

LE REMANIEMENT DU PLAN D'ÉCHANTILLONNAGE DE L'ENQUÊTE ANNUELLE SUR LES MANUFACTURES

José Gaudet et Isabelle Marchand¹

RÉSUMÉ

L'Enquête annuelle sur les manufactures (EAM) est une enquête de Statistique Canada couvrant tous les établissements des secteurs manufacturier et forestier. L'information recueillie inclut les revenus, les dépenses, l'emploi, les coûts des matériaux utilisés, les livraisons et les stocks. Les matériaux et les livraisons sont rapportés par type de marchandises. Les principales statistiques financières sont estimées à partir d'un fichier de micro-données provenant d'établissements échantillonnés, de données fiscales et de données modélisées.

Dans un contexte de rationalisation ainsi que dans un but d'uniformisation avec l'Enquête unifiée sur les entreprises de Statistique Canada avec laquelle l'EAM est progressivement intégrée, un remaniement du plan de sondage de l'enquête a été entrepris pour l'année de référence 2004. Le nouveau plan de sondage comporte des modifications à plusieurs niveaux. Nous décrivons ici les principales modifications apportées au plan sondage, en particulier au plan d'échantillonnage, de même que les justifications derrière celles-ci.

1. INTRODUCTION

L'Enquête annuelle sur les manufactures (EAM) est une enquête de Statistique Canada menée par la Division de la fabrication, de la construction et de l'énergie (DFCE). L'EAM couvre tous les établissements du secteur manufacturier et de l'exploitation forestière identifiés à l'aide du Système de classification des industries de l'Amérique du Nord (SCIAN). Ces secteurs sont regroupés en 23 industries (codes SCIAN à 3 chiffres ou SCIAN3). L'information recueillie porte sur des variables financières (revenus, dépenses, salaires et traitements) et des variables de production (détails sur les marchandises fabriquées et utilisées comme matière première). Les estimations produites par l'EAM sont une composante importante du Produit Intérieur Brut (PIB) canadien.

Au cours des derniers mois, suite à un projet de rationalisation, l'EAM a subi des changements considérables affectant plus particulièrement l'année de référence 2004. La section 2 décrit les principales décisions liées à l'initiative de rationalisation qui affectent directement le plan de sondage de l'année de référence 2004. La section 3 illustre le plan de sondage avant et après le remaniement, la section 4 présente quelques résultats pour illustrer l'impact de ces changements.

2. ORIGINE DU REMANIEMENT

Au cours des dernières années, Statistique Canada a mis sur pied un projet de rationalisation des enquêtes économiques dont l'objectif est de diminuer les coûts et les ressources à l'échelle du bureau tout en maintenant un bon niveau de qualité. Différents comités ont ainsi fait des recommandations affectant toutes les étapes de l'EAM. Voici, en résumé, une brève description des recommandations principales qui ont affecté le plan de sondage de l'enquête pour l'année de référence 2004.

2.1. Utilisation accrue de données fiscales

Afin de diminuer les coûts et le fardeau de réponse, l'EAM utilise les données fiscales de différentes façons depuis un certain nombre d'années, que ce soit pour l'imputation ou la modélisation des données d'enquête. Le remplacement direct des données d'enquête a été utilisé pour la première fois pour l'année de référence 2002 et ce, pour une petite portion de l'échantillon. En 2004, le questionnaire remanié de l'EAM est le premier à suivre les principes du plan comptable de Statistique Canada, i.e. que le questionnaire est maintenant structuré comme l'état des résultats d'une entreprise et utilise la terminologie de la comptabilité d'entreprise pour être mieux

¹ José Gaudet, Statistique Canada, 120 Parkdale, Ottawa (Ontario), Canada, K1A 0T6, jose.gaudet@statcan.ca, Isabelle Marchand, Statistique Canada, 120 Parkdale, Ottawa (Ontario), Canada, K1A 0T6, isabelle.marchand@statcan.ca.

compris par les comptables, les répondants. Environ 10% des variables financières peuvent être directement extraites des fichiers de données fiscales. Une des contraintes qui affecte directement le plan de d'échantillonnage est que les données fiscales ne sont présentement pas allouées à tous les niveaux d'une entreprise complexe. Ceci restreint leur utilisation aux entreprises ayant une structure opérationnelle très simple.

2.2. Harmonisation et intégration avec l'Enquête unifiée auprès des entreprises (EUE)

L'EUE existe depuis 1997 et a été développée pour produire des estimations annuelles pour différents secteurs économiques. Le Système de Comptabilité Nationale (SCN) utilise ces estimations pour mesurer la production et la consommation totale de biens et services dans les provinces et territoires (Smith, 1998). L'EUE couvre pratiquement tous les secteurs économiques, sauf le secteur manufacturier. L'EAM utilise la même base de sondage que l'EUE, le Registre des entreprises (RE) de Statistique Canada, mais diffère en terme de concepts (ex : différentes unités d'échantillonnage), de méthodologie et de système opérationnel et informatique. Le travail sur l'intégration de l'EAM à l'EUE a commencé il y a deux ans et se fait de façon progressive. Lors de la révision du plan de sondage pour l'année de référence 2004, une attention plus particulière a été portée à celui de l'EUE pour maximiser l'harmonisation des concepts, de la méthodologie et des systèmes autant que possible.

2.3. Révision des besoins des utilisateurs principaux

Une des ententes avec certains utilisateurs, soit l'Institut de la statistique du Québec (ISQ) et Ressources Naturelles Canada (RNCAN) était de s'assurer que l'échantillon couvre un pourcentage suffisant des livraisons totales de la population manufacturière pour chaque industrie dans chacune des régions économiques du Québec et dans chacune des provinces et territoires. Pour l'année de référence 2004, suite à des consultations entre Statistique Canada et ces utilisateurs, cette exigence n'a pas été retenue.

Le SNC est l'utilisateur principal des données de l'EAM. En plus d'être une composante importante du PIB, les variables de production de l'EAM sont les données principales des tableaux d'entrées-sorties produits par le SCN. Avant 2004, le système de classification des marchandises comportait environ 9 000 codes de marchandises, alors que le SCN utilisait un agrégat d'environ 1 000 codes. Pour 2004, l'EAM utilise un nouveau système qui est maintenant basé sur un modèle américain adapté pour les besoins canadiens et est composé d'environ 1 400 codes de marchandises. Aussi, afin de réduire les coûts de l'enquête, un fichier de micro-données ne sera requis que pour les variables financières. Pour le SCN, l'enquête devra produire des estimations des variables de production au niveau des secteurs industriels SCIAN6 pour chaque région géographique.

3. LE PLAN DE SONDRAGE AVANT ET APRÈS LE REMANIEMENT

En passant en revue les différentes composantes du plan de sondage, nous allons maintenant décrire les principales différences entre celui utilisé avant le remaniement et le nouveau plan qui est utilisé pour 2004, en mettant l'emphase plus particulièrement sur le plan d'échantillonnage.

3.1. Plan de sondage

La seule modification apportée à la population cible de l'EAM vise un sous-ensemble de petites unités, les établissements incorporés sans employé, qui étaient auparavant considérés comme ne faisant pas partie de la population cible de l'EAM et qui sont maintenant inclus. Cette décision a pour but d'harmoniser la définition de population cible de l'EAM avec celle de l'EUE.

L'EAM tire sa base de sondage du RE de Statistique Canada. Le RE contient toutes les entreprises connues opérant au Canada. Sur le RE, la structure de chaque entreprise est hiérarchisée en fonction des besoins statistiques et comporte quatre niveaux qui sont dans l'ordre l'entreprise, la compagnie, l'établissement et l'emplacement. Une entreprise se compose d'une ou plusieurs compagnies. Une compagnie se compose d'un ou plusieurs établissements et ainsi de suite. Pour plusieurs entreprises canadiennes, les 4 entités coïncident, ce qu'on appelle une structure simple. Pour l'EAM, l'unité d'échantillonnage est l'établissement.

L'enquête utilise un plan stratifié aléatoire simple sans remise où les strates sont un croisement des industries, des régions géographiques et d'une variable de taille. Dans le passé, l'EAM comportait deux types de questionnaires et la sélection de l'échantillon se faisait en deux phases. Tous les établissements devant recevoir un questionnaire étaient identifiés lors de la première phase. Ceux devant recevoir un questionnaire court étaient

identifiés lors de la deuxième phase. En 2003, avec l'introduction d'un projet de remplacement des données par les données fiscales, lors de la deuxième phase, les établissements visés pour faire du remplacement de données d'enquête par des données fiscales ont été identifiés. En 2004, l'approche pour le remplacement par les données fiscales a été étendue de façon à couvrir tous les établissements éligibles à l'échantillonnage, excepté les unités complexes (voir section 3.2). La deuxième phase d'échantillonnage devenait inutile et la sélection a donc pu se faire en une seule phase. Aussi, une des contraintes budgétaires était que la taille finale de l'échantillon devait demeurer inchangée en 2004, c'est-à-dire autour de 15 500 établissements.

L'estimation se fait au niveau de l'emplacement. Avant 2004, les données étaient estimées à partir d'un fichier de micro-données au niveau de la population combinant des données provenant d'établissements échantillonnés, de données fiscales et de données modélisées. Pour 2004, les variables financières continueront d'être estimées à l'aide d'un fichier de micro-données au niveau de la population. Par contre, les variables de production seront maintenant estimées à l'aide d'un système de pondération. Plus précisément à l'aide d'un estimateur par calage, où la somme des marchandises fabriquées (total des livraisons) et utilisées comme matière première (total des entrées) sont deux variables financières et dont l'estimation devra coïncider pour chaque croisement de secteurs industriels SCIAN6 par région géographique aux données du fichier de micro-données des variables financières.

3.2. Stratification, répartition et sélection de l'échantillon

Comme la plupart des enquêtes auprès des entreprises, l'EAM utilise une strate à tirage complet constituée d'unités qui sont toutes enquêtées. Elle comprend entre autres toutes les unités de type complexe sur le RE, pour diverses raisons. Par exemple, comme seules les structures simples ont des données fiscales disponibles au niveau de l'emplacement, les structures complexes doivent être enquêtées puisque l'EAM doit produire un fichier de micro-données au niveau de l'emplacement pour les variables financières qui couvrent toute la population.

Avec la disparition des ententes qui exigeaient une certaine couverture accrue des livraisons manufacturières (voir section 2.3), la taille de la strate à tirage complet a été considérablement réduite. Pour 2004, les cinq plus grands établissements par industrie (SCIAN6) en termes de livraison sont maintenant inclus dans la strate à tirage obligatoire et ce afin d'augmenter les chances de représenter l'industrie adéquatement en termes de répartition des marchandises. La stratification étant au niveau industriel SCIAN3 et l'EAM devant produire des estimations des variables de production au niveau industriel SCIAN6 pour le SCN, cette spécification permet d'assurer un nombre minimal d'établissements dans l'échantillon pour produire les estimations au niveau requis.

Dans le but de réduire les coûts ainsi que le fardeau de réponse, l'EAM utilise une strate à tirage nul. La définition de cette strate a subi de légères modifications au cours du remaniement. La première étape dans la détermination des établissements admissibles à l'échantillonnage est l'application de bornes d'exclusion (Maranda et Royce, 1998). Ces bornes visent à exclure du processus d'échantillonnage l'ensemble des établissements représentant moins d'un certain pourcentage du revenu par secteur industriel et région géographique. Le secteur industriel est défini au niveau du SCIAN4 pour le secteur manufacturier et du SCIAN6 pour le secteur forestier. Auparavant, le pourcentage utilisé était de 10% avec certains secteurs industriels à 5%. Les seuils sont maintenant tous à 10%, comme l'EUE.

Tous les établissements appartenant à la population cible ne faisant pas partie de la strate à tirage nul constituent la population éligible à l'échantillonnage. La méthode utilisée pour la stratification et la répartition de l'échantillon de l'EAM est celle de Lavallée-Hidiroglou (Lavallée et Hidiroglou, 1988). Cette méthode sépare d'abord la population éligible à l'échantillonnage de chacun des croisements de secteur industriel SCIAN3 et de régions géographiques en strates d'unités homogènes. En plus de la strate à tirage obligatoire définie plus haut, l'algorithme définit une strate à tirage complet (recensée) et deux strates à tirage partiel (échantillonnées). Par la suite, l'algorithme détermine la fraction de sondage nécessaire dans chaque strate pour atteindre la précision voulue tout en minimisant la taille totale de l'échantillon. Cette précision est déterminée par les coefficients de variation visés dans chaque croisement de secteur industriel SCIAN3 et de région géographique. La répartition dans chaque strate se fait de façon proportionnelle à la racine carrée de la variable de taille. Cette méthode continue d'être utilisée pour 2004.

Une des modifications apportées au cours du remaniement concerne la variable de taille utilisée pour la stratification. Avant 2004, le revenu de l'établissement, qui est disponible sur le RE pour toute la population, était utilisé. Mais pour l'EAM, le total des livraisons est une variable jugée plus pertinente au contexte de

l'enquête, c'est d'ailleurs pour cette raison que les ententes avec les utilisateurs utilisaient cette variable pour la détermination de la couverture de l'échantillon. Le total des livraisons est une variable financière disponible au niveau de la population pour l'année de référence précédente et modélisée pour les naissances de l'année en cours. Suivant une approche conservatrice, il a été décidé d'utiliser le maximum de ces deux variables, le revenu et le total des livraisons, comme variable de stratification.

La méthode utilisée pour déterminer les coefficients de variation visés a été modifiée pour 2004. Anciennement, les coefficients étaient fixes pour chaque croisement de secteur industriel SCIAN3 et de région géographique, spécifiés selon une entente avec les utilisateurs. La nouvelle méthode utilise un coefficient de variation global qui est alloué aux régions géographiques et aux secteurs industriels SCIAN3 de façon à ce que les coefficients soient inversement proportionnels à la contribution de l'industrie au total des industries à l'intérieur d'une province ou d'un territoire. Cette méthode vise aussi à ce que les coefficients soient semblables entre les régions géographiques. Cette méthode est celle utilisée par l'EUE et suite à des consultations avec le SCN, elle a été adoptée par l'EAM.

La sélection de l'échantillon de l'EAM s'effectue toujours avec le Système généralisé d'échantillonnage de Statistique Canada. L'échantillonnage aléatoire simple stratifié est utilisé, avec une maximisation du chevauchement avec l'échantillon de l'année précédente. Il est à noter cependant que puisque la stratification a été largement modifiée (principalement en ce qui concerne la strate à tirage obligatoire) et que les coefficients de variation visés ont changé, le chevauchement est moins grand que par le passé.

4. QUELQUES RÉSULTATS

Voici d'abord une comparaison des tailles de la population cible, de la population admissible à l'échantillonnage (population cible moins la partie à tirage nul), de même que les tailles d'échantillon des strates de l'EAM en 2003 et en 2004. Tel que mentionné précédemment, le nombre d'établissements à tirage obligatoire a considérablement diminué.

Tableau 1 : Comparaison des tailles de population et d'échantillon de l'EAM 2003 et 2004. (En nombre d'établissements)

	Taille en 2003	Taille en 2004
Population cible	116 249	122 930
Population admissible à l'échantillonnage	37 198	39 512
Échantillon : strates à tirage complet (tirage obligatoire)	12 871 (12 144)	12 275 (10 354)
Échantillon : strates à tirage partiel	2 615	3 325
Échantillon total	15 486	15 600

On peut remarquer que même si la taille de la population cible a changé, le nombre d'établissements admissibles à l'échantillonnage est demeuré sensiblement le même. Les modifications à la strate à tirage complet combinées à la nouvelle méthode pour les coefficients de variation visés lors de la répartition de l'échantillon ont contribué à une nouvelle répartition provinciale de l'échantillon comme on peut le voir au tableau 2. De la même façon, l'échantillon pour les secteurs forestier, du cuir et des ordinateurs a augmenté aux dépens du secteur du vêtement. En maximisant le chevauchement, seulement 3 000 nouvelles unités ont fait leur entrée dans l'échantillon.

Tableau 2 : Comparaison des tailles d'échantillons de l'EAM 2003 et 2004 par région. (En nombre d'établissements)

Région	Taille d'échantillon en 2003	Taille d'échantillon en 2004
Maritimes	1 433	1857
Québec	3 999	3474
Ontario	5 166	5090
Prairies	2 925	3136
Colombie-Britannique	1 917	1967
Territoires du nord	46	76
Total	15 486	15 600

5. CONCLUSION

Le remaniement du plan de sondage, et en particulier celui du plan d'échantillonnage, de l'EAM fut l'occasion de mettre la stratégie à jour de façon à ce qu'elle cadre mieux avec les besoins des utilisateurs tout en tentant de l'uniformiser du mieux possible avec celle de l'EUE. Il subsiste encore certaines différences entre la stratégie de l'EUE et celle de l'EAM, tant pour le plan de l'échantillonnage que pour les étapes subséquentes; des travaux futurs permettront d'harmoniser davantage les deux enquêtes. Un des défis futurs sera de faire une meilleure utilisation des données fiscales.

L'initiative a aussi affecté les étapes subséquentes de l'enquête qui feront probablement l'objet de nouvelles études. Les données selon le nouveau plan de sondage seront disponibles vers le printemps 2006, différentes analyses seront alors effectuées.

REMERCIEMENTS

Les auteurs tiennent à remercier Hélène Bérard, François Gagnon et Paul Labelle pour leur aide à la révision de cet article.

RÉFÉRENCES

Smith, Philip (1998), « Trousse d'information sur le PASEP », rapport non publié, Ottawa, Canada : Statistique Canada.

Maranda, F. et D. Royce (1998), « Groupe de travail sur l'acquisition des données auprès des entreprises », Document technique de Statistique Canada, Ottawa, Canada.

Lavallée, P. et M. A. Hidioglou (1988), « Sur la stratification de populations asymétriques », *Techniques d'enquête*, juin 1988, col. 14, no. 1, p. 35-45.