

11P0019F no 59



Direction des études analytiques



Ans Years of
d'excellence Excellence



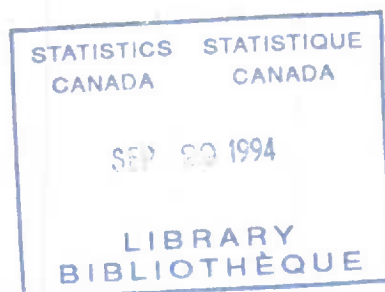
Documents de recherche



Statistique
Canada

Statistics
Canada

Canada



**MAINTENIR LA COMPARABILITÉ DANS LE TEMPS
DES CLASSIFICATIONS PAR INDUSTRIE**

par

John S. Crysdale

N° 59

Division de l'industrie
Secteur de la statistique
du commerce et des entreprises
Statistique Canada
1993

L'auteur assume seul la responsabilité des opinions formulées dans le présent document qui ne représentent pas nécessairement le point de vue de Statistique Canada.

Also available in English

Maintenir la comparabilité dans le temps des classifications par industrie

par John S. Crysdale, Statistique Canada

*[...] j'ai connu que notre nature n'était qu'un
continuel changement, et je n'ai plus changé depuis;
et si je changeais, je confirmerais mon opinion.*

Blaise Pascal, Pensées

Résumé

Les utilisateurs de données relatives aux établissements sont toujours aux prises avec le même problème : s'accoutumer aux changements qui surviennent dans la classification des industries. La solution la plus courante est de refondre la Classification type des industries (CTI). Cet article a pour but d'évaluer plusieurs méthodes automatisées par lesquelles Statistique Canada peut refaire cette classification. Parmi ces méthodes, on note 1) l'utilisation des données détaillées sur les produits, auxquelles on applique un ensemble de règles de résistance; 2) l'application d'une concordance d'un à un; et 3) une combinaison des deux premières. Pour les besoins de l'évaluation, on utilise chacune des méthodes pour reclasser chaque établissement manufacturier qui a déclaré des données sur les produits en 1982, puis on compare les résultats avec les codes officiels de cette année-là. En 1982, les séries de données officielles étaient classées et publiées selon la CAÉ de 1970 et la CTI de 1980. La méthode jugée la meilleure est celle qui reproduit le plus fidèlement les codes officiels de 1982; elle peut ensuite servir à reclasser les données d'autres années. La principale conclusion de notre étude est que la troisième méthode est supérieure aux deux autres, surtout lorsqu'elle redéfinit la classification selon la CAÉ de 1970.

Dans ce même article, nous examinons une méthode qui n'a pas rapport à la CTI. Cette méthode consiste à trouver des groupes d'industries équivalents dans la CAÉ de 1970 et la CTI de 1980 et à leur attribuer un identificateur numérique. Celui-ci peut ensuite servir à reclasser les données de n'importe quelle année, peu importe leur mode de classement initial. En épurant les données des liens interindustriels inhabituels ou incertains, on peut conserver des groupes petits et homogènes. Le principal inconvénient de cette méthode est de définir des industries qui ne sont pas aussi connues que celles de la CTI.

Mots-clés : Classification par industrie automatisée; Classification type des industries (CTI).

Remerciements

Cet article a été rédigé par suite d'une demande qu'avaient faite John Baldwin et Paul Gorecki, qui étaient chercheurs invités à la Direction des études analytiques, pour refondre la classification des industries. Par ses commentaires et ses conseils, Shaila Nijhowne, de la Division des normes, a contribué grandement à la réalisation de cette étude. Nos adressons aussi nos remerciements à Jack Bailey, Gérard Côté, Gil Elliot et Ken Young, tous de la Division des normes, à Katherine Blais, de la Division des petites entreprises et des enquêtes spéciales, à John S. McVey, du Secteur de la statistique du commerce et des entreprises, à Garnett Picot, du Groupe de l'analyse des entreprises et du marché du travail, et à Harley Potter, anciennement de la Division de l'industrie. D'autres remerciements vont à George Andrusiak (Division de l'industrie), Bruce Cooke (Division des mesures et analyse des industries), Brenda Hutchinson (Division de l'industrie) et Bruce Mitchell (Division des comptes nationaux et de l'environnement), pour avoir accepté de commenter au préalable la communication que nous devons présenter au colloque international sur les enquêtes-établissements, qui s'est tenu à Buffalo (New York). Toute erreur est notre responsabilité propre.

INTRODUCTION

Au Canada, la dernière révision de la classification des industries remonte à 1983, année où est entrée en vigueur la version 1980 de la CAÉ. Par cette révision, les 171 industries manufacturières de la CAÉ de 1970 (plus une industrie non manufacturière) devenaient 236 industries (plus trois industries non manufacturières) selon la CTI de 1980. Dans beaucoup de cas, la transition était simple : 79 industries de la CAÉ de 1970 ont été converties "d'une en une" et deux ont connu une conversion "de plusieurs en une". Cependant, il y a aussi beaucoup de cas où la transition était plus complexe : onze industries de la CAÉ de 1970 ont subi une conversion "d'une en plusieurs" et quatre-vingts autres ont connu une conversion "de plusieurs en plusieurs"; dans un cas en particulier, un groupe qui comptait 59 industries sur la base de 1970 en comptait 84 une fois la conversion effectuée.^{1 2}

¹ Pour plus de détails, voir «Notes sur les industries manufacturières dans la classification type des industries de 1980» dans *Industries manufacturières du Canada : niveaux national et provincial, 1983*, n° 31-203 au catalogue, pp. xxiii-xcviii.

² En ce qui concerne les industries qui font partie du secteur manufacturier selon une classification mais qui en sont exclues selon l'autre, notre analyse se limite au caractère manufacturier de ces industries; nous n'établissons pas de correspondance entre les industries qui ne font pas partie du secteur manufacturier selon l'une et l'autre classifications.

Le but de cet article est de comparer différentes méthodes - toutes automatisées - qui s'offrent à l'expérimentateur qui dispose de micro-données lisibles par machine pour composer avec les différences de classification et exprimer les données sur une base comparable. Nous traitons les établissements manufacturiers qui déclarent des données détaillées sur les produits³; en 1982, ces établissements constituaient 58.7% de l'ensemble des unités statistiques et représentaient 97.6% des livraisons manufacturières.⁴ 5

Trois méthodes fondamentales sont proposées ici pour obtenir des classifications comparables : 1) prolonger la CAÉ de 1970 en l'appliquant aux établissements qui sont classés à l'heure actuelle selon la CTI de 1980. Les expérimentateurs pourraient ainsi mettre à jour les calculs statistiques qui ont été entrepris selon la CAÉ de 1970. 2) Étendre rétrospectivement la CTI de 1980 en l'appliquant aux établissements qui sont classés à l'heure actuelle selon la CAÉ de 1970. On aurait ainsi une représentation plus juste de la structure industrielle courante. 3) Trouver des groupes d'industries équivalents dans l'une et l'autre classifications. Ces industries ne répondent ni à l'ancienne classification ni à la nouvelle, mais elles se rapprochent beaucoup de l'une et l'autre.

Nous utilisons trois méthodes pour prolonger les classifications de 1970 et de 1980. La première consiste à utiliser les données détaillées sur les produits, avec un ensemble de règles de résistance conçu pour empêcher les établissements de passer brusquement d'une industrie à l'autre; la deuxième méthode consiste à utiliser une concordance d'un à un optimale et la troisième est une combinaison des deux précédentes.

³ Par produits nous entendons ici les biens de propre fabrication et les services effectués sur des biens qui appartiennent à d'autres fabricants (travail donné à contrat et travail de réparation).

⁴ Livraisons manufacturières : la somme des livraisons de produits, redressée au niveau de l'établissement pour la défalcation des taxes de vente et des frais de transport (entre autres choses).

⁵ Dans la version de cet article qui a été publiée dans les Actes du colloque international sur les enquêtes-établissements, un établissement donné n'était pas réputé avoir déclaré des données sur les produits si le questionnaire portant sur ses activités avait été rempli par un établissement connexe et si les données des deux unités avaient été combinées et qu'elles n'avaient pas été redistribuées par le personnel spécialisé. En ce qui concerne la présente version de l'article, la redistribution des données a été effectuée au complet (par l'auteur) et les établissements correspondants ont été inclus dans le groupe qui est réputé déclarer des données détaillées sur les produits. (Dans l'opération de redistribution, on s'est servi de l'emploi manufacturier -- le seul poste existant sous une forme non combinée pour chacun de ces établissements -- pour répartir proportionnellement les données. Les résultats sont conformes au fait que, à l'intérieur de chaque combinaison, tous les établissements constituants exercent la même activité.) Ont été aussi inclus les établissements dont les données ont été entièrement estimées par l'organisme statistique. Par suite de ces ajouts, le pourcentage des établissements réputés produire des données détaillées sur les produits et qui, par conséquent, sont sujets à un reclassement est passé de 57 à 58.7% pour ce qui a trait aux unités statistiques et de 96 à 97.6% pour ce qui est des livraisons manufacturières.

Même lorsqu'on dispose de toutes les micro-données voulues, il existe plusieurs façons de reclasser des industries. Cette multiplicité s'explique par les aspects subjectifs de la classification qui sont traités dans la section suivante. La concordance d'un à un sous-entend les considérations subjectives contenues dans la série qui sert de point de départ au reclassement, alors que la méthode des données détaillées doit les expliciter.

Dans la première section de cet article, nous décrivons le processus de classement utilisé pour créer la série officielle. La deuxième section expose à grands traits trois méthodes pour prolonger des classifications d'industries. Dans la troisième section, nous évaluons ces méthodes en les appliquant à un reclassement des établissements manufacturiers qui ont déclaré des données détaillées sur les produits en 1982, puis en comparant les résultats à la série officielle pour cette année-là. En 1982, les données avaient été recueillies selon la CAÉ de 1970, mais avaient été publiées selon les deux classifications. Notre principale constatation est la suivante : au niveau à 4 chiffres, on obtient des résultats très proches de la série officielle en prolongeant la CAÉ de 1970 à l'aide d'une combinaison de méthodes. Dans la quatrième section, nous examinons une méthode qui n'a pas rapport aux classifications d'industries, soit l'agrégation. C'est une méthode qui est facile à appliquer mais son principal inconvénient est de définir des industries qui ne sont pas aussi reconnues que celles de la CTI.

I. CLASSIFICATION UTILISÉE POUR LA SÉRIE OFFICIELLE

Comme cet article porte principalement sur des méthodes qui visent à reproduire la classification officielle des établissements manufacturiers qui déclarent des données détaillées sur les produits, il est utile de voir comment s'opère le classement de la série officielle.

La classification des industries s'opère au niveau à 4 chiffres. Chaque industrie à 4 chiffres est définie selon le ou les produits qui sont réputés ressortir principalement à cette industrie.⁶ Au niveau de l'établissement, on effectue une classification provisoire en groupant des produits selon l'industrie à laquelle

⁶ Ces rapports sont décrits sommairement dans *Classification des activités économiques, révisée en 1970*, n° 12-501 au catalogue, hors série et dans *Classification type des industries, 1980*, n° 12-501F au catalogue, hors série. Ils sont exposés plus en détail dans les concordances «produits-activités économiques» : la concordance entre la Classification des produits industriels (CPI) et la Classification des activités économiques (CAÉ) de 1970 est publiée dans *Concepts et définitions du recensement des manufactures*, n° 31-528 au catalogue, hors série, 1979; la concordance entre la CPI et la Classification type des industries de 1980 se trouve au tableau C de *Industries manufacturières du Canada : niveaux national et provincial, 1983*, n° 31-203 au catalogue. Pendant la rédaction de cet article, un certain nombre de changements ont été apportés à la concordance. C'est pourquoi il peut exister des différences entre des codes pour 1982 calculés maintenant et des codes calculés antérieurement.

ils ressortissent principalement, puis en déterminant quel groupe représente la plus grosse part des livraisons de produits.⁷

Depuis l'année de référence 1982, cette opération est automatisée.⁸ Dans un deuxième temps, on compare le code obtenu à celui qu'a déjà l'établissement (le plus souvent, le code attribué l'année précédente, ou, dans le cas de nouveaux établissements, un code qui dépend de la nature de l'entreprise). Si la comparaison révèle que l'établissement en question devrait être transféré à une autre industrie, un rapport est imprimé en vue d'une vérification manuelle. On peut devoir modifier parfois les codes de produits ou la valeur des livraisons. Si la vérification manuelle confirme la nécessité d'un transfert, un certain nombre de considérations subjectives entrent en ligne de compte pour déterminer si le transfert se fera immédiatement.

Les règles de résistance sont un exemple de considération subjective. Ces règles ont pour but d'empêcher les transferts d'établissement injustifiés, c'est-à-dire les transferts qui découleraient d'un faible changement dans les proportions, à moins que ce changement ne soit considéré comme permanent. Les transferts occasionnés par de faibles changements ont un effet démesuré sur les groupes d'industries. Par exemple, si un établissement dont le chiffre des livraisons est \$100 est transféré à une autre industrie à cause d'une variation de \$1 dans sa production, l'industrie "donneuse" perdra une valeur égale à 100 fois cette variation de \$1 tandis que l'industrie "receveuse" gagnera une valeur équivalente. Si le changement n'est que temporaire et que l'établissement est "retourné" dans l'industrie d'origine, l'effet se fera sentir une seconde fois. Il est possible d'éviter ces transferts injustifiés grâce à une connaissance approfondie de la conjoncture des industries. Cependant, il n'existe pas de règles explicites à cet égard.

La question de la couverture est un autre exemple de considération subjective. Il arrive qu'un établissement soit classé dans une industrie dont le produit principal ne représente pas la plus grande part des livraisons de l'établissement : en effet, l'établissement occupe une place tellement importante dans l'industrie que s'il en était exclu, on perdrait une grande partie des renseignements concernant cette industrie. Ce genre de situation est plus susceptible de se produire lorsque l'industrie dont le produit principal représente la plus grande part des livraisons de l'établissement englobe aussi des activités non déterminées et que l'activité de l'établissement en question ne

⁷ On se fonde sur la valeur des livraisons car il n'est pas toujours possible de calculer la valeur ajoutée pour le produit.

⁸ Voir J.S. Crysdale, «La classification par industrie dans le recensement canadien des manufactures : vérification automatisée à l'aide des données sur les produits», document de recherche n° 20, Direction des études analytiques, Statistique Canada.

peut être répartie artificiellement entre les diverses industries concernées.

Le classement peut aussi être influencé par des considérations de confidentialité. Par exemple, si le fait de transférer un gros établissement à une petite industrie stable a pour conséquence de dévoiler les données confidentielles de cet établissement, il vaut mieux retarder le transfert jusqu'à ce que les données de cette industrie soient publiées.

La taille de l'établissement est un facteur qui peut aussi entrer en ligne de compte dans l'évaluation de la question. Ainsi, on peut choisir de retarder le transfert d'un établissement si on juge que celui-ci a un effet négligeable sur le total de l'industrie et surtout si ce transfert risque de retarder la publication des données.

Bref, la classification officielle des établissements manufacturiers qui fournissent des données détaillées sur les produits dépend de règles objectives et de considérations subjectives.

II. PROLONGEMENT DE LA CLASSIFICATION DES INDUSTRIES : ANALYSE GÉNÉRALE

Dans cette section, nous abordons de façon générale les trois méthodes destinées à prolonger la classification des industries.

Méthode n° 1 : Codage détaillé par produit

Cette méthode suppose que l'on parte des micro-données pour déterminer le code des industries. Elle ressemble beaucoup à la méthode utilisée pour produire la série officielle, à deux différences près.

La première différence concerne le traitement des produits qui sont classés à un niveau trop global pour être considérés comme le produit principal d'une seule industrie à 4 chiffres. Par exemple, les services exécutés sur des biens qui appartiennent à d'autres fabricants (travail donné à contrat et réparations) font partie de classes qui ne sont pas suffisamment détaillées. En 1982, les produits classés à un niveau trop peu détaillé représentaient 5% des livraisons des établissements manufacturiers qui déclaraient des données détaillées sur les produits.⁹ Selon la méthode de classification officielle, une activité comme celle mentionnée ci-dessus est soit définie comme l'activité principale de l'industrie

⁹ 5.0% dans le cas d'une classification selon la CAÉ de 1970 et 5.1% dans le cas d'une classification selon la CTI de 1980.

dans laquelle se trouve l'établissement déclarant ou n'est associée à aucune industrie en particulier. Ce procédé suppose qu'il existe déjà un code d'industrie ou qu'une intervention manuelle soit possible. Selon la méthode automatisée que l'on utilise ici, les produits mentionnés ci-dessus sont soit définis comme le produit principal de l'industrie dans laquelle l'établissement déclarant est classé par une conversion "d'une en une", soit définis comme le produit principal de l'industrie à laquelle appartenait l'établissement déclarant l'année précédente.^{10 11}

La seconde différence a trait aux facteurs subjectifs dont il a été question dans la section précédente. Seules les règles de résistance sont décrites explicitement ici. Elles ont été codifiées de manière que l'opération de classement puisse être entièrement automatisée.

D'une façon générale, les règles utilisées sont les suivantes : 1) si on observe un changement *significatif* pour un établissement, celui-ci est transféré sur-le-champ; 2) sinon, le transfert sera effectué lorsque le changement sera considéré comme *permanent*.

Pour mesurer le changement, on soustrait de 100 la valeur calculée par la formule :

¹⁰ Voici le déroulement de la méthode :

Étape n° 1 : Si l'établissement déclarant fait partie, selon la classification initiale, d'une industrie qui peut être classée sous un seul et même code dans la nouvelle classification (avec moins de 3% d'erreur), le produit sera défini comme l'élément principal de cette industrie. (Le seuil de 3% correspond au seuil utilisé dans la méthode hybride, de sorte que la caractérisation des paires d'industries soit la même dans la concordance "d'une à une".)

Étape n° 2 : Si la condition initiale n'est pas respectée et que le reclassement se fait selon la CAÉ de 1970, le produit sera défini comme l'élément principal de l'industrie (selon la CAÉ de 1970) à laquelle appartenait l'établissement l'année précédente. Si le reclassement se fait selon la CTI de 1980, le produit sera défini comme l'élément principal de l'industrie (selon la CTI de 1980) dans laquelle avait été classé l'établissement en 1983.

Étape n° 3 : Si aucune des conditions préalables ci-dessus n'est respectée, le produit sera défini comme l'élément principal de l'industrie déterminée à l'étape 1 même si le taux d'erreur dépasse les 3% pour la conversion.

En 1982, 29.7% des établissements déclaraient le produit principal parmi leurs produits si le reclassement se faisait selon la CAÉ de 1970, contre 30.4% si le reclassement se faisait selon la CTI de 1980. Dans l'un et l'autre cas, les établissements étaient responsables de 32.5% des livraisons. Les produits classés à un niveau trop peu détaillé constituaient la production entière de 10.6% des établissements qui déclarent des données détaillées sur les produits selon la CAÉ de 1970 (10.8% des établissements selon la CTI de 1980); dans l'un et l'autre cas, ces établissements étaient responsables de 1.7% des livraisons. Dans le cas d'un reclassement selon la CAÉ de 1970, 89.9% des produits étaient traités à l'étape 1, 9.4% à l'étape 2 et 0.7% à l'étape 3. Dans le cas d'un reclassement selon la CTI de 1980, les pourcentages correspondants étaient 39.5, 58.0 et 2.5.

¹¹ Les produits classés à un niveau trop peu détaillé ne comprennent pas les produits qui permettent d'identifier une industrie dont la définition comporte la notion de procédé mais qui ne laissent pas voir un tel aspect : 1) l'intégration verticale dans certaines des industries des pâtes et papiers, 2) la production conjointe dans les industries combinées de l'imprimerie et de l'édition. Ces groupes d'industries étaient responsables de 5.6% des livraisons totales de 1982 selon la CAÉ de 1970 et de 5.5% des livraisons totales selon la CTI de 1980. Comme pour la série officielle, on commence par vérifier l'existence de certains intrants (dans le cas de l'intégration verticale) ou de certains extrants (dans le cas de la production conjointe).

Valeur des livraisons de l'année courante qui ressortissent principalement à l'industrie dans laquelle a été classé l'établissement l'année précédente

x 100

Valeur des livraisons de l'année courante qui ressortissent principalement à l'industrie à laquelle correspond le gros de l'activité de l'établissement pour l'année courante

Cette formule donne des valeurs de 0 à 100. Plus la valeur est élevée, plus le changement est important. Le changement est jugé *significatif* si la valeur calculée est égale ou supérieure à 67. Le même seuil vaut pour les deux classifications (celles de 1970 et de 1980). En outre, le changement est qualifié de *permanent* (même s'il n'est pas significatif) si le code d'industrie déterminé pour l'établissement demeure le même pendant deux années consécutives; dans ce cas, le transfert est effectué la seconde année.^{12 13}

Tableau 1 : Effet des règles de résistance, 1982
Pondéré par les livraisons de produits manufacturés

Reclassement selon:	CAÉ	CTI
	de 1970	de 1980
Code de l'industrie principale inchangé	93.5	94.9
Code de l'industrie principale modifié, test		
retarder le transfert (< 67)	0.6	0.6
transférer immédiatement (≥ 67)	0.5	0.8
Changement permanent, transférer	2.5	1.1
Nouvel établissement, attribuer le code de l'industrie principale	3.0	2.6
Total	100.0	100.0

Afin de raccorder les codes obtenus de cette manière avec ceux des séries officielles des autres années, on effectue *rétrospectivement* le reclassement des industries selon la CTI de 1980; pour les mêmes

¹² Une des conséquences de ces règles est qu'un établissement peut conserver son code d'industrie initial dans le reclassement malgré des changements mineurs continus dans la production, pourvu que ces changements ne touchent pas la même industrie dans deux années consécutives. Autrement dit, s'il se produit un changement dans l'industrie qui correspond à la plus grande part des livraisons de l'établissement mais que ce changement n'est pas significatif, le code d'industrie de l'établissement demeurera le même, et si, l'année suivante, c'est une tout autre industrie qui correspond à la plus grande part des livraisons de l'établissement et que le changement n'est pas significatif non plus, le code initial sera conservé.

¹³ Lorsqu'un établissement ne peut fournir de données sur les produits pendant deux années consécutives, le transfert est immédiat.

raisons, on effectue *prospectivement* le reclassement des industries selon la CAÉ de 1970. Cela signifie que dans l'application des règles de résistance (et dans le traitement des produits classés à un niveau trop peu détaillé), le terme *année précédente* doit être compris par rapport au processus de reclassement; il ne s'agit pas nécessairement de l'année civile.

Pour montrer l'effet de ces règles de résistance, nous allons calculer plus loin des taux d'erreur selon deux scénarios : sans l'application des règles de résistance et avec l'application des règles. Le premier scénario correspond au mode de classification automatisé, sauf en ce qui concerne le traitement des produits classés à un niveau trop peu détaillé.

Méthode n° 2 : Codage d'un à un optimal

Selon cette méthode, on part de codes existants en se servant de tableaux de codes d'industrie (voir annexes A et B) qui associent chaque élément de la CAÉ de 1970 à un seul élément de la CTI de 1980, et chaque élément de la CTI de 1980 à un seul élément de la CAÉ de 1970.¹⁴ À titre d'exemple, la classe 271 de la CAÉ de 1970 (Usines de pâtes et papiers), qui forme dans la CTI de 1980 les classes 2711 (Industrie des pâtes à papier), 2712 (Industrie du papier journal), 2713 (Industrie du carton), 2714 (Industrie des panneaux de construction) et 2719 (Autres industries du papier), est associée uniquement à la classe 2712 de la CTI de 1980 (celle qui représente la majeure partie de la valeur ajoutée de 2710 dans

¹⁴ Nous avons construit chaque concordance en appliquant la CAÉ de 1970 et la CTI de 1980 à la série officielle de 1982, puis en comparant la composition (en établissements) de chaque industrie définie selon la CAÉ de 1970 avec celle de chaque industrie définie selon la CTI de 1980. Nous avons pu ainsi mettre par couple les industries qui avaient la même composition en établissements. Nous avons introduit des pondérations afin de tenir compte du degré de chevauchement des couples d'industries. Ainsi, lorsqu'il y a une conversion "d'une en plusieurs", on ne retient que le couple d'industries qui représente la plus grande part de l'industrie de la classification de départ. Deux autres méthodes s'offrent à nous : 1) l'agrégation (traitée dans la section IV de cet article) et 2) la répartition proportionnelle. Celle-ci consiste à répartir les données de tous les établissements d'une industrie de la classification de départ entre les industries correspondantes de la classification d'arrivée selon (par exemple) la valeur des livraisons des données recueillies de 1982. Cela implique que l'on fractionne les données sur les établissements. Puisque cette méthode suppose l'énoncé d'hypothèses particulières concernant les proportions d'intrants et que les données reclassées sont destinées à l'analyse au niveau de l'établissement, nous ne nous y attacherons pas. On trouvera des exemples de concordances basées sur l'établissement dans les tableaux A et B de *Industries manufacturières du Canada : niveaux national et provincial, 1983*.

On peut aussi établir des concordances basées sur le produit (plutôt que sur un ensemble de produits comme l'établissement). Cela peut se faire en comparant les produits qui sont les plus importants pour chaque industrie codée selon la CAÉ de 1970 avec ceux qui sont les plus importants pour chaque industrie codée selon la CTI de 1980. On obtient ainsi une liste de couples d'industries qui ont des produits définitoires en commun. Cette liste peut faire l'objet d'une pondération de manière à tenir compte des livraisons réelles de produits, ou peut demeurer telle quelle. Un exemple de concordance non pondérée basée sur le produit est celle que supposent conjointement la concordance entre la CPI et la CAÉ de 1970 et la concordance entre la CPI et la CTI de 1980; un autre exemple de concordance non pondérée est celle que l'on trouve dans *Classification type des industries, 1980*. La concordance basée sur l'établissement et celle basée sur le produit peuvent donner des résultats différents; par exemple, un couple d'industries quelconque peut figurer dans une concordance basée sur l'établissement mais être absent d'une concordance (pondérée ou non) basée sur le produit. Cette situation peut se produire lorsque par suite d'une redistribution et d'un regroupement des produits rendus nécessaires par une révision de la classification, la principale activité d'un établissement ne correspond plus à la même industrie. Un autre exemple de divergence entre la concordance basée sur l'établissement et celle basée sur le produit : un couple d'industries qui, théoriquement, peut figurer dans une concordance (non pondérée) basée sur le produit peut ne correspondre à rien sur le plan des données, soit parce qu'il n'existe rien sur les livraisons de produits correspondantes (c.-à-d. que le couple d'industries ne figure pas dans une concordance pondérée basée sur le produit), soit parce que le couple en question n'est pas né d'un regroupement comme celui mentionné dans le premier exemple (c.-à-d. que le couple ne figure pas dans une concordance basée sur l'établissement).

les données recoupées de 1982¹⁵). Par conséquent, tous les établissements contenus dans la classe 271 de la CAÉ de 1970 seront transférés dans la classe 2712 de la CTI de 1980; aucun ne sera transféré dans les classes 2711, 2713, 2714 ou 2719.

Le codage d'un à un optimal est peut-être la manière la plus simple d'effectuer un reclassement au niveau à 4 chiffres. On n'a pas besoin des données détaillées sur les produits. Un expérimentateur qui a entre les mains la liste des établissements qui constituent une industrie peut reclasser tous ces établissements. De fait, l'opération ne doit pas nécessairement porter sur l'établissement comme tel; elle peut se faire à l'aide de données globales publiées. Cette méthode de reclassement a aussi l'avantage de refléter les décisions subjectives qui sous-tendent la série officielle. Par exemple, elle dénote l'application de règles de résistance sans qu'il soit nécessaire de formuler ces règles. En revanche, une des faiblesses de cette méthode est que, à proprement parler, la concordance basée sur des données ne concerne que l'année de laquelle elle est issue, quoiqu'on ne soit pas porté à utiliser cette concordance pour reclasser les données de cette même année. Et même si c'était le cas, l'application de cette méthode peut produire des erreurs de classement (comme les deux autres méthodes d'ailleurs).

Méthode n° 3 : Combinaison de méthodes

Il s'agit d'une combinaison du codage d'un à un optimal et du codage détaillé par produit (avec règles de résistance). Cette méthode réunit les avantages des deux précédentes, d'une part en reproduisant des considérations d'ordre subjectif et d'autre part en reflétant la pratique courante.

Pour une industrie donnée de la classification de départ, on choisira de recourir au codage détaillé par produit ou au codage d'un à un optimal selon qu'il est difficile ou non d'associer, avec un niveau d'erreur maximum donné (calculé comme un pourcentage du total des livraisons de l'industrie), cette industrie à une seule classe de la classification d'arrivée. Si l'association est facile à faire, on recourra au codage d'un à un optimal; sinon, on optera pour le codage détaillé par produit.

Dans la présente étude, le taux d'erreur maximum est de 3%.¹⁶ Cette valeur a été déterminée par suite d'une expérimentation.

¹⁵ Lorsque l'achat de services de sous-traitance n'est pas déduit de la valeur ajoutée totale, comme c'est le cas pour les établissements qui remplissent le questionnaire détaillé, on parle, à juste titre, de valeur ajoutée totale recensée.

¹⁶ Les annexes A et B font la distinction entre les industries qui peuvent être transposées avec moins de 3% d'erreur et celles qui ne le peuvent pas.

Dans le cas d'un reclassement selon la CAÉ de 1970 avec la méthode n° 3, le codage d'un à un optimal recouvre 92.0% des livraisons des industries visées. Dans le cas d'un reclassement selon la CTI de 1980, la proportion correspondante est de 52.3%.

III. PROLONGEMENT DE LA CLASSIFICATION DES INDUSTRIES : ÉTUDE EMPIRIQUE

Nous allons évaluer les trois méthodes décrites ci-dessus en les utilisant à tour de rôle pour classer tous les établissements qui ont déclaré des données détaillées sur les produits en 1982. Nous déterminons tout d'abord des codes selon la CAÉ de 1970 et selon la CTI de 1980. Nous comparons ensuite ces codes aux codes officiels de 1982, qui existent dans les deux versions (1970 et 1980).¹⁷ Les codes officiels sont considérés comme exacts. La méthode qui reproduit le plus fidèlement la série officielle de 1982 sera jugée la meilleure. On pourra ensuite l'utiliser pour prolonger la classification des industries.

Calcul du taux d'erreur

Nous désignerons le taux d'erreur calculé ici comme le pourcentage de classement erroné. Sa valeur varie de zéro à cent et la façon de le calculer est la suivante :

$$\frac{\text{Inclusion erronée} + \text{exclusion erronée}}{\text{Livraisons totales selon la série officielle} + \text{livraisons totales selon la méthode étudiée}} \times 100$$

L'inclusion erronée correspond à la valeur des livraisons des établissements qui ont été classés par erreur dans une industrie donnée au moyen de la méthode étudiée; l'exclusion erronée correspond à la valeur des livraisons des établissements qui ont été exclus par erreur de cette industrie.

Afin d'illustrer le calcul du taux d'erreur, supposons que les établissements qui constituent officiellement une industrie déclarent des livraisons totales de \$100 et que les établissements

¹⁷ Pour les besoins de cet article, les établissements qui avaient déclaré des données détaillées sur les produits en 1982 et qui, en 1983, appartenaient à la classe 3721 de la CTI de 1980 (Industrie des engrais chimiques et des matières pour engrais), laquelle classe n'était pas représentée dans la série officielle avant 1983, portent le code de 1983 dans la série qui est reproduite. Les livraisons de ce groupe d'établissements représentent moins de 1% des livraisons manufacturières pour 1982. Les établissements qui appartenaient à la classe 893 de la CAÉ de 1970 (Photographie, n.c.a.) et qui ont été transférés dans la classe 2821 de la CTI de 1980 (Industrie du clichage, de la composition et de la reliure) en 1983 ne figurent pas dans la base de sondage de 1982 et ne sont donc pas inclus dans la série qui est reproduite. On ne peut identifier ces établissements que de façon imparfaite, par exemple en notant les établissements qui faisaient partie de la classe 2821 en 1983 et pour lesquels il n'existait pas de données pour 1982 (cet algorithme est imparfait car il peut aussi bien identifier les établissements qui venaient d'être créés en 1983); les livraisons de ce groupe d'établissements pour 1983 équivalaient à moins de 0.5% des livraisons manufacturières pour 1982.

qui sont classés dans cette même industrie au moyen de la méthode étudiée déclarent des livraisons totales de \$110. De plus, supposons que les livraisons des établissements classés par erreur dans cette industrie se chiffrent à \$40 et que celles des établissements exclus par erreur de cette industrie totalisent \$30. Dans ces conditions, le pourcentage de classement erroné est de 33.3, c.-à-d. $((40 + 30)/(100 + 110)) \times 100$.

Une autre façon de calculer le taux d'erreur est de comparer le chiffre des livraisons des établissements qui constituent une industrie selon la série officielle avec le chiffre des livraisons des établissements qui sont classés dans cette même industrie selon la méthode étudiée (en l'occurrence, \$100 et \$110 respectivement). On aurait donc un taux d'erreur de 10%. Or, cette manière de comparer des totaux nous fait oublier que ces chiffres sont la somme de totaux d'établissements. Elle peut donc produire des résultats trompeurs. Par exemple, si le total des livraisons des établissements classés dans une industrie selon la méthode étudiée avait été de \$100 au lieu de \$110 et si les livraisons des établissements classés par erreur dans cette industrie s'étaient chiffrées à \$100 et que celles des établissements exclus par erreur de cette industrie avaient aussi totalisé \$100, on aurait obtenu un taux d'erreur nul. C'est pourquoi nous n'insistons pas davantage sur cette seconde méthode.

Comme les utilisateurs de données s'intéressent souvent aux niveaux à 2 et à 3 chiffres, nous évaluons aussi les méthodes de classement pour ces niveaux en nous servant du pourcentage de classement erroné. À cette fin, nous comparons les trois (ou deux) premiers caractères du code à 4 chiffres produit par la méthode étudiée avec les caractères correspondants du code à 4 chiffres de la classification officielle.

Résultats

Le tableau 2 donne le pourcentage de classement erroné pour les niveaux à 2, à 3 et à 4 chiffres; ces pourcentages sont des moyennes calculées pour l'ensemble des industries manufacturières, pondérées en fonction des livraisons (voir l'annexe D pour les taux d'erreur moyens pour le niveau à 2 chiffres).

**Tableau 2 : Pourcentage de classement erroné, 1982
Ensemble des industries manufacturières**

Reclassement selon	CAÉ de 1970	CTI de 1980
Niveau à 4 chiffres		
Codage détaillé par produit (sans règle de résistance)	2.8	2.8
Codage détaillé par produit (avec règles de résistance)	2.5	2.3
Codage d'un à un optimal	1.7	25.7
Combinaison de méthodes	0.8	1.6
Niveau à 3 chiffres		
Codage détaillé par produit (sans règle de résistance)	2.5	2.2
Codage détaillé par produit (avec règles de résistance)	2.3	1.7
Codage d'un à un optimal	1.1	2.9
Combinaison de méthodes	0.8	1.1
Niveau à 2 chiffres		
Codage détaillé par produit (sans règle de résistance)	1.4	1.1
Codage détaillé par produit (avec règles de résistance)	1.3	0.8
Codage d'un à un optimal	0.4	0.4
Combinaison de méthodes	0.5	0.5

Nous voyons d'après ce tableau que la combinaison de méthodes donne les meilleurs résultats. En effet, cette solution surclasse les autres méthodes pour les niveaux à 3 et à 4 chiffres, quel que soit le mode de reclassement. L'introduction de règles de résistance dans la méthode du codage détaillé par produit a pour effet de réduire le taux d'erreur.¹⁸

¹⁸ En ce qui concerne cette méthode de reclassement, les écarts qui pourraient subsister par rapport à la série officielle sont dus à quatre facteurs possibles : 1) différences dans le traitement des produits classés à un niveau trop peu détaillé, 2) différences dans le traitement des facteurs subjectifs, 3) changements dans les concordances entre les produits et les industries et 4) erreurs dans la série officielle.

On constate que toutes les méthodes donnent de meilleurs résultats aux niveaux d'agrégation plus élevés. C'est particulièrement le cas de la méthode du codage d'un à un optimal, dont l'efficacité progresse très sensiblement entre les niveaux à 4 et à 2 chiffres (ce qui dénote que la plupart des erreurs, pour cette méthode, sont associées aux industries à 2 et à 3 chiffres). Au niveau à 2 chiffres, le codage d'un à un optimal est supérieur à la combinaison de méthodes.

En ce qui concerne le reclassement selon la CTI de 1980, le codage d'un à un optimal donne des résultats médiocres au niveau à 4 chiffres. Le taux d'erreur élevé s'explique par les 82 classes vides de la CTI (par opposition aux 15 de la CAÉ de 1970) ainsi que tous les cas d'inclusion erronée auxquels correspondent autant de cas d'exclusion erronée. S'il y a tant de classes vides dans la classification d'arrivée, c'est que nous avons voulu garder telles quelles des industries de la classification de départ qui, en fait, devaient être fractionnées.

IV. UNE MÉTHODE QUI N'A PAS RAPPORT À LA CLASSIFICATION DES INDUSTRIES : AGRÉGATION

Les deux grandes méthodes que nous avons étudiées dans cet article consistaient à prolonger l'ancienne classification et à étendre rétrospectivement la nouvelle classification. Une troisième approche consisterait à créer une toute nouvelle classification en trouvant des groupes d'industries complètes dans la CAÉ de 1970 et la CTI de 1980 qui auraient une composition en établissements équivalente. On numéroterait ensuite ces groupes pour obtenir une concordance d'agrégation. On pourrait ainsi reclasser n'importe quel établissement de la CAÉ de 1970 ou de la CTI de 1980 en attribuant à sa classe le code du nouveau groupe.

Cette méthode du *regroupement* présente tous les avantages du codage d'un à un optimal. En outre, elle n'engendre aucune erreur de classement si la concordance sert au reclassement d'établissements dans l'année d'où elle tire son origine.

En revanche, la méthode présente trois inconvénients : 1) les nouvelles classes ne sont pas aussi connues que celles de la CTI; 2) il n'y a pas de structure hiérarchique simple et 3) il y a une perte de détail : les 172 classes de la CAÉ de 1970 et les 239 de la CTI de 1980 (mentionnées dans l'introduction) se réduisent à 97 groupes seulement, dont un comprend à lui seul 59 industries de la CAÉ de 1970 et 84 industries de la CTI de 1980.¹⁹

¹⁹ La concordance obtenue en utilisant les données recoupées officielles de 1982 et en trouvant des groupes dont la composition est parfaitement équivalente compte 97 groupes répartis ainsi : 79 conversions "d'une en une", 1 conversion "de plusieurs en une", 11 conversions "d'une en plusieurs" et 6 conversions "de plusieurs en plusieurs".

Cette perte de détail s'explique, en partie du moins, par le fait que les groupes sont constitués à partir de données recoupées réelles. Cela veut dire que des industries peuvent s'ajouter à un groupe donné si des conditions de production exceptionnelles sont observées ou s'il se produit des erreurs de classement. On peut limiter l'élargissement des groupes en faisant abstraction, dans les données pertinentes, des liens interindustriels inhabituels ou incertains. Dans cet article, l'existence de liens de ce genre est reconnue lorsque la production commune de deux industries représente moins de 15% de la valeur ajoutée de chaque industrie. Grâce à l'élimination de ces liens, on peut obtenir une concordance beaucoup plus détaillée (voir l'annexe C) : 147 groupes d'industries, aucune industrie de la CTI absente, et aucun groupe de taille démesurée. Cependant, la suppression de liens interindustriels peut avoir pour conséquence d'introduire des erreurs dans la nouvelle classification. L'erreur équivaut à la valeur des livraisons des établissements dont l'activité commune correspond aux liens réputés inhabituels ou incertains; cette erreur représente moins de un demi de 1% de la valeur totale des livraisons des industries manufacturières.

On trouve une concordance semblable dans les tableaux d'entrées-sorties du système de comptabilité nationale du Canada.²⁰ Les groupes d'industries, appelés industries de niveau commun ou de niveau de raccordement historique, mettent en relation les industries des classifications de 1960, 1970 et 1980. Toutefois, ce n'est pas une véritable concordance d'agrégation (au sens où nous l'entendons dans cet article) puisque les groupes ne comprennent pas toujours des industries entières (au sens de la CTI). Dans plusieurs cas, les industries de la CTI peuvent être associées à plus d'une industrie de niveau commun. C'est pourquoi l'opération de reclassement ne se réduit pas toujours à attribuer un nouveau code à une industrie de la CTI.

CONCLUSIONS

Après avoir expérimenté trois méthodes permettant de prolonger la classification des industries, nous avons constaté que la combinaison de la méthode du codage détaillé par produit et de la méthode du codage d'un à un optimal donnait de meilleurs résultats que chacune des autres méthodes prise individuellement. La version combinée produisait des résultats légèrement meilleurs lorsqu'elle servait à prolonger la CAÉ de 1970 que lorsqu'elle servait à étendre rétrospectivement la CTI de 1980.

²⁰ Voir Statistique Canada, *La structure par entrées-sorties de l'économie canadienne, 1961-1981 (données révisées)*, n° 15-510 au catalogue, Ottawa, 1987.

Le prolongement d'une classification des industries se heurte à plusieurs petits obstacles. Le premier est le fait qu'un certain nombre d'industries de la CAÉ de 1970 ont changé de définition pendant que cette classification était en vigueur. Cela a causé des solutions de continuité dans la série officielle qui n'ont rien à voir avec le reclassement dont il est question dans cet article.²¹ Une manière d'éliminer ces solutions de continuité est de reclasser les données pertinentes selon la version de 1982 de la CAÉ de 1970. Une deuxième difficulté est que la définition de la fabrication et, par conséquent, de la composition des industries manufacturières a changé avec l'adoption de la CTI de 1980. Ce n'était qu'un changement mineur cependant : moins de 0.5% des industries manufacturières définies selon la CAÉ de 1970 étaient supprimées et moins de 0.5% des industries manufacturières de la CTI de 1980 étaient nouvelles. Une troisième difficulté est la nécessité de raccorder la nouvelle classification des produits, qui est une extension du Système harmonisé de désignation et de codification des marchandises, à la CAÉ de 1970 avant de prolonger cette classification au-delà de 1987.

Par ailleurs, un certain nombre de modifications pourront faciliter dans l'avenir les opérations de reclassement. Premièrement, on devrait codifier les règles de résistance utilisées dans la série officielle. Deuxièmement, on devrait aussi codifier tous les autres facteurs subjectifs tels que le champ d'observation et l'importance de la taille. Troisièmement, on devrait adopter une classification des services manufacturiers qui soit suffisamment détaillée pour que l'on puisse raccorder ces services aux industries à 4 chiffres.

Une autre façon, simple et très précise, d'assurer la comparabilité dans le temps des classifications d'industries est d'utiliser une concordance d'agrégation. En supprimant les liens inter-industriels inhabituels ou incertains dans les données pertinentes, on fait en sorte que les groupes demeurent petits et homogènes. Le principal inconvénient de cette méthode est que les industries ainsi formées ne sont pas aussi connues que celles de la CTI.

En résumé, il est possible de présenter les données des vingt dernières années sur les industries manufacturières selon des classifications comparables en utilisant une méthode hybride pour prolonger la classification d'industries ou en appliquant la méthode d'agrégation décrite plus haut.

²¹ Voir *Industries manufacturières du Canada : niveaux national et provincial, 1983* (n° 31-203 au catalogue), p. 338.

BIBLIOGRAPHIE

- [1] John S. Crysdale, 'Extending Historical Comparability in Industrial Classification' *Proceedings of the International Conference on Establishment Surveys*, (Buffalo, New York, June 1993). American Statistical Association, November 1993.
- [2] John S. Crysdale, 'Industrial Classification in the Canadian Census of Manufactures: Towards Less Art and More Science' *Statistical Journal of the United Nations Economic Commission for Europe*, December 1988, Volume 5, No 4., 377-392. Also available as 'Industrial Classification in the Canadian Census of Manufactures: Automated Verification Using Product Data' *Research Paper Series*, Analytical Studies Branch, Statistics Canada, Discussion Paper #20, January 1989.
- [3] Statistics Canada, *Concepts and definitions of the census of manufactures*, Catalogue 31-528, Ottawa, 1979.
- [4] Statistics Canada, *The input-output structure of the Canadian economy, 1961-1981 (Revised data)*, Catalogue 15-510, Ottawa, 1987.
- [5] Statistics Canada, 'Notes on the 1980 Standard Industrial Classification in the Manufacturing Industries' in *Manufacturing industries of Canada: national and provincial areas, 1983*, Catalogue 31-203, Ottawa, 1986, xxiii-xcviii.
- [6] Statistics Canada, *Standard Industrial Classification Manual, Revised 1970*, Catalogue 12-501, Ottawa, 1970.
- [7] Statistics Canada, *Standard Industrial Classification 1980*, Catalogue 12-501E, Ottawa, 1980.

Annexe A : Concordance d'un à un optimale, entre la CTI de 1980 et la CAÉ de 1970²²

Error < 3								Error >=3			
1011	1011	1993	1899	2733	2733	3121	3160	3512	3512	1052	1050
1012	1012	1994	1899	2791	2740	3191	3150	3521	3520	1099	1089
1021	1020	1995	1832	2792	2740	3192	3150	3541	3541	1631	1650
1031	1031	2431	2431	2793	2740	3193	3150	3542	3542	1712	1740
1032	1032	2432	2431	2799	2740	3194	3150	3549	3549	1719	1792
1041	1040	2435	2432	2811	2860	3199	3150	3551	3550	1829	1810
1049	1040	2441	2441	2819	2860	3211	3210	3561	3561	1911	1852
1051	1050	2443	2441	2831	2880	3231	3230	3571	3570	1999	1899
1053	1060	2444	2441	2839	2880	3241	3241	3581	3580	2433	2431
1061	1083	2445	2442	2841	2890	3242	3243	3591	3591	2434	2431
1071	1071	2491	2392	2849	2890	3243	3242	3592	3599	2442	2441
1072	1072	2495	2460	2911	2910	3244	3242	3593	3599	2451	2450
1081	1082	2496	2480	2912	2910	3251	3250	3594	3599	2492	2431
1082	1081	2511	2511	2919	2910	3252	3250	3611	3651	2493	1750
1083	1081	2512	2513	2921	2920	3253	3250	3612	3652	2494	2310
1091	1089	2521	2520	2941	2940	3254	3250	3699	3690	2499	2441
1092	1089	2522	2520	2951	2950	3255	3250	3712	3783	2599	2599
1093	1089	2541	2543	2959	2950	3257	1880	3722	3720	2821	2870
1093	1089	2542	2544	2961	2960	3259	3250	3729	3799	3021	3010
1094	1089	2543	2541	2962	2960	3261	3260	3731	3730	3022	3010
1111	1091	2549	2541	2971	2970	3271	3270	3741	3740	3023	3020
1121	1092	2561	2560	2999	2980	3281	3280	3751	3750	3053	3051
1131	1093	2581	2580	3011	3010	3299	3290	3761	3760	3256	1650
1141	1094	2591	2591	3029	3020	3311	3310	3771	3770	3372	3360
1211	1510	2592	2593	3031	3031	3321	3320	3791	3791	3562	3562
1221	1530	2593	2593	3032	3039	3321	3320	3792	3799	3599	3530
1511	1623	2593	2593	3039	3039	3331	3330	3913	3912	3711	3782
1521	1629	2611	2619	3041	3041	3332	2680	3914	3914	3721	3782
1599	1629	2612	2619	3042	3042	3333	3399	3921	3920	3911	3911
1611	1650	2619	2619	3049	3042	3341	3340	3922	3920	3912	3911
1621	1650	2641	2640	3051	3059	3351	3350	3931	3931	3999	3999
1621	1650	2649	2640	3052	3059	3352	3350	3932	3932		
1691	2733	2691	2660	3059	3059	3359	3350	3971	3970		
1699	1650	2692	2660	3061	3060	3361	3180	3991	3991		
1711	1720	2699	2660	3062	3060	3362	3180	3992	3992		
1713	1799	2711	2710	3063	3060	3369	3180	3993	3993		
1811	1831	2712	2710	3069	3060	3371	3360	6012	1072		
1821	1820	2713	2710	3071	3070	3379	3360	6213	2611		
1831	2391	2714	2710	3081	3080	3381	3380	9213	1072		
1921	1860	2719	2710	3091	3090	3391	3391				
1931	1872	2721	2720	3092	3090	3392	3399				
1991	1892	2731	2731	3099	3090	3399	3399				
1992	1894	2732	2732	3111	3110	3511	3511				

²² Les noms d'industries se trouvent dans *Classification des activités économiques, révisée 1970* et dans *Classification type des industries, 1980*. Les relations présentées ici sont conformes à celles que renferme la concordance officielle de *Industries manufacturières du Canada : niveaux national et provincial, 1983*, n° 31-203 au catalogue. Les deux concordances reposent sur les données relatives à tous les enregistrements et non seulement celles concernant les établissements qui déclarent des données détaillées sur les produits. En outre, il s'agit de données de 1982 et de 1983, de sorte que l'applicabilité soit générale. L'usage des données de 1983 est limité à deux fins : 1) définir les établissements qui, ayant déclaré des données détaillées sur les produits en 1982, sont classés dans l'industrie 3721 en 1983, de sorte qu'ils soient représentés dans la série officielle de 1982 et 2) définir les établissements qui, en 1983, ont été transférés de la classe 893 de la CAÉ de 1970 à une classe d'industries manufacturières et les prendre en considération (voir la note 16). Ce tableau est divisé en deux groupes, selon un critère d'erreur. Cette valeur seuil est égale à la proportion de l'industrie de la classification de départ qui, d'après les données recoupées desquelles est issue la concordance, se rattache en propre à d'autres industries que celle à laquelle l'industrie initiale a été associée strictement dans la classification d'arrivée. L'établissement d'une concordance d'un à un optimale a amené l'élimination de seize industries (dont une industrie non manufacturière) -- pourtant pertinentes -- dans la classification d'arrivée; ces industries sont (selon la CAÉ de 1970) : 1624, 1840, 1851, 1871, 1891, 1893, 2491, 2492, 2499, 2592, 3781, 3913, 3915, 3996, 3998, 8930.

Annexe B : Concordance d'un à un optimale, entre la CAÉ de 1970 et la CTI de 1980²³

Error < 3	1880 3257	2970 2971	3570 3571	Error >=3	3020 3029
1011 1011	1891 1999	2980 2999	3580 3581	1040 1049	3031 3031
1012 1012	1892 1991	3041 3041	3591 3591	1050 1052	3039 3039
1020 1021	1894 1992	3051 3053	3651 3611	1072 1072	3042 3049
1031 1031	2310 2494	3080 3081	3652 3612	1081 1083	3059 3059
1032 1032	2391 1831	3110 3111	3690 3699	1089 1099	3060 3062
1060 1053	2432 2435	3160 3121	3720 3722	1629 1599	3070 3071
1071 1071	2442 2445	3210 3211	3730 3731	1650 1699	3090 3099
1082 1081	2450 2451	3230 3231	3740 3741	1799 1713	3150 3199
1083 1061	2460 2495	3241 3241	3750 3751	1832 1829	3180 3361
1091 1111	2480 2496	3243 3242	3760 3761	1893 1999	3242 3243
1092 1121	2491 2493	3260 3261	3770 3771	1899 1994	3250 3251
1093 1131	2492 2499	3270 3271	3781 3711	2392 2491	3350 3351
1094 1141	2511 2511	3280 3281	3791 3791	2431 2433	3360 3379
1510 1211	2513 2512	3290 3299	3912 3913	2441 2442	3399 3333
1530 1221	2543 2541	3310 3311	3913 3999	2499 2499	3599 3594
1623 1511	2544 2542	3320 3321	3914 3914	2520 2522	3782 3711
1624 1712	2560 2561	3330 3331	3915 3999	2541 2543	3783 3712
1720 1711	2580 2581	3340 3341	3931 3931	2593 2592	3799 3799
1740 1712	2591 2591	3380 3381	3932 3932	2619 2611	3911 3911
1750 2493	2592 2599	3391 3391	3970 3971	2640 2641	3920 3921
1792 1719	2599 2599	3511 3511	3991 3991	2660 2692	
1810 1829	2611 6213	3512 3512	3992 3992	2710 2712	
1820 1821	2680 3332	3520 3521	3993 3993	2733 1691	
1831 1811	2720 2721	3530 3599	3994 3994	2740 2799	
1840 1999	2731 2731	3541 3541	3996 3999	2860 2819	
1851 1911	2732 2732	3542 3542	3998 3999	2880 2839	
1852 1911	2870 2821	3549 3549	3999 3999	2890 2841	
1860 1921	2920 2921	3550 3551	8930 2821	2910 2919	
1871 1999	2940 2941	3561 3561		2950 2959	
1872 1931	2960 2961	3562 3562		3010 3011	

²³ Les noms d'industries se trouvent dans *Classification des activités économiques, révisée 1970* et dans *Classification type des industries, 1980*. Les relations présentées ici sont conformes à celles que renferme la concordance officielle de *Industries manufacturières du Canada : niveaux national et provincial, 1983*, n° 31-203 au catalogue. Les deux concordances reposent sur les données relatives à tous les enregistrements et non seulement celles concernant les établissements qui déclarent des données détaillées sur les produits. En outre, il s'agit de données de 1982 et de 1983, de sorte que l'applicabilité soit générale. L'usage des données de 1983 est limité à deux fins : 1) définir les établissements qui, ayant déclaré des données détaillées sur les produits en 1982, sont classés dans l'industrie 3721 en 1983, de sorte qu'ils soient représentés dans la série officielle de 1982 et 2) définir les établissements qui, en 1983, ont été transférés de la classe 893 de la CAÉ de 1970 à une classe d'industries manufacturières et les prendre en considération (voir la note 16). Ce tableau est divisé en deux groupes, selon un critère d'erreur. Cette valeur seuil est égale à la proportion de l'industrie de la classification de départ qui, d'après les données recoupées desquelles est issue la concordance, se rattache en propre à d'autres industries que celle à laquelle l'industrie initiale a été associée strictement dans la classification d'arrivée. L'établissement d'une concordance d'un à un optimale a amené l'élimination de 82 industries (dont deux industries non manufacturières) -- pourtant pertinentes -- dans la classification d'arrivée; ces industries sont (selon la CTI de 1980) : 1041, 1051, 1082, 1091, 1092, 1093, 1094, 1521, 1611, 1621, 1631, 1993, 1995, 2431, 2432, 2434, 2441, 2443, 2444, 2492, 2521, 2549, 2593, 2612, 2619, 2649, 2691, 2699, 2711, 2713, 2714, 2719, 2733, 2791, 2792, 2793, 2811, 2831, 2849, 2911, 2912, 2951, 3021, 3022, 3023, 3032, 3042, 3051, 3052, 3061, 3063, 3069, 3091, 3092, 3191, 3192, 3193, 3194, 3244, 3252, 3253, 3254, 3255, 3256, 3259, 3352, 3359, 3362, 3369, 3371, 3372, 3392, 3399, 3592, 3593, 3721, 3729, 3792, 3912, 3922, 6012, 9213.

Annexe C (suite et fin)

n° id. CAÉ 1970

CTI 1980

090	3150	3191	3192	3193	3194	3199
091	3160	3121				
092	3180	3361	3362	3369		
093	3210	3211				
094	3230	3231				
095	3241	3241				
096	3242	3243	3244			
097	3243	3242				
098	3260	3261				
099	3270	3271				
100	3280	3281				
101	3290	3299				
102	3310	3311				
103	3320	3321				
104	3330	3331				
105	3340	3341				
106	3350	3351	3352	3359		
107	3360	3371	3372	3379		
108	3380	3381				
109	3391	3391				
110	3399	3333	3392	3399		
111	3511	3511				
112	3512	3512				
113	3520	3521				
114	3530	3592	3593	3594	3599	
115	3541	3541				
116	3542	3542				
117	3549	3549				
118	3550	3551				
119	3561	3561				
120	3562	3562				
121	3570	3571				
122	3580	3581				
123	3591	3591				
124	3651	3611				
125	3652	3612				
126	3690	3699				
127	3720	3722				
128	3730	3731				
129	3740	3741				
130	3750	3751				
131	3760	3761				
132	3770	3771				
133	3781	3711	3712	3721		
134	3791	3791				
135	3799	3729	3792	3799		
136	3911	3911	3912			
137	3912	3913				
138	3913	3999				
139	3914	3914				
140	3920	3921	3922			
141	3931	3931				
142	3932	3932				
143	3970	3971				
144	3991	3991				
145	3992	3992				
146	3993	3993				
147	3994	3994				

Crysdale / Comparabilité des classifications

Annexe D : Pourcentage de classement erroné (moyenne pour les industries du niveau chiffres)²⁶

	niveau à 4 chiffres				niveau à 3 chiffres				niveau à 2 chiffres			
	srr	arr	con	comb	srr	arr	con	comb	srr	arr	con	comb
<u>CAÉ 1970</u>												
10 Aliment & Boissons	0.6	0.7	0.7	0.3	0.3	0.4	0.4	0.3	0.0	0.0	0.0	0.0
15 Tabac	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16 Caoutchouc/Plstc	5.1	5.4	4.1	2.2	5.2	5.4	3.2	2.2	5.0	5.2	3.2	2.2
17 Cuir	1.5	1.5	6.1	1.3	1.5	1.5	4.4	1.2	1.3	1.3	4.6	1.1
18 Textiles	3.6	3.3	21.2	2.4	2.7	2.6	14.1	2.0	1.2	1.1	0.1	0.6
23 Bonneterie	18.1	17.4	16.7	16.0	18.8	17.7	14.4	15.6	19.7	17.9	12.4	15.6
24 Habillement	7.4	8.9	10.7	7.5	7.4	8.8	10.7	7.4	6.2	5.5	3.0	4.6
25 Bois	4.2	3.4	0.8	0.4	4.4	3.4	0.0	0.1	0.6	0.4	0.0	0.0
26 Meuble	4.8	4.8	0.3	0.3	4.8	4.8	0.3	0.3	1.4	1.1	0.0	0.0
27 Papier	2.0	2.0	0.3	0.1	2.0	1.9	0.3	0.1	0.8	0.7	0.3	0.1
28 Imprimerie, édition	3.0	2.9	0.2	0.7	3.0	2.9	0.2	0.7	0.9	0.9	0.0	0.2
29 1 ^{er} trans. métaux	2.4	2.3	0.0	0.1	2.4	2.3	0.0	0.1	1.2	1.3	0.0	0.1
30 Fab. prod. métaux	8.3	6.8	1.8	1.8	7.3	6.5	1.5	1.6	3.6	3.2	0.5	0.4
31 Fab. Machines	3.3	2.8	0.3	0.3	3.3	2.8	0.3	0.3	3.3	2.8	0.3	0.3
32 Fab. équip. trans.	1.5	1.4	0.3	0.3	1.3	1.2	0.3	0.3	1.1	1.1	0.3	0.3
33 Prod. Electriques	4.8	3.8	0.8	1.3	4.8	3.9	0.8	1.3	2.6	1.9	0.2	0.4
35 Prod. Min. Non-met	3.7	3.3	1.7	0.6	1.8	1.3	1.7	0.6	1.5	1.1	0.4	0.5
36 Prod. Pétr./Char.	0.6	0.6	0.0	0.0	0.6	0.6	0.0	0.0	0.1	0.1	0.0	0.0
37 Industrie chimique	2.8	2.5	3.0	0.8	2.5	2.3	0.0	0.5	0.8	0.6	0.0	0.5
39 Ind. Manuf. diverse	8.7	7.0	6.2	4.2	7.5	5.6	2.5	3.7	7.1	5.3	0.9	3.4
99 Total	2.8	2.5	1.7	0.8	2.5	2.3	1.1	0.8	1.4	1.3	0.4	0.5
<u>CTI 1980</u>												
10 Aliments	1.0	0.9	22.7	0.5	0.9	0.8	1.0	0.5	0.2	0.2	0.2	0.2
11 Boissons	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12 Tabac	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15 Prod. caoutchouc	2.4	2.0	14.5	2.0	2.4	2.0	14.5	2.0	2.3	1.9	0.0	1.9
16 Prod. Plastique	7.7	6.9	58.4	6.3	7.7	6.9	58.3	6.3	4.3	3.5	7.6	2.9
17 Cuir	0.8	0.6	1.9	0.5	0.6	0.4	0.1	0.3	0.6	0.4	0.1	0.3
18 Textiles (1 transf)	2.3	0.5	1.8	0.3	2.3	0.5	1.8	0.3	1.0	0.3	1.7	0.2
19 Prod. textiles	2.3	1.7	35.6	1.1	1.9	1.3	2.5	0.9	1.6	1.1	2.5	0.9
24 Habillement	4.1	1.7	56.8	1.7	2.8	1.2	14.8	1.1	0.5	0.2	0.1	0.2
25 Bois	1.8	1.5	11.5	0.6	1.4	1.2	0.0	0.3	0.6	0.5	0.0	0.3
26 Meuble	6.2	5.4	62.8	5.3	3.4	2.4	0.3	2.3	1.4	1.2	0.0	1.1
27 Papier	2.4	2.3	60.7	2.0	2.1	1.9	0.8	1.7	0.6	0.5	0.8	0.5
28 Imprimerie, édition	3.6	3.6	24.0	3.4	2.8	2.7	0.2	2.5	1.0	0.9	0.0	0.7
29 1 ^{er} trans. métaux	2.5	1.7	24.0	0.4	2.4	1.6	0.0	0.2	1.2	0.6	0.0	0.2
30 Fab. prod. métaux	8.1	7.6	43.2	6.2	6.5	5.9	4.1	4.5	3.4	3.0	0.5	1.9
31 Machinerie	6.4	6.0	58.8	4.2	3.6	3.3	0.3	1.4	3.5	3.2	0.3	1.4
32 Métal de transport	1.9	1.5	19.8	0.9	1.1	0.8	0.4	0.4	0.9	0.6	0.4	0.4
33 Prod. électriques	3.8	3.5	46.0	2.7	3.2	3.0	5.0	2.2	1.7	1.5	0.1	0.9
35 Prod. Min. Non-met	3.6	3.2	10.5	0.9	1.5	1.1	0.4	0.9	1.2	1.0	0.4	0.9
36 Prod. Pétr./Char.	0.6	0.4	0.0	0.0	0.6	0.4	0.0	0.0	0.1	0.0	0.0	0.0
37 Industrie chimique	3.6	2.3	11.6	1.6	3.6	2.2	10.0	1.5	0.6	0.3	0.0	0.2
39 Ind. Manuf. diverse	6.6	5.4	30.3	2.5	4.8	3.5	0.9	1.4	4.1	3.0	0.9	1.1
99 Total	2.8	2.3	25.7	1.6	2.2	1.7	2.9	1.1	1.1	0.8	0.4	0.5

²⁶ Abréviations : srr = codage détaillé par produit (sans règle de résistance), arr = codage détaillé par produit (avec règles de résistance), con = codage d'un à un, comb = combinaison de méthodes. Dans ce tableau, on ne fait pas la distinction entre des valeurs nulles réelles et des valeurs nulles obtenues par arrondissement.

**DIRECTION DES ÉTUDES ANALYTIQUES
DOCUMENTS DE RECHERCHE**

N°

1. *Réaction comportementale dans le contexte d'une simulation micro-analytique socio-économique, Lars Osberg*
2. *Chômage et formation, Garnett Picot*
3. *Des pensions aux personnes au foyer et leur répartition sur la durée du cycle de vie, Michael Wolfson*
4. *La modélisation des profils d'emploi des Canadiens au cours de leur existence, Garnett Picot*
5. *Perte d'un emploi et adaptation au marché du travail dans l'économie canadienne, Garnett Picot et Ted Wannell*
6. *Système de statistiques relatives à la santé: proposition d'un nouveau cadre théorique visant l'intégration de données relatives à la santé, Michael C. Wolfson*
7. *Projet-pilote de raccordement micro-macro pour le secteur des ménages au Canada, Hans J. Adler et Michael C. Wolfson*
8. *Notes sur les groupements de société et l'impôt sur le revenu au Canada, Michael C. Wolfson*
9. *L'expansion de la classe moyenne: données canadiennes sur le débat sur la déqualification, John Myles*
10. *La montée des conglomérats, Jorge Niosi*
11. *Analyse énergétique du commerce extérieur canadien: 1971 et 1976, K.E. Hamilton*
12. *Taux nets et bruts de concentration des terres, Ray D. Bollman et Philip Ehrensaft*
13. *Tables de mortalité en l'absence d'une cause pour le Canada (1921 à 1981): une méthode d'analyse de la transition épidémiologique, Dhruva Nagnur et Michael Nagrodski*
14. *Distribution de la fréquence d'occurrence des sous-séquences de nucléotides, d'après leur capacité de chevauchement, Jane F. Gentleman et Ronald C. Mullin*

15. *L'immigration et le caractère ethnolinguistique du Canada et du Québec, Réjean Lachapelle*
16. *Intégration de la ferme au marché extérieur et travail hors ferme des membres des ménages agricoles, Ray D. Bollman et Pamela Smith*
17. *Les salaires et les emplois au cours des années 1980: éolutin des salaires des jeunes et déclin de la classe moyenne, J. Myles, G. Picot et T. Wannell*
18. *Profil des exploitants agricoles dotés d'un ordinateur, Ray D. Bollman*
19. *Répartitions des risques de mortalité: une analyse de tables de mortalité, Geoff Rowe*
20. *La classification par industrie dans le recensement canadien des manufactures: vérification automatisée à l'aide des données sur les produits, John S. Crysdale*
21. *Consommation, revenus et retraite, A.L. Robb et J.B. Burbridge*
22. *Le renouvellement des emplois dans le secteur manufacturier au Canada, John R. Baldwin et Paul K. Gorecki*
23. *La Dynamique des marchés concurrentiels, John R. Baldwin et Paul K. Gorecki*
 - A. *Entrée et sortie d'entreprises dans le secteur manufacturier au Canada*
 - B. *Mobilité à l'intérieur des branches d'activité dans le secteur manufacturier au Canada*
 - C. *Mesure de l'entrée et de la sortie dans le secteur manufacturier au Canada: méthodologie*
 - D. *Effet de la libre concurrence sur la productivité: rôle de la rotation des entreprises et des usines*
 - E. *Les fusions et le processur concurrentiel*
 - F. *À venir*
 - G. *Lews statistiques de concentration comme prédicteurs du degré de concurrence*
 - H. *Le rapport entre la mobilité et la concentration dans le secteur manufacturier au Canada*
24. *Améliorations apportées au SAS de l'ordinateur central en vue de faciliter l'analyse exploratoire des données, Richard Johnson et Jane F. Gentleman*
25. *Aspects de l'évolution du marché du travail au Canada: mutations intersectorielles et roulement de la main-d'oeuvre, John R. Baldwin et Paul K. Gorecki*
26. *L'écart persistant: étude de la différence dans les gains des hommes et des femmes qui ont récemment reçu un diplôme d'études postsecondaires, Ted Wannell*

27. *Estimation des pertes de sol sur les terres agricoles à partir des données du recensement de l'agriculture sur les superficies cultivées, Douglas F. Trant*
28. *Les bons et les mauvais emplois et le déclin de la classe moyenne: 1967-1986, Garnett Picot, John Myles, et Ted Wannell*
29. *Données longitudinales sur la carrière relatives à certaines cohortes de fonctionnaires, Garnett Picot et Ted Wannell*
30. *L'incidence des revenus sur la mortalité sur une période de vingt-cinq ans, Michael Wolfson, Geoff Rowe, Jane F. Gentleman et Monica Tomiak*
31. *Réaction des entreprises à l'incertitude des prix: la stabilisation tripartite et l'industrie des bovins dans l'ouest du Canada, Theodore M. Horbulyk*
32. *Méthodes de lissage pour microdonnées longitudinales simulées, Jane F. Gentleman, Dale Robertson et Monica Tomiak*
33. *Tendances des investissements directs canadiens à l'étranger, Paul K. Gorecki*
34. *POHEM - une approche inédite pour l'estimation de l'espérance de vie corrigée en fonction de l'état de santé, Michael C. Wolfson*
35. *Emploi et taille des entreprises au Canada: les petites entreprises offrent-elles des salaires inférieurs?, René Morissette*
36. *Distinguer les caractéristiques des acquisitions étrangères en haute technologie dans le secteur manufacturier canadien, John R. Baldwin et Paul K. Gorecki*
37. *Efficiencce des branches d'activité et roulement des établissements dans le secteur canadien de la fabrication, John R. Baldwin*
38. *Le vieillissement de la génération du baby boom: effets sur le secteur public du Canada, Brian B. Murphy et Michael C. Wolfson*
39. *Tendances dans la répartition de l'emploi selon la taille des employeurs: données canadiennes récentes, Ted Wannell*
40. *Les petites collectivités du Canada atlantique: structure industrielle et caractéristiques du marché du travail au début des années 80, Garnett Picot et John Heath*
41. *La répartition des impôts et des transferts fédéraux et provinciaux dans le Canada rural, Brian B. Murphy*
42. *Les multinationales étrangères et les fusions au Canada, John Baldwin et Richard Caves*

43. *Recours répétés à l'assurance-chômage, Miles Corak*
44. *POHEM -- Un cadre permettant d'expliquer et de modéliser la santé de populations humaines, Michael C. Wolfson*
45. *Analyse de modèle de l'espérance de vie en santé de la population: une approche fondée sur la microsimulation, Michael C. Wolfson et Kenneth G. Manton*
46. *Revenue de carrière et décès: une analyse longitudinale de la population âgée masculine du Canada, Michael C. Wolfson, Geoff Rowe, Jane Gentleman et Monica Tomiak*
47. *La modélisation des profils d'emploi des canadiens au cours de leur existence, Miles Corak*
48. *La dynamique du mouvement des entreprises et le processus concurrentiel, John Baldwin*
49. *Élaboration de données-panel longitudinales à partir de registres des entreprises: Observations du Canada, John Baldwin, Richard Dupuy et William Penner*
50. *Le calcul de l'espérance de vie ajustée sur la santé pour une province canadienne à l'aide d'une fonction d'utilité multiattribut: Un premier essai, J.-M. Berthelot, R. Roberge et M. C. Wolfson*
51. *Mesure de la robustesse des barrières à l'entrée, J. R. Baldwin et M. Rafiquzzaman*
52. *Les multinationales au Canada : Caractéristiques et facteurs déterminants, Paul K. Gorecki*
53. *La persistance du chômage : Dans quelle mesure l'attribuer aux prestations d'assurance-chômage de prolongation fondée sur le taux de chômage régional, Miles Corak et Stephen Jones*
54. *Variations cycliques de la durée des périodes de chômage, Miles Corak*
55. *Licenciements et travailleurs déplacés: Variations cycliques, secteurs les plus touchés et expériences après le licenciement, Garnett Picot et Wendy Pyper*
56. *La durée du chômage en période d'expansion et de récession, Miles Corak*
57. *Obtenir un emploi en 1989-1990 au Canada, René Morissette*
58. *L'appariement de données échantillonales et administratives en vue d'étudier les déterminants de la santé, P. David, J.-M. Berthelot et C. Mustard*

59. *Maintenir la comparabilité dans le temps des classifications par industrie, John S. Crysdale*

Pour de plus amples renseignements, s'adresser au Président, Comité d'études des publications, Direction des études analytiques, Édifice, R.H. Coats, 24ième étage, Statistique Canada, Parc Tunney, Ottawa, Ontario, K1A 0T6, (613) 951-8213.

STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA



1010173807

OOS