NBER WORKING PAPER SERIES

QUALITY ADJUSTMENT AT SCALE:
HEDONIC VS. EXACT DEMAND-BASED PRICE INDICES

Gabriel Ehrlich
John C. Haltiwanger
Ron S. Jarmin
David Johnson
Ed Olivares
Luke W. Pardue
Matthew D. Shapiro
Laura Zhao

Working Paper 31309
http://www.nber.org/papers/w31309

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2023, Revised October 2024

We acknowledge financial support of the Alfred P. Sloan Foundation and the additional support of the Michigan Institute for Data Science, the Michigan Institute for Teaching and Research in Economics and the U.S. Census Bureau. We thank David Byrne, Erwin Diewert, Robert Feenstra, Melvin Hochster, Robert Martin, Ariel Pakes, Stephen Redding, Marshall Reinsdorf, David Weinstein, and participants at multiple seminars and conferences for helpful comments. Ed Olivares, Luke Pardue, and Laura Zhao worked on this project as doctoral students at the University of Maryland (and Laura Zhao subsequently as a post-doc). Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. We also use the NPD data housed at the U.S. Census Bureau. All results using the NPD Group (Circana) data have been reviewed to ensure that no confidential information has been disclosed (CBDRB-FY19-122, CBDRB-FY21-074, and CBDRB-FY23-067, CBDRB FY24 0464). Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the view of the U.S. Census Bureau. The findings, conclusions, views, and opinions are solely those of the authors and do not necessarily represent the views of the U.S. Department of the Treasury or the United States government. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Quality Adjustment at Scale: Hedonic vs. Exact Demand-Based Price Indices
Gabriel Ehrlich, John C. Haltiwanger, Ron S. Jarmin, David Johnson, Ed Olivares, Luke W. Pardue, Matthew D. Shapiro, and Laura Zhao
NBER Working Paper No. 31309
June 2023, Revised October 2024
JEL No. C81, E31

ABSTRACT

This paper explores methods for constructing price indices from item-level transactions data on prices, quantities, and product attributes. The paper evaluates approaches that are feasible at scale, i.e., across the wide range of products, disparate encoding of attributes, and rapid product turnover inherent in "big data" on economic transactions, while producing improved cost-of-living indices that reflect both substitution effects and quality change. The paper presents hedonic methods that estimate changing valuations of both observable and unobservable characteristics in the presence of product turnover. It also considers demand-based methods that account for product turnover and changing appeal of continuing products. The paper provides evidence of substantial quality-adjustment in prices for a wide range of goods, including food and high-tech consumer products. The paper also shows that hedonics can be implemented with well-encoded attributes using standard econometrics and with unstructured attribute data using machine learning.

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

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

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

David Johnson
National Academies of Sciences,
Engineering and Medicine
500 Fifth Street, NW
Washington, DC 20001
djohnson@nas.edu

Ed Olivares
Department of Economics
University of Maryland
College Park, MD 20742
and U.S. Treasury
edwardo@umd.edu

Luke W. Pardue
Aspen Economic Strategy Group
2300 N St NW, Ste 700
Washington, DC 20037
and University of Maryland
luke@econstrategygroup.org

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

Laura Zhao
Bank of Canada
234 Wellington Street
Ottawa, ON K1A 0G9, Canada
and University of Maryland
YZhao@bank-banque-canada.ca

1 Introduction

Retail businesses create item-level data on the prices and quantities of the goods that they sell. These can form the basis for re-engineering key economic indicators by building internally consistent aggregates of value, volume, and price directly from item-level transactions. Aggregation of transactions data could supplant traditional surveys and enumerations conducted by statistical agencies. This paper implements and evaluates hedonic superlative price indices—price indices that lever the promise of item-level transactions data to deliver cost of living indices that account both for substitution effects that arise from relative price changes and for quality change embodied in the rapid turnover of goods.

Systemically accounting for substitution effects and quality change has long been a goal for the statistical system, as outlined for instance in the recommendations of the Boskin Commission summarized in Boskin et al. (1998) and recent National Academy reports such as Sichel and Mackie (2022). This paper shows—across multiple data sources and with multiple techniques for accommodating how attributes of products are encoded in source data—that hedonic superlative price indices are feasible to implement and have desirable properties.

We construct our hedonic price indices using both econometric and machine learning approaches. Our hedonic approach builds on the insights of Erickson and Pakes (2011, hereafter "EP"), who develop a novel method of calculating hedonic price indices that accounts for changing valuations of both observable and unobservable product characteristics. High-frequency, item-level transactions data with prices, quantities, and attributes permit the implementation of the EP hedonics approach with superlative price indices (such as the Fisher or Tornqvist) in real time using internally consistent expenditure weights. We discuss the motivation and assumptions underlying superlative price indices in the conceptual framework. Our approach is consistent with the recommendations of the Boskin commission, which called for the use of price indices that account for substitution in response to changes in relative prices and quality adjustment using hedonic methods. With the availability of scanner data that include prices, quantities and product attributes, we show that these recommendations are now feasible at scale.

This paper also evaluates exact demand-based price indices alongside the hedonic superlative indices. The demand-based approach discussed in this paper builds on the exact price indices developed from theoretical models of consumer demand: the Sato-Vartia price index (Sato, 1976; Vartia, 1976); the Feenstra (1994) adjustment to the Sato-Vartia index, which adjusts for quality change from product entry and exit (denoted the Feenstra index hereafter); and the Constant Elasticity of Substitution (CES) Unified Price Index (CUPI) developed in Redding and Weinstein (2020), which extends the Feenstra index for changing product appeal within narrow product groups. The demand-based approaches have the attractive feature that they yield exact price indices under specific assumptions. Moreover, these methods impose sufficient structure that they can be implemented without attribute data beyond a product taxonomy. Ehrlich et al. (2022) provides a preliminary examination of the exact demand-based price indices we consider here. We extend that work in this paper and highlight that, in contrast to hedonic superlative indices, the demand-based approach to account for changing product appeal leans heavily on assumptions about the economic environment beyond the functional form of the utility function.

A common feature of the frontier research methods using both hedonic and demand-based approaches is that they can account for changing consumer valuations of products or product characteristics. In principle, the CUPI captures both quality change due to product turnover and time-varying product appeal over the course of products' time in the marketplace without directly using detailed product attributes. The EP approach also reflects changing consumer valuations of various product attributes over time as they translate into the changing mapping between prices and characteristics.

Our goal is to provide a framework for re-engineering a wide range of economic statistics leading to substantial improvements in how they measure the economy. These aims include

- Having internally consistent measures of inflation and real expenditure;

- Increased timeliness of economic statistics, which means not only shorter lags for initial releases, but greatly mitigating the need for revisions that occur in the current infrastructure as source data slowly become available;
- Improved granularity of official statistics (frequency, product/industry detail, geographic detail); and
- More accurate measures of inflation.

The last point—more accurate measures of inflation—is the main aim advanced in this paper.

Our approach therefore seeks to establish a new "ground truth" for inflation. While we compare our measures to current official measures, the official statistics cannot be the main metric for validating re-engineered inflation measures. Instead, we articulate the following principles for re-engineered inflation measures. A re-engineered infrastructure for measuring inflation should

- Account for product entry and exit and accompanying changes in product quality;
- Approximate the true change in the cost of living under reasonable assumptions;
- Enable the computation of bounds for the cost of living that account for product turnover and quality change under general conditions;
- Not depend critically on small changes in assumptions or implementation; and
- Be feasible to implement at scale.

Specifically, statistical agencies need to be able to produce relevant components of the CPI and personal consumption expenditure price indices with at least the same timeliness as current statistical releases. Agencies need to maintain the reliability, consistency, and credibility of official statistics while producing much-improved measures of inflation. Inflation and nominal expenditures should be measured simultaneously and consistently from the same source data.

We evaluate multiple methods for measuring quality change at scale across different sources for a broad range of consumer goods. Our aim is to demonstrate that it is feasible to implement a new infrastructure for measuring price and quantity that would systematically take advantage of the real-time, item-level transaction data. That is, our goal is to establish that there are methodologies for turning the promise of "big data" for economic statistics into practice for official statistics. Given that there is no "ground truth" for evaluating all-new measures—in particular those that systematically account for quality change—we will evaluate the methods under the principles discussed above. To this end, the paper is organized as follows.

Section 2 introduces the two data sets we use in the paper. First, data from the NPD Group cover a wide range of general merchandise goods from both bricks and mortar and online retailers. We study disparate product groups: memory cards, headphones, coffee makers, boys' jeans, and work and occupational footwear. Second, data from NielsenIQ Marketing data provided by the Kilts Center at the University of Chicago Booth School of Business cover a wide range of food products from grocery stores, discount stores, pharmacies and liquor stores. In addition to covering different products and sectors, the NPD data and the NielsenIQ data that we use encode attributes very differently. The NPD data have a rich set of labeled attributes that makes them ready to analyze using the EP econometric approach. The NielsenIQ data have only sparse information on product attributes, which we address by adapting the EP methodology to a machine learning (ML) framework. Our findings show that the EP method can be made operational across distinct transaction data platforms with very different information about product attributes. In Section 2, we also present results illustrating the joint dynamics of market share and price over the life-cycle of products at the item level, which document interrelated changes in product quality, product demand, and price that must be captured by the price index methodology. Finally, Section 2 makes the case for focusing on food when using the NielsenIQ data.

Section 3 presents the conceptual framework for constructing price levels that meet the criteria outlined in the introduction when confronted by the massive product turnover and interrelated changes in price, quantity, and quality in item-level transactions data. We first review the matched-model price indices that are the basis for much of the current practice for measuring inflation and for deflating nominal expenditure and income. We then present in parallel two approaches for measuring quality change at scale: hedonics and demand-based approaches. This conceptual framework allows us to compare and evaluate these approaches, particularly as they are applied to item-level transactions data.

Section 4 presents the results, which show that there are substantial impacts from accounting for both substitution effects and quality change using hedonic superlative price indices. The desirable properties of these indices can be evaluated against the rubrics discussed in the introduction based both on bounding arguments for index numbers and results for superlative price indices. Our results show that the hedonic superlative has these desirable properties and is indeed feasible to implement at scale. We also present results for constant elasticity of substitution demand-based price indices, including the Feenstra (1994) index, which generalizes a CES price index to allow for entry and exit, and the Redding and Weinstein (2020) CUPI, which generalizes Feenstra to allow for changing product appeal. Feenstra shows quality adjustment of similar magnitude to hedonics, while the CUPI shows much greater quality adjustment. We explore the sources of the difference between the hedonic and CUPI approach in detail.

The final section offers our conclusions. Our results shows that it is both practical and advisable to move forward with re-engineering the statistical system to lever the huge amount of item-level transactions data generated in the twenty-first century economy. Doing so will produce measures of inflation and real activity that better capture changes in the cost of living and better capture quality change embodied in the rapid turnover of goods.

2 Data

This section provides an overview of the two data sets that we use to compute price indices. The first comes from the NPD Group and the second comes from NielsenIQ. For both data infrastructures, we aggregate the item-level transactions data to the quarterly, national level and focus on quarterly price indices.

2.1 NPD Data

We use proprietary data that the NPD Group provided to the U.S. Census Bureau.[1] They consist of monthly sales and quantity data at the product-store level from 2014 through 2018.[2] The NPD group tracks more than 65,000 retail stores, including widespread coverage of general merchandise, department and speciality stores (e.g., electronic and clothing stores) along with online retailers. The NPD data analyzed here consist of five broad product groups, within which we conduct our analysis separately: memory cards, coffee makers, headphones, boys' jeans, and work/occupational footwear (hereafter simply "occupational footwear").[3] The NPD data have unique item-level identifiers that are consistent cross-sectionally and over time. We aggregate the item-by-store-level observations to the national product-quarter level and calculate total quantity sold and average price for each product-quarter. The item-level data cover tens of thousands of product-quarter-level observations.

The NPD data contain detailed and well-coded information on the characteristics of each product. Beyond basic information such as product category and brand, these characteristics include details on different types of products within the broader categories (e.g., on-ear vs. in-ear headphones; coffee vs. espresso machines) and the features or attributes of different products (e.g. built-in grinders or auto-on/off settings for coffee makers). In some cases, the attributes include continuous variables, such as memory size for memory cards. We use the detailed product characteristics in the estimation of hedonic price indices and to group products into subcategories in our estimation of nested CES models.

Table 1 displays average item-level product turnover rates for each product group. Each of the groups exhibits a high degree of product turnover, ranging from 4.5 percent to 13.5

[1] The NPD Group is now part of Circana. We use the label NPD because the company was named NPD when we received the data.

[2] Month definitions follow the National Retail Federation (NRF) calendar (National Retail Federation, 2023). The NRF calendar is a guide for retailers that ensures sales comparability between years by dividing a year into months based on a 4 weeks-5 weeks-4 weeks format. The layout of the calendar lines up holidays and ensures the same number of Saturdays and Sundays in comparable months across years. The NRF calendar thus ensures the comparability of the aggregated sales over time.

[3] The NPD categories we investigate are narrow relative to the reported categories in the PCE, so we cannot make the same comparisons we do below for Food using the NielsenIQ data vs the PCE.

percent in terms of quarterly entry and exit rates. There is a distinction between overall entry and true entry – where the former represents a product not sold in the prior quarter but sold in the current quarter while the latter reflects the subset of overall entry where it is the quarter of the first product introduction. Similar remarks apply to overall exit and true exit.

Evidence on the life cycle dynamics of products is presented in Figure 1. The illustrated statistics are mean log differences from the product-specific initial values upon entry. Prices decline steadily after entry, while market shares exhibit a hump-shaped pattern post-entry. Memory Cards are characteristic: the share of a new memory card increases by about 200 percent over a year and then gradually declines—presumably because new memory cards, with superior attributes, have growing shares. The price peaks at entry and then falls gradually, but significantly. Because share and price are charted on the same scale, the 50-percent price decline in memory cards over their lifecycle may not be apparent. These share/price dynamics are systematically inter-related, so accounting for them is critical for accurate price index construction.[4]

Taken together, these findings highlight three important features of the data. First, there is considerable item-level product turnover that is a potentially important source of changing product quality. Second, there is a period of product rollout and decline over the life-cycle of a product. Third, there are inter-related movements in price and share over the product cycle.

2.2 NielsenIQ Data

We use the NielsenIQ Retail Scanner data (also referred to as RMS) from the Kilts Center for Marketing at the University of Chicago Booth School of Business for the period 2006 to 2015. The data consist of over 2.6 million products identified by the finest level of aggregation–12-

[4]Further evidence about the lifecycle dynamics of products is presented in appendix Figure D.1. We find that part of the hump-shaped pattern in market share is related to the relative number of stores where a product is available, which also exhibits a hump shape. We present related evidence for food products using NielsenIQ in Appendix B.6.

digit universal product codes (UPCs) that uniquely identify specific goods. The NielsenIQ Retail Scanner data are collected from over 40,000 individual stores from approximately 90 retail chains in over 370 metropolitan statistical areas (MSAs) in the United States. Total sales in NielsenIQ RMS are worth over $200 billion per year and represent 53% of all sales in grocery stores, 55% in drug stores, 32% in mass merchandisers, and 2% in convenience stores.

NielsenIQ organizes item-level goods into 10 departments, over 100 product groups, and over 1,000 product modules. A typical department is, for example, dry grocery, which consists of 41 product groups, such as baby food, coffee, and carbonated beverage. Within the carbonated beverage product group, there are product modules such as soft drinks and fountain beverage. We have classified the product groups into food and nonfood categories based on our own judgment in communication with researchers at the BLS. Tables D.1 and D.2 show our classifications into the food and nonfood categories. Appendix B.1 describes how we clean and prepare the data for analysis at the product group-quarter level.

To assess the merits of the NielsenIQ Scanner for our aim of improving inflation measurement for the aggregate economy, we compare its properties to official statistics. We conduct the comparison separately for food and nonfood categories because we find important differences across these two groups.

Food versus Nonfood

Figure 2 presents comparisons of nominal expenditures for the food and nonfood categories of the NielsenIQ data compared to the Bureau of Economic Analysis (BEA) personal consumption expenditure (PCE) data. In the top panel, for food, we find nominal sales indices for the NielsenIQ Scanner data tracks the PCE quite closely. We find the close tracking of food expenditures in the NielsenIQ Scanner data reassuring but not surprising. The NielsenIQ Scanner data has very good coverage of grocery stores.

For nonfood in the bottom panel, the scanner data exhibits less of an increase over time than the PCE. This finding for the scanner data, though little noticed, is not surprising. The scanner data tracking of stores that sell the nonfood items tracked by NielenIQ has weaker coverage. The non-food items sold, for example, at grocery stores, are highly idiosyncratic. While grocery stores sell non-food items such as cameras, small appliances, and housewares (among the listed nonfood product groups in NielsenIQ data), the type and extent of such goods sold at these outlets is far from representative of sales in the overall economy, and likely changing dramatically over time.

Figure 3 provides more evidence on these issues by comparing the growth of nominal expenditures for detailed categories in the NielsenIQ Scanner data from 2008:1 to 2015:4 to the growth in nominal sales for the PCE over the same period. Nominal expenditures grew at the same rate in the NielsenIQ Scanner data and the PCE in most of the detailed food product categories. In contrast, the growth rates for NielsenIQ's nonfood categories tend to be slower than and much more variable relative to the PCE.

In Appendix B.3, we also show that computing an arithmetic Laspeyres index from the NielsenIQ Scanner for food yields levels and fluctuations in food prices that track the CPI for food; so for food, the differences between the official price indices and our results arises from the methods for constructing the indices, not differences in the price quotations. For nonfood, we find substantially larger discrepancies between the CPI and the ScannerIQ arithmetic Laspeyres.

In summary, the nonfood data in the NielsenIQ Scanner do not track the official statistics aggregates for nonfood. In contrast, for food, the evidence supports the use of the NielsenIQ Scanner data for our analysis of price indices.[5]

[5] In Apppendix B.3 we also compare the NielsenIQ Consumer Panel data to official statistics. We find that NielsenIQ Consumer Panel data does not match total sales patterns for food as well as the Scanner data and does equally poorly for nonfood. Also, an arithmetic Laspeyres price index from the Consumer Panel is substantially different from the CPI for food and for nonfood also exhibits substantial differences with the CPI in a manner similar to the Scanner data.

3 Conceptual Framework

Time series price indices aim to measure approximately or exactly the change in the cost of living between two or more time periods. Building price indices from item-level transactions data addresses multiple challenges in tracking changes in of the cost of living.

1. By simultaneously and consistently measuring price and quantity at the item level, the use of transactions data makes it feasible to construct price indices that adjust for how consumers economize on the cost of living by substituting across goods in response to relative price changes.

2. With item-level data, price index construction needs, as a practical matter, to confront the rapid turnover of goods documented in the previous section, and use techniques to measure the quality change associated with the new and disappearing goods.

This section lays out the conceptual framework for using item-level transactions data for re-engineering price indices. First, it addresses continuing-goods price indexes which, with the simultaneous measurement of price and quantity, can account for consumer substitution. Second, it presents two approaches—hedonic and demand-based indices—that can account for quality change both from product turnover and from changing valuation of products or their

In this section, we review the basics of price indices—starting with indices currently used by statistical agencies and moving to indices that can be constructed with the item-level transactions data described above. In Section 3.1, we consider matched-model price indices, that is, indices which follow goods that are available in both the base and reference period. In Section 3.2.1, we consider approaches that account for entering and exiting goods.

3.1 Matched-Model Price Indices

Matched-model price indices are expenditure-share-weighted averages of price changes. Fundamental to re-engineering price indices using transactions data is the ability to align, in real-time and at the item-level, the price changes and expenditure shares. It is useful to define sets of goods. Let $\mathbb{C}_t$ denote the set of goods sold both in period t-1 and in period t, which we will call "continuing" or "common" goods. Further letting q_{kt} be the quantity of good k purchased in period t, s_{kt} and $s^{\star}_{kt}$ are defined as

$$s_{kt} \equiv \frac{p_{kt}q_{kt}}{\sum_{l\in\Omega_t} p_{lt}q_{lt}} \quad \text{and} \quad s^{\star}_{kt} \equiv \frac{p_{kt}q_{kt}}{\sum_{l\in\mathbb{C}_t} p_{lt}q_{lt}}. \tag{1}$$

Here, $s^{\star}_{kt}$ refers to the share of expenditure among goods that were common to the current and previous period.

Consider first the *arithmetic Laspeyres* price index,

$$\Phi_t^{L,Arith.} \equiv \sum_{k\in\mathbb{C}_t} \frac{q_{kt\text{-}1}p_{kt}}{q_{kt\text{-}1}p_{kt\text{-}1}} = \sum_{k\in\mathbb{C}_t} s^{*}_{kt\text{-}1}\frac{p_{kt}}{p_{kt\text{-}1}}, \tag{2}$$

that is, an arithmetic average of price relatives. The arithmetic Laspeyes index is of interest for two reasons. First, it is the formula used by the BLS and most other statistical agencies to construct the CPI. Components of the CPI are also used to deflate personal consumption expenditure as constructed by the BEA for the National Income and Product Accounts.[6] Second, there are important theoretical results discussed below that the arithmetic Laspeyes is an upper bound on the true cost of living index (COLI).

While we present the arithmetic Laspeyes index for comparison with the CPI and because it is an (the loosest) upper bound on the true COLI, our empirical work in this paper focuses on geometric price indices, which are weighted averages of log price changes. Geometric indices, while still generally bounds, mitigate substitution biases. The geometric Laspeyes index is defined as

[6]Equation 2 is a stylized approximation to the CPI. First, the CPI is technically a Lowe index rather than a Laspeyres index because its shares are lagged by more than one period. They are currently lagged by one year but were lagged by two years in the period that we study. Second, and not to be confused with the geometric indices we also discuss in this subsection, the elementary price observations at the item-area level in the CPI are unweighted geometric averages of price quotations across outlets. This use of geometric means to compute elementary prices was adopted by the BLS subsequent to the Boskin Commission to help address the "low-level" substitution problem.

$$\ln \Phi_t^{L,Geo.} \equiv \sum_{k \in \mathbb{C}_t} s^*_{kt\text{-}1} \ln \frac{p_{kt}}{p_{kt\text{-}1}}. \tag{3}$$

The geometric Paasche index, which uses current rather than base-period weights, is defined as

$$\ln \Phi_t^{P,Geo.} \equiv \sum_{k \in \mathbb{C}_t} s^*_{kt} \ln \frac{p_{kt}}{p_{kt\text{-}1}}. \tag{4}$$

As is familiar from textbook economics, the Laspeyres overstates changes in the cost of living because it holds spending weights fixed and the Paasche understates the change in the cost of living because it overstates the adjustment to relative price shocks.[7] We focus in particular on the Tornqvist index, given by

$$\ln \Phi_t^{TQ} \equiv \sum_{k \in \mathbb{C}_t} \left(\frac{s^*_{kt\text{-}1} + s^*_{kt}}{2} \right) \ln \frac{p_{kt}}{p_{kt\text{-}1}}, \tag{5}$$

which is a weighted average of the same log price changes as the Laspeyres and Paasche, but where the weights are a moving average of the base and current expenditure shares (a Divisia index).

The Tornqvist index has multiple attractive properties in addition to being a sensible average of the upper and lower bounds of the COLI. First, it is a "superlative" price index, meaning that it is a second-order approximation to the cost of living under a wide class of utility functions (Diewert, 1976).[8] Second, as a superlative price index, the Tornqvist is

[7] The formal bounding arguments are proven for the arithmetic indices. Following Erickson and Pakes (2011) and Diewert (2021) and general practice, we apply the same logic to the geometric indices. We present results for both the arithmetic and geometric Layspeyes.

[8] Every superlative price index is the change in the unit expenditure function (i.e., the exact price index) for flexible functional forms that are second-order approximation for a wide class of utility functions. The formal arguments are developed under homothetic preferences but research has shown that the Tornqvist is a reasonable approximation of the cost of living under nonhomothetic preferences (see, e.g., Boskin et al. (1998) and Jaravel (2024)). The Fisher and Sato-Vartia are also superlative indices, and, in our data, are very close to the Tornqvist. We generally discuss the Sato-Vartia in the context of the demand-based CES indices because it is exact for CES preferences under certain assumptions and because of our interest in contrasting it with other demand-based CES price indices. Reinsdorf and Dorfman (1999) and Barnett and Choi (2008) demonstrate that the Sato-Vartia index is superlative.

also approximately consistent in aggregation, meaning that it is not sensitive to changes in product categorization or nesting strategies (Diewert, 1978). The properties of superlative indices are derived under more restrictive assumptions than the bounding results for the Laspeyres. Hence, we regard the bounding results (e.g., Konüs, 1939; Pakes, 2003) and the second-order approximation results (Diewert, 1976) as complementary.

A key benefit of basing official price indices on item-level transaction data is the ability to implement superlative price indices at all levels of aggregation essentially in real time. Economists have long called for using superlative price indexes, for example, in the recommendations of the Boskin Commission (see Boskin et al., 1998). These recommendations have not been fully embraced by the statistical agencies largely because the contemporaneous data on sales or quantities needed to calculate expenditure share are not readily available. This practical limitation motivates the frequent use of the Laspeyres index in official statistics (e.g. for the CPI), which is subject to potentially large substitution bias relative to the superlative indices.

Where agencies have used superlative indices or approximations to them, they typically do so at a relatively high level of aggregation where expenditure data from disparate sources are combined with elementary price indices collected by the BLS. Example of this practice include the BLS's chained CPI (C-CPI-U) and the BEA's use of chaining throughout the national income and product accounts. The FOMC's use of the BEA's PCE price index as its preferred measure of inflation is in part motivated by the index's effort to reflect substitution effects. A major benefit of the use of transactions data in price indices is to allow implementation at scale of the superlative indices long understood to have desirable properties.

3.2 Accounting for Product Turnover and Quality Change

An important limitation of traditional price indices—whether using transactions data or prices collected by statistical agencies—is that they are "matched-model" indices: they calculate price changes across the goods that were sold both in the base and in the current periods. Hence, the indices discussed in Section 3.1 do not account directly for goods that enter or exit across periods. We show product turnover is an important source of changing product quality and is ubiquitous in item-level data. Moreover, even absent quality change, transactions data requires automatic handling of product turnover because of the high number of products that experience frequent turnover.[9] Obviously, this case-by-case approach is not feasible for the millions of goods confronted in transactions data. The main contribution of this paper is to consider two price-index methods for dealing with this product turnover at scale: (1) hedonics and (2) demand-based.

3.2.1 Hedonic Price Indices

Hedonic imputation allows a price index to account for product turnover by using product characteristics and an estimated hedonic relationship between characteristics and prices to impute the "missing" prices for entering and exiting products. Pakes (2003) shows that under general conditions, the hedonic Laspeyres index provides an upper bound to the true change in the consumer's cost of living following the standard Konüs (1939) bounds arguments. Moreover, he shows that this hedonic Laspeyres generally provides a tighter upper bound relative to the matched-model Laspeyres because it accounts for the selection bias from omitting exiting goods in the matched-model. The Könus bounds approach does not rely on a representative consumer; indeed it applies even in the presence of nonhomothetic preferences (e.g., Pakes, 2005; Diewert, 2021).

In this section, we will first discuss how the missing prices of entering and exiting goods

[9]Statistical agencies also need to handle product turnover in the current architecture. BLS enumerators aim to follow the items for up to four years. When a product inevitably disappears, they make a judgement—sometime in consultation with an analyst in Washington—about whether a similar item can be substituted because its characteristics have *de minimus* changes. The analyst uses three methods – direct comparison, direct quality adjustment, or imputation. If the products are comparable, any price change with an item replacement is inflation. If the goods are non-comparable, i.e., they substitute a product with substantial changes, they can use imputation to estimate the price change or "link out" the price change. For direct quality adjustment, they can use hedonic adjustments, which are made for about 7.5 percent of goods (mainly apparel and tech goods). In a limited number of cases, BLS makes adjustments based on costs of changing components (e.g., when an option on an automobile becomes standard equipment).

are imputed in the price index formulas analogous to the matched-model indices presented in the previous section. We will then discuss approaches to carrying out the imputation in item-level transactions data to produce hedonic price indices at scale.

Defining the Hedonic Tornqvist

Our hedonic Tornqvist index is

$$\ln \Phi_t^{TQH} \equiv \sum_{k \in \{\Omega_{t\text{-}1} \cup \Omega_t\}} \left(\frac{s_{kt\text{-}1} + s_{kt}}{2} \right) \widehat{\ln \frac{p_{kt}}{p_{kt\text{-}1}}}. \tag{6}$$

where $\widehat{\ln \frac{p_{kt}}{p_{kt\text{-}1}}}$ is the imputed value of the log change in prices. Hedonic Laspeyres and Paasche indices are calculated similarly. Note several features of equation (6), all of which are consistent with Erickson and Pakes (2011). First, the hedonic index is based on the imputed price change (not the ratio of imputed price levels). Second, the hedonic index is based on imputations of all observations (including the continuing goods for which actual observations are available). This procedure, called "full imputation," defines all goods in terms of their attributes, not just the ones whose prices must be imputed. Third, the Divisia index of the shares is the average of the shares of goods at time t and t-1, both of which are observable notwithstanding entry and exit. That is, the expenditure shares used reflect observed data (no imputation is required).

Given our use of scanner data that permits measuring expenditure shares on a timely basis, we focus on the hedonic Tornqvist index. As noted above, the traditional Tornqvist index is a superlative index that addresses the substitution bias present in the Laspeyres index. In turn, the hedonic Tornqvist permits us to account for both product entry and exit, avoiding the selection bias inherent in the traditional matched-model Tornqvist index. We interpret the hedonic Tornqvist index as an approximation of the cost-of-living index defined over product characteristics inclusive of product entry and exit.[10] We also report

[10] Hill and Melser (2008) provide further support for the use of the hedonic Tornqvist index. They show that a price index constructed from log-linear hedonic estimates of the type on which we focus is numerically equivalent to a Fisher characteristics price index, i.e., a Fisher index constructed using the quantities of characteristics embedded in products and those characteristics' hedonic prices.

results for the traditional matched-model Laspeyres index without product exit and a hedonic Laspeyres with product exit, following Pakes (2003) and Erickson and Pakes (2011), as a useful point of reference given the bounding properties under very general conditions.

Approaches to Hedonic Estimation

The log-level hedonic price model common in the literature takes the form

$$\ln p_{kt} = h_t(Z_k) + \eta_{kt}, \tag{7}$$

where Z_k is a vector of observable characteristics for good k. The function $h_t()$ is often linear in parameters, and the hedonic equation is estimated with with ordinary or weighted least squares regression. An important feature of equation (7) is that the hedonic function varies over time, i.e., the function $h_t()$ is estimated separately period-by-period. Underlying the hedonic approach is the assumption that value can be specified as a function of the goods' characteristics. The time-varying estimation allows the hedonic function to capture changing consumer valuations, markups, or other changing aspects of market structure (Pakes, 2003). Pakes (2003) emphasizes that the estimated coefficients are not readily interpretable as marginal valuations of characteristics.

A core limitation of the log-level hedonic estimation approach outlined in equation (7) is that there are likely to be product characteristics that are relevant to the formation of prices but that the econometrician cannot observe. Erickson and Pakes (2011) introduce hedonic methods that can account for such unobserved characteristics. An important first step is to estimate hedonic models of price changes rather than price levels

$$\Delta \ln p_{kt} = Z_k' \beta_t + v_{kt}. \tag{8}$$

This log-difference hedonic model estimates the change in hedonic price coefficients directly, which "differences out" any unobservable item-level characteristics that have a fixed influence on prices over time. This basic log-difference hedonic model does not, however, account for the influence of time-varying unobservable characteristics. We call this approach the "EP-F" approach for short to indicate that it accounts for only fixed unobservables.

Erickson and Pakes (2011) propose a modified approach that can account for the time-varying influence of unobservable characteristics. We call this approach the "EP-TV" approach for short. Implementing the EP-TV approach requires two steps. First, estimate the log-level hedonic specification in equation (7) for period t-1. Second, estimate a log-difference hedonic model including the lagged residuals from the first stage, which we denote $\hat{\eta}_{kt\text{-}1}$.[11] The second estimating equation is then

$$\Delta \ln p_{kt} = Z_k' \beta_t + \kappa_t \hat{\eta}_{kt\text{-}1} + v_{kt}. \tag{9}$$

Including the initial residuals $\hat{\eta}_{kt\text{-}1}$ from equation (7) in equation (9) allows the model to capture the influence of time-varying valuations of unobservable product characteristics to the extent that the initial residuals are correlated with price changes. In our analysis, we consider log-level, EP-F, and EP-TV approaches.

Our hedonic methods—both the econometric approach with NPD data implemented in this paper and the ML approach with NielsenIQ data—use quantity-share weights *for the estimation*. The indices are, of course, constructed using expenditure weights. In Appendix A, we also consider EP-TV hedonics estimation using expenditure weights. Our main conclusions are broadly similar using both sets of estimation weights (see Table D.3). Bajari et al. (2021) also use quantity-share weights in their implementation of ML methods for hedonic price indices using item-level transactions data. Broda and Weinstein (2010) and Redding and Weinstein (2020) advocate for quantity weighting in estimation using scanner data based on the argument that unit values calculated based on a large number of purchases

[11] This characterization is equivalent to the time varying unobservables specification in Erickson and Pakes (2011). In that paper, they describe a closely related multi-step procedure. First, estimate the log levels hedonics and recover the residual. Second, estimate the log price relative on characteristics. Third, estimate the change in the residuals from the the log levels on the characteristics. Using the sum of the predictions from the latter two steps, as described in Erickson and Pakes (2011), is equivalent to using the predictions from equation (9).

are better measured than those based on a small number of purchases.

The initial lagged residual for an entering product cannot be recovered from equation (7) because the price is not observable prior to entry. Erickson and Pakes (2011) do not face this issue because they consider only hedonic Laspeyres indices, which account for exiting goods but not entering goods. In our main analysis using the EP-TV approach, we assume the sales weighted initial lagged residual for entering goods in the period prior to entry has a mean of zero. We consider alternative approaches in our analysis of NPD data. First, we impute the unmeasured initial residual for an entering product to be equal to the observed current period residual for the entering good (using the first step log level model for the current period to measure the current period residual). Second, we construct a backcasted initial residual for entering goods taking advantage of the high persistence in the residuals from the log level first step for continuing goods (details explained below). These two alternatives produce very similar results to the baseline specification setting the sales-weighted mean of the initial residual for entering goods to zero.

As already noted, we use full-imputation price indices, which can be interpreted as characteristic price indices (Hill and Melser, 2008; De Haan, 2008).[12] Erickson and Pakes (2011), who also use full-imputation indices, argue that methods that use observed rather than predicted price relatives for continuing goods, are subject to a form of selection bias because they treat the hedonic estimation error for continuing, entering, and exiting goods in an asymmetric manner. Benkard and Bajari (2005), Diewert et al. (2008), and Bajari et al. (2021) also use full-imputation indices. Finally, as we show below, full-imputation hedonic indices are less subject to chain drift than traditional matched-model price indices.

Hedonic price indices use the mapping between prices or price relatives and product characteristics among continuing goods to impute the "missing" prices or price relatives for entering and exiting goods. Characteristics turnover, i.e., the appearance or disappearance of particular product attributes in the data, is distinct from product turnover. We find that

[12] Using full-imputation indices also facilitates comparison with the commonly used time dummy method discussed below, as highlighted by De Haan (2008) and Diewert et al. (2008).

characteristics entry and exit rates are much smaller than the product entry and exit rates reported in Table 1. For boys' jeans, occupational footwear, and memory cards, we observed essentially zero characteristics entry over our sample period, although we do observe brand entry and exit (at rates less than 1 percent) for occupational footwear and boys' jeans. For headphones and coffee makers, we observe very low characteristics entry and exit rates (less than 0.1 percent), with slightly higher sales-weighted rates (as high as 0.5 percent for coffee makers). This evidence is consistent with the view that new characteristics diffuse slowly through the entry of new goods, and characteristics disappear from the available bundle slowly through product exit. Relatedly, new goods often have *more* of an important characteristic (e.g., size and speed of memory cards), while exiting goods often have *less* of those characteristics, so that product turnover involves upgrading of existing characteristics rather than the entry and exit of characteristics themselves.

We also consider the related, but distinct, *time dummy* method that has been actively used in the research literature and by the BLS (see, e.g., Triplett, 2004). We follow the recent literature (e.g., Byrne et al., 2019) using adjacent-period, weighted least squares estimation with Tornqvist market-share weights. Specifically, we pool observations from the adjacent periods t-1 and t and estimate hedonic regressions of the form

$$\ln p_{k\tau} = \alpha_{t\text{-}1,t} + \delta_t + Z_k'\gamma_{t\text{-}1,t} + \epsilon_{k\tau}, \quad \tau = \{t\text{-}1, t\}, \tag{10}$$

where $\alpha_{t\text{-}1,t}$ is a constant, Z_k is the vector of characteristics for good k, $\gamma_{t\text{-}1,t}$ is a vector of estimated hedonic coefficients held fixed across periods t-1 and t, and δ_t is a fixed effect for period t.[13] Exponents of the resulting coefficients δ_t can be interpreted as the quality-adjusted change in the price level between periods t-1 and t. Some limitations of the time

[13] Letting T denote the total number of periods in the data, we estimate $T-1$ separate pooled two-period regressions. We specify the hedonic regression equation (10) using the same vector of characteristics Z_k in each pair of adjacent periods. Occasionally, new features are introduced to the data. In pairs of adjacent periods entirely prior to the introduction of a new characteristic, it will be omitted from the regression because of collinearity with the intercept term. In pairs of adjacent periods in which the new feature is absent during period t-1 and present during period t, the feature will be included in the estimated regression. Symmetric arguments apply for characteristics that exit.

dummy method are that it does not account for unobservable product characteristics and that it imposes constant coefficients on characteristics in adjacent periods. Appendix A provides additional discussion.

Our implementations of the EP-TV approach and the time dummy method with the NPD data use standard econometric methods to estimate the hedonic function $h_t()$. This approach is feasible with the NPD data because of the enormous value-added the NPD group provides in terms of item-level attributes.

Machine Learning Hedonics

To implement the hedonic EP-TV approach with the NielsenIQ Scanner data we need to overcome the challenge that unlike the NPD data the NielsenIQ data do not contain pre-coded product attribute data for most products aside from short textual product descriptions.[14] To illustrate this point, consider two product descriptions for soft drinks: `ZR DT LN/LM CF NBP CT` and `NATURAL R CL NB 12P`. A product description for toilet paper is: `DR W 1P 308S TT 6PK`. A human analyst could decipher portions of these descriptions: `DT` means "diet," `12P` means "twelve pack," `1P` means "one ply," `308S` means "308 sheets," etc. It would not be feasible for human analysts to encode such data at scale, and simple dictionaries would be fooled (e.g., the `P`-suffix means "pack" for soft drinks and "ply" for toilet paper).

To address this challenge, we implement deep neural networks to predict product prices and price changes from the product descriptions in the NielsenIQ Kilts Center data. The machine learning (ML) approach parallels the EP-TV approach we use for NPD, in that it first predicts price levels and then, to capture time-varying unobservable effects, uses the prediction error from the first-stage in a second-stage neural net predicting price changes. In many ways, the EP-TV method can incorporate ML methods quite naturally. The key innovation is to use ML methods rather than standard regression techniques we used with

[14] Cafarella, Ehrlich, Gao, Haltiwanger, Shapiro, and Zhao (2023) provide technical details and evaluation of this methodology. This working paper is provided as supplemental on-line material. Appendix B.5 also provides more detail.

NPD data to estimate the hedonic functions for log price levels and changes in equations (7) and (9).[15]

3.2.2 Demand-Based Price Indices

In this section, we describe our use of exact cost-of-living indices for Constant Elasticity of Substitution (CES) demand systems. They provide tractable, implementable price indices that can account for quality change and product turnover. Redding and Weinstein (2020) generalize the unit expenditure function for a representative consumer with CES preferences as

$$P_t = \left[\sum_{k \in \Omega_t} \left(\frac{p_{kt}}{\varphi_{kt}} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}, \tag{11}$$

where $\sigma > 1$ is the consumer's elasticity of substitution between products, and φ_{kt} is an appeal parameter for product k. Both the set of products sold Ω_t and product-level appeal φ_{kt} may vary over time. Equation (11) is not directly empirically implementable, because the appeal parameters φ_{kt} are unobservable. The standard Sato-Vartia and Feenstra indices (Sato, 1976; Vartia, 1976; Feenstra, 1994) are also based on equation (11), but with the φ_{kt} restricted to have constant values over time (i.e., $\varphi_{kt} = \varphi_k$).

The Sato-Vartia index is exact for CES preferences in the absence of product turnover or time-varying product appeal. The log Sato-Vartia index is defined as

$$\ln \Phi^{SV}_{t\text{-}1,t} \equiv \sum_{k \in \mathbb{C}_t} \omega_{kt} \ln \left(\frac{p_{kt}}{p_{kt\text{-}1}} \right), \quad \omega_{kt} \equiv \frac{s^*_{kt} - s^*_{kt\text{-}1}}{\ln(s^*_{kt}) - \ln(s^*_{kt\text{-}1})} \Bigg/ \left(\sum_{k \in \mathbb{C}_t} \frac{s^*_{kt} - s^*_{kt\text{-}1}}{\ln(s^*_{kt}) - \ln(s^*_{kt\text{-}1})} \right). \tag{12}$$

The Feenstra (1994) index generalizes the Sato-Vartia index to account for turnover in

[15] In related work, Bajari et al. (2021) use an advanced machine learning approach that includes encoding image data as inputs into price predictions. Bajari et al. (2021) provide novel methodology for encoding images via machine learning. However, they estimate hedonic models of price levels period by period – not using the EP-TV method that allows for unobserved characteristics.

the set of goods sold Ω_t. We define the terms $\lambda_{t,t-1}$ and $\lambda_{t-1,t}$ as

$$\lambda_{t,t-1} \equiv \frac{\sum_{k\in\mathbb{C}_t} p_{kt}q_{kt}}{\sum_{k\in\Omega_t} p_{kt}q_{kt}}, \qquad \lambda_{t-1,t} \equiv \frac{\sum_{k\in\mathbb{C}_t} p_{kt-1}q_{kt-1}}{\sum_{k\in\Omega_{t-1}} p_{kt-1}q_{kt-1}}. \tag{13}$$

The log Feenstra index is then defined as

$$\ln \Phi^{Feenstra}_{t\text{-}1,t} \equiv \frac{1}{\sigma-1}\ln\left(\frac{\lambda_{t,t-1}}{\lambda_{t\text{-}1,t}}\right) + \ln \Phi^{SV}_{t\text{-}1,t}. \tag{14}$$

Letting $ER_{t\text{-}1,t}$ and $XR_{t\text{-}1,t}$ represent the sales-weighted product entry and exit rates, the log Feenstra adjustment term can be approximated as $\ln\left(\frac{\lambda_{t,t\text{-}1}}{\lambda_{t\text{-}1,t}}\right)^{\frac{1}{\sigma-1}} \approx \frac{1}{\sigma-1}\left(XR_{t\text{-}1,t} - ER_{t\text{-}1,t}\right)$. The Feenstra adjustment factor for product turnover (or "Lambda Ratio") thus indicates a downward adjustment to the Sato-Vartia index when the sales share of entering products is higher than the sales share of exiting products; it collapses to one in levels, or zero in the log version in equation (14), in the absence of product turnover with quality change. We use the actual Feenstra adjustment and not the approximation in our implementation.

The CUPI generalizes the Feenstra index to allow for time-varying product-level appeal. Redding and Weinstein (2020) emphasize that including time-varying product appeal is essential for the CES demand system to be consistent with the observed micro variation in prices and quantities because quantities purchased often change even when relative prices do not. They specify a normalization on the changes in the appeal shocks so that there is no change in geometric average tastes at the product group level for common goods. This assumption, combined with their assumption, which we also maintain, that consumers have Cobb-Douglas preferences across product groups, guarantees that product-level appeal shocks do not spill across product groups.

The CUPI in logs is given by

$$\ln \Phi^{CUPI}_{t\text{-}1,t} \equiv \frac{1}{\sigma-1}\ln\left(\frac{\lambda_{t,t-1}}{\lambda_{t\text{-}1,t}}\right) + \ln \Phi^{SV}_{t\text{-}1,t} - \sum_{k\in\mathbb{C}_t}\omega_{kt}\ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right). \tag{15}$$

where φ_{kt} is the time-varying product appeal. The last term is the taste shock bias. The sign of the taste shock bias term depends on the sign of correlation between product appeal shocks and the Sato-Vartia expenditure weights. The Feenstra and Sato-Vartia indices are upward (downward) biased relative to the CUPI (the exact price index under the assumptions) when there is a positive (negative) correlation.

Redding and Weinstein (2020) derive an empirically implementable version of the CUPI (in logs) as

$$\ln \Phi_{t-1,t}^{CUPI} = \frac{1}{\sigma - 1} \ln \left(\frac{\lambda_{t,t-1}}{\lambda_{t-1,t}} \right) + \frac{1}{N_{\mathbb{C}_t}} \sum_{k \in \mathbb{C}_t} \ln \left(\frac{p_{kt}^*}{p_{kt-1}^*} \right) + \frac{1}{\sigma - 1} \frac{1}{N_{\mathbb{C}_t}} \sum_{k \in \mathbb{C}_t} \ln \left(\frac{s_{kt}^*}{s_{kt-1}^*} \right). \quad (16)$$

where the variables defined over the set of common goods are denoted by an asterisk. Equation (16) shows that two of the CUPI's three terms are unweighted geometric means. This property is important for the CUPI's empirical performance. We call the CUPI's second term the "P^* ratio" and its third term the "S^* ratio." The second term is the traditional Jevons index. The third term depends on the average of the log expenditure shares in the two time periods.

These CES price indices exactly recover the change in the consumer's cost of living under different assumptions. The Sato-Vartia price index is exact if there is no product turnover and no time variation in product appeal.[16] The Feenstra-adjusted Sato-Vartia index is exact in the presence of product turnover but the absence of time-varying product appeal. The CUPI is exact under the more general conditions of product turnover and time variation in product appeal. We find that these generalizations of the Sato-Vartia index are empirically relevant.

Although the CUPI nests the Sato-Vartia and Feenstra, it leans more heavily on the CES functional form in following sense. The last two terms in equation (16) are unweighted indices. The Sato-Vartia and the Feenstra, in contrast, include only expenditure-weighted

[16] Feenstra and Reinsdorf (2007) show the Sato-Vartia index is unbiased in expectation with randomness in tastes under restricted conditions. Appendix C.3 contains a further related discussion of this topic.

indices. Because the CUPI's unweighted terms are sensitive to products with very small expenditure shares, the CUPI can feature large measured price changes from what appear to be economically minor products. The CUPI requires very large taste shocks to rationalize the low demand for products with small shares.[17]

Redding and Weinstein (2020) adjust their empirical implementation of the CUPI by applying what we call a "common goods rule," which defines the set of goods over which the P^* and S^* ratio terms are calculated (i.e., the goods included in the set $\mathbb{C}_t$). The goods excluded from the set of common goods are reallocated to the product turnover component (Feenstra adjustment factor), which is expenditure weighted. A common goods rule of this sort can be motivated by the argument that it takes time for goods to enter and exit the market. Consistent with this argument, Redding and Weinstein (2020) restrict the set of common goods in their empirical CUPI to those that are sold for a sufficiently long duration both prior to period t-1 and subsequent to period t.[18] A limitation of this duration-based common goods rule is that it requires forward-looking information to implement, and thus it is not feasible to implement in real time. We find that we can mimic Redding and Weinstein's results using a purely backward-looking rule that can be implemented in real time. As we will see, the empirical effect of the S^* ratio varies significantly depending on the implementation of the common goods rule. We discuss potential reasons for the common goods rule's importance in Appendix C.2.

4 Results

In this section, we present and discuss the traditional matched-model, hedonic, and demand-based exact price indices we calculate in the item-level data. We focus first on our results

[17]We thank Rob Feenstra for first bringing this point to our attention in his discussion of Ehrlich et al. (2022).

[18]Redding and Weinstein (2020) measure annual CUPI inflation from the fourth quarter of one year to the fourth quarter of the next year. Defining those quarters as periods t-1 and t, they define common goods as those sold in both of those quarters as well as in the 3 quarters prior to t-1 and the 3 quarters subsequent to t. In addition, they require the good be sold for at least 6 years total (although not necessarily consecutively).

from the NPD data, because the richness of the data permits more exploration of alternative methods.

4.1 NPD Results

4.1.1 Hedonics

We consider a wide variety of hedonic specifications as discussed in Section 3.2.1. We find that predicting log price changes directly produces significantly better model fit than estimating log price levels in periods t-1 and t separately and then forming a predicted log price change. Using the EP-TV approach, which accounts for the time-varying influence of unobservable characteristics by including the lagged residual from a log-level regression, further increases the model fit across all product groups. The results therefore support the argument in Erickson and Pakes (2011) that estimating price changes helps to account for unobservable product characteristics, and that including the first-stage residual from predicting price levels in the estimation of log price changes provides a further advantage.

In terms of specifics, the EP-TV approach yields average R^2 values that range from 0.13 to 0.50 across product groups. This reflects an improvement on the results using the EP-F approach – with values that range between 0.09 to 0.47. Both of these approaches dominate using the log level hedonic specification for implied price changes – where the implied R^2 ranges from 0.05 to 0.24. The log level specifications exhibit much higher R^2 (in the range of 0.62 to 0.72 across product groups) in accounting for the pooled cross sectional variation in prices. Predicting price changes is inherently a much more difficult task than predicting price levels, because the latter reflect cross-sectional differences in product characteristics, while the former reflect changes in the mapping between prices and characteristics over time.[19]

In the main text, we focus on the results using the EP-TV estimation approach. In

[19]The discussion in this paragraph summarizes results in Table D.4. This table also shows results from expenditure-weighted estimation in the columns labeled "EW" in the row labeled "Estimation Weights." Quantity-weighted estimation yields higher R^2 for log price changes on average than expenditure-weighted estimation.

Appendix A, we also present results for level models, time-dummy models and EP-F models. We find larger measured quality adjustment using the EP-TV approach compared to these other methods. Figure 4 shows a comparison of annual rates of inflation for the matched-model Tornqvist and the EP-TV Tornqvist. The gap between the matched-model Tornqvist and the EP-TV approach indices varies considerably across product groups, with the largest average differences for memory cards (-2.9 percentage points annually) and headphones (-2.5), and smaller differences for coffee makers (-0.70), boys' jeans (-1.30), and occupational footwear (-0.42). Figure 4 also provides further evidence on the efficacy of the EP-TV approach by displaying results with key observable characteristics left out of the hedonic estimation. Specifically, for memory cards the memory size is omitted, and for the other product groups, the large brand dummy variables are omitted. Omitting these informative characteristics from the estimation equation has a minimal effect on the resulting price indices. Appendix Figure D.3 presents additional analyses showing that omitting those characteristics has a much larger effect on the hedonic indices using a log-level estimation approach.

Table 2 reports the price index in 2018:4 (2014:4=1) for a wider range of matched model and hedonic indices (all using the EP-TV estimation method). In this table we include the arithmetic and geometric Laspeyres for both matched-model and hedonics along with Tornqvist matched-model and hedonic. We confirm the expected ordering for matched-model indices: Arithmetic Laspeyres > Geometric Laspeyres > Tornqvist. In all cases we find that the corresponding hedonic index for each of these indices is substantially lower while maintaining the same order. The hedonic Laspeyres includes the contribution of exiting products and the hedonic Tornqvist the contribution of entering and exiting products. If anything, the gaps between matched model and hedonics are larger in magnitude for Laspeyres than the Tornqvist. We obtain similar patterns for the Laspeyres to those in Erickson and Pakes (2011). That is, using the EP-TV method yields significantly lower inflation for the Laspeyres (either arithmetic or geometric) than the matched model index. Moreover, because the hedonic Laspeyres does not account for entering goods, the EP-TV Laspeyres results do not depend on our methodology for measuring the lagged residual for entrants.

We examine two alternative methods for treating the *unmeasured* lagged initial residual for entering goods in the hedonic Tornqvist. In Table 3, we compare these two alternative methods to the baseline approach for handling the unobserved initial lagged residual. The first row in each panel sets the sales weighted mean of the missing lagged residual to zero (the baseline specification). The second row, labeled *current*, uses the current period residual of entering goods as a proxy for the missing lagged residual. The third row, labeled *backcast*, uses the tight relationship between current period and lagged residuals to backcast the lagged residual (using an auxiliary regression model estimated each quarter for continuing goods as reported in appendix Table D.5). Results across the baseline and two alternative specifications are very similar. This reassuring robustness stems from several factors. First, the predicted price relative for entrants depends on the contribution of observable characteristics for entrants using estimated coefficients from the second stage of the EP-TV methodology. This desirable property holds regardless of the treatment of the lagged residual for entrants. Second, the sales share of entry is relatively modest at a quarterly frequency. Over longer horizons (e.g., direct year-over-year, or YOY) the sales share of entrants over a four-quarter horizon is substantially larger.[20] The second panel of Table 3 shows more sensitivity to alternative methods over this longer horizon, but the patterns remain quite similar.[21]

We also examine whether the hedonic estimates are sensitive to differences in the range of values of characteristics of entering or exiting goods relative to continuing goods.[22] We conduct this robustness check for memory cards because their two most important characteristics—

[20]See appendix Table D.6 for sales shares at quarterly and annual frequencies.

[21]The YOY index estimates the annual rate of inflation by defining common goods as those present in quarters t and $t-4$, entering products as those present in t but not $t-4$, and exiting products as those present in $t-4$ but not in t. The hedonic EP-TV second stage regression specification is re-estimated over this horizon for common goods and the predicted price relatives for entering and exiting goods use these estimates over this horizon.

[22]This robustness check follows the spirit of a similar exercise in Pakes (2003). Appearance or disappearance of entire characteristics is exceedingly rare, so we focus on changes in the support of the value of characteristics.

size and speed—have well-defined ranges. Specifically, we construct an alternative EP-TV-based hedonic price index excluding entering products in a period if the size or speed of the entering products is outside the range of continuing products. The resulting average annual cumulative chained EP-TV Tornqvist price index is -20.09% – which is very similar to the baseline estimate of -20.12%. This robustness holds even though about 50% of entering products are excluded, and those products account for about 25% of the sales of entering products.

Our main results are consistent with the findings in Erickson and Pakes (2011). They present results for televisions in which standard log-level hedonic estimation suggests higher rates of inflation than traditional matched models. They show, however, that using their methodologies to account for unobservable product characteristics (both using fixed valuation of unobservables and time-varying unobservables) produces systematically lower estimated inflation than the traditional matched models. Our analysis is a significant extension of their findings to superlative price indexes, which, unlike with the data they had available, we can compute because the scanner data contain both prices and quantities.

4.1.2 CES Demand-Based Price Indices

We turn now to CES demand-based price indices. The Feenstra index and the CUPI require estimates of the elasticities of substitution in their empirical implementation. Our baseline approach is to estimate a single elasticity for each of the NPD product groups. We employ the method used by Feenstra (1994) and Redding and Weinstein (2020) for this purpose.[23] Appendix Table D.7 reports the estimated elasticities from about 5.2 to 7.8, which is consistent with the literature.

Figure 5 plots the Sato-Vartia, Feenstra, and CES unified (CUPI) price indices, as well as the components of the latter two indices. The baseline CUPI is calculated without a

[23]This method double-differences the demand and supply curves sweeping out time and product group effects. The double-differenced demand and supply shocks are assumed to be uncorrelated but heteroksedastic across products. This yields a GMM specification for estimation. As in Redding and Weinstein (2020), the weighting matrix is based on quantity weights.

common goods rule and without any nesting within product groups. The Lambda Ratio and S^* ratio components in the figure are scaled by $\frac{1}{\hat{\sigma}-1}$ so that the CUPI is the sum of the three components (see equation (16)). We find that the CUPI shows lower inflation than the Feenstra index and very low inflation in absolute terms. In all goods but occupational footwear, the CUPI produces an estimate of 30 to 40 percent annual declines in the price level, and it often falls 10 to 30 percentage points more quickly than the Feenstra index.

The large differences between the Feenstra index and the CUPI in these product groups arise from two sources. The first is the difference between the P^* ratio (Jevons index) and the Sato-Vartia index. The Sato-Vartia is a weighted average log price change among common goods, whereas the P^* ratio is an unweighted average. In boys' jeans, for instance, the P^* ratio is far below the Sato-Vartia. The difference between the weighted and unweighted log price ratios for common goods suggests there are a large number of low-share goods experiencing price declines that are driving down the CUPI. The second source is the introduction of the S^* ratio in the CUPI, intended to account for changing consumer tastes. Almost uniformly, the S^* ratio contributes a large downward shift to the CUPI. It is also an unweighted geometric mean that is sensitive to low-share goods.

The CUPI's sensitivity to low-share goods led Redding and Weinstein (2020) to introduce a common goods rule (or CGR) to the index. We use a related but distinct methodology that can be implemented in real time using only current and backward looking information available in quarter t. For our NPD analysis, we specify a market share threshold for goods present in periods t and t-1 to be considered as common goods for the Jevons and the S^* ratio terms of the CUPI.[24] Goods below this threshold are excluded from the set of common goods, but they still enter the CUPI through their inclusion in the Feenstra adjustment term (Lambda ratio). We consider alternative market share percentile thresholds.[25]

[24]The details of the procedure are as follows. Compute the Xth percentile of the expenditure shares within product groups in both period t-1 and period t. A common good must exceed the Xth percentile in both periods.

[25]In our analysis of the NielsenIQ data, which is a longer panel, we consider further alternative approaches to define common goods. In our analysis of chain drift below, we also consider the impact of the CGR implemented over a longer horizon in the NPD data.

Figure 6 illustrates the CUPI's sensitivity to the CGR for different market share thresholds. We consider market share thresholds at the 10th, 30th, and 50th percentiles for continuing goods in t and t-1. We depict the CUPI for these different CGRs alongside the Sato-Vartia index, the Feenstra index, and the CUPI without a CGR. Implementing the restriction on the set of common goods by market share raises the CUPI by cutting off the low end of the market share distribution from unweighted relative comparisons and shifting it to the weighted entry/exit adjustment term. In that sense, applying a stricter CGR moves the CUPI closer to the Feenstra-adjusted Sato-Vartia index, which combines a traditional matched model index with an adjustment for entry and exit.

The resulting price indices generally shift up as successively stricter definitions of common goods are imposed. For some product groups, such as memory cards, the CUPI using the CGR at the 30th or 50th percentile yields inflation measurements similar to the Feenstra index. For products groups such as headphones and boys' jeans, however, the CUPI shows noticeably lower inflation than the Feenstra index even using a 50th-percentile CGR threshold (i.e., shifting half of products from the set of common goods to entering and exiting goods).

The key takeaway from this analysis is that the CUPI is sensitive to the specific definition of the CGR, and that sensitivity varies across product groups. In contrast to the finding in Redding and Weinstein (2020) that the CUPI eventually stabilizes as successively stricter duration-based CGRs are applied, we do not find evidence that the CUPI stabilizes as stricter share-based CGRs are applied. A 50th-percentile threshold for the market share of goods present in t and t-1 implies that an entering good does not count as a common good until it reaches the top half of the market share distribution. Similarly, a good that is on its way to exit and that falls below the 50th percentile of market share is put into the entry/exit group (and becomes part of the Feenstra adjustment term).

Empirical Performance of the CUPI with NPD Data

We are sympathetic to the view that some form of a CGR is a sensible and necessary component of empirically implementing the CUPI. The primary inference we draw from our analysis using the NPD data is that the CUPI is sensitive to the specification of the CGR in a manner that varies across product groups, and more research is necessary on best empirical practice in implementing the index. Further research into the dynamic process of the entry and exit of goods should be a part of such research. Our analysis in Figure 1 above, which shows the product lifecycle dynamics of prices and shares, is a step in that direction. It is likely that process varies by product group, consistent with our results showing the CUPI's differential sensitivity to various CGRs across product groups in the NPD data.

A critical insight underlying taste shock bias is that product appeal shocks are needed to reconcile the micro evidence on prices and quantities. The theory implies that if a product exhibits a low price but low expenditures it must be that this is because of low product appeal. However, there are factors outside the model underlying the CUPI that imply that small expenditure shares are not mediated by price and appeal alone. These factors are especially relevant for how entering and exiting products are treated in the CUPI. For example, consider the close-out of products about to be discontinued (e.g., clearance racks). Consumers might be very happy to buy such goods at the low price offered, but might not do so because they are not widely available or consumers might not be aware of their availability. Hence, search frictions, lack of widespread product availability, etc., can explain low expenditure given price rather than low appeal.

It is these type of frictions that underlie the need for the common goods rule for empirical implementation of the CUPI. The challenge is that the nature of these frictions arguably varies across product groups. Our finding of differential sensitivity of the CUPI to common goods rules across product groups in the NPD data is consistent with this interpretation. We explore frictions in Appendix C that may be especially relevant for entering and exiting goods that help to justify a CGR. This analysis is illustrative, with the primary inference that more research is needed to provide practical guidance for implementation of the CUPI with a CGR at scale.

The NPD data offers additional opportunities to explore sources of variation in the novel taste shock component of the CUPI. One hypothesis we investigate is that the large measured taste shock bias without a CGR reflects in part the need for a richer nested CES. That is, it might be that within product groups there are sub-groups with greater elasticities of substitution than between sub-groups. The measured contribution of the S* ratio will be smaller within sub-groups with larger elasticities. We explore this hypothesis with two different approaches to nesting: a characteristics-based approach including brands and an approach based on predicted prices using hedonic regressions. Details are in Appendix C.1. Neither nesting method materially changes the inflation measurements of the CUPI, suggesting that aggregation issues are not the primary driver of the large measured taste shock bias in the NPD product groups we consider.[26]

Another possible source of large systematic positive taste shock bias is rapidly rising dispersion of relative product appeal.[27] In Appendix C.2.4 we investigate the patterns of product appeal dispersion for each of the NPD product groups. We find some evidence of rising product appeal shocks for headphones and memory cards (especially without a CGR) but not for the other product groups including product groups with very large measured taste shock bias. The need for a CGR that varies across product groups appears to be independent of patterns of changing dispersion of relative product appeal.

4.1.3 Comparing Matched-Model, Hedonic, and Exact Price Indices

Figure 7 presents the key matched-model, hedonic, and demand-based price indices that we have considered for all five product groups. Table 2 reports the chained index levels in 2018:4, reflecting the cumulative price changes since 2014:4, when all indices are normalized to one. This table includes the arithmetic Laspeyres for comparison with the BLS's CPI approach.

[26]Martin (2020) also has suggested the possible importance of the need for a nested preference structure in implementing the CUPI.

[27]Redding and Weinstein (2020) highlight this is a possible source of positive taste shock bias and we present simulation evidence in Appendix C.2.3.1 in support of this hypothesis.

The price indices follow a roughly similar pattern of relative orders across these product groups. The Laspeyres matched model index typically shows the least deflation. The matched model Tornqvist and Sato-Vartia tend to track each other closely and to show more rapid deflation than the Laspeyres, as expected given that they are both superlative price indices that account for substitution. The Feenstra and hedonic Tornqvist using the EP-TV method in turn tend to show greater deflation than their unadjusted counterparts, indicating the importance of product entry and exit. Finally, the CUPI (both baseline and nested by product characteristics) shows the greatest deflation, especially for headphones and boys' jeans. The substantial gap between the CUPI and the Feenstra index in headphones and boys' jeans is especially striking given our imposition of a 30th-percentile CGR. It is also evident from Figure 7 that the CUPI exhibits very different time series patterns relative to the other indices.

The gap between the matched-model Laspeyres and Tornqvist indices for most product groups highlights the advantages of using item-level scanner data, which permits construction of a superlative price index with internally consistent prices and expenditure shares in adjacent periods. The cumulative gaps are on the order of 5–10 percentage points from 2014 to 2018 in coffee makers, occupational footwear, and boys' jeans, and significantly larger in the "high-tech" categories of memory cards and headphones. The gap also varies over time, consistent with the Laspeyres matched-model index exhibiting a time-varying substitution bias. Thus, using scanner data can produce substantial improvements in price measurement even without performing quality adjustment.

We also find that the hedonic Tornqvist using the EP-TV method tends to indicate larger cumulative quality adjustment than the Feenstra index. The Feenstra index indicates approximately 2 percentage points of cumulative disinflation beyond the Sato-Vartia index across all five product groups. The hedonic Tornqvist indicates roughly the same adjustment (relative to the traditional Tornqvist) for coffeemakers and occupational footwear, but larger adjustments, of roughly 5–7 percentage points, in memory cards, headphones, and boys' jeans.

4.2 NielsenIQ Results

The machine learning approach used to implement the EP-TV method with the NielsenIQ Scanner data works well in mapping the relationship between prices to the abbreviated text fields. The median in-sample R^2 of the hedonic (log-level) price predictions is roughly 85% for the food product groups. The median out-of-sample R^2 is roughly 75% for food product groups. The model's predictive performance is comparable to that of Bajari et al. (2021), who report out-of-sample R^2 values of 80–90% in their best-performing specifications using the rich product text and image information in their data set. For log price changes, our median in-sample R^2 is above 50% for the food product groups. The median out-of-sample R^2 values decline to nearly 20% for the food product groups. We consider our procedure to be very successful in light of the limited attribute information available in the data set.[28]

For the CES exact price indices, our empirical implementation in the NielsenIQ data largely follows our strategy in the NPD data. The Feenstra index and the CUPI require estimates of elasticities of substitution within product groups. As in the NPD data, we use the Feenstra (1994) procedure to estimate those elasticities. The estimated elasticities for the 50+ product groups in food display considerable variation. The median elasticity is about 6, the 10th percentile is about 4, and the 90th percentile is 12. These patterns are similar to those reported in Redding and Weinstein (2020).

We again explore alternative CGRs to calculate the CUPI. The NielsenIQ data provides a longer panel than the NPD data, which allows the exploration of alternative CGRs that depend on the duration of goods' time in the market to date. We implement a modified approach to defining the CGR rule in the NielsenIQ data as follows. We first compute percentiles of the pooled sales distribution within a narrow product group for pooled sales in periods t-1 and t. Common goods are defined as goods sold in both periods, and which have

[28]See appendix B.5 for more discussion of goodness of fit.

sales in period t above the Xth percentile of this pooled sales distribution. This alternative approach to defining the CGR allows us to consider longer duration-based alternatives.[29] Specifically, we consider specifications of the CGR using market thresholds using percentiles of sales pooled over over the current and prior 4 quarters. In addition and critically, a common good is defined in this context if it is present in periods t and $t-4$. Using a duration component in the CGR puts more weight on goods present for the longer horizon, yielding greater comparability with the duration-based CGR used by Redding and Weinstein (2020).

Figure 8 presents a full set of price indices for the NielsenIQ scanner food product groups in change and level forms. For all indices, we aggregate across product groups using a Tornqvist aggregator with Divisia-style product group market share weights. The panels of the figure include the BLS CPI computed for the same NielsenIQ product groups.[30] We find that the CPI and the matched model Laspeyres index track each other closely in NielsenIQ's food product groups for the first part of the sample period, with a discrepancy arising towards the end of the period. The Tornqvist and Sato-Vartia indices are lower and track each other closely. The quality-adjusted indices (Feenstra, hedonic Tornqvist using the EP-TV approach, and CUPI) are even lower.

Using a 25th-percentile CGR, the CUPI is more than 40 percentage points lower than the Feenstra index in 2015; using a 50th percentile CGR reduces the difference to 20 percentage points. Alternatively, using the longer $t-4$ to t horizon described above, the 10th-percentile CGR yields a difference of about 25 percentage points.[31]

The cumulative level implications highlight that the hedonic Tornqvist is about 4 per-

[29] In unreported results, we have found that the NielsenIQ results using the identical CGR used in the NPD data yields very similar results to those reported here using a two-quarter horizon.

[30] We thank the BLS for producing these calculations.

[31] In appendix B.4, we also show the sensitivity of the CUPI to the CGR using the NielsenIQ Consumer Panel. We explore the sensitivity of the CUPI to the CGR in the Consumer Panel as this is the data used by Redding and Weinstein (2020). Results are broadly consistent. We note that Redding and Weinstein (2020) focus on a year-over-year (YOY) index rather than a chained quarterly index (although they use quarterly data). The use of a YOY index restricts entering (exiting) goods to those entering (exiting) 4 quarters ago mitigating the need for a CGR. Even so, they impose a strict CGR as noted above.

centage points lower in 2015 than the matched-model Tornqvist, and the Feenstra index is about 5 percentage points lower than the Sato-Vartia. These substantial cumulative differences for the food product groups suggest that quality improvement via product turnover has not been limited to products where technological progress is most visible. It is also striking that these two distinct relative comparisons yield such similar quantitative implications.

We consider the patterns in the NielsenIQ data to be broadly similar to the patterns in the NPD data. Quality adjustment, either via hedonic approaches or the Feenstra product turnover adjustment, imparts a substantial downward adjustment on price indices. The CUPI suggests an even larger quality adjustment, but we note again its sensitivity to the CGR. This sensitivity manifests across alternative approaches to defining the CGR thresholds for common goods.

4.3 Chain Drift

Chain drift occurs when price changes cumulated ("chained") over multiple periods do not return to the same value between a base period and a subsequent period even if relative prices have returned to base period values (e.g., Diewert, 2021). Chain drift arises from high-frequency volatility in price changes. Chain drift is particularly prevalent with indices computed from local transactions data, where prices are particularly volatile (e.g., De Haan and Van Der Grient, 2011). Our analysis uses national data at a quarterly frequency, which mitigates this issue. Given our focus on comparing alternative approaches for computing price indices, we examine whether GEKS-type indices (Gini, 1931; Eltetö and Köves, 1964; Sculz, 1964), which involve taking the geometric mean of price indices calculated over multiple horizons, preserve the implications of our core findings.

We follow Bajari et al. (2021) by computing a GEKS-type index (which we denote "GEKS-lite") that is the geometric mean of the chained year-over-year index for the 4th quarter of each year and the directly computed (unchained) year-over-year price index for the 4th quarter. GEKS-lite offers an easily implementable alternative to a full GEKS procedure, which involves computing price indices over many possible horizons. Given that we are including hedonic indices which require re-estimation of models for alternative time horizons, the computational burden of implementing a full GEKS procedure is substantial. Table 4 reports average annual chained and GEKS-lite indices for the five NPD product groups and alternative indices. The GEKS-lite price change index tends to show less deflation than the chained price index for traditional matched-model price indices. For the hedonic indices, the GEKS-lite indices actually show faster deflation in three out of the five product groups.[32] Among the demand-based indices, the GEKS-lite indices typically show greater deflation, but these differences are modest quantitatively. The GEKS-lite CUPI shows substantially less deflation, but this difference also reflects the effects of applying a common good rule over a longer horizon.[33]

The key result of this analysis is that applying the GEKS-lite procedure does not change the rank ordering of the various indices we have considered. The Laspeyres matched-model index yields higher inflation than the matched-model Tornqvist, which in turn is higher than the hedonic Tornqvist. Likewise, the Sato-Vartia yields higher inflation than the Feenstra, which in turn is higher than the CUPI. Notably, the traditional matched-model Tornqvist is more sensitive to chain drift than the hedonic Tornqvist. The reduced sensitivity of full-imputation hedonic indices to chain drift is intuitive since hedonic prices are less subject to transitory price volatility (which is the source of chain drift).

We also implement the rolling-year GEKS method (Ivancic et al., 2011) as a point of comparison. Given that this procedure is computationally burdensome, we only implement this for matched-model price indices. In appendix Table D.8, we show that results for rolling year GEKS are similar to those using the GEKS-lite approach.

Table 5 reports analogous chained and GEKS-lite indices for the aggregated food indices.

[32] The results in Table 3 imply that these patterns are robust to using alternative methods for imputing the missing lagged residual for entrants.

[33] The longer horizon affects the CGR because for the year-over-year measure the good must not only be above the Xth percentile in the appropriate samples, but also be present in quarters t and $t-4$, as opposed to quarters t and t-1.

The table reports average annual indices for both specifications. The results for NielsenIQ's food product groups show that we obtain similar, albeit slightly higher, rates of average inflation using the GEKS-lite compared to the chained indices. This pattern is especially noticeable for the CUPI, a result which again reflects the effects of applying the CGR over a longer horizon. Importantly, the rank ordering and the quantitative differences across alternative indices are preserved using the GEKS-lite based indices. Focusing on the GEKS-lite indices, inflation in NielsenIQ's Food product groups is higher using the Laspeyres index than the Tornqvist, higher using the Sato-Vartia than the Feenstra, and higher using the Feenstra than the CUPI.[34]

5 Conclusion

Incorporating transactions data into official statistics requires making several methodological decisions. Putting aside the important issues of quality change and product turnover, matched model price indices constructed from item-level data possess several advantages relative to the current system: the expenditure shares from the item-level data are internally consistent with the price data, and they are also available in real time. The data therefore permit the construction of superlative price indices such as the Tornqvist in real time. We find that the matched-model Tornqvist index tends to measure systematically lower inflation than the matched-model Laspeyres, with the gap varying over time and product groups.

The item-level data also importantly permit addressing product turnover and quality change in real time at scale. We find that implementing the time-varying unobservables hedonic estimation approach from Erickson and Pakes (2011) is (i) feasible at scale and (ii) yields evidence of systematic quality change. We show that this method can be applied across multiple platforms and estimation approaches—econometric and machine learning.

A systematic pattern emerges: hedonic indices yield significantly lower rates of inflation

[34] We do not report the hedonic indices using the GEKS-lite procedure for NielsenIQ's food product groups because of the large computational burden that would be required to apply our machine learning procedure to additional comparison periods.

than associated matched model indices. These patterns hold for both Laspeyres indices, which provide bounds for the cost of living index under very general conditions, and for the Tornqvist superlative index, which provides a second-order approximation to the cost of living index under specific assumptions. We find evidence that product turnover and quality change affect inflation across a wide range of consumer goods. These effects are, not surprisingly, largest for tech goods. Nonetheless, for food sold in grocery stores, we find significant quality change not captured by conventional price indices, which leads them to meaningfully overstate the increase in the cost of living over time. We thus demonstrate that item-level transaction data present the opportunity for statistical agencies to adjust for product turnover and quality change at scale without selecting only product groups for hedonic adjustment that are deemed *ex ante* to be the most important for this type of adjustment. The machine learning approach shows that these adjustments can be made in data sources that lack well-encoded product attributes, a situation that often arises in item-level transactions data.

We also find that the demand-based indices that incorporate quality adjustment—particularly the Feenstra index—provide useful benchmarks that should be used for purposes of comparison with the hedonic and matched-model indices. The recently developed CUPI is a major conceptual breakthrough incorporating the potential contribution of taste shock bias on the cost of living, yet its current empirical implementations feature some limitations. The key limitation is that the measurement of taste shock bias is sensitive to the definition of common goods – i.e., the Common Goods Rule (CGR). While it makes sense that the CUPI is especially sensitive to the definition of common goods, we have found that its sensitivity to CGRs varies substantially across product groups. This variation likely reflects differences across product groups in the product life cycle and associated entry and exit dynamics of products.

Item-level transactions data with price, quantity, and attribute information enable considerable advances in the production of price indices. They permit accounting for product turnover and quality change along with substitution effects in real time. The availability of contemporaneous price and quantity data and of rich item descriptions is fundamental to implementing these innovations. The increased availability of "big data" therefore has considerable promise for improving price measurement.

There are several important issues to consider when assessing the applicability of our findings for the statistical agencies. Statistical agencies are necessarily and justifiably conservative in making changes to official statistics. They are responsible for feasibly producing statistics that are reliable, credible, and consistent across long periods of time. Any changes must be transparent and understandable to multiple constituencies—including policymakers, businesses, the news media, researchers, and the public. Changing the CPI is especially sensitive because of its legal status for indexing a variety of government programs, inflation-protected securities, and some private contracts. The agencies have nevertheless made significant changes in methodology that affect headline inflation, such as using geometric means for low-level aggregation of prices or introducing hedonics for selected goods. Hence, even in the current infrastructure, price index methods are not static, and agencies have shown willingness to modify them.

Because of the massive turnover of goods at the item level, agencies need to go beyond traditional matched-model approaches that require interventions and judgments by price enumerators or analysts when items disappear and appear. It is simply not feasible to implement the current practice for the number of goods available at the item-level. This paper suggests that hedonic approaches—including using machine learning for encoding product attributes and estimating hedonic models—are feasible and have desirable properties. Introducing hedonics at scale, i.e., for essentially all categories of goods and services rather than the small fraction of categories agencies currently select for hedonic adjustment, can replace the current, category-by-category approach to hedonics. Nevertheless, given that selective use of hedonic adjustment is already a feature of official statistics, incorporating hedonic adjustment at scale can be accurately communicated to data users as a natural advance on current practice.

Quite apart from the promise of improved measurement of the cost of living, agencies must consider the innovations such as those studied in this paper because of the increased cost of a measurement infrastructure based on surveys and enumerations and the growing misalignment between how information is processed in the economy and by the agencies. Our paper is an important step in demonstrating both the value and the feasibility of agencies aligning their infrastructure with that of the twenty-first century information economy.

This paper examines consumer goods. We view this as a point of departure, not a limitation of the approach. While scanning of retail transactions and the availability of item-level transactions data makes the retail goods sector a natural starting point for re-engineering, many consumer services, such as utilities and telecommunications, transportation services, and health care, are mediated by electronic records with the same granularity as retail scanner data. Similarly, most business-to-business transactions and government transactions have electronic records. The practical challenges and institutional arrangements for using these data for statistical purposes are of course substantial. Nonetheless, the methods evaluated in this paper are not specific to retail goods, and could be straightforwardly (if not easily) applied across much of the economy.

This paper is a step in demonstrating that using item-level transactions data at scale can lead to a re-engineering of key national indicators. It shows that accounting both for substitution effects and for quality change substantially lowers the estimates of inflation rates across a wide range of goods. These innovations therefore should have important implications for understanding the average rate of price change, with further implications for estimates of the rate of growth of output and productivity.

References

P. Bajari, Z. Cen, V. Chernozhukov, M. Manukonda, J. Wang, R. Huerta, J. Li, L. Leng, G. George Monokroussos, S. Vijaykunar, and S. Wan. Hedonic prices and quality adjusted price indices powered by ai. Technical report, The Institute for Fiscal Studies Department of Economics, UCL, CENMAP working paper CWP 04/21, 2021.

W. A. Barnett and K.-H. Choi. Operational identification of the complete class of superlative index numbers: An application of Galois theory. *Journal of Mathematical Economics*, 44(7):603–612, 2008. ISSN 0304-4068. doi: https://doi.org/10.1016/j.jmateco.2006.05.002. URL https://www.sciencedirect.com/science/article/pii/S0304406806000589. Special Issue in Economic Theory in honor of Charalambos D. Aliprantis.

C. L. Benkard and P. Bajari. Hedonic price indexes with unobserved product characteristics, and application to personal computers. *Journal of Business & Economic Statistics*, 23(1): 61–75, 2005.

R. Blundell and S. Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87:115–143, 1998.

M. J. Boskin, E. R. Dulberger, R. J. Gordon, Z. Griliches, and D. W. Jorgenson. Consumer prices, the consumer price index, and the cost of living. *Journal of economic perspectives*, 12(1):3–26, 1998.

C. Broda and D. Weinstein. Product creation and destruction: Evidence and price implications. *American Economic Review*, 100:691–723, 2010.

D. M. Byrne, D. E. Sichel, and A. Aizcorbe. Getting Smart About Phones: New Price Indexes and the Allocation of Spending Between Devices and Services Plans in Personal Consumption Expenditures. *Finance and Economics Discussion Series*, 2019(012), 2019. ISSN 19362854. doi: 10.17016/feds.2019.012.

M. Cafarella, G. Ehrlich, T. Gao, J. C. Haltiwanger, M. D. Shapiro, and L. Zhao. Using machine learning to construct hedonic price indices. Working Paper 31315, National Bureau of Economic Research, June 2023. URL http://www.nber.org/papers/w31315.

J. De Haan. Hedonic price indexes: a comparison of imputation, time dummy and other approaches. 2008.

J. De Haan and H. A. Van Der Grient. Eliminating chain drift in price indexes based on scanner data. *Journal of Econometrics*, 161(1):36–46, mar 2011. doi: 10.1016/j.jeconom.2010.09.004.

E. Diewert. Quality adjustment and hedonics: A unified approach. Working paper, University of British Columbia, 2019.

E. Diewert. The Economic Approach to Index Number Theory. Technical Report Discussion Paper 20-05, Vancouver School of Economics, University of British Columbia, 2021. URL https://econ.cms.arts.ubc.ca/wp-content/uploads/sites/38/2020/04/pdf_.paper_erwin-diewert_DP2005EconomicApproachtoINdexNumberTheory.pdf.

E. Diewert, S. Heravi, and M. Silver. Hedonic Imputation versus Hedonic Time Dummy Indexes. Technical report, NBER WP 14018, 2008.

W. Diewert. Exact and superlative index numbers. *Journal of Econometrics*, 4(2):115–145, 1976. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(76)90009-9. URL https://www.sciencedirect.com/science/article/pii/0304407676900099.

W. E. Diewert. Superlative index numbers and consistency in aggregation. *Econometrica: Journal of the Econometric Society*, pages 883–900, 1978.

G. Ehrlich, J. Haltiwanger, R. Jarmin, D. Johnson, and M. D. Shapiro. Re-engineering key national economic indicators. *Big Data for Twenty-First Century Economic Statistics*, 2022.

L. Einav, E. Leibtag, and A. Nevo. Recording discrepancies in nielsen homescan data: Are they present and do they matter? *QME*, 8:207–239, 2010.

O. Eltetö and P. Köves. On a problem of index number computation relating to international comparison. *Statisztikai Szemle*, 42(10):507–518, 1964. (in Hungarian).

T. Erickson and A. Pakes. An experimental component index for the CPI: From annual computer data to monthly data on other goods. *American Economic Review*, 101(5): 1707–1738, aug 2011. ISSN 00028282. doi: 10.1257/aer.101.5.1707.

R. C. Feenstra. New Product Varieties and the Measurement of International Prices. *The American Economic Review*, 84(1):157–177, 1994.

R. C. Feenstra and M. B. Reinsdorf. *Should Exact Index Numbers Have Standard Errors? Theory and Application to Asian Growth*. Number October. 2007. ISBN 9780226044491. doi: 10.3386/w10197. URL http://www.nber.org/papers/w10197{%}5Cnhttp://www.nber.org/papers/w10197.pdf.

C. Gini. On the circular test of index numbers. *Metron*, 9(9):3–24, 1931.

A. Gorajek. Generalizing the stochastic approach to price indices. *Review of Income and Wealth*, pages 1–22, October 2022.

R. J. Hill and D. Melser. Hedonic imputation and the price index problem: an application to housing. *Economic Inquiry*, 46(4):593–609, 2008.

C. J. Hottman, S. J. Redding, and D. E. Weinstein. Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics*, 131(3):1291–1364, 2016.

L. Ivancic, W. E. Diewert, and K. J. Fox. Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics*, 161(1):24–35, 2011.

X. Jaravel. Distributional consumer price indices. 2024.

A. A. Konüs. The problem of the true index of the cost of living. *Econometrica*, pages 10–29, 1939.

R. Martin. Changing Tastes Versus Specification Error in Cost-of-Living Measurement. BLS Working Paper 531, 2020.

National Retail Federation. 4-5-4 calendar, 2023. URL https://nrf.com/resources/4-5-4-calendar#category.name. Accessed February 10, 2023.

A. Pakes. A reconsideration of hedonic price indexes with an application to PC's. *American Economic Review*, 93(5):1578–1596, 2003. ISSN 00028282. doi: 10.1257/000282803322655455.

A. Pakes. Hedonics and the consumer price index. *Annales d'Économie et de Statistique*, (79/80):729–748, 2005. ISSN 0769489X, 22726497. URL http://www.jstor.org/stable/20777594.

S. J. Redding and D. E. Weinstein. Measuring Aggregate Price Indexes with Demand Shocks: Theory and Evidence for CES Preferences. 2016.

S. J. Redding and D. E. Weinstein. Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences. *The Quarterly Journal of Economics*, 135(1): 503–560, 2020. ISSN 0033-5533. doi: 10.1093/qje/qjz031.

S. J. Redding and D. E. Weinstein. Accounting for trade patterns. *Journal of International Economics*, 150:103910, 2024. ISSN 0022-1996. doi: https://doi.org/10.1016/j.jinteco.2024.103910. URL https://www.sciencedirect.com/science/article/pii/S0022199624000345.

M. B. Reinsdorf and A. H. Dorfman. The sato–vartia index and the monotonicity axiom. *Journal of Econometrics*, 90(1):45–61, 1999. ISSN 0304-4076. doi: https://doi.org/10.1016/S0304-4076(98)00027-X. URL https://www.sciencedirect.com/science/article/pii/S030440769800027X.

K. Sato. The Ideal Log-Change Index Number. *The Review of Economics and Statistics*, 1976. ISSN 00346535. doi: 10.2307/1924029.

B. Sculz. Index numbers for multiregional comparisons. *Przeglad Statystyczny*, 3:239–254, 1964.

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

J. Triplett. Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes. 2004. doi: 10.1787/643587187107. URL http://dx.doi.org/10.1787/643587187107.

Y. Vartia. Ideal Log-Change Index Numbers. *Scandinavian Journal of Statistics*, 3(3): 121–126, 1976.

Table 1: Rates of Product Turnover: NPD Data

	Entry Rate		Exit Rate	
	All	Initial	All	Final
Memory Cards	5.8%	3.0%	6.0%	3.3%
Coffee Makers	5.7%	3.4%	4.5%	2.1%
Headphones	6.4%	3.8%	5.5%	2.9%
Boys' Jeans	11.5%	8.3%	7.8%	4.3%
Occupational Footwear	13.5%	9.1%	10.6%	5.5%

Notes: Average quarterly rates of product turnover. Entry/exit rates are computed as the number of entering/exiting goods as a percentage of common goods in the previous period. "Initial" entries are those for which the product was never observed in the data prior to the quarter. "All" entries include entries in which the product was previously observed prior to a spell of absence and the re-entered the data (i.e., "re-entries"). "Final" exits are those for which the product was never again observed in the data after the quarter. "All" exits include exits for which the product is subsequently observed after a temporary spell of absence (i.e., "temporary exits"). Transition quarter between data vintages excluded. Data come from NPD Group.

Table 2: Alternative Price Indices, Level in 2018:4 Relative to 2014:4=1

	Memory Cards	Coffeemakers	Headphones	Boys' Jeans	Occupational Footwear
Laspeyres (arith)	0.639	0.796	0.847	0.820	0.916
Hedonic Laspeyres (arith), EP-TV	0.430	0.697	0.614	0.748	0.870
Laspeyres (geo)	0.539	0.749	0.605	0.735	0.887
Hed. Laspeyres (geo), EP-TV	0.414	0.683	0.494	0.709	0.859
Tornqvist	0.467	0.688	0.607	0.726	0.872
Hed. Tornqvist, EP-TV	0.399	0.666	0.541	0.680	0.856
Sato-Vartia	0.481	0.706	0.602	0.773	0.879
Feenstra	0.469	0.685	0.582	0.749	0.857
CUPI, CGR 30p	0.389	0.625	0.332	0.181	0.777
CUPI-N, CGR 30p	0.367	0.640	0.349	0.173	0.780

Notes: CUPI-N is nested CUPI using characteristics approach.

Table 3: Alternative Treatment of Initial Residual for Entrants, Tornqvist (EP-TV) Price Growth, NPD Data

	Memory Cards	Coffeemakers	Headphones	Boys Jeans	Work/Occ Footwear
Method for Initial Residual					
		Chained Cumulative 4 quarter Annual Rates			
Zero	-20.12	-9.57	-14.13	-9.16	-3.79
Current	-20.13	-9.56	-13.93	-9.35	-3.82
Backcast	-20.13	-9.56	-13.93	-9.28	-3.82
		Direct Year over Year			
Zero	-20.52	-10.33	-16.43	-6.32	-3.45
Current	-20.55	-10.11	-16.89	-6.07	-3.33
Backcast	-20.63	-10.12	-16.81	-6.28	-3.40

Notes: Table shows alternative estimates of average annual price growth from 2014 to 2018.

Table 4: Alternative Price Growth Indices, Chained (C) vs GEKS-Lite (GL): NPD Data

	Memory Cards	Coffeemakers	Headphones	Boys' Jeans	Occupational Footwear
Laspeyres (C)	-13.89	-6.87	-11.74	-7.36	-2.92
Laspeyres (GL)	-12.16	-5.63	-11.77	-6.11	-2.15
Tornqvist (C)	-16.90	-8.86	-11.58	-7.63	-3.35
Tornqvist (GL)	-15.41	-6.64	-11.55	-5.56	-2.31
Hed.Tornqvist, EP-TV (C)	-20.1	-9.57	-14.13	-9.16	-3.79
Hed.Tornqvist, EP-TV (GL)	-20.6	-10.06	-14.51	-7.94	-3.76
Sato-Vartia (C)	-16.24	-8.24	-11.75	-6.20	-3.14
Sato-Vartia (GL)	-14.32	-6.36	-11.34	-4.13	-2.10
Feenstra (C)	-16.78	-8.92	-12.47	-6.92	-3.76
Feenstra (GL)	-16.46	-9.43	-13.06	-5.51	-3.80
CUPI,CGR 30p (C)	-20.64	-11.05	-24.08	-34.74	-6.08
CUPI,CGR 30p (GL)	-19.94	-9.41	-22.55	-26.91	-5.20

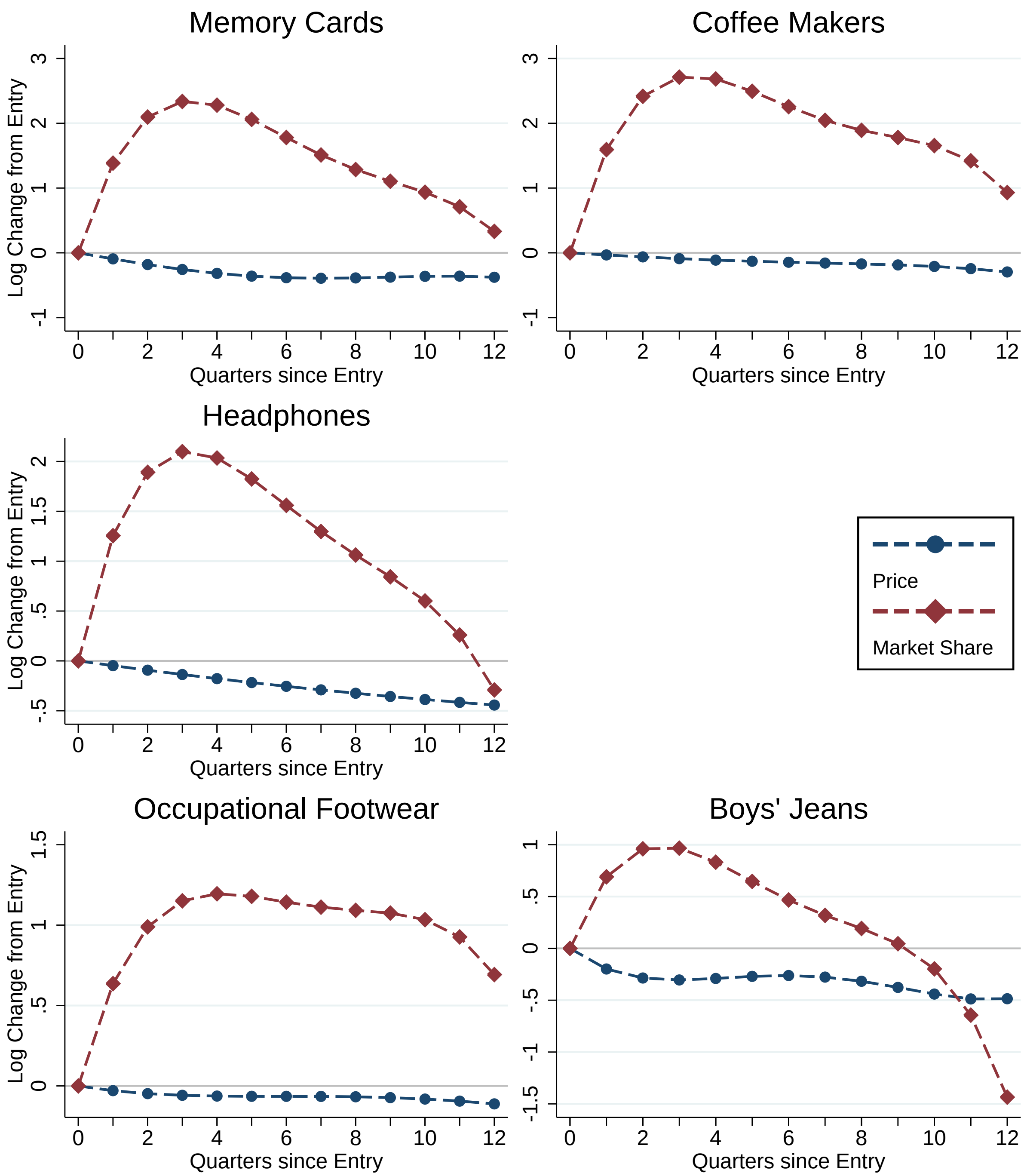
Notes: Chained values are averages of cumulative quarterly rates for year. Laspeyres is the geometric index. GEKS-lite is the average of the geometric mean of the chained values and the YoY price indices for q4 for each year.

Table 5: Alternative Price Growth Indices (%), Chained vs GEKS-Lite: NielsenIQ Food

Index	Chained	GEKS-Lite
Laspeyres	1.4	144
Tornqvist	0.5	0.9
Sato-Vartia	0.7	1.0
Feenstra	0.3	0.5
CUPI	-3.4	-2.0

Notes: Chained values are averages of cumulative quarterly rates for year. GEKS-lite is the average of the geometric mean of the chained values and the YoY price indices for q4 for each year. Laspeyres is the geometric Laspeyres. CUPI uses 25th percentile CGR. Data come from NielsenIQ.

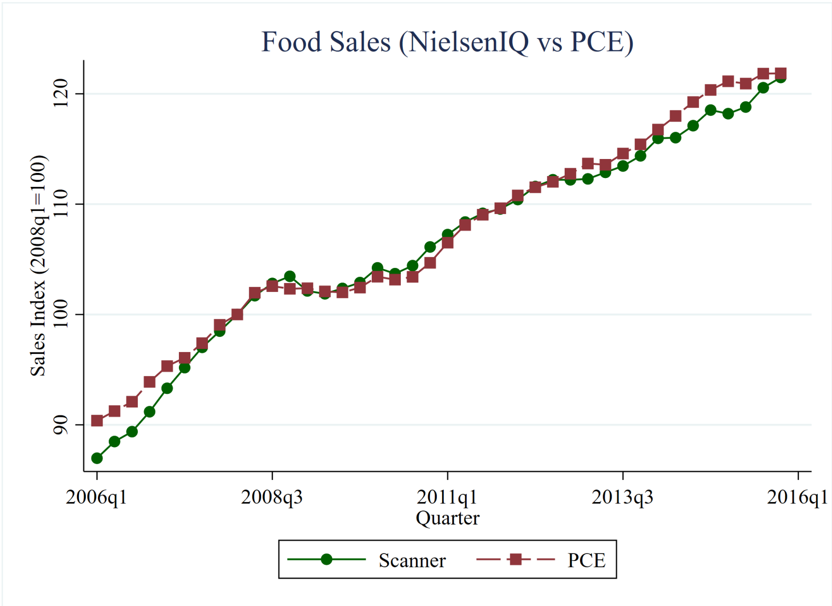
Figure 1: Product Lifecycle Dynamics, Market Share and Prices: NPD Data



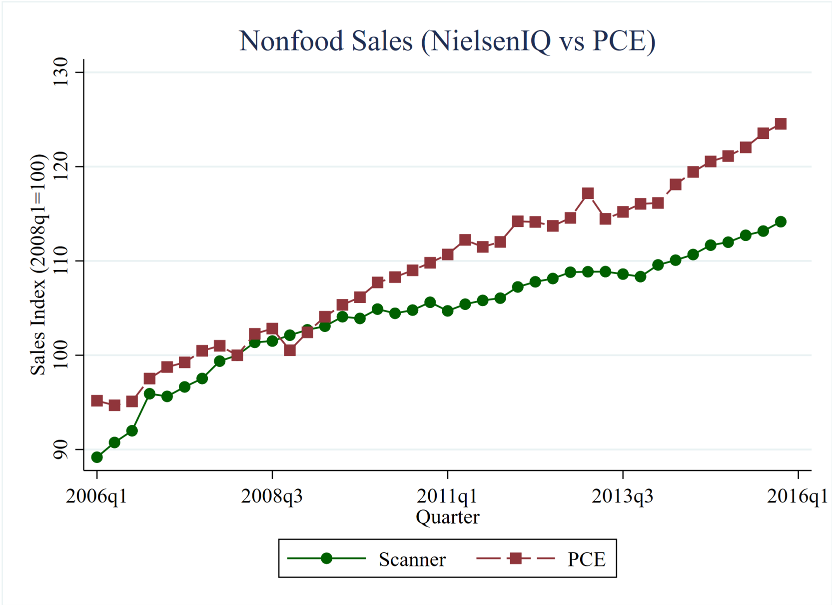
Notes: Unweighted average market share and prices relative to their value in the period of their initial entry. Entry occurs in period 0. All series are smoothed with a quartic spline.

Figure 2: PCE vs NielsenIQ Scanner Data Sales, Food and Nonfood

(a) Food



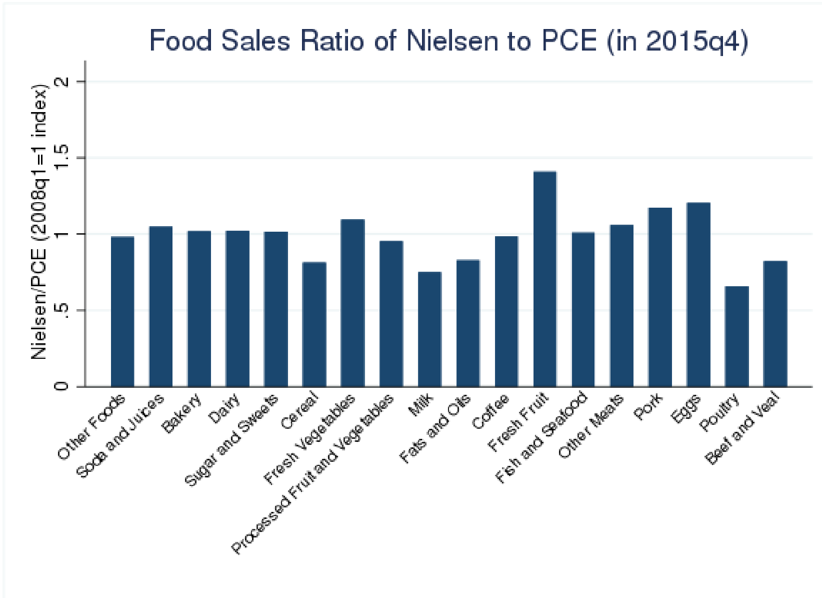
(b) Nonfood



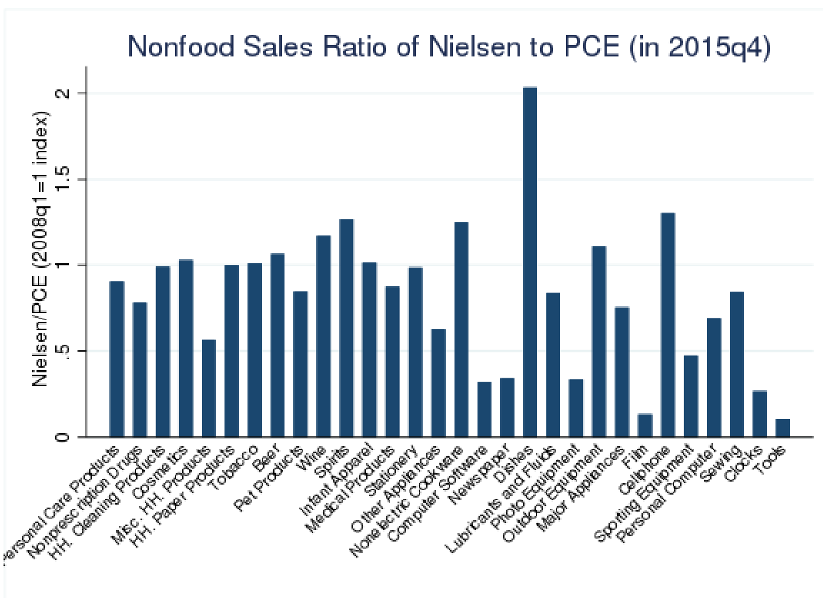
Notes: Figures uses NielsenIQ Scanner data for Food and Nonfood (aggregated) product groups. PCE is personal consumption expenditures (nominal) from BEA. All series indexed to 100 in 2008q1.

Figure 3: PCE vs NielsenIQ Scanner Data Sales, By Category within Food and Nonfood

(a) Food

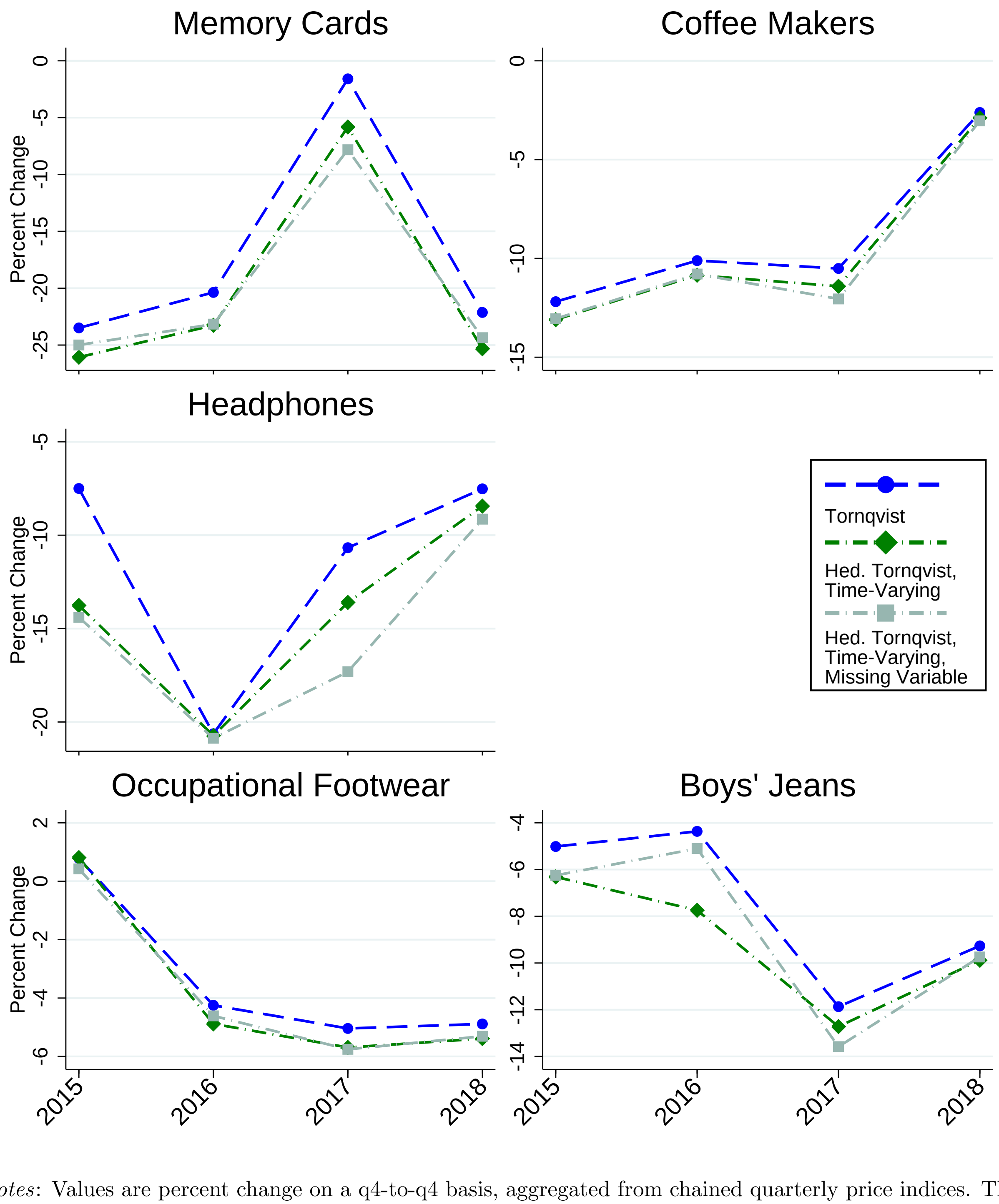


(b) Nonfood



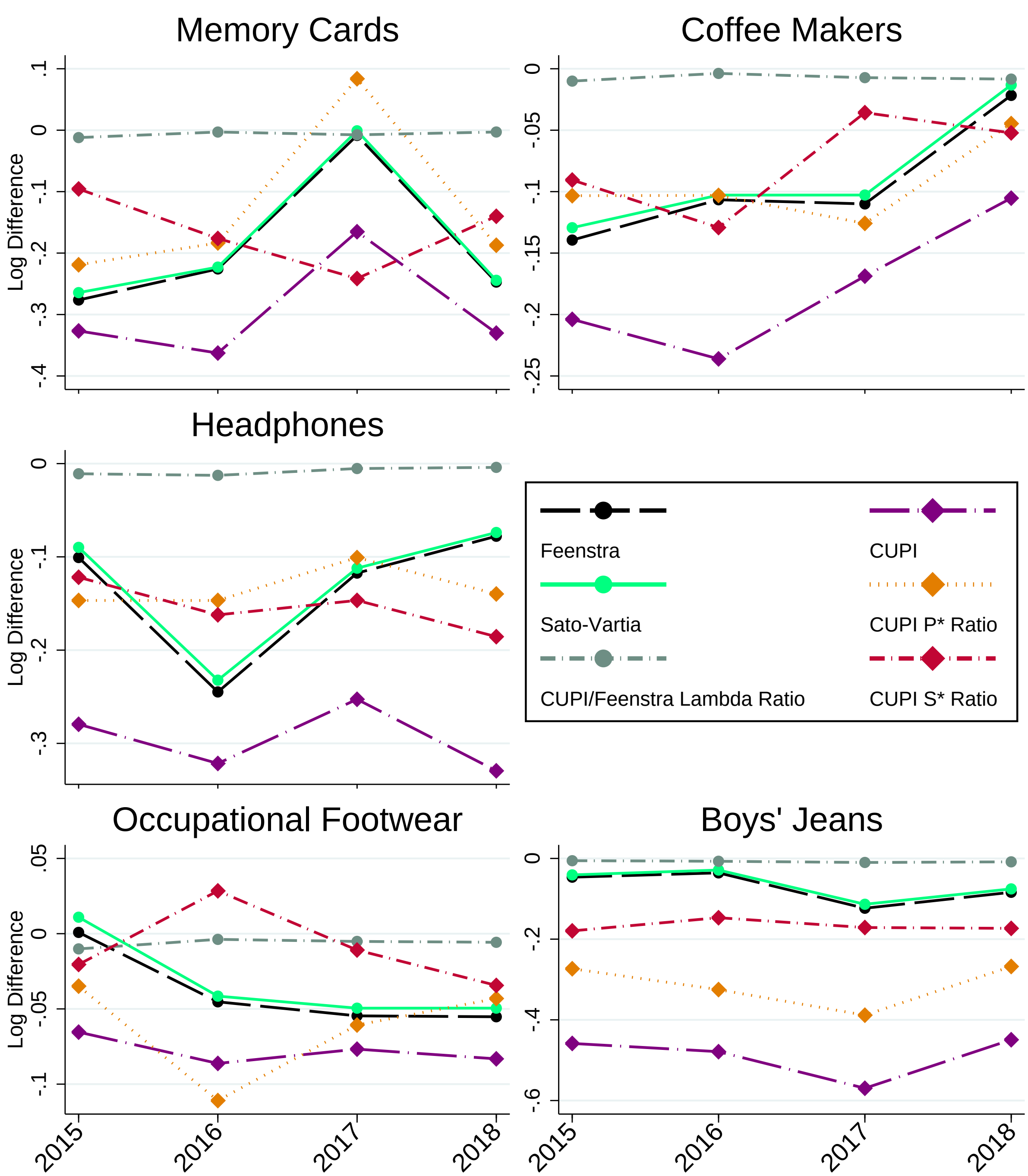
Notes: Figures use NielsenIQ Scanner data for Food and Nonfood (aggregated) product groups. PCE is personal consumption expenditures (nominal) from BEA. Each bar displays the ratio of nominal expenditures for a product group in the NielsenIQ Retail Scanner Panel to the corresponding PCE measure calculated as the ratio in 2015q4 to 2008q1. A value of one for a product group indicates that expenditures in the NielsenIQ data grew at the same rate as the PCE data over that period; values below one indicate slower growth in the NielsenIQ data.

Figure 4: Hedonic Specifications, Evaluating Time-Varying Unobservable Specification: NPD Data



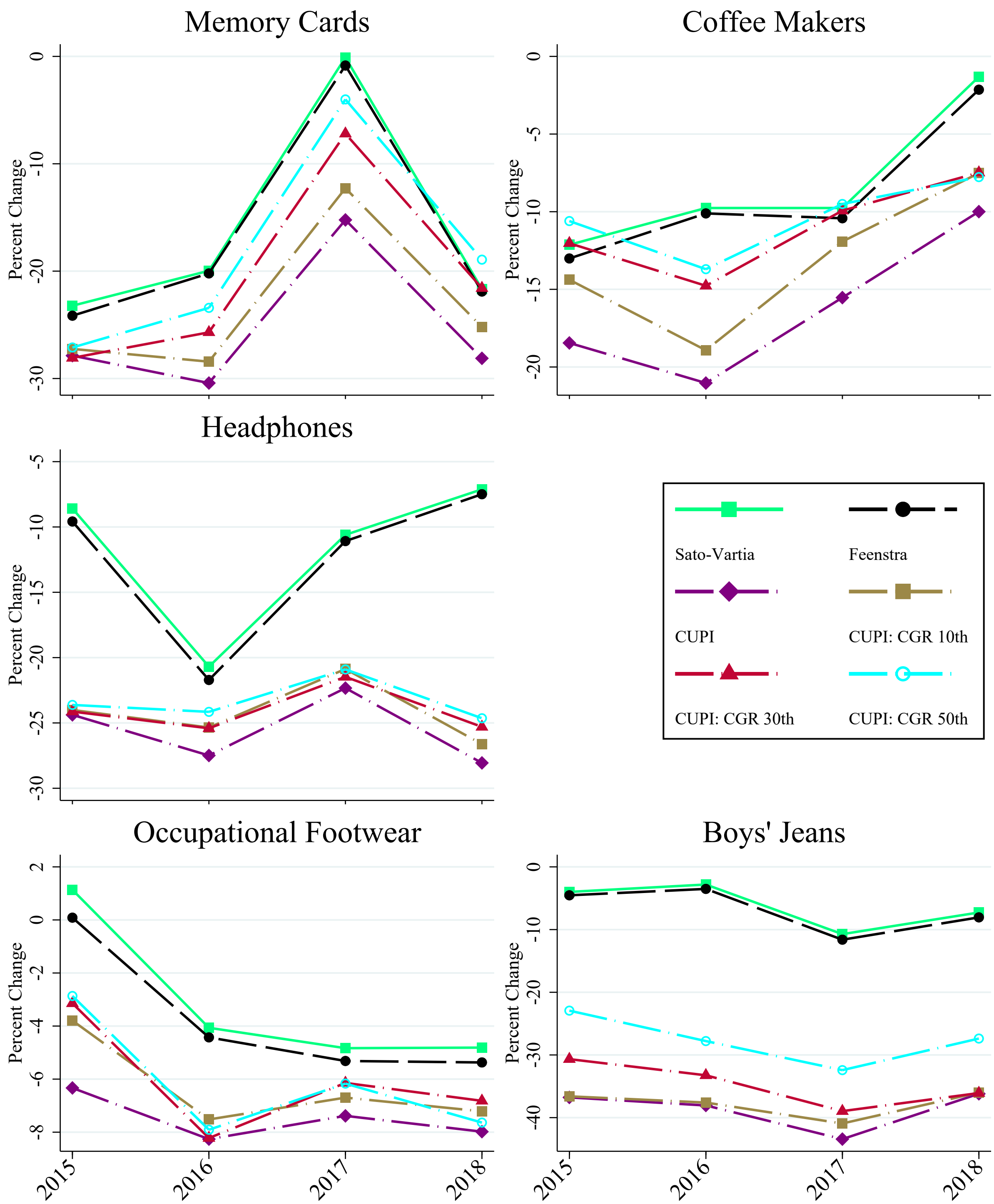
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. The time-varying unobservable model estimates hedonic models of log change in price using WLS and average quantity-share weights, including lagged hedonic level residuals. The "Missing Variable" series displays full imputation hedonic Tornqvist indices estimated using the time-varying unobservables approach, omitting key variables from the estimation.

Figure 5: Components of the Feenstra Index and CUPI: NPD Data



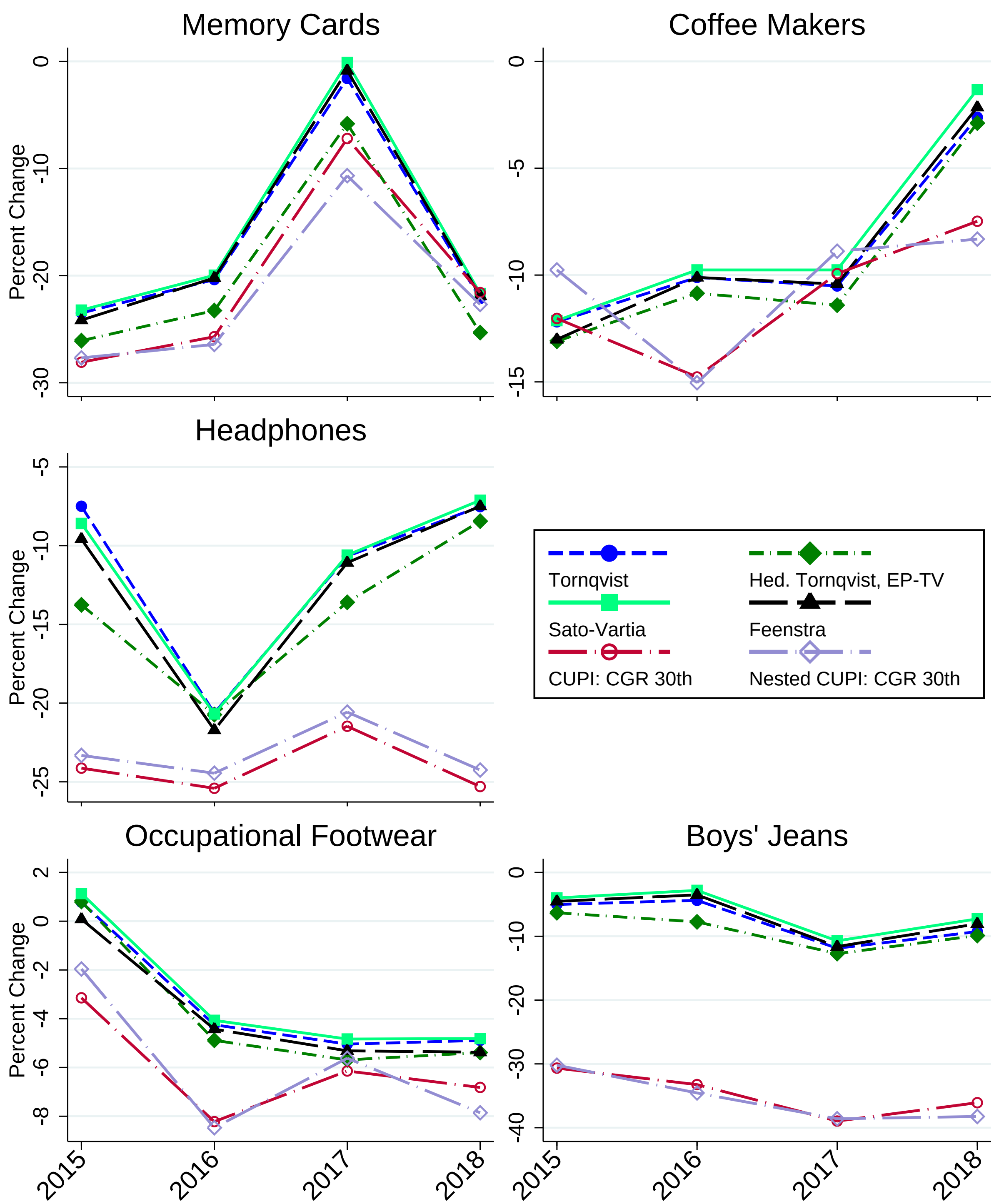
Notes:Values are log differences on a q4-to-q4 basis, aggregated from chained quarterly price indices. Units are reported in log-differences to allow for an additive decomposition of price indices. The Feenstra index is the sum of the Sato-Vartia and CUPI/Feenstra Lambda Ratio. The CUPI is the sum of the Lambda ratio, P^*-ratio, and S^*-ratio.

Figure 6: CUPI, Alternative Common Goods Rules (CGRs): NPD Data



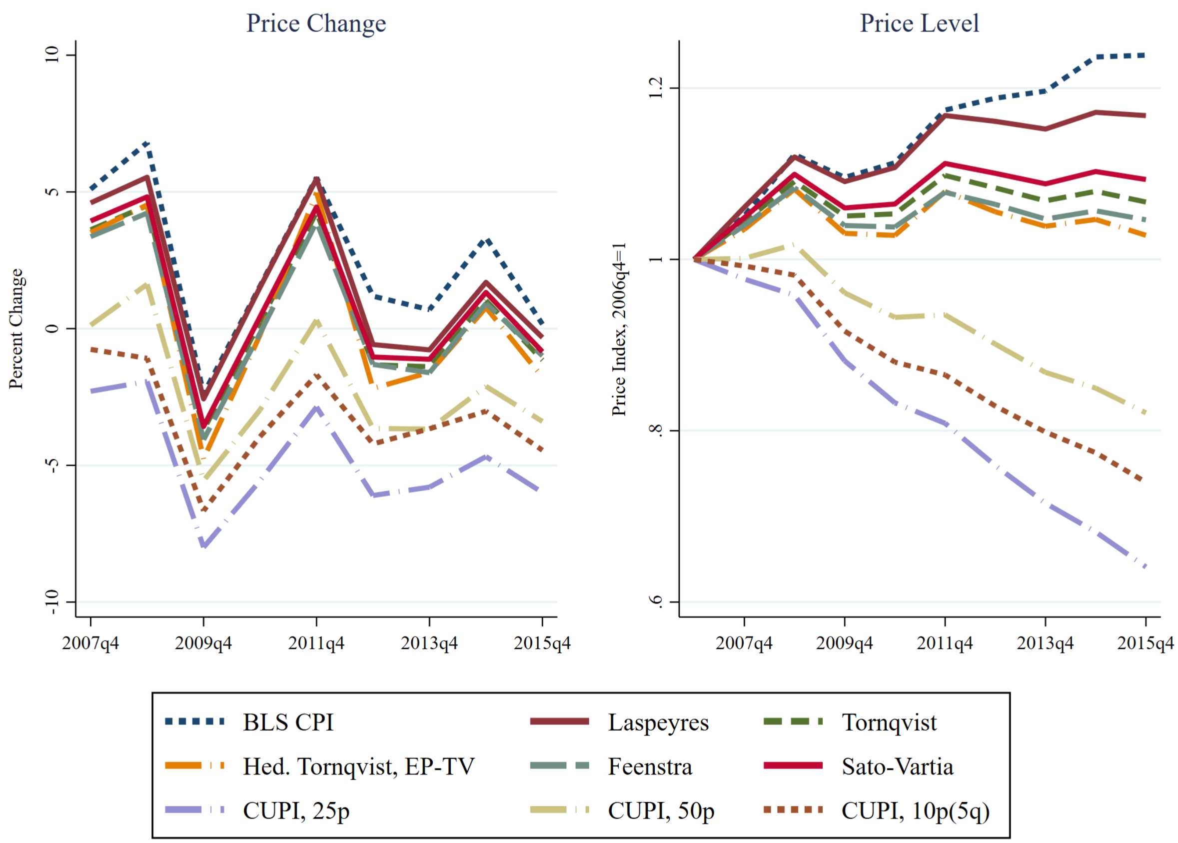
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. Common goods market share rules for the CUPI exclude from the group of common goods those products with market shares below the noted percentile in both periods. The Feenstra index is included for reference.

Figure 7: Comparison of Main Price Index Specifications: NPD Data



Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. "Hed. Tornqvist, EP-TV" is the hedonic time-varying unobservables model estimated over log price differences using WLS and with weights that are average quantity-shares in adjacent periods. The Nested CUPI uses within-product-group nests based on observable characteristics. For the Nested CUPI, the 30th-percentile market share common goods rule is applied within nests.

Figure 8: Main Price Index Specifications, Price Changes and Levels: NielsenIQ Food



Notes: Price changes are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. The price levels reflect chained quarterly values of each price index in the fourth quarter of each year, with the price level in the fourth quarter of 2006 normalized to one for each index. The Laspeyres index is geometric. “CUPI, 25p” and “CUPI, 50p” use 25th-percentile and 50th-percentile quantity shares, respectively. “CUPI, 10p (5q)” uses a 10th-percentile quantity share threshold computed over quarters t and $t-4$.

Appendix

A Hedonic Estimation Details: Additional Results with NPD Data

We estimate hedonic imputation models in both log-levels and log-differences as well as using the time dummy method. For the hedonic imputation models, we also consider alternative weighting approaches.

Figure D.2 presents results comparing the log-level relative to log-difference (EP-F and EP-TV) specifications. The log-level specification yields more erratic patterns than the log-difference specifications. We use quantity weights in the results presented in Figure D.2. For single-period log-level estimation, we use contemporaneous quantity shares. Intuitively, this specification only uses information from the current period to produce hedonic estimates. For estimation of the specifications proposed by Erickson and Pakes (2011), in which the dependent variable is the change in log prices, we use weights that are the average of the quantity shares in the previous and current periods. The results using the EP methods presented in the main text take the same approach.

The log-level specifications are sensitive to omitted unobservable characteristics. To illustrate this point, Figure D.3 presents a version of Figure 4 that shows the sensitivity of the levels specification to intentionally omitted key observable characteristics. Unlike the EP-TV approach, the log-levels specification is very sensitive to omitting these observable characteristics.

For the time dummy method, we specify the hedonic regression equation (10) using the same vector of characteristics Z_k in each pair of adjacent periods. Occasionally, new features are introduced to the data. In pairs of adjacent periods entirely prior to the introduction of a new characteristic, it will be omitted from the regression because of collinearity with the intercept term. In pairs of adjacent periods in which the new feature is absent during period t-1 and present during period t, the feature will be included in the estimated regression. Symmetric arguments apply for characteristics that exit.

Intuitively, the period-t fixed effect δ_t reflects the difference in average price of a "generic" good between $t-1$ and t because the contributions of all of the product characteristics have been partialled out. The hedonic time dummy specification includes goods entering in period t and exiting after period t-1 through its use of the Tornqvist weights, which are average market shares between the two periods. Nonetheless, a limitation of the time dummy method relative to the EP-TV approach is that the former does not account for unobservable product characteristics. Another issue emphasized by Pakes (2003) and Diewert et al. (2008) is that this method imposes constant coefficients on characteristics in adjacent periods, a restriction that is often rejected by the data.

Figure D.4 compares the time-dummy price indices to EP-F, EP-TV and matched model Tornqvist price indices. The values displayed in the figure are annual percent changes in the 4th quarter of each year from chained cumulative quarterly indices. The various price indices track each other broadly, but they also display some systematic differences. For all product groups, the EP-TV approach yields a lower rate of price inflation compared to the matched-model Tornqvist, the time dummy based index, or the first-difference based index.

The time dummy method suggests a notable hedonic adjustment for coffee makers relative to the matched-model Tornqvist index, but for other products the difference is modest or is positive rather than negative. Our finding of limited quality adjustment using the time dummy method is broadly consistent with the discussion in Erickson and Pakes (2011). As they emphasize, traditional hedonic approaches cannot account for the changing valuations of unobservable product characteristics, and in particular, how those changing valuations interact with product turnover. For example, if entering goods have desirable unobserved characteristics and correspondingly high prices, then the time dummy method may erroneously suggest a higher index value relative to the matched-model Tornqvist.[35]

The results presented thus far use quantity-share weights in our hedonic specifications following Bajari et al. (2021). As noted in the text, the motivation for using quantity weights in estimation using unit prices is consistent with Broda and Weinstein (2010) and Redding and Weinstein (2020). The argument is that unit prices based on a large number of purchases are better measured than those based on a small number of purchases.

Diewert (2019) favors using expenditure-weights in hedonic estimation based on the argument that this provides more weight on items with more economic importance. He also notes that this approach facilitates comparisons of the time dummy and full imputation hedonic approaches. He acknowledges, however, that his preference for expenditure weighting is based more on index-number issues than econometric issues (footnote 14, p. 6 of Diewert, 2019).[36] We report sensitivity of results using expenditure-weights in the regression-based EP-TV results for NPD in Table D.3.

We present a large number of statistics including comparisons with related demand-based approaches to quality adjustment in Table D.3. Given this, we focus on the results that incorporate chain drift using the GEKS-Lite procedure of Section 4.3. Such results reflect the many different issues including chain drift relevant for comparing alternatives. We report means and standard deviations of annual chained price indices as well as correlations. We included for comparison purposes the matched-model Tornqvist and Sato-Vartia indices, the Feenstra index, and the hedonic Tornqvist using the EP-TV approach estimated using both quantity weights and expenditure weights.

We find that using the expenditure-weighted and quantity-weighted approaches yield broadly similar results. For four of the five product groups, the quantity-weighted and expenditure-weighted EP-TV price indices yield lower rates of annual price inflation than the matched-model Tornqvist. The exception is headphones, for which the expenditure-weighted EP-TV mean is slightly above the matched-model Tornqvist. We also report and compare the hedonic Tornqvist indices' differences versus the matched-model indices with the differences between the Sato-Vartia and Feenstra indices. For the latter, all five product

[35]For headphones, the matched-model Tornqvist is notably lower in 2016 compared to the Hedonic Tornqvist using the time dummy method. This is a year when the share-weighted average price per item increases substantially. This pattern is consistent with entering goods having higher prices than existing goods. The time dummy method still yields a negative price change in that year, but not as negative as the standard Tornqvist. The hedonic Tornqvist EP-TV method yields a more negative price index than the standard Tornqvist.

[36]Gorajek (2022) highlights that using expenditure weights yields potential issues with consistency of the estimation given that the expenditure shares are direct functions of the dependent variable (prices). He proposes an alternative transformation of the dependent variable to address these issues.

groups exhibit a lower rate of inflation for Feenstra than Sato-Vartia. The patterns of these differences are more similar to the Tornqvist and Hedonic Tornqvist, EP-TV using the quantity weighting. As discussed in the main text, we regard the demand-based quality approaches as a relevant benchmark to compare with the hedonic-based indices.[37]

We also find that the quantity-weighted and expenditure-weighted indices are very highly correlated with each other and have similar variation as measured by the standard deviation. The matched-model Tornqvist tends to have slightly lower correlations with the EP-TV-based indices (especially for expenditure-weighted EP-TV vs. matched-model Tornqvist for Coffeemakers). The Feenstra and EP-TV-based indices have high correlations with the exception of coffeemakers, for which the correlation is especially low using expenditure-weights.

We report goodness of fit statistics for alternative specifications in Table D.4. As expected, the log-level estimation models account for a large share of variation in product price levels, as measured by R^2. This high explanatory power reflects the fraction of the cross-sectional variation in prices accounted for by the observable characteristics. Those same models account for a small fraction of the variation in price relatives as can be seen in the second column. The EP-F and EP-TV methods yield much higher R^2 values for the price relatives. The expenditure-weighted and quantity-weighted specifications yield similar R^2 values, although for occupational footwear and boys' jeans they are lower using expenditure-weighted estimation.

B Using the NielsenIQ Data

B.1 NielsenIQ Data Preparation

The NielsenIQ RMS data consists of more than 100 billion unique observations at the week-store-UPC level. We first aggregate the weekly data to the monthly frequency according to the NRF calendar and then aggregate the monthly data to quarterly. Following procedures used by Hottman et al. (2016) and Redding and Weinstein (2020), we drop outliers from the monthly data before aggregating to the data to quarterly frequency. Specifically, we drop observations with prices above 3 times or below one-third the module-level median for each UPC in a given month. We also drop product-month observations with quantities sold that are more than 24 times that product's median quantity sold per month. One feature of barcoded products is that goods of different sizes and packaging have different barcodes, even if the product contained in the packaging is the same. To ensure comparability between prices, we follow Hottman et al. (2016) and normalize UPC prices to the same units (e.g., ounces), utilizing the size and packaging information provided by NielsenIQ. Consistent with the literature, we winsorize monthly price changes at the top and bottom 1% of each product group. Tables D.1 and D.2 show our classification of the product groups in the NielsenIQ Retail Scanner data into Food and Nonfood categories, respectively. Our normalization of

[37] One note of caution about making such comparisons is that in principle, the Feenstra (1994) adjustment to the Sato-Vartia index captures pure love of variety effects in addition to quality improvement via entry and exit. The hedonic indices do not feature a direct role for gains from increasing product variety, although if increasing variety brings about lower prices or higher quality, it will affect the hedonic price indices indirectly.

units carries over to our measurement of quantities (e.g., all quantities within a product module are measured in consistent units such as ounces). In this appendix, we also use the NielsenIQ Consumer Panel and apply the same procedures from Hottman et al. (2016) and Redding and Weinstein (2020) to prepare the data. In the next subsection we discuss differences between the Scanner and Consumer Panel data.

B.2 Retail Scanner versus Consumer Panel

The NielsenIQ data are also available in a Consumer Panel, where a sample of households scan their purchases for the same consumer package product groups covered by the Scanner data. When the Consumer Panel is used for price index research (e.g., Redding and Weinstein, 2020), the Consumer Panel observations are typically restricted to include only purchases of goods that have UPC codes. There are potential challenges to tracking economywide patterns in both the Retail Scanner and Consumer Panel databases. The Retail Scanner data are a census of all transactions at outlets covered by NielsenIQ. However, the Scanner data has much better coverage of grocery stores than mass merchandisers including general merchandise stores. As we document below, this results in significant limitations in using the Scanner data for nonfood items.

The Consumer Panel has distinct limitations. First, it is a challenge for participants to accurately record item-level prices from their transactions – the protocol NielsenIQ uses is to use the item-level price from the Scanner data for all purchases from participating retail partners in the Scanner data. Second, the Consumer Panel sample is highly nonrepresentative prior to the application of sample weights: it is dominated by individuals that are older than 55 and are White, Non-Hispanic. For example, in the 2016 Consumer Panel the share of panelists that are 55+ is 56% and are White, Non-Hispanic is 81%. In comparison, Census population statistics for that year show the share of household heads that are 55+ is 42% and are White, Non Hispanic is 72%. While the projection factors make adjustments to make the weighted sample statistics representative, the projection factors are highly right skewed with a skewness statistic of 1.55. To highlight the potential issues here, the CPS monthly sample (in March 2016) even in raw form is much more representative with the share of household heads that are 55+ at 42% and are White, Non-Hispanic at 71%. Moreover, the CPS sample weights (using household weights) used to make the CPS representative exhibit essentially no skewness: skewness statistic is -0.05. With the highly skewed projection factors in the Consumer Panel, a small number of observations can have very large influence.

The validation study by Einav et al. (2010) highlights related issues. They find that a significant (about 10%) of transactions are not reported and the extent of non-reporting varies systematically across demographic groups. The projection factors do not take this type of non-reporting into account. In addition, Einav et al. (2010) find significant misreporting of prices and quantities.

In order to compare the relative merits of the Scanner and Consumer Panel data, we build on the analysis in the main text comparing the aggregate properties of both to official statistics in the next section. Again, we do this comparison distinctly for food and nonfood.

B.3 Comparisons of the NielsenIQ Data to Official Statistics

This section provides a comparison of both the NielsenIQ Scanner and Consumer Panel with official statistics. Figure D.5 presents comparisons of nominal expenditures for the food and nonfood categories of the NielsenIQ data compared to the Bureau of Economic Analysis (BEA) personal consumption expenditure (PCE) data.[38] As we reported in the main text, for food, we find nominal sales indices for the NielsenIQ Scanner data tracks the PCE quite closely. The NielsenIQ Consumer Panel tracks the PCE reasonably well through 2012, but it rises less rapidly than either the NielsenIQ Scanner or PCE thereafter. The poorer performance of the Consumer Panel likely reflects the limitations discussed above.

For nonfood in the bottom panel, both the Scanner and Consumer Panel exhibit less of an increase over time than the PCE. As we discussed in the main text, this finding for the Scanner data is not surprising. The Scanner data tracking of stores that sell the nonfood items tracked by NielenIQ has weaker coverage. For the Consumer Panel, the limitations discussed above are again presumably at work.

Figure D.6 presents the relationship between the BLS CPI and corresponding Laspeyres indices from the NielsenIQ Scanner and Consumer Panel data sets.[39] We show both arithmetic and geometric Laspeyres. The CPI is a two-stage index with a geometric unweighted index at the MSA level and arithmetic Laspeyres to the national level. For food, both the NielsenIQ Scanner and Consumer Panel Laspeyres indices are highly correlated with the CPI. In terms of inflation levels, however, the NielsenIQ Scanner more closely matches the CPI. The BLS CPI for Food averages 2.4% over our sample period, the arithmetic Laspeyres for the Retail Scanner averages 2.9% and the arithmetic Laspeyres for the Consumer Panel averages 4.2%. The correlations between Laspeyres indices for the nonfood product groups and the CPI are much weaker than for food (0.53 and 0.67 for the Scanner and Consumer Panel data sets, respectively, using the arithmetic Laspeyres). The average inflation level is closer to the CPI in the Scanner data than in the Consumer Panel.

The results of our comparisons with official statistics show that the NielsenIQ Scanner data matches expenditure patterns and price indices with official statistics well for food at the aggregate and disaggregate levels. For non-food, the NielsenIQ Scanner displays discrepancies with the official statistics at both the aggregate and disaggregate levels. The Consumer Panel exhibits substantially larger differences for expenditure and price patterns for food at the aggregate and disaggregate levels and performs equally poorly in matching official statistics for non-food.

B.4 Common Goods Rules – Consumer Panel and Retail Scanner

This section presents sensitivity results to alternative common good rule approaches for both the NielsenIQ Scanner and NielsenIQ Consumer Panel data sets. We consider alternative specifications of the CGR using market thresholds using percentiles of sales pooled over a two-quarter and a five-quarter horizon (that is, the current and prior 4 quarters). For the latter, a common good is defined in this context if it is present in periods t and $t-4$. Using

[38] PCE data used in main text and in this appendix downloaded from BEA in September 2020.

[39] We thank the BLS for providing us with CPI for Food and NonFood using product group categories consistent with the NielsenIQ data.

a duration component in the CGR puts more weight on goods present for the longer horizon, yielding greater comparability with the duration-based CGR used by Redding and Weinstein (2020).

To start, we use the 2-quarter horizon. Figure D.7 shows the results for the aggregated food categories of the CUPI and its components using various CGRs defined by different sales-based percentiles using the Scanner data. The Feenstra adjustment and S^* ratio terms have again been scaled by $\frac{1}{\hat{\sigma}-1}$ for each constituent product group so that the components sum to the CUPI. The Feenstra adjustment (Lambda ratio) and Jevons index (P^* ratio) components of the CUPI show very little sensitivity to the alternative CGRs. Indeed, the plots for the different values are nearly indistinguishable. In contrast, the S^* ratio is very sensitive to the CGR in the NielsenIQ data, which leads directly to sensitivity in the CUPI. The baseline CUPI without a CGR percentile threshold has average four-quarter price inflation about 10 percentage points below the Feenstra. Using a 50th percentile for the CGR yields a price index that is much closer to the Feenstra index.

Again using the Scanner data, Figure D.11 compares the results of imposing common goods rules using the 2-quarter horizon vs. a 5-quarter horizon (i.e., computing percentiles for sales pooled over quarters t and $t-4$). These alternative CGRs impose different duration-based restrictions on products to be included in the set of common goods. The figure shows that the 5-quarter CGR using a 10th-percentile share threshold leads the CUPI to measure inflation between what is measured using 25th and 50th percentile thresholds using the 2-quarter horizon. The longer-horizon CGR puts additional weight on the goods that have been present in the marketplace for a longer time, which moves our approach in the direction of the duration-based CGR approach of Redding and Weinstein (2020).

Figure D.12 shows the sensitivity of the CUPI to different CGRs using the NielsenIQ Consumer Panel for food. Here, we focus on 5-quarter horizon CGRs. While the results differ quantitatively, the same general pattern holds as in the NielsenIQ Scanner data, with the CUPI increasing in the percentile of the CGR.

To facilitate comparison of our results to Redding and Weinstein (2020), who report pooled results for food and nonfood product groups, Figure D.13 shows various price indices calculated using all product groups in the NielsenIQ Consumer Panel data. The results are broadly consistent with Redding and Weinstein (2020). However, importantly our analysis focuses on chained quarterly annual indices while Redding and Weinstein (2020) focus on year-over-year indices for fourth quarters of each year. In Figure D.14, we show we can closely mimic their results for the CUPI using a market share common goods rule at the 5th percentile if we calculate a year over year (YOY) price index instead of the chained quarterly price indices that have been the focus of this paper. As we have noted in the preceding discussion, the use of a YOY index imparts a duration-based component to the CGR in addition to the expenditure share-based thresholds.[40]

Figure D.15 shows related indices, using the NielsenIQ Scanner data, pooling all product

[40] We note that we do not impose a CGR in computing the other price indices shown in Figure D.13. In contrast, Redding and Weinstein (2020) apply the same common goods rule for all of the price indices they display. In unreported analysis, we have found that the Sato-Vartia and Feenstra are not very sensitive to the CGR. This inference is also evident in Figure D.7, which shows that the Lambda ratio is not sensitive to the CGR for NielsenIQ Food data. Because our objective is to compare demand-based indices with the hedonic indices, we aim to treat entry and exit symmetrically across these indices.

groups, and using various CGRs based on sales percentiles computed over the 5-quarter horizon.[41] These results are therefore suggestive of the results that applying the empirical approach in Redding and Weinstein (2020) to the Scanner data would produce. The CUPI with no CGR suggests deflation of 10 percent or more per year. Even the CUPI with a 25th-percentile cutoff rule shows persistent deflation in the Retail Scanner data; imposing a 50th-percentile CGR brings the CUPI closer in line with the Laspeyres index. The series labeled "CUPI, RW CP" shows results from applying the market share threshold in the 5th-percentile CGR from the Consumer Panel to the Scanner data, rather than calculating a percentile-based threshold directly from the Scanner data. Using the Consumer Panel share threshold for the CGR produces results similar to using the 50th-percentile CGR calculated directly in the Scanner data.

The main message from this analysis is that the CUPI is very sensitive to the specification of the CGR, both in the NielsenIQ Consumer Panel and in the NielsenIQ Scanner data, and to the horizon over which the threshold is computed. Using the longer horizon market share threshold moves the CGR towards the Redding and Weinstein (2020) duration-based approach. It is worth reiterating that any duration-based approach has greater data requirements for practical implementation.

B.5 Machine Learning and Hedonics for Non-Encoded Product Descriptions

As we discuss in Section 3.2.1, NielsenIQ product descriptions are not encoded as variables that can readily be incorporated into econometric models. Instead they are strings which, though they are human interpretable, would be extremely tedious to encode. (See examples in the text.) Encoding such strings is a standard application of machine learning (ML).

In Cafarella, Ehrlich, Gao, Haltiwanger, Shapiro, and Zhao (2023), we describe in detail the procedure we use for encoding the NielsenIQ product descriptions and for implementing a neural net machine learning approach that parallels the EP econometric model. In brief, the steps in our ML algorithm are as follows: First, to convert text-based product descriptions into numerical characteristic representations, we use a hybrid feature encoding architecture that allows the system to incorporate "pre-trained" word embeddings (numerical representations) trained from an external corpus of text as well as specifically trained or "text-tailored" embeddings trained specifically on the product descriptions in the NielsenIQ Kilts Retail Scanner Data set. Second, our architecture does not predict prices or price changes directly, but rather predicts a set of probabilities that the price or price change lies in each of a set of price or price-change bins that partition the observed range. Third, the ML system minimizes the weighted cross-entropy loss function for the products' true price and price change distributions in the hedonic estimation.[42] Fourth, because of the noise in the estimated probabilities, it may not be optimal to calculate price predictions as the simple probability-weighted expected price. We use a receiver operating characteristic (ROC) curve procedure to determine the optimal number of bins to include in the price prediction.

[41]Figure D.15 also displays the Bureau of Labor Statistics' Consumer Price Index for all of the product groups included in the NielsenIQ data as a point of reference.

[42]In this application, the cross-entropy loss objective function is equivalent to maximizing the likelihood of assigning the highest probability to the correct bin.

We assess the ML procedure's performance as measured by the prediction "near accuracy" across every product group-quarter. We define the model's near accuracy as the proportion of products for which it assigns the highest probability to the correct or an adjacent bin. The median in-sample near accuracy for food price change bins is well above 80%. The out-of-sample near accuracy for the median product group-quarter is nearly 60% for the food product groups. In other words, the median-performing model predicts the the correct bin or an adjacent bin more than 80% of the time. We view these model performances as remarkable: in the median product group-quarter, the system is able to closely predict a product's price change over half the time based on the short, nonstandard product descriptions. In the main text, we report related in sample and out of sample R^2 statistics.

B.6 Assessing the National Market Assumption in the NielsenIQ Data

Our empirical implementation of the CUPI relies on the assumption of a unified national market. Redding and Weinstein (2016) note that "Our CES price index is not superlative, because it does not approximate any continuous and differentiable utility function." Although Redding and Weinstein (2024) show how to aggregate the CUPI in a theoretically consistent manner, the fact that it is not superlative means we cannot rely on the standard "approximate consistency in aggregation" property of superlative indices (Diewert, 1978) in its empirical implementation. Deviations from the aggregation structure assumed in the model can have potentially important effects on the measured index. Relatedly, because the CUPI includes unweighted geometric mean terms, any failure of this national market assumption has the potential to affect the CUPI more than other indices. In this section, we assess the realism of the national market assumption in the NielsenIQ Scanner data.

In Figure D.10, we pool the NielsenIQ item-level data for food product groups at the weekly frequency from 2006–15. We then compute the market penetration of items in the pooled data both on an unweighted basis (i.e., all items get the same weight) and on a sales-weighted basis. Market penetration is defined as the share of NielsenIQ metro areas in which the item-level week is observed to have positive sales.

On an unweighted basis, the distribution is very skewed to the left, with most item-level week observations having very low market penetration. Almost all of the unweighted distribution has less than a 20 percent market penetration. In unreported results, we find that the mass of the unweighted distribution with the lowest market penetration reflects entering and exiting goods. Even on a sales-weighted basis, only 15 percent of sales are for items with a truly national market, although much of the mass of the distribution has market penetration of over 80 percent of metro areas. These patterns raise questions about applying a national market based CES price index for most items. In Appendix C.2, we show that the CUPI can be badly biased when the national market assumption fails, suggesting these patterns may be important for understanding the empirical behavior of the CUPI.

C Additional Evidence on the Behavior of the CUPI

In this appendix, we examine the properties of the CUPI in more detail. First, in section C.1, we consider nested versions of the CUPI. Second, in section C.2 we consider the properties of taste shock bias further using numerical simulations. Our objectives to provide more guidance for why the CUPI is so much lower than the Sato-Vartia or Feenstra price indexes. This analysis provides useful context as we then explore market frictions potentially associated with the entry and exit dynamics of products to help provide further perspective on the need for a CGR. In section C.3, we offer an analytical characterization of the taste shock bias in a simple setting.

C.1 Behavior of the CUPI with a Nested Preference Structure

We begin by exploring the issue of nesting in the CUPI using two methods that rely on the product attributes in the data to define a nested product substitution structure. First, we define nests within product groups with a heuristic-based approach. Using this approach, we assign products to subgroups based on a set of key variables that we as analysts hypothesize define market strata. Because this procedure is labor-intensive and relies on our subjective judgments regarding strata, we also construct alternative subgroups by allocating products to groups based on the deciles of their predicted price from a log-level hedonic model. Intuitively, in the first approach, we implicitly assume that substitutability is constant within market strata (for example, drip coffee makers versus espresso machines), while in the second approach we assume that price tiers (for example, low-end versus high-end coffee makers) define the substitution structure.

The nested approach requires estimation of elasticities of substitution for products within the same nest and across nests. We follow the approach of Hottman et al. (2016) to estimate within- and between-nest elasticities for each product group. The within-group estimation uses a modified Feenstra (1994) estimator that double-differences market shares and prices with respect to time and a time-varying nest-level mean.[43] The between-nest estimator of the elasticity of substitution uses an instrumental variable (IV) approach building on Hottman et al. (2016).[44]

Table D.9 reports the estimated elasticities for the nested specifications. The results are broadly similar across the two nested approaches. As expected, the within-nest elasticities are estimated to be larger than the between-nest elasticities.

[43]The identifying assumption of the Feenstra (1994) estimator is that supply and demand shocks are orthogonal when sales growth and price growth are differenced with respect to a time-varying mean. The Hottman et al. (2016) assumption is arguably more natural, as differencing with respect to a within-nest mean more plausibly identifies orthogonal supply and demand shocks.

[44]We follow Hottman et al. (2016) by specifying the between-group relationship between the nest-level price index and expenditure share. The former is endogenous, and Hottman et al. (2016) overcome this by using variation in the nest-level price index caused by changes in within-nest expenditure share dispersion. We innovate on the procedure of Hottman et al. (2016) by using the S^* ratio (i.e., changes in common goods expenditure share dispersion) from the within-nest CUPI as the instrument, which removes changes in expenditure-share dispersion induced by product turnover. The identifying assumption is that within-nest demand shocks are uncorrelated with between-nest demand shocks. This innovation integrates the insights of Hottman et al. (2016) with those of Redding and Weinstein (2020).

In principle, these within-nest vs. between-nest elasticity estimates could produce significantly different results for the Feenstra index and the CUPI, but in our application the differences are modest. Figure D.8 plots nested versions of the CUPI using our two nesting strategies alongside un-nested versions of the CUPI and Feenstra index. Both versions of the CUPI are implemented using a 30th-percentile CGR, applied at the within-nest level in the nested version.[45] The alternative nesting approaches yield similar results, with the nested CUPI tending to show slightly less deflation than the un-nested (or "flat") CUPI. In unreported results, we find that the relationship between the nested and flat CUPIs is robust to using alternative CGR cutoffs.

We conclude that that the CUPI's large negative inflation readings are not meaningfully mitigated by considering a nested preference structure.

C.2 Simulation Evidence on the Behavior of the Sato-Vartia Index and CUPI

In this section, we present numerical simulation evidence on the taste shock bias highlighted by Redding and Weinstein (2020). We build on the core insight of Redding and Weinstein (2020) that the sign of the taste shock bias will depend on the sign of the correlation between Sato-Vartia expenditure weights and taste shocks. We start with some simple examples that illustrate cases with positive, negative and no bias. In all of these examples and simulations the CUPI is the exact price index – with no need for a CGR. This analysis provides useful context for exploring frictions that may underlie the need for a CGR.

C.2.1 Conditional (Non-Stochastic) Examples

We begin by presenting three non-stochastic examples of the Sato-Vartia index's behavior in the presence of shocks to product appeal. In each of these examples, we consider a representative consumer with an elasticity of substitution $\sigma = 5$ across products. The examples feature two products, A and B, which are both sold in each of two periods, 1 and 2. Both products have prices equal to \$1.00 in both periods. The patterns of the product appeal parameters differ between the three examples.

Example 1: In period 1, both products have appeal parameters φ_{kt}=1. In period 2, product A's appeal parameter doubles, while product B's appeal parameter falls in half. Note that the changes in the appeal parameters obey the normalization that log appeal changes are mean zero. Exhibit C.1 displays the products' prices and appeal parameters in each of the two periods, along with the resulting behavior of the Sato-Vartia index and unit expenditure function. Because prices never change in this example, it is straightforward to see that the Sato-Vartia index will measure zero inflation. The log change in the unit expenditure function is -0.52.

The consumer's true cost of living declines from period 1 to period 2 because the dispersion of the product appeal parameters rises. The Sato-Vartia index does not measure this change in the cost of living correctly, because it does not capture changes in the distribution

[45] Nests are weighted by the number of products to ensure differential product group sizes. This ensures that the effective weight assigned to each good is equal to that in the non-nested CUPI, exclusive of the effect of changes to the sample through the different common goods rules.

Exhibit C.1: Taste-Shock Bias, Non-Stochastic Example 1

	Period 1		Period 2	
	Product A	Product B	Product A	Product B
Price p_{kt}	\$1.00	\$1.00	\$1.00	\$1.00
Appeal Parameter φ_{kt}	1.0	1.0	2.0	0.5
Expenditure Share s_{kt}	0.5	0.5	0.996	0.004
Sato-Vartia Weight $\omega_{k,t\text{-}1,t}$	–	–	0.88	0.12
$\Delta \ln p_{kt}$	–	–	0	0
$\Delta \ln \Phi_t^{\mathrm{SV}}$	0.0			
$\Delta \ln \left(\frac{\tilde{S}_t^*}{\tilde{S}_{t-1}^*}\right)^{\frac{1}{\sigma-1}}$	-0.52			
Corr. $\left(\omega_k, \ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right)\right)$	1.0			
$\Delta \ln P_t^{\text{Unit Exp. Fn.}}$	-0.52			
Taste Shock Bias	0.52			

of product appeal. The Sato-Vartia index's upward bias in this example is consistent with the argument in Redding and Weinstein (2020) that the Sato-Vartia index will tend to be biased upward in the presence of changing product appeal. The CUPI, on the other hand, would capture the decrease in cost of living in this example correctly through the change in the S^* ratio (assuming the correct elasticity of substitution were used in the calculation). We note that this illustration is consistent with the condition elucidated in Redding and Weinstein (2020) that the sign of the taste-shock bias will be the same as the sign of the correlation between the Sato-Vartia weights and the log appeal shocks. The Sato-Vartia weights are perfectly positively correlated with the log appeal shocks, and taste-shock bias is positive.

Example 2: We now consider a second example. In this example, product A has appeal parameter 2.0 in period one, and product B has appeal parameter 0.5. In period two, product A's appeal parameter falls in half, while product B's appeal parameter doubles, so that both products have appeal parameters $\varphi_{k2} = 1$. We note that the changes in the appeal parameters again obey the normalization underlying the CUPI that log appeal changes are mean zero. Exhibit C.2 displays the results of this thought experiment.

It is again straightforward to see that the Sato-Vartia index will measure zero inflation. The log change in the unit expenditure function is +0.52, so the taste-shock bias is precisely the opposite of the bias in the previous example. The unit expenditure function shows that the consumer's cost of living increased from period 1 to period 2 because the dispersion of the product appeal parameters declined. The CUPI would again measure the change in the consumer's cost of living correctly through the change in the S^* ratio. In this example, the taste-shock bias is negative, reflecting the negative correlation between the Sato-Vartia weights and the log appeal shocks.

Example 3: In our third example, product A again has appeal parameter 2.0 in period

Exhibit C.2: Taste Shock Bias, Non-Stochastic Example 2

	Period 1		Period 2	
	Product A	Product B	Product A	Product B
Price p_{kt}	\$1.00	\$1.00	\$1.00	\$1.00
Appeal Parameter φ_{kt}	2.0	0.5	1.0	1.0
Expenditure Share s_{kt}	0.996	0.004	0.5	0.5
Sato-Vartia Weight $\omega_{k,t\text{-}1,t}$	–	–	0.88	0.12
$\Delta \ln p_{kt}$	–	–	0	0
$\Delta \ln \Phi_t^{\mathrm{SV}}$		0.0		
$\Delta \ln \left(\frac{\tilde{S}_t^*}{\tilde{S}_{t-1}^*}\right)^{\frac{1}{\sigma-1}}$		0.52		
Corr. $\left(\omega_k, \ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right)\right)$		-1.0		
$\Delta \ln P_t^{\text{Unit Exp. Fn.}}$		0.52		
Taste Shock Bias		-0.52		

one, and product B has appeal parameter 0.5. In period two, the appeal parameters switch places, so that product A has appeal parameter 0.5 and product B has appeal parameter 2.0. In this example, the Sato-Vartia index correctly measures that the consumer's true cost of living does not change, despite the changing product appeal parameters. In this example, the dispersion of the product appeal parameters did not change across periods. Moreover, the taste-shock bias is zero consistent with the zero correlation between the Sato-Vartia weights and the log appeal shocks.

We believe these simple examples highlight a few key points. First, the taste-shock bias in the Sato-Vartia index can be positive, negative, or zero, even conditional on the first period's demand parameters being held fixed. Second, the taste-shock bias takes the same sign as the correlation between the Sato-Vartia weights and the log appeal shocks, consistent with the logic in Redding and Weinstein (2020). Third, the taste-shock bias is symmetrical, in the sense that reversing a given change in the appeal parameters generates precisely the opposite bias.

This result carries through to our first two stochastic simulations in the next section, in which the unconditional expectation of the taste-shock bias is zero. We offer an analytical examination of conditions under which the unconditional expectation of the taste-shock bias will be zero in Appendix C.3.

C.2.2 Stochastic Simulations with Zero Expected Taste-Shock Bias

In this section we present simulation evidence from simple, frictionless settings in which prices and product appeal are drawn from stochastic distributions but the taste-shock bias has an unconditional expected value of zero. Specifically, we focus on cases where the correlation between Sato-Vartia weights and product appeal shocks is (at least approximately) zero.

We note that the realized value of the taste-shock bias will generally be non-zero in these

Exhibit C.3: Taste-Shock Bias, Non-Stochastic Example 3

	Period 1		Period 2	
	Product A	Product B	Product A	Product B
Price p_{kt}	\$1.00	\$1.00	\$1.00	\$1.00
Appeal Parameter φ_{kt}	2.0	0.5	0.5	2.0
Expenditure Share s_{kt}	0.996	0.004	0.004	0.996
Sato-Vartia Weight $\omega_{k,t\text{-}1,t}$	–	–	0.5	0.5
$\Delta \ln p_{kt}$	–	–	0	0
$\Delta \ln \Phi_t^{\mathrm{SV}}$		0.0		
$\Delta \ln \left(\frac{\tilde{S}_t^*}{\tilde{S}_{t-1}^*}\right)^{\frac{1}{\sigma-1}}$		0.0		
Corr. $\left(\omega_k, \ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right)\right)$		0.0		
$\Delta \ln P_t^{\text{Unit Exp. Fn.}}$		0.0		
Taste Shock Bias		0.0		

simulations due to the appeal shocks; the CUPI's ability to capture these shocks' effects on the unit expenditure function is an important strength even when the expected value of the taste-shock bias is zero. Put differently, the CUPI will capture volatility in the unit expenditure function that is not captured by the Sato-Vartia price index.

In these simulations, as in all of the stochastic simulations we consider, the representative consumer is assumed to have CES preferences with elasticity of substitution $\sigma = 5$, and the correct value of σ is used in the calculation of the CUPI.

C.2.2.1 Baseline Simulations with Different Trend Inflation Rates Figure D.16 explores the behavior of the Sato-Vartia index and CUPI when log prices and appeal parameters are are independently and identically distributed (i.i.d.) random variables with a stationary distribution over time. The data generating process for prices and appeal parameters is

$$\ln p_{kt} = \mu t + \varepsilon_{kt} \tag{C.1}$$

$$\ln \varphi_{kt} = \nu_{kt} \tag{C.2}$$

$$\begin{bmatrix} \varepsilon_{kt} \\ \nu_{kt} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \chi_p^2 & \rho_{p\varphi} \\ \rho_{p\varphi} & \chi_\varphi^2 \end{bmatrix}\right). \tag{C.3}$$

We consider different rates of trend inflation μ and different degrees of correlation between price and appeal draws $\rho_{p\varphi}$ along the x-axis. We set $\chi_p = \chi_\varphi = 0.2$ across all of the simulations, which feature 1,000 individual products simulated for 12 time periods. We run 50 simulations for each combination of parameters μ and $\rho_{p\varphi}$.

The Sato-Vartia index does not display a meaningful average taste shock bias under these conditions, regardless of the trend inflation rate μ or the correlation between log prices and appeal parameters $\rho_{p\varphi}$.

Figure D.17 illustrates the mechanics of the taste-shock by plotting the average correlations between the Sato-Vartia weights and the log appeal shocks in each simulation. Redding and Weinstein (2020) show that the taste-shock bias will have the same sign as this correlation, and will equal zero when this correlation is zero. The correlation is centered on and generally near zero in these simulations, suggesting that time variation in product appeal does not, on its own, introduce an unconditional expected taste-shock bias to the Sato-Vartia index.

Again, we emphasize that an unconditional expected taste-shock bias of zero does not mean the CUPI has no advantages relative to the Sato-Vartia index. In this frictionless environment, the CUPI with no CGR exactly replicates the unit expenditure function when it is calculated using the correct value of σ. The Sato-Vartia index, on the other hand, will almost surely display a taste-shock bias given the realizations of product prices and appeal parameters.

C.2.2.2 Baseline Simulations with AR(1) Prices and Appeal Figure D.18 shows that the same pattern for the taste-shock bias extends to the case in which the log prices and log appeal parameters follow $AR(1)$ processes. Consumers again have CES preferences with elasticity of substitution $\sigma = 5$. The data generating process for prices and appeal parameters is now given by

$$\ln p_{kt} = \eta \ln p_{kt\text{-}1} + \varepsilon_{kt} \tag{C.4}$$

$$\ln \varphi_{kt} = \eta \ln \varphi_{kt\text{-}1} + \nu_{kt} \tag{C.5}$$

$$\begin{bmatrix} \varepsilon_{kt} \\ \\ \nu_{kt} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}, \begin{bmatrix} \chi_p^2 & \rho_{p\varphi} \\ \\ \rho_{p\varphi} & \chi_\varphi^2 \end{bmatrix} \right). \tag{C.6}$$

We consider different rates of persistence in the $AR(1)$ processes η and different degrees of correlation between price and appeal draws $\rho_{p\varphi}$ along the x-axis.[46] We set $\chi_p = \chi_\varphi = 0.2$ across all of the simulations, which again feature 1,000 individual products. This set of simulations features 36 time periods, as opposed to the 12 in the baseline simulations, to reduce the volatility introduced by the persistence in the $AR(1)$ processes. We run 50 simulations for each combination of parameters η and $\rho_{p\varphi}$.

Figure D.19 illustrates the correlations between the Sato-Vartia weights and the log appeal shocks in these simulations. As in the previous simulations, they are centered around zero, consistent with the logic in Redding and Weinstein (2020) that a zero correlation between the weights and the appeal shocks implies zero taste-shock bias.

[46] To keep the number of parameters to be displayed manageable, we do not consider non-zero intercepts in the process for log prices, and we restrict the degree of persistence in the processes for prices and appeal parameters to be equal. We draw log prices and appeal parameters from their implied unconditional stationary distribution in the first period of each simulation.

C.2.3 Stochastic Simulations with Nonzero Expected Taste-Shock Bias

In this section we present simulation results from data generating processes in which the Sato-Vartia index displays a nonzero unconditional expected taste-shock bias.

The representative consumer is again assumed to have CES preferences with elasticity of substitution $\sigma = 5$, and the correct value of σ is used in the calculation of the CUPI. As in our previous set of simulations, there are no market frictions that introduce a gap between the CUPI and the true unit expenditure function, so a CGR is unnecessary in these settings.

C.2.3.1 Time-Varying Variance of Product Appeal In this example, we examine simulations in which the variance of the product appeal parameters changes over time. The data generating process is given by

$$\begin{bmatrix} \ln p_{kt} \\ \ln \varphi_{kt} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \chi_p^2 & \rho_{p\varphi} \\ \rho_{p\varphi} & (\chi_{\varphi 0} + \mu t)^2 \end{bmatrix} \right). \tag{C.7}$$

The standard deviation of log appeal draws thus grows or shrinks at rate μ each period.

Figure D.20 shows the results of these simulations, which again use an elasticity of substitution of $\sigma = 5$, with 1,000 products and 12 periods per simulation. We once again conduct 50 simulations per value of the parameters we consider. In this case, we vary only the time trend in the variance of log product appeal μ. The set of columns labeled "Constant Dispersion" uses $\mu = 0$, the set labeled "Falling Dispersion" uses $\mu = -0.01$, and the set labeled "Rising Disperion" uses $\mu = 0.01$. The starting value of the standard deviation of log appeal draws is set to $\chi_{\varphi_0} = 0.2$, as is the (constant) standard deviation of log price draws χ_p.

The results in Figure D.20 reinforce the illustrative non-stochastic examples in section C.2.1: the Sato-Vartia index has an unconditional upward bias when the dispersion of product appeal shocks is rising over time but an unconditional downward bias when the dispersion of product appeal shocks is falling. The CUPI, in contrast, accurately captures these trends' effects on the consumer's cost of living in the absence of other frictions.[47]

Figure D.21 shows that the correlation between the Sato-Vartia weights and the log appeal shocks is centered around zero in the case of constant dispersion in the product appeal parameters, but is systematically negative when the variance of the product appeal parameters is falling over time and systematically positive when the variance of appeal is rising over time. These patterns mimic the patterns of the non-stochastic examples in section C.2.1.

C.2.3.2 Unit Root in Product Appeal This set of simulations explores the taste-shock bias in a baseline setting with no frictions but with a unit root in the data generating process for product appeal parameters. The simulation environment is the same as in Section

[47] In unreported results, we find that a similar pattern holds in the presence of time-varying correlations between the price and appeal draws. The Sato-Vartia index has an upward unconditional bias relative to the unit expenditure function when this correlation is falling over time and a downward unconditional bias when this correlation is rising over time.

C.2.2.1, but the data generating process for price and appeal parameters is now

$$\ln p_{kt} = \varepsilon_{kt} \tag{C.8}$$

$$\ln \varphi_{kt} = \ln \varphi_{kt\text{-}1} + \nu_{kt} \tag{C.9}$$

$$\begin{bmatrix} \varepsilon_{kt} \\ \nu_{kt} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \chi_p^2 & 0 \\ 0 & \chi_\varphi^2 \end{bmatrix}\right). \tag{C.10}$$

Figure D.22 displays the results of these simulations, where we consider different values for χ_φ along the x-axis. The unit root in the data generating process for $\ln \varphi_{kt}$ means that the dispersion of product appeal will rise over time in the simulations, as in the right-hand set of columns in Figure D.20. This rising dispersion produces an upward unconditional taste shock bias in the Sato-Vartia index, which is larger when the variance of the product appeal shocks χ_φ^2 is greater. The theoretical CUPI with no CGR tracks the unit expenditure function. The unit root ensures that the dispersion of product appeal parameters cannot fall over time in this set of simulations.

Figure D.23 displays the correlations of the Sato-Vartia weights and the log appeal shocks in these simulations; they are consistently positive, consistent with the positive taste-shock bias. Figures D.22 and D.23 reinforce the results from the previous simulations and the logic in Redding and Weinstein (2020) that the sign of this correlation will correspond to the direction of the taste-shock bias. They likewise reinforce the argument that the taste-shock bias will be positive when the dispersion of product appeal is rising over time.

C.2.4 Empirical Evidence

We examine the empirical relevance of rising dispersion in product appeal in Figure D.9. The figure plots the measured dispersion (standard deviation) in normalized product appeal for each of the NPD product groups. We find evidence of rising relative product appeal dispersion for memory cards and headphones. This increase is more pronounced without a common goods rule. For other goods, we observe a nonmonotonic change in dispersion even when the S^* ratio is highly negative (see, e.g., boys' jeans in Figure 5). In other words, rising dispersion of product appeal does not obviously drive the CUPI's ubiquitous finding of rapid deflation without using a common goods rule.

We also examine the persistence of the log product appeal in Table D.10. We consider two specifications. The first is a simple AR1 specification using OLS. Results are reported in the first column of D.10. We also consider a specification with product-level fixed effects as the persistence in the first column may not reflect the within product stochastic process but product-level fixed effects. For this second specification we use the Blundell and Bond (1998) methodology. We find that log product appeal shocks are highly persistent. However, some of that apparent high persistence using OLS estimation reflects the presence of product-level fixed effects. Using the more general Blundell and Bond (1998) methodology, the AR1 coefficient estimates range from 0.66 to to 0.77.

C.2.5 Possible Reasons a Common Goods Rule Might Be Needed

In this section, we investigate potential market frictions that could contribute to the CUPI's measurements of rapid deflation and thereby motivate the use of a CGR. We consider market frictions that are likely present for entering and exiting products. For entering goods, we have presented evidence that it takes time for new products to become available on a widespread basis (and some goods remain local). To provide perspective on this type of friction, we consider an environment with segmented markets. This case also captures the more general idea that deviations from the assumption of a frictionless national market in which all goods can be bought in unlimited quantities in any location can have a large effect on the CUPI's inflation measurements. The CUPI is sensitive to these frictions because it includes ratios of unweighted geometric mean prices and expenditure shares. A CGR reduces the influence of small-share goods, which in our simulations are more likely to be affected by the market frictions we consider. Applying a CGR therefore mitigates the large deflation measured by the CUPI in these cases, motivating a CGR's use in empirical practice.

Relately but distinctly, as products exit it is likely they become unavailable in some local markets relative to others. We consider an environment where there are partial stockouts of products prior to exit. A CGR again helps in reducing the CUPI's sensitivity to these frictions.

C.2.5.1 Segmented Markets The first set of frictions we consider is segmentation between markets. Figure D.24 displays results from a set of simulations in which the market is segmented into five distinct submarkets. Consumers have Cobb-Douglas preferences over their composite consumption aggregate from each of the submarkets. Those consumption aggregates are defined by CES preferences over the products consumed in each submarket with appeal shocks subject to the standard normalization. One of the markets is "large," and has a weight of 0.8 in the consumer's aggregate utility function, while the other four markets are "small," and have weights of 0.05 each.

Products enter and exit the market deterministically, with each product being sold for four periods. The data generating process in the large market, which we conceptualize as the "national" market, is given by

$$\ln p_{kt}^{\text{large}} = \varepsilon_{kt}^{\text{large}} \tag{C.11}$$

$$\ln \varphi_{kt}^{\text{large}} = \nu_{kt}^{\text{large}} \tag{C.12}$$

$$\begin{bmatrix} \varepsilon_{kt}^{\text{large}} \\ \nu_{kt}^{\text{large}} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \chi_{p,\text{large}}^2 & 0 \\ 0 & \chi_{\varphi,\text{large}}^2 \end{bmatrix} \right). \tag{C.13}$$

The data generating process in the small markets, which we conceptualize as "local" markets, is given by

$$\ln p_{kt}^{\text{small}} = \varepsilon_{kt}^{\text{small}} \tag{C.14}$$

$$\ln \varphi_{kt}^{\text{small}} = \mu t_{\text{entry}} + \nu_{kt}^{\text{small}} \tag{C.15}$$

$$\begin{bmatrix} \varepsilon_{kt}^{\text{small}} \\ \nu_{kt}^{\text{small}} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \chi_{p,\text{small}}^2 & 0 \\ 0 & \chi_{\varphi,\text{small}}^2 \end{bmatrix} \right). \tag{C.16}$$

The term μt_{entry} gives entering products a trend μ in average appeal; if $\mu > 0$, products in the small market will have rising appeal over time.

The simulations assume that the elasticity of substitution in the representative consumer's lower-level utility aggregator is $\sigma = 5$, and we set $\chi_{p,\text{large}} = \chi_{p,\text{small}} = \chi_{\varphi,\text{large}} = \chi_{\varphi,\text{small}} = 0.2$. Figure D.24 shows results for simulations with different values of μ, the time trend in the average appeal of entering products in the small markets, across the x-axis. The simulations present price indices measured assuming that the econometrician is unaware of the market segmentation and measures prices assuming a unified marketplace.

The assumptions of market segmentation in these simulations are meant to evoke the empirical pattern documented in Figure D.10. which shows that although most sales are concentrated among products sold in nearly all metro areas nationally, on a UPC basis, most products are sold in relatively few areas.

Figure D.24 shows that the CUPI is downward biased when the entering products in the small market have rising appeal over time. The intuition for the bias in the CUPI is that its $P^\star$ and $S^\star$ ratio terms are unweighted geometric means. The theoretical CUPI therefore assigns the price movements in the small markets, driven by product turnover, equal importance to the price movements in the large market. Although that equal weighting scheme would be theoretically justified in a unified marketplace under CES preferences with appeal shocks, it implicitly overweights the small markets in the segmented market environment. A CGR reduces the bias in the theoretical CUPI by reallocating products from the unweighted geometric mean terms to the lambda ratio term, which is weighted.

Figure D.24 thus provides a theoretical justification for the use of a CGR in Redding and Weinstein (2020) and our own empirical work. The key intuition is that the pace and pattern of product upgrading may differ between geographic markets and between national and localized products. This exercise highlights that choosing an appropriate CGR will depend on the pace of product upgrading and degree of market segmentation. In addition, in practice entering goods can transition to becoming national goods, and that process will influence the nature of the CGR. Future research on this topic could help to provide more detailed guidance on the nature of the appropriate CGR in the presence of such frictions.

C.2.5.2 Partial Stock-outs (Rationing) Prior to Exit Figure D.25 examines the behavior of the CES exact price indices when there are partial product stock-outs in the period prior to exit. Consumers have CES preferences over products in a single market with elasticity of substitution $\sigma = 5$. Products enter and exit the market deterministically, with each product being sold for four periods. The data generating process for prices and appeal parameters is given by

$$\ln p_{kt} = \varepsilon_{kt} \tag{C.17}$$

$$\ln \varphi_{kt} = \nu_{kt} \tag{C.18}$$

$$\begin{bmatrix} \varepsilon_{kt} \\ \\ \nu_{kt} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}, \begin{bmatrix} \chi_p^2 & 0 \\ \\ 0 & \chi_\varphi^2 \end{bmatrix} \right). \tag{C.19}$$

The simulations feature a stylized version of stock-outs, or a "clearance rack," in which product sales are rationed in the period before they exit the marketplace. Products' expenditure shares in the period prior to exit are set equal to a given fraction of their level in a frictionless CES demand system. We assume that consumers optimally reallocate their demands toward the unconstrained products in response to the rationing.[48] We set $\chi_p = \chi_\varphi = 0.2$ in these simulations, and consider different "in-stock proportions" for shortly exiting goods along the x-axis of the figure.

The unit expenditure function in Figure D.25 shows that stock-outs do not have a meaningful effect on the consumer's cost of living, because there is no trend in the average appeal of goods over time. Yet stock-outs introduce a substantial bias to both the CUPI and the Feenstra index. The intuition for the bias is subtle. Stock-outs lower expenditure shares on goods just prior to their exit from the marketplace, with the expenditure reallocated to unconstrained goods. Stock-outs therefore raise the dispersion of expenditure shares on continuing goods relative to the frictionless case, leading to a negative log $S^\star$ ratio. The Feenstra adjustment to the Sato-Vartia index is also negative, because new goods enter the market un-rationed, allowing consumers to buy whatever quantities they please; prior to exit, quantities are constrained below consumers' desired levels. The expenditure share on exiting products is therefore lower than the expenditure share on entering products. The Feenstra index and the CUPI are therefore both downward-biased, with the bias being larger for the CUPI.

Imposing a CGR reduces the influence of the stocked-out goods, thereby mitigating the bias introduced by this friction. Future research on this topic providing estimates of the extent and nature of product rationing could help to provide guidance on the empirically appropriate CGR.

[48] The assumption that consumers have homothetic CES preferences makes it straightforward to calculate their re-optimized demands in the presence of rationing; consumers reallocate their expenditure to each of the non-rationed goods in proportion to their unconstrained demands had there been no rationing. The unit expenditure function under rationing can then be computed as the ratio of indirect utilities provided by a unit of expenditure between periods.

C.3 Analytical Characterization of the Taste Shock Bias

In this section, we consider analytically the expected taste shock bias in the Sato-Vartia index under a simple data generating process with no market frictions.

$$\begin{bmatrix} \ln p_{kt} \\ \ln \varphi_{kt} \end{bmatrix} \sim F\left(\begin{bmatrix} \mu_p \\ \mu_\varphi \end{bmatrix}, \begin{bmatrix} \chi_p^2 & \rho_{p\varphi} \\ \rho_{p\varphi} & \chi_\varphi^2 \end{bmatrix} \right) \quad \forall\, t, \tag{C.20}$$

where F is an arbitrary distribution function, μ_p and μ_φ are the means of $\ln p_{kt}$ and $\ln \varphi_{kt}$, and χ_p^2 and χ_φ^2 are their variances, and $\rho_{p\varphi}$ is their covariance, which can be nonzero. We abstract from product entry and exit, so the set and number of goods is fixed over time, and we assume that the number of goods present in the market is large, i.e., $N_{\mathbb{C}_t} \to \infty$. Despite the simplicity of this setting, we believe it helps to illustrate the mechanics of the taste shock bias in a transparent and tractable way.

Denote the joint probability distribution function for (p_{kt}, φ_{kt}) as $f_{P_t \Phi_t}$. Our distributional assumption allows p_{kt} and φ_{kt} to be correlated within each period t but implies that (p_{kt}, φ_{kt}) are distributed identically to and independently from $(p_{kt\text{-}1}, \varphi_{kt\text{-}1})$, i.e.,

$$f_{P_t \Phi_t}(\cdot, \cdot) = f_{P_{t\text{-}1} \Phi_{t\text{-}1}}(\cdot, \cdot). \tag{C.21}$$

Also denote the joint probability distribution function of $(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1})$ as $f_{P_t P_{t\text{-}1} \Phi_t \Phi_{t\text{-}1}}$, and note that because (p_{kt}, φ_{kt}) are distributed independently from $(p_{kt\text{-}1}, \varphi_{kt\text{-}1})$,

$$f_{P_t P_{t\text{-}1} \Phi_t \Phi_{t\text{-}1}}(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1}) = f_{P_t \Phi_t}(p_{kt}, \varphi_{kt}) f_{P_{t\text{-}1} \Phi_{t\text{-}1}}(p_{kt\text{-}1}, \varphi_{kt\text{-}1}). \tag{C.22}$$

Redding and Weinstein (2020) show that the taste shock bias in the Sato-Vartia index can be written as

$$\ln \Phi_t^{\star CCV} - \ln \Phi_t^{\star SV} = \sum_k \omega_{kt} \ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right), \tag{C.23}$$

where ω_{kt} are the Sato-Vartia weights defined by

$$\omega_{kt} = \frac{\frac{s_{kt} - s_{kt\text{-}1}}{\ln(s_{kt}) - \ln(s_{kt\text{-}1})}}{\sum_k \frac{s_{kt} - s_{kt\text{-}1}}{\ln(s_{kt}) - \ln(s_{kt\text{-}1})}},$$

and $\Phi_t^{\star CCV}$ is the CES common varieties index, which abstracts from product entry and exit, as does the Sato-Vartia index.

The unconditional expected value of the taste shock bias is therefore

$$\mathbb{E}\left[\text{Taste Shock Bias}_t\right] = \mathbb{E}\left[\sum_k \omega_{kt} \ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right)\right] = N_{\mathbb{C}_t}\big(\mathbb{E}\left[\omega_{kt} \ln \varphi_{kt}\right] - \mathbb{E}\left[\omega_{kt} \ln \varphi_{kt\text{-}1}\right]\big). \tag{C.24}$$

Equation (C.24) shows that the unconditional expected taste shock bias will be zero when

$\mathbb{E}[\omega_{kt}\ln\varphi_{kt}] = \mathbb{E}[\omega_{kt}\ln\varphi_{kt\text{-}1}]$:

$$\mathbb{E}[\omega_{kt}\ln\varphi_{kt}] = \mathbb{E}[\omega_{kt}\ln\varphi_{kt\text{-}1}] \rightarrow \mathbb{E}[\text{Corr.}(\omega_{kt},\Delta\ln\varphi_{kt})] = 0 \rightarrow \mathbb{E}[\text{Taste Shock Bias}_t] = 0, \tag{C.25}$$

where the middle condition follows from the normalization that $\mathbb{E}[\Delta\ln\varphi_{kt} = 0]$. We will demonstrate that $\mathbb{E}[\omega_{kt}\ln\varphi_{kt}] = \mathbb{E}[\omega_{kt}\ln\varphi_{kt\text{-}1}]$ in the simple setting we consider in this section, thereby demonstrating that the taste-shock bias also has an unconditional expectation of zero. Our logic therefore connects to the condition emphasized in Redding and Weinstein (2020) that the sign of the taste shock bias will be the same as the sign of the correlation between the Sato-Vartia weights and the log appeal shocks, and if that correlation is zero, the taste-shock bias will also be zero.

We can write the Sato-Vartia weights as

$$\omega_{kt} = \frac{\xi_{kt}}{\sum_{l\in\Omega_t}\xi_{lt}}, \quad \text{where} \quad \xi_{kt}(s_{kt}, s_{kt\text{-}1}) = \frac{s_{kt} - s_{kt\text{-}1}}{\ln s_{kt} - \ln s_{kt\text{-}1}}. \tag{C.26}$$

It is straightforward to see that ξ_{kt} is symmetric in its arguments, i.e., $\xi_{kt}(s_{kt}, s_{kt\text{-}1}) = \xi_{kt}(s_{kt\text{-}1}, s_{kt})$. Furthermore, the denominator of ω_{kt}, $\sum_{l\in\Omega_t}\xi_{kt}$, must always equal one by construction. The CES demand system implies that expenditure shares s_{kt} are given as

$$s_{kt} = \frac{\left(\frac{p_{kt}}{\varphi_{kt}}\right)^{1-\sigma}}{\sum_{l\in\Omega_t}\left(\frac{p_{lt}}{\varphi_{lt}}\right)^{1-\sigma}}.$$

Our assumption that $N_{\mathbb{C}_t} \rightarrow \infty$ implies that the proportional change in the denominator, $\sum_{l\in\Omega_t}\left(\frac{p_{lt}}{\varphi_{lt}}\right)^{1-\sigma}$, with respect to any individual good's price or appeal parameter is approximately zero, so that

$$\frac{\partial \frac{1}{\left(\sum_{l\in\Omega_t}\left(\frac{p_{lt}}{\varphi_{lt}}\right)^{1-\sigma}\right)}}{\partial p_{kt}} \rightarrow 0 \quad \text{and} \quad \frac{\partial \frac{1}{\left(\sum_{l\in\Omega_t}\left(\frac{p_{lt}}{\varphi_{lt}}\right)^{1-\sigma}\right)}}{\partial \varphi_{kt}} \rightarrow 0.$$

The symmetry of the logarithmic mean function ξ_{kt} with respect to the expenditure shares combined with this property implies that in the limiting case as $N_{\mathbb{C}_t} \rightarrow \infty$, the Sato-Vartia weights will converge to symmetric functions of (p_{kt}, φ_{kt}) and $(p_{kt\text{-}1}, \varphi_{kt\text{-}1})$, i.e.,

$$\omega(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1}) = \omega(p_{kt\text{-}1}, p_{kt}, \varphi_{kt\text{-}1}, \varphi_{kt}), \tag{C.27}$$

where we have suppressed the kt subscript on ω and the limit notation for simplicity.

We can define $\mathbb{E}\left[\omega\left(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1}\right) \ln\left(\varphi_{kt}\right)\right]$ as

$$\begin{aligned}
&\mathbb{E}\left[\omega\left(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1}\right) \ln\left(\varphi_{kt}\right)\right] \equiv \\
&\iiiint \omega\left(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1}\right) \ln\left(\varphi_{kt}\right) f_{P_t P_{t\text{-}1} \Phi_t \Phi_{t\text{-}1}}\left(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1}\right) dp_{kt} dp_{kt\text{-}1} d\varphi_{kt} d\varphi_{kt\text{-}1} = \\
&\iiiint \omega\left(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1}\right) \ln\left(\varphi_{kt}\right) f_{P_t \Phi_t}(p_{kt}, \varphi_{kt}) f_{P_{t\text{-}1} \Phi_{t\text{-}1}}(p_{kt\text{-}1}, \varphi_{kt\text{-}1}) dp_{kt} dp_{kt\text{-}1} d\varphi_{kt} d\varphi_{kt\text{-}1}.
\end{aligned} \tag{C.28}$$

Likewise, we can write

$$\begin{aligned}
&\mathbb{E}\left[\omega\left(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1}\right) \ln\left(\varphi_{kt\text{-}1}\right)\right] = \\
&\iiiint \omega\left(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1}\right) \ln\left(\varphi_{kt\text{-}1}\right) f_{P_t \Phi_t}(p_{kt}, \varphi_{kt}) f_{P_{t\text{-}1} \Phi_{t\text{-}1}}(p_{kt\text{-}1}, \varphi_{kt\text{-}1}) dp_{kt} dp_{kt\text{-}1} d\varphi_{kt} d\varphi_{kt\text{-}1}.
\end{aligned} \tag{C.29}$$

Note that the only difference between these expressions is that in the former, the Sato-Vartia weights ω are multiplied by $\ln\left(\varphi_{kt}\right)$, and in the latter, they are multiplied by $\ln\left(\varphi_{kt\text{-}1}\right)$. Invoking two results,

- The equality of the joint distributions $f_{P_t \Phi_t}(p, \varphi) = f_{P_{t\text{-}1} \Phi_{t\text{-}1}}(p, \varphi)$, and
- The symmetry of the Sato-Vartia weights $\omega\left(p_{kt}, p_{kt\text{-}1}, \varphi_{kt}, \varphi_{kt\text{-}1}\right) = \omega\left(p_{kt\text{-}1}, p_{kt}, \varphi_{kt\text{-}1}, \varphi_{kt}\right)$,

it is clear that the definitions in equations (C.28) and (C.29) must be equal.[49]

This result is consistent with the logic in Redding and Weinstein (2020) and in our examples and simulations that the sign of the expected taste shock bias is equal to the expected sign of the correlation between the Sato-Vartia weights and the log appeal shocks. Although the taste-shock bias will be zero when the data generating process for log prices and log appeal parameters features a constant joint distribution, factors such as time-variation in the variance of product appeal or a unit root in the product appeal parameters can introduce an unconditional expected taste-shock bias.

[49] A formal derivation of this result, demonstrating the equality step-by-step using the full definitions of the expectations, is available upon request.

D Appendix Tables and Figures

Table D.1: Food Product Groups: NielsenIQ Retail Scanner Data

Product Group	Product Group Code	Product Group	Product Group Code
Baby Food	501	Juice, Drinks - Canned, Bottled	507
Baked Goods-Frozen	2001	Juices, Drinks-Frozen	2006
Baking Mixes	1001	Milk	2506
Baking Supplies	1002	Nuts	1011
Bread And Baked Goods	1501	Packaged Meats-Deli	3002
Breakfast Food	1004	Packaged Milk And Modifiers	1012
Breakfast Foods-Frozen	2002	Pasta	1013
Butter And Margarine	2501	Pickles, Olives, And Relish	1014
Candy	503	Pizza/Snacks/Hors Doeurves-Frzn	2007
Carbonated Beverages	1503	Prepared Food-Dry Mixes	511
Cereal	1005	Prepared Food-Ready-To-Serve	510
Cheese	2502	Prepared Foods-Frozen	2008
Coffee	1006	Pudding, Desserts-Dairy	2507
Condiments, Gravies, And Sauces	1007	Salad Dressings, Mayo, Toppings	1015
Cookies	1505	Seafood - Canned	512
Cot Cheese, Sour Cream, Toppings	2503	Shortening, Oil	1016
Crackers	1506	Snacks	1507
Desserts, Gelatins, Syrup	1008	Snacks, Spreads, Dips-Dairy	2508
Desserts/Fruits/Toppings-Frozen	2003	Soft Drinks-Non-Carbonated	1508
Dough Products	2504	Soup	513
Dressings/Salads/Prep Foods-Deli	3001	Spices, Seasoning, Extracts	1017
Eggs	2505	Sugar, Sweeteners	1018
Flour	1009	Table Syrups, Molasses	1019
Fresh Meat	3501	Tea	1020
Fresh Produce	4001	Unprep Meat/Poultry/Seafood-Frzn	2009
Fruit - Canned	504	Vegetables - Canned	514
Fruit - Dried	1010	Vegetables And Grains - Dried	1021
Gum	505	Vegetables-Frozen	2010
Ice Cream, Novelties	2005	Yogurt	2510
Jams, Jellies, Spreads	506		

Table D.2: Nonfood Product Groups: NielsenIQ Retail Scanner Data

Product Group	Product Group Code	Product Group	Product Group Code
Automotive	5501	Housewares, Appliances	5513
Baby Needs	6001	Ice	2004
Batteries And Flashlights	5502	Insecticds/Pesticds/Rodenticds	5514
Beer	5001	Kitchen Gadgets	5515
Books And Magazines	5503	Laundry Supplies	4506
Canning, Freezing Supplies	5504	Light Bulbs, Electric Goods	5516
Charcoal, Logs, Accessories	5505	Liquor	5002
Cosmetics	6002	Medications/Remedies/Health Aids	6012
Cough And Cold Remedies	6003	Men's Toiletries	6013
Deodorant	6004	Oral Hygiene	6014
Detergents	4501	Paper Products	4507
Diet Aids	6005	Personal Soap And Bath Additives	4508
Disposable Diapers	4502	Pet Care	4509
Electronics, Records, Tapes	5507	Pet Food	508
Ethnic Haba	6006	Photographic Supplies	5517
Feminine Hygiene	6007	Sanitary Protection	6015
First Aid	6008	Sewing Notions	5519
Floral, Gardening	5508	Shaving Needs	6016
Fragrances - Women	6009	Shoe Care	5520
Fresheners And Deodorizers	4503	Skin Care Preparations	6017
Glassware, Tableware	5509	Stationery, School Supplies	5522
Grooming Aids	6010	Tobacco & Accessories	4510
Hair Care	6011	Vitamins	6018
Hardware, Tools	5511	Wine	5003
Household Cleaners	4504	Wrapping Materials And Bags	4511
Household Supplies	4505		

Table D.3: Summary Statistics for Alternative Price Chained Indices, GEKS-Lite and Alternative Estimation Weights, NPD Data

	Memory Cards	Coffeemakers	Headphones	Boys' Jeans	Occupational Footwear
Mean (Tornqvist)	-15.41	-6.64	-11.55	-5.55	-2.31
SD (Tornqvist)	10.85	4.13	4.20	3.40	3.05
Mean (Tornqvist (EP-TV,QW))	-20.33	-9.96	-15.31	-7.76	-3.63
SD (Tornqvist (EP-TV,QW))	7.97	4.47	5.21	3.13	3.54
Mean(Tornqvist (EP-TV,EW))	-16.60	-8.27	-11.24	-6.45	-3.02
SD(Tornqvist (EP-TV,EW))	8.90	4.48	4.86	3.11	3.28
Mean(Sato-Vartia)	-14.32	-6.36	-11.34	-4.13	-2.08
SD(Sato-Vartia)	11.03	4.25	4.15	2.85	3.06
Mean(Feenstra)	-16.46	-9.43	-13.06	-5.51	-3.76
SD(Feenstra)	9.69	4.09	5.35	2.88	3.46
Difference(Tornqvist, Tornqvist (EP-TV,EW))	1.19	1.63	-0.31	0.9	0.71
Difference(Tornqvist, Tornqvist (EP-TV,QW))	4.92	3.32	3.76	2.21	1.32
Difference(Sato-Vartia, Feenstra)	2.14	3.07	1.72	1.38	1.68
Corr(Tornqvist, Tornqvist (EP-TV,QW))	0.99	0.95	0.91	0.97	0.99
Corr(Tornqvist, Tornqvist (EP-TV,EW))	1.00	0.92	0.95	0.99	0.98
Corr(Tornqvist (EP-TV,QW), Tornqvist (EP-TV,EW))	1.00	0.98	0.98	1.00	1.00
Corr(Feenstra, Tornqvist)	0.97	0.98	0.99	0.98	0.93
Corr(Feenstra, Tornqvist (EP-TV,QW))	0.96	0.90	0.94	0.96	0.97
Corr(Feenstra, Tornqvist (EP-TV,EW))	0.97	0.82	0.92	0.98	0.98
Corr(Sato-Vartia, Tornqvist)	0.99	1.00	0.98	1.00	1.00
Corr(Sato-Vartia, Tornqvist (EP-TV,QW))	0.97	0.92	0.95	0.95	0.99
Corr(Sato-Vartia, Tornqvist (EP-TV,EW))	0.98	0.82	0.91	0.97	0.98
Corr(Sato-Vartia, Feenstra)	0.99	0.98	1.00	0.97	0.92

Notes: GEKS-lite is the average of the geometric mean of the chained values and the YoY price indices for q4 for each year. QW and EW indicate, respective, quantity weights and expenditure weights in the hedonic estimation. Data come from the NPD Group.

Table D.4: R^2 for Hedonic Models, NPD Data

	Log Price Level	Log Price Relative				
Estimation Method:	Log-Level	Log-Level	EP-F		EP-TV	
Estimation Weights:	QW	QW	EW	QW	EW	QW
Coffee Makers	0.62	0.05	0.20	0.20	0.21	0.23
Headphones	0.89	0.24	0.47	0.47	0.51	0.49
Memory Cards	0.71	0.05	0.10	0.09	0.15	0.13
Work/Occ Footwear	0.73	0.10	0.33	0.40	0.34	0.41
Boy's Jeans	0.72	0.22	0.36	0.46	0.42	0.50

Notes: This table reports average quarterly R^2s for hedonic regression models. The first column shows the R^2 for log price levels. The second column shows the R^2 for log price relatives that are calculated from estimated log price levels in consecutive quarters. The reported R^2 from a regression of these inferred price relatives on actual price relatives. The EP-F and EP-TV estimation methods are the Erickson and Pakes (2011) "fixed unobservables" and "time-varying unobservables" methods described in section 4.1.1, respectively. For the log-level models, weights are the quantity shares in the current period. For the EP-F and EP-TV models, weights are the average quantity shares (QW) or the average expenditure shares (EW) in the current and lagged periods. The timing of the estimation weights aligns with the timing of the weights used to construct the index numbers. The EP-TV model includes lagged residuals from a log-level hedonic regression.

Table D.5: Estimated Relationship Between Lagged and Current Residual: NPD Data

Product	Estimate
Headphones	0.89 (0.016)
Memory Cards	0.87 (0.016)
Coffeemakers	0.99 (0.016)
Occupational Footwear	0.91 (0.033)
Boys' Jeans	0.61 (0.049)

Notes: Regression of lagged residual from log(level) hedonic model on current period residual. Standard errors in parentheses. Data come from NPD Group.

Table D.6: Sales-Weighted Rates of Product Entry: NPD Data

	1 Quarter	4 Quarter
Memory Cards	1.1	18.2
Coffee Makers	1.5	28.6
Headphones	1.9	22.7
Boys' Jeans	2.2	15.5
Occupational Footwear	2.0	17.8

Notes: Average rates of product entry, sales weighted. Entry rate computed as the sales of entering goods as a percentage of sales of common goods in the previous period. For 1 quarter, entry reflects goods with positive sales in current quarter and zero in prior quarter. For 4 quarter, entry reflects goods with positive sales in current quarter and zero 4-quarters ago.

Table D.7: Estimated Elasticities of Substitution: NPD Data

Product	Elasticity of Substitution
Headphones	7.634 (0.748)
Memory Cards	5.623 (0.484)
Coffeemakers	5.183 (1.289)
Occupational Footwear	7.31 (0.533)
Boys' Jeans	7.861 (0.565)

Notes: Elasticities of substitution for Feenstra index and CUPI estimated using approach of Feenstra (1994) and Redding and Weinstein (2020). Standard errors in parentheses. Data come from NPD Group.

Table D.8: Comparison of GEKS-Lite and Rolling-Year GEKS, Annual Chained Price Indices, NPD Data

	Memory Cards	Coffeemakers	Headphones	Boys' Jeans	Occupational Footwear
Tornqvist (Chained)	-17.34	-8.48	-10.93	-7.35	-3.29
Tornqvist (GEKS-Lite)	-15.89	-6.42	-10.91	-5.40	-2.28
Tornqvist (Rolling-Year GEKS)	-15.89	-6.47	-8.72	-4.66	-1.94

Notes: The chained Tornqvist indices in the first row are calculated from chained quarter-over-quarter price changes. The GEKS-Lite indices in the second row are the geometric means of the chained year-over-year indices and the directly computed (unchained) year-over-year indices, as described in section 4.3. The rolling-year GEKS indices in the third row implements the method of Ivancic et al. (2011). The data reports annual average price indices for Q4. Data come from the NPD Group.

Table D.9: Nested Estimated Elasticities of Substitution: NPD Data

Product	Groups	Elasticity of Substitution			
		Within		Across	
Headphones	Manual	8.609	(0.544)	7.704	(0.491)
	Hedonic	9.537	(0.969)	8.958	(0.423)
Memory Cards	Manual	6.31	(0.675)	4.534	(0.298)
	Hedonic	6.621	(0.657)	5.25	(0.586)
Coffeemakers	Manual	5.495	(0.791)	3.42	(0.63)
	Hedonic	5.345	(0.99)	5.306	(0.374)
Occupational Footwear	Manual	5.545	(0.509)	3.057	(0.493)
	Hedonic	6.199	(0.548)	4.135	(0.769)
Boys' Jeans	Manual	7.439	(1.5)	3.234	(0.734)
	Hedonic	8.156	(1.82)	3.418	(0.657)

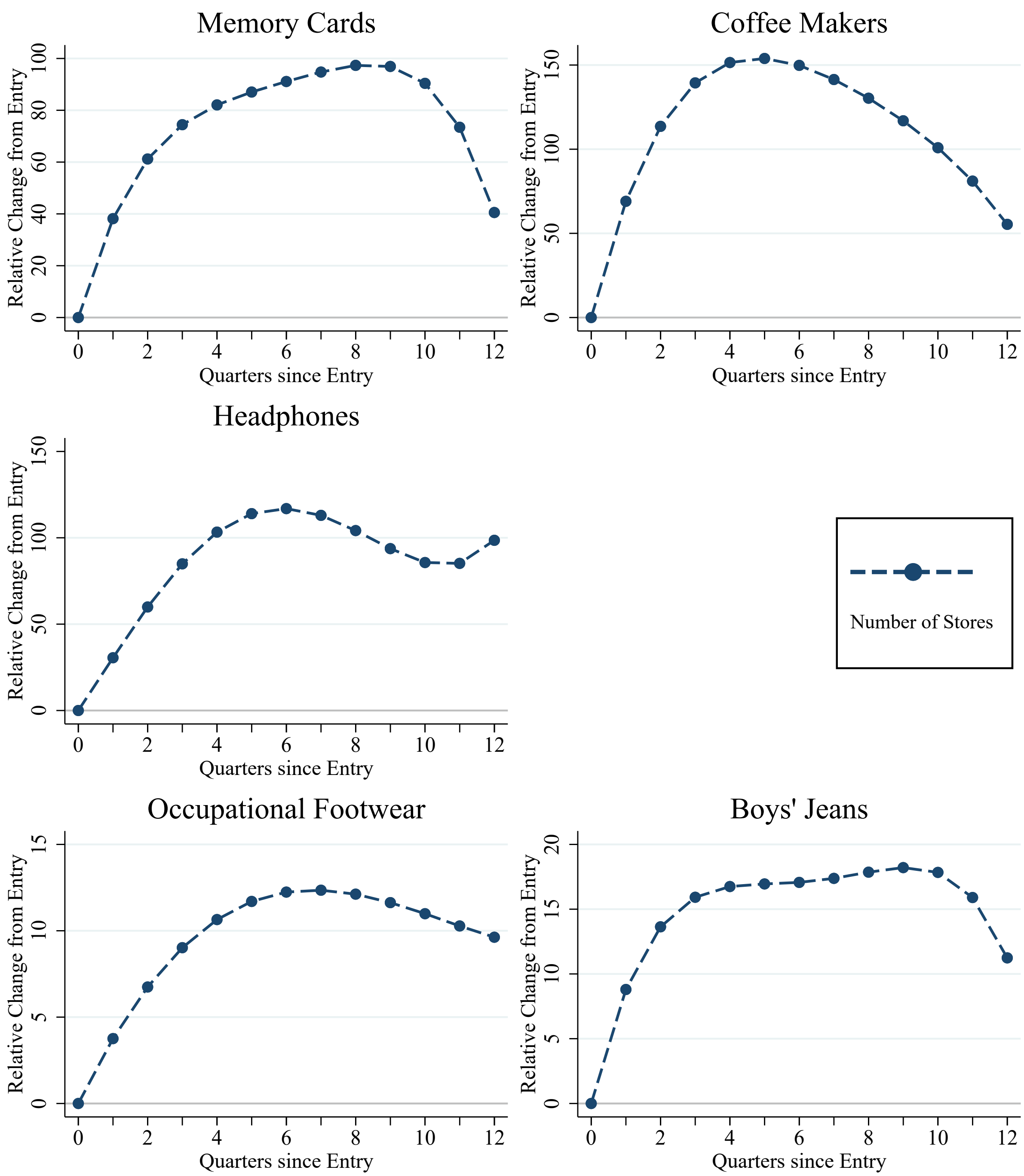
Notes: Estimated elasticities of substitution for nested CES models. Standard errors in parentheses. Data come from NPD Group. Within-nest elasticities are estimated using the methodology of Feenstra (1994) and Redding and Weinstein (2020). Across-nest elasticities are estimated using the nested CES estimation procedure of Hottman et al. (2016) modified to be robust to product entry and exit.

Table D.10: Estimated Autocorrelation in Relative Product Quality: NPD Data

Product	OLS Estimate	Blundell-Bond Estimate
Headphones	0.973 (0.002)	0.656 (0.027)
Memory Cards	0.970 (0.004)	0.718 (0.034)
Coffeemakers	0.982 (0.004)	0.765 (0.036)
Occupational Footwear	0.978 (0.004)	0.665 (0.030)
Boys' Jeans	0.896 (0.010)	0.727 (0.032)

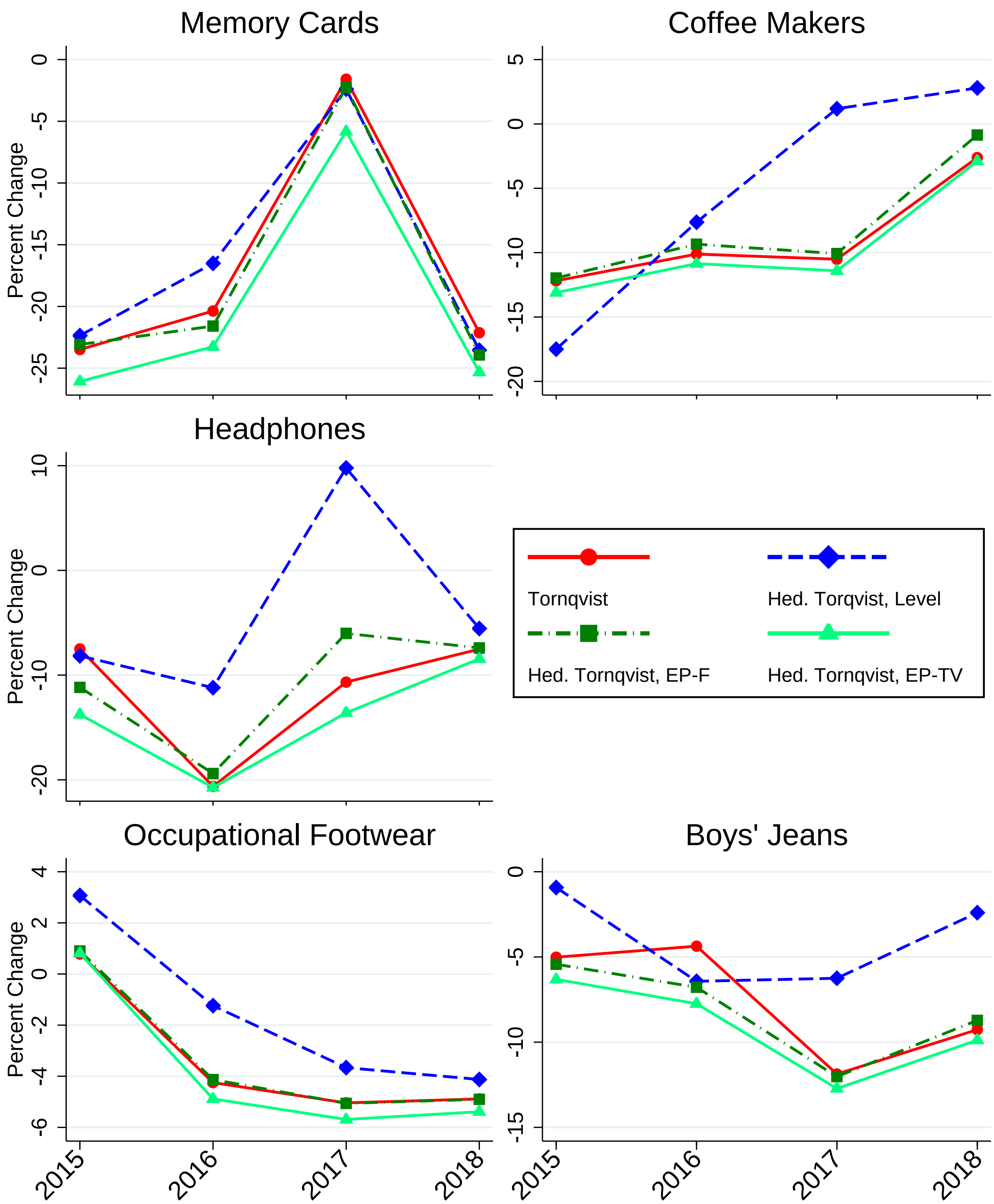
Notes: Regression of current relative quality on lagged relative quality from CUPI model. Standard errors in parentheses. Blundell-Bond allows for item-level fixed effects. Data come from NPD Group.

Figure D.1: Product Lifecycle Dynamics, Number of Stores: NPD Data



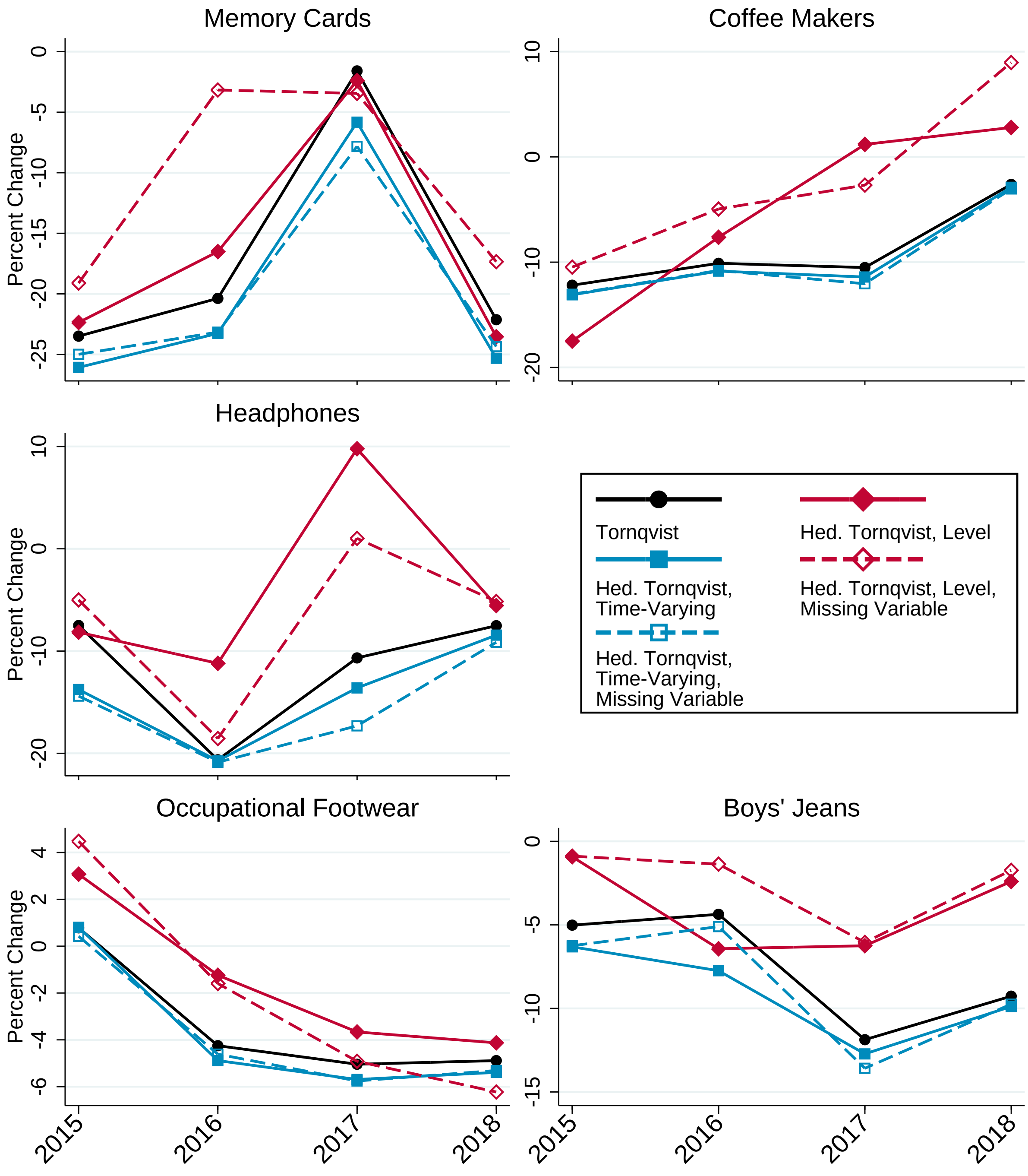
Notes: Number of stores relative to their value in the period of their initial entry. Entry occurs in period 0. All series are smoothed with a quartic spline.

Figure D.2: Alternative Hedonic Estimation Strategies: NPD Data



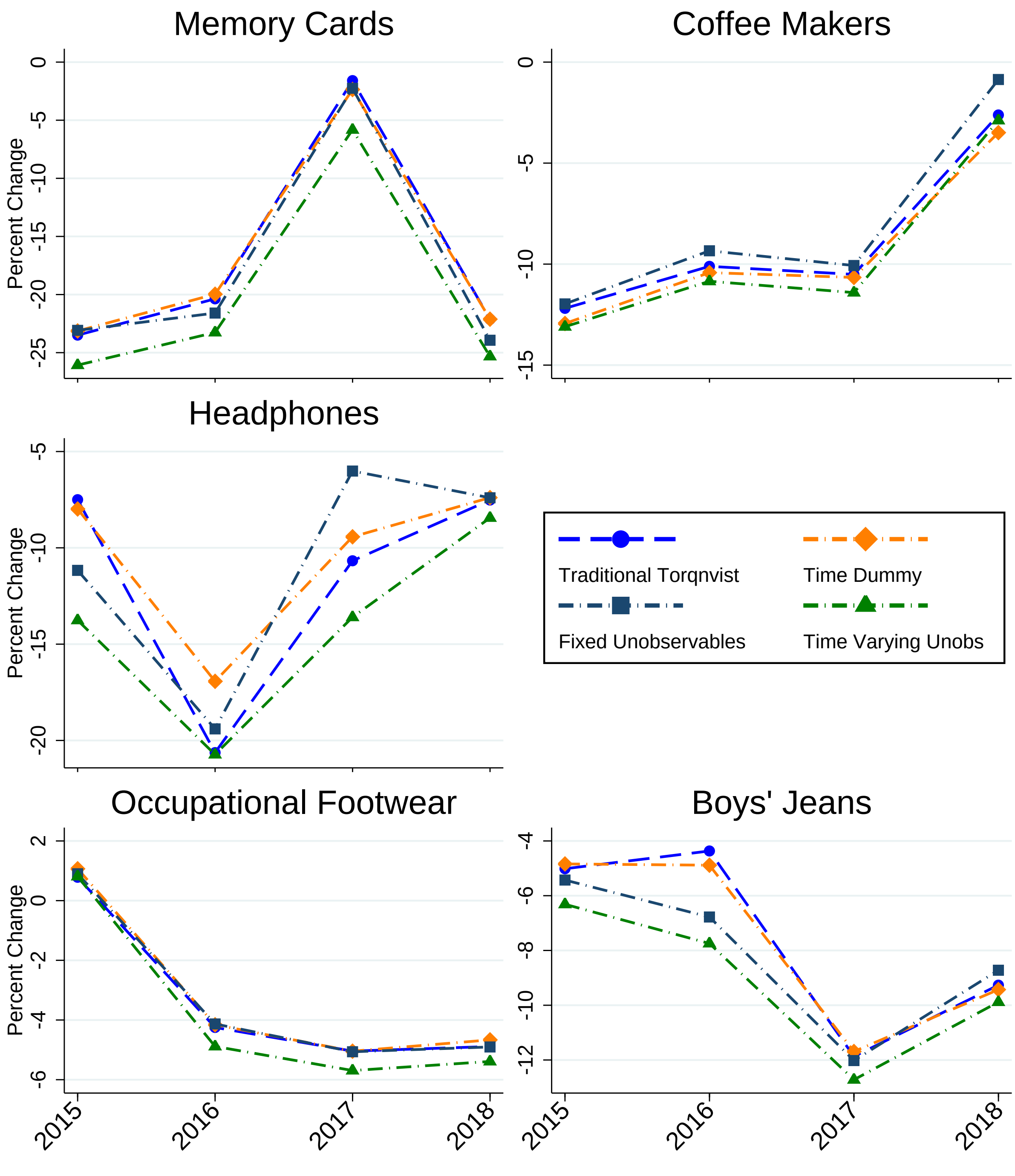
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices.

Figure D.3: Test of Time-Varying Unobservable Hedonic Specification
First-Differences and Levels Estimation: NPD Data



Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices.

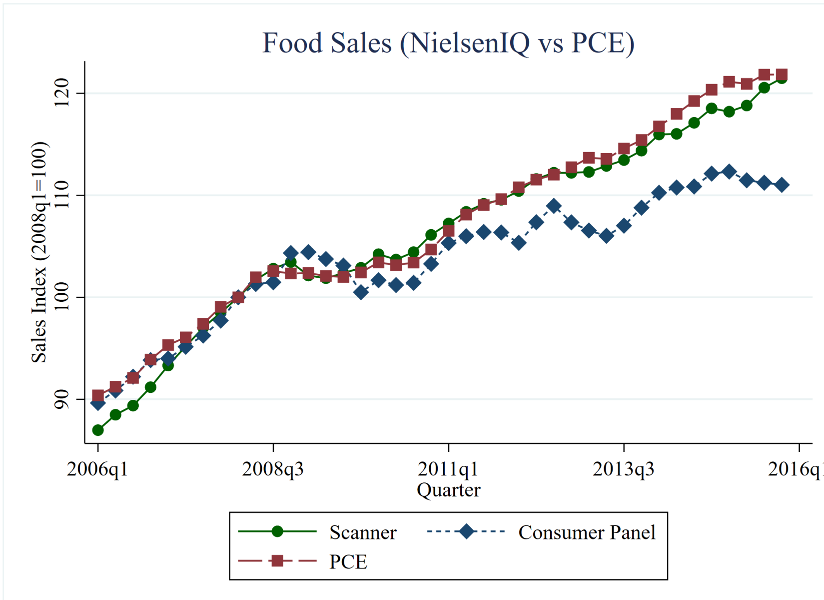
Figure D.4: Hedonic Specifications, Time Dummy, Fixed vs. Time-Varying Unobservables: NPD Data



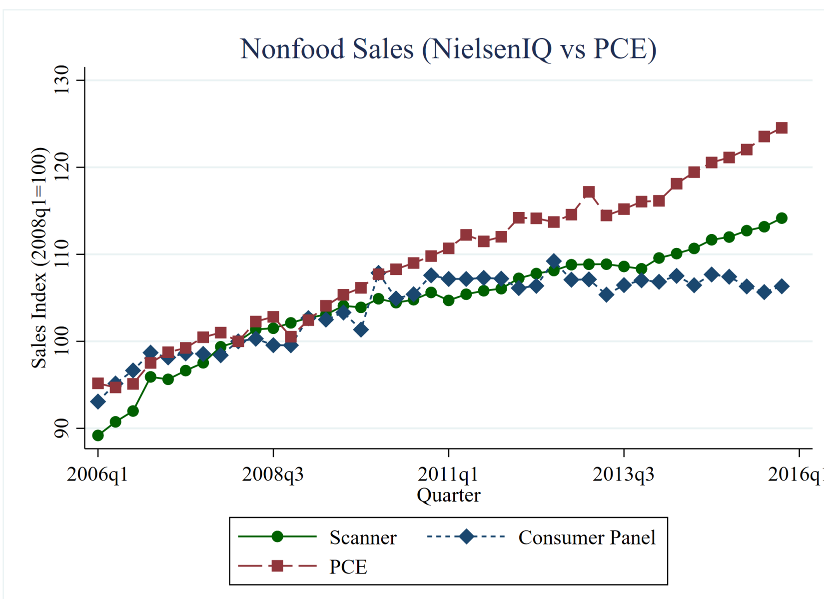
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. The time-dummy Tornqvist index uses adjacent period estimation with Tornqvist market share weights. The fixed unobservables model estimates hedonic models of log change in price using WLS and average quantity-share weights. The time-varying unobservables model adds lagged hedonic level residuals to the log-difference specification.

Figure D.5: PCE vs NielsenIQ Sales for Scanner and Consumer Panel, Food and Nonfood

(a) Food



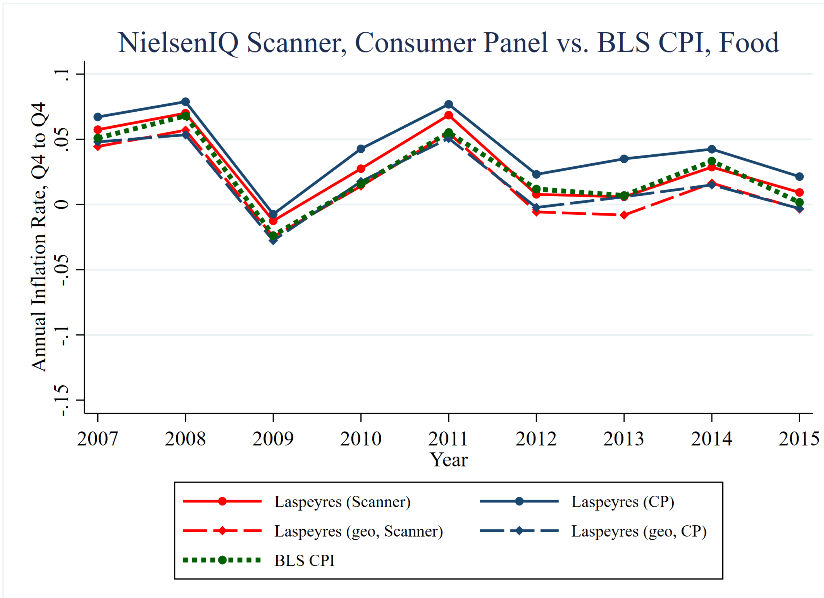
(b) Nonfood



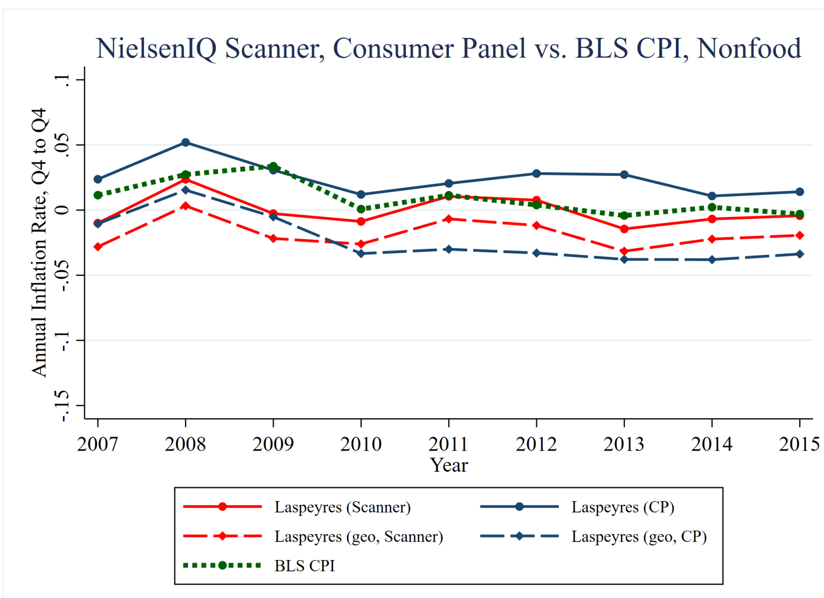
Notes: NielsenIQ values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. Figures uses NielsenIQ Scanner and Consumer Panel data for Food and Nonfood (aggregated) product groups. PCE is personal consumption expenditures (nominal) from BEA. All series indexed to 1 in 2008:1.

Figure D.6: BLS CPI vs NielsenIQ Laspeyres, Food and Nonfood

(a) Food

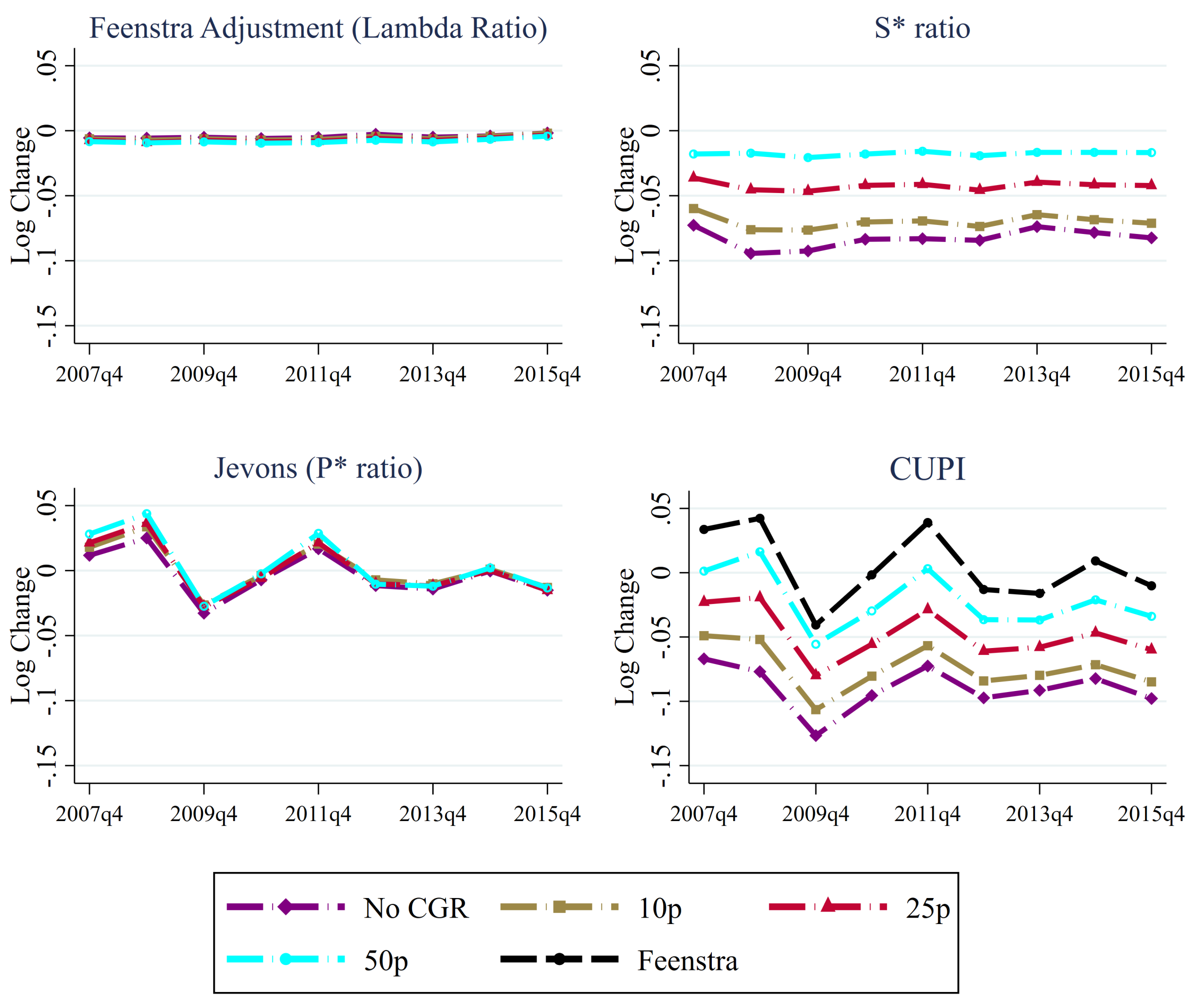


(b) Nonfood



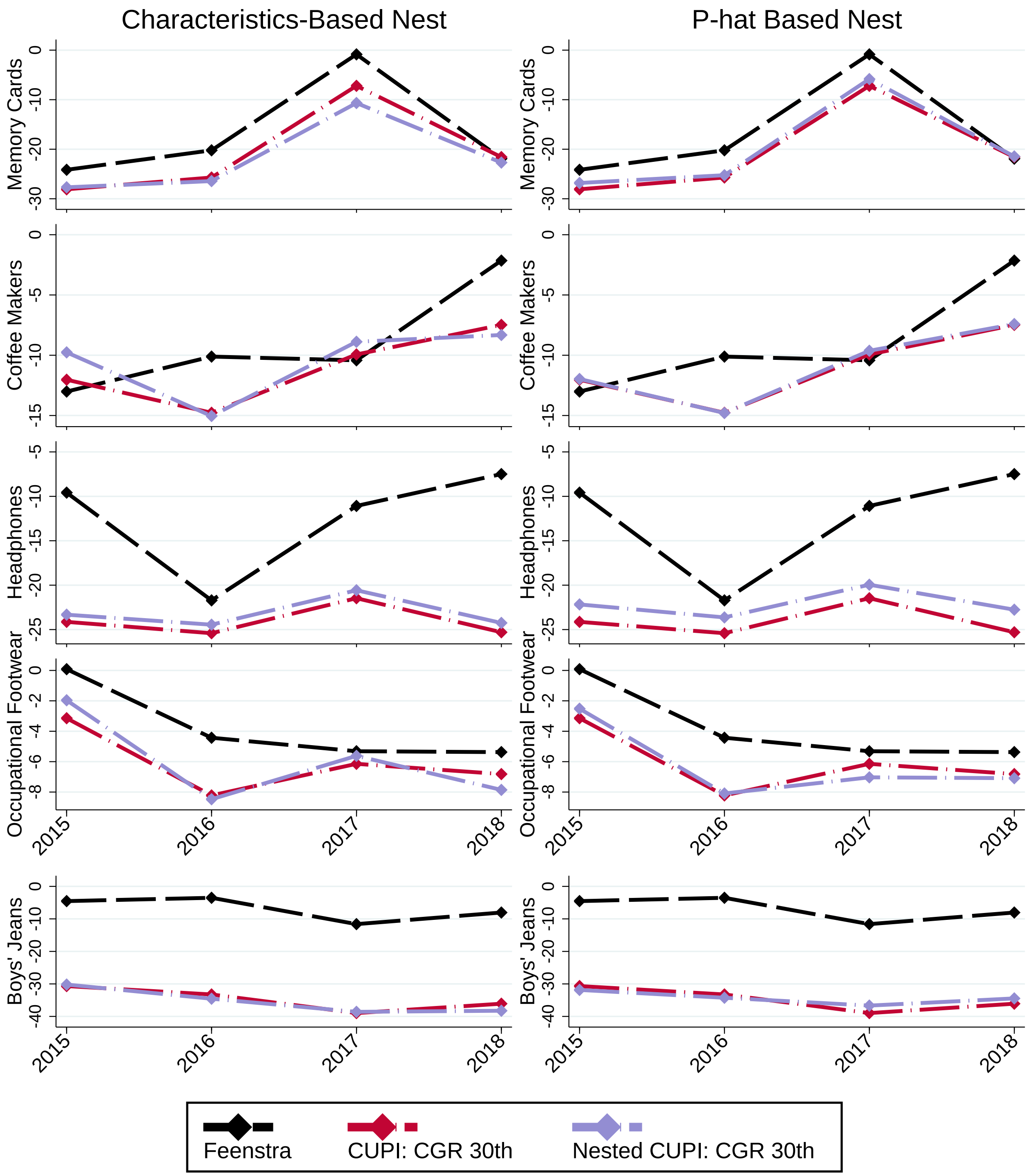
Notes: Nielsen IQ values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. The panels use NielsenIQ Scanner and Consumer Panel data for Food and Nonfood (aggregated) product groups. The BLS CPI was computed by BLS staff.

Figure D.7: CUPI and Its Components with Alternative CGRs: NielsenIQ Food



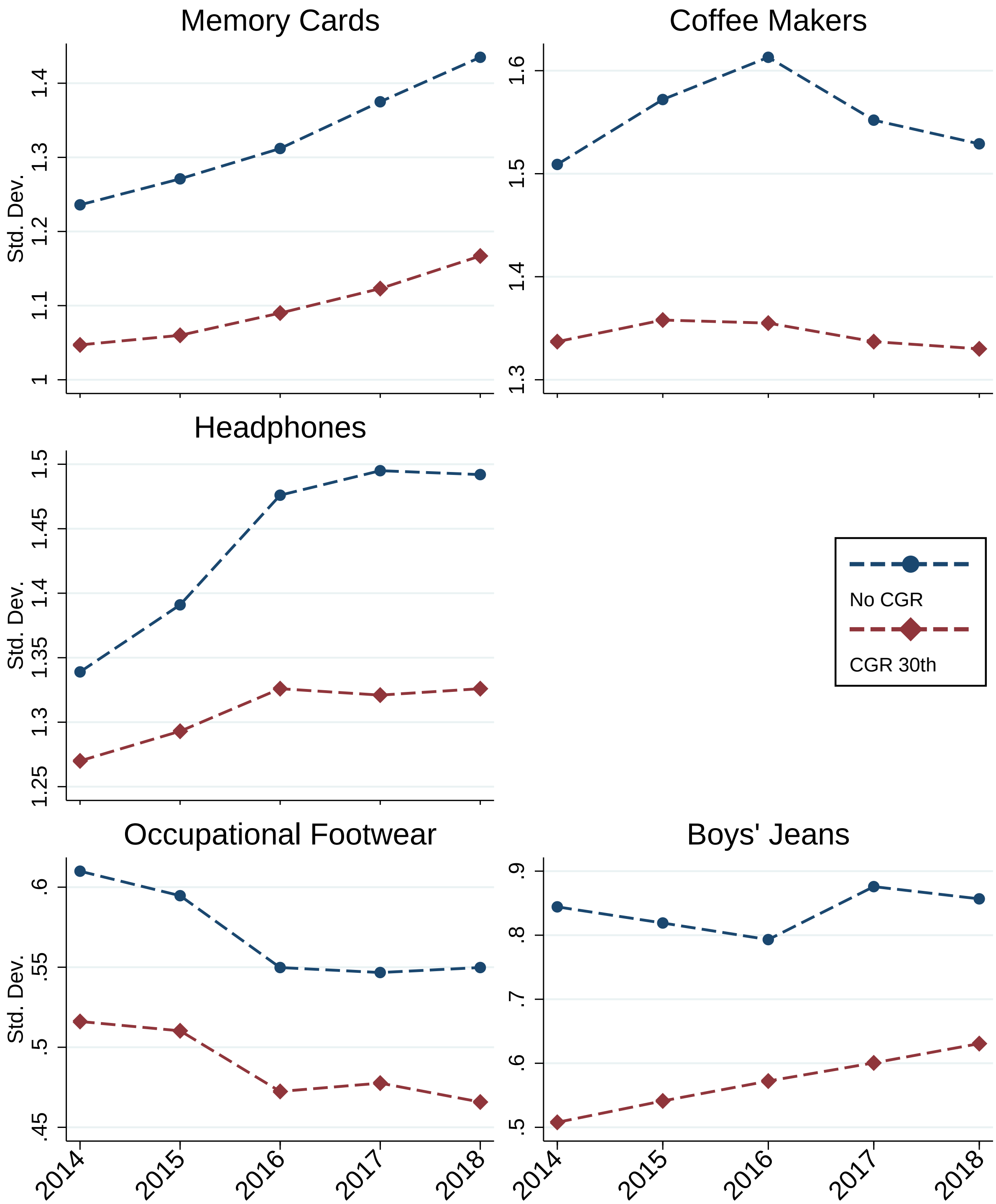
Notes: The figure shows NielsenIQ Retail Scanner data for food product groups. Each plot shows log changes from the fourth quarter of the previous year to the fourth quarter of the labeled year. The values are cumulative changes from chained quarterly indices. The Feenstra adjustment and S^* ratio panels show the adjustments scaled by $\frac{1}{\hat{\sigma}-1}$ for each product group, so that the sum of the those two components and the Jevons index (P^* ratio) equals the CUPI.

Figure D.8: Nested CUPI: Characteristics- and P-Hat- Based Nests
Percent Changes: NPD Data



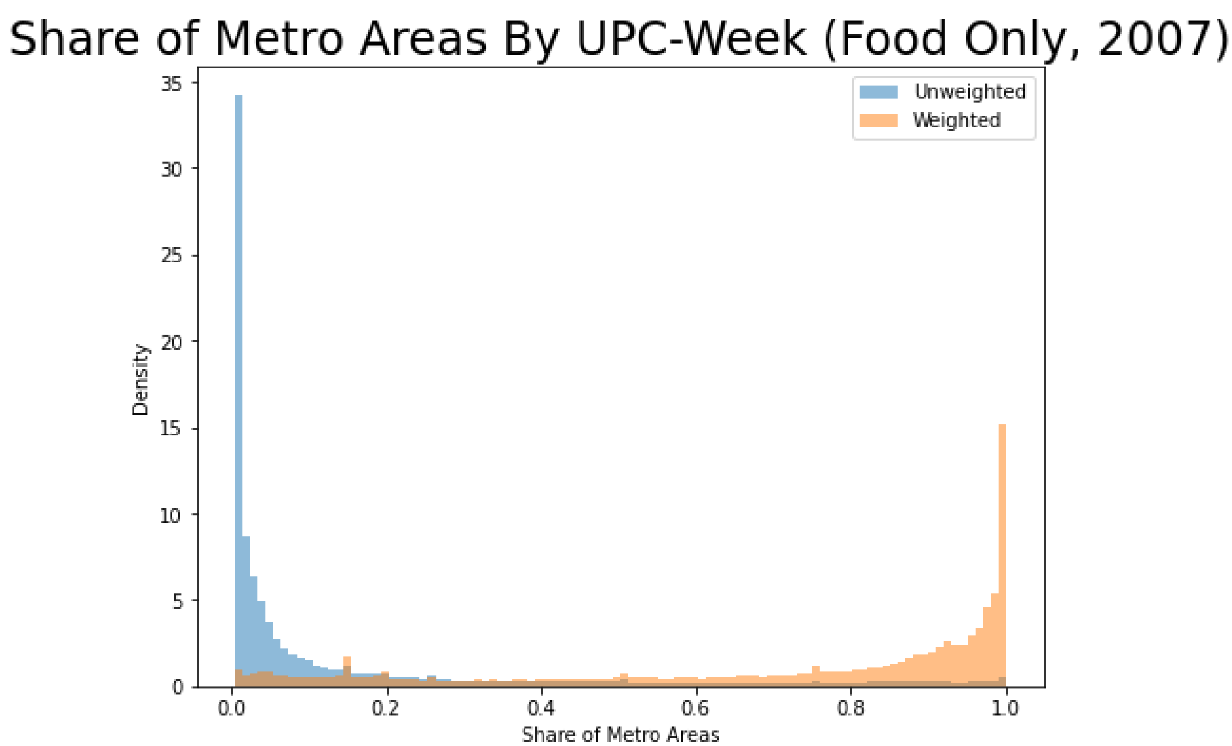
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. Laspeyres is the geometric Laspeyres. For the characteristics-based nests, we assign items to groups based on shared observable characteristics. The p-hat based nests are based on the decile of predicted prices from unweighted hedonic log-level models. We estimate period-by-period hedonic models and assign items their most common decile over all periods.

Figure D.9: Dispersion of Relative Product Appeal: NPD Data



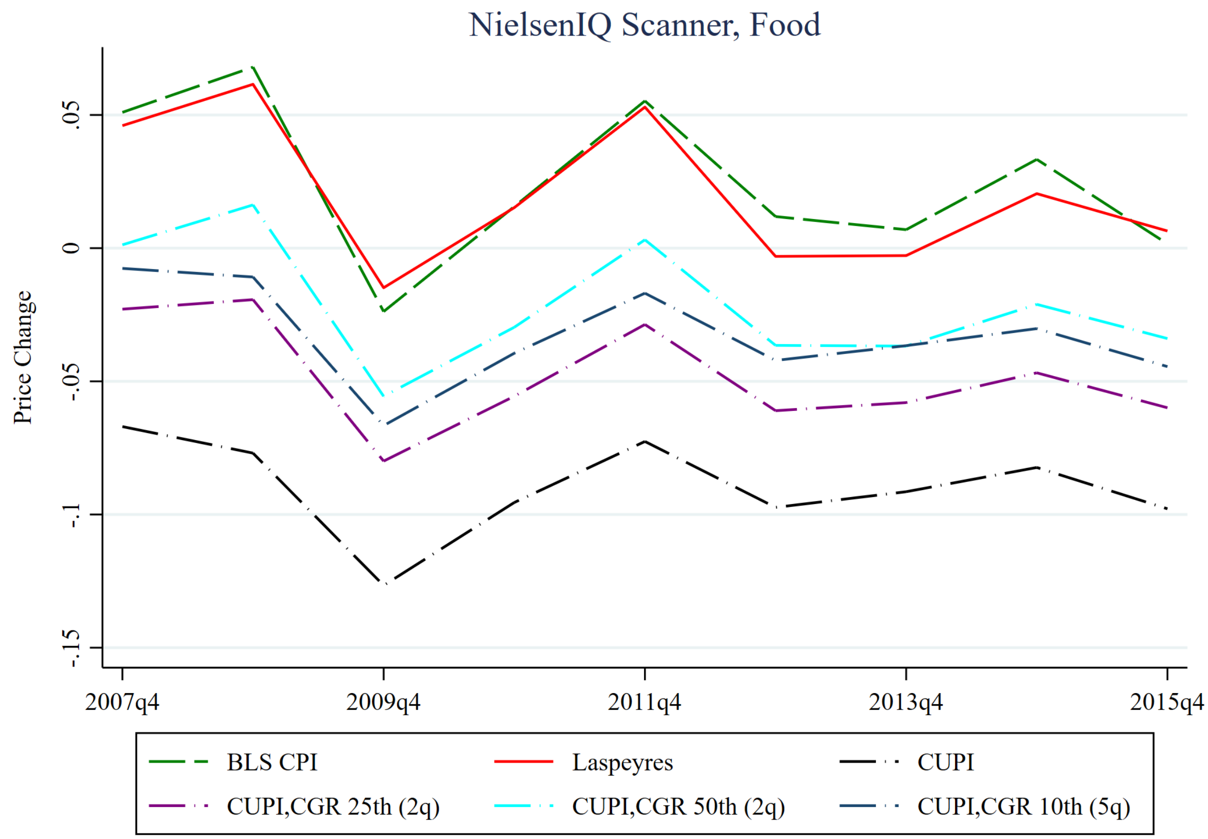
Notes: Values are annual averages of quarterly dispersion (standard deviation) in normalized log relative product appeal for common goods. Reported are the annual averages without imposing a common goods rule and also those with a 30th percentile common goods rule.

Figure D.10: Sales-weighted and Unweighted Distributions of Market Penetration of Items: NielsenIQ Food



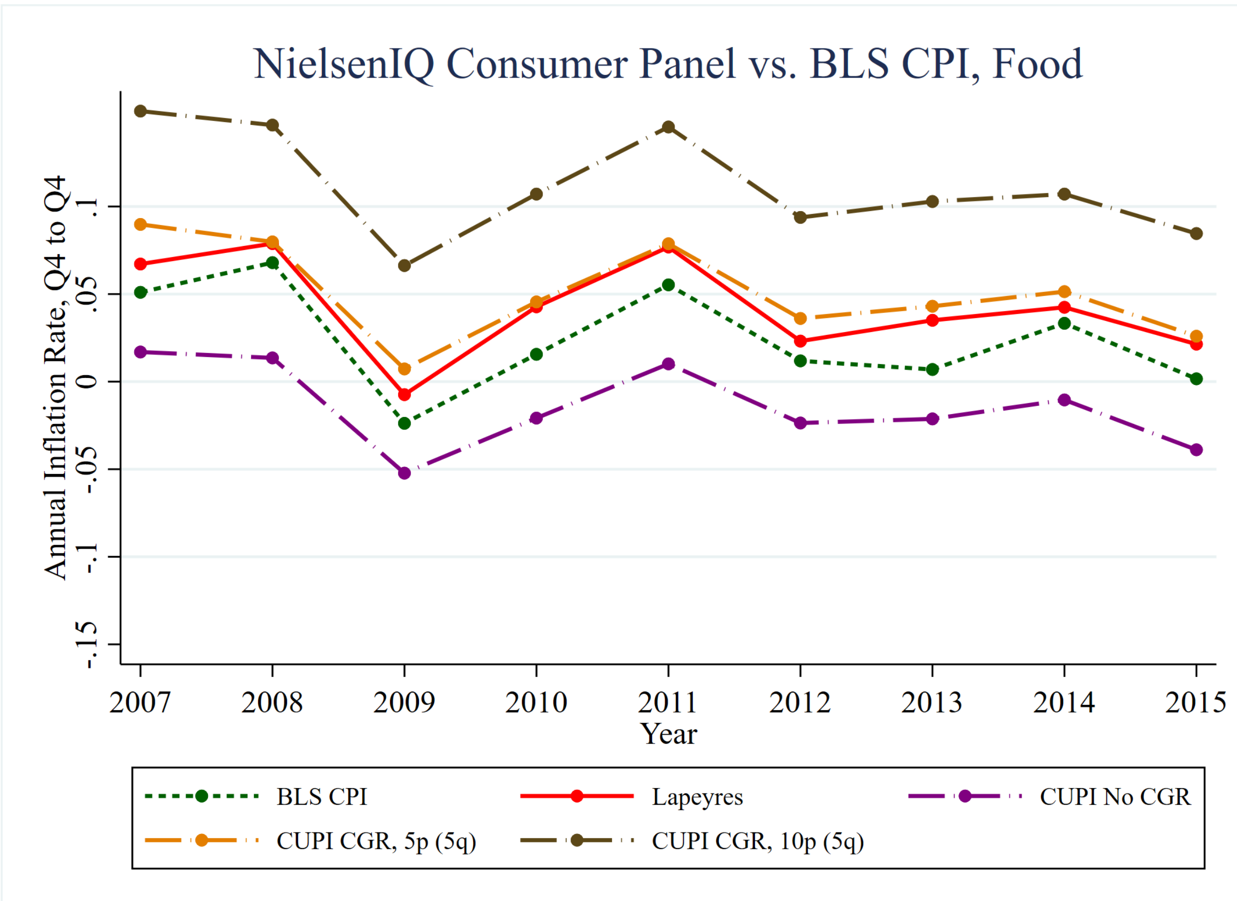
Notes: All UPC items at a weekly frequency are used from 2006-2015. Unweighted shows the market penetration at the metro area of the unweighted pooled distribution. Sales-weighted shows the equivalent using sales weights. Figure uses NielsenIQ Retail Scanner data for food product groups.

Figure D.11: Common Goods Rules–2-quarter vs 5-quarter Horizons: NielsenIQ Food



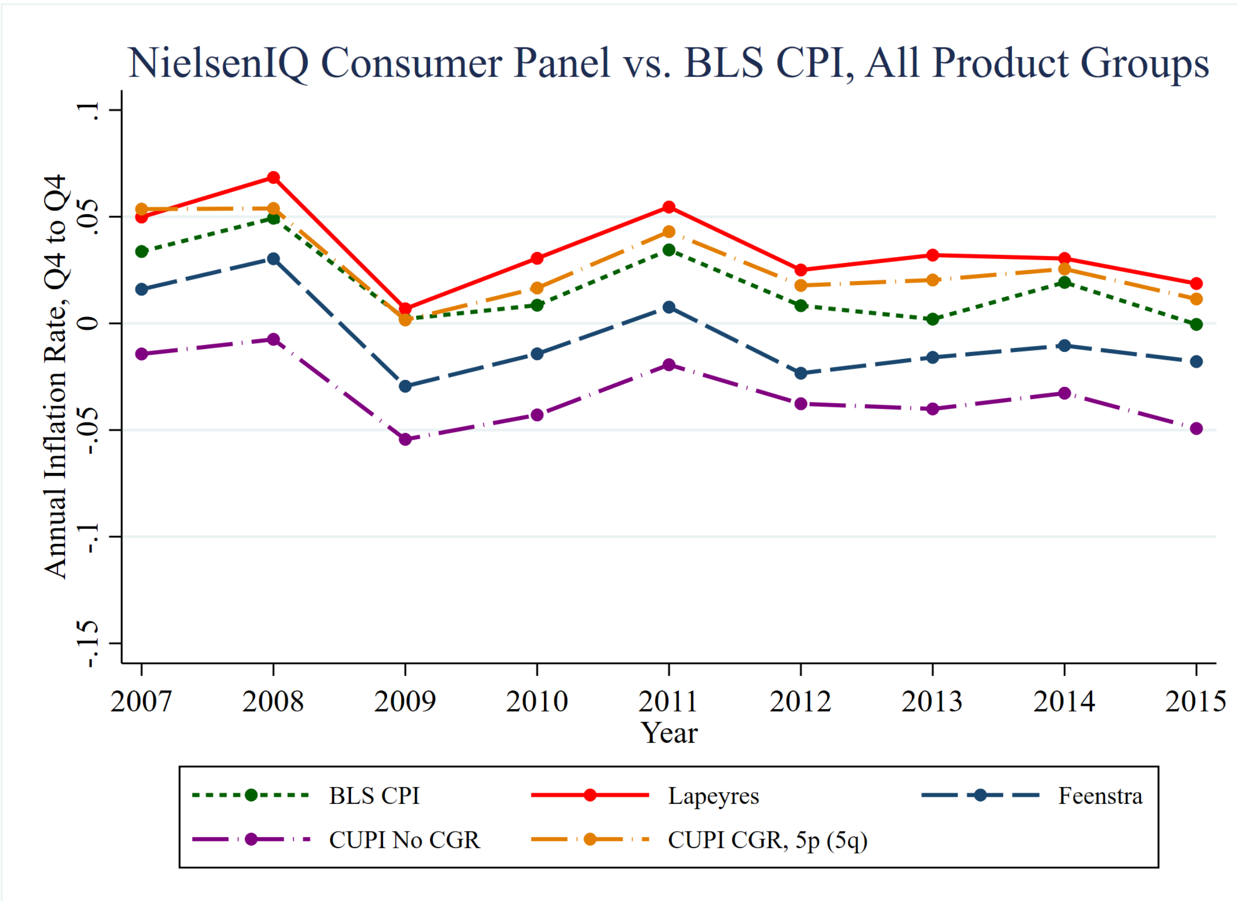
Notes: Figure uses NielsenIQ Scanner data for food. The $2q$ CUPI computes CGR percentile thresholds using sales pooled over a two quarter horizon (t and $t-1$). The $5q$ CUPI computes CGR percentile thresholds using sales pooled over a 5 quarter horizon (current and prior 4 quarters). Laspeyres is arithmetic.

Figure D.12: Common Goods Rules: NielsenIQ Food, Consumer Panel



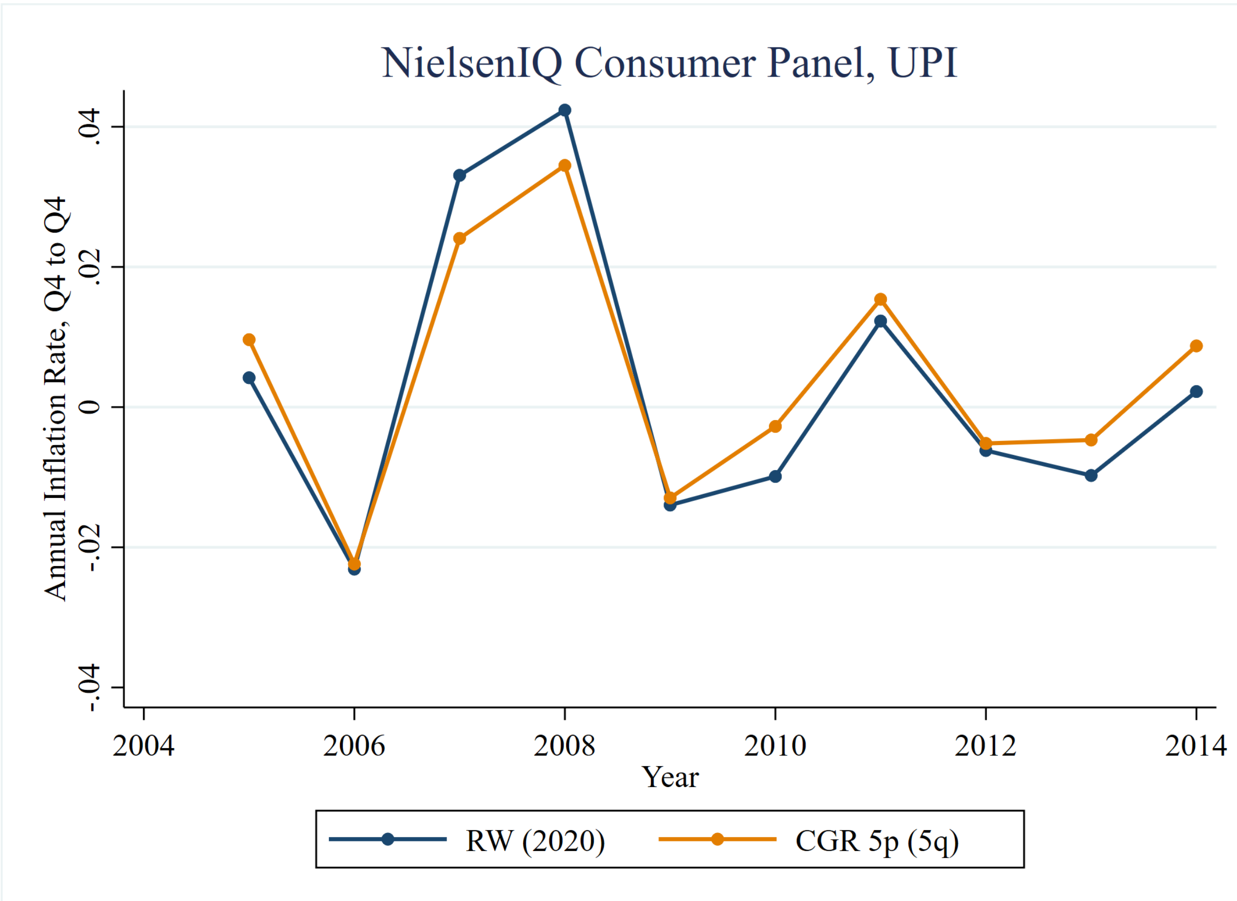
Notes: Figure uses NielsenIQ Consumer Panel data for food. The 5q CUPI computes CGR percentile thresholds using sales pooled over a five quarter horizon (t and $t-1$). (current and prior 4 quarters). Laspeyres is arithmetic.

Figure D.13: Common Goods Rules: NielsenIQ Consumer Panel



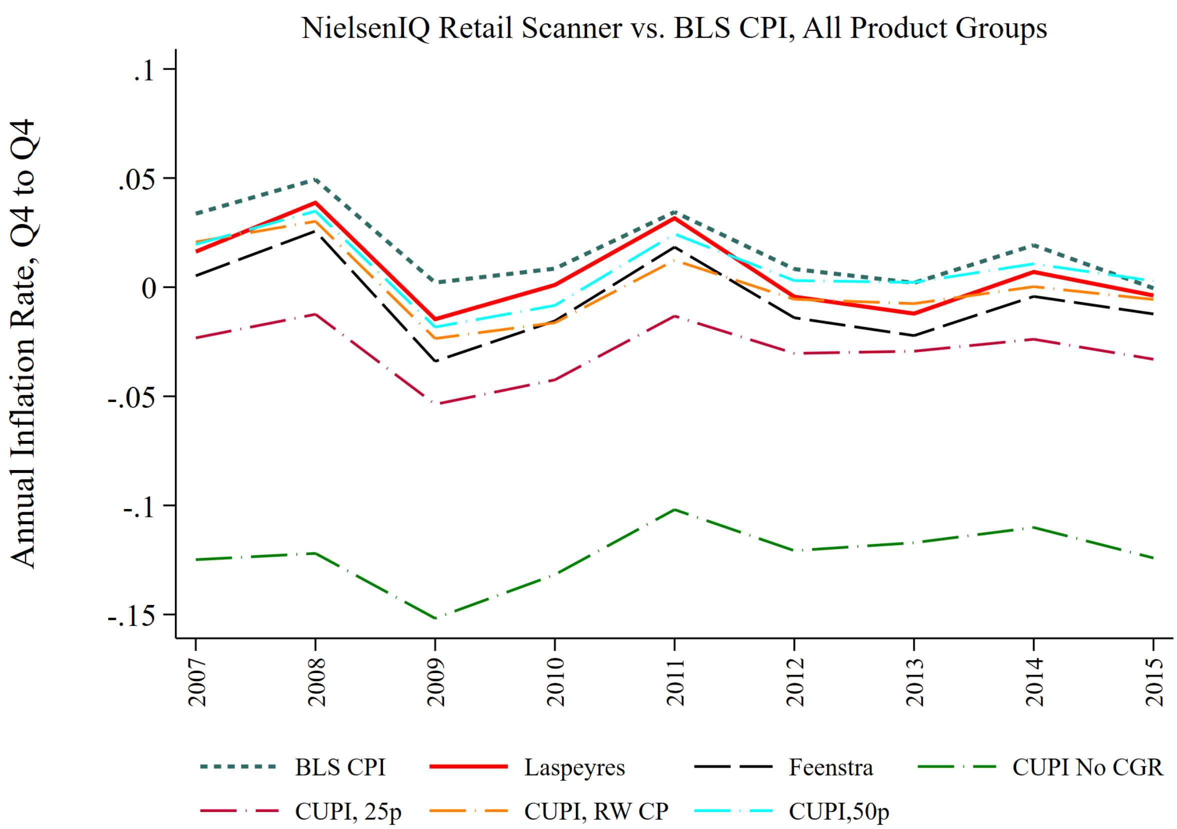
Notes: Figure uses NielsenIQ Consumer Panel data for food and nonfood product groups. The series "CUPI CGR RW" uses a 5th-percentile sales cutoff for the common goods rule. Percentile computed from sales pooled over 5 quarter horizon (current and prior 4 quarters). Laspeyres is arithmetic.

Figure D.14: Replication of Redding and Weinstein (2020): NielsenIQ Consumer Panel



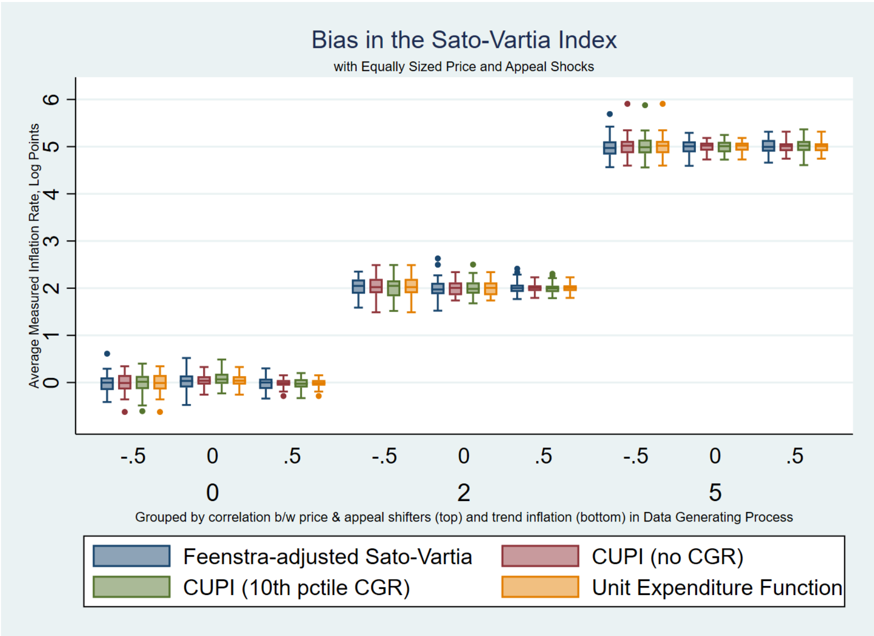
Notes: Figure uses NielsenIQ Consumer Panel for food and nonfood product groups. The indices are YoY for Q4. The series RW(2020) uses the same CGR duration rule as in Redding and Weinstein (2020). The series 5p(5q) use percentiles based on sales pooled over 5 quarter horizon (current and prior 4 quarters.

Figure D.15: Common Goods Rules: NielsenIQ Scanner Data, Food and Nonfood



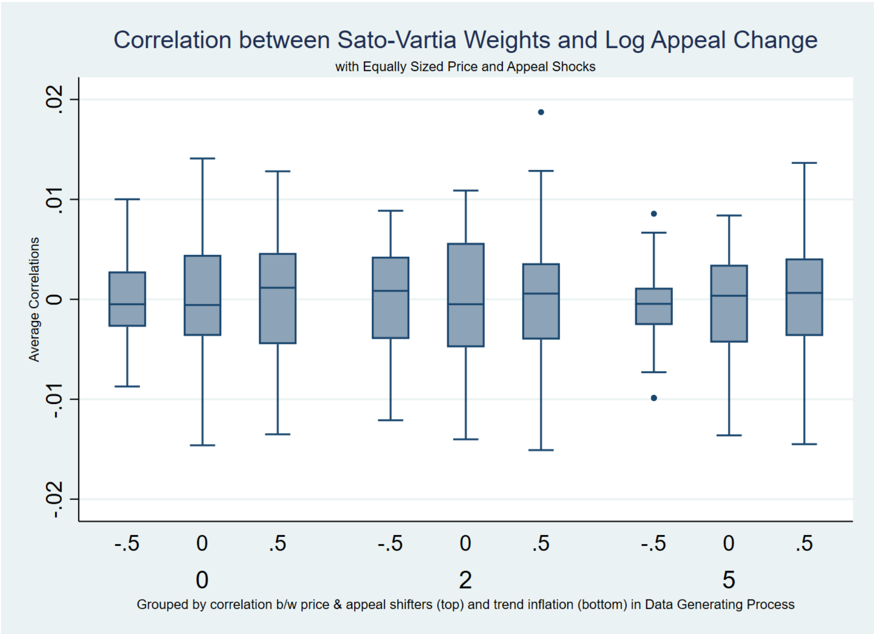
Notes: Figure uses NielsenIQ Retail Scanner data for food and nonfood product groups. The "CUPI, 25p" and "CUPI, 50p" series use 25th- and 50th-percentile cutoffs for the common goods rule, respectively. The series "CUPI, RW CP" uses the CGR 5th percentile threshold from the consumer Panel data for the common goods rule. Percentiles based on sales pooled over 5 quarter horizon (current and prior 4 quarters). Laspeyres is arithmetic.

Figure D.16: Taste Shock Bias: Baseline Simulations with Different Trend Inflation Rates



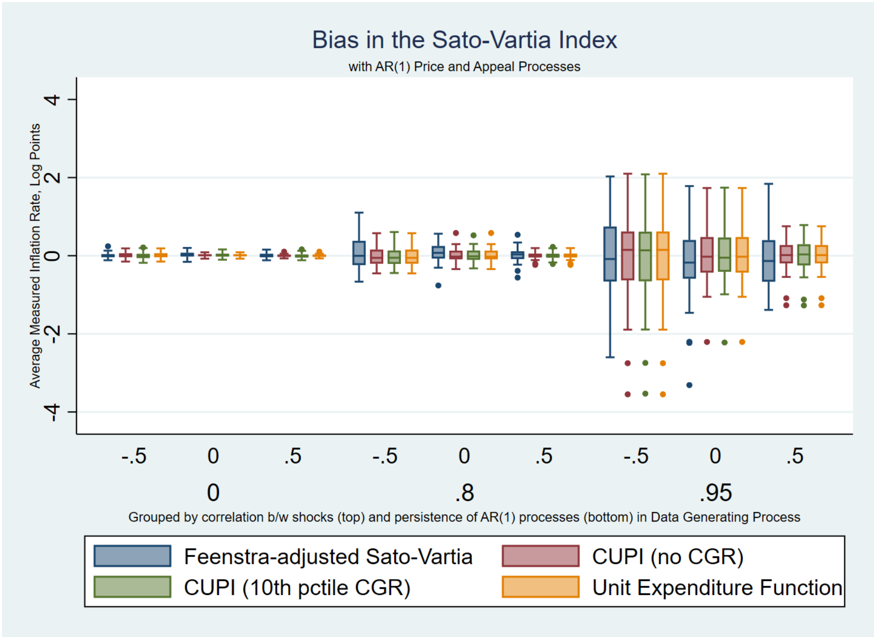
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section C.2.2.1. The 0, 2, and 5 along the horizontal axis specify the average trend rates of log price growth in log points, while the -0.5, 0, and 0.5 along the horizontal axis specify the correlations between log price innovations and log appeal parameters. The vertical axis displays the average log inflation rate across simulation periods.

Figure D.17: Correlation between Sato-Vartia Weights and Log Appeal Shocks: Baseline Simulations with Different Trend Inflation Rates



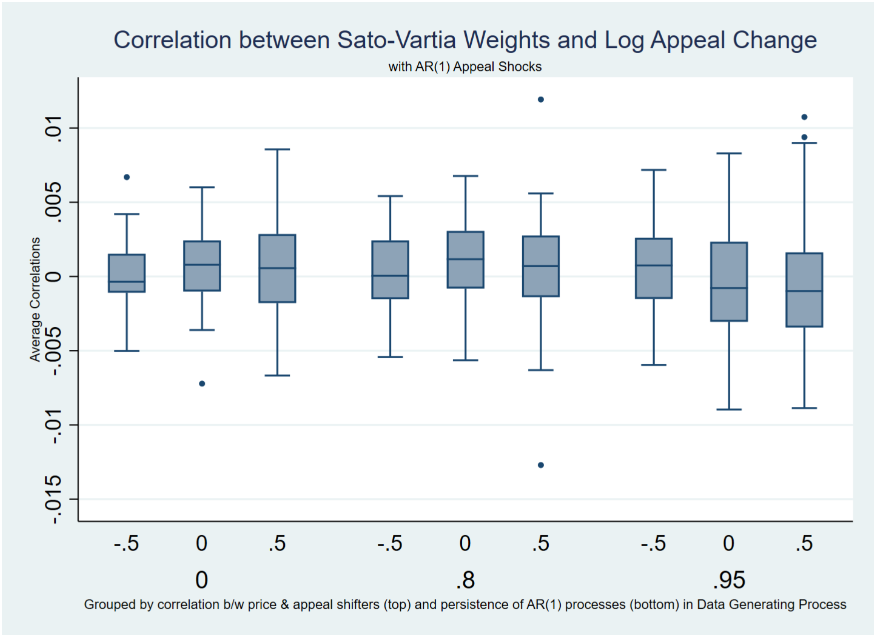
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section C.2.2.1. The 0, 2, and 5 along the horizontal axis specify the average trend rates of log price growth in log points, while the -0.5, 0, and 0.5 along the horizontal axis specify the correlations between log price innovations and log appeal parameters. The vertical axis displays the average correlation between the Sato-Vartia weights and the log appeal shocks across simulation periods.

Figure D.18: Taste Shock Bias: AR(1) Processes for Prices and Appeal



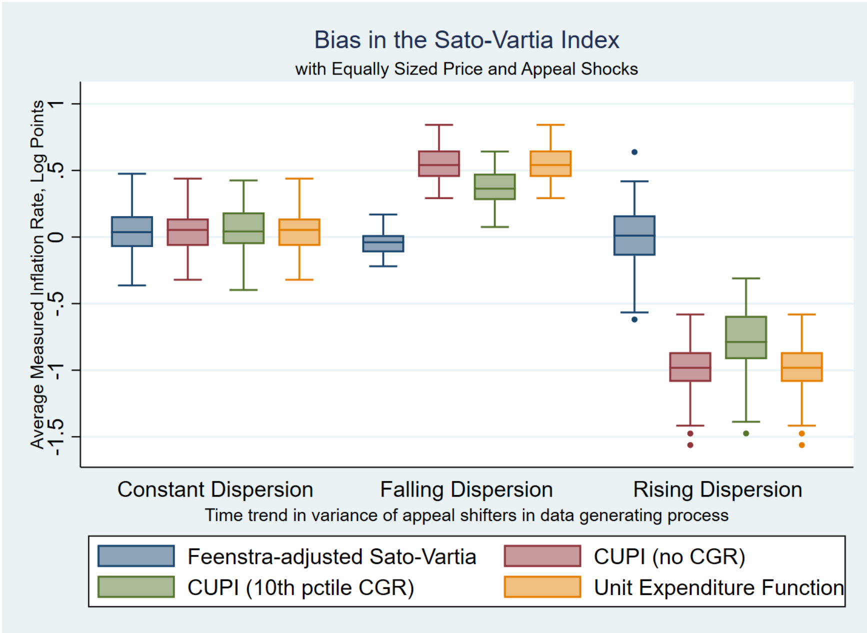
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section C.2.2.2. The 0, 0.8, and 0.85 along the horizontal axis specify the autocorrelations of log prices and log appeal parameters, while the −0.5, 0, and 0.5 along the horizontal axis specify the correlations between log price and appeal innovations. The vertical axis displays the average log inflation rate across simulation periods.

Figure D.19: Correlation between Sato-Vartia Weights and Log Appeal Shocks: AR(1) Processes for Prices and Appeal



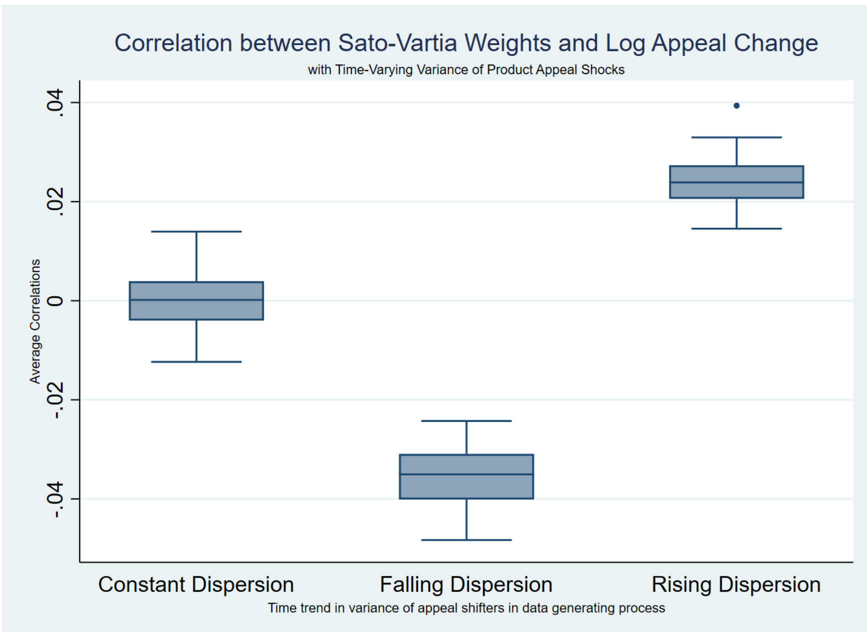
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section C.2.2.2. The 0, 0.8, and 0.85 along the horizontal axis specify the autocorrelations of log prices and log appeal parameters, while the −0.5, 0, and 0.5 along the horizontal axis specify the correlations between log price and appeal innovations. The vertical axis displays the average correlation between the Sato-Vartia weights and the log appeal shocks across simulation periods.

Figure D.20: Taste Shock Bias with Time-Varying Appeal Variance



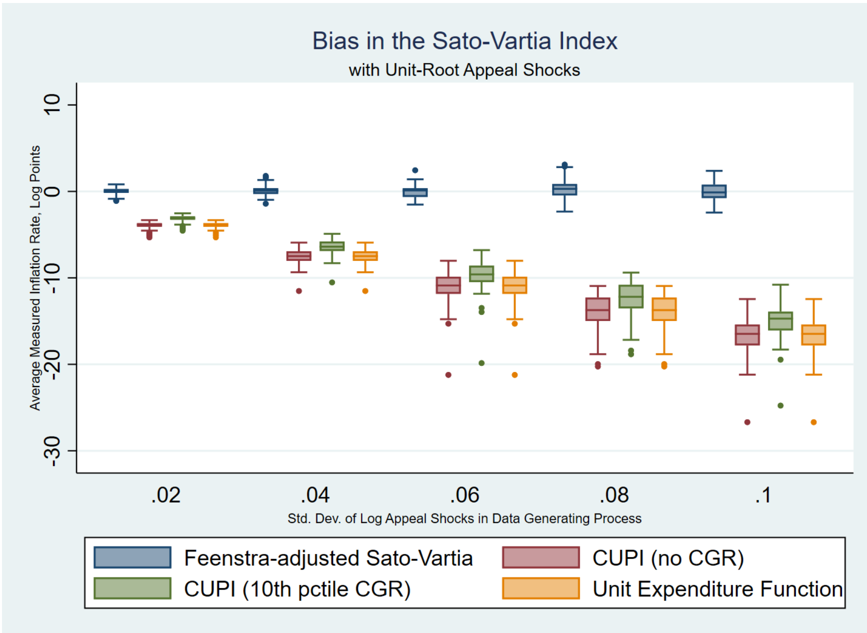
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section C.2.3.1. The variance of log product appeal draws is constant over time in the simulations labeled "Constant Dispersion," falling over time in the simulations labeled "Falling Dispersion," and rising over time in the simulations labeled "Rising Dispersion." The vertical axis displays the average log inflation rate across simulation periods.

Figure D.21: Correlation between Sato-Vartia Weights and Log Appeal Shocks: Time-Varying Appeal Variance



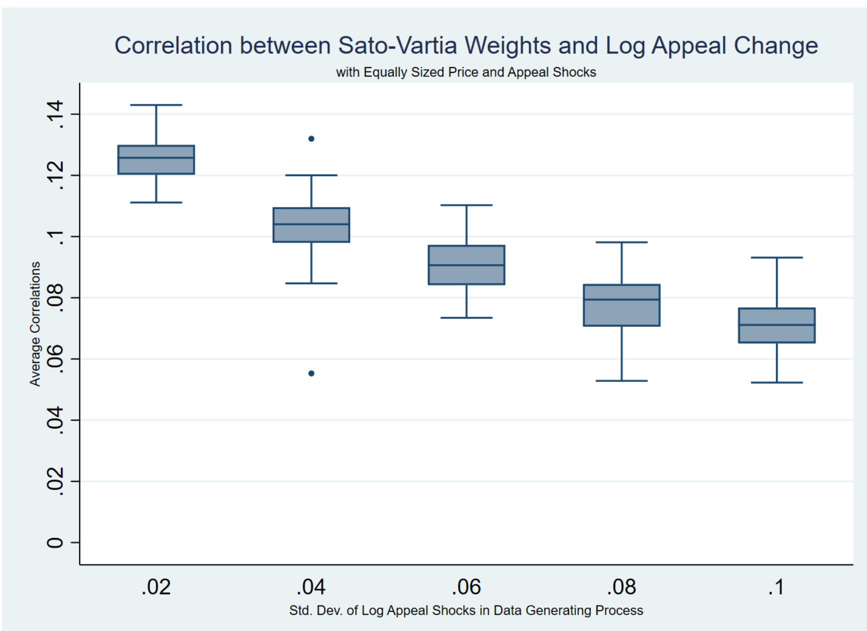
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section C.2.3.1. The variance of log product appeal draws is constant over time in the simulations labeled "Constant Dispersion," falling over time in the simulations labeled "Falling Dispersion," and rising over time in the simulations labeled "Rising Dispersion." The vertical axis displays the average correlation between the Sato-Vartia weights and the log appeal shocks across simulation periods.

Figure D.22: Taste Shock Bias with Unit Root in Product Appeal



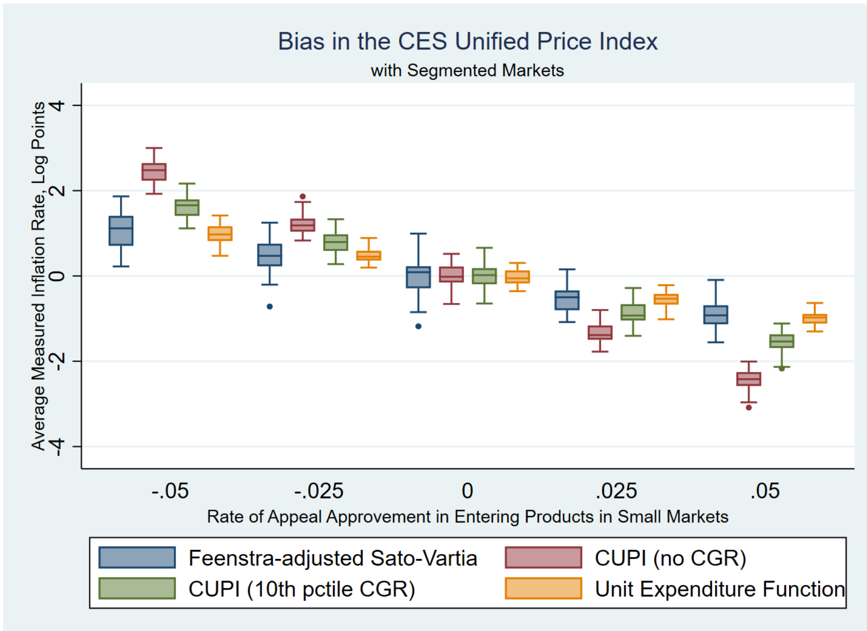
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section C.2.3.2. The horizontal axis displays the standard deviations of log appeal shocks. The vertical axis displays the average log inflation rate across simulation periods.

Figure D.23: Correlation between Sato-Vartia Weights and Log Appeal Shocks: Unit Root in Product Appeal



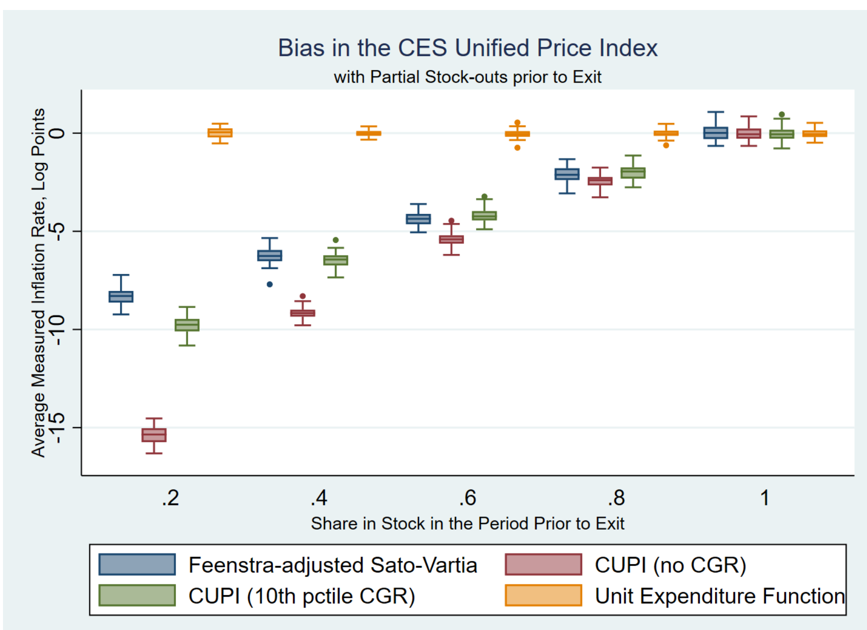
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section C.2.3.2. The horizontal axis displays the standard deviations of log appeal shocks. The vertical axis displays the average correlation between the Sato-Vartia weights and the log appeal shocks across simulation periods.

Figure D.24: Simulated CES Exact Price Indices with Segmented Markets



Notes: The figure displays measured inflation from the Monte Carlo simulations described in section C.2.5.1. The horizontal axis displays the trend rate of change in log product appeal in the "small" markets. The vertical axis displays the average log inflation rate across simulation periods.

Figure D.25: Simulated CES Exact Price Indices with Stockouts



Notes: The figure displays measured inflation from the Monte Carlo simulations described in section C.2.5.2. The horizontal axis displays the shares of products that are in stock in the period prior to exit. The vertical axis displays the average log inflation rate across simulation periods.