# Séance 7 - Instrument shift-share en panel : aide alimentaire et conflits

Support étudiant
Pierre BEAUCORAL - Magistère 3

**Question directrice**
Identifier la variation exploitée par un instrument construit comme le produit d'un choc commun et d'une exposition propre à chaque unité, puis éprouver la validité de l'inférence.

## 1 Objectifs et déroulement

- Construire un instrument shift-share à un seul choc (exposition × choc commun) et dire ce que les effets fixes absorbent.
- Répliquer une estimation publiée avec des effets fixes de grande dimension.
- Montrer comment des tendances différentielles menacent l'identification.
- Mener un exercice placebo pour juger la fiabilité des tests usuels.
- Choisir le niveau de regroupement des écarts types lorsque la variation vient d'un choc commun.

| Temps | Travail |
|---|---|
| 0 à 20 min | Construction de l'instrument et variation résiduelle |
| 20 à 50 min | Réplication du tableau 2 de Nunn et Qian |
| 50 à 80 min | Tendances différentielles |
| 80 à 110 min | Placebo et niveau de regroupement |
| 110 à 120 min | Synthèse écrite individuelle |

## 2 Repères avant l'ordinateur

Nunn et Qian (2014) demandent si l'aide alimentaire américaine prolonge les guerres civiles. L'aide reçue est endogène : elle répond aux crises. Les auteurs l'instrumentent par la production américaine de blé de l'année précédente (les excédents sont écoulés sous forme d'aide), interagie avec la fréquence à laquelle le pays reçoit de l'aide. Les effets fixes pays et région-année absorbent chacun des deux termes pris séparément ; seule leur interaction identifie l'effet. C'est un instrument shift-share (« de Bartik ») à un seul choc : faute de nombreux chocs, l'identification repose sur l'exogénéité de l'exposition (Goldsmith-Pinkham et al. 2020), et non sur celle des chocs (Borusyak et al. 2022). Christian et Barrett (2024) montrent, notamment par des simulations où la production américaine est remplacée par une marche aléatoire, que ce type d'instrument peut produire des résultats fallacieux lorsque le choc commun est persistant.

**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.** Nunn et Qian (2014) : une hausse de l'aide alimentaire américaine augmente-t-elle l'incidence des conflits dans les pays bénéficiaires ?

**Échantillon.** 125 pays non membres de l'OCDE, 1971–2006 ; 4 089 observations pays-année (échantillon du tableau 2).

**Ce que nous allons faire.** Répliquer la colonne 1 du tableau 2 (MCO, première étape, forme réduite, 2SLS), introduire des tendances propres aux pays souvent aidés, puis remplacer la production américaine par une marche aléatoire, dans une version simplifiée de l'exercice de Christian et Barrett (2024).

**Références de l'étude.** (Nunn et Qian 2014 ; Christian et Barrett 2024 ; Goldsmith-Pinkham et al. 2020 ; Borusyak et al. 2022 ; Adão et al. 2019).

## 4 Données et variables

`nunn_qian_td7.dta` est extrait du paquet de réplication des auteurs (AER) : il reprend l'échantillon du tableau 2 et les variables utiles à la séance.

| Variable | Définition et unité |
|---|---|
| `risocode`, `year` | Code ISO du pays, année |
| `wb_region` | Région de la Banque mondiale |
| `any_war` | Incidence d'un conflit armé, tout type (0/1) |
| `intra_state` | Incidence d'une guerre civile (0/1) |
| `wheat_aid` | Aide en blé reçue des États-Unis (milliers de tonnes) |
| `l_US_wheat_production` | Production de blé des États-Unis en $t-1$ (milliers de tonnes) |
| `fadum_avg` | Part des années 1971–2006 où le pays reçoit de l'aide américaine |
| `instrument` | `l_US_wheat_production` × `fadum_avg` |
| `oil_fadum_avg`, ... | Contrôles américains interagis avec `fadum_avg` (colonne 2) |

## Logiciels : Stata ou R

**Stata** (version 17 ou plus) : `ssc install ivreg2`, `ranktest`, `reghdfe`, `ftools`, `require`, `ivreghdfe`. Do-file : `TD7_etudiant.do`.
**R** (4.3 ou plus) : `install.packages(c("fixest", "haven", "dplyr", "tidyr", "ggplot2", "patchwork", "here"))`. Script : `TD7_etudiant.R`.

| Stata | R (`fixest`) |
|---|---|
| `reghdfe y x, absorb(cid i.year#i.region) cluster(cid)` | `feols(y ~ x \| risocode + wb_region^year, cluster = ~risocode)` |
| `ivreghdfe y (x = z), absorb(...) cluster(cid)` | `feols(y ~ 1 \| ... \| x ~ z, cluster = ~risocode)` |
| `cluster(year)` | `summary(m, cluster = ~year)` |
| `set seed` ; `runiform()` | `set.seed()` ; `runif()` |
| `postfile` / `post` | liste pré-allouée puis `bind_rows()` |

## 5 Questions de préparation

1. Quelle est l'unité d'observation et quelle population le fichier représente-t-il ?
2. Pourquoi l'aide reçue serait-elle corrélée au terme d'erreur de l'équation de conflit ? Dans quel sens ?
3. Pourquoi interagir la production américaine avec `fadum_avg` plutôt que l'utiliser seule ?
4. Quelle hypothèse doit vérifier `fadum_avg` pour que l'instrument soit valide ?

# 6 Exercice 1 Décrire le panel et construire l'instrument

1. Déclarez le panel (`encode risocode, gen(cid)`, puis `xtset cid year`). Est-il équilibré ?
2. Vérifiez que `instrument` est bien le produit de `l_US_wheat_production` et `fadum_avg`.
3. Les effets fixes pays et région-année absorbent quels termes de l'instrument ? Quelle variation reste-t-il ?
4. Tracez la production américaine retardée et la part des pays en conflit, séparément pour les pays au-dessus et au-dessous de la médiane de `fadum_avg`. Que suggère le graphique ?

# 7 Exercice 2 Répliquer le tableau 2, colonnes 1 et 2

Estimez avec `ivreg2` et les indicatrices des auteurs, puis avec `reghdfe` et `ivreghdfe`, effets absorbés `absorb(cid i.year#i.region)`, écarts types groupés par pays.

1. Retrouvez la première étape, la forme réduite et le coefficient 2SLS publiés. Vérifiez que le rapport forme réduite sur première étape redonne le coefficient 2SLS.
2. Les deux commandes donnent les mêmes coefficients. Expliquez l'écart entre leurs écarts types et leurs statistiques F.
3. Interprétez le coefficient 2SLS pour une hausse de 10 % de l'aide à la moyenne de l'échantillon.
4. Commentez le F de Kleibergen-Paap et le test d'Anderson-Rubin.

# 8 Exercice 3 Tendances différentielles

1. Expliquez pourquoi une tendance de long terme dans la production américaine, combinée à une trajectoire de conflit propre aux pays souvent aidés, pourrait produire un coefficient positif sans effet causal.
2. Ajoutez `fadum_avg` $\times$ (année $-$ 1971) comme contrôle. Que deviennent le coefficient et le F ?
3. Ajoutez en plus le terme quadratique. Même question.
4. Concluez sur la variation qui identifie l'effet dans l'article.

# 9 Exercice 4 Placebo et niveau de regroupement

Remplacez la production américaine par une série annuelle simulée, sans rapport avec le conflit : (a) une marche aléatoire, (b) un bruit blanc. Fixez la graine une seule fois en tête du do-file. Pour chaque tirage, construisez l'instrument fictif, estimez la première étape, la forme réduite et le 2SLS, et conservez les statistiques $t$ (`postfile`).

1. Sous l'hypothèse nulle, quelle part de rejets à 5 % attend-on ?
2. Calculez les parts de rejets avec des écarts types groupés par pays. Commentez.
3. Recommencez avec des écarts types groupés par année. Qu'est-ce qui change pour le bruit blanc ? Pour la marche aléatoire ?
4. Réestimez la colonne 1 avec des écarts types groupés par année. Quelle conclusion tirer sur la fiabilité de l'inférence dans l'article ?

## 10 Production attendue

- Un tableau de réplication : MCO, première étape, forme réduite, 2SLS, avec la valeur publiée en regard.
- Un tableau de sensibilité aux tendances différentielles.
- Un tableau des parts de rejet du placebo et un paragraphe d'interprétation de huit lignes.

**Billet de sortie**
En dix lignes, dire quelle variation identifie l'effet de l'aide, quelle hypothèse doit vérifier l'exposition `fadum_avg`, et pourquoi un test significatif ne suffit pas lorsque le choc commun est persistant.

## Espace pour la synthèse

## Références

Adão, Rodrigo, Michal Kolesár et Eduardo Morales (2019). "Shift-Share Designs : Theory and Inference". In : *Quarterly Journal of Economics* 134.4, p. 1949-2010. doi : 10.1093/qje/qjz025.

Borusyak, Kirill, Peter Hull et Xavier Jaravel (2022). "Quasi-Experimental Shift-Share Research Designs". In : *Review of Economic Studies* 89.1, p. 181-213. doi : 10.1093/restud/rdab030.

Christian, Paul et Christopher B. Barrett (2024). "Spurious Regressions and Panel IV Estimation : Revisiting the Causes of Conflict". In : *The Economic Journal* 134.659, p. 1069-1099. doi : 10.1093/ej/uead091.

Goldsmith-Pinkham, Paul, Isaac Sorkin et Henry Swift (2020). "Bartik Instruments : What, When, Why, and How". In : *American Economic Review* 110.8, p. 2586-2624. doi : 10.1257/aer.20181047.

Nunn, Nathan et Nancy Qian (2014). "US Food Aid and Civil Conflict". In : *American Economic Review* 104.6, p. 1630-1666. doi : 10.1257/aer.104.6.1630.