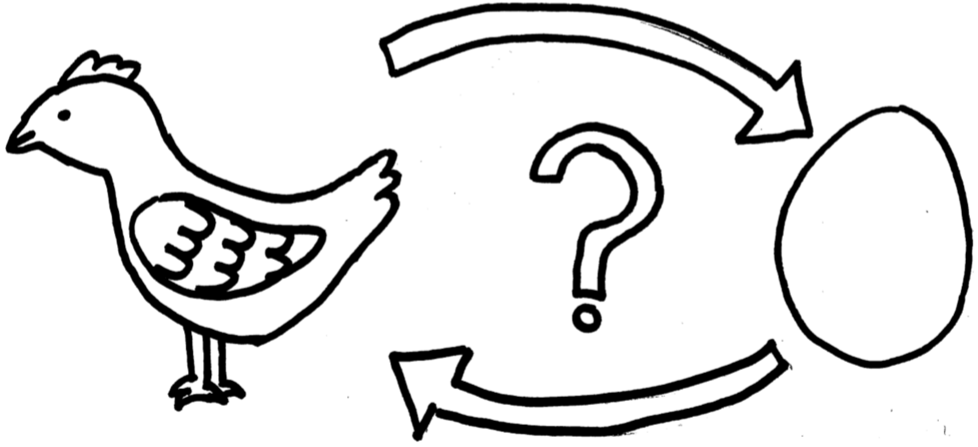
Biais de simultanéité

Vincent Arel-Bundock

Biais de simultanéité

- Biais de simultanéité
- Causalité inversée
- Causalité bidirectionnelle

Certains disent "endogénéité", mais je n'aime pas ce terme.



Analyse graphique

Analyse graphique

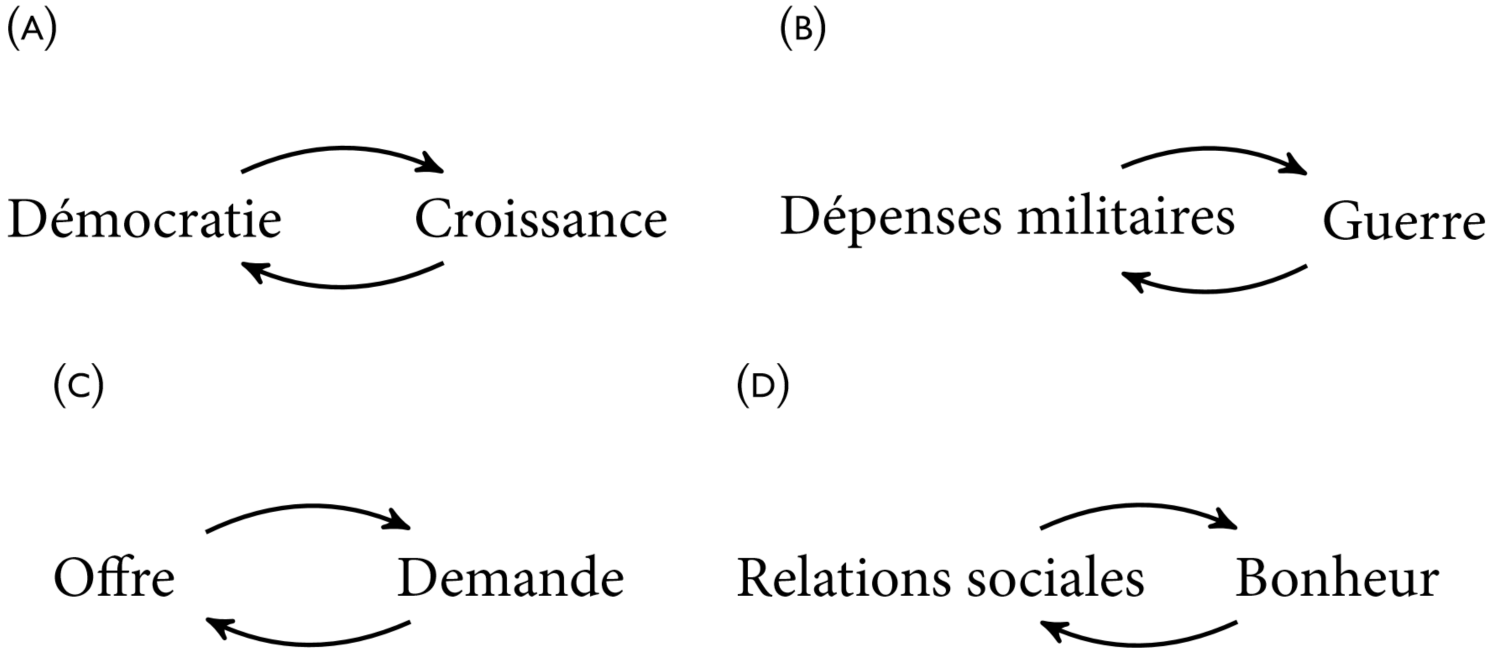
Objectif:

- Estimer l'effet causal de X sur Y dans un modèle avec flèches bidirectionnelles

Problème:

- Un GOA est orienté et acyclique.
- Pas de circuits.
- 2 conditions de l'identification causale ne s'appliquent pas ici!
    - Pas contrôler pour les descendants
    - Pas de chemin ouvert par la porte arrière

Analyse graphique



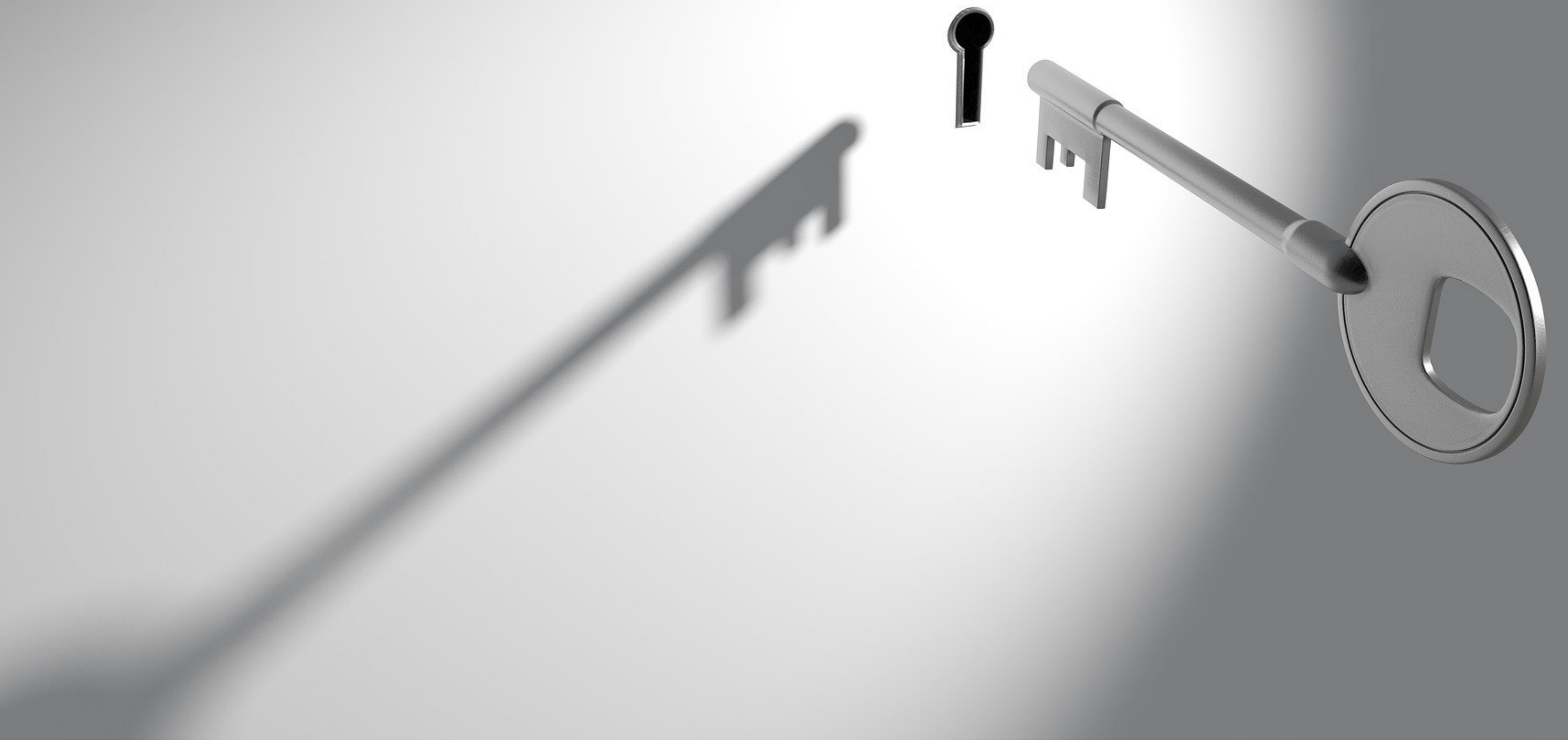
Avant de jeter les GOA à la poubelle...

Est-ce qu'il y a vraiment simultanéité?

$$\begin{tikzcd}
X_{t=1} \arrow[dr] & X_{t=2} \arrow[dr] & X_{t=3} \\
Y_{t=1} \arrow[ur] & Y_{t=2} \arrow[ur] & Y_{t=3}
\end{tikzcd}$$

Quand un mécanisme causal est défini avec suffisamment de granularité temporelle, il est possible de le représenter à l'aide d'un GOA valide (orienté et sans cycle).

Analyse algébrique



Analyse algébrique

On veut estimer l'effet causal de X sur Y, mais il y a causalité bidirectionnelle.

Cette simultanéité peut être représentée par un système de deux équations :

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{1}$$
$$X = \alpha_0 + \alpha_1 Y + \nu \tag{2}$$

Peut-on estimer seulement le 1er modèle?

Analyse algébrique

On veut estimer l'effet causal de X sur Y, mais il y a causalité bidirectionnelle.

Cette simultanéité peut être représentée par un système de deux équations :

$$\begin{align} Y &= \beta_0 + \beta_1 X + \varepsilon \tag{1} \\ X &= \alpha_0 + \alpha_1 Y + \nu \tag{2} \end{align}$$

Peut-on estimer seulement le 1er modèle?

L'estimé du coefficient de régression par les moindres carrés est non biaisé lorsque la variable explicative est indépendante du terme d'erreur.

$$E[\hat{\beta}_1] = \beta_1 \textbf{ si } X \perp \varepsilon$$

En présence de simultanéité, cette condition n'est pas satisfaite.

Analyse algébrique

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$
$$X = \alpha_0 + \alpha_1 Y + \nu \quad (2)$$

On substitue l'équation 2 dans l'équation 1, et on substitue à nouveau l'équation 1 dans le résultat:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + \varepsilon \\ &= \beta_0 + \beta_1(\alpha_0 + \alpha_1 Y + \nu) + \varepsilon \\ &= \beta_0 + \beta_1(\alpha_0 + \alpha_1(\beta_0 + \beta_1 X + \varepsilon) + \nu) + \varepsilon \end{aligned}$$

L'estimé de β_1 par les moindres carrés risque d'être biaisé, car:

$$X \not\perp \varepsilon$$

Analyse algébrique

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{1}$$
$$X = \alpha_0 + \alpha_1 Y + \nu \tag{2}$$

Il est difficile de quantifier la force du biais de simultanéité, ou même d'anticiper sa direction.

Si on estime le modèle 1 en ignorant le modèle 2, il est raisonnable de croire que la taille du biais sur l'estimé de β_1 est proportionnelle à la force de la relation causale inverse, soit α_1.

Pour des modèles plus complexes avec plus de variables explicatives, cela n'est pas nécessairement le cas.

Merci!

Slide title