



Rapport sur la méthodologie utilisée pour convertir les séries chronologiques de l'Enquête trimestrielle sur les marchandises vendues au détail (ETMVD) de la CTI au SCIAN

Date de création : 23 août 2005
Créé par : Norman Fyfe

Dernière mise à jour : 7 octobre 2005
Mis à jour par : Norman Fyfe

Date d'impression : 7 octobre 2005



1. Introduction

La signature de l'Accord de libre-échange nord-américain (ALENA) entre le Canada, les États-Unis et le Mexique en janvier 1994 a créé le besoin d'un système de classification économique commun. 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 axé sur la production ou l'offre; les établissements y sont groupés par classe en fonction de la similitude des procédés utilisés pour la production de biens et de services. Le nouveau système permet une meilleure comparaison des statistiques industrielles entre les trois pays¹.

L'Enquête trimestrielle sur les marchandises vendues au détail (ETMVD) est réalisée par Statistique Canada depuis le premier trimestre de 1997. Les ventes au détail y sont ventilées en plus de 100 groupes de marchandises. L'ETMVD est un sous-échantillon de l'Enquête mensuelle sur le commerce de détail (EMCD), enquête parallèle conçue pour évaluer les ventes selon le type de magasin, et est étalonnée d'après cette dernière. Le remaniement de l'EMCD et son adoption subséquente, en 2004, de la version 2002 du SCIAN ont mené au remaniement de l'ETMVD et à sa conversion au SCIAN, puisque son échantillon représente un sous-ensemble des détaillants participant à l'EMCD. Depuis 1948, Statistique Canada a utilisé principalement la version canadienne de la Classification type des industries (CTI) et ses diverses révisions (1960, 1970 et 1980). L'ETMVD a été fondée sur la version de 1980 de la CTI depuis son lancement en janvier 1997. La conversion au SCIAN en 2004 a fait que les nouvelles séries chronologiques de l'ETMVD fondée sur le SCIAN ont effectivement débuté au premier trimestre de 2004 et ne contenaient aucune donnée historique antérieure.

Le présent document décrit dans les grandes lignes la méthodologie qui a été utilisée pour convertir les séries chronologiques de l'ETMVD de la CTI au SCIAN. En particulier, il décrit comment de nouvelles séries de données sur les biens et services selon le type de magasin fondées sur le SCIAN ont été générées d'après les données d'enquêtes chronologiques fondées sur la CTI, comment ces séries ont été ajustées à la nouvelle ETMVD fondée sur le SCIAN et comment les séries rétropolées ajustées ont été étalonnées à l'EMCD pour assurer la convergence avec l'EMCD.

2. Méthodologie utilisée pour convertir les séries chronologiques de l'ETMVD au SCIAN.

Dans le cadre des efforts en vue de convertir les séries chronologiques fondées sur la CTI au SCIAN, l'équipe de l'EMCD a créé des microfichiers à codage double pour les établissements compris dans son échantillon. Cet exercice a été

¹ SCIAN Canada 1997, Statistique Canada, n° 12-501-XPF au catalogue.

possible grâce au fait que les établissements statistiques inscrits dans le Registre des entreprises (RE) de Statistique Canada sont classés, depuis 1998, selon la CTI ainsi que selon le SCIAN. Le RE est la base de sondage de l'EMCD et de la plupart des enquêtes-entreprises réalisées par Statistique Canada. Cette double classification a permis de concevoir pour l'EMCD une méthode en vue d'obtenir des estimations par domaine conformément au SCIAN, même si la base de sondage était fondée sur la CTI.

Les fichiers à double codage comprenaient des enregistrements contenant le numéro d'identification unique de l'établissement, son poids de sondage, la valeur totale des ventes et les codes du SCIAN et de la CTI. Pour les établissements pour lesquels il existait une correspondance du code de la CTI avec plus d'un code du SCIAN (appelés « unités scindées »), l'enregistrement contenait un marqueur ainsi que la proportion des ventes totales de l'établissement imputables à chaque code du SCIAN.

Comme l'échantillon de l'ETMVD est un sous-échantillon de celui de l'EMCD, les mêmes fichiers à codage double que ceux utilisés pour la rétopolation des données de l'EMCD ont été utilisés pour attribuer un code du SCIAN à chaque enregistrement dans son ancien échantillon fondé sur la CTI. Suivant cette approche, chaque unité de l'échantillon a été convertie individuellement aux codes du SCIAN.

Des estimations des ventes de marchandises selon le groupe de commerce² fondées sur le SCIAN ont alors été obtenues pour l'ETMVD en totalisant les données sur les marchandises pondérées et recodées par domaine du SCIAN et en agrégeant les estimations par domaine du SCIAN au niveau du groupe de commerce.

2.1 Conversion des données historiques de l'ETMVD : problèmes de qualité et sources d'erreur

Les séries de données chronologiques converties doivent être analysées et interprétées avec prudence. Nous exposons à la présente section plusieurs problèmes dont il faut tenir compte en ce qui concerne la qualité des données.

2.1.1 – Erreurs de codage

Diverses sources d'erreur ont une incidence sur l'efficacité de l'approche de recodage de données micro qui a été adoptée pour convertir au SCIAN les estimations historiques fondées sur la CTI. Une première source possible d'erreur est la base de sondage de l'enquête proprement dite. Une erreur de classification dans le Registre des entreprises sous la CTI produirait un code du

² Les groupes de commerce sont des agrégations spéciales des industries du Système de classification des industries de l'Amérique du Nord (SCIAN). Il existe 19 groupes de commerce qui sont agrégés pour former 8 secteurs de groupes de commerce.

SCIAN correspondant incorrect. En outre, comme la double classification a été introduite en 1998, le taux d'erreurs de codage risque d'être plus élevé pour les années remontant loin dans le passé que pour celles plus récentes, pour lesquelles la plupart des problèmes de classification ont été résolus. Afin de réduire l'effet des unités classées incorrectement, les unités dont la contribution aux estimations était importante ont fait l'objet d'une vérification manuelle et ont été recodées au besoin.

En règle générale, si le code CTI d'un établissement dans le microfichier de l'ETMVD n'était pas le même que celui figurant dans le fichier à double codage de l'EMCD, le code CTI de l'ETMVD a été considéré comme étant la valeur par défaut, car il reflétait l'information d'échantillonnage la plus récente. L'une des raisons des différences de classification vient du fait que les microfichiers de l'ETMVD sont révisés annuellement et contiennent l'information la plus à jour, tandis que les fichiers à double codage reflètent l'information la plus récente disponible dans la base de sondage au moment de la création des fichiers à double codage.

2.1.2 – Traitement des établissements présentant des liens CTI-SCIAN multiples

Comme dans le cas de toute mise en concordance, un certain nombre de « scissions d'unité » ont eu lieu. Les unités scindées sont des établissements codés selon la CTI pour lesquels il n'existe pas de relation unique avec un code SCIAN. La plupart des unités scindées les plus importantes ont été reclassifiées, après des recherches, sous un code du SCIAN dominant simple, reflétant les modalités de déclaration courante de l'ETMVD. Les unités scindées importantes sont définies comme étant des unités représentant au moins 1 % de la valeur totale d'un groupe de commerce. En général, environ une douzaine de ces grandes unités ont été observées pour un mois donné. Elles représentaient environ la moitié des estimations imputables aux unités scindées. Le nombre de ces grandes unités a été relativement constant tout au long de la période de réropolation.

Les unités scindées restantes étaient des unités trop petites pour avoir une incidence sur les estimations (ou faisaient toutes partie du même groupe de commerce) et ont donc été ignorées. Ces unités scindées plus petites ne représentaient que 2 % environ de l'estimation globale des ventes.

Le groupe de commerce contenant le plus grand nombre d'unités scindées était celui des magasins de vêtements, et une part importante des estimations (45 % en décembre 2003) pour ces magasins était attribuable à des unités scindées. Cependant, comme la majorité des unités scindées ont été relevées dans le groupe de commerce des magasins de vêtements, les effets des scissions sur les estimations se sont annulés au niveau du groupe de commerce. Pour tous les autres groupes de commerce, les unités scindées représentaient moins d'un quart de l'estimation totale pour le groupe de commerce en 2003.

2.1.3 – Prise en compte des établissements qui passent du secteur de la vente en gros au secteur de la vente au détail

Suite à l'adoption de la version 2002 du SCIAN, certains établissements couverts par l'ETMVD qui étaient classés antérieurement comme étant des grossistes en vertu de la CTI 80 sont maintenant classés comme étant des détaillants ou des « **nouveaux entrants** ». Les magasins d'ordinateurs, les centres de rénovation, les magasins de matériel de construction et les magasins de fournitures de bureau et de papeterie en sont des exemples.

Selon une étude sur le secteur du commerce de détail menée par l'équipe de l'EMCD pour la période allant de 1998 à 2001, dans l'ensemble, 96,3 % des ventes de détail codées sous la CTI 80 sont restés dans le secteur de la vente au détail en vertu de la classification conformément au SCIAN, 3,7 % ayant été transférés aux secteurs de la fabrication et des services. De même, 94,3 % des ventes de gros codées selon la CTI 80 sont demeurés dans le secteur de la vente en gros sous le SCIAN, 4,2 % ayant été transférés dans le secteur de la vente au détail, et 1,5 %, dans d'autres secteurs.

Pour tenir compte du fait que l'ETMVD ne peut fournir, pour les estimations par domaine sous le SCIAN, des renseignements sur les nouveaux entrants, les données de l'ETMVD de 2004 ont été utilisées pour recréer des ventilations des marchandises et des profils saisonniers pour chaque période de réropolation, pour les établissements qui chevauchaient l'ancien échantillon d'établissements de vente en gros³ selon la CTI et le nouvel échantillon d'établissements de vente au détail selon le SCIAN.

Seules les données sur certains types d'établissement ont été retenues. Elles comprenaient les établissements ayant un code du SCIAN associés à une forte proportion de nouveaux entrants, à savoir les magasins de pièces automobiles et d'accessoires, les marchands de pneus, les magasins d'ordinateurs et de logiciels, les centres de rénovation, les magasins de peinture et de papier peint, les marchands de matériel de construction, les magasins d'équipement énergétique extérieur, les pépinières et les magasins de fournitures de bureau et de papeterie. Ensemble, ces neuf groupes de code du SCIAN représentaient la majorité des nouveaux entrants. Les magasins d'ordinateurs et de logiciels et les magasins de matériaux de construction et de produits extérieurs pour la maison ont été les plus touchés, près de 100 % et de 79 % des estimations pour ces groupes de commerce, respectivement, étant imputables aux nouveaux entrants en 2003.

³ Les unités chevauchantes sont les enregistrements d'établissements qui participaient à l'Enquête Mensuelle sur le Commerce de Gros (EMCG) sous la CTI et qui furent sélectionnées pour l'échantillon de la nouvelle EMCD sous le SCIAN

La proportion d'estimations de l'ETMVD attribuable aux nouveaux entrants a varié durant la période de rétropolation, mais, en général, est demeurée élevée pour les neuf codes du SCIAN susmentionnés. Les estimations par domaine pour d'autres codes du SCIAN affectés par les nouveaux entrants, mais n'ayant pas été retenus pour l'exercice de rétropolation ont été ajustées durant le processus d'étalonnage d'après l'EMCD. Cette option était la meilleure disponible, car il n'existait aucune ventilation fiable des ventes de marchandises pour les nouveaux entrants au cours de la période de rétropolation. L'inconvénient de cette méthode réside dans l'application des distributions mensuelles des ventes de marchandises de 2004 aux années antérieures.

Pour les unités non chevauchantes, la distribution des ventes de marchandises a été imputée d'après une moyenne sur les 12 mois de 2004 des enregistrements de collecte provenant des établissements regroupés sous le même code du SCIAN. L'effet de ces unités sur la série rétropolée était négligeable, car elles contribuaient peu aux estimations, étant donné leur taille assez petite. Puis, les unités chevauchantes et non chevauchantes sélectionnées à partir de l'échantillon de l'Enquête mensuelle sur le Commerce de Gros (EMCG) ont été ajoutées aux microfichiers de l'ETMVD.

3. Ajustement des séries chronologiques rétropolées aux niveaux de l'enquête courante

Une analyse graphique des séries rétropolées a révélé un changement de niveau important pour un certain nombre de marchandises entre la dernière période rétropolée et la première période de l'échantillon codé selon le SCIAN. Ces changements de niveau ont plusieurs causes, y compris les changements d'échantillon, les changements de classification et d'autres changements méthodologiques. La méthode utilisée pour corriger ces changements de niveau est exposée en détail à la présente section.

Novembre 2003 a été le dernier mois de collecte des données sur la base de la CTI. Afin d'obtenir un ensemble complet publiable d'estimations fondées sur la CTI pour le dernier trimestre de 2003 et pour évaluer partiellement l'effet du nouvel échantillon sur les estimations de l'ETMVD, un ensemble d'estimations fondées sur la CTI ont été calculées d'après l'échantillon codé selon le SCIAN pour décembre 2003 (premier mois de collecte sous le SCIAN) et le premier trimestre de 2004. Cet ensemble d'estimations a servi d' « exécution parallèle » de l'ETMVD qui a été utilisé pour estimer l'effet de l'introduction d'un nouvel échantillon sur l'estimation.

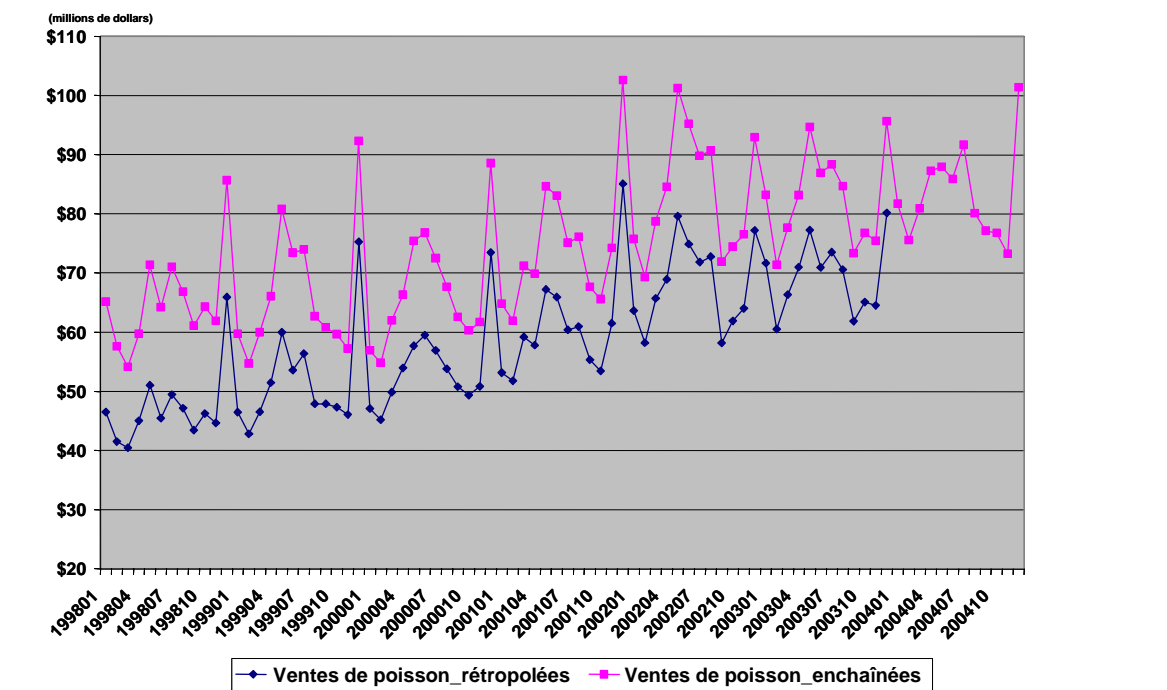
Ces estimations ont été générées en commençant par copier les données provenant des enregistrements identifiés comme étant des « unités chevauchantes » de l'échantillon selon le SCIAN dans les enregistrements correspondant de l'échantillon selon la CTI. Pour les « unités non

chevauchantes », le processus a été plus complexe. Le chiffre total des ventes d'un établissement a été obtenu d'après les données de l'EMCD recueillies par la collecte en parallèle (c'est à dire que les données recueillies par l'EMCD sous le nouvel échantillon sur la base SCIAN furent à la fois codées selon la CTI et le SCIAN pour une période de cinq mois) et sa répartition par marchandise a été dérivée par imputation en utilisant les données historiques les plus récentes/appropriées pour l'établissement.

Les estimations des ventes de marchandises fondées sur la CTI résultantes ont été converties au SCIAN par recodage des microdonnées selon un processus similaire à celui utilisé pour la rétropolation des séries chronologiques de l'ETMVD.

Les données d'« exécution en parallèle » et les données courantes codées selon le SCIAN ont été utilisées pour ajuster les séries chronologiques rétropolées aux niveaux publiés d'après la nouvelle enquête. Les séries chronologiques rétropolées ont été ajustées au niveau de la nouvelle enquête fondée sur le SCIAN au moyen de ratios multiplicatifs, constants dans le temps. Les ratios ont été calculés en divisant les *estimations courantes selon le SCIAN* par les *estimations rétropolées correspondantes* pour chaque combinaison de marchandise et groupe de commerce. Chaque série chronologique rétropolée a ensuite été multipliée par le ratio d'ajustement, constant dans le temps, correspondant. La figure 1 illustre graphiquement la façon dont les niveaux d'une série rétropolée sur les ventes de marchandises ont été ajustés de façon à ce qu'ils coïncident avec les nouveaux niveaux de 2004 selon le SCIAN en utilisant les ratios d'ajustement décrits plus haut.

Figure 1 : Ventes de poisson frais et d'autres fruits de mer au Canada : comparaison des données enchaînées et non enchaînées



Il a été décidé d'ajuster les séries réropolées sur les niveaux de l'enquête courante de mars 2004, car les ratios de mars donnaient de meilleurs ajustements pour les marchandises saisonnières, comme les produits de jardinage et d'horticulture.

Ce processus d'ajustement de niveau a enlevé son « additivité » au système de séries, ce qui a produit des écarts entre les totaux de groupe de commerce et la somme des ventes de marchandises dans les groupes. Pour rétablir l'additivité, on a recouru à un processus de réconciliation des séries visant à répartir proportionnellement l'écart entre un agrégat de marchandises et ses composantes. Pour illustrer ce processus, examinons les ventes de livres, de journaux et de périodiques. Les livres, journaux et périodiques constituent un groupe de marchandises pour lequel sont calculées des estimations. Cependant, ces estimations des ventes peuvent être ventilées en deux marchandises : (1) les ventes de livres et (2) les ventes de journaux et de périodiques. L'estimation ajustée pour les marchandises individuelles devrait être égale à l'estimation ajustée pour le groupe de marchandises. Cependant, des écarts surviennent, car le processus d'ajustement décrit plus haut produit des facteurs d'ajustement distincts pour le groupe et pour les marchandises composantes.

Compte tenu du niveau de confiance plus élevé de l'estimation plus agrégée (meilleure déclaration, taux d'imputation plus faible, etc.), il a été décidé de préserver les chiffres au niveau du groupe de marchandises. Par conséquent, on a appliqué à chaque marchandise un ratio du total pour le groupe de marchandises à la somme des deux marchandises. Le total global a été obtenu indirectement par sommation des marchandises sur les groupes de commerce.

Après le processus de réconciliation, les séries ont été soumises à un processus d'étalonnage. Les totaux des ventes réropolées de l'ETMVD ont été étalonnés au niveau du groupe de commerce de l'EMCD, sauf pour le groupe de commerce des grands magasins (l'ETMVD comprend les ventes des concessions des grands magasins, mais l'EMCD ne les comprend pas).

4. Corrections des valeurs extrêmes et des changements de niveau dans les séries réropolées.

Faute de temps et de ressources, il n'aurait pas été faisable d'analyser chacune des 2 736 séries chronologiques réropolées (19 groupes de commerce et 144 marchandises) de l'ETMVD en vue de déceler les anomalies. Par conséquent, on a dressé la liste des séries de données chronologiques sur les marchandises les plus importantes par groupe de commerce. Figuraient sur cette liste les séries chronologiques représentant au moins 1 % du total estimé des ventes dans un groupe de commerce donné. Dans la plupart des cas, les dix marchandises les plus importantes représentaient la grande majorité des

estimations dans un groupe de commerce donné. Le nombre final de séries rétrolées nécessitant une méthode de dépistage des valeurs extrêmes et des changements de niveau était d'environ 225.

Ces séries principales ont été soumises à un dépistage minutieux des valeurs extrêmes et des changements de niveau réalisé au moyen du programme X-12-ARIMA. Ce programme statistique complexe a permis de déceler des anomalies dans les données rétrolées en tenant compte des jours ouvrables, des jours fériés mobiles et des variations saisonnières. Les valeurs aberrantes et les changements de niveau dans les séries chronologiques peuvent avoir de nombreuses causes et compliquer l'interprétation des données au fil du temps. Un changement de niveau concerne toutes les observations faites avant un point particulier dans le temps. Les observations passent d'un niveau à un autre, et demeurent à ce nouveau niveau. En revanche, une valeur extrême additive ne comporte qu'une seule observation ayant une valeur anormalement grande ou faible.

Dans le cas des séries économiques sur le commerce de détail, les valeurs extrêmes et les changements de niveau peuvent être causés par la restructuration d'un secteur (par exemple, ouverture ou fermeture d'une chaîne de magasins) ou par le déplacement d'une grande entreprise d'une industrie dans une autre. Les événements inhabituels, comme l'épidémie du SRAS, la maladie de la vache folle ou une panne d'électricité générale, peuvent aussi causer des valeurs inhabituelles. La modification de l'échantillon, due à la « création de nouvelles entreprises » ou à la « fermeture d'entreprises existantes », peut aussi causer des changements de niveau. Dans d'autres cas, des erreurs d'enquête non dues à l'échantillonnage, comme des problèmes de couverture ou de mesure, peuvent avoir lieu. Ces dernières erreurs sont celles qui doivent être corrigées afin d'éviter que soit faussée la prochaine dessaisonnalisation des séries chronologiques de l'ETMVD. Les autres valeurs extrêmes ou changements de niveau causés par des événements inhabituels ou une restructuration économique sont maintenus afin de refléter la réalité économique, mais sont pris en compte dans le calcul des coefficients saisonniers des séries chronologiques.

Dans la mesure du possible, les écarts non justifiés décelés dans les séries principales ont été éliminés par la correction des erreurs ou l'amélioration des imputations statistiques présentes dans les microdonnées rétrolées en se fondant sur les connaissances antérieures concernant les données. Si un ajustement aux données micro n'était pas faisable, on a recouru à des macroajustements pour corriger les erreurs dans les données rétrolées des séries principales. Les autres séries n'ont pas été corrigées pour ce genre de valeurs atypiques.