d-SEPARATION IN LINEAR MODELS

Theorem 5.2.1
(Verma and Pearl 1988; Geiger et al. 1990)
If sets X and Y are d-separated by Z in a DAG G, then X is independent of Y conditional on Z in every Markovian model structured according to G. Conversely, if X and Y are **not** d-separated by Z in a DAG G, then X and Y are dependent conditional on Z in **almost all** Markovian models structured according to G.

Corrolary 5.2.2
In any Markovian model structured according to a DAG G, the partial correlation $\rho_{XY \cdot Z}$ vanishes whenever the nodes corresponding to the variables in Z d-separate node X from node Y in G, regardless of the model's parameters. Moreover, no other partial correlation would vanish for all the model's parameters.

Theorem 5.2.3
(d-Separation in General Linear Model)
For any linear model structured according to a diagram D, which may include cycles and bidirected arcs, the partial correlation $\rho_{XY \cdot Z}$ vanishes if the nodes corresponding to the set of variables Z d-separate node X from node Y in D.

5.2.1 THE TESTABLE IMPLICATIONS OF STRUCTURAL MODELS

General structural equations

$$x_i = f_i(pa_i, \epsilon_i,) \qquad i = 1, \ldots, n, \qquad \textbf{(5.1)}$$

Linear structural equations

$$x_i = \sum_{k \neq i} \alpha_{ik} x_k + \epsilon_i, \qquad i = 1, \ldots, n, \qquad \textbf{(5.2)}$$

To be distinguished from regression equations

$$x_i = \sum_{k \neq i} r_{ik} x_k + \eta_i$$

where $\eta_i \perp\!\!\!\perp X_k$ is assumed for $k \neq i$, BY DEFINITION!

The coefficient of X in the regression of Y on $X, Z_1, \ldots, Z_k$

$$y = ax + b_1 Z_1 +, \ldots, + b_k z_k$$

is given by

$$a = r_{YX \cdot Z_1 Z_2, \ldots, Z_k}$$

5.2.2 TESTING THE TESTABLE

Definition 5.2.4 (Basis)

Let S be a set of partial correlations. A basis B for S is a set of zero partial correlations where (i) B implies (using the laws of probability) the zero of every element of S and (ii) no proper subset of B sustains such implication.

Theorem 5.2.5 (Graphical Basis)

Let (i, j) be a pair of nonadjacent nodes in a DAG D, and let Z_{ij} be any set of nodes that are closer to i than j is to i and such that Z_{ij} d-separates i from j. The set of zero partial correlations $B = \{\rho_{ij \cdot Z_{ij}} = 0 | i > j\}$, consisting of one element per nonadjacent pair, constitutes a basis for the set of all zero partial correlations entailed by D.

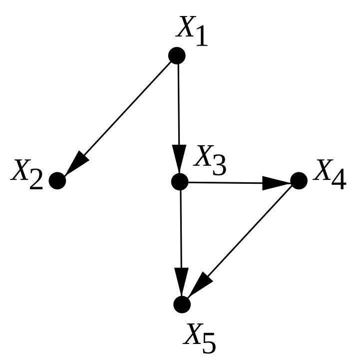


Figure 5.1

$$\begin{aligned} B_1 &= \{\rho_{32\cdot 1} = 0, \rho_{41\cdot 3} = 0, \rho_{42\cdot 3} = 0, \rho_{51\cdot 43} = 0, \\ &\qquad \rho_{52\cdot 43} = 0\} \\ B_2 &= \{\rho_{32\cdot 1} = 0, \rho_{41\cdot 3} = 0, \rho_{42\cdot 1} = 0, \rho_{51\cdot 3} = 0, \\ &\qquad \rho_{52\cdot 1} = 0\} \end{aligned} \tag{5.3}$$

PROBLEMS WITH TRADITIONAL (GLOBAL) TESTING

1. If some parameters are not identifiable, then the first phase may fail to reach stable estimates for the parameters and the investigator must simply abandon the test.

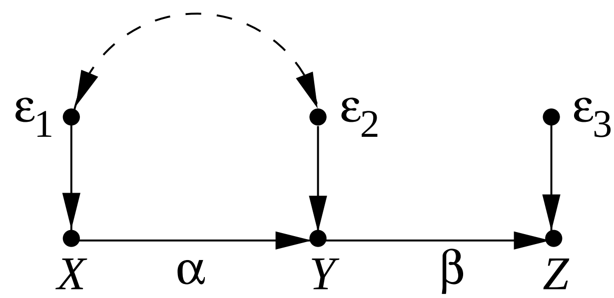


Figure 5.2

2. If the model fails to pass the data-fitness test, the investigator receives very little guidance about which modeling assumptions are wrong.

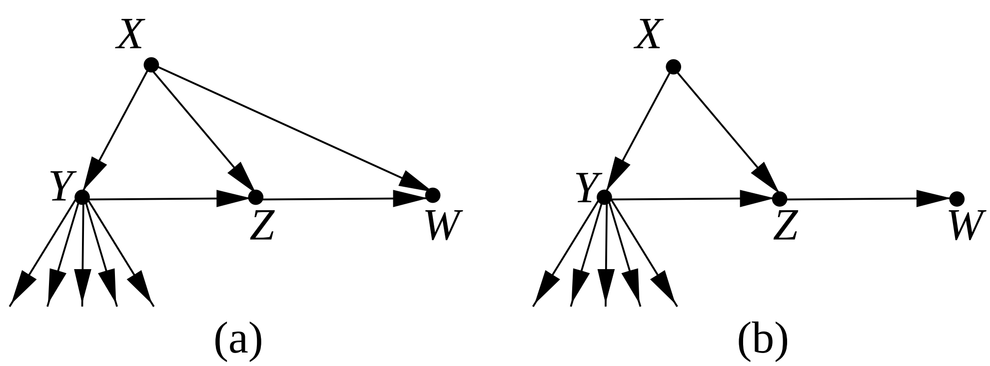


Figure 5.3

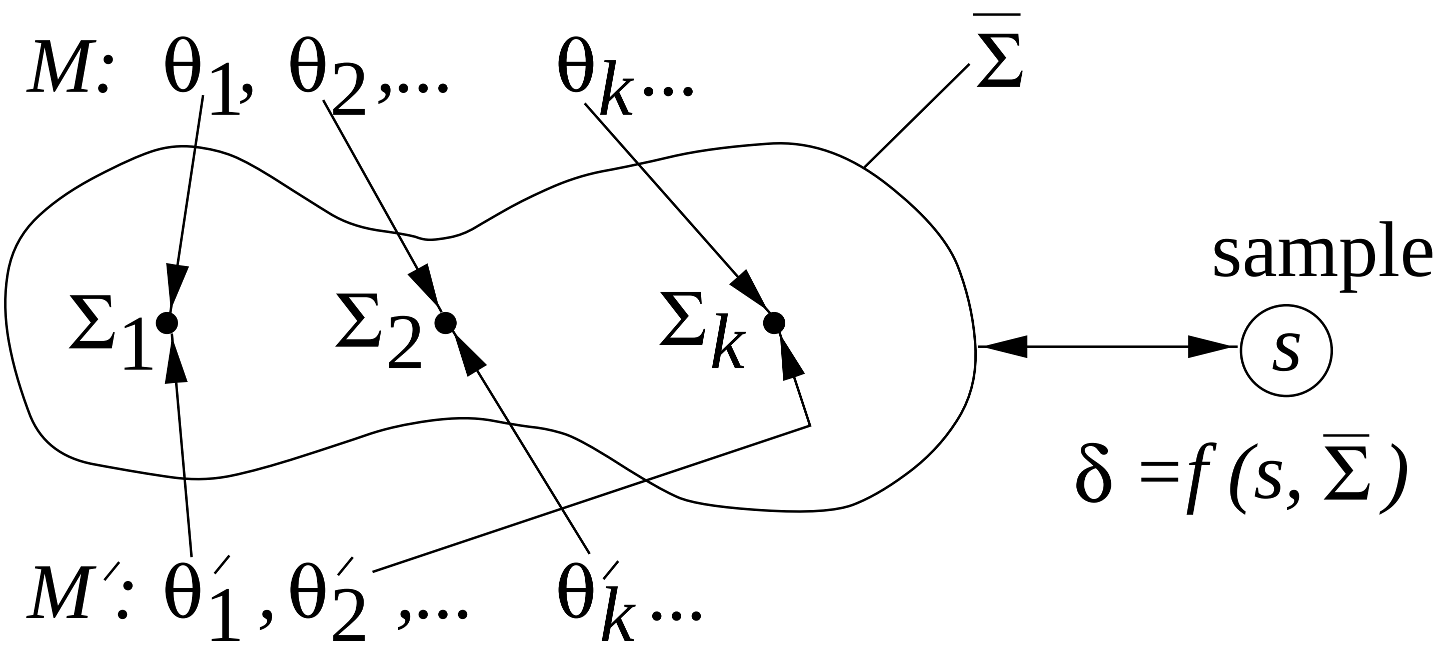
5.2.3 MODEL EQUIVALENCE

Definition
(Observational Equivalence)
Two SEM's are **observationally equivalent** if every probability distribution that is generated by one can also be generated by the other.

Theorem 1.2.8 (Review)
(Verma and Pearl 1990)
Two Markovian models are observationally equivalent iff they entail the same sets of conditional independencies. Moreover, two such models are observationally equivalent iff their corresponding graphs have the **same sets of edges** and the **same sets of v-structures** (two converging arrows whose tails are not connected by an arrow).

Theorem 5.2.6 (Covariance Equivalence)
Two Markovian linear-normal models are covariance equivalent if and only if they entail the same sets of zero partial correlations. Moreover, two such models are covariance equivalent if and only if their corresponding graphs have the same sets of edges and the same sets of v-structures.

WHY COVARIANCE-EQUIVALENCE IMPLIES FITNESS-EQUIVALENCE (FOR ALL DATA)



$$\delta' = f(s, \Sigma')$$

$$\delta = \delta' \quad \text{whenever} \quad \Sigma' = \Sigma$$

THE SIGNIFICANCE OF EQUIVALENT MODELS

What does it mean to "test a model"?

Some models have testable implications.

The same testable implications are shared by a whole class of equivalent models.

"Testing a model M" means "testing the class of models equivalent to M"

Does this make SEM useless for causal modeling?

Qualitative assumptions + data = qualitative conclusions

GENERATING EQUIVALENT MODELS NECESSARY CONDITIONS

Rule 1: An arrow $X \rightarrow Y$ is interchangeable with $X \leftarrow - - - \rightarrow Y$ only if every neighbor or parent of X is inseparable from Y. (By **neighbor** we mean a node connected (to X) through a bidirected arc.)

Rule 2: An arrow $X \rightarrow Y$ can be reversed into $X \leftarrow Y$ only if, before reversal, (i) every neighbor or parent of Y (excluding X) is inseparable from X and (ii) every neighbor or parent of X is inseparable from Y.

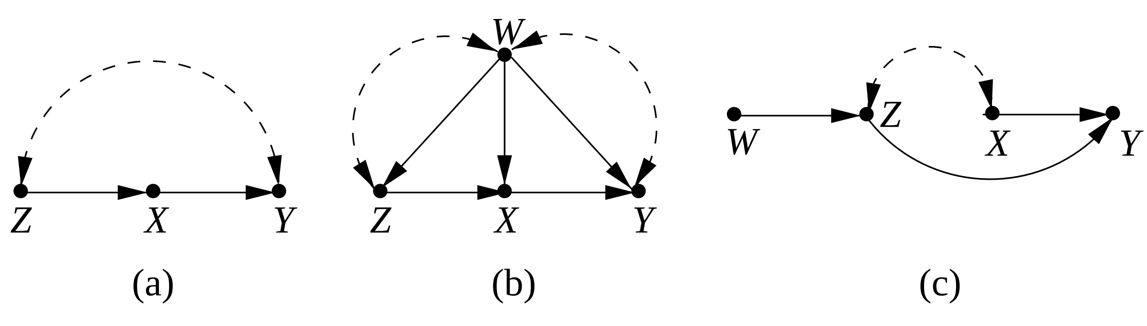


Figure 5.4

5.3.1 PARAMETER IDENTIFICATION IN LINEAR MODELS

Wright Rule (1923):

r_{XY} = Sum of products of path coefficients along all collider-free paths between X and Y.

If there is an edge $X \xrightarrow{\alpha} Y$ in the model then:

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

where I_{YX} is independent of α.

Thus, $\alpha = r_{YX}$ if X and Y are d-separated in G_α

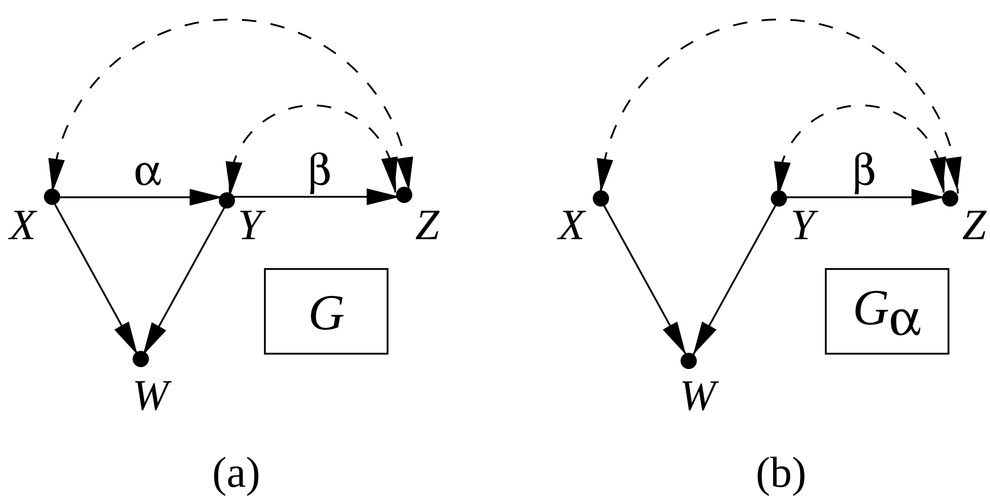


Figure 5.6

SINGLE LINK CRITERION

(for the direct identification
of a structural parameter)

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 $X \to Y$, and let G_α denote the diagram that results when $X \to Y$ 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).

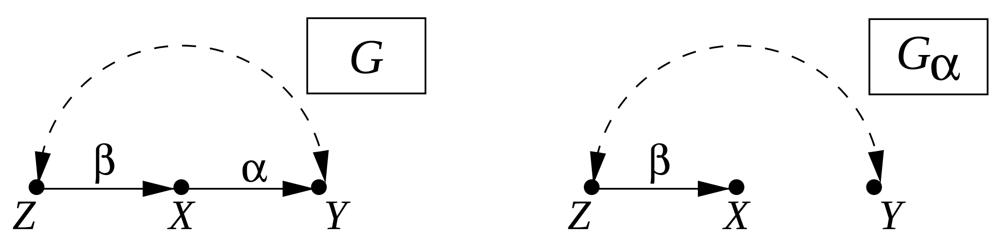


Figure 5.7: The identification of α with $r_{YX \cdot Z}$.

PARAMETER IDENTIFICATION WITH BACK-DOOR CRITERION

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 $r_{YX \cdot Z}$.

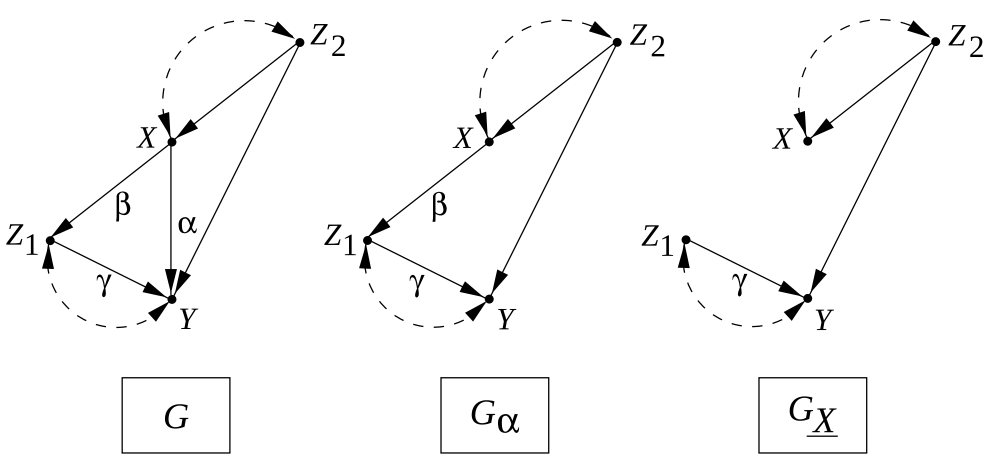


Figure 5.8: $\alpha + \beta\gamma = r_{YX \cdot Z_2}$

INDIRECT IDENTIFICATION OF STRUCTURAL PARAMETERS

e.g, Instrumental variables:

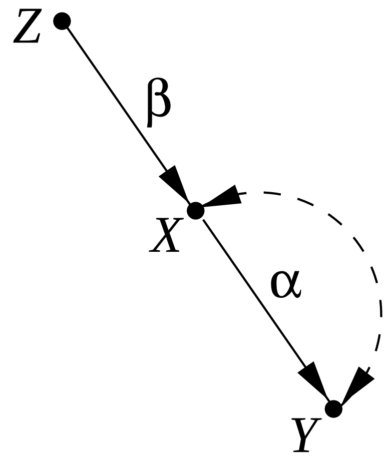


Figure 5.9: Graphical identification of α using instrumental variable Z, $\alpha = r_{YZ}/r_{XZ}$.

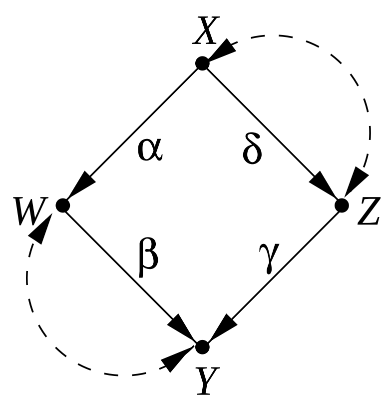


Figure 5.10 Graphical identification of α, β, and γ.

$$\begin{aligned}
\alpha\beta &= r_{YX\cdot Z} \\
\beta &= r_{YX\cdot Z}/r_{WX} \\
\gamma &= r_{YZ\cdot X}
\end{aligned}$$

BUCKET ELIMINATION PROCEDURE

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.

PRACTICAL QUESTIONS THAT FOLLOW

1. When are two structural equation models observationally indistinguishable?

2. When do regression coefficients represent path coefficients?

3. When would the addition of a regressor introduce bias?

4. How can we tell, prior to taking any data, which path coefficients can be identified?

5. When can we dispose of the linearity-normality assumption and still extract causal information from the data?

SETTLED FOUNDATIONAL QUESTIONS

1. Under what conditions can we give causal interpretation to structural coefficients?

2. What are the causal assumptions underlying a given structural equation model?

3. What are the statistical implications of any given structural equation model?

4. What is the operational meaning of a given structural coefficient?

5. What are the policy-making claims of any given structural equation model?

6. When is an equation non-structural?