

Enquête annuelle sur le secteur des services personnels, 1999 – Concepts, méthodologie et qualité des données

Ce document contient des renseignements permettant de bien comprendre les concepts de base définissant les données fournies, les méthodes qui sous-tendent l'enquête et les aspects clés de la qualité des données. Ces renseignements vous aideront à mieux comprendre les points forts et les limites des données ainsi que la façon appropriée de les utiliser et de les analyser. Les renseignements pourront vous être particulièrement utiles au moment de comparer les données à celles d'autres enquêtes ou sources d'information et de tirer des conclusions au sujet du changement au fil des ans.

Définitions

Une **entité commerciale** est un agent économique qui a la responsabilité et le pouvoir d'affecter des ressources pour la production de biens et de services.

Un **établissement statistique** est une entité de production ou le plus petit groupe d'entités de production qui produit un ensemble de biens et/ou de services homogènes, dont les activités ne débordent pas les frontières provinciales et qui est en mesure de fournir des données sur la valeur de la production ainsi que sur le coût des principaux facteurs intermédiaires utilisés et le coût et l'importance de la main-d'œuvre affectée à la production.

Une **entreprise statistique** est une unité organisationnelle d'une unité commerciale qui dirige et contrôle l'affectation de ressources intéressant ses activités nationales et pour laquelle on établit des états financiers et des bilans consolidés à partir desquels il est possible de tirer des renseignements sur les transactions internationales, la situation des investissements internationaux et la situation financière générale de l'entité.

Une **entreprise complexe** est une entreprise qui est active dans plus d'une province et/ou dans plus d'un secteur et/ou qui compte plusieurs entités légales (liens avec plusieurs entités légales).

Une **entreprise simple** est une entreprise qui est active dans une province, dans un secteur et qui compte une seule entité légale.

Les firmes complexes peuvent comporter une unité organisationnelle qui se situe au-dessus du niveau de l'établissement, mais au-dessous du niveau de l'entreprise : la compagnie. La *compagnie* est une unité statistique qui correspond à l'unité organisationnelle pour laquelle on établit des comptes des revenus et des dépenses et des bilans à partir desquels il est possible de calculer les bénéfices d'exploitation et le rendement du capital.

Couverture et base de sondage

La population cible est constituée de tous les établissements statistiques considérés comme des services personnels et services de blanchissage dans le Système de Classification des Industries de l'Amérique du Nord (SCIAN) au cours de l'année de référence 1999. Les utilisateurs de données qui désirent en savoir plus sur le SCIAN, sur les principes sous-jacents et sur de nombreux autres concepts statistiques abordés dans le présent sommaire, pourront consulter l'introduction de la publication de Statistique Canada intitulée « Système de Classification des Industries de l'Amérique du Nord : Canada 1997 » (n° 12-501 au catalogue).

Même si l'objectif de l'enquête consiste principalement à produire des estimations pour l'ensemble de la branche d'activité – établissements constitués et non constitués – on a établi que la partie de la population admissible à l'échantillonnage se composait de tous les établissements constitués ayant un chiffre d'affaires supérieur à 50 000 \$. Certaines unités non constituées ayant une contribution exceptionnellement importante ont été ajoutées à l'étape de la collecte directe des données. Le même principe s'applique aux unités non constituées appartenant à des entreprises complexes. L'exclusion de la collecte directe des données des entreprises non constituées et des entreprises constituées en deçà du seuil de 50 000 \$ a pour objectif principal de réduire appréciablement le fardeau de réponse. Le segment exclu représente une proportion importante du nombre d'entités commerciales (68 %), mais représente seulement une portion modeste de l'estimation finale (22 %). Les entreprises en deçà du seuil d'exclusion font tout de même partie de l'univers, mais elles figurent dans les estimations finales par l'entremise d'information provenant d'enregistrements administratifs. Seuls des renseignements de base peuvent être obtenus de cette source, c'est-à-dire : le total des recettes, les dépenses, l'amortissement et les salaires, traitements et avantages sociaux. Pour l'année de référence de 1999, les données administratives utilisées dans les estimations des entreprises non constituées ont été limitées aux données qui pouvaient être liées directement à la base de sondage. Ceci devrait améliorer la qualité des estimations des entreprises non constituées. Ce changement va nécessiter des révisions aux données historiques pour certaines industries. Des caractéristiques détaillées comme les catégories de clients, les recettes selon le type de service et les postes de dépenses détaillés ne peuvent être obtenues que pour des entreprises relevant de la collecte directe des données.

La base de sondage est la liste servant à identifier les entreprises classées selon la branche d'activité en question. La base de sondage est tenue à jour dans le Registre des entreprises de Statistique Canada à l'aide des données des comptes fiscaux (c.-à-d. les enregistrements de données sur l'impôt sur le revenu, la taxe sur les produits et services et les retenues sur la paie) traitées par l'agence des douanes et du revenu du Canada.

Plan d'échantillonnage

Le plan d'échantillonnage concerne uniquement la partie de la base de sondage relevant de la collecte directe des données. Avant la sélection d'un échantillon aléatoire, les unités sont partagées en groupes homogènes définis en fonction d'attributs sectoriels (SCIAN) et géographiques (province ou territoire). Des exigences de qualité semblables sont prévues pour chaque groupe, avec répartition en quatre sous-groupes appelés strates : tirage obligatoire, tirage complet, tirage partiel (grandes entreprises) et tirage partiel (petites entreprises).

La strate à tirage complet englobe les principales sociétés en termes de rendement industriel. Ces dernières sont incluses dans l'échantillon avec certitude, ce qui en fait des unités autoreprésentatives. La strate à tirage obligatoire est également constituée d'unités autoreprésentatives ayant une structure complexe (entreprises à plusieurs établissements, à plusieurs entités légales, ayant plusieurs codes SCIAN ou se trouvant dans plusieurs provinces). Les unités appartenant aux deux strates à tirage partiel relèvent d'un échantillon aléatoire, chaque entreprise échantillonnée représentant d'autres entreprises semblables de la combinaison secteur/province suivant l'inverse de la probabilité de sélection.

Enfin, la taille de l'échantillon a été augmentée afin de compenser pour des situations comme la non-réponse et les sociétés qui ne peuvent pas être jointes parce qu'elles ont déménagé ou fermé leurs portes. La taille finale de l'échantillon pour cette enquête, une fois exclues les

entreprises qui n'auraient pas dû faire partie de la base de sondage (hors du champ d'observation, enregistrements en double, entreprises qui ne sont plus actives), était de 1497 sociétés.

Collecte des données

Les données ont été recueillies par envoi et retour par la poste, mais les répondants ont eu la possibilité de répondre par téléphone ou d'utiliser d'autres méthodes de déclaration électronique. Même si l'unité d'échantillonnage était l'établissement statistique, la compagnie statistique a été choisie comme unité de collecte afin de réduire le fardeau de réponse et simplifier la collecte. Par conséquent, les sociétés dont la production était répartie en plus d'un endroit ont reçu un seul questionnaire par la poste, avec instruction de regrouper dans leur questionnaire toutes leurs opérations pour la branche d'activité en question. Des données sommaires ont été recueillies pour chaque province ou territoire dans lequel la société exerçait ses activités.

Vérification et imputation

Les données recueillies sont sujettes à plusieurs vérifications afin d'en assurer la cohérence interne et cerner les valeurs extrêmes. En cas de non-réponse, on exécute une imputation à l'aide de la procédure « du plus proche voisin » (imputation par enregistrement donneur). Cette méthode fait appel à l'information auxiliaire dont on dispose sur le non-répondant pour identifier une société répondante ayant des caractéristiques semblables. L'enregistrement identifié selon cette procédure est par la suite utilisé comme substitut. En plus de ce type d'imputation, on a recours à une procédure d'imputation historique qui consiste à inférer une valeur imputée à partir des données provenant d'une enquête ultérieure (pour la même société).

Estimation

Avant l'étape de l'estimation, les données pour des sociétés ayant des activités de production dans plus d'une province ou plus d'un territoire ont été désagrégées au niveau de la province. Les données d'enquête recueillies auprès de l'échantillon sont ensuite pondérées en utilisant l'inverse de la probabilité de sélection de chaque unité échantillonnée afin de produire des estimations représentatives de la population visée. Les estimations relatives aux entreprises exclues de l'enquête (c.-à-d. les entreprises non constituées et les entreprises constituées ayant un chiffre d'affaires inférieur à 50 000 \$) ont été calculées en se servant des données administratives.

Les résultats combinés de l'enquête sont analysés avant d'être publiés; en général, cette opération consiste à examiner en détail les réponses individuelles (surtout celles des grandes sociétés), à revoir la conjoncture générale et les tendances historiques et enfin à comparer ces résultats aux données fiscales et à d'autres sources de données administratives (associations industrielles et de métiers).

Qualité des données

Malgré les efforts déployés afin que les opérations de collecte et de traitement se déroulent en fonction de normes rigoureuses, les estimations qui en résultent comportent inévitablement des erreurs. Celles-ci sont classées en deux grandes catégories : erreurs d'échantillonnage et erreurs non dues à l'échantillonnage.

Les erreurs non dues à l'échantillonnage ne sont pas liées à l'échantillonnage et peuvent survenir pour de nombreuses raisons. Par exemple, la non-réponse est une source importante d'erreurs non dues à l'échantillonnage. La couverture de la population, les écarts dans l'interprétation des questions, les mauvaises réponses des répondants et les erreurs de consignation, de codage et de traitement des données représentent d'autres exemples d'erreurs non liées à l'échantillonnage.

Le taux de réponse pour cette enquête a été de 73,9 %, compte tenu du fait que certaines entreprises n'étaient plus en affaires ou qu'elles avaient modifié leur activité principale.

Des erreurs d'échantillonnage risquent de se produire parce que les estimations proviennent d'un échantillon de la population plutôt que de la population entière. Ces erreurs dépendent de divers facteurs, notamment la taille de l'échantillon, le plan d'échantillonnage et la méthode d'estimation. Une des propriétés importantes de l'échantillonnage probabiliste tient au fait que les erreurs d'échantillonnage peuvent être calculées à partir de l'échantillon à l'aide d'une mesure statistique appelée le coefficient de variation (CV). Pour des enquêtes menées à plusieurs reprises, l'écart relatif entre l'estimation d'un échantillon et l'estimation qu'on aurait dû obtenir d'un dénombrement de toutes les unités serait inférieur à deux fois le coefficient de variation, 95 fois sur 100. Les intervalles de confiance peuvent être édictés autour des estimations en utilisant les CV. Premièrement, nous calculons l'erreur type en multipliant l'estimation d'un échantillon par le CV. L'estimation d'un échantillon plus ou moins deux fois l'erreur type ce qu'on appelle l'intervalle de confiance de 95 %.

Pour l'Enquête annuelle de 1999 sur le secteur des services personnels, les CV ont été calculés pour chaque estimation produite. De façon générale, les variables déclarées plus fréquemment ont résulté en de très bons CV (10 % ou moins), tandis que les variables déclarées moins souvent ont été associées à des CV plus élevés mais néanmoins acceptables (moins de 25 %). On peut obtenir ces CV sur demande.