Minding Your Ps and Qs:

Going from Micro to Macro in Measuring Prices and Quantities

by

Gabriel Ehrlich, John Haltiwanger, Ron Jarmin, David Johnson, and Matthew D. Shapiro*

December 2018 (rev December 31, 2018)

Abstract

Key macro indicators such as output, productivity, and inflation are based on a complex system across multiple statistical agencies using different samples and different levels of aggregation. The Census Bureau collects nominal sales, the Bureau of Labor Statistics collects prices, and the Bureau of Economic Analysis constructs nominal and real GDP using these data and other sources. The price and quantity data are integrated at a high level of aggregation. This paper explores alternative methods for re-engineering key national output and price indices using item-level data. Such re-engineering offers the promise of greatly improved key economic indicators along many dimensions.

*University of Michigan (gehrlich@umich.edu); University of Maryland and NBER (haltiwan@econ.umd.edu); U.S. Census Bureau (ron.s.jamin@census.gov); University of Michigan (johnsods@umich.edu); and University of Michigan and NBER (shapiro@umich.edu).
We acknowledge financial support of the Alfred P. Sloan Foundation and additional support from the Michigan Institute for Data Science. The results in this presentation are based on researchers own analyses calculated (or derived) based in part on data from the Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the view of the U.S. Census Bureau. We thank Jamie Fogel, Diyue Guo, Dyanne Vaught and Laura Zhao for superb assistance.

Key national economic statistics produced by the U.S. statistical agencies rely primarily on surveys of businesses and households. The current methods for measuring GDP were developed by Simon Kuznets and the National Bureau of Economic Research (NBER) in the mid-20th century. Many improvements have been made since then, but the basic concepts and methodology remains the same. We think we are at a point where measurement of key economic concepts like GDP will begin to change in critical ways. There are fundamental changes in the information available on households and businesses given the digitization of most transactions in the economy. For key sectors of the U.S. economy, digitized data offer the opportunity to improve the measurement of key national economic indicators while also drastically reducing the respondent burden on households and businesses from responding to a variety of surveys conducted across multiple statistical agencies. In addition, declining response rates and increasing costs of household and business surveys provide additional incentives to explore new source data for economic measurement.

The digitization of economic activity yields a vast and rapidly increasing pool of data, residing mostly within the private sector that can potentially be used for improving, enhancing, and in many respects re-inventing the way we measure two key building blocks of our national economic statistics: output and prices. The core motivation for our approach is that prices and quantities should be measured in an integrated and consistent manner at the micro and the macro levels. The current system of economic measurement does not do so and suffers from substantial limitations as a result.

The methods used to measure GDP in the U.S. are similar to those in other advanced economies, though data collection in the U.S. is divided across multiple agencies. To measure real output, the Census Bureau collects the source data for the numerator (revenue), the Bureau of Labor Statistics (BLS) collects the source data for the denominator (prices), and the Bureau of Economic Analysis (BEA) divides the revenue by price data to measure real output.

The constraints built into this system impose limitations on the quality and utility of the resulting data products. One important limitation is that it is extremely difficult for economic measurement to keep up with the changes in the structure of the economy. Changes in information technology have yielded a rapid turnover of goods as well as changes in the way that persons acquire goods and services. Rapid product turnover is an inherent challenge for price index measurement, which in turn impacts estimates of GDP and productivity growth. Some have speculated that mismeasurement underlies the post-2000 decline in measured productivity growth. While evidence supporting this hypothesis is limited, there is widespread concern that measures of productivity and real wage growth do not adequately capture improvements in goods and services.

Incorporating data from digitized transactions on a widespread basis has the potential to overcome these and related limitations. First, the approach we advocate implies that macro indicators can be generated in an internally consistent manner with underlying micro data so that micro-macro based statistics and research can become the norm. Second, data based on item-level transactions have the potential to enable the incorporation of rapid product turnover as goods are tracked at the product code and outlet levels. Item-level data on prices, quantities, and attributes of goods and services allow for new techniques to measure the value of new goods, new outlets, and quality change. Third, the dense nature of item-level data provides data with more granularity (e.g., by industry and geography or time). In addition to providing more detail, more granular data can provide distributional measures beyond means or totals. Fourth, processing of digitized source data from firms or aggregators could reduce the already short lags between reference periods and the preliminary release of official statistics. Fifth, digital data can provide comprehensive data covering a large fraction of sales and prices almost immediately. Hence, a new architecture for economic statistics could obviate the need for some of the substantial revisions of statistics that take place over the long intervals in the current system.

I. Starting with Retail

A. *Challenges Facing Current Methods*

Measurement of GDP is a complex multi-step process involving many components with source data from both statistical agencies and private sector sources.[1] We focus our attention on the expenditure side approach of measuring GDP, and in particular its personal consumption expenditure (PCE) component. Digitized data could also be used extensively for the other components of GDP, as well as on the income side, so our approach has wider applicability. Within the PCE component of GDP, we focus on expenditure on goods through the retail sector, with the goal of eventually incorporating additional components of PCE that are transacted digitally.

To illustrate the current decentralized approach to data collection, processing and provision, we consider the measurement of real GDP for the retail sector using the expenditure approach.[2] The U.S. Census Bureau collects revenue (sales) information from businesses at various frequencies and levels of detail using different surveys. These provide key inputs to measuring nominal output (nominal GDP). The BLS conducts surveys measuring prices to produce the Consumer Price Index (CPI). The CPI integrates multiple data collections: the Consumer Expenditure (CE) Survey for expenditure weights, the Telephone Point-of-Purchase

[1] Much of this discussion follows the BEA methodology papers http://www.bea.gov/methodologies/. We also draw the BLS Handbook of Methods http://www.bls.gov/cpi/cpi_methods.htm and the Census Bureau https://www.census.gov/retail/index.html.

[2] Figure A.1 in the online appendix provides an overview of the data and measures.

Survey (TPOPS) for location of expenditure, and then enumeration of prices from a probability sample of goods within locations.[3] Inflation measures are complex and subject to considerable sampling and non-sampling measurement errors both in terms of statistical methodology and challenges in integrating prices from BLS surveys and expenditure/output weights from BLS and Census surveys. Finally, the BEA measures real output by integrating price indices from BLS with the nominal output measures from Census and augmenting with additional data sources.[4] A notable feature of the current architecture is that data collection for spending (Census) and for prices (BLS) are largely independent.

An additional challenge is that response rates on the retail trade surveys have declined substantially over time. The response rate for the monthly retail trade survey has recently fallen to 52 percent in 2015, from 66 percent in 2009. Jarmin (2019) discusses challenges facing surveys and the urgency for finding alternative sources of data for economic measurement, especially for sectors such as retail where survey performance has declined sharply.

B. *Using Item-level Transactions Data for Economics Measurement: Nominal Revenue Indexes*

Recent research on the use of digitized data for retail trade focuses on improving price indices, either from use of scanner data or from web-scraped posted prices. The potential advantages of digitized data do not stop at improved price indices. Our approach aims to develop integrated nominal expenditure, price, and real expenditure indices from the same sources rather than focusing on price indices alone. Our research aims to combine data from multiple sources including individual retailers and aggregators. In this paper, we present indexes based on

[3] In 2020, the CE Survey will begin collecting data on the location of expenditure, replacing the TPOPS.
[4] For an early discussion of the challenges of using scanner data for price measurement, see BEA and the Census Bureau have been actively evaluating the use of commercial data (see Bostic, Jarmin and Moyer (2016)).

transactions data aggregated to the item-level across time and outlets from Nielsen Scanner data to produce prices, *P*, quantity, *Q*, and total revenue, *P*Q*.

The Nielsen Retail Scanner data (made available from the Kilts center) provides weekly item-level data on sales and units sold for a large panel of grocery stores and other mass merchandisers.[5] Since items are defined very narrowly (i.e., the UPC level), dividing sales by units sold gives a good measure of price. The ability to infer prices from unit values is central to approaches to *P* and *Q* measurement using scanner data.[6] We aggregate to monthly data using the National Retail Federation calendar. For current purposes, we aggregate items to the national level and further time-aggregate to a quarterly frequency.

Figure 1 depicts the quarterly level of nominal food sales from the scanner data compared to the Monthly Retail Trade Survey (MRTS) estimates for Grocery Stores and to nominal BEA Personal Consumption Expenditure (PCE) for Off-premises food and non-alcoholic beverages.[7] The figure shows index numbers with calendar year 2010=1. Despite their completely different source data, the scanner and the MRTS have very similar trends. They trend together for most of the period in the sample, though they diverge somewhat early and late in the sample. The PCE and MRTS have very similar trends, but PCE is based in part on MRTS, so this is the case by construction. The Nielsen data likely has some coverage changes early in the sample.

There are important differences in the data sources for the series that highlight the value of item-level transactions data for measuring nominal volumes. Census monthly and annual retail sales are measured across all retail establishments within a firm. Census monthly retail

[5] We have also analyzed the Nielsen Consumer Panel. Results for the indexes presented in this paper are similar (see Figures A.4 and A.5 in the online appendix.)

[6] Price is not exactly measured by unit value if there is within-week variation in prices and other complications in pricing.

[7] Figure A.2 in the online appendix displays the same information in growth rates.

sales are based on a relatively small sample of firms (13,000 for the entire retail trade sector), while the scanner data covers about 35,000 stores. Census retail sales at grocery stores include sales of many non-food items, but can exclude sales of food at, for example, general merchandise stores. In contrast, the scanner data, which we aggregated based on product codes, include only sales of food regardless of the type of outlet and contain information on more than 650,000 item-level products per month. The different seasonality of the Census MRTS data on grocery stores relative to the scanner data (see Figure A.3 in the online appendix) likely reflects the non-food items at grocery stores.

The advantages of item-level data that yield detailed product class information at high frequency are highlighted when one considers estimates of PCE. The high-frequency data underlying PCE come from the MRTS, which as we have seen, provides estimates by type of outlet, not by product. Every five years the Economic Census (EC) yields information on sales at the establishment level by detailed product class. In the intervening time periods, the Annual Retail Trade survey (ARTS) and the MRTS survey firms for their total sales, classifying firms into major kind of business (e.g., Food and Beverage or Grocery Stores). The revenue growth and quantity indices developed by BEA using the integrated data from Census and BLS require extrapolating the detailed EC information at the product class level with the more current information by outlet type from the ARTS and MRTS. The EC uses an annual reference period, so it provides the BEA no information on the within-year composition of products sold by outlets. Thus, the EC provides no information for the BEA to produce non-seasonally adjusted PCE at the detailed goods level at high frequencies. Hence, BEA uses within-year composition information from scanner data in combination with the PCE reported in Figure 1 to produce statistics on more detailed food products (e.g., poultry).

This example highlights the extrapolative nature of high-frequency GDP estimation given the current architecture. Data users might not be too concerned about the fact that GDP statistics abstract from the shifting seasonal mix of goods sold by grocery stores. But the same issue will apply at business cycle frequency and for business cycle shocks, with the potential for the current system to either overstate or understate cyclical fluctuations depending on the product mix across outlets and their cyclicality sensitivity.

C. *Using Item-level Transactions Data for Economics Measurement: Price Indices*

Both BLS and BEA, along with some international statistical agencies, have already begun to incorporate item-level transactions scanner data in their price indices. To provide some perspective on the relationship between the current methods and using scanner data, Figure 2 compares the BLS CPI for the product groups covered by the Nielsen scanner data to a variety of price indices computed from the scanner data.[8] The scanner indices in Figure 2 are constructed from quarterly unit prices and expenditure shares covering more than 100 product groups. Figure 2 reports four-quarter averages of the quarterly indices of price change (measured using the log differences of the indices). The number of item-level price quotes monthly in the BLS CPI for these product groups is about 40,000, compared to the 650,000 item-level prices in the scanner data.

The top panel shows the results for the food product groups and the lower panel the nonfood product groups. The price indices from the scanner data are computed at the product group level, and then Divisia expenditure share weights by product groups are used to aggregate to the food and nonfood indices. Each panel displays three indices calculated from the scanner

[8] We thank the BLS for producing food and nonfood CPI indices using the product groups in the Nielsen data. The BLS data provided should be interpreted with care because they do not meet BLS's standard publication criteria. Figure 2 uses the retail scanner data. Broadly similar patterns are exhibited by the consumer panel data. See Appendix Figures A.4 and A.5.

data: a Laspeyres index; the constant elasticity of substitution (CES) demand-based price index with the adjustment for product turnover proposed by Feenstra (1994, hereafter Feenstra); and the Unified Price Index (UPI) proposed by Redding and Weinstein (2018). The Feenstra and UPI are modified price indices based upon a CES expenditure function approach that incorporates (i) product quality changes from product turnover (Feenstra and UPI) and (ii) product quality/appeal from what Redding and Weinstein denote the "consumer valuation bias" (UPI only).

Computing the Feenstra and UPI requires elasticities of substitution, which we estimate for each product group using the Feenstra (1994) method applied to the item-level data following Redding and Weinstein (2018). To calculate the Laspeyres index using the item-level data, we use previous-quarter expenditure weights updated for each quarter.

For food, the average rate of price change using the BLS CPI is very similar to (albeit slightly lower than) the Laspeyres index from the scanner data, and the two price indices track each other well (with a correlation about 0.97). The Feenstra shows a notably lower average price change and a correlation with the CPI of 0.98. The UPI has a much lower average and a correlation with the CPI of 0.78. The finding that the CPI and the Laspeyres from scanner data track each other so well is reassuring, but also not surprising given that the quality adjustments used in the CPI for Food are modest at best. The close relationship between the CPI and Laspeyres for food provides a benchmark to gauge the impact of the quality adjustments via Feenstra and UPI, which like the Laspeyres use the scanner data.

The lower panel shows greater differences across price indices for nonfood. Here the CPI inflation rate is slightly higher than the scanner Laspeyres, but their correlation is weaker (0.54). The Feenstra has a substantially lower mean and the UPI a much lower mean. The CPI's correlation with the Feenstra is 0.37 and with the UPI is negative (-0.53). The larger gap across price indices for nonfood than for food is consistent with the hypothesis that quality adjustments from product turnover and changes in product appeal for continuing goods (i.e., consumer valuation) are likely to be more important for nonfood. Also consistent with that hypothesis, there is a more substantial gap between the Feenstra and UPI than there is between the Laspeyres and Feenstra.

The results suggest the CPI is missing substantial quality adjustments, especially for nonfood. Appropriate caution is required in drawing this inference because both the Feenstra and UPI require specification of a utility function and estimates of the elasticity of substitution parameters. Although estimating the elasticities at a product group level (e.g., carbonated beverages for food and electronic products for nonfood) permits allowing for over 100 different elasticities within the scanner data, this may still be a very high level of aggregation. Within scanner's product groups are arguably goods that are very close substitutes, while others are more differentiated. For product turnover and expenditure share volatility with close substitutes, the appropriate quality adjustment factors in Feenstra and UPI become very small. The procedure used in Figure 2 is to assume the same elasticity of substitution for all products within a product group.[9]

An alternative approach is to use attribute data with transactions data on prices and quantities to create hedonic price indices. Hedonic methods (e.g. Pakes (2003) and Bajari and Benkard (2005)) can be used with price, quantity, and attribute data, but the practicality of using the hedonic approach at scale remains to be seen.[10] The BLS implements hedonics on a limited

[9] A nested CES approach within product groups has the potential to overcome these aggregation issues, but it brings its own challenges (see, e.g., Hottman, Redding and Weinstein (2016)).
[10] The relationship between quality adjusted price indices using the UPI and hedonic approaches is explored in much more detail in Ehrlich et. al. (2019).

basis with careful attention to the measurement of attributes for products. Although the expenditure function approaches promise to overcome the scale issues, they rely on an appropriate structure of grouping and nesting products. The problem of finding the correct grouping or nesting has elements of the "house to house combat" of conventional approaches to quality adjustment (see Shapiro and Wilcox (1996)), and the solution may rely on similar measurement of attributes to justify the classification structure of items. Machine learning methods could potentially be used with the rich text and image data on products that are increasingly available in digital format for either estimating hedonic models of specifying the nests for expenditure functions.

II. Re-Engineering Key Economic Indicators

Given the availability of item-level transaction data and the advantages they present relative to survey data along multiple dimensions, the time is ripe for re-engineering the data collection and measurement of key economic indicators such as real output and inflation. Beyond the conceptual challenges discussed in this paper, there are many practical challenges to address in finding ways to tap into the item-level data in a manner that is cost effective for both businesses and the statistical agencies. Getting buy-in from companies to harvest their granular data is a key challenge. An open question is whether information aggregators (such as Nielsen) are more desirable means of tapping into this data or alternatively whether "harvesting" apps could be developed for firms to implement on their individual data platforms.

References

Bajari, Patrick and C. Lanier Benkard (2005). "Hedonic Price Indexes With Unobserved Product Characteristics, and Application to Personal Computers." *Journal of Business and Economic Statistics* 23(1), 61-75.

Bostic, William, Ron Jarmin, and Brian Moyer (2016). "Modernizing Federal Economic Statistics." *American Economic Review Papers and Proceedings* 106(5).

Ehrlich, Gabriel, John Haltiwanger, Ron Jarmin, David Johnson, and Matthew D. Shapiro (2019). "Re-engineering Key National Indicators." working paper.

Feenstra, Robert C. (1994). "New Product Varieties and the Measurement of International Prices." American Economic Review, 84(1), 157-177.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. (2016). "Quantifying the Sources of Firm Heterogeneity." *Quarterly Journal of Economics* 131(3), 1291-1364.

Jarmin, Ron (2019). "Evolving Measurement for an Evolving Economy: Thoughts on 21st Century US Economic Statistics." *Journal of Economic Perspectives*, forthcoming.

Pakes, Ariel (2003). "A Reconsideration of Hedonic Price Indexes with an Application to PCs." *American Economic Review* 93(5) 1578-1596.

Redding, Stephen, and David Weinstein (2018). "A Unified Approach to Estimating Demand and Welfare." Unpublished paper, Columbia University.

Shapiro, Matthew D. and David W. Wilcox (1996). "Mismeasurement in the Consumer Price Index: An Evaluation." *NBER Macroeconomics Annual* 11, 93-142.

Figure 1. Nominal Sales of Food: Scanner, Census Retail Sales, and BEA PCE

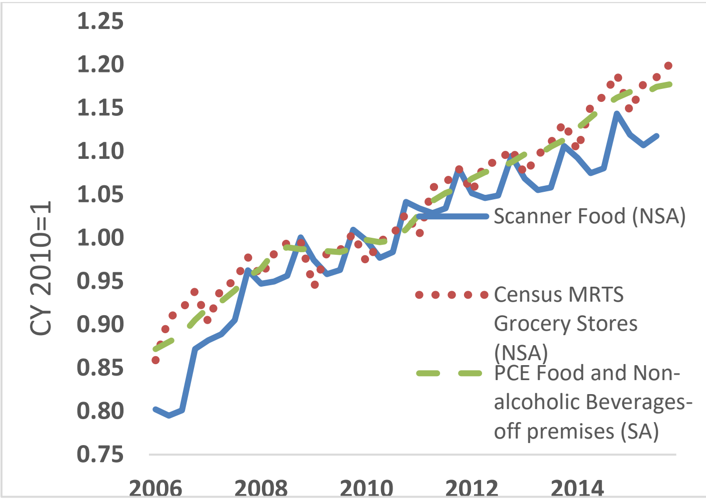
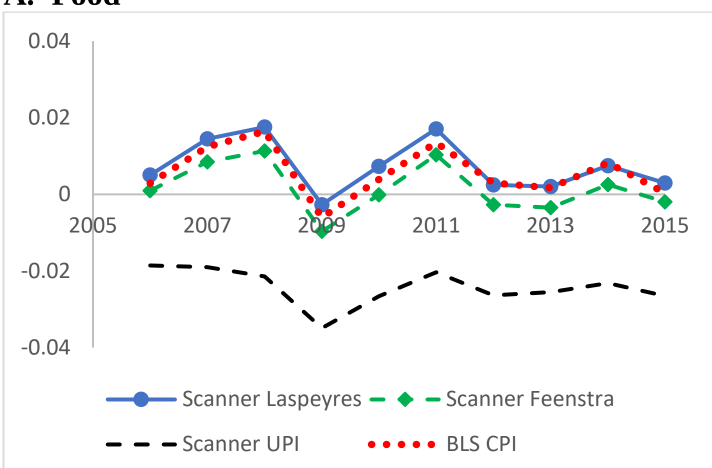
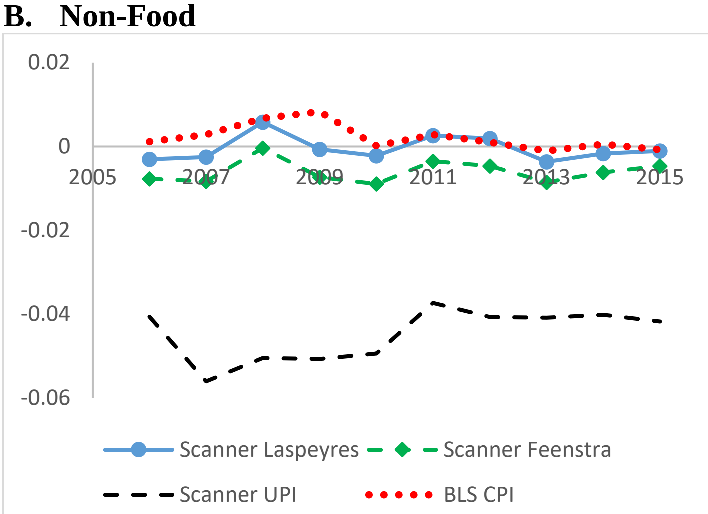


Figure 2. BLS CPI and Nielsen-Scanner Based Price Indices (Annual Averages of Quarterly Changes)

A. Food



B. Non-Food



Online Appendix

Minding Your Ps and Qs:

Going from Micro to Macro in Measuring Prices and Quantities

by

Gabriel Ehrlich, John Haltiwanger, Ron Jarmin, David Johnson, and Matthew D. Shapiro*

December 2018 (rev December 31, 2018)

Figure A.1. Measuring Real and Nominal Consumer Spending: Current Architecture

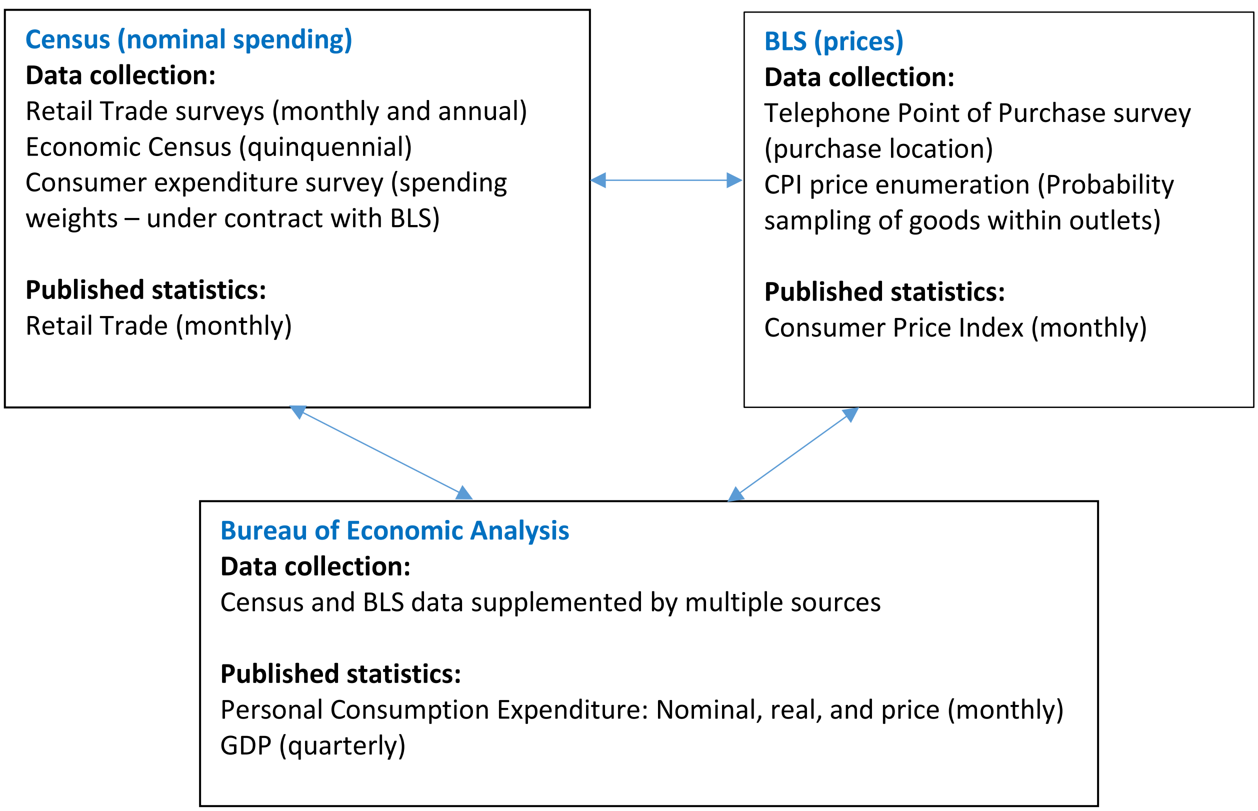


Figure A.2. Quarterly Growth Rates of Nominal Sales of Food: Scanner, Census Retail Sales, and BEA PCE

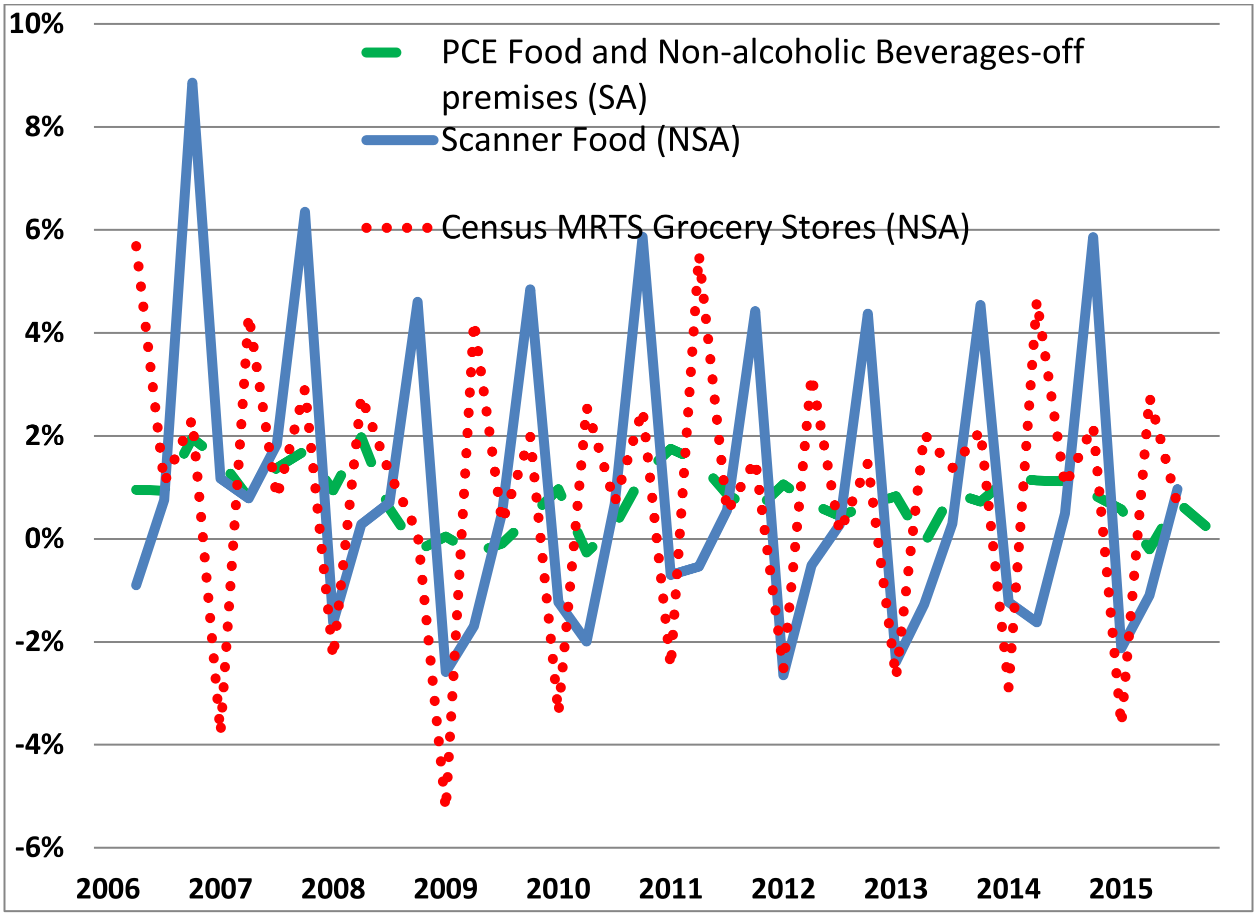
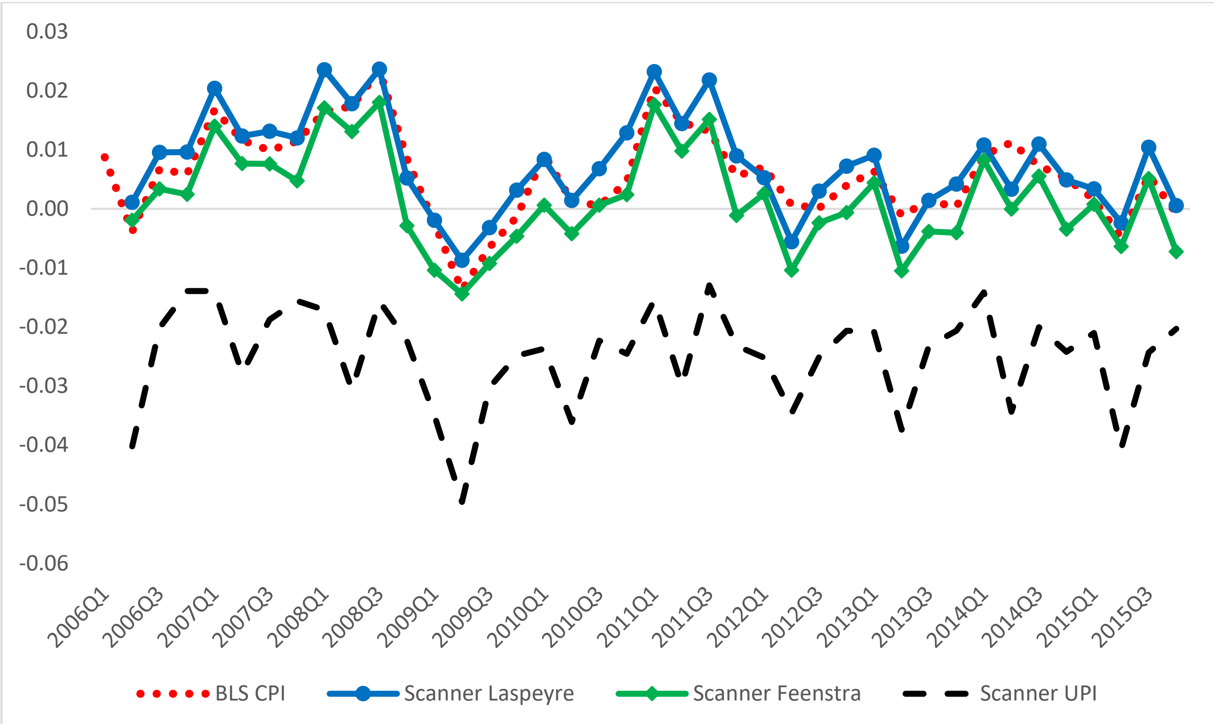


Figure A.3. Comparisons of BLS CPI and Nielsen-Scanner Based Price Indices, Quarterly Changes

A. Food



B. Non-Food

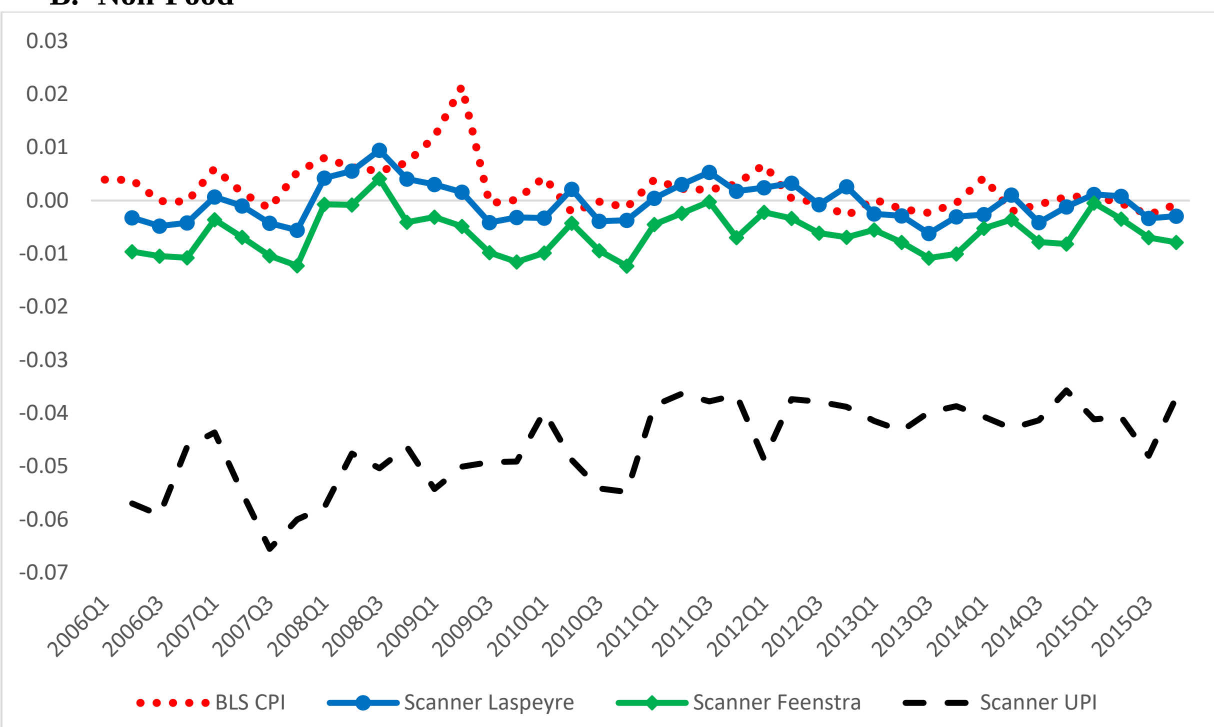
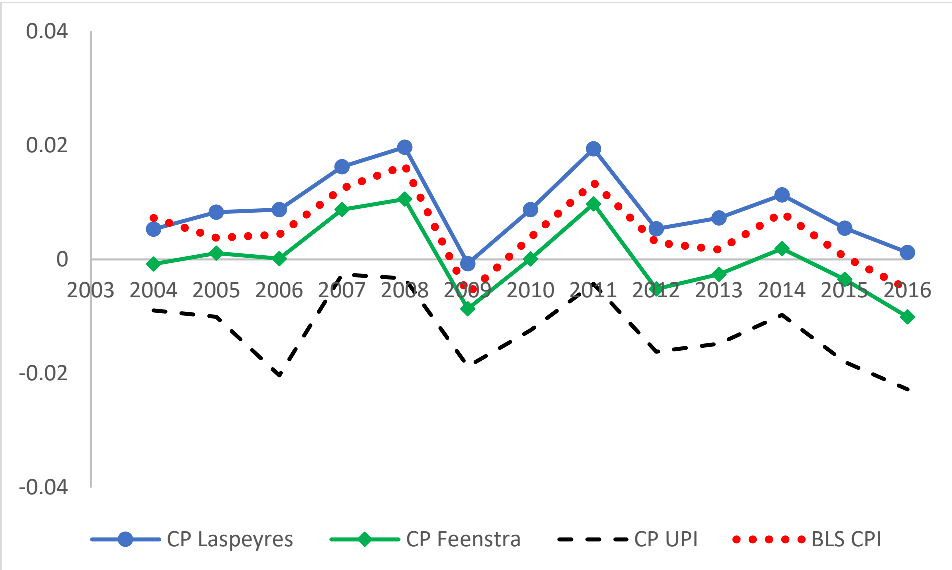


Figure A.4. Comparisons of BLS CPI and Nielsen-Consumer Panel (CP) Based Price Indices, Annual Averages of Quarterly Changes

A. Food



B. Non-Food

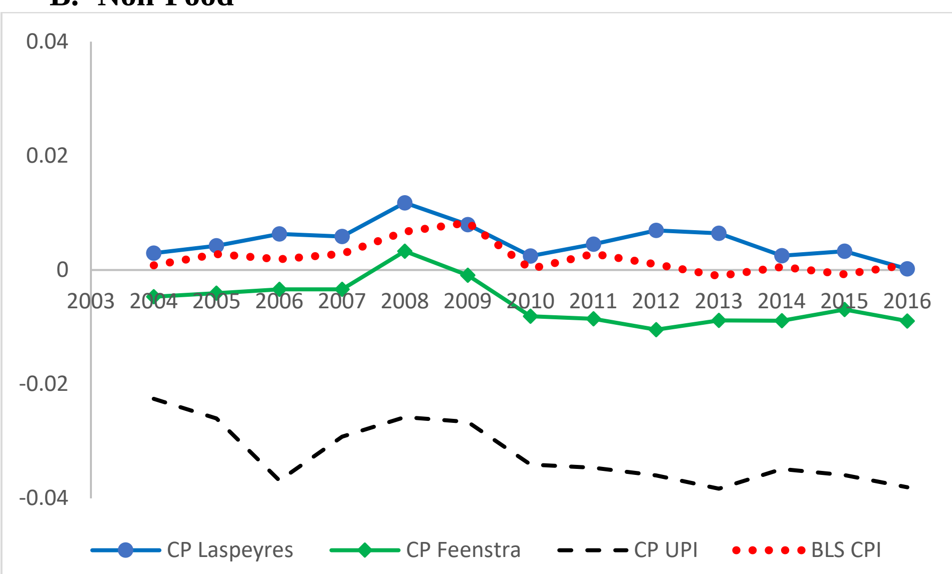
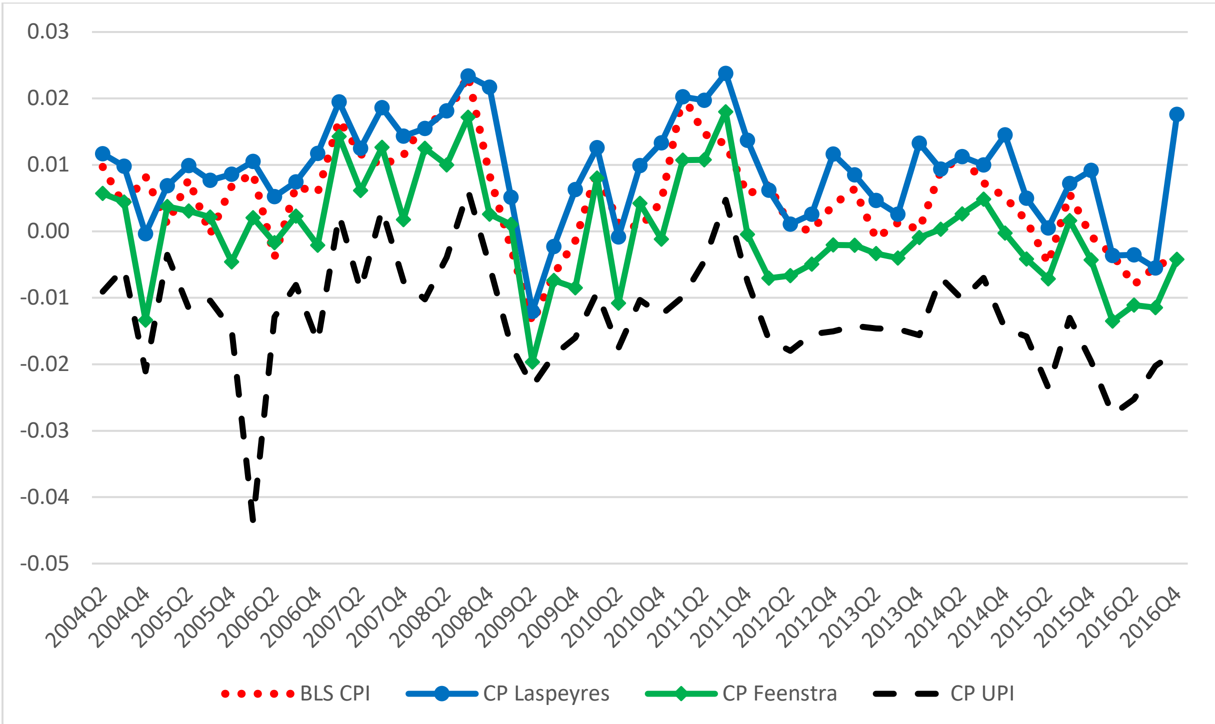


Figure A.5. Comparisons of BLS CPI and Nielsen-Consumer Panel (CP) Based Price Indices, Quarterly Changes

A. Food



B. Non-Food