

Recueil du Symposium de 2024 de Statistique Canada : Le futur des statistiques officielles

Analyse du contrôle de la divulagation statistique pour l'estimation sur petits domaines

par Cissy Tang

Date de diffusion : le 8 septembre 2025



Statistique
Canada

Statistics
Canada

Canada

Analyse du contrôle de la divulgation statistique pour l'estimation sur petits domaines

Cissy Tang¹

Résumé

À l'heure actuelle, Statistique Canada n'a pas de directives officielles sur les règles de confidentialité pour la diffusion d'estimations sur petits domaines. Au cours des dernières années, il y a eu une demande croissante de la part des chercheurs des centres de données de recherche (CDR) pour l'élaboration de lignes directrices complètes en matière de confidentialité afin qu'ils puissent publier des estimations sur petits domaines dans leurs recherches. La présente analyse de confidentialité s'applique à l'estimation sur petits domaines au niveau de la région.

Une étude par simulation est menée dans R pour créer des populations simulées à partir desquelles les échantillons sont sélectionnés. La population simulée contient une variable auxiliaire, une variable d'intérêt et des informations sur le domaine. La force du lien entre la variable auxiliaire et la variable d'intérêt est contrôlée grâce à une variable « d'erreur » comportant une composante aléatoire. Des échantillons aléatoires stratifiés sont tirés et des estimations sur petits domaines au niveau de la région sont calculées à l'aide d'un paquet R « sae » (Molina et Marhuenda, 2015). La simulation est exécutée pour différents taux d'échantillonnage et divers niveaux de force du lien des variables auxiliaires et des variables de réponse, afin d'observer l'incidence sur le risque de divulgation et de déterminer les domaines qui pourraient présenter des risques de divulgation. On a comparé le risque de divulgation de l'estimation sur petits domaines avec celui de l'estimation directe de Horvitz-Thompson pour démontrer que les estimations sur petits domaines sont intrinsèquement moins risquées que les estimations directes, surtout lorsque les taux d'échantillonnage sont très faibles. Les résultats sont ensuite analysés et, enfin, des lignes directrices exhaustives en matière de confidentialité pour la diffusion d'estimations sur petits domaines au niveau de la région sont proposées. L'article décrit le processus de simulation et présente des justifications quant aux lignes directrices proposées en matière de confidentialité.

Mots-clés : Contrôle de la divulgation statistique; confidentialité; estimation sur petits domaines.

1. Introduction

1.1 Contexte

À l'heure actuelle, Statistique Canada n'a pas de directives officielles sur les règles de confidentialité pour la diffusion d'estimations sur petits domaines et aucune étude officielle n'a encore été menée sur le sujet. Au cours des dernières années, les chercheurs des centres de données de recherche (CDR) ont transmis un nombre croissant de demandes au Centre de confidentialité et d'accès aux données pour l'élaboration de lignes directrices complètes en matière de confidentialité, afin qu'ils puissent publier des estimations sur petits domaines dans leurs recherches.

Les nombres minimaux typiques de données d'enquête et de données administratives utilisés pour effectuer des estimations directes sont respectivement 5 et 10. Ces nombres minimaux correspondent au nombre minimal de répondants ou d'enregistrements nécessaires pour être en mesure de diffuser une fréquence pondérée (ou une fréquence non pondérée dans le cas de données administratives) et des statistiques telles que la moyenne, la variance, la proportion et le centile. Dans le cadre de l'estimation sur petits domaines (EPD), l'estimation est une combinaison linéaire d'une estimation directe et d'une estimation fondée sur un modèle qui est établie à partir de données auxiliaires. Par conséquent, on observe davantage de bruit dans l'estimation résultante que dans les estimations directes, ce qui suggère que les nombres minimaux de données utilisées pour effectuer une EPD ne devront pas être aussi stricts qu'ils le sont pour les estimations directes. De plus, la taille des domaines utilisés dans l'EPD peut être parfois si petite qu'on y retrouve très peu ou pas de répondants. Le risque de divulgation de l'identité, de divulgation d'attribut ou de divulgation par déduction ne serait toutefois pas nécessairement plus élevé dans une telle situation,

¹Cissy Tang, Statistique Canada, 100, promenade Tunney's Pasture (Ottawa), Canada, K1A 0T6
(cissy.tang@statcan.gc.ca)

comme ce serait le cas lorsque le nombre de répondants est faible dans l'estimation directe. Dans le cadre de l'EPD, lorsque le nombre de répondants dans un domaine est extrêmement faible, la variance de l'estimation directe est élevée. Une EPD peut ainsi s'appuyer davantage sur l'estimation fondée sur un modèle que sur l'estimation directe. Nous examinons ainsi, dans le cadre de la présente étude par simulation, la possibilité de supprimer totalement l'exigence relative au nombre minimal à inscrire pour diffuser une EPD. La suppression proposée de l'exigence relative au nombre minimal à utiliser pour effectuer une EPD permettra aux chercheurs d'examiner des sous-domaines de population rares lorsque la taille de l'échantillon direct n'est pas assez grande. La présente étude par simulation vise à créer des lignes directrices sur le contrôle de la divulgation pour les EPD, tant pour utiliser celles-ci dans les centres de données de recherche de Statistique Canada que pour diffuser des données officielles.

1.2 Types de divulgation statistique

Il existe trois principaux types de divulgation statistique. Ceux-ci servent de cadre pour évaluer le risque de contrôle de la divulgation dans le cadre de l'EPD. Le premier type est la divulgation de l'identité. Il y a divulgation de l'identité lorsqu'il est possible d'identifier un individu de la population à partir d'un tableau (Hundepool, 2024, p. 4). On observe généralement une telle situation en la présence d'une petite cellule comportant un ou deux répondants ayant une caractéristique unique. Dans le cadre de l'estimation directe, la mise en œuvre d'une exigence de nombre minimal, généralement « 5 » ou « 10 », empêche toute divulgation de l'identité. La divulgation d'attribut constitue le deuxième type de divulgation. Il y a divulgation d'attribut lorsque des caractéristiques au sujet d'une sous-population sont découvertes, et ce, même si l'identité des membres de la sous-population est inconnue (Hundepool, 2024, p. 214). La divulgation d'attribut survient généralement lorsqu'on diffuse des cellules vides ou des cellules pleines, ce qui permet aux intrus de découvrir que tout le monde ou personne dans une sous-population présente une certaine caractéristique. Pour éviter de divulguer des attributs, on peut ajouter du bruit afin de s'assurer de ne jamais diffuser des cellules vides ou pleines dans les tableaux statistiques. Le troisième type de divulgation est celui de la divulgation par déduction. Il y a divulgation par déduction lorsque des renseignements précédemment inconnus peuvent être déduits avec un degré élevé de certitude d'après des données diffusées, parfois en combinaison avec des données diffusées précédemment ou des données disponibles auprès d'autres sources (Hundepool, 2024, p. 218).

2. Étude par simulation

2.1 Portée

L'étude par simulation portera sur l'estimation sur petits domaines (Fay-Herriot) au niveau de la région d'une enquête probabiliste. Le meilleur prédicteur linéaire sans biais empirique (MPLSBE) de l'estimation sur petits domaines (Fay-Herriot) au niveau de la région consiste en un rapport entre l'estimation directe $\hat{\theta}$ et une estimation synthétique $z'_i\hat{\beta}$. Le coefficient est déterminé par la variance ainsi que l'erreur de l'estimation directe et de l'estimation synthétique, respectivement. Il s'agit de l'estimateur qui servira à évaluer le risque de divulgation dans le cadre de la diffusion d'une EPD au niveau de la région.

$$\hat{\theta}_i^{MPLSBE} = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) z'_i \hat{\beta}, \quad \hat{\gamma}_i = \frac{b_i^2 \hat{\sigma}_v^2}{b_i^2 \hat{\sigma}_v^2 + \hat{\psi}_i},$$

où $\hat{\theta}_i$ est l'estimation directe, $z'_i \hat{\beta}$ est l'estimation synthétique, $\hat{\sigma}_v^2$ est la variance de modèle, b_i sont des constantes fixes utilisées pour tenir compte de l'hétéroscédasticité et $\hat{\psi}_i$ est le prédicteur de la variance lissée du plan de sondage.

2.2 Méthodologie de la simulation

Un fichier de microdonnées contenant 50 groupes de population mutuellement exclusifs (les domaines) est simulé dans R. Des échantillons aléatoires stratifiés sont sélectionnés, en stratifiant selon les domaines. La variable d'intérêt est un indicateur qui est égal à 1 lorsque la personne possède la caractéristique, et égal à 0 dans le cas contraire. Pour évaluer son incidence sur le risque de divulgation, on utilise différentes fractions de sondage. En supposant un taux de réponse de 100 %, les estimations d'Horvitz-Thompson et les variances de strate sont calculées à l'aide du poid

de sondage de chaque unité échantillonnée. On lisse ensuite les variances de strate à l'aide d'une fonction logarithmique linéaire ordinaire, ce qui améliore la qualité des estimations sur petits domaines (Hidiroglou et You, 2024).

L'EPD repose sur la présence d'informations auxiliaires pour tous les domaines. Ces informations servent à établir un modèle pour les estimations synthétiques. Les informations auxiliaires sont importantes, surtout lorsque l'estimation directe est de mauvaise qualité. En pareil cas, l'estimation sur petits domaines repose davantage sur l'estimation synthétique que sur l'estimation directe. La variable auxiliaire sert à générer l'estimation synthétique sous la forme d'un modèle. Pour s'assurer de contrôler tous les aspects du modèle de régression linéaire, y compris l'ordonnée à l'origine, la pente et l'erreur, on génère d'abord la variable auxiliaire à l'aide de paramètres arbitraires. On utilise ensuite la variable auxiliaire pour générer le « vrai » taux de la caractéristique d'intérêt dans la population, lequel servira à simuler la population appariée à ces taux. Cette méthode permet de faire varier le degré de corrélation entre le « vrai » taux de la population et la variable auxiliaire en spécifiant l'ordonnée à l'origine, la pente et l'erreur afin de refléter les diverses situations à étudier. Autrement dit, on peut augmenter le terme d'erreur pour simuler une variable auxiliaire ayant une très faible capacité de prédiction ou le diminuer pour simuler une variable auxiliaire ayant une forte capacité de prédiction.

L'EPD est calculée à partir de l'estimation directe et de l'estimation synthétique mentionnées ci-dessus afin de l'utiliser dans l'évaluation du risque de divulgation. On a utilisé le paquet « sae » de R (Molina et Marhuenda, 2015) pour calculer l'EPD. La fonction utilisée pour calculer l'estimateur de Fay-Herriot est eblupFH(). La méthode d'estimation utilisée pour estimer l'erreur du modèle dans eblupFH() est la méthode du maximum de vraisemblance avec contrainte (MVC), laquelle produit des estimateurs sans biais des variances. Le tableau 2.2-1 résume les paramètres de simulation qui feront l'objet d'une analyse et d'une discussion à la section 3.

Tableau 2.2-1
Paramètres de la simulation et scénarios

Variables	Paramètre(s) de la simulation	
	Scénario A	Scénario B
Nombre de domaines	50	
Taille de la population	Choisie au hasard dans la fourchette [25000, 75000]	
Taux de variable auxiliaire (x)	Variable binaire au niveau de l'enregistrement dans chaque domaine généré pour apparier au taux du domaine (distribution uniforme où min = 0,03 et max = 0,1)	
Pente (b_1)	1,01	
Ordonnée à l'origine, (b_0)	0,005	
Erreur	Distribution uniforme avec...	
	... min = 0,005 et max = 0,05	... min = 0 et max = 0
Taux caractéristique de la population	$= b_1 x + b_0 + erreur$	
Fractions de sondage	0,1, 0,01, 0,001	

3. Résultats de la simulation et analyse

3.1 Résultats de la simulation et analyse

On exécute d'abord la simulation avec les paramètres du « scénario A » du tableau 1, en utilisant des fractions de sondage de 0,1, de 0,01 et de 0,001. Dans les quatre visualisations de données suivantes, on utilise un terme d'erreur généré de manière aléatoire pour effectuer l'estimation synthétique.

Au fur et à mesure que la fraction de sondage augmente, l'estimation de Horvitz-Thompson (directe) se rapproche de la valeur réelle de la population. La figure 3.1-1 est un histogramme des fréquences de la différence absolue entre les estimations d'Horvitz-Thompson (directes) et les vraies valeurs de la population utilisées dans la simulation.

Figure 3.1-1
Histogramme des fréquences de la différence absolue entre les estimations d'Horvitz-Thompson (directes) et les vraies valeurs de la population utilisées dans la simulation (scénario A)

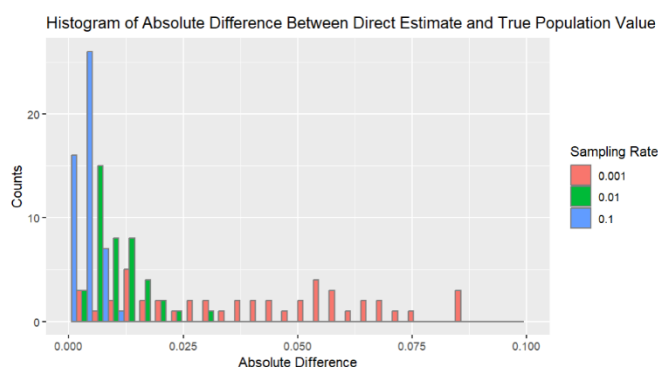
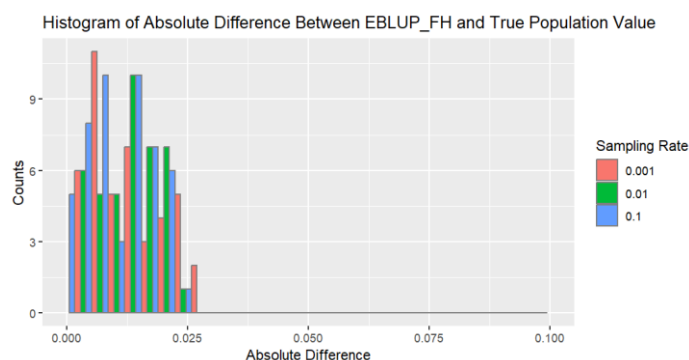


Figure 3.1-2
Histogramme des fréquences de la différence absolue entre les estimations de Fay-Herriot (petits domaines) selon l'approche MPLSBE et les vraies valeurs de la population utilisées dans la simulation



La figure 3.1-2 montre que l'EPD est plus efficace que l'estimation directe pour estimer le véritable paramètre de la population dans ce scénario de simulation. La visualisation des données de la figure 3.1-2 est un histogramme des fréquences de la différence absolue entre les estimations de Fay-Herriot (petits domaines) selon l'approche MPLSBE et les vraies valeurs de la population. L'axe des X est le même pour les deux histogrammes. La différence absolue dans le cas d'un taux d'échantillonnage = 0,001 est beaucoup plus faible lorsque la méthode d'EPD est utilisée plutôt que l'estimation directe.

L'EPD est beaucoup moins susceptible d'être identique ou proche de l'estimation directe au fur et à mesure que le taux d'échantillonnage diminue. La figure 3.1-3 montre la différence absolue entre l'estimation de Fay-Herriot (petits domaines) selon l'approche MPLSBE et l'estimation directe. Même lorsque la taille de l'échantillon est extrêmement réduite (moins de cinq répondants) pour un domaine donné, la diffusion de l'EPD ne révélera probablement pas la valeur de l'estimation directe. Par ailleurs, même s'il est impossible de garantir que l'EPD sera différente de l'estimation directe dans chaque simulation, selon la simulation, lorsque le taux d'échantillonnage diminue, la différence absolue entre l'EPD et les estimations directes augmentent, ce qui ajoute du bruit à l'estimation directe.

Figure 3.1-3
Histogramme des fréquences de la différence entre les estimations de Fay-Herriot (petits domaines) selon l'approche MPLSBE et les estimations Horvitz-Thompson (directes)

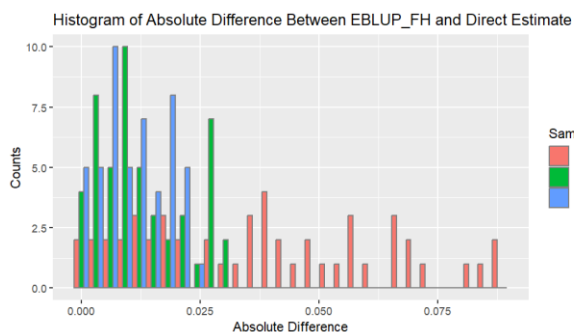
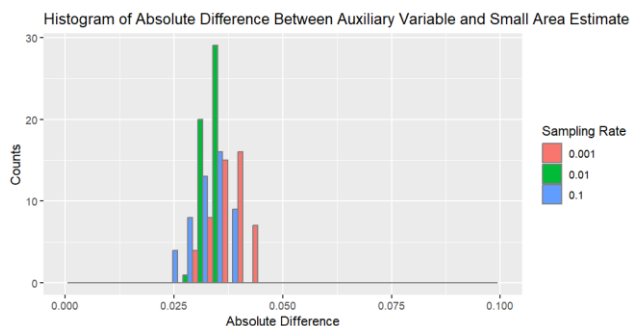


Figure 3.1-4
Histogramme des fréquences de la différence entre la variable auxiliaire et les estimations de Fay-Herriot (petits domaines) selon l'approche MPLSBE (scénario A)



L'étape suivante consiste à analyser l'incidence de la taille du terme d'erreur dans le modèle synthétique sur le risque de divulgation. Au fur et à mesure que la fraction de sondage diminue, l'EPD se rapproche de l'estimation synthétique. Par conséquent, si l'estimation synthétique est extrêmement proche ou identique à la variable auxiliaire, la diffusion de l'EPD pourrait présenter un risque de divulgation de la variable auxiliaire si la fraction de sondage est faible.

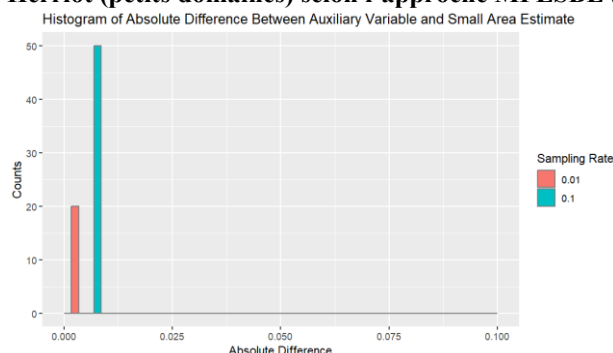
En ce qui concerne le scénario A, la figure 3.1-4 montre la différence absolue entre la variable auxiliaire et l'EPD en cas de composante d'erreur aléatoire dans l'estimation synthétique. Même si la différence absolue diminue à mesure que le taux d'échantillonnage augmente, il y a beaucoup de variation et de bruit. Il serait ainsi difficile de déduire à nouveau la variable auxiliaire d'après l'EPD.

Toutefois, si l'on supprime la composante d'erreur aléatoire présente dans l'estimation synthétique, l'estimation sur petits domaines s'approche beaucoup plus de la variable auxiliaire. Considérons maintenant une simulation similaire (scénario B, tableau 2.2-1) dans laquelle on supprime le terme d'erreur généré aléatoirement (erreur = 0 pour chaque enregistrement simulé).

Dans la figure 3.1-5, les histogrammes du taux d'échantillonnage = 0,001 n'ont pas été produits. Au cours d'une simulation aléatoire, un ou plusieurs domaines ne comportaient aucun répondant présentant la caractéristique ciblée. Ce n'était pas un problème pour le scénario A, car le terme d'erreur étant positif pour tous les domaines, les taux de la caractéristique d'intérêt dans le scénario A sont strictement supérieurs aux taux dans le scénario B. Ainsi, le scénario B était plus susceptible de tirer des échantillons qui aboutissent à une valeur d'estimation directe de 0 pour chaque domaine.

La fonction `eblup_FH()` du paquet R « sae » ne fonctionne pas lorsque l'estimation de la variance est nulle dans l'un des domaines. Lorsque le taux d'échantillonnage est si faible qu'on obtient des estimations nulles, la solution pratique à envisager consisterait à fixer la variance du domaine à un niveau si élevé que la fonction `eblup_FH()` serait utilisée de manière à produire l'estimation synthétique comme l'estimation sur petits domaines pour ce domaine. Dans la figure 3.1-5, même si les histogrammes du taux d'échantillonnage = 0,001 n'ont pas été produits, la différence absolue entre la variable auxiliaire et l'estimation sur petits domaines est déjà très proche de 0, même lorsque les taux d'échantillonnage sont relativement élevés à 0,01 et 0,1. La figure 3.1-5 montre que lorsque la variable auxiliaire a un pouvoir prédictif extrêmement élevé pour la variable d'intérêt, la diffusion de l'estimation sur petits domaines présente un risque de divulgation pour la variable auxiliaire.

Figure 3.1-5
Histogramme des fréquences de la différence absolue entre la variable auxiliaire et les estimations de Fay-Herriot (petits domaines) selon l'approche MPLSBE (scénario B)



3.2.1 Risque de divulgation lié à l'estimation directe

La simulation appuie l'hypothèse selon laquelle, lorsque très peu de répondants (moins de cinq) sont échantillonnés dans un domaine, il est peu probable que l'EPD de ce domaine révèle des renseignements pouvant être attribués à des répondants réels. Par conséquent, le risque de divulgation de l'estimation sur petits domaines dans cette situation est faible et il n'est pas nécessaire d'imposer un nombre minimal au moment de diffuser une EPD.

Un cas particulier à prendre en considération serait celui où l'estimation directe est égale à 0 ou n'existe pas pour un domaine. Dans la pratique, il s'agirait d'un domaine dans lequel le nombre de répondants est si faible que l'estimation directe est égale à 0, c'est-à-dire dans lequel aucun répondant ne présente la caractéristique en question. Dans le cadre de l'estimation directe classique, on ne diffuserait pas une cellule nulle, car cela pourrait entraîner une divulgation d'attribut. Toutefois, dans le cadre de l'EPD, lorsque la variance de l'estimation directe est égale à 0, la fonction eblupFHQ ne fonctionne pas comme prévu et doit être utilisée en fixant la variance à un nombre très élevé pour produire l'estimation sur petits domaines souhaitée, qui est égale à l'estimation synthétique. Par conséquent, lorsqu'il n'y a aucun répondant présentant la caractéristique d'intérêt dans le domaine cible de l'échantillon direct, il est peu probable que l'estimation sur petits domaines entraîne une divulgation d'attribut par le biais de sa composante d'estimation directe. Ainsi, dans des cas comme celui-ci, il n'est nécessaire d'évaluer que le risque de divulgation de l'estimation synthétique.

Il serait également prudent de ne pas diffuser l'estimation directe et les nombres non pondérés dans l'échantillon en même temps que l'estimation sur petits domaines. Cela rehaussera le niveau d'incertitude et constituera une mesure supplémentaire de protection.

3.2.2 Risque de divulgation lié à l'estimation indirecte

Selon la simulation, l'estimation sur petits domaines est identique ou extrêmement proche de l'estimation synthétique lorsque l'estimation directe est composée de très peu de répondants. Il est donc important de s'assurer que l'estimation synthétique n'est pas divulgatrice en soi.

Un problème de divulgation que l'on rencontre fréquemment dans la modélisation avec la régression linéaire est la présence d'un modèle saturé, c'est-à-dire un modèle dans lequel il existe un paramètre pour chaque combinaison unique de covariables. Un tel modèle permet de prédire parfaitement les variables auxiliaires. Dans la terminologie du contrôle de la divulgation statistique, les modèles saturés sont dits « équivalents aux tableaux », puisque le modèle lui-même suppose des tableaux descriptifs des variables auxiliaires utilisées dans le modèle. Si le modèle utilisé pour prédire l'estimation synthétique est saturé, la diffusion d'une estimation sur petits domaines pour un domaine comprend la diffusion d'informations auxiliaires sous forme de tableau. Dans ce scénario, en supposant qu'un nombre minimal de répondants n'est pas imposé pour diffuser l'EPD, la diffusion de données sur un domaine ne comporte aucun ou seulement quelques répondants présentant la caractéristique d'intérêt pourrait entraîner une divulgation de l'identité ou d'attribut des variables auxiliaires associées aux répondants dans le domaine. Il s'agirait alors d'une divulgation de la variable auxiliaire, laquelle doit également rester confidentielle en vertu de la *Loi sur la statistique*. Il est donc recommandé de ne pas diffuser d'EPD comportant des estimations synthétiques dérivées d'un modèle saturé.

Un autre type de modèle dont il faut se méfier est celui des modèles de régression à fort pouvoir prédictif. Pour mesurer la force de la variable auxiliaire z_i aux fins de la prédiction de θ_i , Hidiroglou, Beaumont et Yung (2019) ont dérivé une formule pour calculer le coefficient R^2 . $0 < R^2 < 1$, où une valeur proche de 0 représente un modèle faible et une valeur proche de 1 représente un modèle fort. Comme l'indique Miller (2024), lorsque la valeur R^2 est très proche de 1, le lien entre la variable auxiliaire et le prédicteur est quasi déterministe. Ce type de modèle peut entraîner une divulgation par déduction. Si $\hat{\theta}_i$ est diffusé pour un domaine comptant très peu de répondants, de sorte que l'EPD repose presque entièrement sur l'estimation du modèle, la variable auxiliaire z_i pourrait alors être déduite avec une grande précision pour la population de ce domaine. Cela pose problème si la variable auxiliaire est considérée comme confidentielle, notamment dans le cas de données fiscales ou sanitaires. Pour réduire le risque de divulgation lié à ce scénario dans la pratique, la diffusion des estimations sur petits domaines peut être limitée lorsque le modèle utilisé pour calculer les estimations synthétiques présente un coefficient R^2 supérieur à 0,95 (Miller, 2024). Le seuil de 0,95

est choisi pour être cohérent avec la modélisation des règles de vérification de la confidentialité dans les centres de données de recherche.

4. Lignes directrices proposées en matière de confidentialité et conclusion

4.1 Lignes directrices proposées en matière de confidentialité pour les EPD

Compte tenu des résultats de la simulation, voici les lignes directrices que nous proposons en matière de confidentialité pour une estimation sur petits domaines au niveau de la région (Fay-Herriot) :

- Restreindre la diffusion des estimations sur petits domaines lorsque l'estimation synthétique est dérivée d'un modèle saturé.
- Restreindre la diffusion des estimations sur petits domaines lorsque l'estimation synthétique est dérivée d'un modèle dans lequel le coefficient R^2 est supérieur à 0,95.
- N'imposer aucune restriction sur le nombre minimal de répondants. C'est-à-dire qu'il n'y a pas de nombre minimal d'unités requis dans l'estimation directe pour diffuser une estimation sur petits domaines.
- Il est recommandé de ne pas diffuser l'estimation directe, les renseignements sur l'échantillonnage et les données descriptives non pondérées en même temps que les estimations sur petits domaines.

4.2 Conclusion

En conclusion, l'étude par simulation a démontré que les estimations sur petits domaines peuvent être diffusées en toute sécurité, même lorsque la taille de l'échantillon est extrêmement réduite, sauf lorsque l'estimation synthétique est basée sur un modèle saturé ou lorsque l'estimation synthétique a un pouvoir prédictif extrêmement élevé. Pour appliquer officiellement les règles de confidentialité pour les EPD au sein des centres de données de recherche (CDR) de Statistique Canada, il convient de consulter le Comité de la confidentialité des CDR et d'obtenir son approbation. Par ailleurs, il est primordial de consulter les utilisateurs d'EPD, tant au sein de Statistique Canada que dans le milieu universitaire externe, pour s'assurer que les règles sont claires, complètes et suffisamment efficaces avant de les mettre en œuvre officiellement.

Bibliographie

- Hidiroglou, M.A., Beaumont, J.-F. et Yung, W. (2019), "Élaboration d'un système d'estimation sur petits domaines à Statistique Canada". *Technique d'enquêtes*, Statistique Canada, Catalogue n° 12-001-X, 45(1). Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/pub/12-001-x/2019001/article/00009-fra.htm>.
- Hidiroglou, M. A. et You, Y. (2024), 'Application de méthodes de lissage de la variance due à l'échantillonnage aux fins d'estimation sur petits domaines', *recueil du Symposium de 2022 de Statistique Canada*.
- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Naylor, J., Nordholt, E. S., Seri, G., Wolf, P.-P. D., Tent, R., Młodak, A., Gussenbauer, J., et Wilak, K. (2024), *Handbook on Statistical Disclosure control*, Centre of Excellence on SDC.
- Miller, J. (2024) 'Disclosure Risk of Parametric Regression Output', article interne de Statistique Canada non publié.