
02 CLASSIFICATIONS



Doc. Class/05/40

1 July 2005

**The conversion of historical time series according to a revised
classification in the wholesale and retail sale monthly survey**

(Canadian Statistical Society, June 2003)

EN

**NACE/CPA Implementation Task Force meeting 4-6 July 2005
Agenda item 6**

STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES

02 Statistical Governance, quality and evaluation
Tel. (352) 4301-32023, Fax (352) 4301-33899

Assemblée annuelle de la SSC, juin 2003
Recueil de la Section des méthodes d'enquête

The DATA CONVERSION HISTORICAL ACCORDING TO A NEW SYSTEM OF CLASSIFICATION FOR the MONTHLY SURVEY INTO WHOLESALE AND OF RETAIL

Susie Fortier

SUMMARY

The systems of industrial classification provide a common framework of concepts making it possible to describe the economic activity. During years, Statistics Canada used various versions of the standard Classification of industries (CTI) and, more recently, the System of classification of industries of North America (SIC). The changes of system of classification create cracks in the time series. In order to preserve the continuity of these series, it becomes necessary to convert the historical estimates obtained under an old system into historical estimates according to more recent classification. A major investigation of Statistics Canada, the Monthly survey of wholesale and RETAIL (MWRTS), currently undergoes a complete redesign which includes the use of the new classification and must take up the challenge of such a conversion. The current version of the MWRTS was developed with the end of 1980 in order to produce mainly estimates on the sales and stocks for industrial sectors defined by the CTI of 1980. The altered investigation will henceforth produce estimates according to SIC'S since 2004. Various options were considered in order to convert the data of the sectors of the CTI into estimates for the sectors of the SIC. The advantages and disadvantages of these options will be discussed. The option chosen, being given the constraints connected to the MWRTS, is presented.

Assemblée annuelle de la SSC, juin 2003
Recueil de la Section des méthodes d'enquête

LA CONVERSION DE DONNÉES HISTORIQUES SELON UN NOUVEAU SYSTÈME DE CLASSIFICATION POUR L'ENQUÊTE MENSUELLE SUR LE COMMERCE DE GROS ET DE DÉTAIL

Susie Fortier

RÉSUMÉ

Les systèmes de classification industrielle fournissent un cadre commun de concepts permettant de décrire l'activité économique. Au cours des années, Statistique Canada a utilisé différentes versions de la Classification type des industries (CTI) et, plus récemment, le Système de classification des industries de l'Amérique du Nord (SCIAN). Les changements de système de classification créent des brisures dans les séries chronologiques. Afin de préserver la continuité de ces séries, il devient nécessaire de convertir les estimations historiques obtenues sous un ancien système en estimations historiques selon la classification plus récente. Une enquête majeure de Statistique Canada, l'Enquête mensuelle du commerce de gros et de détail (MWRTS), subit actuellement un remaniement complet qui inclut l'utilisation de la nouvelle classification et doit relever le défi d'une telle conversion. La version actuelle de l'MWRTS a été développée à la fin des années 1980 afin de produire principalement des estimations sur les ventes et les stocks pour des secteurs industriels définis par la CTI de 1980. L'enquête remaniée produira dorénavant des estimations selon le SCIAN à partir de 2004. Diverses options ont été considérées afin de convertir les données des secteurs de la CTI en estimations pour les secteurs du SCIAN. Les avantages et inconvénients de ces options seront discutés. L'option choisie, étant donné les contraintes reliées à l'MWRTS, est présentée.

MOTS CLÉS : Rétropolation ; système de classification ; coefficient de conversion.

ABSTRACT Industrial classification systems provide a conceptual framework that allows the description of economic activities. Throughout the years, Statistics Canada has used different versions of the Standard Industrial Classification (SIC)

KEY WORDS: Rétropolation; system of classification; coefficient of conversion.

ABSTRACT Industrial classification systems provide has conceptual framework that allows the description of economic activities. Throughout the years, Statistics Canada has used different versions of the Standard Industrial Classification (SIC) system and the North American Industry Classification System (NAICS) for industrial classification. Exchanges in industrial classification systems create disruptions in associated time series of population estimates. To maintain continuity, existing series under the previous classification must be converted according to the new classification. The Monthly Wholesale and Retail Trade Survey (MWRTS), has major survey conducted by Statistics Canada, is in the midst of has redesign and faces the challenges of classification conversion. The current dirty MWRTS was developed in the late 1980' S to produce and inventory estimates for SIC-based industrial sectors. The redesigned survey will produce NAICS-based estimates starting in 2004. This paper summarizes the options considered for backcasting the SIC-based dated into NAICS-based estimates. The pros and idiots of each method final are discussed and the strategy is presented within the specific context of the MWRTS.

KEY WORDS: Backcasting; Classification system; Agreement coefficient

1. INTRODUCTION

The Agreement of free trade, signed in January 1994, created the need for a classification of the economic activities common to Canada, Mexico and the United States. The System of classification of industries of North America (SIC) of 1997 was conceived for this purpose by the statistical organizations of the three countries (Statistical

system and the North American Industry Classification System (NAICS) for industrial classification. Changes in industrial classification systems create disruptions in associated time series of population estimates. To maintain continuity, existing series under the previous classification must be converted according to the new classification. The Monthly Wholesale and Retail Trade Survey (MWRTS), a major survey conducted by Statistics Canada, is in the midst of a redesign and faces the challenges of classification conversion. The current MWRTS was developed in the late 1980's to produce sales and inventories estimates for SIC-based industrial sectors. The redesigned survey will produce NAICS-based estimates starting in 2004. This paper summarizes the options considered for backcasting the SIC-based data into NAICS-based estimates. The pros and cons of each method are discussed and the final strategy is presented within the specific context of the MWRTS.

KEY WORDS: Backcasting; Classification system; Concordance coefficient

1. INTRODUCTION

L'Accord de libre-échange, signé en janvier 1994, a créé le besoin d'une classification des activités économiques commune au Canada, au Mexique et aux États-Unis. 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 (Statistique Canada 2002). 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

Canada 2002). The SIC is pressed on a conceptual framework founded on the production or the offer; the establishments are grouped there by class according to the similarity of the processes which they apply to the production of goods and services. The use of the new system allows a better comparison of the industrial statistics between the three countries. However, this stops the historical continuity of the national series based on other systems of classification. Since 1948, Statistique Canada used mainly the system of the standard Classification of industries (CTI) and its various revisions (in 1960, 1970 and 1980). Conversion with the SIC touches several major investigations of Canada Statistics, of which the Monthly survey into wholesale and of RETAIL (MWRTS).

Since 1988, the data of the MWRTS are produced according to the definitions of the CTI version 1980. The investigation was recently altered to produce estimates of quality according to SIC'S at reduced costs. This redesign also makes it possible to reduce the burden of answer, to establish more powerful information processing systems and to support the use of administrative data lately available, such as the sales according to the Tax on the products and services (TPS) perceived (Bérard 2001). The first results of the altered investigation will be published in 2004. They will be accompanied by monthly historical estimates according to SIC'S obtained thanks to the conversion of the historical data collected according to CTI'S.

This paper presents the challenges associated with the production with historical data according to a new system of classification. Section 2 defines the wholesale and the Retail trade according to each of the two systems of classification. The various methods considered for the retropolation are briefly discussed in section 3. Section 4 identifies selected method and draft of its establishment.

2. DEFINITIONS OF WHOLESALE AND RETAIL TRADE

Under the CTI and the SIC, the sectors wholesale and Retail trade have the same principal function;

classification. Depuis 1948, Statistique Canada utilisait principalement le système de la Classification type des industries (CTI) et ses différentes révisions (en 1960, 1970 et 1980). La conversion au SCIAN touche plusieurs enquêtes majeures de Statistique Canada, dont l'Enquête mensuelle sur le commerce de gros et de détail (MWRTS).

Depuis 1988, les données de l'MWRTS sont produites en fonction des définitions de la CTI version 1980. L'enquête a récemment été remaniée pour produire des estimations de qualité selon le SCIAN à coûts réduits. Ce remaniement permet également de réduire le fardeau de réponse, d'implanter des systèmes informatiques plus performants et de favoriser l'utilisation de données administratives nouvellement disponibles, telles que les ventes selon la Taxe sur les produits et services (TPS) perçue (Bérard 2001). Les premiers résultats de l'enquête remaniée seront publiés en 2004. Ils seront accompagnés d'estimations historiques mensuelles selon le SCIAN obtenues grâce à la conversion des données historiques recueillies selon la CTI.

Cet article présente les défis associés à la production de données historiques selon un nouveau système de classification. La section 2 définit le commerce de gros et le commerce de détail selon chacun des deux systèmes de classification. Les différentes méthodes considérées pour la rétropolation sont brièvement discutées dans la section 3. La section 4 identifie la méthode choisie et traite de son implantation.

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 (Meyer 2001). Selon la CTI, la distinction est fondée sur la catégorie de clients. Les marchandises des détaillants sont destinées au public, pour usage

that is to say the purchase of goods for purposes of resale. The distinction between the two sectors depends on the other hand on the system of classification used (Meyer 2001). According to the CTI, the distinction is founded on the category of customers. The goods of the Retailers are intended for the public, for personal or domestic use. The wholesalers resell goods with Retailers, industrialists, tradesmen, publicly-owned establishments, farmers, professionals or with other wholesalers. The difference between wholesale and the Retail trade, according to SIC'S, resides rather on the production process used, namely the fact that the sales are carried out in a store or not. This change of concept involves a change of sectors for certain stores. For example, the stores of computers, the building material merchants, including the centers of restoration, as well as the stores of paper mill and office supplies, which all were of the wholesalers according to CTI'S, become Retailers under the SIC. The establishments whose principal activity is the installation and repair, activity belonging to the sector Retail trade according to CTI'S, are now classified in the sector of the services according to SIC'S. Table 1 presents an outline of the movements between the sectors of the CTI and the SIC. One presents to it the distribution of the sales of the sectors wholesale and Retail trade of the CTI according to the new classification of the SIC.

Table 1 - Average percentage of sales according to the sector of industry (1998-2001)

Secteur selon la classification	SIC :
commerce de détail	
commerce de gros	autres
secteurs	total
CTI : commerce de détail	96,3%
CTI : commerce de gros	4,2%
	94,3%

The activity remains strongly concentrated in the same industrial sector. On the other hand, it should be stressed that the MWRTS publishes its data on a level more refined than the sector, that is to say the groups of trade. Whereas the current investigation under the CTI has 11 groups of trade of interest for wholesale and 18 for the Retail trade, the

personnel ou ménager. Les grossistes revendent des marchandises à des détaillants, des industriels, des commerçants, des établissements publics, des agriculteurs, des professionnels ou à d'autres grossistes. La différence entre le commerce de gros et le commerce de détail, selon le SCIAN, réside plutôt sur 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 – Pourcentage moyen de ventes selon le secteur d'industrie (1998-2001)

Secteur selon la classification	SIC :
commerce de détail	
commerce de gros	
autres secteurs	total
CTI : commerce de détail	96,3%
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'MWRTS 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 aura respectivement 15 et 19. Les changements entre

investigation altered under the SIC will respectively have of them 15 and 19. The changes between the old ones and the new groups of trade many and are varied. In order to allow the analysis of the 34 new series according to SIC'S, it is necessary to consider their values historical using the retropolation of the data obtained under the CTI.

3. METHODS CONSIDERED

3.1 Setting in context

The Register of the companies (RE) of Statistics Canada is a data base on the population of the Canadian companies. The establishments of these companies are indexed on RE and are classified inter alia according to their type of industrial activities and their geographical classification. The MWRTS resorts to a simple random sampling design laminated without handing-over which is in place since 1988 and which uses RE as bases survey. The industrial stratification is according to the CTI. Since 1998, the establishments on RE are classified according to the CTI and SIC'S. This double classification thus makes it possible to obtain easily estimates according to SIC'S even if the sample design is based on the CTI. It is enough to assign classification SIC with the units sampled according to information available on RE and to carry out an estimate by field. The precision of the estimates is however difficult to control.

Since all the units of the population are classified according to two systems', it would have also been possible to proceed by postratification in order to improve the precision of the estimates according to SIC'S. This option was rejected because classification SIC of the units except sample is of worse quality than that of the sampled units; and this particularly in 1998. The estimates according to SIC'S will be rather produced using the estimate by going field for the period of 1998 until the setting in?uvre of the altered investigation. Certain users, like the system of national accounting, need historical series former to January 1998. The

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

3.1 Mise en contexte

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'MWRTS recourt à un plan d'échantillonnage stratifié aléatoire simple sans remise qui est en place depuis 1988 et qui utilise le RE comme base de sondage. La stratification industrielle est en fonction de la CTI. Depuis 1998, les établissements sur le RE sont classés selon la CTI et le SCIAN. Cette double classification permet donc d'obtenir facilement des estimations selon le SCIAN même si le plan de sondage est basé sur 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. La précision des estimations est toutefois difficile à contrôler.

Puisque toutes les unités de la population sont classées selon les deux systèmes, il aurait également été possible de procéder par postratification afin d'améliorer la précision des estimations selon le SCIAN. 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 seront plutôt 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. Certains utilisateurs, comme le système de comptabilité nationale, ont besoin de séries historiques antérieures à janvier 1998. La rétopolation doit commencer en janvier 1991 pour le commerce de détail, alors que pour le commerce de gros, elle peut commencer en janvier 1993. Différentes

retropolation must start in January 1991 for the Retail trade, whereas for wholesale, it can start in January 1993. Various options for the data conversion of the CTI to the SIC were studied for several investigations of Canada Statistics. (Hidiroglou, Quenneville and Huot 2001). For the EMCDG, two approaches were considered.

3.2 Approach "micro": adjustments of microdonnées

In the "micro" approach, the objective is to assign with each unit sampled according to CTI'S a group of trade according to SIC'S then to add up the estimates. The challenge lies in reclassification for the period prior to 1998. Certain codes of the CTI have a simple bond (one by one) with a code of the SIC. An automated reclassification can then be carried out. For the codes with multiple bonds, a manual reclassification by specialists on the matter would have been ideal but impracticable, being given the time constraints and budget. Various methods of charge used at ends of reclassification were studied. For example, for the units which are still present on the basis of survey in 1998, we considered the possibility of charging the same codes of the SIC for the former years. This option is effective if no change of activity occurred since 1991. However, only 55 % of the units aimed to the beginning of the period of retropolation (that is to say 1991 or 1993) are present in January 1998. Moreover, the always alive units very often underwent important changes of structure questioning the assumption of stability of the principal activity. For the units which are not any more on the basis of survey in 1998, a code of the SIC can be allotted in a probabilistic way. For each code of the CTI, the probabilities of assignment to a code of the SIC are given in an empirical way, i.e. by the frequency of each bond CTI-SIC present in 1998 and the following years. Lastly, the assignment of a code of the SIC for the units classified under a code of the CTI with a bond "to several" can be done using a method of division. A certain percentage of the variable of interest is then re-coded in each code of the eligible SIC. The factors of division are also derived from the data for which classification is known under the two systems.

options pour la conversion de données de la CTI au SCIAN ont été étudiées pour plusieurs enquêtes de Statistique Canada. (Hidiroglou, Quenneville et Huot 2001). Pour l'EMCDG, deux approches ont été considérées.

3.2 Approche « micro » : ajustements des microdonnées

Dans l'approche « micro », l'objectif est d'assigner à chaque unité échantillonnée selon la CTI un groupe de commerce selon le SCIAN puis de totaliser les estimations. Le défi réside dans la reclassification pour la période antérieure à 1998. Certains codes de la CTI ont un lien simple (un à un) avec un code du SCIAN. Une reclassification automatisée peut alors être effectuée. Pour les codes avec liens multiples, une reclassification manuelle par des spécialistes en la matière aurait été idéale mais impraticable, étant donné les contraintes de temps et de budget. Diverses méthodes d'imputation utilisées à des fins de reclassification ont été étudiées. Par exemple, pour les unités qui sont encore présentes sur la base de sondage en 1998, nous avons considéré la possibilité d'imputer les mêmes codes du SCIAN pour les années antérieures. Cette option est efficace si aucun changement d'activité ne s'est produit depuis 1991. Toutefois, seulement 55 % des unités visées au début de la période de réropolation (soit 1991 ou 1993) sont présentes en janvier 1998. De plus, les unités toujours vivantes ont bien souvent subi des changements de structure importants mettant en doute l'hypothèse de stabilité de l'activité principale. Pour les unités qui ne sont plus sur la base de sondage en 1998, un code du SCIAN peut être attribué de manière probabiliste. Pour chaque code de la CTI, les probabilités d'assignation à un code du SCIAN sont déterminées de manière empirique, c'est-à-dire par la fréquence de chacun des liens CTI-SCIAN présents en 1998 et les années suivantes. Enfin, l'assignation d'un code du SCIAN pour les unités classées sous un code de la CTI avec un lien « un à plusieurs » peut se faire à l'aide d'une méthode de partage. Un certain pourcentage de la variable d'intérêt est alors recodé dans chacun des codes du SCIAN éligibles. Les facteurs de partage sont également dérivés des données pour lesquelles la

The principal advantage of the "micro" approach is on the level of the precision. The methods considered are however rather complex. Moreover, one supposes that microdonnées under the CTI agree perfectly with the aggregate data. It is not the case for the MWRTS. Inter alia, an important restratification took place in December 1997 and caused a jump in the historical series. This jump was smoothed gradually in the months former to the restratification using adjustments the "macro" use of the "micro" approach would thus require an additional treatment for the "macro" adjustments related to the restratification and for other "macro" adjustments tiny but considerable. We rejected this approach.

3.3 "macro" approach: coefficients of conversion

In the "macro" approach, the units do not need to be reclassified individually. For each lobby according to SIC'S, one rather uses a linear combination balanced of the total of each lobby according to CTI'S. The total $X_j(a, m)$ of the group of trade j according to SIC'S for the year a and the month m is given by

$$X_j(a, m) = \sum_i \alpha_{ij}(a, m) X_i(a, m)$$

where $X_i(a, m)$ is the sum of the group of trade i according to CTI'S. The weights of the linear combination, are the coefficients of conversion $\alpha_{ij}(a, m)$, represent the percentage of the total of the group CTI i which is allotted to group SIC j . They are estimated using the data for which double classification is known. For the MWRTS, these data are those of January 1998 and the following months.

The major advantage of this approach is that no additional treatment is necessary for the adjustments for aggregate levels (smoothing or

classification est connue sous les deux systèmes.

L'avantage principal de l'approche « micro » est au niveau de la précision. Les méthodes considérées sont cependant assez complexes. De plus, on suppose que les microdonnées sous la CTI concordent parfaitement aux données agrégées. Ce n'est pas le cas pour l'MWRTS. Entre autres, une restratification importante a eu lieu en décembre 1997 et a causé un saut dans les séries historiques. Ce saut a été lissé graduellement dans les mois antérieurs à la restratification à l'aide d'ajustements « macro » L'utilisation de l'approche « micro » nécessiterait donc un traitement supplémentaire pour les ajustements « macro » liés à la restratification et pour d'autres ajustements « macro » minimes mais non négligeables. Nous avons rejeté cette approche

3.3 Approche « macro » : coefficients de conversion

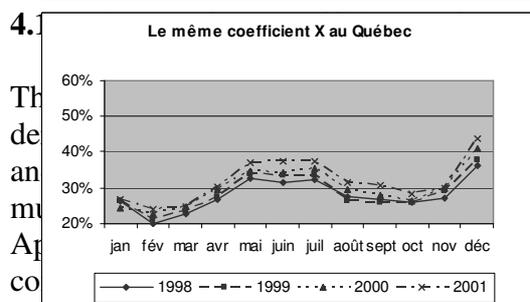
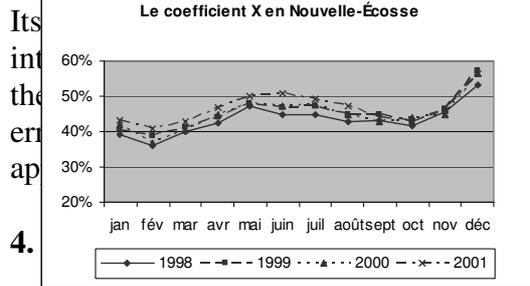
Dans l'approche « macro », les unités n'ont pas besoin d'être reclassifiées individuellement. Pour chaque groupe d'intérêt selon le SCIAN, on utilise plutôt une combinaison linéaire pondérée du total de chacun des groupes d'intérêt selon la CTI. Le total $X_j(a, m)$ du groupe de commerce j selon le SCIAN pour l'année a et le mois m est donné par

$$X_j(a, m) = \sum_i \alpha_{ij}(a, m) X_i(a, m)$$

où $X_i(a, m)$ est la somme du groupe de commerce i selon la CTI. Les poids de la combinaison linéaire, soit les coefficients de conversion $\alpha_{ij}(a, m)$, représentent le pourcentage du total du groupe CTI i qui est attribué au groupe SCIAN j . Ils sont estimés à l'aide des données pour lesquelles la double classification est connue. Pour l'MWRTS, ces données sont celles de janvier 1998 et des mois suivants.

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

others) carried out with the series under the CTI.



4. The de an mu Ap co of the CTI and the groups of the SIC, were studied using 48 observations (January 1998 at December 2001). The coefficients considered to be invalid by specialists on the matter or in absolute value lower than 0,3 % were eliminated and reallocated. The 230 series of nonnull coefficients remaining were analyzed graphically to detect the presence of regional differences, or outliers. Since the differences regional and were easily detectable for it (graph 2), it was decided to consider the coefficients monthly for the years 1991 at 1997 using the average of the coefficients calculated over the months homologous with the years 1998 to 2001 and this, for each area.

Graph 2 - Coefficients of conversion in time

conversion. Les sources d'erreurs sont détaillées dans la section 4.2. L'approche « macro » a été retenue pour l'MWRTS.

4. IMPLANTATION

4.1 Estimation des coefficients de conversion

Les valeurs des coefficients de conversion peuvent être dérivées des données échantillonnées pour janvier 1998 et les mois suivants. Les coefficients antérieurs doivent être estimés à l'aide des coefficients connus. Environ 1000 séries de coefficients de conversion, obtenues par le croisement des groupes de la CTI et des groupes du SCIAN, ont été étudiées à l'aide de 48 observations (janvier 1998 à décembre 2001). Les coefficients jugés invalides par des spécialistes en la matière ou en valeur absolue inférieurs à 0,3 % ont été éliminés et réalloués. Les 230 séries de coefficients non nuls restantes ont été analysées graphiquement pour détecter la présence de différences régionales, de saisonnalité ou de valeurs aberrantes. Puisque les différences régionales et la saisonnalité étaient facilement décelables (graphique 2), il a été décidé d'estimer les coefficients mensuels pour les années 1991 à 1997 à l'aide de la moyenne des coefficients calculés sur les mois homologues des années 1998 à 2001 et ce, pour chaque région.

Graphique 2 - Coefficients de conversion dans le temps

The coefficients of conversion can thus be written in the form

$$\hat{\alpha}_{ij}^r(1991, m) = \dots = \hat{\alpha}_{ij}^r(1997, m) = \frac{1}{k} \sum_{a=1998}^{2001} \delta_{ij}^r(a, m) \alpha_{ij}^r(a, m)$$

where the variable k is the sum over the 4 years of the indicatrices $\delta_{ij}^r(a, m)$ defined by

$$\delta_{ij}^r(a, m) = \begin{cases} 0 & \text{si } \alpha_{ij}^r(a, m) \text{ est jugé aberrant;} \\ 1 & \text{sinon.} \end{cases}$$

The outliers are thus withdrawn from the calculation of the average. The coefficients obtained are readjusted to sum to 100% for each combination of year, of month m , of area r and group of trade i according to CTIS.

4.2 Sources of errors

Various sources of errors affect the effectiveness of the approach "macro" (Hidioglou et al., 2001). A first source of potential error is the base of survey itself. An error of classification according to the CTI or SIC'S for one month given between 1998 and 2001 assigns obviously the month in question like every month homologous with the years 1991 to 1997. In order to reduce the impact of the badly classified units, the large contributors were manually checked and re-coded with the need. The corrections made to the base of survey since 1998 were indexed with an aim of evaluating their impact; when considered to be necessary, of the adjustments to the series estimated under the SIC were brought.

The second type of error comes from the use of the

Les coefficients de conversion peuvent donc s'écrire sous la forme

$$\hat{\alpha}_{ij}^r(1991, m) = \dots = \hat{\alpha}_{ij}^r(1997, m) = \frac{1}{k} \sum_{a=1998}^{2001} \delta_{ij}^r(a, m) \alpha_{ij}^r(a, m)$$

où la variable k est la somme sur les 4 années des indicatrices $\delta_{ij}^r(a, m)$ définies par

$$\delta_{ij}^r(a, m) = \begin{cases} 0 & \text{si } \alpha_{ij}^r(a, m) \text{ est jugé aberrant;} \\ 1 & \text{sinon.} \end{cases}$$

Les valeurs aberrantes sont ainsi retirées du calcul de la moyenne. Les coefficients obtenus sont réajustés pour sommer à 100% pour chaque combinaison d'année a , de mois m , de région r et de groupe de commerce i selon la CTI.

4.2 Sources d'erreurs

Différentes sources d'erreurs affectent l'efficacité de l'approche « macro » (Hidioglou et al. 2001). Une première source d'erreur potentielle 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.

Le second type d'erreur provient de l'utilisation des coefficients de conversion calculés sur des années récentes (1998-2001) pour estimer les coefficients de conversion des années antérieures.

coefficients of conversion calculated over recent years (1998-2001) to estimate the coefficients of conversion of the former years. This method is effective if the distribution according to SIC'S is stable from one year to another; if it is not it, we consider nevertheless who the risk of error is weaker in 1997 than in 1991. The assumption of stability was accepted in the majority of the cases. It is however necessary to proscribe the use of the coefficients of conversion only estimated by the average when an industry underwent an important change such as the centers of restoration on great surface. Since that the sector of the restoration knew a growth marked during last years and that the centers on great surfaces are recent in the Canadian economy, their contribution to the group of trade of the CTI in 1998 badly represents the contribution which they had at the beginning of the years 1990. One must thus adjust the coefficients of conversion which are associated to them the fall for the years 1991 to 1997. The centers of restoration represent a part of a group of trade of the CTI. When their contribution is revised to the fall, the contribution of the residual units must be revised with the rise. The value of the adjustment itself is based on the analysis of the experts and the results of a partial classification under the SIC at the micro level. This type of adjustment makes it possible to model the variations in the time of the coefficients. Other elements related to coefficients the such working days and the effect Easter were not considered being given the little of observations available to the moment of the analysis.

A source of additional error rises from the use of coefficients of conversion based on a variable and used on another. The MWRTS has two variables of interest; the sales and stocks - let us note that stocks are published at present only for the wholesalers according to CTI'S. All the analysis of the coefficients of conversion was carried out on the level of the sales. Instead of converting according to SIC'S the series of stocks according to CTI'S, we converted the retrofitted series of the sales into stocks using ratios. Ratios stocks/sales are calculated for each month since 1993 on the level of the groups of trade CTI and are applied to

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 important tels que les centres de rénovation à grande surface. Puisque que le secteur de la rénovation a connu une croissance marquée au cours des dernières années et que les centres à grandes surfaces sont récents 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. On doit donc ajuster les coefficients de conversion qui leur sont associés à la baisse pour les années 1991 à 1997. Les centres de rénovation représentent une partie d'un groupe de commerce de la CTI. Lorsque leur contribution est révisée à la baisse, la contribution des unités résiduelles doit être révisée à la hausse. 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 permet 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.

Une source d'erreur additionnelle découle de l'utilisation de coefficients de conversion basés sur une variable et utilisés sur une autre. L'MWRTS a deux variables d'intérêt ; les ventes et les stocks – notons que les stocks ne sont présentement publiés que pour les grossistes selon la CTI. 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 retrofitted 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

the groups of corresponding trade SIC. The correspondences are established according to the following rule: a group of trade according to SIC'S corresponds to a group of trade according to CTI'S if this last contributes to more than 99 % of the total of the group according to SIC'S. Seven groups of wholesale according to SIC'S have a group corresponding according to CTI'S. The groups of the SIC without group corresponding according to CTI'S are treated using a total ratio stocks/sales.

4.3 Continuity of the series under the SIC

The series under the classification of the SIC are divided into three parts. A first from January 1991 to December 1997 when the estimates are obtained using estimated coefficients of conversion. The second part starts in January 1998 and will finish at the time of the stop of the current investigation. In this part, the series under the SIC are obtained by estimate by field, in other words using coefficients of conversion observed. The third part begins at the same time as the altered investigation. The second and the third parts overlap few months when the old one and the new investigation is both in production (test in parallel). One envisages the presence of a breaking in the series at the time of the passage to the altered investigation. This jump is explained by the change of classification but also by other methodological changes. The results of the test in parallel will make it possible to adjust the level of the retrofitted series. One envisages a constant multiplicative adjustment in time to adjust the historical series at the levels published in the new investigation. By holding account of, of breakings in the series under the SIC were also observable in January 1998. One passes then from coefficients estimated in December 1997 at coefficients observed. To reduce the effect, all the retrofitted data of 1998 were recomputed by using the estimated coefficients, i.e. by calculating the four years average, including 1998. Let us note that the coefficients of 1998 more often differ from the average than the three other years and than the very aberrant coefficients were removed calculation of the average. By lengthening the first part of the series until December 1998, one cancels the presence of breaking between the first two parts.

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

4.3 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 terminera lors de l'arrêt de l'enquête actuelle. Dans cette partie, les séries sous le SCIAN sont obtenues par estimation par domaine, autrement dit à l'aide de coefficients de conversion *observés*. La troisième partie débute en même temps que l'enquête remaniée. La deuxième et la troisième parties se chevauchent de quelques mois lorsque l'ancienne et la nouvelle enquête seront toutes les deux en production (test en parallèle). On prévoit la présence d'un bris dans les séries lors du passage à l'enquête remaniée. Ce saut s'explique par le changement de classification mais aussi par d'autres changements méthodologiques. Les résultats du test en parallèle permettront d'ajuster le niveau des séries retrofitted. On prévoit un ajustement multiplicatif constant dans le temps pour ajuster les séries historiques aux niveaux publiés dans la nouvelle enquête.

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 retrofitted 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

5. FUTURE WORK

The first estimates under the SIC of the altered investigation are expected in 2004. The former historical data will be adjusted for the level using the results of the test in parallel. Thanks to the retrofitted data, the new estimates could de-seasonalised and be analyzed.

The SIC was first updated into 2002 which did not touch the sectors wholesale and Retail trade. One second revision, expected for 2007, will affect the wholesale. A retropolation of the data according to the new version of the SIC is expected.

THANKS

The author wishes to thank François Lavoie, Benoît Quenneville, Julie Trépanier and particularly Helene Bérard for their extremely useful comments and their suggestions.

RÉFÉRENCES

Bérard, H.(2001), The Redesign of the Monthly Wholesale and Retail Trade Survey of Statistics Canada, *Recueil 2001 de la section des méthodes d'enquêtes*, Société statistique du Canada, pp 81-86.

Hidiroglou, M., Quenneville, B. et Huot G. (2001), *Methodological Problem and Options for SIC-NAICS Conversion*, Document de travail de Statistique Canada, Division des méthodes d'enquêtes auprès des entreprises, octobre 2001.

Meyer, B.(2001), *Conversion de la CTI de 1980 au SIC; Commerce de gros et de détail; aperçu dans la perspective d'un secteur d'enquête*, Document de travail de Statistique Canada, Division de la statistique du commerce, septembre 2001.

Statistique Canada (2002), *Système de classification des industries de l'Amérique du*

premières parties.

5. ACTIVITÉS À VENIR

Les premières estimations sous le SCIAN de l'enquête remaniée sont prévues en 2004. Les données historiques antérieures seront ajustées pour le niveau à l'aide des résultats du test en parallèle. Grâce aux données retrofitted, les nouvelles estimations pourront être désaisonnalisées et analysées.

Le SCIAN a connu une première mise à jour en 2002 qui ne touchait pas les secteurs du commerce de gros et du commerce de détail. Une seconde révision, attendue en 2007, affectera le commerce de gros. Une rétopolation des données selon la nouvelle version du SCIAN est prévue.

REMERCIEMENTS

L'auteure désire remercier François Lavoie, Benoît Quenneville, Julie Trépanier et particulièrement Hélène Bérard pour leurs commentaires et leurs suggestions fort utiles.

RÉFÉRENCES

Bérard, H.(2001), The Redesign of the Monthly Wholesale and Retail Trade Survey of Statistics Canada, *Recueil 2001 de la section des méthodes d'enquêtes*, Société statistique du Canada, pp 81-86.

Hidiroglou, M., Quenneville, B. et Huot G. (2001), *Methodological Problem and Options for SIC-NAICS Conversion*, Document de travail de Statistique Canada, Division des méthodes d'enquêtes auprès des entreprises, octobre 2001.

Meyer, B.(2001), *Conversion de la CTI de 1980 au SCIAN; Commerce de gros et de détail; aperçu dans la perspective d'un secteur d'enquête*, Document de travail de Statistique Canada, Division de la statistique du commerce, septembre 2001.

Statistique Canada (2002), *Système de classification des industries de l'Amérique du*

N... X... C... 501-

Nord : SCIAN, Canada 2002, produit n° 12-501-XPF au catalogue de Statistique Canada, Ottawa, Ministre de l'industrie, avril 2002.

