

# **LA CONVERSION DE DONNÉES HISTORIQUES DE L'ENQUÊTE MENSUELLE SUR LE COMMERCE DE GROS ET DE DÉTAIL**

## **1. INTRODUCTION**

L'Accord du libre-échange nord-américain entre le Canada, les États-Unis et le Mexique, signé en janvier 1994, a créé un contexte favorable à une classification des activités économiques commune aux trois pays signataires. Le Système de classification des industries de l'Amérique du Nord (SCIAN) de 1997 a été conçu à cette fin par les organismes statistiques des trois pays. Le SCIAN s'appuie sur un cadre conceptuel fondé sur la production ou l'offre ; les établissements y sont groupés par classe en fonction de la similitude des procédés qu'ils appliquent à la production de biens et de services. L'utilisation du nouveau système permet une meilleure comparaison des statistiques industrielles entre les trois pays. Toutefois, ceci interrompt la continuité historique des séries nationales basées sur d'autres systèmes de classification. Depuis 1948, Statistique Canada utilisait principalement le système de la Classification type des industries (CTI) et ses différentes révisions (1960, 1970 et 1980).

Depuis 1988, les données de l'Enquête mensuelle sur les commerces de gros et de détail (EMCGD) ont été produites en fonction des définitions de la CTI version 1980. La conception du SCIAN a mené à un remaniement de l'enquête qui a diffusé ses premières estimations sur une base SCIAN lors du mois de référence avril 2004 en juin 2004. Pour maintenir la continuité historique des estimations, il a fallu produire des estimations historiques mensuelles selon le SCIAN obtenues grâce à la conversion des données historiques recueillies selon la CTI.

## **2. DÉFINITIONS DU COMMERCE DE GROS ET DU COMMERCE DE DÉTAIL**

Sous la CTI et le SCIAN, les secteurs du commerce de gros et du commerce de détail ont la même fonction principale, soit l'achat de marchandises aux fins de revente. La distinction entre les deux secteurs dépend par contre du système de classification utilisé. Selon la CTI, la distinction est fondée sur la catégorie de clients. Les marchandises des détaillants sont destinées au public, pour usage personnel ou ménager. Les grossistes revendent des marchandises à des détaillants, des industriels, des commerçants, des établissements publics, des agriculteurs, des professionnels et à d'autres grossistes. La différence entre le commerce de gros et le commerce de détail, selon le SCIAN, réside plutôt dans le processus de production utilisé, à savoir le fait que les ventes sont effectuées dans un magasin ou non. Ce changement de concept entraîne un changement de secteurs pour certains magasins. Par exemple, les magasins d'ordinateurs, les marchands de matériaux de construction (y compris les centres de rénovation), ainsi que les magasins de fournitures de bureau et de papeterie, qui étaient tous des grossistes selon la CTI, deviennent des détaillants sous le SCIAN. Les établissements dont l'activité principale est l'installation et la réparation, activité appartenant au secteur du commerce de détail selon la CTI, sont maintenant classés dans le secteur des services selon le SCIAN. Le tableau 1 présente un aperçu des mouvements entre les secteurs de la CTI et

du SCIAN. On y présente la distribution des ventes des secteurs du commerce de gros et du commerce de détail de la CTI selon la nouvelle classification du SCIAN.

Tableau 1 – Ventes selon le secteur d'industrie de la CTI sous le SCIAN (1998-2001)

Secteur selon la classification	SCIAN :	SCIAN :	SCIAN :	total
	Commerce de détail	Commerce de gros	Autres secteurs	
CTI : commerce de détail	96,3%	-	3,7%	100%
CTI : commerce de gros	4,2%	94,3%	1,5%	100%

L'activité reste fortement concentrée dans le même secteur industriel. Par contre, il faut souligner que l'EMCGD publie ses données à un niveau plus raffiné que le secteur, soit les groupes de commerce. Alors que l'enquête courante sous la CTI a 11 groupes de commerce d'intérêt pour le commerce de gros et 18 pour le commerce de détail, l'enquête remaniée sous le SCIAN en a respectivement 15 et 19. Les changements entre les anciens et les nouveaux groupes de commerce sont nombreux et variés. Afin de permettre l'analyse des 34 nouvelles séries selon le SCIAN, il est nécessaire d'estimer leurs valeurs historiques à l'aide de la rétopolation des données obtenues sous la CTI.

### 3. MÉTHODES CONSIDÉRÉES

Le Registre des entreprises (RE) de Statistique Canada est une base de données sur la population des entreprises canadiennes. Les établissements de ces entreprises sont répertoriés sur le RE et classés entre autres selon leur type d'activités industrielles et leur classification géographique. L'EMCGD sous la classification selon la CTI avait recourt à un plan d'échantillonnage stratifié aléatoire simple sans remise qui était en place depuis 1988 et qui utilisait le RE comme base de sondage. Depuis 1998, les établissements sur le RE ont été classés selon la CTI et le SCIAN. Cette double classification permet donc d'obtenir facilement des estimations selon le SCIAN à partir des estimations selon la CTI. Il suffit d'assigner la classification SCIAN aux unités échantillonnées selon l'information disponible sur le RE et d'effectuer une estimation par domaine.

Puisque toutes les unités de la population sont classées selon les deux systèmes, il aurait également été possible de procéder par post-stratification. Cette option a été rejetée car la classification SCIAN des unités hors échantillon est de moins bonne qualité que celle des unités échantillonnées, et ce particulièrement en 1998. Les estimations selon le SCIAN ont plutôt été produites à l'aide de l'estimation par domaine pour la période allant de 1998 jusqu'à la mise en œuvre de l'enquête remaniée. Pour satisfaire les besoins de séries historiques des utilisateurs, la rétopolation a commencé en janvier 1991 pour le commerce de détail et en janvier 1993 pour le commerce de gros.

L'approche retenue pour la rétopolation des estimations de l'EMCGD est une approche dite « macro ». Selon cette approche, les unités échantillonnées pour les estimations sur

la base de la CTI n'ont pas besoin d'être reclassifiées individuellement. Pour chaque groupe d'intérêt selon le SCIAN, on utilise plutôt une combinaison linéaire pondérée du total de chacun des groupes d'intérêt selon la CTI. Le total  $X_j^{(t)(geo)}$  du groupe de commerce  $j$  selon le SCIAN, au temps  $t$  pour la région géographique  $geo$  est donné par

$$X_j^{(t)(geo)} = \sum_i \alpha_{ij}^{(t)(geo)} * X_i^{(t)(geo)}$$

où  $X_i^{(t)(geo)}$  représente les ventes du groupe de commerce  $i$  selon la CTI, au temps  $t$  pour la région géographique  $geo$  et  $\alpha_{ij}^{(t)(geo)}$  représente le coefficient de conversion entre le groupe de commerce  $i$  selon la CTI et le groupe de commerce  $j$  selon le SCIAN, au temps  $t$  pour la région géographique  $geo$ . Ils sont estimés à l'aide des données pour lesquelles la double classification est connue.

L'avantage majeur de cette approche est qu'aucun traitement supplémentaire n'est nécessaire pour les ajustements à des niveaux agrégés (lissage ou autres) effectués aux séries sous la CTI. Son désavantage principal découle de l'introduction d'erreurs lors de l'estimation des coefficients de conversion.

#### 4. L'ESTIMATION DES COEFFICIENTS DE CONVERSIONS

De janvier 1998 à mars 2004, tous les établissements couverts par l'EMCGD ont été codés à la fois sur la base de la CTI et du SCIAN. Cette double classification nous permet de produire des estimations selon le SCIAN à partir des estimations de la CTI et de calculer des facteurs de conversion lorsque cette double classification n'était pas possible.

Un coefficient de conversion représente le pourcentage du total d'un groupe de commerce sous la CTI qui a été attribué à un groupe de commerce SCIAN. Les coefficients jugés invalides par les spécialistes du commerce de détail et de gros ou de valeur absolue inférieure à 0.3% ont été éliminés et réalloués. Les séries de coefficients non nuls ont été analysées pour détecter la présence de différences régionales, de saisonnalité ou de valeurs aberrantes.

Par ce fait même nous avons décidé d'utiliser pour la période de 1991 à 1997 des coefficients de conversion mensuels *moyens*. Ces facteurs de conversion ont été estimés à l'aide de la moyenne des coefficients calculés sur les mois homologues des années 1998 à 2001 et ce pour chaque province et territoire. Aucune rétropolation n'a été effectuée pour les Régions métropolitaines de recensement (RMR) de Montréal, Toronto et Vancouver, pour lesquelles des données sont disponibles pour le commerce de détail à partir de janvier 2004.

## **5. ASSURANCE DE LA QUALITÉ**

Différents facteurs affectent l'efficacité de l'approche « macro » utilisée pour la rétopolation des estimations de l'EMCGD. Un facteur est la base de sondage elle-même. Une erreur de classification selon la CTI ou le SCIAN pour un mois donné entre 1998 et 2001 affecte évidemment le mois en question ainsi que tous les mois homologues des années 1991 à 1997. Afin de réduire l'impact des unités mal classées, les gros contributeurs ont été vérifiés manuellement et recodés au besoin. Les corrections apportées à la base de sondage depuis 1998 ont été répertoriées dans le but d'évaluer leur impact ; lorsque jugés nécessaires, des ajustements aux séries estimées sous le SCIAN ont été apportés.

Un autre facteur affectant l'efficacité de l'approche « macro » provient de l'utilisation des coefficients de conversion calculés sur des années récentes (1998-2001) pour estimer les coefficients de conversion des années antérieures. Cette méthode est efficace si la répartition selon le SCIAN est stable d'une année à l'autre ; si elle ne l'est pas, nous considérons quand même que le risque d'erreur est plus faible en 1997 qu'en 1991. L'hypothèse de stabilité a été acceptée dans la majorité des cas. Il faut cependant proscrire l'utilisation des coefficients de conversion estimés uniquement par la moyenne lorsqu'une industrie a subi un changement structurel important. Un exemple est l'industrie des centres de rénovation (SCIAN 444110). Puisque que le secteur de la rénovation a connu une croissance marquée au cours des dernières années et que l'avènement des centres à grandes surfaces est récent dans l'économie canadienne, leur contribution au groupe de commerce de la CTI en 1998 représente mal la contribution qu'ils avaient au début des années 1990. Nous avons donc dû ajuster les coefficients de conversion qui leur sont associés à la baisse pour les années 1991 à 1997. La valeur de l'ajustement lui-même est basée sur l'analyse des experts et les résultats d'une classification partielle sous le SCIAN au niveau micro. Ce type d'ajustement a permis de modéliser les variations dans le temps des coefficients. D'autres éléments liés à la saisonnalité des coefficients tels les jours ouvrables et l'effet de Pâques n'ont pas été considérés étant donné le peu d'observations disponibles au moment de l'analyse.

## **6. LE COMMERCE DE GROS – LES INVENTAIRES**

L'EMCGD a deux variables d'intérêt: les ventes et les stocks. Cette dernière n'est diffusée que pour le commerce de gros par groupe de commerce pour tout le Canada. Toute l'analyse des coefficients de conversion a été effectuée au niveau des ventes. Au lieu de convertir selon le SCIAN les séries des stocks selon la CTI, nous avons converti les séries rétopolées des ventes en stocks à l'aide de ratios. Des ratios stocks/ventes sont calculés pour chaque mois depuis 1993 au niveau des groupes de commerce CTI et appliqués aux groupes de commerce SCIAN correspondants. Les correspondances sont établies en fonction de la règle suivante : un groupe de commerce selon le SCIAN correspond à un groupe de commerce selon la CTI si ce dernier contribue à plus de 99 % du total du groupe selon le SCIAN. Sept groupes du commerce de gros selon le SCIAN

ont un groupe correspondant selon la CTI. Les groupes du SCIAN sans groupe correspondant selon la CTI sont traités à l'aide d'un ratio stocks/ventes global.

## 7. CONTINUITÉ DES SÉRIES SOUS LE SCIAN

Les séries sous la classification du SCIAN se divisent en trois parties. Une première de janvier 1991 à décembre 1997 où les estimations sont obtenues à l'aide de coefficients de conversion *estimés*. La seconde partie commence en janvier 1998 et se termine avec mars 2004. Dans cette partie, les séries sous le SCIAN sont obtenues par estimation par domaine ou à l'aide de coefficients de conversion *observés*. La troisième partie débute en même temps que l'enquête remaniée. La deuxième et la troisième partie se sont chevauchés de quelques mois lorsque l'ancienne et la nouvelle enquête étaient toutes les deux en production.

Le passage de l'enquête basée sur la CTI à l'enquête basée sur le SCIAN cause un bris dans les séries. Ce saut s'explique par le changement de classification mais aussi par d'autres changements méthodologiques. Les résultats du test en parallèle ont permis d'ajuster le niveau des séries rétopolées. Un ajustement multiplicatif constant dans le temps a été effectué pour ajuster les séries historiques aux niveaux publiés provenant de l'enquête remaniée.

Cet étalonnage enlève son additivité au système de séries parce que les séries sont étalonnées individuellement. Par exemple, ce processus entraîne des écarts entre la somme des groupes de commerce et la somme des provinces et territoires. Afin d'en restituer l'additivité, un procédé de réconciliation est appliqué aux séries étalonnées.

Le processus de réconciliation est tel que : a) les composantes du système sont modifiées le moins possible en pourcentage et b) l'additivité du système de séries est respectés pour chaque mois.

Ainsi, toutes les séries sont réconciliés de telle sorte que pour chaque mois : (1) l'addition des groupes de commerces par province et territoire égale les totaux provinciaux et territoriaux, (2) l'addition des provinces et territoires par groupes de commerce égale les totaux par groupes de commerces et (3) l'addition des totaux provinciaux et territoriaux égale la somme des totaux par groupes de commerces.

En tenant compte de la saisonnalité, des bris dans les séries sous le SCIAN étaient également observables en janvier 1998. On passe alors de coefficients estimés en décembre 1997 à des coefficients observés. Pour amoindrir l'effet, toutes les données rétopolées de 1998 ont été recalculées en utilisant les coefficients estimés, c'est-à-dire en calculant la moyenne des quatre années, incluant 1998. Notons que les coefficients de 1998 diffèrent plus souvent de la moyenne que les trois autres années et que les coefficients très aberrants étaient enlevés du calcul de la moyenne. En allongeant la première partie de la série jusqu'en décembre 1998, on annule la présence de bris entre les deux premières parties.