

Appariement selon le score de propension pour l'évaluation de programmes

Programme Service de croissance accélérée

Atelier sur l'évaluation quantitative des impacts

Mahamat Hamit-Haggar

Centre des Projets Spéciaux sur les Entreprises

9 Février 2026



Éclairer grâce aux données, pour bâtir un Canada meilleur



Statistics
Canada

Statistique
Canada

Canada

Apeçu

- ▶ Aperçu de l'appariement selon le score de propension
- ▶ Programme Service de croissance accélérée
- ▶ Utilisation de l'Environnement de fichiers couplables – Entreprises (EFC-E) et Soutien de la croissance et de l'innovation en entreprise (SCIE) pour l'analyse d'impact



Qu'est ce que l'Appariement selon le score de propension ?

Une méthode développée par Rosenbaum et Rubin (1983)

Conçue pour être utilisée dans des études avec observations

Appliquée lorsque l'affectation aléatoire n'est pas possible

- ▶ Évaluation de politique et programme
- ▶ Analyse de l'impact économique
- ▶ Recherche en santé, travail, éducation et sciences sociales
- ▶ Situations où les essais contrôlés aléatoires ne sont pas réalisables

Réduit ou élimine le biais de sélection des échantillons lors des comparaisons entre groupes (par exemple, traitement/contrôle) sur la base de caractéristiques **observables**.



Appariement selon le score de propension | Méthodologie

Supposons que nous avons un groupe de traitement et un groupe de contrôle

Le status de traitement D est une variable binaire :

$D = 1$: observations traitées

$D = 0$: observations de contrôle

Qu'est ce que le score de propension ?

Le score de propension est la probabilité prédite de recevoir un traitement, conditionnelle aux covariables observées pré-traitement x

$$p(x) = \text{prob}(D = 1|x) = E(D|x)$$

Estimation du score de propension

Estimé grâce à un modèle probit/logit

D est la variable dépendante

x sont les covariables qui affectent à la fois l'attribution de traitement et le résultat



Appariement selon le score de propension | Méthodes d'Appariement

Étape 1: Obtenir la probabilité prédite de recevoir le traitement (le score de propension) pour chaque observation

Étape 2: Choisir une méthode d'Appariement

Approches courantes d'Appariement

- ▶ Appariement selon la correspondance du plus proche voisin
- ▶ Appariement Kernel
- ▶ Appariement par calibres (rayons)
- ▶ Appariement par stratification
- ▶ Appariement selon la distance Mahalanobis
- ▶ Appariement génétique



Méthode d'Appariement | Plus proche voisin

Pour chaque observation traitée i , une observation de contrôle j est sélectionnée, puisqu'elle est la plus proche de x

$$\min \|p_i - p_j\|$$

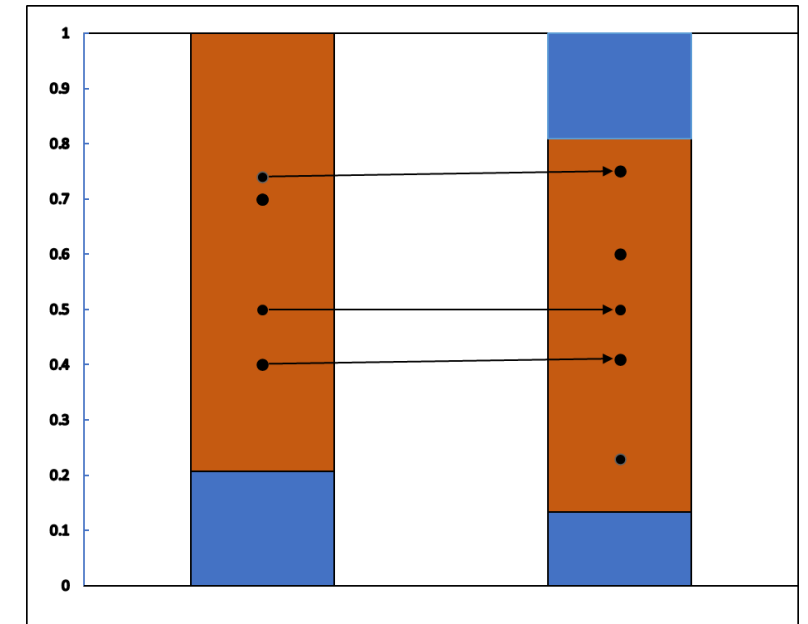
Appariement avec ou sans remplacement

Sans remplacement

- Chaque observation de contrôle est utilisée au maximum une fois comme correspondance pour une observation de traitement

Avec remplacement

- Une observation de contrôle peut être appariée avec plusieurs observations traitées (de traitement)



L'appariement est limité aux observations situées dans la plage commune des scores de propension.

Méthode d'Appariement | Appariement de Kernel

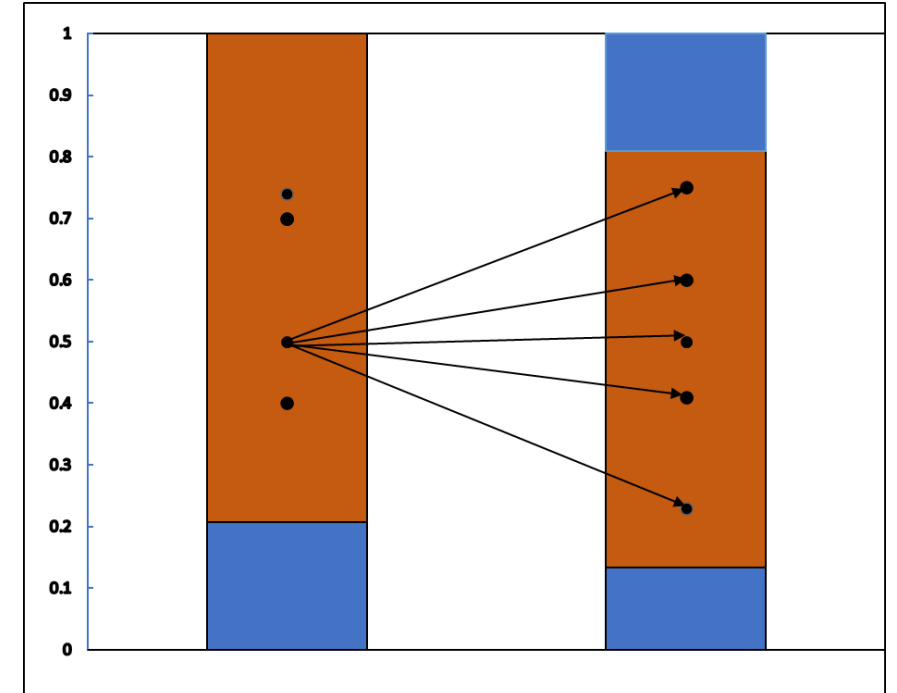
Chaque observation traitées i est appariée avec une multitude d'observations de contrôle

Les observations de contrôle sont pondérées inversement à la distance entre les observations traitées et les observations de contrôle.

$$W(i,j) = \frac{k\left(\frac{p_j - p_i}{h}\right)}{\sum_{j=1}^{n_0} k\left(\frac{p_j - p_i}{h}\right)}$$

Le paramètre de Bandwidth

h est le paramètre de Bandwidth



Comparer les résultats | Effet du traitement moyen sur les traités (ETMT)

ETMT mesure la différence entre les résultats des observations traitées et les résultats des observations traitées si elles n'avaient pas été traitées

$$ATET = E(y_1|x, D = 1) - E(y_0|x, D = 1)$$

$E(y_0|x, D = 1)$ est un élément *contrefactuel* (non directement observable) et qui doit être estimé.

$$ATET = E(y_1|p(x,), D = 1) - E(y_0|p(x), D = 0)$$

Après avoir apparié les scores de propension et vérifié le balancement, les résultats des groupes traités et de contrôle sont comparés.

Un modèle de différence en différences est appliqué lorsque des données de panel sur les résultats sont disponibles (Wooldridge, 2010).

$$ATET = E(y_{1a} - y_{1b}|x, D = 1) - E(y_{0a} - y_{0b}|x, D = 1)$$



Application de l'Appariement selon le score de propension | Programme service de croissance accélérée (SCA)

N° 18-001-X au catalogue
ISBN x

Rapports sur les projets spéciaux sur les entreprises

Évaluation de l'incidence économique du programme Service de croissance accélérée chez les entreprises participantes, 2017 à 2019

par Mahamat Hamit-Haggar

Date de diffusion : le 7 octobre 2024



Canada

Évaluation commandée des services consultatifs du programme SCA

Objectif: Évaluer si le soutien du SCA conduit à des résultats commerciaux mesurables.

- ▶ Performance financière
- ▶ Activité d'exportation
- ▶ Résilience des entreprises

Données fournies:

1000 entreprises ayant bénéficié du soutien SCA

Période de référence de 2017 à 2023



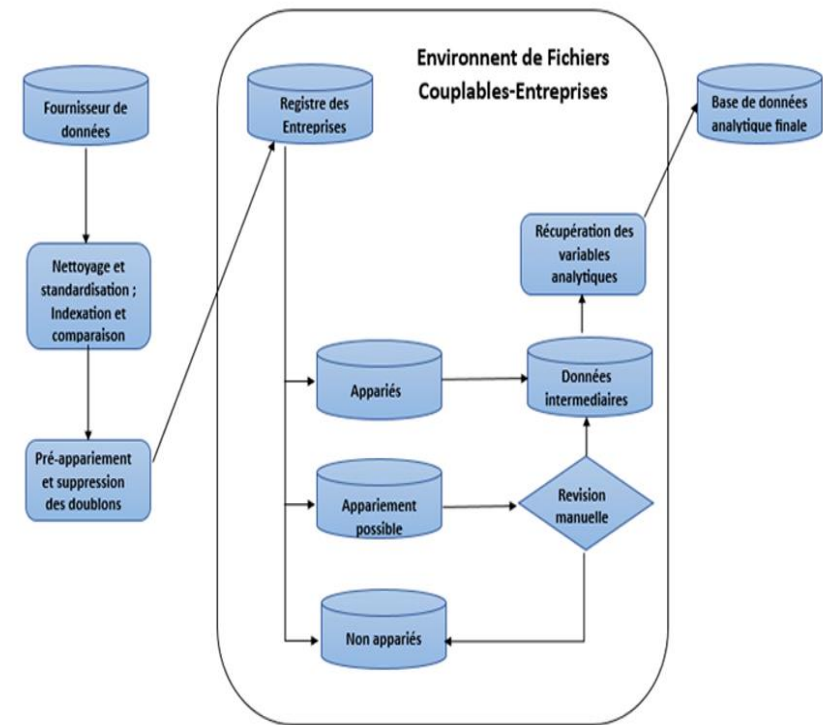
Appariement des clients du SCA avec le Registre des entreprise de Statistique Canada

Taux d'Appariement : 99%

Appariement avec l'Environnement de fichiers couplables – Entreprises (EFC-E)

Registre des entreprises de Statistique Canada	2014 to 2023
Relevé de compte de retenues à la source courantes (RCRSC)	2014 to 2023
Index général des renseignements financiers (IGRF)	2014 to 2023
Enquête annuelle sur la recherche et le développement dans l'industrie canadienne (RDIC)	2014 to 2021
Registre des exportateurs (XPTR)	2014 to 2023
Entreprises appariées avec la base de données IGRF	99%
Entreprises appariées avec la base de données RCRSC	90%

Processus général de couplage de fichiers avec le Registre des entreprises



Source : Adapté de Christen et al., 2012

Approche analytique

1. Exécuter un modèle probit pour obtenir la probabilité prédite du traitement, sur la base d'un ensemble de covariables pré-traitement.
2. Examiner la région commune de soutien.
3. Appliquer l'appariement selon la méthode du plus proche voisin.
4. Évaluer le balancement (équilibre) des covariables dans l'échantillon apparié.
5. Estimer les effets du traitement.

Logiciels et outils

L'appariement selon le score de propension est effectué à l'aide du **package R MatchIt** (Stuart et al, 2011)



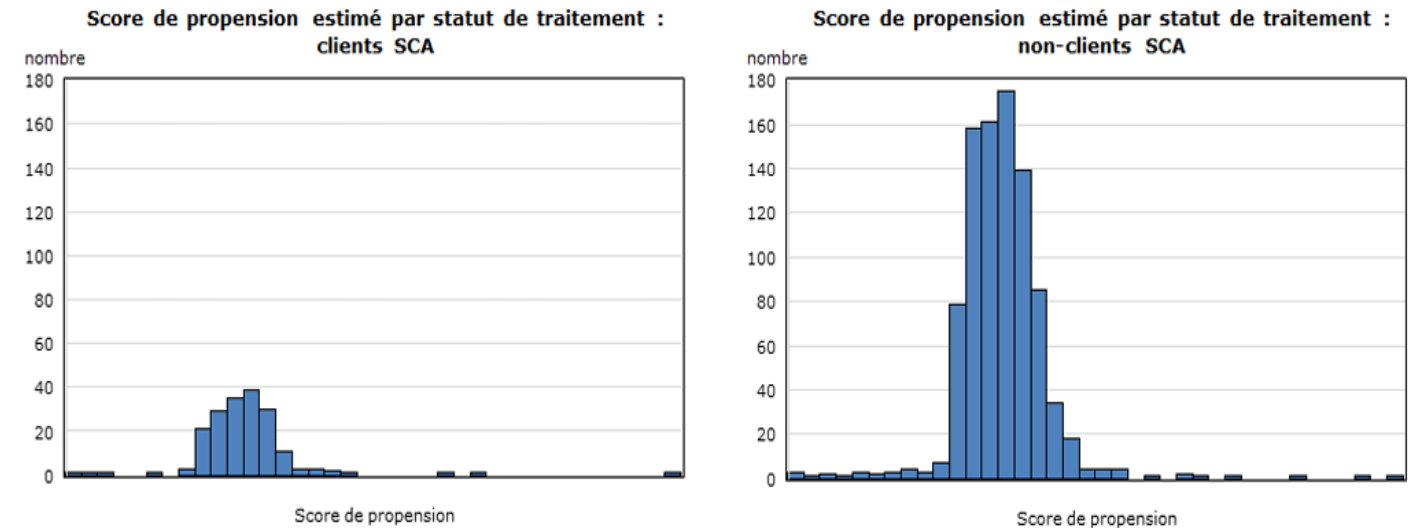
Résultats empiriques | Distribution des scores de propension (cohorte de 2018)

La distribution des scores de propension des entreprises bénéficiant des services SCA recoupe celle du groupe de comparaison potentiel.

Ce recoupement indique la présence **d'une région de soutien commune**.

Identifie les paires d'observations ayant des scores de propension similaires.

Répartition des scores de propension, cohorte de 2018



Source : Calculs de l'auteur.

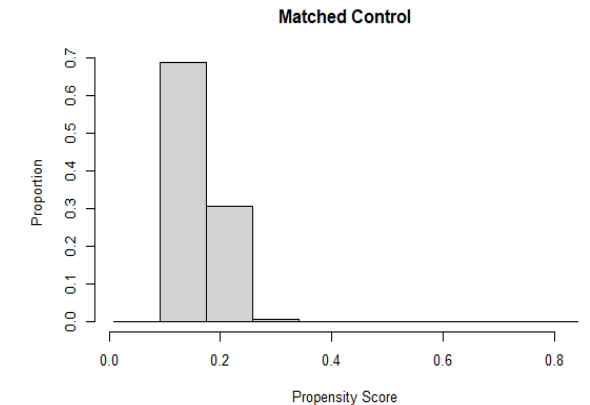
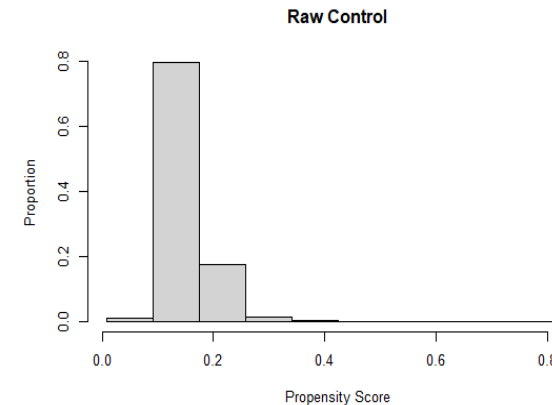
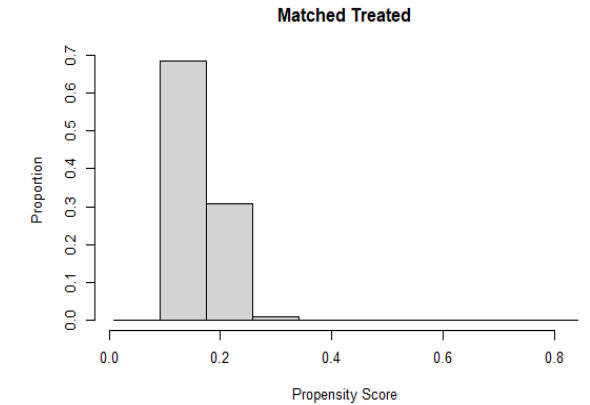
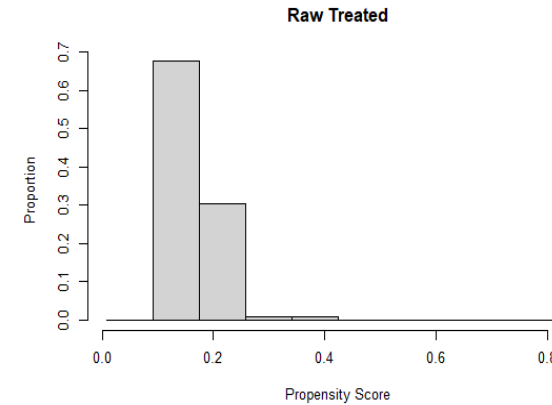
Résultats empiriques | Distribution des scores de propension après l'appariement (cohorte de 2018)

Appariement selon la méthode du plus proche voisin avec calibre est appliqué

Chaque entreprise-cliente du SCA est appariée avec une unique entreprise non-bénéficiaire avec un score de propension similaire

Cette approche combine deux méthodes d'appariement (Cochran & Rubin, 1973):

1. Appariement du plus proche voisin
2. Appariement selon les calibres (rayons)



Résultats empiriques | Balancement des covariables après appariement (cohort de 2018)

Balancement général (Cohorte de 2018)

	Balancement général pour tous les enregistrements			Balancement général pour les enregistrements appariés		
	Moyennes des traités	Moyennes des contrôlés	Différences moyennes type	Moyennes des traités	Moyennes des contrôlés	Différences moyennes type
Distance	0,1750	0,1644	0,1748	0,1692	0,1692	0,0003
Revenu	0,3312	0,2715	0,1064	0,3439	0,3504	-0,0116
Emplois	0,2638	0,2011	0,0939	0,2942	0,2589	0,0529
Profits	0,1091	0,3893	-0,1384	0,3139	0,2768	0,0183
Dépenses RD	0,7809	0,8586	-0,0121	0,8667	0,6756	0,0296
Exportations/Revenu	0,0233	0,0194	0,0166	0,0176	0,0196	-0,0085
Taille des échantillons						
Total	155	778
Appariés	153	153	...
Non-appariés	2	625

... n'ayant pas lieu de figurer

Source : Calculs de l'auteur

Les résultats indiquent qu'un niveau satisfaisant de balancement des covariables a été atteint.

Les différences moyennes standardisées pour les covariables sont inférieures au seuil recommandé de **0,1** (Stuart et al. 2013).

Des différences moyennes standardisées supérieures à 0,1 entraîneraient des effets biaisés.



Résultats empirique | Effets du traitement (cohorte de 2018)

Effets de traitement (cohorte de 2018 premium)

	Effet de traitement moyen	Statistique t
Revenu		
1 an de croissance	3,3590	2,9350
3 ans de croissance	5,2700	2,8560
Emplois		
1 an de croissance	1,2380	1,6400
3 ans de croissance	2,1990	2,3130
Profits		
1 an de croissance	-0,3260	1,6380
3 ans de croissance	0,9580	2,0400
Exportations/Revenu		
1 an de croissance	-0,3330	0,0330
3 ans de croissance	2,2680	1,0560
Dépenses RD		
1 an de croissance	5,7290	2,0900
3 ans de croissance	12,1110	3,7660

Source : Calculs de l'auteur

La croissance du revenu parmi les entreprises-clients du SCA était **supérieur de 3,35 % après un an, et jusqu'à 5,27 % après trois ans** par rapport aux entreprises comparables appariées.

La croissance de l'emploi était **supérieur de 1.23% après un an et de 2.19% après trois ans** pour les entreprises-clients du SCA.

Les dépenses en R&D avaient augmenté parmi les entreprises-clients du SCA, en comparaison avec les entreprises non-bénéficiaires



Résultats empiriques | Principaux points à retenir

Les entreprises-clientes du SCA sont plus enclines à innover.

Les entreprises-clientes du SCA obtiennent de meilleurs résultats en comparaison avec les entreprises non-bénéficiaires pour la plupart des indicateurs de performance.

Les entreprises-clientes du SCA font preuve d'une plus grande résilience face au marché, que les entreprises non-bénéficiaires (résultats non présentés).



Utilisation de EFC-E et SCIE pour l'analyse d'impact

Contexte

En 2017, le budget fédéral a lancé un examen pangouvernemental des programmes de soutien aux entreprises.

En 2022, 5,9 milliards de dollars ont été versés à plus de 39 000 entreprises dans le cadre de 172 programmes fédéraux (*p. ex. le Programme d'aide à la recherche industrielle du Conseil national de recherches (national) et le Programme de développement des entreprises de l'Agence de promotion économique du Canada atlantique (régional)*)

Défi

Les entreprises bénéficient souvent de plusieurs programmes, ce qui rend l'évaluation de l'efficacité des interventions difficile.

La combinaison de programmes nationaux et régionaux aide-t-elle les entreprises à devenir plus compétitives, innovantes et productives ?



Utilisation de EFC-E et SCIE pour l'analyse d'impact

Objectif

Distinguer empiriquement et analyser simultanément les effets des programmes nationaux et régionaux sur l'innovation dans les petites et moyennes entreprises canadiennes.

Méthodologie

Utilisation des base de données d l'EFC-E et du SCIE pour construire l'analyse

Appliquer l'appariement selon le score de propension dans un contexte de traitements multiples
(Gerfin et Lechner 2002, Czarnitzki et Lopes-Bento, 2014 et Guerzoni et Raiteri, 2015)

Pertinence

L'évaluation de l'ensemble des mesures de soutien public permet de déterminer dans quelle mesure les programmes atteignent les objectifs politiques visés.

Les résultats contribuent à orienter les décisions futures en matière de politique et de financement pour le soutien aux entreprises.



Restez branchés!

Application StatsCAN

Balado Hé-coutez bien!

StatsCAN Plus

Le Quotidien

Site Web

Enquêtes et programmes statistiques

Centre de service de données

Mon StatCan



Vous avez des questions? Communiquez avec nous : infostats@statcan.gc.ca

Statistique Canada – Votre organisme national de statistique



Éclairer grâce aux données, pour bâtir un Canada meilleur