Re-engineering Key National Economic Indicators

by

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

March 4, 2019

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.

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. We advocate, in particular, that price and quantity be collected or aggregated simultaneously from the retailer.

---

[1] Erhlich, Haltiwanger, Jarmin, Johnson, and Shapiro (2019) gives a shorten 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 data summarized to the item level should, in principle, allow greatly improved data along many dimensions. First, price and quantity should be based on the same observations. Second, the granularity of data could be greatly increased on many dimensions. 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. High-

frequency data could be constructed at a very fine level of geographical detail. Similarly,

product-level detail could be greatly refined. 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), who build on the traditional Feenstra (1994) product turnover-adjusted Sato-Vartia price index. This approach requires sales and price (i.e., price and quantity data) at the individual product level. Redding and Weinstein's results suggest that 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

4

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. Another, related, lesson from our analysis is that although the two approaches are distinct, both seek to capture quality adjustment from changing product mix and with appropriate implementation may yield similar price and quantity indices.

The first section of this paper reviews the existing paradigm where 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 will require a re-engineering of 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 that 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 take 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 better quality data at potentially 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.[2] A notable feature of the current architecture is that data collection for total 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 survey and Census collects data), (2) selection of outlets based on the Telephone Point of Purchase Survey, and (3) probabilistic selection of a relatively small sample of goods at selected outlets (via the Commodities and Services Survey). The Census Bureau collects sales data from retailers in its monthly and annual surveys. The monthly survey is

---

[2] Table 1 is a vast 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.

voluntary and has suffered from declining response rates. The monthly survey also has irregular responses across 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 the nominal sales and price level by BEA is 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.

The 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 we present 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.[3] 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.[4] To accomplish this, we first aggregate the weekly data to monthly data using the

National Retail Federation (NRF) calendar, which controls for changes in the timing of holidays

and the number of weekends per month. 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 second major data source we use is the NPD scanner data. The NPD data covers

more than 65,000 general merchandise stores, including online retailers, and includes products

that are not included in the Nielson scanner data. We currently restrict ourselves to the analysis

of one detailed product module: memory cards.[5] The NPD raw data are at the item-by-store-by-

---

[3] 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.
[4] 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.
[5] The NPD data also includes coffeemakers and headphones, which we plan to explore in future research.

month level. (NPD has aggregated the monthly data from weekly data using the NRF calendar, as we do ourselves in 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 approaches to *P* and *Q* measurement 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.[6]

A. *Nominal Revenue Indices*

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.

---

[6] 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.

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 the table 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.[7] 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

---

[7] 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.

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 likely reflects nonfood items.

Considering the estimates of PCE highlights the advantages of item-level data that yield detailed product class information at high. The high-frequency data underlying PCE come from the MRTS, which as we have seen, provides estimates by type of outlet, not by product. Every five years the Economic Census (EC) yields information on sales at the establishment level by detailed product class. In the intervening time periods, the Annual Retail Trade survey (ARTS) and the MRTS survey firms for their total sales, classifying firms into major kind of business (e.g., 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.

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 Figure 1 to produce statistics on more detailed food products (e.g., poultry).[8]

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

---

[8] 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.

apply at business cycle frequency and for business cycle shocks, with the potential for the current system to either overstate or understate cyclical fluctuations depending on the product mix across outlets and their cyclicality sensitivity.

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 item that it is matched to, then an attempt is made to adjust the prices to account for the change in features). The sample of products in the CPI consumption 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 Redding and Weinstein (2018) who build on the earlier work by Feenstra (1994). The

13

second approach uses hedonic methods, following the insights of Pakes (2003, 2005) and Erickson and Pakes (2010). While these hedonic approaches are already partly in use by BLS and BEA, the item-level transactions data offers the potential for implementing these approaches at scale. Bajari et al. (2018) is an initial attempt to implement hedonics at scale using a rich set of product attributes.

It is useful to introduce some notation to clarify our empirical analysis of these two approaches. (Appendix A provides further technical details.) Redding and Weinstein (2018) use a CES demand structure at the product group level to generate the UPI, which is given by:

$$\text{UPI} = PV_{adj}^{\frac{1}{\sigma-1}} \ CV_{adj}^{\frac{1}{\sigma-1}} \ \text{RPI}$$

where PV_{adj} is a product variety adjustment bias term based on Feenstra (1994), CV_{adj} is a consumer valuation bias adjustment term that is novel to the UPI, 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*, and σ is the elasticity of substitution across goods within a product group. The UPI is designed to be implemented at a product group basis, which the item-level transactions data permit. Critical issues for the implementation of this approach include determining the appropriate number of nested product groups that are appropriate and defining such product groups.

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. 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.

14

Feenstra (1994) modified the Sato-Vartia price index that is consistent with a CES demand structure to include the PV_{adj} term.[9]

The consumer valuation bias adjustment term in Redding and Weinstein's UPI 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 consistent with consumer demand theory, which recognizes that relative product appeal can change over time, even for a given item.[10]

The elasticity of substitution is again a critical factor for the quantitative relevance of the CV term. If relative product appeal shifts between very close substitutes, the CV adjustment term becomes quantitatively less important. In interpreting observed changes in relative product appeal, they are related to the evolution of product quality via entry and exit. In this respect, the observed changes in product appeal may reflect factors such as learning dynamics about new products.

Implementation of the UPI requires an estimate of the elasticity of substitution at the product group level. Redding and Weinstein (2018) consider multiple approaches for estimation including the reverse weighting, generalized reverse-weighting and the Feenstra (1994) estimators. In our analysis below, we focus on the Feenstra (1994) estimator although note that

---

[9] The Feenstra price index is not simply the UPI holding the CV term constant. Instead, as discussed in Appendix A, 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).

[10] 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.

Redding and Weinstein (2018) find sensitivity of the distribution of estimated elasticities to alternative estimators. We discuss these estimation challenges below.

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

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

where X_i is vector of characteristics or attributes of good i. Note that the attributes are time-invariant 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 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 quality-adjusted by hedonics index at the product group level is given by

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

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. 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 incorporating product turnover.

III. Results

We implement both the UPI and the hedonics approach and compare the 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 permits us to implement both the UPI and hedonics approaches. We also have obtained BLS CPI estimates that correspond to the product groups in the Nielsen scanner data aggregated to food and nonfood items. Thus, for the analysis of the Nielsen scanner, we compare the BLS CPI to the price indices from the transactions data.

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. Since the quantitative implications of the PV and CV terms depend critically on these estimates, Table 3 presents the summary statistics for the estimated elasticities for the more than 100 product groups in the Nielsen

scanner data. The median elasticity is about 8 but there is considerable variation, with the 10th

percentile at about 4 and the 90th percentile about 16.[11]

Table 4 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.[12] 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 4 provides summary statistics using the

quarterly food and nonfood price indices.

The top panel of Table 4 shows the results for the food product groups and the lower

panel the nonfood product groups. Each panel displays 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.

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

---

[11] The median estimate is similar to that reported in Table 2 of Redding and Weinstein (2018) using either the Feenstra or their generalized reverse weighting estimator.
[12] 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.

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 B.1 presents the analogous statistics to Table 4 comparing the BLS CPI for food and nonfood items (covering the same product groups) to the Consumer Panel-based price indices.[13] 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

---

[13] We estimate the elasticities of substitution separately for the Consumer Panel data.

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.[14]

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 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. Within scanner's product groups are arguably goods that are very close

substitutes, while others are more differentiated. For product turnover and expenditure share

volatility with close substitutes, the quality adjustment factors in the Feenstra and UPI indices

become very small. The procedure used in Table 4 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

---

[14] 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.

memory cards. We have estimated quarterly hedonic regressions relating an item's log price (at the national quarterly level) to its attributes. We use a quadratic in memory card size and speed and dummy variables capturing card types (e.g., flash cards, memory chips, etc.). The dataset contains about 12,000 product items by quarter observations at the national level.

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. Table 5 shows the quadratic regression results, demonstrating that prices are increasing in size and speed but at a decreasing rate. Figure 2 shows that the marginal value of additional size and speed appears to be declining over time. An R-squared of about 0.8 in each quarter suggests that there could be other unobserved characteristics that affect the prices.

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 (see Table 6). Following Redding and Weinstein (2018), 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).[15] As expected, Figure 3 shows that entering goods have substantially higher average quality than exiting goods, but they also have much more dispersion in quality. Greater dispersion at entry suggests potentially interesting post entry dynamics of entry that may involve selection and learning dynamics. We explore this below.

Figure 4 summarizes means and standard deviations of the alternative price indices for memory cards and Table 7 shows their correlations.[16] Quality-adjusted prices are declining

---

[15] See also Appendix A. The value at entry or exit is defined as the φ_{kt} in the first or last quarter the item is sold.
[16] We compute these indices for each quarter. Figure 4 and Table 7 present statistics calculated using the quarterly indices.

21

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 hedonics approach also permits changing valuations of characteristics that could capture the variation in the CV. Even though we find that the hedonics 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 is capturing. To explore this issue, we conduct some further exploratory analysis. Figure 5 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). 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 6 shows the UPI components for two narrow product modules that highlight the variation depicted in Figure 6. The top panel shows the (log of) UPI and its components (PV, CV and RPI—again these are in logs and for PV and CV are multiplied by $\frac{1}{\sigma-1}$) for soft drinks. 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. For video games, the PV is large in magnitude and negative on average, as is the CV. Video games thus fits the pattern of Figure 5 in that both PV and CV are large in magnitude.

The scale of Figure 6 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.[17] 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 6).

Figure 5 and Figure 6 suggest 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

---

[17] 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.

simulations of product entry and exit with associated changes in product quality presented in

Figures 7, 8, and 9. 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. 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. Thus, all of the variation in

appeal shows up in appeal-adjusted prices. In each figure, Panel A shows the relative quality and

Panel B shows the implied price indices.

In Figure 7, 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 Panel A. Panel B

of Figure 7 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 8, new products exhibit post-entry dynamics as upon entry their relative

product appeal is 1.0 and then in the 2nd period after entry their appeal rises to 1.2. Product exits

exhibit pre-exit dynamics going from relative product appeal of 1.0 two periods before exit to 0.8

in the period prior to exit. These patterns are depicted in panel A.   The bottom panel of Figure 8

shows the implied Laspeyres, Feenstra and UPI log price indices (making the same assumption

of constant unadjusted prices). 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 9 depicts an alternative simulation where the post entry buildup of appeal is slower (i.e., products enter at relative appeal 0.8 and build to 1.2 by period 3). 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 captures a larger share of the dynamics.

While these are simple illustrative simulations, they highlight a potential important driving force distinguishing the UPI from the Feenstra is 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 7-9 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.

Figure 10 and Figure 11 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.[18] Both figures display the distributions of relative appeal of continuing, entering, and

exiting goods. The upper-left-hand panel of Figure 10 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 11 shows the analogous patterns for soft

drinks. Qualitatively, the patterns are similar, but they are substantially less pronounced,

---

[18] As with Figure 3, this uses the estimate of φ_{kt} as discussed in Appendix A.

consistent with the notion that 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 12 and Figure 13. The figures display patterns for 11 quarters after entry and 11 quarters prior to exit.[19] 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 7 through 9 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 dynamics relative to the median. At the mean, relative quality rises more than 30 log points in

---

[19] 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.

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.

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. [20] 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 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

---

[20] In Redding and Weinstein (2018), the demand residual for good k a time t is denoted by φ_{kt}. As they show, once one has an estimate of the elasticity of substitution for a product group (σ) then φ_{kt} can be measured as the residual given measures of expenditures shares and prices. See Appendix A for more details.

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 estimation 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

30

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 14 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.[21] For each UPI considered, all products within an aggregation (e.g., text descriptions or attributes) are treated as perfect substitutes. Strikingly, but perhaps not surprisingly, the UPI becomes substantially less negative using these more aggregated product definitions. While this exercise is simple, it demonstrates that important components of the UPI (via the PV and CV terms) reflect variation within 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

---

[21] 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.

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 by

using the same attributes approach as in the hedonics to generate product classification and nests

the UPI and hedonic indices will 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 is available 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, and it is unclear that this approach can be used for more than two nests.[22]

Finally, the approach to hedonics with item-level transactions data based on Bajari and Benkard (2005) 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, may have less readily available information on item attributes.

Machine learning approaches as in Bajari et. al. (2018) could in principle be used to overcome these issues. Bajari et. al. (2018) convert text and images to vectors along with dimensionality reduction techniques to estimate hedonic adjusted prices at scale for millions of products at a high frequency. They show that their approach yields high R-square in the estimation of hedonic functions. An open question in this area is how to interpret the nature of the bounds Pakes (2003) emphasizes for hedonics. The tightness of the bounds depends on the extent to which the estimated relationship between prices and attributes reflects quality variation vs. markups. The literature has not yet grappled with how to evaluate such issues at scale.

IV. Re-engineering the Data Architecture

---

[22] Redding and Weinstein (2018) argue that the reverse weighting and generalized reverse weighting estimators they develop can be used for multiple nests. The nature of the biases from these estimators are still not well understood. The reverse weighting estimator is consistent only under the implausible assumption that prices are independent of demand shocks.

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. These include issues of privacy, confidentiality, and value of business data; cost to businesses and the statistical agencies of the new architecture; 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 of 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 system of firms. Doing so would, as is typical for IT innovations, 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 the producing the indices discussed in this paper. In particular. The first option of direct feed of transactions and, potentially, the third option of third-party aggregators, potentially allows 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 business 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 surveys of statistical agencies. 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.

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. Implementing this

architecture where the collection and aggregation of data components is spread across three

agencies is 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 fundamental separates 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. One agency negotiating access to transaction level data will

be difficult enough. Multiple agencies doing so, unduly burdens both businesses and the

taxpayer. Importantly, under the current statistical system, no agency has the mandate both to

collect data on price and quantities, so implementing the data architecture to simultaneously

measure price and quantity is out of scope of each of the agencies.[23]

What sort of legal and policy structure is needed govern how statistical agencies access

private data assets for statistical uses? For instance, a key question is whether companies would

---

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

seek to charge for access to the type of data described above and, if so, would the associated fees be within the budgetary resources of the statistical agencies or other public-spirited organizations committed to high quality economic statistics?

To further test, develop and implement a solution like 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 get more staff with 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 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 statistical products from blended data. Information systems 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 decisionmakers.

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. Our review and exploration of the methodological

innovations is that integration of 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.

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)

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

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.

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* (forthcoming).

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,* forthcoming.

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

Redding, Stephen and David Weinstein. 2018. "Measuring Aggregate Price Indexes with
Demand Shocks: Theory and Evidence for CES Preferences," Unpublished paper, Columbia
University.

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

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

Census (nominal spending)	**BLS (prices)**
Data collection: Retail Trade surveys (monthly and annual) Economic Census (quinquennial) Consumer expenditure survey (conducted for BLS)	**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)
Published statistics: Retail Trade (monthly and annual) by firm type Retail Trade (quinquennial) by product class	**Published statistics:** Consumer Price Index (monthly) by product class

BEA (aggregation and deflation)

Data collection:
Census and BLS data; supplemented by multiple other sources

Published statistics:
Personal Consumption Expenditure: Nominal, real, and price (monthly)
GDP (quarterly)

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—Surveys vs. Scanner Data

	Scanner	Census MRTS (Grocery)	PCE
A. Seasonally Adjusted			
Mean	0.87	0.74	0.78
Standard deviation	0.98	0.64	0.61
Correlations:			
Scanner	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 3: Estimates of Elasticity of Substitution from Scanner Data

Mean	9.09
Median	8.06
10th percentile	4.14
90th percentile	16.10
Standard deviation	4.98

Notes: The elasticities of substitution are estimated using the Feenstra (1994) method for 110 product groups in the Nielsen scanner data.

Table 4: Summary Statistics on Comparisons of 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.

Table 5: Summary Statistics from Hedonic Regressions by Quarter: Memory Cards

	Size	Size squared	Speed	Speed Squared
Mean Estimated Coefficient	0.01888	-0.00003309	0.005487	-0.000002349
Mean Std Error	0.001368	0.000004341	0.001517	0.000005885
Std Dev of Estimated Coefficients Across Quarters	0.005935	0.00002273	0.001409	0.000003915

Notes: Estimates based on NPD data for memory cards. Hedonic regressions of log(price) regressed on quadratic in size and speed along with attribute dummies (not reported) estimated by quarter from 2014 to 2016. Reported statistics are means of estimated coefficients, standard errors and standard deviations of coefficients over time.

Table 6: Estimate of Elasticity of Substitution: Memory Cards

Coefficient Estimate	4.207
Std Error	(0.5835)
R-squared	0.2973

Note: Estimates from NPD data using the Feenstra (1994) method. Pooled item level quarterly data from 2014 to 2016.

Table 7: Correlations of Alternative Price Indices: Memory Cards

	Laspeyres	Feenstra	Hedonic (Laspeyres)	Hedonic (Paasche)	UPI
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
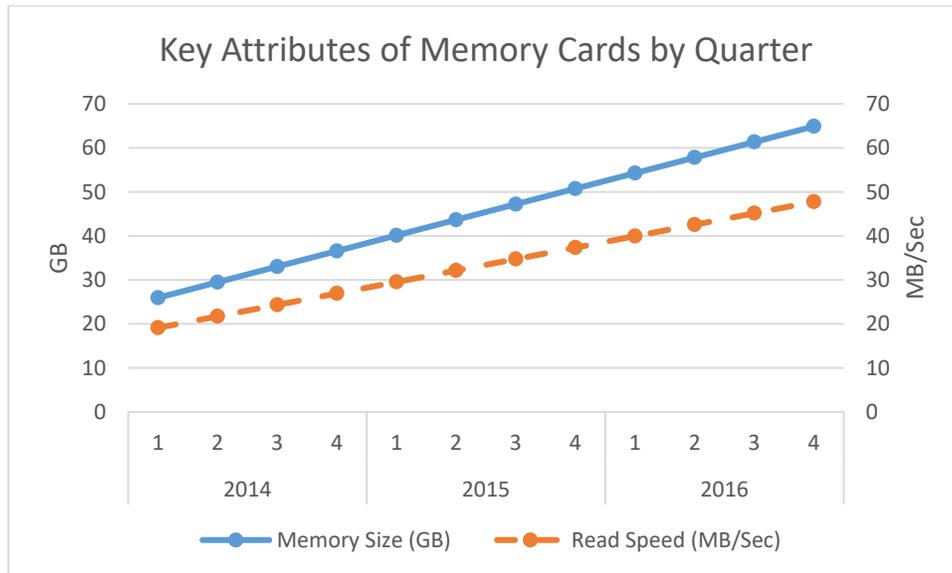
Note: NDP data. The figure shows estimated linear trends in sales-weighted national memory size and read speeds used to produce the estimates in Table 5.

Figure 2

Changing Marginal Value of Attributes by Quarter
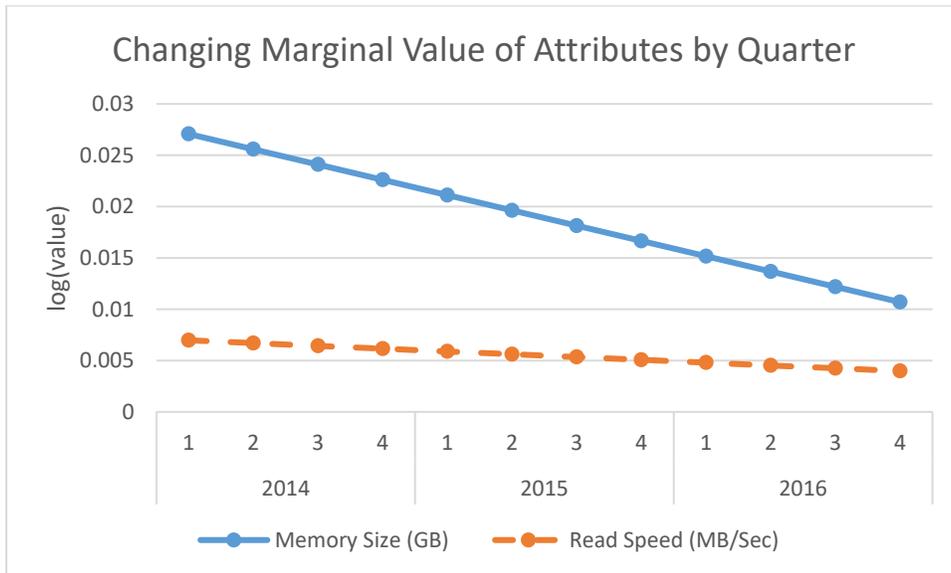
Note: NPD data. This figure shows the changing marginal value (from linear term from Table 5) in estimation of hedonic specification for memory cards

Figure 3: Relative Quality of Common, Entering and Exiting Goods: <mark>Memory Cards</mark>

A. Mean Differences with Common Goods
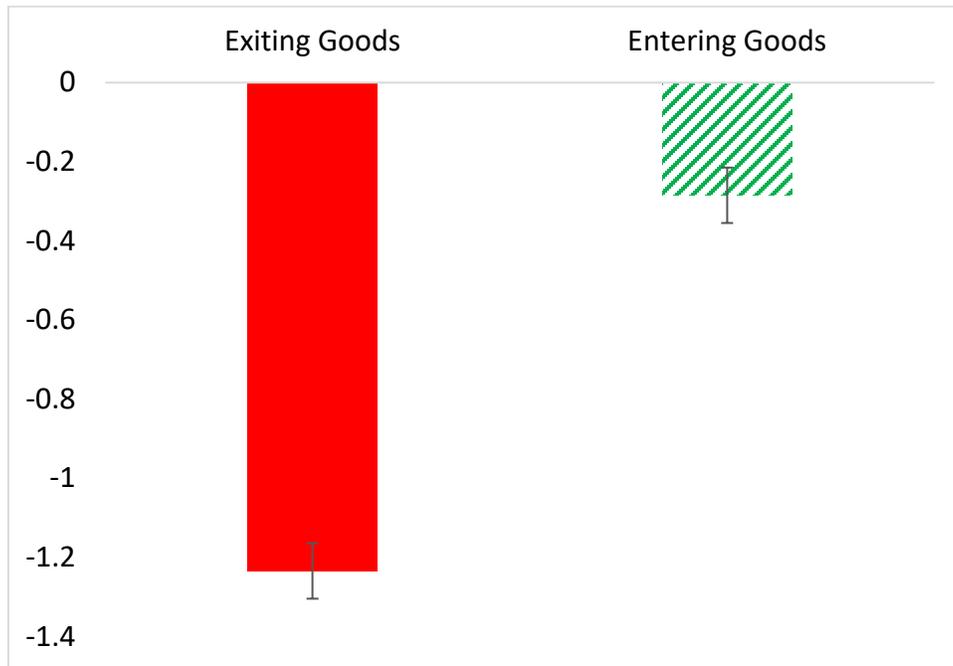
B. Dispersion (Std Dev) in Relative Quality
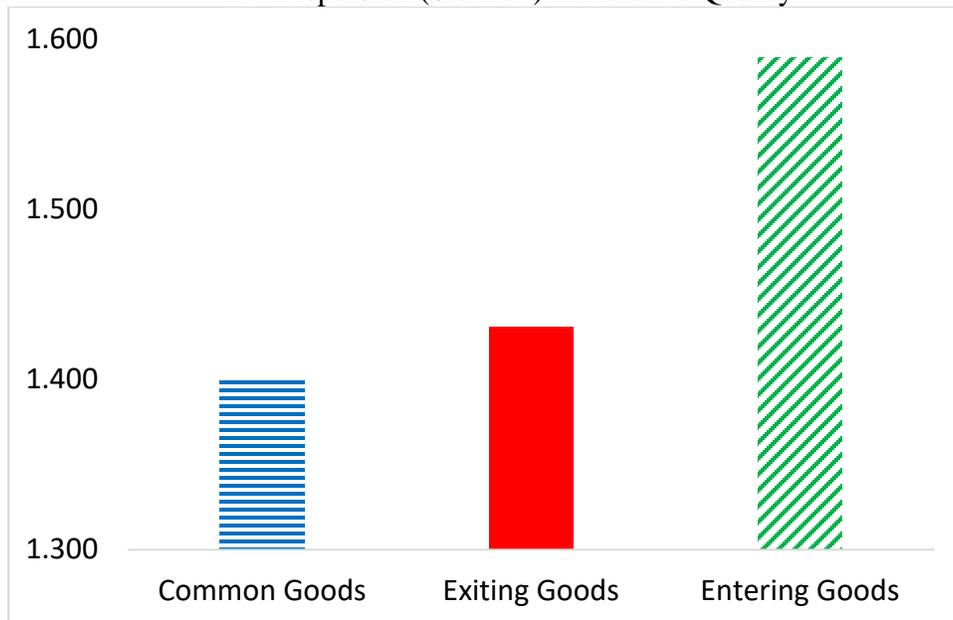
Note: NPD data. Reported summary statistics of the demand quality/appeal residual (φ_{kt}) as described in Appendix A.

Figure 4: Alternative Price Indices: Memory Cards

A. Means

B. Standard Deviations
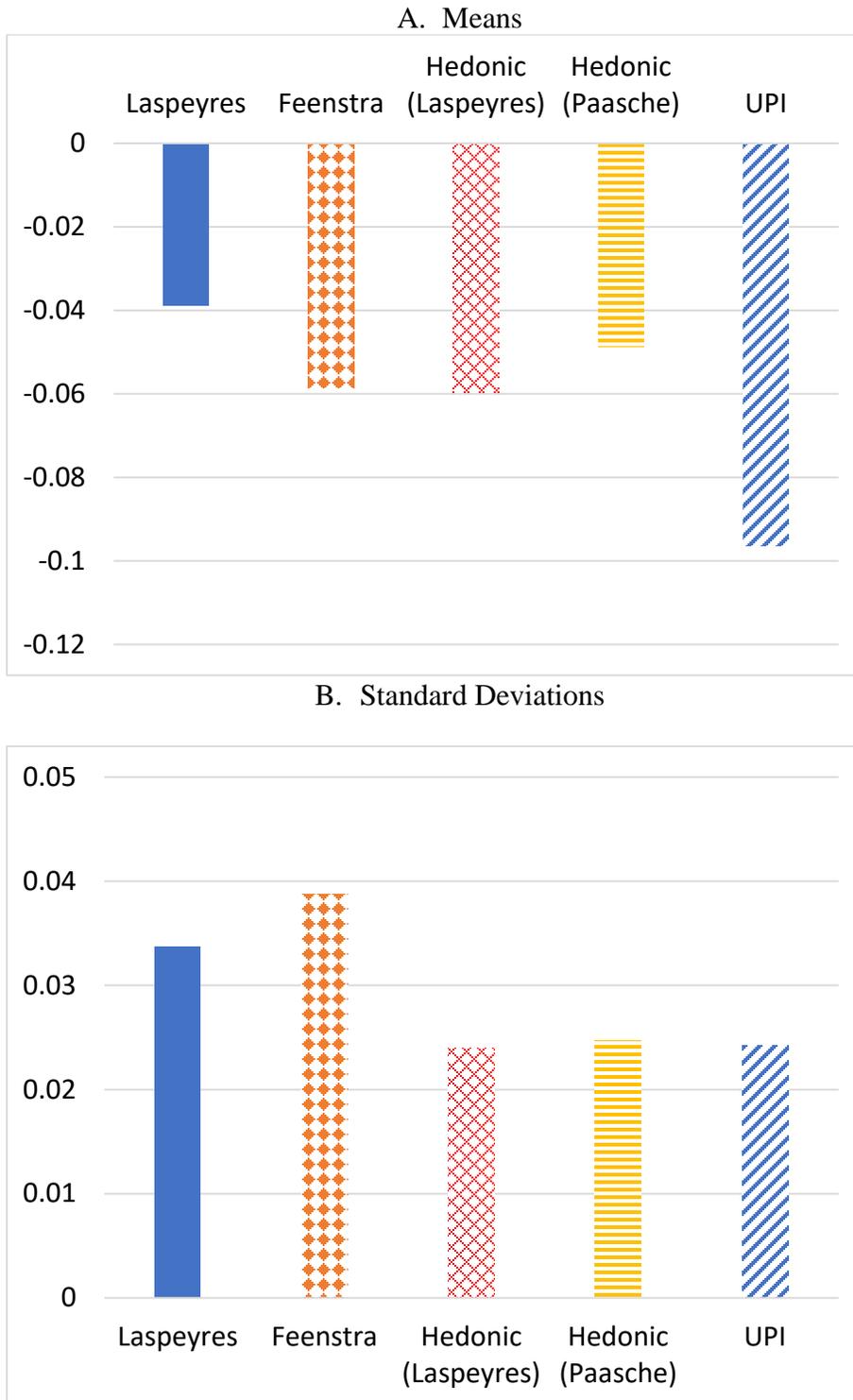
Note: NPD data at item-level quarterly from 2014 to 2016. Price indices constructed at a quarterly frequency. Reported statistics are summary statistics from quarterly indices (not seasonally adjusted).

Figure 5: The Relationship between Product Variety Bias and Consumer Valuation Bias
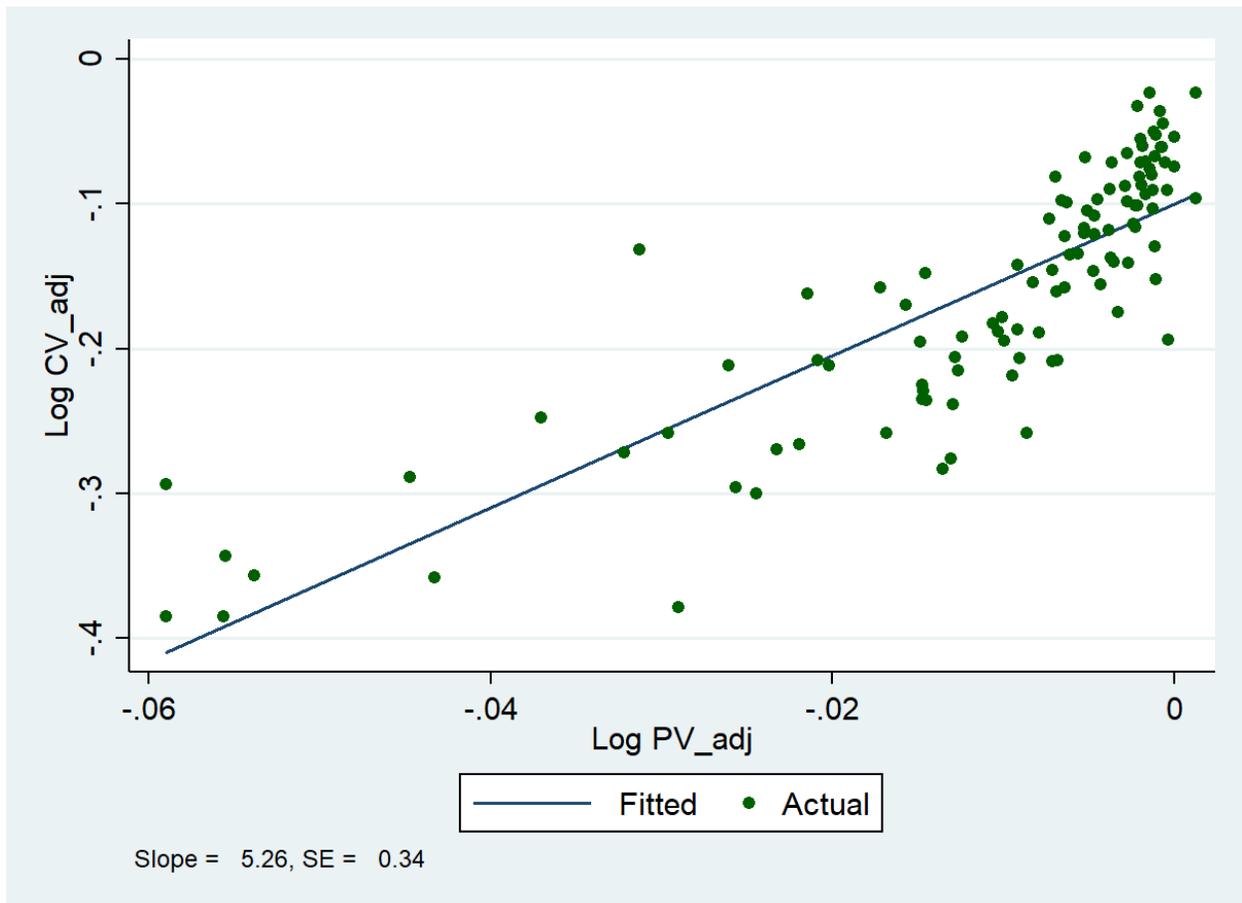
Note: Nielsen scanner data. Each dot represents a product group showing average PV and CV adjustment factors from quarterly measures of PV and CV. Quarterly series from 2006:2-2015:4.

Figure 6: UPI Components for Specific Product Modules

A. Soft Drinks

B. Video Games
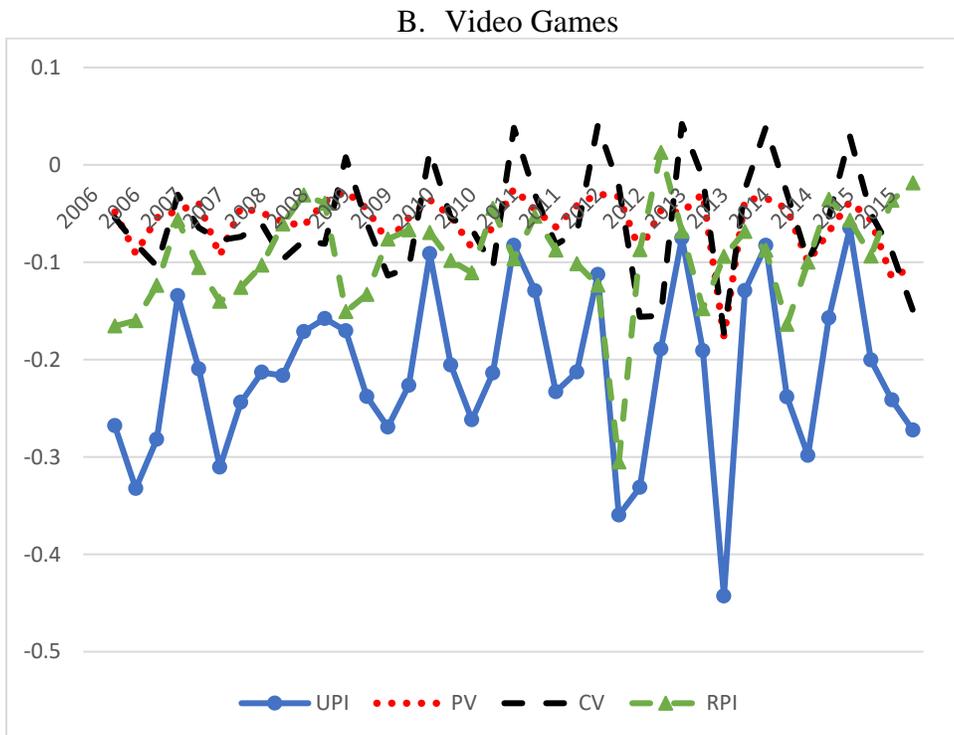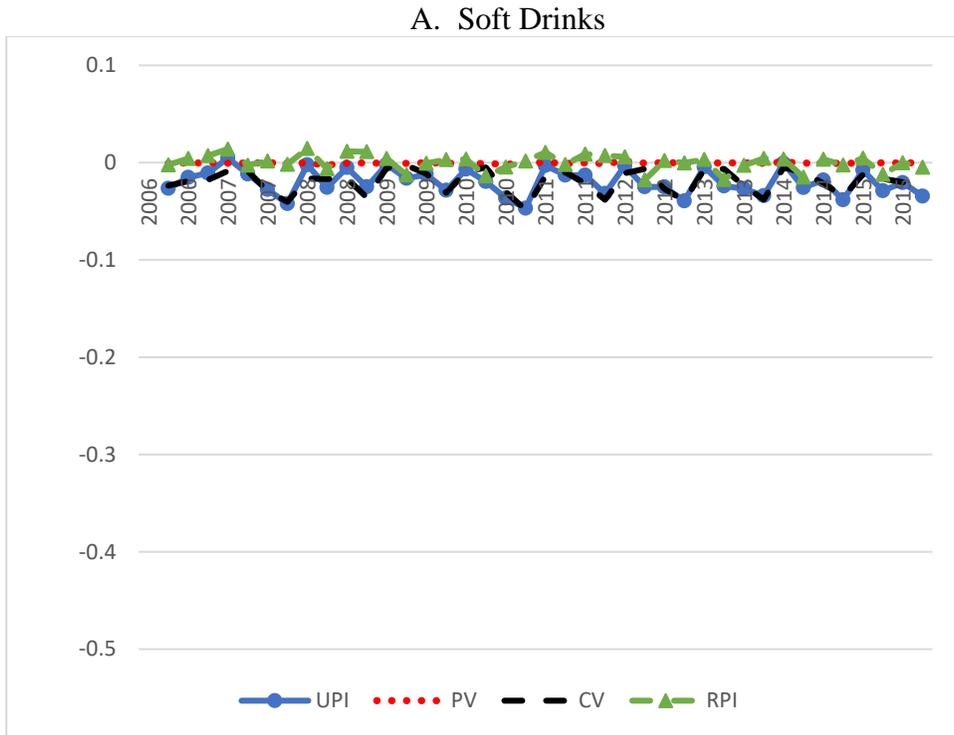
Note: Nielsen scanner data. Log differences at quarterly rate.

Figure 7: Simulated Product Entry and Exit with Quality Change: No Continuing Good Quality Change

A. Relative Product Quality
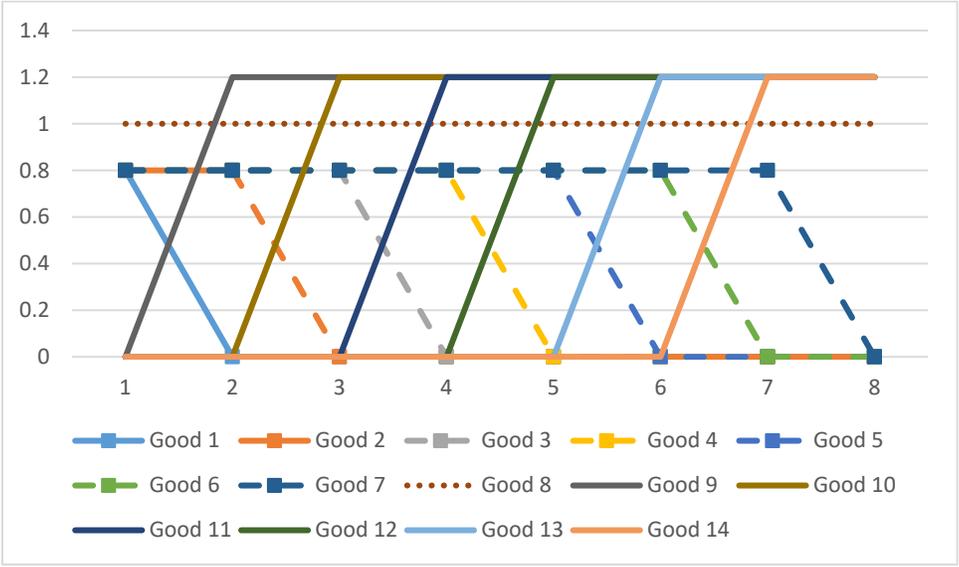
B. Implied Price Indices
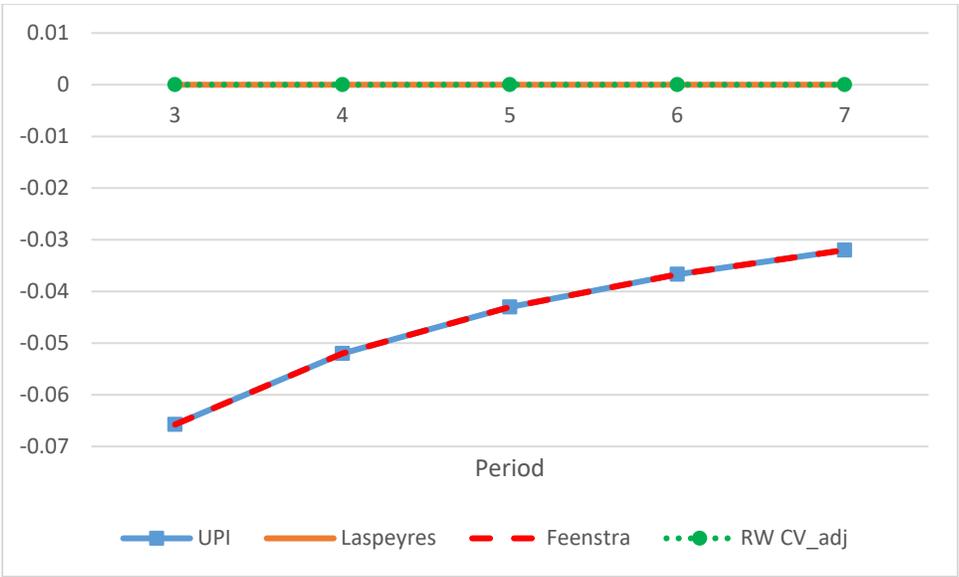
Figure 8: Simulated Product Entry and Exit with Quality Change: Post Entry and Pre-Entry
Quality Dynamics

A. Relative Product Quality
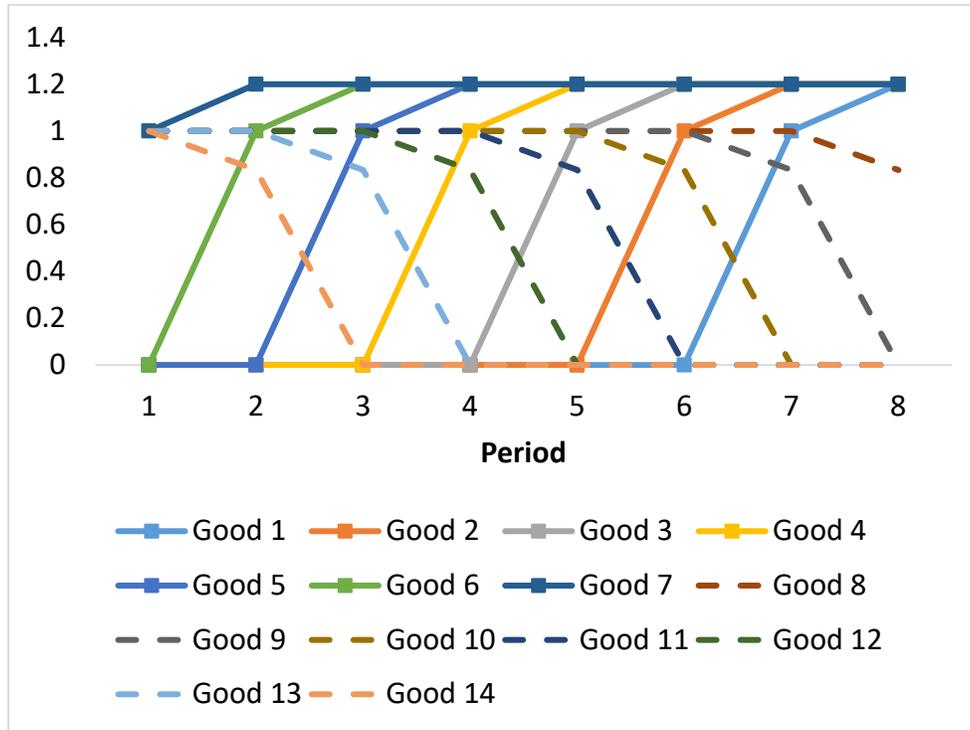
B. Implied Price Indices
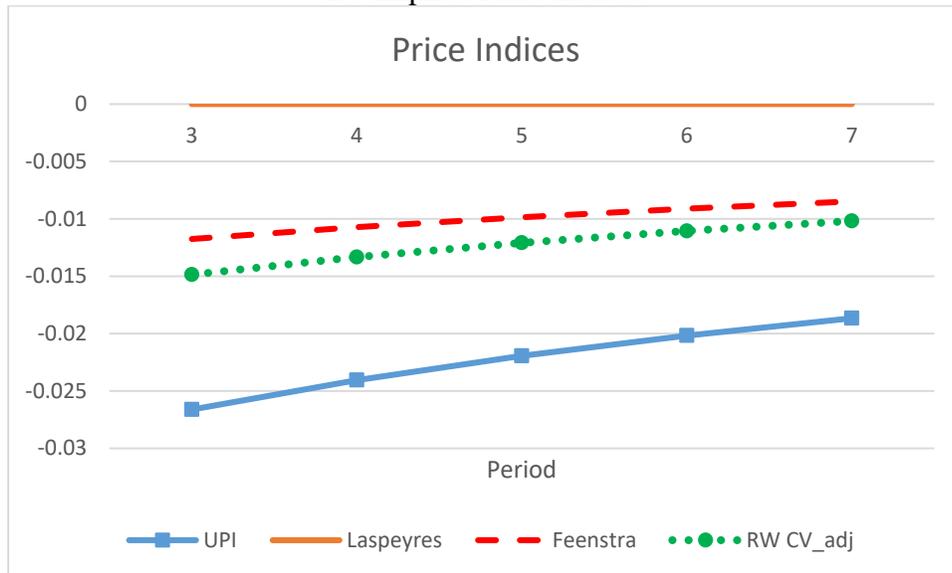
Figure 9: Simulated Product Entry and Exit with Quality Change: Slower Post-Entry and Pre-Exit Quality Dynamics.

A. Relative Product Quality

B. Implied Price Indices
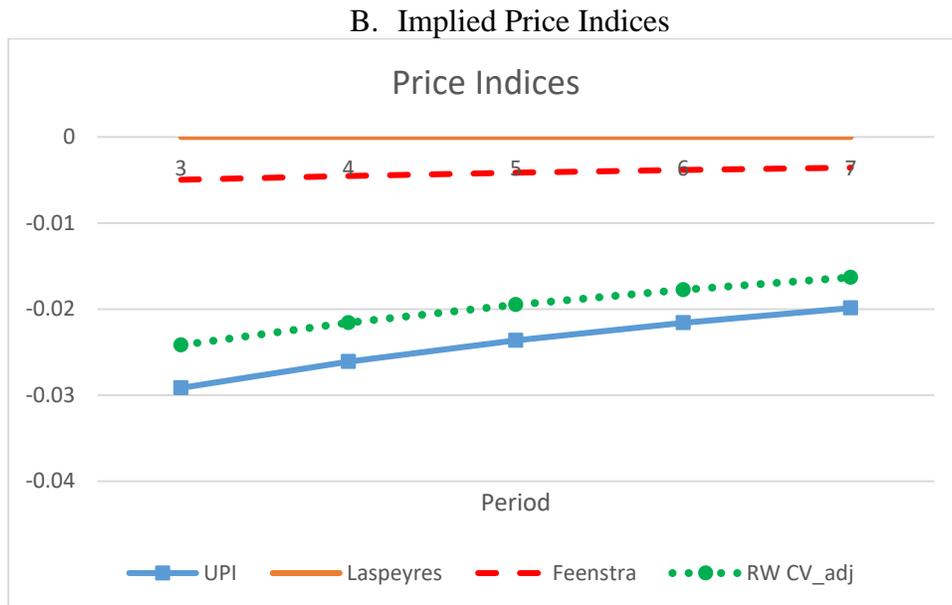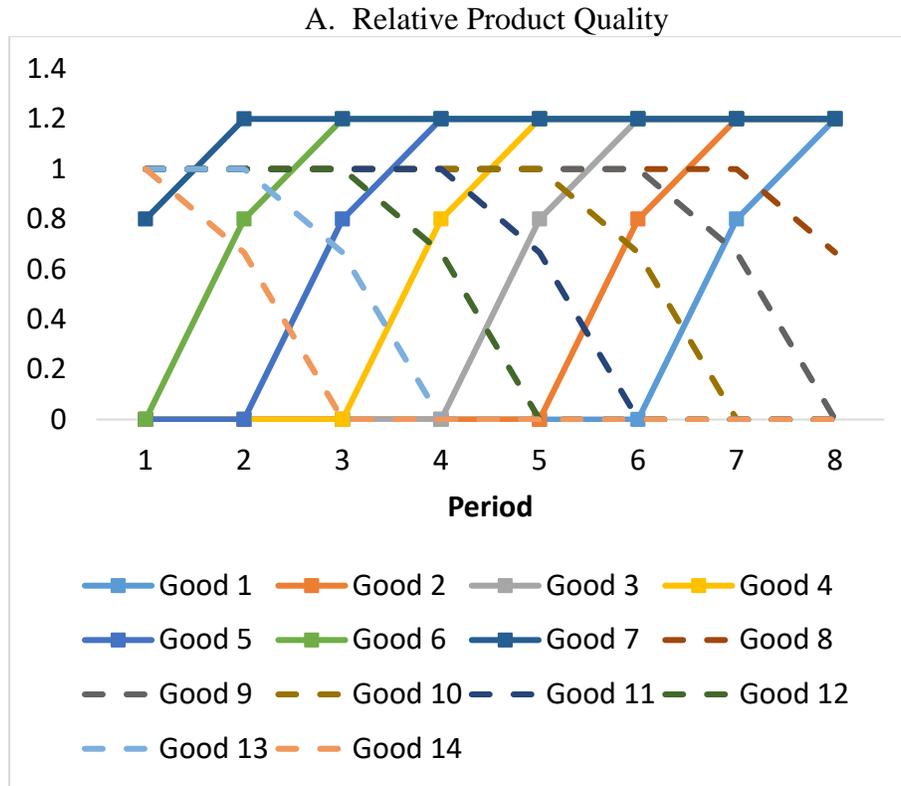
Figure 10: Relative Product Quality by Entry, Exit: Video Games
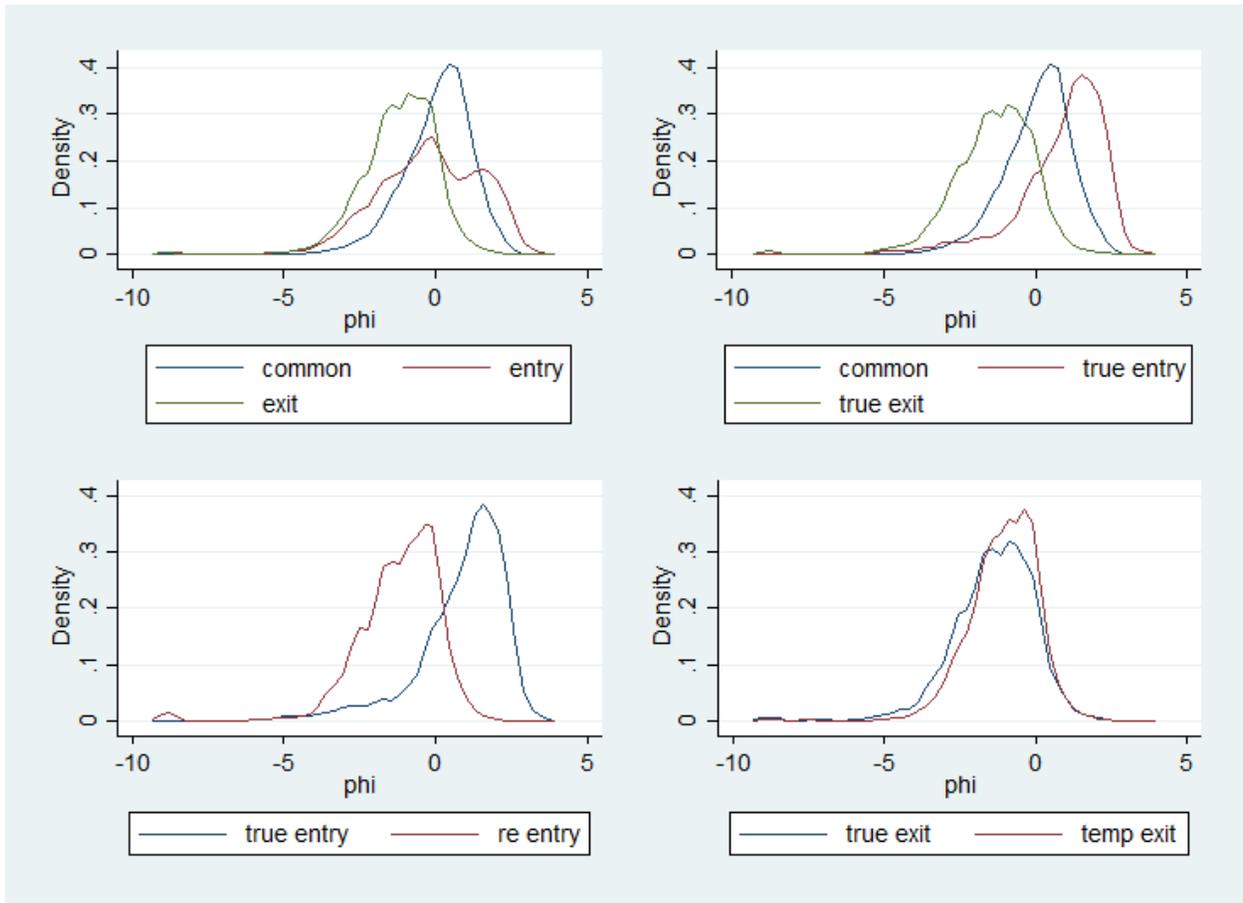
Note: Nielsen scanner data. True entry is the first quarter a good appears, true exit is the last period a good appears, re-entry is for a good that changed from zero to positive sales in the current period but not true entry and temp exit is for a good that changed from positive sales in the prior period 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}) as described in Appendix A for these categories.

Figure 11: Relative Product Quality by Entry, Exit: Soft Drinks
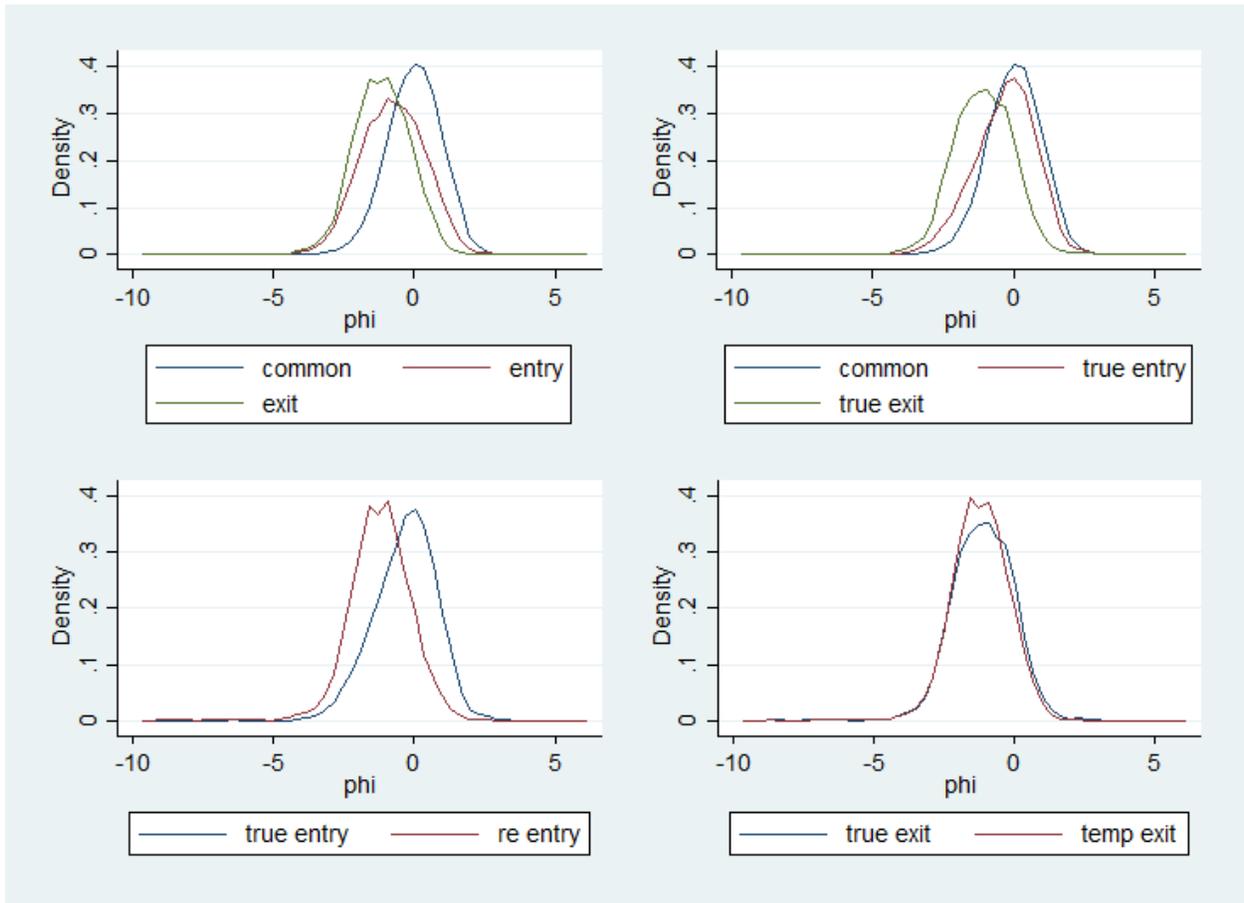
Note: Nielsen scanner data. True entry is the first quarter a good appears, true exit is the last period a good appears, re-entry is for a good that changed from zero to positive sales in the current period but not true entry and temp exit is for a good that changed from positive sales in the prior period 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}) as described in Appendix A for these categories.

Figure 12: Post-Entry and Pre-Exit Dynamics of Relative Product Appeal, Price and Market Share for Video Games

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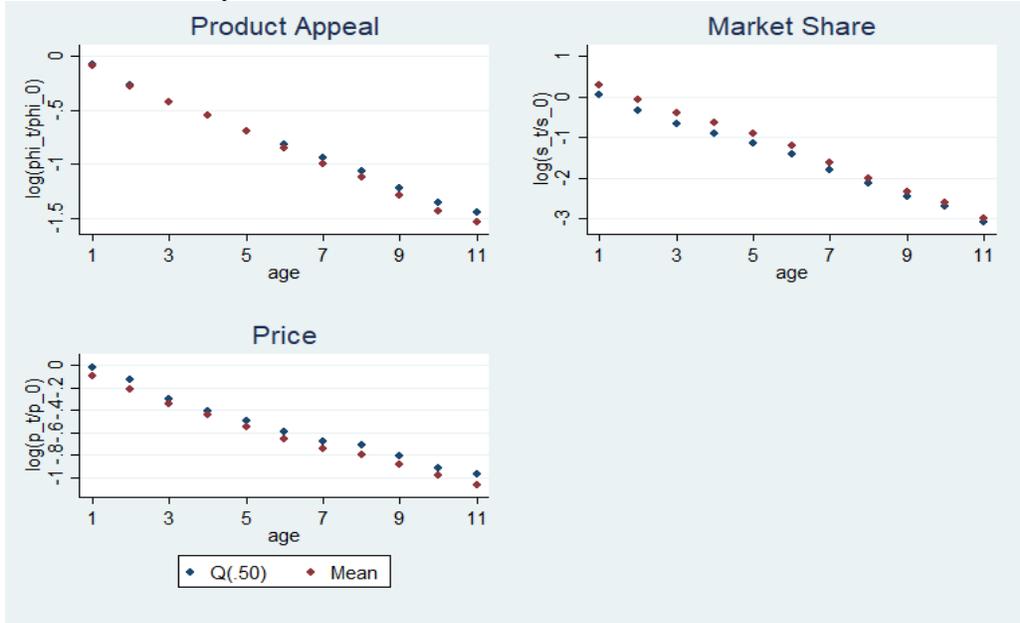
Note: Nielsen scanner data. Age is number of quarters since entry. Reported statistics are relative to the product's value in its first quarter. Analysis is restricted to items that survive for 11 quarters.

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 13: 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 12A.

B. Pre-Exit
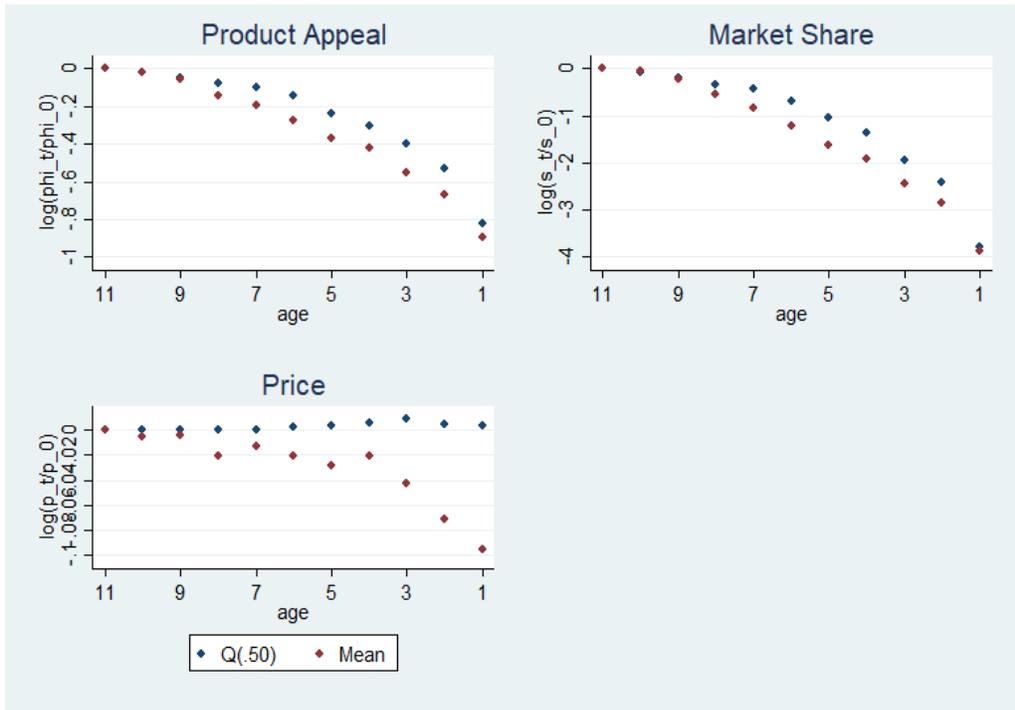
Note: See notes to Figure 12B.

Figure 14: Sensitivity of UPI to Product Classification
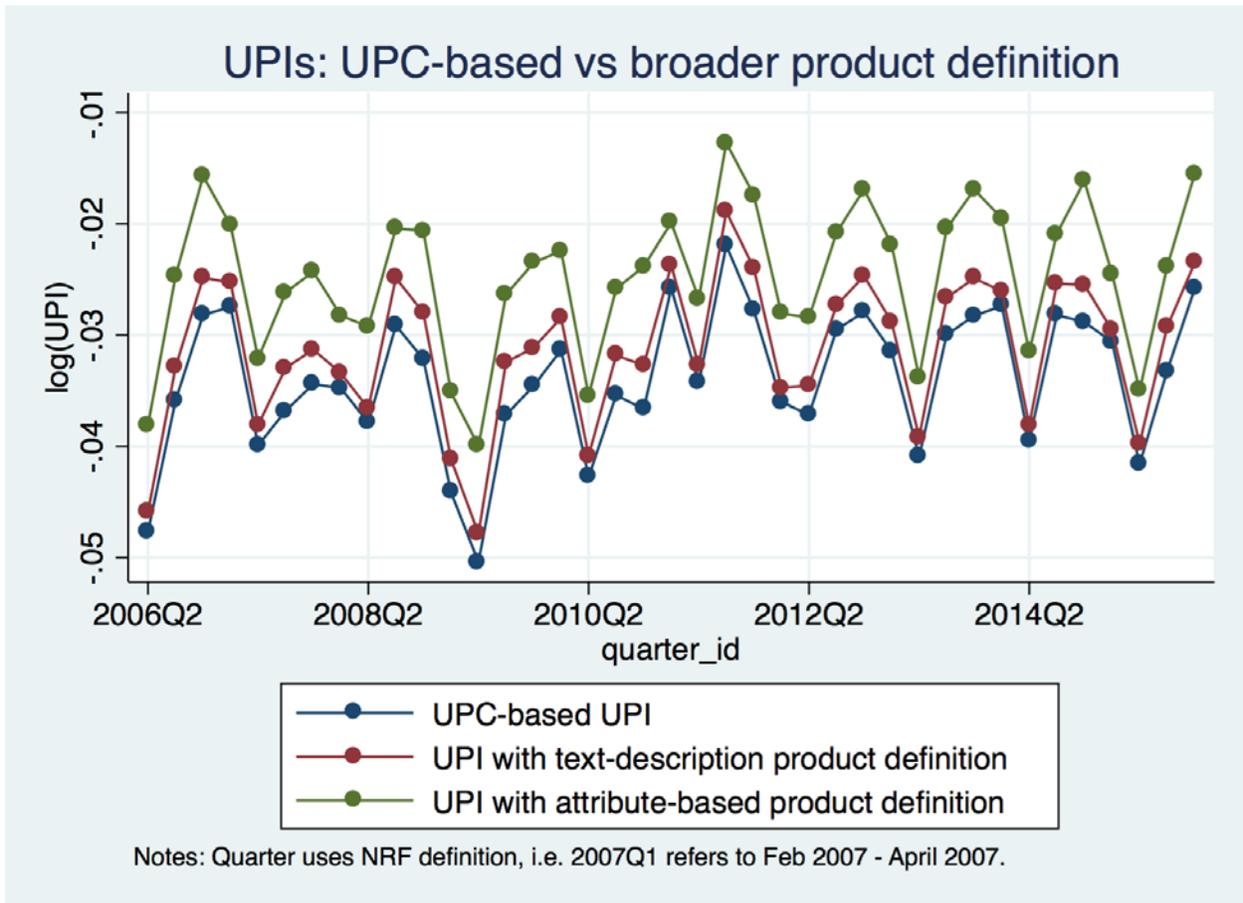
Appendix A. The Mechanics of the CES Unified Price Index

This appendix briefly lays out the mechanics of Redding and Weinstein's (2018) CES Unified Price Index, or UPI. Beginning with some additional notation, let:

- C_{kt} be the consumption of each good k at time t

- P_{kt} be the price of good k at time t

- Ω_t be the set of goods with positive sales at time t

- $\Omega_{t,t-1}$ be the set of goods with positive sales in times t-1 and t

- σ be the elasticity of substitution across goods

- $S_{lt} \equiv \frac{P_{lt}C_{lt}}{\sum_{k\in\Omega_t} P_{kt}C_{kt}}$ be the expenditure share for each good l and time period t

- $S_{lt}^* \equiv \frac{P_{lt}C_{lt}}{\sum_{k\in\Omega_{t,t-1}} P_{kt}C_{kt}}$ be the share of each common good in expenditure on all common goods

- $\lambda_{t,t-1} \equiv \frac{\sum_{k\in\Omega_{t,t-1}} P_{kt}C_{kt}}{\sum_{k\in\Omega_t} P_{kt}C_{kt}}$ be expenditure on all common goods as a share of total expenditure in periods t and t-1, evaluated using period t prices and quantities

- $\lambda_{t-1,t} \equiv \frac{\sum_{k\in\Omega_{t,t-1}} P_{kt-1}C_{kt-1}}{\sum_{k\in\Omega_{t-1}} P_{kt-1}C_{kt-1}}$ be expenditure on all common goods as a share of total expenditures in t and t-1, evaluated using period t-1 prices and quantities

- $\tilde{x}_t^* = \left(\prod_{k\in\Omega_{t,t-1}} x_{kt}\right)^{1/N_{t,t-1}}$ be the geometric average of x_{kt}

Then the ratio of price t to t-1 prices as defined by the UPI is:

$$\Phi_{t-1,t}^U = \left(\frac{\lambda_{t,t-1}}{\lambda_{t-1,t}}\right)^{\frac{1}{\sigma-1}} \left[\frac{\tilde{P}_t^*}{\tilde{P}_{t-1}^*}\left(\frac{\tilde{S}_t^*}{\tilde{S}_{t-1}^*}\right)^{\frac{1}{\sigma-1}}\right]$$

where $\left(\frac{\lambda_{t,t-1}}{\lambda_{t-1,t}}\right)^{\frac{1}{\sigma-1}}$ is the Variety or Feenstra Adjustment;

$\left(\frac{\tilde{S}_t^*}{\tilde{S}_{t-1}^*}\right)^{\frac{1}{\sigma-1}}$ is the Consumer Valuation Adjustment;

$\frac{\tilde{P}_t^*}{\tilde{P}_{t-1}^*}$ is the Relative Price Index or RPI;

and $\left[\frac{\tilde{P}_t^*}{\tilde{P}_{t-1}^*}\left(\frac{\tilde{S}_t^*}{\tilde{S}_{t-1}^*}\right)^{\frac{1}{\sigma-1}}\right]$ is the Common-Goods Price Index $\Phi_{t-1,t}^{CG}$.

The above shows the m Reported are kernel density estimates of the distributions of the demand quality/appeal residual (φ_{kt}) as described in Appendix A for these categories.mechanics of computing the UPI. It is worth noting that underlying this derivation is a CES consumer demand specification where the expenditure share of a good can be expressed in terms of quality adjusted prices (see equation (3) of Redding and Weinstein) as

$$S_{lt} \equiv \frac{P_{lt}C_{lt}}{\sum_{k \in \Omega_t} P_{kt}C_{kt}} = \frac{(\frac{P_{lt}}{\varphi_{lt}})^{1-\sigma}}{\sum_{k \in \Omega_t}(\frac{P_{kt}}{\varphi_{kt}})^{1-\sigma}} = \frac{(\frac{P_{lt}}{\varphi_{lt}})^{1-\sigma}}{P_t^{(1-\sigma)}}$$

where φ_{kt} is product quality/appeal for product k at time t. With an estimate of σ, φ_{kt} can be measured as a residual given measures of expenditure shares and prices. We explore the properties of the evolution of the distribution of φ_{kt} since this underlies the contribution of the PV and CV adjustment bias terms in the UPI.

Appendix B

Table B.1: Summary Statistics on Comparisons of 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 4 of text using the Nielsen Consumer Panel [CP] data instead of the Nielsen retail scanner data used in the main text. Data are quarterly from 2004-2016.