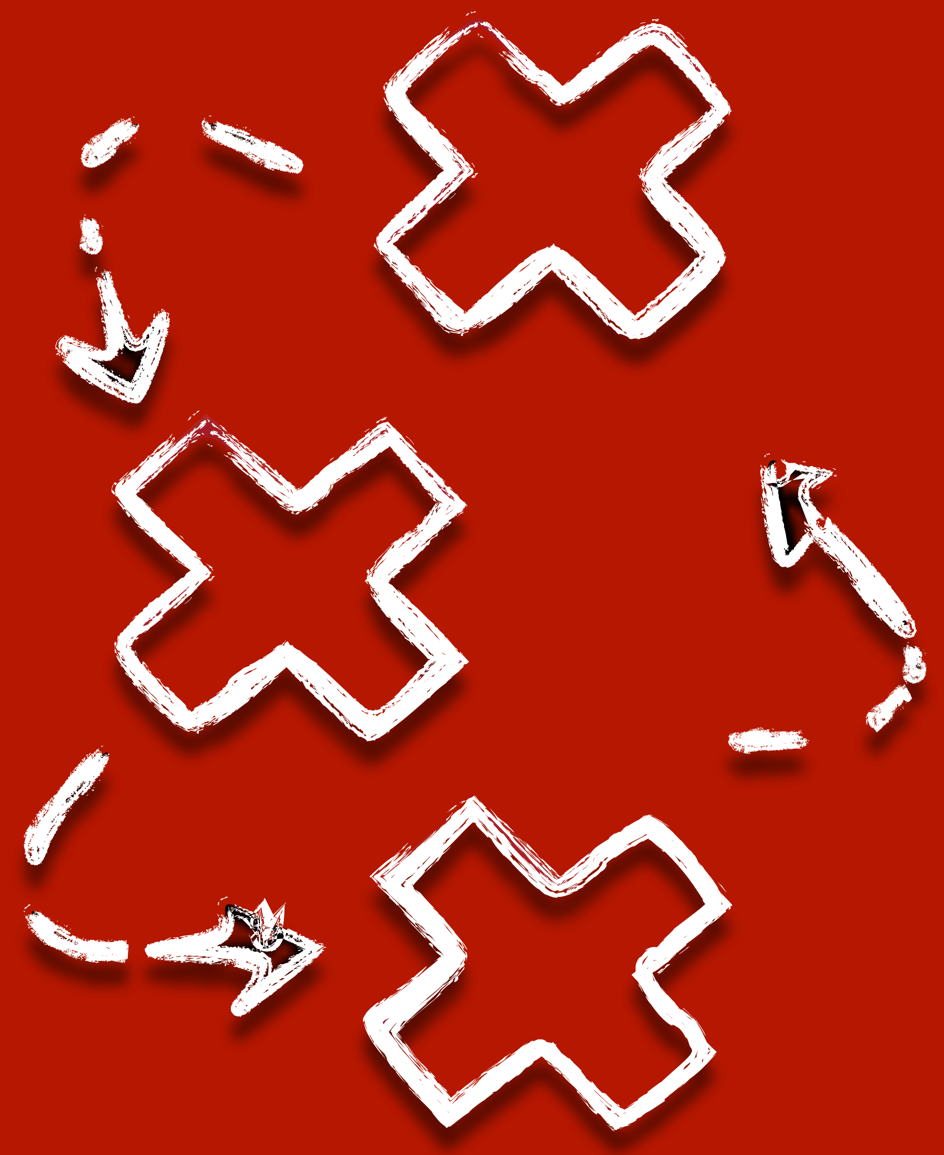


Causal Data Science

Lecture 6:2 Frontdoor criterion, Instrumental variables

Lecturer: Sara Magliacane

UvA - Spring 2024

Adjustment criterion [Shpitser et al, Perkovic et al]

- Find all **valid adjustment sets** for estimating the causal effect of X_i on X_j with adjustment sets $\mathbf{Z} \subseteq \mathbf{V} \backslash \{i, j\}$:

$$p(x_j \,|\, \mathrm{do}(x_i)) = \int_{x_{\mathbf{Z}}} p(x_j \,|\, x_i, x_{\mathbf{Z}}) p(x_{\mathbf{Z}}) dx_{\mathbf{Z}}$$

- There are other **identification strategies,** i.e. ways to estimate the interventional distribution from the observational ones:
  - **Frontdoor criterion**, Instrumental variables
- **None of these are complete (they find all formulas),** do-calculus is

Advanced topics: Do-calculus (last week)

- Given a causal graph G, an **identification strategy** is a formula to estimate an interventional distribution from a combination of observational ones
- **Do-calculus is complete:** if a formula (**an estimand**) exists for estimating an interventional distribution from observational data, it will find it

Advanced topics: Do-calculus (last week)

- Given a causal graph G, an **identification strategy** is a formula to estimate an interventional distribution from a combination of observational ones
- **Do-calculus is complete:** if a formula (**an estimand**) exists for estimating an interventional distribution from observational data, it will find it
- **Calculus**: a set of rules that you can use to rewrite interventional distributions in equivalent expressions using interventional and observational distributions
  - Keep applying rules until there are no interventional distributions left

Advanced topics: Do-calculus (last week)

- Rule 1: insertion/deletion of observations

$$\mathbf{A} \perp_d \mathbf{B} \mid \mathbf{C}, \mathrm{do}(\mathbf{W}) \implies P(X_{\mathbf{A}} \mid X_{\mathbf{B}}, X_{\mathbf{C}}, \mathrm{do}(X_{\mathbf{W}})) = P(X_{\mathbf{A}} \mid X_{\mathbf{C}}, \mathrm{do}(X_{\mathbf{W}}))$$

- Rule 2: action/observation exchange

$$\mathbf{A} \perp_d I_{\mathbf{B}} \mid \mathbf{B}, \mathbf{C}, \mathrm{do}(\mathbf{W}) \implies P(X_{\mathbf{A}} \mid \mathrm{do}(X_{\mathbf{B}}), X_{\mathbf{C}}, \mathrm{do}(X_{\mathbf{W}})) = P(X_{\mathbf{A}} \mid X_{\mathbf{B}}, X_{\mathbf{C}}, \mathrm{do}(X_{\mathbf{W}}))$$

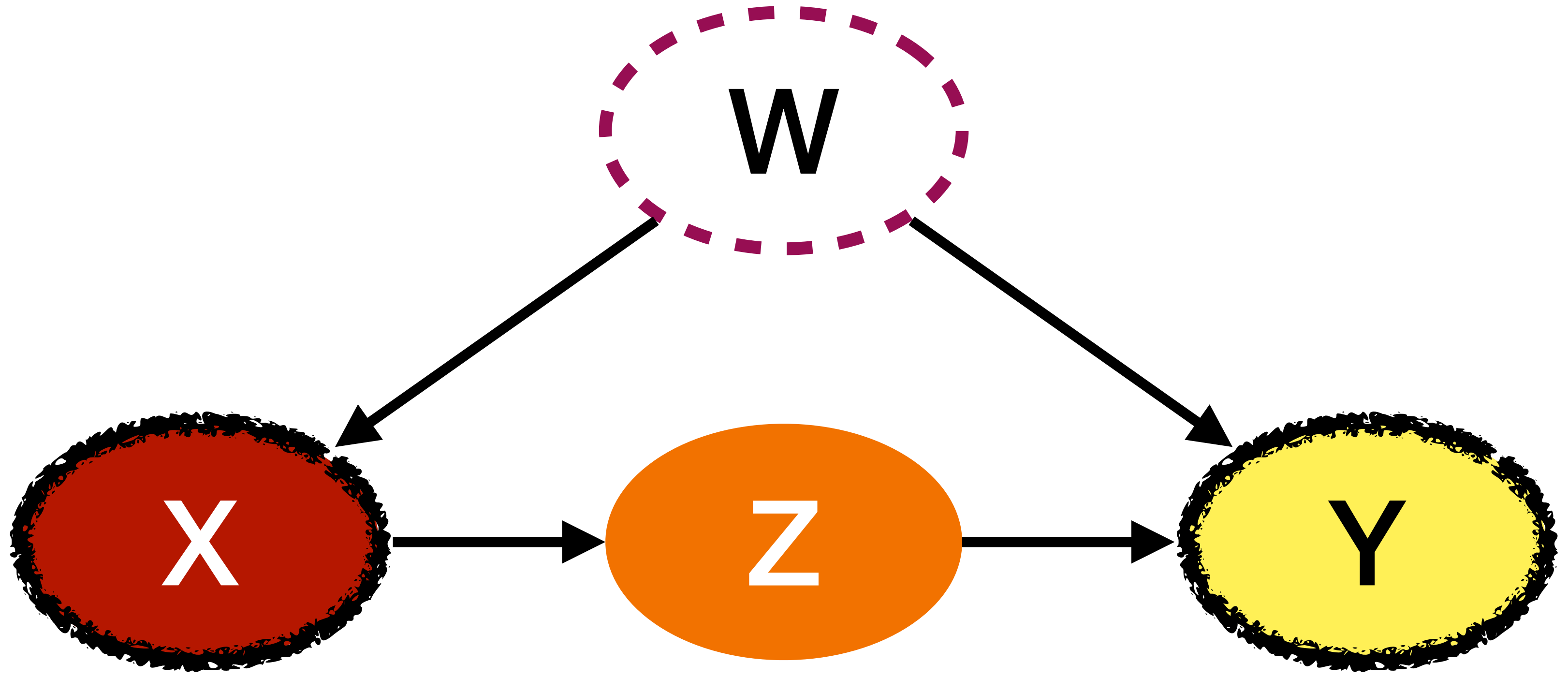
- Rule 3: insertion/deletion of actions

$$\mathbf{A} \perp_d I_{\mathbf{B}} \mid \mathbf{C}, \mathrm{do}(\mathbf{W}) \implies P(X_{\mathbf{A}} \mid \mathrm{do}(X_{\mathbf{B}}), X_{\mathbf{C}}, \mathrm{do}(X_{\mathbf{W}})) = P(X_{\mathbf{A}} \mid X_{\mathbf{C}}, \mathrm{do}(X_{\mathbf{W}}))$$

- One can show that these rules + probability axioms are **complete**
- There is a polytime algorithm for do-calculus (**ID algorithm** [Shpitser and Pearl 2006], https://cran.r-project.org/web/packages/causaleffect/index.html

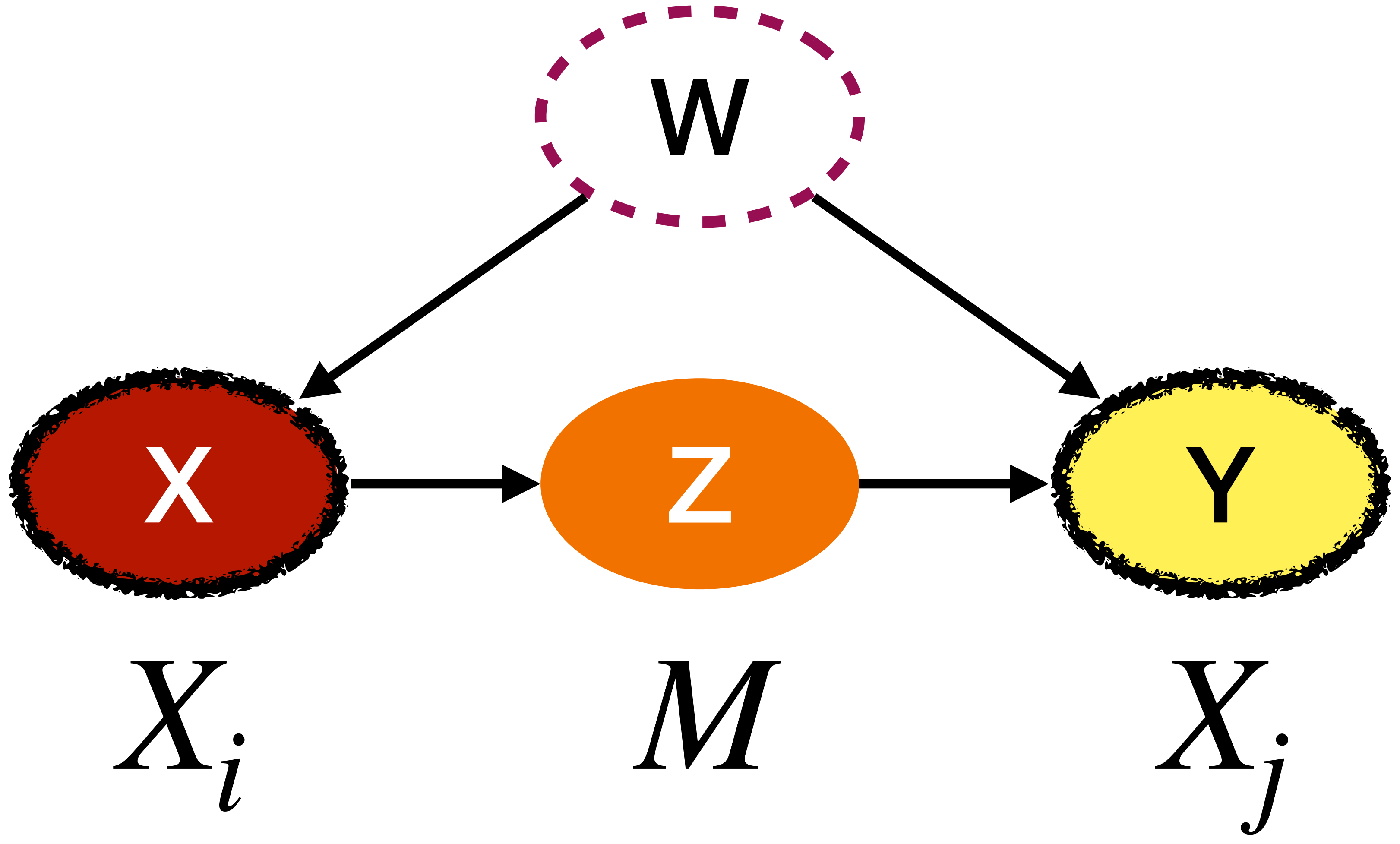
Example - cannot use backdoor/adjustment criteria

- We cannot use the backdoor/adjustment criteria, because W is unobserved



Example - cannot use backdoor/adjustment criteria

- We cannot use the backdoor/adjustment criteria, because W is unobserved
- Frontdoor criterion intuition:
  1. Find all **mediator variables $\mathbf{M}$** on the directed paths between X_i and X_j

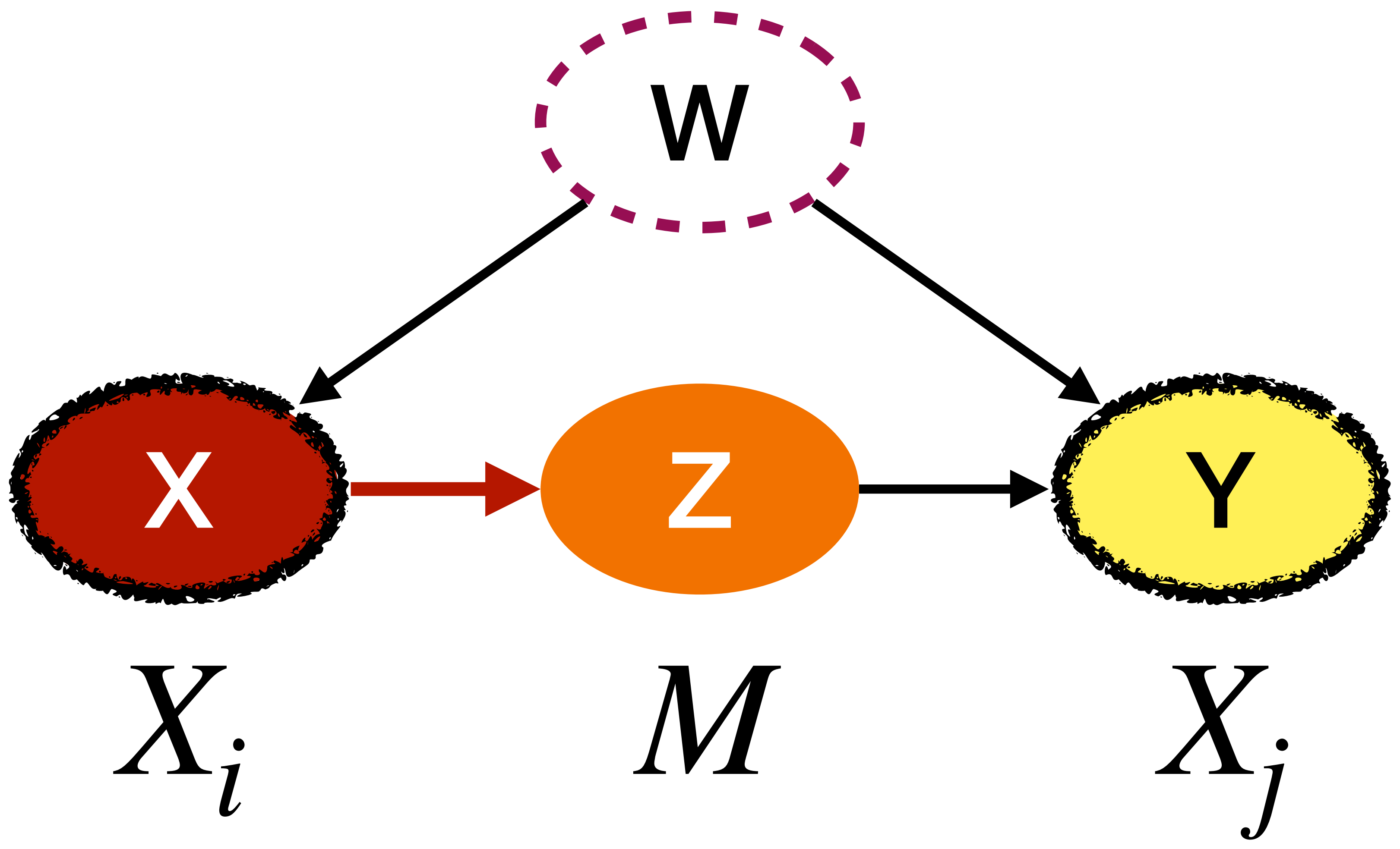


$\mathbf{M}$ blocks all directed paths from X_i to X_j

Example - cannot use backdoor/adjustment criteria

- We cannot use the backdoor/adjustment criteria, because W is unobserved

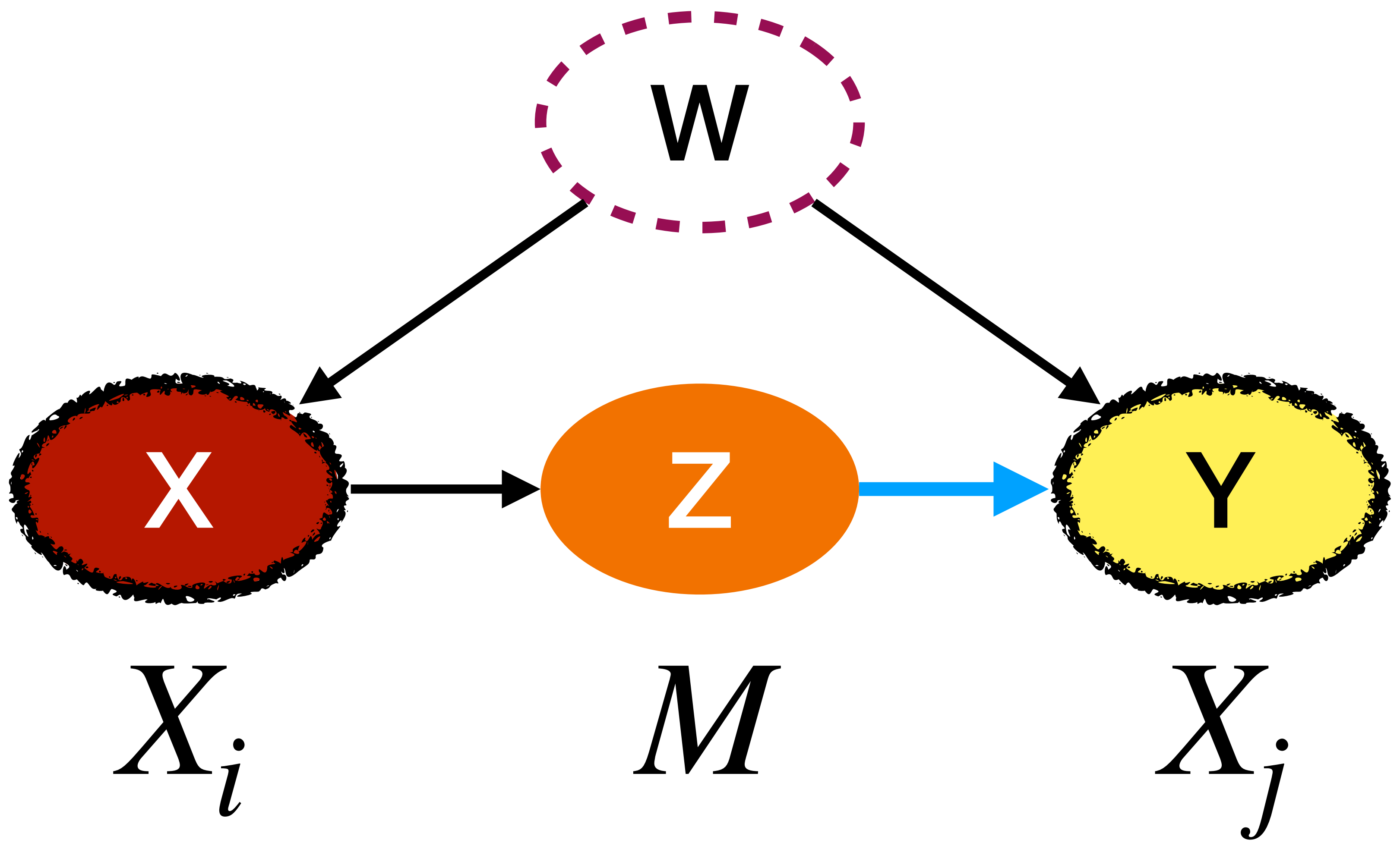
- Frontdoor criterion intuition:
  1. Find all **mediator variables** $\mathbf{M}$ on the directed paths between X_i and X_j
  2. Estimate the effect of X_i on $\mathbf{M}$



Example - cannot use backdoor/adjustment criteria

- We cannot use the backdoor/adjustment criteria, because W is unobserved

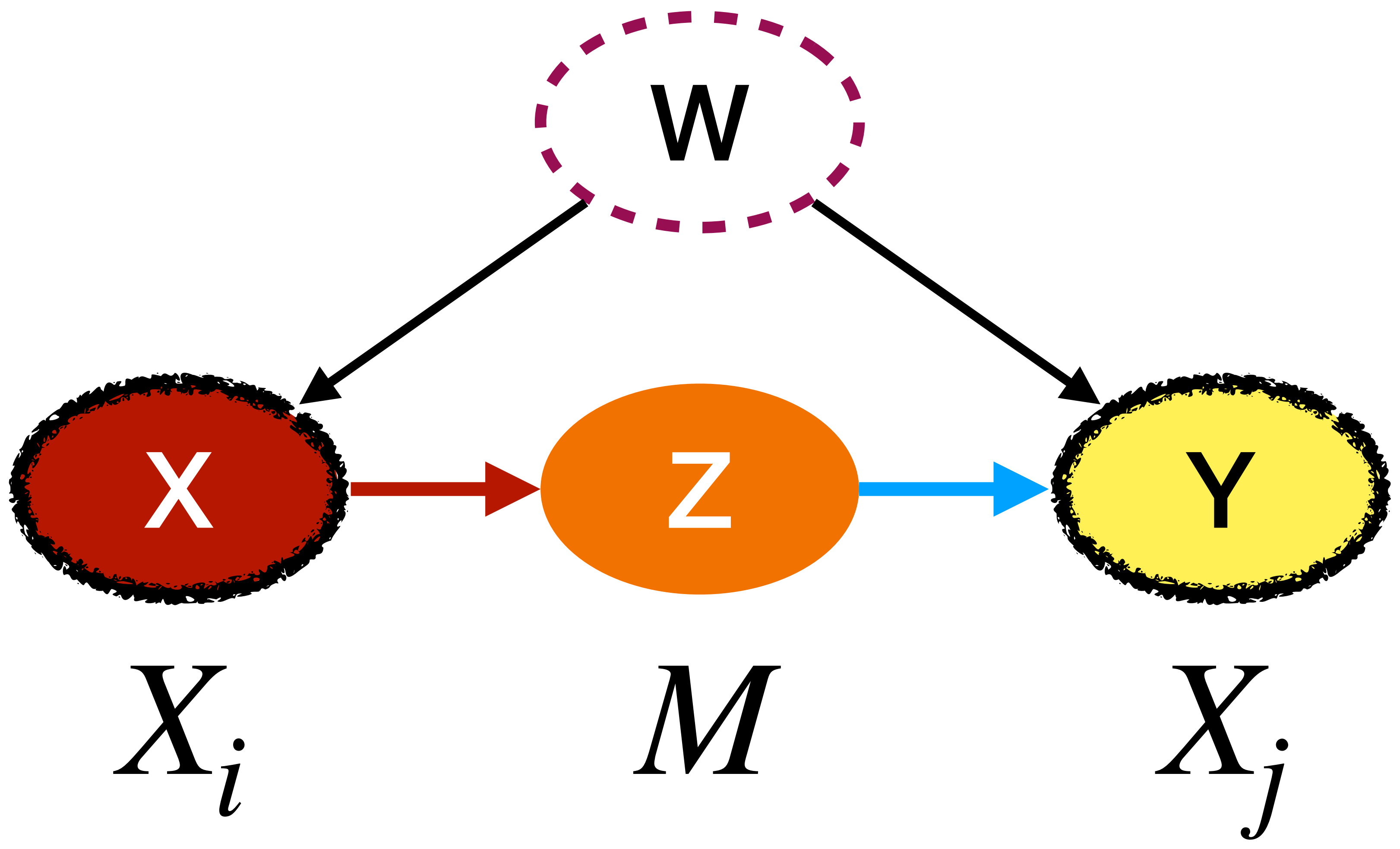
- Frontdoor criterion intuition:
  1. Find all **mediator variables** $\mathbf{M}$ on the directed paths between X_i and X_j
  2. Estimate the effect of X_i on $\mathbf{M}$
  3. Estimate the effect of $\mathbf{M}$ on X_j



Example - cannot use backdoor/adjustment criteria

- We cannot use the backdoor/adjustment criteria, because W is unobserved

- Frontdoor criterion intuition:
  1. Find all **mediator variables $\mathbf{M}$** on the directed paths between X_i and X_j
  2. Estimate the effect of X_i on $\mathbf{M}$
  3. Estimate the effect of $\mathbf{M}$ on X_j
  4. ~~Profit~~ Combine the two effects



Frontdoor criterion

- Given a CBN (G, p) with $G = (\mathbf{V}, \mathbf{E})$, a set $\mathbf{M} \subset \mathbf{V} \backslash \{i, j\}$ satisfies the **front door criterion** for estimating the causal effect of X_i on X_j with $i \neq j$ if:
  1. $\mathbf{M}$ blocks all directed paths from i to j, **and**

Frontdoor criterion

- Given a CBN (G, p) with $G = (\mathbf{V}, \mathbf{E})$, a set $\mathbf{M} \subset \mathbf{V} \backslash \{i, j\}$ satisfies the **front door criterion** for estimating the causal effect of X_i on X_j with $i \neq j$ if:
  1. $\mathbf{M}$ blocks all directed paths from i to j, **and**
  2. No unblocked backdoor paths from $i \leftarrow \dots$ to $\mathbf{M}$ with $Z = \emptyset$, **and**
  - So we can estimate $P(X_{\mathbf{M}} \mid \mathrm{do}(X_i)) = P(X_{\mathbf{M}} \mid X_i)$ without adjustment

Frontdoor criterion

- Given a CBN (G, p) with $G = (\mathbf{V}, \mathbf{E})$, a set $\mathbf{M} \subset \mathbf{V} \backslash \{i, j\}$ satisfies the **front door criterion** for estimating the causal effect of X_i on X_j with $i \neq j$ if:
  1. $\mathbf{M}$ blocks all directed paths from i to j, **and**
  2. No unblocked backdoor paths from $i \leftarrow \dots$ to $\mathbf{M}$ with $Z = \emptyset$, **and**
     - So we can estimate $P(X_{\mathbf{M}} \,|\, \mathrm{do}(X_i)) = P(X_{\mathbf{M}} \,|\, X_i)$ without adjustment
  3. $Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \dots$ to j
     - So we can estimate $P(X_j \,|\, \mathrm{do}(X_{\mathbf{M}}))$ with backdoor adjustment $Z = \{i\}$

Frontdoor criterion

- Given a CBN (G, p) with $G = (\mathbf{V}, \mathbf{E})$, a set $\mathbf{M} \subset \mathbf{V} \backslash \{i, j\}$ satisfies the **front door criterion** for estimating the causal effect of X_i on X_j with $i \neq j$ if:
  1. $\mathbf{M}$ blocks all directed paths from i to j, **and**
  2. No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, **and**
     - So we can estimate $P(X_{\mathbf{M}} \mid \mathrm{do}(X_i)) = P(X_{\mathbf{M}} \mid X_i)$ without adjustment
  3. $Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \ldots$ to j
     - So we can estimate $P(X_j \mid \mathrm{do}(X_{\mathbf{M}}))$ with backdoor adjustment $Z = \{i\}$

Frontdoor criterion formula

- Given a CBN (G, p) with $G = (\mathbf{V}, \mathbf{E})$, and a set $\mathbf{M} \subset \mathbf{V} \setminus \{i, j\}$ that satisfies the **front door criterion,** we can estimate the causal effect of X_i on X_j as:

$$P(X_j = x_j \mid do(X_i = x_i)) = \sum_{x_\mathbf{M}} P(X_\mathbf{M} = x_\mathbf{M} \mid X_i = x_i) \sum_{x_i'} P(X_j = x_j \mid X_\mathbf{M} = x_\mathbf{M}, X_i = x_i') P(X_i = x_i')$$

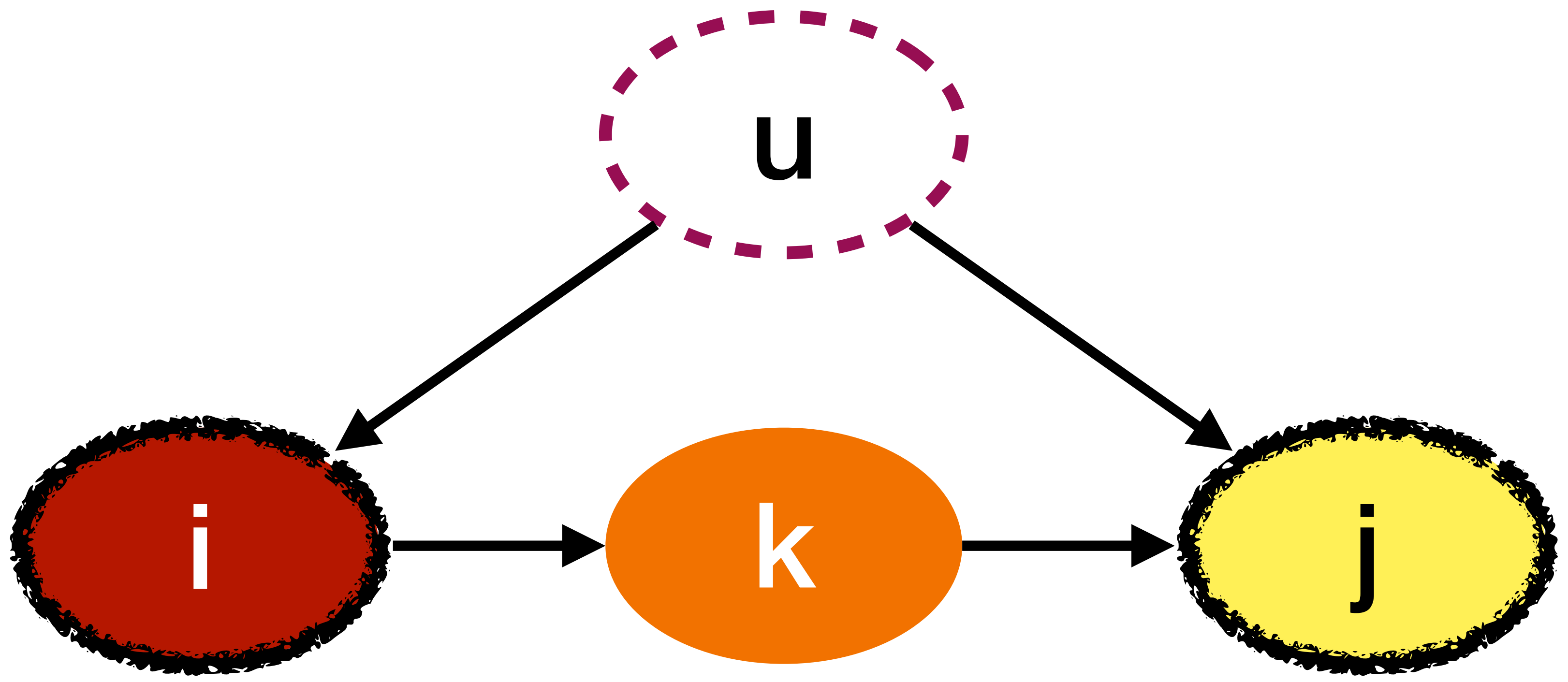
$P(X_\mathbf{M} \mid do(X_i))$ $Z = \emptyset$

$P(X_j \mid do(X_\mathbf{M}))$ $Z = \{i\}$

$$P(X_\mathbf{M} \mid do(X_i)) = P(X_\mathbf{M} \mid X_i)$$

$$P(X_j \mid do(X_\mathbf{M})) = \sum_{x_\mathbf{Z}} P(X_j \mid X_\mathbf{M}, X_\mathbf{Z} = x_\mathbf{Z}) P(X_\mathbf{Z} = x_\mathbf{Z})$$

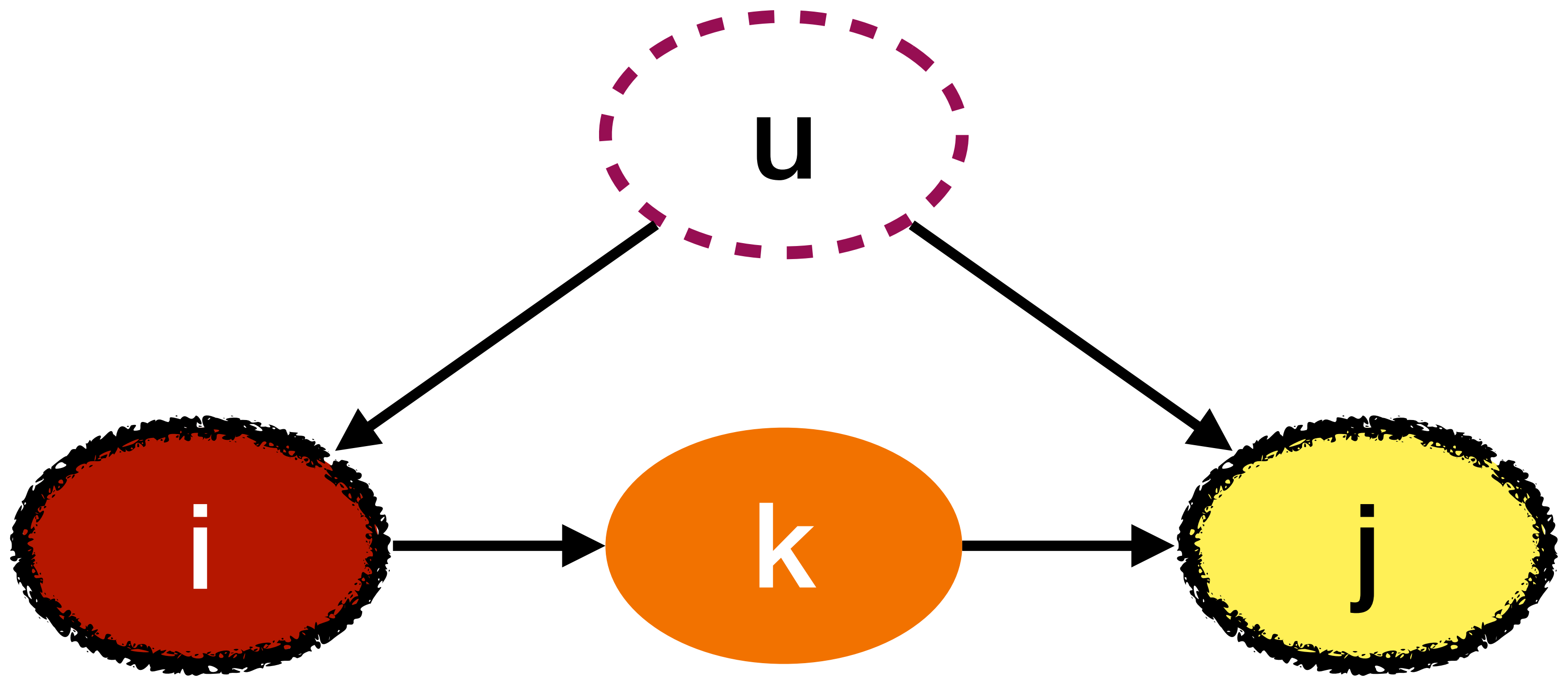
Frontdoor criterion example



$$\mathbf{M} = \{k\}?$$

1. $\mathbf{M}$ blocks all directed paths from i to j, **and**
2. No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, **and**
3. $Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \ldots$ to j

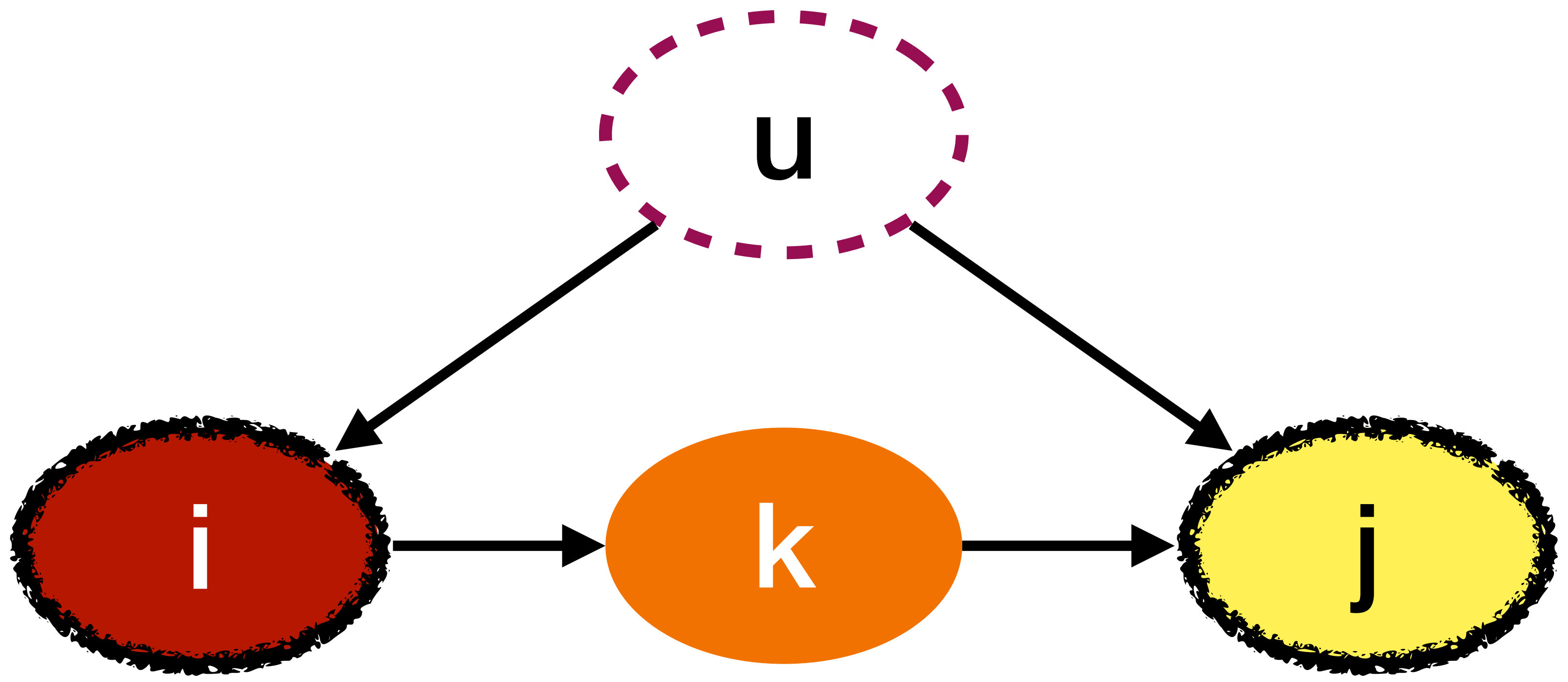
Frontdoor criterion example



$$\mathbf{M} = \{k\}?$$

1. **$\mathbf{M}$ blocks all directed paths from i to j, and**
2. No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, **and**
3. $Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \ldots$ to j

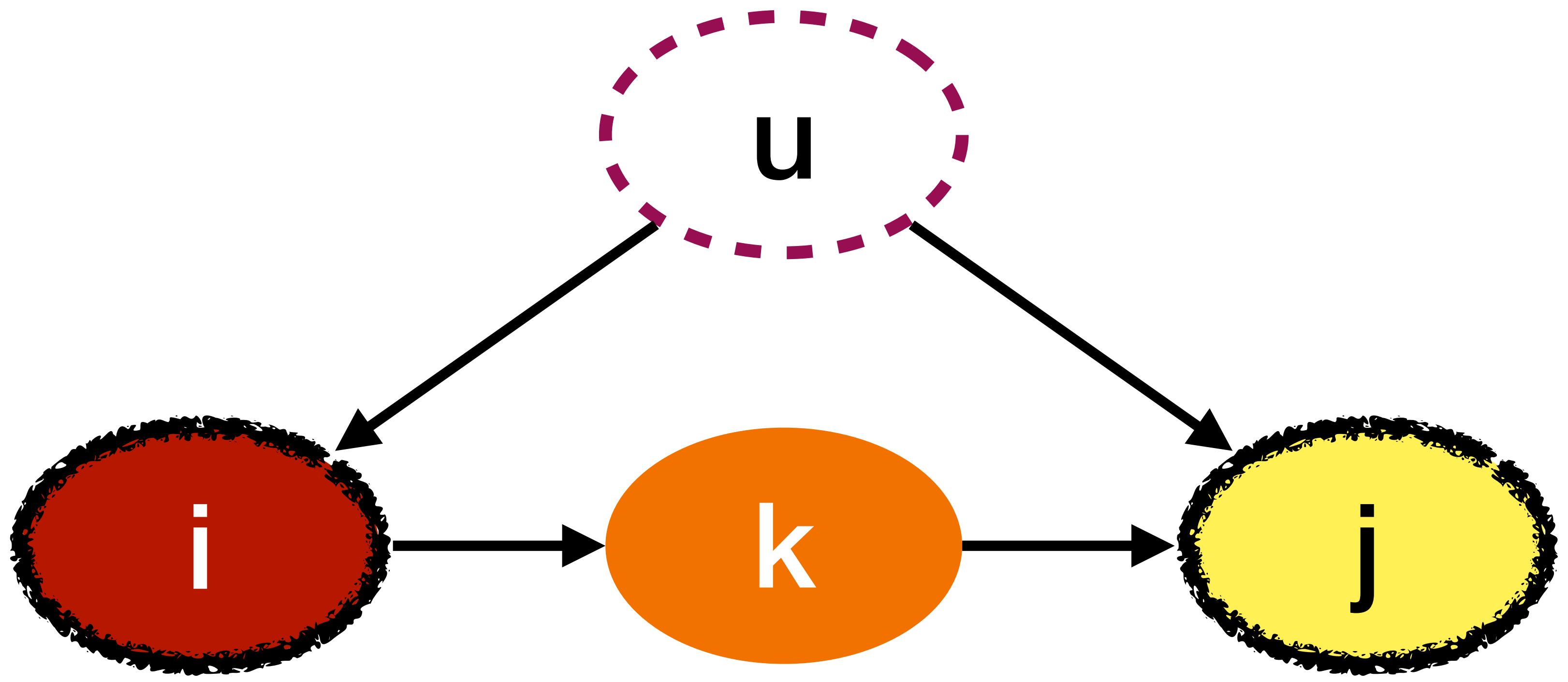
Frontdoor criterion example



$$\mathbf{M} = \{k\}?$$

1. $\mathbf{M}$ blocks all directed paths from i to j, **and**
2. **No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, and**
3. $Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \ldots$ to j

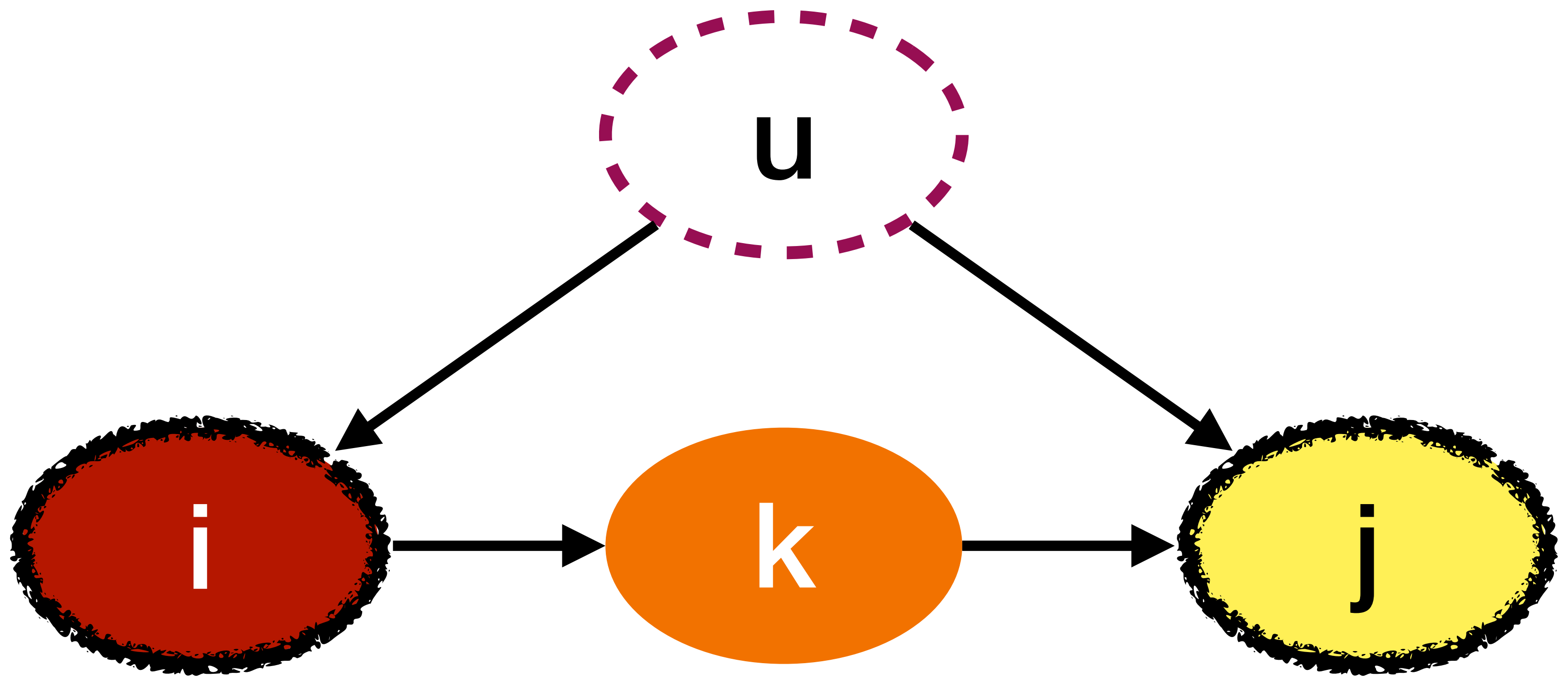
Frontdoor criterion example



$$\mathbf{M} = \{k\}?$$

1. $\mathbf{M}$ blocks all directed paths from i to j, **and**
2. No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, **and**
3. **$Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \ldots$ to j**

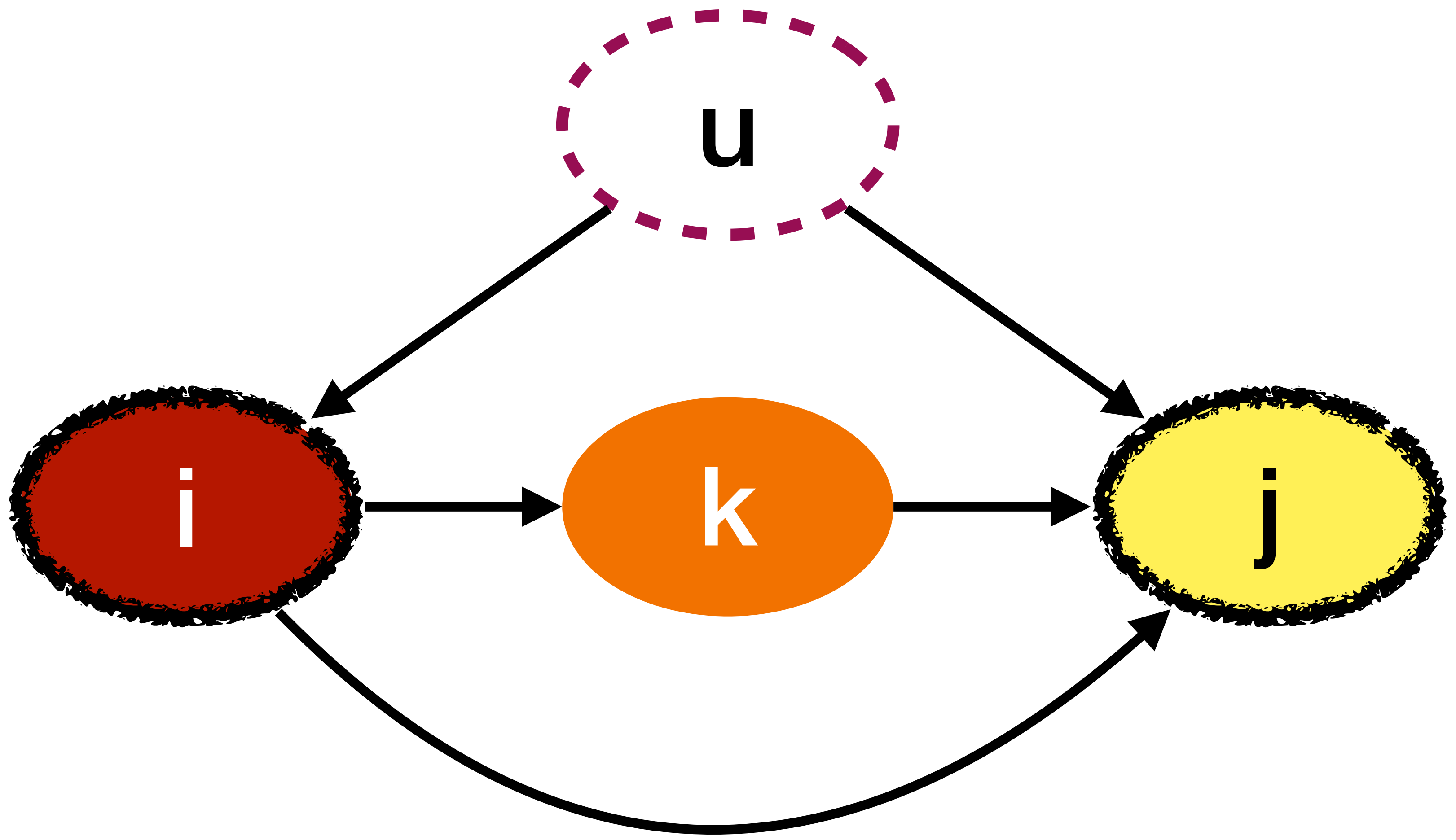
Frontdoor criterion example



$$\mathbf{M} = \{k\}$$

1. $\mathbf{M}$ blocks all directed paths from i to j, and
2. No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, and
3. $Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \ldots$ to j

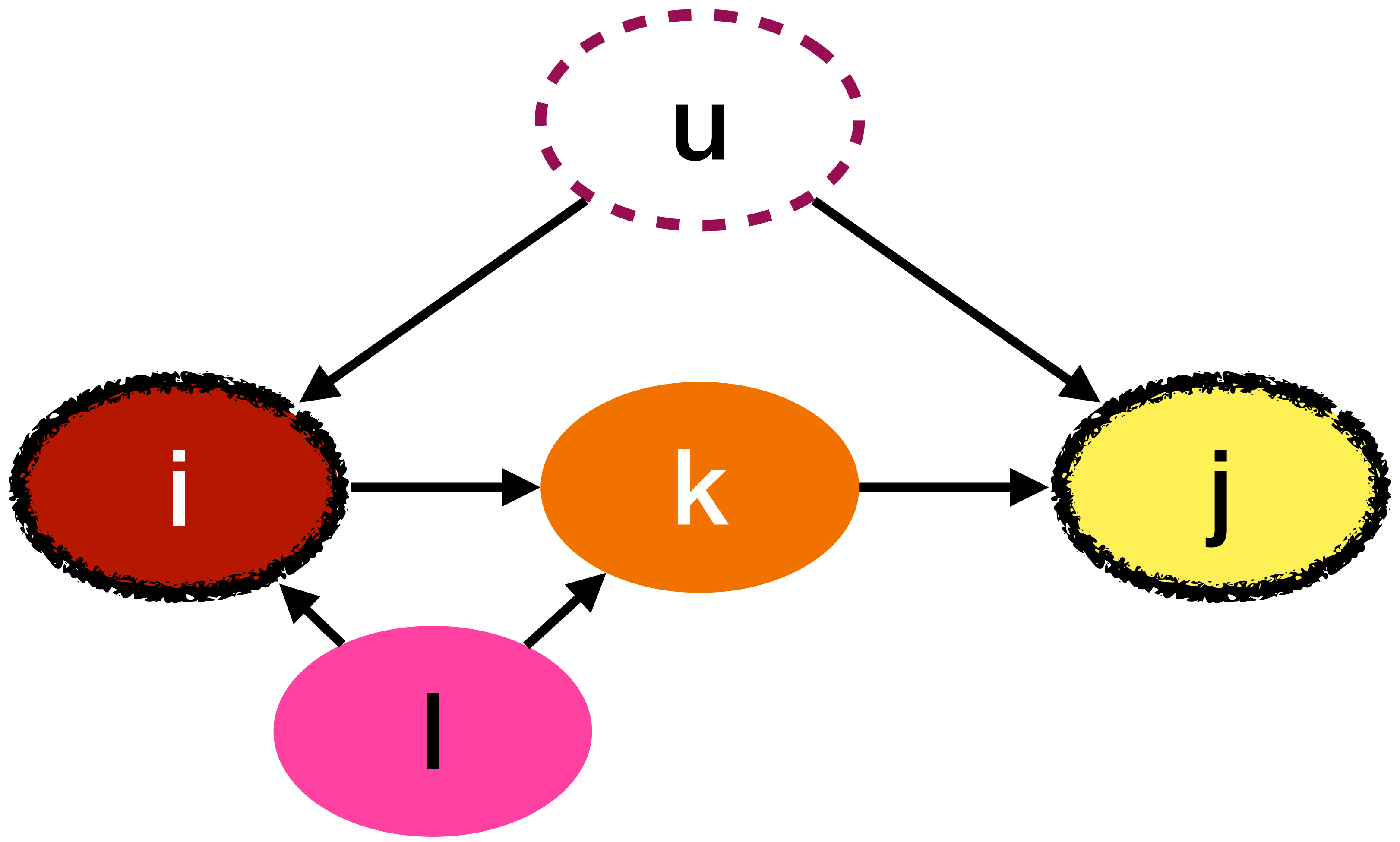
Frontdoor criterion example 2



$$\mathbf{M} = \{k\}?$$

1. ~~$\mathbf{M}$ blocks all directed paths from i to j, and~~
2. No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, **and**
3. $Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \ldots$ to j

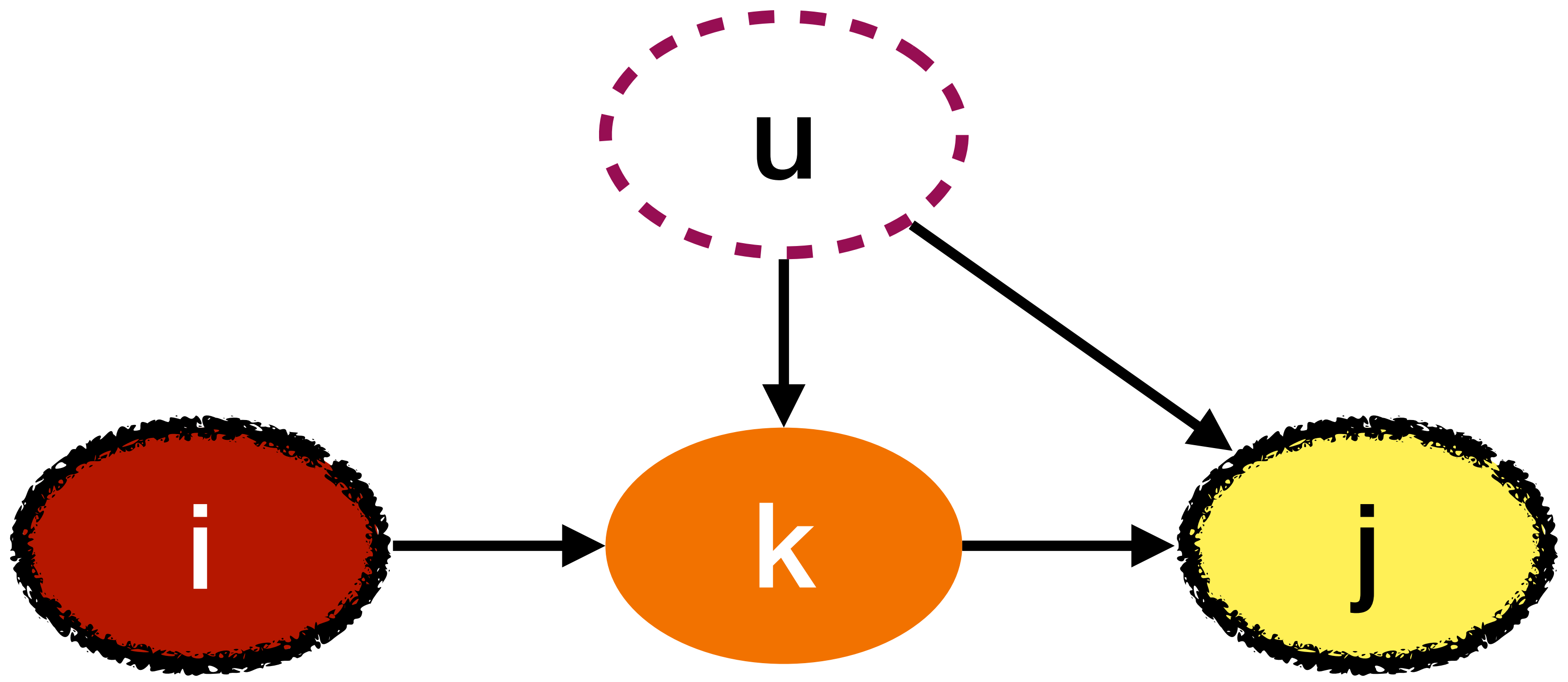
Frontdoor criterion example 2



$$\mathbf{M} = \{k\}?$$

1. $\mathbf{M}$ blocks all directed paths from i to j, **and**
2. ~~**No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, and**~~
3. $Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \ldots$ to j

Frontdoor criterion example 2

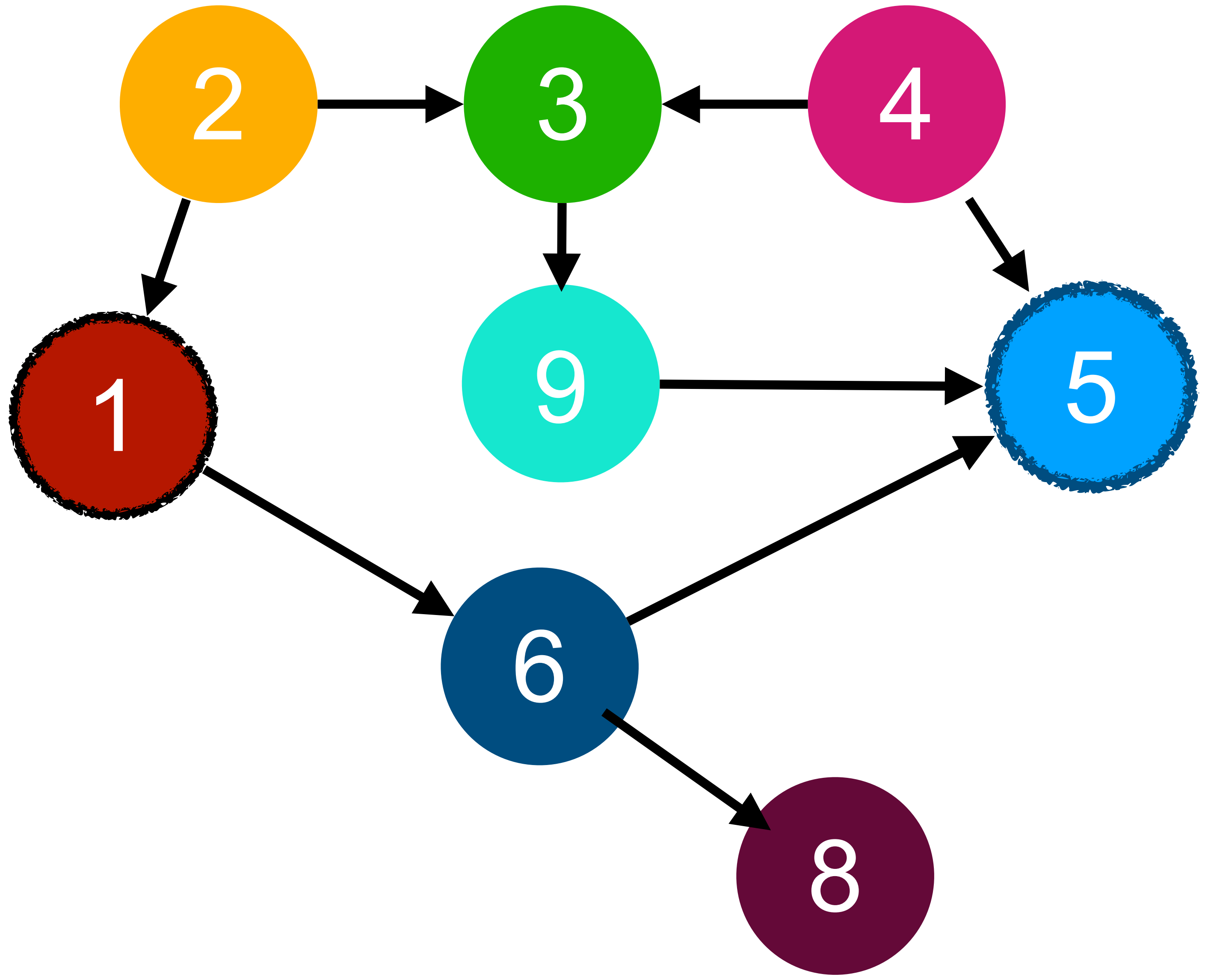


$$\mathbf{M} = \{k\}?$$

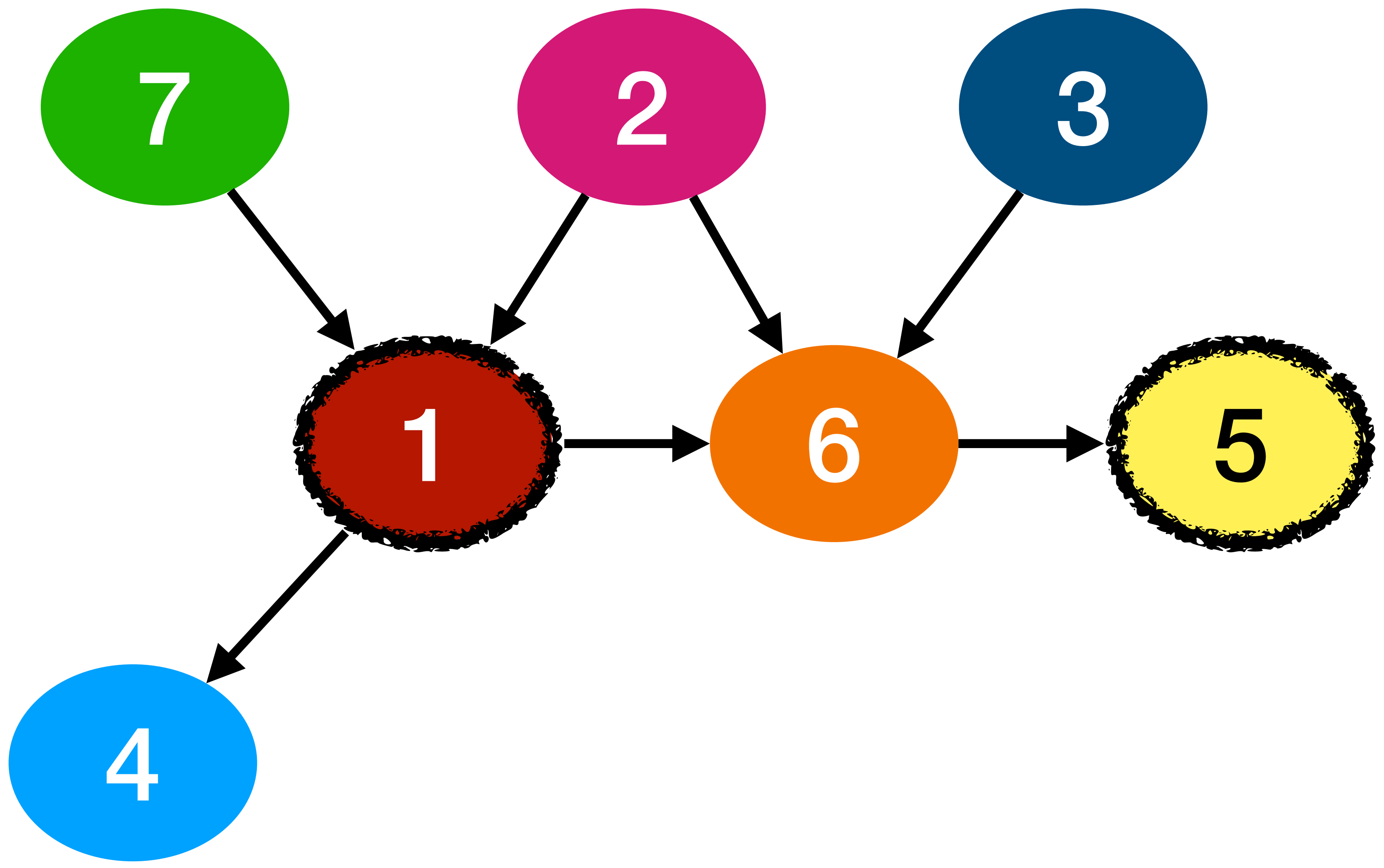
1. $\mathbf{M}$ blocks all directed paths from i to j, **and**
2. No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, **and**
3. ~~$Z = \{i\}$ **blocks all backdoor paths from** $\mathbf{M} \leftarrow \ldots$ **to** j~~

Frontdoor exercise in Canvas for the effect of X_1 on X_5

1. $\mathbf{M}$ blocks all directed paths from i to j, and
2. No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, and
3. $Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \ldots$ to j



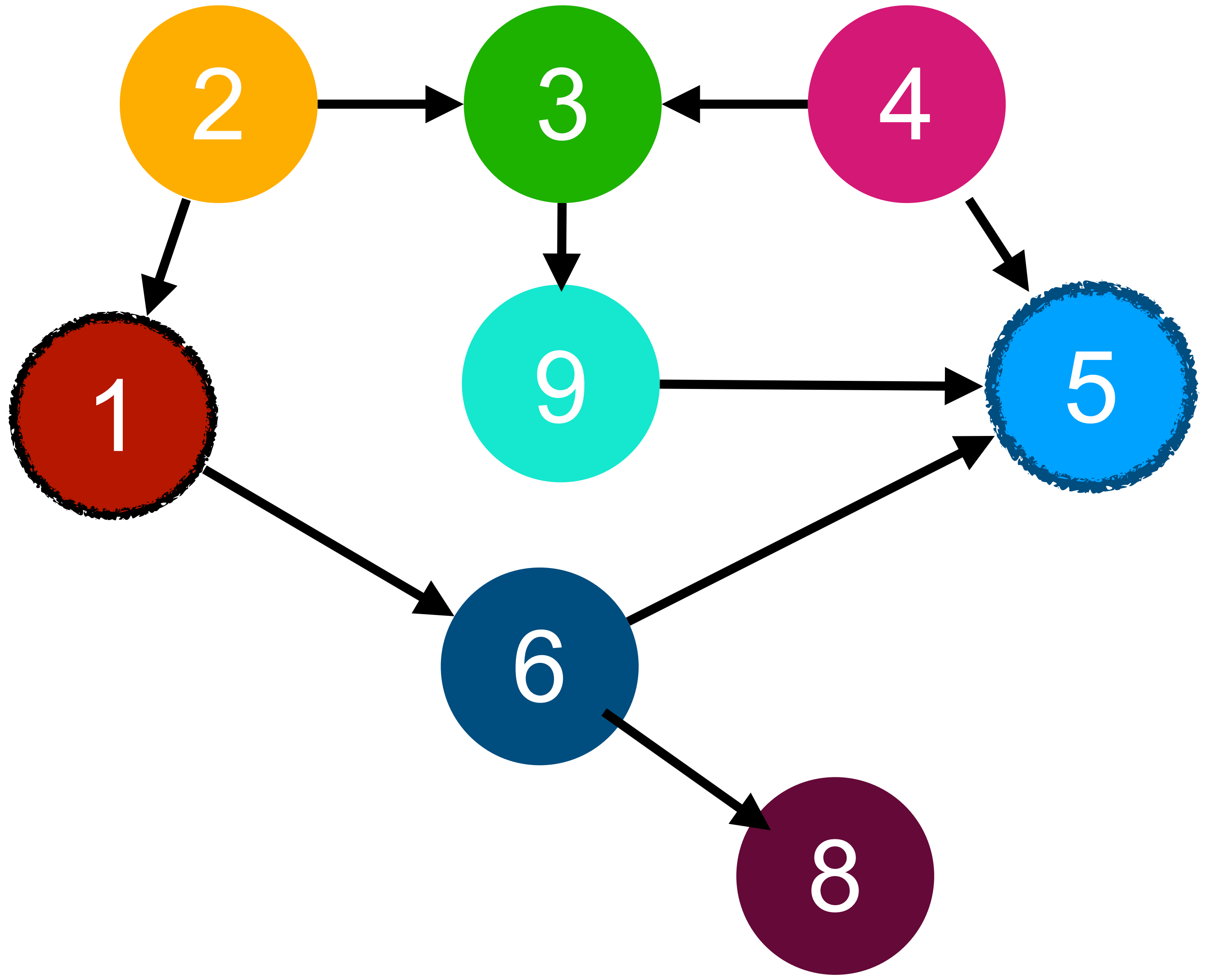
Graph 1



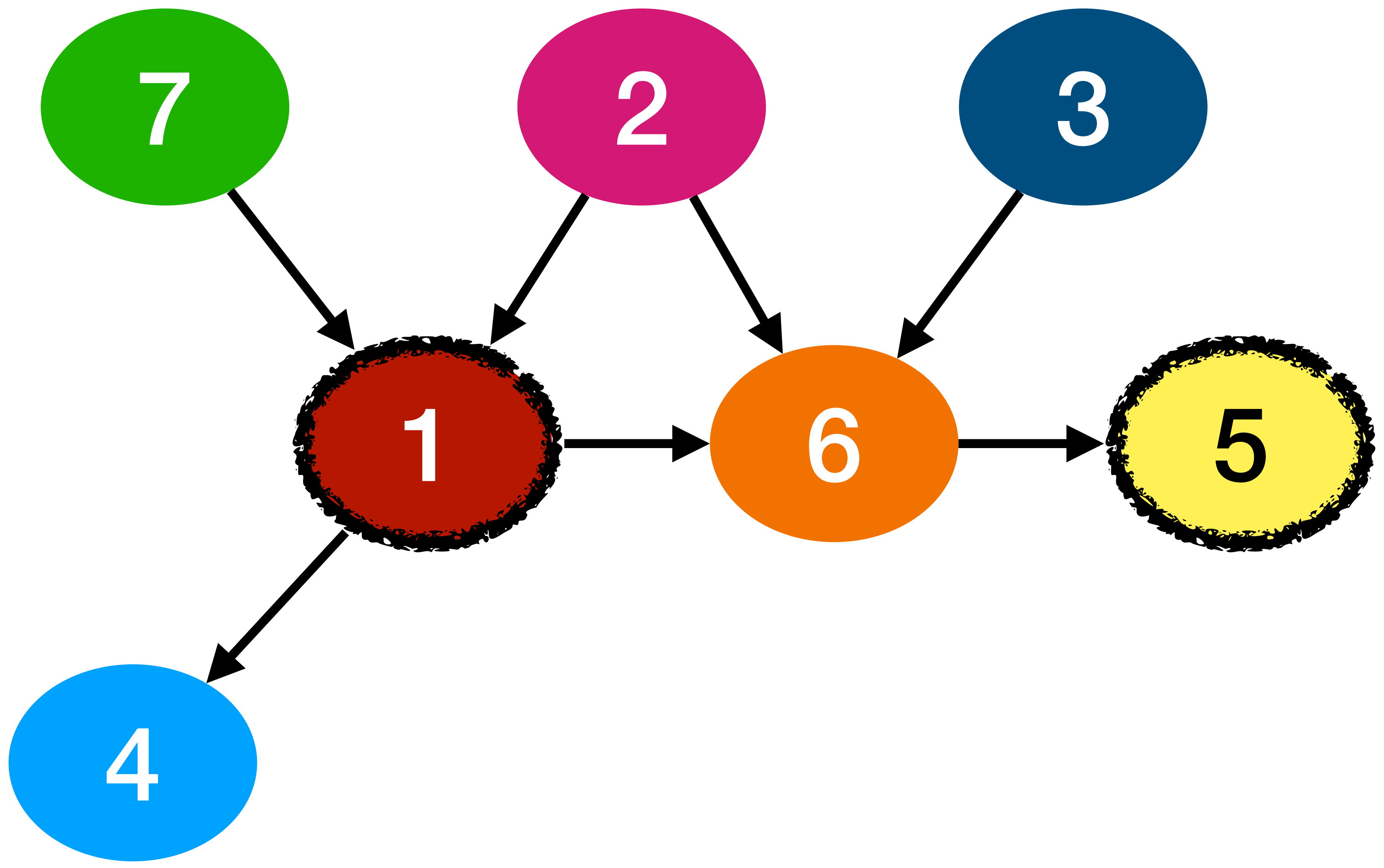
Graph 2

Frontdoor exercise in Canvas for the effect of X_1 on X_5

1. $\mathbf{M}$ blocks all directed paths from i to j, and
2. No unblocked backdoor paths from $i \leftarrow \ldots$ to $\mathbf{M}$ with $Z = \emptyset$, and
3. $Z = \{i\}$ blocks all backdoor paths from $\mathbf{M} \leftarrow \ldots$ to j



Graph 1



Graph 2

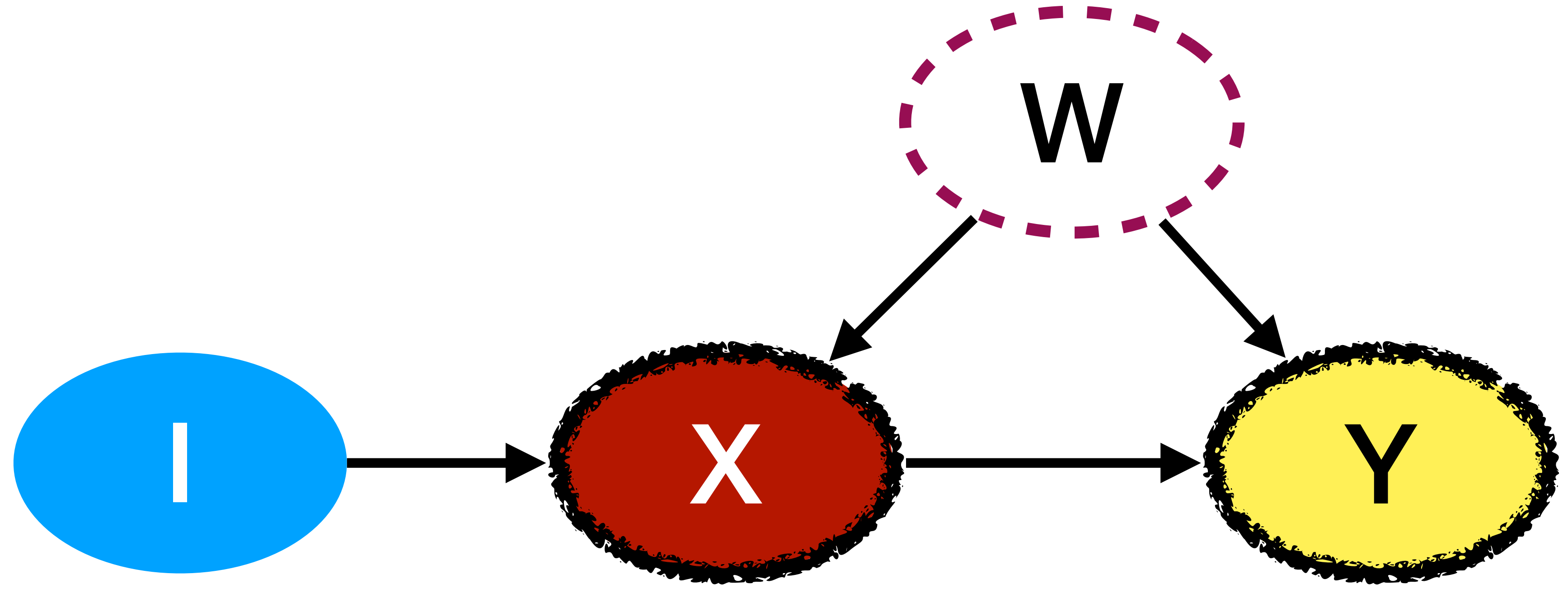
Adjustment criterion [Shpitser et al, Perkovic et al]

- Find all **valid adjustment sets** for estimating the causal effect of X_i on X_j with adjustment sets $\mathbf{Z} \subseteq \mathbf{V} \backslash \{i, j\}$:

$$p(x_j \mid \mathrm{do}(x_i)) = \int_{x_\mathbf{Z}} p(x_j \mid x_i, x_\mathbf{Z}) p(x_\mathbf{Z}) dx_\mathbf{Z}$$

- There are other **identification strategies,** i.e. ways to estimate the interventional distribution from the observational ones:
  - Frontdoor criterion, **Instrumental variables**

Instrumental variables



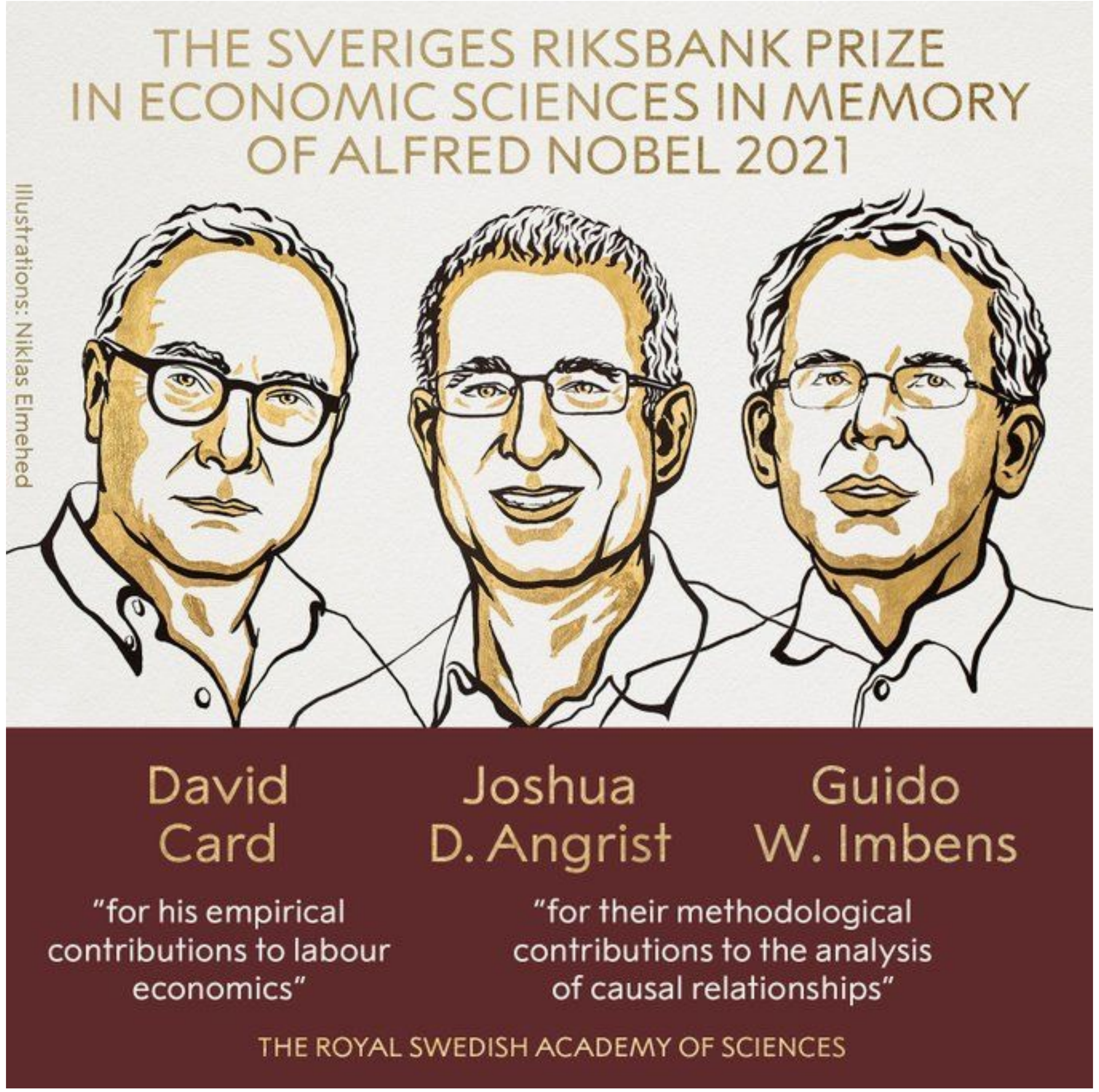
- We want to estimate the effect of X on Y
- We cannot use the backdoor/adjustment criteria, because W is unobserved
- We cannot use frontdoor because there is no mediator
- We can exploit the **instrumental variable (IV)** I iff all these hold:
    - $I \rightarrow X$, $I \perp\!\!\!\perp W$, $I \not\rightarrow Y$ directly

Substantial first stage **Independence Assumption** **Exclusion Restriction**

(there exist also conditional IVs...)

Instrumental variables

- Instrumental variables are a big topic in econometrics
  - Natural experiments
  - Economics Nobel prize 2021
- Similarly to other topics, in this course we will cover a broad range of topics, so we will only see a simple example
- https://www.mostlyharmlesseconometrics.com/



Causal inference example: instrumental variables

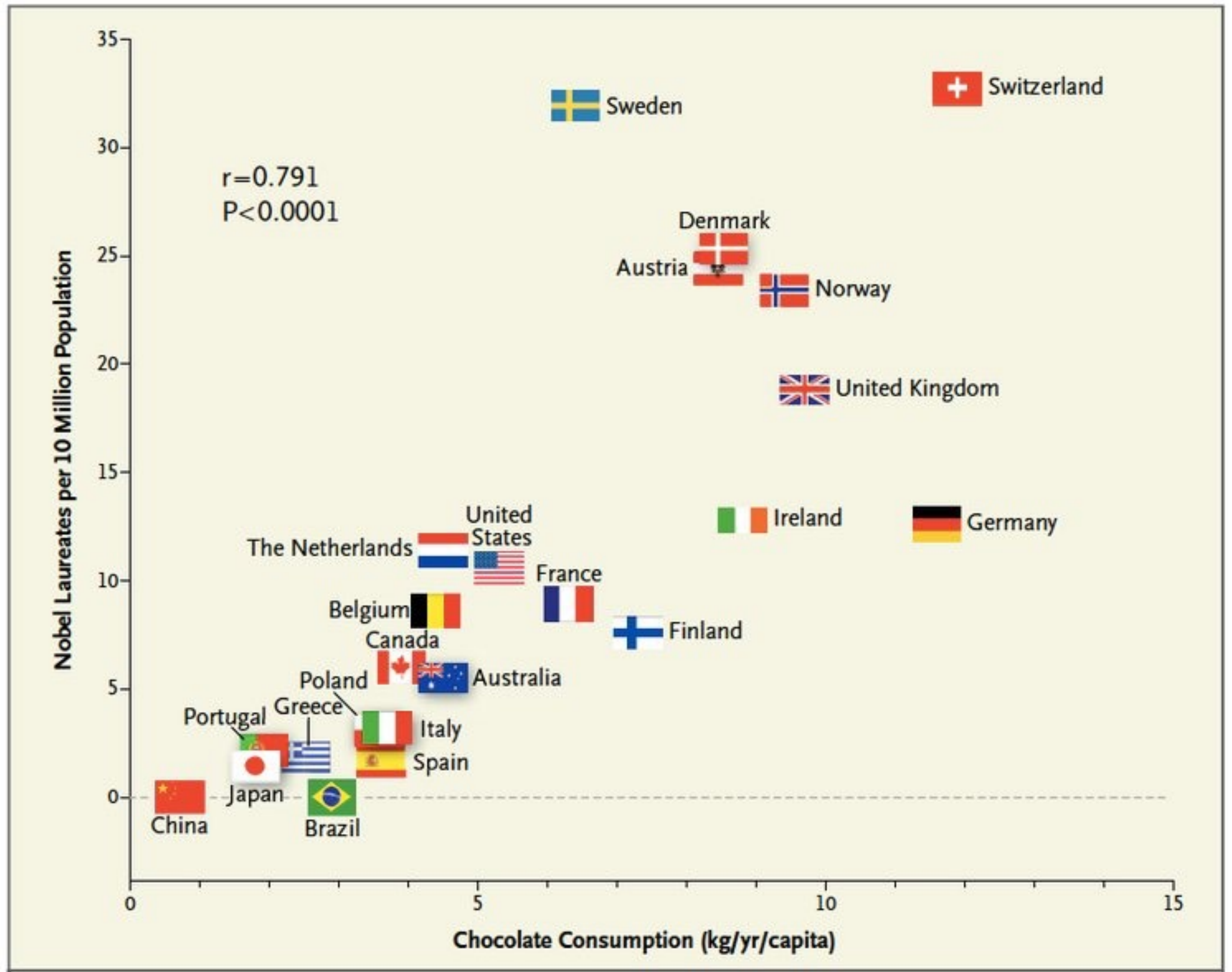
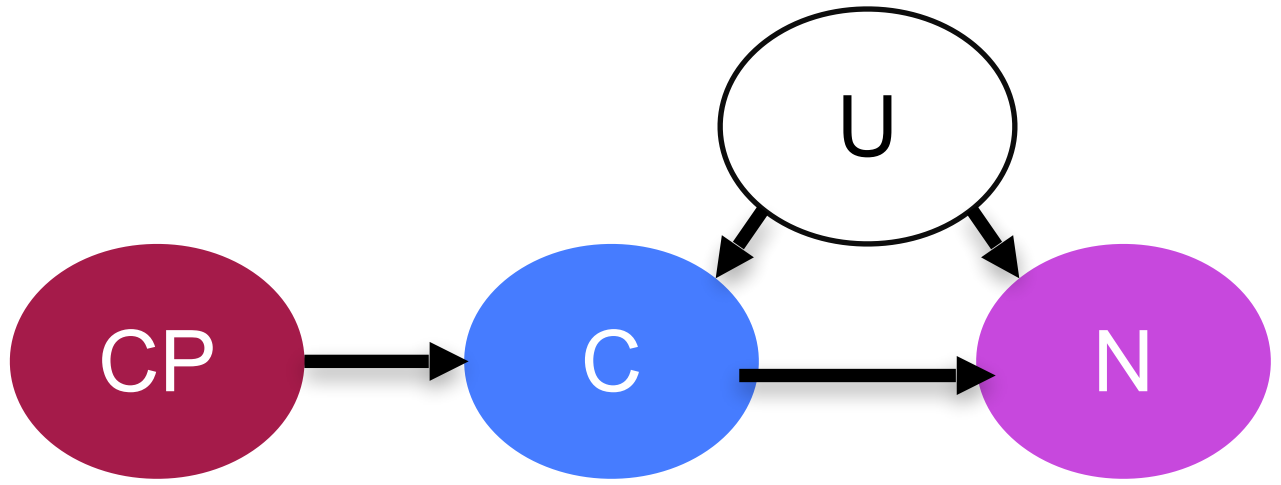


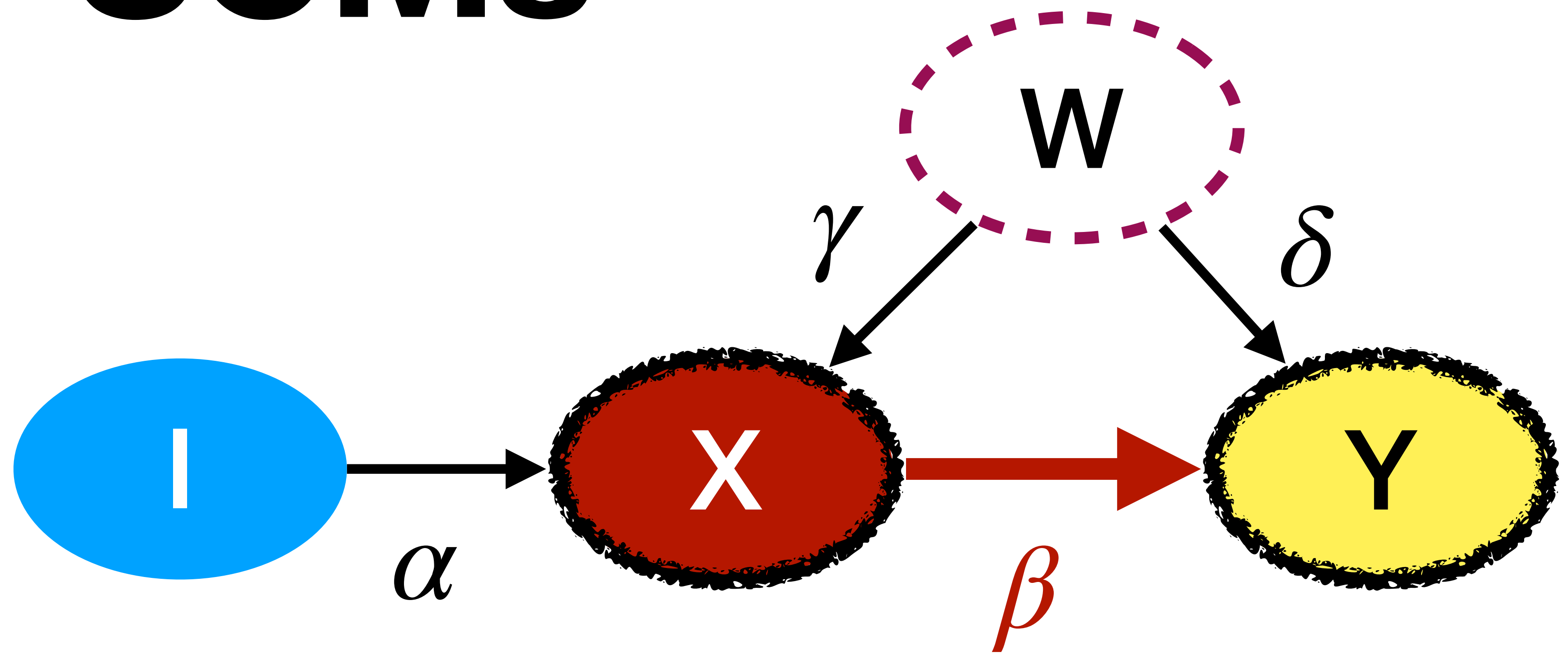
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



- Chocolate price CP does not cause Nobel Prizes N directly
- CP causes consumption of chocolate C
- CP is independent of U *(in this case, unclear)*

Instrumental variables - linear SCMs

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



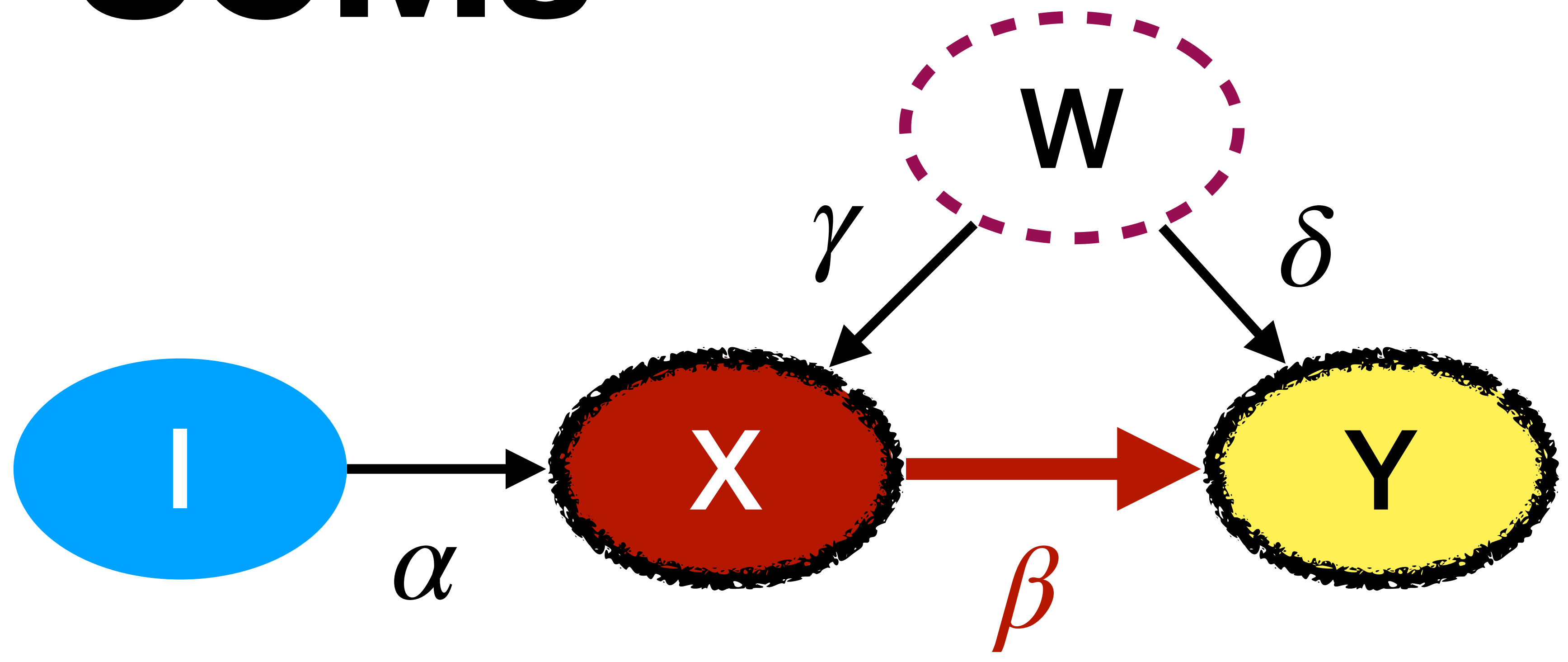
$$E[Y|\mathrm{do}(X=1)] - E[Y|\mathrm{do}(X=1)] = \beta$$

- Naive estimation, predict Y from X with OLS, take the coefficient

$$\hat{\beta}^{OLS} = \frac{\mathrm{Cov}(X,Y)}{Var(X)}$$

Instrumental variables - linear SCMs

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



$$E[Y|\text{do}(X=1)] - E[Y|\text{do}(X=1)] = \beta$$

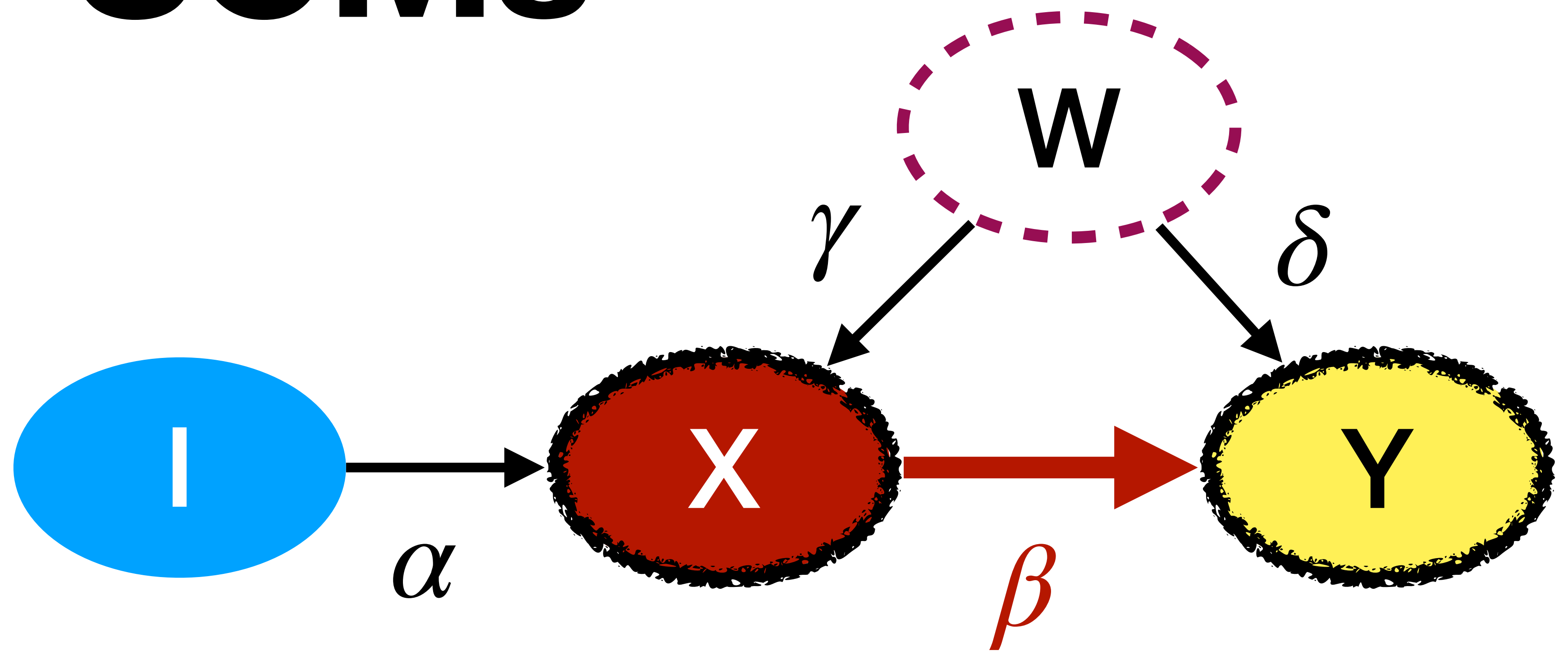
- Naive estimation, predict Y from X with OLS, take the coefficient

$$\hat{\beta}^{OLS} = \frac{\text{Cov}(X, Y)}{Var(X)} = \frac{\text{Cov}(X, \beta X + \delta W + \epsilon_Y)}{Var(X)}$$

$$Cov(aA + bB, C) = a \cdot Cov(A, C) + bCov(B, C)$$

Instrumental variables - linear SCMs

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



$$E[Y|\mathrm{do}(X=1)] - E[Y|\mathrm{do}(X=1)] = \beta$$

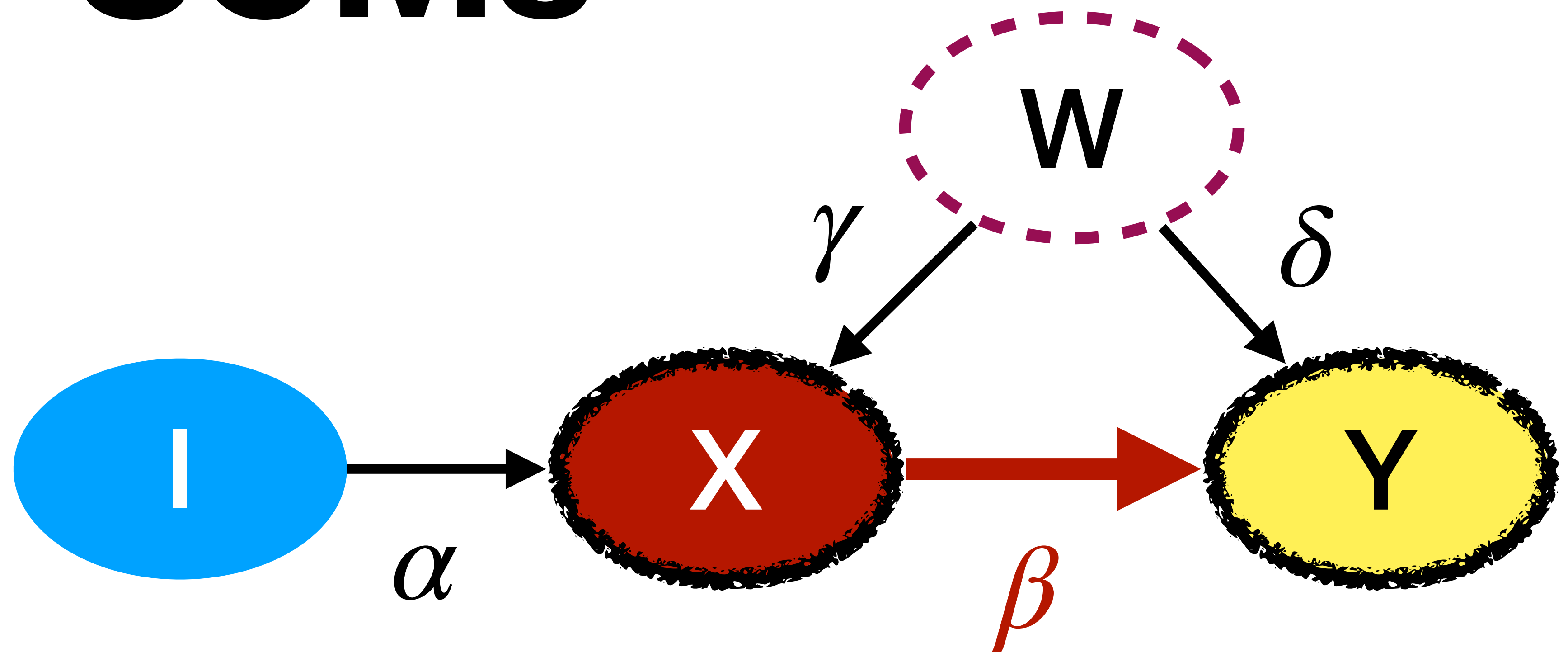
- Naive estimation, predict Y from X with OLS, take the coefficient

$$\hat{\beta}^{OLS} = \frac{\mathrm{Cov}(X,Y)}{Var(X)} = \frac{\mathrm{Cov}(X, \beta X + \delta W + \epsilon_Y)}{Var(X)} = \beta \frac{\mathrm{Cov}(X,X)}{Var(X)} + \delta \frac{\mathrm{Cov}(X,W)}{Var(X)} + \frac{\mathrm{Cov}(X,\epsilon_Y)}{Var(X)}$$

$$Cov(aA + bB, C) = a \cdot Cov(A,C) + bCov(B,C)$$

Instrumental variables - linear SCMs

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



$$E[Y|\mathrm{do}(X=1)] - E[Y|\mathrm{do}(X=1)] = \beta$$

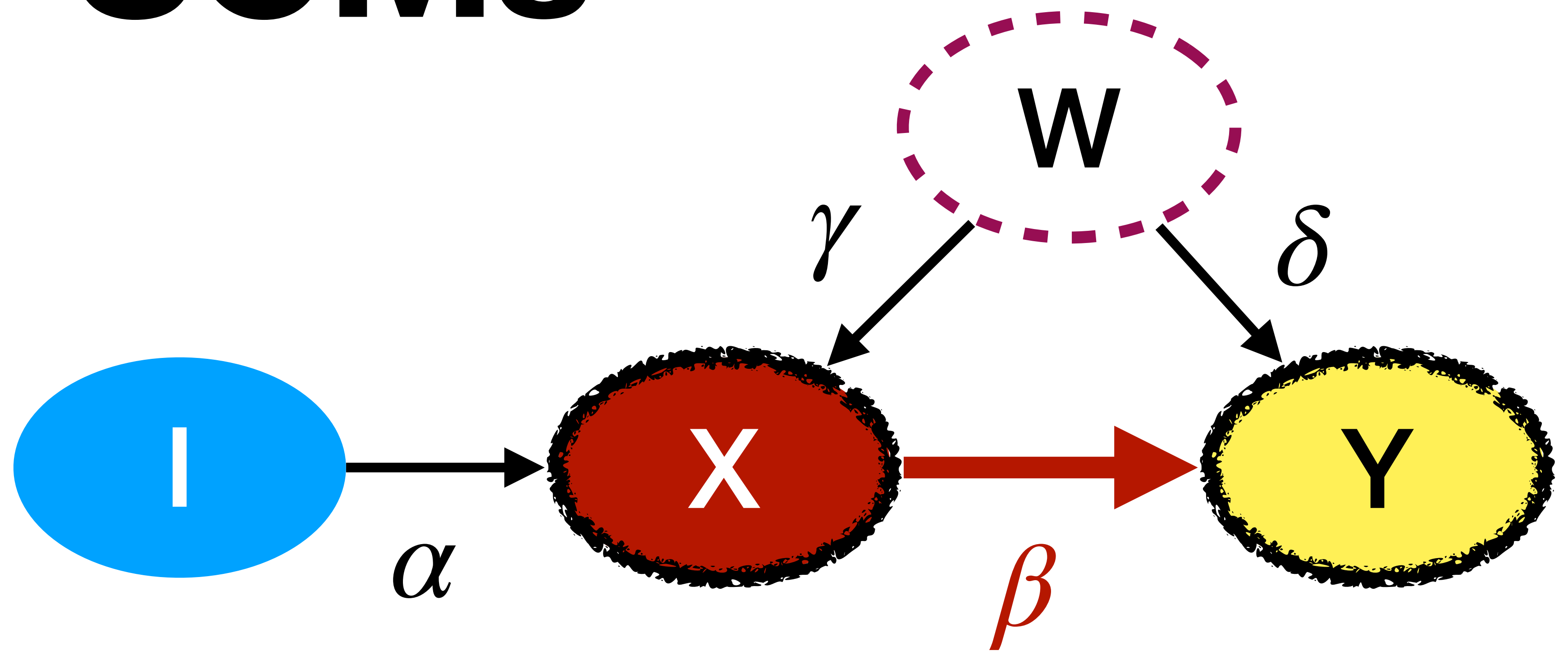
- Naive estimation, predict Y from X with OLS, take the coefficient

$$\hat{\beta}^{OLS} = \frac{\mathrm{Cov}(X,Y)}{Var(X)} = \frac{\mathrm{Cov}(X, \beta X + \delta W + \epsilon_Y)}{Var(X)} = \beta \frac{\mathrm{Cov}(X,X)}{Var(X)} + \delta \frac{\mathrm{Cov}(X,W)}{Var(X)} + \cancel{\frac{\mathrm{Cov}(X,\epsilon_Y)}{Var(X)}}$$

$$Cov(A,B) = 0 \text{ if } A \perp\!\!\!\perp B$$

Instrumental variables - linear SCMs

$$
\begin{cases}
I = \epsilon_I \\
X = \alpha I + \gamma W + \epsilon_X \\
Y = \beta X + \delta W + \epsilon_Y \\
\epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1)
\end{cases}
$$



$$E[Y \mid \mathrm{do}(X=1)] - E[Y \mid \mathrm{do}(X=1)] = \beta$$

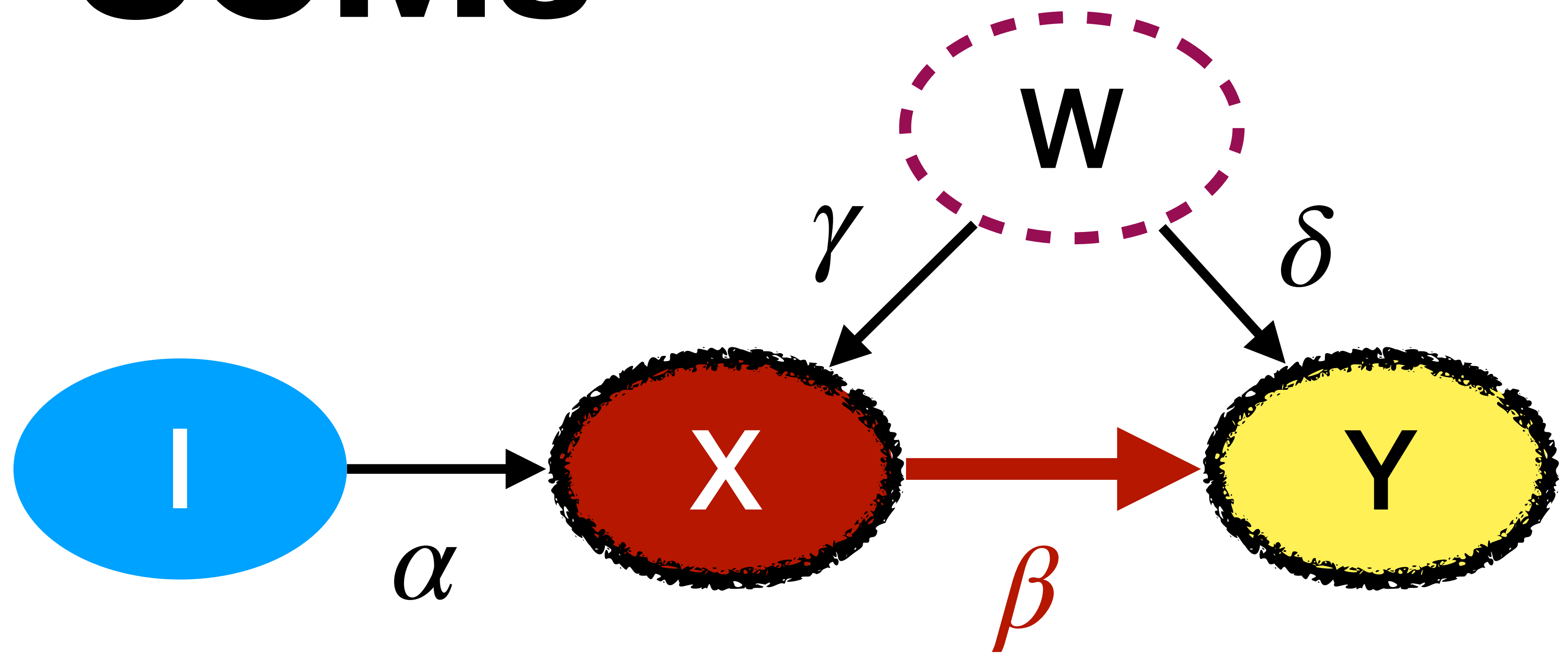
- Naive estimation, predict Y from X with OLS, take the coefficient

$$\hat{\beta}^{OLS} = \frac{\mathrm{Cov}(X,Y)}{Var(X)} = \frac{\mathrm{Cov}(X, \beta X + \delta W + \epsilon_Y)}{Var(X)} = \beta \frac{\mathrm{Cov}(X,X)}{Var(X)} + \delta \frac{\mathrm{Cov}(X,W)}{Var(X)}$$

$$Cov(A,A) = Var(A)$$

Instrumental variables - linear SCMs

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$

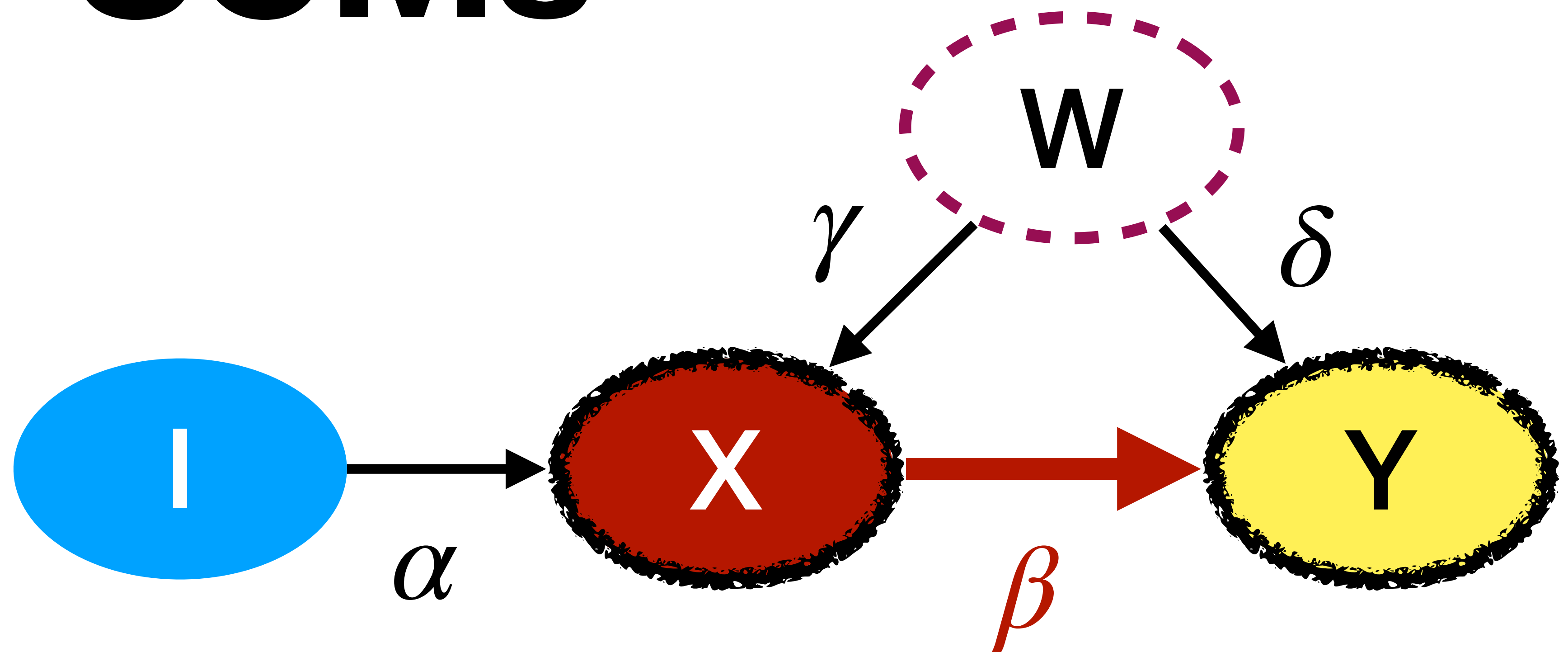


$$E[Y | \mathrm{do}(X = 1)] - E[Y | \mathrm{do}(X = 1)] = \beta$$

- Naive estimation, predict Y from X with OLS, take the coefficient

$$\hat{\beta}^{OLS} = \frac{\mathrm{Cov}(X, Y)}{Var(X)} = \frac{\mathrm{Cov}(X, \beta X + \delta W + \epsilon_Y)}{Var(X)} = \beta \cancel{\frac{\mathrm{Var}(X)}{Var(X)}} + \delta \frac{\mathrm{Cov}(X, W)}{Var(X)}$$

Instrumental variables - linear SCMs



$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$

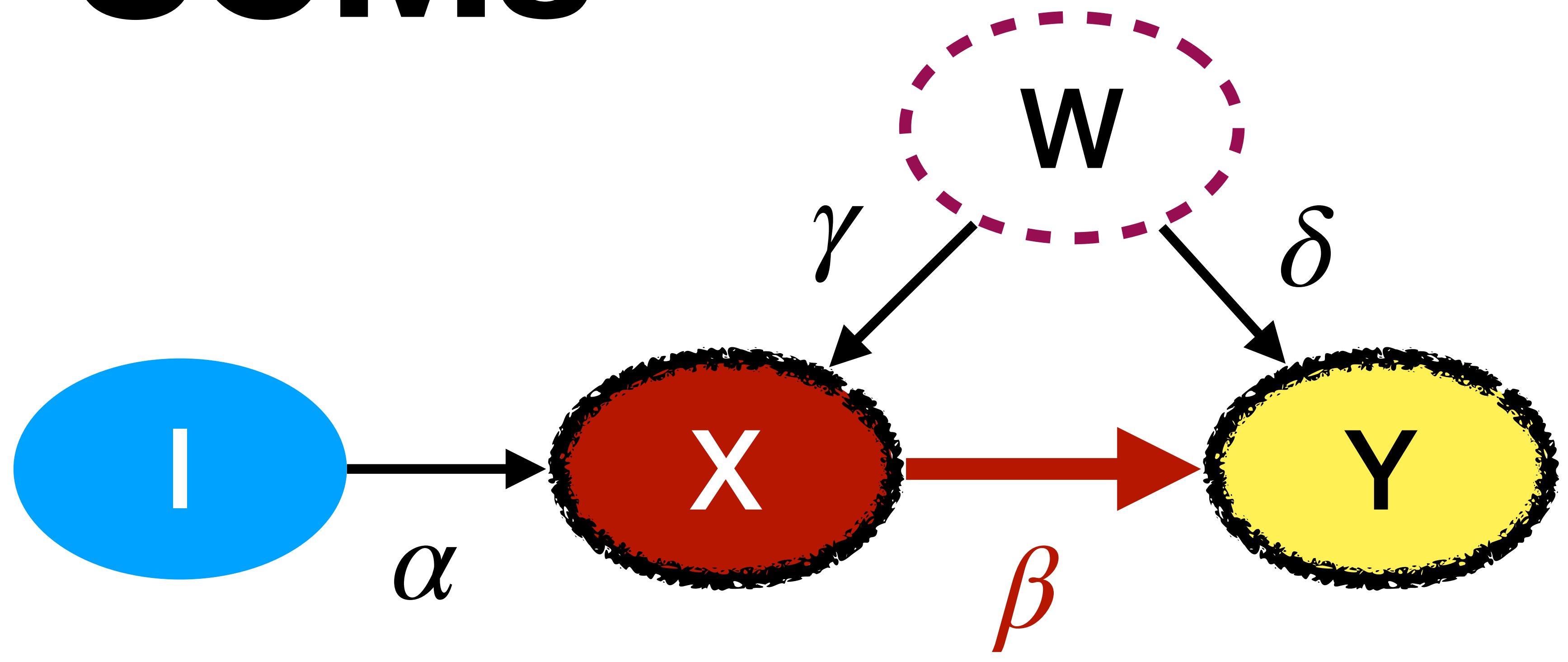
$$E[Y|\mathrm{do}(X=1)] - E[Y|\mathrm{do}(X=1)] = \beta$$

- Naive estimation, predict Y from X with OLS, take the coefficient

$$\hat{\beta}^{OLS} = \beta + \delta \frac{\mathrm{Cov}(\alpha I + \gamma W + \epsilon_X, W)}{Var(X)} = \beta + \delta\gamma \frac{\mathrm{Var}(W)}{Var(X)} + \delta \cancel{\frac{\mathrm{Cov}(\alpha I, W)}{Var(X)}} + \delta \cancel{\frac{\mathrm{Cov}(\epsilon_X, W)}{Var(X)}}$$

$$Cov(aA + bB, C) = a \cdot Cov(A, C) + bCov(B, C) \qquad Cov(A, B) = 0 \text{ if } A \perp\!\!\!\perp B$$

Instrumental variables - linear SCMs



$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$

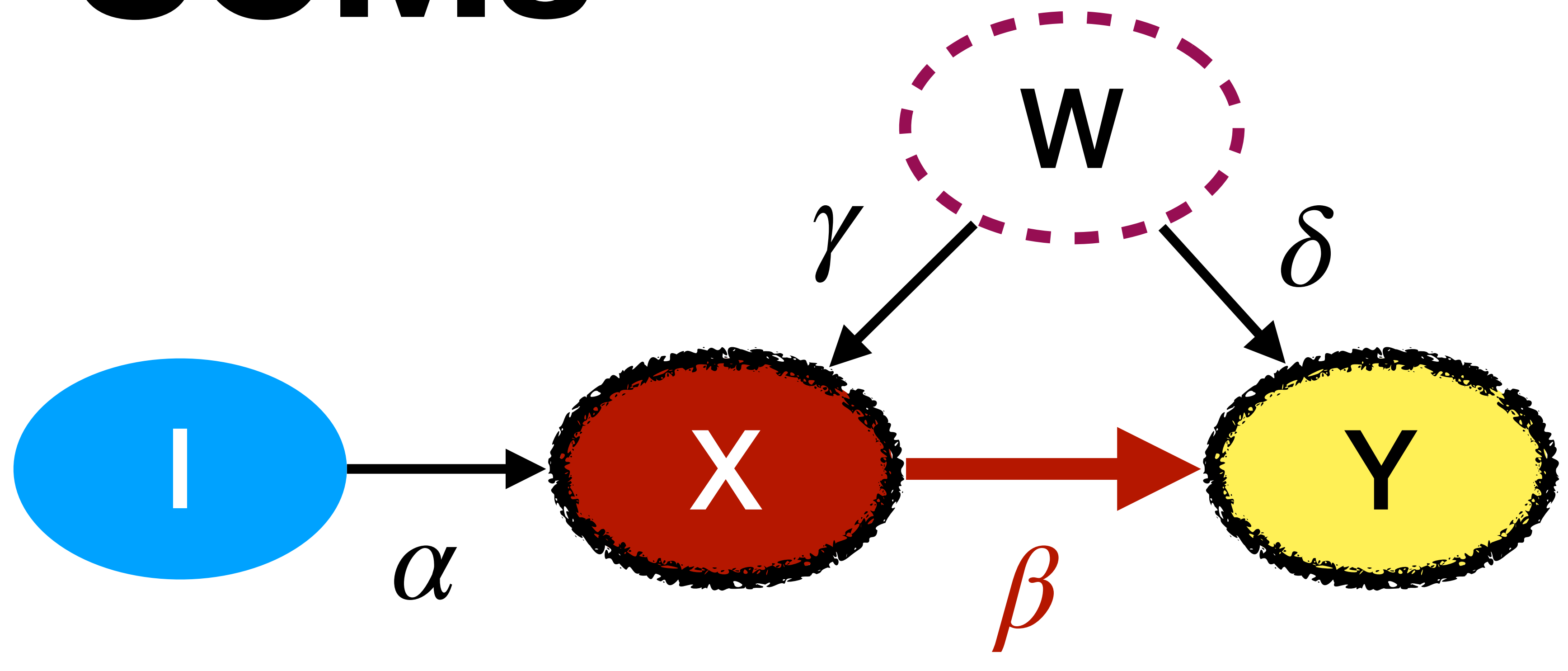
$$E[Y \,|\, \mathrm{do}(X=1)] - E[Y \,|\, \mathrm{do}(X=1)] = \beta$$

- Naive estimation, predict Y from X with OLS, take the coefficient

$$\hat{\beta}^{OLS} = \beta + \delta \frac{\mathrm{Cov}(\alpha I + \gamma W + \epsilon_X, W)}{Var(X)} = \beta + \delta\gamma \frac{\mathrm{Var}(W)}{Var(X)}$$

Instrumental variables - linear SCMs

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



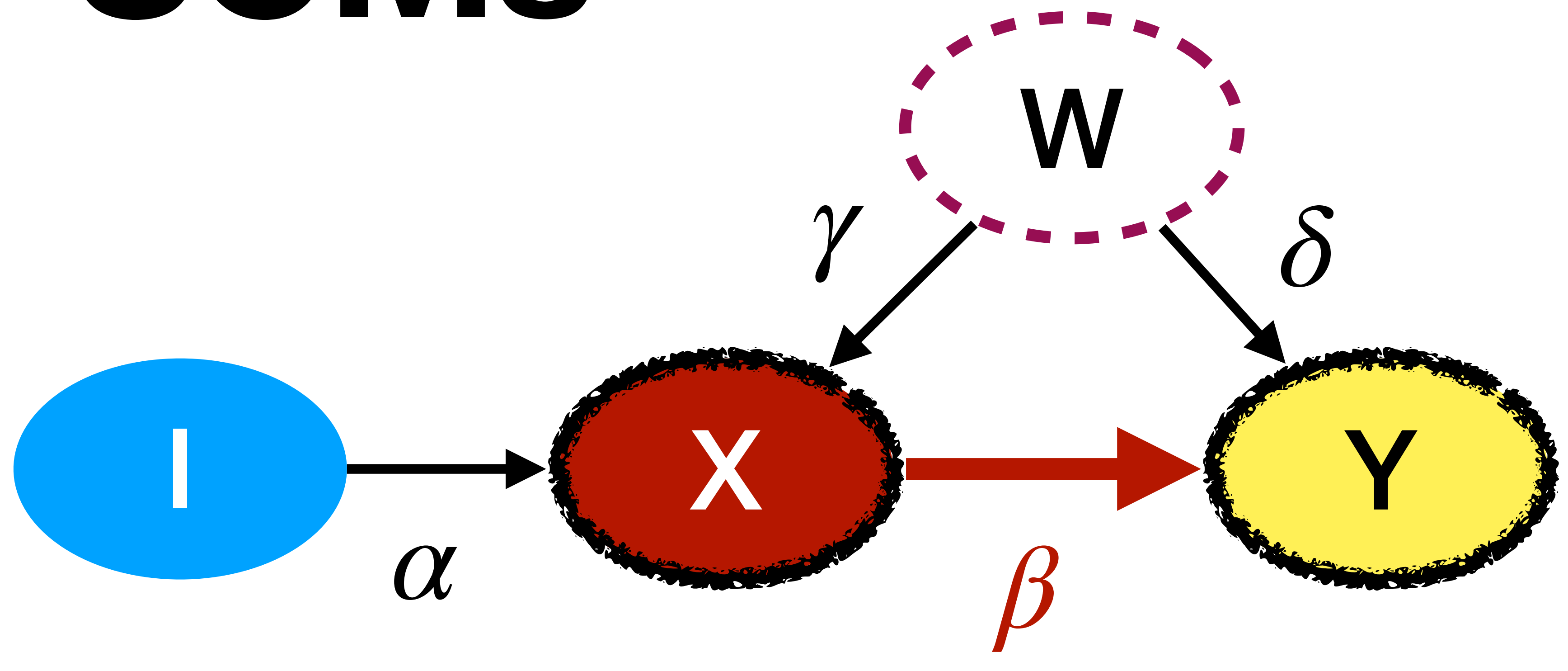
$$\hat{\beta}^{OLS} = \beta + \gamma\delta \frac{\text{Var}(W)}{Var(X)}$$

$$\text{Cov}(I, X) = \text{Cov}(I, \alpha I + \gamma W + \epsilon_X) = \alpha \text{Cov}(I, I) + \gamma \cancel{\text{Cov}(I, W)} + \cancel{\text{Cov}(I, \epsilon_X)}$$

$Cov(aA + bB, C) = a \cdot Cov(A, C) + bCov(B, C)$ $Cov(A, B) = 0 \text{ if } A \perp\!\!\!\perp B$

Instrumental variables - linear SCMs

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



$$\hat{\beta}^{OLS} = \beta + \gamma\delta \frac{\text{Var}(W)}{Var(X)}$$

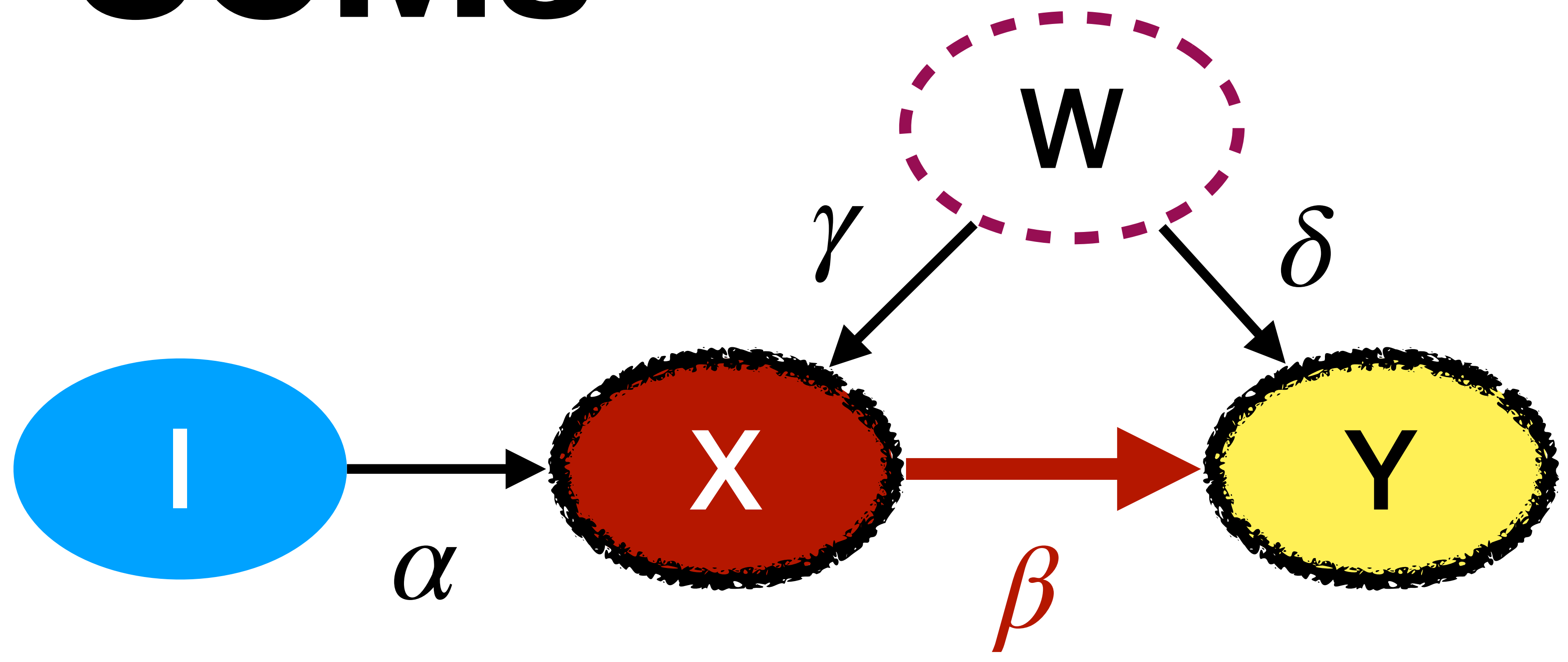
$$\text{Cov}(I, X) = \text{Cov}(I, \alpha I + \gamma W + \epsilon_X) = \alpha \text{Cov}(I, I) = \alpha \text{Var}(I)$$

$$Cov(A, A) = Var(A)$$

$$Cov(aA + bB, C) = a \cdot Cov(A, C) + bCov(B, C) \qquad Cov(A, B) = 0 \text{ if } A \perp\!\!\!\perp B$$

Instrumental variables - linear SCMs

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



$$\hat{\beta}^{OLS} = \beta + \gamma\delta \frac{\mathrm{Var}(W)}{Var(X)}$$

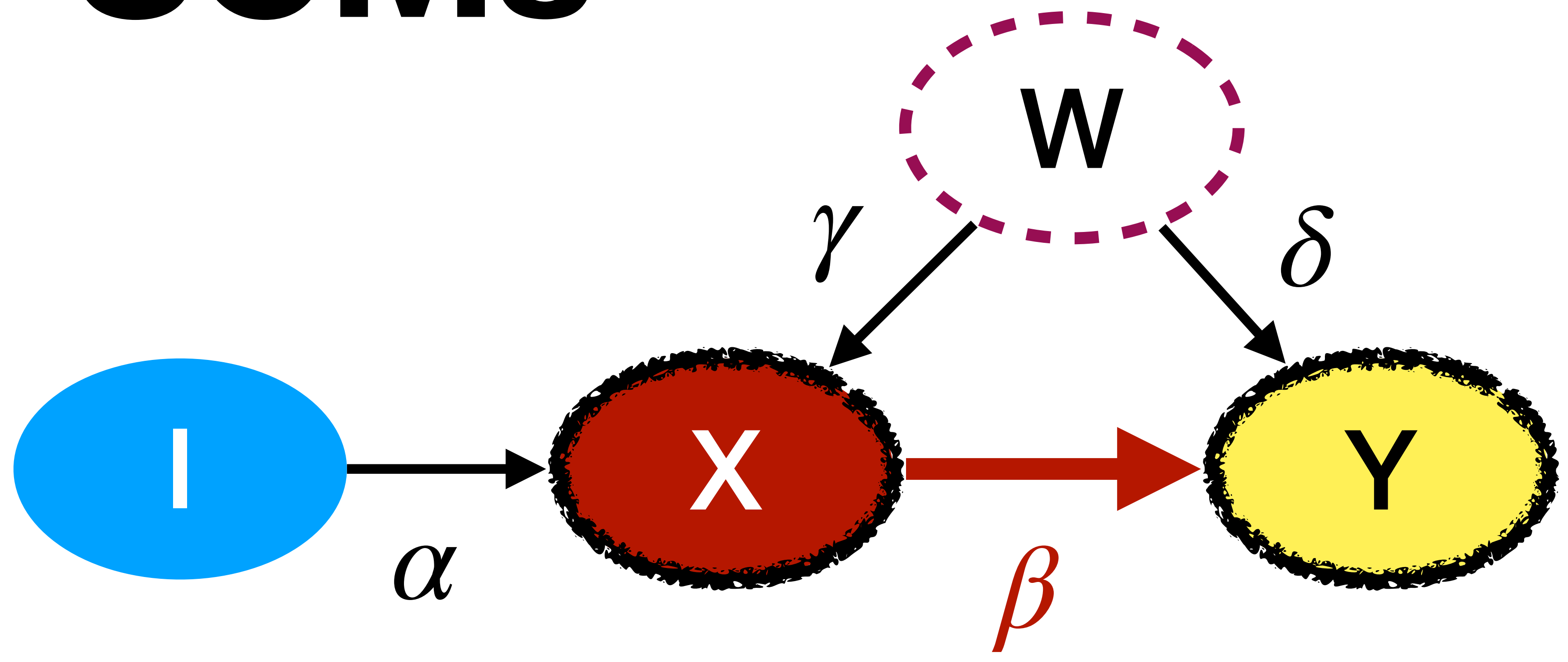
$$\mathrm{Cov}(I, Y) = \mathrm{Cov}(I, \beta X + \delta W + \epsilon_Y) = \beta \mathrm{Cov}(I, X) + \delta \mathrm{Cov}(I, W) + \mathrm{Cov}(I, \epsilon_Y)$$

$$Cov(A, A) = Var(A)$$

$$Cov(aA + bB, C) = a \cdot Cov(A, C) + bCov(B, C) \qquad Cov(A, B) = 0 \text{ if } A \perp\!\!\!\perp B$$

Instrumental variables - linear SCMs

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



$$\hat{\beta}^{OLS} = \beta + \gamma\delta \frac{\text{Var}(W)}{Var(X)}$$

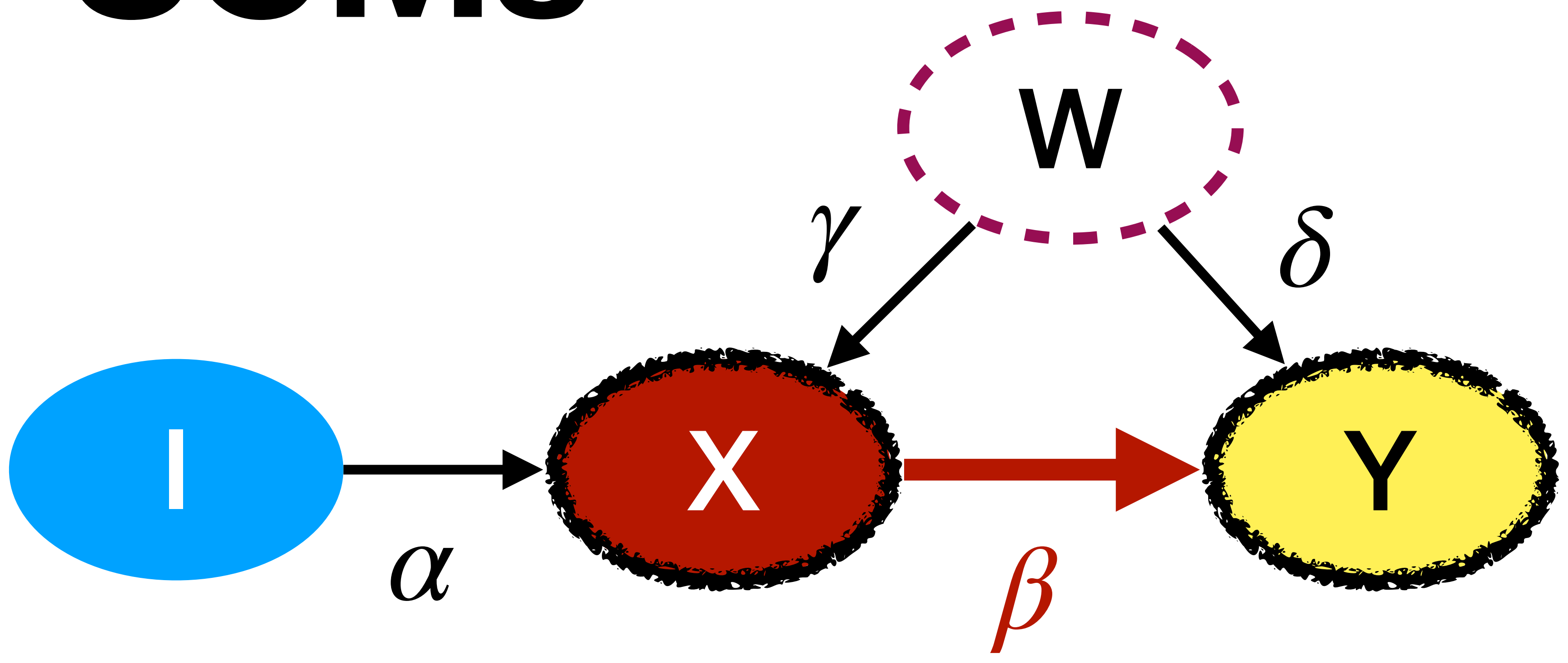
$$\text{Cov}(I, Y) = \text{Cov}(I, \beta X + \delta W + \epsilon_Y) = \beta \text{Cov}(I, X) + \delta \text{Cov}(\cancel{I}, W) + \text{Cov}(\cancel{I}, \epsilon_Y)$$

$$Cov(A, A) = Var(A)$$

$$Cov(aA + bB, C) = a \cdot Cov(A, C) + bCov(B, C) \qquad Cov(A, B) = 0 \text{ if } A \perp\!\!\!\perp B$$

Instrumental variables - linear SCMs

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



$$\hat{\beta}^{OLS} = \beta + \gamma\delta \frac{\text{Var}(W)}{Var(X)}$$

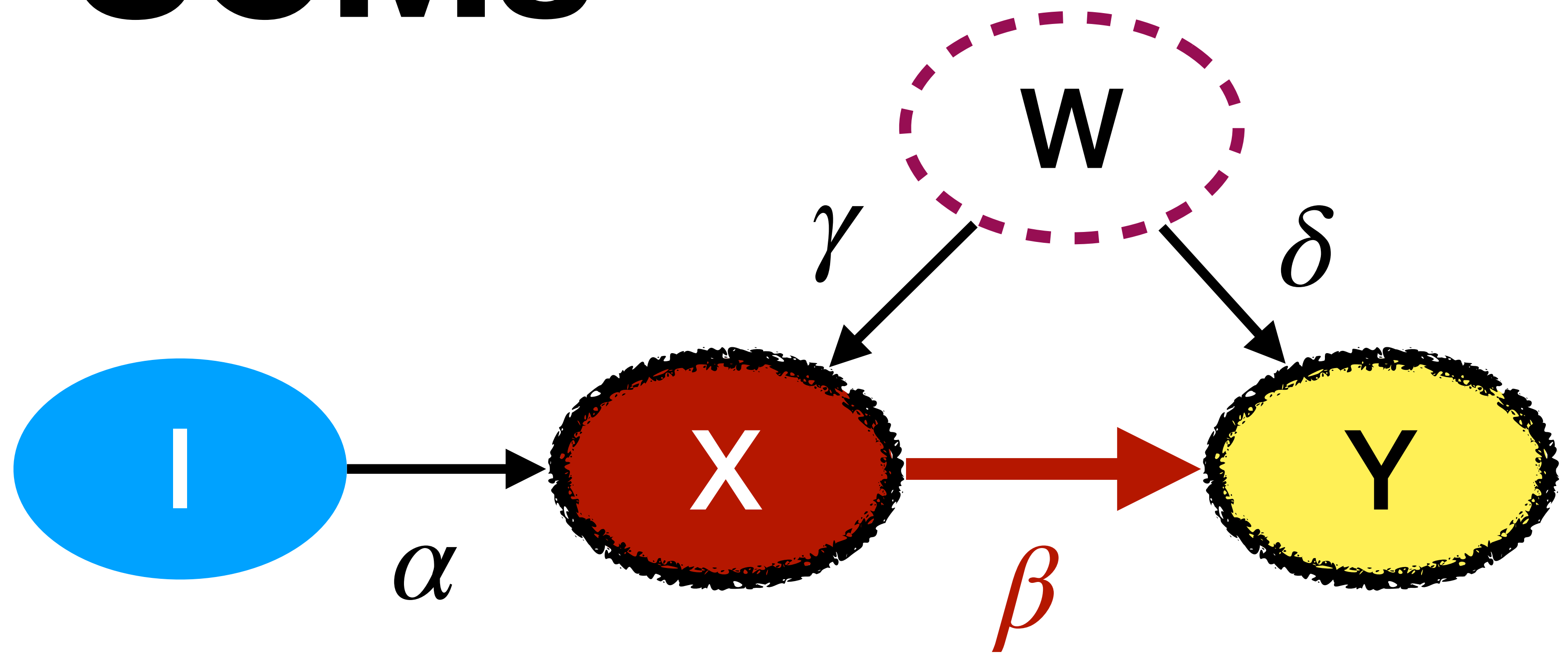
$$\text{Cov}(I, Y) = \text{Cov}(I, \beta X + \delta W + \epsilon_Y) = \beta \text{Cov}(I, X) = \beta \cdot \alpha Var(I)$$

$$Cov(A, A) = Var(A)$$

$$Cov(aA + bB, C) = a \cdot Cov(A, C) + bCov(B, C)$$

$$Cov(A, B) = 0 \text{ if } A \perp\!\!\!\perp B$$

Instrumental variables - linear SCMs



$$
\begin{cases}
I = \epsilon_I \\
X = \alpha I + \gamma W + \epsilon_X \\
Y = \beta X + \delta W + \epsilon_Y \\
\epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1)
\end{cases}
$$

$$\hat{\beta}^{OLS} = \beta + \gamma\delta \frac{\mathrm{Var}(W)}{Var(X)}$$

$$\mathrm{Cov}(I, X) = \alpha \mathrm{Var}(I)$$

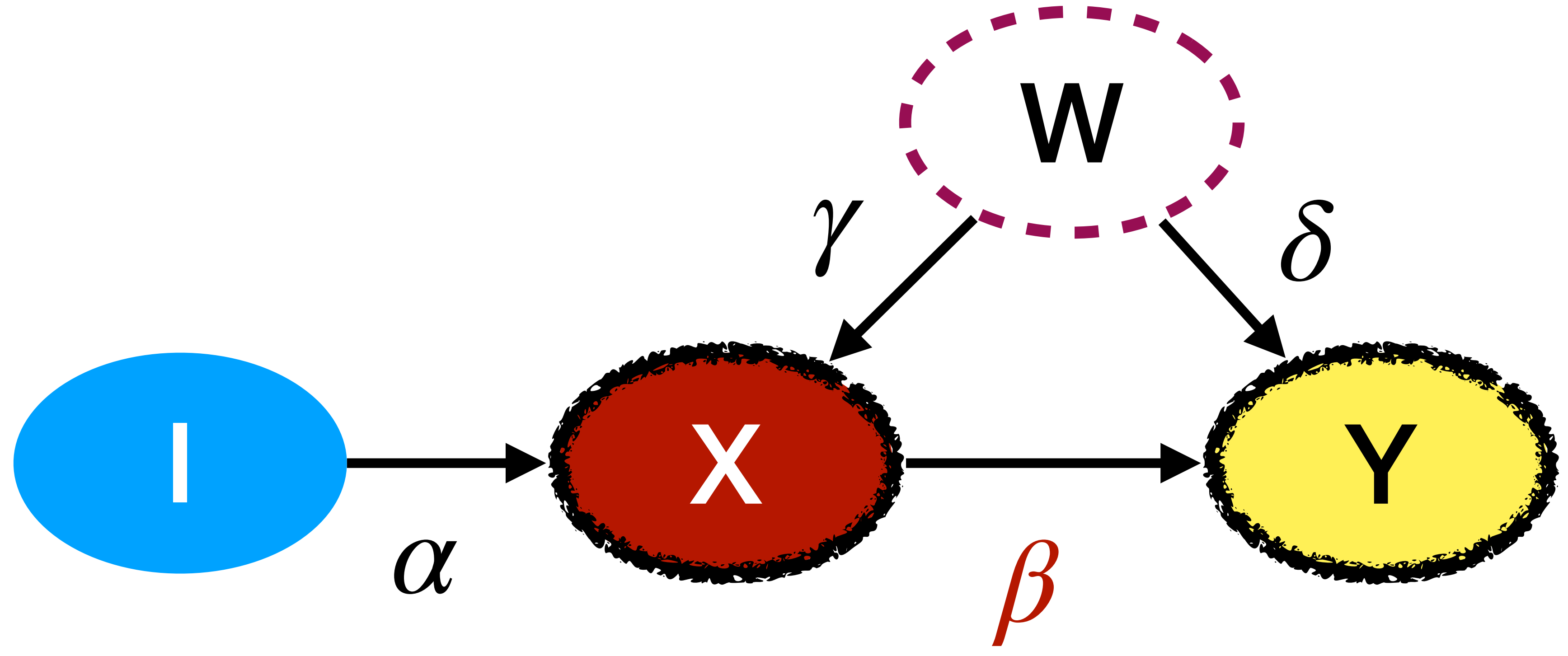
$$\mathrm{Cov}(I, Y) = \beta \cdot \alpha Var(I)$$

$$\frac{\mathrm{Cov}(I, Y)}{\mathrm{Cov}(I, X)} = \frac{\beta \cdot \cancel{\alpha \mathrm{Var}(I)}}{\cancel{\alpha \mathrm{Var}(I)}} = \beta$$

A more general approach: Two Stage Least Squares (2SLS)

1. Regress X on I

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



$$\hat{\alpha} = \frac{\mathrm{Cov}(X, I)}{Var(I)} = \frac{\mathrm{Cov}(\alpha I + \gamma W + \epsilon_X, I)}{Var(I)} = \alpha \frac{\mathrm{Cov}(I, I)}{Var(I)} = \alpha \cancel{\frac{\mathrm{Var}(I)}{Var(I)}}$$

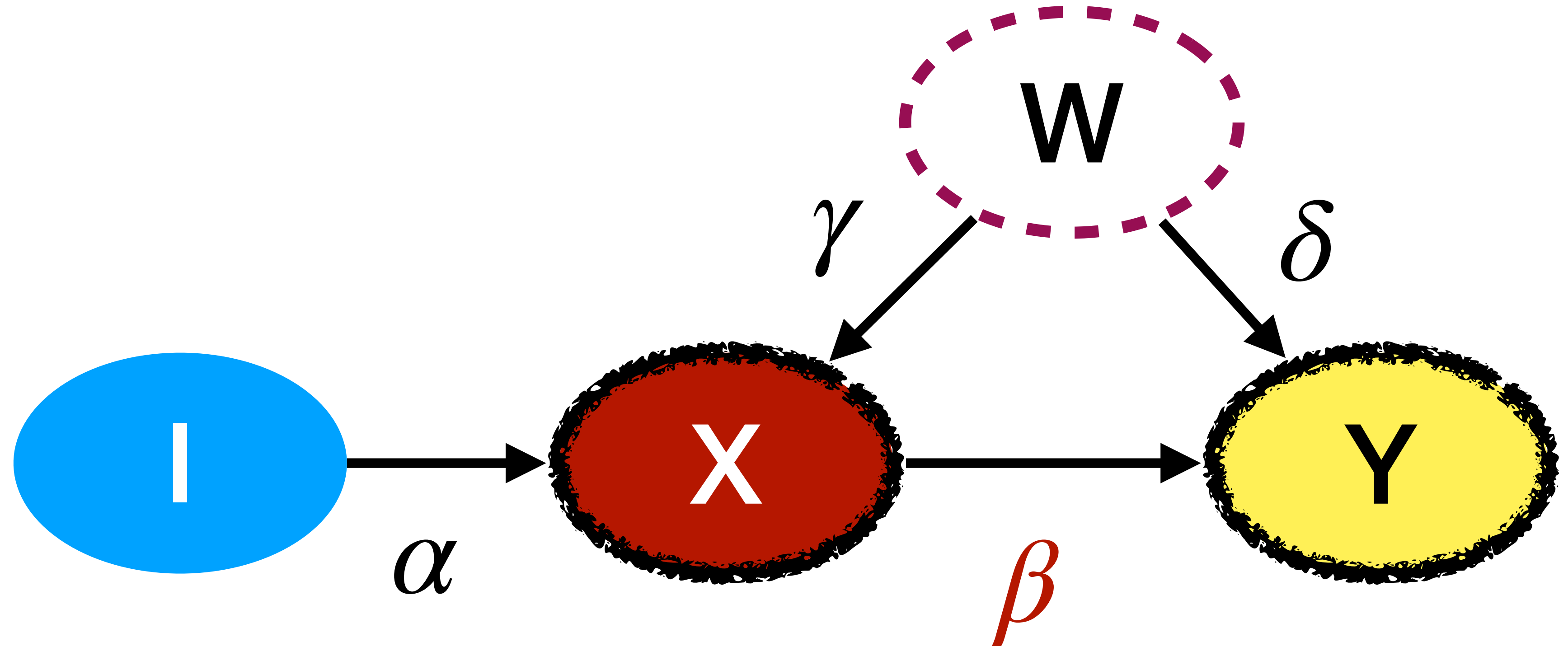
2. Construct $\hat{X} = \hat{\alpha} I$

3. Regress Y on $\hat{X}$ to obtain β

$$\hat{\beta} = \frac{\mathrm{Cov}(\hat{X}, Y)}{Var(\hat{X})} = \frac{\mathrm{Cov}(\alpha I, Y)}{Var(\alpha I)} = \alpha \frac{\mathrm{Cov}(I, \beta X + \delta W + \epsilon_Y)}{\alpha^2 \mathrm{Var}(I)} = \alpha \frac{\mathrm{Cov}(I, \beta \alpha I + \gamma W + \epsilon_X)}{\alpha^2 \mathrm{Var}(I)}$$

A more general approach: Two Stage Least Squares (2SLS)

$$\begin{cases} I = \epsilon_I \\ X = \alpha I + \gamma W + \epsilon_X \\ Y = \beta X + \delta W + \epsilon_Y \\ \epsilon_I, \epsilon_X, \epsilon_Y, W \sim N(0,1) \end{cases}$$



1. Regress X on I

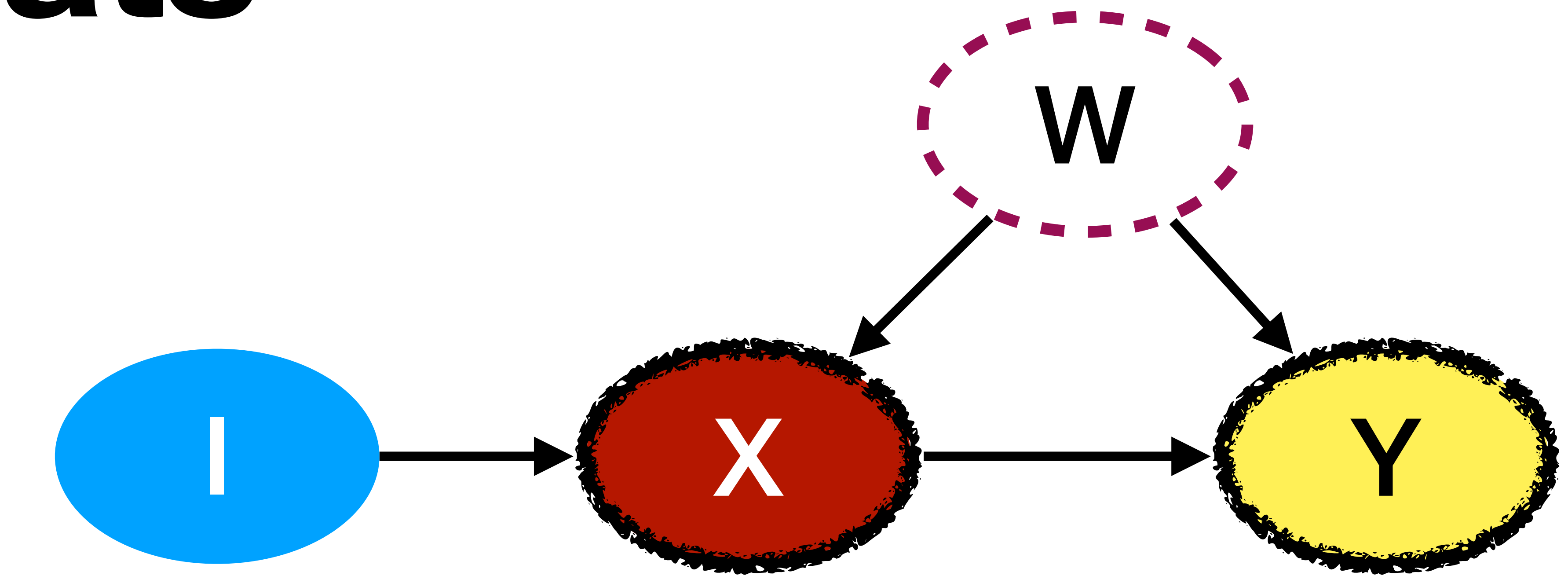
$$\hat{\alpha} = \frac{\mathrm{Cov}(X, I)}{Var(I)} = \frac{\mathrm{Cov}(\alpha I + \gamma W + \epsilon_X, I)}{Var(I)} = \alpha \frac{\mathrm{Cov}(I, I)}{Var(I)} = \alpha \cancel{\frac{\mathrm{Var}(I)}{Var(I)}}$$

2. Construct $\hat{X} = \hat{\alpha} I$

3. Regress Y on $\hat{X}$ to obtain β

$$\hat{\beta} = \frac{\mathrm{Cov}(\hat{X}, Y)}{Var(\hat{X})} = \alpha \frac{\mathrm{Cov}(I, \beta\alpha I + (\beta\delta + \gamma)W + \beta\epsilon_X + \epsilon_Y)}{\alpha^2 \mathrm{Var}(I)} = \alpha^2 \cdot \beta \cancel{\frac{\mathrm{Var}(I)}{\alpha^2 \mathrm{Var}(I)}} = \beta$$

Instrumental variables - caveats

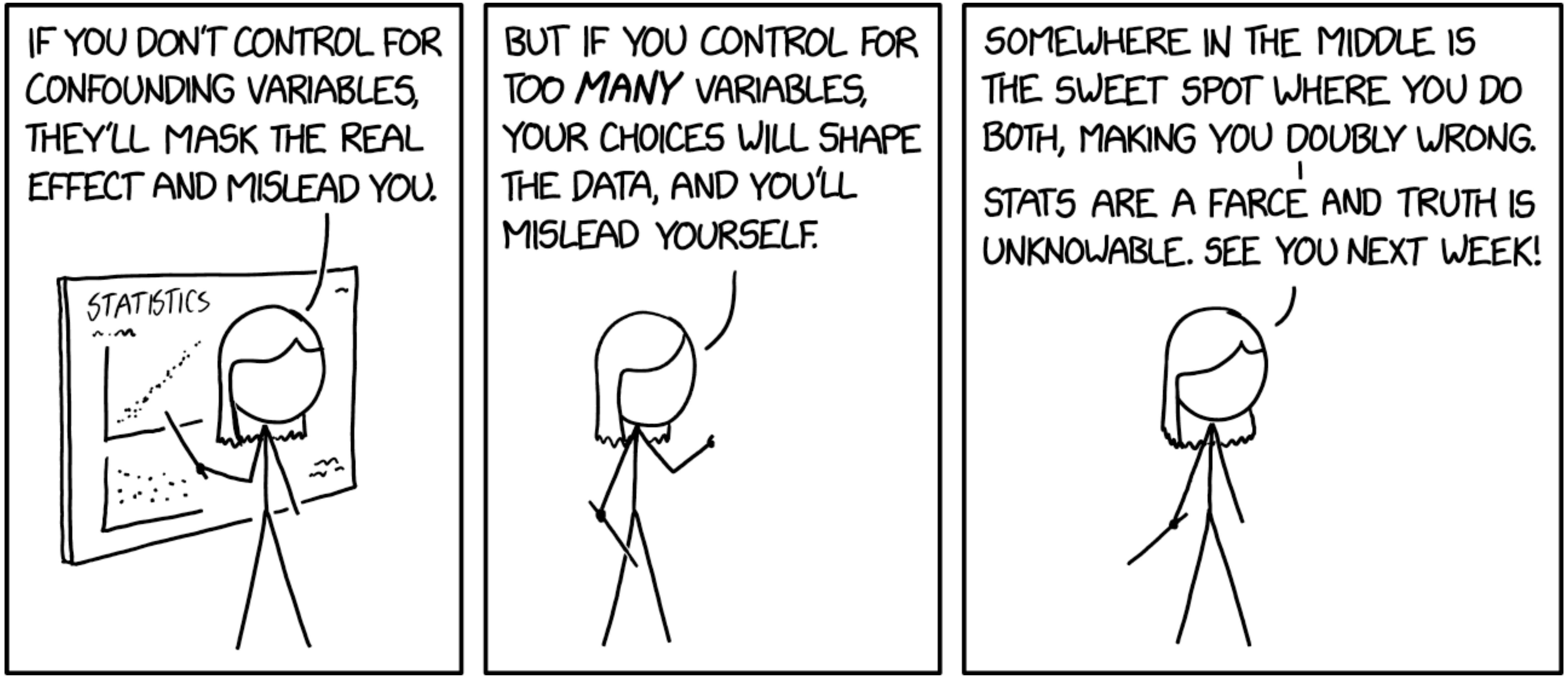


- **Weak instrument:** $cov(I, X)$ is small
  - Small errors in $cov(I, X)$ lead to big errors in β^{IV}
  - High variance in the estimator

- Strong requirements: $I \to X$, but $I \nrightarrow Y$ directly, $I \perp\!\!\!\perp W$

- **Conditional IV:** $\mathbf{S}$ s.t. $I \perp\!\!\!\perp X | \mathbf{S}$ and every path from I to Y that is not blocked by $\mathbf{S}$ has an edge into X , **can use 2SLS**

Questions?



https://xkcd.com/2560/