



**Institut
canadien
des actuaires**

**Canadian
Institute
of Actuaries**



**CASUALTY
ACTUARIAL
SOCIETY**

Évaluation des effets de la décriminalisation du cannabis sur l'expérience d'accidents automobiles

Vyacheslav Lyubchich

Décembre 2022

Document 222163

Table des matières

Sommaire	3
1 Introduction	4
2 Revue de la littérature	5
Effets du cannabis et de la décriminalisation	5
Vue d'ensemble	6
3 Aperçu des données	6
3.1 Canada	6
3.2 États-Unis	8
4 Méthodes	10
4.1 Comparaisons avant-après	10
4.2 Appariement par score de propension	11
4.3 Régression par forêt aléatoire	12
5 Résultats	13
5.1 Canada	13
5.2 États-Unis	20
6 Bibliographie	23
7 À propos de l'auteur	25



Sommaire

Ce rapport répond à une demande de l'Institut canadien des actuaires (ICA) et de la Casualty Actuarial Society (CAS) d'analyser les effets de la décriminalisation du cannabis sur l'expérience d'accidents automobiles. Une revue de la littérature montre que même si l'intoxication au cannabis influe sur le comportement au volant, ce comportement n'est pas toujours plus dangereux; par exemple, on a signalé que les conducteurs qui avaient les facultés affaiblies enregistraient des vitesses moindres et maintenaient des interdistances plus grandes. Les études observationnelles des accidents de la route font état de résultats mitigés, le plus souvent ne décelant pas d'effets significatifs, particulièrement à long terme.

Ce rapport s'appuie sur les progrès techniques des algorithmes d'apprentissage machine pour trouver des tendances dans les données, sur des méthodes rigoureuses de sélection des unités témoins sur la base de données d'observation et sur les voies nouvellement tracées pour interpréter de façon causale les modèles à boîte noire.

L'étude repose sur les données canadiennes et américaines des années 2016 à 2019, y compris les rapports officiels sur les sinistres et les collisions de voitures de tourisme au Canada, les accidents mortels et les facteurs météorologiques aux États-Unis. Pour chaque source de données, des modèles statistiques et des modèles d'apprentissage machine ont été choisis pour tenir compte des différentes sources de variabilité.

Une analyse de dix régions canadiennes qui tenait compte des différences régionales a permis de modéliser une tendance linéaire de base qui avait également été observée dans les seules données précédant la légalisation. Cette analyse n'a révélé aucun changement statistiquement significatif du coût moyen par réclamation et de la fréquence des réclamations après la légalisation du cannabis au Canada. Les données trimestrielles disponibles pour le Québec ont donné lieu à des constatations semblables.

Aux États-Unis, les résultats variaient selon la paire d'États choisie comme groupe témoin dans les comparaisons statistiques. Les tests des effets de la décriminalisation sur les décès n'ont pas permis de déceler un changement statistiquement significatif. Les techniques d'apprentissage machine ont permis à l'auteur de tenir compte d'autres facteurs, dont la météo et les tendances annuelles et hebdomadaires des décès.

Dans l'ensemble, l'auteur de cette étude a surmonté les inconvénients des recherches antérieures sur les effets de la décriminalisation du cannabis en intégrant de nouvelles méthodes qui ne reposent pas sur la linéarité des relations ou l'inférence paramétrique. L'étude n'a pas permis de détecter d'effets persistants statistiquement significatifs de la décriminalisation.



1 Introduction

Les voitures de tourisme sont essentielles pour un grand nombre de personnes vivant aux États-Unis et au Canada et ont été un symbole de prospérité matérielle selon le rêve américain. Les risques d'assurance automobile ont changé au cours des dernières années en raison des changements législatifs régionaux touchant la consommation de cannabis. La décriminalisation du cannabis dans plusieurs régions a été associée à un plus grand nombre de conducteurs d'automobiles sous l'influence de la drogue (Pollini et coll., 2015) et à une hausse des collisions mortelles (Cook et coll., 2020). Tandis que les débats et les décisions en matière de décriminalisation du cannabis se poursuivent, les compagnies d'assurance doivent évaluer les changements connexes dans l'expérience d'accidents automobiles pour établir le prix du risque et ajuster le coût d'assurance.

Il existe des rapports mitigés sur les implications de la décriminalisation du cannabis (voir par exemple l'aperçu qu'en donnent Hall et Lynskey, 2020) qui pourraient s'expliquer en partie par le fait que les études ne prennent en compte qu'une seule région (limitation des données) ou ne s'appuient que sur une seule méthode statistique. D'autres publications ont examiné le changement relatif sans considérer l'impact global (p. ex., Pollini et coll. ont étudié en 2015 la proportion de conducteurs ayant obtenu un résultat positif au delta-9-tétrahydrocannabinol [THC], mais non les variations dans le nombre ou la gravité des accidents de voiture).

La présente étude propose des outils fiables guidés par les données pour évaluer l'incidence sur l'expérience d'accidents automobiles des nouvelles lois sur le cannabis, en exploitant les données accessibles au public grâce à la puissance des techniques statistiques modernes et d'apprentissage machine dans un cadre intégratif alliant l'analyse d'impact et la quantification de l'incertitude. Le cadre obtenu servira à l'analyse d'autres changements politiques, comme la décriminalisation de la psilocybine ou l'adoption future de véhicules autonomes. L'étude jette un nouveau regard sur les effets de la décriminalisation du cannabis sur l'expérience d'accidents automobiles et sur les implications pour le secteur de l'assurance, en plus de concevoir une approche systématique pour évaluer des changements de ce type dans l'avenir, ce qui permettra aux assureurs de couvrir les risques connexes et d'élaborer des stratégies plus efficaces de tarification des produits d'assurance. Les objectifs de la présente étude sont donc de tirer des conclusions rigoureuses en analysant le plus grand nombre possible de régions du Canada et des États-Unis et en abordant l'analyse au moyen de plusieurs techniques non paramétriques modernes issues des domaines de la statistique et de l'apprentissage machine.

Voici un résumé des principales étapes :

1. Établir des groupes témoins pertinents pour quantifier les effets de la décriminalisation du cannabis. Le groupe témoin pourrait être le même territoire (comparaison avant-après, si des séries chronologiques de qualité de longueur suffisante sont disponibles pour les deux périodes) ou des territoires similaires pour des comparaisons concomitantes.
2. Recueillir des données sur les accidents de voiture, notamment la fréquence, la gravité et le type d'accidents. Recueillir des informations sur les facteurs parasites comme la météo, l'heure du jour et le jour de la semaine (voir par exemple Bailey et coll., 2020) ainsi que l'âge et le sexe du conducteur et l'altération de ses capacités de conduite.
3. Appliquer des techniques statistiques pour quantifier les effets de la décriminalisation du cannabis, tout en contrôlant les facteurs parasites.
4. Élargir l'ensemble de modèles linéaires et paramétriques classiques en tirant parti de la puissance des techniques statistiques et d'apprentissage machine modernes non paramétriques pour régulariser les coefficients du modèle, calibrer empiriquement les valeurs p et les intervalles de confiance au moyen de procédures *bootstrap*, et construire des modèles non linéaires flexibles avec des forêts aléatoires et des réseaux de neurones artificiels (voir Peters et coll. (2017), Zhao et Hastie [2021] pour connaître les résultats prometteurs de ces modèles en inférence causale).
5. Faire la synthèse des résultats des différentes méthodes et régions géographiques.

Les approches utilisées dans l'étude sont guidées par les données et peuvent servir à évaluer l'incidence d'autres types de changements progressifs, comme la décriminalisation d'autres substances ou l'adoption de voitures autonomes.



2 Revue de la littérature

Effets du cannabis et de la décriminalisation

Des études ont porté sur : 1) les effets du cannabis sur la capacité de conduire (études comparatives et études observationnelles); 2) les effets de la décriminalisation sur la proportion de conducteurs ayant obtenu un résultat positif à la consommation de cannabis (études observationnelles); 3) les effets de la décriminalisation sur la santé et la sécurité publique (études observationnelles). Dans le premier groupe d'études, Gelmi et coll. (2021) ont fait des essais comparatifs auprès de 33 volontaires en Suisse (tous des fumeurs ou des ex-fumeurs) et n'ont détecté aucun effet significatif sur la capacité de conduite des sujets qui fumaient du cannabis riche en cannabidiol (CBD), alors que tous ces participants dépassaient la limite légale de THC dans le sang pour conduite avec facultés affaiblies, après avoir fumé du cannabis riche en CBD. Hartman et coll. (2016) ont rapporté les résultats d'une autre expérience (19 volontaires dans l'Iowa, aux États-Unis) qui a montré que les sujets dont les facultés étaient affaiblies par le THC réduisaient leur vitesse et maintenaient des interdistances plus grandes, ce qui laisse penser que les conducteurs étaient conscients de l'altération de leurs facultés et se montraient plus prudents. Par contre, Bédard et coll. (2007) ont utilisé les données des conducteurs âgés de 20 à 50 ans du Fatality Analysis Reporting System (FARS) sur la période 1993-2003 et ont constaté que les conducteurs qui avaient obtenu un résultat positif au test de dépistage du cannabis (soit 5 % des conducteurs) présentaient des rapports de cotes plus élevés pour l'existence de « facteurs liés au conducteur » (il s'agit de 40 facteurs codés dans les rapports de police, notamment les excès de vitesse, la conduite dangereuse et le non-respect des panneaux de signalisation).

Dans le deuxième groupe d'études, Chung et coll. (2019) ont signalé une proportion accrue de présence de cannabis chez des patients admis dans des hôpitaux du Colorado avec une lésion traumatique, après la décriminalisation du cannabis dans l'État. Pollini et coll. (2015) ont utilisé les données du FARS relatives à la Californie (2008-2012) et détecté une proportion statistiquement plus élevée de prévalence du CBD en 2012 comparativement à la période 2008-2010 qui précédait la décriminalisation. Certaines études ont permis de compléter les données du FARS par des données hors collisions. Romano et coll. (2017) ont résumé les résultats de deux études (Li et coll., en 2013, et Romano et coll., en 2014) qui ont utilisé des données très similaires, tirées du FARS (cas de décès) et de la National Roadside Survey de 2007 (NRS; cas témoins hors décès), et ils montrent que même si les rapports de cotes bruts révèlent des effets importants de l'intoxication au cannabis, les rapports ajustés (corrigés pour l'âge, le sexe et la race) dans tous les sous-ensembles considérés d'États ne témoignent plus de la contribution du cannabis aux collisions.

Tandis que les deux premiers groupes d'études constituent des bases importantes pour l'étude de l'impact du cannabis, le présent projet et le troisième groupe d'études portent principalement sur l'estimation des effets totaux sur les collisions et les décès. Les études de ce groupe font état de résultats mitigés, le plus souvent ne détectant pas les effets de la décriminalisation du cannabis sur les collisions routières (Hall et Lynskey, 2020). Roy et coll. (2020) ont étudié les données semestrielles du FARS sur la période 2010-2017 provenant de villes dont la population est supérieure à 100 000 et situées dans des États qui n'avaient pas promulgué de lois sur le cannabis à des fins médicales en date de 2010. S'appuyant sur la méthode des doubles différences, cette étude a révélé une légère réduction (9 %) des collisions mortelles après la légalisation du cannabis à des fins médicales dans les États, mais aucun effet de la décriminalisation n'a été détecté, sauf dans le cas des jeunes conducteurs de sexe masculin (augmentation de 13 % des accidents mortels). Lane et Hall (2019) ont procédé à une analyse de séries chronologiques mensuelles interrompues (2009-2016) et ont détecté une augmentation à court terme des décès enregistrés dans le FARS au Colorado, dans l'État de Washington, en Oregon et dans les États voisins (une augmentation de 1 décès par million de résidents dans l'année suivant la légalisation du cannabis). Favero et coll. (2022) ont étudié les taux de collision trimestriels par mille parcouru sur la période 2009-2019 dans onze États, tout en tenant compte des effets temporels, du taux de port de la ceinture de sécurité, du chômage et des taux de consommation d'alcool. Les auteurs ont conclu que les estimations des effets de la légalisation du cannabis ne sont pas toujours statistiquement significatives et varient selon les États. Wilkie et coll. (2021a) et Windle et coll. (2021b) ont étendu au Canada les résultats obtenus avec la base de données FARS des États-Unis au sujet des effets de la décriminalisation sur les décès et les taux de collisions mortelles. Wilkie et coll. (2021a) ont constaté une augmentation des taux de collisions mortelles lorsqu'ils ajustent le modèle pour tenir compte de la tendance temporelle linéaire. Toutefois, un commentaire dans l'article publié en ligne par Jiang et coll. rapporte que le modèle de Poisson utilisé donne un mauvais ajustement. Enfin, le rapport publié en 2017 par la NHTSA du département des Transports (DOT) donne un excellent aperçu des recherches connexes et des questions pratiques d'actualité relatives aux processus de dépistage des drogues, aux dispositifs utilisés et à la formation du personnel.

La présente étude s'appuie conjointement sur les progrès techniques des algorithmes d'apprentissage machine pour trouver des tendances dans les données, sur des méthodes rigoureuses de sélection des unités témoins sur la base de données d'observation et sur les voies nouvellement tracées pour interpréter de façon causale les modèles à boîte noire. Aucune étude à ce jour sur les effets de la décriminalisation du cannabis n'avait réuni toutes ces étapes.

Plus précisément, cette étude tente de surmonter les inconvénients des études antérieures qui reposaient sur l'hypothèse de linéarité (l'auteur suggère d'utiliser des modèles non linéaires lorsque cela est approprié, tels que des forêts aléatoires) et l'inférence paramétrique (l'auteur propose l'utilisation du bootstrap pour calculer les intervalles de confiance sans s'appuyer sur une distribution particulière).

Vue d'ensemble

Il y a d'autres questions à explorer concernant la recherche sur le cannabis, qui pourraient influencer sur la façon dont les gens perçoivent la drogue, tenir compte des différents niveaux de tolérance et même consister à tester les gens pour savoir s'ils sont sous l'influence de la drogue. Par exemple, le chapitre 12 de Potter et Weinstock (2019) établit une comparaison du cannabis avec l'industrie du tabac, de l'alcool et des drogues (opioïdes), tandis qu'Ammerman et coll. (2015) donnent une vue d'ensemble des expériences internationales et soutiennent que certains effets du cannabis, comme la dépendance, sont négligés par rapport à l'alcool et au tabac. Alexander (2003) fait valoir l'insuffisance des méthodes existantes de dépistage du cannabis. Bernard et coll. (2021) donnent un nouvel aperçu des questions connexes concernant la relation dose-effet (concentration de THC dans le sang et comportement réel au volant; incertitude quant à la concentration de THC qui entraîne systématiquement une altération de la conduite), la tolérance après une utilisation fréquente, les effets des ratios CBD:THC et les différents modes de prise de la drogue (inhalation par combustion, par vapeur ou ingestion par voie orale).

3 Aperçu des données

Pour ne pas confondre les effets de la décriminalisation avec ceux de la pandémie de COVID-19, les données de 2020 ou des années ultérieures ne sont pas utilisées. Plus précisément, aucune province au Canada ne peut agir comme groupe témoin lorsqu'il s'agit d'estimer les effets de la décriminalisation du cannabis (voir la section 3.1); les effets de la pandémie de COVID-19 seraient donc confondus avec d'autres effets. Aux États-Unis, les États servent de groupes témoins lorsqu'il s'agit d'estimer les effets de la décriminalisation (voir la section 3.2) et certains chercheurs soutiennent qu'il n'est pas nécessaire de procéder à un ajustement explicite pour tenir compte des facteurs parasites lorsqu'on utilise un groupe témoin (Elvik, 2002). Toutefois, les effets de la COVID-19, comme les restrictions de mobilité du fait de l'imposition de confinements de sévérité diverse, ont varié considérablement au sein des États (Barnes et coll., 2020), entre les États (Lin et coll., 2021) et à l'échelle mondiale (Gupta et coll., 2021, Rahman et coll., 2021, Yasin et coll., 2021). De plus, la reprise de la mobilité humaine après le relâchement des confinements était spatialement hétérogène. Ces tendances poseraient des difficultés supplémentaires si elles étaient incluses dans l'étude (voir par exemple Choi et coll., 2018), car les facteurs parasites peuvent influencer sur les emplacements différemment dans le temps (Smith et coll., 1993).

3.1 Canada

La légalisation de la consommation récréative de cannabis est entrée en vigueur partout au Canada le 17 octobre 2018¹. Dans le passé (avant 2018), la consommation de cannabis à des fins récréatives dans la province de la Colombie-Britannique était plus tolérée que dans le reste du pays². Toutefois, en raison de différences importantes avec d'autres provinces, ces dernières ne peuvent pas servir de bons groupes témoins lorsqu'il s'agit d'étudier les effets de la consommation de cannabis en Colombie-Britannique.

Les données utilisées dans cette étude proviennent des rapports annuels de l'Agence statistique d'assurance générale³ (ASAG), qui recueille des informations sur l'assurance auprès de la plupart des régions canadiennes (sauf la Colombie-Britannique, le Manitoba, le Québec et la Saskatchewan; voir les tableaux AUTO1003, AUTO1005 et AUTO1010 sur le site Web de l'ASAG), et du Groupement des assureurs automobiles⁴ (GAA), qui fournit les informations pour le Québec (voir le Rapport 1A.2 – Sommaire sur le site Web du GAA). Des informations sur la *collision de véhicules de tourisme* par année de survenance ont été extraites de ces rapports. Les réclamations de cette catégorie sont caractérisées par des facteurs de développement des sinistres proches de 1; par conséquent, on ne s'attend pas à ce qu'ils changent considérablement à mesure que se matérialiseront les sinistres d'accidents survenus au cours des dernières années. Les montants en dollars ont été ajustés aux prix de 2019 pour tenir compte de l'évolution de la valeur de l'argent, au moyen de l'indice des prix à la consommation (tous les groupes de produits) de Statistique Canada⁵.

¹ <https://sencanada.ca/fr/sencaplus/nouvelles/loi-cannabis/> (consulté le 11 juillet 2021)

² https://en.wikipedia.org/wiki/Cannabis_in_British_Columbia (consulté le 11 juillet 2021)

³ <https://www.gisa.ca/> (consulté le 1^{er} juillet 2021)

⁴ <https://gaa.qc.ca/fr/statistiques/en-bref/> (consulté le 1^{er} juillet 2021)

⁵ Statistique Canada. Tableau 18-10-0005-01, *Indice des prix à la consommation, moyenne annuelle, non désaisonnalisé*, <https://doi.org/10.25318/1810000501-fra> (date de diffusion : 20 janvier 2021)

La figure 1 montre la dynamique par région des variables étudiées. La population varie d'une région à l'autre; par conséquent, le nombre annuel d'années-véhicules acquis a été utilisé pour pondérer les observations dans des analyses plus poussées (figure 2).

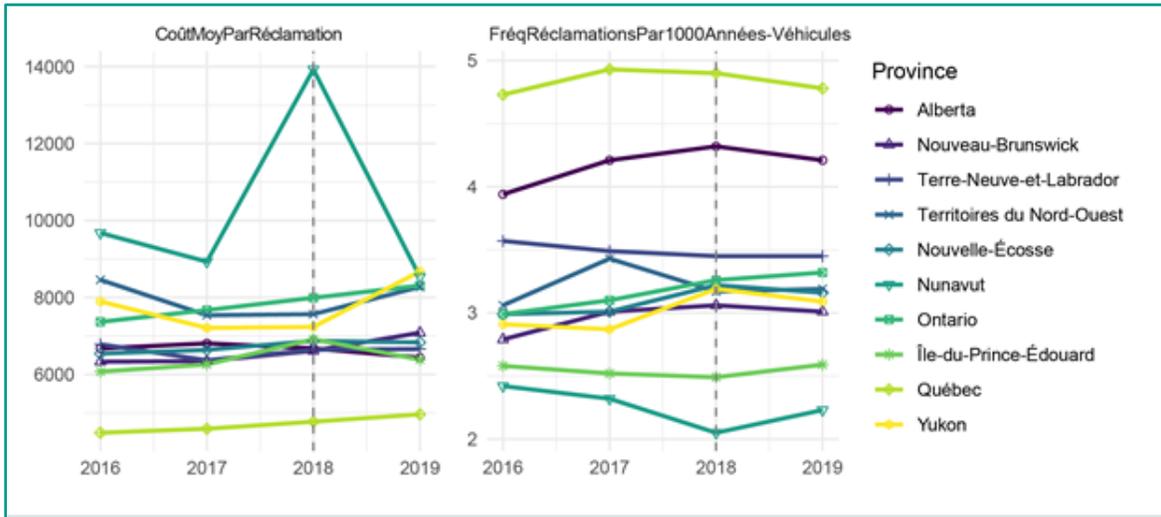


Figure 1 : Coût annuel moyen par réclamation (en dollars de 2019) et fréquence des collisions. La ligne pointillée indique l'année de décriminalisation du cannabis.

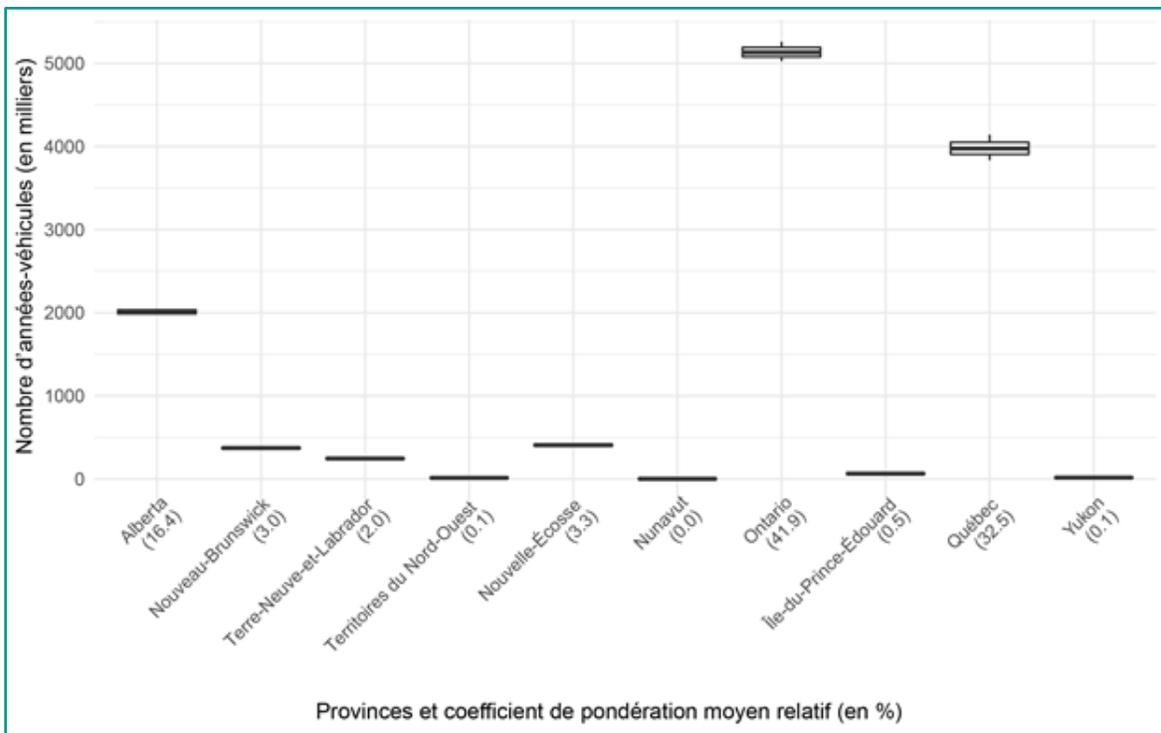


Figure 2 : Nombre d'années-véhicules acquis en 2016-2019, utilisé comme coefficient de pondération régional. Les points d'inflexion inférieur et supérieur correspondent aux premier et troisième quartiles.

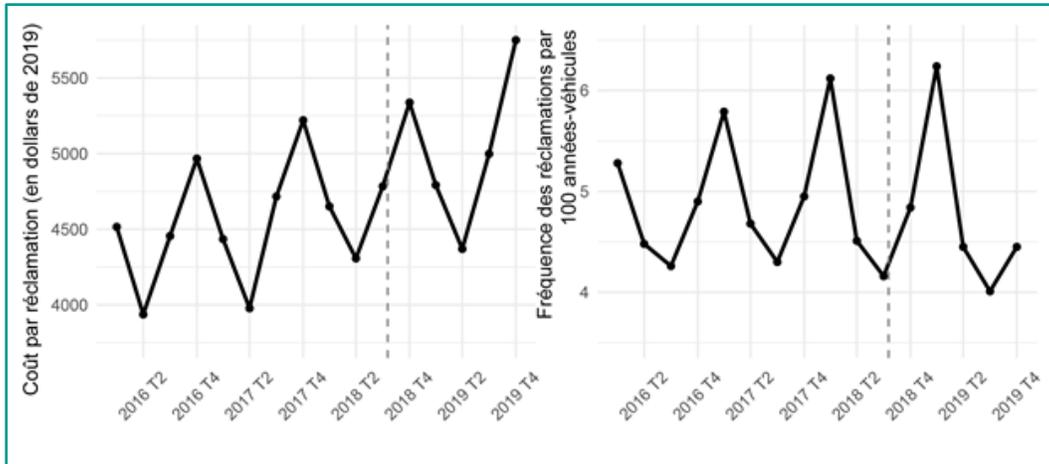


Figure 3 : Coût moyen trimestriel par réclamation (en dollars de 2019) et fréquence des collisions au Québec. La ligne pointillée indique la date de décriminalisation du cannabis.

3.2 États-Unis

Aux États-Unis, la légalité du cannabis varie selon les États. Le tableau 1 résume l'historique de la légalisation par les États durant la période à l'étude pour laquelle les données sur les accidents sont disponibles (voir ci-après). Les statistiques suivantes sur les États ont servi à sélectionner les États composant le groupe témoin : urbanisation (pourcentage de la population totale)⁶, population⁷, milles-véhicules par conducteur titulaire d'un permis⁷, milles parcourus par 1 000 personnes (soit le résultat de la division du millage total sur la route et en ville par la population)⁷ et nombre de véhicules par 1 000 personnes⁸. Le nombre de conducteurs titulaires d'un permis a également été pris en considération, mais il est corrélé trop fortement ($r = 0,99$) à la population.

Pour étudier les accidents d'automobile aux États-Unis, nous utilisons la quatrième version du Countrywide Traffic Accident Dataset (CTAD) (Moosavi et coll., 2019). Le CTAD couvre les États contigus des États-Unis de février 2016 à décembre 2020 (environ 4,2 millions d'accidents). Toutefois, une analyse plus détaillée laisse penser que les données spatiotemporelles ne sont pas bien représentées également et qu'on manque de données pour certains États au début de la période d'échantillonnage. Néanmoins, les avantages de cet ensemble de données comprennent des résolutions spatiales et temporelles élevées et une collecte universelle de données à travers les États-Unis au moyen des interfaces de programmation d'applications (API) de MapQuest Traffic et de Microsoft Bing Map Traffic. L'un des inconvénients de cet ensemble de données est que la gravité des accidents est mesurée en fonction de la gravité de leur impact sur la circulation, et non en fonction des dommages matériels ou corporels.

Le CTAD contient aussi des informations sur des variables parasites telles que les conditions météorologiques enregistrées dans un aéroport voisin au moment approximatif de l'accident. Toutefois, pour intégrer aux modèles les effets de la météo, il faut un historique météorologique complet (pas seulement au moment des accidents). Les informations sur la température de l'air et les précipitations ont été obtenues de deux autres sources :

1. Les données horaires ont été obtenues du réseau américain ASOS d'observations météorologiques automatisées des aéroports⁹, à l'aide des codes des aéroports du CTAD. Lorsqu'on ne pouvait trouver d'aéroport figurant dans le CTAD, nous avons déterminé son emplacement de façon approximative par l'utilisation des latitudes et des longitudes des accidents déclarés et nous avons utilisé les distances géodésiques pour remplacer l'aéroport manquant par un aéroport situé à proximité et représenté dans la base de données ASOS.
2. Données quotidiennes sur grille de Daymet¹⁰. Le choix des cellules de la grille a été fait en fonction des coordonnées des stations météorologiques des aéroports du réseau ASOS.

⁶ https://en.wikipedia.org/wiki/Urbanization_in_the_United_States (consulté le 1^{er} novembre 2021)

⁷ <https://www.fhwa.dot.gov/ohim/onh00/onh2p11.htm> (consulté le 25 décembre 2021)

⁸ https://en.wikipedia.org/wiki/List_of_U.S._states_by_vehicles_per_capita (consulté le 1^{er} novembre 2021)

⁹ <https://mesonet.agron.iastate.edu/request/download.phtml?network=AWOS> (consulté le 2 janvier 2022)

¹⁰ Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4 <https://doi.org/10.3334/ORNLDAAC/1840> (consulté le 2 janvier 2022)

Le début de l'échelle de temps du tableau 1 (début 2016) est défini par la disponibilité des données dans le CTAD. Étant donné que certains États (dont le Colorado et Washington) avaient déjà légalisé le cannabis en 2016, leurs données ne peuvent servir aux comparaisons avant-après de cette étude qui portent sur la fréquence des accidents de voiture. Nous avons choisi la fin de l'échelle de temps du tableau 1 (décembre 2019, comme pour le Canada) de manière à ne pas confondre les effets estimés avec ceux de la pandémie de COVID-19. Au Michigan, les ventes de cannabis récréatif avec l'autorisation de l'État ont commencé le 1^{er} décembre 2019, soit un mois avant la fin de la période à l'étude. Afin que toutes les périodes avant et après soient d'au moins un an et que nous puissions ainsi tirer des conclusions plus rigoureuses, nous avons retranché les données du Michigan.

Outre le CTAD, nous avons utilisé le FARS¹¹ de la National Highway Traffic Safety Administration (NHTSA) du département des Transports des États-Unis (DOT) pour extraire les données de cette étude.

La base de données a été complétée notamment par les variables suivantes: Année (variable numérique), Mois (1 à 12, variable catégorielle), Jour de la semaine (1 à 7, variable catégorielle), Fin de semaine (0/1, variable catégorielle), Jour férié (0/1, variable catégorielle, selon le calendrier des jours fériés 2022 de Wuertz et coll. établi pour la Bourse de New York).

Les indicateurs de l'expérience d'accidents automobiles ont été normalisés en fonction du nombre de véhicules immatriculés par année et par État, disponible à l'adresse <https://www.fhwa.dot.gov>.

Tableau 1 : Groupes d'États dont les données servent à l'étude de la légalisation du cannabis aux États-Unis (de mars 2016 à décembre 2019). Les dates du début des ventes commerciales sont utilisées dans l'analyse et indiquées entre parenthèses.

Groupe	États
Cannabis légalisé	CA (1er janvier 2018), MA (20 novembre 2018), NV (1er juillet 2017)
Complètement illégal	ID, KS, NE, NC, SC, TN, WY
Illégal à des fins récréatives	AL, AR, FL, IN, IA, KY, OK, TX, UT, WV, WI
Légalisé après la période à l'étude	AZ, MT, NJ, NY, SD, VA
Situation mixte (les données de ces États n'ont pas été utilisées dans l'étude)	AK, CO, CT, DE, DC, GA, HI, IL, LA, ME, MD, MI, MN, MS, MO, NH, NM, ND, OH, OR, PA, RI, VT, WA

¹¹ <https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars> (consulté le 1^{er} juillet 2021)

4 Méthodes

Utiliser des tests bilatéraux avec un niveau de signification $\alpha = 0,05$ sauf indication contraire ($1 - \alpha = 0,95$ ou 95 % correspond au niveau de confiance). Autrement dit, les résultats sont statistiquement significatifs lorsque les valeurs p correspondantes sont inférieures à α .

4.1 Comparaisons avant-après

Pour effectuer de simples comparaisons avant-après lorsqu'aucun groupe témoin concomitant n'est disponible (au Canada, par exemple), les modèles de régression suivants avec inférence paramétrique et bootstrap résiduel (semi-paramétrique) sont appliqués.

Modèles paramétriques

Considérons le modèle à effets mixtes (Zuur et coll., 2009)

$$y_{tp} = a + bX_t + ct + \beta_p + e_{tp}, \quad (1)$$

où y est la variable dépendante (fréquence des réclamations ou coût moyen par réclamation), t est l'indice numérique du temps (année), p est l'indice spatial (province), a est l'ordonnée à l'origine principale b et c sont les coefficients des effets fixes, X est l'indicateur indiquant si t se situe avant ou après la date de décriminalisation du cannabis, $\beta_p \sim N(0, \delta^2)$ sont des ordonnées à l'origine aléatoires qui tiennent compte des différents niveaux moyens de la variable dépendante selon la province et $e_{tp} \sim N(0, \sigma^2)$ sont les résidus du modèle. Les coefficients de pondération w_{tp} peuvent être appliqués aux variances résiduelles pour l'estimation pondérée, de façon que $e_{tp} \sim N(0, \sigma^2 \times w_{tp})$ (section 4.1.2 dans Zuur et coll., 2009). Plus particulièrement, $w_{tp} = (\text{Nombre d'années véhicules acquis})_{tp}^{-1}$ peut être utilisé pour assigner des coefficients de pondération plus élevés aux observations correspondant à des années-véhicules acquis. Nous souhaitons estimer et tester la signification statistique du coefficient b qui résume les effets de la décriminalisation du cannabis.

L'inclusion d'une tendance temporelle dans le modèle de régression est l'une des façons de tenir compte de cette tendance et d'éviter ainsi d'obtenir des résultats factices (voir la section 10.5 dans Wooldridge, 2013). Cela est possible en raison de l'interprétation conditionnelle des coefficients dans la régression multiple qui peut différer des relations marginales (voir par exemple la section 1.3.1 dans Chatterjee et Simonoff, 2013). Par exemple, le coefficient b du modèle (1) peut être interprété comme le changement attendu de la fréquence des réclamations du fait de la décriminalisation du cannabis, tout le reste du modèle étant inchangé. Dans cette analyse, la tendance linéaire est adéquate pour prendre en compte le comportement tendanciel d'une série chronologique (d'après la figure 1) et constitue donc une option privilégiée d'analyse de régression d'une série non stationnaire, comparativement à l'utilisation dans le modèle de régression de la série différenciée corrigée de la tendance (p. ex., $\Delta y_{t,p} = y_{t,p} - y_{t-1,p}$) ou des écarts par rapport à une tendance estimée \tilde{y}_{tp} (c.-à-d. $\hat{e}_{tp} = y_{tp} - \tilde{y}_{tp}$).

Pour les données saisonnières d'une seule province, considérons le modèle

$$y_t = a + bX_t + ct + \sum_{i=1}^{s-1} d_i S_{it} + e_t, \quad (2)$$

où S_{it} est une variable indicatrice qui prend la valeur 1 si l'instant t appartient à la saison i et la valeur 0 autrement ($i = 1, \dots, 4$ dans le cas des données trimestrielles; pour des raisons d'identification, seulement trois coefficients d_i sont estimés). Autrement dit, le modèle (2) comprend une tendance linéaire avec saisonnalité additive modélisée par l'utilisation d'une variable catégorielle pour les saisons (trimestres).

Bootstrap

Pour que l'inférence basée sur les modèles à effets mixtes soit robuste à la non-normalité, appliquez l'algorithme du *bootstrap* résiduel semi-paramétrique (Carpenter et coll., 2003, Loy et coll., 2022) :

1. Estimer les paramètres et les termes d'erreur du modèle.
2. Prélever de façon indépendante des échantillons avec remise parmi les ensembles de termes d'erreur.
3. Modifier l'échelle des termes d'erreur obtenus par *bootstrap* afin que leur variance empirique soit égale aux estimations du modèle.
4. Obtenir des échantillons *bootstrap* en combinant les échantillons obtenus par *bootstrap* au moyen de l'équation du modèle ajusté.
5. Ajuster à nouveau le modèle et extraire les statistiques d'intérêt.

6. Répéter les étapes 2 à 5 B fois.

À partir des distributions obtenues par *bootstrap*, obtenir des intervalles de confiance pour le niveau de confiance souhaité $1 - \alpha$, tel qu'il est défini au chapitre 5 dans Davison et Hinkley (1997). Pour l'estimation \hat{b} et ses équivalents *bootstrap* \hat{b}^* , les limites de confiance *bootstrap* de base sont données par

$$2\hat{b} - \hat{b}_{[(B+1)(1-\alpha/2)]}^*, \quad 2\hat{b} - \hat{b}_{[(B+1)\alpha/2]}^*, \quad (3)$$

les centiles limites sont donnés par

$$\hat{b}_{[(B+1)\alpha/2]}^*, \quad \hat{b}_{[(B+1)(1-\alpha/2)]}^*, \quad (4)$$

et les limites avec approximation normale sont données par

$$\hat{b} - v^{1/2}z_{1-\alpha/2}, \quad \hat{b} + v^{1/2}z_{1-\alpha/2}, \quad (5)$$

où v est la variance approximative et z est le quantile de la distribution $N(0,1)$.

4.2 Appariement par score de propension

Lorsque des essais comparatifs randomisés ne peuvent être mis en œuvre, l'appariement par score de propension (ASP) sert à appairer les sujets à l'étude en fonction de la probabilité estimée qu'ils reçoivent le traitement, sur la base de certaines covariables de base X_1, X_2, \dots, X_k qui ne sont pas touchées par le traitement (Austin, 2011, Olmos et Govindasamy, 2015, Zhao et coll., 2021). L'ASP cherche donc à créer des groupes homogènes pour éviter la possibilité de confondre les effets étudiés avec les effets des covariables de base. Par conséquent, étant donné des scores de propension semblables, l'affectation de sujets au traitement peut être considérée comme étant aléatoire, ce qui est important pour satisfaire aux hypothèses de nombreux modèles statistiques et de l'inférence causale (Damrongplisit et coll., 2010). Par exemple, il est souhaitable de comparer l'État qui a décriminalisé le cannabis avec l'État ou les États qui présentent des caractéristiques de base similaires, comme la population et l'urbanisation, afin de réduire au minimum les effets de ces caractéristiques sur les écarts de l'expérience d'accidents automobiles entre les États. Entre autres exemples pertinents de mise en œuvre de l'ASP, mentionnons Lopez Bernal et coll. (2018), Damrongplisit et coll. (2010) et Stuart et coll. (2014).

Les scores de propension sont les probabilités conditionnelles de se voir assigner un traitement, qui sont estimées au moyen du modèle.

$$\text{Traitement} \sim X_1 + X_2 + \dots + X_k,$$

qui est, pour l'essentiel, un modèle de classification pour lequel le traitement assigné (décriminalisation ou non) est la variable dépendante¹². Dans cette étude, les covariables X_1, X_2, \dots, X_k sont représentées par

1. urbanisation;
2. population;
3. milles-véhicules par conducteur titulaire d'un permis;
4. milles parcourus par 1 000 personnes;
5. nombre de véhicules par 1 000 personnes.

La régression logistique binaire est la forme la plus utilisée de ce classificateur, mais les techniques d'apprentissage machine permettent une plus grande flexibilité dans la modélisation des relations et des interactions non linéaires (effets combinés) de plusieurs covariables. D'après les méthodes arborescentes et les réseaux neuronaux, la forêt aléatoire est une méthode d'arbre de décision qui offre une performance concurrentielle et qui, tout à la fois, ne comporte que quelques hyperparamètres que l'utilisateur doit définir (Hastie et coll., 2009). Par conséquent, la présente étude utilise le classificateur de forêt aléatoire pour prédire la probabilité qu'un État se voie assigner le traitement.

Dans l'algorithme de forêt aléatoire mis en œuvre, l'ensemble de données original est rééchantillonné avec remise 500 fois, et un arbre de classification est construit sur chaque échantillon *bootstrap* de ce type. Chaque arbre utilise les covariables X_1, X_2, \dots, X_k pour diviser les données en sous-ensembles homogènes plus petits, selon les valeurs de la variable dépendante. Pour plus de robustesse (décorrélation des arbres), chaque division de l'arbre utilise

¹² Une autre façon de considérer l'ASP est de la définir comme une technique de réduction de dimensionnalité permettant d'utiliser k covariables en même temps pour trouver des sujets qui sont proches des sujets traités dans cet espace à k dimensions (Olmos et Govindasamy, 2015).

seulement un sous-ensemble aléatoire de \sqrt{k} covariables du total de k covariables. Ensuite, chaque arbre de la forêt est utilisé pour les prédictions, et la proportion d'arbres qui classe l'État comme étant décriminalisé est utilisée comme probabilité de traitement ou comme score de propension. Pour plus de détails sur l'algorithme de forêt aléatoire, voir Breiman (2001) et le chapitre 15 dans Hastie et coll. (2009).

À l'étape de l'appariement, les scores de propension servent à sélectionner les bons sujets du groupe témoin à partir d'un nombre suffisant d'unités non traitées. Les techniques d'appariement varient, allant de l'identification des plus proches voisins aux algorithmes d'optimisation et d'analyse par groupe. L'appariement avec remise permet d'apparier un sujet du groupe non traité à plus d'un sujet du groupe traité. Plusieurs méthodes d'appariement sont habituellement appliquées, car les résultats ultimes peuvent être sensibles à la sélection des sujets du groupe témoin (Zhao et coll., 2021). Par exemple, Damrongplasit et coll. (2010) ont utilisé l'appariement stratifié selon lequel on a d'abord sélectionné une fourchette de scores de propension avec des sujets traités et des sujets non traités, avant de les séparer en intervalles plus petits (chaque intervalle comprenait aussi les deux types de sujets). Dans chaque strate, l'effet moyen du traitement a été estimé pour évaluer la variabilité des estimations. Dans cette étude, la sélection du plus proche voisin est utilisée avec différents nombres de sujets du groupe témoin à assigner à chaque sujet traité. Avec l'appariement 1:1, les éléments appariés sélectionnés sont très similaires du point de vue de leurs scores de propension, mais on peut utiliser des proportions plus grandes pour obtenir de meilleures estimations du contrefactuel (Olmos et Govindasamy, 2015). Les différences moyennes standardisées sont habituellement utilisées pour évaluer l'équilibre des covariables dans les groupes ajustés de traitement et de contrôle (témoin), avec un seuil empirique de 0,1 ou 0,25 (Zhao et coll., 2021).

Appariement conventionnel

Comme solution de rechange aux scores de propension, la proximité des États dans l'espace à k dimensions des covariables peut servir à sélectionner des États similaires, sans lier les covariables aux résultats de la décriminalisation. Cette approche est parfois appelée « appariement conventionnel » (Zhao et coll., 2021). Dans un tel cas non supervisé, on peut encore utiliser une forêt aléatoire pour extraire une matrice de proximité (Shi et Horvath, 2006) afin d'identifier les voisins dans l'espace à k dimensions. La proximité dans une forêt aléatoire est quantifiée comme suit : si une paire d'observations hors sac (non utilisées pour construire un arbre, car en raison de l'échantillonnage avec remise, certaines observations sont laissées de côté) aboutit dans le même nœud terminal de l'arbre, leur proximité est augmentée de 1 (voir la section 15.3 dans Hastie et coll., 2009).

4.3 Régression par forêt aléatoire

La méthode de la forêt aléatoire a été présentée dans la section sur l'ASP comme étant un classificateur (pour prédire l'attribution d'une unité au traitement; la variable dépendante est binaire) et un outil pour trouver des distances dans un espace de grande dimension pour effectuer un appariement plus conventionnel des États en fonction de leur proximité dans cet espace (Breiman, 2001, Hastie et coll., 2009). Les forêts aléatoires peuvent également être utilisées dans un contexte de régression lorsque la variable dépendante modélisée est numérique, comme le taux d'accident de voiture (voir par exemple Bailey et coll., 2020).

Les forêts aléatoires comportent seulement quelques hyperparamètres, dont le nombre d'arbres, le nombre de variables choisies au hasard pour sélectionner la meilleure division, et la taille minimale des sous-ensembles (soit les nœuds terminaux d'un arbre de régression) qui ne font plus l'objet d'une division. Le nombre d'arbres influe sur la stabilité des résultats, mais tant qu'il est relativement grand, l'erreur d'estimation se stabilise. Dans cette étude, 500 arbres sont utilisés dans chaque forêt aléatoire et les deux autres hyperparamètres sont calibrés au moyen d'une validation croisée à dix blocs qui minimise la racine carrée de l'erreur quadratique moyenne.

Vu que les forêts aléatoires sont constituées d'un grand nombre d'arbres de régression, l'interprétabilité directe du modèle est perdue (la méthode devient une boîte noire). Toutefois, le post-traitement de la forêt aléatoire entraînée permet d'obtenir de précieuses informations, y compris celles qui sont semblables aux informations tirées des modèles statistiques. Ainsi, Altmann et coll. (2010) ont proposé des permutations des covariables pour calculer des valeurs p évaluant la signification statistique de la contribution des covariables à la précision prédictive d'une forêt aléatoire. Pour visualiser les relations apprises par la forêt aléatoire, on construit des graphiques de dépendance partielle (GDP) en variant les valeurs des covariables sélectionnées, tout en maintenant les autres covariables fixes (chapitre 10 dans Hastie et coll., 2009). Cette méthode s'applique à de nombreux autres types de modèles à boîte noire. Les GDP obtenus montrent les effets marginaux des variables, et la façon dont ils sont estimés correspond à l'ajustement *back-door* de Pearl (1993) qui est utilisé pour identifier les effets de causalité à partir des données d'observation (Zhao et Hastie, 2021). Plus précisément, lorsque le prédicteur est une variable binaire comme la décriminalisation du cannabis, le GDP montre l'effet moyen du traitement (Zhao et Hastie, 2021).

5 Résultats

5.1 Canada

Dans cette section, le nombre d'années-véhicules acquis par région (figure 2) a servi de coefficient de pondération.

5.1.1 Coût moyen par réclamation

Les résultats du coût moyen par réclamation s'établissent comme suit. La valeur p des effets de la décriminalisation du cannabis au Canada dans le modèle à effets mixtes est supérieure au seuil de signification (tableau 2), ce qui veut dire qu'il n'y a pas assez de preuves des effets de la décriminalisation. La distribution *bootstrap* de ce paramètre est centrée autour de zéro (figure 4) et toutes les versions des intervalles de confiance contiennent zéro (tableau 3), témoignant ainsi de l'absence d'effets significatifs.

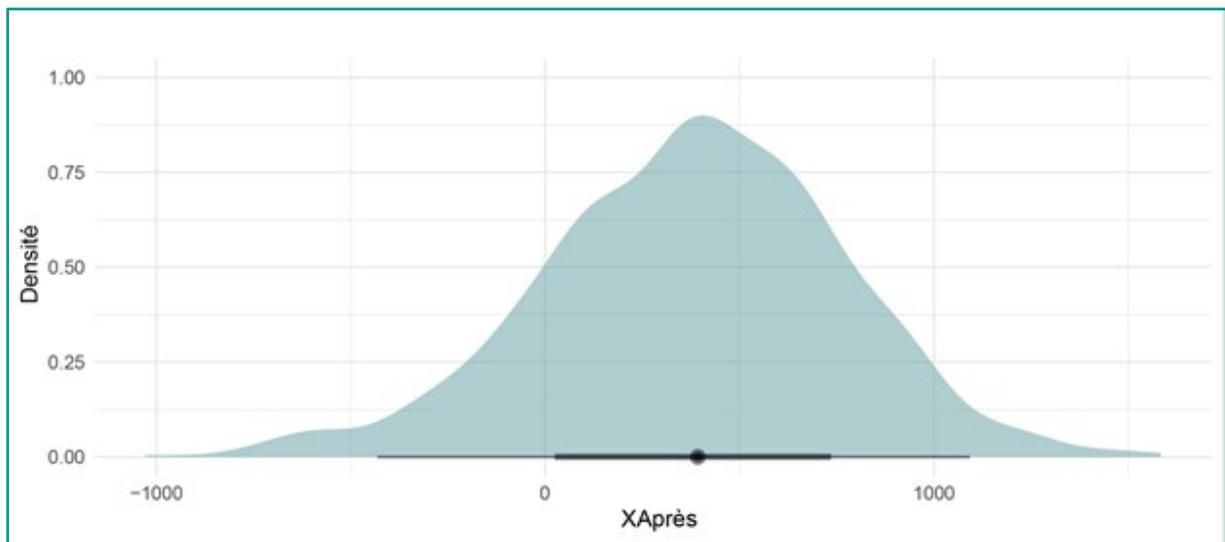


Figure 4 : Graphique de la densité de la distribution bootstrap du coefficient b dans le modèle à effets mixtes des effets de la décriminalisation du cannabis sur le coût moyen par réclamation au Canada. Les lignes horizontales correspondent à des niveaux de confiance de 66 % et 95 %.

Tableau 2 : Estimations paramétriques tirées du modèle à effets mixtes des effets de la décriminalisation du cannabis sur le coût moyen par réclamation au Canada

	Valeur	Erreur type	Degr. lib.	Valeur t	Valeur p
XAfter	376,912	380,70	18	0,99005	0,33527
Année	-12,204	144,01	18	-0,08475	0,93340

Tableau 3 : Intervalles de confiance bootstrap pour le coefficient b du modèle à effets mixtes des effets de la décriminalisation du cannabis sur le coût moyen par réclamation au Canada

	Terme	Estimation	Borne inf.	Borne sup.	Type	Niveau
1	XAfter	376,91	-384,09	1 134,1	Normal	0,95
2	XAfter	376,91	-338,37	1 185,4	De base	0,95
3	XAfter	376,91	-431,57	1 092,2	Centile	0,95

5.1.2 Fréquence des réclamations

Les résultats de la fréquence des réclamations par 100 années-véhicules acquis donnent lieu aux mêmes conclusions que les résultats du coût moyen par réclamation présentés à la section précédente. Plus précisément, la valeur p des effets de la décriminalisation du cannabis au Canada dans le modèle à effets mixtes est supérieure au seuil de signification (tableau 4), ce qui veut dire qu'il n'y a pas assez de preuves des effets de la décriminalisation. La distribution *bootstrap* de ce paramètre est centrée autour de zéro (figure 5) et toutes les versions des intervalles de confiance contiennent zéro (tableau 5), témoignant ainsi de l'absence d'effets significatifs.

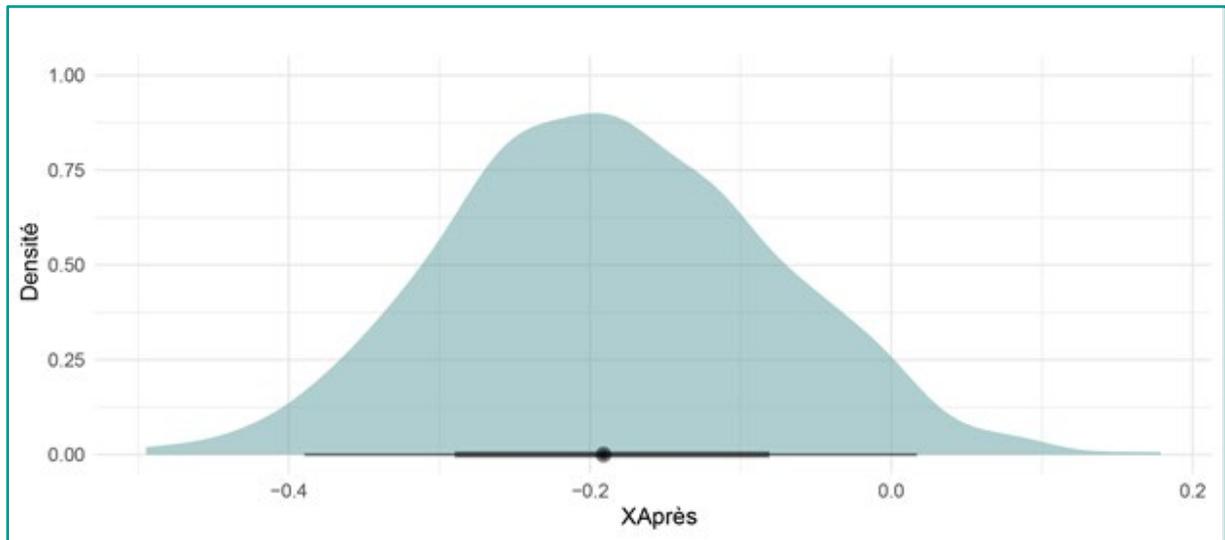


Figure 5 : Graphique de la densité de la distribution bootstrap du coefficient b dans le modèle à effets mixtes des effets de la décriminalisation du cannabis sur la fréquence des réclamations au Canada. Les lignes horizontales correspondent à des niveaux de confiance de 66 % et 95 %.

Tableau 4 : Estimations paramétriques tirées du modèle à effets mixtes des effets de la décriminalisation du cannabis sur la fréquence des réclamations au Canada

	Valeur	Erreur type	Degr. lib.	Valeur t	Valeur p
XAfter	-0,19125	0,13038	18	-1,4668	0,159684
Année	0,10657	0,04930	18	2,1616	0,044364

Tableau 5 : Intervalles de confiance bootstrap pour le coefficient b du modèle à effets mixtes des effets de la décriminalisation du cannabis sur la fréquence des réclamations au Canada

	Terme	Estimation	Borne inf.	Borne sup.	Type	Niveau
1	XAfter	-0,19125	-0,40626	0,0160142	Normal	0,95
2	XAfter	-0,19125	-0,39958	0,0071157	De base	0,95
3	XAfter	-0,19125	-0,38961	0,0170906	Centile	0,95

5.1.3 Coût moyen par réclamation au Québec

Ici, nous appliquons le modèle (2) aux coûts moyens trimestriels par réclamation au Québec. Compte tenu de la tendance et de la saisonnalité observées, il n'y a pas assez de preuves de l'évolution des coûts moyens après la décriminalisation du cannabis (tableau 6, valeur p supérieure à 0,05). Voyez l'ajustement de ce modèle à la figure 6; toutes les valeurs observées appartiennent à l'intervalle de prédiction à 95 %.

Tableau 6 : Résultats de l'estimation des effets de la décriminalisation du cannabis sur le coût moyen par réclamation au Québec au moyen d'un modèle à tendance saisonnière

	Estimation	Erreur type	Valeur t	Valeur p
(Ordonnée à l'origine)	-331 924,566	79 125,587	-4,19491	0,0018435344
XAfter	12,824	97,316	0,13177	0,8977778234
Année	166,800	39,229	4,25193	0,0016843695
Trimestre 2	-450,174	73,391	-6,13388	0,0001106417
Trimestre 3	139,939	73,391	1,90675	0,0856636605
Trimestre 4	716,961	77,319	9,27280	0,0000031612

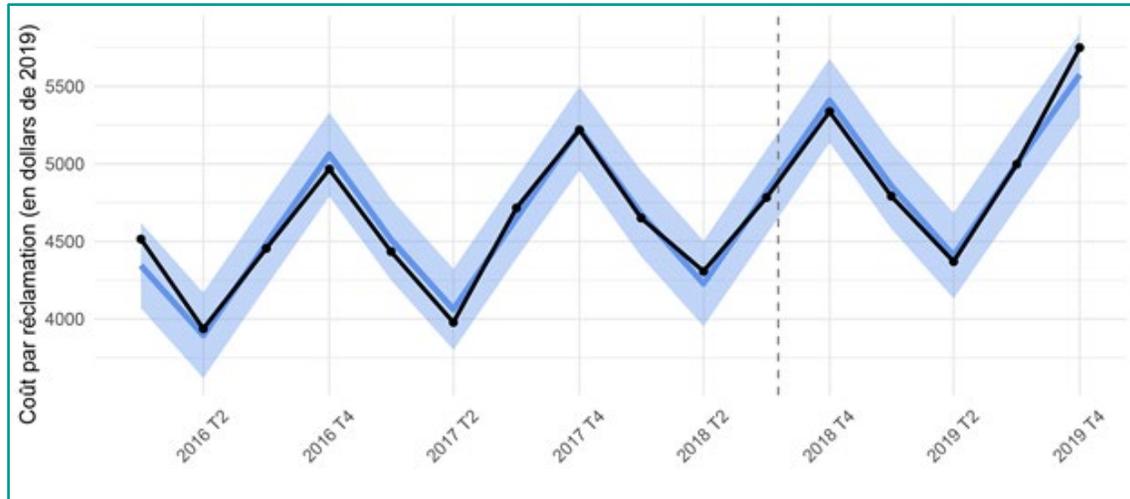


Figure 6 : Coût moyen trimestriel par réclamation au Québec (en dollars de 2019). Le noir correspond aux valeurs observées, le bleu, aux valeurs ajustées et aux intervalles de prédiction à 95 %. La ligne pointillée indique la date de décriminalisation du cannabis.

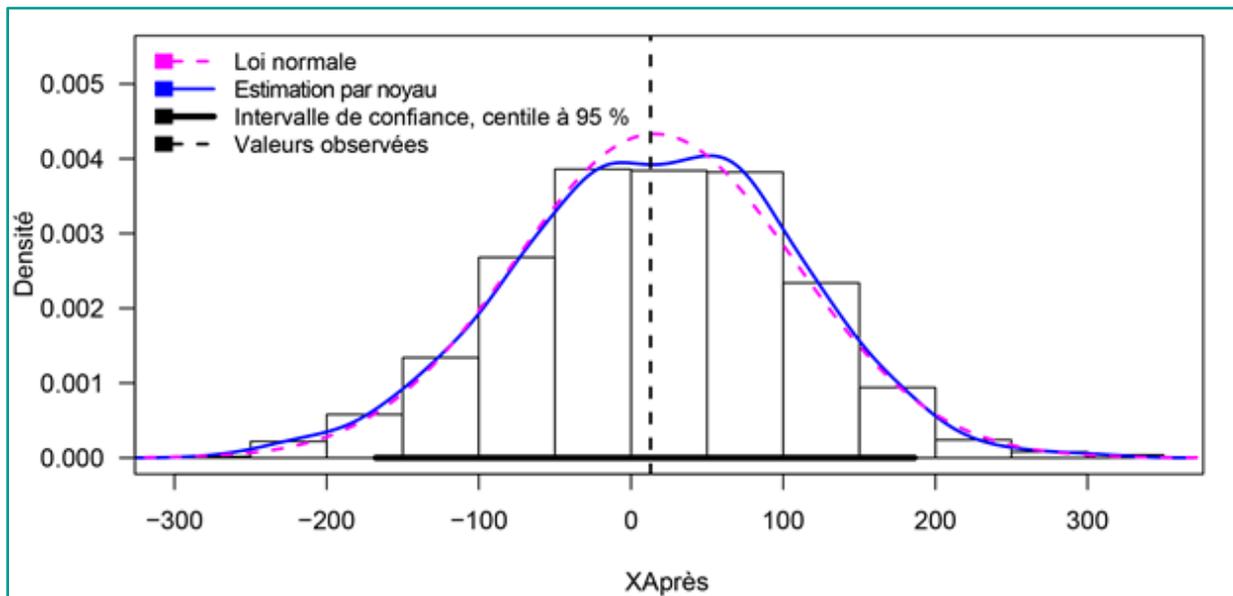


Figure 7 : Graphique de la densité de la distribution bootstrap du coefficient b dans le modèle linéaire des effets de la décriminalisation du cannabis sur le coût moyen par réclamation au Québec.

La figure 7 montre la distribution *bootstrap* du paramètre avec un intervalle centile. Toutes les versions des intervalles à 95 % contiennent zéro; par conséquent, il n'y a pas assez de preuves des effets de la décriminalisation : intervalle normal (-170,50; 190,57), intervalle de base (-160,59; 193,20) et intervalle centile (-167,55; 186,24).

En outre, un modèle restreint (sans la variable indicatrice représentant les périodes avant/après),

$$y_{t'} = a + ct' + \sum_{i=1}^{s-1} d_i S_{it'} + e_{t'} \quad (6)$$

a été estimé à l'instant t' avant la date de décriminalisation, et des données de sortie similaires ont été obtenues de ce modèle (tableau 7, figure 8). La figure 8 montre que les tendances avant la décriminalisation extrapolées dans le futur peuvent prédire avec précision les coûts moyens par réclamation, sans modéliser les effets de la décriminalisation.

Tableau 7 : Résultats de l'estimation du coût moyen par réclamation au Québec obtenue au moyen du modèle restreint à tendance saisonnière et des données avant la décriminalisation

	Estimation	Erreur type	Valeur t	Valeur p
(Ordonnée à l'origine)	-293 708,61	76 830,157	-3,8228	0,00873184
Année	147,86	38,091	3,8818	0,00815333
Trimestre 2	-459,57	79,293	-5,7958	0,00115583
Trimestre 3	117,92	79,293	1,4871	0,18754294
Trimestre 4	633,56	90,675	6,9871	0,00042774

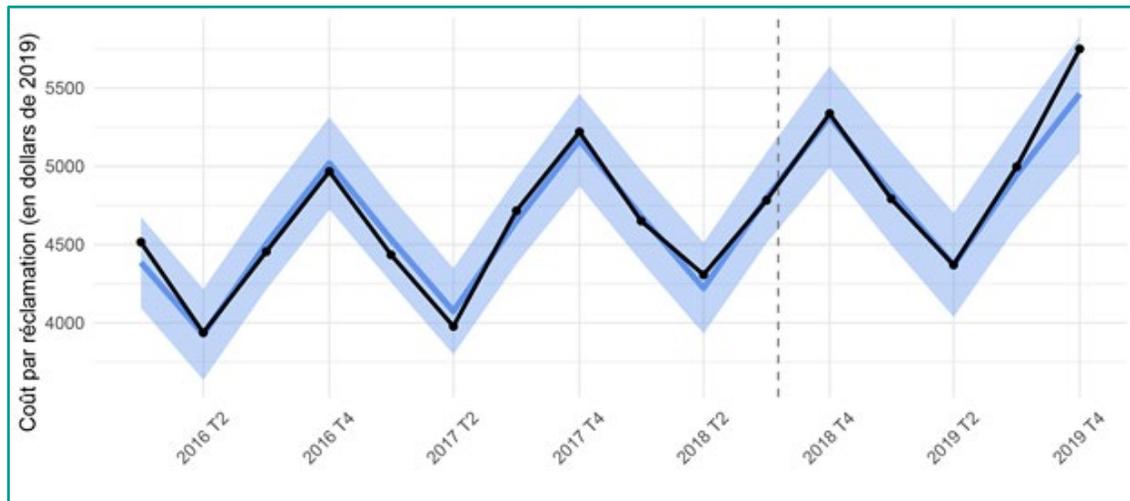


Figure 8 : Coût moyen trimestriel par réclamation au Québec (en dollars de 2019). Le noir correspond aux valeurs observées, le bleu, aux valeurs ajustées (modèle restreint à tendance saisonnière basé sur les données avant la décriminalisation) et aux intervalles de prédiction à 95 %. La ligne pointillée indique la date de décriminalisation du cannabis.

5.1.4 Fréquence des réclamations au Québec

Ici, nous appliquons le modèle (2) à la fréquence trimestrielle des réclamations au Québec. Les résultats reflètent ceux de la section précédente relatifs au coût moyen par réclamation.

Compte tenu de la tendance et de la saisonnalité observées, il n'y a pas assez de preuves de l'évolution des fréquences des réclamations après la décriminalisation du cannabis (tableau 8, valeur p supérieure à 0,05). Voyez l'ajustement de ce modèle à la figure 9; toutes les valeurs observées appartiennent à l'intervalle de prédiction à 95 %.

Tableau 8 : Résultats de l'estimation des effets de la décriminalisation du cannabis sur la fréquence des réclamations au Québec au moyen d'un modèle à tendance saisonnière

	Estimation	Erreur type	Valeur t	Valeur p
(Ordonnée à l'origine)	-167,016676	206,32901	-0,80947	0,437082930
XAfter	-0,217582	0,25376	-0,85743	0,411287857
Année	0,085714	0,10229	0,83791	0,421654958
Trimestre 2	-1,327500	0,19138	-6,93659	0,000040119
Trimestre 3	-1,675000	0,19138	-8,75239	0,000005314
Trimestre 4	-1,018104	0,20162	-5,04968	0,000499480



Figure 9 : Fréquence trimestrielle des réclamations par 100 années-véhicules acquis au Québec. Le noir correspond aux valeurs observées, le bleu, aux valeurs ajustées et aux intervalles de prédiction à 95 %. La ligne pointillée indique la date de décriminalisation du cannabis.

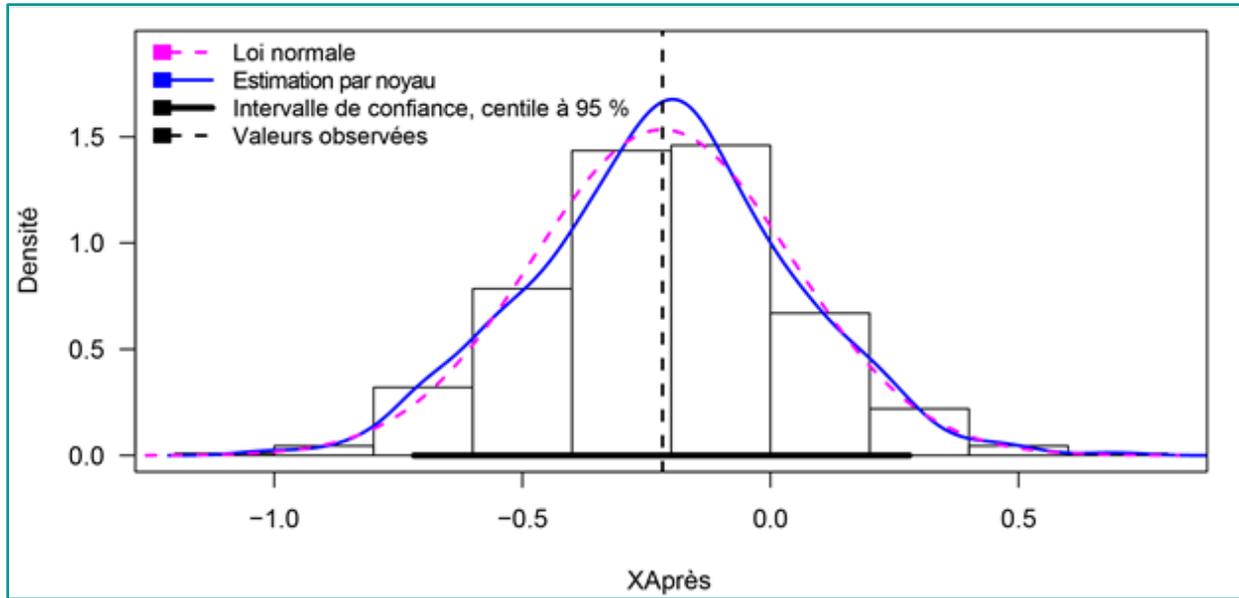


Figure 10 : Graphique de la densité de la distribution bootstrap du coefficient b dans le modèle linéaire des effets de la décriminalisation du cannabis sur le coût moyen par réclamation au Québec.

La figure 10 montre la distribution *bootstrap* du paramètre avec un intervalle centile. Toutes les versions des intervalles à 95 % contiennent zéro; par conséquent, il n'y a pas assez de preuves des effets de la décriminalisation : intervalle normal (-0,7272; 0,2924), intervalle de base (-0,7148; 0,2828) et intervalle centile (-0,7179; 0,2797).

En outre, des résultats similaires ont été obtenus d'un modèle restreint (6) sans la variable indicatrice représentant les périodes avant/après (tableau 9, figure 11). La figure 11 montre que les tendances avant la décriminalisation extrapolées dans le futur peuvent prédire avec précision les réclamations.

Tableau 9 : Résultats de l'estimation de la fréquence des réclamations par 100 années-véhicules acquis au Québec au moyen du modèle restreint à tendance saisonnière et des données avant la décriminalisation

	Estimation	Erreur type	Valeur t	Valeur p
(Ordonnée à l'origine)	-240,96462	175,684328	-1,3716	0,21927370
Année	0,12231	0,087102	1,4042	0,20985112
Trimestre 2	-1,17333	0,181317	-6,4712	0,00064639
Trimestre 3	-1,49000	0,181317	-8,2177	0,00017525
Trimestre 4	-0,74385	0,207344	-3,5875	0,01153847



Figure 11 : Fréquence trimestrielle des réclamations par 100 années-véhicules acquis au Québec. Le noir correspond aux valeurs observées, le bleu, aux valeurs ajustées (modèle restreint à tendance saisonnière basé sur les données avant la décriminalisation) et aux intervalles de prédiction à 95 %. La ligne pointillée indique la date de décriminalisation du cannabis.

5.2 États-Unis

5.2.1 Appariement par score de propension

L'ASP a été effectué pour plusieurs autres ratios du nombre d'États témoins sélectionnés sur le nombre d'États ayant décriminalisé le cannabis. Sur la base des différences moyennes standardisées absolues de la figure 12, le meilleur appariement est obtenu au ratio 1:1 (les différences d'échantillon ajustées, qui correspondent aux États témoins sélectionnés et aux États ayant décriminalisé le cannabis, sont les plus faibles). L'effet de l'appariement est encore perceptible au ratio 3:1, mais au ratio de 5:1, qui est plus élevé, les États témoins sélectionnés n'obtiennent pas la similarité souhaitée dans les variables de base (figure 12). La distribution des scores de propension apporte des informations générales sur les comportements observés. Les scores de propension des États traités (États ayant décriminalisé le cannabis) varient d'environ 0,15 à 0,30 et il n'y a que quelques États non traités dont les scores de propension se situent dans cette fourchette. L'algorithme ASP étant chargé de sélectionner plus d'éléments à appairer à chaque État traité (à mesure que le ratio augmente), un plus grand nombre d'États dont les scores de propension se situent en dehors de la fourchette et autour de 0 sont sélectionnés comme groupes témoins. Voici les éléments appariés à chacun des États ayant décriminalisé le cannabis :

- État de la CA :
 - WV
 - FL, MT, WV
 - FL, MT, SD, TN, WV
- État du MA :
 - UT
 - KS, UT, VA
 - ID, IN, KS, UT, VA
- État du NV :
 - NY
 - AZ, NY, TX
 - AZ, NC, NE, NY, TX

L'appariement 1:1 semble donc optimal dans cette étude, car il produit des échantillons beaucoup plus homogènes (figure 12A) et il se trouve qu'il suit la méthode d'appariement stratifié (Damrongplasit et coll., 2010) selon laquelle les unités témoins sont sélectionnées dans la fourchette de scores de propension des unités traitées.

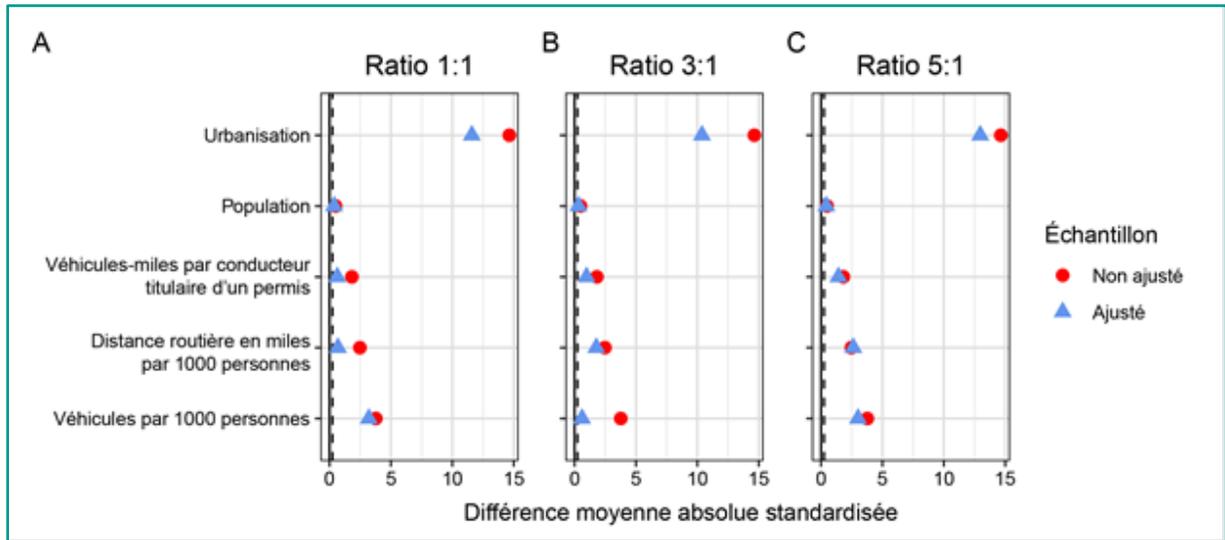


Figure 12 : Différences moyennes standardisées pour l'ASP avec différents ratios des États témoins sélectionnés sur les États traités. La ligne pointillée correspond au seuil de 0,25.

Appariement conventionnel

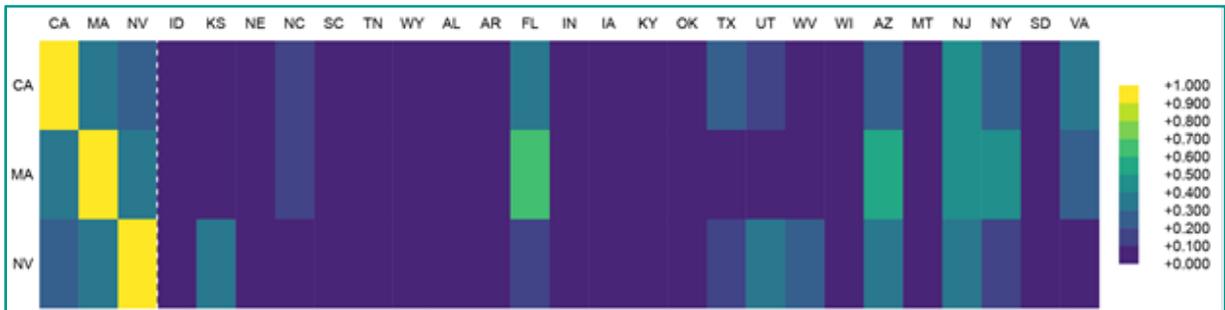


Figure 13 : Matrice de similarité basée sur une forêt aléatoire et sur les cinq variables relatives aux États.

Des similitudes entre les États obtenues d'une forêt aléatoire (figure 13) ont servi à trouver d'autres éléments à appairer. Chaque État appartenant au groupe « Cannabis légalisé » était apparié à l'État le plus similaire du groupe témoin si ce dernier n'avait pas été utilisé comme élément à appairer :

- État de la CA : NJ
- État du MA : FL
- État du NV : UT

5.2.2 Estimations des effets

Nous résumons ici les résultats de l'analyse du taux quotidien moyen de létalité et du nombre d'accidents mortels (provenant de l'ensemble de données du FARS) et du taux d'accidents de voiture (provenant de l'ensemble de données du CTAD). Tous les taux sont calculés à partir du nombre annuel de véhicules immatriculés dans les États visés par l'étude. Les prédicteurs suivants ont été pris en compte dans les modèles : Légalisation (variable binaire indiquant l'éventuelle décriminalisation du cannabis), État (variable catégorielle), Mois (variable catégorielle), Jour de la semaine (variable catégorielle), Jour férié (variable catégorielle), Température quotidienne moyenne, Plage de température quotidienne moyenne et Précipitation.

Nous avons estimé séparément les modèles dans le cas des États appariés (imitant en cela la méthode ASP stratifié de Damrongplasit et coll., 2010) en combinant tous les États sélectionnés, puis en répétant la procédure pour les États qui ont été appariés manuellement. Le tableau 10 résume les effets estimés de la décriminalisation; la figure 14 contient les GDP du taux de létalité (nous avons omis les GDP du nombre d'accidents mortels par manque d'espace et du fait que les résultats sont très semblables).

Tableau 10 : Estimations des effets de la décriminalisation basées sur la régression par forêts aléatoires; les valeurs *p*d'Altmann sont indiquées entre parenthèses. La dernière colonne des résultats exclut UT et utilise des sous-ensembles de ± 1 an à partir de la date de décriminalisation dans d'autres États.

États	Taux de létalité	Taux d'accidents mortels	Taux d'accidents de la route
CA, WV	-0,032 (1,00)	-0,015 (0,58)	2,444 (0,01)
MA, UT	-0,026 (0,88)	-0,020 (0,89)	–
NV, NY	-0,015 (0,24)	-0,018 (0,99)	-0,387 (0,02)
CA, WV, MA, UT, NV, NY	-0,011 (0,95)	-0,009 (0,99)	0,642 (0,01)
CA, NJ	-0,014 (0,97)	-0,015 (0,98)	1,196 (0,01)
MA, FL	-0,066 (0,01)	-0,060 (0,01)	-0,943 (0,01)
NV, UT	-0,015 (0,95)	-0,012 (0,81)	–
CA, NJ, MA, FL, NV, UT	-0,012 (0,50)	-0,012 (0,54)	0,321 (0,01)

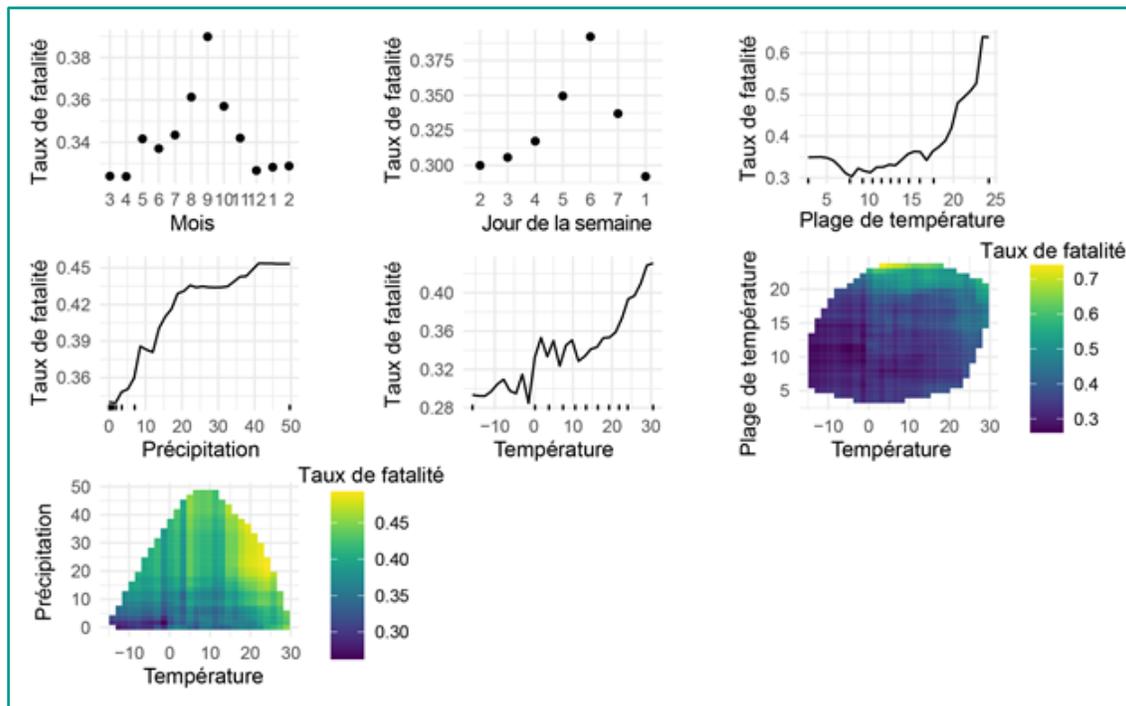


Figure 14 : GDP <https://doi.org/10.15288/jsad.2022.83.494>

6 Bibliographie

- Gelmi, T.J., Weinmann, W. et M. Pfäffl. « Impact of smoking cannabidiol (CBD)-rich marijuana on driving ability », *Forensic Sciences Research*, 2021, p. 1-13. <https://doi.org/10.1080/20961790.2021.1946924>
- Gupta, M., Pawar, N.M. et N.R. Velaga. « Impact of lockdown and change in mobility patterns on road fatalities during COVID-19 pandemic », *Transportation Letters*, vol. 13, 2021, p. 447-460. <https://doi.org/10.1080/19427867.2021.1892937>
- Hall, W. et M. Lynskey. « Assessing the public health impacts of legalizing recreational cannabis use: The US experience », *World Psychiatry*, vol. 19, 2020, p. 179-186. <https://doi.org/10.1002/wps.20735>
- Hartman, R.L., Brown, T.L., Milavetz, G. et coll. « Cannabis effects on driving longitudinal control with and without alcohol », *Journal of Applied Toxicology*, vol. 36, 2016, p. 1418-1429. <https://doi.org/10.1002/jat.3295>
- Hastie, T.J., Tibshirani, R.J. et J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2^e éd., Springer, New York, NY, 2009.
- Lane, T.J. et W. Hall, W. « Traffic fatalities within US states that have legalized recreational cannabis sales and their neighbours », *Addiction*, vol. 114, 2019, p. 847-856. <https://doi.org/10.1111/add.14536>
- Li, G., Brady, J.E. et Q. Chen. « Drug use and fatal motor vehicle crashes: A case-control study », *Accident Analysis & Prevention*, vol. 60, 2013, p. 205-210. <https://doi.org/10.1016/j.aap.2013.09.001>
- Lin, L., Shi, F. et W. Li. « Assessing inequality, irregularity, and severity regarding road traffic safety during COVID-19 », *Scientific Reports*, vol. 11, 2021, p. 1-7. <https://doi.org/10.1038/s41598-021-91392-z>
- Lopez Bernal, J., Cummins S. et A. Gasparrini. « The use of controls in interrupted time series studies of public health interventions », *International Journal of Epidemiology*, vol. 47, 2018, p. 2082-2093. <https://pubmed.ncbi.nlm.nih.gov/29982445/>
- Loy, A., Steele, S. et J. Korobova. *Lmeresampler: Bootstrap methods for nested linear mixed-effects models*, progiciel R, version 0.2.2, 2022. <https://github.com/aloy/lmeresampler>
- Moosavi, S., Samavatian, M.H., Parthasarathy, S. et coll. « Accident risk prediction based on heterogeneous sparse data: New dataset and insights », *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2019, p. 33-42
- NHTSA DOT. *Cannabis-Impaired Driving – A Report to Congress*, National Highway Traffic Safety Administration, département des Transports des États-Unis, Washington, D.C., 2017.
- Olmos, A. et P. Govindasamy. « Propensity scores: A practical introduction using R », *Journal of MultiDisciplinary Evaluation*, vol. 11, 2015, p. 68-88.
- Pearl, J. « [Bayesian analysis in expert systems]: Comment: Graphical models, causality and intervention », *Statistical Science*, vol. 8, 1993, p. 266-269.
- Peters, J., Janzing, D. et B. Schölkopf. *Elements of Causal Inference*, MIT Press, Cambridge, MA, 2017.
- Pollini, R.A., Romano, E., Johnson, M.B. et J.H. Lacey. « The impact of marijuana decriminalization on California drivers », *Drug and Alcohol Dependence*, vol. 150, 2015, p. 135-140. <https://doi.org/10.1016/j.drugalcdep.2015.02.024>
- Potter, A. et D.M. Weinstock. *High Time – The Legalization and Regulation of Cannabis in Canada*, McGill-Queen's University Press, Montréal, 2019.
- Rahman, M., Paul, K.C., Hossain, A. et coll. « Machine learning on the COVID-19 pandemic, human mobility and air quality: A review », *IEEE Access*, vol. 9, 2021, p. 72420-72450. <https://doi.org/10.1109/ACCESS.2021.3079121>

Romano, E., Torres-Saavedra, P., Voas, R.B. et J.H. Lacey. « Drugs and alcohol: Their relative crash risk », *Journal of Studies on Alcohol and Drugs*, vol. 75, 2014, p. 56-64. <https://doi.org/10.15288/jsad.2014.75.56>

Romano, E., Torres-Saavedra, P., Voas, R.B. et J.C Lacey. « Marijuana and the risk of fatal car crashes: What can we learn from FARS and NRS data? », *Journal of Primary Prevention*, vol. 38, 2017, p. 315-328. <https://doi.org/10.1007/s10935-017-0478-3>

Shi, T. et S. Horvath. « Unsupervised learning with random forest predictors », *Journal of Computational and Graphical Statistics*, vol. 15, 2006, p. 118-138. <https://doi.org/10.1198/106186006X94072>

Smith, E.P., Orvos, D.R. et J. Cairns Jr. « Impact assessment using the before-after-control-impact (BACI) model: Concerns and comments », *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 50, 1993, p. 627-637. <https://doi.org/10.1139/f93-072>

Stuart, E.A., Huskamp, H.A., Duckworth, K. et coll. « Using propensity scores in difference-in-differences models to estimate the effects of a policy change », *Health Services and Outcomes Research Methodology*, vol. 14, 2014, p. 166-182. <https://doi.org/10.1007/s10742-014-0123-z>

Windle, S.B., Eisenberg, M.J., Reynier, P. et coll. « Association between legalization of recreational cannabis and fatal motor vehicle collisions in the United States: An ecologic study », *Canadian Medical Association Open Access Journal*, vol. 9, 2021a, p. E233-E241. <https://doi.org/10.9778/cmajo.20200155>

Windle, S.B., Sequeira, C., Filion, K.B. et coll. « Impaired driving and legalization of recreational cannabis », *CMAJ*, vol. 193, 2021b, p. E481-E485. <https://doi.org/10.1503/cmaj.191032>

Wooldridge, J.M. *Introduction Econometrics: A Modern Approach*, 5^e éd., Cengage Learning, Mason, OH, 2013.

Wuertz, D., Setz, T. et Y. Chalabi. *timeDate: Rmetrics – Chronological and Calendar objects*, progiciel R, version 4021.104, 2022. <https://CRAN.R-project.org/package=timeDate>

Yasin, Y.J., Grivna, M. et F.M. Abu-Zidan. « Global impact of COVID-19 pandemic on road traffic collisions », *World Journal of Emergency Surgery*, vol. 16, 2021, p. 1-14. <https://doi.org/10.1186/s13017-021-00395-8>

Zhao, Q. et T. Hastie. « Causal interpretations of black box models », *Journal of Business & Economic Statistics*, vol. 39, 2021, p. 272-281. <https://doi.org/10.1080/07350015.2019.1624293>

Zhao, Q.-Y., Luo, J.-C., Su, Y. et coll. « Propensity score matching with R: Conventional methods and new features », *Annals of Translational Medicine*, vol. 9, n° 9, 2021. <https://doi.org/10.21037/atm-20-3998>

Zuur, A., Ieno, E.N., Walker, N.J. et coll. *Mixed Effects Models and Extensions in Ecology with R*, Springer, New York, NY, 2009.



7 À propos de l'auteur

Vyacheslav Lyubchich est professeur associé de recherche au University of Maryland Center for Environmental Science (UMCES). Il a obtenu un doctorat en statistique de l'Université d'État d'Orenbourg, en Russie, en 2011. La même année, il a reçu une bourse postdoctorale du gouvernement du Canada pour poursuivre ses recherches sur les méthodes de séries chronologiques au Département de statistique et d'actuariat de l'Université de Waterloo, au Canada. Depuis 2015, M. Lyubchich est professeur de recherche et membre fondateur de l'Environmental Statistical Collaborative de l'UMCES.



© 2022 Institut canadien des actuaires et Casualty Actuarial Society

Institut canadien des actuaires

360, rue Albert, bureau 1740

Ottawa, ON K1R 7X7

613-236-8196

siege.social@cia-ica.ca

cia-ica.ca

voiraudeladurisque.ca



L'Institut canadien des actuaires (ICA) est l'organisme de qualification et de gouvernance de la profession actuarielle au Canada. Nous élaborons et maintenons des normes rigoureuses, partageons notre expertise en gestion du risque et faisons progresser la science actuarielle pour améliorer la vie des gens au Canada et à l'échelle du monde. Nos plus de 6 000 membres utilisent leurs connaissances en mathématiques, en statistiques, en analyse de données et en affaires dans le but de prodiguer des services et des conseils de la plus haute qualité afin d'aider les personnes et les organisations canadiennes à faire face à leur avenir en toute confiance.

Casualty Actuarial Society

4350 N. Fairfax Drive, bureau 250

Arlington, VA 22203

<https://www.casact.org>

La Casualty Actuarial Society (CAS) est un organisme international de pointe en matière de titres de compétence et de formation professionnelle. Fondée en 1914, la CAS est le seul organisme actuariel au monde qui traite exclusivement de biens et de risques divers et qui dessert plus de 9 500 membres aux quatre coins du monde. Les membres de la CAS sont des experts en assurances IARD, réassurance, finance, gestion des risques et gestion du risque d'entreprise. Les professionnels et professionnelles formés par la CAS habilite les entreprises et le gouvernement à prendre des décisions stratégiques, financières et opérationnelles éclairées.