

**L'EXPÉRIMENTATION TZCLD ET L'ENQUÊTE « QUALITÉ DE VIE DANS LES TERRITOIRES »
L'INTRODUCTION D'UNE VARIABLE ENDOGÈNE INOBSERVABLE
MODIFIE-T-ELLE LES RÉSULTATS ?**

Félix BONNETÊTE (*)

(*) Dares, Sous-direction Suivi et Évaluation des Politiques de l'Emploi et de la Formation professionnelle

felix.bonnetete@travail.gouv.fr

Mots-clés : Évaluation, Enquête, Matching, Tests de robustesse

Domaine concerné : Évaluation des Politiques Publiques, Algorithmes

Résumé

Entrée en vigueur en janvier 2017 dans dix territoires de France métropolitaine, l'expérimentation « Territoires Zéro Chômeurs de Longue Durée » se donne pour objectifs d'éradiquer le chômage de longue durée en embauchant en CDI dans des structures dédiées (entreprises à but d'emploi, EBE) d'anciennes personnes privées durablement d'emploi (dénommées ci-après « bénéficiaires »).

La Dares, pour le compte du comité scientifique chargé de l'évaluation de cette expérimentation, a mené l'enquête « Qualité de vie dans les territoires » (QVT) en deux vagues sur les dix territoires pilotes de l'expérimentation et des territoires témoins qu'elle a sélectionnés afin de déterminer l'impact de l'expérimentation sur les bénéficiaires.

L'estimation de plusieurs modèles d'appariements sur score de propension a fait ressortir un effet positif de l'expérimentation sur l'insertion en emploi des bénéficiaires (+36,8 points) ainsi que sur leur santé (+9,0 points) et leur bien-être (jusqu'à +17,2 points), par comparaison à des faux-jumeaux constitués de personnes éligibles (individus présentant des caractéristiques susceptibles de les rendre éligibles à une embauche en EBE) dans des territoires témoins.

La crédibilité de ces résultats dépend très fortement de la validité de l'hypothèse d'indépendance conditionnelle : une fois que toutes les variables de contrôle observables possible ont été utilisées lors du calcul du score de propension, la sélection dans le traitement ne dépend pas des résultats attendus de ce dernier pour un individu donné. Ici, on cherche à tester la robustesse des résultats présentés plus haut en introduisant de manière artificielle une variable inobservable et on détermine dans quelle mesure sa prise en compte modifie les résultats des matchings.

La démarche proposée ici est inspirée de la méthode développée par Rosenbaum et Rubin en 1982, qui cherchent à répondre aux critiques faites aux résultats d'expériences cliniques non randomisées selon lesquelles les patients à plus haut risque peuvent être surreprésentés dans certains groupes de traitement. Les auteurs mettent en place une technique évaluant la sensibilité des conclusions aux hypothèses concernant une covariable binaire non observée et pertinente pour l'affectation du traitement, via une étude médicale sur la maladie coronarienne.

Dans le cadre de l'évaluation de l'expérimentation TZCLD, un exemple d'une telle variable inobservée serait la motivation des individus comparés dans l'évaluation : elle serait corrélée positivement avec le fait d'être bénéficiaire, car cette caractéristique est déterminante dans l'entrée dans l'expérimentation et elle a de bonnes chances d'être corrélée positivement avec le fait d'être employé (ici la variable d'intérêt) : on peut supposer que toutes choses égales par ailleurs, le fait d'être motivé accroît les chances d'accéder à un emploi.

Un algorithme est ensuite implémenté pour simuler cette variable inobservable (appelée **motiv**) selon différents niveaux de corrélation avec la probabilité d'entrée dans le dispositif d'une part, et celle d'accéder à un emploi d'autre part. Dans le détail, cette variable est simulée et prise en compte de la manière suivante :

- Sélection d'un couple de facteurs $(x;y)$, où x est l'intensité de la corrélation entre le fait d'être bénéficiaire et l'inobservable, et y l'intensité de la corrélation entre le fait d'être en emploi et l'inobservable.
- Pour un couple $(x;y)$ donné, création d'une variable **seuil** égale à 0,8.
 - o Si l'individu n'est pas bénéficiaire alors **seuil**=**seuil**/ x , sinon **seuil**=**seuil** ;
 - o Si l'individu n'est pas en emploi alors **seuil**=**seuil**/ y , sinon **seuil**=**seuil**.
- Création d'une variable u suivant une loi uniforme entre 0 et 1.
- Création de la variable inobservable **motiv** de manière binaire : si **seuil**> u alors **motiv**=1, sinon **motiv**=0.
- Calcul du score de propension sur la base des variables de contrôle (incluant **motiv**) et appariement.
- Calcul de la différence pondérée moyenne entre bénéficiaires et faux-jumeaux sur la variable d'intérêt.
- Pour un couple $(x;y)$ donné, u est simulée 120 fois, on aboutit donc à 120 différences pondérées moyennes.
- Il est alors possible d'aboutir, pour chaque couple $(x;y)$, à un coefficient moyen et à un intervalle de confiance à 90 %.

Pour la variable « être en emploi », les conclusions sont les suivantes : plus l'intensité de la corrélation entre le fait d'être un bénéficiaire et l'inobservable (facteur x) et celle entre le fait d'être en emploi et l'inobservable (facteur y) augmentent, plus la valeur du coefficient diminue. Ainsi, on réduit la valeur du coefficient initialement calculé en lui retirant la part liée à l'inobservable ici introduite. Cependant, ce coefficient semble finir par atteindre un palier. En outre, quelle que soit l'ampleur de la corrélation entre l'inobservable et le fait d'être bénéficiaire et/ou d'être en emploi, les coefficients sont toujours significatifs à 10 %. Ceci peut laisser penser que l'introduction d'une variable inobservable ne **discrédite pas la validité globale des résultats sur l'emploi**.

Le même travail a été effectué mais cette fois-ci en choisissant une variable de bien-être (« ne pas avoir de mal à se concentrer ») comme variable d'intérêt. L'effet observé est moins

net : au fur et à mesure que les niveaux de corrélation augmentent, le coefficient diminue plus vite et les intervalles de confiance deviennent non significatifs plus rapidement. Cependant, un effet positif et significatif à 10% est toujours observé pour des niveaux de corrélation élevés, l'analyse ne semble donc pas perturbée outre mesure par l'introduction d'une inobservable.

Bibliographie

- [1] Rosenbaum and Rubin « Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome», *Journal of the Royal Statistical Society. Series B*, vol.45 n°2, 1983.
- [2] Joshua Angrist and Jorn-Steffen Pischke, « Mostly harmless econometrics », *Princeton Libri*, 2009.
- [3] Pauline Givord, « Méthodes économétriques pour l'évaluation des politiques publiques », *Insee Documents de travail*, 2010.
- [4] Simon Bunel, Clémence Lenoir, and Simon Quantin, « Evaluation du dispositif Jeune entreprise innovante (JEI), un exemple d'application du modèle d'analyse de sensibilité de Rosenbaum », 2021.