

Distr.  
GÉNÉRALE

CES/AC.49/2001/8  
16 août 2001

FRANÇAIS  
Original: ANGLAIS

**COMMISSION DE STATISTIQUE et  
COMMISSION ÉCONOMIQUE POUR  
L'EUROPE**

**ORGANISATION INTERNATIONALE  
DU TRAVAIL (OIT)**

**CONFÉRENCE DES STATISTICIENS  
EUROPÉENS**

**Réunion commune CEE/OIT  
sur les indices des prix à la consommation  
(Genève, 1<sup>er</sup> et 2 novembre 2001)**

**L'ÉCHANTILLONNAGE DANS LES INDICES DES PRIX À  
LA CONSOMMATION: COMMENT LES DONNÉES SAISIES  
PAR LECTURE OPTIQUE POURRAIENT Y CONTRIBUER**

Communication sollicitée de l'Office for National Statistics du Royaume-Uni\*

**Résumé**

La qualité d'un indice des prix à la consommation dépend dans une très large mesure de la qualité des données de base, et en particulier de la représentativité tant de l'échantillon de commerces de détail utilisé pour observer l'évolution des prix que du choix des articles dont les prix sont collectés. Le présent document étudie les possibilités d'améliorer la qualité d'un indice des prix en prenant comme référence des données saisies par lecture optique pour vérifier la représentativité de l'échantillon obtenu, contrôler le choix initial des composants de l'échantillon et corriger a posteriori les défauts de cet échantillon. Le document commence par examiner les fondements sur lesquels repose la sélection des éléments de l'échantillon, ainsi que les options s'offrant dans la pratique à la personne chargée d'établir un indice et les problèmes qui peuvent s'ensuivre. Il passe en revue les méthodes actuelles d'échantillonnage utilisées pour le calcul de l'indice des prix de détail (IPD), en comparant les principes sur lesquels cet indice est fondé avec

---

\* Document établi par David Fenwick et Adrian Ball, Office for National Statistics du Royaume-Uni, 1 Drummond Gate, Londres SW1V 2QQ, Royaume-Uni. Adresse électronique: [david.fenwick@ons.gov.uk](mailto:david.fenwick@ons.gov.uk) et [adrian.ball@ons.gov.uk](mailto:adrian.ball@ons.gov.uk).

ceux qui régissent la compilation des données saisies par lecture optique. Les auteurs envisagent ensuite les problèmes concrets que pourrait susciter l'utilisation éventuelle de données saisies par lecture optique pour améliorer les méthodes actuelles d'échantillonnage. Le document traite séparément de deux aspects des méthodes d'échantillonnage employées dans le cadre de l'IPD: le choix des articles et le choix des points de vente. S'agissant du choix des articles, le document se concentre sur les biens de consommation durables qui sont ceux pour lesquels le maintien de la représentativité s'avère, depuis toujours, le plus épineux. Des données saisies par lecture optique sont utilisées pour procéder à une analyse comparative de l'échantillon actuel de biens de consommation durables présentant un taux de substitution élevé, par exemple les télévisions, en mettant en lumière les différences et en proposant des solutions basées sur un contrôle de l'échantillon par le biais de quotas. Les auteurs envisagent en outre l'utilisation potentielle de données saisies par lecture optique pour choisir des articles de substitution lorsque la variété d'origine a disparu des magasins et lorsqu'il se produit un changement de qualité connexe. Le moment où s'effectue le choix et la question du renouvellement des articles sont également pris en considération. En ce qui concerne le choix des points de vente, le document met l'accent sur les écarts de prix qui peuvent exister entre différents types de commerces, et souligne les avantages qu'apporterait l'adoption d'une stratification affinée sur le plan de la représentativité. Le document s'achève par la présentation de lignes directrices et de méthodes de contrôle de la qualité améliorées pour le relevé des prix.

**Mots clefs:** données scannées, points de vente, articles, produits anciens et nouveaux, stratification, échantillonnage aléatoire et échantillonnage par choix raisonné, représentativité, modélisation, repondération, rééchantillonnage, quotas, analyse comparative, contrôle de la qualité, lignes directrices.

## **I. Introduction**

1. Un certain nombre d'études antérieures ont fait état de la possibilité d'utiliser des données saisies par lecture optique (ci-après dénommées «données scannées») pour l'élaboration des indices des prix à la consommation, soit comme source directe de données sur les prix, soit comme instrument d'estimation des ajustements qui s'imposent pour tenir compte des changements de qualité lorsqu'un article est remplacé par un autre, et que les caractéristiques des articles dont les prix sont observés se modifient. Par ailleurs, il semblerait que les données scannées pourraient contribuer à améliorer l'efficacité des méthodes traditionnelles d'échantillonnage probabiliste.
2. Les avantages que pourrait procurer l'utilisation de données scannées ne sont pas négligeables, notamment si les régressions hédoniques et les données scannées sont utilisées pour compléter les pratiques actuelles en vue de mieux répondre, grâce à une approche intégrée, aux exigences tant de la représentativité que de l'ajustement de qualité.
3. C'est dans ce contexte qu'un projet commun de recherche a été mis sur pied entre l'ONS et la Cardiff Business School, Cardiff University, en vue, à la fois, d'étudier les possibilités d'utiliser des données scannées comme instrument de diagnostic d'éventuelles insuffisances dans la collecte de données au titre de l'IPD et d'y apporter des solutions. Le présent document rend compte des résultats des travaux accomplis à ce jour.

## II. Renseignements d'ordre général: population cible de l'IPD et méthodes d'échantillonnage

### Population cible

4. L'IPD est une mesure moyenne de l'évolution des prix des biens et services achetés à des fins de consommation par la grande majorité des ménages du Royaume-Uni. La population de référence comprend l'ensemble des ménages ordinaires à l'exception des catégories suivantes: a) ménages de retraités dont trois quart au moins du revenu total proviennent de pensions et d'allocations versées par l'État et b) les «ménages à haut revenu» dont le revenu total se situe dans les 4 % supérieurs de l'ensemble des catégories de ménages. Les postes de dépenses de référence sont les achats de biens et services effectués par la population de référence pour sa consommation. Les prix retenus aux fins du calcul de l'indice devraient refléter les prix au comptant généralement payés par la population de référence pour les biens et services en question. L'indice est calculé essentiellement au coût d'acquisition, c'est-à-dire sur la base de la valeur totale des biens et services achetés durant une période considérée, qu'ils soient ou non intégralement payés au cours de la période en question. La principale exception concerne les logements occupés par leurs propriétaires pour lesquels on s'appuie sur le coût d'utilisation.

### Jour de référence pour les prix

Le jour de référence pour les prix est le deuxième ou le troisième mardi du mois.

### Principes généraux régissant l'échantillonnage et la collecte des prix

5. L'Office for National Statistics emploie actuellement une méthode traditionnelle d'échantillonnage selon laquelle les prix sont relevés localement auprès de magasins individuels, et centralement, en se servant des tarifs nationaux des services publics de distribution ou des chiffres communiqués par les sièges de chaînes de magasins de vente au détail appliquant une politique centralisée des prix. La principale difficulté que l'on rencontre à cet égard est l'absence de base de sondage appropriée correspondant à l'univers ciblé sur le plan géographique et pour ce qui est des points de vente, de la ligne de produits et des différents articles considérés. Il en résulte que les instituts nationaux de statistique sont souvent contraints soit d'établir leurs propres bases de sondage et méthodes d'échantillonnage aléatoire, soit de recourir à un système de sondage procédant d'un choix raisonné. Bien entendu, ces méthodes doivent satisfaire aux critères de représentativité dans le temps. On considère généralement que ce dernier point pose moins de problèmes que la représentativité sur le plan géographique ou en termes de points de vente, de ligne de produits ou d'articles, en tous cas dans le contexte de la période de référence des prix. Il convient de noter à ce propos que le jour de référence pour les prix aux fins du calcul de l'IPD a été choisi en fonction des résultats d'une étude sur les habitudes d'achat. Cette étude a abouti à la conclusion qu'un mardi du milieu du mois serait vraisemblablement le jour le plus représentatif. Cependant, la dimension chronologique comporte un autre élément, à savoir l'amointrissement de la représentativité de l'échantillon à mesure que le panier «type» perd de son actualité du fait que de nouveaux produits sont introduits et que les consommateurs remplacent les commerces qu'ils fréquentaient et leurs achats antérieurs par d'autres magasins, articles ou variétés. Par conséquent, la dimension chronologique est présente dans tous les aspects du processus d'échantillonnage se rapportant à un indice des prix à la consommation.

### Méthodes d'échantillonnage utilisées pour les relevés des prix sur place

6. La méthode actuellement suivie pour la sélection des lieux de relevé des prix locaux a été introduite en 2000. Elle vise à donner à chaque localité où sont effectués des achats au Royaume-Uni une probabilité d'être sélectionnée pour le relevé des prix égale à la proportion des dépenses totales de consommation qui lui correspond. À cette fin, on utilise une base de sondage hiérarchique à deux degrés fondée sur les régions géographiques. Un total de 141 localités sont nécessaires pour les relevés des prix et le nombre à sélectionner dans chacune des régions est déterminé en fonction de la part des dépenses totales du Royaume-Uni qu'absorbe la région. Il s'agit là du premier degré du processus d'échantillonnage basé sur des informations émanant des enquêtes sur les dépenses des ménages. Dans chacune des régions, les localités sont sélectionnées selon une probabilité proportionnelle à leur taille, en prenant le nombre d'employés du secteur du commerce du détail comme critère d'appréciation approximative du volume des dépenses. Des considérations pratiques amènent à modifier de deux manières ce principe fondamental. Premièrement, du fait qu'il n'est pas rentable de relever les prix dans des zones trop restreintes pour que l'on y trouve une proportion raisonnable de la liste complète des articles, nous faisons abstraction des localités comportant moins de 250 points de vente. Deuxièmement, pour des raisons analogues liées à la rentabilité, les centres commerciaux de la périphérie où s'effectue une large part des dépenses, mais où il n'est pas possible de trouver tous les articles, sont appariés avec des lieux d'achat plus petits situés à proximité où l'on peut se procurer le reste des articles. Ce lieu d'achat composite est alors considéré comme une localité unique dans l'échantillonnage probabiliste.

7. Dans chaque localité sélectionnée les agents chargés de relever les prix (ci-après dénommés «enquêteurs») procèdent à une énumération pour constituer une base de sondage à partir de laquelle les points de vente sont sélectionnés au hasard, en séparant les commerces de détail à succursales multiples des commerces indépendants. Ce processus de sélection s'effectue sur la base d'un roulement, de sorte que l'ensemble des unités composant l'échantillon sont renouvelées tous les cinq ans.

8. Contrairement à l'échantillonnage des points de vente, le choix des articles représentatifs à prendre en compte dans le calcul de l'IPD se fait à dessein (c'est-à-dire par choix raisonné et non pas de manière aléatoire). Toutes les catégories de dépenses, auxquelles les ménages consacrent des sommes importantes d'après les résultats de l'enquête sur les dépenses des ménages, sont réparties en 80 rubriques environ et on choisit des articles jugés représentatifs pour chaque rubrique. Le nombre d'articles retenus dans chaque rubrique dépend à la fois du coefficient de pondération attribué à la rubrique en question et de la variabilité des prix des articles qu'elle comprend. Près de 650 articles représentatifs sont choisis au niveau central par des spécialistes des biens et services marchands et ce choix d'articles est réexaminé chaque année en janvier pour veiller à ce qu'il demeure représentatif de la rubrique considérée. Des articles nouveaux sont choisis pour représenter des catégories de dépenses nouvelles ou en expansion, ou pour réduire l'instabilité des agrégats de niveau plus élevé. D'autres articles sont éliminés si l'on constate que les dépenses qui y sont consacrées deviennent insignifiantes. Les décisions sont fondées sur des informations tirées d'études de marché, des journaux ou des revues spécialisées ou sur des données recueillies par les enquêteurs sur le terrain. Ainsi, le panier est tenu à jour mais ce processus ne garantit pas à lui seul la représentativité de l'échantillon. Les descriptions sont générales plutôt que normatives de sorte qu'il incombe à l'enquêteur de choisir le produit précis ou la variété spécifique dont le prix va être relevé.

9. Le choix par l'enquêteur des produits et variétés qui représenteront les articles sélectionnés s'effectue aussi par choix raisonné et sur le terrain. Les enquêteurs ont pour instruction de choisir, dans le magasin sélectionné, le produit ou la variété qui représente le mieux les ventes de l'article en question dans ce magasin. Dans la pratique, l'enquêteur se fait généralement aider par le commerçant auquel il demande quel est le produit ou la variété qu'il vend le mieux. Dans la plupart des cas, c'est ce produit ou cette variété qui est choisi comme article représentatif aux fins du relevé des prix. Cette méthode d'échantillonnage basée sur les ventes effectives a l'avantage d'accroître la dimension de l'échantillon obtenu en réglant le problème lié au fait que certains magasins peuvent ne pas avoir en stock un produit ou une variété déterminés. Par ailleurs, elle permet théoriquement d'inclure dans l'échantillon une plus large gamme de produits et de variétés que si l'on utilisait une description très précise.

#### Méthodes d'échantillonnage utilisées pour les prix relevés au niveau central et les prix obtenus par téléphone

10. Parfois, les prix font l'objet d'une collecte centralisée, moins coûteuse que l'envoi d'enquêteurs sur place. On a recours à une collecte centralisée des prix dans deux circonstances bien distinctes:

- ?? Dans le cas des **magasins «centralisés»** où, pour des raisons d'économie, on relève les prix directement auprès du siège des magasins à succursales multiples pratiquant des politiques des prix nationales. Ces prix sont ensuite combinés avec des prix relevés localement auprès d'autres points de vente, proportionnellement au nombre de points de vente initialement choisis dans les localités sélectionnées;
- ?? Dans le cas des **articles «centralisés»** pour lesquels il existe un nombre limité de fournisseurs et qui en principe ne sont pas achetés auprès de commerces locaux, par exemple le gaz, l'électricité ou l'eau, dont les prix sont extraits des tarifs communiqués directement par les sièges des sociétés de distribution concernées. Ces données sont utilisées pour créer des sous-indices qui sont combinés avec d'autres sous-indices en vue de construire l'IPD d'ensemble.

11. En outre, les prix de certains articles sont relevés par téléphone, les visites personnelles aux commerçants n'étant effectuées qu'occasionnellement pour s'assurer du maintien de la qualité des réponses. Il s'agit, entre autres, des tarifs des électriciens, pour lesquels il n'y a pas de point de vente proprement dit et du prix d'une entrée dans un centre de loisirs qui ne risque guère de susciter une équivoque, de sorte qu'un déplacement pour relever un seul prix se révélerait être une perte de temps. Les prix en question sont combinés, le cas échéant, avec des données recueillies localement.

#### Facteurs critiques

12. Les méthodes d'échantillonnage des localités et des magasins sont, dans l'ensemble, statistiquement rigoureuses et ne posent donc pas beaucoup de problèmes. Par conséquent, on part de l'hypothèse que c'est surtout en rapport avec le choix des articles que des difficultés liées à la non-représentativité peuvent survenir – d'autant que pour certains produits, le taux de renouvellement de l'offre est relativement élevé. Il s'ensuit que la constitution d'un échantillon

représentatif dans le cadre de l'IPD du Royaume-Uni dépend en particulier des éléments suivants:

- ?? Les méthodes d'échantillonnage initial par choix raisonné des articles sur le terrain;
- ?? Les méthodes utilisées pour sélectionner les articles de substitution que l'on est forcé d'introduire lorsque des articles disparaissent des rayons;
- ?? Les méthodes d'actualisation de l'échantillon retenu en vue de refléter le renouvellement général des variétés et produits en vente.

13. C'est dans cet esprit que l'on a entrepris une analyse comparative entre l'échantillon constitué aux fins du calcul de l'IPD, pour une série d'appareils électriques ou de haute technicité, et des données scannées correspondantes, ainsi qu'une comparaison des niveaux de prix et variations des prix relatifs.

14. Avant de présenter les résultats de cette étude, il est utile de se remettre en mémoire les principales caractéristiques des données scannées, étant donné que ces données ne sont pas spécifiquement destinées à l'élaboration des indices des prix à la consommation et posent donc des problèmes qui leur sont propres. Les caractéristiques des données scannées sont passées en revue dans la section suivante du présent document.

### **III. Caractéristiques des données scannées**

15. Les données scannées sont compilées à partir de données recueillies aux points de vente informatisés (appelés «EPOS», sigle d'«electronic point of sale»), c'est-à-dire des données enregistrées par des lecteurs de codes-barres au moment et sur le lieu de l'achat. À mesure qu'un nombre croissant de magasins s'équipent de lecteurs de codes-barres, les données scannées offrent de plus en plus de possibilités de fournir des informations actuelles et précises sur les éléments suivants:

- ?? Le nombre des ventes sur une période déterminée de différentes variétés d'un produit uniquement identifiées par leur code-barre;
- ?? La valeur totale des ventes en question et, par déduction, le «prix» moyen d'une opération;
- ?? Une liste des caractéristiques individuelles des différentes variétés de produits concernées;
- ?? Des caractéristiques géographiques et autres du point de vente.

16. En réalité, le champ actuel des données scannées varie d'un type de magasin et d'un groupe de marchandises à l'autre; par ailleurs, la quantité et le degré de détail des données effectivement disponibles peut varier selon la source commerciale et selon le produit ou groupe de produits considéré. En outre, comme les données scannées constituent un produit dérivé d'un système de comptabilité financière et d'inventaire, elles ne sont pas spécifiquement conçues en fonction des besoins d'un spécialiste des statistiques des prix, ce qui a des incidences sur leur

utilisation pour l'élaboration de l'indice. Premièrement, les définitions employées peuvent ne pas être compatibles avec la définition de l'indice. Par exemple, le «prix» moyen d'une opération, tel qu'enregistré dans les données scannées, tient compte des rabais comme ceux accordés sur des marchandises endommagées qui, en principe, ne sont pas pris en considération dans les indices des prix à la consommation. Deuxièmement, le codage des données peut se présenter sous une forme qui n'est pas immédiatement exploitable ni compatible avec les normes internationales, pour ce qui est, par exemple, de la classification des produits.

17. Par ailleurs, et de manière plus générale, l'expérience nous apprend qu'une grande compétence et beaucoup d'efforts sont nécessaires pour épurer les données scannées, afin d'éliminer, par exemple, les incidences de la réutilisation de codes-barres, et pour les rendre ainsi utilisables à des fins statistiques.

#### Principales différences de définitions entre les données scannées et les données recueillies sur place pour l'établissement de l'indice des prix de détail

18. Les principales différences entre les deux séries de données sont les suivantes:

- ?? Les données de l'IPD couvrent les opérations réalisées dans les commerces de détail par des ménages ordinaires aux fins de leur consommation personnelle. Les données scannées couvrent uniquement les ventes EPOS, généralement complétées par des données d'enquête destinées à prendre en compte les magasins qui n'emploient pas de code-barre. Elles font souvent abstraction des produits de la «propre» marque du magasin, mais tiennent compte des ventes aux clients appartenant au secteur commercial;
- ?? Les données de l'IPD ne tiennent pas compte des ristournes conditionnelles (notamment celles dont les clients en possession de la carte du magasin sont les seuls à pouvoir bénéficier), ni des offres de deux produits pour le prix d'un, des rabais personnels accordés à titre exceptionnel par le directeur du magasin, ou des réductions sur le prix des articles de fin de série ou les marchandises endommagées. Les données scannées, par contre, mesurent les recettes moyennes diminuées des ristournes en général, quelles qu'en soient les modalités; elles prennent en considération les rabais sur les fins de séries ou les marchandises abîmées en magasin et les ristournes y sont comptabilisées en fonction du code-barre plutôt que de la nature de l'opération (par exemple, le coût des cassettes vidéo offertes lors de l'achat d'un magnétoscope viendra en déduction des recettes moyennes provenant de la vente de cassettes vidéo);
- Les données de l'IPD se rapportent à un ensemble fixe de points de vente et ne tiennent donc pas compte des effets de la substitution de points de vente à d'autres. Les données scannées concernent les opérations courantes et prennent donc en considération la substitution des points de vente;
- Au niveau de l'«article», les données sur les prix de l'IPD ne sont pas pondérées tandis que les données scannées reflètent les différentes quantités vendues de chaque modèle ou variété.

19. Bien que l'on ne sache pas quel est l'impact numérique de ces différences, il est clair que cet impact ne sera pas nécessairement constant dans le temps et évoluera selon les conditions du marché et les types de produit.

20. Il faut garder présentes à l'esprit d'autres caractéristiques des deux sources de données, lorsque l'on compare les prix affichés dans les magasins et les données scannées correspondantes, notamment:

- L'erreur d'échantillonnage associée aux enquêtes par sondage, en particulier au niveau de la variété de produit qui est considéré dans le présent document (l'échantillon de l'IPD n'est pas conçu pour fournir des informations fiables pour une ventilation aussi fine). En revanche, les données scannées assurent une couverture totale des segments du commerce de détail pris en considération;
- L'IPD enregistre les prix, un jour du mois bien précis, tandis que les données scannées utilisées dans la présente analyse comparative portent sur un mois entier;
- Les données scannées font la distinction entre différents types de détaillants, par exemple les magasins à succursales multiples et les commerces indépendants; ce n'est pas le cas, en revanche, des données de l'IPD (cette distinction étant superflue dans l'IPD puisque l'échantillon pour le relevé des prix sur place est conçu pour s'autopondérer). On risque donc d'être confronté à un manque d'homogénéité entre les deux sources de données dans les comparaisons si la combinaison des types de points de vente n'est pas la même dans les deux sources et se modifie au fil du temps.

#### **IV. Plan de l'étude**

21. L'étude comportait trois phases:

- L'analyse comparative du choix des produits et des variétés au titre de l'IPD et des données scannées correspondantes. Cette phase comprenait une comparaison des répartitions relatives des ventes, ainsi que des proportions des relevés de prix;
- Une comparaison des prix unitaires moyens et variations de prix dans le cadre de l'IPD avec les valeurs unitaires (c'est-à-dire la production moyenne de recettes) et variations des valeurs unitaires correspondantes telles qu'elles ressortent des données scannées;
- Un examen des options envisageables pour renforcer l'efficacité des techniques d'échantillonnage traditionnelles en utilisant les données scannées dans le cadre des méthodes classiques de collecte de données et pour adopter une approche intégrée de la représentativité et de l'ajustement de qualité.

22. L'étude s'est concentrée sur cinq articles présélectionnés: téléviseurs; machines à laver; aspirateurs; lave-vaisselle et appareils photo. Des travaux connexes ont également été entrepris à partir de la même base de données pour étudier l'application des méthodes de régression hédonique en vue d'effectuer des ajustements de qualité explicites et de mettre en évidence les



caractéristiques clefs des articles à prendre en compte lorsqu'il fallait remplacer des articles ayant disparu des rayons. Les liens qui existent entre la représentativité de l'échantillon et les ajustements de qualité sont apparus de plus en plus clairement au cours des travaux. Nous reviendrons sur ce dernier point dans la dernière partie du présent document.

## V. Représentativité du choix des produits et des variétés

23. Cette phase de l'étude avait pour objet de déterminer la mesure dans laquelle les méthodes de sélection actuelles pourraient amener à choisir, pour le relevé des prix, des échantillons non représentatifs de produits et de variétés. On a examiné les distributions générales obtenues avec les méthodes de sélection utilisées dans le cadre de l'IPD et on les a comparées à celles obtenues avec les données scannées. Des données mensuelles ont été comparées pour la période allant d'août 1999 à octobre 1999. Cette comparaison s'est effectuée au niveau global et on n'a pas procédé à des mises en correspondance individuelles de données.

### Résumé des résultats

24. Dans le tableau 1 ci-après, la distribution par modèle des données scannées sur les prix (colonnes de gauche) montre quels sont les dix articles qui se sont le mieux vendus pour chaque groupe de produits en septembre 1999, d'après les volumes des ventes indiqués par les données scannées. Dans les colonnes de droite, on trouve les proportions correspondantes des relevés de prix effectués au titre de l'IPD pour chacun des articles en question.

Tableau 1: les dix articles qui se sont le mieux vendus selon les données scannées et pourcentages correspondants des relevés de prix effectués au titre de l'IPD, en septembre 1999 (les pourcentages cumulatifs sont mentionnés entre parenthèses)

Modèle	Téléviseurs à écran de 36 cm		Téléviseurs à écran de 53 cm		Aspirateurs	
	Pourcentage des données scannées	Pourcentage des relevés de prix pour l'IPD	Pourcentage des données scannées	Pourcentage des relevés de prix pour l'IPD	Pourcentage des données scannées	Pourcentage des relevés de prix pour l'IPD
Modèle 1	17,7 (17,7)	1,0 (1,0)	16,2 (16,2)	10,5 (10,5)	30,1 (30,1)	18,7 (18,7)
Modèle 2	13,9 (31,6)	25,0 (26,0)	12,8 (29,0)	4,4 (14,9)	13,2 (43,3)	3,0 (21,7)
Modèle 3	11,0 (42,6)	1,9 (27,9)	11,7 (40,7)	1,8 (16,7)	8,7 (52,0)	1,2 (22,9)
Modèle 4	8,5 (51,1)	28,6 (56,5)	10,2 (50,9)	8,8 (25,5)	5,7 (57,7)	1,2 (24,1)
Modèle 5	8,2 (59,3)	3,8 (60,3)	10,1 (61,0)	31,6 (57,1)	4,4 (62,1)	0,6 (24,7)
Modèle 6	6,9 (66,2)	4,8 (65,1)	10,1 (71,1)	3,5 (60,6)	4,1 (66,2)	20,5 (45,2)
Modèle 7	6,6 (72,8)	1,9 (67,0)	6,1 (77,2)	8,8 (69,4)	4,1 (70,3)	0,6 (45,8)
Modèle 8	4,9 (77,7)	4,8 (71,8)	5,6 (82,8)	0,8 (70,2)	3,8 (74,1)	1,2 (47,0)
Modèle 9	4,4 (82,1)	1,0 (72,8)	4,1 (86,9)	1,7 (71,6)	3,5 (77,6)	0,6 (47,6)
Modèle 10	3,9 (86,0)	3,8 (76,6)	1,8 (88,7)	1,7 (73,6)	3,4 (81,0)	6,6 (54,2)

Modèle	Appareils photo		Lave-vaisselle		Machines à laver	
	Pourcentage des données scannées	Pourcentage des relevés de prix pour l'IPD	Pourcentage des données scannées	Pourcentage des relevés de prix pour l'IPD	Pourcentage des données scannées	Pourcentage des relevés de prix pour l'IPD
Modèle 1	28,4 (28,4)	38,4 (38,4)	17,2 (17,2)	2,2 (2,2)	12,0 (12,0)	6,5 (6,5)
Modèle 2	13,6 (42,0)	1,2 (39,6)	17,1 (34,3)	16,3 (18,5)	11,2 (23,2)	20,3 (26,8)
Modèle 3	11,9 (53,9)	12,8 (52,4)	9,4 (43,7)	11,9 (30,4)	11,2 (34,4)	2,3 (29,1)
Modèle 4	7,6 (61,5)	3,5 (55,9)	7,8 (51,5)	5,9 (36,3)	9,8 (44,2)	5,8 (34,9)
Modèle 5	6,7 (68,2)	1,2 (57,1)	7,3 (58,8)	6,7 (43,0)	6,9 (51,1)	1,4 (36,3)
Modèle 6	5,6 (73,8)	2,3 (59,4)	5,8 (64,6)	0,7 (43,7)	5,1 (56,2)	4,3 (40,6)
Modèle 7	4,4 (78,2)	15,1 (74,5)	5,1 (69,7)	23,0 (66,7)	5,1 (61,3)	2,9 (43,5)
Modèle 8	4,3 (82,5)	3,5 (78,0)	5,1 (74,8)	0,7 (67,4)	4,4 (65,7)	1,4 (44,9)
Modèle 9	4,0 (86,5)	1,2 (79,2)	4,8 (79,6)	3,0 (70,4)	4,2 (69,9)	1,4 (46,3)
Modèle 10	3,4 (89,9)	1,2 (80,4)	4,1 (83,7)	0,7 (71,1)	4,1 (74,0)	4,3 (50,6)

25. Il convient de noter que l'échantillon constitué pour l'IPD de septembre résulte de l'effet conjugué du choix de l'échantillon initial (qui peut en théorie remonter à cinq ans au maximum), de l'actualisation annuelle du panier (en l'occurrence, les nouveaux relevés de prix introduits en janvier 1999 lorsqu'un quart des points de vente ont été réapprovisionnés) et des substitutions qu'il a fallu effectuer depuis cette date à mesure que des modèles anciens disparaissaient des magasins.

26. Les résultats révèlent des tendances très intéressantes. De manière générale, les enquêteurs ont choisi des articles qui se vendaient bien, encore qu'il leur soit souvent arrivé de recueillir un nombre excessif de prix pour des modèles qui n'étaient que moyennement populaires. Certains des exemples les plus frappants de disparité concernaient les lave-vaisselle. En effet, le modèle le mieux vendu, qui représentait un cinquième des ventes, n'a fait l'objet que de 2 % des relevés de prix, tandis que le modèle venant au septième rang, qui ne représentait que 4 % des ventes, comptait pour plus de 20 % dans les relevés de prix. La même tendance a été observée pour d'autres articles.

27. Même si nous analysons une distribution cumulative, les problèmes demeurent évidents. Dans l'ensemble des cas, la proportion des relevés des prix pour l'IPD qui correspondent aux dix modèles les mieux vendus est bien inférieure à la place qu'ils occupent dans les chiffres de vente. Dans le cas des machines à laver, les dix modèles les mieux vendus qui représentent 74,0 % des ventes d'après les données scannées ne comptent que pour 50,6 % dans les relevés de prix au titre de l'IPD. Sur les trois mois considérés, ces résultats sont relativement stables, encore que les variations constatées soient suffisantes pour révéler une certaine diminution de la validité de l'échantillon au cours de la période considérée.

28. Les raisons de ces anomalies apparentes ne sautent pas aux yeux. Elles feront l'objet d'une analyse plus approfondie dans la suite du document. Cela dit, les anomalies en question ne sont pas nécessairement liées uniquement aux faiblesses des données de l'IPD. Par exemple,

un modèle particulier de machine à laver représente près de 10 % des relevés de prix pour l'IPD de septembre, tandis que d'après les données scannées, aucune vente de ce modèle n'a été enregistrée au cours de ce mois. Comme il est difficile de croire que les enquêteurs se soient attachés à relever le prix d'une machine qui ne s'est pas du tout vendue au cours d'un mois donné, on peut se demander si, en réalité, les ventes de la machine ne se sont pas effectuées sur un segment du marché qui n'était pas couvert par les données scannées. Malheureusement, nous n'avons pas pu vérifier cette hypothèse, faute d'informations concernant les points de vente sur lesquels portaient effectivement les données scannées.

### Interprétation

29. Bien entendu, l'interprétation des résultats dépend tout autant de la qualité des données scannées et du champ qu'elles couvrent que de la représentativité de l'échantillon de l'IPD. Toutefois, ces résultats semblent indiquer qu'une trop grande importance est parfois attribuée dans les relevés de prix à des produits ou variétés qui se vendent relativement peu, si l'on en juge d'après les données scannées, et ce malgré les instructions données aux enquêteurs de choisir une variété de produit représentative des ventes de cet article dans le magasin considéré.

Inversement, il arrive que des articles qui se vendent très bien ne soient pas sélectionnés.

Au nombre des causes de ce phénomène figurent:

- La méthode fondée sur un panier fixe, selon laquelle l'ensemble des produits et variétés ainsi que des articles composant ce panier est réexaminé une fois par an tout au plus. En effet, elle se traduit par une diminution toujours plus marquée de la représentativité de l'échantillon, à mesure que le «choix fixe» de biens composant le panier perd de son actualité au cours de la durée de vie utile de l'échantillon. Cette constatation n'est pas surprenante mais elle amène à se demander si, pour certains articles, dont les modèles sont très rapidement renouvelés, le panier ne devrait pas être actualisé plus souvent qu'une fois par an. En tout état de cause, on peut en déduire qu'un modèle d'article devrait être remplacé avant que le volume des ventes correspondant se réduise à presque rien ou que le modèle disparaisse des rayons;
- Les défauts de l'approche selon laquelle un enquêteur choisit un produit ou une variété «analogue» lorsqu'il se voit contraint de procéder à une substitution parce qu'un article n'est plus représentatif et ne se trouve plus en magasin. Cette approche peut contribuer à l'obsolescence de l'échantillon mais elle présente l'avantage de ne pas nécessiter autant d'ajustements de qualité. Elle appelle l'attention sur l'opportunité d'adopter une approche intégrée de la représentativité et de l'ajustement de qualité;
- Le fait que l'efficacité d'un choix approprié de produits et de variétés peut être amoindrie par le manque de représentativité de l'échantillon des points de vente. Cette cause est jugée la moins vraisemblable compte tenu du mode d'échantillonnage utilisé. Il est intéressant de noter, toutefois, que les données scannées font apparaître une différence marquée entre les types de points de vente dans les valeurs unitaires et les variations mensuelles des valeurs unitaires. Ainsi, un déséquilibre relativement mineur dans la composition de l'échantillon des points de vente pourrait avoir un retentissement sensiblement plus grand sur la fiabilité de l'inflation mesurée (voir la section VII ci-après).

30. Le degré d'inquiétude que doivent inspirer ces résultats dépend, en partie du moins, de l'effet constaté ou non sur l'indice publié et sur la mesure du taux d'inflation. La section suivante du document rend compte de la deuxième phase de l'étude qui avait pour but de déterminer si un tel effet se faisait sentir.

## **VI. Prix unitaires moyens et variations de prix moyennes**

31. Cette partie de l'étude consistait à observer, pour des variétés déterminées de produits, la mesure dans laquelle les niveaux et variations des prix observés à partir de données recueillies par les enquêteurs sur le terrain différaient de ceux que faisaient apparaître les données scannées. À cette fin, des données scannées relatives à des modèles spécifiques de chacun des produits couverts ont dû être soigneusement appariées avec des données pour le même modèle recueillies dans le contexte de l'IPD. Ces travaux ont exigé un volume considérable de ressources car il a fallu extraire des données détaillées des fichiers informatiques contenant les archives de l'IPD et effectuer une série de mises en concordance et de contrôles de validité avant de pouvoir exploiter les données. C'est pour cette raison que la comparaison a été limitée à une période de trois mois (août à octobre 1999).

### Limites pratiques du processus d'appariement et degré de succès obtenu

32. Il convient de noter que certains problèmes n'ont pas été résolus malgré les processus de vérification mentionnés plus haut. Ces problèmes sont essentiellement imputables au fait que les descriptions des enquêteurs ne sont pas assez complètes pour le processus d'appariement (bien qu'elles soient généralement suffisantes pour identifier les variétés de produits dans les magasins). Par exemple, le nom du fabricant et un nombre restreint de caractéristiques peuvent suffire pour identifier une variété de produit dans un magasin mais on aura besoin du numéro du modèle, qui dans de nombreux cas ne sera pas mentionné par l'enquêteur, pour pouvoir appairer, sans erreur possible, une variété de produit avec l'une de celles figurant sur la liste des données scannées.

### Niveaux des prix

33. Le tableau 2 donne un aperçu général de l'efficacité du processus d'appariement. Il convient de noter que le degré de succès des appariements n'est pas le même pour les cinq articles sélectionnés. Le processus s'est révélé le plus efficace pour les lave-vaisselle, les machines à laver et les aspirateurs, où plus de 70 % des observations au titre de l'IPD (représentant près de 50 % des variétés de produits considérées aux fins de l'IPD) ont pu être mis en correspondance avec les données scannées. C'est dans le cas des appareils photo qu'il s'est avéré le plus difficile puisque environ la moitié seulement des relevés de prix au titre de l'IPD (soit à peu près un tiers des variétés de produits considérées aux fins de l'IPD) a pu être appariée. Il va de soi que ces différences pouvaient influencer sur les conclusions de l'étude. En particulier, les différences entre les niveaux de prix et variations de prix pour l'échantillon apparié, d'une part, et la série complète de données de l'IPD, d'autre part, pouvaient entraîner des erreurs systématiques si, en raison de la manière dont il était constitué, l'échantillon apparié n'était pas représentatif.

34. Il y a lieu de formuler un certain nombre d'observations:

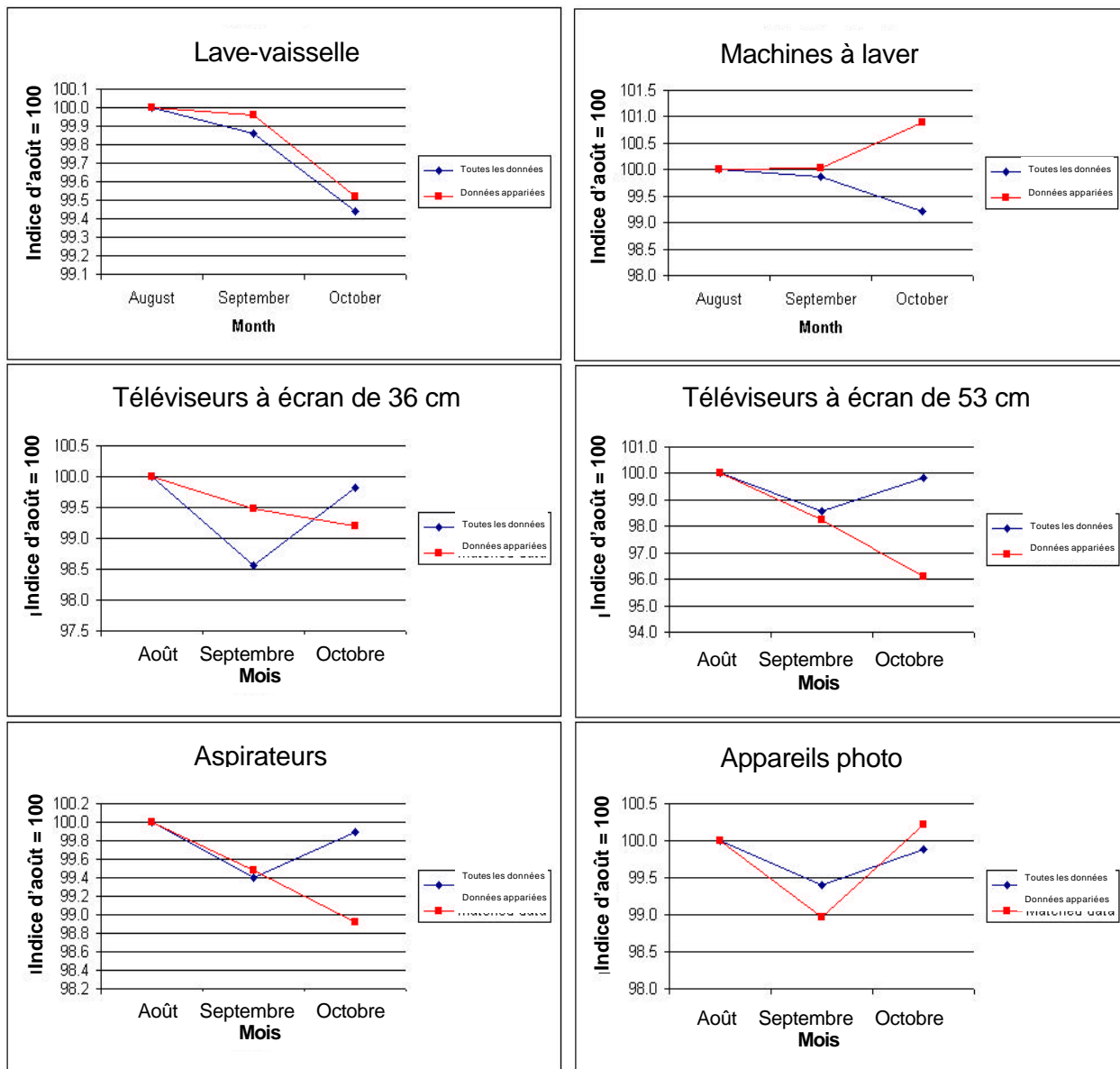
- Des différences significatives peuvent exister en ce qui concerne le **niveau des prix moyen** pour une variété de produit entre la série complète de relevés de prix de l'IPD et le sous-ensemble efficacement apparié avec les données scannées. C'est pour les téléviseurs et les machines à laver que ces différences sont les plus marquées;
- En général, on ne relève pour l'ensemble des articles aucune tendance indiquant que le prix moyen pour l'échantillon apparié est systématiquement plus élevé ou plus bas que pour la série complète des relevés de prix de l'IPD. Toutefois, pour un article donné, le sens de la différence restait le même au fil du temps, sauf dans le cas des appareils photo pour lesquels l'écart était faible. On peut en déduire qu'un effet non aléatoire se manifeste au niveau de chaque type d'articles, bien qu'il soit difficile de le vérifier à l'aide d'une moyenne pondérée, et d'un échantillon autocorrélé;
- Des différences apparaissent entre les **variations de prix moyennes** indiquées par la série complète des données scannées et celles qui ressortent des données appariées. Cet aspect a été étudié en calculant les indices de Laspeyres<sup>1</sup>, de Paasche<sup>1</sup> et de Fisher<sup>1</sup> pour la série complète de relevés des prix de l'IPD et pour le sous-échantillon correspondant aux observations appariées. Les résultats obtenus avec un indice de Fischer montrent que les variations de prix pour le sous-échantillon suivaient des tendances analogues, mais pas forcément identiques, à celles constatées pour la série complète de données scannées (voir la figure 1).

35. Ces résultats indiquent clairement qu'il existe des différences nettes entre les séries de données complètes et les séries de données appariées, tout particulièrement en rapport avec le prix de l'article. Il est difficile de déterminer avec certitude les raisons de ces différences car tester des hypothèses à partir du système de l'IPD s'avère problématique. Toutefois, il est possible que les données émanant de certains types de magasins soient définies de manière plus précise et ce facteur, associé aux différences de prix décrites dans les analyses ultérieures, peut être à l'origine de l'effet. Toujours est-il que, quelle qu'en soit la cause, il existe un effet réel et il faut en tenir compte lorsqu'on analyse les résultats des comparaisons.

Tableau 2: Pourcentage de données appariées et comparaison entre la moyenne des prix pour l'ensemble des relevés effectués au titre de l'IPD et l'échantillon apparié. Août à octobre 1999 (moyennes en livres)

	Août			Septembre			Octobre		
	Pourcentage de données appariées	Moyenne pour tous les relevés IPD	Moyenne pour l'échantillon apparié	Pourcentage de données appariées	Moyenne pour tous les relevés IPD	Moyenne pour l'échantillon apparié	Pourcentage de données appariées	Moyenne pour tous les relevés IPD	Moyenne pour l'échantillon apparié
Téléviseurs à écran de 36 cm	39	135,5	146,7	46	130,8	148,9	46	129,2	150,7
Téléviseurs à écran de 53 cm	48	249,7	291,3	56	246,5	283,8	58	240,1	268,4
Aspirateurs	76	129,5	129,1	77	130,0	130,9	78	128,9	130,2
Appareils photo	55	55,4	56,9	50	56,5	59,9	53	57,3	56,4
Lave-vaisselle	71	339,5	332,3	73	337,9	330,8	69	333,3	328,5
Machines à laver	81	345,3	349,7	75	354,0	323,2	76	348,9	317,8

**Figure 1. Indices des prix de chaque article calculés à partir de l'ensemble des données scannées et du sous-ensemble de données appariées, au moyen d'un indice de Fisher. Août à octobre 1999**



## Les résultats

36. Malgré les limites de cet exercice dues aux problèmes d'appariement, les résultats sont néanmoins instructifs. Il y a lieu d'observer, tout d'abord, que de façon générale, le prix moyen calculé sur la base des relevés des prix effectués au titre de l'IPD est supérieur à la valeur unitaire correspondante calculée sur la base des données scannées. Ceci n'a rien de surprenant et découle de la différence entre les principes régissant la collecte de données. L'IPD est fondé sur la collecte de données pour un panier fixe de biens, en faisant abstraction des substitutions de produits ou de points de vente. En outre, il n'est tenu compte dans ce contexte que des rabais qui sont de nature à influencer les prix et on exclut expressément les fins de série ou les soldes, par exemple. En revanche, les données scannées fournissent une estimation directe des prix effectivement payés par les consommateurs pour les marchandises qu'ils achètent en mesurant la valeur et le volume des achats. Il en résulte que ces estimations reflètent les efforts déployés par les consommateurs pour acheter un produit au meilleur prix, et tiennent compte par conséquent des effets des substitutions. De ce fait, elles produisent systématiquement un prix moyen inférieur. En outre, tous les rabais, quelles qu'en soient les modalités, sont pris en considération, ce qui réduit aussi le prix moyen dérivé du coût unitaire.

37. En examinant les données plus en détail, on a constaté que non seulement les prix moyens enregistrés par les enquêteurs aux fins de l'établissement de l'IPD, pour chaque produit, étaient généralement supérieurs à la valeur unitaire moyenne calculée à partir des données scannées, mais, très fréquemment, le prix moyen calculé par les enquêteurs pour une variété d'un produit donné était lui aussi plus élevé que les valeurs unitaires correspondantes tirées des données scannées. Toutefois, une analyse comparative des écarts absolus et des écarts absolus en pourcentage entre les relevés de prix de l'IPD et les valeurs unitaires indiquées par les données scannées (tableau 3) montre que cette différence est due pour une large part à un nombre relativement restreint de prix ou valeurs unitaires élevés ou faibles apparaissant dans la comparaison. Ainsi, les écarts des médianes sont dans tous les cas sensiblement inférieurs aux écarts correspondants des moyennes arithmétiques. Cette différence peut aussi s'expliquer partiellement par le fait que les valeurs unitaires des données scannées reflètent les quantités vendues.

Tableau 3: Écarts absolus et écarts absolus en pourcentage entre les moyennes pour les relevés effectués au titre de l'IPD et les valeurs unitaires des données scannées, calculés à la fois sur la base des moyennes arithmétiques et sur la base des médianes: moyenne d'août à octobre 1999

	Écart absolu (en livres)		Écart absolu en pourcentage	
	Moyenne	Médiane	Moyenne	Médiane
Lave-vaisselle	29,4	21,1	9,99	6,35
Machines à laver	34,8	21,3	10,45	7,58
Aspirateurs	13,3	7,7	9,71	6,07
Télévisions à écran de 36 cm	14,9	9,7	13,95	7,84
Télévisions à écran de 53 cm	30,0	16,6	9,60	6,05
Appareils photo	9,2	5,9	16,10	10,36

38. Les coefficients de variation figurant dans le tableau 4 donnent une vue d'ensemble utile, car ils font abstraction de l'impact des différents niveaux de la moyenne pour les divers produits. Les lave-vaisselle affichent le coefficient de variation le plus élevé pour la différence entre le prix moyen et la valeur unitaire moyenne, exprimée en pourcentage de la valeur unitaire moyenne. Les aspirateurs et les téléviseurs à écran de 53 cm présentent des coefficients de variation élevés à la fois pour la différence de prix exprimée en valeur monétaire et pour la différence exprimée en pourcentage. De toute évidence, il existe des arguments en faveur de l'utilisation d'échantillons élargis lorsque, comme dans les cas évoqués précédemment, les moyennes sont particulièrement sensibles aux valeurs aberrantes.

Tableau 4: Coefficients de variation

	Coefficients de variation	
	Écarts absolus en valeur monétaire	Écarts absolus en pourcentage
Lave-vaisselle	0,92	1,32
Machines à laver	1,09	0,99
Aspirateurs	1,41	1,19
Télévisions à écran de 36 cm	1,07	1,12
Télévisions à écran de 53 cm	1,23	1,23
Appareils photo	1,04	1,04

#### Variations des prix

39. Une analyse correspondante des variations mensuelles des prix (tableau 5) indique que rien ne prouve que **les variations** des prix relevés dépassent constamment **les variations** des valeurs unitaires ou inversement sauf dans le cas des articles suivants:

- Les machines à laver et les aspirateurs dont les baisses de prix enregistrées par les données scannées sont systématiquement plus marquées que celles apparaissant dans l'échantillon de l'IPD;
- Les appareils photo, pour lesquels les données de l'IPD révèlent une évolution des prix de même type, mais en plus prononcé.



Tableau 5: Indices (août = 100) et variations des prix mensuelles pour les relevés de prix de l'IPD et les données scannées appariées. Août à octobre 1999

	Août		Septembre		Octobre	
	Indice	Variation par rapport au mois précédent	Indice	Variation par rapport au mois précédent	Indice	Variation par rapport au mois précédent
<b>Lave-vaisselle</b>						
Relevés de l'IPD	100	–	102,2	+2,2 %	104,7	+2,5 %
Données scannées	100	–	101,6	+1,6 %	106,0	+4,4 %
<b>Machines à laver</b>						
Relevés de l'IPD	100	–	98,4	-1,6 %	97,0	-1,4 %
Données scannées	100	–	96,6	-3,4 %	98,3	-1,7 %
<b>Télévisions à écran de 36 cm</b>						
Relevés de l'IPD	100	–	99,9	-0,1 %	140,6	+4,4 %
Données scannées	100	–	100,5	+0,5 %	101,2	+0,6 %
<b>Télévisions à écran de 53 cm</b>						
Relevés de l'IPD	100	–	93,5	-6,5 %	91,5	-2,1 %
Données scannées	100	–	94,5	-5,5 %	99,2	+5,0 %
<b>Aspirateurs</b>						
Relevés de l'IPD	100	–	97,1	-2,9 %	94,3	-2,9 %
Données scannées	100	–	96,6	-3,4 %	92,5	-4,3 %
<b>Appareils photo</b>						
Relevés de l'IPD	100	–	109,8	+9,8 %	101,5	-7,6 %
Données scannées	100	–	105,5	+5,5 %	100,8	-4,5 %

40. Dans certains cas, les divergences qui se manifestent entre les tendances des prix et celles des valeurs unitaires peuvent être dues au petit nombre d'observations des prix réalisées dans le cadre de l'IPD pour le modèle considéré – dans ces conditions, le prix peut fluctuer considérablement d'un mois à l'autre avec l'introduction de prix soldés et d'offres spéciales. Ce phénomène ne devrait pas nécessairement devenir un sujet de préoccupation car l'IPD n'est pas conçu pour mesurer les variations de prix des différentes variétés d'un produit. Cependant, dans d'autres cas, la différence est difficile à expliquer. Elle peut être imputable, entre autres, aux différences dans la combinaison de points de vente retenue et en particulier au fait que les données scannées prennent en compte le remplacement de points de vente par d'autres, c'est-à-dire les variations des prix d'achat moyens qui résultent de la recherche du meilleur prix par la clientèle. Ce problème de manque d'homogénéité a été évoqué précédemment et est susceptible d'avoir un effet significatif parce qu'il existe entre les divers types de commerce d'importantes variations observables des niveaux de prix et de l'évolution des prix, ainsi qu'il ressort clairement de l'analyse des valeurs unitaires obtenues à partir des données scannées qui figure dans le tableau 6.

**Tableau 6: Effet du type de magasin considéré sur les valeurs unitaires pour différentes marques de lave-vaisselle, telles qu'elles apparaissent dans les données scannées**

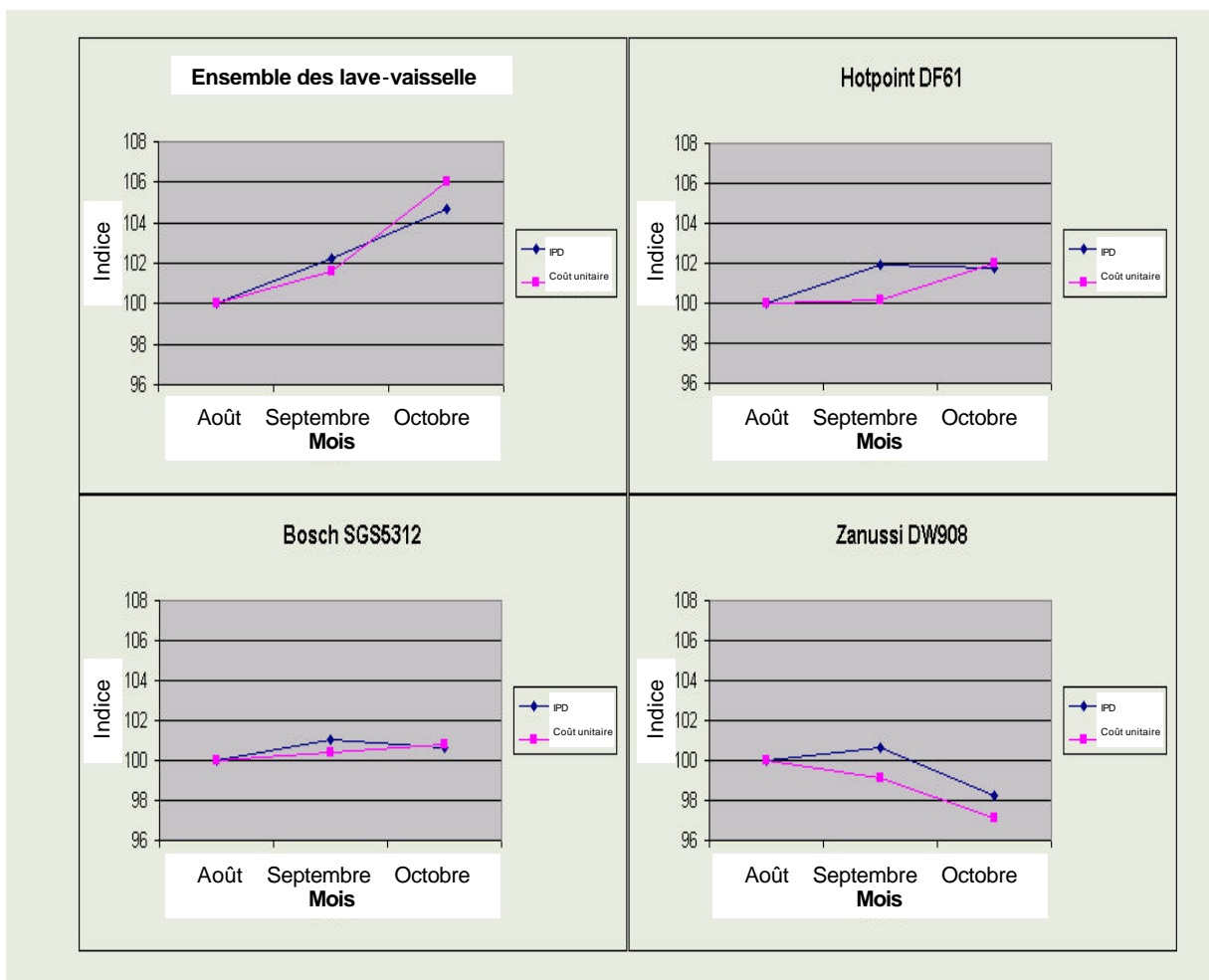
	Valeur unitaire (en £)			Variation d'août à septembre (%)	Ventes (%)		
	Août	Septembre	Octobre		Août	Septembre	Octobre
Bosch SGS5312 Grands magasins à succursales multiples	370,1	374,9	379,1	2,5	707 (31,5 %)	853 (34,4 %)	681 (35,7 %)
Grandes surfaces	364,0	364,8	363,1	-0,3	1 195 (53,3 %)	1 288 (51,9 %)	944 (49,5 %)
Commerces indépendants	386,5	382,3	386,0	-0,1	341 (15,2 %)	342 (13,8 %)	281 (14,7 %)
Vente sur catalogue	-	-	-	-	0 (0 %)	0 (0 %)	0 (0 %)
Ensemble des commerces	369,2	370,8	372,2	0,8	2 243	2 483	1 906
Hotpoint DF61 Grands magasins à succursales multiples	309,2	310,7	314,6	1,7	1 190 (35,9 %)	2 361 (54,4 %)	1 756 (53,5 %)
Grandes surfaces	288,1	296,4	307,8	6,8	364 (11,0 %)	458 (10,6 %)	310 (9,4 %)
Commerces indépendants	326,9	328,7	332,9	1,8	1 756 (52,9 %)	1 513 (34,9 %)	1 211 (36,9 %)
Vente sur catalogue	400,0	400,0	346,7	-1,2	6 (0,2 %)	5 (0,1 %)	6 (0,2 %)
Ensemble des commerces	315,0	315,6	321,2	1,4	3 316	4 337	3 283
Zanussi DW908 Grands magasins à succursales multiples	258,2	261,5	242,2	-6,2	740 (49,3 %)	705 (54,0 %)	780 (51,5 %)
Grandes surfaces	264,6	260,6	263,4	-0,5	236 (15,7 %)	287 (22,0 %)	210 (13,9 %)
Commerces indépendants	282,1	275,5	286,8	1,7	463 (30,9 %)	265 (20,3 %)	476 (31,4 %)
Vente sur catalogue	313,4	307,9	309,6	-1,2	61 (4,1 %)	48 (3,7 %)	49 (3,2 %)
Ensemble des commerces	268,2	265,9	260,6	-2,8	1 500	1 305	1 515

#### Examen détaillé des modèles de lave-vaisselle

41. Pour mieux comprendre les raisons de ces différences, il convient de procéder à un examen détaillé de chaque produit et de chaque variété de produit. Dans la figure 2, on trouve une

comparaison entre un indice pour l'ensemble des lave-vaisselle, et les indices calculés pour les différents modèles dans ce groupe d'articles. Bien que, dans leur ensemble, les lave-vaisselle n'accusent pas de différence systématique des variations de prix entre l'IPD et les données scannées et que les variations soient relativement proches, certains écarts intéressants sont observés pour les différents modèles.

**Figure 2: Variations des prix entre août 1999 et octobre 1999 pour certaines marques de lave-vaisselle**



### Bosch SGS5312

Ce lave-vaisselle est celui pour lequel on constate le moins de différence entre les variations de prix enregistrées sur la base des relevés de l'IPD et les variations des coûts unitaires. Les raisons de cette faible différence se dégagent de l'analyse des prix par type de magasin figurant dans le tableau 6. En l'occurrence, les prix et les variations de prix, pour les divers types de magasin, sont semblables, les variations ne s'écartant jamais de plus de 1,7 % de la moyenne. Ces raisons, associées au fait que seuls des changements mineurs sont intervenus dans la répartition des ventes par type de magasin, ont produit un indice par article semblable pour les deux sources de données.

### **Hotpoint DF61**

Pour cet article, l'indice d'octobre fondé sur les données de l'IPD est très proche de celui basé sur les données scannées. Toutefois, les indices de septembre sont sensiblement différents. On trouve également une explication partielle de cette différence dans l'analyse par type de magasin figurant dans le tableau 6. Entre août et septembre, une tendance marquée à la baisse des achats auprès des commerces indépendants les plus chers au profit des grands magasins à succursales multiples meilleur marché s'est manifestée, en même temps qu'une augmentation générale du volume des ventes. Cette évolution s'est traduite par une baisse de l'indice en septembre. Toutefois, un autre facteur entre en jeu étant donné que la remontée de l'indice en octobre ne s'est pas accompagnée d'un retour à la répartition des ventes antérieure. Il y a tout lieu de penser que ce phénomène est partiellement lié aux augmentations différentielles des prix observées dans les différents groupes, mais ce n'est peut-être pas la seule explication.

### **Zanussi DW908**

Pour ce modèle de lave-vaisselle, nous constatons une divergence entre l'indice issu des données scannées et l'indice calculé en fonction des données de l'IPD, d'août à septembre, et quoique l'écart se rétrécisse légèrement de septembre à octobre, ces indices demeurent différents. Là aussi, la différence initiale est due, du moins partiellement, à une baisse des achats dans les magasins coûteux au profit de magasins pratiquant des prix moins élevés. Toutefois, en l'occurrence, la répartition des ventes revient pratiquement à sa configuration initiale, sans un retour consécutif de l'indice pour les données scannées au niveau de l'IPD. Il est clair également que cette évolution de l'indice ne doit pas être attribuée à des variations de prix différentielles dans les magasins car les commerces où les consommateurs sont revenus faire leurs achats ont augmenté leurs prix davantage que les autres types de magasins. On ne sait pas très bien à quoi tient la différence, encore qu'une partie du mouvement soit peut-être imputable à des offres spéciales qui n'ont pas été prises en compte dans les données scannées. Nous analyserons ces différences plus en détail dans le cadre des travaux en cours.

42. Ces travaux montrent clairement que le choix des points de vente est important pour veiller à ce que l'IPD produise un ensemble de prix représentatif. Certes, nous sommes convaincus que le système actuel fonctionne bien, mais il est essentiel que nous ne perdions pas de vue les changements qui peuvent se produire dans la répartition des ventes entre les détaillants, particulièrement à plus long terme. À plus court terme, la substitution de points de vente à d'autres est plus difficile à prendre en compte et cet aspect de la question n'entre pas du tout dans le champ de l'IPD actuel. Cependant, nous devons avoir conscience de ces changements pour pouvoir mieux interpréter les variations de l'IPD.

## **VII. La question des pondérations implicites et des formules d'agrégation**

43. Les indices des prix des produits, sur lesquels est axé le présent document, sont calculés sur la base de la moyenne de rapports<sup>1</sup>. On n'utilise pas de pondération explicite dans ce calcul mais on part de l'hypothèse, s'agissant de la moyenne de rapports, que tous les relevés de prix ont la même importance, c'est-à-dire qu'on leur attribue un poids égal dans l'agrégat élémentaire. Or, cette hypothèse n'est vraiment exacte que si la combinaison de relevés de prix

considérée est représentative des marques et modèles vendus pour chaque article. Une autre démarche consisterait à utiliser des pondérations explicites dérivées des volumes des ventes de chaque modèle, tels qu'ils ressortent des données scannées. Le tableau 7 compare les indices des prix basés sur la méthodologie actuelle de l'IPD avec une moyenne pondérée calculée selon une formule de Laspeyres<sup>1</sup>, en utilisant à la fois des données sur les prix de l'IPD et des données scannées relatives au mois d'août pour les pondérations.

44. Ces comparaisons révèlent des différences assez importantes (par exemple 4,5 points de pourcentage pour les machines à laver en septembre) mais ne font apparaître aucune tendance systématique que ce soit dans l'ordre de grandeur ou dans l'orientation de l'écart, et reflètent en grande partie la disparité des proportions de relevés de prix par modèle que l'on constate entre les données de l'IPD et les données scannées. De toute évidence, ces résultats traduisent l'effet qu'exercent sur les indices pour ces articles les différences de répartition mises en lumière dans les sections précédentes du document. Là aussi, nous devons nous montrer prudents en appliquant ces résultats à l'indice d'ensemble, compte tenu des différences constatées entre les données appariées et la série complète de données de l'IPD. Quoi qu'il en soit, il est clair que nous pourrions obtenir des résultats sensiblement différents pour les divers groupes de produits si nous adoptions une autre optique pour le choix des articles.

Tableau 7: Comparaison des indices reposant sur un rapport non pondéré des moyennes et une formule de Laspeyres pondérée: août à octobre 1999

	août	septembre	octobre
<b>Lave -vaisselle</b>			
Rapport des moyennes	100,0	99,2	97,2
Laspeyres	100,0	100,8	100,4
<b>Machines à laver</b>			
Rapport des moyennes	100,0	103,3	99,7
Laspeyres	100,0	98,7	99,7
<b>Aspirateurs</b>			
Rapport des moyennes	100,0	102,1	101,6
Laspeyres	100,0	101,4	100,2
<b>Télévisions à écran de 36 cm</b>			
Rapport des moyennes	100,0	100,9	100,4
Laspeyres	100,0	101,4	100,0
<b>Télévisions à écran de 53 cm</b>			
Rapport des moyennes	100,0	100,2	94,6
Laspeyres	100,0	96,9	97,2
<b>Appareils photo</b>			
Rapport des moyennes	100,0	100,7	100,0
Laspeyres	100,0	99,2	97,9

### VIII. Une approche intégrée de la représentativité et de l'ajustement de qualité

45. Jusqu'ici, le présent document a porté essentiellement sur la question de la représentativité de l'échantillon et la manière dont on pouvait vérifier cette représentativité par une analyse comparative avec des données scannées. Dans la pratique, il est difficile de dissocier une réflexion sur la représentativité de l'échantillon des questions relatives à l'ajustement de qualité.

En particulier, il faut trouver un compromis, à la fois en termes de ressources et de qualité technique de l'indice, entre des ajustements de qualité peu fréquents mais substantiels et des ajustements plus fréquents mais aussi plus limités:

- Le maintien de la représentativité de l'échantillon peut imposer un surcroît de travail dans la mesure où il suppose des ajustements de qualité explicites. Par exemple, si l'on actualise plus fréquemment le panier d'articles pour ce qui concerne les produits de haute technicité en y introduisant des substitutions forcées «programmées» entre les actualisations générales, il faudra procéder à des ajustements de qualité plus fréquents;
- L'ajustement de qualité devient plus difficile sur le plan technique à mesure que la représentativité du panier va décroissant. Les variables hédoniques deviennent moins fiables et moins pertinentes;
- Certains changements dans l'évaluation de la qualité par les consommateurs n'auront pas été saisis au moment voulu. Par exemple, dans le cas où des particularités de l'ancien modèle auront été ramenées, au fil du temps, à une valeur nominale;
- La même source de données scannées peut fournir des informations sur les ventes à partir desquelles on peut constituer un échantillon et des informations sur les caractéristiques permettant d'effectuer des régressions hédoniques aux fins d'un ajustement de qualité;
- Les mêmes régressions hédoniques peuvent à la fois apporter aux enquêteurs des renseignements au sujet de la marque et de ses principales caractéristiques lorsqu'un article de substitution doit être sélectionné, et procurer la base nécessaire pour un ajustement de qualité explicite.

#### **IX. Aspects pratiques de l'utilisation de données scannées en vue d'améliorer l'échantillonnage**

46. Suite à la présentation du rapport, l'ONS s'est efforcée de traduire les conclusions de l'étude en une méthode pratique visant à améliorer la collecte des prix aux fins de l'établissement de l'IPD. La démarche suivie en l'occurrence a consisté à examiner les moyens d'utiliser les données scannées disponibles pour élaborer un plan de sondage par quotas, en donnant aux enquêteurs des instructions claires en ce qui concerne les modèles qu'il fallait prendre en considération.

47. Selon les modalités mises au point pour constituer un échantillon par la méthode des quotas, chaque enquêteur dispose d'une liste de modèles classés par ordre de priorité, dans laquelle il choisit son article. Par exemple, dans le cas des machines à laver, une telle liste pourrait se présenter comme suit:

Choix 1	BOSCH	WFL2000 FSA AUTO W FL 1000
Choix 2	ZANUSSI	FLA1001 FSB AUTO WASH FL 1000
Choix 3	HOOVER	AM120 FSA AUTO W.FL 1200
Choix 4	ZANUSSI	FJS1225 FSA AUTO W FL 1200
Choix 5	BOSCH	WFL226/2260 FSA AUTO W.FL 1100
Choix 6	SERVIS	M3510 FSB AUTO W.FL 1000

48. L'enquêteur se rend ensuite sur place pour la mise à jour du panier et vérifie si le premier modèle de la liste est disponible; dans l'affirmative, il le sélectionne, sinon, il passe au deuxième modèle sur la liste et répète l'opération. Ce processus de sélection se poursuit jusqu'à ce qu'un modèle soit choisi ou que la liste soit épuisée, auquel cas l'enquêteur choisit alors lui-même un modèle en fonction de sa popularité auprès de la clientèle du magasin considéré.

49. Les échantillons par la méthode des quotas ne sont pas personnalisés, mais ils sont établis pour l'ensemble des enquêteurs. La première étape consiste à sélectionner le modèle qui constitue le premier choix pour chaque enquêteur. Cette sélection se fait de manière aléatoire, en utilisant une méthode de probabilité proportionnelle à la taille. La mesure de la taille employée est la proportion des ventes totales qui correspond aux ventes de ce modèle, telle qu'indiquée par les données scannées. La deuxième étape consiste à sélectionner le modèle qui constituera le deuxième choix sur la liste et pour cela, on procède de la même manière si ce n'est que pour chaque enquêteur, le modèle sélectionné au cours de la première étape est exclu et que les probabilités sont ajustées en conséquence. L'opération est répétée jusqu'à ce qu'une liste de six modèles ait été établie pour chaque enquêteur.

50. L'efficacité pratique de cette méthode a été vérifiée de deux manières. Premièrement, on a testé la commodité de la méthode en procédant à une étude pilote avec la participation de 20 enquêteurs du Royaume-Uni et en élaborant des échantillons par la méthode des quotas pour cinq articles: machines à laver, lave-vaisselle, appareils photo, téléviseurs et aspirateurs. Deuxièmement, on a simulé des conditions de vente au détail et évalué leur effet sur la répartition finale par modèle pour déterminer dans quelle mesure les distributions obtenues correspondaient à la distribution idéale définie par les ventes, telles qu'elles ressortaient des données scannées.

### Étude pilote

51. Les enquêteurs ont été priés de tester les modalités mises au point pour l'introduction de l'échantillonnage par la méthode des quotas dans la collecte des prix de certains articles aux fins de l'établissement de l'IPD. Des relevés de prix ont été entrepris en mars, en avril et en mai 2001, et les enquêteurs ont fait part de leurs commentaires sur la commodité de la méthode et son impact sur les commerçants.

52. L'accueil a été généralement favorable: en effet, la plupart des enquêteurs préféraient être dirigés vers des modèles précis plutôt que d'avoir à demander l'avis des directeurs de magasin ou à s'appuyer sur leurs propres connaissances pour choisir les modèles les plus populaires.

53. Cependant, des problèmes se sont posés essentiellement dans le cas des revendeurs spécialisés dans une marque particulière (par exemple, les magasins Sony) ou dans le cas de marques ayant passé des contrats d'exclusivité avec des grands magasins. Ces problèmes seront analysés dans le cadre des travaux futurs pour déterminer quelle en est l'étendue, quels en sont les effets sur l'échantillon et quelles solutions sont envisageables. Les résultats de cette analyse influenceront sur les décisions qui seront prises concernant l'application de la méthode en situation réelle.

### Simulations

54. Lorsque les distributions initiales ont été établies, la stabilité de la méthode a été vérifiée à l'aide de simulations reproduisant les différents degrés de difficultés qu'éprouveraient les enquêteurs à trouver les produits considérés dans les magasins. Dans chaque cas, un total de 40 simulations ont été réalisées pour veiller à ce qu'un seul résultat inattendu ne fausse pas les conclusions. On a pris comme hypothèses quatre proportions différentes de relevés manquants: 5 %, 10 %, 20 % et 50 %. Des distributions hybrides ont alors été établies, en utilisant le deuxième, le troisième, le quatrième, etc. choix, selon les besoins, pour remplacer les valeurs manquantes.

55. Les distributions obtenues ont été soumises à un test du khi-carré en prenant comme référence la distribution idéale utilisée pour établir la distribution initiale fondée sur le premier choix.

56. Les résultats étaient sans surprise. Dans tous les cas, il est apparu que les distributions initiales n'étaient pas sensiblement différentes de la distribution idéale, ce qui confirme essentiellement la validité de notre première méthode de sélection.

57. Nous avons aussi constaté qu'à mesure que nous nous éloignons des distributions pures de la première série, le taux d'insuccès augmentait. On compte, cependant, un très petit nombre de tests de signification infructueux pour les zones où le taux d'insuccès dans la recherche de modèles est faible (selon toute attente, puisque les taux de substitution sont bas). L'effet le plus marqué en termes de différence entre l'échantillon obtenu et l'échantillon escompté est lié au degré d'insuccès de la recherche de modèles. À mesure que ce taux augmente, nous observons une accentuation constante de la différence, encore que certaines distributions correspondant à un taux d'insuccès de 50 % produisent des résultats acceptables par rapport à la distribution idéale.

58. Ce qui précède nous amène donc à penser que la méthode employée constitue une excellente base théorique pour la mise en place d'un système pratique de sondage par quotas. En outre, les distributions laissaient apparaître des défaillances prévisibles, avec une tendance à se rapprocher des distributions initiales lorsque les enquêteurs étaient libres de choisir eux-mêmes les modèles. Ces résultats sont rassurants, puisqu'ils indiquent que l'issue la moins favorable serait le statu quo. Une analyse plus approfondie est en cours pour examiner les taux d'insuccès enregistrés dans le cadre de l'étude pilote, en vue de déterminer à quoi ressembleraient des distributions effectives dans une situation réelle.

## **X. Conclusions de l'étude et ses incidences sur l'échantillonnage, le relevé des prix et l'ajustement de qualité**

59. L'étude décrite dans le présent document a soulevé un certain nombre de questions relatives aux pratiques actuelles en matière d'échantillonnage et de relevé des prix aux fins de l'établissement de l'indice des prix de détail du Royaume-Uni. Elle a également mis en lumière un certain nombre de solutions envisageables pour utiliser les données scannées afin d'obtenir un choix plus représentatif d'articles et de produits dans le cadre du processus traditionnel de relevé des prix, c'est-à-dire l'observation des prix dans les magasins. L'étude ne laisse pas nécessairement entendre que les méthodes actuelles d'échantillonnage entraînent des erreurs



systématiques mais elle incite à envisager la possibilité d'instaurer des contrôles et procédures supplémentaires pour maîtriser le risque d'erreur systématique.

60. Tout examen des incidences pratiques doit partir du principe que pour refléter la situation du marché, les variétés de produit représentatives devraient correspondre non seulement à des proportions importantes des ventes de l'article considéré, mais aussi accuser, dans l'ensemble, des variations de prix semblables. Sur cette base, nous pouvons formuler les observations pratiques suivantes, tout en gardant à l'esprit la nécessité d'étudier plus en détail l'échantillonnage par la méthode des quotas:

- La validité d'un panier «représentatif» peut se dégrader au cours de sa durée de vie utile, même s'il est mis à jour chaque année. Ce phénomène se produit, par exemple, dans le cas des produits de haute technicité, pour lesquels le taux de renouvellement des modèles est élevé. On peut alors se servir de données scannées, si leur couverture est satisfaisante, pour suivre l'évolution de la représentativité du panier au fil du temps et déterminer s'il doit être actualisé plus fréquemment et à quel moment. La mise à jour pourrait se faire sur la base de substitutions «forcées» programmées pour éviter les problèmes d'erreurs systématiques potentielles liés à des enchaînements fréquents. Ces mises à jour pourraient être déclenchées soit par un algorithme fondé sur des données scannées, ou, ce qui serait plus pratique, s'effectuer à intervalles réguliers;
- Dans les cas où des substitutions forcées demeurerait nécessaires, à cause de la disparition de certaines variétés de produits des magasins, le choix des produits de substitution pourrait se faire à l'aide de données scannées. On pourrait, par exemple, repérer les articles dont les caractéristiques sont les plus proches de celles du modèle qui disparaît ou encore utiliser une régression hédonique pour déterminer les caractéristiques qui influent le plus sur les décisions d'achat des consommateurs;
- Les mêmes régressions hédoniques peuvent être utilisées aux fins de l'ajustement de qualité explicite, à la fois lors des substitutions de type traditionnel, et lors des substitutions «forcées» programmées;
- Les données scannées par type de magasin indiquent qu'il faut veiller tout particulièrement à assurer une répartition appropriée des points de vente dans l'échantillon de l'IPD et que les données scannées peuvent être utilisées pour une stratification a posteriori lorsqu'il y a lieu de croire que l'échantillon obtenu selon les méthodes actuelles d'échantillonnage de l'IPD n'est pas entièrement autopondéré.

61. L'Office for National Statistics étudiera ces questions plus en détail dans le cadre de son programme de recherche méthodologique à plus long terme.

**Appendice 1: formules des agrégats élémentaires et formules de calcul des indices**

$$\text{Laspeyres} = \frac{P_t Q_0}{P_0 Q_0}$$

où  $P_t$  = le prix à la période t  
 $Q_t$  = la quantité vendue à la période t  
 Période 0 = le mois de référence

$$\text{Paasche} = \frac{P_t Q_t}{P_0 Q_t}$$

où  $P_t$  = le prix à la période t  
 $Q_t$  = la quantité vendue à la période t  
 Période 0 = le mois de référence

$$\text{Fisher} = \sqrt{\frac{P_t Q_0 P_t Q_t}{P_0 Q_0 P_0 Q_t}}$$

où  $P_t$  = le prix à la période t  
 $Q_t$  = la quantité vendue à la période t  
 Période 0 = le mois de référence

$$\text{Moyenne des rapports} = \frac{1}{n} \sum_{i=1}^n \frac{P_i^t}{P_i^0}$$

où  $P_i^t$  = le prix de l'article I à la période t  
 Période 0 = le mois de référence

$$\text{Rapport des moyennes} = \frac{\frac{1}{n} \sum_{i=1}^n P_i^t}{\frac{1}{n} \sum_{i=1}^n P_i^0}$$

où  $P_i^t$  = le prix de l'article I à la période t  
 Période 0 = le mois de référence

**NOTE**

<sup>1</sup> Voir l'appendice.

**RÉFÉRENCE**

Bradley, Cook, Leaver & Moulton [1997] *An overview of Research on Potential Uses of Scanner Data in the US CPI.*

*Genève 277*

-----