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

May 2023

Abstract

This paper explores alternative methods for adjusting price indices for quality change at scale. These methods can be applied to large-scale item-level transactions data that includes information on prices, quantities, and item attributes. The hedonic methods can take into account the changing valuations of both observable and unobservable characteristics in the presence of product turnover. The paper also considers demand-based approaches that take into account changing product quality from product turnover and changing appeal of continuing products. The paper provides evidence of substantial quality-adjustment in prices for a wide range of goods, including both high-tech consumer products and food products.

*Ehrlich: University of Michigan; Haltiwanger: University of Maryland and NBER; Jarmin: U.S. Census Bureau; Johnson: University of Michigan and National Academies of Sciences, Engineering and Medicine; Olivares: University of Maryland; Pardue: University of Maryland and Gusto; Shapiro: University of Michigan and NBER; Zhao: Bank of Canada and University of Maryland. 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, Robert Martin, Ariel Pakes, Stephen Redding, Marshall Reinsdorf, David Weinstein, and participants at multiple seminars and conferences for helpful comments. Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business.The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. We also use the NPD data housed at the U.S. Census Bureau. All results using the NPD data have been reviewed to ensure that no confidential information has been disclosed (CBDRB-FY19-122, CBDRB-FY21-074, and CBDRB-FY23-067). Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the view of the U.S. Census Bureau.

1 Introduction

Retail businesses create item-level data on the prices and quantities of the goods that they sell. Such data form the basis for re-engineering key economic indicators by building internally consistent aggregates of value, volume, and price directly from item-level transactions. Aggregation of transactions data could supplant traditional surveys and enumerations conducted by statistical agencies (see, e.g., Ehrlich et al., 2022). There are many potential advantages to such a re-engineering. One is to address the issue of rapid product turnover, which is likely to be associated with quality improvements. Current statistical agency procedures for measuring prices inadequately address such turnover. This paper considers scalable procedures for constructing price indices using item-level transactions data that can account for entering and exiting goods as well as changing consumer valuations of product attributes.

The use of high-frequency, item-level sales data to produce accurate inflation measures also requires incorporation of advances in index number and economic theory. We consider two complementary approaches: hedonics and demand-based models. Both approaches suggest that quality improvement is widespread across a large range of consumer goods, including in categories in which technological progress is not immediately visible.

Our preferred hedonic approach builds on the insights of Erickson and Pakes (2011, hereafter "EP"), who develop a novel method of calculating hedonic price indices that can account for changing valuations of both observable and unobservable product characteristics. High-frequency, item-level transactions data with prices, quantities, and attributes permit the implementation of the EP hedonics approach with superlative price indices (such as the Fisher or Tornqvist) in real time using internally consistent expenditure weights. We show that the EP methodology has important advantages relative to commonly used alternative hedonic methods such as the time dummy method.

Our 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 product entry and exit (denoted the Feenstra index hereafter); and the Constant Elasticity of Substitution (CES) Unified Price Index (CUPI) developed in Redding and Weinstein (2020), which adjusts the Feenstra index for changing consumer preferences. The demand-based approaches have the attractive feature that they yield exact price indices under certain sets of assumptions. Moreover, in principle these methods impose sufficient structure that they can be implemented without attribute data beyond a product taxonomy. That said, a challenge for the demand-based models is that they assume a specific functional form for consumer preferences, which may or may not hold in the data.

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

We implement the hedonic and demand-based approaches at scale using item-level transactions data from two major sources. The first is from the NPD Group, which covers a wide range of general merchandise goods from bricks and mortar and online retailers. In this paper, we construct price indices for a select number of product groups: memory cards, headphones, coffee makers, boys' jeans, and work and occupational footwear. The NPD data include rich product attributes, which facilitate the implementation of the EP methodology. The second platform we use is the Nielsen Marketing data provided by the Kilts Center at the University of Chicago Booth School of Business, which covers a wide range of food products from grocery stores, discount stores, pharmacies and liquor stores. The Nielsen data have only sparse information on product attributes, which we address by adapting the EP methodology to a machine learning (ML) framework. Our companion paper, Cafarella, Ehrlich, Gao, Shapiro, and Zhao (2023) develops our machine learning approach.[1]

Consistent with the literature using scanner data, we find rapid product turnover, along with rich post-entry product life-cycle dynamics. Products' market shares peak several quarters after entry, while on average prices decline monotonically after entry. We also find evidence of substantial quality adjustment in price indices using either hedonic or demand-based approaches across the full range of product groups we consider. The magnitude of the quality adjustment is greater for high-tech goods such as memory cards and headphones, but we find that quality adjustment is pervasive for food product groups as well. While the latter result might be surprising, our findings are consistent with the changing and increasing variety of food products available over time.

We find that the EP method of hedonic adjustment, which can account for unobservable product characteristics, yields more systematic evidence of pervasive quality adjustment than the time dummy hedonic method. The EP method also produces a meaningful improvement in our hedonic regressions' goodness of fit. Currently, the Bureau of Labor Statistics (BLS) uses hedonic adjustment for only about 7.5 percent of goods and commodities in the CPI (our estimates based on Bureau of Labor Statistics, 2023). Our results extend the EP methodology from the CPI database, which does not contain item-level quantities or expenditures, to transactions data with internally consistent prices and quantities. We show that the advantages of the EP approach extend to our data environment, bolstering the case for the applicability of these data and methods in official statistics. With the techniques used in this paper and with the ubiquitous availability of item-level transactions data, there is increasingly little reason why statistical agencies should not quality-adjust most goods and services.

In evaluating these alternatives, we also consider the impact of chain drift, which is a known issue for chained price indices from scanner data. We find that using standard (e.g.,

[1] Cafarella et al. (2023) focuses on the methodological aspects of our architecture for applying machine learning to hedonic price indices. The comparison of those results to demand-based methods and all of the NPD-based results are original to this paper. Cafarella et al. (2023) is available at http://www-personal.umich.edu/~shapiro/papers/ML_Hedonics.pdf.

GEKS) adjustments to address chain drift tends to increase the computed rates of inflation, especially for traditional price indices. Interestingly, the EP-based hedonic price indices are substantially less subject to chain drift.

Among the demand-based methods, we find that the Feenstra index systematically measures lower price inflation than the Sato-Vartia. This result suggests that product turnover is associated with quality improvement. We find that the CUPI, which generalizes the Feenstra index to allow for changing product appeal over product life cycles, implies substantial additional quality adjustment relative to the Feenstra index. We find that this rank ordering holds with adjustments for chain drift.

A challenge in implementing the CUPI is that two of its three components are unweighted geometric means. These terms are sensitive to the inclusion of goods with very small quantities sold or market shares, which is one reason that unweighted indices are generally discouraged in the index number literature. Redding and Weinstein (2020) employ a reallocation procedure by which they move a subset of goods out of the CUPI's unweighted geometric mean terms and into the Feenstra (1994) adjustment term using what we term a *common goods rule* based on the durations of products' time in the marketplace. Applying a common goods rule brings the CUPI's measurement of price changes closer in line with other indices. Our results suggest more research is needed to provide guidance about how to define common goods rules.

2 Data

This section provides an overview of the two data sets that we use to compute price indices. The first comes from the NPD Group and the second comes from Nielsen. For both data infrastrucures, we aggregate the item-level transactions data to the quarterly, national level and focus on quarterly price indices. This approach facilitates comparing traditional, hedonic, and demand-based price indices in a manner consistent with the recent literature.

2.1 NPD Data

We use proprietary data that the NPD Group provided to the U.S. Census Bureau. They consist of monthly sales and quantity data at the product-store level from 2014 through 2018.[2] The NPD group tracks more than 65,000 retail stores, including online retailers. The retail stores cover a wide range of general merchandise products. 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"). The NPD data have unique item-level identifiers that are consistent cross-sectionally and over time. We aggregate the item-by-store-level observations to the national product-quarter level and calculate total quantity sold and average price for each product-quarter. The item-level data cover tens of thousands of product-quarter-level observations.

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

Table 1 displays average item-level product turnover rates for each product group. Each of the groups exhibits a high degree of product turnover, ranging from 4.5 percent to 13.5 percent in terms of quarterly entry and exit rates. Figure 1 presents life-cycle dynamics of product market shares and prices within these product groups. The illustrated statistics

[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.

are mean log differences from the product-specific initial values upon entry. Prices decline steadily after entry, while market shares exhibit a hump-shaped pattern post-entry. The post-entry patterns of market shares differ considerably across product groups. For example, while memory cards, coffeemakers and headphones all peak after 3 quarters, headphones decline much more rapidly than memory cards or coffeemakers. Magnitudes at the peak are large but also differ by product group. For memory cards and coffeemakers the peak is about 300 log points relative to the first quarter while for headphones the peak is about 200 log points.

Taken together, these findings highlight two important features of the data. First, there is considerable item-level product turnover that is a potentially important source of changing product quality. Second, post-entry dynamics suggest that it may be important for methods of quality adjustment to account for time-varying product appeal. Both the hedonic and demand-based approaches we consider can account for such variation.

2.2 Nielsen Data

We use the Nielsen Retail Scanner data (also referred to as RMS) from the Kilts Center for Marketing at the University of Chicago Booth School of Business for the period 2006 to 2015. The data consist of over 2.6 million products identified by the finest level of aggregation–12-digit universal product codes (UPCs) that uniquely identify specific goods. The Nielsen Retail Scanner data are collected from over 40,000 individual stores from approximately 90 retail chains in over 370 metropolitan statistical areas (MSAs) in the United States. Total sales in Nielsen RMS are worth over $200 billion per year and represent 53% of all sales in grocery stores, 55% in drug stores, 32% in mass merchandisers, and 2% in convenience stores.

Nielsen 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.6 and D.7 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.

We focus on results for the Nielsen data's food product groups because we estimate that the data's coverage is more extensive and tracks economywide time trends more closely for those groups than for the nonfood product groups. Appendix B.2 describes the analysis underlying that conclusion. A related implication of that analysis is that the Nielsen Retail Scanner data more closely tracks economy-wide trends in both sales and prices for overall food than the Nielsen Consumer panel data.

3 Conceptual Framework

Time series price indices aim to measure approximately or exactly the change in the cost of living between two or more time periods. One important challenge in constructing price indices from item-level data is the substantial pace of product turnover that we documented in the previous section. Another important challenge is that consumer preferences over products or valuations of different product characteristics may vary over time. Traditional "matched-model" price indices do not capture quality change from such product turnover or from changing relative product appeal. In contrast, the hedonic and demand-based indices we construct have the potential to capture these changes.

3.1 Traditional Price Indices

Our empirical work in this paper focuses on so-called geometric price indices, which are weighted averages of log price changes. We will focus in particular on the Tornqvist index, given by

$$\ln \Phi_t^{TQ} \equiv \sum_{k \in \mathbb{C}_t} \frac{s^*_{kt\text{-}1} + s^*_{kt}}{2} \ln \frac{p_{kt}}{p_{kt\text{-}1}}. \tag{1}$$

The set $\mathbb{C}_t$ in equation (1) is the set of all "continuing" or "common" goods that are sold both in period t and in period t-1, while s^*_{kt} and $s^*_{kt\text{-}1}$ denote product k's share of expenditures in those periods, respectively, among the set of common goods $\mathbb{C}_t$. p_{kt} and $p_{kt\text{-}1}$ denote product k's average unit prices in those two periods; their log ratio is often called a log price relative.

The Tornqvist index has multiple attractive properties. First, as a "superlative" price index, it is closely related to other superlative price indices such as the Fisher and Sato-Vartia.[3] We show that the Tornqvist and Sato-Vartia indices track each other closely in our empirical results. 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. Appendix A.1 describes the interpretation of traditional price indices in more detail.

An important limitation of traditional price indices in the context of transactions data 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. Traditional indices therefore do not account directly for goods that enter or exit across periods, which we argue is an important source of changing product quality and is ubiquitous in item-level data. Another important limitation of traditional price indices in current practice is that statistical agencies' data on sales and expenditure shares is often limited to disparate sources at higher levels of aggregation and lower frequency. For instance, the BLS uses expenditure shares from the Consumer Expenditure survey, with infrequently updated weights, to produce the Consumer Price Index (CPI). High-frequency scanner data connect the prices and quantities sold for each product, which allows for the construction of superlative price indices using internally

[3] Every superlative price index is the change in the unit expenditure function (i.e., the exact price index) that is a second-order approximation for a wide class of utility functions in the absence of product turnover and taste shocks (Diewert, 1978). 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.

consistent price and quantity data. We explore this advantage in our empirical analysis.

3.2 Hedonic Price Indices

Hedonic imputation allows a price index to account for product turnover by using product characteristics and an estimated hedonic relationship between characteristics and prices to impute the "missing" prices for entering and exiting products.

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

$$\ln p_{kt} = h_t(Z_k) + \eta_{kt}, \tag{2}$$

where Z_k is a vector of observable characteristics for good k. The function $h_t()$ is often linear in parameters, and the hedonic equation is estimated with with ordinary or weighted least squares regression. An important feature of equation (2) 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).[4]

A core limitation of the log-level hedonic estimation approach outlined in equation (2) is that there are likely to be product characteristics that are relevant to the formation of prices but that the econometrician cannot observe. Erickson and Pakes (2011) introduce hedonic methods that can account for such unobserved characteristics. An important first step is to estimate hedonic models of price changes rather than price levels, e.g.

$$\Delta \ln p_{kt} = Z_k' \beta_t + v_{kt}. \tag{3}$$

This log-difference hedonic model estimates the change in hedonic price coefficients directly,

[4]Although Pakes (2003) emphasizes that the estimated coefficients are not readily interpretable as marginal valuations of characteristics, the indices that emerge can be used as quality-adjusted estimates of changes in the cost of living.

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) therefore 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, we estimate the log-level hedonic specification in equation (2) for period t-1. Second, we estimate a log-difference hedonic model including the lagged residuals from the first stage.[5] The second estimating equation is then

$$\Delta \ln p_{kt} = Z_k' \beta_t + \kappa_t \hat{\eta}_{kt\text{-}1} + \upsilon_{kt}. \tag{4}$$

Including the initial residuals from equation (2) in equation (4) 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.

The Nielsen data provided by the Kilts Center for Marketing at the University of Chicago does not contain pre-coded product attribute data for most products aside from short textual product descriptions. To address these challenges, Cafarella et al. (2023) implement deep neural networks to predict product prices and price changes from the product descriptions in the Nielsen Kilts Center data. We draw on results from that paper for the quality-adjusted price indexes for the Nielsen food data reported in this paper. Appendix B.4 briefly summarizes the ML approach. It parallels the EP-TV approach, in that it first predicts price

[5] It can be shown that 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 (4).

levels and then, to capture time-varying unobservable effects, uses the prediction error in a second-stage neural net predicting price changes. In related work, Bajari et al. (2021) use an advanced machine learning approach that includes encoding image data as inputs into price predictions.[6]

Our hedonic methods—both the econometric approach with NPD data implemented in this paper and the ML approach with Nielsen data implemented in the companion paper—use quantity-share weights *for the estimation*. The indices are, of course, constructed using expenditure weights. In Appendix A.2, we also consider EP-TV hedonics using expenditure weights. Our main conclusions are broadly similar using both sets of estimation weights (see Table D.3). Bajari et al. (2021) also uses 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.

In our main analysis using the EP-TV approach, we assume the lagged residual for an entering good in the period prior to entry is zero. As a robustness check, in Table D.1, we find very similar results if we replace the predicted price relatives for entering goods with those from a hedonic regression that uses current period rather than lagged residuals and is otherwise equivalent to equation (4).[7]

We focus on full-imputation versions of the hedonic imputation indices that use predicted price relatives for all observations, including for common goods. Pakes (2003) shows that the hedonic Laspeyres imputation index bounds the exact change in the cost of living under a

[6]Bajari et al. (2021) provide novel methodology for encoding images via machine learning. They estimate hedonic models of price levels period by period.

[7]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. We also consider the difference between the traditional and hedonic Laspeyres using the EP-TV method below. We find the differences are similar to the analogous differences using the Tornqvist indices. The Laspeyres indices have other limitations, but they are not sensitive to this assumption about the residual prior to product entry.

relatively weak set of assumptions. The key assumption is that consumers have preferences over the characteristics embodied in goods, rather than over the goods themselves. Indeed, full-imputation indices can be interpreted as characteristic price indices (Hill and Melser, 2006; De Haan, 2008). Using full-imputation indices also facilitates comparison with the time dummy method, as highlighted by De Haan (2008) and Diewert et al. (2008).[8] In addition, Erickson and Pakes (2011) argue that double-imputation indices, which use observed rather than predicted price relatives for continuing goods, are subject to a form of selection bias, because they treat the hedonic estimation error for continuing, entering, and exiting goods in an asymmetric manner. Benkard and Bajari (2005), Diewert et al. (2008), and Bajari et al. (2021) have also used full-imputation indices. As we show below, full-imputation hedonic indices are less subject to chain drift than traditional 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 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

[8] De Haan (2008) argues that, in the absence of unobserved characteristics, these indices are "strikingly similar." Diewert et al. (2008) note the similarities and also derive the conditions under which they are identical. They note the full imputation approach is more flexible and in practice yields different results than the time dummy method. Neither of these papers highlights the importance of unobserved characteristics, as do Erickson and Pakes (2011). Incorporating the EP-TV approach developed by Erickson and Pakes (2011) to address unobserved characteristics in the full-imputation indices produces additional advantages over the time dummy method. For these reasons, we favor the full-imputation EP-TV approach of Erickson and Pakes (2011) in our hedonic indices.

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. We follow the recent literature (e.g., Byrne et al., 2019) using adjacent-period, weighted least squares estimation with Tornqvist market-share weights. Specifically, we pool observations from the adjacent periods t-1 and t and estimate hedonic regressions of the form

$$\ln p_{k\tau} = \alpha_{t\text{-}1,t} + \delta_t + Z_k' \gamma_{t\text{-}1,t} + \epsilon_{k\tau}, \quad \tau = \{t\text{-}1, t\}, \tag{5}$$

where $\alpha_{t\text{-}1,t}$ is the constant, Z_k is the vector of characteristics for good k, $\gamma_{t\text{-}1,t}$ is a vector of estimated hedonic coefficients held fixed across periods t-1 and t, and δ_t is a fixed effect for period t.[9,10] 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.2 provides additional discussion.

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

[9] Letting T denote the total number of periods in the data, we estimate $T-1$ separate pooled two-period regressions.

[10] We specify the hedonic regression equation (5) 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.

3.3 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) characterize the unit expenditure function for a representative consumer with CES preferences can be characterized as

$$P_t = \left[\sum_{k \in \Omega_t} \left(\frac{p_{kt}}{\varphi_{kt}} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}, \tag{6}$$

where $\sigma > 1$ is the consumer's elasticity of substitution between products, φ_{kt} is an appeal parameter for product k, and Ω_t is the set of products sold in period t. Both the set of products sold Ω_t and product-level appeal φ_{kt} may vary over time. Equation (6) 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 (6), 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. We denote product k's expenditure share among all goods sold in period t as s_{kt} and its expenditure share among common goods $\mathbb{C}_t$ as $s^{\star}_{kt}$. Letting c_{kt} be the quantity of good k purchased in period t, s_{kt} and $s^{\star}_{kt}$ are defined as

$$s_{kt} \equiv \frac{p_{kt} c_{kt}}{\sum_{l \in \Omega_t} p_{lt} c_{lt}} \quad \text{and} \quad s^{\star}_{kt} \equiv \frac{p_{kt} c_{kt}}{\sum_{l \in \mathbb{C}_t} p_{lt} c_{lt}}. \tag{7}$$

The log Sato-Vartia index is then defined as

$$\ln \Phi^{SV}_{t\text{-}1,t} \equiv \sum_{k \in \mathbb{C}_t} \omega_{kt} \ln \left(\frac{p_{kt}}{p_{kt\text{-}1}} \right), \quad \omega_{kt} \equiv \frac{s^{*}_{kt} - s^{*}_{kt\text{-}1}}{\ln(s^{*}_{kt}) - \ln(s^{*}_{kt\text{-}1})} \Bigg/ \left(\sum_{k \in \mathbb{C}_t} \frac{s^{*}_{kt} - s^{*}_{kt\text{-}1}}{\ln(s^{*}_{kt}) - \ln(s^{*}_{kt\text{-}1})} \right). \tag{8}$$

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

$$\lambda_{t,t-1} \equiv \frac{\sum_{k\in\mathbb{C}_t} p_{kt}c_{kt}}{\sum_{k\in\Omega_t} p_{kt}c_{kt}}, \qquad \lambda_{t-1,t} \equiv \frac{\sum_{k\in\mathbb{C}_t} p_{kt-1}c_{kt-1}}{\sum_{k\in\Omega_{t-1}} p_{kt-1}c_{kt-1}}. \tag{9}$$

The log Feenstra index is then defined as

$$\ln \Phi^{Feenstra}_{t\text{-}1,t} \equiv \frac{1}{\sigma-1}\ln\left(\frac{\lambda_{t,t-1}}{\lambda_{t\text{-}1,t}}\right) + \ln \Phi^{SV}_{t\text{-}1,t}. \tag{10}$$

Letting $ER_{t\text{-}1,t}$ and $XR_{t\text{-}1,t}$ represent the sales-weighted product entry and exit rates as, 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 the absence of product turnover.[11]

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

Redding and Weinstein (2020) derive an empirically implementable exact price index in

[11] We use the actual Feenstra adjustment and not the approximation in our implementation.

this setting (the CUPI) as

$$\Phi_{t-1,t}^{CUPI} = \left(\frac{\lambda_{t,t-1}}{\lambda_{t-1,t}}\right)^{\frac{1}{\sigma-1}} \frac{\tilde{P}_t^*}{\tilde{P}_{t-1}^*} \left(\frac{\tilde{S}_t^*}{\tilde{S}_{t-1}^*}\right)^{\frac{1}{\sigma-1}}. \tag{11}$$

where we denote the geometric mean of a variable x_{kt} across goods k in period t as $\tilde{X}_t$ and denote the geometric mean over the set of common goods with an asterisk. $\tilde{P}_t^*$ therefore represents the geometric mean of prices, i.e., $\tilde{P}_t^* \equiv (\Pi_{k\in\mathbb{C}_t} p_{kt}^*)^{\frac{1}{N_{\mathbb{C}_t}}}$, where where $N_{\mathbb{C}_t}$ denotes the number of common goods (i.e., products sold both in period t and in period t-1). Likewise, $\tilde{S}_t^*$ represents the geometric mean expenditure shares on common goods in period t, i.e., $\tilde{S}_t^* \equiv (\Pi_{k\in\mathbb{C}_t} s_{kt}^*)^{\frac{1}{N_{\mathbb{C}_t}}}$. We call the CUPI's second term the "P^* ratio" and its third term the "S^* ratio."

The log version of the CUPI is given by

$$\ln \Phi_{t-1,t}^{CUPI} = \frac{1}{\sigma-1} \ln\left(\frac{\lambda_{t,t-1}}{\lambda_{t-1,t}}\right) + \frac{1}{N_{\mathbb{C}_t}} \sum_{k\in\mathbb{C}_t} \ln\left(\frac{p_{kt}^*}{p_{kt-1}^*}\right) + \frac{1}{\sigma-1}\frac{1}{N_{\mathbb{C}_t}} \sum_{k\in\mathbb{C}_t} \ln\left(\frac{s_{kt}^*}{s_{kt-1}^*}\right). \tag{12}$$

Equation (12) clarifies that two of the CUPI's three terms are unweighted geometric means. This property is important for the CUPI's empirical implementation. Equation (12) also shows that the P^* ratio is simply the traditional Jevons index.

Each of these price indices exactly recovers 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.[12] The Feenstra-adjusted Sato-Vartia index is exact in the presence of product turnover but the absence of time-varying product appeal. The CUPI is exact under the more general conditions of product turnover and time variation in product appeal. We find that these generalizations of the Sato-Vartia index are empirically relevant.

The CUPI, the Sato-Vartia, and the Feenstra indices are all exact CES indices. Although

[12] Feenstra and Reinsdorf (2007) show the Sato-Vartia index is unbiased in expectation with randomness in tastes under restricted conditions. Appendix C.2 contains a further related discussion of this topic.

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 (12) 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.[13]

Redding and Weinstein (2020) adjust their empirical implementation of the CUPI by applying what we call a "common goods rule," which defines the set of goods over which the P^* and S^* ratio terms are calculated (i.e., the goods included in the set $\mathbb{C}_t$). The goods excluded from the set of common goods are reallocated to the product turnover component (Feenstra adjustment factor), which is expenditure weighted. A common goods rule of this sort can be motivated by the argument that it takes time for goods to enter and exit the market. Consistent with this argument, Redding and Weinstein (2020) restrict the set of common goods in their empirical CUPI to those that are sold for a sufficiently long duration both prior to period t-1 and subsequent to period t.[14] A limitation of this duration-based common goods rule is that it requires forward-looking information to implement, and thus it is not feasible to implement in real time. We find that we can mimic Redding and Weinstein's results using a purely backward-looking rule that can be implemented in real time. As we will see, the empirical effect of the S^* ratio varies significantly depending on the implementation of the common goods rule. We discuss potential reasons for the common goods rule's importance in Section 4.4 and Appendix C below.

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

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

4 Results

In this section, we present and discuss the traditional, hedonic, and demand-based exact price indices we have calculated in the item-level data. We focus first on our results from the NPD data, because the richness of the data permits more exploration of alternative methods.

4.1 NPD Results

4.1.1 Hedonics

We consider a wide variety of hedonic specifications as discussed in Section 3.2. We find that predicting log price changes directly produces significantly better model fit than estimating log price levels in periods t-1 and t separately and then forming a predicted log price change. Using the EP-TV approach, which accounts for the time-varying influence of unobservable characteristics by including the lagged residual from a log-level regression, further increases the model fit across all product groups. The results therefore support the argument in Erickson and Pakes (2011) that estimating price changes helps to account for unobservable product characteristics, and that including the first-stage residual from predicting price levels in the estimation of log price changes provides a further advantage.

Table D.4 presents on the goodness of fit for our primary specifications. The column under the sub-header "Log Price Level" shows the average quarterly R^2 for products' predicted log price levels. The average R^2 values range from 0.62 to 0.72 across product groups. The columns labeled "EP-F" display results from predicting log price changes directly, without using the lagged residual, so these columns correspond to a fixed unobservables specification. 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. The average R^2 values using the EP-F approach range from 0.09 to 0.47. Finally, the columns labeled "EP-TV" show results using the time-varying unobservables approach, which includes the lagged first-stage hedonic residual to predict log price changes. The average R^2 values using this approach range from 0.13 to 0.50 across product groups, an improvement on the results using the EP-F approach.[15]

Figure 2 compares several hedonic price indices for the five NPD product groups to the traditional Tornqvist index. We focus primarily on hedonic Tornqvist indices using the EP-F and EP-TV approaches. For comparison we also present results from the time dummy hedonic approach. 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 traditional Tornqvist, the time dummy based index, or the first-difference based index. The gap between the traditional Tornqvist and the EP-TV approach indices varies considerably across product groups, with the largest average differences for memory cards (-2.9 percentage points annually) and headphones (-2.5), and smaller differences for coffee makers (-0.70), boys' jeans (-1.30), and occupational footwear (-0.42).

The time dummy method suggests a notable hedonic adjustment for coffee makers relative to the traditional Tornqvist index, but for other products the difference is modest or is positive rather than negative. Our finding of limited quality adjustment 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 traditional Tornqvist.[16] The previous literature

[15] The discussion above focuses on R^2 values using quantity-weighted estimation. Table D.4 also shows results from expenditure-weighted estimation in the columns labeled "EW" for the weights. Quantity-weighted estimation yields higher R^2 for log price changes on average than expenditure-weighted estimation.

[16] For headphones, the traditional 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

has highlighted other limitations of the time dummy approach; for instance, Pakes (2003) raises questions about the bound implied by the time dummy method, while Diewert et al. (2008) point out that the time dummy method requires more restrictive assumptions than the other hedonic approaches.

Figure 2 illustrates the contrasting results from using the fixed and time-varying unobservable estimation strategies. The hedonic Tornqvist indices using the fixed unobservables specification show inconsistent patterns of hedonic adjustment across product groups. In contrast, the hedonic Tornqvist using the EP-TV method shows consistently greater deflation than the traditional index across product groups. This difference suggests that it is important to permit time-varying valuations of unobservable characteristics.

Figure 3 provides further evidence on the efficacy of the EP-TV approach by displaying results with key observable characteristics left out of the hedonic estimation. Specifically, for memory cards the memory size is omitted, and for the other product groups, the large brand dummy variables are omitted. Omitting these informative characteristics from the estimation equation has a minimal effect on the resulting price indices. Appendix Figure D.2 presents additional analyses showing that omitting those characteristics has a much larger effect on the hedonic indices using a log-level estimation approach.

Our results are consistent with the findings in Erickson and Pakes (2011). They present examples (e.g., 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.

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.

4.1.2 CES Demand-Based Price Indices

We turn now to CES demand-based price indices. The Feenstra index and the CUPI require estimates of the elasticities of substitution in their empirical implementation. Our baseline approach is to estimate a single elasticity for each of the NPD product groups. We employ the method used by Feenstra (1994) and Redding and Weinstein (2020) for this purpose.[17] Table 2 reports the estimated elasticities, which range from about 5.2 to 7.8, consistent with the literature. The table also reports estimates from nested specifications, which we discuss below.

Figure 4 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 are scaled by $\frac{1}{\hat{\sigma}-1}$ so that the CUPI is the sum of the three components; see equation (12). We find that the CUPI shows low inflation relative to the Feenstra index and quite low inflation in absolute terms. In all goods but occupational footwear, the CUPI produces an estimate of 30%–40% annual declines in the price level annually, and it often falls 10–30 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.

[17] 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.

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 implement a related but distinct methodology that can be implemented in real time using only current and backward looking information available in quarter t. For our NPD analysis, we specify a market share threshold for goods present in periods t and t-1 to be considered as common goods for the Jevons and the S^* ratio terms of the CUPI.[18] Goods below this threshold are excluded from the set of common goods, but they still enter the CUPI through their inclusion in the Feenstra adjustment term (lambda ratio). We consider alternative market share percentile thresholds.[19]

Figure 5 illustrates the CUPI's sensitivity to the CGR for different market share thresholds. We consider market share thresholds for continuing goods in t and t-1 of the 10th, 30th, and 50th percentiles. We depict the CUPI for these different CGRs alongside the Feenstra index and the CUPI without a CGR. Implementing the restriction on the set of common goods by market share raises the CUPI by cutting off the low end of the market share distribution from unweighted relative comparisons and shifting it to the weighted entry/exit adjustment term. In that sense, applying a stricter CGR moves the CUPI closer to the Feenstra-adjusted Sato-Vartia index, which combines a traditional matched model index with an adjustment for entry and exit.

The resulting price indices generally shift up as successively stricter definitions of common goods are imposed. For some product groups, such as memory cards, the CUPI using the CGR at the 30th or 50th percentile yields inflation measurements similar to the Feenstra index. For products groups such as headphones and boys' jeans, however, the CUPI shows

[18]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.

[19]In our analysis of the Nielsen 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.

noticeably lower inflation than the Feenstra index even using with a 50th-percentile CGR threshold (i.e., excluding half of products from the set of common goods).

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

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

Martin (2020) notes that the S^* ratio can reflect not only shifting preferences, but also any model misspecification, including a nested preference structure. The CUPI's assumed CES preference structure imposes an equal elasticity of substitution within product groups, and violations of this assumption could lead to biased measures of inflation. Furthermore, the CUPI is more vulnerable to this issue than the other CES price indices we consider.[20]

[20]More precisely, Martin (2020) shows that the CUPI is not consistent in aggregation. Vartia (1976) defines consistency in aggregation as the equality of a single-stage or two-stage index number. In the single-stage of an index number, all goods are included in a single aggregation. In a two-stage construction, the index is computed for a number of subgroups, and the subgroups are aggregated using the same index number formula. Diewert (1978) shows that the Sato-Vartia index is consistent in aggregation.

We have explored nesting products into subgroups in the NPD data, but we have consistently found that nesting does not overturn the CUPI's substantially negative inflation readings relative to the Feenstra index. We describe our analysis in detail in Appendix C.1. To summarize, we have explored two different approaches to nesting, a characteristics-based approach and an approach based on predicted prices using hedonic regressions. Neither nesting method materially changes the inflation measurements of the CUPI, suggesting that aggregation issues are not the primary driver of its estimates of rapid deflation in the product groups we consider.

4.1.3 Comparing Traditional, Hedonic, and Exact Price Indices

Figure 6 presents the main traditional, hedonic, and demand-based price indices that we have considered for all five product groups. Because the CUPI indices are outliers for some groups, Figure 7 displays price indices without the CUPIs. It also adds the Laspeyres index for comparison with the BLS' CPI approach. Figures 8 and D.3 in Appendix D present plots of chained price index levels calculated by chaining the quarterly price indices underlying Figures 6 and 7. They therefore illustrate the cumulative effects of the differences between the various indices. Likewise, Table 3 reports the chained index levels in 2018:4, reflecting the cumulative price changes since 2014:4, when all indices are normalized to one.

The price indices follow a roughly similar pattern of relative orders across these product groups. Figure 7 shows that the Laspeyres index typically shows the least deflation. The traditional 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, in Figure 6, the CUPI (both baseline and nested by product characteristics) shows the greatest deflation, especially for headphones and boys' jeans. The substantial gap between the CUPI and the Feenstra index in headphones and boys' jeans is especially striking given our imposition of a 30th-percentile CGR.

The gap between the traditional Laspeyres and Tornqvist indices for most product groups highlights the advantages of using item-level scanner data, which permits construction of a superlative price index with internally consistent prices and expenditure shares in adjacent periods. The gap is especially striking in the cumulative price indices shown in Figure D.3. The cumulative gaps are on the order of 5–10 percentage points from 2014 to 2018 in coffee makers, occupational footwear, and boys' jeans, and significantly larger in the "high-tech" categories of memory cards and headphones. The gap also varies over time, consistent with the Laspeyres index exhibiting a time-varying substitution bias. Thus, using scanner data can produce substantial improvements in price measurement even without performing quality adjustment.

Figure D.3 also shows that the hedonic Tornqvist using the EP-TV method tends to indicate larger cumulative quality adjustment via product turnover than the Feenstra index. The Feenstra index indicates approximately 2 percentage points of cumulative disinflation beyond the Sato-Vartia index across all five product groups. The hedonic Tornqvist indicates roughly the same adjustment (relative to the traditional Tornqvist) for coffeemakers and occupational footwear, but larger adjustments, of roughly 5–7 percentage points, in memory cards, headphones, and boys' jeans.

4.2 Nielsen Results

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

For the hedonic EP-TV approach, we combine the insights of Erickson and Pakes (2011) with the machine learning approach developed in our companion paper, Cafarella et al. (2023), and summarized in Appendix B.4. The machine learning approach allows us to exploit the Neilsen scanner data's unstructured information on item-level attributes.

The median in-sample R^2 of the hedonic (log-level) price predictions is roughly 85% for the food product groups.[21] The median out-of-sample R^2 is roughly 75% for food product groups. The model's predictive performance is comparable to that of Bajari et al. (2021), who report out-of-sample R^2 values of 80–90% in their best-performing specifications using the rich product text and image information in their data set. For log price changes, our median in-sample R^2 is above 50% for the food product groups. The median out-of-sample R^2 values decline to nearly 20% for the food product groups. We consider our procedure to be very successful in light of the limited attribute information available in the data set.

We again explore alternative CGRs to calculate the CUPI. The Nielsen data provides a longer panel than the NPD data, which allows the exploration of alternative CGRs that depend on the duration of goods' time in the market to date. We implement a modified approach to defining the CGR rule in the Nielsen data as follows. We first compute percentiles of the pooled sales distribution within a narrow product group for pooled sales in periods t-1 and t. Common goods are defined as goods sold in both periods, and which have sales in period t above the Xth percentile of this pooled sales distribution. This alternative approach to defining the CGR allows us to consider longer duration-based alternatives.[22]

Figure 9 shows the results for the aggregated food categories of the CUPI and its components using various CGRs defined by different sales-based percentiles. 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 al-

[21] See Cafarella et al. (2023) for details.

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

ternative CGRs. Indeed, the plots for the different values are nearly indistinguishable. In contrast, the S^* ratio is very sensitive to the CGR in the Nielsen data, which leads directly to sensitivity in the CUPI. The baseline CUPI without a CGR percentile threshold has average four-quarter price inflation about 10 percentage points below the Feenstra. Using a 50th percentile for the CGR yields a price index that is much closer to the Feenstra index.

We consider alternative specifications of the CGR using market thresholds using percentiles of sales pooled over over the current and prior 4 quarters. In addition and critically, a common good is defined in this context if it is present in periods t and $t-4$. Using a duration component in the CGR puts more weight on goods present for the longer horizon, yielding greater comparability with the duration-based CGR used by Redding and Weinstein (2020).[23] Appendix B.3 shows that using this longer horizon approach for computing sales percentiles, a CGR with a 10th-percentile sales threshold yields results comparable to a CGR with a sales threshold between the 25th and 50th percentiles using a two quarter horizon.[24]

Figure 10 presents a full set of price indices for the Nielsen scanner food product groups in change and level forms.[25] The panels of the figure include the BLS CPI computed for the same Nielsen product groups.[26] We find that the CPI and the traditional Laspeyres index track each other closely in Nielsen'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.

The cumulative level implications highlight that the hedonic Tornqvist is about 4 percentage points lower in 2015 than the traditional Tornqvist, and the Feenstra index is about

[23]An advantage of this alternative duration-based CGR for the purposes of producing real-time statistics is that it does not require forward-looking information.

[24]Appendix B.3 also explores the use of the Nielsen Consumer Panel and CGR sales-based percentile rules. Using the Consumer Panel enables us to more readily compare our results to those in Redding and Weinstein (2020). The Consumer Panel requires smaller CGR adjustments than the Retail Scanner data. However, we note that the Retail Scanner data matches economy-wide trends for prices and sales than the Consumer Panel.

[25]For all indices, we aggregate across product groups using a Tornqvist aggregator with Divisia-style product group market share weights.

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

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

Using a 25th-percentile CGR, the CUPI is more than 40 percentage points lower than the Feenstra index in 2015; using a 50th percentile CGR reduces the difference to 20 percentage points. Alternatively, using the longer $t-4$ to t horizon described above, the 10th-percentile CGR yields a difference of about 25 percentage points.[27]

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

4.3 Chain Drift

A potential challenge to using transactions data to compute price indices is chain drift. This issue is particularly problematic with high-frequency indices computed from local transactions data (e.g., De Haan and Van Der Grient, 2011). Our analysis uses national data at a quarterly frequency, which mitigates this issue. Given our focus on comparing alternative approaches for computing price indices, we consider whether GEKS-type indices (Gini, 1931; Eltetö and Köves, 1964; Sculz, 1964) preserve the implications of our core findings.

Primarily, we follow Bajari et al. (2021) by computing a GEKS-type index (which we

[27]Results for the nonfood product groups, described in Appendix B.2, show substantially greater departures between the BLS CPI and the Nielsen Laspeyres, consistent with our concerns about the Nielsen scanner data's representativeness for the nonfood product groups. The CUPI for nonfood is very low. With a 30th-percentile CGR, the CUPI price level (indexed to 2006) is almost 70 percentage points lower in 2015 compared to the Feenstra (the difference shrinks slightly to 40 percentage points with a 50th-percentile CGR). These results may arise partly from the limited coverage of nonfood items in the Nielsen scanner data.

denote "GEKS-lite") that is the geometric mean of the chained year-over-year index for the 4th quarter of each year and the directly computed (unchained) year-over-year price index for the 4th quarter. GEKS-lite offers an easily implementable alternative to a full GEKS procedure, which involves computing price indices over many possible horizons. Given that we are including hedonic indices which require re-estimation of models for alternative time horizons, the computational burden of implementing a full GEKS procedure is substantial. Table 4 reports average annual chained and GEKS-lite indices for the five NPD product groups and alternative indices. The GEKS-lite price change index tends to show less deflation than the chained price index for traditional price indices. For the hedonic indices, the GEKS-lite indices actually show faster deflation in three out of the five product groups. Among the demand-based indices, the GEKS-lite indices typically show greater deflation, but these differences are modest quantitatively. The GEKS-lite CUPI shows substantially less deflation, but this difference also reflects the effects of applying a common good rule over a longer horizon.[28]

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

We also implement the rolling-year GEKS method (Ivancic et al., 2011) as a point of comparison. Given that this procedure is computationally burdensome, we only implement this for traditional price indices. In appendix Table D.2, we show that results for rolling

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

year GEKS are similar to those using the GEKS-lite approach.

Table 5 reports analogous chained and GEKS-lite indices for the aggregated food indices, which we generate following the same procedure as in prior sections.[29] The table reports average annual indices for both specifications. The results for Nielsen's food product groups show that we obtain similar, albeit slightly higher, rates of average inflation using the GEKS-lite compared to the chained indices. This pattern is especially noticeable for the CUPI, a result which again reflects the effects of applying the CGR over a longer horizon. Importantly, the rank ordering and the quantitative differences across alternative indices are preserved using the GEKS-lite based indices. Focusing on the GEKS-lite indices, inflation in Nielsen's Food product groups is higher using the Laspeyres index than the Tornqvist, higher using the Sato-Vartia than the Feenstra, and higher using the Feenstra than the CUPI.[30]

4.4 Taking Stock

The current system of price measurement implemented by the BLS samples a relatively small quantity of goods and then aims to follow them for a number of periods. When goods disappear, the BLS uses various approaches to account for turnover, which often involve expert judgment. The BLS uses hedonic methods on a case-by-case basis for a small fraction of the market basket. Thus, while procedures to address quality changes and product turnover do currently exist, they are designed around a data collection and processing architecture based on hand-collection of price quotations and non-uniform approaches to addressing quality change. These procedures do not readily scale and do not take advantage of data housed in the information systems of businesses.

Incorporating transactions data into official statistics requires making several methodological decisions, such as whether to use traditional, hedonic, or exact price indices, and

[29] That is, we compute the indices at the product group level and then use Divisia weights to aggregate to the food level.

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

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

If the item-level data contain information on product attributes, as they commonly do, hedonic methods can also be applied at scale in real time. We have found that the most robust approach for implementing hedonics at scale is to use the time-varying unobservables approach from Erickson and Pakes (2011). Our results provide ample support for their argument that it is important to correct for the reevaluation of the unmeasured characteristics of continuing, entering, and exiting goods. Our approach also extends their results by demonstrating that their methodology can be implemented at scale using transactions data and superlative price indices with internally consistent prices and expenditure weights.

Demand-based indices offer a useful alternative for comparison to hedonic indices. These indices are exact under certain sets of assumptions, and in the most general case (the CUPI), they can account both for quality change via product turnover and for time-varying product appeal for continuing goods. Redding and Weinstein (2020) argue that neglecting the latter issues can bias cost-of-living price indices.

The limitation of the CES demand-based approaches we have considered is their sensitivity to the strong assumptions of their underlying models, which may omit empirically important market imperfections. A central assumption of these approaches is the existence of a unified national market where all goods are available. In Appendix B.5, we assess the realism of this assumption in the Nielsen data. We show that most products have less than 20 percent penetration across Nielsen metro areas. Items with greater sales volumes are sold across more areas, so the national market assumption is more realistic on a sales-weighted basis. Nonetheless, even on a sales-weighted basis, a distinct minority of items reach a truly national market.

We believe the failure of the national market assumption is likely to have a much larger effect on the CUPI than on the other price indices we have considered. Superlative price indices such as the Tornqvist and Sato-Vartia are approximately consistent in aggregation (Diewert, 1978), so the failure of the national market assumption is less troubling for those indices; a similar argument applies to the hedonic Tornqvist index. The Feenstra index generalizes the Sato-Vartia index with an expenditure-weighted term to correct for product turnover, so it also contains only expenditure-weighted terms. In contrast, the CUPI contains multiple unweighted geometric mean terms, which implies that goods with small expenditure shares can have an outsize effect on the index.

In Appendix C, we examine the behavior of the CUPI empirically, analytically, and in simulations under various assumptions about preferences and market structures. A few key results emerge from our examination. First, as noted above, in Appendix C.1 we examine the CUPI's sensitivity to assumptions about nested preference structures. The various nesting structures we have explored, using characteristics-based and predicted price-based nests, do not meaningfully change the pattern that the CUPI measures significantly lower inflation than the Feenstra index even with the application of a stringent common goods rule.

Next, we present an analytical argument in Appendix C.2 that the presence of time-varying product appeal does not, on its own, produce a bias in the Sato-Vartia and Feenstra indices. We show that, although the force that Redding and Weinstein (2020) argue will tend to impart an upward bias to those indices does exist, if appeal shocks are independently and identically distributed (i.i.d.) over time, there is a symmetrical and offsetting force that will impart a downward bias. We present simulation evidence to support our analysis in Appendix C.3. The simulations show that, although time-varying product appeal does not produce an average bias in the Sato-Vartia index in the presence of i.i.d. product appeal shocks, the Sato-Vartia is noisier than the CUPI in the presence of appeal shocks. The simulation results also show that rising dispersion in product appeal will cause the CUPI to measure lower inflation than the Sato-Vartia, which will be upward-biased. We show in Figure D.5, however, that rising product dispersion appears empirically unable to account for the ubiquitously low inflation rates measured by the CUPI. Finally, we present simulation evidence from cases of market imperfections such as localized markets or product stockouts that may drive a downward bias in the CUPI, and that a CGR can help to ameliorate the bias in those cases. These simulations suggest that the failure of the national market assumption documented in Appendix B.5 may be an important driver of the empirical behavior of the CUPI.

Taking stock of the empirical and theoretical evidence, we believe that the hedonic methods we have explored provide a sensible and data-driven approach to quality adjustment that can be applied at scale in transactions-level data. We believe that the demand-based indices that incorporate quality adjustment—particularly the Feenstra—provide useful benchmarks that should be used for purposes of comparison with the hedonic and traditional indices. On the other hand, current implementations of the CUPI involve strong assumptions about market structure and preferences that can yield anomalous results when taken to the data. Although modifications such as a CGR can ameliorate or eliminate those anomalies, there is limited theoretical or empirical guidance on what an appropriate correction might be. This topic should be a high priority for future research. As this paper demonstrates, item-level transactions data with prices, quantities and attributes permits exploration and comparison of a wide range of potential price indices, which should facilitate such future research.

5 Conclusion

Item-level transactions data with price, quantity, and attribute information enable considerable advances in the production of price indices. These include price indices that are

- granular with respect to product, geography, and frequency,

- available in close to real time,
- adjusted for substitution across goods using superlative index formulas, and
- adjusted for quality change and rapid turnover of goods.

The availability of contemporaneous price and quantity data and of rich item descriptions is fundamental to implementing these innovations. While the increased availability of "big data" therefore has considerable promise for improving price measurement, there are also considerable challenges to the production of quality-adjusted price indices. In particular, because of the rapid turnover of products in item-level data, quality change must be addressed at scale.

We address the challenge of implementing price indices in big data by exploring and evaluating two alternative approaches for quality adjustment at scale: hedonic methods and demand-based methods. We find that it is important for hedonic methods to account for time-varying changes in the valuation of goods' unobservable characteristics. We do so using the methodology of Erickson and Pakes (2011). Our results show that traditional matched-model methods substantially overstate the rate of inflation. We find that these patterns are pervasive, that is, not limited to goods such as electronics where technological progress is most visible. Hence, the practice of statistical agencies of focusing hedonic adjustment in sectors where technological change is most visible can obscure pervasive quality change.

Among the demand-based methods we have considered, we find that the Feenstra (1994) index produces a relatively stable magnitude of quality adjustment across product groups that is also broadly consistent with the implications of our preferred hedonic methods. The CUPI developed in the path-breaking work of Redding and Weinstein (2020) implies very rapid deflation without the imposition of a common goods rule. Our analysis suggests the need for further research on how to choose such a rule while preserving the CUPI's theoretical appeal and suggests some potential paths forward on that topic.

Finally, and importantly, item-level transaction data allow superlative indices to be constructed in real time using internally consistent prices and quantities. We find that using superlative indices to account for substitution effects among continuing goods with such data has a large effect on estimated inflation.

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

References

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

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

C. L. Benkard and P. Bajari. Hedonic price indexes with unobserved product characteristics, and application to personal computers. *Journal of Business and Economic Statistics*, 23 (1):61–75, jan 2005. ISSN 07350015. doi: 10.1198/073500104000000262.

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

Bureau of Labor Statistics. Quality adjustment in the cpi, 2023. URL https://www.bls.gov/cpi/quality-adjustment/home.htm. Accessed February 9, 2023.

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

M. J. Cafarella, G. Ehrlich, T. Gao, M. D. Shapiro, and L. Y. Zhao. Using machine learning to construct hedonic price indices. Unpublished working paper, Census, Maryland and Michigan, 2023.

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

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

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

E. Diewert. Chapter 6: The Chain Drift Problem and Multilateral Indices. 2021.

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

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

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

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

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

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

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

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

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

R. J. Hill and D. Melser. The Hedonic Imputation Method and the Price Index Problem. Technical report, 2006.

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

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

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

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

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

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

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

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

H. Schultz. A Misunderstanding in Index-Number Theory: The True Konus Condition on Cost-of-Living Index Numbers and Its Limitations. *Econometrica*, 1939. doi: 10.2307/1906996.

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

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

Table 1: Rates of Product Turnover: NPD Data

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

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

Table 2: 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 3: Alternative Price Indices, Levels in 2018q4 Relative to 2014q4: NPD Data

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

Notes: Values are cumulative changes in 2018:4 relative to the 2014 price level, with 2014 price level set to 1. CUPI-N is nested CUPI using characteristics approach. Data come from the NPD Group.

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

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

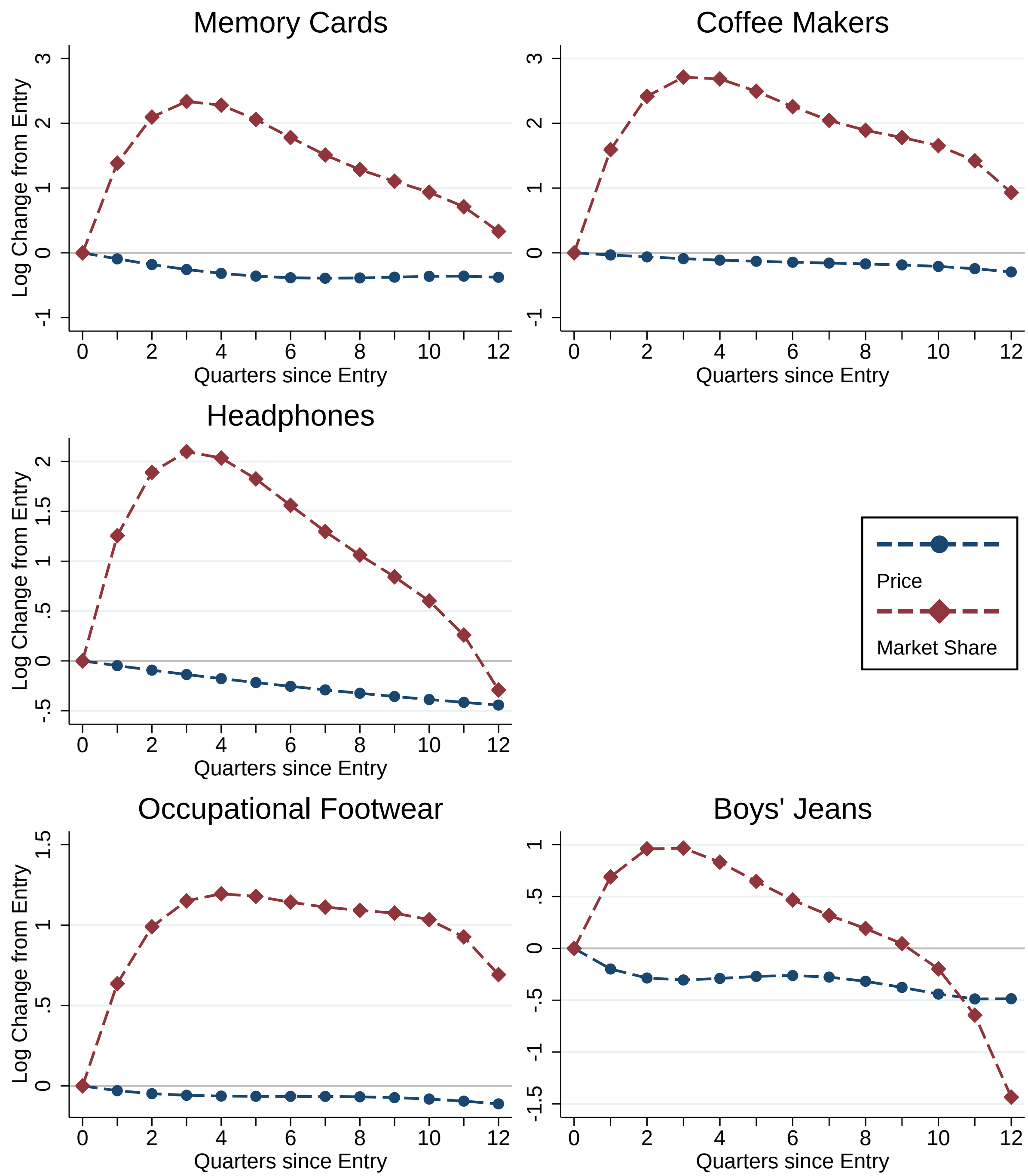
Notes: Chained values are averages of cumulative quarterly rates for year. GEKS-lite is the average of the geometric mean of the chained values and the YoY price indices for q4 for each year. Data come from the NPD Group.

Table 5: Alternative Price Change Indices, Chained vs GEKS-Lite: Nielsen Food

Index	Chained	GEKS-Lite
Laspeyres	.014	.014
Tornqvist	.005	.009
Sato-Vartia	.007	.010
Feenstra	.003	.005
CUPI	-.034	-.020

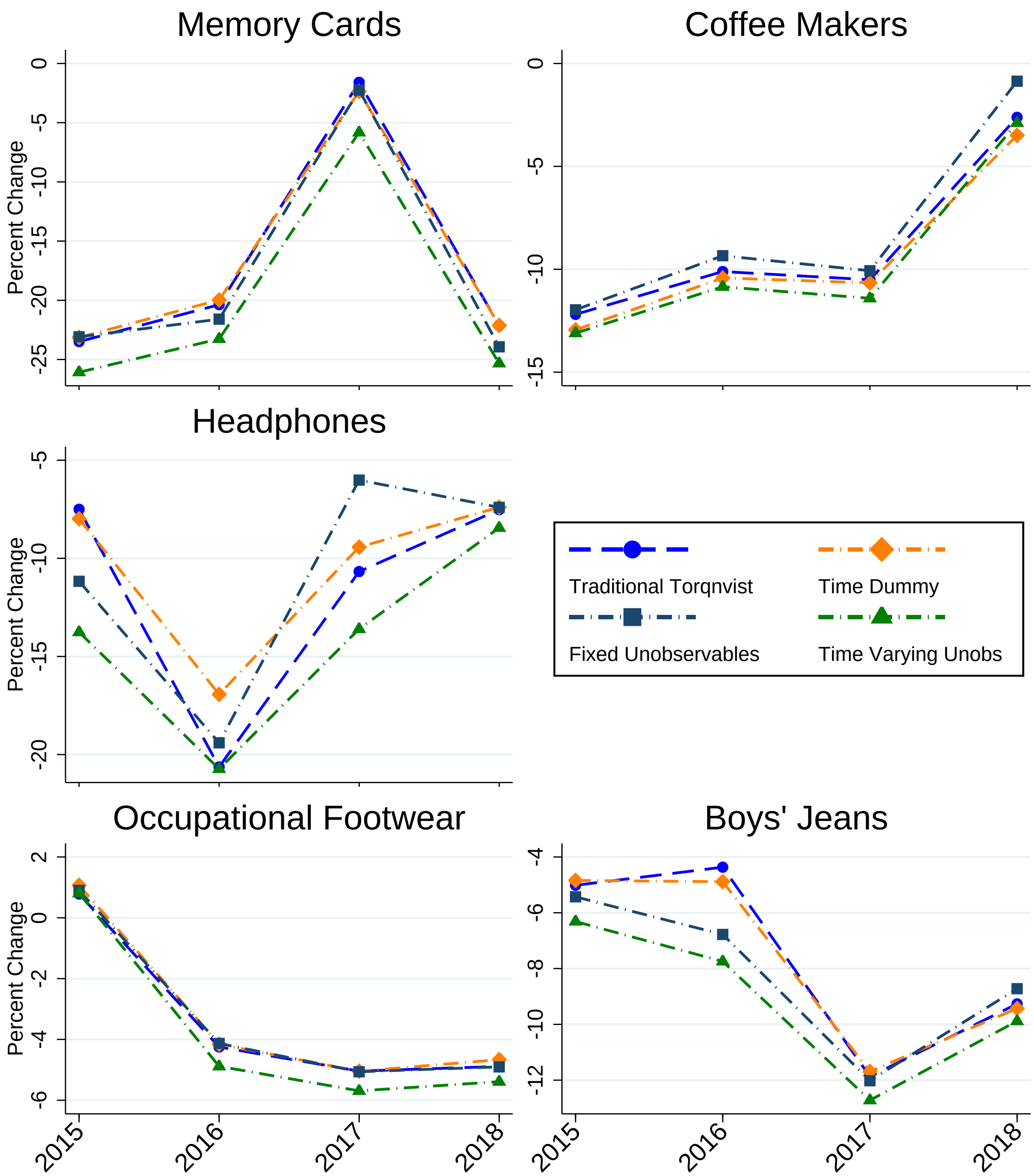
Notes: Chained values are averages of cumulative quarterly rates for year. GEKS-lite is the average of the geometric mean of the chained values and the YoY price indices for q4 for each year. Laspeyres is the geometric Laspeyres. CUPI uses 25th percentile CGR. Data come from Nielsen.

Figure 1: Product Lifecycle Dynamics: NPD Data



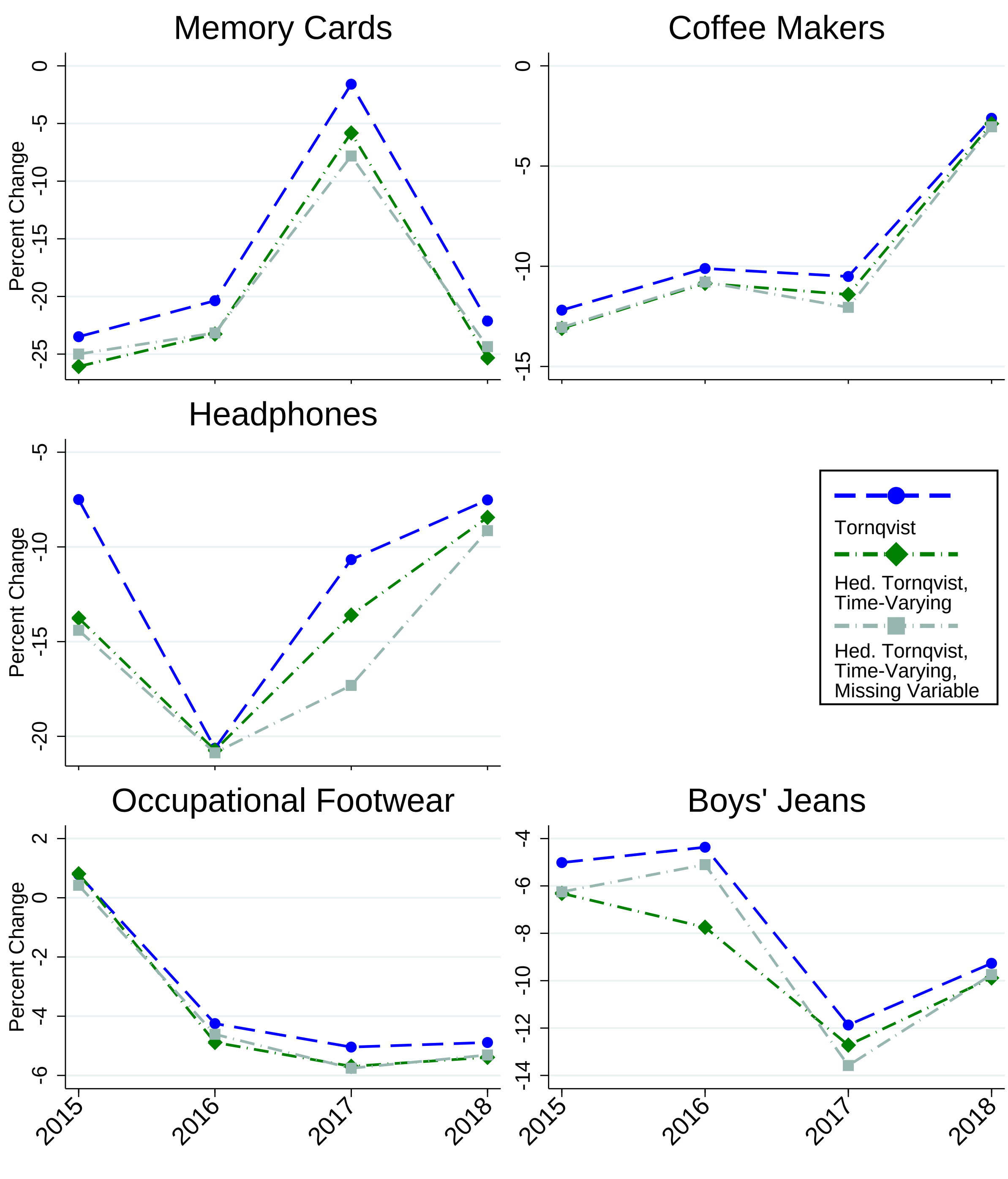
Notes: Unweighted average market share and prices relative to their value in the period of their initial entry. Entry occurs in period 0. All series are smoothed with a quartic spline. Data comes from the NPD Group.

Figure 2: Hedonic Specifications, Fixed vs. Time-Varying Unobservables: NPD Data



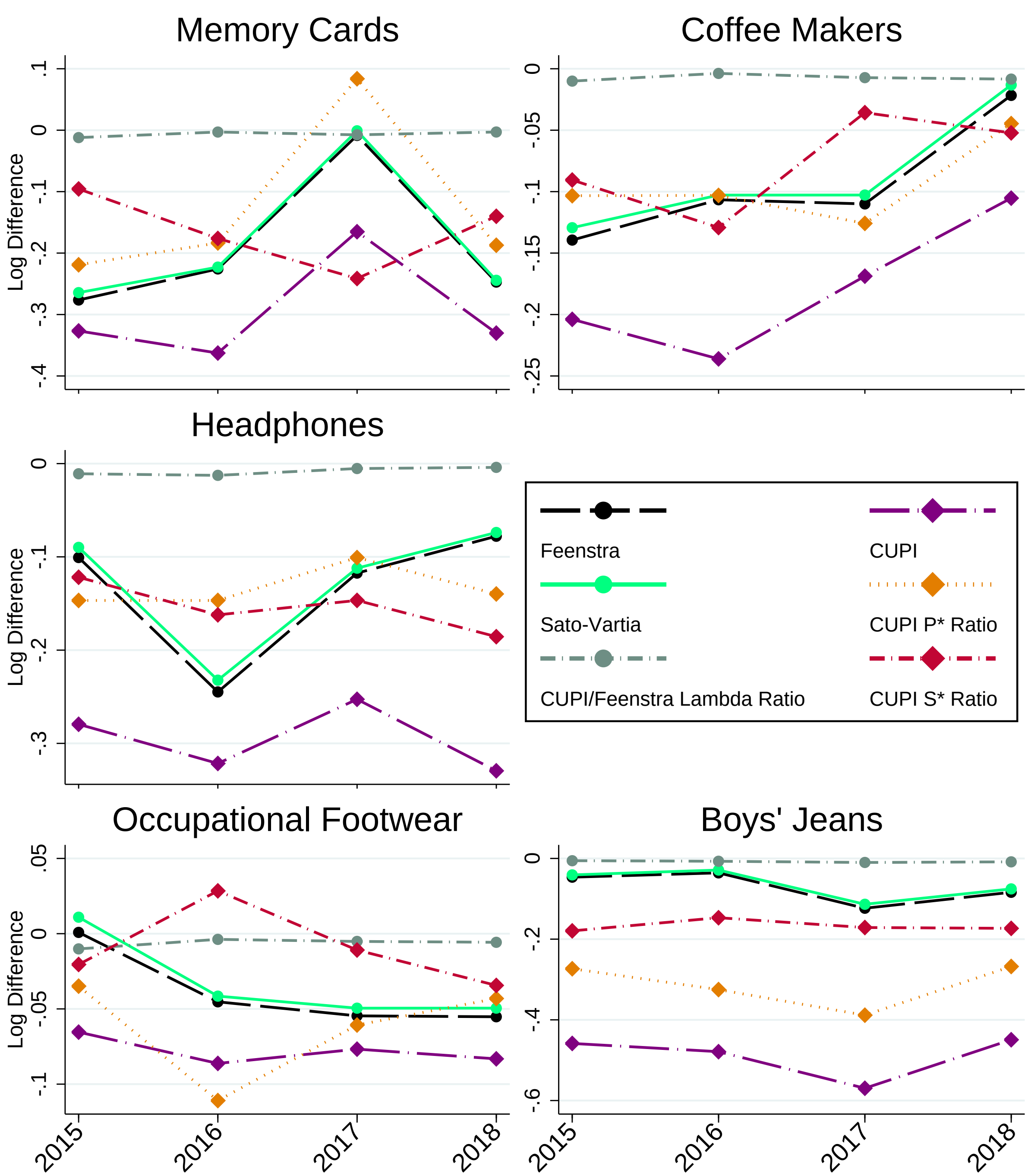
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. The time-dummy Tornqvist index uses adjacent period estimation with Tornqvist market share weights. The fixed unobservables model estimates hedonic models of log change in price using WLS and average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

Figure 3: Hedonic Specifications, Evaluating Time-Varying Unobservable Specification: NPD Data



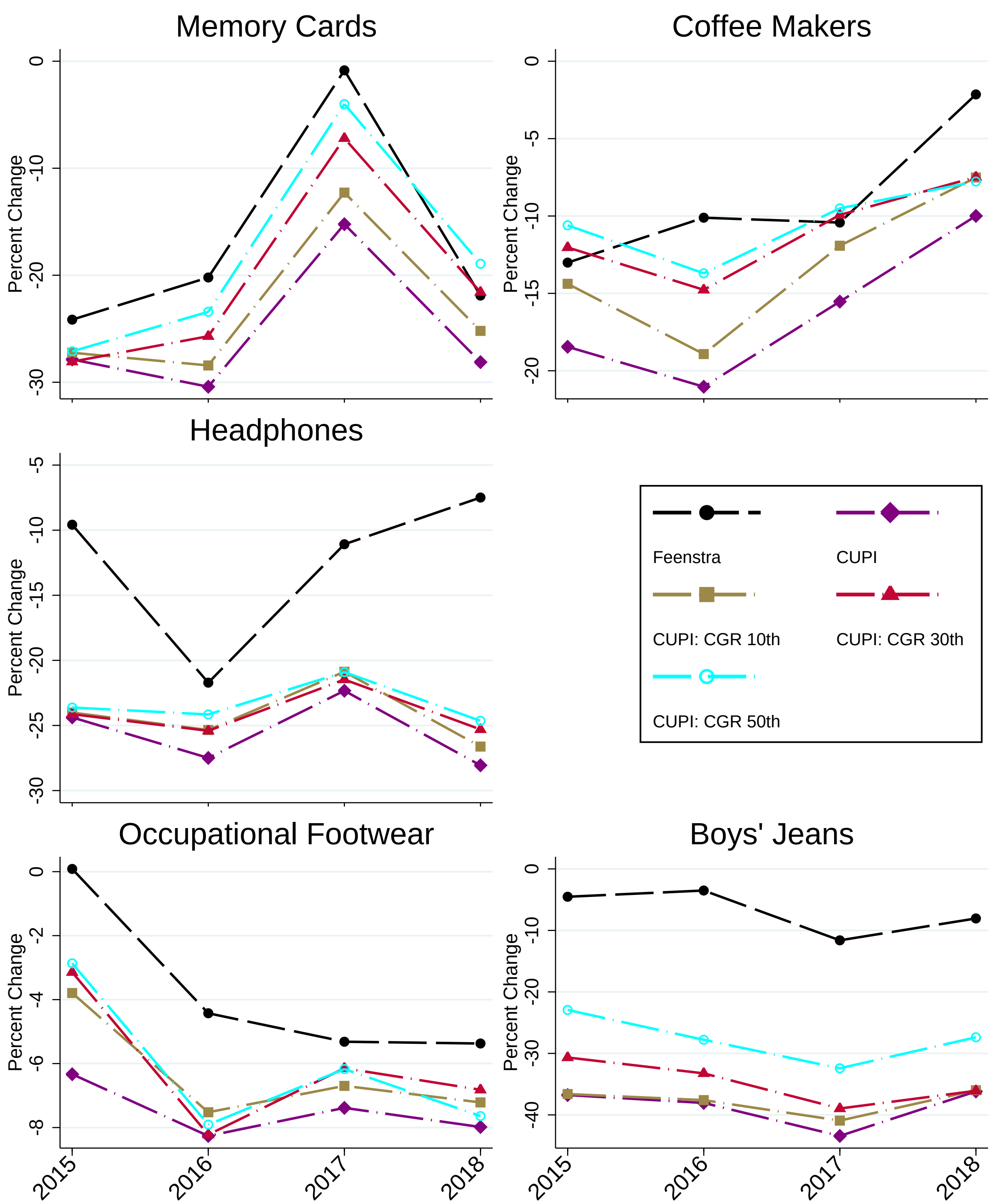
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. The time-varying unobservable model estimates hedonic models of log change in price using WLS and average quantity-share weights, including lagged hedonic level residuals. The "Missing Variable" series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

Figure 4: Components of Feenstra and UPI: NPD Data



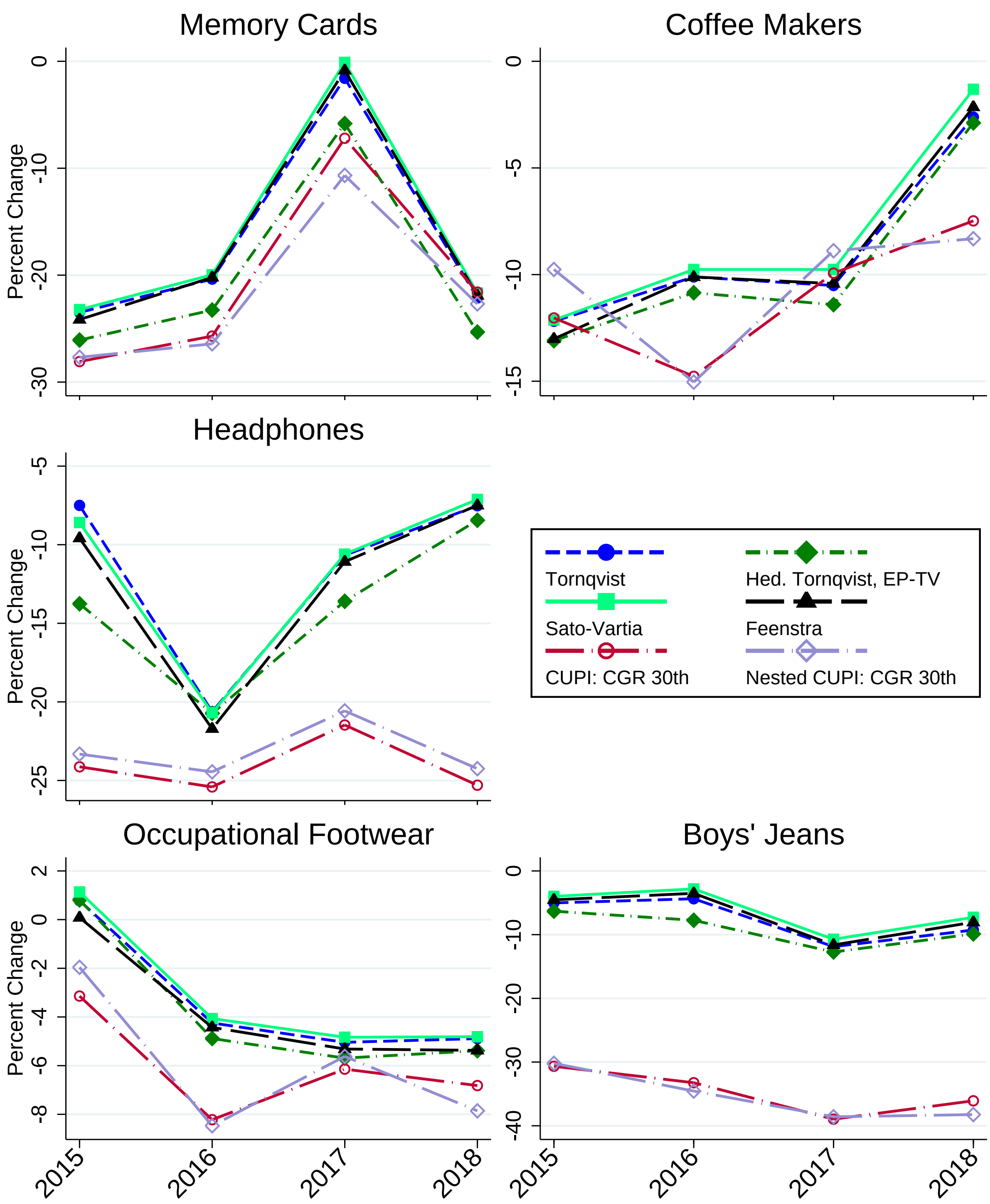
Notes:Values are log differences on a q4-to-q4 basis, aggregated from chained quarterly price indices. Units are reported in log-differences to allow for an additive decomposition of price indices. The Feenstra index is the sum of the Sato-Vartia and CUPI/Feenstra Lambda Ratio. The CUPI is the sum of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

Figure 5: CUPI, Alternative Common Goods Rules (CGRs): NPD Data



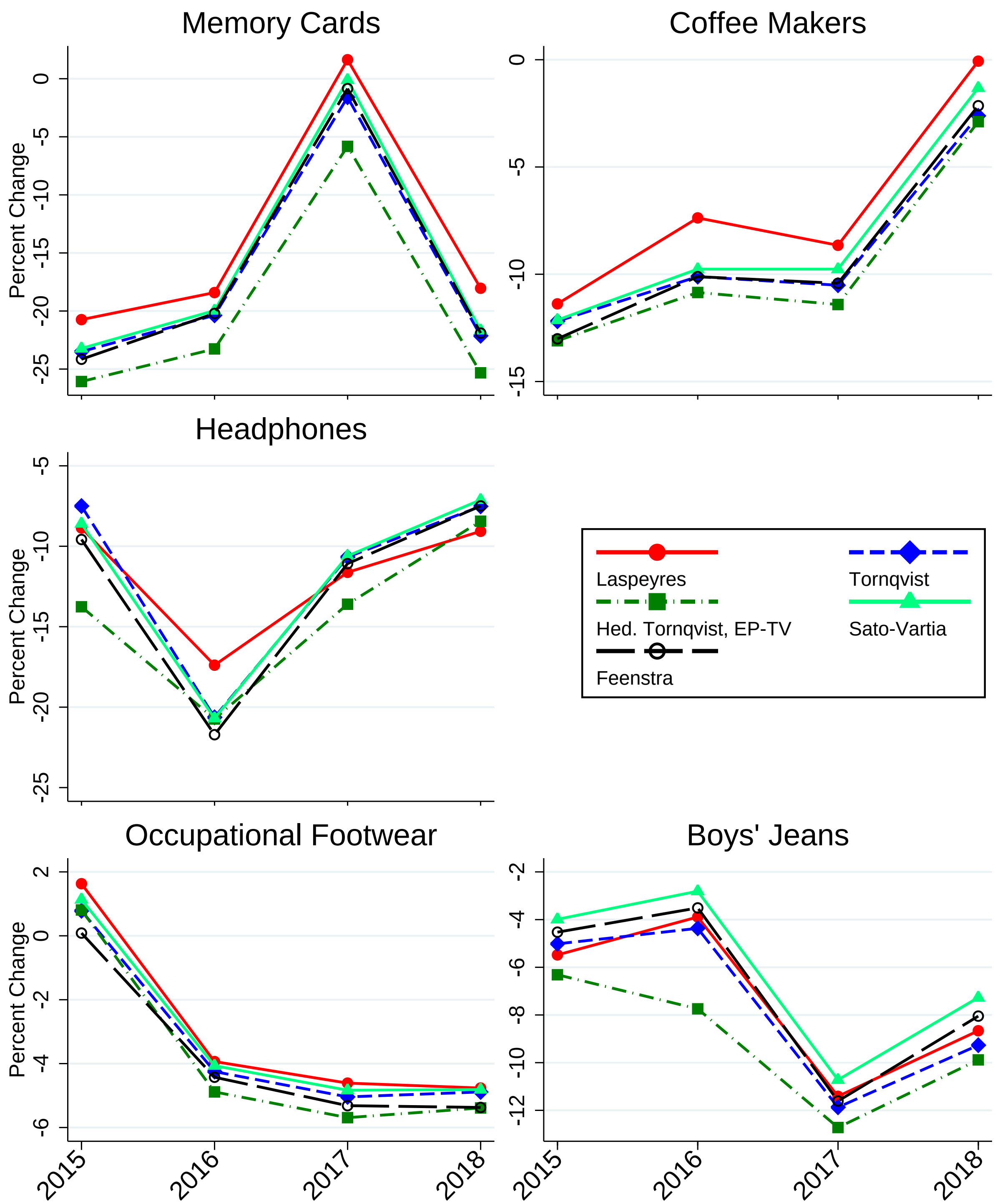
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. Common goods market share rules for the CUPI exclude from the group of common goods those products with market shares below the noted percentile in both periods. The Feenstra-adjusted Sato-Vartia index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

Figure 6: Comparison of Main Price Index Specifications: NPD Data



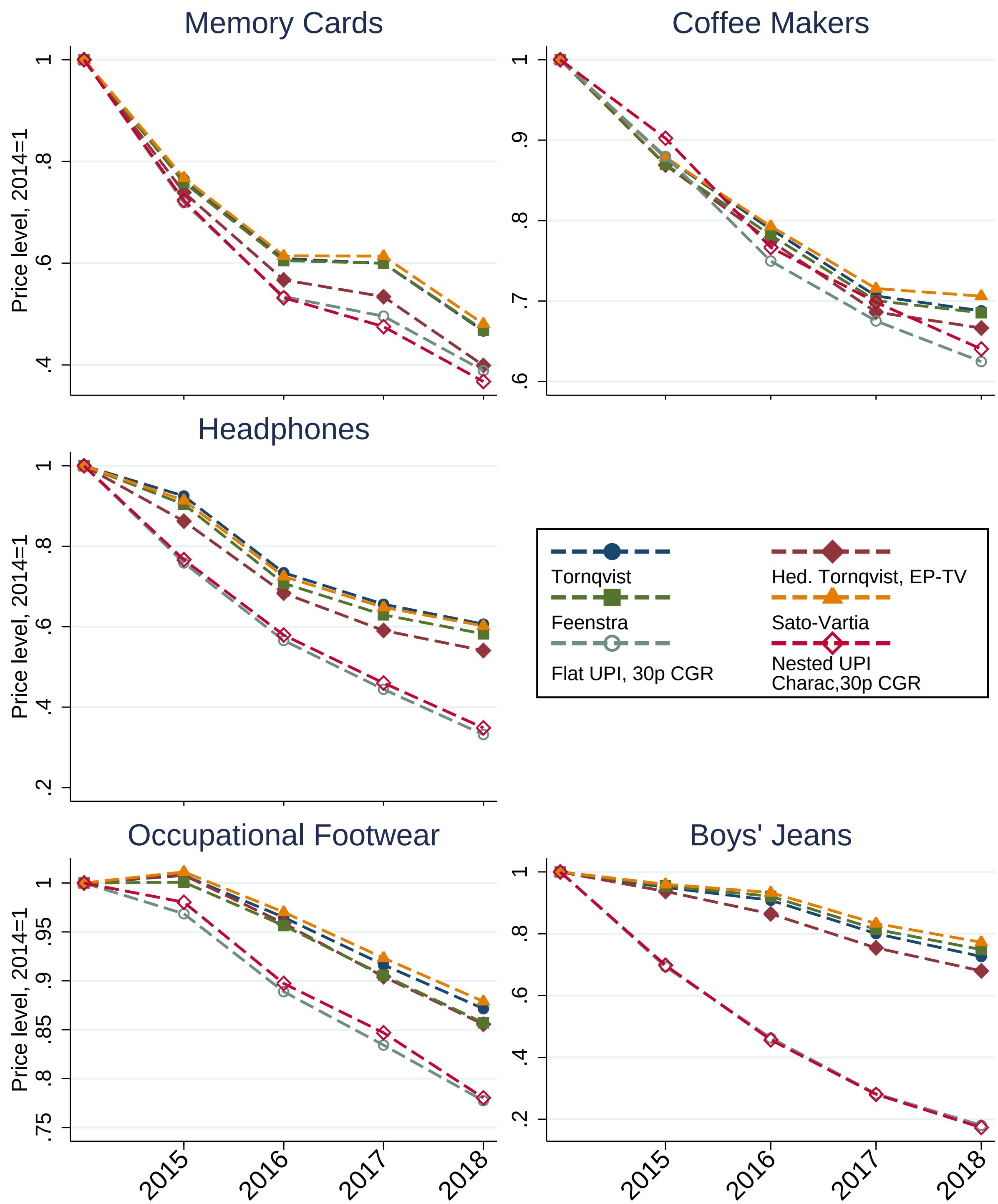
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. “Hed. Tornqvist, EP-TV” is the hedonic time-varying unobservables model estimated over log price differences using WLS and with weights that are average quantity-shares in adjacent periods. Data comes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

Figure 7: Main Price Index Specifications, Without CUPI: NPD Data



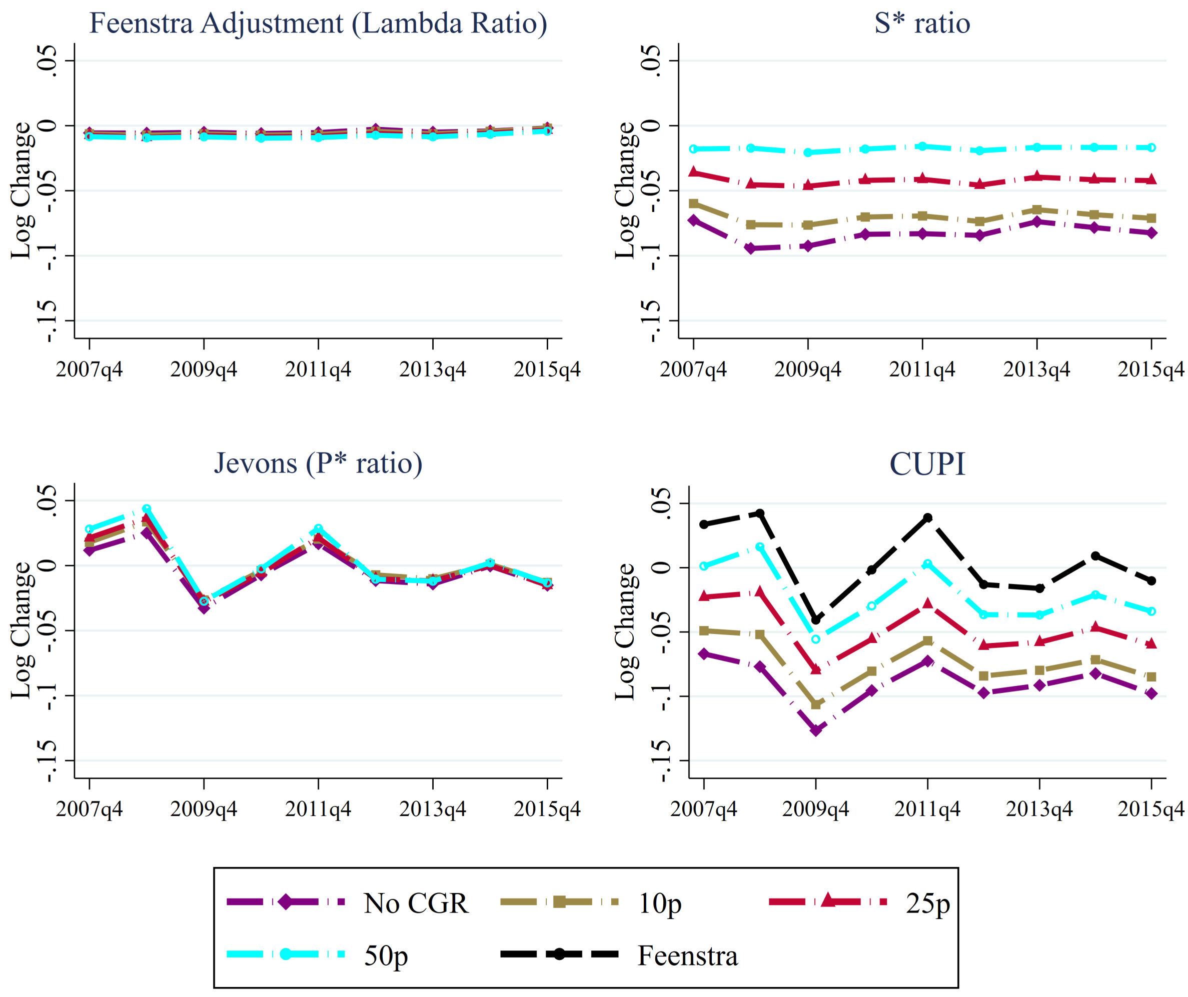
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. The Laspeyres series reports a geometric mean Laspeyres index. "Hed. Tornqvist, EP-TV" is the hedonic time-varying unobservables model estimated over log price differences using WLS and with weights that are average quantity-shares in adjacent periods. Data comes from the NPD Group.

Figure 8: Main Price Index Specifications, Cumulative Price Level Changes: NPD Data



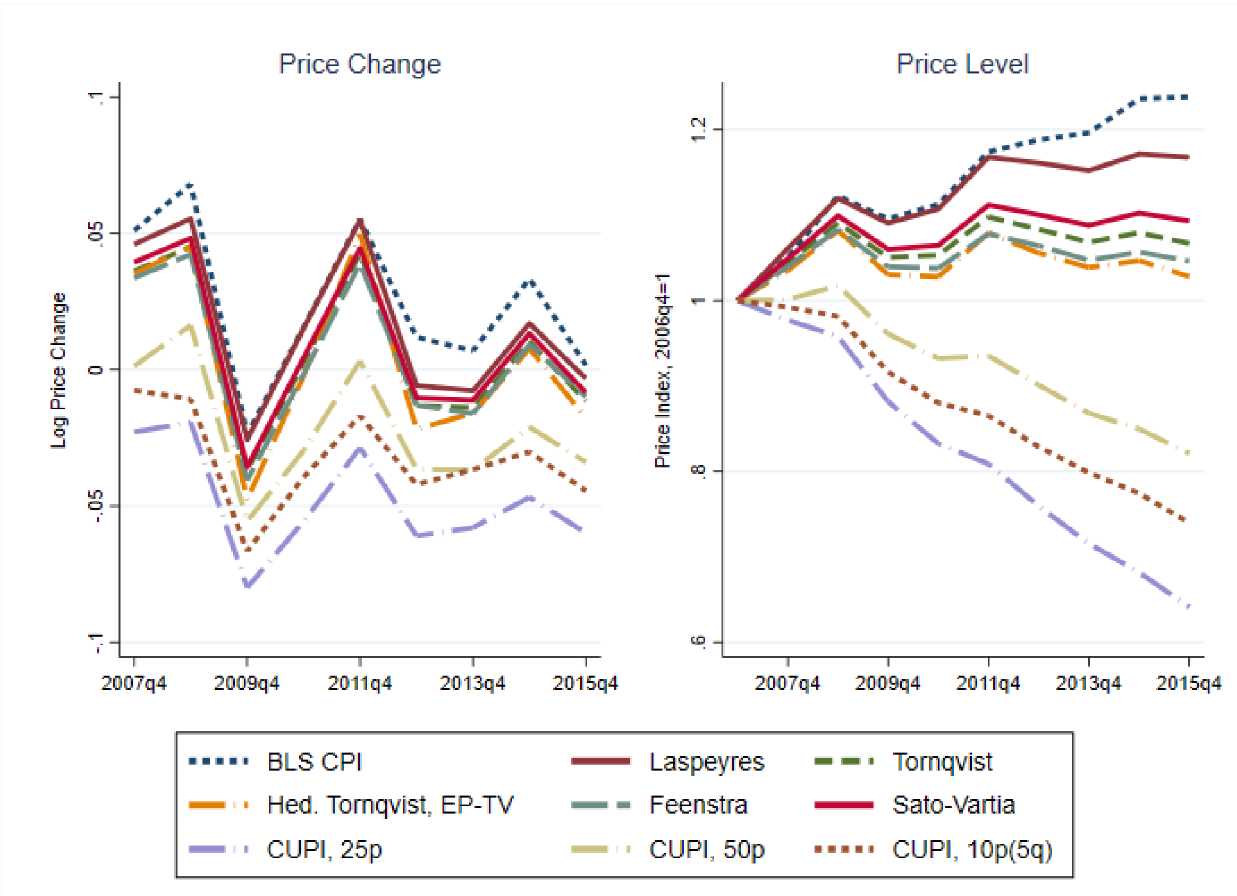
Notes: Values are cumulative changes relative to the 2014 price level, with 2014 price level set to 1. "Hed. Tornqvist, EP-TV" is the hedonic time-varying unobservables model estimated over log price differences using WLS and with weights that are average quantity-shares in adjacent periods. Data comes from the NPD Group.

Figure 9: CUPI and Its Components with Alternative CGRs: Nielsen Food



Notes: The figure shows Nielsen 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 10: Main Price Index Specifications, Price Changes and Levels: Nielsen Food



Notes: The figure shows Nielsen Retail Scanner data for food product groups. Price changes show annual log differences 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 price levels chained quarterly values of each price index in the fourth quarter of each year, with the price level in the fourth quarter of 2006 normalized to one for each index. The Laspeyres index is geometric. "CUPI, 25p" and "CUPI, 50p" use 25th-percentile and 50th-percentile quantity shares, respectively. "CUPI, 10p (5q)" uses a 10th-percentile quantity share threshold computed over quarters t and $t-4$.

Appendix

A Additional Price Index Details

A.1 Traditional Price Indices

Generalizing equation (1), we can define every traditional geometric price index as a weighted average of log price changes. Specifically the log geometric price index, $\ln \Psi_t^G$, is given by

$$\ln \Phi_t^G \equiv \sum_{k \in \mathbb{C}_{t\text{-}1,t}} w_{kt} \ln \frac{p_{kt}}{p_{kt\text{-}1}},$$

where w_{kt} is a weight assigned to product k. The choice of weights determines the index. The Laspeyres index uses lagged expenditure shares as weights ($w_{kt} = s^*_{kt\text{-}1}$), the Paasche index uses current expenditure shares ($w_{kt} = s^*_{kt}$), and as seen in equation (1), the superlative Tornqvist index uses average expenditure shares ($w_{kt} = \frac{s^*_{kt\text{-}1}+s^*_{kt}}{2}$). Hence, the Tornqvist lies between the geometric Paasche and geometric Laspeyres. Diewert (2021) shows that for price indices at a detailed level of aggregation (so that the goods are sufficiently close substitutes), the "standard" ordering occurs - geometric Paasche < Tornqvist < geometric Laspeyres. The Sato-Vartia index uses logarithmic mean expenditure shares, as defined in equation (8).

Traditional price indices have a theory-free interpretation as weighted-average changes in product prices. While this statistical interpretation is valuable on its own, there is also an economic interpretation of these indices dating back to the seminal work of A.A. Konus (Konüs, 1939; Schultz, 1939). The arithmetic Laspeyres and Paasche indices provide upper and lower bounds, respectively, on the exact change in the cost of living between two periods in the absence of product turnover and associated quality change.[31] Superlative indices, including the Fisher and Tornqvist, have more desirable theoretical properties: they are the change in the unit expenditure function (i.e., the exact price index) that is the second-order approximation for a wide class of utility functions in the absence of product turnover and taste shocks (Diewert, 1978).[32]

The traditional superlative price indices require both the *price* and *sales or expenditure share* of each good in either one or both time periods to calculate weighted price changes. As noted in the main text, however, in current practice, statistical agencies' data often does not permit the calculation of such weights. 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.

[31] In the case of strictly normal goods, the arithmetic Paasche is a lower bound of the equivalent variation, and the arithmetic Laspeyres is an upper bound to the compensated variation, so we have that Paasche $\leq$ EV $\leq$ CV $\leq$ Laspeyres. Paasche < Laspeyres typically holds in the data, and will be the case when substitution is, on net, away from goods that have the highest change in price and towards those with the lowest.

[32] A longstanding question in the literature concerned whether the Sato-Vartia index is superlative, until Barnett and Choi (2008) demonstrated that it is. Like the Torqnvist, the Sato-Vartia index is also an expenditure-weighted average of log price changes; it differs from the Tornqvist by using the logarithmic mean of period t-1 and period t expenditure shares instead of the arithmetic mean.

A.2 Hedonic Estimation Details

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.

For the time dummy method, we specify the hedonic regression equation (5) using the same vector of characteristics Z_k in each pair of adjacent periods. Occasionally, new features are introduced to the data. In pairs of adjacent periods entirely prior to the introduction of a new characteristic, it will be omitted from the regression because of collinearity with the intercept term. In pairs of adjacent periods in which the new feature is absent during period t-1 and present during period t, the feature will be included in the estimated regression. Symmetric arguments apply for characteristics that exit.

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

Turning to the hedonic imputation indices, Figure D.1 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.1. 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.2 presents a version of Figure 3 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.

The results presented thus far use quantity-share weights in our hedonic specifications following Bajari et al. (2021). As noted in the text, 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).[33]

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 . Given this, we focus on the results that incorporate chain drift using the GEKS-Lite procedure of Section 4.3. Such results reflect the many different issues including chain drift relevant for comparing alternatives. We report means and standard deviations of annual chained price indices as well as correlations. We included for comparison purposes the traditional 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 traditional Tornqvist. The exception is headphones, for which the expenditure-weighted EP-TV mean is slightly above the traditional Tornqvist. We also report and compare the hedonic Tornqvist indices' differences versus the traditional 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.[34]

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 traditional Tornqvist tends to have slightly lower correlations with the EP-TV-based indices (especially for expenditure-weighted EP-TV vs. traditional 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.

[33]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.

[34]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.

B Using the Nielsen Data

B.1 Nielsen Data Preparation

The Nielsen RMS data consists of more than 100 billion unique observations at the week-store-UPC level. We first aggregate the weekly data to the monthly frequency according to the NRF calendar and then aggregate the monthly data to quarterly. Following procedures used by Hottman et al. (2016) and Redding and Weinstein (2020), we drop outliers from the monthly data before aggregating to the data to quarterly frequency. Specifically, we drop observations with prices above 3 times or below one-third the module-level median for each UPC in a given month. We also drop product-month observations with quantities sold that are more than 24 times that product's median quantity sold per month. One feature of barcoded products is that goods of different sizes and packaging have different barcodes, even if the product contained in the packaging is the same. To ensure comparability between prices, we follow Hottman et al. (2016) and normalize UPC prices to the same units (e.g., ounces), utilizing the size and packaging information provided by Nielsen. Consistent with the literature, we winsorize monthly price changes at the top and bottom 1% of each product group. Tables D.6 and D.7 show our classification of the product groups in the Nielsen 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 Nielsen Consumer Panel and apply the same procedures from Hottman et al. (2016) and Redding and Weinstein (2020) to prepare the data.

B.2 Comparisons of the Nielsen Data to Official Statistics

In this section, we compare patterns of sales and prices for the Nielsen Scanner and Consumer Panel with official statistics. Using the Economic Census data from 2012, we have calculated that the types of retailers that the Nielsen scanner data tracks have very high coverage of food items (about 90%). Moreover, using a back-of-the-envelope calculation based on Nielsen's coverage of different types of retailers, we estimate that Nielsen scanner data accounts for about 41% of total food sales in the U.S. In contrast, the data's coverage is meaningfully lower for several nonfood categories. The types of stores Nielsen tracks accounts for about 53% of small appliance sales. However, Nielsen's coverage of general merchandise stores is only 32%. Using our back-of-the-envelope calculation, these figures imply that the Nielsen scanner data accounts for only about 19% of total small appliance sales in the U.S. Coverage in other categories is substantially lower. For instance, we estimate that the Nielsen scanner data accounts for only about 5% of total sales of hardware and tools.

We have also compared patterns of total expenditures for harmonized categories from Nielsen and Personal Consumption Expenditures data (PCE) from the Bureau of Economic Analysis. We have constructed a concordance between Nielsen and PCE categories at a detailed level (e.g., Bakery) and for broader categories–Food and Nonfood. For prices, we thank the BLS for preparing CPI indices for the broader categories of food and nonfood in a harmonized fashion.[35]

[35] For the broad food and nonfood comparisons with PCE we use a concordance of the 100 plus product

Figure D.7 presents comparisons of nominal expenditures for the broad food and nonfood categories. It is drawn from Cafarella et al. (2023) and reproduced here for convenience. For food, we find nominal sales for the Nielsen Scanner data tracks the PCE closely. The Nielsen Consumer Panel tracks the PCE reasonably well through 2012, but it rises less rapidly than either the Nielsen Scanner or PCE thereafter. For nonfood, both the Scanner and Consumer Panel exhibit less of an increase over time than the PCE.[36]

These patterns are consistent with the discussion in the main text that the Nielsen data's coverage of nonfood items has deteriorated over time. Cafarella et al. (2023) provides additional evidence on this point by comparing the growth of nominal expenditures for detailed categories in the Nielsen Scanner Data from 2008:1 to 2015:4 to the growth in nominal sales for the PCE over the same period.[37] Nominal expenditures grew at the same rate in the Nielsen Scanner Data and the PCE in many of the food product categories. In contrast, the growth rates for Nielsen's nonfood categories tend to be slower than much more variable relative to the PCE.[38]

Figure D.8 presents the relationship between the BLS CPI and corresponding Laspeyres indices from the Nielsen Scanner and Consumer Panel data sets. 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 Nielsen Scanner and Consumer Panel Laspeyres indices are highly correlated with the CPI. In terms of inflation levels, however, the Nielsen Scanner more closely matches the CPI (especially for the arithmetic Laspeyres using the Scanner data). The correlations between Laspeyres indices for the nonfood product groups and the CPI are much weaker than for food (0.53 and 0.67 for the Scanner and Consumer Panel data sets, respectively, using the arithmetic Laspeyres). The average inflation level is closer to the CPI in the Scanner data than in the Consumer Panel.

We interpret these results as providing justification for our focus on food results using the Nielsen Scanner data in the main text. The results also support the view that the Nielsen Scanner data tracks the official statistics as well as, if not more closely than, the Nielsen Consumer Panel.

B.3 Common Goods Rules – Consumer Panel and Retail Scanner

This section presents sensitivity results to alternative common good rule approaches for both the Nielsen Scanner and Nielsen Consumer Panel data sets. Using the scanner data, Figure D.9 compares the results of imposing common goods rules using the 2-quarter horizon, as in the main text (i.e., using percentiles from sales pooled over the current and prior periods),

groups in the Nielsen data with the PCE. When we examine more detailed categories we 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.

[36] For our analysis of the Retail Scanner we use the NRF calendar, while for the Consumer Panel we use the regular calendar. This difference is not important for the patterns reported in this and the next sections. The NRF calendar is especially relevant at the monthly frequency.

[37] Even though our concordance is at the product module level, the categories are more aggregated than Nielsen product groups.

[38] In unreported results, we find similar patterns for the Nielsen Consumer Panel data.

vs. a 5-quarter horizon (i.e., computing percentiles for sales pooled over quarters t and $t-4$).[39] These alternative CGRs impose different duration-based restrictions on products to be included in the set of common goods. The 2-quarter horizon CGR requires goods to be present in periods t and t-1, while the 5-quarter horizon requires goods to be present in periods t and $t-4$. The figure shows that the 5-quarter CGR using a 10th-percentile share threshold leads the CUPI to measure inflation between what is measured using 25th and 50th percentile thresholds using the 2-quarter horizon. The longer-horizon CGR puts additional weight on the goods that have been present in the marketplace for a longer time, which moves our approach in the direction of the duration-based CGR approach of Redding and Weinstein (2020).

Figure D.10 shows the sensitivity of the CUPI to different CGRs using the Nielsen 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 Nielsen 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.11 shows various price indices calculated using all product groups in the Nielsen 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.12, 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 Y-o-Y 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 Y-o-Y index imparts a duration-based component to the CGR in addition to the expenditure share-based thresholds.[40]

Figure D.13 shows related indices, using the Nielsen Scanner data, pooling all product groups, and using various CGRs based on sales percentiles computed over the 5-quarter horizon.[41] These results are therefore suggestive of the results applying the empirical approach in Redding and Weinstein (2020) to the Scanner data would produce. The CUPI with no CGR suggests deflation of 10 percent or more per year. Even the CUPI with a 25th-percentile cutoff rule shows persistent deflation in the Retail Scanner data; imposing a 50th-percentile CGR brings the CUPI closer in line with the Laspeyres index. The series labeled "CUPI, RW CP" shows results from applying the market share threshold in the 5th-percentile CGR from the Consumer Panel to the Scanner Panel data, rather than calculating a percentile-based threshold directly from the Scanner Panel data. Using the Consumer Panel share threshold

[39] In many of the figures of this appendix, we include the arithmetic Laspeyres as this facilitates comparison with Redding and Weinstein (2020). The prior section shows arithmetic and geometric Laspeyres yield similar patterns.

[40] We note that we do not impose a CGR in computing the other price indices shown in Figure D.11. 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 9 that shows that is sensitive to the CGR for Nielsen 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.

[41] Figure D.13 also displays the Bureau of Labor Statistics' Consumer Price Index for all of the product groups included in the Nielsen data as a point of reference.

for the CGR produces results similar to using the 50th-percentile CGR calculated directly in the Scanner Panel data.

The lower inflation rates the CUPI measures in the Nielsen Retail Scanner data relative to the Consumer Panel data highlight the scanner data's large number of very low-market share products. This long tail disproportionately impacts the CUPI. In contrast, the Laypeyres and Feenstra indices are much more consistent between the Nielsen Consumer Panel and Nielsen Retail Scanner data.

Figure D.14 displays for the nonfood product groups the analogous plots to Figure 10, which displays results for food product groups. For comparability purposes to the those in the main text, the CGR rules in this figure are based on sales percentiles over the 2-quarter horizon.[42]

The main message from this analysis is that the CUPI is very sensitive to the specification of the CGR, both in the Nielsen Consumer Panel and in the Nielsen Scanner data. This sensitivity applies both to the market share threshold used 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.4 Machine Learning and Hedonics

This appendix summarizes the Cafarella et al. (2023) procedure for incorporating machine learning into hedonic estimation.

Using machine learning (ML) methods to estimate hedonic price indices requires making several practical choices regarding the architecture of the ML system used for prediction and the conversion of those predictions into price indices. As discussed in the main text of this paper, our preferred approach to constructing hedonic price indices is the "time-varying unobservables" hedonic imputation approach of Erickson and Pakes (2011). The core of this method is to estimate price *levels* for each product in each period in a first step. In a second step, this approach estimates price *changes*, using the hedonic residual (or prediction error) from the first step as a predictor. This methodology allows the hedonic predictions partially to capture unobserved product characteristics' influence on price changes.

In many ways, the "EP-TV" approach of Erickson and Pakes (2011) can incorporate ML methods quite naturally. The key innovation is to use ML methods rather than standard regression techniques to estimate the hedonic functions for log price levels and changes in equations (2) and (4). Another important difference from the more standard econometric procedures we employ in the NPD data is that the Nielsen data available from the Kilts Center does not include pre-coded item-level product attributes. Attribute information is limited to short, non-standard text descriptions. We use deep neural networks to predict product prices and price changes from these product descriptions. Cafarella et al. (2023) describes our approach in detail.

[42]To be consistent with the results for food reported in the main text, Laspeyres is geometric in this figure.

B.5 Assessing the National Market Assumption in the Nielsen Data

As noted in Section 4.4, our empirical implementation of the CUPI relies on the assumption of a unified national market. Because the CUPI is not consistent in aggregation and includes unweighted geometric mean terms, any failure of this assumption may affect the CUPI more than other indices. In this section, we assess the realism of the national market assumption in the Nielsen scanner data.

In Figure D.6, we pool the Nielsen item-level data for food product groups at the weekly frequency from 2006–15. We then compute the market penetration of items in the pooled data both on an unweighted basis (i.e., all items get the same weight) and on a sales-weighted basis. Market penetration is defined as the share of Nielsen metro areas in which the item-level week is observed to have positive sales.

On an unweighted basis, the distribution is very skewed to the left, with most item-level week observations having very low market penetration. Almost all of the unweighted distribution has less than a 20 percent market penetration. In unreported results, we find that the mass of the unweighted distribution with the lowest market penetration reflects entering and exiting goods. Even on a sales-weighted basis, only 15 percent of sales are for items with a truly national market, although much of the mass of the distribution has market penetration of over 80 percent of metro areas. These patterns raise questions about applying a national market based CES price index for most items. In Appendix C.3.2, we show that the CUPI can be badly biased when the national market assumption fails, suggesting these patterns may be important for understanding the empirical behavior of the CUPI.

C Additional Evidence on the Behavior of the CUPI

In this appendix, we examine the behavior of the CUPI in additional detail. We begin in Section C.1 by examining empirically whether implementing a nested structure in the CUPI modifies its extremely negative inflation readings. We find that the nesting nesting approaches we have explored do not meaningfully modify the CUPI's measurement of inflation. We then examine analytically and via simulation studies whether time-varying product appeal shocks generate an expected bias in the Sato-Vartia index relative to the consumer's exact price index under CES preferences. In section C.2, we examine the mathematical source of the taste shock bias highlighted by Redding and Weinstein (2020). We conclude that the presence of time-varying product appeal on its own will not generate an expected bias in the Sato-Vartia index. On the other hand, time trends in the *dispersion* of product appeal shocks do introduce an expected bias. We show that this explanation is empirically unlikely, however, to account for the extremely low inflation measured by the CUPI.

A natural question that arises from that conclusion is why the CUPI measures consistently lower inflation than the Sato-Vartia and Feenstra indices. In section C.3, we present simulation evidence showing that geographical segmentation of entering goods and limited availability of existing goods can cause the CUPI to measure significantly lower inflation than is implied by the consumer's unit expenditure function. A common goods rule helps alleviate such biases. We believe these simulations point the way toward future research on the implementation of the CUPI.

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

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

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

Table D.5 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.4 plots nested versions of the CUPI using our two nesting strategies alongside un-nested versions of the CUPI and Feenstra index. Both versions of the CUPI are implemented using a 30th-percentile CGR, applied at the within-nest level in the nested version.[45] The alternative nesting approaches yield similar results, with the nested CUPI tending to show slightly less deflation than the un-nested (or "flat") CUPI. In unreported results, we find that the relationship between the nested and flat CUPIs is

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

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

[45]Nests are weighted by the number of products to adjust for differential product group sizes.

robust to using alternative CGR cutoffs.

C.2 Analytical Characterization of the Taste Shock Bias

We consider a representative consumer with CES preferences. For simplicity, in this subsection we examine a market with no product turnover, and we assume the consumer has non-nested preferences over the set of available products. Let N denote the number of products present in each period and P_t denote the unit expenditure function in period t. Redding and Weinstein (2020) show that, in the presence of product appeal shocks, the change in the log Sato-Vartia price index equals the change in the log unit expenditure function plus an additional term

$$\ln \Phi_t^{SV} = \ln \frac{P_t}{P_{t\text{-}1}} + \left[\sum_k \omega_{kt} \ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right)\right], \tag{C.1}$$

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

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

Redding and Weinstein (2020) label the term in the square brackets of equation (C.1) the "taste shock bias," as it represents the difference between the Sato-Vartia index and the true change in the cost of living index. It is easy to see that when product appeal is constant over time, so that $\varphi_{kt} = \varphi_{kt\text{-}1}$ for every product k, the taste shock bias term will be zero and the Sato-Vartia index will exactly recover the true change in the cost of living. Redding and Weinstein (2020) argue that when product appeal is time varying, however, the taste shock bias term will be positive in expectation, so that the Sato-Vartia index will tend to overstate the true rate of inflation.

The expected taste shock bias can be written as

$$\begin{aligned}\mathbb{E}\left[\ln \Phi_t^{SV} - \ln \frac{P_t}{P_{t\text{-}1}}\right] &= N\mathbb{E}\left[\omega_{kt} \ln\left(\frac{\varphi_{kt}}{\varphi_{kt-1}}\right)\right] \\ &= N\text{Cov}\left[\omega_{kt}, \ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right)\right] + N\mathbb{E}\left[\ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right)\right].\end{aligned} \tag{C.2}$$

The second term in equation (C.2) will be zero due to the normalization. Redding and Weinstein (2020) note, however, that the Sato-Vartia weights ω_{kt} are an increasing function of the appeal parameters φ_{kt}, so

$$\frac{\partial \omega_{kt}}{\partial \phi_{kt}} = \frac{\partial \omega_{kt}}{\partial s_{kt}} \frac{\partial s_{kt}}{\partial \phi_{kt}} > 0 \implies \text{Cov}\left[\omega_{kt}, \ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right)\right] > 0. \tag{C.3}$$

Other factors equal, consumers will devote a greater share of expenditure to goods that experience favorable appeal shocks. In isolation, that tendency would lead the Sato-Vartia taste-shock bias in equation (C.2) to be positive. As Redding and Weinstein (2020) argue in their abstract:

> In the presence of relative taste shocks, the Sato-Vartia price index is upward biased because an increase in the relative consumer taste for a variety lowers its taste-adjusted price and raises its expenditure share. By failing to allow for this association, the Sato-Vartia index underweights drops in taste-adjusted prices and overweights increases in taste-adjusted prices, leading to what we call a "taste-shock bias."

We believe that this intuition, while correct on its own, is also incomplete: there is a symmetrical and offsetting tendency for appeal shocks to induce a downward bias in the Sato-Vartia index when the appeal parameters φ_k are independently and identically distributed across periods t-1 and t. The offsetting bias comes from the fact that the Sato-Vartia weights ω_{kt} are also an increasing function of the *previous* period's appeal parameters $\varphi_{kt\text{-}1}$, which enter the second term in the covariance, $\ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right)$, in the opposite direction from the current period's appeal parameters. Hence,

$$\frac{\partial\omega_{kt}}{\partial\varphi_{kt\text{-}1}} = \frac{\partial\omega_{kt}}{\partial s_{kt\text{-}1}}\frac{\partial s_{kt\text{-}1}}{\partial\varphi_{kt\text{-}1}} > 0 \implies \mathrm{Cov}\left[\omega_{kt}, \ln\left(\frac{\varphi_{kt}}{\varphi_{kt\text{-}1}}\right)\right] < 0. \tag{C.4}$$

This offsetting tendency would lead the Sato-Vartia taste-shock bias to be negative in isolation. The upward and downward biases will offset each other in expectation when the appeal parameters are identically distributed across periods t-1 and t, so the Sato-Vartia index will not exhibit a generic taste-shock bias under those assumptions.

Nonetheless, if the assumption of idiosyncratically and identically distributed appeal parameters does not hold precisely, for instance, because the dispersion of product appeal changes over time, the Sato-Vartia price index may exhibit a taste-shock bias. In particular, as noted by Redding and Weinstein (2020), increasing dispersion in product appeal will induce an upward bias in the Sato-Vartia index, while the CUPI will remain unbiased.

We examine the empirical relevance of this explanation for the empirical gap in Figure D.5. 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 where the S^* ratio is highly negative (see, e.g., boys' jeans in Figure 4). In other words, the CUPI's ubiquitous finding of rapid deflation without using a common goods rule is not readily justified by the observed patterns of dispersion in product appeal.

C.3 Simulation Evidence on the Behavior of the CES Exact Price Indices

C.3.1 Simulation Model Environment

We base our simulations on the general equilibrium environment of Hottman et al. (2016).[46] A set of firms, indexed by f, each produces multiple products, indexed by u. Consumers have nested preferences, with preferences over the total output of each firm in the upper-level nest and preferences over the individual products supplied by each firm in the lower-level nests.

The bottom-level CES consumption index over the products supplied by firm f, C^F_{ft}, is given by

$$C^F_{ft} = \left[\sum_{u \in \Omega^U_{ft}} \left(\varphi_{ut} C_{ut} \right)^{\frac{\sigma^U - 1}{\sigma^U}} \right]^{\frac{\sigma^U}{\sigma^U - 1}}, \tag{C.5}$$

where C_{ut} represents the quantity consumed of product u in period t, φ_{ut} is a product-level appeal shifter for product u, Ω^U_{ft} is the set of products supplied by firm f in period t, and σ^U is the elasticity of substitution among the products supplied by a firm.

The consumer's utility from consuming the output supplied by all firms, U_t, is given by

$$U_t = \left[\sum_{f \in \Omega^F_t} \left(\varphi_{ft} C^F_{ft} \right)^{\frac{\sigma^F - 1}{\sigma^F}} \right]^{\frac{\sigma^F}{\sigma^F - 1}}, \tag{C.6}$$

where C^F_{ft} is the firm-level consumption aggregate defined in equation C.5, φ_{ft} is a firm-level appeal shifter for firm f, Ω^F_t is the set of firms supplying products in the marketplace in period t, and σ^F is the elasticity of substitution across firm-level consumption aggregates.

It is necessary to provide a normalization for the product-level and firm-level appeal shifters φ_{ut} and φ_{ft}. We follow Redding and Weinstein (2020) in assuming that both product-level and firm-level appeal shifters for continuing products have an average log change of zero in every period. That normalization allows for the possibility that entering or exiting products have higher or lower average appeal levels than continuing products.

The supply side of the market is populated by a set of firms that produce output using a composite input factor that serves as the economy's numeraire good. Firms' cost functions are assumed to be additively separable across products supplied. The total variable cost of producing Y^U_{ut} units of product u at time t, A_{ut}, is given by

$$A_{ut} \left(Y^U_{ut} \right) = a_{ut} \left(Y^U_{ut} \right)^{1+\delta}, \tag{C.7}$$

where a_{ut} is a marginal cost shifter of producing product u at time t, and δ is the elasticity of marginal costs with respect to output. We assume that product entry and exit is exogenous

[46] Hottman et al. (2016) consider consumers with Cobb-Douglas preferences over a number of different product groups and constant elasticity of substitution (CES) preferences within each product group. For simplicity, we restrict our attention to consumers with preferences over products within a single group.

in our simulations.

Firms choose prices under Bertrand competition. Each firm's decisions affect other firms' decisions only through their effects on the economywide price index. In equilibrium, firms choose product prices to maximize profits and consumers choose quantities demanded of each product. We will generally assume that the market clears so that $C_{ut} = Y_{ut}^{U}$ for every product u and time t. Certain simulations will feature market imperfections that prevent this market-clearing condition.

Hottman et al. (2016) derive analytical formulas for consumers' product demands, firms' pricing rules, and firm-level and aggregate price indices in this environment, and provide computer code to solve for the market-clearing general equilibrium numerically. They also characterize the economics of the market environment in depth. We build our numerical simulations on the code provided by Hottman et al. (2016), so our environment will parallel theirs except for the differences that we highlight to explore the behavior of the CES exact price indices in various market environments.

Each simulation contains 50 firms and lasts for 40 periods.[47] Unless otherwise noted, each firm sells 50 products in each period. 100 Monte Carlo simulations were run for each set of model parameters considered. To abstract from issues of within-firm vs. between-firm substitution, we set the elasticity of substitution between a firm's individual products σ^U equal to the elasticity of substitution across firms' composite output σ^F.[48] We choose a value of 5 for both elasticities, between the values of σ^U and σ^F in Hottman et al. (2016) of 7 and 4, respectively. We set the elasticity of marginal costs with respect to quantity supplied δ to 0.15, consistent with the Monte Carlo simulations in Hottman et al. (2016).

Each period, the log product appeal shifters are drawn from normal distributions with product-specific means and variances. The product-specific means are drawn from a standard normal distribution, and the product-specific variances are drawn from a uniform distribution between one and two. Product-specific means and variances are constant over time, unless otherwise noted. Each period, the log firm appeal shifters are drawn from normal distributions with zero means and firm-specific variances. The firm-specific variances are drawn from a uniform distribution between one and two, and are constant over time. Finally, the log marginal cost shifters are drawn each period from normal distributions with zero means and product-specific variances. The product-specific variances are drawn from a uniform distribution between one and two, and are constant over time. The product appeal shifters, firm appeal shifters, and marginal cost shifters are mutually independent.

The econometrician is assumed to be able to observe the elasticities of substitution σ^U and σ^F exactly without estimation in constructing the price indices. The CES exact price indices are calculated without considering nesting of preferences among products and firms, with the exception of the unit expenditure function, which is calculated according to the consumer's exact preference structure.

[47]We initialize all stationary variables by drawing from their steady-state distributions, so the simulations do not include a burn-in period.

[48]Hottman et al. (2016) found evidence of such differences and we find related evidence of differences in elasticities within and across nests. We found in section 4.1.2 this did not matter much for the properties of the CUPI. More work is needed in this area, but we do not explore this issue in our simulation analysis . Relatedly, an interesting and open question is how much the CUPI is sensitive to any biases in the estimation of the elasticities. We leave that question for future work.

C.3.2 Simulation Evidence

We consider five sets of simulations in this section. In each set of simulations, we vary one key parameter and run 100 Monte Carlo simulations as described in the previous section for each value of the key parameter we consider in the set of simulations. The figures display inflation as measured by the unit expenditure function and various CES price indices; the lines represent the average realization of measured inflation using each price index, while the shaded regions represent 95-percent asymptotic confidence intervals. The first three sets of simulations consider frictionless markets in which the assumptions underlying the CUPI hold exactly, so it coincides identically with the unit expenditure function in those exercises. The fourth and fifth sets of simulations introduce market imperfections that drive a wedge between the CUPI and the unit expenditure function.

Trends in Marginal Costs

Figure D.15 explores the behavior of the Sato-Vartia index and CUPI in the environment of Hottman et al. (2016) when there is a trend in the marginal cost shifter a_{ut}. On the left-hand side of the graph, marginal costs are falling at a rate of 5 log points per period; in the middle of the graph, marginal costs have no trend; and on the right-hand side of the graph, marginal costs are rising at a rate of 5 log points per period. These trends in marginal costs drive non-zero average inflation. In this frictionless environment, the CUPI exactly replicates the unit expenditure function. The Sato-Vartia index is substantially less precise than the CUPI, as seen in its wider 95-percent simulation bands for estimated inflation. The Sato-Vartia index is noisier than the CUPI because it does not account for changes in product appeal; despite the normalization that average appeal levels are steady over time in these simulations, appeal shocks may affect the consumer's cost of living in any particular simulation. Generally speaking, if goods with large expenditure shares experience positive appeal shocks on average, the cost of living will fall, but if they experience negative appeal shocks on average, the cost of living will rise. Consistent with the logic in Section C.2, however, the Sato-Vartia index does not display an average bias relative to the unit expenditure function.

Trends in Variance of Product Appeal

Figure D.16 explores the behavior of the Sato-Vartia index and CUPI when there is a trend in the variance of the product appeal parameters φ_{ut}. The horizontal axis of the graph shows different growth rates for the variance of appeal; on the left-hand side of the graph, appeal is becoming more compressed over time, while on the right-hand side of the graph, appeal is becoming more dispersed over time. The unit expenditure function shows that the consumer's cost of living is falling over time when the variance of product appeal is rising, and conversely the cost of living is rising when the variance of product appeal is falling over time. This pattern is consistent with the logic in Redding and Weinstein (2020) that increasing dispersion in product appeal is valuable to consumers when products are substitutes, because it provides greater opportunities for substitution to preferred varieties. In contrast to the results in Figure D.15, the Sato-Vartia index does exhibit an average bias in the presence of time trends in the variance of product appeal, which is especially evident in the right-hand portion of the figure where the variance is growing over time. This figure helps illustrate the potential benefits of using the CUPI.

As noted above, however, rising dispersion in product appeal seems empirically unable to account for the rapid deflation implied by the CUPI. Figure D.5 shows that several product groups in the NPD data exhibit nonmonotonic patterns in measured appeal dispersion, even where there is a large gap between the CUPI and Feenstra index. We therefore conclude that rising dispersion in product appeal is unlikely to account for the empirical gap between the CUPI and the other exact CES price indices.

Product Upgrading and Downgrading via Turnover

Figure D.17 displays results from simulations featuring product entry and exit. For simplicity, we assume that products are present in the market place for a deterministic number of periods (set to five in these simulations) after which they exit. Equal numbers of products enter and exit the market in every period.

The key feature of the simulations is that the average appeal parameter φ_{ut} for entering products can differ from the average for continuing products.[49] The horizontal axis of the graph shows different trends in the average appeal of entering products. On the left-hand side of the graph, entering products are less appealing on average than existing products, while on the right-hand side of the graph, entering products are more appealing.

Figure D.17 shows inflation as measured by the Sato-Vartia index, the CUPI, and the Feenstra index, which is equal to the Sato-Vartia index in the absence of product entry and exit. The CUPI again tracks the true unit expenditure function exactly, showing inflation from product downgrading and deflation from product upgrading. The Sato-Vartia index captures these effects directionally, because product turnover affects the prices of continuing products via competition. Because it considers only continuing products, however, the Sato-Vartia index quantitatively understates product turnover's effects on the cost of living. The Feenstra index augments the Sato-Vartia index with an adjustment term that captures the effects of product turnover directly. Figure D.17 shows that it is unbiased on average relative to the true unit expenditure function, despite the presence of relative appeal shocks in the simulations. Echoing the results of Figure D.15, the Sato-Vartia index and the Feenstra index are noisier than the CUPI because they do not account for the effect of product appeal shocks. This figure helps make the case for using an index such as the Feenstra or CUPI to incorporate product turnover that yields quality change.

Segmented Markets

Figure D.18 displays results from a set of simulations in which the market is segmented into five distinct submarkets. Consumers have nested CES preferences over the products consumed in each submarket and firms compete within each submarket as described in Section C.3.1. Consumers have Cobb-Douglas preferences over their consumption across the various submarkets. 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

[49]Recall that the normalization on product appeal in Redding and Weinstein (2020) applies only to continuing products, so product upgrading or downgrading does not violate the normalization.

each. Product entry and exit within each market otherwise proceeds as in the previous set of simulations.

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 are meant to mimic the pattern documented in Figure D.6, which shows that although most sales are concentrated among products sold in nearly all metro areas nationally, on a UPC basis, most products are sold in relatively few areas.

As in Figure D.17, the horizontal axis of Figure D.18 shows different trends in the average appeal parameter φ_{ut} of entering products. Only the small markets feature a trend in the average appeal of entering products; there is no trend in the large market. Figure D.17 displays inflation as measured by five price indices in addition to the unit expenditure function: the Sato-Vartia; the Feenstra; the CUPI with no common goods rule, which we have called the "theoretical CUPI"; the CUPI implemented with a 40th-percentile common goods rule; and the CUPI implemented with an 80th-percentile common goods rule.

Figure D.18 conveys a few key messages. First, the theoretical CUPI is significantly biased in the presence of product upgrading or downgrading in the small markets. The intuition for this bias is that the $P^{\star}$ and $S^{\star}$ ratio terms in the CUPI 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. The second key message is that the Sato-Vartia and Feenstra indices fare better in these simulations than the theoretical CUPI because all of their components are expenditure-share weighted. The third key message is that a common goods rule (CGR) can help reduce 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.18 thus provides a theoretical justification for the use of a CGR in Redding and Weinstein (2020) and our own empirical work. We interpret this segmented markets case as broadly capturing the intuition for a CGR given that goods may first enter local markets. While this exercise helps justify a CGR, it highlights that choosing the 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. Put differently, although this exercise provides theoretical motivation for a CGR, it does not provide precise guidance as to the nature of the appropriate CGR.

Partial Stock-outs (Rationing) Prior to Exit

Figure D.19 examines the behavior of the CES exact price indices when there are partial product stock-outs in the period prior to exit. 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. Product entry and exit within each market otherwise proceeds as in the previous two sets of simulations.

The horizontal axis of the figure shows various shares of rationing prior to exit. On the left-hand side of the figure, consumers are only able to purchase 10 percent of their desired (unconstrained) product demands; on the right-hand side of the figure, there is no rationing. We assume that firms do not adjust stocked-out products' prices to clear the market, but instead price all products as they would in the flexible price equilibrium. We assume that consumers optimally reallocate their demands toward the unconstrained products in response to the rationing.[50]

The unit expenditure function in Figure D.19 shows an approximately constant cost of living in the presence of stock-outs. Although the simulations feature product turnover, they do not feature any trend in average appeal of entering products. As the figure shows, though, stock-outs introduce a substantial bias to the CUPI and the Feenstra index. The intuition for the bias is subtle. Rationing lowers expenditure shares on goods just prior to their exit from the marketplace, with the expenditure reallocated to unconstrained goods. Rationing therefore raises the dispersion of expenditure shares on continuing goods relative to the un-rationed case, leading to a negative log $S^{\star}$ ratio.[51] Likewise, the Feenstra adjustment to the Sato-Vartia index is 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, producing a negative adjustment to both the Feenstra index and the CUPI.

The key messages from Figure D.19 are similar to those from Figure D.18. The theoretical CUPI is significantly biased in the presence of this market friction, while the Sato-Vartia index is approximately unbiased. Imposing a CGR helps move the CUPI closer to the true unit expenditure function. Again, though, the simulations do not provide guidance on the empirically appropriate CGR. Estimates of the extent and nature of rationing are needed to yield guidance for the appropriate CGR.

[50]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.

[51]The entry of the unconstrained goods does not affect this calculation, because the expenditure shares in the CUPI's consumer valuation adjustment term are calculated over continuing goods only.

D Appendix Tables and Figures

Table D.1: Impact of Alternative Imputation for Missing Price of Entrants, NPD Data

	Memory Cards	Coffeemakers	Headphones	Boys' Jeans	Occupational Footwear
Hedonic Tornqvist (EP-TV)	-20.12	-9.56	-14.13	-9.16	-3.79
Hedonic Tornqvist (EP-TV, alternative imputation)	-20.06	-9.57	-13.74	-8.92	-3.78

Notes: The "EP-TV" method in the first row uses the (Erickson and Pakes, 2011) method imputing the missing price relative for entrants assuming the lagged residual is zero. The "EP-TV, alternative imputation" method in the second row imputes the missing price relative using the current period residual. Reported are annual averages of chained price indices for Q4. Data come from the NPD Group.

Table D.2: Comparison of GEKS-Lite and Rolling-Year GEKS, Annual Chained Price Indices, NPD Data

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

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

Table D.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))	-20.60	-10.06	-14.51	-7.94	-3.76
SD (Tornqvist (EP-TV,QW))	8.57	4.53	5.39	3.23	3.28
Mean(Tornqvist (EP-TV,EW))	-16.89	-8.21	-11.07	-6.72	-3.14
SD(Tornqvist (EP-TV,EW))	9.45	4.60	5.05	3.20	3.01
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))	1.48	1.57	-0.48	1.17	0.83
Difference(Tornqvist, Tornqvist (EP-TV,QW))	5.19	3.42	2.96	2.39	1.45
Difference(Sato-Vartia, Feenstra)	2.14	3.07	1.72	1.38	1.68
Corr(Tornqvist, Tornqvist (EP-TV,QW))	0.99	0.91	0.96	0.97	0.99
Corr(Tornqvist, Tornqvist (EP-TV,EW))	1.00	0.84	0.98	0.98	0.98
Corr(Tornqvist (EP-TV,QW), Tornqvist (EP-TV,EW))	1.00	0.97	1.00	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	0.83	0.98	0.96	0.97
Corr(Feenstra, Tornqvist (EP-TV,EW))	0.97	0.71	0.95	0.97	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	0.87	0.99	0.94	0.99
Corr(Sato-Vartia, Tornqvist (EP-TV,EW))	0.98	0.80	0.94	0.96	0.98
Corr(Sato-Vartia, Feenstra)	0.99	0.98	1.00	0.97	0.92

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

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

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

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

Table D.5: Nested Estimated Elasticities of Substitution: NPD Data

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

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

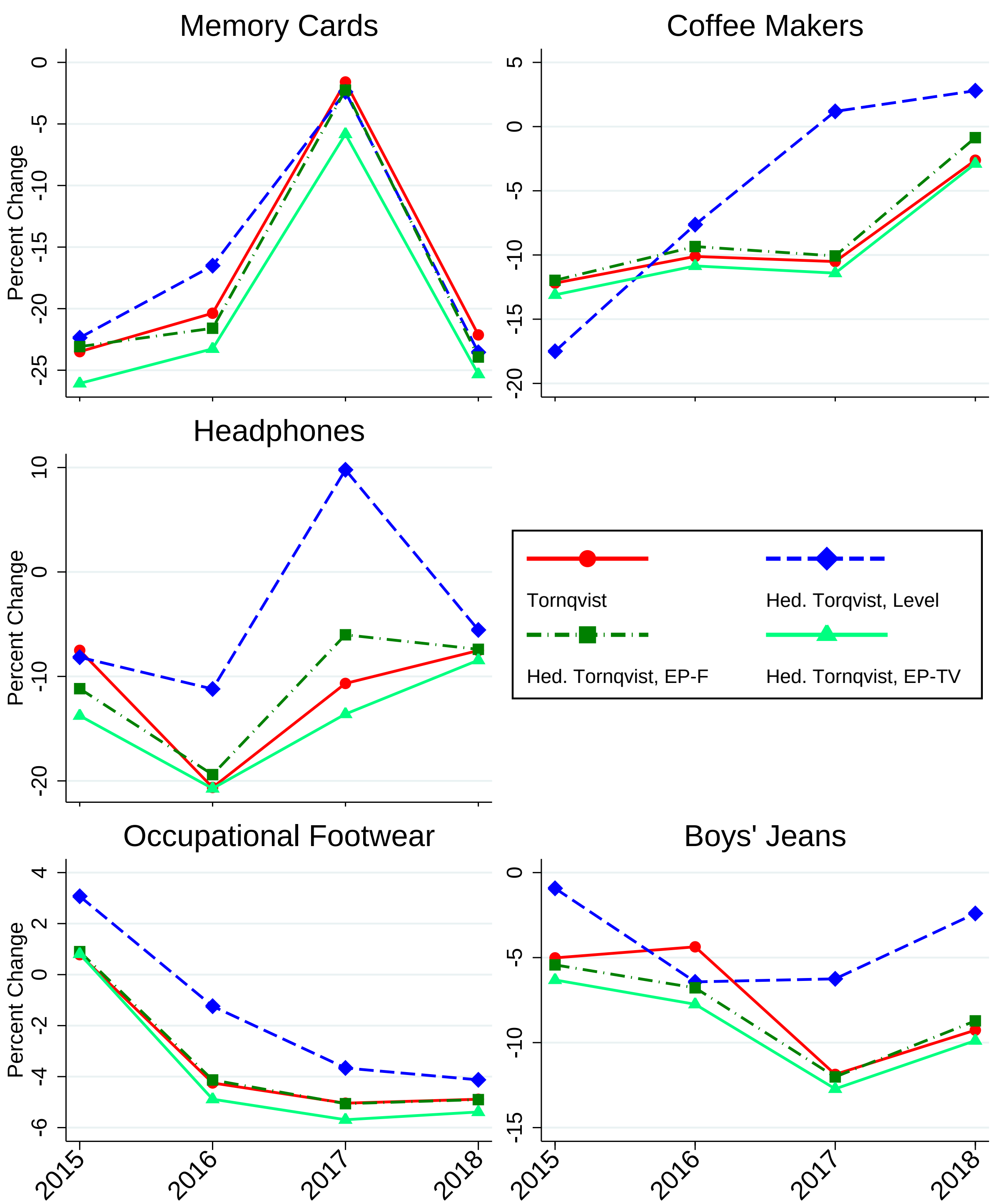
Table D.6: Food Product Groups: Nielsen Retail Scanner Data

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

Table D.7: Nonfood Product Groups: Nielsen Retail Scanner Data

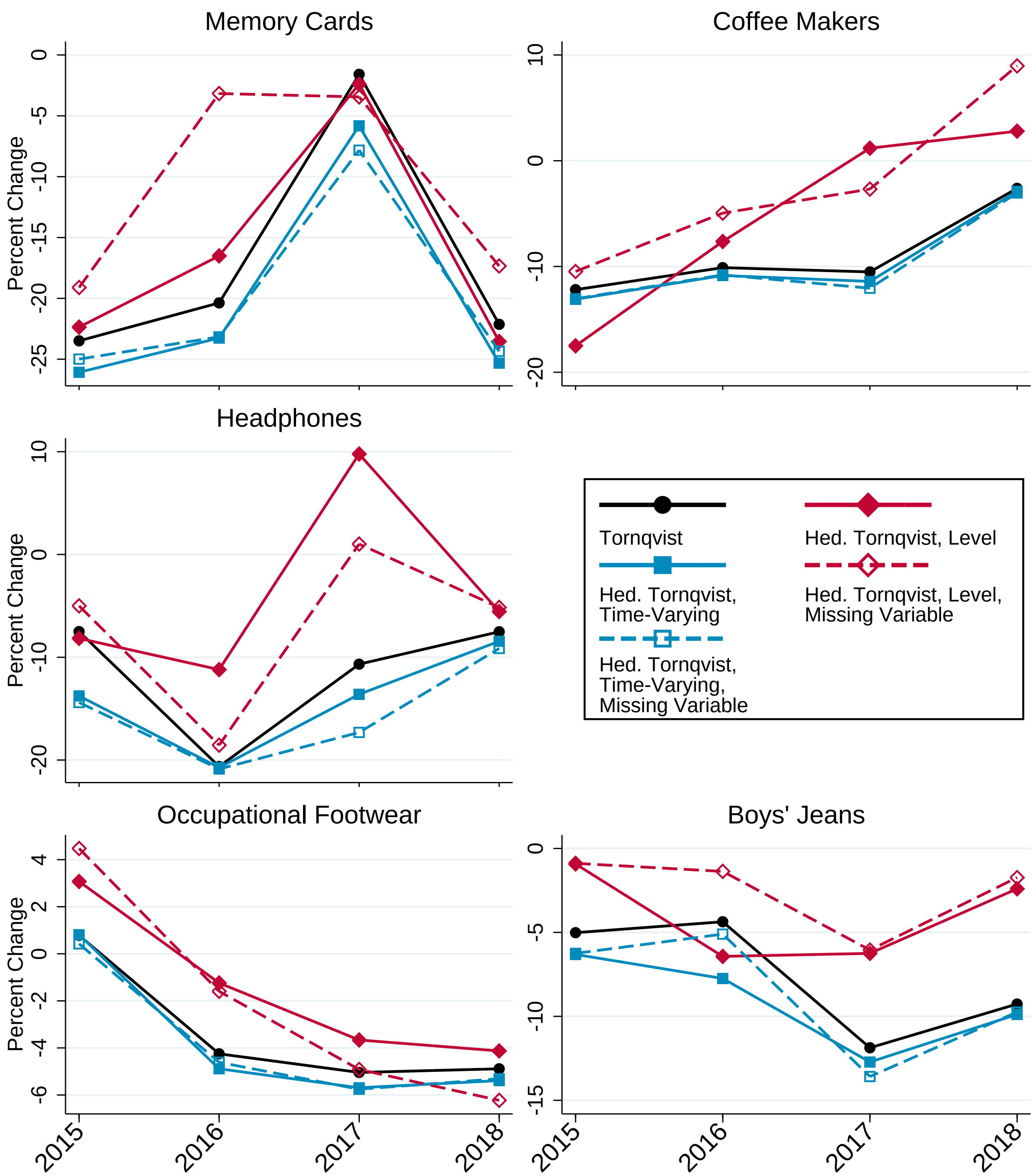
Product Group	Product Group Code	Product Group	Product Group Code
Automotive	5501	Housewares, Appliances	5513
Baby Needs	6001	Ice	2004
Batteries And Flashlights	5502	Insecticds/Pesticds/Rodenticds	5514
Beer	5001	Kitchen Gadgets	5515
Books And Magazines	5503	Laundry Supplies	4506
Canning, Freezing Supplies	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		

Figure D.1: Alternative Hedonic Estimation Strategies: NPD Data



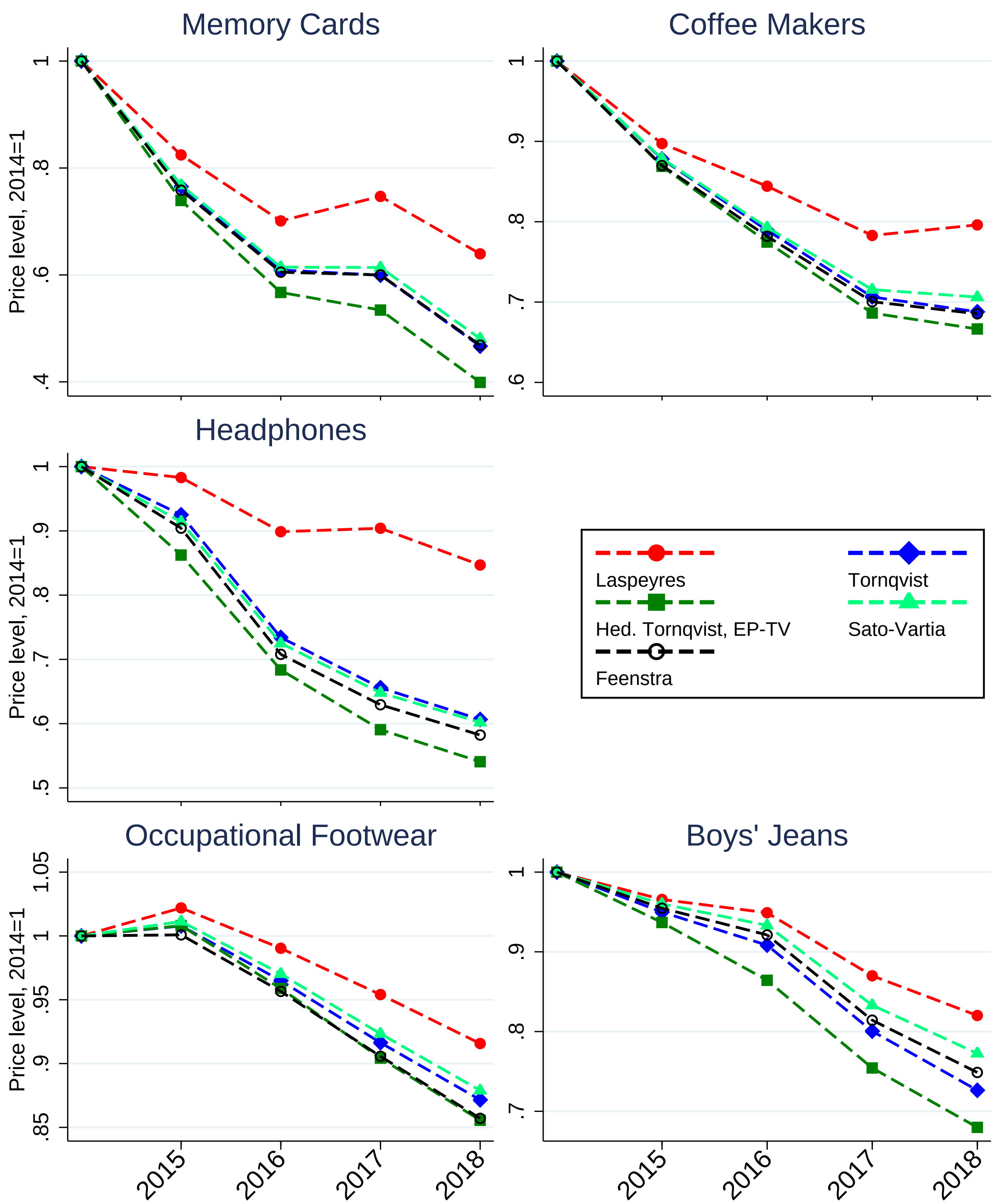
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. Data comes from the NPD Group.

Figure D.2: Test of Time-Varying Unobservable Hedonic Specification
First-Differences and Levels Estimation: NPD Data



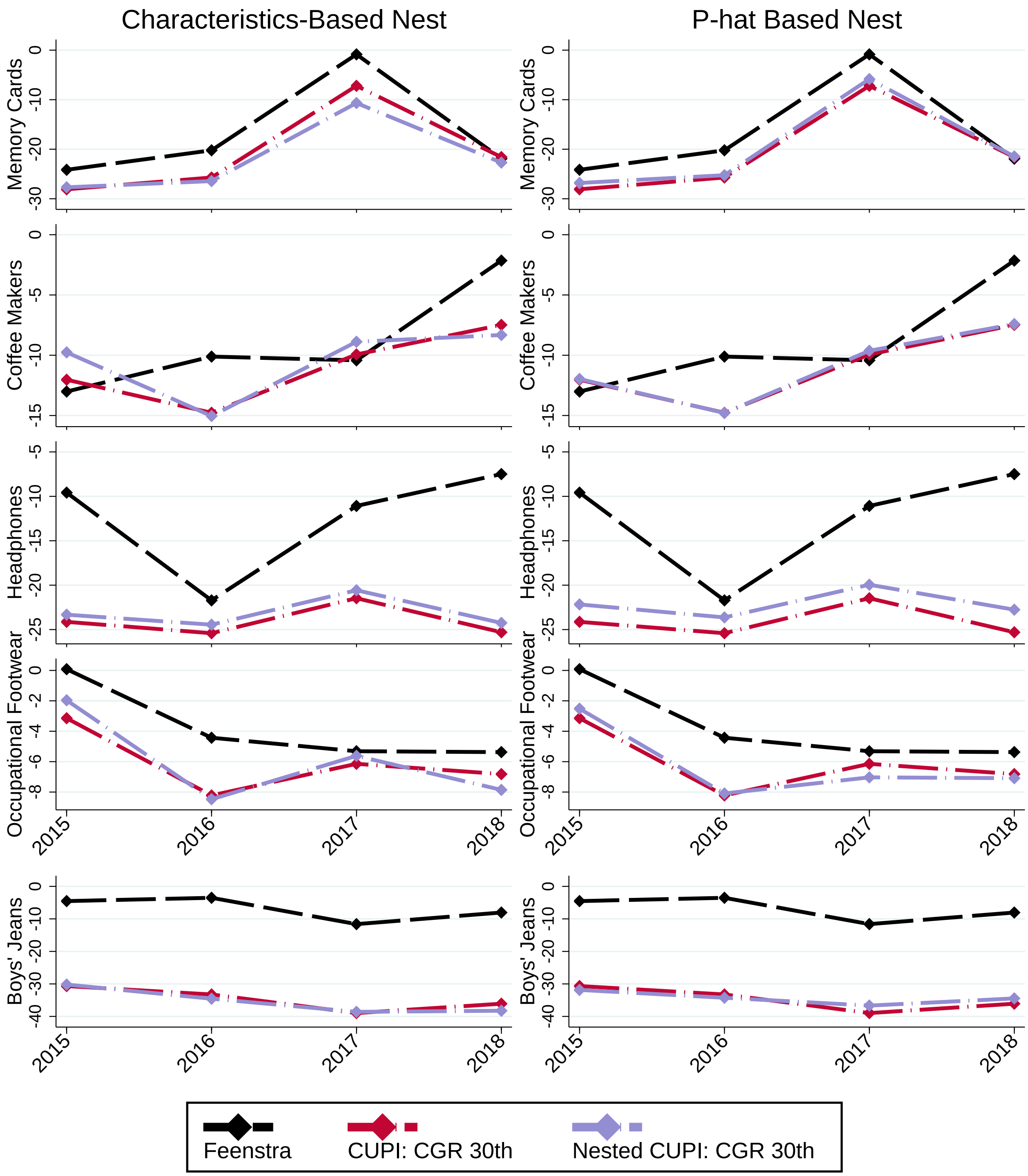
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. Data comes from the NPD Group.

Figure D.3: Main Price Index Specifications
Cumulative Price Level Changes, No CUPI: NPD Data



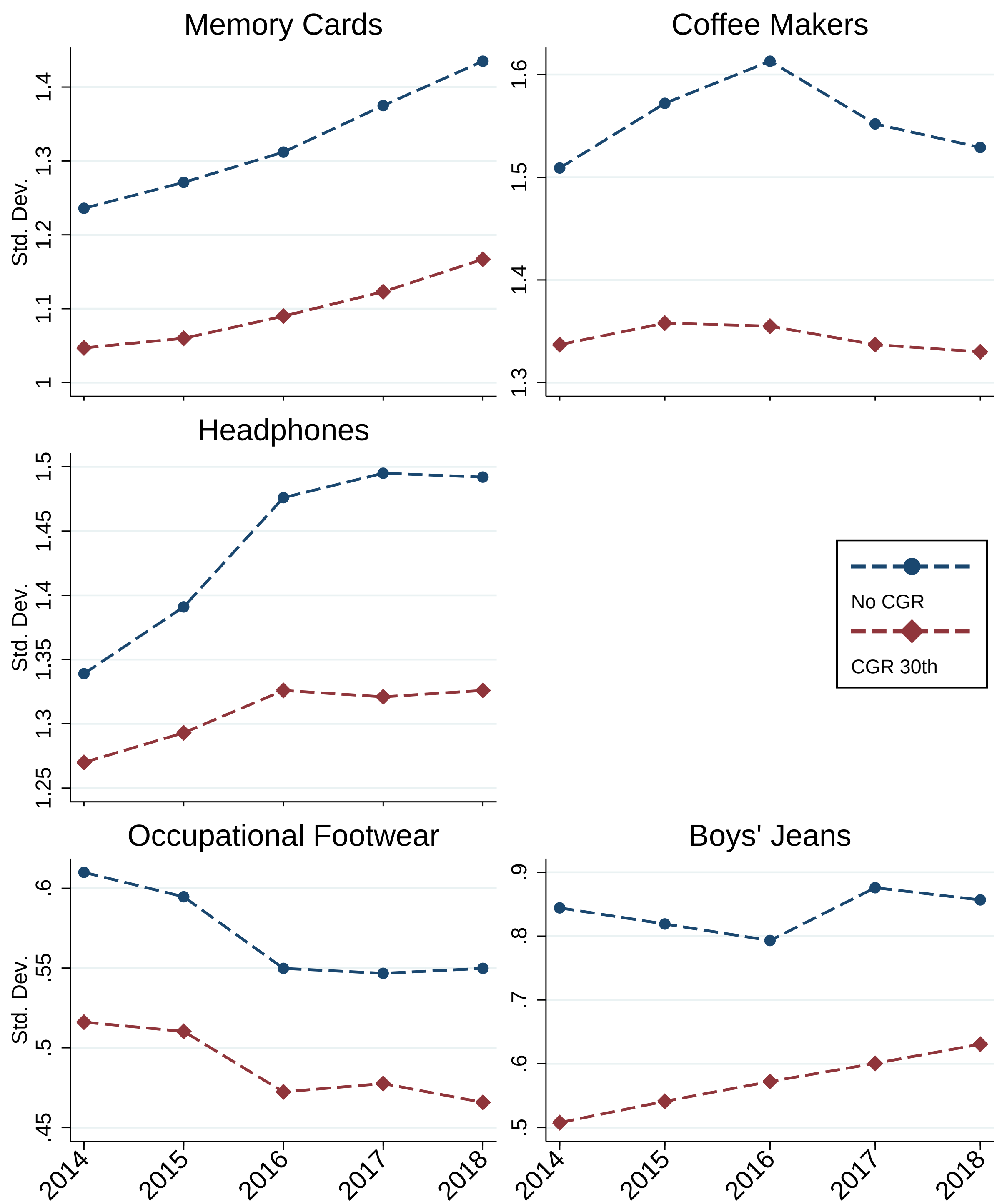
Notes: Values are cumulative changes relative to the 2014 price level, with 2014 price level set to 1. The hedonic time-varying unobservables model is estimated over log price differences using WLS and with weights that are average quantity-shares in adjacent periods. Data comes from the NPD Group.

Figure D.4: Nested CUPI: Characteristics- and P-Hat- Based Nests
Percent Changes: NPD Data



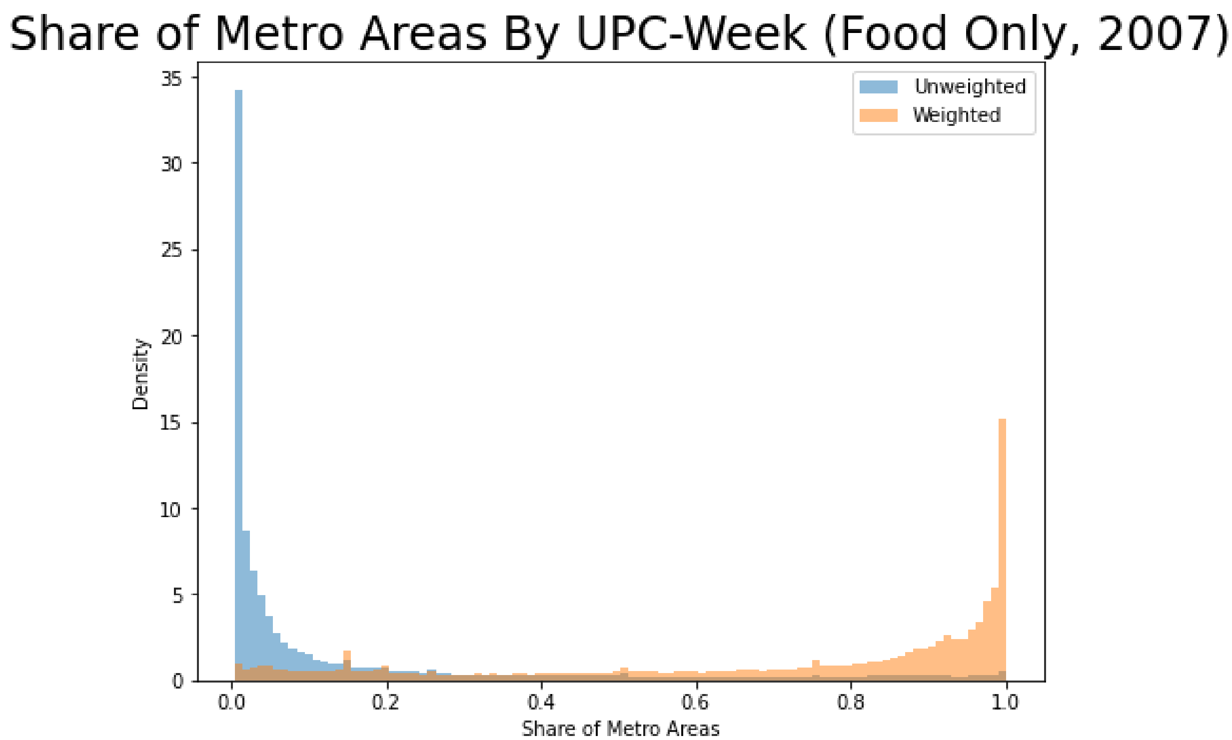
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. Laspeyres is the geometric Laspeyres. For the characteristics-based nests, we assign items to groups based on shared observable characteristics. The p-hat based nests are based on the decile of predicted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

Figure D.5: 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. Data comes from the NPD Group.

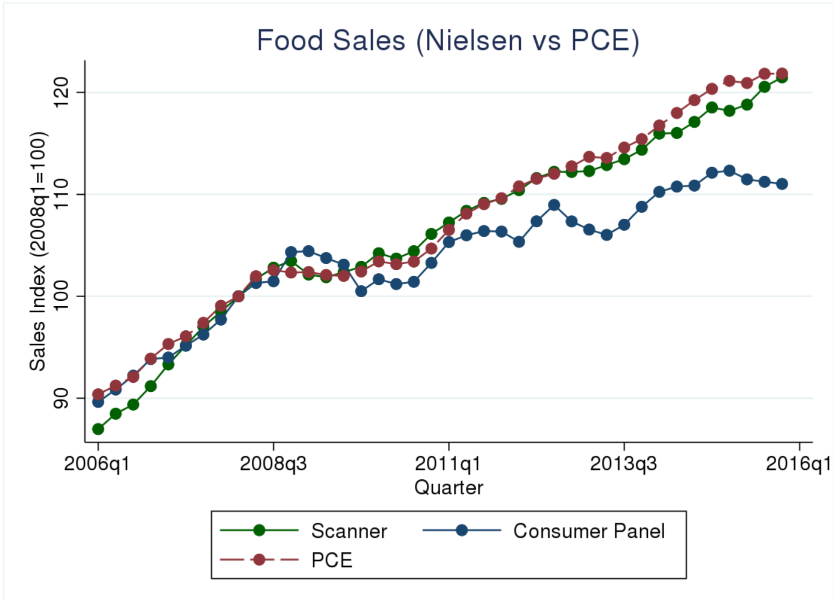
Figure D.6: Sales-weighted and Unweighted Distributions of Market Penetration of Items: Nielsen Food



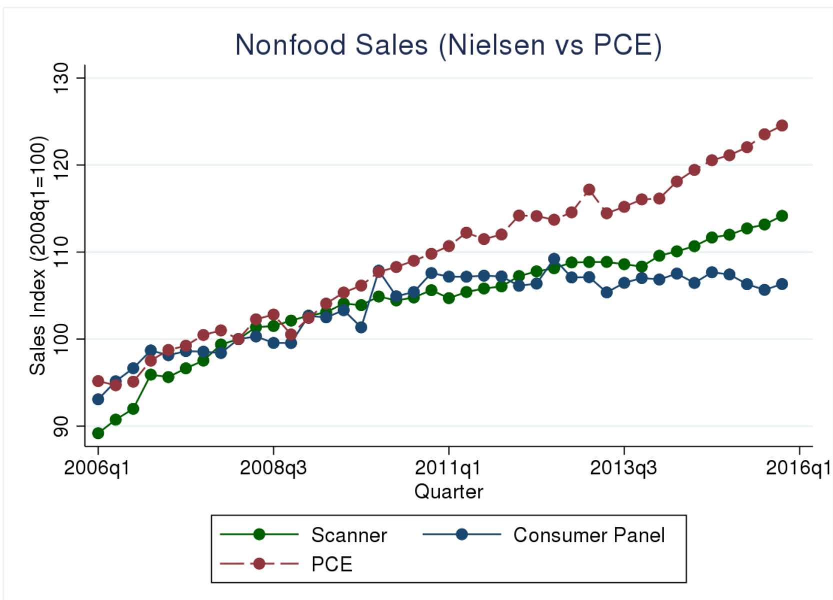
Notes: All UPC items at a weekly frequency are used from 2006-2015. Unweighted shows the market penetration at the metro area of the unweighted pooled distribution. Sales-weighted shows the equivalent using sales weights. Figure uses Nielsen Retail Scanner data for food product groups.

Figure D.7: PCE vs Nielsen Sales for Scanner and Consumer Panel, Food and Nonfood

(a) Food



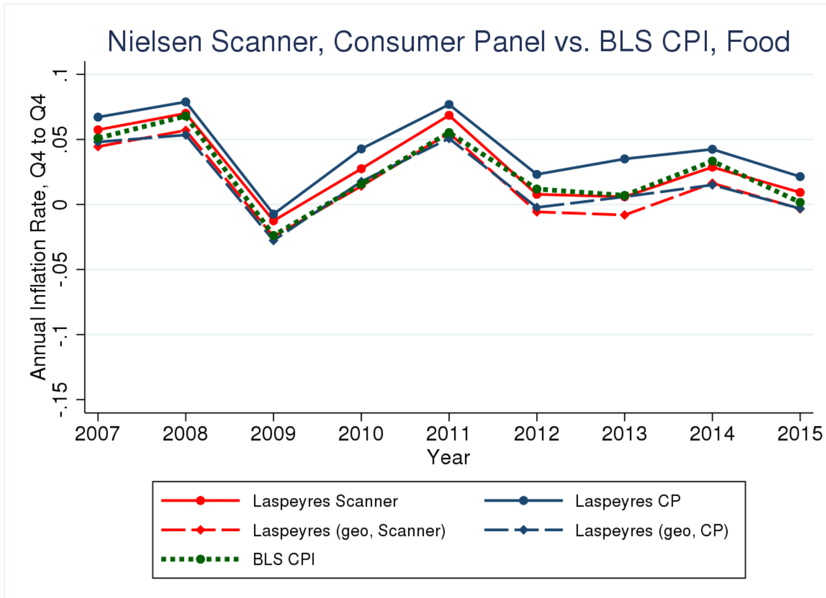
(b) Nonfood



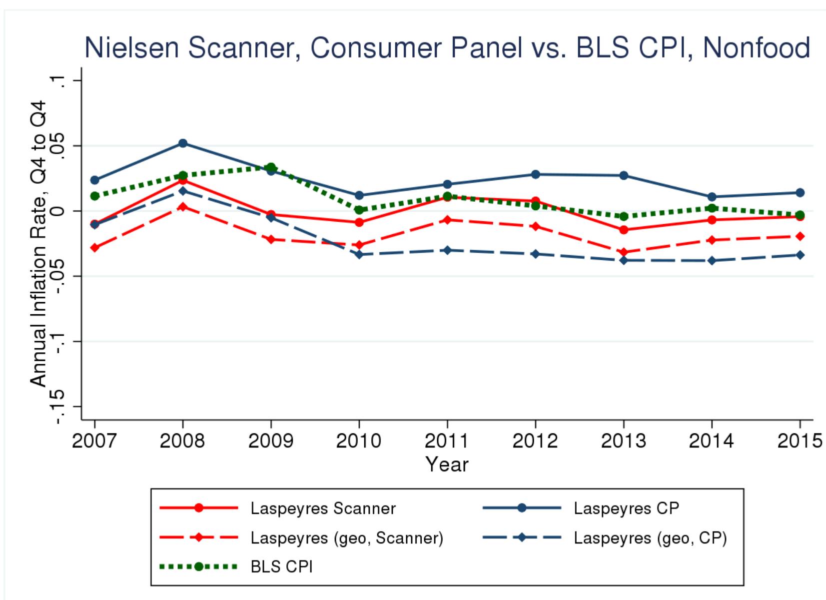
Notes: Figures uses Nielsen Scanner and Consumer Panel data for Food and Nonfood (aggregated) product groups. PCE is personal consumption expenditures (nominal) from BEA. All series indexed to 1 in 2008:q1.

Figure D.8: BLS CPI vs Nielsen Laspeyres, Food and Nonfood

(a) Food

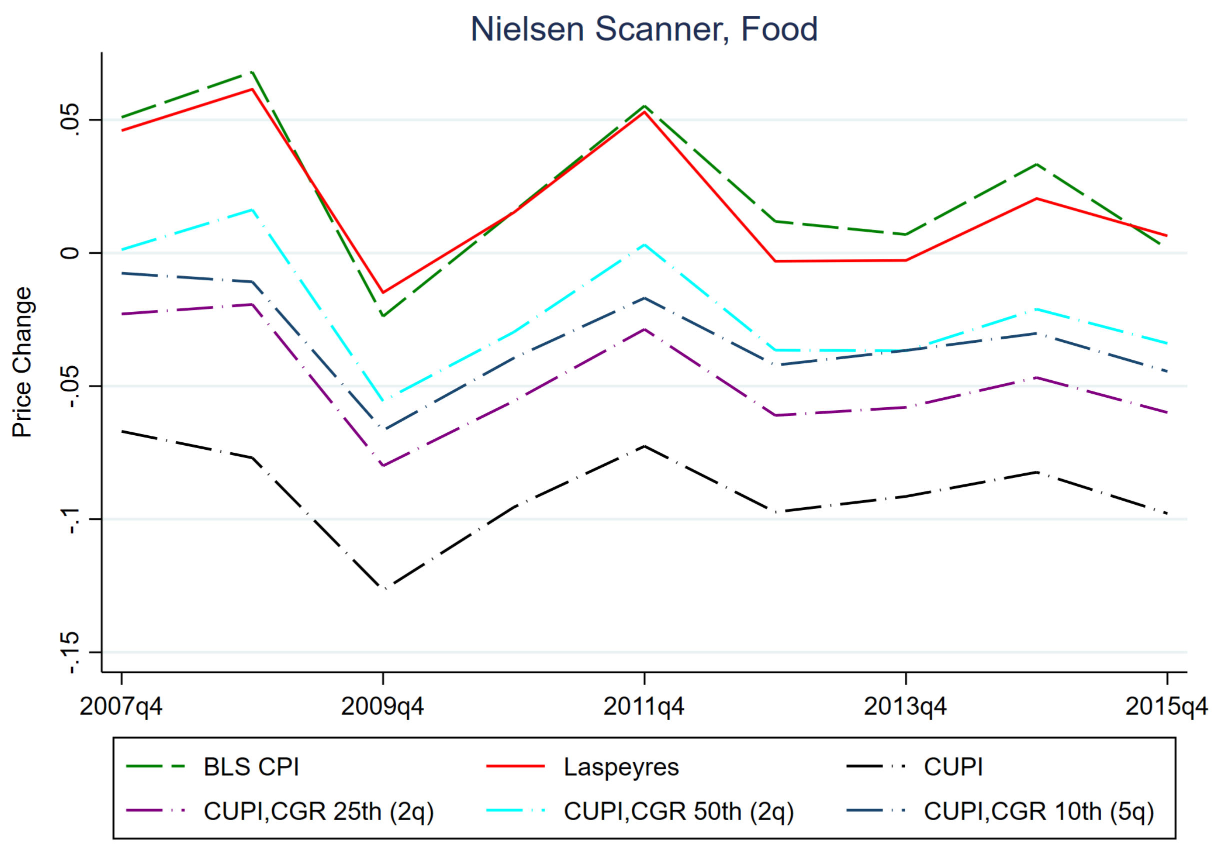


(b) Nonfood



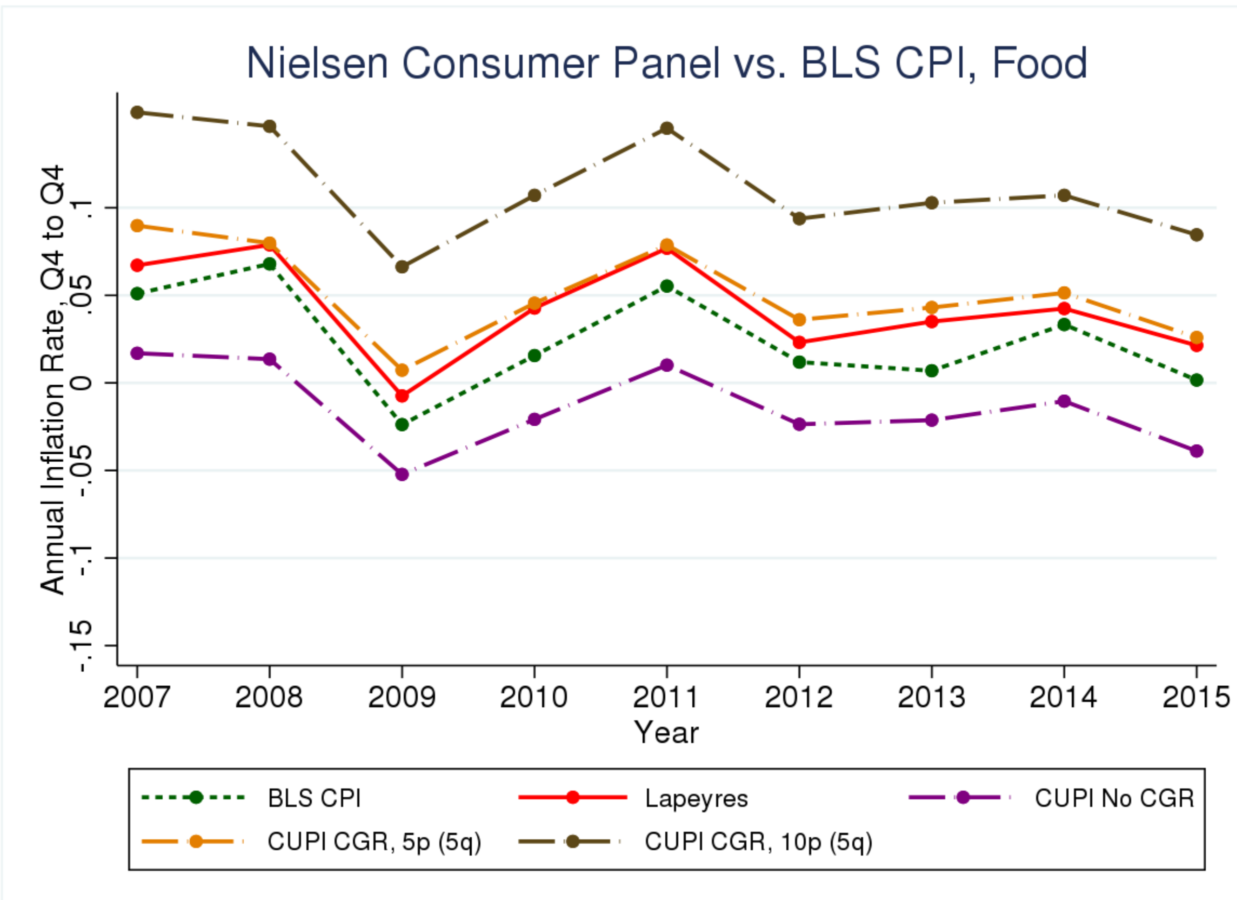
Notes: The figures show annual inflation from the fourth quarter of the previous year to the fourth quarter of the labeled year calculated from chained quarterly price indices. The panels use Nielsen Scanner and Consumer Panel data for Food and Nonfood (aggregated) product groups. The BLS CPI was computed by BLS staff.

Figure D.9: Common Goods Rules–2-quarter vs 5-quarter Horizons: Nielsen Food



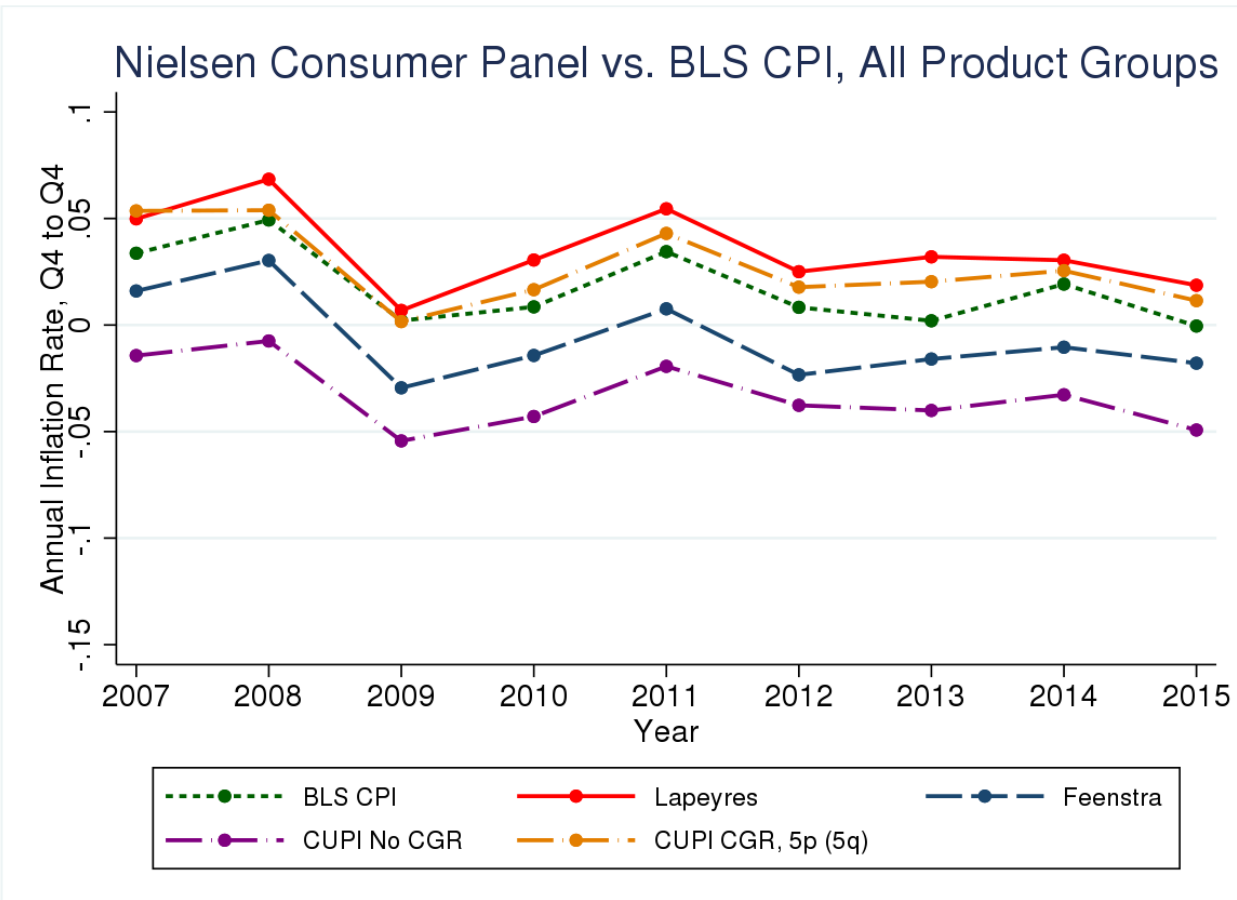
Notes: Figure uses Nielsen Scanner data for food. The $2q$ CUPI computes CGR percentile thresholds using sales pooled over a two quarter horizon (t and $t-1$). The $5q$ CUPI computes CGR percentile thresholds using sales pooled over a 5 quarter horizon (current and prior 4 quarters). Laspeyres is arithmetic.

Figure D.10: Common Goods Rules: Nielsen Food, Consumer Panel



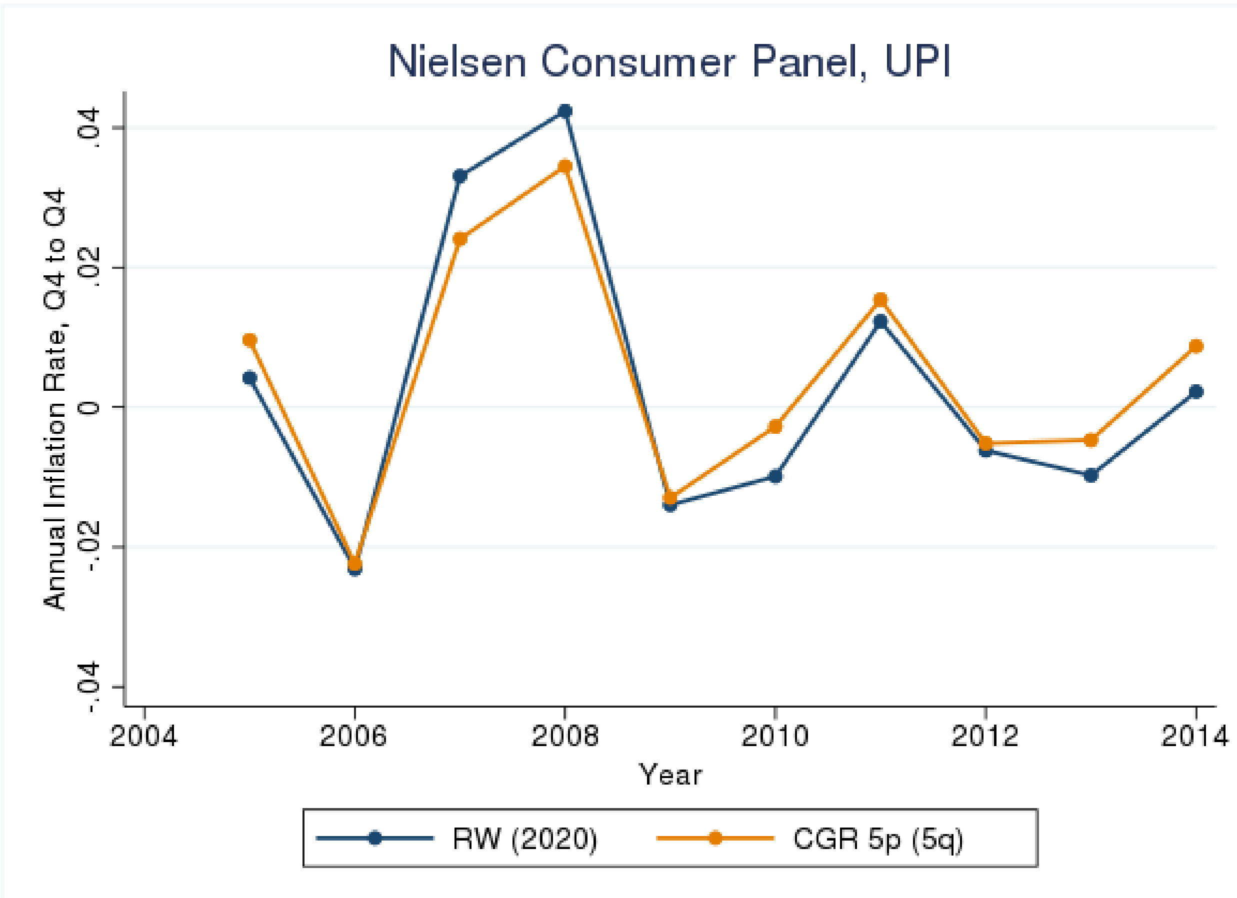
Notes: Figure uses Nielsen Consumer Panel data for food. The $5q$ CUPI computes CGR percentile thresholds using sales pooled over a five quarter horizon (t and $t-1$). (current and prior 4 quarters). Laspeyres is arithmetic.

Figure D.11: Common Goods Rules: Nielsen Consumer Panel



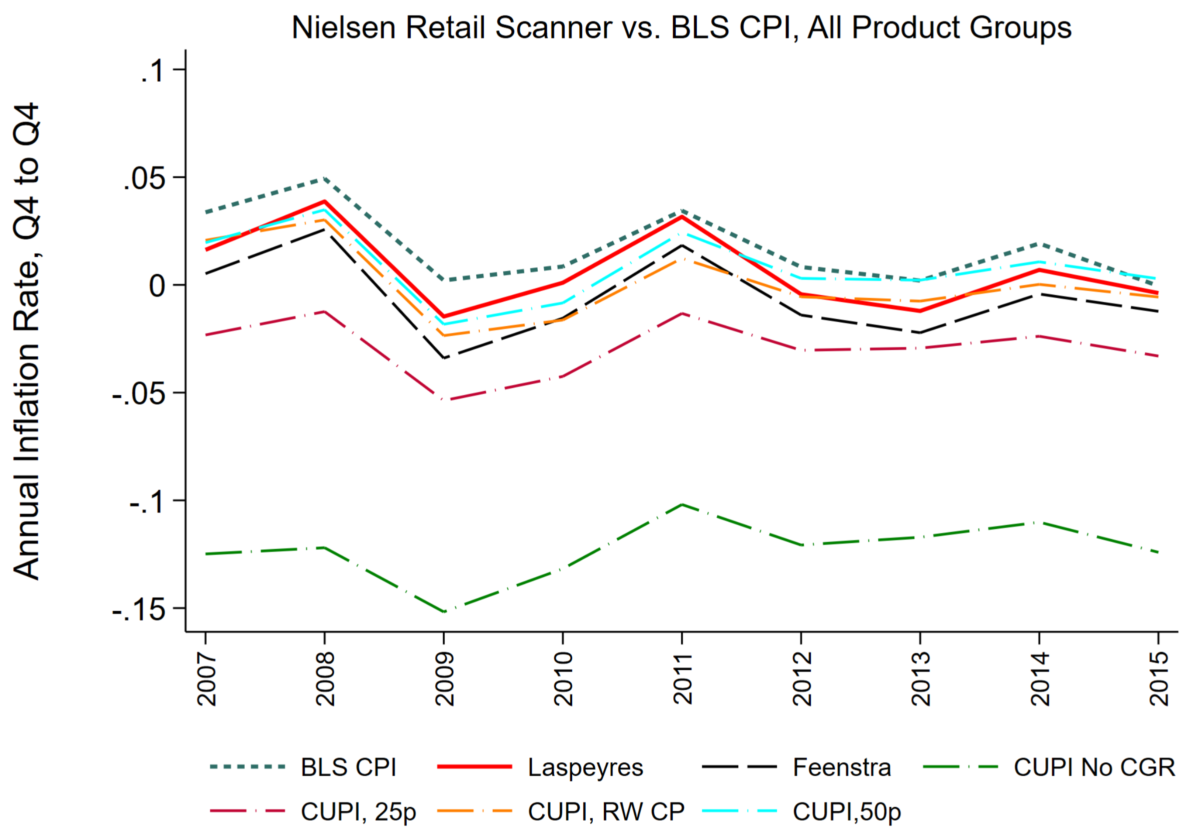
Notes: Figure uses Nielsen Consumer Panel data for food and nonfood product groups. The series "CUPI CGR RW" uses a 5th-percentile sales cutoff for the common goods rule. Percentile computed from sales pooled over 5 quarter horizon (current and prior 4 quarters). Laspeyres is arithmetic.

Figure D.12: Replication of Redding and Weinstein (2020): Nielsen Consumer Panel



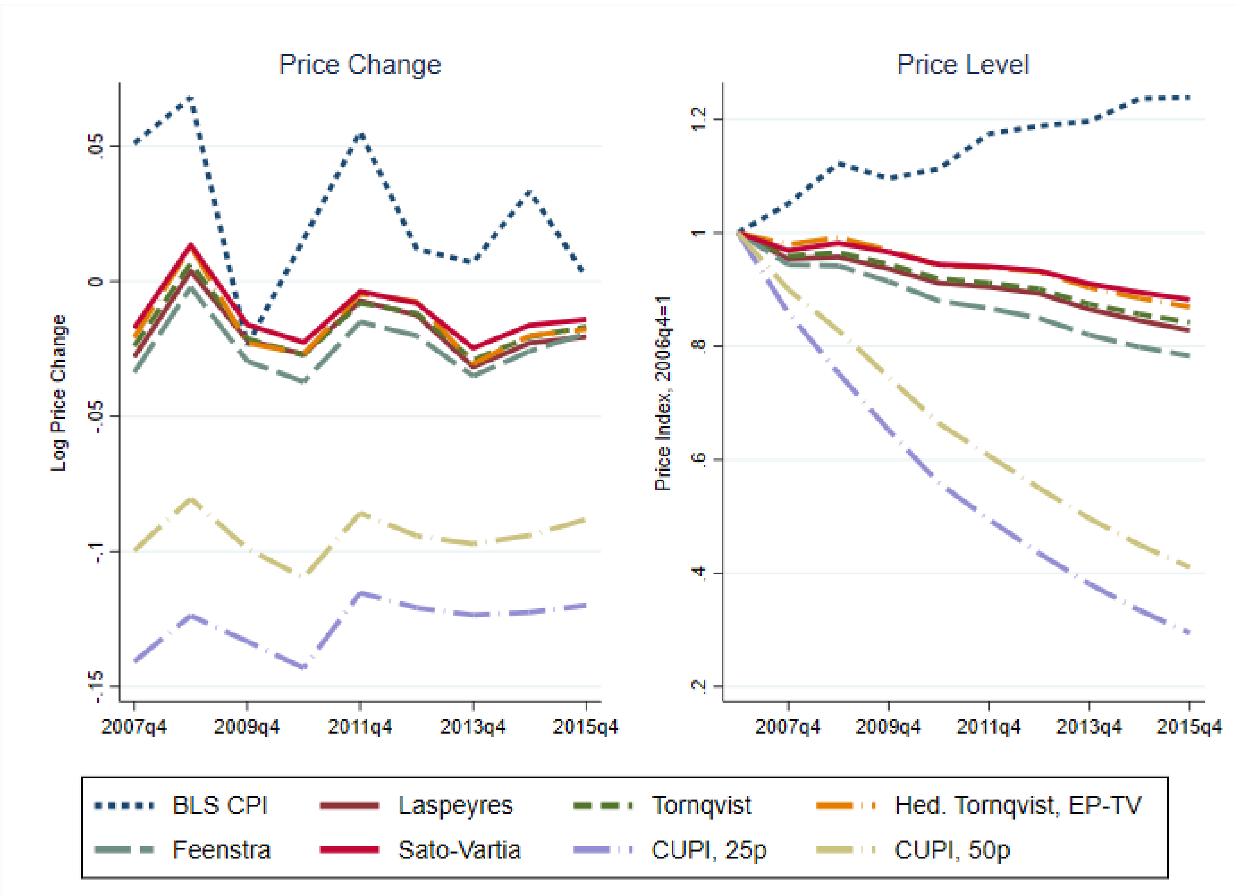
Notes: Figure uses Nielsen Consumer Panel for food and nonfood product groups. The indices are YoY for Q4. The series RW(2020) uses the same CGR duration rule as in Redding and Weinstein (2020). The series 5p(5q) use percentiles based on sales pooled over 5 quarter horizon (current and prior 4 quarters.

Figure D.13: Common Goods Rules: Nielsen Scanner Data, Food and Nonfood



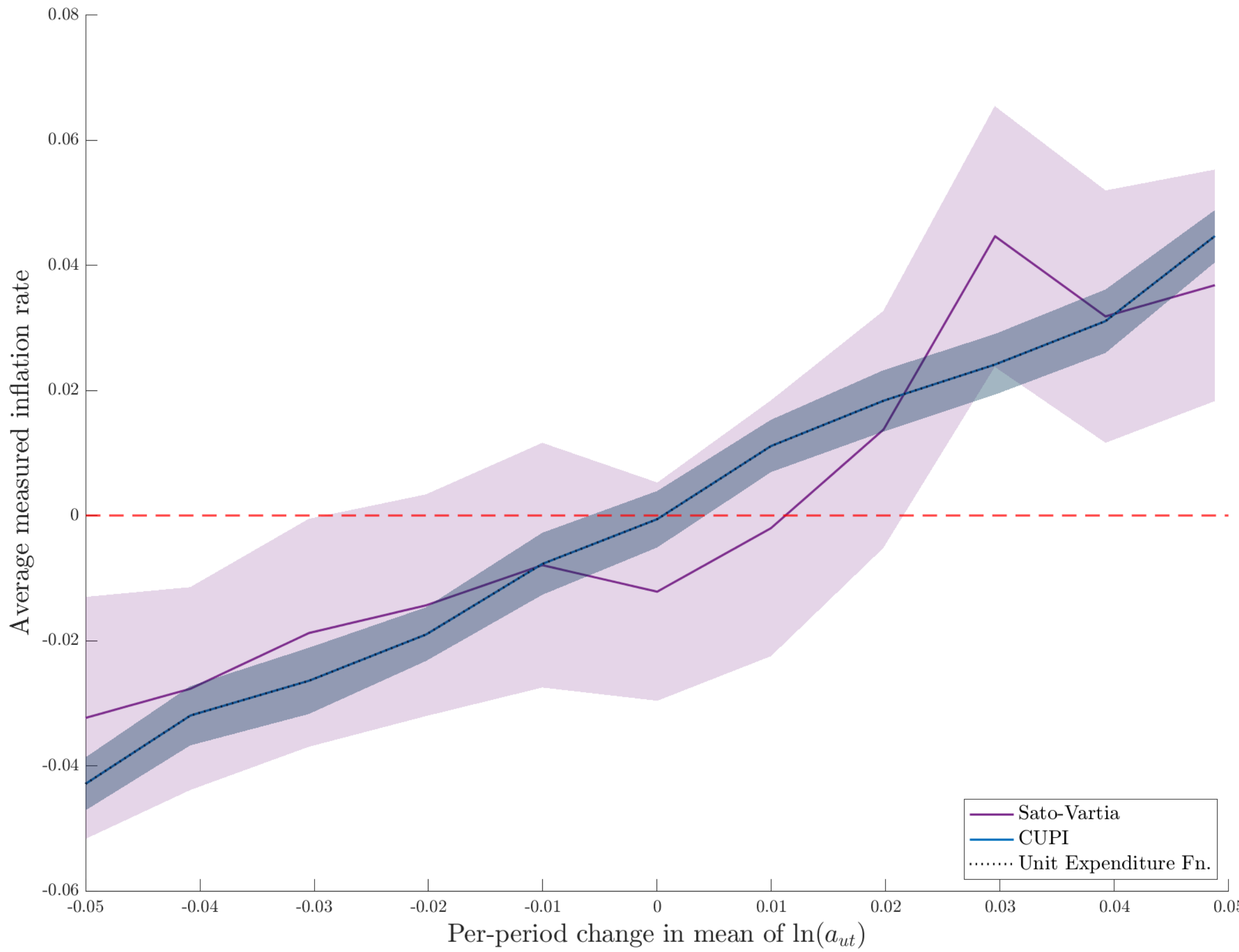
Notes: Figure uses Nielsen Retail Scanner data for food and nonfood product groups. The "CUPI, 25p" and "CUPI, 50p" series use 25th- and 50th-percentile cutoffs for the common goods rule, respectively. The series "CUPI, RW CP" uses the CGR 5th percentile threshold from the consumer Panel data for the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

Figure D.14: Main Price Index Specifications, Price Changes and Levels: Nielsen Nonfood



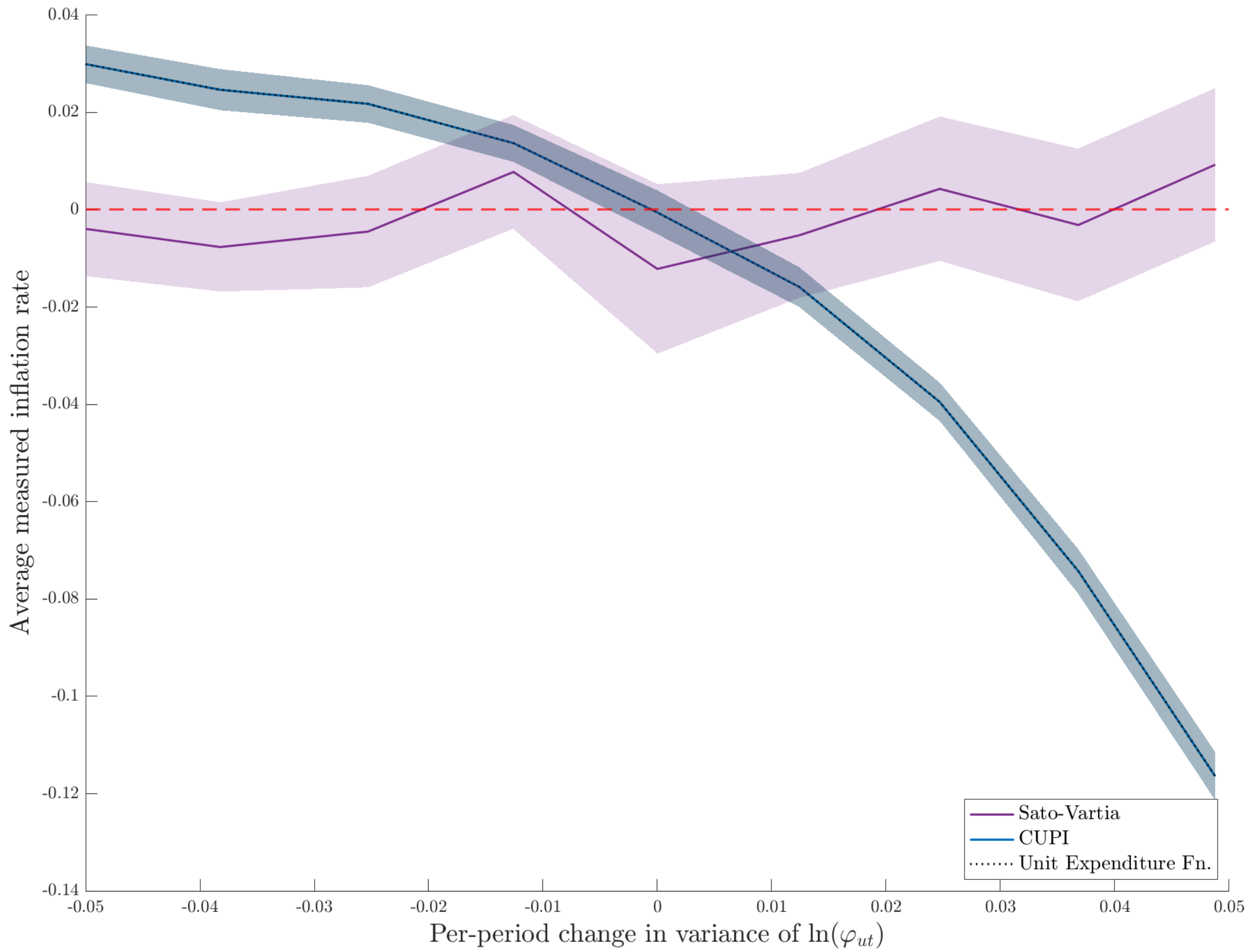
Notes: The figure shows Nielsen Retail Scanner data for nonfood product groups. Price changes show annual log differences from the fourth quarter of the previous year to the fourth quarter of the labeled year. The values are cumulative changes from chained quarterly indices. The price levels chained 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.

Figure D.15: Simulated CES Exact Price Indices with Trends in Cost Shifters



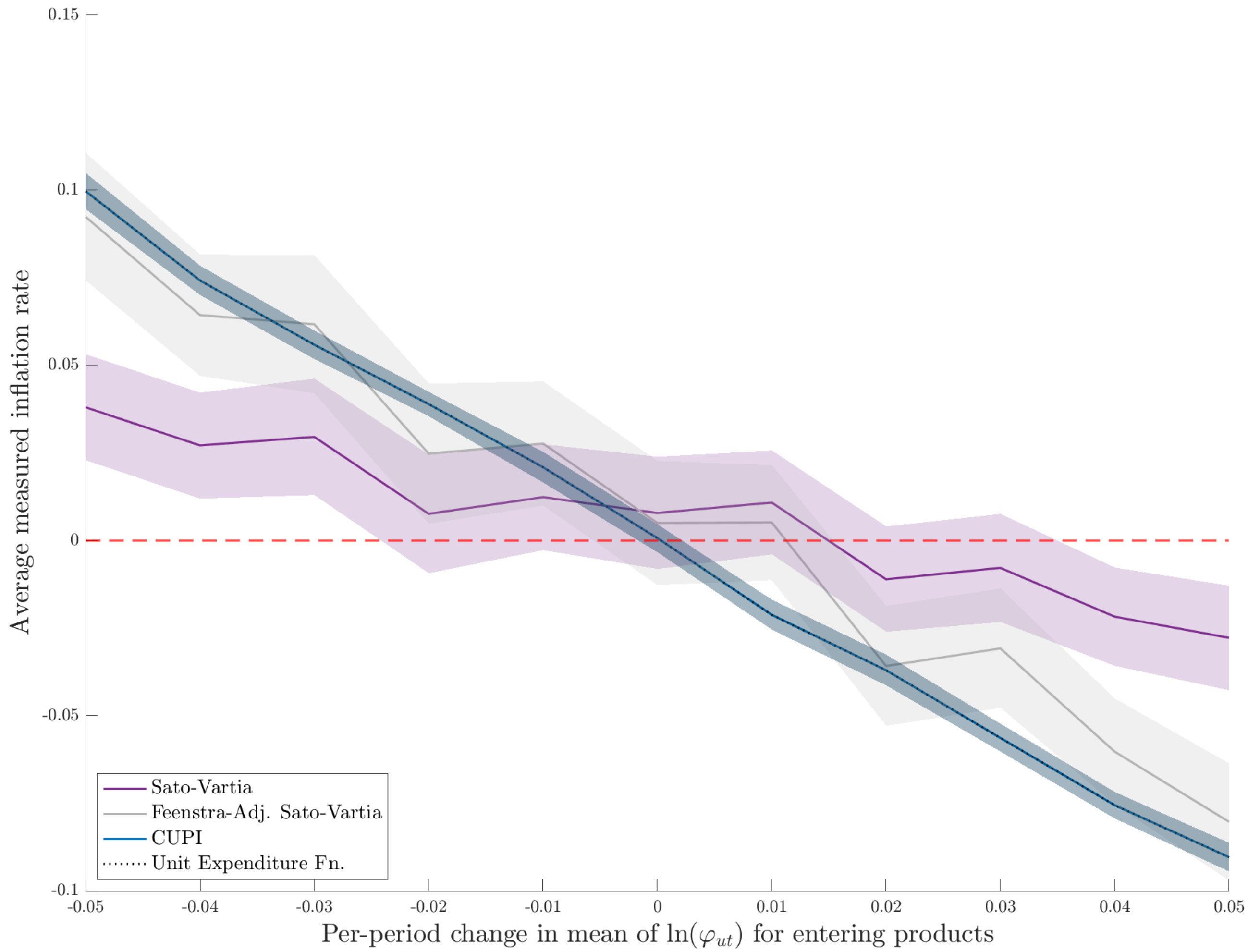
Notes: The figure displays inflation as measured by various CES exact price indices from Monte Carlo simulations of the general equilibrium environment of Hottman et al. (2016). See Section C.3 for simulation details. The horizontal axis displays different average growth rates for the products' marginal cost shifters. The vertical axis displays the average log inflation rate across simulation periods; solid lines represent simple averages across simulations and shaded regions represent 95-percent asymptotic confidence intervals. The CUPI coincides exactly with the unit expenditure function in these simulations.

Figure D.16: Simulated CES Exact Price Indices with Trends in Dispersion of Product Appeal



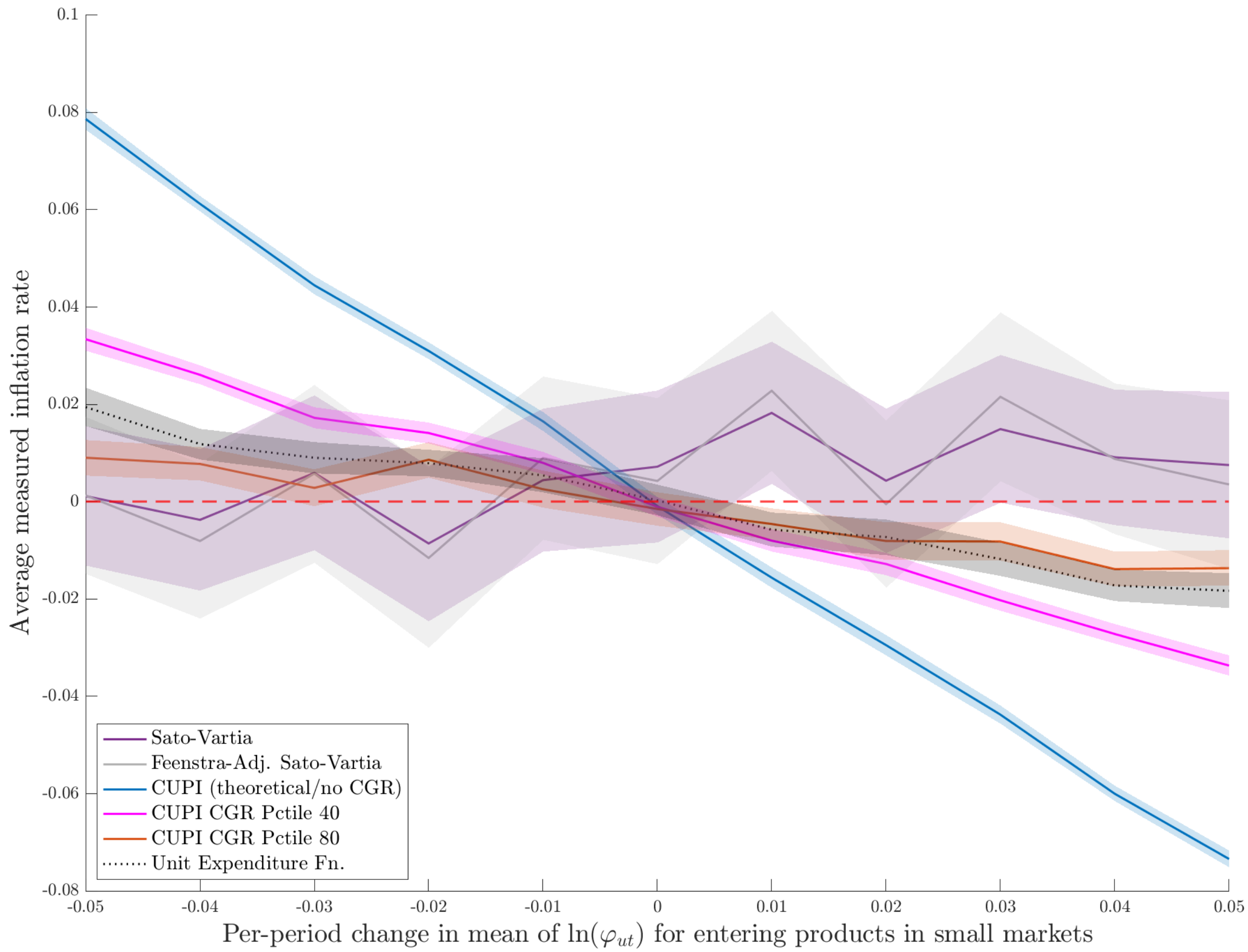
Notes: The figure displays inflation as measured by various CES exact price indices from Monte Carlo simulations of the general equilibrium environment of Hottman et al. (2016). See Section C.3 for simulation details. The horizontal axis displays different average growth rates for the variance of the product appeal parameters. The vertical axis displays the average log inflation rate across simulation periods; solid lines represent simple averages across simulations and shaded regions represent 95-percent asymptotic confidence intervals. The CUPI coincides exactly with the unit expenditure function in these simulations.

Figure D.17: Simulated CES Exact Price Indices with Product Turnover



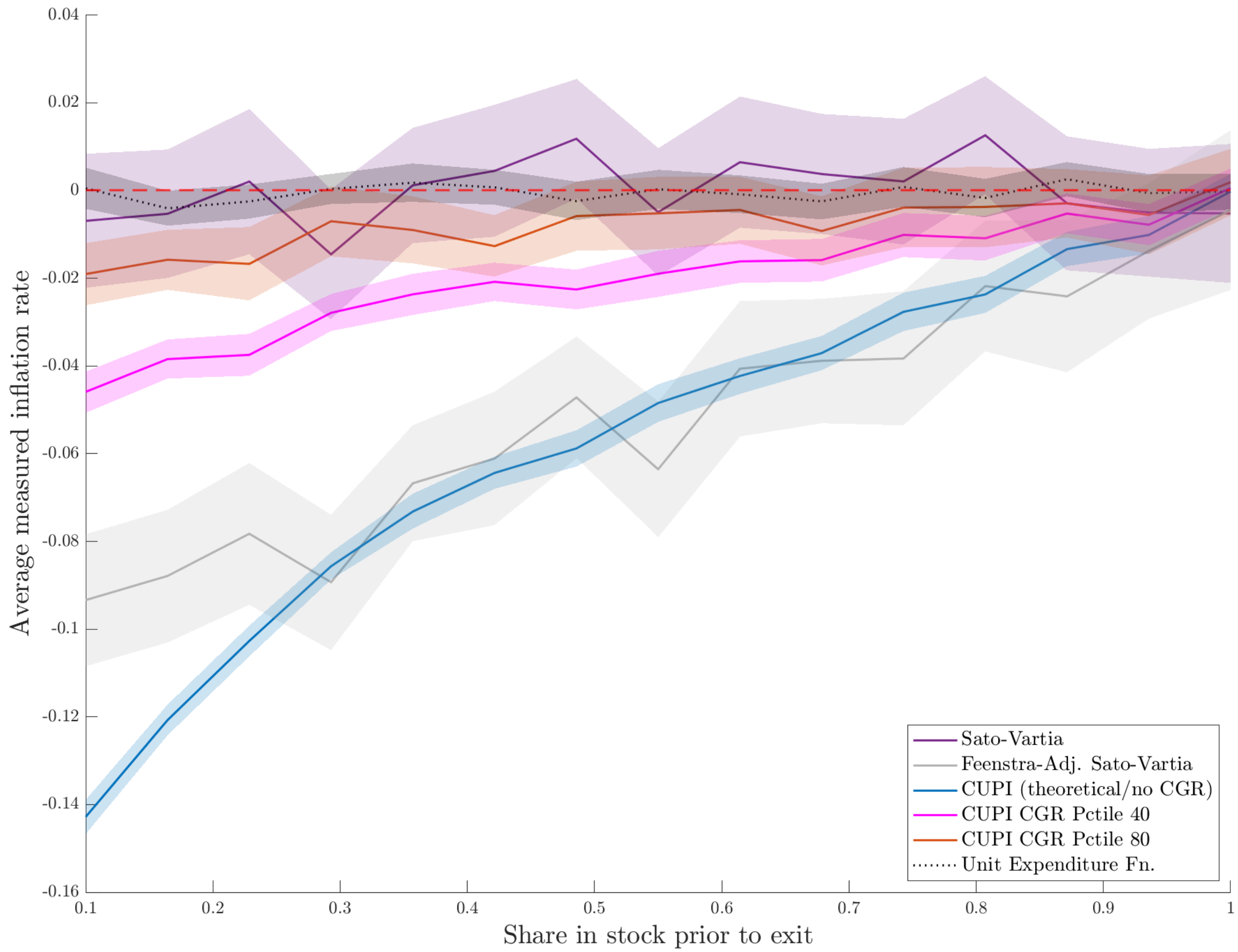
Notes: The figure displays inflation as measured by various CES exact price indices from Monte Carlo simulations of the general equilibrium environment of Hottman et al. (2016). The simulations feature product turnover, with equal numbers of products entering and exiting the market each period. Each product spends five periods in the market. See Section C.3 for simulation details. The horizontal axis displays different average growth rates for the product appeal parameters of entering products. The vertical axis displays the average log inflation rate across simulation periods; solid lines represent simple averages across simulations and shaded regions represent 95-percent asymptotic confidence intervals. The CUPI coincides exactly with the unit expenditure function in these simulations.

Figure D.18: Simulated CES Exact Price Indices with Segmented Markets



Notes: The figure displays inflation as measured by various CES exact price indices from Monte Carlo simulations of the general equilibrium environment of Hottman et al. (2016). The simulations feature segmented markets, with one large "national" market and four small "local" markets. See Section C.3 for simulation details. The horizontal axis displays different average growth rates for the product appeal parameters of entering products in the small markets. The vertical axis displays the average log inflation rate across simulation periods; solid lines represent simple averages across simulations and shaded regions represent 95-percent asymptotic confidence intervals.

Figure D.19: Simulated CES Exact Price Indices with Partial Stock-outs prior to Exit



Notes: The figure displays inflation as measured by various CES exact price indices from Monte Carlo simulations of the general equilibrium environment of Hottman et al. (2016). The simulations feature partial stock-outs in the period prior to products' exit. See Section C.3 for simulation details. The horizontal axis displays the share of the desired quantities available for purchase in the period prior to exit. The vertical axis displays the average log inflation rate across simulation periods; solid lines represent simple averages across simulations and shaded regions represent 95-percent asymptotic confidence intervals.