Causal Inference and Data-Fusion in Econometrics

(Preliminary and incomplete version)

Elias Bareinboim* Paul Hünermund†

10 February 2019

1. INTRODUCTION

Causal inference is arguably the most important goal in applied econometric work. In order to make informed decisions, policy-makers and managers need to be able to forecast the likely impact of the actions they consider. Providing said knowledge by uncovering quantitative relationships in statistical data is the goal of econometrics since the beginning of the discipline (Frisch, 1933). In recent decades, after a decline of interest in the postwar period (Hoover, 2004), causal inference was again brought to the forefront of the methodological discussion by the emergence of the potential outcome framework (Rubin, 1974; Imbens and Rubin, 2015) and advances in structural econometrics (Heckman and Vytlacil, 2005, 2007; Matzkin, 2013).

Woodward (2003) defines causal knowledge as "knowledge that is useful for a very specific kind of prediction problem: the problem an actor faces when she must predict what would happen if she or some other agent were to act in a certain way [...]".[1] This association of causation with control in a stimulus-response-type

*Purdue University, Department of Computer Science. 305 N. University Street 2142L West Lafayette, IN, 47907-2107. Email: eb@purdue.edu

†Maastricht University, School of Business and Economics. Tongersestraat 53, 6211 LM Maastricht, Netherlands. Email: p.hunermund@maastrichtuniversity.nl

[1] Woodward continues: "[...] on the basis of observations of situations in which she or the other agent have not (yet) acted" (p. 32).

relationship is likewise foundational for econometric methodology. According to Strotz and Wold (1960), "z is a cause of y if [...] it is or 'would be' possible by *controlling* z indirectly to control y, at least stochastically" (p. 418; emphasis in original).

Although already implicit in earlier treatments (e.g., in Haavelmo, 1943), Strotz and Wold (1960) were the first to express actions and control of variables as *"wiping out"* of equations in an economic system (Pearl, 2009, p. 32). To illustrate this idea, consider the two-equation model

$$z = f_1(x, \varepsilon_1) \tag{1.1}$$

$$y = f_2(z, x, \varepsilon_2) \tag{1.2}$$

in which y might represent earnings obtained in the labor market, z the years of education an individual received, and x refers to other relevant socio-economic factors. Since x enters in both equations, it creates a correlation between z and y that is not due to a causal influence of schooling on earnings. Thus, in order to predict how y reacts to changes in z, the causal mechanism that usually determines schooling needs to interrupted in order to create exogenous variation in z. This is achieved by replacing $f_1(\cdot)$ from the model and instead fixing z at a constant value $z = z_0$. Subsequently, the quantitative impact of such an intervention on y can be traced in order to pin down z's causal effect.

The notion of "wiping out" equations as a fundamental building block of the conceptualization of causality eventually received full formal treatment in form of the *do*-operator (Pearl, 1995). Consider the task of predicting the post-intervention distribution of a random variable Y that is the result of a change in X. In mathematical notation this can be written as $Q = P(Y = y|do(X = x))$. Causal inference requires the analyst to answer the query Q with the help of information that is available to her at a certain point in time. For example, if experimental manipulation is currently not feasible, $P(Y = y|do(X = x))$ will not be directly computable. Instead, the expression first needs to be transformed into a standard probability object only involving ex-post observable quantities before estimation can proceed. The symbolic language that warrants such kinds of syntactic transformations is called *do-calculus* (Pearl, 1995).

Do-calculus, as a causal inference engine, takes three inputs: (1) a causal query Q; (2) a model G that encodes assumptions about the causal dependencies between the economic variables under study; and (3) the type of data $P(v)$ that are available to the analyst (e.g., observational, experimental, measured with or without selection bias, etc.). It consists of three inference rules for transforming probabilistic sentences involving do-expressions. By repeatedly applying these rules, do-calculus then returns a solution to the identification problem of expressing Q as a function of $P(v)$, whenever this syntactic target can be reached.

Do-calculus complements standard approaches in econometrics in two important ways. First, it builds on a mathematical formalism borrowed from graph theory, which describes causal models as a set of nodes in a network, connected by directed edges (Neuberg, 2003). This description does not rely on any functional-form restrictions imposed on the relationships between economic variables. Therefore, the approach provides a formal treatment of nonparametric causal inference in full generality. Second, as a subfield of computer science, the literature on graph-theoretic treatments of causality has developed algorithmic solutions to a wide variety of causal inference problems arising in applied work. These algorithms are able to solve the syntactic transformation described above – mapping a query to available data – completely automatically. From do-calculus the algorithms inherit the property of *completeness* (Shpitser and Pearl, 2006; Huang and Valtorta, 2006). This means that the procedure is guaranteed to return a solution whenever one exists.[2] Consequently, for the class of models in which these algorithmic formulations of do-calculus are applicable, the identification problem is fully solved (Pearl, 2013).

The development of do-calculus gave the literature on causal inference within the field of computer science a tremendous boost and many significant contributions have been made since Pearl (2000) published his seminal work. The aim of this paper is to review these more recent developments. In particular, we show how do-calculus can be applied to solve many recurrent problems in causal inference. Thereby, our hope is to convince the reader that graph-theoretic approaches to

[2] Conversely, and importantly, if the procedure fails to provide an answer to a causal query, no such answer will be obtainable unless the assumptions imposed on the model are strengthened (Pearl, 2013).

causation also make an important contribution to the econometric toolbox. The three main topics we cover are: (1) dealing with confounding bias; (2) recovering from selection bias; and (3) transporting causal knowledge across populations.

Confounding bias. In most applied settings, the distribution of Y following an intervention on X, $P(Y|do(X))$, does not coincide with the conditional distribution $P(Y|X)$. This is the result of confounding influence factors, which produce a correlation among the variables irrespective of any causal relationship. Nevertheless, the inference rules of do-calculus can be used to transform $P(Y|do(X))$ into an expression – generally different from $P(Y|X)$ – that is estimable from available data. If a solution containing standard probability objects can be reached, the confounding is solvable with the help of observational data alone. Additionally, sometimes the analyst actually is able to experimentally manipulate a third variable Z that is causally related to the treatment of interest. In such situations (one example would be the classic *encouragement design*; Duflo et al., 2008), the identification problem can be relaxed because an estimable transformation of $P(Y|do(X))$ can now also involve a $do(Z)$-operator.

Sample selection bias. A common threat to valid inference in practice is sample selection bias, which occurs if the analyst is only able to observe information for members of the population that possess specific characteristics or fulfill certain requirements (e.g., market wages are only observable if individuals are employed; Heckman, 1979). Selection-biased data aggravate the identification problem since $P(Y|do(X))$ needs to be transformed into an expression solely comprised of probabilities from a non-random sample (inclusion in the selected sample is usually denoted by an indicator S, which implies that only probabilities conditional on $S = 1$ are observable). The inference rules of do-calculus provide a principled and complete solution for carrying out this task.

Transportability of causal knowledge. Another important topic in econometric practice is the question of external validity. Causal knowledge is usually acquired in a specific population (e.g., for participants in a laboratory setting), but needs to be brought to productive use in other domains in order to be most valuable. What permits such a transportation of causal knowledge across settings, however, if the underlying populations differ structurally in important ways? To answer this question, inference rules are required that make it possible to express a causal query for a target domain in terms of causal knowledge from a source domain. These inference rules are provided by the do-calculus. Moreover, the transportability problem can be extended to combine causal knowledge from several, possibly heterogeneous source domains (a strategy generally known under the name *"meta-analysis"*). Thereby, do-calculus opens up entirely new possibilities for leveraging results from a whole body of empirical literature in order to address policy questions arising in new, still understudied contexts.

The aforementioned topics may seem diverse, yet they share a common structure. Data, which are created in various different ways (e.g., from observational or experimental studies, from non-random sampling, or from heterogeneous underlying populations), are combined in order to answer a causal query of interest. For this strategy of *"data fusion"* to be viable, the analyst needs to be equipped with a powerful inference engine that licenses this kind of information transfer. We will describe such an inference engine in detail in the remainder of the paper.

2. PRELIMINARIES: STRUCTURAL CAUSAL MODELS, GRAPHS, AND INTERVENTIONS

In this section, we introduce a graph-theoretic framework for causal inference and put forward a theory of interventions in structural causal models as a fundamental element of causality. Following Pearl (2009, p. 203) a structural causal model (SCM) is defined as follows[3]

Definition 2.1. *(Structural causal model) A structural causal model is a 4-tuple $M = \langle U, V, F, P \rangle$ where*

1. *U is a set of background or exogenous variables, representing factors that are determined outside of the model and affect relationships within the model.*
2. *$V = \{V_1, \dots, V_n\}$ is a set of endogenous variables that are determined within the model. Each V_i is functionally dependent on some subset PA_i of $U \cup V$.*
3. *F is a set of functions $\{f_1, \dots, f_n\}$ such that each f_i determines the value of $V_i \in V$, $v_i = f_i(pa_i, u_i)$.*

[3] Structural causal models are nonparametric versions of structural equation models (SEM). The term SCM is used to avoid confusion with the vast literature on SEM that usually assumes parametric or even linear functional forms.

4. $P(u)$ is a joint probability distribution over U.

Each f_i constitutes a causal process or mechanisms that determines the value of the left-hand side variable given the variables on the right-hand side. They represent the naturally occurring data generating process (DGP) and are assumed to be invariant unless explicitly intervened on. In a fully specified model $\langle U, V, F, P(u)\rangle$, any counterfactual quantity is well-defined and computable from the model. In many social science contexts, however, precise knowledge of the functional relationships that govern the DGP is not available. In the following, we will thus advocate an approach that relies on a smaller set of assumptions, building on a graphical representation of a structural model.

Every SCM can be associated with a directed graph $G(M)$. Nodes in G correspond to variables and directed edges point from a set of parent nodes PA_i and U_i toward V_i.[4] An example is given by Figure 1a, which refers to the following SCM[5]

$$\begin{aligned} z &= f_Z(u_z), \\ x &= f_X(z, u_X), \\ y &= f_Y(x, z, u_Y). \end{aligned} \tag{2.1}$$

In this example, U_X, U_Z, and U_Y are assumed to be jointly independent. Dependence between exogenous variables due to unaccounted confounding influences can be denoted by bidirected dashed arrows as in Figure 2a. For the sake of visibility, one usually omits the background factors U_i from the graph and we shall do so in the following.[6] A graph is called *acyclic* if it contains no directed cycles. Figure 1a is an example for such a directed acyclic graph (DAG), as it contains no sequences of edges that point from a variable back to itself (i.e., no feedback loops).

Working with the graphical representation of M entails a deliberate choice by the analyst to refrain from parametric and functional form assumptions, as the shape of f_i and the distribution of background factors U remain completely unspecified.

[4] It is customary to use the notation of kinship relations (parents, children, ancestors, descendants, etc.) to describe the relative position of nodes in directed graphs.

[5] We follow the usual notation and denote realized values of variables by lowercase letters.

[6] These exogenous background factors correspond to what is often referred to as “error terms” in classical econometrics. However, we deliberately avoid this terminology to emphasize that the U_i’s in an SCM have a causal nature, in contrast to the purely statistical notion of a prediction error or deviation from the conditional mean.

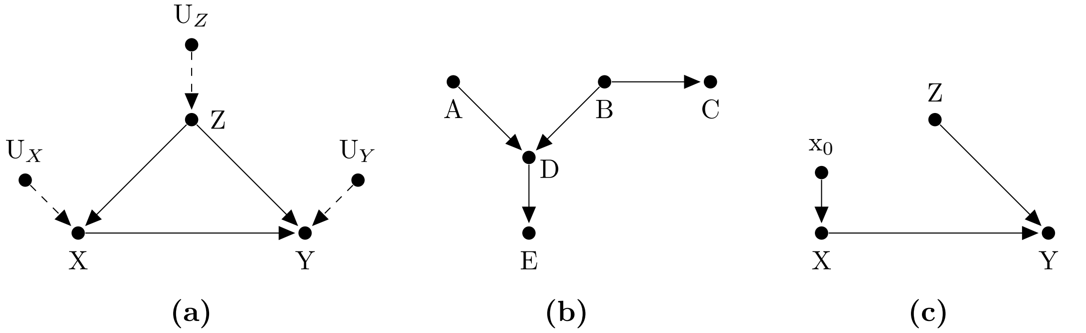


Figure 1: *(a) Directed acyclic graph corresponding to SCM in equation (2.1) with background variables U_i explicitly depicted. (b) Graphical illustration of* d-*separation. D acts as a collider that opens up the path between A and C, whereas B blocks it. (c) Post-intervention graph of (a) for $do(X = x_0)$.*

Consequently, graphical models are fully nonparametric. This constitutes an important difference to the "structural econometrics" literature, which often assumes specific parametric error distributions (such as the normal or logistic distribution) or imposes shape restrictions on functions (such as separability, monotonicity, or differentiability) in order to achieve identification (Heckman and Vytlacil, 2007; Matzkin, 2007, 2013). In certain applications, these distributional and fuctional-form assumptions might be licensed by economic theory (Matzkin, 2013). If they are not, however, we concur with Manski (2003) that it is a more robust research approach to start with the most flexible model possible and only later resort to parametric and functional form assumptions, if the explanatory power of nonparametric approaches has been exhausted. In line with this philosophy, the techniques we present in the following explore the possibilities to identify causal effects from data when only knowledge about the graph G is available, and G is known to be acyclic.[7]

One key feature of graphical causal models is their ability to efficiently encode conditional independence relationships between variables in the model, which are characterized by the following definition (Pearl, 1988).

[7] For extensions to the SCM framework that incorporate cyclic graphs the interested reader is referred to discussions in Spirtes et al. (2001, ch. 12) and Pearl (2009, ch. 3.6). The requirement of acyclicity implies that the underlying structural model is recursive. Appendix A.1 provides a brief discussion of the diverging notions of "causality" in recursive versus nonrecursive economic systems.

Definition 2.2. *(D-separation) A set Z of nodes is said to block a path p if either*

1. *p contains at least one arrow-emitting node that is in Z*
2. *p contains at least one collision node that is outside Z and has no descendant in Z*

If Z blocks all *paths from set X to set Y, it is said to "d-separate X and Y", and then it can be shown that variables X and Y are independent given Z, written $X \perp\!\!\!\perp Y|Z$.*[8]

Conditional independence licensed by d-separation (d for "directional") holds for any distribution $P(v)$ compatible with the causal assumptions encoded in the graph; *regardless* of the parametrization of the arrows. An example is given in Figure 1b, where the path $A \rightarrow D \leftarrow B \rightarrow C$ is blocked by $Z = \{D\}$, since B emits arrows on that path. Consequently, we can infer the conditional independencies $A \perp\!\!\!\perp C|B$ and $D \perp\!\!\!\perp C|B$. In fact, A and D are independent conditional on the empty set $\{\emptyset\}$ too. Because B acts as a so-called *collider* node, due the two arrows pointing into it, it blocks the path between A and C, according to the second condition of Definition 2.2. Conversely, conditioning on a collider would open up a path that has been previously blocked; thus, $A \not\!\perp\!\!\!\perp C|D$. The same holds for descendants of colliders such as E in Figure 1b.

D-separation allows one to read the conditional independencies implied by the model in the observed data, which constitutes a way of testing whether the model and data are actually compatible (testable implications of the model). The full list of conditional independence relations implied by the graph in Figure 1b is given by

$$\begin{array}{cccc} A \perp\!\!\!\perp B; & A \perp\!\!\!\perp C; & A \perp\!\!\!\perp E|D; & B \perp\!\!\!\perp E|D; \\ & C \perp\!\!\!\perp D|B; & C \perp\!\!\!\perp E|D; & C \perp\!\!\!\perp E|B. \end{array} \tag{2.2}$$

If these relations do not hold in the data, the hypothesized model can be rejected. Moreover, in contrast to global goodness-of-fit tests to discriminate between different models, violations of particular conditional independence relations provide the analyst with concrete clues on where the model is incompatible with the observed data.

[8] See Verma and Pearl (1988). A path is any consecutive sequence of edges in a graph regardless of their orientation.

Conditional independence relationships licensed are a main building block in causal inference, as we will further explicate in Section 3. With the help of d-separation they can be discerned simply based on the topolgy of the graph. DAGs are thus a valuable complement to the treatment effects literature, where independence assumptions for counterfactuals, such as "ignorability" or "unconfoundedness" (Imbens and Rubin, 2015), are invoked without a reference to a underlying structure describing the phenomenon that is being investigated. As a consequence, the analyst has no guidance for scrutinizing the plausibility of crucial identifying assumptions – a task which is rendered much more transparent with the help of causal diagrams.

2.1. Interventions and identification of causal effects

Causal inference aims at predicting the effects of interventions, such as those resulting from policy actions, social programs, or management initiatives (Woodward, 2003). Based on early ideas from the econometrics literature (Haavelmo, 1943; Strotz and Wold, 1960; Pearl, 2015b), interventions in structural causal models are carried out by deleting certain functions from the model and replacing them with a constant $X = x$.[9] This action is denoted by a mathematical operator called $do(x)$. For example, in model M of Figure 2.1, the action $do(X = x_0)$ results in the post-intervention model M_{x_0},

$$\begin{aligned} z &= f_Z(u_z), \\ x &= x_0, \\ y &= f_Y(x, z, u_Y). \end{aligned} \tag{2.3}$$

The graph associated with M_{x_0} is depicted in Figure 1c, where all incoming arrows into X from the original graph are deleted and replaced by $x_0 \to X$. This captures the notion that an intervention interrupts the original treatment assignment mechanism and thus eliminates all naturally occurring influence factors on the now intervened variable. Differences between the two probability distributions

[9] The early literature on Bayesian networks relied entirely on probabilistic models, which were unable to answer counterfactual queries (Pearl and Mackenzie, 2018, p. 284f.). A major intellectual breakthrough was achieved in the early 1990s by changing over to the quasi-deterministic functional relationships of the sort that are ubiquitous in econometrics (Pearl, 2009, p. 104f.).

associated with M_{x_0} and M_{x_1} capture the variations of an outcome Y that is exactly due to the causal variations of X, irrespective of the original confounding that is present in the natural regime. Not coincidentally, these distributions are also the ones entailed by a randomization device in an experiment that overwrites the original function f_X and assigns values to X based on an external source of variation.

The post-intervention distribution of Y can also be defined in counterfactual notation[10]

$$P(y|do(x)) \triangleq P(Y_x = y). \tag{2.4}$$

Definition 2.4 illustrates the connection to the potential outcome framework (Neyman, 1923; Rubin, 1974; Imbens, 2004), where counterfactuals such Y_{x_0} and Y_{x_1} are taken as primitives. Alternatively, we can see that in a structural model

$$\begin{aligned} Y_{x_0} &= f(x_0, z, u_Y), & (2.5)\\ Y_{x_1} &= f(x_1, z, u_Y), & (2.6) \end{aligned}$$

which follow immediately from M_{x_0} and M_{x_1}. In other words, in an SCM counterfactuals are *derived* from first principles, i.e., the structural mechanisms that underlie the system that is being investigated.

Equipped with clear semantics of causal effects in terms of the underlying causal processes, we can now consider the problem of nonparametric identification.

Definition 2.3. *(Identifiability[11], Pearl, 1993) A causal query Q is identifiable ($\mathcal{ID}$, for short) from distribution $P(v)$ compatible with a causal graph G, if for any two (fully specified) models M_1 and M_2 that satisfy the assumptions in G, we have*

$$P_1(v) = P_2(v) \Rightarrow Q(M_1) = Q(M_2). \tag{2.7}$$

Thus, for any two structural models M_1 and M_2, if the induced distributions $P_1(v)$ and $P_2(v)$ coincide, both models need to provide the same answers to query Q. If this is the case, identifiability entails that Q depends solely on $P(v)$ and G and can be expressed in terms of parameters of $P(v)$, *regardless* of the underlying

[10] $Y_x = y$ is interpreted as "Y would be equal to y, if X had been x" (Pearl, 2009, def. 7.1.5).

[11] This definition of identification is similar to the one used in Matzkin (2007).

mechanisms F and randomness $P(u)$.

Finally, once identification of the post-intervention distribution $P(y|do(x))$ for any value of x is achieved, the average causal effect (as well as any other quantity, such as risk ratios, odds ratios, quantile effects, etc.) can be computed as[12]

$$\mathbb{E}\left[Y|do(X=x_1)\right] - \mathbb{E}\left[Y|do(X=x_0)\right] = \sum_y y\left[P(y|do(x_1)) - P(y|do(x_0))\right]. \quad (2.8)$$

3. COUNFOUNDING BIAS

One of the biggest threats to the identification of causal effects, and the one which usually receives the greatest attention from methodologists, is confouding bias. The suspicion that a correlation might not reflect a genuine causal link between two variables, but is instead driven by a common cause, gives rise to the maxim *"correlation does not imply causation"* (List, 2011). In this section we present strategies for dealing with confounding bias in structural causal models. First, we present graphical criteria for finding appropriate covariate adjustment sets that eliminate confouding. We next show that the task of establishing identification in causal graphs can be fully automated with the help of a procedure called *do-calculus*. Eventually, we discuss identification strategies if covariate adjustment proves to be ineffective but a surrogate experiment, similar to an instrumental variable affecting an endogenous treatment, is available.

3.1. Covariate selection and the backdoor criterion

We consider the well-known example of estimating college wage premia from labor economics (Angrist and Pischke, 2009, ch. 3.2.3). Let the the structural causal model be represented by the graph G in Figure 2a. C is a dummy for college degree and the outcome of interest are earnings Y. W is a binary indicator denoting whether an individual has a "white-collar" or "blue-collar" job. W is causally affected by C since many white-collar jobs require a college degree. At the same time, the effect of W is partially mediated by an individual's work-related health H. This assumption captures the idea that blue-collar jobs might

[12] For ease of exposition, we assume random variables to be discrete throughout the text. Summations should be replaced by integrals if variables with continuous support are considered.

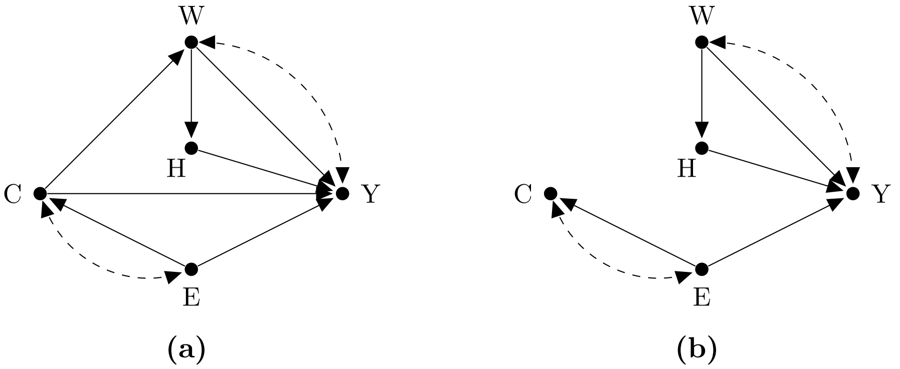


Figure 2: *(a) College wage premium example of Section 3.1. Variables: college degree (C), earnings (Y), occupation (W), work-related health (H), socio-economic factors (E). (b) Graph $G_{\underline{C}}$ for the previous example in which all arrows emitted by C are deleted.*

be associated with higher adverse health effects, which ultimately reduce life-time earnings. Finally, E represents a set of socio-economic variables that influence both an individual's college graduation probability as well as future earning potential. Dashed bidirected arrows depict unmeasured common causes that lead to dependence between the background characteristics U_i of the connected variables.

In order to estimate the causal effect of a college degree on earnings, the following graphical criterion can be used to find admissible adjustment sets that eliminate any confounding influence between C and Y.

Definition 3.1. *(Admissible sets – the backdoor criterion) Given an ordered pair of variables (X, Y) in a directed acyclic graph, a set Z is backdoor admissible if it blocks every path between X and Y in the graph $G_{\underline{X}}$.*

$G_{\underline{X}}$ in definition 3.1 denotes the graph that is obtained when all edges emitted by node X are deleted in G. Figure 2b depicts $G_{\underline{C}}$ for the college wage premium example. The intuition behind the backdoor criterion is simple. Unblocked paths between X and Y pointing into X (i.e., "entering through the backdoor") create an association between X and Y that is not due to any causal influence exerted by X. By conditioning on variables along these spurious paths, this association can be canceled such that only the causal path from X to Y remains.

$Z = \{E\}$ in Figure 2a satisfies the backdoor criterion and is thus an admissible adjustment set. W can be left unaccounted for because it does not lie on a backdoor path between X and Y. In fact, the graph illustrates why conditioning on occupation would create rather than reduce estimation bias in this case. Because W is a collider node on $C \rightarrow W \dashleftarrow\dashrightarrow Y$, it would open up the path and produce spurious correlation when conditioned on.

If a set of variables satisfies the backdoor criterion, the causal effect of X on Y can be identified from observational data by the adjustment formula (Pearl, 2009, Theorem 3.3.2)

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z). \tag{3.1}$$

Practically, estimation can be carried out by propensity score matching (Rosenbaum and Rubin, 1983; Heckman et al., 1998), inverse probability weighting (Robins, 1999), or linear regression. The similarity with the treatment effects literature at this point is no coincidence as the backdoor criterion implies ignorability (Pearl, 2009, Theorem 4.3.1).

Theorem 3.1. *(Counterfactual interpretation of backdoor) If a set Z of variables satisfies the backdoor condition relative to (X, Y), then for all x, the counterfactual Y_x is conditionally independent of X given Z*

$$Y_x \perp\!\!\!\perp X|Z \tag{3.2}$$

Other than in the potential outcome framework, which provides the analyst with no guidance to identify confounding causal paths, the search for appropriate adjustment sets via the backdoor criterion and d-separation can easily be automated (Textor and Liśkiewicz, 2011). This is particularly useful in larger graphs such as in Figure 3a. The set of all admissible adjustment sets for identifying $P(y|do(x))$ in Figure 3a is given by

$$\begin{aligned} Z = &\{\{W_2\}, \{W_2, W_3\}, \{W_2, W_4\}, \{W_3, W_4\}, \\ &\{W_2, W_3, W_4\}, \{W_2, W_5\}, \{W_2, W_3, W_5\}, \{W_4, W_5\}, \\ &\{W_2, W_4, W_5\}, \{W_3, W_4, W_5\}, \{W_2, W_3, W_4, W_5\}\}. \end{aligned} \tag{3.3}$$

This example illustrates that it is neither necessary nor sufficient to adjust for

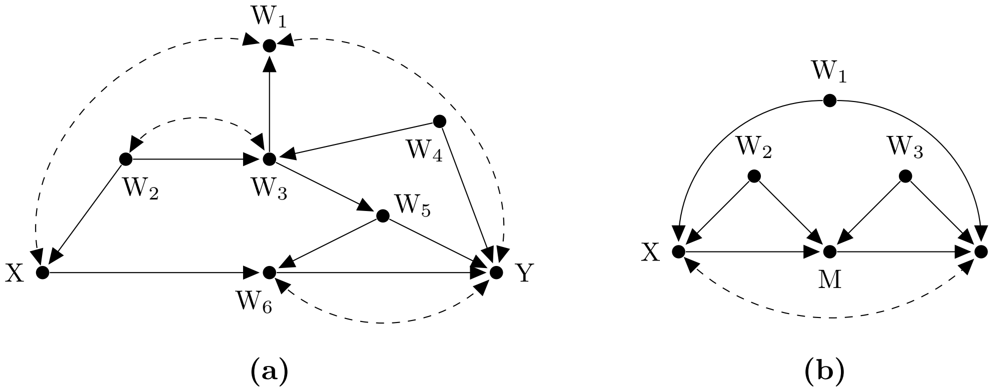


Figure 3: *(a) Application of the backdoor criterion in larger graphs. (b) The presence of M on the directed path from X to Y allows for identification via the front-door criterion.*

all variables in a model. Thus, the analyst could decide, for example, to save costs on collecting data for W_4 and estimate the effect of X by conditioning on $\{W_2, W_3\}$ instead. Likewise, it would be a serious mistake to condition on W_1 since conditioning on W_1 would introduce collider bias on the path $X \dashleftarrow\dashrightarrow W_1 \dashleftarrow\dashrightarrow Y$. These intricacies in finding appropriate adjustment sets cast serious doubt on a researcher's ability to judge the plausibility of the ignorability assumption without the help of a graph and simply based on introspection.

3.2. Front-door adjustment in the case of unmeasured confounders

Identification via backdoor adjustment requires that all backdoor paths can be closed by a set of observed variables. This is not always feasible. However, in situations where no set of observables is backdoor admissible, another identification strategy, which is less familiar to economists, might be applicable. Think of a situation as in Figure 3b, where adjusting for a set of observable variables W_1 is not sufficient to close all backdoor paths between X and Y. There are remaining unobserved confounders that are represented by the bidirected arc $X \dashleftarrow\dashrightarrow Y$. At the same time, the entire effect of X is assumed to be mediated by a third variable M. This assumption is plausible, for example, if a policy intervention in the educational system affects the job market prospects of high school graduates solely by raising test scores. If the data allows to adjust for confounders at the mediator, i.e., W_2 and W_3 in Figure 3b are observed, the causal effect of X on Y is identifiable with the help of the following criterion (Pearl, 1995, 2009).

Definition 3.2. *(The front-door criterion) A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if, after adjusting for a set of variables W*

1. *Z intercepts all directed paths from X to Y,*
2. *there is no unblocked path from X to Z,*
3. *and all backdoor paths from Z to Y are blocked by X.*

The causal effect in Figure 3b is then given by the following formula

$$P(y|do(x)) = \sum_m P(m|x, w_2, w_3)P(w_2)P(w_3) \sum_{x'} P(y|x', m, w_2, w_3)P(x'|w_2). \tag{3.4}$$

Front-door adjustment amounts to a sequential application of the backdoor criterion. First, the effect of X on M can be identified by adjusting for W_2. Second, the backdoor path $M \leftarrow X \dashleftarrow\dashrightarrow Y$, which remains open after adjusting for W_3, can be blocked by conditioning on X. The front-door adjustment formula then chains these individual causal effect estimates together to arrive at the overall effect of X on Y (Pearl, 2009, p. 82). Because the front-door criterion is applicable even in the presence of unobserved confounders between treatment and outcome, it is a good example of how causal graphs provide new sources of identification that go beyond the traditional requirement of ignorability in the treatment effects literature.[13]

3.3. Causal calculus and the algorithmization of identification

In directed acyclic graphs, identifiability of queries of the form $P(y|do(x))$ can be decided systematically by using an algebraic procedure called the do-calculus

[13] Glynn and Kashin (2017) present another interesting application of the front-door criterion for evaluating the effect of the National Job Training Partnership Act (JTPA; Heckman et al., 1997) program on earnings, where X measures the (self-selected) sign-up for the program and M whether an individual actually showed up for the training. They are able to relax the assumptions given in Definition 3.2 by complementing the front-door criterion with a difference-in-differences-type identification approach, which is able to deal with bias stemming from possibly unobserved confounders between M and outcome Y.

(Pearl, 1995). Do-calculus consists of three inference rules that permit to transform probabilistic sentences involving interventions and observations whenever certain separation conditions hold in a causal graph G. Let X, Y, Z, and W be arbitrary disjoint sets of nodes in G. The graph that is obtained by deleting from G all arrows pointing to nodes in X is denoted by $G_{\overline{X}}$. Similarly, $G_{\underline{X}}$ results from deleting in G all arrows that are emitted by X. Furthermore, the deletion of both arrows incoming in X and arrows outgoing from Z is denoted by $G_{\overline{X}\underline{Z}}$. Given this notation, the following three rules, which are valid for every interventional distribution compatible with G, can be formulated.

Rule 1. *(Insertion/deletion of observations)*

$$P(y|do(x), z, w) = P(y|do(x), w) \qquad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}. \tag{3.5}$$

Rule 2. *(Action/observation exchange)*

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \qquad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}. \tag{3.6}$$

Rule 3. *(Action/observation exchange)*

$$P(y|do(x), do(z), w) = P(y|do(x), w) \qquad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\overline{Z(W)}}}, \tag{3.7}$$

where $Z(W)$ is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

Identifiability of causal query Q can be decided by repeatedly applying the rules of do-calculus, until Q is transformed into a final expression that no longer contains a do-operator. This renders Q consistently estimable from nonexperimental data. In appendix A.2 we show an application of the do-calculus for the college wage premium example from the previous section.

Do-calculus was proved to be complete for queries of the form $Q = P(y|do(x), z)$ (i.e., all unconditional and covariate-specific interventions; Shpitser and Pearl, 2006; Huang and Valtorta, 2006), which implies that if the procedure fails to transform Q into an expression only containing observable quantities, the causal effect is guaranteed to be non-identifiable. In this case, point identification will only be achievable by imposing stronger functional-form assumptions, such as linearity, monotonicity, or additivity. Shpitser and Pearl (2006, building on earlier work

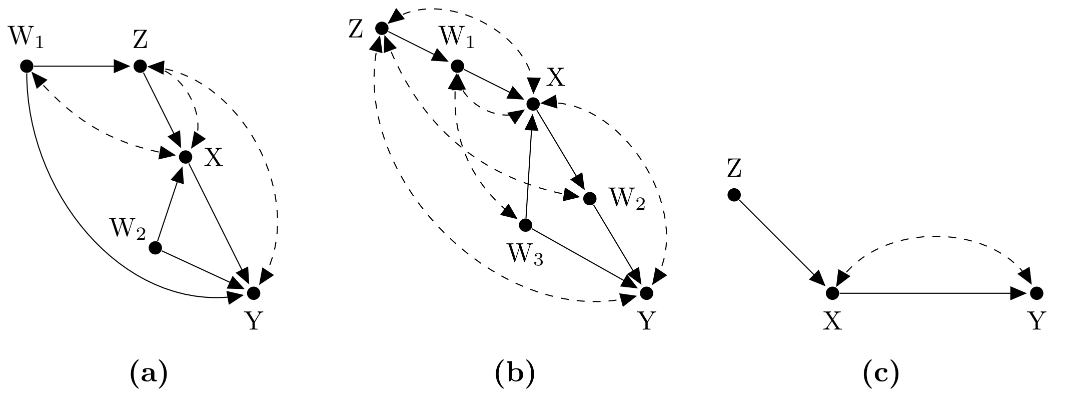


Figure 4: *(a) $P(y|do(x))$ is not identifiable with observational data but z-identifiable if experimental variation in Z is available. (b) Example of $zI\mathcal{D}$ in the presence of unobserved confounders between X and Y and Z affecting X only indirectly. (c) The standard instrumental variable setting.*

by Tian and Pearl, 2002) show that transforming a causal query into a do-free expression can be fully automated by an algorithm that inherits from do-calculus its completeness property of being able to return estimable expressions for all identifiable queries in a DAG. This automatization converts the often tedious task of identification into a straightforward exercise in graphical models.

3.4. Identification by surrogate experiments

In practical applications it is not uncommon that identification of causal queries based on observational data alone is unattainable. In these cases, a frequently applied strategy is to use experimental variation in other variables than the treatment under study in order to identify its causal effect. In development economics and economic policy this strategy is known under the name *"encouragment design"* (Duflo et al., 2008). One example are Duflo and Saez (2003) who study the effect of information on retirement planning decisions by conducting a randomized control trial in which financial incentives to attend an information session on tax deferred account (TDA) retirement plans are provided to university employees. Thus, their study design uses a surrogate experiment (providing financial incentives) to *encourage* exogenous variation in an otherwise endogenous treatment variable (information about TDA retirement plans).

To make this idea concrete, we turn to Figure 4a in which several backdoor paths passing through Z confound the relationship between X and Y. At the same time, adjusting for Z is not possible because it would open up the path $X \dashleftarrow\dashrightarrow Z \dashleftarrow\dashrightarrow Y$ on which Z is a collider. Consequently, $Q = P(y|do(x))$ is not identifiable with purely observational data. If, however, it is feasible to manipulate Z experimentally, it can be shown that Q is identified from the interventional distribution $P(v|do(z))$ instead. This leads to a straightforward extension of the identification problem that was formulated in Definition 2.3.

Definition 3.3. *(Causal effects z-identifiability) Let X, Y, Z be disjoint sets of variables, and let G be the causal diagram. The causal effect of an action $do(X = x)$ on a set of variables Y is said to be **z**-identifiable (**z**$\mathcal{ID}$, for short) from P in G, if $P(y|do(x))$ is (uniquely) computable from $P(V)$ together with the interventional distributions $P(V \setminus Z'|do(Z'))$, for all $Z' \subseteq Z$, in any model that induces G.*

Bareinboim and Pearl (2012a) prove that the problem of **z**$\mathcal{ID}$ can be solved by applying the rules of do-calculus to transform Q into an expression that only contains $do(Z)$.

Theorem 3.2. *Let X, Y, Z be disjoint sets of variables, and let G be the causal diagram, and $Q = P(y|do(x))$. Q is **z**$\mathcal{ID}$ from P in G if the expression $P(y|do(x))$ is reducible, using the rules of do-calculus, to an expression in which only elements of Z may appear as interventional variables.*

It can be further be shown that the do-calculus is complete also for the problem of **z**-identification (Bareinboim and Pearl, 2012a, Corrolary 3). The following theorem provides a complete characterization of the **z**$\mathcal{ID}$ class in terms of graphical conditions (Bareinboim and Pearl, 2012a, Theorem 3).

Theorem 3.3. *Let X, Y, Z be disjoint sets of variables and let G be the causal graph. The causal effect $Q = P(y|do(x))$ is **z**$\mathcal{ID}$ in G if and only if one of the following conditions hold:*

(i) Q is identifiable in G; or.

(ii) There exists $Z' \subseteq Z$ such that the following conditions hold,

a. X intercepts all directed paths from Z' to Y and *b. Q is identifiable in* $G_{\overline{Z'}}$.

Manipulation of Z leads to a post-interventional graph $G_{\overline{Z}}$ in which all incoming arrows into Z are deleted. In Figure 4a this breaks all backdoor paths except the one going through W_2. Thus, if W_2 is observable identification can be achieved. Bareinboim and Pearl (2012a) extend the work of Shpitser and Pearl (2006) on the standard $\boldsymbol{\mathcal{ID}}$ problem and develop a complete $\boldsymbol{z\mathcal{ID}}$ algorithm that returns an estimable expression for $P(y|do(x))$ using experimental variation in Z, if such an expression can be deduced.

$\mathcal{Z}$-identification bears close resemblance to instrumental variable estimation, but they are not equivalent. Take the canonical IV setting, following Angrist (1990), with an exogenous instrument and unobserved confounders between treatment and outcome, which is depicted in Figure 4c. In this graph, $P(y|do(x))$ is not $\boldsymbol{z\mathcal{ID}}$ since the bidirected arc between X and Y violates condition *ii*a in Theorem 3.3.[14]

The fact that $P(Y|do(X))$ remains unidentifiable in Figure 4c is not surprising, as it is a well-known result that point identification is not possible in the nonparametric IV setting (Manski, 1990; Balke and Pearl, 1995). Introducing additional functional form restrictions as a remedy, such as monotonicity or linearity, would only permit to identify a local average treatment effect for the latent subgroup of compliers (Imbens and Angrist, 1994). $\mathcal{Z}$-identification, by contrast, preserves the fully nonparametric nature of graphical causal models. Thus, if a casual effect is $\boldsymbol{z\mathcal{ID}}$ in G, the average treatment effect, instead of its local variety, can be computed from data. Moreover, z-identification is applicable in more complicated settings than the canonical IV; for example, if Z exerts and indirect effect on X such as in Figure 4b. Theorem 3.3 thus provides a complete characterization of the class of encouragement designs in which the ATE of a policy initiative is identifiable.

4. SAMPLE SELECTION BIAS

The previous section discussed strategies to control for confounding bias, which arises due to nonrandom assignment into treatment. In applied empirical work,

[14] Notwithstanding the restriction of condition *ii*a in Theorem 3.3, z-identification is possible if there are unobserved confounders between treatment and outcome. Figure 4b presents such an example, where identification in $G_{\overline{Z}}$ can be achieved via the front-door criterion.

however, researchers often encounter another source of bias that stems from preferential selection of units into the data pool. Such sample selection poses a serious threat to both statdistical as well as causal inference, because it jeopardizes the representativeness of the data for the underlying population. A seminal discussion of this problem in an economic context is by Heckman (1976, 1979). He estimates a model of female labor supply in a sample of 2,253 working women interviewed in 1967. The challenge to valid inference in this setting arises due to the fact that market wages are only observable for women who actually choose to work. His model is described by the following two equations

$$s_i = \mathbb{1}[Z_i'\delta - \eta_i > 0] \tag{4.1}$$

$$y_i = \begin{cases} x_i\beta + Z_i'\gamma + \varepsilon_i & \text{if } s_i = 1, \\ \text{unobserved} & \text{if } s_i = 0. \end{cases} \tag{4.2}$$

The first equation characterizes the sampling mechanism. Wages y_i are unobserved if $(Z_i'\delta - \eta_i)$ stays below a threshold of zero. The economic interpretation of this equation is that individuals will choose to remain unemployed, if the income they are able to earn on the market does not exceed their reservation levels. Selection bias may then be the result of a correlation between reservation wages and unobservables in the market wage equation, i.e., $Corr(\eta_i, \varepsilon_i) \neq 0$. Similar settings of sample selection are widespread in economics. Examples are discussed, e.g., by Levitt and Porter (2000), who estimate the effectiveness seat belts and airbags in a sample of fatal crashes, and by Ihlanfeldt and Martinez-Vazquez (1986), who note the difficulty of assessing the determinants of house prices when using data on recently sold homes.

In causal diagrams, such cases can be represented by a special selection variable S that takes on two values – one if a unit is part of sample and zero otherwise. The selection mechanism is thereby specified by the parent nodes of S, which identify the causes of why certain units appear more frequently in the sample than others.[15] Figure 5a depicts such an augmented graph G_S for the female labor supply example. Here, sample selection is driven by socio-economic factors Z.

[15] Sample selection nodes are only allowed to have incoming arrows. They do not emit arrows by themselves.

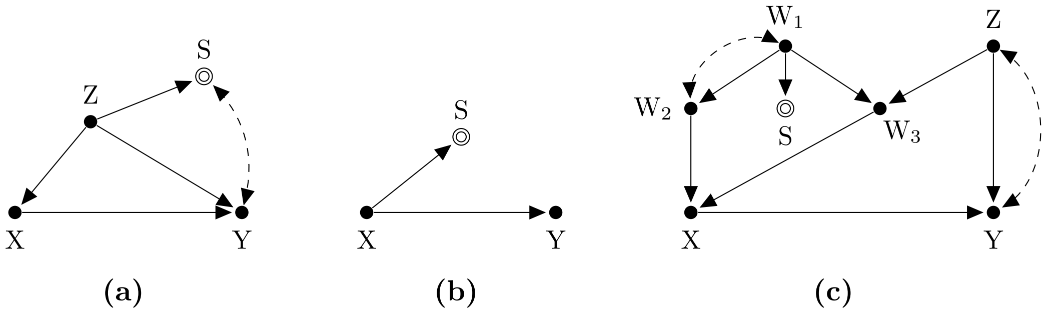


Figure 5: *(a) A model of female labor supply (Heckman, 1976, 1979). Variables: hours worked (X), earnings (Y), socio-economic factors (Z), sampling mechanism (S). (b) $P(y|do(x))$ is recoverable from selected data as $P(y|x, S = 1)$. (c) $\{W_1, W_3\}$, $\{W_2, W_3\}$ and $\{Z\}$ are all backdoor admissible, but the causal effect is only recoverable with $\{Z\}$.*

Moreover, the source of error correlation in the Heckman model is made explicit by the unobserved common cause of S and Y (denoted by the bidirected dashed arc).

Simultaneously controlling for confounding and selection bias introduces a new challenge in the do-calculus. Not only do we need to transform interventional distributions into do-free expressions, but the probabilities that make up these expressions need to be conditional on $S = 1$, because that is all we are able to observe. This additional restriction explains why dealing with selection bias is such a hard problem in practice. At the same time, the literature on recovering causal effects from selection-biased data (Bareinboim and Pearl, 2012b; Bareinboim et al., 2014; Bareinboim and Tian, 2015) aims at preserving the fully nonparametric nature of causal graphs. Consequently, the proposed approaches refrain from making any functional form assumptions related to the selection-propensity score $P(s_i|pa_i)$, such as monotonicity or (joint) normality, which are ubiquitous in the econometrics literature (Angrist, 1997). Nevertheless, even with this limited set of assumptions as a starting point, several positive results for the recoverability of causal effect estimates can be derived, which we will present in the following.

Bareinboim et al. (2014) provide a complete condition for recovering conditional probabilities from selected samples.

Theorem 4.1. *The conditional distribution $P(y|t)$ is recoverable from G_S (as*

$P(y|t, S = 1)$*) if and only if* $(Y \perp\!\!\!\perp S|T)$.

Sufficiency of this condition follows immediately. However, its necessity is less obvious and implies that if Y is not d-separated from S in G_S, its conditional distribution is not recoverable. Combining Theorem 4.1 with do-calculus suggests a straightforward strategy for recovering causal effects from biased data (Bareinboim and Tian, 2015).

Corollary 4.1. *The causal effect* $Q = P(y|do(x))$ *is recoverable from selection-biased data if using the rules of the do-calculus,* Q *is reducible to an expression in which no do-operator appears, and recoverability is determined by Theorem 4.1.*

Take Figure 5b as an example. Here, the relationship between X and Y is unconfounded and therefore $P(y|do(x)) = P(y|x)$ (by the second rule of do-calculus). Moreover, since S and Y are d-separated by X, we find the causal effect to be recoverable as $P(y|x, S = 1)$.

As an immediate consequence of Theorem 4.1, causal effects will not be recoverable if Y is directly connected to S (or shares a common parent with S). Thus, without invoking stronger functional form assumptions there is no possibility to control for selection bias in the female labor supply model of Figure 5a. In general, selection-biased data restrict the possibilities for identification in observational studies. For example, in Figure 5c there are several (minimal sufficient) backdoor admissible adjustment sets: $\{W_1, W_3\}$, $\{W_2, W_3\}$ and $\{Z\}$. However, recoverability can only be achieved with $\{Z\}$. That is because in the adjustment formula of equation (3.1) also the prior distribution needs to be recovered and $\{Z\}$ is the only conditioning set that is d-separated from S. Thus, the only feasible backdoor adjustment in this case, which is estimable from biased data, is

$$P(y|do(x)) = \sum_z P(y|x, z, S = 1)P(z|S = 1). \tag{4.3}$$

It is important to note that although Theorem 4.1 provides a necessary condition for recovering conditional probabilities, the same does not hold for Corollary 4.1 with respect to causal effects. This is exemplified by the graph in Figure 6a. Due to unobserved confounders between X and Y and the fact that Z is a collider,

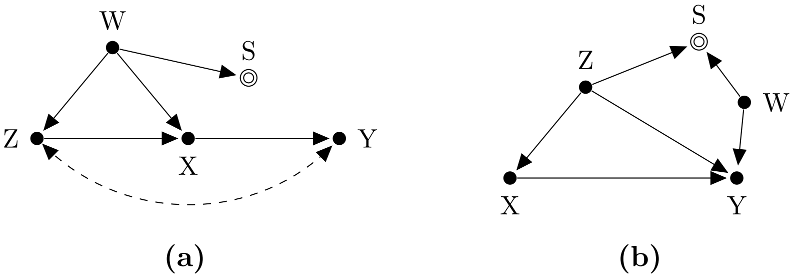


Figure 6: *(a) $P(y|do(x))$ is not recoverable from selection bias following the procedure described in Corollary 4.1. Nevertheless recovery can be achieved by applying the algorithm in Bareinboim and Tian (2015). (b) Adaption of the sample selection model in Figure 5a where the set $\{Z, W\}$ is s-backdoor admissible.*

identification via the backdoor criterion would require to adjust for both Z and W in order to close all backdoor paths. But $P(z, w)$ is not recoverable and, thus, an attempt to apply Corollary 4.1 will fail. Nevertheless, $P(y|do(x))$ in Figure 6a can be recovered with the help of do-calculus. To see this, first note that $(S, W \perp\!\!\!\perp Y)$ in $G_{\overline{X}}$ (the resulting graph when all incoming arrows in X are deleted, see Section 3.3). Then, according to the first rule of do-calculus

$$P(y|do(x)) = P(y|do(x), w, S = 1) \tag{4.4}$$

$$= \sum_z P(y|do(x), z, w, S = 1)P(z|do(x), w, S = 1), \tag{4.5}$$

where the second line follows from conditioning on Z. Applying rule 2, since $(Y \perp\!\!\!\perp X|W, Z)$ in $G_{\underline{X}}$, we can eliminate the do-operator in the first term

$$= \sum_z P(y|x, z, w, S = 1)P(z|do(x), w, S = 1). \tag{4.6}$$

Finally, because $(Z \perp\!\!\!\perp X|W)$ in $G_{\overline{X(W)}}$, it follows from rule 3 that

$$= \sum_z P(y|x, z, w, S = 1)P(z|w, S = 1). \tag{4.7}$$

Bareinboim and Tian (2015) provide algorithmic conditions for the recoverability of interventional distributions in arbitrary graphs. Based on these results, derivations such as the one just performed can be automated, which releases analysts from the tedious task of applying do-calculus on a case-by-case basis.[16]

4.1. Combining biased and unbiased data

Another promising strategy for recovering causal effects from selection-biased data can be pursued if several data sources are combined. The distributions of socio-economic factors such as age, sex, and education, for example, can usually be estimated without bias at the population level (e.g., from census data). To illustrate how this helps for recoverability, we revisit the female labor supply example from above, but now assume that the common parent of wage and the selection node is observable as W (see Figure 6b). Then, conditioning on the set $\{Z, W\}$ simultaneously closes all backdoor paths between X and Y and d-separates Y from S. From the adjustment formula we can thus derive

$$\begin{aligned} P(y|do(x)) &= \sum_{z,w} P(y|x,z,w)P(z,w) && (4.8) \\ &= \sum_{z,w} P(y|x,z,w,S=1)P(z,w), && (4.9) \end{aligned}$$

where the second line follows from $(Y \perp\!\!\!\perp S|Z,W)$. As $P(z,w)$ cannot be recovered, Corollary 4.1 is not applicable. However, if in addition to the selected data, unbiased measurements of $P(z,w)$ are available from secondary data source, equation (4.9) becomes estimable.

Bareinboim et al. (2014) generalize this idea and present the following extension to the backdoor criterion, which can be invoked if a subset Z of the data is measured without bias.

Definition 4.1. *(Selection backdoor criterion) Let a set Z of variables be partitioned into $Z^+ \cup Z^-$ such that Z^+ contains all non-descendants of X and Z^- the descendants of X, and let G_S stand for the graph that includes sampling mechanism S. Z is said to satisfy the selection backdoor criterion (s-backdoor, for short) if it satisfies the following conditions:*

[16] Note, however, that in contrast to the identification algorithms described in Section 3, this algorithm is not complete. To this date, no necessary and sufficient conditions for automatically deciding about the recoverability of causal effects when only selection-biased data is available are known.

(i) *Z^+ blocks all backdoor paths from X to Y in G_S;*

(ii) *X and Z^+ block all paths between Z^- and Y in G_S, namely, $(Z^- \perp\!\!\!\perp Y|X, Z^+)$;*

(iii) *X and Z block all paths between S and Y in G_S, namely, $(Y \perp\!\!\!\perp S|X, Z)$; and*

(iv) *Z and $Z \cup \{X, Y\}$ are measured in the unbiased and biased studies, respectively.*

Theorem 4.2. *If Z is s-backdoor admissible, then causal effects are identified by*

$$P(y|do(x)) = \sum_z P(y|x, z, S = 1)P(z) \tag{4.10}$$

The s-backdoor criterion is a sufficient condition for generalized adjustment, which deals with both confounding and selection bias simultaneously, when unbiased measurements of covariates are available. Conditions that are both necessary and sufficient are presented by Correa et al. (2017), who also develop a complete algorithm that efficiently finds all sets of variables admissible for generalized adjustment.

5. TRANSPORTABILITY OF CAUSAL KNOWLEDGE

In this section we consider the task of extrapolating experimental knowledge across domains (e.g., settings, populations, environments) that differ in their causal characteristics. This problem is called "transportability" in the computer science literature but is often – in a broader context – referred to as "external validity" in the social sciences (Pearl and Bareinboim, 2014).[17] It is fundamental in scientific discovery because experimental knowledge is ultimately intended to be used in other environments than it was obtained in, where conditions are likely to be different. Duflo et al. (2008) allude to this fact when asking: *"If a [development] program worked for poor rural women in Africa, will it work for middle-income*

[17] In econometrics the term "external validity" is sometimes used in the narrower sense of extrapolating local average treatment effect estimates to the group of always- and never-takers within the same empirical domain (Kowalski, 2018). The remainder of this section deals with the more challenging topic of transporting causal effects across causally heterogeneous domains.

urban men in South Asia?". Nakamura and Steinsson (2018) discuss external validity in a macroeconomic context and come to the conclusion that *"even very cleanly identified monetary and fiscal natural experiments give us, at best, only a partial assessment of how future monetary and fiscal policy actions—which may differ in important ways from those in the past—will affect the economy."*

It is often implicitly assumed that estimated effects of an intervention in one population Π are similar – or at least provide a good approximation – to what would be found in another population Π^*. This would render it possible to use results from Π for policy decisions in Π^*. However, such *direct transportability* (Pearl and Bareinboim, 2011) and is likely to be violated in many empirical settings.

Definition 5.1. *(Direct Transportability) A causal relation R is said to be directly transportable from Π to Π^*, if $R(\Pi^*) = R(\Pi)$.*

For an example, consider Banerjee et al. (2007) who study the effects of a remedial education program in two major cities in Western India, Mumbai and Vadodara. The randomized intervention provided schools with an extra teacher to work with children in the third and fourth grades who had been lagging behind their peers. The program showed substantial positive effects on children's academic achievements, at least in the short-run. Interestingly, however, while treatment effects on math scores were similar in both cities, the effect on language proficiency was weaker in Mumbai compared to Vadodara. The authors explain this finding with basic reading skills that are higher in Mumbai, where students are from more wealthy families and schools are better equipped. In math, by contrast, baseline achievement levels were more similar, and thus the remedial education program, which targets the most basic competencies in the curriculum, was equally effective.

Structural causal models offer the possibility formally address these kinds of external validity questions in a rigorous way. The graph in Figure 7a provides a stylized representation of the Banerjee et al. (2007) setting. Assume that we have experimental knowledge about the causal effect of an educational program on students' test scores from a randomized control trial conducted in Vadodara. We would like to generalize these results to to the population in Mumbai but are aware of the fact that income levels of families, Z, which are an important causal factor influencing children's academic achievements, are higher in Mumbai. In an SCM

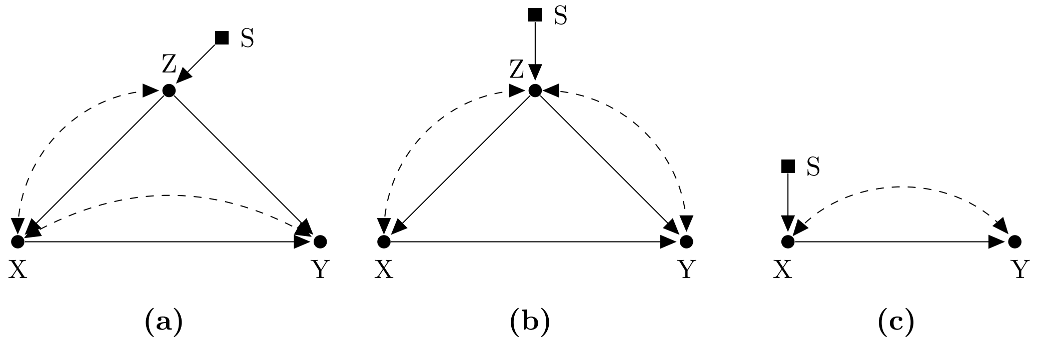


Figure 7: *(a) Z d-separates S and Y in $D_{\overline{X}}$. The causal effect of X on Y is thus transportable. (b) The unobserved confounder between Z and Y prevents transportability. (c) The causal effect $P(y|do(x))$ is directly transportable.*

we can incorporate this knowledge by adding a set of *selection variables S* that indicate where both populations under study differ, either due to the distribution of background factors $P(U_i)$ or varying causal mechanisms f_i. These S variables thus locate the source of structural discrepancies that threaten transportability and switching between two populations Π and Π^* is denoted by conditioning on different values of S.[18]

Definition 5.2. *(Selection Diagram) Let $\langle M, M^* \rangle$ be a pair of structural causal models (Definition 2.1) relative to domains $\langle \Pi, \Pi^* \rangle$, sharing a causal diagram G. $\langle M, M^* \rangle$ is said to induce a selection diagram D if D is constructed as follows:*

(i) Every edge in G is also and edge in D.

(ii) D contains an extra edge $S_i \rightarrow V_i$ whenever there might exist a discrepancy $f_i \neq f_i^*$ or $P(U_i) \neq P^*(U_i)$ between M and M^*.

Alternatively, the absence of a selection node represents the assumption that the causal mechanism that assigns values to these variables is the same in both populations. In the extreme case, one could add S nodes to all variables in the graph to express the belief that the populations are maximally heterogeneous.

[18] For clarity, S variables related to transportability are depicted by squares in the graph (■) to distinguish them from the selection bias case. Note also that S is pointing into other variables in the transportability setting, whereas selection nodes that indicate preferential inclusion into the sample only receive incoming arrows.

Naturally, this would undermine any hope for information exchange across the two domains, however.

Equipped with the definition of a selection diagram we can now state the following theorem (Pearl and Bareinboim, 2011).

Theorem 5.1. *Let D be the selection diagram characterizing two populations, Π and Π^*, and S the set of selection variables in D. The strata-specific causal effect $P^*(y|do(x), z)$ is transportable from Π to Π^* if Z d-separates Y from S in the X-manipulated version of D, that is, Z satisfies $(Y \perp\!\!\!\perp S|Z, X)_{D_{\overline{X}}}$.*

If it is possible to d-separate the set of factors S that cause disparities between the two domains from the outcome variable, tansportability can can be achieved. The intuition behind this approach is very similar to the slection-bias case discussed in Section 4 (Theorem 4.1) where the selection indicator likewise needs to be d-separated from Y by a set T to recover the conditional distribution $P(y|t)$ (Pearl, 2015a). This similarity is reflected in the following definition.

Definition 5.3. *(S-admissibility) A set T of variables satisfying $(Y \perp\!\!\!\perp S|T, X)$ in $D_{\overline{X}}$ will be called s-admissible (with respect to the causal effect of X on Y).*

Looking at the selection diagram in Figure 7a, we note that the set Z satisfies s-admissibility as it d-separates S and Y when X is experimentally manipulated. We now show that this implies that the causal effect is transportable across the two domains characterized by S.

$$\begin{aligned} P^*(y|do(x)) &= P(y|do(x), s) && (5.1) \\ &= \sum_z P(y|do(x), z, s)P(z|do(x), s) && (5.2) \\ &= \sum_z P(y|do(x), z, s)P(z|s) && (5.3) \\ &= \sum_z P(y|do(x), z)P^*(z). && (5.4) \end{aligned}$$

The first equation follows from the definition that distributions in the target domain Π^* are denoted by conditioning on S. The last line is then derived by using the s-admissibility of Z and the third rule of do-calculus. As long as Figure 7a provides an accurate model for the setting in Banerjee et al. (2007), the causal effect of the remedial education program in Mumbai can thus be computed by reweighting the stratum-specific causal effect (for every income level of Z) obtained in Vadodara by the income distribution $P^*(z)$ in Mumbai. This result is stated in full generality in the following corollary.

Corollary 5.1. *The causal effect $P^*(y|do(x))$ is transportable from Π to Π^* if there exists a set Z of observed pretreatment covariates that is S-admissible. Moreover, the transport formula is given by the weighting*

$$P^*(y|do(x)) = \sum_z P(y|do(x), z)P^*(z) \tag{5.5}$$

As a graphical criterion for transportability, s-admissibility can be easily checked in SCMs. Figure 7b provides a cautionary tale in that regard. Apart from the unobserved confounder between Z and Y, it is identical to 7a. However, s-admissibility is violated because conditioning on Z would open up the path $S \rightarrow Z \dashleftarrow\dashrightarrow Y$. Without the rigorous description of the structural assumptions imposed on the DGP that causal graphs provide, checking the sometimes intricate conditions for successful transportability of causal effect estimates across populations appears to be an extremely difficult undertaking.

It is an immediate consequence of Theorem 5.1 that any S variable that points into X can be ignored. The causal effect $P(y|do(x))$ is thus directly transportable in Figure 7c. The same holds for S nodes that are d-separated by the empty set in $D_{\overline{X}}$.

The transport formula presented in equation (5.5) is well known in the econometrics literature (Hotz et al., 2005; Dehejia et al., 2015; Andrews and Oster, 2018).[19] However, transportability in causal graphs, based on the rules of do-calculus, are applicable in a much broader range of settings (Pearl and Bareinboim, 2011; Bareinboim and Pearl, 2012c).

[19] Because these studies rely on the potential outcome framework, they express s-admissibility in terms of ignorability relations; i.e., domain heterogeneity S is assumed to be ignorable given pretreatment covariates X. Note, however, how easily this assumption can fail, for example, when adding an unobserved confounder between X and Y to the model as in Figure 7b. Causal graphs thus provide valuable guidance for judging the validity of s-admissibility, which is not offered by the potential outcome framework.

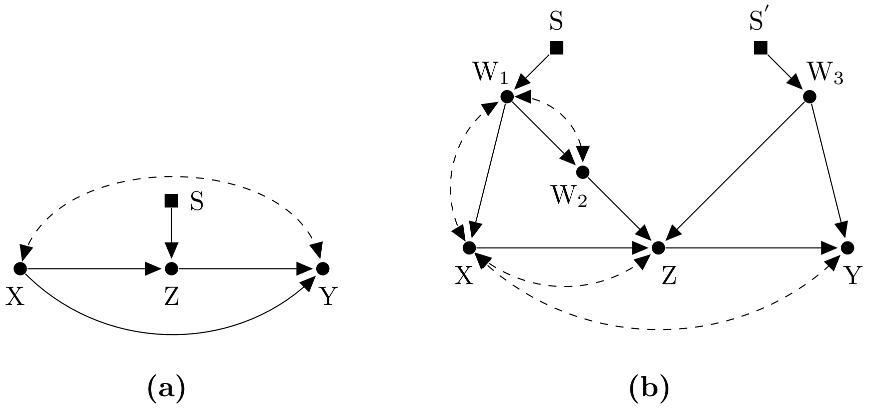


Figure 8: *(a) $P(y|do(x))$ is transportable even though S points into a post-treatment variable. (b) A more complex graph in which transportability can be decided algorithmically by the criteria developed in Bareinboim and Pearl (2013b).*

Theorem 5.2. *Let D be the selection diagram characterizing two populations, Π and Π^*, and S as set of selection variables in D. The relation $R = P^*(y|do(x), z)$ is transportable from Π to Π^* if the expression $P(y|do(x), z, s)$ is reducible, using the rules of do-calculus, to an expression in which S appears only as a conditioning variable in do-free terms.*

One such class of models is given when domains differ due to variables that are themselves causally affected by the treatment, as in Figure 8a. Here, the effect of X on Y is partly transmitted by Z and domains differ either according to the distribution of background factors U_Z or the mechanism f_Z that determines Z. Such a situation arises, for example, if the effect of a development program depends on the level of care that is taken during implementation. Duflo et al. (2008) acknowledge that pilot programs are often implemented particularly carefully using high-quality program officials, which is difficult to replicate on a wider scale and thus threatens the generalizability of results.

Gordon et al. (2018) provide another example from a completely different context.The effectiveness of advertising campaigns on social media platforms depends on how frequently clients are exposed to the ads. Since exposure is determined by user behavior it is difficult to control for the advertiser. If a social media company running advertising experiments wants to transport results obtained on a desktop version of the platform to users with mobile devices, it will thus need to take into account that user demographics differ across both domains, which is likely to lead to differences in exposure.

In Figure 8a, and similar settings in which post-treatment variables are s-admissible, causal effects are transportable as (Pearl and Bareinboim, 2014)

$$\begin{aligned}
P^*(y|do(x)) &= P(y|do(x), s) && (5.6)\\
&= \sum_z P(y|do(x), z, s)P(z|do(x), s) && (5.7)\\
&= \sum_z P(y|do(x), z)P^*(z|do(x)), && (5.8)
\end{aligned}$$

where the last line follows from s-admissibility. Transportability of $P^*(y|do(x))$ then depends on transforming $P^*(z|do(x))$ into a do-free expression, since by definition no manipulation can be carried out in the target domain. Noting that X and Z are unconfounded in Figure 8a, this is achieved by $P^*(z|do(x)) = P^*(z|x)$.[20]

The resulting transport formula thus different from the simple expression in equation (5.5) and instead prescribes to reweight the z-specific effects by the conditional distribution $P^*(z|x)$, estimated in the target population

$$P^*(y|do(x)) = \sum_z P(y|do(x), z, s)P^*(z|x). \qquad (5.9)$$

Theorem 5.2 was proven to be a necessary and sufficient criterion for transporting causal effect estimates across domains by Bareinboim and Pearl (2012c). Although it is procedural, however, the theorem does not specify the sequence of do-calculus steps that need to be taken in order to arrive at the desired expression. Also it provides no guidance for deciding whether transportability is possible in the first place. In order to fill this gap, Bareinboim and Pearl (2013b) develop a complete algorithmic solution for carrying out these tasks, based on the earlier work on identification algorithms by Tian and Pearl (2002) and Shpitser and Pearl (2006).

The benefits of solving the transportability problem algorithmically become particularly apparent for more complex graphs such as in Figure 8b. Here, the correct

[20] If causal effects are ordinarily identified in the target domain, they are called *trivially transportable* (Pearl and Bareinboim, 2011).

transport formula is found to be

$$P^*(y|do(x)) = \sum_{z,w2,w3} P(y|do(x), z, w2, w3)P(z|do(x), w2, w3)P^*(w2, w3) \quad (5.10)$$

Note furthermore that this expression does not contain W_1. It thus helps to economize on data collection efforts by allowing the analyst to decide which measurements need to be taken in the target and source domain in order to transport causal effects.

5.1. Transportability with surrogate experiments

Bareinboim and Pearl (2013a) combine the idea of transportability with the previously introduced concept of z-identification, to what is called *z-transportability.* With this extension it becomes possible to not only transfer causal knowledge obtained from randomized control trials, but also from the encouragement designs discussed in Section 3.4, which rely on surrogate experiments. Researchers are thus given a high degree of flexibility to learn from knowledge across domains even in cases when direct manipulation of a variable of interest would be prohibitively costly, both in the target and source domain.

It is important to note that z-transportability is a distinct problem and reduces neither to ordinary transportability nor to z-identification. Bareinboim and Pearl (2013a) demonstrate this fact by presenting examples of causal queries which are $z\mathcal{ID}$ in the source domain Π, but that may or may not be z-transportable to the target domain Π^*. Still, analogous to Theorem 5.2, the rules of do-calculus can be used to transfer causal knowledge from surrogate experiments in the following way (Bareinboim and Pearl, 2013a).

Theorem 5.3. *Let D be the selection diagram characterizing two populations, Π and Π^*, and S as set of selection variables in D. The relation $R = P^*(y|do(x), z)$ is z-transportable from Π to Π^* in D if the expression $P(y|do(x), z, s)$ is reducible, using the rules of do-calculus, to an expression in which all do-operators apply to subsets of Z, and the S-variables are separated from these do-operators.*

However, Theorem 5.3 provides no indication on the sequence of do-calculus steps that need to be taken to establish z-transportability. To this end, Bareinboim

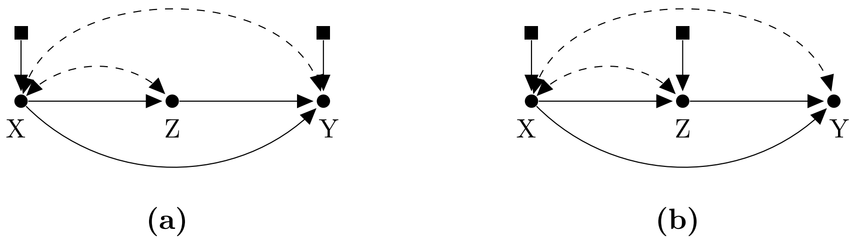


Figure 9: *Selection diagrams relative to two heterogeneous source domains π_a and π_b. Square nodes indicate discrepancies between the source and target domains. Meta-transportability entails to combine causal knowledge from both π_a and π_b to arrive at an estimate for $P^*(y|do(x))$ in the target domain.*

and Pearl (2013a) develop a complete algorithm, which takes the selection diagram D and information on the variable that has been intervened on in the source domain as inputs and returns a transport formula expression whenever such an expression exists.

5.2. Combining causal knowledge from several heterogeneous source domains

Transportability techniques are particularly valuable in situations in which it is possible to combine empirical knowledge from several source domains. Dehejia et al. (2015) discuss the case of a policy maker who is faced with the decision to either learn about a desired treatment effect from extrapolation of an existing experimental evidence base, or to commission a costly new experiment. The challenge in this situation is that previous experiments have possibly been conducted in very different contexts than the one of interest and underlying populations might be quite heterogeneous. It thus remains unclear how accurate such an extrapolation of prior results will be. Based on the approaches presented in the previous sections, Bareinboim and Pearl (2013c) introduce the concept of *meta*-transportability (or μ-transportability, for short), which provides a principled solution to this problem.

Let $\mathcal{D} = \{D_1, \dots, D_n\}$ be a collection of selection diagrams relative to source domains $\Pi = \{\pi_1, \dots, \pi_n\}$. An example is given in Figure 9. Panel (a) depicts the selection diagram that corresponds to source domain π_a, while panel (b) refers to π_b. Square nodes indicate where discrepancies between the target domain π^* and the source domains arise.[21] In accordance with Definition 5.2, these discrepancies can occur due to differences in causal mechanisms as well as background factors related to the the variables a square node is pointing into.

Figure 9 is a simple extension of a graph that was presented earlier (see Figure 8a). In contrast to the previous example, the unobserved confounder between X and Z (denoted by the dashed bidirected arc $X \dashleftarrow\dashrightarrow Z$) that was added to the diagram renders individual transportability impossible.[22] Interestingly, however, by combining information from both source domains, μ-transportability is feasible. The post-intervention distribution in the target domain π^* can written as

$$P^*(y|do(x)) = \sum_z P^*(y|do(x), z)P^*(z|do(x)) \tag{5.11}$$

$$= \sum_z P^*(y|do(x), do(z))P^*(z|do(x)) \tag{5.12}$$

where the second line follows from rule 2 of do-calculus, given that $(Z \perp\!\!\!\perp Y|X)_{D_{\overline{X}\underline{Z}}}$. Using this representation, it can be seen that each component is individually transportable from one of the source domains. First, note that the empty set is s-admissible in D_a, $(S \perp\!\!\!\perp Z|X)_{D^{(a)}_{\overline{X}}}$, and thus $P^*(z|do(x))$ is directly transportable from π^a. Second, $P^*(y|do(x), do(z))$ is directly transportable from π^b, since $(S \perp\!\!\!\perp Y|X, Z)_{D^{(b)}_{\overline{X,Z}}}$. The individual components of equation (5.12) can therefore be written as $P^*(z|do(x)) = P^{(a)}(z|do(x))$ and $P^*(y|do(x), do(z)) = P^{(b)}(y|do(x), do(z))$. This leads to the final transport formula

$$P^*(y|do(x)) = \sum_z P^{(a)}(z|do(x))P^{(b)}(y|do(x), do(z)). \tag{5.13}$$

Note that multiple pairwise transportability is not a necessary condition for μ-transportability to hold.[23] Bareinboim and Pearl (2013c) develop a complete

[21] The causal diagram for the target domain is accordingly obtained by deleting all square nodes from the selection diagrams.

[22] The complete transportability algorithm by Bareinboim and Pearl (2013b) would exit without returning a transport formula expression for both selection diagrams. Intuitively, looking at equation (5.8), the unobserved confounder prevents transporting $P^*(z|do(x)) = P^*(z|x)$ in (a). In (b), transportability is prohibited by the selection node pointing directly into Y.

[23] Meta-transportability is related to the ideas concerning "data combination" presented in Rid-

algorithmic solution for deciding about μ-transportability. The approach is further extended by Bareinboim et al. (2013) who combine μ-transportability with z-transportability, to allow for combining causal knowledge from multiple heterogeneous sources when the possibility to conduct experiments is limited to a subset Z_i of the variables in $\mathcal{D}$. They call this task *mz-transportability* and propose an algorithm for its automatization, which is proved to be complete by Bareinboim and Pearl (2014).

Combining experimental knowledge from several independent studies receives more an more attention in economics. Examples for recent meta-analytic studies can be found in Card et al. (2010) and Dehejia et al. (2015). However, existing approaches do not take domain heterogeneity in terms of causal mechanisms and background factors into account. Instead, they attempt to "average out" differences across populations. Data fusion techniques, based on graphical representations of structural causal models, by contrast, make it transparent how discrepancies in effect estimates across studies come about. They discipline the analyst to think carefully about the type of knowledge can actually be shared between domains. These tools therefore enable the research community to devise an effective strategy for leveraging the evidence base related to a specific context that exists at any given point in time. Causal knowledge obtained by individual experiments does not need to be regarded in isolation but rather as contributing to a shared body of knowledge, which can be recombined to tackle entirely new policy problems that were unimagined at the time when the original study was conducted. Combined with undergoing efforts in economics to make data sets of published papers openly available[24], data fusion techniques thus bear the potential to save on discipline-wide data collection costs and to render causal inference a truly collective endeavor.

REFERENCES

Andrews, I. and E. Oster (2018): "Weighting for External Validity," NBER Working Paper No. 23826.

der and Moffitt (2007). In this case, however, the goal is to combine causal knowledge from several heterogeneous populations that share at least some causal mechanisms.

[24] See `https://www.econometricsociety.org/publications/econometrica/information-authors/instructions-submitting-articles#replication`

ANGRIST, J. D. (1990): "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *The American Economic Review*, 80, 313–336.

——— (1997): "Conditional independence in sample selection models," *Economics Letters*, 54, 103–112.

ANGRIST, J. D. AND J.-S. PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

BALKE, A. AND J. PEARL (1995): "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 92, 1171–1176.

BANERJEE, A. V., S. COLE, E. DUFLO, AND L. LINDEN (2007): "Remedying Education: Evidence from Two Randomized Experiments in India," *The Quartely Journal of Economics*, 122, 1235–1264.

BAREINBOIM, E., S. LEE, V. HONAVAR, AND J. PEARL (2013): "Transportabilityfrommultipleenvironmentswithlimited experiments," in *Advances in Neural Information Processing Systems*, ed. by C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, vol. 26, 136–144.

BAREINBOIM, E. AND J. PEARL (2012a): "Causal Inference by Surrogate Experiments: z-identifiability," in *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, 113–120.

——— (2012b): "Controlling Selection Bias in Causal Inference," in *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, 100–108.

——— (2012c): "Transportability of Causal Effects: Completeness Results," in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence.*

——— (2013a): "Causal Transportability with Limited Experiments," in *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, 95–101.

——— (2013b): "A general algorithm for deciding transportability of experimental results," *Journal of Causal Inference*, 1, 107–134.

——— (2013c): "Meta-Transportability of Causal Effects: A Formal Approach," in *Proceedings of the 16th International Con- ference on Artificial Intelligence and Statistics (AISTATS)*, Scottsdale, AZ, vol. 31.

——— (2014): "Transportability from Multiple Environments with Limited Experiments: Completeness Results," in *Advances of Neural Information Processing Systems*, ed. by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, vol. 27, 280–288.

Bareinboim, E. and J. Tian (2015): "Recovering Causal Effects from Selection Bias," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, ed. by S. Koenig and B. Bonet, Association for the Advancement of Artificial Intelligence, Palo Alto, CA: AAAI Press.

Bareinboim, E., J. Tian, and J. Pearl (2014): "Recovering from Selection Bias in Causal and Statistical Inference," in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence.*

Basman, R. L. (1963): "The Causal Interpretation of Non-Triangular Systems of Economic Relations," *Econometrica*, 31, 439–448.

Bentzel, R. and B. Hansen (1954): "On Recursiveness and Interdependency in Economic Modeks," *The Review of Economic Studies*, 22, 153–168.

Bentzel, R. and H. Wold (1946): "On Statistical Demand Analysis from the Viewpoint of Simulatneous Equations," *Skandivavisk Aktuarietidskrift*, 29, 95–114.

Card, D., J. Kluve, and A. Weber (2010): "Active Labour Market Policy Evaluations: A Meta-Analysis," *The Economic Journal*, 120, 452–477.

Cartwright, N. (2007): *Hunting Causes and Using Them*, Cambridge University Press.

Correa, J. D., J. Tian, and E. Bareinboim (2017): "Generalized Adjustment Under Confounding and Selection Biases," Tech. Rep. R-29-L.

Dehejia, R., C. Pop-Eleches, and C. Samii (2015): "From Local to Global: External Validity in a Fertility Natural Experiment," NBER Working Paper No. 21459.

Duflo, E., R. Glennerster, and M. Kremer (2008): "Using Randomization in Development Economics Research: A Toolkit," in *Handbook of Development Economics*, Elsevier, vol. 4, chap. 61.

Duflo, E. and E. Saez (2003): "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *Quarterly Journal of Economics*, 118, 815–842.

Frisch, R. (1933): "Editor's Note," *Econometrica*, 1, 1–4.

Glynn, A. N. and K. Kashin (2017): "Front-Door Difference-in-Differences Estimators," *American Journal of Political Science*, 61, 989–1002.

Gordon, B. R., F. Zettelmeyer, N. Bhargava, and D. Chapsky (2018): "A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook," .

Haavelmo, T. (1943): "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11, 1–12.

Heckman, J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *The Annals of Economic and Social Measurement*, 5, 475–492.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

Heckman, J. J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator," *The Review of Economic Studies*, 65, 261–294.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64, 605–654.

Heckman, J. J. and R. Pinto (2013): "Causal Analysis after Haavelmo," *Econometric Theory*, 31, 115–151.

Heckman, J. J. and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738.

Heckman, J. J. and E. J. Vytlacil (2007): "Econometric Evaluation of Social Programs, Part 1: Causal Models, Structural Models and Econometric Policy Evaluation," in *Hanbook of Econometrics*, Elsevier B.V., vol. 6B.

Hoover, K. D. (2004): "Lost Causes," *Journal of the History of Economic Thought*, 26.

Hotz, V. J., G. W. Imbens, and J. H. Mortimer (2005): "Predicting the efficacy of future training programs using past experiences at other locations," *Journal of Econometrics*, 125, 241–270.

Huang, Y. and M. Valtorta (2006): "Pearl's Calculus of Interventions Is Complete," in *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI2006)*.

Ihlanfeldt, K. R. and J. Martinez-Vazquez (1986): "Alternative Value Estimates of Owner-Occupied Housing: Evidence on Sample Selection Bias and Systematic Errors," *Journal of Urban Economics*, 20, 356–369.

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *The Review of Economics and Statistics*, 86, 4–29.

——— (2014): "Instrumental Variables: An Econometrician's Perspective," *Statistical Science*, 29, 323–358.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Imbens, G. W. and D. B. Rubin (2015): *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press.

Kowalski, A. E. (2018): "How to examine External Validity Within an Experiment," NBER Working Paper 24834.

Levitt, S. D. and J. Porter (2000): "Sample Selection in the estimation of air bag and seat belt effectiveness," *The Review of Economics and Statistics*, 83, 603–615.

List, J. A. (2011): "Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off," *Journal of Economic Perspectives*, 25, 3–16.

Maddala, G. S. (1986): *Limited-Dependent and Qualitative Variables in Econometrics*, Econometric Society Monographs.

Manski, C. F. (1990): "Nonparametric bounds on treatment effects," *American Economic Review, Papers and Proceedings*, 80, 319–323.

——— (2003): *Partial Identification of Probability Distributions*, New York: Springer.

Matzkin, R. L. (2007): "Nonparametric Identification," in *Handbook of Econometrics*, vol. 6B.

——— (2013): "Nonparametric Identification in Structural Economic Models," *Annual Review of Economics*, 5, 457–486.

Morgan, M. S. (1991): "The Stamping Out of Process Analysis in Econometrics," in *Appraising Economic Theories: Studies in the Methodology of Research Programs*, ed. by N. D. Marchi and M. Blaug, Aldershot, UK: Edward Elgar, 237–265.

Nakamura, E. and J. Steinsson (2018): “Identification in Macroeconomics,” *Journal of Economic Persepctives*, 32, 59–86.

Neuberg, L. G. (2003): “Review of ’Causality: Models, Reasoning, and Inference’,” *Econometric Theory*, 19, 675–685.

Neyman, J. (1923): “Sur les applications de la thar des probabilities aux experiences agraricales: Essay des principle,” English translation of excerpts (1990) by D. Dabrowska and T. Speed in *Statistical Science*, 5:463-472.

Pearl, J. (1988): *Probabilistic Reasoning in Intelligent Systems*, San Mateo, CA: Morgan Kaufmann.

——— (1993): “Graphical models, causality, and interventions,” *Statistical Science*, 8, 266–269.

——— (1995): “Causal diagrams for empirical research,” *Biometrika*, 82, 669–709.

——— (2000): *Causality: Models, Reasoning, and Inference*, New York, United States, NY: Cambridge University Press, 1st ed.

——— (2009): *Causality: Models, Reasoning, and Inference*, New York, United States, NY: Cambridge University Press, 2nd ed.

——— (2013): “Reflections on Heckman and Pinto’s ’Causal analysis after Haavelmo’,” Tech. Rep. R-420, Univiversity of California, Los Angeles.

——— (2015a): “Generalizing Experimental Findings,” *Journal of Causal Inference*, 3, 259–266.

——— (2015b): “Trygve Haavelmo and the Emergence of Causal Calculus,” *Econometric Theory*, 31, 152–179.

Pearl, J. and E. Bareinboim (2011): “Transportability of Causal and Statistical Relations: A Formal Approach,” in *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, Menlo Park, CA: AAAI Press, 247–254.

——— (2014): “External Validity: From Do-Calculus to Transportability Across Populations,” *Statistical Science*, 29, 579–595.

Pearl, J. and D. Mackenzie (2018): *The Book of Why: The New Science of Cause and Effect*, New York: Basic Books.

Richardson, T. S. and J. M. Robins (2014): “ACE Bounds; SEMs with Equilibrium Conditions,” *Statistical Science*, 29, 363–366.

RIDDER, G. AND R. MOFFITT (2007): *The Econometrics of Data Combination*, Elsevier B.V., vol. 6B, chap. 75, 5470–5547.

ROBINS, J. M. (1999): "Testing and estimation of of directed effects be reparameterizing directed acyclic with structural nested models," in *Computation, Causation, and Discovery*, ed. by C. N. Glymour and G. F. Cooper, Cambridge, MA: AAAI/MIT Press, 349–405.

ROSENBAUM, P. R. AND D. B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

RUBIN, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

SHPITSER, I. AND J. PEARL (2006): "Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models," in *Twenty-First National Conference on Artificial Intelligence.*

SPIRTES, P., C. GLYMOUR, AND R. SCHEINES (2001): *Causation, Prediction, and Search*, Cambride, MA: The MIT Press, 2nd ed.

STROTZ, R. H. AND H. O. A. WOLD (1960): "Recursive vs. Nonrecursive Systems: An Attempt At Synthesis (Part I of a Triptych on Causal Chain Systems)," *Econometrica*, 28, 417–427.

TEXTOR, J. AND M. LIŚKIEWICZ (2011): "Adjustment Criteria in Causal Diagrams: An Algorithmic Perspective," in *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*, AUAI press, 681–688.

TIAN, J. AND J. PEARL (2002): "A general identification condition for causal effects," in *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, Menlo Park, CA: AAAI Press/The MIT Press, 567–573.

VERMA, T. AND J. PEARL (1988): "Causal networks: Semantics and expressiveness," in *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*, Mountain View, CA, 352–359.

WOLD, H. (1954): "Causality and Econometrics," *Econometrica*, 22, 162–177.

——— (1981): *The Fix-point Approach to Interdependent Systems*, North-Holland Publishing Company, chap. The Fix-point Approach to Interdependent Systems: Review and Current Outlook, 1–36.

WOLD, H. O. A. (1960): "A Generalization of Causal Chain Models (Part III of a Triptych on Causal Chain Systems)," *Econometrica*, 28, 443–463.

Woodward, J. (2003): *Making Things Happen*, Oxford Studies in Philosophy of Science, Oxford University Press.

Zellner, A. (1979): “Causality and econometrics,” *Carnegie-Rochester Conference Series on Public Policy*, 10, 9–54.

APPENDIX

A.1. Causality in recursive and interdependent systems

In this paper, attention is restricted to a class of models that can be described by directed acyclic graphs and in which the rules of do-calculus apply. The requirement of acyclicity gives rise to what economists commonly denote as *recursive* systems (Wold, 1954; Pearl, 2009, p. 231). At the same time, many standard models in economics, such as the canonical supply and demand relationship, as well as game theoretic models, are nonrecursive or *interdependent*. In the aftermath of Haavelmo's celebrated paper on simultaneous equation models (Haavelmo, 1943), an intensive discussion about the conceptual interpretation of recursive versus interdependent models emerged in the econometrics literature (see Morgan, 1991, for an excellent historical account). The debate was particularly motivated by practical concerns of estimation. Haavelmo showed for the first time in full clarity that the method of least squares does not lead to unbiased parameter estimates in interdependent simultaneous equation models.[25] However, it also touched on the causal interpretation of interdependent models and the adequacy of cyclic causal relationships as a representation of economic processes. One central argument, most notably formulated in Bentzel and Hansen (1954) and Strotz and Wold (1960), was that individual equations in an interdependent model do not have a causal interpretation *"in the sense of a stimulus-response relationship"* (Strotz and Wold, 1960, p. 417).[26] Rather, interdependent systems with equilibrium conditions are regarded as *"shortcut"* descriptions (Wold, 1960; Imbens, 2014) of the underlying dynamic behavioral processes (Herman Wold coined the term *causal chain* for the latter).

In this context, Strotz and Wold (1960) discuss the example of the cobweb model, a particular form of dynamic supply and demand model based on Jan Tinbergen's

[25] As a matter of fact, Haavelmo never made a distinction between recursive and interdependent models in his 1943 paper. Starting from a interdependent simultaneous equation model, he demonstrated that OLS is biased in this context. Later, Bentzel and Wold (1946; as cited in Wold, 1981) were able to show that least squares estimation is indeed appropriate if the system is recursive.

[26] More than two decades later, in his influential textbook Maddala (1986, p. 111) presents a similar point of view.

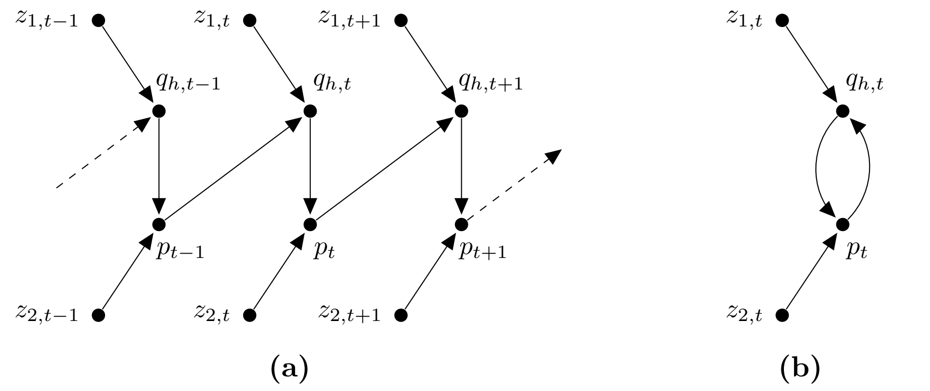


Figure 10: *(a) Dynamic, recursive model of a market for crops. (b) Nonrecursive model of the same market after imposing an equilibrium constraint.*

microeconometric work in the 1920s (see Morgan, 1991)

$$q_{h,t} = \gamma + \delta p_{t-1} + \nu z_{1,t} + u_{1,t}, \tag{A.1}$$
$$p_t = \alpha - \beta q_{h,t} + \varepsilon z_{2,t} + u_{2,t}. \tag{A.2}$$

This model is recursive. The first equation determines the quantity of a particular crop harvested at time t, based on the crop's price p_{t-1} in the previous period. The second equation describes crop demand and pins down prices in t, depending on current supply. Moreover, the model incorporates exogenous supply and demand shifters z_1 and z_2. By imposing an equilibrium assumption on the system, such that prices are required to remain constant over time

$$p_{t-1} = p_t, \tag{A.3}$$

the model becomes interdependent, as price and quantity now affect each other simultaneously in the same period.

$$q_{h,t} = \gamma + \delta p_t + \nu z_{1,t} + u_{1,t}, \tag{A.4}$$
$$p_t = \alpha - \beta q_{h,t} + \varepsilon z_{2,t} + u_{2,t}. \tag{A.5}$$

Figure 10 illustrates the step from the fully dynamic model to a nonrecursive equilibrium model graphically.[27] Note, however, that the equilibrium assumption

[27] Bentzel and Hansen (1954) point out that interdependency can also be the result of an ag-

(A.3) carries no behavioral interpretation and may or may not describe the data adequately. Likewise, the individual equations of the interdependent system do not represent autonomous causal relationships in the stimulus-response sense, since the endogenous variables are determined jointly by all equations in the system (Matzkin, 2013; Heckman and Pinto, 2013). Thus, it would not be possible, for example, to directly use p_t in equation (A.4) to bring about a desired change in $q_{h,t}$.

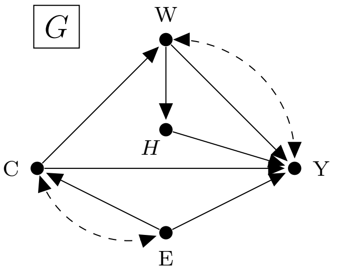
That is not to say – as these authors have stated repeatedly – that equilibrium models cannot be useful for learning about individual causal parameters (Strotz and Wold, 1960, p. 426), nor that a causal interpretation cannot be given to a nonrecursive model as a whole (Bentzel and Hansen, 1954; Basman, 1963; Zellner, 1979). However, if individual functions of an economic model are supposed to be interpreted as stimulus-response relationships, cyclic patterns need to be excluded. Otherwise, stimuli would be permitted to be causes of themselves, which would violate the notion of asymmetry usually attached to them (Woodward, 2003; Cartwright, 2007).[28] Incidentally, the potential-outcomes framework in the treatment effects literature also interprets the link between treatment and outcome as a stimulus-response relationship and therefore implicitly maintains the assumption of acyclicty (Heckman and Vytlacil, 2007).

A.2. Do-calculus derivation of backdoor adjustment formula

In this section we derive the causal effect for the college wage premium example given in Figure 2a of Section 3.1 by using the inference rules of do-calculus. In addition, this derivation provides a justification for the adjustment formula in equation (3.1) when backdoor-admissible sets are available. For illustration purposes, subgraphs used in the respective steps of the do-calculus are placed alongside.

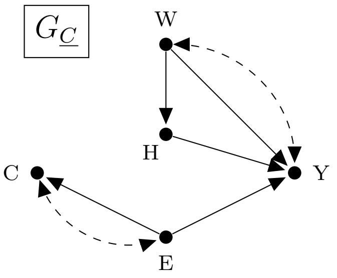
gregation of variables measured at an inappropriate frequency, even if the underlying data generating process is fully recursive.

[28] To emphasize the interpretation of individual functions in an SCM as a stimulus-response relationships, equality signs should be better replaced by assignment operators "←" (similar to the syntax of computer languages; Richardson and Robins, 2014), which change meaning under solution-preserving algebraic operations (Pearl, 2009, p. 27).



There are two backdoor paths in the post-intervention graph $G_{\underline{C}}$ that can both be blocked by E. Conditioning and summing over all values of E yields

$$P(y|do(c)) = \sum_{e} P(y|do(c), e)P(e|do(c)).$$

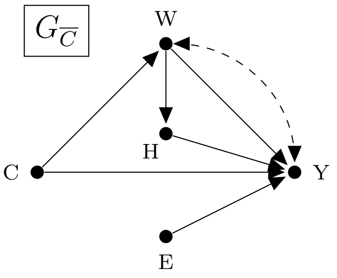


By rule 2 of do-calculus

$$P(y|do(c), e) = P(y|c, e), \quad \text{since } (Y \perp\!\!\!\perp C|E)_{G_{\underline{C}}}.$$

By rule 3 of do-calculus

$$P(e|do(c)) = P(e), \quad \text{since } (E \perp\!\!\!\perp C)_{G_{\overline{C}}}.$$



It follows that

$$P(y|do(c)) = \sum_{e} P(y|c, e)P(e).$$

The right-hand side expression is do-free and can therefore be estimated from observational data.