

Distr.  
GENERAL

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

Original: ENGLISH

**STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR  
EUROPE**

**INTERNATIONAL LABOUR  
ORGANISATION (ILO)**

**CONFERENCE OF EUROPEAN  
STATISTICIANS**

**Joint UNECE/ILO Meeting  
on Consumer Price Indices**  
**(Geneva, 4-5 December 2003)**

## **ANALYSIS ON ITEM SAMPLING USING SCANNER DATA**

Invited paper submitted by Ministry of public Management,  
Home Affairs, Posts and Telecommunications, Japan\*

### **Summary**

This paper deals with an analysis of item sampling using scanner data. We have calculated and conducted a comparative study for indices by three types of item sampling – (1) all items, (2) top item, (3) cut-off items in 8 groups and 24 categories of foods and daily goods.

## **I. INTRODUCTION**

1. The Consumer Price Index (CPI) is usually calculated by breaking down household

---

\* Prepared by Mr. Nobuyuki Sakashita and Mr. Shin-ichi Takahashi, Price Statistics Office, Consumer Statistics Division, Statistics Bureau.

consumption expenditure into groups and categories; giving weights to categories, which formulate bottom-level classification, and aggregating indices for categories to a general index. Categories are created by grouping the purchased by the households items into families, which are generally assumed to have similar price movements. In making indices for categories, representative items in the categories are shown to enumerators and prices of the same items are captured continuously. While it would be ideal to include all items within a category, it is difficult for both expense and physical reasons to collect prices of all items. On the other hand, scanner data are recently becoming ready for use. In most countries, real CPI uses scanner data solely for quality adjustment by hedonic method in IT equipment such as personal computers, due to restriction in cost and limitation of outlets. In Japan, the hedonic method is applied for PCs and digital cameras. Meanwhile, we studied the general issues related to the selection of outlets and items, and making indices for categories using scanner data in the past (Okamoto 1999). Studies using scanner data are also conducted in many other countries (e.g. de Haan 1999). The merit of using scanner data is that the prices and quantities for all items actually sold in the sample outlets are available. In this experiment, we make category indices, group indices and a total index for 8 groups and 24 categories of foods and daily goods using scanner data. Three methods for making category indices are examined – (1) “superlative” index formula (Törnqvist index) using all items, (2) indices by top 1 item and (3) geometric mean using plural items topping in sales amount. From these results, we consider the issues in item sampling.

2. Section 1 outlines the scanner data used, section 2 explains each sampling method and formula used for the experimented indices, section 3 presents the result of the experiment and section 4 is the conclusion.

## **II OUTLINE OF THE SCANNER DATA**

3. The outline of the scanner data used in this experiment is as follows:

- About 1,000 outlets are selected by stratified sampling from each region and each type of business (GMS, supermarket, drugstore, CVS) across the country.
- Scanner data are collected continuously from January 2000 to December 2001.
- 8 groups and 24 categories of foods and daily goods (see Table 1).
- Monthly average prices and value of sales for all items, which have sales performance in every week between Jan. 2000 and Dec.2001, in each category<sup>1</sup>.

**Table 1. Groups and categories**

Groups		Categories	
1	Dairy products	1	Fresh Milk
		2	Yogurt
		3	Cheese
2	Fats and oils	4	Edible Oil
		5	Butter
		6	Margarine
3	Seasonings	7	Vinegar
		8	Ketchup
		9	Mayonnaise
4	Confectionary	10	Biscuits
		11	Chocolate
		12	Ice cream
5	Beverages	13	Instant coffee
		14	Tea beverages
		15	Fruit and vegetable juice
6	Domestic utensils	16	Kitchen detergent
		17	Laundry Detergent
		18	Fabric softener
7	Medical supplies	19	Disposable diapers
		20	Sanitary products
		21	Contact lens care products
8	Toilet articles	22	Soap
		23	Shampoo
		24	Toothpaste

### III SAMPLING METHODS AND FORMULAE

#### All-item indices

4. In all-item indices, we use all items actually sold in the sample. There would be a variety of formulae to calculate indices – i.e. simple arithmetic mean or geometric mean, weighted arithmetic mean or geometric mean. Here we use the Törnqvist index formula, which is one of the “superlative” index formulae, as we can get the total value of sales for every month from scanner data. To be more precise, for each category, we adopt a chain index formula which calculates indices as the average price of each item weighted by the average sales for the current month and the month before, and chaining these indices as factorials every month.

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

where  $i$  stands for item,  $j$  for category,  $t$  for current period,  $\bar{p}$  for average price and  $w$  stands for the value of sales.

5. Group indices and the total index are calculated from the indices for the relevant categories. For integrated indices in upper level, the Laspeyres formula is adopted, weighted by the total value of sales in the year 2000 for each item<sup>2</sup>.

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

#### Top-item indices

6. For each category, we select one item which represents the category and calculate a top item index from that item. The item to be selected is the one, which has the largest annual value of sales in the year 2000. Indices for each month are calculated as the ratio of average prices of the corresponding item.

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

where  $\bar{p}_{01}$  is the average price of the top item in the base period and  $\bar{p}_{t1}$  is the corresponding price in the current period. The calculation of group indices and of the total index is the same as (1).

#### Cut-off-item indices

7. For the cut-off-item indices, unlike PPS-method, plural topping items, which are highly representative, are selected. Representativeness is assessed on the basis of the annual value of sales in the year 2000, and the top 5 items are selected (n=5). The indices for each month are calculated as a geometric mean of the ratio of the average prices of the corresponding items.

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

The calculation of group indices and of the total index is the same as (1).

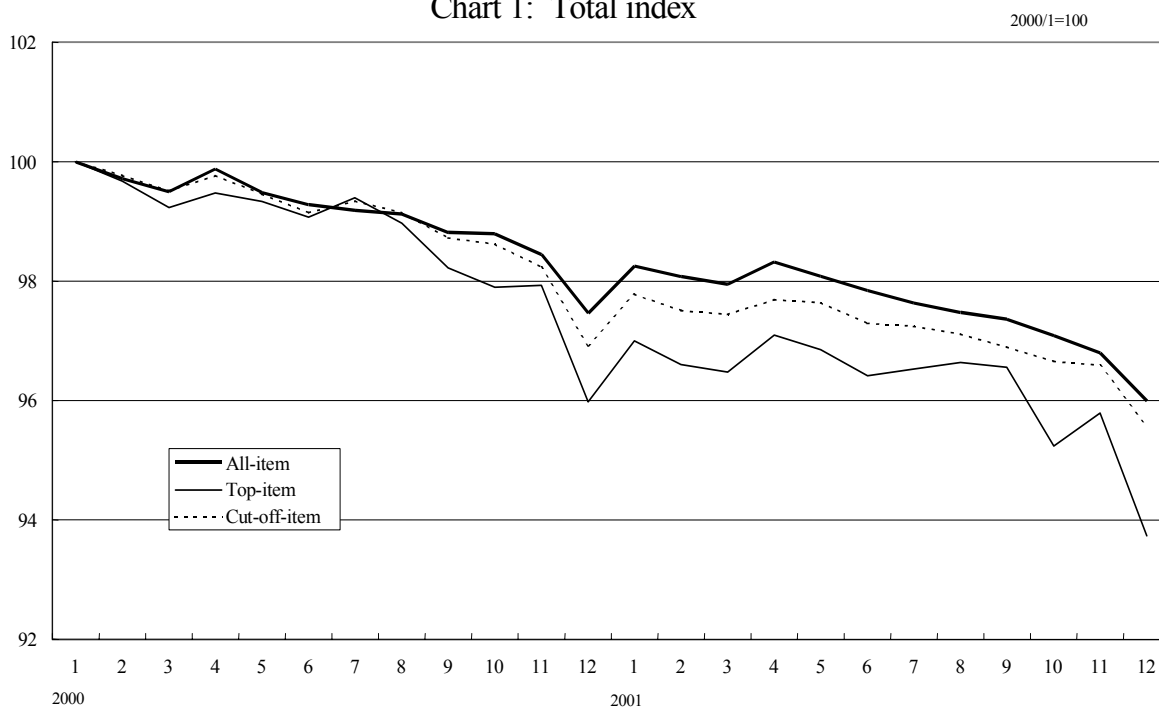
## **IV RESULTS**

8. The all-item index is declining, reflecting the recent falling trend of prices in Japan. For the top-item index and the cut-off-item index, the direction of movement is almost the same as for the all-item index. The top-item index, however, has a higher rate of decline than that of the all-item index. The cut-off item index, also has a higher rate of decline than that of the all-item index, but the discrepancy is narrower compared with the top-item index.

**Table 2. Total index (Jan.2000 = 100)**

	Index		Y2001/Y2000 (% change)	Discrepancy
	Y2000	Y2001		
All item	99.1	97.6	-1.6	-
Top item	98.8	96.2	-2.6	-1.0
Cut-off item	99.1	97.1	-2.0	-0.4

**Chart 1: Total index**



9. For groups and categories, although both top-item and cut-off-item indices seems to have similar directions of movements as the all-item index, most of them have downward discrepancy. Especially in groups and categories with strong and continuous downward trend, discrepancies tend to become wider. In the background of this phenomenon there may be a tendency, that popular items tend to be subject to bargain sale as an action to increase sales.

## V CONCLUSION

10. If we could fully apply scanner data clearing cost problems, it would be ideal to calculate elementary index by the “superlative” index formula using all items. The problems of new goods and quality adjustment, however, have to be solved. Although elementary indices are generally calculated from one or more particular items, our experiment shows that this method may cause a downward bias however in restrictive cases. This result may be caused by the strong effect of Japan’s recent economic situation i.e. depressed demand under deflationary trend, and it is not clear whether a similar tendency can be seen in the case of price increases. In the actual CPI, apart from scanner data, survey dates are specified and

downward biases are expected to be constrained by omitting short-time bargain price. In some categories, the cut-off-item index, sometimes even the top-item index are quite similar to the all-item index. By using scanner data like this, the selection of items in real CPI may be validated, and thus improvement of accuracy in the real CPI may be expected.

**Table 3. Group index (Jan.2000 = 100)**

		The number of items	Total value of sales in Y2000 (yen)	Weight	Coverage	Index Y2000	Index Y2001	Y2001/Y2000 (% change)	Discrepancy
Total Index	All-item	3864	41,796,766,865	100	100.0	99.1	97.6	-1.6	-
	Top-item	24	5,141,421,721		12.3	98.8	96.2	-2.6	-1.0
	Cut-off-item	120	14,631,885,377		35.0	99.1	97.1	-2.0	-0.4
Dairy Products	All-item	1014	13,618,824,793	32.6	100.0	99.6	98.6	-1.0	-
	Top-item	3	1,142,983,006		8.4	98.6	97.3	-1.3	-0.3
	Cut-off-item	15	3,595,664,055		26.4	99.5	98.1	-1.3	-0.3
Fats and oils	All-item	204	2,741,996,717	6.6	100.0	99.0	97.0	-2.1	-
	Top-item	3	454,622,841		16.6	98.9	96.6	-2.3	-0.2
	Cut-off-item	15	1,289,051,282		47.0	99.2	96.7	-2.5	-0.4
Seasoning	All-item	146	1,535,349,250	3.7	100.0	99.0	97.8	-1.3	-
	Top-item	3	606,664,598		39.5	99.1	97.1	-2.0	-0.7
	Cut-off-item	15	1,085,520,990		70.7	99.1	97.9	-1.2	0.1
Confectionary	All-item	757	6,564,047,569	15.7	100.0	99.7	99.6	-0.2	-
	Top-item	3	353,449,533		5.4	100.7	100.1	-0.6	-0.4
	Cut-off-item	15	1,115,877,192		17.0	100.0	99.4	-0.6	-0.4
Beverages	All-item	933	8,768,307,077	21.0	100.0	98.8	96.9	-2.0	-
	Top-item	3	1,149,131,891		13.1	98.5	95.3	-3.3	-1.3
	Cut-off-item	15	3,591,096,391		41.0	98.8	96.4	-2.4	-0.4
Domestic utensils	All-item	248	3,119,623,423	7.5	100.0	98.9	97.0	-1.9	-
	Top-item	3	870,951,264		27.9	99.3	96.0	-3.3	-1.3
	Cut-off-item	15	2,108,358,789		67.6	98.4	96.5	-1.9	0.0
Medical supplies	All-item	252	2,889,921,568	6.9	100.0	97.6	92.8	-4.9	-
	Top-item	3	261,125,145		9.0	94.9	84.4	-11.0	-6.1
	Cut-off-item	15	963,555,916		33.3	96.6	91.0	-5.8	-0.9
Toilet articles	All-item	310	2,558,696,468	6.1	100.0	98.6	96.4	-2.3	-
	Top-item	3	302,493,443		11.8	98.8	96.5	-2.3	0.0
	Cut-off-item	15	882,760,762		34.5	98.8	96.0	-2.9	-0.6

### NOTES

<sup>1</sup> The available scanner data is a simple mean price and the total value of sales for each type of business across the county. In this experiment, average prices for all types of business across the country are calculated by arithmetic mean weighted by the number of sample outlets.

<sup>2</sup> Superative index using scanner data is not adopted for the upper level integration of indices. This is because, in this experiment, the Laspeyres formula with fixed weight would be closer to the ongoing CPI as household consumption expenditure obtained from Family Income and Expenditure Survey etc. is used as weights for categories there. Thus, total value of sales calculated from scanner data used as weights for this purpose has a meaning of a proxy for household consumption expenditure.

Chart 2: Dairy products

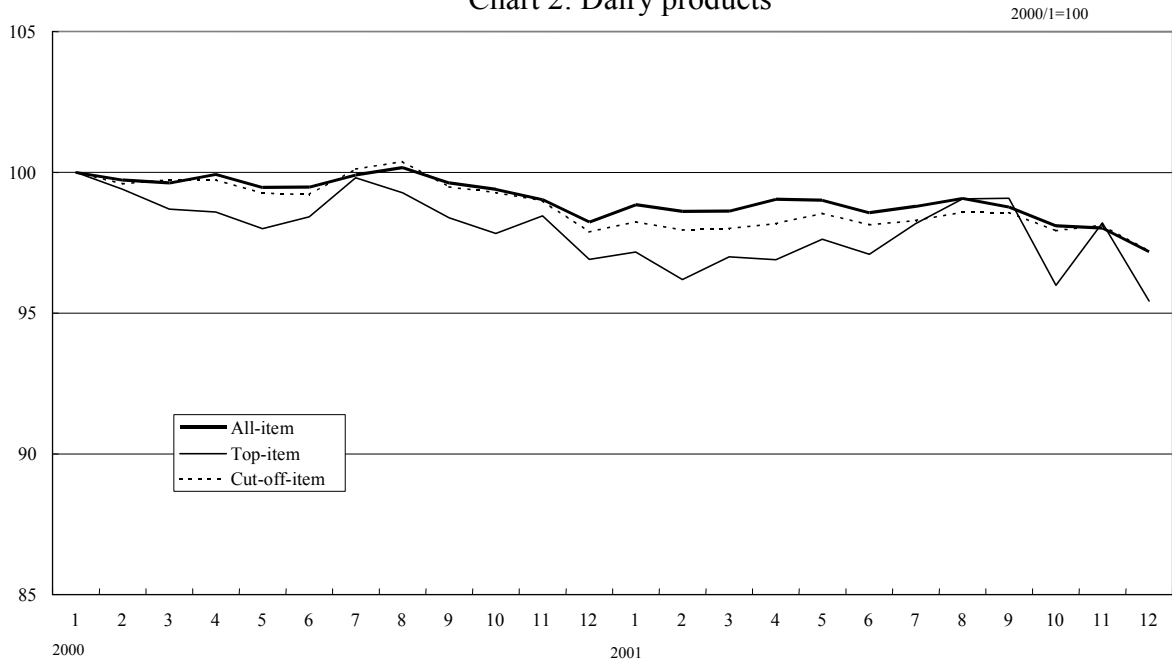


Chart 3: Fats and oils

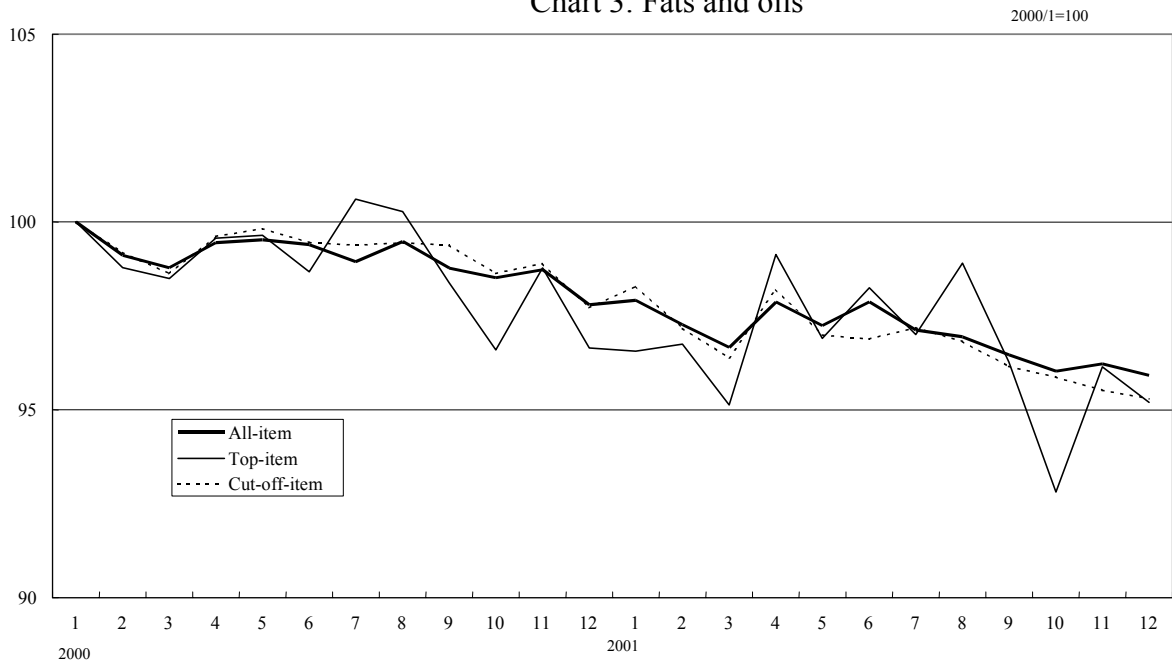


Chart 4: Seasonings

2000/1=100

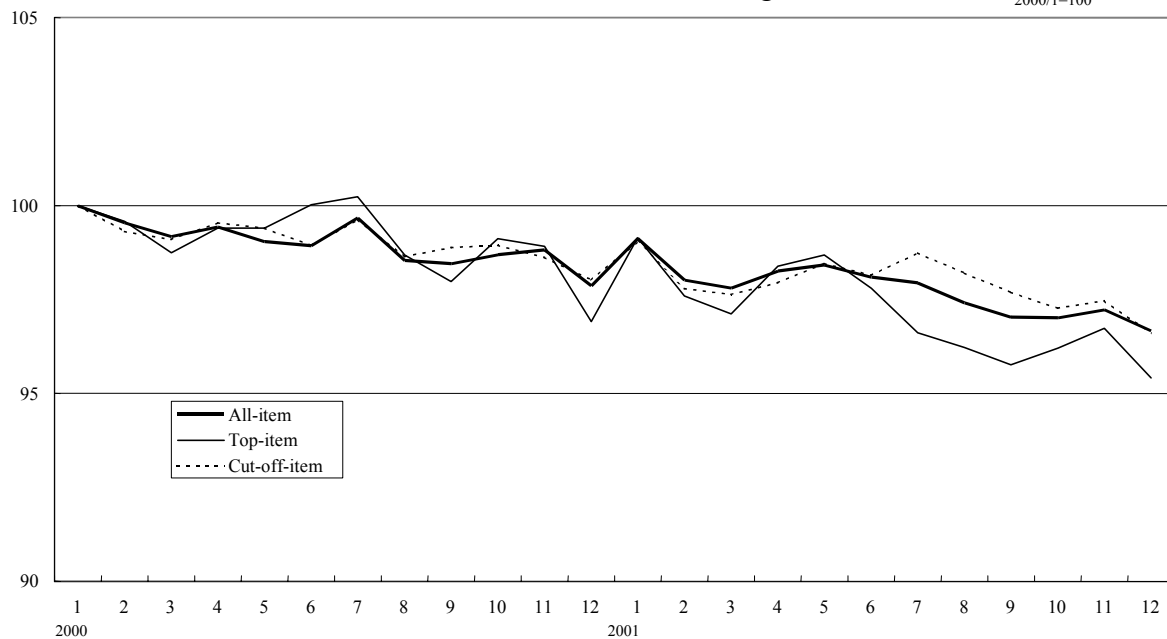


Chart 5: Confectionery

2000/1=100

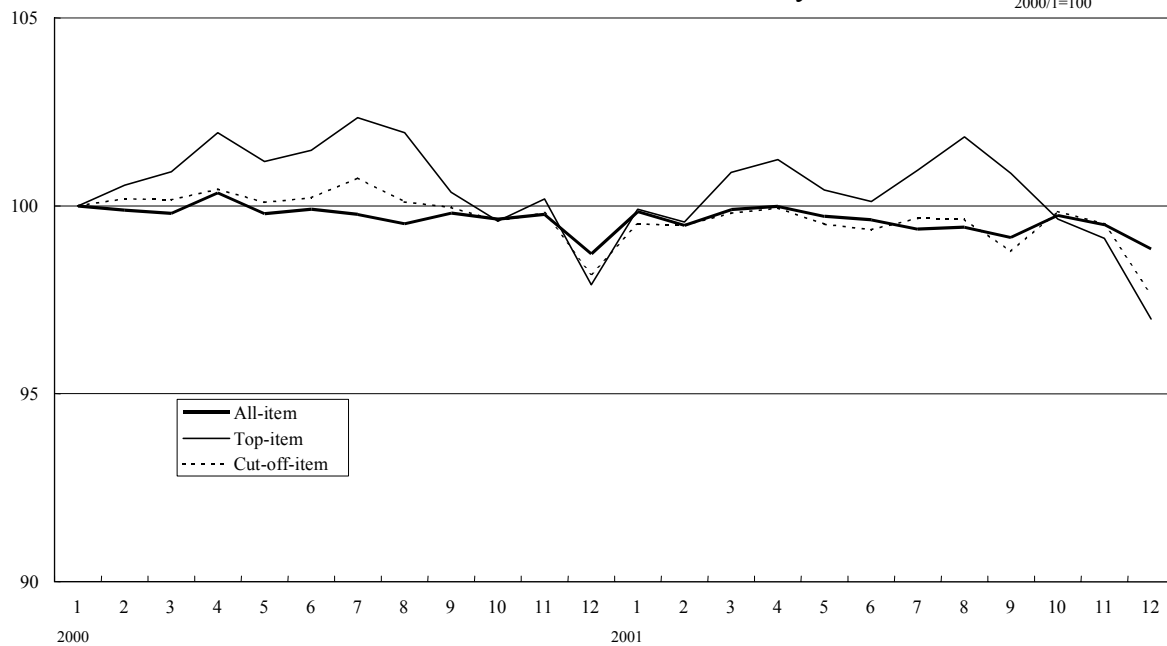


Chart 6: Beverages

2000/1=100

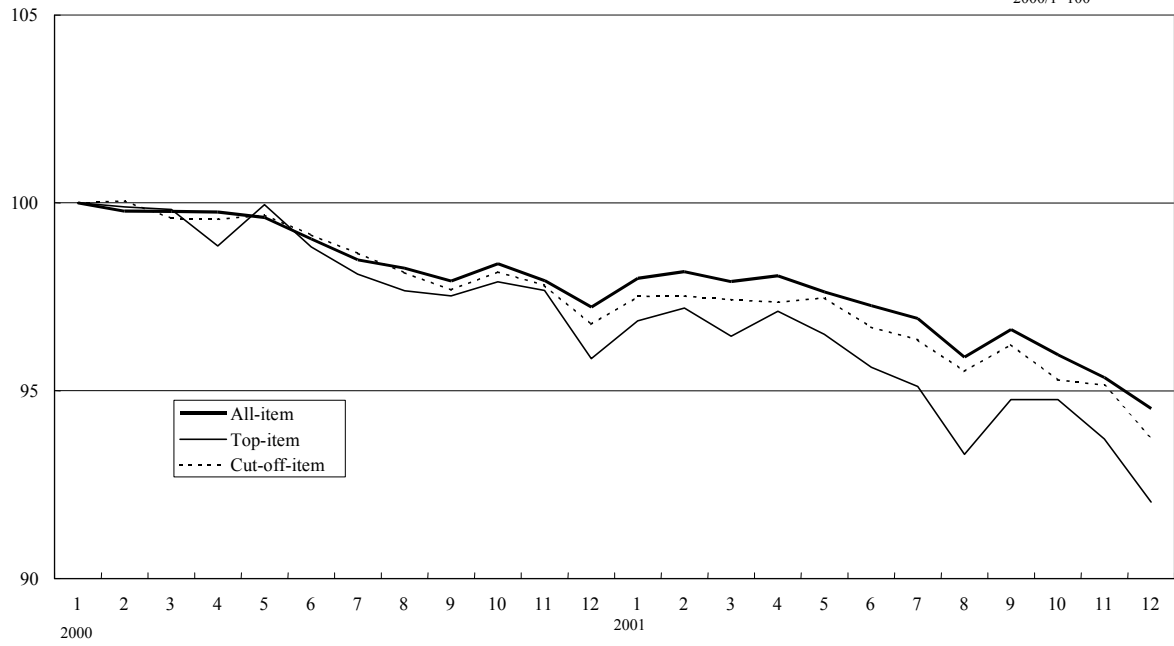


Chart 7: Domestic utensils

2000/1=100

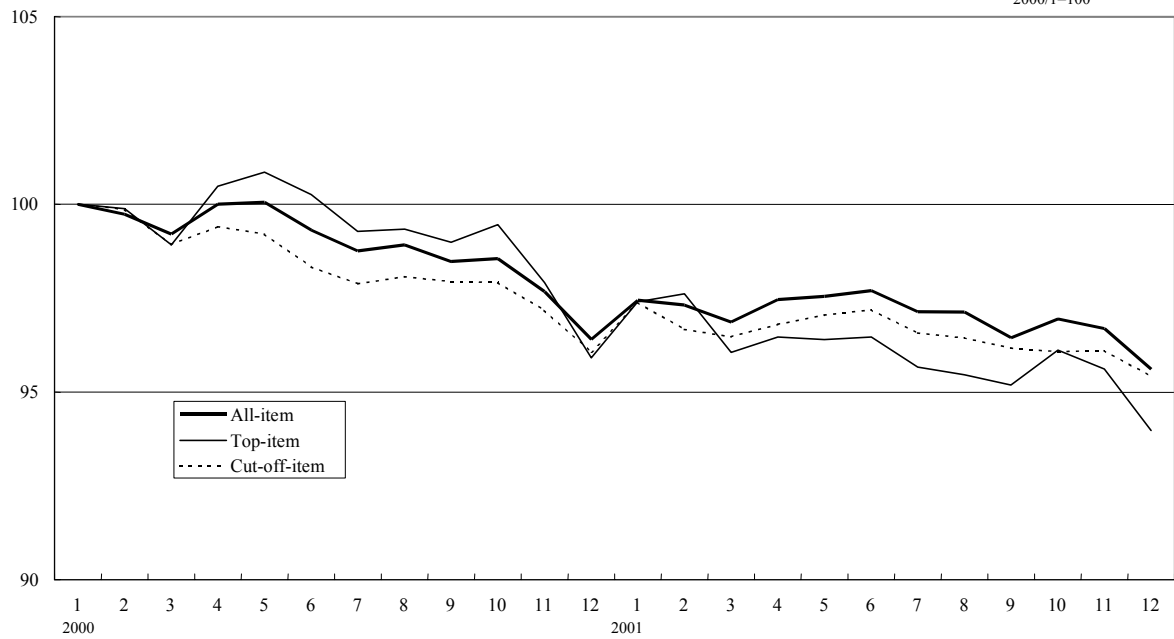


Chart 8: Medical supplies

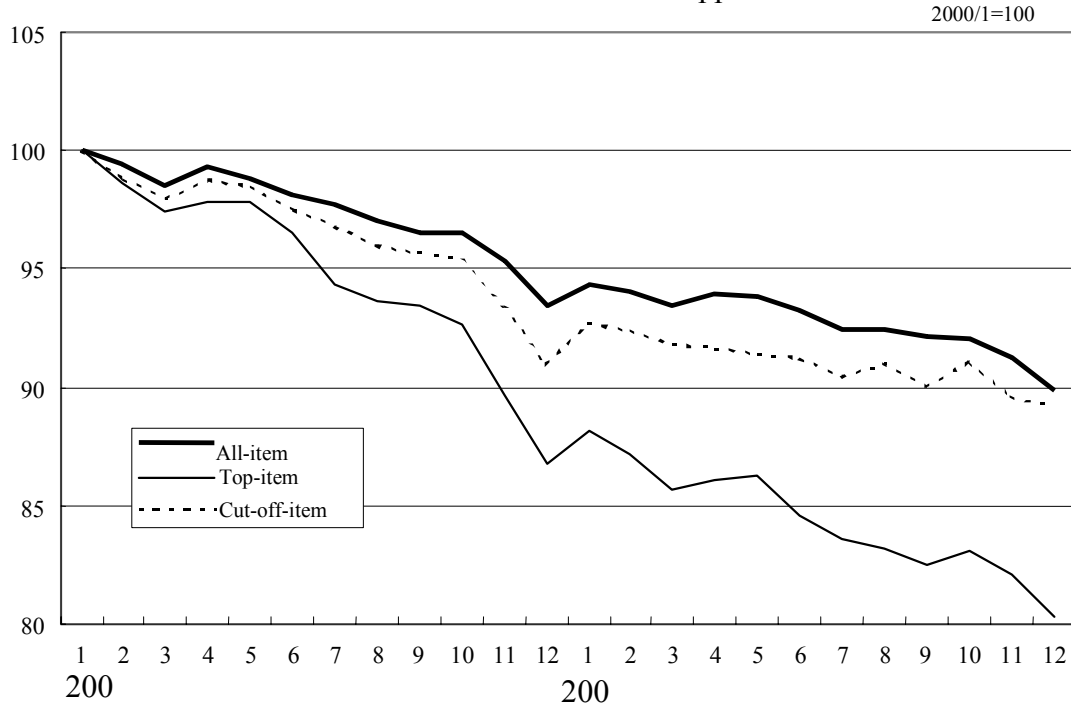
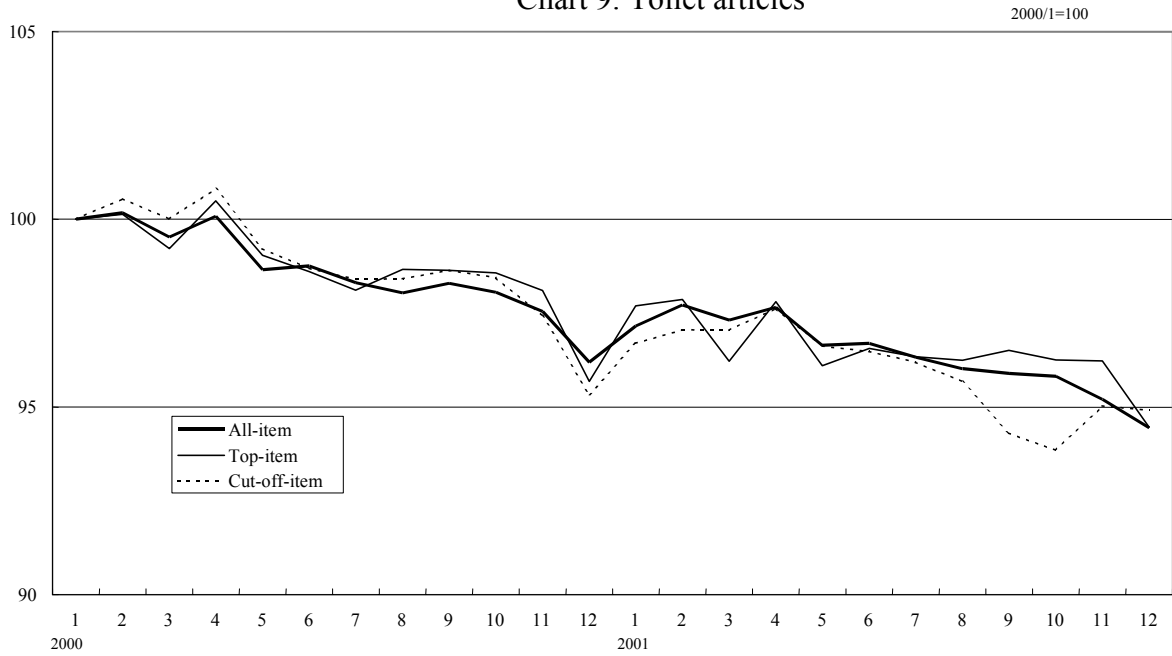


Chart 9: Toilet articles



### 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.

-----