



Research on Improving the CPI Statistical Survey System in the Information Age

Qian Wang *

Beijing Municipal Bureau of Statistics, Beijing, China – wangqian@bjstats.gov.cn

Yan Feng

Beijing Municipal Bureau of Statistics, Beijing, China – fengyan@bjstats.gov.cn

Min Zuo

NBS Beijing Survey Office in Beijing, Beijing, China – zuomin@bjstats.gov.cn

Zidong Zhao

Beijing Municipal Bureau of Statistics, Beijing, China – zhaozd@bjstats.gov.cn

Abstract

The development of information and technology has brought tremendous changes to consumer activities. The methodology and thinking adopted in consumer price statistical analysis must be adjusted to suit such changes by using new technologies and highlighting some new features. This research mainly focuses on collecting price data of commodities sold online, acquiring price information through scanned data and using big data as an effective approach in this process. This analysis compares the advantage of each approach by using the example of book prices and housing rental with a feasible survey scheme designed according to China's current economic and social conditions. In light of the technical difficulties we currently face in data collection, this research put forward some countermeasures and pointed out the direction of future research, in the hope that future consumer price index (CPI) statistical survey can become more improved and scientific with the help of new technologies in the information age.

Keywords: survey; price data collection; online shopping; big data.

1. Introduction

Over the past half century, the information revolution triggered a revolution in consumption which tremendously shifted people's consuming attitudes, created some new consuming patterns and means and dramatically increased the number of consumers. The advancement in information and technology has brought multiple impacts to consumer price index (CPI) statistical survey. On the one hand, the advent of the "Big Data Time" challenged the traditional sampling collection in price survey, the transformation from collecting only some parts of data to collecting as much as or even all the data has become possible; one the other hand, the diversified consuming pattern challenged the traditional methodology of selecting price survey points, and the increased proportion of online shopping entails that online consuming activities must also be included in CPI survey; moreover, electronic data storage challenged the traditional price collection pattern as obviously collecting data electronically is more direct and reliable than collecting data manually.

2. Section 2

(I) Preliminary exploring and using "Big Data" in the housing rental survey of Beijing

1. "Big Data" adopted as a new approach in the price survey of housing rental

Instead of following the traditional approach of sample collection from survey points, we selected 26 housing intermediary companies in Beijing as the new survey points. These intermediaries all have a relatively large business amount and we collect all their rental transaction records on a monthly basis, which reached an average of 12,000. In addition, instead of making strict quality comparison of the price data, we compared the price data of rental properties at the same quality level and location. The

new approach was adopted in 2013 and the rental price index we thereafter formulated is able to better reflect the rental price changes of Beijing in a more objective and accurate manner.

2. Analyzing data in great depth

We calculated the yearly rental transaction record of these survey points in 2012, and analyzed the influencing factors on rental price through the simple box plot. The conclusion we reached indicates that location is the largest influencing factor whereas decoration and floor number only pose an insignificant influence on rental price.

(II) Online shopping price included in the price survey system

1. Advantages of including online shopping price in the CPI survey system

First, the inclusion expanded the survey scope and made the index more representative. At present, books, magazines and home appliances are registered a huge online sales volume, and including online prices into the CPI survey can expand the CPI data source of the survey; second, low cost and high benefit. The current survey pattern mainly relies on human, who is costly in terms of both time and money, and the cost can be significantly dropped if the price collection process is replaced by the transaction information query function offered by various E-shopping sites.

2. Applicable range of online price collection

According to the CPI survey categories, some commodities such as food, and in particular fresh food, only register a small proportion of online sales and therefore the price collection for these commodities should follow the traditional method of on-site surveys. For the category of clothing which features high seasonal fluctuation and diverse fabrics and styles, the traditional on-site survey should be followed given also the fact that some brands are not sold online and there may be discrepancies between the clothing sold online and their online descriptions. The price information of books can be collected online given its huge proportion of online sales, clear definition of commodity nature and the unlikelihood of quality discrepancy between books sold online and on site. The price information of train tickets, air tickets and travel agency charges can also be collected online as the price is easily identified online and the ticket/service can be easily purchased online. The applicable range can be expanded to electronic products such as home appliances and communication devices after further study.

3. Method of price collection for books sold online

First, include new online bookstores as price survey points; second, designate bestsellers as survey samples. If a bestseller is re-published, the publisher can re-verify the price according to the cost. Bestsellers are sold with high volume at both physical bookstores and online stores and the price is very sensitive; third, conduct dynamic sample survey for bestsellers. Top bestsellers register a huge sales volume with significant market influences. We focused on the Top 10 bestsellers sold at the survey points (excluding special categories such as teaching assistance books and legal policy books, etc.) and we calculated the mixed average price; four, we processed the data using the standard mixed average price method, and obtained the price index of bestsellers. In order to remove the influence of structural price changes to the index and other non-comparable factors, we calculated the actual sales price of each printed sheet of the bestseller according to its actual price and number of printed sheets; we then calculated the mixed average price of the book price and actual price of each printed sheet, and value we obtained represents the price of the bestseller. Our result shows that the book price index obtained through the above method can more objectively reflect the price change trend of books, as compared to the existing calculation method.

(III) Collecting price information using scanned data

1. Advantages of the approach

In the *Consumer Price Index Manual: Theory and Practice*, a manual co-authored by the International Labor Organization, Eurostat and other organizations, the approach of using scanned data for CPI data collection was mentioned. Records of large amount, frequent electronic transaction can provide rich and detailed information about price. With more and more supermarkets and malls are using POS machines and other price scanning system, and given the fact in particular that large-sized retail corporations have mostly established their business database to collect the data of their retail outlets, it has now become more convenient to use scanned data in the collection of price information.

Compared to the traditional manual collection process, scanned data contains more timely information with high accuracy.

2. Applicable range of scanned data

The following conditions must be met if scanned data is used in CPI survey: the retailer uses POS machines and has a business database; the commodity has a large sales volume that is barely affected by different quality levels; and commodities sold by different sellers have the same quality. Therefore, the price survey for fresh vegetables, fruits, meat, eggs and other agricultural products, vehicles with small sales volume but large price difference amongst different categories, restaurant meals cannot use scanned data. Food, alcohol, daily necessities and home appliances sold at large-sized shopping malls and electronic malls can use scanned data for CPI survey.

3. The idea of using scanned data in consumer price survey

The ideal scenario for using scanned data in consumer price survey requires that the cashing machine has a data output port which uploads all the data to the database on a real time basis. But currently it is still unrealistic to use POS machines to transmit data. This research mainly delves into the possibility of acquiring scanned data through the business database. We designed three schemes using the example of home appliances in order to work out a reasonable plan according to actual market practice. Scheme 1: Designate no sample product but acquire the sales data of all commodities being sold under certain categories. Take television for example, we can obtain the sales record of all types of TVs from Gome, Dazhong and Suning and other retailers of electric appliance on a monthly basis, calculate the mixed average price and the month-on-month (MoM) index based on the mixed average price of the previous month. The MoM index can be viewed as the price index of the TV category.

Scheme 2: Designate a sample product under the basic categorization, and acquire all the sales record of the product in a month. Take television for example, designate several samples according to the existing survey system, and acquire the sales record of all samples on a monthly basis, calculate the average price of the samples and then calculate the CPI using the Laspeyres Index.

Scheme 3: Designate a sample group and obtain all the data of the sample group. Take television for example again, obtain the total sales volume and amount of the Top 10 TV models under each brand on a monthly basis, calculate the average sales price of each model and adopt the figure as the average price of the sample group, and then calculate the CPI using existing methods.

In general, each of the three schemes has its advantages and disadvantages. Scheme 1 is the most ideal scenario to adopt scanned data in the CPI survey as it features a large amount of comprehensive data, and the price change of certain television models upon the index can be ignored. But the most prominent problem is that it's barely feasible with the current technologies we have. The large amount of sales record also poses certain challenge in terms of data output, storage and processing. On the other hand, the index calculated based on Scheme 1 might be subject of structural influences, as sometimes price drop is not a result of pure price change, but the promotional activities of certain models. The price drop can stimulate the increase of sales and further decrease the overall average price. Of course this influence can be ignored when the amount of records is large enough, but it's very difficult to realize this given the current technologies.

Compared to Scheme 1, the data amount of Scheme 2 is significantly dropped, but this approach, though offering more accurate sample prices, is not conducive to making the sample more representative. And in real practice when designating product samples, it is very difficult to match the sample with the business database. As the business databases of retailers are separate from one another, retailers oftentimes independently set the commodity code which is used as the only identification code of the commodity in the business database. In the database of some retailers, the commodity code is set randomly without any patterns, thus it is very difficult to increase the survey efficiency through data programming, and manual help is required to identify each commodity code from the database. Moreover, the models of home appliances are frequently updated, and matching commodity codes with the database requires a huge amount of manual input, and is very low efficient as well.

Compared to Scheme 1 and 2, the advantage of Scheme 3 is quite obvious as it's very easy to put it in practice. First, the workload on the retailer's side is fairly light. The statistical data required by the scheme is exactly the set of data used by the retailer in daily sales analysis, and the survey object



pressure were not increased largely; Second, the retailer is not required to submit specific transaction



coding that features uniform data format, or if could adopt the bar coding system which is internationally accepted, then it would make it a lot easier to use scanned data in CPI survey.

4. Provide more hardware support

The amount of data is quickly increasing in the process of CPI survey, no matter it is online data or scanned data. This will require us to re-construct a data transmission network for data collecting and processing. We need to set up platforms, improve the configuration of equipments and provide greater technical support to the processing of large amount of data so as to make it possible to adopt modern technologies in the work of CPI survey.

5. Conclusions

In combination with today's rapid development of information technology, this project proposes a reform program for the current method and system of price statistics and revolutionizes the survey method of partial categories. First, it integrates the Big Data idea into the statistics of house rental, reform the existing methods for collecting prices and setting product specifications, improve the quality of rental data, and provide data support for further conducting research on the house rental market. Second, it includes online sales in statistics of book prices. Meanwhile, a complete set of book price statistics survey system has been developed through reforming the methods of setting product specifications and compiling indexes, causing the book price index to reflect the price change trends in a more realistic way. Third, it takes electrical appliances as example to explore how to use scanned data, innovatively proposes the concept of using specification product groups, and develops a reasonable and operational survey program in combination with the reality.

Because IT applications in the CPI statistics survey system is an ambitious field, the project only selects three aspects of applications feasible in practice to explore the idea of reform, and select a specific category of goods in each aspect for empirical research. However, specific researches in Big Data, online collecting prices, and scanned data and other fields have yet to be further standardized and refined. In the next step, we can launch an in-depth research on categories that have the criteria to conduct information reforms, like books and electrical appliances, develop detailed investigation programs, study data processing methods and index-compiling methods, and expect them to be applied in practical work as soon as possible. Meanwhile, with the extensive application of information technology, we can put forward feasible reform ideas in more fields of price statistics and surveys, thus constantly improving the household consumption price statistics system.

References

- [1] ILO, IMF, OECD, UNECE, Eurostat, and the World Bank. Consumer Price Index Manual: Theory and Practice. Geneva. International Labor Office. 2004.
- [2] City Department of National Statistics Bureau: China circulation consumer price work manual, 2010.
- [3] ILO et al: Consumer price index manual: theory and practice [M]. Beijing: China Financial & Economic Publishing House, 2008.
- [4] National Statistics Bureau of the People's Republic of China: Statistical statement system for circulation and consumer price (2012 annual statistics journal and 2012 annual periodical statement), 2012.
- [5] NSB Survey Office in Beijing, Beijing Municipal Bureau of Statistics: Statistical statement system for circulation and consumer price (2012 annual statistics journal and 2012 annual periodical statement) , 2012.
- [6] Bernard Baumohl. Interpreting economic indicators. China Renmin University Press, 2009
- [7] Chen Xin. Research on china's book pricing system. SDX Joint Publishing Company, 2011
- [8] Liu Lei. Book market and marketing strategy. Bianjizhiyou, 2003
- [9] Chen Xiangcheng, Qiao Han and Wen Suqing: CPI survey empowered by scanned data. Market Research, 2012.