Chapter 9

Producing CPI During Covid-19 Period in Türkiye

Furkan Metin[1]

İbrahim Demir[2]

Cem Baş[3]

Abstract

The first Covid-19 national lockdown was decelerated on 11 March 2020 in Türkiye. The full lockdown process has made it difficult to collect the data necessary for both household budget surveys and consumer price index in Türkiye. Nevertheless, these challenges of Covid-19 have forced national statistical offices to produce reliable, qualified and timely consumer price statistics through using alternative data sources such as scanner data and web scraping data. In this paper, experiences gained by the Turkish Statistical Institute to overcome the difficulties arise by Covid-19 have been shared. It is noteworthy that during the Covid-19 pandemic, Turkish Statistical Institute (TurkStat) successfully raised the proportion of prices essential for the Consumer Price Index (CPI) sourced from alternative data from 0% to approximately 49% within a span of just one to two years.

1. Introduction

Since December 2019, the spread of the SARS-Cov-2 virus has reached the level of a global epidemic. In this period, which is called the Covid-19 epidemic, almost all countries around the world have experienced processes called "full lockdown". Citizens, except for the limited number of sector employees who are essential for human life, especially health, safety and food, are prohibited from leaving their homes, except for compulsory situations.

1 Associated Professor, TurkStat Vice President, furkanmetin@gmail.com, ORCID –ID 0000-0001-5782-6581

2 Associated Professor, TurkStat Vice President, ibrahimdemir@tuik.gov.tr, ORCID –ID 0000-0002-2734-4116

3 Dr., TurkStat Expert, cembas@tuik.gov.tr, ORCID –ID 0000-0003-4488-2562

https://doi.org/10.58830/ozgur.pub518.c2131

With these restrictions, the spread of the SARS- Covid-2 virus, which caused the Covid-19 epidemic, among humans was tried to be reduced. Nevertheless, the full lockdown process has made it difficult to collect the data necessary for both household budget surveys and consumer price index (CPI) (Andersen, 2020).

During the Covid-19 pandemic, in order to protect public health, face-to-face survey method has been temporarily suspended in all household surveys of Turkish Statistical Institute (TurkStat). Therefore, the Household Budget Survey could not be completed for 2020. CPI weights for 2021 according to the methods given in the EUROSTAT guide; Expenditure shares (HBS data) for the years 2017, 2018, and 2019 have been updated using the 2020 National Accounts Household Final Consumption Expenditure growth rates (EUROSTAT, 2020). In following years EUROSTAT have also advised to use the latest National Accounts Household Final Consumption Expenditure growth rates (EUROSTAT, 2021-2022). This is due to the fact that consumption habits of households have changed during the period of Covid-19 pandemic (Baker et al, 2020).

Furthermore, some outlets were closed or working hours changed as a result of the restrictions of Covid-19; some services were not provided (Dunn et al, 2020). In this period, field studies continued as much as possible in order to produce the CPI in the best way possible (Destatis, 2022). When prices were unavailable due to restrictions, alternative data sources, such as scanner data and web scraping information, were recommended in EUROSTAT's guidance published in April 2020 (EUROSTAT, 2020). As a necessary result, utilizing new alternative digital data sources, such as scanner data and web scraping data, became crucial for statistical offices in calculating the Consumer Price Index (CPI).

Accordingly, TurkStat implemented several measures to address the challenges posed by Covid-19. TurkStat makes great efforts to minimize the impact of Covid-19 measures on CPI quality and takes the necessary measures in a timely manner. Difficulties in data collection, especially during quarantine periods, accelerated the studies on the use of alternative data sources in TurkStat. Within a year, scanner data and web scraping data were incorporated into the production of the CPI by TurkStat.

2. Scanner data

Scanner data, as one of the Big Data sources, has become reasonable popular in national statistical systems to improve the accuracy of CPI. Scanner data is "high in volume and contains information about individual

transactions or summaries, date, quantities and values of products sold, and product descriptions". Since scanner data is an enhanced data sources for national statistical organizations (NSOs) that can be used to enhance their statistical products and improved, more efficient ways of working by reducing provider burden and also reducing cost of physically collecting data (International Monetary Fund, 2020).

While scanner data sets provide several positive outcomes, there are also several challenges that need to be overcome by NSOs to compile the CPI. Some of these challenges were outlined as follows,

- establishing a sufficient sustainable communication channels among the inside and outside partners,
- developing an IT system to asses and prepare data
    - o such as controlling the scope of items included in the data sets (unit of measurements, date, quantities, price values etc.)
    - o adaptation of the classifying systems of each retailer to classifications of individual consumption (COICOP) quality assurance of the scanner data sets (International Monetary Fund, 2020).

3. The Challenges Faced by the Turkstat During Covid-19 Pandemic

3.1. Obtaining Scanner Data Sets

In fact, obtaining comprehensive data sets from different retailers is require expertise and skills in many different fields like communication, statistics, information technologies and law. Effective planning and coordination of all processes involving multiple units within the National Statistical Office (NSO) is essential (UnECE, 2021a).

Scanner data can be compiled directly from companies or through intermediary institutions that keep invoice information electronically. In the TurkStat version, data is compiled directly from the companies. The advantages of obtaining data directly from the company are explained in the Consumer Price Index Handbook published by the International Monetary Fund (IMF) as follows:

- Easier negotiations about data (Experiences in the Netherlands, Switzerland and Australia suggest these negotiations will take approximately six months to complete;
- The supply of the data set at no (or minimal) cost

- The scope of items included in the data set
- The level of item aggregation to ensure homogenous information (UnECE,2021b)
- An agreed timetable for the supply of the data set to meet CPI processing requirements
- A contact officer within the retail business who is familiar with the data set to answer NSO data queries (International Monetary Fund, 2020).

A key determinant of share prices and profitability within the sector is the pricing strategy employed by companies. Product features and prices, linked to unique barcode identifiers, are typically available to consumers, and are not regarded as confidential information. Nevertheless, when this pricing data is aggregated from multiple companies, the pricing strategies of these firms become apparent. Additionally, confidential information, or trade secrets, plays a significant role in determining the competitive strength of firms. Thus, safeguarding trade secrets is essential for businesses to maintain their competitive edge (Karataş, 2017).

For these reasons, although barcode scanning data is obtained from firms and utilized by national statistical offices, aggregated data pertaining to individual companies is not disclosed. This aggregated data is used solely for statistical purposes, specifically for the production of the Consumer Price Index (CPI). Four important benefits of using scanner data instead of sampling data have been summarized by the American Bureau of Labor Statistics as follows:

- the ability to account for product and outlet substitutions in some categories of consumer spending,
- the ability to produce indexes with greatly reduced sampling error for some item strata,
- the ability to identify new products and quality enhancements,
- and the ability to ensure a more accurate probability sampling of items with fewer manual steps, a sampling procedure known as "disaggregation" (Bradley et al.,1997).

The TurkStat takes the following steps while compiling the scanner data:

i. **Examination of legal terms:** Statistics Law of Türkiye Article 7 enacted in 2005, authorized The TurkStat to compile the information serve the needs of Institute. Based on this article, The TurkStat has

the right to request all data to be used in the production of official statistics from all units.

ii. **Clarifying data characteristics:** It is very significant to complete internal processes before requesting data from companies. It should be determined how often and what kind of data is needed in order to decrease response burden. The data pattern should be clarified by determining all units within the Institute that need to use the data. During the scanner data project conducted by TurkStat, it was identified that two departments were the main users of this data source. These departments were both the Price Statistics Department, responsible for calculating the CPI, and the National Accounts Department, which monitors the gross domestic product calculations.

iii. **Selection of companies:** When compiling scanner data, ensuring data flow from all companies operating in the country will be very costly, considering the transfer, processing and storage of the data. Therefore, a small subset of nationwide retailers was selected to begin the scanner data Project. In recent years, the TurkStat was entitled to access to detailed information about the companies through the protocols with various public institutions, in the context of efforts to increase the use of administrative records. This allowed, companies that are predominant in their own sectors for certain variables (turnover, added value, etc.) were determined easily.

iv. **Contacting the companies:** The TurkStat has widespread organizational structure, the expert staff and advanced information technology applications on collecting data. In this way, the Institute communicates with respondents from all over the country quickly and effectively. the TurkStat produces meaningful statistics from compiled data by the authorization of Statistical Law of Türkiye and describes to the respondents how they can take advantage of these statistics that they contributed. In this regard, summary outcomes were planned to be prepared in order to share them merely with the nationwide retailers themselves.

v. **Planning and Performing Data Transfers:** For data transfer, the most appropriate method for the company and the TurkStat infrastructures is determined by negotiating with information technology units of the companies. While determining the appropriate method, attention is paid to protect data privacy and prevent data loss. A platform (FTP, SFTP, etc.) is decided for

transferring of compiled scanner data, and provide data transfer to Institute through this platform. The data is prepared for use by the relevant units through ETL (Extract, Transform, Load) processes conducted following its transfer.

vi. **Data Editing:** To prepare the transferred data from companies to TurkStat for analysis, potential inconsistencies and data entry errors are identified and reported. These reports include assigning multiple barcodes, definitions or amount to a product, and assigning incorrect amount to the products. As a result of the reports, some imputations such as correcting the number of products, and excluding from relevant population the incorrect barcodes and products are made. After reporting and imputation stages, product codes are assigned to each product so that products can be used in calculations.

vii. **Data Analyze:** In the analysis phase of code assigned and error free data set, the unit prices that formed on the basis of barcodes and products, and monthly turnover of each company and distribution of this turnover by products are calculated. The unit prices in the scanner data are compared with the unit prices in the Consumer Price Index, and price differences are reported.

3.2. Challenges to Compile Scanner Data

As stated above, the difficulties experienced in transferring this data which is of great commercial importance for companies, and the measures taken can be listed as follows:

- **Exhaustive data:** By defining the selection criteria for both sectors and companies, collaboration has begun with the limited number of identified nationwide retailers.
- **Reluctance of companies:** An effective communication strategy is to implement, minimizing the response burden to the lowest possible level, while promptly addressing the needs of the companies. Statistics producers who support and encourage a customer-friendly culture that is mindful of respondent concerns are more likely to be rewarded with improved response rates and careful, accurate responses from respondents (Karataş, 2017).
- **Difficulty in compiling requested data:** Keeping data in different databases, franchises data, etc. are some of the most common problems. Companies should be given sufficient time to complete

these works and technical support should also be provided when it is necessary.

- **Ensuring data privacy:** Privacy provisions that located in The Statistics Law of Türkiye are followed with precision, access authorizations to received data are restricted.
- **Failure to transfer data within specified periods**: In order to accelerate the processes protocols were prepared for each company and submitted to the companies before the ultimate data transfer. Additionally, an e-mail notification system has been developed to ensure scanner data transfers within the specified periods. In the case of exceeding the deadline of data transfer, the e-mail notifications system automatically sends notifications to the people who are responsible for data transfer in both TurkStat and companies.
- **Inadequate product classification of companies:** Assignment studies of the classification code (E-COICOP) are continued. Codes are assigned manually, by programming or by machine learning techniques (ONS,2021).

  In the year 2020, the analysis processes of the data obtained from a total number of seven national chain stores (markets), which have a significant share in the retail trade sector, have been completed. Various studies (classification, code matching, etc.) have been carried out to make the data ready for use for CPI calculations. Barcode sales data obtained from chain markets started to be used in 2021 CPI calculations. In this way, prices and quantities covering the whole month were included in the index by using daily sales data on the basis of branches obtained from seven national chain companies instead of prices compiled from the field on certain days of the month. As a result, 21 per cent of the prices for CPI were gathered by scanner data sources and used in CPI calculations in 2021.
- **Local chain market data:** In the 2022 CPI study, sales data obtained from a total of 42 companies are included in the index calculation. In this context, the prices of food, glassware and personal care products from 9 chain markets and 26 local markets in the market sector are followed. Currently, approximately 84.000 item numbers are compiled from 9 chain markets and 8.000 item numbers from 26 local markets.

  Nevertheless, the number of item numbers obtained from 26 local markets corresponds to approximately one-tenth of the number of item numbers obtained from 9 national markets. Therefore, covering

26 local markets has significantly increased our communication cost. In fact, national markets are more likely to have their own information technologies (IT) units. Therefore, scanner data transfers from the national markets to TurkStat have been completed usually in a timely manner. On the other hand, most of the time, problems were encountered in terms of timeliness in data transfers made from local markets., which has shortened the CPI analysis periods before the CPI bulletin (ONS,2021).

A further problem regarding to scanner data of local markets has been that there are a few chain stores in local markets and most of them are in just one city.

- In order for the barcode numbers in the sales data to be used in the index calculation, first of all, each product is assigned to the appropriate COICOP codes in accordance with the item basket and detailed description. Nevertheless, considering the number of local markets and the number of items tracked from local markets in COICOP assignment processes, workload weight occurs.

  The problem of product continuity in local markets increases the need for substitution and COICOP codes must be assigned quickly and accurately. Apart from these, significant errors such as quantity errors are frequently detected in local market sales data. In addition, the fact that the data transfer requested from the local markets covering 1-7 and 1-14 periods of each month is not done regularly and the monthly final sales data requested to be sent on the 26th of the month are delayed then it makes difficult for the CPI calculations to reach the bulletin date. In the year 2022, 44.3% of the prices have been obtained directly by this method (Please see the Figure 1).

3.3. Web scrapping data

The prices compiled over the internet for the prices of household appliances (refrigerator, washing machine, dishwasher, etc.), electronic products, clothing, furniture, brand new cars and bus tickets, which are suitable for pricing using the web scraping method in terms of item definitions, were used in the CPI calculations in 2022. The number of prices to be obtained by this method constitutes approximately 5.3% of the total number of prices to be compiled (Please see the Figure 1).

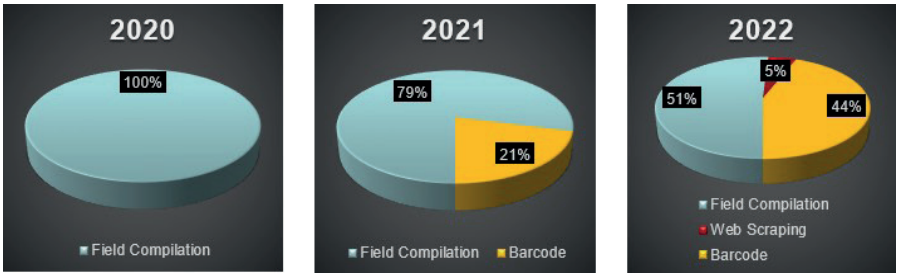

Figure 1. The significant change in the share of CPI data sources during Covid-19

4. Conclusion

It can be considered a significant achievement that Turkish Statistical Institute (TurkStat) increased the proportion of prices necessary for the Consumer Price Index (CPI) collected from alternative data sources from 0% to nearly 49% in a few years during the Covid-19 pandemic (Please see the Figure 1).

The scope of this study, which has initially commenced with food and clothing retailers, is set to expand in the coming years to encompass telecommunications companies, insurance firms, entities in the healthcare sector, airline companies, and organizations operating within the energy market. Following an evaluation of product information obtained from approximately 90 websites, this data will be utilized for the production of official statistics (TurkStat, 2023).

Using information from all processed products, rather than just from the selected items, will enhance the quality of the generated statistics, mitigate errors stemming from pollster or tagging mistakes, facilitate advanced analyses based on price and quantity, and simplify the selection of substitute products. Although engaging with companies during the establishment of the system, implementing IT and coding systems, and fostering a working culture centred around big data may increase the burden on both TurkStat and the companies involved, it has been observed that, in the medium term, this approach leads to higher-quality statistics with reduced costs and lower response burdens.

References

Andersen, A.L., Hansen, E.T., Johannesen, N. and Sheridan, A. (2022), Consumer responses to the COVID-19 crisis: evidence from bank account transaction data*. Scand. J. of Economics, 124: 905-929. https://doi.org/10.1111/sjoe.12512

Scott R Baker, Robert A Farrokhnia, Steffen Meyer, Michaela Pagel, Constantine Yannelis, How Does Household Spending Respond to an Epidemic? Consumption during the 2020 COVID-19 Pandemic, The Review of Asset Pricing Studies, Volume 10, Issue 4, December 2020, 834–862. https://doi.org/10.1093/rapstu/raaa009

Bradley, R., Cook, B.A., Leaver, S.G., & Moulton, B.R. (1997). An Overview of Research on Potential Uses of Scanner Data in the US CPI. https://stats.unece.org/ottawagroup/download/f46.pdf

Dunn AC, Hood KK, Driessen A. Measuring the effects of the COVID-19 pandemic on consumer spending using card transaction data. BEA Working Paper Series. 2020;5. https://www. bea.gov/system/files/papers/BEA-WP2020-5_0.pdf.

Destatis (2022). Impact of the corona crisis on the CPI/HICP price collection. https://www.destatis.de/EN/Themes/Economy/Prices/Consumer-Price-Index/Methods/corona-cpi-hicp.pdf?__blob=publicationFile

EUROSTAT (2020). Guidance on the Compilation of the HICP in the Context of the Covid-19 Crisis. https://ec.europa.eu/eurostat/documents/10186/10693286/HICP_guidance.pdf

EUROSTAT (2021). Derivation of HICP Weights for 2022. https://ec.europa.eu/eurostat/documents/10186/10693286/Derivation-of-HICP-weights-for-2022.pdf/c47503af-e526-0e40-def2-370c501bb595?t=1638266939151

EUROSTAT (2022). Derivation of HICP Weights for 2023. https://ec.europa.eu/eurostat/documents/272892/11336726/Derivation+of+HICP+weights+for+2023.pdf/

International Monetary Fund. (2020). Consumer Price Index Manual-Concepts and Methods, pp.365

Karataş D.N. (2017). Tax Confidentiality within the Context of Trade Secret and the Protection of Personal Data, pp.39, 383.

ONS (2021). Classification of new data in UK consumer price statistics-https://www.ons.gov.uk/economy/inflationandpriceindices/articles/classificationofnewdatainukconsumerpricestatistics/2021-04-06

TurkStat (2023). Consumer Price Index February. https://data.tuik.gov.tr/Bulten/Index?p=Consumer-Price-Index-February-2023-49656

UnECE (2021a). Guide on producing CPI under lockdown. https://unece.org/sites/default/files/2021-12/Guide%20on%20producing%20CPI%20under%20lockdown%20WEB%20version.pdf

UnECE (2021b). Index Compilation Techniques for Scanner Data: An Overview. https://unece.org/sites/default/files/2021-05/Session_1_Eurostat_Paper.pdf