Techniques d'enquête, décembre 1993
Vol. 19, n° 2, pp. 175–182
Statistique Canada

Estimation d'un modèle d'étalonnage à biais multiplicatif constant par la méthode du maximun de vraisemblance et application

IJAZ U.H. MIAN et NORMAND LANIEL[1]

RÉSUMÉ

Nous considérons le cas de l'estimation d'un modèle d'étalonnage non linéaire par la méthode du maximum de vraisemblance, tel que le présentent Laniel et Fyfe (1989; 1990). Ce modèle tient compte des biais et des erreurs d'échantillonnage rattachés à la série originale. Comme on ne peut exprimer les estimateurs du maximum de vraisemblance des paramètres du modèle sous une forme analytique fermée, nous examinons deux méthodes itératives permettant de calculer les estimations du maximum de vraisemblance. Nous donnons aussi les expressions en forme analytique fermée pour les variances et les covariances asymptotiques des séries étalonnées et des valeurs ajustées. Pour illustrer les méthodes, nous nous servons de données publiées sur le commerce de détail au Canada.

MOTS CLÉS: Autocorrélations; modèle à biais; moindres carrés généralisés; erreurs d'échantillonnage.

1. INTRODUCTION

Les méthodes d'étalonnage servent très souvent à améliorer les estimations tirées d'enquêtes infra-annuelles; on utilise à cette fin les estimations correspondantes tirées d'enquêtes annuelles et que l'on appelle des estimations repères. Cette opération se traduit généralement par une réduction du biais et de la variance des estimations infra-annuelles. Par exemple, les estimations tirées des enquêtes annuelles sur le commerce de détail peuvent servir à améliorer les estimations mensuelles du commerce de détail. Les estimations infra-annuelles renferment souvent un biais à cause des lacunes de la base de sondage au point de vue de la couverture. Le sous-dénombrement est attribuable au fait que les nouvelles entreprises tardent à être incluses dans la base de sondage et à l'absence, dans cette base, des entreprises sans salariés. En outre, la variance des estimations infra-annuelles est souvent plus élevée que celle des estimations annuelles correspondantes et on observe des covariances d'échantillonnage pour les estimations de périodes différentes à cause du chevauchement des échantillons. En revanche, on peut supposer les estimations annuelles non biaisées parce que, dans la pratique, les bases de sondage correspondantes présentent peu de lacunes au point de vue de la couverture. On peut trouver des discussions détaillées sur l'étalonnage dans Laniel et Fyfe (1989; 1990), Cholette (1987; 1988) et d'autres ouvrages.

La littérature statistique présente plusieurs méthodes d'étalonnage des séries chronologiques. Dans une perspective de minimisation quadratique, Denton (1971) propose plusieurs méthodes pour étalonner une série temporelle unique. Cholette (1984), à son tour, propose une variante de la méthode proportionnelle d'ordre un de Denton, par laquelle il supprime la condition de départ dans le but d'éviter les effets transitoires. Les hypothèses que posent les auteurs ont très peu de chances d'être vérifiées avec la plupart des séries économiques. En effet, les modèles que ces auteurs élaborent supposent que le biais rattaché aux estimations infra-annuelles suit une marche aléatoire et que les observations infra-annuelles et annuelles ne renferment aucune erreur d'échantillonnage. En règle générale, les estimations proviennent d'enquêtes par sondage et sont donc exposées à l'erreur d'échantillonnage.

Hillmer et Trabelsi (1987) ont proposé une autre approche pour l'étalonnage, qui repose sur un modèle ARMMI (voir, par exemple, Box et Jenkins, 1976). Bien que cette approche tienne compte des covariances d'échantillonnage des estimations infra-annuelles et annuelles, elle fait abstraction du biais dans les estimations infra-annuelles. Cholette et Dagum (1989) ont modifié l'approche de Hillmer et Trabelsi en remplaçant le modèle ARMMI par un modèle d'"intervention". Cette variante permet la modélisation d'effets systématiques dans les séries temporelles mais elle présente néanmoins les mêmes lacunes que le modèle de Hillmer et Trabelsi (Laniel et Fyfe 1990).

Afin de suppléer les lacunes mentionnées plus haut, Laniel et Fyfe (1989; 1990) ont proposé un modèle d'étalonnage non linéaire pour les mesures de niveau. Ils ont élaboré un algorithme complexe pour calculer les estimations par les moindres carrés généralisés (MCG) (et les covariances asymptotiques correspondantes) des paramètres du modèle. Ce modèle tient compte des covariances d'échantillonnage des estimations infra-annuelles et annuelles et peut être utilisé lorsque les données repères proviennent soit de recensements ou d'échantillons chevauchants annuels. Il suppose aussi l'existence d'un biais (relatif) multiplicatif constant dans les estimations infra-annuelles de niveau. D'autres modèles d'étalonnage à biais multiplicatif constant ont été proposés par Cholette (1992) et Laniel et Mian (1991). Cholette suppose un modèle dans

[1] Ijaz U.H. Mian et Normand Laniel, Division des méthodes d'enquêtes sociales, Statistique Canada, Ottawa, Ontario, K1A 0T6, Canada.

lequel non seulement le biais mais aussi les erreurs sont multiplicatives. L'auteur se sert de la théorie des MCG pour calculer les estimations des paramètres du modèle après avoir soumis ce modèle à une transformation logarithmique. Laniel et Mian (1991) ont défini un algorithme pour calculer les estimations du maximum de vraisemblance d'un modèle d'étalonnage à biais multiplicatif constant avec données-repères mixtes (c.-à-d. un mélange de repères fermes et de repères non fermes). Les données-repères fermes sont en l'occurrence des estimations tirées d'un recensement (c.-à-d. des estimations à variance nulle) tandis que les données-repères non fermes sont des estimations fondées sur un échantillon. L'hypothèse de l'existence d'un biais multiplicatif constant se vérifiera dans la pratique si la fréquence de mise à jour de la base de sondage est relativement stable, c'est-à-dire si le taux de sous-dénombrement pour cette base varie très peu d'une année à l'autre. Cette hypothèse implique aussi que le rapport moyen des valeurs de la variable étudiée d'une période à l'autre est le même pour les entreprises qui sont incluses dans la base de sondage et celles qui ne le sont pas. La nature du biais rattaché aux estimations infra-annuelles peut varier d'une série temporelle à l'autre. Cholette et Dagum (1991) ont proposé une méthode d'étalonnage qui suppose l'existence d'un biais additif constant dans les estimations infra-annuelles.

L'objet de cet article est d'estimer les paramètres du modèle de Laniel et Fyfe par la méthode du maximum de vraisemblance (MV); les résultats s'inspirent du rapport de Mian et Laniel (1991). Le modèle en question est décrit dans la section suivante. Ensuite, nous étudions deux processus itératifs qui permettent de calculer les estimations du maximum de vraisemblance (EMV) des paramètres du modèle. Dans la section 4, nous donnons les expressions en forme analytique fermée pour les covariances asymptotiques des estimateurs des paramètres du modèle et des valeurs ajustées. À titre d'exemple, nous nous servons des données publiées par Statistique Canada sur le commerce de détail au Canada.

2. MODÈLE A BIAIS MULTIPLICATIF CONSTANT (MBMC)

Afin de répondre aux exigences des enquêtes économiques en matière d'étalonnage, Laniel et Fyfe (1989; 1990) ont proposé le modèle à biais multiplicatif constant (MBMC) défini ci-dessous. Ce modèle suppose que les estimations infra-annuelles biaisées y_t répondent à la relation

$$y_t = \beta\theta_t + a_t, \quad t = 1, 2, \ldots, n \tag{2.1}$$

et que les estimations annuelles non biaisées z_T répondent à la relation

$$z_T = \sum_{t \in T} \theta_t + b_T, \quad T = 1, 2, \ldots, m, \tag{2.2}$$

où les indices t et T désignent respectivement les périodes infra-annuelles et les périodes annuelles, θ_t est le paramètre (inconnu) infra-annuel fixe, β est un paramètre (inconnu) de biais constant rattaché à y_t, et a_t et b_T sont les erreurs d'échantillonnage rattachées à y_t et à z_T respectivement. Le modèle ci-dessus est un modèle de type hybride (mixte), dans lequel le biais est multiplicatif mais les erreurs, additives.

Avant d'aller plus loin, définissons les vecteurs colonnes $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$, $\boldsymbol{z} = (z_1, z_2, \ldots, z_m)'$, $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)'$, $\boldsymbol{b} = (b_1, b_2, \ldots, b_m)'$, et $\boldsymbol{\Theta} = (\theta_1, \theta_2, \ldots, \theta_n)'$. Le MBMC, défini par les équations (2.1) et (2.2), peut être réexprimé sous la forme

$$\begin{aligned} \boldsymbol{w} &= \boldsymbol{X}_\beta \boldsymbol{\Theta} + \boldsymbol{u} \\ &= \boldsymbol{X}_\Theta \boldsymbol{\beta} + \boldsymbol{X}_D \boldsymbol{\Theta} + \boldsymbol{u}, \end{aligned} \tag{2.3}$$

où

$$\boldsymbol{X}_\beta = (\beta \boldsymbol{I}_n : \boldsymbol{D}')', \quad \boldsymbol{X}_\Theta = (\boldsymbol{\Theta}' : \boldsymbol{0}')', \quad \boldsymbol{X}_D = (\boldsymbol{0}' : \boldsymbol{D}')',$$
$$\boldsymbol{w} = (\boldsymbol{y}' : \boldsymbol{z}')', \quad \boldsymbol{u} = (\boldsymbol{a}' : \boldsymbol{b}')', \quad \boldsymbol{D} = (d_{Tt}), \tag{2.4}$$

$\boldsymbol{I}_n$ étant une matrice unité d'ordre n, $\boldsymbol{0}$ étant un vecteur ou une matrice nuls d'ordre approprié et d_{Tt} étant une fonction indicatrice égale à 1 pour $t \in T$ et égale à 0 dans le cas contraire. On suppose que les vecteurs d'erreurs d'échantillonnage $\boldsymbol{a}$ et $\boldsymbol{b}$ suivent une distribution normale multidimensionnelle telle que $\boldsymbol{a} \sim MN(\boldsymbol{0}, \boldsymbol{V}_{aa})$ et $\boldsymbol{b} \sim MN(\boldsymbol{0}, \boldsymbol{V}_{bb})$. De plus, dans le cas général, $\boldsymbol{a}$ et $\boldsymbol{b}$ sont corrélés, ce qui signifie que $\text{Cov}(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{V}_{ab} = \boldsymbol{V}_{ba}' \neq \boldsymbol{0}$. Nous allons voir dans la section suivante que, pour ce modèle, l'estimateur du maximum de vraisemblance (MV) de $\boldsymbol{\Theta}$ et de β est identique à l'estimateur des moindres carrés généralisé (MCG) des mêmes paramètres. Par conséquent, l'hypothèse concernant la normalité de $\boldsymbol{a}$ et $\boldsymbol{b}$ n'est nécessaire que pour obtenir la matrice d'information de Fisher (et, donc, les variances) des EMV.

3. ESTIMATION PAR LA MÉTHODE DU MAXIMUM DE VRAISEMBLANCE

Suivant le MBMC, la fonction de vraisemblance logarithmique peut s'écrire

$$\ln(L) = -\frac{(n+m)}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{V}| - \frac{1}{2}Q, \tag{3.1}$$

où

$$Q = (\boldsymbol{w} - \boldsymbol{X}_\beta \boldsymbol{\Theta})' \boldsymbol{V}^{-1} (\boldsymbol{w} - \boldsymbol{X}_\beta \boldsymbol{\Theta}) \tag{3.2}$$

et

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{aa} & \boldsymbol{V}_{ab} \\ \boldsymbol{V}_{ba} & \boldsymbol{V}_{bb} \end{pmatrix}.$$

On peut obtenir les estimations du maximum de vraisemblance des paramètres du modèle $\boldsymbol{\Theta}$ et β (si l'on suppose $\boldsymbol{V}$ connu) en maximisant la fonction de vraisemblance

logarithmique (3.1) ou, ce qui revient au même, en minimisant le terme quadratique Q (3.2). Pour ce modèle en particulier, les estimateurs du MV et les estimateurs des MCG des paramètres du modèle sont les mêmes et nous n'en ferons la distinction que si le besoin s'en fait sentir. Si nous calculons les dérivées partielles d'ordre un de $\ln(L)$ par rapport à $\boldsymbol{\Theta}$ et à β respectivement et si nous posons le résultat égal à zéro, nous avons

$$\frac{\partial \ln(L)}{\partial \boldsymbol{\Theta}} = X_\beta' V^{-1}(w - X_\beta \boldsymbol{\Theta}) = \mathbf{0},$$
$$\frac{\partial \ln(L)}{\partial \beta} = X_\Theta' V^{-1}(w - X_\beta \boldsymbol{\Theta}) = 0. \tag{3.3}$$

Puisque $\mathrm{E}(w) = X_\beta \boldsymbol{\Theta}$ selon le modèle (2.3), les équations ci-dessus sont des équations d'estimation au sens de Godambe (1960) et elles sont non biaisées du point de vue de l'information. Il est intéressant de noter que $X_\beta' V^{-1}$ et $X_\Theta' V^{-1}$ ne dépendent pas de w, de sorte que les équations (3.3) convergent vers zéro et ont, par conséquent, des racines consistantes pourvu que $\mathrm{E}(w) = X_\beta \boldsymbol{\Theta}$. Autrement dit, même si V dans les équations ci-dessus est remplacé par une de ses estimations consistantes, les équations donneront des estimations consistantes des vecteurs $\boldsymbol{\Theta}$ et β. Il convient aussi de noter que ces équations sont non linéaires par rapport aux paramètres à estimer et qu'on ne peut obtenir d'expression formelle pour les estimateurs de $\boldsymbol{\Theta}$ et de β. On peut donc recourir à une méthode itérative comme la méthode bien connue de Fisher-Newton-Raphson (aussi appelée méthode "de la fonction de caractérisation" de Fisher) pour calculer les estimations. La section 4 donne les éléments de la matrice d'information de Fisher espérée nécessaire à l'application de la méthode de Fisher-Newton-Raphson.

Une autre manière de calculer les EMV des paramètres du modèle est de résoudre successivement les équations d'estimation (3.3). Si nous résolvons la première expression de (3.3), l'estimateur de $\boldsymbol{\Theta}$, comme fonction de β, est défini

$$\hat{\boldsymbol{\Theta}} \equiv \hat{\boldsymbol{\Theta}}(\beta) = (X_\beta' V^{-1} X_\beta)^{-1} X_\beta' V^{-1} w. \tag{3.4}$$

De même, si nous résolvons la seconde expression de (3.3), l'estimateur de β, comme fonction de $\boldsymbol{\Theta}$, est défini

$$\hat{\beta} \equiv \hat{\beta}(\boldsymbol{\Theta}) = \left[\boldsymbol{\Theta}' V_{aa.b}^{-1}(y - V_{ab} V_{bb}^{-1}(z - D\boldsymbol{\Theta}))\right] / \left[\boldsymbol{\Theta}' V_{aa.b}^{-1} \boldsymbol{\Theta}\right], \tag{3.5}$$

où

$$V_{aa.b} = V_{aa} - V_{ab} V_{bb}^{-1} V_{ba}.$$

On obtient les estimations du MV de $\boldsymbol{\Theta}$ et de β en calculant successivement les équations (3.4) et (3.5) jusqu'à convergence. L'avantage de cette méthode par rapport à la méthode de Fisher-Newton-Raphson est qu'elle est facile à appliquer. Cependant, la convergence est ordinairement très lente pour ce genre d'algorithme. Dans la section 6, nous comparerons ces deux méthodes dans le but de vérifier leur vitesse de convergence.

Une fois que l'on connaît les estimations du MV des paramètres du modèle, on peut déterminer les valeurs infra-annuelles ajustées $\hat{y} = \hat{\beta}\hat{\boldsymbol{\Theta}}$ et les valeurs annuelles ajustées $\hat{z} = D\hat{\boldsymbol{\Theta}}$.

Estimation initiale de $\boldsymbol{\Theta}$ et β

Afin d'établir une estimation initiale pour β, disons $\hat{\beta}_0$, réécrivons le modèle (2.3) comme suit:

$$w^* = X_\Theta^* \beta + u^*,$$

où $w^* = ((Dy)' : (z - D\boldsymbol{\Theta})')'$, $X_\Theta^* = \left((D\boldsymbol{\Theta})' : \mathbf{0}'\right)'$ et $u^* = \left((Da)' : b'\right)'$. L'EMV de β est donc défini par l'expression

$$\hat{\beta} = \left[X_\Theta^{*\prime} (V^*)^{-1} w^*\right] / \left[X_\Theta^{*\prime} (V^*)^{-1} X_\Theta^*\right], \tag{3.6}$$

où

$$V^* = \mathrm{Cov}(u^*) = \begin{pmatrix} DV_{aa} D' & DV_{ab} \\ V_{ba} D' & V_{bb} \end{pmatrix}.$$

Si l'on tient compte de ce que $\mathrm{E}(z) = D\boldsymbol{\Theta}$ et si on remplace $D\boldsymbol{\Theta}$ par z dans (3.6), on peut établir une estimation initiale pour β au moyen de la formule

$$\hat{\beta}_0 = \left[\begin{pmatrix} z \\ \mathbf{0} \end{pmatrix}' (V^*)^{-1} w^*\right] \Big/ \left[\begin{pmatrix} z \\ \mathbf{0} \end{pmatrix}' (V^*)^{-1} \begin{pmatrix} z \\ \mathbf{0} \end{pmatrix}\right]$$
$$= \left[z' (DV_{aa.b} D')^{-1} Dy\right] / \left[z' (DV_{aa.b} D')^{-1} z\right]. \tag{3.7}$$

On peut établir une estimation initiale pour à l'aide de l'équation (3.4) en remplaçant β par $\hat{\beta}_0$.

4. COVARIANCES DES ESTIMATEURS

Dans cette section, nous établissons les formules des covariances asymptotiques des estimateurs du MV des paramètres du MBMC en inversant la matrice d'information de Fisher, dénotée par $\boldsymbol{\Omega}$. Les covariances asymptotiques des valeurs annuelles et infra-annuelles ajustées sont calculées au moyen de la méthode delta. Considérons tout d'abord le calcul des covariances des estimateurs du MV de $\boldsymbol{\Theta}$ et β. Les éléments de $\boldsymbol{\Omega}$ (c.-à-d. les espérances négatives des dérivées partielles d'ordre deux de $\ln(L)$) sont définis par les expressions

$$\Omega_{11} = -\mathrm{E}\left[\frac{\partial^2 \ln(L)}{\partial \boldsymbol{\Theta}\, \partial \boldsymbol{\Theta}'}\right] = X_\beta' V^{-1} X_\beta,$$

$$\Omega_{22} = -\mathrm{E}\left[\frac{\partial^2 \ln(L)}{\partial \beta^2}\right] = \Theta' V_{aa.b}^{-1} \Theta$$

et

$$\Omega_{12} = \Omega_{21}' = -\mathrm{E}\left[\frac{\partial^2 \ln(L)}{\partial \Theta \, \partial \beta}\right] = X_\beta' V^{-1} X_\Theta.$$

Par conséquent, la matrice d'information de Fisher d'ordre $(n+1) \times (n+1)$ est définie

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}. \tag{4.1}$$

Si nous inversons Ω au moyen de l'algèbre de partition des matrices, nous avons

$$\begin{aligned} \mathrm{Cov}(\hat{\Theta}) &= \Omega_{11.2}^{-1}, \\ \mathrm{Var}(\hat{\beta}) &= \Omega_{22.1}^{-1}, \\ \mathrm{Cov}(\hat{\Theta}, \hat{\beta}) &= -\Omega_{11.2}^{-1} \Omega_{12} \Omega_{22}^{-1} \\ &= -\Omega_{11}^{-1} \Omega_{12} \Omega_{22.1}^{-1}, \end{aligned} \tag{4.2}$$

où

$$\begin{aligned} \Omega_{11.2} &= \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21}, \\ \Omega_{22.1} &= \Omega_{22} - \Omega_{21} \Omega_{11}^{-1} \Omega_{12}. \end{aligned} \tag{4.3}$$

Une fois que la matrice de covariances Ω^{-1} est obtenue, on peut calculer les covariances asymptotiques des valeurs infra-annuelles ajustées $\hat{y}$ à l'aide de la méthode delta (voir, par ex., Rao 1973). Soit Δ la matrice des dérivées partielles d'ordre un de y par rapport aux éléments de $(\Theta' : \beta)'$. De toute évidence, la matrice $n \times (n+1)$ est $\Delta = (\beta I_n : \Theta)$. Si on utilise la méthode delta, la matrice des covariances asymptotiques de $\hat{y}$ est définie par l'expression

$$\mathrm{Cov}(\hat{y}) = \Delta \Omega^{-1} \Delta'. \tag{4.4}$$

En outre, la matrice des covariances des valeurs annuelles ajustées $\hat{z}$, d'après la théorie de la distribution normale centrée réduite à plusieurs variables, est définie par l'expression

$$\mathrm{Cov}(\hat{z}) = D \Omega_{11.2}^{-1} D', \tag{4.5}$$

où D et $\Omega_{11.2}$ sont définies en (2.4) et en (4.3) respectivement.

5. ESTIMATION PAR LA MÉTHODE DU MAXIMUM DE VRAISEMBLANCE LORSQUE $V_{ab} = 0$

Dans cette section, nous examinons l'estimation par le maximum de vraisemblance des paramètres du modèle dans le cas particulier où les vecteurs d'erreurs $\boldsymbol{a}$ et $\boldsymbol{b}$ sont non corrélés (c.-à-d., $\mathrm{Cov}(\boldsymbol{a}, \boldsymbol{b}) = V_{ab} = V_{ba}' = 0$). Cette situation est observée normalement dans les enquêtes par sondage lorsque les échantillons infra-annuels et annuels sont tirés de façon indépendante. On peut voir un changement dans les résultats des sections 3 et 4 en substituant $V_{ab} = V_{ba}' = 0$ dans les équations. Par exemple, dans ce cas particulier, les estimateurs du MV de Θ et de β, définis en (3.4) et en (3.5), se réduisent à

$$\hat{\Theta}^* \equiv \hat{\Theta}^*(\beta) = \left(\beta^2 V_{aa}^{-1} + D' V_{bb}^{-1} D\right)^{-1} \left(\beta V_{aa}^{-1} y + D' V_{bb}^{-1} z\right)$$

et

$$\hat{\beta}^* \equiv \hat{\beta}^*(\Theta) = \left[\Theta' V_{aa}^{-1} y\right] / \left[\Theta' V_{aa}^{-1} \Theta\right],$$

respectivement. On doit résoudre ces équations successivement pour obtenir les estimations voulues.

De la même manière, les éléments de la matrice d'information de Fisher se ramènent à

$$\begin{aligned} \Omega_{11}^* &= \beta^2 V_{aa}^{-1} + D' V_{bb}^{-1} D, \\ \Omega_{22}^* &= \Theta' V_{aa}^{-1} \Theta, \\ \Omega_{12}^* &= \Omega_{21}^{*\prime} = \beta V_{aa}^{-1} \Theta. \end{aligned}$$

6. APPLICATION

Dans cette section, nous présentons un exemple qui utilise des données publiées sur le commerce de détail au Canada; ces données viennent des enquêtes mensuelles et annuelles qu'effectue Statistique Canada sur le commerce de détail. Les estimations mensuelles sur le commerce de détail et les coefficients de variation (CV) correspondants se trouvent dans la publication de Statistique Canada intitulée Commerce de détail (n° 63-005 au catalogue, mensuel). Il existe deux catégories d'estimations mensuelles sur le commerce de détail, à savoir les estimations provisoires et les estimations révisées. Dans l'exemple qui suit, nous avons utilisé les estimations révisées mais non désaisonnalisées (brutes). Comme les CV des estimations révisées ne sont pas disponibles, nous nous sommes servis des CV des estimations provisoires pour calculer une approximation de la variance des estimations mensuelles révisées. Les données utilisées portent sur la période de janvier 1985 à décembre 1988. Une autre difficulté que comporte cet exemple est de déterminer les coefficients d'autocorrélation

entre les estimations mensuelles. Se fondant sur des données mensuelles sur le commerce de détail, Hidiroglou et Giroux (1986) ont calculé les valeurs estimées des coefficients d'autocorrélation aux décalages 1, 3, 6, 9 et 12 pour trois sortes de strate dans plusieurs provinces du Canada. La moyenne des estimations des coefficients d'autocorrélation entre les estimations mensuelles sur le commerce de détail pour les strates formées de la province de l'Ontario et du code 60 de la Classification type des industries (industries des aliments, boissons et médicaments) a servi d'approximation pour les coefficients d'autocorrélation entre les estimations mensuelles sur le commerce de détail. Ces coefficients (moyens) d'autocorrélation, c'est-à-dire $\rho(k)$, figurent dans le tableau 1.

Tableau 1

Coefficients d'autocorrélation approximatifs $\rho(k)$ pour les estimations mensuelles sur le commerce de détail

Décalage k	1	3	6	9	12
$\rho(k)$	0.970	0.940	0.918	0.914	0.962

La méthode des moindres carrés ordinaires ainsi qu'un algorithme de McLeod (1975) pour le calcul de coefficients d'autocorrélation théoriques pour des séries temporelles autorégressives à moyennes mobiles ont servi à réviser les coefficients d'autocorrélation observés. On a ajusté un modèle multiplicatif saisonnier ARMA $(1,0)(1,0)_{12}$ aux cinq coefficients observés en minimisant la somme des carrés des écarts entre le coefficient d'autocorrélation observé et le coefficient d'autocorrélation théorique. On a ensuite calculé les coefficients d'autocorrélation pour tous les autres décalages pertinents à l'aide des paramètres du modèle estimés et de l'algorithme de McLeod mentionné plus haut. Étant donné que le modèle ARMA est correct pour les coefficients d'autocorrélation théoriques, cette méthode produit une estimation consistante de la fonction d'autocorrélation. Le tableau 2 donne les coefficients d'autocorrélation approximatifs révisés pour 47 décalages; ces coefficients ont servi à calculer approximativement les covariances des estimations mensuelles sur le commerce de détail par multiplication avec les écarts-types.

Au moment de l'étude des estimations annuelles sur le commerce de détail existaient uniquement pour les années 1985 à 1988. On peut trouver ces estimations dans la publication de Statistique Canada intitulée Commerce de détail annuel (n° 63-223 au catalogue, annuel). Les variances correspondantes ne sont pas publiées; on a donc dû les calculer à l'aide des données de l'enquête. Les covariances entre les estimations mensuelles et annuelles sont nulles parce que les échantillons des enquêtes mensuelle et annuelle sur le commerce de détail ont été tirés indépendamment l'un de l'autre. Comme les estimations annuelles sur le commerce de détail reposent sur des échantillons qui ne sont pas indépendants, leur covariance est non nulle. Or, les opérations normales de traitement des données d'enquête ne permettent pas de connaître les estimations des covariances; il faudrait une étude pour les obtenir. Par conséquent, pour les besoins de notre exemple, nous avons supposé que les covariances des estimations annuelles sur le commerce de détail sont nulles.

Tableau 2

Coefficients d'autocorrélation approximatifs révisés $\rho^*(k)$ pour les estimations mensuelles sur le commerce de détail (calculés pour 47 décalages)

Décalage k	$\rho^*(k)$	Décalage k	$\rho^*(k)$	Décalage k	$\rho^*(k)$	Décalage k	$\rho^*(k)$
0	1.0000	12	0.9602	24	0.8896	36	0.8100
1	0.9758	13	0.9345	25	0.8647	37	0.7869
2	0.9555	14	0.9126	26	0.8433	38	0.7669
3	0.9391	15	0.8943	27	0.8253	39	0.7501
4	0.9266	16	0.8798	28	0.8107	40	0.7363
5	0.9177	17	0.8687	29	0.7994	41	0.7254
6	0.9126	18	0.8612	30	0.7913	42	0.7176
7	0.9113	19	0.8572	31	0.7864	43	0.7126
8	0.9136	20	0.8567	32	0.7843	44	0.7106
9	0.9196	21	0.8595	33	0.7862	45	0.7114
10	0.9293	22	0.8661	34	0.7909	46	0.7151
11	0.9429	23	0.8760	35	0.7989	47	0.7217

Un des arbitres a soulevé une question intéressante: qu'arrive-t-il lorsque les variances et les covariances des estimations de l'enquête sont inconnues? C'est une question difficile, à laquelle on ne peut répondre sans une réflexion approfondie. Cependant, le modèle présenté ici suppose que ces variances et covariances sont connues. En règle générale, il suffit que les équations d'estimation utilisées pour calculer les estimations du MV soient des estimateurs consistants des variances et des covariances. Il est courant, en étalonnage, d'estimer ces variances et ces covariances à l'aide des données de l'enquête puisqu'on ne connaît jamais les valeurs théoriques (voir, par exemple, Hillmer et Trabelsi 1987).

Les calculs requis pour notre exemple sont exécutés par un algorithme rédigé dans le langage de programmation GAUSS pour micro-ordinateurs. L'estimation initiale de β pour le processus itératif, d'après (3.7), est $\hat{\beta}_0 = 0.9162$. L'estimation initiale du vecteur de paramètres $\boldsymbol{\theta}$ est calculée à l'aide de (3.4), après que l'on a remplacé β par $\hat{\beta}_0$. La méthode de Fisher-Newton-Raphson et la méthode d'itération successive, dont il a été question dans la section 3, servent toutes deux à calculer les estimations du MV des paramètres du modèle. L'estimation du MV finale de β se rapproche sensiblement de l'estimation initiale et $\hat{\beta}$ est égale à 0.9016, avec un CV de 0.0065. Il est intéressant de constater que, dans cet exemple, la méthode de Fisher-

Tableau 3

Estimations mensuelles sur le commerce de détail, estimations du MV de Θ_t's et des valeurs ajustées (en millions de dollars), et CV correspondants

Année	Mois	y_t*	CV(y_t)*	$\hat{\Theta}_t$	CV($\hat{\Theta}_t$)	$\hat{y}_t$	CV($\hat{y}_t$)
1985	1	8,689.668	0.008	9,686.630	0.00210	8,733.384	0.00667
	2	8,390.380	0.008	9,350.078	0.00210	8,429.951	0.00665
	3	10,107.485	0.006	11,248.048	0.00233	10,141.146	0.00496
	4	10,541.145	0.008	11,741.785	0.00200	10,586.294	0.00656
	5	11,763.659	0.007	13,094.151	0.00198	11,805.576	0.00570
	6	11,067.487	0.008	12,321.326	0.00189	11,108.803	0.00647
	7	10,810.755	0.008	12,029.467	0.00184	10,845.666	0.00643
	8	11,289.656	0.009	12,554.808	0.00206	11,319.309	0.00726
	9	10,336.540	0.009	11,484.216	0.00205	10,354.073	0.00728
	10	11,213.751	0.010	12,447.696	0.00256	11,222.737	0.00809
	11	11,935.495	0.010	13,234.412	0.00258	11,932.034	0.00808
	12	13,300.288	0.008	14,734.891	0.00188	13,284.853	0.00643
1986	1	9,753.373	0.009	10,794.009	0.00221	9,731.787	0.00716
	2	9,249.279	0.009	10,227.777	0.00224	9,221.277	0.00709
	3	10,609.952	0.008	11,729.293	0.00207	10,575.031	0.00622
	4	11,637.936	0.008	12,860.626	0.00206	11,595.032	0.00614
	5	12,695.108	0.008	14,024.139	0.00205	12,644.046	0.00605
	6	11,826.254	0.008	13,059.556	0.00202	11,774.385	0.00598
	7	11,940.908	0.010	13,164.500	0.00233	11,869.002	0.00740
	8	11,866.547	0.010	13,070.205	0.00232	11,783.987	0.00743
	9	11,540.397	0.009	12,712.283	0.00202	11,461.287	0.00670
	10	12,208.845	0.010	13,430.932	0.00235	12,109.215	0.00747
	11	12,201.498	0.010	13,418.219	0.00240	12,097.753	0.00747
	12	14,479.170	0.009	15,933.951	0.00215	14,365.916	0.00670
1987	1	10,271.723	0.012	11,276.676	0.00357	10,166.956	0.00891
	2	9,951.105	0.010	10,945.319	0.00261	9,868.208	0.00737
	3	11,492.162	0.008	12,663.849	0.00230	11,417.620	0.00584
	4	12,867.443	0.009	14,172.605	0.00235	12,777.901	0.00652
	5	13,508.434	0.012	14,850.145	0.00343	13,388.765	0.00862
	6	13,608.274	0.011	14,973.985	0.00287	13,500.418	0.00786
	7	13,278.474	0.023	14,483.340	0.01066	13,058.057	0.00165
	8	12,728.196	0.008	14,028.998	0.00227	12,648.426	0.00577
	9	12,616.239	0.009	13,888.982	0.00233	12,522.188	0.00659
	10	13,760.829	0.008	15,156.409	0.00227	13,664.890	0.00592
	11	13,380.142	0.008	14,733.240	0.00227	13,283.365	0.00597
	12	16,269.757	0.007	17,928.148	0.00241	16,163.867	0.00525
1988	1	11,134.013	0.010	12,234.529	0.00274	11,030.548	0.00753
	2	10,959.374	0.010	12,042.761	0.00276	10,857.651	0.00754
	3	13,177.788	0.008	14,508.565	0.00233	13,080.800	0.00602
	4	13,666.311	0.009	15,035.737	0.00243	13,556.094	0.00676
	5	14,267.530	0.006	15,742.039	0.00379	14,192.890	0.00448
	6	14,432.944	0.009	15,884.130	0.00240	14,320.997	0.00673
	7	13,960.825	0.009	15,363.957	0.00240	13,852.014	0.00673
	8	13,691.315	0.008	15,073.691	0.00233	13,590.312	0.00606
	9	13,773.109	0.008	15,159.075	0.00235	13,667.294	0.00613
	10	13,900.743	0.009	15,279.950	0.00255	13,776.282	0.00696
	11	14,453.461	0.009	15,884.279	0.00260	14,321.132	0.00700
	12	17,772.990	0.009	19,529.791	0.00267	17,607.895	0.00702

*__Source__: Commerce de détail, publication de Statistique Canada (n° 63-005 au catalogue, mensuel).

Newton-Raphson a convergé très rapidement vers une solution. De fait, elle a convergé en 6 itérations seulement (environ 1 minute) - avec une précision de dix chiffres - tandis que la méthode des calculs successifs a convergé, pour le même degré de précision, au bout de 500 itérations et plus (plus de 45 minutes) sur un ordinateur personnel 386DX-25Mhz. Cependant, comme c'était prévu, les deux méthodes ont convergé vers la même solution finale. On obtient la matrice des covariances du vecteur estimé $(\hat{\boldsymbol{\theta}}' : \hat{\beta})'$ en inversant la matrice d'information de Fisher $\boldsymbol{\Omega}$, définie en (4.1), après avoir remplacé les paramètres par les estimations du MV correspondantes. Le tableau 3 donne la série originale des estimations mensuelles sur le commerce de détail de même que la série étalonnée des estimations du MV avec les CV correspondants. On y trouve aussi la série des valeurs infra-annuelles ajustées et les CV correspondants (deux dernières colonnes). La figure 1 représente graphiquement la série originale et la série étalonnée. Ce graphique permet de constater que l'étalonnage ne modifie pas la tendance initiale de la série et qu'il s'opère une très forte réduction du CV des estimations infra-annuelles. Le tableau 4 donne la série originale des estimations annuelles sur le commerce de détail de même que la série des valeurs annuelles ajustées avec les CV correspondants. On calcule les variances des valeurs ajustées des tableaux 3 et 4 au moyen des expressions (4.4) et (4.5) respectivement, après avoir remplacé les paramètres par les estimations du MV correspondantes. Les résultats relatifs aux valeurs ajustées témoignent aussi d'une forte réduction du CV des estimations originales. Autrement dit, l'étalonnage a pour effet d'accroître la précision des séries mensuelles et annuelles.

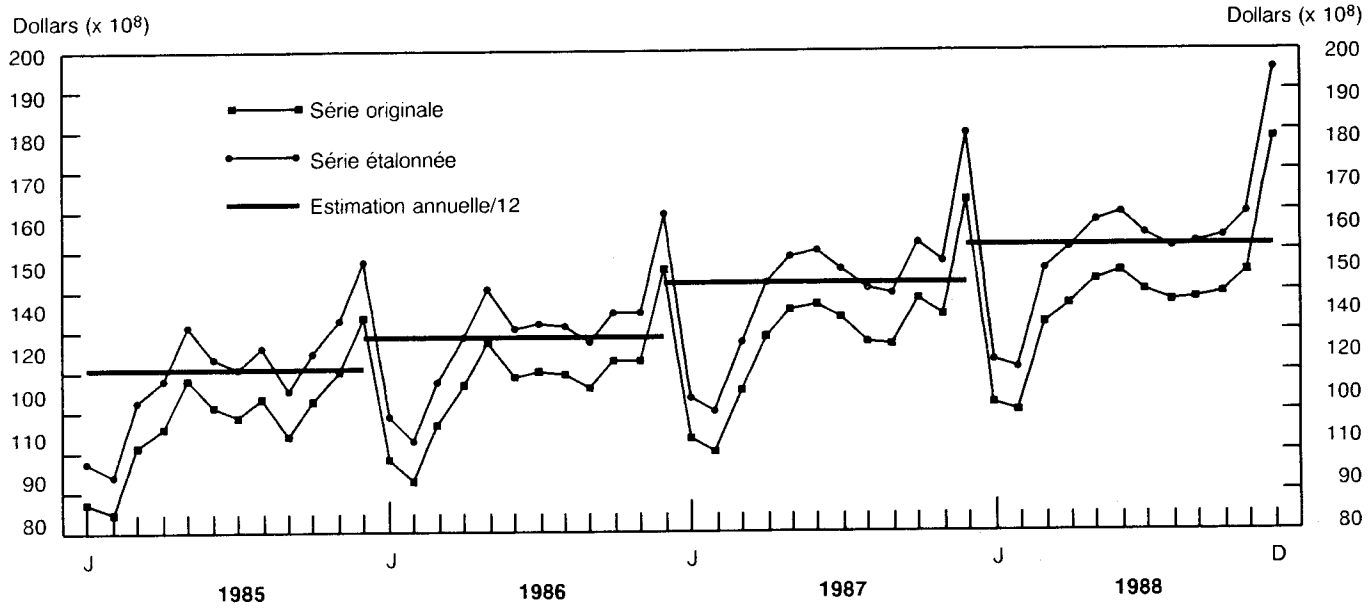

Figure 1. Estimations mensuelles sur le commerce de détail Canada, tous les magasins

Tableau 4

Estimations annuelles sur le commerce de détail et valeurs annuelles ajustées (en millions de dollars) et CV correspondants

Année	z_T^*	$CV(z_T)$	$\hat{z}_T$	$CV(\hat{z}_T)$
1985	143,965.400	0.00033	143,927.507	0.00032
1986	154,377.100	0.00031	154,425.491	0.00030
1987	169,944.600	0.00193	169,101.697	0.00128
1988	181,594.000	0.00137	181,738.512	0.00127

* **Source:** Commerce de détail annuel, publication de Statistique Canada (n° 63-223 au catalogue, annuel).

7. CONCLUSIONS

Le modèle non linéaire traité dans cet article semble se prêter très bien à l'étalonnage de séries économiques construites à partir de grandes enquêtes par sondage. Les méthodes itératives proposées pour calculer les estimations du maximum de vraisemblance des paramètres du modèle sont très faciles à appliquer. Cependant, la convergence se fait beaucoup plus lentement avec la méthode des calculs successifs des équations d'estimation qu'avec la méthode de Fisher-Newton-Raphson. Nous avons présenté dans cet article les expressions en forme analytique fermée pour les covariances des estimateurs du MV. Ces estimateurs et les covariances correspondantes peuvent servir à faire des inférences sur les paramètres du modèle. Nous avons aussi

défini les expressions pour les valeurs infra-annuelles et annuelles ajustées et les covariances asymptotiques correspondantes. La méthode décrite dans cet article semble offrir un ajustement précis pour les données sur le commerce de détail au Canada. Cependant, il est nécessaire d'élaborer des tests de validité de l'ajustement pour ce modèle et pour d'autres modèles d'étalonnage.

REMERCIEMENTS

Nous tenons à remercier le rédacteur en chef, le rédacteur associé et deux arbitres pour leurs précieux conseils qui ont contribué à hausser grandement la qualité de cet article. Nous voulons aussi exprimer notre gratitude à messieurs J. Gambino et M. Kovačević, de Statistique Canada, pour les commentaires qu'ils ont faits sur cette étude.

BIBLIOGRAPHIE

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

CHOLETTE, P.A. (1984). L'ajustement des séries infra-annuelles aux repères annuels. *Techniques d'enquête*, 10, 39-53.

CHOLETTE, P.A. (1987). Benchmarking and interpolation of time series. Direction de la méthodologie, document de travail, Statistique Canada.

CHOLETTE, P.A. (1988). Benchmarking systems of socio-economic time series. Direction de la méthodologie, document de travail, Statistique Canada.

CHOLETTE, P.A. (1992). Users' manual of programmes BENCH and CALEND to benchmark, interpolate and Calendarize time series data on micro computers. Direction de la méthodologie, document de travail, Statistique Canada.

CHOLETTE, P.A., et DAGUM, E.B. (1989). Benchmarking socio-economic time series data: a unified approach. Direction de la méthodologie, document de travail, Statistique Canada.

CHOLETTE, P.A., et DAGUM, E.B. (1991). Benchmarking time series with autocorrelated sampling errors. Direction de la méthodologie, document de travail, 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.

GODAMBE, V.P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 13, 1208-1211.

HIDIROGLOU, M.A., et GIROUX, S. (1986). Composite estimation for the Retail Trade Survey. Direction de la méthodologie, document de travail, Statistique Canada.

HILLMER, S.C., et TRABELSI, A. (1987). Benchmarking of economic time series. *Journal of American Statistical Association*, 82, 1064-1071.

LANIEL, N., et FYFE, K. (1989). Benchmarking of economic time series. Direction de la méthodologie, document de travail, Statistique Canada.

LANIEL, N., et FYFE, K. (1990). Étallonnage des séries économiques. *Techniques d'enquête*, 16, 283-289.

LANIEL, N., et MIAN, I.U.H. (1991). Maximum likelihood estimation for the constant bias model with mixed benchmarks. Direction de la méthodologie, document de travail, Statistique Canada.

McLEOD, I. (1975). Derivation of the theoretical autocovariance function of autoregressive-moving average time series. *Applied Statistics*, 24, 255-256.

MIAN, I.U.H., et LANIEL, N. (1991). Maximum likelihood estimation for the constant bias benchmarking model. Direction de la méthodologie, document de travail, Statistique Canada.

RAO, C.R. (1973). *Linear Statistical Inference and Its Applications*, (2ième Éd.). New York: John Wiley.