**Study question 3.8.1.**
*Consider the causal model of Figure 3.10.*

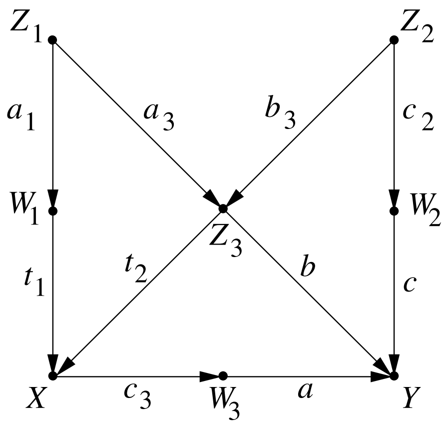


Figure 3.10

(a) *Identify three testable implications of this model.*

(b) *Identify a testable implication assuming that only $X, Y, W_3$, and $Z_3$ are observed.*

(c) *For each of the parameters in the model, write a regression equation in which one of the coefficients is equal to that parameter. Identify the parameters for which more than one such equation exists.*

(d) *Suppose $X, Y$ and $W_3$ are the only variables observed. Which parameters can be identified from the data? Can the total effect of $X$ on $Y$ be estimated?*

(e) *If we regress $Z_1$ on all other variables in the model, which regression coefficient will be zero?*

(f) *The model in Figure 3.10 implies that certain regression coefficients will remain invariant when an additional variable is added as a regressor. Identify five such coefficients with their added regressors.*

(g) *Assume that variables $Z_2$ and $W_2$ cannot be measured. Find a way to estimate $b$ using regression coefficients. [Hint: Find a way to turn $Z_1$ into an instrumental variable for $b$.]*

**Solution to study question 3.8.1**

**Part (a)**

Testable implications are conditional independence relationships implied by the structure of the graph. These conditional independences translate into vanishing regression coefficients in the data. Examining Figure 3.10, three regression equations that could be used to test the model could be:

1. $W_3$ is independent of $W_1$ given $X$, giving us the regression equation: $W_3 = r_X X + r_{W_1} W_1$ with $r_{W_1} = 0$. This means that if we fit the data to the line $W_3 = r_X X + r_{W_1} W_1$, we expect to find $r_{W_1} = 0$, or else the model is wrong.
2. $W_1$ is independent of $Z_3$ given $Z_1$, giving us the regression equation: $W_1 = r_{Z_1} Z_1 + r_{Z_3} Z_3$ with $r_{Z_3} = 0$
3. $Y$ is independent of $Z_1$ given $W_1, Z_2$, and $Z_3$, giving us the regression equation: $Y = r_{Z_1} Z_1 + r_{W_1} W_1 + r_{Z_2} Z_2 + r_{Z_3} Z_3$ with $r_{Z_1} = 0$

**Part (b)**

The only conditional independence that involves the measured variables is the one between $Z_3$ and $W_3$ given $X$, which leads to $r_{Z_3} = 0$ in the corresponding regression equation:

$$W_3 = r_{Z_3} Z_3 + r_X X \text{ with } r_{Z_3} = 0$$

**Part (c)**

(i) If we regress a variable on its parents, we get a regression equation whose coefficients equal the model parameters. Therefore:

1. $a = r_{W_3}, b = r_{Z_3}, c = r_{W_2}$ in the equation:

$$Y = r_{W_3} W_3 + r_{Z_3} Z_3 + r_{W_2} W_2$$

2. $a_1 = r_{Z_1}$ in:

$$W_1 = r_{Z_1} Z_1$$

3. $a_3 = r_{Z_1}, b_3 = r_{Z_2}$ in:

$$Z_3 = r_{Z_1} Z_1 + r_{Z_2} Z_2$$

4. $c_2 = r_{Z_2}$ in:

$$W_2 = r_{Z_2} Z_2$$

5. $c_3 = r_X$ in:

$$W_3 = r_X X$$

6. $t_1 = r_{W_1}, t_2 = r_{Z_3}$:

$$X = r_{W_1} W_1 + r_{Z_3} Z_3$$

(ii) The "Regression Rule for Identification" tells us that, if $G_\alpha$ has several backdoor sets, each would lead to a regression equation in which $\alpha$ is a coefficient. Therefore, $a, b, c$ can be identified by:

1. $a = r_{W_3}, b = r_{Z_3}, c = r_{W_2}$ in:

$$Y = r_{W_3}W_3 + r_{Z_3}Z_3 + r_{W_2}W_2$$

Or, by $a = r_{W_2}, b = r_{Z_3}, c = r_{Z_2}/c_2$ in:

$$Y = r_{W_3}W_3 + r_{Z_3}Z_3 + r_{Z_2}Z_2$$

2. Likewise, $t_1$ can be identified either by $t_1 = r_{W_1}$ in:

$$X = r_{W_1}W_1 + r_{Z_1}Z_1$$

Or, by $t_1 = r_{W_1}$ in:

$$X = r_{W_1}W_1 + r_{Z_3}Z_3$$

**Part (d)**

To determine which parameters are estimable from data, we consult "The Regression Rule for Identification." For example, the parameter $c_3$ can be estimated from data because $W_3 = r_X X + U_3' = c_3 X + U_3'$, since $W_3$ is $d$-separated from $Y$ given $X$ in $G_{W_3}$. Likewise, $a = r_{YW_3 \cdot X}$.

Lastly, we note that $W_3$ is a front-door admissible variable for attaining the total effect of $X$ on $Y$, and so the effect is estimable. Indeed the total effect of $X$ on $Y$ is simply the product of $a * c_3$, which we identified above.

**Part (e)**

Regressing $Z_1$ on all other variables in the model gives:

$$Z_1 = r_{Z_2}Z_2 + r_{Z_3}Z_3 + r_X X + r_{w_1}W_1 + r_{W_2}W_2 + r_{W_3}W_3 + r_Y Y$$

By $d$-separation, we see that $Z_1$ is independent of $\{X, W_3, Y, W_2\}$ given $W_1, Z_3, Z_2$. Therefore, $r_X = 0, r_{W_2} = 0, r_{W_3} = 0, r_Y = 0$

**Part (f)**

In order for a coefficient to remain invariant under the addition of a new regressor, the dependent variable must be independent of the added regressor given all of the old regressors.

Thus, for example, if we regress $W_1$ on $Z_3$ and $X$, adding $W_3$ would keep all regression coefficients in tact, but adding $Y$ or $Z_2$ would change them, because of the path: $Y \leftarrow W_2 \leftarrow Z_2 \rightarrow Z_3 \leftarrow Z_1 \rightarrow W_1$ is opened by conditioning on $Z_3$. If we regress $W_1$ on $Z_1$, then we can add $Z_3, Z_2$, or $W_2$ without changing the regression coefficient.

**Part (g)**

Note that if we condition on $W_1$, we turn $Z_1$ into an instrument relative to the effect $\tau$ of $Z_3$ on $Y$. Using this idea, we can write the regression of $Y$ on $Z_1$ given $W_1$, as the product $\tau a_3$ where $\tau = t_2 c_3 a + b$. Since each of $t_2, c_3, a$ can be separately identified (see Part a above), we can then solve for $b$. Formally, we have:

$$t_2 c_3 a + b = r_{Z_1} / r_{Z_1'}$$

Where $r_{Z_1}$ and $r_{Z_1'}$ are the regression coefficients in the following equations:

$$Y = r_{Z_1} Z_1 + r_{W_1} W_1 + \epsilon$$
$$Z_3 = r_{Z_1'} Z_1 + \epsilon$$