

# Biais de mesure

Vincent Arel-Bundock



# Types de mesure

Grande variété!

Observation (quasi-)directe:

- la température
- la consommation
- les données démographiques
- les indicateurs économiques

Observation indirecte:

- le bonheur, mesuré par des sondages
- les réactions émotionnelles, mesurées par des capteurs de conductance cutanée
- la teneur des discours, mesurée par l'analyse textuelle
- l'embarras, mesuré par des jeux stratégiques en laboratoire

# Qualités d'un instrument de mesure

Un instrument est bon si la mesure qu'il produit est *valide* et *fidèle*.

Validité de construit:

Adéquation entre un concept et la mesure employée pour l'opérationnaliser.

Une mesure est valide si :

- Elle offre une bonne "traduction" du concept, c.-à-d. si elle permet de généraliser à partir de l'observation concrète jusqu'au concept abstrait.
- Un changement dans cette mesure implique un changement dans le concept qu'elle représente.

Deux types de validité de construit :

1. Validité convergente: plusieurs mesures d'un même concept convergent vers un même résultat.
2. Validité discriminante: une mesure capture un seul et unique concept.

# Qualités d'un instrument de mesure

Fidélité:

Notre capacité à mesurer le concept d'intérêt sans faire d'erreurs accidentnelles ou aléatoires.

Lorsque'un instrument de mesure est fidèle :

- Il capture le concept d'intérêt avec précision, sans faire (trop) d'erreurs.
- Mesurer le même phénomène à répétition produirait approximativement le même résultat à chaque fois.

Note: Un instrument de mesure pourrait être valide sans être fidèle. Ici, nous nous intéressons aux conséquences du manque de fidélité: erreur de mesure.

# Analyse graphique

# Analyse graphique du biais de mesure

Objectif:

- Estimer l'effet causal de  $X$  sur  $Y$ .

Problème:

- $X$  et  $Y$  sont impossibles à observer directement.

La variable observée  $\tilde{X}$  pourrait être construite ainsi:

$$X \longrightarrow \tilde{X} \longleftarrow U_X$$

# Analyse graphique du biais de mesure

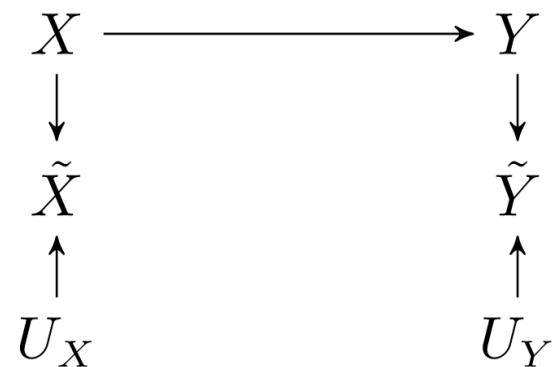
Il est utile de distinguer trois types d'erreurs de mesure:

1. Erreur indépendante,
2. Erreur différentielle, et
3. Erreur sur les variables de contrôle.

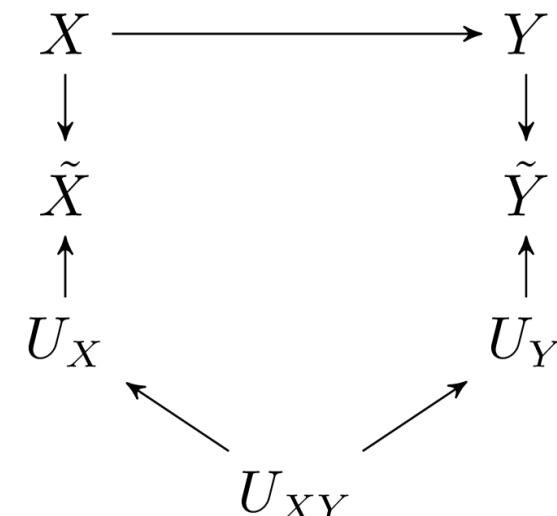
# Erreur de mesure indépendante

Les erreurs de mesure qui affectent  $X$  et  $Y$  sont "indépendantes" si il n'existe aucun chemin ouvert entre elles.

(A)  
Erreurs indépendantes



(B)  
Erreurs dépendantes

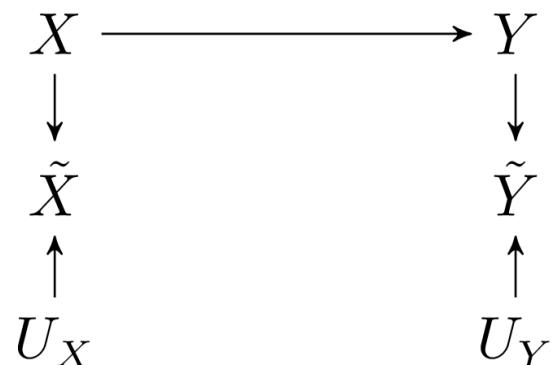


# Erreur de mesure indépendante

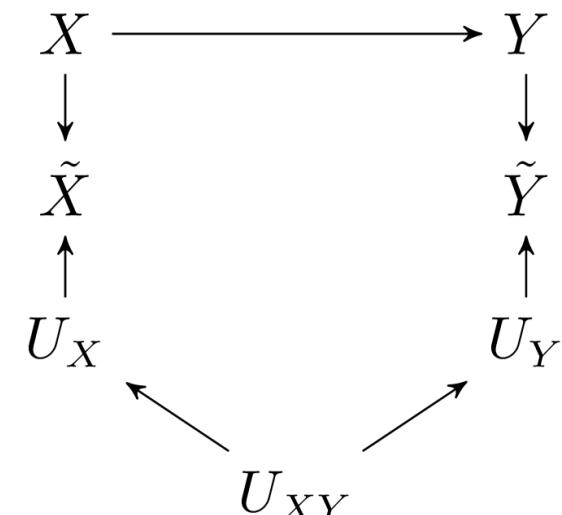
Exemple:

- Questions de sondage controversées et biais de désirabilité sociale.
- Si certains individus ont plus le souci de bien paraître que d'autres, les mesures prises sur ces individus pourraient être systématiquement biaisées, et ces erreurs de mesure pourraient être liées d'une mesure à l'autre.

(A)  
Erreurs indépendantes



(B)  
Erreurs dépendantes



# Erreur de mesure différentielle

L'erreur de mesure sur la cause  $X$  est "non différentielle" si :

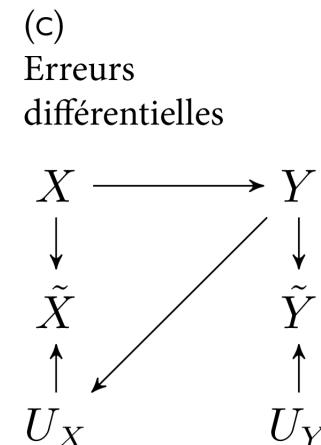
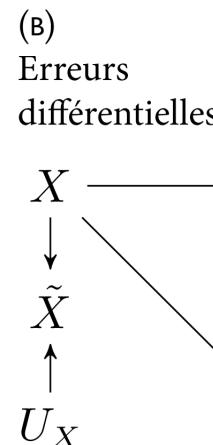
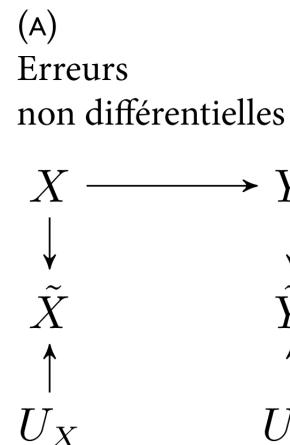
- Elle est indépendante de l'effet  $Y$ .
- Elle est indépendante de la cause  $X$ .

# Erreur de mesure différentielle

Exemple :

- Étude sur les effets de la consommation de drogues ( $X$ ) sur la santé ( $Y$ ).
- On mesure  $X$  à l'aide d'un test sanguin.
- On mesure  $Y$  en administrant un questionnaire aux participants de l'étude.

Si la consommation de drogue pousse les répondants à faire des erreurs systématiques en rapportant leur état de santé, l'erreur de mesure sera différentielle.



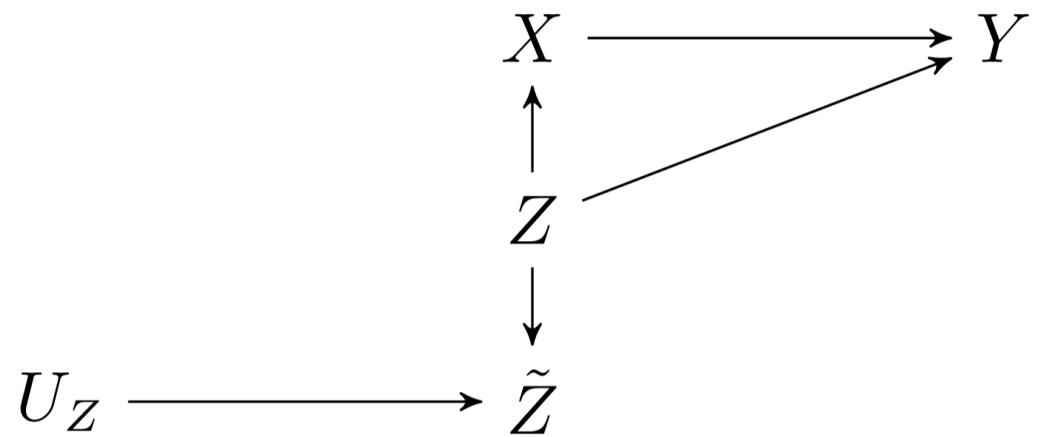
# Erreur de mesure dans les variables de contrôle

Objectif:

- Contrôler pour  $Z$  pour estimer l'effet causal de  $X$  sur  $Y$ .

Problèmes:

- $Z$  est non-observable
- $\tilde{Z}$  est mesurée avec erreur
- Le chemin par la porte arrière n'est pas entièrement fermé



# Mauvaises nouvelles

Quand il y a erreur de mesure, l'estimé de l'effet causal de  $X$  sur  $Y$  risque d'être biaisé.

La force du biais est proportionnelle à la taille de l'erreur de mesure.

Au-delà de ça, la taille ou la direction du biais de mesure est généralement très difficile à anticiper en pratique, surtout lorsque les erreurs sont dépendantes et/ou différentielles.

Conclusion vexante, parce que l'erreur de mesure est omniprésente en sciences sociales.

# Analyse algébrique



# Analyse algébrique

Deux cas d'erreur de mesure dans un modèle de régression linéaire bivarié:

1. Erreur de mesure dans la variable dépendante  $Y$ 
  - Pas de biais; hausse de l'incertitude
2. Erreur de mesure dans la variable indépendante  $X$ 
  - Biais d'atténuation sur l'estimé du coefficient de régression.

# Erreur dans la variable dépendante: Incertitude

Objectif: estimer  $\beta$

$$Y = \alpha + \beta \cdot X + \varepsilon$$

Problème: Erreur dans la variable dépendante

$$\tilde{Y} = Y + \eta,$$

où  $\eta$  représente une erreur de mesure centrée à zéro et indépendante de  $X$  et  $Y$ .

$$\tilde{Y} = \alpha + \beta \cdot X + \tilde{\varepsilon}$$

Cette équation peut-être réexprimée ainsi:

$$\begin{aligned} Y + \eta &= \alpha + \beta \cdot X + \tilde{\varepsilon} \\ Y &= \alpha + \beta \cdot X + (\tilde{\varepsilon} - \eta) \end{aligned}$$

# Erreur dans la variable dépendante: Incertitude

La seule différence entre le modèle qu'on *aimerait* estimer et celui qu'on *peut* estimer est le terme d'erreur:

$$\begin{aligned}Y &= \alpha + \beta \cdot X + \varepsilon \\Y &= \alpha + \beta \cdot X + (\tilde{\varepsilon} - \eta)\end{aligned}$$

Variance d'une somme:

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

Erreur type de  $\hat{\beta}$ :

$$\sigma_{\beta} = \frac{\sigma_{\varepsilon}}{\sigma_X \cdot \sqrt{n}}$$

L'erreur de mesure  $\eta$  enflle la variance du résidu de régression. Pas de biais, mais les erreurs types seront (correctement) plus grandes.

# Erreur dans la variable indépendante: Biais d'atténuation

On veut estimer le modèle suivant:

$$Y = \alpha + \beta \cdot X + \varepsilon$$

La vraie valeur de  $X$  est impossible à mesurer précisément. On peut seulement mesurer  $\tilde{X}$ , égale à :

$$\tilde{X} = X + v,$$

où  $v$  représente un terme d'erreur aléatoire.

On estime donc le modèle suivant:

$$Y = \tilde{\alpha} + \tilde{\beta} \cdot \tilde{X} + \tilde{\varepsilon}$$

# Erreur dans la variable indépendante: Biais d'atténuation

Si l'erreur de mesure est indépendante de  $X$  et de  $Y$ , alors :

$$\text{Cov}(v, X) = \text{Cov}(v, Y) = 0$$

En exploitant ces égalités, on réexprime le coefficient de régression:

$$\begin{aligned}\tilde{\beta} &= \frac{\text{Cov}(\tilde{X}, Y)}{\text{Var}(\tilde{X})} \\ &= \frac{\text{Cov}(X + v, Y)}{\text{Var}(X + v)} \\ &= \frac{\text{Cov}(X, Y) + \text{Cov}(v, Y)}{\text{Var}(X) + \text{Var}(v) + 2 \cdot \text{Cov}(v, X)} \\ &= \frac{\text{Cov}(X, Y)}{\text{Var}(X) + \text{Var}(v)}\end{aligned}$$

# Erreur dans la variable indépendante: Biais d'atténuation

Le coefficient du modèle avec erreur dans  $X$  est  $\tilde{\beta}$  :

$$\tilde{\beta} = \frac{\text{Cov}(X, Y)}{\text{Var}(X) + \text{Var}(v)}$$

Le coefficient du modèle sans erreur dans  $X$  est  $\beta$ :

$$\beta = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

À moins que l'erreur de mesure soit constante ( $\text{Var}(v) = 0$ ), le dénominateur  $\text{Var}(X) + \text{Var}(v)$  est trop grand, et le coefficient de régression aura tendance à être trop près de zéro.

Biais d'atténuation: La valeur absolue de l'estimé de  $\tilde{\beta}$  sera plus petite que la valeur absolue du vrai coefficient  $\beta$ .

Merci!

