

# ÉTUDE DU CHAÎNAGE D'INDICES DE PRIX À L'AIDE DE MICRO-DONNÉES

*F. Magnien et J. Pournard*

## 1. Introduction

Les substitutions que les consommateurs effectuent entre produits ou entre lieux d'achats constituent l'une des difficultés majeures de la construction d'indices de prix. Comment, en effet, mesurer entre deux dates des évolutions différentes - voire divergentes - de prix, tout en prenant en compte les substitutions induites ? Il faut, d'une façon ou d'une autre, agréger les prix au moyen des quantités vendues alors que ces quantités ne cessent d'évoluer. Le plus simple consiste à les fixer. On obtient ainsi les indices de Laspeyres et de Paasche directs suivant que sont retenues les quantités initiales ou bien finales :

$$L_{T/0}^D = \frac{\sum_s q_0^s p_T^s}{\sum_s q_0^s p_0^s} = \sum_s w_0^s \frac{p_T^s}{p_0^s} \text{ où } w_0^s = \frac{q_0^s p_0^s}{\sum_s q_0^s p_0^s}$$

$$P_{T/0}^D = \frac{\sum_s q_T^s p_T^s}{\sum_s q_T^s p_0^s} = \left[ \sum_s w_T^s \left( \frac{p_T^s}{p_0^s} \right)^{-1} \right]^{-1} \text{ où } w_T^s = \frac{q_T^s p_T^s}{\sum_s q_T^s p_T^s}$$

Dans ces expressions,  $s$  désigne un produit dans un point de vente (une « série »), 0 la période initiale, dite de « base » et  $T$  la période finale (sous revue). Malgré leur mise en oeuvre particulièrement simple, ces indices ont rapidement montré leurs limites sur longue période. En effet, ils interdisent catégoriquement la prise en compte des substitutions.

Une autre approche consiste à agréger les prix à l'aide des quantités courantes. Elle conduit à la notion d'indice de prix moyen, encore appelé indice de "valeurs unitaires", apprécié des professionnels :

$$M_{T/0} = \frac{\sum_s q_T^s p_T^s / \sum_s q_T^s}{\sum_s q_0^s p_0^s / \sum_s q_0^s}$$

La critique immédiate soulevée par ce type d'indice est la suivante : si des substitutions s'opèrent des points de vente les plus chers vers les moins chers bien que les prix restent fixes, l'indice baissera. Ainsi, alors que l'on reproche aux indices de Laspeyres et de Paasche de ne pas prendre en compte les substitutions, c'est la critique inverse qui est faite aux indices de valeurs unitaires. A. Saglio (1995) et W. J. Hawkes (1995) ont analysé l'écart entre indice de Laspeyres et indice de prix moyen.

La théorie économique n'a certes pas encore trouvé l'indice idéal ; elle propose cependant une analyse satisfaisante de la prise en compte des substitutions : l'approche dite de l'*indice à utilité constante* (IUC). L'IUC<sup>1</sup> est le rapport de deux dépenses : la dépense initiale du consommateur et une dépense courante fictive celle qui, compte tenu des changements de prix, assure au moindre coût, par des substitutions appropriées entre produits ou points de vente, un niveau d'utilité à la période courante égal à ce qu'il était initialement. L'écart entre l'indice de Laspeyres direct et l'IUC est souvent (ce sera le cas ici) appelé "biais de substitution".

La mise en oeuvre de l'IUC est délicate puisqu'elle suppose connue la fonction d'utilité des consommateurs. Deux méthodes ont été développées. La première, que l'on peut qualifier de *paramétrique*, consiste à faire l'hypothèse que cette fonction d'utilité est d'une forme spécifiée : quadratique, Cobb-Douglas, ... et à considérer l'IUC associé, dit "exact" pour la forme d'utilité retenue (suivant la terminologie de Diewert (1976)). On montre ainsi que l'indice de Fisher :

$$F_{T/0}^D = \sqrt{L_{T/0}^D P_{T/0}^D}$$

est exact pour la fonction d'utilité quadratique homogène ; l'indice de Törnqvist

$$T_{T/0}^D = \prod_s \left( \frac{p_T^s}{p_0^s} \right)^{\frac{w_0^s + w_T^s}{2}}$$

est exact pour la fonction de coût translog, alors que la moyenne géométrique pondérée :

---

1. Cost-of-living (COL) index en anglais.

$$G_{T/0}^D = \prod_s \left( \frac{p_T^s}{p_0^s} \right)^{w_0^s}$$

est exacte pour la fonction d'utilité de Cobb-Douglas (Cf. Diewert (1976)). De fait, la formule géométrique simple :

$$GS_{T/0} = \prod_s \left( \frac{p_T^s}{p_0^s} \right)^{1/n}$$

où  $n$  désigne le nombre de séries, est utilisée en France dans le calcul de l'indice des prix à la consommation (IPC) pour certaines variétés<sup>2</sup>. Les indices de Fisher et de Törnqvist sont considérés comme de bonnes approximations de l'IUC : c'est souvent par rapport à eux que l'on mesure le biais de substitution de l'indice de Laspeyres.

La seconde méthode utilisée pour le calcul de l'IUC, dite non paramétrique, s'appuie sur la "théorie des préférences révélées" développée par Afriat (1967)<sup>3</sup> : plutôt que de se donner une fonction d'utilité d'une forme particulière à partir de laquelle on obtient une expression analytique de l'IUC, on infère cette fonction d'utilité des données (prix et quantités) en exploitant l'hypothèse sous-jacente à l'IUC selon laquelle les agents sont rationnels, c'est à dire cherchent à atteindre un niveau d'utilité donné au moindre coût.

Une approche du problème des substitutions alternative à celle de l'IUC est possible : le *chaînage d'indices*. Le principe consiste à découper le temps en intervalles de longueurs égales et à remettre à jour les pondérations au début de chacune de ces périodes ; à l'intérieur des périodes, on procède au calcul d'un indice direct :

$$I_{T/0}^C = \prod_{t=1}^T I_{t/t-1}^D$$

L'une des questions fondamentales dans le chaînage d'indices est le choix de la durée séparant deux chaînages successifs. Alors que les États-Unis attendent plus de dix ans pour réviser leurs pondérations et que les Allemands chaînent leur indice tous les cinq ans, la France le fait tous les ans. On peut penser que cette durée doit être la plus courte possible. Le recours progressif à des micro-données (scanner data), dont l'un des avantages est de fournir avec une fréquence élevée (elle peut être hebdomadaire) les quantités en plus des prix, permet d'envisager de réduire encore la durée séparant deux chaînages successifs, et d'approcher ainsi l'indice de Divisia<sup>4</sup>.

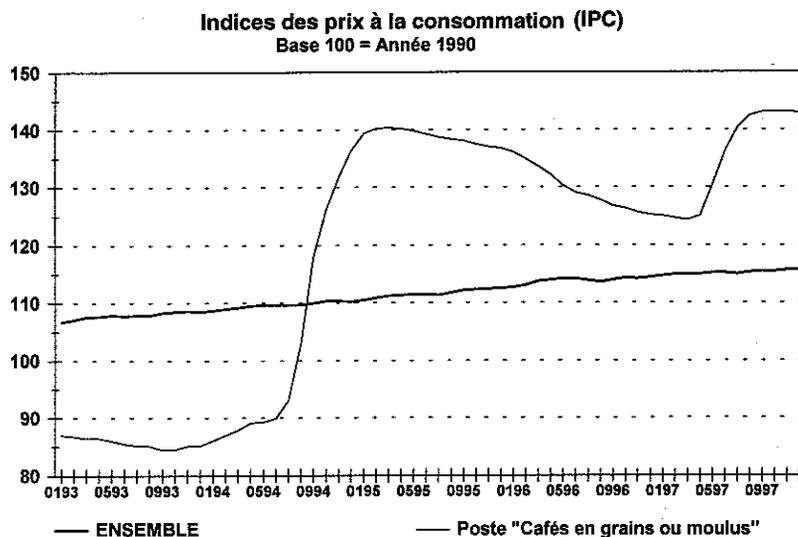
2. Pour le calcul mensuel de l'IPC, on ne dispose pas des quantités mais seulement des prix.

3. Voir aussi Diewert (1973).

4. Rappelons - résultat classique - qu'avec une fonction d'utilité homothétique, l'IUC coïncide avec l'indice de Divisia (Cf. Huhten (1973), Reinsdorf (1998)).

Certains travaux, notamment de B. Szulc (1983), laissent penser qu'en période de prix chahutés, un chaînage trop fréquent peut aller à l'encontre d'une meilleure prise en compte des substitutions. Szulc exprime mathématiquement l'écart à l'unité du ratio "Laspeyres chaîné/Laspeyres direct" comme une accumulation de taux (propres à chaque période de chaînage) dont les signes dépendent d'une part des substitutions opérées par les consommateurs et, d'autre part, du "rebond" des prix.

Des micro-données fournies par la société AC Nielsen nous ont permis de tester cette analyse. Ces données portent sur le café pour une période, 1994-1996, durant laquelle les cours ont connu de très fortes variations :



En effet, l'année 1994 a été le théâtre d'une flambée des cours par suite de gelées au Brésil<sup>6</sup>, premier producteur mondial. En 1995, les cours se sont stabilisés puis ont baissé régulièrement en raison notamment d'une réduction générale des stocks. En 1996, la conjonction d'une mauvaise récolte et d'une restriction des exportations chez les deux principaux producteurs, le Brésil et la Colombie, a enrayé cette baisse.

L'utilisation de données scannées est complexe. Ces données contiennent en effet une masse très importante d'informations : outre les prix et les quantités, une description extrêmement détaillée des points de vente et surtout des produits est fournie. Il en résulte un nombre considérable de séries : environ 30 000 pour le café contre 500 seulement dans l'IPC français !

5. D'autres études ont été menées à partir de données scannées sur le café : Haan-Opperdoes (1997a), Hawkes (1995), Reinsdorf (1995) notamment.

6. Phénomène amplifié par le mécanisme du marché à terme sur lequel s'échange la production mondiale.

Le problème des séries manquantes devient alors aigu. A priori, la multiplicité des séries manquantes peut être imputée à la multiplicité des produits suivis : relativement, le nombre des points de vente de la base AC Nielsen n'est guère plus élevé que dans l'IPC. Il s'avère que c'est la dispersion entre points de vente des taux de distribution des produits<sup>7</sup> qui explique d'abord la disparition de séries. En outre, la dispersion des prix résulte largement de la diversité des produits, de sorte que c'est d'abord entre produits que les substitutions ont lieu (il convient de préciser ici que le panel AC Nielsen utilisé dans cette étude ne contient que des hypermarchés et des supermarchés). Ces considérations nous ont amenés à résoudre le problème des séries manquantes par agrégation des lieux d'achats selon quatre formes de vente. Les données, hebdomadaires, ont également été agrégées mensuellement.

Le calcul des indices définis précédemment a pu ainsi être mené pour le "poste" café et pour chacune des "variétés" qui le constituent. Les résultats obtenus mettent en lumière le phénomène de dérive du chaînage :

(i) pour le poste et la plupart des variétés, une dérive régulière des indices chaînés par rapport aux indices directs, l'indice de Laspeyres (de Paasche) chaîné majorant (minorant) l'indice direct ;

(ii) cette dérive, légère avec un chaînage annuel, devient importante avec un chaînage mensuel ; à plusieurs reprises, l'indice chaîné mensuellement *augmente* quand l'indice direct *diminue* (phénomène inverse avec l'indice de Paasche).

Il était naturel de soumettre ces faits, confirmés par d'autres études récentes<sup>8</sup>, au crible de l'analyse proposée par B. Szulc. Au niveau du poste, les résultats sont partagés : l'explication de la dérive des indices chaînés par une corrélation négative entre la variation courante des prix et leur variation depuis la période de base (le phénomène de « rebond ») n'est vraiment acceptable qu'en 1996 ; toutefois, pour certaines variétés, l'adéquation des résultats obtenus avec l'analyse de Szulc s'avère excellente, dans le cas du café moulu décaféiné en 1996 notamment.

L'organisation de la suite de cette étude est la suivante :

2. Description du panel AC Nielsen sur le café
3. L'instabilité des séries
4. Gestion des séries manquantes
5. Calculs d'indices directs et chaînés
6. Analyse de la dérive du chaînage
7. Conclusion

---

7. On entend par là la proportion de l'ensemble des produits qui sont disponibles dans un point de vente.

8. Cf. note de bas de page 5.

## 2. Description du panel AC Nielsen<sup>9</sup>

La base de données sur laquelle s'appuie cette étude est une extraction du panel SCANTRACK de la société AC Nielsen relative au café torréfié. Elle contient, pour un *échantillon* important d'hypermarchés et supermarchés, et pour la quasi-totalité des produits offerts sur le marché, les *chiffres d'affaires* et les *quantités* vendues chaque semaine d'une période allant de janvier 1994 à décembre 1996.

### 2.1. Les points de ventes

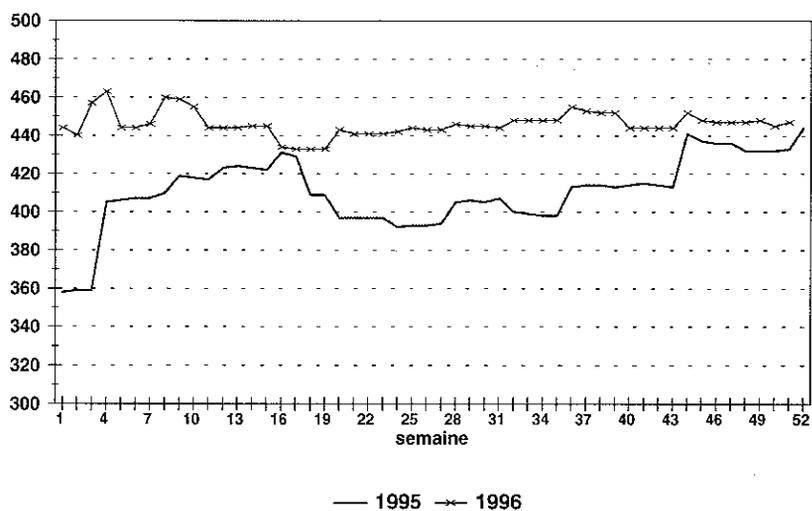
Dans le panel AC Nielsen, un point de vente est décrit à l'aide de trois critères : la *forme de vente*, l'*enseigne* et la *localisation*. Les seules formes de vente retenues sont les hypermarchés et les supermarchés. Les hypermarchés de moins de 6 500m<sup>2</sup> (HM-) ont été différenciés des autres (HM+) ; de même, les supermarchés de moins de 1 200m<sup>2</sup> (SM-) ont été distingués des autres (SM+). Bien que retenues par AC Nielsen comme critère de stratification pour le tirage de leur échantillon de points de vente, les enseignes auxquelles appartiennent les points de vente ne nous ont pas été communiquées. Quant à leur localisation géographique, seuls les départements dans lesquels sont situés les points de vente nous ont été notifiés. La localisation des points de vente n'intervient pas dans cette étude.

Sur la période 1994-1996, 575 points de ventes différents ont fait partie au moins une fois du panel. En tendance, le nombre de points de vente n'a cessé d'augmenter, surtout en 1994, première année de suivi du panel. En 1996, l'effectif des points de vente s'est stabilisé, autour de 440 (graphique 1).

---

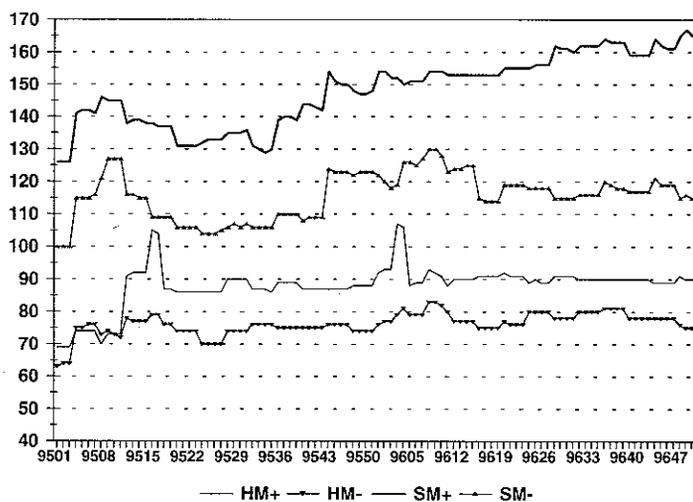
9. Utilisé dans cette étude.

**Graphique 1 : Evolution du nombre de points de vente**



Par forme de vente, on retrouve les mêmes évolutions. Cependant, le nombre de supermarchés a augmenté plus fortement, notamment les plus grands (SM+).

**Graphique 2 : Evolution du nombre de points de vente par forme de vente**



La répartition par formes de vente du chiffre d'affaires ("extrapolé" - cf. Section 2.3.) du café est restée stable au cours des trois années : environ 27 à 28 % ont été réalisés par les grands hypermarchés, 20 % par les plus petits, 35 % par les grands supermarchés et entre 15 et 16 % par les autres.

## 2.2. Les produits élémentaires

Un *produit élémentaire* est défini de façon extrêmement fine par le croisement d'un ensemble de caractéristiques. Les caractéristiques retenues dans cette étude<sup>10</sup> sont au nombre de 10, regroupées, pour la présentation ci-dessous en 3 catégories :

*Production et commercialisation :*

1. Fabricant (ou distributeur)
2. Marque
3. Référence

*Conditionnement :*

4. Type d'emballage d'un paquet (bocal, boîte, carton, ...)
5. Nombre de paquets vendus ensemble
6. Poids de l'ensemble

*Description du "contenu" :*

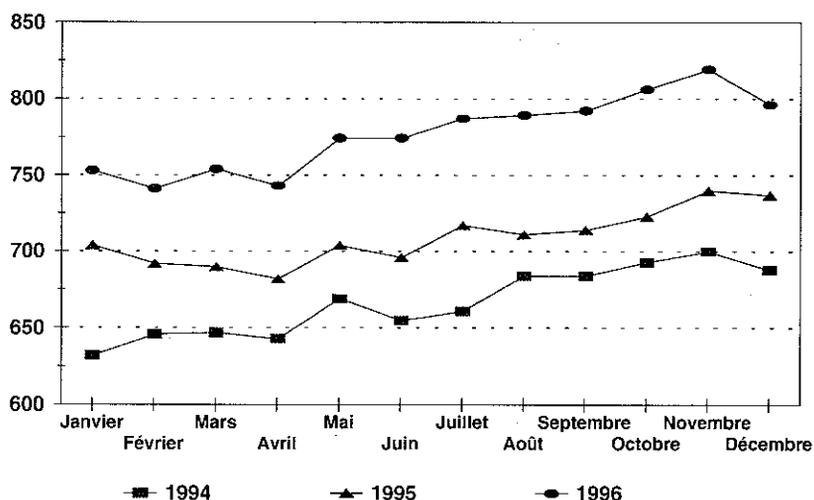
7. Le type (grain, moulu normal, moulu expresso)
8. La qualité (normal, décaféiné)
9. La gamme (arabica, robusta, mélange)
10. L'origine (Brésil, Colombie, ...)

A chacune de ces 10 caractéristiques est associé un certain nombre de modalités (de 2 pour la qualité à 30 pour l'origine). Il s'ensuit que le nombre de produits élémentaires envisageables est très élevé. Seule une partie de ces produits est fabriquée et commercialisée : près de 1 200 produits élémentaires différents ont appartenu au panel au moins une fois au cours des trois années. Ces produits élémentaires se répartissent entre les différents types de café : 871 pour le café moulu normal, 121 pour le café moulu expresso et 176 pour le café en grains. Selon la gamme, il y a 709 arabica, 381 mélanges et 78 robusta.

---

10. Les autres caractéristiques sont relatives au suivi technique du panel. Elles ne sont pas liées à la nature du produit. Elles n'ont donc pas été prises en considération, grâce à un travail d'agrégation préalable (cf. Section 4).

**Graphique 3 : Evolution du nombre de produits élémentaires**  
(nombre moyen présent par mois)



Le nombre de produits élémentaires suivis dans le panel a cru régulièrement, passant de 630 en janvier 1994 à 800 en décembre 1996.

Dans l'IPC, on regroupe les produits élémentaires par *variétés*. On calcule alors des indices de prix pour chaque variété puis on agrège ces indices pour obtenir les indices des *postes*. Il est en effet intéressant, pour l'analyse de l'évolution des prix, de disposer d'indice pour des regroupements étroits de produits à l'intérieur de chaque poste ; en outre, l'absence de relevés de quantités dans l'IPC *nécessite* de passer par ce niveau et même par le niveau plus élémentaire encore des « variété-agglomération ». Avec les micro-données, cette contrainte technique ne se pose pas : le calcul d'indice par variété n'a alors pour objet que de mettre en évidence des évolutions de prix différentes à l'intérieur d'un poste.

Le poste « café en grains ou moulu » de l'IPC est représenté par trois variétés. La richesse du panel AC Nielsen permet d'en introduire deux autres. Au total, les variétés suivantes ont été retenues :

- (1) Café moulu normal, non décaféiné, robusta ou mélange
- (2) Café moulu normal, décaféiné
- (3) Café moulu normal, non décaféiné, arabica
- (4) Café en grains
- (5) Café moulu expresso

Par exemple, la variété (1) est constituée des produits élémentaires qui prennent la modalité "moulu normal" de la caractéristique 7, la modalité "non décaféiné" de la

caractéristique 8 et l'une ou l'autre des modalités "robusta" ou "mélange" de la caractéristique 9, les modalités des autres caractéristiques étant quelconques.

**Tableau 1 : Pondération annuelle<sup>(1)</sup> des variétés selon leur chiffre d'affaires**

	1994	1995	1996
Variété 1	24,6	28,1	25,8
Variété 2	7,1	7,1	7,3
Variété 3	50,5	48,4	50,0
Variété 4	4,8	4,3	4,1
Variété 5	13,0	12,1	12,8
Poste	100	100	100

(1) En %

Dans l'IPC, les variétés<sup>11</sup> sont classées selon deux catégories : les *homogènes* et les *hétérogènes*. Les variétés homogènes réunissent des produits très proches par leurs caractéristiques, donc de prix voisins. Il en est ainsi des variétés 1 et 3. Les autres (2, 4 et 5) sont hétérogènes. L'absence de relevés de quantités conduit à utiliser des formules de calcul différentes (au niveau des agglomérations), selon que l'on a affaire à une variété homogène ou hétérogène. Cette distinction n'est pas nécessaire pour le calcul d'indices dès lors que l'on dispose, comme c'est le cas avec le panel AC Nielsen, de relevés de quantités.

### 2.3. Quantités et prix

Chaque semaine  $t$ , pour chaque couple  $s = (i, j)$ , où  $i$  désigne un produit élémentaire et  $j$  un point de vente, sont relevés une quantité  $q_t^s$  et un chiffre d'affaires  $ca_t^s$ . Le panel contient, en outre, des *coefficients d'extrapolation*  $\lambda_t^s$  permettant, pour chaque produit élémentaire  $i$  et chaque semaine  $t$ , d'estimer par

$$\sum_j \lambda_t^{i,j} q_t^{i,j} \text{ et } \sum_j \lambda_t^{i,j} ca_t^{i,j}$$

la quantité vendue et le chiffre d'affaires réalisé dans l'*univers* des points de vente (dans les expressions précédentes,  $j$  parcourt l'*échantillon* des points de vente de la semaine  $t$ ). Intuitivement, l'application de ces coefficients d'extrapolation revient à

11. dites "ordinaires" ; il y a aussi les *produits frais*, les *biens durables* et les *tarifs*. Le poste « café » de l'IPC ne contient que des variétés ordinaires,

considérer qu'à la semaine  $t$ , chaque point de vente  $j$  du panel représente un ensemble de  $\lambda_t^s$  ( $s = (i, j)$ ) points de vente pour la vente du produit  $i$ .

En fait, ce coefficient d'extrapolation est le même pour tous les produits élémentaires, ne dépend des points de vente que par les "strates" auxquelles ils appartiennent et des semaines que par les mois dont elles font partie. En effet, le panel AC Nielsen résulte d'un sondage aléatoire stratifié des points de ventes (avec allocation optimale au sens de Neyman à l'intérieur de chaque strate). Les critères de stratification retenus sont la taille de la forme de vente (HM+, HM-, SM+, SM-), l'enseigne, la localisation régionale et la taille de l'agglomération.

Pour un produit élémentaire  $i$ , la quantité vendue au cours d'une semaine  $t$  d'une année donnée  $A$  dans l'ensemble (et non l'échantillon) des points de vente  $j$  d'une strate  $J$  est

$$Q_t^i(J) = \sum_{j \in J} q_t^{i,j}$$

Comme le chiffre d'affaires annuel pour l'ensemble (exhaustif) des points de vente  $j$  de la strate  $J$  et de l'ensemble des produits  $i'$  (pas seulement les produits élémentaires du café) :

$$\sum_{t \in A} \sum_{i'} \sum_{j \in J} q_t^{i',j} p_t^{i',j}$$

est connu, la quantité  $Q_t^i(J)$  peut être estimée suivant la méthode du ratio par :

$$\sum_{j \in S_t(J)} \frac{\sum_{t \in A} \sum_{i'} \sum_{j \in J} q_t^{i',j} p_t^{i',j}}{\sum_{t \in A} \sum_{i'} \sum_{j \in S_t(J)} q_t^{i',j} p_t^{i',j}} q_t^{i,j}$$

où  $S_t(J)$  est l'échantillon des points de vente de la strate  $J$  suivi la semaine  $t$ . Ainsi, le coefficient d'extrapolation pour une année donnée

$$\lambda_t^J = \frac{\sum_{t \in A} \sum_{i'} \sum_{j \in J} q_t^{i',j} p_t^{i',j}}{\sum_{t \in A} \sum_{i'} \sum_{j \in S_t(J)} q_t^{i',j} p_t^{i',j}}$$

ne dépend pas des produits et ne dépend que des strates auxquelles les points de vente appartiennent. En outre, l'échantillon  $S_t(J)$  n'est révisé que toutes les quatre

semaines : il ne dépend que du mois  $m$  auquel  $t$  appartient (on écrira donc  $S_m(J)$  et  $\lambda_m^J$ ).

Par forme de vente, le nombre de points de vente que contient le panel diffère de ce que l'on observe dans l'univers des points de vente : les formes de vente les plus petites sont sous-représentées (il y avait réellement 347 HM+, 683 HM-, 2 784 SM+ et 3 887 SM- en janvier 1996 contre respectivement 97, 80, 152 et 122 dans le panel). Les multiplicateurs corrigent cette distorsion puisqu'ils sont d'autant plus grands que la taille de la forme de vente est petite.

Dans cette étude, il a été décidé de travailler sur des données mensuelles. On retrouve ainsi le rythme de calcul de l'IPC et de révision de l'échantillon des points de vente du panel AC Nielsen. Ceci permet en outre d'avoir une base de données moins lourde et donc plus maniable. Surtout, le problème de la perte de « séries » (cf. Section 3) s'en trouve notablement réduit. Ainsi, pour chaque couple  $s = (i, j)$  constitué d'un produit élémentaire  $i$  et d'un point de vente  $j$ , le nombre d'unités vendues et le chiffre d'affaires des quatre semaines d'un même mois  $m$  ont été agrégés, ce qui a permis de calculer un prix mensuel moyen :

$$p_m^s = \frac{\sum_{t \in m} ca_t^s}{\sum_{t \in m} q_t^s} = \frac{ca_m^s}{q_m^s} \quad (2.3.)$$

Ce prix est *unitaire* (les statisticiens parlent de "valeurs unitaires"). Or, suivant les modalités de la caractéristique 6 d'un produit élémentaire, la quantité (exprimée en grammes) de café contenue dans une *unité vendue* en magasin est variable d'un produit à un autre. D'où le choix d'une *unité de compte* commune : 250 grammes. C'est relativement à cette unité de compte que les prix et quantités seront dorénavant mesurés.

Si l'on revient un instant sur les formules de calcul des différents indices présentées dans l'introduction, il apparaît que toutes sont fondées sur des ratios  $p_T^s / p_0^s$  où

$p_m^s$  ( $m = 0, T$ ) est calculé selon 2.3. Il s'agit donc d'indices de valeurs unitaires. Balk (1998) a montré que ces indices sont très proches de l'IUC si les préférences sont représentées par une fonction d'utilité simplement additive (i.e. égale à la somme des quantités consommées). Cette hypothèse est raisonnable au niveau de désagrégation des produits élémentaires (si les quantités consommées d'une semaine à l'autre n'évoluent pas trop).

## 2.4. Les promotions

Le café est évidemment l'objet de *promotions*<sup>12</sup> ; c'est, entre autre, par ces promotions que la baisse des prix, amorcée en début d'année 1995, a été répercutée aux prix de détail. Les promotions prennent essentiellement la forme de quantités supplémentaires ou de réductions de prix.

Lorsqu'un produit élémentaire  $i$  est l'objet, une semaine  $t$  dans un point de vente  $j$ , d'une promotion, les modalités des caractéristiques 5 et 6 restent *inchangées* : ce sont les modalités hors promotion. Les promotions ne se traduisent donc jamais par l'apparition d'un nouveau produit élémentaire. Lorsqu'une promotion prend la forme d'une quantité supplémentaire offerte, celle-ci entraîne<sup>13</sup> une augmentation de la quantité  $q_i^s$ . Une promotion offrant une réduction de prix induit<sup>14</sup> une baisse du chiffre d'affaires  $ca_i^s$ .

## 3. L'instabilité des séries

### 3.1. La notion de série

Les Sections 2.3 et 2.4 ont mis en évidence l'importance des couples formés d'un produit élémentaire et d'un point de vente. Un tel couple s'appelle une *série*. Il s'agit d'un concept essentiel dans la construction d'indices de prix. Dans le panel AC Nielsen relatif au café, il y avait 26 770 séries en décembre 1994, 34 930 en décembre 1995 et 35 683 en décembre 1996. Ce nombre est considérable puisqu'il n'était que de l'ordre de 500 dans le poste « café » de l'IPC aux mêmes dates. La notion de série recouvre les deux types de substitutions auxquelles procèdent les consommateurs en réaction aux variations de prix, à l'absence des produits ou à d'autres facteurs (campagnes publicitaires, ...) : changements de produit et changements de lieu d'achat. La répartition des séries entre les cinq variétés du poste café est donnée dans le tableau suivant :

---

12. Une promotion est nécessairement indiquée en rayon.

13. En supposant le nombre d'achats inchangé.

14. Cf. note précédente.

**Tableau 2 : Répartition des séries entre variétés**

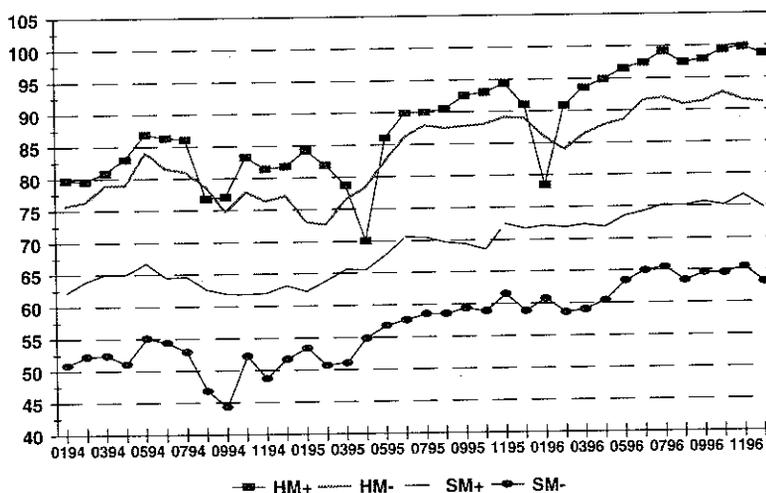
	Décembre 1994		Décembre 1995		Décembre 1996	
	Nombre	%	Nombre	%	Nombre	%
Variété 1	4 845	18,1	5 799	16,6	6 221	17,4
Variété 2	3 332	12,4	4 367	12,5	4 271	12,0
Variété 3	11 462	42,8	15 598	44,7	16 068	45,0
Variété 4	2 647	9,9	3 029	8,7	2 909	8,2
Variété 5	4 484	16,8	6 137	17,6	6 214	17,4
Poste	26 770	100	34 930	100	35 683	100

La variété 3 regroupe à elle seule près de la moitié du nombre des séries. Ceci reflète sa part dans le chiffre d'affaires du poste (cf. tableau 1).

### 3.2. Evolution du nombre de séries

La forte augmentation du nombre de séries résulte à la fois de l'accroissement du nombre de points de vente et du nombre de produits élémentaires suivis dans le panel (graphiques 1 et 3). L'augmentation du nombre moyen de produits élémentaires par point de vente s'observe quel que soit la forme de vente :

**Graphique 4 : Evolution du nombre moyen de séries par point de vente, selon la forme de vente**

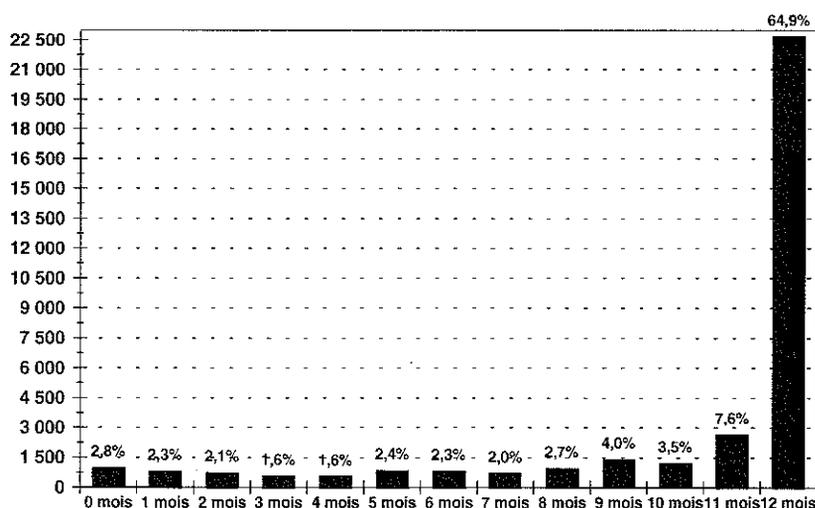


Le graphique 4 est relatif au poste « café » dans son ensemble. A l'intérieur de chaque variété, le nombre moyen de produits élémentaires par point de vente est d'autant plus élevé que la taille de la forme de vente est importante (toutefois, les formes HM+ et HM- ont à peu près le même nombre de séries). Le nombre moyen de séries par produit élémentaire augmente moins vite ; c'est donc d'abord l'élargissement de la gamme de produits offerts qui explique l'accroissement du nombre de séries.

### 3.3. Instabilité des séries dans l'échantillon

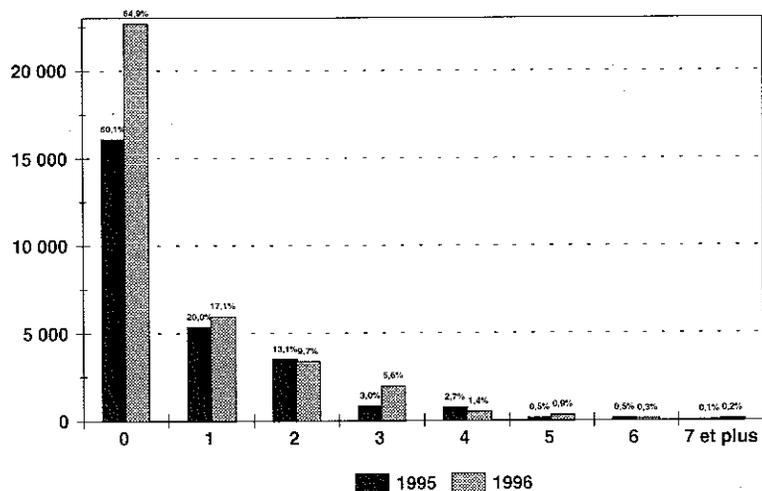
Comme dans l'IPC, les séries du panel ne sont pas systématiquement observables les douze mois d'une année. Le graphique 5 donne par exemple la répartition, en fonction du nombre de mois de présence en 1996, de l'échantillon constitué des séries présentes en décembre 1995.

**Graphique 5 : Répartition, en 1996, des séries du poste « café » selon leur durée de présence dans le panel**



Il apparaît que 65 % de ces séries ont été observées sans discontinuer (60 % en 1995), 7,6 % ont pu être suivies pendant 11 mois, etc ... La présence d'une série est plus ou moins continue. Le graphique 6 montre ainsi que 17,1 % des séries (présentes en décembre 1995) ont quitté l'échantillon pour ne plus y revenir (du moins en 1996), 9,7 % des séries ont quitté une fois l'échantillon puis l'ont réintégré, 5,6 % l'ont quitté, puis réintégré et quitté à nouveau mais définitivement, etc ...

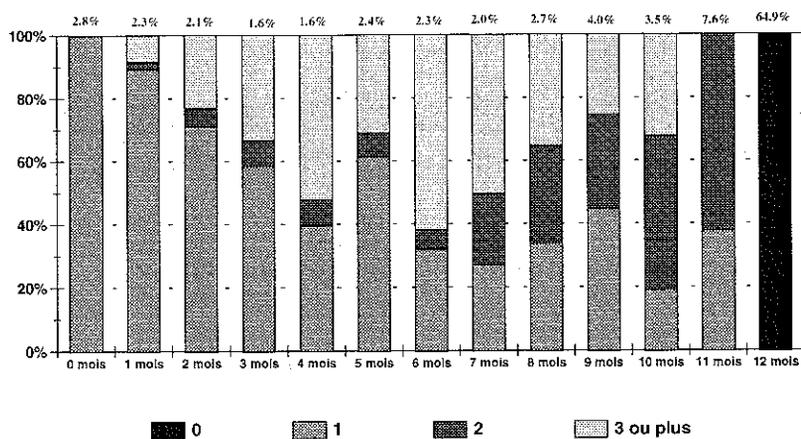
**Graphique 6 : Répartition des séries selon leur nombre de ruptures**



Le graphique 7 précise les deux précédents. Il indique, par exemple parmi les séries dont la durée de présence a été de 10 mois en 1996, qu'environ 20 % ont été observées durant les 10 premiers mois de l'année (elles étaient donc manquantes en novembre et décembre), un peu moins de 50 % ont été manquantes deux mois consécutifs mais étaient présentes en décembre ; les autres (un peu plus de 30 %) ont quitté l'échantillon un mois, l'on réintégré puis l'ont quitté en décembre (où elles n'ont pas été observées).

**Graphique 7 : Stabilité des séries dans le panel en 1996**

(Nombre de ruptures selon la durée de présence)



Plus généralement, on tire trois enseignements de ce graphique :

- dans la limite d'une demi-année environ, plus la durée de présence d'une série est longue, plus cette présence est irrégulière ;
- la probabilité d'une unique absence temporaire (i.e. deux ruptures) augmente nettement lorsque la présence dans l'échantillon est supérieure à six mois ;
- ce sont les séries dont la durée de présence est d'environ six mois qui sont les plus instables dans l'échantillon (au moins trois ruptures).

### 3.4. Une analyse de l'instabilité des séries

Il est difficile, à partir du seul panel, de savoir pourquoi une série, auparavant observée, n'est plus observable un mois donné. Il se peut que le point de vente soit fermé ou bien que la chaîne de distribution à laquelle ce point de vente appartient ne souhaite plus diffuser le produit ; la raison peut également se situer du « côté des produits » : le fabricant est, par exemple, en rupture de stock, ou bien cesse la production d'un produit.

On peut toutefois essayer de mesurer la part de chacun de ces deux effets. Considérons l'ensemble  $S_{t-1}$  des  $N_{t-1}$  séries observables au cours du mois  $t-1$  (les séries de *référence*). Le mois suivant  $t$ , la proportion de séries observables par rapport au mois  $t-1$  est :

$$\delta_t = \frac{1}{N_{t-1}} \sum_{(i,j) \in S_{t-1}} \delta_t^{i,j}$$

où l'indicateur  $\delta_t^{i,j}$  prend la valeur 1 si la série  $(i,j)$  est observable en  $t$  et 0 sinon ( $i$  désigne un produit élémentaire et  $j$  un point de vente). Le *taux d'absence* des séries est donc  $\tau_t = 1 - \delta_t$ . Puisque  $\delta_t$  reste proche de 1 (cf. graphique 8), on a l'approximation :

$$\tau_t \approx \delta_t(1 - \delta_t) = \text{variance empirique de la distribution } \delta_t^{i,j}$$

Elle permet d'expliquer la valeur du taux d'absence  $\tau_t$ . Il suffit en effet de procéder à l'analyse de la variance de la distribution  $\delta_t^{i,j}$ .

Soit, pour cela,  $S_{t-1}^i$  l'ensemble (à  $N_{t-1}^i$  éléments) des points de vente de référence d'un produit élémentaire  $i$ , c'est à dire l'ensemble des lieux d'achats dans lesquels le produit  $i$  était disponible le mois précédent ( $S_{t-1}^i = \{j, (i,j) \in S_{t-1}\}$ ). Le taux de présence du produit élémentaire  $i$  le mois  $t$  dans ses points de vente de référence est :

$$\delta_t^i = \frac{1}{N_{t-1}^i} \sum_{j \in S_{t-1}^i} \delta_t^{i,j}.$$

Soit :

$$\pi_{t-1}^i = N_{t-1}^i / N_{t-1}$$

la part des séries de référence relative au produit élémentaire  $i$ . On a :

$$\delta_t = \sum_i \pi_{t-1}^i \delta_t^i$$

et la quantité :

$$V_t^{produits} = \sum_i \pi_{t-1}^i (\delta_t^i - \delta_t)^2$$

mesure la dispersion, *entre les produits élémentaires*, de leur taux de présence le mois  $t$  dans leurs points de vente de référence. De l'analyse de la variance de la distribution  $\delta_t^{i,j}$  il résulte que

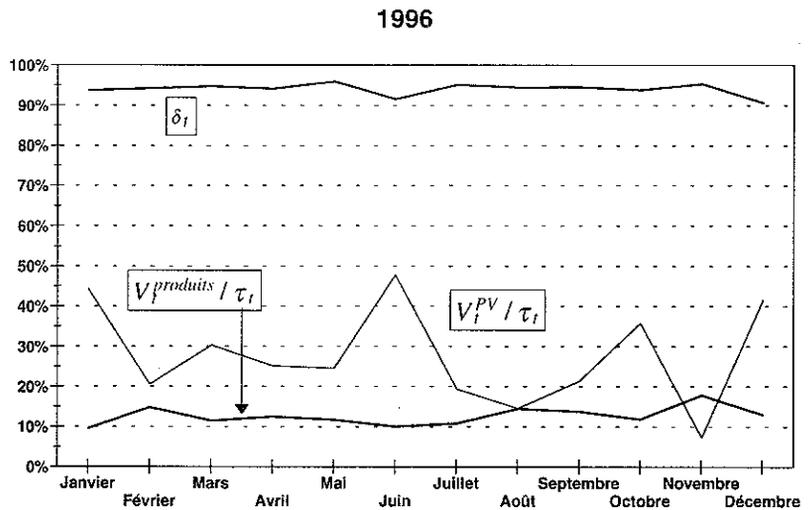
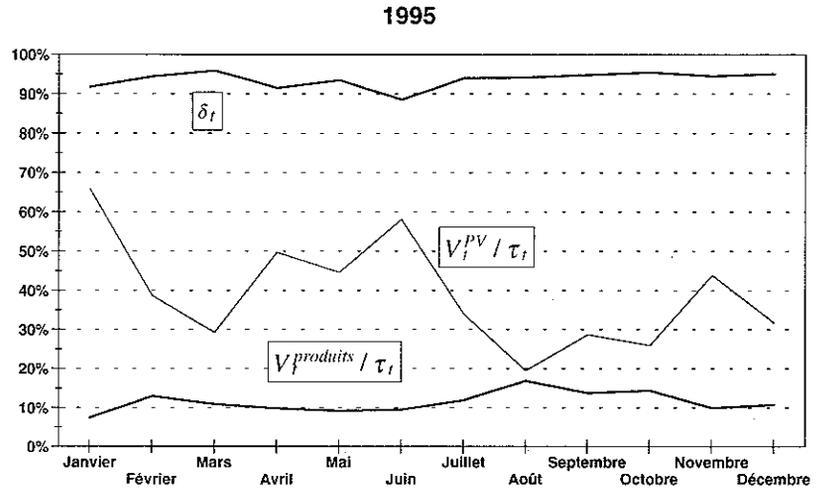
$$V_t^{produits} \leq \tau_t$$

Cette inégalité montre que, mécaniquement, la dispersion  $V_t^{produits}$  joue à la hausse du taux d'absence  $\tau_t$ . Une valeur élevée du ratio  $V_t^{produits} / \tau_t$  signifie que la dispersion, *entre les produits élémentaires*, de leur taux de présence dans leurs points de vente de référence explique largement le taux d'absence des séries.

En inversant les rôles des produits et des points de vente, on obtient une mesure de l'effet "point de vente" sur le taux d'observabilité des séries. Cet indicateur

$V_t^{PV} / \tau_t$  est d'autant plus grand que la dispersion le mois  $t$ , entre les points de vente, des taux de distribution des produits qu'ils distribuait le mois précédent explique largement le taux d'absence des séries.

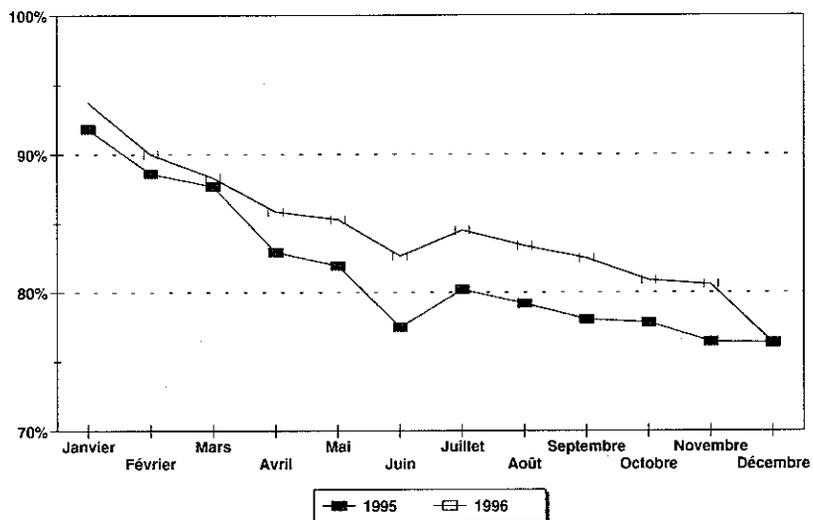
**Graphique 8 : Evolution de  $\delta_t$ , de  $V_t^{produits} / \tau_t$  et  $V_t^{PV} / \tau_t$**



Le graphique 8 montre que l'absence de séries par rapport au mois précédent résulte d'abord de valeur de la dispersion  $V_t^{PV}$ . On observe ainsi, sur ce graphique, que  $\tau_t$  et  $V_t^{PV}$  fluctuent en phase (les fluctuations de  $\tau_t$  sont opposées à celles de  $\delta_t$ , qui apparaissent sur le graphique) : l'accroissement de la dispersion entre points de vente des taux de distribution des produits élémentaires, en juin et novembre 1995 puis en juin, octobre et décembre 1996, s'accompagne chaque fois d'une hausse du taux d'absence des séries. Un élément d'explication peut être avancé : alors que les fabricants cherchent à écouler leur production par tous les canaux de distribution, les distributeurs semblent se montrer plus sélectifs quant aux produits qu'ils choisissent d'offrir sur le marché.

Si l'on prend maintenant pour séries de référence, l'ensemble "fixe" des séries observables le mois de décembre de l'année précédente (34 930 séries en décembre 1995) plutôt que celles du mois précédent, on constate que le taux de présence de ces séries baisse assez *régulièrement* tout au long de l'année (cf. graphique 9). En effet, d'après le graphique 6, en gros 60 % des séries de référence restent présentes toute l'année, le reste se partageant à peu près également entre les séries qui quittent définitivement l'échantillon après une présence continue, et celles qui ont un comportement plus complexe. La première catégorie de séries stabilise le taux de présence des séries, la seconde catégorie tend à sa baisse *régulière*, quant à la troisième catégorie, elle ajoute un bruit à cette tendance.

**Graphique 9 : Evolution annuelle de  $\delta_t$**



## 4. La gestion des séries manquantes

L'instabilité des séries est un problème classique dans la construction d'indices de prix : on doit suivre l'évolution des prix des séries d'un échantillon choisi à la période de base. Il faut donc mettre en place, d'une façon ou d'une autre, une procédure de remplacement. Dans l'IPC, l'une des techniques consiste à affecter à une série manquante l'évolution moyenne des prix des séries relatives à la même variété. C'est également l'approche retenue par certains auteurs lors d'études utilisant des micro-données (De Haan-Opperdoes (1997a)<sup>15</sup>). Une autre approche consiste à « agréger » les séries. Cette opération concerne ici les points de vente.

### 4.1. Agrégation des points de vente par formes de vente

Le nombre de points de vente (hypermarchés et supermarchés) est élevé dans le panel. Or, on a vu à la section 3.4. que la multiplicité des points de vente contribue notablement à accroître le nombre des séries manquantes. En outre, comme on va le voir maintenant, la dispersion des prix des séries entre points de vente est faible, ce qui accrédite l'idée que les substitutions sont essentiellement réalisées entre produits.

Soient  $N_m$  le nombre total de séries  $(i,j)$  observables un mois  $m$  donné et

$$\bar{p}_m = \frac{1}{N_m} \sum_{i,j} p_m^{i,j}$$

leur prix moyen (dans le développement qui suit, chaque série  $(i,j)$  est comptée un nombre de fois égal à  $\lambda_j^m$  où  $J$  est la strate à laquelle  $j$  appartient). On peut décomposer la variance empirique du prix des séries :

$$V_m = \frac{1}{N_m} \sum_{i,j} (p_m^{i,j} - \bar{p}_m)^2$$

en deux termes :

$$V_m = V_m^{\text{produits}} + v_m^{\text{produits}}$$

---

15. Comme dans l'IPC Néerlandais, De Haan-Opperdoes (1997a) considèrent des micro-indices qui n'intègrent aucune pondération.

Dans cette relation,  $V_m^{produits}$  désigne la variance empirique « inter » produits élémentaires :

$$V_m^{produits} = \sum_i \pi_m^i (\bar{p}_m^i - \bar{p}_m)^2$$

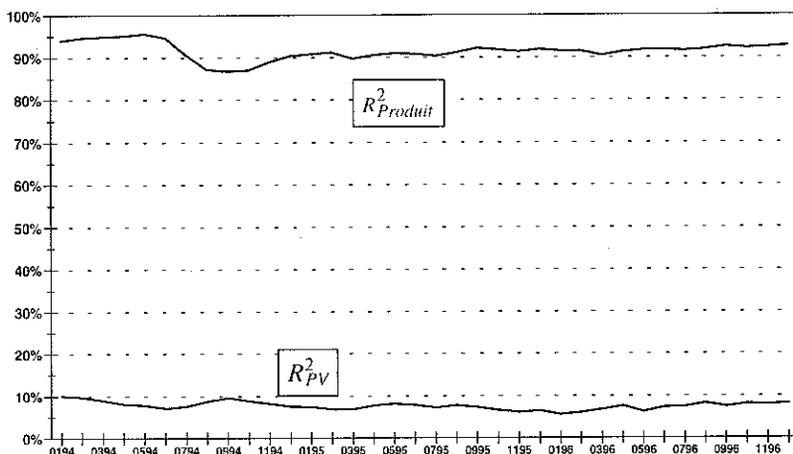
où

$$\bar{p}_m^i = \frac{1}{N_m^i} \sum_j p_m^{i,j}$$

est le prix moyen du produit élémentaire  $i$  entre les différents points de vente  $j$  où il est vendu ( $N_m^i$  est le nombre de points de vente où  $i$  est vendu durant le mois  $m$  et  $\pi_m^i = N_m^i / N_m$ ). Quant au terme  $V_m^{produits}$ , il est égal à la moyenne des variances empiriques « intra » des produits élémentaires entre points de vente.

Le graphique 10 indique la part  $R_{produits}^2 = V_m^{produits} / V_m$  de la variance totale des prix expliquée par la dispersion des prix moyens des produits élémentaires et la part  $R_{PV}^2$  expliquée par la dispersion entre points de vente des prix moyens des produits qui y sont vendus.

**Graphique 10 : Analyse de la variance des prix des séries**



La variance des prix expliquée par les écarts de prix entre points de vente est très faible. On a donc agrégé les points de vente à l'intérieur de chacune des quatre formes de vente (FV). On obtient ainsi 3 127 séries (produits élémentaires, FV). Les prix mensuels moyens de ces séries ont été calculés comme suit :

$$p_m^{i,FV} = \frac{\sum_{j \in FV \cap S^m(J)} \lambda_m^{J(j)} ca_m^{i,j}}{\sum_{j \in FV \cap S^m(J)} \lambda_m^{J(j)} q_m^{i,j}} \quad (FV = HM+, HM-, SM+, SM-)$$

où  $J(j)$  désigne la strate du point de vente  $j$  (cf. Section 2.3.). Grâce à l'utilisation des coefficients d'extrapolation  $\lambda_m^{J(j)}$  (cf. Section 2.3.), l'estimation des prix des séries est plus précise (de façon évidente, dans l'expression 2.3. l'utilisation des coefficients d'extrapolation est inutile).

## 4.2. Elimination des produits nouveaux ou bien retirés de la vente

Afin de pouvoir calculer des indices, on a éliminé les couples (produits élémentaires, FV) qui ne sont pas restés présents pendant 36 mois dans le panel. Il ne restait plus alors que 895 séries (représentant tout de même 92 % du chiffre d'affaires (extrapolé) total des points de vente sur les trois années). Compte tenu de l'agrégation par forme de vente, les séries éliminées correspondaient à des produits nouveaux ou bien retirés, parfois temporairement, du marché. Les produits nouveaux induisent un biais différent du biais de substitution (cf. Turvey (1998)) ; il était donc naturel, dans cette étude, de les écarter.

La répartition des séries entre variétés est donnée dans le tableau 3.

**Tableau 3 : Nombre de séries (produit élémentaire, FV) par variété**

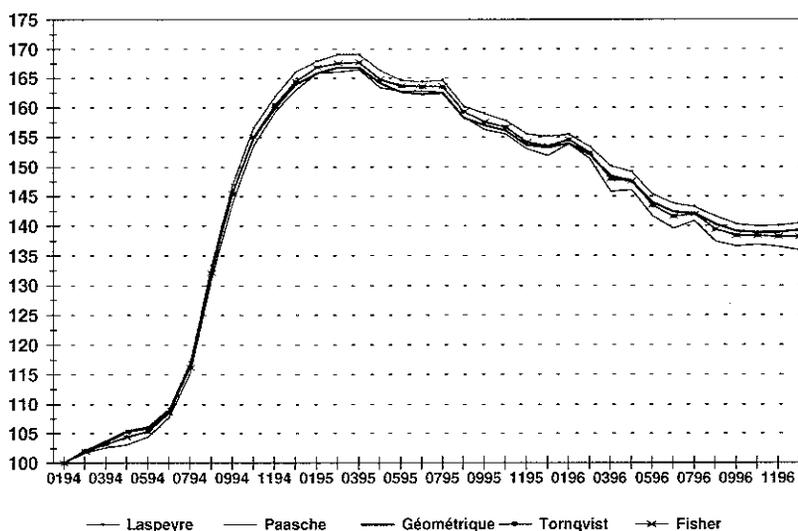
Variétés	nombre de séries(produit élémentaire, FV)
1	153
2	109
3	377
4	135
5	121
Poste	895

## 5. Calculs d'indices directs et chaînés

### 5.1. Les indices directs

La définition des indices de Laspeyres, Paasche, Fisher, Törnqvist et Géométrique a été donnée dans l'introduction (dans les formules de définition,  $s$  parcourt l'ensemble des couples (produit élémentaire, FV) relatifs à une variété ou au poste). Le graphique suivant présente ces différents indices *directs* au niveau du poste :

**Graphique 11 : Les indices directs du poste « café »**  
(Base 100 en Janvier 1994)



Les indices de Laspeyres et de Paasche encadrent les trois autres indices, réputés être proches de l'IUC :

$$P^D < G^D, F^D, T^D < L^D$$

L'indice de Laspeyres sur-pondère en effet les prix qui augmentent le plus alors que l'indice de Paasche sur-pondère ceux qui augmentent le moins. Les indices de Fisher et de Törnqvist sont souvent considérés comme de très bonnes approximations de l'IUC. L'un ou l'autre est pris comme référence pour estimer le biais de substitution, qui mesure l'écart entre l'indice de Laspeyres direct et l'IUC. En décembre 1996, les indices directs de Laspeyres, Törnqvist et Fisher valaient respectivement 140,49, 138,14 et 138,22, soit un biais de 2,35 et 2,27 du Laspeyres par rapport aux deux

autres indices. En pourcentage annuel, le biais est de 0,56 % par rapport à l'indice de Törnqvist et de 0,54 % par rapport à l'indice de Fisher. Mensuellement, le biais de substitution est de 0,047 % et 0,045 %. A l'intérieur du poste, le biais est assez variable (Tableau 4).

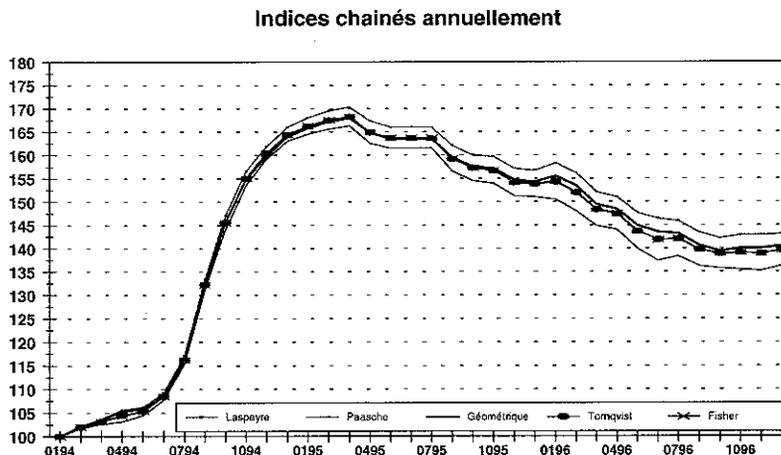
**Tableau 4 : Biais de l'indice de Laspeyres direct, pour le poste et les variétés**

	par rapport à l'indice de Fisher		par rapport à l'indice de Törnqvist	
	par mois (%)	par an (%)	par mois (%)	par an (%)
Variété 1	0,040	0,48	0,04	0,48
Variété 2	0,062	0,75	0,064	0,78
Variété 3	0,036	0,43	0,038	0,45
Variété 4	0,005	0,06	0,003	0,04
Variété 5	0,024	0,28	0,023	0,28
Poste	0,045	0,54	0,047	0,56

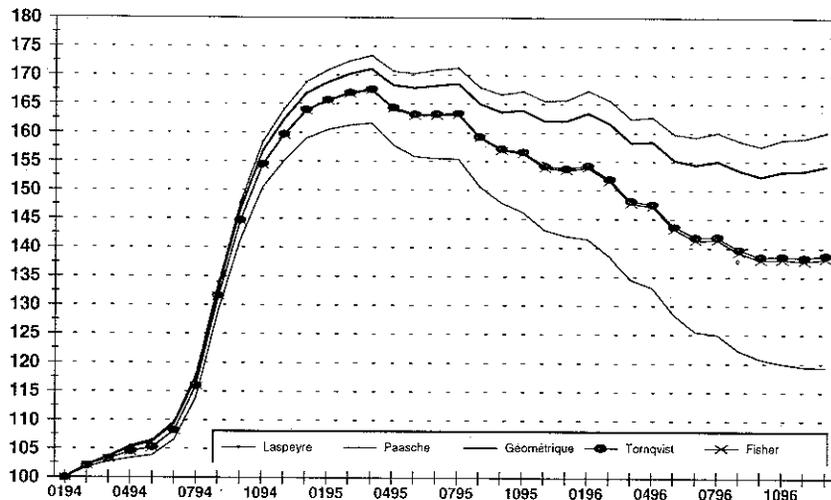
## 5.2. Les indices chaînés

Les faisceaux d'indices directs et chaînés annuellement (comme l'IPC) sont proches. Par contre, ce faisceau s'ouvre largement avec le chaînage mensuel :

**Graphique 12 : Les indices du poste chaînés annuellement et mensuellement (Base 100 en Janvier 1994)**



### Indices chaînés mensuellement

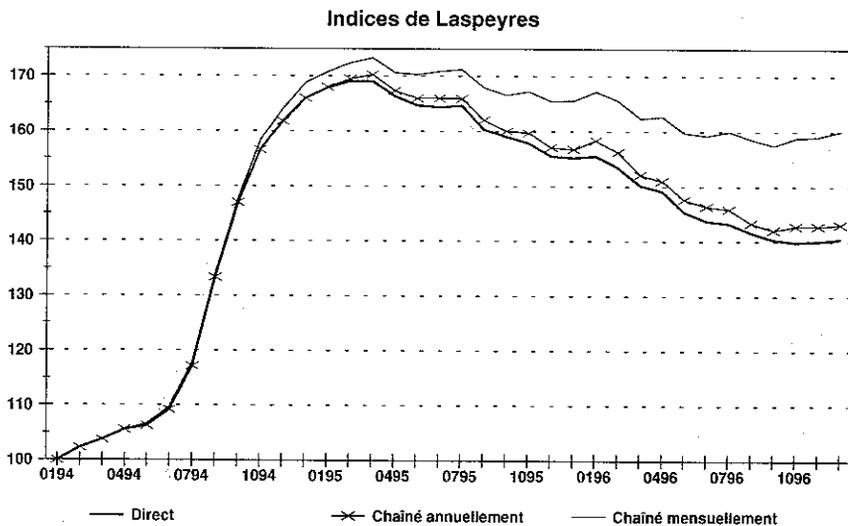


Le constat est le même au niveau des variétés, à l'exception de la variété 2 et, dans une moindre mesure, de la variété 4.

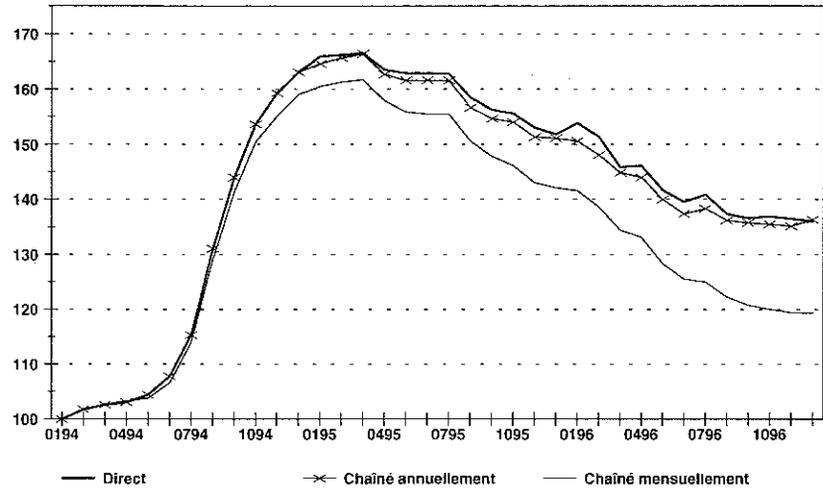
L'indice de Laspeyres chaîné est supérieur à l'indice de Laspeyres direct (l'indice de Paasche chaîné est inférieur au direct). Cette dérive devient très importante avec un chaînage mensuel (graphique 13).

### Graphique 13 : Les indices de Laspeyres et de Paasche, directs et chaînés, pour le poste

(Base 100 en Janvier 1994)



### Indices de Paasche



On observe, à plusieurs reprises, des évolutions contraires des indices direct et chaîné mensuellement. Ce phénomène est assez régulier (tous les 3 à 5 mois), il ne se produit que dans la phase de baisse de l'indice (après mars 1995) ; il se caractérise par une *hausse* (baisse) de l'indice de Laspeyres (Paasche) chaîné *mensuellement* (et parfois annuellement) quand l'indice *direct* enregistre une *baisse* (hausse).

Comme le biais de substitution, la dérive du chaînage est très variable d'une variété à l'autre, notamment avec le chaînage mensuel (Tableau 5). Pour ce type de chaînage, la variété 2 (café moulu décaféiné) est atypique : après une légère dérive passagère, l'indice chaîné mensuellement reste inférieur aux indices directs et chaînés annuellement. Ce phénomène est étudié en détail à la section 6.

**Tableau 5 : Dérive du chaînage de l'indice de Laspeyres, pour le poste et les variétés**

	Dérive du chaînage annuel		Dérive du chaînage mensuel	
	par mois (%)	par an (%)	par mois (%)	par an (%)
Variété 1	0,113	1,369	0,677	8,433
Variété 2	0,028	0,335	-0,029	-0,353
Variété 3	0,033	0,402	0,331	3,795
Variété 4	0,037	0,439	0,090	1,089
Variété 5	0,020	0,241	0,213	2,583
Poste	0,051	0,616	0,364	4,455

## 6. Analyse de la dérive du chaînage

Afin d'analyser le phénomène de dérive de l'indice de Laspeyres chaîné, B. Szulc (1983) a utilisé la relation suivante :

$$L_{T/0}^C / L_{T/0}^D = \prod_{t=1}^T (1 + \rho_t) \text{ avec } \rho_t = cv(x_t)cv(y_t) \cdot r_{xy}^t$$

où

$$x_t = (x_t^s)_x \text{ et } y_t = (y_t^s)_s \quad \text{avec } x_t^s = \frac{p_t^s}{p_{t-1}^s} \text{ et } y_t^s = \frac{q_{t-1}^s}{q_0^s}$$

( $s$  désigne une série) et

$$r_{xy} = (r_{xy}^t), \text{ avec } r_{xy}^t = \frac{\sum_s \tilde{c}_t^s (x_t^s - \bar{x}_t)(y_t^s - \bar{y}_t)}{\sigma(x_t)\sigma(y_t)}$$

$$cv(x_t) = \frac{\sigma(x_t)}{\bar{x}_t}; \quad cv(y_t) = \frac{\sigma(y_t)}{\bar{y}_t}$$

en posant

$$\bar{x}_t = \sum_s \tilde{c}_t^s x_t^s \text{ et } \bar{y}_t = \sum_s \tilde{c}_t^s y_t^s$$

$$\sigma(x_t) = \sqrt{\sum_s \tilde{c}_t^s (x_t^s - \bar{x}_t)^2} \text{ et } \sigma(y_t) = \sqrt{\sum_s \tilde{c}_t^s (y_t^s - \bar{y}_t)^2}$$

Les pondérations  $\tilde{c}_t^s$  sont définies de la façon suivante :

$$c_t^s = \frac{p_{t-1}^s q_0^s}{\sum_s p_0^s q_0^s}$$

$$\tilde{c}_i^s = \frac{c_i^s}{\sum_s c_i^s}$$

Comme  $cv(x_t) > 0$  et  $cv(y_t) > 0$ , l'ordre entre indices de Laspeyres direct et chaîné dépend du signe de la séquence des corrélations  $r_{xy}^t$  entre les variables  $x_t$  et  $y_t$ , c'est à dire entre la variation des quantités vendues entre la période de base et la période précédente, et la variation de prix depuis la période précédente. Szulc a introduit la variable auxiliaire :

$$z_t = (z_t^s) \text{ où } z_t^s = \frac{p_{t-1}^s}{p_0^s}$$

puis considéré que le coefficient de corrélation  $r_{xy}^t$  est généralement de même signe que le produit du coefficient de corrélation  $r_{yz}^t$  entre  $y_t$  et  $z_t$  et du coefficient de corrélation  $r_{xz}^t$  entre  $x_t$  et  $z_t$ . Il fonde cette considération sur l'inégalité :

$$r_{xz} r_{yz} - \sqrt{(1-r_{xz}^2)(1-r_{yz}^2)} \leq r_{xy} \leq r_{xz} r_{yz} + \sqrt{(1-r_{xz}^2)(1-r_{yz}^2)} \quad (6.1)$$

(on omet l'indice  $t$ ) et l'hypothèse que les carrés des corrélations  $r_{yz}$  et  $r_{xz}$  sont relativement proches de 1. Puisqu'en général  $r_{yz} < 0$  (il est à priori raisonnable de penser que les produits dont les prix ont le plus augmenté, sont ceux dont les quantités vendues ont le plus diminué), un indice de Laspeyres chaîné supérieur à l'indice direct résulterait alors de la persistance ou de la prépondérance d'une situation dans laquelle les produits dont les prix ont le moins augmenté depuis la période de base sont ceux dont les prix augmentent le plus à la période courante.

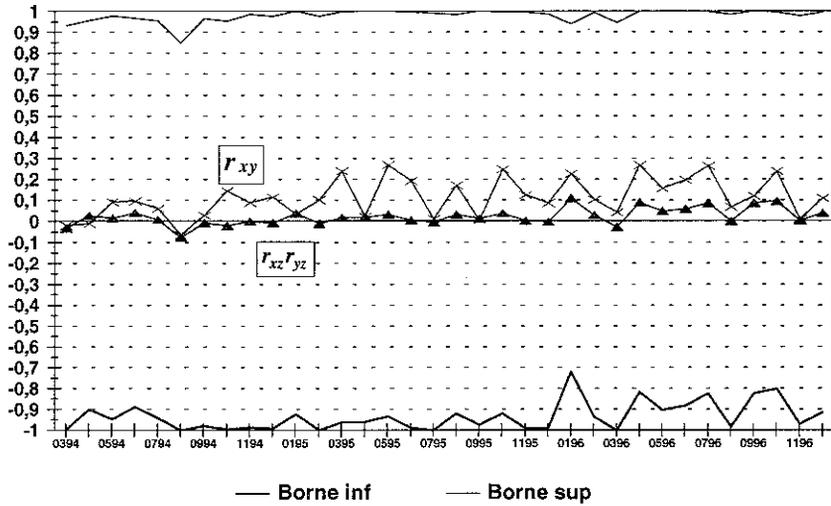
L'examen de cette analyse à partir du panel AC Nielsen sur le café est intéressant puisque les indices de Laspeyres chaînés, annuellement et plus encore mensuellement, majorent l'indice direct (graphique 13).

Pour le poste, on observe - graphique 14 - une positivité quasi-systématique de  $r_{xy}$ . Par contre, selon le même graphique, si l'hypothèse d'évolutions parallèles de  $r_{xy}$  et  $r_{xz} r_{yz}$  (ou du moins de leurs signes respectifs) est satisfaite en 1996, elle ne s'explique en aucune façon par l'inégalité (6.1) : dans le cas du café, les racines carrées sont largement plus proches de 1 que de 0. Cette inégalité semble peu

efficace pour fonder l'hypothèse d'une évolution conjointe de  $r_{xy}$  et  $r_{xz}r_{yz}$  : si

$$r_{yz}r_{xz} = \pm 0,7 \text{ alors } \sqrt{(1-r_{xz}^2)(1-r_{yz}^2)} = 0,5.$$

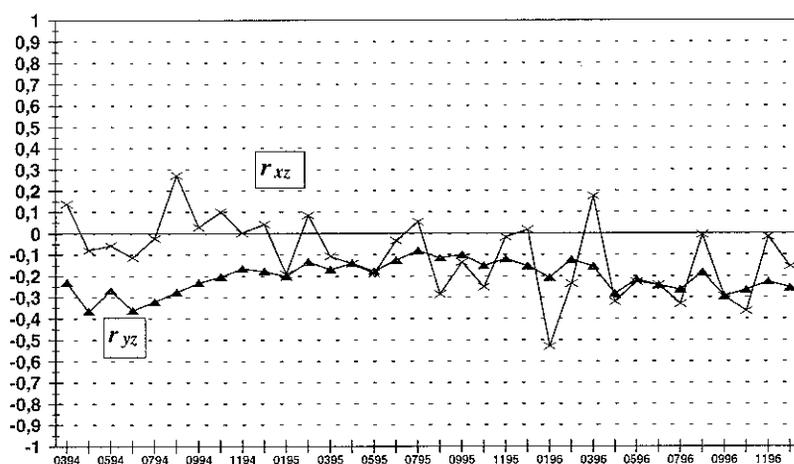
**Graphique 14 : Evolutions de  $r_{xy}$  et  $r_{xz}r_{yz}$  pour le poste (chaînage mensuel)**



$$(\text{Borne inf} = r_{xz}r_{yz} - \sqrt{(1-r_{xz}^2)(1-r_{yz}^2)} \text{ et Borne sup} = r_{xz}r_{yz} + \sqrt{(1-r_{xz}^2)(1-r_{yz}^2)})$$

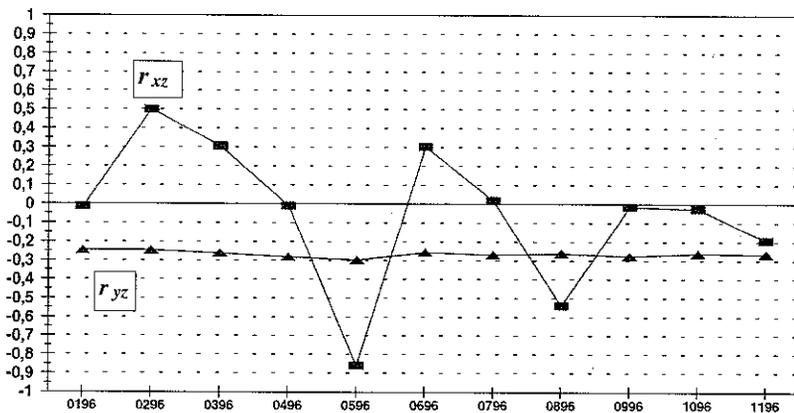
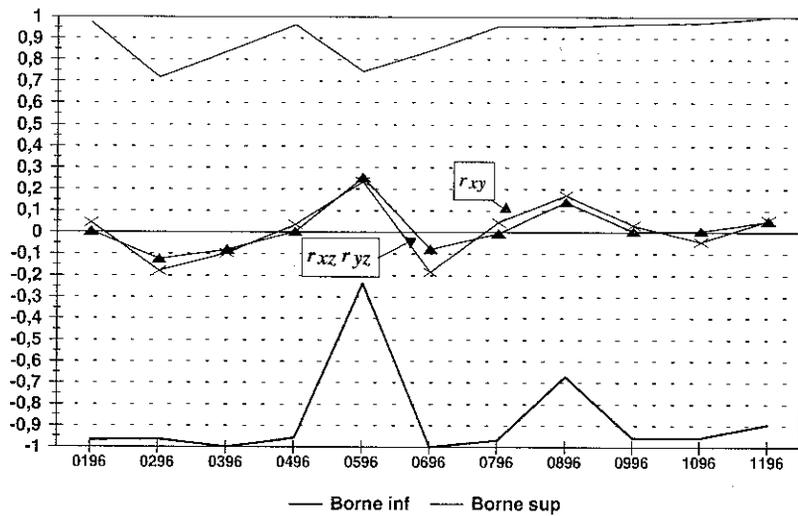
Le coefficient de corrélation de y et z est toujours négatif (graphique 15), de sorte que la hiérarchie entre indices directs et chaînés dépend de l'évolution de la corrélation entre x et z, du moins sur la période où les évolutions de  $r_{xy}$  et  $r_{xz}r_{yz}$  sont parallèles. C'est la situation observée en 1996 ; alors  $r_{xz}$  est le plus généralement négatif.

Graphique 15 : Evolutions de  $r_{xz}$  et  $r_{yz}$  pour le poste (chaînage mensuel)



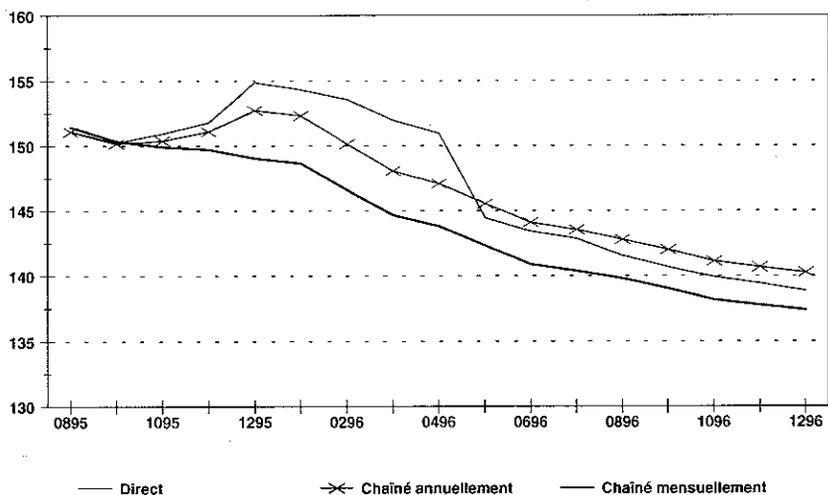
L'adéquation de l'analyse de Szulc aux données relatives au café est encore plus frappante pour certaines variétés (donc plus mauvaise pour d'autres). A cet égard, la variété 2 est tout à fait remarquable. Cette variété est constituée du café moulu décaféiné. Tout au long de l'année 1996, les coefficients  $r_{xz}$  et  $r_{yz}$  restent pratiquement égaux. Là encore ce n'est pas l'inégalité (6.1.) qui justifie cet état de fait : dans le meilleur des cas,  $\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)} \approx 0,3$ . Le coefficient  $r_{yz}$  reste stable aux environs de -0,3 (graphique 16).

Graphique 16 : Evolutions de  $r_{xy}$ ,  $r_{yz}$ ,  $r_{xz}$ , et  $r_{xz}r_{yz}$  pour la variété 2 (chaînage mensuel)



Les évolutions relatives des indices directs et chaînés sont donc complètement déterminées par celle de la corrélation  $r_{xz}$ . En mai 1996, celle-ci est voisine de -1. Si l'indice de Laspeyres direct est supérieur à l'indice chaîné mensuellement, il s'en rapproche brutalement à cette date (graphique 17). Il passe d'ailleurs à ce moment là en dessous de l'indice chaîné annuellement.

**Graphique 17 : Les indices de Laspeyres de la variété 2**  
(base 100 en janvier 1994)



Une mise en garde s'impose : les corrélations calculées dépendent des coefficients  $c_t^s$ . Ces coefficients :

$$c_t^s = \frac{p_{t-1}^s q_0^s}{\sum_s p_0^s q_0^s}$$

sont égaux aux parts respectives des séries dans la dépense *initiale* des consommateurs, *actualisées par les prix* :

$$c_t^s = \frac{p_0^s q_0^s}{\sum_s p_0^s q_0^s} \frac{p_{t-1}^s}{p_0^s}$$

Elles diffèrent des pondérations *courantes* :

$$\frac{p_{t-1}^s q_{t-1}^s}{\sum_s p_{t-1}^s q_{t-1}^s}$$

Il ne faut pas perdre de vue ce fait dans l'interprétation des corrélations. Ainsi, pour la variété 2, la corrélation  $r_{x_t}$ , voisine de -1 en mai 1996, est proche de 0 (-0,1) lorsqu'on la recalcule avec les pondérations courantes.

## 7. Conclusion

L'étude comparative des indices de Laspeyres directs et chaînés a été menée par de nombreux auteurs.

D'abord sur "macro-données", c'est à dire des données utilisées dans le calcul des IPC : prix au mieux relevés mensuellement, pondérations actualisées au mieux annuellement, mais champ étendu à la plus grande partie de la consommation des ménages. Les résultats obtenus sont divers. Les uns montrent que le chaînage a permis de réduire le biais de substitution. Il en est ainsi de Diewert (1978) et Manser-McDonald (1988). Diewert utilise des données en prix et quantités relatives aux dépenses de consommation au Canada sur la période 1947-1971 ; il chaîne tous les cinq ans. Le chaînage élimine la quasi-totalité du biais de substitution évalué avec les indices de Törnqvist ou de Fisher. Manser et McDonald (1988), qui suivent l'approche non paramétrique du calcul de l'IUC (cf. Introduction), obtiennent des résultats analogues pour les Etats-Unis sur la période 1959-1985 : l'écart entre l'indice direct et l'indice chaîné représente près des trois quarts du biais de substitution. Au contraire, les résultats obtenus par Aizcorbe-Jackman (1993), qui portent sur une période plus récente (1982-1991), montrent une légère dérive du chaînage.

L'analyse comparative des indices de Laspeyres directs et chaînés n'a été effectuée sur "micro-données" que récemment, par J. de Haan-E. Opperdoes (1997b), M. Reinsdorf (1995) et J. Dalen (1997) notamment<sup>16</sup>. Les résultats obtenus mettent *tous* en évidence une *importante* dérive du chaînage *mensuel* : loin de réduire le biais de substitution, il l'accroît fortement. Les études de de Haan-Opperdoes et de Reinsdorf portent sur des données scannées relatives au café ; l'étude de Dalen sur les matières grasses, les lessives, les céréales pour le petit déjeuner et le poisson congelé.

L'utilisation, dans la présente étude, du cadre d'analyse de B. Szulc montre que le phénomène de "rebond" des prix autour de leur tendance, c'est à dire leurs fluctuations décalées avec une fréquence élevée, est à l'origine de la dérive du chaînage mensuel. Ceci suggère l'idée d'une valeur optimale de la durée séparant les chaînages successifs. Une durée d'une ou quelques années serait proche de cet optimum : suffisamment longue pour échapper au phénomène de rebond des prix, elle assure une bonne prise en compte des substitutions opérées par les consommateurs.

---

16. Aucun de ces auteurs n'a testé l'analyse de Szulc sur le rebond des prix.

---

## *Bibliographie*

---

- S.N. Afriat (1967), *The construction of a utility function from expenditure data*, International Economic Review, 8, 125-1333.
- A.M. Aizcorbe and P.C. Jackman (1993), *The commodity substitution effect in CPI data, 1982-91*, Monthly Labor Review.
- B.M. Balk (1998), *On the use of unit value indices as consumer price subindices*, Departement of statistical methods, Statistics Netherlands, Voorburg.
- J. Dalen (1997), *Experiments with swedish scanner data*, International Working Group on Price Indices, Statistics Sweden.
- W.E. Diewert (1973), *Afnat and the Revealed Preference Theory*, Review of Economic Studies 40, 419-426.
- W.E. Diewert (1976), *Exact and Superlative Index Numbers*, Journal of Econometrics, vol. 4, n°2.
- W.E. Diewert (1978), *Superlative Index Numbers and Consistency in Aggregation*, Econometrica, Vol. 46, n°4.
- W.J. Hawkes (1995), *Reconciliation of fixed-weight price index trends with corresponding trends in average prices for quasi-homogeneous goods using scanning data*, Nielsen USA, Working paper.
- J. de Haan - E. Opperdoes (1997a), *Estimation of the coffee price index using scanning data : simulation of official practices*, Statistics Netherlands.
- J. de Haan - E. Opperdoes (1997b), *Estimation of the coffee price index using scanning data : the choice of the micro index*, Statistics Netherlands.
- C.R. Hulten (1973), *Divisia index numbers*, Econometrica 41, n°6, 1017-25.
- F. Lequiller (1997), *L'indice des prix à la consommation surestime-t-il l'inflation ?*, Economie et Statistique, n°303.
- M.E. Manser and R.J MacDonald (1988), *An analysis of substitution bias in measuring inflation, 1959-1985*, Econometrica, Vol. 56, n° 4, 909-930.
- M.B. Reinsdorf (1995), *Constructing basic component indexes for the US CPI from scanner data : a test using data on coffee*, U.S. Bureau of Labor Statistics.
- M.B. Reinsdorf (1998), *Divisia Indexes and the Representative Consumer Problem*, Working Paper.

A. Saglio (1995), *Changement de tissu commercial et mesure de l'évolution des prix*, Economie et Statistiques, n° 285-286, pp 9-33.

B.J. Szulc (1983), *Enchaînement des indices de prix*, in *La mesure du niveau des prix*, actes du colloque tenu sous l'égide de Statistic Canada, W.E. Diewert et C. Marquette eds, Canada.

R. Turvey (1998), *New outlets and new products*, Proceedings of the Third Meeting of the International Working Group on Prices Indices, B.M. Balk editor, Statistics Netherlands.