

Glossaire – Évaluation d’Impact et DiD

Termes clés du cours Master GPE UCA FERDI IHEDD

Pierre Beaucoral – Master GPE UCA FERDI IHEDD

2026-03-24

i Comment utiliser ce glossaire ?

Ce glossaire rassemble tous les termes techniques du cours, organisés par thème. Pour chaque concept, vous trouverez :

- Une définition en français accessible
- L’équivalent anglais (indispensable pour lire la littérature internationale)
- Un exemple concret tiré de politiques publiques africaines
- Le module du cours où le terme est introduit

Revenez à ce glossaire chaque fois qu’un terme vous échappe dans les slides ou les exercices.

Cadre causal

ATT – Effet moyen du traitement sur les traités (*Average Treatment Effect on the Treated*)

L’ATT est l’effet moyen d’un programme ou d’une politique, calculé **uniquement pour les unités qui ont effectivement bénéficié de ce programme**. C’est la quantité d’intérêt principale en évaluation de politiques publiques : on cherche à savoir si le programme a fonctionné pour ceux qui en ont bénéficié. Il se distingue de l’ATE (voir ci-dessous) en ce qu’il ne porte pas sur l’ensemble de la population, mais seulement sur les bénéficiaires.

Exemple : Un programme de formation professionnelle est déployé dans 30 communes du Mali. L’ATT mesure l’effet moyen de cette formation sur le taux d’emploi des participants, c’est-à-dire dans ces 30 communes. On ne cherche pas à extrapoler à l’ensemble du pays – uniquement à évaluer si le programme a été efficace là où il a été mis en œuvre.

Module d’introduction : Module 1 – Introduction à l’évaluation d’impact.

ATE – Effet moyen du traitement (*Average Treatment Effect*)

L’ATE est l’effet moyen d’un programme calculé sur **l’ensemble de la population** (bénéficiaires et non-bénéficiaires confondus). Il répond à la question : « Si l’on appliquait ce programme à toute la population, quel serait l’effet moyen attendu ? » L’ATE est plus difficile à estimer que l’ATT car il nécessite de faire des hypothèses sur les effets pour des populations qui n’ont jamais participé.

Exemple : Si l'on envisage d'étendre un programme de subvention agricole à l'ensemble du Burkina Faso (et non plus seulement aux régions pilotes), l'ATE est la quantité pertinente à estimer. Elle permet de prédire ce que produirait l'extension nationale.

Module d'introduction : Module 1 – Introduction à l'évaluation d'impact.

Biais de sélection (*Selection bias*)

Le biais de sélection apparaît lorsque les unités qui participent à un programme sont **systematiquement différentes** des unités qui n'y participent pas, indépendamment de tout effet du programme. Si l'on compare directement les deux groupes sans corriger ce biais, on obtient une estimation biaisée de l'effet causal. C'est le principal obstacle à l'évaluation d'impact non expérimentale.

Exemple : Un programme d'accès au crédit agricole au Sénégal cible en priorité les exploitants les plus actifs et les mieux organisés. Si l'on compare leurs revenus à ceux des non-participants, on observera une différence – mais cette différence existait probablement déjà avant le programme. La DiD permet de corriger ce biais à condition que les tendances soient parallèles.

Module d'introduction : Module 1 – Introduction à l'évaluation d'impact.

Contrefactuel (*Counterfactual*)

Le contrefactuel est ce qui se serait passé pour une unité traitée **si elle n'avait pas bénéficié du programme**. C'est une quantité fondamentalement inobservable : on ne peut pas observer simultanément une commune avec et sans programme. Toute la difficulté de l'évaluation d'impact consiste à construire une estimation crédible de ce contrefactuel. La méthode DiD utilise le groupe de contrôle comme approximation du contrefactuel du groupe traité.

Exemple : La commune de Ségou a bénéficié d'un programme de construction de puits. On observe son taux de couverture en eau en 2023. Le contrefactuel est : quel aurait été ce taux si le programme n'avait pas existé ? On ne peut pas l'observer directement – on l'estime à partir des communes similaires sans programme.

Module d'introduction : Module 1 – Introduction à l'évaluation d'impact.

Effet causal (*Causal effect*)

L'effet causal d'un programme pour une unité donnée est la différence entre le résultat observé avec le programme et le résultat contrefactuel (sans le programme). C'est la définition formelle de l'impact. La difficulté fondamentale – appelée « problème fondamental de l'inférence causale » (Holland, 1986) – est qu'on ne peut observer qu'un seul de ces deux états pour chaque unité. Les méthodes d'évaluation d'impact (DiD, RDD, IV, etc.) sont des stratégies pour estimer cet effet malgré cette contrainte.

Exemple : Pour une école qui a reçu des manuels scolaires supplémentaires, l'effet causal sur le taux de réussite au BEPC est : taux observé avec manuels – taux qu'on aurait observé sans manuels (contrefactuel). Ce second terme est inconnu, et c'est ce que la méthode d'évaluation cherche à estimer.

Module d'introduction : Module 1 – Introduction à l'évaluation d'impact.

Inférence causale (*Causal inference*)

L'inférence causale est le cadre statistique et conceptuel permettant d'estimer des effets causaux à partir de données observationnelles ou expérimentales. Elle s'appuie sur des outils formels (modèle des résultats potentiels, graphes causaux dirigés acycliques – DAGs) pour clarifier les hypothèses

nécessaires à l'identification d'un effet causal. Elle se distingue de la simple corrélation statistique : « deux variables corrélées » ne signifie pas que l'une cause l'autre.

Exemple : Le taux de vaccination et le taux de mortalité infantile sont corrélés dans les données. Mais cela ne signifie pas forcément qu'augmenter la vaccination réduira la mortalité — d'autres facteurs (revenus, accès aux soins, qualité nutritionnelle) sont liés aux deux. L'inférence causale permet de démêler ces relations.

Module d'introduction : Module 1 — Introduction à l'évaluation d'impact.

Résultats potentiels (*Potential outcomes / Rubin causal model*)

Le cadre des résultats potentiels (développé par Donald Rubin à partir des années 1970) formalise la définition de l'effet causal. Pour chaque unité i , on définit deux résultats potentiels : $Y_i(1)$ (résultat si traitée) et $Y_i(0)$ (résultat si non traitée). L'effet causal individuel est $Y_i(1) - Y_i(0)$. Comme on n'observe qu'un seul de ces deux résultats, l'évaluation revient à estimer la moyenne de cette différence sur un groupe (ATT, ATE). Ce cadre est la fondation théorique de tout ce cours.

Exemple : Pour un agriculteur du Niger qui a reçu une subvention aux intrants, on observe $Y_i(1)$ = son rendement avec subvention. On n'observe pas $Y_i(0)$ = son rendement sans subvention. L'ATT est la moyenne de $[Y_i(1) - Y_i(0)]$ sur tous les agriculteurs subventionnés — une quantité qu'on estime grâce au groupe de contrôle.

Module d'introduction : Module 1 — Introduction à l'évaluation d'impact.

La DiD

Différence-en-Différences (*Difference-in-Differences, DiD*)

La différence-en-différences est une méthode d'évaluation d'impact quasi-expérimentale qui exploite la variation temporelle (avant/après) et la variation entre groupes (traité/contrôle) pour estimer l'effet causal d'une intervention. L'idée fondatrice est simple : on soustrait la tendance du groupe de contrôle à celle du groupe traité, ce qui élimine les facteurs communs aux deux groupes. Elle est particulièrement adaptée aux politiques publiques déployées progressivement ou de façon ciblée.

Exemple : Le Cameroun introduit en 2020 une allocation familiale dans 5 régions pilotes. On compare l'évolution de la scolarisation dans ces 5 régions (groupe traité) à l'évolution dans les 5 régions sans allocation (groupe contrôle), entre 2019 et 2022. La DiD est la différence entre ces deux évolutions.

Module d'introduction : Module 2 — Le modèle DiD 2×2.

Estimateur DiD (*DiD estimator*)

L'estimateur DiD est la formule qui calcule l'effet causal dans le cadre DiD. Dans le cas le plus simple (2 groupes, 2 périodes) : $DiD = (\bar{Y}_{\text{traité_après}} - \bar{Y}_{\text{traité_avant}}) - (\bar{Y}_{\text{contrôle_après}} - \bar{Y}_{\text{contrôle_avant}})$. Il peut être calculé manuellement à partir de moyennes, ou estimé par régression (TWFE). Dans le cas de traitements échelonnés, des estimateurs plus sophistiqués sont nécessaires (Callaway & Sant'Anna, de Chaisemartin & D'Haultfoeuille).

Exemple numérique : Revenu moyen avant/après dans le groupe traité : 850 → 1 050 ($\Delta = +200$). Revenu moyen avant/après dans le groupe contrôle : 820 → 870 ($\Delta = +50$). $DiD = 200 - 50 = 150$ millions de FCFA — c'est l'effet estimé du programme.

Module d'introduction : Module 2 — Le modèle DiD 2×2.

Groupe de contrôle (*Control group*)

Le groupe de contrôle est l'ensemble des unités qui **n'ont pas bénéficié du programme** et qui servent de référence pour estimer le contrefactuel du groupe traité. La qualité de l'évaluation dépend en grande partie de la pertinence du groupe de contrôle : il doit être suffisamment similaire au groupe traité pour que l'hypothèse de tendances parallèles soit plausible. Un mauvais groupe de contrôle produit des estimations biaisées même avec la méthode DiD.

Exemple : Pour évaluer un programme de routes rurales dans les provinces du nord du Togo, les provinces du sud (très différentes économiquement) seraient un mauvais groupe de contrôle. Les provinces du centre, avec des caractéristiques proches, seraient plus appropriées.

Module d'introduction : Module 2 – Le modèle DiD 2×2.

Groupe traité (*Treatment group*)

Le groupe traité est l'ensemble des unités qui ont **bénéficié du programme ou de la politique évaluée**. Dans le cadre DiD, on s'intéresse principalement à l'évolution de ce groupe. L'ATT est l'effet moyen estimé pour ce groupe. Dans un traitement échelonné (*staggered adoption*), le groupe traité est lui-même hétérogène : différentes unités entrent dans le programme à des dates différentes, formant des « cohortes de traitement ».

Exemple : Dans une évaluation de l'impact des centres de santé communautaires en Côte d'Ivoire, le groupe traité est constitué des villages qui ont reçu un centre entre 2018 et 2020. Les villages qui n'en ont pas reçu constituent le groupe de contrôle.

Module d'introduction : Module 2 – Le modèle DiD 2×2.

Hypothèse de tendances parallèles (*Parallel trends assumption*)

C'est l'**hypothèse d'identification centrale** de la méthode DiD. Elle stipule qu'en l'absence du programme, le groupe traité et le groupe de contrôle auraient connu des **évolutions parallèles** de la variable d'intérêt. Cette hypothèse est inobservable (car on ne peut pas savoir ce qui se serait passé sans le programme) mais on peut la tester partiellement sur la période pré-traitement. Si les deux groupes avaient des tendances similaires avant le programme, on a de bonnes raisons de penser que cette hypothèse est satisfaite.

Exemple : Si les taux de scolarisation dans les régions traitées et contrôles évoluaient de façon similaire entre 2015 et 2019 (avant le programme lancé en 2020), l'hypothèse de tendances parallèles est plausible. Un graphique d'event study permettra de le visualiser.

Module d'introduction : Module 2 – Le modèle DiD 2×2.

Période avant/après (*Pre/post period*)

La période **avant** (*pre-period*) est l'ensemble des périodes précédant le démarrage du programme — elle sert à vérifier l'hypothèse de tendances parallèles et à caractériser la situation initiale. La période **après** (*post-period*) est l'ensemble des périodes suivant le démarrage du programme — elle permet d'estimer l'effet causal et d'analyser sa persistance. Dans les méthodes modernes (Callaway & Sant'Anna), ces effets sont estimés séparément pour chaque période post-traitement, produisant un « profil temporel » de l'impact.

Exemple : Un programme de transfert monétaire est lancé en janvier 2021. Les données 2018–2020 constituent la période avant (utile pour les tests de pré-tendances) ; les données 2021–2023 constituent la période après (pour estimer l’effet et sa persistance).

Module d’introduction : Module 2 – Le modèle DiD 2×2.

Traitement (*Treatment*)

En évaluation d’impact, le « traitement » désigne le programme, la politique ou l’intervention dont on cherche à estimer l’effet. Ce terme ne désigne pas un traitement médical, mais toute intervention qui modifie la situation d’une unité. Le traitement peut être binaire (bénéficiaire ou non), continu (intensité de la subvention), ou s’étaler dans le temps (*staggered treatment*). La variable indicatrice de traitement est notée D_{it} dans les modèles de panel.

Exemple : Dans une évaluation de l’impact de l’électrification rurale en Tanzanie, le « traitement » est la connexion au réseau électrique. Une commune est traitée ($D_{it} = 1$) à partir de l’année où elle est raccordée.

Module d’introduction : Module 2 – Le modèle DiD 2×2.

TWFE et effets fixes

Effets fixes (*Fixed effects*)

Les effets fixes sont des paramètres dans un modèle de régression qui capturent les caractéristiques **non observées et constantes** d’une unité (commune, individu, pays) ou d’une période. En les incluant dans le modèle, on contrôle automatiquement toutes les différences permanentes entre unités, sans avoir besoin de les mesurer explicitement. C’est une des principales forces du modèle de panel par rapport à la simple coupe transversale.

Exemple : En estimant l’effet d’un programme agricole sur les rendements, l’effet fixe commune absorbe toutes les caractéristiques permanentes de la commune (qualité des sols, distance aux marchés, traditions culturelles) qui influencent les rendements mais restent stables dans le temps.

Module d’introduction : Module 3 – TWFE et effets fixes.

Effets fixes bidirectionnels (*Two-way fixed effects, TWFE*)

Le modèle TWFE inclut **deux séries d’effets fixes** : des effets fixes par unité (qui absorbent les différences permanentes entre unités) et des effets fixes par période (qui absorbent les chocs communs à toutes les unités à une date donnée). C’est le modèle de régression standard pour l’analyse DiD avec données de panel : $Y_{it} = \alpha_i + \lambda_t + \beta \times D_{it} + \varepsilon_{it}$, où α_i sont les effets fixes unité et λ_t les effets fixes temps.

Exemple : Dans une analyse de l’impact de la décentralisation fiscale sur les dépenses d’infrastructure de 50 communes africaines sur 10 ans, le TWFE inclut 50 effets fixes communes (absorbant les caractéristiques permanentes) et 10 effets fixes années (absorbant les chocs macroéconomiques communs comme une crise ou une sécheresse).

Module d’introduction : Module 3 – TWFE et effets fixes.

Effet fixe individuel (*Unit fixed effect*)

L'effet fixe individuel (ou « effet fixe unité ») est un paramètre spécifique à chaque unité d'observation (commune, ménage, pays) qui capte toutes les caractéristiques **non observées et permanentes** de cette unité. Il est estimé implicitement par « démoymoyennage » (*within transformation*) : on soustrait la moyenne de chaque unité avant d'estimer le modèle. Cela contrôle, sans les mesurer, des variables comme la localisation géographique, la culture locale, ou la qualité du management institutionnel.

Exemple : Pour une commune sénégalaise, l'effet fixe individuel absorbera des facteurs comme sa position côtière ou enclavée, son histoire coloniale particulière, ou la compétence historique de son administration — tous des éléments qui influencent les résultats mais ne varient pas dans le temps.

Module d'introduction : Module 3 — TWFE et effets fixes.

Effet fixe temporel (*Time fixed effect*)

L'effet fixe temporel (ou « effet fixe année ») est un paramètre commun à **toutes les unités pour une même période**. Il absorbe les chocs macroéconomiques, climatiques ou politiques qui affectent toutes les unités simultanément : une récession mondiale, une sécheresse régionale, une élection nationale. Sans cet effet fixe, ces chocs communs pourraient être confondus avec l'effet du programme.

Exemple : En 2020, la pandémie de Covid-19 a réduit l'activité économique dans tous les pays d'Afrique subsaharienne. L'effet fixe année 2020 absorbera cet impact commun, permettant d'isoler l'effet spécifique d'un programme national de soutien aux PME sur les entreprises bénéficiaires.

Module d'introduction : Module 3 — TWFE et effets fixes.

Erreurs standard groupées (*Clustered standard errors*)

Les erreurs standard groupées (*clustered*) tiennent compte du fait que les observations appartenant au même groupe (même commune, même région, même pays) sont **corrélées entre elles** — elles ne sont pas indépendantes comme le suppose l'estimateur OLS classique. Ne pas regrouper les erreurs standard conduit à sous-estimer l'incertitude et à obtenir des p-values trop faibles (on « trouve » des effets significatifs qui ne le sont pas vraiment). En pratique, on regroupe au niveau où s'applique le traitement.

Exemple : Si un programme est assigné au niveau des districts (tous les villages d'un district reçoivent ou non le programme), les erreurs standard doivent être regroupées au niveau du district — pas au niveau des villages individuels. Ne pas le faire gonflerait artificiellement la précision des estimations.

Module d'introduction : Module 3 — TWFE et effets fixes.

Modèle de panel (*Panel data model*)

Un modèle de panel (ou données longitudinales) est un modèle qui analyse des données où les **mêmes unités sont observées à plusieurs périodes**. Il combine une dimension transversale (différentes unités) et une dimension temporelle (plusieurs dates). Les données de panel permettent d'inclure des effets fixes et d'identifier des effets causaux qui seraient impossibles à estimer en coupe transversale. La DiD est fondamentalement un estimateur de panel.

Exemple : Un suivi de 200 ménages agricoles dans la région de Mopti (Mali) sur 5 ans (2017–2022), avec des données annuelles sur les revenus, la surface cultivée et la participation à des programmes d'appui. C'est une structure de panel : 200 unités × 5 périodes = 1 000 observations.

Module d'introduction : Module 3 — TWFE et effets fixes.

Variable de traitement *time-varying* (*Time-varying treatment*)

Une variable de traitement *time-varying* (qui varie dans le temps) indique si une unité est traitée à chaque période. Contrairement au cas simple où le traitement est permanent et simultané, la variable D_{it} peut valoir 0 pour les périodes pré-traitement et 1 pour les périodes post-traitement, avec des dates d'entrée différentes selon les unités. Cette structure est au cœur du problème du traitement échelonné (*staggered adoption*).

Exemple : Dans une base de données de communes sur 8 ans, la commune A entre dans un programme en 2019, la commune B en 2021, la commune C en 2023. La variable D_{it} vaut 0 avant la date d'entrée et 1 à partir de cette date — elle est donc « *time-varying* » et diffère entre communes.

Module d'introduction : Module 3 — TWFE et effets fixes.

DiD échelonnée moderne

ATT(g,t) — Effet moyen du traitement pour le groupe g à la période t

L'ATT(g,t) est la quantité d'intérêt centrale de la méthode Callaway & Sant'Anna (2021). Il mesure l'effet moyen du traitement pour les unités qui ont été traitées pour la première fois à la période g (la « cohorte »), évalué à la période t . Cette décomposition fine permet de détecter une **hétérogénéité temporelle et entre cohortes** que le TWFE standard masquerait en les agrégeant dans un seul coefficient.

Exemple : Dans une évaluation de l'universalisation de l'état civil en Afrique de l'Ouest, ATT(2018, 2020) est l'effet moyen pour les communes traitées en 2018, mesuré deux ans après leur traitement (2020). On peut ainsi distinguer si les « premières adoptantes » ont eu de meilleurs ou moins bons résultats que les « adoptantes tardives ».

Module d'introduction : Module 4 — Callaway & Sant'Anna.

Cohorte de traitement (*Treatment cohort*)

Une cohorte de traitement est un groupe d'unités qui ont été **traitées pour la première fois à la même période**. Dans le cadre du traitement échelonné, on distingue autant de cohortes que de dates d'entrée dans le traitement. L'approche Callaway & Sant'Anna estime un ATT séparé pour chaque cohorte à chaque période, avant d'agréger ces estimations en mesures synthétiques (effet moyen, profil temporel).

Exemple : Un programme de vaccination scolaire au Ghana est déployé progressivement : 50 écoles en 2018, 80 autres en 2019, 100 autres en 2020. Il y a donc 3 cohortes de traitement (2018, 2019, 2020). L'analyse Callaway & Sant'Anna estime l'effet séparément pour chacune de ces cohortes.

Module d'introduction : Module 4 — Callaway & Sant'Anna.

Sun & Abraham (2021) — Estimateur par interaction (*interaction-weighted estimator*)

L'approche de Sun & Abraham (2021) propose une alternative au TWFE pour les traitements échelonnés. Elle s'appuie sur les termes d'interaction entre les indicateurs de cohorte et les indicateurs de période relative, puis les agrège avec des poids basés sur la taille des cohortes. Elle est implémentée dans R via la fonction `sunab()` du package `fixest`, ce qui la rend très accessible. Comme Callaway & Sant'Anna, elle évite les poids négatifs du TWFE.

Exemple d'utilisation :

```
library(fixest)
es_sa <- feols(
  lemp ~ sunab(first.treat, year) | countyreal + year,
  data = mpdta,
  cluster = ~countyreal
)
iplot(es_sa)
```

Module d'introduction : Module 3 – TWFE et effets fixes (exercices).

DID_M (de Chaisemartin & D'Haultfoeuille estimator)

Le DID_M est l'estimateur proposé par de Chaisemartin & D'Haultfoeuille (2020, 2023). Il est conçu pour les situations de traitement échelonné et hétérogène. Il identifie les « switchers » (unités qui changent de statut de traitement) et calcule une DiD locale pour chaque changement, pondérée de façon à éviter les poids négatifs qui affectent le TWFE. Il est implémenté dans R avec le package DIDmultiplegtDYN.

Exemple : Dans une évaluation de l'impact de la décentralisation budgétaire sur les investissements locaux en Tanzanie, certaines communes obtiennent plus d'autonomie (switchers de 0 vers 1), d'autres en perdent (switchers de 1 vers 0). Le DID_M exploite ces deux types de variation pour identifier l'effet.

Module d'introduction : Module 5 – de Chaisemartin & D'Haultfoeuille.

Event study / Graphique événementiel

Un graphique d'évent study (ou graphique événementiel) représente les coefficients DiD estimés **pour chaque période** autour de la date de traitement. Les coefficients pré-traitement (périodes avant l'intervention) permettent de tester visuellement l'hypothèse de tendances parallèles : s'ils sont proches de zéro et non significatifs, l'hypothèse est plausible. Les coefficients post-traitement décrivent le profil temporel de l'effet (immédiat, progressif, persistant ou décroissant).

Exemple : Un graphique d'évent study pour un programme d'irrigation au Maroc montre des coefficients proches de zéro pour les années -3 à -1 (avant le programme) puis une hausse progressive des rendements de +8 % à +15 % sur les années +1 à +3 (après le programme). Ce profil suggère un effet significatif et croissant dans le temps.

Module d'introduction : Module 4 – Callaway & Sant'Anna ; Module 5 – de Chaisemartin & D'Haultfoeuille.

Poids négatifs (Negative weights)

Le problème des poids négatifs est une défaillance du TWFE en présence de traitement échelonné avec effets hétérogènes. Goodman-Bacon (2021) a montré que le coefficient TWFE peut être décomposé en une moyenne pondérée d'estimateurs DiD « 2×2 » – mais certains de ces poids peuvent être **négatifs**, ce qui signifie que certains groupes déjà traités servent de « faux contrôle » pour d'autres groupes. Si les effets hétérogènes sont suffisamment différents entre cohortes, le coefficient TWFE peut avoir le mauvais signe.

Exemple : Un programme de transferts conditionnels est déployé progressivement sur 4 ans dans un pays. Si les premiers bénéficiaires (cohorte 2017) ont des effets plus forts que les derniers (cohorte

2020), le TWFE peut attribuer des poids négatifs à la cohorte 2020, biaisant l'estimateur global vers le bas — voire vers un signe négatif.

Module d'introduction : Module 3 — TWFE (limites) ; Module 4 — Callaway & Sant'Anna.

Staggered adoption / Traitement échelonné

On parle de traitement échelonné (*staggered adoption*) lorsque différentes unités entrent dans un programme à des dates différentes, et que le traitement est permanent (une fois traitée, une unité reste traitée). C'est le cas le plus courant en politiques publiques réelles : les réformes sont rarement déployées simultanément sur tout un territoire. Le traitement échelonné pose des problèmes spécifiques pour le TWFE et motive l'usage des méthodes modernes (Callaway & Sant'Anna, de Chaisemartin & D'Haultfoeuille).

Exemple : La réforme de la gestion des finances publiques au niveau des collectivités locales au Bénin est déployée progressivement : 15 communes pilotes en 2016, 40 communes supplémentaires en 2018, le reste en 2020. C'est un cas typique de staggered adoption.

Module d'introduction : Module 4 — Callaway & Sant'Anna.

Switchers / Stayers (de Chaisemartin terminology)

Dans la terminologie de de Chaisemartin & D'Haultfoeuille, les **switchers** sont les unités dont le statut de traitement **change** d'une période à l'autre (de non-traité à traité, ou inversement). Les **stayers** sont les unités dont le statut de traitement **ne change pas** entre deux périodes consécutives (restent toujours traitées, ou toujours non traitées). La méthode DID_M s'appuie sur les switchers pour identifier l'effet causal et utilise les stayers comme groupe de contrôle local.

Exemple : Dans un suivi de 100 entreprises camerounaises sur 5 ans, une entreprise qui obtient un agrément pour des avantages fiscaux en 2020 est un switcher (0→1). Une entreprise qui a l'agrément depuis le début de l'observation est un stayer traité. Une entreprise qui n'a jamais eu l'agrément est un stayer non traité.

Module d'introduction : Module 5 — de Chaisemartin & D'Haultfoeuille.

Tests et validation

Bootstrap

Le bootstrap est une méthode de rééchantillonnage qui permet d'estimer la distribution d'un estimateur statistique **sans faire d'hypothèses sur la distribution des données**. On tire avec remise un grand nombre d'échantillons (typiquement 999 ou 1 999) de la même taille que l'échantillon original, on recalcule l'estimateur sur chaque tirage, et on utilise la distribution empirique obtenue pour calculer des intervalles de confiance. Les méthodes Callaway & Sant'Anna et DIDmultiplegDYN utilisent le bootstrap multiplier pour l'inférence.

Exemple : Avec seulement 30 communes dans votre évaluation, les hypothèses asymptotiques des erreurs standard classiques sont fragiles. Le bootstrap permet d'obtenir des intervalles de confiance valides même avec ce petit nombre de clusters, en simulant 999 fois le processus d'échantillonnage.

Module d'introduction : Module 4 — Callaway & Sant'Anna.

Intervalle de confiance (*Confidence interval*)

Un intervalle de confiance à 95 % est une plage de valeurs telle que, si l'on répétait l'expérience un grand nombre de fois avec différents échantillons, 95 % des intervalles construits de cette façon contiendraient la vraie valeur du paramètre. En pratique, c'est la plage de valeurs compatible avec les données. Un intervalle de confiance étroit indique une estimation précise ; un intervalle large indique une grande incertitude.

Exemple : Une évaluation estime que le programme de bourses scolaires au Niger a augmenté le taux de scolarisation de 8 points de pourcentage, avec un intervalle de confiance à 95 % de [3,5 ; 12,5]. Cela signifie que l'effet est statistiquement différent de zéro et que l'ampleur plausible va de +3,5 à +12,5 points.

Module d'introduction : Module 2 – Le modèle DiD 2×2.

P-value / Valeur-p

La p-value (ou valeur-p) est la probabilité d'obtenir un résultat au moins aussi extrême que celui observé, **si l'hypothèse nulle était vraie** (typiquement : si l'effet était nul). Une p-value inférieure à 0,05 est conventionnellement considérée comme « statistiquement significative » au seuil de 5 %. Attention : une p-value faible indique une évidence contre l'hypothèse nulle, mais ne dit rien sur l'ampleur ou l'importance pratique de l'effet.

Exemple : Une évaluation d'un programme de microcrédit trouve un coefficient de +45 000 FCFA sur les revenus annuels, avec une p-value de 0,03. Cela signifie : si le programme n'avait aucun effet, il y aurait 3 % de chances d'observer un résultat aussi extrême. On rejette donc l'hypothèse nulle d'« effet nul » au seuil de 5 %.

Module d'introduction : Module 2 – Le modèle DiD 2×2.

Placebos pré-traitement (*Pre-treatment placebos*)

Les tests placebo consistent à estimer la DiD **sur des périodes antérieures au traitement**, comme si le programme avait commencé plus tôt (à une date fictive). Si on trouve des effets significatifs sur ces périodes fictives, cela suggère que les groupes traité et contrôle avaient des tendances différentes avant même le programme – ce qui invalide l'hypothèse de tendances parallèles. Ces tests sont une forme de validation cruciale de l'analyse.

Exemple : Un programme est lancé en 2020. On estime l'effet DiD « fictif » pour les années 2016, 2017, 2018, 2019 (comme si le programme avait commencé à ces dates). Si ces coefficients fictifs sont proches de zéro et non significatifs, l'hypothèse de tendances parallèles est crédible.

Module d'introduction : Module 4 – Callaway & Sant'Anna.

Test de pré-tendances (*Pre-trends test*)

Le test de pré-tendances est un test statistique formel visant à vérifier si les groupes traité et contrôle avaient des évolutions parallèles **avant le démarrage du programme**. Il est visuellement représenté par un graphique d'event study où les coefficients pré-traitement devraient être proches de zéro. Un test formel consiste à tester la nullité jointe de tous les coefficients pré-traitement. Un échec à ce test remet en cause la validité de l'analyse DiD.

Exemple : Avant de conclure que l'électrification rurale a augmenté les revenus agricoles au Mozambique, on estime les coefficients DiD pour les 3 années précédant l'électrification. Si ces coefficients sont

statistiquement nuls et que le graphique montre des tendances superposées, la crédibilité de l'analyse est renforcée.

Module d'introduction : Module 4 — Callaway & Sant'Anna ; Module 5 — de Chaisemartin & D'Haultfoeuille.

Significativité statistique vs. significativité pratique

Un résultat est **statistiquement significatif** lorsque la p-value est inférieure à un seuil conventionnel (0,05 ou 0,01). Un résultat est **pratiquement significatif** (ou économiquement significatif) lorsque l'ampleur de l'effet est suffisamment grande pour avoir une importance pour la décision publique. Ces deux notions sont indépendantes : avec de très grands échantillons, des effets minuscules peuvent être statistiquement significatifs mais pratiquement négligeables. Inversement, avec de petits échantillons, des effets importants peuvent ne pas être statistiquement significatifs.

Exemple : Une évaluation d'un programme de cantines scolaires trouve un effet de +0,3 points sur les résultats scolaires (sur 100), statistiquement significatif à 1 % ($p < 0,01$). L'effet est statistiquement solide mais pratiquement modeste. Le décideur doit se demander si ce gain justifie le coût du programme par rapport à des alternatives (bourses, manuels scolaires, etc.).

Module d'introduction : Module 2 — Le modèle DiD 2×2.

R et outils

feols() — package fixest

`feols()` est la fonction principale du package `fixest` pour estimer des modèles de régression linéaire avec effets fixes, y compris le TWFE. Elle est très rapide (optimisée pour les grands panels) et gère automatiquement les erreurs standard groupées. La syntaxe est : `feols(y ~ D | id + annee, data = ..., cluster = ~id)`, où `id + annee` spécifie les effets fixes unité et temps.

Exemple d'utilisation :

```
library(fixest)
resultat <- feols(
  revenu ~ traitement | commune + annee,
  data = donnees_panel,
  cluster = ~commune
)
summary(resultat)
```

Module d'introduction : Module 3 — TWFE et effets fixes.

att_gt() — package did

`att_gt()` est la fonction centrale du package `did` (Callaway & Sant'Anna). Elle estime les ATT(g,t) pour chaque combinaison de cohorte g et période t . Elle requiert de spécifier : la variable d'outcome, la variable de groupe de traitement (première date de traitement, 0 pour les jamais-traités), la variable de temps, l'identifiant des unités, et le groupe de contrôle à utiliser (« `nevertreated` » ou « `notyettreated` »).

Exemple d'utilisation :

```
library(did)
resultats_gt <- att_gt(
  yname = "revenu",
  gname = "annee_traitement",
  idname = "id_commune",
  tname = "annee",
  data = donnees_panel,
  control_group = "nevertreated"
)
```

Module d'introduction : Module 4 – Callaway & Sant'Anna.

aggte() – package did

aggte() est la fonction du package did qui **agrège** les estimateurs ATT(g,t) individuels en mesures synthétiques plus interprétables. Elle propose plusieurs types d'agrégation : type = "simple" (moyenne pondérée), type = "dynamic" (profil temporel – utilisé pour l'évent study), type = "group" (effet moyen par cohorte), type = "calendar" (effet moyen par période calendaire).

Exemple d'utilisation :

```
# Profil temporel (event study)
event_study <- aggte(resultats_gt, type = "dynamic")
summary(event_study)

# Effet moyen global
effet_global <- aggte(resultats_gt, type = "simple")
summary(effet_global)
```

Module d'introduction : Module 4 – Callaway & Sant'Anna.

ggdid() – package did

ggdid() est la fonction de visualisation du package did. Elle prend en entrée le résultat de aggte() et produit automatiquement un graphique d'évent study bien formaté, avec les intervalles de confiance. Elle est construite sur ggplot2 et peut donc être personnalisée avec les fonctions habituelles de ce package.

Exemple d'utilisation :

```
# Graphique d'event study automatique
ggdid(event_study) +
  labs(
    title = "Event study – Programme de vaccination scolaire",
    x = "Périodes relatives au traitement",
    y = "Effet estimé (ATT)"
  ) +
  theme_minimal(base_size = 13)
```

Module d'introduction : Module 4 – Callaway & Sant'Anna.

did_multiplegt_dyn() — package DIDmultiplegtDYN

`did_multiplegt_dyn()` est la fonction principale du package `DIDmultiplegtDYN`, qui implémente l'estimateur de de Chaisemartin & D'Haultfoeuille pour les traitements échelonnés et hétérogènes. Elle identifie les switchers, estime des DiD locales, et produit un graphique d'évent study. Elle est robuste aux effets hétérogènes entre cohortes et aux traitements non monotones (une unité peut entrer et sortir du traitement).

Exemple d'utilisation :

```
library(DIDmultiplegtDYN)
resultat_dCdH <- did_multiplegt_dyn(
  df      = donnees_panel,
  outcome = "revenu",
  group   = "id_commune",
  time    = "annee",
  treatment = "traitement",
  effects  = 4 # nombre de périodes post-traitement à estimer
)
summary(resultat_dCdH)
```

Module d'introduction : Module 5 — de Chaisemartin & D'Haultfoeuille.

Comment lire un article en DiD

Lorsque vous lisez un article empirique ou un rapport d'évaluation utilisant la méthode DiD, voici les cinq questions à vous poser systématiquement pour évaluer la crédibilité des résultats.

Grille de lecture d'une évaluation DiD

1. Quelle est l'hypothèse d'identification ?

L'auteur explique-t-il pourquoi les groupes traité et contrôle auraient eu des tendances parallèles en l'absence du programme ? Sur quoi repose cette affirmation ? La construction du groupe de contrôle est-elle justifiée ? Un groupe de contrôle mal choisi invalide toute l'analyse, quelle que soit la sophistication de la méthode.

2. Le graphique d'event study est-il présenté ?

Tout article DiD sérieux doit présenter un graphique d'event study montrant les coefficients pré et post-traitement. Si les coefficients pré-traitement sont proches de zéro et non significatifs, l'hypothèse de tendances parallèles est crédible. Si ce graphique est absent ou si les coefficients pré-traitement sont significatifs, soyez prudent quant aux conclusions.

3. Les erreurs standard sont-elles regroupées au bon niveau ?

Le niveau de regroupement (*clustering*) des erreurs standard doit correspondre au niveau où le traitement est assigné. Si le programme est assigné au niveau du district, il faut regrouper au niveau du district – pas au niveau des individus ou des communes. Des erreurs standard non regroupées (ou regroupées à un niveau trop fin) produisent des p-values artificiellement faibles.

4. L'auteur utilise-t-il le TWFE seul, ou une méthode robuste à l'hétérogénéité ?

Si le traitement est échelonné (dates d'entrée différentes selon les unités), le TWFE seul est potentiellement biaisé. Vérifiez si l'auteur signale ce problème et utilise une méthode robuste (Callaway & Sant'Anna, de Chaisemartin & D'Haultfoeuille, Borusyak et al.) ou s'il se contente du TWFE sans discussion des limites.

5. Les résultats sont-ils robustes à d'autres spécifications ?

Un résultat crédible doit résister à des variations raisonnables : changement du groupe de contrôle, inclusion ou exclusion de variables de contrôle supplémentaires, modification du niveau de regroupement des erreurs standard, choix d'une fenêtre temporelle différente. Si les résultats changent radicalement selon la spécification, la robustesse de l'analyse est discutable.

Les signaux d'alerte dans un rapport DiD

- Absence de graphique d'event study ou de test de pré-tendances
- Groupe de contrôle très différent du groupe traité sans justification
- Erreurs standard non regroupées avec des données de panel
- Utilisation du TWFE seul sur des données de traitement échelonné sans discussion des poids négatifs
- Résultats présentés uniquement pour la spécification la plus favorable
- Taille d'effet extraordinairement grande sans analyse de robustesse

i Pour aller plus loin dans la lecture critique

La lecture critique d'une évaluation d'impact s'améliore avec la pratique. Quelques références utiles pour approfondir :

- **Cunningham (2021)** – *Causal Inference: The Mixtape* (disponible gratuitement en ligne) – une introduction intuitive et très accessible à la DiD et aux méthodes d'identification causale.
- **Roth et al. (2023)** – « What's Trending in Difference-in-Differences? » *Journal of Econometrics* – une revue technique des développements récents sur la DiD.
- **Callaway & Sant'Anna (2021)** – « Difference-in-Differences with Multiple Time Periods » *Journal of Econometrics* – l'article fondateur de la méthode du Module 4.
- **de Chaisemartin & D'Haultfoeuille (2020)** – « Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects » *American Economic Review* – l'article fondateur du Module 5.