Rule 3 (*Insertion/deletion of actions*):

$$P(y \mid \hat{x}, \hat{z}, w) = P(y \mid \hat{x}, w) \text{ if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \overline{Z(W)}}}, \tag{3.33}$$

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

Each of these inference rules follows from the basic interpretation of the "hat" $\hat{x}$ operator as a replacement of the causal mechanism that connects X to its preaction parents by a new mechanism $X = x$ introduced by the intervening force. The result is a *submodel* characterized by the subgraph $G_{\overline{X}}$ (named "manipulated graph" in Spirtes et al. 1993).

Rule 1 reaffirms *d*-separation as a valid test for conditional independence in the distribution resulting from the intervention $do(X = x)$, hence the graph $G_{\overline{X}}$. This rule follows from the fact that deleting equations from the system does not introduce any dependencies among the remaining disturbance terms (see (3.2)).

Rule 2 provides a condition for an external intervention $do(Z = z)$ to have the same effect on Y as the passive observation $Z = z$. The condition amounts to $\{X \cup W\}$ blocking all back-door paths from Z to Y (in $G_{\overline{X}}$), since $G_{\overline{X}\underline{Z}}$ retains all (and only) such paths.

Rule 3 provides conditions for introducing (or deleting) an external intervention $do(Z = z)$ without affecting the probability of $Y = y$. The validity of this rule stems, again, from simulating the intervention $do(Z = z)$ by the deletion of all equations corresponding to the variables in Z (hence the graph $G_{\overline{X}\,\overline{Z}}$). The reason for limiting the deletion to nonancestors of W-nodes is provided with the proofs of Rules 1–3 in Pearl (1995a).

Corollary 3.4.2

A causal effect $q = P(y_1, \ldots, y_k \mid \hat{x}_1, \ldots, \hat{x}_m)$ *is identifiable in a model characterized by a graph G if there exists a finite sequence of transformations, each conforming to one of the inference rules in Theorem 3.4.1, that reduces q into a standard* (*i.e., "hat"-free*) *probability expression involving observed quantities.*

Rules 1–3 have been shown to be *complete*, namely, sufficient for deriving all identifiable causal effects (Shpitser and Pearl 2006a; Huang and Valtorta 2006). Moreover, as illustrated in Section 3.4.3, symbolic derivations using the hat notation are more convenient than algebraic derivations that aim at eliminating latent variables from standard probability expressions (as in Section 3.3.2, equation (3.24)). However, the task of deciding whether a sequence of rules exists for reducing an arbitrary causal effect expression has not been systematized, and direct graphical criteria for identification are therefore more desirable. These will be developed in Chapter 4.

3.4.3 Symbolic Derivation of Causal Effects: An Example

We will now demonstrate how Rules 1–3 can be used to derive all causal effect estimands in the structure of Figure 3.5. Figure 3.6 displays the subgraphs that will be needed for the derivations that follow.

Task 1: Compute $P(z \mid \hat{x})$

This task can be accomplished in one step, since G satisfies the applicability condition for Rule 2. That is, $X \perp\!\!\!\perp Z$ in $G_{\underline{X}}$ (because the path $X \leftarrow U \rightarrow Y \leftarrow Z$ is blocked by the converging arrows at Y), and we can write

$$P(z \mid \hat{x}) = P(z \mid x). \tag{3.34}$$

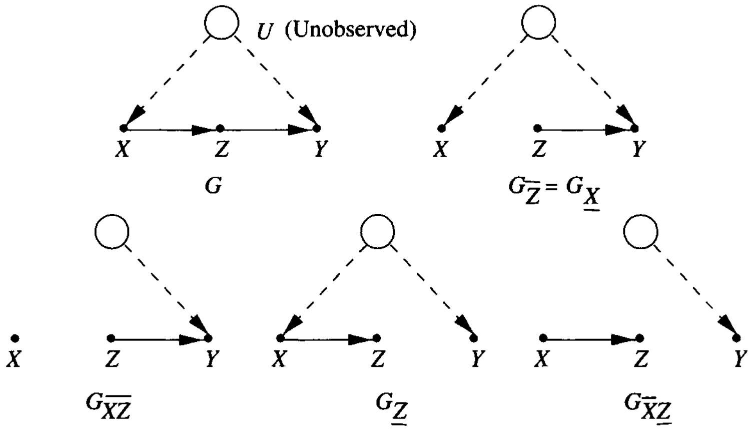

Figure 3.6 Subgraphs of G used in the derivation of causal effects.

Task 2: Compute $P(y \mid \hat{z})$

Here we cannot apply Rule 2 to exchange $\hat{z}$ with z because $G_{\underline{Z}}$ contains a back-door path from Z to Y: $Z \leftarrow X \leftarrow U \rightarrow Y$. Naturally, we would like to block this path by measuring variables (such as X) that reside on that path. This involves conditioning and summing over all values of X:

$$P(y \mid \hat{z}) = \sum_x P(y \mid x, \hat{z})\, P(x \mid \hat{z}). \tag{3.35}$$

We now have to deal with two terms involving $\hat{z}$, $P(y \mid x, \hat{z})$ and $P(x \mid \hat{z})$. The latter can be readily computed by applying Rule 3 for action deletion:

$$P(x \mid \hat{z}) = P(x) \quad \text{if} \quad (Z \perp\!\!\!\perp X)_{G_{\overline{Z}}}, \tag{3.36}$$

since X and Z are d-separated in $G_{\overline{Z}}$. (Intuitively, manipulating Z should have no effect on X, because Z is a descendant of X in G.) To reduce the former term, $P(y \mid x, \hat{z})$, we consult Rule 2:

$$P(y \mid x, \hat{z}) = P(y \mid x, z) \quad \text{if } (Z \perp\!\!\!\perp Y \mid X)_{G_{\underline{Z}}}, \tag{3.37}$$

noting that X d-separates Z from Y in $G_{\underline{Z}}$. This allows us to write (3.35) as

$$P(y \mid \hat{z}) = \sum_x P(y \mid x, z)\, P(x) = E_x P(y \mid x, z), \tag{3.38}$$

which is a special case of the back-door formula (equation (3.19)). The legitimizing condition, $(Z \perp\!\!\!\perp Y \mid X)_{G_{\underline{Z}}}$, offers yet another graphical test for a set X to be sufficient for control of confounding (between Y and Z) that is equivalent to the opaque "ignorability" condition of Rosenbaum and Rubin (1983).

Task 3: Compute $P(y \mid \hat{x})$

Writing

$$P(y \mid \hat{x}) = \sum_z P(y \mid z, \hat{x})\, P(z \mid \hat{x}), \tag{3.39}$$

we see that the term $P(z \mid \hat{x})$ was reduced in (3.34) but that no rule can be applied to eliminate the hat symbol ^ from the term $P(y \mid z, \hat{x})$. However, we can legitimately add this symbol via Rule 2:

$$P(y \mid z, \hat{x}) = P(y \mid \hat{z}, \hat{x}), \tag{3.40}$$

since the applicability condition $(Y \perp\!\!\!\perp Z \mid X)_{G_{\overline{X}\underline{Z}}}$ holds (see Figure 3.6). We can now delete the action $\hat{x}$ from $P(y \mid \hat{z}, \hat{x})$ using Rule 3, since $Y \perp\!\!\!\perp X \mid Z$ holds in $G_{\overline{X}\overline{Z}}$. Thus, we have

$$P(y \mid z, \hat{x}) = P(y \mid \hat{z}), \tag{3.41}$$

which was calculated in (3.38). Substituting (3.38), (3.41), and (3.34) back into (3.39) finally yields

$$P(y \mid \hat{x}) = \sum_{z} P(z \mid x) \sum_{x'} P(y \mid x', z)\, P(x'), \tag{3.42}$$

which is identical to the front-door formula of (3.28).

Task 4: Compute $P(y, z \mid \hat{x})$
We have

$$P(y, z \mid \hat{x}) = P(y \mid z, \hat{x})\, P(z \mid \hat{x}').$$

The two terms on the r.h.s. were derived before in (3.34) and (3.41), from which we obtain

$$\begin{aligned} P(y, z \mid \hat{x}) &= P(y \mid \hat{z})\, P(z \mid x) \\ &= P(z \mid x) \sum_{x'} P(y \mid x', z)\, P(x'). \end{aligned} \tag{3.43}$$

Task 5: Compute $P(x, y \mid \hat{z})$
We have

$$\begin{aligned} P(x, y \mid \hat{z}) &= P(y \mid x, \hat{z})\, P(x \mid \hat{z}) \\ &= P(y \mid x, z)\, P(x). \end{aligned} \tag{3.44}$$

The first term on the r.h.s. is obtained by Rule 2 (licensed by $G_{\underline{Z}}$) and the second term by Rule 3 (as in (3.36)).

Note that, in all the derivations, the graph G has provided both the license for applying the inference rules and the guidance for choosing the right rule to apply.

3.4.4 Causal Inference by Surrogate Experiments

Suppose we wish to learn the causal effect of X on Y when $P(y \mid \hat{x})$ is not identifiable and, for practical reasons of cost or ethics, we cannot control X by randomized experiment. The question arises of whether $P(y \mid \hat{x})$ can be identified by randomizing