

N° 18-001-X au catalogue
ISBN 978-0-660-70426-5

Rapports sur les projets spéciaux sur les entreprises

Améliorer les données pour le Canada rural : estimation sur petits domaines des possibilités de travail à distance

par Stanley Yu Su

Date de diffusion : le 22 avril 2024



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-514-283-9350 |

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « [Normes de service à la clientèle](#) ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté le Roi du chef du Canada, représenté par le ministre de l'Industrie, 2024

L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

This publication is also available in English.

Table des matières

Remerciements	5
Sommaire	6
1 Introduction	7
2 Domaine d'intérêt	8
3 Indicateur d'entreprise cible	9
4 Estimation directe et variance	10
4.1 Estimateur ponctuel direct	10
4.2 Estimateur de variance	10
5 Estimation sur petits domaines	11
5.1 Modèle au niveau du domaine pour estimer la proportion	11
5.2 Lissage de la variance	12
5.3 Données auxiliaires	14
5.4 Progiciel R	15
5.5 Résultats du modèle	15
6 Évaluation du rendement du modèle et mise au point du résultat	16
6.1 Test diagnostique de l'estimation sur petits domaines	16
6.2 Établissement de données repères : Résultats de l'estimation sur petits domaines agrégée par rapport aux estimations directes	19
6.3 Estimation finale	20
7 Résultats analytiques	23
8 Conclusions et pistes de recherche futures	27
Références	29

Liste des tableaux

Tableau 1	Résultats du modèle d'EPD pour le premier trimestre, 2022	15
Tableau 2	EPD agrégée par rapport à l'estimation directe, premier trimestre de 2022.....	19
Tableau 3	Statistiques sommaires de l'estimateur final par région rurale et urbaine	25
Tableau 4	Statistiques sommaires de l'estimateur final en région rurale par grand secteur d'activité.....	26
Tableau 5	Tranche de 5 %, supérieure et inférieure des domaines ruraux figurant dans au moins trois trimestres de 2022	27

Liste des figures

Graphique 1	Variance échantillonnale et taille de l'échantillon, premier trimestre de 2022	11
Graphique 2	Variance de l'échantillon et variance lissée finale, premier trimestre de 2022.....	14
Graphique 3	Estimation directe et estimation composite du modèle d'EPD, premier trimestre de 2022	17
Graphique 4	Résidu normalisé du modèle et logarithme du risque prédit normalisé, premier trimestre de 2022.....	18
Graphique 5	Variance estimée moyenne, estimations directes et EPD, premier trimestre de 2022	19
Carte 1	Estimation directe et estimation finale pour le secteur des services, premier trimestre 2022	21
Carte 2	Estimation directe et estimation finale pour le secteur des biens, premier trimestre 2022.....	22
Graphique 6	Estimations finales réparties différemment selon les groupes sélectionnés	25

Remerciements

Les auteurs remercient Alessandro Alasia, Mahamat Hamit-Haggar, Anne Munro, Jacqueline Ha, Rachana Devkota, Mahdia Khodja et les personnes qui ont participé à l'examen par les pairs pour leurs commentaires et leur rétroaction inestimables.

Améliorer les données pour le Canada rural : estimation sur petits domaines des possibilités de travail à distance

par Stanley Yu Su

Sommaire

Le Canada rural est caractérisé par des collectivités dispersées et diversifiées ayant une faible densité de population. Les estimations générées au moyen des enquêtes sont généralement produites par région métropolitaine; mais dans les régions rurales, les échantillons sont habituellement trop petits pour permettre d'obtenir des estimations détaillées sur le plan géographique. Par conséquent, il y a souvent un manque de statistiques économiques et sociales à une petite échelle géographique. La diversité des conditions rurales peut alors être masquée par une moyenne agrégée pour les régions rurales.

La présente analyse illustre la façon de pallier la rareté des statistiques rurales en utilisant des approches d'estimation à partir des échantillons d'enquête actuels combinés à d'autres fonds de données existants. L'analyse porte sur l'application de méthodes d'estimation sur petits domaines (EPD) à certains indicateurs de télétravail produits à l'aide de l'Enquête canadienne sur la situation des entreprises (enquête trimestrielle). Le concept géographique utilisé pour obtenir des estimations détaillées pour les régions rurales est celui de la zone de travail autonome (ZTA), qui représente un marché du travail en région rurale ou une zone fonctionnelle rurale.

- Les méthodes d'EPD sont des techniques statistiques utilisées pour estimer les caractéristiques d'une population sur de petits domaines, comme des quartiers ou des régions rurales, où la taille des échantillons est généralement trop petite pour produire des estimations fiables. Dans ce cas, l'EPD permet de caler les modèles statistiques appropriés pour qu'ils couplent les données d'enquête avec les données auxiliaires disponibles pour l'ensemble de la population. L'EPD permet alors de produire des estimateurs de paramètres qui couvrent à la fois les domaines d'intérêt échantillonnés et ceux non échantillonnés.
- Dans la présente analyse, la variable d'intérêt est la proportion d'entreprises où le télétravail est possible pour un domaine donné, lequel correspond à une combinaison de région géographique et de grand secteur d'activité. Les données-échantillons sont tirées de l'Enquête canadienne sur la situation des entreprises, pour les quatre trimestres de 2022. Les sources de données auxiliaires sont le Fichier générique de l'univers de l'enquête, créé à partir du Système du Registre des entreprises de Statistique Canada, et le Recensement de la population de 2021.
- Pour chaque trimestre de 2022, les données d'enquête couvrent près de 570 domaines, comparativement aux 1 263 domaines disponibles dans les estimations finales du modèle d'EPD, ce qui représente un supplément important aux données d'enquête. Les domaines couvrent la quasi-totalité des régions géographiques au Canada, à savoir les régions métropolitaines de recensement et les agglomérations de recensement, qui représentent les régions urbaines et, pour les régions rurales, les ZTA, qui sont des groupes de subdivisions de recensement sont connectées entre elles par des déplacements importants.
- S'appuyant sur une couverture assez complète des régions géographiques canadiennes et une ventilation des grands secteurs d'activité des fournisseurs de services et des producteurs de biens, l'étude relève des éléments d'analyse clés sur le comportement des entreprises relativement aux modalités de télétravail et présente une tendance trimestrielle en 2022.
 - ▶ Les régions rurales sont aussi diversifiées que les régions urbaines pour ce qui est des entreprises offrant des possibilités de télétravail. Sans surprise, compte tenu de la diversité des conditions socioéconomiques et de la base d'activité dans les régions rurales, leurs marchés du travail dépendent de diverses conditions.
 - ▶ Comme le montrent les résultats regroupés pour les quatre trimestres et les domaines, les entreprises dans les régions rurales et urbaines et dans les secteurs des biens et des services ont eu des comportements différents en ce qui concerne les modalités de télétravail. Ces données mettent en évidence l'importance de faire une distinction entre les régions rurales et les régions urbaines, de même qu'entre les secteurs d'activité, lorsque l'on analyse l'intention des entreprises d'offrir des possibilités de télétravail.

- ▶ Les préférences des entreprises en ce qui concerne les modalités de télétravail offertes à leurs employés ont varié d'un trimestre à l'autre en 2022. Leur prévision de modalités de télétravail a changé au fil du temps, selon des facteurs comme les améliorations apportées par les entreprises à leurs conditions de sécurité technologique et numérique et la diminution de la gravité de la pandémie. Par conséquent, les tendances des entreprises peuvent varier au fil du temps en réponse à la situation courante.
- ▶ Selon l'évaluation par écart-type, les régions rurales présentaient moins de variabilité dans les estimations finales des proportions que les régions urbaines, pour tous les trimestres de 2022. Par conséquent, les possibilités de télétravail offertes sur les marchés du travail dans les régions rurales avaient tendance à être relativement plus stables que dans les régions urbaines.
- ▶ Pour les régions rurales, d'après la mesure de la valeur médiane, les proportions d'entreprises offrant des possibilités de télétravail dans le secteur des services étaient plus élevées que celles de leurs homologues du secteur des biens pour trois trimestres de 2022, et elles diminuaient légèrement d'un trimestre à l'autre.
- ▶ Dans une analyse fondée sur les centiles des domaines ruraux, les estimations finales de la tranche de 5 % supérieure et de la tranche de 5 % inférieure, groupées par trimestre, ont été calculées; la tranche supérieure étant définie comme les domaines les plus susceptibles d'offrir des possibilités de télétravail et la tranche inférieure étant définie comme les domaines les moins susceptibles de le faire. L'analyse révèle que certains domaines parmi les plus susceptibles et les moins susceptibles de le faire dans les régions rurales n'étaient pas les mêmes pour les quatre trimestres. En fait, les estimations trimestrielles présentaient des fluctuations entre les intervalles supérieur et inférieur, et la majorité (70 %) des entreprises de ces domaines appartenaient au secteur des services.

1 Introduction

Dans une enquête sur échantillon, la taille des échantillons n'est généralement pas suffisante pour générer des estimations directes de précision adéquate pour les petites régions ou les petits domaines. Ce constat représente un défi pour les régions rurales en particulier puisque leurs populations sont petites et dispersées sur le territoire. Dans ce cas, une estimation sur petits domaines (EPD) permet de caler les modèles statistiques appropriés pour qu'ils couplent les données d'enquête avec les données auxiliaires disponibles pour l'ensemble de la population. Autrement dit, l'EPD « emprunte la force » en utilisant les renseignements auxiliaires pour produire des estimations indirectes fiables.

Dans la présente étude expérimentale, l'application du modèle d'EPD au niveau du domaine a permis d'estimer la proportion d'entreprises où le télétravail est possible au niveau du domaine, pour chaque trimestre de 2022 de l'Enquête canadienne sur la situation des entreprises (ECSE). Dans le cadre de cette étude, un domaine correspond à une combinaison de région géographique et de grand secteur d'activité. La désagrégation par région géographique qui a été utilisée dans cette analyse permet de générer des données plus détaillées pour les régions rurales en utilisant un concept géographique nouveau, à savoir la zone de travail autonome (ZTA) (OCDE, 2020). Les régions géographiques (subdivisions de recensement [SDR]) sont classées comme des régions urbaines lorsqu'elles font partie d'une région métropolitaine de recensement (RMR) ou d'une agglomération de recensement (AR), et elles sont définies comme des régions rurales lorsqu'ensemble elles forment une ZTA.

La présente étude est motivée par deux éléments. Premièrement, les régions rurales présentent une diversité de conditions économiques et sociales; néanmoins, on reconnaît généralement qu'il y a souvent des lacunes en matière de renseignements actuels et détaillés sur le plan géographique pour ces régions. Deuxièmement, depuis le début de la pandémie de COVID-19, en mars 2020, les entreprises canadiennes ont pris des mesures pour offrir des formules de télétravail à leurs employés. Sur ce point, les entreprises situées dans les petites régions rurales et les régions urbaines peuvent avoir des comportements différents, et nous présumons que les entreprises rurales spécialisées dans le secteur des biens et le secteur des services agiront également différemment.

Le télétravail, un élément de cette diversité rurale/urbaine, a reçu une importante attention politique depuis le début de la pandémie. Au fil de l'évolution de pandémie, les entreprises ont ajusté leurs politiques pour offrir des possibilités de télétravail, et les changements apportés peuvent varier selon les domaines et les trimestres. La

présente étude s'intéresse donc à la proportion d'entreprises où le télétravail était possible, en ciblant les petites régions rurales et les différents trimestres.

Le présent article comporte plusieurs sections. Le domaine et l'indicateur d'entreprise d'intérêt provenant de l'ECSE sont définis et abordés dans la [section 2](#) et la [section 3](#). La [section 4](#) présente l'estimateur ponctuel direct et l'estimateur de variance. La [section 5](#) présente une description d'un modèle au niveau du domaine pour estimer la proportion, du lissage de la variance, des sources de données auxiliaires et du progiciel R utilisé pour caler le modèle d'EPD. Les résultats du modèle sont également analysés dans cette section. L'évaluation du rendement du modèle et les résultats analytiques des estimations finales sont examinés dans la [section 6](#) et la [section 7](#). La [section 8](#) conclut l'article en proposant des pistes de recherche futures.

2 Domaine d'intérêt

Comme le font remarquer Rao et Molina (2015), les domaines peuvent être définis par des régions géographiques ou des groupes sociodémographiques ou d'autres sous-populations. Un domaine est vu comme grand (ou important) si l'échantillon propre au domaine est suffisamment grand pour produire des estimations directes de précision adéquate. Un domaine est vu comme petit si l'échantillon propre au domaine n'est pas suffisamment grand pour permettre des estimations directes de précision adéquate.

Dans la présente étude, un domaine est défini comme une combinaison de région géographique et de grand secteur d'activité. Les types de régions géographiques sont les RMR, les AR et les ZTA rurales, qui sont toutes délimitées à partir des SDR détaillées, d'après la Classification géographique type de 2016¹.

Une RMR a une population totale d'au moins 100 000 habitants, dont 50 % ou plus doivent vivre dans le noyau. Quant à l'AR, son noyau doit compter au moins 10 000 habitants, mais elle n'a pas d'autre exigence de population totale ([Statistique Canada, 2022](#)). Les municipalités (SDR) appartiennent à une RMR ou à une AR si elles sont fortement intégrées au centre, comme le démontre une mesure des déplacements. Les ZTA représentent les régions géographiques qui n'appartiennent pas aux RMR ou aux AR. Selon Munro et coll. (2011), les ZTA sont créées en regroupant les SDR qui présentent d'importants déplacements réciproques entre elles. L'algorithme de groupement cible l'importance réciproque des déplacements provenant des données du Recensement de la population de 2016. L'algorithme au centre de la procédure de groupement montre un lien plus fort entre deux régions si les déplacements entre ces régions sont proportionnellement importants pour les deux régions. Il a des fonctionnalités propres qui le rendent utile aux fins de découverte des zones de travail rurales. La procédure délimite les ZTA comme étant formées d'au moins deux SDR où la majorité des travailleurs vivent et travaillent dans la même région, ce qui répond au critère d'autonomie. Selon la définition donnée par l'[Organisation de coopération et de développement économiques \(OCDE, 2020\)](#), l'autonomie comporte deux aspects : 1) l'autonomie des travailleurs (pourcentage des emplois de la région qui sont occupés par des résidents de la région); 2) l'autonomie des résidents (pourcentage des résidents de la région qui travaillent dans la région). Le seuil minimal d'autonomie est 75 %.

Les RMR et les AR sont essentiellement des régions urbaines, et les ZTA sont des régions rurales. Si une région chevauche deux provinces, on considère qu'il y a deux zones distinctes. Au total, il y a 36 RMR, 120 AR et 504 ZTA, qui sont délimitées en utilisant les 5 162 SDR comme unités élémentaires, d'après la Classification géographique type de 2016.

Du point de vue de la classification des industries, les entreprises enregistrées dans chaque SDR sont classées selon deux grands secteurs d'activité en fonction des codes du Système de classification des industries de l'Amérique du Nord (SCIAN)² : les producteurs de biens (SCIAN 11 à 33) et les fournisseurs de services (SCIAN 41 à 91). Chacun comptait certains sous-secteurs hors du champ de l'enquête couplés aux données d'enquête correspondantes. Au niveau supérieur, la combinaison de RMR, d'AR ou de ZTA et de grand secteur d'activité forme le domaine d'intérêt de la présente étude. Le nombre total de domaines est déterminé par les données disponibles dans le fichier de données auxiliaires qui est utilisé comme population d'entreprises. Cet aspect est abordé plus en détail à la [section 5.3](#). Au total, 1 263 domaines sont établis à partir des 640 régions géographiques.

1. [Statistique Canada. Classification géographique type \(CGT\) 2016 – Volume I, La classification.](#)

2. [Statistique Canada. Variante du SCIAN 2017 version 3.0 - Industries productrices de biens et services.](#)

3 Indicateur d'entreprise cible

L'ECSE a pour but de recueillir des renseignements sur les attentes et la situation des entreprises au Canada ainsi que sur les enjeux émergents. La population cible de cette enquête est l'ensemble des établissements actifs du Registre des entreprises (RE) ayant une adresse au Canada et comptant des employés. Les secteurs et sous-secteurs suivants du SCIAN sont exclus de la population cible. Par conséquent, la définition des grands secteurs d'activité exclut également ces catégories de cette étude :

- 22 : services publics
- 523990 : toutes les autres activités d'investissement financier
- 55 : gestion de sociétés et d'entreprises
- 611 : services d'enseignement
- 6214 : centres de soins ambulatoires
- 6215 : laboratoires médicaux et d'analyses diagnostiques
- 6219 : autres services de soins de santé ambulatoires
- 622 : hôpitaux
- 814 : ménages privés
- 91 : administration publique

La présente étude porte sur l'indicateur d'entreprise relatif aux modalités de travail, obtenu à l'aide de l'ECSE trimestrielle. Plus précisément, pour un domaine donné, quelle est la proportion d'entreprises où le télétravail est possible? Cette question cible n'est pas directement posée dans l'ECSE de 2022. Les questions de l'enquête étaient plutôt formulées ainsi ([Statistique Canada, questionnaire de l'ECSE](#)) :

Au cours des trois prochains mois, quel pourcentage des employés de cette entreprise ou cet organisme devrait accomplir chacune des tâches suivantes?

Exclure les employés attirés principalement à la conduite de véhicules ou à la livraison, et ceux qui travaillent principalement chez des clients.

Fournir votre meilleure estimation arrondie au pourcentage près. Si les pourcentages sont inconnus, laissez la question vide.

- a. Travailler exclusivement sur place
- b. Travailler la plupart des heures sur place
- c. Travailler autant d'heures sur place qu'à distance
- d. Travailler la plupart des heures à distance
- e. Travailler exclusivement à distance

Les répondants pouvaient donner un pourcentage à chaque option de réponse. Pour toute réponse valide, si les points de pourcentage de A et de B totalisaient jusqu'à 100 %, il en découlait que le télétravail n'était pas une possibilité dans l'entreprise échantillonnée. Alors, un indicateur « Non » était attribué. Autrement, un indicateur « Oui » sera codé. Cette procédure couvre les cinq variables numériques en une variable binaire, qui mène à une estimation de la proportion d'entreprises où le télétravail est possible, pour un domaine donné.

4 Estimation directe et variance

4.1 Estimateur ponctuel direct

L'estimateur ponctuel direct pour un domaine donné d est dérivé à l'aide des données-échantillons recodées et du poids d'échantillonnage de l'enquête. Il est donné par l'équation (1) ci-dessous :

$$\hat{P}_d^{DIR} = \frac{\sum_{i \in d} W_i I[\text{entreprise}_i = \text{Oui}]}{\sum_{i \in d} W_i} \quad (1)$$

où \hat{P}_d^{DIR} désigne l'estimateur ponctuel direct de la proportion d'entreprises où le télétravail est possible dans le domaine d , W correspond au poids d'échantillonnage de l'enquête, i désigne la i ème entreprise dans le domaine d et $I[\text{entreprise}_i = \text{Oui}]$ indique que la i ème entreprise est recodée « Oui » en variable binaire.

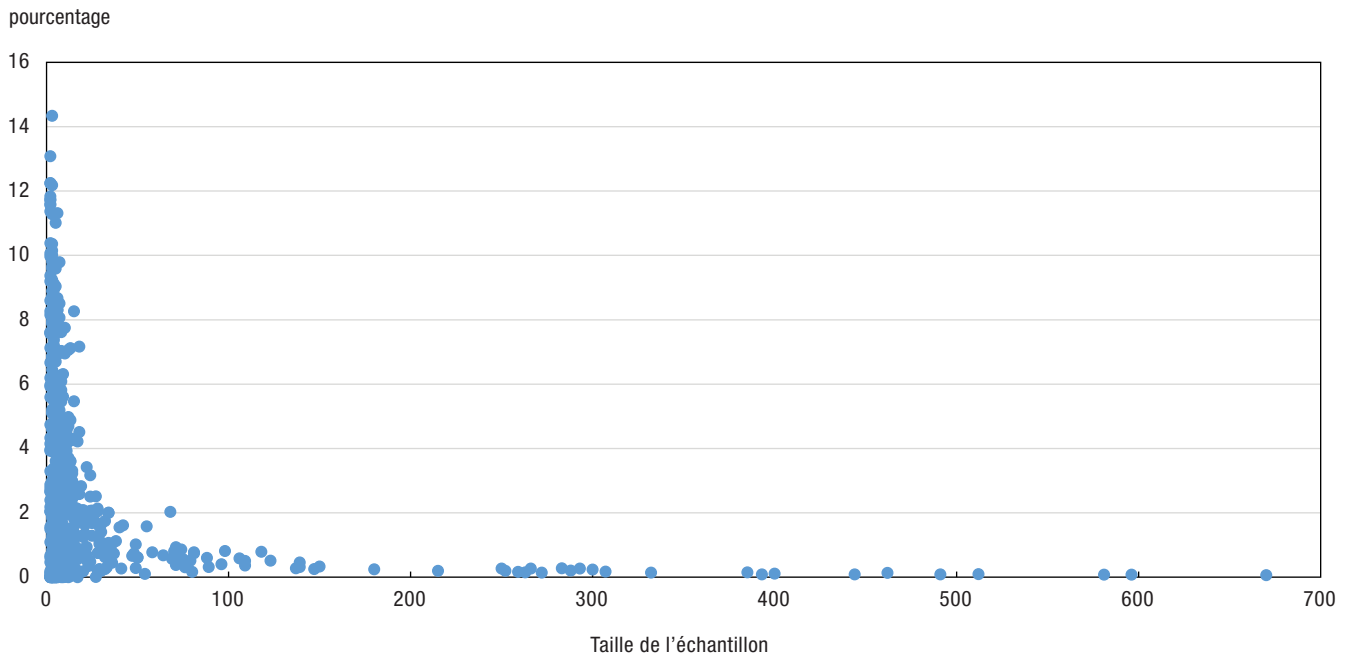
4.2 Estimateur de variance

Les estimations dérivées des échantillons sont sujettes à une erreur d'échantillonnage, mesurée communément par la variance échantillonnale, qui dépend de plusieurs éléments, dont la méthode d'échantillonnage, la méthode d'estimation, la taille de l'échantillon et la variabilité de la caractéristique estimée (Statistique Canada, 2021). L'ECSE s'appuie sur un échantillonnage aléatoire stratifié des établissements commerciaux classés par lieu géographique, secteur d'activité et taille (pour obtenir des renseignements détaillés, consulter [Statistique Canada, Base de métadonnées intégrée de l'ECSE](#)). La méthode applique un concept d'échantillonnage aléatoire simple stratifié à deux phases, où la deuxième phase tient compte de non-réponse à l'enquête. Dans la deuxième phase, la nouvelle pondération est effectuée sur les unités déclarantes de sorte que leurs facteurs de pondération finaux représentent encore la population cible entière. Après la nouvelle pondération, un processus de calage est mené pour que les totaux pondérés par classe de calage correspondent aux totaux de la population. Puisque l'estimateur ponctuel de la proportion, tel que donné par l'équation (1), est un estimateur non linéaire, et que le calage est utilisé, l'approche de linéarisation par série de Taylor³ est adoptée pour estimer la variance échantillonnale. L'estimation de la variance est calculée en utilisant le Système généralisé d'estimation (Statistique Canada, 2019).

L'estimation de la variance est une fonction de la taille de l'échantillon. Comme l'illustre la **graphique 1** en utilisant le premier trimestre de 2022, pour un domaine donné, plus la taille de l'échantillon est grande, plus la variance échantillonnale estimée est petite et, par conséquent, plus l'estimateur ponctuel sera précis.

3. À l'aide de la linéarisation, l'estimateur de calage non linéaire est approximé par une fonction linéaire, puis la variance de la fonction linéaire est obtenue (Statistique Canada, 2019).

Graphique 1
Variance échantillonnale et taille de l'échantillon, premier trimestre de 2022



Source : ECSE (premier trimestre de 2022).

5 Estimation sur petits domaines

Dans le contexte de la présente étude, l'EPD s'entend des techniques qui sont appliquées pour produire des estimateurs fondés sur un modèle. L'essence de l'EPD est le calage des modèles statistiques appropriés reliant les données d'enquête aux données auxiliaires disponibles au niveau d'un petit domaine. Selon Rao et Molina (2015), les modèles sont classés en deux grandes catégories : 1) les modèles au niveau du domaine qui apparentent les statistiques sommaires d'un petit domaine aux covariables au niveau du domaine des sources de données auxiliaires; 2) les modèles au niveau de l'unité qui apparentent les valeurs de l'unité d'une variable d'étude à des variables auxiliaires propres à l'unité. Différentes formes fonctionnelles peuvent être précisées, selon que le type de variable réponse est de nature continue ou discrète. Dans l'étude actuelle, la variable d'intérêt est la proportion d'entreprises, dans chaque domaine, où le télétravail est possible, par rapport au scénario opposé, soit les entreprises où le télétravail n'est pas possible. Puisqu'une loi binomiale s'applique, un modèle de régression logistique est utilisé.

5.1 Modèle au niveau du domaine pour estimer la proportion

Le modèle au niveau du domaine pour estimer la proportion est une application du modèle classique de Fay-Herriot (Fay et Herriot, 1979), où la variable dépendante est définie comme la transformation logarithmique additive des estimateurs directs. Ce modèle suit le cadre d'estimation établi par Esteban et coll. (2020), qui peut être précisé en deux étapes. À la première étape, le modèle d'échantillonnage est précisé :

$$\log\left(\frac{\hat{P}_d^{DIR}}{1 - \hat{P}_d^{DIR}}\right) = \mu_d + e_d \quad (2)$$

Où μ_d est la proportion de la population des entreprises dans un domaine d où une formule de télétravail est possible. Il dépend de paramètres de régression inconnus et des variables auxiliaires. e_d représente les erreurs d'échantillonnage qui sont présumées distribuées normalement, dont la moyenne est nulle, et une variance connue de $\hat{\sigma}_d^2$.

À la deuxième étape, le modèle de liaison est défini par :

$$\mu_d = X_d \beta + u_d \quad (3)$$

où X_d est un vecteur des variables auxiliaires mesurables pour chaque domaine d , β est le vecteur des coefficients de régression qui doivent être estimés et u_d est l'effet aléatoire propre au domaine, dont la moyenne est nulle et la variance est $\hat{\sigma}_u^2$.

En pratique, une estimation synthétique est obtenue en relevant les coefficients estimés $\hat{\beta}$ sur l'inverse de la transformation logarithmique additive de sorte que:

$$\hat{\mu}_d = \frac{e^{X_d \hat{\beta}}}{1 + e^{X_d \hat{\beta}}} \quad (4)$$

L'estimation composite est ensuite calculée en utilisant le mécanisme de liaison qui combine l'estimateur direct et l'estimateur synthétique :

$$\hat{P}_d^{SAE Composite} = \gamma_d \hat{P}_d^{DIR} + (1 - \gamma_d) \hat{\mu}_d \quad (5)$$

où γ_d désigne le facteur d'échelle qui lie la variance de l'estimation directe propre au domaine à la variance de l'effet aléatoire estimé ($\hat{\sigma}_u^2$) sur tous les domaines. Soit,

$$\gamma_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_d^2} \quad (6)$$

où $\hat{\sigma}_u^2$ est estimée en utilisant l'algorithme de maximum de vraisemblance restreint (Rao et Molina, 2015) pour calculer par approximation $\hat{\sigma}_u^2$ et $\hat{\sigma}_d^2$ correspond à l'estimateur de variance de l'échantillon. Pour un domaine où aucun échantillon n'est pris, γ_d se réduit à zéro et l'estimation composite devient l'estimation synthétique $\hat{\mu}_d$.

5.2 Lissage de la variance

Au niveau du domaine, la variance de l'échantillon peut être instable, plus particulièrement pour les domaines où la taille des échantillons est petite. Il est donc recommandé de lisser les variances de quelque façon (Bocci et coll., 2022). Les étapes présentées ci-après illustrent la façon d'appliquer le lissage de la variance dans la présente étude, selon les méthodes proposées par You et Hidirolou (2022).

Étape 1 : Utiliser la fonction de variance généralisée (GVF) et le facteur de correction de Rivest et Belmonte (2000).

Présumer qu'il y a m domaines et préciser le modèle log-log:

$$\log(\text{Sample.variance}_d) = \beta_0 + \beta_1 \log(\text{Sample.size}_d) + \varepsilon_d \quad (7)$$

À l'aide du modèle, l'estimateur lissé GVF naïf de la variance d'échantillonnage, \hat{V}_d^{naive} , est obtenu par $\exp(\hat{\beta}_0 + \hat{\beta}_1 \log(\text{Sample.size}_d))$ pour chaque domaine d . De plus, la variance résiduelle estimée du modèle est définie comme $\hat{\tau}^2$, où la valeur résiduelle du modèle pour le domaine d est $\log(\text{Sample.variance}_d) - (\hat{\beta}_0 + \hat{\beta}_1 \log(\text{Sample.size}_d))$.

La variance lissée corrigée en tenant compte de la variance résiduelle (proposée par Rivest et Belmonte [2000]),

$\hat{V}_d^{GVF.RB}$, est calculée par $\hat{V}_d^{naive} * \exp\left(\frac{\hat{\tau}^2}{2}\right)$.

Étape 2 : Utiliser un GVF et le facteur de correction proposé par Hidirolou et coll. (2019).

Dans cette autre approche, la variance lissée, $\hat{V}_d^{GVF.HBY}$, est calculé par $\hat{V}_d^{naïve} * Factor^{HBY}$, où le $Factor^{HBY}$ est

le facteur de correction sur tous les domaines, obtenu en calculant $\frac{\sum_{d=1}^m Sample.variance_d}{\sum_{d=1}^m \hat{V}_d^{naïve}}$.

Étape 3 : Utiliser la méthode d'effet de plan (DEFF).

Dans cette autre approche, il n'est pas nécessaire de caler le modèle log-log. Pour chaque domaine, l'effet de plan, $DEFF_d$, est donné par :

$$DEFF_d = Sample.variance_d * \left(\frac{\hat{P}_d^{DIR} (1 - \hat{P}_d^{DIR}) + Sample.variance_d}{Sample.size_d} \right)^{-1} * \left(\frac{Sample.size_d + 1}{Sample.size_d} \right) \quad (8)$$

où \hat{P}_d^{DIR} est l'estimateur ponctuel direct tiré de (1).

De plus, la moyenne de $DEFF_d$, désignée par \overline{DEFF} , est agrégée sous $\frac{1}{m} \sum_{d=1}^m DEFF_d$, et la moyenne de \hat{P}_d^{DIR} , désignée par \bar{P} , est $\frac{1}{m} \sum_{d=1}^m \hat{P}_d^{DIR}$. La variance lissée, \hat{V}_d^{DEFF} , est alors donnée par :

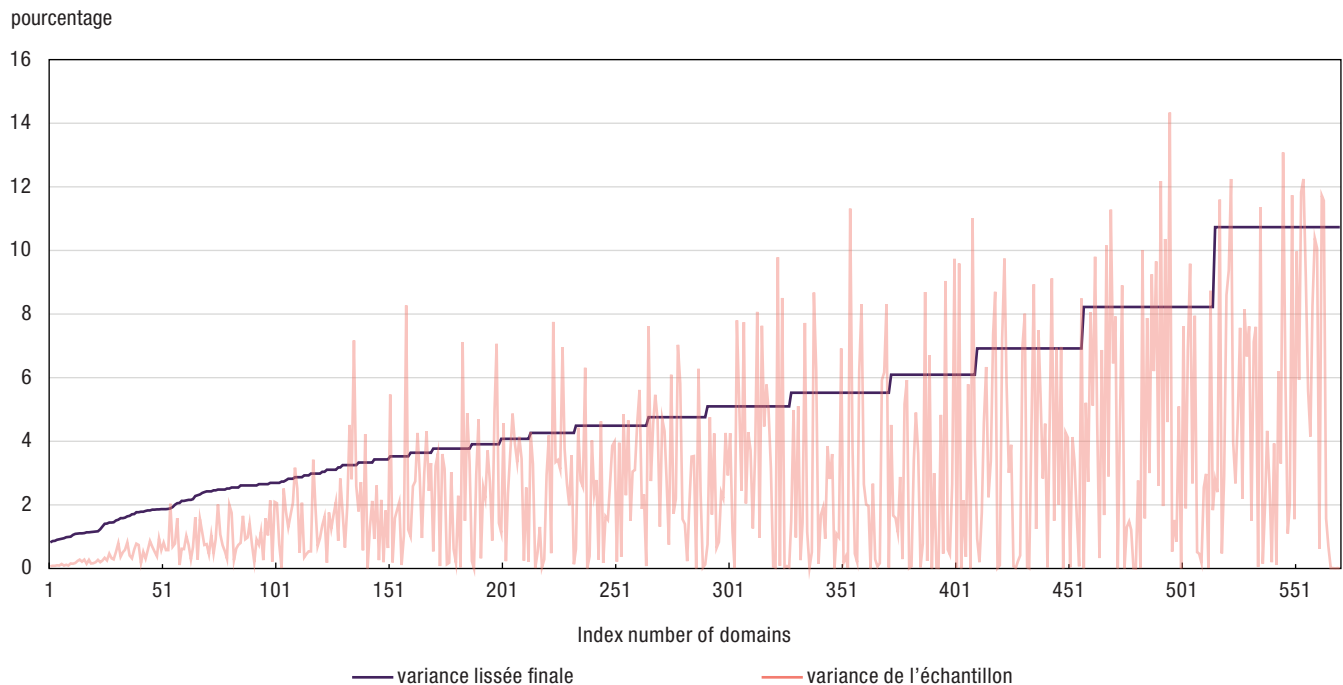
$$\hat{V}_d^{DEFF} = \overline{DEFF} * \frac{\bar{P}(1 - \bar{P})}{Sample.size_d} * \left(1 + \frac{1 - \overline{DEFF}}{Sample.size_d} \right)^{-1} \quad (9)$$

Étape 4 : Prendre la moyenne des résultats des trois approches présentées.

Dans cette application pratique, nous utilisons les estimations lissées GVF et DEFF, selon You et Hidirolou (2022), et l'estimation de variance lissée, \hat{V}_d^{FINAL} , est calculée par $\frac{1}{3} (\hat{V}_d^{GVG.RB} + \hat{V}_d^{GVG.HBY} + \hat{V}_d^{DEFF})$.

Les variances lissées finales sont portées sur un graphique en utilisant les variances calculées selon l'approche de linéarisation par série de Taylor (**graphique 2**), où l'axe X est l'indice des domaines et les variances lissées finales sont triées selon un ordre croissant. Le graphique montre que les variances lissées finales sont moins volatiles, par rapport aux variances originales. Les variances lissées finales seront alors utilisées comme $\hat{\sigma}_d^2$ dans le modèle d'EPD comme une des importantes entrées.

Graphique 2
Variance de l'échantillon et variance lissée finale, premier trimestre de 2022



Source : ECSE (premier trimestre de 2022), calculs de l'auteur.

5.3 Données auxiliaires

Après une analyse exploratoire minutieuse, les variables de deux ensembles de données sont sélectionnées comme données auxiliaires dans le processus d'EPD. Le premier correspond à l'ensemble de données de la population d'entreprises, qui est appelé le Fichier générique de l'univers de l'enquête (FGUE). Les données sont disponibles dans l'Environnement de fichiers couplables, qui est géré par le Centre des projets spéciaux sur les entreprises de Statistique Canada. Dans le cadre de la présente étude, les données V18 de l'année de référence 2021 sont utilisées. Ces données couvrent plus de 1,4 million d'entreprises ayant des employés, comprennent tous les niveaux d'entreprises, soit les organismes, les établissements, les sociétés et les emplacements. De cet ensemble, seuls les établissements s'inscrivant dans le champ de l'enquête⁴ (un peu plus de 1 million d'inscriptions) sont sélectionnés pour bâtir le fichier auxiliaire afin de constituer une population d'intérêt raisonnable.

Pour chaque domaine, l'âge de l'entreprise et la moyenne des ventes modélisées selon les recettes, assujetties à la taxe sur les produits et services (TPS) sont moyennés et pondérés par le nombre d'employés. De même, le carré de l'âge moyen de l'entreprise est utilisé pour saisir le motif de non-linéarité de l'effet de l'âge de l'entreprise sur les décisions d'affaires. Même après le groupement, les ventes moyennes assujetties à la TPS sont asymétriques et ne suivent pas une distribution normale. Le log des ventes est utilisé comme solution. De plus, le type de domaine est créé pour distinguer la zone géographique, soit une RMR, une AR ou une ZTA rurale. Comme déjà mentionné, une RMR et une AR définissent conjointement les régions urbaines, et les ZTA rurale représentent la région rurale. Finalement, un indicateur de grand secteur d'activité est généré pour subdiviser davantage une région géographique, à l'image de la division entre le secteur des biens et le secteur des services.

La deuxième source de données auxiliaires est le Recensement de 2021. Ce choix a été fait dans le but de tenter d'expliquer les comportements des entreprises d'un point de vue démographique. Pour chaque RMR, AR et ZTA rurale, les variables démographiques clés du Recensement de 2021 sont groupées. Ces données comprennent le pourcentage de jeunes (15 à 34 ans) dans la population totale en âge de travailler (15 à 64 ans), le pourcentage des

4. Certains secteurs du SCIAN sont exclus du fichier de population parce qu'ils ne s'inscrivent pas dans le champ d'enquête de l'ECSE. Voir [Statistique Canada, information détaillée de l'ECSE](#).

ménages vivant en appartement par rapport au nombre total des ménages et la moyenne du revenu d'emploi du ménage en milliers de dollars⁵.

Toutes ces variables forment le vecteur X_d de l'équation (3).

5.4 Progiciel R

L'EPD est traitée à l'aide du progiciel R « sae.prop » (Sholihin et Sumarni, 2022), qui implémente l'EPD en utilisant les modèles Fay-Herriot avec la transformation logistique additive (ALR). Plus précisément, la fonction mseFH.ns.uprop() est employée. Cette fonction comprend l'erreur quadratique moyenne bootstrap paramétrique de la méthode du meilleur prédicteur linéaire sans biais empirique fondée sur un modèle Fay-Herriot univarié avec l'ALR pour les données non échantillonnées.

Pour les domaines dont les estimations directes se situent entre 0 et 1, les estimations composites sont calculées à l'aide de l'équation (5). Pour les domaines qui ont une estimation directe extrême de 0 ou 1⁶ ou les domaines sans donnée-échantillon, les estimations synthétiques sont calculées à l'aide de l'équation (4). Pour la présente étude, 500 itérations bootstrap sont appliquées pour estimer l'EQM par domaine, permettant de calculer les erreurs types.

5.5 Résultats du modèle

Le modèle d'EPD est calé pour chaque trimestre de 2022, en utilisant le même ensemble de variables. Le résultat du modèle pour le premier trimestre est présenté dans le tableau 1.

Tableau 1
Résultats du modèle d'EPD pour le premier trimestre, 2022

Variables	Coefficient	Erreur type	Valeur t	Valeur p
Ordonnée à l'origine	5,1091**	2,429	2,103	0,036
Grand secteur (services)	0,2788*	0,145	1,922	0,055
Âge moyen de l'entreprise	-0,3356*	0,183	-1,839	0,066
Carré de l'âge moyen de l'entreprise	0,0082*	0,004	1,876	0,061
Moyenne de ventes assujetties à la TPS en logarithme	-0,1723**	0,083	-2,090	0,037
Pourcentage de jeunes (15 à 34 ans) par rapport à la population totale en âge de travailler (15 à 64 ans)	-3,8501**	1,534	-2,511	0,012
Pourcentage du nombre de ménages vivant en appartement par rapport au nombre total des ménages	1,7827**	0,827	2,156	0,031
Revenu moyen d'emploi du ménage en milliers de dollars	0,0112**	0,005	2,449	0,014
Type de domaine (RMR)	0,4695**	0,234	2,005	0,045
Type de domaine (ZTA)	0,2053	0,197	1,044	0,296

* significativement différent $p < 0,1$

** significativement différent $p < 0,05$

Source : ECSE (premier trimestre de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

D'après le modèle présenté, d'une part, les fournisseurs de services ont un $e^{0,2788}$ ou 1,32 fois plus de probabilités que les producteurs de biens d'offrir des possibilités de télétravail. Autrement dit, au premier trimestre de 2022, les entreprises du secteur des services avaient 32 % plus de probabilités que les entreprises du secteur des biens de proposer des modalités de télétravail (en gardant les autres variables constantes). Ce résultat est statistiquement significatif à un seuil de signification de 10 %, selon la valeur de p. Les activités du secteur des services, comme le traitement des données et les services bancaires, offrent un potentiel élevé de télétravail, et les employeurs sont plus susceptibles de proposer de telles modalités. D'autre part, le secteur des biens, comme le secteur manufacturier et les cultures agricoles, doit habituellement compter sur la présence physique des travailleurs et il est donc moins propice au télétravail.

Selon ce même raisonnement, à un seuil de signification de 5 %, les entreprises situées dans une RMR seraient 60 % ($e^{0,4695}-1$) plus susceptibles d'offrir des modalités de télétravail que la catégorie de référence des entreprises d'une AR. Les entreprises dans les régions rurales (ZTA) ont une probabilité d'offrir des modalités de télétravail qui est significativement différente de celle des AR.

5. Deux variables supplémentaires ont également été analysées, mais ont été retirées du modèle final pour des raisons de colinéarité multiple. Ces variables comprenaient le pourcentage de la population en âge de travailler (15 à 64 ans) ayant au moins un baccalauréat par rapport à la population totale en âge de travailler et le pourcentage des ménages comptant au moins quatre membres par rapport aux ménages totaux.

6. Les domaines où l'estimation directe extrême est 0 ou 1 ont une taille d'échantillon très petite et le progiciel R « sae.prop » les traite comme des domaines sans échantillon.

L'âge de l'entreprise, groupé au niveau du domaine, est associé négativement au logarithme du risque du revenu (c.-à-d., le télétravail est possible), mais cette relation n'est pas linéaire. Le logarithme du risque du revenu augmente à un taux décroissant lorsque la moyenne d'âge de l'entreprise augmente. Sur le plan quantitatif, une augmentation d'un an de la moyenne d'âge de l'entreprise est associée à une diminution de 29 % ($1 - e^{-0,3356}$) de sa probabilité d'offrir des possibilités de télétravail. Ce résultat suggère que les entreprises les plus anciennes sont moins susceptibles de favoriser les formules de télétravail, par rapport aux entreprises plus récentes. Toutefois, l'effet de l'âge sur le logarithme du risque du revenu semble être marginalement positif alors que l'entreprise avance dans le temps et qu'elle est mieux préparée sur le plan technologique à offrir des modalités de télétravail.

En maintenant les autres facteurs constants, une augmentation d'une unité des ventes assujetties à la TPS remise à l'échelle logarithmique est associée à une diminution de 16 % ($1 - e^{-0,1723}$) de la volonté des entreprises d'offrir des possibilités de télétravail. Ce constat est remarqué à un seuil de signification de 5 %. Les domaines où les ventes moyennes assujetties à la TPS sont plus élevées sont possiblement occupés par des entreprises ayant une culture traditionnelle qui accorde de l'importance au travail sur place pour favoriser la collaboration et l'innovation. Toutefois, sans information supplémentaire sur les caractéristiques de l'entreprise, il n'est pas possible de justifier une telle conclusion.

Du point de vue démographique, un pourcentage plus élevé de jeunes dans la population totale en âge de travailler est associé à une plus faible probabilité de possibilités de télétravail offertes par les entreprises. On estime qu'une augmentation de 1 % de la population de jeunes est associée à une diminution de 98 % ($1 - e^{-3,8501}$) de la probabilité que les entreprises offrent des possibilités de télétravail. Sur le plan du type de construction résidentielle, une augmentation de 1 % du pourcentage de ménages vivant dans un appartement est corrélée à une probabilité environ cinq fois plus élevée ($e^{1,7827} - 1$) de formule de télétravail, en maintenant toutes les autres variables constantes. Sur le plan du revenu, une augmentation de 1 000 \$ du revenu moyen d'emploi du ménage est associée à une augmentation de 1 % ($e^{0,0112} - 1$) de la probabilité de formule de télétravail.

L'interprétation illustrée ci-dessus vaut pour chaque variable, en tenant les autres variables constantes. Les estimateurs actuels sont le résultat de coefficients de notation de toutes les variables (y compris l'ordonnée à l'origine), calculées par le mécanisme de liaison illustré dans l'équation (5), s'il y a lieu. Également, le même ensemble de variables est utilisé pour caler des modèles pour les autres trimestres, où les seuils de signification ne sont pas nécessairement les mêmes que ceux du modèle du premier trimestre.

6 Évaluation du rendement du modèle et mise au point du résultat

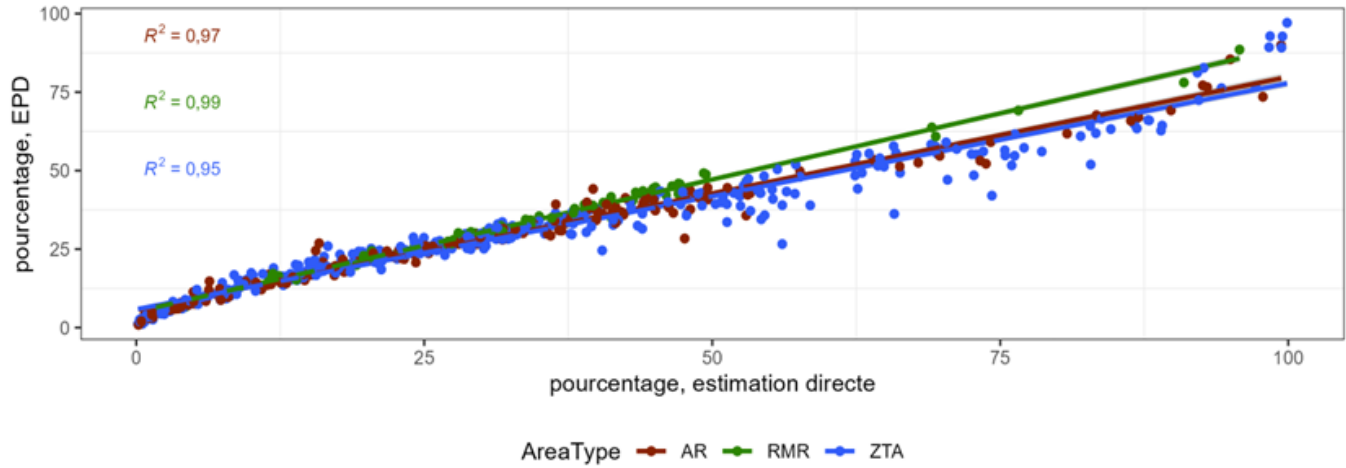
Dans la présente étude, l'évaluation du rendement du modèle d'EPD est menée en adoptant deux points de vue. Le premier consiste à faire un test diagnostique pour examiner le pouvoir prédictif du modèle et analyser les résidus. Le deuxième est un processus d'établissement de données repères selon lequel les estimations de l'EPD sont groupées au plus haut niveau (comme au niveau provincial), pour lequel les estimations directes sont plus fiables. Dans ce cas, les estimations directes sont utilisées comme valeurs de données repères auxquelles les estimations d'EPD groupées sont comparées. Une dernière mise au point est faite pour déterminer les estimations finales à produire.

6.1 Test diagnostique de l'estimation sur petits domaines

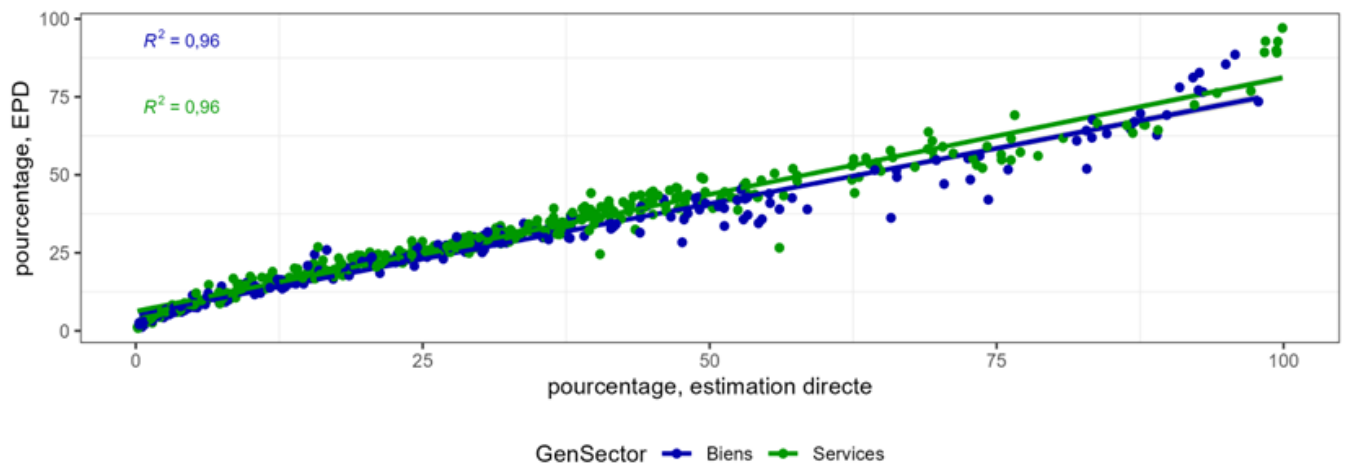
Les résultats de l'ECSE pour le premier trimestre de 2022 sont utilisés pour présenter le test diagnostique du rendement du modèle. Des 1 263 domaines fondés sur le fichier de la population d'entreprises (FGUE 2021 V18), 570 présentent des estimations composites du modèle d'EPD. Les nuages de points des estimations d'EPD par rapport aux estimations directes sont présentés à la **graphique 3**, groupées par type de domaine (graphique 3.A) et grand secteur d'activité (graphique 3.B), où R au carré mesure la proportion de la variation de l'estimation directe saisie par le modèle d'EPD.

Graphique 3
Estimation directe et estimation composite du modèle d'EPD, premier trimestre de 2022

A - Par type de domaine



B - Par grand secteur d'activité



Notes : AR $R^2=0,97$, RMR $R^2=0,99$, ZTA $R^2=0,95$.

Biens $R^2=0,96$, Services $R^2=0,96$.

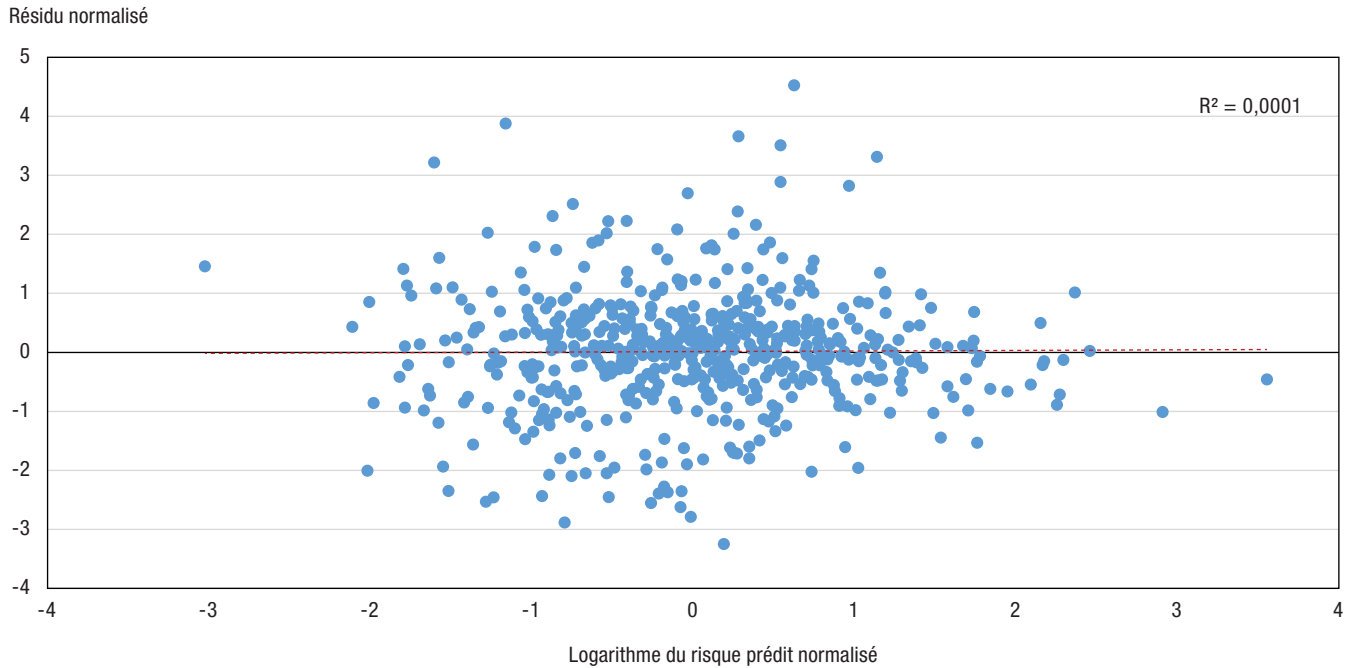
Source : ECSE (premier trimestre de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

Les couleurs de R au carré dans la graphique 3 sont associées aux différentes catégories examinées. Classé selon le type de domaine ou le grand secteur, le modèle d'EPD peut englober une grande proportion ($\geq 95\%$) de la variation de l'estimation directe. Le rendement du modèle est meilleur pour les domaines RMR que pour les domaines AR et ZTA, en raison de la grande taille des échantillons. Également, le modèle a le même pouvoir prédictif (voire supérieur) pour le secteur des biens et le secteur des services.

Pour l'analyse du résidu, le résidu du modèle est donné par $\log\left(\frac{\hat{P}_d^{DIR}}{1-\hat{P}_d^{DIR}}\right) - X_d\hat{\beta}$, où $X_d\hat{\beta}$ est le logarithme

du risque prédit. Le résidu normalisé du modèle est représenté par rapport au logarithme du risque normalisé (**graphique 4**). Le graphique indique que les résidus sont dispersés aléatoirement autour de zéro, sans motif remarquable pour la majorité des domaines, mis en évidence par la petite valeur de R au carré ($< 0,01$). Cette illustration suggère qu'il n'y a pas de preuve solide d'hétéroscédasticité. Le modèle linéaire généralisé est bien adapté aux données et les estimations d'EPD corroborent les déviations modérées des résidus par rapport à la normalité.

Graphique 4
Résidu normalisé du modèle et logarithme du risque prédit normalisé, premier trimestre de 2022

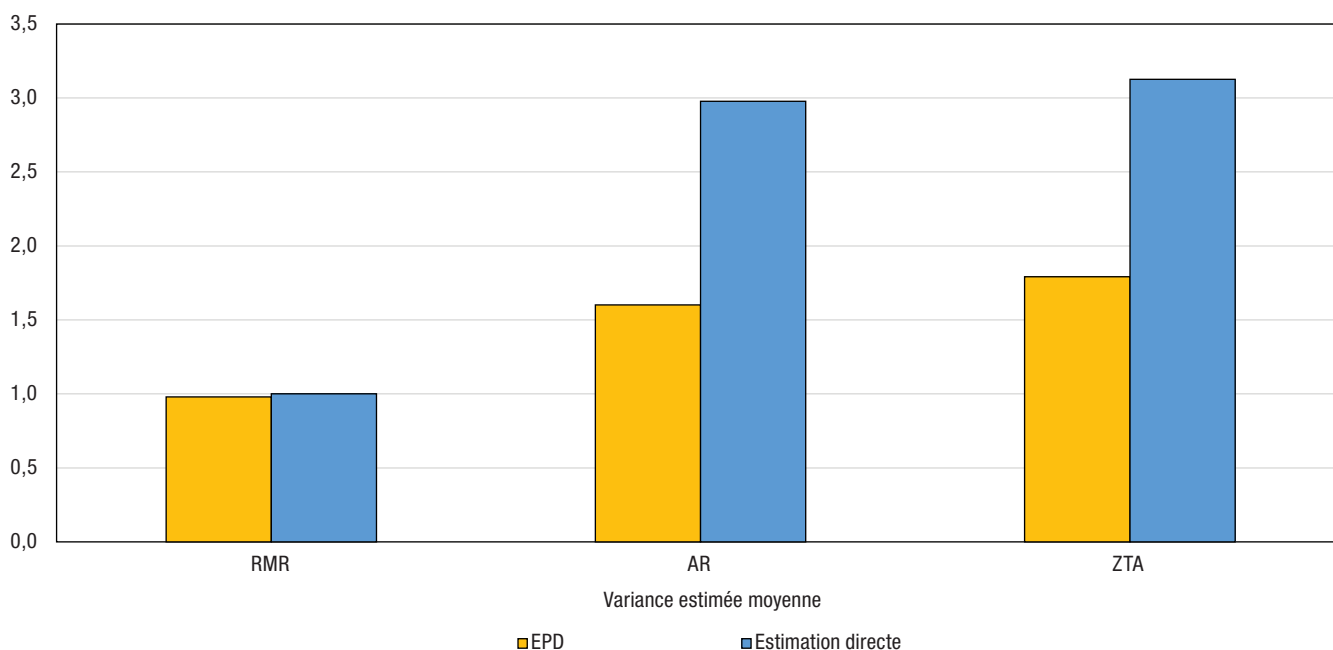


Note : La ligne pointillée rouge est la ligne d'ajustement linéaire associée à $R^2=0,0001$.
Source : ECSE (premier trimestre de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

De plus, la variance moyenne des estimations directes est comparée à la moyenne de l'EQM par rapport à l'EPD, par type de domaine (**graphique 5**). Pour les RMR où de grands échantillons sont disponibles, les estimations directes et l'EPD tendent à avoir des variances similaires (et faibles), ce qui suggère que les deux approches sont sujettes à de plus petites erreurs d'échantillonnage. Pour les AR et les ZTA, toutefois, l'EPD présente clairement une variance moyenne inférieure (EQM) par rapport aux estimations directes. Ce constat mène à la conclusion que les proportions estimées par l'EPD sont plus fiables que celles des estimations directes, pour les plus petits domaines.

Graphique 5**Variance estimée moyenne, estimations directes et EPD, premier trimestre de 2022**

pourcentage



Source : ECSE (premier trimestre de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

6.2 Établissement de données repères : Résultats de l'estimation sur petits domaines agrégée par rapport aux estimations directes

Pour le premier trimestre de 2022, les estimations du modèle d'EPD par domaine sont groupées au niveau provincial, au niveau territorial et national combiné, ainsi qu'aux niveaux urbain (RMR et AR) et rural (ZTA). Ainsi, pour chaque domaine détaillé, l'estimation du modèle d'EPD est multipliée par le nombre total d'entreprises dans le champ d'enquête pour calculer le nombre d'entreprises où le télétravail est possible. Le résultat est ensuite groupé à un niveau plus élevé et divisé par le nombre total d'entreprises dans le champ d'enquête à ce niveau afin de refaire la conversion pour obtenir une proportion. Cette proportion est ensuite comparée à l'estimation directe au même niveau, où la taille de l'échantillon est suffisamment grande pour produire une estimation fiable (**tableau 2**).

Tableau 2**EPD agrégée par rapport à l'estimation directe, premier trimestre de 2022**

	EPD agrégée			Estimation directe		
	Moyenne	Intervalle de confiance de 95 % ¹		Moyenne	Intervalle de confiance de 95 % ¹	
	pourcentage	de	à	pourcentage	de	à
Canada	36,7	32,5	39,5	36,8	35,0	38,7
Terre-Neuve-et-Labrador	30,9	26,8	33,4	28,1	24,7	31,6
Île-du-Prince-Édouard	27,9	18,2	34,0	27,8	24,2	31,5
Nouvelle-Écosse	30,2	21,3	36,6	31,9	28,3	35,2
Nouveau-Brunswick	26,7	21,0	31,4	25,6	22,6	29,1
Québec	35,5	28,8	38,7	34,1	30,7	37,3
Ontario	38,8	31,2	42,2	39,5	35,7	43,7
Manitoba	28,7	20,0	32,1	28,9	25,2	32,5
Saskatchewan	29,4	24,6	33,2	27,8	24,5	31,1
Alberta	37,6	29,8	42,4	38,9	34,7	43,0
Colombie-Britannique	38,6	27,4	44,8	38,6	34,6	42,5
Territoires	40,0	35,4	44,5	40,3	34,8	45,5
Région rurale	28,1	26,4	30,2	26,0	22,7	29,5
Région urbaine	38,6	34,0	41,1	39,1	37,1	41,1

1. Les intervalles de confiance sont obtenus à partir de 1 000 rééchantillonnages bootstrap.

Source : ECSE (premier trimestre de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

D'une part, les estimations agrégées du modèle d'EPD à l'échelon du Canada sont comparables aux valeurs repères des estimations directes. Une relation similaire est observée pour les régions urbaines. Ce résultat est attendu parce que, lorsque les échantillons sont suffisamment grands, l'estimation directe est fiable et l'EPD groupée reproduit essentiellement l'estimation directe. D'autre part, l'EPD est légèrement surestimée pour les régions rurales, ainsi que pour les provinces de l'Atlantique (à l'exception de la Nouvelle-Écosse), le Québec et la Saskatchewan. L'EPD est légèrement sous-estimée pour la Nouvelle-Écosse, l'Ontario et l'Alberta. De plus, l'EPD pour le Manitoba, la Colombie-Britannique et les trois territoires combinés (Nunavut, Territoires du Nord-Ouest et Yukon) est comparable aux estimations directes. Il n'est pas nécessaire de faire d'autres ajustements puisque les écarts entre les EPD et les estimations directes sont modestes et même proches pour les plus grands domaines (comme le Canada et les régions urbaines).

6.3 Estimation finale

Les estimations finales sont préparées en combinant trois sources de résultats. Parce que les échantillons des RMR sont relativement grands, les estimations directes et les estimations du modèle d'EPD pour ces domaines sont très proches (**graphique 3.A**) et les variances moyennes des estimations directes et des EPD sont également similaires (**graphique 5**). Par conséquent, les estimations directes sont adoptées pour les domaines de RMR, si elles sont disponibles. Autrement, les estimations du modèle d'EPD sont utilisées. Pour le modèle d'EPD pour lequel un échantillon d'enquête est disponible, l'estimation finale est l'estimation composite qui lie directement l'estimation directe à la valeur prédite par le modèle par le biais du facteur d'échelle (équation [6]); pour les domaines sans échantillon ou ceux où l'estimation directe extrême est 0 ou 1, les estimations finales sont des estimations synthétiques parce qu'elles sont calculées par le modèle seulement.

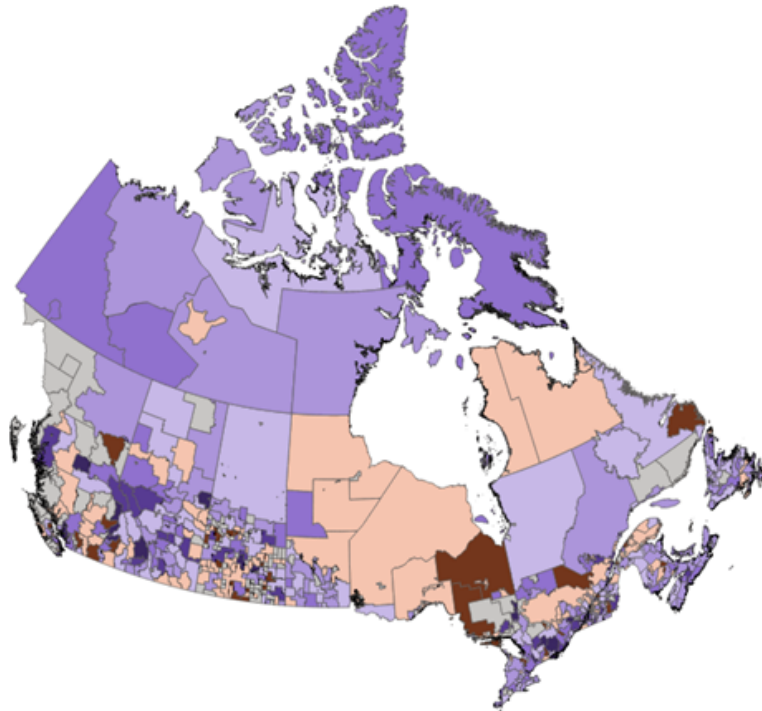
Les estimations finales pour le premier trimestre de 2022, classées par secteur de services ou de biens, sont portées sur une carte (**carte 1. B et carte 2. B**) et comparées à leurs contreparties d'estimation directe (**carte 1. A et carte 2. A**). Globalement, il y a 693 domaines n'ayant pas d'estimation directe (en gris) ou ayant une estimation directe extrême. Parmi ces domaines, 609 (88 %) sont des ZTA (régions rurales). Visuellement, les estimations finales donnent une couverture assez complète de la majorité des régions géographiques du Canada, grâce à l'approche d'EPD employant les données auxiliaires pour compléter l'ECSE.

Carte 1

Estimation directe et estimation finale pour le secteur des services, premier trimestre 2022

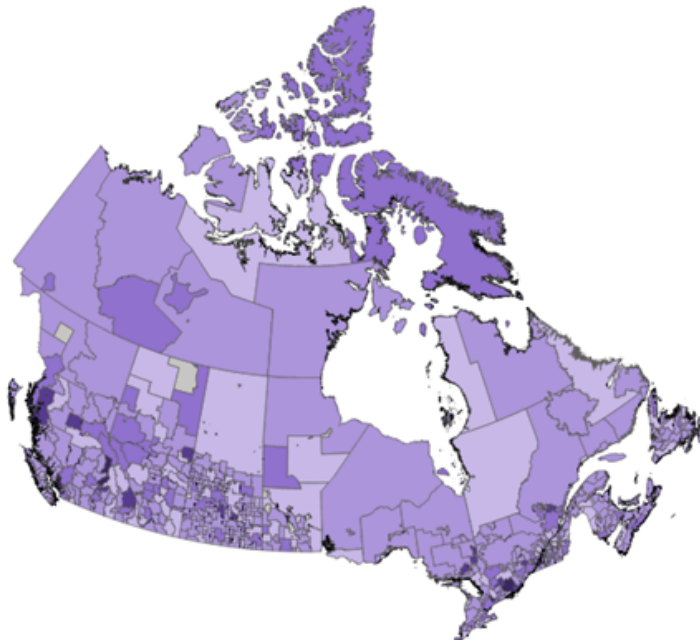
A - Estimation directe (secteur des services)

● 0 % ● 100 % ● > 0 à 20 % ● 20 à 40 % ● 40 à 60 % ● 60 à 80 % ● 80 à < 100 %



B - Estimation finale (secteur des services)

● > 0 à 20 % ● 20 à 40 % ● 40 à 60 % ● 60 à 80 % ● 80 à < 100 %

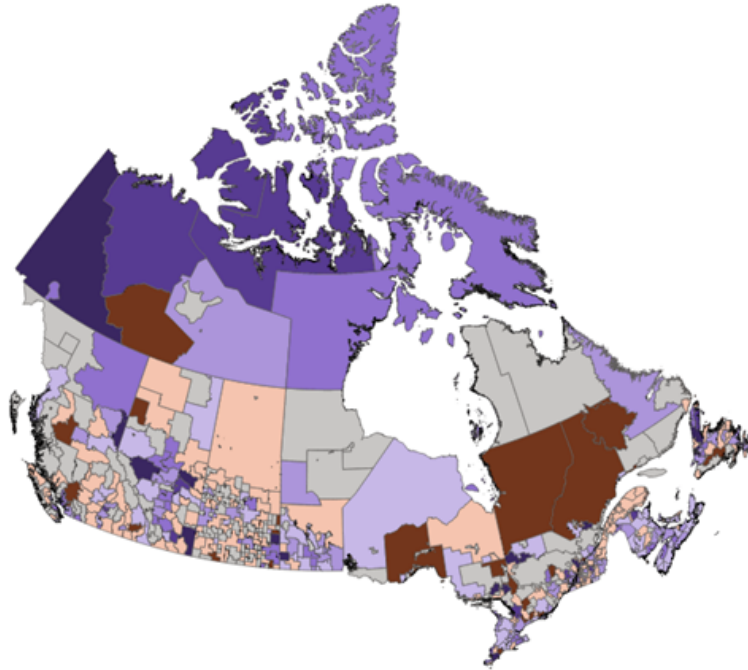


Source : ECSE (premier trimestre de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

Carte 2 Estimation directe et estimation finale pour le secteur des biens, premier trimestre 2022

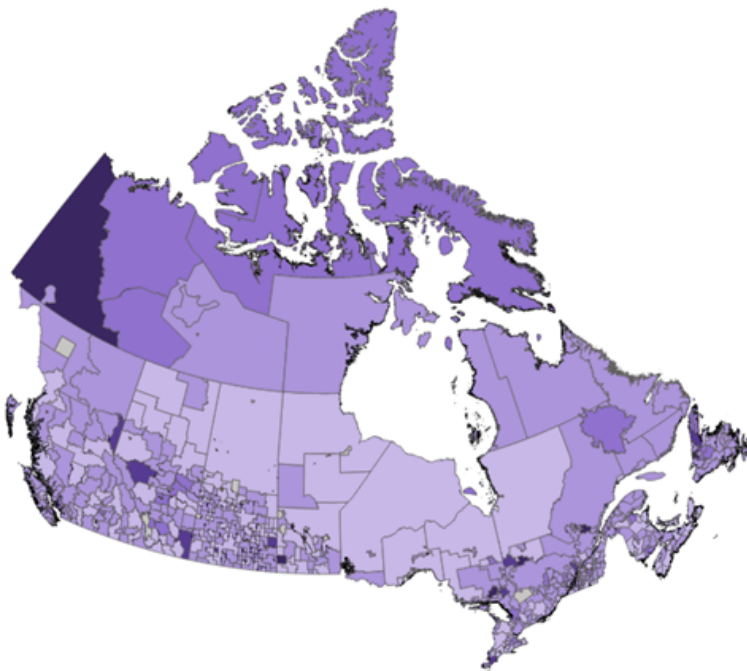
A - Estimation directe (secteur des biens)

● 0 % ● 100 % ● > 0 à 20 % ● 20 à 40 % ● 40 à 60 % ● 60 à 80 % ● 80 à < 100 %



B - Estimation finale (secteur des biens)

● > 0 à 20 % ● 20 à 40 % ● 40 à 60 % ● 60 à 80 % ● 80 à < 100 %



Source : ECSE (premier trimestre de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

7 Résultats analytiques

L'analyse des estimations finales cible leur distribution et leur variance par groupe et permet d'examiner la composition des activités dans les régions rurales. Les principaux résultats analytiques sont abordés ci-après.

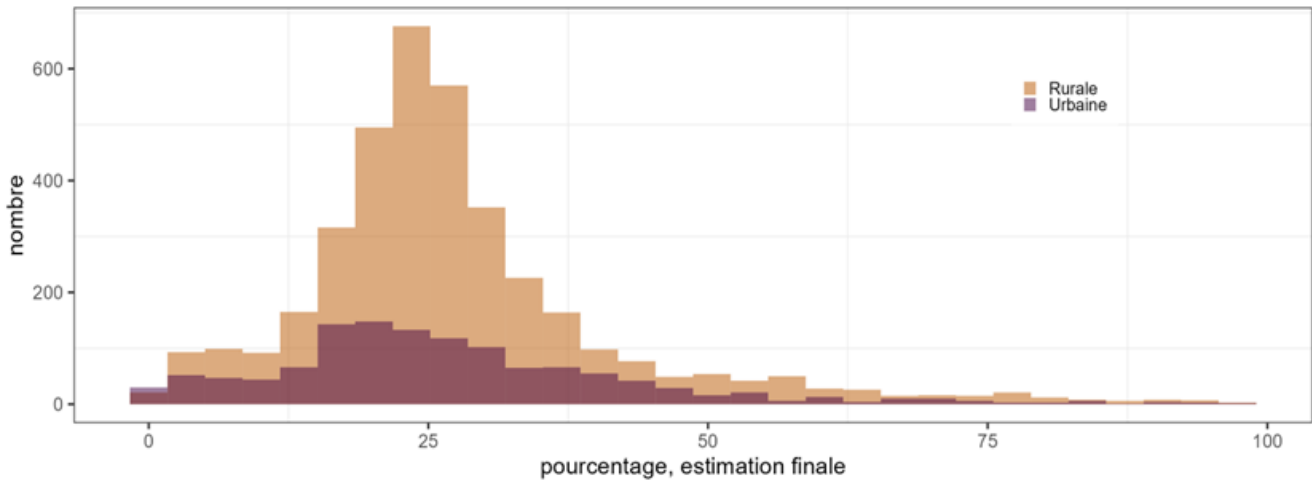
1. Les entreprises dans les régions rurales et urbaines et dans les secteurs des biens et des services offrent des modalités très diversifiées de télétravail

Pour les estimations finales des quatre trimestres de 2022, des tests statistiques regroupés⁷ sont exécutés pour détecter la différence de distribution des proportions estimées pour les régions, rurales et urbaines, et les producteurs de biens et les fournisseurs de services. Dans les deux cas, les résultats des tests mènent au rejet de l'hypothèse nulle d'une même distribution de variable cible entre les deux groupes. Les **graphique 6.A** et **7.B** illustrent la distribution groupée des deux variables catégoriques. Ces données soulignent l'importance de faire une distinction entre les régions rurales et les régions urbaines, et les secteurs d'activité au moment d'analyser l'intention des entreprises d'offrir des possibilités de télétravail. Plus précisément, la **graphique 6.A** indique qu'il y a une diversité importante de possibilités de télétravail dans les régions rurales. Les trois quarts des domaines (951 sur 1 263) sont ruraux et ont une estimation moyenne de 0,274 et un écart-type de 0,144, pour tous les trimestres de 2022. Ce résultat ne peut être obtenu qu'à l'aide de l'approche d'EPD.

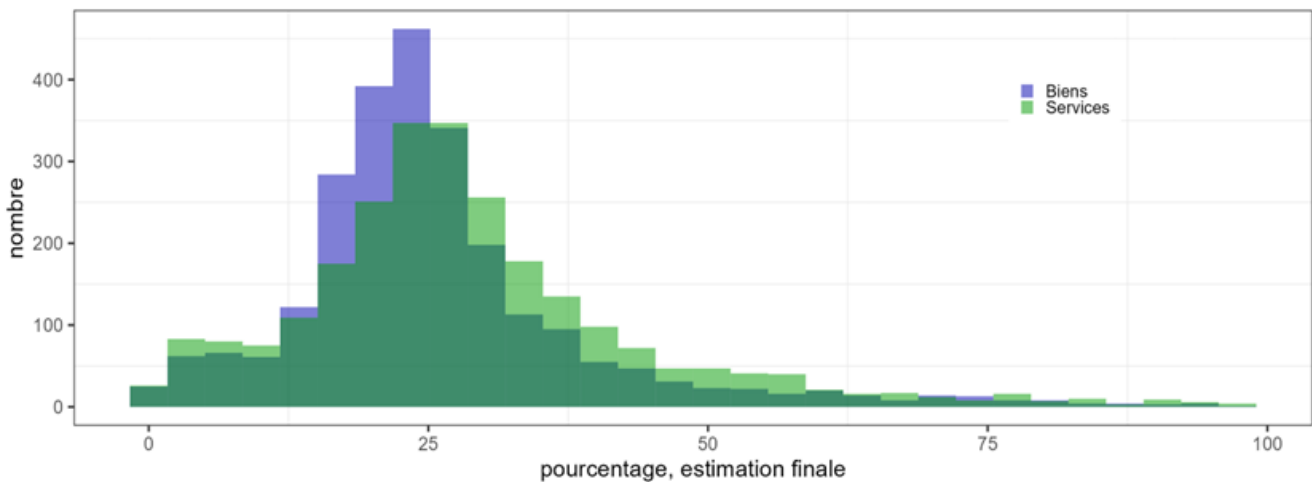
7. Les tests Mann-Whitney U sont exécutés, et les valeurs de p sont inférieures à 0,01.

Graphique 6
Estimations finales réparties différemment selon les groupes sélectionnés

A - Par indicateur de région rurale ou urbaine



B - Par indicateur de grand secteur d'activité



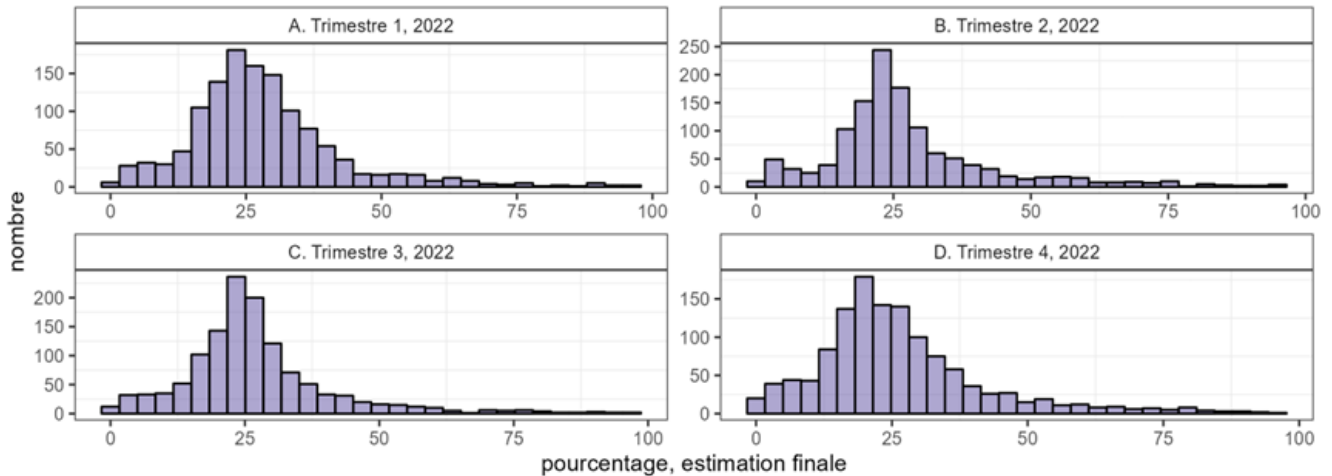
Source : ECSE (tous les trimestres de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

2. Les préférences des entreprises en ce qui concerne les modalités de télétravail offertes à leurs employés varient d'un trimestre à l'autre

Pour les mêmes données regroupées, un test statistique⁸ est exécuté pour déterminer les différences de distribution des proportions estimées pour les quatre trimestres de l'ECSE de 2022. Dans ce cas, l'hypothèse nulle de la même distribution pour les quatre trimestres est également rejetée. La **graphique 7** ci-dessous montre la répartition pour chaque trimestre. Puisque les questions de l'enquête originale portent sur l'anticipation des entreprises relativement aux modalités de travail au cours des trois mois subséquents, on s'attend à des changements au fil du temps, selon des facteurs comme les améliorations apportées par les entreprises à leurs conditions de sécurité technologiques et numériques, et la diminution de la gravité de la pandémie. Par conséquent, les tendances des entreprises peuvent parfois varier pour réagir à la situation courante.

8. Le test Mann-Whitney H est exécuté, et les valeurs de p sont inférieures à 0,01.

Graphique 7
Estimations finales réparties différemment d'un trimestre à l'autre



Source : ECSE (tous les trimestres de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

3. Les possibilités de télétravail offertes sur les marchés du travail en région rurale ont tendance à être relativement plus stables qu'en région urbaine

Selon l'évaluation par écart-type, les ZTA en régions rurales présentaient moins de variabilité dans les estimations finales des proportions que les régions urbaines, pour tous les trimestres de 2022 (**tableau 3**). Le fichier sur la population des entreprises du FGUE comprend environ 176 000 établissements des régions rurales, totalisant environ 18 % de tous les établissements qui étaient dans le champ d'enquête de cette analyse. En revanche, le nombre de domaines ruraux (951) géographiques était beaucoup plus grand que sa contrepartie urbaine (312). En termes de composition des activités, en 2021, les régions rurales comptaient 813 secteurs d'activité (codes SCIAN à six chiffres) qui étaient dans le champ d'enquête, soit moins que leurs contreparties urbaines (848). La moins grande couverture des activités dans les régions rurales entraîne une moins grande diversité de comportements des entreprises et, par conséquent, influence l'écart-type des estimations finales. Finalement, le facteur de pondération du nombre d'employés dans les régions rurales (63) était beaucoup moins grand que dans les régions urbaines (251). Les entreprises situées dans les régions rurales ayant un nombre d'employés relativement petit peuvent offrir plus de souplesse en matière de modalités de travail. Les estimations individuelles d'une entreprise rurale peuvent être davantage regroupées autour de la moyenne du groupe, ce qui donne un écart-type plus faible.

Tableau 3
Statistiques sommaires de l'estimateur final par région rurale et urbaine

	Rurale				Urbaine			
	Moyenne	Écart-type	Intervalle de confiance de 95 %		Moyenne	Écart-type	Intervalle de confiance de 95 %	
	pourcentage		de	à	pourcentage		de	à
Trimestre 1, 2022	28,1	13,5	27,3	29,0	28,4	15,9	26,6	30,2
Trimestre 2, 2022	27,8	14,7	26,9	28,7	26,3	17,2	24,4	28,2
Trimestre 3, 2022	27,6	13,9	26,7	28,5	25,6	15,1	23,9	27,2
Trimestre 4, 2022	26,1	15,3	25,1	27,1	27,6	17,1	25,7	29,4

Source : ECSE (tous les trimestres de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

4. Les entreprises rurales du secteur des services sont plus susceptibles d'offrir des possibilités de télétravail que celles du secteur des biens, mais la tendance est à la baisse d'un trimestre à l'autre

Dans les régions rurales, comme en témoigne la valeur médiane de la proportion estimée, les entreprises du secteur des services étaient plus susceptibles d'offrir des modalités de télétravail que celles du secteur des biens. L'écart le plus notable entre les deux secteurs a été observé au premier trimestre de 2022. Toutefois, alors que la situation de l'entreprise s'améliorait au cours de la pandémie, cette différence diminuait modérément au fil des trimestres (**tableau 4**). Le troisième trimestre était une exception, qui semblait être une réaction exagérée du secteur des services à la diminution des restrictions liées à la pandémie, de telle façon que les entreprises étaient (un peu) moins susceptibles d'offrir le télétravail que celles du secteur des biens.

Globalement, les fournisseurs de services ont légèrement réduit leur offre de possibilité de télétravail, du premier au quatrième trimestre de 2022. Ce résultat est cohérent avec le fait qu'un plus grand nombre d'entreprises ont encouragé leurs employés à revenir travailler sur place lorsque les restrictions liées à la pandémie ont diminué. La réponse des producteurs de biens à ce changement de situation était variée, malgré qu'au quatrième trimestre, on puisse remarquer le plus faible point de pourcentage de volonté des entreprises à permettre le télétravail.

Tableau 4

Statistiques sommaires de l'estimateur final en région rurale par grand secteur d'activité

	Biens				Services			
	Médiane	Moyenne	Intervalle de confiance de 95 %		Médiane	Moyenne	Intervalle de confiance de 95 %	
	pourcentage		de	à	pourcentage		de	à
Trimestre 1, 2022	23,4	25,2	24,1	26,2	29,6	31,1	29,8	32,4
Trimestre 2, 2022	23,6	26,7	25,4	27,9	25,7	28,9	27,5	30,3
Trimestre 3, 2022	25,8	28,0	26,8	29,1	24,4	27,2	25,9	28,6
Trimestre 4, 2022	22,7	24,8	23,6	26,0	23,6	27,4	25,9	28,9

Source : ECSE (tous les trimestres de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

5. En région rurale, les domaines les plus susceptibles d'offrir des possibilités de télétravail et les domaines les moins susceptibles de le faire ne sont pas les mêmes pour les quatre trimestres

Dans une analyse centile des domaines ruraux, les estimations finales de la tranche de 5 % supérieure et de la tranche de 5 % inférieure, groupées par trimestre, ont été calculées; la tranche supérieure étant définie comme les domaines les plus susceptibles d'offrir des modalités de télétravail et la tranche inférieure étant définie comme les domaines les moins susceptibles de le faire. Tous les domaines ruraux dont les estimations finales étaient supérieures ou égales à la tranche de 5 % supérieure, et inférieures ou égales à la tranche de 5 % inférieure, ont ensuite été triés. De l'ensemble des 951 domaines ruraux, il y en avait 275, soit près de 30 %, qui se trouvaient dans cet intervalle, pour un trimestre ou pour plusieurs trimestres. De ce nombre, 24 s'y trouvaient au moins trois trimestres (**tableau 5**).

Tableau 5
Tranche de 5 %, supérieure et inférieure des domaines ruraux figurant dans au moins trois trimestres de 2022

Domaine	Estimation finale de 2022			
	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
	pourcentage			
10_6717_Services	...	5,6	7,9	2,5
13_6532_Biens	1,9	4,4	2,6	5,9
24_4591_Biens	57,7	74,9	61,0	69,5
24_6025_Services	5,8	...	6,2	3,5
24_6238_Services	82,3	67,8	...	58,8
24_6252_Services	...	3,8	79,1	3,9
24_6588_Biens	...	66,1	61,3	3,2
24_6734_Services	58,3	...	64,8	4,6
24_6830_Biens	...	4,5	5,3	1,7
35_1000031_Biens	62,9	76,9	59,0	81,2
35_6804_Services	8,3	83,7	4,0	58,9
46_6778_Biens	7,4	...	3,0	1,8
47_1000038_Biens	58,0	69,6	57,1	77,2
47_4618_Biens	59,0	73,2	58,3	80,4
47_5183_Services	...	76,3	58,8	87,4
47_5907_Services	5,9	2,3	...	1,4
47_6343_Biens	64,2	74,4	56,6	75,3
48_1000059_Biens	...	71,0	59,4	85,2
48_6337_Services	59,0	69,8	...	4,0
48_6763_Biens	...	81,7	5,0	0,8
48_6763_Services	72,5	...	7,0	2,6
59_5524_Services	5,6	67,7	72,8	...
59_6036_Services	66,5	0,8	...	3,1
59_6522_Biens	...	69,2	83,2	2,1

... n'ayant pas lieu de figurer

Source : ECSE (tous les trimestres de 2022), FGUE 2021 V18, Recensement de 2021, calculs de l'auteur.

Une analyse plus approfondie de ces domaines reliés au fichier de la population du RE du FGUE révèle qu'ils étaient dominés par le secteur des services. Il y avait 335 entrées de codes SCIAN à six chiffres du secteur des services qui s'étendaient à plus de 3 810 établissements dans le champ de l'enquête, par rapport à 167 entrées de codes SCIAN pour 2 099 entreprises du secteur des biens.

Certains de ces domaines, toutefois, ne se retrouvaient pas de manière cohérente dans l'intervalle de 5 % supérieur ou inférieur. Leurs estimations trimestrielles fluctuaient plutôt entre les intervalles de la tranche inférieure ou de la tranche supérieure. De nouveau, d'après le fichier de la population du RE du FGUE, ce sous-ensemble était dominé par les entreprises du secteur des services parce que 70 % des entrées de codes SCIAN à six chiffres appartenaient au secteur des services. Par conséquent, les entreprises du secteur des services étaient plus instables dans leur offre de possibilités de télétravail dans ces cas extrêmes.

8 Conclusions et pistes de recherche futures

Dans la présente étude expérimentale, l'application du modèle d'EPD au niveau du domaine a permis d'estimer la proportion d'entreprises où le télétravail est possible, pour chaque trimestre de 2022. Les domaines d'intérêt correspondaient à une combinaison de régions géographiques et de grands secteurs d'activité, et les régions géographiques étaient classées en RMR, AR et ZTA. Les ZTA étaient définies comme des régions rurales ou des marchés du travail en région rurale.

L'approche d'EPD complétait les données-échantillons de l'ECSE, auxquelles étaient ajoutées les données auxiliaires du FGUE, et le Recensement de la population de 2021. Les estimations du modèle d'EPD sont des estimations composites lorsqu'elles sont alimentées par les estimations directes de l'enquête et par les résultats du modèle. D'une part, pour les domaines pour lesquels aucun échantillon n'a été pris dans l'enquête, les estimations d'EPD sont alimentées seulement par les données auxiliaires; par conséquent, ce sont des valeurs synthétiques. Lorsqu'un indicateur d'intérêt est une proportion, les valeurs extrêmes (c.-à-d., 0 ou 1) de l'estimation directe sont ignorées par le modèle puisqu'elles pourraient ne pas être fiables en raison de la petite taille de l'échantillon. Les estimations d'EPD sont alors également synthétiques. Les estimations finales sont

préparées en combinant trois sources de résultats. Pour les domaines de RMR, les estimations directes sont adoptées lorsqu'elles sont pertinentes. Autrement, les estimations du modèle d'EPD sont utilisées.

Quant aux conclusions d'analyse, les entreprises, selon qu'elles sont situées en région rurale ou en région urbaine et qu'elles appartiennent au secteur des biens ou au secteur des services, ont eu des comportements différents en matière de modalités de télétravail. De plus, il y a une diversité importante de possibilités de télétravail dans les régions rurales. La pandémie a entraîné des changements, et les préférences des entreprises quant aux possibilités de télétravail variaient également d'un trimestre à l'autre en 2022. Les possibilités de télétravail offertes sur les marchés du travail dans les régions rurales avaient tendance à être relativement plus stables que dans les régions urbaines. Les proportions médianes indiquent également que les entreprises du secteur des services, situées en région rurale, avaient légèrement mieux réussi que celles du secteur des biens à offrir des possibilités de télétravail, mais la tendance était à la baisse chaque trimestre. Sur les marchés du travail dans les régions rurales, tant les entreprises les plus susceptibles que les moins susceptibles d'offrir des possibilités de télétravail variaient d'un trimestre à l'autre.

Puisque les données de l'ECSE sont publiées trimestriellement, les estimations d'EPD pour les nouveaux trimestres pourraient être générées pour poursuivre la série temporelle. D'autres améliorations aux méthodologies d'EPD sont aussi envisagées pour de futures analyses. Cette évolution pourrait comprendre le choix d'autres indicateurs de la situation des entreprises, l'exploration d'autres sources de données auxiliaires et l'incorporation d'un plus grand nombre de prédicteurs dans le modèle.

Références

- Bocci, C., J.-F. Beaumont et K. Bosa. (2022). Small area estimation methodology of the unemployment rate in special labour areas updated with experiments conducted in 2022. *Document interne de Statistique Canada*.
- Esteban, M. D., M. J. Lombardía, E. López-Vizcaíno, D. Morales et A. Pérez. (2020). [Small area estimation of proportions under area-level compositional mixed models](https://doi.org/10.1007/s11749-019-00688-w). *TEST*, vol. 29, n° 3, p. 793 à 818. <https://doi.org/10.1007/s11749-019-00688-w>
- Fay, R.E. et R.A. Herriot. (1979). [Estimates of income for small places: An application of James-Stein procedures to census data](https://doi.org/10.1080/01621459.1979.10482505). *Journal of the American Statistical Association*, vol. 74, p. 269 à 277. <https://doi.org/10.1080/01621459.1979.10482505>
- Hidiroglou, M.A., J.-F. Beaumont et W. Yung. (2019). Élaboration d'un système d'estimation sur petits domaines à Statistique Canada. *Techniques d'enquête*, vol. 45, n° 1, p. 107 à 133.
- Munro, A., A. Alasia et R. D. Bollman. (2011). [Zones de travail autonomes : une proposition de délimitation et de classification selon le degré de ruralité](https://www150.statcan.gc.ca/n1/fr/pub/21-006-x/21-006-x2008008-fra.pdf?st=_DwqMAqOi). *Bulletin d'analyse : régions rurales et petites villes au Canada*, vol. 8, n° 8 (Ottawa, n° 21-006-X au catalogue de Statistique Canada). https://www150.statcan.gc.ca/n1/fr/pub/21-006-x/21-006-x2008008-fra.pdf?st=_DwqMAqOi
- OCDE (2020). [Delineating Functional Areas in All Territories](https://www.oecd.org/publications/delineating-functional-areas-in-all-territories-07970966-en.htm). <https://www.oecd.org/publications/delineating-functional-areas-in-all-territories-07970966-en.htm>
- Rao, J.N.K. et I. Molina. (2015). *Small Area Estimation*. John Wiley & Sons, Inc., Hoboken (New Jersey).
- Rivest, L.P. et E. Belmonte. (2000). Une erreur quadratique moyenne conditionnelle des estimateurs régionaux. *Techniques d'enquête*, vol. 26, n° 1, p. 79 à 90.
- Sholihin, M. R. et C. Sumarni. (2022). [Package "sae.prop"](https://cran.r-project.org/web/packages/sae.prop/sae.prop.pdf). <https://cran.r-project.org/web/packages/sae.prop/sae.prop.pdf>
- Statistique Canada. [Base de métadonnées intégrée de l'ECSE](https://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&Id=1330334). https://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&Id=1330334
- Statistique Canada. [Questionnaire de l'ECSE](https://www.statcan.gc.ca/fr/programmes-statistiques/instrument/5318_Q1_V8). https://www.statcan.gc.ca/fr/programmes-statistiques/instrument/5318_Q1_V8
- Statistique Canada (2019). Système généralisé d'estimation, version 2.03. Guide méthodologique – Sampling variance with Taylor linearization. *Document interne de Statistique Canada*.
- Statistique Canada (2022). [Glossaire illustré du Recensement de 2021](https://www150.statcan.gc.ca/n1/fr/pub/92-195-x/92-195-x2021001-fra.pdf?st=ARo4Ds4J). <https://www150.statcan.gc.ca/n1/fr/pub/92-195-x/92-195-x2021001-fra.pdf?st=ARo4Ds4J>
- Statistique Canada. [Classification géographique type \(CGT\) 2016 – Volume I, La classification](https://www.statcan.gc.ca/fr/sujets/norme/cgt/2016/index). <https://www.statcan.gc.ca/fr/sujets/norme/cgt/2016/index> (consulté le 28 avril 2023)
- Statistique Canada (2021). [Les statistiques : le pouvoir des données!](https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-fra.htm) <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-fra.htm>
- Statistique Canada. [Variante du SCIAN 2017 version 3.0 – Industries productrices de biens et services](https://www23.statcan.gc.ca/imdb/p3VD_f.pl?Function=getVD&TVD=1206276). https://www23.statcan.gc.ca/imdb/p3VD_f.pl?Function=getVD&TVD=1206276
- You, Y. et M. Hidiroglou. (2022). Application de méthodes de lissage de la variance d'échantillonnage pour des estimations de proportion pour petits domaines. *Recueil du Symposium 2022 de Statistique Canada*.