Signed directed acyclic graphs for causal inference

Tyler J. VanderWeele
University of Chicago, Chicago, USA

James M. Robins
Harvard School of Public Health, Boston, USA

1Address for correspondence: Tyler J. VanderWeele, Department of Health Studies, University of Chicago, 5841 S. Maryland Ave., MC 2007, Chicago, IL, 60637, USA. E-mail: vanderweele@uchicago.edu
Summary. By introducing the notions of a monotonic effect, a weak monotonic effect and a signed edge, the directed acyclic graph causal framework can be extended so as to allow not only for the graphical representation of causal relations amongst variables but also for the sign of these causal relations. Results are developed relating monotonic effects to the sign of the causal effect of an intervention in the presence of intermediate variables. The incorporation of signed edges into the directed acyclic graph causal framework furthermore allows for the development of rules governing the relationship between monotonic effects and the sign of the covariance between two variables and also rules governing the sign of the bias which arises when control for confounding is inadequate.

Keywords: Bias; Causal inference; Confounding; Directed acyclic graphs; Structural equations.

1. Introduction

With very few exceptions (Lauritzen and Richardson, 2002), the use of graphical models in the field of causal inference has been restricted to directed acyclic graphs and graphs allowing for bidirected edges which represent unobserved common causes. The directed acyclic graph causal framework allows for the representation of causal and counterfactual relations amongst variables (Pearl, 1995; Robins, 1997; Pearl, 2000); the estimation of causal effects through the g-formula (Robins, 1986, 1987; Spirtes et al., 1993; Pearl, 1993); the detection of independencies through the d-separation criterion (Verma and Pearl, 1988; Geiger et al., 1990; Lauritzen et al., 1990); and the implementation of algorithms to determine whether conditioning on a particular set of variables, or none at all, is sufficient to control for confounding as well as algorithms which identify such a set of variables (Pearl 1995; Galles and Pearl, 1995; Pearl and Robins, 1995; Kuroki and Miyakawa, 1999, 2003; Geng et al., 2002; Tian and Pearl, 2002).

This paper introduces into the directed acyclic graph causal framework the notion of a monotonic effect and along with its graphical counterpart, a signed edge. Considerations of monotonicity have been incorporated into many types graphical models (Wellman, 1990; Archer and Wang, 1993; Bioch, 1998; Potharst and Feelders, 2002; van der Gaag et al., 2004). Here our focus will be the causal interpretation of monotonicity relationships within the directed acyclic graph framework. We will say that some variable has a positive (or negative) monotonic effect on another if intervening to increase the former will never for any individual decrease (or increase) the latter. By incorporating these notions of monotonic effects, the directed acyclic graph causal framework can be extended in various directions. Signs can be added to the edges of the directed acyclic graph to indicate the presence of a particular positive or negative monotonic effect. Using the signs of these edges, one may then determine the sign of the causal effect of an intervention in the presence of intermediate variables. Similarly one may also determine the sign of the covariance between various nodes on the signed causal directed acyclic graph and in certain circumstances one may determine the sign of the bias resulting when control for confounding is inadequate. These results generalize for the case of non-parametric structural equations corresponding results that are more straightforward in the multivariate normal setting. Although the utility of the results presented in this paper is in part limited by the strong assumptions which must be made about monotonic effects, we argue that a theory based only on average causal effects will not work. We furthermore provide a weaker set of conditions, which we express in terms of weak monotonic effects, under which the major theorems presented still hold.

Before formally introducing the concepts of a monotonic effect, a weak monotonic effect and a signed edge, we review definitions and results concerning causal directed acyclic graphs. Following Pearl (1995), a causal directed acyclic graph is a set of nodes \((X_1, ..., X_n)\) and directed edges amongst nodes such that the graph has no cycles and such that for each node \(X_i\) on the graph the corresponding variable is given by its non-parametric structural equation \(X_i = f_i(pa_i, \epsilon_i)\) where \(pa_i\) are the parents of \(X_i\) on the graph and the \(\epsilon_i\) are mutually independent. These non-parametric structural equations can be seen as a generalization of the path analysis and linear structural equation models (Pearl 1995, 2000) developed by Wright (1921) in the genetics literature and Haavelmo (1943) in the econometrics literature. Robins (1995, 2003) discusses the
close relationship between these non-parametric structural equation models and fully random-
ized causally interpreted structured tree graphs (Robins 1986, Robins 1987). Unlike their linear
counterpart, non-parametric structural equations are entirely general – \( X_i \) may depend on any
function of its parents and \( \epsilon_i \). The non-parametric structural equations encode counterfactual
relationships amongst the variables represented on the graph. The equations themselves repre-
sent one-step ahead counterfactuals with other counterfactuals given by recursive substitution.
A node \( A \) will be a parent of \( Y \) if there is some level of all variables that precede \( Y \) such that
intervening to set \( A \) to different levels will allow \( Y \) to vary even after intervening to fix all other
variables that precede \( Y \). The requirement that the \( \epsilon_i \) be mutually independent is essentially
a requirement that there is no variable absent from the graph which, if included on the graph,
would be a parent of two or more variables (Pearl, 1995, 2000).

A path is a sequence of nodes connected by edges regardless of arrowhead direction; a directed
path is a path which follows the edges in the direction indicated by the graph’s arrows. A node
\( C \) is said to be a common cause of \( A \) and \( Y \) if there exists a directed path from \( C \) to \( Y \) not
through \( A \) and a directed path from \( C \) to \( A \) not through \( Y \). A collider is a particular node on a
path such that both the preceding and subsequent nodes on the path have directed edges going
into that node i.e. both the edge to and the edge from that node have arrowheads into the node.
A path between \( A \) and \( B \) is said to be blocked given some set of variables \( Z \) if either there is
a variable in \( Z \) on the path that is not a collider or if there is a collider on the path such that
neither the collider itself nor any of its descendants are in \( Z \). It has been shown that if all paths
between \( A \) and \( B \) are blocked given \( Z \) then \( A \) and \( B \) are conditionally independent given \( Z \)
(Verma and Pearl, 1988; Geiger et al., 1990; Lauritzen et al., 1990). The directed acyclic graph
causal framework has proven to be particularly useful in determining whether conditioning on a
given set of variables, or none at all, is sufficient to control for confounding. The most important
result in this regard is the back-door path criterion (Pearl, 1995). A back-door path from some
node \( A \) to another node \( Y \) is a path which begins with a directed edge into \( A \). Pearl (1995)
showed that for intervention variable \( A \) and outcome \( Y \), if a set of variables \( Z \) is such that no
variable in \( Z \) is a descendent of \( A \) and such that \( Z \) blocks all back-door paths from \( A \) to \( Y \) then
conditioning on \( Z \) suffices to control for confounding for the estimation of the causal effect of \( A \)
on \( Y \). The counterfactual value of \( Y \) intervening to set \( A = a \) we denote by \( Y_{A=a} \).

We present an example taken from Greenland et al. (1999) which illustrates the use of the
back-door path criterion and which also motivates the development of some of the theory in this
paper. Consider a study of the relation of antihistamine treatment, denoted by \( E \), and asthma
incidence, denoted by \( D \), among first-grade children attending various public schools. Suppose
that air pollution levels, denoted by \( A \), is independent of sex, denoted by \( B \), among first-grade
public school children. Suppose further that sex influences the administration of antihistamine
only through its relation to bronchial reactivity, denoted by \( C \), but that sex directly influences
asthma risk; suppose also that air pollution leads to asthma attacks only through its influence on
antihistamine use and bronchial reactivity; and that there are no important confounders beyond
air pollution, bronchial reactivity and sex. The causal relationships amongst these variables are
then those given in Figure 1.

![Fig. 1. Example illustrating the use of the back-door path criterion](image_url)

Under the assumptions given above, conditioning on \( A \), \( B \) and \( C \) suffices to control for con-
 founding; conditioning on \( A \) and \( C \) alone also suffices to control for confounding; conditioning
on \( B \) and \( C \) alone also suffices to control for confounding. However conditioning on any other
subset of \( A \), \( B \) and \( C \) does not suffice. Thus if data were available only on antihistamine use
\( E \) (recorded as a binary variable: yes or no), asthma \( D \) and sex \( B \) then we could not produce
valid estimates of the causal effect of $E$ on $D$. We will return to this example at the end of the paper and show that under assumptions concerning weak monotonic effects, the estimate of the causal effect of $E$ on $D$ controlling only for $B$ is in fact conservative.

One further note merits attention concerning the theoretical developments in this paper. We have presented the directed acyclic graph framework as developed by Pearl in terms of non-parametric structural equations. The description given above presumed the existence of counterfactuals corresponding to hypothetical interventions for every variable on the graph. Certain authors have objected to the use of counterfactuals in causal reasoning on philosophical grounds (Dawid, 2000). There has thus developed other interpretations of the directed acyclic graph framework which do not make reference to counterfactuals and which rely principally on conditional independence relations. For example, Spirtes et al. (1993) present a causal interpretation of directed acyclic graphs that is agnostic about the existence of counterfactuals. Dawid (2002) advocates for an interpretation of directed acyclic graphs involving decision-nodes. The results we present in this paper do not require a definite stance on the desirability of using counterfactuals in causal inference. As will be seen below, the paper’s results are stated in terms of monotonic effects and weak monotonic effects. Although the definition of a monotonic effect is given in terms of counterfactuals, the definition of a weak monotonic effect is given in terms of conditional probabilities and makes no reference to counterfactuals. All the results given in this paper are true under both monotonic effects and weak monotonic effects. Consequently, the results given below, interpreted in terms of weak monotonic effects, are applicable to approaches such as those of Spirtes et al. (1993) and Dawid (2002) which make no assumptions about the existence of counterfactuals and which only assume that interventions are possible upon particular nodes on the graph.

The remainder of the paper is organized as follows. Section 2 presents the definitions of a monotonic effect and a weak monotonic effect along with related definitions concerning the signs of a graph’s edges and paths; a number of technical lemmas and a graph theoretic result are also given. Section 3 presents a result which allows for the determination of the sign of the causal effect of an intervention in the presence of intermediate variables; in this section it is also argued that a theory of signed directed acyclic graphs based only on average causal effects will fail. In section 4, a probability lemma is given which is needed to prove the section’s theorem concerning the rules governing signed edges and covariance. Section 5 gives a result concerning the use of monotonic effects to determine the sign of the bias when control for confounding is inadequate. Throughout sections 3 through 5, examples are provided to illustrate the use of the sections’ theorems. Some final comments are given in section 6.

### 2. Monotonic effects and signed edges

Various extensions to the directed acyclic graph causal framework are made possible by introducing the idea of a monotonic effect. The definition of a monotonic effect is given in terms of a directed acyclic graph’s nonparametric structural equations.

**Definition 1 (Monotonic Effect).** The non-parametric structural equation for some node $Y$ on a causal directed acyclic graph with parent $A$ can be expressed as $Y = f(pa_Y^*, A, \epsilon_Y)$ where $pa_Y^*$ are the parents of $Y$ other than $A$; the variable $A$ is said to have a positive monotonic effect on $Y$ if for all $pa_Y^*$ and $\epsilon_Y$, $f(pa_Y^*, a_1, \epsilon_Y) \geq f(pa_Y^*, a_2, \epsilon_Y)$ whenever $a_1 \geq a_2$; the variable $A$ is said to have a negative monotonic effect on $Y$ if for all $pa_Y^*$ and $\epsilon_Y$, $f(pa_Y^*, a_1, \epsilon_Y) \leq f(pa_Y^*, a_2, \epsilon_Y)$ whenever $a_1 \geq a_2$.

**Remark 1.** We first note that the definition of a monotonic effect requires that the intervention variable $A$ and the response variable $Y$ be ordered. The definition is inapplicable to any intervention variable or response variable that is categorical but not also ordinal (e.g. hair color, place of birth, etc.). Beyond this, the definition of a monotonic effect essentially requires that some intervention $A$ either increase or decrease some other variable $Y$ not merely on average over the entire population but rather for every individual in that population, regardless of the inventions made on the other parents of $Y$. The requirements for the attribution of a
monotonic effect are thus considerable. However whenever a particular intervention is always beneficial or neutral for all individuals with respect to a particular outcome, one will be able to attribute a positive monotonic effect; whenever the intervention is always harmful or neutral for all individuals with respect to a particular outcome, one will be able to attribute a negative monotonic effect. With data, one might be able to reject the hypothesis of a positive monotonic effect if for some \( a_1 > a_2 \) the average causal effect over the population of setting \( A = a_1 \) is less than that of setting \( A = a_2 \). However, because for any individual we observe the outcome only under one particular value of the intervention variable, the presence of a monotonic effect is not identifiable. We must thus rely on substantive knowledge of the problem under consideration in order to attribute a monotonic effect. The theorems presented in this paper are in fact true under slightly weaker conditions which are identifiable when data on all of the directed acyclic graph’s variables are observed. We thus introduce the concept of a weak monotonic effect.

**Definition 2 (Weak Monotonic Effect).** A parent \( A \) of some node \( Y \) on a causal directed acyclic graph is said to have a weak positive monotonic effect on \( Y \) if the survivor function \( S(y|pa_Y, a) = \text{pr}(Y \geq y|pa_Y, a) \) is such that whenever \( a_1 \geq a_2 \) we have \( S(y|pa_Y, a_1) \geq S(y|pa_Y, a_2) \) for all \( y \) and all \( pa_Y \); the variable \( A \) is said to have a weak negative monotonic effect on \( Y \) if \( S(y|pa_Y, a_1) \leq S(y|pa_Y, a_2) \) whenever \( a_1 \geq a_2 \) for all \( y \) and all \( pa_Y \).

**Remark 2.** Because the conditional probabilities are identifiable from data, the presence of a weak monotonic effect is also identifiable. A weak monotonic effect is not only identifiable but it also constitutes a substantially less stringent condition. In the case of a binary outcome \( Y \) all that is required for a weak monotonic effect is that a higher value of \( A \) makes the outcome \( Y \) at least as likely regardless of the value of the parents of \( Y \) other than \( A \). When \( Y \) is not binary the presence of a weak monotonic effect is equivalent to the statement that for all \( y \) a higher value of \( A \) makes the event \( \{ Y \geq y \} \) more likely regardless of the value the parents of \( Y \) other than \( A \). If intervening to increase \( A \) led to a decrease in \( Y \) for only a single individual the strong conditions for a monotonic effect would fail. The less stringent conditions required for attributing a weak monotonic effect circumvent this difficulty. Consider, for example, an analysis comparing the effect on thyroid cancer of no radiation exposure to a high level of radiation exposure. For most individuals the exposure to a high level of radiation will increase the likelihood of developing thyroid cancer. However, exposure to a high level of radiation may, for a few individuals, destroy already existing thyroid cancer cells and thereby prevent the cancer’s development. Within joint strata of particular sets of background variables on a causal directed acyclic graph, the exposure to radiation will increase the overall likelihood of thyroid cancer but it may not do so for every individual in the population. In such a scenario the high level of radiation exposure would not have a monotonic effect on the development of thyroid cancer but it would have a weak monotonic effect. It was noted above that the presence of a weak monotonic effect is identifiable. Although it is thus possible to empirically verify the presence of a weak monotonic effect, in practice a researcher will still likely rely on substantive knowledge of the problem in attributing a weak monotonic effect. Unless the variable \( Y \) and all its parents are ordinal with a small number of levels, a very large data set would be required to empirically verify the presence of a weak monotonic effect. Furthermore, the presence of a weak monotonic effect is only identifiable when data are observed on \( Y \) and all of its parents; however, as will be seen below, the theorems given in the paper are principally useful precisely when some of this data is missing.

**Remark 3.** The idea of a weak positive monotonic effect is closely associated with that of positive qualitative influence in Wellman’s qualitative probabilistic networks (Wellman, 1990). A weak positive monotonic effect and positive qualitative influence coincide for parent \( A \) and child \( Y \) when the "context" for qualitative influence is chosen to be the parents of \( Y \) other than \( A \). If hypothetical interventions are conceivable only for certain nodes on a graph then it may be more appropriate to speak of "qualitative positive influence" than of "weak monotonic effects" for those variables for which no intervention is possible. We provide a few additional comments relating our work to that of Wellman (1990). A simplified version of Lemma 4 below and the
first half of the statement of Theorem 2 follow in a fairly straightforward manner from Wellman’s Theorems 4.2 and 4.3. Lemma 4 and Theorem 2, however, give the results in much greater generality. Theorem 1 and Theorems 3-5 in the present paper all contain wholly new material. Wellman (1990) gives considerable attention to the preservation of monotonicity under edge reversal, to the necessity of first order stochastic dominance for "propagating influences" and to the "propagation" of sub-additive and super-additive relationships on probabilistic networks.

A monotonic effect is a relation between two nodes on a directed acyclic graph and as such it is associated with an edge. The definition of the sign of an edge can be given either in terms of monotonic effects or weak monotonic effects. We can define the sign of an edge as the sign of the monotonic effect or weak monotonic effect to which the edge corresponds; this in turn gives rise to a natural definition for the sign of a path.

**Definition 3 (Sign of an Edge).** An edge from $A$ to $Y$ on a causal directed acyclic graph is said to be of positive sign if $A$ has a (weak) positive monotonic effect on $Y$. An edge from $A$ to $Y$ is said to be of negative sign if $A$ has a (weak) negative monotonic effect on $Y$. If $A$ has neither a (weak) positive monotonic effect nor a (weak) negative monotonic effect on $Y$, then the edge from $A$ to $Y$ is said to be without a sign.

**Definition 4 (Sign of a Path).** The sign of a path on a causal directed acyclic graph is the product of the signs of the edges that constitute that path. If one of the edges on a path is without a sign then the sign of the path is said to be undefined.

We will call a causal directed acyclic graph with signs on those edges which allow them a signed causal directed acyclic graph. The theorems in this paper are given in terms of signed paths so as to be applicable to both monotonic effects and weak monotonic effects. Theorems 3, 4 and 5 are proved for the case of weak monotonic effects and the case of monotonic effects thereby follows immediately since the presence of a monotonic effect clearly implies the presence of a weak monotonic effect. The conclusions of Theorem 2, however concern statements about monotonic effects themselves and thus must be proved for the case of monotonic effects and weak monotonic effects separately. Theorem 1 is a purely graph-theoretic result and as such does not concern monotonic effects. The statements about signed paths may thus be interpreted throughout as corresponding to either monotonic effects or weak monotonic effects. One further definition will be useful in Theorem 4 concerning the sign of the covariance of the graph's variables.

**Definition 5 (Monotonic Association).** Two variables $X$ and $Y$ are said to be positively monotonically associated if all directed paths between $X$ and $Y$ are of positive sign and all common causes $C_i$ of $X$ and $Y$ are such that all directed paths from $C_i$ to $X$ are of the same sign as all directed paths from $C_i$ to $Y$; the variables $X$ and $Y$ are said to be negatively monotonically associated if all directed paths between $X$ and $Y$ are of negative sign and all common causes $C_i$ of $X$ and $Y$ are such that all directed paths from $C_i$ to $X$ are of the opposite sign as all directed paths from $C_i$ to $Y$.

Before proceeding to the central results governing monotonic effects, causal effects, covariance and confounding, we present two lemmas which will be useful throughout the development of the theory. The proofs follow almost immediately from the definitions and are suppressed.

**Lemma 1.** If $A$ has a positive monotonic effect on $Y$ then $-A$ has a negative monotonic effect on $Y$; and $A$ has a negative monotonic effect on $-Y$. If $A$ has a negative monotonic effect on $Y$ then $-A$ has a positive monotonic effect on $Y$; and $A$ has a positive monotonic effect on $-Y$.

**Lemma 2.** If on a path between $A$ and $B$ there is a node $V$ other than $A$ or $B$ then replacing $V$ by its negation $-V$ does not change the sign of that path.
The application of these two lemmas can be further extended by the graph theoretic result given in Theorem 1. The proof of Theorem 1 and those of all subsequent results are given in the Appendix unless otherwise indicated.

**Theorem 1.** If the sign of every directed path from \( A \) to \( B \) is positive then there exist nodes \( W_1, \ldots, W_t \) on directed paths between \( A \) and \( B \) such that if \( W_1, \ldots, W_t \) are replaced by their negations then the sign of every edge on every directed path from \( A \) to \( B \) is positive.

In subsequent sections we will also need a number of technical lemmas in order to prove the paper's major theorems. A weak monotonic effect is defined in terms of survivor functions and it is thus useful to develop certain results concerning the survivor functions of variables governed by a causal directed acyclic graph. We will assume throughout the remainder of this paper that the variables under consideration satisfy regularity conditions that allow for the integration by parts required in the proof of Lemma 3 (which is in turn used in the proofs of most subsequent results). Hardle et al. (1998, p72) give relatively weak conditions under which such integration by parts is possible; alternatively, the existence of the Lebesgue-Stieltjes integrals found in the proof of Lemma 3 also suffices to allow integration by parts. Lemma 3 relates non-decreasing functions of random variables and non-decreasing survivor functions to non-decreasing conditional expectations. Lemma 4 relates weak monotonic effects to non-decreasing survivor functions and non-decreasing conditional expectations and is used in the proofs of the theorems throughout this paper. A generalization of Lemma 4, which is needed in the proofs of Theorems 4 and 5, is stated in the Appendix. The proofs of Lemmas 3 and 4 are given elsewhere (VanderWeele and Robins, 2006).

**Lemma 3.** If \( h(y, a, r) \) is non-decreasing in \( y \) and in \( a \) and \( S(y|a, r) = pr(Y > y|A = a, R = r) \) is non-decreasing in \( a \) for all \( y \) then \( E[h(Y, A, R)|A = a, R = r] \) is non-decreasing in \( a \).

**Corollary.** Suppose that the \( A \to Y \) edge, if it exists, is positive. Let \( X \) denote some set of non-descendants of \( Y \) that includes \( pa_Y \), the parents of \( Y \) other than \( A \), then \( E(Y|X = x, A = a) \) is non-decreasing in \( a \) for all values of \( x \).

**Lemma 4.** Let \( X \) denote some set of non-descendants of \( A \) that blocks all backdoor paths from \( A \) to \( Y \). If all directed paths from \( A \) to \( Y \) are positive then \( S(y|a, x) \) and \( E(y|a, x) \) are non-decreasing in \( a \).

**3. Monotonic effects and the sign of the causal effect of an intervention**

Having established the properties of the previous section, we can now prove a result related to the preservation of monotonic effects and weak monotonic effects when marginalizing over certain variables on a causal directed acyclic graph. This theorem allows us in turn to prove a result relating monotonic effects to the sign of the causal effect of an intervention in the presence of intermediate variables. Theorem 2 is stated in terms of positive directed paths but has an obvious analogue for negative directed paths. The theorem will be used in the proofs of subsequent theorems. We note that the first part of the theorem follows immediately from repeated application of Theorems 4.2 and 4.3 in Wellman (1990); it also follows immediately from Lemma 4. The second part of the theorem is a generalization of a monotonicity result given by Cox and Wermuth (2003).

**Theorem 2.** If the sign of every directed path from \( A \) to \( Y \) on a causal directed acyclic graph \( G \) is positive then the \( A \to Y \) edge is positive on the causal directed acyclic graph \( H \) formed by marginalizing \( G \) over all variables on directed paths between \( A \) and \( Y \); furthermore the \( A \to Y \) edge is positive on the causal directed acyclic graph \( J \) formed by marginalizing \( H \) over the ancestors of \( A \) or of \( Y \) which are not common causes of \( A \) and \( Y \).

**Remark 4.** In the simple case of only a single path between two variables, say \( A \) and \( C \), Theorem 2 implies that the relation of manifesting a positive monotonic effect or a weak positive
monotonic effect is transitive. For example, if there is no direct $A \rightarrow C$ edge, and if $A$ has a positive monotonic effect on $B$ and $B$ has a positive monotonic effect on $C$ then it follows that $A$ has a positive monotonic effect on $C$ on the causal directed acyclic graph with only $A$ and $C$. We will show below that if the definition of a monotonic effect is made to depend on only average causal effects, this transitivity property fails to hold. One further comment about the transitivity of monotonic effects merits attention. The definition of a positive monotonic effect required $f(pa_Y^*, a_1, \epsilon_Y) \geq f(pa_Y^*, a_2, \epsilon_Y)$ whenever $a_1 \geq a_2$; it did not require that the inequality hold strictly for any $a_1, a_2$. Of course if for all values of $pa_Y^*, \epsilon_Y$ we had $f(pa_Y^*, a_1, \epsilon_Y) = f(pa_Y^*, a_2, \epsilon_Y)$ for all $a_1, a_2$ then we generally would not attribute any causal effect of $A$ on $Y$; in such circumstances, on the causal directed acyclic graph, $A$ would not be a parent of $Y$. However, if the attribution of a monotonic effect required that the inequality hold strictly for some $a_1, a_2$ then the presence of a monotonic effect would not be a transitive relation. It is possible to construct examples in which $A$ has a positive monotonic effect on $B$ with the inequality holding strictly for at least two values of $A$ and in which $B$ has a positive monotonic effect on $C$ with the inequality holding strictly for at least two values of $B$ but for which $A$ has no causal effect on $C$ (i.e. intervening to set $A$ to any value would always leave $C$ unchanged). However, such examples in which causal effects are not transitive are a feature of causal directed acyclic graphs generally and are not unique to the setting of monotonic effects.

We can now state and prove the result relating monotonic effects and weak monotonic effects to the sign of the causal effect of an intervention in the presence of intermediate variables.

**Theorem 3.** If $A$ is an ancestor of $Y$ and the sign of every directed path from $A$ to $Y$ is positive then $E(Y_{A=a})$ is non-decreasing in $a$.

Theorem 3 states that if the sign of every directed path from $A$ to $Y$ is positive then intervening to increase $A$ will always increase or leave unchanged the average value of $Y$ over the population. The theorem also has an obvious analogue if the sign of every directed path from $A$ to $Y$ is negative rather than positive. Note that the theorem requires that the sign of every directed path from $A$ to $Y$ is positive (or negative). If two directed paths from $A$ to $Y$ are of different sign or if any edge on some directed path from $A$ to $Y$ is without sign, we cannot determine the sign of the causal effect of an intervention from Theorem 3.

**Example 1.** We illustrate the use of the theorem by considering the signed causal directed acyclic graph given in Fig. 2. Note that no sign is present on the edge $A \rightarrow E$.

![Fig. 2. Example illustrating the relationship between monotonic effects and the sign of the causal effect of an intervention](image_url)

By Theorem 3, intervening to increase $B$ will increase the average value of $E$ over the population since all directed paths from $B$ to $E$ (i.e. $B \rightarrow E$ and $B \rightarrow C \rightarrow D \rightarrow E$) are of positive sign. Intervening to increase $A$ will decrease the average value of $D$ over the population since all directed paths from $A$ to $D$ (i.e. $A \rightarrow Q \rightarrow D$ and $A \rightarrow B \rightarrow C \rightarrow D$) are of negative sign. However we cannot determine from the signed causal directed acyclic graph whether intervening to increase $A$ will increase or decrease the average value of $E$ over the population because the $A \rightarrow E$ edge is without a sign. Also we cannot determine from the signed causal directed acyclic graph whether intervening to increase $B$ will increase or decrease the average value of $F$ over the population.
because the paths $B - C - D - E - F$ and $B - E - F$ are of positive sign but the path $B - F$ is of negative sign.

Signs have sometimes informally been given to edges on a causal directed acyclic graph when intervening on the parent increases the average value of the child over the population. However when signs are given to edges in this informal manner, there are cases in which the sign from $A$ to $B$ might be positive and the sign from $B$ to $C$ might be positive (i.e. when all directed paths from $A$ to $C$ informally have positive sign) but intervening to increase $A$ in fact decreases $C$ on average over the population. In fact, even very slight departures from the requirements of a monotonic effect suffice to give counterintuitive examples. Even when intervening to increase $A$ will increase $B$ for all individuals in a population with the exception of only a single individual and intervening to increase $B$ will increase or leave unchanged $C$ for every individual in the population we may still have cases in which intervening to increase $A$ will in fact decrease $C$ on average over the population. The Example 3 below illustrates such a case. The example illustrates first the danger of informally giving signs to the edges on a causal directed acyclic graph when intervening on the parent increases the average value of the child over the population. The example furthermore demonstrates that even very slight departures from the requirements of a monotonic effect render false the conclusions of Theorems 2 and 3 (and consequently also Theorems 4 and 5 below). Although the requirements for attributing monotonic effects or weak monotonic effects are considerable, a theory based only on average causal effects will not work. We note that in the case of linear structural equations and multivariate normality, the requirements for a weak monotonic effect are satisfied and so counterintuitive situations, like that illustrated in Example 2, do not arise.

Example 2. Consider the causal directed acyclic graph given in Figure 3 with positive signs given informally to the $A \rightarrow B$ and the $B \rightarrow C$ edges. We denote these informal signs by a plus sign in quotation marks.

Fig. 3. Example illustrating the informal use of positive and negative signs on edges

Suppose that the variables $A$, $B$ and $C$ are all binary and that $P(A = 1) = P(A = 0) = 1/2$. Suppose also that if $A = 1$ then $P(B = 1) = 1$ and that if $A = 0$ then $P(B = 2) = p$ and $P(B = 0) = 1 - p$ where $p < 1/2$ so that $E[B_{A=1}] = 1$ and $E[B_{A=0}] = 2p < 1$. Finally suppose that if $B = 2$ then $C = 1$ and if $B = 0$ or $B = 1$ then $C = 0$. We then have that increasing $A$ increases $B$ on average and increasing $B$ increases $C$ on average but in this example, when $A = 1$ then $B = 1$ and $C = 0$ but when $A = 0$, $P(B = 2) = p$ and $P(B = 0) = 1 - p$ and so $P(C = 1) = p$ and $P(C = 0) = 1 - p$. Thus $E[C_{A=1}] = 0$ but $E[C_{A=0}] = p > 0$. Thus intervening to increase $A$ decreases $C$ on average. Suppose that in a population of $n$ individuals, $p = 1/n$ then even though the departure from the requirements of a monotonic effect concerns only a single individual, intervening to increase $A$ from 0 to 1 will decrease $C$ on average over the entire population.

4. Covariance and monotonic effects

The notions of monotonic effects and weak monotonic effects introduced above can be used to develop rules that govern monotonic effects and covariance. When the signs of directed paths relating two variables satisfy certain conditions, it is possible to determine the sign of the covariance of these variables. However, in order to prove the central result which provides these rules we need to make use of an additional probability lemma presented below. Lemma 5 is essentially a restatement of Theorem 2.1 in Esary et al. (1967).

Lemma 5. Let $f$ and $g$ be functions with $n$ real-valued arguments such that both $f$ and $g$ are non-decreasing in each of their arguments. If $X = (X_1, ..., X_n)$ is a multivariate random variable with $n$ components such that each component is independent of the other components then $cov\{f(X), g(X)\} \geq 0$. 

9
This probability lemma above allows us to prove Theorem 4 concerning the rules governing covariance and monotonic effects. Theorem 4 may be seen as a generalization of the corresponding result for recursive linear structural equation models which would follow quite simply from elementary path analysis (Duncan 1975).

**Theorem 4.** If \( X \) and \( Y \) are positively monotonically associated then \( \text{Cov}(E_1, E_2) \geq 0 \). If \( X \) and \( Y \) are negatively monotonically associated then \( \text{Cov}(E_1, E_2) \leq 0 \).

Theorem 4 concerns the sign of covariances. In related work we have derived rules governing the sign of conditional covariance in the presence of monotonic effects; however these rules require the development of theory concerning minimal sufficient causation and assumptions beyond simply that of the presence of monotonic effects. If the directed paths relating two variables satisfy the conditions of Theorem 4, then this yields implications concerning the sign of the covariance of these variables. Obviously these conditions will not always hold; it will not always be possible to determine the sign of the covariance between any two variables simply from the signed causal directed acyclic graph. For a particular signed causal directed acyclic graph, Theorem 4 will yield implications concerning the signs of covariances for some pairs of variables and may fail to do so for others. Theorem 4 may fail to yield implications concerning the signs of covariances either because the signs of different directed paths are not congruent in the way required by the conditions of the theorem or because certain edges are without sign.

**Example 3.** Consider the signed directed acyclic graph given in Fig. 4 and note that no sign is present on the edge \( A - F \).

By Theorem 4, the covariance between \( B \) and \( C \) will be positive since the sign of the only directed path from \( B \) to \( C \) is positive and \( B \) and \( C \) have no common causes except through \( B \). The covariance between \( C \) and \( D \) will be negative since the sign of the only directed path from \( C \) to \( D \) is negative and because for the only common cause of \( C \) and \( D \), namely \( A \), all directed paths from \( A \) to \( C \) are positive and thus of the opposite sign as all directed paths from \( A \) to \( D \) which are negative. The covariance between \( E \) and \( F \) is undetermined because \( A \) is a common cause of \( E \) and \( F \) and the path from \( A \) to \( F \) consisting of the edge \( A - F \) is without sign. The covariance between \( D \) and \( E \) is also undetermined because the \( D - E \) edge is positive but \( A \) is a common cause of \( D \) and \( E \) and the directed path \( A - B - C - D \) from \( A \) to \( D \) is negative and so of opposite sign as the directed path from \( A \) to \( E \) consisting of the edge \( A - E \) which is positive.

5. **Confounding and monotonic effects**

Our final result relates monotonic effects with the sign of the bias which results when control for confounding is inadequate. Manksi (1997) has considered bounds for the treatment effect when the outcome is a monotonic function of the treatment. Here we consider monotonic relationships between confounding variables and the treatment and outcome. Pearl (1995) showed that for intervention variable \( A \) and outcome \( Y \), if a set of variables \( Z \) such that no variable in \( Z \) is a descendent of \( A \) blocks all back-door paths from \( A \) to \( Y \) then the causal effect on \( Y \) of intervening to set \( A = a \) can be determined by

\[
E(Y_{A=a}) = \sum_z E(Y|A=a, Z=z)\Pr(Z=z).
\]
Note that this result is a graphical generalization of Theorem 4 of Rosenbaum and Rubin (1983a) and of the g-formula (Robins 1986, 1987; Spirtes et al. 1993; Pearl 1993). Now if \( X \) is some set of variables that does not block all backdoor paths from \( A \) to \( Y \) and an attempt is made by means of Pearl’s result to estimate the causal effect on \( Y \) of intervening to set \( A = a \) controlling only for \( X \), one would obtain

\[
\sum_x E(Y|A = a, X = x)pr(X = x).
\]

This latter expression will in general differ from the true causal effect. Theorem 5 relates monotonic effects and weak monotonic effects to the sign of the bias which arises when control for confounding is inadequate. The theorem requires that the intervention variable \( A \) be binary. We provide further discussion of this requirement following the statement of the theorem.

**Theorem 5.** Suppose that for some binary intervention \( A \) and some outcome \( Y \), the set of variables \( X \) does not block all backdoor paths from \( A \) to \( Y \) but does not open any backdoor paths from \( A \) to \( Y \) which were blocked without conditioning on \( X \). Let \( S_a = \sum_x E(Y|A = a, X = x)pr(X = x) \). If all unblocked backdoor paths from \( A \) to \( Y \) are of positive sign then \( S_1 \geq E(Y_{A=1}) \) and \( S_0 \leq E(Y_{A=1}) \). If all unblocked backdoor paths from \( A \) to \( Y \) are of negative sign then \( S_1 \leq E(Y_{A=1}) \) and \( S_0 \geq E(Y_{A=1}) \).

It follows from Theorem 5 that if all unblocked backdoor paths from \( A \) to \( Y \) are of positive sign then we will overestimate the true causal risk difference and if all unblocked backdoor paths from \( A \) to \( Y \) are of negative sign then we will underestimate the true causal risk difference. Control for confounding is often inadequate when certain variables which are known to be confounders are not measured in a particular study. Various attempts have been made to provide sensitivity bounds for unmeasured confounders (Rosenbaum and Rubin, 1983b; Lin et al., 1998; Brumback et al., 2004). Theorem 5 allows the researcher in certain circumstances to determine the sign of the bias, thereby making clear whether the estimate under consideration is generous or conservative due to lack of control for certain confounding variables.

**Example 4.** We return to the example given in the Introduction illustrated by Figure 1 above. Recall that antihistamine treatment is denoted by \( E \), asthma incidence by \( D \), air pollution levels by \( A \), sex by \( B \) and bronchial reactivity by \( C \). Data is available only on antihistamine use \( E \) (recorded as a binary variable: yes or no), asthma \( D \) and sex \( B \). Conditioning only on \( B \) does not suffice to produce valid estimates of the causal effect of \( E \) on \( D \). In this example it is arguably quite reasonable to assume that air pollution has a weak positive monotonic effect on bronchial reactivity and on antihistamine use and that bronchial reactivity has a weak positive monotonic effect on antihistamine use and on asthma. We may then add to Figure 1 the positive signs indicated in Figure 5.

**Fig. 5.** Example illustrating the relationship between bias, confounding and monotonic effects

Conditioning on only \( B \), there are two unblocked backdoor paths from \( E \) to \( D \): \( E - C - D \) and \( E - A - C - D \). Both of these unblocked backdoor paths from \( E \) to \( D \) are positive. Thus we could conclude from Theorem 5 that the estimate of the risk difference controlling only for \( B \) is an overestimate of the true causal risk difference. If the estimate controlling only for \( B \) were negative we could then conclude that the estimate was conservative and thus that antihistamine use did indeed lower asthma risk. We use simulated data to further illustrate the application of Theorem 5. In the example we will assume that \( D \) is recorded as a continuous variable.
relationships were given by the causal directed acyclic graph given in Fig. 6. Clearly the true causal effect, \( E(D_{E=1}) - E(D_{E=0}) \), in this example is \(-8\). Data simulated from the structural equations above gives the following: a regression of \( D \) on \( B \) and \( E \) gives a coefficient estimate for \( E \) of \(-6.61\) with a 95% confidence interval of \((-7.75, -5.67)\) and this estimate coincides almost exactly with the estimate of \( S_1 - S_0 = \sum_b \{ E(Y|B=b, E=1) - E(Y|B=b, E=0) \} pr(B=b) \) calculated by exact stratification. Thus, in this example, we conclude that anti-histamine use lowered asthma from data just on \( B \), \( E \) and \( D \) since it is a priori known that the estimate controlling only for \( B \) is conservative. If in this example data were available for \( A \) and \( C \) as well, a regression of \( D \) on \( A \), \( B \), \( C \) and \( E \) with this simulated data gives a coefficient estimate for \( E \) of \(-7.76\) with 95% confidence interval of \((-8.65, -6.87)\).

Note that Theorem 5 requires that the exposure variable under consideration be binary. The theorem also holds when \( A \) is not binary if \( A = 1 \) is replaced with the maximum value of \( A \) and if \( A = 0 \) is replaced with the minimum value of \( A \). However counterexamples can be constructed to demonstrate that the theorem cannot be generalized beyond the extreme values of the intervention variable.

**Example 5.** Suppose that \( A \) were ternary and took on values in the set \{1, 2, 3\} and suppose that a single binary variable \( C \) confounded the relationship between \( A \) and \( Y \) so that the causal relationships were given by the causal directed acyclic graph given in Fig. 6.

![Fig. 6.](image)

Fig. 6. Example demonstrating that Theorem 5 may fail when \( A \) is not binary

If \( f_Y(A = 3, c) = 1 \) for \( c = 0, 1 \) and \( f_Y(A = 1, c) = 0 \) for \( c = 0, 1 \) and \( f_Y(A = 2, C = 1) = 1 \) and \( f_Y(A = 2, C = 0) = 0 \) then both \( A \) and \( C \) will have a positive monotonic effect on \( Y \). Suppose \( pr(C = 0) = pr(C = 1) = 1/2; \) and \( pr(\epsilon_A = 1) = pr(\epsilon_A = 2) = pr(\epsilon_A = 3) = 0.01 \) and \( pr(\epsilon_A = 4) = 0.97 \); and suppose also \( f_A(c, \epsilon_A = 1) = 3 \) for \( c = 0, 1 \), \( f_A(c, \epsilon_A = 2) = 1 \) for \( c = 0, 1 \), \( f_A(C = 1, \epsilon_A = 3) = 3 \), \( f_A(C = 0, \epsilon_A = 3) = 2 \), \( f_A(C = 1, \epsilon_A = 4) = 2 \) and \( f_A(C = 0, \epsilon_A = 4) = 1 \) then \( C \) also has a positive monotonic effect on \( A \). We then have \( pr(C = 1|A = 2) = pr(C = 1, A = 2)/pr(A = 2) = (0.97)(0.5)/(0.97)(0.5) = 0.970 \) and \( pr(C = 2|A = 2) = 0.03 \) whereas the estimate of this causal effect without controlling for \( C \) will be given by \( E[Y|A = 2, C = c]pr(C = c) = (1)(0.5) + (0)(0.5) = 0.5 \) whereas the estimate of this causal effect without controlling for \( C \) will be given by \( E[Y|A = 2, C = c]pr(C = c) = (1)(0.5) + (0)(0.5) = 0.5 \) whereas the estimate of this causal effect without controlling for \( C \) exceeds the true causal effect on \( Y \) of intervening to set \( A = 2 \). If however the probabilities for \( \epsilon_A \) had been \( pr(\epsilon_A = 1) = pr(\epsilon_A = 2) = pr(\epsilon_A = 4) = 0.01 \) and \( pr(\epsilon_A = 3) = 0.97 \) then we would still have the positive monotonic effects indicated in Fig. 6 but
the estimate of the causal effect on $Y$ of intervening to set $A = 2$ without controlling for $C$ would then be 1/98 which would be less than the true causal effect of 0.5. For intermediate values of the intervention variable we thus see that the bias when control for confounding is inadequate may be in either direction even in the presence of monotonic effects.

Although Theorem 5 does not hold for intermediate values of the intervention variable, the theorem can still be useful when the intervention variable $A$ is ordinal or continuous. A non-binary intervention variable may be dichotomized at various cut-off points. The analysis may proceed with this dichotomized intervention variable with Theorem 5 employed to assess the sign of the bias. The analysis may then be repeated a number of times at different dichotomization points and conclusions drawn from the resulting analyses.

6. Discussion

The extensions introduced in this paper provide the researcher with tools useful in drawing causal inferences. We have formalized the conditions under which signs can be added to the edges of a causal directed acyclic graph and have provided rigorous rules governing their use. We have also argued that a theory based only on average causal effects will not work. The result given in Theorem 3 governing the relationship between monotonic effects and causal effects in the presence of intermediate variables allows the researcher in certain cases to determine a priori whether a particular intervention will on average have a positive or negative effect. The rules governing monotonic effects and covariance given in Theorem 4 may assist the researcher in assessing whether assumptions being made about the causal structure of variables and about monotonic effects are in fact valid. Because unmeasured variables, control for confounding is often inadequate and various attempts have been made to provide sensitivity bounds for unmeasured confounders. Theorem 5 allows the researcher under certain circumstances to determine the sign of the bias, thereby making clear whether the estimate under consideration is generous or conservative. The directed acyclic graph causal framework has proved to be a useful tool in thinking carefully about questions of estimation and confounding. It is hoped that these contributions will extend yet further the framework’s utilization and applicability.

Appendix A

A.1. Proof of theorem 1

Let $P = \{V_1, \ldots, V_n\}$ be all the nodes on directed paths between $A$ and $B$ including $A$ and $B$ such that $A = V_1$ and $B = V_n$ and let the nodes be ordered so that if $i < j$ then $V_i$ is not a descendent of $V_j$. Consider the parents of $V_i$ in $P$ and suppose $W_1^i, \ldots, W_n^i$ are such that the edges from $W_1^i$ to $V_n$ are negative. Then the directed acyclic graph with $W_1^i, \ldots, W_n^i$, replaced by their negations has all edges from nodes in $P$ into $V_n$ positive. We prove the general result by proving inductively the following statement, which we denote by $(\ast)$: For any $k$, there exists a set of nodes $W_1^k, \ldots, W_t^k$ in $P$ such that if these nodes are replaced by their negations then all edges on all directed paths between any of the nodes $V_k, \ldots, V_n$ are positive and all edges from nodes in $P$ into $V_k, \ldots, V_n$ are positive. Clearly this holds for $k = n$ as shown above. We will show that if it holds for $k = l$ then it holds also for $k = l - 1$. Let $G'$ be the graph with $W_1^k, \ldots, W_t^k$, replaced by their negations. Consider the parents of $V_{l-1}$ and let $U_1, \ldots, U_m$ be the parents of $V_{l-1}$ with negative edges into $V_{l-1}$. Let $G''$ be the graph $G'$ with $U_1, \ldots, U_m$ replaced by their negations. We will prove that $G''$ satisfies the properties required by $(\ast)$ for $k = l - 1$; i.e. that taking $W_1^{l-1}, \ldots, W_t^{l-1} = U_1, \ldots, U_m, W_1^l, \ldots, W_t^l$ satisfies $(\ast)$ for $k = l - 1$. On $G''$, none of $V_l, \ldots, V_n$ can be parents of $V_{l-1}$ so clearly all edges on all directed paths between any of the nodes $V_l, \ldots, V_n$ are positive. Furthermore, any edge from $V_{l-1}$ to one of $V_l, \ldots, V_n$ must be positive by virtue of $(\ast)$ holding for $k = l$. Thus all edges on all directed paths between any of the nodes $V_{l-1}, \ldots, V_n$ are positive. Clearly all edges into $V_{l-1}$ are positive in $G''$. All edges into $V_l, \ldots, V_n$ were positive in $G'$ so the only possible way in which an edge into $V_l, \ldots, V_n$ in $G''$ could be negative is if it were from one of $U_1, \ldots, U_m$. Suppose there exist $i, j$ such that the edge from the negation of $U_i$, $-U_i$, to $V_j$ is negative in $G''$, where $V_j$ is one of $V_l, \ldots, V_n$. Then there
exists a directed path from \(-U_i\) through \(V_i\) to \(V_i\) that is positive and also a directed path from \(-U_i\) through \(V_i\) to \(V_i\) that is negative but this contradicts the hypothesis that the sign of all directed paths between \(A\) and \(B\) are positive. Thus none of \(U_1,\ldots,U_m\) can be a parent of any of \(V_i,\ldots,V_n\) and so all edges into \(V_i,\ldots,V_n\) are positive in \(G''\). Furthermore, it follows that \(U_1,\ldots,U_m\) is distinct from \(W_i^1,\ldots,W_i^l\) and that taking \(W_i^{l-1},\ldots,W_i^l = U_1,\ldots,U_m, W_i^1,\ldots,W_i^l\) satisfies (\(\ast\)) for \(k = l - 1\). The result follows.

A.2. Generalization of lemma 4

Lemma 4. Let \(X\) denote some set of non-descendents of \(A\) that blocks all backdoor paths from \(A\) to \(Y\). Let \(R = (R_1,\ldots,R_m)\) denote an ordered list of some set of nodes on directed paths from \(A\) to \(Y\) such that for each \(i\) the backdoor paths from each element of \(R_i,\ldots,R_m\) to \(Y\) are blocked by \(R_i,\ldots,R_{i-1}, A\) and \(X\). If all directed paths from \(A\) to \(Y\) are positive except possibly through \(R\) then \(S(y|a,x,r)\) and \(E(y|a,x,r)\) are non-decreasing in \(a\).

A.3. Proof of theorem 2

We prove Theorem 2 in the case of weak monotonic effects. The case of monotonic effects is reasonably straightforward and follows by recursive substitution. By the properties of causal directed acyclic graphs, the original graph \(G\) can be marginalized to the causal directed acyclic graph \(H\). Let \(C\) denote the set of common causes of \(A\) and \(Y\). Let \(Q\) denote the set of nodes that are ancestors of \(A\) or of \(Y\) but are not descendents of \(A\) and not common causes of \(A\) and \(Y\). By Lemma 4 with \(X = \{C,Q\}\), it follows that \(S(y|a,c)\) is non-decreasing in \(a\) and that the edge between \(A\) and \(Y\) is positive on \(H\). Let \(Q_1\) denote the subset of \(Q\) which are ancestors \(A\); let \(Q_2\) denote the subset of \(Q\) which are ancestors \(Y\). As noted above, \(S(y|a,c)\) is non-decreasing in \(a\). It remains to show that \(S(y|a,c)\) is non-decreasing in \(a\). Because variables in \(Q_1\) are ancestors of \(Y\) only through \(A\) we have that

\[
S(y|a,c,q) = S(y|a,c,q_2).
\]

Now variables in \(Q_2\) are neither ancestors nor descendents of \(A\); furthermore, \(C\) will contain all common cause of \(A\) and variables \(Q_2\). From this it follows that \(A\) and \(Q_2\) are d-separated given \(C\). Thus

\[
S(y|a,c) = E[S(y|a,c,Q_2)|a,c] = E[S(y|a,c,Q_2)] = E[S(y|a,c,Q)] = E[S(y|a,c,Q)|c]
\]

and since \(S(y|a,c,Q)\) is non-decreasing in \(a\) for all values of \(y,c,q\) we have that

\[
S(y|a,c) = E[S(y|a,c,Q)|c]
\]

is non-decreasing in \(a\).

A.4. Proof of theorem 3

Let \(C\) be the parents of \(A\). By Lemma 4, \(E(Y|A = a, C = c)\) is non-decreasing in \(a\) for all values of \(c\) and by the back-door path criterion we have for \(a_1 \geq a_2\) that

\[
E(Y_{A=a_1}) = \int E(Y|A = a_1, C = c)dF(c) \geq \int E(Y|A = a_2, C = c)dF(c) = E(Y_{A=a_2}).
\]

A.5. Proof of theorem 4

We prove (i) and note that (ii) follows by replacing the variable \(A\) with \(-A\). We employ several times through this proof the result that

\[
cov(Y,Z) = cov\{E(Y|X), E(Z|X)\} + E\{cov(Y,Z|X)\}.
\]
By Theorems 1 and 2 and by replacing certain variables by their negations if necessary, the original causal directed acyclic graph can be collapsed to one with only $A$ and $Y$ and their common causes $C$ with the edge between $A$ and $Y$ positive on the resulting causal directed acyclic graph and with every edge between each $C_i$ and $A$ and between each $C_i$ and $Y$ positive. We then have that

\[ \text{cov}(A, Y) = \text{cov}\{E(A|C), E(Y|C)\} + E\{\text{cov}(A, Y|C)\}. \]

We will first show that $E\{\text{cov}(A, Y|C)\}$ is non-negative. We have that

\[ \text{cov}(A, Y|C) = E\{[A - E(A|C)] [Y - E(Y|C)]|C\} \]
\[ = E(E\{[A - E(A|C)] [Y - E(Y|C)]|A, C\}|C) \]
\[ = E\{[A - E(A|C)] [E(Y|A, C) - E(Y|C)]|C\} \]
\[ = \text{cov}\{A, E(Y|A, C)|C\}. \]

Given $C$, and $E(Y|A, C)$ is a non-decreasing function of $A$ by Lemma 4. Furthermore, given $C$, $A$ is an non-decreasing function of $A$ and thus by Lemma 5 we have that for each $C$, $\text{cov}(A, Y|C) \geq 0$ and so

\[ E\{\text{cov}(A, Y|C)\} \geq 0. \]

We will now show that $\text{cov}\{E(A|C), E(Y|C)\}$ is non-negative. For any component $C_i$ of $C$ let $X$ be the non-descendants in $C$ of $C_i$ and let $R$ be the descendants in $C$ of $C_i$ then by Lemma 4, $E(A|C)$ and $E(Y|C)$ are non-decreasing in each component of $C$. Let $f(C) = E(A|C)$ and $g(C) = E(Y|C)$ so that $\text{cov}\{E(A|C), E(Y|C)\} = \text{cov}\{f(C), g(C)\}$ where $f(C)$ and $g(C)$ are non-decreasing in each component of $C$. Let $S_i$ denote the subset of $C$ which has no ancestors; let $S_2$ denote the subset of $C$ which has no ancestors other than those in $S_1$; let $S_i$ denote the subset of $C$ which has no ancestors other than those in $S_1, ..., S_{i-1}$ and let $k$ be such that $C = S_1, ..., S_k$. Note that given $S_1, ..., S_{i-1}$ the components of $S_i$ are independent of one another. We then have that

\[ \text{cov}\{E(A|C), E(Y|C)\} = \text{cov}\{f(C), g(C)\} \]
\[ = \text{cov}[E\{f(C)|S_1\}, E\{g(C)|S_1\}] + E[\text{cov}\{f(C), g(C)|S_1\}]. \]

Applying the conditional covariance result again to the second of these two terms, conditioning on $S_2$ we have

\[ \text{cov}\{f(C), g(C)\} = \text{cov}[E\{f(C)|S_1\}, E\{g(C)|S_1\}] \]
\[ + E(\text{cov}[E\{f(C)|S_1, S_2\}, E\{g(C)|S_1, S_2\}|S_1]) + E[\text{cov}\{f(C), g(C)|S_1, S_2\}]. \]

And continuing to iteratively apply the conditional covariance result to the final term gives

\[ \text{cov}\{f(C), g(C)\} = \text{cov}[E\{f(C)|S_1\}, E\{g(C)|S_1\}] \]
\[ + E(\text{cov}[E\{f(C)|S_1, S_2, S_3\}, E\{g(C)|S_1, S_2, S_3\}|S_1, S_2]) \]
\[ + E(\text{cov}[E\{f(C)|S_1, ..., S_{k-1}\}, E\{g(C)|S_1, ..., S_{k-1}\}|S_1, ..., S_{k-2}]) \]
\[ + E[\text{cov}\{f(C), g(C)|S_1, ..., S_{k-1}\}]. \]

Consider the $i$th term of this expression,

\[ E(\text{cov}[E\{f(C)|S_1, ..., S_i\}, E\{g(C)|S_1, ..., S_i\}|S_1, ..., S_{i-1}]). \]

Now

\[ E\{f(C)|S_1, ..., S_i\} = E[E[...E[E\{f(C)|S_1, ..., S_{k-1}\}|S_1, ..., S_{k-2}]...|S_1, ..., S_{i+1}]|S_1, ..., S_i]. \]
and \( f(C) \) is non-decreasing in \( S_1, \ldots, S_{k-1} \) and \( S_k \). Also, \( S(s_k|s_1, \ldots, s_{k-1}) = S(s_k|pa_{S_k}) \) is non-decreasing in \( s_1, \ldots, s_{k-1} \) since all edges from \( S_1, \ldots, S_{k-1} \) to \( S_k \) are of positive sign and so by Lemma 3 we have that \( E\{f(C)|S_1, \ldots, S_{k-1}\} \) is non-decreasing in \( S_1, \ldots, S_{k-1}; \) furthermore \( S(s_k-1|s_1, \ldots, s_{k-2}) \) is non-decreasing in \( s_1, \ldots, s_{k-2} \) and so \( E[E\{f(C)|S_1, \ldots, S_{k-1}\}|S_1, \ldots, S_{k-2}] \) is non-decreasing in \( S_1, \ldots, S_{k-2} \). Carrying the argument forward we have that

\[
E(f(C)|S_1, \ldots, S_i) = E[E(...E[E(f(C)|S_1, \ldots, S_{k-1})|S_1, \ldots, S_{k-2}]|...|S_1, \ldots, S_i+1]|S_1, \ldots, S_i]
\]

is non-decreasing in \( S_1, \ldots, S_i \). Similarly \( E\{g(C)|S_1, \ldots, S_i\} \) is non-decreasing in \( S_1, \ldots, S_i \) and so conditional on \( S_1, \ldots, S_{i-1}, E\{f(C)|S_1, \ldots, S_i\} \) and \( E\{g(C)|S_1, \ldots, S_i\} \) are both non-decreasing functions of the independent random variables \( S_i \) and thus by Lemma 5,

\[
cov[E\{f(C)|S_1, \ldots, S_i\}, E\{g(C)|S_1, \ldots, S_i\}|S_1, \ldots, S_{i-1}] \geq 0
\]

and so

\[
E(cov[E\{f(C)|S_1, \ldots, S_i\}, E\{g(C)|S_1, \ldots, S_i\}|S_1, \ldots, S_{i-1})] \geq 0.
\]

Since each term of (2) is non-negative we have that

\[
cov\{E(A|C), E(Y|C)\} = cov\{f(C), g(C)\} \geq 0.
\]

This completes the proof.

A.6. Proof of theorem 5

We prove (i) and note that (ii) follows by replacing the variable \( A \) with \( 1 - A \). Let \( S_1 \) denote the estimate of the causal effect on \( Y \) of intervening to set \( A = 1 \) conditioning on \( X \) only and let \( S_0 \) denote the estimate of the causal effect on \( Y \) of intervening to set \( A = 0 \) conditioning on \( X \) only. Let \( W \) be the set of ancestors of \( A \) through which, conditioning on \( X \), there exists an unblocked backdoor path from \( A \) to \( Y \) a component of which consists of a directed path from the node in \( W \) to \( Y \). Let \( C \) be the set of ancestors of \( A \) from which there exists a directed path to \( Y \) not through \( A \). Let \( X' = C \setminus W \). Conditioning on \( C \) suffices to control for confounding. The estimate of the causal effect of \( A \) on \( Y \) conditioning on \( X \) will be the same as the estimate of the causal effect of \( A \) on \( Y \) conditioning on \( Z = X \cup X' \) because all backdoor paths blocked by \( X' = C \setminus W \) are blocked also by \( X \). For the true causal effect of \( A \) on \( Y \) it suffices to condition on \( Z \) since conditioning on \( W \) blocks all backdoor paths not blocked by conditioning on \( Z \). For each \( W_i \in W \), by replacing \( W_i \) with its negation if necessary, we may assume that all directed paths from \( W_i \) to \( A \) are of positive sign and that all directed paths from \( W_i \) to \( Y \) except possibly through \( A \) are of positive sign. The true causal effect on \( Y \) of intervening to set \( A = a \) can be given by

\[
E(Y_{A=a}) = \sum_w E(Y|A = a, Z = z, W = w)pr(W = w, Z = z) = \sum_z (\sum_w E(Y|A = a, Z = z, W = w)pr(W = w|Z = z))pr(Z = z).
\]

Let

\[
g(a, z, w) = E(Y|A = a, Z = z, W = w)
\]

and let \( F_{W|Z=z} \) be the distribution function of \( W \) given \( Z = z \) then

\[
E(Y_{A=a}) = \sum_z F_{W|Z=z}(g(a, z, W))pr(Z = z).
\]

The estimate of the causal effect conditioning on \( Z \) only can be given by

\[
S_1 = \sum_z E(Y|A = a, Z = z)pr(Z = z) = \sum_z (\sum_w E(Y|A = a, Z = z, W = w)pr(W = w|A = a, Z = z))pr(Z = z).
\]

Let \( F_{W|Z=z,A=a} \) be the distribution function of \( W \) given \( Z = z \) and \( A = a \) so that

\[
S_1 = \sum_z E_{F_{W|Z=z,A=a}}(g(a, z, W))pr(Z = z).
\]
Let $V_1, ..., V_k$ be an ordered list of the variables in $W$. Consider some $W_i$ in the set $W$. Since all directed paths from every $W_i$ to $Y$ not through $A$ are of positive sign and since all backdoor paths from $A$ to $Y$ are blocked by variables in $Z \cup W$ and for all $j > i$ all backdoor paths from $W_j$ to $Y$ are blocked by variables in $\{W_1, ..., W_{j-1}\}$, we have by Lemma 4 that

$$g(a, z, w) = E(Y|A = a, Z = z, W = w)$$

is non-decreasing in $w_i$ and since this argument holds for all $i$ we have that $E[Y|A = a, Z = z, W = w]$ is non-decreasing in $w$. Now,

$$pr(W = w|A = 1, Z = z) = \frac{pr(W = w, A = 1, Z = z)}{pr(A = 1|Z = z)} = \frac{pr(A = 1|W = w, Z = z)pr(W = w|Z = z)pr(Z = z)}{pr(A = 1|Z = z)pr(Z = z)} = \frac{pr(A = 1|W = w, Z = z)pr(Z = z)}{pr(A = 1|Z = z)pr(W = w|Z = z)}.$$

By Lemma 4 it follows that

$$pr(A = 1|W = w, Z = z) = E(A|W = w, Z = z)$$

is non-decreasing in each component of $w$ since all directed paths from any component of $W$ to $A$ are of positive sign and since all backdoor paths from $W_j$ to $A$ are blocked by variables in $\{W_1, ..., W_{j-1}\}$ and $Z$. Let

$$v_z(w) = \frac{pr(A = 1|W = w, Z = z)}{pr(A = 1|Z = z)}$$

which we have shown is non-decreasing in $w$. The distribution functions $F_{W|Z=z}$ and $F_{W|Z=z,A=1}$ are related by

$$pr(W = w|A = 1, Z = z) = v_z(w)pr(W = w|Z = z)$$

and so

$$E_{F_{W|Z=z,A=1}}\{g(a, z, W)\} = E_{F_{W|Z=z}}\{g(a, z, W)v_z(W)\}.$$

Since $E_{F_{W|Z=z}}\{v_z(W)\} = 1$ we have

$$E_{F_{W|Z=z}}[g(a, z, W)\{v_z(W) - 1\}] = Cov_{F_{W|Z=z}}\{g(a, z, W), v_z(W)\}$$

and since both $g(a, z, w)$ and $v_z(w)$ are increasing in $w$ it follows by an argument very similar to that in Theorem 4 that $Cov_{F_{W|Z=z}}\{g(a, z, W), v_z(W)\} \geq 0$. Thus

$$E_{F_{W|Z=z,A=1}}\{g(a, z, W)\} = E_{F_{W|Z=z}}\{g(a, z, W)v_z(W)\} \geq E_{F_{W|Z=z}}\{g(a, z, W)\}$$

and so

$$S_1 = \sum_z E_{F_{W|Z=z,A=1}}\{g(a, z, W)\}pr(Z = z) \geq \sum_z E_{F_{W|Z=z}}\{g(a, z, W)\}pr(Z = z) = E(Y_{A=1}).$$

The proof that $S_0 \leq E(Y_{A=0})$ is similar.

References


