
Master 2 ADEPP
Travaux Dirigés
Différences de Différences (DID)
Enoncé

Analyse de l'impact de la construction d'écoles secondaires sur le niveau d'éducation primaire

Cet exercice est consacré à la méthode des différences de différences (DID) comme outil d'évaluation d'impact. Nous allons tenter de reproduire les résultats de l'article de Beegle et al. (2006). Nous présentons d'abord le contexte de l'article ainsi que les enjeux statistiques et économétriques.

Éléments de contexte

Dans cet exercice, il vous est demandé d'analyser des données collectées par la Banque Mondiale, portant sur des communautés et des ménages de la région de Kagera en Tanzanie. En utilisant la méthode des différences de différences, vous cherchez à analyser l'impact de la construction d'écoles secondaires sur le niveau d'éducation primaire des enfants. En effet, un grand nombre de nouvelles écoles secondaires ont été construites dans les années 1980 dans la région de Kagera. Cependant, tous les villages n'en ont pas bénéficié, permettant la comparaison d'individus provenant de villages avec et sans écoles secondaires, quelques années plus tard.

La base de données est issue d'une enquête menée en 1995. Elle concerne des individus vivant dans des villages avec et sans écoles secondaires. Deux niveaux d'analyse sont donc pertinents : la comparaison entre villages (avec et sans écoles, 1^{ère} différence), et la comparaison entre individus (en âge ou non d'être scolarisés au moment de la construction des écoles, 2^{ème} différence).

Les villages où les nouvelles écoles secondaires ont été construites seront le « groupe de traitement ». Les autres villages (sans école secondaire sur la période) seront le « groupe de contrôle ». Pour chaque groupe de villages, on définira deux classes ou cohortes d'âges : la cohorte des jeunes individus en âge d'être scolarisés (âgés de 6 à 16 ans en 1985), et la cohorte des individus plus âgés (âgés de 20 à 40 ans en 1985). L'intuition sous-jacente à la stratégie d'identification que vous allez utiliser est la suivante : l'impact de la construction de nouvelles écoles ne devrait affecter que les individus qui étaient encore en âge d'être scolarisés au moment de leur construction. La comparaison des deux cohortes d'individus jeunes ou plus âgés, combinée à la construction d'écoles dans certains villages et pas dans d'autres, permet ainsi de mettre en œuvre une méthode d'estimation en différences de différences.

Description des variables

- $treat = 1$ si une école secondaire a été construite dans le village ;
- $ycohort = 1$ si l'individu avait entre 6 et 16 ans en 1995 ;
- $ocohort = 1$ si la personne était âgée de 21 à 41 ans en 1995 ;
- $ycohort \times treat$ = termes d'interaction entre cohorte et groupe de traitement ;
- $primary = 1$ if la personne a achevé l'école primaire ;
- $electric = 1$ si le village est électrifié ;
- $pipwater = 1$ si le village a accès à un réseau d'eau potable ;
- $distcapital$ = distance entre le village et la capitale (Dar es Salam) ;
- $cluster$ = identifiant du village ;
- $age1995$ = Age de l'individu en 1995 ;
- $sex1995 = 1$ si la personne est un homme.

1. Première partie : Statistiques descriptives

Cette partie est destinée à décrire les caractéristiques de votre échantillon. Il s'agit tout d'abord de répondre simplement aux 2 questions suivantes : (i) quelle est l'importance du problème considéré dans votre échantillon (niveau d'éducation) ? et (ii) quelles sont les différences qui existent entre le groupe de traitement et le groupe de contrôle ?

- (a) Présentez tout d'abord les caractéristiques moyennes des individus (âge, sexe, éducation primaire) issus des cohortes jeunes (en 1995) selon qu'ils appartiennent au groupe de traitement ou au groupe de contrôle. Pour chaque variable considérée, réalisez un test statistique (d'égalité des moyennes) pour évaluer s'il existe des différences significatives entre les deux groupes. Commentez vos résultats. En particulier, si vous trouvez des différences significatives, indiquez en quoi cela peut affecter les résultats de l'évaluation d'impact.
- (b) Présentez des statistiques descriptives équivalentes au niveau village (c'est-à-dire en n'utilisant qu'une observation par village) pour les variables d'accès à l'électricité, d'accès à l'eau potable, et de distance à la capitale selon que le village appartienne au groupe de traitement ou au groupe de contrôle. Commentez vos résultats. Si vous trouvez des différences significatives, indiquez en quoi cela peut affecter les résultats de l'évaluation d'impact
- (c) Sur la base de vos réponses aux questions (a) et (b), pensez-vous que les villages traités sont différents des villages contrôle ? Pourquoi pensez-vous que les écoles ont d'abord été construites dans les villages du groupe de traitement ?

2. Deuxième partie : Estimation en simple différence

- (a) Comparez les niveaux d'éducation des individus des cohortes jeunes (mesurés en 1995) selon qu'ils résident dans un village traité ou contrôle. Pour ce faire, utilisez une régression de la probabilité d'avoir achevé le cycle primaire sur la variable binaire de traitement, en restreignant l'échantillon aux individus des cohortes jeunes. Commentez la valeur et la significativité du paramètre estimé.
- (b) Sur la base de ces résultats, peut-on dire que la construction d'écoles secondaires a eu un impact direct sur l'éducation primaire ? Ou pourrait-on attribuer ces résultats à d'autres facteurs ?
- (c) Reproduisez la régression de la question (a), en ajoutant des variables de contrôle. Est-ce que la valeur du coefficient estimé de la variable de traitement change ? Commentez vos résultats.

3. Troisième partie : Estimation en différences de différences (comparaison de moyennes simples)

- (a) Calculez les moyennes simples de la probabilité d'avoir achevé le cycle primaire pour chacun des groupes suivants :
- Quelle proportion de la cohorte « jeune » a achevé l'école primaire dans les villages de traitement ?
 - Quelle proportion de la cohorte « âgée » a achevé l'école primaire dans les villages de traitement ?
 - Quelle proportion de la cohorte « âgée » a achevé l'école primaire dans les villages de contrôle ?
 - Quelle proportion de la cohorte « âgée » a achevé l'école primaire dans les villages de contrôle ?
- (b) Remplissez ensuite le tableau suivant avec la proportion d'individus ayant achevé l'école primaire, selon le groupe considéré :

	Villages traités	Villages contrôle	
Cohorte jeune	A =	C =	
Cohorte âgée	B =	D =	
1ère différence	$(A - B) =$	$(C - D) =$	Double différence (DID) : $(A - B) - (C - D) =$

4. Quatrième partie : Estimation en différences de différences (régression)

- (a) Estimez l'impact du programme de construction d'écoles en utilisant l'estimateur en différences de différences (DID) à l'aide de 3 régressions linéaires (MCO) (i) sans variable de contrôle, (ii) avec variables de contrôle et (iii) avec effets fixes village. Quels résultats obtenez-vous pour l'impact du traitement ? Ces résultats sont-ils significatifs ? Comparez ces résultats à ceux obtenus dans la section précédente par une simple comparaison de moyennes.
- (b) Reproduisez l'estimation DID de l'impact du programme en distinguant les individus selon leur sexe. Commentez.
- (c) Testez l'hypothèse d'identification de l'estimateur DID, en utilisant comme groupe « placebo » les individus des cohortes encore plus âgées, de 40 à 50 ans (donc non-affectées par le programme).