

Enquête mensuelle sur le commerce de gros : Utilisation de données administratives

L'Enquête mensuelle sur le commerce de gros (EMCG) figure parmi plusieurs enquêtes-entreprises menées par Statistique Canada (SC) qui produisent des estimations servant à mesurer l'activité économique du Canada. Cette information est essentielle au processus décisionnel des gouvernements et du secteur privé. Le fardeau du répondant est un aspect important dont tient compte SC dans ses enquêtes. Afin d'atténuer ce fardeau, SC a fait de gros efforts ces dernières années pour rationaliser les données recueillies et orienter les enquêtes, dans la mesure du possible, vers l'utilisation de données administratives comme substitut aux données d'enquête.

L'EMCG est un des programmes d'enquête pour lesquels l'utilisation des données administratives existantes (données fiscales) pour remplacer les données d'enquête a été vérifiée et jugée appropriée. L'objectif du projet de la taxe sur les produits et services (TPS) consistait à utiliser ces données afin d'alléger le fardeau du répondant et de réduire les coûts associés à la collecte des données.

Ainsi, à partir du mois de référence d'octobre 2005, l'EMCG utilisera à la fois les données d'enquête et les données modélisées à partir des déclarations de TPS afin d'estimer le niveau des ventes de gros. Voici une brève description de la méthodologie vérifiée et maintenant utilisée pour ce processus.

À propos de la taxe sur les produits et services et de l'Enquête mensuelle sur le commerce de gros

Taxe sur les produits et services

La TPS, entrée en vigueur en 1991, est une taxe fédérale prélevée sur la consommation de biens et de services au Canada. La TPS est perçue par l'Agence du revenu du Canada (ARC) dans toutes les provinces, sauf au Québec. Toutes les provinces, sauf Terre-Neuve-et-Labrador, la Nouvelle-Écosse et le Nouveau-Brunswick, ont établi un taux de taxe de 7 % sur la valeur de la vente. À Terre-Neuve-et-Labrador, en Nouvelle-Écosse et au Nouveau-Brunswick, la taxe est une taxe de vente harmonisée (TVH) de 15 %, qui comprend la TPS et la taxe de vente de chaque province.

Toutes les entreprises, sauf celles dont les revenus sont inférieurs à 30 000 \$, doivent effectuer des versements de TPS. Les entreprises dont le chiffre d'affaires annuel est supérieur à 6 millions de dollars doivent effectuer des versements mensuels. Celles dont les revenus se situent entre 500 000 \$ et 6 millions de dollars par année doivent produire une déclaration trimestrielle et celles qui ont des revenus de 30 000 \$ à 500 000 \$ doivent le faire annuellement. Les déclarants mensuels et trimestriels doivent effectuer leurs versements dans les 30 jours suivant la date de fin de la période de déclaration, tandis que les déclarants annuels doivent faire leur versement dans les trois mois suivant la fin de la période de déclaration.

Le fichier de la TPS est transmis directement par l'ARC à la Division des données fiscales (DDF) de SC. La DDF traite ensuite le fichier uniquement à des fins statistiques à SC. Ce traitement procure une base de données complète et épurée qui est accessible pour les divers programmes d'enquêtes-entreprises de SC. Le traitement fait par la DDF comprend la correction des données erronées, la détection des valeurs aberrantes et le remplacement des données manquantes au moyen d'une calendrialisation et d'une extrapolation. Ce traitement ne vise pas à administrer ou à surveiller le programme de la TPS et aucune modification n'est transmise à l'ARC.

Enquête mensuelle sur le commerce de gros

L'EMCG est une enquête par sondage qui livre de l'information mensuelle sur les ventes et les stocks des grossistes canadiens.

Afin d'alléger le fardeau de réponse et de réduire les coûts de collecte des données, les plus petites unités de la population observée (environ la tranche inférieure de 5 %, d'après la valeur monétaire des ventes de chaque groupe de commerce par province) sont exclues de l'enquête. Cela signifie que des quelques 110 000 grossistes canadiens, seulement environ 35 000 ont la possibilité d'être choisis pour l'EMCG.

L'échantillon de l'EMCG est stratifié par secteur d'activité, province ou territoire et taille (fondée sur la valeur monétaire annuelle des ventes). Environ 8 000 des 35 000 unités de la base de sondage sont échantillonnées. Les unités restent les mêmes d'un mois à l'autre, sauf les nouvelles unités (naissances), qui sont échantillonnées avec la même probabilité que les unités de la base de sondage initiale.

Remplacement possible des données d'enquête par les données de la taxe sur les produits et services

L'échantillon de l'EMCG contient deux types d'unités : simples et complexes. Dans le cas des unités complexes (unités comprenant plus d'un établissement), une seule valeur de la TPS englobe souvent tous les établissements. Par conséquent, les données de la TPS disponibles actuellement ne permettent pas une ventilation par province ou territoire ou selon le SCIAN. Les unités complexes actives représentent environ 5 400 des unités de l'échantillon de l'EMCG et ont donc été jugées non admissibles à un remplacement par les données de la TPS. Les autres unités de l'échantillon sont des unités simples (unités qui représentent un seul établissement) et ont été jugées potentiellement admissibles à un remplacement par les données de la TPS (environ 2 600 unités actives).

La définition d'unité simple s'applique à la fois aux petits et aux grands établissements. Par conséquent, ce ne sont pas toutes les unités désignées comme simples qui peuvent être remplacées. Pour faire en sorte que l'utilisation des données de la TPS n'ait pas une incidence considérable sur les estimations globales, seules les unités simples, actives et non comprises dans les unités qui forment la tranche supérieure de 25 % des ventes d'un

groupe de commerce, la tranche supérieure de 50 % des stocks d'un groupe de commerce particulier ou la tranche supérieure de 50 % des ventes ou des stocks dans une province donnée servent au remplacement. Ainsi, les établissements importants ou dominants simples continueront d'être sondés.

Lien entre l'Enquête mensuelle sur le commerce de gros et la base de données fiscales

Il a fallu établir un lien entre le fichier de la TPS et l'échantillon de l'EMCG afin de pouvoir remplacer les données sur les ventes de l'EMCG par des données fiscales modélisées. Le Numéro d'entreprise (NE) a procuré ce lien et un certain nombre d'essais ont été réalisés afin de voir combien d'unités de l'échantillon étaient reliés individuellement à un fichier de la TPS, ce qui s'est révélé être le cas de la majorité des unités de l'échantillon. Toutefois, dans certains cas, les revenus de la TPS étaient nuls pour un établissement, tandis que les données déclarées pour les ventes étaient supérieures à zéro, et vice-versa. Ces cas particuliers ont été examinés en détail afin de déterminer la source de la différence.

Les unités de remplacement d'octobre 2005 ont été choisies de manière à ce qu'elles soient toutes reliées à un fichier de la TPS.

Corrélations entre l'Enquête mensuelle sur le commerce de gros et les revenus de la TPS

Afin de déterminer si les données sur les revenus de la TPS pouvaient permettre de remplacer adéquatement les données sur les ventes de l'EMCG, une bonne corrélation entre les ventes de gros et les revenus de la TPS était nécessaire. Les corrélations établies lors de la comparaison des valeurs de la TPS pour un mois particulier avec les valeurs de l'EMCG pour le même mois sont de bonne qualité et s'améliorent même lorsque les valeurs nulles et aberrantes sont supprimées.

Toutefois, en raison des dates de diffusion de l'EMCG et d'obtention des données fiscales de l'ARC, les données de la TPS n'étaient pas disponibles à temps afin de pouvoir être utilisées pour l'EMCG pour le présent mois de référence. Les données du fichier de la TPS du mois précédant le mois de référence de l'EMCG (p. ex., les données de février pour la TPS et le mois de référence de mars pour l'EMCG) sont reçues suffisamment tôt pour être intégrées dans le processus de l'EMCG. Il a donc été décidé de vérifier s'il était possible de concevoir un modèle utilisant les données du fichier de la TPS du mois précédant le mois de référence de l'EMCG. Comme l'illustre le tableau 1, la corrélation entre les ventes du mois courant de l'EMCG et les revenus de la TPS du mois précédent était de bonne qualité.

Tableau 1. Corrélations entre les ventes (mars 2004) et les revenus (février 2004)

Type d'unités	Coefficient de corrélation
Toutes les unités	0,8020
Répondants à valeur nulle supprimés	0,8621
Répondants à valeur nulle et données aberrantes supprimés	0,9214

Simulation du processus de remplacement fiscal

Bien que les corrélations entre les ventes de l'EMCG et les revenus de la TPS aient été très bonnes, le principal problème soulevé par l'utilisation des données de la TPS était la conservation des tendances et des niveaux de production actuels de l'EMCG. À cette fin, on a procédé à une simulation du processus pour la période allant de mars 2004 à février 2005.

Détermination des unités à remplacer

La première étape a consisté à déterminer les établissements se prêtant à un remplacement. Il s'agit des unités simples, actives et qui ne sont pas comprises dans la tranche supérieure de 25 % d'un groupe de commerce pour les ventes, dans la tranche supérieure de 50 % d'un groupe de commerce pour les stocks ou dans la tranche supérieure de 50 % d'une région géographique pour les ventes ou les stocks. Aux fins de la simulation, les domaines comprenaient les groupes de commerce (GC) et la géographie (GEO). Pour ce dernier domaine, les trois territoires ont été exclus en raison du faible nombre d'unités simples présentes dans ces régions géographiques.

Selon ces critères, 1 915 établissements simples ont été retenus pour le premier mois de la simulation (mars 2004). Les unités simples appartenant à un domaine autre que les GC ou la GEO ont été exclus en raison de leur importance sectorielle. Une fois ces unités retranchées, 920 unités simples ont constitué l'échantillon du remplacement à l'aide d'un modèle de la TPS. Les autres unités simples ont servi à construire ce modèle. Durant chaque mois de la simulation, seules les unités simples restantes qui ne sont pas aberrantes ont été utilisées pour le modèle de la TPS le mois en question pour produire les valeurs modélisées pour les 920 unités remplacées. Les 920 unités ont été sélectionnées à l'aide d'un échantillon systématique relativement au chiffre d'affaires annuel des 1 915 unités simples.

Modélisation

Modélisation des ventes

Pour chaque mois de la simulation, on a créé un modèle fondé sur les valeurs des ventes pour le mois en cours (M) et sur les valeurs des revenus de la TPS pour le mois précédent (M-1). Le modèle reposait sur la relation entre les ventes et les revenus de la TPS pour les unités retenues pour le remplacement qui faisaient partie de l'enquête et étaient toujours

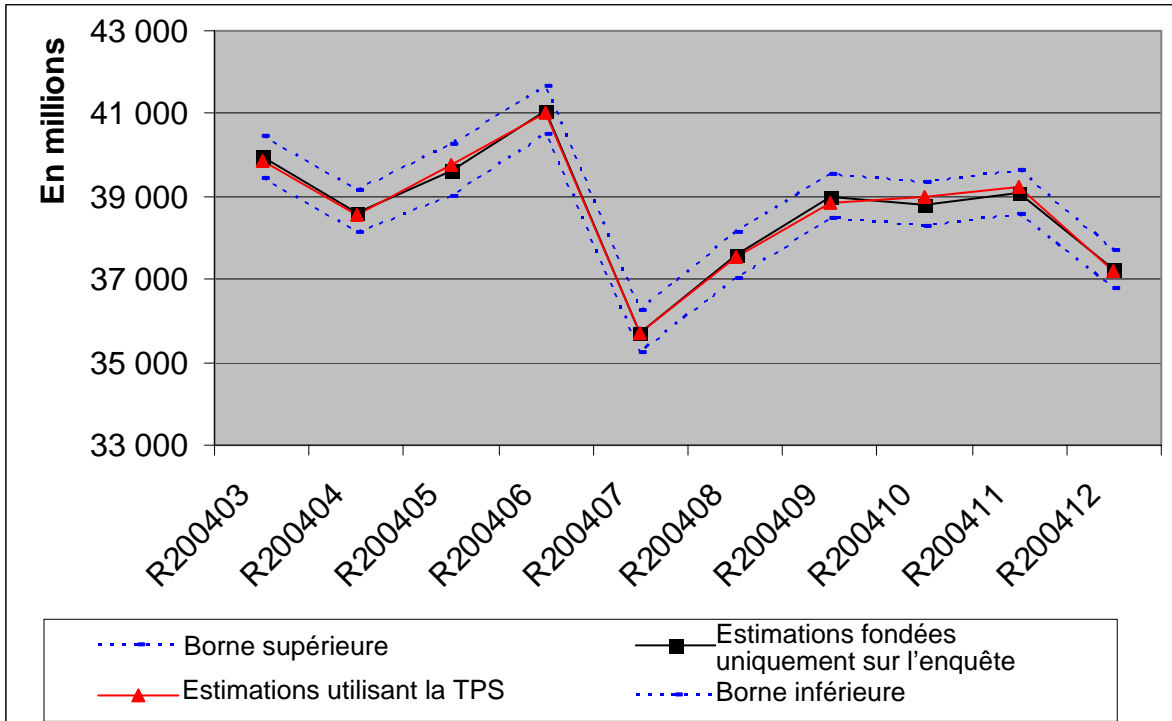
actives. Le modèle a ensuite été appliqué aux valeurs des revenus de la TPS pour les unités sélectionnées.

Pour faire en sorte qu'un modèle approprié soit créé, les données aberrantes ont été supprimées du modèle. Il s'agissait d'observations où la relation entre les valeurs des revenus de la TPS et celle des ventes déclarées divergeait de la majorité des unités du même groupe modèle. Pour tous ces cas, la valeur de la TPS pouvait être nulle, tandis que la valeur d'enquête correspondante pour les ventes était relativement importante, ou vice-versa.

Une fois le modèle élaboré, il a été appliqué aux unités à remplacer. Les données aberrantes produites par le modèle (unités entièrement hors limites comparativement à toutes les unités modélisées) ont été supprimées et remplacées par des données imputées.

Un des principaux avantages de cette approche est qu'après la modélisation des valeurs au micro niveau, le processus d'estimation actuel pouvait être utilisé. Comme le montre le graphique 1, les résultats de la simulation indiquent peu de différence entre les estimations contenant les données administratives et celles tirées exclusivement de l'enquête. La ligne des estimations fondées uniquement sur l'enquête correspond aux estimations fondées sur le sondage de toutes les unités, tandis que la ligne des estimations utilisant la TPS correspond aux estimations fondées sur le sondage d'unités autres que celles qui seront modélisées à l'aide des données de la TPS. L'intervalle de confiance de 95 % (indiqué par la borne inférieure et par la borne supérieure) donne à penser que les estimations utilisant la TPS restent bien en deçà des bornes de l'estimation initiale. La même hypothèse peut être faite pour des estimations plus détaillées, mais il convient de souligner qu'avec des renseignements plus détaillés sur les industries, la modélisation peut devenir moins fiable en raison d'une plus grande influence des données aberrantes.

Graphique 1 – Résultats de la simulation faite pour les ventes de gros totales au Canada



Modélisation des stocks

À l'heure actuelle, les données manquantes pour la variable des stocks en raison la non réponse ou de réponses reçues en retard sont imputées par le système d'imputation de l'EMCG. Les tendances des établissements de groupes géographiques ou du SCIAN semblables sont utilisées pour l'imputation des stocks de ces unités.

Diverses études exploratoires ont montré peu ou pas de corrélations entre les stocks et les variables tirées du fichier de la TPS. Par conséquent, il a été décidé d'utiliser le système d'imputation actuel de l'EMCG afin d'imputer les stocks de tous les établissements sélectionnés pour un remplacement par les données de la TPS.

Conclusion

La preuve a été faite que le remplacement des données d'enquête par des données administratives, plus précisément des données de la TPS, permet de conserver la qualité des estimations de l'EMCG, tout en atténuant le fardeau des répondants, en particulier pour les petites entreprises. Le nouveau projet offre également une plus grande souplesse dans le choix des unités à sonder annuellement, puisque les unités sondées peuvent être remplacées par des données administratives sans nuire aux estimations globales.

À partir d'octobre 2005, environ 879 unités (représentant 11 % de l'échantillon de l'EMCG) ne recevront pas de questionnaire. Leurs valeurs seront plutôt modélisées à l'aide des données de la TPS pour la variable des ventes. Les résultats qui ont été obtenus à l'aide du modèle actuel seront surveillés étroitement à l'avenir afin de déterminer s'il est possible d'utiliser davantage les données administratives.