

Méthode des Doubles Différences (DiD)

Pratiques de la Recherche en Économie

Florentine Oliveira

2025-04-11

- Les méthodes abordées précédemment (régression OLS simple, variable instrumentale) concernaient des données en **coupe instantanée** (*cross-section*) ou des échantillons répétés dans le temps (*repeated cross-section*) sans suivi individuel
- Cette séance: données de panel
 - Effets fixes individus (très rapidement)
 - Méthode des doubles différences

Données de Pannel et Effet Fixe Individu

Dans un jeu de données de panel, on observe:

- de multiples observations t (année, trimestre, etc)
- pour de multiples individus (i)

Le modèle de régression devient alors, pour tout individu i et période t :

$$y_{it} = \alpha + \beta D_{it} + \varepsilon_{it}$$

Effet fixe

En incluant un **effet fixe** dans notre modèle de régression, on contrôle par tous les facteurs individuels **constants au cours du temps**, à la fois observables et non-observables.

⇒ on n'utilise plus la variation entre les unités d'observations, les i (*between*), mais la variation intra-individuelle (variation d'un même individu dans le temps, *within*).

Le modèle s'écrit:

$$y_{it} = \alpha + \beta D_{it} + \alpha_i + \varepsilon_{it}$$

Données de Pannel et Effet Fixe Individu

Exemple: quel est l'effet de la taille de la ville sur le salaire (*urban wage premium*)?

Données de Pannel et Effet Fixe Individu

Exemple: quel est l'effet de la taille de la ville sur le salaire (*urban wage premium*)?

Pourquoi ne peut-on pas déduire l'effet de la taille de la ville en comparant le salaire des individus qui vivent en ville au salaire des individus qui vivent dans des zones peu densément peuplées?

Données de Pannel et Effet Fixe Individu

Exemple: quel est l'effet de la taille de la ville sur le salaire (*urban wage premium*)?

Pourquoi ne peut-on pas déduire l'effet de la taille de la ville en comparant le salaire des individus qui vivent en ville au salaire des individus qui vivent dans des zones peu densément peuplées?

Endogeneous sorting: le lieu d'habitation n'est pas aléatoire (eg: les individus les plus diplômés choisissent souvent de vivre dans de grandes villes!)

Données de Pannel et Effet Fixe Individu

Exemple: quel est l'effet de la taille de la ville sur le salaire (*urban wage premium*)?

Pourquoi ne peut-on pas déduire l'effet de la taille de la ville en comparant le salaire des individus qui vivent en ville au salaire des individus qui vivent dans des zones peu densément peuplées?

Endogeneous sorting: le lieu d'habitation n'est pas aléatoire (eg: les individus les plus diplômés choisissent souvent de vivre dans de grandes villes!)

Solution: Effet fixe individu! Cela veut dire que l'on exploite pour l'identification la variation au sein d'un individu, autrement dit uniquement sur les individus qui sont amenés à déménager (*movers*).

Données de Pannel et Effet Fixe Individu

Exemple: quel est l'effet de la taille de la ville sur le salaire (*urban wage premium*)?

Pourquoi ne peut-on pas déduire l'effet de la taille de la ville en comparant le salaire des individus qui vivent en ville au salaire des individus qui vivent dans des zones peu densément peuplées?

Endogeneous sorting: le lieu d'habitation n'est pas aléatoire (eg: les individus les plus diplômés choisissent souvent de vivre dans de grandes villes!)

Solution: Effet fixe individu! Cela veut dire que l'on exploite pour l'identification la variation au sein d'un individu, autrement dit uniquement sur les individus qui sont amenés à déménager (*movers*).

Problème: les effets fixes individu ne nous permettent pas de contrôler pour les caractéristiques des individus qui varient dans le temps...

Méthode des Doubles Différences

- Méthode des doubles-différences s'appuie sur des **expériences naturelles**, i.e. une politique ou un changement institutionnel qui induit un traitement pour un groupe donné à un moment précis
- Contrairement aux "vraies expériences" où les groupes de contrôle et de traitement sont constitués par assignation aléatoire, les groupes sont déterminées par le changement de politique particulier.
- Les données de panel et la méthode des DiD permettent de tenir compte de cette sélection sur des caractéristiques non-observables, du moment que ces **caractéristiques sont constantes dans le temps**

1. Intuition

Si, avant la mise en place du traitement, le groupe d'individus traités et contrôles diffèrent, cela est probablement dû à des éléments non observés (biais de sélection)

⇒ on utilise cette différence pré-traitement comme référence pour "corriger" le biais initial

1. Intuition

Si, avant la mise en place du traitement, le groupe d'individus traités et contrôles diffèrent, cela est probablement dû à des éléments non observés (biais de sélection)

⇒ on utilise cette différence pré-traitement comme référence pour "corriger" le biais initial

On compare donc la différence entre les deux groupes (première différence), avant et après traitement (deuxième différence).

- en soustrayant la différence initiale, on obtient l'effet 'net' du traitement

Cela suppose que l'**hypothèse de tendance parallèle** est valable: en l'absence de politique, l'écart entre les deux groupes serait resté constant dans le temps.

Estimation

Supposons qu'il y a:

- t périodes $\in \{1, T\}$
- un groupe d'individus $i \in \llbracket 1; n \rrbracket$ traités $D_i = 1$, et non traités $D_i = 0$
- les individus sont traités en $t = 1$

Le modèle s'écrit:

$$Y_{it} = \alpha + \beta_{DiD}(D_i \times Post_t) + \delta D_i + \eta Post_t + \varepsilon_{it}$$

Généralisation d'un modèle à plusieurs périodes et traitement (et traitement pouvant être staggered - *event-study*):

$$Y_{it} = \alpha + \sum_{t=1}^T \beta_t^{TWFE} D_{it} + \gamma_i + \eta_t + \varepsilon_{it}$$

NB: pour la généralisation: TWFE et discussions récentes autour des estimateurs DiD.

Hypothèse d'identification

Parallel trends

- En l'absence de traitement, l'outcome moyen des individus traités aurait évolué de la même manière que celui des individus non traités
- Formellement, $\underbrace{\mathbb{E}[Y_{i2}(0) - Y_{i1}(0)|D_i = 1]}_{\text{Changement de } Y(0) \text{ pour les traités}} = \underbrace{\mathbb{E}[Y_{i2}(0) - Y_{i1}(0)|D_i = 0]}_{\text{Changement de } Y(0) \text{ pour les contrôles}}$
- Ou encore, $\underbrace{\mathbb{E}[Y_{i2}(0)|D_i = 1] - \mathbb{E}[Y_{i2}(0)|D_i = 0]}_{\text{Biais de sélection en période 2}} = \underbrace{\mathbb{E}[Y_{i1}(0)|D_i = 1] - \mathbb{E}[Y_{i1}(0)|D_i = 0]}_{\text{Biais de sélection en période 1}}$

Absence d'anticipation

- Le traitement n'a pas d'effet causal avant qu'il ne soit implanté
- Formellement, $Y_{i1}(1) = Y_{i1}(0)$

Application: Card and Krueger (1994)

Question de Recherche: quel est l'effet causal d'une augmentation du salaire minimum sur le taux d'emploi?

Application: Card and Krueger (1994)

Question de Recherche: quel est l'effet causal d'une augmentation du salaire minimum sur le taux d'emploi?

Question: pourquoi ne peut-on pas simplement comparer le taux de chômage de deux états ayant des niveaux de salaire minimum différents?

Application: Card and Krueger (1994)

Question de Recherche: quel est l'effet causal d'une augmentation du salaire minimum sur le taux d'emploi?

Question: pourquoi ne peut-on pas simplement comparer le taux de chômage de deux états ayant des niveaux de salaire minimum différents?

Biais de sélection/OVB:

Application: Card and Krueger (1994)

Question de Recherche: quel est l'effet causal d'une augmentation du salaire minimum sur le taux d'emploi?

Question: pourquoi ne peut-on pas simplement comparer le taux de chômage de deux états ayant des niveaux de salaire minimum différents?

Biais de sélection/OVB:

Angrist et Krueger (1994)

- expérience naturelle: augmentation du salaire minimum dans le New Jersey au 1er Avril 1992 de \ \$4,25 à \ \$5,05 par heure
- focus sur l'industrie des fast-foods
- comparaison avec l'état de Pennsylvanie

Application: Card and Krueger (1994)

1) Importez les données sur `R` et stockez dans `df`. Créer les variables:

- `fte_before` (équivalent temps-plein) = `empft + nmgrs + (0.5*emppt)`
- `fte_after` = `empft2 + nmgrs2 + (0.5*emppt2)`
- Ré-encoder `state` de pour que ce soit égal à Pennsylvanie si `state` = 0 et New Jersey si `state` = 1

2) Pourquoi s'intéresser à l'industrie des fast-food, et à quel effet s'attend-t-on?

3) Calculer le nombre moyen d'employés en équivalent temps-plein par store dans chaque état, avant et après la réforme, et l'évolution de cette moyenne. Interpréter.

4) Calculer la différence de changement d'heures entre New Jersey et Pennsylvanie. Interprétez.

5) Proposez un modèle économétrique qui estime l'effet du traitement.

Solution : Card and Krueger (1994)

1) Importez les données sur R et stockez dans `df`. Créer les variables:

- `fte_before` (équivalent temps-plein) = `empft + nmgrs + (0.5*emppt)`
- `fte_after` = `empft2 + nmgrs2 + (0.5*emppt2)`
- Ré-encoder `state` de pour que sa valeur soit "Pennsylvanie" si `state` = 0 et "New Jersey" si `state` = 1

```
df = readRDS("data/ck1994.rds") %>%  
  mutate(fte_before = empft + nmgrs + (0.5*emppt),  
         fte_after = empft2 + nmgrs2 + (0.5*emppt2),  
         state = ifelse(state == 0, "Pennsylvanie", "New Jersey"))
```

Solution : Card and Krueger (1994)

2) **Pourquoi s'intéresser à l'industrie des fast-food**, et à quel effet s'attend-t-on?

Solution : Card and Krueger (1994)

2) **Pourquoi s'intéresser à l'industrie des fast-food**, et à quel effet s'attend-t-on?

- Secteur dont une forte proportion de travailleurs est rémunérée au salaire minimum \implies terrain d'observation idéal pour étudier l'effet d'une augmentation du salaire minimum.

Solution : Card and Krueger (1994)

2) **Pourquoi s'intéresser à l'industrie des fast-food**, et à quel effet s'attend-t-on?

- Secteur dont une forte proportion de travailleurs est rémunérée au salaire minimum \implies terrain d'observation idéal pour étudier l'effet d'une augmentation du salaire minimum.
- Marché très concurrentiel: les fast-foods présentent des caractéristiques homogènes en termes de produits et de processus, ce qui permet de réduire les variations non liées au salaire minimum et d'isoler plus clairement l'effet de la politique salariale.

Solution : Card and Krueger (1994)

2) Pourquoi s'intéresser à l'industrie des fast-food, et **à quel effet s'attend-t-on?**

Solution : Card and Krueger (1994)

2) Pourquoi s'intéresser à l'industrie des fast-food, et **à quel effet s'attend-t-on?**

- Secteur dont une forte proportion de travailleurs est rémunérée au salaire minimum \implies terrain d'observation idéal pour étudier l'effet d'une augmentation du salaire minimum.

Solution : Card and Krueger (1994)

2) Pourquoi s'intéresser à l'industrie des fast-food, et **à quel effet s'attend-t-on?**

- Secteur dont une forte proportion de travailleurs est rémunérée au salaire minimum \implies terrain d'observation idéal pour étudier l'effet d'une augmentation du salaire minimum.
- Marché très concurrentiel: les fast-foods présentent des caractéristiques homogènes en termes de produits et de processus, ce qui permet de réduire les variations non liées au salaire minimum et d'isoler plus clairement l'effet de la politique salariale.

Solution : Card and Krueger (1994)

2) Pourquoi s'intéresser à l'industrie des fast-food, et **à quel effet s'attend-t-on?**

- Secteur dont une forte proportion de travailleurs est rémunérée au salaire minimum \implies terrain d'observation idéal pour étudier l'effet d'une augmentation du salaire minimum.
- Marché très concurrentiel: les fast-foods présentent des caractéristiques homogènes en termes de produits et de processus, ce qui permet de réduire les variations non liées au salaire minimum et d'isoler plus clairement l'effet de la politique salariale.

Théoriquement, on s'attend à un effet négatif:

- une hausse du coût de la main d'œuvre \implies

Solution : Card and Krueger (1994)

3) Calculer le nombre moyen d'employés en équivalent temps-plein par store dans chaque état, avant et après la réforme, et l'évolution de cette moyenne. Interpréter.

```
df %>%  
  group_by(state) %>%  
  summarise(mean_before = mean(fte_before, na.rm = T),  
            mean_after = mean(fte_after, na.rm = T)) %>%  
  ungroup() %>%  
  mutate(diff = mean_after - mean_before)
```

```
## # A tibble: 2 × 4  
##   state      mean_before mean_after  diff  
##   <chr>         <dbl>      <dbl> <dbl>  
## 1 New Jersey      20.4        21.0  0.588  
## 2 Pennsylvanie    23.3        21.2 -2.17
```

Solution : Card and Krueger (1994)

4) Calculer la différence de changement d'heures entre New Jersey et Pennsylvanie. Interprétez.

$$0.59 - (-2.17) = 2.76$$

Solution : Card and Krueger (1994)

4) Calculer la différence de changement d'heures entre New Jersey et Pennsylvanie. Interprétez.

$$0.59 - (-2.17) = 2.76$$

5) Proposez un modèle économétrique qui estime l'effet du traitement.

$$FTE = \alpha + \beta \text{ New Jersey}_i + \delta \text{ Post April 1st 1992}_t + \gamma (\text{New Jersey}_i \times \text{Post April 1st 1992}_t) + \varepsilon_{it}$$

Recap: Méthode des Doubles Différences

Data: Données de pannel

Hypothèses d'identification:

- **Parallel Trends:** en l'absence du traitement, les groupes traités et contrôles auraient évolué de manière similaire ($\mathbb{E}[Y_{i2}(0) - Y_{i1}(0)|D_i = 1] = \mathbb{E}[Y_{i2}(0) - Y_{i1}(0)|D_i = 0]$)
- **Absence d'anticipation:** le traitement n'a pas d'effet avant sa mise en place $Y_{i1}(1) = Y_{i1}(0)$

Modèle simple à deux périodes / deux groupes : pour tout individu i et période t ,

$$Y_{it} = \alpha + \beta(D_i \times Post_t) + \delta D_i + \eta Post_t + \varepsilon_{it}$$

Estimateur : Différence des différences de moyennes :

$$\hat{\beta}_{DiD} = (\bar{Y}_{\text{traités, après}} - \bar{Y}_{\text{traités, avant}}) - (\bar{Y}_{\text{contrôles, après}} - \bar{Y}_{\text{contrôles, avant}})$$

Implémentation sur R: package `fixest`

- `feols(y ~ i(post, treat, ref = 0) | FE, data = data)`

Sources

[Causal inference: The Mixtape, Scott Cunningham](#)

[Panel Data, Peter Hull](#)

[Introduction to DiD with Multiple Time Periods, Callaway and Sant'Anna](#)

[Differences-in-Differences, Florian Oswald, Gustave Kennedy, Pierre Villedieu et Mylène Feuillade](#)