

Distr.
GÉNÉRALE

CES/AC.49/2003/4
24 septembre 2003

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, 4-5 décembre 2003

ANALYSE DE LA CONSTITUTION D'ÉCHANTILLONS AU MOYEN DE DONNÉES SCANNÉES

Communication sollicitée du Ministère de la gestion publique, de l'intérieur,
des postes et des télécommunications du Japon*

Résumé

La présente communication est consacrée à une analyse de la méthode d'échantillonnage au moyen de données scannées. Nous avons comparé les indices obtenus pour trois types d'échantillons comprenant respectivement 1) tous les articles, 2) l'article le plus important, 3) un certain nombre d'articles représentatifs (retenus sur la base d'un seuil d'inclusion) de 8 groupes et 24 catégories de produits alimentaires et de biens de consommation courante.

I. INTRODUCTION

1. L'indice des prix à la consommation (IPC) est en général calculé en décomposant la consommation des ménages en un certain nombre de groupes et de catégories, en affectant chaque catégorie d'un coefficient de pondération de façon à établir une classification de base et en procédant à l'agrégation des indices de chacune des catégories en un indice d'ensemble. Pour constituer les catégories, on groupe les articles achetés par les ménages en un certain nombre de familles au sein desquelles les prix sont, dans l'ensemble, supposés évoluer de façon analogue. Pour calculer l'indice de chaque catégorie, on sélectionne un certain nombre d'articles représentatifs signalés comme tels aux enquêteurs et dont les prix sont relevés de façon continue.

* Élaborée par M. Nobuyuki Sakashita et M. Shin-ichi Takahashi, Bureau des statistiques des prix, Division des statistiques de la consommation, Bureau de statistique.

L'idéal serait certes d'inclure tous les articles d'une catégorie, mais il est financièrement et matériellement difficile de collecter les prix de tous les articles. Par ailleurs, il est depuis peu possible d'utiliser des données scannées. Dans la plupart des pays, les données scannées ne sont utilisées pour la construction de l'IPC réel qu'aux fins d'ajustements de qualité concernant le matériel informatique, comme les ordinateurs, et ce parce que le coût de l'opération et le nombre de points de vente sont limités. Au Japon, la méthode hédonique est appliquée pour les ordinateurs et les appareils-photo numériques. Nous avons cependant étudié les questions générales soulevées par l'utilisation passée de données scannées pour la sélection des articles et des points de vente et le calcul d'indices catégoriels (Okamoto, 1999). Des études sur l'utilisation des données scannées sont également menées dans plusieurs autres pays (par exemple de Haan, 1999). L'avantage de l'utilisation de données scannées est qu'elles permettent de connaître les prix et les quantités de tous les articles effectivement vendus dans les points de vente inclus dans l'échantillon. L'expérience dont nous rendons compte consiste à calculer des indices de catégories, de groupes et un indice global pour 8 groupes et 24 catégories de produits alimentaires et de produits de consommation courante grâce à des données scannées. Trois méthodes différentes de calcul des indices de catégories sont passées en revue: 1) la formule de l'indice «superlatif» de tous les articles (indice de Törnqvist), 2) l'établissement d'indices de l'article le plus important, et 3) le calcul de la moyenne géométrique de plusieurs articles représentant les plus fortes ventes. Nous examinons ensuite, sur la base des résultats obtenus, les questions que pose la constitution de l'échantillon.

2. La première section ci-après décrit brièvement l'utilisation qui a été faite de données scannées, la deuxième section expose les différentes méthodes d'échantillonnage et formules utilisées pour le calcul des indices, et les sections 3 et 4 présentent les résultats et les conclusions de l'expérience.

II. UTILISATION DE DONNÉES SCANNÉES

3. L'utilisation de données scannées pour l'expérience peut se résumer comme suit:

- On a procédé à la sélection par échantillonnage stratifié d'un millier de points de vente dans chaque région et pour chaque type de commerce (grands magasins, supermarchés, drogueries, etc.) dans l'ensemble du pays.
- Des données scannées ont été relevées de façon continue de janvier 2000 à décembre 2001.
- Ces données concernaient 8 groupes et 24 catégories de produits alimentaires et de biens de consommation courante (voir tableau 1).
- On a calculé, pour chaque catégorie la moyenne mensuelle des prix et de la valeur ventes de tous les articles, cette valeur ayant été enregistrée chaque semaine entre janvier 2000 et décembre 2001¹.

¹ Les données scannées dont on dispose sont le prix moyen et la valeur totale des ventes pour chaque type d'établissement commercial dans l'ensemble du pays. Dans le cadre de la présente expérience, la moyenne des prix pratiquée dans l'ensemble des établissements commerciaux de l'ensemble du pays est à une moyenne arithmétique pondérée par le nombre de points de vente.

Tableau 1. Groupes et catégories

Groupes		Catégories	
1	Produits laitiers	1	Lait frais
		2	Yogourt
		3	Fromage
2	Matières grasses et huiles	4	Huiles comestibles
		5	Beurre
		6	Margarine
3	Condiments	7	Vinaigre
		8	Ketchup
		9	Mayonnaise
4	Confiserie	10	Biscuits
		11	Chocolat
		12	Crèmes glacées
5	Boissons	13	Café instantané
		14	Thé
		15	Jus de fruits et de légumes
6	Produits d'entretien	16	Nettoyants cuisine
		17	Produits de lessive
		18	Adoucissants
7	Fournitures médicales	19	Couches jetables
		20	Produits sanitaires
		21	Produits d'entretien des lentilles
8	Articles de toilette	22	Savon
		23	Shampooing
		24	Dentifrice

III. MÉTHODES ET FORMULES D'ÉCHANTILLONNAGE

Indices de tous les articles

4. La méthode consiste ici à inclure dans l'échantillon la totalité des articles effectivement vendus. Diverses formules peuvent être utilisées pour calculer les indices correspondants: moyenne arithmétique ou géométrique simple, moyenne arithmétique ou géométrique pondérée. Nous utilisons en l'occurrence l'indice de Törnqvist, qui est l'un des indices «superlatifs», car les données scannées nous permettent de connaître le montant total des ventes de chaque mois. Plus précisément, nous utilisons la méthode consistant, pour chaque catégorie, à rattacher en chaîne les indices calculés chaque mois en pondérant le prix moyen de chaque article de la moyenne des ventes pendant le mois en cours et le mois précédent, selon la formule suivante:

$$I_j^T = \prod_t \prod_i \left[\frac{\bar{p}_{ti}}{\bar{p}_{t-1i}} \right]^{\frac{w_{ti} + w_{t-1i}}{2}},$$

où i représente l'article, j la catégorie, t la période en cours, p le prix moyen et w la valeur des ventes.

5. Les indices de groupes et l'indice global sont calculés à partir des indices des catégories pertinentes. Les indices de niveau supérieur sont établis selon la formule de Laspeyres, en utilisant comme coefficient de pondération la valeur totale des ventes de l'année 2000 pour chaque article².

$$I^T = \frac{\sum_j I_j^T w_j}{\sum_j w_j}$$

Indices de l'article le plus important

6. Dans chaque catégorie, nous sélectionnons un article représentatif dont nous calculons l'indice. L'article retenu est celui qui s'est le mieux vendu pendant l'année 2000, l'indice pour chaque mois représentant le ratio des prix moyens de l'article.

$$I_j^1 = \frac{\bar{p}_{t1}}{\bar{p}_{01}},$$

où p_{01} représente le prix moyen de l'article le plus important pendant la période de base et p_{t1} le prix correspondant pendant la période en cours. La méthode de calcul des indices de groupe et de l'indice d'ensemble est la même qu'en (1).

Indices des articles représentatifs

7. Pour les indices des articles représentatifs, contrairement à la méthode de sélection proportionnelle à la taille, on sélectionne plusieurs articles phares hautement représentatifs. La représentativité est évaluée sur la base de la valeur annuelle des ventes en 2000, les cinq articles qui se sont le mieux vendus étant sélectionnés ($n=5$). Les indices pour chaque mois correspondent à la moyenne géométrique du ratio des prix moyens de ces articles.

² L'agrégation des indices au niveau supérieur n'a pas été faite selon la formule de l'indice superlatif en recourant à des données scannées. La raison en est que, dans cette expérience, l'indice de Laspeyres à pondérations fixes se rapproche davantage de l'IPC courant, puisque les pondérations utilisées dans le calcul de ce dernier sont établies à partir des résultats de l'enquête sur les dépenses des consommateurs. Le montant total des ventes calculé à partir de données scannées, utilisé comme pondération a donc valeur de substitut du montant des dépenses de consommation des ménages.

$$I_j^C = \prod_{i=1}^n \left[\frac{\bar{p}_{t1}}{\bar{p}_{01}} \right]$$

La méthode de calcul des indices des groupes et de l'indice d'ensemble est la même qu'en (1).

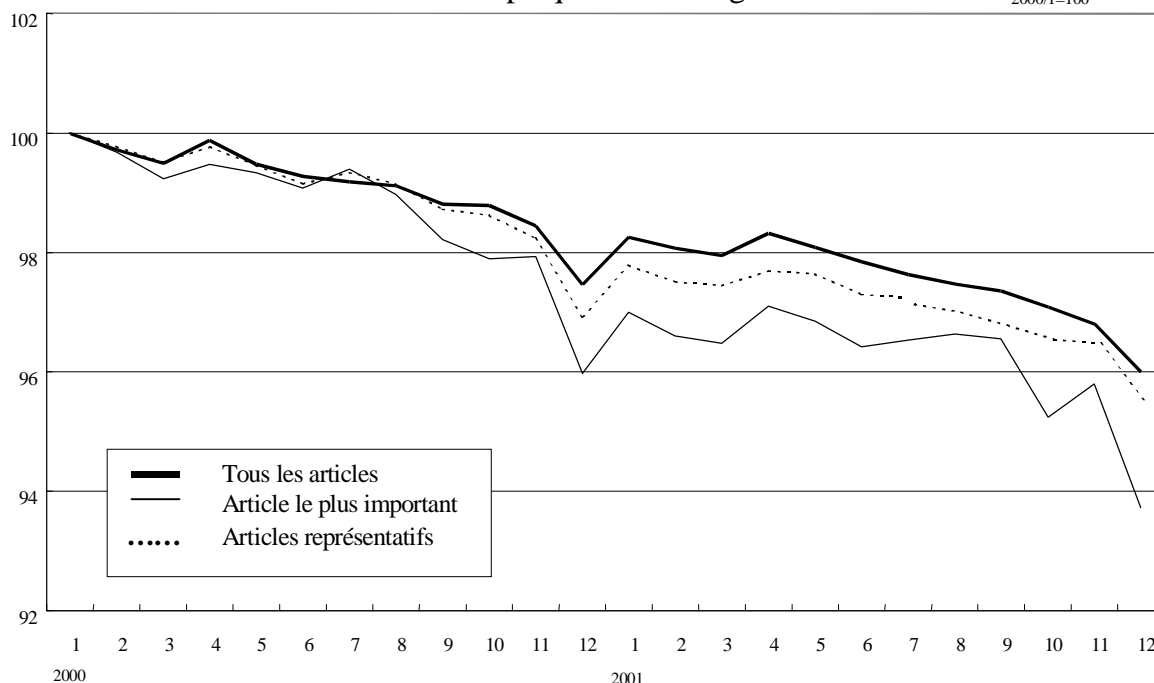
IV. RÉSULTATS

8. L'indice de tous les articles est en baisse, reflétant la récente tendance à la baisse des prix au Japon. L'indice de l'article le plus important et l'indice des articles représentatifs évoluent quasiment dans le même sens que l'indice de tous les articles. L'indice de l'article le plus important baisse toutefois plus fortement que celui de tous les articles, de même que l'indice des articles représentatifs, la divergence dans ce dernier cas étant toutefois moindre.

Tableau 2. Indice global (janv. 2000 = 100)

	Indice		2001/2000 (% de variation)	Divergence
	2000	2001		
Tous les articles	99,1	97,6	-1,6	-
Article le plus important	98,8	96,2	-2,6	-1,0
Articles représentatifs	99,1	97,1	-2,0	-0,4

Graphique 1: Indice global



9. En ce qui concerne les groupes et les catégories, bien que l'indice de l'article le plus important et l'indice des articles représentatifs évoluent apparemment dans le même sens que

l'indice de tous les articles, on constate dans la plupart des cas une divergence quant à l'importance de la baisse, divergence qui tend à s'accroître en particulier pour les groupes et catégories caractérisés par une tendance à la baisse forte et continue. Il se peut que ce phénomène soit en partie dû au fait que les articles les plus populaires font souvent l'objet d'offres spéciales destinés à augmenter les ventes.

V. CONCLUSION

10. S'il était possible d'utiliser pleinement les données scannées en éliminant les problèmes de coût, la méthode idéale consisterait à calculer un indice élémentaire en construisant un indice «superlatif» prenant en compte tous les articles. Cela suppose toutefois que soient résolus les problèmes posés par les nouveaux articles et l'ajustement de qualité, ce qui n'est pas le cas. Bien que les indices élémentaires soient généralement calculés pour un ou plusieurs articles particuliers, notre expérience montre que cette méthode peut aboutir, même si ce n'est que dans un nombre restreint de cas, à un indice biaisé à la baisse. Ce résultat peut être dû à la forte influence de la situation économique récente du Japon, caractérisée par un ralentissement de la demande dans un contexte déflationniste, et il n'est pas certain que l'on observe un phénomène analogue en cas d'augmentation des prix. L'IPC réel prend en compte, outre les données scannées, les dates précises auxquelles des relevés sont effectués et néglige les rabais consentis pendant de courtes périodes, ce qui en principe, doit limiter le biais à la baisse. Pour certaines catégories, l'indice des articles représentatifs et parfois même celui de l'article le plus important sont tout à fait semblables à celui de l'indice de tous les articles. Utiliser de cette façon les données scannées peut permettre de valider la sélection des articles dans l'IPC réel et, partant, améliorer l'exactitude de ce dernier.

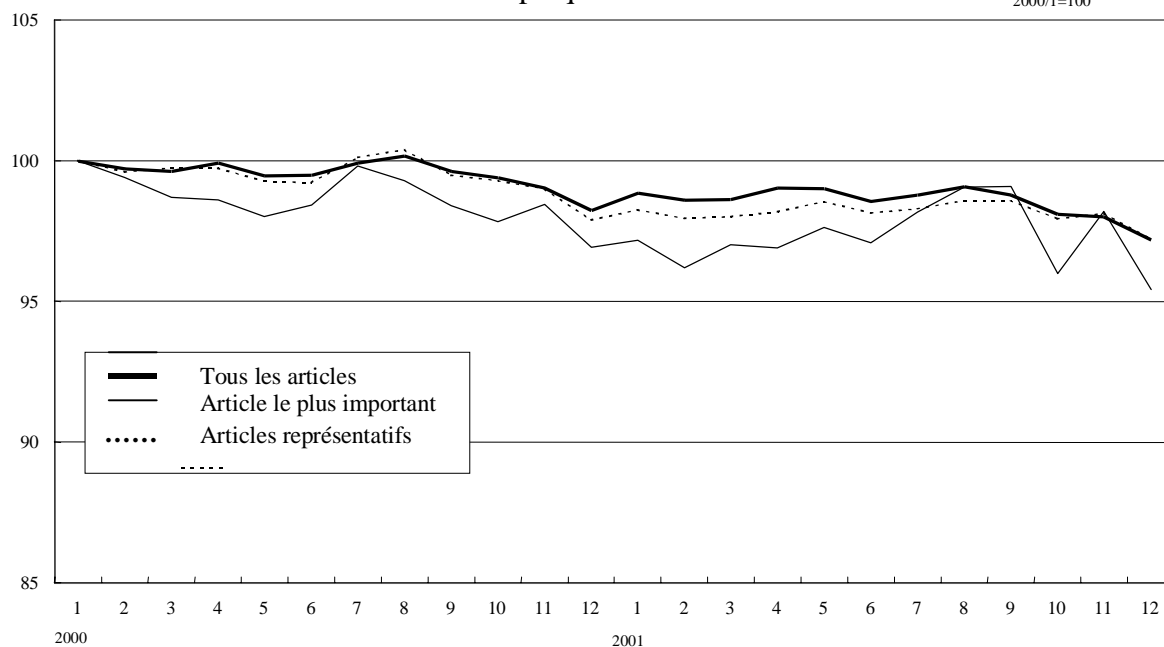
Tableau 3. Indice de groupe (janv. 2000 = 100)

		Nombre d'articles	Valeur totale des ventes en 2000 (yen)	Coefficient de pondération	Couverture	Indice 2000	Indice 2001	2001/2000 (% de variation)	Divergence
Indice global	Tous les articles	3 864	41 796 766 865	100	100,0	99,1	97,6	-1,6	-
	Article le plus important	24	5 141 421 721		12,3	98,8	96,2	-2,6	-1,0
	Articles représentatifs	120	14 631 885 377		35,0	99,1	97,1	-2,0	-0,4
Produits laitiers	Tous les articles	1 014	13 618 824 793	32,6	100,0	99,6	98,6	-1,0	-
	Article le plus important	3	1 142 983 006		8,4	98,6	97,3	-1,3	-0,3
	Articles représentatifs	15	3 595 664 055		26,4	99,5	98,1	-1,3	-0,3
Matières grasses et huiles	Tous les articles	204	2 741 996 717	6,6	100,0	99,0	97,0	-2,1	-
	Article le plus important	3	454 622 841		16,6	98,9	96,6	-2,3	-0,2
	Articles représentatifs	15	1 289 051 282		47,0	99,2	96,7	-2,5	-0,4
Condiments	Tous les articles	146	1 535 349 250	3,7	100,0	99,0	97,8	-1,3	-
	Article le plus important	3	606 664 598		39,5	99,1	97,1	-2,0	-0,7
	Articles représentatifs	15	1 085 520 990		70,7	99,1	97,9	-1,2	0,1

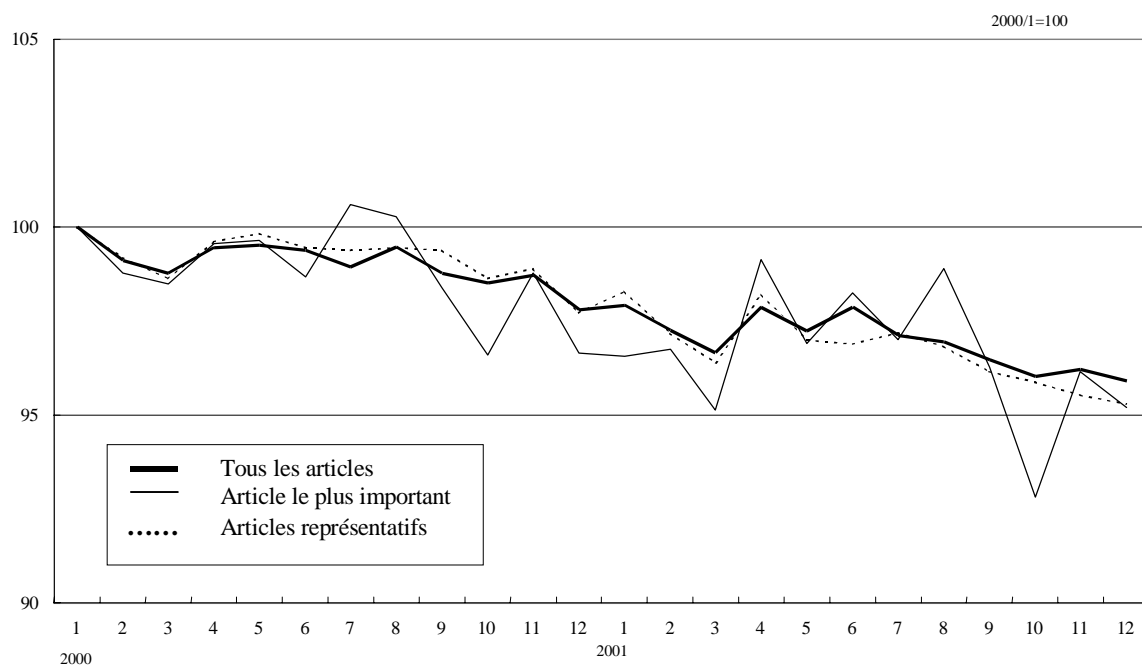
		Nombre d'articles	Valeur totale des ventes en 2000 (yen)	Coefficient de pondération	Couverture	Indice 2000	Indice 2001	2001/2000 (% de variation)	Divergence
Confiserie	Tous les articles	757	6 564 047 569	15,7	100,0	99,7	99,6	-0,2	-
	Article le plus important	3	353 449 533		5,4	100,7	100,1	-0,6	-0,4
	Articles représentatifs	15	1 115 877 192		17,0	100,0	99,4	-0,6	-0,4
Boissons	Tous les articles	933	8 768 307 077	21,0	100,0	98,8	96,9	-2,0	-
	Article le plus important	3	1 149 131 891		13,1	98,5	95,3	-3,3	-1,3
	Articles représentatifs	15	3 591 096 391		41,0	98,8	96,4	-2,4	-0,4
Produits d'entretien	Tous les articles	248	3 119 623 423	7,5	100,0	98,9	97,0	-1,9	-
	Article le plus important	3	870 951 264		27,9	99,3	96,0	-3,3	-1,3
	Articles représentatifs	15	2 108 358 789		67,6	98,4	96,5	-1,9	0,0
Fournitures médicales	Tous les articles	252	2 889 921 568	6,9	100,0	97,6	92,8	-4,9	-
	Article le plus important	3	261 125 145		9,0	94,9	84,4	-11,0	-6,1
	Articles représentatifs	15	963 555 916		33,3	96,6	91,0	-5,8	-0,9
Articles de toilette	Tous les articles	310	2 558 696 468	6,1	100,0	98,6	96,4	-2,3	-
	Article le plus important	3	302 493 443		11,8	98,8	96,5	-2,3	0,0
	Articles représentatifs	15	882 760 762		34,5	98,8	96,0	-2,9	-0,6

Graphique 2: Produits laitiers

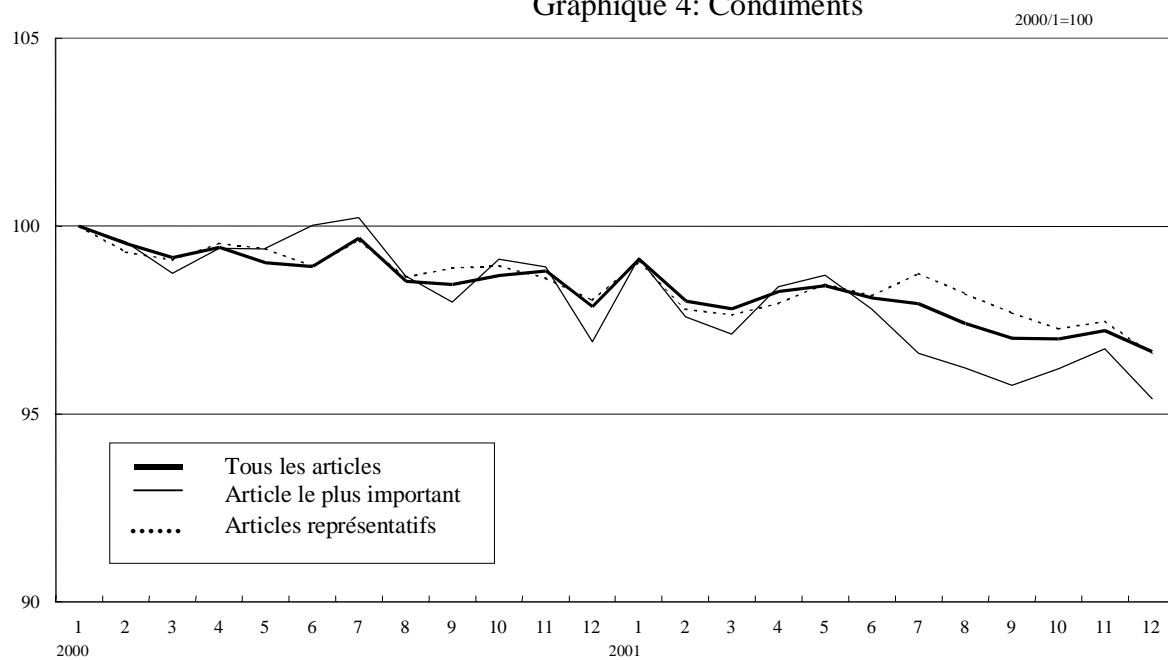
2000/1=100



Graphique 3: Matières grasses et huiles

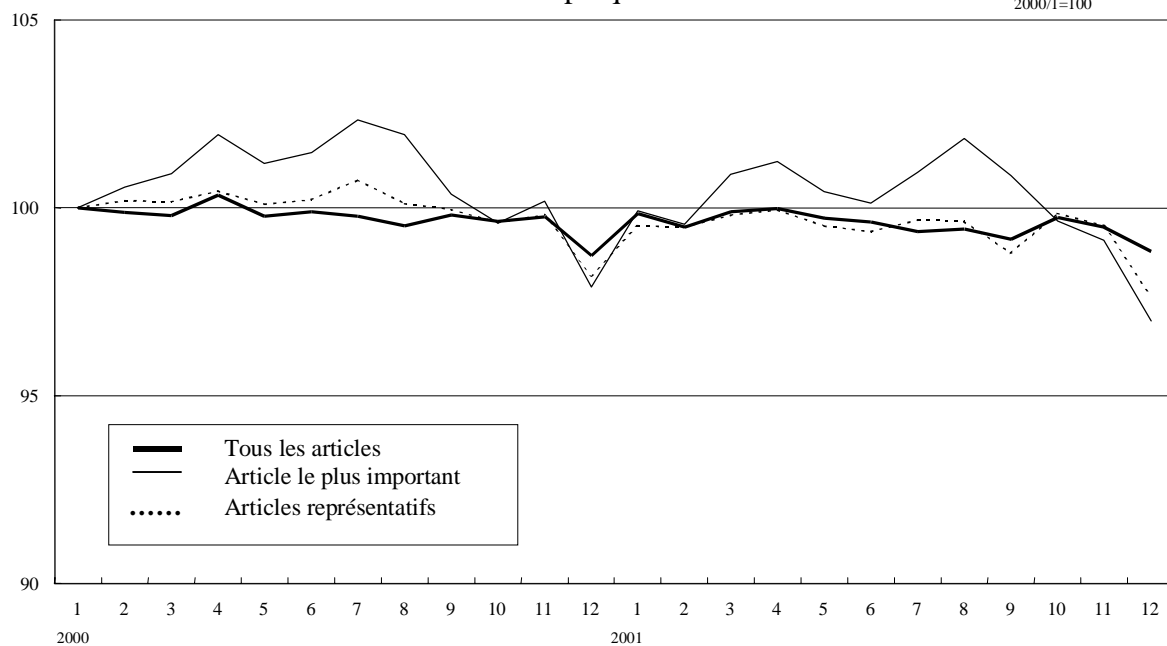


Graphique 4: Condiments



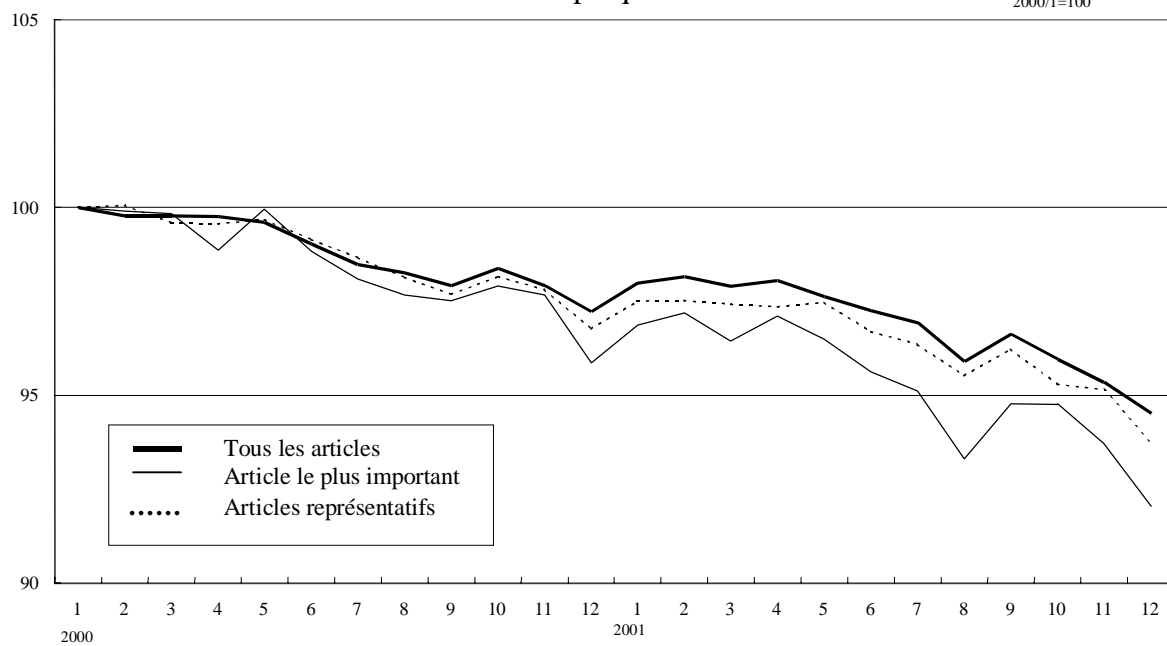
Graphique 5: Confiserie

2000/1=100



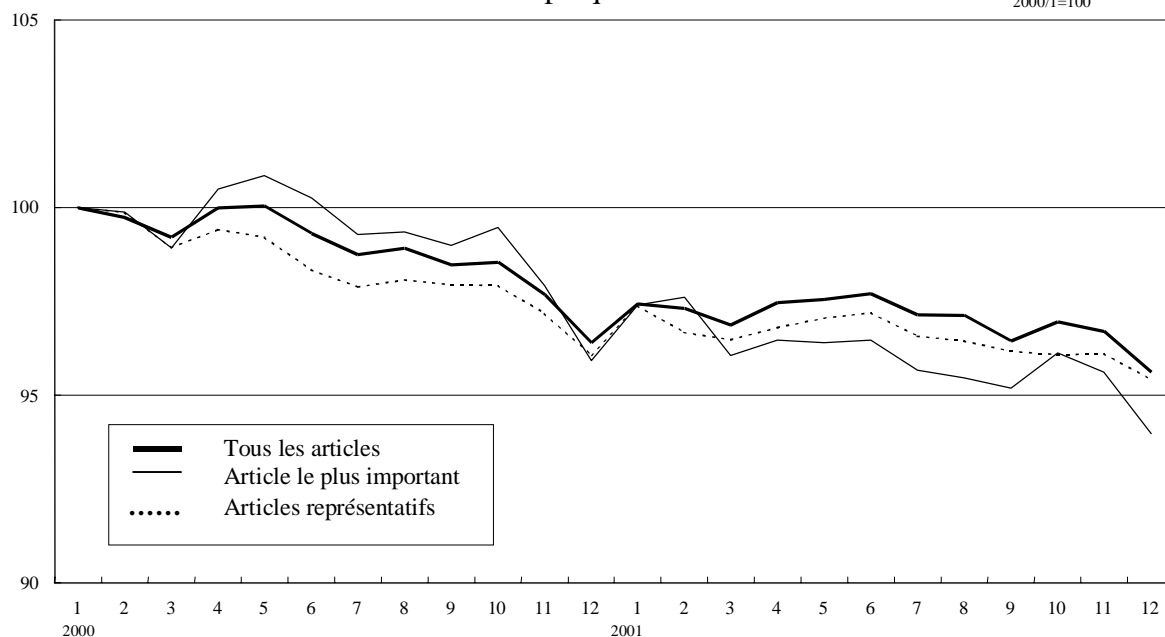
Graphique 6: Boissons

2000/1=100



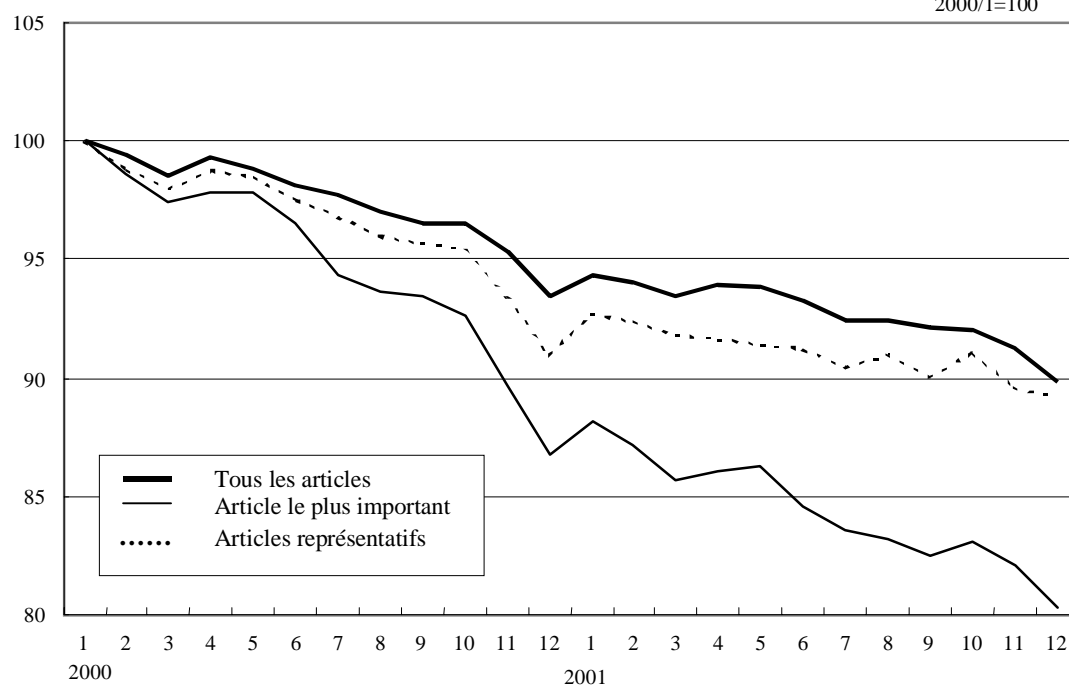
Graphique 7: Produits d'entretien

2000/1=100

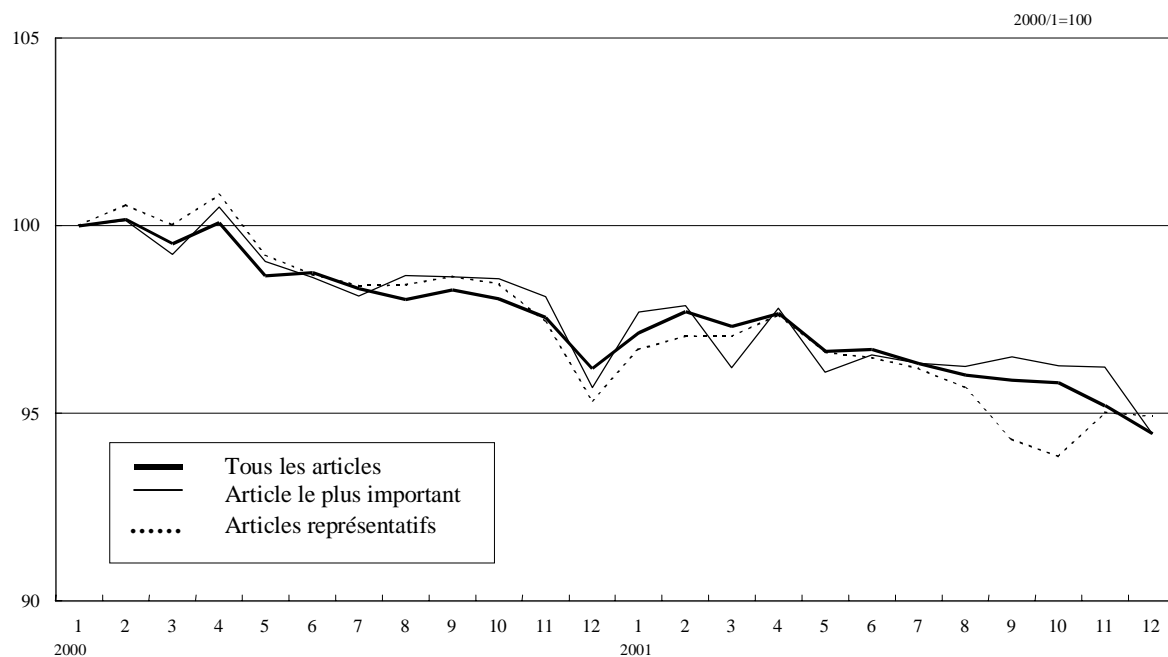


Graphique 8: Fournitures médicales

2000/1=100



Graphique 9: Articles de toilette



References

Okamoto, Masato, 1999 «Empirical Study of Outlet Sampling Using Scanner Data», presented at the Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, November 1999.

Haan, Jan de, 1999, «Item Sampling in the CPI : The Success of Cut-off Selection Methods», presented at the Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, November 1999.
