working paper

15 05

A Distinction between Causal Effects in Structural and Rubin Causal Models

Dionissi Aliprantis

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 15-05 March 2015

A Distinction between Causal Effects in Structural and Rubin Causal Models

Dionissi Aliprantis

Structural Causal Models define causal effects in terms of a single Data Generating Process (DGP), and the Rubin Causal Model defines causal effects in terms of a model that can represent counterfactuals from many DGPs. Under these different definitions, notationally similar causal effects make distinct claims about the results of interventions to the system under investigation: Structural equations imply conditional independencies in the data that potential outcomes do not. One implication is that the DAG of a Rubin Causal Model is different from the DAG of a Structural Causal Model. Another is that Pearl's do-calculus does not apply to potential outcomes and the Rubin Causal Model.

JEL codes: C00, C01, C31, C45.

Keywords: Structural Equation, Potential Outcome, Invariance, Autonomy.

Suggested citation: Aliprantis, Dionissi, 2015. "A Distinction between Causal Effects in Structural and Rubin Causal Models," Federal Reserve Bank of Cleveland, working paper no. 15-05.

Dionissi Aliprantis is at the Federal Reserve Bank of Cleveland, and he can be reached at dionissi.aliprantis@clev.frb.org. He thanks Francisca G.-C. Richter for helpful comments.

1 Introduction

Since several formal results have been proven about their relationship, Structural Causal Models (SCMs) and Rubin Causal Models (RCMs) are often viewed as analogues (Pearl (2014b), Pearl (2012)).[1] For example, a latent index model of selection is equivalent to selection under the assumptions of independence and monotonicity (Vytlacil (2002)). In addition, SCMs are formally equivalent to RCMs in the following sense: There exists a SCM generating the hypothetical contingency tables represented by any RCM (Pearl (2009b) p 244, Halpern (2000), Galles and Pearl (1998)). This paper shows that different approaches to modeling outcomes make SCMs and RCMs distinct in the following sense: There does not exist a unique SCM generating the hypothetical contingency tables represented by a given RCM.

A simultaneous strength and weakness of potential outcomes is that they do not represent an "all causes" model (Heckman and Vytlacil (2007)). Thus, while potential outcomes are able to characterize Data Generating Processes (DGPs) without knowing all relevant causes (Freedman (1987), Imai et al. (2010)), they require similar assumptions to structural equations when used to make out of sample forecasts (Heckman and Vytlacil (2005), Aliprantis (2014)). One attempt to connect internal and external validity characterizes potential outcomes in terms of structural equations (Heckman (2010), Pearl (2014a)).

When comparing potential outcomes and structural equations, it is important to note that there are at least two definitions of a structural equation. Under these alternatives, structural equations represent relationships with varying degrees of autonomy, or invariance when other relationships are subject to external influence (Aldrich (1989), Woodward (2000), Pearl (2009b) Section 2.9.3). Definition 1 is that the outcome variable can be written as a structural equation of the treatment variable if the specified relationship describes changes from interventions manipulating only treatment. Definition 2 is that the outcome variable can be written as a structural equation of the treatment variable if the specified relationship describes changes from interventions manipulating any variables other than the outcome.

This paper shows that distinguishing between these definitions of structural equation matters for inductive inference with causal effects. Specifically, while potential outcomes can be considered structural equations under Definition 1 (Angrist et al. (1996a)), they are not structural equations under Definition 2 (Pearl (2009b), Definition 5.4.1). These alternative definitions create a divergence between the modeling of outcomes in SCMs and RCMs.[2]

Specifically, SCMs define causal effects in terms of a single DGP specified by nature, while potential outcomes define causal effects in terms of a model specified by the researcher that can accommodate many DGPs. As a result, outcome equations in SCMs imply conditional independencies in the data that similarly specified outcome equations in RCMs do not. Two implications are that: (1) The Directed Acyclic Graph (DAG) of a RCM is quite different from the DAG of a

[1] Alternatively labeled as structural equation modeling (SCMs) and the potential outcome framework (RCMs).

[2] Note that the distinction here is due to the outcome equations, and not the selection models whose typical specifications are equivalent (Vytlacil (2002)).

SCM. And (2) Pearl's *do*-calculus does not apply to potential outcomes and the RCM.

2 Defining Structural Equations

Distinguishing between definitions of structural equations, and the resulting differences in causal effects, differentiates between an "all causes" model assumed to represent the DGP and a model capable of representing counterfactuals from any number of DGPs. To be clear about this point, I distinguish between the model at time t, $\mathcal{M}_t$, which is a set of relationships between variables specified by the econometrician, and the DGP $\mathcal{D}_t$, which is the set of relationships between variables specified by nature. The following definitions of structural equation might be associated with Angrist et al. (1996a) (p 450) and Pearl (2009b) (Definition 5.4.1):

Definition 1 $Y_i = Y_i(D)$ is a structural equation *if there exists some variable* Z such that

$$Z \cap Y = Z \cap D = \emptyset \quad \text{and} \quad Y_i(d, z') = Y_i(d, z) \quad \text{for all} \quad z' \neq z$$

Definition 2 $Y_i = Y_i(D, \epsilon)$ is a structural equation *if for any observable variable* M,

$$M \cap Y = M \cap D = M \cap \epsilon = \emptyset \quad \text{implies that} \quad Y_i(D, m', \epsilon) = Y_i(D, m, \epsilon) \quad \text{for all} \quad m' \neq m$$

To investigate the implications of adopting one of these definitions rather than the other, consider the following class of DGPs, which represent a typical mediation system without confounders (Pearl (2014a)).[3] Suppose that at time $t \in \mathbb{N}$ data are generated by a DGP in which the outcome variable (Y_{ti}) for each individual i is causally determined by two observed variables, treatment (D_{ti}) and observed covariates (X_{ti}), as well as unmeasured covariates (U_{ti}^Y), or additional factors not observed by the econometrician. The unmeasured covariates can be broken down into those factors that are unobserved (E_{ti}) and those that are unobservable (ϵ_{ti}) at the given level of measurement. In order to focus on unobserved mediators and outcome equations, I consider DGPs without unobserved confounders.

If Z is an instrument mimicking an observable intervention manipulating treatment and U represents unmeasured variables, then the DGP $\mathcal{D}_t$ is characterized by the following structural equations:[4]

$$Z_{ti} \leftleftarrows f_t^Z(U_{ti}^Z) \tag{1}$$
$$D_{ti} \leftleftarrows f_t^D(Z_{ti}, U_{ti}^D) \tag{2}$$
$$X_{ti} \leftleftarrows f_t^X(D_{ti}, U_{ti}^X) \tag{3}$$
$$Y_{ti} \leftleftarrows f_t^Y(D_{ti}, X_{ti}, U_{ti}^Y) \tag{4}$$

[3] These DGPs are specified to represent simple systems amenable to potential outcomes (Pearl et al. (2014)).

[4] In Pearl (2009b)'s Definition 7.1.1 of a structural causal model, the triple $< U, V, F >$ is defined here as $U \equiv (U_{ti}^Z, U_{ti}^D, U_{ti}^X, U_{ti}^Y)$, $V \equiv (Z_{ti}, D_{ti}, X_{ti}, Y_{ti})$, $F \equiv (f_{ti}^Z, f_{ti}^D, f_{ti}^X, f_{ti}^Y)$. In this specification Z_{ti} randomizes D_{ti} so as to satisfy the standard ignorability assumption (Imbens (2014), White and Lu (2011)), and Z_{ti} separately satisfies the relevant exclusion restriction from the outcome equation (Angrist et al. (1996b)).

The $\Leftarrow$ notation is used to indicate that these are structural equations under Definition 2. Under both definitions, a structural equation contains information about counterfactual manipulations to the right hand side variables, and not only information about what is passively observed in the data. This represents an asymmetric relationship between the variables on the left and right hand sides of the equation.[5] For example, structural Equation 4 provides information about the counterfactual values of Y_{ti} if we were to control the variables on the *rhs*, but makes no claim about how any of the variables on the *rhs* would behave if we were to control the *lhs* variable Y_{ti} (Pearl (2009b), p 160). A standard equation makes no distinction between interventions counterfactually manipulating right hand side variables and information about what is passively observed in the data.[6]

Under Definition 2, a further implication of the $\Leftarrow$ notation is that given control over the specified observable (but not necessarily observed) variables on the *rhs* of the equation, changes made to additional observable variables would provide no further change to the outcome variable. This is what distinguishes Definition 2: All variables at the given level of observation not included on the right hand side satisfy an exclusion restriction (Pearl (2009b), Definition 5.4.1). Definition 1 requires only that an exclusion restriction must hold for *one* additional variable not included on the *rhs* (the one manipulating treatment), not *all* variables not included as arguments of the outcome function (Angrist et al. (1996a)).

3 Defining Causal Effects

3.1 Defining Causal Effects as Changes from Interventions to a DGP

One definition of causal effects is as a quantitative characterization of the change in the outcome variable that would result from an intervention to the DGP. Such interventions to the DGP can be characterized by how they would, or would not, impact covariates, especially unmeasured variables. In order to be precise about which features of the DGP are changed, and which are not, under specific interventions, I use Pearl (2009b)'s *do*-operator throughout the remainder of the analysis.

Direct effects characterize the change in the outcome variable from a specific type of intervention to the DGP. Specifically, the controlled direct effect of D_t on Y_t, $\Delta_{ti}^{CDE}(d', d)$, is the change in Y_t that would result from an intervention setting D_t from d to d' while setting all other variables entering as arguments in f^Y to fixed values:

$$\begin{aligned}\boldsymbol{\Delta}_{\mathbf{t}}^{\mathbf{CDE}}(\mathbf{d}', \mathbf{d}) &\equiv \mathbb{E}[f_t^Y(d', x, U_{ti}^Y)] - \mathbb{E}[f_t^Y(d, x, U_{ti}^Y)] \\ &= \mathbb{E}[Y_{ti}|do(D_{ti} = d', X_{ti} = x)] - \mathbb{E}[Y_{ti}|do(D_{ti} = d, X_{ti} = x)]\end{aligned}$$

Following Pearl (2014a), this definition is made at the population level, with individual-level ef-

[5]Some examples of the asymmetry of "directions of influence" (Strotz and Wold (1960)) include the fact that symptoms do not cause disease (Pearl (2009a)), a child's height does not cause her father's height (Goldberger (1991)), and rainfall determines crop yields but not the reverse.

[6]Chalak and White (2012) study these distinctions in terms of settings and responses, and White and Chalak (2009) allow for causal effects in systems of symmetric equations.

fects given by the expressions under the expectation. Expectations are taken over U_{ti}^Y for the $\Delta_{ti}^{CDE}(d', d)$.

Direct effects are often defined relative to a reference set of variables that are set to fixed values by an intervention, and the values to which the reference variables are set (Spirtes et al. (2001)). Under such a definition, direct effects need not be invariant to interventions changing variables outside that reference set. This direct effect, however, is invariant to changes to any variables outside the parents of Y_{ti}, as the reference set in this definition is all other variables at a given level of measurement. This is because the definition of structural equation adopted in this paper indicates that *all* variables at the given level of observation not included on the *rhs* of Equation 4 satisfy an exclusion restriction. In other words, because we are examining the DGP and not a model of it, mediators outside the reference set can only be found at a finer level of observation (A process which, as noted by Holland (1988), always appears to be possible.).

A second useful definition of causal effect is the total effect:

$$\begin{aligned}\mathbf{\Delta_t^{TE}}(\mathbf{d}', \mathbf{d}) &\equiv \mathbb{E}\big[\, f_t^Y\big(\; d', \quad f_t^X(d', U_{ti}^X), \quad U_{ti}^Y\big)\,\big] \; - \; \mathbb{E}\big[\, f_t^Y\big(\; d, \quad f_t^X(d, U_{ti}^X), \quad U_{ti}^Y\big)\,\big] \\ &= \mathbb{E}[Y_{ti}|do(D_{ti} = d')] - \mathbb{E}[Y_{ti}|do(D_{ti} = d)].\end{aligned}$$

As with the $\Delta_{ti}^{CDE}(d', d)$, this definition is also made at the population level, with individual-level effects given by the expressions under the expectation. Expectations are taken over U_{ti}^X and U_{ti}^Y for the $\Delta_{ti}^{TE}(d', d)$.

Defining the vector $S \equiv (D, X)$ and re-writing Equation 4 more compactly as $Y_{ti} \stackrel{\leftarrow}{=} Y_{ti}(S_{ti})$, both direct and total effects can be written in terms of the econometric or graphical definitions given in Heckman (2008) and Pearl (2009b) (Definition 3.2.1).

3.2 Defining Causal Effects as Changes from Interventions to a Model

The Rubin Causal Model (Rubin (2005), Angrist et al. (1996a)) defines causal effects in terms of the counterfactual outcome variable that would be observed under interventions to treatment. These counterfactual outcomes are also known as potential outcomes,

$$Y_{ti}(D_{ti}),$$

where $Y_{ti}(d)$ is the outcome of individual i at time t if treatment were set to $D_{ti} = d$ by an intervention setting D_{ti} but affecting none of the mediators of the total effect of D_{ti} on Y_{ti} (ie, none of the other parents of Y_{ti}). Although potential outcomes are generated by the DGP, they are defined as features of a model $\mathcal{M}_t$ that can describe many DGPs. That is, the average causal effect in the Rubin Causal Model is defined as

$$\mathbf{\Delta_t^{RCM}}(\mathbf{d}', \mathbf{d}) \equiv \mathbb{E}[Y_{ti}(d') - Y_{ti}(d)].$$

The expectation in $\Delta_t^{RCM}(d', d)$ is taken over individuals in the given population, allowing for any number of underlying functional forms and distributions. In contrast, the expectation in $\Delta_t^{TE}(d', d)$ is taken with respect to the single set of functional forms and distributions specified by the DGP.

4 Interventions to a DGP versus Interventions to a Model

The causal effect from the potential outcome definition represents the change in the outcome variable from a randomized variation in treatment. The causal effect from the structural equation definition represents the change in the outcome variable from a randomized *and* controlled variation in treatment. This difference can be illustrated by the fact that the inclusion or exclusion of a variable X_{ti} from the potential outcome $Y_{ti}(D_{ti})$ has no implication for how an intervention manipulating X_{ti} would or would not change Y_{ti}. This contrasts with the inclusion or exclusion of a variable X_{ti} from the structural equation f_t^Y. Inclusion implies that manipulating X_{ti} would change Y_{ti}, while exclusion from f_t^Y implies that manipulating X_{ti} would not change Y_{ti}.

More formally, consider the structural and potential outcome equations:

$$Y_{ti} \leftleftarrows f_t^Y(D_{ti}, X_{ti}, \epsilon_{ti}), \tag{5}$$

$$Y_{ti} = Y_{ti}(D_{ti}, X_{ti}). \tag{6}$$

The distinction between these equations is illustrated in Figures 1 and 2, in which it is useful to note that I denote unobserved variables with empty nodes and dashed lines. The potential outcome $Y(D, X)$ is consistent with many DGPs, each containing any number of unobserved mediators. The structural equation $f^Y(D, X, U^Y)$ is only consistent with DGPs with the specified parents of Y.[7]

The structural outcome equation 5 implies that for any observable variable that could potentially be a mediator at the given level of observation, M_{ti} , it is excluded from f_t^Y. That is, $Y(D, X, M) = Y(D, X)$, or

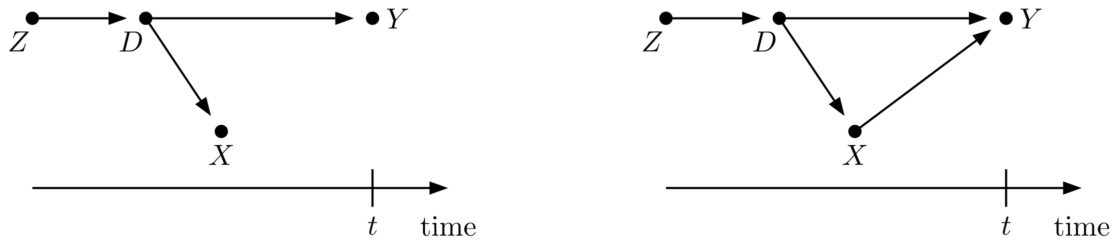
$$Y \perp\!\!\!\perp M | D, X \quad \forall \quad M. \tag{7}$$

Equation 6 makes no such claim about the data. This matters because Equation 7 implies the first rule of Pearl's *do*-calculus, that insertion/deletion of observations is valid (Pearl (2009b), p 85):

Rule 1 $P(y|do(d), m, x) = P(y|do(d), x) \quad \text{if} \quad \left(Y \perp\!\!\!\perp M|D, X\right)_{G_{\overline{D}}}.$

Since Equation 6 does not imply Rule 1, neither do potential outcomes. The exclusion restrictions implied by Definition 2 of structural equation (Pearl (2009b), p 101 and p 160) are required for Pearl's *do*-calculus.

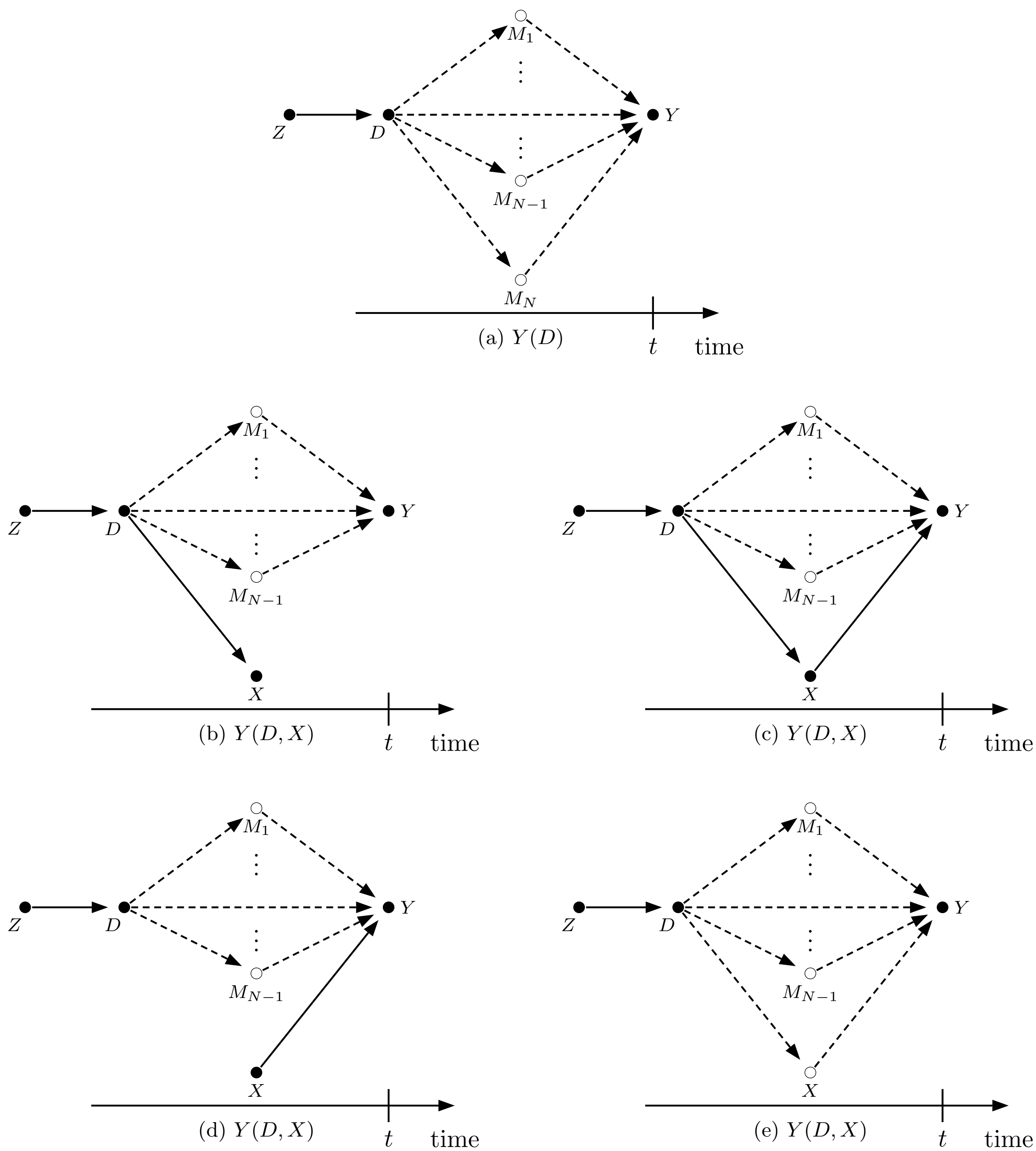
[7] Whether inclusion of a variable as an argument in a structural equation requires that it is a causal variable hinges on whether we interpret Definition 5.4.1 in Pearl (2009b) as a conditional or biconditional statement (ie, an "if" or an "if and only if" statement). For example, the triangular system in White and Chalak (2013) does not require that an argument in a structural equation be a cause, so that in their analysis the converse of the statement in Definition 5.4.1 in Pearl (2009b) need not be true. In this analysis I interpret Definition 5.4.1 as a biconditional statement.



(a) $f^Y(D, U^Y)$

(b) $f^Y(D, X, U^Y)$

Figure 1: Directed Acyclic Graphs of DGPs Accommodated by Structural Outcome Equations



(a) $Y(D)$

(b) $Y(D, X)$

(c) $Y(D, X)$

(d) $Y(D, X)$

(e) $Y(D, X)$

Figure 2: Directed Acyclic Graphs of DGPs Accommodated by Potential Outcomes

References

Aldrich, J. (1989). Autonomy. *Oxford Economic Papers 41*, 15–34.

Aliprantis, D. (2014). Covariates and causal effects: The problem of context. *Federal Reserve Bank of Cleveland Working Paper 13-10R.*

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996a). Identification of causal effects using Instrumental Variables. *Journal of the American Statistical Association 91*(434), 444–455.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996b). Rejoinder. *Journal of the American Statistical Association 91*(434), 465–468.

Chalak, K. and H. White (2012). Causality, conditional independence, and graphical separation in settable systems. *Neural Computation* (24), 1611–1668.

Freedman, D. A. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics 12*(2), 101–128.

Galles, D. and J. Pearl (1998). An axiomatic characterization of causal counterfactuals. *Foundations of Science 3*(1), 151–182.

Goldberger, A. S. (1991). *A Course in Econometrics*. Harvard University Press.

Halpern, J. Y. (2000). Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research 12*(1), 317–337.

Heckman, J. J. (2008). Econometric causality. *International Statistical Review 76*(1), 1–27.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of the Economic Literature 48*(2), 356–398.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. Vytlacil (2007). Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 70, pp. 4779 – 4874. Elsevier.

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology 18*(1), 449–484.

Imai, K., L. Keele, and D. Tingley (2010). A general approach to causal mediation analysis. *Psychological Methods 15*(4), 309–334.

Imbens, G. W. (2014). Instrumental Variables: An econometrician's perspective. *Statistical Science 29*(3), 323–358.

Pearl, J. (2009a). Causal inference in statistics: An overview. *Statistics Surveys 3*, 96–146.

Pearl, J. (2009b). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press.

Pearl, J. (2012). The causal foundations of Structural Equation Modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling*, Chapter 5, pp. 68–91. New York: Guilford Press.

Pearl, J. (2014a). The deductive approach to causal inference. *Journal of Causal Inference 2*(2), 115–129.

Pearl, J. (2014b). Interpretation and identification in causal mediation analysis. *Psychological Methods 19*(4), 459–481.

Pearl, J., G. Imbens, B. Chen, and E. Bareinboim (2014, October-November). Are economists smarter than epidemiologists? (Comments on Imbens' recent paper). *UCLA Causality Blog*. http://www.mii.ucla.edu/causality/?p=1241.

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association 100*(469), 322–331.

Spirtes, P., C. Glymour, and R. Scheines (2001). *Causation, Prediction, and Search* (2nd ed.). Bradford.

Strotz, R. H. and H. O. Wold (1960). Recursive vs. nonrecursive systems: An attempt at synthesis. *Econometrica 28*(2), 417–427.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica 70*(1), 331–341.

White, H. and K. Chalak (2009). Settable systems: An extension of Pearl's Causal Model with optimization, equilibrium, and learning. *Journal of Machine Learning Research* (10), 1759–1799.

White, H. and K. Chalak (2013). Identification and identification failure for treatment effects using structural systems. *Econometric Reviews 32*(3), 273–317.

White, H. and X. Lu (2011). Causal diagrams for treatment effect estimation with application to efficient covariate selection. *Review of Economics and Statistics 93*(4), 1453–1459.

Woodward, J. (2000). Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science 51*, 197–254.