Les données de caisse : vers des indices de prix à la consommation à utilité constante

Patrick SILLARD(*)
(*)INSEE, Division des prix à la consommation

Introduction

Les indices de prix à la consommation sont désormais des indicateurs économiques normalisés au niveau international. En particulier, les méthodologies appliquées par les différents pays européens sont comparables. Les indices de prix à la consommation sont dits à "qualité constante". En d'autres termes, l'indice vise à refléter les évolutions de prix à services rendus et qualité constants. Du point de vue économique, la notion sous-jacente est celle d'indice de prix à utilité constante (voir par exemple [21, 28]). Ceci étant, hormis sous des hypothèses restrictives de comportement du consommateur, les indices de prix à la consommation, tels qu'ils sont construits aujourd'hui, ne sont pas des indices utilité constante. En effet, raisonner à utilité constante suppose d'observer la substitution qui s'opère entre les consommations de biens au cours des états successifs de l'économie. Autrement dit, on ne peut envisager de construire des indices à utilité constante dans le cas général, que si l'on observe les fonctions de demande.

Sur le plan microéconomique, la question de la substituabilité de biens se pose pour des biens de même nature, vendus dans des lieux d'extension géographique restreinte, marchés pour lesquels la substitution de produits est réalisable concrètement pour un consommateur. La question de savoir de combien s'écartent les indices à qualité constante des indices à utilité constante ne se pose donc véritablement qu'au niveau du calcul des micro-indices, c'est-à-dire de l'opération qui consiste à passer de prix de produits à un indice élémentaire se rapportant à un ensemble d'articles de produits de même nature. Dans la mécanique de calcul d'indice, l'indice élémentaire est ensuite agrégé par agrégation de Laspeyres pour fournir un indice d'ensemble. La présente étude porte donc uniquement sur le calcul de micro-indices de prix.

En pratique, l'observation conjointe des prix de vente et quantités repose sur la disponibilité de données de caisse. Dans le cadre du projet d'intégration des données de caisse dans l'indice des prix à la consommation actuellement conduit par l'Insee, plusieurs enseignes de la grande distribution lui ont donné accès à trois années d'observations hebdomadaires de prix et quantités vendues sur environ 1000 de leurs points de vente et ce, pour 10 familles de produits.

La question du biais des indices de prix a été au centre des débats de la commission Boskin à la fin des années 1990 [3, 23, 7] aux États-Unis. Ces travaux ont mis en évidence l'écart entre les indices de prix classiques et les indices de prix à utilité constante ("*Cost of Living Index*" en anglais). Ils montrent aussi que lorsque les fonctions de demandes sont observées, la construction d'indices à utilité constante est possible. Plusieurs travaux ont été conduits à partir de cette idée sur les différents biais des indices de prix, notamment les biais liés à l'apparition de nouveau produits ou à leur disparition [14]. L'application empirique de ces travaux est restée limitée jusqu'aux travaux portant sur données de caisse. Les travaux empiriques réalisés sur de telles données jusqu'à présent portent essentiellement sur l'étude de phénomènes microéconomiques de substitution [9, 15, 17]. Ce texte examine plus spécifiquement l'application de ces idées aux indices de prix.

1 Le cadre économique des indices de prix à la consommation

Les indices de prix sont des indicateurs économiques destinés à rendre compte de l'évolution générale des prix. Si les produits échangés étaient les mêmes d'une période à l'autre et s'ils étaient échangés dans les mêmes proportions, il suffirait d'observer les prix à chaque période et l'évolution moyenne des prix serait simplement la moyenne des évolutions observées pour chaque produit, individuellement pondérées par les quantités échangées. Ce serait aussi l'évolution des dépenses que doivent réaliser les consommateurs pour maintenir leur volume de consommation de biens d'une période à l'autre. En pratique, les quantités consommées et les produits offerts évoluent : chaque mois, de nouveau produits apparaissent et d'autres disparaissent, de sorte qu'il n'est pas possible de fonder un indicateur destiné à rendre compte de l'évolution générale des prix sur le suivi d'un ensemble parfaitement stable de biens. Lorsque les produits évoluent, le consommateur lui-même s'adapte : il achète des produits nouveaux qui, le cas échéant, se substituent à des produits qu'il ne consomme plus (voir par exemple [24]). Si on suppose que le consommateur est rationnel, c'est-à-dire qu'il forge sa consommation sur des choix cohérents dans le temps, alors l'utilité que lui procure la consommation de nouveau produits est vraisemblablement différente de celle que lui procurait la consommation des produits antérieurement consommés. Afin de rendre compte de cette évolution pour le consommateur, les indices de prix à la consommation s'appuient sur une démarche empirique qui vise à reproduire dans l'enquête, le choix que pourrait être amené à faire un consommateur rationnel qui a l'habitude d'acheter un produit donné dans un magasin et qui ne le retrouve pas. Ce consommateur se trouve donc confronté, si la consommation d'un tel produit lui est toujours utile, à la nécessité d'acheter un produit de substitution dont les caractéristiques diffèrent du produit qu'il achetait auparavant. Le prix du nouveau produit est généralement différent du prix de l'ancien, donc la dépense du consommateur évolue. Mais si les caractéristiques des produits diffèrent, on peut penser que le consommateur en tire un surcroît d'utilité pour lequel il peut être disposé à payer. Si le consommateur est disposé à payer pour une différence de caractéristiques de produits, alors la différence de dépense qu'il consent entre les deux périodes ne traduit pas seulement une différence de prix ; elle traduit aussi une différence de volume de consommation. En pratique, lors d'un remplacement, la valorisation de la différence de caractéristiques des produits s'opère soit de manière explicite en s'appuyant sur un modèle hédonique, estimé par ailleurs, de dépendance des prix aux caractéristiques des produits, soit de manière implicite en faisant l'hypothèse que, toutes choses égales par ailleurs, si le produit antérieur avait été conservé, son prix aurait suivi une évolution parallèle à celle des prix des produits de même nature vendus dans des magasins proches.

On comprend dés lors que le partage volume/prix dans la construction des indices de prix est un sujet fondamental des indices de prix. L'approche empirique qui prévaut dans la plupart des indices modernes repose toutefois sur un certain nombre d'hypothèses qu'il est possible de réexaminer dans un cadre théorique de comportement microéconomique du consommateur [5].

1.1 Modélisation économique

On postulera avec la théorie microéconomique [26], que le consommateur est capable d'émettre des préférences qui le conduisent, étant donné le budget dont il dispose, à classer les paniers de biens que son budget lui permet d'acheter et d'en choisir un qui lui plait plus que tout autre. Le consommateur réalise des choix qui peuvent se modéliser à l'aide d'une fonction d'utilité, opérant dans l'espace des paniers de biens et à valeur réelle. Cette fonction classe les paniers de biens conformément aux préférences du consommateur. Le consommateur choisit le panier de biens qui minimise sa dépense sous contrainte de niveau d'utilité ou, de manière duale, maximise son utilité sous contrainte budgétaire.

Si on note $\mathbf{p}$ le vecteur des prix et $\mathbf{s}$ celui des quantités, u l'utilité du consommateur représentatif (la fonction est supposée indépendante du temps), le consommateur décide de consommer une quantité $\mathbf{x}$ sur le fondement du programme suivant (dont $\mathbf{x}$ est l'argument) :

$$e(\mathbf{p}, U) = \left| \begin{array}{l} \min\limits_{\mathbf{s}} \mathbf{p}.\mathbf{s} \\ \text{s.c. } u(\mathbf{s}) = U \end{array} \right. \tag{1}$$

Ainsi construite, $e(\mathbf{p}, U)$ est la fonction de dépense : c'est la dépense minimale que doit consentir le consommateur pour atteindre le niveau d'utilité U conditionnellement aux prix $\mathbf{p}$. De manière duale, l'utilité maximale atteinte pour une dépense R conditionnellement aux prix $\mathbf{p}$ est appelée utilité indirecte. On la note $v(\mathbf{p}, R)$.

On considère à présent deux instants : un instant de référence t et un instant t' ($t' > t$), entre lesquels on souhaite construire un indice traduisant l'évolution des prix. Les observables sont les prix aux deux instants et la dépense initiale, notée R_t. A l'instant t, le consommateur-optimisateur atteint un niveau d'utilité $v(\mathbf{p}_t, R_t)$. Pour atteindre le même niveau d'utilité à l'instant t', il devra consentir une dépense $e(\mathbf{p}_{t'}, v(\mathbf{p}_t, R_t))$. Par construction, l'indice des prix à utilité constante est simplement :

$$I_{UC}^{t',t}(R_t) = \frac{e(\mathbf{p}_{t'}, v(\mathbf{p}_t, R_t))}{R_t} \tag{2}$$

Il traduit l'évolution, entre t et t', du budget que le consommateur doit dépenser pour atteindre en t' le niveau d'utilité qu'il atteignait en t avec une dépense R_t. Il est utile à ce stade de définir la fonction correspondant au numérateur de l'expression précédente ($\mathbf{p}$ et $\mathbf{q}$ sont deux vecteur de prix quelconques ; R est un budget quelconque) :

$$\mu(\mathbf{p}; \mathbf{q}, R) = e(\mathbf{p}, v(\mathbf{q}, R)) \tag{3}$$

Cette fonction est appelée utilité indirecte en équivalent monétaire. Elle correspond à la dépense que le consommateur doit consentir avec le vecteur de prix $\mathbf{p}$ pour atteindre le niveau d'utilité qu'il atteint avec le vecteur de prix $\mathbf{q}$ pour une dépense de R.

A l'aide de la fonction d'utilité indirecte en équivalent monétaire, l'indice à utilité constante s'écrit :

$$I_{UC}^{t',t}(R) = \frac{\mu(\mathbf{p}_{t'}; \mathbf{p}_t, R)}{R} \tag{4}$$

1.2 L'indice à utilité constante pour des fonctions d'utilité usuelles

Il est intéressant d'examiner la forme prise par l'indice des prix (4) lorsqu'une fonction d'utilité est explicitement choisie pour modéliser le choix du consommateur. Nous examinons le cas de trois fonctions classiques : l'utilité de Cobb-Douglas, l'utilité de Léontief et l'utilité CES [1]. Chacune de ces fonctions d'utilité est fondée sur des présupposés sur le comportement du consommateur, notamment en ce qui concerne le caractère substituable des produits composant le panier de biens consommés. La fonction CES a été introduite par [1]. La principale caractéristique de l'utilité CES est de reposer sur l'hypothèse que les biens du panier de consommation sont substituables avec une élasticité de substitution identique quel que soit le couple de biens (i, j) considéré : si x_i est la demande de bien i et p_i le prix du même bien, l'élasticité de substitution $d\ln(x_i/x_j)/d\ln(p_i/p_j)$ vaut $-\varepsilon$ pour la fonction d'utilité CES présentée à la table 1. Cette fonction est concave lorsque $\varepsilon > 0$. Nous supposerons dorénavant que cette inégalité est vérifiée.

Les utilités de Cobb-Douglas et Léontief sont des cas limites de l'utilité CES. La première correspond au cas où $\varepsilon \to 1$ et la seconde, au cas où $\varepsilon \to 0$. Les expressions des utilités, demande et indices sont données à la table 1.

Table 1 – Demande et indice des prix pour les utilités CES, de Cobb-Douglas et de Léontief

		CES	Cobb-Douglas	Léontief
utilité :	$u(\mathbf{s}) =$	$\left(\sum_k \alpha_k s_k^{(\varepsilon-1)/\varepsilon}\right)^{\varepsilon/(\varepsilon-1)}$	$s_1^{\alpha_1} \times \ldots \times s_n^{\alpha_n}$	$\min\{\alpha_1 s_1, \ldots, \alpha_n s_n\}$
demande :	$x_i(\mathbf{p}, R) =$	$\frac{R}{p_i} \frac{\alpha_i \left(\frac{p_i}{\alpha_i}\right)^{1-\varepsilon}}{\sum_k \alpha_k \left(\frac{p_k}{\alpha_k}\right)^{1-\varepsilon}}$	$\alpha_i \frac{R}{p_i}$	$R/\left(\alpha_i \sum_k \frac{p_k}{\alpha_k}\right)$
indice :	$I^{t',t} =$	$\frac{\left(\sum_k \alpha_k \left(\frac{p_k^{t'}}{\alpha_k}\right)^{1-\varepsilon}\right)^{1/(1-\varepsilon)}}{\left(\sum_j \alpha_j \left(\frac{p_j^t}{\alpha_j}\right)^{1-\varepsilon}\right)^{1/(1-\varepsilon)}}$	$\prod_{i=1}^n \left(\frac{p_i^{t'}}{p_i^t}\right)^{\alpha_i}$	$\left(\sum_i \frac{p_i^{t'}}{\alpha_i}\right) / \left(\sum_j \frac{p_j^t}{\alpha_j}\right)$

Notes : Le consommateur consomme un ensemble de n biens notés $\{1, \ldots, n\}$ en quantités $\mathbf{s}$ et aux prix $\mathbf{p}$; la fonction d'utilité CES est définie (concave) pour $\varepsilon > 0$; les coefficients α_i pour $i \in \{1, \ldots, n\}$ sont des pondérations strictement positives traduisant les préférences du consommateur ; R est le budget (exogène) que le consommateur consacre à l'achat de l'ensemble des biens$\{1, \ldots, n\}$; p_i^t est le prix du bien i à la période t ; $I^{t',t}$ est l'indice des prix associé au panier de biens considéré de la période t' par rapport à la période t correspondant à la définition (2). Formellement, l'indice dépend du budget consacré à l'achat des biens en période t. Mais ici, les utilités étant homothétiques, le budget de la période t se simplifie dans l'expression des indices. Les coefficients $(\alpha_i)_{i\in\{1,\ldots,n\}}$ sont des paramètres des fonctions d'utilité et de demande. Ils sont positifs.

On note que la fonction d'utilité de Léontief conduit à un indice de type Laspeyres, tandis que la fonction d'utilité de Cobb-Douglas conduit à un indice sous forme de moyenne géométrique de rapports de prix (indice de Laspeyres géométrique, également connu sous

1. Pour *constant elasticity of substitution.*

le nom d'indice de Jevons lorsque les pondérations sont unitaires). Ces deux formules d'indices sont utilisées dans l'indice des prix à la consommation français pour le calcul des micro-indices. Dans le cas de produits parfaitement homogènes, comme la baguette de pain, la formule utilisée est celle d'un indice de Laspeyres. Dans le cas de produits moins homogènes, comme par exemple, les ordinateurs portables, la formule utilisée pour les micro-indices est la moyenne géométrique de rapports de prix.

Le choix de la fonction d'utilité a évidemment des conséquences importantes sur la trajectoire de l'indice à utilité constante. En particulier, plus les produits sont substituables, plus les variations des prix relatifs ont une incidence sur l'indice. En effet, lorsque l'élasticité de substitution est élevée, si le prix relatif d'un bien augmente, sa consommation diminue au profit de la consommation des autres biens, de sorte que l'accroissement de la dépense d'ensemble est atténué par rapport à l'accroissement qui résulterait de la simple augmentation de la dépense pour le bien considéré, les quantités étant fixées par ailleurs. Ce dernier cas correspond d'ailleurs à ce qui se produit lorsque l'élasticité est nulle. L'utilité sous-jacente est alors celle de Léontief et l'indice est celui de Laspeyres (voit table 1) : un accroissement du prix relatif d'un bien se traduit par un accroissement de l'indice proportionnel au taux de croissance du prix, pondéré par le poids du bien considéré dans la dépense totale. A l'inverse, si le prix relatif d'un bien diminue, la substitution entraîne une amplification de la baisse par rapport à la situtation dans laquelle les biens ne sont pas substituables. En résumé, plus la substituabilité des produits est élevée, plus les hausses de prix relatifs vont être atténuées et plus les baisses vont être accentuées.

Un autre point mérite d'être noté sur la cohérence du modèle de comportement. Pour qu'un modèle de comportement permette d'élaborer un indice de prix, il est souhaitable que la formation des choix à des périodes de temps proches soit stable, faute de quoi, ces choix sont difficilement comparables. Autrement dit, dans un modèle à utilité constante, il est souhaitable que l'on puisse faire l'hypothèse que l'utilité sous-jacente est la même sur des périodes adjacentes. Naturellement, dans le commerce, comme cela a déjà été évoqué plus haut, les biens disparaissent et de nouveaux biens apparaissent continuellement. Pour qu'un modèle de comportement soit compatible avec cet état de fait, tout en supposant que la fonction d'utilité est stable dans le temps, il faut que ce modèle permette la description d'une consommation de certains biens nulle. En effet, lors d'une période donnée, la consommation de biens qui vont apparaître dans le futur est nulle, de même que celle de biens qui ont disparu. Nous allons voir que tous les modèles dérivés d'une utilité CES n'autorisent pas une telle description.

Par exemple, dans un modèle de comportement à utilité de Léontief, aucun bien ne devrait apparaître puisque, dans ce modèle, le consommateur est supposé consommer tous les biens en quantité non nulle (voir la fonction de demande de Léontief – table 1). Sur un plan technique, cette impossibilité à décrire une consommation nulle de certains biens se traduit notamment par la nullité de l'utilité lorsque la quantité consommée d'un bien du panier est nulle. Moyennant quoi, l'utilisation d'un tel modèle conduit à abandonner l'hypothèse fondatrice de stabilité dans le temps de la fonction d'utilité [2]. La possibilité de substituer un bien à un autre est donc une condition nécessaire pour qu'un modèle de comportement autorise, dans sa description de la consommation, la nullité des quantités consommées pour certains biens à certaines périodes. Dans le cas d'une élasticité CES, il faut en outre que $\varepsilon > 1$. En effet, prenons le cas d'un modèle de Cobb-Douglas. Pour que

2. C'est ce qui est fait en pratique à travers la mécanique du remplacement de produits évoquée plus haut.

la quantité consommée soit nulle (voir la fonction de demande – table 1), il faut que le prix soit infini et dans ce cas, l'indice, comme l'utilité, sont dégénérés et égaux à 0. Il en est de même pour une utilité CES lorsque $0 < \varepsilon \leqslant 1$.

En revanche, pour une utilité CES, si $\varepsilon > 1$, alors la quantité consommée de bien i est égale à 0 lorsque le prix p_i^t du bien est infini à la période t. Dans ce cas, le bien ne contribue plus à l'indice, hormis lors d'une transition entre deux périodes, l'une où il est consommé et l'autre où il ne l'est pas. Dans un tel modèle, l'apparition à la vente d'un nouveau bien se traduit par un accroissement de l'utilité du consommateur et une baisse des prix. A l'inverse, la disparition se traduit par une hausse des prix.

La possibilité de calculer des indices de prix sur des paniers évoluant dans le temps a été envisagée et traitée sur un plan théorique par Balk [2]. Elle a été appliquée avec succès par Mesler [22] avec une utilité CES, l'une des principale difficultés résidant dans l'estimation de ε qui peut, le cas échéant, donner lieu à des valeurs inférieures à 1 et donc ne permettant pas d'appliquer la méthode.

1.3 De l'observation des fonctions de demande aux indices à utilité constante

Sans faire d'hypothèses économiques autre que le comportement de maximisation de l'utilité du consommateur, il est possible de passer de l'observation de fonctions de demandes empiriques (en l'occurrence pour un consommateur représentatif qui achèterait ses produits dans un point de vente donné au cours de périodes successives) d'un ensemble de produits élémentaires plus ou moins substituables du point de vue du consommateur, à une fonction d'utilité. Le cadre théorique appliqué ici est celui de l'intégrabilité des fonctions de demande [19].

Le passage de fonctions de demande à des fonctions d'utilité ou des fonctions d'utilité en équivalent monétaire a été examiné en détail par Varian [26, 27] et Hausman [12, 13].

En premier lieu, toutes les fonctions de demande ne dérivent pas d'une utilité. On montre [19, 26] que pour dériver de l'optimisation d'une utilité sous contrainte budgétaire, les fonctions de demandes de biens i et j quelconques doivent vérifier la relation de symétrie suivante (appelée condition d'intégrabilité dans la littérature) portant sur les termes de la matrice de Slutsky :

$$\forall (i,j) \in \{1, \ldots, n\} \ , \ \frac{\partial x_i(\mathbf{p}, R)}{\partial p_j} + \frac{\partial x_i(\mathbf{p}, R)}{\partial R} x_j(\mathbf{p}, R) = \frac{\partial x_j(\mathbf{p}, R)}{\partial p_i} + \frac{\partial x_j(\mathbf{p}, R)}{\partial R} x_i(\mathbf{p}, R) \quad (5)$$

où x_i et x_j sont les fonctions de demande [3] des biens i et j.

Une fois cette propriété vérifiée, on montre qu'il existe une relation entre la demande conditionnelle aux prix $\mathbf{p}$ à budget R donné, $\mathbf{x}(\mathbf{p}, R)$, et l'utilité indirecte en équivalent monétaire. La relation dérive du lemme de Shephard et s'écrit sous forme d'une équation différentielle aux dérivées partielles :

$$\forall i \in \{1, \ldots, n\} \ , \ x_i(\mathbf{p}, \mu(\mathbf{p}; \mathbf{q}, R)) = \frac{\partial \mu(\mathbf{p}; \mathbf{q}, R)}{\partial p_i} \quad (6)$$

Dans ce système d'équations, $\mathbf{q}$ et R sont des paramètres, la fonction μ étant une fonction de $\mathbf{p}$. Si la fonction x_i est observée (i.e. sa dépendance vis-à-vis du vecteur de prix $\mathbf{p}$ et de

3. Ces fonctions dépendent du couple de grandeurs $\mathbf{p}$ et R. Elles sont appelées, dans la littérature, demandes Marshaliennes. Pour l'ensemble des biens, le vecteur de fonctions de demande est noté $\mathbf{w}(\mathbf{p}, R)$.

la dépense R sont établies), alors le système d'équations (6) est un système d'équations différentielles partielles en μ, fonction du vecteur $\mathbf{p}$ et de paramètres $\mathbf{q}$ et R. A cette équation différentielle s'ajoute une condition limite

$$\mu(\mathbf{q};\mathbf{q},R) = R \tag{7}$$

Moyennant quoi, à l'aide des relations (5), (6) et (7), il est possible de passer de relations (observables) de demande pour un ensemble $\{1,\ldots,n\}$ de produits élémentaires à un indice à utilité constante sur cet ensemble de biens.

En pratique, deux solutions peuvent être envisagées. La première consiste à spécifier une fonction de demande empirique vérifiant les propriétés de symétrie (5), à estimer économétriquement cette fonction, puis, à calculer un indice de prix en résolvant, pour la spécification retenue, le système d'équations différentielles (6). C'est cette méthode que nous appliquons dans cette étude. Une seconde méthode pourrait être envisagée. Elle consiste à travailler sur des formes non paramétriques de fonctions de demande, puis à résoudre numériquement le système d'équations différentielles (6). Cette méthode a été examinée par Varian [25] en établissant un parallèle avec la notion de préférences révélées. Toutefois les applications concrètes semblent délicates à mettre en œuvre. Hausman et Newey [16] ont appliqué des méthodes non paramétriques à l'évaluation des effets d'une variation de prix sur le bien-être.

Revenons à l'approche paramétrique retenue dans cette étude et traitons un exemple à partir de la spécification d'une fonction de demande et son estimation économétrique. Il existe naturellement différentes spécifications possibles de la demande. La spécification log-linéaire est assez souvent utilisée pour caractériser des fonctions de demande pour un type de produit donné. Nous adoptons donc une équation de demande de bien i de la forme suivante :

$$x_i(\mathbf{p},R) = p_i^{\alpha_i} R^{\beta_i} e^{\gamma_i}$$

Cette spécification est log-linéaire. L'application des conditions d'intégrabilité (5) conduit à (M_{ij} est le membre de gauche de l'équation 5) :

$$M_{ij} = \frac{\beta_i}{R} x_i(\mathbf{p},R) x_j(\mathbf{p},R) \quad \text{et} \quad M_{ij} = M_{ji} \Leftrightarrow \beta_i \equiv \beta$$

Sous cette réserve, la spécification

$$x_i(\mathbf{p},R) = p_i^{\alpha_i} R^{\beta} e^{\gamma_i} \tag{8}$$

vérifie les conditions d'intégrabilité.

La résolution du système (6) est proposée en annexe pour cette fonction de demande. On aboutit finalement à l'utilité indirecte en équivalent monétaire associée à la demande linéaire précédente :

$$\mu(\mathbf{p};\mathbf{q},R) = \left\{ (1-\beta) \left[\frac{R^{1-\beta}}{1-\beta} + \sum_{i=1}^{n} e^{\gamma_i} . \frac{p_i^{1+\alpha_i} - q_i^{1+\alpha_i}}{1+\alpha_i} \right] \right\}^{\frac{1}{1-\beta}} \tag{9}$$

Il en découle une formule d'indice à utilité constante (traduisant l'évolution des prix entre t et t' où $t' \geqslant t$ et R_t est la dépense du consommateur représentatif à la période t) :

$$I_{LogLin}^{t',t} = \left\{ 1 + (1-\beta) R_t^{\beta-1} \sum_{i=1}^{n} e^{\gamma_i} . \frac{p_{it'}^{1+\alpha_i} - p_{it}^{1+\alpha_i}}{1+\alpha_i} \right\}^{\frac{1}{1-\beta}} \tag{10}$$

A la différence des indices présentés à la table 1, cet indice fait intervenir le budget consacré par le consommateur à l'achat des biens lors de la période de référence. L'utilité sous-jacente n'est donc pas homothétique, ce qui en soi est assez conforme à l'intuition puisqu'il n'y a aucune raison pour qu'une croissance du budget se traduise par une croissance uniforme des quantités consommées. Au contraire, il est assez vraisemblable que, lorsque le budget augmente, la consommation se déplace vers des produits de meilleure qualité, et sur un plan global, vers des biens non alimentaires. On retrouve ce type de résultats dans les travaux empiriques sur les courbes d'Engel [4, 18].

2 Les données

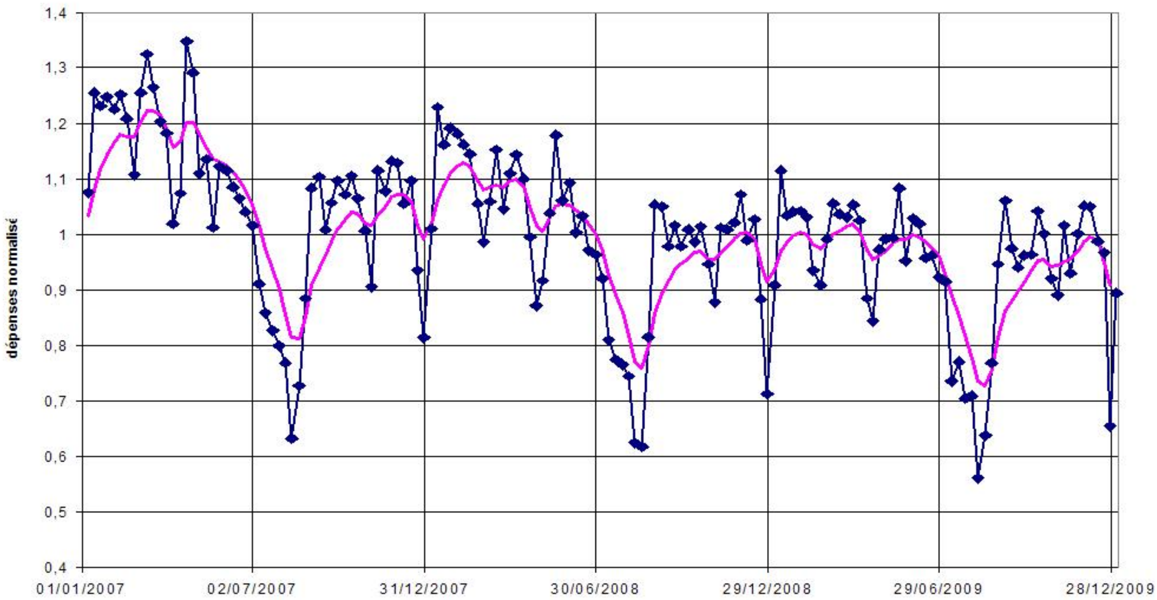
Le modèle proposé à la section précédente repose sur des hypothèses de comportement rationnel d'un consommateur représentatif. La notion de consommateur représentatif n'est pas évidente. On sait par ailleurs qu'en général, l'agrégation des préférences ne repose pas l'optimisation d'une fonction d'utilité collective de sorte qu'il n'est pas certain que le comportement agrégé des consommateurs dérive de l'optimisation d'une fonction d'utilité. Néanmoins, si on souhaite conserver le fondement théorique exposé à la section précédente, on est conduit à s'appuyer sur la notion de consommateur représentatif. En d'autres termes, sur le marché analysé, tout se passe comme si un consommateur unique achète tous les biens consommés sur ce marché, substitue un bien à un autre selon une décision rationnelle découlant de l'optimisation d'une fonction d'utilité et que ce consommateur représentatif est *price taker* sur le marché considéré. Naturellement, l'enchaînement de ces hypothèses rend assez peu probable l'existence d'une marché d'ensemble qui fonctionnerait exactement de cette manière. En revanche, on peut imaginer que plus on descend à un niveau microéconomique fin, plus on se rapproche de conditions de fonctionnement proches de celles exposées. Afin de se placer dans un cadre réaliste, nous devons donc travailler sur un marché local où les consommateurs effectuent un choix entre des produits réellement substituables. Nous avons donc choisi de travailler sur le marché des yaourts vendus dans un magasin de la grande distribution. Les données couvrent trois ans sur lesquels on dispose, pour ce magasin, des quantités vendues et des prix hebdomadaires de tous les yaourts vendus au moins une fois dans la semaine dans ce point de vente.

Sur un tel marché, les hypothèses de consommateur représentatif sont assez vraisemblables puisque la possibilité de substituer un produit à un autre est réalisable concrètement. Par ailleurs, pour le consommateur qui consomme un yaourt dans ce magasin, il est plus aisé de substituer un yaourt à un autre dans ce magasin plutôt que de se rendre dans un autre magasin pour acheter un produit remplaçant. Ainsi, un modèle agrégé de comportement sur le marché des yaourts vendus dans le magasin étudié, s'il ne résulte pas de l'agrégation des préférences individuelles, constitue néanmoins une représentation plausible du fonctionnement du micro-marché étudié. Par conséquent, un indice à utilité constante fondé sur cette modélisation reflétera certainement très convenablement l'évolution des prix sur ce marché.

Le lot de données étudiées comprend environ 35000 observations réparties sur 157 semaines (de janvier 2007 à décembre 2009). Sur cette période, environ 600 références différentes (codes-barres) de yaourts ont été vendues au moins une fois. Dans cette étude, le code-barres constitue l'élément d'identification du produit. En moyenne, chaque référence est vendue 950 fois sur les trois années d'observation ; la médiane se situe à 300, tandis que les 1$^{\text{er}}$ et 9$^{\text{ème}}$ déciles valent respectivement 30 et 2300.

La figure 1 donne le tracé du chiffre d'affaires (normalisé par la moyenne des trois

Figure 1 – Chiffre d'affaire associé au marché des yaourts pour le magasin étudié



Note : Les "dépenses normalisées" sont calculées comme le rapport du chiffre d'affaires hebdomadaire et du chiffre d'affaire hebdomadaire moyen observé pour l'ensemble des trois années d'observation. La courbe pleine est la version filtrée du graphe hebdomadaire par un filtre de Butterworth passe-bas [11].

ans) du marché des yaourts pour le magasin considéré au cours des trois années d'étude. Sur la période, le chiffre d'affaires varie pratiquement du simple au double avec des effets saisonniers très marqués, le point haut se situant approximativement au premier semestre de l'année, tandis que le point bas se situe au mois d'août. Il est aussi caractérisé par une légère décroissance tendancielle sur la période.

3 Application

Trois indices ont été calculés sur les bases exposées à la section 1 : un indice de Laspeyres chaîné annuellement, un indice à utilité constante de type CES (formule de la table 1) et un indice fondé sur l'inversion des fonctions de demande (formule 10).

L'indice de Laspeyres repose sur un panier fixe de biens mis à jour annuellement. Le panier est la réunion des biens vendus lors de la *période de base* pour l'année, couvrant les quatre premières semaines de l'année (i.e. le mois de janvier). Les prix de base correspondent à la moyenne des prix des produits calculée sur la période de base et les quantités de base correspondent à la somme des quantités hebdomadaires vendues, calculée sur la période de base. La formule de calcul appliquée est du type de celle de l'indice associée à l'utilité de Léontief avec des coefficients de pondérations correspondant aux quantités de base. Pour l'exercice, lorsqu'un produit n'est pas vendu lors d'un mois donné, son prix est estimé en appliquant au dernier prix observé la variation mensuelle effectivement observée sur les produits présents au cours des deux mois successifs. Cet algorithme est appliqué itérativement en cas de disparition définitive du produit. C'est aux remplacements près, la procédure appliquée dans l'IPC lorsqu'un produit est absent du magasin où il est suivi. Dans l'alimentaire, les effets qualité sont très faibles – estimés à moins de 0,1% de croissance annuelle [10] – de sorte qu'on peut penser que l'absence de mécanique de remplacement ne biaise probablement que faiblement l'indice obtenu. En revanche, l'indice ainsi calculé ne prend pas en compte les promotions des fabricants (comme par exemple la vente de deux yaourts pour le prix d'un). En effet, au passage en caisse, les promotions des fabricants apparaissent avec un code-barres sans relation avec le produit vendu hors promotion. Le lien entre les deux produits n'est pas fait dans cette étude et par la mécanique du panier fixe annuel, les promotions de fabricants qui n'auraient pas lieu en janvier ne sont pas prises en compte dans l'indice. Cette limite implique que l'indice de Laspeyres ainsi construit n'est pas sensible aux promotions des fabricants, ce qui constitue donc un biais. A noter que dans l'IPC, un tel biais n'existe pas puisque les promotions des fabricants sont suivies et, le cas échéant, reliées par l'enquêteur au produit élémentaire concerné, le prix suivi dans l'IPC étant un prix unitaire (par unité de volume, de poids, ou autre selon ce qui est pertinent pour le produit concerné).

Le calcul d'un indice CES suppose d'abord que les paramètres de pondération individuelle des produits et d'élasticité de substitution soient fixés. Ils sont estimés économétriquement en considérant l'ensemble des achats de yaourts (prix et quantités) réalisés au cours des trois années d'observation. Cette estimation permet en particulier de fixer, pour chaque produit i, un coefficient de préférence élémentaire α_i (voir formule d'indice table 1) qui traduit, à tout instant, la préférence du consommateur pour le bien considéré, que ce bien soit ou non disponible à la vente. Ainsi construite, l'utilité qui conduit les choix du consommateur représentatif est stable dans le temps : les nouveau produits peuvent donc être intégrés au fil du temps et les produits qui disparaissent de la vente sortent de l'indice. Les promotions des fabricants s'insèrent naturellement dans le calcul.

Le calcul d'un indice à utilité constante fondé sur une demande log-linéaire (formule 10) exige, comme pour l'utilité CES, l'estimation de paramètres. Comme pour l'utilité CES, l'utilité est stable dans le temps : selon les valeurs estimées de paramètres, les nouveaux produits peuvent être intégrés naturellement dans l'indice en considérant que les prix des produits absents de la vente sur un mois donné sont infinis.

3.1 Estimation économétrique des paramètres des modèles

La détermination des paramètres de l'utilité CES se fonde sur la relation de base de l'équation de demande (voir table 1). En prenant le logarithme de la fonction de demande, nous obtenons une équation structurelle de la forme (bien i, période t) :

$$\ln(x_{it}) = \underbrace{\varepsilon \ln(\alpha_i)}_{c_i} - \varepsilon \ln(p_{it}) + \underbrace{\ln R_t - \ln \left[\sum_k \alpha_k \left(\frac{p_{kt}}{\alpha_k} \right)^{1-\varepsilon} \right]}_{\nu_t} \tag{11}$$

Les coefficients c_i et ε peuvent donc être estimés en régressant $\ln(x_{it})$ sur $\ln(p_{it})$, le terme ν_t étant considéré ici comme l'aléa. Plusieurs méthodes d'estimation sur données de panel peuvent être envisagées (estimateur en différences ou à effet fixe). Le point essentiel est qu'il est hautement vraisemblable que le prix p_{it} soit endogène dans l'équation. En effet, on peut imaginer que le vendeur adapte son prix à la demande, ce qui engendre dans l'estimation par moindres carrés ordinaires un biais de simultanéité. Afin de lever cette endogénéité, il convient d'utiliser un instrument. L'équation structurelle qui nous intéresse ici est une équation de demande. Pour identifier cette équation, nous cherchons donc un instrument du prix, c'est-à-dire une variable corrélée avec le prix et qui, dans l'équation de demande, est exogène. Concrètement, toute variable liée, par exemple, aux coûts de production est un candidat potentiel. Nous utilisons la variable d'indice des prix de gros alimentaires pour le lait de consommation publiée par l'Insee tous les mois. Cette variable est ré-échantillonnée afin de disposer d'un indice hebdomadaire. L'élasticité (et les effets fixes produits le cas échéant) sont estimés par doubles moindres carrés. Les résultats des différentes régressions sont détaillés dans la table 2. Ils indiquent de manière claire que le prix est une variable endogène dans l'équation de demande et que l'instrument retenu est valide : l'estimation de l'élasticité passe de $0,373$ dans le cas des moindres carrés sans effets fixes à $1,83$ pour l'estimation en doubles moindres carrés sur les différences. Les régressions en niveau et en différences conduisent aux même résultats, s'agissant des doubles moindres carrés. Nous retenons la régression en niveau par double moindres carrés comme estimation de référence. Outre le fait que l'estimation est plus précise, l'estimation d'effets fixe produits (c_i dans l'équation 11) nous permet d'estimer les coefficients α_i. En effet, dans ce cas, un estimateur sans biais est $\hat{\alpha_i} = \exp(\hat{c_i}/\hat{\varepsilon})$.

TABLE 2 – Estimation des paramètres de l'utilité CES

Régression	$\hat{\varepsilon}$	effet fixe produit	AUX
OLS	$0,373^{***}$ (0,014)	non	
Est. à effet fixe	1.21^{***} (0,039)	oui	
OLS sur les Δ	1.27^{***} (0.057)	non	
2SLS sur les niveaux	1.82^{***} (0.090)	oui	$0,36^{***}$ (0.005)
2SLS sur les Δ	1.83^{***} (0.222)	non	

Note : Δ désigne les différences premières. 3 étoiles indiquent une significativité du paramètre à 1%. Les écart-types sont indiquées entre parenthèses. Pour les doubles moindres carrés (indiqués "2SLS" dans le tableau ; les moindres carrés ordinaires sont indiqués "OLS"). La régression auxiliaire consiste à régresser le logarithme du prix de vente du produit sur le logarithme de l'indice des prix de gros alimentaires pour le lait de consommation et des effets fixes produits lorsque ces derniers sont inclus dans l'équation principale (colonne "effets fixes produits"). Quand la régression est en différences, l'équation auxiliaire est également traitée en différences. Environ 34000 observations contribuent au calcul. AUX est le coefficient du logarithme du prix du lait dans la régression auxiliaire du logarithme du prix

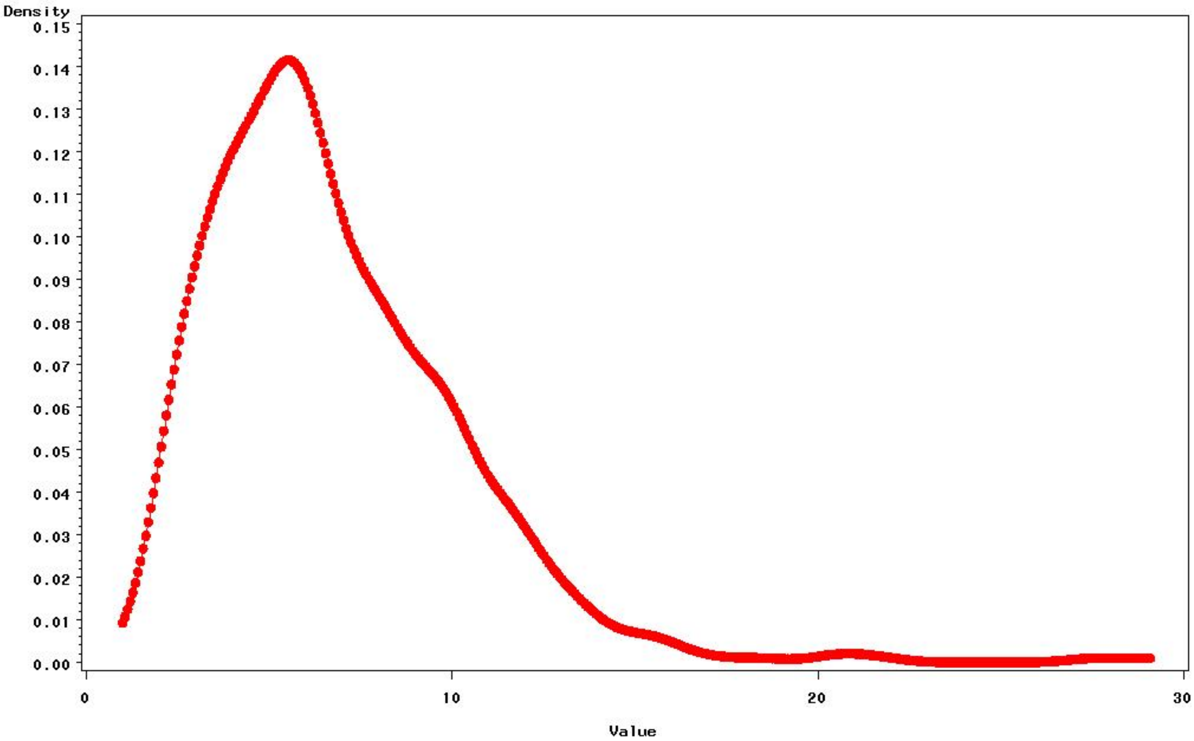
La figure 2 indique la distribution des α_i. Cette distribution nous renseigne sur les différences d'attrait que revêtent les différents yaourts aux yeux du consommateur. Les différences sont assez sensibles et la contribution à l'utilité peut aller d'un rapport de 1 à 20 selon les produits considérés.

Comme pour la détermination des paramètres de l'utilité CES, l'estimation des paramètres de la demande empirique log-linéaire se fonde sur la relation de base de l'équation de demande (8). En prenant le logarithme de la fonction, nous obtenons une équation structurelle :

$$\ln(x_{it}) = \gamma_i + \alpha_i \ln(p_{it}) + \beta \ln(R_t) \tag{12}$$

Les coefficients de l'équation précédente peuvent s'estimer par régression de $\ln(x_{it})$ sur les $\ln(p_{it})$ et $\ln(R_t)$.

FIGURE 2 – Distribution des coefficients α_i figurant dans l'équation 11 (pondération des produits dans l'utilité CES)



Note : Les coefficients α_i sont déduits des effets fixes c_i de la régression 2SLS en niveau et découlent de la relation $\hat{\alpha_i} = \exp(\hat{c_i}/\hat{\varepsilon})$. Estimation par la méthode des noyaux.

Néanmoins, comme précédemment, le logarithme du prix p_{it} est endogène dans la régression, pour les mêmes raisons que celles évoquées à propos de l'utilité CES. Nous n'avons à disposition qu'un seul instrument pour l'ensemble des prix des produits, par conséquent nous ne sommes pas en mesure d'identifier l'ensemble des coefficients α_i. En revanche, si nous faisons l'hypothèse que tous les α_i sont identiques, alors l'équation précédente se simplifie en

$$\ln(x_{it}) = \gamma_i + \alpha \ln(p_{it}) + \beta \ln(R_t) \tag{13}$$

Cette équation est identifiable si la dépense R_t est exogène dans l'équation ou si on dispose d'un instrument pour la dépense R_t. On peut imaginer des causes d'endogénéité de la dépense : par exemple, si le consommateur est *price taker*, alors il est vraisemblable que le vendeur, en situation de monopole, fixe les prix de façon à maximiser sa recette (en l'occurrence la dépense du consommateur) en connaissant la réponse du consommateur aux variations de prix qu'il introduit ; dans ce cas, les quantités vendues résultent d'une détermination simultanée des prix, de la dépense et des quantités. On cherche à tester l'endogénéité de la variable de dépense à l'aide d'un instrument de cette variable dans l'équation de demande, c'est-à-dire à l'aide d'une variable corrélée à la dépense mais qui ne joue pas dans la formation de la demande. Par exemple, une variable de salaire ne convient pas, car si elle est effectivement corrélée avec la dépense, il est vraisemblable qu'elle joue aussi dans la formation de la demande. Donc cette variable ne peut pas constituer un instrument pour une équation de demande. La solution retenue consiste à considérer une variable indicatrice de fin de mois. En effet, sur le plan des déterminants de la demande, rien n'indique que la fin du mois constitue une période privilégiée pour acheter des yaourts. En revanche, la plupart des salariés voient leur paie versée en fin de mois, donc la période de fin de mois est vraisemblablement une période dans laquelle le consommateur dispose de moins de ressources financières à consacrer aux achats de consommation courante. Une variable indicatrice de fin de mois (nous avons retenu la période du 23 au 30) est donc un candidat : empiriquement, cette variable est effectivement corrélée négativement à la dépense (voir colonne "AUX2" de la table 3). En revanche, on constate que l'instrument est assez peu puissant puisque les écart-types du coefficient de la dépense dans les régressions par moindres carrés en deux étapes sont considérablement augmentés. Au final, un test de Hausman indique que la dépense n'est pas endogène dans l'équation de demande. Nous retenons donc la spécification en niveau avec instrumentation du prix seulement (régression "2SLS sur les niveaux (^{2})" de la table 3).

Comme pour l'indice CES, l'indice fondé sur la demande log-linéaire permet de traiter les produits entrants et sortants de manière naturelle en considérant que les produits absents sur une période sont de prix infini. La figure 3 donne le tracé des coefficients γ_i de la spécification (13) pour la régression de référence. Ces coefficients sont tous négatifs et leur dispersion va de 0 à -6, ce qui signifie que le poids de la contribution des produits à l'indice (voir relation 10) va de 1 à $1/400$.

3.2 Résultats sur les indices

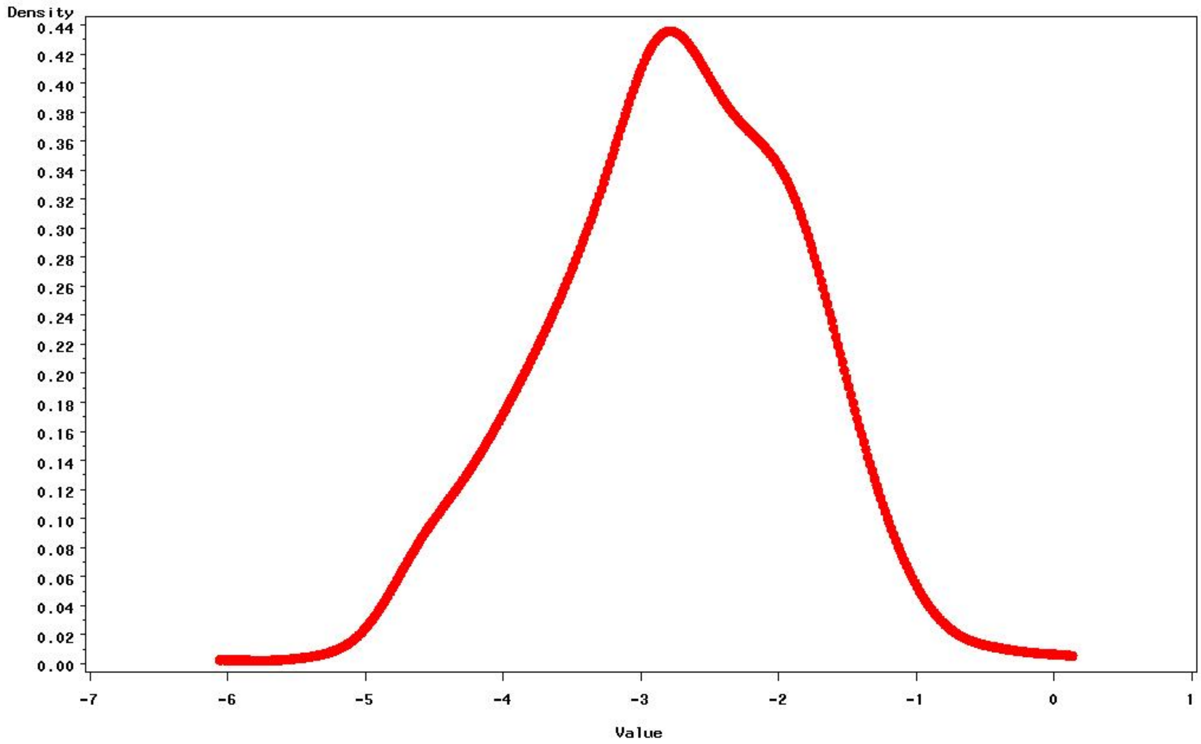
La figure 4 donne le tracé des trois indices étudiés. Les indices sont calculés mensuellement, c'est à dire qu'on considère ici les quantités vendues mensuellement et le prix moyen pratiqué pour chaque article sur le mois considéré. On peut noter que l'indice de Laspeyres est moins bruité que les deux autres. Cela est lié au fait que les indices CES et log-linéaire sont, par construction, sensibles aux entrées-sorties de biens dans le magasin. Ceci vaut pour les produits dont la présence est quasi-continue mais qui peuvent

TABLE 3 – Estimation des paramètres de la demande log-linéaire

Régression	$\hat{\alpha}$	$\hat{\beta}$	effet fixe produit	AUX1	AUX2
OLS sur les niveaux	$-1,13^{***}$ (0,04)	$1,09^{***}$ (0,02)	oui		
OLS sur les Δ	$-1,19^{***}$ (0,06)	$1,00^{***}$ (0,03)	non		
2SLS sur les niveaux	$-2,00^{***}$ (0,18)	$0,67^{*}$ (0,42)	oui	$0,26^{***}$ (0,02)	$-0,017^{***}$ (0,002)
2SLS sur les Δ	$-1,21$ (1,26)	$2,51^{*}$ (1,63)	non	$0,24^{***}$ (0,03)	$-0,005^{***}$ (0,001)
2SLS sur les niveaux (^{2})	$-1,86^{***}$ (0,10)	$1,08^{***}$ (0,02)	oui	$0,26^{***}$ (0,02)	

Note : Δ désigne les différences premières. 3 étoiles indiquent une significativité du paramètre à 1% ; 1 étoile indique une significativité à 15%. Les écart-types sont indiquées entre parenthèses. Pour les doubles moindres carrés (indiqués "2SLS" dans le tableau ; les moindres carrés ordinaires sont indiqués "OLS") les régressions auxiliaires consistent à régresser : 1) le logarithme du prix de vente du produit sur le logarithme de l'indice des prix de gros alimentaires pour le lait de consommation ; 2) le logarithme de la dépense sur le logarithme de l'indice des salaires. Quand la régression est en différences, l'équation auxiliaire est également traitée en différences. Environ 34000 observations contribuent au calcul. AUX1 est le coefficient du logarithme du prix du lait dans la régression auxiliaire du logarithme du prix ; AUX2 est le coefficient de l'indicatrice de fin de mois dans la régression auxiliaire du logarithme de la dépense. (^{2}) : régression sans instrumentation de la dépense

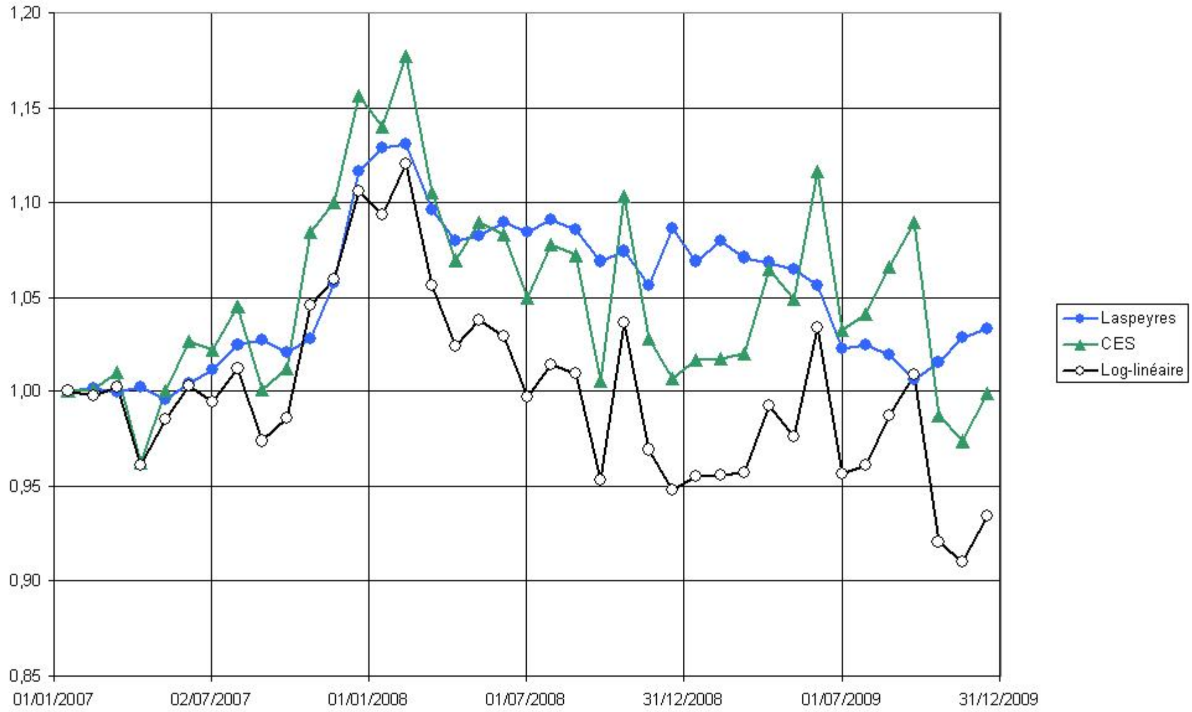
FIGURE 3 – Distribution des coefficients γ_i figurant dans l'équation 13 (multiplicateur constant de la demande log-linéaire)



Note : Les coefficients correspondent aux effets fixes de la regression 2SLS en niveau avec régression auxiliaire pour les prix, la dépense étant considérée comme exogène. Estimation par la méthode des noyaux.

disparaître ponctuellement. Ceci vaut également pour les promotions des fabricants dont la durée de présence dans les données de caisse est relativement brève.

FIGURE 4 – Tracé des indices mensuels de Laspeyres, CES et log-linéaire entre 2007 et 2010



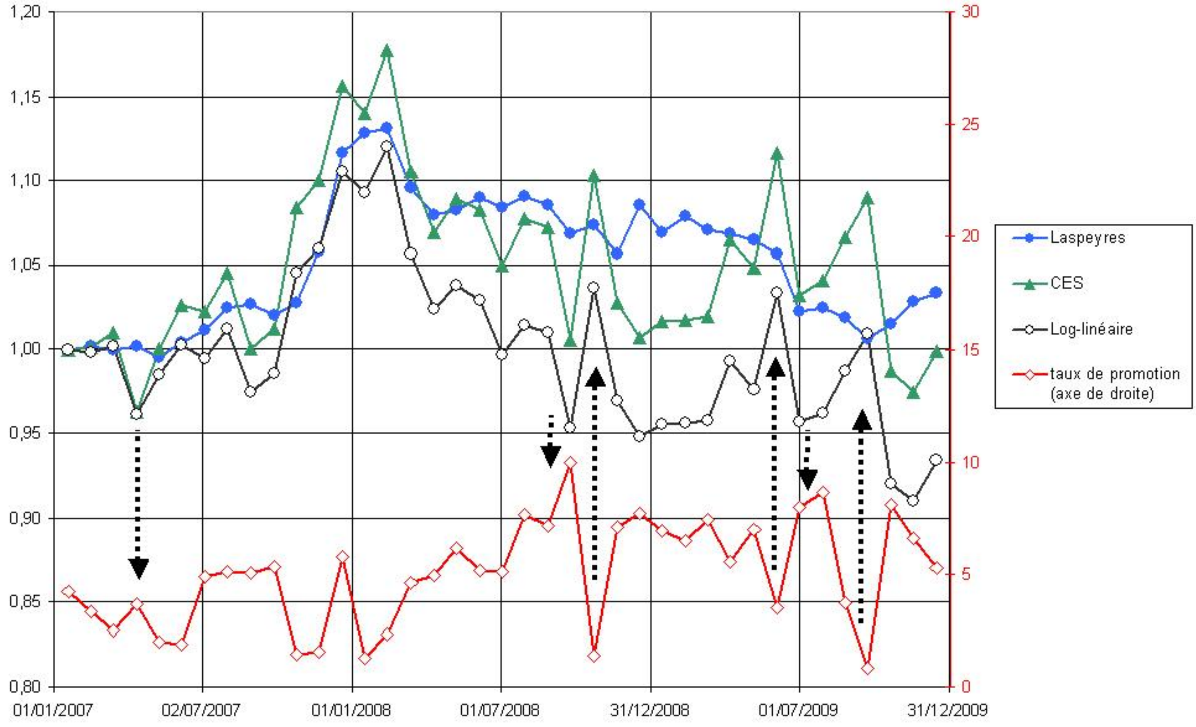
Note : L'indice de Laspeyres est chaîné annuellement et à panier fixe mis à jour annuellement.

Afin de mieux comprendre les différences entre ces indices, la figure 5 met en regard les indices avec les taux de produits en promotion observé chaque mois (pondération en nombre de références d'articles disponibles à la vente). On constate assez nettement la croissance des prix vue par les indices CES et log-linéaire lorsque le taux de promotion est bas, tandis que les prix vus par ces mêmes indices décroissent lorsque le taux de promotion est élevé. Cette mise en regard confirme que l'écart entre l'indice de Laspeyres et l'indice CES est principalement lié aux promotions qui ne sont pas prises en compte dans le premier, tandis qu'elles le sont dans le second.

En dehors des effets promotionnels, on peut noter la bonne cohérence des indices de Laspeyres chaîné et de celui fondé sur l'utilité CES. En revanche, l'indice fondé sur la demande log-linéaire témoigne d'une croissance des prix plus faible. Si on examine plus avant la formule d'indice (10), on peut montrer, sous diverses hypothèses de petitesse de certains termes (voir annexe), que l'indice est approximativement égal à :

$$\begin{aligned} I_{LogLin}^{t',t} \simeq\ & 1 + \sum_{i \in \mathring{I}} \omega_{it} \frac{p_{it'} - p_{it}}{p_{it}} + \\ & \frac{1}{1+\alpha} \left(\sum_{i \in \mathring{I}} \omega_{it} - \sum_{i \in \mathring{I}} \omega_{it'} \right) + \frac{1-\beta}{1+\alpha} \frac{R_{t'} - R_t}{R_t} \left(1 - \sum_{i \in \mathring{I}} \omega_{it} \right) \end{aligned} \tag{14}$$

Figure 5 – Tracé des indices mensuels de Laspeyeres, CES et log-linéaire et taux de produits en promotion



Note : Les flèches indiquent des mois avec des taux de promotion élevés ou faibles selon le cas qui se traduisent par des prix (resp.) bas ou élevés dans les indices CES et log-linéaire.

où $\overset{\circ}{I}$ désigne l'ensemble des produits présents à toutes les dates d'observation, ω_{it} est le poids[4] du produit i dans la dépense à la date t, cette dernière étant notée R_t. Le premier terme de cette expression est la forme de l'indice de Laspeyres restreint aux seuls biens présents sur aux période t (base) et t' (en pratique, nous prenons, pour définir $\overset{\circ}{I}$, l'ensemble des biens présents à toutes les dates). Empiriquement, l'indice de prix correspondant, reproduit à la figure 6, témoigne d'une variation positive des prix sur la période d'observation. Le premier terme de l'expression (14) est donc positif. La trajectoire de l'indice des prix des produits présents à toutes les dates est semblable à celle de l'indice de Laspeyres chaîné présenté à la figure 4 se rapportant à l'ensemble des biens.

Les deux autres termes de l'expression (14) sont liés, d'une part aux variations du poids, dans la dépense contemporaine, des produits présents à toutes les dates d'observation et d'autre part, aux variations de la dépense totale. Empiriquement, le poids des produits présents à toutes les dates reste relativement stable sur la période d'observation (figure 7). On peut toutefois noter une variabilité accrue à partir de l'année 2009 en raison, notamment, d'une plus grande sensibilité du poids au taux de promotion, traduisant vraisemblablement des substitutions plus fréquentes. La baisse du poids des produits présents à toutes les dates se manifeste surtout au dernier trimestre 2009. Le deuxième terme de l'expression (14) est donc empiriquement négatif.

On observe (voir figure 1) enfin une décroissance tendancielle de la dépense sur l'ensemble de la période observée. Comme $\alpha < -1$ et $\beta > 1$, le dernier terme de l'expression (14) est négatif.

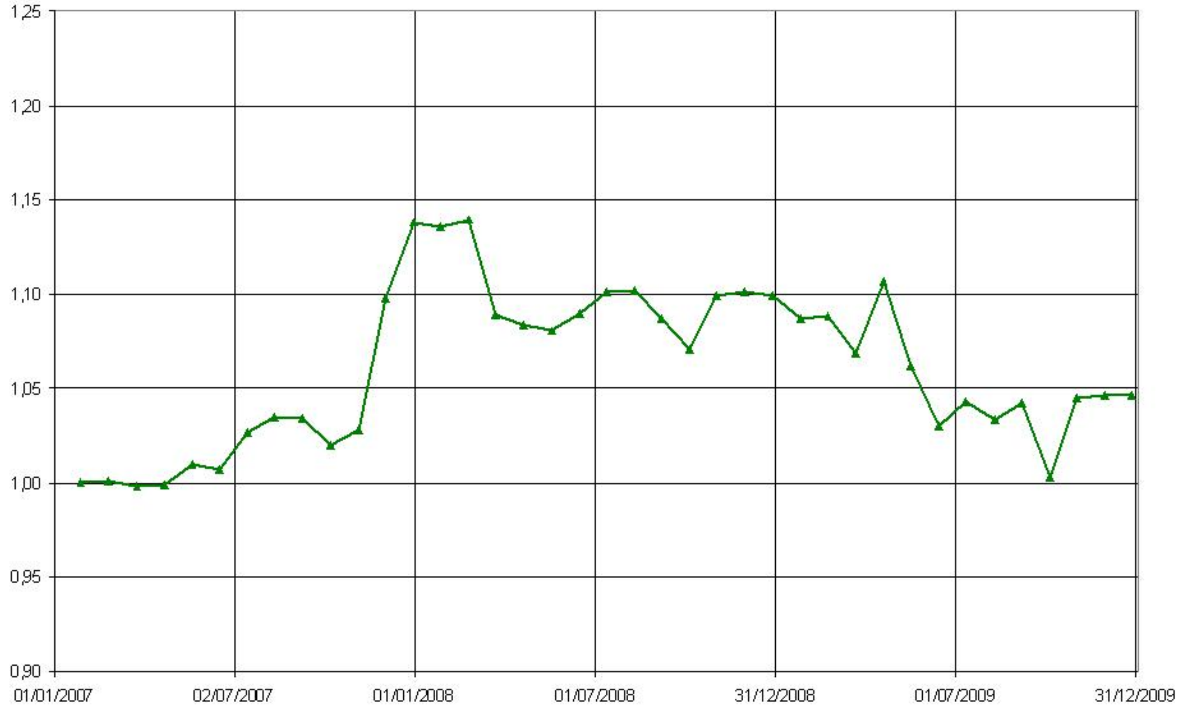
Du point de vue économique, le deuxième terme de l'expression (14) traduit l'effet des substitutions entre produits sur l'indice des prix, tandis que le dernier terme est sensible à l'évolution d'ensemble de la dépense. Lorsque les utilités sont homothétiques, ce dernier terme disparaît. En l'espèce, étant données les valeurs estimées des paramètres α et β, une baisse de la dépense est associée à un gain d'utilité ; dans ce contexte, les yaourths s'apparentent à un bien inférieur. L'application numérique proposée en annexe à partir des données observées montre que les effets de substitution sont environ deux fois plus importants que ceux liés à la baisse tendancielle de la dépense.

4 Conclusion

Nous avons montré trois approches différentes de construction d'un indice de prix. La première, l'approche classique, se fonde essentiellement sur des propriétés *ad hoc* qu'un indice de prix devrait voir vérifiées, ces propriétés étant héritées des propriétés vérifiées par une séquence de prix successifs d'un produit donné [8]. Cette approche, dite approche axiomatique [20], est celle qui prévaut aujourd'hui dans la construction des indices de prix à la consommation "officiels". La deuxième approche consiste à dériver la formule d'indice de l'adoption d'une fonction d'utilité pour un consommateur représentatif et estimer les paramètres de cette fonction à partir de l'observation des demandes (quantités et prix). La troisième approche repose sur l'observation et la paramétrisation des fonctions de demande à partir desquelles on infère un indice à utilité constante. Les deux dernières approches ne sont envisageables qu'à partir de données de caisses. Depuis de nombreuses années les économistes examinent les différences entre les indices fondées sur l'approche axiomatique, d'usage aisé et faciles à expliquer, et les indices à utilité constante,

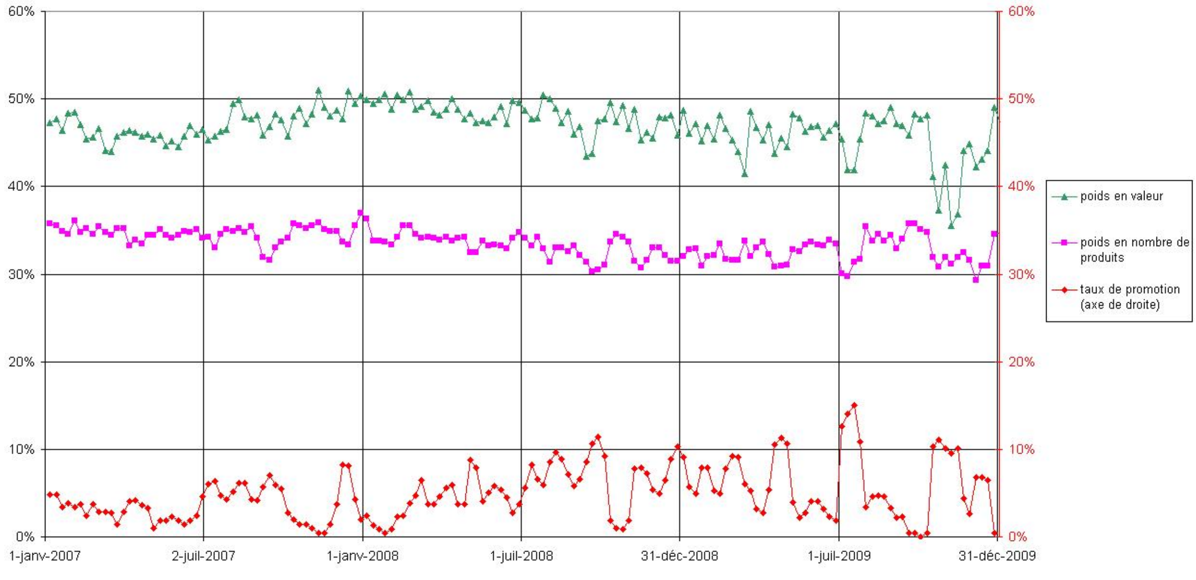
4. À tout instant t, $\sum_i \omega_{it} = 1$.

FIGURE 6 – Indices de prix de Laspeyeres des produits présents à toutes les dates



Note : Il y a environ 200 références de produits présentes chaque semaine d'observation. Les produits présents sur l'ensemble de la période 2007-2009 représentent approximativement 35% des produits (voir figure 7). Indice à base fixe sur les quatre premières semaines d'observation.

FIGURE 7 – Poids en dépense totale des produits présents à toutes les dates



satisfaisants sur un plan économique mais qui nécessitent des éléments rarement disponibles, notamment sur les fonctions de demande, pour pouvoir être calculés. Si le lien entre les deux types d'indice est connu de longue date [6], la disponibilité récente des données de caisse permet réellement d'envisager de mettre en pratique le concept d'indice à utilité constante. Les premiers résultats obtenus montrent que la mécanique de calcul peut s'en trouver simplifiée, notamment s'agissant des remplacements de produits qu'il n'est pas indispensable de traiter de manière spécifique dans le contexte des indices à utilité constante, tandis qu'ils doivent faire l'objet d'une attention particulière dans les calculs d'indices classiques. Au-delà, beaucoup reste à faire pour mesurer en détail les conséquences de l'estimation économétrique de paramètres d'indice sur l'indice calculé. Idéalement, il conviendrait aussi de s'affranchir au maximum d'hypothèses pas toujours justifiées sur la spécification de telle ou telle fonction d'utilité ou de demande en se rapprochant d'une exploitation non paramétrique des données de caisse.

Références

[1] Arrow, K. J., Chenery, H. B., Minhas, B. S., and Solow, R. M. Capital-labor substitution and economic efficiency. *The Review of Economics and Statistics 43*, 3 (1961), 225–250.

[2] Balk, B. On Curing the CPI's Substitution and New Goods Bias. Discussion paper, The Ottawa Group, 1995, 1999.

[3] Boskin, M. J., Dulberger, E. R., Gordon, R. J., Griliches, Z., and W.Jorgenson, D. Consumer Prices, the Consumer Price index, and the Cost of Living. *Journal of Economic Perspectives 12*, 1 (1998), 3–26.

[4] Clerc, E., and Coudin, E. L'IPC, miroir de l'évolution du coût de la vie en France ? Ce qu'apporte l'analyse des courbes d'Engel. *Économie et Statistique 433–434* (2010), 77–105.

[5] Deaton, A., and Muelbauer, J. *Economics and consumer behavior.* Cambridge University Press, 1980.

[6] Diewert, E. Aximoatic and economic approaches to elementary price indexes. Working paper 5104, NBER, 1995.

[7] Diewert, E. Index Number Issues in the Consumer Price Index. *Journal of Economic Perspectives 12*, 1 (1998), 47–58.

[8] Eichhorn, W., and Voeller, J. *Theory of the Price Index.* Springer-Verlag, 1976.

[9] Feenstra, R. C., and Shapiro, M. D. *High-Frequency Substitution and the Measurement of Price Indexes.* University of Chicago Press, 2003, pp. 123–146.

[10] Guédès, D. Impact des ajustement de qualité dans le calcul de l'indice des prix à la consommation. Document de travail F0404, INSEE, 2004.

[11] Hamming, R. *Digital Filters*, 3 ed. Dover, 1997.

[12] Hausman, J. Exact consumer's surplus and deadweight loss. *The American Economic Review 71*, 4 (1981), 662–676.

[13] Hausman, J. Exact Consumer's Surplus and Deadweight Loss. *American Economic Review 71*, 4 (1981), 662–676.

[14] Hausman, J. Sources of Bias and Solutions to Bias ind the Consumer Price Index. *Journal of Economic Perspectives 17*, 1 (2003), 23–44.

[15] Hausman, J., and Leibtag, E. Consumer benefits from increased competition in shopping outlets : measuring the effect of WAL-MART. *Journal of Applied Econometrics 22* (2007), 1157–1177.

[16] Hausman, J. A., and Newey, W. K. Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss. *Econometrica 63*, 6 (1995), 1445–1476.

[17] Hausman, J. A., Pakes, A., and Rosston, G. L. Valuing the Effect of Regulation on New Services in Telecommunications. *Brookings Papers on Economic Activity, Microeconomics* (1997), 1–54.

[18] Herpin, N., and Verger, D. *Consommation et modes de vie en France*, 3 ed. La Découverte, 2008.

[19] Hurwicz, L., and Uzawa, H. On the integrability of demand functions. In *Preferences, Utility, and Demand* (1971), J. S. Chipman, L. Hurwicz, M. K. Richter, and H. F. Sonnenschein, Eds., Harcourt, pp. 114–148.

[20] ILO, IMF, OECD, UNECE, Eurostat, and The World Bank. *Consumer price index manual : Theory and Practise.* International Labour Office, 2004.

[21] Magnien, F., and Pougeard, J. Les indices à utilité constante : une référence pour mesurer l'évolution des prix. *Économie et Statistique 335* (2000), 81 – 94.

[22] Mesler, D. Accounting for the effects of new and disappearing goods using scanner data. *Review of Income and Wealth 52*, 4 (2006), 547–568.

[23] Moulton, B. R., and Stewart, K. J. An Overview of Experimental U.S. Consumer Price Indexes. *Journal of Business & Economic Statistics 17*, 2 (1999), 141–151.

[24] Saglio, A. Comparative changes in average price and a prix index : two case studies. Discussion paper, The Ottawa Group, 1994, 1994.

[25] Varian, H. The Nonparametric Approach to Demand Analysis. *Econometrica 50*, 4 (1982), 945–973.

[26] Varian, H. R. *Microeconomic Analysis.* The MIT Press, 1975.

[27] Varian, H. R. Trois évaluation de l'impact "social" d'un changement de prix. Cahiers du séminaire d'économétrie, 1982.

[28] Viglino, L. Le concept unificateur des indices de prix et proposition d'un nouvel indice. In *Actes des journées de méthodologie statistique* (2000), INSEE.

Annexe : développements calculatoires

Résolution du système d'équations différentielles (6) pour les fonctions de demande (8)

Nous avons donc à résoudre en μ le système d'équations différentielles suivant :

$$1 \leqslant i \leqslant n \ , \ \frac{\partial \mu}{\partial p_i} = p_i^{\alpha_i} \mu^{\beta} e^{\gamma_i}$$

Pour i, nous avons en particulier $\frac{\partial \mu}{\partial p_i} = p_i^{\alpha_i} \mu^\beta e^{\gamma_i}$ qui se résout, partiellement, en intégrant sur p_i :

$$\frac{\mu^{1-\beta}}{1-\beta} = \frac{p_i^{1+\alpha_i}}{1+\alpha_i} e^{\gamma_i} + C_i(\mathbf{p}_{(i)}; \mathbf{q}, R)$$

où C_i est une fonction dépendant de $\mathbf{p}_{(i)}$ (vecteur des prix $\mathbf{p}$ privé de sa $i^{\text{ème}}$ composante), de $\mathbf{q}$ et de R. On peut donc chercher une solution de la forme

$$\mu(\mathbf{p}; \mathbf{q}, R) = \left\{ (1-\beta) \left[\sum_{i=1}^{n} \frac{p_i^{1+\alpha_i}}{1+\alpha_i} e^{\gamma_i} + C(\mathbf{q}, R) \right] \right\}^{\frac{1}{1-\beta}}$$

où C est une fonction dépendant de $\mathbf{q}$ et R. Elle s'obtient en appliquant les conditions limites (voir relation 7) $\mu(\mathbf{q}; \mathbf{q}, R) = R$, soit :

$$C(\mathbf{q}, R) = \frac{R^{1-\beta}}{1-\beta} - \sum_{i=1}^{n} \frac{q_i^{1+\alpha_i}}{1+\alpha_i} e^{\gamma_i}$$

Finalement,

$$\mu(\mathbf{p}; \mathbf{q}, R) = \left\{ (1-\beta) \left[\frac{R^{1-\beta}}{1-\beta} + \sum_{i=1}^{n} e^{\gamma_i} . \frac{p_i^{1+\alpha_i} - q_i^{1+\alpha_i}}{1+\alpha_i} \right] \right\}^{\frac{1}{1-\beta}}$$

Approximation de l'indice log-linéaire (14)

La démonstration est fondée sur l'hypothèse que $p_{it} - p_{it'}$, $R_t - R_t'$, $\omega_{it} - \omega_{it'}$ et le complément à 1 de l'indice $I_{LogLin}^{t',t}$ sont des infiniment petits du premier ordre. On se place ici dans le cas où les α_i sont tous égaux à un réel unique α (cadre retenu pour l'économétrie). Partant de l'expression de (10), au premier ordre, nous avons :

$$I_{LogLin}^{t',t} \simeq 1 + R_t^{\beta-1} \sum_{i=1}^{n} e^{\gamma_i} \frac{p_{it'}^{1+\alpha} - p_{it}^{1+\alpha}}{1+\alpha}$$

Si on note I^t l'ensemble des identifiants de produits présents à la date t, il est possible de définir l'intersection $\overset{\circ}{I}$ de ces ensembles pour toutes les dates d'observation. Ainsi défini, $\overset{\circ}{I}$ correspond à l'ensemble des indentifiants de produits présents à toutes les dates d'observation. À tout instant t, on note $\bar{I}^t$ le complément de $\overset{\circ}{I}$ dans I^t. L'expression précédente peut alors s'écrire :

$$I_{LogLin}^{t',t} \simeq 1 + R_t^{\beta-1} \sum_{i \in \overset{\circ}{I}} e^{\gamma_i} \frac{p_{it'}^{1+\alpha} - p_{it}^{1+\alpha}}{1+\alpha} + R_t^{\beta-1} \sum_{i \in \bar{I}^{t'}} e^{\gamma_i} \frac{p_{it'}^{1+\alpha}}{1+\alpha} - R_t^{\beta-1} \sum_{i \in \bar{I}^{t}} e^{\gamma_i} \frac{p_{it}^{1+\alpha}}{1+\alpha}$$

Puis, à tout instant t, la contrainte budgétaire implique $\sum p_{it} x_{it} = R_t$. En tenant compte de l'expression de la demande (relation 8), il découle que $\omega_{it} = R_t^{\beta-1} p_{it}^{1+\alpha} e^{\gamma_i}$ est le poids du bien i dans la dépense R_t de la date t.

Pour $i \in \overset{\circ}{I}$, par un développement de Taylor au premier ordre en $p_{it'} - p_{it}$, nous avons :

$$p_{it}^{1+\alpha} - p_{it}^{1+\alpha} \simeq (1+\alpha) p_{it}^{\alpha} (p_{it'} - p_{it})$$

Il en découle que :

$$R_t^{\beta-1}\sum_{i\in\overset{\circ}{I}} e^{\gamma_i}\frac{p_{it'}^{1+\alpha}-p_{it}^{1+\alpha}}{1+\alpha} \simeq \sum_{i\in\overset{\circ}{I}}\omega_{it}\frac{p_{it'}-p_{it}}{p_{it}}$$

Puis, les autres termes se simplifient en utilisant la contrainte budgétaire et le fait que les poids se somment à 1 pour une date t donnée. Ainsi :

$$R_t^{\beta-1}\sum_{i\in\bar{I}^t} e^{\gamma_i}\frac{p_{it}^{1+\alpha}}{1+\alpha} = \frac{1}{1+\alpha}\left(1-\sum_{i\in\overset{\circ}{I}}\omega_{it}\right)$$

De même,

$$R_t^{\beta-1}\sum_{i\in\bar{I}^t} e^{\gamma_i}\frac{p_{it'}^{1+\alpha}}{1+\alpha} = \frac{1}{1+\alpha}\left(\frac{R_t}{R_{t'}}\right)^{\beta-1}\left(1-\sum_{i\in\overset{\circ}{I}}\omega_{it'}\right)$$

Cette dernière expression peut être simplifiée en s'appuyant sur les hypothèses de petitesse des différences $R_{t'}-R_t$ et $\omega_{it}-\omega_{it'}$. En effet,

$$\frac{1}{1+\alpha}\left(\frac{R_t}{R_{t'}}\right)^{\beta-1}\left(1-\sum_{i\in\overset{\circ}{I}}\omega_{it'}\right) = \frac{1}{1+\alpha}\left(1+\frac{R_{t'}-R_t}{R_t}\right)^{1-\beta}\left(1-\sum_{i\in\overset{\circ}{I}}\omega_{it}+\sum_{i\in\overset{\circ}{I}}(\omega_{it}-\omega_{it'})\right)$$

En développant au premier ordre et en sommant le résultat avec les deux autres termes déjà explicités, nous avons :

$$\begin{aligned} I_{LogLin}^{t',t} &\simeq 1+\sum_{i\in\overset{\circ}{I}}\omega_{it}\frac{p_{it'}-p_{it}}{p_{it}}+ \\ &\quad \frac{1}{1+\alpha}\left(\sum_{i\in\overset{\circ}{I}}\omega_{it}-\sum_{i\in\overset{\circ}{I}}\omega_{it'}\right)+\frac{1-\beta}{1+\alpha}\frac{R_{t'}-R_t}{R_t}\left(1-\sum_{i\in\overset{\circ}{I}}\omega_{it}\right)\end{aligned}$$

<u>Application numérique</u> : nous adoptons les valeurs numériques suivantes

- $\sum_{i\in\overset{\circ}{I}}\omega_{it}=0,45$
- $\frac{1}{\sum_{i\in\overset{\circ}{I}}\omega_{it}}\sum_{i\in\overset{\circ}{I}}\omega_{it}\frac{p_{it'}-p_{it}}{p_{it}}=0,05$
- $\alpha=-1,86$
- $\beta=1,08$
- $\sum_{i\in\overset{\circ}{I}}(\omega_{it}-\omega_{it'})=0,46-0,44=0,02$ (différence des moyennes du premier et dernier semestres d'observation)

Avec ces valeurs,

$$\begin{aligned} I_{LogLin}^{2009S2,2007S1} &\simeq 1+\underbrace{0,45\times 0,05}_{0{,}02}-\underbrace{\frac{1}{0,86}\times 0,02}_{0{,}02}-\underbrace{\frac{0,08}{0,86}\times 0,2\times(1-0,45)}_{0{,}01} \\ &\simeq 1-0,01\end{aligned}$$

On observe, comme prévu, que l'indice $I_{LogLin}^{t',2007}$ est négatif au dernier semestre 2009. Ce calcul numérique nous permet également d'observer les contributions : si l'indice des prix des biens présents à toutes les dates contribue pour 2 points à la hausse, la baisse de la dépense d'ensemble (un point à la baisse) et surtout, les substitutions (2 points à la baisse) font plus que compenser l'évolution à la hausse des prix.