

L'Enquête mensuelle sur le commerce de détail : utilisation de données administratives

L'Enquête mensuelle sur le commerce de détail (EMCD) figure parmi plusieurs enquêtes-entreprises menées par Statistique Canada (SC) qui produisent des estimations évaluant l'activité économique au Canada. Ces renseignements sont essentiels au processus décisionnel des gouvernements et du secteur privé. Le fardeau du répondant est un problème important auquel SC doit faire face pendant une enquête. Pour atténuer une partie de ce fardeau, SC a déployé de grands efforts au cours des dernières années afin de rationaliser les données recueillies et de substituer des données administratives aux données d'enquête, dans la mesure du possible.

L'EMCD est une des enquêtes dans le cadre desquelles on a tenté d'utiliser des données administratives existantes (données fiscales) comme substitut, données qui se sont avérées convenables pour remplacer les données d'enquête. L'objectif du projet de la taxe sur les produits et services (TPS) était d'utiliser des données administratives existantes au lieu des données d'enquête afin de réduire le fardeau du répondant et de réduire les coûts associés à la collecte.

Ainsi, à partir du mois de référence d'octobre 2005, l'EMCD utilisera des données d'enquête et des données modélisées à partir des déclarations de TPS afin d'estimer les niveaux des ventes au détail. Voici une brève description de la méthode qui a été testée puis utilisée dans ce processus.

À propos de la TPS et de l'Enquête mensuelle sur le commerce de détail

Taxe sur les produits et services

Mise en place en 1991, la TPS est une taxe fédérale prélevée sur la consommation de produits et de services au Canada. Dans toutes les provinces, sauf le Québec, elle est perçue par l'Agence du revenu du Canada (ARC). Toutes les provinces, sauf Terre-Neuve-et-Labrador, la Nouvelle-Écosse et le Nouveau-Brunswick, perçoivent cette taxe en faisant ajouter à la facture 7 % de la valeur de la vente. À Terre-Neuve-et-Labrador, en Nouvelle-Écosse et au Nouveau-Brunswick, la taxe est imposée sous la forme d'une taxe de vente harmonisée (TVH) de 15 % qui comprend la TPS et la taxe de vente de chaque province.

Toutes les entreprises, à l'exception de celles dont le revenu est inférieur à 30 000 \$, doivent effectuer des versements de TPS. Les entreprises qui affichent des revenus annuels supérieurs 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 faire des versements chaque trimestre. Celles qui génèrent des revenus se situant entre 30 000 \$ et 500 000 \$ font des paiements annuels. Les entreprises qui effectuent des versements au mois ou au trimestre doivent le faire au plus tard 30 jours après la fin de la période, tandis que celles qui font des envois annuels disposent d'un délai de trois mois après la fin de la période.

L'ARC envoie le fichier de la TPS à la Division des données fiscales (DDF) de Statistique Canada. La DDF effectue un traitement supplémentaire seulement pour des raisons statistiques. Le traitement fait en sorte que la base de données à laquelle accéderont les différents programmes d'enquêtes-entreprises soit propre et complète. Le traitement effectué par la Division des données fiscales comprend la correction des données inexactes, la détection d'observations aberrantes et le remplacement de données manquantes au moyen d'une calendrialisation et d'une extrapolation. Le traitement de la DDF n'a pas pour objet l'administration ou la surveillance du programme de TPS et aucune modification n'est envoyée à l'ARC.

Enquête mensuelle sur le commerce de détail¹

L'EMCD est une enquête échantillon qui permet d'obtenir des renseignements mensuels sur les ventes des détaillants canadiens.²

Afin de diminuer le fardeau de réponse et de réduire les coûts de collecte, les plus petites unités de la population observées (qui représentaient la valeur monétaire des ventes de chaque groupe de commerce des industries selon la province, soit environ 5 %) étaient exclues. On a ainsi pu réduire la base de sondage, qui est passée d'environ 215 000 points de vente au détail au Canada à 100 000 unités.

L'échantillon de l'EMCD est réparti en fonction de l'industrie, de la région géographique et de la taille (selon la valeur monétaire annuelle des ventes). Les régions de stratification sont les provinces et les territoires, ainsi que trois sous-régions provinciales. Un échantillon d'environ 11 000 unités est tiré sur les 100 000 unités qui restent dans la base de sondage. L'échantillon reste le même de mois en mois, sauf dans le cas des nouvelles unités (naissances), qui sont échantillonnées au moyen de la même probabilité qui est utilisée pour les unités de la population originale.

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

L'échantillon de l'EMCD comporte deux types d'unités : les unités simples et les unités complexes. En ce qui concerne les unités complexes (unités formées de plus d'un établissement), il arrive souvent qu'une seule valeur de la TPS corresponde à tous les établissements. Ainsi, les données de TPS actuellement disponibles ne permettent pas une répartition selon les régions géographiques ou par classe du SCIAN. Les unités complexes en vie représentent environ 4 000 unités de l'échantillon de l'EMCD et ont donc été jugées inadmissibles au remplacement par les données de TPS. Les autres unités de l'échantillon sont des unités simples (unités qui ne représentent qu'un établissement) et on a donc déterminé qu'elles pourraient être remplacées par des données de la TPS (environ 7 000 unités actives).

¹ Pour obtenir de plus amples renseignements sur la méthodologie de l'EMCD, consultez **Définitions, sources de données et méthodes : numéro d'enquête 2406**.

La définition d'une 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 d'incidence considérable sur les estimations en général, seules les unités qui sont simples, actives et qui ne sont pas comprises dans les unités qui représentent la portion supérieure, soit 25 %, d'un groupe de commerce d'une industrie ou la portion supérieure, soit 25 % d'une certaine région géographique, en ce qui concerne les ventes, peuvent être remplacées. Ainsi, les établissements simples dominants ou de grande taille ne seront pas exclus de l'enquête.

Lien entre l'Enquête mensuelle sur le commerce de détail et la base de données de la TPS

Il fallait établir un lien entre le fichier de la TPS et l'échantillon de l'EMCD afin de pouvoir remplacer les données sur la livraison de l'EMCD par les données fiscales modélisées. On a déterminé que ce lien était le Numéro d'entreprise (NE) et on a effectué un certain nombre de tests pour découvrir combien d'unités de l'échantillon étaient liées de façon individuelle à un dossier de la TPS. Il en a été ainsi pour la plupart des unités de l'échantillon. Il y a cependant eu des situations où un établissement n'avait accumulé aucun revenu de la TPS alors que ses données relatives aux ventes étaient supérieures à zéro, et vice versa. Ces cas ont été étudiés en détail afin de trouver la cause de la différence.

Les unités de substitution d'octobre 2005 ont été choisies de sorte qu'elles correspondent toutes à un dossier de la TPS.

Corrélation entre l'Enquête mensuelle sur le commerce de détail et les revenus de la TPS

Afin d'utiliser les données modélisées des revenus de la TPS comme substitut convenable aux données sur les ventes de l'EMCD, il fallait trouver une bonne corrélation entre les ventes au détail et les revenus de la TPS. Lorsque l'on compare les valeurs de la TPS d'un mois donné aux valeurs de l'EMCD du même mois, les corrélations sont de bonne qualité et même s'améliorent lorsqu'on élimine les valeurs nulles et les observations aberrantes.

En raison de contraintes de temps associées aux dates de publication de l'EMCD et à l'obtention des données fiscales auprès de l'ARC, les données fiscales de la TPS ne sont pas disponibles à temps pour être utilisées par l'EMCD pour le mois courant. Les données de la TPS qui sont un mois avant le mois courant (p. ex. données de la TPS pour le mois de février et mois de référence de mars de l'EMCD) sont reçues suffisamment tôt pour être incorporées dans le processus de traitement des données de l'EMCD. Il a été décidé de vérifier si un modèle peut être développé en utilisant les données de la TPS d'un mois d'avance par rapport au mois de référence de l'EMCD. Comme on peut le voir

dans le Tableau 1, la corrélation entre les ventes du mois actuel de l'EMCD et les revenus de la TPS d'il y a un mois était de bonne qualité.

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

Type des unités	Coefficient de corrélation
Total	0,5644
Unités rapportant 0 \$ enlevées	0,9097
Unités aléatoires et rapportant 0 \$ enlevées	0,9632

Simulation du processus de remplacement par les données fiscales

Même si la corrélation entre les ventes de l'EMCD et les revenus de la TPS était très bonne, la principale préoccupation concernant l'utilisation de données de la TPS était la préservation des tendances et des niveaux actuels de production de l'EMCD. Afin de s'assurer de ceci, une simulation de la méthode a été effectuée pour la période allant de mars 2004 à février 2005.

Détermination des unités devant être remplacées

Il fallait premièrement déterminer quels établissements pouvaient être remplacés. Ces unités sont définies comme simples, actives, et ne figurant pas parmi les unités qui forment les premiers 25 % du domaine d'intérêt (groupe de commerce d'industrie ou région géographique). Dans le cadre de la simulation, les domaines étaient formés des groupes de commerce de l'industrie et des géographies. La seule exception était le groupe de commerce 010 (Concessionnaires d'automobiles neuves) qui était exclu en raison de son importante incidence sur l'estimation globale.

On a donc considéré 4 547 établissements comme propices à la substitution au cours du premier mois de la simulation (mars 2004). Les unités simples au sein d'autres groupes de commerce et géographies ont été exemptées du remplacement puisqu'il a été déterminé qu'elles sont importantes pour l'industrie ou qu'il s'agit d'unités faisant également partie de l'échantillon de l'Enquête trimestrielle sur les marchandises vendues au détail (ETMVD). Les unités faisant également partie de l'échantillon de l'ETMVD (2 814 unités simples au total) ne peuvent être remplacées, car il est difficile de déterminer les ventes de marchandises au moyen des données de la TPS. En raison des unités faisant partie de l'échantillon de l'ETMVD et des unités propres à une industrie, la portion finale à être remplacée par des données modélisées de la TPS compte 1 636 unités. Les unités simples restantes servent à établir le modèle TPS. Au cours de chaque mois de la simulation, seules les unités simples restantes qui ne sont pas considérées comme observations aberrantes sont utilisées dans le modèle TPS du mois en question afin de produire les valeurs modélisées pour les 1 636 unités faisant l'objet d'un remplacement.

Modélisation

Modélisation des ventes

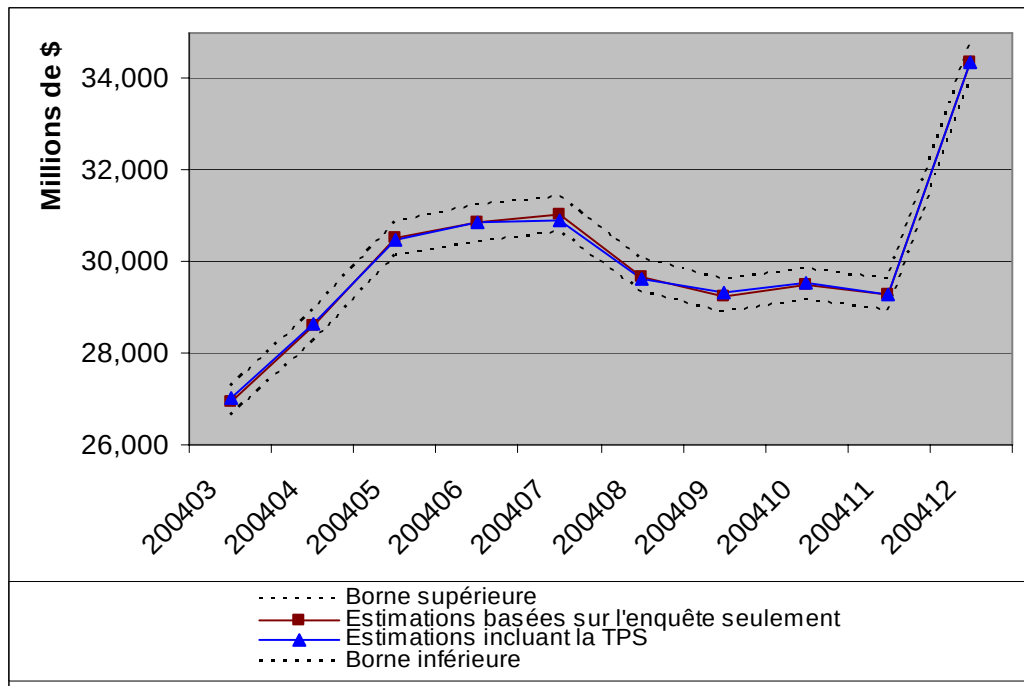
Pour chaque mois de la simulation, on a créé un modèle basé sur les données des ventes du mois actuel (M) ainsi que sur les données des revenus de la TPS du mois précédent (M-1). Le modèle était fondé sur la relation entre les ventes et les revenus de la TPS des unités qui pouvaient être remplacées mais qui sont demeurées dans la portion enquêtée et qui étaient toujours en vie. Ce modèle a ensuite été appliqué aux données des revenus de la TPS des unités devant être remplacées.

De façon à s'assurer de la création d'un modèle convenable, les observations aberrantes ont été retirées du groupe modèle. Il s'agissait d'observations dont la relation entre les revenus de la TPS et les données déclarées des ventes ne correspondaient pas à la majorité des unités du groupe modèle. Il y a eu des situations où un établissement n'avait accumulé aucun revenu de la TPS, tandis que les estimations de la valeur des ventes de l'enquête étaient assez importantes, et vice versa.

Après avoir été produit, le modèle a été appliqué aux unités devant être remplacées. On a également éliminé les observations aberrantes (unités qui sont entièrement hors de portée lorsque comparées avec toutes les unités du modèle) de cette population. Les dossiers d'observations aberrantes générées par le modèle ont été remplacés par le processus normal d'imputation.

L'un des principaux avantages de cette approche était qu'une fois les valeurs modélisées au niveau micro-économique, il était possible d'utiliser le processus d'estimation actuel. Comme le montre le Graphique 1, les résultats de la simulation révèlent peu de différences entre les estimations contenant des données administratives et celles qui ont été calculées uniquement à partir de l'enquête. La ligne "Estimations basées sur l'enquête seulement" correspond aux estimations fondées sur les données de l'ensemble des unités tandis que la ligne "Estimations utilisant la TPS" correspond aux estimations fondées sur les unités d'enquête autres que celles dont les données seront modélisées selon les données de la TPS. L'intervalle de confiance de 95 % (indiquée par les bornes inférieure et supérieure) laisse supposer que les estimations "Estimations utilisant la TPS" demeurent bien à l'intérieur des limites de la première estimation. Cette hypothèse peut être appliquée à des estimations plus détaillées, mais il faut noter qu'avec des renseignements plus détaillés sur les industries, la modélisation peut être moins fiable et les observations aberrantes peuvent avoir une plus grande incidence.

Graphique 1 : Résultats de la simulation du total des ventes au détail au niveau canadien



Conclusion

Il a été démontré que la substitution de données d'enquêtes par des données administratives, particulièrement par des données de la TPS, conservait la qualité élevée des estimations de l'EMCD tout en réduisant le fardeau du répondant, surtout en ce qui a trait aux petites entreprises. Cette nouvelle initiative permet également une plus grande flexibilité par rapport à la sélection d'unités à être interviewées, car celles-ci peuvent être remplacées par des données administratives sans qu'il y ait d'incidence sur les estimations en général.

Dès le mois d'octobre 2005, quelque 1 606 unités (représentant 14 % de l'échantillon de l'EMCD) ne recevront pas de questionnaire. Plutôt, leurs valeurs seront modelées en utilisant des données de la TPS pour la variable des ventes. Les résultats obtenus au moyen du modèle courant seront suivis de près à l'avenir afin de déterminer s'il est possible d'utiliser davantage les données administratives.