Techniques d'enquête, décembre 1998
Vol. 24, n° 2, pp. 151-158
Statistique Canada

Les enquêtes sur les indices de prix en tant qu'études quasi-longitudinales

ALAN H. DORFMAN[1]

RÉSUMÉ

Pour calculer les indices de prix, il faut recueillir des données relatives à un «même article» (en fait, un ensemble d'articles définis avec précision) durant diverses périodes. La question qu'on se pose est celle de savoir si de telles données «quasi-longitudinales» peuvent être modélisées de manière à expliquer ce qu'est un indice des prix. Des chercheurs de pointe spécialisés dans les questions relatives aux indices de prix ont émis des doutes quant à la possibilité d'utiliser la modélisation statistique pour caractériser de tels indices. Dans la présente communication, on propose un simple modèle à espace d'états relatif aux données sur les prix qui donne un indice des prix à la consommation exprimé d'après les paramètres du modèle.

MOTS CLÉS: Modèle à marche aléatoire et bruit; modèle à espace d'états; indice de Laspeyres; indice de Paasche; indice des prix géométrique.

1. INTRODUCTION

L'échantillonnage d'enquête utilisé pour calculer un indice des prix à la consommation comporte la surveillance d'un article donné au fil du temps, afin de déterminer ses prix à diverses époques. Seulement voilà, normalement, on ne suit pas exactement le même article (en effet, ce n'est pas le prix de la boîte de soupe aux tomates de marque Y que l'on trouve au point de vente Z qui est vérifié à plusieurs reprises, car il est probable que la boîte en question aura été vendue et consommée entre deux visites de la personne chargée de l'enquête), mais plutôt d'une succession d'articles qui correspondent tous à la même description («Boîte de 8 oz de soupe aux tomates avec harengs de marque Y, vendue au point de vente Z») et dont le prix est noté à diverses époques. Autrement dit, il s'agit essentiellement d'un groupe d'articles qui correspondent à une description précise suivie au fil du temps. C'est pourquoi les enquêtes qui portent sur les indices de prix peuvent être qualifiées de «quasi-longitudinales», par opposition aux enquêtes longitudinales, dans lesquelles on suit des articles distincts au fil du temps. Il est néanmoins raisonnable d'espérer que le fait d'effectuer des relevés à plusieurs reprises au fil du temps puisse mener à des méthodes d'estimation qui pourraient tirer avantage de l'aspect chronologique des enquêtes de ce type.

Compte tenu de cet espoir, la présente communication se penche sur une question qui a été généralement ignorée par les statisticiens et les économistes, ou qui tout au plus à reçu une réponse négative. Cette question est la suivante: est-ce qu'on peut traiter un indice des prix à la consommation (IPC) d'un point de vue statistique? Autrement dit, est-ce que le paramètre qui caractérise la «variation du coût de la vie» d'une période à une autre, et que les enquêtes sur les indices de prix essaient d'estimer, peut être défini d'après un modèle stochastique?

Aldrich (1992) donne une interprétation historique des premières tentatives, entreprises par Jevons, et surtout par Edgeworth, d'intégrer des hypothèses distributionnelles aux indices des prix à la consommation. Les communications récentes ayant pour objet la modélisation stochastique sont celles de Balk (1980), Clements et Izan (1981, 1987), Bryan et Cechetti (1993), Kott (1984), et Selvanathan et Rao (1994). Diewert (1995) passe en revue et critique ces tentatives en invoquant un argument de Keynes (1930) comme motif déterminant pour rejeter l'approche stochastique.

Dans la présente communication, nous proposons une approche précise en ce qui a trait à la modélisation de l'indice des prix à l'aide de modèles à espace d'états, et nous soumettons provisoirement un tel modèle. Celui-ci est appliqué à des données de balayage afin de montrer la faisabilité d'un indice fondé sur ces mêmes données. La méthode que nous considérons contourne la critique de Keynes d'une façon fondamentale et offre la perspective des nombreux avantages que peut apporter une solide modélisation statistique, y compris, peut-être, une simplification du processus d'échantillonnage d'enquête.

Ci-après, nous allons d'abord passer brièvement en revue la définition d'un indice des prix ainsi que les deux méthodes (non stochastiques) qui sous-tendent de manière prédominante le choix des indices (section 2). Nous examinons notamment l'exemple de modèle statistique pour indices des prix qui est présenté par Bryan et Cecchetti (1993), ainsi que la formulation de l'objection de Keynes par Diewert (section 3). Nous présentons par la suite une méthode de modélisation d'un indice des prix à la consommation qui contourne les difficultés mentionnées

[1] Alan H. Dorfman, U.S. Bureau of Labor Statistics, Room 4915, 2 Massachusetts Ave. N.E., Washington, D.C., 20212-0001, U.S.A.; courrier électronique: dorfman_a@bls.gov.

par Keynes et Diewert, ce qui mène naturellement à l'utilisation de modèles à espace d'états (section 4). Nous présentons notamment les résultats de l'application d'un modèle relativement simple à marche aléatoire et bruit à des données de balayage tirées de l'Academic Data Base d'A.C. Nielsen (section 5). Nous évaluons le nouvel indice dans la section 6, en mentionnant les travaux de recherche ultérieurs qui pourraient être utiles.

2. HISTORIQUE DE LA QUESTION

Par indice des prix à la consommation (IPC) on entend un nombre unique qui indique la variation du pouvoir d'achat des consommateurs entre une période t' et une période t. Les éléments de base bruts de cet indice sont les prix des divers articles que l'on peut se procurer (au moins) aux deux époques

$$p_\tau = (p_{\tau 1}, \ldots, p_{\tau N}), \tau = t', t$$

ainsi que les quantités des articles vendus

$$q_\tau = \left(q_{\tau 1}, \ldots, q_{\tau N}\right), \tau = t', t.$$

(Cependant, dans la pratique, les données quantitatives relatives aux deux époques en question ne sont pas disponibles, et il faut alors recourir à des données de remplacement d'un type ou d'un autre). L'IPC est obtenu à l'aide d'une «formule» dans laquelle on utilise les éléments bruts suivants:

$$I_{t't} = f\left(p_{t'}, p_t, q_{t'}, q_t\right),$$

où $f(\cdot)$ est une fonction de l'une des nombreuses expressions possibles. La plupart de ces expressions existent depuis longtemps et ont été décrites de façon approfondie dans les textes spécialisés qui portent sur les indices de prix.

À titre d'exemple, nous mentionnons ici l'indice de Laspeyres

$$L_{t't} = \frac{\sum_{i=1}^{N} q_{t'i} p_{ti}}{\sum_{i=1}^{N} q_{t'i} p_{t'i}} = \sum_{i=1}^{N} f_{t'i} r_{t't},$$

où $f_{t'i} = q_{t'i} p_{t'i} / \sum_{i=1}^{N} q_{t'i} p_{t'i}$ représente les «dépenses relatives» et $r_{t'i} = p_{ti}/p_{t'i}$ les «rapports de prix». L'indice de Laspeyres fait appel aux quantités relatives à la période qui vient en premier comme base fixe pour la comparaison des prix antérieurs et des prix qui suivent. L'indice Laspeyres (ou une variante proche) a tendance à être l'indice le plus ciblé par les gouvernements, en raison de sa simplicité et de sa transparence aux yeux du profane.

Le pendant naturel de l'indice de Laspeyres est l'indice de Paasche

$$P_{t't} = \frac{\sum_{i=1}^{N} q_{ti} p_{ti}}{\sum_{i=1}^{N} q_{ti} p_{t'i}}$$

qui étalonne les prix d'après les quantités de la période subséquente. La plupart des autres indices qui sont fondés sur d'autres formules se situent généralement entre les indices de Paasche et de Laspeyres.

À titre de référence ultérieure dans la présente communication, nous mentionnons un indice fondé sur la moyenne géométrique, où des poids non négatifs fixes f_i, donnent au total 1:

$$G_{t't} = \prod_{i=1}^{N} \left(\frac{p_{ti}}{p_{t'i}} \right)^{f_i}.$$

On utilise le terme «moyenne géométrique» pour référer à cet indice.

Fisher (1922) analyse ces formules, ainsi que de nombreuses autres formules d'indices. Il introduit notamment l'approche dite «des tests» pour effectuer un choix parmi les diverses possibilités de formule $f(\cdot)$; dans le cas de cette méthode, on décrit les propriétés («tests») qu'un indice raisonnable devrait nécessiter, puis on détermine dans quelle mesure chaque formule d'indice satisfait à ces critères.

Un des tests est la condition de réversibilité dans le temps: $I_{t't} I_{tt'} = 1$. Deux indices qui sont toujours utilisés dans le monde mais qui ne passent pas ce test sont l'indice Carli-Sauerbach $C_{t't} = \sum_{i=1}^{N} f_i p_{ti}/p_{t'i}$ et la moyenne géométrique $\tilde{G}_{t't} = \prod_{i=1}^{N} (p_{ti}/p_{t'i})^{f_{t'i}}$ qui fait appel aux dépenses de la période antérieure plutôt qu'à des poids fixes. On peut montrer facilement que $C_{t't} C_{tt'} \geq 1$, en utilisant l'inégalité de Cauchy-Schwartz, ce qui laisse penser que cet indice sera trop élevé.

Si une augmentation des prix de l'article i a tendance à produire une augmentation des dépenses, alors $\tilde{G}_{t't} \tilde{G}_{tt'} \leq 1$, de sorte que dans de telles conditions, la moyenne géométrique fondée sur la période antérieure tend à être trop basse. D'autre part, si une augmentation des prix de l'article i tend à produire une diminution des dépenses, alors $\tilde{G}_{t't}$ donne des indications trop élevées. En général on peut donc s'attendre à ce que cet indice soit plutôt irrégulier.

Cette constatation suggère l'axiome suivant: les indices de prix qui sont exprimés sous la forme d'une moyenne géométrique ne devraient pas lier des poids aux prix observés dans une des périodes comparées; les indices qui prennent la forme d'une moyenne arithmétique ne devraient pas comporter des poids indépendants des prix en question.

De façon contrastante par rapport à $\tilde{G}_{t't}$, la moyenne géométrique $G_{t't} = \prod_{i=1}^{N} (p_{ti}/p_{t'i})^{f_i}$ qui comporte des poids fixes, est le seul indice qui satisfait aux cinq axiomes relatifs aux indices de prix qui sont énoncés dans Balk (1995) ainsi qu'au «test de circularité», selon lequel $I_{t't} = I_{t't*} I_{t*t}$ lorsque $t' < t^* < t$. L'inversion du temps est une conséquence immédiate.

Les indices qui passent la plupart des tests sont généralement ceux qui intègrent des données quantitatives tirées des deux périodes; c'est le cas, par exemple, de l'indice de Fisher

$$F_{t't} = \left(L_{t't} P_{t't}\right)^{1/2}$$

et de l'indice de Törnqvist

$$T_{t't} = \prod_{i=1}^{N} \left(\frac{p_{ti}}{p_{t'i}}\right)^{f_{t'ti}},$$

où $f_{t'ti} = (f_{t'i} + f_{ti})/2$. Souvent, on ne peut pratiquement pas distinguer l'un de l'autre les indices de Fisher et de Törnqvist. On peut trouver d'autres analyses de la méthode par tests dans Balk (1995), Diewert (1987), ainsi que dans Eichhorn et Voeller (1976).

La deuxième façon d'évaluer les formules relatives aux indices est la méthode «économique». Celle-ci définit un indice générique qui prend la forme suivante:

$$I_{t't} = \frac{C(p_t, U)}{C(p_{t'}, U)},$$

où $U = U(q_1, ..., q_N)$ est une «fonction utilitaire» bien définie et $C(p_t, U)$ est le coût minimal aux prix p_t, de l'atteinte du niveau de vie ou «utilité» U. Dans le cas d'une fonction utilitaire en particulier, on cherche à savoir si une formule donnée peut être considérée comme une bonne approximation de l'indice du coût de la vie correspondant. Comme la méthode des tests, cette deuxième façon de procéder a tendance à donner des indices qui comprennent des données quantitatives relatives aux deux périodes. Voir Diewert (1987) pour plus détails.

3. LA MÉTHODE STOCHASTIQUE

Aldrich (1992) présente l'historique des premiers essais de modélisation des rapports de prix, ou des logarithmes de rapports de prix, à l'aide d'un paramètre courant qui représente le taux global de croissance des prix. Une idée fondamentale contenue dans l'étude de cet auteur est que l'application de la méthode stochastique aux indices de prix, bien qu'elle soit un exemple d'une première application de la statistique à des questions économiques, est morte de sa belle mort. Diewert (1995) analyse lui aussi ces exemples, ainsi que des exemples plus récents de la modélisation statistique de rapports de prix. La difficulté que Diewert, à l'instar de Keynes (1930), voit dans ce genre de modélisation est illustrée par une modèle de Clements et Izan (1987).

La période comprise entre t' et t est divisée en segments de même durée, ce qui donne des intervalles relativement courts représentés de manière générique comme étant compris entre $t-1$ et t. Le logarithme des rapports de prix pour une telle «micro-période» est donné par l'expression suivante:

$$\log\left(\frac{p_{ti}}{p_{t-1,i}}\right) = \pi_t + \beta_i + \varepsilon_{ti}, \tag{1}$$

où $\varepsilon_{ti} \sim (0, \sigma_t^2/f_i)$. Dans le modèle de ces auteurs, les f_i représentent la part de dépenses relative à l'article i, au cours de la période comprise entre t' et t. À des fins d'identification, on présume que $\sum_{i=1}^{N} f_i \beta_i = 0$. Ces présuppositions mènent à un estimateur du maximum de vraisemblance

$$\hat{\pi}_t = \sum_{i=1}^{N} f_i \log\left(\frac{p_{ti}}{p_{t'i}}\right),$$

ce qui donne un estimateur du maximum de vraisemblance (EMV) de la tendance des prix durant la période courte

$$\exp(\hat{\pi}_t) = \prod_{i=1}^{N} \left(\frac{p_{ti}}{p_{t'i}}\right)^{f_i};$$

cela veut donc dire que, d'après le modèle stochastique de ces auteurs, on obtient un indice géométrique, avec des poids f_i, semblable à l'indice de Törnqvist.

On peut également obtenir des estimations de β_i et de σ^2, ainsi que des estimations de la précision, par exemple, de la variance de $\hat{\pi}$. Ainsi, une nouvelle fondation statistique semble être placée sous un vieil estimateur.

Diewert (1995) formule plusieurs objections, dont aucune ne peut être prise à la légère. L'objection principale est la suivante:

> «...l'objection fondamentale de Keynes (Keynes 1930, p. 78): 'La variation du niveau des prix $[\exp(\pi_t)]$ qui devrait avoir eu lieu s'il n'y avait pas eu de variation dans les prix relatifs, n'est plus pertinente si les prix relatifs ont effectivement subi une variation, car cette variation a eu elle-même une incidence sur le niveau des prix'.» (Traduction).

Si, par exemple, le prix du pain varie par rapport au prix des voitures, cette même variation entraîne une variation du niveau global des prix.

L'objection de Keynes n'est pas tout à fait claire. Pourquoi ne peut-il y avoir deux aspects de la variation des prix, un aspect global et un aspect particulier? Cependant, il n'est pas difficile de reconnaître que les tendances des prix considérés individuellement sont primaires, une tendance des prix considérés globalement ne pouvant être qu'une somme pondérée des tendances individuelles. Diewert propose la traduction suivante, sous forme d'un modèle, de l'objection de Keynes: étant donné que nous devons déterminer la tendance globale des prix par l'expression

$$\pi_t^* = \sum_{i=1}^{N} f_i \beta_{ti}^*,$$

le modèle (1) doit être remplacé par

$$\log\left(\frac{p_{ti}}{p_{t-1,i}}\right) = \pi_t + \beta_{ti} + \varepsilon_{ti}, \qquad (2)$$

où $\beta_{ti} = \pi_t - \beta_{ti}^*$ et $\sum_{i=1}^{N} f_i \beta_{ti} = 0$. La différence déterminante entre cette équation et l'équation (1) est que, maintenant, les paramètres d'article β_{ti} sont indexés en fonction du temps. Mais «alors, le modèle qui est obtenu comporte trop de paramètres à identifier». Ce fait devrait suffire à invalider la méthode.

Diewert (1995) n'aborde pas le modèle chronologique, beaucoup plus compliqué, de Bryan et Cechetti (1993). Parmi les communications précédentes, celle-ci est probablement celle qui se rapproche le plus de la nôtre. Elle porte sur l'utilisation d'un modèle complexe à espace d'états et du filtre de Kalman. À l'instar des autres communications examinées par Diewert, elle fait l'objet de l'objection de Keynes.

4. RÉEXAMEN DES INDICES DE PRIX

4.1 Présuppositions courantes

La modélisation stochastique du comportement des prix décrite dans la section précédente, qu'elle soit effectuée à l'aide des équations (1) ou (2), ou d'une équation similaire, présente trois caractéristiques dignes de mention; la modélisation est

1. complète: elle vise directement un «taux d'inflation» global qui comprend tous les articles;
2. atomistique: chaque article est modélisé individuellement et comporte son «propre» paramètre, son propre taux d'inflation [$\exp(\pi_t + \beta_i)$], de manière distincte de tous les autres articles;
3. à périodes multiples: la modélisation pour la période comprise entre $t-1$ et t est distincte de celle s'appliquant à la période qui va de $t-2$ à $t-1$ *etc.*

C'est la combinaison de ces présuppositions qui donne lieu à l'argument de Diewert concernant le trop grand nombre de paramètres. La critique de Keynes vise avant tout la première caractéristique : un taux d'inflation global, pour lequel une augmentation ou une baisse du coût de la vie doit être un mélange pondéré de plusieurs tendances des prix. On peut concéder cela sans aller jusqu'à inclure le point 2. Ce point est accepté tacitement dans le cas de presque toutes les constructions (non stochastiques) d'indices de prix. Cependant, il n'est pas évident du tout que chaque article a sa propre tendance en ce qui a trait au prix. En effet, il est probable que le prix d'articles différents (par exemple, la crème glacée de la marque X dans plusieurs supermarchés) a tendance à augmenter et à diminuer simultanément (du moins à long terme). Il existe des degrés d'homogénéité entre les articles. En tout cas, aucune de ces présuppositions n'est une composante nécessaire d'une approche stochastique à l'égard des indices de prix.

4.2 Un modèle élémentaire à espace d'états

Nous divisons la période comprise entre t' et t en sous-périodes $t', t'+1, ..., t-1, t$, et l'ensemble d'articles hétérogènes en sous-groupes homogènes, où la caractéristique qui définit l'homogénéité est la tendance à une similitude de comportement en ce qui a trait aux variations de prix. Nous présupposons deux choses:

1. $I_{t't}$ est un mélange d'indices $I_{gt't}$ «homogènes»;
2. $I_{gt't}$ peut être obtenu par enchaînement: $I_{gt't} = \prod_\tau I_{g\tau-1,\tau}$, où $\tau = t'+1, ..., t$.

Nous nous concentrons sur un seul indice de groupe $I_{gt't}$, et laissons tomber l'indice inférieur g, afin de simplifier la notation. Par conséquent, pour le reste de la présente communication, nous nous concentrons sur le «sous-indice» $I_{t't} \equiv I_{gt't}$.

Nous allons maintenant créer un modèle élémentaire à espace d'états (Harvey 1990, chapitre 3) pour les logarithmes des rapports de prix à l'intérieur du groupe. Supposons que le groupe comprend un nombre n d'articles. Dans le cas de $i = 1, ..., n$, soient $r_{ti} \equiv p_{ti}/p_{t-1,i}$ les rapports de prix relatifs à la micro-période, et $y_{ti} \equiv \log(p_{ti}/p_{t-1,i}) = \log(p_{ti}) - \log(p_{t-1,i})$, leurs logarithmes. La raison pour laquelle nous utilisons des logarithmes est que d'importants travaux empiriques, en premier lieu ceux d'Edgeworth (voir Diewert (1995)) laissent penser que les logarithmes des rapports de prix seront beaucoup plus susceptibles d'avoir une distribution normale que les rapports de prix eux-mêmes, qui peuvent être considérablement dissymétriques. La distribution normale des erreurs est une présupposition standard dans le cas des modèles à espace d'états. Soit $\boldsymbol{y}_t \equiv (y_{t1}, ..., y_{tn})$ et $\mathbf{1}$ un vecteur de un de longueur n.

Considérons le modèle multidimensionnel à marche aléatoire et bruit (MMAB):

$$\boldsymbol{y}_t = \mathbf{1}\mu_t + \varepsilon_t,\ \varepsilon_t \sim \mathrm{MVN}\left(0, \textstyle\sum_{\varepsilon\varepsilon}\right)$$

$$\mu_t = \mu_{t-1} + \eta_t,\ \eta_t \sim N\left(0, \sigma_{\eta\eta}\right) \qquad (3)$$

où $\varepsilon_\tau, \eta_\tau, \tau \in (t', t'+1, ..., t-1, t)$ sont mutuellement indépendants. Le modèle laisse entendre que le montant de l'augmentation (ou de la baisse) de l'ensemble des prix du groupe dans une micro-période tend à osciller autour du montant de l'augmentation (ou de la baisse) observé généralement au cours de la micro-période précédente. C'est une question d'observation courante: si la hausse des prix au cours d'un mois donné tend à être forte (ou faible), au cours du mois suivant, elle sera forte (ou faible) d'une manière correspondante. Étant donné que nous considérons un ensemble d'articles homogène, il est normal que leurs logarithmes de rapport de prix aient une moyenne commune. Nous renvoyons à des travaux ultérieurs la question de savoir comment intégrer des sous-indices à un indice global.

Le modèle (3) comporte le modèle (plus simple) unidimensionnel à marche aléatoire et bruit:

$$\bar{y}_t = \mu_t + \bar{\varepsilon}_t,\ \bar{\varepsilon}_t \sim N\left(0, \sigma_{\bar{\varepsilon}\bar{\varepsilon}}\right)$$

$$\mu_t = \mu_{t-1} + \eta_t,\ \eta_t \sim N\left(0, \sigma_{\eta\eta}\right) \qquad (4)$$

où $\bar{y}_t = n^{-1}1'y_t, \bar{\varepsilon}_t = n^{-1}1'\varepsilon_t$, et $\bar{\sigma}_{\varepsilon\varepsilon} = n^{-1}\mathbf{1}'\sum_{\varepsilon\varepsilon}\mathbf{1}$. Certains renseignements sont perdus lorsqu'on utilise l'équation (4); en revanche, la présupposition concernant la normalité est encore plus susceptible d'être juste. Pour plus de commodité, les calculs de l'étude décrite dans la section 5 étaient fondés sur le modèle unidimensionnel.

Le filtre de Kalman (Harvey 1990, section 3.2) peut être utilisé pour obtenir les estimations $\hat{\mu}_\tau$, et $\hat{\sigma}_{\bar{\varepsilon}\bar{\varepsilon}}, \hat{\sigma}_{\eta\eta}$ des paramètres d'état μ_τ et des variances $\sigma_{\bar{\varepsilon}\bar{\varepsilon}}, \sigma_{\eta\eta}$ respectivement.

Nous définissons ensuite $I_{t't} \equiv E(G_{t't} \mid S_t)$, où $G_{t't} = \prod_i (p_{ti}/p_{t'i})^{f_i}$ est une moyenne géométrique qui dépend des parts fixes f_i, et où S_t représente la totalité des paramètres d'état μ_τ au cours de la période t, et également les «hyperparamètres» $\sigma_{\bar{\varepsilon}\bar{\varepsilon}}, \sigma_{\eta\eta}$. Autrement dit, nous déterminons ce que nous considérons être le processus sous-jacent au cours de la période t. Il s'ensuit que

$$I_{t't} = \exp\left(\mu_t + \mu_{t-1} + \cdots \mu_{t'+1} + \frac{1}{2}\nu\right), \qquad (5)$$

où $\nu = (t-t')\sum_i\sum_{i'}\sigma_{ii'}f_if_{i'}$, avec $\sigma_{ii'}$ étant la covariance de ε_{ti} et $\varepsilon_{ti'}$, sera habituellement d'un ordre inférieur à celui des paramètres d'état $\mu_{\tau i}$. L'estimateur naturel de $I_{t't}$ est $\hat{I}_{t't} \equiv \exp(\hat{\mu}_t + \hat{\mu}_{t-1} + \ldots \hat{\mu}_{t'+1})$; alors

$$E\left(\hat{I}_{t't} \mid S_t\right) = \exp\left(\mu_t + \mu_{t-1} + \cdots \mu_{t'+1} + \frac{1}{2}\tilde{\nu}\right), \qquad (6)$$

où $\tilde{\nu}$, donné dans l'annexe, n'est généralement pas égal à ν, mais est souvent proche de cette valeur et, en tout cas, présente le même ordre de grandeur. La différence $\Delta(\nu) = \tilde{\nu} - \nu$ peut être estimée par, disons, $\hat{\Delta}(\nu)$, ce qui donne un estimateur corrigé pour le biais $\check{I}_{t't} \equiv \hat{I}_{t't}$ $\exp(-1/2\hat{\Delta}(\nu))$. Les expressions pour ν et $\tilde{\nu}$, ainsi qu'une proposition pour un $\hat{\Delta}(\nu)$ sont donnés dans l'annexe. On peut signaler que $\hat{\Delta}(\nu)$, et donc $\check{I}_{t't}$, dépend des poids f_i, mais pas $\hat{I}_{t't}$.

5. ÉTUDE EMPIRIQUE

Afin de déterminer la faisabilité du calcul d'indices de prix à l'aide du modèle MMAB, et pour avoir une certaine idée du comportement de l'indice fondé sur ce modèle, on a effectué une petite étude empirique en utilisant des données sur les prix et les quantités du thon en boîte tirées de l'Academic Database d'A.C. Nielsen. Le thon en boîte présente un comportement quelque peu instable en ce qui a trait aux prix et aux quantités, en raison de ventes fréquentes, qui s'effectuent parfois à des prix réduits de façon marquée.

Cette étude couvrait le nord-est des États-Unis et les 104 semaines des années 1992 et 1993. L'ensemble de données initial était plutôt volumineux. Afin de faciliter l'enquête, les données hebdomadaires ont été regroupées par périodes de quatre semaines, ce qui a donné 26 périodes réparties sur deux ans. Ainsi, aux fins de l'étude en question, les données qui ont été utilisées étaient des quantités cumulatives et des prix moyens pondérés d'après la quantité, pour des périodes de quatre semaines.

Les groupes homogènes ont été définis d'après la marque et le type de produit, de la façon suivante: 3 marques, désignées ici par les lettres A, B et C, de thon de «première qualité» dans l'eau, les trois mêmes marques de thon «léger» dans l'huile, et de nouveau les trois mêmes marques de thon «léger» dans l'eau, pour un total de 9 groupes.

L'étude était centrée sur 83 points de vente qui présentaient des quantités positives durant la plupart des périodes de 4 semaines, dans le cas de chacun des 9 groupes.

L'indice $\hat{I}_{t't}$ fondé sur le modèle MMAB et l'indice corrigé $\check{I}_{t't}$ fondé sur le même modèle ont été calculés pour quatre intervalles. Dans chaque cas, la période finale $t = 26$, et la période initiale a été prise successivement comme étant $t' = 3, 6, 10, 14$. À des fins de comparaison, nous avons également calculé les indices correspondants de Laspeyres et de Paasche. Ces deux indices standard servent également de base pour une comparaison indirecte avec les indices de Fisher et de Törnqvist, qui se situent environ à moitié chemin entre les deux indices précédents.

Les figures 1 et 2 (thon de première qualité et thon léger respectivement) donnent les valeurs des quatre indices pour les quatre intervalles, les points indiquant les indices d'espace d'états et les lignes les indices de Laspeyres et de Paasche. L'indice corrigé $\check{I}_{t't}$ est invariablement plus élevé que l'indice non corrigé $\hat{I}_{t't}$. Il est à noter que puisque c'est la première période que nous faisons varier lorsque le tracé des indices est monotone vers le haut, cela semble indiquer une tendance à la baisse du prix du thon dans le groupe visé (et vice versa).

Nous constatons que les nouveaux indices ne sont pas aberrants par rapport aux indices classiques, se situant souvent entre les indices de Laspeyres et de Paasche, mais ils ont tendance à être nettement plus stables à mesure que

t' varie, ce qui laisse penser que les indices classiques réagissent à un «bruit» présent dans les données et qu'en fait, très peu de variation a lieu au fond durant cette période de deux ans. On remarquera également, dans la figure 2, que le thon léger dans l'huile et le thon léger dans l'eau d'une même marque ont des comportements semblables, ce qui laisse penser que nous aurions peut-être dû prendre un groupe «homogène» plus large.

6. TRAVAUX ULTÉRIEURS

La recherche décrite dans la présente communication suggère plusieurs sujets pour des travaux de recherche ultérieurs.

Il faut notamment mettre au point des mesures de la précision et des estimations des indices fondés sur le modèle MMAB, sous forme de variances ou d'intervalles de confiance fondés sur le modèle à espace d'états. Même ceux qui doutent de la viabilité de l'application d'une méthode stochastique aux indices de prix trouvent la possibilité d'avoir une mesure de la précision attrayante (Diewert 1995). Il serait également utile d'obtenir des mesures de la précision d'autres indices standard, d'après le modèle à espace d'états.

Il est également souhaitable que du travail empirique soit effectué afin de déterminer avec plus de précision quels groupes d'articles pourraient répondre le mieux au critère d'«homogénéité». Il faudrait également étudier des modèles qui seraient plus sophistiqués que le simple modèle à marche aléatoire et bruit. À cet égard, l'utilisation de données de balayage sera très utile car ces données renseignent de manière très détaillée tant sur les quantités que sur les prix.

La méthode fondée sur des modèles à espace d'états permet de composer avec l'absence de données (Harvey 1990, section 3.4.7). Une question importante que l'on se pose est celle de savoir dans quelle mesure ces modèles permettront de tenir compte de données manquantes dans l'estimation d'indices de prix. Étant donné que, dans la pratique, la plupart des données qui sont utilisée pour calculer des indices de prix portent sur un petit échantillon d'articles offerts, il faut connaître la robustesse des indices fondés sur des modèles à espace d'états à l'égard de l'absence de données.

Les algorithmes de lissage et de prévision des modèles à espace d'états sont bien connus. Leur utilisation dans la révision et la prévision d'indices pourrait être d'un grand intérêt. Enfin, dans la présente communication, nous avons abordé uniquement l'obtention d'un indice pour un seul groupe homogène. Il serait utile de créer un modèle à espace d'états qui réunit plusieurs groupes et qui permet d'obtenir une mesure globale du pouvoir d'achat.

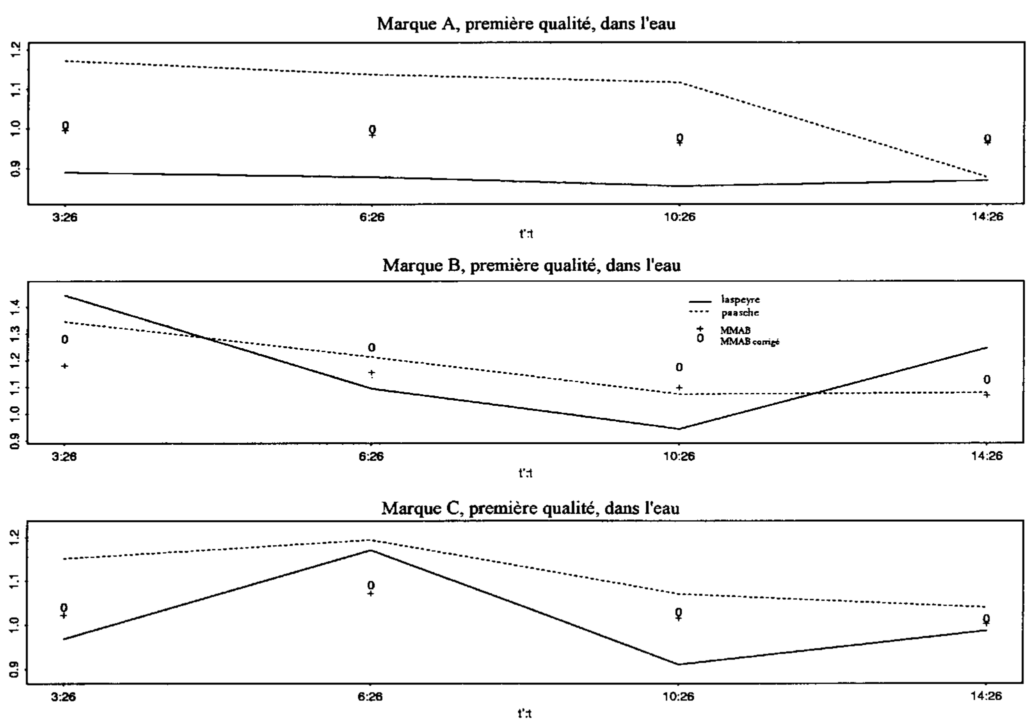

Figure 1. Quatre indices de prix pour quatre périodes de temps, thon de première qualité

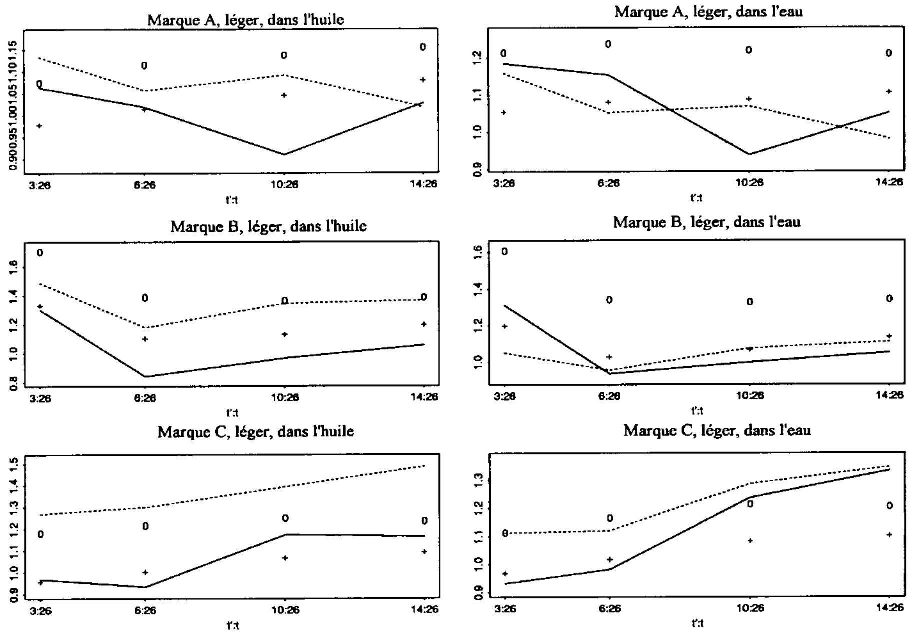

Figure 2. Quatre indices de prix pour quatre périodes de temps, thon léger

REMERCIEMENTS

L'auteur remercie B. Moulton, S. Scott, M. Reinsdorf, R. Tiller, B. Balk, et J. Aldrich pour des entretiens discussions portant sur les idées contenues dans la présente communication.

ANNEXE

Détails des équations (5) et (6).

Nous avons

$$G_{t't} = \prod_i \left(\frac{p_{ti}}{p_{t-1,i}} \frac{p_{t-1,i}}{p_{t-2,i}} \cdots \frac{p_{t'+1,i}}{p_{t'i}} \right)^{f_i}$$

$$= \prod_i \left(r_{ti}\, r_{t-1,i} \cdots r_{t'+1,i}\right)^{f_i},$$

et lorsque

$$H_{t't} = \log(G_{t't}) = \sum_i f_i \log(p_{ti}/p_{t'i}),$$

nous obtenons

$$H_{t't} = \sum_i f_i \log\left(r_{ti}\, r_{t-1,i} \cdots r_{t'+1,i}\right)$$

$$= \sum_i f_i\left(y_{ti} + y_{t-1,i} + \cdots y_{t'+1,i}\right)$$

et également

$$I_{t't} = E(G_{t't}) = \exp(E(H_{t't}) + 1/2\, \mathrm{var}(H_{t't})),$$

où les moments sont calculés en tenant compte de l'état S_t, comme dans la section 4.3. Soit $\nu = \mathrm{var}(H_{t't})$. Alors

$$E(H_{t't}) \equiv E(H_{t't} \mid S_t) =$$

$$\sum_i f_i(\mu_t + \mu_{t-1} + \cdots \mu_{t'+1}) = \mu_t + \mu_{t-1} + \cdots \mu_{t'+1}$$

et

$$\nu = \mathrm{var}(H_{t't}) \equiv \mathrm{var}(H_{t't} \mid S_t) =$$

$$\mathrm{var}\left(\sum_{\tau=t'+1}^{t} \sum_i f_i \varepsilon_{\tau i} \mid S_t \right) = (t - t') \sum_i \sum_{i'} \sigma_{ii'} f_i f_{i'},$$

où $\sigma_{ii'}$ représente la covariance de ε_{ti} et $\varepsilon_{ti'}$. Nous constatons que $\nu = (t - t')\sum_i f_i^2 \sigma_{\varepsilon\varepsilon}$, dans le cas particulier où les erreurs ε_{ti} sont indépendantes et réparties de manière identique dans chaque période.

Nous considérons maintenant l'estimateur $\hat{I}_{t't} \equiv \exp(\hat\mu_t + \hat\mu_{t-1} + \cdots \hat\mu_{t'+1})$. Nous constatons que $E(\hat{I}_{t't}) = \exp(\mu_t + \mu_{t-1} + \cdots \mu_{t'+1} + 1/2\, \tilde{\nu})$, où

$$\tilde{\nu} \equiv \mathrm{var}\left(\sum_{t'+1}^{t} \hat\mu_\tau \mid S_t \right) = \left\{ \sum_{t'+1}^{t} \gamma_\tau^2 \right\} \mathrm{var}(\bar{y}_t \mid S_t) +$$

$$\gamma_{t'}^{*2}\, \mathrm{var}(\hat\mu_{t'} \mid S_t) = \left\{ \sum_{t'+1}^{t} \gamma_\tau^2 + \gamma_{t'}^{*2} p_{t'-1} \right\} \sigma_{\bar\varepsilon\bar\varepsilon}$$

avec

$$\gamma_\tau = k_\tau (1 + \sum_{v=\tau+1}^{t} \prod_{u=\tau+1}^{v} (1 - k_u))$$

et

$$\gamma_{t'}^* = \sum_{v=t'+1}^{t} \prod_{u=t'+1}^{v} (1 - k_u),$$

où

$$k_\tau = p_{\tau|\tau-1} \Big/ (p_{\tau|\tau-1} + 1),$$

et $p_{\tau|\tau-1}$, p_τ représentent les erreurs quadratiques moyennes de $\hat{\mu}_\tau$, dans le cas de données jusqu'à $\tau - 1$, τ respectivement, et ceux-ci sont estimés à l'aide du filtre de Kalman.

Ce résultat découle des équations qui sont utilisées pour estimer μ_τ:

$$\hat{\mu}_t = k_t \bar{y}_t + (1 - k_t) \hat{\mu}_{t-1}$$

$$\hat{\mu}_{t-1} = k_{t-1} \bar{y}_{t-1} + (1 - k_{t-1}) \hat{\mu}_{t-2}$$

$$\vdots$$

$$\hat{\mu}_{t'+1} = k_{t'+1} \bar{y}_{t'+1} + (1 - k_{t'+1}) \hat{\mu}_{t'}$$

(*cf.* Harvey 1990, équation 3.2.8), en exprimant chaque $\hat{\mu}_\tau$ par $\bar{y}_\tau, \bar{y}_{\tau-1}, \ldots, \bar{y}_{t'+1}, \hat{\mu}_{t'}$.

En comparant v et $\tilde{v}$, nous constatons de façon empirique que

$$\sum_{t'+1}^{t} \gamma_\tau^2 + \gamma_{t'}^{*2} p_{t'-1} \approx t - t'.$$

Nous considérons ici le cas simple où $\mathrm{var}(\varepsilon_{ti}) = \sigma_{\varepsilon\varepsilon}$ et $\mathrm{cov}(\varepsilon_{ti}, \varepsilon_{ti'}) = \rho\sigma_{\varepsilon\varepsilon}$, avec $\rho \geq 0$, dans le cas de $i' \neq i$, c'est-à-dire où non seulement les variances, mais toutes les covariances aussi sont égales et non négatives. On peut alors montrer que

$$\sigma_{\bar{\varepsilon}\bar{\varepsilon}} = n^{-2} \sum_i \sum_{i'} \sigma_{ii'} \leq \sum_i \sum_{i'} \sigma_{ii'} f_i f_{i'} \leq n \sum_i f_i^2 \sigma_{\bar{\varepsilon}\bar{\varepsilon}},$$

où n représente le nombre d'articles présents du groupe. La borne inférieure est obtenue dans le cas $f_i = 1/n$, et la borne supérieure dans le cas $\rho = 0$. Dans le premier cas, aucune correction de biais n'est nécessaire; dans le deuxième cas, nous prendrions $\hat{\Delta}(v) = \hat{\tilde{v}} - \hat{v}$, où $\hat{v} = (t - t') n \sum_i f_i^2 \hat{\sigma}_{\bar{\varepsilon}\bar{\varepsilon}}$ et $\hat{\tilde{v}} = \{ \sum_{t'+1}^{t} \gamma_\tau^2 + \gamma_{t'}^{*2} p_{t'-1} \} \hat{\sigma}_{\bar{\varepsilon}\bar{\varepsilon}}$. Ces valeurs correspondraient respectivement à $\hat{I}_{t't}$ et à $\tilde{I}_{t't}$.

BIBLIOGRAPHIE

ALDRICH, J. (1992). Probability and depreciation: a history of the stochastic approach to index numbers. *History of Political Economy*, 24, 657-87.

BALK, B.M. (1980). A method for constructing price indexes for seasonal commodities. *Journal of the Royal Statistical Society, A*, 143, 68-75.

BALK, B.M. (1995). Axiomatic price theory: a survey. *Revue Internationale de Statistique*, 63, 69-93.

BRYAN, M.F., et CECCHETTI, S.G. (1993). The consumer price index as a measure of inflation. *Economic Review, Federal Reserve Bank of Cleveland*, 29, 15-24.

CLEMENTS, K.W., et IZAN, H.Y. (1981). A note on estimating Divisia index numbers. *International Economical Review*, 22, 745-747.

CLEMENTS, K.W., et IZAN, H.Y. (1987). The measurement of inflation: a stochastic approach. *Journal of Business and Economic Statistics*, 5, 339-350.

DIEWERT, W.E. (1987). Index numbers. Dans *The New Palgrave: A Dictionary of Economics*, (Éds. J. Eatwell, M. Milgate, et P. Newman). London: MacMillan.

DIEWERT, W.E. (1995). On the Stochastic Approach to Index Numbers. Document de discussion no. DP 95-31, Department of Economics, University of British Columbia.

EICHHORN, W., et VOELLER, J. (1976). *Theory of the Price Index*. Berlin: Springer-Verlag.

FISHER, I. (1922). *The Making of Index Numbers*. Boston: Houghton Mifflin.

HARVEY, A.C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

KEYNES, J.M. (1930). *A Treatise on Money*. New York: Harcourt, Brace and Company.

KOTT, P.S. (1984). A superpopulation approach to the design of price index estimators with small sampling biases. *Journal of Business and Economic Statistics*, 2, 83-90.

SELVANATHAN, E.A., et RAO, D.S.P. (1994). *Index Numbers: A Stochastic Approach*. Ann Arbor: The University of Michigan Press.