

Statistiques et économétrie appliquées
Printemps 2019

Projet B

Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment

Consigne:

- Nombre maximal de points : 60
- Entre 3-4 participants par groupe
- Votre base de données est spécifique à votre groupe

Ce projet doit être rendu au plus tard le vendredi 31 mai 2019 à 23h55 sur Moodle via la plateforme prévue à cet effet. Une fois le projet envoyé via la plateforme, vous aurez l'occasion de modifier les fichiers envoyés avant la date limite. L'envoi du projet par un membre de votre groupe engage tout le groupe.

Pour la version finale, veuillez rendre un dossier zippé (.zip) contenant les éléments suivants :

- Un fichier **.pdf** contenant vos résultats finaux (**maximum 10 pages, hors tableaux et graphiques**) et vos commentaires
- Un **DO-FILE** Stata contenant les codes que vous avez utilisés
- Un **LOG-FILE** exporté comme un pdf (voir slides complémentaires)
- Votre **BASE DE DONNEES INITIALE** et votre **BASE DE DONNEES FINALE**
- Un fichier **EXCEL** contenant les tableaux de statistiques descriptives et tableaux de régressions qui figurent dans votre PDF.

Veuillez renommer **tous** vos fichiers avec le numéro de votre groupe (p.ex. G-1.do, G-1.dta, G-1.pdf, G-1.xls). Les régressions doivent être effectuées dans Stata et les résultats exportés dans Excel (ou un équivalent).

L'exercice ci-dessous est basé sur l'article : *Macours, Karen, Norbert Schady, and Renos Vakis. "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment." American Economic Journal: Applied Economics 4, no. 2 (April 2012): 247–73 .<https://doi.org/10.1257/app.4.2.247>* téléchargeable sur Moodle. Nous avons pris des éléments simplifiés de l'article, mais nous avons gardé la structure de l'analyse qui est faite dans le papier, vos résultats ne seront donc pas identiques à ceux de l'article.

Cash Transfers, Behavioral Changes and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment (résumé)

Le développement infantile est un prédicteur important de succès au cours de la vie. Dans les pays en voie de développement, un lien a été démontré entre le développements cognitif chez les enfants d'âge préscolaire et les résultats scolaires plus tard.

Cette étude randomisée (RCT) évalue l'impact du programme Nicaraguayen de Conditional Cash Transfers (CCT), *Atención a Crisis*, sur le développement au cours la petite enfance. Le programme *Atención a Crisis* a été mis en place dans six municipalités rurales du Nicaragua entre novembre 2005 et décembre 2006. 56 communautés ont été tirées au sort pour participer au programme et 50 communautés ont été tirées au sort pour faire partie du groupe contrôle.

Au sein des communautés traitées, un nouveau tirage au sort a été réalisé parmi les familles éligibles au programme et les foyers ont été aléatoirement répartis en trois groupes. Nous ne considérons que les deux premiers dans le cadre de ce travail.

Le groupe 1 reçoit un transfert monétaire chaque deux mois. Pour les foyers avec des enfants entre 0 et 5 and, cette subvention est en théorie (non vérifié en pratique) conditionnelle à la participation à des contrôles médicaux réguliers pour les enfants. Les foyers avec des enfants entre 7 et 15 ans reçoivent en plus une subvention éducative, conditionnelle à la participation scolaire (participation effectivement vérifiée). Les montants perçus correspondent à environ 15% des dépenses par habitant du bénéficiaire moyen de l'échantillon. Nous appelons ce traitement « CCT ».

Les familles tirées au sort pour faire partie du groupe 2 reçoivent en plus du groupe 1 une bourse d'éducation pour l'un des membres du foyer, lui permettant de participer à des formations professionnelles. Nous appelons ce traitement « CCT + Formation ».

Une enquête de référence a été effectuée avant le tirage au sort aléatoire en 2005. Une première enquête de suivi a été collectée en 2006, puis une seconde entre 2008 et 2009.

Dans ce projet, nous nous intéressons à l'impact des programmes « CCT » et de « CCT + Formation » lors de la troisième enquête (2008) sur plusieurs indicateurs de développement infantile. La base de données dont vous disposez comporte des données individuelles qui ont été récoltées avant l'intervention (2005), puis trois ans après l'intervention (2008).

1. Nettoyage de la base de données (10points)

1.1. A partir de la variable **gender**, créez une variable binaire **male** prenant la valeur 1 pour les hommes et 0 pour les femmes. **(1 point)**

1.1.1. Donnez à la valeur 1 le label « Masculin » et à la valeur 0 le label « Féminin » **(1 point)**

1.2. Grâce à la variable **group**, créez une variable binaire **cct** prenant la valeur 1 pour les individus du groupe CCT, une variable **ccttraining** prenant la valeur 1 pour les individus du groupe « CCT + Training », et une variable **control** prenant la valeur 1 pour les autres individus. **(1 point)**

1.3. Recodez la variable **group = 1** si « CCT », 2 si « CCT + Formation », 3 si « Contrôle ». **(2 points)**. Ajoutez les labels correspondants. **(1 point)** (Astuce : L'objectif final est d'avoir une variable appelée « group » qui a des valeurs numériques comme indiqué ci-dessus.)

1.4. A partir des variables **male**, **cct** et **ccttraining** créez les variables **cct_male** et **ccttraining_male** en multipliant les variables **male** et **cct** et **male** et **ccttraining** respectivement. **(2 points)**

1.5. Ajoutez les légendes et labels suivantes aux variables. **(2 points)**

Variables 2005	Légende	Labels
group	Groupe de traitement	1: CCT, 2 : CCT + formation, 3 : Contrôle
cct	CCT	1 : CCT, 0 : Non CCT
ccttraining	CCT + formation	1 : CCT + formation 0 : Non CCT + Formation
control	Groupe contrôle	0 : Traitement, 1 : Contrôle
male	Homme	0 : Féminin, 1 : Masculin
age_transfer	Âge (en mois) au premier transfert	
s2mother_inhs_05	Mère vit dans le foyer	0 : Non, 1 : Oui
ed_mom	Années d'éducation de la mère	
ed_dad	Années d'éducation du père	
tvip_05	Score au test de vocabulaire	
bweight	Poids de naissance	
s1male_head_05	Homme chef de ménage	0 : Non, 1 : Oui
s1hhsz_05	Taille du ménage	
s3awater_access_hh_05	Accès à l'eau courante	0 : Non, 1 : Oui
cons_food_pc_05	Consommation de nourriture par tête	
Variables 2008	Légende	
z_language_08	Score au test de langue (standardisé)	
z_memory_08	Score au test de mémoire (standardisé)	
z_martians_08	Score au test de mémoire associative (standardisé)	
z_social_08	Score au test de compétences interpersonnelles (standardisé)	
z_grmotor_08	Score au test de motricité globale (standardisé)	
z_finmotor_08	Score au test de motricité fine (standardisé)	

2. Statistiques descriptives (10 points)

2.1. Créez un tableau de statistiques descriptives les variables suivantes mesurées avant l'intervention en fonction du groupe de traitement. Pour chaque groupe, le tableau (dont le modèle est fourni ci-dessous) doit contenir la moyenne, l'écart-type et le nombre d'observations. La colonne « p-value (cct vs ccttraining) » est discutée au point 2.2. **(2,5 points).** Les variables suivantes apparaissent dans le tableau :

- *male*
- *age_transfer*
- *s2mother_inhs_05*
- *ed_mom*
- *ed_dad*
- *bweight*
- *s1male_head_05*
- *s1hhsz_05*
- *s3awater_access_hh_05*
- *cons_food_pc_05*

Indiquez, à l'aide d'étoiles à côté de la moyenne de CCT et de CCT + formation, si la moyenne est statistiquement significative par rapport à celle du groupe de contrôle (* 0,1 ; ** 0,05 ; *** 0,01). **(2,5 points)**

	Groupe de contrôle			CCT			CCT + formation			p-value (CCT vs CCT + formation)
	Moy.	Ecart- type	Obs.	Moy.	Ecart- type	Obs.	Moy.	Ecart- type	Obs.	
Variables										
Mère vit dans le foyer										
N° d'années d'éducation de la mère										
N° d'années d'éducation du père										
Homme chef de famille										
Taille du ménage										
Masculin										
Âge (en mois) au premier transfert										
Poids de naissance										
Accès à l'eau courante										
Consommation de nourriture par tête										

- 2.2.** Faites un test pour déterminer si les différences entre le groupe de traitement CCT (**cct=1**) et le groupe de traitement CCT + formation (**ccctraining=1**) sont statistiquement significatives avant l'intervention (2005). Faites ce test pour toutes les variables du tableau et indiquez en dernière colonne du tableau le résultat du test. **(2 points)**

Astuce : Pour faire le test, effectuez une régression incluant les deux variables binaires de traitement et testez l'égalité des coefficients.

- 2.3.** Discutez les résultats de vos statistiques descriptives et partagez vos conclusions sur la validité de la méthode. N'oubliez pas d'utiliser les termes économétriques pour répondre à ces questions ! **(3 points)**

3. Analyse d'impact (15 Points)

Nous voulons à présent estimer l'impact de chacun des deux traitements à l'issue de l'intervention (2008) sur les variables suivantes :

- **z_language_08** (Score au test de langue (standardisé))
- **z_memory_08** (Score au test de mémoire (standardisé))
- **z_martians_08** (Score au test de mémoire associative (standardisé))
- **z_social_08** (Score au test de compétences interpersonnelles (standardisé))
- **z_grmotor_08** (Score au test de motricité globale (standardisé))
- **z_finmotor_08** (Score au test de motricité fine (standardisé))

- 3.1.** Effectuez les régressions suivantes (vous présenterez dans un tableau les coefficients de la régression pour les deux groupes traités, l'écart-type de la moyenne (aussi appelé erreurs-types) et la significativité statistique de chaque coefficient (* 0,1 ; ** 0,05 ; *** 0,01)). Chaque colonne reprend l'estimation de la colonne précédente en y ajoutant les éléments indiqués par une croix. La dernière colonne teste si le coefficient cct est statistiquement égal au coefficient ccctraining.

	(1)		(2)		(3)		
Variable dépendante	CCT	CCT + Formation	CCT	CCT + Formation	CCT	CCT + Formation	P-value test CCT=CCTtraining
Score au test de langue (standardisé)							
Score au test de mémoire (standardisé)							
Score au test de mémoire associative (standardisé)							
Score au test de compétences interpersonnelles (standardisé)							
Score au test de motricité globale (standardisé)							

Score au test de motricité fine (standardisé)							
Variables de contrôle			X	X	X	X	X
Cluster					X	X	X

- **Colonne 1** : Commencez par une régression simple avec les variables **cct** et **ccttraining** comme seules variables explicatives. **(1 point)**
- **Colonne 2** : Effectuez la même régression en ajoutant les variables de contrôle suivantes **(1 point)**:
 - **male** (*Masculin*)
 - **age_transfer** (*Âge (en mois) au premier transfert* (une variable binaire par mois))
 - **s2mother_inhs_05** (*Mère vit dans le foyer*)
 - **ed_mom** (*Années d'éducation de la mère*)
 - **ed_dad** (*Années d'éducation du père*)
 - **bweight** (*Poids de naissance*)
 - **s1male_head_05** (*Homme chef de ménage*)
 - **s1hhsz_05** (*Taille du ménage*)
 - **s3awater_access_hh_05** (*Accès à l'eau courante*)
 - **cons_food_pc_05** (*Consommation de nourriture par tête*)

3.1.1. Pourquoi n'est-il en théorie pas nécessaire de rajouter des variables de contrôle dans le cas d'un RCT (Randomized Controlled Trial) ? **(1 point)**

3.1.2. Quel(s) changement(s) constatez-vous à l'ajout de variables de contrôle et pourquoi ? Si vous observez un changement, à quelle(s) variable(s) est-il dû ? **(2 points)**

Astuce : vous pouvez ajouter les variables de contrôle une par une à votre régression et observer si le coefficient du traitement varie.

3.1.3. QUESTION BONUS (2 points) : Pourquoi n'ajoutons-nous que des variables récoltées avant l'expérience (2005) et non des variables récoltées après le traitement (2008) comme variables de contrôle ?

3.1.4. Citez deux raisons pour lesquelles on ajoute généralement tout de même des variables de contrôle ? **(2 points)**

3.1.5. Imaginez que vous observiez les résultats ci-dessous pour la variable **Accès à l'eau courante** au test d'équilibre de l'enquête initiale, et que nous ne contrôlions pas pour cette variable. L'effet des CCT sur la variable **Score au test de langue (standardisé)** serait-il estimé correctement en estimant la régression de la colonne 3 ? Si non, serait-il surestimé ou sous-estimé ? Référez-vous à la formule du Biais de Variable Omise pour répondre à cette question. **(2 points)**

	Groupe de contrôle			CCT			CCT+ Formation			p-value (CCT vs CCT + Formation)
	Moyenne	Ecart-type	Obs.	Moyenne	Ecart-type	Obs.	Moyenne	Ecart-type	Obs.	
Accès à l'eau courante	6.44	2.85	1066	6.23**	2.64	1006	6.48	3.04	1119	0.04

- **Colonne 3** : Ajoutez un clustering par zone (utilisez la variable « **unique_05** »). (1 point)

3.1.6. Qu'est-ce qu'un cluster et en quoi est-il important d'en tenir compte lors des analyses économétriques ? (1 point)

3.1.7. Quels changements observez-vous dans les résultats ? (1 point)

3.2. Tests d'hypothèse (utiliser la dernière spécification):

Peut-on conclure à un impact causal des programmes de CCT et de CCT + Formation et si oui lequel (n'oubliez pas de commenter les coefficients obtenus)? Peut-on conclure à un impact causal différent du programme de CCT et du programme de CCT + Formation. Si oui lequel est le plus efficace ? (3 points)

4. Hétérogénéité (15 points)

L'impact du programme à l'issue de l'intervention (2008) pourrait varier en fonction du sexe. Nous souhaitons tester cette hypothèse. Pour plus de simplicité, nous nous concentrerons à présent sur le programme de CCT uniquement.

Nous allons utiliser deux stratégies différentes afin de tester l'hypothèse selon laquelle l'effet des CCT sur les variables **z_language_08** et **z_grmotor_08** est différent entre les filles et les garçons.

La **première stratégie** est d'effectuer deux régressions différentes, une pour les filles et une pour les garçons. Nous utilisons pour cela la spécification correspondant à la colonne 3 ci-dessus. Attention cependant à supprimer la variable binaire **male** votre liste de variables de contrôle.

4.1. Présentez un tableau contenant les résultats de ces deux régressions. Ce tableau doit contenir le coefficient de la régression pour le groupe CCT, l'écart-type de la moyenne et la significativité statistique du coefficient (* 0,1 ; ** 0,05 ; *** 0,01). La première colonne concerne les filles et la seconde les garçons. (2 points)

Variables dépendantes	Filles	Garçons
Score au test de langue (standardisé)		
Score au test de motricité globale (standardisé)		

4.1.1. Interprétez les coefficients. (2 points)

4.1.2. Peut-on conclure que l'effet du programme est différent en fonction du sexe ? Testez si la différence entre les deux coefficients est significative. Faites l'hypothèse que la covariance entre les deux est nulle. Écrivez l'hypothèse que vous testez et faites le test manuellement dans Stata. Comparez avec les résultats obtenus sur Stata. (3 points)

Pour faire le test manuellement sur Stata, inspirez-vous de l'exemple suivant (code Stata) :

```
// Exemple pour test manuel sur Stata
```

```
sysuse auto, clear // database contenue dans Stata
reg price mpg if foreign == 0
return list
matrix reg = r(table)
matrix list reg
```

```
scalar mpg0 = reg[1,1]
scalar se0 = reg[2,1]
```

```
scalarlist
```

La **deuxième stratégie** consiste à rajouter un terme d'interaction entre la variable de traitement et la variable de sexe. Nous utilisons toujours la dernière spécification, correspondant à la colonne 3 ci-dessus.¹

4.2. Écrivez l'équation de régression à estimer si l'on utilise cette stratégie. **(1 point)**

4.2.1. Présentez un tableau contenant les résultats de l'estimation (ce tableau doit contenir les coefficients de la régression, les écart-types de la moyenne et la significativité statistique de chaque coefficient). **(2 points)**

Variables explicatives	Score au test de langue (standardisé)	Score au test de motricité globale (standardisé)

4.2.2. Interprétez chacun des coefficients. Peut-on conclure que l'effet du programme est différent en fonction du sexe ? **(3 points)**

4.3. Expliquez quelle est la différence théorique entre ces deux stratégies. Dans quel cas doit-on utiliser l'une ou l'autre ? **(2 points)**

5. Graphiques (10 Points)

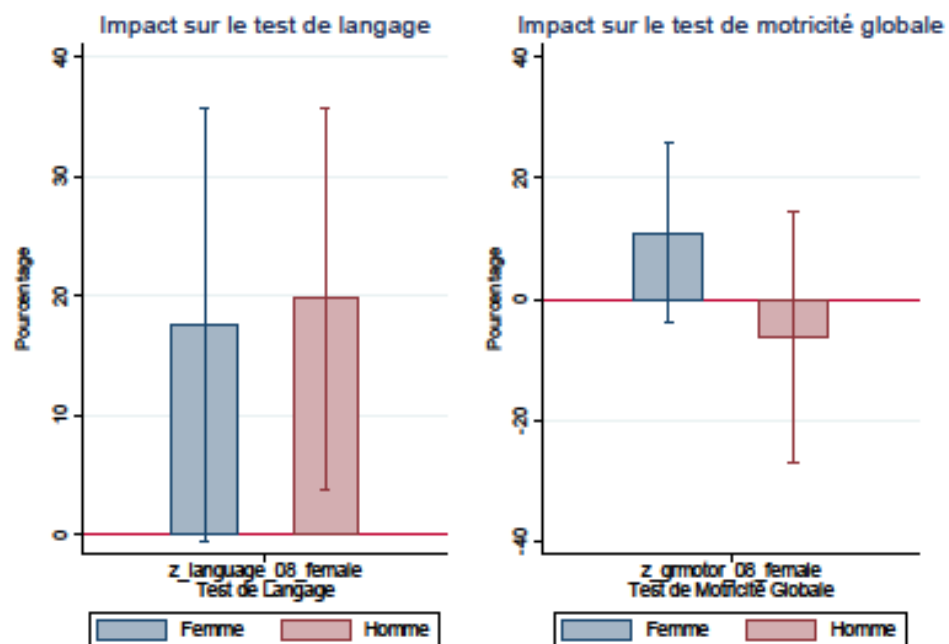
5.1. Faites deux graphiques de type "boîte à moustaches" (similaires à celui ci-dessous) représentant l'effet du programme de **CCT + Formation** sur les variables **z_language_08** et

¹ Une ressource intéressante sur la question des termes d'interaction peut être trouvée à l'adresse suivante : <https://www.theanalysisfactor.com/interpreting-interactions-in-regression/>

z_grmotor_08, l'un pour les filles et l'autre pour les garçons. Utilisez toujours la dernière spécification (colonne 3) **(4 points)**

Astuce : Vous devez faire une régression afin d'obtenir le coefficient de la régression, son intervalle de confiance, puis vous devez enregistrer les résultats de cette régression (commande `eststo`). Finalement, vous pouvez représenter le coefficient de CCT + Formation ainsi que l'intervalle de confiance sur un graphique grâce à la commande `coefplot` (que vous devez préalablement installer grâce à la commande « `ssc install coefplot` »). Utilisez un code du type :

```
Eststo nomdumodele: reg variabledep vardinteret $controls
coefplot nomdumodele, ci keep(vardinteret) autresoptions
```



5.2. Combinez les deux graphiques en un. **(2 points)**

5.3. Comment s'interprète votre graphique ? **(4 points)**