Chapter 8

Self-Selection, Heterogeneity, and Causal Graphs

In this chapter, we will lay the groundwork for our presentation of three strategies to estimate causal effects when simple conditioning on observed variables that lie along back-door paths will not suffice. These strategies will be taken up in Chapters 9, 10, and 11, where we will explain instrumental variable estimators, front-door identification with causal mechanisms, and conditioning estimators that use data on pretreatment values of the outcome variable. Under very specific assumptions, these three strategies will identify average causal effects of interest, even though selection is on the unobservables and treatment assignment is nonignorable.

In this chapter, we will first review the related concepts of nonignorable treatment assignment and selection on the unobservables, using the directed graphs presented in prior chapters. To deepen the understanding of these concepts, we will then demonstrate why the usage of additional posttreatment data on the outcome of interest is unlikely to aid in the point identification of the treatment effects of most central concern. One indirect goal of this demonstration is to convince the reader that oft-heard claims such as "I would be able to establish that this association is causal if I had longitudinal data" are nearly always untrue if the longed-for longitudinal data are additional measurements taken only after treatment exposure. Instead, longitudinal data are most useful, as we will later explain in detail in Chapter 11, when pretreatment measures are available for those who are subsequently exposed to the treatment.

The extended example we will use to make this point involves simple confounding (stable unobserved common causes of the treatment variable and the outcome variable) and more subtle confounding (direct self-selection into the treatment based on accurate perceptions of the individual-level treatment effect). Accordingly, this example will serve as a bridge to the second half of this chapter, where we will expand upon our introduction to causal graph methodology in order to explain how heterogeneity of effects can be accommodated. In Chapter 2, we introduced the potential outcome model while taking the position that individual-level heterogeneity of treatment effects is pervasive. In particular, we have implicitly assumed, and often explicitly

stated, that, in general, individual-level treatment effects do not all equal the average treatment effect (ATE) (i.e., $\delta_i \neq E[\delta]$ for all i). We have offered many examples where individual-level treatment effects are patterned in consequential nonrandom ways, and we have explained the consequences of such heterogeneity for conditioning estimators in Chapters 5, 6, and 7.

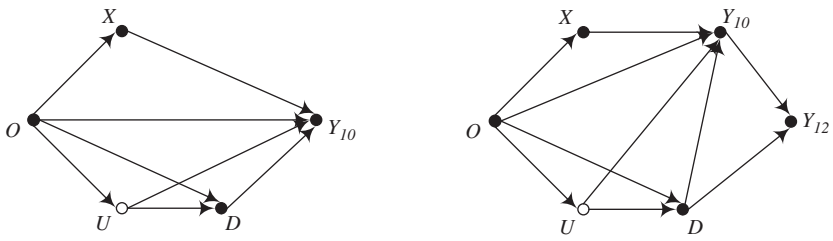
From this chapter onward, we will consider the interconnections between self-selection and individual-level heterogeneity of treatment effects more directly, as these patterns combine to generate selection on the unobservables. We ended our presentation of conditioning by considering these complications, especially for the interpretation of the results for Weighted Regression Demonstration 3 (see pages 257–262). In this chapter, we approach these complications more directly by first showing how to use causal graphs with latent class variables to jointly represent patterns of self-selection and heterogeneity.

8.1 Nonignorability and Selection on the Unobservables Revisited

As demonstrated in Sections 4.3.1 and 4.3.2, the concept of ignorable treatment assignment is closely related to the concept of selection on the observables. In many cases, they can both be represented by the same directed graph. Recall Figure 4.8(a), in which there are two types of paths between D and Y: the causal effect of D on Y represented by $D \rightarrow Y$ and an unspecified set of back-door paths represented collectively by the bidirected edge in $D \dashleftarrow\!\dashrightarrow Y$. If this graph can be elaborated, as in Figure 4.8(b), by replacing $D \dashleftarrow\!\dashrightarrow Y$ with a fully articulated back-door path $D \leftarrow S \rightarrow Y$, then the graph becomes a full causal model. Observation of S ensures that the conditioning strategies of the prior three chapters can be used to generate consistent estimates of the causal effect of D on Y. In this scenario, selection is on the observables – the variables in S – and the remaining treatment assignment mechanism is composed only of random variation that is ignorable.

If, in contrast, as shown in Figure 4.9(b) rather than Figure 4.8(b), only a subset Z of the variables in S is observed, then selection is on the unobservables because some components of S are now embedded in U. Conditioning on Z in this graph leaves unblocked back-door paths represented by $D \leftarrow U \dashleftarrow\!\dashrightarrow Y$ untouched. And, as a result, any observed association between D and Y within strata defined by Z cannot be separated into the genuine causal effect of D on Y and the back-door noncausal association between D and Y that is generated by the unobserved determinants of selection, U.

As explained in Chapter 4, and then as shown in demonstrations in Chapters 5 through 7, the concepts of ignorability and selection on the observables are a bit more subtle when potential outcomes are introduced. Weaker forms of conditional independence can be asserted about the joint distribution of Y^1, Y^0, D, and S than can be easily conveyed as in Figures 4.8 and 4.9, in which the observable variable Y is depicted instead of the underlying potential outcome variables Y^1 and Y^0. For example, the average treatment effect for the treated (ATT) can be estimated consistently



(a) The identification puzzle (b) Coleman's approach

Figure 8.1 Coleman's strategy for the identification of the causal effect of Catholic schooling on achievement.

by asserting only that Y^0 is independent of D conditional on S, even though full ignorability does not hold; see the discussion of Equation (4.4) and Section 4.3.3.[1] These possibilities are not fully revealed by conventional directed graphs, given that for these graphs potential outcome variables are typically not represented (see the discussion of Figure 3.9).

8.2 Selection on the Unobservables and the Utility of Additional Posttreatment Measures of the Outcome

In this section, we present the challenges of modeling causal effects when selection is on unobservables, using as an example the regression equations offered by James Coleman and his colleagues for the estimation of the causal effect of Catholic schooling on high school achievement. The basic estimation challenge that Coleman and his colleagues confronted is depicted in Figure 8.1(a). Although somewhat simplified, the graph represents the basic types of causal variables that Coleman and his colleagues contemplated in their original analyses and then in debate with their critics (Coleman et al. 1982; see also our prior discussion in Section 6.6.2). For Figure 8.1(a), Y_{10} is an observed score on a standardized achievement test given in the tenth grade to second-year high school students. The variable D is again an observed dichotomous causal variable equal to 1 for those who attend Catholic schools and 0 for those who attend public schools. A primary goal of the original research was to estimate the causal effect of D on Y_{10}, conceptualized as the average causal effect of Catholic schooling on achievement.

The remaining variables in the graph represent the basic types of variables that Coleman and his colleagues decided should be specified as adjustment variables in

[1] Furthermore, as discussed in Section 5.2.1 in the presentation of matching techniques, one need not assume full conditional independence of Y^0 in order to offer consistent and unbiased estimates of the ATT. An equality of the two conditional expectations, $E[Y^0|D=1,S]=E[Y^0|D=0,S]$, will suffice; see discussion of Assumption 2-S in Equation (5.2).

their regression equations. The variables in X are determinants of achievement test scores that have no direct causal effects on school sector selection. The variables in O are ultimate background factors that determine all other variables in the graph (i.e., X, U, D, and Y_{10}). Coleman and his colleagues designated many variables that they believed were contained in O and X, although they were unsure of whether to consider them as members of O or X. Finally, the variables in U are the crucial variables that Coleman and his colleagues recognized were probably unobserved. These variables – intrinsic motivation to learn, subtle features of the home environment, and anticipation of the causal effect itself – were thought to determine both school sector choice and achievement. Because these variables were assumed to be unobserved, they are represented collectively in Figure 8.1(a) by the variable U with a node that is a hollow circle $\circ$.

In the initial research, the analysis strategy of Coleman and his colleagues was to condition on observed variables in O and X in order to attempt to remove as much of the confounding of $D \to Y_{10}$ as possible. Reinterpreted with the aid of directed graphs, to identify the causal effect they needed to block five separate back-door paths from D to Y_{10}:

1. $D \leftarrow U \rightarrow Y_{10}$,
2. $D \leftarrow U \leftarrow O \rightarrow Y_{10}$,
3. $D \leftarrow U \leftarrow O \rightarrow X \rightarrow Y_{10}$,
4. $D \leftarrow O \rightarrow Y_{10}$, and
5. $D \leftarrow O \rightarrow X \rightarrow Y_{10}$.

According to the back-door criterion, conditioning on the variables in O and X would block paths 2, 3, 4, and 5 but would leave path 1 unblocked. Coleman and his colleagues recognized that the variables in U rendered the causal effect of D on Y formally unidentified, and they could not convince their critics that their estimates from models that conditioned only on O and X were sufficiently close to the ATE (or the ATT) to offer the conclusion that they did.

Their solution to this predicament is presented in Figure 8.1(b), which became possible when the second round of survey data was released two years later (see Coleman and Hoffer 1987; Hoffer et al. 1985). Rather than focus on explaining variation in the tenth grade test score variable, Y_{10}, they reclassified it as a conditioning variable and instead designated the new twelfth grade test score variable, Y_{12}, as the outcome of interest. They then argued that Y_{10} could serve to screen off the effects of the variables in U on Y_{12}. In particular, their strategy was equivalent to asserting that conditioning on O, X, and Y_{10} blocks all five of the back-door paths between D and Y_{12} in Figure 8.1(b):

1. $D \leftarrow U \rightarrow Y_{10} \rightarrow Y_{12}$,
2. $D \leftarrow U \leftarrow O \rightarrow Y_{10} \rightarrow Y_{12}$,
3. $D \leftarrow U \leftarrow O \rightarrow X \rightarrow Y_{10} \rightarrow Y_{12}$,

4. $D \leftarrow O \rightarrow Y_{10} \rightarrow Y_{12}$, and
5. $D \leftarrow O \rightarrow X \rightarrow Y_{10} \rightarrow Y_{12}$.

Indeed, all of these back-door paths are blocked by conditioning on Y_{10}, and four of them are blocked by supplemental but unnecessary conditioning on O and X as well.[2]

However, notice that Y_{10} does not satisfy Condition 2 of the back-door criterion. It lies on a directed path that begins at the causal variable and reaches the outcome variable, $D \rightarrow Y_{10} \rightarrow Y_{12}$. Conditioning on Y_{10} adjusts away some of the total causal effect of Catholic schooling on achievement measured in the twelfth grade. As a result, by adopting this analysis strategy, Coleman and his colleagues changed the average causal effect of interest from the total effect of Catholic schooling on achievement to a net direct effect, $D \rightarrow Y_{12}$. This effect is best thought of as the average gain in achievement between the tenth and twelfth grades attributable to Catholic schooling, although the specific interpretation depends on the model that is estimated.

Holding aside the fundamental issue of whether narrowing the focus to this particular net direct effect is helpful for the research questions at hand, consider now the graph presented in Figure 8.2, which suggests two more basic criticisms of this approach. The first was seized on by their critics: the variables in U that confound the causal effect of D on Y_{12} are not effectively screened off by Y_{10}. More generally, Figure 8.2 should be seen as more plausible than Figure 8.1(b) because it is unreasonable to rule out three direct effects: $O \rightarrow Y_{12}$, $X \rightarrow Y_{12}$, and $U \rightarrow Y_{12}$ (or, as Coleman and colleagues asserted, that these effect are indirect and transmitted through $Y_{10} \rightarrow Y_{12}$). Because O and X are observed, the claims that $O \rightarrow Y_{12}$ and $X \rightarrow Y_{12}$ should be included in the graph can be accommodated in subsequent analysis. However, the inclusion of $U \rightarrow Y_{12}$ renders back-door adjustment with the observed variables ineffective. The rationale for the inclusion of $U \rightarrow Y_{12}$ in the graph is the following. If the parents of highly motivated students were more likely to pay tuition to enroll in Catholic schools, then enhanced motivation would contribute directly to learning in both the tenth grade and the twelfth grade. Similarly, it is unreasonable to assume that the motivation that is correlated with willingness to pay tuition would exert only a one-time boost early in the Catholic school careers of students, which would then structure subsequent achievement only by way of a twelfth grade knock-on effect from the initial boost in achievement in the tenth grade. Taken together, the critics argued, in effect, that Coleman and his colleagues mistakenly ignored a back-door path, $D \leftarrow U \rightarrow Y_{12}$, that remains unblocked after conditioning on O, X, and Y_{10} because they failed to acknowledge that U causes Y_{12} directly.

The second criticism is more subtle. For Figure 8.2, we have omitted the effect $Y_{10} \rightarrow Y_{12}$ that Coleman and his colleagues seem to have asserted for their models, and which we depicted in Figure 8.1(b). Whether this effect should be included in the graph depends on the richness of the variables in O, X, and U. If these variables, when combined with D, constitute full models of achievement for public and Catholic school students, then no direct relationship between Y_{10} and Y_{12} needs to be represented in

[2] Coleman and his colleagues sometimes wrote as if they allowed for effects such as $O \rightarrow Y_{12}$ and $X \rightarrow Y_{12}$. The back-door paths generated by these additional direct effects would all be blocked by conditioning on O, X, and Y_{10}. The key assumption is their implicit claim that U does not have an effect on Y_{12}, except through Y_{10}.

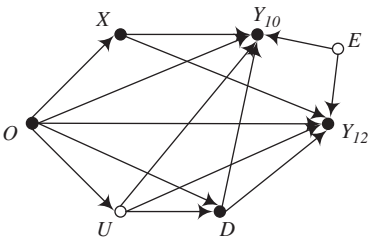


Figure 8.2 Criticism of Coleman's estimates of the effect of Catholic schooling on learning.

the graph. The critics did not take a position on this issue, but we suspect that they would not have been willing to maintain that the measured variables in O and X were rich enough to eliminate the need for including $Y_{10} \rightarrow Y_{12}$ in such a graph. Instead, they would have likely allowed for such an effect so that the causes embedded in $e_{Y_{10}}$ are allowed to contribute to Y_{12} via Y_{10}. The critics only emphasized the implausibility of interpreting models under the faulty assumption that $D \leftarrow U \rightarrow Y_{12}$ does not exist, and this point of criticism stands regardless of what position one takes on the plausibility of $Y_{10} \rightarrow Y_{12}$.

Yet, once one begins to think about the relationship between Y_{10} and Y_{12}, it may seem implausible to assume that these variables are only associated because of their common dependence on O, X, U, and D. Is it reasonable to assume that the structural error terms $e_{Y_{10}}$ and $e_{Y_{12}}$, which are viewable only under magnification, are independent of each other? For Figure 8.1(b), they were assumed to be so. For Figure 8.2, instead it is assumed that they are only independent of each other after we allow for unobserved common direct causes of both Y_{10} and Y_{12}, which are represented by E in Figure 8.2. Although the common cause relationship $Y_{10} \leftarrow E \rightarrow Y_{12}$ is quite general, for now consider E to be an unobserved and completely random characteristic of students that measures their taste for tests. Independent of all else, some students enjoy taking tests and perform well on them.[3] If such a common cause of Y_{10} and Y_{12} exists, it creates four new back-door paths from D to Y_{12} through E:

1. $D \leftarrow U \rightarrow Y_{10} \leftarrow E \rightarrow Y_{12}$,
2. $D \leftarrow O \rightarrow Y_{10} \leftarrow E \rightarrow Y_{12}$,
3. $D \leftarrow U \leftarrow O \rightarrow X \rightarrow Y_{10} \leftarrow E \rightarrow Y_{12}$,
4. $D \leftarrow O \rightarrow X \rightarrow Y_{10} \leftarrow E \rightarrow Y_{12}$.

For these back-door paths, Y_{10} is a collider. In the absence of conditioning, all four of these paths are blocked and do not create additional noncausal associations between D and Y_{12}. But, when O, X, and Y_{10} are used as adjustment variables, the first path

[3]Of course, if this particular variable in E exists, it is unlikely to be independent of both O and X. Were we to add edges to also specify $O \rightarrow E$ or $X \rightarrow E$ as additional causal effects in Figure 8.2, all of what is written in the main text would be the same. These additional back-door paths would be blocked by conditioning on O and X, but the same back-door path, $D \leftarrow U \rightarrow Y_{10} \leftarrow E \rightarrow Y_{12}$, that is unblocked after conditioning on Y_{10} would remain.

becomes unblocked because conditioning on Y_{10} induces an association between U and E (see Chapter 4). The second and third back-door paths remain blocked by the simultaneous conditioning on the observed variables O and X, even though conditioning on Y_{10} induces an association between O and E as well as between X and E. In short, if Figure 8.2 is the appropriate representation of the causal system, Coleman and his colleagues unblocked an already blocked back-door path, thereby creating a new net association between D and Y_{12}. This induced noncausal association is then mixed in with the noncausal association generated by $D \leftarrow U \rightarrow Y_{12}$, which critics argued had been mistakenly assumed not to exist. The following demonstration, which considers the simplest possible causal model consistent with Figure 8.2, explains these complications together, while also leading into the explicit consideration of heterogeneity that we will take up in the section that follows it.

Panel Data Demonstration 1

For this simulated example of the analysis of Catholic school effects by Coleman and his colleagues, we will use the panel data notation introduced in Section 2.8. Nonetheless, because we will consider only posttreatment data on the outcome and because our other variables are fixed in time, we can suppress some of the subscripting and other fine distinctions introduced there.[4] In order to focus on the essential identification issues, we consider only linear specifications and, except when noted otherwise, restrict all causal effects to be constants or to vary across individuals in completely random ways independent of all else in the models. These conditions are consistent with the original research by Coleman and his colleagues, even though they are inconsistent with our extensive treatment of this research question using the more recent literature (as shown in Chapters 5 through 7, where nonlinearities and heterogeneity of effects were both shown to be vital considerations when modeling the Catholic school effect on achievement).

For this demonstration, the potential outcome variables are Y_t^1 and Y_t^0, where $t = \{10, 11, 12\}$ for the three grades that occur during the observation window from the tenth grade through the twelfth grade. Because treatment selection occurs before $t = 10$, we have observed data only for posttreatment time periods. The treatment indicator variable, D, is equal to 1 if a student is enrolled in a Catholic school and 0 if enrolled in a public school. (We again ignore other types of private schools.) The observed outcome variable, Y_t, is defined as $DY_t^1 + (1 - D)Y_t^0$ for $t = \{10, 11, 12\}$.

For simplicity, our other variables O, X, and U should be interpreted as indices of many underlying variables, which we will scale as normally distributed composite variables. Consistent with our discussion of Figures 8.1 and 8.2, X is a composite determinant of achievement test scores, Y_t, that has no direct effect on whether students select Catholic schooling, D. U is a composite variable of unobserved factors that determines both D and Y_t. And O is a composite variable of ultimate background

[4]Most importantly, for this demonstration we do not need to make a distinction between D_{it} (the treatment exposure indicator) and D_i^* (the treatment group indicator). Because we have no pretreatment observations and we ignore (as did Coleman and his colleagues) students who switch between public and Catholic schools between the tenth and twelfth grade, $D_{it} = D_i^*$ for all t. We therefore use our customary D without subscripts for either i or t.

factors that determines U, X, D, and Y_t. To give these composite variables distributions that are familiar, O is a standard normal random variable with mean of 0 and variance of 1. Having defined O as an exogenous root variable, we then set $X = O + e_X$ and $U = O + e_U$, where e_X and e_U are independent standard normal random variables with mean of 0 and variance of 1.

We consider four data setup scenarios, defined as a cross-classification of two binary conditions: (1) the presence of self-selection on accurate expectations of individual-level effects of Catholic schooling and (2) the presence of supplemental dependence between the potential outcomes that is produced by an exogenous variable E (see Figure 8.2). For all scenarios, the probability of Catholic school enrollment is specified as a logistic distribution

$$\Pr[D = 1|O,U] = \frac{\exp(-3.8 + O + U)}{1 + \exp(-3.8 + O + U)}, \tag{8.1}$$

where O and U are as defined above. The probabilities defined by Equation (8.1) are then specified as the parameters for draws from a binomial distribution, yielding the indicator variable D for Catholic schooling defined above. As explained below, we introduce self-selection into two of the four scenarios by allowing U to structure the potential outcomes, such that those with higher values for U have higher average individual-level treatment effects. Because of the presence of U in Equation (8.1), students with higher values for U are then more likely to enter Catholic schooling and more likely to benefit from doing so.

For the scenarios without supplemental dependence on E, Y_t^0 is defined as

$$\begin{aligned} Y_{10}^0 &= 100 + O + U + X + \upsilon_{10}^0, \\ Y_{11}^0 &= 101 + O + U + X + \upsilon_{11}^0, \\ Y_{12}^0 &= 102 + O + U + X + \upsilon_{12}^0, \end{aligned} \tag{8.2}$$

where the υ_t^0 are independent normal random variables with mean of 0. For the scenarios with supplemental dependence, Y_t^0 is instead defined as

$$\begin{aligned} Y_{10}^0 &= 100 + O + U + X + E + \upsilon_{10}^0, \\ Y_{11}^0 &= 101 + O + U + X + E + \upsilon_{11}^0, \\ Y_{12}^0 &= 102 + O + U + X + E + \upsilon_{12}^0, \end{aligned} \tag{8.3}$$

where E is also a normal random variable with mean of 0.[5] On average, Y_t^0 follows a linear time path as determined by the intercept values of 100, 101, and 102. However, the levels of these potential outcomes for individuals are set by the time-invariant values of O, U, X, and E as well as by time-specific shocks to their outcomes, υ_t^0.

[5]To yield a covariance structure similar to real test score data, for Equation (8.2) we specify υ_t^0 as time-period-specific standard normal random variables multiplied by 10. For Equation (8.3), we specify E as a standard normal random variable multiplied by 5 and υ_t^0 as time-period-specific standard normal random variables multiplied by 8.6.

To specify a treatment effect that increases linearly in time, Y_t^1 is defined as

$$\begin{aligned} Y_{10}^1 &= Y_{10}^0 + \delta' + \delta'', \\ Y_{11}^1 &= Y_{11}^0 + (1 + \delta') + \delta'', \\ Y_{12}^1 &= Y_{12}^0 + (2 + \delta') + \delta'', \end{aligned} \tag{8.4}$$

where δ' is a baseline individual-level causal effect, specified as a normal random variable with mean of 10 and variance of 1. The constitution of δ'' depends on whether self-selection is present. In the no-self-selection scenarios, δ'' is additional random individual-level variation, specified as a standard normal random variable with mean of 0 and variance of 1. In the self-selection scenarios, δ'' varies systematically over individuals. The values of δ'' are set as single draws from individual-specific standard normal random variables with expectation equal to each individual's realized value u_i of U.

The first panel of Table 8.1 presents the ATE, ATT, and ATC for the effect of Catholic schooling on test scores in the tenth and twelfth grades, as implied by the data setup just detailed. For the first two columns, students do not self-select into Catholic schooling based on accurate expectations of the individual-level gains from doing so. Accordingly, the ATE, ATT, and ATC are all the same, and they all increase from 10.00 to 12.00 between the tenth and twelfth grades, as determined by Equation (8.4). For the last two columns, students self-select on the causal effect, as explained above. The ATT is larger than the ATE and the ATC. But, again, the average gain is 2.00 for the ATE, ATT, and ATC.

The second panel of Table 8.1 presents estimates for the coefficient on D from a series of regression models, first for cross-sectional estimators that use outcome data from only one point in time and then, as in the research of Coleman and colleagues, for panel data models that use outcome data from two posttreatment points in time. These latter regression models, as we will explain in Chapter 11, are labeled panel data analysis of covariance models. We use generic ordinary least squares (OLS) regression estimation, as in the original research of Coleman and his colleagues, but we have given OLS estimation a best case data setup where linear specifications are reasonable and the pattern of individual-level heterogeneity is simple.

Consider first the results for the regression models in the no-self-selection scenarios. For the first column, we also do not assume that E is a common cause of both Y_{10} and Y_{12}, which does not render Y_{10} a collider on any back-door paths from D to Y_{12}. For the second column, E is a common cause. This distinction for the data setup produces no differences for the cross-sectional regression estimates of the coefficient on D because Y_{10} is not in the conditioning set for these models.

For both sets of cross-sectional estimators in the no-self-selection scenarios, the same pattern of results is present for models that have Y_{10} as the outcome variable and for models that have Y_{12} as the outcome variables. The naive estimator is too large, and adjustment for O and X eliminates only some of the confounding.[6]

[6] Additional adjustment for U would yield estimates that are equal to the ATE, ATT, and ATC. This last result is merely a benchmark, given that we have assumed that U is unobserved, as in the research of Coleman and his colleagues.

Table 8.1 Simulated Results for the Identification Approach Adopted by Coleman and Colleagues

Setup conditions:				
Self-selection on the causal effect	No	No	Yes	Yes
E is a common a cause of Y_t^0	No	Yes	No	Yes
True average treatment effects:				
ATE, tenth grade	10.00	10.00	10.00	10.00
ATT, tenth grade	10.00	10.00	11.85	11.85
ATC, tenth grade	10.00	10.00	9.81	9.81
ATE, twelfth grade	12.00	12.00	12.00	12.00
ATT, twelfth grade	12.00	12.00	13.85	13.85
ATC, twelfth grade	12.00	12.00	11.81	11.81
	Estimated Coefficient on D			
Cross-sectional estimators:				
Regression of Y_{10} on D	14.75	14.75	16.60	16.60
Regression of Y_{10} on O, X, and D	10.80	10.80	12.58	12.58
Regression of Y_{12} on D	16.75	16.75	18.60	18.60
Regression of Y_{12} on O, X, and D	12.80	12.80	14.58	14.58
Panel data analysis of covariance estimators:				
Regression of Y_{12} on Y_{10}, O, X, and D	12.68	9.98	14.41	11.27

For the analysis of covariance models with Y_{12} as the outcome variable, the coefficient estimates on D do not equal the ATE, ATT, or ATC for the twelfth grade, which are all 12.00. The coefficient estimate of 12.68 is too large when E is not a common cause because the unblocked back-door path through U biases the coefficient upward. The estimate is then too small when E is a common cause because the upward bias from the unblocked back-door path through U is then joined by a larger downward bias produced by conditioning on Y_{10}, which is a collider along the new back-door paths through E (see the discussion that precedes the demonstration for an accounting of these back-door paths).

As shown in the third and fourth columns, the basic pattern for the estimated models changes only modestly when self-selection is present. The last two columns are analogous to the first two, after allowing the self-selection term, δ'', to vary with U, as explained above for Equation (8.4). For these last two scenarios, none of the cross-sectional estimators yield estimates equal to any of the average causal effects, either in the tenth grade or the twelfth grade.[7] The analysis of covariance models have

[7]Even models that adjust for U would not suffice. For the self-selection scenarios, the individual-level effect of Catholic schooling varies with U, such that those with higher average values of U have

the same pattern as in the no-self-selection scenarios. When E is not a common cause of Y_{10} and Y_{12}, the estimate is larger than the ATE, ATT, and ATC. When E is a common cause, rendering Y_{10} a collider, the estimate is smaller than the ATE, ATT, and ATC.

Finally, notice also that the analysis of covariance estimates are nowhere near to the average gain in achievement between the tenth and twelfth grades, which is equal to 2.00, and which by construction is equal for both Catholic and public school students and does not depend on whether self-selection is present. As such, under this setup, which is consistent with Figure 8.2, the analysis of covariance models do not estimate a net direct effect of Catholic schooling on achievement in the twelfth grade that can be given an interpretation as an estimate of an average gain in achievement.

Many analysis puzzles and ensuing causal controversies in the social science literature have the same basic features as this demonstration. The plausibility of the criticism in Figure 8.2 should serve as a caution because it demonstrates how additional posttreatment observations of the outcome variable are unlikely to resolve the fundamental identification challenge created by selection on unobservables and nonignorability of treatment assignment. We consider models of this type in detail in Chapter 11, including a more complete discussion of the sorts of processes that generate over-time dependencies, such as the one produced by E in Figure 8.2. However, we also have some more encouraging results to offer in Chapter 11. If, as was the case for this demonstration, the causal effect is evolving in time, such that the trajectory for the outcome in the absence of treatment has the same time path before and after treatment exposure, then it is possible to obtain consistent estimates of some average treatment effects of interest.

Before offering a full explanation of analysis strategies when data are available from pretreatment time periods, we will consider two other types of estimators – instrumental variable estimators in Chapter 9 and mechanism-based estimators in Chapter 10. To motivate our presentation of these methods in subsequent chapters, we conclude this chapter in the next section by enriching our presentation of directed graph methodology, explaining how graphs can be used to represent full patterns of self-selection and heterogeneity using latent class variables.

---

larger treatment effects. As a by-product, the individual-level causal effect also varies with O and X, given that these variables are positively associated with U because of the common-cause path $U \leftarrow O \rightarrow X$. The conditional-variance weighting that is implicit in OLS regression averages these heterogeneous individual-level effects, giving more weight to those whose implicit propensity scores are near to .5. Given the data setup, where only 9 percent of students end up enrolled in Catholic schools – because of the intercept of -3.8 for the logistic distribution in Equation (8.1) – the implicit weighting would move the estimated average effects toward the ATT relative to the ATC. Yet, the coefficient estimates would not equal the ATT. As explained in Section 6.3, these coefficients are best interpreted as estimates of the ATE with supplemental conditional-variance-based weighting, which, for this demonstration, would push the estimates toward the ATT because of the distribution of D and the positive association between U and D. In practice, of course, the variable U would not be available for analysis in the first place.

8.3 Causal Graphs for Complex Patterns of Self-Selection and Heterogeneity

We conclude this chapter with a presentation of how patterns of self-selection and heterogeneity can be represented with directed graphs.[8] We have two goals in this section. First, we want to demonstrate and explain the generality of the graphical approach to the representation of causal relationships, which we will utilize in subsequent chapters. Second, we want to make the following obvious (but frequently forgotten) point: The best solution to an unobserved variable problem is to develop and deploy additional new measures. Too often, the literature implies that self-selection patterns are so complex that estimation in their presence will forever remain infeasible. We see more promise in developing methods to directly confront the limitations of available data than the literature often implies, and we want to make this case explicitly before offering explanations in the next three chapters of estimation strategies that may be feasible while the limitations of available data remain unaddressed.

In this section, we use the charter school effect on learning as the focal example, which we introduced in Section 1.3.2 (see page 24). This example shares some of the same complications as the example of the Catholic school effect that we have considered extensively already, but it offers some new complications as well (and serves as a bridge to the school voucher instrumental variable example that will be considered in Chapter 9). In the material that follows, we start with a simple latent class model of heterogeneity that allows families of different types to differ in their likelihood of choosing charter schools. We then build toward a full directed graph that shows the estimation challenge clearly, at which point we will then elaborate the back-door paths in order to discuss feasible estimation strategies that would be possible if new measures were constructed that would enable direct modeling of choice behavior. We conclude with a graph-aided interpretation of extant empirical research on charter school effects.

8.3.1 A Starting Point: Separate Graphs for Separate Latent Classes

Consider the two causal graphs in Figure 8.3, and suppose that the population is partitioned into two latent classes, each of which has its own graph in panel (a) or panel (b). Suppose that P is a family's parental background, D is charter school attendance, and Y is a score on a standardized test given to all fifth graders in a large metropolitan school district. For these two graphs, the subscripts refer to latent classes, which are also indicated by a latent class membership variable G that takes on values of 1 and 2.[9]

Although surely a gross oversimplification, suppose nonetheless that the population is composed of fifth graders who have been raised in two types of families. Families with $G = 1$ choose schools predominantly for lifestyle reasons, such as proximity to the home and tastes for particular school cultures, assuming that all schools are similar

[8] This section draws on material previously published in Morgan and Winship (2012).

[9] The lower case values x, d, and y for the two causal graphs are meant to connote that these are realized values of X, D, and Y that may differ in their distributions across the two latent classes.

(a) $G = 1$ (b) $G = 2$

Figure 8.3 Separate causal graphs for two groups of individuals ($G = 1$ and $G = 2$) where the effects of parental background (P) and charter schools (D) on test scores (Y) may differ for the two groups.

in instructional impact because achievement is largely a function of individual effort. Families with $G = 2$ choose elementary schools for their children by selecting the school, subject to constraints, that they feel will maximize the achievement of their children, assuming that schools differ in quality and that their children may learn more in some schools than in others. Accordingly, these families are attentive to the national press coverage of educational policy, in which many policymakers have argued for increasing the number of charter schools in the country because some research has claimed that charter schools are more effective than traditional public schools. As a consequence, the second group of families is more likely to send their children to charter schools, such that the mean of D is higher for those families with $G = 2$ than $G = 1$.

Finally, suppose that parents with higher levels of education are more likely to value distinctive forms of education, and as a result are more likely to send their children to charter schools (independent of whether or not highly educated parents are more likely to be found in the latent class for whom $G = 2$, which we will discuss later). They are also more likely to be able to support children in completing homework and otherwise making the most of the educational opportunities that are offered to their children. Accordingly, suppose that in both groups the causal effects $P \rightarrow D$ and $P \rightarrow Y$ are positive and substantial (i.e., that α_1, β_1, α_2, and β_2 in Figure 8.3 are positive and substantial).

The question for investigation is whether the effect of D on Y is positive for both groups and, if so, whether it is the same size for both groups. If one is willing to assume, as some of the literature suggests, that the second group of families is correct in the sense that school quality does matter for student learning, and further that charter schools are higher quality, then we should expect that both δ_1 and δ_2 are more likely positive than not. And, if one believes that parents with $G = 2$ have some sense that this is correct, then not only will more of them send their children to charter schools, they will also sort their children more effectively into charter and non-charter schools. In other words, they will also be more likely to continue to enroll their children in regular public schools if they feel that their children will not benefit from the distinctive characteristics of available charter schools (e.g., if the charter schools that have openings have instructional themes that their children find distasteful). Because both of these self-selection effects are reinforcing, is it likely that $\delta_2 > \delta_1$.[10]

[10] These target parameters, δ_2 and δ_1, are defined implicitly as the average effect of charter schooling for all students from families with $G = 2$ and $G = 1$, respectively.

If this plausible scenario is true in reality, what would happen if a researcher ignored the latent classes (either by mistake or, more realistically, because the membership variable G is unobserved) and simply assumed that a single graph prevailed? In this case, a researcher might estimate the effect of D on Y for each value of P and then average these effects over the distribution of P, yielding a population-level estimate δ. At best, this estimate would be uninformative about the underlying pattern of heterogeneity that suggests that $\delta_2 > \delta_1$. At worst, this estimate would be completely wrong as an estimate of the average causal effect of D on Y, which we have referred to as the ATE in prior chapters. For example, if P predicts latent class membership G, and G predicts the size of the effect of D on Y, then P-stratum-specific effects mix together individual-level causal effects that vary with the conditional distribution of G within the strata of P. Combining P-stratum-specific effects by calculating an average effect across only the distribution of P does not properly weight the G-stratum-specific effects that are embedded in differential patterns within the strata of P.

In order to consider these possibilities, we need to have a model of selection into D that is informed by a model of the traits of families that would cause them to be found in underlying latent classes. It is most natural to pursue such a model in a single causal graph that explicitly represents the latent classes by including the variable G as a node within it.

8.3.2 A Single Graph That Represents All Latent Classes

Consider Figure 8.4, which contains a standard causal triangle where D has an effect on Y but where both D and Y share a common cause P. To this triangle, the latent class membership variable G is introduced as an explicit cause of D. The variable G is given a hollow node, ∘, to indicate that it is unobserved.[11]

The arrow from G to D is present because there are alternative groups of families, coded by the alternative values of the unobserved variable G, that approach differently the decision of whether to send their children to charter schools. As a result, G predicts charter school attendance, D. Although we will continue to write as if G only takes on two values that identify two latent classes, this restriction is no longer necessary. G may take on as many values as there are alternative groups of families who differ systematically in how they approach and enact the decision of whether to send their children to a charter school. (We considered only two values for G in Figure 8.3 to limit the number of G-specific graphs that needed to be drawn.)

The corresponding structural equations for the graph in Figure 8.4 are then

$$P = f_P(e_P), \tag{8.5}$$
$$G = f_G(e_G), \tag{8.6}$$
$$D = f_D(P, G, e_D), \tag{8.7}$$
$$Y = f_Y(P, D, e_Y). \tag{8.8}$$

[11] Here, we follow Elwert and Winship (2010) and introduce a latent class variable G to represent effect heterogeneity.

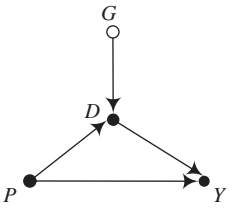


Figure 8.4 A graph where groups are represented by an unobserved latent class variable (G) in a single graph.

The latent class membership variable G only enters these structural equations in two places, on its own in Equation (8.6) and then as an input to $f_D(.)$ in Equation (8.7). Recall also that the unspecified structure of the functions such as $f_Y(.)$ permits the sort of interactions discussed in the last subsection, where, for example, the effect of D depends on the level of P. (See Section 3.3.2 for a full discussion.)

To accept Figure 8.4 as a full representation of the causal relationships between G and the other variables P, D, and Y, we must be able to assume that G shares no common causes with P, D, or Y that have been mistakenly suppressed. The necessary assumptions are that students who have parents with higher levels of education are no more likely to know of the educational policy dialogue that claims that charter schools have advantages and also are no more likely to think that school quality has any effects on achievement. We must also be willing to assume that, within values of P and D, G has no causal effect on Y, which with our example is tantamount to assuming that those who attempt to maximize the learning of their children by selecting optimal schools (1) do not manage to do so well enough so that the obtained effect is any larger on average, conditional on other factors, for their own children than for those who do not attempt to select on the causal effect of schooling and (2) do not do anything else that helps their children to benefit from the learning opportunities provided to them in school or in the home that is not already captured by the direct effect $P \rightarrow Y$. This would require that the impulse to select into charter schools based on beliefs about the size of the charter school effect for one's own child is a completely ignorable process, since it does not result in any actual selection on the variation in the causal effect nor generate any reinforcing behavior that might complement the charter school effect. None of the literature on charter school effects supports such a dismissal of the power of self-selection.

Accordingly, for Figures 8.5(a) and 8.5(b), we add an arrow from G to Y to the graph presented earlier in Figure 8.4. For Figure 8.5(a), which includes only this one additional arrow, the structural equations are

$$P = f_P(e_P), \tag{8.9}$$
$$G = f_G(e_G), \tag{8.10}$$
$$D = f_D(P, G, e_D), \tag{8.11}$$
$$Y = f_Y(P, D, G, e_Y). \tag{8.12}$$

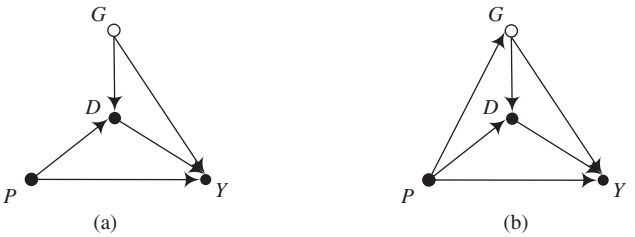


Figure 8.5 Two graphs where selection into charter schools (D) is determined by group (G) and where selection renders the effect of D on Y unidentified as long as G remains unobserved.

For Figure 8.5(b), we drop the assumption that G is independent of P. This elaborated graph now includes an arrow from P to G. As a result, $f_G(e_G)$ is no longer the appropriate function for G. Equation (8.10) must be replaced by

$$G = f_G(P, e_G) \tag{8.13}$$

so that family background is an explicit cause of latent class membership. It is likely that parents with high socioeconomic status are more likely to select on the possible causal effect of charter schooling, which is how the latent classes were discussed for Figure 8.3. Still, how these latent classes emerge is not sufficiently transparent in Figure 8.5. A more explicit causal model that gives structure to the causal pathway from P to G may help to clarify the self-selection dynamic, as we show next.[12]

8.3.3 Self-Selection into the Latent Classes

Suppose that the latent class membership variable G is determined at least in part by a variable that measures a family's subjective expectation of their child's likely benefit from attending a charter school instead of a regular public school. Although we could enter this variable into a graph with a single letter, such as S or E, for Figure 8.6 we use a full mnemonic representation as a variable labeled $\text{Exp}(D \rightarrow Y)$, which represents "the family's subjective expectation of the specific effect of D on Y for their particular child." For Figure 8.6(a), which is a direct analog to Figure 8.5(a), this subjective

[12] In some work, latent class membership variables such as G are not regarded as causal variables. We regard G as a nominal causal variable because we can conceive of an intervention that could move a family from one value of G to another. The goal of introducing G into this graph is to provide a representation of consequential individual-level heterogeneity, and accordingly we allow G to have nominal causal effects on the outcome of interest through $G \rightarrow Y$. We will explain later why a fully elaborated graph that locates the sources of all such heterogeneity in structural variables that lie on directed paths that reach G would thereby render G as a redundant carrier of these fully specified determinants of heterogeneity. In this case, G could then be removed from the graph, as there would be no need for a nominal causal variable to transmit the heterogeneity produced by its genuine determinants.

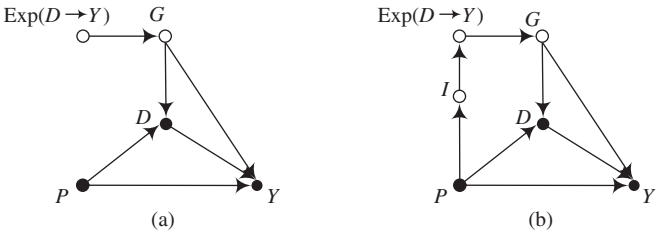


Figure 8.6 Two graphs where selection on the unobservables is given an explicit representation as self-selection on subjective expectations of variation in the causal effect of D on Y. For panel (b), these expectations are determined by information (I) that is differentially available to families with particular parental backgrounds (P).

expectation is the sole determinant of G. The structural equations are then

$$P = f_P(e_P), \tag{8.14}$$
$$\text{Exp}(D \to Y) = f_{\text{Exp}}(e_{\text{Exp}}), \tag{8.15}$$
$$G = f_G[\text{Exp}(D \to Y),\, e_G], \tag{8.16}$$
$$D = f_D(P, G, e_D), \tag{8.17}$$
$$Y = f_Y(P, D, G, e_Y). \tag{8.18}$$

Note that $\text{Exp}(D \to Y)$ is determined solely by e_{Exp} in Equation (8.15). Thus, the graph in Figure 8.6(a) would be an accurate representation of the system of causal relationships if subjective expectations were either completely random or instead based solely on characteristics of families that are independent of the family background variables in P.

Given what we have written about the likelihood that families with different patterns of P will end up in different latent classes represented by G, it seems clear that Figure 8.6(a) is not the proper representation for the research scenario we have already specified. Accordingly, in Figure 8.6(b), e_{Exp} is joined by an unspecified additional input I into the subjective expectation of the child-specific causal effect of D on Y, which is then presumed to be caused, in part, by family background. As a result, there is now a path from P to G through I and $\text{Exp}(D \to Y)$. The structural equations are now augmented as

$$P = f_P(e_P), \tag{8.19}$$
$$I = f_I(P, e_I), \tag{8.20}$$
$$\text{Exp}(D \to Y) = f_{\text{Exp}}(I, e_{\text{Exp}}), \tag{8.21}$$
$$G = f_G[\text{Exp}(D \to Y),\, e_G], \tag{8.22}$$
$$D = f_D(P, G, e_D), \tag{8.23}$$
$$Y = f_Y(P, D, G, e_Y). \tag{8.24}$$

In sociology, the causal effect of P on $\text{Exp}(D \to Y)$ via I follows from the position that privileged positions in social structure are occupied by advantaged families. From these

positions, individuals acquire information I that allows them to recognize benefits that are available to them.[13]

The directed path $P \rightarrow I \rightarrow \text{Exp}(D \rightarrow Y) \rightarrow G$ carries one important systematic source of self-selection through to the latent class variable G. However, this path is not the only directed path that is present, as could be seen under magnification. The differences in information about the charter school effect that are independent of parental background have their effect on G through $e_I \rightarrow I \rightarrow \text{Exp}(D \rightarrow Y) \rightarrow G$. Likewise, differences in expectation formation processes that are not based on information and are independent of parental background have their effects on G through $e_{\text{Exp}} \rightarrow \text{Exp}(D \rightarrow Y) \rightarrow G$.

By building the full graph progressively from Figures 8.4 through 8.6(b), we have explicitly elaborated what is often presumed in models that incorporate self-selection. Background variables in P are related to the cause D by way of a set of latent classes in G that encode subjective evaluations of the individual-specific causal effect of D on Y. These expectations are functions, in part, of characteristics in P by way of the information I that is differentially available to families that differ on P. Yet, even though we now have an elaborate representation of self-selection, we still have not brought what some would label "contextual effects" into the model, such as neighborhood of residence. We consider this complication next, which is clearly important to consider when modeling the effects of charter schools on learning.

8.3.4 Self-Selection into the Treatment and a Complementary Context

How hard is the task of allowing for contextual effects? Consider Figure 8.7, which incorporates a contextual variable N into the causal graph in Figure 8.6(b). N represents all causes of student achievement that can be conceptualized as either features of a student's residential neighborhood or features of a charter school's surrounding neighborhood. The component variables in N might be access to resources not measured by P or specific local cultures that may or may not promote student achievement.[14]

[13] In addition, it may be that there are also additional common causes of P and I, which could be represented by a bidirected edge between P and I in the graph. This would be reasonable if informational advantages that structure expectations for optimal school choices are determined by deeper structural factors that also confer socioeconomic advantages on parents before they arrive at the decision point of whether or not to send their children to charter schools. It is also possible that parental background determines expectation formation processes to some degree, such that different families process information differently, as would be case if the directed path $P \rightarrow \text{Exp}(D \rightarrow Y) \rightarrow G$ were added to the graph. We do not include these additional causal pathways in the graph because they would add additional back-door paths to the subsequent discussion without changing the core identification results.

[14] If the latter are only diffuse cultural understandings that only weakly shape local norms about the appropriateness of enacting the role of achievement-oriented student, then such variables may be difficult to observe. In this case, N might then be coded as a series of neighborhood dummy identifier variables. Analysis of these effects would then only be possible if there were sufficient numbers of students to analyze from within each neighborhood studied. Without such variation, the potential effects of N could not be separated from individual characteristics of students and their families. And, if modeled in this way, only the total effects of N would be identified, since the dummy variables for N would not contain any information on the underlying explanatory factors that structure the neighborhood effects that they identify.

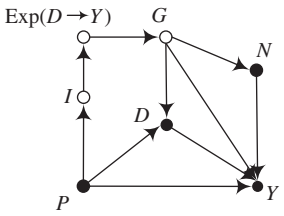


Figure 8.7 A graph where self-selection on the causal effect of charter schooling also triggers self-selection into consequential and interactive neighborhood contexts (N).

With the addition of N, the function for Y is now $f_Y(P,D,G,N,e_Y)$. Recall, again, that N is not restricted by any functional form assumption for $f_Y(.)$. As a result, the causal effect of N can modify or interact with the effects of G, D, and P on Y.[15]

Figure 8.7 also allows for even more powerful effects of self-selection. Suppose that self-selection into the latent classes in G is associated with self-selection into N as well. We see two separate and countervailing tendencies. Parents attuned to the potential benefits of charter schooling are also more likely to choose neighborhood contexts that best allow them to encourage their children to study hard in school. At the same time, after obtaining an attendance offer from a charter school, a family may also decide to move to an alternative neighborhood in the catchment area of a suboptimal regular public school, since attendance at such a school may no longer be a consideration in the family's residential decision. If either effect is present, then the function for N is equal to $f_N(G,e_N)$, and we then have seven structural equations as

$$P = f_P(e_P), \tag{8.25}$$
$$I = f_I(P,e_I), \tag{8.26}$$
$$\text{Exp}(D \to Y) = f_{\text{Exp}}(I,e_{\text{Exp}}), \tag{8.27}$$
$$G = f_G[\text{Exp}(D \to Y),\, e_G], \tag{8.28}$$
$$D = f_D(P,G,e_D), \tag{8.29}$$
$$N = f_N(G,e_N), \tag{8.30}$$
$$Y = f_Y(P,D,G,N,e_Y). \tag{8.31}$$

The nonparametric nature of these structural equations allows for fully interactive effects. Most importantly, the function for Y, $f_Y(P,D,G,N,e_Y)$, allows for the effects of D and N to vary within each distinct combination of values between them, as would be the case if the charter school effect varied based on the neighborhood within which students lived. We should also note that we could enrich the causal model further by drawing from the literature that posits deeper causal narratives for the joint determination of P and N, as well as other causal pathways that link P to N. We see the graph in Figure 8.7 as sufficiently rich to motivate the discussion we have offered so far as well as for the analysis of the extant empirical research to which we turn next.

[15] See VanderWeele (2009b) for an incisive analysis of the difference between an interaction and an effect modification. Our interest, conceptually at least, is in instances of genuine causal interaction, although much of what we write would hold under simpler structures of only effect modification.

8.3.5 A Graph-Aided Interpretation of Extant Empirical Models of Charter School Effects

There are two basic goals of writing down a causal graph: (1) to represent the set of causal relationships implied by the available state of knowledge and (2) to assess the feasibility of alternative estimation strategies. Figure 8.7 represents a causal graph that is a reasonable representation of the causal structure that generates the charter school effect. This is a matter of judgment, and one might contend, for example, that the claim that self-selection on the charter school effect generates movement between neighborhoods is overly complex.

Suppose that one has access to observational data, such as the National Assessment of Education Progress (NAEP) data analyzed by Lubienski and Lubienski (2003), that provide information on standardized test scores, school type, and some family background characteristics. For the sake of our exposition, suppose further that these NAEP data had even better measures of family background and neighborhood characteristics, so that we could conclude that high-quality data are available for all of the variables in Figure 8.7 with solid nodes: Y, D, N, and P. Yet, no data are available for the variables with hollow nodes: I, $\text{Exp}(D \rightarrow Y)$, and G. The primary goal of analysis is to estimate the average causal effect of D on Y, as this effect interacts with the complementary causal effect of N on Y. Can one adjust for confounding in order to estimate these effects?

The first question to consider is the following: Is there a set of observed variables in Figure 8.7 that satisfies the back-door criterion with the respect the causal effect of D on Y and the causal effect of N on Y? The relevant back-door paths are,

for D,

1. $D \leftarrow P \rightarrow Y$,
2. $D \leftarrow P \rightarrow I \rightarrow \text{Exp}(D \rightarrow Y) \rightarrow G \rightarrow Y$,
3. $D \leftarrow P \rightarrow I \rightarrow \text{Exp}(D \rightarrow Y) \rightarrow G \rightarrow N \rightarrow Y$,
4. $D \leftarrow G \rightarrow N \rightarrow Y$,
5. $D \leftarrow G \rightarrow Y$;

and

for N,

6. $N \leftarrow G \rightarrow Y$,
7. $N \leftarrow G \rightarrow D \rightarrow Y$,
8. $N \leftarrow G \leftarrow \text{Exp}(D \rightarrow Y) \leftarrow I \leftarrow P \rightarrow Y$,
9. $N \leftarrow G \leftarrow \text{Exp}(D \rightarrow Y) \leftarrow I \leftarrow P \rightarrow D \rightarrow Y$,
10. $N \leftarrow G \leftarrow \text{Exp}(D \rightarrow Y) \leftarrow I \leftarrow P \rightarrow D \leftarrow G \rightarrow Y$.

How many of these paths can be blocked by conditioning on the observed data? For models that estimate the effect of D on Y, paths 1 through 4 can be blocked by conditioning on P and N. However, path 5 remains unblocked. Likewise, paths 7 through 10 can be blocked by conditioning on P and D, but path 6 remains

unblocked. For the two unblocked paths, the same problematic effect is present: $G \rightarrow Y$. This effect transmits the effects of exogenous causal determinants of I and $\text{Exp}(D \rightarrow Y)$, which are e_I and e_{Exp}, through the directed path $I \rightarrow \text{Exp}(D \rightarrow Y) \rightarrow G \rightarrow Y$. Thus, if there is selection on the causal effect itself, independent of parental background, then it enters the model through G and confounds the conditional associations between D and Y and between N and Y. This confounding cannot be eliminated by conditioning on the observed data, and therefore back-door conditioning for the effects of D and N on Y is infeasible.

This conclusion is hardly revelatory for readers who already know the literature on self-selection bias and/or the empirical literature on charter school effects. Nonetheless, we would argue that there is substantial didactic and communicative value in seeing this result expressed with a causal graph to which the back-door criterion is then applied. To effectively use back-door conditioning to identify all average causal effects of primary interest – the ATC, ATT, and ATE – we would need to specify a more elaborate set of causes of G, which would generate G through and/or alongside the structural equations that determine I and $\text{Exp}(D \rightarrow Y)$. It would be sufficient to have a model for G, $f_G(.)$, where all inputs other than e_G are observed and where these inputs are sufficiently rich such that e_G can be regarded as a constant that applies to everyone. Equivalently, we would need a set of observed measures as variables that point to G in the graph that would allow us to declare that there are no more latent classes precisely because G is a deterministic function of the observed variables in the graph. In this case, all sources of the noncausal association between D and Y and between N and Y would be indexed by the observed variables that determine G, and we could safely remove G from the graph and allow those variables to point directly to D, N, and Y.

Without access to such observed variables, which prevent us from explaining away the existence of the latent classes indexed by G, how can analysis proceed? There are two main choices. First, the analyst can concede that self-selection on the causal effect is present (because G cannot be eliminated from the graph), which may even generate neighborhood-based selection as a by-product. In these circumstances, the presence of the causal relationship $G \rightarrow Y$ renders estimates of D and N on Y unidentified, either in interactive fashion or when averaging one over the other. In this case, analysis must then be scaled back, and we would recommend that set identification results be pursued, as we will explain later in Chapter 12. The new goal would be to estimate an interval within which the average causal effect of interest must fall.

In the actual empirical literature on the charter school effect, this humble option has not been pursued by any of the main combatants in the debates. Instead, this research is a good example of what Manski (2013b) has labeled "policy analysis with incredible certitude." The desire to provide point estimates of causal effects has been too strong, even though it would seem clear to many outside readers that the debate over charter school effects persists simply because the point estimate of the ATE is unidentified. Consider the two dominant positions in the extant literature.

One group of researchers has used the lottery-induced structure of charter school enrollments in order to estimate effects. In cities where the number of charter schools is insufficient to meet demand, most schools are required by their chartering authorities to perform random lotteries and offer admission first to those who win their lotteries. This

admissions mechanism delivers two comparable groups: lottery winners and lottery losers. If the lottery winners attend charter schools and the lottery losers do not, then observation of achievement trajectories following the lottery enables estimation of average treatment effects for those who participate in the lottery.[16]

The researchers who use this research design typically define the charter school effect of interest as solely the ATT: the effect of charter schooling among those who would self-select into charter schooling by applying for the lottery. This is entirely appropriate for the interpretation of lottery-based results. The problem is that these same scholars too frequently forget the bounded nature of their conclusions. For example, in their study of charter schools in New York City, Hoxby, Murarka, and Kang (2009:vii) introduce their results in their "Executive Summary" with three bullet points:

- "Lottery-based analysis of charter schools' effects on achievement is, by far, the most reliable method of evaluation. It is the only method that reliably eliminates 'selection biases' which occur if students who apply to charter schools are more disadvantaged, more motivated, or different in any other way than students who do not apply."

- "On average, a student who attended a charter school for all of grades kindergarten through eight would close about 86 percent of the 'Scarsdale-Harlem achievement gap' in math and 66 percent of the achievement gap in English. A student who attended fewer grades would improve by a commensurately smaller amount."

- "On average, a lotteried-out student who stayed in the traditional public schools for all of grades kindergarten through eight would stay on grade level but would not close the 'Scarsdale-Harlem achievement gap' by much. However, the lotteried-out students' performance does improve and is better than the norm in the U.S. where, as a rule, disadvantaged students fall further behind as they age."

Nowhere in their "Executive Summary" is it stated that these comparisons across lotteried-in and lotteried-out students are only informative about those who self-select into the relevant charter school lottery. Selection bias is not eliminated, as claimed in the first bullet point; the lottery-generated data simply provide a reasonable comparison among those who self-select into the charter school lottery.

More generally, it is not conceded that the results are uninformative about two first-order policy questions:

1. What is the expected charter school effect for a randomly chosen student from New York City, assuming the supply of charter schools is held constant?

2. What is the expected charter school effect for the subset of students in public schools who would be induced to apply to charter schools if a policy intervention were implemented to expand the supply of charter schools?

[16] For simplicity, in this discussion we ignore non-comparability generated by random variation, initial compliance with the lottery, and selective attrition over the observation window following the lottery. In most cases, the researchers who have modeled these effects have carefully adjusted their data to neutralize these threats to inference.

To answer these questions, the authors would need a full model of the charter school effect for students who never applied to charter schools, and lottery-based studies use no data on these students. The ATT is the appropriate parameter to estimate only when addressing a second-order policy question that is not the focus of the debate: "Should charter schools in New York City continue to receive support because those students who have chosen to attend them have learned more than they would have if they had attended regular public schools?" The answer to this question appears to be yes, and yet Hoxby et al. (2009) chose not to limit their conclusions only to this supportable position, at least when putting forward their primary conclusions.

If this were not troubling enough, the alternative is worse. One can simply assume away the unblocked paths that include $G \rightarrow Y$, which is tantamount to assuming that self-selection does not exist. The study of the Center for Research on Education Outcomes (2009) is closer to this position. Its authors offer a complex set of conclusions based on national results where charter school students are matched to students from traditional public schools based on observed characteristics. Their overall conclusion is that charter schools are generally ineffective for a majority of students, even though charter schools may be effective for a minority of students who are not well served by regular public schools:

> In our nationally pooled sample, two subgroups fare better in charters than in the traditional system: students in poverty and ELL [English Language Learner] students. . . . These findings are particularly heartening for the charter advocates who target the most challenging educational populations or strive to improve education options in the most difficult communities. Charter schools that are organized around a mission to teach the most economically disadvantaged students in particular seem to have developed expertise in serving these communities. . . . The flip-side of this insight should not be ignored either. Students not in poverty and students who are not English language learners on average do notably worse than the same students who remain in the traditional public school system. Additional work is needed to determine the reasons underlying this phenomenon. Perhaps these students are "off-mission" in the schools they attend. (Center for Research on Education Outcomes 2009:7)

These conclusions are offered based on models that match students on observable characteristics, leaving unobserved selection on the causal effect unaccounted for and almost entirely ignored in the write-up of the results. The report is written as if variation in the average treatment effect for different types of students is the central interest, and that matching justifies estimation of all such conditional average effects.

It bears noting that the pattern presented in this study by CREO is consistent with the one that we favored for the extended example that concluded Chapter 7, which we see more generally as the signature of an underlying pattern of self-selection. In this case, students from families who are living in poverty but who make their way into charter schools are fleeing poor alternatives in their own neighborhood schools and, furthermore, have extra amounts of motivation to succeed in school. At the same time, it is likely that students from more advantaged families are more likely to be attending charter schools solely for lifestyle reasons. In fact, they may be trading off

academic opportunities in high-quality schools that they have found distasteful for medium-quality charter schools with peer cultures or instructional programs that are more appealing.

8.4 Conclusions

In this chapter, we have made the transition from "easy" to "hard" instances of causal effect estimation. No longer does simple conditioning on the observed determinants of the cause or all other direct causes of the outcome allow for identification and consistent estimation of the causal effect of interest. Instead, we have considered examples in which important variables that might have been used to mount an effective conditioning strategy are unobserved. Thus, selection of the treatment of interest is on unobserved characteristics of individuals that have unknown but suspected relationships to the outcome. We have also expanded our usage of directed graphs to show how patterns of self-selection and heterogeneity can be given explicit representations, and we have made the implicit case that omitted variable problems are best addressed, in the long run, by developing and deploying new measures.

We turn in the next three chapters to a presentation of additional methods to identify causal effects with observed variables when one cannot move beyond the available data: instrumental variable estimators in Chapter 9, causal mechanisms in Chapter 10, and panel data models that utilize pretreatment measures of the outcome variable in Chapter 11. We will explain these strategies in detail, discussing their strengths and weaknesses along the way.