\sup_{x_2^k \le x_1^k} \{\underline{B}(t_1, x_2^k, x^{-k}) - \overline{B}(t_2, x_2^k, x^{-k})\} + \mathrm{E}[y(t_2) \mid x_1^k, x^{-k}]$$

$$\le \mathrm{E}[y(t_1) \mid x_1^k, x^{-k}]$$

$$\le \inf_{x_1^k \le x_2^k} \{\overline{B}(t_1, x_2^k, x^{-k}) - \underline{B}(t_2, x_2^k, x^{-k})\} + \mathrm{E}[y(t_2) \mid x_1^k, x^{-k}] \tag{B.4}$$

and

$$\mathrm{E}[y(t_1) \mid x_1^k, x^{-k}] - \inf_{x_1^k \le x_2^k} \{\overline{B}(t_1, x_2^k, x^{-k}) - \underline{B}(t_2, x_2^k, x^{-k})\}$$

$$\le \mathrm{E}[y(t_2) \mid x_1^k, x^{-k}]$$

$$\le \mathrm{E}[y(t_1) \mid x_1^k, x^{-k}] - \sup_{x_2^k \le x_1^k} \{\underline{B}(t_1, x_2^k, x^{-k}) - \overline{B}(t_2, x_2^k, x^{-k})\}. \tag{B.5}$$

From (B.2)–(B.5), it is clear that

$$\max\left\{\underline{B}(t_1,x_1^k,x^{-k}),\ \sup_{x_2^k\le x_1^k}\ \{\underline{B}(t_1,x_2^k,x^{-k})-\overline{B}(t_2,x_2^k,x^{-k})\}+\underline{B}(t_2,x_1^k,x^{-k})\right\}$$

$$\le \mathrm{E}\left[\,y(t_1)\mid x_1^k,x^{-k}\,\right]$$

$$\le \min\left\{\overline{B}(t_1,x_1^k,x^{-k}),\ \inf_{x_1^k\le x_2^k}\ \{\overline{B}(t_1,x_2^k,x^{-k})-\underline{B}(t_2,x_2^k,x^{-k})\}+\overline{B}(t_2,x_1^k,x^{-k})\right\}$$

and

$$\max\left\{\underline{B}(t_2,x_1^k,x^{-k}),\ \underline{B}(t_1,x_1^k,x^{-k})-\inf_{x_1^k\le x_2^k}\ \{\overline{B}(t_1,x_2^k,x^{-k})-\underline{B}(t_2,x_2^k,x^{-k})\}\right\}$$

$$\le \mathrm{E}\left[\,y(t_2)\mid x_1^k,x^{-k}\,\right]$$

$$\le \min\left\{\overline{B}(t_2,x_1^k,x^{-k}),\ \overline{B}(t_1,x_1^k,x^{-k})-\sup_{x_2^k\le x_1^k}\ \{\underline{B}(t_1,x_2^k,x^{-k})-\overline{B}(t_2,x_2^k,x^{-k})\}\right\}$$

must hold. I show that these bounds are feasible, i.e., consistent with (4.3), whence it follows that they are sharp. Consider the following events:

$$\max\left\{\underline{B}(t_1,x_1^k,x^{-k}),\ \sup_{x_2^k\le x_1^k}\ \{\underline{B}(t_1,x_2^k,x^{-k})-\overline{B}(t_2,x_2^k,x^{-k})\}+\underline{B}(t_2,x_1^k,x^{-k})\right\}$$

$$= \sup_{x_2^k\le x_1^k}\ \{\underline{B}(t_1,x_2^k,x^{-k})-\overline{B}(t_2,x_2^k,x^{-k})\}+\underline{B}(t_2,x_1^k,x^{-k}) > \underline{B}(t_1,x_1^k,x^{-k}),\quad \textbf{(B.6)}$$

$$\min\left\{\overline{B}(t_1,x_1^k,x^{-k}),\ \inf_{x_1^k\le x_2^k}\ \{\overline{B}(t_1,x_2^k,x^{-k})-\underline{B}(t_2,x_2^k,x^{-k})\}+\overline{B}(t_2,x_1^k,x^{-k})\right\}$$

$$= \inf_{x_1^k\le x_2^k}\ \{\overline{B}(t_1,x_2^k,x^{-k})-\underline{B}(t_2,x_2^k,x^{-k})\}+\overline{B}(t_2,x_1^k,x^{-k}) < \overline{B}(t_1,x_1^k,x^{-k}),\quad \textbf{(B.7)}$$

$$\max\left\{\underline{B}(t_2,x_1^k,x^{-k}),\ \underline{B}(t_1,x_1^k,x^{-k})-\inf_{x_1^k\le x_2^k}\ \{\overline{B}(t_1,x_2^k,x^{-k})-\underline{B}(t_2,x_2^k,x^{-k})\}\right\}$$

$$= \underline{B}(t_1,x_1^k,x^{-k})-\inf_{x_1^k\le x_2^k}\ \{\overline{B}(t_1,x_2^k,x^{-k})-\underline{B}(t_2,x_2^k,x^{-k})\} > \underline{B}(t_2,x_1^k,x^{-k}),\quad \textbf{(B.8)}$$

$$\min\left\{\overline{B}(t_2,x_1^k,x^{-k}),\ \overline{B}(t_1,x)-\sup_{x_2^k\le x_1^k}\ \{\underline{B}(t_1,x_2^k,x^{-k})-\overline{B}(t_2,x_2^k,x^{-k})\}\right\}$$

$$= \overline{B}(t_1,x_1^k,x^{-k})-\sup_{x_2^k\le x_1^k}\ \{\underline{B}(t_1,x_2^k,x^{-k})-\overline{B}(t_2,x_2^k,x^{-k})\} < \overline{B}(t_2,x_1^k,x^{-k}).\quad \textbf{(B.9)}$$

It is easy to show that (B.6) $\implies \neg$ (B.8); thus, the lower bounds in (4.4) and (4.5) are consistent with (4.3). Similarly, (B.7) $\implies \neg$ (B.9), and so the upper bounds in (4.4) and (4.5) are consistent with (4.3). ∎

Proof of Proposition 5. The proof of this result is analogous to that of Proposition 9.2 of Manski (2003), where he derived sharp bounds on average potential outcomes using monotone treatment response and a monotone instrumental variable. Addressing the first portion of the claim, Proposition 2 shows that

$$\underline{\Delta B}(t_1,t_2,x_1^k,x^{-k}) \le \mathrm{E}[\,y(t_1)-y(t_2)\mid x_1^k,x^{-k}\,] \le \overline{\Delta B}(t_1,t_2,x_1^k,x^{-k}) \qquad \textbf{(B.10)}$$

and SPMIV implies that

$$E\big[y(t_1) - y(t_2) \mid x_2^k, x^{-k}\big] \leq E\big[y(t_1) - y(t_2) \mid x_1^k, x^{-k}\big] \leq E\big[y(t_1) - y(t_2) \mid x_3^k, x^{-k}\big],$$
(B.11)

when $x_2 \leq x_1 \leq x_3$. The result follows. The derivation is analogous for the second portion of the claim. ∎

Proof of Proposition 6. For a lattice $T = \{t_1, t_2, t_1 \vee t_2, t_1 \wedge t_2\}$ which is not a chain, SPM implies the following inequalities:

$$y(t_2) - y(t_1 \wedge t_2) \leq y(t_1 \vee t_2) - y(t_1),$$
$$y(t_1) - y(t_1 \wedge t_2) \leq y(t_1 \vee t_2) - y(t_2).$$

Since these inequalities hold for all individuals, they imply the following first-order stochastic dominance relationships:

$$F_{t_1 \vee t_2, t_1}(w) \leq F_{t_2, t_1 \wedge t_2}(w),$$
$$F_{t_1 \vee t_2, t_2}(w) \leq F_{t_1, t_1 \wedge t_2}(w).$$

These imply the following bounds on quantile functions:

$$F_{t_2, t_1 \wedge t_2}^{-1}(q) \leq F_{t_1 \vee t_2, t_1}^{-1}(q),$$
(B.12)
$$F_{t_1, t_1 \wedge t_2}^{-1}(q) \leq F_{t_1 \vee t_2, t_2}^{-1}(q).$$
(B.13)

Combining (B.12) and (B.13) with the bounds from Williamson and Downs (1990) yields the result. ∎

Proof of Proposition 7. Trivial. ∎