5.3 Graphs and Identifiability

5.3.1 Parameter Identification in Linear Models

Consider a directed edge $X \longrightarrow Y$ embedded in a path diagram G, and let α stand for the path coefficient associated with that edge. It is well known that the regression coefficient $r_{YX} = \rho_{XY}\sigma_Y/\sigma_X$ can be decomposed into the sum

$$r_{YX} = \alpha + I_{YX},$$

where I_{YX} is not a function of α, since it is computed (e.g., using Wright's rules) from other paths connecting X and Y excluding the edge $X \longrightarrow Y$. (Such paths traverse both unidirected and bidirected arcs.) Thus, if we remove the edge $X \longrightarrow Y$ from the path diagram and find that the resulting subgraph entails zero correlation between X and Y, then we know that $I_{YX} = 0$ and $\alpha = r_{YX}$; hence, α is identified. Such entailment can be established graphically by testing whether X is d-separated from Y (by the empty set $Z = \{\emptyset\}$) in the subgraph. Figure 5.6 illustrates this simple test for identification: all paths between X and Y in the subgraph G_α are blocked by converging arrows, and α can immediately be equated with r_{YX}.

We can extend this basic idea to cases where I_{YX} is not zero but can be made zero by adjusting for a set of variables $Z = \{Z_1, Z_2, \ldots, Z_k\}$ that lie on various d-connected paths between X and Y. Consider the partial regression coefficient $r_{YX\cdot Z} = \rho_{YX\cdot Z}\sigma_{Y\cdot Z}/\sigma_{X\cdot Z}$, which represents the residual correlation between Y and X after Z is "partialled out." If Z contains no descendant of Y, then again we can

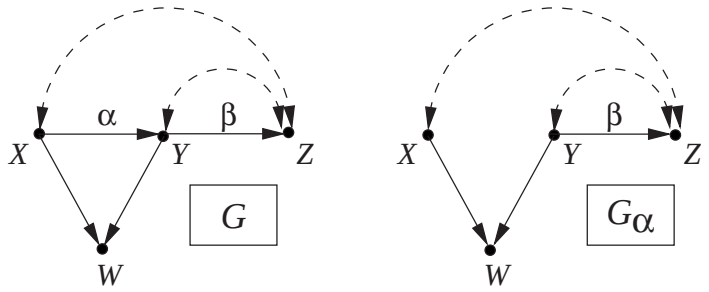

Figure 5.6: Test of whether structural parameter α can be equated with regression coefficient r_{YX}.

write[12]

$$r_{YX\cdot Z} = \alpha + I_{YX\cdot Z},$$

where $I_{YX\cdot Z}$ represents the partial correlation between X and Y resulting from setting α to zero, that is, the partial correlation in a model whose graph G_α lacks the edge $X \longrightarrow Y$ but is otherwise identical to G. If Z d-separates X from Y in G_α, then $I_{YX\cdot Z}$ would indeed be zero in such a model and so we can conclude that, in our original model, α is identified and is equal to $r_{YX\cdot Z}$. Moreover, since $r_{YX\cdot Z}$ is given by the coefficient of x in the regression of Y on X and Z, α can be estimated using the regression

$$y = \alpha x + \beta_1 z_1 + \ldots + \beta_k z_k + \epsilon.$$

This result provides a simple graphical answer to the questions, alluded to in Section 5.1.3, of (i) what constitutes an adequate set of regressors and (ii) when a regression coefficient provides a consistent estimate of a path coefficient. The answers are summarized in the following theorem.[13]

Theorem 5.3.1 (Single-Door Criterion for Direct Effects)
Let G be any path diagram in which α is the path coefficient associated with link $X \longrightarrow Y$, and let G_α denote the diagram that results when $X \longrightarrow Y$ is deleted from G. The coefficient α is identifiable if there exists a set of variables Z such that (i) *Z contains no descendant of Y and* (ii) *Z d-separates X from Y in G_α. If Z satisfies these two conditions, then α is equal to the regression coefficient $r_{YX\cdot Z}$. Conversely, if Z does not satisfy these conditions, then $r_{YX\cdot Z}$ is not a consistent estimand of α* (*except in rare instances of measure zero*).

[12] This can be seen when the relation between Y and its parents, $Y = \alpha x + \sum_i \beta_i w_i + \epsilon$, is substituted into the expression for $r_{YX\cdot Z}$, which yields α plus an expression $I_{YX\cdot Z}$ involving partial correlations among the variables $\{X, W_1, \ldots, W_k, Z, \epsilon\}$. Because Y is assumed not to be an ancestor of any of these variables, their joint density is unaffected by the equation for Y; hence, $I_{YX\cdot Z}$ is independent of α.

[13] This result is presented in Pearl (1998a) and Spirtes et al. (1998).

The use of Theorem 5.3.1 can be illustrated as follows. Consider the graphs G and G_α in Figure 5.7. The only path connecting X and Y in G_α is the one traversing Z,

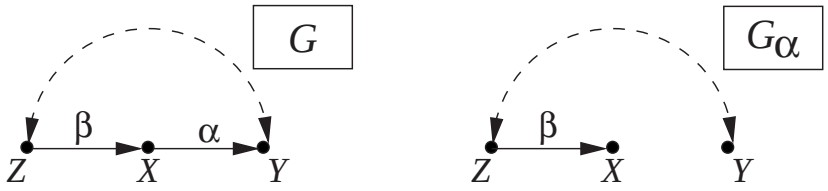

Figure 5.7: The identification of α (Theorem 5.3.1) in confirmed by G_α.

and since that path is d-separated (blocked) by Z, α is identifiable and is given by $\alpha = r_{YX \cdot Z}$. The coefficient β is identifiable, of course, since Z is d-separated from X in G_β (by the empty set $\emptyset$) and thus $\beta = r_{XZ}$. Note that this "single-door" test differs slightly from the back-door criterion for total effects (Definition 3.3.1); the set Z here must block *all* indirect paths from X to Y, not only back-door paths. Condition (i) is identical to both cases, because if X is a parent of Y then every descendant of Y must also be a descendant of X.

We now extend the identification of structural parameters through the identification of total effects (rather than direct effects). Consider the graph G in Figure 5.8. If we

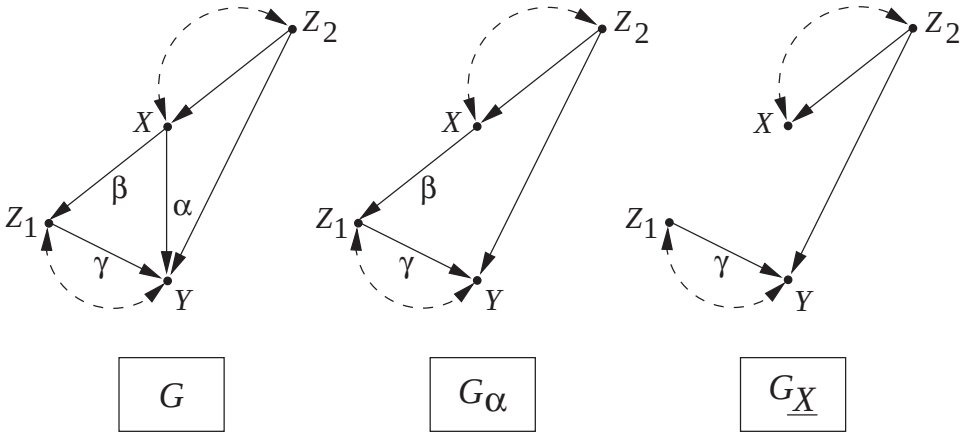

Figure 5.8: Graphical identification of the total effect of X on Y, yielding $\alpha + \beta\gamma = r_{YX \cdot Z_2}$.

form the graph G_α by removing the link $X \longrightarrow Y$, we observe that there is no set Z of nodes that d-separates all paths from X to Y. If Z contains Z_1, then the path $X \longrightarrow Z_1 \dashleftarrow\dashrightarrow Y$ will be unblocked through the converging arrows at Z_1. If Z does not contain Z_1, the path $X \longrightarrow Z_1 \longrightarrow Y$ is unblocked. Thus we conclude that α cannot be identified using our previous method. However, suppose we are interested

in the total effect of X on Y, which is given by $\alpha + \beta\gamma$. For this sum to be identified by r_{YX}, there should be no contribution to r_{YX} from paths other than those leading from X to Y. However, we see that two such paths, called *confounding* or *back-door* paths, exist in the graph—namely, $X \leftarrow Z_2 \rightarrow Y$ and $X \dashleftarrow\dashrightarrow Z_2 \rightarrow Y$. Fortunately, these paths are blocked by Z_2 and so we may conclude that adjusting for Z_2 would render $\alpha + \beta\gamma$ identifiable; thus we have

$$\alpha + \beta\gamma = r_{YX \cdot Z_2}.$$

This line of reasoning is captured by the back-door criterion of Definition 3.3.1, which we restate here for completeness.

Theorem 5.3.2 (Back-Door Criterion)
For any two variables X and Y in a causal diagram G, the total effect of X on Y is identifiable if there exists a set of measurements Z such that:

1. *no member of Z is a descendant of X; and*
2. *Z d-separates X from Y in the subgraph $G_{\underline{X}}$ formed by deleting from G all arrows emanating from X.*

Moreover, if the two conditions are satisfied, then the total effect of X on Y is given by $r_{YX \cdot Z}$.

The two conditions of Theorem 5.3.2, as we have seen in Section 3.3.1, are also valid in nonlinear non-Gaussian models as well as in models with discrete variables. The test ensures that, after adjustment for Z, the variables X and Y are not associated through confounding paths, which means that the regression coefficient $r_{YX \cdot Z}$ is equal to the total effect. In fact, we can view Theorems 5.3.1 and 5.3.2 as special cases of a more general scheme: In order to identify any *partial effect*, as defined by a select bundle of causal paths from X to Y, we ought to find a set Z of measured variables that block all nonselected paths between X and Y. The partial effect will then equal the regression coefficient $r_{YX \cdot Z}$.

Figure 5.8 demonstrates that some total effects can be determined directly from the graphs without having to identify their individual components. Standard SEM methods (Bollen 1989; Chou and Bentler 1995) that focus on the identification and estimation of individual parameters may miss the identification and estimation of effects such as the one in Figure 5.8, which can be estimated reliably even though some of the constituents remain unidentified.

Some total effects cannot be determined directly as a unit but instead require the determination of each component separately. In Figure 5.7, for example, the effect of Z on Y $(= \alpha\beta)$ does not meet the back-door criterion, yet this effect can be determined from its constituents α and β, which meet the back-door criterion individually and evaluate to

$$\beta = r_{XZ}, \qquad \alpha = r_{YX \cdot Z}.$$

There is yet a third kind of causal parameter: one that cannot be determined either directly or through its constituents but rather requires the evaluation of a broader causal

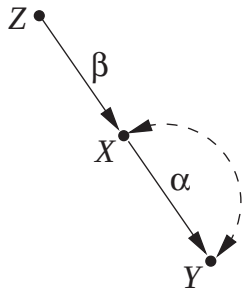

Figure 5.9: Graphical identification of α using instrumental variable Z.

effect of which it is a part. The structure shown in Figure 5.9 represents an example of this case. The parameter α cannot be identified either directly or from its constituents (it has none), yet it can be determined from $\alpha\beta$ and β, which represent the effect of Z on Y and of Z on X, respectively. These two effects can be identified directly, since there are no back-door paths from Z to either Y or X; therefore, $\alpha\beta = r_{YZ}$ and $\beta = r_{XZ}$. It follows that

$$\alpha = r_{YZ}/r_{XZ},$$

which is familiar to us as the *instrumental variable* formula (Bowden and Turkington 1984; see also Section 3.5, equation (3.46)).

The example shown in Figure 5.10 combines all three methods considered thus far. The total effect of X on Y is given by $\alpha\beta + \gamma\delta$, which is not identifiable because it

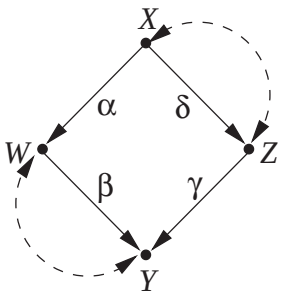

Figure 5.10: Graphical identification of α, β, and γ.

does not meet the back-door criterion and is not part of another identifiable structure. However, suppose we wish to estimate β. By conditioning on Z, we block all paths going through Z and obtain $\alpha\beta = r_{YX \cdot Z}$, which is the effect of X on Y mediated by W. Because there are no back-door paths from X to W, α itself evaluates directly to $\alpha = r_{WX}$. We therefore obtain

$$\beta = r_{YX \cdot Z}/r_{WX}.$$

On the other hand, γ can be evaluated directly by conditioning on X (thus blocking all back-door paths from Z to Y through X), which gives

$$\gamma = r_{YZ \cdot X}$$

The methods that we have been using suggest the following systematic procedure for recognizing identifiable coefficients in a graph.

1. Start by searching for identifiable causal effects among pairs of variables in the graph, using the back-door criterion and Theorem 5.3.1. These can be either direct effects, total effects, or partial effects (i.e., effects mediated by specific sets of variables).
2. For any such identified effect, collect the path coefficients involved and put them in a bucket.
3. Begin labeling the coefficients in the buckets according to the following procedure:
   (a) if a bucket is a singleton, label its coefficient I (denoting *identifiable*).
   (b) If a bucket is not a singleton but contains only a single unlabeled element, label that element I.
4. Repeat this process until no new labeling is possible.
5. List all labeled coefficients; these are identifiable.

The process just described is not complete, because our insistence on labeling coefficients one at a time may cause us to miss certain opportunities. This is shown in Figure 5.11. Starting with the pairs $(X, Z),\ (X, W),\ (X', Z)$, and (X', W), we dis-

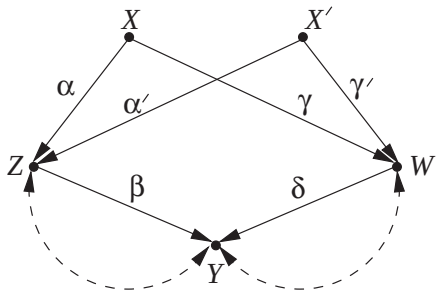

Figure 5.11: Identifying β and δ using two instrumental variables.

cover that $\alpha,\ \gamma,\ \alpha'$, and γ' are identifiable. Going to (X, Y), we find that $\alpha\beta + \delta\gamma$ is identifiable; likewise, from (X', Y) we see that $\alpha'\beta + \gamma'\delta$ is identifiable. This does not yet enable us to label β or δ, but we can solve two equations for the unknowns β and δ as long as the determinant $\begin{vmatrix} \alpha & \gamma \\ \alpha' & \gamma' \end{vmatrix}$ is nonzero. Since we are not interested in

identifiability at a point but rather in identifiability "almost everywhere" (Koopmans et al. 1950; Simon 1953), we need not compute this determinant. We merely inspect the symbolic form of the determinant's rows to make sure that the equations are nonredundant; each imposes a new constraint on the unlabeled coefficients for at least one value of the labeled coefficients.

With a facility to detect redundancies, we can increase the power of our procedure by adding the following rule:

3^*. If there are k nonredundant buckets that contain at most k unlabeled coefficients, label these coefficients and continue.

Another way to increase the power of our procedure is to list not only identifiable effects but also expressions involving correlations due to bidirected arcs, in accordance with Wright's rules. Finally, one can endeavor to list effects of several variables jointly as is done in Section 4.4. However, such enrichments tend to make the procedure more complex and might compromise our main objective of providing investigators with a way to immediately recognize the identified coefficients in a given model and immediately understand those features in the model that influence the identifiability of the target quantity. We now relate these results to the identification in nonparametric models, such as those treated in Section 3.3.

5.3.2 Comparison to Nonparametric Identification

The identification results of the previous section are significantly more powerful than those obtained in Chapters 3 and 4 for nonparametric models. Nonparametric models should nevertheless be studied by parametric modelers for both practical and conceptual reasons. On the practical side, investigators often find it hard to defend the assumptions of linearity and normality (or other functional-distributional assumptions), especially when categorical variables are involved. Because nonparametric results are valid for nonlinear functions and for any distribution of errors, having such results allows us to gauge how sensitive standard techniques are to assumptions of linearity and normality. On the conceptual side, nonparametric models illuminate the distinctions between structural and algebraic equations. The search for nonparametric quantities analogous to path coefficients forces explication of what path coefficients really mean, why one should labor at their identification, and why structural models are not merely a convenient way of encoding covariance information.

In this section we cast the problem of nonparametric causal effect identification (Chapter 3) in the context of parameter identification in linear models.

Parametric versus Nonparametric Models: An Example

Consider the set of structural equations

$$x = f_1(u, \epsilon_1), \tag{5.4}$$
$$z = f_2(x, \epsilon_2), \tag{5.5}$$
$$y = f_3(z, u, \epsilon_3), \tag{5.6}$$

where $X,\ Z,\ Y$ are observed variables, $f_1,\ f_2,\ f_3$ are unknown arbitrary functions, and $U,\ \epsilon_1,\ \epsilon_2,\ \epsilon_3$ are unobservables that we can regard either as latent variables or as disturbances. For the sake of this discussion, we will assume that $U, \epsilon_1, \epsilon_2$, and ϵ_3 are mutually independent and arbitrarily distributed. Graphically, these influences can be represented by the path diagram of Figure 5.12.

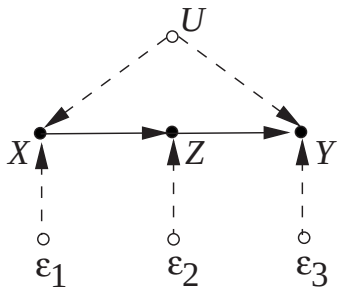

Figure 5.12: Path diagram corresponding to equations (5.4)–(5.6), where $\{X, Z, Y\}$ are observed and $\{U, \epsilon_1, \epsilon_2, \epsilon_3\}$ are unobserved.

The problem is as follows. We have drawn a long stream of independent samples of the process defined by (5.4)–(5.6) and have recorded the values of the observed variables $X,\ Z,\ Y$; we now wish to estimate the unspecified quantities of the model to the greatest extent possible.

To clarify the scope of the problem, we consider its linear version, which is given by

$$x = u + \epsilon_1, \tag{5.7}$$

$$z = \alpha x + \epsilon_2, \tag{5.8}$$

$$y = \beta z + \gamma u + \epsilon_3, \tag{5.9}$$

where $U,\ \epsilon_1,\ \epsilon_2,\ \epsilon_3$ are uncorrelated, zero-mean disturbances.[14] It is not hard to show that parameters α, β, and γ can be determined uniquely from the correlations among the observed quantities X, Z, and Y. This identification was demonstrated already in the example of Figure 5.7, where the back-door criterion yielded

$$\beta = r_{YZ \cdot X}, \quad \alpha = r_{ZX}, \tag{5.10}$$

and hence

$$\gamma = r_{YX} - \alpha\beta. \tag{5.11}$$

Thus, returning to the nonparametric version of the model, it is tempting to generalize that, for the model to be identifiable, the functions $\{f_1, f_2, f_3\}$ must be determined uniquely from the data. However, the prospect of this happening is unlikely,

[14] An equivalent version of this model is obtained by eliminating U from the equations and allowing ϵ_1 and ϵ_3 to be correlated, as in Figure 5.7.

because the mapping between functions and distributions is known to be many-to-one. In other words, given any nonparametric model M, if there exists one set of functions $\{f_1, f_2, f_3\}$ compatible with a given distribution $P(x, y, z)$, then there are infinitely many such functions (see Figure 1.6). Thus, it seems that nothing useful can be inferred from loosely specified models such as the one given by (5.4)–(5.6).

Identification is not an end in itself, however, even in linear models. Rather, it serves to answer practical questions of prediction and control. At issue is not whether the data permit us to identify the form of the equations but, instead, whether the data permit us to provide unambiguous answers to questions of the kind traditionally answered by parametric models.

When the model given by (5.4)–(5.6) is used strictly for prediction (i.e., to determine the probabilities of some variables given a set of observations on other variables), the question of identification loses much (if not all) of its importance; all the predictions can be estimated directly from either the covariance matrices or the sample estimates of those covariances. If dimensionality reduction is needed (e.g., to improve estimation accuracy) then the covariance matrix can be encoded in a variety of simultaneous equation models, all of the same dimensionality. For example, the correlations among X, Y, and Z in the linear model M of (5.7)–(5.9) might well be represented by the model M' (Figure 5.13):

$$x = \epsilon_1, \tag{5.12}$$

$$z = \alpha' x + \epsilon_2, \tag{5.13}$$

$$y = \beta' z + \delta x + \epsilon_3. \tag{5.14}$$

This model is as compact as (5.7)–(5.9) and is covariance equivalent to M with respect to the observed variables X, Y, Z. Upon setting $\alpha' = \alpha$, $\beta' = \beta$, and $\delta = \gamma$, model M' will yield the same probabilistic predictions as those of the model of (5.7)–(5.9). Still, when viewed as data-generating mechanisms, the two models are not equivalent. Each tells a different story about the processes generating X, Y, and Z, so naturally their predictions differ concerning the changes that would result from subjecting these processes to external interventions.

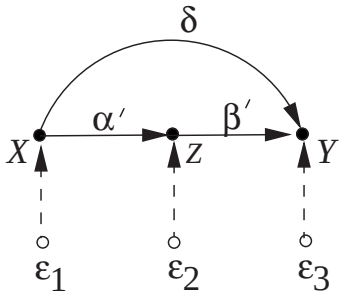

Figure 5.13: Diagram representing model M' of (5.12)–(5.14).

5.3.3 Causal Effects: The Interventional Interpretation of Structural Equation Models

The differences between models M and M' illustrate precisely where the structural reading of simultaneous equation models comes into play, and why even causally shy researchers consider structural parameters more "meaningful" than covariances and other statistical parameters. Model M', defined by (5.12)–(5.14), regards X as a direct participant in the process that determines the value of Y, whereas model M, defined by (5.7)–(5.9), views X as an indirect factor whose effect on Y is mediated by Z. This difference is not manifested in the data itself but rather in the way the data would change in response to outside interventions. For example, suppose we wish to predict the expectation of Y after we intervene and fix the value of X to some constant x; this is denoted $E(Y|do(X = x))$. After $X = x$ is substituted into (5.13) and (5.14), model M' yields

$$E[Y|do(X = x)] = E[\beta'\alpha' x + \beta'\epsilon_2 + \delta x + \epsilon_3] \quad (5.15)$$
$$= (\beta'\alpha' + \delta)x; \quad (5.16)$$

model M yields

$$E[Y|do(X = x)] = E[\beta\alpha x + \beta\epsilon_2 + \gamma u + \epsilon_3] \quad (5.17)$$
$$= \beta\alpha x. \quad (5.18)$$

Upon setting $\alpha' = \alpha,\ \beta' = \beta$, and $\delta = \gamma$ (as required for covariance equivalence; see (5.10) and (5.11)), we see clearly that the two models assign different magnitudes to the (total) causal effect of X on Y: model M predicts that a unit change in x will change $E(Y)$ by the amount $\beta\alpha$, wheras model M' puts this amount at $\beta\alpha + \gamma$.

At this point, it is tempting to ask whether we should substitute $x - \epsilon_1$ for u in (5.9) prior to taking expectations in (5.17). If we permit the substitution of (5.8) into (5.9), as we did in deriving (5.17), why not permit the substitution of (5.7) into (5.9) as well? After all (the argument runs), there is no harm in upholding a mathematical equality, $u = x - \epsilon_1$, that the modeler deems valid. This argument is fallacious, however.[15] Structural equations are not meant to be treated as immutable mathematical equalities. Rather, they are meant to define a state of equilibrium—one that is *violated* when the equilibrium is perturbed by outside interventions. In fact, the power of structural equation models is that they encode not only the initial equilibrium state but also the information necessary for determining which equations must be violated in order to account for a new state of equilibrium. For example, if the intervention consists merely of holding X constant at x, then the equation $x = u + \epsilon_1$, which represents the preintervention process determining X, should be overruled and replaced with the equation $X = x$. The solution to the new set of equations then represents the new equilibrium. Thus, the essential characteristic of structural equations that sets them apart from ordinary mathematical equations is that the former stand not for one but for many sets of equations, each corresponding to a subset of equations taken from the

[15] Such arguments have led to Newcomb's paradox in the so-called evidential decision theory (see Section 4.1.1).

original model. Every such subset represents some hypothetical physical reality that would prevail under a given intervention.

If we take the stand that the value of structural equations lies not in summarizing distribution functions but in encoding causal information for predicting the effects of policies (Haavelmo 1943; Marschak 1950; Simon 1953), it is natural to view such predictions as the proper generalization of structural coefficients. For example, the proper generalization of the coefficient β in the linear model M would be the answer to the control query, "What would be the change in the expected value of Y if we were to intervene and change the value of Z from z to $z+1$," which is different, of course, from the observational query, "What would be the difference in the expected value of Y if we were to *find* Z at level $z+1$ instead of level z." Observational queries, as we discussed in Chapter 1, can be answered directly from the joint distribution $P(x, y, z)$, while control queries require causal information as well. Structural equations encode this causal information in their syntax by treating the variable on the left-hand side of the equality sign as the effect and treating those on the right as causes. In Chapter 3 we distinguished between the two types of queries through the symbol $do(\cdot)$. For example, we wrote

$$E(Y|do(x)) \triangleq E[Y|do(X = x)] \tag{5.19}$$

for the controlled expectation and

$$E(Y|x) \triangleq E(Y|X = x) \tag{5.20}$$

for the standard conditional or observational expectation. That $E(Y|do(x))$ does not equal $E(Y|x)$ can easily be seen in the model of (5.7)–(5.9), where $E(Y|do(x)) = \alpha\beta x$ but $E(Y|x) = r_{YX}x = (\alpha\beta + \gamma)x$. Indeed, the passive observation $X = x$ should not violate any of the equations, and this is the justification for substituting both (5.7) and (5.8) into (5.9) before taking the expectation.

In linear models, the answers to questions of direct control are encoded in the path (or structural) coefficients, which can be used to derive the total effect of any variable on another. For example, the value of $E(Y|do(x))$ in the model defined by (5.7)–(5.9) is $\alpha\beta x$, that is, x times the product of the path coefficients along the path $X \longrightarrow Z \longrightarrow Y$. Computation of $E(Y|do(x))$ would be more complicated in the nonparametric case, even if we knew the functions f_1, f_2, and f_3. Nevertheless, this computation is well-defined; it requires the solution (for the expectation of Y) of a modified set of equations in which f_1 is "wiped out" and X is replaced by the constant x:

$$z = f_2(x, \epsilon_2), \tag{5.21}$$

$$y = f_3(z, u, \epsilon_3). \tag{5.22}$$

Thus, computation of $E(Y|do(x))$ requires evaluation of

$$E(Y|do(x)) = E\{f_3[f_2(x, \epsilon_2), u, \epsilon_3]\},$$

where the expectation is taken over U, ϵ_2, and ϵ_3. Graphical methods for performing this computation were discussed in Section 3.3.2.

What, then, is an appropriate definition of identifiability for nonparametric models? One reasonable definition is that answers to interventional queries are *unique*, and this is precisely how Definition 3.2.3 interprets the identification of the causal effect $P(y|do(x))$. As we have seen in Chapters 3 and 4, many aspects of nonparametric identification can be determined graphically, almost by inspection, from the diagrams that accompany the equations. These include tests for deciding whether a given interventional query is identifiable as well as formulas for estimating such queries.