

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

Nouveau tableau des ventes du secteur manufacturier canadien selon les emplacements d'origine et de destination : défis de l'imputation d'une répartition à partir de données d'une enquête annuelle

par Nicholas Huliganga et Sébastien Landry

Date de diffusion : le 8 septembre 2025



Statistique
Canada

Statistics
Canada

Canada

Nouveau tableau des ventes du secteur manufacturier canadien selon les emplacements d'origine et de destination : défis de l'imputation d'une répartition à partir de données d'une enquête annuelle

Nicholas Huliganga et Sébastien Landry¹

Résumé

Jusqu'à maintenant, les Canadiens n'avaient pas accès à des données détaillées sur l'emplacement des ventes du secteur de la fabrication. Grâce à l'intégration des données de l'enquête annuelle (Enquête annuelle sur les industries manufacturières et de l'exploitation forestière), un tableau de l'emplacement des ventes selon l'industrie et la province d'origine a été élaboré au profit des enquêtes annuelles et mensuelles sur les industries manufacturières à Statistique Canada.

Bien qu'il ne s'agisse pas d'une question de l'enquête mensuelle, les répondants à l'enquête annuelle sont priés d'indiquer leur répartition des ventes en pourcentage dans 15 emplacements. Pour aborder la difficulté de générer une répartition au niveau de l'établissement des répondants œuvrant dans plusieurs provinces, trois approches ont été comparées : utiliser la répartition totale des répondants pour tous leurs établissements; viser l'optimisation; et utiliser les répartitions des répondants d'une seule province. La répartition imputée des ventes selon l'emplacement à partir des données annuelles a ensuite été appliquée à la valeur des ventes mensuelles. Les établissements mensuels non couplés aux répondants annuels ont été imputés à l'aide d'une stratégie comportant la répartition agrégée de leur groupe industrie-province dans l'enquête annuelle.

Enfin, on a réussi à produire, dans le cadre de l'enquête mensuelle, des estimations ponctuelles et une variance d'échantillonnage (ainsi qu'un coefficient de variation) pour tout le tableau des ventes selon les emplacements d'origine et de destination, ainsi que pour chaque industrie.

Le présent exposé aborde les défis liés à l'imputation des ventes selon l'emplacement (particulièrement en ce qui concerne les répondants ayant des établissements dans plusieurs provinces). L'objectif est d'assurer que les ventes correspondent aux totaux marginaux des provinces d'origine et de répartir les emplacements en fonction des données tirées du programme annuel aux estimations mensuelles.

Mots clés : Intégration des données; méthode itérative du quotient; imputation; fabrication.

1. Introduction

Un tableau de l'origine et de la destination des ventes donne une image de la répartition des ventes de chaque origine à un ensemble de destinations disjointes. Ce type de renseignements n'était pas disponible pour les ventes du secteur de la fabrication au Canada avant juillet 2024, lorsque les nouveaux tableaux ont été publiés en ligne (Statistique Canada, 2024a). La production de ces tableaux à l'aide des données de l'Enquête annuelle sur les industries manufacturières de Statistique Canada a posé plusieurs défis, particulièrement en ce qui concerne l'imputation des répartitions pour certains répondants. Trois méthodes d'imputation différentes ont été explorées pour relever ces défis. Les projets de publication d'une version mensuelle comportent également certaines considérations en raison des différences importantes entre les enquêtes annuelles et mensuelles sur les industries manufacturières.

¹Nicholas Huliganga, Statistique Canada, 300 W Georgia St, Vancouver (Colombie-Britannique) Canada, V6B 6B1 (nicholas.huliganga@statcan.gc.ca); Sébastien Landry, Statistique Canada, 150 Pré-Tunney, Ottawa, Canada, K1A 0T6 (sebastien.landry@statcan.gc.ca)

2. Enquêtes sur les industries manufacturières

Les deux enquêtes nécessaires pour recueillir les données d'un tableau origine-destination des ventes étaient l'Enquête annuelle sur les industries manufacturières et de l'exploitation forestière (EAMEF) (Statistique Canada, 2023) et l'Enquête mensuelle sur les industries manufacturières (EMIM) (Statistique Canada, 2024b). Des estimations ponctuelles et des indicateurs de qualité ont été publiés pour les deux enquêtes pour les codes 31 à 33 du Système de classification des industries de l'Amérique du Nord (SCIAN), soit le secteur de la fabrication, y compris la fabrication d'aliments (311), des produits en bois (321), de machines (333) et ainsi de suite. Ces estimations publiées ne fournissent que des renseignements sur l'origine géographique des ventes; aucune information n'a été publiée sur la destination des ventes. Par exemple, les deux enquêtes ont publié les ventes des établissements de fabrication en Ontario, mais pas les ventes à des destinations comme le Québec ou le Manitoba.

Dans l'enquête annuelle (EAMEF), les répondants déclarent leurs ventes pour 15 destinations en pourcentage. Les destinations comprenaient les 13 provinces et territoires du Canada, ainsi que les États-Unis et « Autre », représentant le reste du monde. Ce type de question n'existait pas dans l'EMIM; la seule façon de produire un tableau mensuel origine-destination était donc d'intégrer les données annuelles d'une façon ou d'une autre.

Les réponses reçues dans le cadre de l'EAMEF pourraient servir à estimer les renseignements sur la proportion des ventes destinées à chaque destination au niveau du Canada. Toutefois, pour créer un tableau annuel final sur 13x15 origines-destinations, comprenant les 13 provinces et territoires du Canada comme origine et les 15 destinations, les renseignements doivent se situer au niveau de chaque établissement manufacturier. Cette répartition du niveau du répondant au niveau de l'établissement a été difficile pour les répondants à l'enquête annuelle ayant des établissements dans plusieurs provinces.

3. Répondants avec plusieurs provinces (M-P)

3.1 Description

Les répondants du secteur de la fabrication de l'EAMEF sont échantillonnés à partir d'une base de sondage composée d'une liste d'entreprises figurant dans le Registre des entreprises (RE) de Statistique Canada. Ils peuvent comprendre plusieurs établissements répartis dans plusieurs provinces et territoires. Par conséquent, ils peuvent être classés comme correspondants à une seule province (tous les établissements résident dans la même province ou le même territoire; S-P) ou à plusieurs provinces (établissements résidant dans plusieurs provinces ou territoires; M-P). Pour estimer les totaux agrégés des tableaux annuels, des répartitions au niveau de l'établissement étaient requises au lieu des répartitions globales au niveau du répondant déclarées dans l'EAMEF.

Pour les établissements de répondants S-P, la répartition globale pouvait simplement être copiée pour les établissements individuels. Par contre, pour les répondants M-P, cette façon ne permettait pas nécessairement d'obtenir les répartitions au niveau de l'établissement. Par exemple, si un répondant M-P avait des établissements en Ontario et en Colombie-Britannique (C.-B.) et déclarait que 60 % de ses ventes étaient destinées à l'Ontario, il n'était peut-être pas exact de supposer que ses établissements en C.-B. envoyaient 60 % de leurs propres ventes en Ontario.

Trois approches différentes ont été explorées pour imputer une répartition pour ces répondants M-P. Un exemple fictif d'un répondant annuel peut aider à illustrer les différences entre les approches. Le répondant de cet exemple avait une valeur de ventes totale de 1 000 \$ et était établi dans deux provinces, en ayant des établissements au Québec (pour un total de 780 \$) et en Alberta (pour un total de 220 \$). De plus, il avait déclaré une répartition globale en pourcentage de la destination des ventes (figure 3.1-1).

3.2 Première approche

L'approche la plus élémentaire et la moins souhaitable pour imputer les répartitions pour les répondants M-P consistait à les traiter comme s'ils étaient des répondants S-P et à copier la répartition globale pour chaque établissement. Les résultats de cette approche étaient souvent incompatibles avec une compréhension générale du fonctionnement d'établissements de fabrication. Plus précisément, les établissements sont plus susceptibles de vendre à leur propre

province ou à celles les plus proches d'eux. Toutefois, selon la première approche, le répondant fictif a été imputé de sorte que les établissements de l'Alberta envoient plus du double de leurs ventes au Québec qu'en Alberta (tableau 3.2-1).

Figure 3.1-1
Répartition selon la destination des ventes pour un répondant fictif

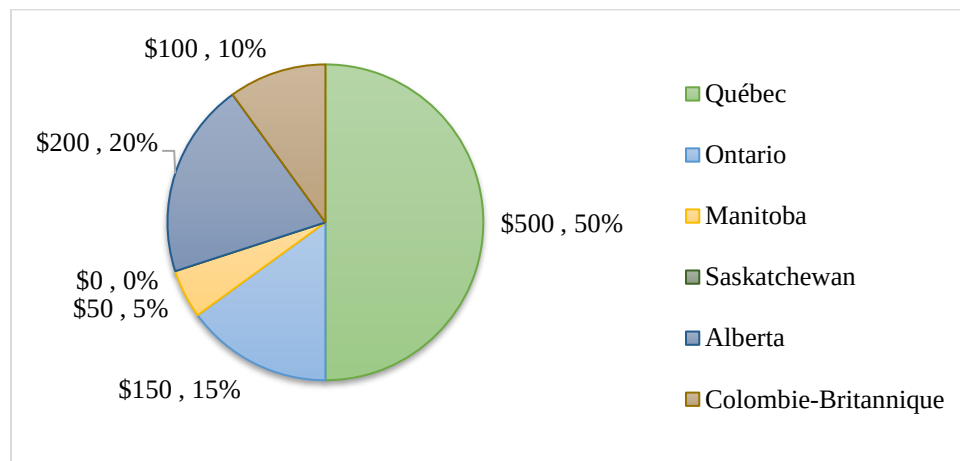


Tableau 3.2-1
Première approche : destination des ventes pour l'exemple de répondant.

	Destination des ventes (\$)						Total
	Québec	Ontario	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	
Répondant	500 (50 %)	150 (15 %)	50 (5 %)	0	200 (20 %)	100 (10 %)	1 000
Origine							
Québec	390 (50 %)	117 (15 %)	39 (5 %)	0	156 (20 %)	78 (10 %)	780
Alberta	110 (50 %)	33 (15 %)	11 (5 %)	0	44 (20 %)	22 (10 %)	220

3.3 Deuxième approche

Pour mieux tenir compte des distances entre les origines et les destinations lors de l'imputation d'une répartition pour les répondants M-P, la deuxième approche a recours à l'optimisation. Plus précisément, le coût de vente d'une origine à une destination a été réduit au minimum, et défini comme la distance moyenne entre chaque origine et chaque destination. La distance pour chaque paire origine-destination a été calculée à l'aide de la distance moyenne entre les codes postaux de l'origine et de la destination; le code postal étant géocodé. Chaque répondant utilisait la même matrice globale coûts-distances, qui comprenait les 13 origines et les 15 destinations, pour le processus d'optimisation.

L'optimisation a également incorporé des contraintes pour s'assurer que les totaux marginaux étaient conservés pour chaque répondant. Cela était particulièrement important pour les totaux des colonnes (destination), de sorte que les répartitions finales au niveau de l'établissement reviendraient à la répartition globale déclarée pour chaque répondant.

Un dernier ajustement avant la procédure d'optimisation consistait à maximiser les ventes intraprovinciales en tenant compte des contraintes marginales. Pour le répondant de l'exemple, les montants maximums de 500 \$ et de 200 \$ ont respectivement été attribués aux paires du Québec au Québec et de l'Alberta à l'Alberta (tableau 3.3-1). Cette étape était nécessaire parce que, sans elle, l'optimisation pouvait parfois attribuer toutes les ventes à une destination comme provenant d'une autre province, même lorsque les ventes des établissements de la province de destination n'étaient pas nulles. La nature générale et imprécise de la mesure des distances en est la cause. Par exemple, l'optimisation pourrait éviter d'envoyer les ventes d'une province comme le Québec sur une longue distance coûteuse vers la Colombie-Britannique, en choisissant plutôt de répartir celles provenant de l'Alberta. Cela pourrait faire en sorte

qu'aucune des ventes de l'Alberta ne reste en Alberta; ce qui est très improbable selon notre compréhension préalable de la façon dont les établissements de fabrication ont tendance à fonctionner.

Tableau 3.3-1

Destination des ventes pour l'exemple de répondant selon la deuxième approche

	Destination des ventes (\$)						
	Québec	Ontario	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	Total
Répondant	500 (50 %)	150 (15 %)	50 (5 %)	0	200 (20 %)	100 (10 %)	1 000
Origine							
Québec	500 (64,1 %)	150 (19,2 %)	50 (6,4 %)	0	0	80 (10,3 %)	780
Alberta	0	0	0	0	200 (90,9 %)	20 (9,1 %)	220

3.4 Troisième approche

Dans la troisième approche, l'imputation au niveau de l'établissement était fondée sur les répartitions agrégées des répondants S-P. Pour chaque établissement d'un répondant M-P, sa répartition initiale entre 15 destinations (exprimée en pourcentage) a été calculée à partir des totaux agrégés des répondants S-P; plus précisément, les établissements de répondants S-P du groupe SCIAN-province d'origine le plus semblable ayant au moins cinq unités.

Un ensemble initial de 15 facteurs correspondant à chaque destination a été calculé en additionnant les ventes attribuées à chaque destination des répondants S-P et en les divisant par la somme des ventes totales des répondants S-P du groupe. Par exemple, si les cinq répondants S-P d'un groupe ont envoyé 50 \$, 150 \$, 0 \$, 0 \$ et 200 \$ au Québec et que leurs ventes totales étaient de 1 000 \$, 500 \$, 50 \$, 50 \$ et 2 000 \$, le facteur initial pour la destination du Québec serait alors de $400/3\ 600$ ou 0,11. Ensuite, pour chaque répondant M-P, les facteurs correspondant aux destinations où les ventes étaient nulles ont été établis à zéro. Par exemple, même si les répondants S-P de leur groupe vendaient à la Saskatchewan, le facteur de destination pour la Saskatchewan du répondant de l'exemple serait établi à zéro.

Un processus itératif de quotient (« raking ») a ensuite été appliqué pour veiller à la conservation des totaux marginaux, comme dans le cas de la deuxième approche. Les résultats finaux tendaient à être semblables à ceux de la deuxième approche; les ventes intraprovinciales dépassant celles de toute autre destination. Pour le répondant de l'exemple, les deuxième et troisième approches étaient assez semblables; les valeurs Québec-Québec et Alberta-Alberta étant les plus élevées pour chaque province d'origine (tableau 3.4-1).

Tableau 3.4-1

Troisième approche : Destination des ventes pour l'exemple de répondant.

	Destination des ventes (\$)						
	Québec	Ontario	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	Total
Répondant	500 (50 %)	150 (15 %)	50 (5 %)	0	200 (20 %)	100 (10 %)	1 000
Origine							
Québec	500 (64,1 %)	150 (19,2 %)	10 (1,3 %)	0	20 (2,6 %)	100 (12,8 %)	780
Alberta	0	0	40 (18,2 %)	0	180 (81,8 %)	0	220

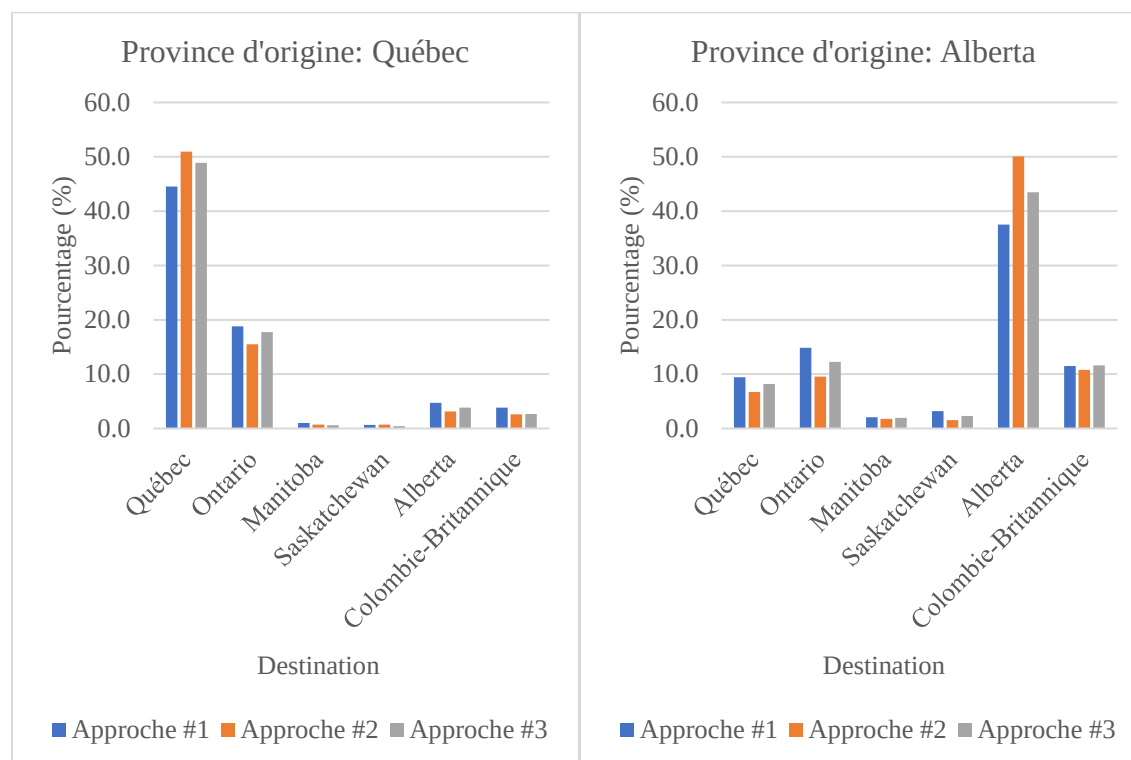
3.5 Comparaisons des approches

Dans le cadre d'une répartition au niveau de l'établissement pour les répondants annuels à l'aide des trois approches, on a comparé ces dernières en effectuant une estimation pour calculer les tableaux finaux à tous les niveaux SCIAN pour chacune d'entre elles. Dans l'ensemble, les résultats étaient semblables entre les trois à l'échelle nationale, puisque 87 % du nombre total de répondants à l'enquête annuelle correspondaient à une seule province ayant les mêmes répartitions pour chaque approche. Les différences sont devenues plus évidentes à des niveaux plus précis SCIAN-province; toutefois, en général, les deuxième (optimisation distance-coût) et troisième approches (agrégats de répondants S-P) étaient les plus semblables; le nombre de ventes le plus élevé ayant tendance à demeurer à l'intérieur de la même province d'origine, tandis que les ventes étaient plus réparties avec la première approche (copie de la

répartition globale des répondants). Cela a, par exemple, été observé pour le code SCIAN 311 (fabrication d'aliments) pour les données de l'EAMEF de 2021 (figure 3.5-1).

Figure 3.5.-1

Comparaison des proportions des ventes pour les données de l'EAMEF de 2021 (code SCIAN 311).



Après avoir consulté des spécialistes du domaine de l'EAMEF, la troisième approche a été choisie comme méthode d'imputation privilégiée. Ce choix a été fait en raison de l'utilisation de données sur les répondants réels pour la troisième approche, alors que la deuxième approche se fondait sur une mesure générale des distances en faisant intervenir des codes postaux et des hypothèses plus difficiles à valider. Même si nous ne pouvons pas prouver de façon définitive les résultats de la troisième approche, elle a été considérée comme une option plus fiable.

4. Estimation annuelle de l'enquête

L'estimation de l'EAMEF était simple une fois les ventes de chaque établissement de la partie échantillonnée réparties entre les 15 destinations (d'après la question initiale de l'enquête ou par imputation). Il était facile de traiter les valeurs pour les 15 destinations comme de nouvelles variables; des estimations étant générées pour chaque niveau du SCIAN-province d'origine existant pour l'EAMEF. Les poids pour les nouvelles variables ont été choisis pour être les mêmes que la valeur initiale des ventes pour chaque établissement. Enfin, les nouvelles variables étaient considérées comme imputées si la valeur globale des ventes était imputée, la réponse de la répartition en pourcentage des ventes entre les 15 destinations était imputée ou si l'établissement relevait d'un répondant correspondant à plusieurs provinces. Les tableaux finaux ont été publiés avec succès en ligne sur le site Web de Statistique Canada pour les données annuelles de 2020 à 2022.

4.1 Partie à tirage nul

Pour produire les estimations finales, des valeurs étaient également nécessaires pour la partie à tirage nul (non échantillonnée) de la population pour les 15 nouvelles variables de destination. Cette partie représentant les plus petits

établissements de la population, il a été décidé de répartir toutes les ventes de ces établissements à la même province d'origine. Par exemple, on supposerait que 100 % des ventes d'un établissement à tirage nul du Québec étaient destinées au Québec.

4.2 Indicateurs de qualité

L'obtention d'un indicateur de qualité était également simple, puisque les 15 variables de destination étaient traitées de la même façon que la valeur globale des ventes. Ainsi, l'indicateur de qualité pour chaque cellule du tableau 13 par 15 a été dérivé exactement de la même façon que les ventes globales, étant fondé sur deux dimensions : le taux de réponse pondéré et le coefficient de variation (CV).

5. Mise à l'essai de l'enquête mensuelle

Des travaux sont en cours pour produire une version semblable des tableaux origine-destination pour l'enquête mensuelle sur les industries manufacturières. Contrairement à l'EAMEF, l'EMIM ne comprend pas de question à laquelle les répondants fournissent des renseignements sur la destination de leurs ventes. Par conséquent, des données annuelles doivent être intégrées pour imputer les renseignements sur la destination pour chaque établissement dans l'enquête mensuelle.

5.1 Défis

Lorsque l'on envisage l'utilisation des données annuelles pour l'enquête mensuelle, deux problèmes sont que les établissements n'existent pas nécessairement dans la partie échantillonnée des deux enquêtes, et qu'il y a des différences dans le calendrier de publication des deux enquêtes.

Une autre considération importante est que, sans calendarisation, les données annuelles seraient les mêmes pour chacun des 12 mois d'une année.

5.2 Intégration des données annuelles

Bien sûr, certains établissements mensuels existent dans l'enquête annuelle; il est donc possible de copier la même répartition au niveau de l'établissement pour les 15 destinations qu'ils ont reçue pour l'EAMEF. La répartition pour le reste des établissements mensuels a été définie au moyen d'une méthode semblable à la troisième approche. Pour chacun de ces établissements, la répartition agrégée annuelle du groupe SCIAN-province ou SCIAN le plus précis comptant au moins cinq unités a été utilisée.

Avec une répartition pour chaque établissement, l'estimation peut être effectuée à l'aide d'un processus semblable à celui de l'EMIM pour la valeur globale des ventes. Lors des essais, on a réussi à produire l'équivalent d'une année de tableaux mensuels. Toutefois, en l'absence de processus de calendarisation, les tableaux sont demeurés très semblables d'un mois à l'autre, car les intrants de l'enquête annuelle étaient les mêmes.

5.3 Différences dans les calendriers de production

En général, pour une période de référence précise, l'EMIM publiera les données beaucoup plus tôt (jusqu'à deux ans) que l'EAMEF. Par exemple, les estimations pour juillet 2022 seraient publiées pour la première fois en septembre 2022 pour l'EMIM. En revanche, les estimations pour l'année 2022 ne seraient pas publiées pour l'EAMEF avant la fin de 2023 ou le début de 2024. Par conséquent, si les tableaux origine-destination étaient publiés selon le même calendrier que l'EMIM, il serait nécessaire d'intégrer les données annuelles de deux années précédentes.

Pour examiner l'effet possible de cette démarche, des tableaux mensuels ont été produits dans le cadre d'essais pour les 12 mois de 2021 en utilisant les données d'enquête annuelles de 2019 et 2021 comme intrants. Des essais préliminaires ont révélé que les différences n'étaient pas importantes, surtout à des niveaux plus généraux du SCIAN, mais une enquête plus approfondie est nécessaire.

6. Conclusions

Le nouveau tableau des ventes selon l'origine et la destination, y compris les indicateurs de qualité, a été publié en juillet 2024 pour les ventes annuelles des industries manufacturières de 2020 à 2022 à l'aide de l'Enquête annuelle sur les industries manufacturières de Statistique Canada. Pour ce faire, la répartition des ventes entre 15 destinations a été imputée au niveau de l'établissement pour les répondants à l'EAMEF correspondant à plusieurs provinces. L'approche d'imputation choisie à cet égard consistait à générer des facteurs initiaux en fonction des répartitions agrégées pour les répondants correspondant à une seule province dont le SCIAN et la province d'origine étaient semblables. Cette approche était préférable à d'autres en raison de son utilisation de données réelles sur les répondants. De plus, un processus itératif de quotient (« raking ») a été appliqué pour s'assurer que les totaux marginaux (ventes totales à chaque destination et ventes totales de chaque origine) étaient conservés pour chaque répondant.

6.1 Travaux futurs

Il est possible d'étendre les tableaux à l'Enquête mensuelle sur les industries manufacturières de Statistique Canada en intégrant les données annuelles. Pour les établissements mensuels qui ne peuvent pas être appariés à une contrepartie annuelle, les répartitions annuelles agrégées d'un groupe SCIAN-province ou SCIAN semblable pourraient être utilisées pour imputer une répartition.

Les effets des différences de calendrier de publication pourraient en outre ne pas être importants, mais une enquête plus approfondie est nécessaire. L'utilisation de la calendarisation serait également utile pour atténuer l'enjeu des tableaux demeurant très semblables d'un mois à l'autre si la même année de données annuelles est utilisée pour chaque mois.

On pourrait également produire un tableau semblable sur l'origine et la destination pour les enquêtes annuelle et mensuelle sur le commerce de gros de Statistique Canada.

7. Remerciements

Les auteurs tiennent à exprimer leur gratitude envers les nombreuses personnes ayant contribué à ce projet, et remercient également les réviseurs de leurs commentaires utiles.

Bibliographie

Statistique Canada (6 mars 2023). Enquête annuelle sur les industries manufacturières et de l'exploitation forestière (EAMEF). https://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&SDDS=2103.

Statistique Canada (2024a, 9 juillet). Tableau 36-10-0698-01 Origine et destination des ventes du secteur de la fabrication par industrie, annuel. Numéro DOI : <https://doi.org/10.25318/3610069801-fra>.

Statistique Canada (2024b, 22 novembre). Enquête mensuelle sur les industries manufacturières (EMIM). https://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&SDDS=2101.