Consumer Price Index Biases*

– Elementary Index Biases vs. Sampling Biases –

Satoshi Imai[†]· Erwin Diewert[‡]· Chihiro Shimizu[§]

May 15, 2015

Abstract

Consumer price indices (CPIs) are one of the economic statistics that is most closely linked with policy. As a result, various policy debates have arisen surrounding the accuracy and definitude of these indices. In this regard, theoretical research has been undertaken to determine the requirements of the ideal price index; for the most part, however, the biases in price indices occur due to practical constraints. In this paper, in terms of the causes of CPI biases, we distinguished between biases based on the calculation formula (elementary index biases) and biases based on sampling (sampling method biases), with the aim of clarifying the extent of the deviation caused by each type of bias. Specifically, using scanner data, we applied nine types of elementary indices and three types of aggregation formulae to price groups sampled using four different methodologies (two types of sample size, two types of product variation) and compared the differences between the various indices. The obtained results made it clear that considerable biases may result based on the price sampling method in addition to the effect of the calculation formula on the elementary index. Furthermore, even if the same formula and the same sampling method is used, considerable biases may result based on the sample size. These findings suggest the importance of weighting and selection of the number of representative products in order to construct more accurate indices.

Key Words: consumer price indices, purposive sampling, Laspeyres, Paasche, Fisher, Lowe, Törnqvist, Carli, Dutot, Jevons, Weighted Jevons

JEL Classification: C43, C81, E01, E31.

* In preparing this paper, we received scanner data from Nikkei Digital Media. In addition, we received many suggestions based on discussion with Tsutomu Watanabe. We hereby express our gratitude to them.

† Statistics Bureau of Japan.

‡ University of British Columbia and University of New South Wales.

§ National University of Singapore and University of British Columbia.

1 Introduction

When measuring economic indices of any kind, there are always biases. In some cases, the issue is that the true index is not measurable, but in many cases, the biases are caused by research design and related issues such as technical limitations and budget constraints. Of course, even if one takes the existence of such research design related issues as a given, it is still necessary to continuously improve indices with the aim of increasing their accuracy and relevance. In particular, given the growing interrelationship of economic indices and policy, if biases are present in economic statistics, it is extremely important to provide index users with some indication about the magnitude of possible biases. In addition, if the factors causing such biases are made clear, it will be easier to make future improvements.

Consumer price indices (CPIs) in particular are one of the indices in economic statistics that is most closely linked with policy. Not only do they measure changes in prices that household is facing, but they also have a strong influence on fiscal policy, the provision of public pensions, and so forth. As a result, many policy debates have arisen surrounding price index biases.

In this regard, theoretical research has been undertaken to determine the requirements of the ideal price index, but in practice, there are still many problems remaining. In many cases, the underlying reasons for the existence of biases in price indices are a result of practical constraints. First, one faces sampling problems—what kind of price data and quantity data should be collected. Then there are problems arising from the question of what formula should be used to aggregate the data, based on the availability of information, technical limitations, and so forth. In this paper, we divided possible CPI biases into biases based on differences in the calculation method (elementary index biases) and biases based on sampling (sampling method biases), with the aim of clarifying the magnitude of each type of bias.

It is important to note here that the purpose of this paper is not to measure the biases possessed by public price indices. Existing public price indices are operated under various systemic restrictions, often relating to consistency with past indices, budget constraints, and technical limitations. For example, since quantity weights are determined independently of price surveys, it is difficult to update them in real time. In such cases, even if it is known that the ideal price index is based on the Fisher formula, calculations must be based on the Laspeyres formula (in practice, the Lowe formula). Given these issues, we measured price index sampling method biases and elementary index biases on an experimental basis using scanner data. Using scanner data, it possible to calculate indices based on a wide variety of index number formulae that have been suggested in the literature. In addition, we were able to run tests relating to sampling by experimentally modifying the number of outlets and selection of products. The purpose of this paper is to observe how changing the sampling method and the choice of elementary index formula changes the resulting price index. While the kind of computations undertaken in this paper do not provide estimates of "biases" in actual Consumer Price Indexes, the results presented in this paper may be helpful in improving actual price indexes in the future.

First, let us consider elementary indices. As indicated previously, many statistical agencies measure price indices based on the Laspeyres formula (in practice, the Lowe formula). However, besides this method, many other formulae have been proposed depending on the index's purpose, such as Carli, Dutot, Jevons, weighted Jevons, Törnqvist, Lowe, Paasche, and Fisher indices. As Diewert (1976[6], 1978[7]) has made clear, indices based on the Fisher formula are known as ideal target indices in terms of consistency with utility theory. It has also been verified that the Törnqvist index approximates indices obtained using the Fisher formula (and moreover, the Törnqvist index can also be justified from the viewpoint of traditional consumer theory). Compared to these "ideal" target indices, in theory, Laspeyres-based indices have an upward bias and Paasche-based indices have a downward bias. This description is purely theoretical, however—with regard to whether these biases actually exist, the results obtained in previous research efforts have been varied. Thus, in this paper, we calculated these indices (and many others) and thus we hope to add to the empirical literature on possible biases in the CPI.

Biases in sampling methods used for CPIs are also discussed in CPI manuals.[*1] With regard to problems based on purposive sampling, significant differences still exist at the item level, but it is said that these may be ignored at upper levels of aggregation. Dalén (1998)[4] indicated that while biases that could not be ignored existed in item level indices for items in 100 groups sold at supermarkets, these biases canceled at higher levels of aggregation. In addition, De Haan, Opperdoes and Schut (1999)[5] tested three categories (coffee, babies' napkins and toilet paper) and indicated that biases were less with non-probability-based sampling and, given that the biases occurred in both directions, they canceled out. These findings suggest that there are still biases in item level indices that cannot be ignored.

These discussions focused mainly on biases when using purposive sampling, not probabilistic random sampling. This is because most national statistics offices have adopted purposive sampling as their method. In this regard, Imai, Shimizu and Watanabe (2012)[12], using similar data to that used in this paper and including not just purposive sampling but also random sampling, which is used by the U.S. BLS, compared and evaluated the effect that index calculation methods have on index results using 64 different procedures. The most outlying results showed annual deflation of around 1%, which was double the figure for the CPI. In other words, this suggests the possibility that significant changes in CPI inflation can occur based on differences in sampling.

Meanwhile, Handbury, Watanabe and Weinstein (2013)[11], using similar data to the data used in this paper, evaluated CPI biases by means of comparison with the Törnqvist index. Their findings indicated that biases were not fixed: they were stable if the index deviated upward from the vicinity of 0, but if the index was in the vicinity of 0 (i.e., less than 2.4% per year), the error effect could no longer be ignored.

In addition, with respect to the Japanese CPI, Broda and Weinstein (2007)[3] and Ariga and Matsui (2003)[2] have pointed out that the traditional Japanese CPI based on the ILO manual has an upward bias that cannot be ignored.

Drawing on the above discussion, this paper will test how price indices are affected by differences in the sampling method and formula and clarify the direction and size of the biases. In particular, we will evaluate what impact the choice of elementary index and sampling method has on indices. The formulae used for the purpose of comparison in this paper are the Carli, Dutot, Jevons, Weighted Jevons, Törnqvist, Lowe (annual weight), Laspeyres, Paasche, and Fisher formulae. Second, we will test whether the price sampling method assigned to the formula has an effect on the index. Based on previous findings, it may be expected that fixed-base elementary index biases will grow larger over time. At the same time, it may be expected that biases due to the price sampling method will be canceled out when aggregated at the upper level.

[*1] http://www.ilo.org/public/english/bureau/stat/download/cpi/corrections/chapter5.pdf 5.29

2 Empirical Analysis

2.1 Data

For this paper, we used a Nikkei Point of Sale (Nikkei POS) dataset. For foods (excluding fresh food) and daily groceries, sales volume and sales total data are recorded daily by outlet and barcode in Nikkei POS. We used from January 2000 to July 2014 as the sample period.*2

The item categories used in the analysis are the 13 categories shown in Table 1. Three of these categories (chocolate bar, dried Japanese vermicelli, and clothing insect repellent) were arbitrarily selected categories whose sales volumes fluctuate based on seasonality.*3

Table 1 Overview of Categories

Code	Category Name
234001	Non-glutinous rice
137001	Cup instant noodle
031031	Bacon
046001	Fresh milk (1l)
041001	Butter
212001	Canned beer
601021	Shampoo (refill)
610001	Facial tissue (boxed)
612042	Liquid detergent for general clothing (refill)
818001	Dog food (dry type)
[Seasonal one]	
191001	Chocolate bar
107008	Dried Japanese vermicelli
620001	Clothing insect repellent

Table 2 shows the number of products and observations by year. For example, in 2000, there were 3,652 unique products and a total of 4,010,804 observations were recorded at the 100 outlets.

Table 3 shows the turnover of products at the 100 outlets. In 2000, there were 3,652 total products, which increased to 3,911 in 2001 due to the introduction of 1,265 new products and elimination of 1,006 products. Roughly speaking, the proportion of new products each year was 30%, while the proportion of eliminated products was 29%.*4

2.2 Data Sampling

To date, non-probability sampling based on the representative method has been used as a sampling method when creating price indices. In Japan, representative outlets are first picked out from 167 survey cities selected by means of stratified sampling. Purposive sampling of products at the selected outlets is then conducted. The Statistics Bureau of Japan specifies

*2 Nikkei POS data is based on data recorded at around 300 outlets across Japan, but in this paper, in order to perform a comparison over time, the analysis was limited to 100 outlets at which certain data were recorded throughout the data period.

*3 These 13 product categories account for about 2.1% of the applicable CPI's weight.

*4 Based on the results for all items, including other products, the turnover for these products was somewhat high. Abe et al. (2015)[1] and Imai et al. (2013)[13] have analyzed the effect of product turnover on price indices.

Table 2 Number of Outlets, Products and Observations

	No. of outlets	No. of products	No. of observations
2000	100	3,652	4,010,804
2001	100	3,911	4,272,381
2002	100	3,790	4,408,992
2003	100	3,851	3,987,877
2004	100	3,951	4,752,444
2005	100	3,832	4,839,600
2006	100	4,029	4,981,847
2007	100	4,139	5,093,593
2008	100	4,369	5,089,374
2009	100	4,220	5,075,707
2010	100	4,246	5,257,439
2011	100	4,592	5,272,547
2012	100	4,630	5,677,755
2013	100	4,943	5,780,280
2014*	100	4,523	3,369,542

* Jan. 2014 – Jul. 2014 only

Table 3 Turnover of Products in the 100 Outlets

	No. of products in the 100 outlets	Entries	Exits	Entry rate	Exit rate
2000	3,652	–	–	–	–
2001	3,911	1,265	1,006	0.323	0.257
2002	3,790	1,086	1,207	0.287	0.318
2003	3,851	1,219	1,158	0.317	0.301
2004	3,951	1,219	1,119	0.309	0.283
2005	3,832	1,013	1,132	0.264	0.295
2006	4,029	1,238	1,041	0.307	0.258
2007	4,139	1,258	1,148	0.304	0.277
2008	4,369	1,496	1,266	0.342	0.290
2009	4,220	1,197	1,346	0.284	0.319
2010	4,246	1,313	1,287	0.309	0.303
2011	4,592	1,463	1,117	0.319	0.243
2012	4,630	1,435	1,397	0.310	0.302
2013	4,943	1,588	1,275	0.321	0.258
2014*	4,523	1,030	1,450	0.228	0.321

* Jan. 2014 – Jul. 2014 only

the product specifications/brands that will be covered by the survey and price collectors carry out the sampling from a list of products that match these specifications/brands, keeping representativeness in mind. In other words, representative outlets are selected in survey cities based on the pre-determined sampling size, then one representative product is selected at each outlet.

Here, in order to compare the effect of sampling methods on indices, we created four categories based on two classes of category: 1) selection of products based on sales volume, and 2) selection of stores with a focus on the number of customers. We then performed a comparative analysis. The reason for this was that for statistical offices in most countries, certain restrictions apply when determining sampling size due to budget constraints, but we will observe what kind of effect occurs if the sample size is increased based on the assumption that no such budget constraints exist.

With regard to product selection, in the existing Japanese survey, one representative product is selected at each selected outlet. We will measure the effect on the price index if this restriction is eliminated and a strategy is adopted whereby multiple products are used from one outlet.

Eliminating the above two restrictions would be feasible if it became possible to use large-scale scanner data. In that sense, we believe that it is extremely important to understand in advance what kind of effect will be produced if large-scale scanner data is incorporated into the practice of price index estimation in the future.

The specifics of the sampling method we used are as follows. With regard to the selection of outlets, in order to compare outlets with a large share of the price data used at a given month t to those with a small share, we sampled outlets according to the size of their monthly customer base. This was done in order to look at the effect of selecting more low-popularity outlets for month t by expanding the number of outlets to be used in creating the index. That is, it is possible that low-popularity outlets may adopt different pricing strategies from high-popularity outlets, and we wanted to examine the effect of including such outlets in the survey.

With regard to selecting products at the target outlets sampled based on the above procedure, we used two methods: 1) selecting only the top-ranked product in terms of sales volume, and 2) selecting the top five products. This was done because we anticipated that the pricing strategy for the most salable product would differ from the pricing strategy for less salable products.

Table 4 shows the number of outlets combined with the number of selected product items. For Sampling Methods A and B, we selected the top five best-selling products at each selected outlet, and by selecting 10 outlets and 20 outlets respectively for each group, we obtained 50 and 100 price observations for month t.[*5] In contrast to the typical method of price collection by price collectors, this was done in order to reflect a broader range of product substitution in the index calculations by using weights corresponding to the sales volume in the available data.

Table 4 Sampling Method

	No. of outlets	No of items from each outlet	Monthly observations
Sampling A	10	Top 5	50
Sampling B	20	Top 5	100
Sampling C	50	Top 1	50
Sampling D	100	Top 1	100

Meanwhile, for Sampling Methods C and D, we selected the best-selling product at the selected outlet, and by selecting 50 outlets and 100 outlets respectively for each group, we obtained 50 and 100 price observations for month t.

This simulates cases where existing national statistics offices calculate indices based on prices collected by conducting price surveys.

Table 5 shows the price count ratio and number of observations for prices collected based on the four sampling approaches. When no sales record exists in a specific category at a specific

[*5] This method makes it possible to look at the effect of selecting many products even though the number of survey outlets is small. If the use of scanner data progresses, there is a high likelihood that it will be possible to use information relating to more products by employing large-scale data.

point in time and no price can be collected, a missing price occurs. The table contents show the average price count ratio and standard deviation for each category over the 175 months of the data period. For some items in Sampling A and B, the variety of products was limited and it was not possible to collect data for the top five products. Due to these cases of missing values, the number of observations is smaller. As well, for category 212001 (beer), there were some outlets that did not stock alcohol, so the price count ratio was somewhat lower.

Table 5 Price Count Ratio and Number of Observations

Category	sampling A		sampling B		sampling C		sampling D	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
234001	0.995	0.015	0.998	0.008	0.998	0.009	0.995	0.010
107008	0.848	0.171	0.838	0.172	1.000	0.000	1.000	0.002
137001	1.000	0.000	0.995	0.015	0.999	0.004	0.999	0.004
031031	1.000	0.000	0.995	0.015	0.995	0.017	0.997	0.011
046001	0.995	0.012	0.997	0.009	1.000	0.000	1.000	0.000
041001	1.000	0.000	1.000	0.002	1.000	0.000	1.000	0.000
191001	1.000	0.004	1.000	0.002	1.000	0.000	1.000	0.000
212001	0.903	0.116	0.872	0.091	0.873	0.114	0.874	0.124
610001	0.999	0.005	0.998	0.005	0.996	0.008	0.998	0.004
612042	1.000	0.003	0.993	0.015	0.989	0.010	0.990	0.006
620001	1.000	0.000	0.998	0.004	0.985	0.009	0.978	0.015
818001	1.000	0.000	1.000	0.000	0.985	0.009	0.983	0.006
601021	1.000	0.000	1.000	0.000	0.986	0.009	0.990	0.006
Total No. of observations	118,343		235,574		118,988		237,706	

As explained above, target outlets and target products are selected for month t and the relevant prices and quantities are sampled. The quantity is the total volume of monthly sales of the given target product at the target outlet in month t. With regard to price, on the other hand, the daily unit price is determined based on daily sales records in scanner data, and the mode value for month t is used as the price for that month.

2.3 Price Index Formulae

The elementary indices which we use to calculate the item level indices in this paper are Carli, Dutot, Jevons, weighted Jevons, Törnqvist, Lowe (annual weight), Laspeyres, Paasche, and Fisher.

The formulae for these elementary indices are described below. The period t price and quantity vectors are defined as $\boldsymbol{p}^t \equiv [p_1^t, ..., p_N^t]$ and $\boldsymbol{q}^t \equiv [q_1^t, ..., q_N^t]$ respectively.

Carli:

$$P_C(\boldsymbol{p}^0, \boldsymbol{p}^t, \boldsymbol{q}^0, \boldsymbol{q}^t) \equiv \sum_{n=1}^{N} (1/N)(p_n^t/p_n^0) \quad \text{(Formula 1)}$$

Dutot:

$$P_D(\boldsymbol{p}^0, \boldsymbol{p}^t, \boldsymbol{q}^0, \boldsymbol{q}^t) \equiv \sum_{n=1}^{N} (1/N)(p_n^t) \Big/ \sum_{n=1}^{N} (1/N)(p_n^0) \quad \text{(Formula 2)}$$

Jevons:

$$P_J(\boldsymbol{p}^0, \boldsymbol{p}^t, \boldsymbol{q}^0, \boldsymbol{q}^t) \equiv \prod_{n=1}^{N} (p_n^t/p_n^0)^{1/N} \quad \text{(Formula 3)}$$

Weighted Jevons:

$$P_J(\boldsymbol{p}^0, \boldsymbol{p}^t, \boldsymbol{q}^0, \boldsymbol{q}^t) \equiv \prod_{n=1}^{N} (p_n^t / p_n^0)^{s_n^0} \qquad \text{(Formula 4)}$$

where $s_n^0 \equiv p_n^0 q_n^t \Big/ \sum_{j=1}^{N} p_j^0 q_j^t$

Törnqvist:

$$\ln P_T(\boldsymbol{p}^0, \boldsymbol{p}^t, \boldsymbol{q}^0, \boldsymbol{q}^t) \equiv \sum_{n=1}^{N} (1/2)(s_n^0 + s_n^t) \ln(p_n^t / p_n^0) \qquad \text{(Formula 5)}$$

Lowe:

$$P_{Lo}(\boldsymbol{p}^0, \boldsymbol{p}^t, \boldsymbol{q}) \equiv \sum_{n=1}^{N} (p_n^t / p_n^0) s_n^{0b} \qquad \text{(Formula 6)}$$

where $s_n^{0b} \equiv p_n^0 q_n^b \Big/ \sum_{j=1}^{N} p_j^0 q_j^b$.

The weight reference period for the Lowe index is one year prior to the price reference period. In other words, the reference quantity is the annual sales quantity from t-11 to 0, including the reference price period.

Laspeyres:

$$P_L(\boldsymbol{p}^0, \boldsymbol{p}^t, \boldsymbol{q}^0) \equiv \sum_{n=1}^{N} p_n^t q_n^0 \Big/ \sum_{n=1}^{N} p_n^0 q_n^0 \qquad \text{(Formula 7)}$$

Paasche:

$$P_P(\boldsymbol{p}^0, \boldsymbol{p}^t, \boldsymbol{q}^t) \equiv \sum_{n=1}^{N} p_n^t q_n^t \Big/ \sum_{n=1}^{N} p_n^0 q_n^t \qquad \text{(Formula 8)}$$

Fisher:

$$P_F^t \equiv [P_L^t P_P^t]^{1/2} \qquad \text{(Formula 9)}$$

Notation: n indicates the n-th price unit with respect to the sample size N (target price determined by the outlet and product). For the n-th price unit, p_n^0 indicates the reference period price and q_n^0 the reference period quantity. p_n^t indicates the commodity n price in month t and q_n^t is the corresponding monthly quantity. In addition, the *hybrid expenditure shares* s_n^{0b} is defined as follows: $s_n^{0b} \equiv p_n^0 q_n^b \Big/ \sum_{j=1}^{N} p_j^0 q_j^b$ where $\boldsymbol{p}^0 \equiv [p_1^0, ..., p_N^0]$ is the month 0 reference price vector and $\boldsymbol{q}^b \equiv [q_1^b, ..., q_N^b]$ is the reference year b quantity vector.

Taking January 2000 as the reference period and using direct comparison without performing quality adjustment, we calculated year-over-year indices from January 2000 to January 2014.[*6]

We used the Lowe (annual weight) and Laspeyres formulae to aggregate item level indices into national indices. As well, in order to conduct comparison at the aggregation stage, we calculated a two-stage Fisher index, too.

Figure 1 to Figure 4 summarize the indices calculated using each sampling method. The results shown are the upper-level aggregate results using the Laspeyres formula.

First, comparing Figure 2 and Figure 4 shows that the indices vary significantly when the top five products are sampled compared to when only the top product is sampled. As well,

[*6] Diewert (2014)[8] used a similar technique to perform a comparison while taking the effect of seasonal goods into account. Since this paper also includes seasonal goods, we used a year-over-year index for the comparison.

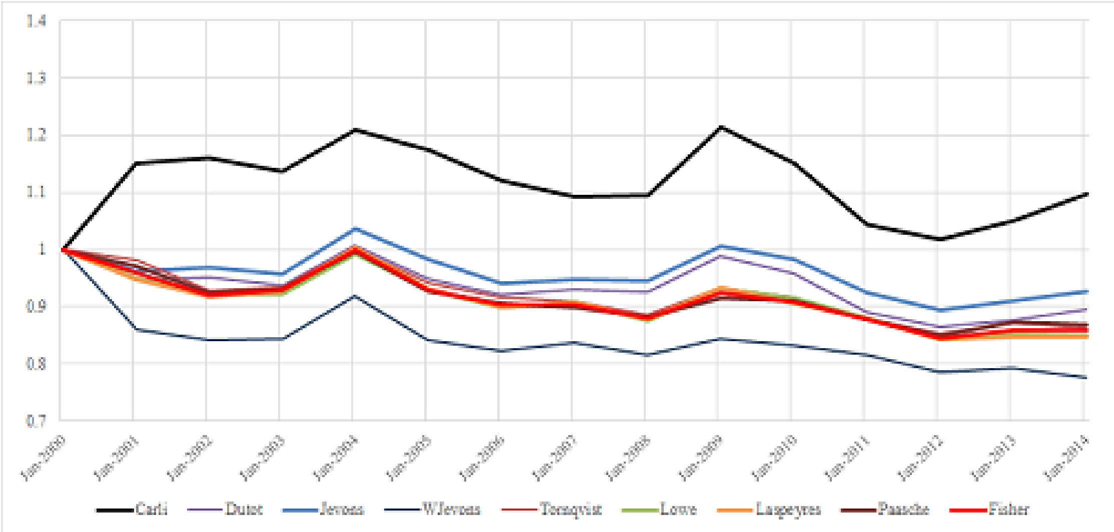


Figure 1 Laspeyres Aggregation Year-over-Year Index—Top 5 Items, 10 Outlets

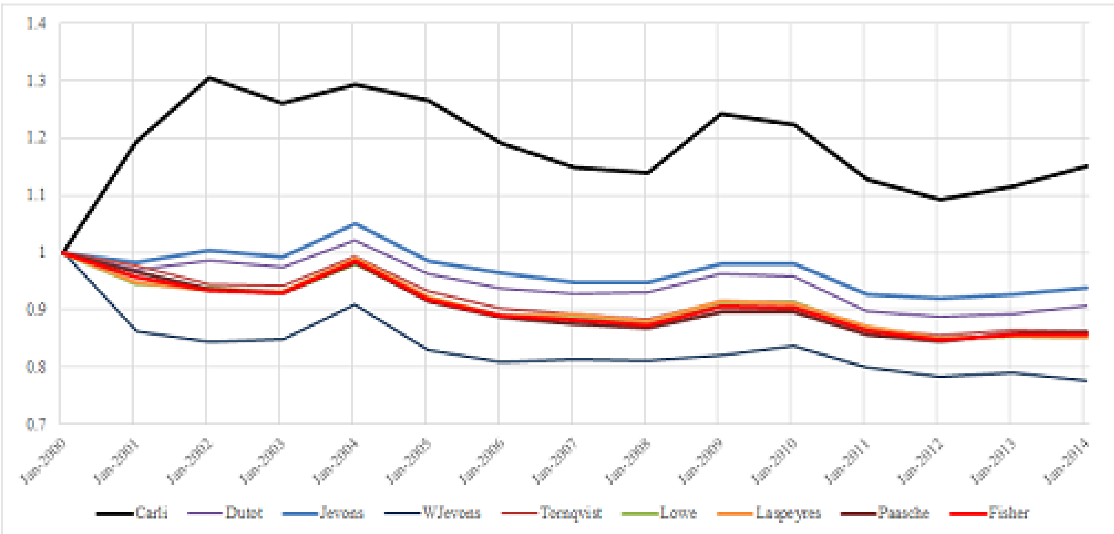


Figure 2 Laspeyres Aggregation Year-over-Year Index—Top 5 Items, 20 Outlets

comparing Figure 1 with Figure 2 and Figure 3 with Figure 4 shows that when only the top product is sampled, the differences between indices shrink due to the number of outlets increasing, and conversely, when the top five products are sampled, the differences expand.

Overall, the Carli index overestimates prices compared to other indices, while the weighted Jevons index underestimates them.

Diewert (2015)[9] has shown that the difference between the Lowe and Laspeyres price indices is equal to the covariance between the relative price and relative quantity at the reference period and observation period divided by the Laspeyres quantity index.*7 The

*7 "the difference between the Lowe and Laspeyres price indexes relating the prices of period 0 to those of period t is equal to the covariance between the relative price and relative quantity vectors, $\mathrm{Cov}(r, t, s^0)$,

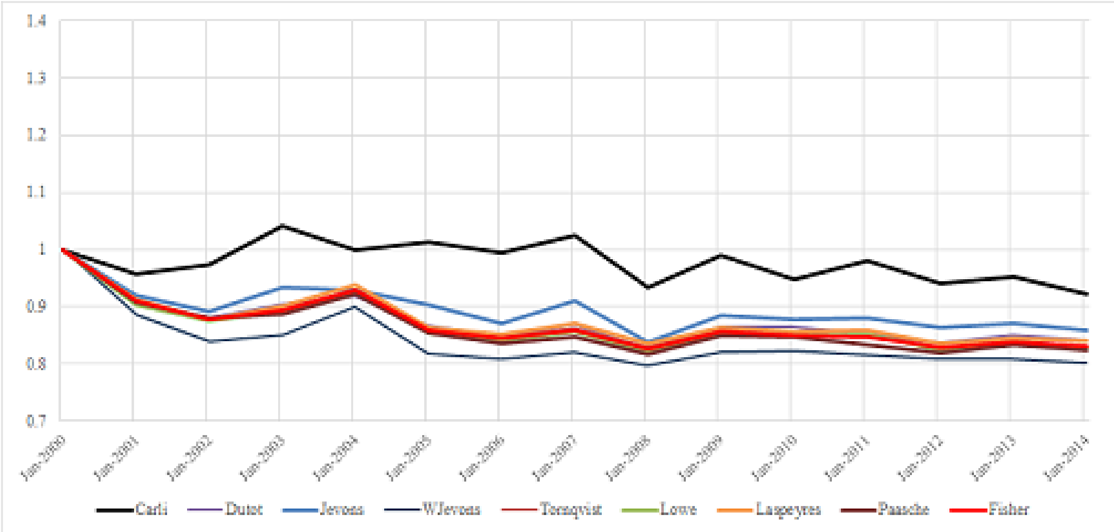


Figure 3 Laspeyres Aggregation Year-over-Year Index—Top 1 Item, 50 Outlets

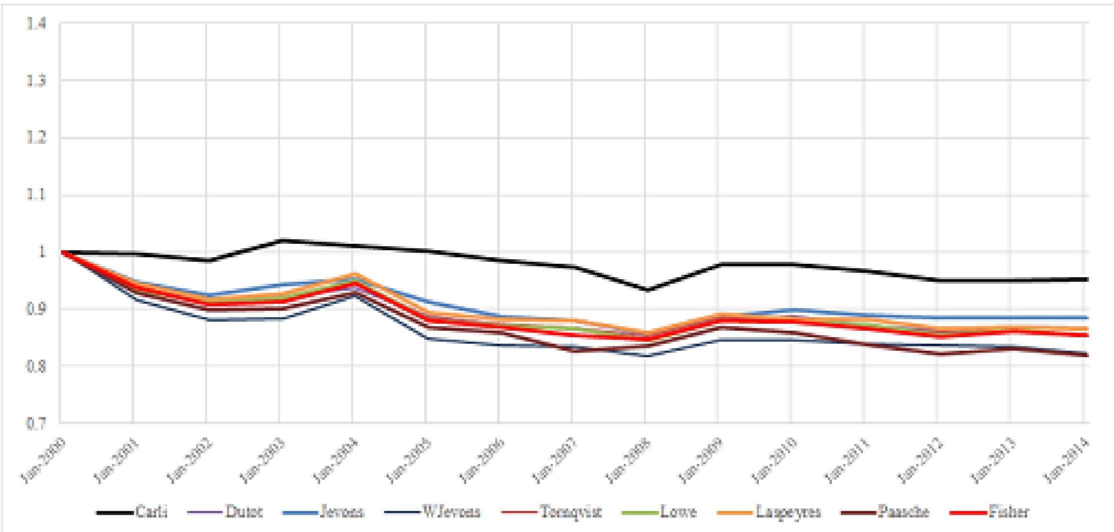


Figure 4 Laspeyres Aggregation Year-over-Year Index—Top 1 Item, 100 Outlets

difference between the Lowe and Laspeyres indices is shown in Table 6 and Figure 5.

The difference between Sampling A and B cannot reject the idea that it is 0, and as Diewert has pointed out, this suggests that the distribution of the relative prices is independent of the distribution of the relative quantities.[*8]

Meanwhile, with regard to Sampling C and D, the difference has a negative value, which

divided by the Laspeyres quantity index, $Q_L(\boldsymbol{q}^0, \boldsymbol{q}^b, \boldsymbol{p}^0)$." (Diewert, W. Erwin (2015)[9], "Index Number Theory and Measurement Economics" Lecture note #580, p7-9)

[*8] Above covariance will be zero ig "the distribution of the relative prices r_m is independent of the distribution of the relative quantities t_m" (Diewert, W. Erwin (2015)[9], "Index Number Theory and Measurement Economics" Lecture note #580, p7-9)

Table 6 The Difference between Lowe and Laspeyres Price Indices

		Sampling A Top 5 items 10 outlets	Sampling B Top 5 items 20 outlets	Sampling C Top 1 item 50 outlets	Sampling D Top 1 item 100 outlets
Difference	mean	−0.0006	−0.0017	−0.0098	−0.0100
	s.d.	0.0044	0.0034	0.0038	0.0034
t-test null hypothesis: true mean is equal to 0					
	t	−0.3522	−2.1161	−9.9360	−11.3897
	p-value	0.7304	0.0542	1.938e-07	3.898e-08
95 percent confidence interval:		−0.0028	−0.0037	−0.0119	−0.0121
		0.0020	0.0000	−0.0076	−0.0082

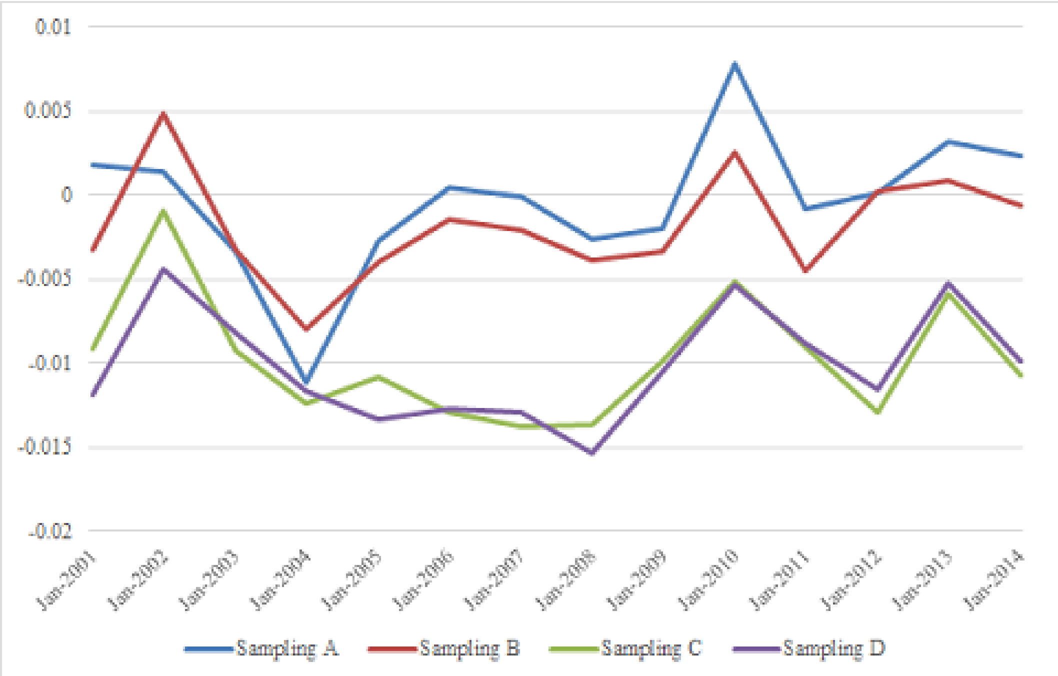


Figure 5 The Difference between Lowe and Laspeyres Price Indices

rejects the idea that it is 0. This is an unusual result, given that the value would normally be expected to be positive. It may be due to the fact there was deflation in Japan during the sample period covered by this paper, so price trends differed from the normal assumptions regarding long-term price trends.

3 Measurement of Price Index Biases

3.1 Biases Based on Choice of Elementary Index

We evaluated elementary index biases based on the procedure explained below. Here, elementary index biases are defined as the differences between calculation results using the Fisher price index and the other elementary indices. Since the elementary indices are calculated using the item level indices before performing upper-level aggregation, the biases were calculated for the 13 items for month t. We calculated the average and standard deviation for these biases by item and created 14 series from January 2001 to January 2014. Furthermore, we calculated the average and standard deviation for the 14 time series. We performed the above calculations for each of the four sampling method groups and then evaluated them.

In this manner, we tested what kind of differences occur in each elementary index's divergence from the Fisher index as a result of changing the sampling method.

3.2 Item Level Sampling Method Bias

Next, in order to evaluate biases in the sampling method, we first performed a comparison at the item level. Here, the sampling method bias is defined as the difference obtained by subtracting the Sampling D (top product at 100 outlets) Fisher index from the Fisher index for each of the other sampling methods (A, B, and C). We calculated the average and standard deviation of the difference for each item for month t. In the same manner as for elementary index biases, we also created 14 series from January 2001 to January 2014 and calculated the average and standard deviation for them. In this manner, we tested the structure of sampling method-based item level biases by evaluating differences between sampling methods for Fisher index item levels.

3.3 Aggregation-Level Sampling Method Bias

Next, we compared sampling method biases at the aggregation level. Here, we compared aggregate indices calculated using the two-stage Fisher method. In the same manner as for item levels, we compared differences between indices calculated with each sampling method, taking Sampling D (top product at 100 outlets) as the reference. We created 14 series from January 2001 to January 2014 and calculated the averages and standard deviations. In this manner, we evaluated aggregation-level sampling method biases and tested whether these biases could be canceled out at the aggregate level.

3.4 Evaluation of Aggregation-Level Sampling Method Bias

Finally, in order to evaluate whether the magnitude of sampling method biases at the aggregation level is at a level that can be ignored, we considered it by means of comparison with upper-level substitution biases. Upper-level substitution bias may be defined as the difference between the upper-level Laspeyres index calculated using Fisher as the elementary index and the index calculated using the two-stage Fisher method. In the same manner, we also calculated the difference between the aggregate Lowe index calculated using Fisher as the elementary index and the index calculated using the two-stage Fisher method. In this way, we tested whether aggregation-level sampling method biases were sufficiently less than upper-level substitution biases.

4 Empirical Analysis Results

4.1 Sampling Method-Based Differences in Biases Due to Choice of Elementary Index

We will first look at elementary index biases at the item level in order. Table 7 and Figure 6 show the results for Sampling A, Table 8 and Figure 7 the results for Sampling B, Table 9 and Figure 8 the results for Sampling C, and Table 10 and Figure 9 the results for Sampling D. As we saw in Figures 1 to 4, the Carli index results significantly exceed the other results, with an average of 0.18 points, while the weighted Jevons results fall below the other results, with an average of -0.07 points. On the other hand, if we exclude the weighted Jevons, the elementary indices that use weights (Törnqvist, Lowe, Laspeyres, and Paasche) all fall within a range of roughly ± 0.02 points. However, regardless of the sampling method, the bias is never fixed over time and fluctuates independently of temporal changes. Furthermore, the fact that the standard deviation grows larger with the passage of time reveals that the characteristics of items fluctuate over time.

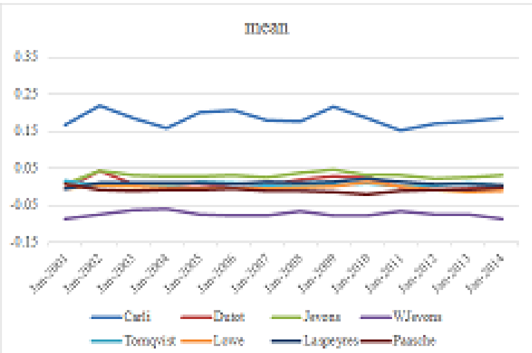
Focusing on the differences between sampling methods, in cases where only the top product is sampled, if the number of observations is increased by expanding the number of outlets, both the mean difference from the Fisher index and the standard deviation become smaller. On the other hand, with the method of sampling the top five products, if the number of observations is increased from 50 to 100 by expanding the number of outlets from 10 to 20, both the mean difference and the standard deviation become larger. It is also necessary to pay particular attention to changes in distribution. In the present comparison, there are 50 observations per month for Sampling A and Sampling C, compared to 100 for Sampling B and Sampling D. In other words, for the latter, the number of observations involved in the calculations is doubled, which may be expected to have a volatility-suppressing effect. However, when comparing Sampling A and B, both of which use the method of sampling the top five products, the volatility is not decreased even when the observations are increased.

Looking at the averages over time, the levels differed: with the Laspeyres index, for example, the mean difference was approximately 0.01 points with Sampling A and Sampling B, compared to 0.022 points for Sampling C and Sampling D.

As this shows, the sampling method produces biases independent of the choice of elementary index. In particular, the difference between selecting the top product only and selecting the top five products seems to have a greater effect than changing the sample size. In order to consider this point in more detail, we next conducted tests to compare differences by sampling method.

Table 7 Elementary Index Bias by Formula over Time (Mean, S.D.) − Top 5 Items, 10 Outlets

	Carli	Dutot	Jevons	WJevons	Törnqvist	Lowe	Laspeyres	Paasche
2001/1	0.167	−0.006	0.007	−0.087	0.017	−0.005	−0.006	0.006
	(0.297)	(0.023)	(0.020)	(0.147)	(0.049)	(0.025)	(0.018)	(0.018)
2002/1	0.219	0.044	0.044	−0.076	0.005	0.004	0.009	−0.008
	(0.306)	(0.050)	(0.043)	(0.103)	(0.011)	(0.032)	(0.025)	(0.024)
2003/1	0.185	0.008	0.030	−0.064	0.010	0.001	0.010	−0.010
	(0.263)	(0.042)	(0.028)	(0.104)	(0.022)	(0.021)	(0.024)	(0.024)
2004/1	0.158	0.001	0.027	−0.059	0.005	−0.001	0.010	−0.009
	(0.199)	(0.033)	(0.034)	(0.085)	(0.011)	(0.021)	(0.016)	(0.016)
2005/1	0.200	−0.002	0.030	−0.074	0.012	−0.004	0.010	−0.009
	(0.348)	(0.042)	(0.057)	(0.121)	(0.029)	(0.024)	(0.018)	(0.018)
2006/1	0.207	0.006	0.032	−0.079	0.011	−0.005	0.007	−0.006
	(0.349)	(0.058)	(0.071)	(0.118)	(0.029)	(0.026)	(0.024)	(0.023)
2007/1	0.180	0.003	0.025	−0.077	0.004	−0.004	0.013	−0.012
	(0.226)	(0.065)	(0.057)	(0.102)	(0.013)	(0.036)	(0.026)	(0.024)
2008/1	0.177	0.018	0.037	−0.066	0.003	−0.003	0.011	−0.010
	(0.248)	(0.065)	(0.086)	(0.085)	(0.008)	(0.018)	(0.024)	(0.024)
2009/1	0.217	0.028	0.046	−0.077	0.005	0.001	0.015	−0.014
	(0.318)	(0.084)	(0.076)	(0.110)	(0.016)	(0.028)	(0.023)	(0.022)
2010/1	0.187	0.024	0.032	−0.077	0.009	0.014	0.021	−0.020
	(0.213)	(0.073)	(0.073)	(0.086)	(0.015)	(0.045)	(0.032)	(0.030)
2011/1	0.151	−0.003	0.031	−0.066	0.001	0.002	0.013	−0.012
	(0.168)	(0.042)	(0.072)	(0.081)	(0.014)	(0.028)	(0.022)	(0.021)
2012/1	0.170	−0.009	0.022	−0.076	0.005	−0.007	0.008	−0.007
	(0.229)	(0.054)	(0.076)	(0.090)	(0.014)	(0.036)	(0.031)	(0.031)
2013/1	0.176	−0.005	0.025	−0.076	0.009	−0.014	0.008	−0.007
	(0.253)	(0.060)	(0.087)	(0.083)	(0.018)	(0.051)	(0.033)	(0.032)
2014/1	0.186	0.000	0.030	−0.086	0.003	−0.012	0.004	−0.003
	(0.315)	(0.073)	(0.089)	(0.122)	(0.014)	(0.034)	(0.021)	(0.020)
mean diff.	0.184	0.008	0.030	−0.074	0.007	−0.002	0.009	−0.009
s.d.	(0.021)	(0.015)	(0.009)	(0.008)	(0.004)	(0.007)	(0.006)	(0.006)



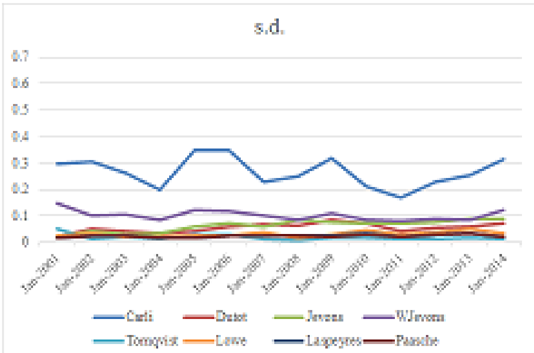


Figure 6 Elementary Index Bias by Formula over Time (Mean, S.D.) − Top 5 Items, 10 Outlets

Table 8 Elementary Index Bias by Formula over Time (Mean, S.D.) − Top 5 Items, 20 Outlets

	Carli	Dutot	Jevons	WJevons	Törnqvist	Lowe	Laspeyres	Paasche
2001/1	0.192	0.006	0.016	−0.088	0.016	−0.005	−0.006	0.006
	(0.342)	(0.023)	(0.022)	(0.141)	(0.044)	(0.017)	(0.017)	(0.017)
2002/1	0.287	0.040	0.042	−0.084	0.007	0.008	0.008	−0.008
	(0.597)	(0.067)	(0.069)	(0.127)	(0.020)	(0.024)	(0.022)	(0.021)
2003/1	0.268	0.024	0.042	−0.070	0.012	0.003	0.009	−0.008
	(0.549)	(0.074)	(0.064)	(0.106)	(0.030)	(0.022)	(0.024)	(0.024)
2004/1	0.224	0.013	0.034	−0.065	0.008	−0.001	0.006	−0.006
	(0.449)	(0.054)	(0.072)	(0.090)	(0.019)	(0.026)	(0.024)	(0.024)
2005/1	0.272	0.020	0.036	−0.080	0.014	0.001	0.008	−0.008
	(0.598)	(0.077)	(0.077)	(0.129)	(0.033)	(0.022)	(0.017)	(0.017)
2006/1	0.250	0.018	0.045	−0.077	0.014	0.004	0.009	−0.008
	(0.489)	(0.086)	(0.100)	(0.123)	(0.030)	(0.028)	(0.024)	(0.023)
2007/1	0.222	0.011	0.029	−0.077	0.008	0.002	0.012	−0.011
	(0.377)	(0.084)	(0.070)	(0.103)	(0.018)	(0.030)	(0.023)	(0.022)
2008/1	0.217	0.029	0.045	−0.066	0.006	0.004	0.012	−0.011
	(0.387)	(0.077)	(0.087)	(0.088)	(0.016)	(0.013)	(0.019)	(0.018)
2009/1	0.257	0.025	0.040	−0.082	0.006	0.000	0.015	−0.014
	(0.490)	(0.089)	(0.070)	(0.120)	(0.016)	(0.031)	(0.019)	(0.018)
2010/1	0.247	0.020	0.027	−0.078	0.012	0.010	0.020	−0.019
	(0.451)	(0.061)	(0.062)	(0.098)	(0.024)	(0.036)	(0.028)	(0.026)
2011/1	0.219	0.009	0.033	−0.070	0.001	0.003	0.019	−0.018
	(0.367)	(0.057)	(0.071)	(0.087)	(0.009)	(0.024)	(0.029)	(0.027)
2012/1	0.214	0.004	0.029	−0.079	0.006	−0.001	0.012	−0.011
	(0.332)	(0.067)	(0.076)	(0.090)	(0.012)	(0.034)	(0.026)	(0.025)
2013/1	0.212	0.006	0.033	−0.075	0.011	−0.007	0.006	−0.006
	(0.352)	(0.059)	(0.077)	(0.082)	(0.021)	(0.034)	(0.028)	(0.027)
2014/1	0.224	0.009	0.035	−0.089	0.003	−0.007	0.006	−0.006
	(0.423)	(0.072)	(0.082)	(0.127)	(0.007)	(0.029)	(0.017)	(0.017)
mean diff.	0.236	0.017	0.035	−0.077	0.009	0.001	0.010	−0.009
s.d.	(0.027)	(0.010)	(0.008)	(0.007)	(0.004)	(0.005)	(0.006)	(0.006)

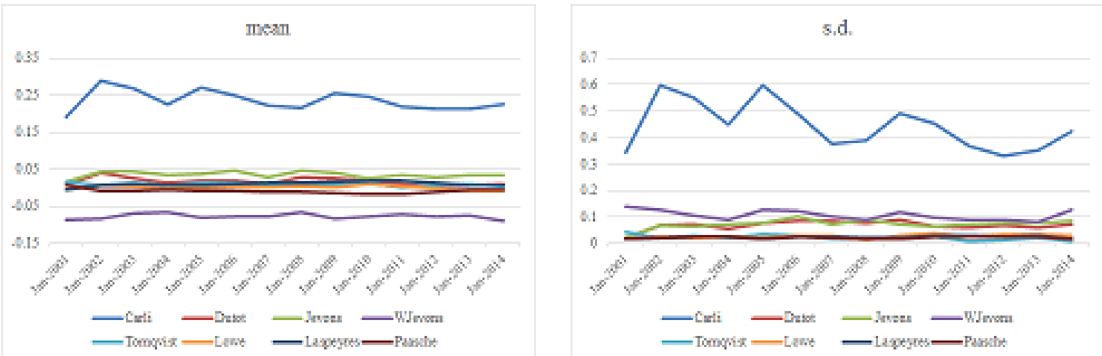


Figure 7 Elementary Index Bias by Formula over Time (Mean, S.D.) − Top 5 Items, 20 Outlets

Table 9 Elementary Index Bias by Formula over Time (Mean, S.D.) − Top 1 Item, 50 Outlets

	Carli	Dutot	Jevons	WJevons	Törnqvist	Lowe	Laspeyres	Paasche
2001/1	0.097	−0.003	0.008	−0.057	0.002	−0.009	0.009	−0.008
	(0.133)	(0.023)	(0.070)	(0.086)	(0.004)	(0.027)	(0.017)	(0.016)
2002/1	0.129	0.005	0.008	−0.070	−0.001	−0.014	0.000	0.001
	(0.225)	(0.054)	(0.063)	(0.122)	(0.014)	(0.071)	(0.037)	(0.040)
2003/1	0.181	0.004	0.033	−0.059	0.002	−0.001	0.029	−0.026
	(0.367)	(0.020)	(0.103)	(0.100)	(0.015)	(0.034)	(0.053)	(0.044)
2004/1	0.111	−0.006	0.009	−0.045	0.002	−0.002	0.015	−0.014
	(0.175)	(0.032)	(0.053)	(0.076)	(0.009)	(0.022)	(0.033)	(0.032)
2005/1	0.210	0.018	0.050	−0.056	0.010	−0.002	0.025	−0.022
	(0.433)	(0.063)	(0.162)	(0.090)	(0.025)	(0.022)	(0.051)	(0.043)
2006/1	0.222	0.023	0.044	−0.056	0.009	0.005	0.036	−0.030
	(0.428)	(0.061)	(0.116)	(0.085)	(0.022)	(0.025)	(0.083)	(0.063)
2007/1	0.231	0.015	0.048	−0.073	0.004	−0.006	0.029	−0.026
	(0.469)	(0.054)	(0.177)	(0.121)	(0.014)	(0.029)	(0.058)	(0.050)
2008/1	0.160	−0.011	0.000	−0.071	0.005	0.002	0.020	−0.018
	(0.307)	(0.047)	(0.101)	(0.106)	(0.015)	(0.023)	(0.053)	(0.047)
2009/1	0.167	−0.010	0.009	−0.069	0.004	−0.010	0.024	−0.023
	(0.322)	(0.066)	(0.119)	(0.124)	(0.021)	(0.042)	(0.031)	(0.029)
2010/1	0.137	0.001	0.006	−0.072	0.003	−0.005	0.026	−0.023
	(0.211)	(0.038)	(0.078)	(0.137)	(0.018)	(0.071)	(0.045)	(0.040)
2011/1	0.193	0.005	0.021	−0.068	0.003	0.001	0.037	−0.032
	(0.377)	(0.041)	(0.113)	(0.103)	(0.010)	(0.036)	(0.076)	(0.065)
2012/1	0.155	−0.005	0.008	−0.069	0.001	−0.008	0.023	−0.021
	(0.277)	(0.058)	(0.120)	(0.144)	(0.023)	(0.031)	(0.051)	(0.045)
2013/1	0.160	−0.003	0.021	−0.066	0.005	−0.013	0.016	−0.015
	(0.272)	(0.062)	(0.082)	(0.117)	(0.012)	(0.033)	(0.043)	(0.040)
2014/1	0.154	−0.017	0.001	−0.080	0.000	−0.020	0.015	−0.015
	(0.233)	(0.076)	(0.098)	(0.130)	(0.020)	(0.072)	(0.029)	(0.028)
mean diff.	0.165	0.001	0.019	−0.065	0.003	−0.006	0.022	−0.019
s.d.	(0.040)	(0.012)	(0.018)	(0.009)	(0.003)	(0.007)	(0.010)	(0.009)

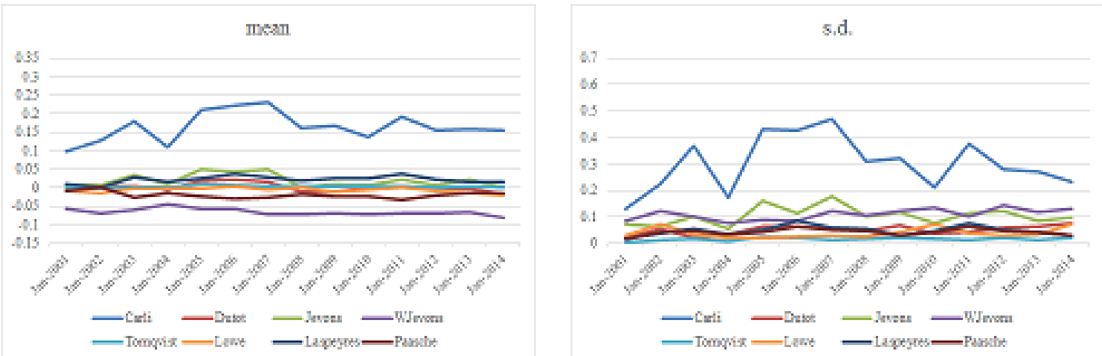


Figure 8 Elementary Index Bias by Formula over Time (Mean, S.D.) − Top 1 Item, 50 Outlets

Table 10 Elementary Index Bias by Formula over Time (Mean, S.D.) − Top 1 Item, 100 Outlets

	Carli	Dutot	Jevons	WJevons	Törnqvist	Lowe	Laspeyres	Paasche
2001/1	0.104	−0.012	−0.003	−0.076	0.005	−0.023	0.002	−0.001
	(0.130)	(0.067)	(0.121)	(0.151)	(0.009)	(0.074)	(0.028)	(0.029)
2002/1	0.121	−0.007	0.000	−0.069	0.003	−0.013	0.007	−0.006
	(0.177)	(0.059)	(0.076)	(0.140)	(0.017)	(0.072)	(0.031)	(0.033)
2003/1	0.163	−0.004	0.020	−0.071	0.003	−0.008	0.027	−0.026
	(0.283)	(0.038)	(0.080)	(0.136)	(0.017)	(0.059)	(0.032)	(0.029)
2004/1	0.124	−0.011	0.006	−0.058	0.007	−0.007	0.020	−0.019
	(0.171)	(0.026)	(0.050)	(0.108)	(0.011)	(0.039)	(0.018)	(0.017)
2005/1	0.188	0.011	0.036	−0.063	0.012	−0.005	0.028	−0.026
	(0.342)	(0.048)	(0.119)	(0.115)	(0.026)	(0.025)	(0.047)	(0.041)
2006/1	0.206	0.020	0.040	−0.061	0.011	0.004	0.036	−0.032
	(0.354)	(0.050)	(0.096)	(0.098)	(0.025)	(0.022)	(0.074)	(0.062)
2007/1	0.194	0.004	0.027	−0.081	0.008	−0.010	0.025	−0.024
	(0.326)	(0.070)	(0.129)	(0.160)	(0.015)	(0.058)	(0.032)	(0.030)
2008/1	0.154	−0.005	0.004	−0.077	0.006	−0.004	0.019	−0.018
	(0.255)	(0.054)	(0.096)	(0.134)	(0.015)	(0.028)	(0.038)	(0.035)
2009/1	0.166	−0.005	0.011	−0.078	0.005	−0.011	0.023	−0.022
	(0.284)	(0.068)	(0.106)	(0.155)	(0.016)	(0.050)	(0.025)	(0.024)
2010/1	0.158	0.007	0.012	−0.082	0.006	−0.012	0.021	−0.019
	(0.244)	(0.059)	(0.096)	(0.158)	(0.018)	(0.082)	(0.035)	(0.032)
2011/1	0.178	0.001	0.015	−0.077	0.003	−0.002	0.034	−0.030
	(0.312)	(0.058)	(0.099)	(0.130)	(0.009)	(0.029)	(0.068)	(0.059)
2012/1	0.160	−0.002	0.015	−0.076	0.001	−0.008	0.025	−0.023
	(0.262)	(0.074)	(0.106)	(0.172)	(0.024)	(0.043)	(0.046)	(0.042)
2013/1	0.142	−0.011	0.011	−0.077	0.003	−0.015	0.012	−0.011
	(0.229)	(0.098)	(0.104)	(0.147)	(0.016)	(0.045)	(0.027)	(0.026)
2014/1	0.165	−0.016	0.008	−0.094	−0.003	−0.022	0.013	−0.012
	(0.269)	(0.099)	(0.105)	(0.160)	(0.017)	(0.072)	(0.024)	(0.024)
mean diff.	0.159	−0.002	0.014	−0.074	0.005	−0.010	0.021	−0.019
s.d.	(0.029)	(0.010)	(0.013)	(0.009)	(0.004)	(0.007)	(0.010)	(0.009)

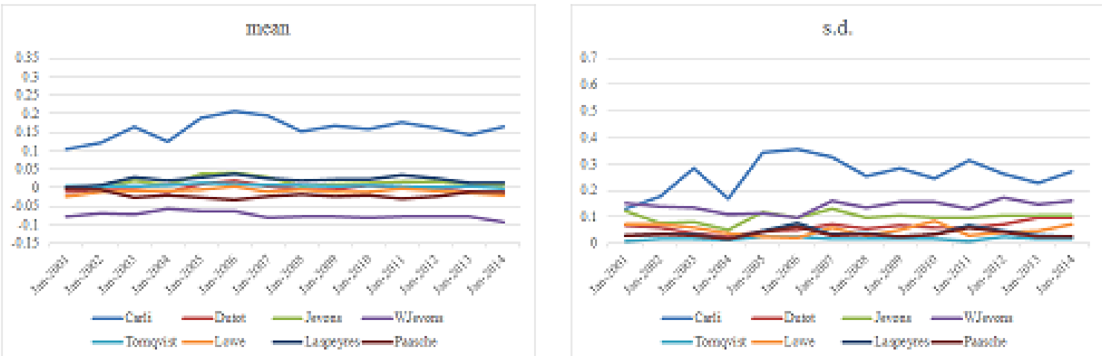


Figure 9 Elementary Index Bias by Formula over Time (Mean, S.D.) − Top 1 Item, 100 Outlets

4.2 Sampling Method-Based Item-Level Bias

As seen in the previous section, the impact of expanding the product sampling range (from one product to five products) appears to be greater than the difference produced by changing the sample size (changing the number of outlets) for a fixed product sampling range. Next, in order to evaluate sampling method biases, we first performed a comparison at the item level. Table 11 and Figure 10 show the biases over time and mean differences.

To begin, having specified Sampling D (top product only, 100 outlets) as the reference, we looked at the divergence that occurs as a result of changes in the sampling method. In comparison to Sampling D, let us first consider Sampling C (top product only, 50 outlets), for which only the sample size is different. With an average of −0.023, Sampling C's item level sampling method bias is not small, but it is relatively stable over time. As well, differences between products are also relatively small and stable over time. What about the fact that Sampling C has a downward bias? When sampling only the top product, increasing the

Table 11 Sampling Method Bias over Time (Mean, S.D.)

	Sampling A Top 5 items 10 outlets	Sampling B Top 5 items 20 outlets	Sampling C Top 1 item 50 outlets
Jan-2001	−0.007	0.003	−0.034
	(0.170)	(0.158)	(0.102)
Jan-2002	−0.005	0.007	−0.011
	(0.137)	(0.132)	(0.057)
Jan-2003	−0.006	0.006	−0.026
	(0.109)	(0.109)	(0.100)
Jan-2004	−0.006	0.000	−0.028
	(0.110)	(0.112)	(0.109)
Jan-2005	0.007	0.011	−0.026
	(0.108)	(0.117)	(0.100)
Jan-2006	−0.004	−0.008	−0.024
	(0.105)	(0.108)	(0.095)
Jan-2007	−0.050	−0.050	−0.025
	(0.222)	(0.233)	(0.113)
Jan-2008	−0.029	−0.024	−0.015
	(0.169)	(0.184)	(0.088)
Jan-2009	−0.044	−0.040	−0.032
	(0.226)	(0.235)	(0.097)
Jan-2010	−0.013	−0.011	−0.026
	(0.173)	(0.188)	(0.077)
Jan-2011	−0.054	−0.045	−0.013
	(0.162)	(0.176)	(0.072)
Jan-2012	−0.049	−0.037	−0.029
	(0.211)	(0.214)	(0.078)
Jan-2013	−0.043	−0.048	−0.020
	(0.258)	(0.268)	(0.056)
Jan-2014	−0.072	−0.065	−0.020
	(0.277)	(0.279)	(0.083)
mean diff.	−0.026	−0.021	−0.024
s.d.	(0.025)	(0.026)	(0.007)

Figure 10 Sampling Method Bias over Time (Mean, S.D.)

number of samples means broadening the scope to include more outlets with relatively low popularity. As a result, in Sampling C, which uses only the 50 top outlets, there is a relatively higher proportion of large-scale stores that adopt a low-margin, high-volume price strategy, which has a downward effect compared to the method which samples 100 outlets.

On the other hand, the initial bias with Sampling A (top five products, 10 outlets) and Sampling B (top five products, 20 outlets), which sample five products, is small compared to Sampling C (top product only, 50 outlets), which samples only one product, but the bias is not stable over time. The standard deviation and difference between products also seem to increase. Sampling B (top five products, 20 outlets) has the same sample size (i.e., 100) as the reference, Sampling D (top product only, 100 outlets), so direct comparison is possible. Sampling A, meanwhile, which has half the sample size of the reference, has a similar structure to Sampling B, and the difference between them is slight. Therefore, the gap between Sampling A and Sampling B, which sample a variety of products, is less than the gap between Sampling C and Sampling B, which sample the top product only. With Sampling A and B, the scope of products sampled within the same store is increased. Thus, with the method of sampling the top five products, we may consider that the variation in products has a significant effect on the results. This demonstrates that the selection of the number of representative products is important. In addition, it reveals that if there is a wide variety of products, the results cannot simply be stabilized by increasing the sample size. This finding is notable in contrast to the fact that, as seen earlier, elementary index biases were stable over time for most indices.

Let us also compare sampling method biases and elementary index biases. Sampling C has an average sampling method bias of -0.023, whereas, as seen earlier, the average elementary index bias for Sampling C is -0.009 with the Lowe index and 0.021 with the Laspeyres index, so the magnitude of the sampling method bias is by no means small.

4.3 Sampling Method-Based Aggregation-Level Bias

Next, we compared the aggregation-level indices calculated based on the two-stage Fisher method and looked at sampling method biases at the aggregation level. Table 12 and Figure 11 show the time series and mean difference results.

The average is 0.025 for Sampling A (top five products, 10 outlets), 0.018 for Sampling B (top 5 products, 20 outlets), and -0.020 for Sampling C, so even at the aggregate level, sampling method biases remain which are not insignificant. There is a 0.05 difference over time in the biases for the sampling methods that involve sampling the top five products, so the difference between these methods is still present at the aggregate level, and it does not become stable over time. This suggests that sampling method bias is not canceled out between products and

Table 12 Sampling Method Bias over Time (Aggregation Level)

	Sampling A Top 5 items 10 outlets	Sampling B Top 5 items 20 outlets	Sampling C Top 1 item 50 outlets
Jan-2001	0.022	0.019	−0.027
Jan-2002	0.010	0.023	−0.030
Jan-2003	0.014	0.014	−0.020
Jan-2004	0.049	0.036	−0.015
Jan-2005	0.049	0.036	−0.018
Jan-2006	0.031	0.017	−0.022
Jan-2007	0.048	0.029	0.006
Jan-2008	0.031	0.027	−0.020
Jan-2009	0.040	0.023	−0.024
Jan-2010	0.029	0.025	−0.027
Jan-2011	0.012	0.000	−0.017
Jan-2012	−0.002	0.000	−0.023
Jan-2013	0.002	−0.001	−0.022
Jan-2014	0.007	0.008	−0.024
mean diff.	0.025	0.018	−0.020
s.d.	(0.018)	(0.013)	(0.009)

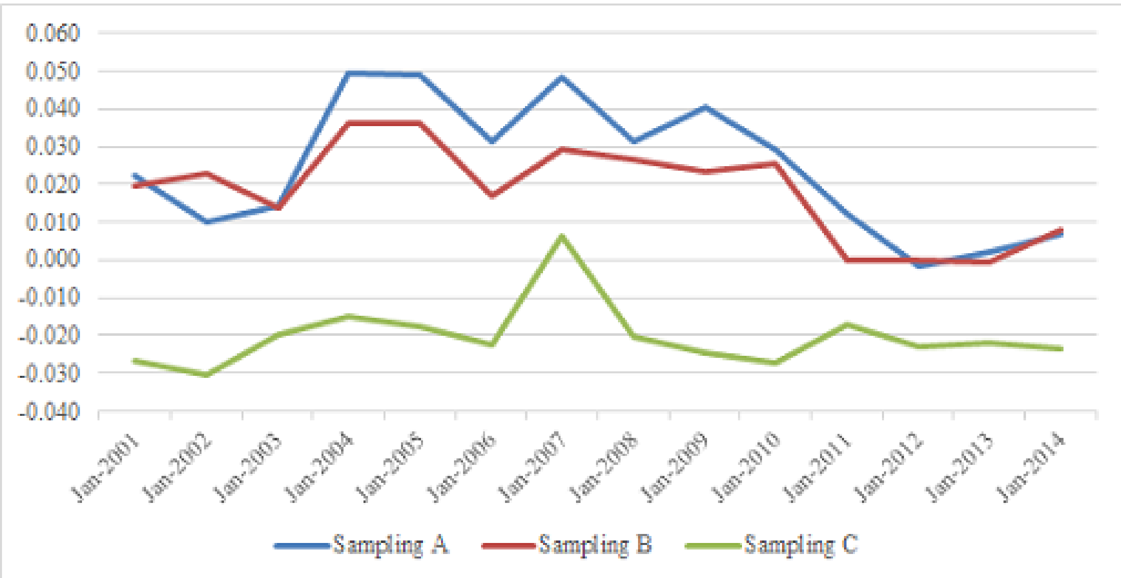


Figure 11 Sampling Method Bias over Time (Aggregation Level)

remains even in upper-level indices, but if we consider this finding in conjunction with the sampling method bias results by item seen above, the cause does not seem to be fixed. That is, for methods that involve sampling the top product only, the sampling method bias by item is a stable downward bias, and the distribution between products is also stable. What's more, this stable, downward sampling bias remains even with upper-level aggregation. As discussed earlier, with the methods that involve sampling the top product only, increasing the sample size from 50 to 100 has the effect of broadening the scope to include more low-popularity outlets. Therefore, if the sample size is small, prices at high-popularity outlets which may adopt a low-margin, high-volume pricing strategy have a greater weight, as a result of which a downward bias is generated. On the other hand, with the methods that involve sampling the top five products, the bias is not stable at either the item level or the aggregate level. This may be because the variation in the sampled products affects the results. In addition, with the methods that involve sampling the top five products, the number of outlets is at most 20, which means they are limited to high-popularity outlets. These findings therefore imply that product variation has a significant effect on prices even at large-scale stores.

4.4 Evaluation of Sampling Method-Based Aggregation-Level Bias

Finally, Table 13, Figure 12, and Figure 13 show upper-level substitution biases. First, starting from the reference period (0), upper-level substitution bias with the Laspeyres index tends to grow larger with the passage of time, which is consistent with past findings. For the Lowe index, on the other hand, upper-level substitution bias varies considerably between sampling methods. Regardless of the sampling method or elementary index, however, upper-level substitution bias mostly falls within the range of −0.01 to 0.01, and the magnitude is large in relation to the average differences seen in the previous section (0.025 for Sampling A, 0.018 for Sampling B, and −0.020 for Sampling C), so even at the aggregation level, sampling method-based biases are at a level that cannot be ignored.

Table 13 Upper-Level Substitution Bias over Time (Elementary: Fisher)

	Sampling A		Sampling B		Sampling C	
	Laspeyres aggregation	Lowe aggregation	Laspeyres aggregation	Lowe aggregation	Laspeyres aggregation	Lowe aggregation
Jan-2001	0.000	0.000	0.000	0.003	−0.001	−0.009
Jan-2002	0.003	0.008	0.003	0.010	0.000	−0.002
Jan-2003	0.000	0.003	0.002	0.004	0.000	0.000
Jan-2004	0.005	0.008	0.004	−0.005	0.000	0.002
Jan-2005	0.000	−0.001	0.001	−0.001	−0.004	−0.007
Jan-2006	0.001	0.003	0.001	0.009	−0.004	−0.005
Jan-2007	0.001	−0.001	0.000	0.010	−0.001	−0.001
Jan-2008	0.001	−0.002	−0.001	0.009	−0.002	−0.003
Jan-2009	0.004	0.004	0.002	0.005	0.001	0.003
Jan-2010	0.010	0.002	0.008	0.017	0.009	−0.004
Jan-2011	0.008	0.000	0.006	0.012	0.006	0.003
Jan-2012	0.005	−0.002	0.005	0.011	0.009	0.003
Jan-2013	0.010	0.005	0.008	0.008	0.013	0.008
Jan-2014	0.009	0.003	0.008	0.009	0.014	0.008
mean diff.	0.004	0.002	0.003	0.007	0.003	0.000
s.d.	(0.004)	(0.003)	(0.003)	(0.006)	(0.006)	(0.005)

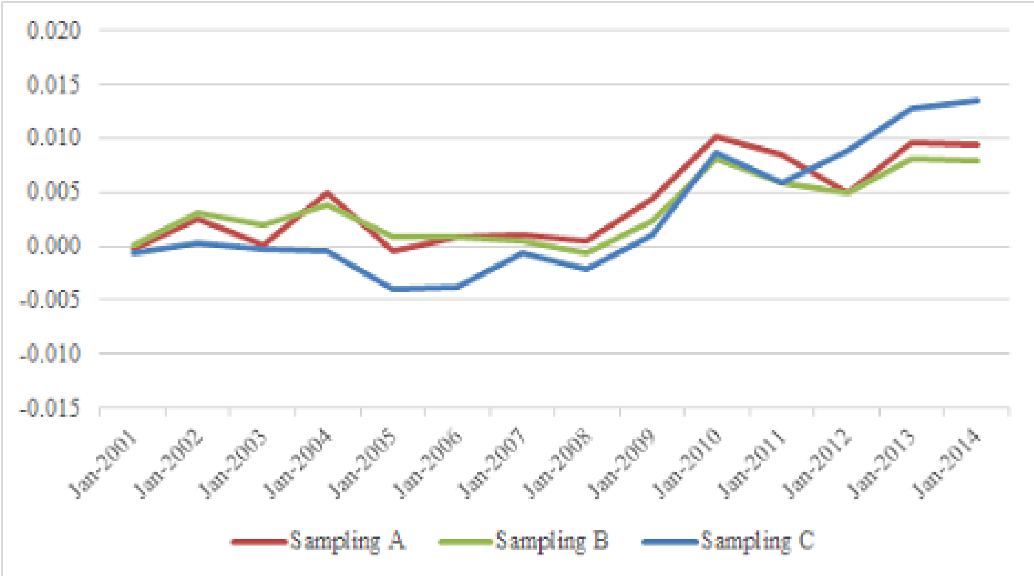


Figure 12 Upper-Level Substitution Bias over Time (Elementary: Fisher, Aggregation: Laspeyres)

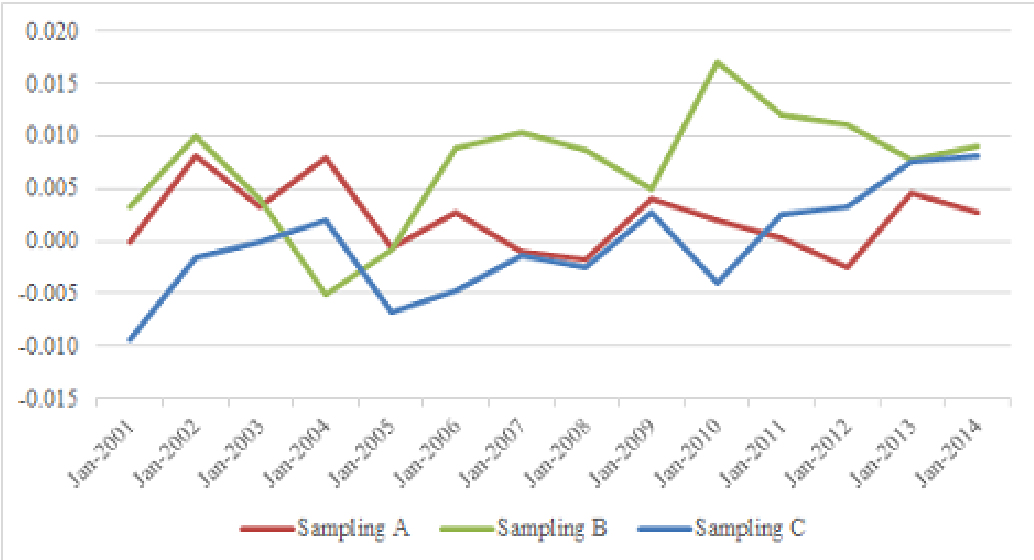


Figure 13 Upper-Level Substitution Bias over Time (Elementary: Fisher, Aggregation: Lowe)

5 Conclusion

There has been much policy-related discussion surrounding the definition of the price index and its accuracy and biases. There are various restrictions when it comes to measuring public price indices, such as budget constraints, but despite these, constant efforts are being made to improve them by changing survey methods, expanding them, or devising creative solutions.

In this paper, we focused on the elementary indices and sampling methods used to measure price indices and experimentally observed what kind of differences occur when these are changed. The following became clear as a result.

In the past, in addition to discussions about upper-level aggregation biases, there has been much discussion regarding elementary indices, with lower-level biases being a point of contention. Our findings, however, make it clear that sampling methodologies for elementary indices that correspond to more lower-level prices may have effects that cannot be ignored at both the item level and at higher levels of aggregation.

With regard to sampling in particular, there are many practical problems. Even if an index is calculated by specifying a uniform sample size based on the budget constraints, we have seen that substantial biases occurred depending on the price sampling method that is used. In addition, it was clear that expanding the sample size produces substantial fluctuations due to differences in the underlying sampling method:

- With the method of sampling only the top product, which is currently used to calculate public price indices in many countries, increasing the sample size seems to have the effect of raising the index's level. In light of this, it would be natural to assume that increasing the number of products that are sampled would make it possible to obtain price levels which are closer to reality, but when we ran tests by sampling the top five products, no structural improvement was observed.
- Similarly, one might also think there is a possibility of over-estimation if one samples the top five products from a few upper-level outlets, but we saw that with methods that involve sampling the top five products, biases do not stabilize over time. This may be because increasing product variation incorporates higher added-value products within the same category that may be expected to have lower sales volumes.[*9]
- In the case of sampling the top product only, there is a possibility of underestimation if there are few outlets. For sampling starting with upper-level outlets, there is a risk of a downward bias due to low-margin, high-volume pricing strategies. That is, if the sample size is small with this method, a downward bias may be generated due to prices at high-popularity outlets which adopt low-margin, high-volume pricing strategies having a greater weight.[*10]

Based on the above discussion, one can see that there is a certain logic to the method of calculating indices by selecting one representative product from survey target outlets, which is the sampling method used to date by various national statistics offices. In Japan, on the other hand, an extremely broad sample size is set, ranging from a minimum of one sample to

[*9] When using scanner data, product variation may easily be increased, but in that case, we may expect that this kind of bias that does not stabilize with the top five-product sampling method will be incorporated into the results. In other words, this suggests that in order to reflect a broader range of product alternatives in indices using scanner data, it is necessary to determine suitable weights.

[*10] There are certain budget constraints on the price surveys conducted by national statistics offices, so there is a limit on the number of outlets from which data can be collected. As a result, it is to be expected that there may be a certain downward bias in effect in published CPIs.

a maximum of 42 samples based on the survey target city and item. Based on this, the effect that differences in sample size by item have on item-level indices and aggregation-level indices should be carefully examined. In other words, we propose that in order to discuss lower-level biases, it is necessary to clearly separate discussion based on the choice of elementary index from discussion relating to the sampling method. Sampling methods that sample the top five products, which we tested in this paper, are another issue that should be carefully examined when using data sources such as scanner data in future. Large-scale sources such as scanner data are recognized for their value in terms of the breadth of their coverage, but when considered from the standpoint of representativeness, price information for products with a low level of representativeness will consistently be included.

This does not mean that scanner data cannot be used for price indices. Rather, it suggests that it is essential to fully understand the attributes of different types of data when using them and also that it is important to perform further quality adjustment and select suitable weights.

References

[1] Abe, Naohito, Toshiki Enda, Noriko Inakura and Akiyuki Tonogi (2015), "Effects of New Goods and Product Turnover on Price Indexes", The Research Center for Economic and Social Risks, DP15-2.

[2] Ariga, Kenn and Kenji Matsui (2003), "Mismeasurement of the CPI", in Blomstrom, Magnus, Jennifer Corbett, Fumio Hayashi, and Anil Kashyap. Structural Impediments to Growth in Japan. Chicago, IL: University of Chicago Press.

[3] Broda, Christian and David E. Weinstein (2007), "Defining Price Stability in Japan: A View from America", Monetary and Economic Studies, Institute for Monetary and Economic Studies, Bank of Japan, vol. 25(S1), pages 29-56, December.

[4] Dalén, Jörgen (1998), "Studies on the Comparability of Consumer Price Indices", International Statistical Review, Vol. 66, No. 1 (Apr., 1998), pp. 83-113

[5] de Haan, Jan, C.M. Schut and E. Opperdoes (1999), "Item selection in the Consumer Price Index: Cut-off versus probability sampling", Survey Methodology, Vol. 32, No. 2, 197-216.

[6] Diewert, W. Erwin (1976), "Exact and Superlative Index Numbers", Journal of Econometrics 4(2), pp.114-145.

[7] Diewert, W Erwin, (1978), "Superlative Index Numbers and Consistency in Aggregation," Econometrica, Econometric Society, vol. 46(4), pages 883-900, July.

[8] Diewert, W. Erwin (2014), "An Empirical Illustration of Index Construction using Israeli Data on Vegetables", Working Paper 14-04, at Vancouver School of Economics.

[9] Diewert, W. Erwin (2015), "Index Number Theory and Measurement Economics", Lecture notes for Economics 580, at University of British Columbia.

[10] Gábor, Enikö and Philip Vermeulen (2014). "New Evidence on Elementary Index Bias", ECB Working Paper No. 1754

[11] Handbury, Jessie, Tsutomu Watanabe and David E. Weinstein (2013). Center for Advanced Research in Finance, Faculty of Economics, CARF-F-328.

[12] Imai, Satoshi, Chihiro Shimizu and Tsutomu Watanabe (2012), "How fast are prices in Japan falling?", CARF Working paper no. CARF-F-298.

[13] Imai, Satoshi and Tsutomu Watanabe (2013), "Product Downsizing and Hidden Price Increases: Evidence from Japan's Deflationary Period", Center for Advanced Research in Finance, Faculty of Economics, CARF-F-320.

[14] Sudo, Nao, Kozo Ueda, Kota Watanabe and Tsutomu Watanabe (2014). "Working Less and Bargain Hunting More: Macro Implications of Sales during Japan's Lost Decades", Graduate School of Economics in its series U Tokyo Price Project Working Paper Series with number 029, at University of Tokyo.