

# Exercices Module 5

## DiD avec hétérogénéité de Chaisemartin & D'Haultfoeuille (2020)

Master GPE – UCA FERDI IHEDD

2026-03-24

### Objectifs de ce module

À la fin de ces exercices, vous saurez :

1. Charger et explorer les données réelles de de Chaisemartin (favara\_imbs)
2. Utiliser `did_multipllegt_dyn()` du package `DIDmultipllegtDYN`
3. Lire les sorties : effets, placebos, ATT agrégé
4. Visualiser les résultats avec `plot()`
5. Comparer les trois approches (TWFE, CS, CdH) sur les mêmes données

**Note** : Le package `DIDmultipllegtDYN` n'est pas encore disponible dans l'environnement WebR. Exécutez ces exercices dans **RStudio** après avoir installé les packages nécessaires.

### Travailler dans RStudio ?

Téléchargez le script R complet du module avec tous les exercices à compléter :

Ouvrez-le dans RStudio et complétez les `___` au fur et à mesure. Exécutez bloc par bloc avec **Ctrl+Enter**.

Les exercices sur `DIDmultipllegtDYN` nécessitent RStudio et les packages installés localement.

### Installation des packages (à faire une seule fois dans RStudio)

```
install.packages(c("DIDmultipllegtDYN", "did", "fixest", "tidyverse"))
install.packages("polars", repos = "https://rpolars.r-universe.dev") #
dépendance de DIDmultipllegtDYN
```

💡 Pourquoi ce concept est-il crucial pour un fonctionnaire ?

Les politiques publiques dans les pays en développement utilisent souvent des traitements gradués : un indice de qualité des services de santé (0 à 5), un score de décentralisation financière (0 à 10), un nombre de réformes adoptées. Les méthodes TWFE et Callaway & Sant'Anna ne gèrent que les traitements *binaires* (0/1). L'approche de Chaisemartin & D'Haultfoeuille est la seule qui s'applique correctement à ces traitements *discrets ou continus*, ouvrant la voie à des évaluations beaucoup plus riches et réalistes des politiques publiques africaines et asiatiques.

## Les données de de Chaisemartin favara\_imbs

### Exercice 5.1 Explorer les données favara\_imbs

**Contexte :** Le dataset favara\_imbs provient de Favara & Imbs (2015) et illustre un cas pratique de traitement discret multi-valeurs. Pour faire le lien avec les contextes africains et asiatiques : imaginez un indice de *décentralisation financière* allant de 0 à 4, mesurant le nombre de compétences fiscales transférées aux collectivités locales (collecte de la taxe foncière, gestion des marchés, redevances minières, etc.). Comme l'indice bancaire américain, cet indice augmente de façon échelonnée et hétérogène selon les collectivités. L'approche de Chaisemartin est exactement la bonne méthode pour ce type de données.

**Sujet original :** Impact de la **déréglementation bancaire** (Interstate Branching) sur les **prix de l'immobilier** aux États-Unis.

```
{webr-r}
# NOTE : Exécutez ce code dans RStudio car DIDmultipleTDYN n'est pas dans webr
# Voici un aperçu des données (simulé pour illustration)

library(tidyverse)

# Structure des données favara_imbs :
# - county      : identifiant du comté (FIPS)
# - year        : année (1994-2005)
# - inter_bra   : indice de déréglementation bancaire (0 à 4) – TRAITEMENT DISCRET
# - DL_hpi      : variation log prix immobilier – RÉSULTAT PRINCIPAL
# - DL_vloans_b : variation log crédit hypothécaire – MÉCANISME
# - state_n     : état (pour le clustering)

cat("Variables principales dans favara_imbs :\n")
cat("  county      : identifiant comté\n")
cat("  year        : 1994 à 2005 (12 périodes)\n")
cat("  inter_bra    : 0, 1, 2, 3, ou 4 (traitement DISCRET)\n")
cat("  DL_hpi       : variation log prix immobilier (outcome)\n")
cat("  DL_vloans_b  : variation log crédit bancaire (mécanisme)\n")
cat("  state_n      : état (pour clustering)\n\n")
cat("Avantage vs Callaway-Sant'Anna :\n")
cat("  CS ne gère que les traitements BINAIRES (0/1)\n")
cat("  CdH gère les traitements DISCRETS et CONTINUS !\n")
```

## 💡 Code RStudio Explorer les données

```
library(DIDmultiplegtDYN)
library(tidyverse)

data(favara_imbs)
glimpse(favara_imbs)

# Distribution du traitement
cat("Distribution de inter_bra :\n")
table(favara_imbs$inter_bra)

# Nombre de comtés et d'années
cat("\nNombre de comtés :", n_distinct(favara_imbs$county), "\n")
cat("Années :", sort(unique(favara_imbs$year)), "\n")
```

## Exercice 5.2 Visualiser les tendances par niveau de traitement

```
{webr-r}
library(tidyverse)

# Simulation d'une version simplifiée pour l'illustration
set.seed(42)
n_c <- 100
favara_sim <- tibble(
  county = rep(1:n_c, 8),
  year = rep(1998:2005, each = n_c),
  state_n = rep(sample(1:10, n_c, replace = TRUE), 8),
  inter_bra = rep(sample(0:3, n_c, replace = TRUE, prob = c(0.3, 0.3, 0.2, 0.2)),
  8)
) |>
mutate(
  traite_apres = pmin(inter_bra * (year >= 2001), 4),
  Dl_hpi = 0.03 + 0.01 * inter_bra + rnorm(n_c * 8, 0, 0.05)
) |>
mutate(groupe_traitement = paste0("inter_bra = ", inter_bra))

# TODO : Visualisez la moyenne de Dl_hpi par niveau de traitement et par année
favara_sim |>
  group_by(____, ____ ) |> # TODO: année et groupe
  summarise(moy = mean(Dl_hpi), .groups = "drop") |>
  ggplot(aes(x = ____, y = ____, color = ____)) + # TODO
  geom_line(linewidth = 1.1) +
  geom_point(size = 2.5) +
  labs(title = "TODO: ajoutez un titre",
        x = "Année", y = "Variation log prix (Dl_hpi)", color = "Déréglementation")
+
  theme_minimal()
```

💡 Voir la solution

```
favara_sim |>
  group_by(year, groupe_traitement) |>
  summarise(moy = mean(Dl_hpi), .groups = "drop") |>
  ggplot(aes(x = year, y = moy, color = groupe_traitement)) +
  geom_line(linewidth = 1.1) +
  geom_point(size = 2.5) +
  labs(title = "Tendances par niveau de déréglementation bancaire",
       subtitle = "Favara & Imbs (2015) – données simulées à titre
d'illustration",
       x = "Année", y = "Variation log prix (Dl_hpi)", color =
"Déréglementation") +
  theme_minimal()
```

**Interprétation** : Avec un traitement discret (0, 1, 2, 3), on observe maintenant *quatre* trajectoires au lieu de deux. Si les courbes sont parallèles avant le traitement pour tous les niveaux, l'hypothèse de tendances parallèles généralisée est soutenue. Dans le contexte fictif de la décentralisation fiscale, vous vérifieriez si les communes avec plus de compétences transférées n'avaient pas déjà une meilleure trajectoire budgétaire avant le transfert.

**Point clé** : La visualisation des tendances pour les traitements multi-valeurs est plus complexe car il y a plus de groupes à comparer. L'œil doit chercher si les trajectoires pré-traitement sont globalement parallèles — pas forcément identiques en niveau, mais évoluant à la même vitesse.

---

## Estimation avec `did_multiplegt_dyn`

### Exercice 5.3 Estimation principale (dans RStudio)

Copiez ce code dans RStudio et complétez les \_\_\_ :

```
{webr-r}
# === À EXÉCUTER DANS RSTUDIO ===
# library(DIDmultiplegtDYN)
# data(favara_imbs)

# TODO : Estimez l'effet de la déréglementation bancaire sur les prix immobiliers
# Complétez les arguments manquants

# result_fi <- did_multiplegt_dyn(
#   df       = ____,          # TODO: le dataset
#   outcome  = "___",        # TODO: variation log prix (Dl_hpi)
#   group    = "___",        # TODO: identifiant comté
#   time     = "___",        # TODO: variable année
#   treatment = "___",       # TODO: indice de déréglementation
#   effects  = ____,         # TODO: essayez 5 périodes post
#   placebo  = ____,         # TODO: essayez 3 placebos pré
#   cluster  = "state_n"     # clustering au niveau état
# )
#
# summary(result_fi)
```

```

cat("Code à compléter et exécuter dans RStudio.\n")
cat("Résultat attendu :\n")
cat(" effect_1 à effect_5 : effets positifs sur Dl_hpi (déréglementation → hausse prix)\n")
cat(" placebo_1 à placebo_3: non significatifs (pré-tendances valides)\n")
cat(" ATT agrégé : environ +0.03 à +0.05 (hausse de 3-5%)\n")

```

### 💡 Solution complète pour RStudio

```

library(DIDmultiplegtDYN)
data(favara_imbs)

result_fi <- did_multiplegt_dyn(
  df      = favara_imbs,
  outcome = "Dl_hpi",
  group   = "county",
  time    = "year",
  treatment = "inter_bra",
  effects = 5,
  placebo  = 3,
  cluster  = "state_n"
)

summary(result_fi)

# Visualisation
plot(result_fi,
      ylab = "Effet estimé (variation log prix immobilier)",
      xlab = "Périodes relatives au changement de déréglementation")

```

#### Interprétation :

La déréglementation bancaire (hausse de l'indice `inter_bra`) est associée à une **hausse des prix immobiliers** dans les comtés concernés. Cet effet reflète l'expansion du crédit hypothécaire qui stimule la demande immobilière.

Les placebos pré-traitement non significatifs valident l'hypothèse de tendances parallèles.

**Transposition dans un contexte africain :** Si vous évaluez un indice de décentralisation fiscale (0 à 4) sur les recettes communales, la même procédure s'applique : `outcome = "recettes_propres"`, `treatment = "indice_decentr"`, `group = "commune_id"`, `time = "annee"`. L'argument `placebo` vous donne les tests de pré-tendances automatiquement.

**Point clé :** L'argument `effects` détermine le nombre de périodes post-traitement estimées, et `placebo` le nombre de périodes pré-traitement testées. Pour des données annuelles sur 10 ans avec traitement à mi-chemin, `effects = 4` et `placebo = 3` sont des valeurs typiques.

**Référence :** de Chaisemartin, C. & D'Haultfoeuille, X. (2020) « Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. » *American Economic Review*, 110(9), 2964-2996.

## Comparaison des trois méthodes

### Exercice 5.4 Comparer TWFE, CS et CdH sur mpdta

Pour la comparaison, on utilise mpdta (traitement binaire) car Callaway-Sant'Anna ne gère pas les traitements multi-valeurs :

```
{webr-r}
library(did)
library(fixest)
library(tidyverse)
data(mpdta)
mpdta$treat_tv <- as.integer(mpdta$first.treat > 0 & mpdta$year >=
mpdta$first.treat)

# 1. TWFE
twfe_est <- feols(lemp ~ treat_tv | countyreal + year,
                 data = mpdta, cluster = ~countyreal)

# 2. Callaway-Sant'Anna
att_gt_res <- att_gt(
  yname = "lemp", tname = "year", idname = "countyreal",
  gname = "first.treat", data = mpdta,
  control_group = "nevertreated", print_details = FALSE
)
cs_est <- aggte(att_gt_res, type = "simple")

# TODO : Complétez le tableau de comparaison
# Extrayez les estimateurs des deux méthodes
twfe_val <- coef(twfe_est)["treat_tv"]
cs_val <- cs_est$overall.att

cat("=== Comparaison TWFE vs Callaway-Sant'Anna ===\n")
cat(sprintf("TWFE : %.4f\n", twfe_val))
cat(sprintf("Callaway-Sant'Anna : %.4f\n", cs_val))
cat("\nNote : CdH sur ces données donnerait un résultat similaire (~-0.03 à
-0.04)\n")
cat("Les trois méthodes convergent sur ces données modérément hétérogènes.\n")
```

---

### Exercice 5.5 Discussion : quelle méthode pour quel contexte ?

```
{webr-r}
# Exercice de réflexion – pas de code à compléter

# Pour chaque scénario ci-dessous, quelle méthode est la plus adaptée ?
# Écrivez votre réponse dans les commentaires.

scenarios <- list(
  "A" = "Programme de transfert conditionnel introduit simultanément dans 20
districts
      → Traitement simultané, résultat binaire (pauvreté)",
  "B" = "Réforme fiscale échelonnée sur 5 ans dans des communes de différentes
provinces"
```

```

→ Traitement échelonné, effets potentiellement hétérogènes",

"C" = "Indice de décentralisation financière (0-10) variant progressivement
selon les capacités des collectivités locales
→ Traitement continu",

"D" = "Loi anti-corruption adoptée à différentes dates selon les régions
→ Traitement échelonné, fort potentiel d'hétérogénéité"
)

for (lettre in names(scenarios)) {
  cat("Scénario", lettre, ":\n")
  cat(scenarios[[lettre]], "\n")
  cat("→ Méthode recommandée : __\n") # TODO: TWFE / CS / CdH
  cat("→ Justification : __\n\n") # TODO: votre explication
}

```

### 💡 Éléments de réponse

**Scénario A : TWFE ou DiD 2×2** – traitement simultané et binaire, TWFE est valide. Exemple concret : évaluation du programme *Filets Sociaux* au Niger, où tous les villages bénéficiaires ont été sélectionnés simultanément en 2018.

**Scénario B : Callaway-Sant’Anna** – traitement échelonné, binaire, effets potentiellement hétérogènes entre cohortes. Exemple : la décentralisation fiscale au Cameroun déployée en 3 vagues (grandes villes 2019, villes moyennes 2020, communes rurales 2021).

**Scénario C : de Chaisemartin** – seul estimateur robuste pour les traitements continus/discrets multi-valeurs. Exemple : un indice composite de gouvernance locale (0 à 10) dans les communes du Vietnam, qui augmente progressivement avec les réformes adoptées.

**Scénario D : Callaway-Sant’Anna ou de Chaisemartin** – traitement binaire échelonné avec forte hétérogénéité attendue. Les deux sont valides ; CS donne des ATT(g,t) plus granulaires, CdH est plus simple. Exemple : lois anti-corruption adoptées à des dates différentes selon les régions du Sénégal.

**Point clé – règle de décision rapide** : - Traitement simultané → TWFE - Traitement échelonné binaire → Callaway-Sant’Anna - Traitement discret ou continu → de Chaisemartin - Doute sur l’hétérogénéité → estimez avec les 3 et comparez

---

## Exercice 5.5 (suite) Réflexion sur votre contexte

Parmi les scénarios ci-dessus (A, B, C, D), lequel ressemble le plus à une politique publique que vous connaissez dans votre pays ?

Réfléchissez à : - La structure temporelle du déploiement (simultané ou échelonné ?) - La nature du traitement (binaire, discret, continu ?) - Les sources potentielles de violation des tendances parallèles dans ce contexte

*Partagez votre exemple avec le groupe en 2 minutes.*

---

**i** Fin du Module 5 et du cours complet

**Récapitulatif des compétences acquises :**

Module	Thème	Package R principal
1	Cadre causal, biais de sélection	ggplot2, base R
2	DiD 2×2, régression	lm()
3	TWFE, effets fixes, event study	fixest::feols()
4	DiD échelonnée, ATT(g,t)	did::att_gt()
5	Hétérogénéité, switchers	DIDmultiplegtDYN

**Pour aller plus loin :**

- HonestDiD : robustesse aux violations de tendances parallèles (Rambachan & Roth, 2023)
- bacondecomp : décomposition de Goodman-Bacon (2021)
- staggered : approche bayésienne pour DiD échelonnée (Roth & Sant’Anna, 2023)