

UPPER-LEVEL SUBSTITUTION AND NEW-GOODS BIAS IN THE KOREAN CONSUMER PRICE INDEX

Daehoon Nahm*

dnahm@efs.mq.edu.au

ABSTRACT

The present paper estimates upper-level substitution and new-goods bias in the Korean Consumer Price Index (CPI). It has been estimated that the upper-level substitution bias in the CPI alone increased the inflation rate by 0.51 percentage points per year over the thirteen-year period between 1990 and 2002. The new-goods bias further increased the inflation rate by 0.17 and 0.13 percentage points per year between 1990–1995 and 1995–2000 respectively. The new Chained Laspeyres index series that is based on annually-updated weights has been found to correct only less than half of the upper-level substitution bias.

JEL Classification: C43

* The BK21 Education and Research Group For the Korean Economy under an Open Economic System, Department of Economics Korea University. The present research was supported by *The BK21 Education and Research Group For the Korean Economy Under an Open Economic System*. I thank David Throsby for his comments.

List of Abbreviations

BLS	Bureau of Labor Statistics
CES	Constant Elasticity of Substitution
CL	Chained Laspeyres
COL	Cost of Living
COLI	Cost of Living Index
CPI	Consumer Price Index
FI	Fisher Ideal
KCPI	Korean Consumer Price Index
KNSO	Korea National Statistical Office
TT	Törnqvist-Theil

1. Introduction

A traditional fixed-weight Consumer Price Index (CPI) is designed to measure changes in the cost of a fixed basket of goods and services that a typical household purchases over different time periods or in separate places with different sets of prices. It has long been known amongst economists that an inflation rate that is based on a traditional CPI might be overstating the rate of change in the true Cost of Living (COL). The main reasons for the overstatement are: i) It does not reflect consumers' behavior in substituting products that have become relatively cheap for those that are relatively dear when their relative prices change (*lower-level substitution bias* arising from the steps leading to elementary aggregation of directly sampled price quotes, and *upper-level substitution bias* stemming from substitutions across elementary aggregates); ii) It does not adequately take account of consumers' behavior to shop around, that is, the behavior to shop at outlets that sell products at cheaper prices with acceptable services (*outlet-substitution bias*); iii) It does not properly handle the effects of changes in the quality of products on the COL (*quality-change bias*); iv) There is usually a long lag before a newly-introduced product is included in the basket of products used for the computation of CPI, and it does not adequately incorporate the effect of a change in the set of products from which consumers can choose (*new-goods bias*).¹ In addition to these sources of bias, Lebow and Rudd (2003) recently revealed another main source of bias, related to the derivation of the weights from a household expenditure survey that is suspected to be incomplete (*weighting bias*).

¹ Neat graphical illustrations of the effect of a change in the variety of choice set can be found in Diewert (1996) and Hausman (2003).

The Final Report to the U.S. Senate Finance Committee from the Advisory Commission to Study the CPI (Boskin et al. 1996, Boskin Report hereafter) has rigorously estimated the bias stemming from the first four sources listed above in the then U.S. CPI, and concluded that the CPI-based inflation rate overstates changes in the COL by 1.1 percentage points per year on average; see Table 1 for a breakdown. The Report further highlighted how significant the effects of such bias would be if it were left uncorrected. The Report estimated the would-be increase in the U.S. annual federal deficit in just over a decade as US\$202 billion in 2008, which would amount to the fourth largest U.S. federal 'program', after social security, health care and defense.

A myriad of further discussions and researches on the issue have followed the Boskin Report. Although opinions are divided over whether or not a national statistical agency should target measuring COL and revise the procedure of CPI construction towards that direction, there appears to be little disagreement that the traditional CPI (and hence inflation rates based on it) is a flawed measure; see, for instance, Triplett (2001), Turvey (1999), Solow (1997), and Tobin (1997). While some statistical agencies, including those of the United States, The Netherlands, and Sweden, have adopted a cost-of-living index (COLI) as their target index and have been gradually modifying CPI construction procedures to make their indices closer to the COLI, many other countries' statistical agencies, including the Korea National Statistical Office (KNSO) and Japan's Statistics Bureau, have rejected a COL index as the target index and still publish CPIs based on the standard Laspeyres index formula. Most of the countries that have not adopted the COL approach appear to be doing so mainly because of practical difficulties and political reasons rather than because of any strong objection to the approach on theoretical grounds.

For this reason, information about the gap between a country's published CPI and a COLI would be important irrespective of whether or not the statistical agency of the country has adopted the COL approach. For the countries that have adopted the COL approach, such bias estimates provide useful information about progress toward correctly measuring the COL. For the countries that have yet to adopt the approach, the information is useful in analyzing the significance of the deviation of their CPI from the cost of living and in estimating the associated social costs arising from the potentially-biased information. In light of such lively interest and importance of the issue, it is surprising to find that very little attention has been paid to the potential bias in the Korean CPI (KCPI) as a measure of COL. The exceptions appear to be Lee and Chung (1997) who constructed a COLI series both including and excluding owner-occupied housing costs, and Lee (2001) who estimated quality-adjusted price indices for personal computers utilizing the hedonic regression approach.

Table 1: Bias Estimates for U.S., Japanese and Korean CPI's*

Country Author(s)	U.S.			Japanese	Korean
	Boskin Report (1996)	Gordon (1999)	Lebow and Rudd (2003)	Shiratsuka (1999)	Lee and Chung (1997)
Upper-level Substitution	0.15	0.05	0.30	0.00 (~0.25)	0.00 (-0.12) ⁺
Lower-level Substitution	0.25	0.05 [#]	0.05	0.10	-
New Goods/ Quality Change	0.60	0.30~0.40	0.37	0.70	-
Outlet Substitution	0.10	0.10	0.05	0.10	-
Weighting	-	-	0.10	-	-
Total	1.10	0.50~0.60	0.87	0.90	-

*: Percentage points per year

#: Based on the statement on page 60 of Gordon (1999).

+: The value in the parentheses is derived from the information in Table 1 of Lee and Chung (1997).

The U.S. CPI would arguably be the most closely scrutinized and extensively studied price index with regard to the bias in it as a measure of COL. As a result, the U.S. CPI construction procedure has gone through several improvements, and much of the bias estimated in the Boskin Report has been corrected. The first three columns of Table 1 show bias estimates for the U.S. CPI at three different points in time. Gordon (1999) revised down each of upper-level and lower-level substitution biases to 0.05 following improvements to the procedure involving more frequent rotation of sampled items, more rapid changes in upper-level weights, and the Bureau of Labor Statistics' (BLS) adoption in January 1999 of geometric-mean price index for individual items. Gordon believed that the improvements up to the point in time of his writing might have halved the total bias of 1.1 estimated by the Boskin Report to 0.50 ~ 0.60 percentage points.² Lebow and Rudd (2003), however, re-estimated the biases and reported a much higher upper-level substitution bias of 0.3 and an additional bias of 0.1 from a new source (weighting bias); see the third column of Table 1. They found that the much higher substitution bias was due to its dramatic increase in the late 1990s. The estimate of upper-level substitution bias in the KCPI by the present paper (reported in Section 4) is much closer to Lebow and Rudd's estimate than to the one by the Boskin Report, but for different reasons. Shiratsuka's (1999) estimates of the biases in the Japanese CPI and Lee and Chung's (1997) estimate of the upper-level substitution bias in the KCPI are shown in the last two columns of the table. It is noteworthy that the bias estimates, especially the estimates for upper-level substitution bias, are significantly different

² His statement on page 62, with regard to the lower-level substitution bias, seems to be inconsistent with that on page 60. The estimate for lower-level substitution bias reported here is based on the information on page 60.

across countries. The very low and negative estimates of the upper-level substitution bias in the Japanese and Korean CPIs respectively are questionable for reasons that will be explained later in Section 4.

The main objective of the present paper is to fill a part of the vacuum in the literature about the bias in the KCPI by estimating and analyzing the bias caused by upper-level substitution and that caused by introduction of new products and disappearance of old products. Although Lee and Chung (1997) have already estimated upper-level substitution bias in the KCPI, the negative estimate of upper-level substitution bias (although reported as negligible in their paper) is not congruent with the general belief that the Laspeyres-index-based CPI is the upper bound of the true COL. The present paper adopts an approach that is explicit in linking, and uses a more recent data set, including different (and presumably superior) estimates of the weights. The technique used to analyze new-goods bias in the present paper was introduced by Nahm (2004), and this is the first attempt to estimate new-goods bias in the KCPI.

The remainder of the paper is organized as follows. Some basic concepts and index number formulas are formally defined in the next section. Section 3 overviews the current procedure to construct the KCPI. Section 4 explains the methodology of the empirical analysis and the data, and then estimates the biases. The final section concludes the paper.

2. Fixed-basket CPI and COLI

To help the expositions in the present and the following sections, some

notations are defined first. Let $t = 0$ be the base period and $t = 1$ be the comparison period. The number of goods and services (product items) available for consumption in either period or both periods is denoted n . Let I^t be the set of variable indices for the product items actually available in period t . Define more sets as $I = (I^0 \cap I^1) \neq \emptyset$, $J = (I^0 \cup I^1)$, $I^{\text{new}} = \{i: i \in I^1, \text{ but } i \notin I^0\}$, and $I^{\text{old}} = \{i: i \in I^0, \text{ but } i \notin I^1\}$. So, there are n elements in J . The product items in I are available in both periods, while the product items in I^{new} are newly introduced to the market in period 1 (and hence unavailable in period 0), and those in I^{old} have disappeared in period 1 (although available in period 0). Let $p^t \gg 0_n$ be an $(n \times 1)$ vector containing strictly positive prices in period t for all product items in J . Note that some items in J are not available in period t and their prices are unobservable. In such cases, their “reservation prices” are assumed to be included in p^t .³ In analogy, $q^t > 0_n$ ⁴ is an $(n \times 1)$ vector of quantities consumed in period t . The elements in q^t for the product items unavailable in period t are zeros. In each period, consumers’ are assumed to find optimal quantities from the choice set J , including those items not available in that period. The reservation prices for those items are assumed to be just high enough to make demand for them equal to zero.

The cost of living is usually measured by the Konüs (1924) index:

$$(1) \quad P_K(p^0, p^1, u, J) = E(p^1, u, J) / E(p^0, u, J)$$

where u is an arbitrary reference level of utility, and $E(p, u, J)$ is an expenditure function

³ The concept of reservation price draws on Hicks (1940).

⁴ The general convention for vector inequalities is followed. That is, $q \gg 0_n$ means that each element q is strictly positive; $q \geq 0$ means that each q is nonnegative; and $q > 0_n$ means that $q \geq 0_n$ but $q \neq 0_n$, that is, at least one element is strictly positive.

that represents the minimum cost to achieve u at prices p when the choice set is given by J . Thus, a COLI measures how much it will cost, in comparison with the base period, to achieve the same level of utility (or standard of living) as in the base period in the face of a change in prices over the two periods. What is implicit in the above definition is that consumers are behaving rationally and efficiently in the sense that they try to minimize the cost of achieving a given level of utility by substituting goods and services they consume, and they do it without spending any more than necessary just to achieve the reference utility.

In contrast, the traditional fixed-basket consumer price index is a Laspeyres index defined by

$$(2) \quad P_L(p^0, p^1, q^0) = p^1 q^0 / p^0 q^0 = w^0 p^*$$

where w^0 is an $(n \times 1)$ vector of weights with a typical element, $p_i^0 q_i^0 / p^0 \cdot q^0$, p_i^0 and q_i^0 are the i -th element of p^0 and q^0 , respectively, and p^* is a vector of price ratios, p_i^1 / p_i^0 . So, the traditional CPI measures the cost, in comparison with the cost in the base period, to purchase the same basket of goods and services as in the base period when there is a change in prices. For this reason, Schultze (2003, p3) calls it a “cost-of-goods index” to contrast it with the term “cost-of-living index”. If the CPI, defined by (2), is used to estimate the COL, defined by (1), it is equivalent to estimating

$$(3) \quad P_K^L = E^L(p^1, u^0, I) / E^L(p^0, u^0, I) = E^L(p^1, F[q^0], I) / E^L(p^0, F[q^0], I) \\ = E^L(p^1, q^0, I) / E^L(p^0, q^0, I) = e^L(p^1, I) / e^L(p^0, I)$$

where $F(q) = \min_i \{q_i/\alpha_i: \alpha_i\text{'s are constants, } i = 1,2,\dots,n\}$, a Leotief type utility function, and $e^L(?) = \sum \alpha_i p_i$ is its dual unit expenditure function. It is well-known that (1) cannot be greater than (3) if $J = I$, the quantities are optimally chosen, and the reference utility level is u^0 ; see, for instance, Diewert (1987). Further, two important restrictions implied by (3) are noteworthy. One is that no substitution between product items in the choice set is allowed, and the other is that the sets of product items available in both periods are identical. The latter is the case because unless $I^0 = I^1 = I$, the index formula defined by (2) is not workable.

If the prices and quantities are measured without a bias, the upper-level substitution and new-goods bias in CPI-based inflation rate will be given by

$$(4) \quad \text{Bias (upper-level substitution \& new-goods)} = \dot{P}_L - \dot{P}_K$$

where \dot{P} denotes the rate of change in P . Note that if the preference ordering has a separability structure such that a valid price aggregator exists for each of the groups of products in I^0, I^1 , and I , the true COLI, (1), can be decomposed into three parts as:

$$(5) \quad P_K(p^0, p^1, u, J) = \frac{E(p^1, u, I^1)}{E(p^0, u, I^0)} = \frac{E(p^1, u, I^1)}{E(p^1, u, I)} \cdot \frac{E(p^1, u, I)}{E(p^0, u, I)} \cdot \frac{E(p^0, u, I)}{E(p^0, u, I^0)}$$

The middle part in the last term measures the effect of price changes for the products that are available in both periods, while the first and the last parts represent the effects of new goods and disappeared goods on the cost of living respectively. So, one will be

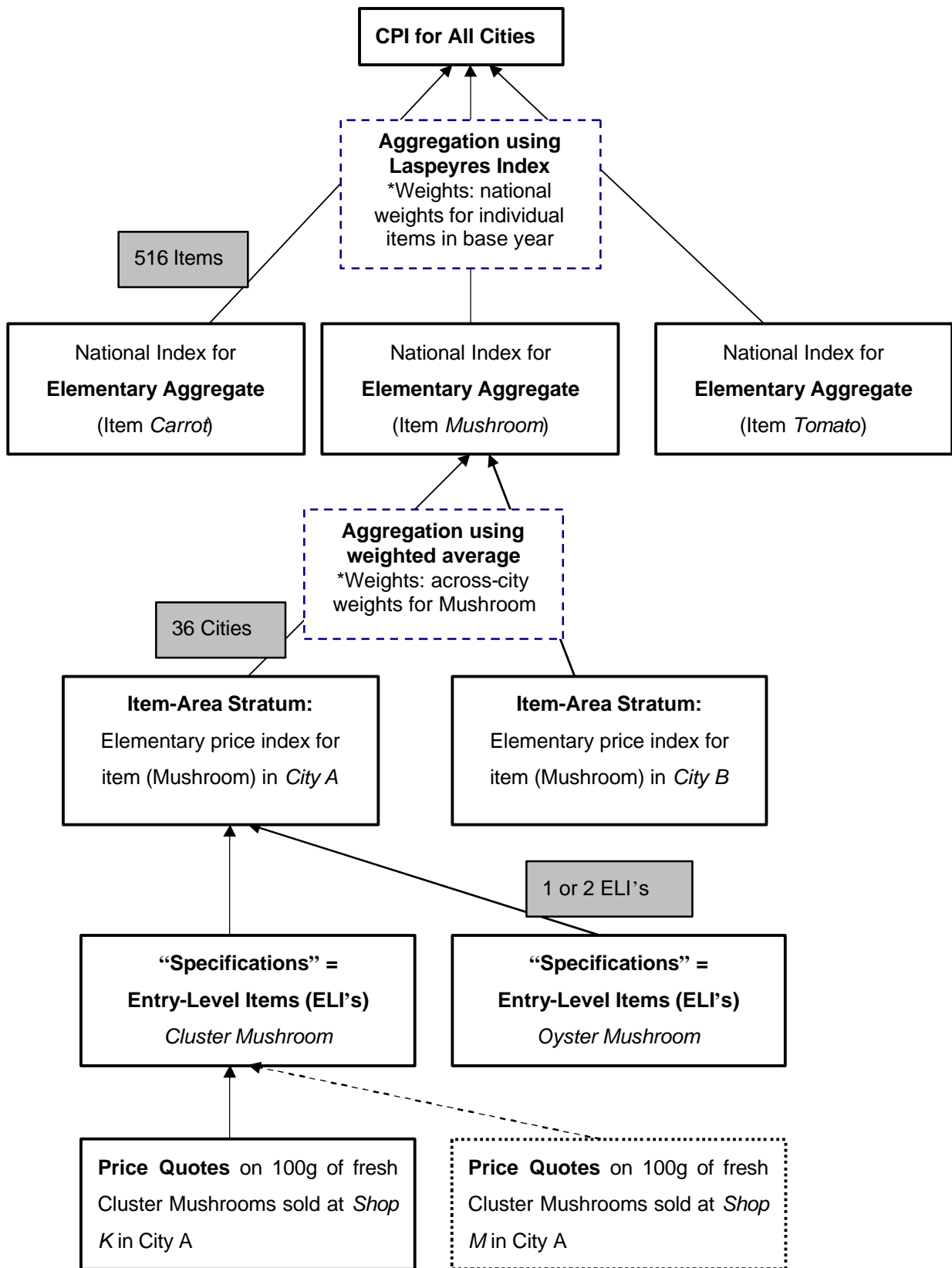
able to separately identify the substitution bias, new-goods bias, and disappeared-goods bias by decomposing the change rate in P_K into the three corresponding parts.

3. How the Korean CPI Is Constructed: An Overview

The Korea National Statistical Office (KNSO) has been publishing CPI every month since it took over the job from the Bank of Korea in 1965. It revises the base period and the weighting system every five years. The current series' base period is 2000, and the next revision is due in 2006. The "specifications" to be priced and the weights for elementary aggregates are determined in line with the results obtained in a Household Income and Expenditure Survey.⁵ Price quotes are collected from about 13,000 retail outlets throughout 36 major cities. Most items are surveyed once a month, but fresh food and petroleum items are sampled three times a month, and school tuition fees are surveyed once a quarter. Owner-occupied shelter costs are not included in the KCPI, but the KNSO publishes an index number series that includes owner-occupied shelter costs as a supplementary series (available from January 1995).

⁵ The KNSO had been carrying out Urban Household Income and Expenditure Surveys on detailed items once per year until 2002. From 2003, however, they have changed the survey system and have been performing a Household Income and Expenditure Survey on broad sub-classes of items every year and one major survey on detailed items every five years, which will be used for the derivation of the weights for the KCPI.

Figure 1: The Hierarchy of the Construction Procedure for the National CPI



The hierarchy of the procedure for the construction of national (All Cities) CPI is shown in Figure 1. In the current weighting basket, there are 516 *elementary aggregates* (indices). For the construction of a national CPI, each of 516 elementary indices for each city is constructed first as a simple arithmetic mean of sampled *price quotes* within each *item-area stratum*.⁶ Then, the 36 city elementary indices for each elementary aggregate item are aggregated into a national elementary item index as a weighted average. The national CPI is constructed by aggregating these 516 national elementary indices using the Laspeyres index formula, defined by (2). The weights for elementary aggregates are allocated in proportion to the corresponding budget shares in the total monthly consumption expenditure by an average urban household in the base year.

Table 2 shows the weight and the number of elementary aggregates for each main group in the latest three baskets (namely, base years of 1990, 1995, and 2000) for all cities. As noted above, owner-occupied housing costs are not included in the KCPI. The nominal monthly consumption expenditure by an average urban household almost doubled over the five years' period from 1990 to 1995, and then increased by just over a quarter over a same period to 2000. It is notable that the budget share of transport and communication has been consistently increasing, while the budget shares of food, furniture and clothing are continuously decreasing. The main sources of the increase in the budget share of transport and communication are private transport and communication, which is congruent with our expectation in the light of recent changes in life style.

⁶ There are 18,576 (=516×36) item-area strata.

Table 2: Weights for Main Groups in the Latest Three Baskets (All Cities)*

Main Groups	Base = 1990	Base = 1995	Base = 2000
Food & Beverages	324.9 (167)	302.9 (171)	271.2 (180)
Housing [#]	141.7 (20)	148.3 (18)	156.4 (15)
Fuels, Electricity & Water	45.3 (9)	41.1 (8)	58.0 (8)
Furniture & Utensils	62.0 (62)	48.1 (63)	37.1 (57)
Clothing & Footwear	88.4 (53)	81.9 (54)	56.5 (43)
Health	54.5 (27)	51.3 (32)	43.9 (42)
Education	92.2 (33)	95.2 (36)	114.6 (33)
Culture & Recreation	50.2 (38)	64.7 (52)	53.6 (58)
Transport & Communication	93.6 (26)	118.1 (40)	159.3 (44)
Miscellaneous	47.2 (35)	48.4 (35)	49.4 (36)
Total Weight (Total Number of Items)	1,000.0 (470)	1,000.0 (509)	1,000.0 (516)
Total Monthly Consumption Expenditure	\679,691	\1,302,702	\1,646,447

*: The numbers in parentheses are the number of items (elementary aggregates) within each group.

#: Owner-occupied housing costs are not included.

Table 3: Decomposition of the Number of Items and the Weights in the Revisions[#]

(a) Officially Classified Old and New[@]				
	Old[*]	Continuing	New	New Basket
1995 Revision	37 (32.5)	433 (905.6)	76 (94.4)	509 (1000.0)
2000 Revision	35 (12.9)	452 (906.4)	64 (93.6)	516 (1000.0)
(b) Combined and Divided Items Included as New and Old				
1995 Revision	60 (73.7)	410 (859.9)	99 (140.1)	509 (1000.0)
2000 Revision	89 (101.2)	420 (842.7)	96 (157.3)	516 (1000.0)

[#]: The numbers in parentheses are weights as a proportion out of 1000.

[@]: In the 1995 revision, Imported Beef is officially classified as a continuing item, but the specifications are different between the 1990-base series and 1995-base series. So, the old and new specifications are classified here as Old and New, respectively. Similar cases in the 2000 revision are: the changes from Traditional Beverage to Blended Beverage, from Dish-Drying Rack to Dish Washer, and Reference Tape to Educational AV Materials.

^{*}: The weight for the “Old” items in the 1995 revision is derived from the 1990-base basket, while that in the 2000 revision is from the 1995-base basket.

The numbers and weights of new, continuing, and old items in the 1995 and 2000 revisions are shown in Table 3. The classification in the upper panel of the table reflects the official classification of newly included items and dropped items,⁷ while in the lower panel those combined into a single item or divided into multiple items are included in Old and newly defined items as a result of the combination or division are included in New. According to the official classification, in each revision, almost 10% of the total weight is for the items that are newly introduced into the basket, heralding

⁷ Refer to KNSO (1997) and KNSO (2002) for the official classifications.

the significance of the size of new-goods bias in the KCPI. The weights for the “Old” items in the two revisions are relatively small, representing 3.3% and 1.3% in the 1995 and 2000 revisions, respectively.

For the purpose of measuring the effect of adding new items or dropping old items from the basket, divided or combined items should also be treated as new or old items because dividing an item into multiple items results in more choices and combining multiple items into a single item leads to less variety. The numbers and weights in the lower panel of Table 3 are based on a classification that takes such effects into account. When those effects are accounted for, the weights representing new items and old items are 14%-16% and 7%-10%, respectively.

As a supplementary series to the official CPI, the KNSO introduced in the 2000 revision an annual Chained CPI series, which is basically a chained Laspeyres index series. It is designed to alleviate, albeit partially, the effect of fixing weights for the five-year period between revisions. Weights for this series are revised every year, based on the annual Urban Household Income and Expenditure Survey, but the items in the basket are revised only in a five-yearly major revision.⁸ So, it does not fully account for the effect of changing baskets over time.

⁸ The KNSO does not publish or make available the weights used for the chained series for some reason. As mentioned earlier in a footnote, the KNSO ceased to survey expenditures on detailed items in 2002, and it is surveying expenditures on broad sub-classes only now. How the weights for the chained series are derived every year from such coarse information is a puzzle.

4. Estimation of the Bias

Methodology

It is well known that a *superlative* index number formula, such as the Törnqvist-Theil (TT) and Fisher Ideal (FI) index, takes account of substitution effect, as it is *exact* for an aggregator function that is *flexible*.⁹ In other words, the true COL index defined by (1) can be locally approximated at least up to the second order if a superlative index number formula is used if the choice set does not change over time.

The TT index number formula is defined by:

$$(6) \quad P_T(p^0, p^1, q^0, q^1, I) = \prod_{i \in I} \left(\frac{p_i^1}{p_i^0} \right)^{[s_i^0(I) + s_i^1(I)]/2}$$

where $s_i^t(I) = p_i^t q_i^t / \sum_{j \in I} p_j^t q_j^t$, the budget share of item i out of total expenditure on the items in set I in period t , for $t = 0$ and 1 . The FI index number formula is the square root of the product of the Laspeyres index and the Paasche index:

$$(7) \quad P_F(p^0, p^1, q^0, q^1, I) = \sqrt{P_L P_P}$$

⁹ A functional form is said to be *flexible* if it provides a second-order local approximation to an arbitrary twice-differentiable function. A price index number formula is said to be *exact* for a price aggregator function if it equals the Konüs index defined by the aggregator function when the quantities are optimal values given the prices. Finally, an index number formula is said to be *superlative* if it is *exact* for an aggregator function that is *flexible*; see Diewert (1976) for more details.

where

$$(8) \quad P_L(p^0, p^1, q^0, I) = \frac{\sum_{i \in I} p_i^1 q_i^0}{\sum_{j \in I} p_j^0 q_j^0} = \sum_{i \in I} s_i^0(I) \left(\frac{p_i^1}{p_i^0} \right), \quad \text{and}^{10}$$

$$(9) \quad P_P(p^0, p^1, q^1, I) = \frac{\sum_{i \in I} p_i^1 q_i^1}{\sum_{j \in I} p_j^0 q_j^1} = \left[\sum_{i \in I} s_i^1(I) \left(\frac{p_i^0}{p_i^1} \right) \right]^{-1}$$

with obvious notations.

Note that neither formula is workable if the choice set I changes over time. So, only the substitution bias in the continuing set I can be measured by the difference between the growth rate of CPI and the growth rate of a superlative index. To estimate the true COL, accounting for the effect of changing choice sets, one needs to impose some restrictions on preference ordering. Nahm (2004) introduces an index number formula that is exact for a constant-elasticity-of-substitution (CES) type aggregator function in the presence of new and disappearing products. The formula is defined by

$$(10) \quad P_N(p^0, p^1, q^0, q^1, I^0, I^1) = P_V(p^0, p^1, q^0, q^1, I) \left[\frac{\lambda^1(I)}{\lambda^0(I)} \right]^\beta$$

where P_V is the Sato (1976)-Vartia (1976) index, which is defined by

¹⁰ Laspeyres index is redefined here to emphasize its definition on set I .

$$(11) \quad P_V(p^0, p^1, q^0, q^1, I) = \prod_{i \in I} \left(\frac{p_i^1}{p_i^0} \right)^{w_i(I)},$$

$$(12) \quad w_i(I) = \frac{[s_i^1(I) - s_i^0(I)] / [\ln s_i^1(I) - \ln s_i^0(I)]}{\sum_{j \in I} \{[s_j^1(I) - s_j^0(I)] / [\ln s_j^1(I) - \ln s_j^0(I)]\}}, \text{ and}$$

$$(13) \quad \lambda^t(I) = \frac{\sum_{i \in I} p_i^t q_i^t}{\sum_{j \in I^t} p_j^t q_j^t} \text{ for } t = 0, 1,$$

Finally, the β is defined by

$$(14) \quad \beta = \frac{\sum_{i \in I} x_i y_i}{\sum_{i \in I} x_i^2}$$

where $y_i = \ln P_V - \ln(p_i^1/p_i^0)$, and $x_i = \ln s_i^1(I) - \ln s_i^0(I)$. This β is computable using only observed prices and quantities, and it replaces the unknown parameter $1/(\sigma-1)$ in Feenstra's (1994) index (with σ denoting the elasticity of substitution), which opened a way to measure effects of new goods using an index number formula. To allow new and disappearing goods and services in the CES preference ordering, it is required that the elasticity of substitution, σ , be greater than unity; see Nahm (2004, footnote 2) for more details. This requirement implies that β should be positive, and it in turn implies that the demands for all items should be elastic with respect to a change in own price. The latter is the case because $\beta = y_i/x_i$ for all $i \in I$, and each y_i/x_i is a linear approximation to the negative of the reciprocal of the own-price elasticity of budget share, which will be

positive only if the own-price elasticity of the quantity demanded is elastic.

The right-hand side of (10) can be decomposed into three parts, namely, P_V , $\lambda^1(I)^\beta$, and $\lambda^0(I)^{-\beta}$. The Sato-Vartia index, P_V , measures the middle part of the right-hand side of equation (5), while $\lambda^1(I)^\beta$ and $\lambda^0(I)^{-\beta}$ respectively measure the first part and the third part.¹¹ Thus, an aggregate price index for the continuing items (set I), accounting for substitution effect, is obtained by P_V , while indices representing the effects of new goods (set I^{new}) and disappeared goods (set I^{old}) can be measured by $\lambda^1(I)^\beta$ and $\lambda^0(I)^{-\beta}$, respectively. Note that $\lambda^i(I)$, for $i = 0,1$, is bounded between 0 and 1 since it is the proportion of spending on continuing items out of total consumption expenditure in each period. Thus, when β is positive as explained above, the multiplier term representing the effect of having new items will be less than unity, while the term representing disappeared items will be greater than unity.

Chained index series are constructed as:

$$(15) \quad P_a(p^0, p^t, q^0, q^t, I^0, I^t) = \prod_{\tau=1}^t P_a(p^{\tau-1}, p^\tau, q^{\tau-1}, q^\tau, I^{\tau-1}, I^\tau)$$

where $P_a(p^{\tau-1}, p^\tau, q^{\tau-1}, q^\tau, I^{\tau-1}, I^\tau)$ is one of the bilateral index formulas defined above.

Data

The 1990-base, 1995-base, and 2000-base price indices and weights for each item in the three (1990, 1995, and 2000) official consumption baskets are obtained from

¹¹ The separability structure of the CES model implies that there exists a valid price aggregator for each of the groups of items in I , I^0 , I^1 , I^{new} , and I^{old} .

various issues of Annual Report on The Consumer Price Index, data CD titled “The Consumer Price Index 2000”, and the KNSO’s website, all published by the KNSO.

Annually-revised weights needed for chained indices are not published or available, and hence they have been estimated using the information from Urban Household Income and Expenditure Surveys in line with the weight-estimation procedure described in the Annual Report of the Consumer Price Index. As detailed information on the expenditure on individual items is not available from 2003, the chained indices are compiled only up to 2002.

To obtain annual price indices for individual items, the published twelve monthly price indices for each item within each year are aggregated as an arithmetic mean. The series for the same item with different base year are linked by dividing old-base index series by the old-base index value in the new base year.¹²

Bias Estimates

Annual percentage changes in the CPI, chained Laspeyres (CL), and Törnqvist-Theil (TT) indices¹³ are reported in the first three columns of Table 4. It can be seen that inflation rates were relatively high until 1999, when the full impact of the Asian financial crisis on the Korean economy manifested itself in the near-zero inflation rate. Inflation rates ever since have been lower than any of the annual inflation rates in the 1990’s until the crisis. The chained indices are computed only up to 2002 as detailed information on the expenditure on individual items has become unavailable from 2003

¹² The price index for each item in the new-base series in the new base year is always one (times 100 in the published data).

¹³ Fisher-Ideal indices are also computed, but not reported as they are very close to the TT indices.

and hence the weights cannot be estimated. The change rates for CL index in 2003 and 2004 are based on the indices that are computed by the KNSO.

Table 4: Upper-Level Substitution Bias *

	% D CPI	% D CL	% D TT	DCPI- DTT	DCL- DTT	s.d.
1990	-	-	-	-	-	-
1991	9.33	9.33	8.78	0.55	0.55	0.031
1992	6.21	5.82	5.58	0.63	0.24	0.028
1993	4.80	4.76	4.58	0.22	0.18	0.022
1994	6.27	5.68	5.30	0.97	0.39	0.027
1995	4.48	3.96	3.57	0.92	0.40	0.032
Average for Sub-Period	6.21	5.89	5.55	0.66	0.35	0.028
1996	4.92	4.84	4.61	0.31	0.23	0.026
1997	4.44	4.28	4.14	0.30	0.15	0.029
1998	7.51	7.01	6.84	0.67	0.16	0.055
1999	0.81	0.53	0.38	0.43	0.15	0.033
2000	2.26	1.92	1.63	0.63	0.30	0.025
Average for Sub-Period	3.97	3.69	3.49	0.47	0.20	0.034
2001	4.07	4.10	3.87	0.20	0.23	0.022
2002	2.76	2.64	2.43	0.34	0.21	0.031
2003	3.51	3.23%	-\$	-	-	-
2004	3.59	3.26%	-	-	-	-
Average for Sub-Period	3.48	3.31	3.15	0.34	0.16	0.03
Average for Whole Period	4.80[#]	4.55[#]	4.29	0.51	0.27	0.03

*: Percentage points per year.

#: Average over the period up to 2002 to make comparable with the other averages.

#: Estimates by KNSO.

\$: Not computable as detailed information on household expenditure is unavailable.

The difference between the percentage changes in CPI and TT index represents upper-level substitution bias in the CPI as a measure of the COL.¹⁴ On average, upper-level substitution bias alone makes the KCPI overstate annual inflation rate by 0.51. This estimate stands in striking contrast to Lee and Chung's (1997) "negligible" estimate, which is based on the data from 1975 to 1993. The bias implied by the TT index number series (excluding owner-occupied shelter costs) in Table 1 of their paper is, however, not negligible; it is in fact -0.12 per year, which is against the general belief that the substitution bias in a CPI is positive.¹⁵ Shiratsuka (1999) also reports that the upper-level substitution bias in the Japanese CPI is negligible, but the validity of the estimate is questionable because the estimate is based on the indices constructed over only less than 15% of all items used for the CPI (88 items out of 598 items).

More in line with the present result are the estimates by Lebow and Rudd (2003) for the U.S. CPI. Their estimates of upper-level substitution bias in the U.S. CPI are 0.26 and 0.54 percentage points per year for the periods 1990–1997 and 1998–2000 respectively. The biases would have been even larger than these if the weights were updated every five years, as is done for the KCPI, instead of biennially. Substitution bias in a fixed-base CPI usually increases as the comparison period moves farther from

¹⁴ Upper-level substitution bias also includes bias caused by substitution between items in different areas (cities for the KCPI). However, one could safely ignore such substitution, as the distances between the 36 cities are far enough to make substitution across cities infeasible.

¹⁵ It is unclear how they handled the changing contents of choice baskets over major revisions in computing the chained indices. Furthermore, their reasoning about the "negligible" estimate of bias in footnote 14 of their paper is based on the incorrect understanding that the 1990-base CPI series are computed using the 1990-base weights throughout the whole period. On the contrary to their belief, different base series are computed using their own weights and then linked later together by normalizing the index for the new base year to 100.

the base period. That is in fact the case in the present result, except for the bias in 1993.

Upper-level substitution bias in the U.S. CPI rose dramatically in the late 1990s. Lebow and Rudd suspect that the increase in the degree of bias might be related to increases in the variance of relative price changes during the same period. Similar increases in the late 1990s are not observed in the KCPI. However, to see if the degree of bias is correlated with the variance of relative price changes, standard deviations of changes in log prices for each bilateral comparison are computed using the following formula, which is similar to the one used by Shapiro and Wilcox (1997):

$$(16) \quad \text{Variance}(\text{change in } \ln[p^t/p^{t-1}]) = \frac{1}{n} \sum_{i=1}^n \left\{ 0.5(s_i^t + s_i^{t-1}) \right\}^2 \left[\ln(p_i^t / p_i^{t-1}) - \ln P_T \right]^2 \left. \right\}$$

where s_i^τ is budget share of item i in period τ ($\tau = t, t-1$), and P_T is the TT index defined by (6). Standard deviations based on this formula are reported in the last column of Table 4. No strong correlation can be found between the degree of bias and the standard deviation of relative price changes, with the sample correlation coefficient between the two variables equal to a meager 0.32.

The second and fifth columns of Table 4 report annual growth rates of the CL index and their deviations from the growth rates of the TT index respectively. For each bilateral comparison between a pair of consecutive periods, the CL index does not allow for the substitution effect because the weights are fixed over the two periods. However, in terms of multilateral comparisons, it alleviates the degree of substitution bias because the weights are updated every year instead of every five years. It reduces bias to the extent that the base period in each bilateral comparison is not separated by as long a time period as for the fixed-base period. The present result shows that, on average,

updating weights more frequently reduces the upper-level substitution bias by 0.24 percentage points per year, which is less than half of the substitution bias estimated above. The correlation between the remaining bias in the CL index and the standard deviation of relative price changes is weaker than that observed for the fixed-base CPI, which is in accordance with prior expectation.

Table 5: New-Goods Bias*

	CPI	P_T[§]	P_V	P_N	Bias[#]	b	s	New	Old
1990-1995	6.21	5.88	5.89	5.72	0.17	0.107	10.35	0.9840	1.0082
1995-2000	3.97	3.80	3.85	3.72	0.13	0.095	11.55	0.9839	1.0102

*: Average changes in percentage points per year for the first five columns; estimates for β and σ in the next two columns; and the multipliers in the last two columns.

§: The Törnqvist-Theil index is recomputed over the items in set I.

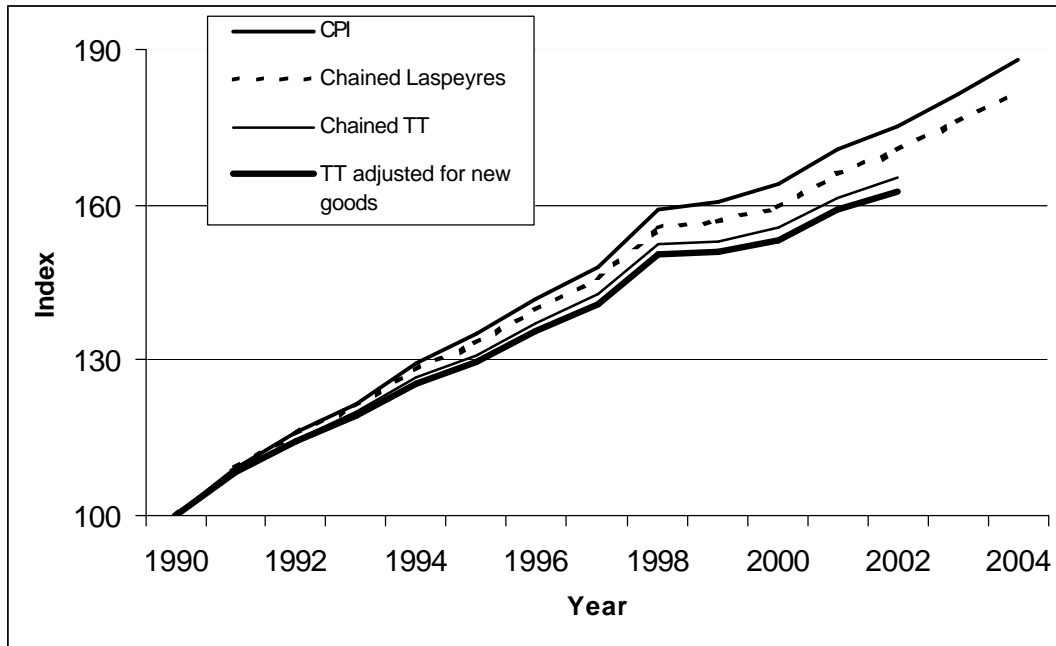
#: Difference between the percentage change in P_V and the percentage change in P_N .

New-goods bias is estimated in Table 5. The figures in the first five columns are average annual change rates, in percentage points, of CPI, P_T , P_V , P_N and the difference between the changes rates of P_V and P_N . The index numbers are computed as a fixed-base bilateral index, comparing 1995 with 1990, and 2000 with 1995; these are all revision years involving inclusion of new items and exclusion of old items. As discussed above, P_N takes into account the effects of new items and disappearing items as well as substitution across continuing items. The Sato-Vartia index (P_V), which is a part of P_N , measures the COL over the items in set I, allowing for substitution. As the Sato-Vartia index is exact for an aggregator function that is restricted in the sense that the CES function is not a flexible functional form, the validity of the restriction might be called into question. To see how closely it estimates the COL over the items in set I,

the TT index, which is a superlative index, is recomputed over set I and reported in the second column. It can be observed that the differences between the index numbers based on P_V and P_T are negligible. It in turn implies that the elasticity of substitution that is implicit in P_V and P_N is close to the average elasticity of substitution implied by the true COLI.

As mentioned earlier, each y_i/x_i has to be positive to be consistent with the existence of new and disappearing products in a choice set. For this reason, the β , defined by (14), has been estimated by the “censored method”, which is explained in Nahm (2004). The estimates of β , and of the elasticity of substitution (σ) implied by β , are reported in the sixth and seventh columns of the table. New-goods bias, which is measured by the difference between the change rates of P_V and P_N , is 0.17 and 0.13 percentage points per year over the periods 1990–1995 and 1995–2000 respectively. The multiplier terms representing the effects of new items and old items are shown in the last two columns. When interpreted in terms of deviations of the multiplier terms from unity, the deflating effect of new items dominates the inflating effect of old items in both revisions, as one would expect. The effect of new items is similar in the two revisions, while the effect of old items is slightly stronger in the 2000 revision than in the 1995 revision, thus explaining the slightly lower bias in the 2000 revision.

Figure 2: Plots of Normalized Indices



To help overview the accumulated deviations between the indices, CPI, CL, TT, and TT adjusted for new items and old items, normalized to 100 in year 1990, are plotted in Figure 2. When the price indices are constructed with 1990 as the base period, the CPI overstates P_N , which accounts for upper-level substitution and new-goods bias, by 12.9% in 2002, of which just over 10% is caused by upper-level substitution bias alone.

5. Conclusions

The present paper has estimated upper-level substitution bias and new-goods bias in the KCPI. The degree of bias stemming from the two sources is found to be significant, in contrast to an earlier finding. When accumulated over the sample period of 1990–2002, upper-level and new-goods bias overstates the cost of living in 2002 in

comparison with 1990 by almost 13%, about 10% of which arises from upper-level substitution bias alone. Although not estimated in the present paper, bias arising from other sources would be also significant. In particular, lower-level substitution bias and outlet-substitution bias would be considerably larger than the counterparts in the U.S. CPI because, unlike the U.S. procedure, entry-level items are not rotated regularly between major revisions, and the price for one item-area stratum is sampled only at a single outlet.¹⁶ In fact, the effect of lower-level substitution and outlet substitution is completely ignored in the construction of KCPI.

It is difficult to estimate the degree of quality-change bias as there exist few studies about such bias in the KCPI, while quality-change bias has to be measured separately for narrowly-defined groups of individual items. The only mechanism that is designed to keep the quality of items constant in the construction procedure of KCPI is using the direct-comparison, the linking, and the cost-estimation methods when there is a change in the specification of a sampled item, hence missing out all the other sources of quality-change bias.

The “Chained Index” published by the KNSO as a reference series is a chained Laspeyres index. As noted earlier, the CL index eliminates only less than a half of upper-level substitution bias. A chained superlative index, which could be computed with little extra cost in addition to that needed to compute the CL index,¹⁷ would contain much more useful information than the CL index, making it easier to observe

¹⁶ Recall that prices are sampled at about 13,000 outlets throughout all the cities while there are 18,576 item-area strata in the present system.

¹⁷ Extra cost would occur only if the KNSO wanted to publish a chained superlative index earlier than the point in time when the information on the weights for the comparison period currently becomes available and thus it had to collect the information more quickly.

the full degree of upper-level substitution bias in the CPI. Alternatively, the KNSO could make the information about the annually-revised weights available to interested parties so that they could use the information to measure substitution bias themselves.

More research, especially on quality-change bias, will have to be done before a reliable estimate for the overall bias in the KCPI can be put forward. However, every piece of information obtained in the present study indicates that the degree of the overall bias will be significantly high.

REFERENCES

Boskin, Michael J., Ellen R. Dulberger, Zvi Griliches, Robert Gordon, and Dale Jorgenson, *Toward a More Accurate Measure of the Cost of Living - Final Report to the Senate Finance Committee from the Advisory Commission To Study The Consumer Price Index*, 1996.

Diewert, W. Erwin, “Exact and Superlative Index Numbers”, *Journal of Econometrics* 4(2), 115–145, 1976.

_____, “Index Numbers”, in J. Eatwell, M. Milgate and P. Newman, ed., *The New Palgrave: A Dictionary of Economics Vol. 2*, 767–780, The Macmillan Press, 1987, reprinted in W.E. Diewert and A.O. Nakamura, ed., *Essays in Index Number Theory Vol. 1*, 71–104, North-Holland, 1993.

_____, “Comment on CPI Biases”, *Business Economics* 31(2), 30–35, 1996.

Feenstra, Robert C., “New Product Varieties and the Measurement of International Prices”, *The American Economic Review* 84(1), 157–177, 1994.

Gordon, Robert J., “The Boskin Commission Report and Its Aftermath”, *Monetary and Economic Studies* 7(3), 41–68, Bank of Japan, 1999.

Hausman, Jerry, “Sources of Bias and Solutions to Bias in the Consumer Price Index”, *Journal of Economic Perspectives* 17(1), 23–44, 2003.

Hicks, J.R., “The Valuation of the Social Income”, *Economica* 7(26), 105–124, 1940.

Konüs, A.A., “The Problem of the True Index of the Cost of Living”, *The Economic Bulletin of the Institute of Economic Conjuncture* 9-10 (36-37), 64–71, (in Russian); English translation was published in 1939 in *Econometrica* 7(1), 10–29, 1924.

Korea National Statistical Office (KNSO), “An Overview of the Revision of '95-Based Consumer Price Index” (in Korean), *Annual Report on the CPI 1996*, 470–487, 1997.

Korea National Statistical Office (KNSO), “Results of the Revision of 2000-Based CPI”

(in Korean), *Media Release*, 23 January 2002.

Lebow, David E. and Jeremy B. Rudd, “Measurement Error in the Consumer Price Index: Where Do We Stand?”, *Journal of Economic Literature* 41(1), 159–201, 2003.

Lee, Eunjeong, “How to Reflect Quality Change in the CPI: The Case of Korea”, paper presented at the 6th Meeting of the International Working Group on Price Indices (Ottawa Group) in Canberra, Australia, 2001.

Lee, Tong Hun, and Keun Jon Chung, “On Measuring the True Cost of Living for the Developing Country: Case of Korea”, *International Economic Journal* 11(1), 1–14., 1997.

Nahm, Daehoon, “Incorporating the Effect of New and Disappeared Products into the Cost of Living: An Alternative Index Number Formula”, *Seoul Journal of Economics* 11(3), 261–282, 1998.

_____, “An Exact Index for the CES Function with Unobservable Prices”, *Journal of Productivity Analysis* 21, 317–336, 2004.

Sato, Kazuo, “The Ideal Log-Change Index Number”, *The Review of Economics and Statistics* 58(2), 223–228, 1976.

Shiratsuka, Shigenori, “Measurement Errors in the Japanese Consumer Price Index”, *Monetary and Economic Studies* 17(3), 69–102, Bank of Japan, 1999.

Schultze, Charles L., “The Consumer Price Index: Conceptual Issues and Practical Suggestions”, *Journal of Economic Perspectives* 17(1), 3–22, 2003.

Shapiro, Matthew D. and David W. Wilcox, “Alternative Strategies for Aggregating Prices in the CPI”, *Federal Reserve Bank of St. Louis Review* 79(3), 113–137, 1997.

Solow, Robert M., “The Nature of Consumer Price Indices”, *Journal of the Federation of American Scientists* 50(2), 1997.

Tobin, James, “Thoughts on Indexing the Elderly”, *Journal of the Federation of*

American Scientists 50(2), 1997.

Triplett, Jack E., “Should the Cost-of-Living Index Provide the Conceptual Framework for a Consumer Price Index?”, *Economic Journal* 111(472), 311–334, 2001.

Turvey, Ralph, “True Cost of Living Indexes”, paper presented at the 5th Meeting of the International Working Group on Price Indices (Ottawa Group) in Reykjavik, Iceland, 1999.

Vartia, Y. O., “Ideal Log-Change Index Numbers”, *Scandinavian Journal of Statistics* 3, 121–126, 1976.