

TD 4 d'analyse d'impact : PSM

Magistère 3

CERDI, Ecole d'Economie, Université Clermont Auvergne

18 novembre 2024

Quelques rappels sur le *Propensity Score Matching*

- ▶ Comme d'habitude, il nous manque un contre-factuel : on tire parti d'une expérience naturelle où certaines unités sont traitées et d'autres non.

Quelques rappels sur le *Propensity Score Matching*

- ▶ Comme d'habitude, il nous manque un contre-factuel : on tire parti d'une expérience naturelle où certaines unités sont traitées et d'autres non.
- ▶ Principe : Constituer un groupe de non-traités similaire au groupe de traités, et le considérer comme un groupe de contrôle viable. **Similaire = en termes de caractéristiques observables seulement !!**

Quelques rappels sur le *Propensity Score Matching*

- ▶ Comme d'habitude, il nous manque un contre-factuel : on tire parti d'une expérience naturelle où certaines unités sont traitées et d'autres non.
- ▶ Principe : Constituer un groupe de non-traités similaire au groupe de traités, et le considérer comme un groupe de contrôle viable. **Similaire = en termes de caractéristiques observables seulement !!**
- ▶ Score de propension :

Quelques rappels sur le *Propensity Score Matching*

- ▶ Comme d'habitude, il nous manque un contre-factuel : on tire parti d'une expérience naturelle où certaines unités sont traitées et d'autres non.
- ▶ Principe : Constituer un groupe de non-traités similaire au groupe de traités, et le considérer comme un groupe de contrôle viable. **Similaire = en termes de caractéristiques observables seulement !!**
- ▶ Score de propension : probabilité conditionnelle d'être traité sachant des caractéristiques observables.

Quelques rappels sur le *Propensity Score Matching*

- ▶ Comme d'habitude, il nous manque un contre-factuel : on tire parti d'une expérience naturelle où certaines unités sont traitées et d'autres non.
- ▶ Principe : Constituer un groupe de non-traités similaire au groupe de traités, et le considérer comme un groupe de contrôle viable. **Similaire = en termes de caractéristiques observables seulement !!**
- ▶ Score de propension : probabilité conditionnelle d'être traité sachant des caractéristiques observables.
- ▶ Condition d'indépendance conditionnelle centrale + *Common support* nécessaire.
- ▶ Deux étapes : estimer le score de propension, puis procéder au *matching* en lui-même.

La question de recherche

La question de recherche

- ▶ Quel sont les effets de la présence de coopératives agricoles sur l'activité des petits producteurs ?

La question de recherche

- ▶ Quel sont les effets de la présence de coopératives agricoles sur l'activité des petits producteurs ?
- ▶ Difficulté majeure de la démarche : mesurer de manière pertinente et surtout **exogène** la présence de coopérative.

La question de recherche

- ▶ Quel sont les effets de la présence de coopératives agricoles sur l'activité des petits producteurs ?
- ▶ Difficulté majeure de la démarche : mesurer de manière pertinente et surtout **exogène** la présence de coopérative.
- ▶ Comment faire ?

La question de recherche

- ▶ Quel sont les effets de la présence de coopératives agricoles sur l'activité des petits producteurs ?
- ▶ Difficulté majeure de la démarche : mesurer de manière pertinente et surtout **exogène** la présence de coopérative.
- ▶ Comment faire ? Les auteurs tirent parti de la création par des partenaires extérieurs (instrument exogène) de coopératives (instrument pertinent) dans certains villages (*kebeles*) uniquement. Intuition → la décision d'implanter une coopérative n'est pas liée aux caractéristiques des *kebeles* ou de ses membres, car, à terme, le gouvernement prévoit d'implanter une coopérative dans chaque *kebele*.

La question de recherche

- ▶ Quel sont les effets de la présence de coopératives agricoles sur l'activité des petits producteurs ?
- ▶ Difficulté majeure de la démarche : mesurer de manière pertinente et surtout **exogène** la présence de coopérative.
- ▶ Comment faire ? Les auteurs tirent parti de la création par des partenaires extérieurs (instrument exogène) de coopératives (instrument pertinent) dans certains villages (*kebeles*) uniquement. Intuition → la décision d'implanter une coopérative n'est pas liée aux caractéristiques des *kebeles* ou de ses membres, car, à terme, le gouvernement prévoit d'implanter une coopérative dans chaque *kebele*.
- ▶ Convaincant ?

Contexte

- ▶ Chaque *kebele* peut être traité (= présence d'une coopérative) ou non. Au sein de chaque *kebele*, chaque producteur peut être traité (= membre de la coopérative) ou non.

Contexte

- ▶ Chaque *kebele* peut être traité (= présence d'une coopérative) ou non. Au sein de chaque *kebele*, chaque producteur peut être traité (= membre de la coopérative) ou non.
- ▶ Il existe donc deux niveaux de sélection: quel *kebele* est traité, et qui est traité au sein de chaque *kebele*.

Contexte

- ▶ Chaque *kebele* peut être traité (= présence d'une coopérative) ou non. Au sein de chaque *kebele*, chaque producteur peut être traité (= membre de la coopérative) ou non.
- ▶ Il existe donc deux niveaux de sélection: quel *kebele* est traité, et qui est traité au sein de chaque *kebele*.
- ▶ Donc, *matching* en deux temps: *kebeles* traités matchés avec *kebeles* non-traités, puis, membres traités d'un *kebele* traité matchés avec membres d'un *kebele* non-traité.

Identification des biais potentiels et réponses

Les participants aux coopératives diffèrent des non-participants en termes de caractéristiques individuelles et communautaires observables.

Identification des biais potentiels et réponses

Les participants aux coopératives diffèrent des non-participants en termes de caractéristiques individuelles et communautaires observables.

Réponse : matching des ménages en fonction des données observables (on ne compare entre eux que des individus avec des observables similaires).

La participation de certains agriculteurs aux coopératives a des externalités sur les non-participants du même *kebele*.

Identification des biais potentiels et réponses

Les participants aux coopératives diffèrent des non-participants en termes de caractéristiques individuelles et communautaires observables.

Réponse : matching des ménages en fonction des données observables (on ne compare entre eux que des individus avec des observables similaires).

La participation de certains agriculteurs aux coopératives a des externalités sur les non-participants du même *kebele*.

Réponse : on matche des ménages similaires en termes d'observables, mais provenant de *kebeles* différents, et opérant sur des marchés locaux différents.

Identification des biais potentiels et réponses, suite

Les participants aux coopératives diffèrent des non-participants en termes de caractéristiques individuelles et communautaires inobservables.

Contexte : L'implantation de coopératives dans chaque *kebele* impulsée à 63% par le gouvernement, 11% par des ONG et 26% par les membres eux-mêmes.

Identification des biais potentiels et réponses, suite

Les participants aux coopératives diffèrent des non-participants en termes de caractéristiques individuelles et communautaires inobservables.

Contexte : L'implantation de coopératives dans chaque *kebele* impulsée à 63% par le gouvernement, 11% par des ONG et 26% par les membres eux-mêmes.

Réponse : Exclusion des *kebeles* initiés par les membres.
L'implantation des *kebeles* est alors considérée exogène.

Est-ce que cela est convaincant ? Et surtout, est-ce que cela suffit ?

Matching des kebeles

- ▶ Echantillon de départ:

Matching des kebeles

- ▶ Echantillon de départ: 293 *kebeles* dans l'échantillon, dont 94 traités (i.e. avec une coopérative au moment de l'enquête).

Matching des *kebeles*

- ▶ Echantillon de départ: 293 *kebeles* dans l'échantillon, dont 94 traités (i.e. avec une coopérative au moment de l'enquête).
- ▶ L'implantation des coopératives dans les *kebeles* doit être exogène :

Matching des *kebeles*

- ▶ Echantillon de départ: 293 *kebeles* dans l'échantillon, dont 94 traités (i.e. avec une coopérative au moment de l'enquête).
- ▶ L'implantation des coopératives dans les *kebeles* doit être exogène : exclusion des *kebeles* traités dont la coopérative a été créée en interne (directement par la communauté), car considérés non-exogènes.
- ▶ On veut des *kebeles* différents (i.e. opérant sur des marchés locaux distincts) pour éviter les externalités :

Matching des *kebeles*

- ▶ Echantillon de départ: 293 *kebeles* dans l'échantillon, dont 94 traités (i.e. avec une coopérative au moment de l'enquête).
- ▶ L'implantation des coopératives dans les *kebeles* doit être exogène : exclusion des *kebeles* traités dont la coopérative a été créée en interne (directement par la communauté), car considérés non-exogènes.
- ▶ On veut des *kebeles* différents (i.e. opérant sur des marchés locaux distincts) pour éviter les externalités : exclusion des *kebeles* non traités dans lesquels les ménages ont accès à une coopérative dans un *kebele* proche.

Matching des *kebeles*, suite

- ▶ On veut des niveaux de développement similaires entre *kebeles*, pour qu'ils soient comparables :

Matching des *kebeles*, suite

- ▶ On veut des niveaux de développement similaires entre *kebeles*, pour qu'ils soient comparables : *matching* au sein de chaque *development domain*. Seuls 5 sur 22 comptent assez de traités pour procéder au matching : exclusion des autres.
- ▶ Trois régions ne comptent pas assez de *kebeles* traités ou témoins, et sont donc exclues de l'échantillon.
- ▶ Échantillon final :

Matching des *kebeles*, suite

- ▶ On veut des niveaux de développement similaires entre *kebeles*, pour qu'ils soient comparables : *matching* au sein de chaque *development domain*. Seuls 5 sur 22 comptent assez de traités pour procéder au matching : exclusion des autres.
- ▶ Trois régions ne comptent pas assez de *kebeles* traités ou témoins, et sont donc exclues de l'échantillon.
- ▶ Échantillon final : 35 *kebeles* traités, 73 non traités.

Treatment and comparison *kebeles*, by development domains

Domain	% Comparison <i>kebeles</i>	% Treatment <i>kebeles</i>
1 Highland, moisture reliable, high market access, high population density	8.21	13.24
2 Highland, moisture reliable, high market access, medium population density	23.13	20.59
3 Highland, moisture reliable, high market access, low population density	2.24	4.41
4 Highland, moisture reliable, low market access, high population density	4.48	0.00
5 Highland, moisture reliable, low market access, medium population density	19.40	10.29
6 Highland, moisture reliable, low market access, low population density	2.99	0.00
7 Lowland, moisture reliable, high market access, medium population density	3.73	1.47
8 Lowland, moisture reliable, high market access, low population density	0.75	0.00
9 Lowland, moisture reliable, low market access, medium population density	1.49	2.94
10 Lowland, moisture reliable, low market access, low population density	5.22	0.00
11 Highland, drought prone, high market access, high population density	1.49	1.47
12 Highland, drought prone, high market access, medium population density	2.99	13.24
13 Highland, drought prone, high market access, low population density	1.49	1.47
14 Highland, drought prone, low market access, high population density	1.49	1.47
15 Highland, drought prone, low market access, medium population density	2.99	14.71
16 Highland, drought prone, low market access, low population density	2.24	2.94
17 Lowland, drought prone, high market access, high population density	1.49	0.00
18 Lowland, drought prone, high market access, medium population density	0.75	2.94
19 Lowland, drought prone, high market access, low population density	2.99	1.47
20 Lowland, drought prone, low market access, medium population density	2.24	2.94
21 Lowland, drought prone, low market access, low population density	5.22	2.94
22 Lowland, pastoralist, high market access, low population density	2.99	1.47

Table 2

Treatment and comparison *kebeles*, distribution by region

	Tigray	Amhara	Oromia	Beneshangul- Gumuz	SNNP	Harari	Total
Comparison <i>kebeles</i>	1	18	33	1	22	1	76
Treatment <i>kebeles</i>	14	9	20	0	6	0	49

Table 3

Balancing tests: Treatment and comparison *kebeles*

	Sample with all regions			Sample with Amhara, Oromia, and SNNP only		
	Comparison <i>kebeles</i>	Treatment <i>kebeles</i>	Difference: <i>P > t</i>	Comparison <i>kebeles</i>	Treatment <i>kebeles</i>	Difference: <i>P > t</i>
Population	4,548.29	5,359.72	0.0966	4,526.18	4,622.91	0.8550
% Female-headed hh	14.01	21.52	0.0116	13.98	18.60	0.1607
% Households Orthodox	45.97	57.77	0.1725	45.08	43.06	0.8298
% Households Muslim	38.09	26.21	0.2208	37.44	33.70	0.7306
% Households speak Amharic	50.95	48.15	0.7010	52.90	61.28	0.2875
Existence commercial bank	9.21	12.24	0.5910	9.58	17.14	0.2630
Existence microfinance institution	21.05	46.93	0.0021	19.17	31.42	0.1607
Importance of traditional institutions*	46.64	46.16	0.9193	46.50	50.91	0.4053
Number of Das	1.83	2.52	0.0023	1.84	2.05	0.3444
Productive safety net <i>woreda</i> **	27.63	38.77	0.1952	26.02	28.57	0.7825
Existence of primary school	86.84	91.83	0.3914	86.30	88.57	0.7452
Direct access to seasonal/dry road	55.26	69.38	0.1161	56.16	60.00	0.7019
Access to safe water	44.73	67.34	0.0131	45.20	57.14	0.2495
Number obs.	76	49		73	35	

Note: Bold *P*-values indicate differences significant at the 10% level or lower.

*Percentage of conflicts resolved through Shimagle (council of elders) as opposed to local courts.

**The Productive Safety Net Program (PSNP) targets to provide food or cash through public work and direct support to 8.3 millions chronically food insecure individuals in 268 *woredas*.

Matching des ménages

- ▶ Le *matching* est fait entre ménages traités (i.e. membres d'une coopérative) des *kebeles* traités (i.e. comportant une coopérative) et ménages (*de facto* non-traités) des *kebeles* non-traités en fonction du score de propension, et au sein d'un même *development domain*.
- ▶ 150 ménages traités, 1702 non-traités.
- ▶ Deux méthodologies : *nonparametric Kernel regression matching* et *five nearest neighbors matching*.
- ▶ Création du score de propension : régression des caractéristiques des ménages sur le fait d'être membre ou non sur le groupe traité seulement (où le choix d'être membre existe). Pour une fois, on cherche à maximiser le R2.
- ▶ Restriction de l'échantillon à la *Common support region* : l'intervalle de score de propension pour lequel à la fois des traités et des non-traités peuvent être trouvés.

Percentage of farmers

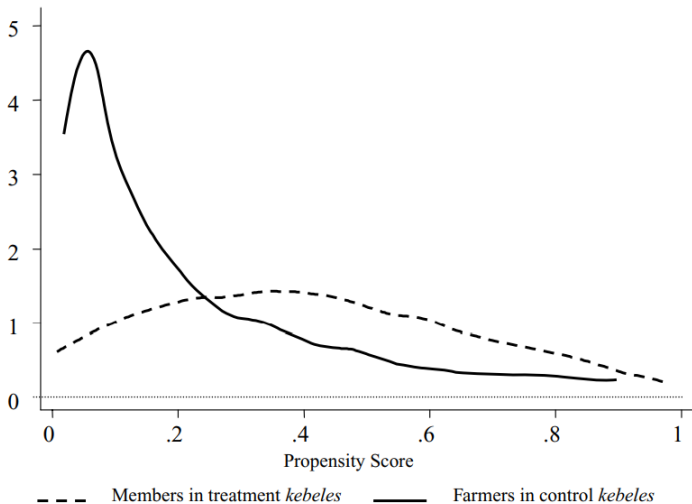


Fig. 2. Propensity scores distribution among treatment and control groups.

Table 6
Balancing tests of matched samples

	(1) Unmatched samples			(2) Kernel-based matching			(3) 5 nearest neighbors matching		
	Treatment <i>kebeles</i>	Comparison <i>kebeles</i>	Diff: <i>P</i> -value	Treatment <i>kebeles</i>	Comparison <i>kebeles</i>	Diff: <i>P</i> -value	Treatment <i>kebeles</i>	Comparison <i>kebeles</i>	Diff: <i>P</i> -value
Age of household head	47.79	42.82	0.000	47.60	47.79	0.915	47.60	48.55	0.608
Gender of household head (1 = male, 2 = female)	1.10	1.17	0.015	1.10	1.06	0.336	1.10	1.05	0.185
Household head reads (1 = yes, 2 = no)	39.59	30.69	0.025	40.00	36.43	0.541	40.00	32.71	0.207
Household size	6.04	5.19	0.000	5.92	5.84	0.768	5.92	5.74	0.513
Radio (1 = yes, 2 = no)	1.40	1.21	0.000	1.37	1.31	0.343	1.37	1.32	0.368
Nonfarm income (1 = yes, 2 = no)	1.48	1.53	0.262	1.48	1.49	0.920	1.48	1.49	0.868
Land owned (hectares)	2.10	1.34	0.000	1.95	2.07	0.449	1.95	2.06	0.475
Oxen (number)	1.63	0.90	0.000	1.47	1.38	0.553	1.47	1.38	0.583
Cattle (number)	5.16	3.50	0.000	4.80	4.55	0.593	4.80	4.54	0.587
Ruminant (number)	3.42	2.44	0.003	2.84	3.73	0.089	2.84	3.70	0.075
Poultry (number)	3.66	2.11	0.000	3.60	3.39	0.688	3.60	3.41	0.715
Cereal production (in kg)	1,148.9	650.94	0.000	1,073.00	896.39	0.135	1,073.0	897.6	0.115

Note: Bold *P*-values indicate differences significant at a 10% level or lower. Samples limited to common support region.

Limites 1

- ▶ L'implantation de coopératives dans chaque *kebele* est considérée exogène, mais on peut penser que cela n'est pas le cas. Dans quel cas est-on si les ONG implantent des coopératives dans les *kebeles* les plus dans le besoin ? Et au contraire, si le gouvernement fait cela dans les *kebeles* les plus dynamiques ?

Limites 1

- ▶ L'implantation de coopératives dans chaque *kebele* est considérée exogène, mais on peut penser que cela n'est pas le cas. Dans quel cas est-on si les ONG implantent des coopératives dans les *kebeles* les plus dans le besoin ? Et au contraire, si le gouvernement fait cela dans les *kebeles* les plus dynamiques ?
- ▶ Que se passe-t-il si les individus qui sont membres des coopératives ont des caractéristiques inobservables différentes des non-membres, par exemple :
 - (i) un esprit entrepreneurial plus développé ?
 - (ii) une aversion au risque plus élevée ?

Limites 2

- ▶ Les coopératives permettent aux participants d'obtenir des prix de vente plus haut. Les auteurs mentionnent un biais en cas de comparaison entre traités et non-traités intra-*kebeles* traités. Ce biais est-il une menace ou un *lower bound* si les prix pour les non-membres augmentent aussi ? S'ils diminuent ?

Limites 2

- ▶ Les coopératives permettent aux participants d'obtenir des prix de vente plus haut. Les auteurs mentionnent un biais en cas de comparaison entre traités et non-traités intra-*kebeles* traités. Ce biais est-il une menace ou un *lower bound* si les prix pour les non-membres augmentent aussi ? S'ils diminuent ?

- ▶ Que peut-on dire de la taille de l'échantillon ?

Limites 2

- ▶ Les coopératives permettent aux participants d'obtenir des prix de vente plus haut. Les auteurs mentionnent un biais en cas de comparaison entre traités et non-traités intra-*kebeles* traités. Ce biais est-il une menace ou un *lower bound* si les prix pour les non-membres augmentent aussi ? S'ils diminuent ?
- ▶ Que peut-on dire de la taille de l'échantillon ?
- ▶ Que peut-on dire de la variation des résultats lorsqu'on prend en compte les régions avec peu de *kebeles* traités ? Est-ce normal ?

Limites 2

- ▶ Les coopératives permettent aux participants d'obtenir des prix de vente plus haut. Les auteurs mentionnent un biais en cas de comparaison entre traités et non-traités intra-*kebeles* traités. Ce biais est-il une menace ou un *lower bound* si les prix pour les non-membres augmentent aussi ? S'ils diminuent ?
- ▶ Que peut-on dire de la taille de l'échantillon ?
- ▶ Que peut-on dire de la variation des résultats lorsqu'on prend en compte les régions avec peu de *kebeles* traités ? Est-ce normal ?
- ▶ Quel aurait pu être un groupe de traités plus exogène et convaincant?

On ouvre Stata

Table 7
Effect of cooperatives on members' cereals commercialization

	Kernel-based matching		Five nearest neighbors matching		Number of observations
	ATT	Std. error	ATT	Std. error	
A. Two-step matching, final sample					
% Price difference (PD)	7.249	3.229**	8.901	4.361**	862
% Production sold (PS)	-0.122	1.862	-1.116	2.202	1,817
B. Checks of robustness					
1. Two-step matching, final sample limited to Oromia region only					
% PD	8.545	3.952**	7.141	6.373	454
% PS	-2.881	3.671	-2.914	3.141	787
2. Two-step matching, final sample excluding <i>kebeles</i> with NGO-created cooperatives					
% PD	5.567	3.674	6.196	4.366	861
% PS	0.761	1.876	-1.402	2.442	1,805
3. Two-step matching, final sample, with propensity scores estimated on whole sample					
% PD	7.206	3.229**	7.562	3.771**	862
% PS	-1.180	1.964	-3.224	2.550	1,817
C. Checks of representativeness					
1. Two-step matching, sample with all regions					
% PD	11.451	2.060***	10.607	2.765***	1,449
% PS	-1.174	1.147	-0.944	1.340	2,993
2. Two-step matching, sample with member-created cooperatives					
% PD	9.294	1.831***	8.626	2.327***	1,471
% PS	0.005	1.07	-0.084	1.325	2,972
3. Two-step matching, sample with all regions and member-created cooperatives					
% PD	13.120	1.923***	12.118	2.206***	1,504
% PS	-1.380	0.923	-1.295	1.093	3,105
4. One-step matching, final sample within treatment <i>kebeles</i> only					
% PD	7.606	3.980*	7.259	3.571**	368
% PS	-1.083	1.811	-1.410	2.135	822

Note: Stratified bootstrap with 100 replications are used to estimate the standard errors.

***Significant at 1% level, **significant at 5% level, and *significant at 10% level.