

LA CALENDRIALISATION DANS LE CADRE DE L'ENQUÊTE UNIFIÉE SUR LES ENTREPRISES

Yanick Beaucage et Peter S. Tarassoff¹

RÉSUMÉ

Dans les enquêtes auprès des entreprises, la plupart des répondants rapportent leurs données selon leur propre année fiscale correspondant à leur exercice financier. La période fiscale diffère fréquemment d'un répondant à l'autre tout en étant différente de l'année civile. Malgré la diversité des périodes couvertes, les gestionnaires de ces enquêtes regroupent habituellement ces données et publient des estimations qualifiées d'« annuelles ». Cependant, le fait de calculer des estimations ou des taux de croissance à partir de données provenant de différentes périodes fiscales engendre plusieurs problèmes. On obtiendrait des résultats différents si les données étaient basées sur l'année civile. La transformation de données fiscales en données civiles aussi appelée calendrialisation vient remédier au problème. Avec la nouvelle enquête unifiée sur les entreprises (EUE) de Statistique Canada, on a maintenant l'opportunité d'appliquer plus généralement la calendrialisation. Sous l'égide du projet d'amélioration des statistiques économiques provinciales (le PASEP), l'enquête pilote de l'EUE se concentrera sur sept industries cibles dont les données rapportées seront calendrialisées.

Ce document montrera l'importance de procéder à la calendrialisation, expliquera la méthode employée ainsi que les entrées nécessaires à son application et finalement présentera les divers aspects reliés à la mise en place d'un tel système. Parmi ces aspects, nous discuterons de l'élaboration des profils saisonniers, du type de variable à calendrialiser, du moment propice pour procéder à la calendrialisation, des contraintes à respecter entre les variables et/ou entre les différentes unités (entreprises/établissements) lors de la calendrialisation, et finalement d'une méthode de vérification des données calendrialisées. Ceci sera fait de façon exploratoire puisque le système ne sera mis en application qu'à l'automne 1998. Des chiffres provenant d'autres enquêtes seront utilisés pour illustrer la méthode.

MOTS CLÉS : Calendrialisation, étalonnage, PASEP, annualisation, minimisation quadratique.

ABSTRACT

Many respondents to business surveys provide data for their respective fiscal years, which often differ both from one respondent to another and from the calendar year. Despite the mix of periods covered, business survey managers commonly aggregate all such data and published so-called « annual » estimates. However, several problems ensue from the fact that these estimates and any related growth rates are calculated from data with different fiscal year-end dates. These results would have been different if all of the data had been based on true calendar years. Calendarization, the transformation of fiscal year to a calendar year basis, is a remedy to this problem. Statistics Canada's new Unified Enterprise Survey (UES) represents an opportunity for a more widespread application of calendarization. Under the aegis of the Project to Improve Provincial Economic Statistics (PIPES), the UES Pilot will target seven industries, and all respondent microdata are to be calendarized.

This document shows the importance of calendarization, explain how it is performed and what the necessary inputs are, and finally discuss certain aspects of the development of calendarization system. These aspects include: the construction of seasonal profiles, the types of variable that are to be calendarized, the place of calendarization in a processing sequence, the constraints among variables and or among units (enterprise/establishment) that calendarization ought to preserve, and finally the evaluation of calendarized values. It will be made in an exploratory way since the UES calendarization system will not be ready until Fall 1998. Data from other surveys will be used to illustrate each point.

KEY WORDS : Calendarization, calibration, PIPES, annualization, quadratic minimization.

¹ Yanick Beaucage, Division des méthodes d'enquêtes auprès des entreprises, Statistique Canada, 11-G édifice R.H.Coats, Ottawa (Ontario), Canada, K1A 0T6, E-Mail : beauyan@statcan.ca.
Peter S. Tarassoff, Division de la statistique des entreprises, Statistique Canada, GE-8 édifice Jean-Talon, Ottawa (Ontario), Canada, K1A 0T6, E-Mail : tarapet@statcan.ca.

1. INTRODUCTION

L'enquête unifiée auprès des entreprises (EUE) est une enquête annuelle qui invite les répondants à divulguer certaines informations financières requises selon leurs années fiscales (correspondant à leurs exercices financiers) pour en faciliter l'obtention. Évidemment, chaque répondant peut avoir des périodes fiscales différentes, au gré de ses besoins. Lors de l'agrégation des données, on fait face à une multitude de périodes se chevauchant l'une et l'autre et, par le fait même, à un éventail d'information couvrant une période allant jusqu'à presque deux ans. La procédure habituelle consiste à ne pas tenir compte de la période exacte de rapport et de tout simplement agréger les données rapportées et ainsi publier des estimations se rapportant soi-disant à une même et unique période, l'année civile par exemple. Mais, malgré la facilité de cette pratique, est-ce rigoureux? Que peut-on faire pour produire des estimations se rapportant réellement à l'année civile?

Le but de cet article est de montrer d'une part le danger de l'agrégation pure et simple des données couvrant différentes périodes. Dans un deuxième temps, nous allons discuter de l'utilisation de deux méthodes de calendrialisation permettant de résoudre le problème. Finalement, on aborde d'autres aspects connexes à la calendrialisation des données de l'EUE.

2. CALENDRIALISATION

Dans le cadre de l'enquête pilote de l'EUE de 1997, chaque questionnaire contient une section « Période de déclaration », indiquant aux répondants la période visée par l'enquête : « Aux fins de la présente enquête, veuillez déclarer les données pour **vosre exercice financier de douze mois** dont le **DERNIER JOUR** se situe le, ou entre le 1^{er} janvier 1997 et le 31 décembre 1997 ». Cette procédure n'est pas inhabituelle, puisque la plupart des renseignements fournis par les entreprises canadiennes le sont selon leur propre exercice financier.

Le répondant est tenu d'indiquer la période (début et fin) pour laquelle ses données sont rapportées en termes de jour/mois/année. On comprend facilement qu'en général, les réponses peuvent s'échelonner de la période « 2 janvier 1996 au 1^{er} janvier 1997 » à la période « 1^{er} janvier 1997 au 31 décembre 1997 ». On a donc un ensemble de réponses qui couvre une période de deux ans, lesquelles on aimerait convertir en une seule et même période soit du 1^{er} janvier 1997 au 31 décembre 1997, appelée l'année civile.

Pour évaluer l'étendue des périodes sur lesquelles les entreprises canadiennes rapportent leurs données, nous devons examiner la distribution de leurs fins d'année fiscale (en supposant que leurs années fiscales couvrent bien 12 mois). Il existe quelques sources de données administratives permettant d'obtenir ces informations. Le tableau 1 présente la distribution du nombre d'établissements (sous forme de pourcentages) selon le mois de fin de période fiscale.

Tableau 1 : Distribution du nombre d'établissements selon le mois de fin de période fiscale

	Construction	Courriers	Courtiers
Janvier	9%	6%	7%
Février	7%	6%	4%
Mars	10%	8%	6%
Avril	8%	8%	6%
Mai	6%	6%	5%
Juin	6%	7%	7%
Juillet	7%	8%	8%
Août	6%	8%	7%
Septembre	7%	8%	8%
Octobre	7%	7%	7%
Novembre	5%	4%	5%
Décembre	22%	25%	31%
Total	100%	100%	100%

Notons que la classification est approximative puisque les codes SCIAN (Système de Classification Industrielle de l'Amérique du Nord) utilisés pour l'enquête pilote ne sont pas encore très répandus. Une correspondance avec l'ancienne classification type des industries (CTI-80) nous a permis de fabriquer le tableau. Évidemment, même si la proportion des répondants ayant une fin d'année fiscale en décembre est d'environ 25%, on peut s'interroger sur la distribution lorsque l'on considère les revenus. Le tableau 2, construit à partir de données fiscales encore une fois, montre cette distribution sous forme de pourcentages.

Tableau 2 : Distribution des revenus annuels selon le mois de fin de période fiscale des établissements

	Construction	Courriers	Courtiers
Janvier	8%	1%	2%
Février	7%	2%	1%
Mars	10%	3%	1%
Avril	7%	1%	2%
Mai	4%	17%	3%
Juin	4%	11%	2%
Juillet	5%	4%	3%
Août	5%	2%	2%
Septembre	7%	7%	3%
Octobre	8%	2%	3%
Novembre	6%	26%	2%
Décembre	30%	26%	78%
Total	100%	100%	100%

Évidemment, les distributions du pourcentage de déclarants et du pourcentage de leurs revenus sont assez différentes. On peut dire que dans l'industrie des courtiers, les déclarants avec gros revenus ont habituellement des fins d'années fiscales se terminant en décembre, ce qui minimise l'impact de ne pas calendrialiser. Mais pour les autres industries, on observe des exercices financiers répartis sur toute l'année avec une plus grande proportion pour le mois de décembre.

Pour ces cas-là, quel dommage cause-t-on aux estimations lorsque l'on ne calendrialise pas? En fait, la réponse à cette question dépend du taux de croissance des variables à calendrialiser entre janvier 1996 et décembre 1997. Évidemment plus ce taux diffère de 0, plus la calendrialisation devient importante. Une façon d'évaluer de manière grossière l'erreur que l'on commet en ne procédant pas à la calendrialisation consiste à supposer que les données sont rapportées pour une période ayant un décalage fixe (de trois mois par exemple) sur l'année civile. En fixant le taux de croissance annuel et en calculant le taux mensuel composé équivalent à celui-ci, on peut évaluer l'erreur commise. Le tableau 3 donne l'erreur commise pour des décalages de 3, 6 et 9 mois i.e. correspondant aux fins d'années fiscales se situant le 31 mars, le 30 juin et le 30 septembre.

Tableau 3 : Erreur commise sans calendrialisation

Taux annuel	Décalage en termes de mois (p.r. au 31 décembre)						
	-9	-6	-3	0	3	6	9
-10,0%	8,2%	5,4%	2,7%	0,0%	-2,6%	-5,1%	-7,6%
-5,0%	3,9%	2,6%	1,3%	0,0%	-1,3%	-2,5%	-3,8%
-2,5%	1,9%	1,3%	0,6%	0,0%	-0,6%	-1,3%	-1,9%
-1,0%	0,8%	0,5%	0,3%	0,0%	-0,3%	-0,5%	-0,8%
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
1,0%	-0,7%	-0,5%	-0,2%	0,0%	0,2%	0,5%	0,7%
2,5%	-1,8%	-1,2%	-0,6%	0,0%	0,6%	1,2%	1,9%
5,0%	-3,6%	-2,4%	-1,2%	0,0%	1,2%	2,5%	3,7%
10,0%	-6,9%	-4,7%	-2,4%	0,0%	2,4%	4,9%	7,4%

* Ce tableau s'applique dans le cas de données de flux

On peut ainsi voir que dans le cas d'un décalage négatif de 3 mois (i.e. période fiscale se termine le 30 septembre de l'année t) et d'un taux de croissance de 5% (appliqué aux 3 derniers mois finalement) on commet une erreur de -1,2% en utilisant les données rapportées de ce répondant pour faire une quelconque tabulation. On comprend facilement que plus le décalage ou le taux de croissance s'éloigne de 0, plus grande est l'erreur. Ces erreurs malheureusement ne s'annulent pas (si le taux de croissance va dans la même direction pour chacun d'entre eux) puisque tous les répondants de l'enquête auront des décalages négatifs par rapport à la fin d'année civile en raison

même de la période de couverture de l'enquête. On peut facilement conclure que la somme de telles erreurs causera une différence indésirable dans les tabulations.

3. MÉTHODE DE CALENDRIALISATION

Conceptuellement, la calendrialisation se divise en 3 étapes :

- (1) Interpoler les données fiscales rapportées en données sous-annuelles (telles des données mensuelles, hebdomadaires ou même quotidiennes) en s'assurant que la somme est toujours égale au total annuel.
- (2) Extrapoler, au début et à la fin des périodes couvertes, le comportement des données fiscales (toujours sous une forme sous-annuelle).
- (3) Sommer sur les valeurs (interpolées ou extrapolées) qui se retrouvent au sein de la période d'intérêt comme, par exemple, l'année civile.

Il existe quelques méthodes de calendrialisation qui peuvent plus ou moins combiner chacune de ces étapes pour produire une valeur calendrialisée. Laniel et Poirier (1992) examinent quelques-unes de ces méthodes tout en les comparant au moyen d'une simulation. De même, Cholette (1998) examine à son tour quelques méthodes.

Certaines méthodes sont simples et faciles à appliquer; d'autres sont plus rigoureuses et s'appuient sur des critères habituels de minimisation. La première méthode présentée fait partie de la première catégorie, c'est la méthode que nous appellerons la « méthode fractionnelle ». Celle-ci divise chacune des valeurs fiscales en deux parties selon la portion de l'année couverte par chacune des parties. Lorsque l'on a ainsi divisé deux valeurs fiscales (pour les années t et t+1), on obtient la valeur ajustée pour l'année t :

$$X_t^c = \frac{kX_t^f + (12-k)X_{t+1}^f}{12}$$

où X_t^c dénote la valeur calendrialisée pour l'année t,

X_t^f dénote la valeur fiscale (collectée) pour l'année t et k le nombre de mois passés depuis le début de l'année civile t. Si, par exemple, un répondant termine son année fiscale le 31 mai 1997, la valeur calendrialisée pour 97 est :

$$X_{97}^c = \frac{5}{12} X_{96-97}^f + \frac{7}{12} X_{97-98}^f$$

Notons tout de suite que cette méthode demande d'attendre une deuxième année de données (97-98 en l'occurrence) ou de trouver au moins une prévision pour 97-98. Avec une prévision pour 97-98, on fait

face à une estimation préliminaire suivie d'une estimation finale lorsque les données fiscales de 97-98 sont finalement disponibles.

Une deuxième méthode faisant partie de la catégorie des méthodes plus rigoureuses est celle présentée dans Cholette (1984). Elle utilise une méthode d'étalonnage d'une série sous-annuelle (mensuelle, par exemple) pour calendrialiser les données sur la période de temps désirée. Puisque cette méthode est une modification de l'approche de Denton (1971), appelons-la la « méthode de Denton ». Cette méthode se sert de données sous-annuelles à étalonner aux données fiscales. On utilise alors une fonction objective à minimiser sous la contrainte de respecter les totaux fiscaux. La fonction objective est la suivante :

$$O(x) = \sum_{m=2}^T ((x_m - x_{m-1}) - (z_m - z_{m-1}))^2$$

où z correspond aux valeurs mensuelles à étalonner, l'indice m fait référence à la valeur mensuelle à étalonner et T représente le total de mois compris dans toutes les années fiscales disponibles pour permettre la calendrialisation. Lorsqu'on a une seule année de données, $T=12$. On cherche donc la série mensuelle des valeurs de x_m qui reproduit le plus fidèlement possible le mouvement de la série des z_m , sous la contrainte :

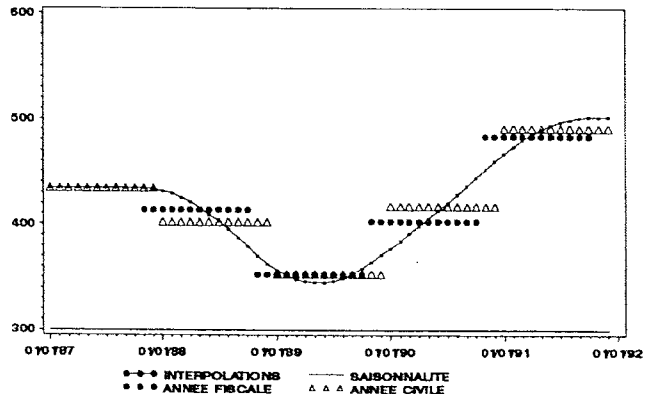
$$X_t^f = \sum_{m=1}^{12} x_m \text{ pour chaque année fiscale } t \text{ impliquée.}$$

Notons que l'on minimise les premières différences ($x_m - x_{m-1}$) des séries de telle sorte que les deux séries soient aussi parallèles que possible. On pourrait vouloir tenter de minimiser les deuxièmes différences ($x_m - x_{m-2}$) ce qui donnerait une fonction objective différente et des valeurs calendrialisées différentes. On peut montrer que le choix de minimiser les deuxièmes différences revient à exécuter la méthode fractionnelle (Cholette (1998)).

Denton propose également une fonction objective proportionnelle où les différences sont divisées par les observation z_m correspondantes. Cholette (1984) a réussi à approximer numériquement la méthode proportionnelle de Denton pour application dans un contexte d'interpolations non pas mensuelles, mais quotidiennes. Un produit dérivé est l'obtention d'extrapolations pour les années non-complétées.

Voici une première illustration où l'on a quatre valeurs fiscales (de nov. 87 à oct. 88, ..., et de nov 90 à oct. 91) représentées par des cercles pleins sur le graphique 1.

Graphique 1 : Application simple de la méthode de Denton



Notons qu'une valeur annuelle est tout d'abord divisée en 12 parties égales avant d'être représentée sur le graphique. Au bas, on retrouve la série des z_m (représentant la saisonnalité) qui est égale, pour les besoins de l'exemple, à la constante 300. La méthode cherche donc la série des interpolations x_m qui minimisent la fonction objective tout en s'assurant que les quatre totaux fiscaux sont respectés. La série des x_m (interpolations) est représentée par la courbe lisse.

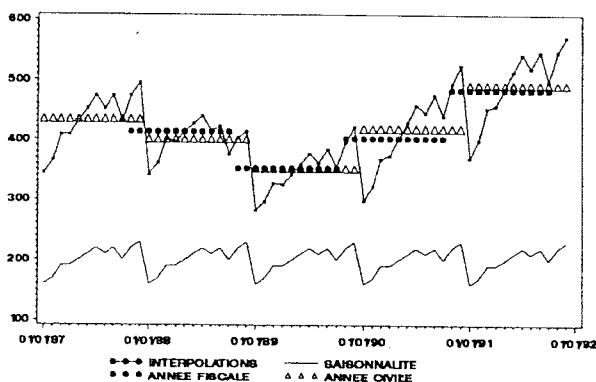
Une fois les interpolations (et les extrapolations pour les périodes de jan. 87 à oct. 87 et de nov 91 à déc. 91) calculées, il ne reste qu'à sommer sur les 12 mois qui nous intéressent, soient ceux de l'année civile. On obtient alors les valeurs calendrialisées (5 dans ce cas-ci) à nouveau divisée en 12 parties égales représentées par des triangles vides. Notons tout de suite que les extrapolations sont plates et ne font que répéter la dernière interpolation.

Lorsqu'on veut calendrialiser une seule observation annuelle basée sur une période fiscale de douze mois selon la méthode telle que présentée ici, on retombe exactement sur la valeur fiscale que l'on voulait calendrialiser. Comme dans le cas de la méthode fractionnelle, on devra attendre d'obtenir plus d'une année financière pour produire une « vraie » valeur calendrialisée. Cette notion implique que les données calendrialisées seront révisées et on pourrait vouloir plus d'une révision. Cholette (1984) indique cependant qu'après la deuxième révision, les changements deviennent très mineurs.

La beauté de cette méthode est son adaptation à l'apport d'information auxiliaire. En effet, au lieu d'utiliser une série des z_m constante, on peut améliorer la méthode à l'aide d'une série d'indices mensuels. Ceci évitera d'une part d'obtenir des extrapolations plates en début et en fin de série et d'autre part une

meilleure calendrialisation dans le cas de données fiscales avec une période différente de douze mois. Si la période fiscale excède 12 mois, la saisonnalité permet de connaître la répartition à l'intérieur de la période fiscale de façon à extraire uniquement la valeur correspondant à l'année civile. Si la période est inférieure à 12 mois, la connaissance du profil saisonnier permet une meilleure inférence sur le reste de l'année (s'il y a lieu). Le graphique 2 illustre le même exemple auquel la série des z_m (saisonnalité) n'est plus constante, mais obéit à un profil donné.

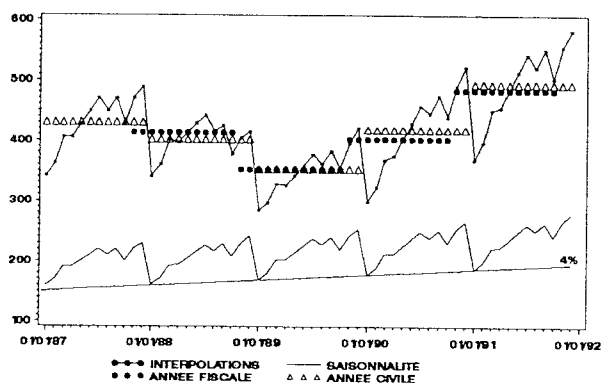
Graphique 2 : Ajout d'un profil saisonnier



Notons encore une fois que dans le cas d'une seule donnée fiscale avec période de douze mois, l'ajout du profil saisonnier à la méthode ne changera pas la valeur calendrialisée. Celle-ci sera en effet égale à la valeur fiscale puisque sur une période de douze mois consécutif, la saisonnalité s'annule.

On peut encore améliorer le processus en greffant un paramètre de tendance aux données. Dans le graphique 3, on a greffé un paramètre de croissance annuelle de 4% en calculant le taux mensuel composé équivalent. De plus, on utilise le même taux de croissance d'une année à l'autre. La droite sous la série des z_m montre cette croissance.

Graphique 3 : Ajout d'un paramètre de tendance



Encore une fois, on utilise une information auxiliaire (la tendance) pour améliorer la méthode. Cette fois-ci, si on a une seule donnée fiscale d'une longueur de douze mois à calendrialiser, la valeur obtenue sera différente.

Notons finalement qu'une quatrième variante de la méthode consiste à ne spécifier que le paramètre de tendance sans utiliser de série d'indices mensuels. La série des z_m serait alors représentée par la droite (de 4%) qui apparaît dans le graphique 3. Cette variante correspond en fait à procéder à de la macro-calendrialisation (calendrialisation au niveau d'un agrégat ou groupement) et à l'appliquer au niveau des micro-données. Un désavantage habituel de la macro-calendrialisation est de considérer une période commune pour toutes les unités du groupement (ce qui est rarement le cas en pratique). Dans ce cas-ci cependant, connaissant pour un groupement donné (industrie par province par exemple) le taux de croissance annuel, on l'applique (sous forme de calendrialisation) individuellement à chacune des données faisant partie de ce groupe. On conserve donc les avantages de la macro-calendrialisation tout en tenant compte de la période individuelle au moment de la calendrialisation.

Un autre avantage de la micro-calendrialisation est l'utilisation des micro-données calendrialisées pour construire toute tabulation requise sans avoir besoin de re-calendrialiser pour le nouveau total désiré. Évidemment l'idéal est de connaître la tendance au niveau de chaque unité, mais celle-ci peut être très difficile à obtenir. Trouver la tendance au niveau du groupe s'avère un bon compromis tout en étant facile à appliquer.

La méthode de Denton a donc les avantages suivants:

- 1) C'est une méthode rigoureuse basée sur une fonction objective à minimiser sous certaines contraintes.
- 2) Elle s'adapte à l'apport d'information auxiliaire sous différentes formes (tendance, profils, ...).
- 3) Elle s'adapte facilement à des années fiscales de longueur autre que douze mois.
- 4) Elle peut s'utiliser avec une seule donnée.

Par ailleurs, la méthode peut estimer les valeurs manquantes lorsqu'il y a un (ou plusieurs) trou dans les données et peut reproduire la méthode fractionnelle en ajustant la fonction objective. Cette méthode semble donc prometteuse et nous espérons pouvoir parvenir à l'implanter pour pouvoir engendrer des données calendrialisées de façon routinière.

4. AUTRES ASPECTS

Évidemment, le fait de choisir une méthode plutôt qu'une autre n'élimine pas la totalité des problèmes reliés à la calendrialisation. Parmi ces problèmes, notons tout de suite le choix du moment auquel procéder à la calendrialisation. Laniel et Poirier (1992) suggèrent d'appliquer la calendrialisation directement sur les données saisies, avant de procéder à la vérification et à l'imputation. Dans le cadre de l'EUE, les données recueillies entrent dans un processus de vérification automatisé qui permet d'évaluer la qualité de l'information collectée ainsi que la possibilité de faire face à des données aberrantes. Les enregistrements rejetés sont alors confiés à la revue manuelle pour être ensuite réintégrés au processus de vérification et d'imputation. On ne peut donc pas sortir du système après la vérification automatisée/manuelle et avant l'imputation pour faire la calendrialisation. Celle-ci a donc lieu après, à la fin des étapes de traitement, juste avant l'estimation.

Évidemment, si on pouvait calendrialiser tout de suite après la vérification cela nous permettrait d'imputer les non-répondants avec des valeurs déjà calendrialisées. Ceci diminuerait le nombre d'enregistrements à calendrialiser tout en évitant d'avoir à évaluer la période de référence des non-répondants. En procédant à la calendrialisation après l'imputation, il faut tout d'abord trouver une source fiable pour établir la période fiscale du non-répondant et ensuite tenir compte de cette période lors du choix du donneur avant de procéder à l'imputation. Une fois l'imputation complétée, les données rapportées (répondants) et les données imputées (non-répondants) sont calendrialisées. D'un point opérationnel, la tâche s'avère traitée plus efficacement avec le module de calendrialisation à la fin, même si des étapes sont ainsi ajoutées au processus.

L'EUE récolte plus de 500 variables différentes à calendrialiser. Traiter toutes ces variables séparément entraînerait des coûts de temps et de développement formidables. De plus, la plupart des variables sont en faits, des « détails » des variables globales : les revenus sont divisés en catégories, les dépenses aussi, etc... Les variables qui correspondent aux « totaux » seront donc les variables calendrialisées et les variables détaillées seront ajustées au prorata.

De plus, l'EUE récolte des données au niveau des entreprises et des établissements appartenant à ces mêmes entreprises. On veut donc s'assurer que si la somme des valeurs non-calendrialisées des établissements est égale à la valeur rapportée au niveau de l'entreprise, il en soit de même pour les

valeurs calendrialisées. Cette contrainte n'est pas facile à formuler dans le cadre de la méthodologie proposée et nous espérons trouver une solution à ce problème graduellement.

Les illustrations présentées jusqu'à présent ne considèrent que des variables de flux tels les revenus, les dépenses,... dont la valeur représente une accumulation sur la période fiscale. Cependant, l'EUE collecte également des variables de stocks (comme les inventaires) dont la valeur, à la différence des variables de flux, représente les stocks au dernier jour de la période fiscale. Heureusement, l'algorithme de Cholette peut traiter de façon similaire les variables de flux et de stocks, mais dans le second cas, il faut nécessairement utiliser un profil saisonnier pour assurer une bonne calendrialisation. Sans profil, la valeur calendrialisée est tout simplement une interpolation à partir d'une droite entre deux valeurs de stocks.

Tout au long de la section précédente, on a vu que la méthode peut fonctionner sous quatre modes différents : aucune information auxiliaire (mode de base), profils saisonniers (mode amélioré I), tendance (mode amélioré II) et profils saisonniers avec tendance (mode ultime). Il est donc important d'utiliser le plus d'information auxiliaire possible et de tenter d'appliquer la méthode en mode ultime. Ceci est d'autant vrai que lorsque l'on a une seule donnée, le mode de base produit une valeur calendrialisée égale à la valeur fiscale originale dans le cas des données de flux et que la valeur n'est pas fiable dans le cas des données de stocks. Ceci a un impact sur l'importance des révisions qui prendront plus d'ampleur une fois l'enquête à sa deuxième année si la calendrialisation est minime lors de la première année.

On voudra donc tenter de trouver un profil saisonnier ainsi que la tendance dans chaque groupement (industrie/province) jugé différent et ce pour chacune des variables à calendrialiser. Ce genre d'information peut devenir très fastidieux à obtenir. L'utilisation de données auxiliaires sera la principale source utilisée pour créer ces profils et calculer ces tendances. De plus, on devra utiliser des variables « proxi » et souhaiter que celles-ci recourent chacun des types de variables à calendrialiser, que ce soit des variables de flux ou de stocks. Ces profils et tendance seront emmagasinés pour fins d'observation et de révision.

Notons qu'au moment de procéder à la calendrialisation des données de l'année de référence 97, l'année 98 sera déjà assez avancée. Les séries de 97 et 98 disponibles à ce moment deviendront très utiles pour évaluer les tendances ainsi que les profils.

Finalement, comme dans tout bon système informatique, on devra procéder à la vérification du processus de calendrialisation. On aimerait donc mettre en place la méthodologie nécessaire pour suivre un enregistrement tout au long du processus de calendrialisation. On pourrait s'assurer de la qualité de la calendrialisation et modifier la valeur là où ce serait nécessaire. Dans un deuxième temps, on aimerait instaurer un processus permettant de confirmer la valeur calendrialisée. Selon la tendance observée dans le groupement (industrie/province) pour la variable calendrialisée, on aimerait utiliser des limites déduites de façon semblable à celles du tableau 3. On pourrait alors déterminer un voisinage à l'intérieur duquel la valeur calendrialisée devrait se trouver. Les valeurs hors du voisinage se verraient envoyées à un processus de revue manuelle où l'intelligence humaine permettra de décider du traitement à adopter.

5. CONCLUSION

Il reste bon nombre de problèmes opérationnels à régler avant la mise en marche complète (et même partielle) du processus de calendrialisation dans le cadre de l'EUE. Nous savons cependant que la calendrialisation a sa place dans l'enquête. Celle-ci permet d'éviter les biais trop souvent ignorés tout en donnant aux gestionnaires une confiance accrue dans la représentativité de leurs chiffres sur une base annuelle civile. Le choix de la méthode est assez clair, mais, pour faire un bon travail lors de la calendrialisation, nous devons nous assurer d'avoir en main les données auxiliaires nécessaires pour

fonctionner au moins en mode amélioré I dans le cas des données de stocks et au moins en mode amélioré II dans le cas des données de flux. Tout ceci pour s'assurer d'éviter de trop grandes révisions si l'on utilise seulement le mode de base pour la première année. La plupart des autres problèmes reliés à la calendrialisation ont été résolus de façon simple ou à l'aide d'idées prometteuses qui attendent encore d'être mises à l'essai. La programmation de celles-ci deviendra un élément crucial dans la mise en marche du processus, mais, si le budget et les ressources s'y prêtent, l'EUE aura le système de calendrialisation le plus sophistiqué au monde.

6. RÉFÉRENCES

Cholette, P.A. (1984). L'ajustement des séries infra-annuelles aux repères annuels. *Technique d'enquête*, vol 10. No. 1, 39-53.

Cholette, P.A. (1998). Comparison of Methods to Calendarize Fiscal Year Data (draft). Document interne, Statistique Canada.

Denton, F.T. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, vol 66, no 333, 99-102.

Laniel, N. et Poirier C. (1992). *Ajustement des données d'enquêtes annuelles à l'année civile: Comparaison des méthodes*. Document interne de Statistique Canada. Version préliminaire seulement.