How Fast Are Prices in Japan Falling?

Satoshi Imai[*] Chihiro Shimizu[†] Tsutomu Watanabe[‡]

First draft: May 17, 2012

Revised version: June 7, 2012

Abstract

The consumer price inflation rate in Japan has been below zero since the mid-1990s. However, despite the presence of a substantial output gap, the rate of deflation has been much smaller than that observed in the United States during the Great Depression. Given this, doubts have been raised regarding the accuracy of Japan's official inflation estimates. Against this background, the purpose of this paper is to investigate to what extent estimates of the inflation rate depend on the methodology adopted. Our specific focus is on how inflation estimates depend on the method of outlet, product, and price sampling employed. For the analysis, we use daily scanner data on prices and quantities for all products sold at about 200 supermarkets over the last ten years. We regard this dataset as the "universe" and send out (virtual) price collectors to conduct sampling following more than sixty different sampling rules. We find that the officially released outcome can be reproduced when employing a sampling rule similar to the one adopted by the Statistics Bureau. However, we obtain numbers quite different from the official ones when we employ different rules. The largest rate of deflation we find using a particular rule is about 1 percent per year, which is twice as large as the official number, suggesting the presence of substantial upward-bias in the official inflation rate. Nonetheless, our results show that the rate of deflation over the last decade is still small relative to that in the United States during the Great Depression, indicating that Japan's deflation is moderate.

Keywords: consumer price index; scanner data; deflation; outlet sampling; product sampling; purposive sampling; random sampling; sampling bias

1 Introduction

The consumer price index (CPI) inflation rate in Japan has been below zero since the mid-1990s, clearly indicating the emergence of deflation over the last 15 years. However, the rate of deflation measured by headline CPI in each year was around 1 percent, which is much smaller than the rates observed in the United States during the Great Depression. Some suggest that this simply reflects the

[*]Statistics Bureau of Japan. s2.imai@soumu.go.jp
[†]Reitaku University and University of British Columbia. cshimizu@reitaku-u.ac.jp
[‡]University of Tokyo and Canon Institute for Global Studies. watanabe@e.u-tokyo.ac.jp

fact that although Japan's deflation is persistent, it is only moderate. Others, both inside and outside the country, however, argue that something must be wrong with the deflation figures, questitioning Japan's price data from a variety of angles. One of these is that, from an economic perspective, the rate of deflation, given the huge and persistent output gap in Japan, should be higher than the numbers released by the government suggest. Fuhrer et al. (2011), for example, estimating a NAIRU model for Japan, conclude that it would not have been surprising if the rate of deflation had reached 3 percent per year. Another argument focuses on the statistics directly. Broda and Weinstein (2007) and Ariga and Matsui (2003), for example, maintain that there remains non-trivial mismeasurement in the Japanese consumer price index, so that the officially released CPI inflation rate over the last 15 years contains substantial upward bias.

Against this background, the purpose of the present paper is to investigate to what extent estimates of the CPI inflation rate depend on the methodology adopted. Our specific focus in this paper is on how the inflation rate depends on the sampling method; that is, how estimates of inflation depend on store sampling (i.e., from which stores prices are collected), product sampling (i.e., for which products prices are collected), and price sampling (i.e., which day of the month prices are collected; whether they are regular or sales prices; etc.). To conduct such an investigation, we employ daily scanner data on prices and quantities for *all* products sold at about 200 stores (chained or independent supermarkets) over the last ten years (January 2000 to April 2010). We regard this dataset as the "universe." We then send (virtual) price collectors to this universe to conduct sampling following more than sixty alternative sampling rules. Specifically, with regard to store sampling, we instruct the price collectors to choose stores based on the quantities sold at the store or, alternatively, on the number of customer visits to the store. With regard to product sampling, we conduct purposive and random sampling. In purposive sampling, we first define the set of candidate products that meet the prespecified product type specifications, and then choose products out of the set using the quantities sold as a criterion. In random sampling, products are randomly chosen among all products belonging to an item category (i.e., without specifying a set of candidate products). Finally, with regard to price sampling, we instruct the price collectors not to collect prices of products that are on sale, with a sale alternatively defined as a temporary price reduction that lasts less than eight days or less than three days.

Our main findings are as follows. First of all, we successfully reproduce the outcome released by the Statistics Bureau when we employ a sampling rule quite similar to the one adopted by the Statistics

Bureau itself. Using the officially released CPI, the overall change in the price level over the last 10 years is -5.6 percent (for an annualized rate of -0.5 percent), while our estimate turns out to be -6 percent. This result indicates that our universe consisting of products sold at 200 stores is not substantially different from the actual universe to which the Statistics Bureau sends price collectors. However, when we employ rules different from the baseline rule, we obtain quite different figures from the official ones. Specifically, the largest rate of deflation we find using a particular sampling rule is about 10 percent over the decade, which is twice as large the figure based on the official CPI data. This finding lends support to the argument that the official inflation rate may be upward-biased. At the same time, though, even in the case of our most extreme result, which is equivalent to an annualized rate of 1 percent over the decade, deflation was still relatively moderate compared to the rates of up to 7 percent observed during the Great Depression in the United States.

The rest of the paper is organized as follows. In Section 2, we conduct purposive sampling. Specifically, we use the product type specifications employed by the Statistics Bureau and define the set of candidate products that meet the product type specifications. We then choose some products out of the set following 64 alternative sampling rules, each of which differs in terms of the method of outlet, product, and price sampling. In Section 3, we proceed to random sampling, in which we choose products among all products belonging to a particular item category with the choice probability for each product determined by the quantities sold over the last one month. We propose a new methodology to quantitatively evaluate the size of the sampling bias, which we define as the difference between the inflation rates calculated based on purposive sampling and based on random sampling. Section 4 concludes the paper.

2 Price Indexes Based on Purposive Sampling

Consumer price indexes in different countries are constructed following a set of common rules, which are described in various documents such as ILO et al. (2004). Nevertheless, there still remain several important methodological differences, one of which is differences in product sampling, with some countries employing purposive and others random sampling. In purposive sampling, the statistic agency of a country defines product type specifications for each of the item categories. Products are sampled only from a set of candidate products with these specifications. On the other hand, in random sampling, products are randomly chosen among *all* products belonging to an item category (i.e., without specifying a set of candidate products).

From a statistical perspective, purposive sampling has some undesirable features, including sampling bias (i.e., the prices of sampled products may not come from the true price distribution) and lower sampling efficiency (i.e., the variance of prices of sampled products may be larger than the corresponding variance in the case of random sampling).[1] However, purposive sampling has advantages from a practical perspective in that the process of narrowing the range of candidate products makes sampled products more homogeneous, thereby making estimated price indexes less volatile even in the case of high product substitution. As a result, many countries, including Japan, have adopted purposive sampling, while only a limited number of countries, including, however, the United States, have adopted random sampling.

In this section, we employ purposive sampling to collect prices from the universe of scanner data, while in the next section we employ random sampling. The purposive sampling conducted in this section is based on the list of product types, with product type specifications, used by Japan's Statistics Bureau (JSB), which we refer to as the JSB product type specifications.[2] Using this list, we conduct sampling to see how much the resulting price indexes differ depending on the specifics of the sampling method eomployed.

2.1 Methodology

2.1.1 Sampling

Outlet sampling Our dataset covers about 200 outlets in Japan. While some of them are located in large cities like Tokyo or Osaka, others are located in smaller cities. We sample outlets in two different ways. The first is based on the assumption that all of the 200 outlets are located in a single large commercial area, while the second is based on the assumption that there are six different commercial areas and that each of the 200 outlets is located in one of the six areas. In the first case, we pick 42 outlets for each item using several alternative criteria, including the number of customer visits to the outlet and the quantity sold at the outlet. In the second case, we choose 8 outlets based on similar criteria for each item in each of the six areas, so we choose 48 outlets in total for each item. Specifically, we employ four different criteria for outlet sampling: (1) the number of customer visits to the outlet over the last one month; (2) the number of customer visits to the outlet over the last

[1] See the Appendix for more on these issues.
[2] The complete list of product type specifications is available in Statistics Bureau (2@@@).

three months; (3) the quantity sold at the outlet over the last one month for products belonging to an item category; (4) the quantity sold at the outlet over the last three months for products belonging to an item category.[3] Note that we pick the same set of outlets for all items when we use the first and second criteria, while we may pick different outlets depending on the item concerned when we use the third and fourth criteria.

Product sampling Once an outlet is picked, we then choose a product out of the set of products that meet the JSB specifications, based on the quantity sold at that outlet over the last one month or the last three months. Let us explain how we specify the set of candidate products taking butter as an example. According to the JSB list of product types, the product type specifications for butter are as follows:

JSB Product Type Specifications for Butter

Jul 1996 - Jan 2001	"Snow Brand Hokkaido Butter"
Jan 2001 - present	200g. Packed in a paper container. Excluding unsalted butter.

Note that only a single product, "Snow Brand Hokkaido Butter," was on the list from July 1996 to January 2001, while multiple products were allowed in the more recent period. Based on this information, we produce the list of product barcodes, which are called JAN (Japan Article Number) codes, that meet the JSB product type specifications. Our sample period is January 2000 to April 2010. Our task is very simple for the period from January 2000 to January 2001: we just look for the unique JAN code corresponding to "Snow Brand Hokkaido Butter." On the other hand, for the period from February 2001 to April 2010, we look for the JAN codes of products that meet the specifications described above. Specifically, we do so using supplementary information on each JAN code, including the name of a product, brand, model number, net quantity, and ingredients. This process is done by using a text matching technique ("regular expression"). We find that the number of products (i.e., the number of JAN codes) that meet the above specifications is 31. Among these 31 products, we choose a single product based on the quantities sold over the last one month or the last three months at a particular outlet chosen through the process of outlet sampling.

Note that in the example given here, "unsalted butter" can be regarded as a negative characteristic in the sense that products with that characteristic ("unsalted butter") are *excluded* from the product

[3] Note that when we count the quantity sold at the outlet, we only count products that meet the JSB product type specifications.

specification, while "200g" and "packed in a paper container" can be seen as positive characteristics. For each product type, we calculate the number of products that meet the product specifications based on positive characteristics only and the number of products based on the full range of both positive and negative characteristics. For butter, the number of products based on the full range of characteristics (including the negative characteristic "excluding unsalted butter") is 31, as mentioned above, while the number of products based on positive characteristics only (i.e., "unsalted butter" is not excluded) is 123. In our simulation exercises, we compare the outcomes obtained when we use the full range of characteristics and when we use only positive characteristics to see how much the estimated rate of deflation differs depending on how tightly product type specifications are defined.

Table 1 presents the number of products that meet the JSB product type specifications. For example, the total number of products (i.e., the number of JAN codes) for item code 1321 ("Butter") is 369, and the number of products that meet the JSB product type specifications is 31. The share of products that meet the JSB specifications is very small (8.1 percent), although the sales share of those products is not that small (45.8 percent). The number of products belonging to all of the item categories covered by our scanner data is 462,906, among which 70,966 products meet the JSB specifications (15.3 percent).[4]

Price sampling Price collectors are instructed by the Statistics Bureau not to collect sale prices. Specifically, price collectors are instructed to exclude "extra-low prices due to bargain, clearance, or discount sales, and quoted for less than eight days" (Statistics Bureau (2@@@)). To mimic this practice, we treat temporary price reductions as follows. First, we define a temporary price reduction as a price reduction where the price goes back to its original level. Next, we then identify such temporary price reductions for each product at each outlet. If the duration of a temporary price reduction is equal to or more than κ days, we do not apply any special treatment; however, if it is less than κ days, we do not use that price and instead look for the "regular" price. Specifically, we assume that the regular price is equal to either the price level just before the temporary price reduction (i.e., we use forward imputation) or the price level just after the temporary price reduction ends (backward imputation). In our simulation exercises, we replace a temporary price with a duration of less than κ days with

[4]The JSB list of product type specifications is updated every five years, although for some items minor modifications are made more frequently. In our simulation exercises, we employ an overlap method to eliminate price changes resulting from product substitution when the JSB list is updated.

the regular price calculated in this manner. We set $\kappa = 8$ and $\kappa = 3$. The former corresponds to the current rule employed by the Statistics Bureau, while the latter case implies that our (virtual) price collectors are allowed to collect prices that are lowered only for, say, four days. Note that since for $\kappa = 3$ our price collectors tend to collect more sale prices than in the case of $\kappa = 8$, we would expect the estimated rates of deflation to be potentially greater.

As for the timing of price collection, we follow the current practice adopted by the Statistics Bureau. That is, our price collectors are instructed to collect prices on Wednesday, Thursday, and Friday of the week which includes the 12th of the month. The order of priority regarding the three days is Thursday, Wednesday, and Friday. If no transaction is recorded during these three days for a particular product in a particular month, we search for a record of transactions retroactively from that date to the 1st of that month.

2.1.2 Aggregation at lower and upper levels

Aggregation at the lower level For aggregation at the lower level, we employ the unweighted arithmetic mean of price levels across product-outlet combinations (i.e., the Dutot index). That is, the purposive-sampling (PS) price index for item i in region r in month t, $P_{r,i}^{PS}(t)$, is defined by

$$P_{r,i}^{PS}(t) \equiv n^{-1} \sum_{(o,j)\in A_{r,i}} P_{r,i,o,j}(t) \tag{2.1}$$

where $P_{r,i,o,j}(t)$ represents the price in month t of product j, which belongs to item i, quoted at outlet o located in region r, n is the number of products collected for an item in a region, and $A_{r,i}$ is the set of product-outlet combinations obtained through the process of outlet and product sampling explained earlier.

Note that a similar procedure is adopted by the Statistics Bureau for aggregation at the lower level. However, it is often pointed out that the adoption of the Dutot index for lower level aggregation is a source of measurement bias in the Japanese CPI (see, for example, Broda and Weinstein (2007)). We stick to the arithmetic mean of price levels in this section, but will compare this with the geometric mean of price relatives (i.e., the Jevons index) in the next section, where we conduct random sampling similar to that adopted by the Bureau of Labor Statistics in the United States.

Aggregation at the upper level Next, we construct a fixed-base Laspeyres index by aggregating the lower level indexes. The price index in region r is defined as

$$I_r^{PS}(t) \equiv \sum_i \omega_{r,i} \frac{P_{r,i}^{PS}(t)}{P_{r,i}^{PS}(t_0)} \tag{2.2}$$

where $\omega_{r,i}$ is the consumption weight for item i in region r in the base year ($t = t_0$), satisfying $\sum_i \omega_{r,i} = 1$. The weight $\omega_{r,i}$ is taken from the Family Income and Expenditure Survey conducted by the Japanese government. Finally, we construct a price index for the entire country by aggregating the regional indexes:

$$I^{PS}(t) \equiv \sum_r \phi_r I_r^{PS}(t). \tag{2.3}$$

where ϕ_r represents the consumption weight for region r with $\sum_r \phi_r = 1$.

Previous studies on upper level aggregation argue that the adoption of the fixed-base Laspeyres index by the Statistics Bureau is another important source of upward bias in the Japanese CPI. In particular, empirical studies on this issue, including Shiratsuka (@@@@), find a substantial difference between the fixed-base Laspeyres index and other indexes, including the chained Tornqvist index, and argue that the fixed-base Laspeyres index has some undesirable features. In this paper, however, we adopt the fixed-base Laspeyres index unless otherwise mentioned. Our focus in this paper is on the sampling issue, which has not been discussed much in previous studies, rather than upper level aggregation. Therefore, in order to ensure that the alternative indexes obtained by different sampling methods are directly comparable with the officially released results, we stick with the fixed-base Laspeyres index.

2.1.3 List of simulation exercises

In sum, we conduct various types of simulation, which differ in the following respects:

- Two alternative definitions of commercial areas: the 200 outlets covered by our scanner data are located in a single region or in six different regions.
- Four alternative criteria for outlet selection: (1) the number of customer visits to the outlet over the last one month; (2) the number of customer visits to the outlet over the last three months; (3) the quantity sold at the outlet over the last month; (4) the quantity sold at the outlet over the last three months.

- Two alternative criteria for product selection: the quantity sold of the product over the last one month; and the quantity sold of the product over the last three months.
- Two alternative approaches to product type specifications: based on a full range of product characteristics and based on positive characteristics only.
- Two alternative definitions of a sale: the duration of a temporary price reduction is less than 8 days or less than 3 days.
- Two alternative definitions of the regular price: backward or forward imputation.

The total number of simulation exercises we conduct is 64, all of which are presented in Table 2.

2.2 Data

The dataset we use consists of store scanner data compiled jointly by Nikkei Digital Media Inc. and the Research Center for Price Dynamics. This dataset contains daily sales data for more than 200,000 products sold at about 200 supermarkets in Japan from 2000 to 2010. The products consist mainly of food, beverages, and other domestic nondurables (such as detergent, facial tissue, shampoo, soap, toothbrushes, etc.), which make up 125 items of the consumer price statistics compiled by the Statistics Bureau.[5] Their sales are recorded through the so-called point-of-sale system. Each product is identified by the JAN code, the equivalent of the UPS code in the United States.

Tables 3 and 4 show the number of outlets and products for each year, as well as the turnover (entry and exit) of outlets and products during the sample period. The number of outlets covered in 2009 is 260, and the total number of different products sold in 2009 is about 230,000. The total number of observations for 2009 is about 422 million (no. of articles $\times$ no. of outlets $\times$ no. of days), while the total for the entire sample period is approximately 3.6 billion observations.

The number of outlets that are included in the dataset throughout the entire sample period is 103. The number of products sold by those 103 outlets in 2000 was approximately 203,000 and has subsequently risen steadily, reaching roughly 256,000 in 2009. During this period, tens of thousands of products were newly launched each year, but about the same number of products were also withdrawn. The ratio of the number of newly launched products relative to existing products was about 30 percent,

[5]The total number of items in the consumer price statistics is 584. Our dataset covers about 20 percent of the entire items of the consumer price statistics in terms of consumption weight.

while the withdrawal rate was about 27 percent, indicating that the turnover in products was quite rapid.

2.3 Empirical Results

Table 5 shows the results of the simulation exercises for the mean and standard deviation of estimated monthly inflation rates for each of the 64 cases. As for the mean of monthly inflation rates, the highest value we obtain is -0.035 percent in simulation #12, which is equivalent to an annualized rate of deflation of 0.43 percent. On the other hand, the lowest value we obtain (in simulation #54) is -0.081 percent, which translates into an annualized deflation rate of 0.97 percent. The officially released inflation rate for the same set of items is -0.045 percent per month, or an annualized rate of deflation of 0.54 percent. Thus, the estimated rate of deflation based on simulation #12 is slightly smaller than the official figure, while the estimate from simulation #54 is almost twice as large as the official figure. In fact, a careful examination of Table 5 reveals, the mean of monthly inflation rates is lower (deflation is higher) than the official figure in 41 out of the 64 cases, suggesting that inflation rates obtained from our simulation exercises tend to be lower than the officially released inflation rate. Turning to the standard deviation, the smallest value is 0.616 percent (simulation #43). Interestingly, this is much greater than the corresponding figure for the officially released inflation rate, which is 0.228 percent, indicating that the inflation rates obtained in the simulation exercises are much more volatile than the official inflation rate.

Figure 1 shows the movement in the monthly price index based on simulations #12 and #54, with the upper panel depicting the log of the price index levels and the lower panel the year-on-year change in the price index. In the upper panel, we see that the estimated index for #12 moves quite similarly to the official index. That is, the two indexes were on the same downward trend from 2000 to mid-2007, then simultaneously started to rise at the end of 2007, and continued to rise until the end of 2008. The two indexes again embarked on a downward trend in the fall of 2008, when the output gap widened substantially due to the global financial crisis. On the other hand, the estimated index for #54 exhibits a more rapid decline in 2000-2007 than the other two indexes. Turning to the lower panel of Figure 1, this shows that the year-on-year inflation rate on the basis of #54 fell below -2 percent numerous times during the sample period, while for the official inflation rate this occurred only twice.

Next, we proceed to investigating which elements of the sampling method have greater influence on the estimated rate of inflation. Figure 2 shows how the mean of monthly inflation depends on the way outlet and product sampling is conducted. Panel (a) shows the result obtained when we use only positive product characteristics, while panel (b) shows the result obtained when we use the full range of product characteristics to define products. The results suggest the following. First, the rate of deflation tends to be greater when we use the quantity sold as the criterion of outlet selection than when we use customer visits as the criterion. Note that we tend to pick large outlets when we use the number of customer visits as the criterion, since the number of customer visits to the store is greater for large outlets than for medium-sized or small outlets. Our results indicate that prices decline less rapidly in these large outlets, which is inconsistent with claims repeatedly made by researchers and practitioners that prices have been declining more rapidly in large outlets. According to our results, prices decline more rapidly in medium-sized (or even small) outlets, which are not very large in terms of customer visits, but specialize in certain particular items, offering cheaper prices and selling more of these items (as a result of which they are chosen when the quantity sold of these items is used as the criterion).

Second, the rate of deflation tends to be greater when we assume a single commercial area than when we assume six heterogeneous commercial areas. Note that the assumption of a single commercial area results in outlets located in large cities such as Tokyo to be more likely to be picked. Therefore, our results indicate that prices tend to decline more rapidly at outlets located in such large cities, suggesting the presence of non-trivial heterogeneity across regions in terms of the rate of deflation, which requires regional stratification.

Third, the rate of deflation tends to be greater when we use only positive characteristics to define product types than when we use the full range of characteristics. This is shown by comparing Figures 2(a) and (b), which show the results when using the full range of characteristics (Figure 2(b)) and when using only positive characteristics (Figure 2(a)). Specifically, Figure 2(b) shows that when the full range of characteristics is employed, the mean of monthly inflation rates is almost the same as that based on the officially released figures (indicated by the horizontal broken line). In contrast, as can be seen in Figure 2(a), when only positive characteristics are used, the mean of monthly inflation rates is substantially below the value based on officially released data. It may not be particularly surprising that the results come closer to the official figure when we use the full range, which comprises exactly the same range of characteristics used by the Statistics Bureau. However, what is surprising is that simply adjusting the set of product characteristics yields such a substantial difference. In this sense, our results suggest that the estimated rate of deflation depends crucially on how product types are specified.[6]

Finally, the standard deviation of monthly inflation rates tends to be higher when outlet sampling is based on quantities sold than when it is based on customer visits. Our interpretation of this result is that outlet sampling based on quantities sold results in more frequent outlet substitution, and thus more frequent product substitution, which leads to higher volatility in the estimated inflation rate. A somewhat interesting finding is that even when we use customer visits as the criterion, the standard deviation of monthly inflation rates is still substantially higher than the standard deviation of the official inflation rates.[7]

Figure 3 investigates the effect of how sales are treated on monthly inflation rates, with Figure 3(a) showing the results when only positive characteristics are employed and Figure 3(b) showing those when the full range of characteristics is used. It is frequently pointed out that the practice of excluding sale prices with a short duration (i.e., sale prices that last less than 8 days) in the consumer price statistics has substantially reduced the CPI rate of deflation. However, Figure 3 shows that the estimated rate of deflation does not depend much on how sale prices are treated. Specifically, the rate of deflation is slightly greater when sale prices with shorter duration are included (i.e., $\kappa = 3$), especially when combined with the assumption of a single commercial area. However, the difference is not very large and, more importantly, there are some cases in which the rate of deflation is smaller when sale prices with a shorter duration are included. Moreover, the figure also shows that the rate of deflation does not depend on how regular prices are estimated (i.e., whether forward or backward imputation is used).

[6] Note that the standard deviation of monthly inflation rates is slightly lower when the full range is used, but the difference is not as substantial as in the case of the mean of monthly inflation rates.

[7] In fact, as can be seen in Table 5, we fail to produce even a single case in which the standard deviation of monthly inflation rates comes close to the corresponding official figure. This is in a sharp contrast with our results for the mean of monthly inflation rates, for which we are able to produce numbers close to those based on the official data. We are not quite sure why this is the case, but this finding seems to suggest that the sampling rules we consider in this paper may differ in some important respects from those employed by the Statistics Bureau.

3 Price Indexes Based on Random Sampling

3.1 Methodology

In the purposive sampling implemented in the previous section, we first determined the set of candidate products that meet the product type specifications and then chose a product out of the set following a particular sampling rule. Random sampling, which we consider in this section, differs from purposive sampling in that no set of candidate products is determined; instead, specific products are chosen randomly from among *all* products belonging to a particular item category.

The issue we focus on specifically here is the sampling bias introduced by purposive sampling, which we measure as the difference between the different between purposive sampling index and the random sampling index. To obtain an accurate estimate of the sampling bias, we need to construct the two price indexes in such a manner that they differ in terms of the sampling method employed but are identical in all other respects. To do so, we employ the same procedure for constructing the two indexes. Specifically, we use specification #3 listed in Table 2: that is, for each of the two indexes, the number of regions is assumed to be six; outlet sampling is conducted based on the number of customer visits over the last one month;[8] a sale is defined as a temporary price reduction that lasts less than eight days ($\kappa = 8$), and regular prices are estimated by forward imputation. As for the product sampling proceedure, we randomly choose products among all products belonging to an item category, with the choice probability for each product determined by the quantity sold over the last one month. Note that we do not resample products every month; instead, we conduct resampling only when product substitution is inevitable (i.e., when a product disappears or when outlet substitution occurs).

For aggregation at the lower level (i.e., aggregation of prices of products within an item category), we take the unweighted geometric mean of price relatives (i.e., the Jevons index) instead of the ratio of the unweighted arithmetic means of the price level in months t and $t-1$ (i.e., the Dutot index) described in equation (2.1). That is,

$$\frac{P_{r,i}^{RS}(t)}{P_{r,i}^{RS}(t-1)} \equiv \prod_{(o,j)\in B_{r,i}} \left[\frac{P_{r,i,o,j}(t)}{P_{r,i,o,j}(t-1)}\right]^{1/n} \tag{3.1}$$

where $B_{r,i}$ represents the set of products chosen by random sampling. Finally, we aggregate this over

[8] We assume that (forced) outlet substitution takes place four times a year (March, June, September, and December) and that each time one fourth of the outlets are replaced. Note that a similar (but less frequent) outlet rotation is conducted in the United States.

i, and then over r to obtain

$$I_r^{RS}(t) \equiv \sum_i \omega_{i,r} \prod_{T=t_0+1}^{T=t} \frac{P_{i,r}^{RS}(T)}{P_{i,r}^{RS}(T-1)} \tag{3.2}$$

and

$$I^{RS}(t) \equiv \sum_r \phi_r I_r^{RS}(t) \tag{3.3}$$

3.2 Empirical Results

3.2.1 Year-on-year inflation rates

Figure 4 shows the year-on-year rate of inflation based on the price index constructed using random sampling, which we will refer to as the "*RS* index." The index shown here is based on specification #3 in Table 2. We produced 78 replications of the time series for the *RS* index over the entire sample period and calculated the year-on-year inflation rates for each of the 78 replications. The blue line in the figure represents the mean of the year-on-year inflation rates of the 78 replications, while the shaded area shows the confidence interval defined by the mean ± 1 standard deviation, where the standard deviation is calculated based on the 78 replications. On the other hand, the green line in the figure represents the year-on-year inflation rate based on the index constructed using purposive sampling ("*PS* index"), which is again based on specification #3 in Table 2. The blue and green lines differ in terms of product sampling (i.e., random vs. purposive sampling) and lower level aggregation (i.e., Dutot index vs. Jevons index), but are identical in all other respects.

The *RS* and *PS* indexes exhibit similar trends over time, but differ in some respects. Specifically, the rate of deflation in 2000-2003 was below 0.5 percent for the *RS* index, while it was more than one percent for the *PS* index. During this period, inflation based on the *PS* index was closer to that based on the official CPI, which is represented by the red line in the figure. Another significant difference can be observed in 2008, when the inflation rate turned positive. The peak of the inflation rate in this year was 2 percent for the *RS* index, while it was above 3 percent for the *PS* index. Again, the inflation rate based on the *PS* index was closer to that based on the official CPI. The rate of deflation over the entire sample period was 0.622 percent per year for the *RS* index, 0.537 percent for the *PS* index, and 0.543 percent for the official CPI.

In the rest of this section, we compare the *RS* and *PS* indexes in more detail. Our main interest is in the sampling bias; more specifically, we are interested in the impact that the product sampling approach (i.e., random vs. purposive sampling) has on the inflation rate. To focus on this issue, we need to eliminate the effect of different methods of lower level aggregation. We do so by using the geometric rather than the arithmetic mean of prices levels when conducting lower level aggregation for the *PS* index. Figure 5 compares the item-level inflation rates constructed in this manner, shown on the vertical axis, with the item-level inflation rates for the original *PS* index, shown on the horizontal axis. We see that most of the dots are below the 45 degree line, indicating that the inflation rates based on the geometric mean of price levels tend to be lower by around 0.3 percent per year.

3.2.2 Sampling bias at the item level

Figure 6 compares the *RS* and *PS* indexes for margarine (item code 1602). The log of the price relatives for this item collected by random sampling in region r is $\pi^{RS}_{r,1602,o,j}(t)$, and the item level inflation rate is given by $n^{-1}\sum_{(o,j)\in B_{r,1602}} \pi^{RS}_{r,1602,o,j}(t)$, where $B_{r,1602}$ is the set of products belonging to item 1602, which in this case consists of 416 products, as shown in Table 1. Note that the number of margarine prices collected in each of the six regions is 16, so that the total number of prices collected is 96. We repeat random sampling to produce 78 replicates of the time series for the margarine index. On the other hand, the inflation rate based on purposive sampling is given by $n^{-1}\sum_{(o,j)\in A_{r,1602}} \pi^{PS}_{r,1602,o,j}(t)$, where $A_{r,1602}$ is the set of products that meet the JSB product type specifications, which in this case consists of 12 products, as shown in Table 1.

We define the measure of the difference between the two indexes for margarine, $\delta_{1602}(t)$, as

$$\delta_{1602}(t) \equiv \sum_r \phi_r \left(n^{-1} \sum_{(o,j)\in A_{r,1602}} \pi^{PS}_{r,1602,o,j}(t) - n^{-1} \sum_{(o,j)\in B_{r,1602}} \pi^{RS}_{r,1602,o,j}(t) \right). \tag{3.4}$$

Note that we have 78 replications of $\delta_{1602}(t)$, each of which corresponds to the 78 replications for the time series of the *RS* index for margarine. We then calculate the mean of $\delta_{1602}(t)$ over the 78 replications, which is denoted by $\hat{\bar{\delta}}_{1602}(t)$.

The result is shown in Figure 6, where the blue line represents the probability density function (PDF) for $\delta_{1602}(t)$, while the green line represents the PDF for $\hat{\bar{\delta}}_{1602}(t)$. As can be seen, the two PDFs are almost identical, implying that the variance of $\delta_{1602}(t)$ over the 78 replications is small relative to the variance of $\delta_{1602}(t)$ over t, and thus each of the 78 replications of $\delta_{1602}(t)$ follows an almost identical distribution. The mean of $\hat{\bar{\delta}}_{1602}(t)$ is -0.0018 per month, indicating that the sampling bias is, on average, very close to zero. On the other hand, the standard deviation of the sampling bias is quite large at 0.038 per month, implying that it is not unlikely that the RS and PS inflation rates deviate substantially. Specifically, the probability that the sampling bias exceeds 30 percent on the basis of annualized inflation rates is more than @@ percent.

We conduct a similar calculation for each of the 125 items that make up the CPI, and the result is shown in Figure 7. Starting with the mean of the sampling bias, we find that this is close to zero for some items as in the case of margarine, but there are many other items for which the mean is substantially above or below zero, suggesting that margarine is not typical in this respect. On the other hand, the standard deviation of the sampling bias exceeds 0.02 per month for most items, indicating that the large standard deviation observed for margarine is not an exception.

Why is there such a large sampling bias at the item level? One potential reason is that products chosen by random sampling and those chosen by purposive sampling do not overlap very much. To see whether this is the case, we count how many products meet the JSB product type specifications among all the products picked by random sampling. For example, in the case of margarine, the total number of prices collected in constructing the RS index is 76,128, among which 21,362 prices are for products that meet the JSB product type specifications, corresponding to a share of 28.1 percent. We conduct the same calculation for each item, and the result is presented in Figure 8. Referring to the share of prices for products that meet the JSB product type specifications in the total number of prices collected in constructing the RS index as the degree of overlap, it can be readily seen that in Figure 8 the number of items for which this overlap exceeds 30 percent is relatively limited, and for many items this overlap is considerably lower. The average overlap for all items is only 11.7 percent, suggesting that the difference in products picked by random and purposive sampling may be one source of the large sampling bias at the item level.[9]

[9] One may wonder why the overlap is so low. One possible explanation is that the JSB product type specifications are very tight, so that only a limited number of products meet these JSB specifications. This seems to be consistent with what we saw in Table 1. To examine this possibility in more detail, we plot each of the 125 items in Figure 9 where the horizontal axis measures the degree of overlap, while the vertical axis measures the share of products that meet the JSB specifications in the total number of products, which is taken from Table 1. For margarine, the share of products that meet the JSB specifications is 2.9 percent, while the overlap in the case of purposive sampling and random sampling is 28.1 percent, so that the observation for margarine is located far above the 45 degree line. On the other hand, as the figure shows, there are also many items located below the 45 degree line. Interestingly, most of the items for which the share of products that meet the JSB specifications in the total number of products is above 10 percent are located far below the 45 degree line. These results suggest that it may not be appropriate to ascribe the observed low overlap to the tightness of JSB product type specifications.

3.2.3 Sampling bias at the aggregate level

Let us turn to the sampling bias at the aggregate level. We aggregate $\delta_i(t)$ and $\hat{\delta}_i(t)$ over i to define $\delta(t)$ and $\hat{\delta}(t)$; that is, $\delta(t) \equiv \sum_i \omega_i \delta_i(t)$ and $\hat{\delta}(t) \equiv \sum_i \omega_i \hat{\delta}_i(t)$, where ω_i is the consumption weight for item i. Figure 10 shows the PDFs of $\delta(t)$, represented by the blue line, and $\hat{\delta}(t)$, represented by the green line. Note that, as we saw for margarine in Figure 6, the two PDFs overlap, indicating that each of the 78 replications of $\delta(t)$ comes from an almost identical distribution.

Figure 10 shows that sampling bias $\hat{\delta}(t)$ has a mean of 0.00011 or 0.13 percent on an annualized basis, indicating that the sampling bias is, on average, quite close to zero. On the other hand, the standard deviation of the sampling bias is 0.0053 or 6.4 percent at an annualized basis. These results indicate that the distribution of the sampling bias at the aggregate level differs substantially from that at the item level. Specifically, at the item level, as seen in Figure 7, the mean of the sampling bias exhibits a substantial deviation from zero for a large number of items; however, at the aggregate level, the positive and negative deviations from zero at the item level cancel each other out, resulting in an almost zero mean at the aggregate level. In other words, heterogeneity at the item level in terms of the mean of the sampling bias disappears at the aggregate level.

In addition, the standard deviation of the sampling bias at the aggregate level is only about one-tenth of the standard deviation at the item level. This can be interpreted as a direct consequence of the central limit theorem. Under the assumption that sampling bias is independent across items, the central limit theorem implies

$$\sum_i \omega_i \hat{\delta}_i \xrightarrow{d} N\left(\sum_i \omega_i \mu_i, \sum_i \omega_i^2 \sigma_i^2\right) \tag{3.5}$$

where μ_i and σ_i are the mean and the standard deviation of the sampling bias for item i shown in Figure 7. To see what equation (3.5) means, suppose that the consumption weight ω_i is equal across i. Then, according to equation (3.5), aggregation of the sampling biases over 125 items yields a standard deviation at the aggregate level that is smaller by $1/\sqrt{125}$. Note that the standard deviation at the aggregate level declines for a similar reason even in the case of unequal weights. To check whether this argument based on the central limit theorem applies in this case, we calculate $\sum_i \omega_i \mu_i$ and $\sum_i \omega_i^2 \sigma_i^2$ using the numbers underlying Figure 7. The PDF shown by the red dotted line in Figure 10 is a normal distribution with mean $\sum_i \omega_i \mu_i$ and variance $\sum_i \omega_i^2 \sigma_i^2$. We see that the empirical distribution is close to the normal distribution implied by the central limit theorem, although the central part of the former is slightly higher than that of the latter. In sum, the above findings regarding the mean and the standard deviation of the sampling bias indicate that the sampling bias is much smaller at the aggregate level than at the item level.

To check the robustness of these findings, we calculate the sampling bias using different sampling methods. The results are presented in Figure 11, where the PDF shown by the green line corresponds to the one shown in Figure 10. The red line represents the PDF of the sampling bias obtained using specification #19 in Table 2; that is, the number of regions is assumed to be six; outlet sampling is conducted based on the number of customer visits over the last one month; a sale is defined as a temporary price reduction that lasts less than eight days ($\kappa = 8$), and regular prices are estimated by forward imputation. Note that specification #19 is the same as specification #3 in Figure 10 in these respects, but differs in that it uses product type specifications based on positive characteristics only. On the other hand, the blue line represents the PDF of the sampling bias obtained using specification #7, which differs from specification #3 in that outlet sampling is based on the quantities sold over the last one month rather than the number of customer visits over the last one month. The figure indicates that the mean and the standard deviation for each of the three PDFs are almost identical, confirming that the findings at the aggregate level are robust to changes in sampling methods.

3.2.4 Sampling bias with different time intervals

The sampling bias in Figure 10 is calculated based on month-on-month inflation rates, that is, the ratio of the price level in month t to the price level in month $t-1$, i.e., $P_{i,t}/P_{i,t-1}$, where $P_{i,t}$ is the price level in month t for item i. However, just like aggregation across items reduced the sampling bias, as shown above, so the sampling bias might be smaller if inflation rates are calculated for lower frequencies, that is, if we calculate quarterly or annual inflation rates. To examine this issue, let us define the rate of inflation over time interval τ months by $(P_{i,t}/P_{i,t-\tau})^{12/\tau}$. Note that inflation rates are annualized in this definition so as to make comparison easier between inflation rates for different time intervals. The inflation rate $(P_{i,t}/P_{i,t-\tau})^{12/\tau}$ can be rewritten as

$$\left(\frac{P_{i,t}}{P_{i,t-\tau}}\right)^{12/\tau} = \left[\left(\frac{P_{i,t}}{P_{i,t-1}}\right)^{12}\left(\frac{P_{i,t-1}}{P_{i,t-2}}\right)^{12}\left(\frac{P_{i,t-2}}{P_{i,t-3}}\right)^{12} \times \cdots \times \left(\frac{P_{i,t-\tau-2}}{P_{i,t-\tau-1}}\right)^{12}\left(\frac{P_{i,t-\tau-1}}{P_{i,t-\tau}}\right)^{12}\right]^{1/\tau}, \quad (3.6)$$

so that $(P_{i,t}/P_{i,t-\tau})^{12/\tau}$ can be seen as the geometric mean of monthly inflation $(P_{i,t}/P_{i,t-1})^{12}$ over time. Importantly, since a larger interval implies that the mean is calculated based on more monthly terms, we would expect that the standard deviation of the sampling bias decreases with τ due to the central limit theorem. Figure 12 presents the results of this exercise, with the horizontal axis representing the length of the time interval τ, while the vertical axis represents the sampling bias on an annualized basis. For example, $\tau = 1$ corresponds to month-on-month rates of inflation, based on which Figure 10 was generated. Using the PDF for the sampling bias shown in Figure 10, we calculate the 10th, 20th, 40th, 50th, 60th, 80th, 90th percentiles of the distribution, and plot these percentiles. We repeat this calculation for different values of τ to produce Figure 12, which clearly shows that the standard deviation of the sampling bias monotonically declines with time interval τ.

Let us measure the dispersion of the sampling bias by the distance between the 10th and 90th percentiles. In other words, we look at an 80 percent confidence interval of the sampling bias. The distance is 15.7 percent with $\tau = 1$ but declines to 4.1 percent with $\tau = 6$ and to 2.9 percent with $\tau = 12$. For the case of $\tau = 12$, which represents the sampling bias based on year-on-year inflation rates, the 10th and 90th percentiles are -1.2 percent and 1.7 percent, respectively, while the 50th percentile is 0.04 percent. This implies that the sampling bias with $\tau = 12$ follows a distribution whose upper tail is slightly thicker than its lower tail.[10] Importantly, it also implies that, even when one obtains an estimate of -1.2 percent for the year-on-year inflation rate using purposive sampling, it is still possible, with a probability of 10 percent, that the inflation rate estimated using random sampling is above zero. Similarly, even when one obtains an estimate of +1.7 percent for the year-on-year inflation rate using purposive sampling, it is still possible, with a probability of 10 percent, that the estimated inflation rate using random sampling is below zero, implying that one cannot be confident that deflation is over, even if one observes an inflation rate of 1.7 percent using purposive sampling.

4 Conclusion

[to be completed]

[10] This implies that it is considerably more likely that PS inflation rates are higher than RS than vice versa.

References

[1] Ariga, Kenn, and Kenji Matsui (2003), "Mismeasurement of the CPI," in M. Blomström, J. Corbett, F. Hayashi, A. Kashyap (eds.), *Structural Impediments to Growth in Japan*, University of Chicago Press, 89-128.

[2] Broda, Christian, and David E. Weinstein (2007), "Defining Price Stability in Japan: A View from America," *Monetary and Economic Studies*, Special Edition, Bank of Japan, December 2007, 169-189.

[3] Dalén, Jörgen (1998), "Studies on the Comparability of Consumer Price Indices," *International Statistical Review* 66, 1, 83-113.

[4] De Haan, Jan, Eddy Opperdoes, and Cecile Schut (1999), "Item Selection in the Consumer Price Index: Cut-off versus probability sampling," *Survey Methodology*, Vol. 25, No. 1, 31-41.

[5] Dorfman, Alan H., Janice Lent, Sylvia G. Leaver, and Edward Wegman (2006), "On Sample Survey Designs for Consumer Price Indexes," *Survey Methodology*, Vol. 32, No. 2, 197-216.

[6] Fenwick, David, Adrian Ball, Peter Morgan, and Mick Silver (2003), "Price Collection and Quality Assurance of Item Sampling in the Retail Prices Index: How Can Scanner Data Help?" in R. C. Feenstra and M. D. Shapiro (eds.), *Scanner Data and Price Indexes*, University of Chicago Press, 67-87.

[7] Fuhrer, Jeffrey C., Giovanni P. Olivei, and Geoffrey M. B. Tootell (2011), "Inflation Dynamics when Inflation is Near Zero," Federal Reserve Bank of Boston Working Paper No. 11-17, September 2011.

A Purposive versus Random Sampling

In the purposive sampling implemented in Section 2, we first determine the set of candidate products that meet the product type specifications and then choose some products from the set of candidates following a particular sampling rule. Random sampling, which we consider in Section 3, differs from purposive sampling in that no set of candidate products is determined; instead, specific products are chosen randomly from among all products belonging to a particular item category.

Let us use a simple statistical model to explain how purposive and random sampling differ. Let $\Pi_{i,j}(t)$ denote the price relative for product j, which belongs to item i, in period t. To simplify the exposition, we assume in this appendix that there are only one region and one outlet. We assume that the price relative consists of three different components, and is given by

$$\Pi_{i,j}(t) = U(t)V_i(t)Z_{i,j}(t) \tag{A.1}$$

where $U(t)$ is a component common to all products, while $V_i(t)$ is a component specific to item i (but common to all products belonging to item i), and $Z_{i,j}(t)$ is an idiosyncratic component for product j belonging to item i. We take the log of both sides of the above equation to obtain

$$\pi_{i,j}(t) = u(t) + v_i(t) + z_{i,j}(t) \tag{A.2}$$

where $\pi_{i,j}(t)$, $u(t)$, $v_i(t)$, and $z_{i,j}(t)$ are the logs of $\Pi_{i,j}(t)$, $U(t)$, $V_i(t)$, and $Z_{i,j}(t)$, respectively. Note that $\pi_{i,j}(t)$, $u(t)$, $v_i(t)$, $z_{i,j}(t)$ are all random variables. For simplicity, we assume that $u(t)$ is distributed with mean zero and variance σ_u^2. Similarly, we assume that $v_i(t)$ follows a distribution with mean zero and variance σ_v^2. Note that we assume, for simplicity, that the mean and variance of $v_i(t)$ do not depend on i. Finally, $z_{i,j}(t)$ is distributed with mean μ_j and variance σ_z^2. Again, we assume that the mean of $z_{i,j}(t)$ depends only on j (and not on i) and that the variance depends on neither i nor j.

Let us start with purposive sampling. We assume that there is only a single product, which is indexed by $j = j_0$, in the candidate set for each item category. The log of the relative price constructed using purposive sampling, denoted by $\pi_{i,j}^{PS}(t)$, is then given by

$$\pi_{i,j}^{PS}(t) \equiv u(t) + v_i(t) + z_{i,j_0}(t). \tag{A.3}$$

On the other hand, the log of the relative price constructed using random sampling, $\pi_{i,j}^{RS}(t)$, is given by

$$\pi_{i,j}^{RS}(t) \equiv u(t) + v_i(t) + z_{i,j}(t). \tag{A.4}$$

We assume that the number of prices collected is identical for each item category, which is denoted by n_J. Then the price index based on purposive sampling, denoted by $\pi^{PS}(t)$, is given by

$$\pi^{PS}(t) \equiv \sum_i \omega_i \left(n_J^{-1} \sum_j \pi_{i,j}^{PS}(t) \right) = u(t) + \sum_i \omega_i v_i(t) + \sum_i \omega_i z_{i,j_0}(t) \tag{A.5}$$

where ω_i is the weight for item i. Similarly, the price index based on random sampling, $\pi^{RS}(t)$, is given by

$$\pi^{RS}(t) \equiv \sum_i \omega_i \left(n_J^{-1} \sum_j \pi_{i,j}^{RS}(t) \right) = u(t) + \sum_i \omega_i v_i(t) + \sum_i \omega_i \left(n_J^{-1} \sum_j z_{i,j}(t) \right) \tag{A.6}$$

Using equations (A.5) and (A.6), we calculate the expectation and variance of the price indexes as follows:

$$E\left(\pi^{PS}(t)\right) - E\left(\pi^{RS}(t)\right) = \mu_{j_0} - n_J^{-1} \sum_j \mu_j \tag{A.7}$$

$$V\left(\pi^{PS}(t)\right) - V\left(\pi^{RS}(t)\right) = \sigma_z^2 \left(1 - n_J^{-1}\right) \sum_i \omega_i^2 \tag{A.8}$$

Equation (A.7) indicates that the two price indexes are, on average, different unless the μ's are all identical, implying the presence of sampling bias stemming from non-random sampling in constructing the purposive sampling price index. Equation (A.8) shows that, as n_J becomes greater, the variance of the purposive sampling index increases relative to the variance of the random sampling index. This happens because idiosyncratic shocks to product prices, z, cancel each other out when random sampling is employed, while such an effect is absent in purposive sampling. Note that this relative increase in the variance of the purposive sampling index implies lower measurement efficiency.

To focus on the difference between the purposive sampling index and the random sampling index, we define the measure of the difference for item i, $\delta_i(t)$, as follows:

$$\delta_i(t) \equiv n_J^{-1} \sum_j \pi_{i,j}^{PS}(t) - n_J^{-1} \sum_j \pi_{i,j}^{RS}(t) = z_{i,j_0}(t) - n_J^{-1} \sum_j z_{i,j}(t) \tag{A.9}$$

We then sum this up over i to obtain the measure of the difference at the aggregate level, $\delta(t)$, i.e., $\delta(t) \equiv \sum_i \omega_i \delta_i(t)$. The empirical distributions of $\delta_i(t)$ and of $\delta(t)$ are presented in Section 3 to quantitatively evaluate the size of the sampling bias.

Table 1: Number of Products that Meet the JSB Product Type Specifications

Item code	Description	No. of JAN codes (A)	No. of JAN codes that meet the product specifications (B)	(B/A)	Sales share of products that meet the product specifications
1001	Rice-A (domestic)	11962	1649	0.138	0.179
1002	Rice-B (domestic)	11962	1905	0.159	0.178
1011	Glutinous rice	477	321	0.673	0.935
1031	Boiled noodles	4944	1213	0.245	0.456
1041	Dried noodles	2194	37	0.017	0.002
1042	Spaghetti	1410	237	0.168	0.277
1051	Instant noodles	6879	6	0.001	0.063
1052	Uncooked Chinese noodles	8042	2439	0.303	0.268
1071	Wheat flour	199	71	0.357	0.597
1081	Mochi (rice cakes)	1687	1296	0.768	0.895
1151	Agekamaboko	20029	5129	0.256	0.291
1152	Chikuwa	3556	311	0.087	0.035
1153	Kamaboko	5917	4925	0.832	0.843
1161	Dried bonito fillets	897	9	0.010	0.001
1163	Shiokara (salted fish guts)	1870	989	0.529	0.645
1166	Fish prepared in soy sauce	1236	364	0.294	0.345
1173	Canned fish	1022	108	0.106	0.358
1252	Ham	2245	2065	0.920	0.973
1261	Sausages	5351	4753	0.888	0.940
1271	Bacon	2189	1936	0.884	0.906
1303	Milk	2144	1337	0.624	0.832
1311	Powdered milk	453	3	0.007	0.008
1321	Butter	369	30	0.081	0.458
1331	Cheese	599	23	0.038	0.242
1332	Cheese, imported	442	110	0.249	0.029
1333	Yogurt	557	174	0.312	0.610
1451	Azuki (red beans)	504	243	0.482	0.638
1453	Shiitake mushrooms	3700	57	0.015	0.006
1463	Dried tangle	980	536	0.547	0.482
1471	Bean curd	2914	2581	0.886	0.868
1472	Fried bean curd	2762	181	0.066	0.025
1473	Natto (fermented soybeans)	3809	3271	0.859	0.908
1481	Konnyaku (devil's tongue)	2705	2088	0.772	0.813
1482	Umeboshi, pickled plums	6743	5338	0.792	0.829
1483	Pickled radishes	4544	1383	0.304	0.317
1485	Tangle prepared in soy sauce	5339	2375	0.445	0.806
1486	Pickled Chinese cabbage	2818	1760	0.625	0.694
1487	Kimchi	5155	807	0.157	0.197
1491	Canned sweet corn	643	21	0.033	0.106
1591	Canned fruits	579	83	0.143	0.227
1601	Edible oil	1022	142	0.139	0.567
1602	Margarine	416	12	0.029	0.268
1611	Salt	1005	1	0.001	0.135
1621	Soy sauce	1793	24	0.013	0.234
1631	Soybean paste	5042	530	0.105	0.303
1632	Sugar	197	29	0.147	0.638
1633	Vinegar	636	2	0.003	0.222
1642	Ketchup	397	8	0.020	0.552

Item code	Description	A	B	B/A	Sales share of products that meet the product specifications
1643	Mayonnaise	451	3	0.007	0.205
1644	Jam	3823	5	0.001	0.081
1652	Instant curry mix	743	34	0.046	0.260
1653	Instant soup	1658	7	0.004	0.063
1654	Flavor seasonings	796	2	0.003	0.131
1655	Liquid seasonings	1758	9	0.005	0.339
1656	Granular flavor seasonings	776	2	0.003	0.000
1701	Yokan (sweet bean jelly)	3444	16	0.005	0.006
1711	Castella (sponge cakes)	2185	174	0.080	0.057
1714	Pudding	5280	4	0.001	0.171
1721	Biscuits	13130	4	0.000	0.021
1732	Candies	2067	22	0.011	0.162
1741	Sembei (Japanese crackers)	8314	453	0.054	0.035
1761	Chocolate	1238	8	0.006	0.257
1772	Peanuts	3651	705	0.193	0.124
1781	Chewing gum	1185	18	0.015	0.083
1782	Ice cream	1494	1	0.001	0.125
1791	Box lunch	21254	905	0.043	0.021
1793	Rice balls	7647	467	0.061	0.145
1794	Frozen pilaf	999	36	0.036	0.163
1811	Salad	11165	513	0.046	0.069
1812	Boiled beans	808	639	0.791	0.883
1851	Frozen croquettes	1167	64	0.055	0.039
1871	Cooked curry	3321	18	0.005	0.316
1881	Gyoza	3201	626	0.196	0.196
1891	Mazegohan no moto	303	3	0.010	0.367
1902	Green tea	5614	4329	0.771	0.602
1911	Black tea	1469	8	0.005	0.211
1914	Tea beverages	505	48	0.095	0.379
1921	Instant coffee	975	27	0.028	0.162
1922	Coffee beans	678	16	0.024	0.148
1923	Coffee beverages	3576	1184	0.331	0.620
1930	Fruit juice	2689	185	0.069	0.162
1931	Beverages which contain juice	2202	17	0.008	0.210
1941	Vegetable juice	353	2	0.006	0.307
1951	Carbonated beverages	400	4	0.010	0.047
1971	Fermented lactic drinks, sterilized ("Calpis")	231	3	0.013	0.657
1981	Sports soft drinks	341	15	0.044	0.311
1982	Mineral water	1887	14	0.007	0.233
2003	Sake	6747	168	0.025	0.372
2011	Shochu (distilled spirits)	6691	32	0.005	0.172
2021	Beer	2430	246	0.101	0.391
2026	Low-malt beer	1389	157	0.113	0.308
2033	Whisky	1689	8	0.005	0.169
2041	Wine	21123	249	0.012	0.092
4401	Food wrap	993	14	0.014	0.180
4412	Facial tissue	1295	81	0.063	0.503
4413	Rolled toilet paper	2944	415	0.141	0.214
4431	Liquid detergent, kitchen	1212	21	0.017	0.076
4441	Detergent, laundry	866	144	0.166	0.457
4442	Fabric softener	836	43	0.051	0.410

Item code	Description	A	B	B/A	Sales share of products that meet the product specifications
4451	Insecticide	132	7	0.053	0.114
4461	Moth repellent for clothes	736	57	0.077	0.232
4471	Fragrance	1034	70	0.068	0.186
6095	Bath preparations	8648	54	0.006	0.059
6101	Sanitary napkins	2155	33	0.015	0.045
9111	Ball-point pens	15380	53	0.003	0.026
9115	Marking pens	1604	32	0.020	0.127
9121	Notebooks	13805	23	0.002	0.004
9124	Cellophane adhesive tape	1262	4	0.003	0.015
9127	Papers for office automation	518	97	0.187	0.766
9193	Dog food	2049	190	0.093	0.067
9195	Dry batteries	112	31	0.277	0.762
9196	Cat food	4250	580	0.136	0.332
9611	Toothbrushes	2388	32	0.013	0.102
9621	Toilet soap	2802	35	0.012	0.228
9622	Shampoo	4410	238	0.054	0.230
9623	Toothpaste	1255	21	0.017	0.110
9624	Hair conditioner	2932	138	0.047	0.185
9625	Hair dye	4200	37	0.009	0.077
9631	Hair liquid	380	2	0.005	0.255
9641	Hair tonic	233	5	0.021	0.192
9652	Face cream-B	1982	10	0.005	0.021
9661	Toilet lotion	5251	63	0.012	0.023
9672	Foundation-B	12600	74	0.006	0.024
9682	Lipsticks-B	18723	262	0.014	0.041
9692	Milky lotion-B	2157	18	0.008	0.018

Table 2: List of Purposive Sampling Simulations

	Regions	Outlet sampling	Product sampling	Range of product characteristics	Treatment of sale prices
#1	Single	One month customer visits	One month sales	Full range	8 days & forward imputation
#2	Single	One month customer visits	One month sales	Full range	3 days & forward imputation
#3	Six	One month customer visits	One month sales	Full range	8 days & forward imputation
#4	Six	One month customer visits	One month sales	Full range	3 days & forward imputation
#5	Single	One month sales	One month sales	Full range	8 days & forward imputation
#6	Single	One month sales	One month sales	Full range	3 days & forward imputation
#7	Six	One month sales	One month sales	Full range	8 days & forward imputation
#8	Six	One month sales	One month sales	Full range	3 days & forward imputation
#9	Single	One month customer visits	One month sales	Full range	8 days & backward imputation
#10	Single	One month customer visits	One month sales	Full range	3 days & backward imputation
#11	Six	One month customer visits	One month sales	Full range	8 days & backward imputation
#12	Six	One month customer visits	One month sales	Full range	3 days & backward imputation
#13	Single	One month sales	One month sales	Full range	8 days & backward imputation
#14	Single	One month sales	One month sales	Full range	3 days & backward imputation
#15	Six	One month sales	One month sales	Full range	8 days & backward imputation
#16	Six	One month sales	One month sales	Full range	3 days & backward imputation
#17	Single	One month customer visits	One month sales	Positive only	8 days & forward imputation
#18	Single	One month customer visits	One month sales	Positive only	3 days & forward imputation
#19	Six	One month customer visits	One month sales	Positive only	8 days & forward imputation
#20	Six	One month customer visits	One month sales	Positive only	3 days & forward imputation
#21	Single	One month sales	One month sales	Positive only	8 days & forward imputation
#22	Single	One month sales	One month sales	Positive only	3 days & forward imputation
#23	Six	One month sales	One month sales	Positive only	8 days & forward imputation
#24	Six	One month sales	One month sales	Positive only	3 days & forward imputation
#25	Single	One month customer visits	One month sales	Positive only	8 days & backward imputation
#26	Single	One month customer visits	One month sales	Positive only	3 days & backward imputation
#27	Six	One month customer visits	One month sales	Positive only	8 days & backward imputation
#28	Six	One month customer visits	One month sales	Positive only	3 days & backward imputation
#29	Single	One month sales	One month sales	Positive only	8 days & backward imputation
#30	Single	One month sales	One month sales	Positive only	3 days & backward imputation
#31	Six	One month sales	One month sales	Positive only	8 days & backward imputation
#32	Six	One month sales	One month sales	Positive only	3 days & backward imputation

	Regions	Outlet sampling	Product sampling	List of product types	Treatment of sale prices
#33	Single	Three month customer visits	Three month sales	Full range	8 days & forward imputation
#34	Single	Three month customer visits	Three month sales	Full range	3 days & forward imputation
#35	Six	Three month customer visits	Three month sales	Full range	8 days & forward imputation
#36	Six	Three month customer visits	Three month sales	Full range	3 days & forward imputation
#37	Single	Three month sales	Three month sales	Full range	8 days & forward imputation
#38	Single	Three month sales	Three month sales	Full range	3 days & forward imputation
#39	Six	Three month sales	Three month sales	Full range	8 days & forward imputation
#40	Six	Three month sales	Three month sales	Full range	3 days & forward imputation
#41	Single	Three month customer visits	Three month sales	Full range	8 days & backward imputation
#42	Single	Three month customer visits	Three month sales	Full range	3 days & backward imputation
#43	Six	Three month customer visits	Three month sales	Full range	8 days & backward imputation
#44	Six	Three month customer visits	Three month sales	Full range	3 days & backward imputation
#45	Single	Three month sales	Three month sales	Full range	8 days & backward imputation
#46	Single	Three month sales	Three month sales	Full range	3 days & backward imputation
#47	Six	Three month sales	Three month sales	Full range	8 days & backward imputation
#48	Six	Three month sales	Three month sales	Full range	3 days & backward imputation
#49	Single	Three month customer visits	Three month sales	Positive only	8 days & forward imputation
#50	Single	Three month customer visits	Three month sales	Positive only	3 days & forward imputation
#51	Six	Three month customer visits	Three month sales	Positive only	8 days & forward imputation
#52	Six	Three month customer visits	Three month sales	Positive only	3 days & forward imputation
#53	Single	Three month sales	Three month sales	Positive only	8 days & forward imputation
#54	Single	Three month sales	Three month sales	Positive only	3 days & forward imputation
#55	Six	Three month sales	Three month sales	Positive only	8 days & forward imputation
#56	Six	Three month sales	Three month sales	Positive only	3 days & forward imputation
#57	Single	Three month customer visits	Three month sales	Positive only	8 days & backward imputation
#58	Single	Three month customer visits	Three month sales	Positive only	3 days & backward imputation
#59	Six	Three month customer visits	Three month sales	Positive only	8 days & backward imputation
#60	Six	Three month customer visits	Three month sales	Positive only	3 days & backward imputation
#61	Single	Three month sales	Three month sales	Positive only	8 days & backward imputation
#62	Single	Three month sales	Three month sales	Positive only	3 days & backward imputation
#63	Six	Three month sales	Three month sales	Positive only	8 days & backward imputation
#64	Six	Three month sales	Three month sales	Positive only	3 days & backward imputation

Table 3: Number of Outlets, Products, and Observations

	No. of outlets	Entries	Exits	No. of products	No. of observations
2000	185	21	5	174,928	242,357,320
2001	185	1	1	176,504	274,319,027
2002	186	14	13	180,355	283,433,216
2003	185	2	3	172,150	290,910,066
2004	168	14	31	182,661	282,074,675
2005	183	19	4	190,256	309,888,190
2006	186	7	4	206,287	329,139,639
2007	266	93	13	236,825	386,389,129
2008	257	4	13	234,660	419,941,109
2009	260	7	4	230,483	422,389,029
2010	256	0	4	223,810	410,358,552

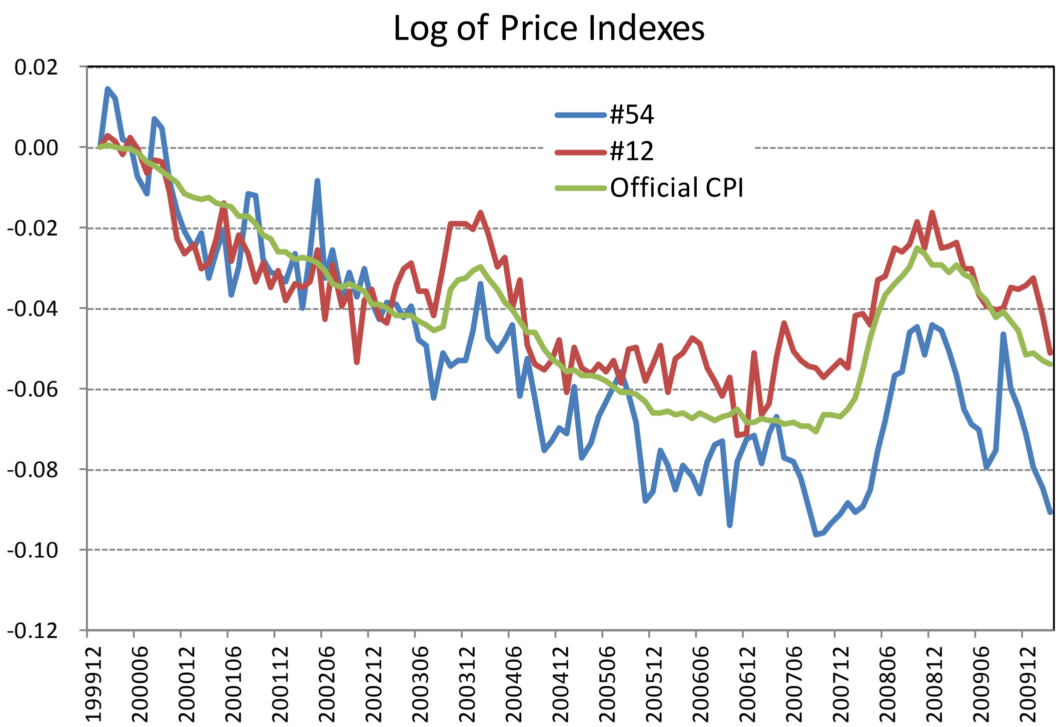
Table 4: Turnover of Products in the 103 Outlets

	No. of products in the 103 outlets	Entries	Exits	Entry rate	Exit rate
2000	203,563	-	-	-	-
2001	208,164	57,526	52,925	0.276	0.254
2002	217,139	66,035	57,060	0.304	0.263
2003	206,172	51,696	62,663	0.251	0.304
2004	222,486	74,655	58,341	0.336	0.262
2005	224,705	62,158	59,939	0.277	0.267
2006	242,669	80,361	62,397	0.331	0.257
2007	254,887	78,060	65,842	0.306	0.258
2008	268,541	89,557	75,903	0.333	0.283
2009	256,824	75,495	87,212	0.294	0.340

Table 5: Results of Purposive Sampling Simulations

#	Mean of monthly inflation (percent)	Std. dev. of monthly inflation (percent)	#	Mean of monthly inflation (percent)	Std. dev. of monthly inflation (percent)
1	-0.041	0.669	33	-0.043	0.711
2	-0.037	0.721	34	-0.041	0.737
3	-0.039	0.673	35	-0.051	0.630
4	-0.036	0.720	36	-0.050	0.701
5	-0.047	0.790	37	-0.047	0.711
6	-0.048	0.844	38	-0.047	0.758
7	-0.038	0.843	39	-0.041	0.736
8	-0.037	0.875	40	-0.039	0.761
9	-0.038	0.660	41	-0.040	0.695
10	-0.036	0.696	42	-0.040	0.725
11	-0.037	0.647	43	-0.049	0.616
12	-0.035	0.710	44	-0.049	0.692
13	-0.044	0.784	45	-0.044	0.707
14	-0.047	0.837	46	-0.045	0.752
15	-0.036	0.834	47	-0.038	0.734
16	-0.035	0.867	48	-0.038	0.755
17	-0.075	0.815	49	-0.072	0.815
18	-0.074	0.830	50	-0.070	0.827
19	-0.055	0.833	51	-0.068	0.708
20	-0.054	0.805	52	-0.066	0.708
21	-0.077	1.087	53	-0.080	0.854
22	-0.080	1.126	54	-0.081	0.889
23	-0.061	0.998	55	-0.062	0.837
24	-0.058	0.986	56	-0.060	0.844
25	-0.073	0.797	57	-0.070	0.815
26	-0.079	0.795	58	-0.069	0.823
27	-0.054	0.807	59	-0.067	0.692
28	-0.053	0.800	60	-0.066	0.699
29	-0.075	1.075	61	-0.077	0.845
30	-0.079	1.119	62	-0.079	0.889
31	-0.058	0.986	63	-0.059	0.835
32	-0.057	0.980	64	-0.058	0.843

Figure 1: Price Indexes Based on Purposive Sampling



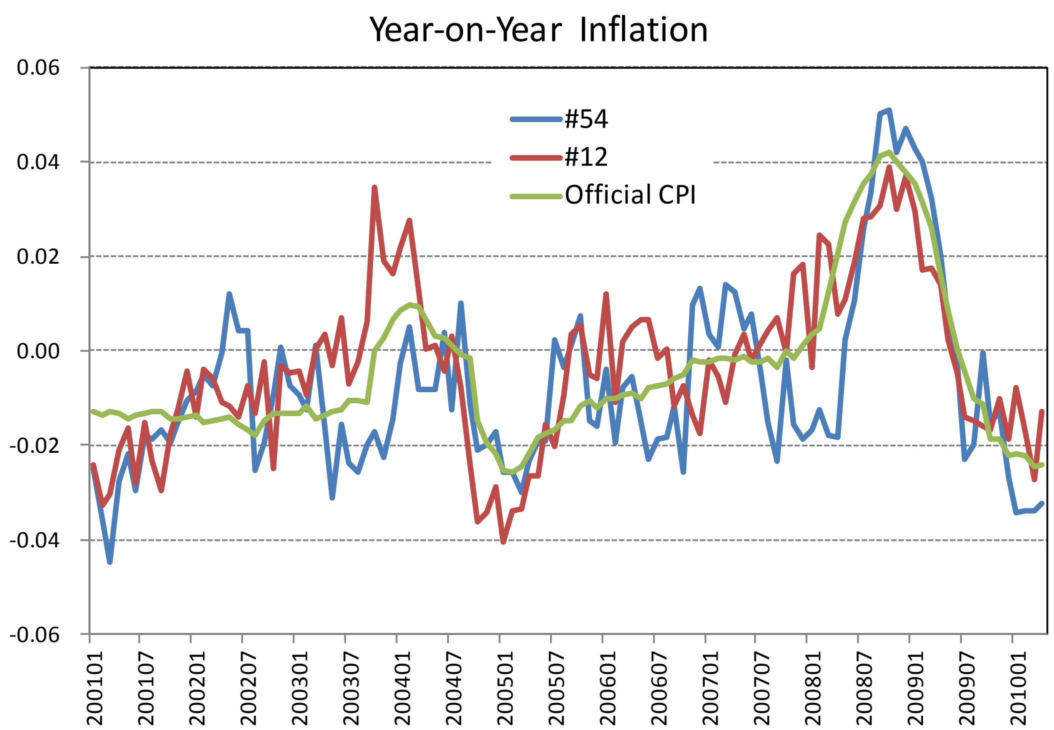
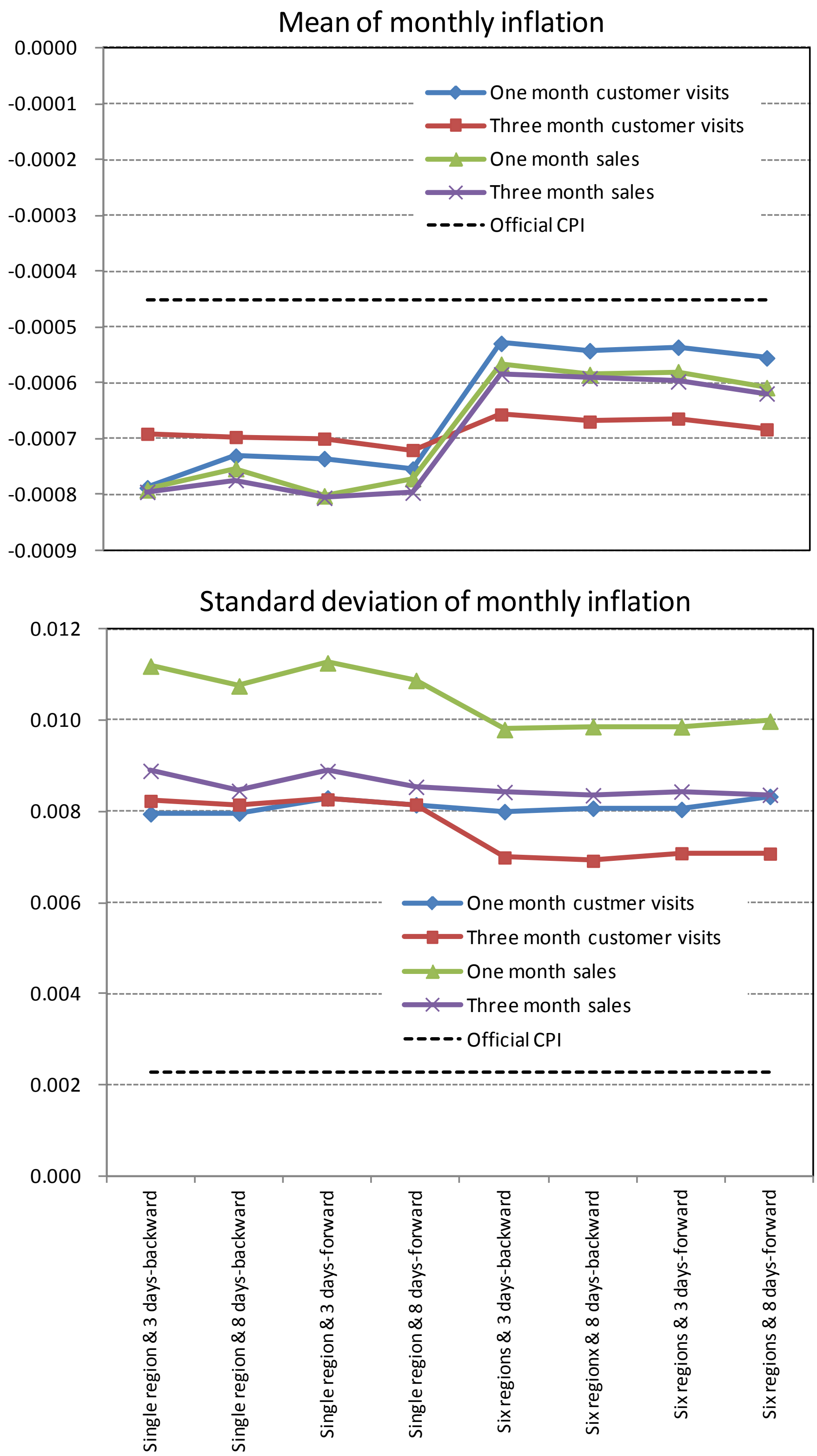


Figure 2: Outlet and Product Sampling

(a) Sampling with positive characteristics only



(b) Sampling with full range of characteristics

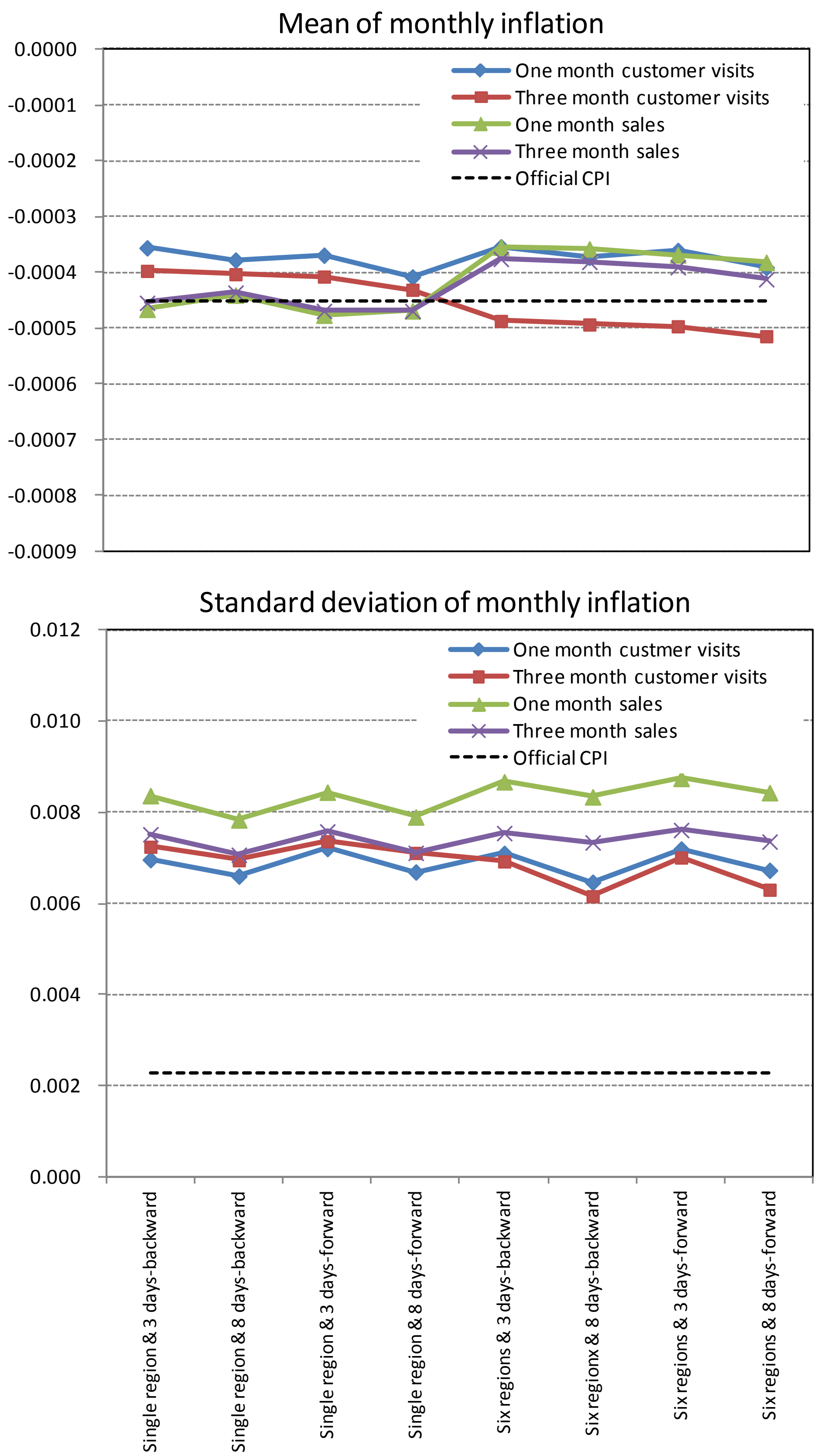
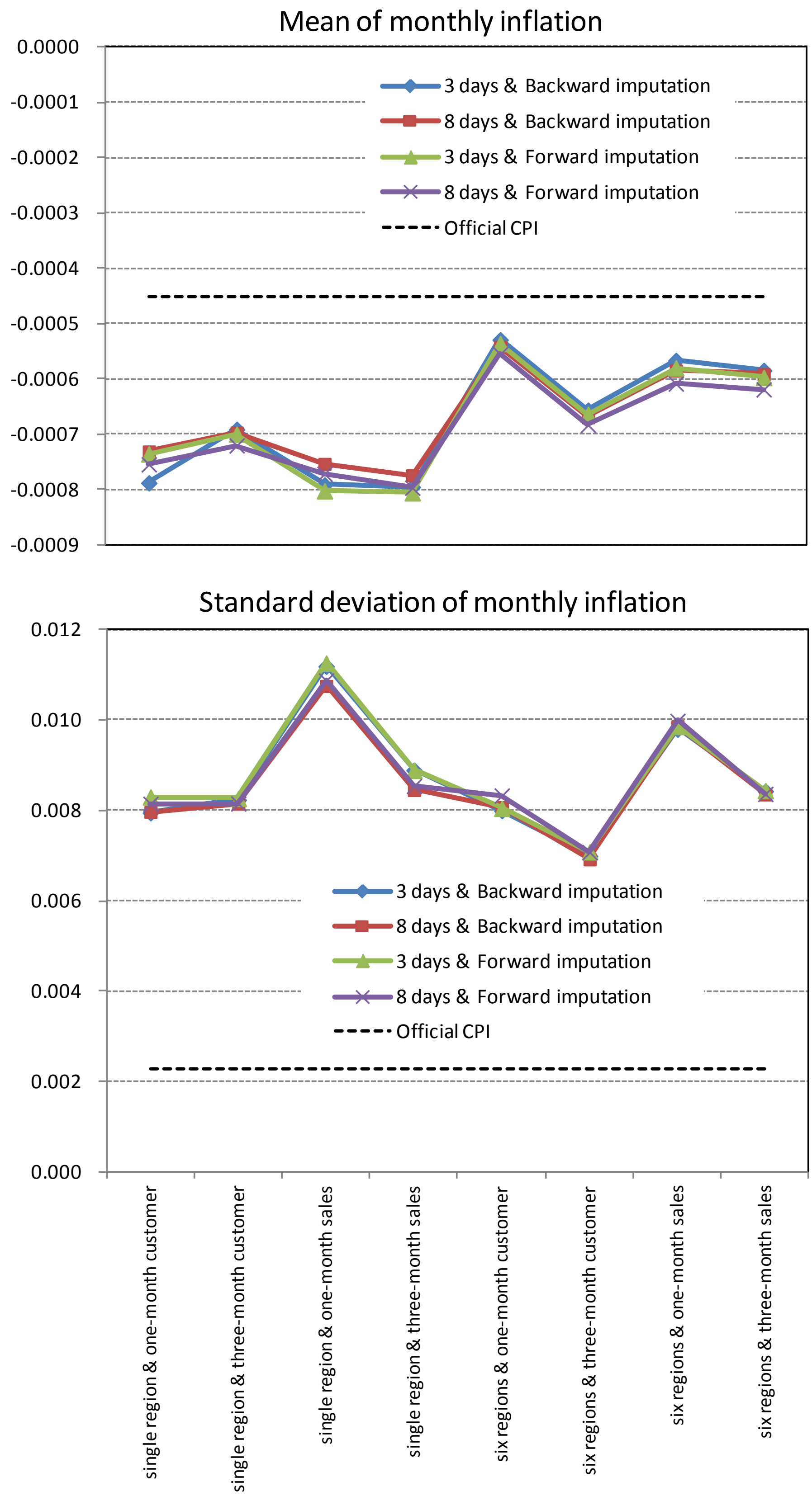
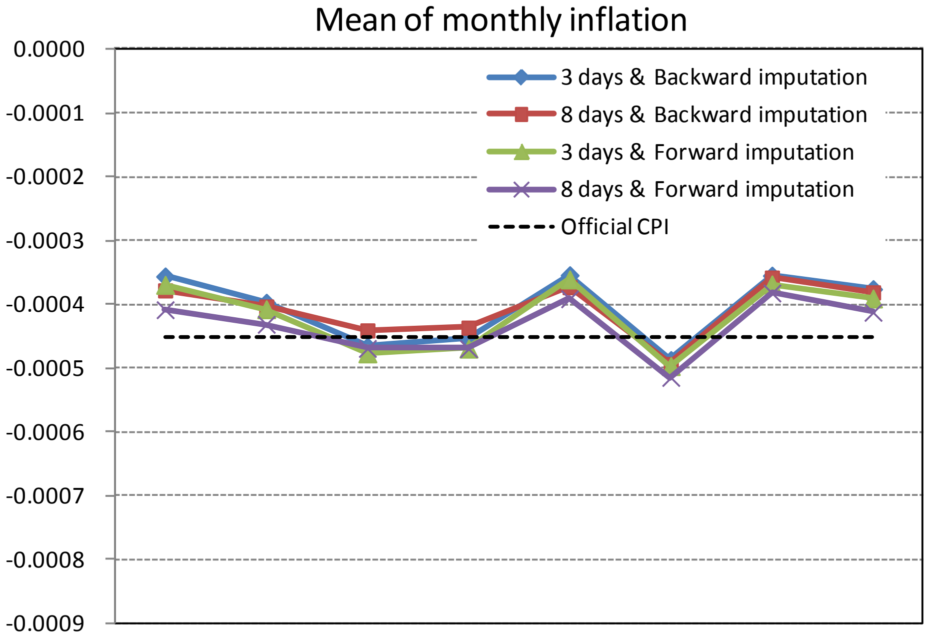


Figure 3: Treatment of Sale Prices

(a) Sampling with positive characteristics only



(b) Sampling with full range of characteristics



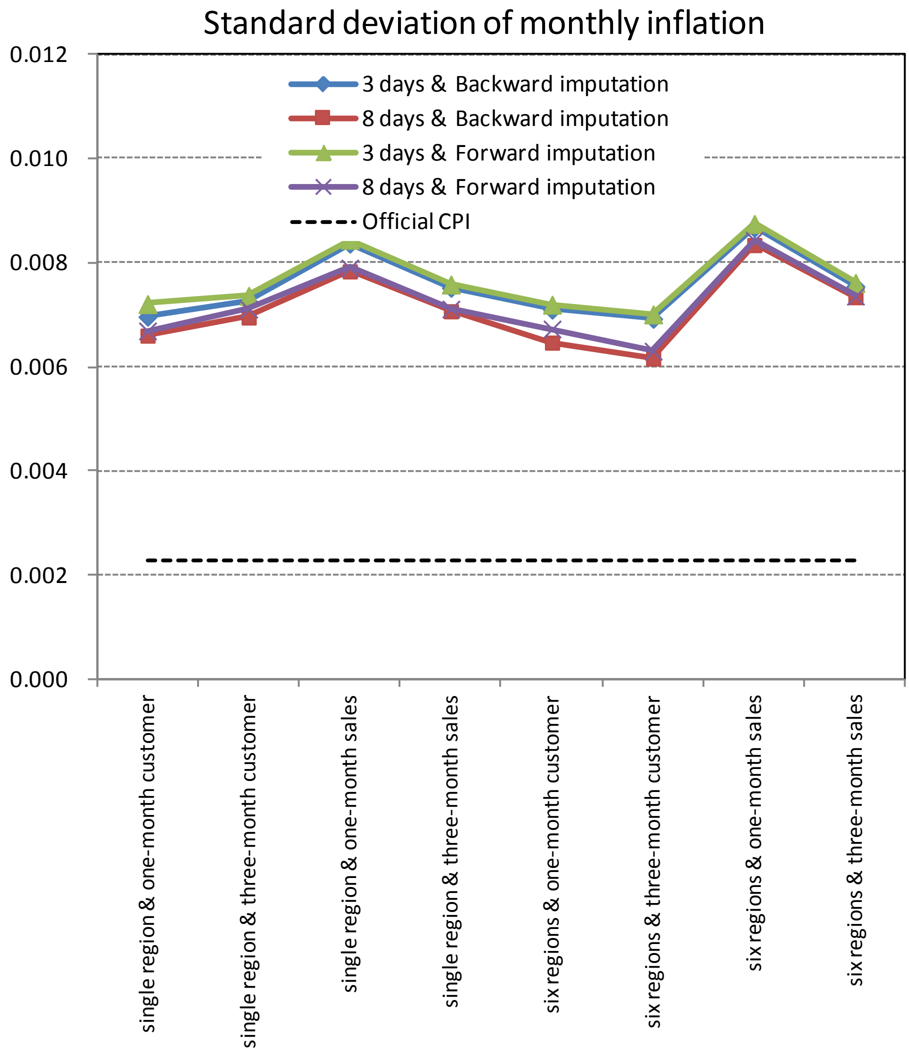


Figure 4: Year-on-Year Inflation Estimated Using Random Sampling

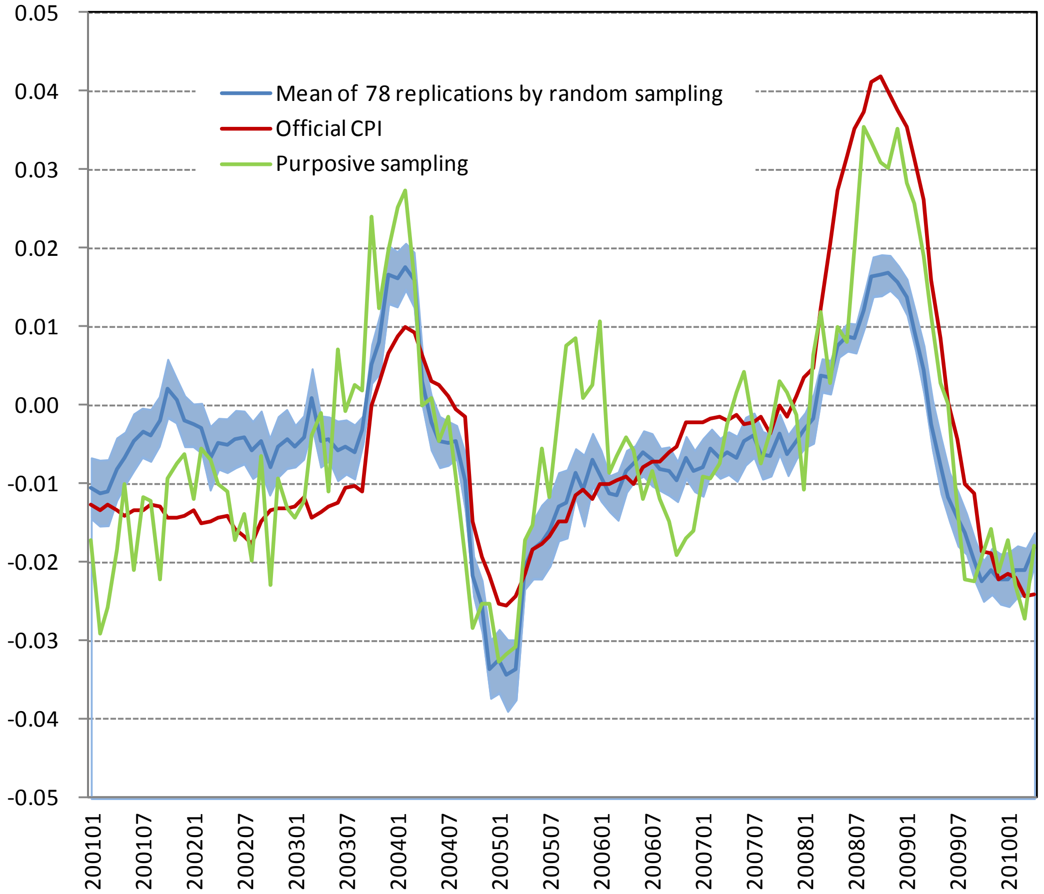


Figure 5: Arithmetic vs. Geometric Mean for Lower Level Aggregation

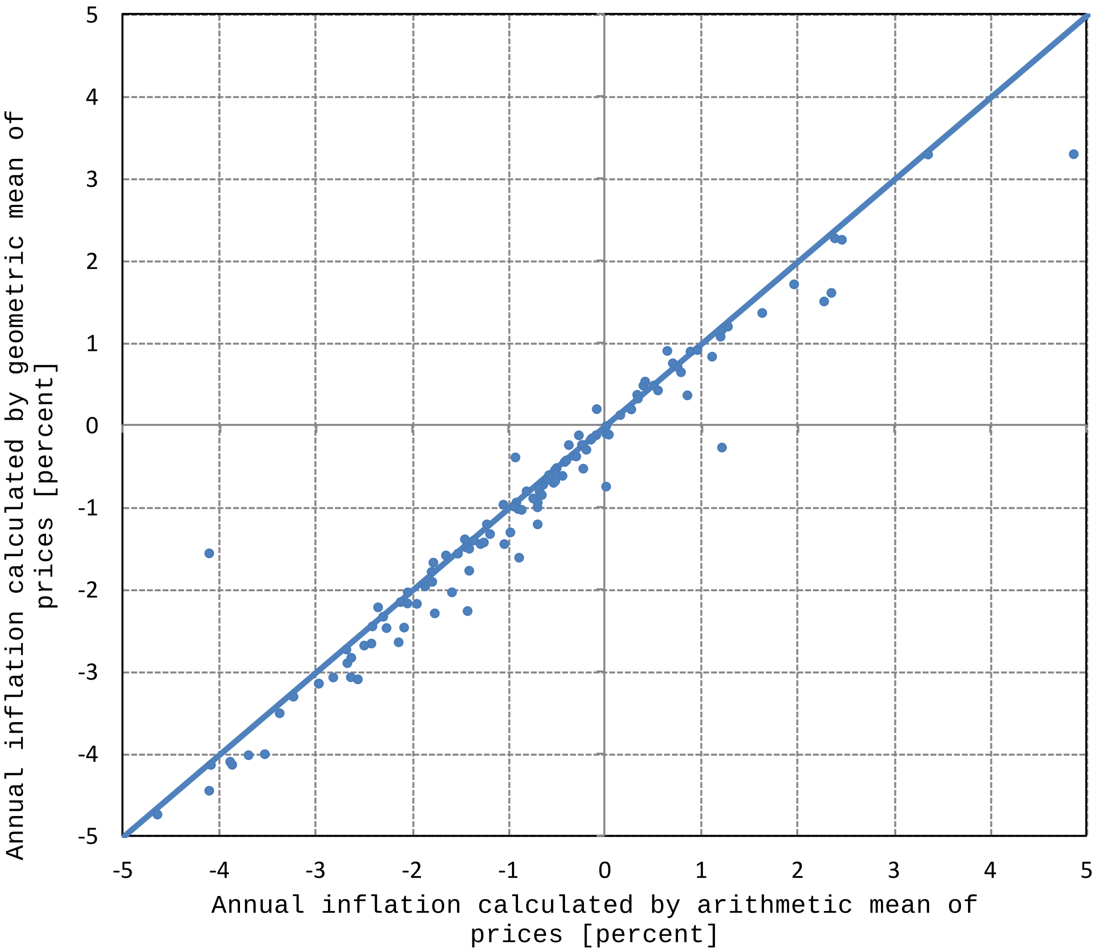


Figure 6: Sampling Bias for Margarine

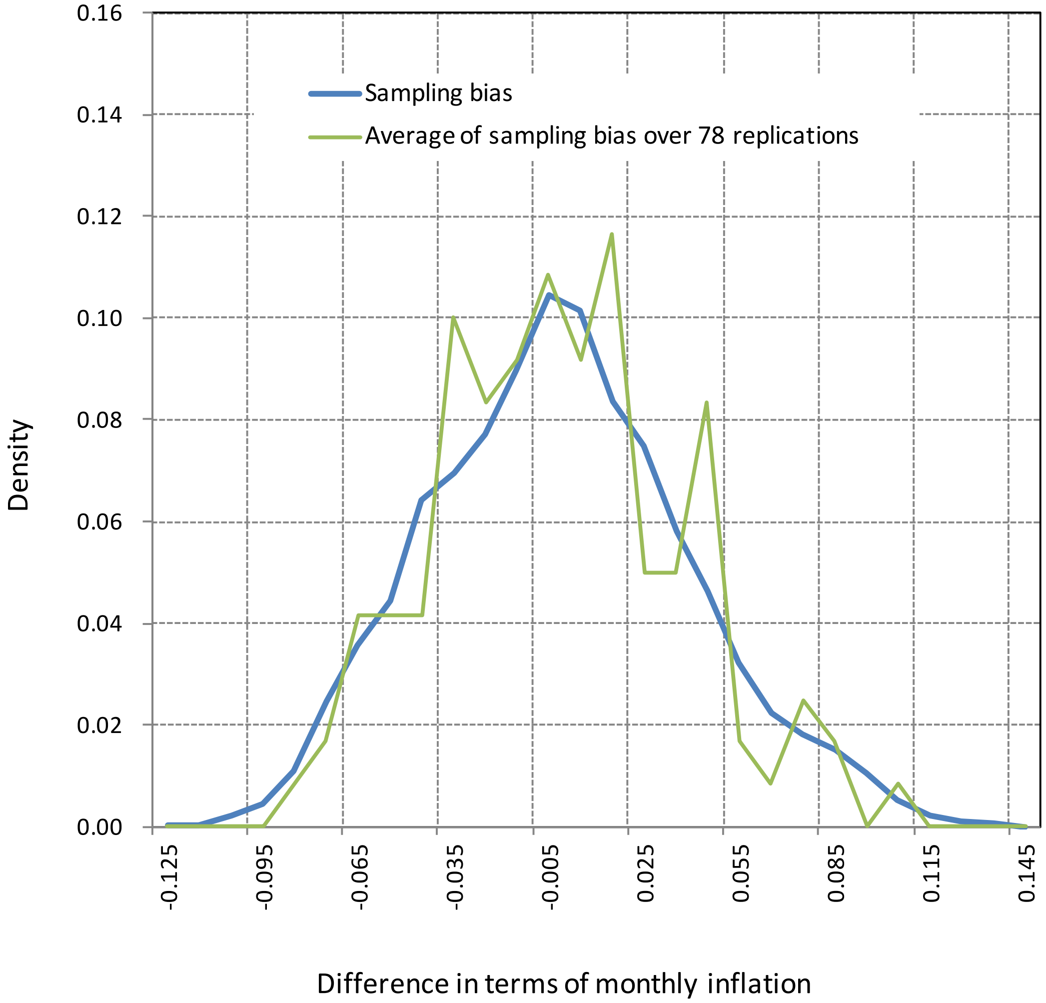


Figure 7: Sampling Bias by Item

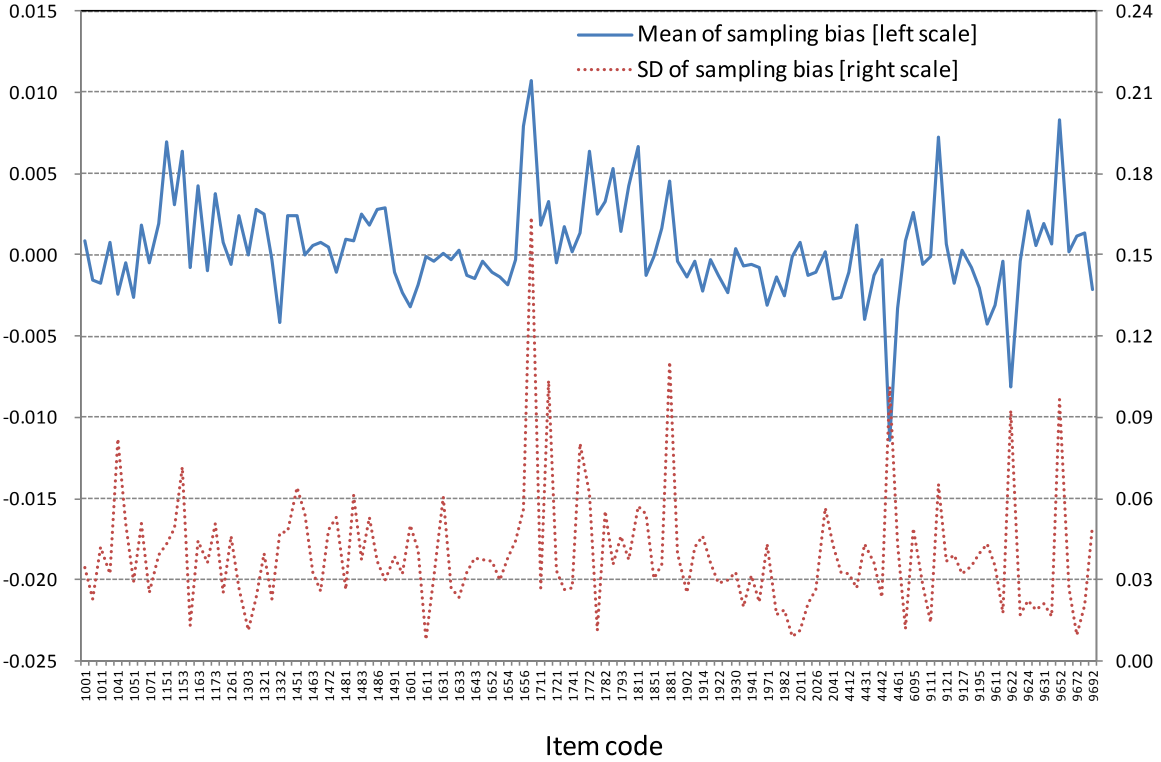


Figure 8: Overlap of Products that Meet the JSB Product Type Specifications among Those Picked by Random Sampling

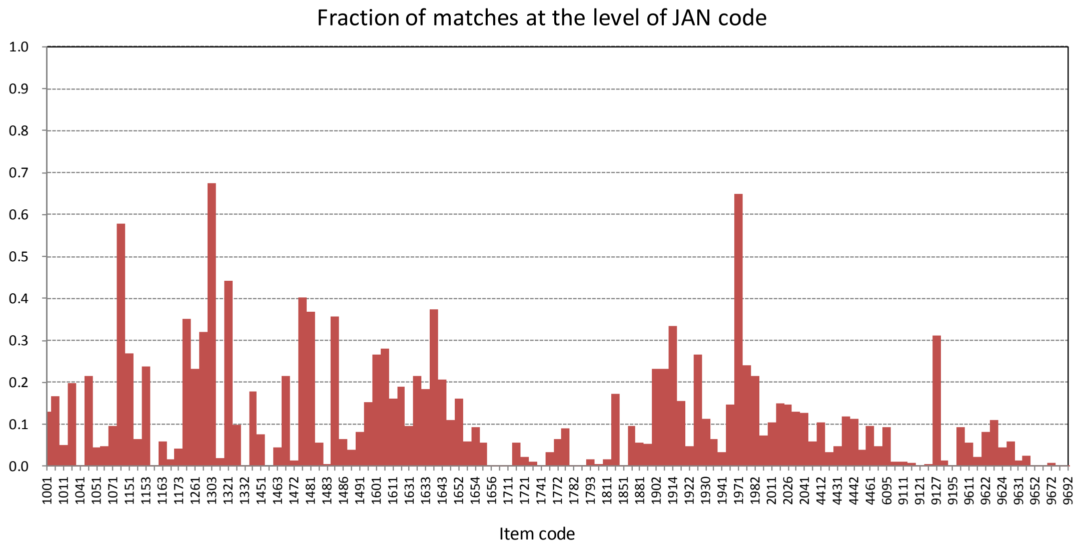


Figure 9: Is the Small Overlap Due to Tight JSB Product Type Specifications?

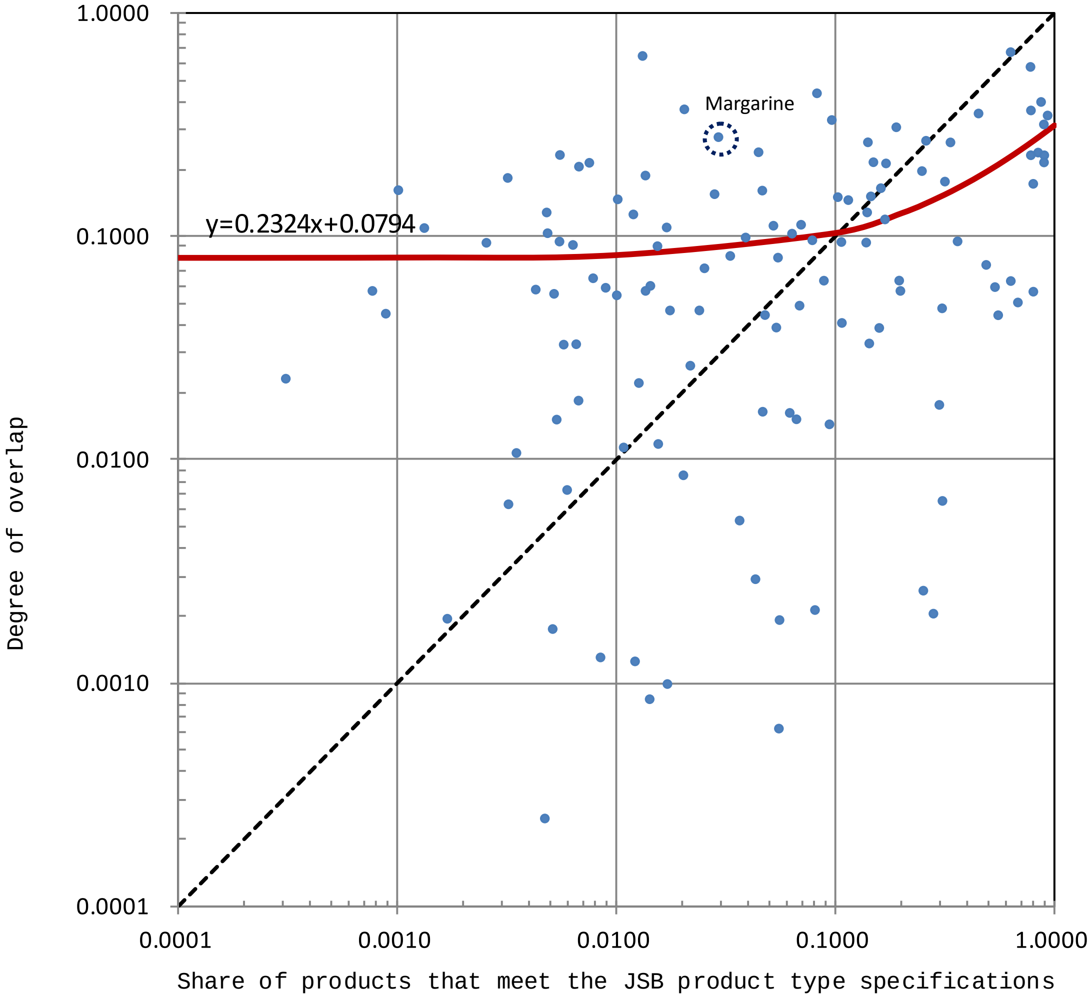


Figure 10: Sampling Bias at the Aggregate Level

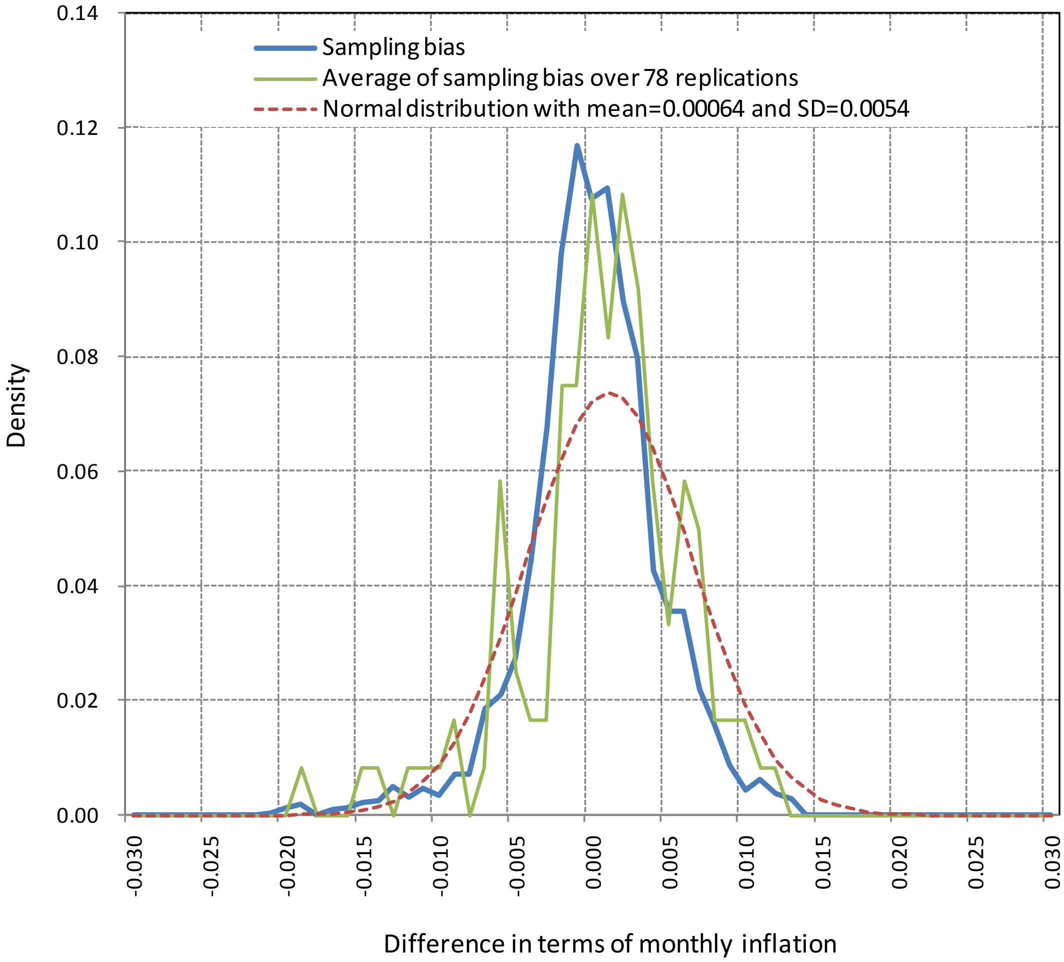


Figure 11: Sampling Bias for Alternative Price Index Definitions

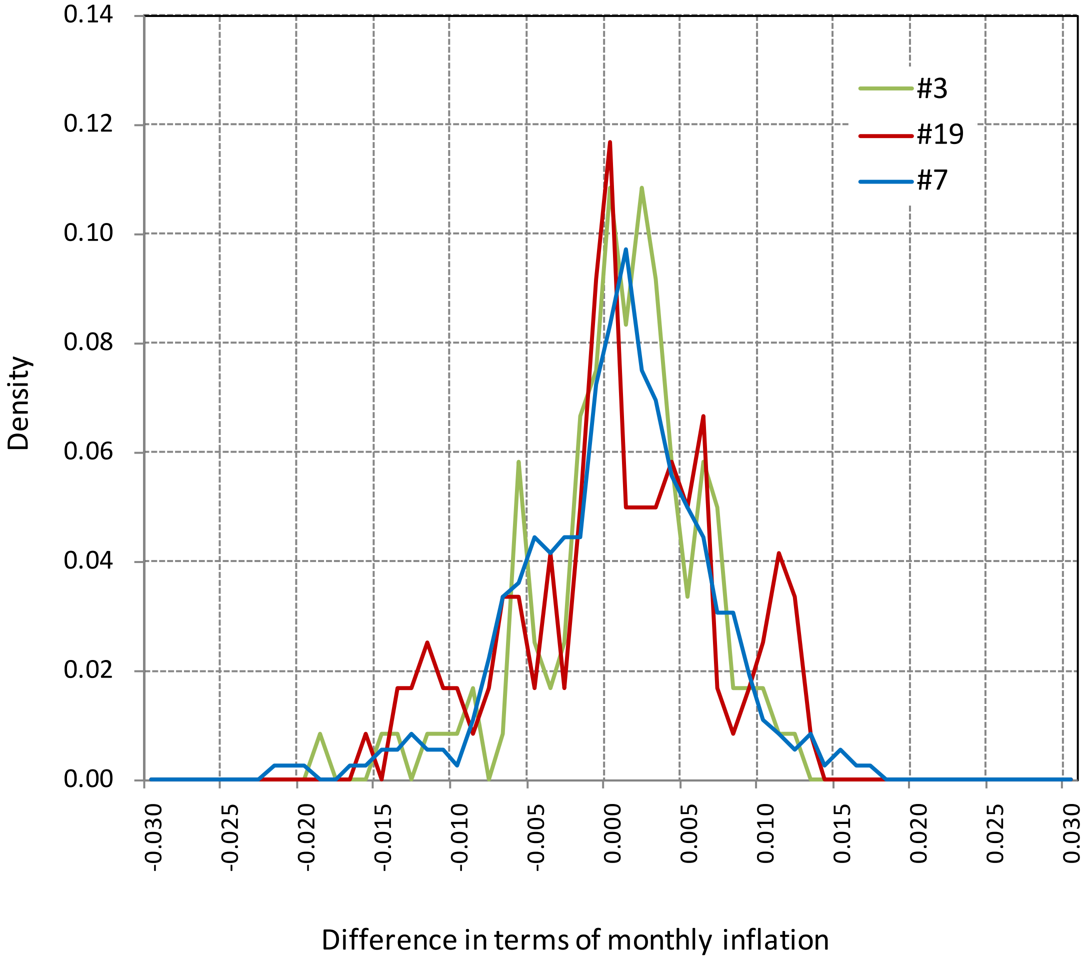


Figure 12: Sampling Bias for Different Time Intervals