American Economic Review, Conditionally Accepted

Quality Adjustment at Scale: Hedonic vs. Exact Demand-Based Price Indices*

Gabriel Ehrlich John Haltiwanger Ron Jarmin David Johnson

Ed Olivares Luke Pardue Matthew D. Shapiro Laura Yi Zhao

November 2025

Item-level transactions data yield cost-of-living indices that can account for quality change and consumer substitution. Transactions data require confronting the rapid turnover of items because prices of new and existing products are interrelated in equilibrium. This paper evaluates multiple approaches to measuring quality change at scale. It shows that a hedonic superlative approach—using econometrics or machine learning for hedonic estimation combined with index formulas that require simultaneous observation of item-level price and expenditure—yields improved measures of the cost of living. Accounting for ubiquitous quality change and for consumer substitution yields lower measures of inflation than traditional, official methods.

*Ehrlich: University of Michigan; Haltiwanger: University of Maryland and NBER; Jarmin: U.S. Census Bureau; Johnson: University of Michigan and International Association for Research in Income and Wealth; Olivares: University of Maryland and U.S. Treasury; Pardue: University of Maryland and Aspen Economic Strategy Group; Shapiro: University of Michigan and NBER; Zhao: University of Maryland and Bank of Canada. Ed Olivares, Luke Pardue, and Laura Zhao worked on this project as doctoral students at the University of Maryland (and Laura Zhao subsquently as a post-doc). 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. We thank Neil Christy, Diyue Guo, Hari Jayashankar, Seula Kim, Michael Navarette, and Ben Rodriguez for outstanding research assistance. 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. There is no commingling of NPD and NielsenIQ data and associated analysis. All results using the NPD Group (currently Circana) data have been reviewed to ensure that no confidential information has been disclosed (CBDRB-FY19-122, CBDRB-FY21-074, CBDRB-FY23-067,CBDRB-FY24-0464, CBDRB-FY25-0156, and CBDRB-FY25-ESMD002-020). 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.

Retail businesses create item-level data on the prices and quantities of the goods that they sell. These data can serve as the foundation for re-engineering key economic indicators by constructing internally consistent aggregates of value, volume, and price directly from transactions. Aggregating such data has the potential to supplant traditional surveys and enumerations conducted by statistical agencies as the source data for official statistics. This paper implements and evaluates price indices that leverage item-level transactions to construct cost-of-living indices that incorporate both substitution effects from relative price changes and quality change from rapid product turnover.

Systematically addressing substitution effects and quality change has long been a goal of the statistical system, as emphasized in the Boskin Commission's recommendations (Boskin et al., 1998) and recent National Academy reports (Sichel and Mackie, 2022). This paper demonstrates that both substitution bias and quality change can be systematically addressed at scale by applying state-of-the-art hedonic techniques to item-level transactions data.

We construct hedonic price indices using both econometric and machine learning methods. Our approach builds on Erickson and Pakes (2011, "EP"), who propose a method for estimating hedonic indices that accounts for evolving valuations of observable and unobservable characteristics. High-frequency, item-level transactions with prices, quantities, and attributes (referred to as "scanner data" or "point of sales [POS] data") enable implementation of the EP approach with superlative indices (e.g., Tornqvist) in real time using expenditure weights measured consistently and simultaneously with prices. Superlative indices use weights that reflect how consumers economize on the cost of living by changing expenditures in response to relative price changes. We refer to the combination of hedonics with superlative price index formulas as hedonic superlative price indices. Our approach aligns with the Boskin Commission's recommendations both on substitution effects and on quality adjustments using hedonics. With the availability of rich scanner data, we show these recommendations are now implementable at scale with the EP methodology.

We also evaluate exact demand-based price indices alongside the hedonic superlative indices. The demand-based approach 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 item entry and exit ("Feenstra index"); 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 continuing goods in 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 for grouping entering, exiting, and continuing goods.

Our goal in this paper is to provide a framework for re-engineering a wide range of economic statistics leading to substantial improvements in how we measure the economy. Re-engineered price indices should mitigate biases from substitution, quality change, and new goods, as well as measurement and sampling errors. We consider all these goals in evaluating alternative approaches. In considering these alternatives we seek 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 because of their conceptual and measurement limitations.

An advantage of the EP hedonic methodology is that it is designed to address time-varying valuation of the relationship between prices and both observable and unobservable characteristics. The latter is of critical importance since a common concern is that hedonic adjustments require high-quality item-level attributes. When using traditional econometric methods, the EP method incorporates unobservable characteristics with a two-step estimation approach. The first step is to estimate a hedonic model of log price levels and compute the residuals, which are used as a proxy for unobserved characteristics. The second step is to estimate a hedonic model of log price changes that includes the lagged residuals from the first step to capture time-varying valuations of both observed and unobserved product characteristics. Both steps are estimated separately every period. Pakes (2003) emphasizes that period-by-period estimation is required to capture the evolution of the complex mapping between attributes and prices that reflects many factors including shocks to demand, cost, and markups.

We apply the EP hedonic methodology to two very different item-level transactions databases. The first is from the NPD group for selected narrow product groups in consumer electronics and apparel. The NPD data have the advantage of item-level prices and quantities along with highly curated item-level product attributes. We use traditional econometric methods with the NPD data. We also apply the EP methodology to the NielsenIQ Kilts item-level transactions data for food. The NielsenIQ data have price and quantity information, but attribute data are limited to abbreviated text descriptions. We use machine learning (ML) methods to apply the EP procedure to the NielsenIQ data. Natural language processing is used to encode the abbreviated text descriptions into numerical summaries. We feed these encodings into neural networks that predict log prices in the first stage and log price changes in the second stage, again using the first-stage residual as an additional predictor. This procedure has the capacity to implement hedonics at scale because it does not require encoded product attributes.

We also apply the demand-based approaches separately to the NPD and NielsenIQ data. Given its use of a CES demand structure, this approach requires estimation of the elasticity of substitution for narrow product groups for indices that take into account product turnover and quality change. We implement an estimation approach based on Feenstra's (1994) methodology. The CUPI is sensitive to the CES economic environment beyond the functional form of the utility function. We examine this issue through a "common goods rule" (CGR) with alternative parameterizations of the sets of "common" and "continuing" items used in constructing the components of the CUPI.

Across the wide range of goods this paper studies, both hedonic superlative and demand-based indices constructed from item-level transactions data show significantly lower rates of inflation than traditional matched-model indices. In the consumer goods categories we consider, the hedonic Tornqvist index runs between 0.9 and 6.2 percentage points per year lower than the traditional Laspeyres index. Especially large differences are evident for consumer technology goods. The CUPI, even implemented with a relatively strict common goods rule, runs between 3.2 and 12.4 percentage points lower per year than the Laspeyres index.

The hedonic Tornqvist indices that we compute using the EP methodology are robust to small changes in assumptions or implementation. This desirable property contrasts with the demand-based indices—especially the CUPI. The CUPI's novel feature is its treatment of time-varying product appeal, but its performance depends heavily on how "continuing" items are defined. Using a baseline definition (presence in both current and prior periods) often results in implausibly low inflation rates. Redding and Weinstein (2020) acknowledge this issue and suggest that only "established" goods should be included in the set of continuing items. We find that the CUPI's measures of inflation are highly sensitive to the parameterization of the CGR, and this sensitivity varies substantially by product group, likely due to differences in product life cycles and turnover dynamics. The sensitivity of the CUPI to the CGR limits its scalability. Other demand-based indices—especially the Feenstra-adjusted Sato-Vartia—do not require a CGR, so they remain useful benchmarks.

The paper proceeds as follows. Section I describes the NPD and NielsenIQ data infrastructures. Section II presents our conceptual framework, reviews traditional matched-model indices, and compares them with the hedonic superlative and demand-based indices. Section III presents results—first for the product groups from NPD, and then for the food product groups from NielsenIQ. Section IV concludes.

I 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.

I.A 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 item-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 item-quarter level and calculate total quantity sold and average price for each item-quarter. These data cover tens of thousands of item-quarter-level observations.

The NPD data contain detailed and well-coded information on the characteristics of each item. 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. In other cases, we create discrete variables from the brand, product type, features and attributes information. We use the detailed product characteristics in the estimation of hedonic price

[1] The NPD Group became part of Circana in 2022.

[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.

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, with quarterly entry and exit rates ranging from 4.5 percent to 13.5 percent. There is a distinction between all entry and initial entry – where the former represents any item not sold in the prior quarter but sold in the current quarter, while the latter reflects the subset of all entry limited to the first quarter the item is sold. Similar remarks apply to all exit and final exit.

Evidence on the life cycle dynamics of products is presented in Figure 1. Product entry is defined as a new item (i.e., new UPC) within a product group. The new item could represent a minor change in packaging or specifications, or a significant change in product characteristics (e.g., the entry of coffee makers with pods). Figure 1 shows the cumulative mean log change from initial entry of market share and unweighted price. Prices decline steadily after entry, while market shares exhibit a hump-shaped pattern post-entry. Memory Cards are typical: 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.

[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 an item is available, which also exhibits a hump shape. We present related evidence for food products using NielsenIQ in Appendix B.5.

I.B 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.[5] The data consist of over 2.6 million products identified by the finest level of aggregation–12-digit universal product codes (UPCs) that uniquely identify specific goods. The NielsenIQ 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 the NielsenIQ data 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 NielsenIQ data 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. Figure 2 compares 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, nominal sales indices for the NielsenIQ data tracks the PCE quite closely. That food expenditures in the NielsenIQ data closely track the BEA data is reassuring, but not surprising. The NielsenIQ data has very good coverage of grocery stores.

[5] We aggregate to the quarterly frequency using the same NRF calendar described in Section I.A.

For nonfood in the bottom panel, the scanner data exhibits less of an increase over time than PCE. This finding for the scanner data, though little noticed in the large literature using this database, is not surprising. The nonfood products sold in the stores covered by the NielsenIQ data distributed by the Kilts Center are not representative of the overall economy. The nonfood items sold, for example, at grocery stores, are highly idiosyncratic. While grocery stores sell nonfood 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 data from 2008:1 to 2015:4 to the growth in nominal sales for PCE over the same period. Nominal expenditures grew at the same rate in the NielsenIQ data and 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 PCE.

In Appendix B.3, we also show that computing an arithmetic Laspeyres index from the NielsenIQ data 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 NielsenIQ arithmetic Laspeyres. Hence, the BLS price quotations are similar to the scanner data for food. For nonfood, since the NielsenIQ data are not representative of nonfood goods sold in the economy overall, we cannot tell whether the issue is differences in BLS price quotations or representativeness of the nonfood items sold in outlets covered by NielsenIQ.

In summary, the nonfood data in the NielsenIQ data do not track the official statistics aggregates for nonfood. In contrast, for food, the evidence supports the use of the NielsenIQ data for our analysis of price indices.[6]

II Conceptual Framework

Consider time series price indices that 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 exiting 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 indices 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 characteristics.

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 II.A, we consider matched-model price

[6] In the results in the main paper, we use the NielsenIQ Retail Scanner data. 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. Furthermore, 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.

indices, that is, indices which follow goods that are available in both the base and reference period. In Sections II.B and II.C, we consider approaches that account for entering and exiting goods.

II.A 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 Ω_t denote the set of all goods sold in period t and $\mathbb{C}_t$ denote the set of goods sold both in period t-1 and in period t, so $\mathbb{C}_t \equiv \Omega_{t\text{-}1} \cap \Omega_t$. We will call the set $\mathbb{C}_t$ "continuing" or "common" goods. Letting q_{kt} be the quantity of item k purchased in period t and p_{kt} be its price, item k's share of expenditure among goods sold at time t is

$$s_{kt} \equiv \frac{p_{kt}q_{kt}}{\sum_{l\in\Omega_t} p_{lt}q_{lt}}, \tag{1}$$

where the summation is over Ω_t, the set of all goods sold in period t. It is useful to define item k's share of expenditure among continuing goods as

$$s^*_{kt} \equiv \frac{p_{kt}q_{kt}}{\sum_{l\in\mathbb{C}_t} p_{lt}q_{lt}}, \tag{2}$$

where the summation is over $\mathbb{C}_t$, the set of continuing goods from period t-1 to period t.

Now consider first the *arithmetic Laspeyres* price index,

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

that is, an arithmetic average of price relatives. The arithmetic Laspeyres 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.[7] Second, there are important theoretical results discussed below that the arithmetic Laspeyres is an upper bound on the true cost-of-living index.

While we present the arithmetic Laspeyres index for more direct comparability with the CPI, our empirical work in this paper focuses on geometric price indices, which are weighted averages of log price changes. The *geometric Laspeyres* index is defined as

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

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

$$\ln \Phi_t^{P,Geo.} \equiv \sum_{k \in \mathbb{C}_t} s^*_{kt} \ln \frac{p_{kt}}{p_{kt\text{-}1}}. \tag{5}$$

The matched-model indices by construction ignore product turnover, and as such, their properties as cost-of-living indices apply to environments without such turnover. Under these conditions, 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 changes.[8]

[7]Equation 3 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 discuss next, where the elementary price observations at the item-area level in the CPI are equally weighted geometric averages of price quotations of the selected products 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. Because the selection probabilities are based on in-store expenditure shares, the BLS procedure for the low-level indices is implicitly weighted based on lagged expenditure shares.

[8]These 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 Layspeyres. Note further that if utility is Cobb-Douglas, then expenditure shares will be constant, the Laspeyres and Paasche indices will be equal, and they will be exact

To address substitution effects in traditional price indices, we focus in particular on the *Tornqvist* index, given by

$$\ln \Phi_t^{TQ} \equiv \sum_{k \in \mathbb{C}_t} \left(\frac{s_{kt\text{-}1}^* + s_{kt}^*}{2} \right) \ln \frac{p_{kt}}{p_{kt\text{-}1}}, \tag{6}$$

which is a weighted average of the same log price changes as the Laspeyres and Paasche, but with weights that 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 cost-of-living index. 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) in the absence of product turnover.[9] Second, as a superlative price index, the Tornqvist is 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 in real time. Economists have long called for using superlative price indices, for example, in the recommendations of the Boskin Commission (see Boskin et al., 1998). These recommendations have not been

indices.

[9]Under certain assumptions, a 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 continuously differentiable preferences (the latter requires no product turnover). 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.

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. Examples 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 Federal Open Market Committee'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.[10]

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. The use of matched-model indices is not a desideratum, but rather owes to the fact that the price changes in the formulas given above require goods to be observed in adjacent periods. Hence, the indices discussed in Section II.A 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 as it is ubiquitous in item-level data.

[10]There is the challenge of chain drift in high frequency chained price indices. Chain drift occurs when price indices cumulated (*chained*) over multiple periods do not return to the same index value between a base period and a subsequent period even if relative prices have returned to base period values (e.g., Diewert, 2021). This issue is more likely to arise in the presence of transitory price changes, with accompanying changes in expenditure shares that may not return to original values. We discuss chain drift in detail in Section III.E.

Item Replacement in Matched-Model Indicies–. Statistical agencies currently address product turnover in limited ways. The BLS selects a *sample* of goods to be priced, and then prices the sampled item for a period of time. The price of a sampled item in the CPI is tracked for up to four years. When an item disappears, analysts choose among three methods to address it: comparable item replacement, non-comparable item replacement, or (less frequently) quality adjustment. If the analyst selects an item that is comparable (deemed to have essentially the same quality), any price change is counted as inflation or deflation. If the item replacement is judged non-comparable, the price change is *linked out*, so none of the price change is reflected in the price index (that is, the changes in the item priced are implicitly assumed to be quality change). In a limited number of cases, BLS analysts adjust the new good for changes in quality. Quality adjustments may use hedonic methods (applied to about 7.5% of goods, mostly apparel and tech), or in rare cases, cost-based adjustments (e.g., when a car option becomes standard). BLS also makes adjustments for package size or item count. These *ad hoc* processes do not scale to millions of items in transactions data. Also, the BLS applies them only when an item exits during its four-year sample window. Hence, importantly, when an item is replaced as scheduled after four years, the price change is always linked out of the price index.

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. In principle, both of these methods overcome the limitations of the standard methods of item replacement, as the universe of goods is tracked with product turnover and quality change fully incorporated into the price indices.

II.B Hedonic Price Indices

A hedonic price index uses imputation 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. Symmetric arguments establish that the hedonic Paasche is a lower bound on the true cost of living. Both the argument in Pakes (2003) for the Laspeyres index and the symmetric argument for the Paasche index abstract from estimation concerns; i.e., they assume knowledge of the true hedonic pricing function. 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).

We advocate using the hedonic Tornqvist index for measuring price change using scanner data because it has multiple desirable properties. The hedonic Tornqvist is the mean of the hedonic geometric Laspeyres (an upper bound on the cost of living) and the hedonic geometric Paasche (a lower bound). The Tornqvist is a *symmetric index* because it gives equal importance to the prices and expenditure data in both adjacent periods.[11] Moreover, Crawford and Neary (2019) highlight that a hedonic price index can be interpreted as a characteristics-based price index. From this perspective, they extend the second-order approximation results of (Diewert, 1976) to the hedonic Tornqvist under the assumption that product entry and exit occur without corresponding entry and exit of characteristics. Under these conditions, the hedonic Tornqvist can be interpreted as a second-order approximation to the true cost-of-living index.[12] We acknowledge that this interpretation of the hedonic Tornqvist entails more stringent assumptions than advocating for the use of a hedonic Tornqvist as the average of Könus-type upper and lower bounds, but we regard these motivations

[11]Because the geometric price indices are defined in logs, the Tornqvist—defined as the simple average of the Laspeyres and Paasche—weights the price relatives by the moving average of the expenditure shares, i.e., Divisia weights. Because it is the average of an upper and lower bound, the hedonic Tornqvist has been widely advocated in the literature (see, e.g., Diewert, Heravi and Silver 2008, and Hill and Melser 2008).

[12]That is, turnover that involves compositional shifts in the distribution of continuing characteristics rather than the introduction or disappearance of entirely new attributes. Homotheticity and continuous differentiability of preferences with respect to characteristics are also required.

as complementary.

Beyond the stricter theoretical assumptions for this interpretation, in practice we must estimate the hedonic function in the presence of unobservable characteristics. While the Erickson and Pakes (2011) methodology is well suited for this purpose, as discussed below it has potential limitations. We explore these limitations with a number of robustness analyses. We also report the hedonic Laspeyres compared to the matched-model Laspeyres to provide more guidance about the hedonic Laspeyres being a tighter upper bound than the matched-model Laspeyres.

Defining the Hedonic Indices–. Hedonic versions of the matched-model indices (4)–(6) are a general solution to dealing with the high rate of item turnover in scanner data. By imputing the price changes of goods that disappear and appear, hedonic indices allow using *all* prices, not just the prices of goods that are continuing.

Consider the *hedonic Laspeyres* index, defined as

$$\ln \Phi_t^{L,H} \equiv \sum_{k \in \Omega_{t-1}} s_{kt\text{-}1} \widehat{\ln \frac{p_{kt}}{p_{kt\text{-}1}}}, \tag{7}$$

where the summation is over the set of goods sold at time t-1, $\Omega_{t\text{-}1}$, that is, the continuing goods plus the exiting goods at time t. Similarly, the *hedonic Paasche* is defined as

$$\ln \Phi_t^{P,H} \equiv \sum_{k \in \Omega_t} s_{kt} \widehat{\ln \frac{p_{kt}}{p_{kt\text{-}1}}}, \tag{8}$$

where the summation is over the set of goods sold at time t, Ω_t, that is, the continuing goods plus the entering goods at time t.

The *hedonic Tornqvist* index is then defined as the average of the hedonic Laspeyres and hedonic Paasche,

$$\ln \Phi_t^{TQ,H} \equiv \sum_{k \in \{\Omega_{t-1} \cup \Omega_t\}} \left(\frac{s_{kt\text{-}1} + s_{kt}}{2} \right) \widehat{\ln \frac{p_{kt}}{p_{kt\text{-}1}}}, \tag{9}$$

where the summation is over the union of $\Omega_{t\text{-}1}$ and Ω_t, that is, the set of continuing, entering, and exiting goods.

Note several features of equations (7)–(9), 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 relevant observations for the index in question (including the continuing goods for which actual observations are available). Third, the Divisia index of the shares in the hedonic Tornqvist index 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). Fourth, the hedonic indices use expenditure weights defined over all goods sold in either period t-1 or period t, i.e., $s_{kt\text{-}1}$ and s_t, rather than weights defined only over continuing goods ($s^*_{kt\text{-}1}$ and s^*_t), as in the matched-model indices.

Note that "full imputation"—in which quality adjustment for continuing goods is combined with imputation for entering and exiting goods—is intimately linked with the principle that the hedonic model should be estimated period-by-period. The valuation of quality change of continuing goods cannot be separated economically from entry and exit. The value of entering, exiting, and continuing goods is jointly determined and, indeed, as we saw in Section I.A, is closely related to the lifecycle of products. New and better goods affect the prices and market shares of continuing goods and the market value of their attributes. Hence, quality adjustment to account for product turnover and quality adjustment of continuing goods are jointly determined in market equilibrium and cannot be readily separated. Relatedly, as Pakes (2003) has stressed, the estimated hedonic coefficients are generally not the consumer valuations of attributes, but rather equilibrium outcomes.

Approaches to Hedonic Estimation–. The log-level hedonic price model common in the literature takes the form

$$\ln p_{kt} = h_t(Z_k) + \eta_{kt}, \tag{10}$$

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 ordinary or weighted least squares regression. An important feature of equation (10) 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).

A core limitation of the log-level hedonic estimation approach outlined in equation (10) 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 element of this approach is to estimate hedonic models of price changes rather than price levels

$$\Delta \ln p_{kt} = Z_k' \beta_t + v_{kt}. \tag{11}$$

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 (10) for period t-1. Second, estimate a log-difference hedonic model including the lagged residuals from the first stage, which we denote

$\hat{\eta}_{kt\text{-}1}$.[13] The second estimating equation is then

$$\Delta \ln p_{kt} = Z_k' \beta_t + \kappa_t \hat{\eta}_{kt\text{-}1} + \upsilon_{kt}. \tag{12}$$

Including the initial lagged residuals $\hat{\eta}_{kt\text{-}1}$ from equation (10) in equation (12) 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 using weighted least squares with NPD data 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 are better measured than those based on a small number of purchases. More generally, if the underlying prices for each individual transaction are noisy, using quantity weights increases the efficiency of the estimator.

Accounting for Unobserved Quality of Entering Items—. The initial lagged residual for an entering item cannot be recovered from equation (10) 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

[13] 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 (12).

main analysis using the EP-TV approach, we set the sales weighted initial lagged residual for entering goods in the period prior to entry to its mean of zero. We consider alternative approaches in our analysis of NPD data. First, we impute the unmeasured initial lagged residual for an entering item to be equal to the observed current period residual for the entering item (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 items 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. We also consider an alternative in which we drop the item in the index in the first quarter of its entry.

Interpreting Hedonic Indices—. The hedonic bounding arguments above assume that the true hedonic pricing function was known, whereas in practice it must be estimated. Erickson and Pakes (2011) show that their method preserves the upper bound properties of the Laspeyres as long as the projection of exiting goods price relatives from continuing goods is greater than or equal to the expected value of the true hedonic price relative for exiting goods. They argue that this is a reasonable assumption because the true valuation of the unobserved characteristics for exiting goods likely is declining faster than that for continuing goods. The analogous condition for entering goods is that the projection of entering goods price relatives from continuing goods is less than or equal to the expected value of the true price relative of entering goods. In other words, if entering goods have unobservable characteristics with valuations that are improving more quickly than for continuing goods, the estimated hedonic Paasche will be below the true hedonic Paasche, preserving the bounding arguments. Our view is that the Erickson and Pakes (2011) approach provides a much more robust approach to addressing the potential role of unobservable factors than standard log-level hedonic models. We conduct a number of exercises below that support this view.

As already noted, we use full-imputation price indices, which can be interpreted as characteristic price indices (Hill and Melser, 2008; De Haan, 2008).[14] 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, Heravi and Silver (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 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, Sichel and Aizcorbe, 2019) using adjacent-period, weighted least

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

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

$$\ln p_{k\tau} = \alpha_{t\text{-}1,t} + \delta_t + Z_k' \gamma_{t\text{-}1,t} + \epsilon_{k\tau}, \quad \tau = \{t\text{-}1, t\}, \tag{13}$$

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, δ_t is a fixed effect for period t, and τ indexes the two periods in the pooled regression, t-1 and t.[15] 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 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 implementation of the hedonic specifications (log level, EP-F, EP-TV, and time dummy) with the NPD data use weighted least squares with the vector of observable characteristics Z_k based on the product attributes described in section I.A. This approach is feasible with the NPD data because of the enormous value-added the NPD group provides by encoding item-level attributes.

Machine Learning Hedonics–. To implement hedonics with the NielsenIQ data we need to overcome the challenge that unlike the NPD data the NielsenIQ data do not contain product attribute data for most products aside from short textual product descriptions.[16] To illustrate this point, consider two product descriptions for soft drinks: `ZR DT LN/LM CF`

[15] 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 (13) 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.

[16] The hedonic Tornqvist index for the NielsenIQ food products that we present in this paper is from Cafarella et al. (2023). See that paper, which is included in the supplementary files, for development of the algorithm summarized in this section. Appendix C provides detailed specific steps of the algorithm. In related work, Bajari et al. (2021) use an advanced machine learning approach that includes encoding image data as inputs into price predictions. They estimate hedonic models of price levels period by period—not using the EP-TV method that allows for unobserved characteristics.

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).

For the NielsenIQ data, we encode the abbreviated text descriptions and map these encoded characteristics to prices using a neural net machine learning approach that parallels the EP econometric model. 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 Retail Scanner data at the Kilts Center. 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 into deciles. Third, the ML system minimizes the weighted cross-entropy loss function for the products' true price and price change distributions in the hedonic estimation. In this application, the cross-entropy loss objective function is equivalent to maximizing the likelihood of assigning the highest probability to the correct bin. 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.

Using this algorithm, 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 NPD data to estimate the hedonic functions for log price levels and changes in equations (10) and (12).

The ML procedure thus enables the implementation of hedonic superlative indices in transactions data at scale. The traditional econometric approach we implement with the NPD data relies on having well-encoded product attributes, which ultimately were produced by human analysts. The approach also required significant human judgment and attention. At-scale implementation of hedonic superlative indices in the statistical system cannot rely on the presence of well-encoded attributes or substantial human judgment. The ML approach is agnostic regarding the structure or presence of well-encoded attribute data and does not require human judgment to specify sensible hedonic regressions across large range of product groups.

II.C 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

$$P_t = \left[\sum_{k \in \Omega_t} \left(\frac{p_{kt}}{\varphi_{kt}} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}, \tag{14}$$

where $\sigma > 1$ is the consumer's elasticity of substitution between products, and φ_{kt} is an appeal parameter for item k. Both the set of products sold Ω_t and product-level appeal φ_{kt} may vary over time. Equation (14) 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 (14), 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

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

where

$$\omega_{kt} \equiv \frac{s^*_{kt} - s^*_{kt\text{-}1}}{\ln(s^*_{kt}) - \ln(s^*_{kt\text{-}1})} \Bigg/ \left(\sum_{k \in \mathbb{C}_t} \frac{s^*_{kt} - s^*_{kt\text{-}1}}{\ln(s^*_{kt}) - \ln(s^*_{kt\text{-}1})}\right). \tag{16}$$

The Feenstra (1994) index generalizes the Sato-Vartia index to account for turnover in the set of goods sold Ω_t. Define the terms $\lambda_{t,t-1}$ and $\lambda_{t-1,t}$ as

$$\lambda_{t,t-1} \equiv \frac{\sum_{k \in \mathbb{C}_t} p_{kt} q_{kt}}{\sum_{k \in \Omega_t} p_{kt} q_{kt}}, \qquad \lambda_{t-1,t} \equiv \frac{\sum_{k \in \mathbb{C}_t} p_{kt-1} q_{kt-1}}{\sum_{k \in \Omega_{t-1}} p_{kt-1} q_{kt-1}}. \tag{17}$$

The log Feenstra index is then defined as

$$\ln \Phi^{Feenstra}_{t\text{-}1,t} \equiv \frac{1}{\sigma - 1} \ln\left(\frac{\lambda_{t,t-1}}{\lambda_{t\text{-}1,t}}\right) + \ln \Phi^{SV}_{t\text{-}1,t}. \tag{18}$$

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 (18), in the absence of product turnover. 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 over across product groups.

The CUPI in logs is given by

$$\ln \Phi_{t\text{-}1,t}^{CUPI} \equiv \frac{1}{\sigma-1} \ln\left(\frac{\lambda_{t,t-1}}{\lambda_{t\text{-}1,t}}\right) + \ln \Phi_{t\text{-}1,t}^{SV} - \sum_{k \in \mathbb{C}_t} \omega_{kt} \ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right) \tag{19}$$

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

$$\begin{aligned} \ln \Phi_{t-1,t}^{CUPI} &= \frac{1}{\sigma-1} \ln\left(\frac{\lambda_{t,t-1}}{\lambda_{t-1,t}}\right) + \frac{1}{N_{\mathbb{C}_t}} \sum_{k \in \mathbb{C}_t} \ln\left(\frac{p_{kt}}{p_{kt-1}}\right) + \frac{1}{\sigma-1} \frac{1}{N_{\mathbb{C}_t}} \sum_{k \in \mathbb{C}_t} \ln\left(\frac{s_{kt}^*}{s_{kt-1}^*}\right) \\ &\equiv \quad \text{Lambda Ratio} \quad + \quad P^* \text{ Ratio} \quad + \quad S^* \text{ Ratio} \end{aligned} \tag{20}$$

where the variables defined over the set of common goods are denoted by an asterisk. Equation (20) 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.[17] 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.

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 (20) are unweighted indices. The Sato-Vartia and the Feenstra, in contrast, include only expenditure-weighted 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.[18]

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. The CGR does not change the structure of the CUPI defined in equation (20). Instead, it merely changes how goods are categorized between the sets of continuing and non-continuing goods. 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

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

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

duration both prior to period t-1 and subsequent to period t.[19] A limitation of this duration-based common goods rule is that it requires forward-looking information to implement, so 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 D.2.

III Results

We present and discuss the traditional matched-model, hedonic, and demand-based exact price indices we calculate in the item-level data. In this section, we focus first on our results from the NPD data, because the richness of the data permits more exploration of alternative methods. In subsequent sections, we consider results for NielsenIQ data and consider the issue of chain drift.

III.A Hedonics: NPD Data

We consider a wide variety of hedonic specifications as discussed in Section II.B. 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

[19] 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 item be sold for at least 6 years total (although not necessarily consecutively).

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.[20]

In the main text, we focus on the results using the EP-TV estimation approach. In 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 (-3.2 percentage points annually) and headphones (-2.6), and smaller differences for coffee makers (-0.71), boys' jeans (-1.53), and occupational footwear (-0.44).[21] 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

[20] 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.

[21] The quality change estimate of -1.53 percentage points annually for boys' jeans may seem surprising, as BLS finds that hedonic adjustments in apparel item replacement generally raise measured inflation (Brown and Stockburger, 2006). This pattern arises because BLS prevents linking out price increases when items return at the start of seasonal cycles. Our hedonic Tornqvist method fully accounts for product turnover, including returning items in the index. However, as discussed below, boys' jeans is the only product group for which the hedonic Tornqvist exhibits substantial chain drift.

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.[22] We confirm the expected ordering for matched-model indices: Arithmetic Laspeyres > Geometric Laspeyres > Tornqvist. The pattern that the geometric Laspeyres indicates lower inflation than the arithmetic Laspeyres reflects the tendency for geometric indices to mitigate substitution biases. Likewise, the Tornqvist generally indicating lower inflation than the geometric Laspeyres reflects the Tornqvist's accounting for item-level substitution as the average of the geometric Laspeyres and Paasche.

In all cases, we find that the corresponding hedonic version of each of these indices is substantially lower than the traditional version, while maintaining the same order.[23] This pattern is consistent with the intuition that accounting for product turnover in the presence of quality change should reduce measured inflation. The hedonic Laspeyres includes the contribution of exiting products, while the hedonic Tornqvist the contributions both of entering and of exiting products.[24]

For entering goods in the hedonic Tornqvist, we examine two alternative methods for treating the *unmeasured* initial lagged residual. Table 3 shows 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

[22]We do not report the matched model or hedonic Paasche indices, but their patterns can be inferred from the Laspeyres and Tornqvist indices.

[23]For headphones, the hedonic Tornqvist is lower than the Tornqvist and the hedonic geometric Laspeyres is lower than the (geometric) Laspeyres but the hedonic Tornqvist is slightly above the hedonic Laspeyres.

[24]Note that, because the hedonic Laspeyres does not account for entering goods, the EP-TV Laspeyres results do not depend on the method for measuring the lagged residual for entrants.

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 (and in the hedonic Tornqvist the entrants' share weight is divided by two). Over longer horizons (e.g., direct year-over-year [YOY]) the sales share of entrants over a four-quarter horizon is substantially larger.[25] The second panel of Table 3 shows more sensitivity to alternative methods over this longer horizon, but the patterns remain quite similar.[26]

As an additional robustness check, we estimate a specification that excludes product entrants during their first quarter of market entry. The resulting EP-TV average annual chained inflation estimates for the adjusted hedonic Tornqvist index are –20.0% for memory cards, –9.5% for coffeemakers, –13.8% for headphones, –9.0% for boys' jeans, and –3.8% for occupational footwear. Relative to the estimated inflation rates in Table 3, these values are generally somewhat higher, indicating that including product entrants in their initial quarter tends to lower the measured inflation rate. The magnitude of the difference is modest, which is consistent with the relatively small expenditure shares for entering products during their first quarter.

We also examine whether the hedonic estimates are sensitive to differences in the range

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

[26]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.

of values of characteristics of entering or exiting goods relative to continuing goods.[27] We conduct this robustness check for memory cards because their two most important 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 -20.1%, the same as the baseline estimate. 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 indices, which, unlike with the data they had available, we can compute because the scanner data contain both prices and quantities.

III.B CES Demand-Based Price Indices: NPD Data

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.[28] Appendix Table D.7 reports the estimated elasticities from about 5.2 to 7.8, which

[27] 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.

[28] 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.

is consistent with the literature.

Table 2 reports the price index in 2018:4 relative to 2014:4 = 1 for the CES demand-based indices as well as the matched model and hedonic indices. The Sato-Vartia index is generally similar to the traditional Tornqvist index, which is intuitive because both are superlative indices. The Feenstra index is uniformly lower than the Sato-Vartia, reinforcing our finding from comparing the hedonic and matched model indices that accounting for quality change via product turnover reduces measured inflation. Finally, both the non-nested and nested CUPI show substantially lower inflation than the other indices. In this section, we examine the sources of these differences.

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 common goods rule and without any nesting within product groups. The Lambda Ratio and S^* ratio components in the figure reflect scaling by $\frac{1}{\hat{\sigma}-1}$ so that the CUPI is the sum of the three components (see equation (20)). We find that the CUPI shows lower inflation than the Feenstra index and very low inflation in absolute terms. In all goods except occupational footwear, the CUPI produces an estimate of 20 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 lagged 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 P^* and S^* ratio terms of the CUPI.[29] 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).[30]

Figure 6 illustrates the CUPI's sensitivity to the CGR. We show these patterns alongside the Sato-Vartia index and the Feenstra index. Implementing the restriction on the set of common goods by market share raises the CUPI by moving the products in the low end of the market share distribution from the unweighted components to the weighted entry/exit adjustment term. Applying a 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 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 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 with a 30th-percentile CGR threshold. As shown in Appendix Figure D.5, we find that even with a 50th percentile threshold (i.e., shifting half of products from the set of common goods to entering and exiting goods) the CUPI for headphones and boys' jeans yields considerably lower inflation.

In contrast to the finding in Redding and Weinstein (2020) that the CUPI eventually

[29] 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.

[30] 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.

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. We are sympathetic to the view that incorporating some form of a CGR is both sensible and necessary for empirically implementing the CUPI. Our main inference from the NPD analysis is that CUPI results are sensitive to the CGR specification, with the degree of sensitivity varying across product groups. This highlights the need for further research into open issues for implementing the index. In particular, understanding the dynamics of product entry and exit should be a key focus. Our analysis in Figure 1, which documents lifecycle patterns in prices and market shares, takes a step in this direction. These patterns vary by product group, consistent with our finding that the CUPI's sensitivity to CGRs differs across groups in the NPD data.

Interpreting "Product Appeal"–. A central insight of the CUPI is that product appeal shocks are required to reconcile observed patterns of prices and expenditures. The theoretical framework implies that products with low prices and low expenditures must have low appeal. In practice, factors outside the CUPI model—especially concerning product entry and exit—can generate low expenditures independent of price and appeal. For instance, close-out sales of discontinued products (e.g., clearance racks) may offer attractive prices, but limited consumer awareness or availability may lead to low expenditure despite low price. In such cases, search frictions and limited availability, not low appeal, explain low expenditure shares.

In Appendix D, we discuss simulations featuring some of these factors. We distinguish between the conceptually distinct issues of clearance sales and product rationing near the end of the product lifecycle. End-of-lifecycle discounting does not cause bias on its own; rather, rationing is necessary to break the link between prices, expenditure shares, and product appeal implied by the simple CES model. Boys Jeans are potentially subject to such rationing with clearance sales. The chained CUPI with a 30 percent common goods rule measures deflation in Boys Jeans of nearly 35 percent per year, a deflation rate more than 25 percentage points lower than any other price index reported in table 4 including the hedonic Tornqvist index. This pattern is consistent with our finding in Appendix D that the

CUPI is especially sensitive to clearance sales.

The NPD data also allow exploration of variation in the CUPI's taste shock component. One hypothesis we examine is whether the large measured taste shock bias (without a CGR) reflects the need for a richer, nested CES structure. Within product groups, subgroups may exhibit higher elasticities of substitution than between groups, damping the measured contribution of the S^* ratio. We explore this by nesting products using two approaches: one based on observable characteristics (e.g., brands), and another based on predicted prices from hedonic regressions. Appendix D.1 provides details. Neither approach substantially alters CUPI inflation measures, suggesting that aggregation structure is not the primary driver of the observed taste shock bias in the NPD groups.[31]

Another mechanism for the large positive taste shock bias is increasing dispersion in relative product appeal over time.[32] In Appendix D.2.4, we examine the evolution of product appeal dispersion across NPD groups. We find evidence of increasing appeal dispersion for headphones and memory cards (especially without a CGR), but not for other groups, including some with high measured taste shock bias. Thus, the need for a CGR that varies by product group appears distinct from changes in product appeal dispersion.

III.C Comparing Matched-Model, Hedonic, and Exact Price Indices: NPD Data

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 as an approximation to the BLS's CPI

[31] Martin (2020) also suggests the importance of a nested preference structure in implementing the CUPI.

[32] Increasing dispersion in relative product appeal over time introduces a positive taste-shock bias because consumers have greater opportunities for substitution when appeal is more dispersed. This effect is analogous to the well-known result that consumers benefit from more dispersed prices all else equal. Redding and Weinstein (2020) highlight this possibility, and we present supporting simulation evidence in Appendix D.2.3.1.

formula.

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. As discussed above, 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 arithmetic 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 at all levels of aggregation. The cumulative gaps are on the order of 6–12 percentage points from 2014 to 2018 in coffee makers, occupational footwear, and boys' jeans, and significantly larger in the 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 lower cumulative inflation relative to 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 for other categories, especially for memory cards and headphones (about 7 percentage points).

Our estimated adjustments are similar in average magnitude to those in Bils (2009) and Bils and Klenow (2001).[33] Bils and Klenow (2001) and Bils (2009) use methods designed to measure average quality change. Bils and Klenow (2001) estimate average quality change based on cross-section variation of income of consumers, and Bils (2009) considers price-entry dynamics. Our approach is designed to give a period-by-period quality adjusted price index by detailed product group, i.e., what is needed to implement quality adjustment at scale in real time for price indices. The similarity of our hedonic estimates to these distinct methods of quantifying the average impact of quality change on price inflation is reassuring. In contrast, the average quality adjustments implied by the CUPI are much larger and inconsistent with this literature.

III.D NielsenIQ Results

In this section, we present results comparing matched-model indices, the hedonic Tornqvist, and demand-based indices for NielsenIQ food products. As discussed above, nonfood items in the NielsenIQ data are not representative of nonfood items in the overall economy, so we do not present them in the main text.[34]

The hedonic Tornqvist index presented in this section is from Cafarella et al. (2023), as described above. Notwithstanding the lack of encoded product attributes in the NielsenIQ data distributed through the Kilts Center, the algorithm provides reasonably fitting predictions of price changes that are necessary for inputting the prices of entering and exiting items. 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

[33] Our hedonic adjustments range from 0.44 percent per year for Occupational Footwear to 3.2 percent for Memory Cards. Bils (2009) estimates that "...CPI inflation for durables has been overstated by almost 2 percentage points per year..." and Bils and Klenow (2001) estimate "...about 2.2-percent upward bias in BLS inflation for our consumer durables because of failure to fully net out quality growth."

[34] The appendix contains results that include nonfood items in Appendix Figures D.7 and D.13–D.15.

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.

For the CES exact price indices, our empirical implementation in the NielsenIQ data largely follows our procedure for 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. For our core results, we implement the same approach as in the NPD data, but we also consider an alternative by adding a duration-based rule as an additional restriction (i.e., we add the restriction that a good is a common good 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 price indices for the NielsenIQ food product groups in levels.[35] For all indices, we aggregate across product groups using a Tornqvist aggregator with Divisia-style product group market share weights. The figure also includes the BLS CPI computed for the same NielsenIQ product groups.[36] We find that the CPI and the matched-model Laspeyres

[35] Appendix Figure D.11 shows results for price changes. Appendix Figure D.8 shows results for alternative CGRs parallel to the results for NPD products. In contrast to the NPD products, for the NielsenIQ food products, the CGR affects the S^* ratio substantially, but has little effect on the P^* ratio.

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

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.[37]

Using a 30th-percentile CGR, the CUPI (CUPI, 30p) yields more than 25 percentage points lower cumulative inflation than the Feenstra index in 2015:4 (with base period 2006:4). This large difference reflects about a 3 percentage point annual inflation rate difference cumulated over 9 years. Adding a duration component to the CGR implies that a smaller percentile threshold is needed to achieve a similar pattern for CUPI. With a 4 quarter duration requirement for common goods, only a 10th percentile market share threshold is needed to achieve about the same inflation pattern (CUPI 10p(5q)). Using a 50th-percentile CGR (CUPI, 50p), the CUPI yields about 15 percentage points lower cumulative inflation than the Feenstra index in 2015:4.[38] This analysis of the CUPI's sensitivity to the CGR does not allow the CGR to vary across the almost 60 product groups in food. Yet our analysis of the NPD data suggests that the sensitivity of the CUPI to the CGR is product-group specific. Developing an approach for product-group specific CGRs is an open area for future research that would enhance the scalability of the CUPI.

The hedonic Tornqvist yields cumulative inflation that is about 3 percentage points lower in 2015 than the matched-model Tornqvist, and the Feenstra index about 4 percentage points lower than the Sato-Vartia. The hedonic Tornqvist cumulative inflation is only about two-thirds of the Tornqvist and the Feenstra only about three-fifths of the Sato Vartia.

[37] Broda and Weinstein (2010) find a similar rank ordering and magnitudes for the non-hedonic matched-model indices using NielsenIQ Consumer Panel data.

[38] 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 because it is the data used by Redding and Weinstein (2020). The results are broadly consistent. 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 the set of common goods to those that were present four quarters apart, effectively imposing a duration component on the CGR. We find in our analysis of chain drift below that the implied YOY CUPI index for food in Table 5 yields substantially higher inflation than the chained CUPI, which is consistent with the idea that the YOY index imposes a meaningful additional restriction on the set of common goods.

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. The hedonic Tornqvist yields cumulative inflation that is about 1 percentage point lower that the Feenstra.

The patterns in the NielsenIQ data are 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 the interpretation of this finding is dependent on the sensitivity of the CUPI to the CGR. This sensitivity manifests across alternative approaches to defining the CGR thresholds for common goods.

III.E Chain Drift

An advantage of chained price indices constructed from transactions data is their timely incorporation of entering and exiting products. A potential drawback is chain drift, which can result from high-frequency transitory price volatility. This issue is particularly pronounced in indices computed from transactions data in a specific local area, where prices tend to be more volatile (e.g., De Haan and Van Der Grient, 2011). Our analysis uses national data at a quarterly frequency, which mitigates these concerns. Nonetheless, we examine the importance of chain drift in our data. Given our focus on comparing alternative price index methodologies, we examine whether GEKS-type indices (Gini, 1931; Eltetö and Köves, 1964; Sculz, 1964)—which take the geometric mean of bilateral indices over multiple horizons—preserve the core implications of our findings.

Following Bajari et al. (2021), we compute a GEKS-type index (denoted "GEKS-lite") as the geometric mean of the chained year-over-year index and the directly computed (unchained) year-over-year index for the fourth quarter. GEKS-lite offers a computationally feasible alternative to a full GEKS procedure, which requires estimation of price indices across many time horizons. This is particularly relevant in our setting, where hedonic indices must be re-estimated for each comparison period.

Table 4 reports average annual chained and GEKS-lite indices for five NPD product groups and across index types. For traditional matched-model indices, GEKS-lite typically shows less deflation than the chained index. For hedonic indices, GEKS-lite shows greater deflation in three of five product groups. Table 3 shows these results are robust to alternative methods for imputing missing lagged residuals for entrants. For demand-based indices, GEKS-lite generally shows greater deflation, though differences are quantitatively modest. For the CUPI, GEKS-lite indicates substantially less deflation, largely due to the longer horizon over which the common goods rule is applied.

The key result is that applying GEKS-lite does not change the rank ordering of index measures. The Laspeyres matched-model index shows more inflation than the matched-model Tornqvist, which exceeds the hedonic Tornqvist. Similarly, the Sato-Vartia exceeds the Feenstra, which in turn exceeds the CUPI. Notably, the matched-model Tornqvist is more sensitive to chain drift than the hedonic Tornqvist. This reduced sensitivity is intuitive, as hedonic indices are less affected by transitory price volatility, the main source of chain drift.

We also compute the rolling-year GEKS index (Ivancic, Diewert and Fox, 2011) as a benchmark. Due to computational intensity, we limit this to matched-model indices. Appendix Table D.8 shows results that are similar to those from GEKS-lite.

Table 5 presents analogous chained and GEKS-lite indices for aggregated NielsenIQ food product groups. The table reports average annual rates under both specifications. As with the NPD data, GEKS-lite generally shows higher inflation than the chained index, especially for the CUPI—again reflecting the impact of applying the CGR over a longer horizon. Importantly, the rank ordering and relative magnitudes of inflation across indices remain consistent under GEKS-lite. For NielsenIQ food products, the Laspeyres exceeds the Tornqvist, the Sato-Vartia exceeds the Feenstra, and the Feenstra exceeds the CUPI.[39]

[39] We do not report GEKS-lite hedonic indices for the NielsenIQ food groups due to the high computational cost of applying our machine learning framework to additional periods.

In this context, it is worth emphasizing the value of high-frequency chained price indices that incorporate product quality and turnover. Policymakers, businesses, and households rely on high-frequency indices for a range of decisions. While year-over-year indices are informative, they miss key aspects of product turnover. This limitation is especially acute for matched-model indices, but also affects hedonic indices, which must increasingly rely on imputed price relatives for entrants and exits based on longer-horizon estimates from continuing products.

IV Conclusion

This paper addresses a substantial challenge of using item-level transaction data to construct price indices. The availability of these data in the 21st century information economy permits long-sought innovations in price measurement (Boskin et al., 1998). Yet, the use of item-level transactions data presents considerable challenges. Measuring inflation requires longitudinally-consistent measures of prices in order to compute price changes. In item-level transactions data, one needs to confront that turnover of items is frequent and ubiquitous. Moreover, this turnover is not just a measurement problem. The introduction of new items often includes both quality changes and price changes. Exit of products is often driven by competition from newer products. And the prices and sales of all products—entering, exiting, and continuing—are determined jointly in equilibrium. Hence, the measurement framework for building price indices from transactions data needs to account for quality change for all goods, not just goods experiencing significant technical progress. This paper evaluates two frameworks for building price indices from item-level transactions data—hedonic superlative and demand-based indices. Both frameworks leverage the ability of indices from item-level data to reflect both substitution effects and quality change.

We find that the hedonic Tornqvist index—a member of the family of hedonic superlative indices—can be implemented at scale to account for quality change as well as consumer substitution in response to relative price changes. 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 infrastructures and estimation approaches—econometric and machine learning.

A systematic pattern emerges: hedonic indices yield significantly lower rates of inflation than associated matched-model indices. These patterns hold both for Laspeyres indices, which provide bounds for the cost-of-living index under very general conditions, and for the Tornqvist superlative index, which is a mid-point of upper and lower bounds and has additional desirable properties. 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. Our 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 CUPI results in large downward adjustments to the cost of living relative to both the hedonic Tornqvist and the Feenstra indices. The size of these adjustments arise because the CUPI strongly relies on the CES economic environment—not just the CES functional form. Specifically, its empirical implementation includes two equally-weighted geometric indices that are jointly determined under strong economic assumptions. These indices are particularly sensitive to small expenditure shares. If these small shares arise because of limited availability of products rather than demand of the representative agent, as is likely to be the case when products have limited availability at the beginning and end of their lifecycle, the CUPI will over-adjust for product appeal because it presumes that low demand arises because of high price or low appeal, not limited availability. For this reason, the CUPI is sensitive to interrelated dynamics of price and expenditure associated with the entry and exit of products. This paper establishes the empirical importance of this feature of the CUPI by demonstrating its sensitivity to the parameterization of the Common Goods Rule (CGR), which controls how entering and exiting products are distributed across the components of the CUPI.

We have found that the 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. Further research is needed to establish best practices for implementing the CGR in the CUPI on a product-group specific basis. Although this challenge, along with the strong assumptions underlying the index, constrains the scalability of the CUPI, demand-based indices—particularly the Feenstra-adjusted Sato-Vartia—continue to serve as valuable benchmarks.

Although this paper focuses on consumer goods, we view this focus as a point of departure rather than a limitation. The retail goods sector is a natural starting point due to the availability of scanner data and item-level transaction records. Many consumer services—such as utilities, telecommunications, transportation, and health care—are also mediated through electronic records with similar granularity. The same is true for many business-to-business and government transactions. While institutional and logistical challenges to using these data for official statistics remain significant, the methods evaluated in this paper are not limited to retail goods and could, in principle, be applied widely across the economy.

This paper illustrates that large-scale use of item-level transactions data can support a re-engineering of key national economic indicators. Accounting simultaneously for substitution and quality change lowers measured inflation rates substantially across a broad range of goods. These methodological advances have important implications for measuring inflation, and consequently for estimates of output and productivity growth.

References

Bajari, Patrick, Zhihao Cen, Victor Chernozhukov, Manoj Manukonda, Jin Wang, Ramon Huerta, Junbo Li, Ling Leng, George George Monokroussos, Suhas Vijaykunar, and Shan Wan. 2021. "Hedonic Prices and Quality Adjusted Price Indices Powered by AI." CENMAP Working Paper 04-21; forthcoming, *Journal of Econometrics.*

Barnett, William A., and Ki-Hong Choi. 2008. "Operational Identification of the Complete Class of Superlative Index Numbers: An Application of Galois Theory." *Journal of Mathematical Economics*, 44(7): 603–612. Special Issue in Economic Theory in Honor of Charalambos D. Aliprantis.

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

Bils, Mark. 2009. "Do Higher Prices for New Goods Reflect Quality Growth or Inflation." *Quarterly Journal of Economics*, 124(2): 637–679.

Bils, Mark, and Peter Klenow. 2001. "Quantifying Quality Growth." *American Economic Review*, 91(4): 1006–1030.

Boskin, Michael J, Ellen R Dulberger, Robert J Gordon, Zvi Griliches, and Dale W Jorgenson. 1998. "Consumer Prices, the Consumer Price Index, and the Cost of Living." *Journal of Economic Perspectives*, 12(1): 3–26.

Broda, Christian, and David Weinstein. 2010. "Product Creation and Destruction: Evidence and Price Implications." *American Economic Review*, 100(3): 691–723.

Brown, Craig, and Anya Stockburger. 2006. "Item Replacement and Quality Change in Apparel Price Indexes." *Monthly Labor Review*, 35–45.

Bureau of Economic Analysis. Accessed 2018. "Personal Consumption Expenditures." https://www.bea.gov/data/consumer-spending/main.

Bureau of Labor Statistics. Accessed, 2018. "Consumer Price Index." special tabulation provided to the authors by BLS.

Byrne, David M., Daniel E. Sichel, and Ana Aizcorbe. 2019. "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.

Cafarella, Michael, Gabriel Ehrlich, Tian Gao, John C Haltiwanger, Matthew D Shapiro, and Laura Zhao. 2023. "Using Machine Learning to Construct Hedonic Price Indices." NBER Working Paper 31315.

Crawford, Ian, and J Peter Neary. 2019. "New Characteristics and Hedonic Price Index Numbers." *The Review of Economics and Statistics*, 105(3): 665–682.

De Haan, Jan. 2008. "Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Other Approaches." CAER Working Paper 2008-01, CAER, Centre for Applied Economic Research, Faculty of Commerce and Economics, UNSW, The University of New South Wales.

De Haan, Jan, and Heymerik A. Van Der Grient. 2011. "Eliminating Chain Drift in Price Indexes Based on Scanner Data." *Journal of Econometrics*, 161(1): 36–46.

Diewert, Erwin. 1976. "Exact and Superlative Index Numbers." *Journal of Econometrics*, 4(2): 115–145.

Diewert, Erwin. 1978. "Superlative Index Numbers and Consistency in Aggregation." *Econometrica*, 46(4): 883–900.

Diewert, Erwin. 2021. "The Economic Approach to Index Number Theory." University of British Columbia Vancouver School of Economics Discussion Paper 20-05.

Diewert, Erwin, Saeed Heravi, and Mick Silver. 2008. "Hedonic Imputation versus Hedonic Time Dummy Indexes." NBER Working Paper 14018.

Ehrlich, Gabriel, John C. Haltiwanger, Ron S. Jarmin, David Johnson, and Matthew D. Shapiro. 2022. "Reengineering Key National Economic Indicators." In *Big Data for Twenty-First-Century Economic Statistics.* , ed. Katharine G. Abraham, Ron S. Jarmin, Brian C. Moyer and Matthew D. Shapiro. Chicago:University of Chicago Press.

Eltetö, Oded, and Pal Köves. 1964. "On a Problem of Index Number Computation Relating to International Comparison." *Statisztikai Szemle*, 42(10): 507–518. (In Hungarian).

Erickson, Tim, and Ariel Pakes. 2011. "An Experimental Component Index for the CPI: From Annual Computer Data to Monthly Data on Other Goods." *American Economic Review*, 101(5): 1707–1738.

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

Feenstra, Robert C., and Marshall B. Reinsdorf. 2007. "Should Exact Index Numbers Have Standard Errors? Theory and Application to Asian Growth." In *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches.* University of Chicago Press.

Gini, Corrado. 1931. "On the Circular Test of Index Numbers." *Metron*, 9(9): 3–24.

Google Code Archive. 2013. "word2vec." Downloaded March 2022 from https://code.google.com/archive/p/word2vec/. Introduced and described in Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013), "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781.*

GS1. 2018. "GS1 DataHub." https://www.gs1us.org/tools/gs1-us-data-hub.

Hill, Robert J, and Daniel Melser. 2008. "Hedonic Imputation and the Price Index Problem: An Application to Housing." *Economic Inquiry*, 46(4): 593–609.

Ivancic, Lorraine, Erwin Diewert, and Kevin J. Fox. 2011. "Scanner Data, Time Aggregation and the Construction of Price Indexes." *Journal of Econometrics*, 161(1): 24–35.

Jaravel, Xavier. 2024. "Distributional Consumer Price Indices." CEPR Working Paper 19802.

Kilts Center for Marketing at the University of Chicago Booth School of Business. Accessed, 2018. "NielsenIQ Consumer Panel Data and Retail Scanner Data."

Konüs, Alexander A. 1939. "The Problem of the True Index of the Cost of Living." *Econometrica*, 7(1): 10–29.

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

National Retail Federation. 2023. "4-5-4 Calendar." Accessed February 10, 2023.

Pakes, Ariel. 2003. "A Reconsideration of Hedonic Price Indexes With an Application to PC's." *American Economic Review*, 93(5): 1578–1596.

Pakes, Ariel. 2005. "Hedonics and the Consumer Price Index." *Annales d'Économie et de Statistique*, , (79-80): 729–748.

Redding, Stephen J., and David E. Weinstein. 2020. "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences." *The Quarterly Journal of Economics*, 135(1): 503–560.

Reinsdorf, Marshall B., and Alan H. Dorfman. 1999. "The Sato–Vartia Index and the Monotonicity Axiom." *Journal of Econometrics*, 90(1): 45–61.

Sato, Kazuo. 1976. "The Ideal Log-Change Index Number." *The Review of Economics and Statistics*, 58(2): 223–228.

Sculz, B. 1964. "Index Numbers for Multiregional Comparisons." *Przeglad Statystyczny*, 3: 239–254. (In Polish).

Sichel, Daniel E., and Christopher Mackie. 2022. *Modernizing the Consumer Price Index for the 21st Century.* National Academies of Sciences, Engineering, and Medicine. Washington, DC: The National Academies Press.

Triplett, Jack. 2004. "Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes." OECD Working Paper.

US Bureau of the Census. 2018. "Point of Sales Data for Select Product Groups from NPD." Third party data provided to the U.S. Bureau of the Census.

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

Table 1: Rates of Product Turnover: NPD Data, Percent

	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: NPD Data

	Memory Cards	Coffee-makers	Head-phones	Boys' Jeans	Occ. Footwear
A. Traditional and Hedonic Indices					
Arithmetic Laspeyres	0.64	0.80	0.85	0.82	0.92
Arithmetic Hedonic Laspeyres	0.43	0.70	0.61	0.75	0.87
Geometric Laspeyres	0.54	0.75	0.61	0.74	0.89
Geometric Hedonic Laspeyres	0.41	0.68	0.49	0.71	0.86
Tornqvist	0.47	0.69	0.61	0.73	0.87
Hedonic Tornqvist	0.40	0.67	0.54	0.68	0.86
B. Demand-based Indices					
Sato-Vartia	0.48	0.71	0.60	0.77	0.88
Feenstra	0.47	0.69	0.58	0.75	0.86
CUPI	0.39	0.62	0.33	0.18	0.78
Nested CUPI	0.37	0.64	0.35	0.17	0.78

Notes: All indices are chained quarterly indices. Hedonic indices use the Erickson and Pakes (2011) time-varying unobservables approach. Both CUPIs use a 30th-percentile common goods rule. Nested CUPI is nested using characteristics approach.

Table 3: Alternative Treatment of Initial Residual for Entrants: Hedonic Tornqvist Price Index Growth: NPD Data, Percent

Method for Initial Residual	Memory Cards	Coffee-makers	Head-phones	Boys' Jeans	Occ. Footwear
A. Chained Cumulative Four-Quarter Annual Rates					
Zero mean	-20.1	-9.6	-14.1	-9.2	-3.8
Current	-20.1	-9.6	-13.9	-9.3	-3.8
Backcast	-20.1	-9.6	-13.9	-9.3	-3.8
B. Direct Year-over-Year Rates for Fourth Quarters					
Zero mean	-20.5	-10.3	-16.4	-6.3	-3.4
Current	-20.6	-10.1	-16.9	-6.1	-3.3
Backcast	-20.6	-10.1	-16.8	-6.3	-3.4

Notes: Table shows alternative estimates of average annual price growth from 2014 to 2018. Zero mean is the baseline approach for imputing initial residual for entering products as described in text, current uses current period residual, backcast uses a backcast imputation from the current period residual. Hedonic indices use Erickson and Pakes (2011) time-varying unobservables approach.

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

	Memory Cards	Coffee-makers	Head-phones	Boys' Jeans	Occ. Footwear
A. Traditional and Hedonic Indices					
Laspeyres (C)	-13.9	-6.9	-11.7	-7.4	-2.9
Laspeyres (GL)	-12.2	-5.6	-11.8	-6.1	-2.1
Tornqvist (C)	-16.9	-8.9	-11.6	-7.6	-3.3
Tornqvist (GL)	-15.4	-6.6	-11.6	-5.5	-2.3
Hedonic Tornqvist (C)	-20.1	-9.6	-14.1	-9.2	-3.8
Hedonic Tornqvist (GL)	-20.3	-10.0	-15.3	-7.8	-3.6
B. Demand-based Indices					
Sato-Vartia (C)	-16.2	-8.2	-11.7	-6.2	-3.1
Sato-Vartia (GL)	-14.3	-6.4	-11.3	-4.1	-2.1
Feenstra (C)	-16.8	-8.9	-12.5	-6.9	-3.8
Feenstra (GL)	-16.5	-9.4	-13.1	-5.5	-3.8
CUPI (C)	-20.6	-11.1	-24.1	-34.7	-6.1
CUPI (GL)	-19.9	-9.4	-22.6	-26.9	-5.2

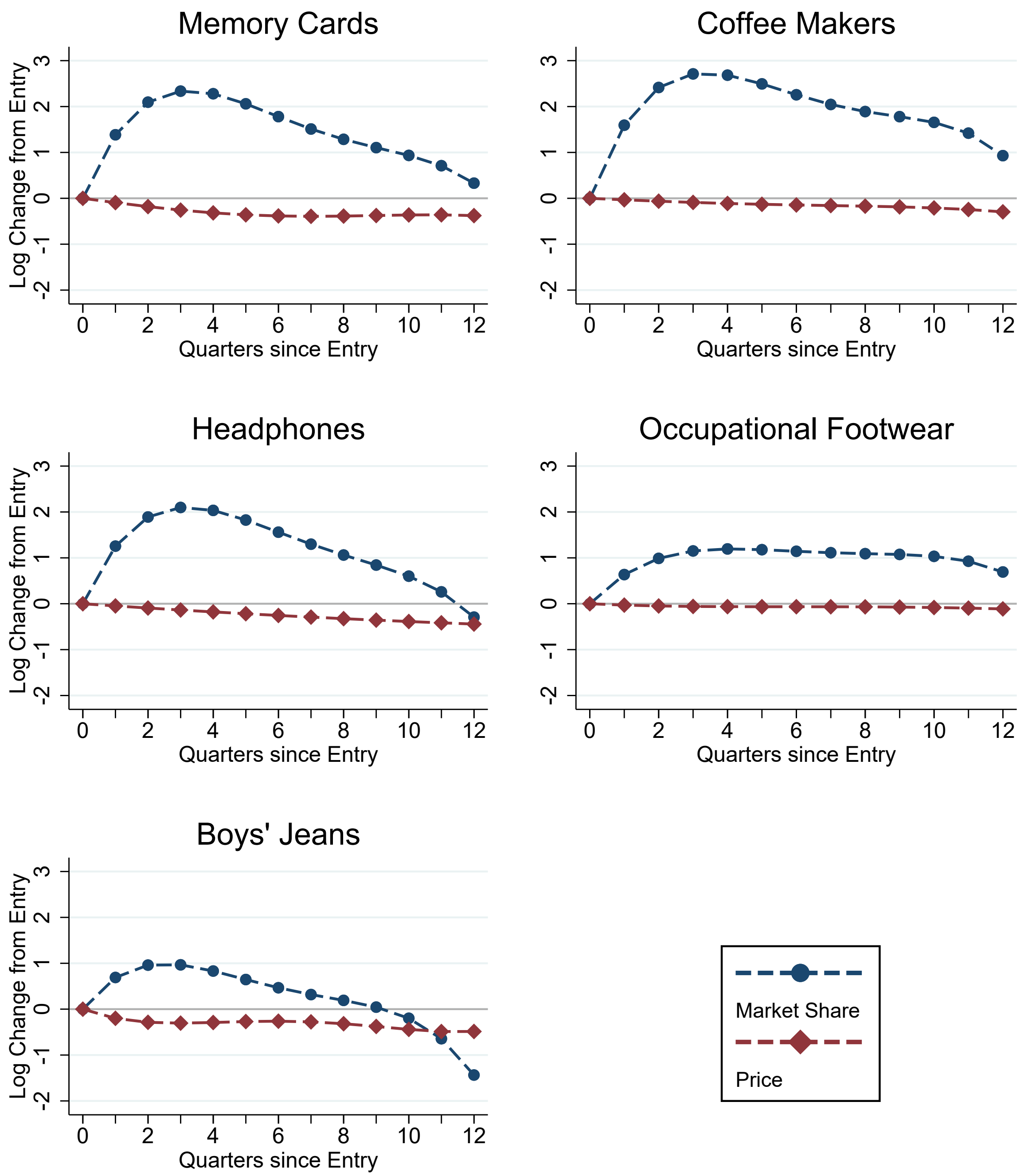
Notes: The rows for chained indices show averages of cumulative quarterly inflation rates for the year. The Laspeyres indices are geometric. GEKS-lite is the average of the geometric mean of the chained values and the year-over-year price indices for the fourth quarter of each year. Hedonic indices use Erickson and Pakes (2011) time-varying unobservables approach. CUPIs use a 30th-percentile common goods rule.

Table 5: Chained vs GEKS-Lite Price Index Growth: NielsenIQ Food, Percent

Index	Chained	GEKS-Lite
Laspeyres	1.9	1.8
Tornqvist	0.9	1.3
Sato-Vartia	1.2	1.4
Feenstra	0.7	0.8
CUPI	-2.5	-1.2

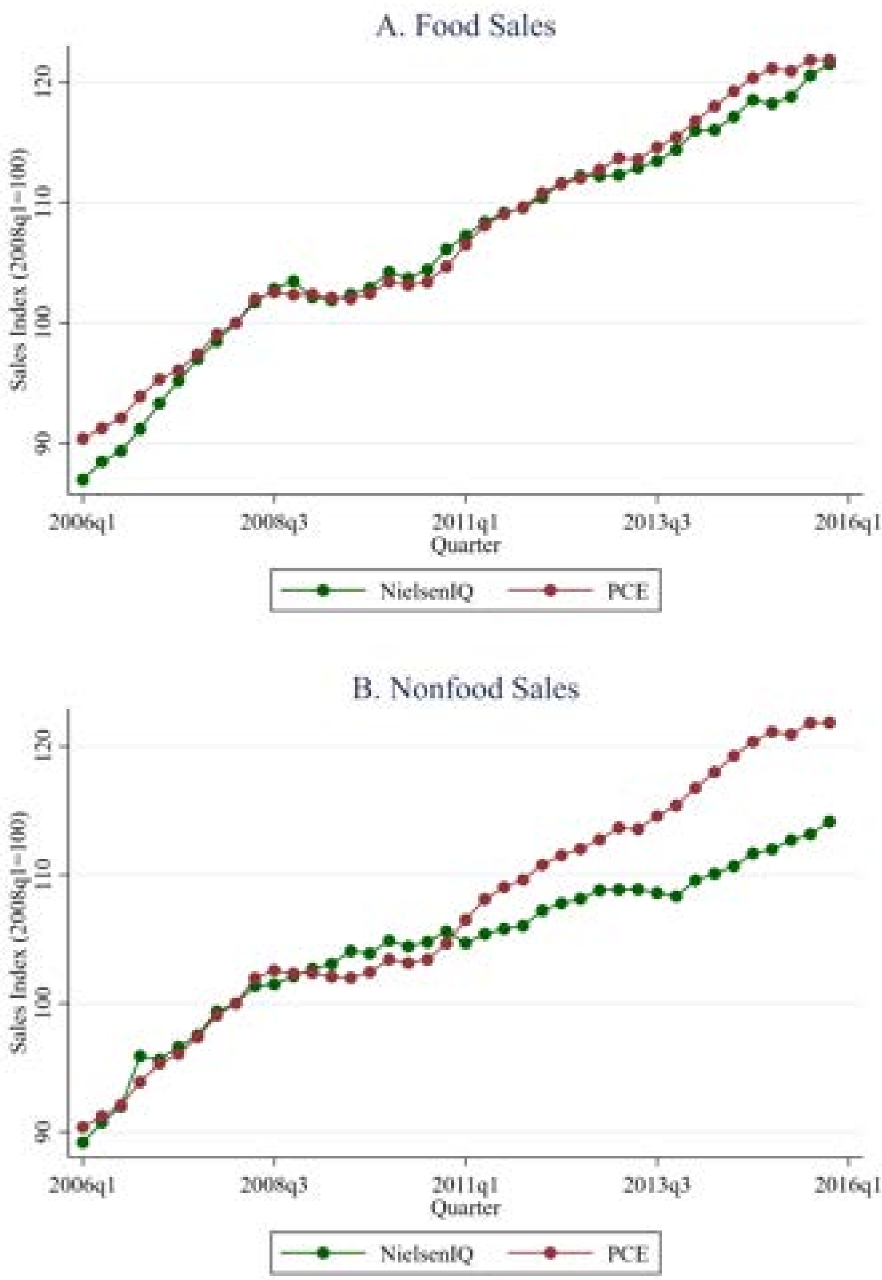
Notes: Chained values are averages of cumulative quarterly rates for year. GEKS-lite is the average of the geometric mean of the chained values and the YoY price indices for q4 for each year. Laspeyres is the geometric Laspeyres. CUPI uses 30th percentile CGR. Data come from NielsenIQ.

Figure 1: Product Lifecycle of Market Share and Prices: NPD Data



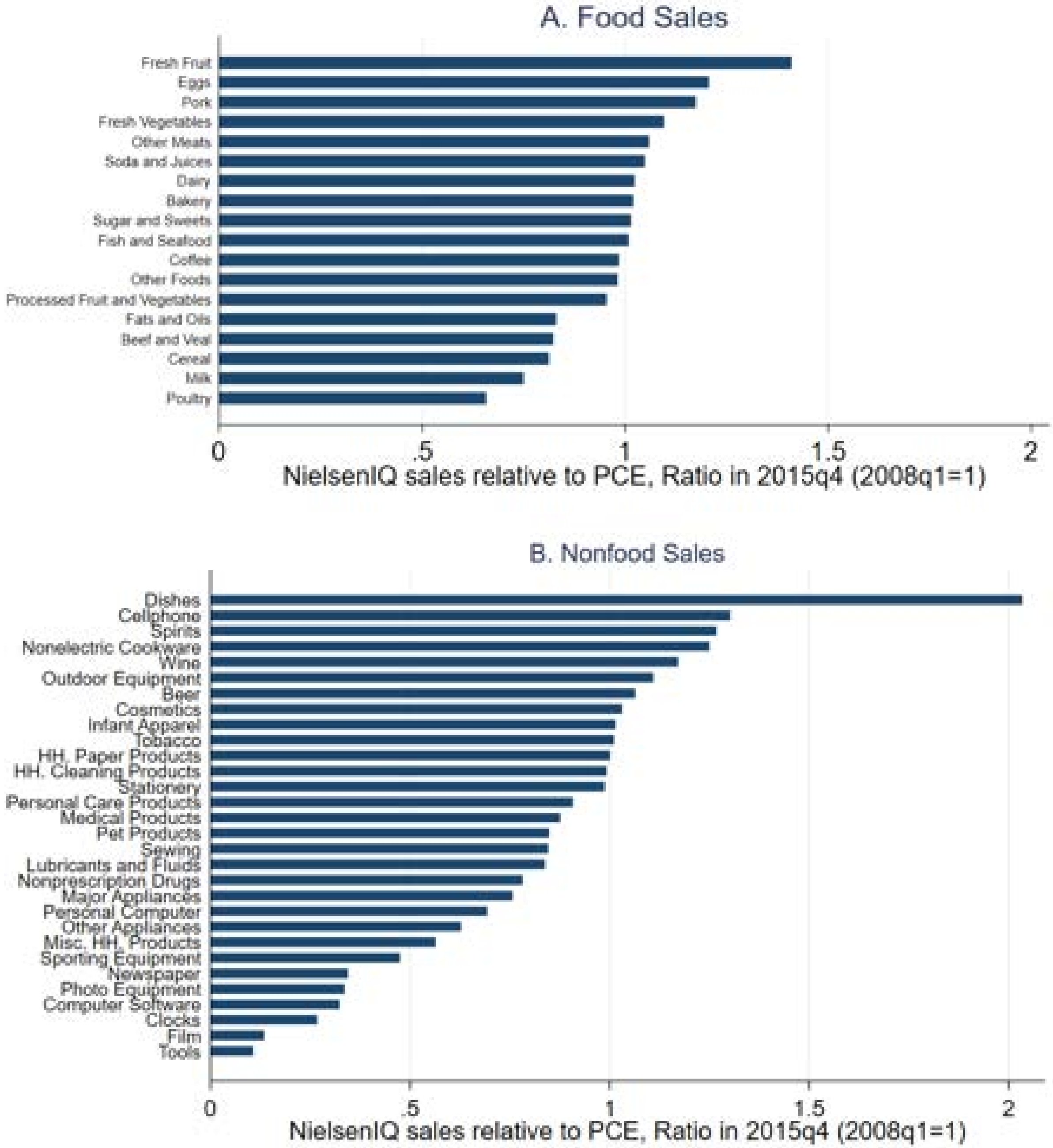
Notes: The plots display cumulative unweighted mean changes from the period of initial entry for each product's market share and price relative to their values in the period of their initial entry. Initial entry occurs in period 0. All series are smoothed with a quartic spline.

Figure 2: PCE versus NielsenIQ Sales



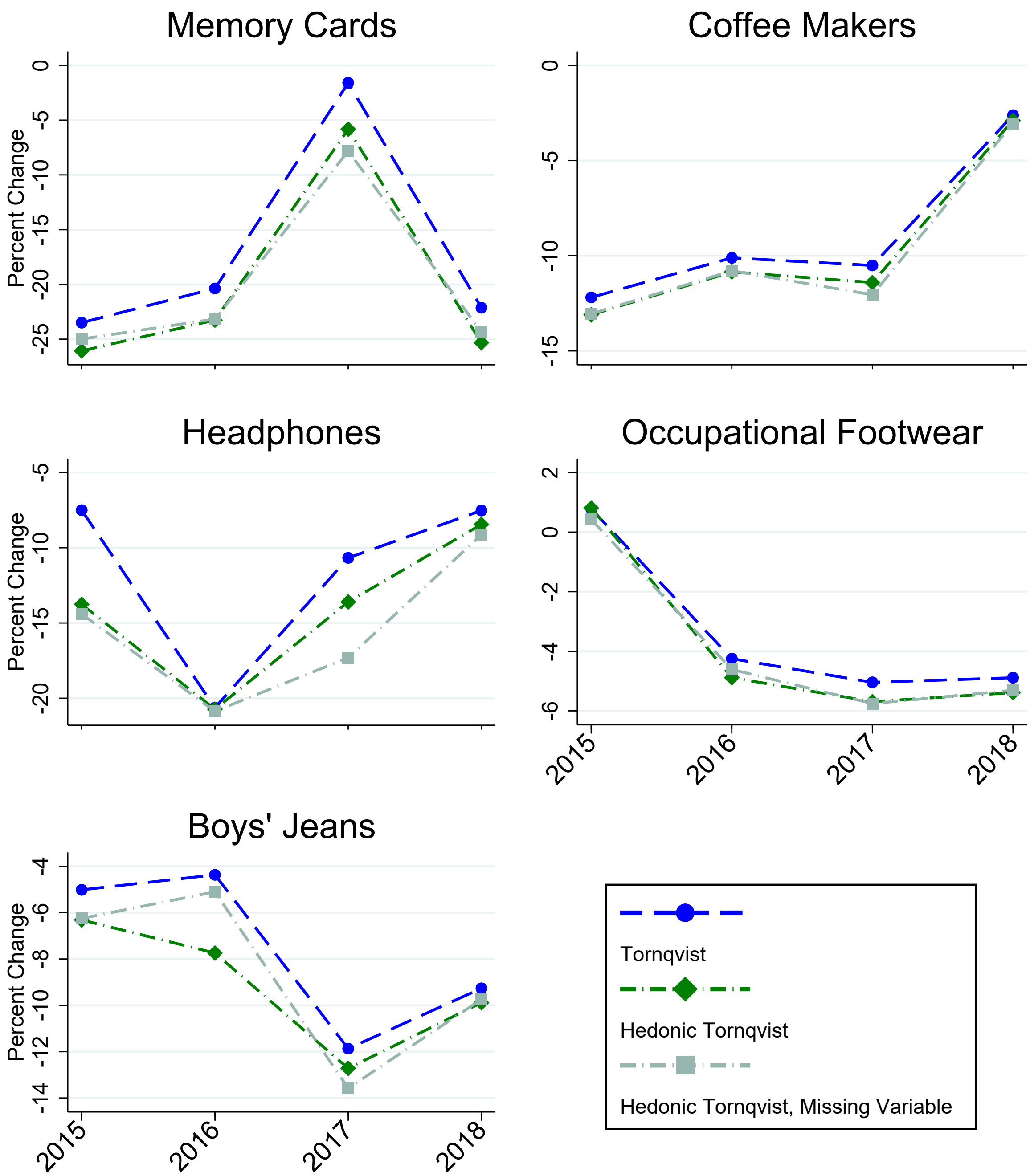
Notes: NielsenIQ Retail Scanner data. PCE is personal consumption expenditures (nominal) from BEA. All series normalized to 100 in 2008q1.

Figure 3: PCE versus NielsenIQ Sales, By Category within Food and Nonfood



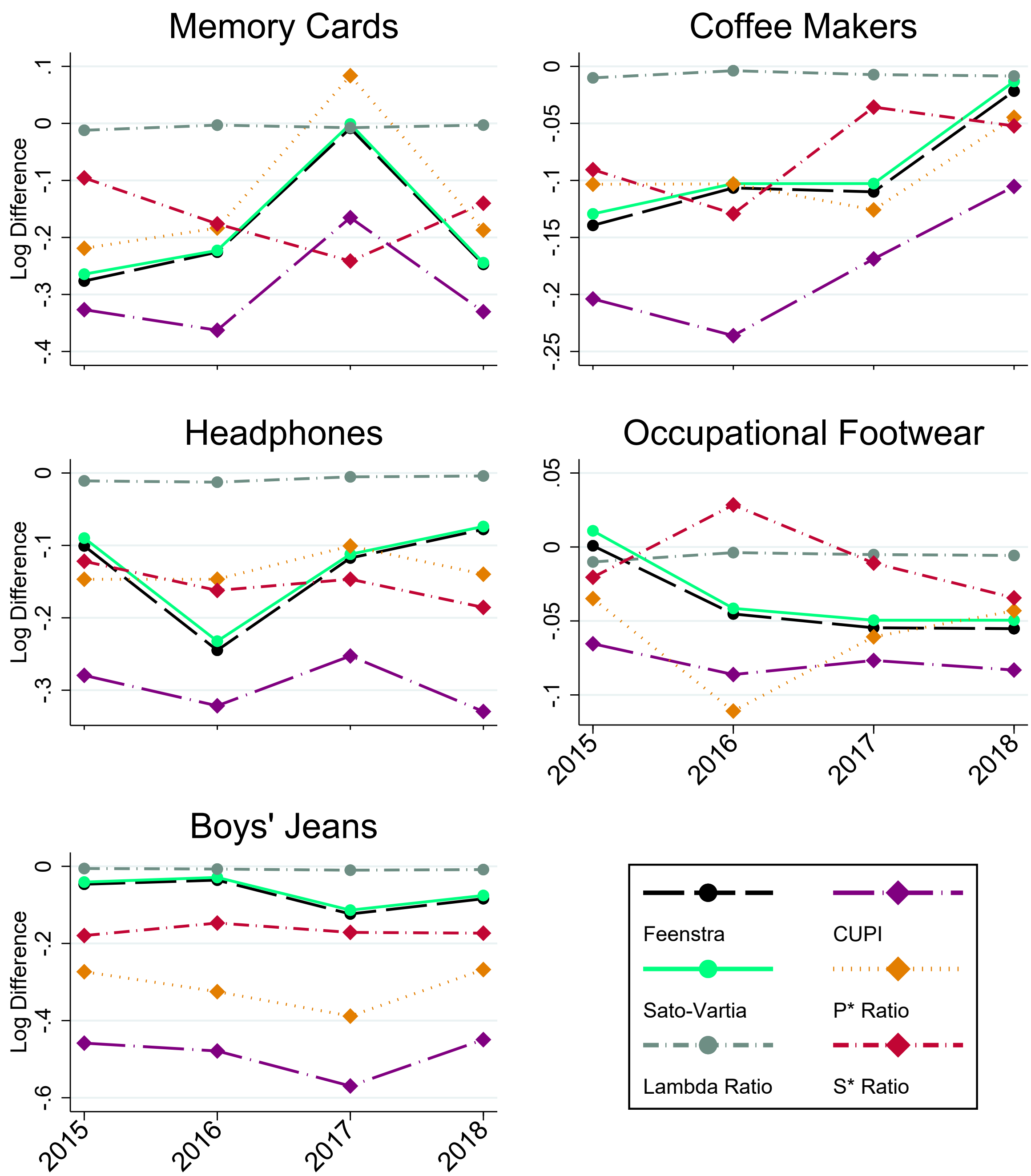
Notes: Figure shows the ratio of NielsenIQ Retail Scanner sales to BEA sales by product categories in 2015q4 (2008q1=1). A value of one for a product group indicates that expenditures in the NielsenIQ data grew at the same rate as the PCE data over that period; values below one indicate slower growth in the NielsenIQ data, and values above one indicate faster growth.

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



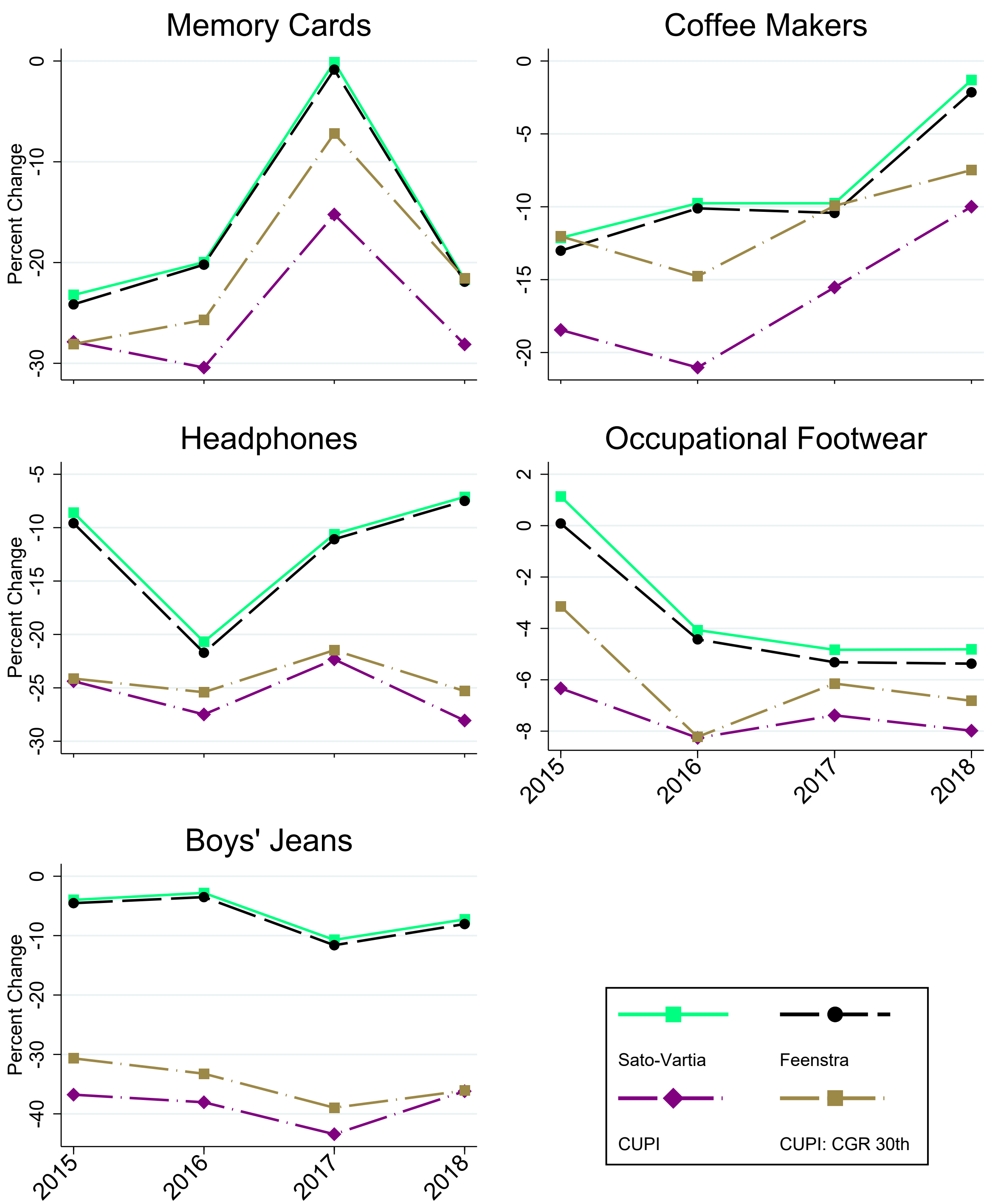
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. The hedonic Tornqvist indices display full imputation indices estimated using the time-varying unobservables approach, estimating hedonic models of log change in price using WLS and average quantity-share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

Figure 5: Components of the CUPI: NPD Data



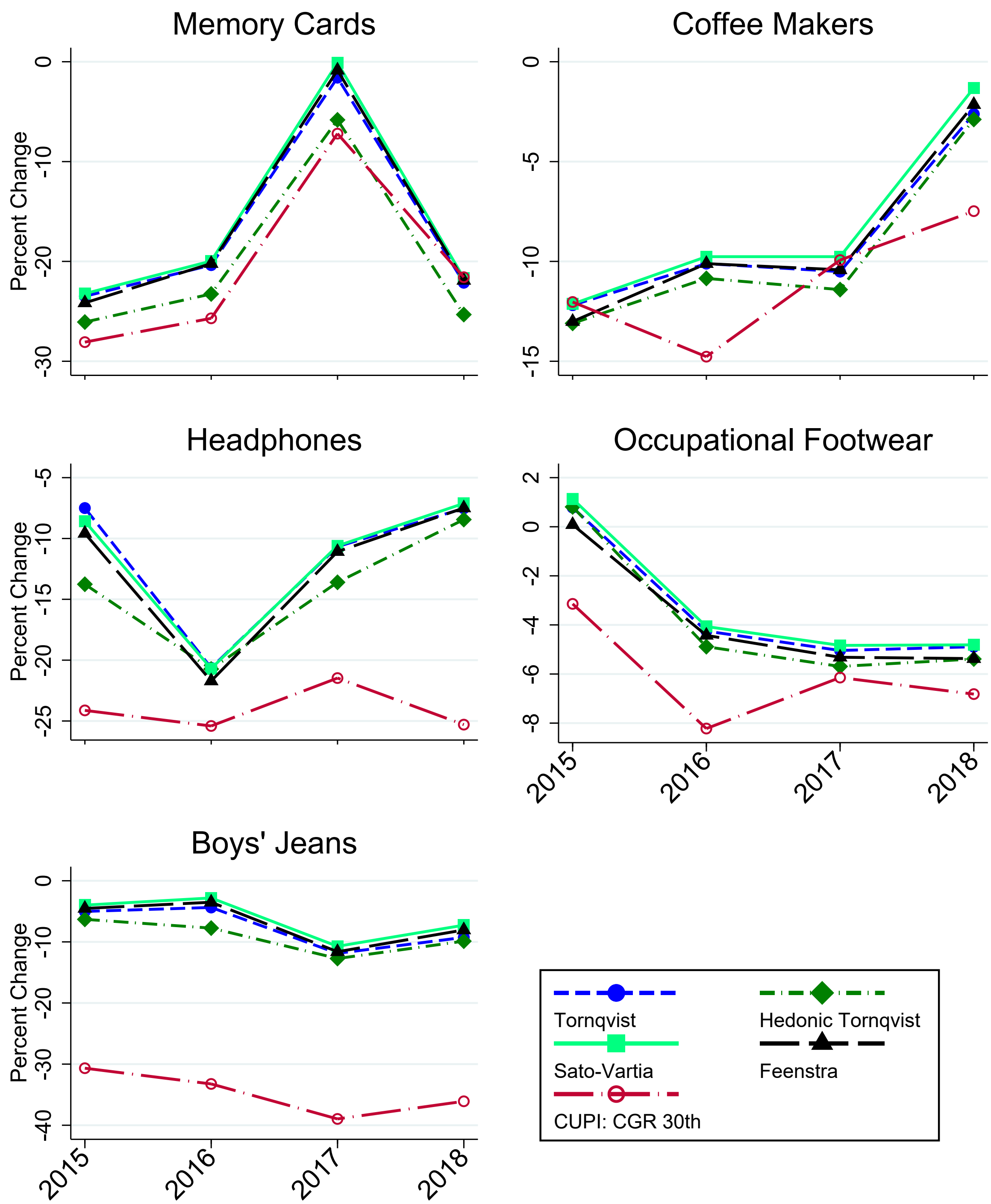
Notes: Values are log differences on a q4-to-q4 basis, aggregated from chained quarterly price indices. Units are reported in log-differences to allow for an additive decomposition of price indices. The plot displays the elements of the CUPI, which is the sum of the Lambda ratio, P^*-ratio, and S^*-ratio.

Figure 6: CUPI, With and Without Common Goods Rules (CGRs): NPD Data



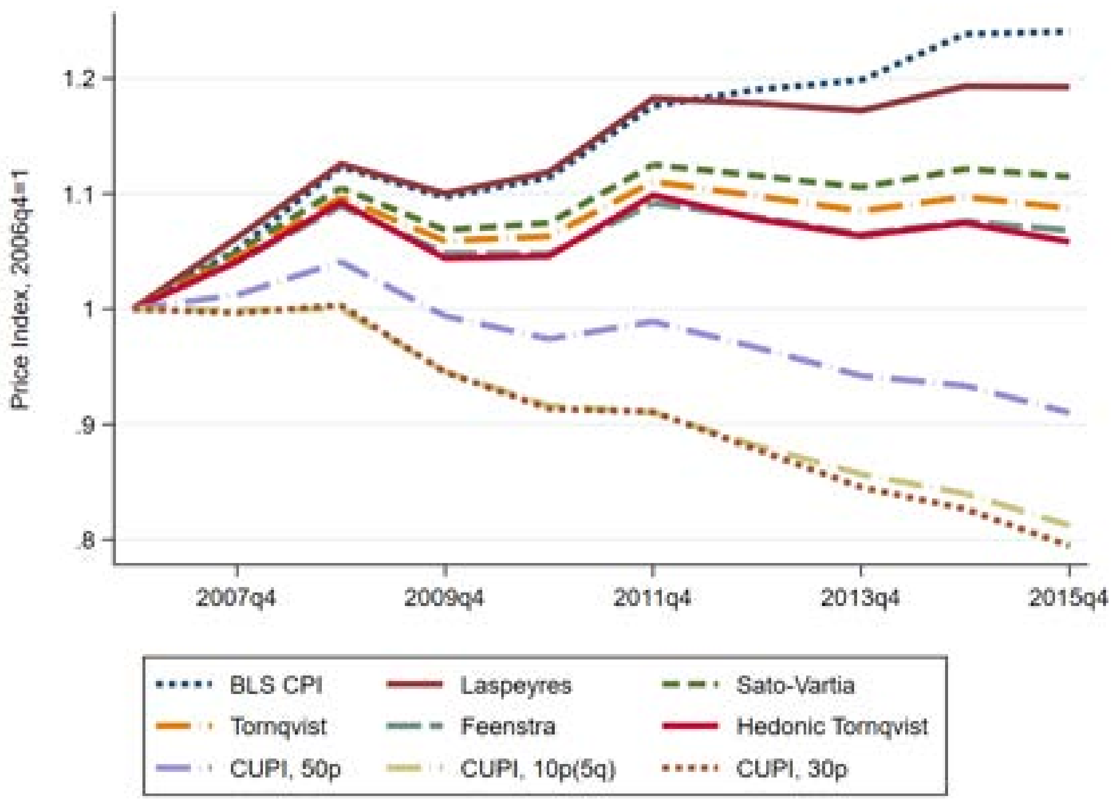
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. "CUPI: CGR 30th" excludes from the group of common goods those products with market shares below the 30th percentile in both periods. The Feenstra and Sato Vartia indices are included for reference.

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. "Hedonic Tornqvist" 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. "CUPI: CGR 30th" excludes from the group of common goods those products with market shares below the 30th percentile in both periods.

Figure 8: Main Price Index Specifications, Price 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, 30p" uses 30th-percentile expenditure shares that must be met for common goods in t and $t-1$. "CUPI, 10p (5q)" uses a 10th-percentile expenditure share threshold computed over quarters t and $t-4$ that a common good must exceed in t along with the requirement that the common good must be present in t and $t-4$.

Supplementary Appendix

Quality Adjustment at Scale: Hedonic vs. Exact Demand-Based Price Indices

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

American Economic Review

November 2025

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 (13) 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, Heravi and Silver (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.[1]

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).[2] 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 III.E. Such results reflect the many different issues including chain drift relevant for comparing alternatives. We

[1]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.

[2]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.

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 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.[3]

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 Retail Scanner (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, Redding and Weinstein (2016) and Redding and

[3]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.

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, Redding and Weinstein (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 price and sales changes at the top and bottom 1%. We also winsorize prices at the 99th percentile at the product module by quarter level. 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 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, Redding and Weinstein (2016) and Redding and Weinstein (2020) to prepare the data. In the next subsection we discuss differences between the Retail 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 Retail 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 Retail 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 Retail 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 Retail Scanner data for all purchases from participating retail partners in the Retail 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, Leibtag and Nevo (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, Leibtag and Nevo (2010) find significant misreporting of prices and quantities.

In order to compare the relative merits of the Retail 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 Retail Scanner and Consumer Panel with official statistics. Figure D.6 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.[4] As we reported in the main text, for food, we find nominal sales indices for the NielsenIQ Retail 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 Retail 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 Retail 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 Retail Scanner data is not surprising. The Scanner data tracking of stores that sell the nonfood items tracked by NielsenIQ has weaker coverage. For the Consumer Panel, the limitations discussed above are again presumably at work.

Figure D.7 presents the relationship between the BLS CPI and corresponding Laspeyres indices from the NielsenIQ Retail Scanner and Consumer Panel data sets.[5] 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 Retail Scanner and Consumer Panel Laspeyres indices are highly correlated with the CPI. In terms of inflation levels, however, the NielsenIQ Retail Scanner more closely matches the CPI. The BLS CPI for Food averages 2.5% over our sample period, the arithmetic Laspeyres for the Retail Scanner averages 3.1% and the arithmetic Laspeyres for

[4] PCE data used in main text and in this appendix downloaded from BEA in September 2020. For the broad food and nonfood comparisons with PCE we use a concordance of the 100 plus product groups in the Nielsen data with the PCE. When we examine more detailed categories in Figure 3 use a concordance provided to us by BLS between PCE categories and the 1000 or so Nielsen product modules. We have found that at the aggregate food and nonfood levels using the product level concordance or product module concordance is not important.

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

the Consumer Panel averages 4.2%. The correlations between Laspeyres indices for the non-food product groups and the CPI are much weaker than for food (0.54 and 0.71 for the Retail Scanner and Consumer Panel data sets, respectively, using the arithmetic Laspeyres). The average inflation level is closer to the CPI in the Retail Scanner data than in the Consumer Panel (CPI mean is 0.9%, Retail Scanner mean is -0.3%, Consumer Panel Mean is 2.3%).

The results of our comparisons with official statistics show that the NielsenIQ Retail Scanner data matches expenditure patterns and price indices with official statistics well for food at the aggregate and disaggregate levels. For nonfood, the NielsenIQ Retail 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 nonfood.

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 Retail 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 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.8 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 Retail Scanner data. In this figure, all series are based on log differences so the decomposition is exact. 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 log points below the Feenstra. Using a 50th percentile for the CGR yields a price index that is substantially closer to the Feenstra index.

Again using the Retail Scanner data, Figure D.11 shows the CUPI with alternative common goods rules compared to matched-model, Feenstra, and hedonic indices. The figure highlights that the CUPI without common goods rules yields much lower inflation than matched-model, Feenstra, and hedonic indices. The common goods rules bring the CUPI more in line with other indices. Either a duration component or a hight percentile threshold is required.

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 Retail 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.[6]

Figure D.15 shows related indices, using the NielsenIQ Retail Scanner data, pooling all product groups, and using various CGRs based on sales percentiles computed over the 5-quarter horizon.[7] These results are therefore suggestive of the results that applying the empirical approach in Redding and Weinstein (2020) to the Retail Scanner data would produce. The CUPI with no CGR suggests deflation of 10 percent or more per year. The CUPI yields less deflation with stricter common goods rules based on expenditure share percentile thresholds. 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 Retail Scanner data, rather than calculating a percentile-based threshold directly from the Retail Scanner data. Using the Consumer Panel share threshold for the CGR produces an index very similar to the Feenstra index.

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 Retail 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 Discussing the National Market Assumption for the CUPI

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)

[6] 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.8, 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.

[7] 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.

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.

The literature on spatial variation in brands (see, e.g., Bronnenberg, Dahr and Dube (2007) and Bronnenberg, Dahr and Dube (2009)) highlights that there are important roles for local, regional and national brands (they focus on consumer packaged goods like those covered by NielsenIQ). Moreover, even for national brands, this analysis shows a slow roll-out of new items geographically. These patterns raise questions about applying a national market based CES price index for most items. In Appendix D.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 Details of Machine Learning Algorithm

This appendix details the steps in the machine learning algorithm developed in Cafarella et al. (2023) used to produce the hedonic Tornqvist index for the NielsenIQ data. See that paper for further discussion and evaluation of procedure. See the codebase in the replication package for the exact implementation of the procedure.

Step 1: Convert text to embeddings. The product descriptions from the Nielsen data provided by the Kilts Center for Marketing at the University of Chicago are generally not coded to be human-intelligible. For instance, two product descriptions for soft drinks are: `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`. (Brand is included in the text fed into the algorithm, though we have suppressed brand name in reporting these examples.)

We convert the individual words into numeric vectors called embeddings using a hybrid feature encoding architecture. We use two separate embeddings for each word in the product descriptions. The "words" are not all English words, and are sometimes called "tokens" or "grams." The first is a set of pretrained embeddings from the Word2Vec algorithm (Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean, 2013). The second set uses customized embeddings, which are initialized with random vectors, one for each of the unique words in the set of NielsenIQ product descriptions, and fine-tuned during the training process described in steps 2 and 3. Both sets of word-level embeddings are 300-dimension vectors. The network learns the meanings of these vectors over the course of its training (while simultaneously tuning the customized embeddings).

Step 2: Convert the embeddings into full-sequence representations. We next combine the word-by-word embeddings into numeric summaries of the entire "sentence-long" product descriptions by feeding them into a "Long Short-Term Memory" component, or "LSTM," which offers a way to represent sequences of words, not just standalone words. An LSTM takes as an input a sequence of word-by-word input representations and emits a single "full sequence" representation (Hochreiter and Schmidhuber, 1997). We use two LSTM layers in our architecture. The first takes the 300-element embedding vectors for each word in the product description and processes them sequentially to produce a 128-element vector called a "hidden state" corresponding to each word.

Our training process uses dropout layers at three points in the algorithm. The first dropout layer applies to the two 300-dimension embeddings for each word that is fed into the first LSTM layer. It applies a dropout rate of 20%, i.e., during model training, a random 20% of the 300 elements in the embedding vectors are set to zero. No dropout is applied when we use the model to make predictions after training is complete.

The second LSTM layer accepts these hidden states as inputs, after applying the second dropout layer to the output of the first LSTM layer with an 80% dropout rate during model training. Again, no dropout is applied when we use the model to make predictions after training is complete. The second LSTM layer also processes each word sequentially. We preserve and concatenate the final, average, and maximum values of these 128-element hidden state vectors, where the average and maximum are taken across the sequence of the LSTM's hidden states after processing the vectors corresponding to each word in the product description. We perform all of these steps separately on the pretrained embeddings and on the customized embeddings. We use the PyTorch package, version 1.12.1 (PyTorch Contributors, 2022), to create and train the LSTM components, with a learning rate of 5×10^{-4}.

Step 3: Train neural net to use the full-sequence representations to predict log price bin probabilities. We train a neural network to predict prices and price changes based on the output of step 2. Our algorithm does not predict product prices directly. Rather, we feed the full-sequence representations of the product descriptions into a neural network in order to predict the probabilities that each product's log price lies in a set of B bins, where we set $B = 10$ in this application. Consistent with the traditional econometric approach to implementing the Erickson and Pakes (2011) algorithm that we use with the NPD data, we produce these predictions separately period by period.

Our machine learning procedure is optimized to minimize the cross-entropy loss function. The system uses the embeddings X^k to produce a bin classification Y^k, where Y^k is a B-dimensional vector with classifier scores for each of the B equally-sized bins that partition the observed range of product prices or price changes. The classifier scores Y_b^k, $b = 1 \cdots B$, are then translated into probabilities using the softmax function

$$\underbrace{P(Y = Y^k | X = X^k)}_{B \times 1} = \frac{e^{Y^k}}{\sum_{b=1}^{B} e^{Y_b^k}}.$$

Denote the bin in which the price for product k truly lies as c. We can define the cross-entropy loss for product k as:

$$L^k = -\log P(Y = Y_c^k | X = X^k) = -\log \left(\frac{e^{Y_c^k}}{\sum_{b=1}^{B} e^{Y_b^k}} \right).$$

We define the objective function for our ML optimizer as the weighted sum of product-level cross-entropy losses:

$$L = \sum_k (w_k L^k),$$

where the weights w_k are product k's share of total unit sales.

Our neural network in this stage features two fully connected linear layers. In the first layer, we concurrently train a single-layer neural net over the output from the LSTM trained on the pretrained embeddings and another single-layer neural net over the output from the LSTM trained on the customized embeddings. Both of these fully connected layers emit 256-dimensional numeric vectors as outputs. In the second layer, we feed the output from both of those layers into a combined single-layer fully connected linear layer. This second fully connected linear layer produces a B-element numeric vector of classifier scores for each product as its output. We apply a third dropout layer with a dropout rate of 60% between the first and second fully connected linear layers during training. Again, no dropout is applied when we use the model to make predictions after training is complete.

We split the data into training, validation, and test sets using proportions of 70%, 20%, and 10%, respectively. We optimize the total loss function individually for each product group-year-quarter using the Adam gradient descent algorithm (Kingma and Ba, 2014). We train each model on the training data set for a number of "epochs," periods in which each example in the training set (i.e., a product description and observed price bin in the product group-quarter) is presented to the classifier one time. We randomly initialize the system to begin the first epoch. We begin subsequent epochs with the system in the state that it concludes the previous epoch.

We again use the PyTorch package, version 1.12.1 (PyTorch Contributors, 2022), to create and train the fully connected layers. We likewise again use a learning rate of 5×10^{-4}. The LSTM and the first set of of fully connected layers use Rectified Linear Unit (ReLU) activation functions. No activation function is applied to the output of the second fully connected layer; implicitly, the softmax function for converting the classifier scores into probabilities acts similarly to an activation function.

We train the model for 50 epochs, a significantly larger number than is likely to be optimal to avoid overfitting. We then apply the models trained in each of the epochs to the validation data set to assess the models' out-of-sample accuracy. We select the model from the best-performing epoch using the model's unweighted "near accuracy" as our selection criterion. We define the "near accuracy" as the unweighted proportion of products in the validation data set for which the model assigns the highest probability to the correct price bin or an adjacent bin (i.e., three bins, unless the highest-probability bin is the bottom or the top bin, in which case near accuracy is defined over two bins).

Step 4: Convert classifier output to price predictions and initial lagged residuals. The machine learning classifier estimates the probabilities of each item falling into individual log price bins. To predict product prices, we must convert those probabilities into continuous point estimates. A practical complication that arises is that products will often have very low estimated probabilities of falling into some bins. Including very low-probability bins in the calculation of predicted prices may add more noise than signal to the estimate.

We develop a procedure to calculate a set of modified bin probabilities to use for price prediction as follows. For each bin in each product group-quarter, we choose a cutoff probability $\overline{P}^{\star}$ to minimize the sum of the classifier's squared false negative rate and squared false positive rate. If the estimated probability from Step 3 for a bin is lower than $\overline{P}^{\star}$, we set the modified probability to zero. We then re-normalize the non-zero modified probabilities to sum to one. We calculate the inner product of the re-normalized bin probabilities and the averages of the log prices within each bin to form the predicted log price.

We then calculate the initial lagged residuals $\hat{\eta}_{kt\text{-}1}$ used in the second stage of the Erickson and Pakes (2011) procedure as each product's actual log price minus its predicted log price in each quarter that it is sold in the data.

Step 5: Repeat Step 2 (retrain customized embeddings sub-step), Step 3, and Step 4 to predict log price changes, using the initial lagged residuals as an additional feature in Step 3. We include the initial lagged residuals $\hat{\eta}_{kt\text{-}1}$ as additional features alongside the full-sequence embeddings of the product descriptions that come out of Step 2 to predict log price changes $\Delta \ln p_{kt}$. To repeat Step 3 with the lagged residual, we augment the full-sequence embeddings X^k with the residuals $\hat{\eta}_{kt\text{-}1}$. Specifically, we pass this residual into an 8-element fully connected layer (single-stage neural network) that otherwise has the same architecture and parameters as the fully connected layers described in Step 3. We concatenate the output from this fully connected layer for the residual with the output from the fully connected layers for the pretrained and customized embeddings in the final fully connected layer described in Step 3 that emits the classifier scores. To predict price changes, we define the weights w_k in the cross-entropy loss function as product k's average shares of total unit sales within the product group between periods t-1 and t. Apart from our handling of the lagged initial residuals and loss function weights, we use the same model architecture and training procedure for log price changes as for log price levels.

We repeat our procedure from Step 4 to calculate a set of modified bin probabilities from the raw log price change bin probabilities. Finally, we insert the predicted log price changes $\widehat{\ln \frac{p_{kt}}{p_{kt\text{-}1}}}$ into equations (7)–(9) to produce our hedonic price indices.

D Additional Evidence on the Behavior of the CUPI

In this appendix, we examine the properties of the CUPI in more detail. First, in section D.1, we consider nested versions of the CUPI. Second, in section D.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 D.3, we offer an analytical characterization of the taste shock bias in a simple setting.

D.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, Redding and Weinstein (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.[8] The between-nest estimator of the elasticity of substitution uses an instrumental variable (IV) approach building on Hottman, Redding and Weinstein (2016).[9]

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.

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.9 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.[10] 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.

D.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

[8]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, Redding and Weinstein (2016) assumption is arguably more natural, as differencing with respect to a within-nest mean more plausibly identifies orthogonal supply and demand shocks.

[9]We follow Hottman, Redding and Weinstein (2016) by specifying the between-group relationship between the nest-level price index and expenditure share. The former is endogenous, and Hottman, Redding and Weinstein (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, Redding and Weinstein (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, Redding and Weinstein (2016) with those of Redding and Weinstein (2020).

[10]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.

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.

D.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 D.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.

Exhibit D.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		

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 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 D.2 displays the results of this thought experiment.

Exhibit D.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
$\Delta \ln \Phi_t^{\text{SV}}$		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		

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 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.

Exhibit D.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		

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 D.3.

D.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 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.

D.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

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

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

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

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.

D.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

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

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

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

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.[11] 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.

D.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.

D.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

$$\begin{bmatrix} \ln p_{kt} \\ \\ \ln \varphi_{kt} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}, \begin{bmatrix} \chi_p^2 & \rho_{p\varphi} \\ \\ \rho_{p\varphi} & (\chi_{\varphi_0} + \mu t)^2 \end{bmatrix} \right). \tag{D.7}$$

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

[11] 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.

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 D.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.[12]

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 D.2.1.

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

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

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

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

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

[12] 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.

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.

D.2.4 Empirical Evidence

We examine the empirical relevance of rising dispersion in product appeal in Figure D.10. 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.

D.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.

D.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

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

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

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

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

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

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

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

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 spatial variation in brands documented in Bronnenberg, Dahr and Dube (2007) and Bronnenberg, Dahr and Dube (2009).

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.

D.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

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

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

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

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.[13] 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. Likewise, the Sato-Vartia index is not meaningfully affected, because goods that exit from periods t-1 to t are excluded. Yet stock-outs introduce a substantial bias to both the CUPI and the Feenstra index. The intuition for the bias is subtle. Stock-

[13] 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.

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.

D.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.

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

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.,

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

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 $(p_{kt\text{-}1}, \varphi_{kt\text{-}1})$,

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

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

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

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

(D.24)

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

Equation (D.24) shows that the unconditional expected taste shock bias will be zero when $\mathbb{E}\left[\omega_{kt}\ln\varphi_{kt}\right] = \mathbb{E}\left[\omega_{kt}\ln\varphi_{kt\text{-}1}\right]$:

(D.25)

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

where the middle condition follows from the normalization that $\mathbb{E}\left[\Delta\ln\varphi_{kt} = 0\right]$. We will demonstrate that $\mathbb{E}\left[\omega_{kt}\ln\varphi_{kt}\right] = \mathbb{E}\left[\omega_{kt}\ln\varphi_{kt\text{-}1}\right]$ 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

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

It is straightforward to see that ξ_{kt} is symmetric in its arguments, i.e., $\xi_{kt}\left(s_{kt}, s_{kt\text{-}1}\right) = \xi_{kt}\left(s_{kt\text{-}1}, s_{kt}\right)$. 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 $N_{\mathbb{C}_t} \to \infty$, 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.,

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

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

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

Likewise, we can write

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

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 (D.28) and (D.29) must be equal.

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.

References

Bajari, Patrick, Zhihao Cen, Victor Chernozhukov, Manoj Manukonda, Jin Wang, Ramon Huerta, Junbo Li, Ling Leng, George George Monokroussos, Suhas Vijaykunar, and Shan Wan. 2021. "Hedonic Prices and Quality Adjusted Price Indices Powered by AI." CENMAP Working Paper 04-21; forthcoming, *Journal of Econometrics.*

Blundell, Richard, and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics*, 87(1): 115–143.

Broda, Christian, and David Weinstein. 2010. "Product Creation and Destruction: Evidence and Price Implications." *American Economic Review*, 100(3): 691–723.

Bronnenberg, Bart, Sanjay Dahr, and Jean-Pierre Dube. 2007. "Consumer Packaged Goods in the US: National Brands, Local Branding." *Journal of Marketing Research*, 44(1): 4–13.

Bronnenberg, Bart, Sanjay Dahr, and Jean-Pierre Dube. 2009. "Brand History, Geography, and the Persistence of Brand Shares." *Journal of Political Economy*, 117(1): 87–115.

Cafarella, Michael, Gabriel Ehrlich, Tian Gao, John C Haltiwanger, Matthew D Shapiro, and Laura Zhao. 2023. "Using Machine Learning to Construct Hedonic Price Indices." NBER Working Paper 31315.

Diewert, Erwin. 1978. "Superlative Index Numbers and Consistency in Aggregation." *Econometrica*, 46(4): 883–900.

Diewert, Erwin. 2019. "Quality Adjustment and Hedonics: A Unified Approach." University of British Columbia Vancouver School of Economics Discussion Paper 19-01.

Diewert, Erwin, Saeed Heravi, and Mick Silver. 2008. "Hedonic Imputation versus Hedonic Time Dummy Indexes." NBER Working Paper 14018.

Einav, Liran, Ephraim Leibtag, and Aviv Nevo. 2010. "Recording Discrepancies in Nielsen Homescan data: Are They Present and Do They Matter?" *Quantitative Marketing and Economics*, 8: 207–239.

Erickson, Tim, and Ariel Pakes. 2011. "An Experimental Component Index for the CPI: From Annual Computer Data to Monthly Data on Other Goods." *American Economic Review*, 101(5): 1707–1738.

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

Gorajek, Adam. 2022. "Generalizing the Stochastic Approach to Price Indices." *Review of Income and Wealth*, 70(1): 1–22.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Comput.*, 9(8): 1735–1780.

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

Ivancic, Lorraine, Erwin Diewert, and Kevin J. Fox. 2011. "Scanner Data, Time Aggregation and the Construction of Price Indexes." *Journal of Econometrics*, 161(1): 24–35.

Kingma, Diederik P, and Jimmy Ba. 2014. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980.*

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781.*

Pakes, Ariel. 2003. "A Reconsideration of Hedonic Price Indexes With an Application to PC's." *American Economic Review*, 93(5): 1578–1596.

PyTorch Contributors. 2022. "PyTorch."

Redding, Stephen J., and David E. Weinstein. 2016. "Measuring Aggregate Price Indexes with Demand Shocks: Theory and Evidence for CES Preferences." NBER Working Paper 22479.

Redding, Stephen J., and David E. Weinstein. 2020. "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences." *The Quarterly Journal of Economics*, 135(1): 503–560.

Redding, Stephen J., and David E. Weinstein. 2024. "Accounting for Trade Patterns." *Journal of International Economics*, 150: 103910.

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		

Notes: This table contains the 59 Food product groups in the NielsenIQ Retail Scanner data for the price indices we use in this paper.

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		

Notes: This table contains the 51 Nonfood product groups in the NielsenIQ Retail Scanner data for the price indices we use in this paper.

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 III.A, 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)	-16.89	-8.86	-11.58	-7.63	-3.35
Tornqvist (GEKS-Lite)	-15.41	-6.64	-11.55	-5.55	-2.31
Tornqvist (Rolling-Year GEKS)	-15.57	-6.69	-9.12	-4.77	-1.96

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 III.E. The rolling-year GEKS indices in the third row implements the method of Ivancic, Diewert and Fox (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)

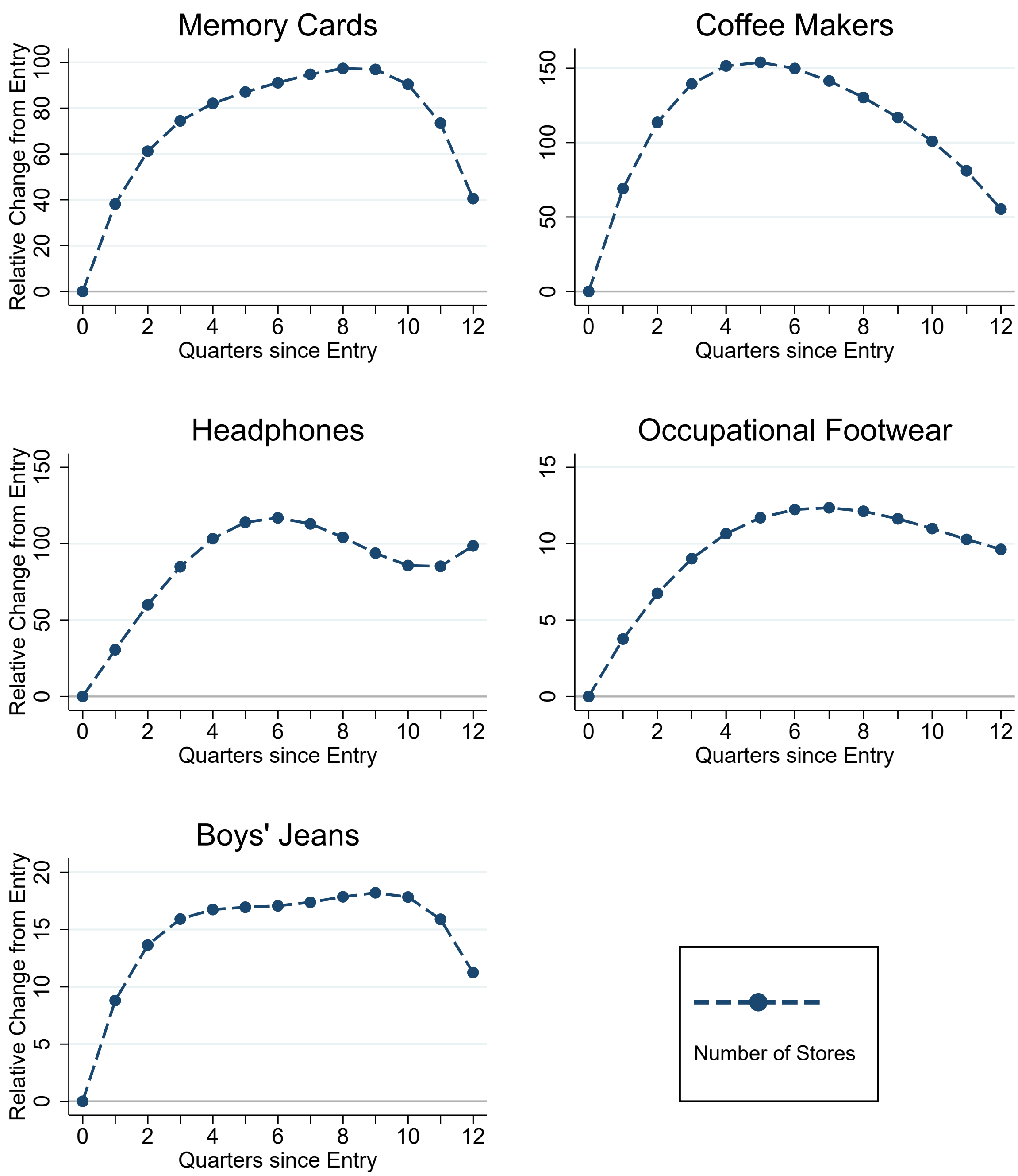
Notes: Estimated elasticities of substitution for nested CES models. Standard errors in parentheses. Data come from NPD Group. Within-nest elasticities are estimated using the methodology of Feenstra (1994) and Redding and Weinstein (2020). Across-nest elasticities are estimated using the nested CES estimation procedure of Hottman, Redding and Weinstein (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)

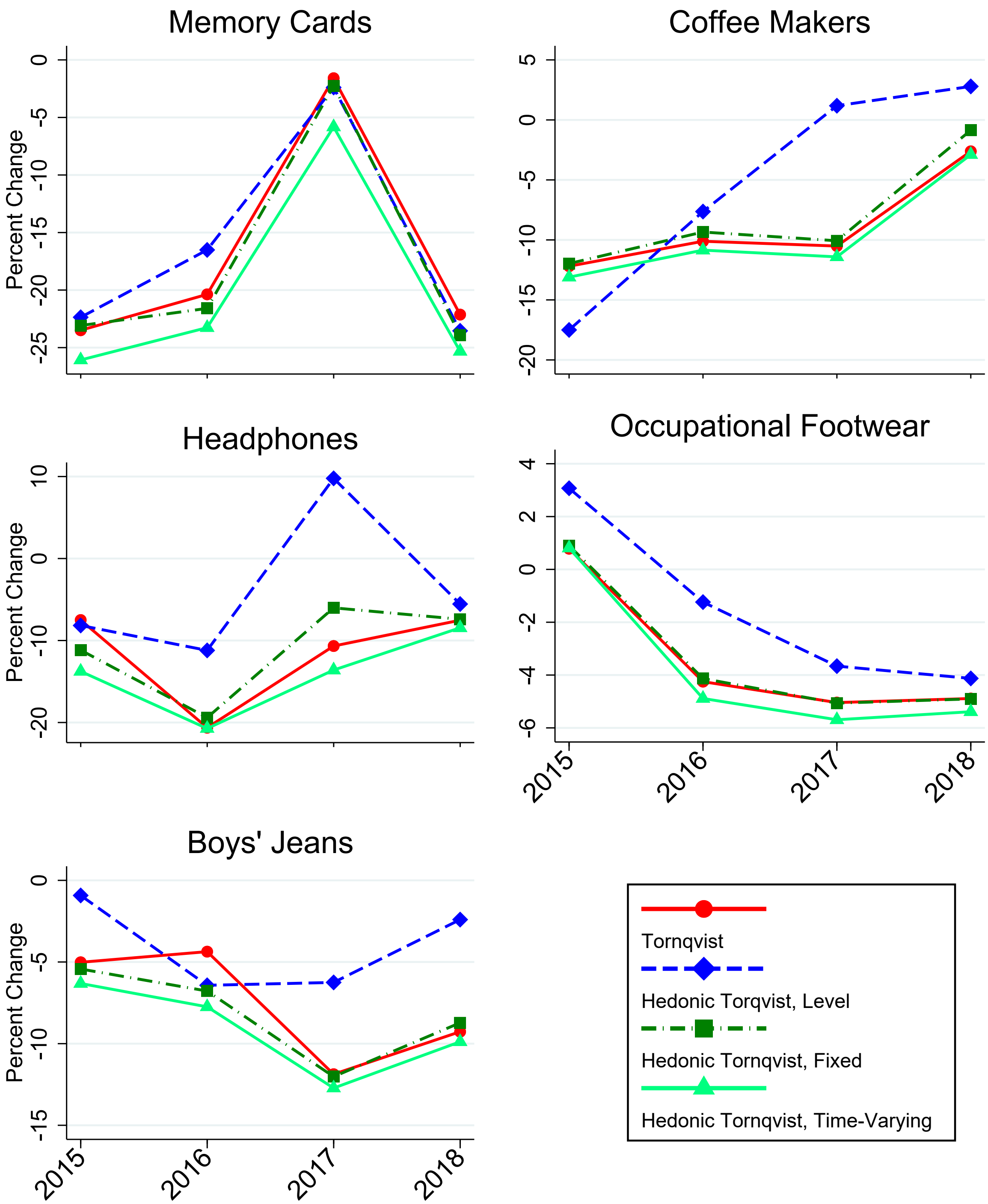
Notes: Regression of current relative quality on lagged relative quality from CUPI model. Standard errors in parentheses. Blundell-Bond allows for item-level fixed effects. Data come from NPD Group.

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



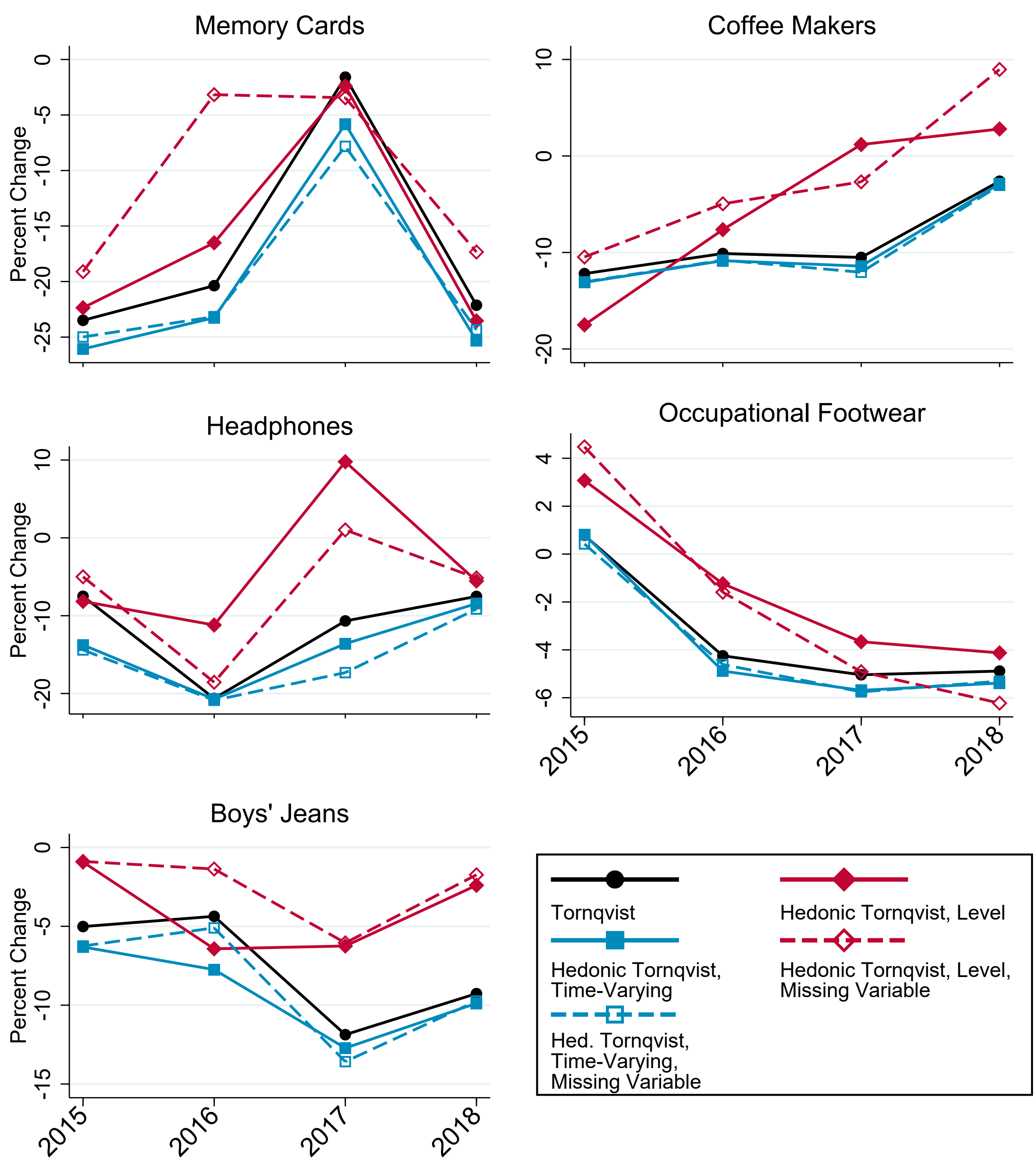
Notes: Number of stores relative to their value in the period of their initial entry. Entry occurs in period 0. All series are smoothed with a quartic spline.

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



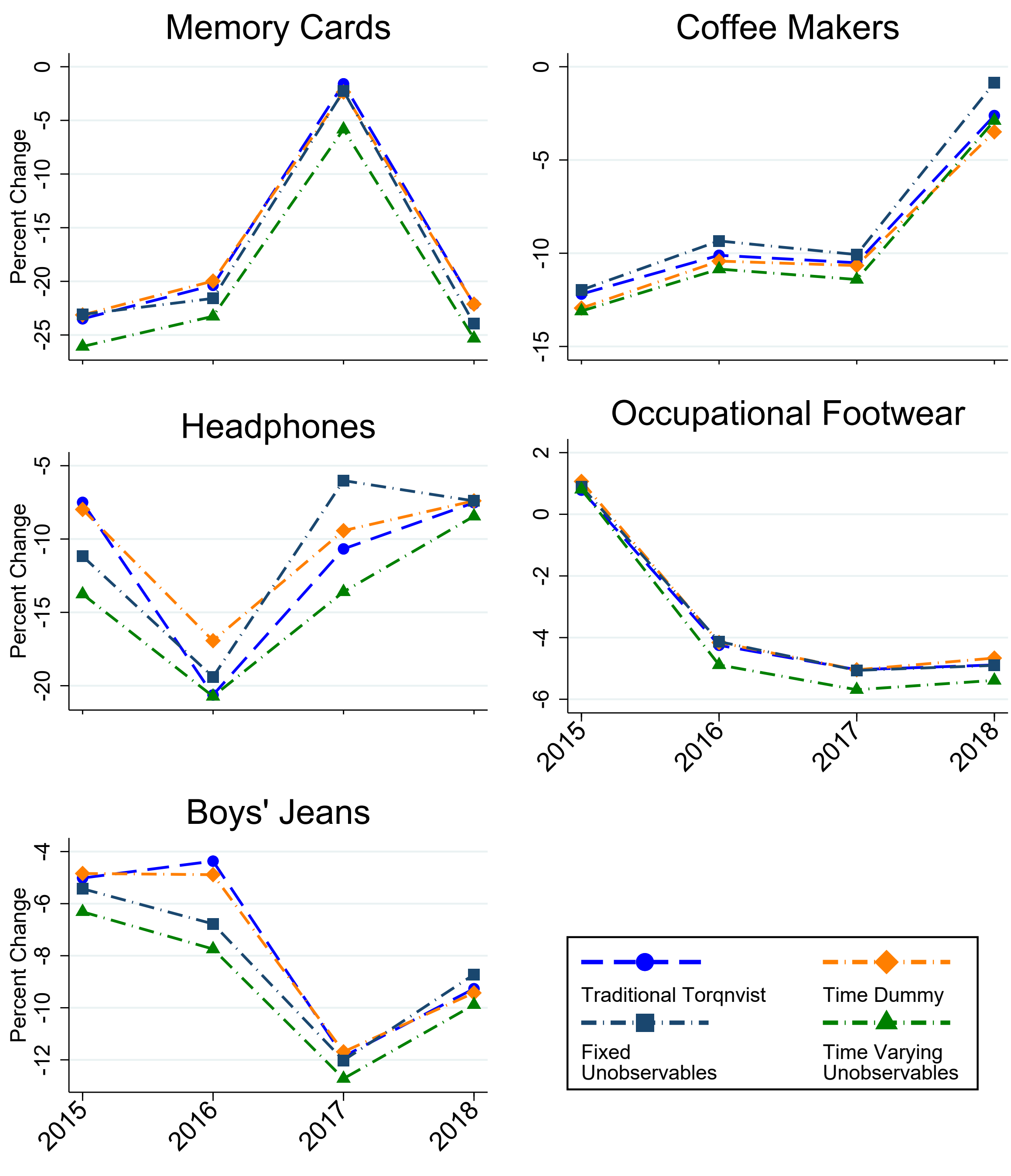
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices.

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



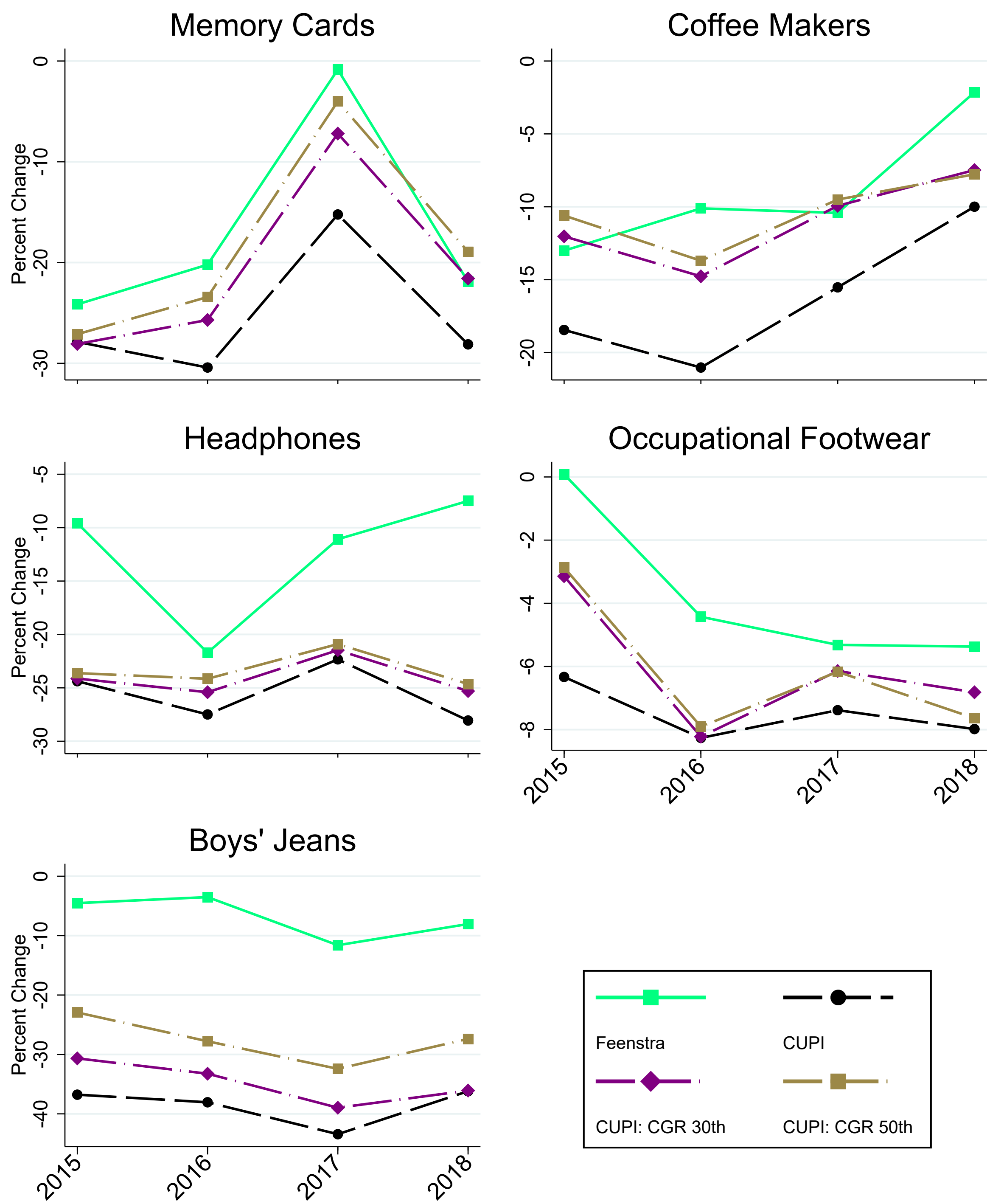
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices.

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.

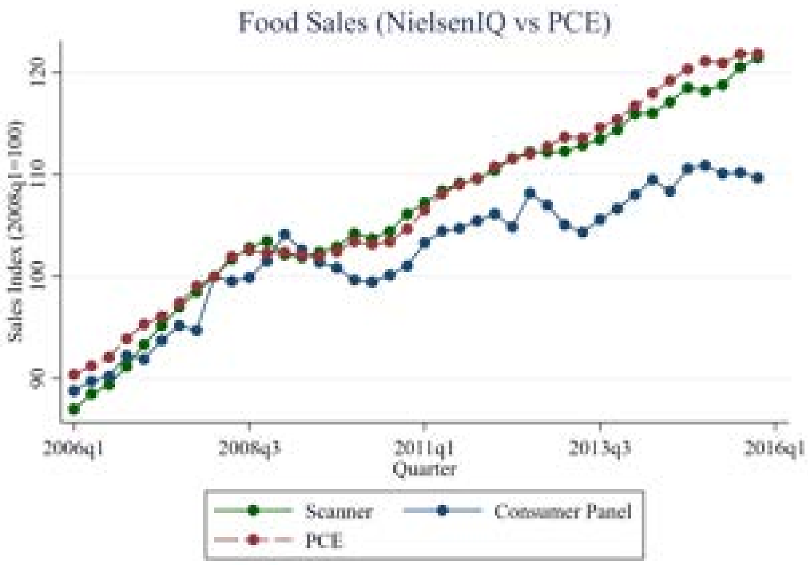
Figure D.5: CUPI, Sensitivity to Alternative Common Goods Rules (CGRs): NPD Data



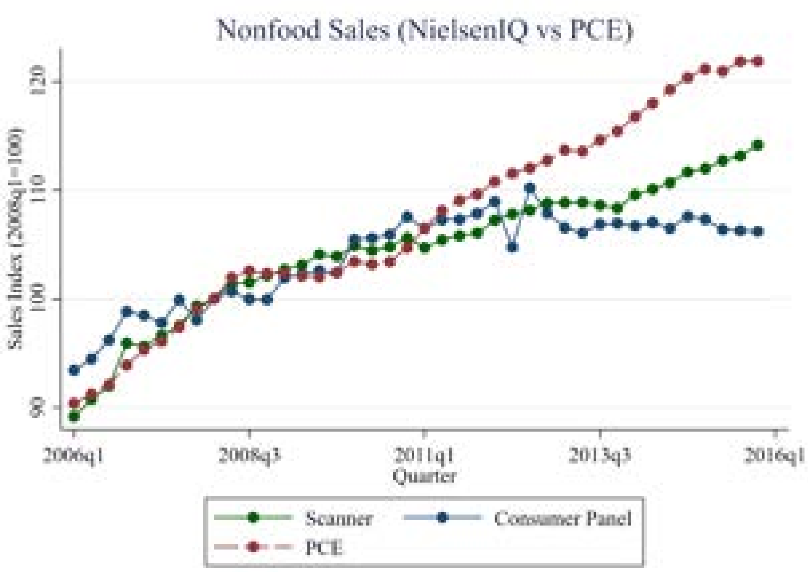
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. "CUPI: CGR Xth" excludes from the group of common goods those products with market shares below the Xth percentile in both periods. The Feenstra index is included for reference.

Figure D.6: PCE vs NielsenIQ Sales for Retail Scanner and Consumer Panel, Food and Nonfood

(a) Food

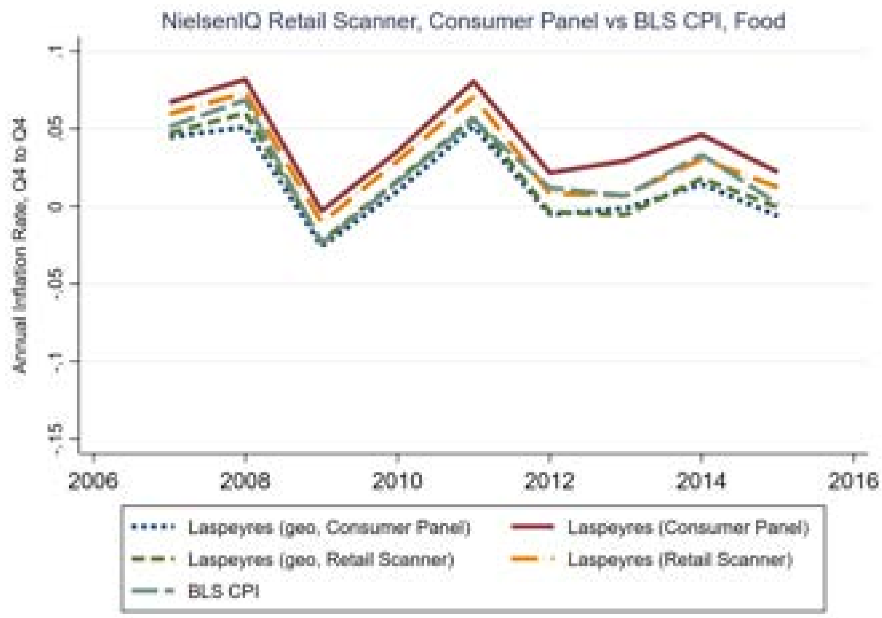


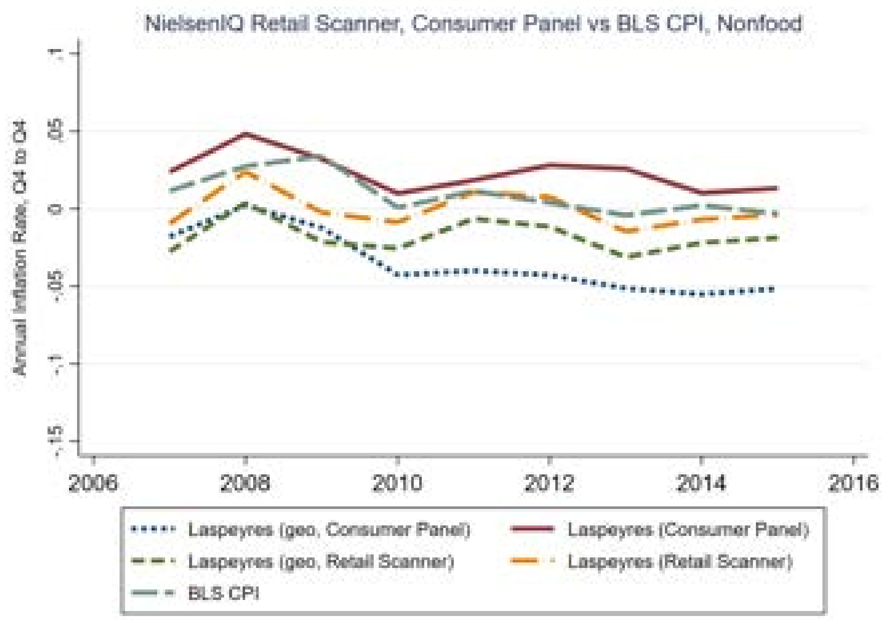
(b) Nonfood



Notes: NielsenIQ values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. Figures uses NielsenIQ Retail 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.

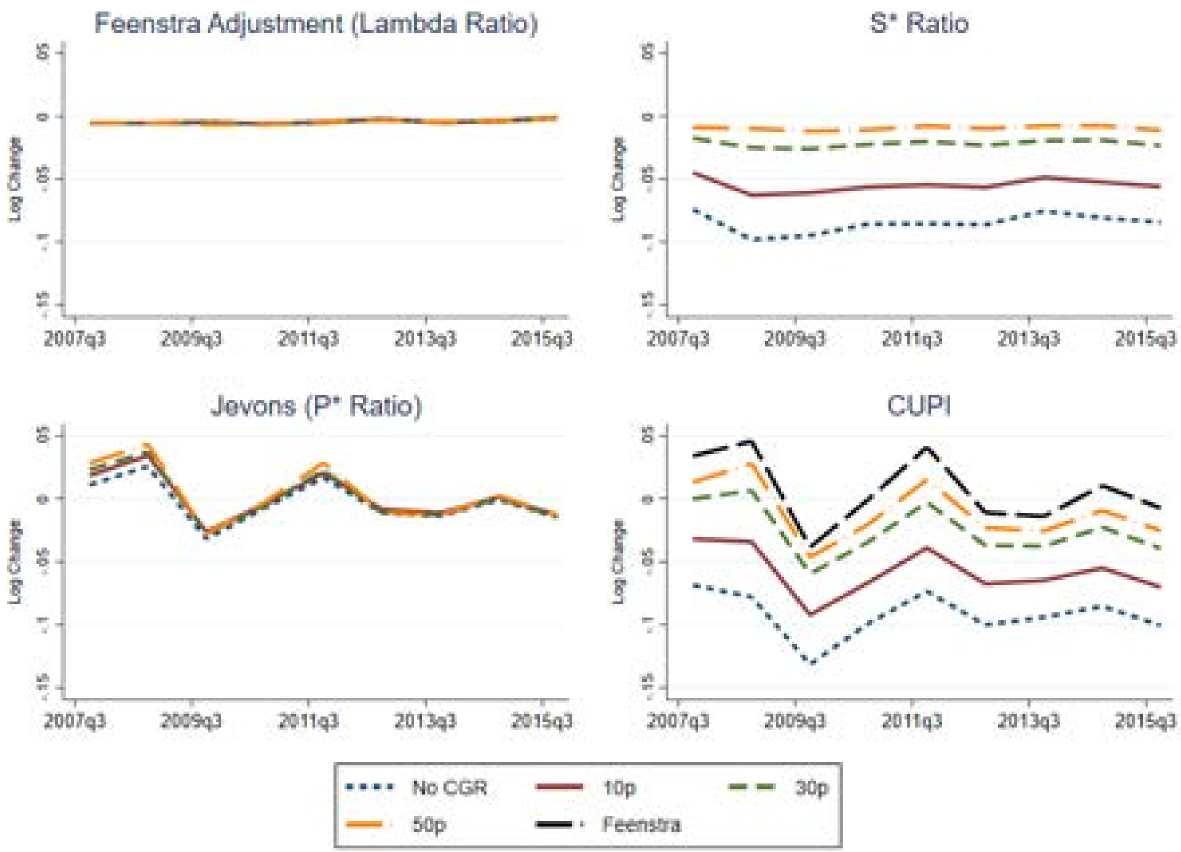
Figure D.7: BLS CPI vs NielsenIQ Laspeyres, Food and Nonfood





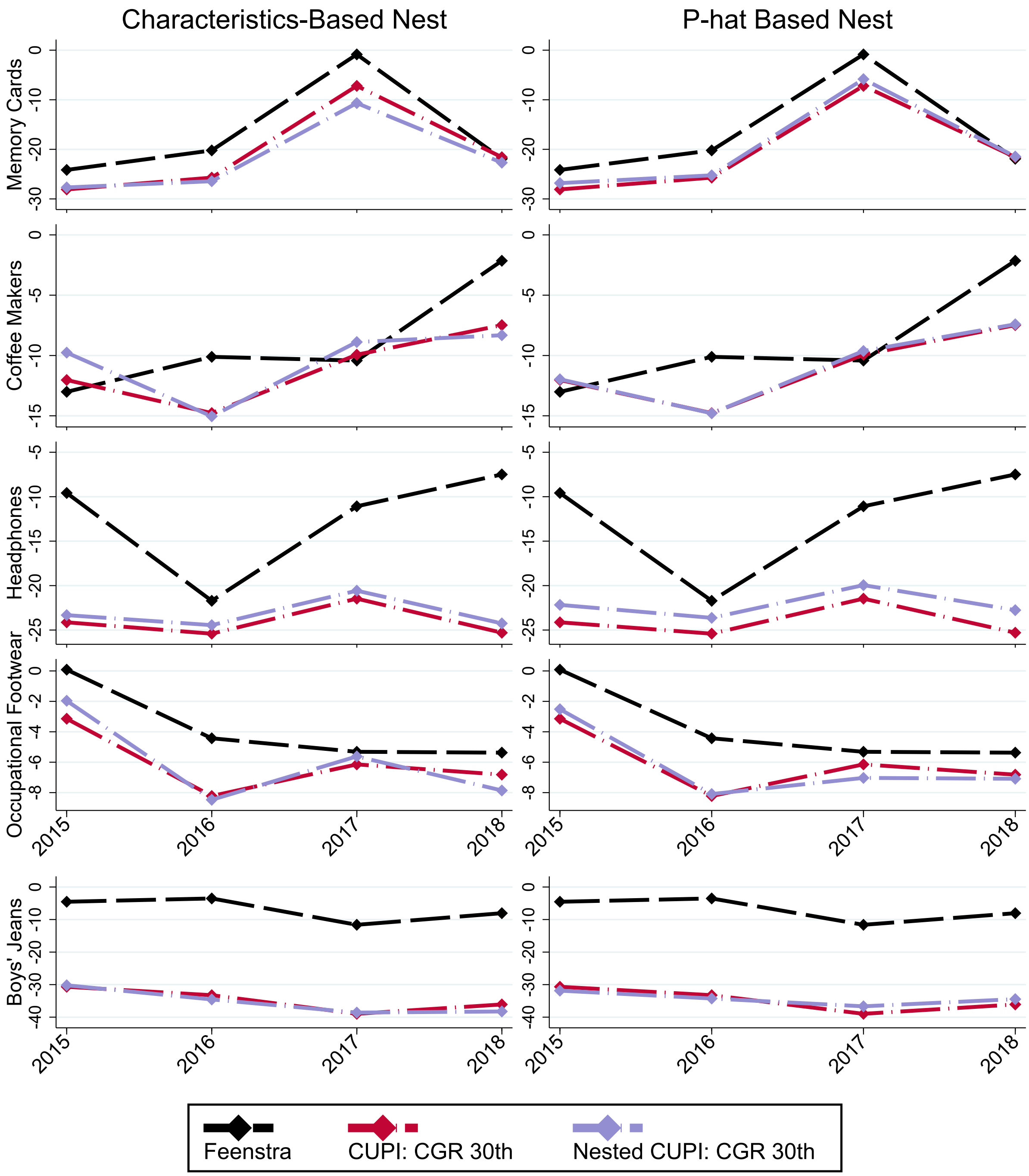
Notes: NielsenIQ values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. The panels use NielsenIQ Retail Scanner and Consumer Panel data for Food and Nonfood (aggregated) product groups. The BLS CPI was computed by BLS staff.

Figure D.8: CUPI and Its Components with Alternative CGRs: NielsenIQ Food



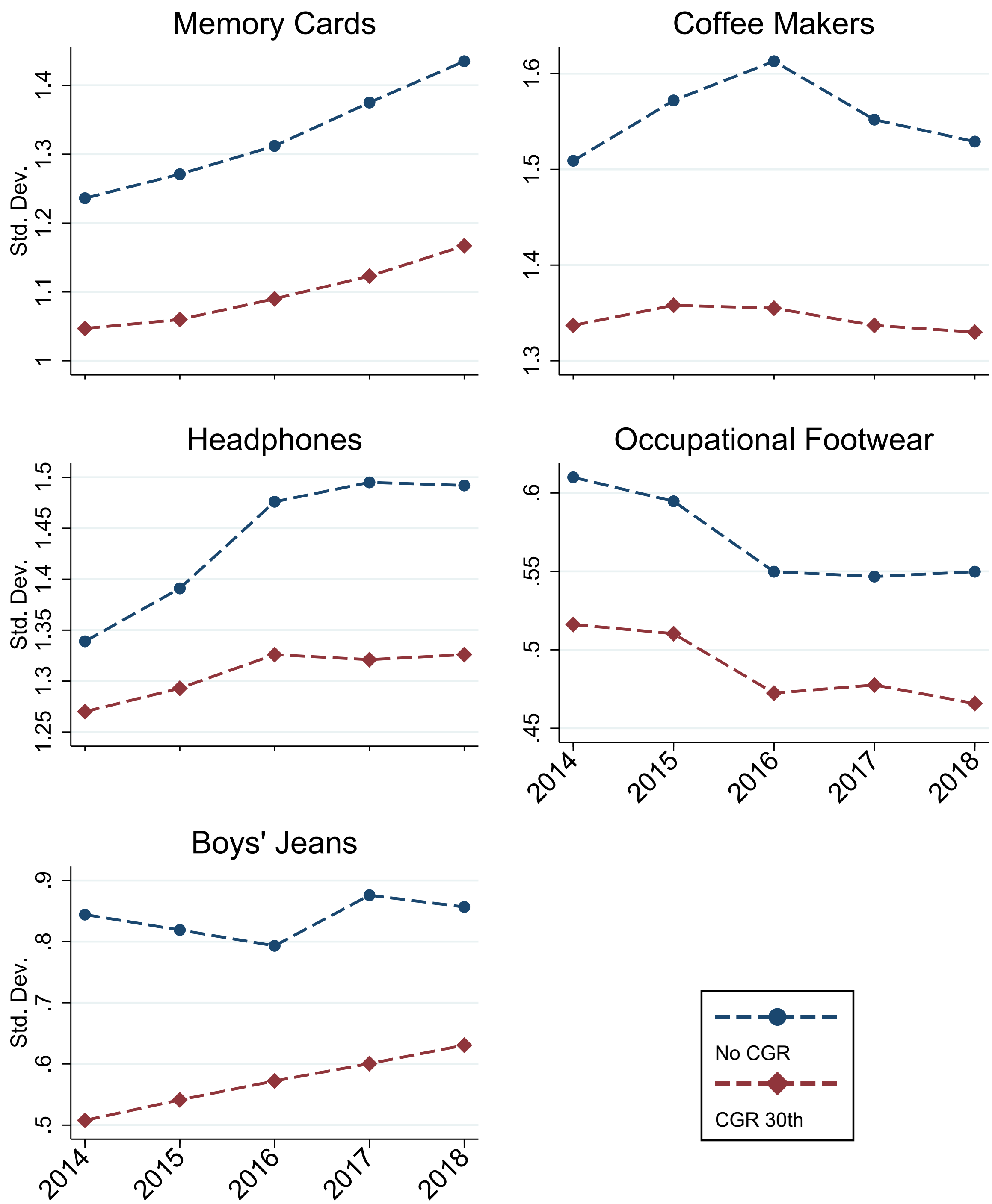
Notes: The figure shows NielsenIQ Retail Scanner data for food product groups. Each plot shows log changes from the fourth quarter of the previous year to the fourth quarter of the labeled year. The values are cumulative changes from chained quarterly indices. The Feenstra adjustment and S^* ratio panels 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.9: Nested CUPI: Characteristics- and P-Hat- Based Nests
Percent Changes: NPD Data



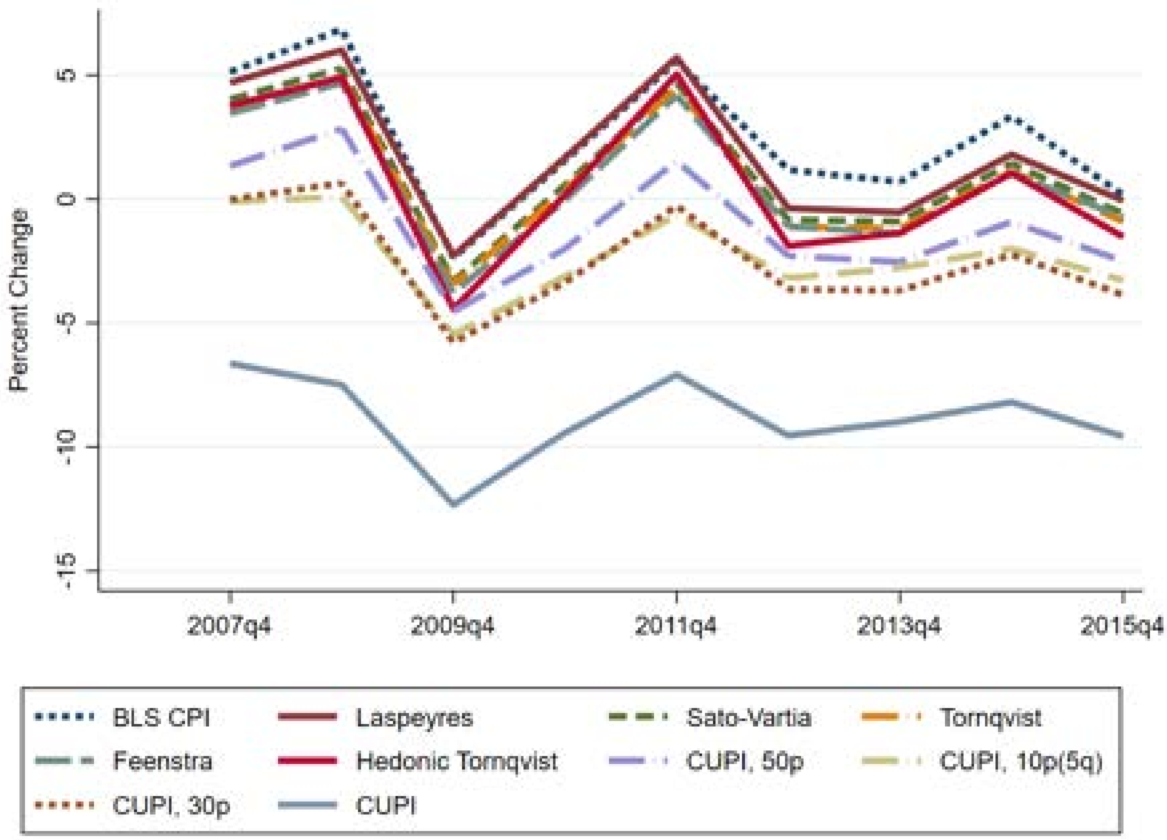
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. Laspeyres is the geometric Laspeyres. For the characteristics-based nests, we assign items to groups based on shared observable characteristics. The p-hat based nests are based on the decile of predicted 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.10: Dispersion of Relative Product Appeal: NPD Data



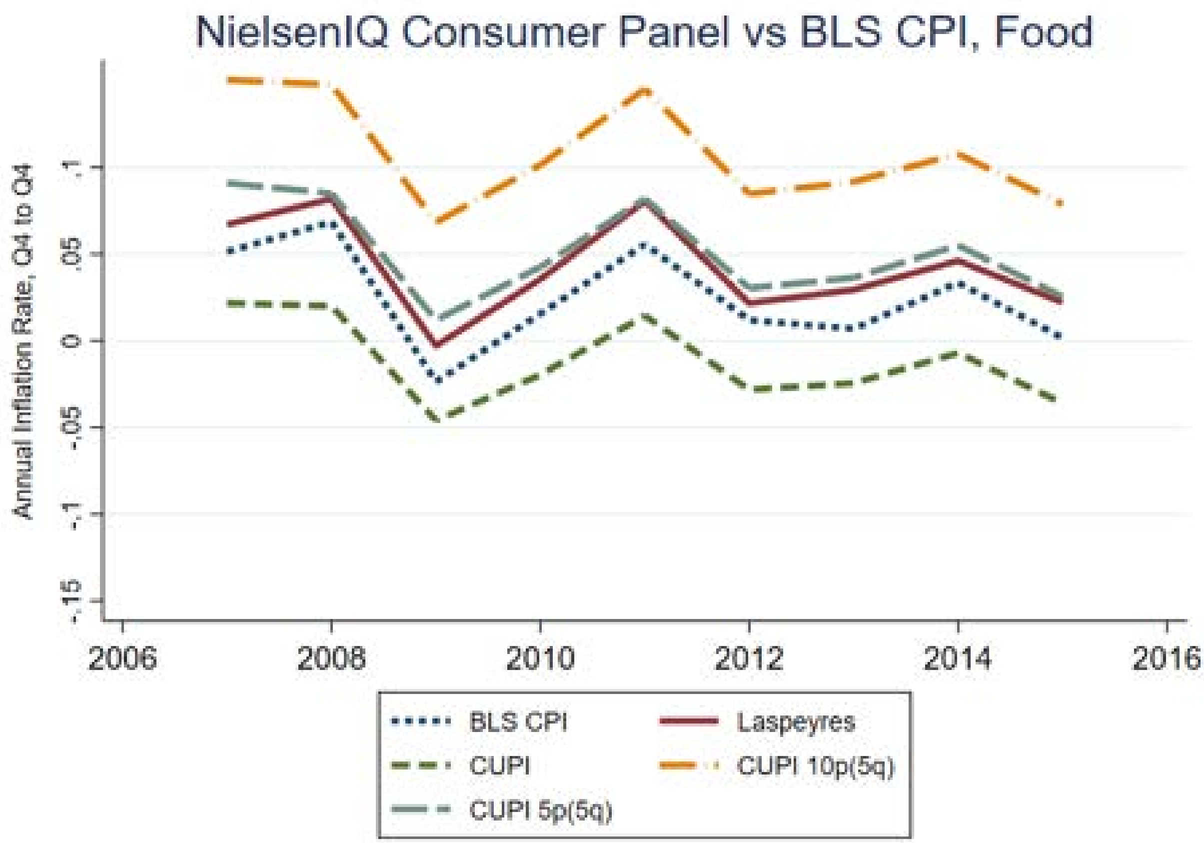
Notes: Values are annual averages of quarterly dispersion (standard deviation) in normalized log relative product appeal for common goods. Reported are the annual averages without imposing a common goods rule and also those with a 30th percentile common goods rule.

Figure D.11: Annual Inflation Rates: NielsenIQ Retail Scanner Data, Food



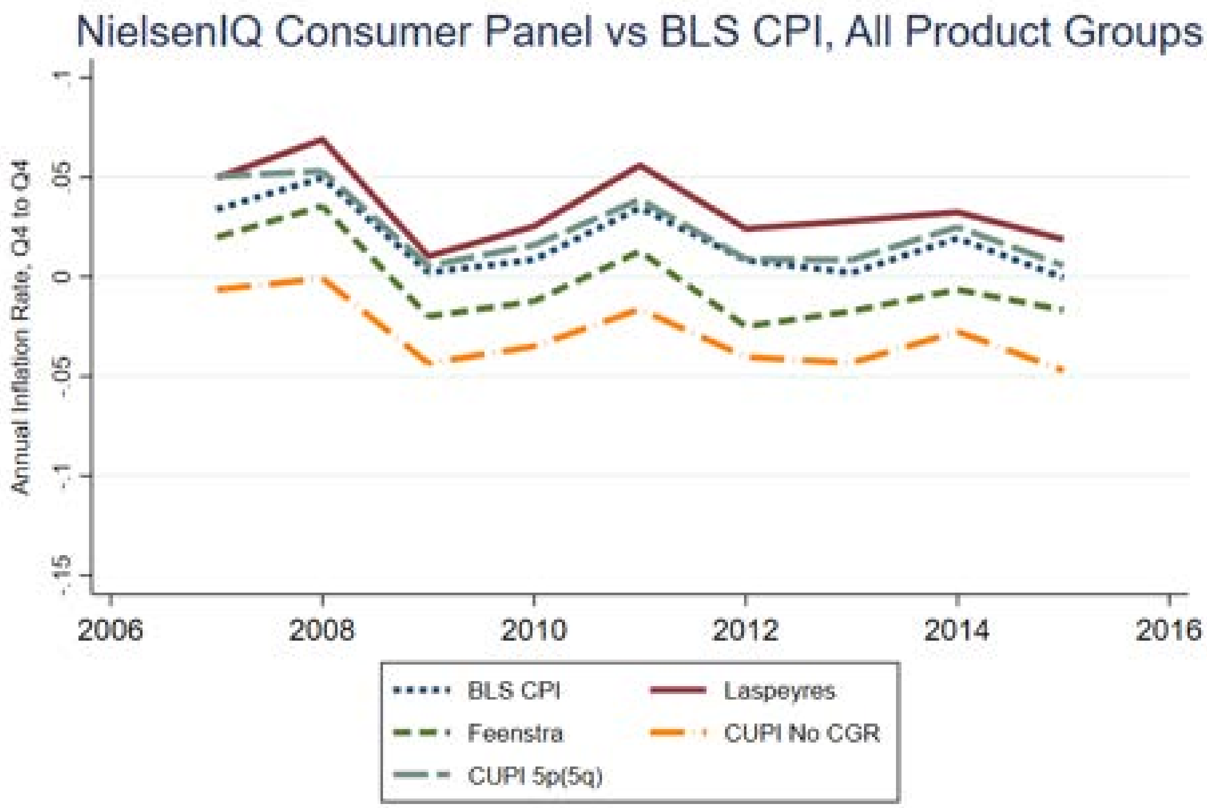
Notes: Figure uses NielsenIQ Retail Scanner data for food. The $2q$ CUPI computes CGR percentile thresholds using sales pooled over a two quarter horizon (t and $t-1$). The $5q$ CUPI computes CGR percentile thresholds using sales pooled over a 5 quarter horizon (current and prior 4 quarters). Laspeyres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

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



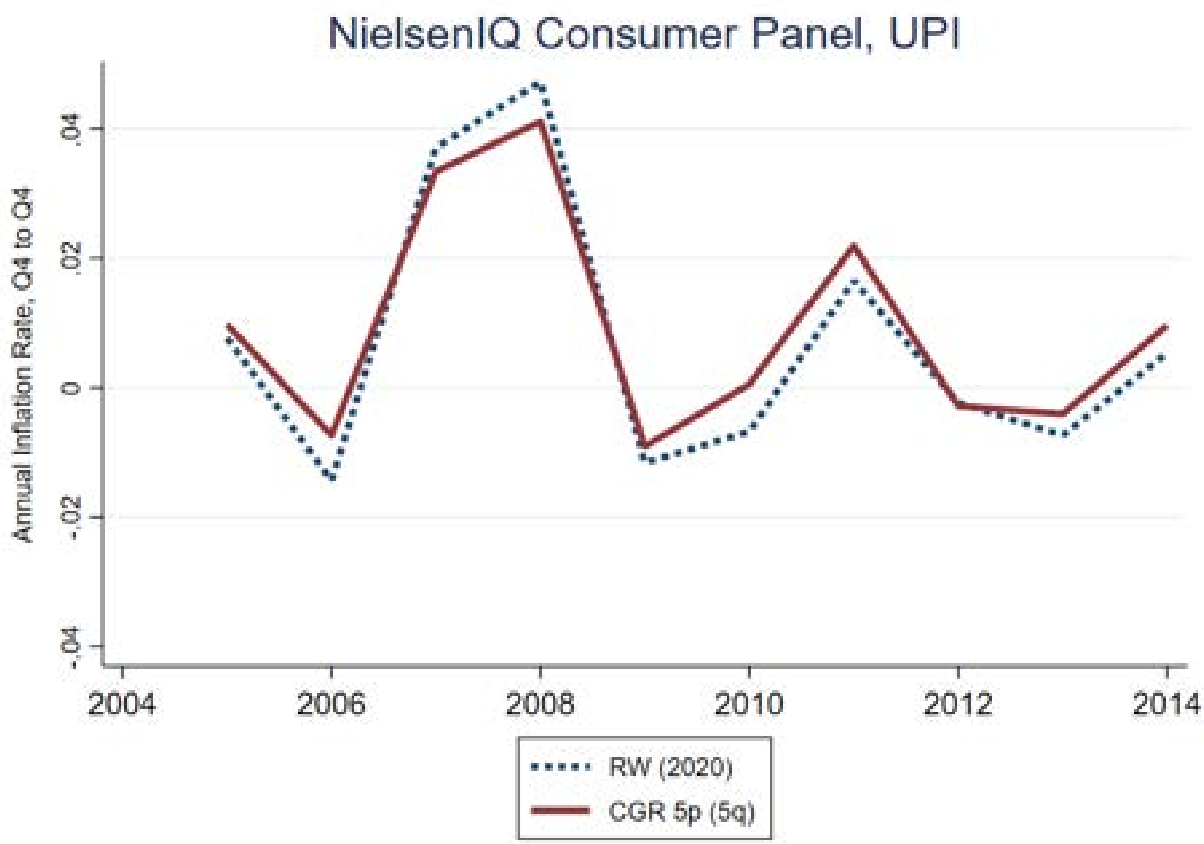
Notes: Figure uses NielsenIQ Consumer Panel data for food. The $5q$ CUPI computes CGR percentile thresholds using sales pooled over a five quarter horizon (t and $t-1$) (current and prior 4 quarters). Laspeyres is arithmetic.

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



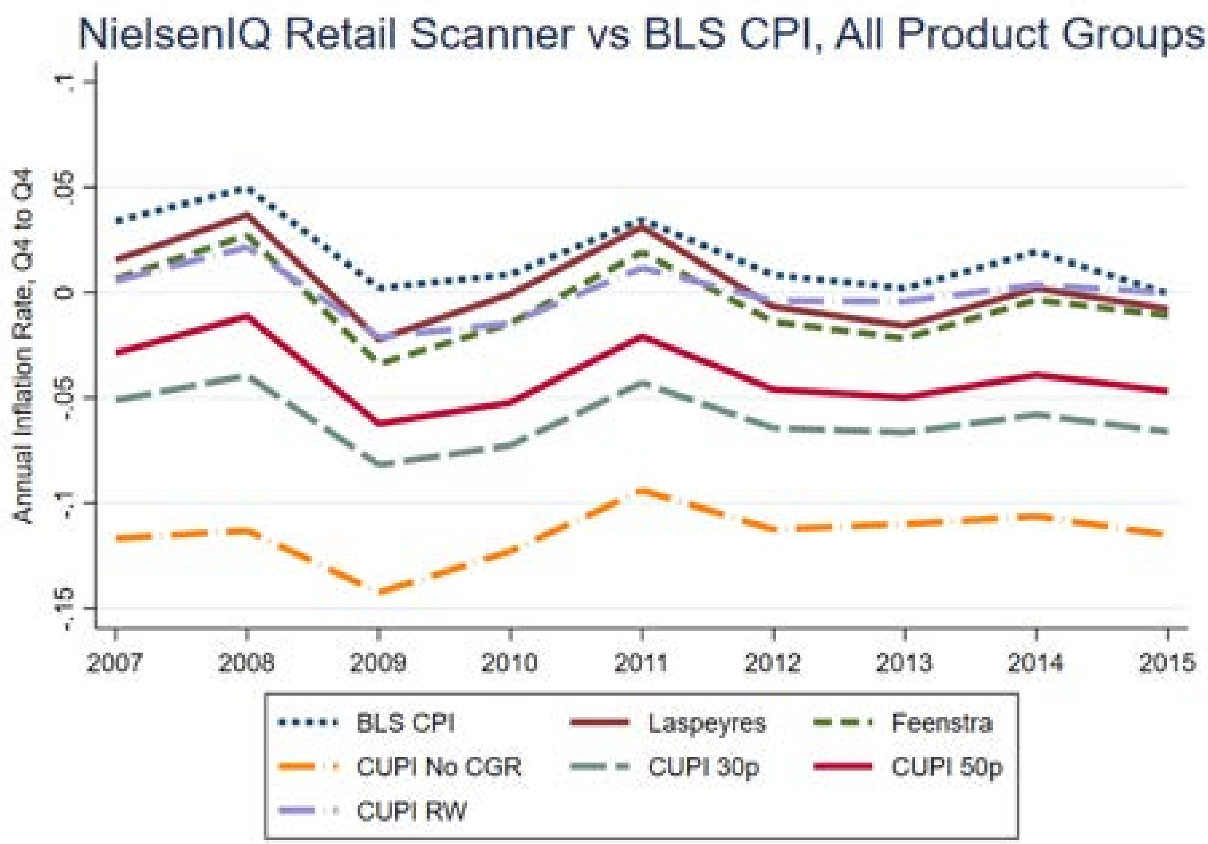
Notes: Figure uses NielsenIQ Consumer Panel data for food and nonfood product groups. The series “CUPI CGR RW” uses a 5th-percentile sales cutoff for the common goods rule. Percentile computed from sales pooled over 5 quarter horizon (current and prior 4 quarters). Laspeyres is arithmetic.

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



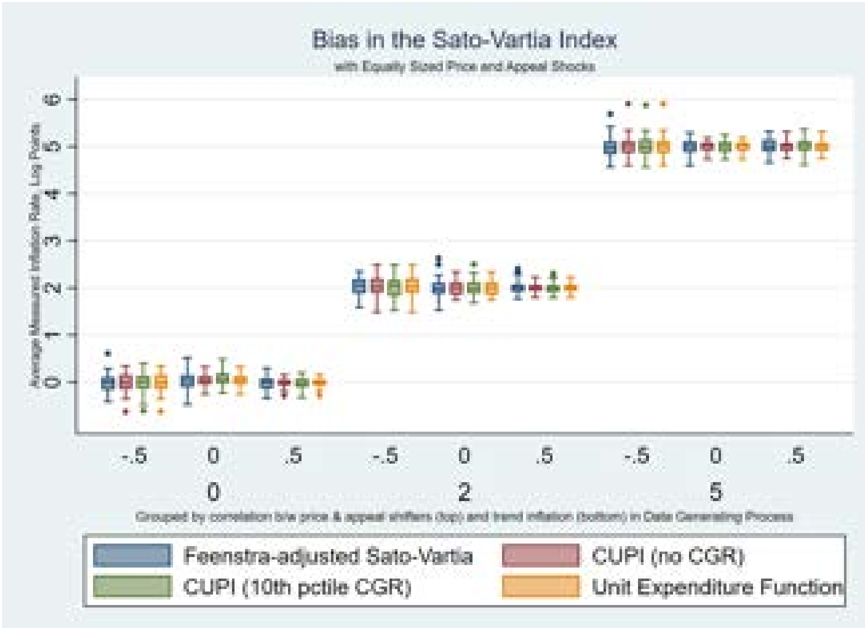
Notes: Figure uses NielsenIQ Consumer Panel for food and nonfood product groups. The indices are YoY for Q4. The series RW(2020) uses the same CGR duration rule as in Redding and Weinstein (2020). The series 5p(5q) use percentiles based on sales pooled over 5 quarter horizon (current and prior 4 quarters.

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



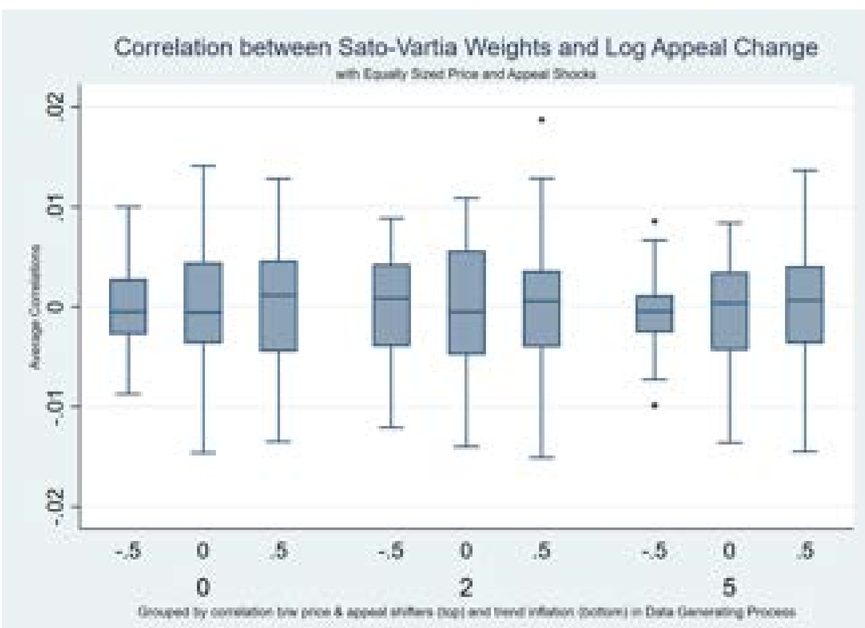
Notes: Figure uses NielsenIQ Retail Scanner data for food and nonfood product groups. The "CUPI, 25p" and "CUPI, 50p" series use 25th- and 50th-percentile cutoffs for the common goods rule, respectively. The series "CUPI, RW CP" uses the CGR 5th percentile threshold from the consumer Panel data for the 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



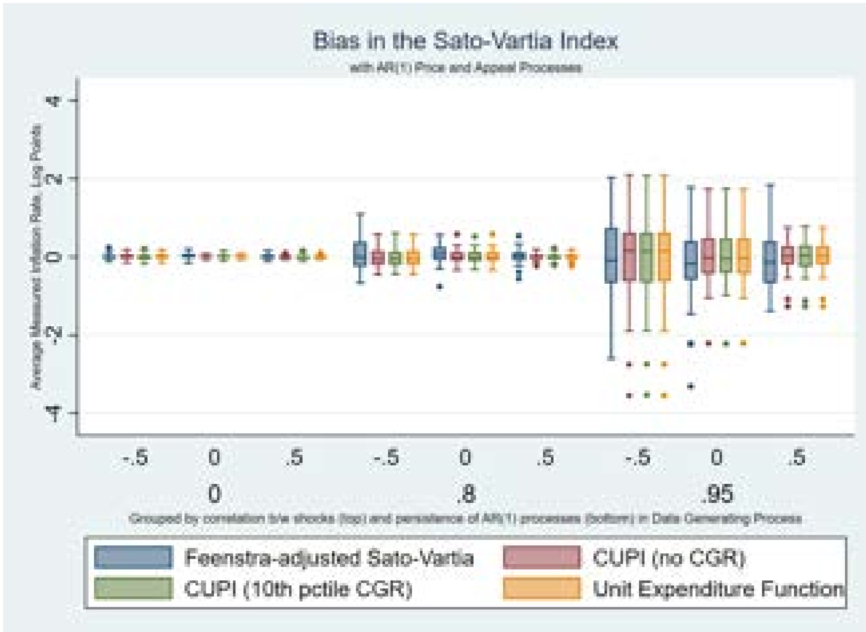
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.2.1. The 0, 2, and 5 along the horizontal axis specify the average trend rates of log price growth in log points, while the −0.5, 0, and 0.5 along the horizontal axis specify the correlations between 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



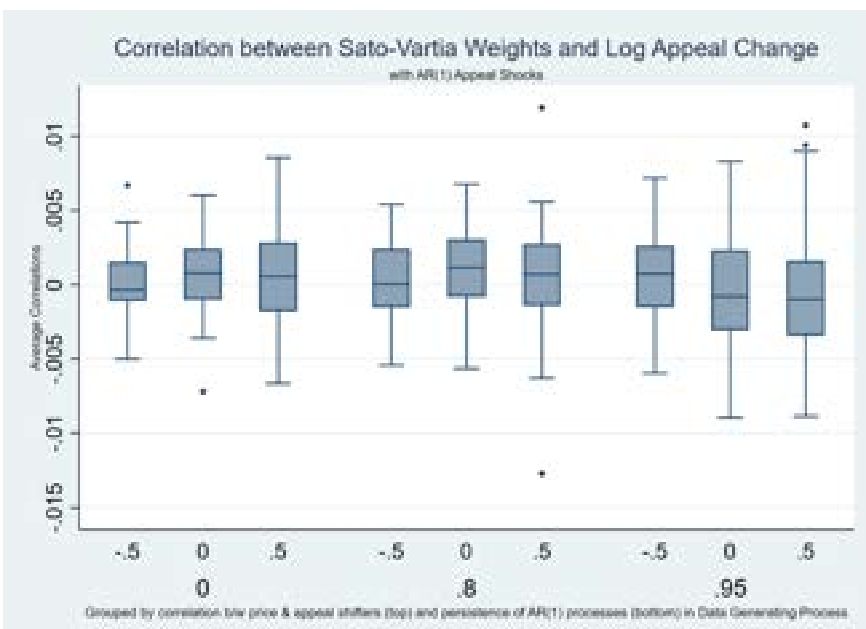
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.2.1. The 0, 2, and 5 along the horizontal axis specify the average trend rates of log price growth in log points, while the −0.5, 0, and 0.5 along the horizontal axis specify the correlations between 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



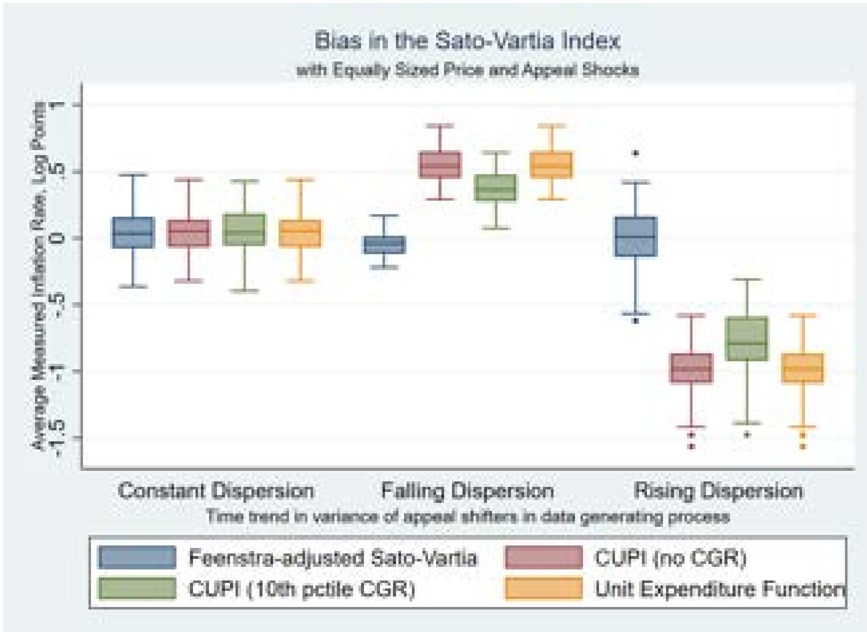
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.2.2. The 0, 0.8, and 0.85 along the horizontal axis specify the autocorrelations of log prices and log appeal parameters, while the −0.5, 0, and 0.5 along the horizontal axis specify the correlations 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



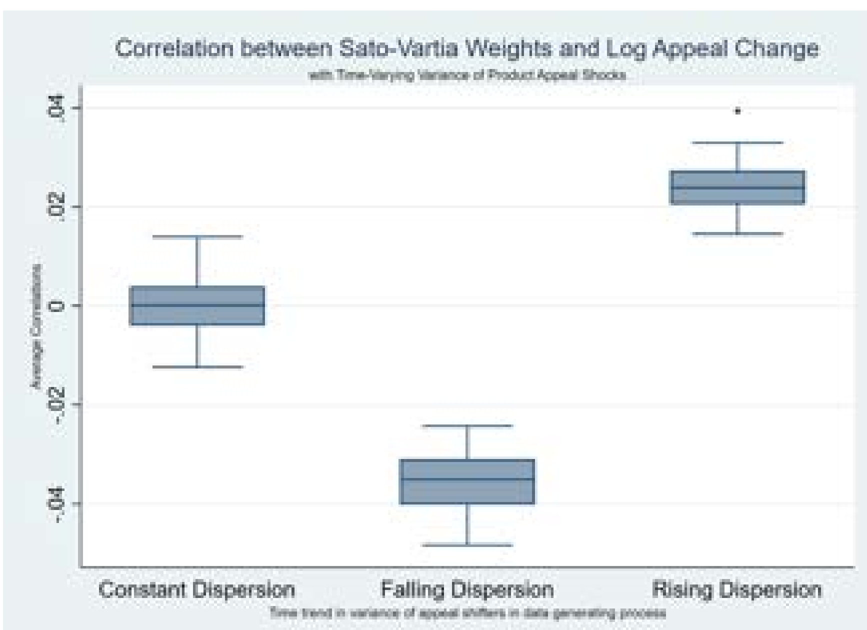
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.2.2. The 0, 0.8, and 0.85 along the horizontal axis specify the autocorrelations of log prices and log appeal parameters, while the −0.5, 0, and 0.5 along the horizontal axis specify the correlations 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



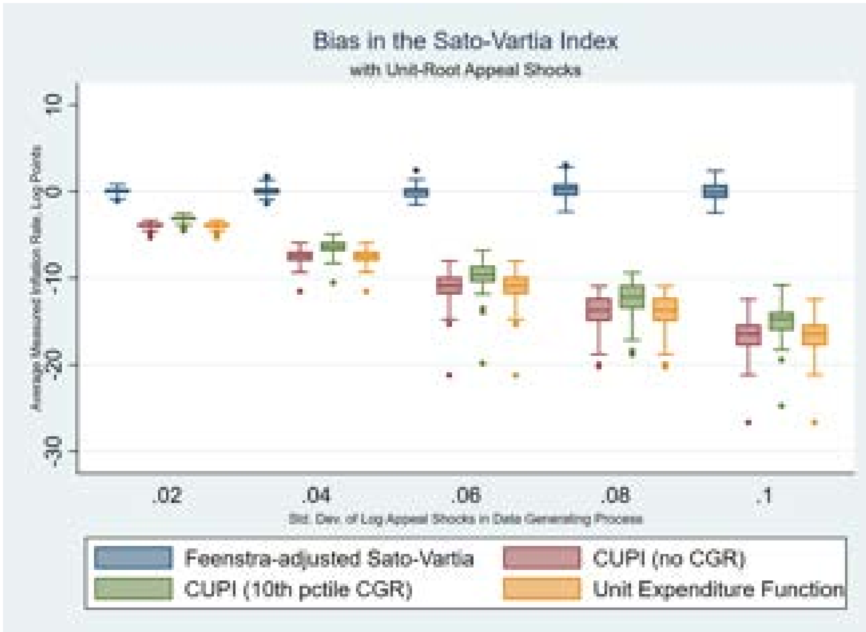
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.3.1. The variance of log product appeal draws is constant over time in the simulations labeled "Constant Dispersion," falling over time in the simulations labeled "Falling Dispersion," and rising over 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



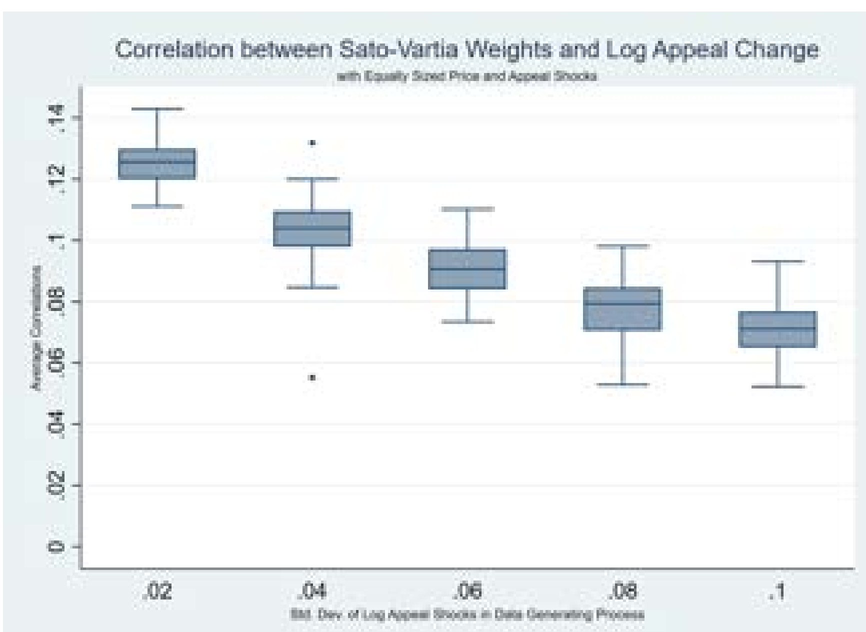
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.3.1. The variance of log product appeal draws is constant over time in the simulations labeled "Constant Dispersion," falling over time in the simulations labeled "Falling Dispersion," and rising over 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



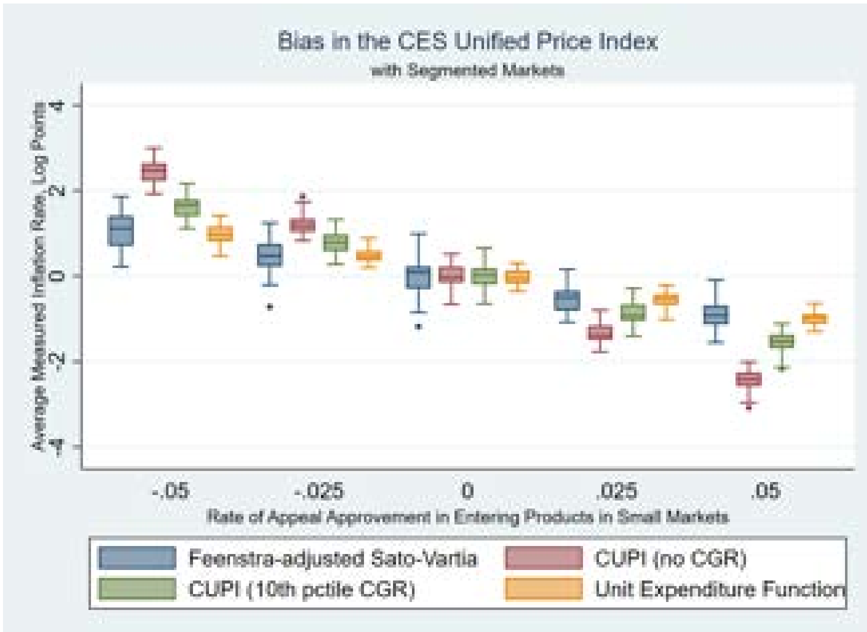
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.3.2. The horizontal axis displays the standard deviations of log appeal shocks. The vertical axis displays the average log inflation rate across simulation periods.

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



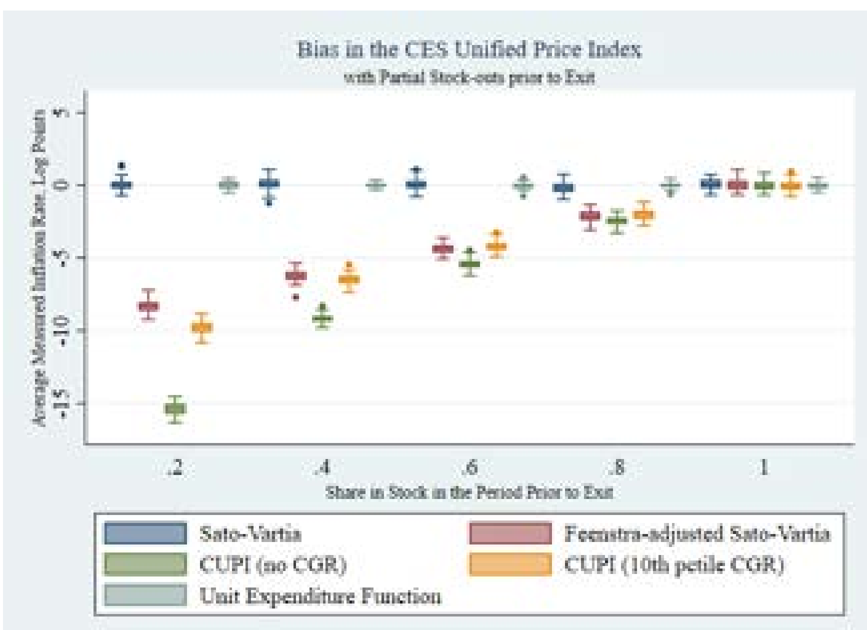
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.3.2. The horizontal axis displays the standard deviations of log appeal shocks. The vertical axis displays the average correlation between the Sato-Vartia weights and the log appeal shocks across simulation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