Graduate School of Economics

Hitotsubashi University

Discussion Paper Series No.2014-21

Estimating Upward Bias of Japanese Consumer Price Index Using Engel's Law

by

Kazuhito Higa

December 2014

Estimating Upward Bias of Japanese Consumer Price Index Using Engel's Law

Kazuhito Higa*

Graduate School of Economics, Hitotsubashi University

December 17, 2014

Abstract

The Japanese Consumer Price Index (CPI) is considered to be upwardly biased. This paper estimated the Engel curve based on National Survey of Family Income and Expenditure data to measure the bias. The estimated bias for the period 1989 to 2004 was 0.53 percentage points per annum. Correcting the bias led to a lower inflation rate of 0.14 percent per year, against the official inflation rate of 0.65 percent during the period. A demographic analysis showed that a household with a non-working spouse faced a larger bias suggesting that the opportunity cost of shopping determines the size of the bias.

JEL Classification Numbers: C10 E31 D12
Keywords: consumer price index bias; Engel's Law

* I appreciate comments from Daiji Kawaguchi, Naohito Abe, Miki Kohara, and attendances of 7th Applied Econometrics Conference. The data used in this paper are provided by Research Centre for Information and Statistics of Social Science Institute of Economics Research, Hitotsubashi University. This study is supported by Global COE project of Hitotsubashi Unviersity.

1 Introduction

The general consumer price index (CPI) represents a cost of living and an inflation. The nature of CPI as the cost of living index entails many policies, such as public pension benefit, index-linked to the CPI. Despite the importance of the CPI as the cost of living index, it has been criticized for its upward bias based on several grounds including consumers' substitution behavior and improper sampling of surveyed items (Shiratsuka, 1999). The overstatement of the cost of living index directly leads to the overcompensation of the beneficiaries and the excess government spending. Moreover, the Bank of Japan recently adopted the 2 percent annual inflation targeting measured by CPI. This policy change makes the analysis of CPI bias even more important.

Several studies have estimated the bias of the Japanese CPI. Shiratsuka (1999) has estimated the size of the overall bias as 0.9 percentage points per year. Ariga and Matsui (2003) have concluded that the CPI is upwardly biased by at least 0.5 percentage points per year. Measuring the bias of the Japanese CPI based on the calculation methodology and bias estimates of the US CPI, Broda and Weinstein (2007) have concluded that the Japanese CPI overstates the cost of living by around 1.8 percentage points per year. These papers have analyzed the CPI bias by its causes such as a limitation of the price survey, a calculation method, and the exclusion of the effects of a new product rotation and a quality improving. Then adding up each cause, they estimated the size of the CPI bias. Watanabe and Watanabe (2014) constructed a daily price index using scanner data. They have found that the CPI inflation rate is 0.5 percentage points higher than the inflation rate based on their price index.

This paper, following a different approach from previous studies, applied the Engel curve method suggested by Hamilton (2001) and Costa (2001), which study the bias of the US CPI, to estimate the over all bias of the Japanese CPI. The Engel curve method exploits the fact that the change of food expenditure share represents the change of real income after accounting for the change of relative food price. Then, if the CPI correctly represents the cost of living and household preference is stable, the CPI-deflated Engel curve should be stable across different periods after the relative price change in food and household characteristics are controlled for. In this approach, a shift of the Engel curve from one year to another indicates CPI bias. Several previous studies rely on this method (Beatty and Larsen, 2005 for Canada; Barrett and Brzozowski, 2010 for Australia; Chung et al., 2010 for Korea). In contrast to previous Japanese studies in which the CPI bias is analyzed in terms of its causes, this approach estimates the overall CPI bias.

In addition, this paper analyzed the cause of bias from the point of the opportunity cost of shopping. Aguiar and Hurst (2007) and Abe and Shiotani (2012) have analyzed the relationship between prices that households face and the frequency of shopping. Both studies suggested that households, such as the elderly, who have enough time to find lower prices for commodities face a lower inflation. Based on the same idea, therefore, this paper estimated bias between the households with a working or a non-working spouse. A household with enough time for shopping would be able to seek out low-priced or discount goods and substitute them for expensive goods. As a result, the household might face lower prices that result in a larger upward bias of the official CPI.

This paper obtained the following results. The overall bias between 1989 and 2004 was 0.53 percentage points per year for households with a household head aged 20 to 59 years old. This bias is substantial considering the fact that the official inflation rate during the analysis period was 0.65 percent per year. In addition, from an analysis based on household characteristics, households with a non-working spouse faced an upward bias of 0.55 percentage points per year, substantially larger than the bias of 0.45 percentage points per year for households with a working spouse. Thus, households with a working spouse faced a lower inflation. This finding is presumably because those household with a working spouse can spend time on shopping searching for bargains.

2 Model

This paper followed the approach of Hamilton (2001) and Costa (2001) to estimate the Japanese CPI bias. This method depends on five assumptions. First, the income elasticity of food is assumed to be below unity. This assumption assures that the food share declines as real income increases and allows an inference of the real income from the food share. Second, food has no durability so that food expenditure is identical to food consumption. This implies that food expenditure in a period cannot produce a flow of food consumption to another period. Third, the food component can be separated from the non-food component in a utility function. This assumes that CPI bias of non-food items does not affect the food share through an unexpected mechanism. Fourth, household preference is stable over time. This condition ensures that the shift of the Engel curve is caused only by CPI bias. To control the change in household preference, I include several household demographic variables in a regression equation. Fifth, this paper assumes that the relative CPI bias between food and non-food is constant over time. This implies that the bias between food price and non-food price are the same size. Since this might be a strong assumption, I will discuss the effect of violation of this assumption on the bias estimates in the discussion section.

To estimate CPI bias, this paper goes through the following procedure. Firstly, an Engel curve is derived from the Almost Ideal Demand System by Deaton and Muellbauer (1980), and it can be written as a linear function of the food share and the log of real income:

$$\omega_{it} = \phi + \gamma(\ln P_t^f - \ln P_t^{nf}) + \beta(\ln Y_{it} - \ln P_t) + X'\theta + u_{it}, \tag{1}$$

where ω_{it} is the food-income ratio of a household i at time t and Y is the nominal household income. The terms P^f, P^{nf}, and P are true but unobservable price indices of food, non-food items, and all goods, respectively. The vector X represents the household characteristics, and u is the error term.

Secondly, the true price index of all goods is defined as the weighted average of the true price indices of food and non-food items:

$$\ln P_t = \alpha \ln P_t^f + (1-\alpha) \ln P_t^{nf}. \tag{2}$$

Each true price index is obtained by removing the error from the observed price index:

$$\ln P_t^G = \ln P_t^{'G} + \ln(1 + E_t^G), \tag{3}$$

where $P'$is the observed price index and G represents food, non-food items, and all goods. The term E_t is the percent cumulative measurement error in the cost of living index from time 0 to t. In this paper, the term E_t represents the cumulative CPI bias.

Thirdly, by defining $y \equiv \ln Y$, $\pi \equiv \ln P'$, and $\varepsilon \equiv \ln(1+E)$ and substituting two price assumptions, I get the Engel curve to be estimated:

$$\begin{aligned}\omega_{it} &= \phi + \gamma(\pi_t^f - \pi_t^{nf}) + \beta(y_{it} - \pi_t) + X'\theta + \gamma(\varepsilon_t^f - \varepsilon_t^{nf}) - \beta\varepsilon_t + u_{it} \\ &= \phi + \gamma(\pi_t^f - \pi_t^{nf}) + \beta(y_{it} - \pi_t) + X'\theta + \sum_{t=1}^{T} \delta_t D_t + u_{it} \end{aligned} \tag{4}$$

where D_t represents the year dummy and $\delta_t = \gamma(\varepsilon_t^f - \varepsilon_t^{nf}) - \beta\varepsilon_t$. Then, I have

$$\varepsilon_t = \frac{-\delta_t}{\beta} + \frac{\gamma}{\beta}(\varepsilon_t^f - \varepsilon_t^{nf}). \tag{5}$$

Then, based on the fifth assumption the constant relative bias between food and non-food price indices $\varepsilon_t^f = \varepsilon_t^{nf}$, I have

$$\varepsilon_t = \ln(1 + E_t) = \frac{-\delta_t}{\beta}. \tag{6}$$

Finally, I can calculate the cumulative percentage point of the CPI bias at time t

$$BIAS_t^{linear} = 1 - \exp\left(\frac{-\delta_t}{\beta}\right). \tag{7}$$

Costa (2001) applies the quadratic functional form of the Engel curve. The quadratic form of the Engel curve can be written as follows:

$$\omega_{it} = \mu + \gamma(\pi_t^f - \pi_t^{nf}) + \beta_1(y_{it} - \pi_t) + \beta_2(y_{it} - \pi_t)^2 + X'\theta + \sum_{t=1}^{T}\delta_t D_t + v_{it}. \tag{8}$$

When the relative bias is constant, the cumulative bias is calculated as follows:

$$BIAS_t^{quadratic} = 1 - \exp\left(\frac{\beta_1 + \sqrt{\beta_1^2 + 4\beta_2\delta_t}}{2\beta_2}\right). \tag{9}$$

The annual bias was calculated as follows:

$$annual\ bias = ((1 + BIAS)^{\frac{1}{period}} - 1) * 100 \tag{10}$$

where *BIAS* is as defined in equation (7) or (9), and *period* is the sample period from the base year 1989.

3 Data

This section introduces the data set used in this study and explains the construction of the analysis sample. The cross-sectional data set used in this paper was from National Survey of Family Income and Expenditure, conducted by the Ministry of Internal Affairs and Communication. This survey has been conducted to collect comprehensive data on family income, expenditure, saving, debt, durable consumer goods, and assets. The survey started in 1959 and has been conducted every five years thereafter. It covered households with two or more people and single person households. The survey of households with two or more people was conducted for three months, from September to November. Single person households were surveyed in October and November. A household was requested to fill in the household consumption diary, household questionnaire, consumer durables questionnaire, and annual income and saving questionnaire.

This paper used 80% re-sampled and anonymized data sets for 1989, 1994, 1999, and 2004, focusing on the sample of households with two or more people. Each data set contained almost 45,000 households in each year. I restricted the sample to households headed by persons aged 20 to 59 years old because of the Retirement-

Savings Puzzle. As shown in Banks, Blundell, and Tanner (1998), household which a household head experienced a retirement would change the consumption behaivior. Also as discussed in Aguiar and Hurst (2007) and Abe and Shiotani (2012), elderly people would have enough time to seek lower price food. For these reasons, this paper focuses on estimating the CPI bias for those household with a household head aged between 20 and 59. I eliminated households that headed by persons falling under farmer and fisher categories because of the personal consumption matter. I excluded expenditures on housing because of the matter of the imputed rent. In addition, I excluded expenditures on alcohol and food eaten out because the Engel curves for these items would be different from other food items and would have an increasing or inverted U-shape Engel curve (Banks, Blundell, and Lewbel, 1997).

The food share, obtained by dividing food expenditure by living expenditure, was used as a dependent variable that took a value between zero and one. From the definition of the expenditure categories, this paper used living expenditure to represent household permanent income, whilst Hamilton (2001) has used family income and Costa (2001) has used total expenditure. The exclusion of alcohol and food eaten out produced negative food expenditure and living expenditure values.[*1] Thus, I restricted the sample to food shares between zero and one. To exclude observations

[*1] Some households reported zero for expenditure, food expenditure, or income even if they reported non-zero annual income, food consumption, rent, or saving. For unknown reasons, respondents did not expend during the sample period.

where food demand might be unusual, the sample was trimmed at the 3rd and the 97th percentile values of the cross-sectional distributions of food share and real living expenditure for each survey year. Without the trimming, the Engel curve predicted that the food share took a negative value at some expenditure levels.

The description of the independent variable was as follows. The independent variable included the real living expenditure, the relative price of food and non-food items, family characteristics, year dummy variables, and the region dummy variable. The real living expenditure was deflated by the official Japanese CPI. The demographic controls were as follows: the number of adult family members aged 18 to 59 years, the number of elderly people, that is, family members aged 60 years and more, the number of pre-school children, the number of elementary-school-going children, and the number of children going to junior high or high school.

Control variables also included dummy variables of being employed, which took one if a household head was employed, job type of a household head, and age of a household head. Industry dummy variables that indicated the industry of household head's job were included. The industry category included following: mining; construction; manufacturing; electricity, gas, heat supply, and water; transport and communication; wholesale and retail trade; finance and insurance; real estate; service; government; and others. The year dummy variables were for 1989, 1994, 1999, and 2004. The region dummy variable took one if a household lived in an urban area.

The construction of price indices was as follows. I used the official 2010-base price indices reported by the government: the Japanese CPI, the food price index, and the non-food price index. Since the definition of urban area in the analysis sample contained small cities around the main big cities and was different from that in the price indices sample, nationwide price indices were used for households living in rural areas; price indices of cities with a population of more than 50,000 were used for households living in urban areas. I constructed three price indices, excluding alcohol, food eaten out, and housing.[*2]

4 Empirical Analysis

4.1 Analysis of Summary Statistics

Table 1 shows summary statistics of the food share and the real living expenditure per month for each sample period. The food share decreased by 15 percent during the sample period, from 25.96 percent in 1989 to 22.06 percent in 2004. The substantial decrease of the food share implies the improvement of living standard and necessitates the increase of the real living expenditure. However, the real living expenditure decreased by about 6 percent during the sample period, from 242,950 yen in 1989 to 229,450 yen in 2004.[*3] The twisted relationship suggests the underestimation of CPI

[*2] The calculation method is explained in Appendix A.

[*3] The National Health and Nutrition Survey showed the trend in nutrient intake of energy between 1990 and 2004 dropped by 124 kilo calories (6 percent). This suggests a decrease in

deflated real living expenditure because of upward bias of the CPI.

Table 1 is inserted here.

Table 2 shows the summary statistics for household characteristics: job type of a household head, age of a household head, rate of home owners, and proportion of households living in the urban area. The table shows that more than 30 percent of the household heads were engaged in manufacturing and service industries. The share of manufacturing jobs decreased by 4 percentage points from 1989 to 2004, in contrast to a 3.5 percentage point increase in service jobs. In this data set, household heads aged 20 to 24 years constituted less than 1 percent of the total. Most household heads were more than 30 years old.

Table 2 is inserted here.

Table 3 shows the changes in the numbers of children.[*4] The table shows the proportion of childless households increased over time. Almost 30 percent of households in 1989, and 40 percent in 2004, had no children. Although the proportion of single child households was almost stable during the sample period, that of households with two children declined by about 10 percentage points.

the importance of food. However, if this results in higher expenditure on other goods, it might imply that living standards have increased.

[*4] The definition of children included pre-school children and those who go to elementary school and junior high or high school.

Table 3 is inserted here.

Figure 1 shows Engel curves, specified as quadratic functions of the real living expenditure, for each year. If the Japanese CPI accurately corresponded to the changes in the true cost of living, the Engel curve would be stable across different periods. However, the figure shows the curves shifted downward in different periods. This implies the CPI was potentially upwardly biased.

Figure 1 is inserted here.

4.2 Analysis of Overall Bias

To estimate the Engel curve, I regressed the food share on the log of real living expenditure, year dummy variables, log of relative price, and dummy variables of family characteristics- being a home owner, living in urban areas, being employed, job type of household head, age of household head, and number of adults, elderly people, pre-school children, children in elementary school, and children in junior high or high school. To estimate the quadratic Engel curve, I included the quadratic term of the log of real living expenditure as an independent variable. Table 4 shows regression results for the linear and quadratic Engel curves. With the linear model, the coefficient of the log of real living expenditure was statistically significant at the 1 percent significance level. The result showed the food share declined as the real living expenditure increased. A 1 percent increase in real living expenditure resulted in a 0.17 percent decrease in the food share. With the quadratic specification, the coefficient of the quadratic term was statistically significant at the 1 percent significance level. This result showed the quadratic Engel curve nested the linear specification and was more appropriate.[*5] This implies that the linear Engel curve might overestimate the bias. The income elasticities of the two specifications were about 0.440, evaluated at the sample means of the food share and the log of real living expenditure.

The crucial variable of this study was the year dummy variable. Again, if the Japanese CPI was not biased and corresponded to the true cost of living, the Engel curve would be stable across the sample period. If so, the parameters of the year dummy variables would not differ from zero. However, regression results from both the linear and the quadratic models showed that the dummy variables, dummy-94, dummy-99, and dummy-04, were statistically significant and negative. This finding corroborates an upward bias in the official Japanese CPI.

Table 4 is inserted here.

Estimates of the annual bias were reported in Table 5. The overall upward bias

[*5] Unayama (2008) estimated the quadratic almost ideal demand system for food and other goods using the Japanese Family Income and Expenditure Survey from 1982 to 2000. The paper did not obtain evidence of a significant quadratic form for food. The paper focused on households with two married adults but no child, living in their own houses. The difference of the results might be due to sample construction differences.

between 1989 and 2004 based on the linear Engel curve was 1.7 percentage points per year. The annual bias increased every five years: 1.6 percentage points during 1989-1994, 2.0 percentage points during 1994-1999, and 2.5 percentage points during 1999-2004. Based on the quadratic Engel curve, the overall bias was 0.5 percentage points per year from 1989 to 2004-a lower size of the bias compared to the linear model. In addition, the bias increased in later periods, although it was much lower compared to the linear model: 0.4 percentage points for 1989-1994, 0.5 percentage points for 1994-1999, and 0.7 percentage points for 1999-2004.

The 1.7 percentage point annual bias based on the linear Engel curve was larger than the 0.9 percentage points estimated by Shiratsuka (1999) but was close to the 1.8 percentage points reported by Broda and Weinstein (2007). On the other hand, the 0.5 percentage point CPI bias based on the quadratic Engel curve was much lower compared to the previous two studies. This bias estimate was similar to the lower bound reported by Ariga and Matsui (2003) and the estimate of Watanabe and Watanabe (2014). As the regression result showed that the quadratic term was statistically significant, the estimated annual bias from the linear model could be overestimated. The estimate of Broda and Weinstein (2007) based on the US CPI bias was much higher than the results of Shiratsuka (1999), Ariga and Matsui (2003), and this paper. As in the comment on the paper by Shigenori Shiratuka, several differences such as the price surveys between Japan and the US might result in the large difference. Shiratsuka (1999) has obtained 0.9 percentage point CPI bias per year. His key cause of the CPI bias was the quality adjustment. Adjusting the quality of items resulted in 0.7 percentage points of the CPI bias per year. To obtain the estimated result, the paper applied the hedonic approach for personal computers, camcorders, automobiles, and apparel. The quality improvement for these items might presumably be much higher than other items, and so the bias estimates might be overestimated.

Table 5 is inserted here.

The estimation method employed in this paper contains any bias drives such as the calculation method and the price survey. If the CPI is upwardly biased because of any reasons, the real living expenditure (or real income) would be under estimated. As a result, this method estimates the CPI bias as the downward shift of the Engel curve. Thus this approach would include the points argued in previous studies and might estimate the CPI bias more comprehensively.

Estimates of annual bias indicated that the official Japanese CPI and the inflation rate are likely to be upwardly biased. Figure 2 shows the official Japanese CPI and the revised CPI. The revised CPI was obtained from Equation 3 and the estimated annual bias. The figure shows the official CPI was placed above the revised CPI. In 2004, the cumulative bias amounted to almost 7 percentage points. The revised CPI implied lower inflation and more serious deflation. It showed moderate inflation from 1992 to 1998 and severe deflation, more acute than suggested by the official CPI, thereafter. The inflation rate based on the official CPI was 0.65 percent from 1989 to 2004. This contrasted with the 0.14 percent inflation rate for the same period based on the revised CPI. Moreover, bias correction suggested severe deflation after 1998. After 1998, the deflation rate was 0.95 percent based on the corrected CPI whereas it was 0.36 percent according to the official CPI.

Figure 2 is inserted here.

4.3 Analysis by Demographics

The above discussion showed that the CPI was upwardly biased and presented the degree of bias across the sample period. This subsection analyzed the cause of bias by examining the difference in bias across demographic groups that presumably represented the difference in the opportunity costs of shopping time.

The Ministry of Internal Affairs and Communications, in principle, collects normal retail prices or normal service charges, and so outlet or discount sale prices are excluded in its Retail Price Survey. In addition, the survey specifies the name of brand for specific goods. This implies that lower-level substitution is implicitly assumed away (Broda and Weinstein, 2007). Aguiar and Hurst (2007) and Abe and Shiotani

(2012) have analyzed the relationship between prices that households face and the frequency of shopping. Both studies suggested that households, such as the elderly, who have enough time to find lower prices for commodities face a lower inflation. Thus, this paper estimates the CPI bias between households with a working or a non-working spouse. Compared to households with a working spouse, households with a non-working spouse might have more time to seek out low-priced or discount goods and might substitute them for expensive goods. According to the Survey on Time Use and Leisure Activities in 2006, the weekly average time spent on shopping by a household with both working head and spouse is 44 minutes. On the other hand, the weekly average time spent on shopping by a household with a working head and a non-working spouse is 66 minutes. This would mean that a household with a working spouse cannot spend on shopping as much as a household with a non-working spouse. Therefore, this would result in a larger upward bias for household with a non-working spouse.

Table 6 shows bias estimates for households with a working or non-working spouse. Households with a non-working spouse faced an overall bias of 0.55 percentage points per year from 1989 to 2004, compared to 0.45 percentage points for those with a working spouse. This means that households with a working spouse presumably had less shopping time, resulting in lower upward bias. A household with a non-working spouse could spend more time on looking for cheaper goods and face lower inflation, resulting in a larger CPI bias for the household. The result implies the similar conclusion with previous studies. If a household can spend more time on shopping or go shopping more frequently, a household faces lower prices and an lower inflation. Therefore, excluding outlet and discounted sale prices from the survey and not allowing for the lower-level substitution consist a part of the reasons for the upward bias of the official CPI.

Table 6 is inserted here.

4.4 Robustness Check of Annual Bias Estimates

The above analysis assumes that the relative bias between food and non-food is consistent. This section estimates the bias when this assumption is violated. This means that equation (9) is changed as follows:

$$BIAS_t^{quadratic} = 1 - \exp\left(\frac{\beta_1 + \frac{\gamma(1-k)}{1-\alpha(1-k)} + \sqrt{\left(\beta_1 + \frac{\gamma(1-k)}{1-\alpha(1-k)}\right)^2 + 4\beta_2\delta_t}}{2\beta_2}\right),$$

where k is the size of the relative bias between food and non-food ($\varepsilon_t^f/\varepsilon_t^{nf}$) and α is the weight between food and non-food prices. If the relative bias is not consistent and if the bias of the non-food prices is larger than the bias of the food prices, the true bias estimates would be larger than the estimates of the previous sections.

Table 7 calibrates the annual bias under several values of k, the relative bias between food and non-food prices, because we do not have the exact number of the degree of the relative bias. If the relative bias was not existed, which means k was 100 percent, the estimate of the CPI bias was 0.52 percentage points. If the relative bias was 99 percent, the estimate of the CPI bias was 0.52 percentage points. These estimates are very close to the size of the bias under the assumption. If there was 90 percent of the relative bias, the size of the CPI bias was 0.56 percentage points. If there was 80 percent of the relative bias, the estimate was 0.61 percentage points. These estimates mean that if the constant relative bias assumption is violated, the size of the CPI bias were larger than the bias estimate under the asumption. However, if the relative bias was within the reasonable range, the results in the previous sections are feasible estimates.

Table 7 is inserted here.

5 Summary and Conclusion

In this paper, I estimated the upward bias of the Japanese CPI based on Engel's law, using National Survey of Family Income and Expenditure data. The estimated overall bias was 0.53 percentage points per year between 1989 and 2004. This estimate was close to the lower bound reported by Ariga and Matsui (2003) and the estimate by Watanabe and Watanabe (2014) but lower than the result reported by Shiratsuka

(1999) and Broda and Weinstein (2007). Previous studies have analyzed the CPI bias in terms of its causes; on the other hand, this paper applied a different approach of estimating the Engel curve. This approach estimated the over all CPI bias, and so all causes of the CPI bias would be included. The method relied on several assumptions, especially on a relative bias between food and non-food items. If the relative bias of non-food items is larger than that of food items, the estimates of this paper were underestimated. However, there were not so much differences in bias estimates when the relative bias between food and non-food items is within the plausible range.

The 0.53 percentage point upward bias implies that the official CPI overstates the true cost of living. Consequently, the bias-corrected annual inflation rate from 1989 to 2004 was estimated to be 0.14 percent, whereas the official inflation rate was 0.65 percent during the same period. The deflation during the period was severe, especially after 1998, when the rate of deflation was 0.95 percent compared to 0.36 percent based on official statistics. These results suggest that government policy payments index-linked to the Japanese CPI, such as pension benefit, might overcompensate the recipients.

A demographic analysis showed that, compared to households with a working spouse, those with a non-working spouse faced a larger bias and lower inflation since they presumably have enough time to visit different stores and can seek out low-cost goods. The result offers suggesting evidence that the ignorance of sales, discount or brand substitution is an important source of the CPI bias.

References

[1] Abe, Naohito and Kyosuke Shiotani, "Who faces higher prices? -An empirical analysis based on Japanese homescan data-." Research Center for Price Dynamics Working Paper Series, No. 3, 2012.

[2] Aguiar, Mark and Erik Hurst, "Life-Cycle Prices and Production." *American Economic Review* 97(5), pp. 1533-1559, 2007.

[3] Ariga, Kenn and Kenji Matsui, "Mismeasurement of the CPI." *Structural Impediments to Growth in Japan* (Eds) M. Blomstom, J. Corbett, F. Hayashi, and A. Kashyap, University of Chicago Press, pp. 89-154, 2003.

[4] Banks, James, Richard Blundell, and Arthur Lewbel, "Quadratic Engel Curves and Consumer Demand." *Review of Economics and Statistics*, 79(4), pp. 527-539, 1997.

[5] Banks, James, Richard Blundell, and Sarah Tanner, "Is There a Retirement-Savings Puzzle?" *American Economic Review*, 88(4), pp. 769-788, 1998.

[6] Barrett, Garry F. and Matthew Brzozowski, "Using Engel Curves to Estimate the Bias in the Australian CPI." *Economic Record* 86(272), pp. 1-14, 2010.

[7] Beatty, Timothy K. M. and Erling Roed Larsen, "Using Engel Curves to Estimate Bias in the Canadian CPI as a Cost of Living Index." *Canadian Journal of Economics* 28(2), pp. 482-499, 2005.

[8] Broda, Christian M. and David E. Weinstein, "Defining Price Stability in Japan: A View from America." Monetary and Economic Studies 25(S1), pp. 169-206. 2007.

[9] Chung, Chul, John Gibson, and Bonggeun Kim, "CPI Mismeasurements and Their Impacts on Economic Management in Korea." *Asian Economic Papers* 9(1), pp. 1-15, 2010.

[10] Costa, Dora L., "Estimating Real Income in the United States from 1888 to 1994: Correcting CPI Bias Using Engel Curves." *Journal of Political Economy* 109(6), pp. 1288-1310, 2001.

[11] Deaton, Angus S. and John Muellbauer, "An Almost Ideal Demand System." *American Economic Review* 70(3), pp. 312-326, 1980.

[12] Hamilton, Bruce W., "Using Engel's Law to Estimate CPI Bias." *American Economic Review* 91(3), pp. 619-630, 2001.

[13] Shiratsuka, Shigenori, "Measurement Errors in the Japanese Consumer Price Index." *Monetary and Economic Studies* 17(3), pp. 69-102, 1999.

[14] Unayama, Takashi, "Measuring the substitution bias in Japan: the demand system approach and a superlative index." *Applied Economics* 40(14), pp. 1795-1806, 2008.

[15] Watanabe, Kota and Tsutomu Watanabe, "Estimating Daily Inflation Using Scanner Data: A Progress Report." CARF Working Paper Series, CARF-F-342,

2014.

AppendixA Price Data

Since this analysis eliminates the effect of alcohol, food eaten out, and cost of housing, the data of the Japanese CPI, the food price, and the non-food price are calculated as follows:

For CPI,

$$\text{CPI} = \frac{p_T \times w_T - p_A \times w_A - p_O \times w_O - p_H \times w_H}{w_T - w_A - w_O - w_H}.$$

For food price,

$$\text{food-price} = \frac{p_F \times w_F - p_A \times w_A - p_O \times w_O}{w_F - w_A - w_O}.$$

For non-food price,

$$\text{non-food-price} = \frac{p_T \times w_T - p_F \times w_F - p_H \times w_H}{w_T - w_F - w_H}.$$

The variables p_T, p_F, p_A, p_O, and p_H are the price indices of total, food, alcohol, food eaten out, and cost of housing. The variables w_T, w_F, w_A, w_O, and w_H represent the weights of these variables.

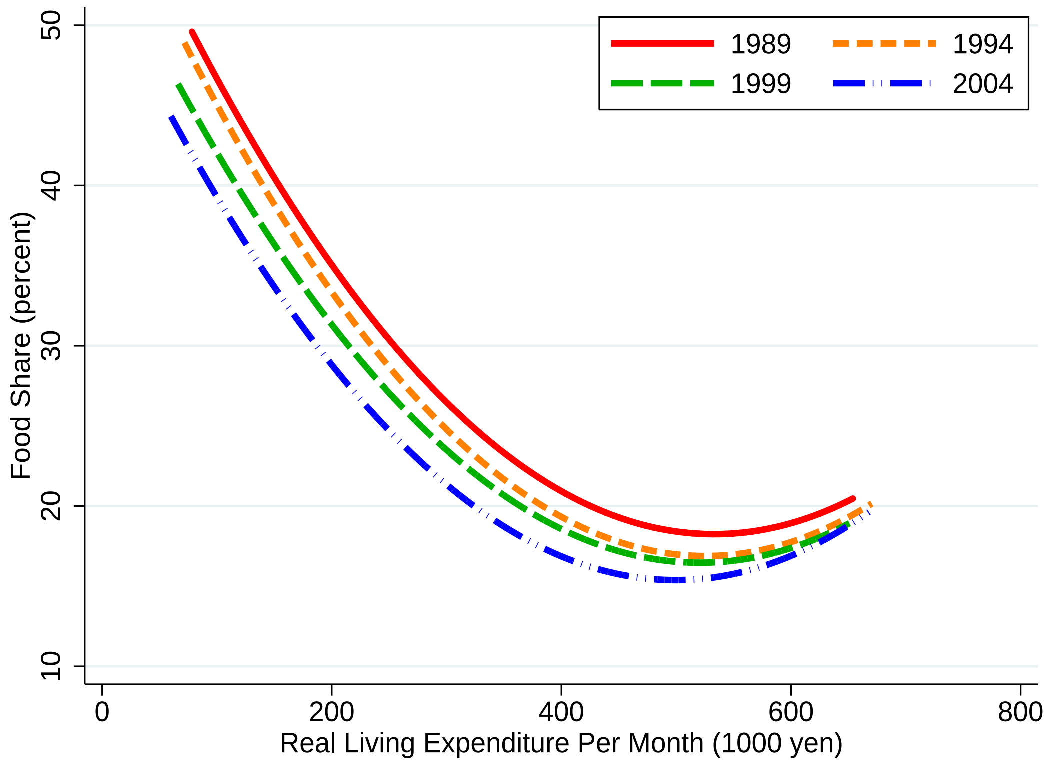

Figure1 Engel curves

Source: National Survey of Family Income and Expenditure.

Note: The lines are simple Engel curves for each year.

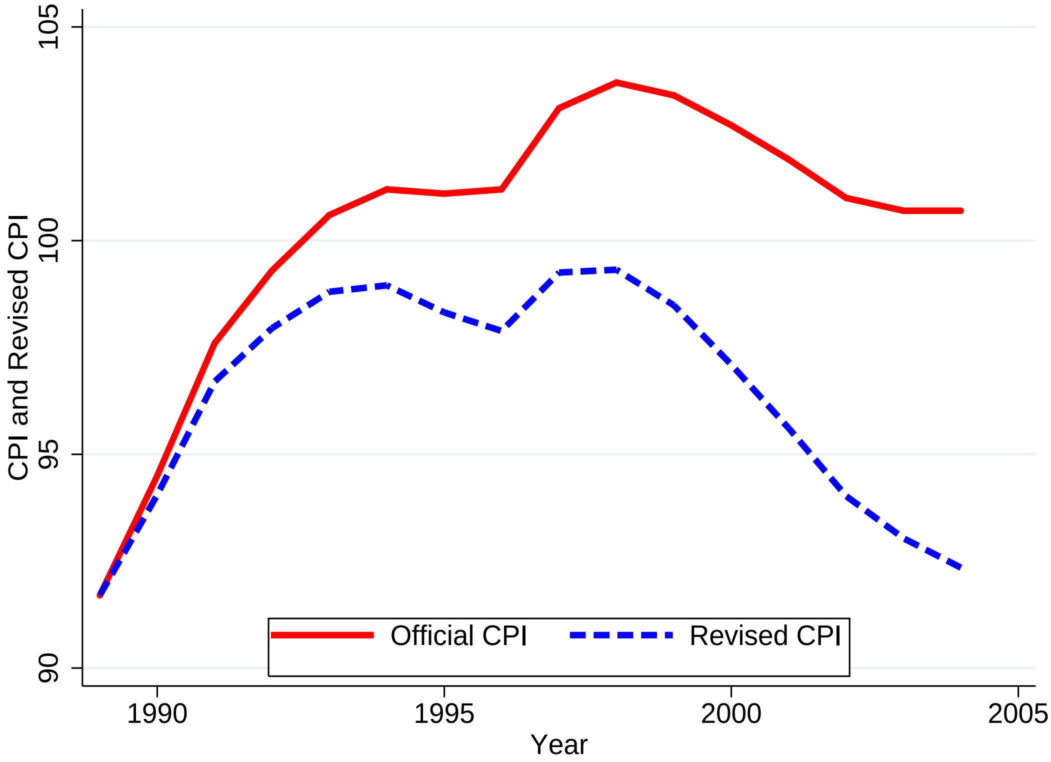

Figure2   CPI and revised CPI

Source: Official CPI statistics. Revised CPI is based on author's calculation.

Table1 Summary Statistics 1

Year	1989	1994	1999	2004
Observations	29,594	28,745	26,723	23,820
Food share (%)	25.96	23.34	22.75	22.06
	(10.85)	(10.45)	(10.82)	(10.86)
Real living expenditure	242.95	245.74	235.68	229.45
per month (1000 yen)	(109.03)	(118.16)	(117.03)	(119.71)

Source: National Survey of Family Income and Expenditure.

Notes: The food share is obtained by dividing the food expenditure by the real living expenditure.

The real living expenditure per month is deflated by the official 2010-base CPI.

Standard deviations are in parentheses.

Table2 Summary Statistics 2

	1989	1994	1999	2004
Variable	Percent	Percent	Percent	Percent
Household head employed	98.00	98.18	97.47	97.28
Household head's job				
Mining	0.32	0.22	0.16	0.18
Construction	12.53	12.43	13.05	12.31
Manufacturing	25.54	24.45	23.40	21.57
Electricity, gas, heat supply and water	1.29	1.41	1.29	1.48
Transport and IT	9.08	9.12	8.58	9.39
Wholesale and retail trade	15.58	15.34	14.72	14.41
Finance and insurance	3.56	3.66	3.39	3.23
Real estate	0.78	0.81	0.93	0.99
Service	18.46	20.14	21.43	22.04
Government	10.78	10.55	10.59	11.82
Others	2.06	1.85	2.45	2.58
Household head's age				
20-24	0.75	0.79	0.84	0.63
25-29	5.06	4.69	5.19	4.32
30-34	11.46	11.15	10.80	10.38
35-39	18.47	15.19	14.03	14.13
40-44	20.64	19.43	16.68	16.05
45-49	17.73	19.31	18.28	17.20
50-54	13.74	16.57	18.08	18.30
55-59	12.16	12.88	15.98	18.85
Home owner	70.44	65.67	67.82	70.31
Living in urban area	40.60	42.37	42.70	40.92

Source: National Survey of Family Income and Expenditure.

Table3 Summary Statistics 3

	1989	1994	1999	2004
	Percent	Percent	Percent	Percent
Total children				
0	27.88	33.71	37.31	38.82
1	21.93	22.12	21.78	22.88
2	36.05	31.37	29.29	27.90
3	13.09	11.95	10.59	9.50
4	0.98	0.80	0.98	0.76
5	0.07	0.07	0.06	0.14

Source: National Survey of Family Income and Expenditure.

Table4 Results of Selected Regression Coefficients

	Linear model coefficient	Quadratic model coefficient
Constant	2.49***	5.79***
	(0.01)	(0.17)
Log (real living expenditure)	-0.17***	-0.71***
	(0.00)	(0.03)
(Log (real living expenditure))^2		0.02***
		(0.00)
Dummy-94	-0.04***	-0.04***
	(0.00)	(0.00)
Dummy-99	-0.05***	-0.05***
	(0.00)	(0.00)
Dummy-04	-0.07***	-0.07***
	(0.00)	(0.00)
sample size	108,882	108,882
Adjusted R^2	0.40	0.40

Source: National Survey of Family Income and Expenditure.

Note: The results are obtained by regressing the food share on the log of the real living expenditure, the log of the relative price, and dummy variables of home owner, living in the urban area, employed, job types, ages, adults, elderly people, pre-school children, children in elementary school, children in junior high or high school, and year dummy variables.

The robust standard errors are in parentheses.

Nonlinear least-squares estimation is used.

$***$: Significant at the 1 percent level

Table5 Annual Bias (percentage point)

	1989-2004	1989-1994	1994-1999	1999-2004
Linear model	1.73	1.69	2.04	2.55
	(0.02)	(0.09)	(0.08)	(0.09)
Quadratic model	0.53	0.45	0.53	0.69
	(0.02)	(0.03)	(0.03)	(0.03)

Source: National Survey of Family Income and Expenditure.

Note: Standard errors from 500 bootstrap replications are in parentheses.

Table6 Annual Bias by Demographics (percentage point)

	1989-2004	1989-1994	1994-1999	1999-2004
Household head is working and	0.55	0.53	0.50	0.71
spouse is not working	(0.03)	(0.04)	(0.04)	(0.05)
Household head is working and	0.45	0.33	0.48	0.59
spouse is working	(0.02)	(0.03)	(0.04)	(0.04)

Source: National Survey of Family Income and Expenditure.

Note: The results are obtained by regressing the food share on the log of the real living expenditure, the quadratic term of the log of the real living expenditure, the log of the relative price, and dummy variables of home owner, living in the urban area, employed, job types, ages, adults, elderly people, and year dummy variables.

Standard errors from 500 bootstrap replications are in parentheses.

Table7 Robustness Check of Annual Bias Estimates (percentage point)

Relative bias ($\varepsilon_t^f/\varepsilon_t^{nf}$)	1.00	0.99	0.95	0.90	0.85	0.80
CPI bias estimates	0.52	0.52	0.54	0.56	0.58	0.61
with a quadratic model	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)

Source: National Survey of Family Income and Expenditure.

Note: Standard errors from 500 bootstrap replications are in parentheses.

The term α is 0.25, and this is taken from the weight of the 2010-base official CPI.