CAUSAL EFFECT

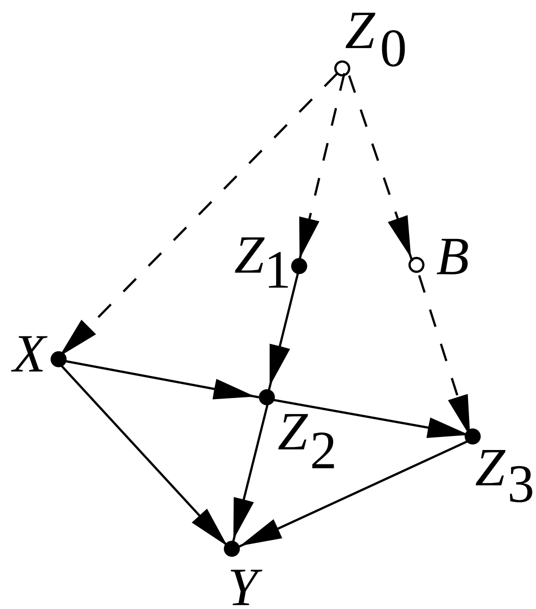

Figure 3.1: A causal diagram representing the effect of fumigants (X) on yields (Y).

$$\begin{aligned} Z_0 &= f_0(\epsilon_0), & B &= f_B(Z_0, \epsilon_B), \\ Z_1 &= f_1(Z_0, \epsilon_1), & X &= f_X(Z_0, \epsilon_X), \\ Z_2 &= f_2(X, Z_1, \epsilon_2), & Y &= f_Y(X, Z_2, Z_3, \epsilon_Y), \\ Z_3 &= f_3(B, Z_2, \epsilon_3). \end{aligned} \quad \textbf{(3.3)}$$

$$P(x_1, ..., x_n) = \prod_i P(x_i \mid pa_i), \quad \textbf{(3.5)}$$

$$\begin{aligned} P(z_0, x, z_1, b, z_2, z_3, y) &= P(z_0)P(x|z_0)P(z_1|z_0) \\ &\quad \times P(b|z_0)P(z_2|x, z_1) \\ &\quad \times P(z_3|z_2, b)P(y|x, z_2, z_3). \end{aligned} \quad \textbf{(3.6)}$$

Find $P(y|\hat{x})$ given $P(y, x, z_1, z_2, z_3)$

CAUSAL EFFECT

Definition 3.2.1 (Causal Effect)
Given two disjoint sets of variables, X and Y, the **causal effect** of X on Y, denoted either as $P(y|\hat{x})$ or as $P(y|do(x))$, is a function from X to the space of probability distributions on Y.

For each realization x of X, $P(y|\hat{x})$ gives the probability of $Y = y$ induced by deleting from the model of (3.4) all equations corresponding to variables in X and substituting $X = x$ in the remaining equations.

INTERVENTIONS AS VARIABLES

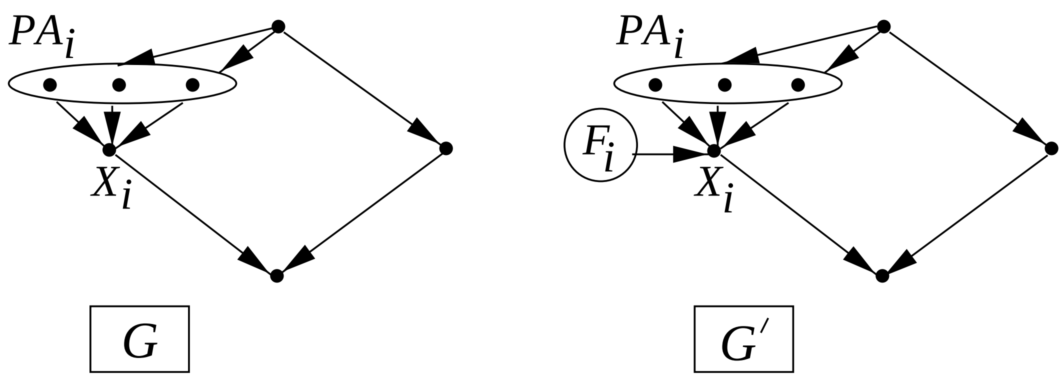

Figure 3.2: Representing external intervention F_i by an augmented network $G' = G \cup \{F_i \to X_i\}$.

$$P(x_i \mid pa_i') = \begin{cases} P(x_i \mid pa_i) & \text{if } F_i = \text{idle}, \\ 0 & \text{if } F_i = do(x_i') \\ & \text{and } x_i \neq x_i', \\ 1 & \text{if } F_i = do(x_i') \\ & \text{and } x_i = x_i'. \end{cases} \quad \textbf{(3.8)}$$

$$P(x_1, ..., x_n | \hat{x}_i') = P'(x_1, ..., x_n \mid F_i = do(x_i')), \quad \textbf{(3.9)}$$

where P' is represented by G'.

THE TRUNCATED FACTORIZATION FORMULA

$$P(x_1, ..., x_n|\hat{x}'_i) \;=\begin{cases} \prod_{j\neq i} P(x_j|pa_j) & \text{if} \quad x_i = x'_i, \\ 0 & \text{if} \quad x_i \neq x'_i. \end{cases} \qquad \textbf{(3.10)}$$

$$P(x_1, \ldots, x_n|\hat{x}'_i) \;=\begin{cases} \dfrac{P(x_1, \ldots, x_n)}{P(x'_i \mid pa_i)} & \text{if} \quad x_i = x'_i, \\ 0 & \text{if} \quad x_i \neq x'_i. \end{cases} \qquad \textbf{(3.11)}$$

$$P(x_1, \ldots, x_n|\hat{x}'_i) =\begin{cases} P(x_1, \ldots, x_n|x'_i, pa_i)P(pa_i) \\ \qquad\qquad\qquad\qquad \text{if} \quad x_i = x'_i, \\ 0 \qquad\qquad\qquad\quad\;\; \text{if} \quad x_i \neq x'_i. \end{cases} \qquad \textbf{(3.12)}$$

Theorem 3.2.2 (Adjustment for Direct Causes) Let PA_i denote the set of direct causes of variable X_i, and let Y be any set of variables disjoint of $\{X_i \cup PA_i\}$. The effect of the intervention $do(X_i = x'_i)$ on Y is given by

$$P(y|\hat{x}'_i) = \sum_{pa_i} P(y|x'_i, pa_i)P(pa_i), \qquad \textbf{(3.13)}$$

where $P(y|x'_i, pa_i)$ and $P(pa_i)$ represent preintervention probabilities.

EXAMPLE: PROCESS CONTROL

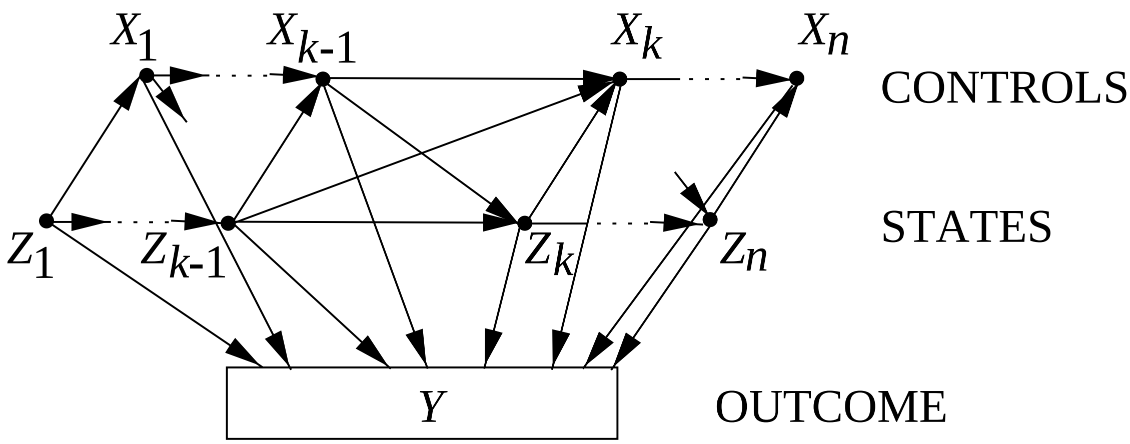

Figure 3.3:

Given samples from $P(y, z_1, \ldots, z_n, x_1, \ldots, x_n)$, find $P^*(y)$ where P^* obtains under a new strategy $S^* : P^*(x_k | x_{k-1}, z_k, z_{k-1})$

If $S^* : do(X_k = x_k)$, then

$$\begin{aligned} P^*(y) &= P(y | \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) \\ &= \sum_{z_1,..,z_n} P(y | z_1, z_2, \ldots, z_n, x_1, x_2, \ldots, x_n) \\ &\quad \prod_k P(z_k | z_{k-1}, x_{k-1}) \end{aligned} \tag{3.18}$$

IDENTIFIABILITY

Definition 3.2.3 (Identifiability)
Let $Q(M)$ be any computable quantity of a model M. We say that Q is identifiable in a class $\boldsymbol{M}$ of models if, for any pairs of models M_1 and M_2 from $\boldsymbol{M}$, $Q(M_1) = Q(M_2)$ whenever $P_{M_1}(v) = P_{M_2}(v)$.

If our observations are limited, and permit only a partial set F_M of features (of $P_M(v)$) to be estimated, we define Q to be identifiable from F_M if $Q(M_1) = Q(M_2)$ whenever $F_{M_1} = F_{M_2}$.

Definition 3.2.4 (Causal Effect Identifiability)
The **causal effect** of X on Y is said to be **identifiable** from a graph G if the quantity $P(y|\hat{x})$ can be computed uniquely from any positive probability of the observed variables—that is, if $P_{M_1}(y|\hat{x}) = P_{M_2}(y|\hat{x})$ for every pair of models M_1 and M_2 with $P_{M_1}(v) = P_{M_2}(v) > 0$ and $G(M_1) = G(M_2) = G$.

Theorem 3.2.5
Given a causal diagram G of any Markovian model in which a subset V of variables are measured, the causal effect P$(y|\hat{x})$ is identifiable whenever $\{X \cup Y \cup PA_X\} \subseteq V$, that is, whenever X, Y, and all parents of variables in X are measured. The expression of $P(y|\hat{x})$ is then obtained by adjusting for PA_x, as in (3.13).

Corollary 3.2.6
Given the causal diagram G of any Markovian model in which all variables are measured, the causal effect $P(y|\hat{x})$ is identifiable for every two subsets of variables X and Y and is obtained from the truncated factorization of (3.14).

THE BACK-DOOR CRITERION

Definition 3.3.1 (Back-Door)
A set of variables Z satisfies the **back-door** criterion relative to an ordered pair of variables (X_i, X_j) in a DAG G if:

(i) no node in Z is a descendant of X_i; and

(ii) Z blocks every path between X_i and X_j that contains an arrow into X_i.

Theorem 3.3.2 (Back-Door Adjustment)
If a set of variables Z satisfies the back-door criterion relative to (X, Y), then the causal effect of X on Y is identifiable and is given by the formula

$$P(y|\hat{x}) = \sum_z P(y|x, z)P(z). \qquad \textbf{(3.19)}$$

THE FRONT-DOOR CRITERION

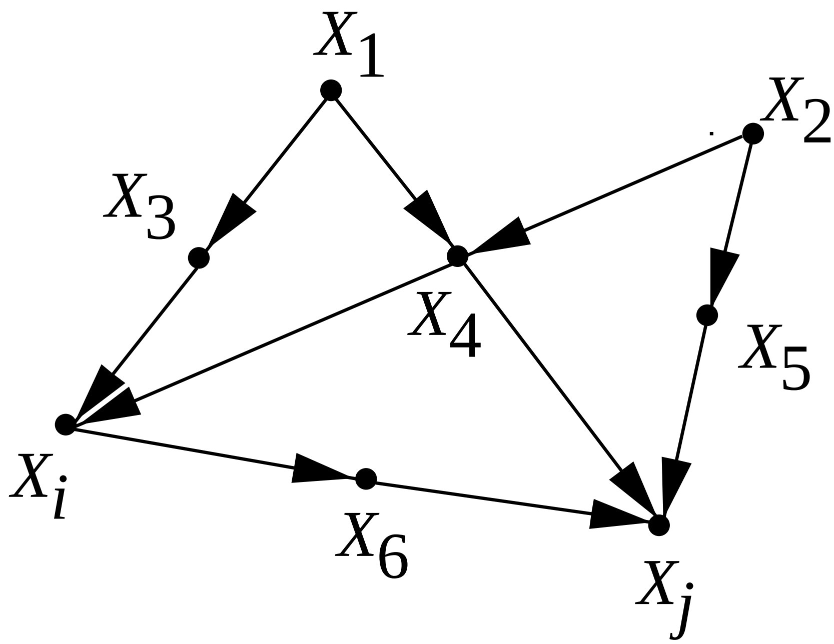

Figure 3.4

Suppose X_1, X_2, X_3, X_4, and X_5 are unobserved.
Can we find $P(x_j|\hat{x}_i)$?

THE FRONT-DOOR CRITERION (Cont.)

Definition 3.3.3 (Front-Door)
A set of variables Z is said to satisfy the **front-door** criterion relative to an ordered pair of variables (X, Y) if:

(i) Z intercepts all directed paths from X to Y;

(ii) there is no back-door path from X to Z; and

(iii) all back-door paths from Z to Y are blocked by X.

Theorem 3.3.4 (Front-Door Adjustment)
If Z satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by the formula

$$P(y|\hat{x}) = \sum_{z} P(z|x) \sum_{x'} P(y|x', z)P(x'). \quad \textbf{(3.29)}$$

PROOF OF FRONT-DOOR CRITERION

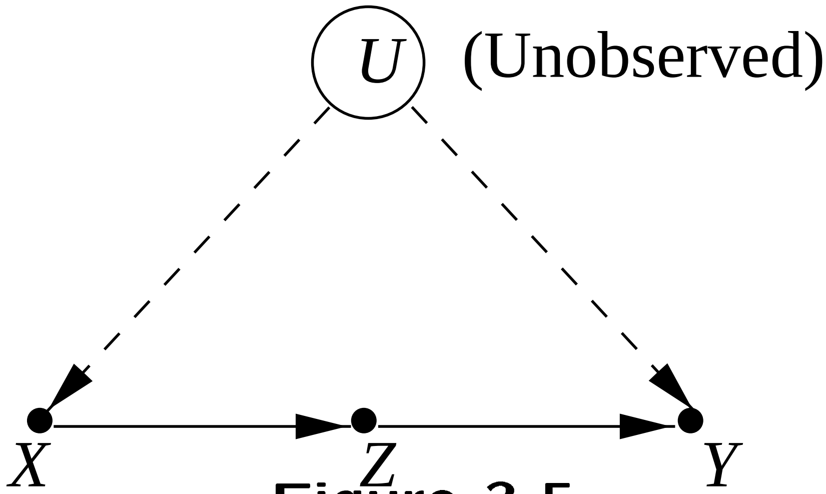

Figure 3.5

$$P(x, y, z, u) = P(u)P(x|u)P(z|x)P(y|z, u). \quad \textbf{(3.22)}$$

$$P(y, z, u|\hat{x}) = P(y|z, u)P(z|x)P(u). \quad \textbf{(3.23)}$$

$$P(y|\hat{x}) = \sum_z P(z|x) \sum_u P(y|z, u)P(u). \quad \textbf{(3.24)}$$

Eliminate u from this expression, using

$$P(u|z, x) = P(u|x), \quad \textbf{(3.25)}$$

$$P(y|x, z, u) = P(y|z, u). \quad \textbf{(3.26)}$$

yielding

$$P(y|\hat{x}) = \sum_z P(z|x) \sum_{x'} P(y|x', z)P(x'). \quad \textbf{(3.28)}$$

$= \sum_z P(y|\hat{z})P(z|\hat{x})$

Theorem 3.4.1 (Rules of *do* Calculus) Let G be the directed acyclic graph associated with a causal model as defined in (3.2), and let $P(\cdot)$ stand for the probability distribution induced by that model. For any disjoint subsets of variables X, Y, Z, and W we have the following rules.

Rule 1 (Insertion/deletion of observations) :

$$P(y|\hat{x}, z, w) = P(y|\hat{x}, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}. \tag{3.31}$$

Rule 2 (Action/observation exchange) :

$$P(y|\hat{x}, \hat{z}, w) = P(y|\hat{x}, z, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}. \tag{3.32}$$

Rule 3 (Insertion/deletion of actions) :

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

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

IMPLICATIONS OF *do*-CALCULUS

Corollary 3.4.2
A causal effect $q = P(y_1, ..., y_k|\hat{x}_1, ..., \hat{x}_m)$ 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.

NOTATION FOR *do* CALCULUS

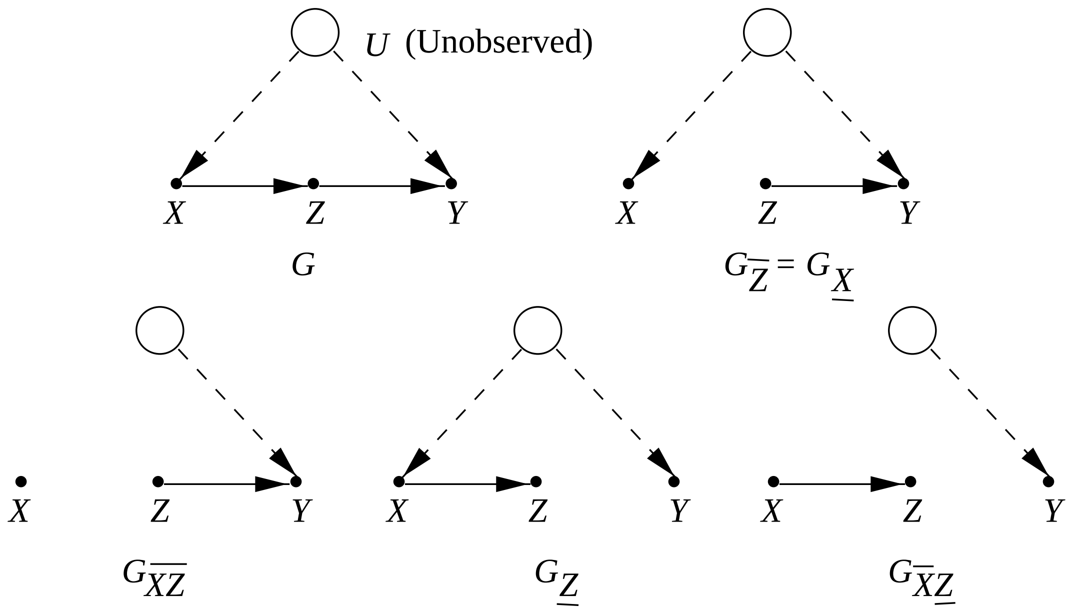

Figure 3.6: Subgraphs of G used in the derivation of causal effects.

$G_{\overline{X}}$ remove arrows pointing to X

$G_{\underline{X}}$ remove arrows emanating from X

$G_{\overline{X}\underline{Z}}$ remove ears of X and legs of Z

$$P(y|\widehat{x}, z) \triangleq \frac{P(y, z|\widehat{x})}{P(z|\widehat{x})}$$

NONIDENTIFYING MODELS (EXAMPLES)

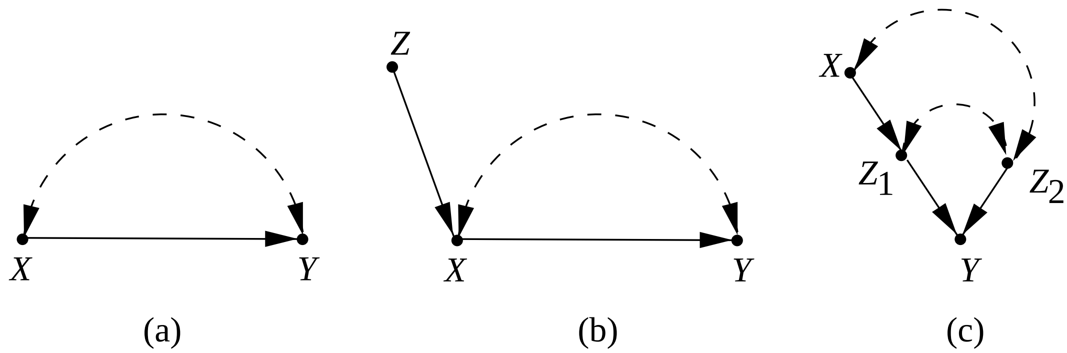

Figure 3.7: (a) A bow pattern: a confounding arc embracing a causal link $X \rightarrow Y$, thus preventing the identification of $P(y|\hat{x})$ even in the presence of an instrumental variable Z, as in (b). (c) A bowless graph that still prohibits the identification of $P(y|\hat{x})$.

$$\begin{aligned} P(y|\hat{x}, \hat{z_2}) &= \sum_{z_1} P(y|z_1, \hat{x}, \hat{z_2}) P(z_1|\hat{x}, \hat{z_2}) \\ &= \sum_{z_1} P(y|z_1, x, z_2) P(z_1|x). \end{aligned} \qquad \textbf{(3.47)}$$

$P(z_1|\hat{x}, z_2)$ is not identified.

IDENTIFYING MODELS

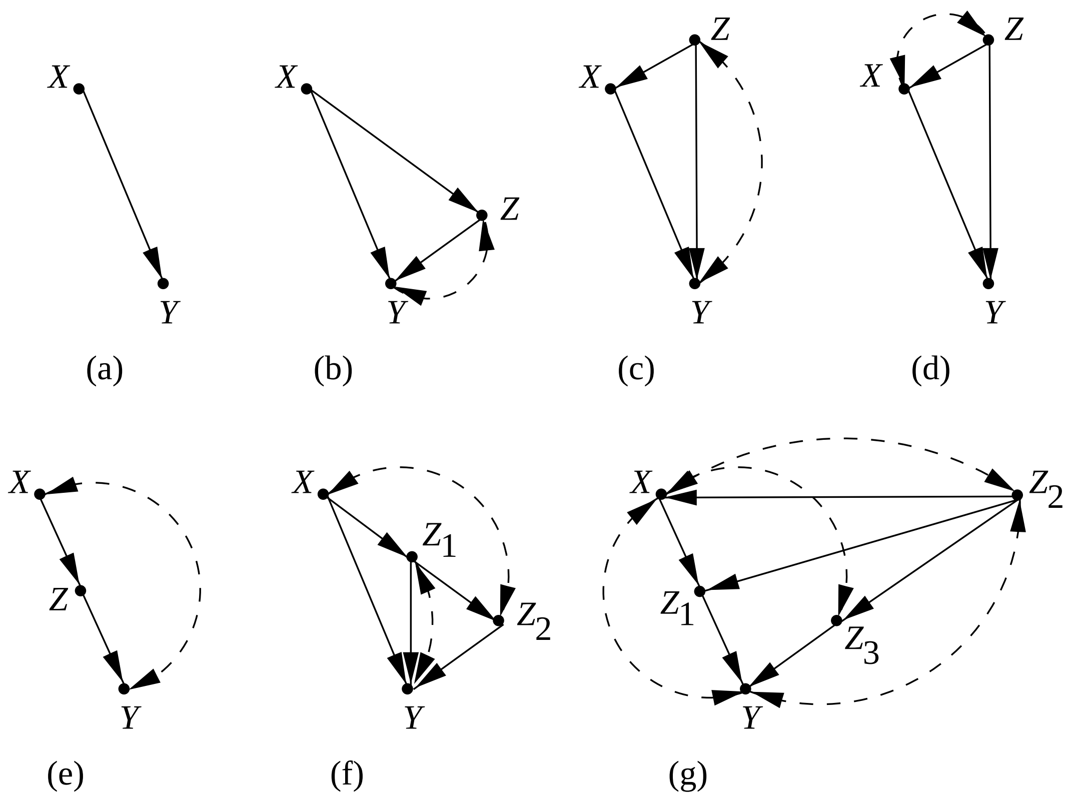

Figure 3.8: Typical models in which the effect of X on Y is identifiable. Dashed arcs represent confounding paths, and Z represents observed covariates.

NONIDENTIFYING MODELS

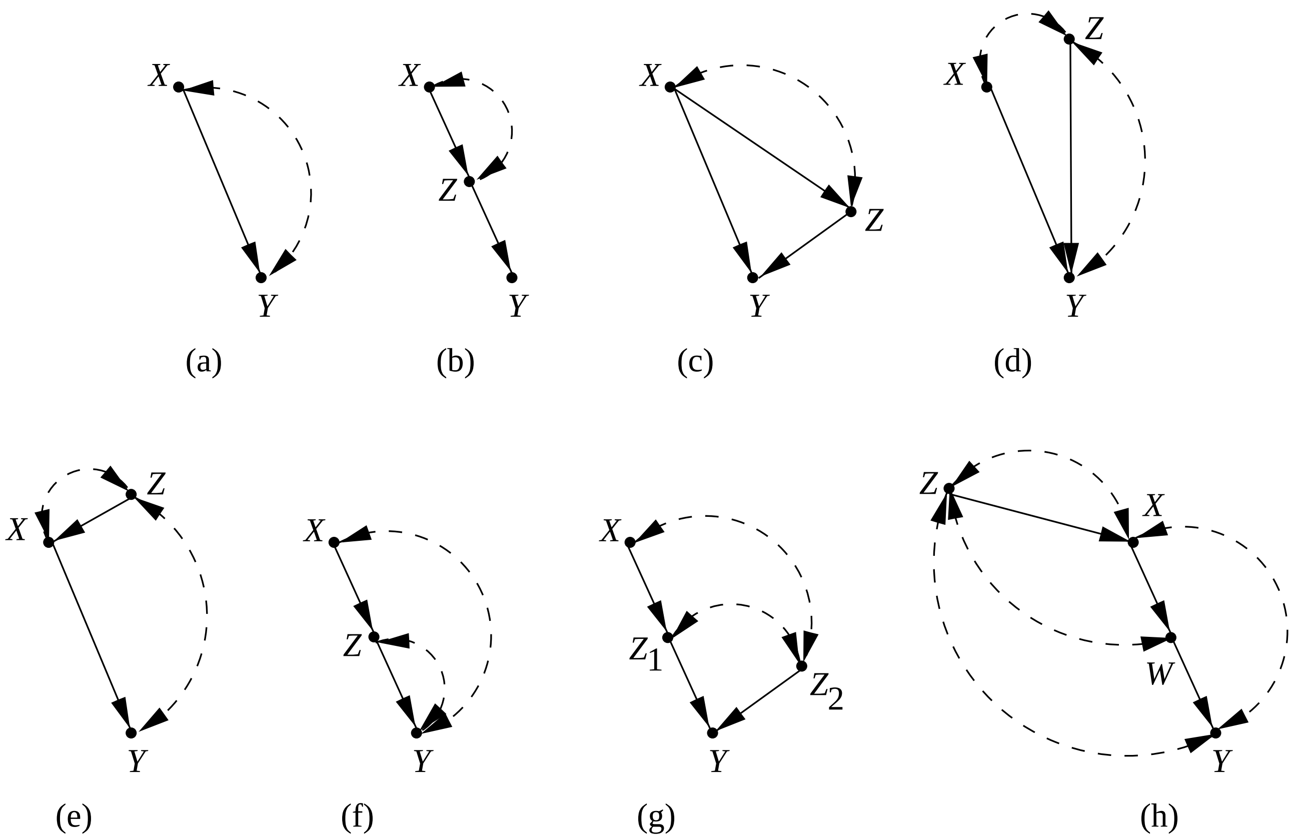

Figure 3.9: Typical models in which $P(y|\hat{x})$ is not identifiable.