

A Stochastic Monotonicity Assumption for the Instrumental Variables Method

Dylan S. Small

*Department of Statistics, The Wharton School of the University of Pennsylvania,
3730 Walnut St., Philadelphia, PA 19104, U.S.A.*

dsmall@wharton.upenn.edu

Zhiqiang Tan

*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health,
615 North Wolfe St., Baltimore, MD 21205, U.S.A.*

ztan@jhsph.edu

January 4, 2007

Summary. The instrumental variables (IV) method is a method for making causal inferences about the effect of a treatment based on an observational study in which there are unmeasured confounding variables. The method requires a valid IV, a variable that is independent of the unmeasured confounding variables and is associated with the treatment but which has no effect on the outcome beyond its effect on the treatment. An additional assumption that is often made for the IV method is deterministic monotonicity, which is an assumption that for each subject, the level of the treatment that a subject would take if given a level of the IV is a monotonic increasing function of the level of the IV. Under deterministic monotonicity, the IV method has certain desirable properties including that it identifies the sign of the average treatment effect when the sign of each subject's treatment effect is the same (the no sign reversal property). However, deterministic monotonicity is sometimes not

realistic. In particular, we show that deterministic monotonicity is sometimes not plausible for the use of a clinician’s preference for different treatments as an IV for comparing medical treatments. We introduce a stochastic monotonicity condition which does not require that a monotonic increasing relationship hold within subjects between the levels of the IV and the level of the treatment that the subject would take if given a level of the IV, but only that a monotonic increasing relationship hold across subjects between the IV and the treatment in a certain manner. We show that under stochastic monotonicity, certain desirable properties of the IV method for deterministic monotonicity still hold, including the no sign reversal property. For the clinician preference IV example, we give plausible conditions for stochastic monotonicity to hold.

Keywords: Causal inference; Wald estimator; Observational study; Clinician preference instrumental variable; Confounding by indication.

1. Introduction

1.1. The IV method and the deterministic monotonicity condition

The instrumental variables (IV) method is designed to estimate the effects of treatments when the level of the treatment is confounded by unmeasured covariates that cannot be controlled by adjustments. The method requires a valid IV, a variable that is independent of the unmeasured confounding variables and that manipulates a treatment but which does not have any effect beyond its manipulation of the treatment. For a review of many applications of the IV method, see Angrist and Krueger (2001).

We consider a binary IV and a binary treatment. For this setting, a usual estimator associated with the IV method is the Wald (1940) estimator. If all subjects have the same treatment effect and the IV is valid, then the Wald estimator is a consistent estimator of the average treatment effect (Holland, 1988). However, constant treatment effects are often an implausible assumption (Imbens and Angrist, 1994; Angrist, Imbens and Rubin, 1996; Hérnan and Robins, 2006). When treatment effects are heterogeneous, Imbens and Angrist (1994) and Angrist *et al.* (1996) (hereafter AIR) showed that the Wald estimator has a disturbing *sign reversal* property: it is possible for the treatment effect to be positive for every subject but for the Wald estimator to converge in probability to a negative number. Because sign reversal is an undesirable property, Imbens and Angrist formulated a condition,

which we call *deterministic monotonicity*, under which the Wald estimator converges to the average treatment effect for a certain subpopulation and hence satisfies the *no sign reversal* property:

Definition 1. *An estimator satisfies the no sign reversal property if whenever the sign of the treatment effects (+, 0 or -) is the same for every subject in the population, then the sign of the estimator converges in probability to the sign of the treatment effects.*

When the Wald estimator satisfies the no sign reversal property and the sign of the treatment effect is the same for every subject, then a valid IV enables the sign of the treatment effects to be consistently estimated.

Suppose that the higher level of the IV provides more encouragement than the lower level of the IV for a subject to take the higher level of the treatment. The deterministic monotonicity condition is that for *all* subjects, the level of the treatment the subject would take if given a level of the IV is a monotone increasing function of the level of the IV. Deterministic monotonicity is not realistic in some settings in which the IV method is commonly used, such as comparing medical treatments using clinician preference as an IV (see Section 1.3). In this paper, we introduce a *stochastic monotonicity* condition, a relaxation of deterministic monotonicity, that does not require the monotone increasing relationship between the IV and treatment taken to hold within subjects for all subjects but only that a monotone increasing relationship hold across subjects in a certain manner. We show that under stochastic monotonicity, the Wald estimator converges in probability to a weighted average of the treatment effects in the population and consequently satisfies the no sign reversal property. Because these properties of the Wald estimator are highly desirable, our work expands the scope of settings in which the IV method is known to be useful. First, we review the AIR framework for the IV method and the Wald estimator.

1.2. Review: The AIR Framework for the IV Method and the Wald Estimator

AIR formulate assumptions for the IV method in the Neyman-Rubin potential outcomes framework (Neyman, 1923; Rubin, 1974). We consider here a binary treatment D with levels 0 and 1, a binary IV Z with levels 0 and 1 and an outcome Y that could be continuous or discrete. We assume that we have an i.i.d. sample from a population. Let $Y_{z,d=0}(\omega)$ denote

the outcome that a subject ω would experience if the subject were given level z of the IV and took treatment level 0, $Y_{z,d=1}(\omega)$ denote the outcome that a subject ω would experience if the subject were given level z of the IV and took treatment level 1 and $Y(\omega)$ denote the observed outcome; the $Y_{z,d}(\omega)$'s are called potential outcomes. Also, let $D_{z=0}(\omega)$ denote the treatment that a subject ω would take if the subject were given level 0 of the IV, $D_{z=1}(\omega)$ denote the treatment that a subject ω would take if the subject were given level 1 of the IV and $D(\omega)$ denote the observed treatment; the $D_z(\omega)$'s are called potential treatment takens. Implicit in this notation for the potential outcomes and potential treatment takens is the stable unit treatment value assumption (Rubin, 1986) that the potential outcomes and potential treatment takens for subject ω are affected only by the level of the IV and treatment taken for subject ω and not by the IV and treatment taken for any other subject ω' . We let level 1 of the IV be the level that encourages the subject to take level 1 of the treatment, i.e., $P(D_{z=1} = 1) \geq P(D_{z=0} = 1)$. AIR classify subjects by how they would respond in terms of treatment taken to the encouragement and no encouragement levels of the IV: subjects with $D_{z=1}(\omega) = 1, D_{z=0}(\omega) = 1$ are called always takers; subjects with $D_{z=1}(\omega) = 1, D_{z=0}(\omega) = 0$ are called compliers; subjects with $D_{z=1}(\omega) = 0, D_{z=0}(\omega) = 1$ are called defiers; and subjects with $D_{z=1}(\omega) = 0, D_{z=0}(\omega) = 0$ are called never takers.

AIR make four assumptions (in addition to the stable unit treatment value assumption) for a variable Z to be a valid IV. We use the first three assumptions in our framework and call a variable satisfying these three assumptions a valid IV, but present an alternative to the fourth assumption.

1. *Assumption A1: Ignorability of the IV:* $(Y_{z=0,d=0}, Y_{z=0,d=1}, Y_{z=1,d=0}, Y_{z=1,d=1}, D_{z=0}, D_{z=1}) \perp\!\!\!\perp Z$, where $\perp\!\!\!\perp$ denotes independence. This ignorability might require conditioning on a covariate vector \mathbf{X} , i.e., $(Y_{z=0,d=0}, Y_{z=0,d=1}, Y_{z=1,d=0}, Y_{z=1,d=1}, D_{z=0}, D_{z=1}) \perp\!\!\!\perp Z | \mathbf{X}$. We suppress such conditioning throughout the rest of the paper to reduce the complexity of notation.
2. *Assumption A2: Nonzero average causal effect of Z on D :* $E[D_{z=1} - D_{z=0}] > 0$. This assumption states that level 1 of the IV on average provides more encouragement than level 0 of the IV for subjects to take level 1 of the treatment.
3. *Assumption A3: Exclusion Restriction:* $Y_{z,d}(\omega) = Y_{z',d}(\omega)$ for all z, z', d and all subjects

ω . This assumption states that the IV has no direct effect on the outcome besides its effect on encouraging or not encouraging the subject to take level 1 of the treatment. Under the exclusion restriction assumption, we can write $Y_{z=0,d=0}(\omega) = Y_{z=1,d=0}(\omega) \equiv Y_{d=0}(\omega)$ and $Y_{z=0,d=1}(\omega) = Y_{z=1,d=1}(\omega) \equiv Y_{d=1}(\omega)$. The treatment effect for subject ω is $Y_{d=1}(\omega) - Y_{d=0}(\omega)$.

4. *Assumption A4: Deterministic Monotonicity:* $D_{z=1}(\omega) \geq D_{z=0}(\omega)$ for every subject ω , i.e., there are no defiers.

Vytlacil (2002) shows that Assumptions A1-A4 are equivalent to the assumptions of a latent index model; see also Tan (2005) for related results.

The Wald estimator of the average treatment effect is

$$\frac{\hat{E}(Y|Z = 1) - \hat{E}(Y|Z = 0)}{\hat{E}(D|Z = 1) - \hat{E}(D|Z = 0)}, \quad (1)$$

where \hat{E} denotes the sample mean. Imbens and Angrist (1994) show that under Assumptions A1-A4, the Wald estimand converges in probability to $E[Y_{d=1} - Y_{d=0} | D_{z=1} - D_{z=0} = 1]$, i.e., the average treatment effect for the subpopulation of compliers. Consequently, the Wald estimator satisfies the no sign reversal property under Assumptions A1-A4. If Assumptions A1-A3 hold but Assumption A4 (deterministic monotonicity) does not hold, then Imbens and Angrist (1994) show that the Wald estimator does not satisfy the no sign reversal property. For example, if the treatment effect is positive for every subject (i.e., $Y_{d=1}(\omega) - Y_{d=0}(\omega) > 0$ for every ω), but the average treatment effect for defiers is greater than the average treatment effect for compliers, then it is possible for the Wald estimator to converge to a negative number. We now present an example in which deterministic monotonicity is implausible.

1.3. An example in which deterministic monotonicity is implausible: comparing medical treatments using physician preference as an IV

Observational studies to compare the intended effects of medical treatments are difficult to conduct because of confounding by indication – a clinician’s decision to treat differently two patients who have the same recorded (in the data set) characteristics may be influenced by unrecorded clinical indicators (Miettinen, 1983). Confounding by indication is particularly problematic in studies comparing treatments based on healthcare claims data that

lack detailed clinical indicators (Brookhart *et al.*, 2006). To overcome the problem of confounding by indication, Korn and Baumrind (1998) propose to use a clinician’s preference for one treatment versus the others as an IV; see also Brookhart *et al.* (2006) and Hernán and Robins (2006) for further applications and discussion. Korn and Baumrind specifically consider using clinician preference as an IV for comparing the effect of tooth extraction versus nonextraction for treating orthodontic patients with crowding and irregularities of their teeth and jaws. To illustrate main points, we consider here a study in which there are two orthodontists, A and B , and the IV Z equals 1 if the patient was treated by orthodontist A and 0 if treated by orthodontist B ; we consider issues arising from there being more than two orthodontists in the study in Section 2.3. The ignorability of the IV assumption (Assumption A1) for Z to be a valid IV corresponds to the orthodontists seeing comparable patient populations. This is reasonable in some settings such as a university clinic (Korn and Baumrind, 1988). The nonzero average causal effect of Z on D assumption (Assumption A2) corresponds to one orthodontist having a greater preference for tooth extraction than the other. Baumrind, Korn, Boyd and Maxwell (1996a) found strong evidence that orthodontists differ in their preferences for extraction. The exclusion restriction assumption (Assumption A3) means that if both orthodontists were to treat a patient in the same way, the outcome would be the same; this implies that the orthodontists are of equal skill. The deterministic monotonicity assumption (Assumption A4) corresponds to the following. Suppose orthodontist A has a greater preference for tooth extraction than orthodontist B . Then deterministic monotonicity means that there is no patient whom orthodontist A would treat with nonextraction but orthodontist B would treat with extraction. Korn and Baumrind (1998) found evidence that deterministic monotonicity was violated among orthodontists. They asked orthodontists to independently decide whether they preferred tooth extraction or nonextraction for different patients. In 15 out of 16 pairs of orthodontists considered, there was at least one patient for whom orthodontist A preferred extraction and orthodontist B preferred nonextraction and one patient for whom orthodontist B preferred extraction and orthodontist A preferred nonextraction (where the total number of patients evaluated by each pair ranged from 32 to 58).

One reason why deterministic monotonicity might fail to hold for the clinician preference

IV is that clinicians might use multiple patient characteristics in deciding between treatments and place different emphases on these characteristics. For example, Baumrind, Korn, Boyd and Maxwell (1996b) found that among orthodontists making decisions on whether or not to extract teeth, the factors crowding, incisor protrusion, improvement of profile, Class II severity/anteroposterior discrepancy and stability of outcome all played substantial roles, and the importance put on these factors differed among orthodontists. To illustrate how differences in the emphases clinicians place on different patient factors can destroy deterministic monotonicity, consider Figure 1. The figure plots the treatment preferences of two clinicians based on two patient characteristics I and II that range from 0 to 1. Clinician A will treat a patient with treatment level 1 if the patients' characteristics are to the right and above the dashed lines and otherwise treat the patient with treatment level 0. Clinician B will treat a patient with treatment level 1 if the patients' characteristics are to the right and above the dotted lines and otherwise treat the patient with treatment level 0. Suppose patients' characteristics are uniformly distributed on the unit square. Then clinician A will treat 75% of patients with treatment level 1 and clinician B will treat 25% of patients with treatment level 1, but for patients with characteristics in the rectangle D , clinician B will treat them with treatment level 1 and clinician A will treat them with treatment level 0.

INSERT FIGURE 1 HERE

Concerns about deterministic monotonicity being violated have been raised in several other settings. Ten Have *et al.* (2004) present evidence that deterministic monotonicity is violated in a randomized physician encouragement trial for treating depression. Heckman and Vytlačil (2005) discuss economic applications of the IV method in which deterministic monotonicity is likely to be violated. Dawid (2000) expresses general concern about the fatalism underlying the deterministic monotonicity condition. In the rest of this paper, we formulate a stochastic monotonicity condition which is plausible for the clinician preference IV described above and other settings, and show the desirable properties of the IV method under the stochastic monotonicity condition.

2. Stochastic Monotonicity Condition

Our stochastic monotonicity condition concerns the relationship between D , Z and a latent variable \mathbf{U} (\mathbf{U} could be a vector) that is assumed to satisfy certain conditions. The

latent variable \mathbf{U} is assumed to satisfy (a) $\mathbf{U}, Y_{d=1}, Y_{d=0}, D_{z=1}, D_{z=0}$ are jointly independent of Z , i.e.,

$$\mathbf{U}, Y_{d=1}, Y_{d=0}, D_{z=1}, D_{z=0} \perp\!\!\!\perp Z \quad (2)$$

and (b) conditional on \mathbf{U} , the potential treatment taken are independent of the potential outcomes,

$$Y_{d=1}, Y_{d=0} \perp\!\!\!\perp D_{z=1}, D_{z=0} | \mathbf{U}. \quad (3)$$

One choice of \mathbf{U} that always satisfies (2)-(3) under Assumption A1 is $\mathbf{U} = (Y_{d=0}, Y_{d=1})$.

Our stochastic monotonicity condition is the following condition about $(P(D_{z=1} | \mathbf{U}), P(D_{z=0} | \mathbf{U}))$ for a \mathbf{U} that satisfies (2)-(3):

$$P(D_{z=1} = 1 | \mathbf{U} = \mathbf{u}) \geq P(D_{z=0} = 1 | \mathbf{U} = \mathbf{u}) \text{ for all } \mathbf{u}. \quad (4)$$

The stochastic monotonicity condition (4) states that for every strata of the latent variable \mathbf{U} , the proportion of subjects who would take the higher level of the treatment if encouraged (subjects with $D_{z=1} = 1$) is at least as high as the proportion of subjects who would take the higher level of the treatment if not encouraged (subjects with $D_{z=0} = 1$). Unlike deterministic monotonicity, stochastic monotonicity does not require that a monotone increasing relationship (between the levels of the IV and the level of the treatment that the subject would take if given a level of the IV) hold within subjects for all subjects, but instead only requires that a monotone increasing relationship hold across subjects within the strata of a latent variable. Note that $P(D_{z=1} = 1 | \mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1 | \mathbf{U} = \mathbf{u})$ is the proportion of compliers minus the proportion of defiers in the stratum $\mathbf{U} = \mathbf{u}$. The stochastic monotonicity condition assumes $P(D_{z=1} = 1 | \mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1 | \mathbf{U} = \mathbf{u}) \geq 0$. Thus, stochastic monotonicity permits defiers, but requires at least as many compliers as defiers within each stratum of \mathbf{U} . Stochastic monotonicity includes deterministic monotonicity as a special case: under deterministic monotonicity, $D_{z=1}(\omega) \geq D_{z=0}(\omega)$ for all subjects ω ; consequently, for any \mathbf{U} that satisfies (2)-(3) ($\mathbf{U} = (Y_{d=1}, Y_{d=0})$ always satisfies (2)-(3) under Assumption A1), we have $P(D_{z=1} = 1 | \mathbf{U} = \mathbf{u}) \geq P(D_{z=0} = 1 | \mathbf{U} = \mathbf{u})$ for all \mathbf{u} .

2.1. Implications of stochastic monotonicity condition

Suppose a proposed IV Z satisfies the stochastic monotonicity condition (4) for some \mathbf{U} as well as Assumptions A1-A3. We now derive two implications: (1) the marginal distributions

of the potential outcomes $Y_{d=1}$ and $Y_{d=0}$ from a certain biased sample from the population are identified and (2) the Wald estimator converges in probability to a weighted average of treatment effects in the population and consequently the Wald estimator satisfies the no sign reversal property.

Consider a biased sample from the population with sampling weight for a unit with $\mathbf{U} = \mathbf{u}$ proportional to $P(D_{z=1} = 1|\mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1|\mathbf{U} = \mathbf{u})$ (the proportion of compliers minus the proportion of defiers in the stratum $\mathbf{U} = \mathbf{u}$). Under stochastic monotonicity, we have $P(D_{z=1} = 1|\mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1|\mathbf{U} = \mathbf{u}) \geq 0$, so that these are legitimate sampling weights. We now show that the marginal distributions of potential outcomes for this biased sample are identified.

Theorem 1. *When Assumptions A1-A3 hold and there is a latent variable \mathbf{U} that satisfies (2)-(3) and the stochastic monotonicity condition (4), the marginal distribution of the potential outcomes from a biased sample from the population with weight for a unit with $\mathbf{U} = \mathbf{u}$ proportional to $P(D_{z=1} = 1|\mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1|\mathbf{U} = \mathbf{u})$ are identified from the distribution of the observables (Y, D, Z) as follows. Let G_1 and G_0 denote the CDFs of $Y_{d=1}$ and $Y_{d=0}$ respectively when $(Y_{d=1}, Y_{d=0})$ are obtained from a biased sample with weights for a unit with $\mathbf{U} = \mathbf{u}$ proportional to $P(D_{z=1} = 1|\mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1|\mathbf{U} = \mathbf{u})$. Then*

$$G_1(x) = \int P(Y_{d=1} \leq x|\mathbf{U} = \mathbf{u}) \frac{P(D_{z=1} = 1|\mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1|\mathbf{U} = \mathbf{u})}{\int [P(D_{z=1} = 1|\mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1|\mathbf{U} = \mathbf{u})] dF(\mathbf{u})} dF(\mathbf{u}) = \frac{1}{P(D = 1|Z = 1) - P(D = 1|Z = 0)} \left[\frac{P(D = 1, Z = 1)}{P(Z = 1)} P(Y \leq x|D = 1, Z = 1) - \frac{P(D = 1, Z = 0)}{P(Z = 0)} P(Y \leq x|D = 1, Z = 0) \right] \quad (5)$$

and

$$G_0(x) = \int P(Y_{d=0} \leq x|\mathbf{U} = \mathbf{u}) \frac{P(D_{z=1} = 1|\mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1|\mathbf{U} = \mathbf{u})}{\int [P(D_{z=1} = 1|\mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1|\mathbf{U} = \mathbf{u})] dF(\mathbf{u})} dF(\mathbf{u}) = \frac{1}{P(D = 1|Z = 1) - P(D = 1|Z = 0)} \left[\frac{P(D = 0, Z = 0)}{P(Z = 0)} P(Y \leq x|D = 0, Z = 0) - \frac{P(D = 0, Z = 1)}{P(Z = 1)} P(Y \leq x|D = 0, Z = 1) \right]. \quad (6)$$

Proof. First, note that (2)-(3) imply

$$\begin{aligned}
P(Y_{d=1} \leq x | D = d, Z = z, \mathbf{U} = \mathbf{u}) &= P(Y_{d=1} \leq x | D_z = d, Z = z, \mathbf{U} = \mathbf{u}) \\
&= P(Y_{d=1} \leq x | D_z = d, \mathbf{U} = \mathbf{u}) \\
&= P(Y_{d=1} \leq x | \mathbf{U} = \mathbf{u}).
\end{aligned} \tag{7}$$

The second equality follows from (2) and the properties of conditional expectation (Williams, 1991, pp. 88, property (k)) and the third equality follows from (3). Analogous to (7), under (2)-(3), we have

$$P(Y_{d=0} \leq x | D = d, Z = z, \mathbf{U} = \mathbf{u}) = P(Y_{d=0} \leq x | \mathbf{U} = \mathbf{u}). \tag{8}$$

Using (7)-(8), we have

$$\begin{aligned}
P(Y \leq x | D = d, Z = z) &= \\
&\int P(Y_d \leq x | D = d, Z = z, \mathbf{U} = \mathbf{u}) dF(\mathbf{u} | D = d, Z = z) = \\
&\int P(Y_d \leq x | \mathbf{U} = \mathbf{u}) dF(\mathbf{u} | D = d, Z = z) = \\
&\int P(Y_d \leq x | \mathbf{U} = \mathbf{u}) \frac{P(D = d | Z = z, \mathbf{U} = \mathbf{u}) P(Z = z | \mathbf{U} = \mathbf{u})}{P(D = d, Z = z)} dF(\mathbf{u}) = \\
&\frac{P(Z = z)}{P(D = d, Z = z)} \int P(Y_d \leq x | \mathbf{U} = \mathbf{u}) P(D_z = d | Z = z, \mathbf{U} = \mathbf{u}) dF(\mathbf{u}) = \\
&\frac{P(Z = z)}{P(D = d, Z = z)} \int P(Y_d \leq x | \mathbf{U} = \mathbf{u}) P(D_z = d | \mathbf{U} = \mathbf{u}) dF(\mathbf{u}).
\end{aligned} \tag{9}$$

The second equality follows from (7)-(8), the fourth equality follows from $\mathbf{U} \perp\!\!\!\perp Z$ from (2) and the fifth equality also follows from (2). From (9), we have

$$\begin{aligned}
&\frac{P(D = 1, Z = 1)}{P(Z = 1)} P(Y \leq x | D = 1, Z = 1) - \frac{P(D = 1, Z = 0)}{P(Z = 0)} P(Y \leq x | D = 1, Z = 0) = \\
&\int P(Y_{d=1} \leq x | \mathbf{U} = \mathbf{u}) [P(D_{z=1} = 1 | \mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1 | \mathbf{U} = \mathbf{u})] dF(\mathbf{u})
\end{aligned} \tag{10}$$

and

$$\begin{aligned}
&\frac{P(D = 0, Z = 0)}{P(Z = 0)} P(Y \leq x | D = 0, Z = 0) - \frac{P(D = 0, Z = 1)}{P(Z = 1)} P(Y \leq x | D = 0, Z = 1) = \\
&\int P(Y_{d=0} \leq x | \mathbf{U} = \mathbf{u}) [P(D_{z=1} = 1 | \mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1 | \mathbf{U} = \mathbf{u})] dF(\mathbf{u}).
\end{aligned} \tag{11}$$

Note also that using $\mathbf{U} \perp\!\!\!\perp Z$ from (2), we have

$$\begin{aligned} & P(D = 1|Z = 1) - P(D = 1|Z = 0) = \\ & \int P(D = 1|Z = 1, \mathbf{U} = \mathbf{u})dF(\mathbf{u}|Z = 1) - \int P(D = 1|Z = 0, \mathbf{U} = \mathbf{u})dF(\mathbf{u}|Z = 0) = \\ & \int [P(D = 1|Z = 1, \mathbf{U} = \mathbf{u}) - P(D = 1|Z = 0, \mathbf{U} = \mathbf{u})] dF(\mathbf{u}) \end{aligned} \quad (12)$$

The results (10), (11) and (12) together prove (5)-(6). \square

Note that rather than the joint independence of (3), Theorem 1 only requires the following pairwise conditional independence statements between the potential outcomes and potential treatment taken: $Y_{d=1} \perp\!\!\!\perp D_{z=1} | \mathbf{U}$ and $Y_{d=0} \perp\!\!\!\perp D_{z=0} | \mathbf{U}$.

From Theorem 1, we obtain the following result about the Wald estimand (the quantity the Wald estimator converges in probability to):

Corollary 1. *Under the assumptions of Theorem 1, the Wald estimand*

$$\frac{E(Y|Z = 1) - E(Y|Z = 0)}{P(D = 1|Z = 1) - P(D = 1|Z = 0)}$$

equals

$$\frac{\int E(Y_{d=1} - Y_{d=0} | \mathbf{U} = \mathbf{u}) [P(D_{z=1} = 1 | \mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1 | \mathbf{U} = \mathbf{u})] dF(\mathbf{u})}{\int [P(D_{z=1} = 1 | \mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1 | \mathbf{U} = \mathbf{u})] dF(\mathbf{u})}. \quad (13)$$

Note that (13) is the expected value of the treatment effect $Y_{d=1} - Y_{d=0}$ for a biased sample from the population with weights for a unit with $\mathbf{U} = \mathbf{u}$ proportional to $P(D_{z=1} = 1 | \mathbf{U} = \mathbf{u}) - P(D_{z=0} = 1 | \mathbf{U} = \mathbf{u})$.

Proof of Corollary 1. Using the fact that for a random variable X , $E(X) = \int_0^\infty P(X > x)dx - \int_0^\infty P(X < -x)dx$ and using the notation from Theorem 1, (13) equals

$$\int_0^\infty \{[1 - G_1(x)] - [1 - G_0(x)]\}dx - \int_0^\infty [G_1(-x) - G_0(-x)]dx. \quad (14)$$

From Theorem 1, we have

$$\begin{aligned}
G_1(x) - G_0(x) &= \tag{15} \\
&\left[\frac{1}{P(D = 1|Z = 1) - P(D = 1|Z = 0)} \right] \\
&[P(D = 1|Z = 1)P(Y \leq x|D = 1, Z = 1) + P(D = 0|Z = 1)P(Y \leq x|D = 0, Z = 1) - \\
&P(D = 1|Z = 0)P(Y \leq x|D = 1, Z = 0) - P(D = 0|Z = 0)P(Y \leq x|D = 0, Z = 0)] = \\
&\frac{P(Y \leq x|Z = 1) - P(Y \leq x|Z = 0)}{P(D = 1|Z = 1) - P(D = 1|Z = 0)}. \tag{16}
\end{aligned}$$

Thus, using (14) and (16), (13) equals

$$\begin{aligned}
&\left[\frac{1}{P(D = 1|Z = 1) - P(D = 1|Z = 0)} \right] \\
&\left[\int_0^\infty [P(Y > x|Z = 1) - P(Y > x|Z = 0)]dx - \int_0^\infty [P(Y \leq -x|Z = 1) - P(Y \leq -x|Z = 0)]dx \right] = \\
&\frac{E(Y|Z = 1) - E(Y|Z = 0)}{P(D = 1|Z = 1) - P(D = 1|Z = 0)}.
\end{aligned}$$

□

From Corollary 1, under stochastic monotonicity, the Wald estimator converges to a weighted average of the treatment effects in the population; consequently, under stochastic monotonicity, the Wald estimator satisfies the no sign reversal property.

Corollary 1 shows there is a close connection between the properties of the Wald estimator under deterministic monotonicity and stochastic monotonicity: under deterministic monotonicity, the Wald estimator converges in probability to the average treatment effect for the subpopulation of compliers, while under stochastic monotonicity, the Wald estimator converges in probability to the average treatment effect for a biased sample from the population that samples more heavily from the strata of the latent variable \mathbf{U} with a higher proportion of compliers minus proportion of defiers.

2.2. Example: Stochastic Monotonicity for the clinician preference IV in Figure 1

As an example of when stochastic monotonicity is plausible, we consider the clinician preference IV in the setting of Figure 1. Let $\mathbf{U} = (Y_{d=1}, Y_{d=0})$, for which (2)-(3) are always

satisfied under Assumption A1. Consider the following model for the potential outcomes:

$$\begin{aligned}
Y_{d=0}(\omega) &= g(c_1(\omega), c_2(\omega)) + \epsilon(\omega) \\
Y_{d=1}(\omega) &= h(c_1(\omega), c_2(\omega)) + \nu(\omega) \\
&(\epsilon, \nu) \perp\!\!\!\perp (c_1, c_2).
\end{aligned} \tag{17}$$

The following assumption ensures that stochastic monotonicity is satisfied:

$$\begin{aligned}
P((c_1, c_2) \in C | g(c_1, c_2) = x, h(c_1, c_2) = y) &\geq P((c_1, c_2) \in D | g(c_1, c_2) = x, h(c_1, c_2) = y) \\
&\text{for all } (x, y)
\end{aligned} \tag{18}$$

where C and D are the rectangles in Figure 1. Assumption (18) states that if we partition the patient characteristics (c_1, c_2) by the values of the expected potential outcomes given the patient characteristics (i.e., by $E(Y_{d=0}|c_1, c_2), E(Y_{d=1}|c_1, c_2)$), then for every set in the partition, clinician A is at least as likely as clinician B to treat a randomly chosen patient in the set with treatment level 1. Assumption (18) implies stochastic monotonicity under model (17) because for every (u, v) ,

$$\begin{aligned}
P((c_1, c_2) \in C | Y_{d=0} = u, Y_{d=1} = v) - P((c_1, c_2) \in D | Y_{d=0} = u, Y_{d=1} = v) = \\
\int \int [P((c_1, c_2) \in C | g(c_1, c_2) = u - \epsilon, h(c_1, c_2) = v - \nu) - \\
P((c_1, c_2) \in D | g(c_1, c_2) = u - \epsilon, h(c_1, c_2) = v - \nu)] dF(\epsilon, \nu) ;
\end{aligned} \tag{19}$$

the integrand of (19) is always ≥ 0 under (18).

As an example of assumption (18) holding, suppose that $g(c_1, c_2)$ and $h(c_1, c_2)$ are functions of the sum $c_1 + c_2$ and that (c_1, c_2) are uniformly distributed. Then stochastic monotonicity is satisfied for the setting of Figure 1 because for every line $\ell(x) = \{(c_1, c_2) : c_1 + c_2 = x\}$, the line segment $\ell \cap C$ is longer than the line segment $\ell \cap D$

2.3. Extension to non-binary IV

Suppose there are more than two (say K) clinicians in the clinician preference setting, and that the variable Z , taking on values $1, \dots, K$ for the the clinician a patient sees, satisfies analogues of Assumptions A1-A3, i.e., Z is independent of potential outcomes and potential treatment taken, not all clinicians have the same probability of treating the patient and

the exclusion restriction holds (implying there are no clinician skill effects). The following condition generalizes deterministic monotonicity:

$$D_{Z=K} \geq D_{Z=K-1} \geq \cdots \geq D_{Z=1}.$$

Accordingly, the following condition generalizes stochastic monotonicity:

$$P(D_{Z=K} = 1|\mathbf{U} = \mathbf{u}) \geq P(D_{Z=K-1} = 1|\mathbf{U} = \mathbf{u}) \geq \cdots \geq P(D_{Z=1} = 1|\mathbf{U} = \mathbf{u})$$

for all \mathbf{u} (20)

When the IV takes on more than two values or there is more than one IV (Z can be coded as $K - 1$ dummy variables), the two stage least squares (TSLS) estimator generalizes the Wald estimator (Angrist and Imbens, 1995). Under the generalized stochastic monotonicity condition (20), the TSLS estimator converges to a weighted average of the treatment effects in the population. This follows from our Corollary 1 and Theorem 2 of Angrist and Imbens (1995), which shows that the TSLS estimator converges to a weighted average of $\beta_{j,j-1} = (E[Y|Z = j] - E[Y|Z = j - 1]) / (E[D|Z = j] - E[D|Z = j - 1])$ for $j = 2, \dots, K$.

3. Conclusion

We have introduced a relaxation of the deterministic monotonicity condition for the IV method that does not require a monotone increasing relationship between the levels of the IV and the treatment that a subject would take if given a level of the IV to hold within subjects for all subjects, but only for a monotone increasing relationship to hold across subjects in a certain manner. We have shown that certain desirable properties of the IV method under deterministic monotonicity, such as the no sign reversal property, are preserved under our stochastic monotonicity condition.

References

- Angrist, J.D. and Imbens, G.W. (1995), “Two-stage least squares estimation of average causal effects in models with variable treatment intensity,” *Journal of the American Statistical Association*, 90, 431-442.
- Angrist, J.D., Imbens, G.W. and Rubin, D.R. (1996), “Identification of causal effects using instrumental variables (with discussion),” *Journal of the American Statistical Association*, 91, 444-472.

- Angrist, J.D. and Krueger, A.B. (2001), "Instrumental variables and the search for identification: from supply and demand to natural experiments," *Journal of Economic Perspectives*, 15, 69-85.
- Baumrind, S., Korn, E.L., Boyd, R.L and Maxwell, R. (1996a), "The decision to extract: Part I – Interclinician agreement," *American Journal of Orthodontics and Dentofacial Orthopedics*, 109, 297-309.
- Baumrind, S., Korn, E.L., Boyd, R.L and Maxwell, R. (1996b), "The decision to extract: Part II. Analysis of clinicians' stated reasons for extraction," *American Journal of Orthodontics and Dentofacial Orthopedics*, 109, 393-402.
- Brookhart, M.A., Wang, P.S., Solomon, D.H. and Schneeweiss, S. (2006), "Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable," *Epidemiology*, 17, 268-275.
- Dawid, A.P. (2000), "Causal inference without counterfactuals (with discussion)," *Journal of the American Statistical Association*, 95, 407-448.
- Heckman, J. and Vytlacil, E. (2005), "Structural equations, treatment effects, and econometric policy evaluation," *Econometrica*, 73, 669-738.
- Hernán, M.A. and Robins, J.M. (2006), "Instruments for causal inference: an epidemiologist's dream?," *Epidemiology*, 14, 360-372.
- Holland, P.W. (1988), "Causal inference, path analysis and recursive structural equation models," *Sociological Methodology*, 18, 449-484.
- Imbens, G.W. and Angrist, J.D. (1994), "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467-476.
- Imbens, G.W. and Rubin, D.B. (1997), "Estimating outcome distributions for compliers in instrumental variables models," *The Review of Economic Studies*, 64, 555-574.
- Korn, E.L. and Baumrind, S. (1998), "Clinician preferences and the estimation of causal treatment differences," *Statistical Science*, 13, 209-227.
- Miettinen, O.S. (1983), "The need for randomization in the study of intended effects," *Statistics in Medicine*, 2, 267-271.
- Neyman, J. (1923), "On the application of probability theory to agricultural experiments," trans. D. Dabrowska, *Statistical Science*, 1990, 5, 463-480.

- Rubin, D.B. (1974), "Estimating causal effects of treatments in randomized and non-randomized studies," *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D.B. (1986), "Statistics and causal inference: comment: which ifs have causal answers," *Journal of the American Statistical Association*, 81, 961-962.
- Tan, Z. (2005), "Estimation of causal effects using instrumental variables," *Journal of the American Statistical Association*, in press.
- Ten Have, T., Elliott, M., Joffe M., Zanutto, E. and Datto, C. (2004) "Causal models for randomized physician encouragement trials in treating primary care depression," *Journal of the American Statistical Association*, 99, 8-16.
- Vytlacil, E.J. (2002), "Independence, monotonicity and latent index models: an equivalence result," *Econometrica*, 70, 331-341.
- Wald, A. (1940), "The fitting of straight lines if both variables are subject to error," *Annals of Mathematical Statistics*, 11, 284-300.
- Williams, D. (1991), *Probability with martingales*, Cambridge Press: Cambridge.

Treatment Preferences of Two Clinicians

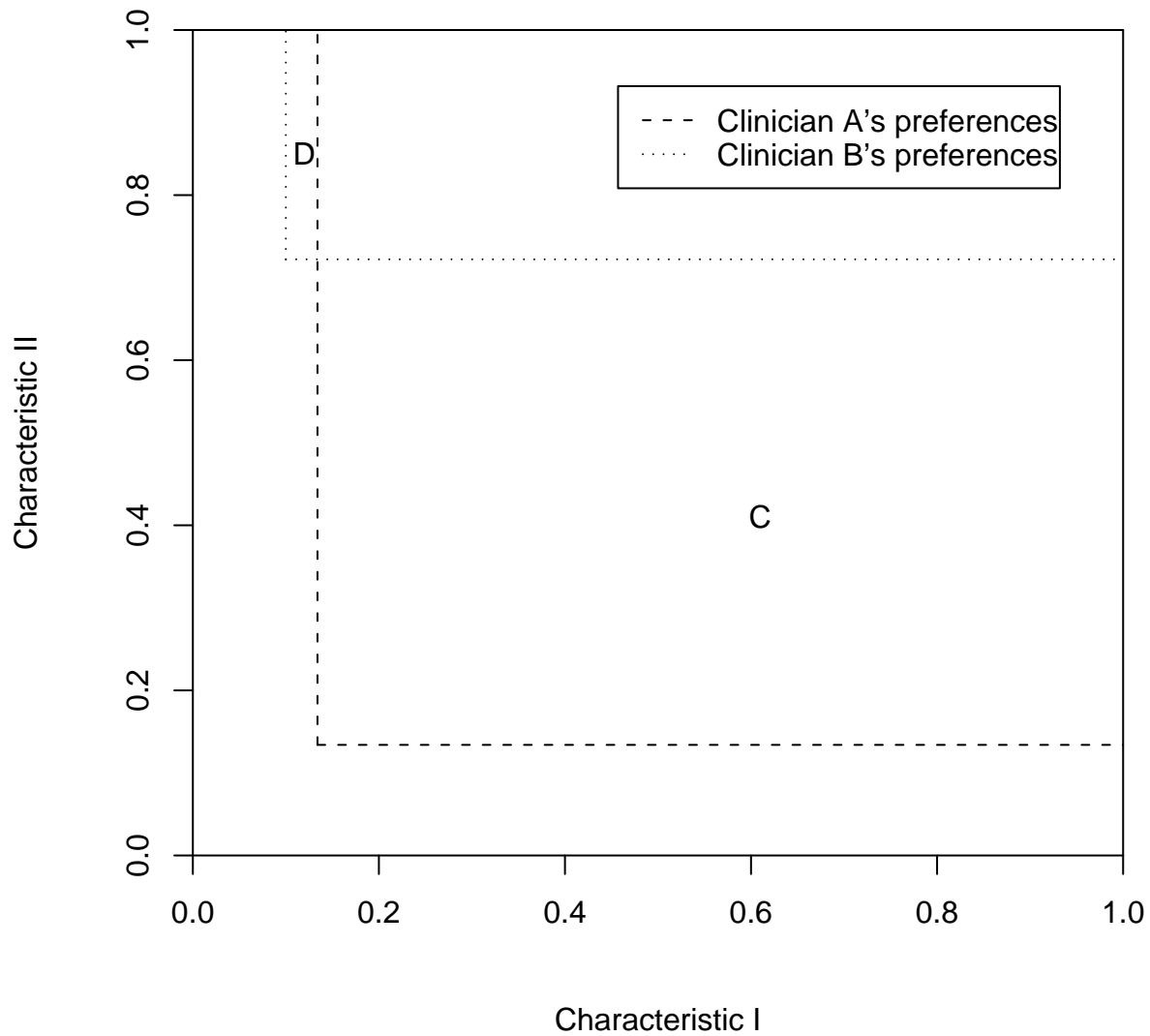


Figure 1: The figure depicts the preferences of two hypothetical clinicians for using treatment level 1 versus treatment level 0. Each clinician will treat patients with characteristics to the right and above the lines that correspond to the clinician with treatment level 1 and otherwise with treatment level 0. The dashed lines correspond to clinician *A* and the dotted lines correspond to clinician *B*. The rectangle *C* shows the compliers and the rectangle *D* shows the defiers.