

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

L'Enquête mensuelle sur le commerce de gros (EMCG) figure parmi plusieurs enquêtes auprès des entreprises menées par Statistique Canada (StatCan) 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 StatCan dans ses enquêtes. Afin d'atténuer ce fardeau, StatCan 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.

## **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 l'Ontario, l'Île-du-Prince-Édouard, Terre-Neuve-et-Labrador, la Nouvelle-Écosse et le Nouveau-Brunswick, ont établi un taux de taxe de 5 % sur la valeur de la vente. À l'Île-du-Prince-Édouard, 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. En Ontario, la TVH est de 13%.

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 1 500 000 \$ et 6 millions de dollars par année doivent produire une déclaration trimestrielle et celles qui ont des revenus de 30 000 \$ à 1 500 000 \$ 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 administratives (DDA) de StatCan. La DDA traite ensuite le fichier uniquement à des fins statistiques à StatCan. Ce traitement procure une base de données complète et épurée qui est accessible pour les divers programmes d'enquêtes auprès des entreprises de StatCan. Le traitement fait par la DDA 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 10 %, d'après la valeur monétaire des ventes de chaque groupe de commerce par province) sont exclues de l'enquête. Cela signifie que des quelques 100 000 établissements du secteur du commerce de gros ciblés par l'enquête, seulement environ 20 000 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 12 000 des 20 000 établissements de la base de sondage sont échantillonnés. 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.

Dans le cadre de l'Enquête mensuelle sur le commerce de gros (EMCG), une utilisation considérable des données de la TPS est faite depuis maintenant plus de dix ans. L'introduction des données fiscales visait principalement à alléger le fardeau de réponse, surtout pour les entreprises de taille moyenne. Étant donné que, dans le cadre de l'EMCG, on communique avec les répondants sur une base mensuelle et que les unités demeurent dans l'échantillon pendant environ cinq ans, on comprend que le fait d'être sélectionné pour cette enquête est assez exigeant.

Au début des années 2000, on a commencé à observer une baisse dans les taux de réponse à l'EMCG, surtout chez les entreprises de taille moyenne. La situation est devenue tellement préoccupante qu'il fallait agir avant que cette enquête doive faire l'objet d'une restratification. Il importe de mentionner que l'EMCG fait l'objet d'une restratification environ tous les cinq ans. En gros, ce processus consiste à recalculer les limites des strates et à sélectionner un nouvel échantillon. Évidemment, les grandes unités dans la population de l'enquête sont si importantes pour l'économie qu'elles feront toujours partie de l'échantillon, même après une restratification. Cependant, les unités de taille moyenne sont retirées de l'ancien échantillon et sont remplacées dans le nouvel échantillon par des unités de la population sélectionnées au hasard. Or, comme le temps n'était pas venu de restratifier cette enquête, il fallait agir autrement pour alléger le fardeau des entreprises de taille moyenne. L'idée a été lancée d'utiliser les données fiscales pour modéliser leurs données d'enquête.

On a mené des études pour déterminer dans quelle mesure il existait une corrélation entre les données fiscales et les données d'enquête pour les entreprises de taille moyenne. Comme on pouvait s'y attendre, la corrélation entre leur revenu tiré des données fiscales et leurs ventes totales déclarées dans le cadre de l'enquête était très forte. D'après ce résultat, la stratégie suivante a été adoptée. Il a été décidé que certaines des entreprises de taille moyenne les plus petites qui avaient déclaré des ventes stables depuis leur sélection

pour l'enquête ne recevraient plus de questionnaire. Ces unités étaient communément appelées les unités S2. Leurs données d'enquête seraient plutôt imputées au moyen d'une valeur obtenue à l'aide d'un modèle de régression linéaire sur des données fiscales et des données d'enquête ajusté aux entreprises de taille similaire qui ont reçu un questionnaire et ont répondu à l'enquête. Ce deuxième groupe d'unités était appelé les unités S1. Au total, environ 1 000 unités faisaient partie du groupe S2. Sans être parfaite du point de vue méthodologique, cette solution provisoire atteignait son but et permettait tout de même d'obtenir des données de très grande qualité compte tenu du lien très étroit observé entre les données fiscales et les données d'enquête. Cependant, il apparaissait clairement que cette solution devrait ultérieurement être remplacée par une approche méthodologique plus solide. Au cours des dernières années, on a déployé des efforts pour utiliser les données fiscales à l'étape de l'estimation par la mise en œuvre d'un estimateur par le quotient.

L'utilisation d'un estimateur par le quotient dans les enquêtes mensuelles n'est pas une idée nouvelle. Dès 2000, des études sur le sujet ont été réalisées. À cette époque, les données de la TPS n'étaient disponibles que depuis quelques années pour être utilisées comme variables auxiliaires dans les enquêtes, et certains concepts sous-jacents à ces données n'étaient pas totalement compris et documentés. De plus, le système de traitement de ces données n'était pas aussi élaboré que le système utilisé aujourd'hui.

Au cours des quinze dernières années, la méthodologie pour traiter les données de la TPS n'a cessé de s'améliorer, tant du côté de la calendarisation, de l'imputation que de la répartition des données des entreprises aux établissements. Le tout est bien documenté et la qualité des données est maintenant excellente. L'utilisation des données de la TPS à l'aide d'un estimateur par le quotient s'avère donc aujourd'hui une avenue des plus prometteuses.

L'estimation par le quotient consiste à remplacer les poids de sondage initiaux (définis comme étant l'inverse de la probabilité de sélection dans l'échantillon) par de nouveaux poids de façon à satisfaire des contraintes de calage (le total d'une variable auxiliaire estimé à partir de l'échantillon doit être égal au total de cette même variable auxiliaire obtenu à partir de toutes les unités de la population) et de sorte que ces nouveaux poids d'estimation soient aussi près que possible (selon une certaine mesure de distance) des poids de sondage initiaux.

Supposons, par exemple, que le total connu de la variable auxiliaire pour la population soit égal à 100 et qu'il soit estimé à 90 en se fondant sur un échantillon; nous obtenons donc une sous-estimation d'environ 10 %. Puisque nous connaissons le total de la variable auxiliaire pour la population, il serait raisonnable d'augmenter les poids des unités échantillonnées afin que l'estimation soit exactement égale à cette valeur. Puisque la variable d'intérêt est liée à la variable auxiliaire, il n'est pas déraisonnable de penser que l'estimation des ventes fondée sur le même échantillon et les mêmes poids que l'estimation de la variable auxiliaire puisse également représenter une sous-estimation d'environ 10 %. Si c'est en effet le cas, les poids ajustés pourraient donc permettre de produire un autre estimateur des ventes totales. Cet autre estimateur s'appelle l'estimateur par le quotient.

L'estimateur par le quotient essaie essentiellement de compenser les échantillons « malchanceux » et rapproche l'estimation du total réel. Le gain de variance dépendra de la solidité du lien entre la variable d'intérêt et les données auxiliaires.

Une caractéristique intéressante de l'estimateur par le quotient est qu'il peut permettre d'obtenir une estimation pour la population entière, y compris les unités les plus petites exclues de l'enquête. Cette estimation se fait en incluant simplement la portion à tirage nul dans les totaux de contrôle pour la portion échantillonnée. Cela augmente les poids de la portion échantillonnée de telle sorte que les estimations seront ajustées afin de prendre en compte la portion à tirage nul.

## 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 être utilisées comme variable auxiliaire dans l'estimation par le quotient, 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 ne sont 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 par contre reçues suffisamment tôt pour être intégrées dans le processus 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 est aussi 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

L'utilisation des données fiscales dans l'EMCG a crû constamment au cours des ans. Cette utilisation a d'abord débuté avec des méthodes simples, puis a culminé avec l'adoption récente d'un estimateur par le quotient.