Techniques d'enquête, décembre 1990
Vol. 16, n° 2, pp. 283–289
Statistique Canada

Étalonnage des séries économiques

NORMAND LANIEL et KIMBERLEY FYFE[1]

RÉSUMÉ

L'étalonnage est l'opération qui consiste à améliorer les estimations tirées d'une enquête infra-annuelle à l'aide des estimations correspondantes tirées d'une enquête annuelle. Par exemple, les estimations de l'enquête annuelle sur les ventes au détail peuvent servir à améliorer les estimations des ventes au détail mensuelles. Dans cet article, nous nous penchons tout d'abord sur le problème que pose l'étalonnage des séries chronologiques issues d'enquêtes économiques et nous analysons les solutions les plus appropriées dans les circonstances. Dans un deuxième temps, nous proposons deux nouvelles méthodes statistiques qui reposent sur un modèle non linéaire pour données infra-annuelles. Finalement, nous obtenons des estimations étalonnées en appliquant la méthode des moindres carrés pondérés.

MOTS CLÉS: Erreurs d'enquête; modèle non linéaire; moindres carrés pondérés.

1. INTRODUCTION

L'étalonnage a toujours été défini comme l'opération qui consiste à ajuster des valeurs mensuelles ou trimestrielles tirées d'une source particulière à des valeurs annuelles (données repères) tirées d'une autre source (voir Denton 1971, Cholette 1988a et Monsour et Trager 1979). Il peut s'agir, par exemple, de corriger les chiffres des expéditions mensuelles des manufacturiers canadiens de manière que leur somme égale la valeur des expéditions établie à l'aide de l'enquête annuelle sur les manufactures. L'étalonnage peut aussi se définir, d'une manière plus générale, comme l'opération qui consiste à améliorer les estimations infra- annuelles tirées d'une source particulière à l'aide d'estimations annuelles provenant d'une autre source (voir Hillmer et Trabelsi, 1987). Contrairement à la première définition, la seconde suppose que les valeurs annuelles peuvent être erronées. Il peut s'agir, par exemple, d'améliorer les estimations des stocks mensuels des détaillants canadiens, établies à l'aide d'une enquête par sondage, au moyen des données sur les stocks de fin d'année tirées de l'enquête annuelle par échantillon sur le commerce de détail. La seconde définition est celle qui s'applique le plus souvent aux séries économiques et c'est à elle que nous nous intéressons dans cet article.

Cet article se divise en deux parties. Premièrement, nous allons exposer en détail le problème de l'étalonnage tel qu'il se présente pour de nombreuses séries chronologiques issues des grandes enquêtes économiques. Ensuite, nous allons présenter et analyser deux méthodes d'étalonnage parmi les plus courantes pouvant traiter une série à la fois. Comme aucune de ces méthodes ne répond parfaitement aux exigences, nous en proposons deux autres basées sur les moindres carrés pondérés avec modèle non-linéaire. Enfin, nous donnons un exemple de deux des méthodes évoquées ci-dessus en nous servant de données simulées, puis nous en faisons l'analyse.

2. POSITION DU PROBLÈME

L'objet de cet article est de tenter d'améliorer des séries d'estimations infra-annuelles à l'aide de séries annuelles tirées des enquêtes-entreprises. Nous allons décrire ici les caractéristiques des données originales et ce que nous attendons d'une méthode d'étalonnage.

[1] Normand Laniel et Kimberley Fyfe, Division des méthodes d'enquêtes-entreprises, Statistique Canada, Ottawa (Ontario), K1A 0T6.

Les estimations infra-annuelles sont souvent entachées d'une erreur systématique à cause de problèmes de couverture. Le sous- dénombrement tient au fait que de nouvelles entreprises tardent à être intégrées à la base de sondage et que les entreprises qui n'ont pas d'employés (il s'agit le plus souvent de petites entreprises) ne figurent pas dans cette base. Les estimations infra-annuelles proviennent habituellement d'échantillons se chevauchants relativement petits, ce qui fait que les variances d'échantillonnage sont assez élevées et qu'il existe des covariances d'échantillonnage pour les estimations infra-annuelles qui se rapportent à des périodes différentes. En outre, la plupart des enquêtes économiques infra-annuelles produisent des séries d'estimations pour un certain nombre d'activités industrielles qui se déroulent dans un certain nombre de régions géographiques. Ces estimations sont publiées infra-annuellement dans des tableaux croisés (industrie $\times$ région géographique) dont il faut étalonner les diverses composantes (fréquences par case, totaux marginaux et total général).

Pour ce qui a trait aux estimations annuelles, on peut supposer qu'elles sont sans biais puisqu'en pratique, les bases de sondage correspondantes sont relativement peu touchées par les problèmes de couverture. En outre, les estimations annuelles proviennent habituellement de recensements ou d'échantillons d'enquête assez grands, ce qui fait que les erreurs d'échantillonnage rattachées à ces estimations sont relativement faibles ou inexistantes alors que les covariances d'échantillonnage tendent à être élevées à cause du fort taux de chevauchement des échantillons d'une année à l'autre. Une autre chose qu'il faut retenir au sujet des estimations annuelles est que ces chiffres sont produits environ deux ans après la période de référence de l'enquête. Par exemple, les données annuelles pour 1988 n'ont été diffusées qu'en 1990 tandis que les données infra-annuelles sont diffusées habituellement quelques mois suivant la période à laquelle elles se rapportent. Ainsi, lorsque vient le moment d'étalonner les estimations infra-annuelles, il y a des périodes infra-annuelles pour lesquelles nous n'avons pas de repères annuels.

Une méthode d'étalonnage doit présenter un certain nombre de caractéristiques pour pouvoir être appliquée à des estimations de grandes enquêtes. Premièrement, elle doit être suffisamment simple pour qu'on puisse l'utiliser sans avoir à recourir trop largement à l'analyse de données. Deuxièmement, elle doit pouvoir produire des facteurs d'étalonnage préliminaire pour des périodes pour lesquelles il n'existe pas encore de données repères. Cette caractéristique permet de faire de l'étalonnage à mesure que sont produites les données infra-annuelles, sinon des discontinuités sont introduites dans ces données. Enfin, il est souhaitable qu'une méthode d'étalonnage assure la concordance des totaux généraux, des totaux marginaux et des estimations par case pour des données étalonnées.

Pour une analyse plus approfondie des deux dernières caractéristiques, voir Laniel et Fyfe (1989, 1990) et Cholette (1988a, 1988b). Dans le reste de cet article, nous voyons comment étalonner une seule série chronologique dans les conditions décrites ci-dessus.

3. ÉTALONNAGE D'UNE SÉRIE CHRONOLOGIQUE

Nous allons exposer ci-dessous quatre méthodes qui peuvent servir à étalonner une série d'estimations infra-annuelles sur les flux ou les stocks.

3.1 Méthode de Denton

Dans son article de 1971, Denton a proposé des méthodes d'étalonnage fondées sur la minimisation quadratique. Chacune de ces méthodes correspond à une fonction de perte particulière et l'une de ces fonctions s'exprime sous forme d'écarts proportionnels entre la série originale et la série étalonnée et est souvent utilisée pour résoudre le problème d'étalonnage exposé dans la section 2. Nous pouvons présenter cette méthode de Denton en termes statistiques en affirmant tout d'abord que les estimations infra- annuelles sont expliquées par le modèle:

$$\frac{\theta_t}{y_t} = \frac{\theta_{t-1}}{y_{t-1}} + \epsilon_t, \qquad t=1, 2, \ldots, n \tag{3.1}$$

sujet à la restriction sur les données annuelles:

$$z_T = \sum_{t \epsilon T} \theta_t, \qquad T = 1, 2, \ldots, m, \tag{3.2}$$

où:

t désigne une période infra-annuelle,
T désigne une période annuelle,
$\{y_t\}$ est une suite d'estimations biaisées des paramètres infra-annuels (niveaux),
$\{\theta_t\}$ est une suite de paramètres infra-annuels fixes (valeurs réelles des niveaux),
$\{\epsilon_t\}$ est une suite d'erreurs non corrélées et identiquement distribuées avec comme vecteur de moyennes et matrice de covariances $(\mathbf{0}, \sigma^2 \boldsymbol{I})$,
$\{z_T\}$ est une suite de données repères annuelles.

Pour obtenir les estimations étalonnées, on applique les moindres carrés au modèle défini ci-dessus.

Précisons que la méthode de Denton suppose que le biais suit une marche aléatoire et que les données annuelles, aussi bien qu'infra-annuelles, ne sont entachées d'aucune erreur d'échantillonnage. Malheureusement, il est peu probable que ces hypothèses se vérifient avec des séries économiques (voir section 2).

3.2 Méthode de Hillmer et Trabelsi

En 1987, Hillmer et Trabelsi ont proposé une méthode d'étalonnage fondée sur les modèles ARMMI de Box et Jenkins (1976). Ils ont supposé que les estimations infra-annuelles étaient expliquées par le modèle:

$$y_t = \theta_t + u_t \qquad t = 1, 2, \ldots, n \tag{3.3}$$

et les estimations annuelles, par le modèle:

$$z_T = \sum_{t \epsilon T} \theta_t + a_T \qquad T = 1, 2, \ldots, m, \tag{3.4}$$

où:

$\{\theta_t\}$ est une suite de paramètres infra-annuels stochastiques (valeurs réelles des niveaux) qui suivent un modèle ARMMI,
$\{y_t\}$ est une suite d'estimations non biaisées des paramètres infra-annuels,
$\{u_t\}$ est une suite d'erreurs d'échantillonnage infra-annuelles corrélées avec comme vecteur de moyennes et matrice de covariances $(\mathbf{0}, \boldsymbol{\Sigma}_u)$,
$\{z_T\}$ est une suite d'estimations annuelles non biaisées,
$\{a_T\}$ est une suite d'erreurs d'échantillonnage annuelles corrélées avec comme vecteur de moyennes et matrice de covariances $(\mathbf{0}, \boldsymbol{\Sigma}_a)$.

Se servant de ces modèles, Hillmer et Trabelsi obtiennent des estimations infra-annuelles étalonnées en appliquant les moindres carrés stochastiques, c'est-à-dire en minimisant l'erreur quadratique moyenne, $E(\hat{\theta}_t - \theta_t)^2$. Dans la terminologie de l'analyse chronologique, cette méthode est aussi appelée extraction de signal et Hillmer et Trabelsi en font l'illustration dans leur article.

Comme l'indiquent les modèles ci-dessus, la méthode de Hillmer et Trabelsi tient compte des variances et covariances d'échantillonnage des estimations annuelles et infra-annuelles. Malheureusement, elle ne tient pas compte des erreurs systématiques que peuvent renfermer les estimations infra-annuelles. En outre, comme il s'agit de modèles ARMMI, il serait trop coûteux d'utiliser cette méthode pour de grandes enquêtes d'où sont tirées des centaines de séries de données. Il serait donc plus profitable de n'utiliser ce type de méthode que pour un petit nombre d'indicateurs économiques majeurs. Si les modèles ARMMI sont mal spécifiés, on risque par ailleurs d'obtenir un lissage excessif des données.

Cholette et Dagum (1989) ont modifié la méthode de Hillmer et Trabelsi en utilisant un modèle "d'intervention" au lieu d'un modèle ARMMI. Cela permet de modéliser les effets systématiques contenus dans la série chronologique. Selon les auteurs toutefois, cette méthode présente les mêmes lacunes que la méthode de Hillmer et Trabelsi.

3.3 Modèle pour tendances

La méthode suivante a été élaborée dans le but de satisfaire aux conditions d'étalonnage des enquêtes économiques. Elle repose sur l'hypothèse que les estimations infra-annuelles sont expliquées par le modèle:

$$\frac{y_t}{y_{t-1}} = \frac{\theta_t}{\theta_{t-1}} + v_t \qquad t = 1, 2, \ldots, n \tag{3.5}$$

et les estimations annuelles, par le modèle (3.4), où:

$\{\theta_t\}$ est une suite de paramètres infra-annuels (valeurs réelles), comme dans la méthode de Denton,

$\{v_t\}$ est une suite d'erreurs d'échantillonnage infra-annuelles corrélées des tendances, avec comme vecteur de moyennes et matrice de covariances $(\mathbf{0}, \boldsymbol{\Sigma}_v)$.

On obtient les estimations étalonnées en appliquant les moindres carrés aux modèles ci-dessus. L'algorithme de Gauss-Newton utilisé à cette fin ainsi que le calcul de la matrice des covariances des estimations étalonnées sont décrits dans Laniel et Fyfe (1989) ou (1990).

On peut utiliser cette méthode lorsque les données repères proviennent d'un recensement ou d'une enquête annuelle avec échantillons chevauchants et lorsque les estimations de niveau infra-annuelles sont biaisées, à la condition que le biais relatif soit fixe. En pratique, l'hypothèse d'un biais relatif constant se vérifiera si les opérations de mise à jour de la base de sondage se font à un rythme régulier, c'est-à-dire lorsque la proportion d'unités absentes de la base de sondage se stabilise avec les années. En outre, cette hypothèse suppose que les entreprises non énumérées ont le même comportement que celles qui figurent dans la base de sondage.

La méthode décrite ci-dessus pose toutefois un problème technique. En effet, on ne peut calculer directement la matrice des variances- covariances d'échantillonnage des tendances infra-annuelles et il faut donc recourir à une formule d'approximation. Après avoir utilisé l'approximation de Taylor du premier degré, on a constaté que dans certains cas, les variances et covariances d'échantillonnage étaient nulles ou négatives alors qu'elles devraient être positives. C'est pourquoi nous présentons dans la section suivante un modèle qui peut remplacer avantageusement le modèle (3.5).

3.4 Modèle pour niveaux

La méthode ci-dessous est une solution de remplacement pour la méthode précédente et est proposée dans le but d'obtenir plus facilement la matrice des variances-covariances d'échantillonnage des estimations infra-annuelles. Elle suppose que les estimations infra-annuelles sont expliquées par le modèle:

$$y_t = \alpha\theta_t + u_t \qquad t = 1, 2, \ldots, n, \tag{3.6}$$

où α est un paramètre fixe qui tient compte du biais relatif constant et u_t est défini comme dans l'équation (3.3). Les estimations annuelles sont expliquées par le modèle (3.4).

On obtient les estimations étalonnées en appliquant les moindres carrés aux modèles ci-dessus. L'algorithme pertinent est le même que celui utilisé pour la méthode décrite en 3.3.

3.5 Analyse

De toutes les méthodes exposées ci-dessus, la dernière est celle qui se prête le mieux à l'étalonnage d'une série chronologique tirée d'une grande enquête. Elle offre une base statistique qui permet de calculer des régions de confiance et de tester la validité de l'ajustement du modèle étalonné. Le choix du test doit être exercé avec soin puisque les estimations étalonnées, $\hat{\theta}_t$, ont un très petit nombre de degrés de liberté, $m - 1$ (nombre d'observations annuelles moins un), par rapport au nombre d'observations, $n + m$. De plus, le faible nombre de degrés de liberté donne à penser que le modèle pour niveaux produira vraisemblablement des estimations étalonnées qui afficheront le même profil chronologique que les estimations infra-annuelles.

Une question pratique que soulèvent actuellement les méthodes d'étalonnage qui tiennent compte des erreurs d'échantillonnage, comme celle décrite dans la section 3.4, est le calcul des covariances d'échantillonnage de deux estimations de niveau se rapportant à deux périodes différentes. Devrait-on modéliser ces covariances ou les calculer directement à l'aide du plan de sondage pour toutes les paires de périodes? Du point de vue théorique, il est préférable de les calculer directement puisque la suite des erreurs d'échantillonnage est en soi un processus stochastique non stationnaire étant donné que les variances-covariances de la population varient dans le temps. Toutefois, comme le calcul de l'ensemble des covariances d'échantillonnage peut s'avérer une entreprise laborieuse, il faut continuer de chercher une solution au problème du calcul des covariances d'échantillonnage.

3.6 Exemple

Afin de comparer la méthode de Denton, décrite dans la section 3.1, à celle qui repose sur le modèle pour niveaux, proposée dans la section 3.4, nous avons utilisé l'une et l'autre pour résoudre un problème d'étalonnage particulier et intéressant, où la variance d'échantillonnage des estimations annuelles est six fois plus élevée que la variance d'échantillonnage des estimations mensuelles correspondantes. Cet exemple fait ressortir clairement, pour ce cas-ci, la supériorité de la méthode fondée sur le modèle pour niveaux par rapport à la méthode de Denton.

Bien que ce problème soit observable en pratique, nous avons eu recours ici à des données simulées. Premièrement, nous avons tiré vingt-quatre estimations mensuelles d'une enquête économique existante, puis avons défini arbitrairement une matrice de covariances d'échantillonnage pour ces estimations. Les variances et les covariances ont été calculées au moyen d'un coefficient de variation unique pour toutes les périodes et du profil de corrélation suivant:

$$\rho_{ij} = 1 - \frac{|j - i|}{24} \quad \text{pour} \quad i = 1, 2, \ldots, 24 \quad \text{et} \quad j = 1, 2, \ldots, 24$$

où i et j sont les indices d'un couple d'estimations mensuelles. Ensuite, nous avons construit deux estimations annuelles correspondantes selon les critères suivants: dans le premier cas, une estimation 25% plus élevée que la somme des douze premières estimations mensuelles; dans le second cas, une estimation seulement 5% plus élevée que la somme des douze dernières estimations mensuelles. Enfin, pour ces deux estimations annuelles, nous avons défini des variances d'échantillonnage six fois plus élevées que celles des estimations mensuelles correspondantes et avons fixé leur coefficient de corrélation à 0.5.

Dans la figure 3.1, les estimations mensuelles sont représentées par la ligne continue et les estimations annuelles, par les traits horizontaux. Ces traits correspondent aux quotients des estimations annuelles par douze. Dans la même figure, la ligne en tirets représente la série d'estimations mensuelles étalonnées au moyen de la méthode fondée sur le modèle pour niveaux, tandis que la ligne en pointillés représente la série d'estimations mensuelles étalonnées à l'aide de la méthode de Denton.

Nous voyons d'après cette figure que la série étalonnée au moyen de la méthode fondée sur le modèle pour niveaux a le même taux de changement annuel que la série mensuelle originale, tandis que celle étalonnée au moyen de la méthode de Denton suit plutôt le mouvement des estimations annuelles. Nous constatons aussi que les deux séries étalonnées sont au-dessus de la série originale.

La différence de mouvement d'une année à l'autre entre les deux séries étalonnées peut s'expliquer de la façon suivante. Selon la méthode fondée sur le modèle pour niveaux, on obtient le mouvement de la série étalonnée en pondérant les estimations annuelles et infra-annuelles avec l'inverse des variances d'échantillonnage de ces estimations. Comme, dans cet exemple, les estimations infra-annuelles sont beaucoup plus précises que les estimations annuelles, la série étalonnée suit le mouvement des estimations mensuelles. En revanche, dans le cas de la méthode de Denton, le mouvement de la série étalonnée ne peut être axé que sur celui de la série d'estimations annuelles, peu importe le degré de précision de ces estimations. En ce sens, la méthode fondée sur le modèle pour niveaux est supérieure à la méthode de Denton.

Enfin, le fait que les deux séries étalonnées se trouvent au-dessus de la série originale confirme simplement que les deux méthodes renferment un mécanisme de correction pour le biais des estimations mensuelles.

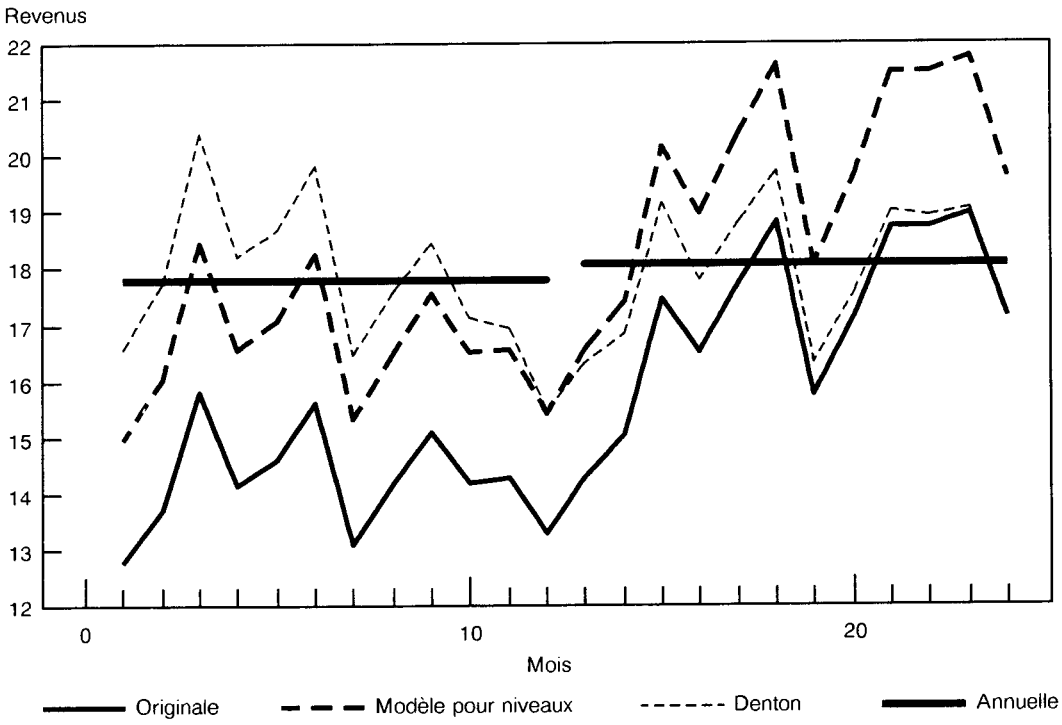

Figure 3.1 La série originale, les deux séries étalonnées, et la série annuelle

4. CONCLUSION

Dans cet article, nous avons vu comment améliorer des estimations d'enquête infra-annuelles au moyen d'estimations d'enquête annuelles. Nous avons présenté un moyen simple et inédit d'étalonner une série chronologique. La méthode en question peut être appliquée à l'aide d'un programme d'informatique de sorte que son exécution soit automatisée. Elle se distingue avantageusement des méthodes plus classiques (ex.: méthode de Denton) par le fait qu'elle tient compte des erreurs d'échantillonnage. Par la même occasion, nous avons abordé certaines questions liées à l'utilisation de cette méthode. Deux grandes questions pratiques ont été soulevées: l'étalonnage d'un tableau de données chronologiques et l'étalonnage préliminaire. Nous devons chercher des façons d'aborder ces deux questions.

BIBLIOGRAPHIE

BOX, G.E.P., et JENKINS, G.M. (1976). *Time Series Analysis, Forecasting and Control.* Holden-Day.

CHOLETTE, P.A. (1988a). Benchmarking and Interpolation of Time Series. Statistics Canada, Working Paper No. TSRA-87-014E.

CHOLETTE, P.A. (1988b). Benchmarking Systems of Socio-Economic Time Series. Statistique Canada, Document de travail N° TSRA-88-017E.

CHOLETTE, P.A., et DAGUM, E.B. (1989). Benchmarking Socio-Economic Time Series Data: A Unified Approach. Document de travail N° TSRA-89-006E, Statistique Canada.

DENTON, F.T. (1971). Adjustment on Monthly or Quarterly Series to Annual Totals: An approach Based on Quadratic Minimization. *Journal of the American Statistical Association*, 66, 99-102.

HILLMER, S.C., et TRABELSI, A. (1987). Benchmarking of Economic Time Series. *Journal of the American Statistical Association*, 82, 1604-1071.

LANIEL, N., et FYFE, K. (1989). Benchmarking of Economic Time Series. *Proceedings of the Business and Economic Statistics Section, American Statistical Association.*

LANIEL, N., et FYFE, K. (1990). Benchmarking of Economic Time Series. Méthodes d'enquêtes-entreprises, Statistique Canada, Document de travail du Projet de remaniement des enquêtes-entreprises.

MONSOUR, N.J., et TRAGER, M.L. (1979). Revision and Benchmarking of Business Time Series. *Proceedings of the Business and Economic Statistics Section, American Statistical Association.*