Intro — Introduction to causal inference and treatment-effects estimation

Description Remarks and examples References

Description

This entry provides an introduction to causal inference and treatment-effects estimation. It presents concepts, frameworks, and assumptions that researchers consider when they wish to draw causal inferences in their analyses.

For information on Stata commands that estimate treatment effects and that are specifically designed for causal inference, see [CAUSAL] **Causal inference commands**.

For more in-depth introductions to causal inference, see Imbens and Rubin (2015), Robins and Greenland (1992), Hernán and Robins (2020), and Pearl (2009).

Remarks and examples

Remarks are presented under the following headings:

Motivation: Causation versus association

Research may be driven by the desire to evaluate causation or association. Causal questions explore changes in the outcome when we change a variable under our control or examine what would happen to the outcome if a variable of interest had not changed. For example:

- Does receiving a treatment cure the illness?
- What would have happened to the inflation rate if the Federal Reserve had not increased interest rates?
- Does smoking reduce fetal growth?
- Does raising the minimum wage decrease unemployment?

In contrast, associative questions observe patterns in data. For example:

- How does the cure rate between patients who received treatment and those who did not receive treatment compare?
- Is there a correlation between interest rates and inflation?
- Is there a difference in the mean birthweight of infants born to mothers who smoke versus those born to mothers who do not smoke?
- What is the difference in the unemployment rate between states that have implemented an increase in minimum wage and those that have not?

The first set of questions asks what happens when there is an intervention or imagine a scenario where a variable changes versus does not change. The second set of questions observe only the pattern without intervention.

To examine some of the considerations we must face when performing causal inference, we consider a hypothetical study, conducted by a software company called Statanium, that examines the relationship between the number of breaks taken by software developers and their productivity. The company wants to find out whether the number of breaks impacts productivity. As a user of Statanium and an expert in causality, you are hired to advise the company whether it should encourage developers to take additional breaks during their workday. The question the company is interested in is causal because it aims to compare the productivity when the number of breaks is increased versus the productivity when the number of breaks (and other possible factors) remains unchanged. If the two outcomes of productivity are different, then the action of increasing the number of breaks has a causal effect. In causal inference literature, the action of increasing the number of breaks is referred to as a treatment or an intervention.

You could estimate an association between increased breaks and productivity via correlation or many other statistical methods that estimate dependence. However, the interest is in estimation of causal effect. The well-known expression "association is not causation" suggests that for any given amount of association, only some part or none of it is causal. Thus, a challenge in causal inference is to identify and eliminate relationships that are only associative. To perform causal inference in our case, we want to create a hypothetical scenario where the number of breaks is increased and all other factors that may influence productivity remain fixed. Then we can determine the causal effect of the additional break.

When it is possible, randomized experiments are a popular method to estimate causal effects because the treatment (number of breaks) is controlled by the experimenter. Therefore, the treatment assignment is guaranteed to be random and unrelated to all other factors that may determine the outcome. This can be represented using causal diagrams, as in figure 1, where T represents the treatment, Y represents the productivity or outcome, and X represents all other factors. The absence of an arrow between X and T indicates that X does not affect T.

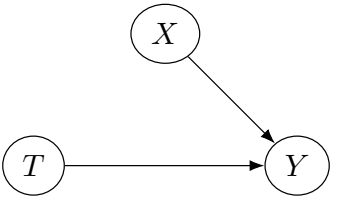

Figure 1.

However, randomized experiments are not always feasible, and causal effects need to be inferred from observational data. In the observational data case, the experimenter does not have control over the treatment assignment, and the assumption that all other factors are held constant in both the observed and hypothetical worlds may not hold. This is due to the presence of confounding factors (for example, job satisfaction) that affect both the treatment and outcome. This is reflected in the causal diagram in figure 2 by arrows from X to both T and Y.

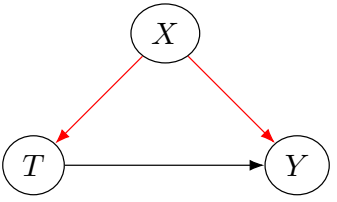

Figure 2.

Here the measured total association between T and Y contains both causal path $T \rightarrow Y$ and association path $T \leftarrow X \rightarrow Y$, highlighted in red. To identify the causal effect and make further causal inference, you need to eliminate the association represented by the red path by accounting or adjusting for other confounding factors.

The following highlights the significance of adjusting for confounding variables. Suppose that in figure 3 you plot the productivity of software developers as a function of the average number of breaks.

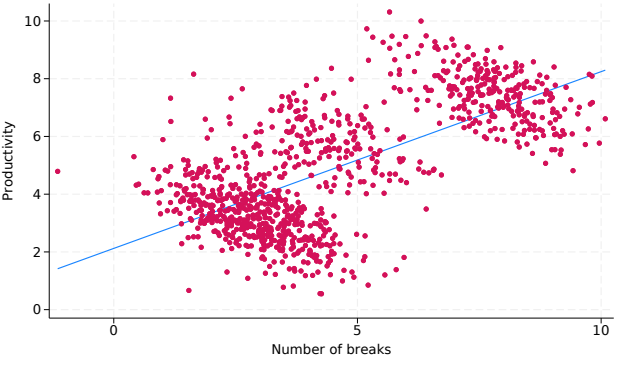

Figure 3.

The plot shows a positive association between the number of breaks and productivity, implying that taking more breaks leads to higher productivity. The questions of interest are whether this association can be considered causal and whether Statanium should motivate its hardworking developers to take more breaks during the workday. The conclusion that more breaks are beneficial would be valid if the model assumed is as shown in figure 1. However, as an expert, you believe that there are confounding factors, such as workload or job satisfaction, that need to be considered, as shown in figure 2. To account for the effect of job satisfaction, in figure 4, you plot the productivity of software developers against the average number of breaks they take, considering different levels of job satisfaction—dissatisfied (orange), satisfied (red), and highly satisfied (yellow).

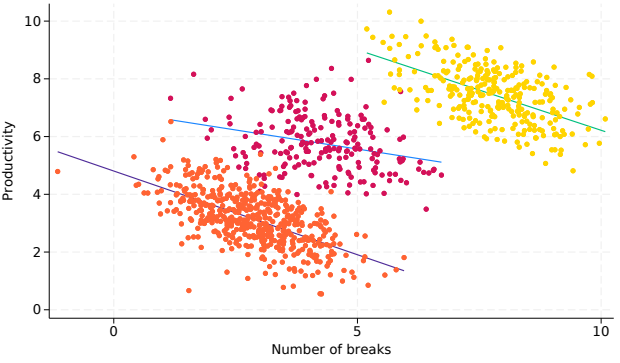

Figure 4.

The plot shows that when job satisfaction is accounted for, the effect of the number of breaks on productivity changes from positive to negative. This phenomenon is known as Simpson's paradox (Blyth 1972), where the overall effect appears to be positive but, when it is adjusted for a confounder, the effect sign changes direction.

The above example illustrates the fundamental difference between causation and association. For example, if for some reason the researcher does not account for the job satisfaction or it is not observed, then the estimated association cannot be interpreted as causal.

Below, we introduce the workflow, popular causal inference frameworks, and assumptions that allow researchers to draw causal inferences rather than merely find associations.

Causal inference workflow

The causal inference literature recommends the following three-phase workflow (Pearl 2009; Imbens 2020; and Heckman and Pinto 2022) when a research question is causal in nature:

1. Hypothetical modeling: Researchers make assumptions about relationships among variables based on their understanding and expertise. These assumptions are related not only to the treatment variable and the outcome of interest but also to any variables that might be related to the treatment or the outcome. The assumptions regarding these relationships cannot be tested from data; therefore, the validity of these assumptions must come from previous theory or the researcher's own expertise.
2. Causal effect identification: Based on the assumptions made in the first phase, the researcher tries to determine whether the causal effect can be identified.
3. Parameter estimation: If the answer to the second phase is positive, the researcher can then use various estimation techniques, such as those provided by the commands discussed in [CAUSAL] **Causal inference commands**, to estimate the causal effect.

Potential-outcomes framework

The potential-outcomes framework is one of the most commonly used theories for understanding and evaluating causal inference. The foundation of this framework is motivated by the idea of a natural experiment (Imbens 2020), which focuses on finding settings and identification strategies under which the assignment of a treatment can be considered as good as random even though the data are observational rather than from a randomized experiment. The roots of the potential-outcomes framework trace back to

a seminal paper by Splawa-Neyman (1923), and the framework was formally introduced in Rubin (1974). Fisher (1925) built upon Neyman's ideas and introduced the idea of physical randomization, which formally defines the concept of a treatment-assignment mechanism. For further reading, see Imbens and Rubin (2015).

As discussed earlier, the goal of causal inference is to estimate the change in an outcome as the treatment varies. In the Statanium example, we now assume that the treatment (number of breaks) and the outcome (productivity) are binary. That is, $T_i = 1$ means the Statanium developer takes additional breaks, and $T_i = 0$ otherwise. Similarly, $Y_i = 1$ if a developer's productivity increases, and $Y_i = 0$ otherwise. Thus, we are interested in estimating the change in productivity if the developer takes additional breaks, $T_i = 1$, versus if the developer does not take additional breaks, $T_i = 0$. However, for a given developer, we can observe only one outcome. This is known as the fundamental problem of causal inference (Holland 1986). The potential-outcomes framework provides tools and assumptions to solve this problem.

The important concepts in potential outcomes are the unit, treatment, and outcome. A unit is the research object to which treatment is assigned. It can be, for example, a person, a company, a school, or a county. Treatment is the action that we apply to a unit. In this entry, treatment is denoted as T. For binary treatment, the units to which treatment is applied ($T = 1$) are called the treated group, and the units to which the treatment is not applied ($T = 0$) are called the control group. The potential outcome of treatment with value t for unit i is denoted by $Y_i(T_i = t)$ or $Y_i(t)$. In our running example, the unit is the developer, treatment is taking additional breaks, and the outcome is whether productivity improves.

The observed outcome is related to the potential outcome through $Y_i = T_i Y_i(1) + (1 - T_i)Y_i(0)$. That is, if the unit receives the treatment $T_i = 1$, then $Y_i = Y_i(1)$ and $Y_i(0)$ otherwise.

A potential outcome that is not observed is called a counterfactual outcome. For example, for the developer who took additional breaks, $Y_i(1)$ is the observed outcome and $Y_i(0)$ is the counterfactual outcome.

Treatment-effect estimands

Individual treatment effect (ITE). For each developer or unit i, the ITE is defined as

$$Y_i(T_i = 1) - Y_i(T_i = 0)$$

Because only one of the potential outcomes is observed, we cannot identify this quantity directly from the data without making assumptions about the unobserved counterfactuals and the assignment to treatment. However, estimates of individual effects can be useful for providing insight into how a treatment may affect an individual. For instance, in epidemiology, estimated ITEs could help determine whether a treatment is likely to be helpful for a particular patient.

We can also define and estimate effects that allow us to draw interesting causal inferences for the population instead of for each individual.

Average treatment effect (ATE). The ATE is also known as the average causal effect. The ATE at the population level can be defined as the mean difference in potential outcomes when units received a treatment versus when units did not receive any treatment,

$$\text{ATE} = E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)] \tag{1}$$

where $Y(t) = Y(T = t)$, for $t = \{0, 1\}$. We say that the ATE of treatment T on outcome Y exists if $E[Y(1)] \neq E[Y(0)]$.

The ATE provides an estimate of the expected average effect in the population and can therefore be used to answer many interesting questions that could help policymakers make decisions. In our Statanium example, the ATE is the expected difference in productivity if all developers took extra breaks versus none taking extra breaks. Statanium could determine whether extra breaks are beneficial and set break policies based on the estimated ATE. Similarly, we interpret the ATE for a couple of our motivating causal questions. When investigating whether a treatment cures an illness, we could interpret the ATE as the expected difference in the proportion of individuals who were cured from an illness when everyone received the treatment versus when no one received the treatment. When evaluating a raise in minimum wage, we could interpret the ATE as the expected change in unemployment rate when the minimum wage is raised for everyone versus when the minimum wage stays the same.

When the ATE is of interest, we must use an appropriate method that leads to an estimate of this quantity. We might be tempted to use association difference to estimate ATE, where association difference is the conditional mean difference of outcomes between treatment and not treated units. The treatment T and Y are associated if $E[Y|T = 1] \neq E[Y|T = 0]$. Thus, one might try to estimate the causal quantity in (1) with the statistical quantity $E[Y|T = 1] - E[Y|T = 0]$. For the Statanium example, this quantity could be estimated from the data by contrasting the sample average of the developers that took the treatment with the ones that did not. Mathematically,

$$\frac{\sum_{i=1}^{N} Y_i T_i}{\sum_{i=1}^{N} T_i} - \frac{\sum_{i=1}^{N} Y_i (1 - T_i)}{\sum_{i=1}^{N} (1 - T_i)}$$

However, in general, ATE and association difference are different; otherwise, association would be causation. Figure 5 highlights the causation–association difference.

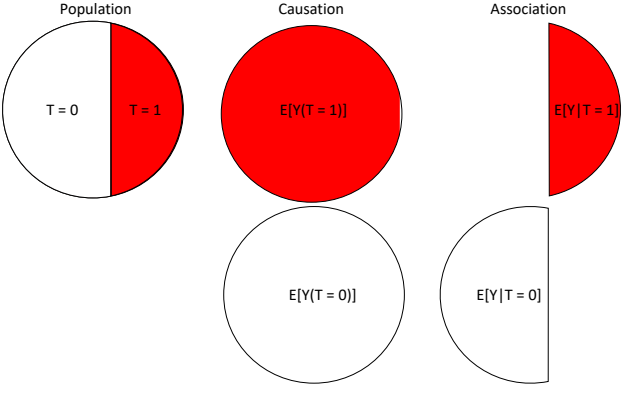

Figure 5.

To infer causation, we imagine that each treatment or intervention is applied to the entire population, and the difference between the red and white circles is observed in the same population. In contrast, to infer association, we condition on $T = t$ and estimate the difference between subsets of populations.

Average treatment effect on the treated (ATET). For the treated group, the treatment-effect estimand is the ATET:

$$\text{ATET} = E[Y(1)|T = 1] - E[Y(0)|T = 1]$$

The ATET is useful when researchers are interested in the effect on those who received the treatment. This effect may be of particular interest when the goal is to understand how a treatment performs for the subpopulation at which the treatment was targeted. The effect is a comparison with what would have happened in this subpopulation if they had not received the treatment. In our Statanium example, we estimate the expected difference in productivity for developers who took extra breaks compared with the productivity of these developers if they had not taken extra breaks. When we investigate whether a treatment cures an illness, the ATET is focused only on those who received the treatment. For this group, what is the expected difference in the proportion of individuals who were cured when given the treatment versus the proportion cured if they had not received the treatment? When evaluating a raise in minimum wage, we might be interested in the effect for states that enacted a minimum wage increase. The ATET is the expected difference in unemployment rate for these states compared with a situation where these states did not raise minimum wage.

Additional estimands. In specific situations, there are several other treatment-effect estimands that may be of interest and that can be defined in the potential-outcomes framework. We briefly mention a few here.

Sometimes, it is assumed that the effect of T on Y may involve both a direct effect and an indirect effect such that T has an effect on another variable M, known as a mediator, and that M in turn has an effect on Y. In the Statanium example, we might believe that increased breaks could improve the developers' focus while working and that improved focus leads to increased productivity. In such a situation, comparisons of average direct treatment effects and average indirect treatment effects may be of interest. See Robins and Greenland (1992), VanderWeele (2015), and Pearl and MacKenzie (2018) for discussions of causal mediation analysis and definitions of applicable direct and indirect effect estimands.

Recently, there has been a surge of interest in estimating the treatment effect when it differs between subgroups, also known as the heterogeneous treatment effect (Athey and Imbens 2016; Künzel et al. 2019 ; and Nie and Wager 2021). At the subgroup level, the treatment-effect estimand is called conditional average treatment effect (CATE),

$$\text{CATE} = E[Y(1)|X = x] - E[Y(0)|X = x]$$

where $Y(t)|X = x$ for $t = \{0, 1\}$ are the potential outcomes of the subgroup $X = x$.

Assumptions required in potential-outcomes framework

At this point, it is natural to ask under which conditions treatment effects can be estimated from observational data. For illustration purposes, our focus will be on ATE. We are interested in assumptions for which

$$\text{ATE} = E[Y(1)] - E[Y(0)] = E[Y|T = 1] - E[Y|T = 0]$$

A causal quantity, that is, ATE, is "identifiable" if it can be computed from a statistical quantity $E[Y|T = t]$. By statistical quantity, we mean an object that can be estimated from data.

In the potential-outcome framework, commonly used assumptions are the stable unit treatment value assumption (SUTVA), unconfoundedness assumption, and overlap assumption.

SUTVA. The SUTVA, along with consistency, states that for a given unit, the treatment of other units does not affect the outcome of the treatment received by that unit. Consequently, there are two different sources in which SUTVA could be violated. The first source is a violation of the consistency condition, which might not hold in some studies (Cole and Frangakis 2009; and Schwartz, Gatto, and Campbell 2011). Usually, the problem arises from the vagueness of the assigned treatment. For example, for the Statanium example, if the treatment is additional breaks, this may be one additional break or three additional breaks. If we observe only whether a treatment has been assigned, the counterfactual $Y(T = t)$ is not well defined because different numbers of breaks have different causal effects. The second source of violation arises if some units are influenced by the assignment of the treatment of other units. For example, if some developers in the control group noticed that the developers in the treatment group are less productive, they might change their lifestyle and start taking fewer breaks, which can lead to an increase in productivity. Typically, interference can occur because of spillover effects or noncompliance or because the units are members of a social network. For details, see Hernán and Robins (2020, chap. 3).

Unconfoundedness assumption. The unconfoundedness assumption goes by many names, including the conditional-independence assumption, the ignorability assumption, and the exchangeability assumption. The impetus behind this assumption is to make the treatment and control group comparable within strata defined by X. It states that the probability of a positive outcome in the control group (white group in the figure above) would be the same as the probability of a positive outcome in the treatment group (red group) had units in the control group received the treatment given to those in the treatment group. In other words, under unconfoundedness, if by accident the treatment were given to the white group instead of the red group, then the ATE would remain the same. Mathematically, this is represented by $(Y(1), Y(0)) \perp\!\!\!\perp T|X$, where $\perp\!\!\!\perp$ denotes (conditional) independence and X are potential confounders. It is important to differentiate between $Y(t) \perp\!\!\!\perp T|X$, which utilizes potential outcomes, and $Y \perp\!\!\!\perp T|X$. The unconfoundedness assumption does not imply that $Y \perp\!\!\!\perp T|X$. On the contrary, if the ATE is not zero, then Y and T are associated. In the context of the Statanium example, the underlying intuition is that for two developers i and j, their potential outcomes should be independent of the treatment assignment $P(Y_i(0), Y_i(1)|T = t_i, X) = P(Y_j(0), Y_j(1)|T = t_j, X)$, for $t_i, t_j \in \{0, 1\}$.

Overlap assumption. Finally, the overlap or positivity assumption $P(T = t|X = x) > 0$ implies that the treatment assignment should be stochastic. For example, if the developers who are highly satisfied are always assigned treatment $T = 1$, then there is no meaning in studying the treatment $T = 0$. In contrast to unconfoundedness, the overlap assumption can sometimes be verified from the data. For details, see Hernán and Robins (2020, chap. 12).

When these three main assumptions in the potential-outcome framework are satisfied, then the ATE is identified, and estimation methods available in the `teffects` suite of commands can be used.

Relaxing causal assumptions

One of the crucial questions in causal inference is whether all confounders have been accounted for in the study. Unfortunately, the unconfoundedness is not testable from data. There are observable and unobservable confounders that, unaccounted for, will lead to incorrect conclusions. This problem of unobserved confounders or endogeneity is usually addressed using estimators that account for an endogenous treatment or instrumental-variable method.

Discussions of these issues can be found in Imbens and Rubin (2015). For a general treatment, see Wooldridge (2010) and Angrist and Pischke (2009).

The potential-outcome framework is not restricted only to models in which identification relies on conditional unconfoundedness, which is sometimes referred to as selection on observables. It also allows us to recover the ATET by controlling for individual and time-varying unobservables, without the need to control for covariates. One example of this is the difference-in-differences method, which can be used to estimate ATET by comparing the change in outcomes between the treatment and control groups over time.

Causal diagrams

In *Motivation: Causation versus association*, we used causal diagrams, specifically directed acyclic graphs (DAGs), to represent our assumptions about causal relationships. DAGs are helpful in phases 1 and 2 described in *Causal inference workflow*. The intellectual predecessor of causal diagrams, or more generally, the structural causal models (SCMs) framework, goes back to the pioneering work of geneticist Wright (1921, 1934) and econometricians Frisch and Waugh (1933) and Haavelmo (1944) and the references therein. Early econometricians were attempting to conceptualize the fact that, unlike correlation, the regression of a variable Y on T has a natural direction and is different from the regression of T on Y. This distinction led to the development of causal frameworks by Pearl (2009) and Heckman and Pinto (2022). Furthermore, this differentiation is linked to the concept known as "the ladder of causality" (Pearl and MacKenzie 2018), where the lowest rung signifies association and higher rungs address causal queries. Here our focus is on causal diagrams. We provide only a brief introduction and explore the usefulness of causal diagrams in determining causal-effect identification before estimating treatment effects using one of Stata's estimation commands. We will mention some of the common terminology used in SCM, but for more detailed descriptions of these terms and additional details on SCM, we refer you to Pearl (2009), Peters, Janzing, and Schölkopf (2017), Bareinboim et al. (2022), and the references therein.

DAGs are characterized by nodes and directed edges between the nodes (shown as circles and arrows, respectively, in diagrams below), which represent causal relationships. The absence of edges between nodes indicates that there is no direct causal effect between those nodes. Unlike path diagrams in SEM, independent unobserved error terms are not depicted in DAGs.

For the Statanium example, we capture our assumptions using the DAG in the left panel of figure 6.

Figure 6. (left) Unconfoundedness and (right) Instrumental variable

Here X is a confounder in the causal relationship for T and Y, and because X is observed, we can control or adjust for it by including it as a covariate in our model when we estimate the causal effect. With this adjustment, we eliminate associative paths and identify the causal relationship between T and Y. In the causal diagram literature, the path $T \leftarrow X \rightarrow Y$ is known as a backdoor path because it contains an arrow that goes to the back of node T (hence, the name "backdoor"). If we do not condition on X, we say that we are leaving the path open, and we cannot estimate the causal effect. In this situation, we meet what is known as the backdoor path criterion because, by adjusting for X, we close all paths entering T.

In comparison, in the DAG shown on the right of figure 6, X is an unobserved confounder, which is indicated by shading the node for this variable. In this case, we cannot condition on X by including it as a covariate when we estimate the causal effect. Thus, the backdoor path criterion is not satisfied, and the causal effect of T on Y cannot be identified in this way. Fortunately, in this case, the assumed DAG

still allows us to estimate the causal effect because Z can be used as an instrumental variable. A proper instrumental variable can be included as a predictor of the treatment to eliminate unobserved sources of confounding such as that from X here. Z can be used as a proper instrumental variable because there is no direct effect of Z on the outcome Y and there are no unobserved confounders between Z and X or Z and Y.

For our simple DAGs, we were able to evaluate identification easily by considering the relationships among the few variables. For more complex situations, there exist several algorithms and criteria that are designed to check and achieve causal effect identification using a graphical approach. We do not discuss those algorithms here; we instead refer the interested reader to Shpitser and Pearl (2008), Pearl (2009), Maathuis and Colombo (2015), Perković et al. (2015), and van der Zander, Liśkiewicz, and Textor (2019). These algorithms can select variables that form a valid adjustment set that can be used for treatment-effect estimation. Once the valid adjustment set is selected, researchers can use their preferred treatment-effect estimation method from the list of commands in [CAUSAL] **Causal inference commands**.

Below, we demonstrate how a DAG can be used in the phases discussed in the *Causal inference workflow*. Consider a study that examines the effect on five-year mortality of polycystic kidney disease (PKD) among patients undergoing peritoneal dialysis (PD). Recall that in phase 1 of the workflow, we construct a hypothetical model that represents researchers' assumptions. We begin our causal analysis by drawing a DAG that represents our assumptions about the causal effects we are interested in estimating and the relationship among all variables, observed and unobserved, that may affect the variables of interest. Here we borrow the hypothetical model from Evans et al. (2012), illustrated in figure 6.

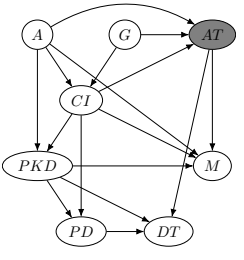

Figure 6.

The goal of phase 2 is to check whether the causal effect of the treatment variable PKD on mortality M is identifiable. For the variable definitions, see table 2. The gray node in the graph indicates that the variable AT is not observed.

Table 2. Variable definitions

PKD - polycystic kidney disease; treatment variable
M - mortality; outcome variable
A - patient age
G - patient gender
C - patient's comorbidities summary index
PD - indicator for patients undergoing peritoneal dialysis
AT - type of medical assistance
DT - type of peritoneal dialysis

Note that if all variables in the causal graph are observed, then the causal effect can be identified using backdoor path criteria (Pearl 2009).

However, in our example, AT is not observed, and there are unblocked noncausal (backdoor) paths that hinder identification:

P1: PKD ← CI → AT → M

P2: PKD ← A → AT → M

P3: PKD ← CI ← G → AT → M

Note that, despite a common misconception, conditioning on all observed variables does not mitigate the identification issue. For example, consider the path PKD → DT ← AT. This type of path is known as a v-structure, and DT is called a collider. PKD and AT are unrelated because the collider closes or blocks the path. However, we would not want to include DT in the model because conditioning on the collider DT introduces a selection bias through PKD → DT ← AT → M. For more examples, see *Importance of identification before estimation*. Interestingly, the above noncausal paths P1, P2, and P3 can be blocked if we condition on variables CI and A. In fact, CI and A are the necessary minimal sufficient set that makes the causal effect of PKD to M identifiable. Now that we know that the causal effect is identified, we are ready to move to phase 3, estimation of the treatment effect. In this case, CI and A can be used as covariates in, for instance, the inverse-probability weighting estimator. For details, see [CAUSAL] **teffects ipw**.

```
. teffects ipw (M) (PKD CI A)
```
(*output omitted*)

Importance of identification before estimation

In this section, we provide numerical examples that reemphasize the importance of the identification step. If a causal effect is not identified, the results of any treatment-effect estimation method are unreliable. We illustrate this point with two simple examples, which also aim to clear up the common misconception that it is harmless to control for more variables in treatment-effect estimation. For more examples, the reader is referred to Cinelli, Forney, and Pearl (2024) and Hünermund, Louw, and Caspi (2021).

▷ Example 1

The presumed relationship among variables in figure 7 is well known for introducing collider bias, sometimes called selection bias, in the estimated effect of T on Y. In the causal diagram literature, the represented DAG is known as a v-structure, and the variable X as a collider. A v-structure $T \rightarrow X \leftarrow Y$ has an interesting characteristic that the causal effect of T on Y is 0 because the collider X blocks the path. If we adjust for X, we introduce a selection bias because the adjustment results in an induced association between T and Y.

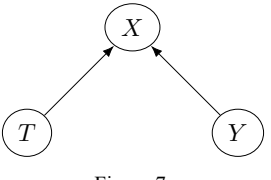

Figure 7.

Suppose the data come from the data-generation process as

$$
\begin{aligned}
T &:= \epsilon_T \\
Y &:= \epsilon_Y \\
X &:= 2 \times T - 0.5 \times Y + \epsilon_X
\end{aligned}
$$

where $\epsilon_T, \epsilon_Y, \epsilon_X \sim N(0,1)$. The regression of Y on T and X shows that T and X are for prediction purposes if we were to determine this on the basis of their p-values. However, the X is a bad control in terms of answering a causal question about the effect of T on Y.

```
. regress Y T X
```

Source	SS	df	MS	Number of obs	=	1,000
				F(2, 997)	=	111.51
Model	187.350778	2	93.675389	Prob > F	=	0.0000
Residual	837.504999	997	.840025074	R-squared	=	0.1828
				Adj R-squared	=	0.1812
Total	1024.85578	999	1.02588166	Root MSE	=	.91653

Y	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
T	.8088171	.0607918	13.30	0.000	.6895225	.9281117
X	-.3942266	.0263994	-14.93	0.000	-.4460314	-.3424219
_cons	.0345212	.0289833	1.19	0.234	-.0223541	.0913965

Fortunately, the true causal effect 0 is recovered if we do not control for X.

```
. regress Y T
```

Source	SS	df	MS	Number of obs	=	1,000
				F(1, 998)	=	0.02
Model	.025504326	1	.025504326	Prob > F	=	0.8748
Residual	1024.83027	998	1.02688404	R-squared	=	0.0000
				Adj R-squared	=	-0.0010
Total	1024.85578	999	1.02588166	Root MSE	=	1.0134

Y	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
T	.0049212	.0312266	0.16	0.875	-.0563562	.0661986
_cons	.0347662	.0320452	1.08	0.278	-.0281175	.0976498

◁

▷ Example 2

Here we demonstrate another type of bias that can be induced in the estimated effect of T on Y.

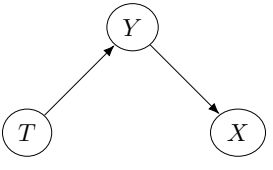

Figure 8.

Compared with the previous example, the harmfulness for controlling X is not explicit. Recall that in the previous section, we mentioned that the nodes for unobserved error terms (ϵ_T, ϵ_Y) are omitted from the causal graph, but they are there. If we imagine that the error term for Y is there, then Y would be a

collider because $T \rightarrow Y \leftarrow \epsilon_Y$. We also see $Y \rightarrow X$, so we can say that X is a descendant of the collider Y. It turns out that by controlling on the descendant X of collider Y, we induce an association between T and ϵ_Y (Pearl 2009, sec. 11.3). This conditioning opens a backdoor path between T and Y, which makes the causal effect of T on Y unidentified.

Suppose the data arise from the following data-generation process:

$$
\begin{aligned}
T &:= \epsilon_T \\
Y &:= 2 \times T + \epsilon_Y \\
X &:= -0.5 \times Y + \epsilon_X
\end{aligned}
$$

Then the causal effect of T on Y is 2. Similar to the previous example, even though X is good for prediction purposes, it is a bad control for causality.

```
. regress Y T X
```

Source	SS	df	MS	Number of obs	=	1,000
				F(2, 997)	=	2631.17
Model	4420.49975	2	2210.24987	Prob > F	=	0.0000
Residual	837.504998	997	.840025073	R-squared	=	0.8407
				Adj R-squared	=	0.8404
Total	5258.00474	999	5.26326801	Root MSE	=	.91653

Y	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
T	1.626137	.0379613	42.84	0.000	1.551644	1.700631
X	-.3942266	.0263994	-14.93	0.000	-.4460314	-.3424219
_cons	.0345212	.0289833	1.19	0.234	-.0223541	.0913965

Again, the true causal effect is recovered if we do not control for X.

```
. regress Y T
```

Source	SS	df	MS	Number of obs	=	1,000
				F(1, 998)	=	4122.35
Model	4233.17448	1	4233.17448	Prob > F	=	0.0000
Residual	1024.83027	998	1.02688404	R-squared	=	0.8051
				Adj R-squared	=	0.8049
Total	5258.00474	999	5.26326801	Root MSE	=	1.0134

Y	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
T	2.004921	.0312266	64.21	0.000	1.943644	2.066199
_cons	.0347662	.0320452	1.08	0.278	-.0281175	.0976498

These examples demonstrate that, before estimating a treatment effect, we must determine whether and how the effect can be identified based on our assumed relationships among variables. Once we have evaluated identification, we can select an appropriate method for estimating the causal effect. In these examples, we used linear regression. Actually, many of Stata's regression commands can be used for estimating treatment effects, provided that identification assumptions hold. Stata also offers estimation commands that are specifically designed for estimating treatment effects in various situations. For more information on these specialized commands, see [CAUSAL] **Causal inference commands**.

◁

References

Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Athey, S., and G. W. Imbens. 2016. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences* 113: 7353–7360. https://doi.org/10.1073/pnas.1510489113.

Bareinboim, E., J. D. Correa, D. Ibeling, and T. Icard. 2022. "On Pearl's hierarchy and the foundations of causal inference". In *Probabilistic and Causal Inference: The Works of Judea Pearl*, edited by H. Geffner, R. Dechter, and J. Y. Halpern, vol. 36: 507–556. New York: ACM Books. https://doi.org/10.1145/3501714.3501743.

Blyth, C. R. 1972. On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association* 67: 364–366. https://doi.org/10.2307/2284382.

Cinelli, C., A. Forney, and J. Pearl. 2024. A crash course in good and bad controls. *Sociological Methods and Research* 53: 1071–1104. https://doi.org/10.1177/00491241221099552.

Cole, S. R., and C. E. Frangakis. 2009. Commentary: The consistency statement in causal inference: A definition or an assumption? *Epidemiology* 20: 3–5.

Evans, D., B. Chaix, T. Lobbedez, C. Verger, and A. Flahault. 2012. Combining directed acyclic graphs and the change-in-estimate procedure as a novel approach to adjustment-variable selection in epidemiology. *BMC Medical Research Methodology* 12: art. 156. https://doi.org/10.1186/1471-2288-12-156.

Fisher, R. A. 1925. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Frisch, R., and F. V. Waugh. 1933. Partial time regressions as compared with individual trends. *Econometrica* 1: 387–401. https://doi.org/10.2307/1907330.

Haavelmo, T. 1944. The probability approach in econometrics. *Econometrica* 12 (Suppl.): 1–115. https://doi.org/10.2307/1906935.

Heckman, J. J., and R. Pinto. 2022. Causality and econometrics. Working Paper 29787, National Bureau of Economic Research. https://doi.org/10.3386/w29787.

Hernán, M. A., and J. M. Robins. 2020. *Causal Inference: What If*. Boca Raton, FL: CRC Press.

Holland, P. W. 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81: 945–960. https://doi.org/10.2307/2289064.

Hünermund, P., B. Louw, and I. Caspi. 2021. Double machine learning and automated confounder selection—a cautionary tale. arXiv:2108.11294 [econ.EM], https://doi.org/10.48550/arXiv.2108.11294.

Imbens, G. W. 2020. Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature* 58: 1129–1179. https://doi.org/10.1257/jel.20191597.

Imbens, G. W., and D. B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139025751.

Künzel, S. R., J. S. Sekhon, P. J. Bickel, and B. Yu. 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences* 116: 4156–4165. https://doi.org/10.1073/pnas.1804597116.

Maathuis, M. H., and D. Colombo. 2015. A generalized back-door criterion. *Annals of Statistics* 43: 1060–1088. https://doi.org/10.1214/14-AOS1295.

Nie, X., and S. Wager. 2021. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* 108: 299–319. https://doi.org/10.1093/biomet/asaa076.

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511803161.

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

Perković, E., J. Textor, M. Kalisch, and M. H. Maathuis. 2015. "A complete generalized adjustment criterion". In *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence*, 682–691. Amsterdam: Association for Uncertainty in Artificial Intelligence. https://auai.org/uai2015/proceedings/papers/155.pdf.

Peters, J., D. Janzing, and B. Schölkopf. 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge, MA: MIT Press.

Robins, J. M., and S. Greenland. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3: 143–155. https://doi.org/10.1097/00001648-199203000-00013.

Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701. https://doi.org/10.1037/h0037350.

Schwartz, S., N. M. Gatto, and U. B. Campbell. 2011. "What would have been is not what would be: Counterfactuals of the past and potential outcomes of the future". In *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*, edited by P. E. Shrout, K. M. Keyes, and K. Ornstein, 25–46. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199754649.003.0006.

Shpitser, I., and J. Pearl. 2008. Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research* 9: 1941–1979.

Splawa-Neyman, J. 1923. On the application of probability theory to agricultural experiments. Essay on principles. *Polish Agricultural Forestry Journal* 10: 29–42.

van der Zander, B., M. Liśkiewicz, and J. Textor. 2019. Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework. *Artificial Intelligence* 270: 1–40. https://doi.org/10.1016/j.artint.2018.12.006.

VanderWeele, T. J. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Oxford University Press.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Wright, S. 1921. Correlation and causation. *Journal of Agricultural Research* 20: 557–585.

———. 1934. The method of path coefficients. *Annals of Mathematical Statistics* 5: 161–215.