MARTÍN BATISTA - SEMINARIO INFERENCIA CAUSAL

# CONTRAFÁCTICOS

- Queremos ir al centro en auto, surgen dos posibilidades:
  - Ir por Av. Italia (X = 0)
  - Ir por la Rambla (X = 1)
- Optamos por ir por Av. Italia, demoramos 1 hora.
  - Capaz que por la Rambla hubiese sido mejor, no?
  - No podemos volver atrás en el tiempo ni replicar exactamente la situación previa y tomar otra decisión.
  - Ésto es un contrafáctico. Tenemos una condición hipotética (haber tomado la Rambla) y un antecedente (habiendo tomado Av. Italia.

- Buscamos algo como:
  - $\mathbb{E}[\text{tiempo de viaje} \,|\, do(\text{Rambla}), \text{tiempo de viaje} = 1 \text{ hora}]$
- Hay dos problemas:
  - No podemos utilizar el operador *do* ya que no podemos intervenir en una situación que ocurrió en el pasado y no es replicable.
  - Un problema de notación, si queremos calcular el tiempo de viaje esperado:
    - Precisamos distinguir entre los distintos *tiempo de viaje*.
    - Si escribimos:
      - $Y_{X=1}$ ó $Y_1$ el tiempo de viaje al haber tomado la rambla (X = 1).
      - $Y_{X=0}$ ó $Y_0$ el tiempo de viaje al haber ido por Av. Italia (X = 0).
    - Tenemos que lo que buscamos es:

$$\mathbb{E}[Y_{X=1} \,|\, X = 0, \, Y = Y_0 = 1]$$

- $\mathbb{E}[Y_{X=1} \,|\, X = 0, \, Y = Y_0 = 1]$
- Tiene sentido esto?
  - $Y_{X=1} = y$ e $X = 0$ son eventos que ocurren en mundos diferentes.
- En que difiere de:
  - $\mathbb{E}[Y \,|\, do(X = x)]$ ?

- Rambla: $X = 1$
- Av. Italia: $X = 0$

- $\mathbb{E}[Y_{X=1} \,|\, X = 0,\ Y = Y_0 = 1]$
- Tiene sentido esto?
  - $Y_{X=1} = y$ e $X = 0$ son eventos que ocurren en mundos diferentes.
- En que difiere de:
  - $\mathbb{E}[Y \,|\, do(X = x)]$ ?
    - Esto es una estimación del tiempo de viaje si tomamos una calle *independientemente* de cualquier otra circunstancia.
- Al no poder tomar ambas calles simultáneamente, no logramos reducir el problema con expresiones *do*.
- Y si al siguiente día tomamos la otra calle? En el mejor de los casos, si las situaciones son bastante similares, es sólo una aproximación.

- Rambla: $X = 1$
- Av. Italia: $X = 0$

- Similar a la clase pasada, podemos utilizar SCM's para encontrar la respuesta, si tenemos un modelo causal completamente especificado y conocemos los valores de las variables exógenas.
- Consideremos la siguiente sentencia contrafáctica:
  - *Y hubiese sido y si X fuese x bajo la situación U = u*
  - Lo escribimos: $Y_x(u) = y$
  - Ejemplo: Y hubiese sido 30 minutos si hubiésemos tomado la rambla.
- "si *X fuese x*":
  - El valor real de *X* sería *X(u)*
  - Podemos analizar el efecto que tiene el fijar *X* = *x* en el SCM, es similar a *do(X* = *x)* pero bastante diferente. El operador *do* captura el comportamiento a nivel poblacional en cuanto los contrafácticos $Y_x(u) = y$ describen el comportamiento a nivel del individuo.

- Un modelo simple con tres variables: *X*, *Y*, *U*

$$X = aU$$

$$Y = bX + U$$

- Primero calculamos $Y_x(u) = y$
  - El valor de *Y* si *X* fuese x bajo la situación *U* =*u*

$$X = x$$

$$Y = bX + U$$

- Sustituyendo *U* = *u* y resolviendo para *Y*

$$Y_x(u) = bx + u$$

**Table 4.1** The values attained by $X(u)$, $Y(u)$, $Y_x(u)$, and $X_y(u)$ in the linear model of Eqs. (4.3) and (4.4)

| $u$ | $X(u)$ | $Y(u)$ | $Y_1(u)$ | $Y_2(u)$ | $Y_3(u)$ | $X_1(u)$ | $X_2(u)$ | $X_3(u)$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 2 | 3 | 4 | 1 | 1 | 1 |
| 2 | 2 | 4 | 3 | 4 | 5 | 2 | 2 | 2 |
| 3 | 3 | 6 | 4 | 5 | 6 | 3 | 3 | 3 |

- Que pasa con el contrafáctico $X_y(u)$ ?
  - El valor que toma *X* si *Y* hubiese sido *y* bajo las condiciones *U* = *u*
  - $X_y(u) = au$
  - *X* no es alterada por los valores que toma y. Tiene sentido?
    - Hipotetizar sobre el futuro no debería afectar el pasado.

## LEY FUNDAMENTAL DE LOS CONTRAFÁCTICOS

- Dado un modelo causal estructural $M$.
- Sean dos variables $X$ e $Y$ no necesariamente conectadas mediante una única ecuación.
- LLamamos $M_x$ (submodelo) a la versión modificada de $M$ con la ecuación de $X$ sustituida por $X = x$.
- La definición formal del contrafáctico $Y_x(u)$ es:

$$Y_x(u) = Y_{M_x}(u)$$

- O sea:
  - El contrafáctico $Y_x(u)$ en el modelo $M$ es la solución de $Y$ en el submodelo $M_x$.
  - Si tenemos un modelo causal que describe una situación dada, podemos utilizarlo responder preguntas hipotéticas de que hubiese ocurrido si X hubiese tomado otro valor. Si hubiésemos ido por la Rambla.

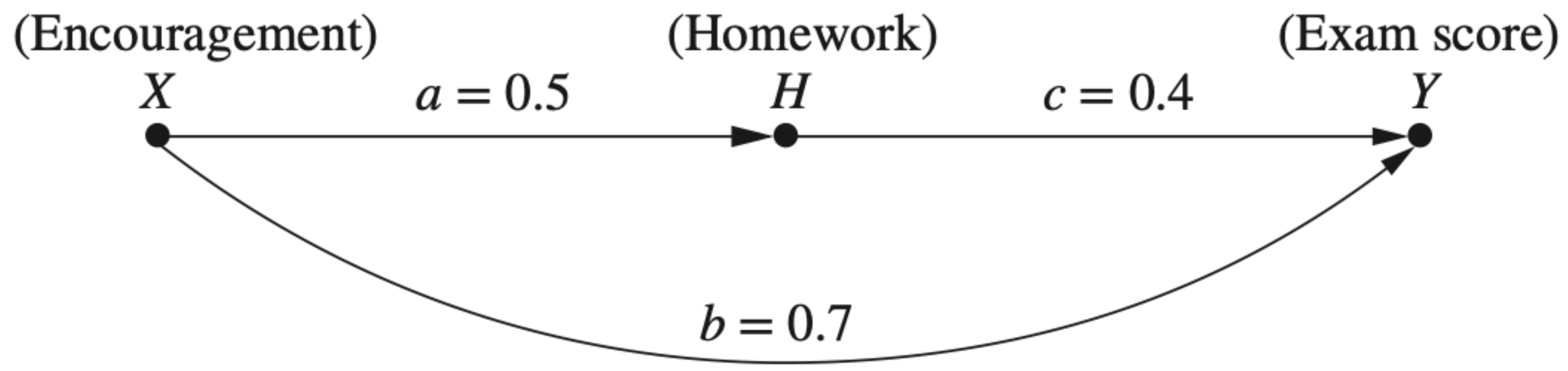



**Figure 4.1** A model depicting the effect of Encouragement ($X$) on student's score

$$X = U_X$$

$$H = a \cdot X + U_H$$

$$Y = b \cdot X + c \cdot H + U_Y$$

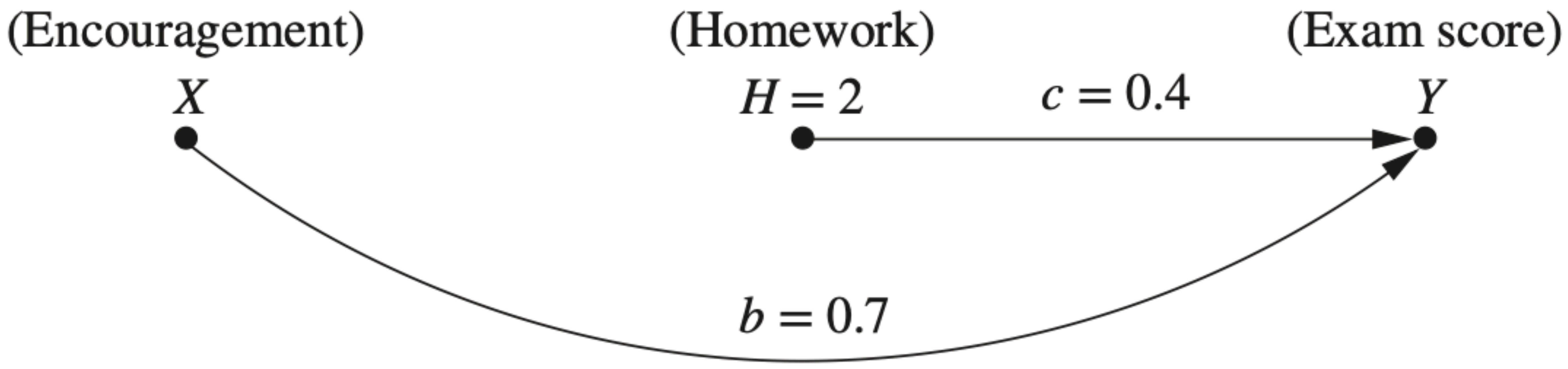
$$\sigma_{U_i U_j} = 0 \quad \text{for all } i, j \in \{X, H, Y\}$$

$$a = 0.5, \quad b = 0.7, \quad c = 0.4$$

# EJEMPLO MOTIVACIONAL

- Hay un estudiante, Juan tal que medimos: *X = 0.5, H =1, Y = 1.5*
- La pregunta es: Cual serian las notas de Juan si hubiese duplicado su tiempo de estudio?

$$
\begin{aligned}
X &= U_X \\
H &= \cancel{a \cdot X + U_H}\ 2 \\
Y &= b \cdot X + c \cdot H + U_Y \\
\sigma_{U_i U_j} &= 0 \quad \text{for all } i, j \in \{X, H, Y\}
\end{aligned}
$$




- Características de Juan (evidencia):

$$
\begin{aligned}
U_X &= 0.5, \\
U_H &= 1 - 0.5 \cdot 0.5 = 0.75, \text{ and} \\
U_Y &= 1.5 - 0.7 \cdot 0.5 - 0.4 \cdot 1 = 0.75.
\end{aligned}
$$

- Sustituyendo *H = 2* en la ecuación SCM.

$$
\begin{aligned}
Y_{H=2}(U_X = 0.5, U_H = 0.75, U_Y = 0.75) &= 0.5 \cdot 0.7 + 2.0 \cdot 0.4 + 0.75 \\
&= 1.90
\end{aligned}
$$

# 3 PASOS PARA CALCULAR CONTRAFÁCTICOS DETERMINISTAS

- **Abducción**

  Usar evidencia $E = e$ para determinar el valor de $U$.

  *Analogía temporal: Esto explica el pasado según la evidencia e.*

- **Acción**

  Modificar el modelo $M$, removiendo las ecuaciones estructurales de la variable $X$ y reemplazándolas por las funciones apropiadas $X = x$ para obtener el submodelo $M_x$.

  *Analogía temporal: Esto cambia el pasado en acorde con el antecedente X = x.*

- **Predicción**

  Usar el modelo modificado $M_x$, el valor de $U$ para calcular el valor de $Y$, la consecuencia del contrafáctico.

  *Analogía temporal: Predecimos el futuro en base a las modificaciones.*

- A veces queremos responder preguntas probabilistas, por ejemplo: ¿Qué probabilidad hay de que Juan hubiese sacado *$Y = y'$ si se hubiese motivado por 5 horas más?*

- El problema es indeterminado ya que hay muchos valores consistentes con la evidencia *e*.

- Al asignar probabilidades *$P(U = u)$* sobre las variables exógenas introducimos indeterminismo al sistema.

- En el caso indeterminado, planteamos las preguntas contrafácticas de la siguiente manera:

$$\mathbb{E}[Y_{X=x} \mid E = e]$$

  - Dado que observamos evidencia $E = e$, qué esperaríamos que fuese el valor de *$Y$ si $X$ hubiese sido $x$?*

(i) **Abduction**: Update $P(U)$ by the evidence to obtain $P(U|E = e)$.

(ii) **Action**: Modify the model, $M$, by removing the structural equations for the variables in $X$ and replacing them with the appropriate functions $X = x$, to obtain the modified model, $M_x$.

(iii) **Prediction**: Use the modified model, $M_x$, and the updated probabilities over the $U$ variables, $P(U|E = e)$, to compute the expectation of $Y$, the consequence of the counterfactual.

- Consideremos $U = \{1, 2, 3\}$ tres tipos distintos de individuos en una población, que ocurren con las probabilidades:

$$P(U=1) = \frac{1}{2}, P(U=2) = \frac{1}{3}, \quad \text{and} \quad P(U=3) = \frac{1}{6}$$

$$X = aU$$

$$Y = bX + U$$

- Con que frecuencia $Y = 3$ *ocurre si* $X$ *hubiese sido 2? O sea:* $Y_2(u) = 3$
- Buscando en la tabla, vemos que es una propiedad de $U = 1$ que ocurre con probabilidad 1/2.

$$P(Y_2 = 3) = \frac{1}{2}$$

- Analogamente:

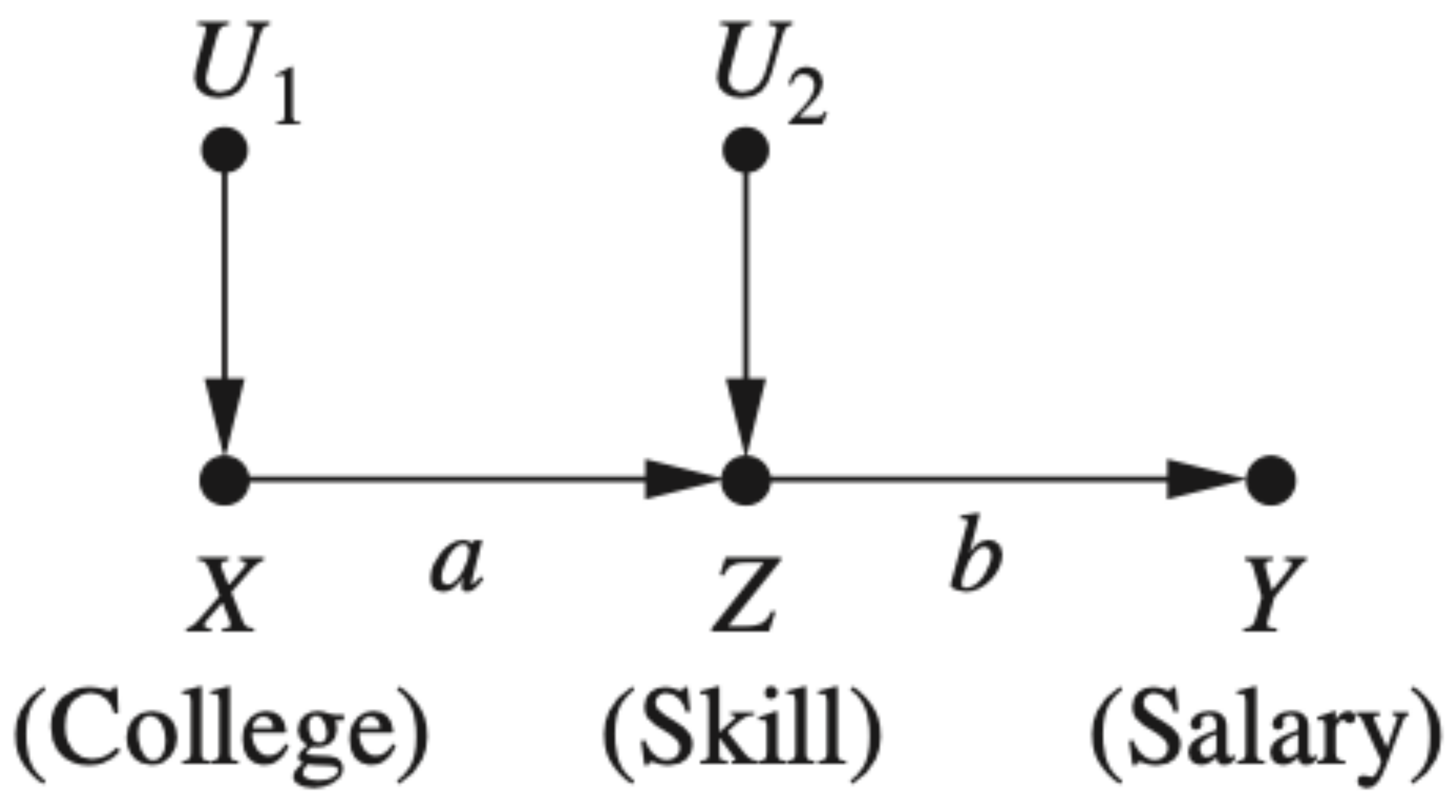
$$P(Y_1 = 4) = \frac{1}{6}, P(Y_1 = 3) = \frac{1}{3}, P(Y_2 > 3) = \frac{1}{2}$$

| $u$ | $X(u)$ | $Y(u)$ | $Y_1(u)$ | $Y_2(u)$ | $Y_3(u)$ | $X_1(u)$ | $X_2(u)$ | $X_3(u)$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 2 | 3 | 4 | 1 | 1 | 1 |
| 2 | 2 | 4 | 3 | 4 | 5 | 2 | 2 | 2 |
| 3 | 3 | 6 | 4 | 5 | 6 | 3 | 3 | 3 |

- Consideremos:

$$X = U_1 \quad Z = aX + U_2, Y = bZ$$

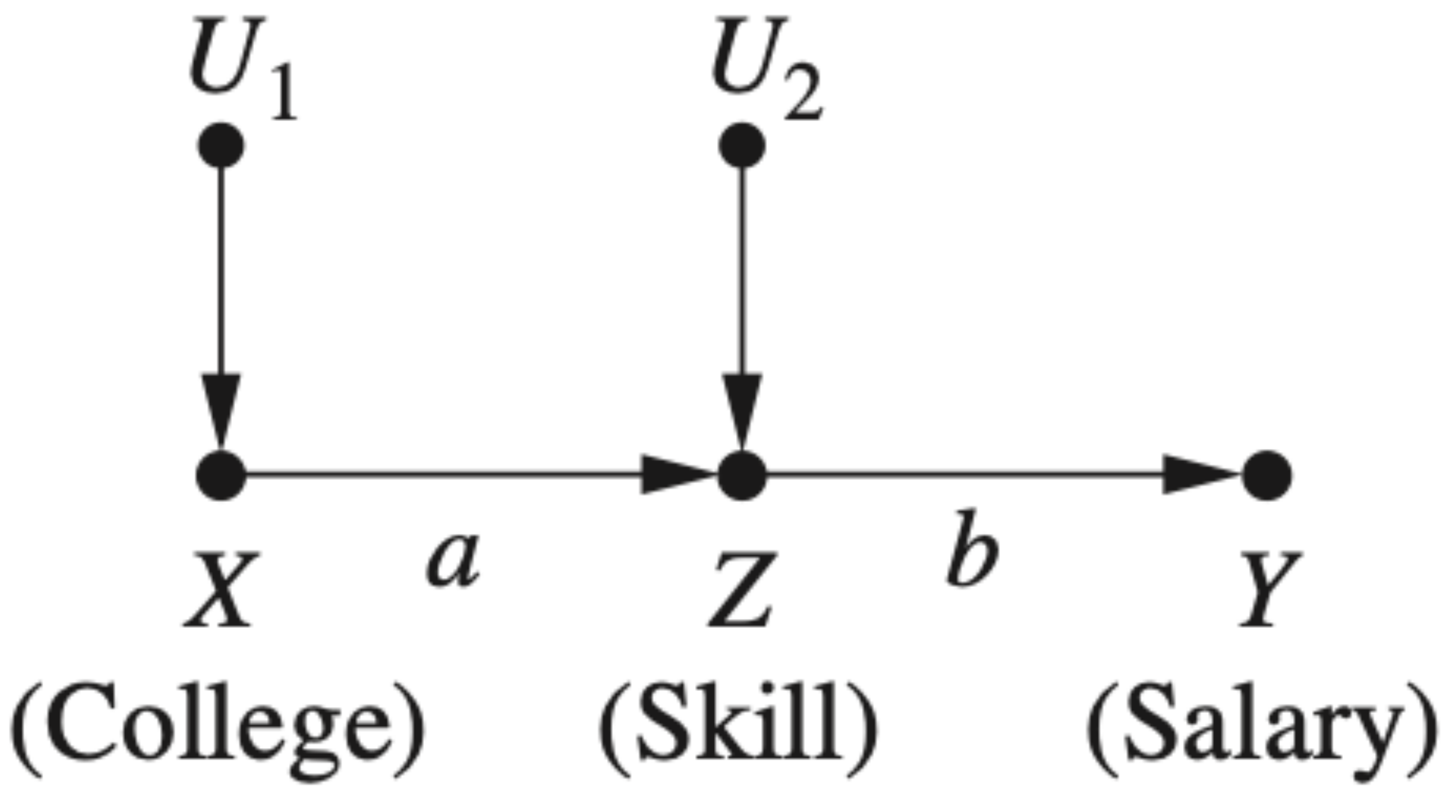
- Sea:
  - $X = 1$ tener una educación terciaria.
  - $U_2 = 1$ experiencia profesional.
  - $Z$ el nivel de habilidad..
  - $Y$ el salario.




- Queremos calcular: $\mathbb{E}[Y_{X=1} | Z = 1]$.
  - El salario de individuos con habilidad Z = 1 si hubiesen tenido una educación terciaria.
- La condición Z = 1 y X = 1 vienen de mundos diferentes. Z = 1es el nivel de habilidad actual, y X =1 es un hipotético de un pasado no realizado.

- Porque no podemos usar $\mathbb{E}[Y|do(X=1), Z=1]$ ?

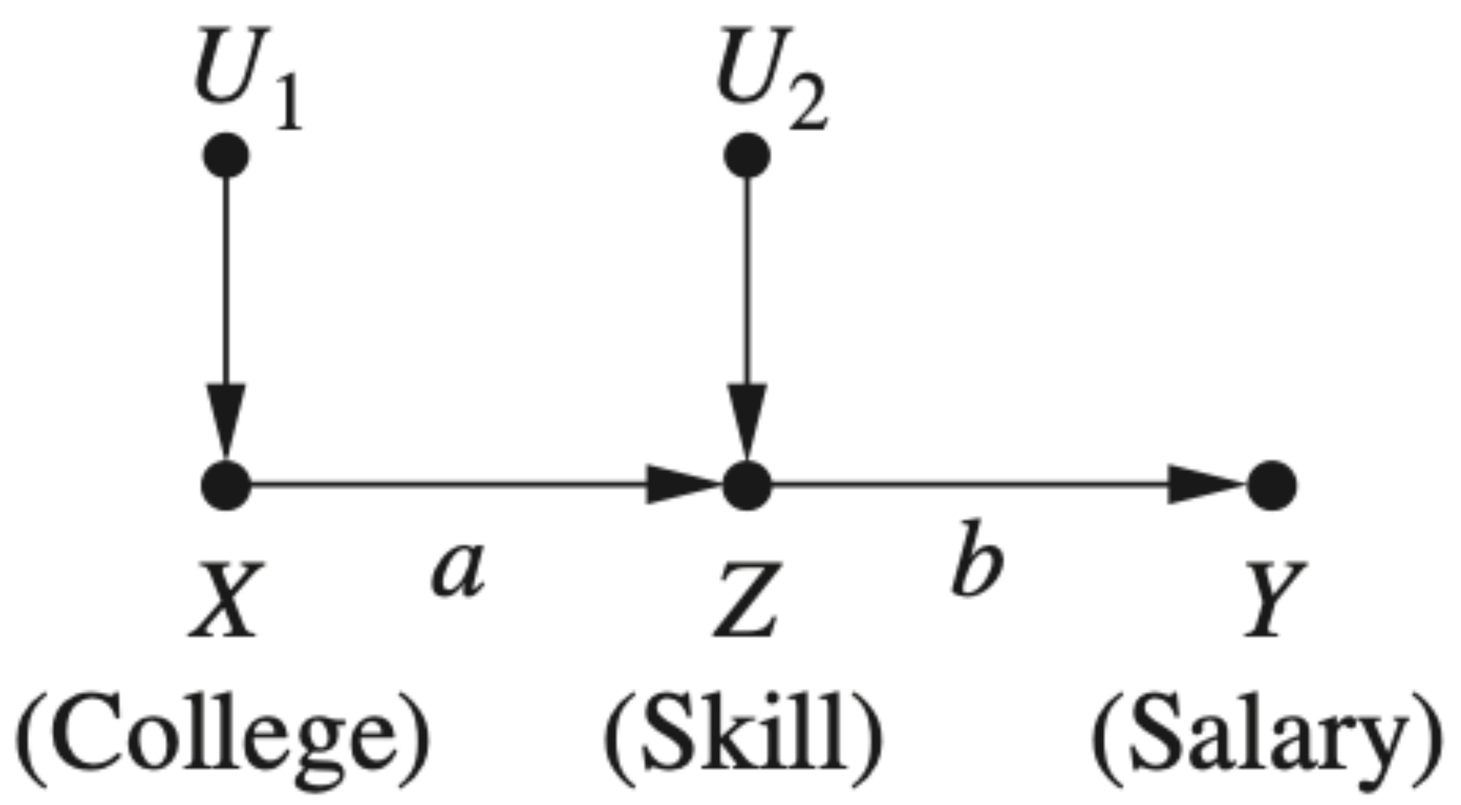
$$X = U_1 \quad Z = aX + U_2, Y = bZ$$



- $\mathbb{E}[Y_{X=1} | Z = 1]$?
  - El salario de individuos con habilidad Z = 1 si hubiesen tenido una educación terciaria.

$$X = U_1 \quad Z = aX + U_2, Y = bZ$$

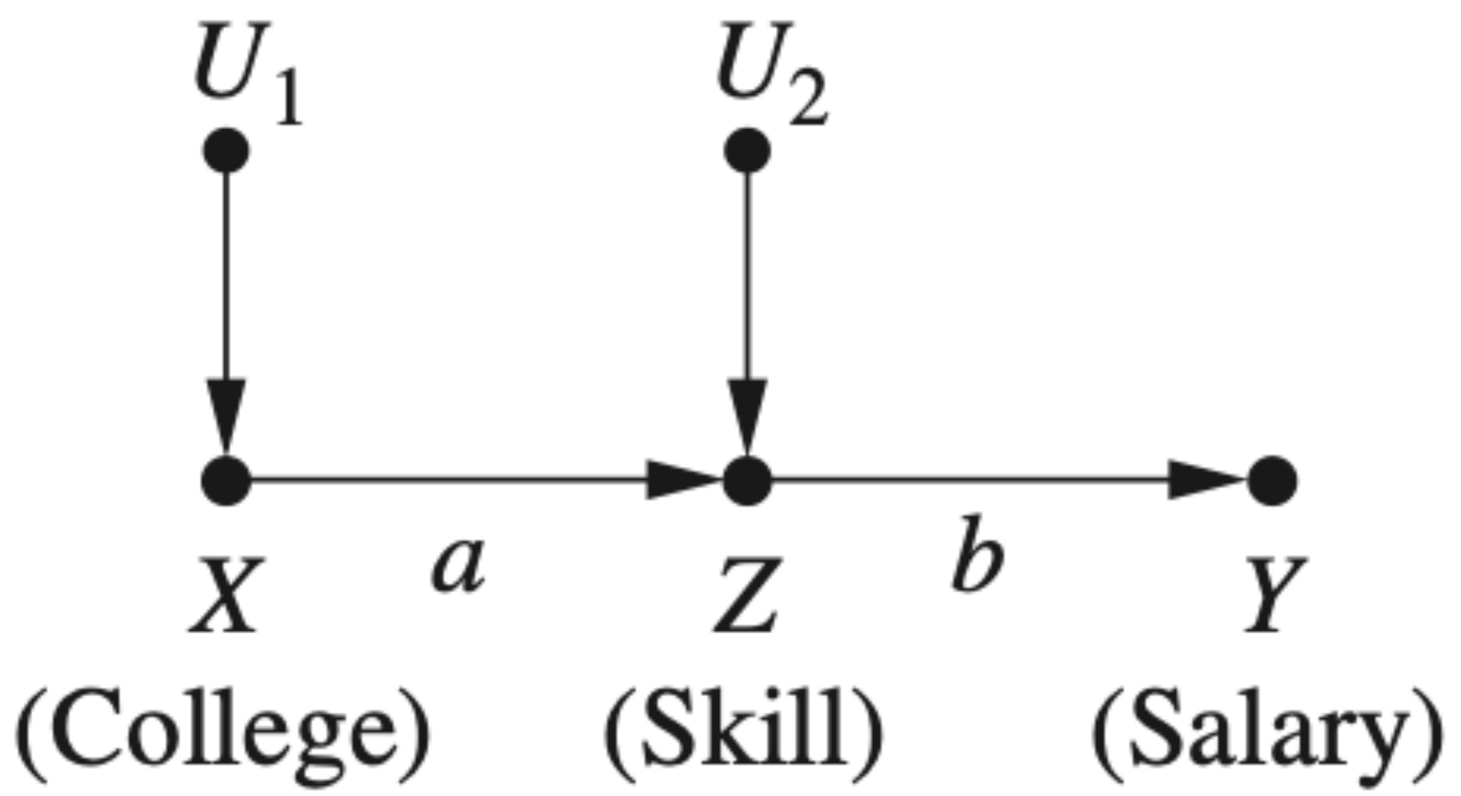
- Porque no podemos usar $\mathbb{E}[Y | do(X = 1), Z = 1]$ ?
  - Seria el salario esperado de los individuos que terminaron la universidad y desde entonces adquirieron un nivel de habilidad Z = 1.
  - Al condicionar en Z = 1 (mediador) en este caso, corta el efecto de la intervención que nos interesa.
  - El salario no depende de la educación directamente, sino de la habilidad que pudo ser adquirida con experiencia previa.



- Generalmente:

$$E[Y|do(X=1),\ Z=1] \neq E[Y_{X=1}|Z=1]$$

- Observemos que:
  - $\mathbb{E}[Y|do(X=1), Z=1] = \mathbb{E}[Y|do(X=0), Z=1]$




- En cambio:
  - $\mathbb{E}[Y_{X=1}|Z=1] \neq \mathbb{E}[Y_{X=0}|Z=1]$
  - Refiere a dos mundos diferentes. Pre-intervención (X =1) y post-intervención.
  - En cambio la expresión *do* sólo puede hacer referencia a eventos post-intervención.

- Al ser más flexibles, los contrafácticos pueden representar la expresión de post-intervención del operador *do:*

$$\mathbb{E}[Y \mid do(X=1), Z=1]$$

- De hecho:
  - $\mathbb{E}[Y \mid do(X=1), Z=1] = \mathbb{E}[Y_{X=1} \mid Z_{X=1} = 1]$
  - $Z_{X=1}$ son los valores que toma Z si X hubiese sido 1.
  - Representa lo mismo cuando ponemos Z = z en una expresión *do,* por la regla de Bayes:

$$P(Y=y|do(X=1), Z=z) = \frac{P(Y=y, Z=z|do(X=1))}{P(Z=z|do(X=1))}$$

  - Si Z es una variable pre-intervención, tenemos $Z_{X=1} = Z$ y se cumple:

$$\mathbb{E}[Y \mid do(X=1), Z=1] = \mathbb{E}[Y_{X=1} \mid Z=1]$$

- Usando la tabla podemos construir:

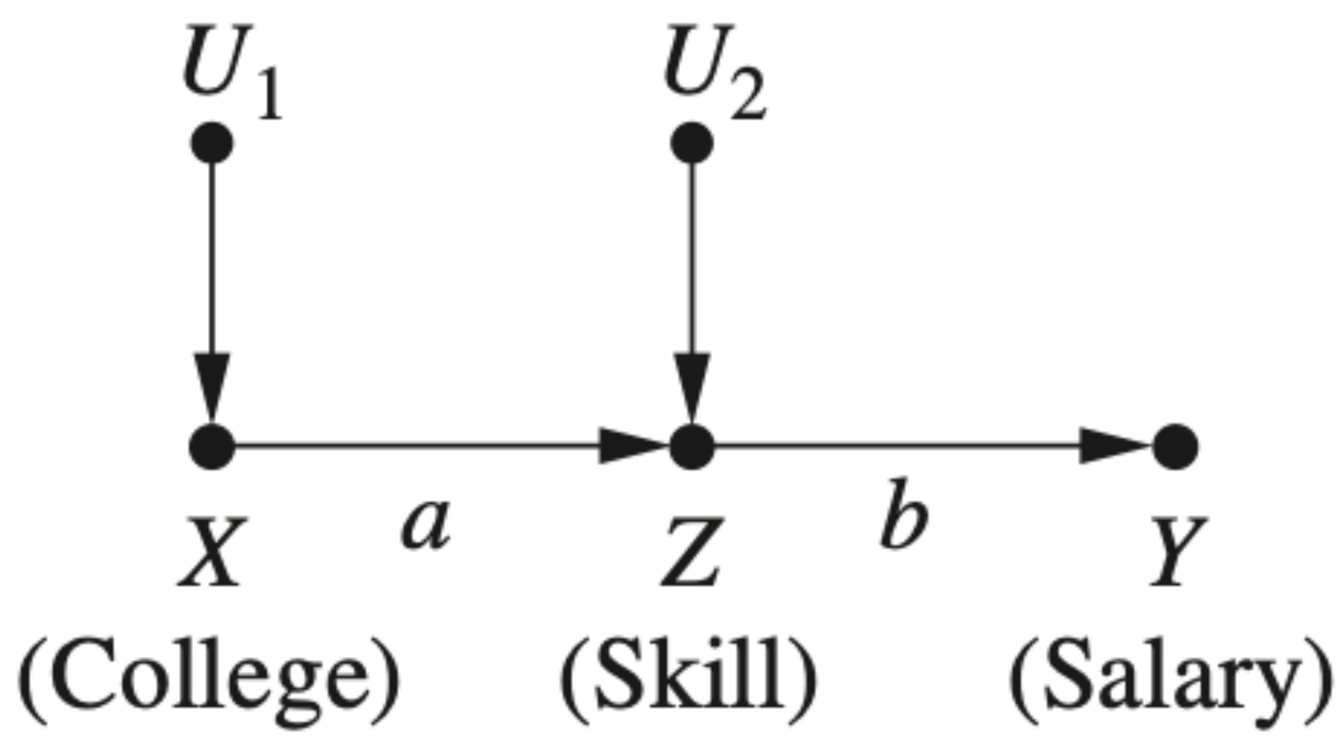
$$X = U_1 \quad Z = aX + U_2, Y = bZ$$

$$E[Y_1|Z=1] = (a+1)b$$

$$E[Y_0|Z=1] = b$$

$$E[Y|do(X=1), Z=1] = b$$

$$E[Y|do(X=0), Z=1] = b$$




- Aunque Z separe a X de Y en el grafo, encontramos un efecto de X en Y cuando Z está por debajo de Z =1:

$$E[Y_1 - Y_0|Z=1] = ab \neq 0$$

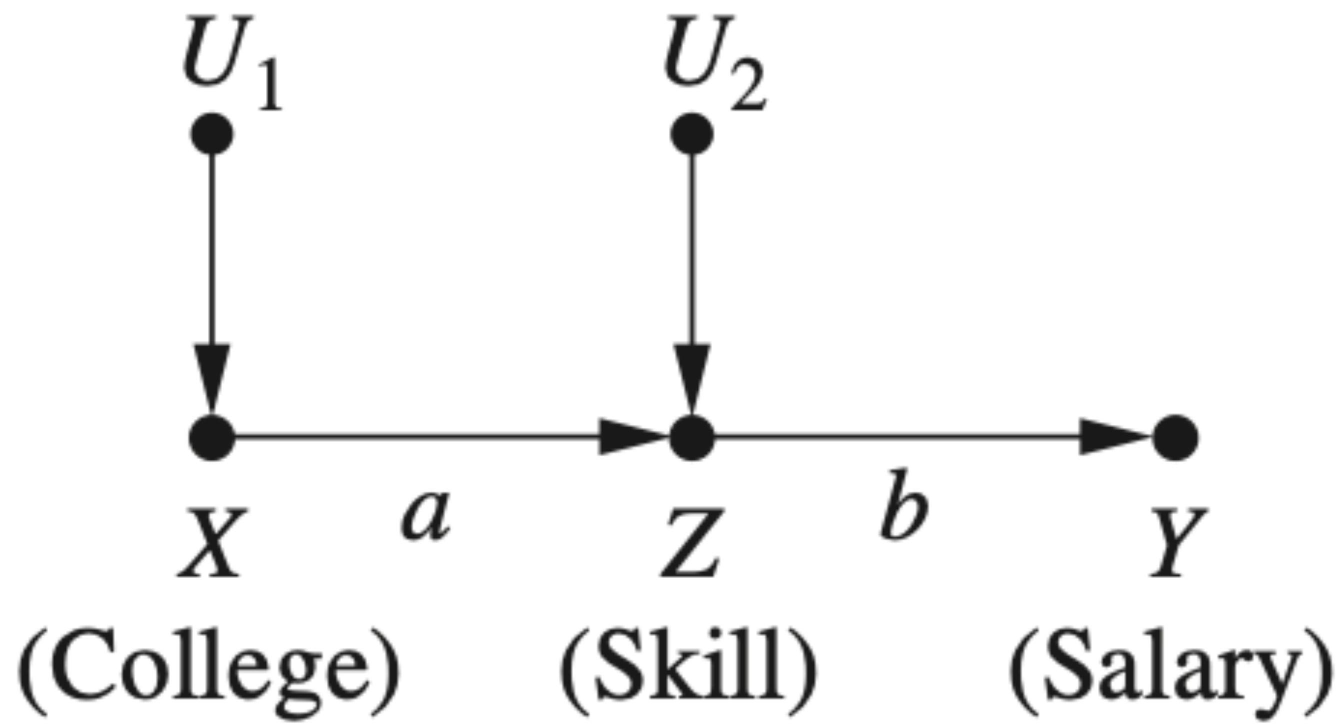
**Table 4.2** The values attained by $X(u), Y(u), Z(u), Y_0(u), Y_1(u), Z_0(u)$, and $Z_1(u)$ in the model of Eq. (4.7)

| $X = u_1 \quad Z = aX + u_2 \quad Y = bZ$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $u_1$ | $u_2$ | $X(u)$ | $Z(u)$ | $Y(u)$ | $Y_0(u)$ | $Y_1(u)$ | $Z_0(u)$ | $Z_1(u)$ |
| 0 | 0 | 0 | 0 | 0 | 0 | $ab$ | 0 | $a$ |
| 0 | 1 | 0 | 1 | $b$ | $b$ | $(a+1)b$ | 1 | $a+1$ |
| 1 | 0 | 1 | $a$ | $ab$ | 0 | $ab$ | 0 | $a$ |
| 1 | 1 | 1 | $a+1$ | $(a+1)b$ | $b$ | $(a+1)b$ | 1 | $a+1$ |

- Aunque Z separe a X de Y en el grafo, encontramos un efecto de X en Y cuando Z está por debajo de Z =1:

$$E[Y_1 - Y_0|Z = 1] = ab \neq 0$$

- El salario de las personas que adquirieron habilidad Z = 1, sólo depende, según el grafo, de Z y no de X. Como puede ser?
- El salario de las personas que están en Z = 1 hubiese sido diferente si hubiesen tenido un pasado diferente (X = 1 o X = 0).
- Estas dependencias del pasado no realizado no aparecen en la gráfica.



- Por la ley fundamental, vimos que $M_x$ (submodelo) es la versión modificada de $M$ con la ecuación de $X$ sustituida por $X = x$.
- La realización de la variable $Y$ en el modelo modificado es el contrafáctico: $Y_x(u) = Y_{M_x}(u)$ del modelo original.
- La modificación consiste en remover todos los vértices incidentes a X

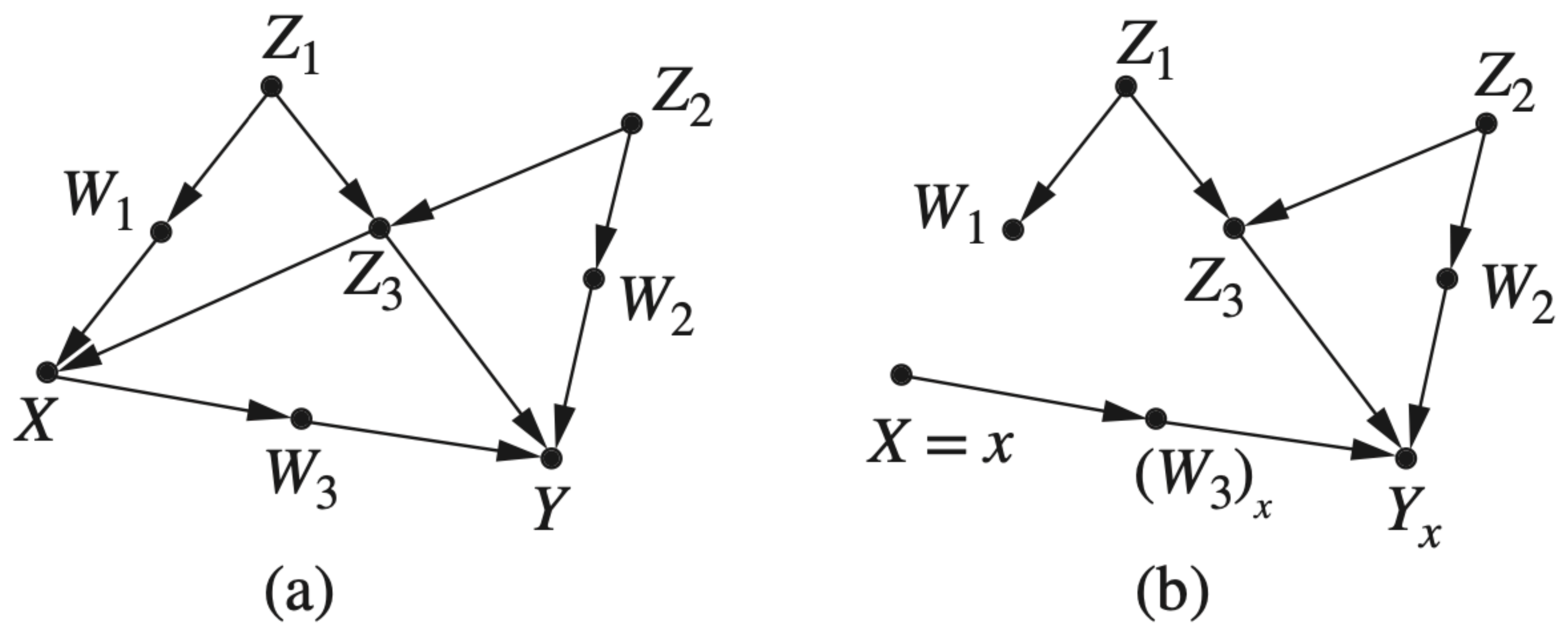



**Figure 4.4** Illustrating the graphical reading of counterfactuals. (a) The original model. (b) The modified model $M_x$ in which the node labeled $Y_x$ represents the potential outcome $Y$ predicated on $X = x$

- Podemos usar la representación gráfica para responder preguntas de $Y_x$ y su dependencia de las otras variables del modelo.

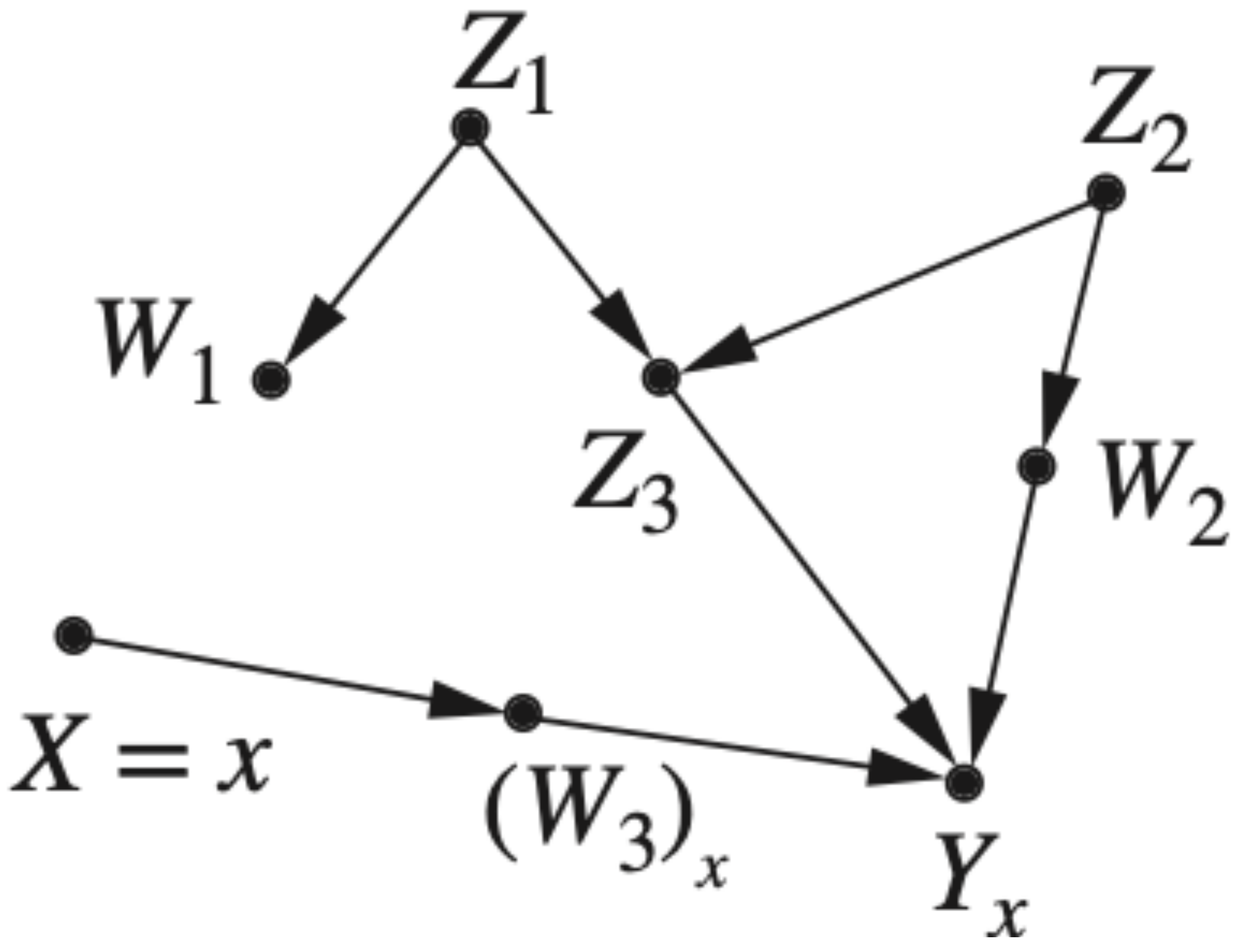
- Que causaria variaciones en $Y_x$ ?

  - Según su definición SCM, $Y_x$ representa los valores de Y donde X se mantiene constante en X = x.

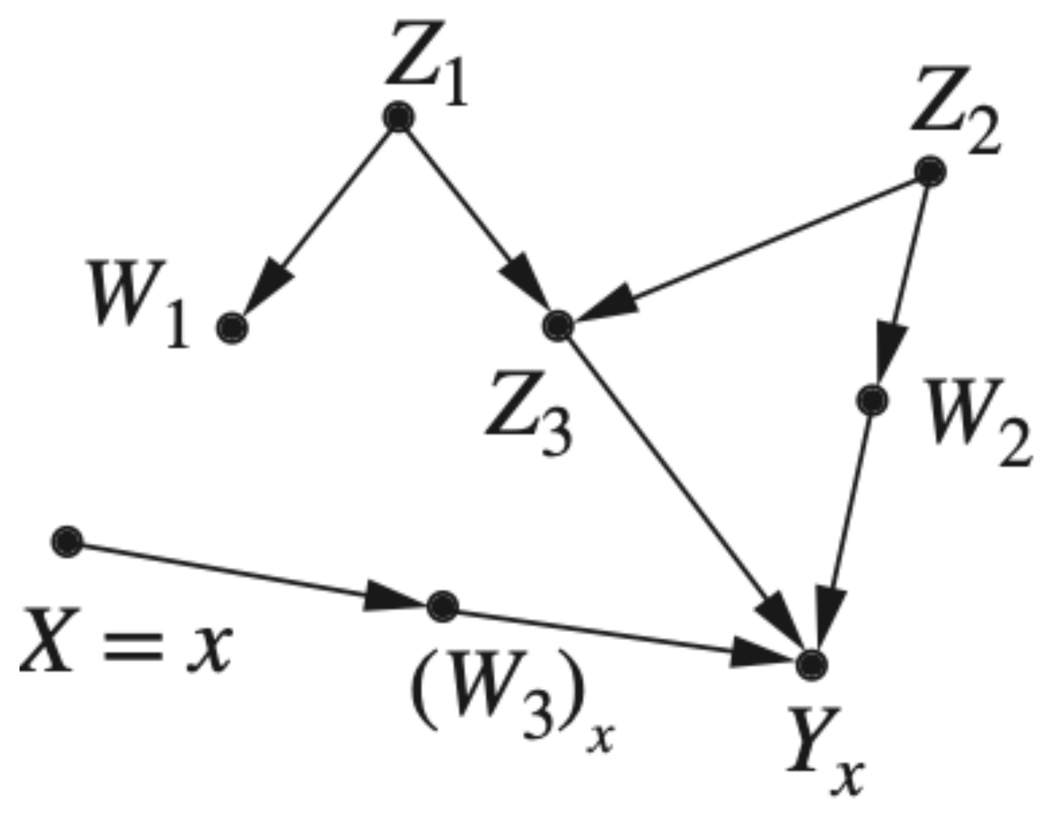
  - Las variaciones en $Y_x$ se deben a todas las variables exógenas capaces de influenciar Y cuando X se mantiene constante.

  - O sea los padres de Y, y los padres de los nodos en el camino entre X e Y.

  - En el caso de la derecha serian: $\{Z_3, W_2, U_3, U_Y\}$ donde $U_3, U_Y$ son términos de error de Y y $W_3$.



- En el caso de la derecha serian: $\{Z_3, W_2, U_3, U_Y\}$ donde $U_3, U_Y$ son términos de error de $Y_x$ y $W_3$.
- Cualquier conjunto de variables que bloquen el camino a estos padres también bloquean el camino a $Y_x$ y resultan en independencia condicional de $Y_x$




**Theorem 4.3.1 (Counterfactual Interpretation of Backdoor)** *If a set Z of variables satisfies the backdoor condition relative to (X, Y), then, for all x, the counterfactual $Y_x$ is conditionally independent of X given Z*

$$P(Y_x|X, Z) = P(Y_x|Z) \tag{4.15}$$

- Podemos explicar porque X tiene un efecto sobre el salario Y cuando condicionamos en Z.

- En el modelo Y depende únicamente de Z.

- Para determinar si el efecto de la educación en el salario ($Y_x$) es independiente, tenemos que ubicar $Y_x$ en el gráfo y ver si está d-separado de X dado Z.

- Podemos ver que $Y_x$ está identificado con $U_2$ el único padre en el camino causal de X a Y.

- Podemos ver que Z es un colíder entre X y $U_2$, al condicionar en Z, creamos una dependencia entre ellos. No están d-separados.

$$E[Y_x|X,Z] \neq E[Y_x|Z]$$

- Aunque:

$$E[Y|X,Z] = E[Y|Z]$$