Les indices à utilité constante : une référence pour mesurer l'évolution des prix

François Magnien et Jacques Pougnard*

Les débats récents sur une possible surestimation de l'inflation ont notamment porté sur l'ampleur du *biais de substitution* dans le calcul des indices de prix. Ce biais résulte de l'insuffisante prise en compte, avec un indice de Laspeyres, des transferts d'achats des consommateurs entre produits ou points de vente en fonction de l'évolution différenciée des prix.

Le biais de substitution peut être important au niveau *détaillé* des produits. Idéalement, il conviendrait de calculer un *indice à utilité constante* (IUC), qui mesure la variation de la dépense assurant au moindre coût le maintien du niveau d'utilité face à la variation des prix. Calculer un IUC est délicat : il est nécessaire de mettre en évidence une fonction d'utilité qui *rationalise* les données. Formellement, ce problème est résolu grâce à la théorie des *préférences révélées*. En pratique, il faut disposer de relevés très fins de prix et de quantités, ce que permettent aujourd'hui les données *scanner*.

Cette étude présente les résultats obtenus avec ce type de données pour des produits de grande consommation : les choix des consommateurs, pris dans leur ensemble, sont effectivement rationnels. Il n'y a pas un, mais toute une *plage* d'IUC, dont les valeurs extrêmes coïncident de temps à autre avec les indices de Laspeyres et de Paasche ; cette plage contient presque toujours l'indice de Fisher.

** François Magnien et Jacques Pougnard appartenaient à la division Prix à la consommation de l'Insee au moment de la rédaction de cet article.*
Les noms et dates entre parenthèses renvoient à la bibliographie en fin d'article.

Récemment, les indices de prix à la consommation (IPC) ont été au cœur d'une polémique sur une possible surestimation de l'inflation. Parti des États-Unis, le débat autour du rapport Boskin (1996) s'est étendu au Canada, au Royaume-Uni et à la France. L'enjeu est important puisque les IPC sont utilisés pour réévaluer des prestations, maintenir le pouvoir d'achat du Smic ou bien indexer les tranches d'impôts. En dehors des nouveaux produits – qui constituent, selon la Commission Boskin et les organismes statistiques nationaux, la source de biais la plus importante – le biais viendrait essentiellement de l'insuffisante prise en compte des *substitutions*, c'est-à-dire des transferts d'achats des consommateurs vers d'autres produits ou points de vente en fonction de l'évolution différenciée des prix au cours du temps.

L'indice des prix est obtenu par agrégations successives de sous-indices. À chaque étape de ce processus, l'agrégation est *laspeyrienne* : les évolutions indiciaires entre deux périodes sont pondérées proportionnellement aux valeurs de la consommation à la période initiale. Si on fait l'hypothèse que les consommateurs déplacent leurs achats vers les produits qui augmentent relativement le moins, ces indices ne tiennent pas compte de leur place croissante dans le panier du consommateur.

Or, les substitutions sont importantes au niveau *détaillé* de l'IPC, là où sont utilisés les prix relevés à partir d'un descriptif très poussé des produits. Ainsi, les prix de plusieurs dizaines de sortes de café sont relevés chaque mois pour le calcul de l'indice des prix. À ce niveau, des *micro-indices* (1) étaient calculés il y a peu à l'aide de moyennes arithmétiques de rapports de prix (2). Il y avait donc là une source de surestimation de l'inflation, un *biais de substitution*, auquel les statisticiens américains comme ceux des États membres de l'Union européenne ont remédié en remplaçant la formule de la moyenne arithmétique des rapports de prix par une moyenne géométrique.

L'indice des prix à utilité constante

La théorie microéconomique justifie ce choix. Elle propose, en effet, une solution séduisante au problème des substitutions : l'*indice à utilité constante* (IUC). Son principe est simple : il consiste à mesurer l'évolution minimale de la dépense des consommateurs entre la période de base et la période courante permettant, face à la variation différenciée des prix des produits offerts, de *maintenir* à la période courante le niveau d'utilité de la période de base.

Cette approche repose toutefois sur une hypothèse fondamentale : l'existence d'une *fonction d'utilité* (3) par rapport à laquelle les choix des consommateurs sont rationnels. Cela signifie qu'à chaque période, les quantités acquises maximisent le niveau d'utilité sous la contrainte de budget. Ce budget est bien sûr égal à ces quantités valorisées par les prix associés (cf. encadré 1). On dit qu'une telle fonction d'utilité *rationalise* les données (prix et quantités).

Lorsque la fonction d'utilité admet une forme « classique », on retrouve les indices connus. Par exemple, si la fonction d'utilité est de type *Cobb-Douglas*, la formule à employer pour obtenir un IUC est la moyenne géométrique ; si la fonction d'utilité est de type *quadratique*, c'est l'indice de Fisher. L'indice de Laspeyres, quant à lui, correspond à une fonction d'utilité à *facteurs complémentaires* pour laquelle il n'y a pas de substitutions (cf. encadré 2). Ces fonctions d'utilité ont la propriété d'être *homothétiques* : si deux paniers sont équivalents, alors ils le restent tant que les quantités *relatives* des produits restent inchangées dans les deux paniers.

L'IUC constitue la référence pour les statisticiens d'indices de prix (4). Ils cherchent à s'en approcher au mieux, compte tenu des moyens de collecte dont ils disposent. C'est pourquoi, en quelques années, la quasi-totalité des pays ont abandonné, dans le calcul des micro-indices, la formule de Laspeyres pour la moyenne géométrique. La première suppose que les agents n'ont, rationnellement, aucune raison de substituer les produits entre eux alors que la moyenne géométrique renvoie à la fonction d'utilité de Cobb-Douglas, donc à l'hypothèse d'une bonne substituabilité des produits et des points de vente (5).

1. *Un micro-indice est calculé à partir de relevés de prix relatifs à une classe étroite de produits, la* variété*, réalisés dans des points de ventes localisés dans la même agglomération. Il y a environ un millier de variétés (réparties dans les 304 postes de l'IPC) et une centaine d'agglomérations. Cependant, toutes les variétés n'étant pas suivies dans toutes les agglomérations, il n'y a « que » 25 000 micro-indices environ (Cf.* Pour comprendre l'indice des Prix, *1998).*
2. *Estimateurs des indices de Laspeyres avec des probabilités d'inclusion proportionnelles aux valeurs des ventes de la période de base.*
3. *Dans cette étude, les fonctions d'utilité sont implicitement continues, concaves et non saturées.*
4. *Ainsi que pour la commission Boskin.*
5. *L'élasticité de substitution d'une fonction d'utilité de Cobb-Douglas est égale à un en tout point.*

Le problème de la rationalité des consommateurs

L'hypothèse d'une rationalité des choix des consommateurs peut sembler très forte. En outre, si elle est satisfaite, comment s'en assurer ?

Dans le cas d'une seule période 0, l'existence d'une fonction d'utilité homothétique rationalisante est triviale : il suffit de la choisir *linéaire* : $U(q) = p_0 q$ (p_0 désigne le vecteur des prix, q_0 celui des quantités). Le schéma 1-A représente la courbe d'indifférence de cette fonction d'utilité passant par q_0. Avec deux périodes, le problème est déjà plus compliqué. Cette fois, la fonction d'utilité n'est plus linéaire que par morceaux : $U(q) = Min\{\alpha_0 p_0 q, \alpha_1 p_1 q\}$ où α_0 et α_1 sont deux nombres positifs. Le schéma 1-B illustre cette construction : il y a autant de façons de choisir le rapport α_1/α_0 que de façons de choisir le rayon OX dans le cône limité par les vecteurs q_0 et q_1. La situation est déjà beaucoup plus complexe que dans le cas d'une seule période : la construction précédente n'est pas toujours possible, comme le montre le schéma 1-C. Le cas de trois périodes est illustré par le schéma 1-D. Avec un nombre quelconque de périodes et de produits, l'existence et la construction d'une fonction d'utilité homothétique rationalisante linéaire par morceaux devient très vite un problème redoutable.

Encadré 1

L'IUC ASSOCIÉ À UNE FONCTION D'UTILITÉ

Pour un ensemble fini E de périodes t on dispose des relevés de prix p_t et de quantités vendues q_t pour un ensemble de produits et un ensemble de points de vente. Les vecteurs p_t et q_t ont autant de composantes qu'il y a de couples « produit-point de vente ».

On suppose qu'il existe une fonction d'utilité U qui *rationalise* les données $(p_t, q_t)_{t \in E}$:

$$U(q_t) = Max\{U(q), p_t q \le p_t q_t\} \quad \text{pour tout } t \in E$$

Autrement dit, à chaque période $t \in E$, le panier q_t doit être optimal (pour U) sous la contrainte budgétaire $p_t q_t$. Dualement (1) :

$$p_t q_t = Min\{p_t q, U(q) \ge U(q_t)\} \quad \text{pour tout } t \in E$$

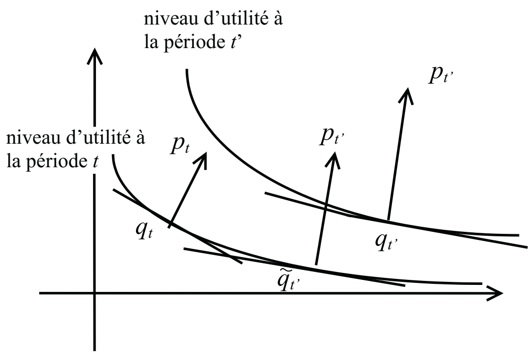
L'IUC à une date t' par rapport à une date t, pour un niveau d'utilité donné u, est le rapport de deux dépenses : la dépense minimale qui permet d'atteindre le niveau d'utilité u à la date t' (respectivement t) face aux prix $p_{t'}$ (respectivement p_t). Pour exprimer mathématiquement l'IUC, il convient donc d'introduire la *dépense minimale* $C_U(u, p)$ permettant d'atteindre un niveau d'utilité donné u face à un système de prix p exogène :

$$C_U(u,p) = Min\{pq, U(q) \ge u\}$$

L'IUC entre deux dates t et t' est alors défini par :

$$IUC_{t'/t}(U,u) = C_U(u, p_{t'}) / C_U(u, p_t)$$

L'IUC présente l'inconvénient de dépendre du niveau d'utilité de référence u. Il est assez naturel de choisir le niveau d'utilité initial $U(q_t)$ comme référence. Cette situation est représentée sur la figure suivante :



Le panier $\tilde{q}_{t'}$ assure au moindre coût face au système de prix $p_{t'}$ le niveau d'utilité initial $U(q_t)$. Alors :

$$\text{IUC}_{t'/t} = \frac{p_{t'} \tilde{q}_{t'}}{p_t q_t} .$$

En général, avec un autre niveau d'utilité de référence, par exemple le niveau final $U(q_{t'})$, l'IUC obtenu est différent. Ce problème disparaît lorsque la fonction d'utilité U est *homothétique,* c'est-à-dire :

$U(q) = U(q') \Rightarrow U(\lambda q) = U(\lambda q')$ pour tout $\lambda > 0$ et tous les paniers q, q'.

Ceci est immédiat dans le cas particulier où U est *homogène* (de degré 1) :

$U(\lambda q) = \lambda U(q)$ pour tout $\lambda > 0$ et tout panier q

car alors $C_U(u, p) = u C_U(1, p)$. Ainsi :

$$IUC_{t'/t}(U,u) = C_U(1, p_{t'}) / C_U(1, p_t)$$

1. Les fonctions considérées dans cette étude sont continues et non saturées.

La théorie micro-économique des *préférences révélées* permet de résoudre ce problème. Développée par Afriat (1967, 1981), elle a également bénéficié des travaux de Diewert (1973) et Varian (1982, 1983). Ils ont dégagé une condition assurant l'existence d'une fonction d'utilité rationalisant des données en prix et en quantités pour un ensemble *quelconque* de périodes. Cette condition, dite HARP (6), est assez simple, du moins dans sa formulation :

Il existe une fonction d'utilité homothétique *rationalisant des données* $(p_t, q_t)_{t \in E}$ *de prix et de quantités sur un ensemble E de périodes si et seulement si* :

$$\frac{p_i q_j}{p_i q_i}\,\frac{p_j q_k}{p_j q_j}\cdots\frac{p_m q_i}{p_m q_m} \geq 1$$

pour tout « cycle » de périodes $i, j, \ldots, m,\ i \in E$.

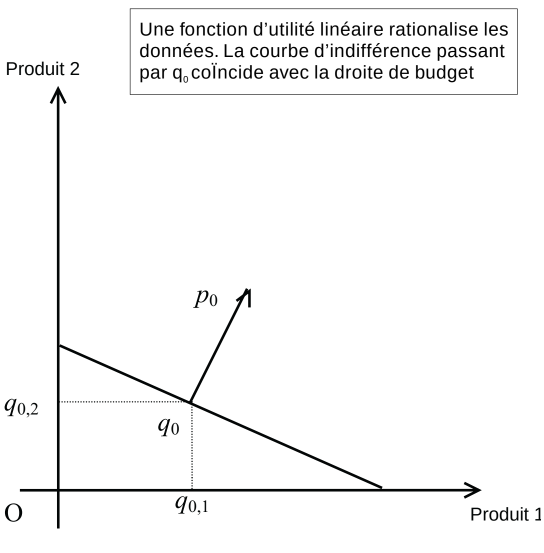
Les composantes du vecteur p_t sont les prix des produits à la période *t*, celles du vecteur q_t les quantités (7). Le cycle de périodes *i, j, k, ... , m, i* n'est pas ordonné chronologiquement, sa longueur est quelconque.

Le membre de gauche de l'inégalité est un indice de *quantités* de type *Laspeyres chaîné*. Il s'agit de

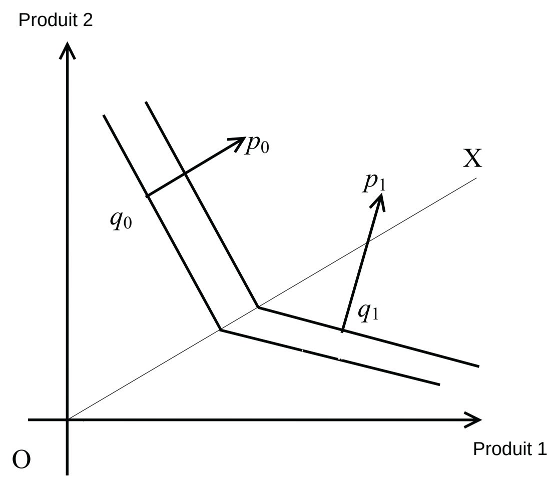
6. *Homothetic Axiom of Revealed Preference.*
7. $p_t q_t$ *désigne le produit scalaire des deux vecteurs.*

Schéma 1
Construction d'une fonction d'utilité linéaire par morceaux rationalisant des données en prix et quantités (cas de deux produits)

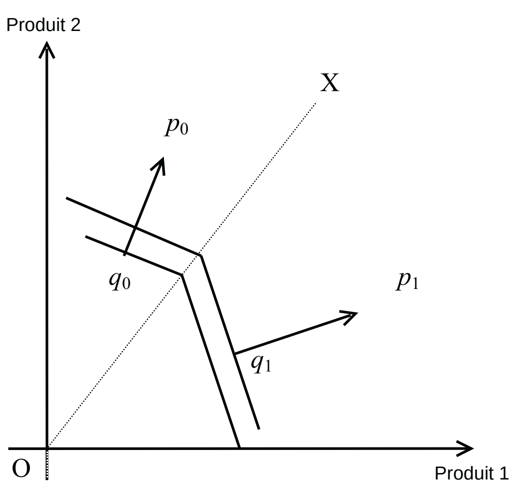
A - Une période



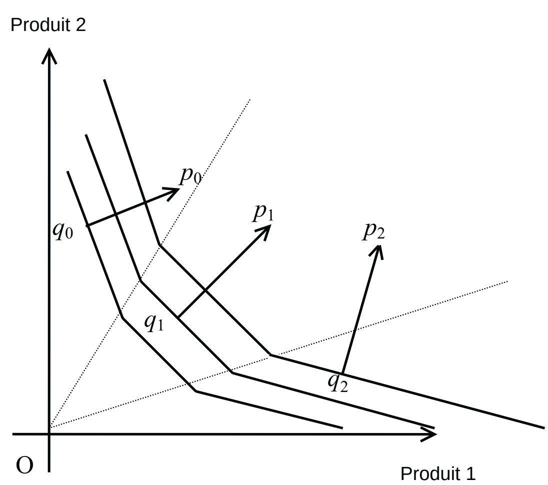
B - Deux périodes (données rationalisables)



C - Deux périodes (données non rationalisables)



D - Trois périodes (données rationalisables)



l'expression classique d'un indice de volume (8). Comme les indices de prix de ce type, cet indice présente l'inconvénient de ne pas revenir à 1 lorsque les quantités retrouvent leur valeur initiale (« défaut de circularité »). La condition HARP limite ce défaut de circularité : l'indice de Laspeyres chaîné vaut *au moins* 1. De façon équivalente, elle exprime que l'indice de volume de *Paasche chaîné* vaut *au plus* 1 lorsque les quantités reviennent à leur valeur initiale. Ainsi, la condition HARP s'exprime agréablement :

Si les quantités retrouvent leur valeur initiale, la fourchette des indices de volume « Paasche chaîné » et « Laspeyres chaîné » contient la valeur 1.

Les fonctions d'utilité rationalisantes dont la condition HARP révèle l'existence sont les fonctions homothétiques. Elles offrent un avantage décisif : l'IUC associé ne dépend pas du niveau d'utilité de référence (cf. encadré 1).

La richesse des données scanner

La vérification de la condition HARP nécessite de connaître les *quantités* en plus des prix. Pour le calcul de l'IPC, on ne dispose de cette information qu'à un niveau agrégé. Encore, ces données ne sont-elles disponibles qu'annuellement. C'est sur de telles données que les travaux empiriques ont été menés jusqu'à présent, notamment par Manser et MacDonald (1988) qui ont utilisé des données agrégées sur la période 1959-1985, aux États-Unis.

Ces travaux sont cependant d'un intérêt limité puisque les substitutions ont lieu essentiellement au niveau détaillé. En effet, en France, le biais de substitution est pratiquement nul au niveau *agrégé* (Lequiller, 1997), c'est-à-dire dans le calcul de l'indice d'ensemble à partir des indices des *postes*. À ce niveau, les pondérations sont revues chaque année, lors du chaînage de l'indice. Le biais de substitution est modéré au niveau *intermédiaire*, lors du calcul des indices des postes par agrégation des micro-indices obtenus au niveau détaillé.

Il est donc judicieux de calculer des IUC au niveau *détaillé*, celui des produits réellement observés, où sur des périodes très courtes à l'intérieur d'une année, de fortes évolutions relatives de prix entre ces produits, au gré des stratégies des marques et des circuits commerciaux, induisent les substitutions les plus importantes. Mais cela nécessite des données

8. Tel qu'en calcule par exemple la comptabilité nationale française.

Encadré 2

QUELQUES EXEMPLES FONDAMENTAUX

Un exemple fondamental de fonction d'utilité homogène est la fonction de *Cobb-Douglas* :

$$U(q) = a\prod_s (q^s)^{\alpha^s} \text{ Avec } a > 0, \alpha^s > 0 \quad \text{et} \quad \sum_s \alpha^s = 1$$

où s désigne un produit dans un point de vente. Lorsqu'une telle fonction d'utilité rationalise les données, l'IUC associé, à une période t par rapport à une période t', est la *moyenne géométrique pondérée* :

$$\prod_s \left(\frac{p^s_{t'}}{p^s_t}\right)^{w^s_t} \quad \text{où} \quad w^s_t = \frac{q^s_t\, p^s_t}{\sum_{s'} q^{s'}_t\, p^s_t}$$

La fonction de *Cobb-Douglas* possède une propriété remarquable : les parts dans la dépense de consommation des différents produits restent constantes :

$w^s_t = \alpha^s$ pour tout s et toute période t

Un autre exemple important est la fonction d'utilité *quadratique* :

$$U(q) = \sqrt{\sum_{s,\,s'} a_{s,\,s'}\, q^s q^{s'}}$$

où $(a_{s,\,s'})$ est une matrice symétrique (dont les coefficients sont convenablement choisis). Si une fonction de ce type rationalise les données, l'IUC correspondant est l'indice de *Fisher* :

$$\sqrt{L_{t'/t}\, P_{t'/t}}$$

Mais l'exemple le plus simple est encore la fonction d'utilité à *facteurs complémentaires* :

$$U(q) = \min_s \left\{\frac{q^s}{a_s}\right\} \quad \text{où} \quad a^s > 0$$

L'indice à utilité constante associé n'est autre que l'indice de *Laspeyres* mais aussi de *Paasche*, qui coïncident alors.

en prix et en quantités au niveau le plus fin, à un rythme mensuel. Seules les *données scanner* (9) le permettent. Elles sont recueillies par des sociétés d'études de marché sur les produits vendus dans les grandes surfaces, grâce à l'exploitation du *code-barre*. Ces sociétés réalisent des études utilisées par les industriels et les distributeurs pour mettre au point de nouveaux produits, mesurer l'impact des campagnes publicitaires, etc.

Les données utilisées pour cette étude ont été fournies à l'Insee par la société AC Nielsen. Elles sont relatives à trois types de produits, huiles alimentaires, lessives et café, vendus dans quatre cents supermarchés et hypermarchés sur l'ensemble de la métropole. Elles recouvrent une période de trois années : 1994, 1995 et 1996.

Ces relevés sont accompagnés d'une description très poussée des produits. Ainsi, on disposait avant apurement de 523 produits élémentaires distincts pour les huiles, 383 pour les lessives et 1168 pour les cafés ! L'avantage est que l'on peut ainsi prendre pleinement en compte les substitutions entre produits. En contrepartie, ces produits élémentaires croisés avec les points de vente donnent lieux à un nombre très élevé de relevés potentiels dont une proportion importante est fréquemment inobservable.

Ce problème est résolu en agrégeant les points de vente selon quatre *formes de vente* : petits ou grands supermarchés et petits ou grands hypermarchés. Pour chacune de ces quatre formes de vente, on retient alors les produits élémentaires qui ont été vendus chaque mois de la période d'étude (de janvier 1994 à décembre 1996) dans l'un au moins des points de vente. On élimine ainsi les produits nouveaux et ceux qui sont retirés temporairement ou définitivement de l'une des quatre formes de vente. Ont ainsi été retenus 138 produits élémentaires pour les huiles, 133 pour les lessives et 353 pour les cafés, représentant respectivement 91 %, 75 % et 92 % du chiffre d'affaires avant apurement. Finalement, on obtient pour l'ensemble des trente-six mois, 333 séries de prix et quantités pour les huiles, 424 séries pour les lessives et 895 pour les cafés (10).

Les consommateurs sont globalement rationnels

Pratiquement, la vérification de la condition HARP sur ces données est très lourde dès que le nombre de mois d'observations dépasse quelques unités. Varian (1982, 1983) a proposé un algorithme qui permet d'y parvenir (cf. encadré 3).

Le résultat obtenu sur les données scannées utilisées est important : les prix observés et les quantités échangées sur le marché sont effectivement rationalisables par une fonction d'utilité *homothétique*. Pour le café, il a même été possible de rationaliser la totalité des données (mois 1 à 36). Pour les huiles et les lessives, seules les sous-périodes 1-26 et 1-27 ont pu être rationalisées (11).

9. *Ou encore données* scannées, *données* scannographiques *ou* micro-données.
10. *On trouvera une description détaillée des données scanner dans Magnien et Pougnard (1998).*
11. *En imposant le mois 1, janvier 1994, parmi les périodes rationalisées.*

Encadré 3

UN ALGORITHME DE CALCUL

L'algorithme de Warshall (1962) associe à une matrice carrée $M = (m_{ij})$ *la matrice carrée* $D = (d_{ij})$ *définie par :*

$$d_{ij} = inf_{i,k,l,\ldots,m,j}\,(m_{ik} + m_{kl} + \ldots + m_{mj})$$

Autrement dit, si m_{ij} représente le « coût » de passage *direct* de i en j alors d_{ij} représente le coût *minimal* de passage de i en j sur l'ensemble des « chemins » allant de i à j. On montre que la matrice D s'obtient comme suit :

(1) faire $k = 1$;

(2) pour tous i, j : si $m_{ij} \geq m_{ik} + m_{kj}$ faire $m_{ij} = m_{ik} + m_{kj}$;

(3) si $k < T$ faire $k = k + 1$ et aller en (2) ; sinon faire $d_{ij} = m_{ij}$ pour tous i, j.

L'algorithme de Warshall permet donc de calculer très facilement les quantités :

$$inf_{i,k,l,\ldots,m,j}\ \ln\left\{\frac{p_i q_k}{p_i q_i}\,\frac{p_k q_l}{p_k q_k}\ldots\frac{p_m q_j}{p_m q_m}\right\}$$

en posant $m_{kl} = \ln \dfrac{p_k q_l}{p_k q_k}$ puis, par application de la fonction exponentielle (croissante, comme la fonction logarithme), les quantités :

$$inf_{i,k,l,\ldots,m,j}\left\{\frac{p_i q_k}{p_i q_i}\,\frac{p_k q_l}{p_k q_k}\ldots\frac{p_m q_j}{p_m q_m}\right\}$$

Les bornes $1/\Delta_{ji}$ et Δ_{ij} de la fourchette des IUC s'en déduisent immédiatement.

L'algorithme de Warshall permet de savoir si la condition HARP est vérifiée : si, lors d'une étape k, on a $m_{ii} < 0$ pour un i, alors la condition HARP n'est pas satisfaite. Sinon elle l'est.

L’indice à utilité constante n’est donc pas seulement un concept théorique : son *existence* est établie sur des produits de grande consommation comme le café, les huiles alimentaires et les lessives.

La détermination des IUC

Sachant qu’il existe des fonctions d’utilité homothétique rationalisant les données scannées, il reste à déterminer les IUC associés à chacune d’elles entre deux dates quelconques. On sait que cet ensemble est majoré, en général strictement, par l’indice de Laspeyres et minoré par l’indice de Paasche. Manser et MacDonald (1988) ont déterminé, lorsque la condition HARP est satisfaite, les bornes inférieure et extérieure de la plage des IUC d’une période *i* par rapport à une période *j* : elles valent respectivement $1/\Delta_{ji}$ et Δ_{ij} où Δ_{ij} est égal à :

$$\frac{p_i q_i}{p_j q_j} Min_{i,k,.l..,m,j}\left\{\frac{p_i q_k}{p_i q_i}\,\frac{p_k q_l}{p_k q_k}\cdots\frac{p_m q_j}{p_m q_m}\right\}$$

pour toute séquence de périodes *i, k, l, ... , m, j* $\in$ *E*. On montre (Magnien et Pougnard, 2000) que toute valeur comprise entre ces bornes est un IUC. Ces IUC sont relatifs aux fonctions d’utilité *homothétiques* qui rationalisent les données.

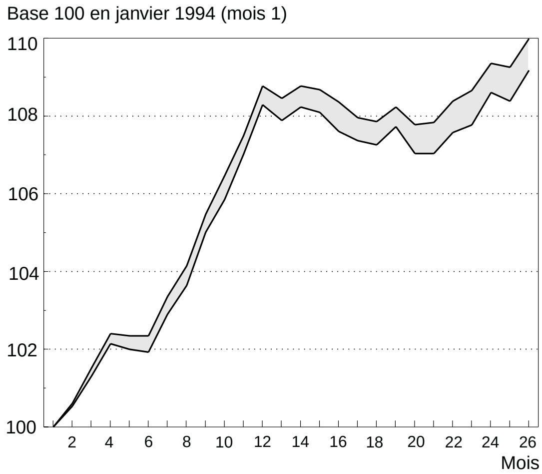
Les résultats du calcul d’IUC obtenus sur les données utilisées apparaissent sur les graphiques I-A, I-B et I-C. Pour chacun des trois produits, la plage des IUC d’un mois donné *par rapport au mois de janvier 1994* y est représentée. Ainsi, pour les huiles alimentaires, les IUC du mois d’août 1995 (mois 20) prennent, en base 100 en janvier 1994 (mois 1), toutes les valeurs comprises entre 107,0 et 107,8. Cela signifie que pour n’importe quel chiffre compris entre 107,0 et 107,8 il existe une fonction d’utilité homothétique qui rationnalise les données de prix et de quantités observées et telle que ce chiffre soit l’IUC en août 1995, base 100 janvier 1994.

Ces IUC sont associés aux différentes fonctions d’utilité homothétiques qui « rationalisent » les données relatives à *un ensemble de périodes le plus grand possible* : il ne s’agit donc pas des fonctions d’utilité qui rationalisent les seules données s’échelonnant du mois de base au mois sous revue ; il s’agit encore moins des fonctions d’utilités ne rationalisant que les données des seuls mois de base et sous revue.

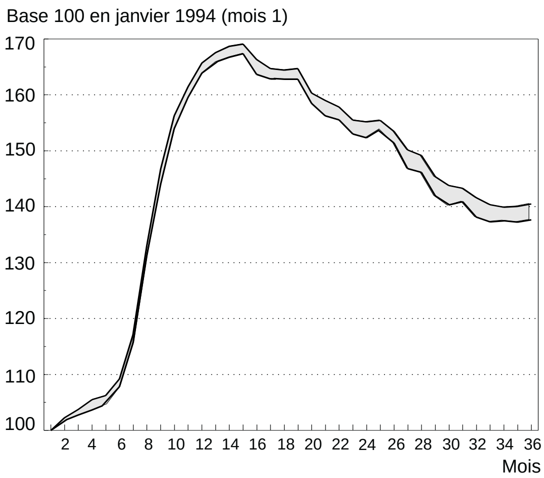
La largeur de la plage fluctue : elle peut se réduire fortement et même s’annuler (mois 11, 12 et 13 pour les lessives) : il y a alors unicité de l’IUC.

Graphique I
La plage des IUC

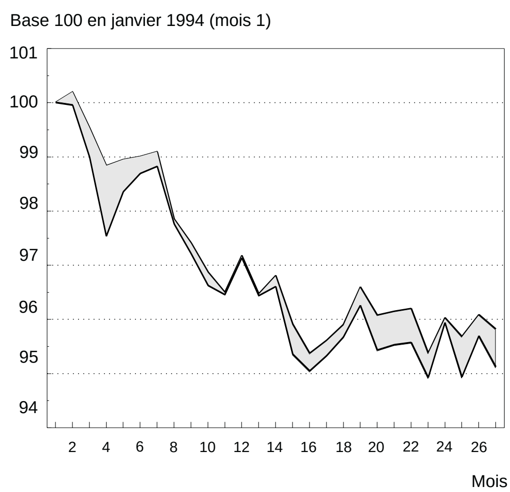
A - Pour les huiles alimentaires



B - Pour le café



C - Pour les lessives



Sources : AC Nielsen, calculs des auteurs.

Les indices de Laspeyres et de Paasche sont de temps à autre des IUC

Lorsque la fonction d'utilité est homothétique, l'IUC est inférieur à l'indice de Laspeyres (12) et supérieur à l'indice de Paasche. Puisqu'il y a toute une plage d'IUC, le biais de substitution résultant de l'usage d'un indice de Laspeyres ne peut être défini de façon unique : seul un encadrement peut en être donné, mesuré par l'écart avec le plus grand et le plus petit des IUC. Le même problème se pose avec l'indice de Paasche.

Le biais de substitution *minimum* des indices de Paasche et de Laspeyres, mesuré par référence au « bas » ou au « haut » de la plage des IUC, évoluent assez différemment l'un par rapport à l'autre mais aussi dans le temps. Ils peuvent *s'annuler* à des périodes différentes ou simultanément. C'est ce que l'on observe avec chacun des produits (cf. graphiques II-A, II-B et II-C).

Diewert (1990) a expliqué comment la *faiblesse* du biais de substitution de l'indice de Laspeyres (Paasche) par rapport à l'IUC résulte d'une relation appropriée entre la variation *relative* des prix d'une part et la *substituabilité* des produits ou des points de vente à la période de base (courante) d'autre part. La *nullité* de ce biais, observée sur les données utilisées ici, correspond à un cas *extrême* de cette relation. Elle résulte de la forme particulière de certaines des fonctions d'utilité homothétiques rationalisantes. On montre (13), en effet, que les fonctions d'utilité *linéaires par morceaux*, dont on a précédemment ébauché la construction (cf. schéma 1), sont « représentatives » (14) de toutes les autres fonctions d'utilité homothétiques rationalisantes. Ces fonctions linéaires par morceaux ont l'expression suivante (15) :

$$U(q) = Min\left\{\alpha_t p_t q,\ t \in E\right\} \quad (\alpha_t > 0)$$

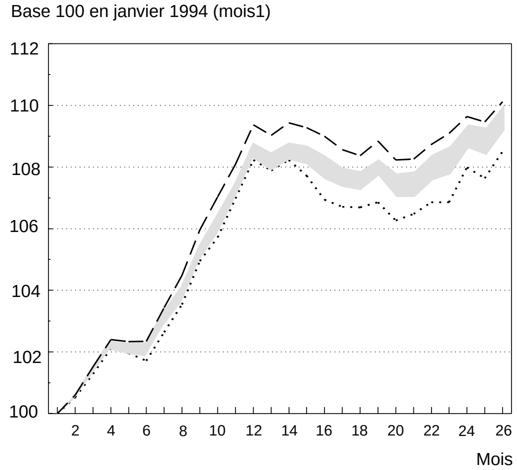
Le schéma 2 représente une courbe d'indifférence propre à ce type de fonction d'utilité dans le cas de deux produits. Ces courbes ont la particularité de posséder des *points anguleux*.

Considérons alors un mois de base *t* correspondant à un tel point anguleux (cf. schéma 3-A). En q_t la substituabilité entre les produits est faible (16).

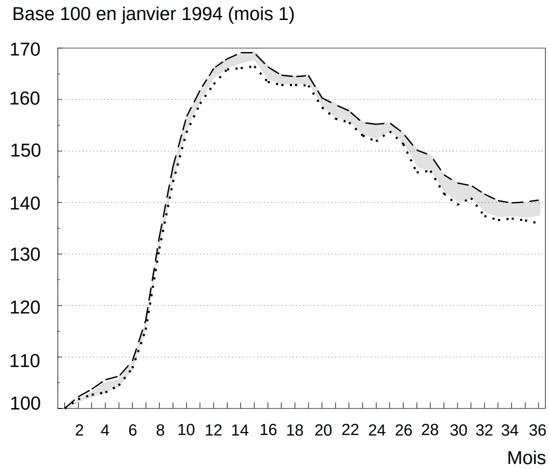
12. *Il suffit, pour s'en assurer, de revenir à la figure de l'encadré 1 : par optimalité du panier* $\tilde{q}_t$*, face au système de prix* p_t*, on a* $p_t \cdot \tilde{q}_t \leq p_t \cdot q_t$.
13. *Cf. Magnien et Pougnard (2000).*
14. *L'IUC calculé avec une fonction d'utilité homothétique rationalisante quelconque est égal à l'IUC calculé avec l'une de ces fonctions.*
15. *Qui généralise la construction ébauchée dans le cas d'une, deux et trois périodes sur le schéma 1.*
16. *L'élasticité de substitution est nulle.*

Graphique II
Plage des IUC et indices de Laspeyres et de Paasche

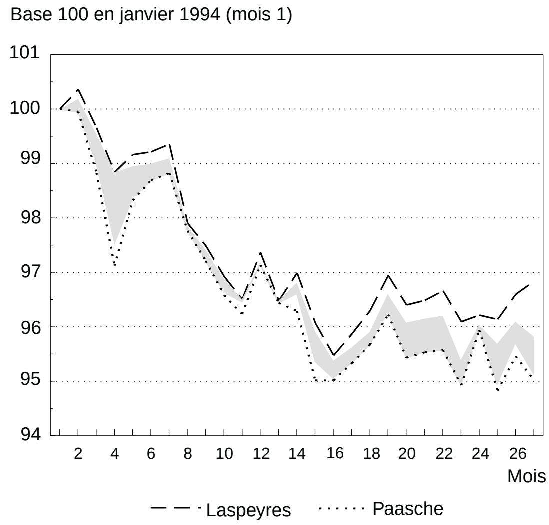
A - Pour les huiles alimentaires



B - Pour le café



C - Pour les lessives



Sources : AC Nielsen, calculs des auteurs.

Schéma 2
Courbes d'indifférence d'une fonction d'utilité homothétique, linéaire par morceaux

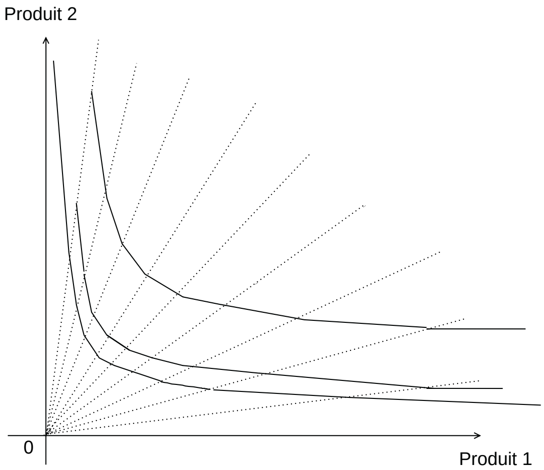
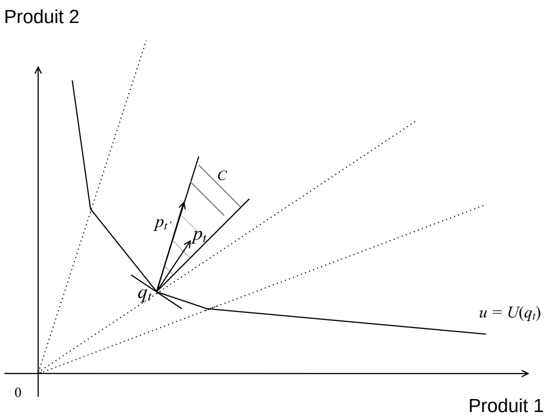
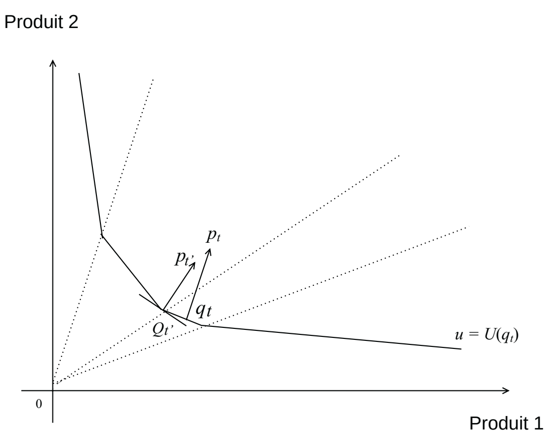


Schéma 3
Substituabilité ou non en un point anguleux d'une courbe d'indifférence

A - L'indice de Laspeyres ne présente pas de biais de substitution



B - L'indice de Laspeyres présente un biais de substitution



Tant que la variation *relative* des prix de ces produits n'est pas trop forte entre le mois de base *t* et le mois sous revue *t'*, le panier de base q_t assure au moindre coût un niveau d'utilité inchangé. L'IUC coïncide alors avec l'indice de Laspeyres. L'ensemble des vecteurs de prix $p_{t'}$ pour lesquels il en est ainsi constitue un *cône C* représenté sur le schéma 3-A (17).

Ainsi, pour les huiles, pendant les six premiers mois de 1994, l'indice de Laspeyres est un IUC : la hausse des prix est plutôt uniforme (18) (cf. graphique II-A). Ensuite, la différentiation des évolutions de prix est suffisamment importante (le vecteur des prix sort du cône *C*) pour que l'indice de Laspeyres surestime l'inflation. Au contraire, avec le café, l'évolution des prix des produits reste essentiellement uniforme : le vecteur des prix reste presque toujours dans le cône *C*, si bien que l'indice de Laspeyres est presque toujours un IUC (cf. graphique II-B).

Si, au contraire, la fonction d'utilité est *linéaire* au voisinage de q_t alors, aussi faible que soit la variation *relative* des prix, la composition du panier ($Q_{t'}$ sur le schéma 3-B) qui assure au moindre coût le même niveau d'utilité sera distincte de celle de q_t : l'IUC sera distinct de l'indice de Laspeyres. Cette linéarité, situation dans laquelle la substituabilité des produits est « parfaite (19) », n'est donc satisfaite ni pour les huiles, ni pour les lessives, ni pour les cafés en janvier 1994 (mois 1) : avec chacun d'eux, il y a plusieurs mois pour lesquels l'indice de Laspeyres est un IUC (20).

Les circonstances dans lesquelles l'indice de Paasche est un IUC sont analogues : le cône *C* est associé au panier de la période courante et il y a coïncidence des deux indices lorsque le vecteur des prix de base appartient à ce cône. La conjonction des deux cas précédents (existence d'une fonction d'utilité homothétique pour laquelle l'IUC est égal au Paasche et d'une autre pour laquelle l'IUC est égal au Laspeyres) se produit avec le café à de très nombreuses reprises (cf. graphique II-B).

---

17. Ce cône est l'ensemble des vecteurs p *tels que* $pq_t = C_U(U(q_t), p)$ *où* $C_U(u, p)$ *désigne la dépense minimale permettant de disposer du niveau d'utilité* u *face au vecteur de prix* p*.*
18. Le mois de base t *est le mois 1 c'est-à-dire janvier 1994.*
19. L'élasticité de substitution est infinie.
20. On imagine difficilement que les prix puissent évoluer de façon parfaitement uniforme.

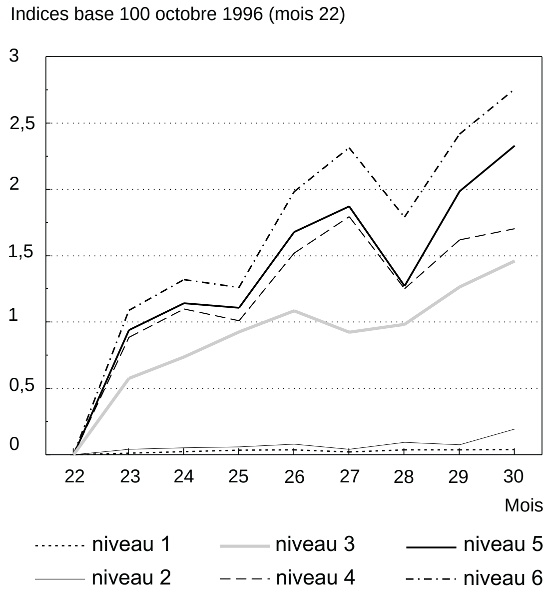
Une bonne désagrégation des données est nécessaire

Il peut sembler surprenant que la condition HARP soit satisfaite sur des périodes aussi longues (les 36 mois de suivi du panel pour le café, 27 mois pour les lessives et 26 pour les huiles), compte tenu du caractère apparemment restrictif de la condition d'homothéticité. Tout d'abord, la plage des IUC est étroite : moins d'un point d'indice environ pour les huiles et les lessives, au plus trois points pour les cafés dont les prix ont enregistré de très fortes variations.

Ensuite, Manser et MacDonald (1988) ont avancé l'idée que l'existence de fonctions homothétiques « rationalisantes » résulterait d'une *désagrégation* suffisante des données. On entend par là que les prix et les quantités sont relatifs à des produits dont la description fait appel à un nombre conséquent de caractéristiques. La grande finesse des données scannées serait donc la raison des bons résultats du test de la condition HARP. Pour le vérifier, on a testé cette condition à différents niveaux d'agrégation des données relatives au café et aux lessives. Ces niveaux sont déterminés par la liste des caractéristiques retenues pour la définition des produits et points de vente. Outre la forme de vente (il y en a quatre), ces caractéristiques pour les produits sont les suivantes : le fabricant (ou distributeur), la marque, la référence, le conditionnement, le contenu. Chacune d'elles prend diverses modalités : pour le café, le conditionnement est ainsi défini par le type d'emballage (bocal, boîte, carton, etc.), le nombre de paquets vendus ensemble et le poids de l'ensemble. Toujours pour le café, la description du contenu est encore plus variée : le type (grain, moulu ou expresso), la qualité (normal ou décaféiné), la gamme (arabica, robusta ou mélange) et l'origine. Ces différents contenus ont été regroupés en quelques variétés (cf. note (1) et Magnien et Pougnard, 1998).

Le tableau indique les différents niveaux d'agrégation retenus. Il indique également le nombre de produits distincts par niveau et la période la plus longue sur laquelle la condition HARP est satisfaite, c'est-à-dire pour laquelle les données sont optimales pour une fonction d'utilité homothétique. Cette période est effectivement d'autant plus longue que les données sont désagrégées. En outre, pour un ensemble donné de périodes, la plage des IUC est d'autant plus large que la désagrégation des données est importante (cf. graphique III) : une plus grande désagrégation rend d'autant plus aisée la rationalisation des données.

Graphique III
Largeur de la plage des IUC selon l'agrégation des données



Lecture : cf. la signification des niveaux *dans le tableau.*
Sources : AC Nielsen, calculs des auteurs.

Tableau
Niveaux d'agrégation et nombre de périodes rationalisées

Niveau d'agrégation des données	Caractéristiques retenues	Café		Lessives	
		Nombre de séries*	Période rationalisée	Nombre de séries*	Période rationalisée
1	Variété	5	22-30	12	13-15
2	Variété et forme de vente (FV)	20	9-32	48	12-16
3	Variété, FV et fabricant	264	8-36	216	1-20
4	Variété, FV, fabricant et marque	329	8-36	320	1-23
5	Variété, FV, fabricant, marque, référence	562	1-36	-	-
6	Toutes	895	1-36	424	1-27

* Croisements des différentes modalités des caractéristiques retenues.

Sources : AC Nielsen, calculs des auteurs.

Avec les lessives, au niveau le plus agrégé (niveau 1), seuls trois mois consécutifs ont pu être rationalisés : le test de la condition HARP a même fréquemment échoué sur deux mois. On observe alors systématiquement que l'indice de Paasche (l'un de ces deux mois par rapport à l'autre) est supérieur à l'indice de Laspeyres (21).

Indices de Paasche et de Laspeyres, et homothéticité

L'homothéticité des fonctions d'utilité, sans laquelle le calcul d'IUC devient problématique puisqu'il dépend alors du choix d'un niveau d'utilité de référence, peut réduire l'ensemble des périodes sur lequel des données sont rationalisées. Ainsi, les données sur les huiles ne sont « homothétiquement » rationalisables que sur 26 mois. Le problème peut d'ailleurs se poser avec deux périodes seulement. Il est alors intimement lié à la relation entre les indices de Paasche et de Laspeyres.

Pour que l'indice de Paasche soit inférieur ou égal à l'indice de Laspeyres, il *suffit* qu'il existe une fonction d'utilité *homothétique* rationalisant les données relatives à *deux* périodes : la période de base et la période courante (mais pas nécessairement les autres). En effet, un IUC s'intercale alors entre les indices de Paasche et de Laspeyres. Réciproquement, si l'indice de Paasche est inférieur ou égal à l'indice de Laspeyres :

$$\frac{p_j q_j}{p_i q_j} \leq \frac{p_j q_i}{p_i q_i}$$

alors :

$$\frac{p_i q_j}{p_i q_i} \frac{p_j q_i}{p_j q_j} \geq 1$$

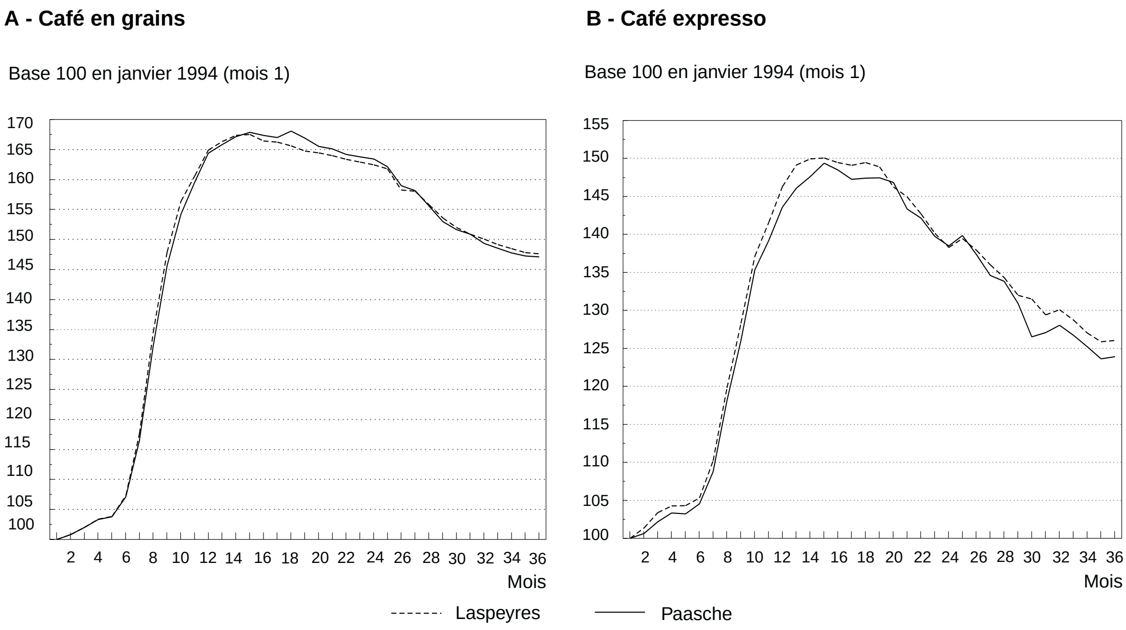
inégalité qui n'est autre que la condition HARP pour l'ensemble réduit aux *deux* périodes *i* et *j*. Ainsi :

L'indice de Paasche de la période j par rapport à la période i est inférieur ou égal à l'indice de Laspeyres si et seulement s'il existe une fonction d'utilité homothétique *qui rationalise les données* (p_i, q_i) *et* (p_j, q_j).

Qu'il s'agisse du café, des huiles ou des lessives, cette inégalité est satisfaite pour tout couple *i*, *j* de mois (22). Cependant, dans le cas du café, si l'on se limite aux « cafés en grains » ou aux « cafés expresso », on observe à plusieurs reprises le phénomène d'inversion des indices de Paasche et de Laspeyres (cf. graphiques IV-A et IV-B).

21. Ces indices étant calculés sur les données agrégées.
22. Pour les huiles et les lessives, cela ne coule pas de source : on ne rationalise pas les données sur 36 mois.

Graphique IV
L'Indice de Paasche peut être supérieur à l'indice de Laspeyres



Source : AC Nielsen, calculs des auteurs.

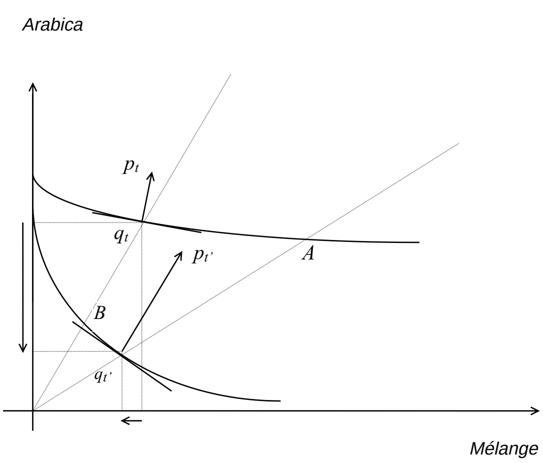
Agrégeons, pour les cafés en grains, les relevés relatifs à l'arabica d'une part (les produits de qualité supérieure, donc de prix plus élevé) et les relevés relatifs aux mélanges d'autre part. Alors, entre le mois de janvier 1994 (mois t=1) et le mois de juin 1995 (mois t'=18), où l'inversion des indices est la plus marquée, c'est le prix des cafés mélangés qui a le plus augmenté (74 % contre 48 %), mais ce sont les quantités vendues d'arabica qui ont le plus reculé (- 40 % contre - 21 %). Le schéma 4 représente cette situation. Une fonction d'utilité rationalise les données relatives aux dates t et t' mais elle ne saurait être homothétique : la configuration des prix et des quantités est telle que la tangente en A à la courbe d'indifférence initiale ne peut être parallèle à la tangente à la courbe d'indifférence en $q_{t'}$ (de même que la tangente en B à la courbe d'indifférence courante ne peut être parallèle à la tangente à la courbe d'indifférence en q_t). C'est en contradiction avec la propriété de « parallélisme » des tangentes aux courbes d'indifférence le long d'un rayon issu de l'origine, caractéristique des fonctions d'utilité homothétiques.

L'indice de Fisher est une meilleure approximation de l'IUC que la moyenne géométrique

Il est possible de juger de la qualité des indices de prix usuellement calculés en fonction de leur degré de proximité avec l'IUC.

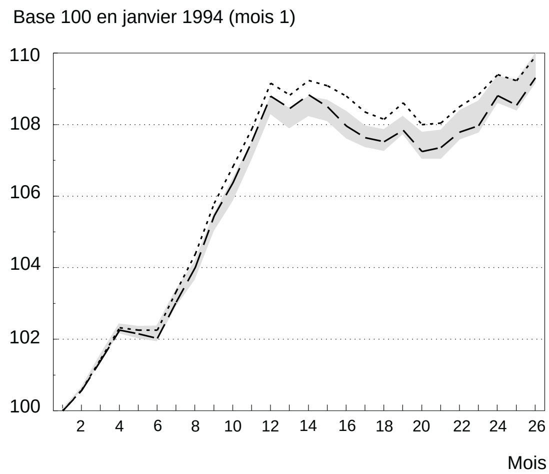
La réputation de l'indice de Fisher comme très bonne approximation de l'indice à utilité constante est confirmée puisqu'il reste pratiquement toujours

Schéma 4
L'indice de Paasche peut être supérieur à l'indice de Laspeyres (exemple du café)

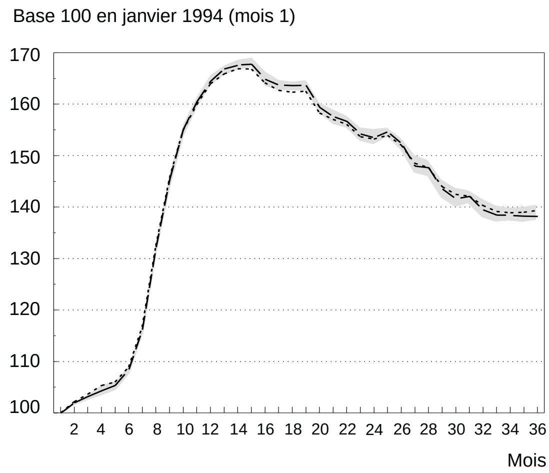


Graphique V
Indices de Fisher, géométrique et IUC

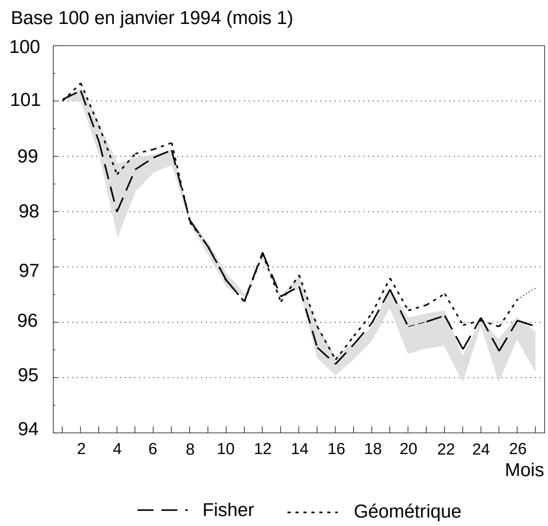
A - Pour les huiles alimentaires



B - Pour le café



C - Pour les lessives



Sources : AC Nielsen, calculs des auteurs.

dans la plage des IUC (23) (cf. graphiques V-A, V-B et V-C). Si l'on considère une valeur de l'indice de Fisher intérieure à cette plage, alors il existe une fonction d'utilité homothétique rationalisant les données et par rapport à laquelle cet indice de Fisher est un IUC. Ceci ne signifie pas pour autant qu'une fonction d'utilité *quadratique* rationalise les données. D'ailleurs, une telle fonction n'existe pas dans le cas des huiles et des lessives puisqu'à plusieurs reprises l'indice de Fisher sort légèrement de la plage des IUC.

Contrairement à l'indice de Fisher, la moyenne géométrique sort fréquemment, le plus souvent par le haut, de la plage des IUC (sauf dans le cas du café). Elle constitue donc une moins bonne approximation de l'IUC. Comme l'indice de Laspeyres – dans une moindre mesure toutefois –, elle surpondère les produits dont les prix augmentent le plus. La fonction de Cobb-Douglas ne prend donc pas toujours convenablement en compte les substitutions opérées entre produits ou points de vente (24).

Les perspectives ouvertes par les données scannées

L'utilisation de données scannées apparaît, à plus d'un titre, comme une perspective d'avenir pour le calcul des IPC. Ces données permettent déjà d'affiner certaines pondérations et de les réviser plus fréquemment. De nombreux travaux, menés récemment, ont montré que les données scannées pouvaient contribuer à améliorer la méthodologie des IPC dans différents domaines : formules de calcul pour les micro-indices (Silver, 1995), échantillonnage des produits (de Hahn, Opperdoes et Schut, 1997), échantillonnage des points de vente (Boon, Opperdoes et Schut, 1997), traitement des changements de qualité (Silver, 1999), en sont quelques exemples.

Ces données apparaissent aussi comme un matériau idéal pour la construction d'indices à utilité constante. Certes, l'IUC conduit à la notion de *plage* d'indices (25). L'existence d'une plage d'indices à utilité constante résulte de la théorie des préférences révélées : une infinité de fonctions homothétiques rationalisent les données. Rien ne peut réduire cette plage à une valeur unique (26). Le choix d'une valeur dans cette plage est lié à des nécessités d'usage. L'indice de Fisher est un excellent candidat : il a l'avantage de « tomber » presque toujours dans la plage des IUC. Son calcul requiert cependant des données aussi détaillées que pour un IUC. □

23. Toujours, dans le cas du café.
24. La moyenne géométrique a cependant un autre objet : éviter le phénomène de dérive de la moyenne arithmétique des rapports de prix dans un contexte de chaînage (Lequiller, 1997).
25. Cette notion est étrangère à celle d'intervalle de confiance, inhérente à l'outil mesurant l'indice des prix. Cet outil opère, en effet, un tirage aléatoire des points de vente et ne suit que certains produits. Si le relevé des prix était exhaustif, cet intervalle disparaîtrait : l'estimation serait certaine.
26. Même un relevé exhaustif *des prix et quantités.*

BIBLIOGRAPHIE

S.N. Afriat (1967), « The Construction of a Utility Function from Expenditure Data », *International Economic Review*, 8, pp. 67-77.

S.N. Afriat (1981), *On the Constructabilty of Consistent Price Indices Between Several Periods Simultaneously*, *in* A. Deaton (ed), *Essays in Applied Demand Analysis* (Cambridge : Cambridge University Press).

M. Boskin, E. Dulberguer, Z. Griliches, R. Gordon et D. Jorgensen (1996), *Toward a More Accurate Measure of the Cost of Living*, Final report to the Senate Finance Committee, décember, États-Unis.

J. de Hahn, E. Opperdoes et C. Schut (1997), *Item Sampling in the Consumer Price Index : a Case Study Using Scanner Data*, Internal Report, Statistic Netherland.

M. Boon, E. Opperdoes et C. Schut (1997), *Effect of Outlet Sampling on the Consumer Price Index : a Case Study Using Scanner Data*, Internal Report, Statistic Netherland.

W.E. Diewert (1973), « Afriat and the Revealed Preference Theory », *Review of Economic Studies*, 40, pp. 419-426.

W.E. Diewert (1990), *The Theory of Cost-of-Living Index and the Measurement of Welfare Change*, Price Level Measurement, W.E. Diewert (Editor).

W.E. Diewert et C. Parkan (1978), « Test for Concistency of Consumer Data and Nonparametric Index Numbers », *Working paper*, 78-27, University of British Columbia.

W.E. Diewert et C. Parkan (1985), « Test for Concistency of Data », *Journal of Econometrics*, 30, pp. 127-147.

Insee (1998), *Pour comprendre l'indice des Prix*, Insee Méthodes, n° 81-82.

F. Lequiller (1997), « L'indice des prix à la consommation surestime-t-il l'inflation ? », *Économie et Statistique* , n° 303, pp. 3-32.

F. Lequiller (1998), *Biais des IPC : où en est-on ?*, Insee Méthodes, n° 84-85-86.

F. Magnien et J. Pougnard (1998), *Étude du chaînage d'indices de prix à l'aide de micro-données*, Insee Méthodes, n° 84-85-86.

F. Magnien et J. Pougnard (2000), « Non-parametric Approach to the Cost-of-Living Index », *Document de travail*, n° 0006 de la série Méthodologie statistique, Insee. À paraître dans *Journal of Official Statistics.*

M.E. Manser et R.J. MacDonald (1988), « An Analysis of Substitution Biais in Measuring Inflation, 1959-1985 », *Econometrica*, vol. 56, pp. 909-930.

H. Scobie (1998), *Utilisations possibles des données scannographiques - Étude de cas à l'aide des données sur le café,* Statistique Canada, n° 62F0014MPB au catalogue, Série n° 6.

M. Silver (1995), « Elementary Agregates, Micro-indices and Scanner Data : Some Issues in the Compilation of Consumer Prices », *Review of Income and Wealth*, 41, pp. 427-438.

M. Silver et S. Heravi (1999), *The Measures of Quality-Adjusted Price Changes*, Proceedings of the Measurement of Inflation Conference, August 31-September 1, 1999.

H.L. Varian (1982), « The Nonparametric Approach to Demand Analysis », *Econometrica*, vol. 50, pp. 945-974.

H.L. Varian (1983), « Nonparametric Tests of Consumer Behaviour », *Review of Economic Studies*, 50, pp. 99-110.

S. Warshall (1962), « A Theorem on Boolean Matrices », *Journal of the American Association for Computing Machinery,* vol. 9-11-12.