

Rapport méthodologique sur l'Enquête sur le secteur de la géomatique, 2004

1. Introduction

Statistique Canada (SC), en partenariat avec Ressources naturelles Canada (RNCCan), a mené une enquête pour l'année de référence 2004 en vue de produire de nouvelles données statistiques sur les entreprises qui produisent des biens géomatiques ou fournissent des services géomatiques au Canada. Les données de cette enquête seront utilisées par Ressources naturelles Canada pour déterminer la position concurrentielle du secteur ainsi que ses besoins en matière de politiques et de programmes du gouvernement fédéral.

2. Population cible et résultats de la collecte

La population visée par cette enquête se compose de tous les établissements qui produisent des biens ou qui fournissent des services géomatiques. La base de sondage élaborée pour l'enquête comprend tous les établissements dans le Registre des entreprises (RE) sous le code 541360 du SCIAN (Services de prospection et de levé géophysiques) ou le code 541370 du SCIAN (Services d'arpentage et de cartographie sauf les levés géophysiques). Les autres unités ajoutées à la base de sondage comprenaient les établissements ayant indiqué mener des activités de géomatique dans le cadre de l'Enquête sur l'innovation de 2003. Enfin, RNCCan a dressé une liste des établissements dont on croit qu'ils fournissent des services géomatiques ou produisent des biens géomatiques. Après élimination des doubles comptes, la base de sondage comprenait au total 4 979 établissements. On a procédé à un contact préliminaire avant l'envoi du questionnaire par la poste pour déterminer si les établissements étaient en exploitation et s'ils faisaient partie du champ de l'enquête. Durant le contact préliminaire, on a déterminé que certains établissements étaient hors du champ de l'enquête, avaient fermé leurs portes ou constituaient des enregistrements en double et, par conséquent, ils n'ont fait l'objet d'aucun envoi par la poste. Un questionnaire a été envoyé à 4 202 établissements, c.-à-d. à tous ceux faisant partie de la base de sondage sauf ceux révélés par le contact préliminaire être hors du champ de l'enquête. Sur les 4 202 établissements, 1 598 (38 %) étaient introuvables ou impossibles à joindre, 1 249 (30 %) ont rempli le questionnaire, 113 (3 %) ont refusé de le remplir mais ont été confirmés comme étant admissibles, 1 074 (26 %) étaient hors du champ de l'enquête ou avaient fermé leurs portes et 168 (4 %) étaient d'autres types d'unités hors du champ de l'enquête. Nous devons souligner que les unités hors du champ de l'enquête se trouvaient principalement en Alberta. En fait, 40 % des unités hors du champ de l'enquête repérées durant le contact préliminaire et 60 % de celles repérées au moyen du questionnaire se trouvaient en Alberta. Dans l'ensemble, le taux de réponse à l'enquête était de 42 %. Il s'agit là d'un taux de réponse habituel pour les enquêtes-entreprises de Statistique Canada. Le taux de réponse visé pour cette enquête était de 40 %.

3. Traitement des données

Pour tous les établissements qui ont rempli le questionnaire, les données ont été saisies et vérifiées afin de détecter les erreurs et les incohérences. Dans le cas de renseignements de qualité douteuse sur les revenus, les données ont été comparées aux données fiscales administratives pour le secteur de la géomatique qui ont été analysées par la Division des industries de service de Statistique Canada. Dans certains cas, il a été nécessaire de communiquer de nouveau avec les répondants pour obtenir des précisions. Les données manquantes ont alors été remplacées par des

valeurs uniformes (imputées) à l'aide de la procédure DONOR dans le système BANFF. Cette procédure utilise la méthode du plus proche voisin pour trouver, pour chaque établissement pour lequel il faut procéder à une imputation, l'établissement valide qui lui ressemble le plus et qui permet d'imputer toutes les variables requises de l'établissement en transférant les valeurs correspondantes de l'enregistrement du plus proche voisin. Il convient de souligner que, pour la variable dépenses, au lieu de transférer les valeurs correspondantes de l'enregistrement du plus proche voisin à l'enregistrement pour lequel une imputation était requise, nous avons utilisé le ratio des dépenses et du revenu total de l'enregistrement donneur et nous avons appliqué ce ratio à l'établissement pour lequel les dépenses devaient être imputées. Les taux d'imputation les plus élevés sont ceux pour les variables recettes et dépenses et pour les questions liées aux obstacles à la R-D et aux exportations. Le taux d'imputation pour la variable recettes est 20 %, le taux pour la variable dépenses, 42 % et les taux pour les obstacles à la R-D et aux exportations sont 25 % et 36 %, respectivement. Pour le reste des variables, les taux d'imputation sont inférieurs à 10 %.

Pour l'année de référence 2004, la Division des industries de service a produit une estimation sectorielle fondée sur un échantillon de données administratives et historiques pour les établissements classés sous les codes 541360 et 541370 du SCIAN.

L'Enquête sur le secteur de la géomatique, 2004, a produit deux estimations fondées sur un recensement des établissements menant des activités de géomatique (tous des établissements connus), soit :

- une estimation sectorielle pour les établissements portant les codes 541360 et 541370 du SCIAN, comparable aux résultats administratifs de la Division des industries de service;
- une estimation de l'activité pour les établissements menant des activités de géomatique.

Une comparaison des estimations sectorielles produites par les deux projets a donné des résultats comparables pour les revenus au niveau national (résultat des données administratives = 1,95 milliard de dollars comparativement au résultat de l'enquête = 1,69 milliard de dollars).

L'écart des revenus pour ces deux codes du SCIAN est attribuable à plusieurs facteurs :

1. Le projet des données administratives prévoyait un traitement spécial pour les entreprises complexes. Par exemple, dans le cas de certaines grandes entreprises complexes, les données au niveau de l'entreprise ont été utilisées lorsque le RE ne reflétait pas comme il se doit l'activité de l'entreprise.
2. Durant l'étape de la collecte, de nombreux établissements ont été codés comme étant hors du champ de l'enquête (c.-à-d. leur code SCIAN dans le RE était incorrect). Même si cela a eu pour effet d'améliorer la qualité des résultats de l'enquête, dans le projet des données administratives tous ces établissements ont été inclus comme faisant partie du champ de l'enquête.
3. Variations mineures entre les processus d'enquête et administratif :
 - Données d'enquête ou données fiscales administratives. Par exemple, dans le cadre de l'enquête, on a recueilli des données sur les revenus bruts tandis que dans les dossiers fiscaux il s'agit des revenus d'exploitation.
 - Recensement ou échantillonnage des données fiscales administratives. Alors que l'enquête visait à recueillir des renseignements de tous les établissements dans le champ de l'enquête, le projet des données administratives était fondé sur un échantillonnage de certains enregistrements admissibles.
 - Repondération de l'enquête pour tenir compte de la non-réponse ou imputation par micro-enregistrement pour les données fiscales administratives. Pour l'enquête, on a traité les enregistrements des établissements non répondants en ajustant les poids attribués aux enregistrements des établissements répondants, de telle sorte qu'un enregistrement d'un établissement répondant puisse aussi représenter d'autres unités non répondantes ayant

des caractéristiques similaires (c.-à-d., taille, province, industrie) à celles de l'unité répondante. Toutefois, le projet des données administratives traitait la non-réponse en tâchant d'imputer les renseignements manquants pour l'ensemble de l'enregistrement.

Même si la variance est plus élevée au niveau provincial, particulièrement en Alberta (estimations sectorielles : résultat des données administratives = 1,19 milliard de dollars comparativement au résultat de l'enquête = 902 millions de dollars environ), cela n'est pas étonnant. Plus le niveau d'agrégation est faible, plus il est probable de constater des différences. La PEES et les Services ont examiné toutes les grandes entreprises pour s'assurer que les estimations sont raisonnables. Néanmoins, les différences entre les deux ensembles de données généralement ne sont pas significatives et les erreurs décelées sont explicables. Statistique Canada est d'avis que les résultats des deux projets concordent.

4. Estimation

Étant donné que seulement 1 249 établissements ont rempli le questionnaire, les valeurs individuelles ont été pondérées de manière à représenter le secteur dans son ensemble. Les poids sont égaux au nombre d'établissements dans la population géomatique que représente chaque établissement qui a rempli le questionnaire. La valeur pour chaque établissement répondant a été multipliée par le poids pour cet établissement, puis on a fait la somme des données pondérées de tous les établissements répondants dans un domaine d'estimation donné (p. ex., code 541360 du SCIAN en Ontario) pour obtenir l'estimation.

Pour calculer les poids, on a d'abord regroupé tous les établissements dans la population en groupes homogènes appelés strates. Les établissements ont été groupés (stratifiés) selon la province, quatre types de groupes d'activité (codes 541330, 541360, 541370 et autres du SCIAN) et deux groupes de taille (les établissements les plus grands et ceux les plus petits), lorsque possible. La taille a été mesurée par le revenu brut d'entreprise annuel obtenu du RE (Registre des entreprises). Comme on souhaitait obtenir un minimum d'environ 10 unités par strate (incluant une ou deux unités hors du champ de l'enquête), il a été impossible d'établir une strate pour chaque combinaison de province, d'industrie et de taille. On a regroupé certaines strates pour obtenir le minimum. On a tâché de procéder de façon semblable pour chaque province, même si on a déployé des efforts pour établir une strate distincte pour les codes 541360 et 541370 du SCIAN pour chaque province. Pour certaines provinces, lorsque le revenu brut d'entreprise annuel n'était pas bien corrélé au revenu brut obtenu de l'enquête sur la géomatique, il a été décidé de ne pas utiliser le groupe de taille pour stratifier la province ou bien de placer les établissements les plus grands selon l'enquête dans une strate distincte.

On a calculé les poids dans chaque strate. Les poids sont le produit de deux termes. Le premier terme correspond à l'ajustement pour tenir compte de l'absence de contact et le deuxième, à l'ajustement pour la non-réponse. L'ajustement pour l'absence de contact représente un ajustement apporté pour compenser pour tous les établissements qu'il a été impossible de joindre et dont on ne sait donc pas s'ils entrent dans le champ de l'enquête ou non. L'ajustement pour la non-réponse représente un ajustement apporté pour compenser pour tous les établissements qui ont refusé de remplir le questionnaire mais dont on sait qu'ils sont admissibles (p. ex., d'après le résultat du contact préliminaire). L'ajustement pour l'absence de contact est égal au nombre d'établissements dans la strate divisé par la somme du nombre d'établissements dans la strate qui ont rempli le questionnaire, du nombre d'établissements dans la strate qui ont refusé de remplir le questionnaire (mais dont on sait qu'ils sont admissibles) et du nombre d'établissements hors du champ de l'enquête dans la strate. L'ajustement pour la non-réponse est égal à la somme du nombre d'établissement dans la strate qui ont rempli le questionnaire et du nombre

d'établissements dans la strate qui ont refusé de remplir le questionnaire (mais dont on sait qu'ils sont admissibles) divisée par le nombre d'établissements dans la strate qui ont rempli le questionnaire. Toutes les estimations ont été calculées à l'aide du Système généralisé d'estimation (SGE). Comme mesure de l'erreur, les coefficients de variation pour toutes les estimations ont été calculés à l'aide du SGE.