



VARIABLES INSTRUMENTALES

Session Technique

Session IV

Emanuela Galasso
Banque Mondiale

Cette présentation a été développée par Christel Vermeersch et modifiée par Emanuela Galasso pour cet atelier.

Un exemple pour commencer...

- Supposons que nous souhaitions évaluer un programme volontaire de formation professionnelle
 - Toute personne sans emploi est éligible
 - Certaines personnes choisissent de s'inscrire (Participants)
 - D'autres choisissent de ne pas s'inscrire (Non participants)
- Quelques moyens simples (mais pas très bons) pour évaluer le programme:
 - Comparer la situation avant et après du groupe de traitement
 - Comparer la situation des participants et des non participants après l'intervention
 - Comparer la situation des participants et des non participants avant et après.



Programme volontaire de formation professionnelle

Supposons que nous décidons de comparer les résultats des participants et des non-participants:

- Soit le modèle simple suivant:

$$y = \alpha + \beta_1 P + \beta_2 X + \varepsilon$$

$$P = \begin{cases} 1 & \text{Si la personne participe à la formation} \\ 0 & \text{Si la personne ne participe pas à la formation} \end{cases}$$

X = Variables de contrôle (exogène & observable)

- Pourquoi ce modèle ne fonctionne pas
2 problèmes:
 - Les variables que nous omettons (pour diverses raisons) mais qui sont importantes
 - La décision de prendre part à la formation est endogène.



Problème #2: La décision de participer est endogène

- Le modèle correct:

$$y = \gamma_0 + \gamma_1 X + \gamma_2 P + \eta$$

avec

$$P = \pi_0 + \pi_1 X + \pi_2 M_2 + \xi$$

M_2 = Vector latent /caracteristiques manquantes
(i.e. nous ne connaissons pas entièrement la raison pour laquelle les gens ont décidé de participer)

- Puisque nous n'observons pas M_2 , nous ne pouvons qu'estimer un modèle simplifié:

$$y = \beta_0 + \beta_1 X + \beta_2 P + \varepsilon$$

- $\beta_{2, OLS}$ est-il un estimateur non-biaisé de γ_2 ?



Problème #2: la décision de participer est endogène

● Supposons: $y = \beta_0 + \beta_1 x + \beta_2 P + \varepsilon$

● Mais le modèle correct est: $y = \gamma_0 + \gamma_1 x + \gamma_2 P + \eta$

Avec $P = \pi_0 + \pi_1 x + \pi_2 M_2 + \xi$

● Est-ce $\beta_{2, OLS}$ un estimateur non-biaisé de γ_2 ?

$$\begin{aligned} \text{Corr}(\varepsilon, P) &= \text{corr}(\varepsilon, \pi_0 + \pi_1 x + \pi_2 M_2 + \xi) \\ &= \pi_1 \text{corr}(\varepsilon, x) + \pi_2 \text{corr}(\varepsilon, M_2) \\ &= \pi_2 \text{corr}(\varepsilon, M_2) \end{aligned}$$

● S'il existe une corrélation entre les variables manquantes déterminant la participation (par exemple le talent) et les résultats non expliqués par les caractéristiques observées, l'estimateur OLS sera alors biaisé.



Que faire pour résoudre ce problème?

● Nous estimons: $y = \beta_0 + \beta_1 x + \beta_2 P + \varepsilon$

● Donc le problème est la corrélation entre P et ε

● Et si nous remplaçons P avec "quelque chose d'autre", que nous allons nommer Z :

- Z doit être similaire à P
- Mais n'est pas corrélé à ε



Revenons au programme de formation professionnel

- P = participation
- ϵ = partie des résultats non-expliqués par la participation au programme ou par les caractéristiques observées
- Je cherche une variable Z c'est-à-dire:
 - (1) Etroitement liée à la participation P
 - (2) Mais n'ayant pas un impact direct sur les résultats Y des participants, *autre que son effet sur la participation.*
- Donc cette variable doit provenir de l'extérieur.



Une variable externe pour le programme de formation professionnelle

- Supposons qu'un travailleur social visite des personnes au chômage afin de les encourager à participer:
 - Elle ne visite que 50% des personnes au cours des tournées, et
 - Elle a par hasard choisit les personnes qu'elle visitera
- Si elle y parvient, davantage personnes vont s'inscrire
Il y aura donc une corrélation entre visite et inscription
- Mais, la visite n'a pas d'effet directe sur les résultats (par exemple: revenu) en dehors de l'effet qu'il a grâce à l'inscription au programme de formation.



Une variable externe ou instrumentale

- Définir une nouvelle variable Z

$$Z = \begin{cases} 1 & \text{Si une personne a été sélectionnée de manière aléatoire} \\ & \text{pour recevoir des encouragements d'un travailleur social} \\ 0 & \text{Si une personne n'a pas été sélectionnée de manière} \\ & \text{aléatoire pour recevoir des encouragement d'un} \\ & \text{travailleur social} \end{cases}$$

- $\text{Corr}(Z, P) > 0$

Les gens qui ont reçu des encouragements ont beaucoup plus de chance de participer que ceux qui n'en ont pas reçus

- $\text{Corr}(Z, \varepsilon) = 0$

Aucune corrélation entre recevoir une visite et bénéficier du programme en dehors de l'impact de la visite sur le participant

- Z est appelé une **variable instrumentale**



Méthode des moindres carrés à deux étapes (2SLS)

Rappelez-vous du modèle original avec la variable endogène P :

$$y = \beta_0 + \beta_1 x + \beta_2 P + \varepsilon$$

Etape 1

Regresser la variable endogène P sur la/les variable/s instrumentale/s Z et les autres variables exogènes

$$P = \delta_0 + \delta_1 x + \delta_2 Z + \tau$$

- Calculer la valeur prédite de P pour chaque observation: P
- Etant donné que Z et x ne sont pas corrélés avec ε , P ne le sera non plus.
- Vous aurez besoin d'une variable instrumentale pour chaque potentiel regresser endogène.



Méthodes des moindres carrés à deux étapes (2SLS)

Étape 2

Regresser y sur la valeur prédite de P et les autres variables exogènes

$$y = \beta_0 + \beta_1 x + \beta_2 P + \varepsilon$$

- **Remarque:** les écarts-types au niveau de la seconde étape OLS doivent être corrigés car P n'est pas une variable explicative fixe.
- **En Pratique:** utiliser la commande `ivreg` de STATA qui effectue les deux étapes en une seule fois et rapporte les écarts-types corrigés.
- **Intuition:** en utilisant Z pour P , nous éliminons P de sa corrélation avec η
- On peut démontrer que (dans certaines conditions) $\beta_{2,IV}$ résulte à un estimateur consistant de γ_2 (théorie des grands nombres)

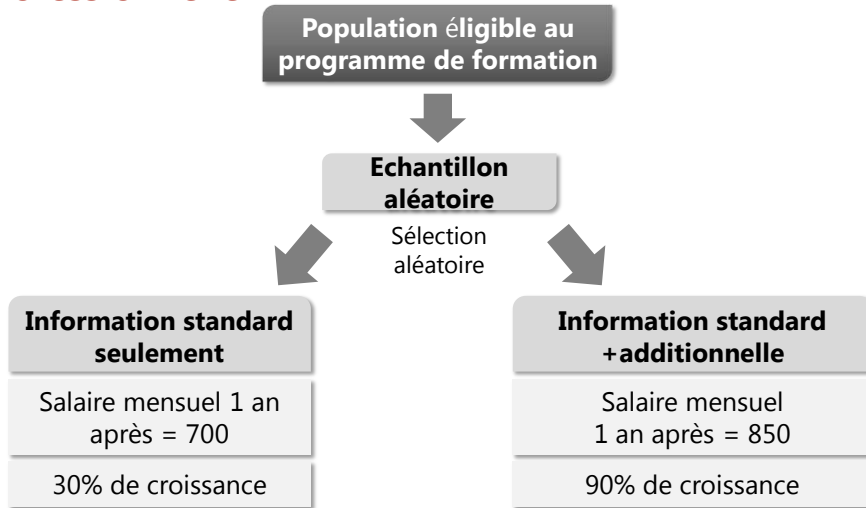


Où trouver des variables instrumentales?

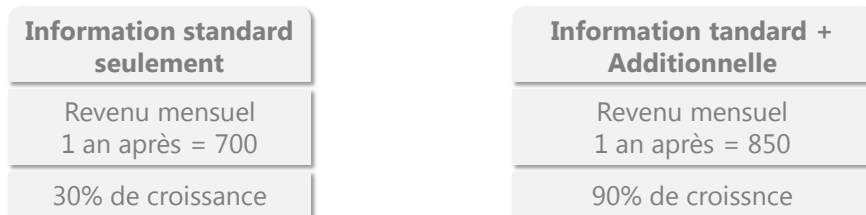
- En rechercher une... **difficile!**
- Créer une variable instrumentale à l'aide d'une campagne d'information ou de promotion:
 - Si chacun est éligible pour participer au traitement
 - Mais certains ont plus d'expérience que d'autres (celui qui a plus d'information a plus de chance de participer)
 - Mise à disposition "d'informations supplémentaires" de manière aléatoire



Exemple 1: programme volontaire de formation professionnelle



Question: quel est l'impact du programme de formation professionnelle sur le revenu?



Question: quel est l'impact du programme de formation professionnelle sur le revenu?

Différence entre le groupe "bien informé" et "moins bien informé":

.....

Corrigé pour le taux différentiel :

.....

Pratiquement
Impact =

Retour à la formule d'estimation

Etape 1

- Régresser la participation à la formation sur une variable indiquant si une personne a reçu de l'information additionnelle (modèle linéaire)
- Calculer la valeur prédite de participation

Etape 2

Régresser les salaires sur la valeur prédite du modèle de participation



Exemple 2: Autonomie des écoles au Népal

Objectif

Evaluer:

- A. La gestion autonome des écoles par les communautés
- B. Les résultats des écoles

Données

- Vous pouvez inclure 1000 écoles dans l'évaluation
- Chaque communauté décide de participer librement ou de ne pas participer
- Les résultats des écoles sont établis par les ONG
- Chaque communauté possède exactement une école

Tâche

Concevoir le programme de telle manière qu'il puisse être évalué –proposer une méthode d'évaluation.



Autonomie des écoles au NEPAL

		Intervention B: résultat des écoles établis par les ONG.		
		Oui	Non	Total
Variable instrumentale pour l'intervention A: Visite de l'ONG pour s'informer des procédures de transfert de la gestion des écoles par la communauté	Oui	300	300	600
	Non	200	200	400
	Total	500	500	1000



Rappel et avertissement ...

- $corr(Z, \varepsilon) = 0$
 - Si $corr(Z, \varepsilon) \neq 0$, "Mauvais instrument"
 - "Il est difficile de trouver un bon instrument"
 - Mais vous pouvez vous-même mettre un en oeuvre grâce à un design de promotion ou encouragement aléatoire.
- $corr(Z, P) \neq 0$
 - "Instruments faibles": la corrélation entre Z et P doit être suffisamment forte.
 - Sinon le biais reste important pour les échantillons de grande taille.



Retour au *ATE* et au *TOT*

- Parfois, les unités admissibles sont sélectionnées au hasard dans le groupe de traitement, sont soumises au traitement, mais toutes ne l'acceptent pas.
- Calculer l'effet moyen du traitement (ATE)
Différence absolue dans les résultats moyens entre le groupe auquel vous offrez un traitement, et le groupe qui n'a pas bénéficié du traitement
- Calculer l'effet du traitement sur les traités (TOT)
Utilisez l'offre randomisée comme une variable instrumentale (Z) pour la participation des gens au traitement (P)



Remarque: *VI* est un effet "local"

- Une variable instrumentale identifie les gains moyens des personnes amenées à changer leur choix par un changement de l'instrument (dénommé « non-complier»)
- ... Nous pouvons toutefois identifier qui sont ces gens ("effet du traitement moyen local" ou LATE)
- ... les différents instruments identifieront les paramètres et répondra aux différentes questions
- Il faut faire attention au moment d'extrapoler les résultats à l'ensemble de la population.





Je vous remercie



Q&R