Modernizing the Consumer Price Index for the 21st Century (2022)

DETAILS

200 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-28698-5 | DOI 10.17226/26485

CONTRIBUTORS

Panel on Improving Cost-of-Living Indexes and Consumer Inflation Statistics in the Digital Age; Committee on National Statistics; Division of Behavioral and Social Sciences and Education; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine 2022. *Modernizing the Consumer Price Index for the 21st Century*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26485.

Prepublication Copy
Uncorrected Proofs

Modernizing the Consumer Price Index for the 21st Century

Panel on Improving Cost-of-Living Indexes and Consumer Inflation Statistics in the Digital Age

Committee on National Statistics

Division of Behavioral and Social Sciences and Education

A Consensus Study Report of

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This activity was supported by a contract between the National Academies of Sciences, Engineering, and Medicine and the U.S. Department of Labor under contract 1625DC-19-C-0008. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

International Standard Book Number-13: 978-0-309-XXXXX-X
International Standard Book Number-10: 0-309-XXXXX-X
Digital Object Identifier: https://doi.org/10.17226/26485

Additional copies of this publication are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2022. *Modernizing the Consumer Price Index for the 21st Century*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26485.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org**.

Prepublication Copy—Uncorrected Proofs

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Consensus Study Reports published by the National Academies of Sciences, Engineering, and Medicine document the evidence-based consensus on the study's statement of task by an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and the committee's deliberations. Each report has been subjected to a rigorous and independent peer-review process and it represents the position of the National Academies on the statement of task.

Proceedings published by the National Academies of Sciences, Engineering, and Medicine chronicle the presentations and discussions at a workshop, symposium, or other event convened by the National Academies. The statements and opinions contained in proceedings are those of the participants and are not endorsed by other participants, the planning committee, or the National Academies.

For information about other products and activities of the National Academies, please visit www.nationalacademies.org/about/whatwedo.

PANEL ON IMPROVING COST-OF-LIVING INDEXES AND CONSUMER INFLATION STATISTICS IN THE DIGITAL AGE

DANIEL E. SICHEL *(Chair)*, Department of Economics, Wellesley College
ANA M. AIZCORBE, U.S. Bureau of Economic Analysis
JAN DE HAAN, Statistics Netherlands (retired)
W. ERWIN DIEWERT, Vancouver School of Economics, University of British Columbia
LISA M. LYNCH, Heller School for Social Policy and Management, Brandeis University
RAVEN S. MOLLOY, Federal Reserve Board of Governors
BRENT R. MOULTON, International Monetary Fund
MARSHALL B. REINSDORF, International Monetary Fund
LAURA ROSNER-WARBURTON, MacroPolicy Perspectives, LLC
LOUISE M. SHEINER, Hutchins Center on Fiscal and Monetary Policy, The Brookings Institution

CHRISTOPHER MACKIE, *Senior Program Officer*
MICHAEL SIRI, *Associate Program Officer*
ANTHONY MANN, *Program Coordinator*

COMMITTEE ON NATIONAL STATISTICS

ROBERT M. GROVES (*Chair*), Office of the Provost, Georgetown University
LAWRENCE D. BOBO, Department of Sociology, Harvard University
ANNE C. CASE, School of Public and International Affairs, Princeton University, *Emeritus*
MICK P. COUPER, Institute for Social Research, University of Michigan
JANET M. CURRIE, School of Public and International Affairs, Princeton University
DIANA FARRELL, JPMorgan Chase Institute, Washington, DC
ROBERT GOERGE, Chapin Hall at the University of Chicago
ERICA L. GROSHEN, School of Industrial and Labor Relations, Cornell University
HILARY HOYNES, Goldman School of Public Policy, University of California-Berkeley
DANIEL KIFER, Department of Computer Science and Engineering, The Pennsylvania State University
SHARON LOHR, School of Mathematical and Statistical Sciences, Arizona State University, *Emeritus*
JEROME P. REITER, Department of Statistical Science, Duke University
JUDITH A. SELTZER, Department of Sociology, University of California-Los Angeles, *Emeritus*
C. MATTHEW SNIPP, School of the Humanities and Sciences, Stanford University
ELIZABETH A. STUART, Department of Mental Health, Johns Hopkins Bloomberg School of Public Health
JEANNETTE WING, Data Science Institute and Computer Science Department, Columbia University

BRIAN HARRIS-KOJETIN, *Director*
MELISSA CHIU, *Deputy Director*

Acknowledgments

This report reflects the contributions of many colleagues who generously gave of their time and expertise to the project. The United States Bureau of Labor Statistics (BLS) sponsored the study with an eye on continuing their decades-long commitment to the production of high-quality price statistics. As a key official statistic, the Consumer Price Index (CPI) provides critical information to a wide range of stakeholders, including other government agencies, private- and public-sector policy makers, financial market participants, researchers, the media, and the general public.

On behalf of my fellow panel members, we thank BLS staff who helped shape the project scope and then provided comprehensive information about the CPI program throughout the study. At the panel's first meeting, BLS Commissioner William Beach articulated the agency's strategy for continuing to modernize the CPI and highlighted priorities where the agency is looking for guidance from the panel. The panel also benefitted from expert coordination and sponsor leadership throughout the project from BLS Economist Anya Stockburger (branch chief, Office of Prices and Living Conditions).

The panel benefited from presentations on topics central to the panel's charge by several BLS staff responsible for delivering CPI statistics to the nation on a regular and timely basis. Rob Cage (assistant commissioner for consumer prices and price indexes) and Anya Stockburger discussed opportunities and challenges in the context of BLS's alternative data initiative. Rob also outlined current and planned work by the agency on price measurement for subpopulation groups, including past efforts to develop indexes defined by income quintiles. Brett Matsumoto (research economist) provided an overview of CPI program approaches to expenditure categories that raise unique measurement challenges such as health care, housing, and high-tech goods. Brendan Williams (economist) provided an overview of the CPI program's approach to quality assessment for alternative data that helps determine which sources to pursue and approve for use in production.

The panel is also grateful to many colleagues conducting research beyond the BLS who presented their work involving innovative use (or planned use) of alternative data for price measurement. This testimony included expert guidance from academic researchers and from practitioners delivering price statistics for agency programs in other countries. Tanya Flower and Helen Sands, Office for National Statistics, discussed ongoing and planned CPI data

transformation for the United Kingdom, including integration of scanner and scraped price data, and the role of these data sources for measuring inflation during the pandemic. Ken Van Loon, Statistics Belgium, described his agency's use of scanner data and web scraping, and implementation of a multilateral methods in CPI production. Leigh Merrington and Catherine Smyth, Australian Bureau of Statistics, provided a history of their country's use of scanner data in the CPI and how the agency had gone about implementing data modernization for the program. Heidi Ertl, Statistics Canada, detailed that agency's CPI modernization plan, highlighting experiences and lessons learned in the process. Claude Lamboray, Eurostat, provided an overview of data quality and statistical issues when alternative data are used as a substitute for, or are blended with, traditional data. Kevin Fox, University of New South Wales, synthesized material presented by all the statistical agencies and offered observations on which approaches are most promising, while highlighting some practical considerations.

Several presenters provided essential input on broad methodological price measurement topics, to whom the panel is grateful. Jens Mehrhoff, Deutsche Bundesbank, provided a statistical assessment of probabilistic (survey-based) datasets versus scanner and web-scraped data sources. Alberto Cavallo, Harvard University, and Pilar Iglesias, PriceStats, provided the panel with detailed information on techniques for measuring inflation using online price data, including how data collection and processing of online data could work in the production environment of a statistical agency. Matthew Shapiro, University of Michigan, John Haltiwanger, University of Maryland, and David Johnson, University of Michigan, provided an overview of current efforts to reengineer key economic indicators for the United States using alternative data. They presented findings from their recent research comparing alternative approaches to quality adjustment of price indexes using item-level price and quantity data and company-specific data. Costa Lasiy, Adobe Digital Insights, discussed his company's Digital Economy Index, which includes a high-frequency index for tracking e-commerce prices. David Byrne, Federal Reserve Board, discussed his research measuring price change for consumer digital access services.

One public session was devoted to invitees sharing their research to construct subgroup price indexes paying particular attention to the data demands and methodological issues confronted during such work. Xavier Jaravel, London School of Economics, discussed his work using scanner data to measure differences in prices paid by population subgroups across the income distribution. Greg Kaplan, University of Chicago, described recent research measuring inflation at the household level that revealed price dispersion across subgroups. Dennis Fixler, Bureau of Economic Analysis (BEA), reported on work by his agency to produce distributional statistics of personal income for U.S. households. Chris Payne, UK Office for National Statistics (ONS), described the household cost indexes published by his agency for population subgroups in the UK.

Another public session was held on the pricing of housing and shelter, a key CPI component given that it accounts for over 20 percent of consumer units' expenditures. Jeremy Moulton, University of North Carolina, discussed his research using Zillow data to measure the price of housing services, and offered suggestions for how the data might be useful to BLS in its current approach. Bettina Aten, BEA, described her work on the rental equivalence method of measuring housing services in the regional indexes and national accounts context. Annie De

Champlain, Elspeth Hazell, and Faouzi Tarkhani, all of Statistics Canada, provided a comprehensive overview of methods used by their agency to measure housing price changes, including use of alternative data sources. Robert Hill, University of Graz, followed with a discussion of the potential advantages of a simplified user-cost method as an alternative to rental equivalence for measuring price changes in owner-occupied housing. Brian Adams, BLS, discussed statistical issues pertaining to representativeness of housing data sources that rely disproportionately on prices of apartment rents. Paul Liegey, BLS, and Randy Verbrugge, Cleveland Federal Reserve Bank, also provided valuable perspectives regarding useful directions for use of alternative data sources in the housing component of the CPI program.

The panel held a similar session to learn about strategies for improving the accuracy, timeliness, and detail of price measurement for medical care services, specifically health insurance, purchased by consumers. After Brett Matsumoto, BLS, reported on the current use of alternative data sources in the medical care portion of the CPI and Abe Dunn, BEA, described medical care price measurement issues of interest to BEA, the panel heard from a roundtable of experts on the topic. Those providing details on decades of relevant leading-edge research included David Cutler, Harvard University, Matt Fiedler, Brookings Institution, and Martin Gaynor, Carnegie Mellon University. This discussion provided essential input for the panel's chapter on pricing medical care services.

Even with the many contributions detailed above, the panel could not have conducted its work efficiently without the capable staff of the National Academies of Sciences, Engineering, and Medicine. Brian Harris-Kojetin, director of the Committee on National Statistics, provided institutional leadership and overarching guidance about the study process; Kirsten Sampson-Snyder, Division of Behavioral and Social Sciences and Education, coordinated the review process flawlessly; and Paula Whitacre provided thorough final editing that improved the readability of the report. We also thank senior program associate Anthony Mann for his well-organized and efficient logistical support of the panel's meetings and Michael Siri, associate program officer, for substantive contributions organizing the panel's meetings and in formatting this report. The panel is especially indebted to Christopher Mackie of the Committee on National Statistics, who was the study director for the project. Chris did an extraordinary job staffing the panel, providing timely coordination and logistical support, critical insights and background on substantive issues, and substantial work in drafting the report as well as drafting the responses to external reviews. Beyond that, the panel chair is especially grateful for Chris's wise counsel throughout the entire panel process.

Finally, and most importantly, I would like acknowledge the collective effort of my fellow panel members who provided deep insights and gave generously of their time throughout the study. This report reflects the collective expertise and commitment of all panel members: Ana M. Aizcorbe, U.S. Bureau of Economic Analysis; Jan de Haan, Statistics Netherlands (retired); W. Erwin Diewert, Vancouver School of Economics, University of British Columbia; Lisa M. Lynch, Heller School for Social Policy and Management, Brandeis University; Raven S. Molloy, Federal Reserve Board of Governors; Brent R. Moulton, International Monetary Fund; Marshall B. Reinsdorf, International Monetary Fund (retired); Laura Rosner-Warburton, MacroPolicy Perspectives, LLC; and Louise M. Sheiner, Hutchins Center on Fiscal and Monetary Policy, The Brookings Institution. This group—chosen for their diverse perspectives, backgrounds, and deep subject matter knowledge—was a pleasure to work with. Each and every member made substantial substantive contributions to the report as well as maintained their good nature and positive attitude, even as all of our meetings had to be carried out remotely**.**

This Consensus Study Report was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets the institutional standards for quality, objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

The panel thanks the following individuals for their review of this report: Alberto F. Cavallo, Business, Government, and the International Economy, Harvard Business School; Julia Coronado, President, MacroPolicy Perspectives, LLC, New York, NY; Edward Coulson, Merage School of Business, University of California, Irvine; Matthew Fiedler, Economic Studies Program, Brookings Institution; Dennis J. Fixler, Chief Economist, Bureau of Economic Analysis; Kevin J. Fox, School of Economics, University of New South Wales, Sydney; Austan D. Goolsbee, Department of Economics, The University of Chicago Booth School of Business; Ariel Pakes, Department of Economics, Harvard University; John J. Stevens, Division of Research and Statistics, Board of Governors of the Federal Reserve System; and Marcel P. Timmer, Growth and Development Economics, University of Groningen.

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of the report was overseen by Charles F. Manski, University of Chicago. He was responsible for making certain that an independent examination of this report was carried out in accordance with the standards of the National Academies and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring panel and the National Academies.

Daniel E. Sichel, *Chair*
Committee on Improving Cost-of-Living Indexes and
Consumer Inflation Statistics in the Digital Age

Contents

Summary

The Consumer Price Index (CPI) produced by the Bureau of Labor Statistics (BLS) is the most widely used measure of inflation in the United States. The CPI is used in calculations of changes in the nation's economic output and living standards, to determine annual cost-of-living allowances for Social Security retirees and other recipients of federal payments, to adjust the federal income tax system for inflation, and to provide the yardstick for U.S. Treasury inflation-indexed bonds. The CPI also factors into determining the appropriate stance of U.S. monetary policy, which affects all Americans and the global economy. In addition, the CPI is monitored by households, businesses, and financial market participants to provide a broad statistic of price changes, and by many other organizations to adjust a wide range of contracts for inflation.

The marketplace for consumer goods and services, and the data available for characterizing it, have changed dramatically in recent decades. What consumers buy, how they buy, and from where is almost unrecognizable when compared with prevailing norms 100 years ago when the CPI was introduced. Reflecting these economic trends, price measurement has become more complex, placing ever greater demands on the data needed to attain accuracy, coverage, and timeliness. In the process, the decades-old survey infrastructure has been pushed beyond its capacity to meet the statistical needs of stakeholders. Field-generated data on which the CPI has traditionally relied have become more challenging and expensive to collect, and likely less representative of the overall population. At the same time, the digital revolution has given rise to vast new data sources that can be leveraged for the purpose of tracking consumer prices. As these trends have continued to unfold, statistical agencies around the world have responded with a sense of urgency to prepare for a world when most transactions leave electronic records that will become the principal source of price and quantity data.

Recognizing these realities, in opening remarks to the panel, BLS stated its objective to "convert a significant proportion of the CPI market basket from traditional collection to non-traditional sources and collection modes, including harnessing large-scale data, by 2024." To assist in this process, the agency commissioned the Committee on National Statistics of the National Academies of Science, Engineering, and Medicine to assemble a panel of experts to provide guidance on how the CPI might be improved by accelerating the transition to an approach that blends multiple data sources.[1] The panel also was asked to consider ways to

[1] The charge to the panel is reproduced in full in Chapter 1.

improve the measurement of traditionally difficult-to-measure sectors such as medical care and housing and, finally, to assess the policy and research value of developing price indexes to track differential inflation rates experienced by population subgroups such as those defined by income level. This summary highlights a subset of the panel's recommendations for modernizing BLS's CPI program.

THE POTENTIAL OF ALTERNATIVE DATA FOR CONSTRUCTION OF ELEMENTARY INDEXES

Point-of-sale data, household-generated home scanner data, and data scraped from the web are the primary "nontraditional" data sources that have been successfully exploited for price measurement. These data are generated for a wide range of items, often in near real time, and provide key information, including the price, source outlet, quantity, and characteristics of the item.

A key motivation driving data modernization is the potential to improve timeliness, relative to survey alternatives, in detecting what consumers are buying and from where. At no time has this need for updated data collection methods been exposed more starkly than during the COVID-19 pandemic. Statistical agencies that systematically use transactions data in their CPI programs—such as the Australian Bureau of Statistics (ABS), where less than 2 per cent of the price quotes (by expenditure weight) are collected by field staff in-stores—were able to provide up-to-date information about prices being paid and about shifts in consumers' buying patterns even as access to outlets became highly limited during lockdowns.

Scanner Data

Integration of scanner data registering retail transactions has been in the research and production pipelines of statistical agencies' price measurement programs for decades. The hope has always been that scanner data on both prices and quantities could vastly expand CPI coverage of product varieties and outlets and, in the process, accelerate the detection of shifts in consumers' buying patterns. A number of statistical agencies around the world have since established the practical feasibility of the concept.

Beyond point-of-sale data, some data vendors produce datasets containing information recorded at home by individuals on their purchases using a scanner. One important benefit of household-based scanner data is that purchases made at retailers that do not participate in point-of-sale programs can, in principle, be captured. Data collection centered around those making purchases also enables information on consumer characteristics to be collected in a way that can be used to construct price indexes for different population subgroups.

Web-Scraped Data

In addition to scanner data, web-scraped data is the other major source of high-frequency information being applied by statistical agencies to price measurement. Scanner data are not available for goods purchased online and for some goods where one firm dominates the market (e.g., Apple smartphones). For those items, online price data provides an alternative. An attractive feature of web-scraped data is that they are often easier to access than are point-of-sale data from retailers. Additionally, when data processing algorithms can be automated, time lags are almost eliminated. The main theoretical challenge with web-scraped data for official price measurement is that methods are still underdeveloped for establishing the relative importance of different items in the consumption basket, given that information on quantities sold is typically not available. Statistical agencies are in the process of overcoming many of these challenges—sometimes by combining survey and nonsurvey data sources—and pushing forward aggressively with use of web-scraping in their CPI programs.

Next Steps

To date, alternative data typically have been integrated incrementally within the existing CPI infrastructure when the opportunity has arisen or when pressure emerged to do so because of a problem with a conventional data source. Going forward, BLS will need to go beyond price quote replacement and make progress in areas where traditional data are still available, but where cost or accuracy benefits may emerge from pursuing alternative sources. **BLS should embark on a broad-based strategy of accelerating and significantly enhancing the use of transactions data and other alternative data sources in CPI compilation. Embracing alternative data sources now, and moving forward aggressively with research for their integration, will ensure that the accuracy and timeliness of the CPI will not be compromised in the future. The data modernization strategy will involve:**

- **Identifying promising alternative data sources and then prioritizing the work needed to evaluate and incorporate these data into the items/strata where they can be applied;**
- **Continuing development of a robust research agenda that supports incorporation of alternative data and associated new methodologies more broadly beyond just price quote replacement;**
- **Continuing research assessing the quality of new types of data;**
- **Developing staff expertise that includes more data scientists and other specialists;**
- **Creating a cross-agency strategy for data acquisition with the possibility of joint contracts across statistical agencies;**
- **Carrying out a strong communication strategy to inform stakeholders of plans and implementation details.**

Testing of indexes constructed from alternative data sources and methodologies will be a key part of data modernization. **Before BLS incorporates alternative data for specific item categories into the official CPI, it will be important to have a significant overlap period (perhaps as long as two years) during which parallel indexes based on new data sources can be tested and compared against their traditionally constructed counterparts.**

BLS has the opportunity to build upon the experiences of statistical agencies that have expedited incorporation of nonsurvey data into their CPIs. These agencies have navigated key challenges such as managing risks associated with the use of privately collected data and assessing the capacity of alternative data sources to track product quality changes. Where rapid quality changes are common, such as for high-tech items, combining datasets that include information about product characteristics expands the opportunities for improving price measurement.

One analytic task where immediate progress can be made is the development of elementary indexes constructed from web-scraped data. Beginning in March 2020, due to initial COVID-19 shutdowns, BLS had to improvise as monthly in-person collection of price data from retailers and businesses by field staff came to a halt. For an extended period, price checkers were limited to filling up virtual carts online to check prices. This process, brought on as a stopgap measure in the face of the immediate crisis, only mimicked the in-store price checking activity. **Converting opportunities for permanently automating web-scraping of price data should be a high priority for the CPI. In evaluating the usefulness of web-scaped data for elementary index estimation, food, electronics, and apparel should be priority categories. Data for these categories are readily available with a large share of transactions already online, and work by other statistical agencies and private-sector organizations have demonstrated feasibility. In the short term, BLS could consider obtaining web-scraped data from outside vendors but ultimately BLS should develop automated web-scraping methodologies within the agency.** BLS should also continue and expand work with large companies to understand how they record data, and then collaborate on building programming interfaces that can be run behind company firewalls to provide the statistics needed by BLS.

HIGHER-LEVEL AGGREGATION AND SHIFTING CONSUMER BEHAVIOR

With the availability of near real-time information from private-sector data sources, coupled with an ever-increasingly dynamic economy, the collection and dissemination of timely data has become a basic expectation of statistical agencies. While the composition of what consumers buy is always evolving, the shifts were especially dramatic during the pandemic. Spending on travel, food away from home, and clothing worn outside the home declined dramatically, while demand for computer and communications equipment needed for remote work and home deliveries surged. Improving the ability to track such shifts is an essential goal.

Revising CE Weights More Frequently

The primary method used by most statistical offices to determine the composition of households' expenditures is to ask them directly. In the United States, this process is carried out using the Consumer Expenditure Survey (CE), which has for decades been the statistical system's most comprehensive source of data on households' income and expenditures. And—because a nationally representative survey conducted by a government statistical agency is needed for benchmarking estimates of consumer expenditures in a way that links buying patterns to households—a version of the CE will continue to have value for the foreseeable future.

However, if the relevance of the CPI is to be maintained, changes in the way market basket shares (weights) are estimated must be a high priority. **To improve the timeliness of the CPI and the accuracy with which it captures changing buying patterns, BLS must (1) update upper-level weights—which currently, on average, lag 36 months behind actual expenditures in a given period—more frequently and rapidly, and (2) improve the**

accuracy of weights applied to specific items that the Consumer Expenditure Survey measures poorly and for which alternative data likely are more accurate.

Ideally, the expenditure data used to calculate CPI weights would come from a single 12-month period ending no more than six months prior to their introduction. For example, new CPI weights introduced in January 2022 would reflect expenditure patterns from July 2020 to June 2021. This production schedule may take time to achieve so, as an interim step in mitigating the timeliness problem, weights should be updated annually using two-year rolling averages of the CE data. Under this setup, the rolling weights would still lag real-time market realities, but not by as much as they do in the current two-year cycle.

Broadening Sources of Data Used for Estimating Expenditure Weights

In the not-to-distant future, detailed price and quantity information likely will be available from a range of nonsurvey sources for almost all products, which will allow much more frequent (and possibly more accurate) updating of CPI weights. **With supplementing and complementing the CE data in mind, BLS should invest in collecting comprehensive data for individual spending using electronic means of payments such as credit/debit cards or electronic payment processors (e.g., PayPal or Stripe). Initially, these new data could be applied to the chained CPI-U or to a new experimental index. Later, after an adequate period of study, expenditure pattern estimates used to construct CPI weights should be derived as a blend of data on spending from (1) the CE, (2) timely private sources, and (3) the national accounts.** Given limitations of electronic transaction data to link information on prices and quantities to specific households—and, in turn, to population subgroups—research into use of alternative data sources for estimating upper-level weights should initially be directed toward production of the national-level CPI.

The national accounts—specifically, Personal Consumption Expenditure (PCE) data—from the Bureau of Economic Analysis can also be used, albeit retrospectively, to produce a more accurately weighted experimental CPI. One advantage of PCE data is that they are benchmarked to a census of retail establishments (conducted every five years) and a variety of other merchant-based sources, so they reflect a more comprehensive accounting of transactions by consumers. A first step might involve creating an experimental CPI that uses PCE-generated weights at the upper (243 item) level but that is otherwise no different from the CPI. **Another option for blending PCE and CE data that BLS should test for the purpose of updating upper-level expenditure weights is to continue using the CE as the benchmark for most CPI categories, then integrating PCE data to adjust the acknowledged weakest categories of the CE.** The experiences of statistical agencies in several other countries that already use this strategy can be drawn upon to expedite this line of research. For example, Statistics Canada (2021) published a special edition price index using credit and debit card data supplied by the Bank of Canada and alternative weights based on national accounts data "to account for pandemic-related expenditure shifts at more detailed levels of geography and CPI components."

Longer-Term Planning

The strategy outlined above presumes that BLS will adopt a data approach for establishing upper-level expenditure weights that blends data from national accounts and private companies. A longer-run alternative to the data purchase model would be for BLS to set up an in-house operation for collecting the needed data. Since point-of-sale scanner data are not perfectly suited for use in the CPI (for reasons discussed in Chapter 2), the project should focus on collecting scanner data directly from households. **BLS should begin exploring development of a household-based scanner recording program that would capture prices, quantities, and item characteristics of purchases made by surveyed respondents. In addition to its value for estimating item strata weights, this method of obtaining spending information would be useful for construction of elementary aggregates.** BLS should consider "leapfrogging" traditional methodologies of handheld scanners that require large initial investments and look to modern approaches using a custom mobile phone app. Technologies are changing fast, so the most durable solutions may be based on flexible, mostly software-based approaches.

MODERNIZING MEASUREMENT OF DIFFICULT-TO-MEASURE EXPENDITURE CATEGORIES

A small subset of goods and services account for a large share of expenditures by households, and therefore have an outsized impact on the CPI. If new data sources can be harnessed for improving price measurement in these categories, the return on investments by BLS may be especially large.

Housing

Housing services represent by far the largest component of most consumers' cost of living. Owner-occupied housing is particularly important, both because it accounts for about three-fourths of the shelter category and because statistical agencies have not fully converged on a standard measurement approach. Prominent reviews of the CPI program have endorsed BLS's rental equivalence method—which essentially estimates the amount that would be required for a homeowner to rent a home with all the characteristics of that household's owned home—because, in many situations, it corresponds closely to a cost-of-living concept for measuring housing services. **BLS should continue using rental equivalence as the primary approach to estimating the price of housing services for owner-occupied units.**

The CPI methodology has traditionally relied on survey data to provide information on rent changes and housing expenditure shares. However, a number of new data sources for rents have emerged in recent decades, resulting from the expansion of large institutional landlords and property management companies. Supplementing the CPI Housing survey with data from such sources could help improve the accuracy of imputations of rent changes to the owner-occupied stock. **BLS should seek to identify new data sources that would allow for improved coverage of single-family homes and of areas where houses are predominantly owner-occupied. New data sources could also improve the CPI's ability to reflect rapid changes in rent growth by allowing rent for a given housing unit to be measured in consecutive**

months, allowing changes to be assessed over short periods of time. By contrast, the CPI Housing Survey only samples each housing unit every six months, which is problematic in times when rents change rapidly.

In addition to their value in measuring rent changes, alternative information sources—including property tax records and American Community Survey data—could be used in conjunction with the CPI Housing Survey. Specifically, **BLS should consider strategies for estimating expenditure shares for owner-occupied housing that would make use of the rich housing characteristics information often available in property tax data.**

In tracking the price of shelter in the United States, geographic detail is especially important because so much variation exists across and within regions. With this in mind, **BLS should publish additional detail on the housing components of the CPI, such as indexes by structure type and for a larger number of metropolitan areas than the roughly 20 areas for which they are currently published. Broadening the geographic scope of the CPI could be facilitated by de-linking the housing sample from the samples of other CPI items.**

Imputing rent for owner-occupied homes works best when there is a high degree of overlap—in terms of geography and housing quality—between the market of homes for sale and the rental market. When overlap is more limited, a user cost approach might be helpful to improve estimates of the price of housing services. Thus, although this panel is of the view that the rental equivalence approach should continue to be the primary method used in the CPI, research conducted on data at the micro level would be valuable for testing where the rental equivalence method is performing well and where it is not (such as for pricing housing services associated with higher-end properties). **As part of its research program, BLS should compare rental equivalence estimates to user cost estimates for individual properties. Research on alternative methods for housing could lay the groundwork for eventually publishing housing indexes using different methodologies.**

Medical Care

Medical care is another large, growing, and rapidly changing consumer expenditure category. Even though the scope of the CPI covers only the share of the sector directly paid for by consumers (and not the share of spending by businesses and government), medical care is an important component of the CPI. Within medical care, health insurance is the largest expenditure made by consumers. For this reason, a pressing methodological decision facing BLS is whether to continue pricing insurance using an indirect method or to migrate to a direct pricing approach. The direct approach involves estimating total health insurance premium prices; the indirect approach involves pricing health insurance using prices of medical care blended with information about retained earnings of insurance providers. The underlying logic of the indirect approach is to separate out the prices of medical care goods and services from the portion of insurance premiums retained by insurance companies to cover administrative costs and profits. Although perhaps second best conceptually, the indirect method represents a pragmatic approach that at least partially reflects quality change and care utilization. **The indirect method has practical advantages and therefore should, in the short to medium run, continue to be the method for pricing health insurance in the CPI.**

The above recommendation notwithstanding, declining response rates in the medical care components of the commodities and services surveys are making reliance on the indirect method increasingly difficult. For this reason, and for its attractiveness conceptually, **BLS should investigate historical differences between the indirect and direct approach doing a true apples-to-apples comparison. A "whole health insurance price deflator" that is a weighted average between the CPI's current health insurance deflator and the various deflators for the medical services financed by insurance should be calculated and compared to the deflator used in the direct approach. If this research reveals that the two approaches do not differ greatly historically, BLS could revisit its reliance on the indirect method.**

One source of concern with the current indirect approach is the volatility exhibited in the insurance services component of the total health insurance price index. For example, a particularly heavy flu season leading to high utilization of health care services will reduce retained earnings and push down that component of the index. For this reason, among others, a number of potential improvements to the indirect approach should be considered. **To better capture what consumers actually pay for insurance, and which does not depend on utilization rates, BLS should explore using a rolling average of retained earnings per unit of health services (where retained earnings equal premiums, less medical expenses), rather than an annual value.** This approach will mean that actual changes in the cost of health insurance—stemming from changes in regulation, market structure, or technology—will appear more slowly in the CPI; this limitation could be an acceptable tradeoff if the problem of excess volatility now built into the index can be mitigated.

Also, there is currently a long lag between the time that prices of health insurance change and their incorporation into the CPI. Shortening this lag would improve the accuracy and timeliness of the index. **For the purpose of tracking changes in health insurance prices, BLS should consider switching from using annual data on profits net of premiums to quarterly data.**

For health insurance, developing methods and identifying data sources for adjusting prices to reflect quality change in insurance policies over time will continue be a top priority. As part of this work, BLS should continue evaluating how to accelerate incorporation of claims data to improve the coverage, detail, and timeliness of price and quantity information in the medical care component of the CPI. Such data may also be useful for research on broader questions, such as about productivity in the medical care sector. A pilot program is currently underway at BLS indicating promising uses of claims data including for the construction of experimental disease-based price indexes.

Supplemental Subgroup Price Indexes

Price indexes and other economic statistics tailored to describe and track the experiences of specific population subgroups are a growing research and policy need. The rationale for producing price indexes for population subgroups is clear for purposes such as adjusting Social Security benefits and marginal tax rates or for specifying transfer payments for which only certain groups are eligible.

One factor that can lead to differential inflation rates—and, as it turns out, the easiest one to measure—is that people purchase different baskets of goods and services. In general, research using a simple reweighting approach to reflect different consumption patterns has tended to detect only minimal differences in inflation rates faced by different groups. To fully portray differential inflation, subgroup price indexes must also account for different prices paid for similar items. Recent research based on diverse data sources has revealed clear patterns of differential price inflation, in particular across income groups. Crucially, this research suggests that the greatest source of heterogeneity in households' inflation rates is variation in prices paid for the same types of goods—not from variation in broadly defined consumption bundles. Research and policy making stand to benefit a great deal if the underlying trends in price inflation faced by different population groups can be more accurately measured and, in turn, better understood. **Because of the urgency of issues related to income and wealth inequality, social welfare, and poverty, developing price indexes for population subgroups along the income distribution should be a high priority for BLS. Identifying data sources that would ultimately allow production of price indexes by income quintile or, if possible, decile is a key part of this work.** The potential return from investments in developing income-defined subgroup price indexes is further enhanced by ongoing work at the Bureau of Economic Analysis to produce prototype statistics on the distribution of personal income across households.

As with other aspects of CPI modernization discussed in this report, long-term promise for creative initiatives for subgroup indexes comes with the increased availability of micro data containing information on prices that households actually pay and on details about the items purchased. However, electronic transactions data as currently generated do not cover all consumer expenditures. For this reason, the next generation of empirical studies on inflation inequality will need to draw on additional, alternative data sources—perhaps most importantly, a household-based scanner recording program that captures prices, quantities, and item characteristics of purchases made by surveyed respondents.

The above-described research linking individuals to their purchases strongly suggests the need for approaches that blend multiple data sources—encompassing survey data that cover the full consumption basket, including item categories for which electronic transaction data are still incomplete, and commercial data sources that allow deep analyses of prices paid and product detail—in a way that account for the full range of consumer expenditures.

ORGANIZATIONAL CONSIDERATIONS

Given that it has performed reliably for decades, the survey-based methodology underlying the CPI is commonly assumed to be the gold standard for estimating price changes. However, as with other economic statistics rooted in the application of a 20th century survey-centric system, the resulting estimates have been affected by falling survey response rates and increasing costs. Accordingly, the data collection model of statistical agencies is shifting. To fully capitalize on emerging data opportunities to improve the quality and timeliness of the CPI, a paradigm shift at BLS will be required that lessens reliance on older survey-based approaches.

In addition to the methodological challenges—such as coverage, representativeness, and scope of variables present—practical considerations hinder development of a mixed data infrastructure that includes public, private, survey, and nonsurvey data. Legal constraints, privacy concerns, and high data acquisition costs—acutely present for the U.S. case—have slowed the incorporation of commercial and even government administrative data sources into social and economic statistics (NASEM, 2020). Modernization of the data system is further complicated by the decentralized statistical system of the United States. Especially within such a system, there are compelling reasons for agencies to pursue integrated data collection and production processes. **More extensive collaboration between the Census Bureau, BLS, and BEA—along with other statistical agencies that collect key economic data, such as the U.S. Department of Agriculture—is needed to advance the acquisition and use of alternative data sources in the production of economic statistics. More specifically, such coordination will allow the statistical system to negotiate common, unified, comprehensive contracts with companies that collect applicable data.**

To navigate the above-described complexities, and to establish authority and accountability within BLS, the agency should build data modernization into its organizational structure. **BLS should designate a single, high-level person within the agency, preferably at the deputy commissioner level, whose job is to lead data transformation efforts. Having this responsibility explicitly designated would facilitate a focused, coordinated effort and would ensure accountability. This person also could be the visible point person in collaborations with other statistical agencies. A key objective is to avoid duplicative efforts that likely would arise if data transformation proceeded in a more decentralized way.**

In addition to facilitating administration of workflows, such formalized institutional arrangements would signal a commitment by senior leadership to expand the use of alternative data sources for statistical purposes. The data transformation lead would also be part of the team to develop communication strategies to work with Congress to seek the necessary resources to implement changes and highlight the value of the task to user communities.

BLS should continue to look externally for data modernization models as well. With price measurement in particular, ample opportunities exist to learn from and adopt innovative approaches pioneered by statistical offices internationally. **BLS should enhance its contacts and collaborations with CPI staff in statistical agencies beyond the U.S. system. Other countries have made significant progress on data transformation—specifically in methods blending scanner and web-scraped data with survey sources—and CPI staff would benefit from more fully investigating successes and failures experienced during these efforts.** Beyond national statistical offices, some of the most innovative work on price measurement—often involving electronic transaction data or web-scraped price data—has taken place in academic settings, so continued collaboration with these experts is likewise encouraged.

The kind of data modernization envisioned in this report will require upfront investments in data acquisition, updating of CPI program production procedures and IT systems, and staff training. In the future, the CPI staff skills will need to shift at least partially away from those needed to analyze structured, survey-based price information and toward those needed to process unstructured price data. **In addition to hiring staff with data science skills, BLS should strive to develop this talent in-house by supporting and rewarding staff who pursue training and educational opportunities to develop the technical expertise that will facilitate data transformation efforts in coming years.**

Additionally, since confidence in and understanding by data users of official statistics is critical, successful modernization of the CPI will require that BLS provide clear and consistent communication to stakeholders about the re-design on an ongoing basis. This includes advance notice of changes in an easy to find location on the website, detail about alternative data sources incorporated, transparency around experimental indexes, and updates on the timelines of project as they evolve.

1
Introduction

Measurement of price inflation in an economy serves a range of essential policy and program purposes, and the Consumer Price Index (CPI) produced by the Bureau of Labor Statistics (BLS) for more than 100 years is the most widely used measure of consumer price changes. As a key economic indicator, the CPI measures the average change in the prices paid by households for a market basket of goods and services. It is used to deflate some components in the calculation of the nation's gross domestic product (GDP) statistics and to calculate inflation-adjusted changes in measures of the nation's living standards. The CPI is also used to determine annual cost-of-living allowances for Social Security retirees (since 1972) and other recipients of federal payments, adjust the federal income tax system for inflation (since 1985), and provide a yardstick for U.S. Treasury inflation-indexed bonds (since 1997). Finally, the CPI factors into determining the appropriate stance of U.S. monetary policy, which affects all Americans and the stability of the global economy. Outside of the government, the CPI is often used by households, businesses, and financial market participants as a broad statistic of price changes and to adjust wages and lease payments.

1.1. THE GOALS OF PRICE MEASUREMENT; RESEARCH AND POLICY NEEDS

Prior to the introduction of the CPI in 1921, several "cost of living" measures had been developed, first by the Bureau of Labor and then by the Bureau of Labor Statistics. In 1917, a price collection effort of family expenditures in 92 industrial centers was begun to provide appropriate weighting patterns for a comprehensive index. By 1919, BLS began publishing semiannual cost-of-living data of retail prices for 32 cities and then extended this index to cities across the country in 1921. A price sampling framework more akin to the modern version was established after the 1935 revision. This revision doubled the number of food items sampled and changed the rent sampling to "better represent the population consisting of urban wage earners and lower salaried workers" (Rippy, 2014). The name "cost-of-living index" was changed to "consumer price index" (CPI) in 1945.

As stated in its *Handbook of Methods* (BLS, 2020, Chapter 17), BLS formally adheres to a "conditional COLI framework," which strikes a compromise between a true cost-of-living index (COLI)—which measures the change in expenditures a household would have to make to maintain a given standard of living—and a cost-of-goods and service index (the more traditional market basket pricing approach). The "conditional" COLI seeks to "measure changes in consumers' costs of living on the assumption of stability in conditions—such as the weather or the quality of publicly provided goods—that are outside the universe of private goods."[1] Statistical agencies around the world have by and large settled on this conceptual basis for price change measurement, and this methodological decision is not revisited here. Rather, this report is about modernizing the price measurement data infrastructure within the context of this accepted framework. New data sources carry the potential to increase accuracy, detail, and timeliness (and possibly reduce costs), as well as present new methodological opportunities.

The most appropriate measure of consumer price change is not the same for all purposes; recognizing this, statistical agencies have developed multiple indexes over the years, with different coverage, weighting schemes, and aggregation formulas. Over its history, BLS has produced consumer price indexes focused on wage earners and clerical workers, on urban consumers, and on elderly consumers, as well as an index that more fully accounts for shifts in consumer buying and product substitution patterns.

Elsewhere in the U.S. statistical system, the Bureau of Economic Analysis (BEA) constructs another measure of consumer inflation, the Personal Consumption Expenditures (PCE) price index. This index has a broader scope than the CPI, tracking all spending by and on behalf of the "personal" sector, which includes both households and nonprofit institutions serving households. In contrast, the CPI mostly tracks households' out-of-pocket expenditures. This difference in scope contributes to differences in weights such that, for example, the weight on medical care in the PCE index is larger than that in the CPI because the PCE index includes expenditures paid for by employer-provided insurance, Medicare, and Medicaid while these expenses are not included in the CPI. The PCE also uses a chained index formula that allows it to better account for consumers substituting between items in response to relative price changes.[2] This chained index formula relies on data available with a lag that are not yet available at the time when the official monthly CPI is published, and the PCE can be substantially revised while the (nonseasonally adjusted) CPI is never revised. As a result, over the last 20 years, while the two prices indexes have following broadly similar trends, they are not identical. The CPI typically indicates somewhat higher inflation.

1.2. MOTIVATION FOR THE STUDY: BUILDING A CPI FOR THE 21ST CENTURY

Since the introduction of the CPI, the marketplace and the data available for characterizing it have changed dramatically. What consumers buy, how they buy it, and from where is almost unrecognizable when compared to prevailing norms 100 years ago. Many more products exist, outlet structures are much more diverse, product turnover including introduction

---

[1]National Research Council (2002, p. 3) provides a detailed comparison of the COLI and the "cost-of-goods" approaches and a full discussion of the data and theoretical realities faced by statistical agencies making the conditional COLI concept suitable for official statistics.

[2]More specifically, the PCE price index is based on the Fisher-Ideal formula and the CPI is based on a modified Laspeyres formula.

of new goods and services has become increasingly rapid, and a relatively higher proportion of the market basket consists of information goods and services. At the same time, “basics” such as food and clothing make up a shrinking proportion of overall consumer expenditures. Where consumers buy has changed as well, as shown in the growth of online shopping since 1999 (Figure 1-1).

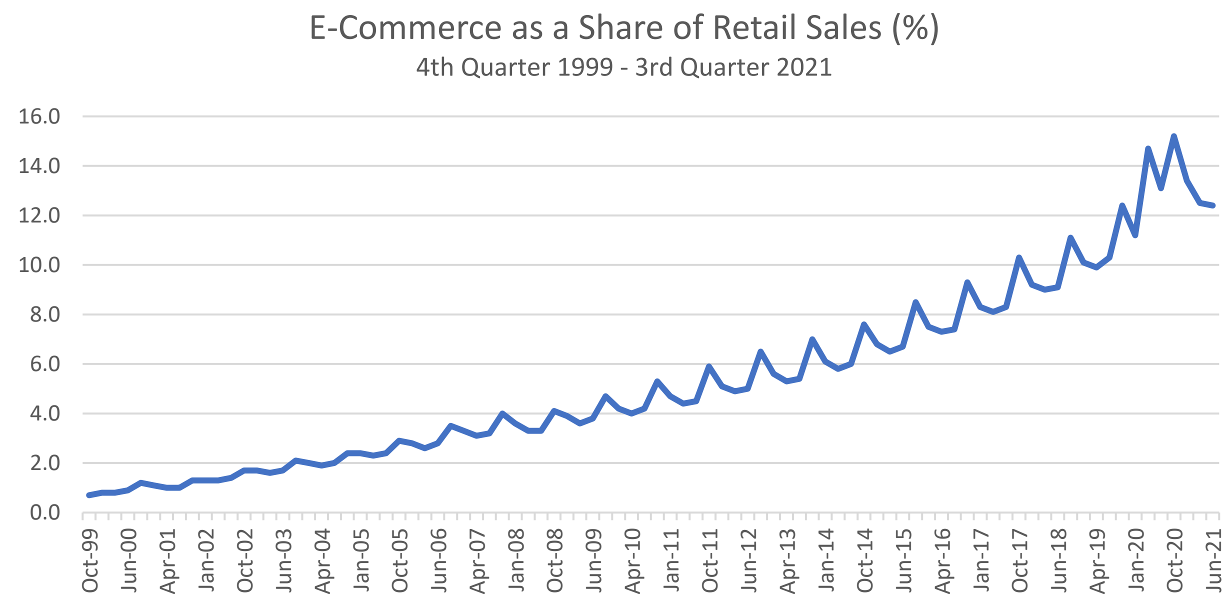

FIGURE 1-1 Growth of online shopping since 1999 in the United States (excluding food services).
SOURCE: Quarterly data from Retail Indicators Branch, U.S. Census Bureau.

As a result of these economic trends, price measurement is much more complex today, which has led to new and greater demands on the data needed for accurate price measurement. In the process, the decades-old survey infrastructure has been pushed beyond its capacity to meet these demands. **Traditional field-collected data sources on which the CPI has relied are becoming more challenging to collect and likely less representative, particularly given changes in consumer and business behavior. Fortunately, at the same time, the digital revolution of recent decades has given rise to vast new alternative data sources that can be leveraged for the purpose of tracking consumer prices.** These new data sources have the potential to improve the accuracy, coverage, and timeliness of the CPI. In particular, a multiple data source paradigm could reduce traditional sources of bias, including from product churn, quality change, lower-level substitution, and outlet churn. For price measurement, transaction data (especially those that can be associated with households’ demographic characteristics) and internet price data are particularly promising, and they are discussed in detail in Chapter 2. More traditional types of alternative data also exist that could be used for some purposes, including surveys undertaken by other government agencies and commercial firms as well as administrative data from government agencies.

Historically in the CPI, a set of interrelated surveys determined the goods and services to be covered, the prices of those items, the outlets from which the prices are obtained, and the relative importance (weight) that is given to each category of goods and services in the index. These methods of price sampling and establishing consumer expenditure shares were more appropriate and more effective in the 20th century than they are now. Today, and even more so 5 to 10 years down the road, BLS will need to take advantage of alternative data sources for the CPI to keep up with measurement demands and to remain relevant.

The current survey infrastructure has been too rigid to capture recent changes in consumer behavior in a timely manner. Item and outlet samples are refreshed infrequently (one full rotation takes four years), leading to long lags in incorporating new types of outlets and goods. More recently, as noted above, a significant share of purchases has shifted to online retailers, creating new complications for tracking prices over time.[3] Indeed, when consumption patterns change rapidly, the CPI may miss these shifts as basket weights are derived from household Consumer Expenditure (CE) surveys with a two-year lag. Meanwhile, as the economy has become increasingly complex and fast-moving, users of the CPI data are demanding more detailed, timely, and high frequency information than ever before (Abraham et al., 2021). To meet modern information needs, statistical agencies and researchers alike are turning to alternative data sources—information beyond that collected through traditional CPI field procedures such as in-store price checking.

The surveys at the core of the CPI have struggled with falling response rates for some time, raising concerns about sample representativeness. While surveys of households' consumption behaviors performed reliably in the 20th century, the survey-based data infrastructure underlying the CPI has come under strain more recently. Widespread survey hesitance has reduced response rates and that, in turn, has boosted the costs of obtaining representative samples.

The challenges of conducting price and expenditure surveys in a timely and reliable manner have been brought into sharp focus by the COVID-19 pandemic. CPI weights, updated every two years using lagged CE survey data, failed to capture dramatic shifts in consumption patterns in 2020 and 2021, leading to biases in measured CPI inflation. Airfares provide a salient example. Figure 1-2 highlights the modest decline in the weight on airfares in the CPI relative to those estimated by the Bureau of Economic Analysis for the PCE price index (which uses more up-to-date weights). The figure illustrates how the CPI placed too much weight on airfares during this period and largely missed the pandemic-induced swings in the amount of air travel purchased by households.

[3]The share of e-commerce in total U.S. retail sales surged by 4 to 5 percentage points during early stages of the COVID-19 pandemic (U.S. Census Bureau, https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf).

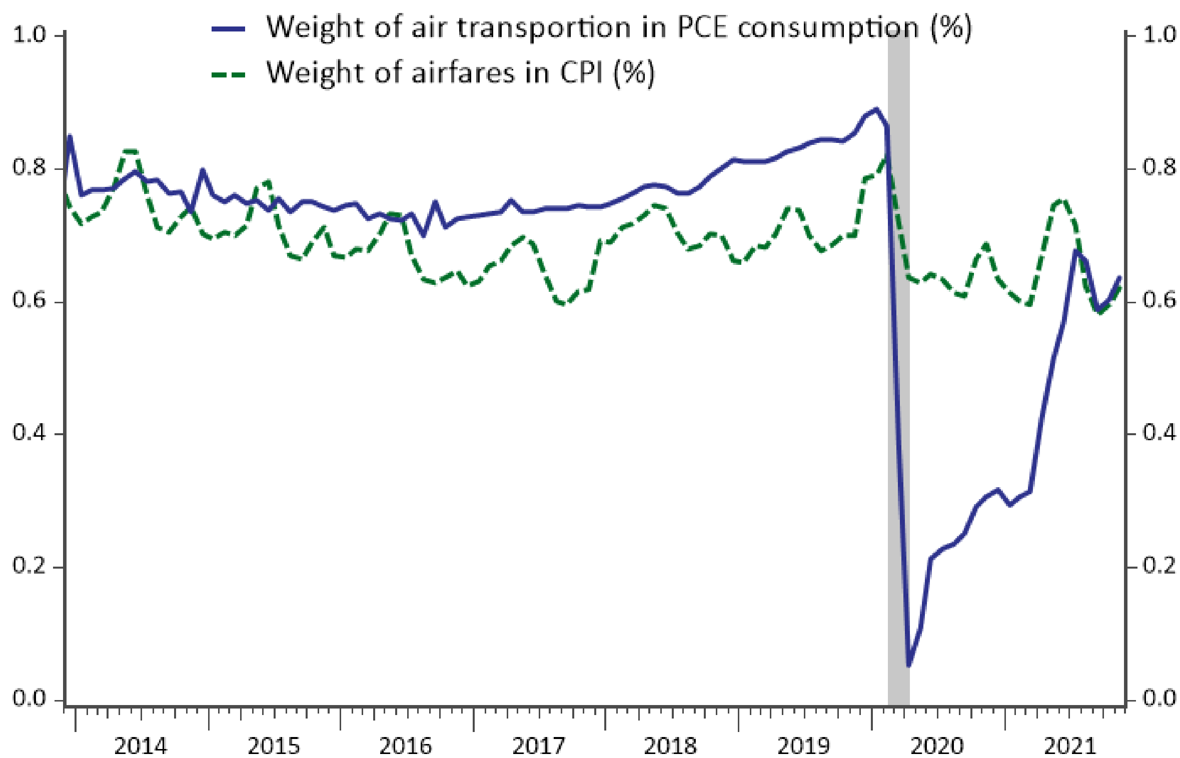

FIGURE 1-2 CPI and PCE weights (expenditure shares) for air transportation, 2013–2021.
SOURCE: Panel-generated, using data from BEA and BLS.

Regarding the collection of price quotes, prior to the pandemic, CPI prices were largely collected through in-person visits from two surveys: the commodities and services (C&S) survey and the housing survey. BLS suspended in-person collection in March 2020, leading to a rapid shift to online and telephone collection; this shift in data collection modes is illustrated in Figure 1-3.

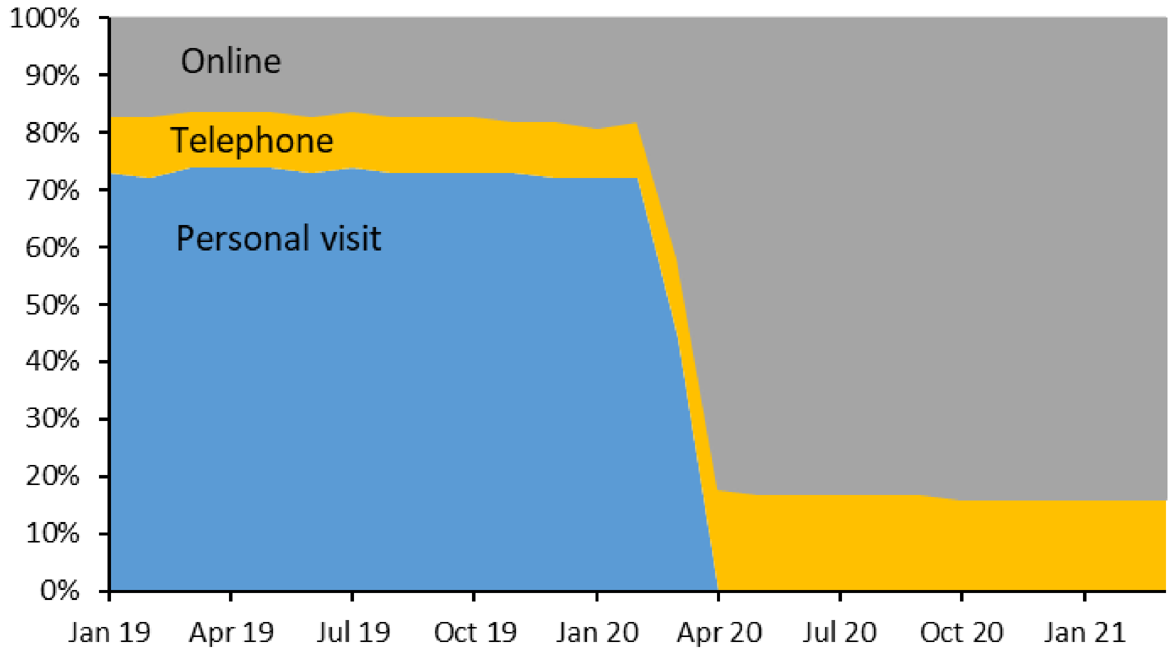

FIGURE 1-3 Shifts in pandemic economy data collection modes for the commodities and services price survey.
SOURCE: Panel-generated, using data from BLS.

While alternative data sources will be useful for many aspects of the CPI, it also is important to acknowledge that it will be difficult to move away fully any time soon from systematic nationwide surveys that track consumer expenditure patterns. Surveys will still be needed for calibration, benchmarking, and testing representativeness, although, supplemented with nonsurvey data, perhaps some surveys could be undertaken less frequently.

A paradigm shift at BLS is needed for the CPI to be modernized. The agency needs to be more flexible about utilizing new data sources, developing new methodologies, and moving away from a strong emphasis on survey error. As noted, BLS has built an infrastructure well-suited for a CPI based on surveys and the types of errors associated with those surveys—and that framework performed well for many decades. Indeed, the CPI was long considered the gold standard of consumer inflation measures. The balance of costs and benefits of this approach, however, has shifted. As noted, the surveys underlying much of the CPI are coming under increasing strains that raise the possibility of greater survey error as well as different types of nonsurvey error. Users are demanding more timely and disaggregated data, and outside organizations already are producing inflation estimates using alternative data sources that are garnering considerable attention. In this environment, BLS will need to develop and invest in new data sources and methodologies as well as develop the complementary staff and expertise required to ensure that the CPI remains a premier measure of inflation for the 21st century.

More broadly, BLS has a vital role to play in the changing data culture in economic statistics brought on by the increasing costs of surveys, reduced response rates, and emergence of data complements and substitutes, including administrative and commercial data sources. The statistical system now finds itself at a promising moment to make headway on modernizing its statistical infrastructure, in part because of the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act), which became law in 2019. This act was intended as a starting point for improving federal data infrastructure. Increasingly, researchers and policy makers benefit from the capability to link across data sources "collected through formal surveys, federal program administration, and non-governmental data sources" (Hart and Potok, 2020, p. 3). As detailed in Chapter 2, statistical agencies around the world are already well along in the process of improving economic statistics while, in some cases, also reducing costs.

1.3. CHARGE TO THE PANEL

As outlined in the statement of task in Box 1-1, the panel was charged by BLS with examining the potential to improve the CPI by supplementing (or in some cases replacing) traditional survey-based data collection with an approach that blends multiple (survey and nonsurvey, government and commercial) data sources. Perhaps most obvious is the role of these data sources to improve elementary indexes (greater timeliness, frequency, and detail of price and quantity information), but it is also important to consider other stages of CPI construction, such as updating of upper-level category weights. The goal of this study is to assess new opportunities created by these alternative data sources to modernize and update CPI methods.

BOX 1-1 Statement of Task

An ad hoc panel of the National Academies of Sciences, Engineering, and Medicine will review measurement issues in the Bureau of Labor Statistics (BLS) Consumer Price Index program and provide guidance for its continued modernization. The study will examine the potential to improve CPI methodology by incrementally transitioning from traditional survey-based data collection to an approach that blends multiple (survey and non-survey, government and commercial) data sources. The panel will consider opportunities to apply new data sources to improve the construction of specific elementary item-area indexes as well as to improve index aggregation along several dimensions.

Many data sources have emerged during the past 20 years (since the last National Academies' review of the CPI) that could be used in the construction of the 7,000+ elementary item-area indexes in a way that improves the accuracy, timeliness, and detail of resulting price statistics, or reduces costs in the CPI program. The panel will identify specific areas where new kinds of data may be harnessed in a relatively straightforward way to improve price measurement of some items such as food and electronics. The panel will also propose solutions for some historically difficult-to-measure expenditure categories, particularly for which the availability of alternative data create opportunities for improved price measurement.

The panel will consider opportunities to use new data sources to improve aggregation of the elementary item-area indexes and also to mitigate upper-level substitution bias in the CPI-U and the CPI-W—for example, by taking advantage of the simultaneous availability of quantity and price information to update baskets and weights with shorter lags. As part of this task, the panel may revisit concerns about data sources used to estimate population item expenditure weights. The panel will also assess the prospects for creating new index aggregates that would present information about prices paid and expenditure weights for goods and services by households across the income distribution (by decile, or perhaps by quintile).

Finally, the panel will offer forward-looking thoughts about what price measurement may look like in 20 years and what BLS can do to anticipate future research and policy needs. As part of its information-gathering activities, the panel will gather input from data users, stakeholders, and survey experts. The panel will produce a consensus report with conclusions and recommendations.

In addressing the charge to the panel, the remainder of this report assesses the potential of alternative data sources to improve the CPI. Chapter 2 focuses on opportunities, challenges, and priorities of transactions (scanner) data from vendors or households, web-scraped data, and other options for improving timeliness and accuracy of elementary indexes. Chapter 3 considers the role of alternative data sources for use in the construction of higher-level aggregation phases of the CPI. As has been made evident by statistical agencies of several peer countries, nonsurvey sources have the potential to allow greater agility when consumer behavior changes rapidly. Chapters 4 and 5 detail how new data sources, combined with new methodologies, might modernize the measurement of difficult-to-measure expenditure categories. Housing services and medical services are, in terms of the link between data and methodological choices and index performance, the most important of these categories and are discussed in that order. The panel also reflected on how various data sources could illuminate price changes for products that experience rapid technological progress (quality change). In Chapter 6 the case for developing supplemental subgroup price indexes, and how the emergence of new data sources could advance that measurement goal, is presented. Finally, in Chapter 7, organizational recommendations pertinent to moving forward with data infrastructure planning are made.

1.4. LOOKING AHEAD

CPI modernization, on which BLS has already embarked, is an enormous undertaking. Indeed, BLS stated in 2019 that its goal is to have a "significant portion of the CPI based on alternative data within 5 years" (Konny et al., 2019) to improve the relevance, timeliness, transparency, and accuracy of the CPI. This work will almost surely require additional resources (and BLS acknowledged in 2019 that progress is contingent on the agency's budget and staffing). Nonetheless, as highlighted in the balance of this report, the panel believes that the case for data modernization in the CPI is urgent, compelling, and feasible.

2
The Potential of Alternative Data Sources to Modernize Elementary Indexes

The Bureau of Labor Statistics (BLS) has stated its strategic objective to "convert a significant proportion of the CPI [Consumer Price Index] market basket from traditional collection to nontraditional sources and collection modes, including harnessing large-scale data, by 2024" (BLS presentation to the panel, October 7, 2020). Implementing this goal will involve a paradigm shift in the way data inputs are evaluated in terms of fitness for use. Even with a continued role for surveys in the CPI data infrastructure (and in economic statistics broadly), BLS will have to take steps to reduce the program's reliance on the "full probability sampling" approach. The agency will also need greater flexibility regarding how closely new data and methods must replicate what has been done historically.

This chapter focuses on the CPI's elementary indexes, the most detailed item-location level at which prices are aggregated. Current methods are briefly reviewed, then alternative data sources—focusing on various types of scanner and web-scraped data—are assessed for their potential to improve the accuracy, coverage, and timeliness of elementary indexes. Challenges to implementing new data and new methods, of which BLS staff are keenly aware, are also considered.

2.1. CURRENT CPI METHODS AND DATA

The CPI's elementary indexes aggregate over groups of goods or services that are "as similar as possible" and that are varieties that may be expected to display similar price movements (IMF Consumer Price Index Manual, 2020, Chapter 8, p. 152). In the U.S. CPI, more than 100,000 items are sampled and aggregated into component indexes. These basic indexes estimate the average price change for each of 243 items (241 commodities and services plus 2 housing item categories) within each of 32 geographic index areas for 7,776 (32x243) area-item combinations.[1] Prices are collected for the more than 100,000 goods and services from about

[1]BLS's *Handbook of Methods*, updated November 2020, provides a complete description of how these data are coordinated in the construction of the CPI (www.bls.gov/opub/hom/cpi/).

23,000 retail establishments in 75 urban areas across the country.[2] Even with this extensive data collection, only a small portion of the many goods, services, or varieties within an elementary index can be sampled.

The outlets that BLS selects to sample for the CPI are chosen independently for each geographic area with a probability proportional to each outlet's reported expenditures from the Consumer Expenditure Survey (CE). The outlet sample is merged with an independent sample of items that consumers buy. The outlet and item samples are updated each year for roughly 25 percent of the item categories (or "strata") in each primary sampling unit (PSU). For most commodities and services, price collection from the selected outlets takes place *monthly* in only the three largest geographic areas (Chicago, Los Angeles, and New York); it is conducted every other month in other PSUs. Expenditure weights are assigned at the item strata level—in the case of the Consumer Price Index for All Urban Consumers (CPI-U), for 32 metropolitan areas. For example, men's shirts and sweaters sold in department stores in Chicago would be an elementary index. The price relative calculation for most of the item strata uses the geometric mean formula; the rest (most notably, for rent) continue to apply a Laspeyres-like formula in which the estimated quantities of the items purchased during the sampling period serve as weights.[3]

The above-described survey-based methodology, introduced in the 1978 CPI revision, had long been viewed as the "gold standard" for estimating price changes—and it has performed reliably for decades. However, as with other economic statistics rooted in the application of a 20th century survey-centric paradigm, the resulting estimates likely have become less precise over time, reflecting a number of factors. Among these factors are falling response rates and sampling errors.[4] Until 2019, the Census Bureau conducted a Telephone Point of Purchase Survey (TPOPS) on behalf of the BLS to identify the places where households purchased various types of goods and services, forming the basis for the CPI outlet sample. In 2007, approximately 13,500 households completed the survey each quarter; by 2017, the number had fallen to 8,600. To address declining response rates to random digit dial telephone surveys and mitigate the associated increase in data collection costs, BLS moved the main questions from TPOPS to the CE.[5] Therefore, in the current setup, information on where households shop (and, as before,

---

[2]Among these 75 primary sampling units (PSUs) are 21 "self-representing" areas with a population greater than 2.5 million, plus Anchorage, AK, and Honolulu, HI.

[3]BLS's *Handbook of Methods* (www.bls.gov/opub/hom/cpi/) details all stages of the process including how the sampling units and stratification variables are determined, as well as the procedure for selecting outlets and sampling items within outlets. Specifications for the geometric mean and Laspeyres formulas used can be found here: www.bls.gov/opub/hom/cpi/calculation.htm. Current weights for detailed item categories—e.g., gasoline = 3.181—can be found in the monthly CPI news releases: www.bls.gov/news.release/archives/cpi_04132021.htm.

[4]BLS publishes standard error estimates for all of its indexes. As described in the IMF manual (IMF, 2020), sampling errors "can be split into a selection error and an estimation error. A selection error occurs when the actual selection probabilities deviate from the selection probabilities as specified in the sample design. The estimation error denotes the effect caused by using a sample based on a random selection procedure" (www.elibrary.imf.org/view/books/069/01345-9789221136996-en/C11.xml). For fuller discussions, see White (1999), which describes the relationship between sampling error and bias estimates, and McClelland and Reinsdorf (1999), which examines the effect that small sample sizes have on indexes. They conclude that it has the effect of raising the expected values of an index based on nonlinear formulas, especially the geometric mean formulae and that more extensive use of large-sample scanner data sources may mitigate the problems.

[5]For an overview of potential biases, see the following BLS article: www.bls.gov/cpi/notices/2017/methodology-changes.htm.

about what they purchase) is obtained from the CE, which, in turn, is used to create the frame of specific outlets from which prices are then obtained and tracked. Unfortunately, over the last 10 years, response rates for the CE have likewise declined significantly. The CE-Interview unit response rate fell from 72.5 percent in October 2010 to 50.3 percent in December 2019, and the CE-Diary fell from 73.6 percent in October 2010 to 47.2 percent in December 2019. Response rates declined to even lower levels during early stages of the COVID-19 pandemic but have since bounced back some.[6] As response rates declined, concerns about the representativeness of the sample grew. Sabelhaus et al. (2015), for example, found that households at the very high end of the income distribution are less likely to respond to the CE.

Second, in addition to nonresponse, lags associated with surveys collecting information about what consumers buy and where are of particular concern as they create some well-known biases in the elementary indexes.[7] New item and outlet samples are selected on a continuous basis with about one-quarter of the sample updated each year. This means that there can be long lags before both new outlet types and new goods enter the CPI in a way that reflects recent changes in buying patterns. Although this problem was acute during the COVID-19 pandemic, it is not a new challenge. One earlier example of this problem was the emergence of lower-cost warehouse retailers from which sales were slow to be reflected in the elementary indexes. According to one study, this delay may have imparted an annual bias in the CPI of about 0.05 percentage points during the late 1980s.[8] More recent estimates of bias originating with outlet sampling are a bit higher (e.g., about 0.08 in Moulton, 2017), reflecting the increased popularity of online retail as a lower-cost option.[9]

Bias also emerges from lags in the introduction of new models or item varieties to the index, since an update of the sample of items to be priced occurs when the outlets are refreshed. Here, the bias is thought to be even higher (0.37 percentage points per year in both Lebow and Rudd, 2003, and Moulton, 2017) than that associated with outlet substitution. A key challenge is that, when samples are updated due either to "forced substitution" or overall sample refreshment, there are likely quality changes that, despite efforts by BLS to capture them, remain unaccounted for.

2.2. HOW ALTERNATIVE DATA SOURCES CAN IMPROVE ACCURACY, COVERAGE, AND TIMELINESS

The digital data revolution has given rise to the availability of information sources that—used in combination with, or in place of, existing surveys—have the potential to modernize price statistics. Alternative data sources that have been explored for price measurement purposes include point-of-sale data (obtained either directly from bricks-and-mortar or online retailers or

---

[6] www.bls.gov/covid19/effects-of-covid-19-pandemic-and-response-on-the-consumer-expenditure-surveys.htm.

[7] www.bls.gov/opub/btn/volume-1/pdf/consumer-price-index-data-quality-how-accurate-is-the-us-cpi.pdf.

[8] This finding by Lebow and Rudd (2003) was admittedly based on "only sketchy evidence"—a single study (Reinsdorf, 1993, 1998) of food and gasoline prices.

[9] Lags in updating the outlet sample can make the outlet sample unrepresentative as well. Outlet substitution bias, however, refers to something more specific: when procedures for bringing in the updated outlet sample assume that all price differentials between outlets are due to quality differences while, in reality, the outlets gaining market share tend to offer lower quality-adjusted prices (Reinsdorf, 1996).

from firms that aggregate the data), data generated from households scanning products at home, and data scraped from the web.

The various data sources amenable for use in price measurement differ in terms of the granularity and coverage of information they contain. Sometimes they only include prices, sometimes they include expenditures as well, and the amount of product detail that can be gleaned varies greatly. But all of these nonsurvey data sources, along with administrative sources, contain types of information that expand opportunities to develop a richer array of price indexes. For example, options for estimating representative statistics at subnational levels have emerged, and sometimes at lower cost than is possible when reliant on traditional establishment and household surveys.

Recognizing their potential, efforts are underway at BLS (and even more so at some peer national statistical offices) to exploit nonsurvey data sources. Transaction data in particular—generated in real time for the universe of goods with product identifiers and with information on the outlet, price, quantity, and characteristics of the item—have the potential "to make small sample sizes an issue of the past and reduce sampling error" (Konny et al., 2019). Such data breadth and detail may help address concerns about sample representativeness heightened by falling response rates in the CE and under-reporting by consumers of their expenditures.[10] As the CPI program moves incrementally away from the current probability sampling approach, standard errors will become less relevant as the metric for assessing data reliability. New types of measurement error will be introduced when traditional in-store collection of prices for a small sample of products are combined with direct electronic capture of large volumes of transaction data.

New data sources also offer the potential to reduce some types of bias resulting from data lags. The traditional sampling framework that, by design, involves less than universal coverage can cause delays in (1) identifying new goods appearing in the market (and quality change associated with those goods); (2) recognizing shifts in outlets frequented by consumers (which may create outlet substitution bias); and (3) updating lower-level weights to reflect the composition of purchases made (which may create substitution bias). Nonsurvey data do not necessarily solve these problems. However, where the pace of new product introduction and old product disappearance are rapid (dramatically highlighted in the COVID-19 economy), surveys will not illuminate trends until the resulting data are processed and incorporated months or years later. In contrast, the arrival and exit of goods is immediately seen in both scanner and web-scraped data when a transaction occurs or is posted online.[11] Likewise, when the places where consumers shop (including new and disappearing stores) are rapidly shifting, current CPI methods for sampling outlets can be inadequate.

---

[10]Problematic elements in the CE, including response rate issues, are documented in Carroll et al. (2015).

[11]Other kinds of corporate data might also prove useful in detecting consumer trends. For example, information on revenues and numbers of rides completed from ridesharing companies could have provided an indication of how fast they were displacing taxi services and other forms of transportation. The issue is that if Uber or Lyft rides are less costly than taxis (as they often, but not always, are), and if BLS tracks taxi prices separately from the rideshare companies, then the drop in the price of urban transit will not be picked up as riders switch modes. Quality differences between the two transit modes figure in as well. The problem is that the "best practices" way to deal with this bias assumes no quality differences (i.e., perfect substitution) between taxi and ride-sharing options.

Statistical agencies in other countries that systematically use transactions (and web-based) data in the CPI were able to provide timely information about shifts in consumer expenditures during the first COVID-19 lockdown in the spring of 2020. The Australian Bureau of Statistics (ABS), for example, was well positioned to handle disrupted access to outlets as most of their direct data collection was already conducted online or over the phone. In fact, less than 2 percent of the Australian CPI (by expenditure weight) is collected by field staff in retail stores.[12] By implementing computer-assisted data collection methods over the years, BLS has done an admirable job mitigating timing and accuracy problems with estimating price relatives for the elementary indexes; however, the agency was less well prepared for lockdown conditions due to its reliance on visiting retail outlets.

2.2.1. Scanner Data

Several companies specialize in producing commercial data based on either point-of-sale transactions from retailers (e.g., IRI Retail data on item-level sales at grocery stores and NPD data on consumer packaged goods) or data provided by households (e.g., item-level IRI Household data on grocery purchases, and Nielsen Homescan price and quantity information on packaged goods scanned by household). Integration of these data sources covering consumer transactions has been in the research (and occasionally) production pipelines of statistical agencies' price measurement programs for decades. Twenty years ago, a CNSTAT panel argued that "scanner technology has the potential to improve the entire process of data collection for the CPI computation" (NRC, 2002, p. 275). That study recommended research into how point-of-sale data could be used "both to select items for pricing and to replace the Commodities and Services Survey [where prices are actually sampled and recorded by BLS data collectors] and to quantify the improvement in the CPI." The report alluded to how household-based scanner data could be used "to record UPCs [Universal Product Codes] and quantities, along with key-entering prices and or store names and addresses" (p. 275). Going back further, Reinsdorf (1996) successfully constructed a basic item-level index for coffee using scanner data. A Conference on Research on Income and Wealth publication on scanner data and price indexes (Feenstra and Shapiro, 2002) also documented opportunities and challenges in using scanner data for the CPI.

Point-of-Sale Data from Retailers or Data Vendors

Scanner data offer simultaneous information on prices and quantities while also making it feasible to vastly expand the coverage of product varieties and outlets. In so doing, small samples of a handful of items to represent broad product categories can, in the process, become a constraint of the past. When obtained from aggregators, the transactions data cover many more retailers than do the samples used by BLS.[13] Many researchers have used point-of-sale data to estimate price indexes. In a recent example, Melser (2021) used Nielsen data covering a large

---

[12]ABS published "a series of notes" to describe the agency's "Methods Changes during COVID-19 period." See www.abs.gov.au/articles/methods-changes-during-covid-19-period.

[13]For a description of coverage of the Nielsen and NPD data, and of challenges created by enormous product turnover, see October 7, 2020, presentations to the panel, available at: www.nationalacademies.org/event/10-07-2020/docs/D124958ED038610E68986C71BEC8EA6D97CBF5F39C35.

share of U.S. national supermarket spending on 20 products; data were disaggregated by week, by store, and by universal product code or barcode.[14]

Their impressive coverage notwithstanding, data obtained from retailers typically require significant processing before they can serve as inputs into price index construction, as the literature using scanner-based methods also indicates. For example, the product codes used in scanner systems need to be matched (or classified) into entry-level item classes used by BLS. Similarly, care must be taken to ensure that the product codes used in the scanner data internally track identical items over time and across retailers.

When scanner data are obtained from aggregator firms, much of this processing is already done, reducing the production burden to a statistical agency. However, the procedures these firms use are proprietary, and it is often difficult to assess whether their classifications hold quality constant as would be needed for the CPI. Moreover, aggregators often calibrate their sample to industry totals from other official data, and information on those methods would also be needed to assess the quality of the data. Thus, effort would be needed to ensure transparency of alternatives data sources in the way that typically exists with the public collection of data. Finally, using aggregators as a data source has the added complication that the vendor could cease supplying the data or the cost of acquiring the data could become prohibitively expensive (the so-called "holdup" problem).

Household Scanner Data

In addition to point-of-sale data, some companies produce datasets containing information recorded by individuals on their purchases at home using a scanner provided by the vendor. For food at home, for example, both IRI and Nielsen have such panels, some of which have been used in price measurement research.[15] One important benefit of these datasets is that they provide information for purchases made at retailers that do not participate in their point-of sale programs. Perhaps more importantly, as detailed in Chapter 6, household scanner data contain information (e.g., demographics) on the characteristics of those participating in data collection that statistical agencies could potentially use to construct price indexes for different population subgroups. Such information is less important for construction of the headline CPI, which includes broader swaths of the population (e.g., all urban consumers). A potential problem with data obtained from consumer panels is that, as noted in Konny et al. (2019), the types of consumers willing to participate and spend time scanning purchases may not be representative of

---

[14]Approximately 32 U.S. retail chains supply data to Nielsen and on average, these chains have 22,870 separate stores. The average number of products in each of the 20 elementary categories was 6,031; on average, 1,690 of these varieties were available in any given month. On average, 196 cities are represented. The total number of price observations was 20x292,586 = 5,851,720. While the size of the Nielsen database is enormous, the number of missing products in any given month is also still enormous.

Nielsen data are made available to researchers through a collaborative arrangement with the Chicago Booth Kilts Marketing Center. Data subscription prices, for individuals and institutions, can be found here: www.chicagobooth.edu/research/kilts/datasets/nielseniq-nielsen/pricing.

[15]Hausman and Leibtag (2010) used the A.C. Nielsen Homescan consumer panel data to "identify the price differentials for twenty food product categories between supercenters, mass merchandisers, and club stores." In so doing, they estimated that, at the time, the CPI measure of inflation of food at home was too high because it failed to completely capture consumer gains from the growth of low-price, high-volume superstores.

consumers overall. For example, they may tend to be shoppers who value the incentives provided to participants.

BLS Experience Using Scanner Data

BLS has historically investigated the role of scanner and web-scraped data mainly as a way of obtaining price quotes, perhaps more easily than in-store price checking by field staff, within the current measurement framework. BLS initiatives incorporating scanner data in the CPI program have focused on the *food at home* category. A decade ago, BLS purchased historical Nielsen Scantrack data to support research comparing the performance of indexes based on the scanner data with those based on traditionally sourced data.[16] The goal of this work was to assess the feasibility of using the Scantrack data—which covered around 2 million UPCs—as a replacement for some food at home item categories in the CPI. The Nielsen data—which omitted convenience stores, bakeries, butchers, smaller grocery stores, warehouse stores, and gas stations—included product descriptors and average prices for each observation. At the time, because of the significant purchase cost of real-time data, BLS concluded that it was "less expensive to collect data in stores than to pay for Nielsen Scantrack for the real-time data and geographic and outlet detail needed to support the monthly CPI" (Konny et al., 2019, p. 17).

However, over the past decade, as in-store pricing and CE surveys have become less sustainable, BLS has improved its capacity to handle transactions data when a company's categorizations do not match CPI item categories. The agency has gained experience through its work with retailers—specifically, a department store (anonymized as CORPX) and a drug store (anonymized as CORPY). The process involves developing concordances between CPI item categories and those available in the alternative data source. In the case of CORPX, BLS has developed a machine learning system to assist in these categorizations, which has "greatly improved [their] ability to handle large datasets with hundreds of thousands of items" (Konny et al., 2019, p. 6). The CORPX data, provided monthly, include price and sales revenue information for each product sold in the store's outlets across the geographic areas covered in the CPI so that match-model price relatives can be estimated. Even though BLS does not refer to the source as "scanner data," the CORPX data are very similar to the data obtained through barcode scanning by national statistics offices in other countries such as the Office for National Statistics (ONS) in the United Kingdom.

Use of Scanner Data by Other Statistical Agencies

Statistical agencies in other countries and academic researchers have led the efforts demonstrating the feasibility of using alternative data sources to replace aspects of the existing sample-based structure for price measurement.[17] Across statistical agencies internationally, the

[16]The Nielsen data covered the period September 2005 to September 2010 and included totals for the quantity and dollar amount of merchandise sold by UPC, www.bls.gov/osmr/research-papers/2013/pdf/st130070.pdf.

[17]This section references only a small sample of the national statistical offices advancing the use of alternative data sources in their CPI programs. A more complete and detailed review of the use of scanner data and web-scraped data for price measurement—in this case focusing on outlier detection methods used—can be found at: www.niesr.ac.uk/sites/default/files/publications/NIESR%20DP%20523.pdf.

motivations behind this work have been diverse, ranging from cost containment to the need to more quickly capture effects associated with the arrival of new goods and outlets, or the changing composition of spending patterns. For BLS, lessons learned from these efforts will have to be adapted to the unique legal and budgetary context of the United States and as discussed in Chapter 1, to the decentralized nature of the statistical system.

Most agencies that have used scanner data have done so, initially, as an alternative for price quotes. For example, in Australia, ABS implemented a three-stage model for integrating transaction data that involved: (1) replacing in-store price quotes by (unit value) prices from transactions datasets without changing methods or samples; (2) enlarging the samples; and (3) using the "universe" of products and implementing new methods. The ABS project took several years, and the third stage required overhauling the agency's IT system. The UK ONS followed a similar timeline: (1) researching the methods required to process alternative data sources in 2020; (2) developing systems for processing alternative and traditional data sources in 2020–2021; (3) conducting index impact analyses for priority items in 2021 and a parallel run to produce experimental aggregate measures planned for 2022; (4) estimating aggregate measures of consumer price statistics in 2023; and (5) rolling out the use of alternative data sources to new items within the inflation basket in 2024 and beyond.[18]

The methodological changes resulting from these research efforts have been dramatic. For example, Statistics Canada reported that, as of March 2021, 50 percent of collected prices originate from alternate data sources (which encompass more than scanner data), representing 20 percent of the CPI Basket Weight. The agency is aiming to collect 70–80 percent of its price quotes from alternative data sources, representing 55 percent of basket weight, by March 2023.[19]

ABS uses scanner data from retailers to obtain prices for about 16 percent of Australia's CPI by item weight. Covering approximately 84 percent of all expenditures at supermarkets, these data offer nearly a "census" of sales at these outlets. The data include product descriptions as well as information on quantity of items sold, dollar value of items sold, and geographical location.[20] Scanner data enabled a chained formula to be constructed for that portion of the CPI as well. As discussed in the next section, ABS also uses web-scraped prices for about 5 percent of the CPI by item weight (alcoholic beverages, clothing, and car parts are major categories) and administrative sources for another 22 percent (electricity, gas, childcare, fuel, pharmaceuticals, and insurance).[21]

The ONS in the United Kingdom is likewise moving forward with incorporating point-of-sale transaction data from retailers, along with web-scraped and administrative data. As has been

---

[18]www.ons.gov.uk/economy/inflationandpriceindices/articles/introducingalternativedatasourcesintoconsumerpricestatistics/may2020.

[19]October 7, 2020, presentation to the panel by Heidi Ertl, Director of Consumer Prices, Statistics Canada, www.nationalacademies.org/event/10-07-2020/improving-cost-of-living-indexes-and-consumer-inflation-statistics-in-the-digital-age-meeting-6.

[20]www.abs.gov.au/statistics/research/recent-applications-supermarket-scanner-data-national-accounts.

[21]October 7, 2020, presentation to the panel by the Australian Bureau of Statistics. An overview of methods used to incorporate scanner data into the ABS's multilateral CPI framework—including how the agency has gone about implementing data changes (e.g., communication with external users, research conducted, input from international experts)—can be found here: www.abs.gov.au/AUSSTATS/abs@.Nsf/39433889d406eeb9ca2570610019e9a5/40fc971083782000ca25768e002c845b!OpenDocument.

the case for other countries, one important research task has been to map item classification. This research focuses on ensuring that the right products are in place for the various datasets to produce an index for specific items as defined in the ONS CPI.[22]

While scanner data provide a useful source of price quotes to fold into the existing production system, such data also contain information on quantities that may be used to estimate elementary price indexes directly. Transaction data can be aggregated by expenditures and quantities sold at the UPC level to form a unit value that serves as the price; scanner data from aggregators are already aggregated to the UPC level. The resulting price and expenditure data can be used to generate superlative indexes (Ehrlich et al., 2021) or to obtain hedonic price indexes.

Indeed, a number of national statistical offices are at comparatively advanced stages of their data modernization programs, bypassing the current survey-based production system and calculating price indexes directly from alternative data sources. Statistics Norway began research to use scanner data to compute the subindex for food and nonalcoholic beverages in 2005. Statistics Netherlands introduced supermarket scanner data into their CPI in 2002 (described in Chessa, 2016; de Haan, Willenborg, and Chessa, 2016). Beginning around 2008, Statistics New Zealand began researching use of scanner data to directly estimate price change for products sold by supermarkets and for consumer electronics.[23] Their research focused on overcoming, with the use of scanner data, volatility of prices and quantities, due to discounting, seasonality, and the churn of new products entering and old products leaving the market.[24]

2.2.2. Web-Scraped Data

Scanner data are not available for all commodities. For some items, notably goods purchased online or goods where one firm dominates the market (e.g., Apple smartphones), scraping price data available on the internet provides an alternative to the traditional survey-based methods. Web-scraping refers to the process whereby price and product information is collected automatically from websites on the internet using software that simulates human web-surfing activity. The objective is to transform unstructured website data into structured data for CPI construction (or other) purposes. The main drawback with the use of web-scraped data for official price measurement is that while prices of available products are known and can be measured almost continuously, methods are lacking for establishing their relative importance in the consumption basket.[25] This means that it is not possible, using web-scraped data alone, to

[22]Details of the UK ONS experience experimenting with multilateral indexes for scanner data can be found in "Using alternative data sources in consumer price indices: May 2019" www.ons.gov.uk/economy/inflationandpriceindices/articles/usingalternativedatasourcesinconsumerpriceindices/may2019.

[23]Statistics New Zealand has since implemented a hedonic multilateral method for consumer electronics based on scanner data purchased from market research company GfK. They have not (yet) implemented scanner data from supermarkets.

[24]unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2014/WS4/WS4_11_New_Zealand_CPI_scanner_data.pdf.

[25]There is, however, some price measurement research geared toward approximating expenditure weights for web-scraped data. See, for example, Thomas and Ayoubkhani (2019) along with foundational work to model sales quantities by Chevalier and Goolsbee (2003).

construct the superlative indexes that are viewed as a superior approach to constructing index numbers.[26]

The most prominent U.S. player in the web-scraping data collection space is not a statistical agency, but MIT's Billion Prices Project and spinoff company PriceStats. PriceStats currently tracks about 25 million prices per day from 1,100 retailers in 50 countries. In the United States alone, it collects 2 million prices per day in real time on a daily basis from not only online retailers such as Amazon.com, but also from the websites of traditional and large multi-channel retailers that sell both online and offline.[27] Product categories include food and beverage, clothing, housing, recreation, household products, and health. Among the data elements collected are price, product description, and product attributes. The country-by-country inflation series published contain daily averages of price changes across multiple categories and retailers, by sector. PriceStats has the kind of expertise collecting and processing online data in a production environment that is similar to what BLS would need to set up if it plans to emulate the approach to construct some of its elementary indexes. Even so, academic research of price measurement does not produce "official statistics" and so there is greater freedom to delve into experimental methods that are only suggestive of approaches that BLS could consider.

Unlike scanner data, the web offers listings of prices (not prices actually paid) with no information on the relative importance of the different offers. An important benefit of web-scraped data is that it is more accessible and is often easier and certainly quicker to obtain than are data from retailers (which can take months or years). Additionally, the approach offers real promise in addressing the timeliness problem—if data processing can be automated, time lags can be almost eliminated.

BLS Experience with Web-Scraping

Until recently, BLS had only used web-scraping in the CPI for research purposes and to collect supplemental observations used in constructing hedonic models (Konny et al., 2019). However, beginning in March 2020, due to initial COVID-19 shutdowns, BLS had to improvise as the monthly, in-person collection of price data from retailers and businesses by field staff came to a halt. Price checkers, who could no longer go to stores, had to switch to filling up virtual carts online to check prices. This process (brought on as a stopgap measure in the face of the immediate crisis, and not fully web-scraping) mimicked the in-store price checking activity, but it would need to be automated (perhaps using PriceStat methods as the model) if timeliness and efficiency gains in data collection are to be realized.

A first step toward integrating web-scraped price data involves performing research to assess the extent to which pricing is the same in-store and online and whether the two sets of prices move in a highly correlated fashion.[28] BLS has some experience with this kind of work when it researched comparisons between CPI's current data collected on the price of motor fuels and web-scraped data from a tech company that crowdsources fuel prices from around 100,000

[26]See the appendix to this chapter on the use of multilateral methods for blending alternative data sources, including web-scraped data, to estimate price relatives.

[27]Similar data can also be found at https://www.pricestats.com/approach/data-composition.

[28]Cavallo (2017), investigating the similarity of online and offline prices using evidence from large multi-channel retailers, found that "price levels are identical about 72 percent of the time."

gas stations across the United States. Preliminary research showed that a Jevons price index based on these data performed almost identically to the conventional CPI's gasoline index, despite the fact that the data were not weighted (at the time of this research, the CPI used TPOPS to weight gas stations) (Konny et al., 2019). BLS has also been engaged in this kind of research, for example, on residential telephone and telecom services and airline prices where, for the latter, web-based pricing "enables the CPI to track a more defined trip month-to-month," according to a BLS fact sheet.[29]

A key question for BLS's web-scraping research is to investigate how alternative data sources can be used to weight products/clusters in the absence of expenditure and quantity information. Questions include: What do the "weights" of the data source or the retailer represent? In turn, how should the price indices within and across data sources be aggregated? (Claude Lamboray, Eurostat, presentation to the panel, October 7, 2020). These limitations of web-scraped data suggest a blended approach. Scanner data can be used to establish weights for some categories. Alternatively, BLS may be able to collect useful information by contacting retailers and asking about their bestselling products in different categories. This is currently done by the price inspectors when they visit a physical store, so it would not mark a drastic change in approach. The advantage of this kind of blending would be that the weights could be obtained at low frequency (e.g., once per quarter or semester), while the scraping provides data at higher frequencies.

Work at Agencies Internationally

Relative to BLS, a number of national statistical offices in other countries have pushed forward more aggressively with web-scraping in their CPI programs. These initiatives, which typically focus on repricing products already in the index, have been motivated in part by the increased share of retail spending that is being transacted online and the need to monitor prices for these outlets. As highlighted in two studies of the UK, the *UK Consumer Price Statistics Review* (Johnson, 2015) and the *Independent Review of UK Economic Statistics* (Bean, 2016), opportunities abound to improve the efficiency and quality of collection methods. On the quality side, price data from the web can be collected in a timelier manner than is possible when relying on surveys or third-party scanner data to be processed. On the cost side, web-scraping can automate price collection for some goods and services, which can potentially reduce costs and increased coverage.

Statistics Belgium scrapes around 6 million prices per month in categories such as clothing, footwear, hotel reservations, airfares, international train travel, secondhand cars, consumer electronics, books, and videogames. Several of these categories are already incorporated into the country's official CPI (Kevin Van Loon, Statistics Belgium, presentation to the panel, October 7, 2020).

ABS has been incorporating web-scraped prices progressively into its CPI since March 2017, currently using primarily a direct replacement strategy. It is the primary data approach for some significant item categories such as alcohol and tobacco (7.3 percent weight), clothing and footwear (3.5 percent), furniture and household equipment (3.7 percent), and recreation and

[29]See BLS Fact Sheet: www.bls.gov/cpi/factsheets/airline-fares.htm).

culture (12.7 percent). Much of the current web-scraping is currently carried out manually but the agency is moving forward toward automation. The agency is now looking into the potential for Application Programming Interfaces to access pricing information that can be more straightforward than maintaining web-scraping code over time.[30]

2.3. FUTURE DIRECTIONS

2.3.1. Challenges

The most obvious obstacle to past efforts to incorporate alternative data sources into the CPI has been the (sometimes) prohibitive cost of acquiring scanner data, particularly those obtained from aggregator firms. As noted above, this has been cited by BLS as the main reason for not moving more quickly to replace in-store price quotes with scanner data from commercial firms (Konny et al., 2019). Even if affordable, using data processed by a third party involves uncertainty about how the data were compiled and processed.

On the methodological front, a pervasive problem with alternative data sources compiled by aggregators is that the data are collected for "nonstatistical" purposes and are not necessarily representative. Many commercial data sources containing price and expenditure information useful for price measurement rely on convenience samples that have coverage patterns that differ from those in currently used sources. For example, an Economic Research Service study by Levin et al. (2018) assessed how well totals from the (unweighted) IRI scanner data for food align with data based on other sources, including products from the Census Bureau (e.g., the Economic Census and County Business Patterns). The researchers found differences that suggest the data would likely benefit from the construction of post-stratified survey weights.

Typically, any adjustments made by the vendor to achieve a representative sample (controlling to totals, weighting, etc.) are not transparent. More generally, statistical agencies do not control the creation and curation of the data and aggregated datasets are manipulated before they are made available. Vendors have different priorities than national statistical offices (NSOs) so may make adjustments that are useful for their own purposes but not so helpful for NSOs. Often, there are no clear incentives for providers to be transparent about methods/changes. Not knowing what scanner data aggregators have done during production of data—there is often a lack of documentation or transparency—is a major shortcoming for use in production of official statistics.

Aside from concerns over representativeness, coverage, and other issues described above, statistical agencies' experience with scanner sources has also revealed methodological challenges that must be tackled. In particular, indexes constructed using high-frequency scanner data can suffer from a "chain drift" problem that introduces biases in the indexes. However, new approaches have been developed (such as multilateral methods, described in the appendix to this chapter) and adopted by some statistical agencies to deal with the chain drift problem.[31]

---

[30] www.abs.gov.au/articles/web-scraping-australian-cpi

[31] Index chain drift is defined by the difference in the performance of a fixed base price index and a chained index (Klick, 2017). Chain drift can trend upward, as found by Feenstra and Shapiro (2003) in a study using scanner data on canned tuna to compile a weekly chained Törnqvist index. It can also trend downward, as found by de Haan (2008) in a study using scanner data from a Dutch supermarket chain on detergents. The international price statistics

Web-scraped data also present challenges. The array of issues that require attention before these data can be routinely integrated in the official CPI are summarized in Table 2-1, reproduced from Auer and Boettcher (2017). Nonsurvey data—whether from retailers or from the web—are typically organized using hierarchies that do not always line up easily with the CPI nomenclature so concordances must be constructed to bridge categories in the new data source to the CPI. Another issue common to both sources is that raw price quotes typically contain outliers so that consistent and transparent methods must be applied to avoid undue volatility in the resulting price indexes.[32] Some of the challenges listed depend on the type of retailers being scraped. For example, issues with the relevance criterion ("are products offered really sold and by whom") can apply to data scraped from online marketplaces such as eBay and Walmart Marketplace Sellers where individual sellers can publish postings; BLS can control this by scraping only goods on, as an example, the Walmart.com website and exclude marketplace sellers altogether.

To extract all available metadata from websites also requires continuous monitoring. For example, scripts need to be adapted when websites change to avoid periods without data.[33] In so doing, federal agencies need to obey legal restrictions on individual websites, such as terms of use. Statistical agencies have begun to grow out staffs with the right skill set to carry out these processes.

TABLE 2-1 Novel Quality Problems and Measurement Methods with Web-Scraped Data

Input data quality criteria	**Web-scraped data**	
	Novel quality problem (for consumer price statistics)	**Measurement method**
Relevance	Representativeness of online data (are products offered really sold and by whom?)	Information by data providers; otherwise unresolved
Accuracy	Website content may be IP-specific (a user who frequently checks a website or a web- scraper might lead to different price displays than first-time users)	Comparison of automatically and manually collected data
Timeliness/Punctuality	The amount of data makes it difficult to judge data quality within a reasonable amount of time	Quantitative instead of qualitative processing of data
Accessibility	Websites might identify web-scrapers and block them	Unresolved

community appears to have reached a consensus that multilateral methods, such as those proposed by Ivancic, Diewert, and Fox (2011), offer an approach that provides drift-free, superlative-type indexes (Kalisch, 2017).

[32]For a detailed description of outlier detection methods for alternative data sources used in price measurement, see www.niesr.ac.uk/sites/default/files/publications/NIESR%20DP%20523.pdf.

[33]In a presentation to the panel, Alberto Cavallo and Pilar Iglesias (PriceStats) illustrated how their company addresses this issue: https://www.pricestats.com/approach/data-composition.

Completeness	Websites change frequently. Relevant variables and URLs might not be identified and scraped	Number and level of target values are measured against historical values from previous data collection activities
Clarity / Interpretability	No new quality problem	

SOURCE: Auer and Boettcher (2017). Reprinted with permission.

Finally, replacing traditional price collection with data obtained from vendors could lead to dependency of the statistical agency on the data providers; even with strong contract provisions, these data could be changed or discontinued without notice. In the future, it may be possible for agencies to set up their own scanner data and web-scraping operations, but such a system is some ways off. The more immediate tasks would be to set up contracting arrangements that make sense for both BLS and data providers, ensure confidentiality given the sensitivity of the data, set up arrangements that ensure reliability of sources, and create contingency plans in the case of disruptions in the supply of CPI input data.

As it moves toward a new paradigm for data quality assessment (Box 2-1), BLS will be able to draw from quality evaluation frameworks developed elsewhere. One example is the framework developed by the Statistical Office of the European Union (Eurostat, 2013), which includes five major output quality components: relevance, accuracy and reliability, timeliness and punctuality, accessibility and clarity, and coherence and comparability. In practice, BLS will continue to perform very granular comparisons to "validate" new data sources by comparing indexes estimated from them with those estimated in the official index.[34]

BOX 2-1 Assessing Data Quality

At the core of efforts by statistical agencies to broaden the sources of information used for purposes of economic measurement is the need to assess the quality of new data. Historically, data quality assessment at statistical agencies has been focused on response rates and variance estimation appropriate to the survey design. However, because the movement by statistical agencies away from a survey-centric system is undeniably underway, and because quality standards must still be maintained, a broader approach is needed. Indeed, a previous expert committee recommended that federal statistical agencies "adopt a broader framework for statistical information than total survey error to include additional dimensions that better capture user needs, such as timeliness, relevance, accuracy, accessibility, coherence, integrity, privacy, transparency, and interpretability" (NASEM, 2017, p. 117).

[34]The Census Bureau has performed similar exercises with NPD scanner data, comparing store-level revenue data to that reported in their trade surveys. As expected, coverage is a major issue. The Census Bureau only purchased data for stores that were "most relevant" for their purposes and, currently, have 20 retailers (each with many outlets) that have given NPD permission to provide the data to Census (for details, see www.nber.org/system/files/chapters/c14270/c14270.pdf).

For most kinds of nonsurvey data, there is little in the way of "agreed-upon techniques for assessing the validity, reliability, and robustness of the inferences made" (NASEM, 2020, p. 128). That said, increased attention is now being given to measuring the quality of administrative and commercial data, and how that quality compares with currently used survey sources. Statistical agencies are being pushed to move beyond frameworks such as Total Survey Error (TSE), which parses potential sources of error and variance broadly into sampling and nonsampling errors (Biemer, 2010; NASEM, 2020, pp. 129–130). TSE metrics of precision (the basis of current quality assessment) are highly focused on response rates and, thus, not relevant for evaluating alternative types of data (e.g., scanner, web-scraped) that are and will become increasingly useful in CPI construction.

An example of a more expansive framework is the Total Error Framework (TEF), which broadens the nonsampling error component to include measures of error associated with commercial and other types of data and suggests methods for comparing errors in big datasets to errors in survey datasets.[35] An important aspect of data quality for transactions data like scanner data and payments data is gaps in the coverage of the population of interest. The set of missing observations due to nonreporting may change from month to month, and information on which observations are missing from the dataset may be unavailable or incomplete. This can make it challenging to control for coverage changes.

2.3.2. Opportunities

Overall Strategy for Integrating Alternative Data

To date, transaction data have typically been integrated incrementally within BLS's existing CPI infrastructure opportunistically or when pressure to do so has arisen because of a problem with a conventional data source. One notable exception is BLS's use of the JDPOWER data on transaction prices and real-time expenditures for light vehicles, which does not rely on the usual sample-based methods for selecting outlets and vehicles to price. The most common application among statistical agencies has been to match and replace price quotes previously obtained by field staff at outlets with electronic point-of-sales data. Going forward, BLS will need to progress in areas where reliable data may already be present, but where benefits in terms of cost, detail, or accuracy may emerge from pursuing alternative sources.

> **Recommendation 2.1:** BLS should embark on a broad-based strategy of accelerating and significantly enhancing the use of transactions data and other alternative data sources in CPI compilation. Embracing alternative data sources now, and moving forward aggressively with research for their integration, will ensure that the accuracy and timeliness of the CPI will not be compromised in the future. The data modernization strategy will involve:

---

[35]For a description of this framework, see *Total Error in a Big Data World: Adapting the TSE Framework to Big Data* (academic.oup.com/jssam/article-abstract/8/1/89/5728725?redirectedFrom=fulltext). For a broad-based discussion of quality assessment frameworks for statistics using multiple data sources, see NASEM, 2017, Chapter 6.

- Identifying promising alternative data sources and then prioritizing the work needed to evaluate and incorporate these data into the items/strata where they can be applied;
- Continuing development of a robust research agenda that supports incorporation of alternative data and associated new methodologies more broadly beyond just price quote replacement;
- Continue research assessing the quality of new types of data;
- Developing staff expertise that includes more data scientists and other specialists;
- Creating a cross-agency strategy for gaining access to data—from third-party providers and, if possible, direct feeds from the largest retailers—with the possibility of joint contracts across statistical agencies;
- Carrying out a strong communication strategy to inform stakeholders of plans and implementation details.

The kind of data modernization envisioned will require upfront new investments in data acquisition, updating of production procedures and IT systems, and staff training. BLS analysts have extensive expertise for conceptualizing and measuring different error sources in conventional data sources but, for nonsurvey data, "expertise and training is also needed in computer science for processing, cleaning, and linking datasets and the errors that can arise in these operations" (NASEM, 2017, p. 127). In the future, CPI staff skills will need to shift (at least partially) away from those needed to obtain structured price information and toward those needed to process unstructured price data.[36]

After these initial investments, once the agency transitions into a routine maintenance phase, cost savings are possible—particularly as transaction and online data allow a shift from labor-intensive manual (field-based) to automated data collection processes.[37] Even if savings are not guaranteed, BLS should not be deterred. Given its wide use by markets and in policy making decisions, the primary objective should be production of an accurate CPI. BLS will need support in the funding process so that near-term costs do not obscure the potential longer-term benefits of developing new data sources. While BLS has certainly made progress using transaction data to replace price quotes, the agency has the opportunity to go much further.

Recommendation 2.2: BLS should accelerate its research identifying alternative data sources that could potentially be integrated to replace price quotes collected within the current framework. As part of a proactive plan for modernizing the data infrastructure used in elementary index construction, BLS should develop, apply, and communicate clear criteria for identifying and prioritizing new data sources for various item categories of the CPI.

[36]Auer and Boettcher (2017) included a detailed discussion of the ways in which price statisticians must re-think index compilation procedures when using web-scraped and scanner data. See, specifically, the section on "Assuring Data Quality of Large New Data Sources." This report lists specific skill areas that need to be covered when migrating toward alternative data-based price measurement programs. These include expertise in big data platforms, analytics engines and programming languages, visualization and reporting applications, data warehousing, security frameworks, web crawling tools, and storage infrastructure.

[37]The field-based labor force for the CPI program includes around 80 full-time and 425 part-time data collectors working in 75 cities in 43 states.

As documented above and in the many references cited, the potential for an expanded role for scanner data in particular has been broadly recognized, including by BLS. Based on documentation from the companies NPD and IRI Nielsen, Table 2-2 provides a sense of the item coverage in some of the large scanner datasets. These product categories are quite broad and sometimes exclude items—for example, NPD does not collect data on cell phones. And even scanner datasets that have good item coverage do not always have comprehensive retailer coverage. For example, Home Depot and some other home improvement stores do not participate with NPD; prior to 2011, Walmart did not participate with IRI and Nielsen, and the data vendors had to visit the retailers (just like BLS does) to get pricing.[38]

TABLE 2-2 Potential Scanner Data Coverage of CPI Items

CPI ITEMS	RELATIVE IMPORTANCE	
	All CPI items	Available in scanner data*
Food	14.2	7.9
Energy	5.8	0.0
Commodities less food and energy commodities	20.3	12.1
Total	40.2	20.0

NOTE: Totals are based on Nielsen and IRI point-of-sale data and Homescan data, and NPD data.

The most obvious limitation of point-of-sale scanner datasets is that their coverage is constrained to goods only (packaged goods, actually) so services, which amount to about 60 percent of the CPI, are not covered. This means that if scanner data cover about half of the CPI relative importance for goods, the total amounts to a bit less than one-fourth of the overall CPI. However, the missing goods are mainly vehicles, nonpackaged food, and energy, where other alternative data sources may be helpful. For food, the Homescan products (done by consumers at home, not the store-based POS data) provide full coverage of retailers that could be used to fill some gaps. Likewise, a big advantage of web-based data is that they include many types of services.

Assessing Data Fitness for Use

Paramount among the challenges in shifting away from traditional data collection is evaluating the quality of the new replacement data. The CPI samples have traditionally been designed with the goal of representative coverage of the full population of goods and services consumed by U.S. urban households, although in practice there have been some gaps in coverage, such as for new goods. New data sources may offer tradeoffs in which there is improved coverage along some dimensions (for example, the number of items or geographic coverage) that must be traded off against reduced coverage along other dimensions (for example, the number or variety of outlets). In some instances, large, quickly available samples that were not designed with representativeness in mind may be preferrable to small samples that were

[38]www.wsj.com/articles/SB10001424053111904233404576460164032135744.

designed to be representative, particularly if they are not timely. A sample representative of the population five years ago, for example, may not be that useful today. Unfortunately, assessing data quality tradeoffs along these dimensions is not simple.

BLS has already stated that, for purposes of expanding the number of expenditure categories to which alternative data sources could be applied, their priority will be "based on factors such as index quality issues, relative importance, size of sample, alternative data source availability" (Konny et al., 2019, p. 25). Other important data characteristics include the following: detail of product coverage,[39] geographic coverage (there is a question of whether BLS should scale back on geographic sampling, especially for items for which activity is shifting online), capturing transaction (as opposed to list) prices, timeliness and frequency, and nature of sample (random versus convenience, census versus subset). Some of the criteria for evaluating data quality—perhaps especially timeliness and other dimensions of granularity—have often been undervalued as indicators of quality but are "increasingly more relevant with statistics based on multiple data sources" (NASEM, 2017, p. 117).

> **Recommendation 2.3:** In the context of CPI construction, which will increasingly rely on data blended from multiple sources, BLS should regularly publish information on the characteristics of alternative data they plan to incorporate. Important quality indicators include the following: number of products covered, number of observations/price quotes, type of price quote (listed price, transaction price, etc.), how many matches of products can be made across periods, extent of coverage within and across expenditure categories, frequency of updates, and level of product detail.

This kind of documentation is a component of the transparent communication strategy identified in Recommendation 2.1. A National Bureau of Economic Research paper (Konny et al., 2019) is a great example of what is needed, but on an ongoing basis—perhaps every six months.

Developing Parallel Series

A strong research program must accompany the transition to a mixed data infrastructure for the CPI.

> **Recommendation 2.4:** BLS should accelerate testing of indexes constructed from alternative data sources and new methodologies. Before BLS incorporates alternative data for specific item categories into the official CPI, it will be important to maintain a significant overlap period (perhaps as long as two years) during which parallel indexes based on new data sources can be tested and compared against their traditionally constructed counterparts.

The overlap period also allows significant changes to CPI methodology and data sources to be vetted with the public and user communities. BLS might also consider a comment period for particularly important changes to methodology.

[39]Product coverage and completeness testing are particularly important given the substrata approach that BLS is thinking of adding, as is capacity to capture rapid item disappearance and appearance or churn.

An illustrative example of parallel series can be found with BLS's own work (described above) using comprehensive transactions data from a department store that included an assessment of how the CPI would have performed if those data had been used in an earlier period. Currently—while the nonsurvey data and survey-centric worlds still very much overlap—statistical agencies have the opportunity to make these kinds of comparisons. If current surveys become obsolete (due to costs, deteriorated response rates, etc.), the opportunity to test parallel series will be lost.

Multilateral Methods; Measuring Quality Change

BLS will no doubt continue its research using already obtained scanner data sources as a laboratory to test how the data perform and methods for blending those data in a way that is statistically sound and useable in the CPI program. The panel recommends that BLS develop a robust research agenda that supports incorporation of alternative data more broadly beyond just price quote replacement. This will require accelerating research evaluating the role of the leading multilateral index approaches designed to maximize and automate the use of alternative data sources for the construction of new elementary indexes that do not require the usual survey-based paradigm (the methods are described in the appendix to this chapter).

> **Recommendation 2.5:** BLS should prioritize experimenting with and getting up to speed on the use of multilateral indexes for scanner data and web-scraped data.

One question that BLS will confront is whether the leading multilateral index approaches (especially those already in use at NSOs) can be applied and used in real-time without the need to revise the indexes. BLS can benefit extensively from the work already done on the topic by Statistics Netherlands, Statistics New Zealand, and ABS.

A key element of the multilateral research program involves assessing the capacity of alternative data sources to identify product attributes and apply quality change estimates. Where rapid quality changes are common, such as for high-tech items, combining datasets that include product codes and identify product characteristics in detail provides rich opportunities for improving measurement. Such data can sometimes be extracted from retailers' or manufacturers' websites to perform quality adjustments.[40] In these cases, alternative data sources can be incorporated into work on hedonic methods, such as those developed by Erickson and Pakes (2011), to adjust for unobserved characteristics and that correct for sample selection effects.[41] These approaches will require gaining expertise in multilateral indexes so that they can continue to be evaluated as they develop.

> **Recommendation 2.6:** A major component of BLS's research effort to experiment with using scanner data and web-scraped data should be assessing their potential for quality change adjustments. Initially, this work could be part of an effort to replace price quotes from traditional data, though ultimately the use of new alternative data likely will lead to

[40]See, for example, Bajari et al. (2021), which uses product descriptions from Amazon to estimate hedonic price functions and, in turn, Fisher price indexes for the period 2013–2017.

[41]As described in the appendix to this chapter, recent research has attempted to perform quality adjustment at scale, often with the use of scanner data.

the need for new methodologies for adjusting for quality change. Top priorities should be items with large expenditure shares and items undergoing rapid technical change.

Methodological improvements along these lines could be consequential when high expenditure items are involved. Accordingly, communications goods and services (internet, streaming, mobile, and cable) is a good example—and appear to be near the top of BLS's priorities.[42] Statistics agencies in other countries have likewise been exploring the value of alternative data sources for measuring quality change.[43]

Automating Web-Scraping

As noted above, during the COVID-19 shutdowns when monthly in-store collection of price data was not possible, price checkers had to switch to filling up virtual carts to check prices.[44] This process—brought on as a stopgap measure in the face of the immediate crisis and which only mimicked the in-store price checking activity—needs to be automated.

Recommendation 2.7: Converting opportunities for permanently automating web-scraping of price data should be a high priority for the CPI. In evaluating the usefulness of web-scaped data for elementary index estimation, food, electronics, and apparel should be priority categories. Data for these categories are readily available with a large share of transactions already online, and work by other statistical agencies and private-sector organizations have demonstrated feasibility. In the short term, BLS could consider obtaining web-scraped data from outside vendors, but ultimately BLS should develop automated web-scraping methodologies within the agency. As progress is made, internet and traditional outlet prices can be compared during a testing period.

Automated methods similar to those developed by PriceStats should be adopted for processing web-scraped data.[45] As alluded to above, most of the data scraped by PriceStats comes from the websites of companies that sell both online and offline. This is important for BLS since it means that the same retailers that its current price inspectors visit physically can also be web-scraped. BLS could focus first on these retailers so that the only thing that changes is the "channel" from which the data are collected, not the retailer type. Later on, BLS can consider scraping online-only retailers, which are likely to have more differences in pricing

---

[42]Brown, Sawyer, and Bathgate (2020) review the "directed substitution approach" used in the CPI for smartphones and the hedonic models used for quality adjustment of telecommunications services. The directed substitution method for smartphones, which BLS began using in the CPI in 2018, rotates in quality-adjusted prices of new models every year, or even every 6 months.

[43]For example, in work measuring price change for consumer electronics using scanner data, Statistics New Zealand has been employing time-dummy hedonic models. www.stats.govt.nz/methods/measuring-price-change-for-consumer-electronics-using-scanner-data. See also the appendix to this chapter on multilateral methods for a discussion of quality-change measurement in the context of scanner and web-scraped data (Léonard, Sillard, Varlet, and Zoyem, 2015).

[44]At the same time, response rates to the Commodities and Services Pricing Survey and to the Housing Survey also dropped off.

[45]PriceStats has already begun collaborating with other statistical agencies about how to operationalize web-scraping in CPI programs. For example, the company has shared data with the UK's ONS and several other (smaller) agencies during 2020; they also have a long-standing contract with Statistics New Zealand.

behaviors. A big opportunity exists for BLS to realize that it can sample the same retailers it does today but using a new technology.

While collection and processing of transaction data can be difficult for a statistical agency to perform internally, a staff with the appropriate skills could soon be web-scraping much more easily. They would have to replicate what firms like PriceStats are doing. Ideally, to become viable for production use—which includes the need to maintain public confidence in the data and to tailor the program specifically to CPI specifications—this capacity would be developed in-house at BLS. During the research phase, however, while internal expertise in web-scraping methods is still being developed, BLS will likely need to contract with outside experts. Likewise, this research will benefit from interaction with other NSOs and measurement economists working outside of statistical offices.

For construction of elementary indexes, item categories of the web-scraped price data also must be mapped to the item strata as defined by statistical agencies. This mapping needs to be automated in the production process—a task at which supervised machine learning methods excel—and extensive data cleaning and maintenance will be needed to keep up with changing websites. Some of this can be done with algorithms that flag irregularities, but considerable human effort is also required at various stages of production.

Research will be needed to test the performance of web-scraped data, particularly how closely online pricing tracks in-store pricing. The testing will need to be sensitive to website content that is IP-specific (e.g., price displays may be different for frequent website visitors than for first-time users). It should be fairly straightforward to periodically check to see how closely a firm's online prices trend with in-store prices. An obvious limitation regarding comparisons of the similarity of online and brick-and-mortar retailer prices is that it requires the presence of both for each firm. However, this point might be deemphasized to the extent that online prices *should* be different from in-person prices due to the costs of delivery (less the costs of making a sale). Both online and in-person purchases are relevant in constructing a CPI, but it would be impractical to devote scarce agency resources to estimating a comprehensive tabulation of all household purchases. Some retail chains have both online and in-store sales, and comparisons can be made to test the nature of any systematic price differences.

Longer-Term Data Visions

For the foreseeable future, representative surveys will continue to play an important role in federal statistics. It is important to also think about what the consumer economy and, in turn, the perfect data for tracking it, will look like in 10 or 20 years. In the not-so-distant future, most transactions in the economy will be electronic and will produce a trail of data useful for measuring prices and quantities of goods and services. In China, for example, virtually all transactions are already electronic, and this vision is quickly becoming a reality.

Beyond the more obvious transactions data sources, peer-to-peer payment platforms (like Venmo and PayPal) are creating additional opportunities for tracking consumer spending, but they come with major challenges with access and privacy issues. Tracking electronic payment data could be especially helpful to identify price trends in new or changing services, as occurred with the rise of the ride-share sector dominated by Uber and Lyft. For example, consumer expenditure survey limits respondents to categorizing these intracity transportation services under the “taxi fare” category. The transaction data would allow BLS to track such rapidly changing categories and perhaps speed up adoption of new or adjusted categories. Market research companies also construct consumer panels to collect timely data on spending. One such company, Traqlinem, conducts 150,000 interviews per quarter and releases spending data within a month of the end of the quarter, along with weights to balance the responses.

Merchant data should also continue to be investigated for use in price measurement. Online transactions collected by the software company Adobe (Lasiy, White, and Pandya, 2020) have been used to produce timely estimates of spending and quantities purchased of certain goods. Launched in 2016, the company’s Digital Price Index (DPI) initially covered a narrow range of goods and services, but now includes product categories including nonprescription medicines, consumer electronics, food, airfares, and furniture. The data behind Adobe’s DPI, sourced through Adobe Experience Cloud, represents 80 percent of all online transactions from the top 100 U.S. retailers, including aggregated, anonymous data from 15 billion website visits and 2.2 million products sold online. Goolsbee and Klenow (2018) accessed data for millions of transactions from Adobe Analytics (a service provider to many of the leading online retailers) to compare inflation rates for online sales with those estimated from traditional matched model, CPI-type indexes. They found online inflation to be lower by about 1 percentage point for the period 2014–2017. And, because the authors had access to quantity data, they were able to examine the importance of several issues raised in this report. For example, using the high-frequency data, they were able to directly test for chain drift and to assess the magnitude of the new goods problem (50 percent of online purchases were found to be of goods that did not exist in the data in the previous year).

The greatest flexibility in producing a wide range of price indexes is possible when transaction-level data for both prices and quantities are available in real time for the universe of goods with product identifiers, information on the outlet, and characteristics of the commodity (good or service). The quantity piece is the most difficult to obtain but—as has been demonstrated in the COVID-19 economy, where baskets have changed extremely rapidly—it is incredibly important information to have in a timely manner. To ready the CPI for this future data environment, modernization will need to focus on integrating multiple (public/commercial, survey/non-survey) data sources. The ability to integrate electronic transactions data—ideally, data that are linked to households making purchases—represents the ideal scenario for price measurement.

APPENDIX 2A: MULTILATERAL METHODS FOR PRICE MEASUREMENT

Scanner Data

Prices and quantities from scanner datasets provide an opportunity to construct timely price indexes using superlative index formulas such as Fisher or Törnqvist. If the datasets also have information on characteristics (attributes) of the product, then these indexes can be improved to better account for quality change with hedonic techniques to impute prices at entry to and exit from the market.

Chained versions of superlative price indexes—the recommended approach in case of high product churn—can suffer from chain drift, for example when consumers stock up goods that are on sale (see, for example, Feenstra and Shapiro, 2003; Ivancic, 2007). Chain drift occurs when the trend of the period-on-period chained version of a price index differs in a systematic fashion from that of the bilateral version of the index that compares directly the prices of two periods.[1] These differences are problematic, because ideally one would like to make comparisons that are transitive, or independent of the order in which periods are compared. While a fixed-weight index, such as the CPI-U, is transitive, it suffers from substitution bias that can be avoided with a superlative index. Multilateral index number methods, which were originally developed for spatial price comparisons, have been adapted to deal with the chain drift problem. These methods have emerged as "best practices" to exploit scanner data for price measurement.

In contrast to bilateral index methods that compare prices across two time periods, multilateral index methods make price comparisons across three or more time periods (Chessa, 2016). Specifically, multilateral index methods use all bilateral product matches across all periods, weighted by their market importance (expenditure share). Usefully, the calculation assigns expenditure weights in a way that automatically gives greater importance to price changes of products with larger sales. Multilateral price indexes are transitive, or path-independent, implying that the chained versions of the indexes are equal to the direct, bilateral indexes. Thus, they are free from chain drift by construction.[2]

GEKS

The GEKS index[3] offers a method to create transitivity in a set of bilateral indexes, for example based on the Fisher formula.[4] Suppose the whole estimation window consists of $T+1$

---

[1]De Haan (2008, p. 15) showed that when the price of a detergent product went on sale in the Netherlands at approximately one-half of the regular price, the volume sold shot up approximately 1,000-fold. van Kints, de Haan and Webster (2019) and de Haan and van der Grient (2011) explored the magnitude of volume fluctuations due to promotional sales which led Ivancic, Diewert, and Fox (2011) to propose the use of multilateral indexes with a rolling estimation window to mitigate the chain drift problem.

[2]Chapter 10 in the recently updated *Consumer Price Index Manual* (International Monetary Fund, 2020) provides a full description.

[3]The acronym GEKS is based on the surnames of the "inventors," Gini (1931), Eltetö and Köves (1964), and Szulc (1964).

[4]There are other transitive multilateral index methods available, such as the weighted Time Product Dummy and Geary-Khamis. The GEKS method, when used with Fisher or Törnqvist bilateral indexes, is more flexible than the weighted Time Product Dummy or Geary-Khamis methods, because it is based on superlative bilateral price indexes, and can deal with different degrees of product substitution.

time periods $t=0,...,T$. So, there are $T+1$ possible base periods *b* for (direct) bilateral price comparisons across the window. In the bilateral price index (which uses both the current and base period quantities for weighting) going from period *b* to period *t* is denoted by P^{bt}; note that *b* can be greater than *t*. Assuming the bilateral index satisfies the time reversal test, i.e. $P^{bt}=1/P^{tb}$,[5] the price change between period 0, the index reference period (where the index=1), and the comparison period *t* $(t=1,...,T)$ can be measured by $P^{0t(b)}=P^{bt}/P^{b0}=P^{0b}\times P^{bt}$ for each *b*. If all base periods *b* are deemed equally valid, then taking the geometric mean of $P^{0t(b)}$ across all possible *b* seems a reasonable thing to do. This leads to the GEKS price index:

$$P_{GEKS}^{0t}=\prod_{b=0}^{T}\left(P^{0t(b)}\right)^{1/(T+1)}=\prod_{b=0}^{T}\left[P^{bt}/P^{b0}\right]^{1/(T+1)}=\prod_{b=0}^{T}\left[P^{0b}\times P^{bt}\right]^{1/(T+1)}.$$

Notice that the GEKS index between period 0 and the last period *T* is based on all the possible bilateral price indexes in the intervening periods. Hence, GEKS makes use of all the matches in the dataset.

For scanner data from supermarkets, the bilateral indexes in GEKS are typically matched model (maximum overlap) superlative price indexes. Ivancic, Diewert, and Fox (2011) used matched-model Fisher price indexes as elements in GEKS. De Haan and van der Grient (2011) used matched-model Törnqvist price indexes instead. The Australian Bureau of Statistics (2016) and the statistical agencies of Norway and Belgium have implemented matched-model GEKS-Törnqvist for the treatment of scanner data from supermarkets. The geometric form of the Törnqvist facilitates decomposition analyses, such as the decomposition of changes in the GEKS(-Törnqvist) index into the contributions of the various products (Webster and Tarnow-Mordi, 2019).

GEKS and Hedonic Imputations

When product churn is high, imputed versions of price indexes are recommended to deal with the “missing prices” of unmatched new and disappearing products. There are several ways to impute the “missing prices,” which are surveyed by Diewert (2021a). One method would be to use inflation-adjusted prices (carried forward or carried backward), as suggested by Diewert, Fox, and Schreyer (2017). This method is similar to traditional imputation methods that do not use any information on product characteristics.

When quality change due to technological improvement is important, it would be preferable to apply hedonic imputations or to estimate reservation prices. If bilateral hedonic imputation price indexes are used in a GEKS context (rather than bilateral matched-model price indexes), the resulting GEKS indexes will be explicitly adjusted for quality changes. De Haan and Krsinich (2012, 2014) proposed using bilateral weighted Time Dummy Hedonic (TDH) regressions, which are estimated on the pooled data of the two periods compared (for each of the bilateral comparisons). They showed that a specific set of expenditure-share weights in the

[5]The time-reversal test is satisfied, for example, by the superlative Fisher and Törnqvist indexes.

hedonic regression produces a bilateral TDH index that equals a bilateral hedonic imputation Törnqvist price index. Using these weighted bilateral TDH indexes as inputs in GEKS thus gives rise to a hedonic imputation GEKS-Törnqvist index. Statistics New Zealand implemented this method for scanner data on consumer electronics goods purchased from market research company GfK.

The choice of method, and whether to impute the missing prices, can be numerically important. Figure 2A-1 shows the performance of three different price indexes for televisions—the chained matched-model Törnqvist index, the matched-model GEKS-Törnqvist index, and the Imputation GEKS-Törnqvist index proposed by de Haan and Krsinich (2012)—estimated from scanner data provided by a large Dutch retailer. The analysis clearly demonstrates that the hedonic imputations had a significant impact as discussed by de Haan and Daalmans (2019). In short, the chained Törnqvist index suffers from chain drift that pulls it down. The index with imputations lies well above the Törnqvist index without imputation. While this result may come as a surprise to those accustomed to hedonic-type adjustments leading to more rapid price declines, de Haan and Daalmans provide a ready explanation in terms of retailers' pricing strategies.

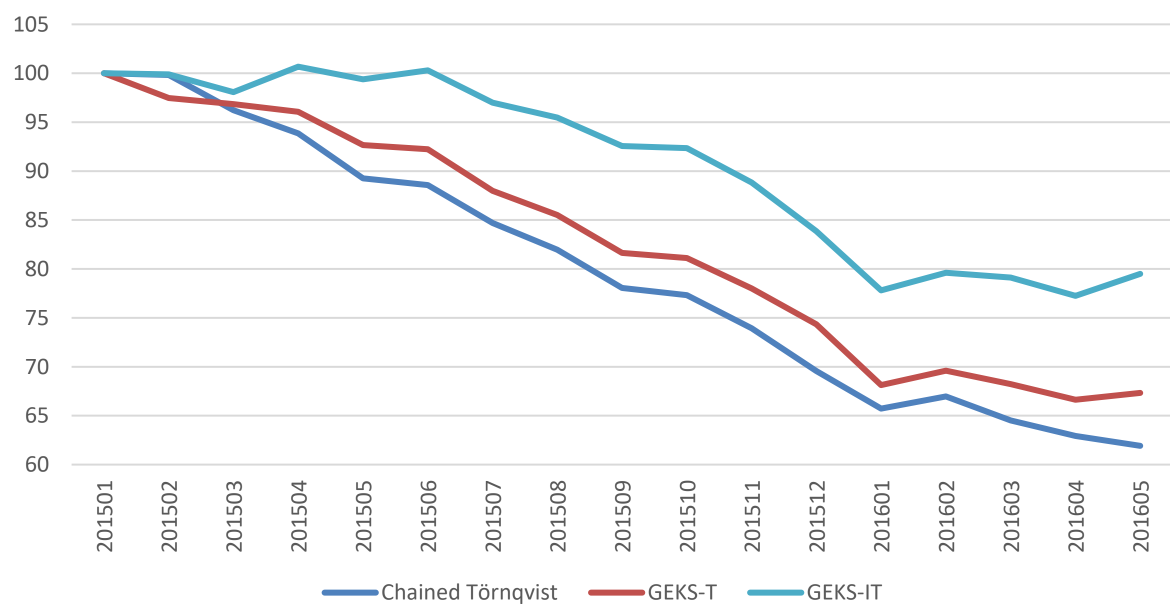

FIGURE 2A-1 Price indexes for televisions utilizing different methods.
SOURCE: de Haan and Daalmans (2019).

Revisions

When the sample period is extended and new data becomes available, previously estimated multilateral indexes will change (though often slightly), which is problematic because the headline CPI cannot be continuously revised. Various approaches have been proposed to extend a multilateral time series without revising the published price index numbers. Rolling window methods are the most popular of these.

Rolling window methods estimate multilateral price indexes on a window of fixed length that is shifted forward each month (or quarter). The results of the latest window are then spliced onto the existing time series. For example, the most recently estimated month-on-month GEKS index movement can be spliced onto the index level of the previous quarter. There are several splice methods available, with the mean splice variant perhaps the most preferred (Diewert and Fox, 2020). As its name suggests, the mean splice method takes the mean of the indexes that result from using all the possible splice periods; hence, it is independent of the choice of splice or link period. For supermarket scanner data, the Australian Bureau of Statistics implemented rolling window matched-model GEKS-Törnqvist with a mean splice (ABS, 2017).

Implementation Issues

Aggregation Level

Like bilateral methods, multilateral methods can be implemented at different levels of product aggregation. The additive Geary-Khamis and approximately weighted time product dummy (TPD) methods should only be applied at detailed aggregation levels where substitution possibilities are high; that is, for relatively homogeneous product categories consisting of products with similar attributes. Low-level Geary-Khamis or TPD indexes should preferably be aggregated up using Fisher or Törnqvist weighting to account for upper-level substitution.

The more flexible GEKS method can be applied either directly at the upper level or at lower levels (combined with Fisher or Törnqvist upper-level weighting). The choice could also depend on practical issues—for example, if there is a lack of meta data to classify products into more or less homogeneous clusters. When supplemented with hedonic imputations, it seems natural to apply GEKS at the same level as where the hedonic models are estimated.

Defining the Product

An important aspect in the construction of price indexes is the choice of product identifier. Individual goods in scanner data are typically identified by barcode. Some products with different barcodes, however, are similar from the consumers' point of view. Also, barcodes often change if unimportant characteristics change, such as type of packaging. In this case, matching at the barcode level would overstate product churn; additionally, price changes due to re-launches of comparable products with different barcodes will not be observed (Dalén, 2017). Such disguised price changes are often upward in which case missing them produces downward bias in the index. This is true for both bilateral and multilateral index number methods.

Statistical agencies sometimes receive Stock Keeping Units (SKUs) from the data providers that allow them to calculate unit value prices across SKUs rather than individual barcodes. This mitigates the above issues. In some instances, even SKUs may be too detailed so that matched-model methods, including GEKS, can yield biased results. Product descriptions in the scanner datasets could potentially be used to identify goods by cross-classifying important categorical attributes. Statistics Netherlands follows this approach when broadly defining products in scanner datasets for a number of product categories where a multilateral method (Geary-Khamis) is used, such as t-shirts and other apparel items (Chessa, 2016).

Similarity Linking

While multilateral methods have emerged as best practices for the treatment of scanner data, they have two potential drawbacks. First, multilateral methods do not satisfy the multiperiod identity test: when prices return to their initial level, multilateral price indexes, including GEKS, are not necessarily equal to 1. At least from a theoretical perspective, violation of the identity test is problematic. Second, the transitivity property is no longer satisfied when a rolling window extension approach is used: that is, rolling window GEKS (or other multilateral) indexes are not necessarily free from "link drift."

To deal with these two drawbacks, Diewert (2021c) proposed using similarity linking. The set of prices in the current (most recent period) is compared with the set of prices in each of the previous periods. Using some measure of relative price (dis)similarity, the prior period with the most similar prices is selected. Then, a bilateral superlative price index going from this period to the current period is constructed and linked to, or spliced onto, the index value in the selected period. To extend the time series, this method simply enlarges the window by adding new data (the "comprehensive window approach" mentioned earlier), not a rolling window approach.

Similarity linking can be seen as an alternative to rolling window GEKS-Fisher or GEKS-Törnqvist with better axiomatic properties. Also, because the time series is extended without a rolling window approach, link drift cannot occur. Different choices of (dis)similarity measure are possible. Diewert (2021c) advocated a predicted share method for price similarity linking. This method takes into account the matched products' expenditure shares, i.e., price comparisons with few matched products, which are likely to be unreliable, will have a small weight. The predicted share similarity linking method thus seems useful when there is a high degree of product turnover. It is also promising for the treatment of strongly seasonal goods, i.e., products that are only available in particular months of the year, such as fresh fruit and vegetables. Diewert, Finkel, and Sayag (2021) applied this method using data from Israel for fresh fruits and compared the resulting indexes with a wide variety of alternative indexes.

Diewert (2021c) showed how similarity linking can be applied when only price information is available, including web-scraped data, as an alternative to rolling window TPD. For the Israeli seasonal data, Diewert, Finkel, and Sayag (2021) compared the modified (for price data only) predicted share indexes to the multilateral TPD and GEKS-Jevons indexes. The seasonal fluctuations in the similarity linked indexes were far smaller than the fluctuations in the two alternative indexes.

So far, no national statistical agency has implemented similarity linking in the CPI, with the exception of Statistics Canada in a specific application.[6] More research is needed to examine how these methods, and in particular the preferred predicted share method, will perform on large

[6]Statistics Canada implemented the predicted share method of linking its Adjusted Consumer Price Index for the current month to a previous month; see O'Donnell and Yélou (2021). The Adjusted CPI was introduced as an analytical series in an attempt to deal with rapidly changing monthly expenditure shares (at the upper level) induced by the COVID-19 pandemic.

scanner data or web-scraped datasets. It would also be interesting to explore how hedonic imputations can be incorporated if explicit quality adjustment is deemed necessary.

APPENDIX 2B: RESEARCH ON EFFORTS TO PERFORM QUALITY ADJUSTMENT AT SCALE

Prices for goods in an elementary index can rise when prices of identical goods change (pure price change) or when goods improve (quality change) and higher prices reflect that higher quality. In a COLI context, the quality change challenge is to identify quality changes in order to construct a price index that only tracks pure price changes that affect welfare and not quality changes.

Some of the recent methodological advances in adjusting for quality change by academics have taken a demand-based approach, making explicit assumptions about the nature of the underlying utility function. In some cases, that approach requires estimated utility parameters for the index construction. For example, Feenstra (1994), and more recently Redding and Weinstein (2020), assume a Constant Elasticity of Substitution (CES) utility specification to construct the implied price indexes.[7] A recent application of the Redding–Weinstein approach found implausible results, suggesting that this approach remains a work in progress. Overall, these demand-based methods are not used by statistical agencies because a price index that is heavily dependent on the assumption that consumers choose expenditures to maximize a CES utility function would not be reliable enough for official purposes.

At the same time, another strand of the literature is based on econometrics rather than a demand model and data on expenditure shares. In this literature the hedonic coefficients are not tied to any underlying consumer preferences and do not necessarily have an intuitive interpretation. Instead, the hedonic regression is viewed as a reduced form whose coefficients reflect changes in both demand- and supply-side factors (Pakes, 2003). Under this view, the primary purpose of a hedonic regression is to predict prices, in which case choices about the specification are all about the predictive power of the regression, not the sign and magnitude of the coefficients.

This approach is related to the "imputation method" that has been around for decades. The chapter on hedonics in Berndt's (1991) econometrics textbook shows how hedonic regressions can be used to predict prices missing in the period before entry or after exit to allow the inclusion of those prices in the index. Recently, these imputation indexes and how they compare to other approaches have been studied, particularly in the context of scanner data (de Haan and Krsinich, 2012, 2014; Silver, 2010) and have been empirically implemented in other "big data" contexts (see, for example, Bajari et al., 2021; Ehrlich et al., 2021).

---

[7]As the name implies, with CES, the ratio between proportional changes in relative prices and proportional changes in relative quantities is always the same. The theoretical appeal of the Redding–Weinstein method has been debated because it violates many of the basic axioms for price indexes by allowing changes in tastes to affect the price index in the same way as price changes.) Moreover, Diewert and Feenstra (2017) argued that the infinitely high implicit "demand reservation prices" of the CES model can result in overadjustment for new and disappearing varieties.

Recent innovations to these imputation methods have centered on (1) handling unobserved characteristics, and (2) improving methods to estimate hedonic equations at scale. Erickson and Pakes (2011) developed a method that allows for accounting for changes in unobserved characteristics. Ehrlich and colleagues (2021) folded this method into their strategy for constructing price indexes with scanner data.[8]

The other direction taken in recent work has been in leveraging new artificial intelligence (AI) techniques to develop improved methods that yield more precise predictions from hedonic regressions. Bajari and colleagues (2021) conducted the seminal work in this area: In the context of superlative index formulas, they show that it is possible to apply these new approaches to obtain reasonable measures of price change at scale. That is, their methods lend themselves to automation and, in principle, do not require the human intervention of the traditional methods. Ehrlich et al. (2021) and Zeng (2021) contain recent implementations of these novel methods.[9]

A very recent paper (Ehrlich et al., 2021) attempts to implement both demand-based and reduced-form approaches to scanner data. The authors applied an approach that Redding and Weinstein applied to several classes of IT goods and found that the approach yielded implausible results. They applied a second approach that combined the machine learning estimation methods introduced by Bajari et al. (2015) with econometric methods that allow for unobservable characteristics based on Erickson and Pakes (2011). Empirical works like these provide much needed perspective on the relative merits of these new contributions.

[8]The ability of the method to measure quality improvements at the time of entry of new varieties is an open question.

[9]Another promising feature of some of these AI methods is that they do not require structured variable with which to represent characteristics. Instead, they can use unstructured text in product descriptions, AI representations of pictures of the products.

3
Higher-Level Aggregation and Shifting Consumer Behavior

3.1. MOTIVATION FOR DATA MODERNIZATION

In today's fast paced economy, statistical agencies face a major challenge to keep published data relevant by reflecting rapidly changing conditions. Timely data have increasingly become a basic expectation, and COVID-19 has reinforced the need for up-to-date statistics. In the context of measuring changes in consumers' costs of living, a price index must track the goods and services that people actually buy and account for the relative amounts spent on them. In other words, the expenditure weights used in a price index should represent current reality as much as possible. For long periods, expenditure patterns may be reasonably stable. However, the recent pandemic and recovery made the importance of updating the weights during unpredictable, turbulent times painfully apparent and showed that failure to keep the weights up to date can lead to lost credibility if households are not able to recognize the market basket. The timeliness of the Consumer Price Index (CPI) weights must be improved.

A primary method used by most statistical offices to determine people's expenditures is asking them directly about their consumption of goods and services in the form of a household survey.[1] In the United States, the Bureau of Labor Statistics (BLS) carries out this process using the Consumer Expenditure Survey (CE), which has for decades been the most comprehensive source of data on households' income and expenditures.[2] The CE is used to establish the relative importance of 243 expenditure items, or item strata (241 commodities and services plus 2 housing strata) for two higher-level indexes: the Consumer Price Index for All Urban Consumers (CPI-U) and the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).[3]

---

[1]As discussed below, Statistical Offices in many countries use national accounts data to estimate upper-level CPI weights, and data from household expenditure surveys to derive the more detailed weights (Eurostat, 2018).

[2]NASEM (2013), *Measuring What We Spend: Toward a New Consumer Expenditure Survey*, includes a detailed description of all of the uses of the CE in administering federal programs.

[3]As discussed in Section 3.3.3, the relevance of the CPI-W has diminished markedly in recent decades as the portion of the economy's workers employed in occupations covered in the index continues to decline.

Additionally, CE weights are used in an experimental price index published by BLS covering urban consumers aged 62 and older (the CPI-E).[4]

One challenge with the current approach for updating weights is that household expenditure surveys are burdensome to respondents and costly to administer, which places practical limits on how large the sample size can be and how often it can be conducted. The limited sample size means that more than one year of data must be pooled to create enough observations to estimate weights, especially for subnational (e.g., regional, population subgroup) indexes. Furthermore, processing the raw survey data and developing new item weights is time-consuming, which means that the weights used for a particular year are based on expenditure patterns from earlier periods and available only with a significant lag. For example, the CPI weights (or "market basket") for 2020 and 2021 were based on the CE expenditure patterns for 2017 and 2018. Similarly, the new CPI weights in 2022 will be based on expenditure patterns from 2019 and 2020. As a result, the CPI weights tend to be outdated in representing consumer purchases taking place in a given period; for example, it is known that consumption behavior in 2020 changed sharply due to COVID-19 and so weights derived from 2019–2020 may not reflect what is going on in the market in subsequent periods.

Additionally, infrequent updating of the market basket delays bringing new goods (which often display distinctive price dynamics) into the CPI or moving obsolete goods out of the index. While the composition of what consumers buy is constantly evolving, the shifts were especially dramatic during the pandemic—to the extent that, during the U.S. lockdowns of 2020, there was a substantial "disappearing products problem" in some expenditure categories.[5] The most noteworthy declines in spending were in the categories of travel/transportation, food away from home, and recreation services (which includes admissions to movies/theater/sports, gambling, and package tours among other things), and clothing worn outside the home (see Figure 3-1) (Cavallo, 2020; Diewert and Fox, 2020; Reinsdorf, 2020). Moving in the opposite direction, demand for food at home and information processing equipment surged. These abrupt changes in spending patterns illustrate the need to keep the CPI basket up to date. Such episodes can cause sudden obsolescence of item weights, or even the basket itself, and call into question the accuracy and relevance of price indexes.

[4]The CPI-E is discussed in greater detail in Chapter 5, on population subgroup price indexes.

[5]Diewert and Fox (2020) and Cavallo (2020) both documented the problems that arose during the pandemic, including goods and services becoming unavailable, for the measurement of price inflation.

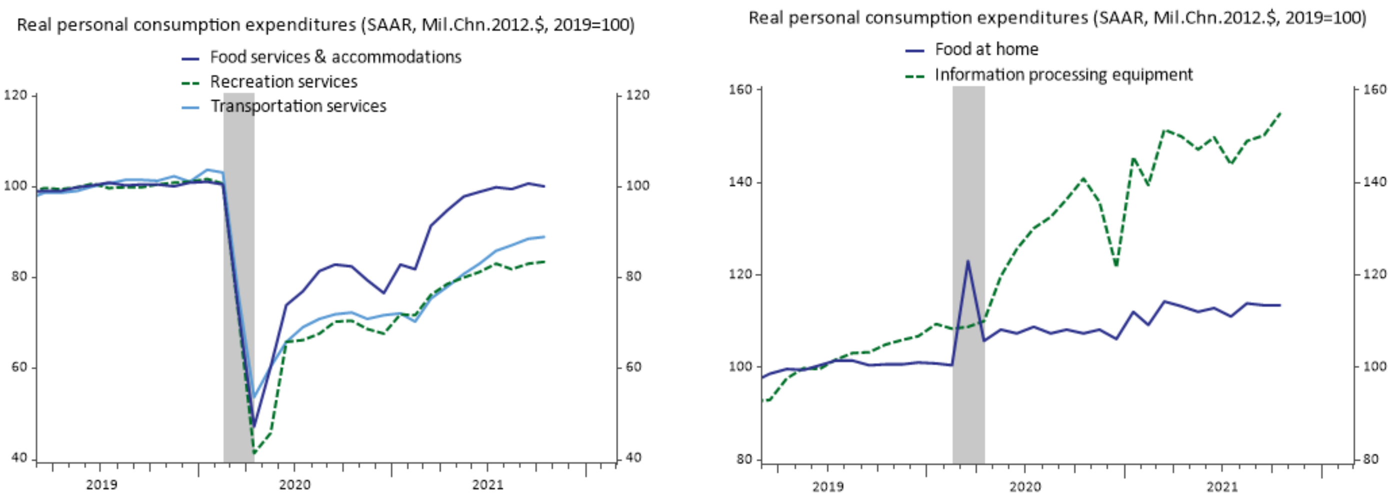

FIGURE 3-1 Consumer expenditure shifts during the COVID-19 economy.
SOURCE: Panel-generated, using BEA data.

The recommendations in this chapter focus on the need to improve the timeliness and accuracy of data on spending patterns and to identify items where survey-based estimates of consumer expenditures are weak. Despite improvements in recent decades—e.g., merging the CE with the Point of Purchase Survey (indicating where households made purchases), a modestly streamlined questionnaire, and an increased survey sample size—the CE-based method of establishing expenditure weights remains problematic. The survey generates high respondent burden and has experienced declining response rates.

The specific problems described in this chapter are: (1) the need to improve the timeliness of upper-level weights, which, on average, lag 36 months behind actual expenditures in a given period; (2) the need to improve the accuracy of weights applied to specific items for which CE does a poor job estimating[6] and for which alternative data could do better; and (3) general concerns about the sustainability of the CE given concerns about respondent burden and falling response rates. The panel is optimistic that problem (2) can be successfully addressed with alternative data. Although problems (1) and (3) are more complicated, alternative data may also allow BLS to shorten the CE survey (which could help with respondent burden and response rate issues) and create other opportunities to shorten the lag between the survey and its incorporation in the CPI.

3.2. APPROACHES TO ESTIMATING CPI WEIGHTS AND MARKET BASKET COMPOSITION

3.2.1. Current BLS Methods

[6]Underreporting of expenditures is especially a problem in the diary survey component. NASEM (2013), which focused on cognitive and motivational issues in the Diary and Interview surveys, documented survey method problems and provided guidance for turning around the deterioration.

In the United States, weights for the 243 component cells that aggregate up to the overall CPI reflect data collected about consumers' spending patterns from a two-year interval that is centered two years back at the time of their introduction. For example, BLS updated expenditure weights for the CPI-U in January 2020 using data collected in the CE during the period 2017–2018.[7] The weights used for 2020 and 2021 continued to use an average of expenditure shares from this two-year period.

A major reason for BLS's practice of pooling data across 2 years is so that sample sizes are large enough to produce accurate weights at the needed item and geographic levels of detail. To calculate the 2017–2018 weights for the urban population, BLS used approximately 21,000 weekly diaries and 44,000 quarterly interviews. However, for 2020, BLS reports that response rates for the Diary part of the CE survey dropped to as low as 28.3 percent after March and the Interview response rate dropped to around 40 percent in July 2020, roughly 15 percentage points lower than April 2019 levels (https://www.bls.gov/cpi/questions-and-answers.htm). Past commenters have argued that an increase in the CE sample size is needed, especially for purposes of estimating subgroup and other subnational indexes.[8] Combining CE records with government administrative data sources (e.g., from the Social Security Administration and the Internal Revenue Service for income) and using household consumption data from the national accounts could also reduce the amount of information requested of respondents.

CPI staff also receive data from the CE program quarterly, which enables the quarterly publication of the final Chained Consumer Price Index (C-CPI-U) indexes. The C-CPI-U, first published in 2002, is a supplemental measure that largely solves the weight timeliness issue, but with the tradeoff of a long revision cycle. In particular, the C-CPI-U better accounts for cost-of-living changes faced by consumers in a world where buying patterns respond to changes in relative prices. It offers a measure of changes in cost of living characterized by reduced substitution bias relative to the Laspeyres formula used in the headline CPI.

Crucially, weights used in the C-CPI-U can be aligned with actual expenditures as opposed to two-year average weights being introduced with a long lag in the CPI-U, and are therefore considerably more coincident with the prices of items in the index than those used in

---

[7]The relative importance of a component is its expenditure or value weight expressed as a percentage of all items within an area or an area within the U.S. For a detailed description of how base period weights are established and brought forward using the Lowe index formula, see the relevant parts of the CPI Handbook of Methods available at www.bls.gov/opub/hom/cpi/calculation.htm. See, also, Chapter 5 in Carroll, Crossley, and Sabelhaus (2015), and background from BLS at https://www.bls.gov/covid19/effects-of-covid-19-pandemic-on-consumer-price-index.htm.

[8]Triplett (1997, p. 15), for example, wrote:

> . . . The [CE] sample size (5,000 consumer units) is certainty too small for almost any use for which one wants consumption data. . . . The recently announced increase from 5,000 to 7,500 [CE] consumer units is a positive, but grossly insufficient, step. . . . The [CE] is the federal government's only general purpose survey of consumer expenditure. . . . For comparison, the Canadian consumer expenditure survey will soon have a sample size of 36,000. . . The [CE]'s small sample size and lack of a benchmarking statistic means that its estimates for smaller components (e.g., household textiles) particularly are not as reliable as one would want for serious research on consumption.

NRC (2002, p. 260), while concluding that research at the time was too incomplete to propose solid recommendations about CE sample size, offered analysis of the impact that "changing the survey sample size would have on the accuracy of expenditure weights and, in turn, the relationship between weight accuracy and index variance."

the CPI-U. A preliminary version of the C-CPI-U uses a constant expenditure shares formula.[9] The *final* chained CPI uses a Törnqvist formula, which accounts for consumer substitution and is the superior index for most applications where index revisions are acceptable. This Törnqvist formula is applied and the index is revised after monthly weights become available. For example, the March 2021 weights were incorporated in the final March 2021 C-CPI-U (released in February 2022).

Differences between the CPI-U and the C-CPI-U reveal how the age of the weights can affect index estimates. Because it captures effects associated with consumer substitutions, the C-CPI-U has typically increased at a slower rate than the CPI-U (or CPI-W).[10] Differences in the two series started out fairly large during the initial period after the C-CPI-U was introduced and then stabilized at more moderate levels. In 2001 (the first year of C-CPI-U calculation), the gap in the 12-month percentage change between the C-CPI-U and the CPI-U was about 0.5 while, in subsequent periods, it was more typically in the 0.15–0.35 range.[11] In contrast to the C-CPI-U, which approximates a superlative index after the final revision, a fundamental issue is that the CPI-U is not revisable and so the lagged weights from the CE cannot be updated to reflect current or recent spending patterns.[12]

In researching options for improving timeliness in the updating of weights to its flagship CPI-U, BLS has used historical CE data to illustrate how the average weights would have looked under different scenarios for estimating expenditure category weights. The alternatives involve either compressing the period from which CE data are drawn for a weight update, reducing the lag between data collection and integration into the CPI, or a combination of the two. Implementing weights based on either a two-year rolling average or on annual estimates from expenditure data are both being considered with the latter appearing to be the frontrunner (Klick, 2021). Figure 3-2 depicts BLS simulation of the annualized percent changes over the period from December 2001 to December 2020 if annual CE data had been used in the CPI. It is important to note that a one-year lag for introducing new expenditure weight estimates into the CPI is not currently feasible due to data collection and processing timelines. Lagging by two years is feasible, and BLS research finds a three-tenth of a percentage point "improvement" over the current methods.

---

[9]This weighted geometric mean index is used to combine individual prices into elementary aggregates for narrow product groups (i.e., the individual constituents of the basket). Higher-level aggregation is done with a Laspeyres formula (which uses a fixed basket from the reference period for expenditure patterns). For a full description of BLS's Chained CPI, see https://www.bls.gov/cpi/additional-resources/chained-cpi.htm. In January 2015, BLS switched to a CES formula for the preliminary version of the C-CPI-U.

[10]BLS research (Kurtzon, 2017) found that the weights, rather than differences in the formula used, account for most of the difference of the Chained CPI from the CPI-U.

[11]Klick (2021) detailed the differences between the 12-month percent changes of the initial C-CPI-U minus the final C-CPI-U from January 2001 to September 2020.

[12]One reason why the index is not revised is its widespread use in escalation of payments.

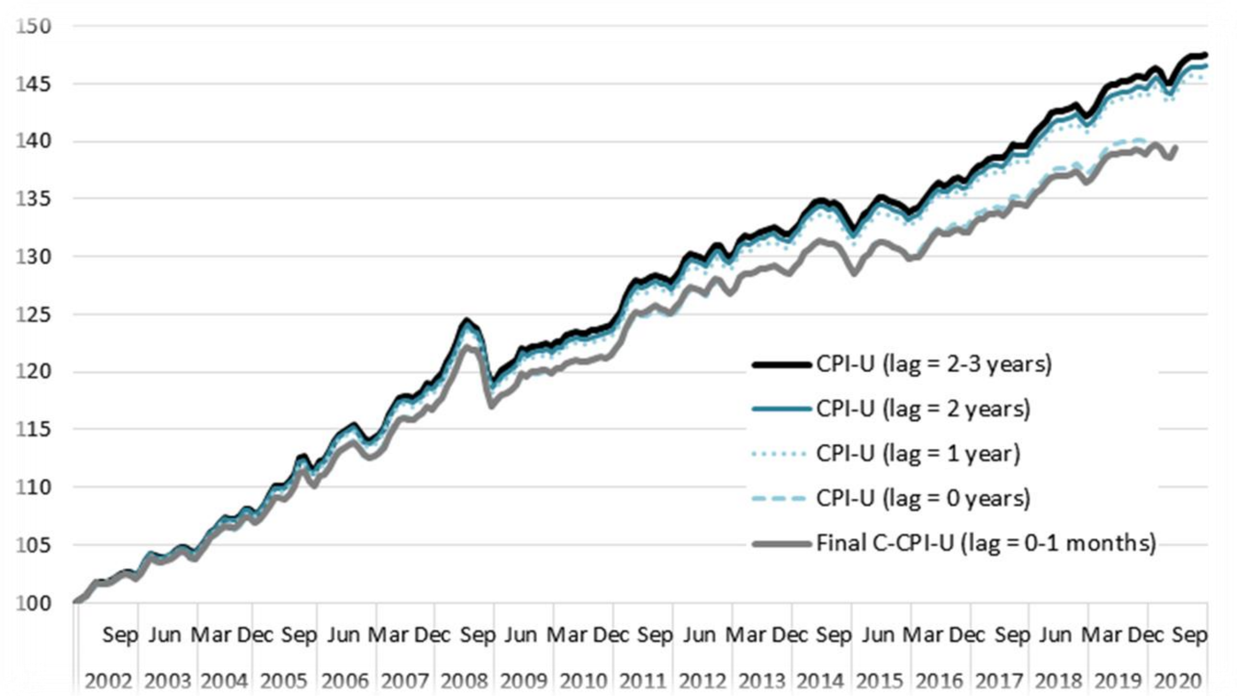

FIGURE 3-2 The impact of reducing the lag between expenditure and index reference periods.
SOURCE: Klick (2021).

The main takeaway from BLS's research is that moving from the current methodology to a 2-year lag in setting annual weights would have reduced the measure of substitution bias (the difference in the 12-month percent change of CPI-U minus C-CPI-U indexes) by 0.036 percentage points per year. Moving from the currently used methodology to a 1-year lag would have reduced substitution bias by 0.061 percentage points (for 2017–2018 estimates).[13] To take a concrete example, calculating the index for April–May 2020 would involve lags of:

- 2–3 years expenditure weights for 2017–2018
- 2 years 2018
- 1 year 2019
- 0 years 2020
- 0–1 month, March-April 2020

Of course, the 2020 weights would not be possible to compile in real time, and even 2019 weights would be challenging from a production standpoint.

It is also worth noting that BLS experimentation with using one year of expenditure data to estimate weights has been focused on the national urban population. It is not clear whether one year of data would be sufficient for lower-level geographic or subgroup indexes (e.g., CPI-E or the income group indexes discussed in Chapter 5); however, it would certainly be an option to estimate the national total using one year of CE data and then use two years of data to allocate of distribute the national total to subgroups.

[13]This sentence was revised after institutional report review to improve its accuracy.

3.2.2. Methods Developed by Other National Statistics Offices and Academic Researchers

Many national statistical offices (NSOs) update their upper-level CPI expenditure weights every year, with a one-year reference period for use in a Laspeyres/Lowe index formula. Due to advantages in terms of timeliness and accuracy relative to household surveys, the United Kingdom Office of National Statistics (ONS), for example, uses information on spending patterns culled primarily from the household final consumption expenditure component of the UK National Accounts, which largely is based on administrative and business data representing the whole economy. These data are used for the annual updates "because the expenditure information is comprehensive and balanced against data collected in other sectors of the economy to create the most accurate picture of consumer spending."[14] In general, Eurostat areas (and ONS) use the national accounts approach for the higher-level weights, and consumer expenditure surveys for the lower-level weights within each higher-level aggregate. Many other NSOs use a similar methodology.

In a U.S. context, the analogous category to household final consumption is personal consumption expenditures (PCE) estimated for the national income and product accounts (NIPAs) produced by the Bureau of Economic Analysis. The implied shares can differ significantly from the lagged shares estimated from CE data.[15] The example cited in Chapter 1 of weights for airfares during the pandemic illustrates the impact of these differences vividly; the weight on air transportation in the CPI only declined from around 0.8 to 0.6 from February 2020 to April 2020, whereas, in the PCE, the share dropped from 0.9 to below 0.1, much more accurately reflecting changed consumption patterns during the lockdowns. CE and PCE discrepancies do not just emerge during extraordinary times; in 1995, the ratio of CE to PCE expenditure shares on alcoholic beverages was a dismal 0.34 (NRC, 2002).[16] Furthermore, there is considerable evidence that the discrepancies between the CE and PCE have been getting larger over time. Passero, Garner, and McCully (2015) found that among items that are totally comparable in definition and coverage between the CE and PCE, the ratio of CE expenditures to PCE expenditures declined from 84 percent in 1992 to 74 percent in 2010. This decline could indicate worsening underreporting of expenditures by CE respondents.

Eurostat guidelines also recommend putting in place procedures to promptly bring important new goods into the index. Extending this concept to the current situation, Eurostat (2020) has advised national statistical agencies to adjust their methodology for producing weights for Harmonized Indices of Consumer Prices so that they better reflect the impact of

[14]https://www.ons.gov.uk/economy/inflationandpriceindices/articles/consumerpriceinflationupdatingweights/2017.

[15]For an assessment of the differences between expenditure data in the National Accounts and Consumer Expenditure Survey data, see Johnson (2017). Earlier, Blair (2015) constructed a PCE-Weighted Consumer Price Index.

[16]As noted below, this observation should be tempered by the fact that the PCE basket has broader coverage of the economy (e.g., it includes business and government spending) than the CPI market basket, so the weight of any single item is usually lower in PCE.

COVID-19. They are recommending a variety of source data, including national accounts, to estimate expenditure shares.[17]

Although somewhat different than their consumer survey counterparts, PCE data underlying the national accounts also come with challenges in terms of their applicability to estimating CPI weights. PCE and CPI item categories are not all comparably defined. One major category in particular—medical care—raises special challenges. For medical care, the PCE covers a wider scope of goods and services than does the CE/CPI. PCE coverage includes government-funded care provided through social insurance—such as Medicare and Medicaid—and the employer-paid portion of medical insurance; in contrast the CE/CPI excludes these categories. BLS itself has published results of the CE/PCE Concordance that compares comparability in detail.[18]

Another challenge in using PCE weights for the CPI is that the procedure for obtaining PCE weights necessitates allocating overall sales data for a particular commodity to business, government, or consumers' spending. Thus, the PCE is an indirect measure, calculated jointly with purchases made by non-consumer sectors. Triplett (1997, p. 16) noted that it is especially difficult to calculate consumption shares at more refined item levels because sales to consumers are not always distinguishable from sales to businesses and government: "The finer the level of detail, the more likely that the long chain of computations necessary to reach the PCE's indirect estimate of consumer spending will have cumulative errors that affect the totals."[19] Even so, it seems implausible that estimates of business and government purchases of consumers goods could be off by enough to account, on their own, for the large differences between NIPA and CE weights for some expenditure categories.[20]

As alluded to above, a major advantage of the CE weights is that they are derived directly from a household survey, which allows household characteristics to be linked to expenditure information. In turn, subpopulation indexes such as the CPI-E and CPI-W can be calculated in a way that reflects expenditure patterns (although not necessarily prices paid) by those groups. A national accounts-based approach to estimating weights is therefore an option for statistical offices for whom producing subpopulation CPIs is less of a priority. The need for information on household/consumer unit characteristics presents a significant stumbling block to moving away from the CE completely, although, precisely because of this issue, some countries (e.g., Netherlands) have discontinued subgroup indexes.[21] This decision has allowed these statistical

[17]See the report appendix, https://ec.europa.eu/eurostat/documents/10186/10693286/Guidance-on-the-compilation-of-HICP-weights-in-case-of-large-changes-in-consumer-expenditures.pdf.

[18]https://www.bls.gov/cex/cepceconcordance.htm.

[19]A main part of the problem is that the sales data used for PCE do not always include product detail. For example, gas station sales combine snacks and drinks, auto parts/fluids, and fuel. BEA has been addressing this problem by supplementing sales data with scanner data providing product detail.

[20]National accounting data provide fairly accurate information on total government spending and business expenses, which limits the scope for large errors to arise in total intermediate spending, though there may be offsetting errors in some detailed spending categories.

[21]And a Netherlands–U.S. comparison of the need for subgroup data is not apples to apples; regional information within Europe is still needed as each country produces its own price data in addition to the 19 EU countries' harmonized index. For the United States, measures of household inequality are likely to feature prominently in the coming years (see Chapter 5, which makes the case for population subgroup price indexes). Household-level information is essential for producing nominal measures of income and consumption inequality.

offices to more easily turn to other sources to estimate upper-level expenditure weights. However, we note again that, even if BLS reduces its reliance on the CE for some CPI item weights, it is still an important source of information on regional and demographic household expenditures.

NSOs are also advancing methods based on alternative, transaction data sources to improve the timeliness of weights. Statistics Canada, for example, published a special edition price index based on an alternative set of weights reflecting spending patterns during the pandemic. As part of the analytic exercise, the agency was able to use credit and debit card data supplied by the Bank of Canada to "to account for pandemic-related expenditure shifts at more detailed levels of geography and CPI components" (Statistics Canada, 2021, p. 4). Even prior to COVID-19, Statistics Canada concluded that expenditure estimates from surveying consumers (using the agency's Survey of Household Spending) were sufficiently inaccurate that a switch to using household final consumption expenditures (HFCE) data from the national accounts was warranted. This approach also helped with timeliness—for example, using the HFCE data, the weights used to estimate the June 2021 CPI were able to be based on expenditure patterns for 2020, a much more recent period that was possible with the Survey of Household Spending. This special analytical index shows the benefits of collaboration and data sharing between government agencies and other organizations—in this case, private data as one of the inputs in estimating the CPI weights, and the use of a timely complementary indicator to provide data users a more complete picture.[22]

Other countries are pursuing similar strategies to those employed by Canada. As noted in Chapter 2, the Australian Bureau of Statistics (ABS) has been turning to a range of more timely data sources such as retail trade and scanner data. For their program's 2020 annual updating, the agency used these alternative data sources for approximately 20 percent of the weight of the CPI.[23] More broadly, evidence has been compiled for many more countries about the changes in consumption patterns in response to COVID-19—see, for example, Seiler (2020) for Switzerland (based on transactions data), Andersen et al. (2020) for Denmark, Bounie et al. (2020) for France, Carvalho et al. (2020) for Spain, and Chronopoulos et al. (2020) for the United Kingdom. BLS is currently exploring these kinds of options for publishing "nearly superlative" indexes based on more timely data used for estimating higher-level aggregation (Bergqvist et al., 2021).

Credit card and other payments records are also being accessed by academic researchers to compile price indexes that reflect current spending patterns. Much of this work, including the references just cited, has been driven by the need to quickly estimate higher-level aggregates during the pandemic. Cavallo (2020), for example, took as his starting point the latest available weights for the CPI and then updated them using credit and debit card transactions data collected

---

[22]See "Consumer expenditures during COVID-19: An exploratory analysis of the effects of changing consumption patterns on consumer price indexes." (https://www150.statcan.gc.ca/n1/pub/62f0014m/62f0014m2020010-eng.htm; and "Adjusting the Consumer Price Index to the new spending realities during the pandemic," The Daily (Oct. 8). https://www150.statcan.gc.ca/n1/en/dailyquotidien/201008/dq201008a-eng.pdf?st=FJEZmhzG

[23]https://www.rba.gov.au/publications/smp/2021/feb/box-a-consumption-patterns-and-consumer-price-index-weights.html.

by the Opportunity Insights Tracker.[24] Cavallo's COVID-19 period reweighting is only applied at the highest level of aggregation (although he does split the "food and beverages" category into three subcategories—food at home, alcoholic beverages, and food away from home[25]—disaggregations that the CPI publishes as a matter of course). Comparing the CPI and COVID-19–adjusted CPI weights yields changes in the expected directions: for example, food at home revised up from 7.58 to 11.28, transportation revised down from 15.74 to 6.25, and food away from home revised down from 6.19 to 3.13.

In the context of market disruptions during COVID-19, Reinsdorf (2020) recommended data blending procedures for updating CPI weights to ensure that they are less distorted by rapid changes in expenditure patterns exhibited by consumers brought on by extraordinary circumstances.[26] As illustrated in the Statistics Canada example above, one such approach would use credit card and other payments data to produce a complementary index of short-term price change with weights reflecting spending patterns during the pandemic. More generally, such timely data on spending would also be valuable for quickly detecting the appearance of important new goods and to determine weights for their prompt incorporation into the basket.

3.3. OPPORTUNITIES, CHALLENGES, AND RECOMMENDATIONS

Leveraging a wider variety of data sources containing information on consumer purchases and changes in how these data are used have the potential to improve the timeliness and accuracy of upper-level weight estimates used to calculate the CPI. The inability of the current CE-based data infrastructure to detect and incorporate shifting purchasing patterns during the COVID-19 pandemic demonstrates well why timely weights are critical.

> **Recommendation 3.1:** To improve the timeliness of the CPI and the accuracy with which it captures changing buying patterns, BLS must (1) update upper-level weights—which currently, on average, lag 36 months behind actual expenditures in a given period—more frequently and rapidly, and (2) improve the accuracy of weights applied to specific items that the Consumer Expenditure Survey measures poorly and for which alternative data are likely more accurate.

Immediate steps that can be taken to improve the accuracy of the weights used in the CPI are to incorporate alternative data to offset the clearest weaknesses of the CE described above and to use two-year rolling averages of CE weights. Indeed, many statistical offices are already aggressively integrating nonsurvey sources into their item weighting and have implemented procedures that incorporate survey data more quickly. The recommendations below are intended to help guide research that is crucial for BLS's capacity to maintain agility and flexibility in its

---

[24]https://opportunityisights.org/.

[25]Cavallo applied the same adjustment for food at home and alcoholic beverages, as he did not have separate spending data on alcohol.

[26]Consistent with normal procedures, weights for the CPI will be refreshed in January 2022 based on CE data from 2019–2020. According to an official notice, BLS considered interventions but decided to maintain normal procedures. https://www.bls.gov/cpi/notices/2021/2022-weight-update.htm.

weight-updating procedures both in ordinary times and to respond quickly to major breaks in consumption patterns such as those experienced during the pandemic.

3.3.1. Revising CE Weights More Frequently

A nationally representative survey conducted by a government statistical agency continues to be needed for benchmarking estimates of consumer expenditures in a way that links buying patterns to households. Thus, for the foreseeable future, a version of the CE—ideally, one optimized for its role in estimating CPI expenditure weights—will be needed. However, for the expenditure estimates to be relevant, particularly during times of rapid change, the lag with which the data become available must be reduced.

> **Recommendation 3.2:** Ideally, the expenditure data used to calculate CPI weights would come from a single 12-month period ending no more than six months prior to their introduction. For example, new CPI weights introduced in January 2022 would reflect expenditure patterns from July 2020 to June 2021. This production schedule may take time to achieve so, as an interim step in mitigating the timeliness problem, weights should be updated annually using two-year rolling averages of the CE data. This approach should become part of the official measure as soon as possible.

Updating the weights every year can be done using the CE data that BLS already collects.[27] Under this setup, the rolling weights would still lag real-time market realities, but not by as much as they do in the current two-year cycle—it would provide an interim immediate improvement even if it does not reach the ideal. In addition to improving the timeliness of the updates, the compressed schedule would also smooth changes in the weights from one period to the next.

The objective in annual updating should be to shrink the lag between the endpoint of the expenditure data collection and weight updating.[28] It should be possible have a cut-off for the expenditure data at the end of Q2 for implementation in January. As noted, annual updating of the weights based on a two-year rolling average of the CE data is meant as an interim solution until BLS can develop methods that allow annual updating based on a single recent year of expenditure data.

Ultimately—in 10 to 20 years—even more detailed price and quantity information will be available for use in the CPI from a range of nonsurvey sources that could allow much more frequent updating of item weights. In the intermediate run, BLS could develop a hybrid approach to estimating consumers' expenditure patterns that includes national accounts and credit card transactions data that could allow the reference period for weight updates to be compressed to

---

[27]In the near term, one reason for continuing to use expenditure estimates based on 2 years of data is to maintain sample sizes needed to support MSA level price indexes.

[28]Currently, 36 months is the time from the middle of the expenditure data collection period to the middle of the index period. For example, the weights for 2017–2018 were used for the CPI from 2020 to 2021. The midpoint of the index period (January 1, 2021) is 36 months after the midpoint of the expenditure period (January 1, 2018). Under the recommended procedure, expenditures for 1 year from July 2018 to June 2019 would be used for the 2020 index, meaning that the endpoint of the expenditure period (June 2019) to the beginning of the index period (January 2020) is 7 months. From midpoint to midpoint, it would run January 1, 2019, to July 1, 2020, or 18 months, cutting the lag in half. If expenditures for 2 years were still used, they would run from July 2017 to June 2019, and the average lag would be 24 months.

just one year and the lag between the period to which they refer and the update to be shortened to a quarter or perhaps even just a month. Indeed, as documented above, some countries' statistical offices—especially those in which high levels of cooperation exist between the business sector and statistical offices—are already moving quickly in this direction of blending different types of data.[29] Typically, these countries use integrated expenditure and price data to construct the elementary aggregates which are in turn combined using an annually chained Laspeyres index with weights from the preceding calendar year.

3.3.2. Broadening Sources of Data Used for Estimating and Updating Expenditure Weights

Comprehensive payment data on individuals' spending from credit card issuers or from electronic payment processors (e.g., PayPal and Stripe) have the potential to substantially improve the timeliness and accuracy of expenditure weights estimated for the CPI; such data could also provide a source of direct information on higher-level purchasing trends that could be useful in assessing substitution and outlet bias. High frequency and timely data on spending can also be used to quickly detect emerging, important (high expenditure) new goods and to determine weights for prompt incorporation of these new goods in the basket.

Transactions Data

As discussed in Chapter 2, the advantages and limitations of transactions data dictate how they can most effectively be used. For example, while timely, credit card data identify only relatively broad expenditure categories, although retailers often maintain and link information on detailed purchases to specific individuals/households. In contrast, transactions data from retailers, such as scanner data or the data used for the Adobe Digital Economy Index,[30] do include product detail. While credit card data would be more useful if they included specific product codes, such detail is not always needed for estimation of upper-level weights.

Research is ongoing at statistical agencies regarding how best to address the challenges of using alternative data. Much of the challenge is in data blending—including, for the CPI, how best to coordinate and map geographic breakdowns and product detail in a way that brings internal consistency across data sources or indexes based on different sources—so that alternative data on consumer purchases can be used most effectively.

> **Recommendation 3.3:** With supplementing and complementing the CE data in mind, BLS should invest in collecting comprehensive data for individual spending using electronic means of payments such as credit/debit cards or other electronic payment processors (e.g., PayPal or Stripe). Initially, these new data could be applied to the chained CPI-U or to a new experimental index. Later, after an adequate period of study, expenditure pattern estimates used to construct CPI weights should be derived as a blend

[29]Eurostat, for example, produces a Lowe index at the higher levels of aggregation (because they can explain it to the public and they are used to it). But, at the lowest level of aggregation where scanner data are available, the agency appears willing to follow the ABS lead and experiment with multilateral methods. For more information, see Appendix 2A at the end of Chapter 2.

[30]The software company Adobe Insights produces online transactions data that have been used to estimate spending on and quantities purchased of certain goods used (see Lasiy, White, and Pandya, 2020).

of data on spending from (1) the CE, (2) timely private sources, and (3) the national accounts.

For this experimental research, it is likely that credit card data will only allow controlling totals (weights) for broad expenditure categories, while CE data can be used to estimate distributions for smaller subcategories. Given constraints about product detail availability and ability to link expenditure, price, and quantity data to households, population subgroups, and geographic areas, research into use of alternative data sources for estimating upper-level weights should be directed toward production of the national level CPI.

Even within the CE program, it may be possible to integrate transaction data. One option would be to redesign the CE such that household respondents have the option of completing the survey conventionally or of allowing access to electronic records (credit card, home-scanned data, etc.) in lieu of completing the eligible categories manually. Such streamlining could have significant impacts on accuracy and respondent burden,

In developing an experimental set of expenditure weights that update CE information with additional data culled from credit card and other transaction records, BLS should explore opportunities to collaborate with other statistical agencies, as Statistics Canada did with the Bank of Canada to produce its special project COVID-19 index. At the same time, as recommended in Chapter 2, BLS should explore options to work with other agencies. For instance, the Federal Reserve, because of its priority on using timely data, has considerable expertise in working with commercial data. Likewise, the Bureau of Economic Analysis has also done work with high frequency “real-time data,” with an emphasis on timely information for early estimates that will later be replaced by other (Census Bureau) data.

National Accounts Data

Because household surveys or other, alternative sources of information on household expenditures become available with a lag, agencies generally use the fixed basket approach that weights current price changes with expenditure share data from an earlier period (the Lowe formula) for their featured consumer price indexes. However, national accounts information can be used to produce a more accurate CPI. An advantage of PCE data, described above, is that the data are benchmarked to a census of retail establishments (conducted every five years) and a variety of other merchant-based sources, so they reflect a more comprehensive accounting of transactions on which to estimate how much was bought and how much was spent.

A disadvantage of the PCE data is that they are subject to revision. Assuming that updated CPI weights are introduced 13 months after the end of the year on which they are based (e.g., the weights introduced in January 2024 would be based on spending data from calendar year 2022), the first annual revision of the NIPAs could be reflected in the PCE data used to estimate the CPI weights. Subsequent revisions of the NIPAs are likely to lead to relatively small changes in the numbers compared to the size of the sampling and nonsampling measurement errors of the CEX.[31]

---

[31]Fixler, de Francisco, and Kanal (2021) describe the magnitude and direction of revisions to PCE and GDP.

The idea of blending national accounts data into CPI weighting has been around for a long time (Triplett, 1997). The National Research Council report conducted through the Committee on National Statistics, *At What Price?* (NRC, 2002), recommended assessing the net advantages of using the BEA's PCE to produce the upper-level weights for the national CPI. The idea would be to set up an experimental CPI that uses PCE-generated weights at the upper (243 item) level but that is otherwise no different from the CPI. PCE data also could be used to improve the official CPI:

> **Recommendation 3.4:** One option for blending PCE and CE data that BLS should test for the purpose of updating upper-level expenditure weights in the CPI is to continue using the CE as the benchmark for most categories, but then integrate PCE data to adjust the acknowledged weakest categories of the CE.

BLS will need to carefully evaluate where to substitute data, but examples of problematic categories for the CE are "clothing" and "alcoholic beverages purchased for off-premises consumption." The ratio of weights for the CE (interview) to PCE is 0.317 for the former and 0.22 for the latter, indicating very large underreporting problems. In contrast, in 2010, the ratio of CE (interview) to PCE exceeded 0.94 for imputed rent, rent and utilities, and new motor vehicles and 0.80 for food at home and communication. In general, larger items estimates (from interview survey) track more consistently between the CE and the PCE (NASEM, 2013). For cases where these discrepancies in weights change rapidly, as in the air travel example cited above, a CPI based on higher frequency information can be invaluable for detecting moments when there has been a disruption.

The experiences of other NSOs that already use national accounts data in their CPIs can be drawn upon to expedite this line of research. An example of the strategy to target problematic expenditure categories is Statistics Canada's use of national accounts data for information on alcohol purchases. ABS also provides a model for between-CE updates: Its Household Expenditure Survey provides the benchmark information on the expenditure weights of Australian households, although, as of 2018, ABS moved to annually re-weighting using Household Final Consumption Expenditure data from the national accounts.

Exploring the use of PCE spending data used by BEA for the national accounts to buttress specific CE categories is a low-hanging fruit, and indeed BLS has had a research program to study level differences in the two sources.[32] BEA and BLS should collaborate on this project, which could also be extended to evaluate whether the PCE data provide a good predictor of the CE data. While improving the timeliness and accuracy of the weights of the national CPI-U should be treated as the most important objective, the extent to which achieving improved timeliness and using transactions or national accounts data entails a sacrifice of accuracy in the detailed expenditure patterns for local areas and demographic groups, such as the demographic groups covered by the CPI-W and CPI-E, might also be explored.

Survey Options

[32]See, for example, BLS (2018); Blair (2015); Carroll, Crossley, and Sabelhaus (2015); Johnson (2017); Passero, Garner, and McCully (2015).

Capturing changing consumption behavior during dramatic circumstances such as pandemic lockdowns could conceivably be done within a survey framework; however, it would require something close to a continuous survey (and one that is processed in near real time). Diewert and Fox (2020) concluded that "for many purposes, it would be useful for statistical agencies to establish a continuous consumer expenditure survey." A continuously (or even monthly) updated set of weights would require a radical shift in methodology and data sources used, and any index built on such an approach would have to be assessed on an experimental basis for a significant period of time.[33] Given that most statistical agencies do not have the resources to estimate representative baskets in anything close to real time, a hybrid data approach—one in which a consumer expenditure survey is still central, but new data sources are integrated for intra-benchmark updates—offers a particularly promising strategy. The hybrid approach entails estimating recent expenditure patterns using the best data source for each level of aggregation and item considering the goals of timeliness and accuracy and the weaknesses of the CE. As described above, national accounts and payments (credit and debit card) data are promising sources for at least some of the high-level weights, perhaps with some adjustments based on CE survey data.

For the foreseeable future, CE-type data will still be relied upon for details about consumer units and for completeness and representativeness, especially for item categories for which its accuracy is believed to be good.[34] Even so, the CE could be retooled to work in conjunction with alternative data sources (e.g., transaction/credit card data) so that weights are updated between CE benchmarks as has been done in some recent academic research.[35]

In the discussion above, it is recommended that BLS move more aggressively toward a blended data approach to establishing upper-level expenditure weights for the CPI. The presumption is that, in addition to national accounts data, BLS would purchase consumer transactions data (including prices and quantities where possible) from private companies on a continuous basis for a sample of households and items. For purposes of estimating expenditure weights, scanner data seem most appropriate as price and quantity information is often available, whereas web-scraped data generally do not offer the latter.

At this point, most official price measurement programs are using scanner data from retailers, mainly covering groceries. But consumer panel data are also produced and sold (usually, at fairly reasonable prices[36]) by companies such as Nielsen that would allow BLS to begin testing their use in the CPI immediately.[37] The advantage of consumer panel data is that

---

[33]The U.S. CE is in the field continuously, but the data are only combined for weight updating sporadically. Monthly weight updates are only worth considering for the Chained CPI.

[34]For an assessment of CE data quality, see Chapter 5 in National Research Council (2013) and Parker, Souleles, and Carroll (2015).

[35]Cavallo (2020) used publicly available data from credit and debit card transactions to update official CPI weights and build an alternative "Covid Basket." Chetty et al. (2020) built a database to measure economic trends more broadly "at a high-frequency, granular level using anonymized data from private companies" to track impacts of COVID-19 on the economy more quickly than is possible using conventional surveys.

[36]Nielsen data are made available to researchers through a collaborative arrangement with the Chicago Booth Kilts Marketing Center. Data subscription prices, for individuals and institutions, can be found here: https://www.chicagobooth.edu/research/kilts/datasets/nielseniq-nielsen/pricing.

[37]As discussed in Chapter 2, and in the context of creating subpopulation price indexes in Chapter 5, the Nielsen Consumer (Homescan) Panel is the most prominent example of this kind of data. The Homescan panel

purchases are linked to the households making them, which allows indexes for subgroups of the population to be calculated more naturally. Currently, consumer panels do not typically collect inventory and consumption information, so there are challenges, for example with new and disappearing goods, to be overcome in using these data. However, assuming the home-scanned data are used primarily for construction of weights that represent spending for the entire year, the lumpiness of purchases and the lags in buying new goods caused by consumption out of inventories will be minimally consequential.

An alternative to the data purchase model to be considered for the longer run would be for BLS to set up an in-house operation for collecting the needed data. Since scanner data from aggregators are not perfectly suited for use in the CPI (for reasons discussed in Chapter 2), this in-house operation would ideally be set up to collect scanner data directly from households.

> **Recommendation 3.5:** BLS should begin exploring development of a household-based scanner recording program that would capture prices, quantities, and item characteristics of purchases made by surveyed respondents. In addition to its value for estimating item strata weights, this method of obtaining spending information would be useful for construction of elementary aggregates.

As a first step in such a project, a retrospective study demonstrating the benefits of a home scanner data approach (using historical data from private companies IRI or Nielsen) could provide insights about the best way forward. If BLS could learn to conduct Homescan-type data collection, it would carry with it the advantage of retaining control over the source, without worrying about prices going up or the source drying up, and the program could be designed in a way to ensure that the data are representative.[38] IRI and Nielsen have already demonstrated that it is feasible to collect and use data in a timely way, and it could be viewed as an alternative survey method similar to how the computer-assisted personal interview (CAPI) method replaced home visits or phone calls years ago. However, thinking longer term, BLS should consider "leapfrogging" traditional methodologies of handheld scanners that require large initial investments and look to modern approaches using a custom mobile phone app. Technologies are changing fast, so the most durable solutions may be based on flexible, mostly software-based approaches. As recommended in Chapter 7, this project would benefit from coordination with the Census Bureau, which would benefit from access to such price and quantity data from retail sales, and BEA, which would also benefit if CPI weights incorporated national account information.

Sorting through these data infrastructure options, particularly the boldest ones, requires statistical agencies to envision the future data environment and what CPI users will need in 10 or 15 years. BLS requires timeliness for the flagship CPI but recognizes the value of population subindexes. In the current data setup, these two objectives have somewhat conflicting data requirements—it would be easy to use alternative data sources to produce a more timely CPI-U if the CPI-E, regional indexes, subgroup indexes, and other measures did not have to be

tracks the expenditures of about 55,000 households who scan the bar codes of their purchased items. Prices are then downloaded from the store where the item was purchased.

[38]For such data collection, cooperation may still be needed from the retailers, which could be an impediment to BLS developing its own home-scan data.

considered. One way for BLS to make a large leap forward is to create its own, ambitious in-house data collection program that combines household scanner data as part of the CE, information from credit cards and other electronic transactions, and web-scraping. The other way BLS could handle the conflict would be to release the headline index in a timely manner but release subgroup indexes with a lag. Because of the need to not revise the headline, the subgroup indexes could be scaled to aggregate up to the headline. BLS would no longer build up the headline from the subindexes. Instead, the agency would estimate and release the headline index and then, later, when it had the detail, it would estimate the subgroup indexes and these indexes would be scaled so that they add up to the previously published headline index.

3.3.3. Increasing the Visibility of the C-CPI-U

For practical reasons—most notably, its advantages in terms of consistency in aggregation, simplicity, and ease of interpretation—it is clear that BLS must use the (nonchained) Laspeyres/Lowe formula for the flagship CPI-U. Further, revisions to the flagship CPI-U are ruled out by its uses for escalation of payments. However, many of the measurement advantages of a more flexible formula and revision policy can be realized by increasing the visibility to data users of the chained CPI (C-CPI-U).

BLS can raise the profile of the C-CPI-U by, among other actions, giving it more prominence in news releases, BLS publications, and on its website, as well as highlighting its advantages for many purposes. To provide researchers and analysts with a long time series, the agency could backcast the C-CPI-U historically to 1960 or 1970 incorporating the corrections from their research price index series (R-CPI-U-RS).

Chained price indexes, precisely because they capture changing consumption patterns, also have some advantages for many purposes including indexation.[39] The chained index also likely would be preferred for many special applications, such as assessing cost-of-living changes during COVID. For many of these applications, a need for updates (especially if modest) to gain timeliness would be favorable a tradeoff. For example, a relevant COVID-19 CPI, becomes possible.[40] Much of the experimental work using alternative data described in this chapter can be carried out in the context of the C-CPI-U. Further improvements of the C-CPI-U as suggested in this chapter would provide useful comparison to the official CPI and in disruptive episodes like COVID-19 would provide one gauge of how far off weights were in the official CPI.

[39]For example, the "Simpson-Bowles" report (The National Commission on Fiscal Responsibility and Reform, 2010) recommended government-wide replacement of the CPI-W and CPI-U with the chained CPI. In fact, a modified version of the Chained CPI-U proposal was included in the Obama Administration's budget for Fiscal Year 2014.

[40]The Fisher formula uses the baskets from both the base and the current period and is well-suited for chaining. For the pandemic period, Diewert and Fox (2020) recommend using a Fisher index, linking the first post-lockdown period to the last pre-lockdown period, for measuring price changes during this aberrant period.

4
Modernizing Difficult-to-Measure Expenditure Categories: Housing/Shelter

This chapter and the next explore several major consumer expenditure categories: housing/shelter, medical care, and several products that display rapid technological progress (and, in turn, quality change) are covered in that order. These are categories of goods and services that represent significant expenditure shares for consumers and have difficult conceptual and measurement issues, but for which new kinds of data could be especially helpful in improving price measurement.

4.1. MOTIVATION

Consumption of housing services, whether rental or owner-occupied, is by far the largest component of most consumers' cost of living. With an expenditure share in the CPI of more than 30 percent, housing far outpaces the weight of other basic necessities such as food (close to 14 percent) and medical care (close to 9 percent).[1] Although shelter remains a CPI category with inherently difficult conceptual and measurement issues, new source data have become available that may create opportunities to improve price measurement for housing services.

The focus of this section is on owner-occupied housing, both because it accounts for about three-fourths of the shelter category and because the issues involved in its measurement have yet to be fully resolved.[2] Reflecting this lack of consensus, as detailed below, the statistical

[1]Shares are from May 2021. See: https://www.bls.gov/news.release/cpi.t01.htm reports the most recent expenditure shares.

[2]This emphasis is not meant to imply that there are not also important issues having to do with measuring price changes faced by renters. For example, several researchers have questioned the CPI method for sampling rental units—particularly the mix of properties that have and have not changed hands. The CPI shelter component mainly surveys existing tenants and may not fully reflect the typically larger rental price changes that take place upon tenant turnover (http://www.personal.psu.edu/juy18/index_files/ACY_inflation_2018.01.13.pdf). Another issue has to do with tenant nonpayments (see, for example, Janson and Verbrugge, 2020, and Diewert and Fox, 2020). This issue—the impact of rent nonpayment and its impact on CPI shelter inflation—has grown markedly in prominence during the COVID-19 economy. On this issue, see: www.clevelandfed.org/newsroom-and-events/publications/working-papers/2020-working-papers/wp-2022-covid19-and-shelter-inflation-reductions.

agencies of various countries use quite different methods to measure price change for shelter. In fact, 11 of 17 OECD members do not include owner-occupied housing in their CPIs at all (OECD, 2020). The measurement challenges for this spending category arise in large part because the price of housing services is not directly observable for owner-occupied dwellings and so must be imputed. In this section we discuss four different approaches to this imputation and how new data may help improve these methods, perhaps changing the relative merits of each. These approaches are the rental equivalence approach, the acquisitions approach, the user cost approach. and the payments approach.[3] Each of these methods have advantages and drawbacks, and fitness for use depends largely on the intended application.

Broadly speaking, in a cost-of-living framework the goal is to estimate the prices of goods and services consumed by households. Therefore, a cost-of-living index should attempt to measure the price of a flow of housing *services* consumed by owner-occupants, which is a concept distinct from the price of purchasing a home in that the latter contains a significant investment component. To give a concrete example, if a family buys a house for $300,000 in year *x* and lives there for the next 10 years, their housing-related cost of living is not $300,000 in year *x* and zero in the subsequent 10 years. Rather, their housing-related cost of living is the amount they would have had to spend in order to consume the same amount of housing services provided by their owner-occupied home.[4] Since the rent paid by renter households is an observable stream of expenditures used to consume housing services, one common approach is to impute the price of housing services for owner-occupied housing using rent data, called the "rental equivalence" approach. The current CPI methodology uses this approach, as do some other national statistical offices such as those in Japan, Mexico, and Switzerland. Consequently, in Section 4.2 below, we review the rental equivalence approach and how it is implemented in the case of the CPI. Section 4.3 then discusses the prominent alternative methods. This discussion is followed by consideration of opportunities for improvement brought about by new data sources and new methods.

4.2. RENTAL EQUIVALENCE APPROACH TO ESTIMATING PRICE CHANGE FOR OWNER-OCCUPIED HOUSING

The basic premise of a rental equivalence approach is that the price of owner-occupied housing services can be imputed by estimating the amount that would be required for a household to rent the same home. This premise guides production of other economic statistics as well; the *System of National Accounts* (International Monetary Fund, 1993, p. 211), for example, states as a guiding principle that "housing services produced are deemed to be equal in value to the rentals that would be paid on the market for accommodation of the same size, quality and type." In the context of the CPI, where the goal is to measure price change, the methodology seeks to estimate changes in rent and then impute these estimates to the stock of owner-occupied

[3]Diewert and Nakamura (2009) provided a detailed description of the specifications and the appropriate application of some of these different approaches to pricing owner-occupied housing (OOH).

[4]To put it another way, owner-occupants are forgoing a flow of rental income by choosing to live in the property themselves. This forgone income could be considered the cost of living in the home.

homes. The appendix to this chapter provides background information on the methodological development of owner equivalent rent (OER) estimation historically at BLS.

Much of the data for BLS's housing price program is generated by the CPI Housing Survey, which collects data on rent payments at a monthly frequency.[5] In part because rents change relatively infrequently, housing units included in the survey are sampled once every six months, as opposed to once every month or two as for most other CPI items. The units in the survey are split into six panels with data collected from each panel in a different month, allowing the BLS to compute overlapping quality-adjusted six-month changes in rent.[6] This quality adjustment accounts for changes in observable features such as number of bedrooms and the presence of air conditioning. BLS also adjusts rent for the estimated loss in quality from the aging of structures.[7] The CPI Housing survey draws from 75 urban areas and selects representative block groups within each area.[8]

For each block group in the sample, changes in average rents are extrapolated to the owner-occupied stock within that block group using information on housing unit and household characteristics from the American Community Survey. A potential concern about this implementation is that, based on a comparison with the 2017–2018 Consumer Expenditure Survey, it appears that single-family rental units are undersampled in the CPI Housing survey.[9] This undersampling might reduce the accuracy of the imputation of rent changes to the owner-occupied stock since many such units are single-family dwellings. Recent research has found that rents of single-family and multifamily units within the same neighborhood did not rise at the same rate from 2013 to 2016 (Adams and Verbrugge, 2021).

Once quality-adjusted rent changes are estimated for the owner-occupied stock in each block group, these changes are aggregated to create an elementary index for each basic area using housing expenditures as weights. The owner-occupied housing expenditures are derived from the Consumer Expenditure Survey, which asks owners to estimate the rent that they could

---

[5]CPI agents identify respondents for each sampled housing unit. The respondent may be its occupant (the renter), its owner (the landlord), a property manager or an authorized representative of the occupant (see, *How the CPI measures price change of Owners' equivalent rent of primary residence (OER) and Rent of primary residence (rent)*, https://www.bls.gov/cpi/factsheets/owners-equivalent-rent-and-rent.pdf).

[6]Detailed information on how BLS adjusts for shelter quality can be found at https://www.bls.gov/cpi/factsheets/owners-equivalent-rent-and-rent.pdf; information on shelter age bias adjustments in particular can be found at https://www.bls.gov/cpi/quality-adjustment/updating-housing-age-bias.pdf.

[7]The depreciation problem is, basically, that the same rental unit does not maintain a constant quality flow of services over time. On a month-to-month basis, depreciation will be tiny but over decades, significant upward bias can result. It is not a simple matter to make a depreciation adjustment since depreciation typically applies only to the structure part of the rental property while the rent covers both the user cost of the structure and the user cost of the land that the structure sits on (and different rental properties will have different land/structure ratios). It is not an easy problem to deal with, but some (Diewert and Shimizu, 2021) argue that a somewhat arbitrary depreciation adjustment may be better than ignoring depreciation of the structure.

[8]The 2018 geographic revision of the CPI Housing survey specified a target sample of 43,000 renters with a goal of collecting 5 renters per block group. Roughly 20 of the urban areas are self-representing, meaning that they each represent a single elementary area. The remaining urban areas are grouped into elementary areas that represent a combination of region and size class. BLS publishes indexes for each elementary area. There are an average of about 275 block groups in each of the 32 elementary areas, with larger strata having more block groups (maximum of about 850) and smaller strata having fewer (minimum of about 80).

[9]See https://www.bls.gov/advisory/fesacp1120905.pdf. Elsewhere, the 2004 ILO CPI manual recommends evaluating stratification of the renter sample by structure type (ILO et al., 2004).

obtain for the property if it were rented. A hedonic regression is used to associate this estimated rent with housing characteristics, and then the characteristics of the owner-occupied stock by block group from the ACS are combined with the hedonic coefficients to calculate the total owner-occupied housing expenditures for each block group.[10] One cause for concern is that owner-occupants have been shown to misjudge the market value of their homes (Benitez-Silva et al., 2015; Chan, Dastrup, and Gould Ellen, 2015; DiPasquale and Somerville, 1995; Goodman and Ittner, 1992; Kiel and Zabel, 1999; Molloy and Nielsen, 2018), and there is no reason to suspect they would be better at assessing rental value.[11] Another concern with this approach is that the extrapolation to block groups can only be done using housing characteristics available in the American Community Survey (ACS). Key housing characteristics that are missing from the ACS include the floor space of the structure and land area. Consequently, block groups with larger homes on more land may not be given enough weight in the CPI if the available characteristics underestimate actual structure size and land area.[12]

One big-picture drawback of the rental equivalence approach is that if owner-occupied and rental markets are segmented, it can be difficult to find rental units that are reasonably comparable to owner-occupied homes in the same block group. This difficulty likely is prevalent specifically in neighborhoods with large, more expensive homes with few rental units. According to the 2014–2018 ACS, 16 percent of block groups in U.S. metropolitan areas had fewer than 10 percent of their occupied homes as rentals. And the few available rental homes could be quite different from the owner-occupied stock along observable and unobservable dimensions. In these cases, extrapolation of rent to the owner-occupied stock could generate a misleading estimate.

4.3. ALTERNATIVE METHODS TO ESTIMATING PRICE CHANGES OF OWNER-OCCUPIED HOUSING

4.3.1. Acquisitions Approach

For some purposes, it could be useful to track changes in the cost of acquiring a housing unit. This approach accounts for the full price of a purchase to the period in which it takes place. In addition to the purchase price of housing units, information on costs associated with maintenance and repair, property taxes, and insurance may also be included. Ideally, the purchase price of the structure and of the land would be tracked separately; this split, however, can be difficult to implement in practice, since most home purchases bundle the structure and land together. Another implementation challenge is that gauging changes in price estimates over time requires accounting for changes in the quality of homes sold. These challenges

[10]The characteristics included in the regression are: number of bedrooms, property value, income, and indicators for whether the housing unit is in an apartment building, high-rise, mobile home, or “other” structure. The coefficients vary over both location and time since the model is estimated in each of the 32 index areas every quarter.

[11]Heston (2009) and Heston and Nakamura (2009) suggested that owner-occupants overestimate the income that they could collect from renting their properties, which would cause imputed rents to be biased upward.

[12]The use of property value and household income in the regression to extrapolate owner-occupied housing expenditures likely mitigates the absence of unit size and lot size information, since larger homes and homes on larger lots will likely have higher values and be occupied by higher-income households.

notwithstanding, some countries use the acquisitions approach. Australia, for example, estimates owner-occupied housing costs based on the purchase price of dwellings (excluding land) and the cost of major improvements (OECD, 2020). It is worth emphasizing, however, that home purchases involve a substantial investment component, and therefore the acquisitions approach is not necessarily a good measure of a household's cost of living over time.

In the United States, the acquisitions approach was given closer consideration during the mid-2000s since house prices were rising much faster than rents during that period. Cecchetti (2007), who concluded that the rental-equivalence approach is the most appropriate measure of changes in costs of living, noted that switching to the acquisition approach as the method for pricing owner-occupied housing would have made a substantial difference for U.S. headline CPI inflation, raising the estimate from an average annual rate of 2.8 percent to an average annual rate near 4 percent. He argued that "[h]ad these been the inflation readings, it is hard to imagine the Fed keeping their federal funds rate target below 2% for three years." Relatedly, for some purposes, Diewert and Nakamura (2009, p. 1) called for further exploration of "more direct measures of inflation for owner-occupied housing services."

Another argument that could be used in support of an acquisitions approach is consistency across CPI categories. Other goods that also provide a flow of future services, such as motor vehicles and washing machines, are included in the index on an acquisitions basis.[13] The EU's Harmonized Index of Consumer Prices Index of Consumer Prices (HICP), which is the EU's most important inflation statistic, excludes services provided by owner-occupied housing but, because it uses the acquisitions approach to measure the contribution of durable goods expenditures in the CPI, an index based on housing acquisition costs is being piloted for possible future use. However, if consistency is the goal, pricing other durable goods on a flow of services basis is another option.[14]

4.3.2. User Cost Approach

Another option for measuring price change in owner-occupied housing is the user cost approach. Sweden and Iceland use variants of the user cost approach (Hill et al., 2019). The user cost is derived from the theory that the return on investing in housing should equal the opportunity cost of investing; otherwise, more investors should bid the price of housing up or down until the return equals the cost. This market equilibration implies that the costs of owning the property—including borrowing costs, the opportunity cost of funds, maintenance, and taxes—should equal rent (the gross income from owning housing) plus the expected capital gain. The user cost is usually defined as the costs of owning minus the expected capital gain; based on the equilibration of cost and return, the user cost should be equal to rental income. Consequently, the user cost is theoretically equal to the price of housing services. With perfect data and a frictionless economy, the rental equivalence method and the user cost method should generate the same result.

[13]For a full discussion of this issue, see the chapter on durables in Diewert (2020b).

[14]The Bureau of Economic Analysis has defined consumer durables as those having an average life of at least 3 years (Katz, 1983).

A key advantage of the user cost approach is that it does not rely on an active rental market for homes of similar quality/location. Therefore, it could provide a reasonable alternative estimate for the price of housing services in cases where rental markets are thin. Notable examples include high-end properties and properties with large land components. The rental equivalence has been shown to be much lower than a user cost estimate in these cases (Garner and Verbrugge, 2009; Heston and Nakamura, 2009), possibly because nearby rents are artificially low to compensate renters for taking good care of these high-end properties.[15] For this reason, Diewert, Nakamura, and Nakamura (2009) and Diewert (2008, 2011) argued that the best approach to valuing the services of owner-occupied housing is to take the maximum of its rental value and its user cost value, using long-run expected inflation rates for expected capital gains in the user cost formula. They refer to this method as the "opportunity cost" approach because it attempts to quantify the opportunity cost of occupying a home instead of renting it out.

However, the user cost is not without its own measurement challenges. Many of the components, such as maintenance and expected capital gains, are difficult to observe and so assumptions must be made for these inputs. Research has found that user cost estimates are quite sensitive to the assumptions made about these unobserved components (Gindelsky, Moulton, and Wentland, 2019; Hill, 2019). Relatedly, measures of user cost tend to exhibit high volatility due to the method's inclusion (sometimes, depending on specification) of an interest rate term, expected capital gains, and a risk premium.

Another issue that arises with the user cost approach is that there could be an adverse interaction with monetary policy (Hill, 2004). When central banks attempt to combat inflation by raising short-term interest rates, long-terms rates can also rise if market participants expect higher rates to persist in the future. Higher long-term interest rates would raise mortgage costs, therefore pushing up owner-occupied housing costs according to the user cost and payments approaches. Thus, the central bank actions would have the unintended consequence of *raising* this large component of inflation.

4.3.3. Payments Approach

A related approach for pricing the consumption of owner-occupied housing is the payments approach, which tracks changes in the ongoing expenditures required to utilize shelter services for homeowners. These payments typically include mortgage payments, maintenance and repair, property taxes, and homeowners' insurance. An appealing aspect of this approach is that it relates directly to observable expenditures by homeowners. Indeed, if the underlying policy objective is to track out-of-pocket costs, it seems sensible to track the prices of the payments required by owner-occupants. The Canadian CPI uses a method similar to the payments approach and includes price changes for six components: mortgage interest, replacement cost (used as a method for measuring the depreciation component), property taxes,

[15] Some owners who rent out high-end homes are looking for "caretakers" to look after the property while the owner is not occupying the structure. Thus, the owners of high-end properties are not able to charge the full user cost to the renters.

homeowners' home and mortgage insurance, maintenance and repairs, and other owned-accommodation expenses.[16]

The payments approach is appealing for some applications because it does not rely on imputations, but rather can be based only on observed payments made by property owners. However, there are significant drawbacks to using this approach. The first is theoretical. Although the expenditures used in the payments approach are components of the user cost, the payments approach does not include all of the elements of the user cost, such as the opportunity cost of owners' equity, depreciation, and expected capital gains. Consequently, the payments approach does not yield a measure that is theoretically equivalent to the price of housing services. To give a concrete example, imagine two identical homes that were both purchased for the same amount at the same time. The owners of the first home financed their purchase with a mortgage for 80 percent of the home value, while the owners of the second home paid for their purchase with cash. The payments approach would assign different prices to the two homes even though the housing services provided by the homes are exactly the same. The payments approach also has implementation challenges. One challenge is how to disentangle the saving and consumption components of mortgage payments, since mortgage payments generally include some repayment of principal.[17] In addition, like the user cost, not all owner payments are easy to measure.[18]

It is noteworthy that, prior to 1983, BLS measured the services of owner-occupied housing in a way that was influenced by mortgage interest costs, specifically the mortgage rate currently being offered on new mortgages. In essence, the index was assuming that the mortgage interest that home buyers agreed to pay over future years was part of their cost of living today. It should be noted that this method of using contracted mortgage payments of consumers buying a home in the current period is substantively different from what is done in the Canadian CPI, which tracks actual mortgage payments.

The high and volatile interest rate environment of the late 1970s made the old BLS approach untenable, as it tended to amplify the volatility in inflation. These issues were large enough that the BLS switched to a rental equivalence approach beginning in 1983 (Gillingham, 1983; Gillingham and Lane, 1982). As described in the appendix to this chapter, the agency experimented with an alternative rental equivalence approach from 1987 to 1997, but then essentially reverted to the methodology used from 1983 to 1987. Since that time, following a recommendation in the report *At What Price*? (NRC, 2002), BLS has considered returning to a "payments approach" for some population subgroup indexes—notably the index used for Social

[16]https://www150.statcan.gc.ca/n1/pub/62f0014m/62f0014m2017001-eng.htm. Because the Canadian method includes depreciation of the structure, it is not a pure payments approach.

[17]This line of reasoning was also articulated by the Reserve Bank of Austria which stated that inclusion of interest charges in a measure of general inflation rates faced by consumers is problematic, conceptually, because it represents "a relative price (that of consumption in the future as opposed to the present), rather than the current price of a good or service." They further note that "in some countries where interest charges are included in the CPI, they are omitted from the CPI measure targeted by the central bank; this was the case in the 1990s in Austria" (Reserve Bank of Austria, *Submission to the 16th Series Review of the Consumer Price Index*, MARCH 2010).

[18]The case against using the payments approach is detailed more completely in the draft chapter by Diewert and Shimizu for the IMF CPI Manual (https://www.imf.org/-/media/Files/Data/CPI/companion-publication/chapter-9-treatment-of-durable-goods-and-housing.ashx).

Security COLAs (CPI-W), the elderly index (CPI-E), and in the production of a new index for low-income households. However, applying a different approach for the housing component of these subgroup indexes than for the headline inflation measure would create inconsistencies that would make differences between indexes difficult to interpret.

4.3.4. Assessment

Prominent reviews of the CPI program have endorsed BLS's rental equivalence method of pricing owner-occupied housing on the basis that it is the most broadly applicable approach for a cost-of-living index. The 1996 Boskin Commission (Advisory Commission to Study the Consumer Price Index, 1996) supported the rental equivalence approach to pricing owner-occupied housing and even argued that the method should be extended to automobiles and other durable goods. Likewise, the 2002 NRC report (p. 72) concluded that "for long-lived items like automobiles or houses...one must use not the purchase price but the consumption price" and "as is the current practice with housing, we believe that using rental rates is probably the best option."

This panel is in general agreement with the overarching recommendation from the 2002 NRC study. One appeal of the rental equivalence approach to valuing owner-occupied housing is that, in many situations, it will correspond closely to the price of housing services. Moreover, it can generally be calculated without making as many assumptions as the alternate approaches discussed above. The current panel thinks that BLS should avoid an owner payments approach since this approach is not theoretically equivalent to the price of housing services.

> **Recommendation 4.1:** BLS should continue using rental equivalence as the primary approach to estimating the price of housing services for owner-occupied units.

The rental equivalence approach is consistent with the cost-of-living index objective of the CPI and the change over time is based on observed price changes. The user cost approach is also conceptually consistent with a cost-of-living index objective but is a model-based approach that has some practical operational constraints.[19] The acquisitions approach and the payments approach are not fully consistent with a cost-of-living index objective.

4.4. OPPORTUNITIES CREATED BY ALTERNATIVE DATA SOURCES

The CPI methodology has traditionally relied on survey data to provide information on rent changes and housing expenditure shares. However, new data sources might provide alternatives or supplements to the existing surveys. Resulting from the expansion of large institutional landlords and property management companies, a number of large data sources for rent have emerged over the past several decades. For example, although institutional ownership remains only a small part of the overall market for single-family homes, the two largest

[19]The user cost approach does work well if long-term average capital gains are incorporated in the user cost formula. Moreover, the use of user costs is nearly universal in production theory, so it has an important role in some contexts.

companies owned about 130,000 units in 2020.[20] Most of these units were owner-occupied prior to 2010 and consequently should be fairly comparable to owner-occupied units. Supplementing the CPI Housing survey with single-family rental data from such sources might therefore help BLS improve its imputations of rent changes to the owner-occupied stock.

Beyond single-family rentals, some property management companies have access to rental data on millions of multifamily rental units, which would also be useful in expanding the BLS's sources on rent changes. Such data would bring a considerable advantage in that rent can be observed for a given housing unit in consecutive months, allowing for an accurate assessment of rent change in a single month. By contrast, the CPI Housing Survey only samples each housing unit every six months, and BLS assumes that the six-month change in rent is evenly spread over the 6-month period. This assumption makes the CPI slow to reflect actual changes in rents paid by households and is particularly problematic in times when rent changes rapidly (Wilcox, 2021). These new data sources may also include information on characteristics—e.g., land plot area of the rental unit, floor space, and structure age—that are essential for imputing appropriate rents for these owned units.

Ambrose, Coulson, and Yoshida (2015) provided a useful example of creating a rent index from private market data on rent contracts. They show that an index using only newly signed leases is much more volatile than the CPI rent index and that fluctuations in their index predict future fluctuations in the CPI series. They interpreted these results as indicating that the CPI methodology smooths out market conditions and reflects the conditions with a lag. They noted, however, that the CPI methodology may be appropriate for an index with a purpose of measuring cost of living, since rental market contracts and frictions do, in fact, reduce the influence of market conditions on the rent that households actually pay.

Beyond the potential of these data sources to improve the CPI's ability to reflect rapid changes in rent growth, resources saved from collecting multifamily data in the CPI Housing survey could be used to expand the survey to include more single-family units in owner-occupied neighborhoods. In addition to data on rent payments, other potential sources for rent data might include asking rents from the Multiple Listings Service or from properties posted for rent on the internet. Although asking rent may deviate from the rent actually paid by households in material ways, it might be possible to develop methods to infer contract rents from asking rents.

Despite the potential of large amounts of data from these alternative data sources, such data would never be a complete substitute for the CPI Housing Survey as these data sources do not cover all parts of the country or all strata of unit quality. It might be possible to use a survey like the ACS as a benchmark for the types of housing units and their geographic distribution across the United States, and then focus the CPI Housing Survey on unit types and areas that are underrepresented or not covered by alternative data sources.

Recommendation 4.2: BLS should seek to identify new data sources that would allow for improved coverage of single-family homes and of areas where houses are

[20]According to their 2020 annual reports, AmericanHomes4Rent owned 53,000 properties and Invitation Homes owned 80,000 properties. Large samples notwithstanding, leasing firms often operate with a limited geographic scale; for example, AmericanHomes4Rent appears to have few listings for the Northeast, Mid-Atlantic, and California. https://s26.q4cdn.com/445305060/files/doc_financial/annual/2020/AMH-2020-Annual-Report.pdf.

predominantly owner-occupied. New data sources could also improve the CPI's ability to reflect rapid changes in rent growth by allowing for the measurement of rent for a given housing unit in consecutive months.

Beyond their use for measuring rent changes, new data sources could also be helpful for estimating the housing expenditure share for owner-occupied housing. One possible source is the ACS. This large, nationally representative dataset could be used to impute rental expenditure to owner-occupied housing, for example following the method that the Bureau of Economic Analysis has begun using to estimate the consumption of owner-occupied housing services (Rassier et al., 2021). Another promising data source is property tax records, which covers nearly all housing units in the nation and contains information on many more housing characteristics than available in the ACS, including building and lot square footage. On their own, property tax records could not be easily used for the purpose of estimating expenditure shares of owner-occupied households because it is difficult to distinguish between rental units and owner-occupied units in these data. However, the property tax records could be merged with rent data from the CPI Housing Survey or the ACS to impute rental expenditure to the owner-occupied stock. With a large enough data source, the BLS could consider a machine learning approach to estimating expenditure shares instead of a single hedonic regression.

Recommendation 4.3: BLS should consider alternative strategies for estimating expenditure shares for owner-occupied housing, especially ones that would make use of the rich housing characteristics information that are often available in property tax data.

Many potential alternative data sources have an advantage of being much larger than the samples currently used for calculating the housing components of the CPI. Effective blending of data sources could ultimately allow BLS to provide additional detail about shelter prices to data users. For example, rent and owners' equivalent rent (OER) growth rates could be published for different housing types, such as single-family and multifamily units. Data on rent growth by structure type would have been particularly helpful during the COVID-19 pandemic since such data would have helped policy makers and researchers assess the shift in housing demand towards single-family structures. Publishing a few "average rent" series for a set of unit types, such as average rent of a one-bedroom or two-bedroom apartment, would also likely be of interest to some stakeholders.[21]

Given growing interest in geographic variation in the price of housing (Diamond and Moretti, 2021; Howard and Liebersohn, 2020), it would also be helpful to a range of stakeholders to publish shelter price indexes for a larger number of metropolitan areas. Geographic detail is especially important for housing because there is so much variation in prices across and within regions (Guerrieri, Hartley, and Hurst, 2013). Currently, BLS publishes shelter price estimates for the roughly 20 metropolitan areas for which it also publishes a headline index. It would be useful to publish shelter indexes for a larger set of locations, even if price changes for other goods and services could not be computed for these areas.[22] Updating the geographic sample of

[21]For example, the OECD is looking into providing detailed housing levels data.

[22]The BEA Regional Economic Accounts (https://www.bea.gov/data/economic-accounts/regional) publishes a variety of data related to output and income for states, metro areas, and counties. Elements of their methodology could be useful in creating geographic price indexes.

the CPI every 10 years may not be sufficient to capture important changes in the geography of housing, for example the reported migrations spurred by the pandemic. There could be advantages to maintaining a larger and more diverse sample of housing units across cities even if other CPI items are not priced in those areas.

> **Recommendation 4.4:** BLS should publish additional detail on the housing components of the CPI, such as indexes by structure type and for a larger number of geographic areas. Broadening the geographic scope of the CPI could be facilitated by de-linking the housing sample from the samples of other CPI items.

Even within the context of the data sources that BLS currently uses to estimate changes in the price of housing services, alternative methodologies have the potential to provide additional insight. One issue emphasized above is that the current method causes the CPI to be slow to reflect changes in rent paid by households because BLS assumes that rent change for the current month is one-sixth of the six-month change. A repeat-sales methodology, which was designed to infer high-frequency price movements from lower-frequency, overlapping price changes (Case and Shiller, 1987), has the potential to help BLS infer the rent change for the current month by comparing the six-month change in rent in one panel of the CPI housing survey to overlapping six-month changes in other panels. Such an approach would provide a timelier estimate of the changes in the price of housing services. However, the standard repeat-sales method would require allowing the index to revise back at least six months because each six-month change provides information about price changes over the previous six months, not only about the current month. Consequently, this approach would not be feasible for the published CPI-U or CPI-W, for which revisions are not allowed.

> **Recommendation 4.5:** BLS should consider publishing a supplementary CPI for housing services that would use a repeat-sales approach for inferring monthly rent changes from all six panels of CPI Housing Survey data. Such an index would provide a more accurate signal of high-frequency changes in the price of housing services.

4.5. OPPORTUNITIES CREATED BY ALTERNATIVE METHODS FOR ESTIMATING PRICE CHANGE IN OWNER-OCCUPIED HOUSING

As outlined above, imputing rent for owner-occupied homes works best when there is a high degree of overlap—in terms of geography and housing quality—between the market of homes for sale and the market of homes for rent. Imputations of rent to owner-occupied homes will be less accurate for situations in which rental and owner-occupied homes are not in the same market. Examples of such market segmentation could occur when (1) most of the owner-occupied homes in a neighborhood are single-family structures while the rental homes are multifamily units, or (2) owner-occupied homes are located in neighborhoods with little rental housing, which often seems to occur in areas with very high-quality housing units. In these cases, a user cost approach might be helpful to improve estimates of the price of housing services. Further research comparing user cost estimates to rental equivalence estimates would be valuable in helping BLS learn about the types of housing units and/or markets for which each approach would be preferable. BLS should also explore the “opportunity cost” approach described by

Diewert, Nakamura, and Nakamura (2009) and Diewert (2011), which advocates for the maximum of the rental equivalence and user cost approaches.

Alternative data sources have the potential to improve user cost estimates. For example, deeds records provide sales price data for the vast majority of property transactions in the United States, while property tax records provide the property characteristics of these homes. These data could be used to estimate the user cost for individual properties, as shown by Gindelsky, Moulton, and Wentland (2019).

Although this panel is of the view that the rental equivalence approach should continue to be the primary method used in the CPI, there could be value to creating alternative indexes using different methods. Such indexes could be useful for research purposes as well as potentially address different policy needs (such as indexation). For example, the creation of an alternative housing index based on the user cost or opportunity cost approaches would help BLS learn about the time series properties of estimating the price of housing services using these methods. Research conducted on data at the micro level would be valuable for testing where the rental equivalence method is performing well and where it is having troubles, such as for pricing housing services associated with higher-end properties.

> **Recommendation 4.6:** As part of its research program, BLS should compare rental equivalence estimates to user cost estimates for individual properties, and also explore the opportunity cost approach. Research on alternative methods for housing could lay the groundwork for eventually publishing alternative housing indexes using different methodologies.

Accurate measurement of the price of housing services is even more important for the creation of price indexes for subpopulations that have a larger share of owner-occupiers, such as the CPI for the elderly. For such indexes, the payments approach may have an even greater appeal since owner payments are, in principle, easy to observe whereas the price of housing services must be imputed. In addition, the payments method may have an appeal for price indexes used to index benefit payments for increases in inflation, like the CPI-W, since the goal of these indexes is to compensate households for inflation in the cost of consuming goods and services purchased with money. However, as discussed above, the payments made by owner-occupants reflect more than the cost to consume housing services—they also capture investment in housing. Moreover, the use of a different method for a subindex than for the headline CPI could cause confusion, because it would be difficult to know whether differential movements were due to the focus on a particular subsample, to the different methodology, or to different data sources.

> **Recommendation 4.7:** For the purpose of learning how the headline inflation measure may differ for various subpopulations (such as the CPI-E), BLS should use the same methodology for the price of housing services as it does for the headline index. For the purpose of creating alternative measures of housing-related inflation, different methodologies could be used depending on the purpose of the index.

The top priority should continue to be improving data sources and methods to improve the rental equivalence approach that is suitable for the flagship CPI-U.

APPENDIX 4A: HISTORICAL DEVELOPMENT OF OWNER EQUIVALENT RENT ESTIMATION AT BLS

To understand the methods currently used by BLS to measure owners' equivalent rent (OER), it is helpful to be aware of the challenges the CPI program encountered with the first attempt to measure this concept during the period 1987 to 1997. BLS's history of measuring OER and how the methods used influenced subsequent BLS sampling and methodologies is described by Ptacek and Rippy (2013). Prior to 1983, BLS measured the services of owner-occupied housing largely based on mortgage interest costs. The high inflation and double-digit interest rates of the late 1970s made that approach untenable, and there was wide consensus that another approach was needed. After investigating the user cost and OER alternatives, in 1981 BLS announced that it would phase in the OER approach beginning in 1983. This involved reweighting the tenant-occupied rents to derive the OER estimates, along with some sample augmentation of tenant-occupied units in heavily owner-occupied neighborhoods (Gillingham, 1983; Gillingham and Lane, 1982).

Initially, as Ptacek and Rippy (2013) described, BLS simply reweighted the tenant-occupied rents to derive the OER estimates, along with some sample augmentation of tenant-occupied units in heavily owner-occupied neighborhoods. Beginning in 1987, however, BLS switched to a sample that was selected specifically to implement the OER concept.

The housing survey that BLS used from January 1987 to January 1998 was designed to implement the OER concept as understood by BLS staff in the mid-1980s. BLS selected a sample of renters to represent the tenant population and a sample of owners to represent the owner-occupant population. The owners were then linked to a supplemental sample of renters that were intended to be matched to those owners and be used to impute their changes in OER, unit by unit. For example, BLS might select an enumeration district (a Census geographical unit that was similar to a block group) that was predominantly owner-occupied—for example, 94 percent owner-occupied. Within that enumeration district, a neighborhood of perhaps 50 homes was selected and screened to find which ones were occupied by owners and tenants. Then, if for example, 47 units were owner-occupied and 3 were tenant-occupied, a sample of 2 owners and 2 tenants from that neighborhood might be selected to participate in the CPI housing survey. The selected units would complete a survey giving a detailed description of the characteristics of the unit. The tenant units would be resampled at 6-month intervals to capture rent changes. Originally, the intention may have been to resample the owner units less frequently (maybe once every couple of years), but because CPI rent changes of the owner-occupied units were being imputed from the tenant-occupied units, they soon decided that it was unnecessary to resample the owner units.

The imputation procedure was complicated, but the essence was that each owner unit in the sample would be matched to one or more tenant units, then the average rent changes of the matched tenant units would be used to impute the rent changes of the owner units. Each owner unit would be imputed in the index calculation at six-month intervals, matching a panel of tenant units sampled at the same frequency. The matching algorithm was run each time, so there was no guarantee that the owner unit would be matched to the same tenant units every time, although the algorithm was designed to usually make the same matches.

The matching algorithm went through a hierarchy of about six levels, first trying to find the closest matches, and then if close matches were not found, dropping to a lower level. The levels emphasized geography. To take the earlier example, if the two owner units and two tenant units were all single family, the algorithm would probably match each owner unit to both tenant units. If one of the tenant units was in a multifamily apartment building, however, then both owner units might be matched to the one single-family tenant unit. The algorithm always favored units from the same neighborhood, but if no tenant units were matched in the same neighborhood, it would try to match units from a broader geography.

In practice, the method encountered several problems. One was that the tenant-occupied units in predominantly owner neighborhoods tended to have high attrition. Owners sometimes rent out their own homes for short periods while away for limited periods, such as an employee on a one- or two-year assignment at a field office. While these short-term rental arrangements are not especially prevalent in aggregate, they may be an important source of rentals in predominantly owner neighborhoods.

As time went by, the imputed changes in OER were being driven by fewer and fewer tenant units, and the few tenant units remaining in owner neighborhoods were carrying much of the weight. Since the introduction of the 1987 CPI housing sample, the OER index had persistently increased at a higher rate than the tenant index, with a difference of about 1 percentage point per year (see Figure A4-1). At the time, the reason for the difference was not understood by the CPI program staff, as they regarded the imputation procedure they were using as essentially a reweighting of the prices used in the tenant index. While it was thought possible that the difference reflected a real difference in the inflation rates of the types of tenant units that were most similar to owner-occupied units, there was also a concern that there may have been a flaw in the matching process, which was considered to be a bit of a black box.

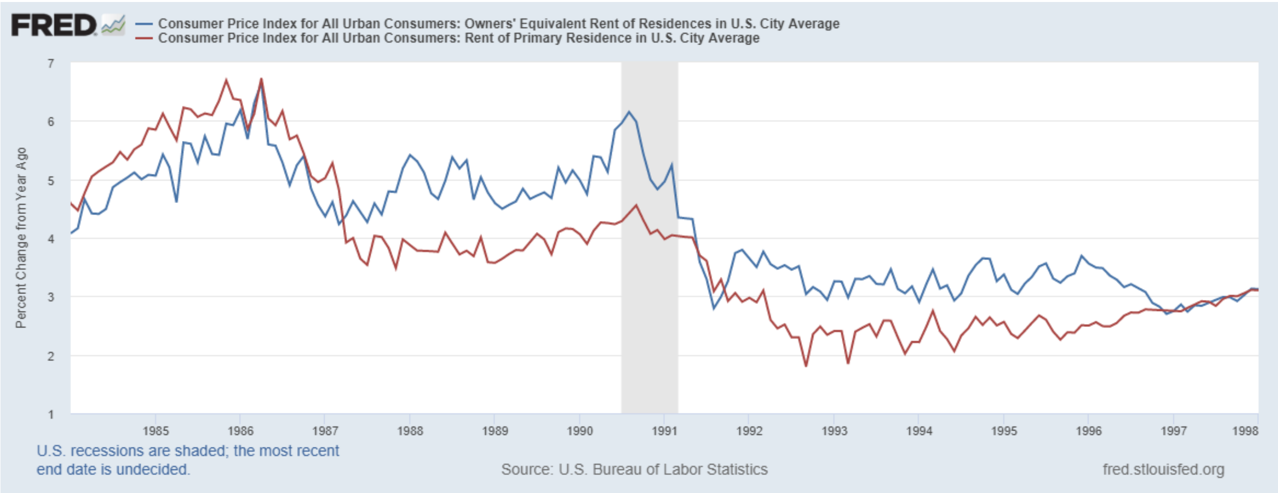

FIGURE 4A-1 CPI for owners' equivalent rent and for rent; percent change from year previous, 1984–1998.
SOURCE: BLS, St Louis Federal Reserve.

In the imputation algorithm, the 6-month rent changes for OER were calculated using a Carli formula—that is, the imputed owner's equivalent rent for owner unit *j* in month *t* was calculated based on the average of the price relatives for the rents of the matched tenant units:

$$OER_{j,t} \;=\; OER_{j,t-6} \;\times\; (1/n_j)\sum_{i\in Q_j}\frac{R_{i,t}}{R_{i,t-6}},$$

where $R_{i,t}$ is the rent for a matched tenant unit *i*, and Q_j is the set of tenant units that are matched to owner unit *j*. It is now well known that the use of the Carli index for such calculations impart an upward bias to the resulting elementary index (which in this case is an imputed rent for an owner unit).

This problem was described by Armknecht, Moulton, and Stewart (1995), which refers to the average-of-relatives formula as a "Sauerbeck" index. Diewert (1995) attributed the index to its earlier discoverer, Carli, and now the term "Sauerbeck index" is typically used only when the relatives have unequal weights.) Beginning in January 1995, BLS switched to using a Dutot formula for the imputation:

$$OER_{j,t} \;=\; OER_{j,t-6} \;\times\; \frac{\sum_{i\in Q_j} R_{i,t}}{\sum_{i\in Q_j} R_{i,t-6}}.$$

The upward bias did not immediately go away. Additional research showed that an upward bias persisted with the Dutot formula when the imputations were based on a very small number of matched tenant units, so about a year later, BLS began requiring a minimum number of units to be matched or else the matching algorithm would move to the next level up. With this additional adjustment, the upward bias of the OER estimator had ceased by 1997.

However, by imposing this requirement on the matching algorithm, the method was no longer routinely able to match owner units to rental units in the same neighborhood, which undercut the rationale for the overall methodology.

In 1998, BLS moved to a new housing survey and switched its OER methodology to resemble what had been used prior to 1987. While the panel does not know the full details of the motivation behind the shift, it seems that, at least in part, the new approach was driven by an understanding that it was problematic to attempt to closely match owner units to similar tenant units—especially in neighborhoods that were predominantly owner occupants—and that the most practical approach was to attempt to reweight the tenant units to more closely resemble the population of owner units.

5
Modernizing Difficult-to-Measure Expenditure Categories: Medical Care

5.1. MOTIVATION

Medical care is a large, growing, and at times rapidly changing expenditure category, and identification and measurement of the sector's prices and quantities are conceptually complex. The domain of medical care in the Consumer Price Index (CPI) is limited to prices for goods and services on which consumers make direct out-of-pocket outlays: (1) health insurance; (2) prescription drugs; (3) over-the-counter drugs and medical supplies; (4) services from physicians, dentists, and other medical professionals; and (5) hospitals and related services.

For the CPI, the expenditure weight assigned to medical care reflects the share of medical services directly purchased by consumers. Health insurance premiums (individually purchased, Medicare Parts B and D premiums, employee share of employer-sponsored insurance) and direct payments (retail purchases, deductibles, copayments, and coinsurance) are therefore in scope while Medicaid, Medicare Part A, Veteran's, and employer-paid health insurance premiums are out of scope.

The focus on consumer purchases renders the weight of health care in the CPI much lower than it is in the National Income and Product Accounts (NIPA), which covers the entire health care sector (not just the goods and services purchased directly by consumers). Reflecting these different objectives, the weight of medical care in Personal Consumer Expenditures (PCE) inflation in 2020 was 22.3 percent whereas the weight in CPI inflation was just 8.9 percent (Figure 5-1).

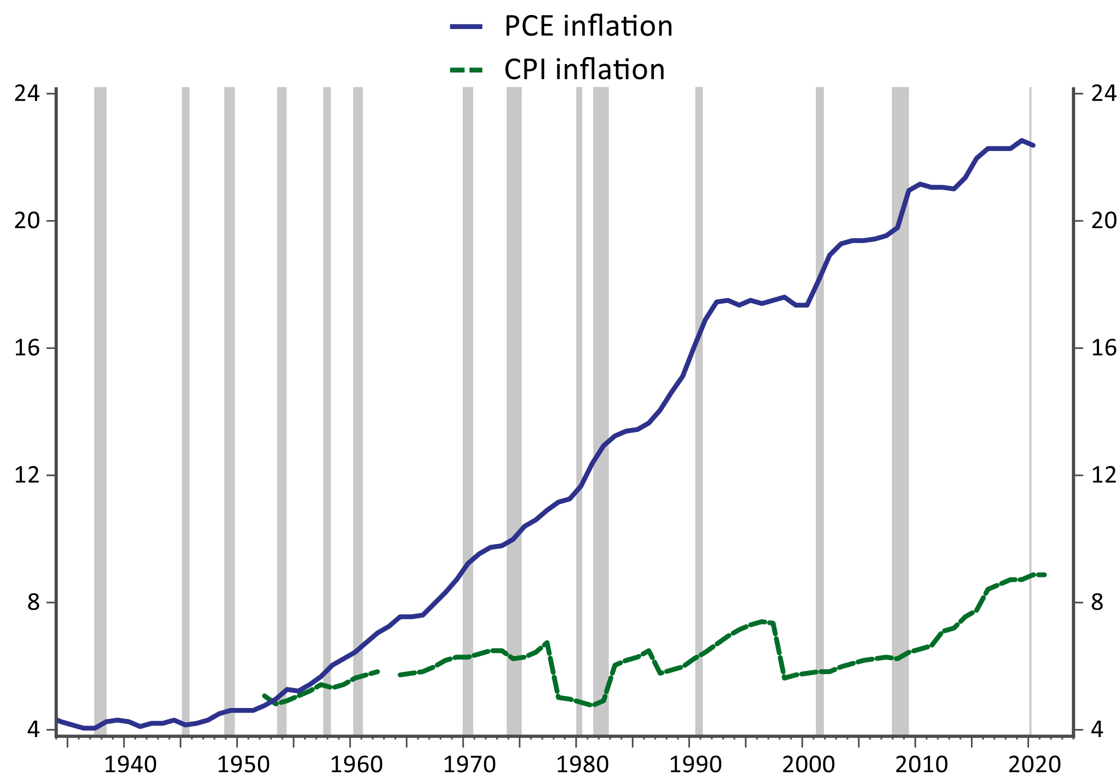

FIGURE 5-1 Weight of medical care (including health insurance) in consumer inflation (%).
SOURCE: Panel-generated using BLS data.

Within the more limited scope of the CPI, health insurance premiums are of primary importance, accounting for around 70 percent of consumers' medical care expenditures (see Figure 5-2). However, as described below, most of consumers' payments for health insurance premiums are reassigned to the other medical expenditure categories. Three other major categories—medical services, drugs, and medical supplies paid directly by consumers—make up the balance.

Accordingly, the primary focus of the panel in the area of medical care is assessing the current methodology of the Bureau of Labor Statistics (BLS) for measuring health insurance costs faced by consumers. In particular, we focus on the pros and cons of two different conceptual frameworks for pricing health insurance: the indirect method, currently used in the CPI, and the direct method, currently used in the Producer Price Index (PPI) for health insurance. This chapter also considers how various data sources (insurance filings, claims data, hospital data, scanner data on drugs, etc.) could be used to improve the coverage, detail, and timeliness of the CPI medical care index more broadly, more or less within the current framework.

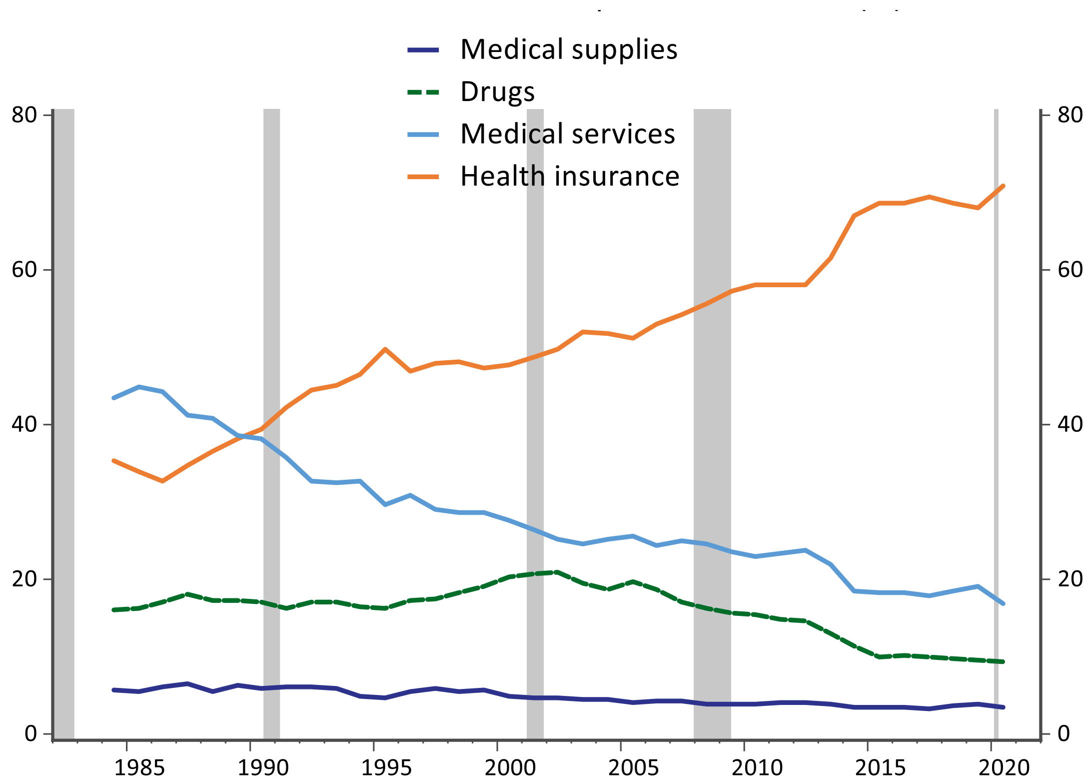

FIGURE 5-2 Share of total health care expenditure in the CPI.
SOURCE: Panel-generated using BLS data.

BLS collects many medical care services prices through its commodities and services survey although it receives a large volume of data on prescription drug prices from one firm.[1] As discussed in Chapter 2, there are growing concerns that BLS's current surveys are becoming less viable and producing less accurate data over time. The medical component of the CPI currently has the lowest response rate among the different commodities and services groups. As shown in Figure 5-3, that rate has dropped significantly—down to around 40 percent over the past decade—reflecting physicians and hospitals that are not responding to requests for price information about their services.

[1]https://www.bls.gov/cpi/factsheets/medical-care.htm.

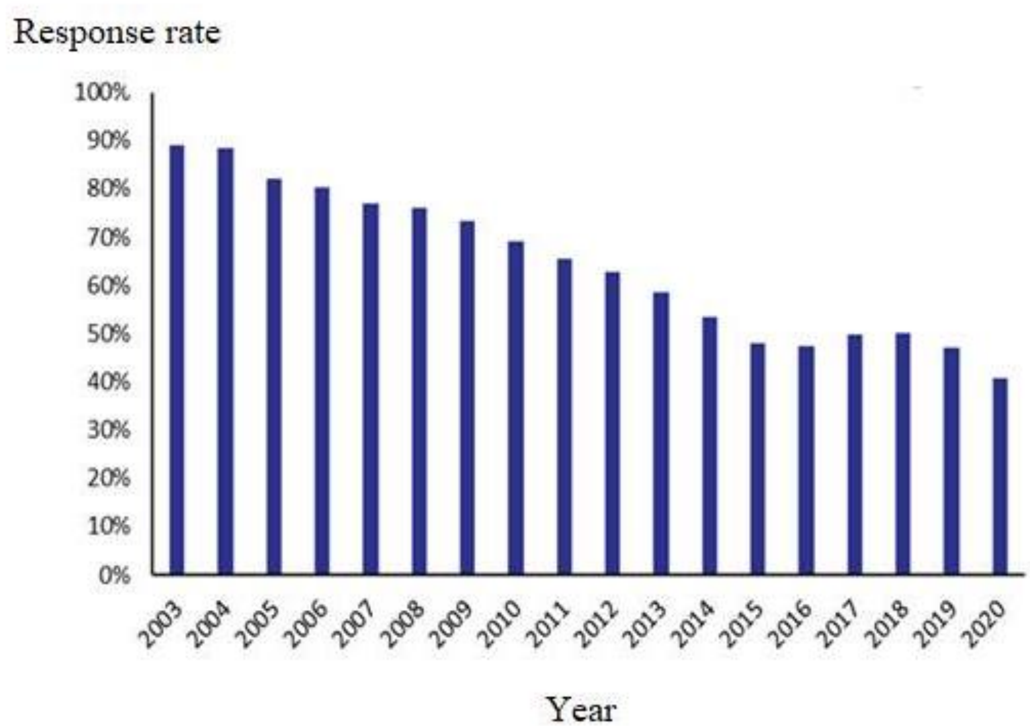

FIGURE 5-3 Medical care response rates (percentage collected out of total eligible).
SOURCE: Panel-generated, from BLS data (https://www.bls.gov/cpi/tables/response-rates/home.htm#Archived%20Response%20Rates).

Much previous work has been devoted to conceptual measurement questions related to how health care prices should adjust for quality change. The productivity literature has developed a conceptual framework for defining the inputs of medical care in such a way that they can be tied to the outputs and then prices appropriately quality adjusted to reflect resulting improved health outcomes.[2] Much of this work has been developed in the context of national income accounting and, for reasons discussed in Section 5.3 and elsewhere (NRC, 2002), may not be completely translatable to the CPI. One argument for a more limited approach to quality adjustment in the CPI is that the uses to which it is put—for example, to index payments to reflect cost-of-living changes on an annual basis—are distinct from that of measuring health care inputs/outputs as covered in the PCE (to reflect welfare).

Because the specifications of a broader health/health care account have been articulated elsewhere, and because the charge here is to provide guidance targeted at data modernization, this report does not delve deeply into quality change measurement for the sector. It is worth reiterating, however, that alternative data sources could be helpful for gauging quality change, allowing for a more careful and detailed tracking of the characteristics of medical care inputs that

[2]Dauda, Dunn, and Hall (2019) examined quality-adjusted prices and productivity for three acute conditions; Dauda, Dunn, and Hall (2020) evaluated 8,000+ cost-effectiveness studies; Romley et al. (2019) conducted event studies for 8 conditions; Sheiner and Malinovskaya (2016) and Cutler et al. (2020) examined productivity for the Medicare population.

consumers purchase—such as insurance policies—and of the outcomes experienced by consumers who undergo various medical procedures and other treatments.

5.2. PRICING HEALTH INSURANCE

Due to its relatively high expenditure share of medical care costs, accurately tracking the prices of medical services purchased through health insurance is of critical importance to accurately estimating price changes faced by consumers. We discuss two different options for pricing health insurance—the indirect pricing method currently used in the U.S. CPI and the direct method used for the PPI, highlighting the pros and cons of each. As explained below, the direct approach involves estimating total health insurance premium prices, while the indirect approach involves pricing health insurance using information about retained earnings blended with changes in the price of medical care.

5.2.1. The Indirect Method

The CPI decomposes spending on health insurance policies into two components: (1) the expenditures related to the provision of services provided by health insurance companies—insurance, claims processing, and utilization management; and (2) the expenditures used to pay medical providers for care, as shown in Figure 5-4. In this conceptualization of health insurance, the insurance company can be viewed as acting as an intermediary that does the actual purchasing of medical services and then "sells" these services to households. Households thus are viewed as purchasing medical services from medical providers, albeit indirectly, and purchasing insurance services from the insurance company. BLS allocates the part of the health insurance premium that goes to paying for medical services—which amounts to roughly 80 percent—away from the "health insurance" category and into the medical services categories. The remaining 20 percent—the premium less the cost of the medical benefits, or what BLS calls "retained earnings"—is labeled as "health insurance."[3] Note that retained earnings include not only profits earned by the insurance company, but also all nonmedical costs, including claims processing, advertising, and taxes.[4]

[3]Sometimes confusion arises in discussions of health insurance prices because, for the PPI, the price of health insurance refers to the price of the whole health insurance policy, without any reallocation, whereas for the CPI, it refers to the prices of only the services directly provided by the health insurer.

[4]There may also be some confusion related to the term "retained earnings," which in this usage differs from its more standard usage in accounting. It can be thought of as the implicit service charge for the services provided directly by the health insurance industry, net of the medical services that are bundled with the premium.

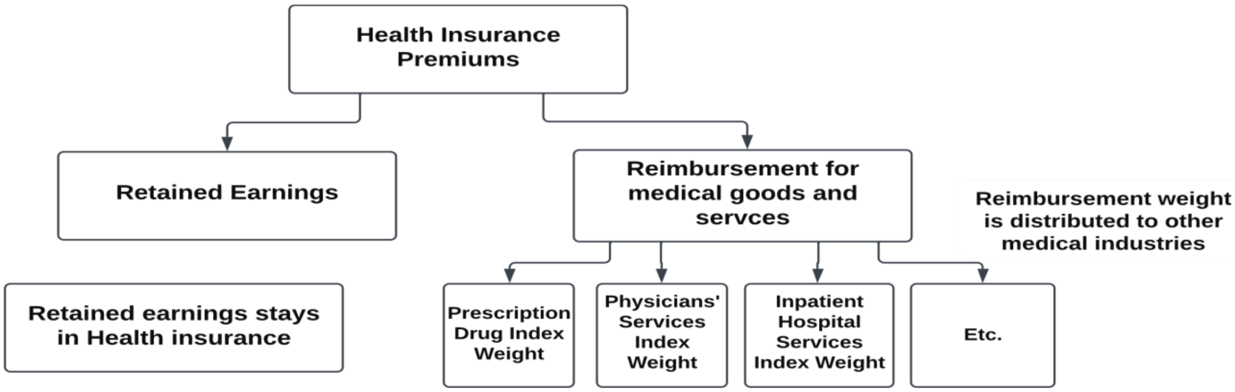

FIGURE 5-4 Health insurance weight redistribution.
SOURCE: BLS, https://www.bls.gov/cpi/factsheets/medical-care.htm.

Figure 5-5 shows the relative importance of the components of health costs in the CPI *after* this reallocation has been done. Once the medical expenses paid for through insurance are relocated to medical care, health insurance is a much smaller share of the medical care CPI—just 14 percent in December 2020 (as opposed to the "pre-reallocation" 70 percent shown in Figure 5-2). To calculate the medical CPI, BLS obtains price quotes for each of the listed components and weights them accordingly. For example, for physician services, BLS obtains prices for the same service at the same physician over time. To calculate the price of health insurance services, BLS uses estimates of changes in the ratio of insurance company retained earnings to medical benefits, which they collect once a year.[5] This annual relative is converted to a monthly relative for the regularly published CPI by assuming a smooth change in prices over the year.

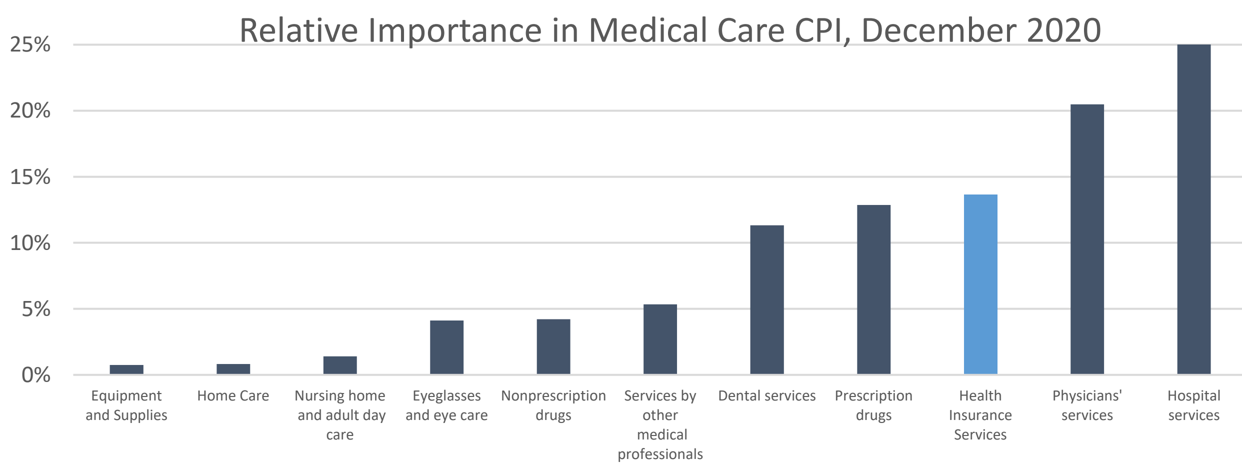

FIGURE 5-5 Relative importance of components in medical care CPI, December 2020.
SOURCE: Panel-generated using BLS data.

[5]The method is described in detail in BLS (2001) and in the CPI Medical Care Fact Sheet.

Several concerns have arisen concerning use of this indirect method. One is that the data on retained earnings are compiled annually and are available only with a considerable lag (about 10 months). To address this problem, the CPI program is considering the use of quarterly data on retained earnings instead of the annual data currently used.

Additionally, the insurance component of medical care is volatile. As shown in Figure 5-6, there is much more volatility in the prices of the health insurance component than in the other medical service components. This volatility likely stems from the fact that retained earnings are a residual and a relatively small share of the health insurance premium. Thus, unexpected changes in utilization from year to year could lead to large swings in the ratio of retained earnings to medical expenditures and therefore to the price of health insurance in the CPI. For example, when utilization increases more than expected, retained earnings will fall and retained earnings as a share of medical expenditures will fall even more. In contrast, expected changes in utilization will boost the premium and not have much effect on retained earnings. This volatility was likely a particularly big issue in recent years given turmoil in the individual and exchange-based health insurance market associated with the Affordable Care Act (ACA).

Finally, the approach does not attempt to adjust for changes in the quality and quantity of the insurance services provided. For example, if the risk of catastrophic medical expenses increases over time, the value of insurance would likewise increase, even holding expected medical expenses constant. Similarly, if a particular policy increased in price because it lowered the probability of "surprise billing," that price increase should be viewed as an increase in quality, not an increase in price as would be recorded in the CPI.

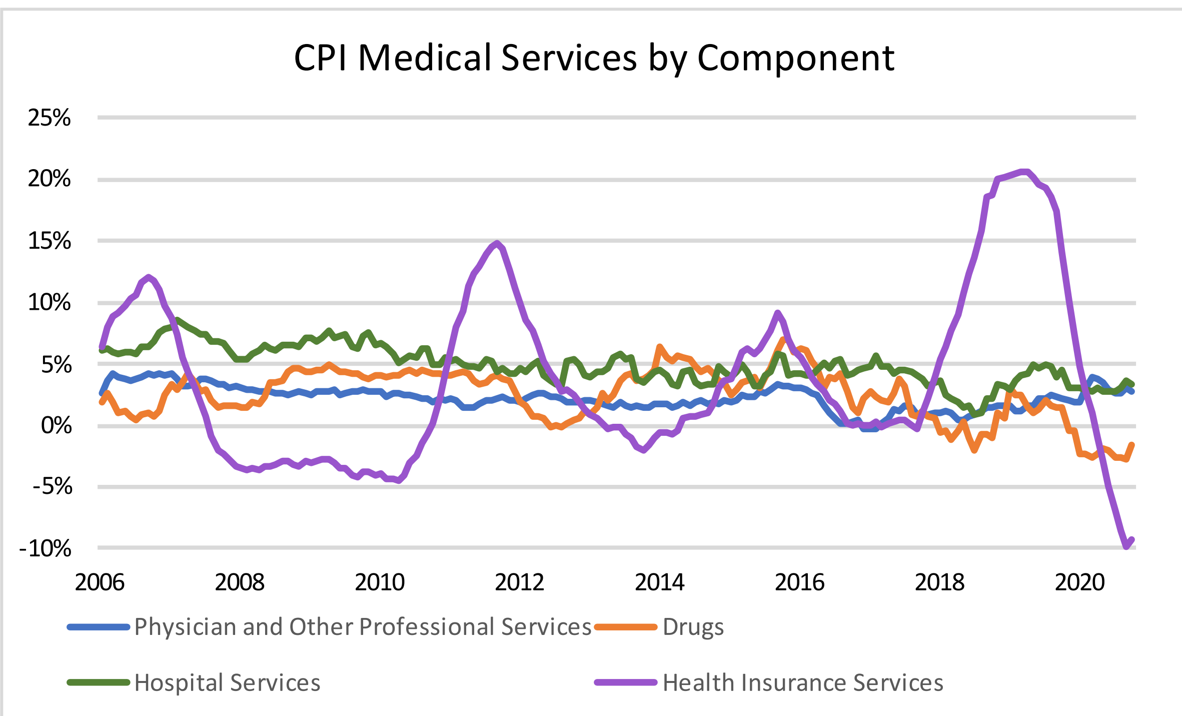

FIGURE 5-6 Prices of medical service components of the CPI, 2006–2020 (percent change).
SOURCE: Panel-generated using BLS data.

5.2.2. The Direct Method

BLS uses the direct method to measure the prices of health and medical insurance carriers in the PPI. Unlike the indirect method, the direct method prices health insurance premiums as a whole rather than as a residual. The method involves, first, selecting a set of health insurance policies to price, then, for each policy, tracking the price of the policies over time holding constant the age, health conditions of the applicant, coverage, deductibles, and copays.

A data issue for the direct method is that identical policies are not always available (or offered) for comparison from one period to the next. In such cases, BLS asks the insurance companies to assign a value to the risk change associated with the change in policy characteristics to maintain constant quality. Since policies are repriced when they are renewed, companies should in principle be able to provide this information in the rate determination for a given year. Broadly speaking, it is a challenge to estimate the health insurance value to consumers in terms of risk management.

5.2.3. Comparing the Methods

A key question is whether BLS should switch its CPI approach of pricing insurance from the current indirect method to a direct method more akin to that used in the PPI. Of course, the appropriate price index will depend on what it is used for. As discussed below, from a cost-of-living perspective, an increase in spending on health insurance that reflects increased health care expenditures from improvements in technology is an increase in real spending, not in prices. Because the CPI is intended to be a conditional cost-of-living measure, an ideal index would fully control for changes in quality and utilization over time, and we evaluate the two methods under this framework.[6]

It is worth noting that BLS has been weighing the relative merits of the different approaches to pricing health insurance for some time. In response to the Boskin Commission's (Advisory Commission to Study the Consumer Price Index, 1996) estimation of an annual upward bias of 3 percentage points for hospital and related services in the CPI, BLS began investigating approaches to addressing problems cited in that report which mainly had to do with the difficulty of capturing the benefits of new and improved technologies and treatment methods. One approach was the direct pricing of health insurance done through the collection and quality-adjustment of health insurance premiums data. But, at the time, BLS decided that "the problems of adjusting premiums for utilization and other quality changes were prohibitive" (Greenlees, 2006, p. 33).

---

[6]Exactly how to control for quality is a very complicated issue, particularly because the health care system tends to allow the consumer limited choice about the quality of health care. Thus, an improvement in quality that is worth its cost to the average consumer and hence counted as a price decrease may be viewed as a price increase by consumers who do not value the improvement as highly.

When it comes to an appropriate index for benefits like Social Security or child tax credits, policy makers may want to ensure that increases in the "costs" of health care—even if stemming from improvements in health care—lead to an increase in benefits such that households would be able to afford both the improved health care and the same basket of non-health goods as before. We view that as a policy question that is beyond the scope of this chapter.

As noted, one source of discomfort with the indirect method is the volatility of the price relative for health insurance in the CPI. Indeed, comparing price change for health insurance in the CPI to that as measured by the PPI (Figure 5-7, left panel), the latter appears less volatile. But this is not a valid comparison because, as noted above, health insurance services in CPI are quite different in scope from health insurance services in the PPI. The PPI prices the whole policy rather than splitting it into health insurance and benefits components as is done in the CPI.[7] A more relevant comparison would be between the PPI and estimates of CPI price changes for the whole insurance policy using a weighted average of the health insurance component and the medical price components. However, BLS does not publish the data necessary to perform that calculation.[8] As shown on the right side of Figure 5-7, when comparing total medical services in the CPI (which include health insurance costs) to the PPI health insurance index, no obvious pattern of differences in variability emerges.

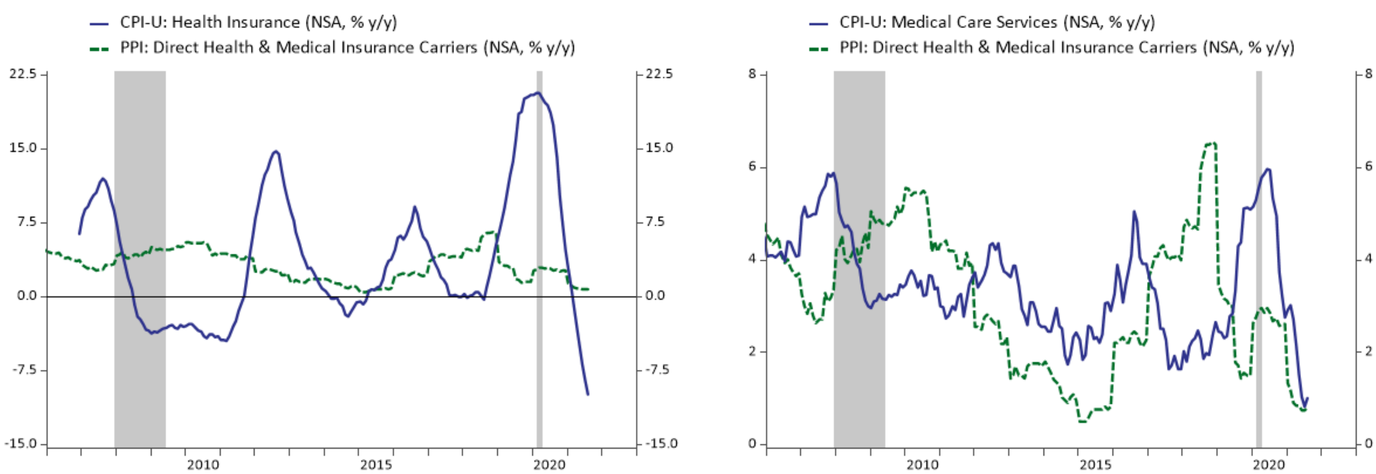

FIGURE 5-7 Performance of CPI and PPI health insurance price indexes.
SOURCE: Panel-generated using BLS data.

Conceptual Comparison

The indirect and direct methods differ conceptually in three ways, which are described mathematically in Appendix 5A. In order to provide a meaningful conceptual comparison between the two methods, we focus on the "whole policy" price under the indirect method—that is, the weighted average change in the price of CPI "health insurance services" and the price of the medical services financed by health insurance. This is the price change the indirect method implicitly uses for the change in price of health insurance policies and is directly comparable to the price change under the direct method.

---

[7]The scope of the PPI and the CPI differ in other ways as well. For example, since the PPI includes all revenue received by insurers, employers' contributions are in scope. These contributions are excluded in the CPI, as is Medicaid, much of which is provided by commercially managed care companies that are within scope of the PPI. And, of course, the CPI Medical Care index also includes prices of items not paid for by insurance.

[8]It is difficult to even do a rough estimate because the shares of different components—hospital, physician, health insurance, drug, etc. in covered medical expenses—have shifted a lot over time.

Three conceptual differences between the direct and indirect methods are as follows:

1. Although the measurement ideal of both methods is the same—to track prices of constant quantity, constant quality insurance policies—the direct method prices a fixed health insurance policy in such a way that adjustment for expected changes in utilization is limited,[9] while the indirect method attempts to capture the pure price changes in health insurance as distinct from quantity changes.
2. The indirect method is an ex-post measure. Because it relies on retained earnings, it asks: What were the costs of insurance per unit of health care that a consumer experienced? The direct method is an ex-ante method, capturing what the consumer actually paid for insurance, which does not depend on whether utilization turned out to be unexpectedly high or low.
3. The indirect method relies on BLS's price measures for underlying medical services while the direct method does not. As discussed below, in some cases, the direct method will do a better job of capturing cost savings that are the result of substitution between medical goods and services.

As we discuss below, neither method is an ideal measure of health insurance price inflation, but the panel believes that the indirect method is on the whole less likely to conflate quality/quantity changes with price changes.[10]

Changes in Utilization

The direct method simply prices health insurance policies, holding fixed aspects like the demographics of the policy holder and the financial attributes of the policy—coinsurance, deductibles, and the like. If health utilization increases over time for a given set of policy attributes, the direct method will attribute those increases to health insurance prices. In contrast, under the indirect method, expected changes in utilization that do not affect retained earnings per unit of medical care are not counted as an increase in the "whole price" of health insurance (the health insurance services plus the medical services financed through insurance). That is because the indirect method is based only on the price of medical services—e.g., the price of a cataract operation—and the retained earnings of insurance companies. An increase in premiums because an insurance company expects to pay for more cataract operations, for example, will not show up as higher medical price inflation using the indirect method, but it will using the direct method.

Indeed, Appendix 5A shows that, abstracting from data issues, **the direct method price relative is equal to the indirect method price relative multiplied by a factor reflecting the growth in utilization**. When utilization increases from one period to the next, the direct method will show a larger increase in prices than the indirect method; the converse is true when utilization falls.

---

[9] The direct method used for the PPI does attempt to hold constant expected utilization tied to individual risk factors. However, exogenous factors shifting utilization independent of enrollee characteristics (e.g., introduction of a new treatment) will generally not be captured.

[10] Indeed, the problem of not measuring the quality/quantity of financial protection offered by the insurance policy is common to both approaches.

A key question is the extent to which changes in health insurance costs over time result from improved quality or increased quantity of the health services covered by that insurance, and the extent to which they represent increases in prices for the same treatments. Most economists believe that the increase in health spending over time is largely the result of improvements in technology, rather than increased costs for the same services (Smith, Newhouse, and Freeland, 2009). Some new technologies are more expensive but also have better outcomes. For these technologies, spending increases over time both because treatments are more expensive and because the improved outcomes may make the treatments helpful for more patients. For example, the technologies used in renal dialysis therapy have gotten more expensive but better over time, leading to shorter treatment times and lower side effects. As a result of these improved outcomes, the clinical criteria for prescribing such therapy have expanded, leading to a dramatic rise in the number of dialysis treatments over the past 30 years (CBO, 2008; CDC, 2021). Other technologies lower treatment costs per patient, but still lead to higher health spending because of their wider use. For example, improvements in cataract surgery greatly reduced the cost of treatment but led to a four-fold increase in the incidence of the surgery (Shapiro, Shapiro, and Wilcox, 2001) and increased spending on cataract treatments overall.

A consumer buying health insurance in 2022 is getting a lot more insurance than they were in 2000—they are getting more and better treatments if they have cancer, have a heart attack, get hepatitis, or have other conditions. These services should be viewed as an increase in the real quantity of health care consumed—or, equivalently, as an increase in the quality of the health insurance policy—not an increase in prices. But because the direct method simply prices a "health insurance policy" with a set of attributes, it will attribute most of these increases in the quantity and quality of health services to health insurance inflation.

When major changes in technology occur that are expected to significantly increase health spending (e.g., the new drugs to treat Hepatitis C), BLS asks insurance companies to price coverage with and without the cost of the new technology so that they do not attribute the increased premium to prices, but it is not clear how effectively this is done. Furthermore, most increases in utilization are less dramatic and unlikely to be picked up in this manner. For example, the improvements in cataract surgery described above occurred gradually throughout the 1970s, 1980s, and 1990s, and it is very unlikely that the direct method would adjust for these types of quality and quantity changes.

Ex-Ante versus Ex-Post Utilization

The indirect method is an ex-post measure of the price of health insurance services—it is based on insurance companies' actual profits after the fact. If utilization is unexpectedly high, insurance companies will have lower retained earnings because they will have paid out more of their earned premium income through claims; if utilization is unexpectedly low, the converse will be the case. (In contrast, if utilization is expected to be high, that will show up in premiums and not retained earnings.)

Thus, the indirect method asks: What were the costs of health insurance services per unit of medical services for someone who bought a health insurance policy in a particular year? If utilization was unexpectedly very high, the measured price of insurance services per unit of health care will be low (both because the numerator of the price [retained earnings] is low and the denominator [quantity of health services] is high), and if utilization was very low, the measured price of insurance will be high.

But the price a consumer faces when choosing whether or not to buy a policy is an ex-ante price—consumers have to purchase insurance before they know whether there will be a bad flu season or some other unexpected occurrence that affects utilization (like COVID-19). When they purchase insurance, they are buying a product that will pay for (some of) their expected health care expenses and for protection from a catastrophic financial loss due to a large medical expense that might happen in the future.

The direct method, in contrast, is an ex-ante price—it simply measures the price a consumer pays for a particular health policy at a given time (holding constant, to the extent possible, the features of that policy like deductibles and coinsurance).[11] If there is a bad unanticipated flu season, the direct method will not show a lower price of health insurance. From this perspective, the direct method seems preferable and akin to how other "composite" goods and services are priced. For example, the cost of a gym membership does not depend on how many times people visit the club.[12] As discussed below, however, health insurance companies are sometimes required to provide rebates to their customers if utilization turns out to be too low relative to premiums.[13]

The ex-post perspective creates volatility in the index, but it does not bias it. Years of higher-than-expected utilization will be balanced by years of lower-than-expected utilization and higher costs.

Quality Adjustment Problems Using the Indirect Method

Although the discussion above suggests that the indirect method is less likely to conflate increases in quantity and quality with prices, it is not perfect. The indirect method relies on CPIs from BLS for physician services, hospital services, prescription drugs, and other costs to capture the changing prices of the medical services provided through health insurance. The prices of medical services as measured by BLS do not fully control for changes in the quality of the services (Dunn, Rittmueller, and Whitmire, 2015), meaning that the indirect method will still ascribe some increases in quality to higher prices. For example, the consumer price index for

[11]BLS uses the direct method to price other types of insurance in the CPI—e.g., auto and homeowners' as well as property and casualty insurance. To price those policies, BLS holds the characteristics of the insured property constant, like the type of car for auto insurance and the size of the house for homeowners' insurance. In these cases, there is less concern that ongoing quality changes in the methods used to address damages—e.g., in the methods used to repair cars—will lead to quality changes being assessed as price changes. One potential counterexample is the effects of climate change on homeowners' insurance. If climate change boosts insurance premiums because it increases the severity and frequency of damages, it is unclear whether the CPI should treat those higher premiums as higher prices, and how much they should be viewed as higher quality, since any given insurance policy will be providing a greater degree of financial protection and expected payouts.

[12]We adhere to the view of the CPI as a "conditional cost of living index" where the conditions would include whether or not there was a bad flu season or pandemic. Thus, we do not think it is a problem for consumers to appear "better off" from the purchase of insurance during a pandemic—the pandemic made people worse off, but the insurance itself made them better off. Our point is more about the knowledge people have when they choose to purchase insurance. If they do not anticipate the pandemic, then they will not view insurance as having a low price even if, ex post, after the pandemic began, it is clear that they are much better off from having the insurance.

[13]The question of how to treat rebates makes this discussion of ex ante versus ex post approaches more complex. We turn to those issues below.

hospital services might price the costs of a hospital stay for a particular cause—say a hip replacement—over time. If the outcomes from hip replacements are improving because of new technologies that cost more, this increase in quality will not be captured and will instead appear as a price increase for hospital services. Of course, both the direct and indirect methods will miss this quality improvement.

But there is one situation in which the direct method may better capture quality changes than the indirect method. In the direct method, a switch from one form of treatment to another—say a switch to medical management from surgery, or a cost-saving switch from inpatient to outpatient care—will count as a lower price of health insurance, whereas in the indirect method (which relies on CPIs from BLS that view inpatient and outpatient care as separate services), it will not. In other words, the direct method solves the outlet substitution bias that plagues many other categories of spending in the CPI.

It is an empirical question which method overstates the true rise in health insurance costs more. The indirect method does not count increases in utilization over time as price increases, but it also does not capture any savings from some cost-saving shifts in treatments. Examining changes in total spending by illness for a group of commercially insured individuals from 2003 to 2007, Dunn, Liebman, and Shapiro (2014) examined changes in the costs of treatments for specific conditions over time—thus capturing any costs changes from modifications in the location or method of treatment—and found that the shift toward greater intensity of service use outweighed the savings from shifting forms of care; however, over this time period, the number of treatment episodes per beneficiary increased 1.7 percent per year. Thus, for this population and time period, the direct method (which would include the increase in the number of treatments as an increase in price) would have overstated inflation in health insurance costs substantially more than the indirect method.

In an attempt to provide more intuition about the two methods, Table 5-1 provides some examples of how different types of changes in health spending affect measured health prices using the indirect and direct methods. We distinguish between changes that were anticipated (and hence are reflected in the premium and do not affect retained earnings) and unanticipated (in which case they are not reflected in the premium and hence affect retained earnings.[14] We compare the treatments under the two methods to those that would be captured by a theoretically ideal CPI.

The indirect method handles anticipated changes in medical care prices and utilization (shown in the first two rows of Table 5-1) appropriately, and it also handles unanticipated changes in medical care prices (row 6) appropriately. However, in the case of an unanticipated increase in utilization (row 5), the indirect method registers a spurious decline, while in the cases of anticipated or unanticipated changes in treatment technologies that reduce the cost of treating a disease (rows 3 and 5), the indirect method *overstates* inflation. Note, however, that the distortions to the indirect method from unanticipated changes do not affect the behavior of the index over the long run, as expectations will eventually catch up with current conditions. Neither the direct nor indirect method will capture changes in consumer welfare arising from quality improvements that exceed their price (rows 4 and 8). To get this right would require a

---

[14]In these examples, we assume retained earnings are a constant share of anticipated medical spending.

very different approach to price measurement, similar to that used in Cutler et al. (2020), Dauda, Dunn, and Hall (2021), and Sheiner and Malinovskaya (2016).

TABLE 5-1 Impact of Changes in Health Spending on Measured Health Prices Using the Indirect and Direct Methods

Effects of Increases in Health Care Spending on "Total Health Insurance" Prices Depending on Source of Increase and Method						
Source of Spending Increase	Effect on Premium	Effect on Retained Earnings	Effect on Measured Prices of Medical Goods and Services	Effect on Health Insurance Price Direct Method	Effect on "Total Health Insurance Price" Indirect Method	Effect on Theoretically Ideal CPI
Anticipated increased utilization (more surgeries, more tests, etc.)	Higher	None	None	Higher Price	None	None
Anticipated higher prices of physicians and other providers	Higher	None	Higher	Higher Price	Higher Price	Higher Price
Anticipated cost-saving change in form of care (inpatient to outpatient, drugs instead of talk therapy, etc.)	Lower	None	None	Lower Price	None	Lower Price
Anticipated increased price and quality/intensity of care *within* a service that BLS prices like a doctor's visit, hip replacement, stent, etc.	Higher	None	Higher	Higher Price	Higher Price	Ambiguous[15]
Unanticipated increased utilization (more surgeries, more tests, etc.) e.g., a bad flu season	None	Lower	None	None	Lower Price	None
Unanticipated higher prices of physicians and other providers	None	Lower	Higher	None	None[16]	None
Unanticipated cost-saving change in form of care (inpatient to outpatient, drugs instead of talk therapy, etc.)	None	Higher	None	None	Higher Price	None
Unanticipated increased price and quality/intensity of care *within* a service—e.g., doctor's visit, hip replacement, stent—that BLS prices	None	Lower	Higher	None	Higher Price	Ambiguous[17]

Choice of Providers Available to the Insured

---

[15]If value of increased quality/intensity exceeds its price, Lower Price. If value is equal to price, No Change. If value is lower than price increase, Higher Price.

[16]When prices are unexpectedly higher, retained earnings are unexpectedly low, offsetting the price increase. This can be seen by noting that prices do not appear in the comparison between direct and indirect methods (see Appendix 5B).

[17]If increased quality is valuable, the effect is a lower price. If increased intensity is not valuable (ineffective treatment), the effect is a higher price.

Neither the direct method nor the indirect method as currently implemented by BLS fully accounts for the quality of health insurance services, which can include the provision of 24-hour hotlines, good customer service, and, most especially, the breadth of the provider network. (The indirect method will capture the difference in claim rejections by different insurers and, to some extent, out-of-network generosity as these will affect the share of the premium that is in retained earnings.)

A Suggested Alternative Indirect Method

As shown in Appendix 5B, the indirect method—as currently implemented by BLS—prices the cost of health insurance services per unit of health care. If the services provided by health insurance are considered as primarily related to claims processing and other administrative costs, this makes sense. The cost of insurance is like a markup on the cost of the health care.

Another approach would be to view the quantity of health insurance purchased as simply 1 per policy. Under that approach, the health insurance costs would simply be the ratio of retained earnings per policy, rather than retained earnings per unit of health care. The appeal of this alternative approach is that it recognizes that health insurance is valuable regardless of whether any health care is used at all: in other words, part of the policy reflects the financial protection it provides. In addition, having health insurance also allows a consumer access to the prices negotiated by insurance companies, which tend to be lower than those charged to uninsured consumers.

With this alternative methodology, it would still be the case that when utilization is unexpectedly low, health insurance would appear expensive, because retained earnings would be high. But the effect would be smaller than under the current implementation, when not only are retained earnings low, but retained earnings per unit of health care is particularly low.

Scope and Compositional Effects

The direct method prices particular health insurance policies for a fixed pool of insured people—that is, BLS asks the insurance company for the price of health insurance for a set of workers (50-year-old smoker, 30-year-old married woman, etc.) for a fixed policy with a fixed structure of copayment, deductible, and out-of-pocket maximums. Thus, changes in the demographic composition of the population or the types of plans offered will not result in changes in the cost of health insurance using the direct method.[18]

By contrast, as currently implemented, the indirect approach takes a unit value approach—it simply lumps together all health insurance policies and measures the ratio of retained earnings to benefits, regardless of whether these policies are for large employer-sponsored insurance, individual plans, plans sold on the ACA marketplace or in the small-group market, or Medicare Advantage Plans (in scope for services paid out of Medicare Part B, etc.). If different plans have different retained earnings ratios, then changes in the composition of plans will lead to changes in measured prices. For example, the ratio of retained earnings to premiums

[18]Some important changes are not well accounted for, however, including changes in the choice of health care providers who are in-network.

is higher for individual plans (around 20 percent) than for group plans (around 15 percent), so changes in the composition of the industry affect the growth rate of the health insurance CPI, which pools all plans (Cicala et al., 2019). To the extent different plans have different retained earnings ratios because they offer products with different levels of service (e.g., if individual plans are more expensive because, on average, they insure higher-risk people), the unit value approach may be biased.

Two additional problems are that the plans for which there are data are (1) not representative of the plans in scope for the CPI and (2) not weighted properly based on their weight in the CPI. The data used by BLS to calculate health insurance prices are based on insurance filings with state insurance commissioners. Self-insured employers (employers that pay for medical costs directly, even if they hire an insurance company to manage their health plan) are not subject to state insurance regulation and do not post information with insurance commissioners. However, they are in scope for the CPI, in that the premiums and out-of-pocket costs of workers with insurance from a self-insured employer are counted in the medical expenditure weight in the CPI. About two-thirds of workers with employer-sponsored insurance are in self-insured plans, so this is a significant omission (Kaiser Family Foundation, 2020a).[19] Because employers with self-insured plans pay the medical expenses of their employees directly, there is no way in which a survey can capture "retained earnings." However, because workers in self-insured plans are almost all working for large employers (because large employers are much more likely to self-insure and because they employ many more workers), retained earnings from large employers' commercial health insurance plans could be useful as a proxy.

An additional problem is that the CPI weights reflect the costs of health insurance plans covered by individuals' out-of-pocket costs—this includes the entire cost of the policy when it is purchased on the individual market, the after-subsidy cost to households who get their health insurance through the ACA exchanges, and the employee's share of the policy when it is purchased in the employer's market.[20] The current method simply lumps together all of the sold policies, regardless of their weight in the CPI. If retained earnings ratios vary systematically with the type of plan purchased, this will distort the CPI.

Timeliness

One disadvantage of the indirect method is that retained earnings are only measurable after the fact. That is, when an insurance company sells a policy, the company does not know what its retained earnings (premiums less benefits) will turn out to be. Furthermore, because of data constraints, the data for the ratio of retained earnings to benefits are calculated with a long lag. In particular, much of the retained earnings data come from the National Association of

[19]Companies that act as a servicer for self-employed plans are likely in the PPI, although their weight is likely to be very low because their revenues will not include any costs of medical care.

[20]An additional complication is the tax treatment of health insurance. Most workers who pay a part of their health insurance premium do so through a tax-preferred vehicle, so that the true out-of-pocket cost is lower than the cost of the premium. The CPI does not typically account for tax subsidies—for example, it does not lower the share of housing in the consumption basket because of its favorable tax treatment. But because people getting their health insurance through the ACA exchanges pay a "premium" that is really a market-based premium less a tax subsidy, the weight in the consumption basket will reflect spending after taxes, whereas the weight for premiums paid for employer-sponsored insurance will be a pre-tax weight, thus overweighting that component.

Insurance Commissioners (NAIC) Statistical Compilation of Annual Statement Information for Health Insurance Companies, which is usually published in October or November of the year following the year in reference. For example, aggregate retained earnings for the year 2020 would not be published by NAIC until October or November of 2021.

As noted above, BLS brings the changes in the measured price of health insurance gradually—by taking the 12th root of the annual change in the ratio of retained earnings to benefits.[21] Thus, it takes another year before the full cost of health insurance is included in the level of the CPI—making it a very lagged indicator. The majority of people with insurance have plans that start in January. For example, 74 percent of people with employer-provided insurance have plans that start in January, and premiums change for ACA and Medicare Advantage plans in January as well (Kaiser Family Foundation, 2020b). Thus, price increases occur in January for most people. However, because of the 9- to 10-month lag in the information on retained earnings, and the 12 months it takes for the prices to be fully incorporated, there is a 21- to 22-month gap between the actual price change and when it is fully reflected in the CPI.

Rebates

The ACA included a provision that limits the amount of premium income that insurers can keep. If the Medical Loss Ratio (ratio of benefit payments to premiums) is less than 80 or 85 percent (depending on the market) over a 3-year period, the insurance company has to rebate the difference to the policy holders in the third year. (For example, rebates paid in 2021 are calculated using insurers' financial data from 2018–2020; the policy holders in 2020 receive the rebates using that calculation.)

Rebates clearly lower the price of health insurance, and policy holders may expect rebates, so rebates may even lower the price of health insurance consumers perceive when they buy it. Thus, it seems that rebates should be deducted from the cost of the policy after the fact. The direct method used by the PPI does subtract rebates from the cost of the policy, but it subtracts them in the year they are paid, rather than the year in which they were "earned." (For example, rebates paid in 2021 are subtracted from 2021 health policy prices, rather than those in 2020.) Currently, the CPI does not include any rebates—the retained earnings used in the calculation are pre-rebate. The data on rebates used in the PPI should be used to adjust retained earnings in the CPI.

Falling Response Rates

As discussed above, survey response rates from health providers have been falling in recent years, making the prices on which the indirect method relies increasingly variable and possibly unrepresentative. An alternative source of data that might in the near future prove useful in countering this deterioration of information stems from recent federal efforts to improve transparency in the health care sector. One such initiative is the Hospital Price Transparency rule, which requires hospitals to display files on their websites containing payer-specific negotiated

[21] If the ratio of retained earnings to benefits increases 6% from one year to the next, the BLS assumes that prices rise roughly 0.5% per month—so that by the end of the year, the total increase in price will be 6%. But policy increases happen at the beginning of the policy year and all at once, so this method somewhat understates the level of prices.

charges and cash prices for at least 300 of their services. To date, compliance with this requirement has been extremely low.[22] However, in a bid to improve compliance, the Biden Administration sharply increased the penalty for noncompliance, with the higher penalties beginning on January 1, 2022.[23] The Health Plan Price Transparency Initiative similarly requires insurers to post prices for a wide variety of medical goods and services; that requirement begins on July 1, 2022. If successful, both of these initiatives could provide a readily available, low-cost source of price information for medical services.

In addition, BLS has been exploring the use of claims data in order to capture prices. That work, though preliminary, shows promise, especially for pricing physician services.

5.3. OPPORTUNITIES AND CHALLENGES

For the medical care component of the CPI, the most immediate methodological decision facing BLS is whether to continue pricing insurance using the indirect (retained earnings) method or begin migrating to a direct pricing approach. As discussed above, neither the direct nor the indirect method is conceptually ideal. The advantages of the direct method over the indirect method include the fact that the direct method reflects the price of insurance that consumers face at the time of purchase, that it is timelier, and that it will capture cost savings arising from shifts in the form of medical care—e.g., from an inpatient hospital to outpatient. However, a major downside of the direct method is that, without corrections, it will mislabel increases in utilization—either reflecting increased usage of existing medical treatments or of newly introduced treatments—as changes in prices rather than quantities; the indirect method will not count increased utilization as a price increase. Given that increased utilization has been an important source of increased health spending over time, this deficiency seems important enough to give the advantage to the indirect method.

In addition, if the ultimate goal is to fully quality adjust changes in health care over time, it would be impossible to do that without "getting under the hood" of the health insurance package and pricing (and quality adjusting) the underlying medical services separately from the health insurance services, which only the indirect method does.

> **Recommendation 5.1:** The indirect method has practical advantages and therefore should, in the short to medium run, continue to be the method for pricing health insurance in the CPI.

The above recommendation notwithstanding, the panel recognizes that falling response rates are making reliance on the indirect method increasingly difficult. Furthermore, the panel was unable to compare the differences between the indirect and direct methods empirically. Thus, this choice needs ongoing attention.

---

[22]Documentation of this point can be found here: https://healthcareexecintelligence.healthitanalytics.com/news/hospitals-fall-short-of-price-transparency-rule-compliance.

[23]https://www.cms.gov/newsroom/press-releases/cms-oppsasc-final-rule-increases-price-transparency-patient-safety-and-access-quality-care

> **Recommendation 5.2:** BLS should explore the historical differences between the indirect and direct methods doing a true apples-to-apples comparison. A "whole health insurance price deflator" that is a weighted average between the CPI's current health insurance deflator and the various deflators for the medical services financed by insurance should be calculated and compared to the deflator used in the direct approach. If this research reveals that the two approaches do not differ greatly historically (particularly after our recommended improvements to the indirect method, discussed below), BLS could revisit its reliance on the indirect method.

Initially, work on the direct pricing of health insurance would be conducted as a research series (as opposed to a published series). Researchers should have access to working papers on the topic.

> **Recommendation 5.3:** BLS should publish the whole health insurance price (the weighted average of the health insurance deflator and the medical services deflator)—either in addition to or perhaps instead of the insurance component separately. The volatility shown in the insurance services component of the total health insurance price index has been a source of concern for users.

We show in Appendix 5A that part of this volatility may reflect unanticipated price changes that do not affect the total health insurance price index but move the price of insurance services and medical services in opposite directions. Furthermore, there is a great deal of confusion about what the health insurance services component of the CPI is capturing, with many people believing it is the "whole health insurance premium"—like the PPI—instead of just the health insurance services component. This is particularly problematic given the differences between what is being priced in the PPI for health insurers and the CPI for health insurance. Publishing the whole health insurance premium index will provide useful information to users and will serve to clear up confusion among users.

> **Recommendation 5.4:** BLS should consider a number of potential improvements to the indirect approach. To better capture what the consumers actually pay for insurance, and which does not depend on utilization rates, BLS should explore using a multiyear rolling average of retained earnings per unit of health services (where retained earnings equal premiums, less medical expenses), rather than an annual value. This approach will mean that actual changes in the cost of health insurance—stemming from changes in regulation, market structure, or technology—will show up more slowly in the CPI. But these could be worth it to solve the problem of excess volatility.

Such a method is used by the Bureau of Economic Analysis (BEA) and indeed in the national accounts of most countries for property and casualty insurance. In particular, the price of these insurance services is approximated by a geometrically weighted moving average of past ratios of retained earnings to benefits. The weight on the most recent year is 0.3, the weight of the preceding year is (0.3)(0.7), and so on (Chen and Fixler, 2003). BLS should assess the effects of such a change on the volatility of both the pure insurance services component of the index and the whole health insurance index that includes the medical components.

Ideally, information on retained earnings should be obtained by type of health insurance plan, instead of using a unit value approach. To the extent that different types of plans have different retained earnings ratios because of the costs to the insurer of covering them, changes in the composition of the insured should not be captured as a price change.

> **Recommendation 5.5:** BLS should conduct research into pricing different plan types separately, and then weighting the price changes by the out-of-pocket payments for each. In addition, the current approach completely omits the price of health insurance provided by self-insured employers, even though costs borne by the employees of these firms is included in out-of-pocket spending. Because most self-insured plans are very large employers, BLS should impute prices for these plans using the retained earnings of large commercial plans.

Also, currently, there is a very long lag between the time the actual prices of health insurance change and their incorporation into the CPI. Shortening this lag would improve the accuracy and timeliness of the index.

> **Recommendation 5.6:** For the purpose of tracking changes in health insurance prices, BLS should consider migrating from using annual data on profits net of premiums to quarterly data. This would allow changes in retained earnings to be incorporated more quickly. The use of multiyear rolling averages (Recommendation 5.4) would help smooth out any noise in the quarterly patterns of retained earnings.

A first step in this work is to test the performance of estimates based on disaggregated financial statement data against those based on the lagged annual compendium published around October or November of the following year.

BLS should also consider changing the medical prices used in the health insurance services part of the health insurance deflator to match the timing of the retained earnings. That is, if the change in the ratio of retained earnings to benefits is lagged, the medical price changes used in the health insurance calculation should be equally lagged. (See the analysis in Appendix 5A.) In this way, the deflator would represent (lagged) health insurance services per unit of real medical expenditures. Alternatively, BLS could consider changing its conception of health insurance services as simply the price for the insurance component of the health care policy. That would just be the changes in retained earnings per policy over time, rather than dividing by the quantity of health services. This would bridge some of the gap between the direct method and the indirect method.

BLS should also account for the effects of rebates on the price of health insurance. The ACA requires that plans provide rebates to beneficiaries if their average medical loss ratio (basically the ratio of medical benefits to premiums) over a 3-year period is less than 80 or 85 percent, depending on the plan. The rebate is provided to people who were policy holders in the third year of that 3-year average. Under the current methodology, that rebate is ignored in the CPI.

> **Recommendation 5.7:** In estimating the price of health insurance, the BLS should subtract rebate payments from retained earnings.

Even the direct method, which is an ex-ante measure, includes rebates, even though they change the price of health insurance ex-post. But because the average rebate is positive (the rebate is one-sided—it is either zero or positive), consumers may expect rebates when they purchase insurance and so subtracting rebates may be reasonable even with an ex-ante perspective. And because the indirect method is already an ex-post measure of health insurance costs, it seems clear that the measure of retained earnings used in the calculation should be after subtracting rebates. BEA has bought information on rebates that should be accessible by BLS as well.

> **Recommendation 5.8:** As the data ecosystem continues to evolve, research by BLS should continue evaluating how to accelerate incorporation of claims data, hospital data, health plan data, and scanner data on drugs to improve the coverage, detail, and timeliness of price and quantity information in the medical care component of the CPI. Given the large cost in both time and money associated with using claims data, BLS should combine the efforts of those producing PPIs with those producing CPIs. In addition, there should be collaboration between the CPI and PPI on adjusting health care prices for quality change.

A pilot program is currently underway at BLS indicating promising uses of claims data. A first study from the program used data from a single insurance company in a single market (Bieler et al., 2019).[24] Building from this work, BLS is expanding the pilot study using a national claims database covering all CPI areas. The research includes multiple insurance companies, and the results look promising for pricing physician services and outpatient hospitals. The next steps will involve evaluating the possibility of using these claims data as a replacement for manually collected private insurance quotes, determined at the regional level. This will likely involve developing and refining a methodology for blending the new and traditional data.

Indeed, even though BLS research has shown that claims data are not representative, the agency has been revisiting use of this source in large part because of declining respondent cooperation. The pilot study was created in part to test blending the new data with the survey-collected data that did not need to rely on a participant company.

> **Recommendation 5.9:** BLS should investigate the use of prices that hospitals and insurers post on their websites to comply with transparency requirements. If hospital compliance improves, and if the insurer initiative is successful, these could be useful sources of data on prices of medical goods and services.

Another motivation for acquiring and developing agility using claims data is that these data are crucial for constructing disease-based price indexes that can eventually be used to quality adjust health care prices. As noted in much of the literature on health care quality, the only way to properly account for the changing quality of health care is by tracking outcomes, and a first necessary step in such an endeavor is to track nominal spending by disease, which can then be deflated by an index that accounts for changes in outcomes from treatments for that disease over

[24]Bieler et al. (2019) reported on BLS work testing the potential of insurance claims data to supplement manual price collection in the CPI medical indexes. The authors constructed price indexes using “data purchased from an insurance company for a large city and compare them to the CPI medical indexes for that city” with the aim of assessing their feasibility for use on a larger scale.

time. Of course, accounting for changes in outcomes over time is fraught with conceptual and data challenges. But there have been some major advances in recent years (e.g., Cutler et al., 2020; Dauda, Dunn, and Hall, 2021) and it is possible that these types of quality adjustments might be feasible in the future.

BEA has led the way with development of satellite health accounts but, as discussed above, the measurement objective of the CPI (tracking prices faced by consumers) differs from that of the national accounts.[25] BLS also has undertaken important initiatives. BLS has created experimental disease-based prices that bundle medical care inputs by treatments for specific diseases or conditions.[26] In addition, BLS has produced experimental indexes at the International Classification of Diseases (ICD) chapter level with monthly updates since 2016. In 2020, BLS began producing experimental indexes at the subchapter level for common diagnoses (114 conditions). Conceptual issues aside, data demands make it unlikely these methods will be brought into the production of the CPI any time soon. The primary data source for the BEA and BLS disease-based price indexes is the Medical Expenditure Panel Survey, which is not timely and has a relatively small sample size particularly for calculating price indexes for less common medical conditions. To address these deficiencies, BLS must acquire and research integration of claims data for this use.

---

[25]For more information on satellite accounts, see https://www.bea.gov/news/blog/2015-01-22/introducing-new-bea-health-care-satellite-account.

[26]Although BLS calls these experimental indexes "prices," they are better understood as a measure of spending by disease that shifts over time because of higher prices and changing quality.

APPENDIX 5A: COMPARISON OF INDIRECT AND DIRECT METHODS OF PRICING HEALTH INSURANCE

Indirect Method Conceptual Framework

The indirect method decomposes the health insurance premium into two types of goods and services: the services provided by health care providers, and the services provided by the insurance company, which include risk bearing, cost management, price negotiation, claims processing, utilization review, etc.

$$\text{HI Premium} = \text{HI Services} + \text{Medical Benefits} \qquad \text{(eqn. 5.1)}$$

Each of these pieces, in turn, has a price component (P) and a quantity component (Q):

$$\text{HI Premium} = P_{HI}Q_{HI} + P_M Q_M, \qquad (5.2)$$

where HI denotes the services provided by health insurers, and M denoted medical services. Implicitly, BLS's methodology assumes that the quantity of health insurance services is equal to the quantity of medical services:

$$Q_{HI} = Q_M \qquad (5.3)$$

Therefore,

$$\text{HI Premium} = (P_{HI} + P_M)Q_M. \qquad (5.4)$$

The interpretation of this formulation is as follows: For every medical service someone purchases through health insurance, they pay a price for the health insurance services provided and a price for the medical services.[27]

From above, HI Services are equal to the HI Premium, *Prem*, less the Medical Benefits. This is the definition of retained earnings (*RE*)—the part of the premium that the health insurance company does not pay out in medical benefits.

Next, call the Medical Benefits *B*, where

$$B = P_M Q_M \qquad (5.5)$$

Then,

$$\text{HI Services} = P_{HI}Q_{HI} = P_{HI}Q_M = Prem - B = RE \qquad (5.6)$$

And:

$$P_{HI} = \frac{RE}{Q_M} \qquad (5.7)$$

The real quantity of medical benefits, Q_M, is calculated by deflating spending on Medical Benefits, B, by the medical price deflator:

$$Q_M = \frac{\mathrm{B}}{P_M} \qquad (5.8)$$

Putting it all together yields the following:

[27]As we discuss in the text, the assumption that the quantity of health insurance services is equal to the quantity of medical services might be reasonable for some types of services (e.g., claims processing), but may not appropriately capture the value of the financial risk protection associated with health insurance (e.g., health insurance provides valuable financial protection even if one never purchases any medical services, just as fire insurance is valuable even if a consumer does not experience a fire).

$$P_{HI} = \left[\frac{RE}{B}\right] P_M. \tag{5.9}$$

The price relative for health insurance would therefore be:

$$\left[\frac{P_{HI,t}}{P_{HI,t-1}}\right] = \left[\frac{\frac{RE_t}{B_t}}{\frac{RE_{t-1}}{B_{t-1}}}\right] * \left[\frac{P_{M,t}}{P_{M,t-1}}\right] \tag{5.10}$$

This price relative is only for the component of health insurance that reflects services provided by the health insurance company.

To compare the indirect method to the direct method, it is necessary to calculate the implied price index for the whole health insurance premium under the indirect method. Call the ratio of retained earnings to the health insurance premium z, where $z = \frac{RE}{Prem}$. Then the price relative for the whole health insurance premium would be[28]:

$$\left[\frac{P_{M,t}}{P_{M,t-1}}\right] * \left[z_{t-1}\left[\frac{\frac{RE_t}{B_t}}{\frac{RE_{t-1}}{B_{t-1}}}\right] + (1 - z_{t-1})\right] \tag{5.11}$$

This expression can be rewritten as follows:[29]

Price Relative for Whole Health Insurance Policy using Indirect Method = $\frac{\frac{Prem_{,t}}{Prem_{,t-1}}}{\frac{Q_{M,t}}{Q_{M,t-1}}}$.

That is, under the indirect method, the increase in the total cost of health insurance is equal to the increase in the premium cost **per unit** of health services.

Indirect Method Implementation

The HI Index in the monthly CPI is calculated as follows:

---

[28]As a Laspeyres aggregation of the component relatives, this expression is an approximation of true premium inflation. The aggregation weights (share of retained earnings and the share of benefits spent on each type of medical service) are fixed at t-1 values so, if these weights change over the period, this approximation will differ from true premium inflation.

[29]The algebra is as follows:

$$\left[\frac{P_{M,t}}{P_{M,t-1}}\right] * \left[z_{t-1}\left[\frac{\frac{RE_t}{B_t}}{\frac{RE_{t-1}}{B_{t-1}}}\right] + (1 - z_{t-1})\right]$$

$$= \left[\frac{P_{M,t}}{P_{M,t-1}}\right] * \left[\frac{RE_{t-1}}{Prem_{t-1}}\left[\frac{\frac{RE_t}{P_{M,t}Q_{M,t}}}{\frac{RE_{t-1}}{P_{M,t-1}Q_{M,t-1}}}\right] + \left(\frac{B_{t-1}}{Prem_{t-1}}\right)\right]$$

$$= \frac{P_{M,t}}{P_{M,t-1}}\frac{RE_{t-1}}{Prem_{t-1}}\frac{RE_t}{P_{M,t}Q_{M,t}}\frac{P_{M,t-1}Q_{M,t-1}}{RE_{t-1}} + \frac{P_{M,t}}{P_{M,t-1}}\frac{P_{M,t-1}Q_{M,t-1}}{Prem_{t-1}}$$

$$= \frac{RE_tQ_{M,t-1}}{Prem_{t-1}Q_{M,t}} + \frac{P_{M,t}Q_{M,t}Q_{M,t-1}}{Prem_{t-1}Q_{M,t}} = \frac{Q_{M,t-1}}{Prem_{t-1}Q_{M,t}}[RE_t + B_t] = \frac{\frac{Prem_t}{Prem_{t-1}}}{\frac{Q_{M,t}}{Q_{M,t-1}}}$$

$$I_{t,t-1} = \left[\frac{\frac{RE_y}{B_y}}{\frac{RE_{y-1}}{B_{y-1}}}\right]^{1/12} * \left[\frac{\sum_i w^i MCPI^i_{t,b}}{\sum_i w^i MCPI^i_{t-1,b}}\right] \qquad (5.12)$$

Where $MCPI^i_{t,b}$ is the medical care price index for medical service i in month t, w^i is the share of insurance payments on service *i*, b is the base period and y is the year for which retained earnings are measured.

The second term is just the price relative for medical services from above, $\frac{P_{M,t}}{P_{M,t-1}}$. Thus, the only difference in the implementation of the framework from equation (5.1) is that retained earnings are measured annually, so the 12th root is taken, and because of data constraints, the data for the ratio of retained earnings to benefits are calculated with a lag. In particular, much of the retained earnings data come from the National Association of Insurance Commissioners Statistical Compilation of Annual Statement Information for Health Insurance Companies, which tends to be published in October or November of the year following the year in reference. For example, aggregate retained earnings for 2020 would not be published by NAIC until October or November of 2021.

As we note in the text, this makes the interpretation of the HI index somewhat more difficult, because it is no longer exactly retained earnings per quantity of benefits (since the benefits from period y are being deflated by prices from month t).

Comparison with Direct Method

The price index for the direct method for a given type of insurance policy is:

$$\text{Direct Price Relative} = \frac{\text{Prem}_t}{\text{Prem}_{t-1}} \qquad (5.13)$$[30]

Calling the indirect price relative for the whole health insurance policy *Ind* and the direct price relative *D*—and abstracting from the fact that the indirect method does not measure the change in retained earnings on a policy-by-policy basis, and instead uses a unit cost method—one can see that:

$$Ind = \frac{D}{\frac{Q_{M,t}}{Q_{M,t-1}}} \qquad (5.14)$$

Apart from the differences in timing due to data constraints discussed above, the indirect method price relative is equal to the direct method price relative divided by the growth rate of real medical services provided through that policy. If the price of a health insurance policy increases 10 percent, and utilization also increases 10 percent, the direct method would show a price increase of 10 percent, whereas the indirect method would show no price increase at all.

Considering Uncertainty

[30]Ideally, the direct method would hold expected utilization constant, so that the indirect and direct method would be conceptually quite similar. And, as noted earlier, BLS makes some effort to adjust the premium when large changes in technology are introduced but doesn't adjust for the small continuous changes that occur in health care. We ignore the BLS quality adjustments for large technological changes in this exposition.

As discussed in the text above, the two methods will produce different results depending on whether changes in prices and quantities are expected or unexpected, because that will determine whether they are incorporated by insurance companies when setting premiums.

Expected price changes: Expected price changes for medical care services increase the premium, thus boosting both direct and indirect method prices.

Unexpected price changes: Unexpected price changes for medical care services typically do not affect premiums or utilization very much.[31] Thus, they have no effect on either the indirect or direct method prices for the whole insurance policy. Note, however, that unexpected increases in price will lower the health insurance component of the whole health insurance policy using the indirect method (because they lower retained earnings), but that will be offset by an increase in the medical care component. In practice, this may not occur right away due to the timing mismatch between the retained earnings and price adjustments. Moreover, changes in one direction in a year may be offset the following year by change in the opposite direction if health insurers respond to the prior period's unexpected change in retained earnings. These unexpected price changes may be a reason that the health insurance component appears so volatile while the whole health insurance policy deflator is not.

Expected increases in utilization: Expected increases in utilization, unless captured in the quality adjustment, will boost premiums and, in turn, the deflator used in the direct pricing method. However, because the indirect captures changes in the premium per unit of real health services, expected increases in utilization will not affect the whole health insurance policy deflator using the indirect method.

Unexpected increases in utilization: Unexpected increases in utilization will not affect premiums, and so will have no effect on the deflator using the direct method. However, because premiums do not increase but utilization does, this will appear as a reduction in indirect method prices.

APPENDIX 5B: AN ALTERNATIVE FORMULATION OF THE INDIRECT METHOD

As noted in Appendix 5A, the current indirect method assesses the price of health insurance as the costs per unit of real health goods and services. That is, spending on health insurance services = $P_{HI}Q_M$, where P_{HI} is the price of health insurance services and Q_M is equal to the real quantity of medical goods and services paid for by a policy. This assumption that the quantity of health insurance services is equal to the quantity of medical services might be reasonable for some types of services (e.g., claims processing), but may not appropriately capture the value of the financial risk protection associated with health insurance (e.g., health

[31]Small effects could be observed since, although patients may not have much exposure to the full price, they have some and demand is not perfectly inelastic.

insurance provides valuable financial protection even if a consumer never purchases any medical services, just as fire insurance is valuable even if the consumer does not have a fire).

Thus, an alternative is to consider the quantity of health insurance services as equal to 1 per policy. Then, when using the indirect method, the change in the price of health insurance services would simply be the change in retained earnings per policy.

In particular, using the notation from Appendix 5A, this would then make the Alternative Price Relative for Whole Health Insurance Policy using the Indirect Method equal to:

$$z_{t-1}\left[\frac{RE_t}{RE_{t-1}}\right] + (1 - z_{t-1})\left[\frac{P_{M,t}}{P_{M,t-1}}\right] \qquad \text{(eqn 5.1)}$$

Or, alternatively, as:

$$z_{t-1}\left[\frac{RE_t}{RE_{t-1}}\right] + (1 - z_{t-1})\frac{\frac{B_{,t}}{B_{,t-1}}}{\frac{Q_{M,t}}{Q_{M,t-1}}}. \qquad (5.2)$$

Conceptually this is simple: If 20 percent of a health insurance policy pays for health insurance services, and 80 percent for medical services, then the deflator would put a 20 percent weight on the percent change in retained earnings per policy and an 80 percent weight on the change in the prices of medical services.

This formulation would partially bridge the gap between the indirect and the direct methods, as can be seen by comparing equation (5.2) with the formulation of the direct method:

$$\text{Direct method:} \quad z_{t-1}\left[\frac{RE_t}{RE_{t-1}}\right] + (1 - z_{t-1})\frac{B_{,t}}{B_{,t-1}}. \qquad (5.3)$$

The comparison between the two equations shows that under this alternative formulation, the pricing of the insurance component of the indirect method and the insurance component implicit in the direct method would be the same—they would both measure the changes in the cost of insurance services to each policyholder.

6
Supplemental Subgroup Price Indexes

New data sources present opportunities to improve the accuracy and timeliness of both the elementary item-area price indexes (Chapter 2) and of the higher index aggregation levels (Chapter 3) in construction of the Consumer Price Index (CPI). One related research and policy need to which the Bureau of Labor Statistics (BLS) has responded over the years is to produce price indexes (as well as other economic statistics) tailored to measuring trends for specific population subgroups. Both conventional data sources, such as micro data from expenditure and budget surveys, and new data sources, such as detailed transaction data that can be linked to shoppers' characteristics, create new opportunities to study price changes faced by different population groups.

6.1. MOTIVATION

The rationale for producing price indexes for population subgroups is clear for purposes such as adjusting Social Security benefits (which are mainly received by older people), setting marginal tax rates (which increase with income level), and establishing consumer unit needs and resource levels that regulate transfer payments of various safety net programs (for which only certain groups are eligible). Broader public policy questions related to income and wealth inequality, social welfare, and poverty could also be informed by more precise measures of differential inflation rates faced by specific groups, such as lower-income households. Of course, the motivation to marshal the resources needed to produce subgroup indexes is more powerful if it can be established that the rates of inflation experienced by different groups or by people in different geographic locations vary significantly. The assessment of evidence by this panel, as well as other experts (including a National Academies' panel, see NRC, 2002, p. 222), is that, at least during some time periods, considerable heterogeneity does exist in the purchasing patterns and shopping behavior of, as well as the prices paid by, consumers with different demographic characteristics.[1]

---

[1]NRC (2002), Chapter 8, provides a comprehensive examination of the conceptual basis of population subgroup indexes as well as practical data issues that complicate implementation.

6.2. RESEARCH FINDINGS

One factor that can lead to differential inflation rates—and, as it turns out, the easiest one to measure—is that different groups of people tend to purchase different baskets of goods and services. In other words, people allocate their consumption budgets differently across CPI item categories in a way that correlates with income and other demographic characteristics. This heterogeneity occurs at an individual consumer unit level, and may also be associated with observed group patterns:

> Some [expenditure patterns] are idiosyncratic among individuals—vegetarians and meat eaters, book lovers and sports enthusiasts, travelers and homebodies. But many of the differences are systematically related to the economic, demographic, and locational characteristics of households. Lower-income households spend, on average, a higher fraction of their income on food and clothing than do higher-income households and a smaller fraction on travel and entertainment. The elderly tend to devote a smaller fraction of their budgets to durable goods and clothing and a larger fraction to travel and medical care than do non-elderly households. People who live in the South spend less on heating fuel and more on air conditioning than those in the North (NRC, 2002, p. 223).

Variation in purchasing patterns, when measurable, leads to a unique set of weights for each identified segment of the population. The mechanics of the calculation are readily illustrated by the several versions of the CPI already published. BLS currently produces official price indexes for two population subgroups—all urban consumers (CPI-U) and urban wage-earners and clerical workers (CPI-W; see Box 6-1). Additionally, on an experimental basis, BLS publishes a price index covering urban consumers aged 62 and older (CPI-E). Although the CPI-E is not currently used for indexing Social Security benefits, BLS has long been interested in a price index that could be used for that purpose; development of the CPI-E was motivated in part by that line of research.[2] These three versions of the CPI differ only in terms of the expenditure weights used to aggregate the component indexes. For example, the 2017–2018 expenditure weight for medical care is considerably higher for the CPI-E (12.20) than it is for the CPI-U (7.29); conversely, the weight for transportation in the CPI-E is lower (12.97) than it is in the CPI-U (15.16). While the weights differ across the CPI-U, CPI-W, and CPI-E, these indexes are constructed using the same set of price changes for each item strata from the same sample of urban areas.

---

[2]Currently, adjustments to Social Security benefits are still based on percent changes in the CPI-W (Cage, Klick, and Johnson, 2018). Arguments against adopting the CPI-E for the purpose are that it is associated with a higher sampling error than the alternatives and fails to address upper-level substitution bias.

BOX 6-1
Role and Future of the CPI-W

The CPI-W was introduced in 1919 to measure price changes faced by working-class Americans (Rippy, 2014). BLS collected data on expenditure shares from moderate-income families of which a (white male) wage earner or clerical worker contributed to the majority of the family's income. The CPI-W has retained this focus on clerical and wage-paying jobs ever since and now covers women and all people regardless of race or ethnicity if they work in the relevant occupations. The CPI-W was the only price index published until 1978 when a broader index—the CPI-U—was introduced to cover all urban consumers. Up until this point, the CPI-W was the only national CPI available and so it was used from the inception of Social Security to index benefits. The Social Security Administration continues to use the CPI-W for this purpose, and the index also is used for cost-of-living adjustments for federal retirement programs.[3] Since 1981, the CPI-W has been constructed using the same sample of geographic areas, outlets, items, and prices as those used for the CPI-U but then reweighted to reflect the expenditure patterns of the relevant subset of Consumer Expenditure Survey (CEX) households.[4]

The CEX respondents used to construct the CPI-W now represent a small and shrinking segment of the population that has become less representative of middle-class Americans (about 29 percent or the CPI-U population).[5] In part because of labor market shifts away from the blue-collar occupations (such as clerical and sales positions) that are surveyed, the current definition used for the CPI-W is becoming less and less representative and relevant. Longer term, a more useful approach to measuring cost-of-living increases experienced by the working-class population could involve constructing an index using expenditure weights for consumer units identified as belonging to a subset of quintiles of the income distribution. As described in this chapter, characterizing price inflation faced by modern working households requires an index that goes beyond a simple reweighting for a subset of households, and also measures price changes faced specifically by this group. Focusing on a range of demographic characteristics that can be segmented in different ways would provide the greatest flexibility for producing a range of subgroup indexes.

Price measurement research using the simple reweighting approach (e.g., Amble and Stewart, 1994; Garner et al., 1996) has tended to detect only minimal differences in inflation rates faced by different groups. Comparison of the CPI-E and CPI-U (and the CPI-W, for that matter) offers a case in point that simple reweighting typically leads only to minimal differences in index performance. As shown in Figure 6-1, a comparison of average 12-month percent changes in the CPI-U and CPI-E reveals a difference that averages only 0.16 percentage point over the entire period (the CPI-E tends to have slightly larger increases).

[3]Further information on the specification and history of the CPI-W can be found at: www.bls.gov/opub/btn/volume-3/why-does-bls-provide-both-the-cpi-w-and-cpi-u.htm.

[4]From 1978 to 1980, BLS conducted an independent but overlapping sampling of items and outlets for both the CPI-U and the CPI-W populations. See U.S. Bureau of Labor Statistics, "Why Does BLS Provide Both the CPI-W and CPI-U?" (Reed and Stewart, 2014).

[5]www.bls.gov/news.release/pdf/cpi.pdf. Expenditure weights for the CPI-U are based on about 65,000 household interviews; weights for the CPI-W are based on a subset of about 16,000 interviews. This number has been declining along with CEX response rates generally and with the shrinking number of respondents engaged in the occupations captured in the CPI-W. A report on retirement security (GAO, 2020) states that BLS was investigating the merits of expanding the CPI-W to include all labor force participants, which could present a complement to a price index meant to represent the lower to middle group of income quintiles or deciles.

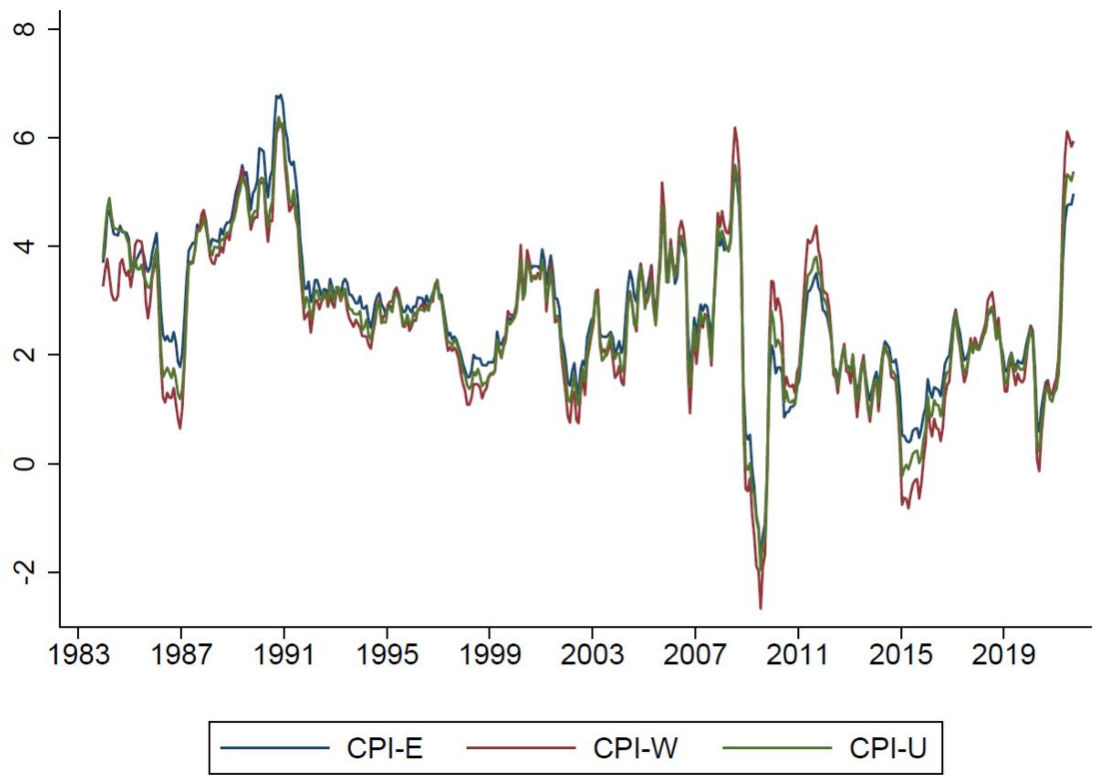

FIGURE 6-1 Twelve-month percent changes in the CPI-U, CPI-E, and CPI-W.
SOURCE: Panel-generated using BLS data.

Statistical agencies in other countries also regularly produce price indexes based primarily on reweighting to reflect subgroup expenditure patterns. For example, the Office for National Statistics in the United Kingdom publishes "Household Cost Indices," developed to measure the change in household costs over time for different population subgroups. These indexes have indicated, for example, that "retired households have experienced higher costs growth than non-retired households since May 2017," mainly due to different weights for electricity and council taxes.[6] The Household Cost Indices are aggregated using democratic weighting, wherein all households are assigned an equal weight, instead of plutocratic weighting, wherein households are implicitly represented in proportion to their total spending levels.[7] Most nations' headline CPIs, including those in the United Kingdom and United States, are plutocratic, meaning that they reflect the consumption patterns of upper-income households more closely than those of lower-income households.

Researchers outside statistical agencies also have developed subgroup indexes that use different patterns of expenditure weights across groups. For example, using microdata from

[6]www.ons.gov.uk/economy/inflationandpriceindices/bulletins/householdcostsindices/thirdpreliminaryestimates2005to2019.

[7]NRC (2002), Chapter 8 on approaches to aggregating across households, provides a complete discussion of the implications (and appropriate uses) of a democratic index in which individual price indexes are estimated for a representative sample of the whole population and then averaged by assigning the same weight to each consumer unit regardless of the magnitude of their total consumption expenditures.

Eurostat's European Household Budget Surveys and the Harmonized Index of Consumer Prices, Gürer and Weichenrieder (2018) found that for the period 2001–2015, the "consumption bundles of the poorest deciles in 25 European countries have, on average, become 11.2 percentage points more expensive than those of the richest deciles" (p. 1). Broadly speaking, the researchers found that price increases had been more rapid for necessities—e.g., food, shelter, and utilities—which constitute a higher expenditure share for lower-income households than for luxury goods and services having to do with, for example, recreation and culture, or car purchases.

Academic researchers have also used microdata on expenditures to calculate U.S. household-level inflation rates. McGranahan and Paulson (2006), for example, used Consumer CEX data and item-specific CPI data to construct the chain-weighted "Chicago Fed Income Based Economic Index" for a variety of different demographic groups. The authors found that, for the period 1983–2005, the inflation experiences of the different groups were "highly correlated with and similar in magnitude to the inflation experiences of the overall urban population" (McGranahan and Paulson, 2006, p. 26). The exception to this pattern found by the authors were the elderly, who faced an 11 percentage point (or 5.5 percent) higher cumulative inflation rate compared with the average over the period of the study.

On the other hand, analyzing the period 1984-2004, Hobijn and Lagakos concluded that the distribution of inflation experiences across households exhibited a large amount of dispersion. Additionally, they found that a democratic index (one that weights price changes faced by each household equally) constructed for the latter part of the period was higher than the plutocratic index (one that weights price changes according to each household's share of aggregate expenditures), suggesting that poorer households (with greater representation in the former) experienced higher inflation than richer households. More rapid price growth for gasoline and food prices contributed to the trend. Michael (1979), Hobijn and Lagakos (2005), and Hobijn and Sahin (2009), applying household-specific consumption bundles to estimate indexes of average prices for relatively broad categories of goods (mainly at the CPI item strata level), reached similar conclusions. These analyses continued to assume that all households pay the same price for specific goods purchased and buy the same mix of goods within each item stratum (Kaplan and Schulhofer-Wohl, 2017).

Their value to research notwithstanding, the deficiency in the alternative CPIs estimated in the way described above is that they do not account for the multiple factors affecting prices paid by different groups. Ideally, for a comprehensive measure of price inflation, any differences in the prices paid by different groups—particularly for big-ticket items that could really make a difference, such as medical care and housing/shelter—would be taken into account along with differences in spending patterns. Housing, discussed at length in Chapter 4, is a particularly important case of inflation differentials faced by subgroups defined by geography since prices (and, at times, changes in prices) vary a great deal from one part of the country to another and between urban and rural areas.

Recent research, some of which is described below, has been based on more diverse data sources that allow factors beyond expenditure shares to be considered. This research has revealed clear patterns of differential price inflation, in particular across income groups. It strongly suggests that if statistical agencies are serious about tackling differential price inflation, they must move beyond the exercise of simply reweighting price quotes from the official CPI. To create meaningful price indexes for demographic subgroups, it will be necessary to combine, for each consumer unit, monthly information on the prices it has paid, the amount expended on each item, and its basic demographic characteristics.

As a first step toward building the capacity to fully portray differential inflation, household- or group-specific inflation rates should reflect the fact that, in addition to buying a different mix of goods and services, households purchase goods and services from different outlets and, therefore, face different prices. Even if all consumer units patronize the same outlets (for example, Amazon during the pandemic) such that prices paid by different groups for the exact same product converge, the rich and poor, old and young, or other different demographic groups will continue to buy different products *within* the elementary index level. For example, high-income and low-income consumer units may tend to frequent different restaurants or hotels. Comprehensive subgroup inflation measures would therefore account for different consumption patterns at a very detailed level. They would also take into account "effective prices paid, spending shares on new and existing goods, and demand elasticities—all of which may vary across households, and in particular along the income distribution" (Jaravel, 2019, p. 7).

Creating a data collection apparatus along the lines implied above is far easier and more cheaply said than done. The key challenge in this proposition, with which BLS is intimately familiar, is that the data system currently underlying the CPI collects prices paid but does not link those prices to specific households. Put another way, information about consumers and how they budget their income is collected from a household survey, while price information is collected predominantly from retail outlets; thus, characteristics of purchasers cannot be linked with the prices they pay. Since retailers generally do not typically have full demographic information about their customers (although this is changing rapidly), the most direct route to a data source that combines these key pieces of information would involve "collecting the monthly price data, as well as expenditure patterns and demographic information, directly from consumers" (NRC, 2002, p. 229).

As described in Chapter 3, household-based data on purchases and prices paid for many items do exist in various commercial data sources that have been used in price measurement research. Research by Kaplan and Schulhofer-Wohl (2017), for example, used data from the Kilts-Nielsen Consumer Panel (KNCP) to estimate differential inflation rates, most prominently by income, at the level of the household.[8] The authors found that

> Households with low incomes, more household members, or older household heads experience higher inflation on average, whereas those in the Midwest and West experience lower inflation [and that] over the nine years from the third quarter of 2004 through the third quarter of 2013, average inflation cumulates to 33 percent for households with incomes below $20,000 but to just 25 percent for households with incomes above $100,000 (p. 3).

[8]The project is a partnership between The University of Chicago Booth School of Business and the Nielsen Company in which marketing datasets are made available to academic researchers. The dataset "records the prices, quantities and specific goods purchased in 500 million transactions by about 50,000 U.S. households from 2004 through 2013." https://www.chicagobooth.edu/research/kilts/datasets/nielsenIQ-nielsen.

Crucially, Kaplan and Schulhofer-Wohl found that the greatest source of heterogeneity in households' inflation rates for the goods they track is variation in prices paid for the same types of goods—not from variation in broadly defined consumption bundles of the type that would be picked up in indexes based only on expenditure weight adjustments. Roughly two-thirds of the variation detected by the authors originated from differences in prices paid for identical goods while only about one-third was found to come from differences in the mix of goods within broad categories; and "only 7 percent of the variation arises from differences in consumption bundles defined by broad categories" (p. 2).

Employing a somewhat different approach, while using the same KNCP data, Jaravel (2019) also estimated inflation rates in the United States as a function of income. For this research, Jaravel assumed that all households within a given range of income (such as all households earning more than $100,000 per year) have the same consumption bundle and pay the same price for each good. Covering the period 2004–2015, Jaravel measured inflation inequality using a linked dataset in which spending shares are based on the CEX and price changes (at the level of product categories) for most goods and services are based on the CPI data series. The matched CEX-CPI dataset provides 256 detailed product categories. However, for consumer goods observable in the KNCP data—including food products, household supplies, and health and beauty products—he incorporated product-level data on both prices and quantities; these goods "account for about 30-40% of expenditure on goods, or about 15% of total expenditures" (Jaravel, 2020, p. 8).[9] A key finding was that annual inflation was approximately 0.65 percentage point lower for households earning above $100,000 a year when compared with households making $30,000 or less per year. The headline finding of Wimer, Collyer, and Jaravel (2019)—who used the Jaravel (2020) estimates based on the linked CEX-CPI sample and Nielsen data to re-estimate recent trends in poverty and income inequality—was that, because of unmeasured inflation inequality, income inequality and poverty rates may be significantly underestimated for the United States over the period 2004 to 2018. For example, using his series for the lowest income quintile, 3 million additional people are found to have been living below the poverty line in 2018 compared with official numbers based on official CPI estimates.

The major limitation of scanner sources such as the KNCP data is that the scope of coverage is confined to goods sold in retail outlets (Kilts Center, 2013a). Unlike research based on the household expenditure surveys—such as noted above by Hobijn and Lagakos (2005)—the scanner-based research is unable to measure the impact on subgroup price inflation of quantitatively important expenditure categories such as medical care and housing "which have all been found to be important sources of inequality in inflation rates" (Kaplan and Schulhofer-Wohl, 2017, p. 3). As detailed in Chapter 4, shelter alone accounts for more than 30 percent of the household spending basket in the Consumer Price Index.

[9]Jaravel identified several key benefits to data that include barcodes. For example, he noted that "meaningful quality change" typically prompts a change in UPCs; additionally, discontinued UPCs (products) can be readily identified. Jaravel's research on differential inflation faced by low- and high-income consumers in the UK uses data collected by the market research firm Kantar FMCG Purchase Panel, which is similar to Homescan in the U.S., in which participants record UPCs for their purchase using a hand-held scanner.

Due to this limitation of products covered by scanner data, researchers have turned to other sources, including surveys not traditionally used in the CPI. Larsen and Molloy (2021) examined differences in quality-adjusted rent growth for households at different points in the income distribution for the period 1985 to 2019. The authors estimated rent changes facing households in different income groups using data from the American Housing Survey (ACS)—a nationally representative panel that tracks housing units over time—on individual housing units, grouped by income of the residents. Their analysis accounts for differences in the price of housing services across groups as well as differences in the fraction of expenditures spent on housing. They did not detect large differences in inflation across income groups over their sample period. This result contrasts somewhat with the results in Moretti (2013), who calculated differences in shelter inflation rates by education group using information on variation across cities in rent growth and variation in location choices by education groups; he found higher shelter inflation for college-educated people over the period 1980 to 2000. However, after adjusting for changes in amenities, Diamond (2016) found lower shelter inflation for the college-educated population.

For medical care (discussed in Chapter 4), another area where prices faced, quantities purchased, and quality received are likely to differ by income group, insurance claims datasets provide a rich source of information. Wimer, Collyer, and Jaravel (2019) documented inflation inequality in health care using comprehensive claims data along with linked employer-employee datasets for Utah between 2012 and 2015, finding that "there was higher inflation for treating conditions that affect low-income groups more, another source of inflation inequality."[10]

BLS and other statistical agencies will also need to develop plans for obtaining price and quantity information needed for improving the measurement of inflation in other areas where electronic transaction data are not typically available. The service sectors represent a sizable part of the economy—and one where higher- and lower-income households often display very different consumer behavior—that currently is not well covered by electronic data sources.

6.3. OPPORTUNITIES AND NEXT STEPS

An empirical consensus is emerging that, at least during some periods, price inflation has varied across population subgroups as well as across locations. Although still nascent, research cited in the previous section convincingly makes the case that high- and low-income groups in particular likely experience differential inflation rates for at least some types of goods and services, and measuring differences in prices paid for the same or similar goods is an important part of the story. Among the possible explanations for this finding, all requiring further investigation, are liquidity constraints inhibiting low-income households from taking advantage of sales and bulk discounts (Orhun and Palazzolo, 2019); greater flexibility of high-income households to substitute toward alternative goods or outlets (Argente and Lee, 2020); and more rapid innovation in product categories that high-income households tend to purchase (Jaravel,

[10]Jaravel notes other potential data sources that could be exploited for measuring inflation inequality in the space of digital "free" goods, including Google, Skype, Wikipedia, maps, messaging, music, and all smartphone apps. Large-scale online choice experiments could be used for this purpose, as in Brynjolfsson et al. (2020).

2019). The digital divide also factors into differential inflation rates faced by higher and lower income consumer units. Data from Adobe's Digital Economy Index found online prices over the period 2014 to 2017 to be more than 3 percentage points lower than the headline inflation rate of 6.8 percent. Crucially, "shopping online is far more common among high-income people…and during the pandemic the practice has grown more prevalent" (Goolsbee, 2021). Another possible source of differential inflation rates is that outlets in low-income neighborhoods have fewer direct competitors given that consumers likely have lower mobility and ability to shop elsewhere. Research and policy making stand to benefit a great deal if these trends can be more accurately measured and, in turn, better understood.

The potential return from investments in developing income-defined subgroup price indexes is further enhanced by ongoing work at the Bureau of Economic Analysis (BEA) to produce prototype statistics on the distribution of personal income across households, for which price deflators will be needed. BEA's goal in estimating the distribution of personal income over time (the initial set of estimates does so by decile, quintile, top 1 percent, and 5 percent) is to "provide a new tool for assessing how households share in the nation's economic growth."[11] Interest in price indexes and inflation rates appropriate for deflating these new measures of income and, more broadly, for measures of well-being (social welfare) will likely continue to grow. BEA also has research underway to develop Personal Consumption Expenditures statistics by income decile.

> **Recommendation 6.1:** Because of the urgency of issues related to income and wealth inequality, social welfare, and poverty, developing price indexes for population subgroups along the income distribution should be a high priority for BLS. Identifying data sources that would ultimately allow production of price indexes by income quintile or, if possible, decile is a key part of this work.

Income-based CPIs, even if only experimental, would inform cost-benefit analyses, taxation policies, and understanding of secular macroeconomic trends, including structural change, changes in the labor share and interest rates, and labor market polarization. If a "Homescan-type" data source can be developed that ties purchases to households (as recommended in Chapter 3), it will open up a wide range of possibilities for creating a range of subgroup indexes that go beyond simply using different expenditure weights.[12]

Twenty years ago, the seemingly obvious first step in creating capacity to estimate accurate expenditure weights at subnational levels, and, in turn, facilitating subgroup CPIs, would have been to expand the CEX. It is well documented that the survey's current sample size does not allow for "production of non-urban-area indexes or regional price-level comparisons; nor does it support accurate price indexes for subpopulations such as the elderly, minorities, or the poor, particularly at subnational levels" (NRC, 2002, p. 252). Beyond price measurement,

---

[11]Full documentation of the distribution of personal income project can be found on BEA's webpage: https://www.bea.gov/data/special-topics/distribution-of-personal-income.

[12] While this chapter emphasizes price indexes for subgroups defined by income level, there are other important population dimensions to consider. Price indexes for subgroups defined by age (e.g., older populations) that account for differences in prices paid should also be a high priority because of their potential applications to policy programs. As revealed by Regional Price Parities developed by BEA, there is also considerable regional dispersion of inflation rates.

other national statistics such as poverty and savings rates that require data from the CEX could also benefit from a larger national sample (Triplett, 1997).

While the above conclusions are no doubt correct, it has become less clear in recent decades that expanding the CEX is the most effective route to building capacity for estimating subgroup price indexes or, as argued in Chapter 3, even to making the expenditure share weights for the flagship CPI more accurate. A clear cause for concern is the future viability of the CEX—specifically, the dual problems of declining response rates and increasing costs—a point made repeatedly in this report.

Even more germane to the question of subgroup indexes, however, is the need to simultaneously collect both expenditure data and timely information on prices paid at the detailed item level. As evidence accumulates of differential price inflation experiences for similar (or identical) goods purchased by different households, the inadequacy of indexes constructed solely by re-weighting expenditure shares has become apparent. Indeed, indexes based on such partial information are just as likely to mislead as to enlighten because, in a sense, they only pretend to answer a question that they are not equipped to answer. For this reason, other directions should be pursued when planning the development of price indexes for population subgroups.

> **Recommendation 6.2:** Even though the marginal cost of such exercises is not high, valuable CPI program resources should not be devoted to developing additional subgroup price indexes that simply entail a re-weighting of upper-level expenditure categories.

The authoring panel of *At What Price?* argued that "BLS should explore collecting prices in a way that allows them to be associated with household characteristics," a proposal that would require additional resources for data expansion (NRC, 2002, p. 241). Although the current panel agrees with this conclusion, the point made above—that putting all resources for developing subgroup price indexes into the CEX would be the wrong way to go—is worth reiterating here. Since data are needed not only on broad expenditure shares, but also on within-strata shopping behavior and effective prices paid, BLS will need to pursue a more expansive data infrastructure for measuring differential inflation across the income distribution.

An important first step in BLS's research agenda should be to identify and report on the most promising data sources for linking prices paid to the households (or groups of households) making the purchases. As described in Chapter 3, an admittedly ambitious long-term vision is to fund establishment of an in-house capacity for BLS to collect and coordinate electronic transaction data for tracking prices and product information for individual purchases. Such a program might include setting up a home scan project—perhaps, initially, as a small-scale pilot within the CEX—in which participants record their purchases or scan their final receipts. But several options are available for linking prices paid to particular households. In the shorter run, during the testing phase, it is likely that BLS will have to buy data[13]; in the longer run the agency may be able to collect its own data directly.

[13]The Nielsen Consumer (Homescan) Panel is the most prominent example of this kind of third party. The Homescan panel tracks the expenditures of about 55,000 households who scan the bar codes of their purchased items. Prices are then downloaded from the store where the item was purchased.

As with other aspects of CPI modernization discussed in this report, the real long-term promise for creative initiatives for subgroup indexes comes with the increased availability of micro data—in particular, household-based scanner data and perhaps transactions data from credit/debit card purchases to help estimate expenditure weights—that contain information on prices that individual households actually pay and on details about the items purchased. These kinds of data allow comparisons of the prices paid by goods with the same barcode across households with different incomes or other characteristics.

Where possible, these data elements should be used in ways that complement the statistical system's most important survey sources in a blended data infrastructure. For example, the ACS is a valuable source of information on household characteristics such as income and geography. Moretti (2013) used ACS data on rents to examine rent inflation by education group. At some point, it may become viable to link ACS records with data on household transactions from credit card and other electronic data sources. This kind of data integration, where prices of purchases are linked to specific households, would revolutionize statistical agencies' capacity to develop subgroup price indexes.

While opening new doors for studying expenditure patterns at the granular levels needed to more fully measure differences in inflation by income group, electronic transactions data, at least as currently generated, do not cover all consumer expenditures; indeed, several key categories are missing. For this reason, as described by Jaravel and O'Connor (2020), the next generation of empirical studies of inflation inequality will need to draw not just from scanner data for tracking fast-moving consumer goods, but also from additional, alternative data sources on other sectors.

> **Recommendation 6.3:** To identify and obtain the data necessary to estimate accurate subgroup price indexes, no one size will fit every category of goods and services. BLS will have to be creative and flexible in finding and blending different data sources. Exploiting commercial datasets on a range of household purchases will be essential.

The above-described research linking individuals to their purchases strongly suggests the need for approaches that blend multiple data sources in a way that account for the full range of consumer expenditures. Especially important are those categories that are likely to impart disproportionate impacts on inflation measures. The goal should be to take advantage of both survey data, typically from statistical agencies, that cover the full consumption basket, including item categories for which electronic transaction data are still incomplete, and commercial data sources that allow deep analyses of price and product detail for specific sectors.

BLS's research program will initially need to focus on a limited set of goods for which data on prices paid and incomes of purchasing households are available. As also recommended by NRC (2002), work on subgroup indexes, designed to investigate several alternative approaches, should be initially conducted for a selection of commodity categories and demographic groups. As this research develops, BLS should be open to publishing a range of experimental price indexes as warranted by research and user needs. As progress is made on developing subgroup price indexes, BLS will need to maintain a strong communications effort to help users understand the best ways to utilize newly developed indexes as well as their limitations. This communication strategy may be all the more important for subgroup index programs since they have significant implications if used for indexing.

7
Organizational Considerations and Overarching Guidance

This report is intended to provide actionable steps as the Bureau of Labor Statistics (BLS) continues modernization of its Consumer Price Index (CPI) program. While recommendations for advancing the data infrastructure supporting elementary price index estimation, higher level index aggregation, and other aspects of CPI construction were the panel's primary focus, a few systemic considerations are presented as well in this concluding chapter.

7.1. COORDINATION WITHIN BLS

A shift toward greater use of alternative and nontraditional data sources is a complex task touching on many dimensions of CPI data acquisition and methodology. This feature creates challenges for organizational authority and accountability within BLS. To meet this challenge, BLS should build data modernization into its organizational structure.

> **Recommendation 7.1:** BLS should designate a single, high-level person within the agency, preferably as the deputy commissioner level, whose job is to lead data transformation efforts. Having this responsibility explicitly designated would facilitate a focused, coordinated effort and would ensure accountability. This person also could be the visible point person for coordination with the Bureau of Economic Analysis, the Census Bureau, U.S. Department of Agriculture, and other statistical agencies that are likewise in the process of data modernization initiatives. A key objective is to avoid duplicative efforts that likely would arise if data transformation proceeded in a more decentralized (siloed) way within BLS.

The data transformation lead would also be part of the team tasked with developing communication strategies to work with Congress to seek the necessary resources to implement changes and highlight the value of the task to user communities and to the general public.

7.2. INTERAGENCY COLLABORATION

The decentralized design of the U.S. statistical system heightens the need for thoughtful collaboration among statistical agencies as data modernization proceeds. Abraham et al. (2021, p. 16)[1] described the situation:

> A central set of challenges for realizing the potential of Big Data for economic statistics arises from the way in which the agencies' collaborations with businesses and with each other are structured. Historically, each of the three main economic statistics agencies—the BLS, Census Bureau, and BEA—has had a well-defined set of largely distinct responsibilities. Although there always has been collaboration among the agencies, each agency collects the information from businesses that it needs for specific statistical series and produces those series independently. In a Big Data world, however, there are compelling reasons for agencies to adopt more integrated data collection and production processes.

Key economic indicators such as national output and income rely on data produced from the multiple statistical agencies identified in the above quote. Although coordination already exists among these agencies, more will be needed for the acquisition and innovative use of alternative data sources in these efforts.

> **Recommendation 7.2:** More extensive collaboration between BLS, the Census Bureau, and the Bureau of Economic Analysis—along with other statistical agencies that collect key economic data, such as the U.S. Department of Agriculture—is needed to advance the acquisition and use of alternative data sources in the production of economic indicators. More specifically, such coordination will allow the statistical system to negotiate common, unified, comprehensive contracts with companies (once, not multiple times) that collect applicable data.

USDA could be a particularly valuable partner. The department's Economic Research Service (Food Economics Division), which collects data to inform policies related to federal nutrition assistance programs, has a history of acquiring proprietary scanner and other transactions data for the purpose of estimating food prices, quantities of sales, and acquisition of food for at-home and away-from-home eating (NASEM, 2020). It might also be useful to include agencies that perform research and use data extensively, such as the Energy Information Administration, the Office for Financial Research, the Federal Reserve, and the Department of Transportation's Bureau of Transportation Statistics.

Ideally, collaboration among the statistical agencies would culminate with the creation of a Joint Office to administer collection of electronic data. For example, if the Census Bureau were interested in a large dataset, it would coordinate with its statistical agency partners who would all be able to access the data. Coordinating such acquisition, while adding interagency complexity, has the potential to reduce duplication, save resources, and enhance cross-agency spillovers. Indeed, formalizing institutional coordination of data acquisition would signal commitment by

---

[1] https://www.nber.org/system/files/chapters/c14265/c14265.pdf.

senior leadership of the statistical agencies to the increased use of alternative data. In the case of the CPI, these efforts would help lay the foundation for a world when most transactions leave an electronic record, which, in turn, may ultimately become the principal source of input data for price measurement.

The call for greater interagency collaboration in securing data from outside the statistical system is consistent with similar recommendations put forth in other reports.[2] An important task of the joint effort by the statistical agencies would be to create incentives for nongovernment data producers to engage in public-private partnerships. The agencies will need to be creative in establishing mutually beneficial agreements that might, for example, offer value-added products back to data providers in return for data access. For the most part, this work will require breaking new ground although some organizations have made progress along these lines. For example, Adobe Insights (Lasiy, White, and Pandya, 2020) collects merchant data on online transactions to produce timely estimates of spending and quantities of varieties of certain goods. Adobe worked out a deal with retailers to provide benchmark indexes that are useful to clients. This initiative created the opportunity to make comparisons across peer firms, and that inducement was enough to persuade retailers to provide data. Similarly, Statistics Netherlands sent selected indexes back to data companies in return for the source data.

7.3. COLLABORATION AND COMMUNICATION

BLS should cast its collaborative efforts more broadly, beyond the U.S. statistical system, as well. With price measurement in particular, ample opportunities exist to replicate (with needed modifications) innovative approaches to the use of alternative data sources that have been developed by various non-U.S. national statistical offices.

> **Recommendation 7.3:** BLS should enhance its contacts and collaborations with CPI staff in statistical agencies beyond the U.S. system. Other countries have made significant progress on data transformation—specifically in methods blending scanner and web-scraped data with survey sources—and CPI staff would benefit from more fully investigating successes and failures experienced during these efforts.

Some of the most innovative use of alternative datasets has taken place in academic settings, so continued collaboration by BLS with academic and other outside experts is likewise encouraged.

> **Recommendation 7.4:** BLS should enhance its interactions with outside experts (in academia, industry, and elsewhere) through collaborative research to leverage the latest advances in research on price measurement methods.

---

[2]NASEM (2017, p. 3) recommended that a higher-level cross-agency entity "should assist federal statistical agencies in identifying data sources that can most effectively inform the creation of national statistics, help develop techniques to use data from these sources to compute national statistics while respecting privacy and other protection obligations on the data, and nurture the expertise required to perform these functions" (https://www.nap.edu/read/24893/chapter/1. The Interagency Council on Statistical Policy (ICSP, https://www.usa.gov/federal-agencies/federal-interagency-council-on-statistical-policy)—established to "improve communication among the statistical agencies" and which includes membership of all officials at U.S. statistical agencies—may provide a good starting place.

Existing venues that could be particularly valuable for these efforts include the Federal Economic Statistics Advisory Committee (FESAC) and the BLS's advisory committees, which could play a larger role in guiding data transformation efforts. BLS also could look to models used by statistical agencies in other countries, such as the Statistics Canada advisory board model, which includes real-time consulting on data modernization issues as they arise.[3]

The type of data modernization for the CPI envisioned in this report will require a technical staff with an expanded set of statistical and computational skills. The panel recognizes that BLS is well aware of these staffing challenges. Accordingly, while BLS should expand collaboration with experts from beyond the agency (and beyond the statistical system), a long-term goal is certainly to expand and redirect its own in-house staff's skill portfolio.

> **Recommendation 7.5:** In addition to hiring staff with data science skills, BLS should strive to develop this talent in-house by supporting and rewarding staff who pursue training and educational opportunities to develop the technical expertise that will facilitate data transformation efforts in coming years.

This message that has been articulated frequently in other recent reports. NASEM (2017, p. 149), for example, recommended that "federal statistical agencies should ensure their technical staff receive appropriate training in modern computer science technology including but not limited to database, cryptography, privacy-preserving, and privacy-enhancing technologies."

Since confidence in and understanding by data users of official statistics is critical, successful modernization of the CPI will require that BLS provide clear and consistent communication about the re-design on an ongoing basis.

> **Recommendation 7.6:** As CPI modernization proceeds, BLS should ensure that key information is readily available to all stakeholders—such as by posting in an easy-to-find location of the website—including advance notice of changes, detail about alternative data sources incorporated, transparency around experimental indexes, and updates on the timeline for the project as it evolves. The agency also should aggressively and frequently communicate with stakeholders in the user and research communities.

Such updates and visibility will be especially important during times of rapid changes, such as the dramatic shift in purchasing patterns during the pandemic and the associated heightened interest in the compiling of CPI relative importance weights during and following the pandemic.

More generally, the key is for BLS to ensure that stakeholders can plan and use data appropriately. A key element of this communication, as alternative data sources are incorporated into the CPI, is to ensure that stakeholders know the specific sources of data and index methodology used for individual components as well as the terms on which BLS obtained the data so that all users are an equal footing in interpreting CPI releases.

One statistical office that undertook aggressive and frequent communication about the use of alternative data sources is the Australian Bureau of Statistics. That agency collaborated with international experts (including statistical offices in other countries) and consulted with key stakeholders (e.g., the Reserve Bank of Australia, the Treasury, and the Department of Social

---

[3]Information about Statistics Canada advisory groups can be found at https://www.statcan.gc.ca/en/about/relevant.

Services) to resolve outstanding methodological issues associated with its data modernization effort, particularly the use of transactions data to compile the CPI.[4] Estimating parallel series based on alternative data for an experimental period, as suggested below, will support data quality and public perceptions of integrity of the data collection changes.

7.4. DATA ACQUISITION AND ACCESS

Beyond content-oriented questions about coverage and representativeness, scope of variables, and transparency regarding methods, the potential of commercial data sources is often limited by legal access hurdles and privacy concerns (NASEM, 2020, p. 124).[5] Indeed, sometimes even data sources (e.g., administrative records) from other agencies are inaccessible. BLS documentation indicates that the main obstacles to adopting the kinds of information that could naturally be applied to price measurement, such as web-scraped data, have been administrative and legal (Konny et al., 2019, p. 7):

> Concerns regarding web scraping have arisen both internally and from respondents…To ensure all alternative data used in research or production is protected under CIPSEA[6], BLS must provide establishments, including those whose data we collect on-line, whether manually or automatically, a pledge of confidentiality promising to use the information for exclusively statistical purposes. In the case of web scraping, BLS cannot proceed without permission of the establishment.

An additional challenge that has been raised is the possibility of a data source that BLS does not directly control disappearing or being significantly altered if the data vendor's motivations for producing the data change. Further, data purchased from vendors may have been sampled, cleaned, and aggregated in a way that does not best serve the purposes of constructing the CPI. BLS and other statistical agencies must be mindful of the high stakes, in terms of policy and market impacts, associated with methodological changes to the CPI.

The panel recognizes these challenges. Nonetheless, the case for CPI modernization and the greater use of alternative data is compelling and the panel believes that the BLS, in coordination with the other statistical agencies, can effectively overcome organizational, administrative, and legal hurdles to move forward while maintaining the high quality the CPI is known for. Moreover, as legal and administrative frameworks become more established and standardized, these efforts should become more routine.

---

[4]For a description of this effort, see www.abs.gov.au/AUSSTATS/abs@.Nsf/39433889d406eeb9ca2570610019e9a5/40fc971083782000ca25768e002c845b!OpenDocument.

[5]See *Democratizing Our Data: A Manifesto*, by Julia Lane, which includes a probe into the bureaucratic and other obstacles that have impeded progress on data modernization in the U.S. statistical system.

[6]The Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002 requires that data collected be used strictly for statistical purposes and promises respondents high levels of data protection against disclosure of confidential information. See *Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002* at: https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/omb/inforeg/proposed_cispea_guidance.pdf.

References

Abraham, K.G., Jarmin, R.S., Moyer, B.C., and Shapiro, M.D. (2021). Big data for twenty-first century economic statistics: The future is now. Pp. 1–22 in *Big Data for Twenty-First Century Economic Statistics*, K.G. Abraham, R.S. Jarmin, B. Moyer, and M.D. Shapiro, editors. Cambridge, MA: National Bureau of Economic Research. Available: https://www.nber.org/system/files/chapters/c14265/c14265.pdf.

Adams, B., and Verbrugge, R. (2021). Location, Location, Structure Type: Rent Divergence within Neighborhoods (February 9, 2021). FRB of Cleveland Working Paper No. 21-03, Available: https://ssrn.com/abstract=3782829 or http://dx.doi.org/10.2139/ssrn.3782829.

Advisory Commission to Study the Consumer Price Index (Boskin Commission). (1996). *Toward a More Accurate Measure of the Cost of Living*. Final Report to the Senate Finance Committee, Washington, DC. Available: https://www.ssa.gov/history/reports/boskinrpt.html.

Ambrose, B.W., Coulson, N.E., and Yoshida, J. (2015). The repeat rent index. *Review of Economics and Statistics 97*: 939-950.

Amaya, A., Biemer, P.P., and Kinyon, D. (2020). Total error in a big data world: Adapting the TSE framework to big data. *Journal of Survey Statistics and Methodology, 8*(1): 89–119.

Amble, N., and Stewart, K. (1994). Experimental price index for elderly consumers. *Monthly Labor Review, 117*(5): 11–16.

Andersen, A.L., Hansen, E.T., Johannesen, N., and Sheridan, A. (2020). Consumer Responses to the COVID-19 Crisis: Evidence from Bank Account Transaction Data. CEBI Working Paper 18/20, Department of Economics, University of Copenhagen. Available: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3609814_code352274.pdf?abstractid=3609814&mirid=1.

Argente, D., and Lee, M. (2020). Cost of living inequality during the great recession. *Journal of the European Economic Association, 19*(2): 913–952.

Armknecht, P.A., Moulton, B.R., and Stewart, K.J. (1995). Improvements to the Food at Home, Shelter, and Prescription Drug Indexes in the U.S. Consumer Price Index. Working Paper No. 263, U.S. Department of Labor, Bureau of Labor Statistics, Washington, DC. Available: https://www.bls.gov/osmr/research-papers/1995/pdf/ec950010.pdf.

Aten, B.H., and Heston, A. (2020). The Owner-Premium Adjustment in Housing Imputations.

Bureau of Economic Analysis Working Paper Series, WP2020-7. Available: https://www.bea.gov/system/files/papers/BEA-WP2020-7_0.pdf.

Auer, J., and Boettcher, I. (2017). From Price Collection to Price Data Analytics: How New Large Data Sources Require Price Statisticians to Re-think Their Index Compilation Procedures. Experiences from Web-scraped and Scanner Data. Paper prepared for the Fifteenth Meeting of the International Working Group on Price Indices, Eltville am Rhein, Germany, May 10–12. Available: https://www.bundesbank.de/en/service/dates/fifteenth-meeting-of-the-ottawa-group-2017-634968.

Australian Bureau of Statistics. (2016). Making Greater Use of Transactions Data to Compile the Consumer Price Index. ABS Information Paper, No. 6401.0.60.003, Canberra. Available: https://www.abs.gov.au/ausstats/abs@.nsf/mf/6401.0.60.003.

Australian Bureau of Statistics (2017). An Implementation Plan to Maximise the Use of Transactions Data in the CPI. ABS Information Paper, No. 6401.0.60.004, D. Kalisch, editor. Available: https://www.abs.gov.au/ausstats/abs@.nsf/mf/6401.0.60.004.

Bajari, P., Cen, Z., Chernozhukov, V., Manukonda, M., Wang, J., Huerta, R., Li, J., Leng, L., Monokroussos, G., Vijaykunar, S., and Wan, S. (2021). Hedonic Prices and Quality Adjusted Price Indices Powered by AI. Cemmap Working Paper, No. CWP04/21, Centre for Microdata Methods and Practice, London. Available: http://dx.doi.org/10.47004/wp.cem.2021.0421.

Balk, B.M. (1980). A method for constructing price indices for seasonal commodities. *Journal of the Royal Statistical Society, Series A, 143*(1): 68–75.

Bean, C. (2016). Independent Review of UK Economic Statistics. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf.

Benitez-Silva, H., Eren, S., Heiland, F., and Jimenez-Martin, S. (2015). How well do individuals predict the selling prices of their homes? *Journal of Housing Economics 29*(C): 12–25.

Bergqvist, T., Cage, R., Hoarty, B., Stanley, S., and Wilson, J. (2021). Options for Publishing Timely, Nearly Superlative Indexes. Presentation at the UNECE CPI Expert Group Meeting, June 9. Available: https://unece.org/sites/default/files/2021-05/Session_4_US-BLS_Presentation.pptx.

Berndt, E.R. (1991). *The Practice of Econometrics: Classic and Contemporary*. Reading, MA: AddisonWesley

Bieler, J., Cho, C., Matsumoto, B., Parker, B., and Wang, D. (2019). Using Insurance Claims Data in the Medical Price Indexes. American Economic Association Conference, Nashville, TN. Available: https://www.aeaweb.org/conference/2020/preliminary/paper/7i7sn2D3.

Biemer, P.P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly, 74*(5): 817–848.

Blair, C. (2015). Constructing a PCE-weighted Consumer Price Index. Pp. 53–74 in *Improving the Measurement of Consumer Expenditures*, Studies in Income and Wealth, Volume 74, C.D. Carroll, T.F. Crossley, and J. Sabelhaus, editors. Chicago: University of Chicago Press.

Bounie, D., Camara, Y., Fize, E., Galbraith, J., Landais, C., Lavest, C., Pazem, T., and Savatier, B. (2020). Consumption dynamics in the Covid crisis: Real time insights from French transaction & bank data. CEPR Discussion Paper No. DP15474 (November).

Brown, C., Sawyer, S., and Bathgate, D. (2020). *A Review of Hedonic Price Adjustment Techniques for Products Experiencing Rapid and Complex Quality Change*. Washington, DC: Bureau of Labor Statistics. Available: https://www.bls.gov/advisory/tac/review-of-hedonic-price-adjustment-techniques-for-products-experiencing-rapid-and-complex-quality-change-11-20-2020.pdf.

Brynjolfsson, E., Collis, A., Diewert, W.E., Eggers, F., and Fox, K.J. (2020). Measuring the impact of free goods on real household consumption. *AEA Papers and Proceedings, 110*: 25–30. Available: https://www.aeaweb.org/articles?id=10.1257/pandp.20201054.

Bureau of Labor Statistics. (2018). *Consumer Expenditure Surveys*. Washington, DC. Available: https://www.bls.gov/cex/cepceconcordance.htm.

Bureau of Labor Statistics. (2020). Consumer Price Index: Overview. Pp. 1–67 in *Handbook of Methods*. Washington, DC. Available: https://www.bls.gov/opub/hom/cpi/pdf/cpi.pdf.

Bureau of Labor Statistics. (2021). Measuring Price Change in the CPI: Medical Care. Available: https://www.bls.gov/cpi/factsheets/medical-care.htm.

Bureau of Labor Statistics. (undated). Disease Based Price Indexes. Available: https://www.bls.gov/pir/diseasehome.htm.

Cage, R., Klick, J., and Johnson, W. (2018). Population Subgroup Price Indexes: Evidence of Heterogeneity or Measurement Error? Presentation to the Meeting of the Group of Experts on Consumer Price Indexes, United Nations Economic Commission for Europe, Geneva, Switzerland. Available: https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/BLS_4.pdf.

Carroll, C.D., Crossley, T.F., and Sabelhaus, J. (2015*). Improving the Measurement of Consumer Expenditures*. Chicago: University of Chicago Press.

Carvalho, V.M., Hansen, S., Ortiz, A., García, J.R., Rodrigo, T., Rodríguez Mora, J.V., and Ruiz, P. (2020). Tracking the Covid-19 crisis with high-resolution transaction data. CEPR Discussion Paper No. DP14642 (July).

Cavallo, A. (2017). Are online and offline prices similar? Evidence from large multi-channel retailers. *American Economic Review, 107*(1): 283–303.

Cavallo, A. (2020). Inflation with Covid Consumption Baskets. NBER Working Paper No. 27352. Cambridge, MA: National Bureau of Economic Research. Available: https://www.nber.org/system/files/working_papers/w27352/w27352.pdf.

Cavallo, A., and Rigobon, R. (2016). The Billion Prices Project: Using online prices for measurement and research. *Journal of Economic Perspectives, 30*(2): 151–178.

Cecchetti, S. (2007). Housing in Inflation Measurement. VoxEU-CEPR, The Centre for Economic Policy Research. Available: https://voxeu.org/article/housing-inflation-measurement.

Centers for Disease Control and Prevention. (2021). Chronic Kidney Disease in the United States, 2021. Available: https://www.cdc.gov/kidneydisease/pdf/Chronic-Kidney-Disease-in-the-US-2021-h.pdf.

Chan, S., Dastrup, S., and Gould Ellen, I. (2015). Do homeowners mark to market? A comparison of self reported and estimated market home values during the housing boom and bust. *Real Estate Economics*, *44*(3): 627–657.

Chen, B., and Fixler, D.J. (2003). Measuring the services of property-casualty insurance in the NIPAs: Changes in concepts and methods. *Survey of Current Business, 83*: 10–26.

Chen, Z., Finkelstein, E.A., Karns, S.A., Leibtag, E., and Zhang, C. (2018). Scanner data-based panel price indexes. *American Journal of Agricultural Economics*, *101(1)*: 311–329. Available: https://doi.org/10.1093/ajae/aay032.

Chessa, A.G. (2016). A new methodology processing scanner data in the Dutch CPI. *Eurona, 1*: 49–69.

Chessa, A.G. (2019). A Comparison of Index Extension Methods for Multilateral Methods. Paper presented at the 16th Meeting of the International Working Group on Price Indices, Rio de Janeiro, Brazil, May 8–10. Available: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+16/$FILE/A%20comparison%20of%20index%20extension%20methods%20paper.pdf.

Chessa, A.G. (2021). Extension of Multilateral Index Series over Time: Analysis and Comparison of Methods. Paper written for the 2021 Meeting of the Group of Experts on Consumer Price Indices, May 7. Available: https://www.unece.org/sites/default/files/2021-05/Session_1_Netherlands_Paper.pdf.

Chetty, R., Friedman, J.N., Hendren, N., Stepner, M., and the Opportunity Insights Team (2020). The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data. National Bureau of Economic Research Working Paper 27431 (November). Available: https://www.nber.org/system/files/working_papers/w27431/w27431.pdf.

Chevalier, J., and Goolsbee, A. (2003). Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics, 1*: 203–222.

Chronopoulos, D.K., Lukas, M., and Wilson, J.O.S. (2020). Consumer spending responses to the COVID-19 Pandemic: An Assessment of Great Britain. Available: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3640244_code2453555.pdf?abstractid=3586723&mirid=1.

Cicala, S., Lieber, E.M.J., and Marone, V. (2019). Regulating markups in US health insurance. *American Economic Journal: Applied Economics 11*(4): 71-104.

Congressional Budget Office (CBO). (2008). Technological Change and the Growth of Health Care Spending. Available: https://www.cbo.gov/sites/default/files/110th-congress-2007-2008/reports/01-31-techhealth.pdf.

Court, A.T. (1939). Hedonic price indexes with automotive examples. Pp. 99-117 in *The Dynamics of Automobile Demand*, V. von Szeliski, S.L. Horner, and C.F. Roos, editors. New York: General Motors Corporation. Available: http://homepages.rpi.edu/~simonk/pdf/gm1939.pdf.

Cutler, D.M., Ghosh, K., Messer, K., Raghunathan, T., Rosen, A.B., and Stewart, S.T. (2020). A Satellite Account for Health in the United States. NBER Working Paper, No. 27848. Cambridge, MA: National Bureau of Economic Research. Available: https://www.nber.org/papers/w27848.

da Silva, L.T., de Oliveira, I.L., Dantas, T.M., and Miranda, V.G. (2019). Studies of New Data Sources and Techniques to Improve CPI Compilation in Brazil. Paper presented at the 16th International Working Group on Price Indices, Rio de Janeiro, Brazil, May 8–10. Available: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+16/$FILE/Studies%20of%20new%20data%20sources%20paper.pdf.

Dalén, J. (2017). Unit Values in Scanner Data: Some Operational Issues. Paper presented at the 15th Meeting of the International Working Group on Price Indices, Eltville am Rhein, Germany, May 10–12. Available: http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/1ab31c25da944ff5ca25822c00757f87/$FILE/Unit%20values%20in%20scanner%20data%20%E2%80%93%20some%20operational%20issues%20-%20J%C3%B6rgen%20Dal%C3%A9n%20-Paper.pdf.

Dauda, S., Dunn, A.C., and Hall, A.E. (2020). Are Medical Care Prices Still Declining? A Systematic Examination of Quality-Adjusted Price Index Alternatives for Medical Care. Hutchins Center Working Paper No. 59. Washington, DC: Brookings Institution. Available: https://www.brookings.edu/wp-content/uploads/2020/03/WP59_Dauda-et-al..pdf.

de Haan, J. (2008). Reducing Drift in Chained Superlative Price Indexes for Highly Disaggregated Data. Paper presented at the Economic Measurement Workshop, Centre for Applied Economic Research, University of New South Wales, December 10. Available: https://www.researchgate.net/publication/229051374_Reducing_Drift_in_Chained_Superlative_Price_Indexes_for_Highly_Disaggregated_Data.

de Haan, J. (2015). A Framework for Large Scale Use of Scanner Data in the Dutch CPI. Paper presented at the 14th Meeting of the International Working Group on Price Indices, Tokyo, Japan, May 20–22. Available: https://www.stat.go.jp/english/info/meetings/og2015/pdf/t6s11p33_pap.pdf.

de Haan, J., and Daalmans, J. (2019). Scanner Data in the CPI: The Imputation CCDI Index Revisited. Paper presented at the 16th Meeting of the International Working Group on Price Indices, Rio de Janeiro, Brazil, May 8–10. Available: https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/presentation_jan_de_haan_og_2019_0.pdf.

de Haan, J., and Hendriks, R. (2013). Online Data, Fixed Effects, and the Construction of High-Frequency Price Indexes. Paper presented at the 13th Economic Measurement Group Workshop, November 28–29, University of New South Wales, Sydney, Australia. Available: https://www.researchgate.net/publication/303486094_Online_Data_Fixed_Effects_and_the_Construction_of_High-Frequency_Price_Indexes.

de Haan, J., and Krsinich, F. (2012). The Treatment of Unmatched Items in Rolling Year GEKS Price Indexes: Evidence from New Zealand Scanner Data. Paper presented at the Economic Measurement Group Workshop, Australian School of Business, University of New South Wales, Sydney, Australia. Available:

https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2012/New_Zealand_-_The_Treatment_of_Unmatched_Items_in_Rolling_Year.pdf.

de Haan, J., and Krsinich, F. (2014). Scanner data and the treatment of quality change in nonrevisable price indexes. *Journal of Business and Economic Statistics, 32*(3): 341–358.

de Haan, J., and Krsinich, F. (2018). Time dummy hedonic and quality-adjusted unit value indexes: Do they really differ? *Review of Income and Wealth, 64*(4): 757–776.

de Haan, J., and van der Grient, H. (2011). Eliminating chain drift in price indexes based on scanner data. *Journal of Econometrics, 161*(1): 36–46.

de Haan, J., Hendriks, R., and Scholz, M. (2021). Price measurement using scanner data: Time-product dummy versus time dummy hedonic indexes. *Review of Income and Wealth, 67*(2): 394–417.

de Haan, J., Willenborg, L., and Chessa, A.G. (2016). An Overview of Price Index Methods for Scanner Data. Paper presented at the Meeting of the Group of Experts on Consumer Price Indices, Geneva, Switzerland, May 2–4. Available: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2016/Session_1_room_doc_Netherlands_an_overview_of_price_index_methods.pdf.

Diamond, R. (2016). The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000. *American Economic Review, 106*(3): 479–524.

Diamond, R., and Moretti, E. (2021). Where Is Standard of Living the Highest? Local Prices and the Geography of Consumption. NBER Working Paper, No. 29533. Cambridge MA: National Bureau of Economic Research.

Diewert, W.E. (1995). Axiomatic and Economic Approaches to Elementary Price Indexes. NBER Working Paper, No. 5104. Cambridge, MA: National Bureau of Economic Research. Available: https://www.nber.org/system/files/working_papers/w5104/w5104.pdf.

Diewert, W.E. (1999). Axiomatic and economic approaches to international comparisons. Pp. 13–87 in *International and Interarea Comparisons of Income, Output, and Prices,* A. Heston and R.E. Lipsey, editors. Chicago: The University of Chicago Press. Available: https://www.nber.org/system/files/chapters/c8385/c8385.pdf.

Diewert, W.E. (2002). Harmonized indexes of consumer prices: Their conceptual foundations. *Swiss Journal of Economics and Statistics, 138*(4): 547–637.

Diewert, W.E. (2004). On the Stochastic Approach to Linking the Regions in the ICP. Discussion Paper No. 04-16, Department of Economics, The University of British Columbia, Vancouver, Canada. Available: http://papers.economics.ubc.ca/legacypapers/dp0416.pdf.

Diewert, W.E. (2008). OECD workshop on productivity analysis and measurement: Conclusions and future directions. Pp. 11–36 in *Proceedings from the OECD Workshops on Productivity Measurement and Analysis*. Paris: OECD.

Diewert, W.E. (2011). The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and future directions. Pp. 87–116 in *Price and Productivity Measurement: Volume 1-Housing,* W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox, and A.O. Nakamura, editors. Victoria, Canada: Trafford Press. Available: http://www.indexmeasures.ca/price_productivity,vol[1].1with%20cover,04,03,09.pdf.

Diewert, W.E. (2018). Scanner Data, Elementary Price Indexes, and the Chain Drift Problem. Discussion Paper No. 18-06, Vancouver School of Economics, The University of British Columbia, Vancouver, Canada. Available: https://econ2017.sites.olt.ubc.ca/files/2018/10/pdf_paper_diewert-DP18-06-ScannerDataElementaryPriceetc_oct2018.pdf.

Diewert, W.E. (2020a). The Economic Approach to Index Number Theory. Chapter 5 in *Consumer Price Index Theory*. Washington, DC: International Monetary Fund. Available: https://www.imf.org/-/media/Files/Data/CPI/companion-publication/chapter-4-economic-approach.ashx.

Diewert, W.E. (2020b). The treatment of durable goods and housing. Chapter 10 in *Consumer Price Index Theory*. Washington, DC: International Monetary Fund. Available: https://econ2017.sites.olt.ubc.ca/files/2021/05/pdf_paper_diewert-erwin_IMFCPIChapter10.pdf.

Diewert, W.E. (2021a). Quality adjustment methods. Chapter 8 in *Consumer Price Index Theory*. Washington, DC: International Monetary Fund. Available: https://econ2017.sites.olt.ubc.ca/files/2021/04/pdf_paper_diewert-erwin_IMFCPIChapter8.pdf.

Diewert, W.E. (2021b). Elementary indexes. Chapter 6 in *Consumer Price Index Theory*. Washington, DC: International Monetary Fund. Available: https://econ2017.sites.olt.ubc.ca/files/2021/04/pdf_paper_diewert-erwin_IMFCPIChapter6.pdf.

Diewert, W.E. (2021c). The chain drift problem and multilateral indexes. Chapter 7 in *Consumer Price Index Theory*. Washington DC: International Monetary Fund. Available: https://econ2017.sites.olt.ubc.ca/files/2021/04/pdf_paper_diewert-erwin_IMFCPIChapter7.pdf.

Diewert, W.E., and Feenstra, R.C. (2017). Estimating the Benefits and Costs of New and Disappearing Products. Discussion Paper 17-10, Vancouver School of Economics, University of British Columbia. Available: https://www.business.unsw.edu.au/research-site/centreforappliedeconomicresearch-site/newsandevents-site/workshops-site/Documents/Erwin-Diewert-Estimating-the-Benefit-%20and-Costs-of-New-and-Disappearing-Products.pdf.

Diewert, W.E., and Feenstra, R. (2021). Estimating the benefits of new products. In *Big Data for Twenty-First-Century Economic Statistics*, NBER Studies in Income and Wealth Vol. 79, K.G. Abraham, R.S. Jarmin, B. Moyer, and M.D. Shapiro, editors. Available: https://www.nber.org/system/files/chapters/c14281/c14281.pdf.

Diewert, W.E., and Fox, K.J. (2020a). Measuring Real Consumption and CPI Bias Under Lockdown Conditions. NBER Working Paper, No. 27144. Cambridge MA: National Bureau of Research. Available: https://www.nber.org/system/files/working_papers/w27144/w27144.pdf.

Diewert, W.E., and Fox, K.J. (2020b). Substitution bias in multilateral methods for CPI construction. *Journal of Business & Economic Statistics*, *40*(1): 355–369.

Diewert, W.E., and Shimizu, C. (2021). The treatment of durable goods and housing. Chapter 10 in *Consumer Price Index Theory*. Washington DC: International Monetary Fund.

Available: https://econ2017.sites.olt.ubc.ca/files/2021/05/pdf_paper_diewert-erwin_IMFCPIChapter10.pdf.

Diewert, W.E., Finkel, Y., and Sayag, D. (2020). Seasonal products. Chapter 9 in *Consumer Price Index Theory*. Washington DC: International Monetary Fund. Available: https://econ2017.sites.olt.ubc.ca/files/2021/04/pdf_paper_diewert-erwin_IMFCPIChapter8.pdf.

Diewert, W.E., Fox, K.J., and Schreyer, P. (2017). The Digital Economy, New Products, and Consumer Welfare. Discussion Paper No. 17-09, Vancouver School of Economics, University of British Columbia, Vancouver, BC, Canada. Available: https://econ2017.sites.olt.ubc.ca/files/2017/12/pdf_paper_erwin-diewert-17-09DigitalEconomy-1.pdf.

Diewert, W.E., Nakamura, A.O., and Nakamura, L.I. (2009). The housing bubble and a new approach to accounting for housing in a CPI. *Journal of Housing Economics, 18*(3): 156–171.

DiPasquale, D., and Somerville, C.T. (1995). Do house price indices based on transacting units represent the entire stock? Evidence from the American Housing Survey. *Journal of Housing Economics, 4*: 195–229.

Dunn, A., Liebman, E., and Shapiro, A.H. (2014). Implications of utilization shifts on medical-care price measurement. *Health Economics*, *24*(5): 530–557. Available: https://doi.org/10.1002/hec.3036.

Dunn, A., Rittmueller, L., and Whitmire, B. (2015). Introducing the new BEA health care satellite account. *Survey of Current Business, 95*(1): 1–21.

Ehrlich, G., Haltiwanger, J. Olivares, E., Shapiro, M.D., and Zhao, L.Y. (2021). Quality Adjustment at Scale: Hedonic vs. Exact Demand-Based Price Indices. Paper presented at the 36th International Association for Research in Income and Wealth Virtual General Conference, August 23–27. Available: https://iariw.org/wp-content/uploads/2021/08/Price_Quantity_Scale_paper.pdf.

Eltetö, Ö., and Köves, P. (1964). On a problem of index number computation relating to international comparisons. *Statisztikai Szemle, 42*(5): 507–518.

Erickson, T., and Pakes, A. (2011). An experimental component index for the CPI: From annual computer data to monthly data on other goods. *American Economic Review, 101*(5): 1707–1738.

Eurostat. (2013). Eurostat Quality Assurance Framework (Based on ESS QAF V1.1). Available: https://ec.europa.eu/eurostat/documents/64157/4372717/Eurostat-Quality-Assurance-Framework-June-2013-ver-1-1-EN.pdf/352234ca-77a0-47ca-93c7-d313d760bbd6.

Eurostat. (2018). *Harmonised Index of Consumer Prices (HICP) Methodological Manual*. Luxembourg: Publications Office of the European Union. Available: https://ec.europa.eu/eurostat/documents/3859598/9479325/KS-GQ-17-015-EN-N.pdf/d5e63427-c588-479f-9b19-f4b4d698f2a2.

Eurostat. (2020). Guidance on the compilation of HICP weights in case of large changes in consumer expenditures. Eurostat Methodological Note. Available: https://ec.europa.eu/eurostat/documents/10186/10693286/Guidance-on-the-compilation-of-HICP-weights-in-case-of-large-changes-in-consumer-expenditures.pdf.

Feenstra, R.C. (1994). New product varieties and the measurement of international prices. *American Economic Review*, *84*(1): 157–177.

Feenstra, R.C., and Shapiro, M.D. (2001). High-Frequency Substitution and the Measurement of Price Indexes. NBER Working Paper, No. 8176. Cambridge, MA: National Bureau of Economic Research. Available: https://www.nber.org/system/files/working_papers/w8176/w8176.pdf.

Feenstra, R.C., and Shapiro, M.D. (2003). High-frequency substitution and the measurement of price indexes. Chapter 5 in *Scanner Data and Price Indexes*, R.C. Feenstra and M.D. Shapiro, editors. Chicago: The University of Chicago Press.

Feenstra, R.C., and Weinstein, D.E. (2017). Globalization, markups and US welfare. *Journal of Political Economy*, *125*: 1040–1074.

Fixler, D.J., de Francisco, E., and Kanal, D. (2021). The revisions to Gross Domestic Product, Gross Domestic Income, and their major components. *Survey of Current Business, 101*(1). Available: https://apps.bea.gov/scb/2021/01-january/0121-revisions-to-gdp-gdi.htm.

Garner, T.I., and Verbrugge, R. (2009). Reconciling user costs and rental equivalence: Evidence from the US Consumer Expenditure Survey. *Journal of Housing Economics, 18*(3): 172–192.

Garner, T.I., Johnson, D., and Kokoski, M. (1996). An experimental consumer price index for the poor. *Monthly Labor Review, 119*(9): 32–42.

Geary, R.G. (1958). A note on comparisons of exchange rates and purchasing power between countries. *Journal of the Royal Statistical Society, Series A, 121*(1): 97–99.

Gillingham, R. (1983). Measuring the cost of shelter for homeowners: Theoretical and empirical considerations. *Review of Economics and Statistics, 65*(2): 254–265.

Gillingham, R., and Lane, W. (1982). Changing the treatment of shelter costs for homeowners in the CPI. *Monthly Labor Review, 105*(6): 9–14.

Gindelsky, M., Moulton, J.G., and Wentland, S.A. (2020). Valuing housing services in the era of big data: A user cost approach leveraging Zillow microdata. In *Big Data for Twenty-First-Century Economic Statistics,* NBER Studies in Income and Wealth, Volume 79, K.G. Abraham, R.S. Jarmin, B. Moyer, and M.D. Shapiro, editors. Chicago: University of Chicago Press.

Gini, C. (1931). On the circular test of index numbers. *Metron,* 9(2): 3–24.

Goodman, J.L., and Ittner, J.B. (1992). The accuracy of home owners' estimates of house value. *Journal of Housing Economics, 2*(4): 339–357.

Goolsbee, A.D. (2021). The missing data in the inflation debate. *The New York Times,* December 30. Available: http://www.nytimes.com/2021/12/30/opinion/inflation-economy-biden-inequality.html.

Goolsbee, A.D., and Klenow, P.J. (2018). Internet rising, prices falling: Measuring inflation in a world of E-Commerce. *AEA Papers and Proceedings, 108*: 488–492.

Greenlees, J.S. (2006). The BLS Response to the Boskin Commission Report. *International Productivity Monitor, no. 12(Spring):* 23–41.

Government Accountability Office. (2020). *Retirement Security: BLS Should Explore Ways to Improve the Accuracy, Timeliness, and Relevance of its Cost-of-Living Measurements*.

Report to Congressional Requestors No. GAO-20-422, Washington, DC: GPO. Available: https://www.gao.gov/assets/gao-20-422.pdf.

Guerrieri, V., Hartley, D., and Hurst, E. (2013). Endogenous gentrification and housing price dynamics. *Journal of Public Economics, Elsevier, 100*(C): 45–60.

Gürer, E., and Weichenrieder, A. (2020). Pro-rich inflation in Europe: Implications for the measurement of inequality. *Journal German Economic Review, 21*(1): 107–138.

Hart, N., and Potok, N. (2020). *Modernizing U.S. Data Infrastructure: Design Considerations for Implementing a National Secure Data Service to Improve Statistics and Evidence Building*. Washington, DC: Data Foundation. Available: https://www.datafoundation.org/modernizing-us-data-infrastructure-2020#executive-summary.

Hausman, J., and Leibtag, E. (2010). CPI bias from supercenters. Pp. 203–231 in *Price Index Concepts and Measurement*, W.E. Diewert, J.S. Greenlees, and C.R. Hulten, editors. Chicago: University of Chicago Press. Available: https://chicago.universitypressscholarship.com/view/10.7208/chicago/9780226148571.001.0001/upso-9780226148557-chapter-6.

Heston, A., and Nakamura, A.O. (2009). Questions about the equivalence of market rents and user costs for owner occupied housing. *Journal of Housing Economics, 18*(3): 273–279.

Hill, R.J. (2004). Accounting for unexpected capital gains on natural assets in net national product. *Empirical Economics, 29*(4): 803–824.

Hill, R.J., Steurer, M., and Waltl, S.R. (2018). Owner Occupied Housing in the CPI and Its Impact On Monetary Policy During Housing Booms and Busts. Luxembourg Institute of Socio-Economic Research Working Paper No. 2018-05.

Hobijn, B., and Lagakos, D. (2005). Inflation inequality in the United States. *Review of Income and Wealth, 51*(4): 581–606.

Hobijn, B., and Şahin, A. (2009). Job-finding and separation rates in the OECD. *Economics Letters, 104*(3): 107–111.

Holt, M., Pham, H.Q., and Webster, M. (2017). An Implementation Plan to Maximise the Use of Transactions Data in the CPI. ABS Information Paper, No. 6401.0.60.004. Canberra, Australia: ABS. Available: https://www.abs.gov.au/ausstats/abs@.nsf/mf/6401.0.60.004.

Howard, G., and Liebersohn, C. (2020). Regional Divergence and House Prices. Fisher College of Business Working Paper No. 2020-03-004 (May 27).

International Labour Office, International Monetary Fund, Organisation for Economic Co-operation and Development, Statistical Office of the European Communities, United Nations, The International Bank for Reconstruction and Development, and The World Bank. (2004). *Consumer Price Index Manual: Concepts and Methods*. Geneva: International Labour Office. Available: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/presentation/wcms_331153.pdf

International Monetary Fund. (2009). *System of National Accounts 2008.* Available: https://www.elibrary.imf.org/downloadpdf/books/071/10402-9789211615227-en/10402-9789211615227-en-book.xml.

International Monetary Fund. (1993). *System of National Accounts 2008.*

International Monetary Fund. (2009). *System of National Accounts 2008.* Available:

https://www.elibrary.imf.org/downloadpdf/books/071/10402-9789211615227-en/10402-9789211615227-en-book.xml.

Ivancic, L. (2007). Scanner Data and the Construction of Price Indices. Ph.D. Thesis, School of Economics, University of New South Wales, Sydney, Australia. Available: http://unsworks.unsw.edu.au/fapi/datastream/unsworks:1404/SOURCE02.

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

Janson, W., and Verbrugge, R. (2020). Will COVID-19–Induced Rental Nonpayment Drive Large Reductions in Shelter Inflation? Hints from the Great Recession. Federal Reserve Bank of Cleveland Working Paper No. 20-22 (July 8).

Jaravel, X. (2019). The unequal gains from product innovations: Evidence from the U.S. retail sector. *The Quarterly Journal of Economics, 134*(2): 715–783.

Jaravel, X., and O'Connell, M. (2020). High-frequency changes in shopping behaviours, promotions and the measurement of inflation: Evidence from the great lockdown. *Fiscal Studies, 41*(3): 733–755. Available: https://doi.org/10.1111/1475-5890.12241

Johnson, N. (2017). A Comparison of PCE and CPI: Methodological Differences in U.S. Inflation Calculation and Their Implications. Office of Survey Methods Research, Bureau of Labor Statistics Research Paper, Washington: DC. Available: https://www.bls.gov/osmr/research-papers/2017/pdf/st170010.pdf.

Jorgenson, D.W., and Schreyer, P. (2017). Measuring individual economic well-being and social welfare within the framework of the system of national accounts. *The Review of Income and Wealth, 63*(S2): S460–S477.

Kaiser Family Foundation (2020a). Section 10: Plan Funding in *2020 Employer Health Benefits Survey*, G. Claxton, M. Rae, G. Young, and D. McDermott, editors. Available: https://www.kff.org/report-section/ehbs-2020-section-10-plan-funding/.

Kaiser Family Foundation (2020b). *2020 Employer Health Benefits Survey*, G. Claxton, M. Rae, G. Young, and D. McDermott, editors. Available: https://files.kff.org/attachment/Report-Employer-Health-Benefits-2020-Annual-Survey.pdf.

Kaplan, G., and Schulhofer-Wohl, S. (2017). Inflation at the household level. *Journal of Monetary Economics, 91*(C): 19–38.

Katz, A.J. (1983). Valuing the services of consumer durables. *Review of Income and Wealth, Series 29*(December): 405–427.

Khamis, S.H. (1970). Properties and conditions for the existence of a new type of index number. *Sankhya: The Indian Journal of Statistics, Series B* 32: 81–98.

Kiel, K.A., and Zabel, J.E. (1999). The accuracy of owner-provided house values: The 1978-1991 American Housing Survey. *Real Estate Economics, 27*(2): 263–298.

Klick, J. (2017). Measurement of Chain Drift in the Chained CPI-U November 2017 U. S. Bureau of Labor Statistics Research Paper.

Klick, J. (2021). Improving Weight Representivity of Fixed Quantity Consumer Price Index Products. BLS economic working paper. Available: https://www.bls.gov/osmr/research-papers/2021/pdf/st210030.pdf.

Klick, J. (2021). Measuring consumer price change during economic downturns: A review of assumptions about consumer spending. *Beyond the Numbers: Prices and Spending 10*(13). Available: https://www.bls.gov/opub/btn/volume-10/measuring-consumer-price-change-during-economic-downturns.htm.

Konny, C.G., Williams, B.K., and Friedman, D.M. (2019). Big Data in the U.S. Consumer Price Index: Experiences & Plans. Paper presented at the National Bureau of Economic Research Conference—Big Data for 21st Century Economic Statistics, Washington, DC. Available: https://www.nber.org/system/files/chapters/c14280/c14280.pdf.

Krsinich, F. (2016). The FEWS index: Fixed effects with a window splice. *Journal of Official Statistics, 32*(2): 375–404.

Kurtzon, G. (2018). How Much Does Formula vs. Chaining Matter for a Cost-of-Living Index? The CPI-U vs. the C-CPI-U. U.S. Bureau of Labor Statistics Working Paper 498. Available: https://www.bls.gov/osmr/research-papers/2017/pdf/ec170060.pdf.

Kuznets, S. Epstein, L., and Jenks, E. (1941). *National Income and Its Composition, 1919-1938*. New York: National Bureau of Economic Research.

Lamboray, C. (2020). Use of High-Frequency and Other Alternative Data Sources in Price Measurement. Presentation to the panel on October 7. Available: https://www.nationalacademies.org/event/10-07-2020/docs/DF89846708A5A69DAD8DC28493B4883F235FAD32E45F

Landais, C., Bounie, D., Camara, Y., Fize, E., Galbraith, J.W., Lavest, C., Pazem, T., and Savatier, B. (2020). Consumption Dynamics in the COVID Crisis: Real Time Insights from French Transaction & Bank Data. CEPR Discussion Paper No. DP15474.

Larsen, D., and Molloy, R. (2021). Differences in Rent Growth by Income 1985-2019 and Implications for Real Income Inequality. FEDS Notes, Washington, DC: Board of Governors of the Federal Reserve System. Available: https://www.federalreserve.gov/econres/notes/feds-notes/differences-in-rent-growth-by-income-1985-2019-and-implications-for-real-income-inequality-20211105.htm.

Lasiy, C., White, L., and Pandya, V. (2020). Tracking the Impact of Covid-19 on the Online Economy in Real-Time. Paper presented at the 8th IMF Statistical Forum: Measuring the Economics of a Pandemic. Available: https://www.imf.org/-/media/Files/Conferences/2020/8th-stats-forum/paper-costa-lasiy-losaunne-white-and-vivek-pandya.ashx.

Lebow, D.E., and Rudd, J.B. (2003). Measurement error in the consumer price index: Where do we stand? *Journal of Economic Literature, 41*(1): 159–201.

Léonard, I., Sillard, P., Varlet, G., and Zoyem, J.-P. (2015). Scanner Data and Quality Adjustment. Ottawa, Canada: Ottawa Group. Available: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/d012f001b8a1cf6cca257eed008074c9/$FILE/Patrick%20Sillard%20(Institut%20National%20de%20la%20Statistique%20et%20des%20Etudes%20Economiques)%20Scanner%20data%20and%20quality%20adjustment.pdf.

Levin, D., Noriega, D. Dicken, C., Okrent, A.M., Harding, M., and Lovenheim, M. (2018). Examining Food Store Scanner Data: A Comparison of the IRI InfoScan Data With Other Data Sets, 2008–2012. Technical Bulletin Number 1949, U.S. Department of Agriculture, Economic Research Service. Available: https://www.ers.usda.gov/webdocs/publications/90355/tb-1949.pdf?v=724.4.

McClelland, R., and Reinsdorf, M. (1999). Small Sample Bias in Geometric Mean and Seasoned CPI Component Indexes. U.S. Bureau of Labor Statistics Working Paper 324. Available: https://www.bls.gov/osmr/research-papers/1999/pdf/ec990050.pdf.

McGranahan, L., and Paulson, A. (2006). Constructing the Chicago Fed Income Based Economic Index – Consumer Price Index: Inflation Experiences by Demographic Group: 1983–2005. Federal Reserve Bank of Chicago Working Paper 2005-20. Available: https://www.chicagofed.org/-/media/publications/working-papers/2005/wp2005-20-pdf.pdf.

Melser, D. (2011). Constructing High Frequency Price Indexes Using Scanner Data. Paper presented at the Twelfth Meeting of the International Working Group on Price Indices, Washington, DC. Available: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+12+-+Wellington+2011.

Michael, R.T. (1979). Variation across households in the rate of inflation. *Journal of Money, Credit and Banking, 11*(1): 32–46.

Molloy, R., and Nielsen, E.R. (2018). How can we measure the value of a Home? Comparing model-based estimates with owner-occupant estimates. FEDS Notes No. 2018-10-11.

Moretti, E. (2013). Real wage inequality. *American Economic Journal: Applied Economics, 5*(1): 65–103.

Moulton, B. (2018). The Measurement of Output, Prices, and Productivity: What's Changed Since the Boskin Commission. The Hutchins Center on Fiscal and Monetary Policy at The Brookings Institution. Available: https://www.brookings.edu/wp-content/uploads/2018/07/Moulton-report-v2.pdf.

Moulton, J.G. (2020). Valuing Housing Services in the Era of Big Data: A User Cost Approach Leveraging Zillow Microdata. Presentation to the panel on December 15.

National Academies of Sciences, Engineering, and Medicine. (2013). *Measuring What We Spend: Toward a New Consumer Expenditure Survey*. Washington, DC: The National Academies Press. Available: https://www.nap.edu/catalog/13520/measuring-what-we-spend-toward-a-new-consumer-expenditure-survey.

National Academies of Sciences, Engineering, and Medicine. (2017). *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps*. Washington, DC: The National Academies Press. Available: https://doi.org/10.17226/24893.

National Academies of Sciences, Engineering, and Medicine. (2020). *A Consumer Food Data System for 2030 and Beyond*. Washington, DC: The National Academies Press. Available: https://doi.org/10.17226/25657.

National Commission on Fiscal Responsibility and Reform. (2010). *The Moment of Truth*: *Report of the National Commission on Fiscal Responsibility and Reform*. Washington, DC: The White House. Available: https://www.ssa.gov/history/reports/ObamaFiscal/TheMomentofTruth12_1_2010.pdf

National Research Council. (2002). *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*. Washington, DC: The National Academies Press. Available:

https://www.nap.edu/catalog/10131/at-what-price-conceptualizing-and-measuring-cost-of-living-and.

National Research Council. (2010). *Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement*. Washington, DC: The National Academies Press.

National Research Council. (2013). *Measuring What We Spend: Toward a New Consumer Expenditure Survey*. Washington, DC: The National Academies Press

O'Donnell, G., and Yélou, C. (2021). *Adjusted Price Index and Monthly Adjusted Consumer Expenditure Basket Weights*. Catalogue No. 62F0014M, Ottawa: Statistics Canada. Available: https://www150.statcan.gc.ca/n1/pub/62f0014m/62f0014m2021016-eng.htm.

Office for National Statistics (ONS). (2020). New Index Number Methods for Consumer Price Statistics. Paper presented at the UNECE Meeting of the Group of Experts on Consumer Price Indices, June 2. https://www.ons.gov.uk/economy/inflationandpriceindices/articles/newindexnumbermethodsinconsumerpricestatistics/2020-09-01/pdf.

Organisation for Economic Co-operation and Development (OECD). (2020). Methodological Notes: Compilation of G-20 Consumer Price Index. Available: https://www.oecd.org/sdd/prices-ppp/CPI-G20-methodology.pdf.

Orhun, A.Y., and Palazzolo, M. (2019). Frugality is hard to afford. *Journal of Marketing Research, 56*(1): 1–17.

Pakes, A. (2003). A reconsideration of hedonic price indexes with an application to PC's. *American Economic Review, 93*(5): 1578–1596.

Pandya, V., and Laisy, C. (2020). Tracking the impact of Covid-19 on the online economy in real-time. Paper presented at the 8th IMF Statistical Forum: Measuring the Economics of a Pandemic. Available: https://www.imf.org/-/media/Files/Conferences/2020/8th-stats-forum/paper-costa-lasiy-losaunne-white-and-vivek-pandya.ashx

Parker, J.A., Souleles, N.S., and Carroll, C.D. (2015). The Benefits of Panel Data in Consumer Expenditure Surveys. Pp. 75–99 in *Improving the Measurement of Consumer Expenditures*, C.D. Carroll, T.F. Crossley, and J. Sabelhaus, editors. Studies in Income and Wealth, Volume 74. Chicago : The University of Chicago Press.

Passero, W., Garner, T.I., and McCully, C. (2015). Understanding the relationship: CE survey and PCE. Pp. 181–203 in *Improving the Measurement of Consumer Expenditures*, C.D. Carroll, T.F. Crossley, and J. Sabelhaus, editors. Studies in Income and Wealth, Volume 74. Chicago : The University of Chicago Press.

Poole, R., Ptacek, F., and Verbrugge, R. (2005). Treatment of Owner-Occupied Housing in the CPI. Office of Prices and Living Conditions, Bureau of Labor Statistics. Available: https://docplayer.net/9019370-Treatment-of-owner-occupied-housing-in-the-cpi.html.

Ptacek, F., and Rippy, D.A. (2013). *Owners' Equivalent Rent and the Consumer Price Index: 30 Years and Counting*. Beyond the Numbers, U.S. Bureau of Labor Statistics, Volume 2, No. 14, Washington, DC. Available: https://www.bls.gov/opub/btn/volume-2/pdf/owners-equivalent-rent-and-the-consumer-price-index-30-years-and-counting.pdf.

Rao, D.S.P. (1995). On the Equivalence of the Generalized Country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Comparisons. Working Paper No. 5, Centre for International Comparisons, University of Pennsylvania, Philadelphia.

Rao, D.S.P. (2004). The Country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power Parities in the ICP. Paper presented at the SSHRC Conference on Index Numbers and Productivity Measurement, June 30-July 3, Vancouver, Canada. Available: https://www.researchgate.net/publication/24119708_The_Country-Product-Dummy_Method_A_Stochastic_Approach_to_the_Computation_of_Purchasing_Power_Parities_in_the_ICP.

Rao, D.S.P. (2005). On the equivalence of the weighted country product dummy (CPD) method and the Rao system for multilateral price comparisons. *Review of Income and Wealth*, *51*(4): 571–580.

Rassier, D.G., Aten, B.H., Figueroa, E.B., Kublashvili, S., Smith, B.J., and York, J. (2021). Improved measures of housing services for the U.S. economic accounts. *Survey of Current Business*, *101*(5). Available: https://apps.bea.gov/scb/2021/05-may/0521-housing-services.htm.

Redding, S.J., and Weinstein D.E. (2020). Measuring aggregate price indices with taste shocks: Theory and evidence for CES preferences. *Quarterly Journal of Economics*, *135*(1): 503–560.

Reed, S.B., and Stewart, K.J. (2014). Why does BLS provide both the CPI-W and CPI-U? *Beyond the Numbers: Prices and Spending* 3(5): 1–5.

Reinsdorf, M.B. (1993). The effect of outlet price differentials on the U.S. Consumer Price Index. In *Price Measurement and Their Uses*, NBER Studies in Income and Wealth, M.F. Foss, M.E. Manser, and A.H. Young, editors. Chicago: University of Chicago Press.

Reinsdorf, M.B. (1996). Constructing Basic Component Indexes for the U.S. CPI from Scanner Data: A Test Using Data on Coffee. Bureau of Labor Statistics Working Paper 277. Available: https://www.bls.gov/osmr/research-papers/1996/ec960260.htm.

Reinsdorf, M.B. (1998). Divisia Indices and the Representative Consumer Problem. Paper presented at the Fourth Meeting of the International Working Group on Price Indices, Washington, DC, April 22–24. Available: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+4/$file/1998+4th+Meeting+-+Reinsdorf++Marshall+-+Divisia+Indexes+and+the+Representative+Consumer+Problem.pdf.

Reinsdorf, M.B. (2020). COVID-19 and the CPI: Is Inflation Underestimated? IMF Working Paper No. 2020/224. Available: https://www.imf.org/-/media/Files/Publications/WP/2020/English/wpiea2020224-print-pdf.ashx

Rippy, D.A. (2014). The First Hundred Years of the Consumer Price Index: A Methodological and Political History. Monthly Labor Review, U.S. Bureau of Labor Statistics, Washington, DC. Available: https://doi.org/10.21916/mlr.2014.13.

Romley, J.A., Dunn, A., Goldman, D., and Sood, N. (2019). Quantifying productivity growth in the delivery of important episodes of care within the Medicare program using insurance claims and administrative data. In *Big Data for Twenty-First-Century Economic Statistics*, K.G. Abraham, R.S. Jarmin, B.C. Moyer, and M.D. Shapiro, editors. Chicago: University of Chicago Press. Available:

https://www.nber.org/system/files/chapters/c14275/c14275.pdf.
Sabelhaus, J., Johnson, D., Ash, S., Swanson, D., Garner, T.I., Greenlees, J., and Henderson, S. (2015). Is the Consumer Expenditure Survey representative by income? Pp. 241–262 in *Improving the Measurement of Consumer Expenditures*, C.D. Carroll, T.F. Crossley, and J. Sabelhaus, editors. Chicago: University of Chicago Press.
Seiler, P. (2020). Weighting bias and inflation in the time of Covid-19: Evidence from Swiss transaction data. *Swiss Journal of Economics and Statistics, 156*(13): 96–115.
Shapiro, I., Shapiro M., and Wilcox D. (2001). Measuring the value of cataract surgery. Pp. 411–438 in *Medical Care Output and Productivity*, D.M. Cutler and E.R. Berndt, editors. Chicago: University of Chicago Press.
Sheiner, L., and Malinovskaya, A. (2016). Measuring Productivity in Healthcare: An Analysis of the Literature. Hutchins Center on Fiscal and Monetary Policy at Brookings, Washington, DC. Available: https://www.brookings.edu/wp-content/uploads/2016/08/hp-lit-review_final.pdf.
Silver, M. (2013). PPP estimates: Applications by the International Monetary Fund. Chapter 23 in *Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program—ICP*. Washington, DC: World Bank.
Smith, S., Newhouse, J., and Freeland, M. (2009). Income, insurance, and technology: Why does health spending outpace economic growth. *Health Affairs*, *28*(5): 1276–1284.
Song, X., Marder, W., Houchens, R., Conklin, J., and Bradley, R. (2009). Can a Disease-Based Price Index improve the Consumer Price Index? Pp. 329–368 in *Price Index Concepts and Measurement*, W.E. Diewert, J.S. Greenlees, and C.R. Hulten, editors. Chicago: University of Chicago Press.
Soumare, A. (2017). Shelter in the Canadian CPI. Statistics Canada Catalogue No. 62F0014F2017001. Available: https://www150.statcan.gc.ca/n1/en/pub/62f0014m/62f0014m2017001-eng.pdf?st=jJEU1Glm.
Statistic Canada. (2020). Housing price measurement: Methods and use of alternative data in the Canadian CPI. Presentation to the panel on December 15.
Statistics Canada. (2021). An Analysis of the 2021 Consumer Price Index Basket Update, Based on 2020 Expenditures. July 21. Available: https://www150.statcan.gc.ca/n1/en/pub/62f0014m/62f0014m2021011-eng.pdf?st=y81EKq_n.
Stockburger, A. (2021). Incongruous Expenditure and Index Reference Periods: BLS Perspective. Presentation at the UNECE CPI Expert Group Meeting, June 10. Available: https://unece.org/sites/default/files/2021-05/Session_5_US-BLS_Presentation.pptx
Summers, R. (1973). International price comparisons based upon incomplete data. *Review of Income and Wealth*, *19*(1): 1–16.
Szulc, B.J. (1964). Indices for multiregional comparisons. *Przeglad Statystyczny, 3*: 239–254.
Thomas, H., and Ayoubkhani, D. (2019). Investigating the use of approximate expenditure weights for web scraped data in consumer price indices. Paper presented at the 16th Meeting of the International Working Group on Price Indices, Rio de Janeiro, May 8–10. Available:

https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+16/$FILE/The%20use%20of%20approximate%20expenditure%20paper.pdf

Triplett, J.E. (1997). Measuring consumption: The post-1973 slowdown and the research issues. *Federal Reserve Bank of St. Louis Review, 79*(3): 9–42. Available: https://doi.org/10.20955/r.79.9-42.

Johnson, P. (2015). UK Consumer Price Statistics: A Review. United Kingdom Statistics Authority. Available: https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-ukconsumerpricestatisticsarevie_tcm97-44345.pdf.

van Kints, M., de Haan, J., and Webster, M. (2019). Utilizing big data and multilateral index methods to produce the Australian CPI: Theory, implementation, and empirical results. *Statistical Journal of the IAOS, 35*(3): 1–19.

van Loon, K. (2020). Scanner Data and Web-Scraping in the Belgian CPI. Presentation to the panel on October 7. Available: https://www.nationalacademies.org/event/10-07-2020/docs/D124958ED038610E68986C71BEC8EA6D97CBF5F39C35.

Webster, M., and Tarnow-Mordi, R.C. (2019). Decomposing multilateral price indexes into the contributions of individual commodities. *Journal of Official Statistics, 35*(2): 461–486.

White, A.G. (1999). Measurement biases in consumer price indexes. *International Statistical Review, 67*(3): 301–325.

Wijaya, A.P., and Mariyah, S. (2019). Study of Consumer Price Index Based on E-Commerce in Indonesia. Paper presented at the Asia-Pacific Economic Statistics Week, Bangkok, Thailand, June 17–21. Available: https://www.researchgate.net/publication/349038751_Study_of_Consumer_Price_Index_based_on_E-Commerce_in_Indonesia.

Wilcox, D. (2021). How to improve the measurement of housing costs in the CPI. Peterson Institute for International Economics, RealTime Economic Issues Watch. Available: https://www.piie.com/blogs/realtime-economic-issues-watch/how-improve-measurement-housing-costs-cpi.

Wimer C., Collyer S., and Jaravel X. (2019). The Costs of Being Poor: Inflation Inequality Leads to Three Million More People in Poverty. The Groundwork Collaborative, Center on Poverty & Social Policy, Columbia University. Available: https://www.povertycenter.columbia.edu/s/The-Costs-of-Being-Poor-CPSP-Groundwork-Collaborative-2019.pdf.

Zhen, C., Finkelstein, E.A., Karns, S., Leibtag, E., and Zhang, C. (2019). Scanner data-based panel price indexes. *American Journal of Agricultural Economics, 101*(1): 311–329.

Appendix
Biographical Sketches of Panel Members

Daniel E. Sichel (chair) is professor of economics at Wellesley College. He also serves as a research associate at the National Bureau of Economic Research, advisory committee member at the Bureau of Economic Analysis, and executive committee member for the Conference on Research in Income and Wealth. Sichel is a member of the American Economic Association and an international advisory board member for the International Productivity Monitor. He has been awarded several honors including the Indigo Prize in 2017 and the Kendrick Prize in 2010. His research interests and publications are in macroeconomics, economic growth, technology, and economic measurement. Sichel has a B.A. degree in economics and an M.P.P., both from the University of Michigan, and a Ph.D. in economics from Princeton University.

Ana M. Aizcorbe is a research economist at the Bureau of Economic Analysis (BEA), where she conducts research into price index and other measurement issues. Prior to this position, she served as BEA's chief economist, where she initiated a Health Satellite Account that allows for the identification of drivers underlying the cost of treating diseases. Aizcorbe also held positions as an ASA/NSF/BLS research fellow, staff economist at the Federal Reserve Board, visiting fellow at The Brookings Institution, and research economist in the Bureau of Labor Statistics. She has published the book *A Practical Guide to Price Index and Hedonic Techniques* and numerous articles on the theoretical issues underlying price measurement, with empirical applications to the high-technology and service sectors. Aizcorbe has a B.A. degree in economics from Georgetown University and a Ph.D. in economics from Boston College.

Jan De Haan is a senior researcher at Statistics Netherlands, the Dutch national statistical institute. Until January 2019, he was also a professor at the Delft University of Technology. He is a member of the steering committee of the International Working Group on Price Indices (Ottawa Group) and an elected member of the U.S. Conference on Research in Income and Wealth. In the past, he has served as an advisor to many statistical institutes, including Statistics Canada, the Australian Bureau of Statistics, and Statistics New Zealand, and is an affiliate of the University of New South Wales' Real Estate Initiative. His main research interest is economic measurement, with a focus on index numbers and applications to scanner data and web-scraped data, real estate, hedonic regression methods, and price and volume measurement of public services. He has published extensively in peer-reviewed journals. He has an M.A. degree in economics from the University of Amsterdam and a Ph.D. in economic measurement from Erasmus University Rotterdam.

W. E. Diewert is professor of economics at the University of British Columbia and the University of New South Wales. He has published over 120 papers in journals and 140 chapters in books. His main areas of research include duality theory, flexible functional forms, index number theory (including the concept of a superlative index number formula), the measurement of productivity, the measurement of property prices, the pure theory of international trade, and the calculation of excess burdens of taxation. He was awarded the Julius Shiskin Memorial Award for Economic Statistics in 2005. He is a founding member of two international groups that study measurement issues: the Ottawa Group on Prices and the Canberra Group on Capital Measurement. He currently serves as chair of the Statistics Canada advisory committee on prices. He has B.A. and M.A. degrees in mathematics from the University of British Columbia and a Ph.D. in economics from the University of California at Berkeley.

Lisa M. Lynch is provost and Maurice B. Hexter Professor of Social and Economic Policy at Brandeis University. She is also a member of the Federal Reserve Bank of New York's Economic Advisory Panel and a research associate at the National Bureau of Economic Research and at IZA Germany. She has researched and published extensively on the impact of technological change and organizational innovation on productivity and wages, the determinants of youth unemployment, and the school-to-work transition. Lynch has an A.B. from Wellesley College and an M.Sc. and Ph.D. in economics from the London School of Economics and Political Science.

Raven S. Molloy is assistant director of the Division of Research and Statistics at the Federal Reserve Board of Governors, where she oversees work related to residential and commercial mortgage credit conditions, real estate prices, and housing markets. Her primary fields of research are housing, and urban and labor economics, and she has written on topics including housing supply regulation, housing affordability and valuation, mortgage credit availability, migration, foreclosure, vacancy, and executive compensation. She serves on the editorial boards of the *Journal of Housing Economics*, *Journal of Urban Economics*, and *Regional Science and Urban Economics*. She is a committee member of the Women in Real Estate Network of the American Real Estate and Urban Economics Association and a fellow of the Weimer School of Advanced Studies in Real Estate and Land Economics of the Homer Hoyt Institute. She has a B.A. degree in economics and Asian studies from the University of Virginia and a Ph.D. in economics from Harvard University.

Brent R. Moulton is senior economist in the Statistics Department of the International Monetary Fund. He has spent 32 years working in federal economic statistics, serving as associate director for national economic accounts at the Bureau of Economic Analysis and as chief of price and index number research at the Bureau of Labor Statistics. He received the Julius Shiskin Award for leadership in implementing major innovations into the U.S. national accounts. Previously, he served as a member of the international advisory expert group on national accounts, which worked on the development of the updated international standards in the System of National Accounts 2008. He is the author of numerous articles on economic measurement including essential research on lower-level substitution bias in the Consumer Price Index that led to the adoption by BLS of the geometric mean formula. He has a B.A. degree in economics from Brigham Young University and a Ph.D. in economics from the University of Chicago.

Marshall B. Reinsdorf is senior economist at the International Monetary Fund, working in areas of statistical methodology, and is president of the International Association for Research in Income and Wealth. He was previously chief of the national accounts research group at the Bureau of Economic Analysis, a financial economist at the FDIC, and a research economist at the Bureau of Labor Statistics. He is an expert in macroeconomic statistics, including national accounts, prices, and productivity. He is an author of more than 40 articles on economics and statistics and an editor of two books on economic measurement. He has a Ph.D. in economics from the University of Maryland, College Park.

Laura Rosner-Warburton is senior economist and founding partner at MacroPolicy Perspectives, LLC, specializing in the areas of monetary policy and inflation. She began her career in the financial services industry, working as an economist for Barclays Capital, and in a policy analysis group in the Markets Division of the Federal Reserve Bank of New York. Her work on inflation spans a range of economic and market environments and benefits from a blend of macro and micro perspectives. In the past, she has contributed to the development and design of the New York Fed's Survey of Primary Dealers. She has a B.A. degree in economics from Columbia University.

Louise M. Sheiner is Robert S. Kerr Senior Fellow and policy director at the Hutchins Center on Fiscal and Monetary Policy at the Brookings Institution in Washington, DC. Prior to this position, she was senior economist at the board of governors of the Federal Reserve System. Her expertise covers a range of disciplines, including public finance, health economics, fiscal policy, public economics, and welfare. She has written widely on subjects such as health care spending, macroeconomic implications of aging, household spending, and Medicare. Her recent publications include *Should America Save for Its Old Age? Fiscal Policy, Population Aging, and National Saving*; *Generational Aspects of Medicare*; and *Demographics and Medical Care Spending*. She has an A.B. in biology, an M.A. degree in economics, and a Ph.D. in economics, all from Harvard University.

COMMITTEE ON NATIONAL STATISTICS

The Committee on National Statistics was established in 1972 at the National Academies of Sciences, Engineering, and Medicine to improve the statistical methods and information on which public policy decisions are based. The committee carries out studies, workshops, and other activities to foster better measures and fuller understanding of the economy, the environment, public health, crime, education, immigration, poverty, welfare, and other public policy issues. It also evaluates ongoing statistical programs and tracks the statistical policy and coordinating activities of the federal government, serving a unique role at the intersection of statistics and public policy. The committee's work is supported by a consortium of federal agencies through a National Science Foundation grant, a National Agricultural Statistics Service cooperative agreement, and several individual contracts.