IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Estimating effective inflation using scanner data[1]

Younghwan Lee,
Bank of Korea

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Estimating Effective Inflation Using Scanner Data

Younghwan Lee[1]

Abstract

This paper presents a cost of living index using Korean scanner data collected between February 2017 and October 2022. As most existing official statistics of consumer price indices measure the cost of living based on changes in the prices of a few selected varieties collected by survey methodology, this can be costly and may mislead policy makers and economists. By contrast, the new index accounts for changes in consumption patterns, such as substitution effects and changes in preferences, as well as changes in the market environment, such as the entry and exit of product varieties, by exploiting the information in the scanner data. Although scanner data are a valuable source of information for improving the consumer price index, there have been few implementation examples in practice owing to technical difficulties. In this paper, I propose some theoretically founded solutions with an implementation example. The proposed solutions can be implemented with manageable computing costs Empirically, the new index not only closely correlates with existing official statistics but it also statistically leads the existing indices, thus serving as a leading indicator for inflation dynamics.

Keywords: cost of living index, scanner data, chain index, Törnqvist index

Table of Contents

[1] Data Research Section, Bank of Korea

E-mail: yhlee@bok.or.kr

Disclaimer: The views expressed in this article do not necessarily reflect those of Bank of Korea.

1. Introduction

Traditionally, many national statistical institutions (hereafter, NSIs) measure inflation using the fixed-base method which assumes that the consumption basket remains constant over time. However, this unrealistic assumption of consumer behavior and the market environment can mislead us measuring the price levels faced by households in their daily lives; thus, it is ineffective to capture economically meaningful price movements. This study presents a method for constructing an effective price index using scanner data. The proposed method allows for changes in the consumption basket by considering substitution effects, changes in preferences, and the entry-exit of items. This approach enables a more accurate measurement of effective prices.

Economists have long acknowledged the bias stemming from the use of traditional fixed-base methods, which are commonly linked to three primary sources: substitution bias, changes in preference, and entry-exit bias. Braun and Lein (2021) quantified the extent of this bias by measuring the inflation originating from these sources. Their findings indicate that overlooking changes in preference and entry-exit bias results in an average bias of 2.6% in the annual inflation rate, which is economically significant. The following three motivating questions illustrate how the fixed-base method can yield biased measurements of price levels in relation to the identified sources of bias.

Q. (Substitution bias) Suppose there is a product with a regular price of $100, but it is available at a 50% discount every Monday, Wednesday, and Friday. What is the effective price of this product?

Assuming rationality among consumers and the storability of the product, consumers will not purchase the product on Tuesdays, Thursdays, Saturdays, or Sundays at a price twice as high. In other words, consumers respond to price changes to minimize spending while maximizing welfare. Therefore, assuming a fixed consumption basket can lead to overestimating the price levels.

Q. (Changes in preference) In 1976, the price of the Apple 1 was $666.66, and in 2014, it was sold for $0.9 million. What is the inflation rate of computer prices between 1976 and 2014?

Although computing power of the Apple 1 lags significantly behind that of modern computers, its price has increased substantially over the last 40 years. This is primarily due to changes in consumer preferences for Apple 1. When Apple 1 was first introduced in the market, consumers purchased it because of its computational ability. However, over time, consumers perceive it as an iconic piece of art rather than a computer. Therefore, interpreting the change in the price of the Apple 1 as indicative of changes in the price levels of computers is generally misleading.

Q. (Entry-exit bias) Did the price of stone axes increase during the COVID-19 period?

The price of stone axes can be a significant indicator of the cost of living during the Stone Age. However, it is irrelevant to track the cost of stone axes as a measure of the cost of living for modern households because we no longer use stone axes. Advances in technology, deterioration of profitability, and changes in government regulations influence the variety of products available in the market. Thus, if such changes in the market environment are not considered when constructing a consumption basket, the measurement of price levels can become distorted.

The primary reason why the fixed-base method is preferred by many National Statistical Institutions is the difficulty in obtaining timely access to consumption quantities. To update the consumption basket to reflect changes in household consumption, data on consumption amounts should be available. However, collecting such information and updating the consumption basket directly in a timely manner can be costly. NSIs, such as Statistics Korea, routinely update the consumption basket indirectly through household surveys.

To address this problem, I propose a method to construct an effective price index using scanner data. Scanner data provides detailed records of real transactions in the market, including the sales price of specific items and the quantities sold. Assuming that the timing of transactions and consumption are sufficiently close, scanner data can shed light on the evolution of household consumption behavior. Specifically, the effective price index is defined as a chain index that operates under the Törnqvist formula derived from the household consumption problem. Based on an economically founded model, the proposed method accounts for changes in consumption behaviour, thereby enabling a more accurate measurement of price levels. Additionally, the Rolling Window GEKS method is employed to manage the measurement error problems associated with the chain method, ensuring index stability.

The remainder of this paper is structured as follows. Section 2 outlines the methodology in detail. Section 3 applies the proposed method is applied to Korean scanner data to construct an effective price index and examines its empirical properties alongside the official consumer price index. Finally, Section 4 concludes the paper.

2. Methodology

The effective price index in this study is defined as the cost of living index (COLI), an economically founded concept. It is defined as a minimum level of expenditure required to sustain a constant level of utility. The Törnqvist formula and the chain identity are derived from the properties of COLI. From the definition of COLI, the consumption basket determined by Hicksian demand which evolves according to the changes in prices and preferences of consumers. In addition, the fact that the COLI has chain identity theoretically supports the use of the rolling windows GEKS (RWGEKS) method which is applied to suppress the chain drift problem in this study.

This section proceeds as follows. First, I derive the COLI from the economic model and investigate how changes in prices and preferences of consumers interact with COLI. Specifically, the COLI and Hicksian demand are derived from the cost minimization problem to shed light on possible sources of bias in the fixed-base method. Second, the theoretical and empirical support for the use of the chain index is provided. Third, the Törnqvist formula is derived as a COLI measurement. This implies that applying Törnqvist formula can alleviate the bias caused by the substitution effect and changes in preferences. Fourth, the RWGEKS method is introduced to suppress measurement error accumulation.

2.1 Cost of Living Index (COLI)

Assume an environment where a representative household minimizes the cost of living for each period t while achieving the level of utility $\bar{U}$. Let the household's utility at t be $u(c_t \mid \alpha_t)$ where the price vector of items and the parameter vector of preference of household are denoted by p_t and α_t, respectively. Then the Hicksian demand $c^*(\bar{U}, p_t|\alpha_t)$ is a cost-minimizing consumption bundle that is the solution to the following cost-minimization problem:

$$c^*(\bar{U}, p_t|\alpha_t) = \mathrm{argmin}_c\, p_t^T \cdot c \quad \text{s.t.} \quad \bar{U} = u(c_t|\alpha_t).$$

Then, $\Pi_{0,t}$, the COLI between period 0 and t is derived as follow:

$$\Pi_{0,t} = \frac{p_t^T \cdot c^*(\bar{U}, p_t|\alpha_t)}{p_t^T \cdot c^*(\bar{U}, p_0|\alpha_0)}. \tag{1}$$

From the above, we observe that COLI monitors changes in living expenditures through a weighted sum of prices, where the weights are determined by Hicksian demand. Two noteworthy observations were made. First, assuming that the other factors remain constant, households respond to changes in their preferences by increasing their consumption of preferred goods. In other words, as a product becomes preferable while its price remains constant, consumers can minimize their cost of living by increasing their consumption of these preferred goods and reducing their consumption of comparatively less preferred goods. This implies that the consumption basket varies with changes in the preferences.

Second, an increase in product price does not necessarily lead to an equivalent increase in COLI. Suppose that the price of a product variety increases while other factors remain constant, and the utility function is sufficiently regular. Thus, it is more advantageous for households to shift their consumption from the product variety with increased prices to other, relatively cheaper products. Specifically, as long as the marginal utility of consuming products decreases, households can maintain their level of utility with reduced expenditures by reallocating their budgets to relatively less expensive products. This finding demonstrates that the consumption basket responds to price changes, which can be a potential source of substitution bias that can overestimate consumer expenditure.

2.2 Chain Index

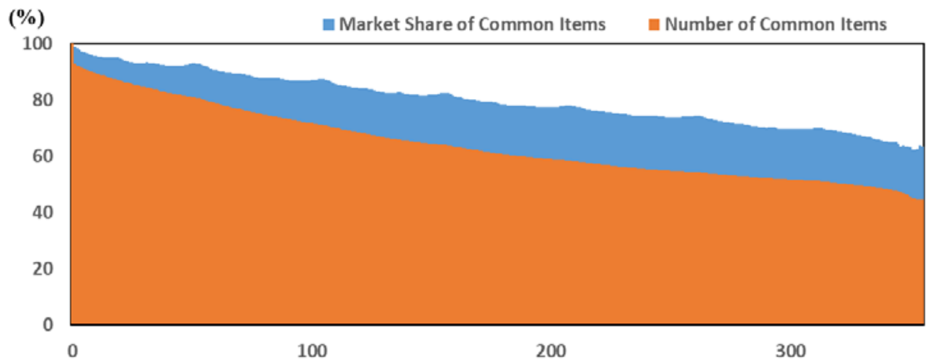
The chain index measures the change in price levels between two distant periods by accumulating changes in price levels over smaller time intervals. This allows us to update the composition of the consumption basket frequently while keeping it realistic. This method is justified by the chain identity property of the COLI, which is a consequence of equation (1). Specifically, the following relationship can be derived directly from equation (1):

$$\Pi_{0,t} = \Pi_{0,1} \times \Pi_{1,2} \times \cdots \times \Pi_{t-1,t}\,.$$

The theoretical implication of chain identity is whether to directly measure changes in the price level between two distant periods or indirectly calculate it by accumulating changes in smaller time intervals, the chain links, the results are equivalent.

Attrition Rate[1]

Figure 1.



[1] The vertical axis represents the percentage share of common items available in both measurement periods, while the horizontal axis represents the time interval between the two measurement periods in weeks.

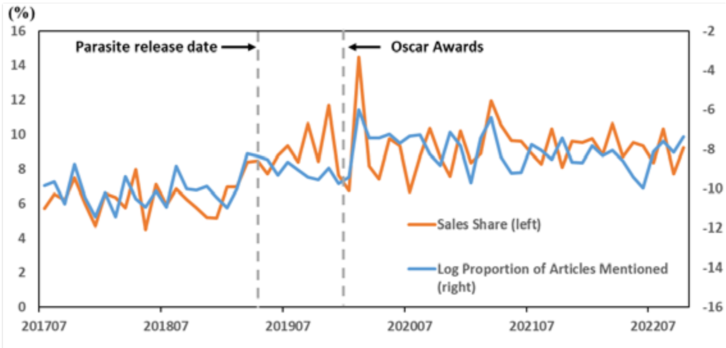
Source: Author's calculation

Although the chain index requires more computation and theoretically provides the same results as the direct comparison method, working with it offers two advantages. First, it can help alleviate entry-exit bias. This is because, as the measurement points become less distant, more common product varieties become available for comparing price levels. Figure 1 illustrates the proportion of common products in terms of the number of product varieties and their sales share relative to the time interval between the two measurement periods in weeks. This shows that the closer the time interval is, the more common the available product varieties are. Specifically, on average, approximately 20% (7% in terms of sales amount) of items disappear from the common product baskets within 1 year, and 45% (26% in terms of sales amount) of items disappear within 5 years. The fact that the proportion of items that disappeared from the common set in terms of the number of items is larger than that in terms of sales amount implies that the entry and exit of product varieties is not a random phenomenon. Rather, it suggests that relatively popular items tend to survive longer. This survival bias can also exacerbate the bias resulting from the entry and exit of product varieties when measuring the COLI.

Second, it is unrealistic to assume the same consumption basket for two distant periods because the composition of the consumption basket is subject to change for various reasons, such as the marketing activities of firms or seasonality. Figure 2 illustrates how the composition of a consumption basket can be altered for various reasons. For example, the Oscar-winning movie "Parasite" features a scene that includes the preparation of "Jjapagetti," one of the representative instant noodle brands in Korea. After the movie's release, the sales share of "Jjapagetti" began to increase and reached its peak after the film won the Oscar. Prior to the release of "Parasite," the average market share was approximately 6.4% in the instant noodle market. However, the average share increased to 9.3% immediately after the Oscar awards, which is economically significant. Interestingly, the proportion of economic news articles mentioning " Jjapagetti" showed a similar increasing pattern during the same period. One plausible hypothesis is that consumers changed their preferences as the product gained popularity because of the movie, leading to adjustments in their consumption baskets. This example illustrates how quickly the composition of a consumption basket can change, even in a short amount of time.

Change in Sales Share of "Jjapagetti" [1]

Figure 2.



[1] The left vertical axis represents the market share of "Jjapagetti" in the instant noodle market, while the right vertical axis represents the logarithmically scaled proportion of economic news articles that mention "Jjapagetti."

Source: Author's calculation

2.3 Törnqvist Formula

The Törnqvist index is the weighted geometric mean of the change in the price of product varieties in a household's consumption basket, where the weight is defined as the average expenditure weight of the product. In addition, it is derived from the definition of the COLI under certain mathematical conditions. The Törnqvist formula is as follows:

$$\Pi_{s,t} = \prod_{i \in I_{s,t}} \left(\frac{p_{i,t}}{p_{i,s}}\right)^{\frac{w_{i,s}+w_{i,t}}{2}}.$$

Here, $I_{s,t}$ denotes the set of available product varieties available both at time s and t, and $p_{i,t}$ and $w_{i,t}$ are price and expenditure share of product i at time t, respectively. Assume following Cobb-Douglas utility function of the household:

$$u(c_t|\alpha_t) = \prod_i c_{i,t}^{\alpha_{i,t}},$$

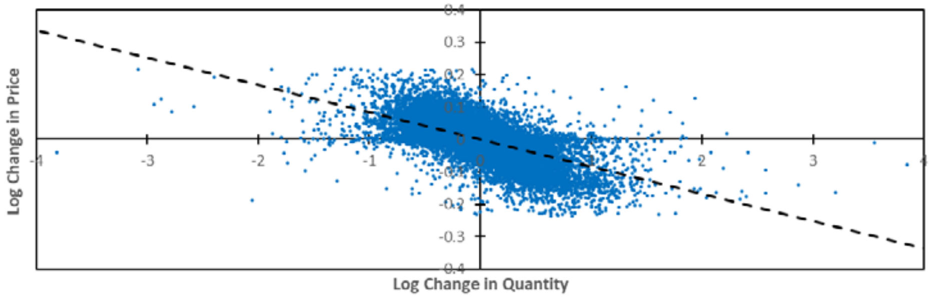
where $\sum_i \alpha_{i,t} = 1$ and $\alpha_{i,t} > 0 \; \forall \, i$. In this case, the COLI is equal to the following:

$$\Pi_{s,t} = \prod_{i \in I_{s,t}} \left(\frac{p_{i,t}}{p_{i,s}}\right)^{\alpha_{i,t}}.$$

In equilibrium, each $\alpha_{i,t}$ equal to the expenditure share of product i at time t. The above COLI is equal to the Törnqvist formula if we substitute $\alpha_{i,t}$ by $\frac{w_{i,s}+w_{i,t}}{2}$. In other words, if we approximate the expenditure share at times s and t by their average expenditure share, the Törnqvist index becomes the COLI.

Relationship between Change in Price and Price in Sales Quantity of "Instant Noodle" [1]

Figure 3.



[1] Each point in the figure shows the weekly log change in price and quantity sold for 'Instant noodle' category items at 90 supermarkets from July 2017 to Oct. 2022.

Source: Author's calculation

The Törnqvist index formula, derived as a COLI that operates under Hicksian demand, effectively addresses the bias stemming from substitution effects. Figure 3 illustrates the statistically significant negative correlation between changes in the sales quantities and prices. This indicates that households adjust their consumption baskets in response to price shifts, decreasing the demand for higher-priced items. Neglecting this substitution effect by assuming a fixed consumption basket could potentially overstate inflation because an increase in the prices of certain product varieties may be misinterpreted as an increase in the aggregate price level. However, the Törnqvist index formula consider changes in the consumption basket owing to substitution effects, allowing for a more precise measurement of the cost of living.

2.4 Rolling Window GEKS (RWGEKS)

The fundamental challenge encountered when working with a chain index is the tendency of measurement errors in each chain link to accumulate and amplify rather than cancel out. This issue is akin to the unit-root problem in a time-series analysis. To address this issue, the Rolling Window GEKS (RWGEKS) method proposed by Ivancic et al. (2011) was employed. Recognizing how the accumulation of measurement errors can lead to excessive volatility in price indices is essential. Denote $\ln \Pi_{s,t}$ by $\pi_{s,t}$. The chain index is defined as follows:

$$\pi_{0,t} = \sum_{s=1}^{t} \pi_{s-1,s}.$$

However, it is important to note that we cannot directly observe $\pi_{s-1,s}$. Instead, the best we can obtain is its noise-contaminated estimate, denoted as $\tilde{\pi}_{s-1,s}$ where $\tilde{\pi}_{s-1,s} = \pi_{s-1,s} + \varepsilon_{s-1,t}$. The error in $\tilde{\pi}_{0,t}$ as calculated by the chain index increases as t advances, as shown below:

$$\begin{aligned} \tilde{\pi}_{0,\mathrm{t}} - \pi_{0,\mathrm{t}} &= \sum_{\mathrm{s}=1}^{\mathrm{t}} \tilde{\pi}_{s-1,s} - \sum_{\mathrm{s}=1}^{\mathrm{t}} \pi_{\mathrm{s}-1,\mathrm{s}} \\ &= \sum_{\mathrm{s}=1}^{\mathrm{t}} \varepsilon_{\mathrm{s}-1,\mathrm{s}}. \end{aligned}$$

This implies that the price index becomes more uncontrollably erratic as the chain grows longer. To mitigate this noise, a is required to suppress it.

The RWGEKS method addresses this issue by identifying multiple potential measurements of the chain link $\pi_{s-1,s}$ using the chain identity. It then computes a proxy $\hat{\pi}_{s-1,s}^{GEKS}$ for $\pi_{s-1,s}$ by averaging these measurements. Specifically, for each $\mathrm{l} > 0$, we can obtain a noisy measurement $\tilde{\pi}_{s-1,s}^{l}$ for $\pi_{s-1,s}$ as follow:

$$\begin{aligned} \tilde{\pi}_{s-1,s}^{l} &= \tilde{\pi}_{l,s} - \tilde{\pi}_{l,s-1} \\ &= \pi_{s-1,s} + \varepsilon_{s-1,s}^{\mathrm{l}}, \end{aligned}$$

where $\varepsilon_{s-1,s}^{l} = \varepsilon_{l,s} - \varepsilon_{l,s-1}$. Therefore, based on the definition of the RWGEKS chain link $\hat{\pi}_{s-1,s}^{GEKS}$, it becomes clear that the bias caused by measurement error can be controlled by selecting the appropriate value for L as follows:

$$\begin{aligned} \hat{\pi}_{s-1,s}^{GEKS} &= \frac{1}{L}\sum_{l=1}^{L} \tilde{\pi}_{s-1,s}^{l} \\ &= \pi_{s-1,s} + \frac{1}{L}\sum_{l=1}^{L} \varepsilon_{s-1,s}^{l}. \end{aligned}$$

The second equation shows that the size of error can be regulated by controlling for the number of averaged samples L. That is, the larger the L, the smaller the error term. However, using a larger L can reduce the sensitivity of capturing changes in the price level. Thus, it is important to determine the proper value of L.

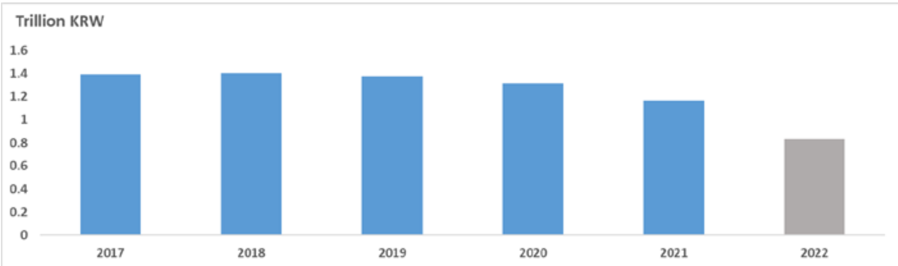
In this study, I constructed a price index with different values of L: 4, 13, and 52 weeks, which correspond to one month, one quarter, and one year, respectively. The empirical results suggest that the price index exhibits the best balance between stability and sensitivity when L is 13 weeks. Therefore, 13 weeks was chosen for this study.

3. Empirical Results

I applied the new method to scanner data collected from 90 supermarkets in Korea from July 2017 to October 2022 to test the empirical validity of the effective price index (EPI). Figure 4 shows that the data covers an average annual sales amount of approximately 1.4 trillion KRW, which is sufficiently large to assess the validity of the new index.[2] The data contains store-level information about the weekly sales amount and quantity of each product variety identified by its corresponding barcode. The barcode values for each product variety are universal across all stores, and their product names and categories are also identifiable. As the source of the scanner data is supermarkets, the variety of items included in the scanner data is limited to those typically sold at supermarkets. Approximately 80% of sales come from food items, and 20% come from living necessities.

Total Sales Amount of Sample Stores [1]

Figure 4.



[1] As data cover sales until Oct. 2022, the sales amount of 2022 is incomplete.

Source: Author's calculation

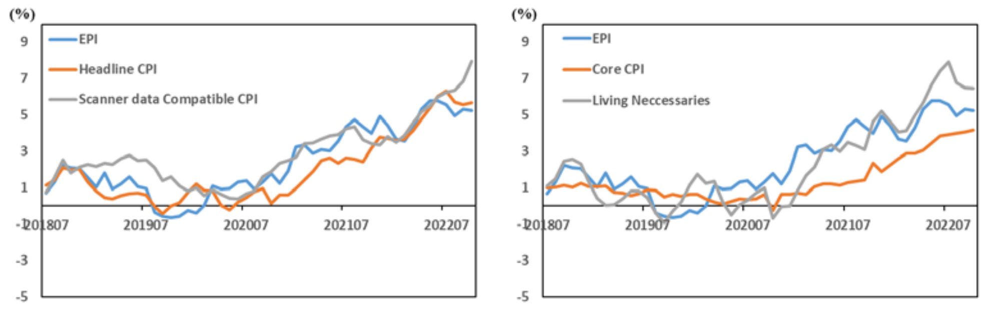
To investigate the empirical properties of the EPI, I compared it with various official CPIs, including the headline CPI, core CPI, living necessity CPI, and their variant, the scanner data-compatible CPI. The scanner data-compatible CPI is designed to serve as a benchmark for assessing the validity of the EPI because the new index covers a subset of the consumption basket owing to the limited coverage of scanner data. It is calculated using CPI values at the subcategory level, along with their respective weights in the CPI. These subcategories align with the items included in the scanner data.

Figure 5 plots the EPI alongside the official CPIs and the scanner data-compatible CPI. The correlation coefficient between the EPI and the headline CPI, core CPI, living necessity CPI, and the scanner data compatible CPI turn out to be 0.9, 0.88, 0.81, and 0.88, respectively. Notably, the correlation between the EPI and the headline CPI is slightly higher than that of the scanner data-compatible CPI, even though the EPI does not cover all product varieties included in the headline CPI. This could be attributed to indirect influences from changes in prices of other categories, such as oil prices, which are not directly reflected in the EPI but are captured through adjustments in consumption behavior and price changes in some product varieties in the scanner data, given the large number of product varieties in the EPI.

[2] The data set has been generously provided by The Korea Chamber of Commerce and Industry.

EPI and official Consumer Price Indexes [1]

Figure 5.



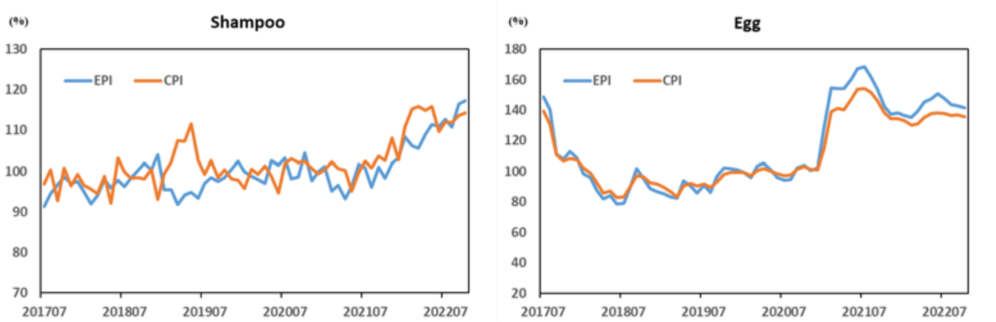
[1] All indexes are year-over-year based.

Source: Author’s calculation, Statistics Korea

Where do these differences originate? One hypothesis is that the EPI reflects a competitive market structure different from the official CPI. The fundamental difference between the methods used to measure the EPI and official CPIs is that the EPI allows consumption baskets to evolve dynamically, whereas the price data of all product varieties in the scanner data are included. In contrast, official CPIs consider only changes in the prices of a few representative varieties. The strategy of relying on a few representative product varieties is effective when the product varieties within the category are highly homogeneous, and the market is close to perfect competition. In such cases, changes in a few representative product varieties provide sufficient information to describe price level changes.

EPI and CPI for “shampoo” and “egg” [1]

Figure 6.



[1] The average price level of 2020 is set to 100.

Source: Author’s calculation, Statistics Korea

However, if the product varieties within a category are highly heterogeneous and the market exhibits monopolistic characteristics to some extent, tracking the prices of a few product varieties to measure the overall price level of the product category can be misleading. Figure 6 illustrates this point. It plots the EPI and CPI for the "shampoo" and "egg" categories. The "shampoo" category is presumed to be more heterogeneous than the "egg" category. Consequently, the deviation of the "shampoo" EPI tends to be more pronounced compared to the official CPI than the "egg" EPI. In contrast, in the case of "egg," the EPI and the official CPI are closely aligned for the entire period. This serves as evidence to support the hypothesis.

Result of Granger Causality test (Z = EPI) [1]

Table 1.

Y	N of Obs.	Null hypothesis	
		Z ≠> Y	Y ≠> Z
Headline CPI	49	4.15 (0.25)	1.87 (0.6)
Scanner data Compatible CPI	49	10.39** (0.02)	1.5 (0.68)
Core CPI	49	20.3*** (1.5E-4)	2.95 (0.4)
Living Necessity CPI	49	5.22 (0.16)	2.91 (0.41)

[1] The values in the third and fourth columns of the table represent the Wald statistics, with the corresponding p-values displayed in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

Source: Author's calculation

To assess whether the EPI can help predict the direction of inflation, a Granger causality test is conducted and the results indicate that the EPI is not only highly correlated with the official CPI but also significantly leads to the core CPI. Specifically, the model is assumed as follows:

$$y_t = \beta_0 + \sum_{m=1}^{M} \beta_m y_{t-m} + \sum_{n=1}^{N} \gamma_n z_{t-n} + \varepsilon_t.$$

The maximum time lags M and N are both set to 3. Using this model, whether z_t Granger cause y_t was tested under the null hypothesis that z_t does not cause y_t. That is, for example, to test whether EPI causes core inflation; z_t is set to EPI, and y_t is set to core inflation. Table 1 summarizes the results. This shows that the EPI Granger cause core inflation significantly while there is no sufficient evidence for the converse. As shown in Table 2, an alternative test was conducted to see if the headline CPI can also predict core inflation, but there is insufficient evidence to support it. This finding implies that the new index provides unique information that allows us to closely monitor and predict the future direction of inflation. To confirm the robustness of the empirical results, the same procedure was applied under alternative specifications, where M and N are set to 1,2, and 3, and the results were consistent over all specifications.

Result of Granger Causality test (Z = Headline CPI) [1]

Table 2.

Y	N of Obs.	Null hypothesis	
		Z ≠> Y	**Y ≠> Z**
EPI	49	1.87 (0.6)	4.15 (0.25)
Scanner data Compatible CPI	49	6.3* (0.1)	11.71*** (0.01)
Core CPI	49	4.69 (0.2)	2.4 (0.49)
Living Necessity CPI	49	4.74 (0.19)	2.83 (0.42)

[1] The values in the third and fourth columns of the table represent the Wald statistics, with the corresponding p-values displayed in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

Source: Author's calculation, Statistics Korea

4. Concluding Remark

Thus far, a method for constructing an effective price index using scanner data has been introduced, and its empirical properties have been investigated. This method allows us to accurately measure the inflation households face in their daily lives by operating under an economic model describing household consumption behavior. The empirical investigation shows that the EPI conveys information similar to that of existing official CPIs while leading the Core CPI. Thus, it can help monitor and forecast inflation closely.

References

Braun, R., & Lein, S. M. (2021). Sources of bias in inflation rates and implications for inflation dynamics. *Journal of Money, Credit and Banking,* 53(6), 1553-1572.

Ivancic, L., Diewert, W. E., & Fox, K. J. (2011). Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics*, 161(1), 24-35.

Estimating Effective Inflation Using Scanner Data

Younghwan Lee

Bank of Korea

Questions

- Suppose there is a product with a regular price of \$100, but it is available at a 50% discount every Monday, Wednesday, and Friday. What is the effective price of this product?
- In 1976, the price of the Apple I was \$666.66, and in 2014, it was sold for \$0.9 million. What is the inflation rate of computer prices between 1976 and 2014?
- Did the price of stone axes increase during the COVID-19 period?

Ideas

- Challenges:
  - The fixed base index with a fixed consumption basket cannot take consumer behavior into account when measuring the effective price index.
  - The lack of quantity data is the root of the problem.
- Strategy:
  - Utilize **scanner data** and the **Tönqvist formula** to construct a **chain index** that captures changes in consumer behavior.
  - Using **RWGEKS** of Ivancic et al. (2011), to mitigate the amplification of measurement noise in the chain index.

Cost of Living Index (COLI)

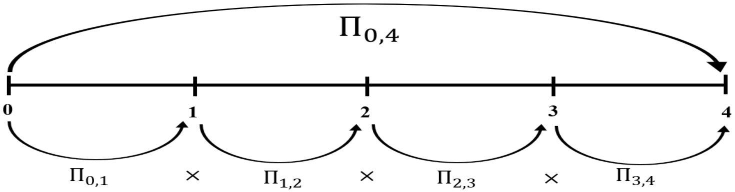
- The **Hicksian demand** $\mathbf{c}^*(\bar{U}, \mathbf{p}_t, |\alpha_t)$ that achieves the utility $\bar{U}$ given the price vector $\mathbf{p}_t$ and the time-varying preference α_t is the solution of the following cost minimization problem:

$$\mathbf{c}^*(\bar{U}, \mathbf{p}_t, |\alpha_t) = \underset{\mathbf{c}}{\operatorname{argmin}}\, \mathbf{p}_t^T \cdot \mathbf{c} \qquad \text{s.t.} \qquad \bar{U} = u(\mathbf{c}|\alpha_t).$$

- Accordingly, $\Pi_{0,t}$, the COLI inflation measure between time 0 and t, is defined as the change in expenditure required to sustain the household's level of utility, as follow:

$$\Pi_{0,t} = \frac{\mathbf{p}_t^T \cdot \mathbf{c}^*(\bar{U}, \mathbf{p}_t, |\alpha_t)}{\mathbf{p}_0^T \cdot \mathbf{c}^*(\bar{U}, \mathbf{p}_0, |\alpha_0)}.$$

Implication of COLI: Chain Identity



- Formally, the Chain Identity implies the following:

$$\Pi_{0,t} = \Pi_{0,s} \times \Pi_{s,t} \qquad \forall s.$$

- Thus, we can measure the COLI inflation by measuring a sequence $\{\Pi_{s-1,s}\}$ instead of directly measuring $\Pi_{0,t}$.

Why Chained Index?: Entry-Exit Bias

- The number common items that exist in both period decreases in the time difference between the measurement points.

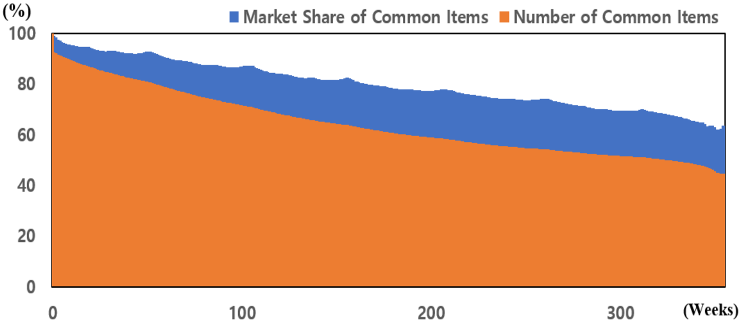


Figure: Attrition Rate

Why Chained Index?: Change in Consumer Preferences

- Consumer preferences can change over time...

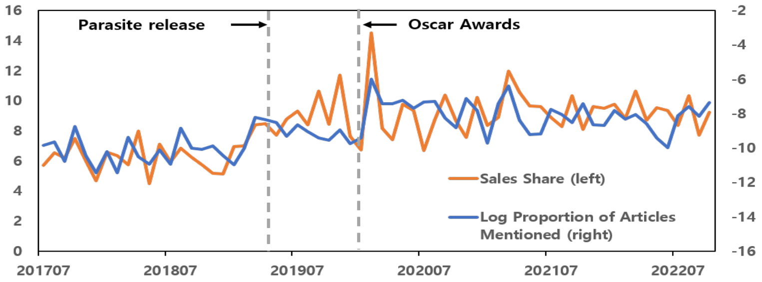


Figure: Sales Share of Jjapagetti

Implication of COLI: Törnqvist Formula

- The Törnqvist formula is defined as follows:

$$\Pi_{s,t}^{\text{Törn}} = \prod_{i \in I} \left(\frac{p_{i,t}}{p_{i,s}} \right)^{\frac{w_{i,s}+w_{i,t}}{2}},$$

  where $w_{i,t}$ is the sales share of the item i at t.

- Assume that the utility function is a Cobb-Douglas function as follow:

$$u(\mathbf{c}_t|\alpha) = \prod_{i \in I} c_{i,t}^{\alpha_i},$$

  where $\alpha_i = \frac{w_{i,s}+w_{i,t}}{2}$. Then, the COLI inflation between s and t coincides with the Törnqvist index.

Why Törnqvist formula?: Substitution bias

- Usually, an increase in price decreases the sales quantity.

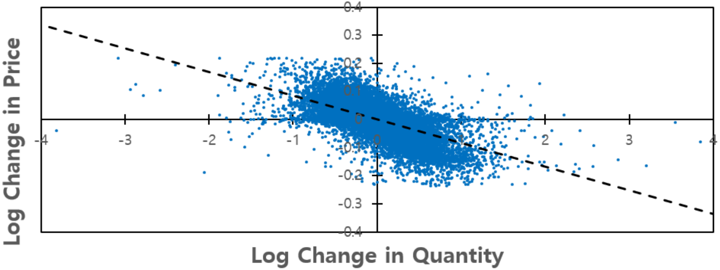


Figure: Substitution Effect in Instant Noodle Market

Why RWGEKS?

- Chain index is vulnerable to noise if the measurement error of each chain link is not suppressed. Let $\pi_{s,t}$ be $\ln \Pi_{s,t}$ and let $\hat{\pi}_{s,t} = \pi_{s,t} + \varepsilon_{s,t}$ be a noisy measurement of $\pi_{s,t}$. Then,

$$\begin{aligned}\sum_{s=1}^{t} \hat{\pi}_{s-1,s} &= \sum_{s=1}^{t} \pi_{s-1,s} + \varepsilon_{s-1,s} \\ &= \pi_{0,t} + \sum_{s=1}^{t} \varepsilon_{s-1,s}.\end{aligned}$$

- The idea of RWGEKS(Vancic et al., 2011) is to suppress the noise of each chain links by imposing chain identity. i.e. since $\pi_{t-1,t} = \pi_{l,t} - \pi_{l,t-1}$ for all l, RWGEKS chain link $\tilde{\pi}_{t-1,t}$ is defined as follow:

$$\tilde{\pi}_{t-1,t} = \frac{1}{|L|} \sum_{l \in L} \left(\hat{\pi}_{l,t} - \hat{\pi}_{l,t-1} \right).$$

Scanner Data

- Collected from 89 supermarkets in Korea from July 2017 to October 2022, the dataset covers an average of 80,000 items on a weekly basis.

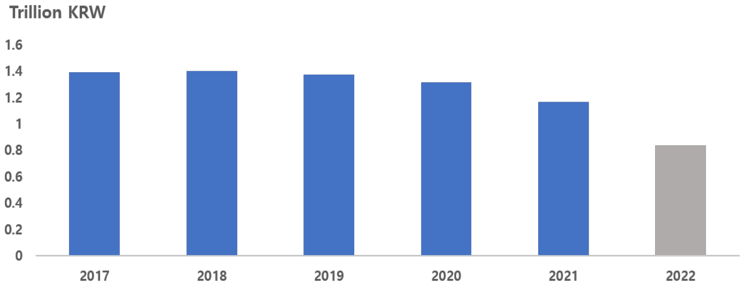


Figure: Total Sales Amount per Year

Comparison with Official CPI (1)

- The new index evolves similarly to the official Consumer Price Index (CPI).

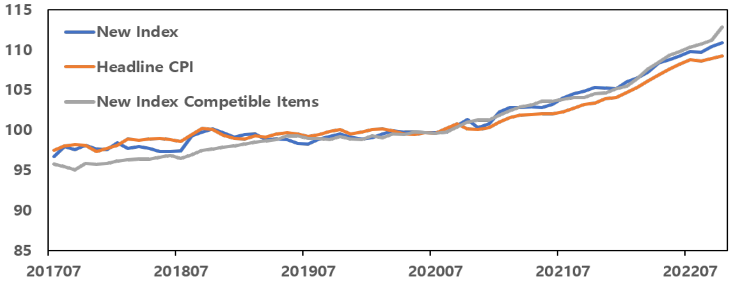


Figure: New Index vs Headline CPI (Level)

Comparison with Official CPI (2)

- Interestingly, the new index statistically more similar with headline CPI.

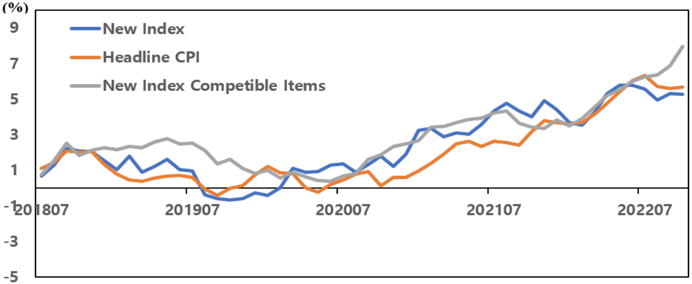


Figure: New Index vs Headline CPI (YoY)

Comparison with Official CPI (3)

- Statistically, new index lead the Core CPI significantly.

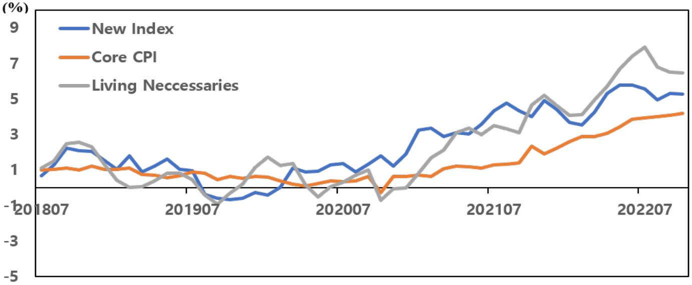


Figure: New Index vs Core Inflation and Living Necessaries (YoY)

Implication

- The fact that the new index evolve similar to and lead official CPI implies that granular data sets can help us to track and predict the official CPI.

- Remaining question is: Where do the differences originate from?

Implication (Cond.)

- The difference of monopolistic power or product heterogeneity could be a possible source of difference.

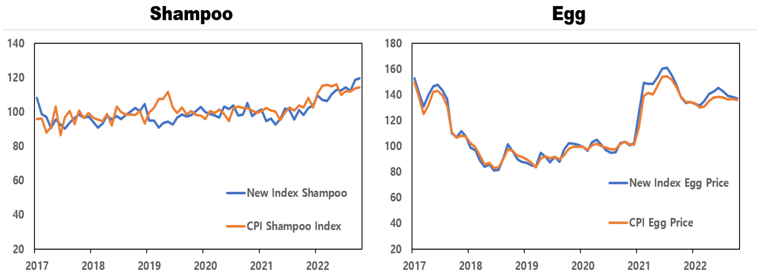


Figure: New Index and Official Core CPI (YoY)

- As the fixed base method does not take into account the change in consumption pattern, it could mislead the measurement of cost of living.

Future Work

- One promising direction of research is utility-based inflation measurement.
- Redding and Weinstein (2020) propose a method to measure the cost of living in an environment where consumer preferences change, assuming a CES utility function.
- In a similar context, Braun and Lein (2021) decompose the sources of bias in the fixed base method and find that the entry-exit of items and changes in preferences account for 2.6% of the bias.
- During the COVID-19 period, the degree of competition may have also changed.
- It is necessary to verify whether the change in market structure is a new source of the recently surging inflation.