United Nations ECE/CES/GE.22/2023/2

Economic and Social Council

Distr.: General
6 July 2023

English only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Consumer Price Indices

Sixteenth session
Geneva, 7-9 June 2023

Report the sixteenth session of the Group of Experts on Consumer Price Indices

Summary

The present document is the report of the meeting of the Group of Experts on Consumer Price Indices on 7–9 June 2023. This report is provided to inform the Conference of European Statisticians of the organization and outcomes of the meeting.

Introduction

1. The meeting was organised by the United Nations Economic Commission for Europe (UNECE) in cooperation with the International Labour Organization (ILO). The agenda of the meeting was prepared by the UNECE Steering Group on Consumer Price Indices.

2. The meeting was attended by representatives from Albania, Angola, Armenia, Australia, Austria, Azerbaijan, Bangladesh, Belarus, Belgium, Belize, Bosnia and Herzegovina, Botswana, Bulgaria, Burundi, Canada, China, Comoros, Denmark, Egypt, Estonia, Ethiopia, Finland, France, Georgia, Germany, India, Indonesia, Iran, Italy, Japan, Kazakhstan, Kuwait, Kyrgyzstan, Latvia, Lesotho, Lithuania, Luxembourg, Mexico, Moldova, Morocco, Mozambique, Namibia, Netherlands, Norway, Philippines, Poland, Singapore, South Africa, Spain, Switzerland, Tajikistan, Thailand, Tunisia, Türkiye, Turkmenistan, Uganda, United Kingdom of Great Britain and Northern Ireland, United States of America and Uzbekistan.

3. The meeting was also attended by representatives from the following organizations: East African Community (EAC), European Central Bank (ECB), European Commission (Eurostat), International Monetary Fund (IMF), International Labour Organization (ILO), Interstate Statistical Committee of the Commonwealth of Independent States (CIS-STAT), Organisation for Economic Co-operation and Development (OECD), Economic Commission for Latin America and the Caribbean (ECLAC), United Nations Economic and Social Commission for Western Asia (UNESCWA), United Nations Statistics Division (UNSD), World Bank, Hitotsubashi University, University of British Columbia, University of Tokyo, University of Tuscia and several independent experts.

4. The meeting was chaired by Christopher Jenkins, Office for National Statistics, United Kingdom.

II. Organization of the meeting

5. The meeting was organised in the following sessions:

(a) Session 1a. Scanner data

(b) Session 1b. Scanner data

(c) Session 2a. Other data sources and machine learning

(d) Session 2b. Other data sources and machine learning

(e) Workshop on Scanner data for CPI

(f) Poster session

(g) Session 3. Implementation of COICOP 2018

(h) Session 4. CPI for groups of households and measurement of well-being

(i) Session 5. Expenditure weights in the CPI

(j) Session 6. Quality adjustments

(k) Session 7. Modernisation of CPI production

(l) Session 8. Other CPI issues – owner-occupied housing

(m) Conclusion and proposals for future work.

6. The conclusion of the meeting and proposals for future work are provided in section III. The main outcomes of the sessions are summarised in section IV.

7. The proceedings of the meeting are available on the UNECE website: https://unece.org/info/Statistics/events/348376

III. Conclusion and proposals for future work

8. The participants of the meeting recommended the next meeting of the Expert Group on CPI be organised to take place in 2025 in Geneva.

9. The following topics were suggested to be included on the agenda of the meeting:

1) Alternative data sources - scanner data, web scraping and administrative data
2) Use of multiple data sources, collection modes and calculations methods; their integration and conceptual implications
3) Modernising production systems for the CPI
4) Communication and transparency; ensuring public trust in the CPI
5) Household cost indices and cost of living indices
6) The treatment of informal markets, quality changes and discounts
7) Housing – rented and owner-occupied dwellings
8) COICOP 2018

10. The participants of the meeting encouraged for timely release of the final version of COICOP 2018 and relevant correspondence tables to support the planning and the implementation of COICOP 2018 in countries.

IV. Summary of sessions

A. Session 1a. Scanner data

11. The session was chaired by Michael Hardie (Office for National Statistics, United Kingdom). The session was based on presentations by Adrien Montbroussous (INSEE, France), Adam Tardos (Statistics Austria) and Ken Van Loon (National Statistical Institute, Belgium).

12. Different versions of GEKS Törnqvist indices applied for products at the COICOP 6-digit level were presented (INSEE). The At that level they behave similarly while at a more aggregated level there is more volatility. Further research is required to understand this.

13. The choice of window length when applying multilateral indices for scanner is crucial. There is a trade-off between using a longer time window, which is recommended in the literature, and the practical challenges of collecting sufficient data before implementation. Experiences from Austria showed that annual inflation rates derived from 25 month and 13 month window indices do not differ significantly, but the annual inflation of the lower COICOP level categories is more often higher for a 25 month window. In case of seasonal products, differences are larger with the longer window length producing higher inflation rates. The differences may be small enough to implement a 13 month window length and therefore saving time and resources.

14. Product relaunches where products get a new unique identifier may lead to ‘shrinkflation’ when relaunches are not captured in the index calculation. Research by the National Institute of Statistics of Belgium showed that multilateral methods which use hedonics can be applied to supermarket scanner data and capture some product relaunches and reduce shrinkflation.

15. The session presented innovate research and insights on scanner data and multilateral indices to tackle methodological and practical challenges in using scanner. It was recommended to continue sharing methodological and practical experiences among countries to facilitate further work and development.

B. Session 1b. Scanner data

16. The session was chaired by Claude Lamboray (Statistics Luxembourg). The session was based on presentations by Amelie Blasius (DESTATIS, Germany), Alessandro Brunetti (ISTAT, Italy) and Tiziana Laureti (University of Tuscia, Italy), Tanya Flower (Office for National Statistics, United Kingdom), Paul Konijn (Eurostat) and Brendan Williams (Bureau of Labor Statistics, United States).

17. The session discussed practical challenges and price index methods suitable for scanner data. Applications were presented for a variety of product categories, including package holidays, grocery products and vehicles. The session focused on multilateral methods and hedonic approaches. The application of these more advanced methods requires extensive testing. A wide range of experimental results were presented during the session. In addition, an update was given on the work of the UN Task Team on Scanner Data whose aim is to develop guidance, code and training material on the use of new data sources for official price statistics. The following comments and conclusions were made during the session:

18. With scanner data, products can be classified in different ways, and this can have an impact on the resulting price indices and the treatment of relaunches.

19. The index methods must cope with both product life cycles and quality change. There is a need to further study and develop methods that treat quality differences in the context of scanner data.

20. Various technical decisions must be made with multilateral methods, for example on window length. It is not always straightforward on which basis these decisions should be made and there is a lack of established best practices.

21. Further work is required on data integration, as multiple data sources and possibly different compilation methods must be combined and incorporated into the CPI.

22. Communication with users is becoming more challenging because of the increased methodological complexity. Further attention must be given to communication and transparency.

C. Session 2a. Other data sources and machine learning

23. The session was chaired by Helen Sands (Office for National Statistics, United Kingdom). It was based on presentations by Lucien May and Botir Radjabov (Statistics Luxembourg), Camilla Rochlenge (Statistics Norway), Tiziana Laureti (University of Tuscia), and Luigi Palumbo (Bank of Italy).

24. The session was dedicated to the use of other (non-scanner) data sources that can be used for the compilation of consumer price statistics. All three presentations recognized innovative uses of alternative data for CPI construction and dissemination. The following comments and conclusions were made:

25. The use of unweighted multilateral methods (in the form of a GEKS-Jevons index) may be suitable for web scraped data when considering products of stable assortment, e.g., with low market churn and stable pricing lifecycles (Statistics Luxembourg). More complex methods may be needed when considering less stable products.

26. Before starting web scraping it is worth investigating if it is possible to use Application Programming Interface (API) that provides access to more stable data.

27. Covering the sharing economy may benefit from the use of alternative data sources. For instance, tax registers may be available. While there may be challenges in using tax registers data it may prove helpful given that companies that have appeared to be using a peer-peer model are now more often using a business-consumer model and therefore should be in the scope of a CPI (Statistics Norway). With the growing importance of the sharing economy, its coverage and measurement should be considered further.

28. Web scraped data can be used to assess the geographical coverage of consumer price statistics (University of Tuscia and Bank of Italy). It may also be considered if and how web scraped data could be used to inform the sample of locations used to construct the CPI.

29. The session highlighted innovative new uses of alternative data sources and continued challenges in the measurement of consumer price indices despite having access to new means of data collection – due to dynamic and emerging economies and challenges with ensuring representativity of data.

D. Session 2b. Other data sources and machine learning

30. The session was chaired by Federico Polidoro, (World Bank). The session was based on presentations by Sameer Nawaz and Fred Hanmer (Australian Bureau of Statistics), Seitaro Tanimich (Statistics Bureau of Japan), Serge Goussev (Statistics Canada), Mostafy Farrokhfal (Statistical Centre of Iran), Mario Spina (Office for National Statistics, United Kingdom) and Divina Gracia Del Prado (Philippine Statistics Authority). The following comments and conclusions were made:

31. The utilisation of new data sources has the potential to increase the coverage and quality of the CPI, make the CPI production more efficient and reduce response burden.

32. Administrative data provides an important alternative data source for CPI, and can be used to replace or complement survey data, e.g. for rentals of housing, There may be issues with getting access to administrative data and obtaining the data in a suitable format. Countries are encouraged to reach out to owners of administrative data and investigate the possibilities to get access.

33. Web scraping comes with challenges on a range of methodological and technical issues and operational risks. For instance, access to web pages may be blocked and there may be legal and ethical issues related to the scraping of some websites. The use of API as a more stable alternative to web scraping was mentioned.

34. New data sources raise issues regarding the classification of individual observations, the use of machine learning and outlier detection that assume a different nature given the bulk-data characteristics of the new sources.

35. Integration of multiple sources and multiple collection modes in the CPI production emerged as a major challenge for statistical offices. This includes process design and sampling issues, data treatment, multiple index calculation methods, aggregation (including a possible combination of bilateral and multilateral indices), weights estimation and weights of retail trade distributional channels. It should also be considered whether or to what extent a multi-source/multi-mode approach may impact the interpretation of the CPI.

E. Workshop on scanner data for CPI

36. The workshop was conducted by Tanya Flower (Office for National Statistics, United Kingdom), Kristiina Nieminen (Statistics Finland), Serge Goussev (Statistics Canada), Jacek Białek (University of Lodz and Statistics Poland). Kjersti Nyborg Hov (Statistics Norway), Federico Polidoro (World Bank) and Jens Mehrhoff (IMF). The workshop was organised to present the work of the UN Task Team on Scanner Data and solicit feedback and comments from participants.

37. The first session of the workshop summarised the forthcoming training module on the acquisition of scanner data for CPI to be published on the UN Learning Hub.[1] The course will cover the end-to-end process of data acquisition, including how to negotiate with data suppliers, what specifications to ask for and basic checks on data quality.

[1] https://learning.officialstatistics.org/

38. The second session provided an overview of the open source PriceIndices package and how it can be used for analysis and research on new data sources. Participants suggested to create notebooks on the UN Global Platform to provide supplementary examples of how the PriceIndices package can be used for analysis and to tie into additional training material when developed.

39. The third and fourth sessions discussed the classification of alternative data for CPIs and provided a first look at some of the content the Task Team classification workstream is developing. The workstream goal is to centralise guidance on available methods and promote best practices, ranging from simple attribute-based to more advanced machine learning.

40. The workshop and the work of the UN Task Team was appreciated. It will be important to ensure that the material that has been developed or is being developed will be made available to countries and feed into further international work and developments on the utilisation of new data sources for price statistics.

F. Poster session

41. The session included presentations of country experiences and good practices in CPI compilation. The following countries and organisations provided poster presentations: Jefte Ochaeta and Melvin Perez (Belize), Ben Hillman (United Kingdom), Rejoyce Mbalakelwa (Botswana), Alemayehu Teferi (Ethiopia), Kabeli Mefane (Lesotho), Narong Kongsung and Tanaporn Sriklay (Thailand), Jens Mehrhoff (IMF), Vincent N. Musoke (East African Community), Kjersti Nyborg Hov (Norway), Zhakypbekuly Kuanyshbek (Kazakhstan), Filipe Amaral Jose Amone (Mozambique), Mohiuddin Ahmed (Bangladesh), Edgar Niyimpa (Uganda) Rafael Gaona López (Mexico), Hamidou Ounais (Comoros), Antonietta D'Amore (Italy), Wilhelmina Nangobe (Namibia) and Nami Takahashi (Japan).

42. The presentations covered a variety of issues, including implementation of COICOP 2018; scanner data; expenditure weights and the impacts of the pandemic on these; use of alternative data sources; web scraping; and modernisation of CPI production.

G. Session 3. Implementation of COICOP 2018

43. The session was chaired by Valentina Stoevska (ILO). The session was based on presentations by Ernestina Perez (UNECLAC) and Paul Konijn (Eurostat).

44. The Classification of Individual Consumption According to Purpose (COICOP) is the international reference classification of household expenditure. It is widely used in national accounts, household budget surveys, CPIs, and PPPs. The most recent update, COICOP 2018, was endorsed as an international statistical standard by the UN Statistical Commission at its forty-ninth session in 2018.

45. Some countries, mainly developing ones, have already implemented COICOP 2018, either by mapping the CPI basked based on COICOP 1999 to the new one, or by first implementing the new COICOP in the household budget survey.

46. Due to the changes at the division level and the introduction of an extra level of detail, the implementation of COICOP 2018 in CPI is facing a number of challenges, including:

- Price collection for new goods and services
- Linking of COICOP 1999 and COICOP 2018 time series and backward calculations
- Deriving weights for some items at a 4-digit level appears difficult (e.g., for preventive care services and other types of outpatient services)
- Treatment of bundled items that are now split into goods and services (e.g., delivery fees for online purchases)
- Representativeness of estimates at the lowest level of classification

47. Also, it was indicated that there seem to be some inconsistencies in the explanatory notes that need corrections and clarifications.

48. The implementation of COICOP 2018 requires careful and timely planning and coordination to address operational, IT, and methodological challenges. The meeting encouraged the final version to be released as soon as possible, together with correspondence tables with COICOP 1999 and CPC.

H. Session 4. CPI for household groups and measurement of well-being

49. The session was chaired by Christopher Jenkins (Office for National Statistics, United Kingdom). The session was based on presentations by Alessandro Brunetti (ISTAT, Italy), Anya Stockburger (Bureau of Labor Statistics, United States), and John Astin and Jill Leyland (independent experts).

50. Headline CPIs are usually designed to provide an aggregate measure of the average price development of consumer goods and services, based on a set of (lagged) expenditure weights that are updated annually or with less frequency. However, households are likely to experience different rates of inflation depending on their consumption patterns and will tend to substitute products in times of high inflation and changes in relative prices. Hence, over the previous years, many countries have experienced growing user demands for CPIs for specific household groups, accentuated by the Covid-19 lockdown period. The following comments and conclusions were made:

51. Traditionally, CPIs for household groups have been compiled by simply adjusting the weights of the headline CPI to reflect different inflation across income groups and households. However, for developing more suitable measures it will be necessary first to clarify the conceptual ground and the intended purpose of an index for specific household groups, e.i. should it target payments or an ‘out of pocket’ price index? An important aspect in this regard is the coverage of such an index. For instance, a household cost index (HCI) could be established that covers, e.g., capital costs and (mortgage) interest payments, that are not typically included in the headline CPI.

52. Issues about data sources and calculation methods were also raised, including, e.g., how to measure households’ expenditures, the use of alternative data sources, use of expenditure-based (plutocratic) or household-based (democratic) weights.

53. Dissemination of CPIs for household groups requires careful planning and communication and explanations that support correct interpretation and use of the indices and avoid/reduce the risk of confusion with the headline CPI. A further discussion centered on the dissemination of many different price indices and how NSOs need to manage the amount, and explanation of differences, in indices being published.

54. In conclusion, the session demonstrated the wide and important uses of CPIs, as well as the efforts of NSOs to address user needs. Given the importance of this topic and the evolving approach seen across countries, further discussion would be needed at future meetings.

I. Session 5. Expenditure weights in the CPI

55. The session was chaired by Jarmila Botev (OECD). It was based on presentations by Patrick Kelly (Statistics South Africa), Fathia Utami Afdi (Statistics Indonesia), and Anya Stockburger (Bureau of Labor Statistics, United States). The session discussed ways in which household budget survey (HBS) data can be complemented either by alternative surveys or national accounts data to update CPI weights more frequently, as well as insights from using more timely weights on CPI developments during the Covid pandemic. The following comments and conclusions were made:

56. Common issues faced when using complementary data sources to update CPI weights are that while the alternative sources use the same (COICOP) classification of expenditures, their granularity differs and the expenditures need to be mapped to existing HBS weights or imputed from other sources. Additional issues are different time and/or geographical and/or population coverage. Nevertheless, alternative data sources can help to improve CPI

expenditure weighs. Alternative data sources also provided important insights into changes in consumer expenditures during the pandemic.

57. Using more up-to-date weights showed shifts in consumer behaviour during the pandemic, notably in categories most affected by pandemic measures, like restaurants and hotels or transport. The differences between the CPI measures derived with timelier weights relative to the standard CPI measure were, however, much smaller. This was due to atypical (negative) consumer substitution of some items, as a result of the policy measures and/or supply bottlenecks, e.g. in personal computers, new vehicles, or owners' equivalent rent (as shown in the Bureau of Labor Statistics presentation).

58. The discussion highlighted that in case several CPI measures with different weights are published (such as in the US), they need to be accompanied by good communication. NSOs, as well as researchers, should carefully consider and test alternative data sources and should use a combination of them where possible (e.g. complemented by administrative data), as they may not have comprehensive coverage and hence may not be representative.

J. Session 6. Quality adjustments

59. The session was chaired by Corinne Becker Vermeulen (Swiss Federal Statistical Office). It was based on presentations by Satu Montonen (Statistics Finland), Sahoko Furuta (Bank of Japan), and Chihiro Shimizu (Hitotsubashi University). The following comments and conclusions were raised during the session.

60. Adjusting for quality changes remains a key difficulty in producing price statistics, for instance for clothing, technological products, products with a short life cycle, and, e.g., new and second-hand cars. Hedonic methods are often used to adjust for quality changes so that only 'pure' price changes are included in the CPI.

61. The presentations provided examples of hedonic modeling and comparison with indices without adjustments for quality change. Statistics Finland has combined hedonic quality adjusting and traditional index calculation for the second-hand cars price index; Bank of Japan has introduced a new estimation method using "sparse estimation" to overcome multicollinearity and omitted variable bias and has tested it on cars; Diewert and Shimizu have used the full potential of scanner data to derive different hedonic models for laptops and have compared the results with traditional method using quantity or expenditure weights in the compilation.

62. Hedonic methods are suitable for quality adjustments for some product groups, e.g., laptops or cars. However, developing them requires a large amount of data, good modeling skills, and time for maintaining and updating the models. Hundreds of sub-indices are produced, so it is not possible to use these methods on a large scale. Research in this area must continue, but it is also necessary to look at other ways of adjusting for quality changes when data needed for modeling are not available or when resources do not allow it.

K. Session 7. Modernisation of CPI production

63. The session was chaired by Giorgi Tetrauli (National Statistics Office of Georgia, Geostat). It was based on presentations by Rohan Draper (Statistics Denmark) and Angela Hernandez Santacoloma (Swiss Federal Statistical Office), Serge Goussev (Statistics Canada), Jefte Ochaeta and Melvin Perez (Statistical Institute of Belize), and Matt Price (Office for National Statistics, United Kingdom).

64. The session was dedicated to the modernisation of CPI production systems, including data collection, processing, analysis and index compilation. The following comments and conclusions were made:

65. To increase efficiency and reduce response burden, the Voorburg Group on Service Statistics suggests using CPIs as proxies for price development in certain areas of production,

such as transport, communication, accommodation and catering services. Collaboration of CPI and PPI producers can be mutually beneficial.

66. With software available for free it is possible also for smaller countries with relatively few resources to develop CPI systems to manage data collection, survey administration, data processing and index construction (Statistical Institute of Belize). Open-source technologies and software are key tools to ensure sustainability and shareability.

67. The move from a traditional survey-based statistics set-up to a production system that facilitates the use of multiple data sources, collection modes and calculation methods and their integration into the CPI will be a major challenge for statistical offices. The Office for National Statistics, United Kingdom, presented a newly developed system for the national CPI that includes sandbox and production lines for analyses and testing of new indices in parallel to the regular CPI production, and makes use of cloud-based technologies.

68. Modernisation and automation of CPI production include investing in machine learning, integration of processes, use of cloud-based computing and equipping price collectors with tablets and development of corresponding software with built-in algorithms for data analysis and validation. Automation of processes, in turn, can increase efficiency and free resources for methodological improvements of CPI.

69. The modernisation of the production system also requires training/upgrading of staff skills in data science, updating of documentation and data management principles, ensuring data confidentiality and making the new processes comprehensible for users.

70. The modernisation of the CPI production will be a major challenge for statistical offices. Because of the crucial importance of countries' development of suitable and up-to-date production systems it should be discussed at future meetings.

L. Session 8. Other CPI issues – owner-occupied housing

71. The session was chaired by Christopher Jenkins, Office for National Statistics, United Kingdom. The session was based on a presentation by Paul Konijn (Eurostat) on owner-occupied housing (OOH) in the Harmonised Index of Consumer Prices (HICP).

72. The inclusion of OOH in HICP has been a long-running discussion and no consensus on the most appropriate method, for EU countries, has been reached. Eurostat recommends further investigation and analysis before preparing a proposal for possible inclusion of OOH in HICP.[2]

73. Given the importance of housing costs, and the continued discussion of methods to measure OOH, it was agreed housing, and in particular the measurement of rental prices, should be included as a topic for a future meeting.

[2] The report of Eurostat's work on owner-occupied housing is available on https://ec.europa.eu/eurostat/web/products-statistical-working-papers/w/ks-tc-23-001