

NBER WORKING PAPER SERIES

RE-ENGINEERING KEY NATIONAL ECONOMIC INDICATORS

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

Working Paper 26116
<http://www.nber.org/papers/w26116>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

July 2019, Revised December 2020

We acknowledge financial support of the Alfred P. Sloan Foundation and the additional support of the Michigan Institute for Data Science and the Michigan Institute for Teaching and Research in Economics. The results here are in part 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. We also use the NPD data housed at the U.S. Census Bureau. All results using the NPD data have been reviewed to ensure that no confidential information has been disclosed (CBDRB-FY19-122). 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 Katharine Abraham, Robert Cage, Robert Feenstra, David Friedman, Greg Kurtzon, Robert Martin, Stephen Redding and David Weinstein for helpful comments. We thank Jamie Fogel, Diyue Guo, Edward Olivares, Luke Pardue, Dyanne Vaught, and Laura Zhao for superb research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Gabriel Ehrlich, John Haltiwanger, Ron Jarmin, David Johnson, and Matthew D. Shapiro. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Re-engineering Key National Economic Indicators

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

NBER Working Paper No. 26116

July 2019, Revised December 2020

JEL No. C81,C82,E31

ABSTRACT

Traditional methods of collecting data from businesses and households face increasing challenges. These include declining response rates to surveys, increasing costs to traditional modes of data collection, and the difficulty of keeping pace with rapid changes in the economy. The digitization of virtually all market transactions offers the potential for re-engineering key national economic indicators. The challenge for the statistical system is how to operate in this data-rich environment. This paper focuses on the opportunities for collecting item-level data at the source and constructing key indicators using measurement methods consistent with such a data infrastructure. Ubiquitous digitization of transactions allows price and quantity be collected or aggregated simultaneously at the source. This new architecture for economic statistics creates challenges arising from the rapid change in items sold. The paper explores some recently proposed techniques for estimating price and quantity indices in large-scale item-level data. Although those methods display tremendous promise, substantially more research is necessary before they will be ready to serve as the basis for the official economic statistics. Finally, the paper addresses implications for building national statistics from transactions for data collection and for the capabilities and organization of the statistical agencies in the 21st century.

Gabriel Ehrlich
Department of Economics
University of Michigan
611 Tappan St
Ann Arbor, MI 48109-1220
gehrlich@umich.edu

David Johnson
ISR
426 Thompson, Rm 3234
Ann Arbor, MI 48106
United States
johnsods@umich.edu

John Haltiwanger
Department of Economics
University of Maryland
College Park, MD 20742
and NBER
haltiwan@econ.umd.edu

Matthew D. Shapiro
Department of Economics
University of Michigan
611 Tappan St
Ann Arbor, MI 48109-1220
and NBER
shapiro@umich.edu

Ron Jarmin
U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233
ron.s.jarmin@census.gov

Statistical agencies face multiple challenges in the present environment. Traditional methods of collecting data—whether asking businesses or individuals to complete surveys, gathering price data by sending enumerators to stores—face increasing challenges.¹ These include declining response rates to surveys, increasing costs to traditional modes of data collection, and, perhaps most importantly, the difficulty of keeping pace with rapid changes in the economy. The information technology revolution is dramatically changing how and where consumers and businesses carry out their transactions. Consumers shop online, summon cars for hire with an app, watch “TV” without television stations or TVs, and “bank” without cash or checks. These technologies are leading to widespread changes in industrial structure and the organization of markets, with implications for prices and real output that the official economic statistics may fail to capture.

The good news for economic measurement is that these transactions inherently create huge amounts of data precisely because they are driven by information technology. Determining how to operate in this data-rich environment is therefore both a major challenge and a great opportunity for the statistical system. The information economy calls for more than using new technologies and new sources of data to improve on existing approaches to data collection. Instead, now is a promising time to explore reengineering the system of national statistics, specifically the National Income and Product Accounts (the NIPAs, which include GDP), productivity and consumer and producer price measurement, by collecting specific product data at source, or as close to the source as is feasible. In particular, we advocate that price and quantity be collected or aggregated simultaneously from retailers.

¹ Ehrlich, Haltiwanger, Jarmin, Johnson, and Shapiro (2019) gives a short introduction to some of the arguments and results presented in this paper.

Before sketching how such a new infrastructure might look, we first briefly describe how the NIPAs and price indices are currently assembled. We focus on consumer spending and prices, but similar issues apply across other components. In brief, nominal sales are collected by the Census Bureau, prices are collected by the Bureau of Labor Statistics (BLS), and real and nominal GDP are constructed by the Bureau of Economic Analysis (BEA) using these and other data sources.

A key point to understand is that prices and sales are currently based on different samples and levels of aggregation. Measurements of retail sales and the prices used to deflate them are not matched at the outlet level, let alone at the item level. A similar mismatch of price and nominal variables pervades the productivity data, in which industry-level producer price indices are used as deflators. This generates great challenges for micro productivity analysis but also is problematic for the industry-level indices.

The information technology revolution brings huge opportunities for replacing this multi-layered, multi-mode, multi-agency methodology with a unified approach to collecting price and quantity information simultaneously at the source. Retail transactions—whether online or at brick and mortar stores—ubiquitously create a record of the sale at the item level. Individual items are defined finely enough—barcode or SKU—that price can be calculated simply by dividing the nominal value of the sale by the quantity sold. Other sectors also increasingly have digitized transactions-level data.

Transactions-level information summarized to the item level should, in principle, allow the production of greatly improved statistics. First, price and quantity could be based on the same observations. Second, the granularity of the data along multiple dimensions could be greatly increased. Statistics could be constructed at a very fine level of geographical detail.

Similarly, product-level detail could be greatly refined. Third, time series could, in principle, be constructed at any frequency—yearly, monthly, weekly, daily, or even hourly. The daily data would be particularly helpful for dealing with “seasonality” relating to trading days and holidays and how that seasonality interacts with paydates. Data could, in principle, be available with a very short lag. Using all transactions rather than a sample should greatly reduce sampling error and data revisions. Additionally, improved measurement of price change and quantities would directly affect the quality and detail of measurement of productivity.

Implementing such a new architecture for measuring economic activity and price change is not, however, without considerable challenges. Our paper explores three general areas of such challenges relating to (1) measurement, (2) data access, and (3) the capabilities and mandates of the statistical agencies. First, consider the measurement challenges. Confronted with the firehose of newly-available data, the economist or official statistician is confronted quickly with a case of “be careful what you wish for.” There are technical and computing challenges for dealing with the volume of data. The statistical system will need to learn from best practices in computer and data science and in business to process the data at scale. Moreover, because the data are created for tracking transactions and other internal purposes, they are not organized for the convenience of official statisticians. In contrast, official statistics are often based on surveys where businesses and households are asked to fit their answers into the paradigm of economists and statistical agencies’ nomenclature. That makes such designed data convenient for official statisticians, but potentially difficult and costly for respondents to prepare. With naturally occurring data, the statistical agency needs to transform the data to suit its purpose. This shift of burden from respondent to agency will be costly, but if done right, can improve data quality because it will reduce reliance on getting accurate responses from businesses and individuals

who might lack incentives for giving accurate responses and may not understand what is being asked.

A related practical measurement and conceptual challenge is that there is enormous turnover of goods. Roughly speaking, if a good defined at the barcode or SKU level is sold today, there is only a fifty percent chance it will be sold a year from today. This turnover of goods is one of the greatest challenges of using the raw item-level data for measurement, but also is an enormous opportunity. When new goods replace old goods there is frequently both a change in price and quality. Appropriately identifying whether changing item-level prices imply changes in the cost of living or instead reflect changes in product quality is a core issue for measuring activity and inflation. The statistical agencies currently perform these judgments using a combination of direct comparisons of substitutes, adjustments, and hedonics that are very hard to scale. Hence, new techniques will be needed to implement quality adjustment at scale. Luckily, such techniques—leveraging the resource made available by big data—may now be coming available.

We explore and compare two proposed approaches to measuring prices and real quantities using item-level data. The first is the Unified Price Index (UPI) approach proposed by Redding and Weinstein (2018, 2020), who build on the traditional Feenstra (1994) product turnover-adjusted Sato-Vartia price index.² This approach requires sales and price (i.e., price and quantity data) at the individual product level. Redding and Weinstein’s results suggest that

² We use the terminology UPI in this paper since it is the terminology of Redding and Weinstein (2018). In their revised paper (Redding and Weinstein (2020)) they denote this price index as the CUPI. This alternative naming convention reflects the CES demand structure underlying this price index. We use the UPI naming convention because our approach more closely follows that of what will call the theoretical UPI in Redding and Weinstein (2018). Redding and Weinstein (2020) implement some changes (to a version we call the seasoned UPI based on our discussant’s comments) to address issues of the slow rollout of goods post-entry and the slow exit process of goods before final exit. We are sympathetic to these issues but as discussed below we think that more research is needed to understand these issues. Using the theoretical UPI permits us to draw out the issues.

traditional (e.g., Paasche, Laspeyres, Sato-Vartia) indices typically miss important components of quality change. The second approach that we explore is the possibility of doing hedonics at scale in the spirit of, e.g., Benkard and Bajari (2005). Such hedonic approaches use the attributes that are available in retailers' information systems or can be scraped from the web. These attributes can include the standard hedonic covariates (size, color) or non-standard data such as images.

One key lesson from our explorations is that, despite these methods' elegance and ingenuity, there are many practical challenges and nuances involved in implementing them at scale and in interpreting the results. We believe more research is necessary to reach consensus on many of these issues before these methods can serve as the basis of official statistics. Indeed, both of these methods are the subject of active research by the academic and statistical communities. Given that research is actively evolving, we expect new developments over the coming years. This paper both examines and advances this evolving state of the art. Digging into the details of these active research agendas helps to reveal the challenges and opportunities of working with the price and quantity data in this context.

The first section of this paper reviews the existing paradigm in which economic statistics are built from disparate sources—often starting with source data that is already substantially aggregated (e.g., firm-level sales) and combining price measurement from samples independent from the nominal values. It also requires substantial interpolation and extrapolation to produce higher-frequency time-series benchmarked to detailed data that are collected infrequently.

Building key national indicators from item-level transactions data requires re-engineering how data are collected and accessed for official statistics. A new architecture for data collection is a requisite for implementing the procedures studied in this paper at scale. The logic and

logistics of building economic statistics from the ground up mandates that there be entirely new procedures for data collection that lever the information systems that already exist in business. In the paper we discuss alternative modes for capturing the data from business information systems. These include direct feeds of transactions data from businesses to agencies, the use of applications interfaces that produce business-level statistics from transactions data that are transmitted to agencies, and the use of commercial data aggregators.

Transactions-level data are sensitive commercial information. Statistical agencies are already gathering sensitive information at the establishment and firm levels on a regular basis and providing privacy and confidential protection for such data. Part of re-engineering the data collection process will involve modifying protocols to assure the continued high level of protection of confidential information and the confidence that information will be used for statistical purposes only is maintained in a modernized architecture built around digitized transactions data.

Finally, to implement this new architecture, there would have to be changes in the organization and capabilities of the statistical agencies. The simultaneous collection of price and quantity data requires combining the data collection activities that are now spread over multiple agencies. The agencies are already undertaking major initiatives to use transactions data to supplement or replace data collected by surveys or enumerations. Yet, these are largely efforts to replace data streams within the existing architecture, so do not create the improvements to measurement of economic activity envisioned in this paper.

The agencies would also need staff with the expertise to do this type of work, which lies at the intersection of data science, economics, and statistics. We recognize that this proposed new architecture for official statistics would be costly to implement. Substantial R&D would be

necessary to put these innovations into action. The current system would have to run in parallel for a period of time to allow consistent time series to be published. While the agencies are already taking steps in these directions, a wholesale reengineering would take a high-level commitment to change and commensurate funding during the transition period. Given the promise of improved data quality together with the potential for lower long-run cost, it is essential to undertake these investments now. Indeed, without them, we risk deterioration of the quality of statistics as response rates continue to erode and the cost of business as usual continues to outpace agency budgets.

Our paper provides an overview of this re-engineering approach, including a discussion of the issues and challenges mentioned above. We also argue, and provide evidence, that while the challenges are great, there are reasons to be optimistic that practical implementation of many components of this approach is relatively close at hand. We provide examples of the implementation of this approach using item-level data for the retail trade sector. Our examples highlight that the data are already being generated and the computational capacity to undertake this approach is readily available.

I. Existing Architecture

Table 1 summarizes the source data and statistics produced to measure real and nominal consumer spending.³ A notable feature of the current architecture is that data collection for total

³ Table 1 is an oversimplification of how economic statistics in general, and the NIPA in particular, are produced. The simplification that Census collects nominal sales, BLS collects prices, and BEA uses them to produce price and quantity is a useful one. This simplification is broadly accurate as a portrayal of the current architecture and conveys why it cannot accommodate the measurement innovations that this paper addresses. Nonetheless, it is important to recognize that each agency does multiple data collections that contribute to the real and nominal national accounts in complex ways. For example, the BLS Housing Survey for rents and owner equivalent rents that enter the NIPA. BEA collects data on prices and transactions from multiple sources to produce the NIPA. The agencies have made substantial strides to bring in new sources of data for official statistics, a number of which are presented in papers in this volume. Nonetheless, as discussed in Sections IV and V of this paper, these efforts are largely aimed at improving measurement within the current paradigm and therefore do not generally lever the advantages of simultaneous collection of price and quantity as advanced in this paper.

retail sales (Census) and for prices (BLS) are completely independent. The consumer price index program collects prices based on (1) expenditure shares from the Consumer Expenditure Survey (BLS manages the survey and Census collects the data), (2) outlets selected based on the Telephone Point of Purchase Survey, and (3) a relatively small sample of goods at these outlets that are chosen probabilistically (via the Commodities and Services Survey). The Census Bureau collects sales data from retailers in its monthly and annual surveys. The monthly survey is voluntary and has suffered from declining response rates. In addition, the composition of the companies responding to the monthly survey can change over time, which complicates producing a consistent time series. Store-level sales data are only collected once every five years as part of the Economic Census.

Integration of nominal sales and prices by BEA is done at a high level of aggregation that is complicated by the availability of product class detail for nominal sales that is only available every five years from the Economic Census. In the intervening periods, BEA interpolates and extrapolates based on the higher frequency annual, quarterly, and monthly surveys of nominal sales by the Census Bureau. These higher frequency surveys are typically at the firm rather than establishment level. Moreover, they classify firms by major kinds of business. For example, sales from the Census Monthly Retail Trade Survey (MRTS) reflect sales from “Grocery Stores” or “Food and Beverage Stores.” Such stores (really firms) sell many items beyond food and beverages, complicating the integration of the price indices that are available at a finer product class detail.

This complex decentralized system implies that there is limited granularity in terms of industry or geography in key indicators such as real GDP. BEA’s GDP by industry provides series for about 100 industries, with some 4-digit (NAICS) detail in sectors like manufacturing,

but more commonly 3-digit and 2-digit NAICS detail. The BEA recently released county-level GDP on a special release basis, a major accomplishment. However, this achievement required BEA to integrate disparate databases at a high level of aggregation with substantial interpolation and extrapolation. Digitized transactions data offer an alternative building up from micro data in an internally consistent manner.

II. Using Item-level Transactions Data

In the results presented here, we focus on two sources of transactions data summarized to the item level. One source is Nielsen retail scanner data, which provides item-level data on expenditures and quantities at the UPC code-level for over 35,000 stores, covering mostly grocery stores and some mass merchandisers.⁴ Any change in product attributes yields a new UPC code so there are no changes in product attributes within the item-level data we use. The Nielsen data cover millions of products in more than 100 detailed product groups (e.g., Carbonated Beverages) and more than 1000 modules within these product groups (e.g., Soft Drinks is a module in Carbonated Beverages). While the Nielsen scanner item-level data are available weekly at the store level, our analysis aggregates the item-level data to the quarterly, national level.⁵ Since the weeks may split between months, we use the National Retail Federation (NRF) calendar to aggregate the weekly data to monthly data using the National Retail Federation (NRF) calendar. The NRF calendar creates places complete weeks into months and controls for changes in the timing of holidays and the number of weekends per month, and

⁴ Nielsen also has a scanner dataset based on household sampling frames called the Consumer Panel (Homescan). We discuss this dataset below and provide estimates based on it in the appendix. The results in the main body of the paper are based on the Nielsen retail scanner data made available through the Kilts Center of the University of Chicago.

⁵ The use of quarterly indices at the national level minimizes the problem of entry and exit of goods owing to stockouts or zero sales. Redding and Weinstein (2018) use quarterly aggregation, partially for this reason. To implement these methods in the statistical agencies, monthly indices would be required.

we use the months to create the quarterly data used in this paper. For more than 650,000 products in a typical quarter, we measure nominal sales, total quantities, and unit prices at the item level. We use the Nielsen scanner data from 2006:1 to 2015:4. The NPD Group (NPD)⁶ data cover more than 65,000 general merchandise stores, including online retailers, and includes products that are not included in the Nielsen scanner data. We currently restrict ourselves to the analysis of one detailed product module: memory cards.⁷ The NPD raw data are at the item-by-store-by-month level; NPD produces the monthly data by aggregating weekly data using the NRF calendar, as we do ourselves with the Nielsen data. Again, for our analysis we aggregate the data to the quarterly, national item level. For example, the item-level data for memory cards tracks more than 12,000 item-by-quarter observations for the 2014:1-2016:4 sample period. As with the Nielsen data, we measure nominal sales, total quantities, and unit prices at the item-level by quarter.

Because items are defined very narrowly (i.e., the UPC level) in both datasets, dividing sales by units sold gives a good measure of unit price. In principle, any changes in product attributes should yield a new UPC code. Both retailers and manufacturers have strong incentives to make UPC codes unique to specific products and the cost of assigning unique codes is minimal. Indeed, the ability to infer prices from unit values is a central advantage of measuring P and Q using scanner data. The unit price within a time interval is an average price for an item that will not capture within-period variation in prices that may be of interest.⁸

A. *Nominal Revenue Indices*

⁶ NPD, formally known as National Purchase Diary Panel Inc., collects, processes and analyzes transactions data from retail locations.

⁷ The NPD data includes a wide variety of product categories. The current analysis exams only one product; however, in future research we also plan to explore additional products.

⁸ For the Nielsen scanner data, our unit prices adjust for product size (e.g., number of ounces) so that the units within a product group are comparable.

Digitized item-level transactions data from individual retailers or data aggregators such as Nielsen and NPD can be used as an alternative source for measuring nominal expenditures. Moreover, such data permit the integration of the nominal expenditure and price measures at a very detailed level (i.e., at the item level). This approach solves many of the data integration and aggregation issues discussed above. In addition, novel approaches to quality adjustment of prices, including capturing the improvements in quality from product turnover, are available. Quality-adjusted prices built up from the same micro-level transactions data for measuring nominal expenditures have great advantages, as discussed above.

To begin, we compare the properties of nominal expenditure measures from survey vs. item-level transactions data. Table 2 presents summary statistics for nominal food sales from the Nielsen scanner data, nominal sales from grocery stores from the MRTS, and nominal BEA Personal Consumption Expenditure (PCE) for off-premises food and non-alcoholic beverages. The PCE data are only available seasonally adjusted while the MRTS are available both not seasonally adjusted and seasonally adjusted. The Nielsen scanner data are not seasonally adjusted. We use a simple quarterly dummy seasonal adjustment procedure to create a seasonally-adjusted series.

The top panel of Table 2 compares seasonally-adjusted statistics for all three series. Despite their completely different source data, the scanner, MRTS and PCE have similar average growth rates. The PCE is based in part on the MRTS, so the latter is not surprising. Consistent with this, the PCE is more highly correlated with the MRTS than with the Nielsen scanner series.

There are nonetheless 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 sales are based on a relatively small sample of firms (13,000 for the entire retail trade sector), while the Nielsen scanner data covers about 35,000 stores for grocery stores and mass merchandisers alone.⁹ Census retail sales at grocery stores include many non-food items but can exclude sales of food at, for example, general merchandise stores. In contrast, the Nielsen 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. Thus, one source of the differing volatility and seasonality of the scanner and the MRTS series (as exhibited in the bottom panel of Table 2) is likely differences in the coverage of nonfood items.

Considering the estimates of PCE highlights the advantages of item-level data that yield detailed product class information at high. Much of the high-frequency data underlying commodities in 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., 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.

⁹ Appropriate caution is needed in comparing firm-level and store-level counts. Large, national firms in retail trade have many establishments (stores). Foster et al. (2015) report that there are about 400 national firms in 2007 in Retail Trade that operate in more than 18 states. These 400 firms operate about 290,000 establishments. Our point is not that the MRTS has limited coverage of retail activity, but rather it is collected at a highly aggregated (firm-level) basis.

A related issue is that 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. BEA uses within-year composition information from scanner data from a commercial aggregator in combination with the PCE reported in Table 1 to produce statistics on more detailed food products (e.g., poultry).¹⁰

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 cyclical sensitivity.

B. Quality- and Appeal-Adjusted Price Indices

The promise of digitized data goes beyond the ability to produce internally consistent price and nominal revenue data. The item-level price and quantity data, which is often accompanied by information on item-level attributes, offer the prospect of novel approaches to quality adjustment. Currently, the BLS CPI implements hedonic quality adjustment on a relatively small share of consumer expenditures (about 5 percent). For the remaining items, a matched model approach is used with ad hoc quality adjustments when feasible (e.g., if a new model of an item has more features than a prior matched item, then an attempt is made to adjust the prices to account for the change in features). The sample of products in the CPI consumption

¹⁰ This use of aggregated scanner data by BEA is an example of how the statistical agencies are incorporating transactions data into the NIPA in the current architecture. Note that this use of aggregated scanner data is a patch to address a limitation of the existing architecture—that the detailed data from Census are only available once every five years for an annual reference period while the BEA is producing statistics at monthly and quarterly frequency.

basket is rotated every four years and no quality adjustment is made to prices when a new good enters the index due to product rotation.

The digitized data offer the possibility of accounting for the enormous product turnover observed in item-level transactions data. For the Nielsen scanner data, the quarterly rates of product entry and exit are 9.62% and 9.57%, respectively. By product entry and exit, we mean the entry and exit of UPCs from the data. Some of the product turnover at the UPC code level in the scanner data involves minor changes in packaging and marketing, but other changes represent important changes in product quality.

We consider two approaches for capturing the variation in quality in price indices using transactions data. The first approach is based on consumer demand theory and has been developed by Redding and Weinstein (2018, 2020) who build on the earlier work by Feenstra (1994). The second approach uses hedonic methods, following the insights of Pakes (2003, 2005) and Erickson and Pakes (2011). While these hedonic approaches are already partly in use by BLS and BEA, the item-level transactions data offer the potential for implementing these approaches with continuously updated weights and with methods to avoid selection bias arising from product entry and exit, and—equally importantly—at scale. Bajari et al. (2018) is an initial attempt to implement hedonics at scale using a rich set of product attributes. We draw out the many different issues that must be confronted for practical implementation of these modern methods by the statistical agencies. Since both of these methods are part of active research agendas, we emphasize that our discussion and analysis is exploratory rather than yielding ultimate guidance for implementation.

Redding and Weinstein (2018, 2020) use a CES demand structure at the product group level to generate the UPI. It is useful to provide a brief overview of the demand structure and the

underpinnings of the derivation of the UPI since this helps draw out conceptual and implementation issues. The CES demand structure for a narrow product group yields the unit expenditure function (the exact price index) given by:

$$P_t = \left[\sum_{k \in \Omega_t} \left(\frac{p_{kt}}{\varphi_{kt}} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \quad (1)$$

where Ω_t is the set of goods available in time t in this product group, p_{kt} is the price of good k at time t (purchased quantities), σ is the elasticity of substitution across goods within the product group, and φ_{kt} are relative product appeal terms.¹¹

The UPI implements this exact price index, which accounts for quality change and product turnover within a product group, using only observable data and an estimate for the elasticity of substitution. The UPI is given by the formula:

$$\ln(\text{UPI}) = \ln\left(\frac{P_t}{P_{t-1}}\right) = RPI + PV_{adj} + CV_{adj} \quad (2)$$

where RPI is a Jevons index given by the ratio of the geometric means of the prices for continuing goods between periods $t-1$ and t , PV_{adj} is a product variety adjustment bias term based on Feenstra (1994), and CV_{adj} is a consumer valuation bias adjustment term that is novel to the UPI. Formally, these three terms are given by:

$$RPI = \frac{1}{N_t^*} \sum_{k \in \Omega_t^*} \ln\left(\frac{p_{kt}}{p_{kt-1}}\right) \quad (3)$$

$$PV_{adj} = \frac{1}{\sigma - 1} \ln\left(\frac{\lambda_t}{\lambda_{t-1}}\right) \quad (4)$$

¹¹ A normalization is made so that average product appeal for a product group remains invariant over time.

$$CV_{adj} = \frac{1}{\sigma - 1} \frac{1}{N_t^*} \sum_{k \in \Omega_t^*} \ln \left(\frac{s_{kt}^*}{s_{kt-1}^*} \right) \quad (5)$$

where a * represents goods that are common in period $t-1$ and t , s_{kt}^* is good k 's share of expenditures on common goods in period t , p_{kt} is the price of good k in period t , σ and is the elasticity of substitution across goods. The product variety adjustment term (PV_{adj}) depends on

$\lambda_t \equiv \frac{\sum_{k \in \Omega_t^*} p_{kt} C_{kt}}{\sum_{k \in \Omega_t} p_{kt} C_{kt}}$, where C_{kt} is the consumption of good k at time t (purchased quantities). A

remarkable and attractive feature of the UPI is that, given an elasticity of substitution, this price index incorporating unobservable quality adjustment factors φ_{kt} is computable using observable information on prices and expenditure shares along with information to define common, entering and exiting goods.

The UPI is designed to be implemented on a narrow product group basis, which the item-level transactions data permit. Critical issues for the implementation of this approach include determining the classification of goods into narrow product groups, estimating the elasticity of substitution for each product group, and defining what constitutes entering, exiting, and common goods. We discuss these implementation issues below as we explore this method.

Before proceeding to the implementation issues, it is helpful to provide some intuition regarding the adjustment factors incorporated in the UPI. The product variety adjustment bias term depends on the relative expenditure shares of entering vs. exiting goods. Following Feenstra (1994), a higher expenditure share devoted to entering goods relative to exiting goods implies improvements in quality from product turnover. Feenstra's procedure adjusts expenditure shares for differences in prices of entering and exiting goods based on a CES demand structure. In particular, Feenstra's exact CES price index adjusts the Sato-Vartia exact

price index for product turnover simply by adding the PV_{adj} term defined in equation (4).¹²

Feenstra's price index is thus given as:

$$\log(\text{Feenstra}) = \ln\left(\frac{P_t}{P_{t-1}}\right) = PV_{adj} + \sum_{k \in \Omega_t^*} \omega_{kt} \ln\left(\frac{p_{kt}}{p_{kt-1}}\right), \quad (6)$$

where the weights ω_{kt} are defined as:

$$\omega_{kt} = \frac{\frac{s_{kt}^* - s_{kt-1}^*}{\ln(s_{kt}^*) - \ln(s_{kt-1}^*)}}{\sum_{k \in \Omega_t^*} \frac{s_{kt}^* - s_{kt-1}^*}{\ln(s_{kt}^*) - \ln(s_{kt-1}^*)}}. \quad (7)$$

Consumer demand theory implies that the quantitative importance of such quality change depends on the elasticity of substitution. Product turnover of very close substitutes (large σ) yields little product variety adjustment bias; mathematically, as the elasticity of substitution σ goes to infinity in equation (4), the PV_{adj} term goes to zero regardless of the amount of product turnover, because the new products are close substitutes with the old products. With a finite elasticity of substitution, the PV_{adj} term will be negative when the expenditure share on entering products is larger than the expenditure on exiting products. Conversely, the PV_{adj} term will be positive when the expenditure share on exiting products is greater. In our analysis below, we consider the Feenstra index along with the UPI.

The consumer valuation bias adjustment term applies similar logic by permitting changes in how consumers value continuing products over time. If the relative appeal of a product increases between periods $t-1$ and t , then consumer demand will shift towards that product. The relevant appeal-adjusted price should take into account consumers' substitution towards more desired products. The inclusion of the CV adjustment in the UPI is therefore internally

¹² The Feenstra price index is not simply the UPI holding the CV term constant. Instead, it is an adjustment to the Sato-Vartia index. The RPI in the UPI is not the Sato-Vartia index, but the simple ratio of geometric means of prices across time periods (the Jevons index).

consistent with consumer demand theory, which recognizes that relative product appeal can change over time, even for a given item.¹³ The elasticity of substitution is again a critical factor for the quantitative relevance of the CV term.

Implementation of the UPI requires an estimate of the elasticity of substitution at the product group level. Estimation of this elasticity is based on the demand function relating expenditure shares to prices, given by:

$$s_{kt} = \frac{p_{kt} c_{kt}}{\sum_l p_{lt} c_{lt}} = \frac{\left(\frac{p_{kt}}{\varphi_{kt}}\right)^{1-\sigma}}{\sum_{l \in \Omega_t} \left(\frac{p_{lt}}{\varphi_{lt}}\right)^{1-\sigma}} = \frac{\left(\frac{p_{kt}}{\varphi_{kt}}\right)^{1-\sigma}}{P_t^{1-\sigma}}, \quad k \in \Omega_t \quad (8)$$

In practice, a common procedure is to use the Feenstra (1994) estimation approach or some related modification. Focusing on the shares of the expenditures on common goods, the expenditure share relationship can be double-differenced (differencing out time effects but also potentially specific group effects like brand or firm effects as in Hottman, Redding and Weinstein (2016)) to yield the relationship:

$$\Delta \ln \bar{s}_{kt}^* = \beta_0 + \beta_1 \Delta \bar{p}_{kt} + u_{kt}, \quad \beta_1 = (1 - \sigma) \quad (9)$$

(where the notation reflects the impact of double-differencing). Feenstra (1994), Hottman, Redding and Weinstein (2016) and Redding and Weinstein (2018, 2020) overcome the potential endogeneity bias in equation (9) by: (i) specifying and double-differencing an analogous supply curve; (ii) assuming the double-differenced demand and supply shocks are uncorrelated; and (iii) assuming heteroskedasticity across individual products in the relative variances of demand and supply shocks. The advantage of this method is that it can be implemented at scale with item-

¹³ Cost of living indices are typically defined by holding utility constant, so normally do not allow for taste shocks. Redding and Weinstein note, however, that there are very large changes in demand for goods that are not accounted for by changes in price, and developed the UPI to account for this fact. This churning in demand is evident from item-level transactions data, so the analytic innovation of Redding and Weinstein is motivated as an approach to accommodate such data in a price index. Note that the shocks are to relative demand within a narrow product group, not to the level of demand given income.

level transactions data, but the disadvantage is that it relies on these strong identifying assumptions.

Another critical issue is defining what constitutes common, entering, and exiting goods. In our analysis, we implement the UPI with an entering good in period t (quarterly) defined as a good that had no expenditures in period $t-1$ but positive expenditures in t ; an exiting good as one with positive expenditures in period $t-1$ but not in t ; and a common good as one that has positive expenditures in both periods. This implementation is consistent with the theory, and we denote our implementation the *theoretical UPI*.¹⁴

In the published version of the paper, Redding and Weinstein (2020) depart from these assumptions by defining common goods based on a much longer horizon in their baseline estimates. Specifically, their baseline calculates the UPI based on changes from the 4th quarter of year $t-1$ to the 4th quarter of year t . They define *common goods* between those two periods as follows: (i) the good must have been present in the three quarters prior to the 4th quarter of $t-1$; (ii) the good must have been present in the three quarters after the 4th quarter of t ; and (iii) the good must be present cumulatively for at least six years. With this definition of common goods, *entering goods* are any goods in the 4th quarter of year t that has positive expenditures in that period but is not a common good. Likewise, *exiting goods* are any goods in the 4th quarter of year $t-1$ that has positive expenditures in that period but is not a common good.

¹⁴ This version is close to the implementation of Redding and Weinstein (2018), with one exception. Redding and Weinstein calculate annual price changes from the 4th quarter of one year to the 4th quarter of the next year. They define common goods as goods that are present in both of those quarters. In our implementation, we calculate price changes on a quarter-over-quarter basis, and define common goods as goods that are present in those two consecutive quarters. We chain the resulting price index to calculate annual inflation. This difference does lead to quantitatively different measures of price change,

The motivation for this alternative baseline definition of common goods is based on practical implementation concerns about the implications of the UPI that we will discuss in detail below.¹⁵ Redding and Weinstein (2020) suggest that the slow roll-out of new goods across stores and geographic areas create complex entry dynamics. Likewise, there is a slow process of exit along the same dimensions. Moreover, the dynamics of product entry and exit may reflect dynamic learning by consumers, along with heterogeneity in the preferences for newer goods across consumers. Since these factors are outside the scope of their theoretical model, they suggest in their 2020 paper that permitting this type of seasoning of entering and exiting goods is a practical way to abstract from these factors. Following Robert Feenstra’s discussion of our paper at the conference suggesting a similar approach, we call this version the *seasoned UPI*.

We are sympathetic to these practical concerns, but given our objective of exploring key issues, in the interest of transparency, we find it instructive to implement the theoretical UPI. In addition, our use of the timing conventions for entering, exiting, and common goods in the theoretical UPI closely align with the definitions of entering, exiting, and common goods used in the hedonics literature. This enables us to draw out the differences between the UPI and hedonic approaches more readily. In addition, as will become apparent, we think there are other implementation issues beyond seasoning that must be addressed, and they are easier to understand using the implementation of the theoretical UPI. Moreover, the seasoned UPI has the limitation that it could not be implemented on a timely basis at scale. One would only know

¹⁵ Technically the issue can be understood by reviewing the terms in the consumer valuation adjustment in equation (8). This term is the unweighted average of log changes in expenditure shares of common goods. Goods with very small shares can have very large log changes that dominate this term. We are grateful to our discussant Robert Feenstra for this observation.

whether a good is common in a period until well after that period is complete.¹⁶ As we discuss, below there are alternative ways of implementing a common good rule, but we think that considerable research is required to develop a practical implementation.

If attribute data are available along with the price and quantity data, then an alternative approach to accounting for product turnover and quality adjustment is to use hedonics.

Following Pakes (2003), Bajari and Benkard (2005), Erickson and Pakes (2011) we estimate hedonic regressions using item-level data every period within a product group of the form

$$\ln(p_{it}) = X_i' \beta_t + \eta_{it} \quad (10)$$

where X_i is vector of characteristics or attributes of good i . Note that the attributes are time-invariant at the item-level in contrast to earlier hedonic approaches that examine how price changes for a broadly-defined good (a car, a computer) as the attribute changes. In the approach we feature here, the goods are narrowly defined at the item level. If the attributes change, it is presumed that the good will be given a new UPC or barcode, and therefore be treated as a different good. A core challenge of implementing hedonics is measuring the relevant set of attributes. As Bajari and Benkard (2005) emphasize, if only a subset of the relevant attributes is included in the regression, then this generates a bias in the hedonics-based price indices.

A Laspeyres index quality-adjusted using hedonics at the product group level is given by

$$\text{Laspeyres Hedonic Index}(t) = \frac{\sum_{i \in A_{it-1}} h^t(X_i) q_{it-1}}{\sum_{i \in A_{it-1}} h^{t-1}(X_i) q_{it-1}} \quad (11)$$

¹⁶ This problem is even more of a challenge since the objective is to generate timely, monthly price indices. There are also some conceptual issues with the definition of the common goods in terms of how long a good must be present to be a common good. It is important to emphasize that the common good rule approach of Redding and Weinstein (2020) blurs the distinction between entering and exiting goods. Their common goods rule implies that goods that don't satisfy the threshold are in practice put into the entry/exit category. This implies that their entry/exit category in period t includes goods that have positive sales in both $t-1$ and t .

where $h^t(X_i) = X_i' \beta_t$ is the period- t estimate of the hedonic function and A_{it-1} is the set of all goods sold in period $t-1$ (including exits). Using hedonics in this manner adjusts for quality and selection bias from exiting goods by imputing the price of the exiting goods in period t using the hedonic function.¹⁷ Transactions data permit use of item-level rather than sample weights from alternative sources. An analogous approach can be used for a Paasche index that adjusts for selection bias for entering goods. The Fisher ideal index using hedonics incorporates both adjustments. An important feature of implementation in this setting with item-level transactions data is the use of continuously updated weights (period $t-1$ weights for Laspeyres and t weights for Paasche). Bajari and Benkard (2005) observe that such chain weighting is readily feasible with item-level transactions data and that such chain weighting accommodates the incorporation of product turnover.

A practical challenge for implementing hedonics at scale is measuring attributes at scale. Machine learning approaches as in Bajari et al. (2018) could in principle be used to overcome this issue. Bajari et al. (2018) convert text and images to vectors and use dimensionality reduction techniques to estimate hedonically adjusted prices at scale for millions of products at a high frequency. They show that their approach yields high R -squared measures in the estimation of hedonic functions.

In comparing the UPI and hedonic approaches, an advantage of the UPI is that it is fully consistent with micro consumer demand theory that reconciles the relationship between expenditure shares and prices. As is apparent from equations (8) and (9), the UPI approach defines the product quality/appeal as the residual from the demand equation. This contrasts with the hedonic approach, which only uses the variation in prices that can be accounted for by

¹⁷ Pakes (2003) emphasizes that this is one of the key advantages of hedonic indices relative to standard price indices.

observable characteristics. As already noted, omitted characteristics will bias the estimates obtained using the hedonic approach. We discuss the strengths and weaknesses of claiming the entire residual (as in the UPI) below.

These two approaches can in principle be combined. Crawford and Neary (2019) build on both approaches in developing a Feenstra (1994) product variety adjustment factor to the standard hedonic approach. For this purpose, their suggested price index is analogous in form to the Feenstra-adjusted Sato-Vartia index but using characteristics in the adjustment factor and the remainder of the index is a Sato-Vartia hedonic index. We have not explored this hybrid approach ourselves, but our results below suggest this is a promising area to pursue. Our separate consideration of the UPI and hedonics does raise issues that need to be confronted by practical implementation of this hybrid approach developed by Crawford and Neary (2019).¹⁸

III. Results

We implement both the UPI and the hedonics approach and compare these quality-adjusted price indices to standard price indices using transactions-level data. For the Nielsen scanner data, we only implement the UPI since attribute data are less readily available. For the NPD scanner data, we have rich attribute data that permit us to implement both the UPI and hedonics approaches. We also have obtained estimates of a CPI-type index restricted to the same product groups in the Nielsen scanner data. BLS created aggregate indexes for the comparable food and non-food items.¹⁹ Thus, for the analysis of the Nielsen scanner, we compare the BLS CPI to the price indices from the transactions data.

¹⁸ The proposed hybrid by Crawford and Neary (2019) does not incorporate the time-varying product appeal terms that are at the core of the UPI. In many respects, this hybrid approach should be interpreted as a hybrid of the Feenstra-adjusted Sato Vartia index and hedonics.

¹⁹ 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.

We begin by examining prices for the Nielsen scanner data classifying the more than 100 product groups into food and nonfood items. To implement the Feenstra index and the UPI, we require estimates of the elasticity of substitution. For our initial analysis with both the Nielsen and NPD data, we use the Feenstra (1994) estimation procedure as modified by Redding and Weinstein (2018) for use with item-level transactions data. The sizes and impacts of the CV and PV depend critically on the elasticity estimates. The estimated elasticities for the 100+ product groups display considerable variation. While the median is about 8, the 10th percentile is 4 and the 90th percentile is 16.²⁰

Table 3a provides summary statistics of alternative price indices for the 2006-2015 period using the BLS CPI and Nielsen scanner data where the 100+ product groups have been classified into food and nonfood items. The number of item-level price quotes each month in the BLS CPI for these product groups is about 40,000, compared to the 650,000 item-level prices in the scanner data. We first create quarterly estimates for the 2006-15 period at the product group level. We then aggregate the quarterly estimates to food and non-food items using Divisia expenditure share weights by product groups. Table 3a provides summary statistics using the quarterly food and nonfood price indices.

The top panel of Table 3a shows the results for the food product groups and the lower panel the nonfood product groups. Each panel displays results for quarterly price changes based on three indices calculated from the scanner data: a Laspeyres index; the CES demand-based price index with the adjustment for product turnover proposed by Feenstra discussed above (hereafter Feenstra); and the UPI. To calculate the Laspeyres index using the item-level data, we use previous-quarter expenditure weights updated for each quarter.

²⁰ The mean is 9 and the standard deviation is 5. The median estimate, of 8, is similar to that reported in Table 2 of Redding and Weinstein (2018) using the Feenstra estimator.

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 0.91). The Feenstra shows a notably lower average price change and a correlation with the CPI that is also 0.91. The UPI has a much lower average and a correlation with the CPI of 0.63. The finding that the CPI and the Laspeyres from the 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, and we made similar adjustments for changes in package size in the Nielsen data. The close relationship between the CPI and Laspeyres for food provides a benchmark to gauge the impact of the quality adjustments via Feenstra and the UPI, which like the Laspeyres use the scanner data.

The lower panel shows greater differences across price indices for nonfood. For this category, the CPI inflation rate is slightly higher than the scanner Laspeyres rate, but their correlation is substantially weaker (0.42) than for food. The Feenstra price index 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 (-0.22). 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 larger gap between the Feenstra and UPI than there is between the Laspeyres and Feenstra.

Results for the UPI presented here differ some from the patterns presented in Redding and Weinstein (2018), who use the Nielsen Consumer (Homescan) Panel in their analysis. The latter tracks the expenditures of about 55,000 households. Households scan the bar codes from purchased items, and prices are either downloaded from the store where the item was purchased

or hand entered. Table 3b presents the analogous statistics to Table 3a comparing the BLS CPI for food and nonfood items (covering the same product groups) to the Consumer Panel-based price indices.²¹ Qualitatively, the patterns are similar between the retail scanner and consumer panel indices, with some exceptions. For food, the correspondence between the BLS CPI and consumer panel Laspeyres is weaker than that between the BLS CPI and retail scanner Laspeyres. The correlation is 0.80 instead of 0.91, and the BLS CPI has a notably lower mean (0.42% lower) than the consumer panel Laspeyres. Additionally, the gap between the UPI and Laspeyres is much smaller for the consumer panel compared to the retail scanner data. Especially notable is that the gap between the Feenstra and UPI indices is substantially smaller using the consumer panel compared to the retail scanner data for both food and nonfood. This implies the CV adjustment is not as large for the consumer panel compared to the retail scanner data.²² Even more dramatic reductions in the magnitude of the CV adjustment using the consumer panel emerge by using the seasoned UPI as developed by Redding and Weinstein (2020).²³

Taken at face value, the results suggest that the UPI captures substantially more quality adjustment than captured by the CPI, 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

²¹ We estimate the elasticities of substitution separately for the Consumer Panel data.

²² It is beyond the scope of this paper to investigate the sources of discrepancies between the results using the retail scanner and consumer panel databases. We focus on the retail scanner data because it is arguably more comprehensive and also more suitable for the objectives of re-engineering key national indicators.

²³ In unreported results, we have found that we can replicate the findings in Redding and Weinstein (2020) using a “seasoned” UPI based on a simpler to implement common goods rule than their longevity rule. Specifically, if we define common goods as those with above the fifth percentile of the market shares over the current and prior five quarters we can closely approximate their findings. As we have discussed, we have not implemented a common good rule in this paper to produce a “seasoned” UPI since it is our objective to draw out the issues associated with dealing with the rich entry and exit dynamics of goods and its impact on the UPI. Imposing such a rule limits the ability to explore such issues.

elasticities at a product group level (e.g., carbonated beverages for food and electronic products for nonfood) allows for over 100 different elasticities within the scanner data, this level of aggregation may still be too high. While each product group contains goods that are very close substitutes, many product groups also contain goods that are quite different. For product turnover and expenditure share volatility with close substitutes, the quality adjustment factors in the Feenstra and UPI indices become very small. The procedure used in Tables 3a and 3b is to assume (and estimate) the same elasticity of substitution for all products within a product group.

Turning now to the analysis with the NPD data, the detailed characteristic data allows us to consider the hedonic approach. For the analysis in this paper, we present estimates for memory cards. We have estimated a separate hedonic regression relating an item's log price (at the national quarterly level) to its attributes for each quarter. We use a quadratic in memory card size and speed and dummy variables capturing card types (e.g., flash cards, memory chips, etc.). At the national level, the dataset contains about 12,000 observations of product items for all quarters over the three year period.

Memory cards have exhibited substantial improvements in quality over our short sample period. Figure 1 shows the sales-weighted linear trend of memory card size and speed, with both size and speed more than doubling over the sample period. Figure 2 shows that the marginal value of additional size and speed appears to be declining over time. Hedonic regressions that include both size and speed and the squares of size and speed show that prices are increasing in size and speed but at a decreasing rate.²⁴ The R-squared is about 0.8 in each quarter, suggesting that there could be other unobserved characteristics that affect the prices.

²⁴ Hedonic regressions of $\log(\text{price})$ regressed on quadratic in size and speed along with attribute dummies (not reported) estimated by quarter from 2014 to 2016.

As with the Nielsen scanner data, we also estimate a UPI using the NPD data. The elasticity of substitution (i.e., demand elasticity estimate) is about 4. We can use this method to estimate product quality levels for entering and exiting goods using the φ_{kt} in Redding and Weinstein equation (4). As expected, entering goods have substantially higher average quality (-0.28) than exiting goods (-1.23), but they also have much more dispersion in quality (standard deviation of 1.59 for entering goods compared to 1.43 for exiting). Greater dispersion at entry suggests potentially interesting post-entry dynamics that may involve selection and learning. We explore this below.²⁵

Table 4 summarizes means, standard deviations and correlations of the alternative price indices for memory cards. Quality-adjusted prices are declining much more rapidly than standard price indices indicate: the Feenstra and hedonic indices are substantially lower than Laspeyres index, and the UPI is even lower. Most of the series are highly correlated, except for the UPI. Interestingly, the price indices most highly correlated with UPI are the hedonic indices. This suggests the hedonics comes the closest to capturing the quality adjustment measured in the UPI.

a. What does the CV term capture, and why is it so large?

Taken at face value, the UPI yields the most comprehensive quality-adjusted prices. Our analysis suggests it yields substantially more quality adjustment than the Feenstra index and the hedonics based indices. The differentiating factor for the UPI is the inclusion of the consumer valuation bias adjustment (CV). In principle, the hedonic approach also permits changing valuations of characteristics that could capture the variation in the CV. Even though we find that

²⁵ We also note that it is not feasible to implement the seasoned UPI using the very long horizons specified in Redding and Weinstein (2020) for the NPD data, since being a common good requires being present for six years.

the hedonic indices are the most highly correlated with the UPI for memory cards, it is apparent that the UPI via the CV captures quality adjustment not captured in the other indices.

This discussion suggests it is critical to understand what the CV adjustment bias (using what we denote as the theoretical UPI) is capturing. To explore this issue, we conduct some further exploratory analysis. Figure 3 shows that (the logs of) PV and CV are negative on average for virtually all products in the Nielsen data (or alternatively PV and CV in levels are below one). The figure also shows that PV and CV are also positively correlated across product groups. However, CV shows much more variation than PV; log PV ranges from 0 to -.06 while CV ranges from 0 to -0.4.

Figure 4 shows the UPI components for two narrow product modules that highlight the variation depicted in Figure 3. The top panel shows the (log of) UPI and its components for soft drinks (RPI, PV, and CV, in logs and, for PV and CV, multiplied by $\frac{1}{\sigma-1}$). The bottom panel shows the analogous components for video games. Both panels show quarterly log price changes, i.e., they are not expressed as annualized values. The scales of the two figures are the same in order to highlight the dramatic differences in the respective roles of PV and CV across these product modules. Video games thus fits the pattern of Figure 3 in that both PV and CV are large in magnitude.

The scale of Figure 4 obscures substantial measured price declines for soft drinks. The UPI implies average price deflation of approximately 2 percent per quarter from 2006:1 to 2015:4. Put differently, the UPI suggests that entry- and appeal-adjusted prices for soft drinks fell by more than half over a period of ten years. Mechanically, almost all of that decline stems

from the CV term.²⁶ The large amount of price deflation for soft drinks implied by the UPI is dwarfed by the massive deflation that is measured for video games. The UPI suggests that (appropriately measured) prices for video games fell by more than 99 percent over the same ten year period.

These large implied rates of quality-adjusted price declines are heavily dependent on the estimated elasticities of substitution. For soft drinks, we use the estimate of $\sigma = 6.22$, the estimate for the product group, carbonated beverages, using the Feenstra estimation method. If we let $\sigma = 12$, then the rate of price decline is less than one percent per quarter (less than half of that reported in Figure 4). Thus, even without changing the definitions of common goods, the UPI delivers more plausible results for higher estimated elasticities of substitution. The Feenstra method for estimating the latter makes strong identifying assumptions and further research is needed in this area.

Figure 3 suggests that CV may reflect post-entry and pre-exit dynamics given its close relationship with PV. To explore this possibility, we conducted some small-scale simulations of product entry and exit with associated changes in product quality presented in Figure 5, 6, and 7. In each of the simulations, we track the evolution of 14 items within a simulated product group with an assumed elasticity of substitution of five. Seven goods enter and seven goods exit; one entering good and one exiting good in each period. To focus on product quality, we keep price for each good equal and constant across time, that is, the goods are produced competitively with a constant and equal marginal cost. To examine the quality and price changes, the average

²⁶ For comparison, the official CPI published by the BLS for carbonated beverages yields an average quarterly price increase of 0.4 percent between 2006:1 to 2015:4, and the Sata-Vartia index for soft drinks yields a similar 0.4 percent average quarterly increase. The RPI term of the UPI for soft drinks tracks the CPI fairly well, yet yields an average quarterly change of zero. The correlation coefficient between the two series is 0.75. The major differences between the UPI and the CPI therefore arise from the PV and CV terms.

quality increases over time. Thus, all of the variation in appeal shows up in appeal-adjusted prices. As a result, the constant quality index represented by the Laspeyres shows no price change, while the appeal-adjusted UPI shows a decreasing price index. In each figure, Panel A shows an example of the relative quality paths for an entering and exiting good along with the increase in average quality and Panel B shows the implied price indices.

In Figure 5, new products enter at higher quality than exiting products, but there is no change in quality post-entry or pre-exit. This results in flat post-entry appeal in the example in Panel A where the new good enters in the 4th period; however, average quality increases due to the higher quality new goods. Panel B of Figure 5 shows the implied Laspeyres, Feenstra, and UPI log price indices along the CV adjustment factor component of the UPI. The Laspeyres exhibits no rate of change given constant unadjusted prices. The Feenstra and UPI show substantial negative price change reflecting the product entry and exit that are identical. The CV adjustment factor component is zero since there is no change in quality for continuing goods. The latter implies that common goods expenditure shares are not changing over time.

In Figure 6, the 14 new products exhibit post-entry dynamics as upon entry and exiting goods exhibit pre-exit dynamics. An example of an entering and exiting good is shown in panel A, with a new good entering in period 4 with a relative appeal of 1.0 and increasing in appeal to 1.2 in the next period, while the exiting good has the reverse pattern. With these entering and exiting goods in each period, Panel B shows the implied Laspeyres, Feenstra and UPI log price indices (making the same assumption of constant unadjusted prices). As with each example, the Laspeyres exhibits no rate of change given constant unadjusted prices. The Feenstra and UPI show substantial negative price change reflecting the product entry and exit but there is a large

gap between the UPI and the Feenstra. The CV adjustment component is large suggesting the CV is capturing post-entry and pre-exit dynamics.

To emphasize this possibility, Figure 7 depicts an alternative simulation where the post entry buildup of appeal is slower. Panel A shows an example of a new good entering in period 4 with a relative appeal of 0.8 and increase to 1.2 in the next period. Alternatively, pre-exit dynamics are faster in that the initial fall in appeal is larger, with the initial fall in appeal to .067. As a result, Panel B shows that the gap between the Feenstra price index and UPI is even larger in this case, with slower and richer post-entry dynamics. The richer post-entry dynamics generate a larger gap between the UPI and the Feenstra price indices with the CV capturing a larger share of the dynamics.

While these are simple illustrative simulations, they highlight a potentially important driving force distinguishing the UPI from the Feenstra, namely, that the UPI captures changes in relative product appeal associated with more complex post-entry and pre-exit dynamics than permitted by the Feenstra index. The latter only captures quality differences at the exact points of product entry and exit. Instead, it may be that there are learning and other adjustment dynamics that imply the product quality changes from product turnover take time to evolve post entry and pre exit. In this respect, our finding that the UPI and the CV play a significant role relative to the Feenstra index highlights the potential advantages of the UPI, but also suggests the need for care in interpreting both the CV and the UPI.

Figures 5-7 emphasize the potential role of post-entry and pre-exit dynamics that may be associated with the life cycle dynamics of products. In a related fashion, a simpler and more basic relationship between the CV and PV may emerge due to measurement and timing issues. At high frequencies (e.g., weekly or monthly) it is likely that entering and exiting products

exhibit some ramp up as products become available in a diffuse manner geographically and some ramp down as the last product is sold in a specific location. Our use of quarterly, national measures mitigates these measurement and timing issues. Nonetheless, these issues remain present, underlying the gap between the theoretical and seasoned UPIs.

Figure 8 and Figure 9 document patterns in relative product appeal among entering and exiting goods in the video game and soft drink modules, respectively, of the Nielsen scanner data.²⁷ Both figures display the distributions of relative appeal of continuing, entering, and exiting goods. The upper-left-hand panel of Figure 8 shows that for video games, as for memory cards, entering goods have a higher mean appeal than exiting goods but a lower mean than continuing or common goods. Entering goods also have more dispersion in relative appeal than either exiting or common goods consistent with what we found for memory cards.

The remaining panels of the figure show, however, that this is an incomplete characterization of entry and exit. In the upper-left-hand panel, entry is defined for any item that had zero sales in the prior quarter and positive sales in the current quarter. Exit is likewise defined for any item that had positive sales in the prior quarter and zero sales in the current quarter. In many cases, these entry and exit dynamics don't represent true entry (the first quarter an item is observed) or true exit (the last period an item is observed). The remaining panels show that the mean product appeal of true entrants is substantially higher than of re-entrants and indeed is higher than the mean for common goods. In contrast, exiting and re-exiting goods have similar means. These patterns suggest that re-entering and temporarily exiting goods are likely part of the end of a product life cycle (hence the similar means between re-entering, temporary exits, and true exits). Figure 9 shows the analogous patterns for soft drinks. Qualitatively, the

²⁷ As with Figure 3, this uses the estimate of φ_{kt} .

patterns are similar, but they are substantially less pronounced, consistent with the notion that changes in product appeal and technological change are less rapid for soft drinks than for video games.

The post-entry and pre-exit dynamics of relative product appeal, price, and market share for video games and soft drinks are depicted in Figure 10 and Figure 11. The figures display patterns for 11 quarters after entry and 11 quarters prior to exit.²⁸ Statistics for post-entry are relative to the first period of entry for each good. Statistics for pre-exit are relative to 11 quarters prior to exit for each good. Both means and medians of these life cycle dynamics are displayed.

For video games, products decline in relative product quality, market share, and price both post-entry and pre-exit, with the means and medians showing similar patterns. The magnitudes of the declines are substantial. Relative quality declines by 150 log points in the first 11 quarters post-entry, price declines by 100 log points, and the market share declines by 300 log points. Similar magnitudes are present for the 11 quarters prior to exit. These patterns of substantial post-entry and pre-exit dynamics help account for the large role of the CV adjustment for video games. Recall our simulations in Figures 5 through 7 show that post-entry and pre-exit dynamics that exhibit substantial changes in relative quality yield a substantial CV adjustment. To help put these patterns into context, it is useful to observe that the probability of exit increases with product age. In unreported statistics, we find that the exit rate rises from under 5 percent in the first five quarters post-entry to over 20 percent after 20 quarters.

For soft drinks, the dynamics are different and the effects are somewhat muted. There is a hump shaped behavior in post entry relative appeal, price, and market share. The gap between the mean and median patterns is also substantial—consistent with the right tail driving the mean

²⁸ We restrict analysis for the post-entry to goods that survive at least 11 quarters and for pre-exit for goods that are present for all 11 quarters prior to exit. Results without these restrictions show similar patterns.

dynamics relative to the median. At the mean, relative quality rises more than 30 log points in the first five quarters, but it falls to 15 log points lower than the initial product quality after 11 quarters. Pre-exit dynamics show monotonic patterns similar to video games but are less steep. Relative product quality is 100 log points lower just before exit than 11 quarters prior to exit. This compares to the 150 log point gap for video games. Even though the magnitudes are somewhat smaller, they are still substantial and help account for the non-trivial CV adjustment for soft drinks. The exit rate also rises with product age for soft drinks but again the patterns are more muted, with the rate increasing from about 4 percent shortly after entry to about 6 percent after 20 quarters.

The CV term's close connection with the product valuation (PV) adjustment term introduced by Feenstra (1994) complicates attempts to interpret the PV term in isolation. For instance, Diewert and Feenstra (2018) argue that the infinite reservation or choke price implied for every good under CES preferences is *a priori* unreasonable. They advocate for a rule of thumb that price indices should reflect one half of the welfare gains implied by a CES utility function. Their setting does not allow for time-varying appeal shocks, however. As we have seen in this section, allowing for such shocks via the CV term substantially increases measured deflation via a channel that is independent of consumers' reservation prices. It may well be that when the consumer valuation channel is considered, the welfare gains from entering products are larger than is implied by the classic Feenstra (1994) approach.

All of this discussion of complex entry and exit dynamics can be used to justify the practical implementation of the seasoned UPI. However, from our vantage point, this appears to be an important area for future research. Instead of simply assuming a long horizon for

seasoning,²⁹ we think it useful to understand the nature of the entry and exit dynamics. The discussion above suggests that the nature of those dynamics likely varies across goods. Simply assuming a common horizon for seasoning is likely to be inadequate; at the least, this topic merits further investigation.

b. Claiming φ : The Demand Residual

The large declines in the UPI, even for product categories such as soft drinks that are not obvious hotbeds of technological innovation, raise the question of whether the implied estimates are reasonable, and if so, how best to interpret them.

Redding and Weinstein (2018) take a strong view in formulating the UPI: they treat *all* of the *measured* residual demand variation not accounted for by changing prices as reflecting changes in product appeal or quality. The UPI exactly rationalizes observed prices and expenditure shares by treating the entire error in an estimated demand system as reflecting such changes. In contrast, other approaches, such as hedonics or the Feenstra (1994) approach, leave an estimated residual out of the price index calculation. Although hedonic approaches can in principle capture much of the variation from changing product quality and appeal, the R-squared in period-by-period hedonic regressions is typically substantially less than one. Conceptually, therefore, although both the UPI and hedonics capture time-varying quality and appeal valuations from both product turnover and continuing products, the UPI is arguably more general because it comprehensively captures the error term from the underlying demand system in the price index.

The debate over whether it is appropriate to treat the entire error term from an estimated consumer demand system as reflecting changes in product quality and appeal that affect the cost

²⁹ Or alternatively some threshold rule for market share for common goods in the implementation of the UPI.

of living is very much in its infancy, however. The measured error term from the estimated demand system may reflect measurement or specification error from several sources. Specification error may reflect not only functional form but also a mis-specified degree of nesting or level of aggregation. Presumably, those errors would ideally be excluded from the construction of a price index.

Another possible source of specification error relates to permitting richer adjustment dynamics in consumer demand behavior. Diffusion of product availability, diffusion of information about products, habit formation, and learning dynamics will show up in the error term from estimation of specifications of static CES demand models. A related but distinct possibility is that the underlying model of price and quantity determination should reflect dynamic decisions of the producing firms (through endogenous investments in intangible capital like customer base as well as related marketing, promotion, and distribution activity by firms). It is important to remember that the approaches being used to estimate the elasticity of substitution are jointly estimating the demand and supply system so misspecification of either the demand or supply equations can yield specification error.

We are not yet able to quantify the importance of these measurement and specification issues. One area we think is especially promising is to explore a more theoretically based definition of product group classification and nesting. In the next section, we examine the UPI's sensitivity to product group classification and nesting.

c. Product Group Classification and Nesting

An inherent challenge for implementing the UPI is the definition of product groups and the associated estimation of the elasticities of substitution. In our implementation (and consistent with the approach taken in Redding and Weinstein (2018)), we have assumed all items within a

product group or module defined by the data provider are equally substitutable. A review of the individual items in these groups quickly suggests this is a very strong assumption. Consider soft drinks. Presumably some soft drinks (e.g., caffeinated, with sugar, colas) are much closer substitutes than others. Moreover, some of the item-level variation in the scanner data for soft drinks reflect changes in packaging associated with marketing during holiday seasons. Similar remarks can be made about nonfood items. For video games, essentially the same game may be released with slightly different features or complementary support products—again suggesting very close substitutes in some cases. Alternatively, vintages of video games of only a few years ago (e.g., PacMan) are quite primitive compared to the latest games with advanced graphics, animation, and audio (e.g., FIFA World Cup Soccer 2018). In spite of the likely substantial differences in the degree of substitutability among sub-groups of products, the estimation of the elasticity of substitution we have considered so far pools items in the cross section and over time, yielding a single elasticity by product group.

To provide some perspective on the potential importance of this issue, Figure 12 illustrates three different versions of the UPI calculated using the Nielsen scanner data for all product groups, but classifying groups at different levels of aggregation. The first version of the UPI is constructed using the item-level (UPC code) variation. The second UPI aggregates items to common text descriptions within the Nielsen scanner data. The third UPI aggregates items based on item attributes defined in terms of product module, brand, size and packaging.³⁰ For each UPI considered, all products within an aggregation (e.g., text descriptions or attributes) are treated as perfect substitutes. The UPI becomes substantially less negative using these more

³⁰ For this exercise, the elasticities of substitution at the product group level are based on the Feenstra estimation procedure for each product group. They are computed as usual for the first version of the UPI. For the second and third versions, we use the same elasticity for the components within the product group. Hence, the exercise highlights the effects of aggregation.

aggregated product definitions. Figure 12 demonstrates that important components of the UPI (via the PV and CV terms) depend on the methods of classifying products and reflect variation between goods that are likely much closer substitutes than with other items in the same pre-defined grouping.

In principle, a nested CES approach can be used to construct the UPI, potentially overcoming this issue. Redding and Weinstein (2018) show in their appendix that this is feasible conceptually. There are two primary challenges for implementing a nested CES. First, what classification of goods should be used? Hottman, Redding and Weinstein (2016) consider a nested CES using a within vs. between firm classification, and they provide evidence that goods produced within a firm are more substitutable than goods produced across firms. It is unclear, though, that this is an ideal or sufficient classification approach. An alternative might be to use product attributes. This possibility raises an interesting question: could the use of product attributes to define nests lead the UPI and hedonic approaches to be more similar in implementation than they first appear in principle? Our initial analysis above already found that the price indices that most closely approximate the UPI are the hedonic indices; perhaps making use of the same attributes as in the hedonic approach to generate product classification and nests for the UPI will yield indices that track each other even more closely.

A second primary challenge for implementing a nested (or even non-nested) CES utility-based index is the estimation of the elasticities of substitution. The various approaches used for estimation in the literature make strong identifying assumptions. These identification issues become that much more complex in a nested environment. For example, Hottman, Redding, and Weinstein (2016) use a nested procedure that is essentially a modified version of the Feenstra approach for the within firm estimation (double differencing firm and time effects) and then use

an instrumental variables procedure for the higher level between firm estimation. The instrument emerges from the structure of the model as the firm-level price index is a UPI at the firm-level with one important term being the within-firm CV adjustment term. The latter reflects changing relative product appeal shocks across goods within the firm. They argue that the latter is orthogonal to the (double differenced) between-firm relative appeal shocks. This procedure uses strong identifying assumptions at both levels.

Finally, the approach to hedonics with item-level transactions data based on Bajari and Benkard (2005) that we pursued for memory cards has some advantage over traditional implementations of hedonic methods, but it requires further enhancements to be scalable. An advantage of using hedonics with transactions data is that the weights can be updated continuously and in turn selection bias from exit (as in the Laspeyres approach above), entry (the Paasche approach), or both (using both Laspeyres and Paasche hedonics adjustment and computing a Fisher hedonics adjustment) is feasible. Still, the approach we used for memory cards relied on high-quality and relevant attributes (memory card size and speed) being readily available from the NPD data. Other datasets, such as the Nielsen scanner data, have less readily available information on item attributes. As noted above, the machine learning approaches of Bajari et al. (2018) show great promise in overcoming these issues.

IV. Re-engineering the Data Architecture

The opportunities created by the ubiquitous digitization of transactions can only be realized with a new architecture for data collection. The aim is for the statistical system to use all the relevant detail provided by transactions data. There are number of issues the new data architecture would need to address (see Jarmin, 2019). These include issues of privacy,

confidentiality, and value of business data; cost to businesses and the statistical agencies of the new architecture; and the technical and engineering issues of building a new architecture.

There are multiple potential modes for businesses providing such data. All have advantages and disadvantages. We expect that the new architecture should support multiple approaches to providing and collecting data. The agencies will need to be flexible.

Direct feed of transactions-level data. The agencies could get transactions-level data directly from firms and do the calculations necessary to aggregate them. This approach has already been implemented by the Australian Bureau of Statistics for its retail food price index. While the agencies should be receptive to such arrangements, it is unlikely to be practical in the U.S. context because of unwillingness of companies to provide such granular data and the difficulty for the agencies of handling the volume of data that it would entail.

Direct feed of (detailed) aggregate measures of price, quantity, and sales via APIs. Alternatively, and probably more practical in the U.S. context, firms (e.g., retailers) could do the calculations needed to produce detailed but aggregated measures of price, quantity, and sales that could then be transmitted to the statistical agencies. Surveys and enumerations could be replaced by APIs. The agencies—in collaboration with businesses—would have to design a large, but finite, number of APIs that would mesh with the information systems of firms. As is typical for IT innovations, doing so would have a substantial fixed cost, but then provide much improved data at low marginal cost.

Third-party aggregators. Third-party aggregators are already collecting much of the relevant data from many firms (especially retailers). These third parties could do the aggregation as part of their service and provide client firms with an option of responding to statistical agency requests using their service.

Note that the choice among these modes is not just a matter of how the data are collected but carries substantive implications for producing the indices discussed in this paper. The first option of direct feed of transactions, and perhaps the third option of third-party aggregators, potentially allow the pooling of observations at the item level across firms. In contrast, the second option would provide price, quantity, and sales measures possibly aggregated into quite detailed products, but would not allow direct pooling at the item level across businesses because the index-number formulas are highly non-linear. Hence, there is an interaction between decisions about nesting in the index number construction discussed in Section III with the data architecture.

These approaches—whether direct data feeds, API, or third-parties—would have many benefits to firms beyond improving the public good provided by official statistics. Firms could save costs by not having to transform their business data to meet the requirements of statistical agencies’ surveys. This approach would reduce the current burden associated with collecting data to replace multiple survey requests from agencies. For example, this approach would replace having BLS price enumerators visit outlets. These visits take the time of store management through queries related to sampling goods and finding replacements for goods that disappear. In addition, obtaining revenue data directly from firms could replace the collection of Census’s monthly retail trade survey.

Notwithstanding these potential benefits, achieving firms’ participation in a new data collection paradigm will be a considerable challenge. Voluntary compliance with current data collection is far from universal, so cooperating with a new paradigm for collecting data certainly cannot be taken for granted. New approaches to data collection may require firms to incur additional costs, at least at the outset. More generally, it may require firms to rethink how and

why they interact with statistical agencies. The next section of this paper and the Introduction to this volume consider some of these challenges.

V. Capabilities and Mandates of the Statistical Agencies

This paper envisions a new architecture for economic statistics that would build consistent measurement of price and quantity from the ground up. Currently, the collection and aggregation of data components is spread across three agencies. Implementing the new architecture we envision undoubtedly will be a challenge. Moving away from a survey-centric form of data collection for retail prices and quantities to computing statistics from detailed transaction level data requires an approach that would have businesses providing their data in a unified way. The institutional arrangements that fundamentally separate the collection of data on prices and quantities would need to be changed. There have long been calls for reorganizing BEA, BLS and Census to help normalize source data access, improve efficiencies and foster innovation. Regardless of whether the agencies are realigned or reorganized, they need to review the current structure given how the production of statistics is evolving. Having one agency negotiate access to transaction level data will be difficult enough. Having multiple agencies doing so unduly burdens both businesses and the taxpayer. Importantly, under the current statistical system structure, no agency has the mandate to collect data on both price and quantities, so implementing the data architecture to measure price and quantity simultaneously is not in scope for any agency.³¹

³¹ The agencies are undertaking important and innovative work using transactions data as part of their ongoing measurement programs, some of which is described in papers in this volume. The Census Bureau pays for access to limited data for experimenting with augmenting the Monthly Retail Trade Survey (Hutchinson 2019). The BLS has multiple efforts to replace or augment CPI enumerations with alternative sources (Friedman, Konny, and Williams 2019). Notably, both of these efforts are focused on using non-survey data to supplant or supplement data collections within the current architecture, so the Census effort does not measure prices and the BLS efforts do not measure quantities. The Census Bureau is, however, supporting this project by providing NPD data that does measure price and quantity simultaneously. An exception to the agencies not considering both price and quantity

There are also difficult questions about the legal and policy structure needed to govern how statistical agencies access private data assets for statistical uses. For instance, a key question is whether companies would seek to charge for access to the type of data described above and, if so, whether the associated fees would be within the budgetary resources of the statistical agencies.

To further test, develop and implement a solution such as we are proposing here, the statistical agencies must expand their general data science capabilities. Whether transaction level data are transmitted to the agencies or whether retailers provide intermediate calculations, an important point of focus for the statistical agencies will be not only the acquisition but the curation of new types of unstructured data. The ingestion, processing and curation of these new sources introduces scalability concerns not present in most survey contexts. Also, negotiating access will require the agencies to hire more staff with the skills to initiate and manage business relationships with data providers.

Clearly, modernization requires significant investments in computer science and engineering expertise at the statistical agencies. This is a major challenge given the competition for attracting talent across other government agencies and with the private sector. Collaboration with academic experts and contracting can be part of the solution, but some internal expertise is essential.

The collective economic measurement system will need to make a number of investments. It will need to invest in building relationships across government agencies and the private sector to secure access to high quality source data. It will need to invest in staff with the skills to acquire, process and curate large datasets and build reliable and privacy-protected

simultaneously when using non-survey data is the BEA's extensive program to address measurement of quality change in health care.

statistical products from blended data. Information systems will need to be redesigned to accommodate both survey and alternative data processing. These are large challenges, but we believe necessary in order to build a 21st century statistical system that can deliver the trusted information needed by private and public sector decision makers.

VI. Concluding Remarks

In the introduction to the 2000 NBER/CRIW conference volume *Scanner Data and Price Indices*, Feenstra and Shapiro (2002) stated, “Scanner data and other electronic records of transactions create tremendous opportunities for improving economic measurement.” Almost two decades after that conference, researchers have made progress using digitized transactions data on many dimensions, but the U.S. statistical agencies have not yet implemented the vision of using such data for dramatic improvements in economic measurement for official statistics. Indeed, many of the papers in that conference pointed to the difficulty in using scanner data for measurement. Both push and pull factors, however, suggest the time is now ripe for full-scale implementation of using transactions-level data that will yield a significant re-engineering of key national indicators. In particular, developments in economics and computer science, such as the UPI and hedonics-at-scale, are innovations that address some of the difficulties with using scanner data for economic measurement under the existing architecture for economic statistics.

On the push side, declining response rates on business and household surveys yield both higher costs and lower quality of economic measurement. Relatedly, the current decentralized system imposes a substantial burden on households and businesses with a multiplicity of surveys. On the pull side, the digitization of virtually everything has been dramatic over the last two decades. Moreover, substantial progress has been made on the technical challenges for implementation. Active research using item-level transactions data has yielded development of

price index methodology that captures quality changes from product turnover and changing product appeal for continuing goods. Based on our review and exploration of the methodological innovations, we conclude that integration of the alternative approaches that have been proposed is likely to be fruitful.

In particular, the UPI methodology developed by Redding and Weinstein (2018) has great promise, but it likely requires refinements that are closely connected to an alternative hedonics-based approach to quality adjustment. We suspect that a successful implementation of the UPI methodology requires a nested product classification approach based on nests defined by product attributes of individual items. Tracking item-level product attributes is at the core of the hedonics-based approach. A limitation of the latter is that implementation has involved intensive study of each product group (e.g., computers) one at a time. Advances in machine learning and other data dimensionality reduction techniques offer the prospect of implementation of either the nested UPI or the hedonics approach with attributes at scale. It remains to be seen what exact method will prove to be conceptually and practically the best approach.

Beyond the issues of developing classification of groups based on hedonics, the theoretical UPI also faces other practical implementation challenges related to the complex dynamics of entering and exiting goods. One practical implementation method is to use what we have denoted as the seasoned UPI to overcome these issues. We think it is premature to settle on this methodology. We instead suggest further investigation into the nature of the entry and exit dynamics of goods. We anticipate substantial progress will be made on this issue in future research.

Active and intensive research on these issues should be a high priority. At the same time, substantial effort needs to be made in exploring how the U.S. statistical agencies can harvest the

firehose of digital data that are increasingly available. The agencies are experimenting with alternative harvesting approaches, but a variety of challenges remain. In addition, implementing this 21st-century approach to using integrated price and quantity collection and measurement will require rethinking the coordination and organization of the U.S. statistical agencies.

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.
- Bajari, P, Chernozhukov, R. Huerta G. Monokrousos, M. Manukonda, A. Mishra, B. Schoelkopf. 2018. "Quality Adjusted Price Indices Powered by ML and AI" Presentation made at FESAC, December (<https://www.census.gov/content/dam/Census/about/about-the-bureau/adrm/FESAC/meetings/Chernozhukov%20Presentation.pdf>).
- Bureau of Economic Analysis. 2016. "NIPA Revisions Table: Selected Components Detail and Major Source Data Incorporated," July 2016.
- Bureau of Labor Statistics. 2018. "Chapter 17. The Consumer Price Index." BLS Handbook of Methods.
- Crawford, Ian and J. Peter Neary. 2019. "New characteristics and hedonic price index numbers." CESifo Working Paper No. 7529.
- Diewert, Erwin and Robert Feenstra. 2018. "Estimating the Benefits of New Products." Unpublished paper, University of British Columbia.
- Ehrlich, Gabriel, John Haltiwanger, Ron Jarmin, David Johnson, and Matthew D. Shapiro. 2019. "Minding Your Ps and Qs: Going from Micro to Macro in Measuring Prices and Quantities." *AEA Papers and Proceedings* 109, 438-443.
- Erickson, Tim and Ariel Pakes. 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.

- Feenstra, Robert C. 1994. “New Product Varieties and the Measurement of International Prices,” *American Economic Review*, 84(1), 157-177.
- Feenstra, Robert C. and Matthew D. Shapiro. 2002. “Introduction” in *Scanner Data and Price Indices*, NBER/CRIW Conference volume (Feenstra and Shapiro, eds.), University of Chicago Press.
- FitzGerald, J., and Shoemaker, O. (2013). Evaluating the Consumer Price Index Using Nielsen’s Scanner Data. In *Joint Statistical Meetings* (pp. 1033–1045).
- Foster, Lucia, John Haltiwanger, Shawn Klimek, C.J. Krizan and Scott Ohlmacher. 2015. “The Evolution of National Retail Chains: How We Got Here” in *Handbook of Retail Trade* (Basker, eds).
- Friedman, David, Crystal Konny, and Brendan William. 2019. “Big Data in the U.S. Consumer Price Index: Experiences & Plans.” This conference.
- Hottman, Colin, Stephen Redding, and David Weinstein. 2016. “Quantifying the Sources of Firm Heterogeneity,” *Quarterly Journal of Economics*, 131(3), 1291-1364.
- Hutchinson, Rebecca. 2019. “Using Nontraditional Data Sources to Reduce Respondent Burden in United States Census Bureau Economic Data Products.” This conference.
- Jarmin, Ron. 2019. “Evolving Measurement for an Evolving Economy: Thoughts on 21st Century US Economic Statistics,” *Journal of Economic Perspectives*, 33(1), pp. 165-184.
- 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. “Measuring Aggregate Price Indexes with Demand Shocks: Theory and Evidence for CES Preferences,” NBER Working Paper 22479, revised May 2018. <https://www.nber.org/papers/w22479.rev2.pdf>

Redding, Stephen and David Weinstein. 2020. "Measuring Aggregate Price Indexes with Demand Shocks: Theory and Evidence for CES Preferences," *Quarterly Journal of Economics* 135:1, 503-560.

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

Table 1: Measuring Real and Nominal Consumer Spending—Current Architecture

<p>Census (nominal spending)</p> <p>Data collection: Retail Trade surveys (monthly and annual) Economic Census (quinquennial) Consumer expenditure survey (conducted for BLS)</p> <p>Published statistics: Retail Trade (monthly and annual) by firm type Retail Trade (quinquennial) by product class</p>	<p>BLS (prices)</p> <p>Data collection: Consumer Expenditure survey (used for spending weights), collected under contract by Census Telephone Point of Purchase survey (purchase location)^a CPI price enumeration (Probability sampling of goods within outlets)</p> <p>Published statistics: Consumer Price Index (monthly) by product class</p>
<p>BEA (aggregation and deflation)</p> <p>Data collection: Census and BLS data; supplemented by multiple other sources</p> <p>Published statistics: Personal Consumption Expenditure: Nominal, real, and price (monthly) GDP (quarterly)</p>	

Note: This table shows key elements of for measurement of real and nominal consumer spending.

^a The TPOPS will be incorporated into the CES.

Table 2: Comparisons of Nominal Quarterly Growth for Food Sales—Surveys vs. Scanner Data

A. Seasonally Adjusted			
	Scanner	Census MRTS (Grocery)	PCE
Mean	0.87	0.74	0.78
Standard deviation	0.98	0.64	0.61
Correlations:			
Scanner	1.00		
Census MRTS (Grocery)	0.49	1.00	
PCE	0.65	0.86	1.00
B. Not Seasonally adjusted			
Standard deviation	2.87	2.70	
Correlations:			
Scanner	1.00		
Census MRTS (Grocery)	0.31	1.00	

Notes: Census MRTS is for Grocery Stores. PCE is for Food and Non-Alcoholic Beverages Off Premises. Period is 2006:2-2015:3. PCE is seasonally adjusted by BEA and MRTS by Census. Scanner seasonally adjusted in top panel using seasonal dummies.

Table 3a: Summary Statistics on Comparisons of Quarterly Price Indices

A. Food				
	BLS CPI	Scanner Laspeyres	Scanner Feenstra	Scanner UPI
Mean	0.57%	0.76%	0.16%	-2.49%
Standard deviation	0.77%	0.82%	0.82%	0.84%
Correlations				
BLS CPI	1.00			
Scanner Laspeyres	0.91	1.00		
Scanner Feenstra	0.91	0.97	1.00	
Scanner UPI	0.63	0.72	0.66	1.00
B. Non-Food				
	BLS CPI	Scanner Laspeyres	Scanner Feenstra	Scanner UPI
Mean	0.22%	-0.05%	-0.62%	-4.59%
Standard deviation	0.46%	0.36%	0.38%	0.77%
Correlations				
BLS CPI	1.00			
Scanner Laspeyres	0.42	1.00		
Scanner Feenstra	0.37	0.90	1.00	
Scanner UPI	-0.22	0.18	0.23	1.00

Note: Nielsen scanner product groups are classified into food and non-food items. BLS CPI is harmonized to these product groups. Quarterly series from 2006:2-2015:4 reflect the log first differences of the price indices

Table 3b: Summary Statistics on Comparisons of Quarterly Price Indices: Consumer Panel [CP] Data

A. Food				
	BLS CPI	CP Laspeyres	CP Feenstra	CP UPI
Mean	0.49%	0.91%	0.01%	-1.27%
Std Deviation	0.76%	0.77%	0.81%	0.86%
Correlations				
BLS CPI	1.00			
CP Laspeyres	0.80	1.00		
CP Feenstra	0.81	0.86	1.00	
CP UPI	0.60	0.66	0.75	1.00
B. Non-Food				
	BLS CPI	CP Laspeyres	CP Feenstra	CP UPI
Mean	0.21%	0.51%	-0.58%	-3.31%
Std Deviation	0.43%	0.42%	0.56%	0.79%
Correlations				
BLS CPI	1.00			
CP Laspeyres	0.48	1.00		
CP Feenstra	0.50	0.64	1.00	
CP UPI	0.37	0.48	0.70	1.00

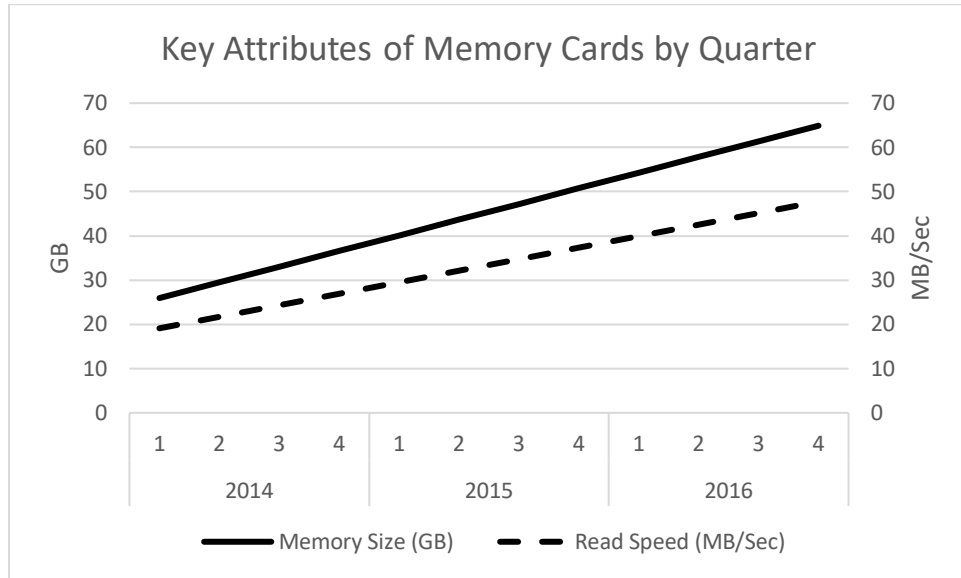
Note: This table replicates the calculations in Table 3a using the Nielsen Consumer Panel [CP] data instead of the Nielsen retail scanner data. Data are quarterly from 2004-2016 reflect the log first differences of the price indices

Table 4: Means, Standard Deviations, and Correlations of Alternative Price Indices: Memory Cards

	Laspeyres	Feenstra	Hedonic (Laspeyres)	Hedonic (Paasche)	UPI
Mean Price Change	-0.039	-0.059	-0.060	-0.049	-0.096
Standard Deviation (price change)	0.034	0.039	0.024	0.025	0.024
Laspeyres	1.00				
Feenstra	0.89	1.00			
Hedonic (Laspeyres)	0.72	0.72	1.00		
Hedonic (Paasche)	0.61	0.72	0.77	1.00	
UPI	0.15	0.07	0.32	0.48	1.00

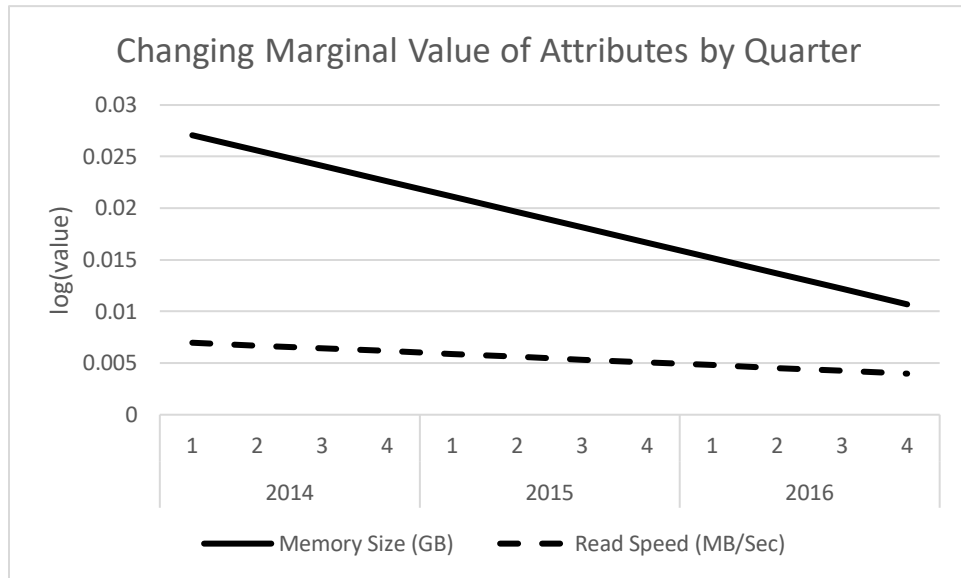
Note: Source is NPD data at item-level quarterly from 2014 to 2016. Price indices constructed at a quarterly frequency. Reported statistics are correlations of quarterly indices (not seasonally adjusted).

Figure 1



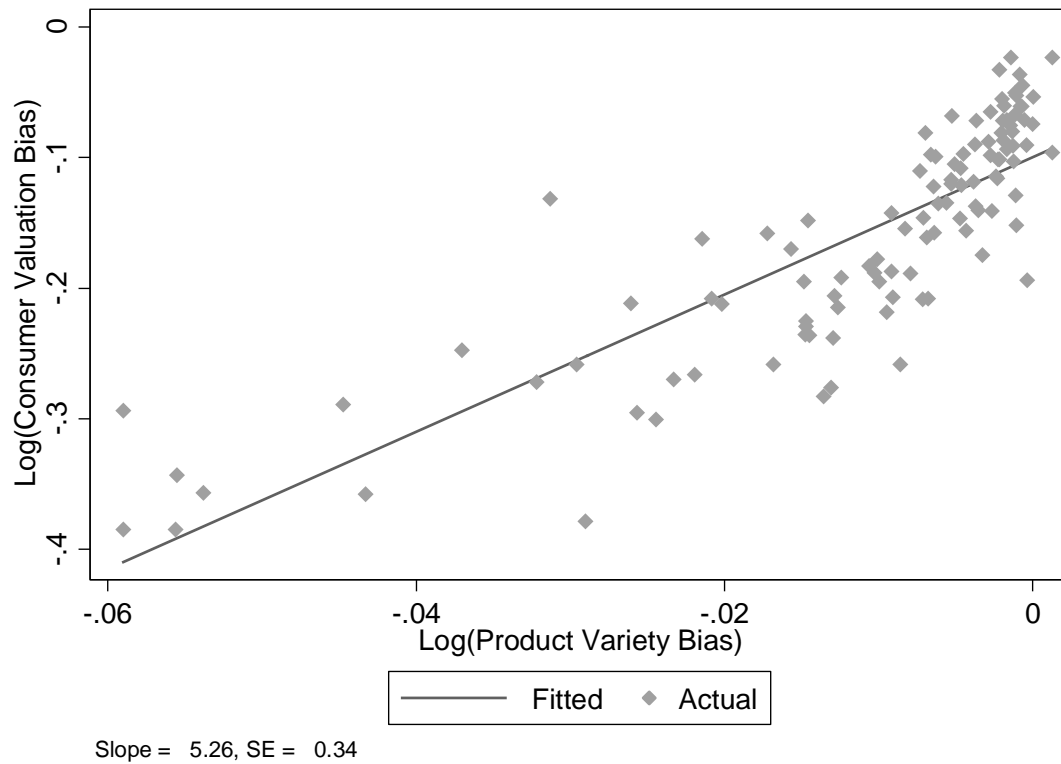
Note: NPD data. The figure shows estimated linear trends in sales-weighted national memory size and read speeds used to produce the estimates in Table 4.

Figure 2



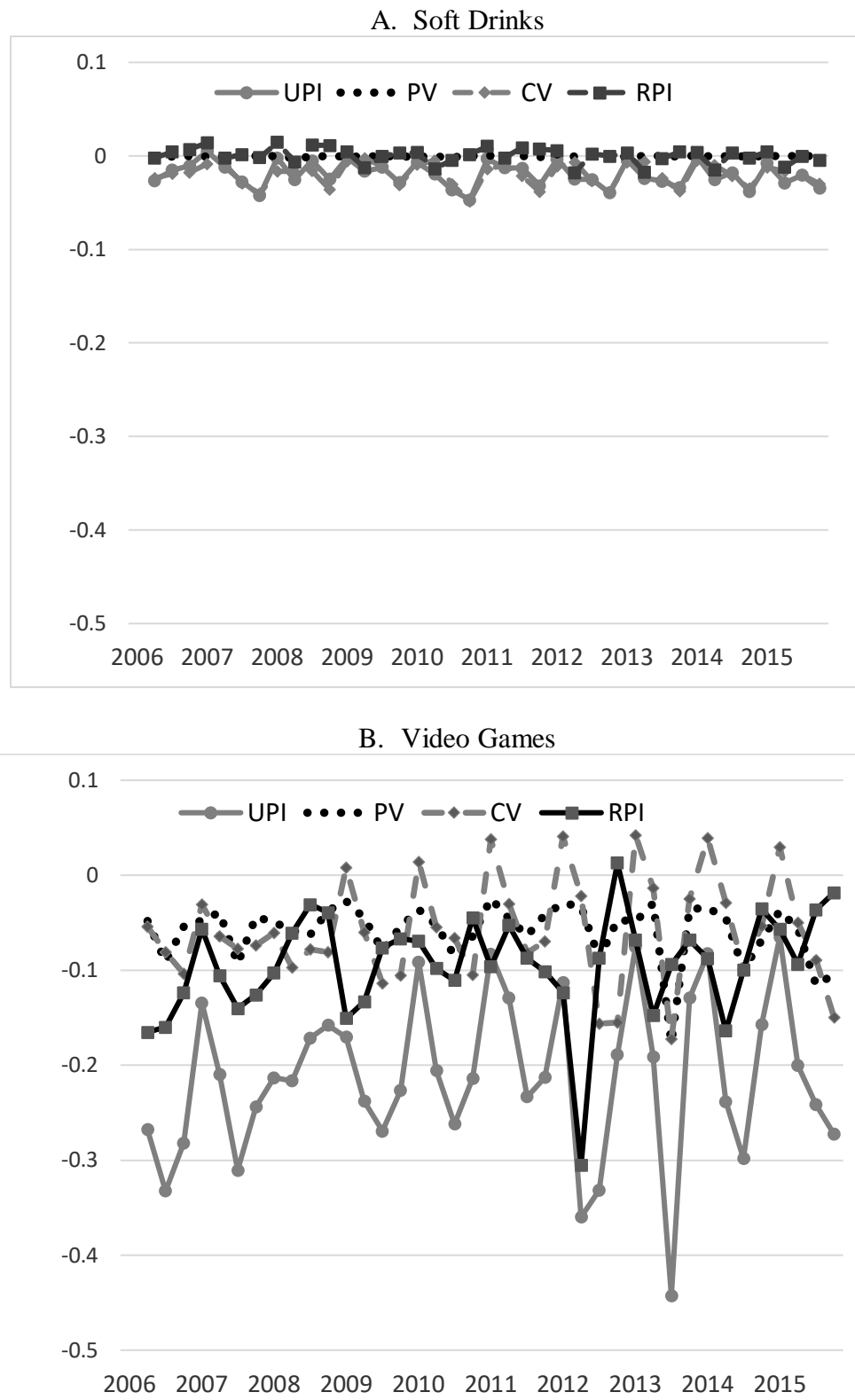
Note: NPD data. This figure shows the changing marginal value (from linear term from Table 4) in estimation of hedonic specification for memory cards

Figure 3: The Relationship between Product Variety Bias and Consumer Valuation Bias



Note: Nielsen scanner data. Each dot represents a product group showing average PV and CV adjustment factors from quarterly measures of PV and CV. Quarterly series from 2006:2-2015:4.

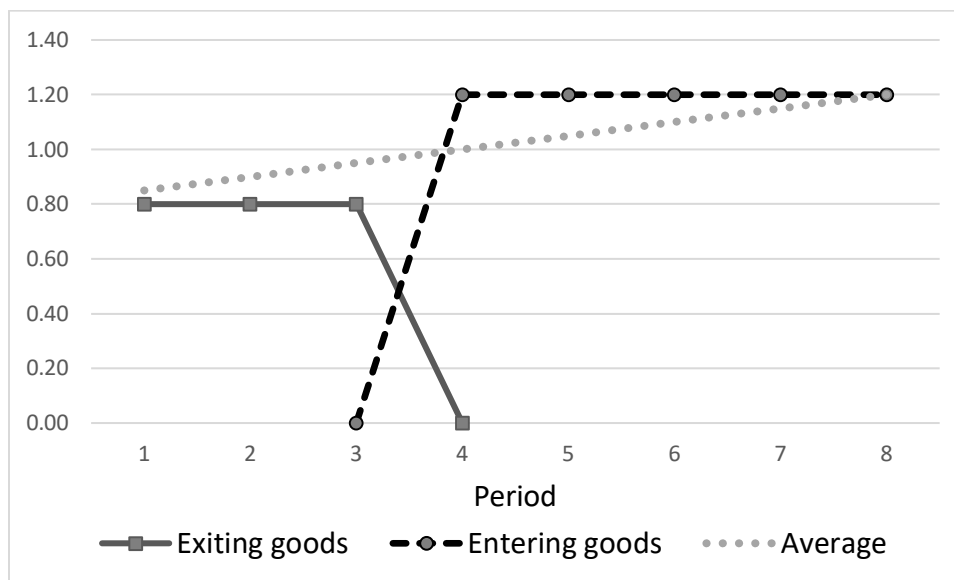
Figure 4: UPI Components for Specific Product Modules



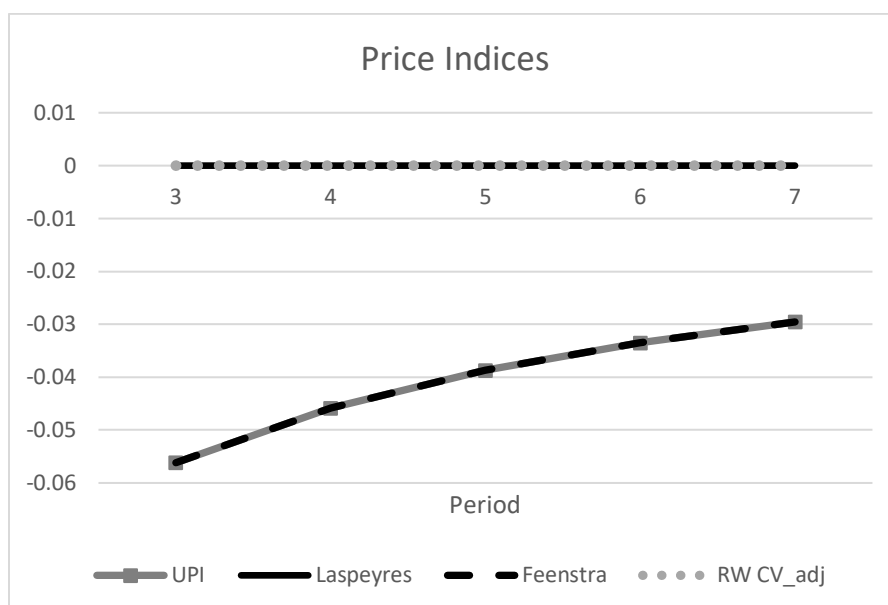
Note: Nielsen scanner data. Log differences at quarterly rate.

Figure 5: Simulated Product Entry and Exit with Quality Change: No Continuing Good Quality Change

A. Relative Product Quality



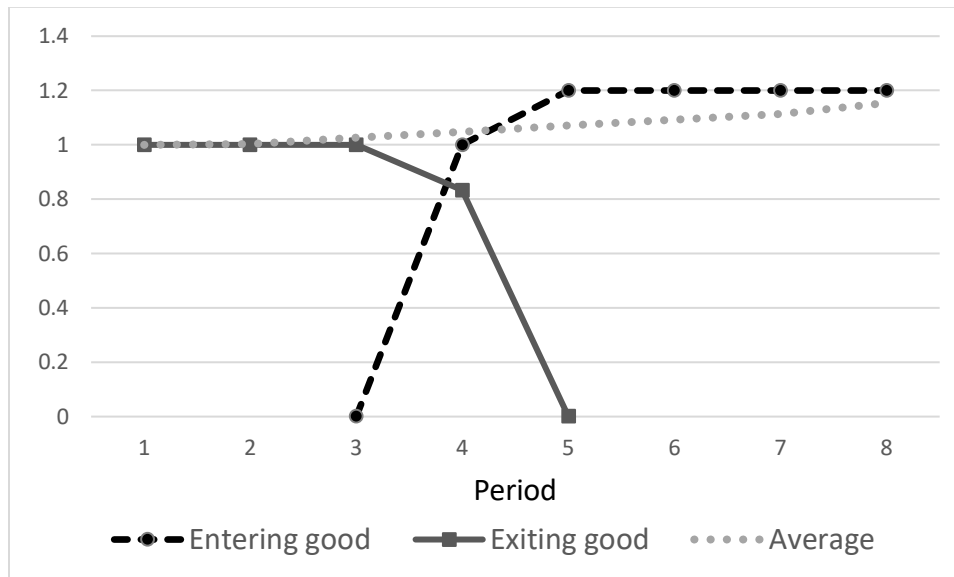
B. Implied Price Indices



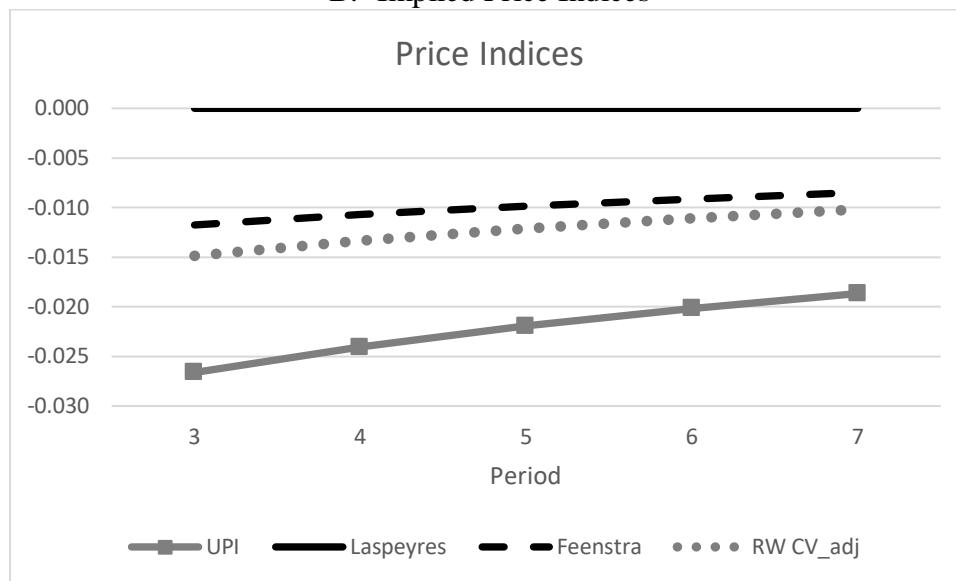
Note: These figures show the results of a simulation in which 14 goods enter and exit with constant price and changing quality. Panel A shows an example of the relative quality paths for an entering and exiting good along with the increase in average quality and Panel B shows the implied price indices.

Figure 6: Simulated Product Entry and Exit with Quality Change: Post Entry and Pre-Entry Quality Dynamics

A. Relative Product Quality



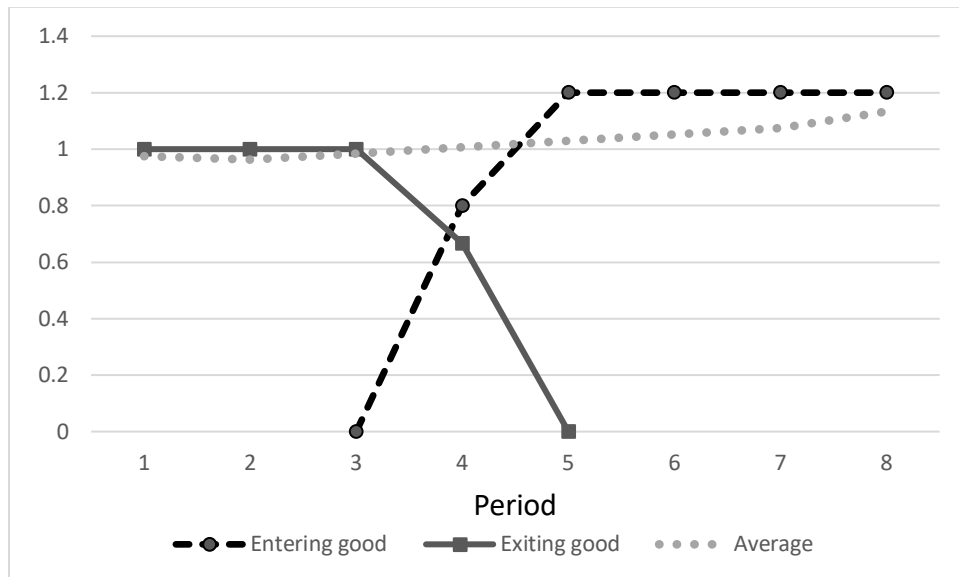
B. Implied Price Indices



Note: These figures show the results of a simulation in which 14 goods enter and exit with constant price and changing quality. Panel A shows an example of the relative quality paths for an entering and exiting good along with the increase in average quality and Panel B shows the implied price indices.

Figure 7: Simulated Product Entry and Exit with Quality Change: Slower Post-Entry and Pre-Exit Quality Dynamics.

A. Relative Product Quality



B. Implied Price Indices

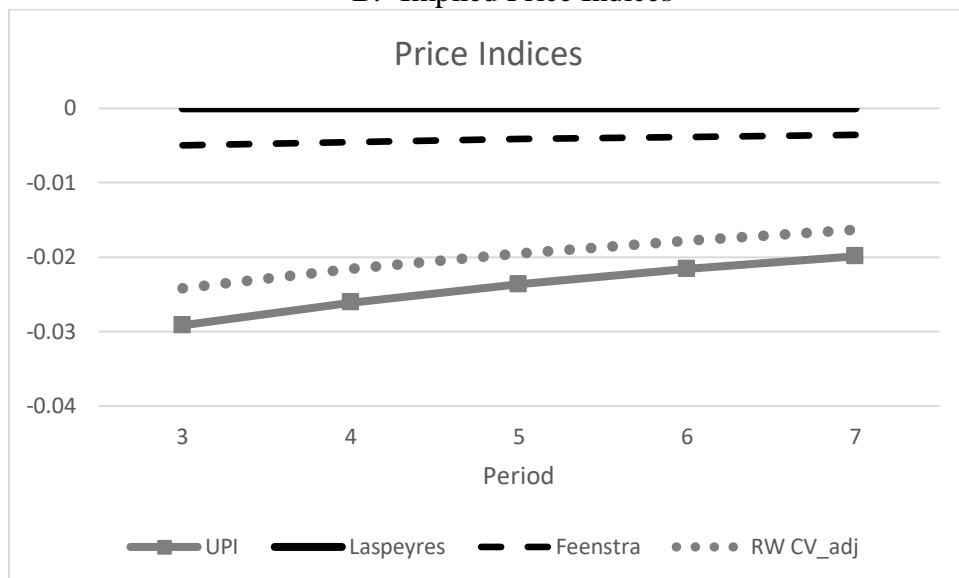
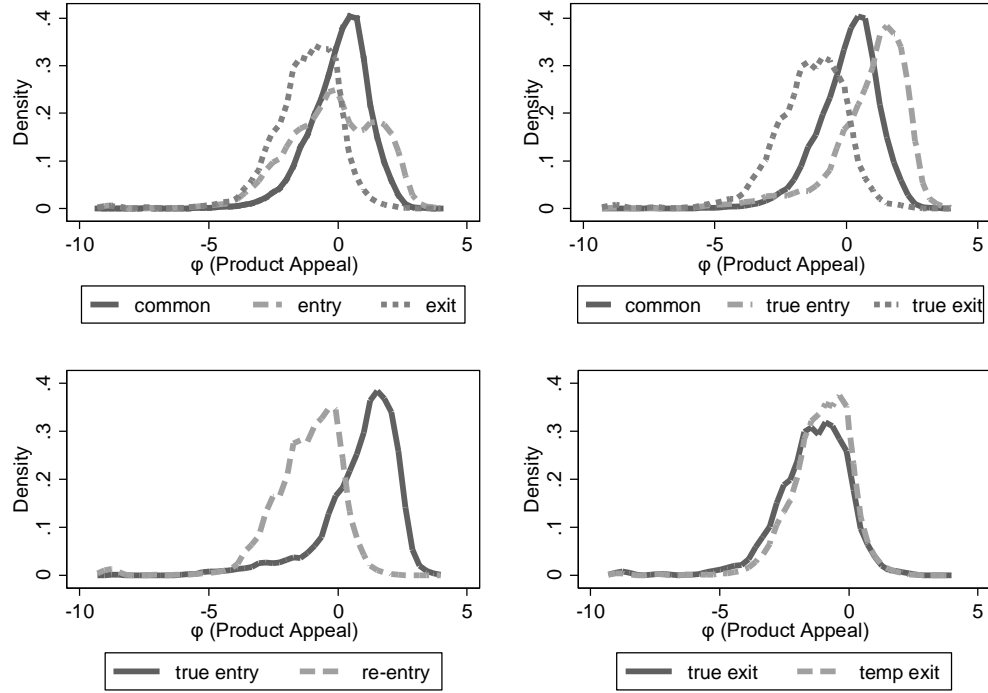
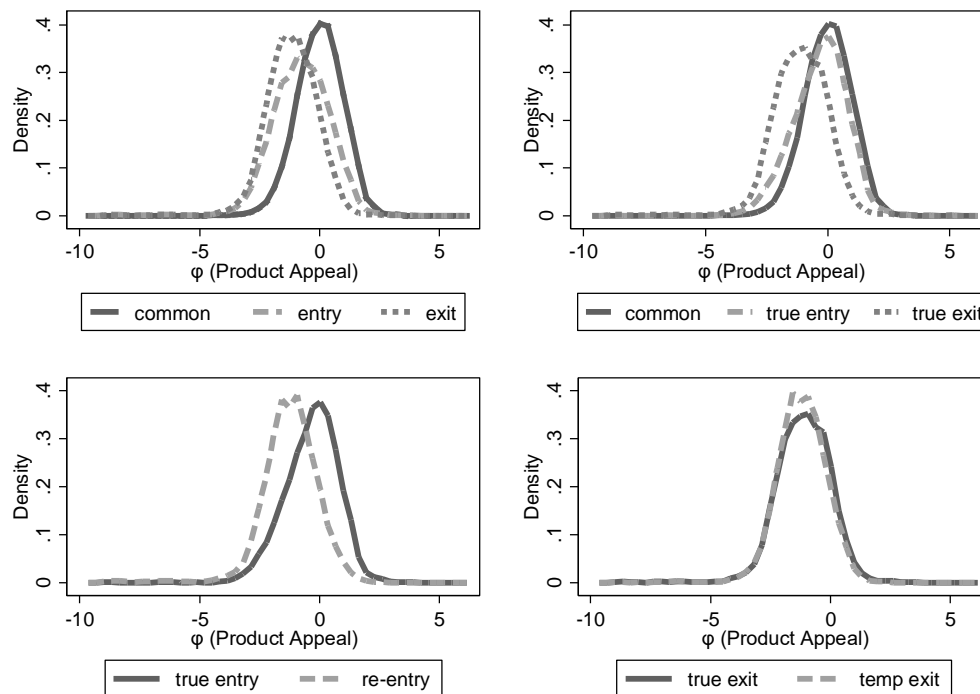


Figure 8: Relative Product Quality by Entry, Exit: Video Games



Note: Nielsen scanner data. True entry is the first quarter a good appears, true exit is the last period a good appears, re-entry is for a good that changed from zero to positive sales in the current period but not true entry and temp exit is for a good that changed from positive sales in the prior period to zero in the current period but the good re-enters at a later period. Reported are kernel density estimates of the distributions of the demand quality/appeal residual (ϕ_{kt}).

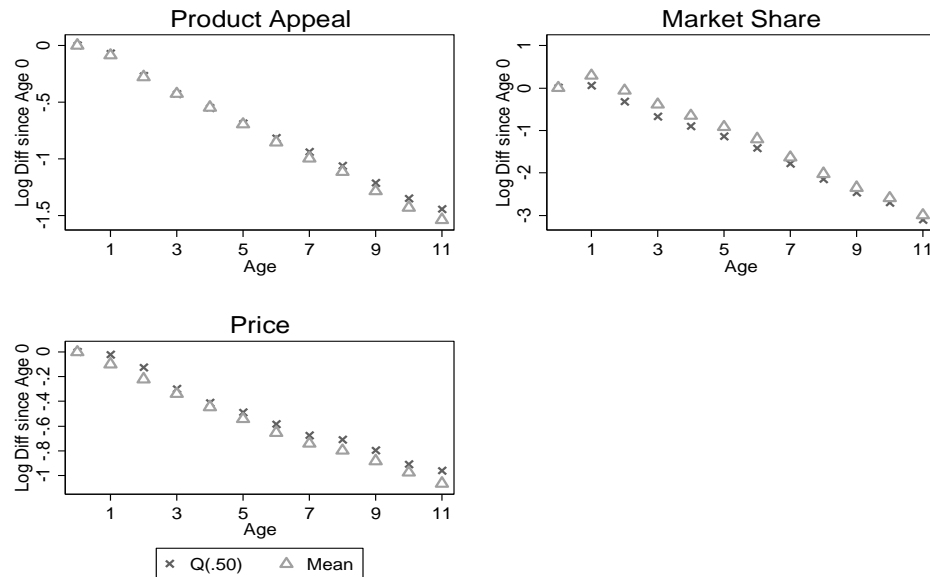
Figure 9: Relative Product Quality by Entry, Exit: Soft Drinks



Note: Nielsen scanner data. True entry is the first quarter a good appears, true exit is the last period a good appears, re-entry is for a good that changed from zero to positive sales in the current period but not true entry and temp exit is for a good that changed from positive sales in the prior period to zero in the current period but the good re-enters at a later period. Reported are kernel density estimates of the distributions of the demand quality/appeal residual (ϕ_{kt}).

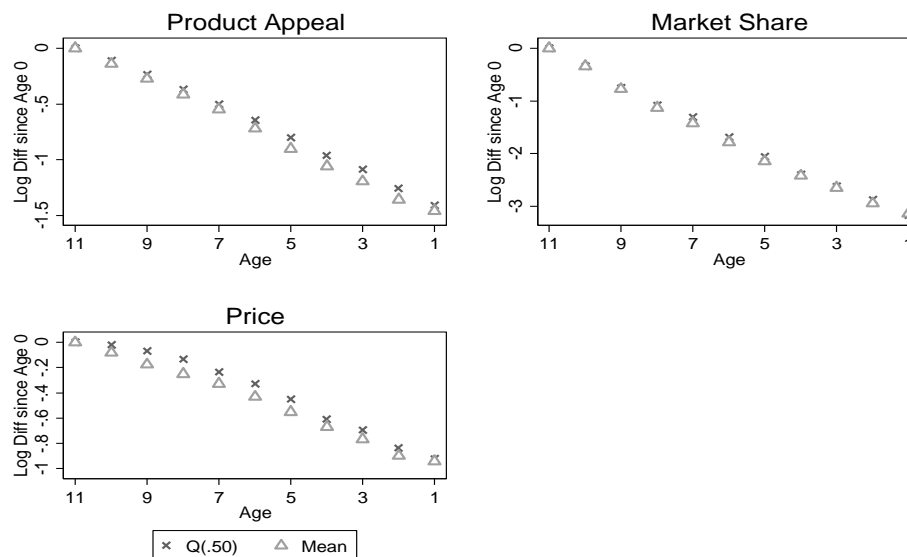
Figure 10: Post-Entry and Pre-Exit Dynamics of Relative Product Appeal, Price and Market Share for Video Games

A. Post-Entry



Note: Nielsen scanner data. Age is number of quarters since entry. Reported statistics are relative to the product's value in its first quarter. Analysis is restricted to items that survive for 11 quarters.

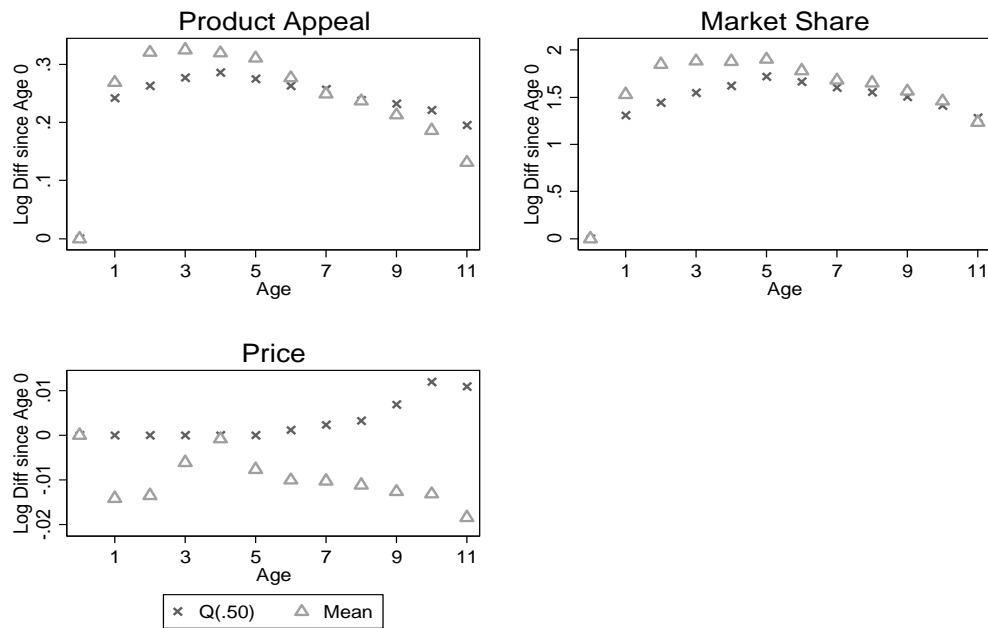
B. Pre-Exit



Note: Nielsen scanner data. Age is number of quarters prior to exit. Reported statistics are relative to the product's value 11 quarters prior to exit. Analysis is restricted to items that are present for 11 quarters prior to exit.

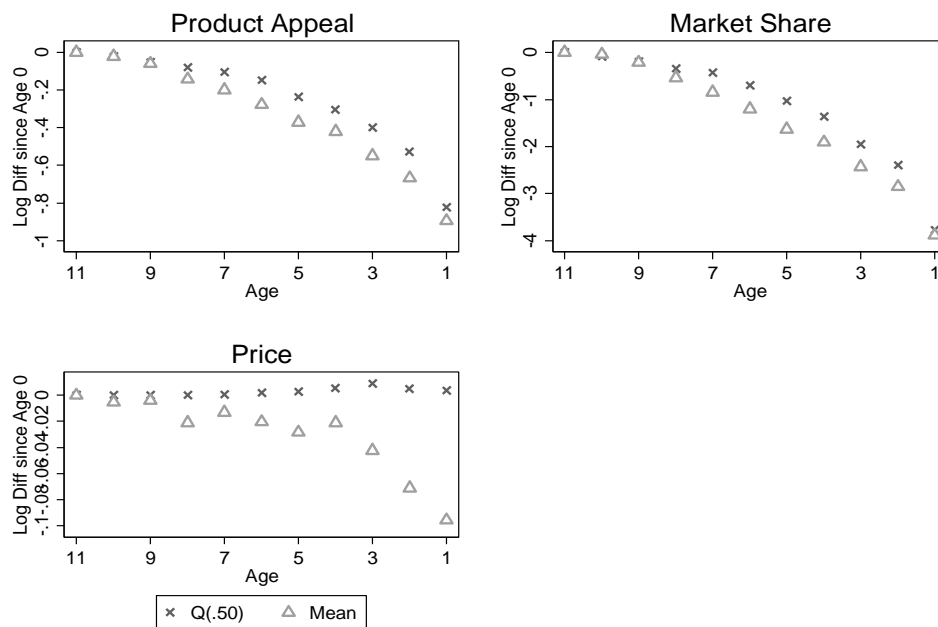
Figure 11: Post-Entry and Pre-Exit Dynamics of Relative Product Appeal, Price and Market Share for Soft Drinks

A. Post-Entry



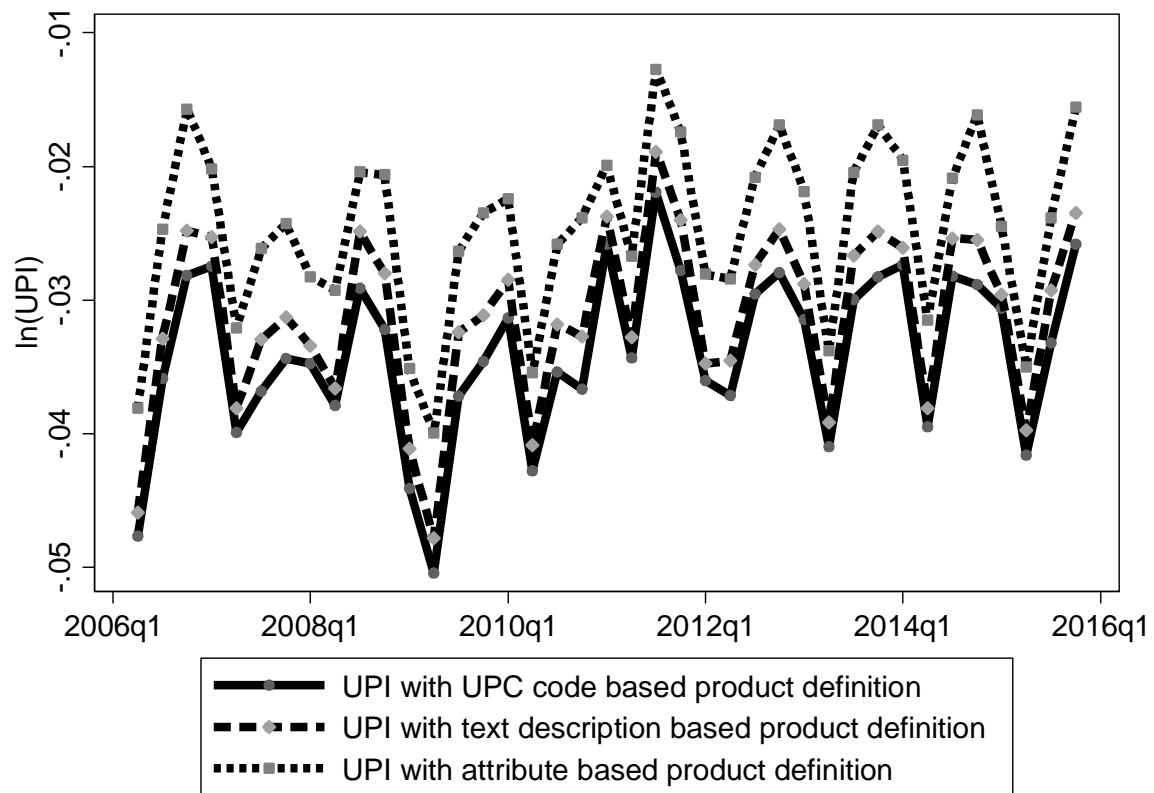
Note: See notes to Figure 10A.

B. Pre-Exit



Note: See notes to Figure 10B.

Figure 12: Sensitivity of UPI to Product Classification



Source: Nielsen scanner data.