Séance 4 - Données de panel et effets fixes

Support étudiant

Ibrahim Goni ABDOULKADIRI - Magistère 3

Question directrice
Identifier la variation within et choisir le niveau de regroupement.

1 Objectifs et déroulement

— Décrire un panel et distinguer variation within et between.

— Dériver et reproduire la transformation within.

— Interpréter un coefficient à effets fixes comme comparaison intra individuelle.

— Comprendre l'hypothèse plus forte du modèle à effets aléatoires.

— Choisir des écarts types groupés au niveau de l'unité d'observation répétée.

Temps	Travail
0 à 20 min	Structure du panel et décomposition de la variation
20 à 50 min	MCO groupés et effets fixes
50 à 80 min	Transformation within manuelle et effets temporels
80 à 105 min	Effets aléatoires et test de Hausman
105 à 120 min	Inférence groupée et synthèse

2 Repères avant l'ordinateur

La séance exploite dix années d'observations sur 532 individus. Elle montre pourquoi les MCO groupés mélangent variation entre individus et variation au cours du temps, puis comment l'estimateur à effets fixes isole les écarts à la moyenne individuelle. Le choix des écarts types est relié à la dépendance des erreurs au sein de chaque individu.

Règle de lecture
Un coefficient décrit d'abord une comparaison statistique. Une interprétation causale exige une hypothèse d'identification formulée séparément.

3 Étude de cas de la séance

Article et question. Ziliak (1997) : offre de travail en panel. L'application étudie la réponse des heures annuelles au salaire en suivant les mêmes personnes. Les effets fixes retirent l'hétérogénéité individuelle permanente.

Échantillon. Panel équilibré de 532 individus suivis de 1979 à 1988. Unité : individu-année.

Ce que nous allons faire. Nous commencerons par définir l'unité d'observation et les variables. Nous écrirons ensuite la spécification, identifierons la variation utilisée, exécuterons le do-file, puis rédigerons une interprétation qui sépare taille de l'estimation, incertitude et portée causale.

Références de l'étude. (Ziliak 1997 ; Mundlak 1978 ; Wooldridge 2010).

4 Données et variables

labor_supply.dta est un panel équilibré de 532 individus observés de 1979 à 1988, soit 5 320 observations. L'équation relie le logarithme des heures annuelles au logarithme du salaire horaire, à l'âge, au nombre d'enfants et à l'état de santé.

Variable	Définition et unité
id	Identifiant individuel
year	Année d'observation
lnhr	Log des heures annuelles travaillées
lnwg	Log du salaire horaire
age	Âge en années
agesq	Carré de l'âge
kids	Nombre d'enfants dans le ménage
disab	Indicatrice de mauvaise santé

5 Questions de préparation

1. Quelle est l'unité d'observation et quelle population le fichier représente-t-il ?
2. Quelle comparaison économique le coefficient principal doit-il résumer ?
3. Quelle hypothèse serait nécessaire pour employer le mot « effet » ?
4. Quelle sortie ou quel graphique permettrait de détecter une interprétation fragile ?

6 Exercice 1 Décrire la variation du panel

Déclarez le panel avec xtset id year et vérifiez son équilibre.

1. Combien d'individus et de périodes observez-vous ?
2. Comparez la variation within et between de lnhr, lnwg, kids et disab avec xtsum.
3. Quelles variables risquent d'être faiblement identifiées par les effets fixes ?
4. Tracez lnhr et lnwg pour quatre individus choisis de manière reproductible.

7 Exercice 2 MCO groupés et effets fixes

Estimez le modèle groupé puis le modèle à effets fixes avec effets temporels et écarts types groupés par individu.

1. Comparez le coefficient de lnwg. Quelles sources de variation alimentent chaque estimateur ?
2. Interprétez le coefficient FE en pourcentage.
3. Pourquoi les effets temporels sont-ils utiles dans cette application ?

4. Quelles caractéristiques individuelles constantes sont absorbées sans être observées ?

8 Exercice 3 Transformation within

Reproduisez manuellement l'estimateur FE en retirant à chaque variable sa moyenne individuelle.

1. Construisez les variables centrées pour lnhr, lnwg, age, agesq, kids et disab.
2. Estimez l'équation centrée sans constante, d'abord sans effets temporels.
3. Comparez à xtreg fe estimé sans effets temporels. Expliquez toute différence résiduelle.
4. Pourquoi une variable constante dans le temps disparaît-elle ?

9 Exercice 4 Effets aléatoires et choix du modèle

Estimez le modèle à effets aléatoires puis comparez FE et RE.

1. Quelle hypothèse porte sur la corrélation entre effet individuel et régresseurs dans le modèle RE ?
2. Exécutez Hausman avec des matrices compatibles puis discutez ses limites après une inférence robuste.
3. Le test peut-il remplacer un argument économique sur l'hétérogénéité individuelle ?
4. Proposez la spécification principale à publier et justifiez le niveau de regroupement.

10 Production attendue

— Un tableau pooled FE RE avec la source de variation annotée.

— Une vérification numérique de la transformation within.

— Une recommandation de spécification en cinq lignes.

Billet de sortie
En cinq lignes, expliquer à quel contrefactuel descriptif correspond le coefficient FE de lnwg et citer une source d'endogénéité qui demeure après effets fixes individuels et temporels.

Espace pour la synthèse

Références

MUNDLAK, Yair (1978). "On the Pooling of Time Series and Cross Section Data". In : *Econometrica* 46.1, p. 69-85. DOI : 10.2307/1913646.

WOOLDRIDGE, Jeffrey M. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2e éd. Cambridge, MA : MIT Press.

ZILIAK, James P. (1997). "Efficient Estimation with Panel Data When Instruments Are Predetermined : An Empirical Comparison of Moment-Condition Estimators". In : *Journal of Business & Economic Statistics* 15.4, p. 419-431. DOI : 10.1080/07350015.1997.10524720.