

REMANIEMENT DE L'ENQUÊTE SUR L'ORIGINE ET LA DESTINATION DES MARCHANDISES TRANSPORTÉES PAR CAMION

François Gagnon et Julie Trépanier, Division des méthodes d'enquêtes auprès des entreprises, Statistique Canada

1. Introduction

L'Enquête sur l'origine et la destination des marchandises transportées par camion (ODMTC), qui est gérée par la Division des transports de Statistique Canada et financée en partie par Transports Canada, sert à mesurer les mouvements de marchandises et la production de l'industrie du camionnage au Canada. Les estimations obtenues incluent le volume de marchandises transportées par type de marchandises, et les recettes générées par les expéditions selon l'origine et la destination. Les estimations sont publiées dans *Le camionnage au Canada*. Les données d'enquête sont utilisées par les gouvernements fédéral et provinciaux, les associations et les entreprises de camionnage, les universités et les établissements de recherche, pour évaluer le taux de croissance de l'industrie et sa contribution à l'économie canadienne, de même que pour mesurer le volume des marchandises transportées dans chaque province et entre les provinces par les compagnies de camionnage. En outre, les statistiques sont utilisées par les comités de planification pour déterminer le volume de circulation sur les routes, et par les compagnies de camionnage qui cherchent des occasions d'expansion. Les estimations de l'Enquête ODMTC sont l'une des données utilisées par le Système de comptabilité nationale de Statistique Canada.

Avant qu'elle ne soit remaniée, cette enquête était fondée sur un plan d'échantillonnage à deux degrés : en premier lieu, un échantillon de compagnies de camionnage était choisi et en deuxième lieu, un échantillon de documents d'expédition était choisi, à l'occasion de visites sur place, dans chaque compagnie échantillonnée.

Depuis quatre ans, l'Enquête ODMTC est remaniée pour en élargir la couverture et pour améliorer la qualité des estimations grâce à un plus vaste échantillon et à une meilleure méthodologie. Il fallait atteindre cet objectif sans augmenter les coûts opérationnels. La nouvelle Enquête ODMTC a été mise en œuvre en 2004. Dans le présent chapitre, nous donnons un aperçu des défis méthodologiques qui se sont posés dans l'élaboration de cette nouvelle enquête qui utilise un plan d'échantillonnage à quatre degrés.

2. CONTEXTE DU REMANIEMENT

L'Enquête ODMTC antérieure a été élaborée au début des années 70. La dernière révision mineure a eu lieu à la fin des années 80. En janvier 2000, Statistique Canada a entériné un projet pluriannuel, devant commencer au cours de l'exercice 2000-2001, en vue de remanier l'Enquête ODMTC. Les buts principaux étaient d'améliorer la qualité des données et de répondre aux nouveaux besoins des utilisateurs des données.

En 2000-2001, les principaux utilisateurs des données de l'Enquête ODMTC ont participé à des séances de consultation et un questionnaire a fait l'objet d'une vaste diffusion. Quatre grands groupes d'utilisateurs ont été consultés : Transports Canada, le Système de comptabilité nationale de Statistique Canada, les gouvernements provinciaux et les associations de camionnage. Voici les points saillants de cette consultation :

- Une majorité d'utilisateurs ont recommandé que toutes les expéditions transportées par camion au Canada soient prises en compte dans la nouvelle enquête. C'est donc que les expéditions faites par les compagnies ci-après devraient être prises en compte :
 - toutes les compagnies canadiennes de camionnage de longue distance (seulement celles qui avaient des recettes supérieures à un million de dollars étaient visées par l'enquête antérieure);
 - toutes les compagnies canadiennes de camionnage local (non visées par l'enquête antérieure);
 - toutes les compagnies canadiennes non associées à l'industrie du camionnage mais qui font du transport par camion dans le cadre de leurs activités, ce qu'on appelle le « camionnage pour compte propre » (non visées par l'enquête antérieure);
 - toutes les compagnies étrangères qui font du transport par camion au Canada (elles n'étaient pas visées par l'enquête antérieure).
- La majorité des utilisateurs ont demandé que les variables recueillies soient les mêmes que celles qui étaient recueillies dans l'enquête antérieure, c.-à-d. le nombre de tonnes transportées, la distance, les marchandises transportées, les recettes provenant du

transport, et l'origine et la destination des expéditions. Certains utilisateurs nous ont demandé de tenir compte de la valeur des marchandises transportées et de recueillir des renseignements sur les expéditions intermodales et le transport de marchandises dangereuses.

- La majorité des utilisateurs ont demandé que les estimations selon l'origine et la destination des expéditions soient produites à l'échelle de la province ou du territoire, plutôt que la région (groupes de provinces) comme dans l'enquête antérieure.
- Les utilisateurs se contenteraient d'une enquête annuelle plutôt que de l'enquête trimestrielle antérieure.
- Pour ce qui est de la qualité des données, toutes les parties intéressées ont convenu que les estimations devraient être plus précises.

Compte tenu des commentaires reçus, il était clair que les objectifs principaux du remaniement viseraient à élargir la couverture de l'enquête et à améliorer la qualité des estimations, surtout à des niveaux détaillés. Il y avait une contrainte importante, soit que ces objectifs devaient être atteints sans générer de coûts supplémentaires de production. Pour cette raison, la nouvelle enquête est menée annuellement et a été rendue plus efficace à chaque étape.

3. MÉTHODOLOGIE DE LA NOUVELLE ENQUÊTE

Dans l'élaboration de la méthodologie de la nouvelle enquête, les besoins des utilisateurs qui se sont exprimés au cours des consultations ont été pris en compte, mais, vu les contraintes budgétaires, certaines demandes n'ont pu être rencontrées. Dans la suite du présent document, nous décrivons comment l'Enquête ODMTC antérieure a été améliorée pour qu'elle soit plus efficace et qu'elle réponde mieux aux besoins des utilisateurs.

3.1 Population observée et base de sondage

L'Enquête ODMTC antérieure et la nouvelle extraient toutes deux leur base de sondage du Registre des entreprises (RE) de Statistique Canada. La structure statistique d'une entreprise du RE comprend, dans l'ordre descendant, quatre niveaux d'entités statistiques : entreprise, compagnie, établissement et emplacement. Dans l'enquête remaniée, toutes les expéditions faites par les compagnies figurant dans la

base de sondage font partie du champ de l'enquête. Les expéditions sur moins de 25 kilomètres ne sont plus exclues comme c'était le cas dans l'enquête antérieure.

Avant le remaniement

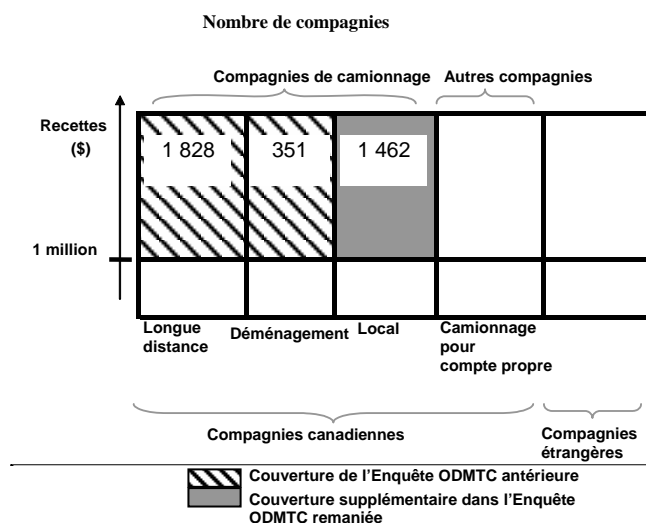
La base comprenait la liste des compagnies de camionnage figurant dans le RE et ayant des recettes annuelles d'un million de dollars ou plus, qui étaient classées aux rubriques Transport par camion de marchandises sur de longues distances (48412, 48423) ou Déménagement de biens usagés de maison et de bureau (48421) dans le Système de classification des industries de l'Amérique du Nord (SCIAN). Les expéditions sur moins de 25 kilomètres faites par ces compagnies étaient réputées être hors du champ de l'enquête. La base était créée le 1^{er} janvier de l'année de référence. Les compagnies qui étaient inscrites au RE plus tard dans l'année de référence n'étaient pas couvertes. Les compagnies qui n'étaient pas classées à la rubrique Transport par camion dans le SCIAN mais qui avaient certains établissements de camionnage n'étaient pas incluses sur la base de sondage.

Après le remaniement

La nouvelle population observée (au premier degré d'échantillonnage) comprend toutes les compagnies figurant dans le RE et qui ont au moins un établissement de camionnage (SCIAN : 484XXX) et des recettes annuelles d'au moins un million de dollars. Le secteur du camionnage local (SCIAN 48411, 48422) a été ajouté à la couverture antérieure de l'enquête. La base de sondage est établie le 1^{er} janvier de l'année de référence pour que les intervieweurs puissent commencer à recueillir les données tôt dans l'année (voir la partie 3.3 concernant le plan d'échantillonnage pour plus de précision). Toutefois, un échantillon de nouvelles compagnies est choisi à la fin de l'année de référence dans la liste des compagnies qui ne faisaient pas partie de la population observée de l'Enquête ODMTC le 1^{er} janvier de l'année de référence, mais qui avaient figuré dans la population observée pendant au moins un jour pendant l'année de référence.

Toutes les expéditions faites par les compagnies faisant partie de la base de sondage font partie de l'enquête remaniée.

Figure 1 : Couverture de l'Enquête ODMTC antérieure et remaniée (source : Registre des entreprises de Statistique Canada, 2004).



La figure 1 illustre le changement dans la couverture de l'enquête. Pour ce qui est du nombre de compagnies, il y a une augmentation de 67 % dans la couverture de l'enquête [100*1462/(1828+351)]. En termes de recettes, ceci correspond à une augmentation de 35 % de la couverture. Bien que la couverture soit beaucoup plus grande dans l'enquête remaniée, elle ne répond pas à tous les besoins des utilisateurs. Les compagnies canadiennes non classées comme compagnies de camionnage mais qui avaient certains établissements de camionnage (on appelle ces compagnies des « compagnies de camionnage pour compte propre »), de même que les compagnies étrangères qui transportent des marchandises au Canada, n'ont pas pu être incluses dans la nouvelle population observée en raison des contraintes budgétaires. Le seuil d'un million de dollars a été maintenu dans l'enquête remaniée afin d'alléger le fardeau de réponse pour les petites compagnies et pour respecter les contraintes budgétaires.

3.2 Méthodes de collecte

Dans l'Enquête ODMTC, qu'il s'agisse de l'enquête antérieure ou de l'enquête remaniée, on utilise trois méthodes de collecte : (1) la collecte électronique des données; (2) les profils; et (3) les visites sur place.

Avant le remaniement

- (1) Collecte électronique des données (CED). Cinq compagnies envoyaient (elles le font encore) leurs données à Statistique Canada sur bande magnétique. Ces bandes magnétiques contiennent

les données des documents d'expédition pour chaque expédition faite pendant la période de référence. Dans l'Enquête ODMTC antérieure, les processus de codage et d'imputation n'étaient pas totalement automatisés et environ 10 % seulement des données sur les expéditions reçues sur bande magnétique étaient utilisées dans l'enquête.

- (2) Profils. Cette méthode de collecte était utilisée pour environ 9 % des compagnies dans l'échantillon. Les intervieweurs de Statistique Canada visitaient chaque compagnie de camionnage choisie dans l'échantillon de premier degré. Toutefois, si la compagnie était spécialisée dans certains types d'expéditions (même origine/destination, même marchandise, même poids, etc.) l'intervieweur ne recueillait des données que sur les « expéditions typiques » et il prenait note du nombre de chaque « expédition typique » qui était faite par la compagnie. À l'étape du traitement, des enregistrements synthétiques étaient produits sur les fichiers à partir de ces données.

- (3) Visites sur place. Cette méthode de collecte était utilisée pour 91 % des compagnies dans l'échantillon de l'Enquête ODMTC antérieure. Les intervieweurs de Statistique Canada se rendaient dans chaque compagnie choisie dans l'échantillon, choisissaient un échantillon de documents d'expédition et transcrivaient les données à partir de ces documents sur des ordinateurs portables. Les données étaient chiffrées et l'intervieweur les envoyait ensuite à Statistique Canada par modem. Cette méthode était coûteuse en raison du temps nécessaire pour chaque visite (habituellement moins d'une journée, mais parfois plus), de même que du coût des déplacements. La présence de l'intervieweur de Statistique Canada dans les bureaux de la compagnie pendant une longue période représentait un fardeau pour la compagnie.

Après le remaniement

Pour l'Enquête ODMTC remaniée, on utilise les trois mêmes méthodes de collecte : (1) la collecte électronique des données; (2) les profils, mais par le biais d'interviews téléphoniques assistées par ordinateur (ITAO); et (3) les visites sur place.

- (1) Collecte électronique des données. Pour la première année de la nouvelle enquête (année de référence 2004), cinq compagnies (les mêmes que dans l'enquête antérieure) enverront leurs données sur bande magnétique. Toutefois, la totalité de leurs données seront traitées (comparativement à

10 % dans l'enquête antérieure), grâce aux nouveaux systèmes de codage et d'imputation maintenant totalement automatisés. Il devrait donc y avoir deux millions d'expéditions de plus dans l'échantillon. Le but est d'augmenter de beaucoup le nombre de compagnies de camionnage qui font une déclaration par voie électronique de sorte que cette méthode devienne à long terme la méthode principale de collecte pour l'enquête remaniée. Cette méthode offre la possibilité de réduire les coûts, et donc d'augmenter la taille globale de l'échantillon des expéditions aux degrés d'échantillonnage suivants. Elle offre aussi la possibilité d'améliorer l'actualité des données (étant donné l'élimination du temps nécessaire pour organiser les interviews, les déplacements et les visites), d'augmenter les taux de réponse (grâce à la réduction du fardeau de réponse) et d'améliorer la qualité des données (p. ex. par l'élimination des erreurs de saisie).

- (2) Les profils avec l'ITAO. Les profils sont utilisés davantage dans l'Enquête ODMTC remaniée. Les données de toutes les compagnies se spécialisant dans certains types d'expédition, c.-à-d. celles qui ont déclaré moins de 50 combinaisons origine/destination/marchandises dans l'année précédente, seront recueillies par le biais d'un profil par ITAO. Au lieu de se rendre dans les bureaux de la compagnie pour recueillir les données, l'intervieweur recueillera, m par une ITAO, les données au sujet de chaque « expédition typique » et il prendra note du nombre de chaque « expédition typique » qui a été faite par la compagnie pendant la période de référence. À l'étape du traitement, des enregistrements synthétiques seront produits sur les fichiers, à partir de ces données. La nouvelle application utilisée pour ITAO réduira beaucoup le coût de la collecte de données pour les compagnies dont les activités peuvent être décrites par des profils. Comme un profil sera établi pour beaucoup plus de compagnies dans l'enquête remaniée, nous comptons ainsi ajouter 2,3 millions d'expéditions dans l'échantillon.
- (3) Visites sur place. Cette méthode de collecte est utilisée lorsque aucune autre méthode (les profils ou la CED) ne peut être utilisée pour une compagnie donnée. Les visites sur place, bien qu'il y en ait moins comparativement à l'Enquête ODMTC antérieure, restent le mode le plus fréquent de collecte dans l'enquête remaniée. Comme dans le cas de l'enquête antérieure, les intervieweurs de Statistique Canada se rendent dans chaque compagnie retenue dans

l'échantillon, choisissent un échantillon systématique de documents d'expédition, puis choisissent un échantillon dans chacun de ces documents d'expédition et transcrivent finalement les données sur un ordinateur portable à partir des documents. Voici les variables recueillies pour chaque expédition : l'origine et la destination de l'expédition, la description de la marchandise transportée, le poids et le revenu produit par l'expédition. Lorsque l'interviewer a transcrit les données, elles sont chiffrées et transmises à Statistique Canada par modem.

3.3 Plan d'échantillonnage

Avant le remaniement

Pour l'Enquête ODMTC antérieure, qui était une enquête trimestrielle, on employait un plan d'échantillonnage à deux degrés : au premier degré, un échantillon de compagnies de camionnage était choisi et, au deuxième degré, un échantillon systématique de documents d'expédition était retenu. Au premier degré, les compagnies faisant partie de la base de sondage étaient stratifiées selon le type d'activité, la région de domicile et la classe de recettes annuelles (classe I – 12 millions de dollars et plus; classe II – de 1 à 12 millions de dollars). Les compagnies de la classe I faisaient partie de strates à tirage complet (recensement) alors que les petites compagnies étaient divisées en quatre parties égales dans chaque strate. Chaque partie représentait l'échantillon pour l'un des quatre trimestres. Par conséquent, chaque échantillon trimestriel était considéré comme un échantillon aléatoire simple stratifié sans remise stratifié.

Au deuxième degré, en vue de choisir un échantillon systématique d'expéditions, l'intervieweur demandait à la compagnie de camionnage une bonne estimation du nombre total d'expéditions faites pendant la période de référence. Avec un logiciel installé sur son ordinateur portable, l'intervieweur prenait ensuite ce nombre estimatif d'expéditions pour établir le pas de sondage K qui permettrait d'obtenir la taille de l'échantillon souhaitée au deuxième degré pour cette compagnie. Au moyen de ce pas de sondage K et d'un début aléatoire P , il tirait un échantillon systématique d'expéditions de taille m en choisissant les expéditions $P, P+K, P+2K, \dots, P+(m-1)K$. Il est intéressant de signaler que, lorsqu'une compagnie avait des documents d'expédition contenant des données au sujet de plus d'une expédition, les intervieweurs devaient extraire chaque document et compter le nombre d'expéditions puisque le pas de sondage était appliqué au niveau de l'expédition, et non au niveau du document d'expédition. Ce travail était exigeant pour les intervieweurs.

Après le remaniement

L'enquête remaniée est une enquête annuelle fondée sur un plan d'échantillonnage à quatre degrés où un échantillon aléatoire simple stratifié de compagnies de camionnage est choisi sans remise au premier degré d'échantillonnage. Les compagnies qui font partie de la base de sondage et qui peuvent fournir leurs données par voie électronique font partie de la strate à tirage obligatoire « CED ». De même, les compagnies qui ont déclaré moins de 50 combinaisons origine/destination/marchandise dans l'année antérieure (et dont les données seront recueillies par conséquent par la voie d'un profil par ITAO) font partie de la strate à tirage obligatoire des « Profils ». Toutes les compagnies faisant partie de la strate à tirage obligatoire font partie de l'échantillon. Les compagnies qui ne font pas partie de l'une ou l'autre des strates à tirage obligatoire sont ensuite stratifiées selon trois secteurs du SCIAN (transport sur de longues distances, transport local et déménagement), 13 types d'activités fondés sur les données historiques (p. ex., le transport de produits forestiers, de marchandises sèches en vrac, etc.), 13 provinces ou territoires de domicile et 2 classes de recettes annuelles (classe I – grandes compagnies; classe II – petites compagnies). Les grandes compagnies font partie de strates à tirage complet (recensement) alors que les plus petites compagnies font partie de strates à tirage partiel. On utilise la méthode de Lavallée-Hidiroglou (1988) pour déterminer de façon optimale le seuil entre les classes de recettes à tirage complet et à tirage partiel.

Au deuxième degré d'échantillonnage, pour chaque compagnie choisie dans l'échantillon du premier degré, une période (p. ex. janvier à juin) est choisie au hasard dans l'année de référence. Il est ainsi possible de répartir le travail de collecte tout au long de l'année. Pour l'année de référence 2004, trois périodes sont possibles. Les compagnies devant être sondées selon un profil ITAO, de même que celles qui transmettent leurs données par voie électronique, se voient attribuer une période de 12 mois (janvier à décembre). Les plus grandes compagnies des strates à tirage complet se voient attribuer une période de collecte de 12 mois également, tandis que les autres compagnies des strates à tirage complet ou à tirage partiel se voient attribuer de façon aléatoire une période de six mois, soit de janvier à juin, soit de juillet à décembre.

Dans l'enquête remaniée, le budget fixe pour la collecte est une contrainte importante. Dans l'enquête antérieure, certains problèmes se posaient lorsque le budget total de la collecte était épuisé avant que les données ne soient recueillies pour toutes les

compagnies dans l'échantillon. Sachant que les compagnies éloignées sont souvent les dernières à être visitées, il y avait une possibilité de biais étant donné qu'il était impossible de rejoindre toutes les compagnies. Pour éviter une telle situation dans l'enquête remaniée, un algorithme a été élaboré pour faire l'estimation des coûts de collecte au moment du choix de l'échantillon afin de s'assurer que le budget de collecte soit suffisant afin de recueillir les données auprès de toutes les compagnies dans l'échantillon. L'algorithme est fondé sur des données historiques sur les coûts de la collecte dans le cas des visites sur place (p. ex. les frais de déplacement aux lieux des compagnies et les coûts associés à la transcription des données sur les expéditions) afin d'assurer que l'échantillon d'expéditions espéré puisse être recueilli totalement dans les limites du budget établi. Toutefois, l'Enquête ODMTC remaniée reste vulnérable car les fluctuations des coûts de la collecte des données ne peuvent être entièrement prédites (p. ex. les frais de déplacement, le transport par avion, etc.).

Au troisième degré d'échantillonnage, toutes les expéditions des compagnies faisant partie de la strate à tirage obligatoire de la CED ou des profils sont choisies. Pour le reste des compagnies échantillonnées, un échantillon systématique de documents d'expédition est tiré à l'occasion de visites sur place auprès de chaque compagnie de camionnage échantillonnée et ce, pour la période choisie au deuxième degré d'échantillonnage. À cette fin, l'intervieweur doit obtenir une bonne estimation R_{it} du nombre total d'expéditions transportées par la compagnie i pendant la période t . Si l'on suppose qu'il y a une seule expédition par document d'expédition, le logiciel sur mesure installé sur un ordinateur portable utilise ensuite ce total estimatif d'expéditions pour dériver le pas de sondage approprié A qui donne la taille souhaitée de l'échantillon au troisième degré pour cette compagnie. Avec le pas de sondage A et un début aléatoire P , on tire un échantillon systématique (de taille approximative r_{it}) en choisissant les documents d'expédition $P, P+A, P+2A, \dots, P+(r_{it}-1)A$. Si $R_{it} < 100$, alors le pas de sondage A est égal à « 1 ». Si $R_{it} \geq 100$, le pas de sondage A est obtenu par $ARRONDI(R_{it} / r_{it})$ où $r_{it} = ARRONDI [11*(R_{it} + 1100)^{0,31}]$. Cette fonction continue remplace la fonction non continue qui était utilisée dans l'Enquête ODMTC antérieure pour déterminer le pas de sondage et la taille de l'échantillon. Cette nouvelle fonction continue a été choisie car elle donne à peu près la même taille d'échantillon que l'ancienne fonction pour un R_{it} donné, mais elle a l'avantage d'être une fonction strictement croissante.

Lorsqu'il y a plus d'une expédition dans le document

d'expédition choisi au troisième degré d'échantillonnage, on procède à un quatrième degré d'échantillonnage, dans lequel un échantillon systématique d'expéditions est choisi dans le document d'expédition. Le pas de sondage B est donné par $B = \text{ARRONDI}(M_{ij} / m_{ij})$ où : M_{ij} est le nombre d'expéditions dans le document d'expédition j d'une compagnie i pour la période t (à indiquer par l'intervieweur dans l'application sur ordinateur portable); $m_{ij} = 1$ si $M_{ij} = 1$; $m_{ij} = 2$ si $2 \leq M_{ij} \leq 10$; $m_{ij} = 3$ if $11 \leq M_{ij} \leq 30$; et $m_{ij} = 4$ si $M_{ij} \geq 31$. Les documents d'expédition contenant des expéditions multiples deviennent de plus en plus la norme (p. ex. une facture mensuelle à un client qui comprend des données sur toutes les expéditions faites pendant ce mois pour ce client). Le quatrième degré d'échantillonnage a été ajouté pour réduire la charge de travail des intervieweurs lorsqu'il y a beaucoup d'expéditions par document. Avec l'ajout de ce quatrième degré, ils n'ont qu'à compter le nombre d'expéditions dans les documents choisis au troisième degré d'échantillonnage (dans l'enquête antérieure, ils devaient compter les expéditions dans tous les documents puisque le pas de sondage était appliqué au niveau de l'expédition seulement).

3.4 Traitement des données

Dans l'Enquête ODMTC remaniée, l'automatisation des processus de codage et d'imputation était une amélioration essentielle pour qu'on puisse faire le traitement de toutes les expéditions recueillies par la CED. Voici les variables recueillies pour chaque expédition : l'origine et la destination, la description de la marchandise transportée, le poids et le revenu produit par l'expédition. Le codage de la description des marchandises, de l'origine et de la destination, est fait par les intervieweurs pour les compagnies dont les données sont recueillies à l'occasion de visites sur place ou par ITAO. Pour coder la description des marchandises, l'intervieweur fait un choix dans une liste de descriptions possibles figurant dans une bibliothèque des marchandises dans le logiciel de son ordinateur portable. Le codage pour l'origine et la destination se fait de la même façon : l'intervieweur choisit dans une liste de villes, de municipalités et de villages.

Le processus de codage est plus complexe pour les compagnies qui transmettent leurs données par voie électronique. Les données fournies par chaque compagnie doivent être automatiquement converties en un format et selon une nomenclature normalisée. Il fallait à cette fin établir une cartographie ou des tableaux de liens pour chaque compagnie. Dans l'enquête antérieure, le processus de codage pour les

données électroniques n'était pas automatisé, de sorte qu'on devait se limiter à un échantillon d'environ 10 % des données électroniques reçues. Dans l'enquête remaniée, toutes les descriptions de marchandises obtenues en version électronique sont automatiquement codées au moyen du système généralisé de Statistique Canada « Codage automatisé par reconnaissance de texte ». De plus, un processus de codage automatique est utilisé pour le codage des variables de l'origine et de la destination. Ce logiciel permet également de dériver la distance entre l'origine et la destination de chaque expédition. En outre, dans l'enquête remaniée, le processus d'imputation a été amélioré et entièrement automatisé.

3.5 Estimation

Avant le remaniement

Pour la période de référence, les estimations des totaux pour les domaines d'intérêt étaient produites au moyen d'un estimateur Horvitz-Thompson pour un plan d'échantillonnage à deux degrés :

$$\hat{Y}(d) = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}(d)$$

où :

- H = nombre de strates au premier degré
- n_h = nombre de compagnies dans l'échantillon de premier degré dans la strate h
- m_{hi} = nombre d'expéditions de la compagnie i dans la strate h , choisies dans l'échantillon de 2^e degré
- w_{hi} = poids de sondage de 1^{er} degré de la compagnie i dans la strate h
- w_{hij} = poids de sondage de 2^e degré de l'expédition j de la compagnie i dans la strate h
- y_{hij} = valeur de la variable d'intérêt pour l'expédition j de la compagnie i dans la strate h
- d = domaine d'intérêt
- $y_{hij}(d) = y_{hij}$ if $hij \in d$, $y_{hij}(d) = 0$ if $hij \notin d$.

On peut décrire les principaux domaines d'estimation par un tableau tridimensionnel. Dans l'enquête antérieure, les deux premières dimensions étaient la région d'origine (au nombre de 5) et la région de destination (au nombre de 5) de l'expédition, respectivement. La troisième dimension était le type de marchandises (au nombre de 60). Étant donné le plan d'échantillonnage choisi, la taille de l'échantillon dans chaque cellule du tableau tridimensionnel était aléatoire, c'est-à-dire non contrôlée. Des estimations annuelles et trimestrielles étaient produites.

La variance avait deux composantes : dans la

composante du premier degré, on tenait compte de la variabilité entre les compagnies, tandis que, dans la composante du deuxième degré, on tenait compte de la variabilité entre les expéditions au sein d'une compagnie. Pour faire l'estimation de ces deux composantes, on utilisait un estimateur de variance Horvitz-Thompson pour un plan d'échantillonnage à deux degrés. La variance de l'estimateur servant à produire les estimations pour un trimestre q avait la forme suivante :

$$\text{Var}_q = \text{Var}_q(\text{intercompagnies}) + \text{Var}_q(\text{interexpéditions}).$$

La variance de l'estimateur servant à produire les estimations annuelles avait la forme suivante :

$$\text{Var} = \sum_{q=1 \text{ to } 4} \text{Var}_q + 2 \sum_{q=1 \text{ to } 4} \sum_{q < r} \text{COV}_{q,r}.$$

Le terme de covariance était essentiel puisque les échantillons trimestriels n'étaient pas indépendants dans une année de référence donnée.

Après le remaniement

Pour la période de référence, les estimations des totaux pour les domaines d'intérêt seront produites au moyen d'un estimateur Horvitz-Thompson pour un plan d'échantillonnage à quatre degrés :

$$\hat{Y}(d) = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{1hi} w_{2hit} \sum_{j=1}^{r_{hit}} w_{3hitj} \sum_{k=1}^{m_{hitj}} w_{4hitjk} y_{hitjk}(d)$$

où :

- H = nombre de strates au 1^{er} degré
- n_h = nombre de compagnies dans l'échantillon de premier degré dans la strate h
- r_{hit} = nombre de documents d'expédition échantillonnés pour la période d'enquête t , la compagnie i dans la strate h
- m_{hitj} = nombre d'expéditions échantillonnées dans le document d'expédition j , pour la période d'enquête t , la compagnie i dans la strate h
- w_{1hi} = poids de sondage de 1^{er} degré de la compagnie i dans la strate h
- w_{2hit} = poids de sondage de 2^e degré de la période d'enquête t pour la compagnie i dans la strate h
- w_{3hitj} = poids de sondage de 3^e degré du document d'expédition j pour la période t de la compagnie i dans la strate h
- w_{4hitjk} = poids de sondage de 4^e degré de l'expédition k dans le document d'expédition j pour la période t de la

- y_{hitjk} = compagnie i dans la strate h
- y_{hitjk} = valeur de la variable d'intérêt pour l'expédition k dans le document d'expédition j pour la période d'enquête t de la compagnie i dans la strate h
- d = domaine d'intérêt $y_{hitjk}(d) = y_{hitjk}$ if $hitjk \in d$, $y_{hitjk}(d) = 0$ if $hitjk \notin d$.

Encore une fois, on peut décrire les principaux domaines d'estimation par un tableau tridimensionnel. Les deux premières dimensions seront la province ou le territoire d'origine (au nombre de 13) et la province ou le territoire de destination (au nombre de 13) de l'expédition, respectivement. La troisième dimension est le type de marchandises (au nombre de 60). En raison du plan d'échantillonnage retenu, la taille de l'échantillon dans chaque cellule du tableau tridimensionnel est encore aléatoire. La taille globale de l'échantillon d'expéditions devrait augmenter de 500 000 à environ 4,9 millions dans l'enquête remaniée, ce qui améliorera la qualité des estimations. Cependant, il pourrait tout de même y avoir des cas où certains domaines comportent trop peu d'expéditions échantillonnées, ce qui affecterait la qualité des estimations.

Dans le cadre d'une entente spéciale de partage des données, Statistique Canada fournit à Transports Canada, qui est l'un des partenaires financiers de l'enquête, un fichier de microdonnées. C'est pourquoi il était important de choisir une méthode d'estimation de la variance qui procurerait à Transports Canada un moyen commode de produire des estimations de la variance pour les estimations par domaine qu'ils produiront à partir du fichier de microdonnées. Nous avons d'abord envisagé la méthode dite « bootstrap » pour le calcul des estimations de variance dans l'enquête remaniée. Cette méthode consiste à tirer des sous-échantillons de l'échantillon initial. À l'intérieur de chaque strate, un échantillon aléatoire simple de $n-1$ grappes (unités de premier degré = compagnies) est choisi avec remise, à partir des n grappes de la strate. On obtient ainsi B nouveaux échantillons (ou répétitions). La même estimation est ensuite calculée pour chacun des B échantillons, ce qui donne B estimations différentes $\hat{\theta}^*_{(1)}, \hat{\theta}^*_{(2)}, \dots, \hat{\theta}^*_{(B)}$. Pour obtenir chacune des B estimations, il faut avoir un poids spécifique pour chaque échantillon. Dans chaque échantillon aléatoire simple, le poids est ensuite recalculé pour chaque enregistrement dans la strate. Ces B poids, dits poids « bootstrap », doivent être produits et ajoutés au fichier de microdonnées. Ensuite, l'estimateur de la variance de $\hat{\theta}$ est donné par :

$$v_B(\hat{\theta}) = \frac{1}{B} \sum_b (\hat{\theta}_{(b)}^* - \hat{\theta}_{(\bullet)}^*)^2 \quad \text{où}$$

$$\hat{\theta}_{(\bullet)}^* = \frac{\sum_n \hat{\theta}_{(b)}^*}{B}.$$

Des essais préliminaires ont montré que, dans le contexte de l'Enquête ODMTC, l'estimateur de la variance « bootstrap » surestime la variance réelle de l'échantillon, ce qui est attribuable aux fractions de sondage élevées au premier degré dans certaines des strates. En raison de ce problème, il a été décidé d'utiliser un estimateur de variance Horvitz-Thompson pour un plan d'échantillonnage à quatre degrés et d'élaborer un module convivial qui permettrait à Transports Canada d'estimer la variance de chacune des estimations que leurs analystes produisent à partir du fichier de microdonnées qu'ils reçoivent.

4. CONCLUSION

La plupart des objectifs du remaniement de l'Enquête ODMTC ont été atteints. La couverture de l'enquête a été augmentée de beaucoup pour ce qui est à la fois des recettes et du nombre de compagnies. Une stratification et une méthode de répartition de l'échantillon plus efficaces ont été mises en place au premier degré d'échantillonnage. Les méthodes de collecte ont été améliorées : plus de données sont maintenant recueillies avec des méthodes moins coûteuses comme l'ITAO et la CED. Ces changements ont permis d'augmenter la taille prévue de l'échantillon d'expédition de 0,5 M à 4,9 M, et de garder en même temps les coûts de la collecte au même niveau que l'Enquête ODMTC antérieure. Le codage et l'imputation de la description des marchandises, et de l'origine et de la destination ont été entièrement automatisés, ce qui a permis d'augmenter beaucoup la quantité de données électroniques pouvant être traitées. Un nouveau système convivial d'estimation de la variance sera élaboré afin que les analystes de Transports Canada puissent produire leurs propres estimations de la variance à partir du fichier de microdonnées qu'ils reçoivent. Dans le futur, l'objectif est de continuer à augmenter le nombre de compagnies qui déclarent leurs données par voie électronique en vue d'augmenter la taille de l'échantillon des expéditions et de réduire les coûts de la collecte.

REMERCIEMENTS

Les auteurs aimeraient remercier Sébastien Landry, Jeannine Claveau, Jean-François Bastien et Windie Gagné de leur contribution précieuse à ce

projet. Ils aimeraient également remercier Ed Hamilton et Jack Lothian de leurs commentaires, lesquels ont servi à améliorer la qualité du document.

RÉFÉRENCES

- Gagnon, F., Rathwell, S. et Gauthier, Y. (2000). Electronic Data Reporting in the Context of the Redesign of the Canadian For-Hire Trucking Origin/Destination Survey. *Proceedings of the Second International Conference on Establishment Surveys*, American Statistical Association, 171-179.
- Lavallée, P., et Hidioglou, M.A. (1988), Sur la stratification de populations asymétriques. *Techniques d'enquêtes*, Volume 14, Numéro 1, 35-45.
- Rao, J.N.K., Wu, C.F.J. et Yue, K. (1992). Quelques travaux récents sur les méthodes de ré échantillonnage applicables aux enquêtes complexes. *Techniques d'enquêtes*, Volume 18, Numéro 2, 225-234.
- Équipe de développement d'une enquête sur le trafic du camionnage (2001), Document sur les besoins et les exigences en matière de données: résultats des consultations auprès des principaux utilisateurs. *Rapport interne*, Statistique Canada.