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

July 5, 2022

Abstract

We explore alternative methods for adjusting price indices for quality change at scale. The methods we consider can be used with large-scale item-level transactions data that has been digitized with price, quantity, and item-attribute data. We consider hedonic methods that take into account the changing valuation of both observable and unobservable characteristics in the presence of product turnover. We also consider demand-based approaches that take into account changing product quality from product turnover and changing product appeal of continuing products. Using these methods, we find evidence of substantial quality-adjustment in prices for a wide range of goods, including 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; Olivares: University of Maryland; Pardue: University of Maryland; Shapiro: University of Michigan and NBER; Zhao: University of Maryland and Bank of Canada. Laura Zhao worked on this project when she was a doctoral student at the University of Maryland. 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, Marshall Reinsdorf and participants at multiple seminars and conferences for helpful comments, The results here are in part based on researchers own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. We also use the NPD data housed at the U.S. Census Bureau. All results using the NPD data have been reviewed to ensure that no confidential information has been disclosed (CBDRB-FY19-122, CBDRB-FY21-074.). 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 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., 2021). While ambitious, such re-engineering has many potential advantages. One is to address the issue of rapid product turnover that is associated with quality improvements. We find quarterly product-level entry and exit rates that range between 5 and 15 percent across product groups. Current statistical agency procedures for measuring prices inadequately address such turnover. This paper considers scalable procedures using item-level transactions data that can be used to measure quality change, and therefore account for entering and exiting goods as well as changing consumer valuations of product attributes.

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 accounts for product quality changes that are unobserved to the econometrician. Higher-frequency, item-level transactions data with prices, quantities, and attributes greatly facilitate the implementation of the EP methodology. These data permit implementing the hedonics approach with superlative price indices (such as the Fisher or Tornqvist) in real time using internally consistent expenditure weights. We compare and contrast the EP methodology with more 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 CES Unified Price Index (CUPI) developed in Redding and Weinstein (2020), which adjusts 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 do not require attribute data beyond a basic product taxonomy to implement. Empirically implementing these methods at scale introduces a number of challenges, however, including classifying the goods in a manner so that the CES assumptions are valid. We have found that, in practice, the attribute data proves helpful in addressing these challenges.

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 of Redding and Weinstein (2020) 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 of characteristics along with the changing mix of characteristics given product entry and exit.

We implement the hedonic and demand-based approaches at scale using item-level transactions data from two major sources. The first platform we use is from NPD, which covers a wide range of general merchandise goods from bricks and mortar and online retailers. In this paper, we construct data 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 for Marketing 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.[1] Two challenges in the Nielsen data are that, first, it contains sparse data on product attributes, and second, it contains far more product groups (over 100) than the select group we used in the NPD data (five). We overcome those challenges by adapting the EP methodology to incorporate machine learning (ML) techniques that can relate product prices to seemingly limited characteristic data.

Consistent with the literature using scanner data, we find enormous product turnover at a quarterly frequency, 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. Currently, the Bureau of Labor Statistics (BLS) uses hedonic adjustment for only about 7 percent of products in the CPI. Our results suggest potentially substantial gains from implementing hedonic methods at scale using the EP method with item-level transactions data.

Among the demand-based methods, we find that the Feenstra (1994) index, which adjusts the Sato-Vartia for product turnover, systematically yields lower price inflation than the Sato-Vartia. This result suggests that product turnover is associated with quality improvement. The most general demand-based index we consider is the CUPI, which generalizes the Feenstra index to allow for changing product appeal over product life cycles. We find the

[1] We focus in this paper on product groups classified as "food" in the Nielsen data to sidestep potential concerns about the representativeness of the nonfood product groups.

CUPI implies substantial quality adjustment beyond what the Feenstra index implies.

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 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.[2] 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 a common goods rule.

We proceed as follows. Section 2 describes the underlying data. The conceptual framework we use for hedonic and demand-based indices is presented in section 3. Section 4 presents our main results and discusses the advantages and drawbacks of the alternative approaches that we have considered in light of those results. Section 6 provides concluding remarks.

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 data and focus on quarterly price indices. This approach is motivated by our objective to compare traditional, hedonic, and demand-based price indices in a manner consistent with the recent

[2]Redding and Weinstein (2020) refer to these rules "as alternative definitions of common varieties." In his insightful discussion of Ehrlich et al. (2021), Robert Feenstra motivated a common goods rule as a necessary recognition that it takes time for goods to enter and exit markets, a concept he denoted *seasoning*. As we discuss below, one limitation of the Redding and Weinstein (2020) specification of the common goods rule is that it requires forward-looking information and therefore cannot be implemented in real time. We find we can mimic their results using only backward-looking and real-time information.

literature, particularly the CUPI developed by Redding and Weinstein (2020).

2.1 NPD Data

We use proprietary data that the NPD Group provided to the U.S. Census Bureau, which consists of monthly sales and quantity data at the product-store level from 2014 through 2018.[3] 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.

An attractive feature of the NPD data is that they 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, which facilitate estimation of hedonic models. 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

[3]Month definitions follow the National Retail Federation (NRF) calendar. 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.

of the five groups exhibits a high degree of product turnover. In unreported results, we find that the turnover rates are lower on a sales-weighted basis, though, suggesting that turnover is more common among goods with smaller market shares.

Figure 1 presents life-cycle dynamics of product market shares and prices within these product groups. The illustrated statistics are mean log differences from the product-specific initial values upon entry. Prices decline steadily after entry, while market shares exhibit a hump-shaped pattern post-entry. 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 consists of over 2.6 million products identified by the finest level of aggregation–12-digit universal product codes (UPCs) that uniquely identify specific goods.[4] The RMS 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,

[4]The Nielsen data contain both UPC codes and "UPC version codes." The unique product identifier used in the analysis is the combination of the UPC code and the UPC version code.

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. The product groups are classified into food and nonfood sectors based on a concordance provided by the BLS.

The RMS 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.

We focus on results for the Nielsen data's food product groups in the main text 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. 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.

As we discuss in Appendix B, we have also compared patterns of total expenditures for harmonized categories from Nielsen and Personal Consumption Expenditures data (PCE). We find evidence suggesting that the Nielsen Retail Scanner data's coverage of nonfood items deteriorated during our study period, potentially driven by the ongoing structural shifts in Retail Trade, especially towards e-commerce. In contrast, we find a close correspondence between total expenditure trend patterns from Nielsen and the PCE for harmonized categories of food items. This correspondence holds for individual food groups as well as aggregated food categories.

3 Conceptual Framework

The goal of any price index is to measure approximately or exactly the change in the cost of living between two time periods—that is, to calculate how much more or less expensive it is to achieve the same standard of living as in some base period given current prices. 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 from the item-level data potentially incorporate quality change from product turnover and changing product appeal over products or valuations of product characteristics.

3.1 Traditional Price Indices

Traditional price indices typically can be conceptualized as weighted-average price changes of some form. Our empirical work in this paper will focus on so-called geometric price indices, which are weighted averages of log price changes. Specifically the log geometric price index, $\ln \Psi_t^G$, is given by:

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

where w_{kt} is a weight assigned to product k (typically based on the product's market share) and the ratio of prices to be aggregated is often called a log price relative. The set $\mathbb{C}_{t\text{-}1,t}$ is the set of all "continuing" or "common" goods that are sold both in period t and in period t-1. The use of different 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 the superlative Tornqvist index uses average expenditure shares ($w_{kt} = \frac{s_{kt\text{-}1}+s_{kt}}{2}$).[5] Hence, the Tornqvist lies between the geometric Paasche and geometric Laspeyres, and 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.

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

[5] The Tornqvist is closely related to the Fisher superlative index, which is the geometric mean of the arithmetic Laspeyres and Paasche indices. 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. 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.

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.[6] So-called 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). We will generally use a superlative index, particularly the Tornqvist, when comparing traditional indices with hedonic or demand-based indices.

In theory, these traditional 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. In current practice, however, statistical agencies' data on sales and expenditure shares is often limited to disparate sources at higher levels of aggregation and lower frequency. For instance, BLS uses expenditure shares from the Consumer Expenditure survey, with infrequently updated weights, to produce the Consumer Price Index (CPI). 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. High-frequency scanner data connect the prices and quantities sold for each product, allowing for the construction of superlative price indices using internally consistent price and quantity data. We explore this advantage in our empirical analysis.[7]

These traditional price indices are all "matched-model" indices: they calculate price changes across the goods that were sold both in the base and in the current period. The traditional indices therefore do not account directly for goods that enter or exit across periods, which may be an important source of changing product quality.

[6] 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.

[7] The limitations of the current system are discussed further in Ehrlich et al. (2021).

3.2 Hedonic Price Indices

In this section, we describe our use of hedonic methods to adjust price indices for quality changes, especially in the context of product turnover. 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:

$$\ln p_{kt} = h_t(Z_k) + \eta_{kt}, \tag{1}$$

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 (1) 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 utility 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). 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.

A core limitation of the log-level hedonic estimation approach outlined in equation (1) 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. One simple, though effective, method is to estimate hedonic models of price changes rather than price levels, by estimating the equation:

$$\Delta \ln p_{kt} = Z_k' \beta_t + v_{kt}. \tag{2}$$

This log-difference hedonic model estimates the change in hedonic price coefficients directly, which "differences out" any unobservable item-level characteristics that have a fixed influence on prices over time. This basic log-difference hedonic model will not account for the influence of time-varying unobservable characteristics, however.

The most general form of the Erickson and Pakes (2011) approach can account for time-varying unobservable characteristics, so we will call this approach the "TV approach" for short. Implementing the TV approach requires two steps. First, we estimate the log-level hedonic specification in equation (1) for period t-1. Second, we estimate a log-difference hedonic model, including the lagged residuals from the first stage. The second estimating equation is then:[8]

$$\Delta \ln p_{kt} = Z_k' \beta_t + \kappa \hat{\eta}_{kt\text{-}1} + v_{kt}. \tag{3}$$

Including the initial residuals from equation (1) in equation (3) 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, log first-difference, and TV approaches.

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 estimate hedonic regression equations pooling observations from the adjacent periods t-1 and t. Letting T denote the total number of periods in the data, we thus estimate $T-1$ separate pooled two-period regressions of the form:

$$\ln p_{k\tau} = \alpha_{t\text{-}1,t} + \delta_t + Z_k' \gamma_{t\text{-}1,t} + \epsilon_{k\tau}, \quad \tau = \{t\text{-}1, t\}, \tag{4}$$

[8] 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 (3).

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]

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. 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 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 which is often rejected by the data.

Our implementations of the 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.

The product descriptions in the Nielsen data provided by the Kilts Center for Marketing at the University of Chicago are generally not coded to be human-intelligible. For instance, two product descriptions for soft drinks are `ZR DT LN/LM CF NBP CT` and `NATURAL R CL NB 12P`, while a product description for toilet paper is `DR W 1P 308S TT 6PK`. A human analyst could decipher portions of these descriptions: `DT` means "diet," `12P` means twelve pack, `1P` means one ply, `308S` means 308 sheets, etc. It would not be feasible for human analysts to encode such data at scale, however, and simple dictionaries would be fooled (e.g., the `P`-suffix

[9] We specify the hedonic regression equation (4) 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.

means "pack" for soft drinks and "ply" for toilet paper).

An additional challenge in the Nielsen data is its sheer scale. The Retail Scanner data contains more than 100 product groups and over 1,000 product modules. It would be difficult for human analysts to specify sensible hedonic regression equations for so many product groups. It would be even more difficult to update those regression equations over time as product mixes and characteristics change.

To address these challenges, we have implemented deep neural networks to predict product prices and price changes from the product descriptions in the Nielsen Kilts Center data. Our approach parallels the TV approach of Erickson and Pakes (2011), in that it first predicts price levels and then, to capture time-varying unobservable effects, uses the prediction error in a second-stage neural net predicting price changes. We provide more details of our approach in the appendix and in a companion paper (Cafarella et al., 2021) that focuses on the machine learning methodology. In related work, Bajari et al. (2021) use an advanced machine learning approach that includes encoding image data as inputs into price predictions.[10]

To make our alternative hedonic approaches as comparable as possible, we use weighted estimation methods in all cases. We follow the recent time dummy literature by using Tornqvist expenditure weights, so that the time dummy method yields a quality-adjusted Tornqvist price index. We apply quantity-share weights for the estimation of the hedonic pricing functions in the hedonic imputation approaches using both econometric and ML methods. This approach follows Bajari et al. (2021), who also use quantity-share weights in their implementation of hedonic price indices using item-level transactions data. Using quantity-share weights focuses the hedonic estimation procedure on accurately mapping the relationship between prices and characteristics for the market basket of goods purchased by the consumer.[11] Note that we use the expenditure-share weights in the construction of

[10]Bajari et al. (2021) provides novel methodology for encoding images via machine learning but do not incorporate the TV approach to constructing hedonic price indices.

[11]See for example the discussion in De Haan (2008). We provide further discussion of the motivation for using weighted results in the appendix. We also show that the first-difference methods we focus on are

the price indices themselves; the quantity-share weights are used only in estimation of the hedonic relationships in equations (1)–(3) and their machine-learning analogues.

We focus on full-imputation versions of the hedonic imputation indices, which use predicted prices for all observations, including for common goods. In other words, we use the predicted price relatives of continuing goods for the *missing* price relatives for entering and exiting goods.[12] Pakes (2003) shows that this form of hedonic imputation index provides a bound to the exact change in the cost of living under a weaker set of assumptions than those commonly used in the literature. 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).[13] Using full-imputation indices also facilitates comparison with the time dummy method, as highlighted by De Haan (2008) and Diewert et al. (2008).[14] In addition, Erickson and Pakes (2011) observe that single- and double-imputation indices are subject to a form of selection bias, because they treat the hedonic estimation residuals for continuing, entering, and exiting goods in an asymmetric manner.[15] Full-imputation indices have also been used in Benkard and Bajari (2005), Diewert et al. (2008) and Bajari et al.

largely robust to using weighted or unweighted specifications. In unreported results, we have explored using expenditure weights and have also found similar results.

[12]For the implementation of the TV approach, we assume the lagged residual for an entering good in the period prior to entry is zero. 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. As a robustness check, we consider the difference between the traditional and hedonic Laspeyres using the 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. In addition, in unreported results we find very similar results if we replace the residual for entering goods based on a first-stage level regression using a current period residual.

[13]For example, Hill and Melser (2006) show that the full-imputation hedonic Tornqvist index estimated with a semi-log model has a dual representation as the Fisher index in characteristics space.

[14]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 in Erickson and Pakes (2011). Incorporating the 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 TV approach of Erickson and Pakes (2011) in our hedonic indices.

[15]See footnote 3 of Erickson and Pakes (2011).

(2021). Our implementation of hedonic indices builds on and integrates the insights of this literature.

The use of the relationship between price relatives for continuing goods to impute the price relatives for entering and exiting goods is justified by the argument that this imputation relies on the relationship between price relatives and characteristics. Importantly, characteristics turnover is distinct from product turnover. New characteristics arguably diffuse slowly through the entry of new goods and characteristics disappear from the available bundle slowly through product exit. Relatedly, new goods often have *more* of an important characteristic (e.g., size and speed of memory cards in one of the product groups studied below), 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.

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. The CES utility function yields a tractable demand system with several computable price indexes that correspond exactly to the theoretical unit cost function faced by a representative consumer in the presence of product turnover and time-varying product appeal. We explore the Sato-Vartia index (Sato, 1976; Vartia, 1976), the Feenstra-Adjusted Sato-Vartia index (Feenstra, 1995), which we will call the "Feenstra index," and the CES Unified Price Index of Redding and Weinstein (2020), which we will call the "CUPI." We focus on CES demand systems as this structure has been developed to provide tractable, implementable price indices that can account for quality change and product turnover.

We start with the CUPI, as it nests the other two indices as special cases. Redding and

Weinstein characterize the CES unit expenditure function as:

$$P_t = \left[\sum_{k \in \Omega_t} \left(\frac{p_{kt}}{\varphi_{kt}} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}, \tag{5}$$

where $\sigma > 1$ is the consumer's elasticity of substitution between products and Ω_t is the set of products sold in period t. φ_{kt} is a product-level appeal parameter that varies over time. Redding and Weinstein (2020) emphasize that including time-varying product appeal is essential to make the CES system consistent with the observed micro variation in prices and quantities. 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.

Consumers' optimally chosen expenditure shares in this system are given as:

$$s_{kt} \equiv \frac{p_{kt} c_{kt}}{\sum_l p_{lt} c_{lt}} = \frac{(p_{kt}/\varphi_{kt})^{1-\sigma}}{\sum_{l \in \Omega_t} (p_{lt}/\varphi_{lt})^{1-\sigma}} = \frac{(p_{kt}/\varphi_{kt})^{1-\sigma}}{P_t^{1-\sigma}}, \tag{6}$$

where c_{kt} is the quantity of good k purchased in period t.

Redding and Weinstein (2020) derive the exact-price index in this setting (the CUPI) as:

$$\Psi_{t-1,t}^{CUPI} = \left(\frac{\lambda_{t,t-1}}{\lambda_{t-1,t}} \right)^{\frac{1}{\sigma-1}} \frac{\tilde{P}_t^*}{\tilde{P}_{t-1}^*} \left(\frac{\tilde{S}_t^*}{\tilde{S}_{t-1}^*} \right)^{\frac{1}{\sigma-1}}. \tag{7}$$

The first term in the CUPI is the Feenstra (1994) adjustment factor for product turnover, with elements defined as:

$$\lambda_{t,t-1} = \frac{\sum_{k \in \mathbb{C}_t} p_{kt} c_{kt}}{\sum_{k \in \Omega_t} p_{kt} c_{kt}}, \qquad \lambda_{t-1,t} = \frac{\sum_{k \in \mathbb{C}_t} p_{kt-1} c_{kt-1}}{\sum_{k \in \Omega_{t-1}} p_{kt-1} c_{kt-1}}, \tag{8}$$

where $\mathbb{C}_t$ is the set of common goods (what Redding and Weinstein (2020) denote as "common varieties").[16] Denoting the sales-weighted product entry and exit rates as $ER_{t\text{-}1,t}$ and $XR_{t\text{-}1,t}$, 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 term thus indicates a downward adjustment to traditional matched price indices 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.[17]

The second term in the CUPI is the traditional Jevons index, defined over the set of common goods. We follow Redding and Weinstein (2020) in denoting the geometric mean of a variable x as $\tilde{x}$ and denoting the geometric mean over the set of common goods with an asterisk, so $\tilde{P}_t^*$ denotes the geometric mean of prices across common goods in period t and $\tilde{P}_{t\text{-}1}^*$ represents the same object in period t-1.[18] The ratio of the two is the Jevons index, which is an unweighted index.

We refer to the third term in the CUPI as the "S^* ratio," with elements defined as the unweighted geometric average expenditure shares on common varieties in periods t-1 and t:

$$\tilde{S}_t^* = \left(\prod_{k\in\mathbb{C}_t} s_{kt}\right)^{\frac{1}{N_{\mathbb{C}_t}}}, \qquad \tilde{S}_{t\text{-}1}^* = \left(\prod_{k\in\mathbb{C}_t} s_{kt\text{-}1}\right)^{\frac{1}{N_{\mathbb{C}_t}}}, \tag{10}$$

where we have denoted the number of common goods (i.e., products sold both in period t and in period t-1) as $N_{\mathbb{C}_t}$. This third term is novel to the CUPI and reflects, amongst other things, changes in dispersion of product appeal shocks for common goods over time. We discuss the factors underlying the contribution of the S^* ratio both conceptually and in practice below.

[16] The simplest definition of this set is products that were sold in both periods t-1 and t, but it is possible to re-define the set so that only goods that are sold in quantities above a threshold or which are sold in the market for a suitably long duration are included in this set, as in Redding and Weinstein (2020).

[17] We use the actual Feenstra term and not the approximation in our implementation.

[18] Using the notation in equation (10), we would write:

$$\tilde{P}_t^* = \left(\prod_{k\in\mathbb{C}_t} p_{kt}\right)^{\frac{1}{N_{\mathbb{C}_t}}}, \qquad \tilde{P}_{t\text{-}1}^* = \left(\prod_{k\in\mathbb{C}_t} s_{kt\text{-}1}\right)^{\frac{1}{N_{\mathbb{C}_t}}}. \tag{9}$$

It is instructive to consider the log version of the CUPI, which is given by:

$$\ln \Phi_{t-1,t}^{CUPI} = \frac{1}{\sigma - 1} \ln \left(\frac{\lambda_{t,t-1}}{\lambda_{t-1,t}} \right) + \frac{1}{N_{\mathbb{C}_t}} \sum_{k \in \mathbb{C}_t} \ln \left(\frac{p_{kt}^*}{p_{kt-1}^*} \right) + \frac{1}{\sigma - 1} \frac{1}{N_{\mathbb{C}_t}} \sum_{k \in \mathbb{C}_t} \ln \left(\frac{s_{kt}^*}{s_{kt-1}^*} \right). \quad (11)$$

Equation (11) clarifies that two of the CUPI's three terms (the Jevons index and the S^* ratio) are unweighted geometric means. As discussed below, this property is important for the CUPI's empirical implementation.

In the absence of time-varying product appeal, the CUPI collapses to the Feenstra (1994) index:

$$\Phi_{t\text{-}1,t}^{Feenstra} = \left(\frac{\lambda_{t,t-1}}{\lambda_{t\text{-}1,t}} \right)^{\frac{1}{\sigma-1}} \left(\Phi_{t\text{-}1,t}^{SV} \right), \quad (12)$$

where $\Phi_{t\text{-}1,t}^{SV}$ is the Sato-Vartia price index defined over common varieties. With no product turnover, the Feenstra index further collapses to the Sato-Vartia index (Sato, 1976; Vartia, 1976), defined as:

$$\ln \left(\Phi_{t\text{-}1,t}^{SV} \right) = \sum_{k \in \mathbb{C}_t} \omega_{kt} \ln \left(\frac{p_{kt}}{p_{kt\text{-}1}} \right), \quad \omega_{kt} = \frac{s_{kt} - s_{kt\text{-}1}}{\ln(s_{kt}) - \ln(s_{kt\text{-}1})} \Bigg/ \left(\sum_{k \in \mathbb{C}_t} \frac{s_{kt} - s_{kt\text{-}1}}{\ln(s_{kt}) - \ln(s_{kt\text{-}1})} \right). \quad (13)$$

Each of these price indexes 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.[19] 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 these generalizations of the Sato-Vartia index are empirically relevant.

Although the CUPI is the most general of CES exact price indices that we consider, its inclusion of two unweighted geometric mean terms contrasts with the Sato-Vartia and Feenstra indices, which include only expenditure-weighted terms. The CUPI's unweighted

[19] Feenstra and Reinsdorf (2007) show Sato-Vartia is unbiased in expectation with randomness in tastes under restricted conditions. Appendix C.1 contains a further related discussion of this topic.

terms are sensitive to products with very small expenditure shares. The CUPI can therefore feature large measured price changes from what would appear to be economically minor products.

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 Jevons index and S^* ratio terms are calculated.[20] 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. They 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). A limitation of this particular 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.

The need to consider common goods rules and more generally the calculation of terms involving unweighted geometric means is due to the CUPI's incorporation of the relative product appeal shocks. As Redding and Weinstein (2020) emphasize, a core motivation for including these shocks is that they enable the reconciliation of micro variation in prices and expenditure shares with price index measurement. It is instructive to explore the bias from neglecting such relative product appeal shocks in terms of mean inflation rates.

[20] That is, the goods included in the set $\mathbb{C}_t$.

We argue in Appendix C.1 that independently and identically distributed relative product appeal shocks alone are not sufficient to generate an expected bias in the Sato-Vartia index's measure of inflation. Rather, the index's expected bias in terms of measured inflation arises either from systematic variation in the dispersion of relative product appeal shocks or the correlation between appeal shocks and prices. Specifically, we find that in the absence of product turnover, the Sato-Vartia is upward-biased if there is rising product appeal dispersion or an increasingly positive correlation between product appeal shocks and prices over time; the Sato-Vartia index is downward-biased in the opposite cases. The practical relevance of these conceptual considerations is an empirical question. As we will see, the S^* ratio is empirically relevant, but its effect may reflect unresolved specification issues for the CGR or other issues.

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, with details reported in Appendix A.1. In the main text, we focus on the Erickson and Pakes (2011) method, using fixed unobservables (first-difference specification) and using first differences along with time-varying unobservables (the TV approach). We find that the average quarterly R-squared for log price changes is the highest using the TV approach (see Table D.1) for each product group, with values ranging from 0.13 for memory cards to 0.47 for headphones and boys' jeans. The first-difference approach without time-varying unobservables has an average R-squared ranging from 0.09 for memory cards to 0.43 for headphones and boys' jeans, which is higher than the other alternatives considered but lower than the TV approach. A noteworthy pattern in Table D.1 is the large gap between the average R-squared values for log price changes and for log price levels, the latter of which range from 0.62 to 0.72. Predicting price changes is inherently a much more difficult task than predicting price levels, because price levels reflect cross-sectional differences in product characteristics, while price changes reflect changes in the mapping between prices and characteristics over time.

Figure 2 presents the results for alternative hedonic price indices for the five NPD product groups and compares these indices to the traditional Tornqvist index. As noted, for hedonics we focus on the Hedonic Tornqvist using fixed unobservables, and the Hedonic Tornqvist with time varying unobservables (the TV approach). For purposes of comparison we also include the Hedonic Time Dummy results. The values displayed in the figure are annual percent changes in the 4th quarter of each year from chained cumulative quarterly indices.[21] All of the price indices track each other closely, but there are systematic differences in the patterns. For all product groups, the TV approach yields the lowest 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 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 does not yield systematic quality-adjustment differences relative to the traditional Tornqvist index. The time dummy method suggests a notable quality adjustment for coffee makers, 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

[21] We explore the potential role of chain drift below.

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.[22] Additional limitations of the time dummy approach have been highlighted by Pakes (2003) and Diewert et al. (2008).[23]

The findings in Figure 2 also do not exhibit a systematic relationship between the Tornqvist and the hedonic Tornqvist accounting for fixed unobservable characteristics (i.e., using a log first-difference hedonic model). Given that we find a systematically lower hedonic Tornqvist when accounting for time-varying unobservable characteristics, the somewhat erratic pattern of the first-difference specification with fixed unobservables suggests it important to permit time-varying valuations of in order to systematically adjust for unobservable product characteristics.

Figure 3 provides further evidence on the efficacy of the 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 broadly 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. However, with the same data,

[22]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 time dummy method still yields a negative price change in that year, but not as negative as the standard Tornqvist. The hedonic Tornqvist TV method yields a more negative price index than the standard Tornqvist.

[23]Pakes (2003) raises questions about the bound implied by the time dummy method. Diewert et al. (2008) highlight that the time dummy method requires more restrictive assumptions than the other hedonic approaches.

they find their methodologies to account for unobservable product characteristics (both using fixed valuation of unobservables and time-varying unobservables) yield systematically lower estimated inflation than the traditional Tornqvist index. They also conduct a test of their methodology similar to the one we present in Figure 3. They also find their time-varying unobservables methodology is robust to leaving out important characteristics but standard (e.g., level hedonic estimations) specifications are not robust to this exercise.

4.1.2 CES Demand-Based Price Indices

We turn now to CES demand-based price indices. For the Feenstra (1994) price index and the CUPI, implementation requires estimates of the elasticities of substitution. 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.[24] 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 Feenstra index comprises the "Lambda Ratio" (Feenstra adjustment term) and the Sato-Vartia index, while the CUPI includes the identical Lambda Ratio, the "P^* ratio" (or Jevons index), and the S^* ratio. 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 (7). 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% declines in the price level annually, and it is often 10%–30% below the Feenstra Index.

[24] 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.

The large differences between the Feenstra index and the CUPI in these product groups arise from two sources. The first source is the difference between P^* ratio (Jevons index) and the Sato-Vartia index. The Sato-Vartia is a weighted average log price change among common goods, and the P^* ratio is an unweighted average. In boys' jeans, for instance, the CUPI P^* ratio is far below the Sato-Vartia. The difference between the weighted and unweighted log price ratios for common goods suggests there are a large number of low-share goods experiencing price declines that are driving down the CUPI. The second source is the introduction of the S^* ratio in the CUPI, intended to account for changing consumer tastes. Almost everywhere, 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 (hereafter often denoted a *CGR*) to the index. The logic, as discussed above, is that it takes some time for goods to break into the market as well to exit from the market. A limitation of the forward-looking duration-based approach for the CGR in Redding and Weinstein (2020) is that it cannot be implemented in real time. Their approach requires information about goods' future presence in or absence from the marketplace. 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.[25] 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). For our main NPD analysis, we consider alternative market share percentile thresholds.[26]

Figure 5 illustrates the CUPI's sensitivity to the CGR for different market share thresh-

[25]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.

[26]As we discuss below, 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 CGR implemented over a longer horizon for our NPD analysis.

olds. Specifically, we consider market share thresholds for continuing goods in t and t-1 of the 10th percentile, the 30th percentile and the 50th percentile. We depict the CUPI for these different CGR rules alongside the Feenstra index and the CUPI without a common goods rule. 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 relative comparisons and shifting it to the entry/exit adjustment term (the Lambda ratio). 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 patterns in the ballpark of the Feenstra index. For products groups such as headphones and boys' jeans, however, the CUPI shows noticeably lower inflation than the Feenstra index even using with a 50th-percentile CGR threshold (i.e., excluding half of products from the set of common goods).

These findings have a number of important implications. First, the CUPI is sensitive to the specific definition of the CGR, in a manner that varies across product groups. 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). Many factors may underlie these patterns. For instance, in boys' jeans, the seasonal product turnover cycle in apparel is arguably at work. Late in any apparel product turnover cycle, bargain racks are often available with low prices but limited supply.

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 is a step in that direction. We think 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. Further research, motivated by theory, is necessary to provide guidance about product-group specific common goods rules. We provide further analysis and discussion of these points below.

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.[27]

We explore this issue by exploiting the detailed product attributes in the data to define a nested product substitution structure using two methods. First, we define nests within product groups with a heuristic-based approach. With this method, we assign products to subgroups based on a set of key variables that we as analysts hypothesize define market strata. As 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 decile 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

[27] 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, which is also exact for CES preferences under stricter assumptions, is consistent in aggregation.

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.[28] The between-nest estimator of the elasticity of substitution uses an instrumental variable (IV) approach building on Hottman et al. (2016).[29]

Table 2 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 6 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.[30] The alternative nesting approaches yield similar results, with the nested CUPI tending to show slightly less deflation than the un-nested (or "flat") CUPI. In unreported results, we find that the relationship between the nested and flat CUPIs is robust to using alternative CGR cutoffs.

[28]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.

[29]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).

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

4.1.3 Comparing Traditional, Hedonic, and Exact Price Indices

Figure 7 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 8 displays price indices without the CUPIs but with the addition of the Laspeyres index. Price indices in these product groups all follow a roughly similar pattern of relative orders: the Laspeyres index is the highest, the Tornqvist, Sato-Vartia, and Feenstra are in the next group, the hedonic Tornqvist using the TV method is systematically lower, and the CUPI (both baseline and nested by product characteristics) is the lowest, especially for headphones and boys' jeans. The Feenstra index is systematically lower than the Sato-Vartia index, consistent with the quality adjustment for product turnover lowering the estimated rate of inflation. 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 substantial gap between the 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 between the Laspeyres and the Tornqvist indices 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.

Figures 9 and D.3 in Appendix D present plots of chained price index levels calculated by chaining the quarterly price indices underlying Figures 7 and 8. Figures 9 and D.3 thus provide perspective on the cumulative effects of the differences between the various indices. 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 traditional Laspeyres is substantially higher than the traditional Tornqvist except for headphones where they are essentially the same. The hedonic Laspeyres or Tornqvist using the TV approach yields systematically larger cumulative declines in prices than the traditional Laspeyres or Tornqvist. The range of differences between the traditional and hedonic indices varies across product groups, with larger differences for memory cards and headphones. The Feenstra index yields systematically larger cumulative declines than the Sato-Vartia index, but the differences are smaller than the differences between the traditional and hedonic Tornqvist indices. The CUPI (with a 30th percentile CGR) yields substantially larger cumulative price declines for headphones and boys' jeans, while for memory cards and coffee makers, the CUPI yields similar declines to the hedonic Tornqvist. For occupational footwear, the cumulative declines from the CUPI are larger than from the other indices, but the gap is relatively modest.

Taking stock of the results from the NPD data, the most robust methodology yielding systematic quality adjustment is the hedonic Tornqvist using the TV method. The Feenstra index is also a useful point of comparison given its systematic relationship with the Sato-Vartia index. The CUPI is the most general demand-based index, but is sensitive to the specification of the CGR. More research is needed to provide guidance about how to specify the CGR on a product group specific basis.

4.2 Nielsen Results

We restrict our analysis of results for the Nielsen scanner data to the food product groups in the main text. Our empirical implementation in the Nielsen data largely follows our strategy in the NPD data for the CES exact price indices. 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 TV approach, we combine the insights of the Erickson and Pakes (2011) with the machine learning approach summarized, which Appendix B.3 describes in more detail.[31] The machine learning approach allows us to exploit the Neilsen scanner data's unstructured information on item-level attributes. We find that the hedonic TV approach using machine learning yields a median R-squared for price relatives across product group-quarters of about 0.50 within-sample and about 0.20 out-of-sample.

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.[32]

Figure 10 shows the results for the aggregated food categories of the CUPI and its components using various CGRs defined by different sales-based percentiles. The S^* ratio is especially sensitive to the CGR in the Nielsen data. Recall that the S^* ratio is an unweighted geometric mean, which is sensitive to small market shares. The S^* ratio's sensitivity to alternative CGR thresholds leads directly to sensitivity in the CUPI. The baseline CUPI without a CGR percentile threshold has average *quarterly* 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 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 to 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

[31] Our companion paper, Cafarella et al. (2021) provides a full description.

[32] 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.

(2020).[33] Appendix B.2 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.[34]

Figure 11 presents a full set of price indices for the Nielsen scanner food product groups, in change and level forms.[35] The panels of the figure include the BLS CPI, computed for the same Nielsen product groups.[36] We find that the CPI and the traditional Laspeyres index track each other closely in Nielsen's food product groups, with a discrepancy arises towards the end of the sample period. The Tornqvist and Sato-Vartia indices are lower than the Laspeyres, and the quality-adjusted indices (Feenstra, hedonic Tornqvist using the 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 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. 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.[37]

[33] 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.

[34] Appendix B.2 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).

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

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

[37] Results for the nonfood product groups, described in Appendix B.1, 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.

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 (or even single-store) transactions data (e.g., De Haan and Van Der Grient, 2011). Our analysis uses national data at a quarterly frequency, which mitigates this issue. However, given our focus on comparing alternative approaches for computing price indices, in this section we consider whether GEKS-type indices (Gini, 1931; Eltetö and Köves, 1964; Sculz, 1964) preserve the implications of our core findings.

We follow Bajari et al. (2021) by computing a GEKS-type index (which we denote "GEKS-lite") based on the geometric mean of the chained quarterly indices for each year for the 4th quarter with the year-over-year (YoY) price indices for the 4th quarter.[38] Table 4 reports chained and GEKS-lite indices for the five NPD product groups and alternative indices. Results reported are average annual indices. More often than not, the GEKS-lite price change index is less negative than the chained price index for traditional price indices, but the differences are not large. For the hedonic indices, the GEKS-lite indices are actually more negative in three out of the five product groups. For the demand-based indices, the GEKS-lite indices are also typically less negative, but again these differences are modest quantitatively. The GEKS-lite CUPI is substantially less negative, but this difference also

[38] Given that we are including hedonic indices, the computational burden of implementing price indices over all possible horizons would be substantial.

reflects the effects of applying a commons good rule over a longer horizon.[39]

From our perspective, 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 yields higher inflation than 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.

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.[40] 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, but this result 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 for Food is higher using Laspeyres than Tornqvist, higher using Sato-Vartia than Feenstra, and higher using Feenstra than the CUPI.[41]

5 Taking Stock

Item-level transactions data with prices, quantities, and attributes offer considerable advantages for computing quality-adjusted price indices compared to the traditional methods the BLS currently uses to compute price indices. The current system draws on disparate data sources for price quotes and expenditure shares among continuing goods. Even traditional

[39]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 $t-1$.

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

[41]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.

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 measures systematically lower inflation than the Laspeyres, with the gap varying over time.

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.

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 not taking into account 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 imposed by 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 Figure D.4 in Appendix D, 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.[42]

On an unweighted basis, the distribution is very skewed to the left, with most item-

[42]Market penetration is defined as the share of Nielsen metro areas in which the item-level week is observed to have positive sales.

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.

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 terms, which mean that goods with small expenditure shares can have an outsize effect on the index.

To provide more perspective on the pros and cons of the CES demand-based price indices given these issues, in Appendix C, we examine the behavior of these CES indices indices analytically and in simulations under various assumptions about market structure. A few key results emerge from our examination.

First, we present an analytical argument in Appendix C.1 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.2. 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, when product appeal is becoming more (less) dispersed over time, the Sato-Vartia index will be biased upwards (downward), but the CUPI will remain unbiased. This bias will also apply to the Feenstra index. Thus, one possible explanation for the CUPI to measure uniformly lower inflation than the Sato-Vartia and Feenstra indices is that product appeal dispersion is rising over time. However, this finding on the CUPI is so ubiquitous that we suspect other forces are at work. Indeed, the CUPI becomes less of an outlier when a common goods rule is applied, and the motivation for such a rule stems largely the time it takes for goods to enter and exit the market. Such *seasoning* effects are outside the scope of the derivation of the CES. In our simulations, we explore environments that mimic the need for a common goods rule. We show that if goods enter in local markets instead of national markets, but prices are measured assuming a unified national market, then the CUPI will deviate from the true expenditure function. The CUPI will be downward biased if entering goods into local markets are of improving quality, and a common goods rule will help alleviate the bias. Relatedly but distinctly, we find that if exiting goods are not widely available (i.e., are rationed or subject to stockouts), then the CUPI will be downward biased, and a common goods rule will alleviate the bias. While these simulations provide promising intuitive justifications underlying a common goods rule, they highlight that the needed common goods rule will depend on product-group specific dynamics (e.g., the rate of increase of quality of entering goods, the extent of rationing) that are not easily observable.

Based both on our empirical findings and this numerical simulation analysis, we believe that the demand-based indices that incorporate quality adjustment (specifically the Feenstra and the CUPI) provide useful benchmarks that should be used for purposes of comparison with indices such as the hedonic Tornqvist using the TV approach. However, the national CES market assumption is too strong, especially for current implementations of the CUPI.[43]

[43] Relatedly, the representative household assumption is too strong. Heterogeneity across consumers is of interest for a variety of reasons, including quantifying differences in changes in the cost of living across

Future research could presumably make progress by developing a framework to distinguish between national and local goods and aggregate indices from local nests as appropriate. Likewise, progress could be made in better understanding the dynamics of product availability in the first periods after entry and before exit, so that common goods rules could be disciplined by the nature of this process. Developing more structure for the dynamics of product entry and exit should also provide guidance about any needed modifications in the Feenstra adjustment and $S^{\star}$ ratio terms in the CUPI in a richer dynamic environment. We think these topics should be high priorities for future research.

6 Concluding Remarks

Using item-level transactions data with price, quantity, and attribute information enables the production of quality-adjustment price indices at scale. This paper employs such an approach to present evidence that traditional matched-model methods overstate the rate of inflation and understate the rates of real expenditure and real output growth substantially. We find that these patterns are pervasive, that is, not limited to goods such as electronics where technological progress is most visible.

We have explored and evaluated two alternative approaches for quality adjustment at scale with item-level transactions data, hedonic methods and demand-based methods. For hedonics, we have found that it is critically important to use the methodology developed by Erickson and Pakes (2011) that takes into account time-varying changes in the valuation of unobservable characteristics of continuing, entering, and exiting goods. Using this methodology, we have found that traditional matched model indices overstate the rate of inflation for a wide range of goods.

We have focused on CES frameworks for demand-based indices, building in particular on the path-breaking work of Redding and Weinstein (2020). The CES unified price index they develop incorporates quality change from product turnover (consistent with Feenstra, 1994)

different groups.

but also time-varying product appeal for continuing goods. A challenge in implementing the CUPI is that, contrary to the sales patterns in the item-level data, it assumes a national market for each CES-based nest of goods. Current implementations of the CUPI address these limitations by imposing common goods rules. While this approach is promising, our results indicate that the construction and suitability of common goods rules that vary by product group is an open area requiring further research.

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. The current paper shows the promise for price measurement. A next step is to explore the promise of using the improved price index measurement with internally consistent measures of sales to improve measurement of real output growth. This topic is a high priority for future research.

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. Globalization and the gains from variery. *Quarterly journal of economics*, (May), 2006. URL https://academic.oup.com/qje/article-abstract/121/2/541/1884019.

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. Technical report, Census, Maryland and Michigan, 2021.

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. Hedonic Regressions: A Review of Some Unresolved Issues. 2002. URL https://www.researchgate.net/publication/228404555.

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*, 2021.

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. Exact Hedonic Price Indexes. 77(4):634–653, 1995.

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.

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.

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.

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.

M. Silver. The use of weights in hedonic regressions: the measurement of quality adjusted price changes. 2003.

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%

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	Estimator	Groups	Elasticity of Substitution			
			Within		Across	
Headphones	Feenstra	-	7.634	(0.748)		
	HRW plus	Manual	8.609	(0.544)	7.704	(0.491)
	HRW plus	Hedonic	9.537	(0.969)	8.958	(0.423)
Memory Cards	Feenstra	-	5.623	(0.484)		
	HRW plus	Manual	6.31	(0.675)	4.534	(0.298)
	HRW plus	Hedonic	6.621	(0.657)	5.25	(0.586)
Coffeemakers	Feenstra	-	5.183	(1.289)		
	HRW plus	Manual	5.495	(0.791)	3.42	(0.63)
	HRW plus	Hedonic	5.345	(0.99)	5.306	(0.374)
Occupational Footwear	Feenstra	-	7.31	(0.533)		
	HRW plus	Manual	5.545	(0.509)	3.057	(0.493)
	HRW plus	Hedonic	6.199	(0.548)	4.135	(0.769)
Boy's Jeans	Feenstra	-	7.861	(0.565)		
	HRW plus	Manual	7.439	(1.5)	3.234	(0.734)
	HRW plus	Hedonic	8.156	(1.82)	3.418	(0.657)

Estimated elasticities of substitution for CES and nested CES models. Standard errors in parentheses. Data come from NPD Group. "HRW Plus" is a modified nested CES estimation procedure from Hottman et al. (2016) that is robust to product entry and exit in the estimation of the between-group elasticity.

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	0.773	0.887
Hed. Laspeyres,TV	0.414	0.683	0.494	0.709	0.859
Tornqvist	0.467	0.688	0.607	0.726	0.872
Hed. Tornqvist,TV	0.406	0.667	0.541	0.687	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)	-10.09	-5.44	-4.00	-4.80	-2.15
Laspeyres (GL)	-10.48	-4.94	-7.98	-4.86	-1.78
Tornqvist(C)	-16.90	-8.86	-11.58	-7.63	-3.35
Tornqvist(GL)	-15.67	-6.67	-11.55	-5.57	-2.33
Hed.Tornqvist,TV(C)	-19.83	-9.56	-14.13	-8.93	-3.77
Hed.Tornqvist,TV(GL)	-21.64	-9.95	-15.64	-7.60	-3.33
Sato-Vartia(C)	-16.24	-8.24	-11.75	-6.20	-3.14
Sato-Vartia(GL)	-14.57	-6.38	-11.34	-4.13	-2.10
Feenstra(C)	-16.78	-8.92	-12.47	-6.92	-3.76
Feenstra(GL)	-16.61	-9.46	-13.10	-5.52	-3.80
CUPI,CGR 30p(C)	-20.64	-11.05	-24.08	-34.74	-6.08
CUPI,CGR 30p(GL)	-20.14	-9.43	-22.56	-26.99	-5.25

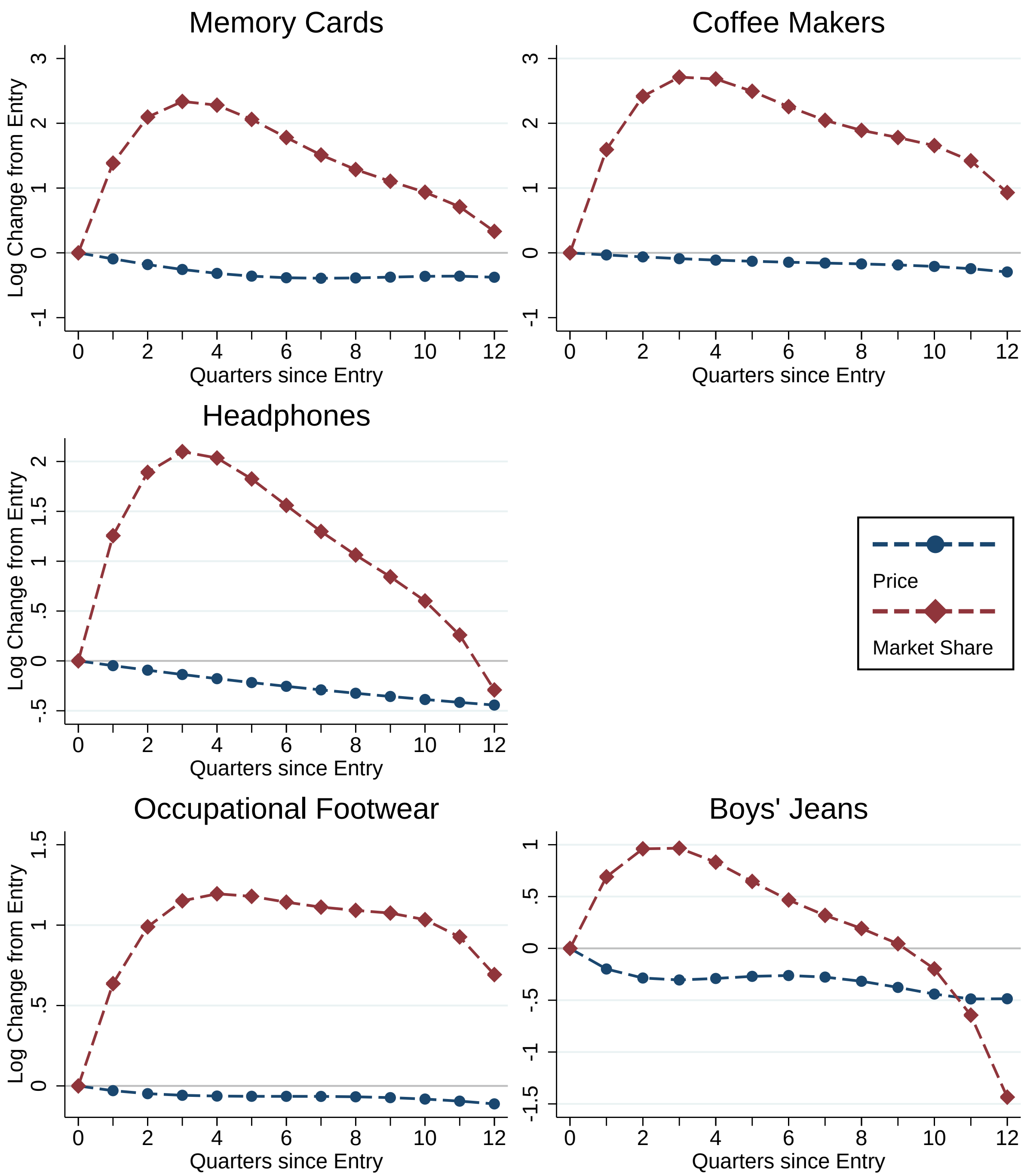
Notes: Chained values are averages of cumulative quarterly rates for year. GEKS-lite is the average of the geometric mean of the chained values and the YoY price indices for q4 for each year. Data come from the NPD Group.

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

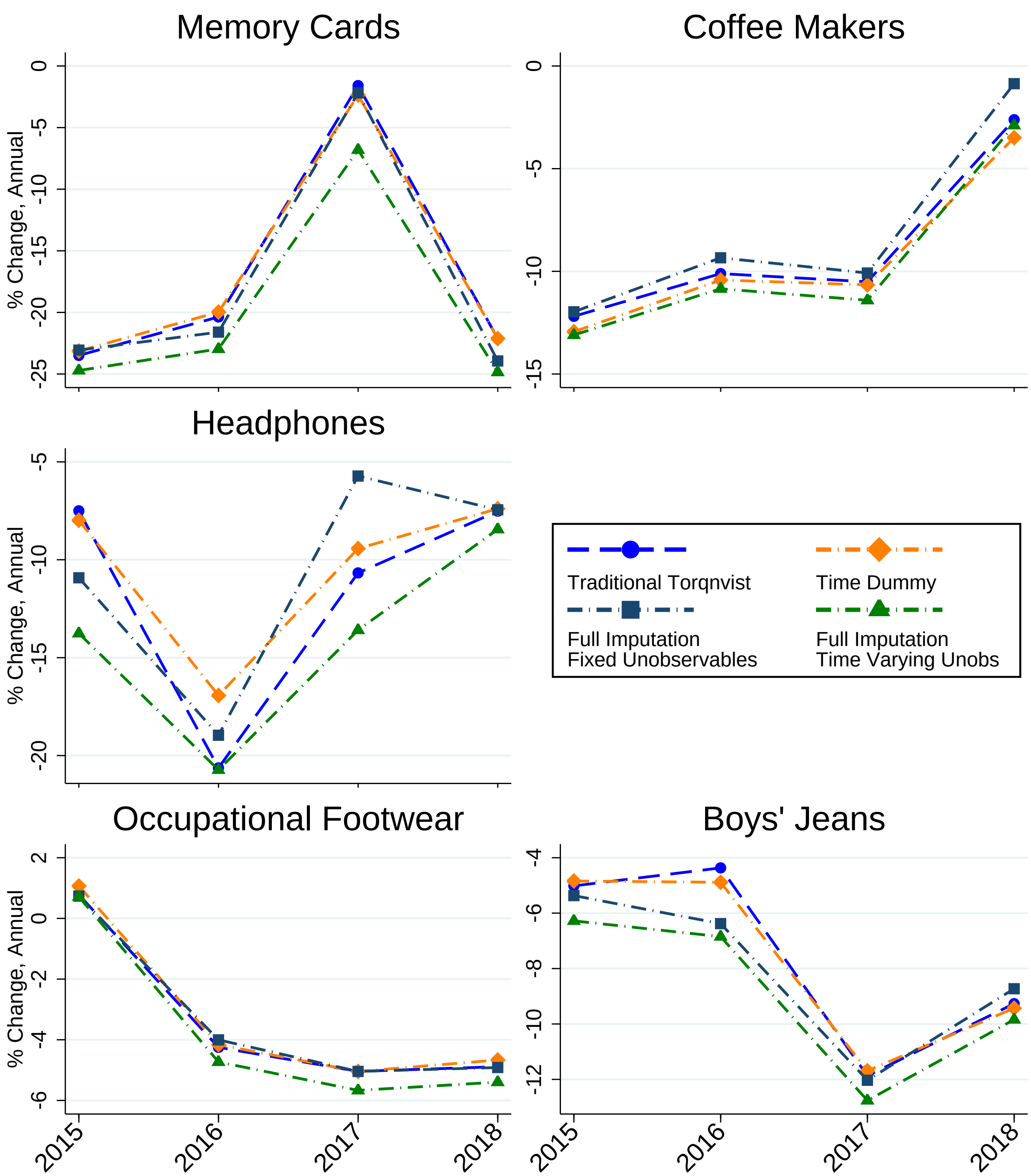
Notes: Chained values are averages of cumulative quarterly rates for year. GEKS-lite is the average of the geometric mean of the chained values and the YoY price indices for q4 for each year. Laspeyres is the geo Laspeyres. CUPI uses 25th percentile CGR. Data come from Nielsen.

Figure 1: Product Lifecycle Dynamics



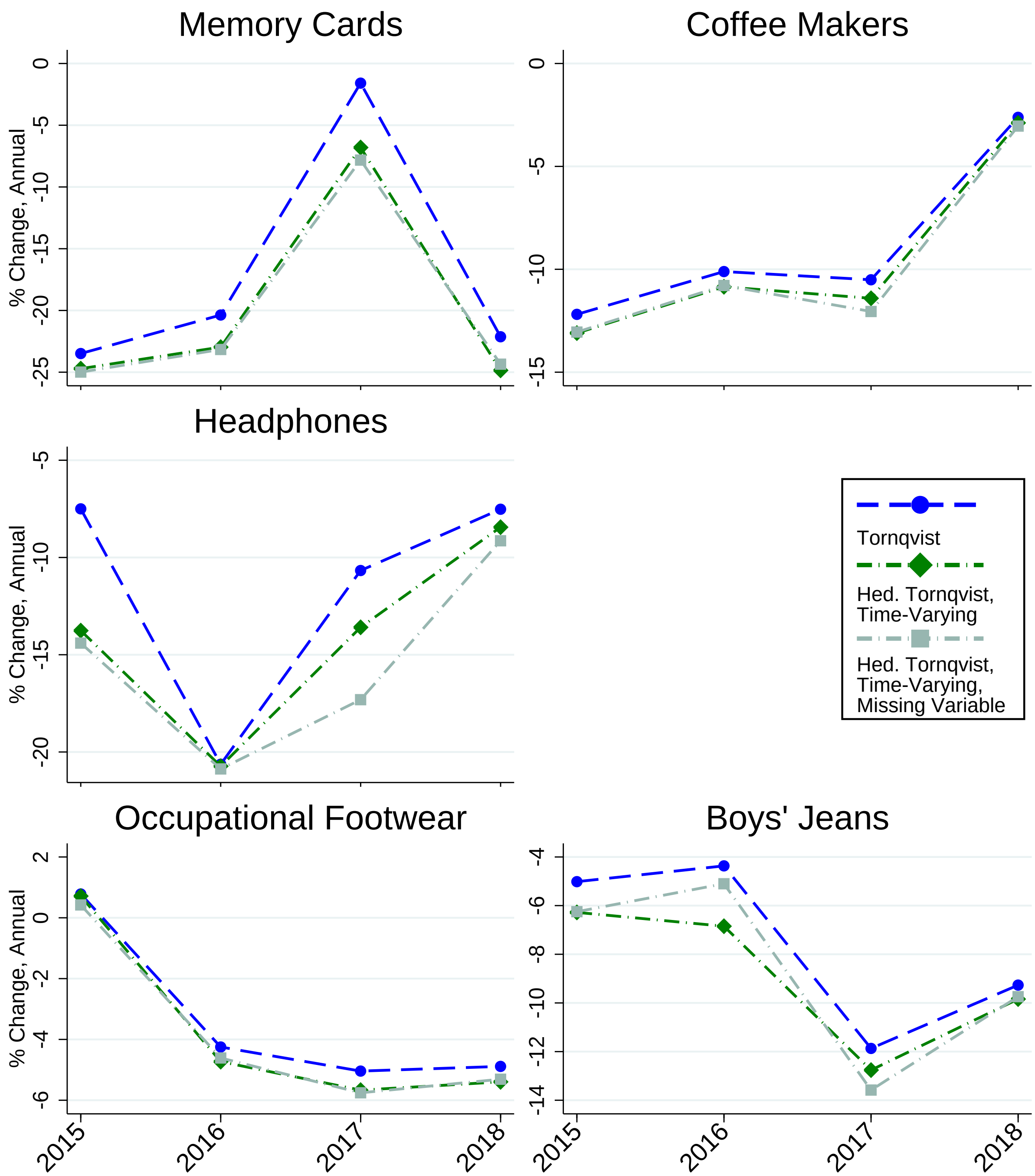
Notes: Unweighted average market share and prices relative to their value in the period of their initial entry. Entry occurs in period 0. All series are smoothed with a quartic spline. Data comes from the NPD Group.

Figure 2: Hedonic Specifications: Fixed vs. Time-Varying Unobserved Characteristics



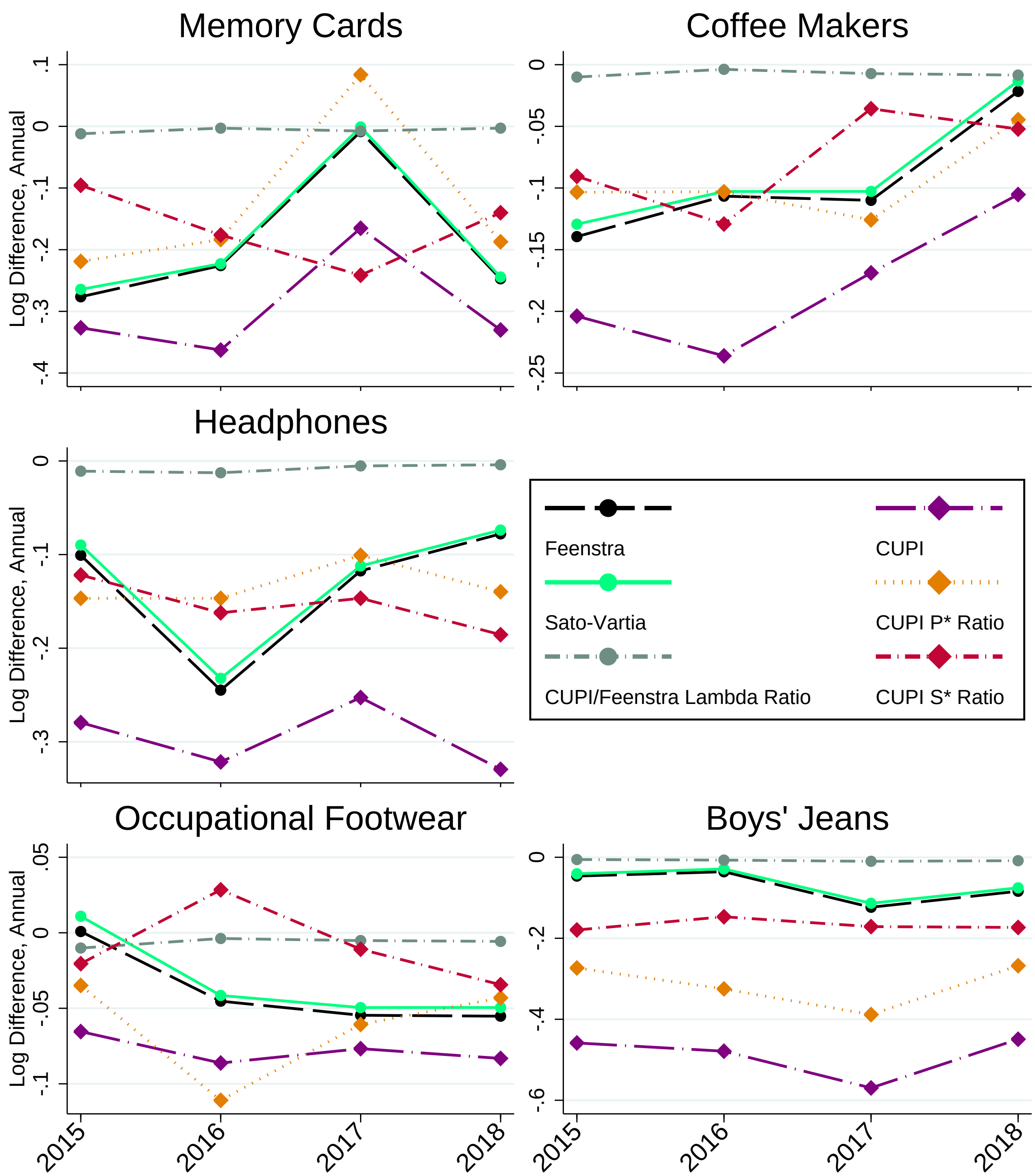
Notes: Traditional and hedonic Tornvqvist indices estimated with different methods. Values are percent change on an annual basis, aggregated from chained quarterly indices. The time-dummy Tornqvist index uses adjacent period estimation with Tornqvist market share weights. The fixed unobservables model 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: Test of Time-Varying Unobservable Specification



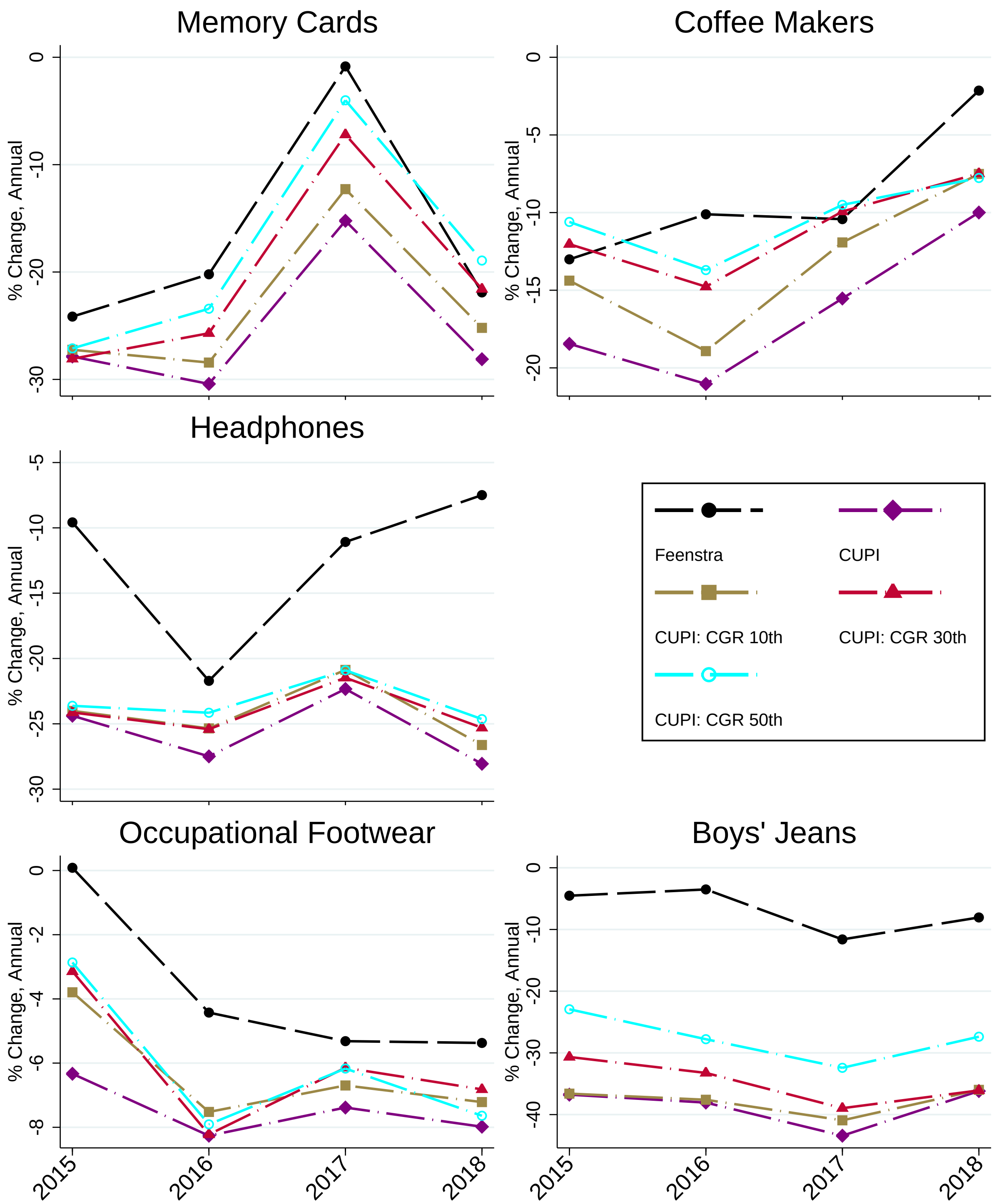
Notes: Values are percent changes on an annual basis, aggregated from chained quarterly indices. The time-varying unobservable model estimates hedonic models of log change in price using WLS and average quantity-share weights, including lagged hedonic level residuals. The "Missing Variable" series displays 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



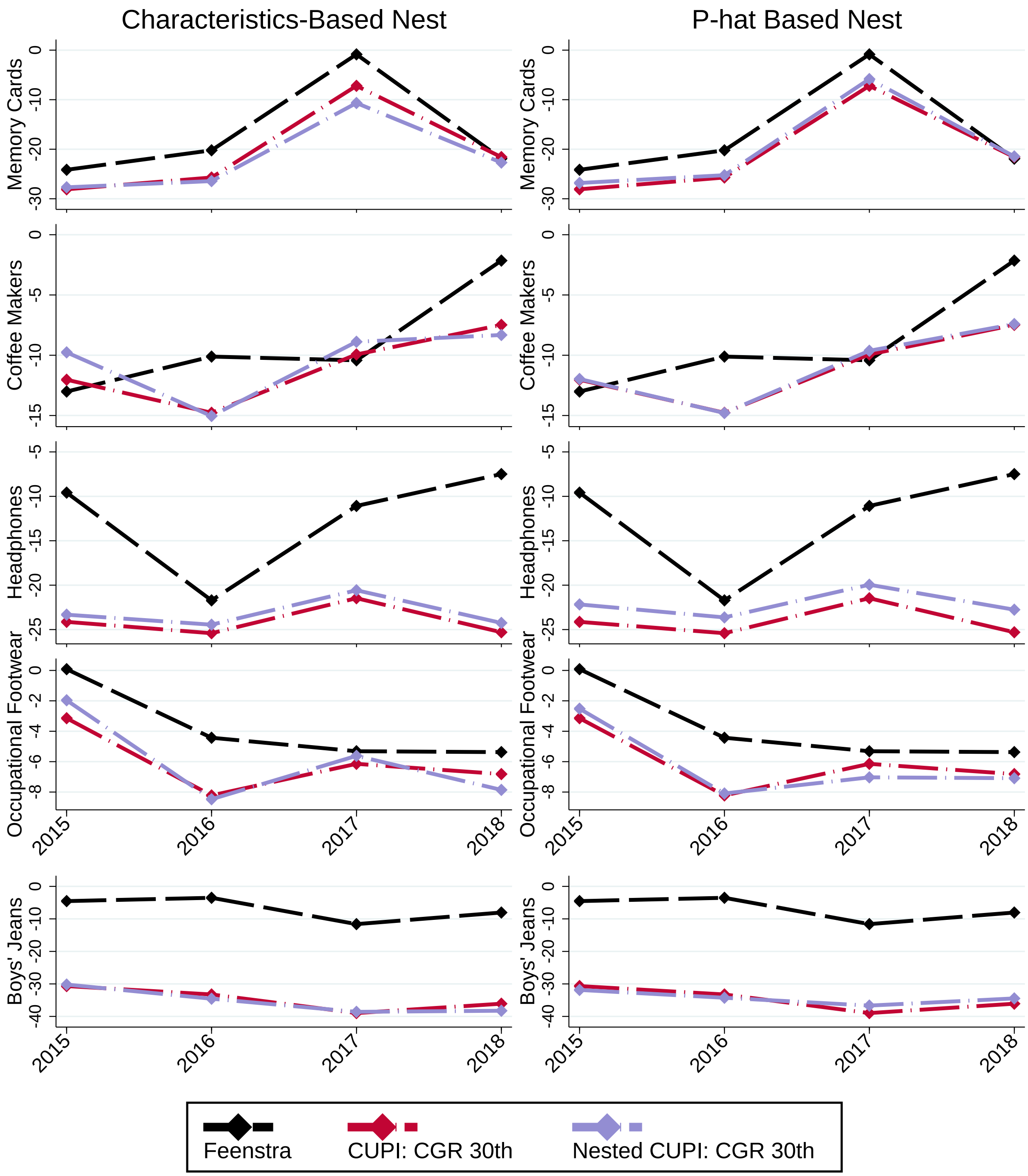
Notes: Values are percent change on an annual basis, aggregated from chained quarterly indices. Units are reported in log-differences to allow for an additive decomposition of price indices. The Feenstra index is the sum of the Sato-Vartia and CUPI/Feenstra Lambda Ratio. The CUPI is the sum of the Lambda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

Figure 5: CUPI: Common Goods Market Share Rules



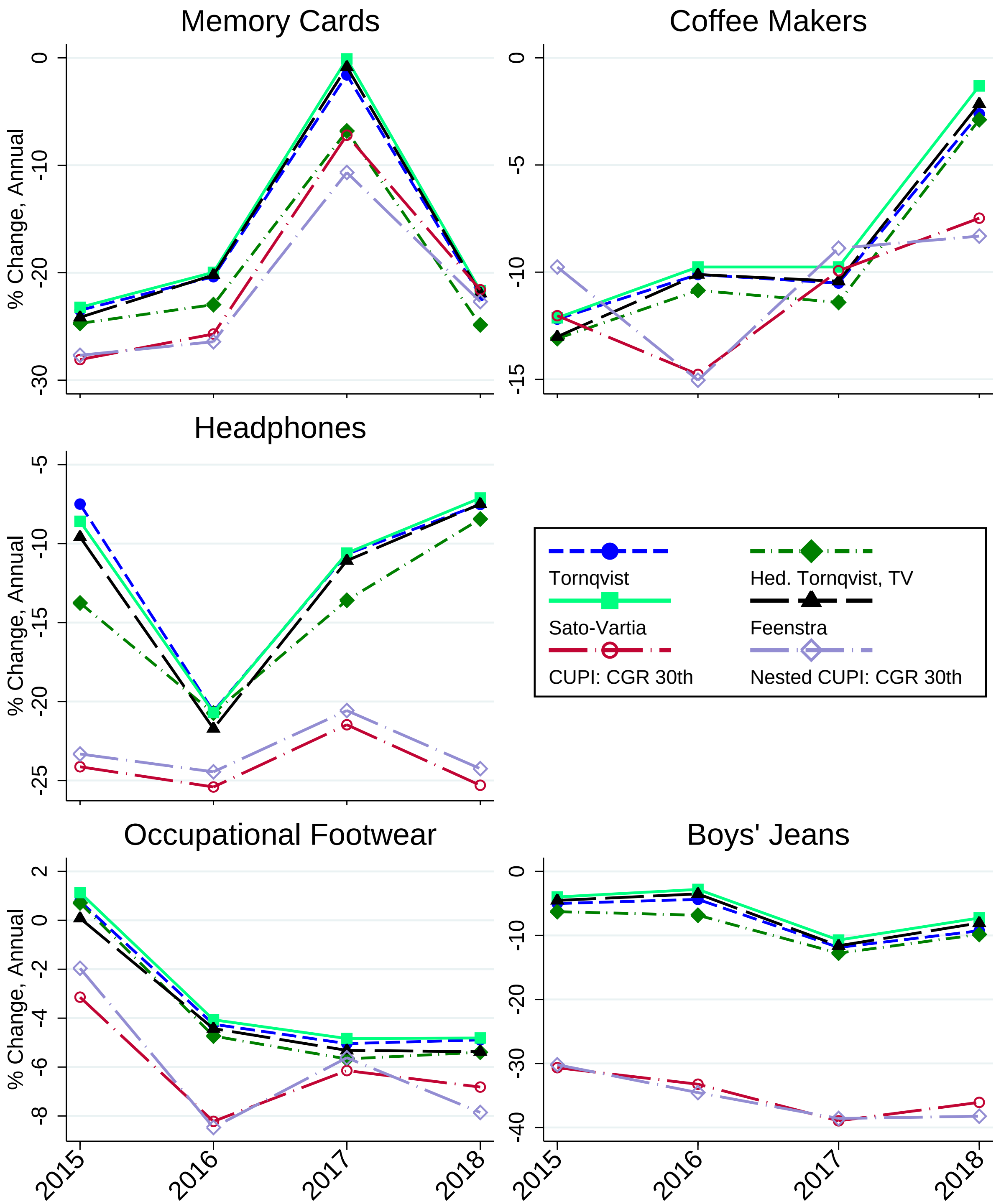
Notes: Values are percent change on an annual basis, aggregated from chained quarterly indices. Common goods market share rules for the CUPI exclude from the group of common goods those products with market shares below the noted percentile in both periods. The Feenstra-adjusted Sato-Vartia index is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

Figure 6: Nested CUPI: Characteristics- and P-Hat- Based Nests
% Changes, Annual



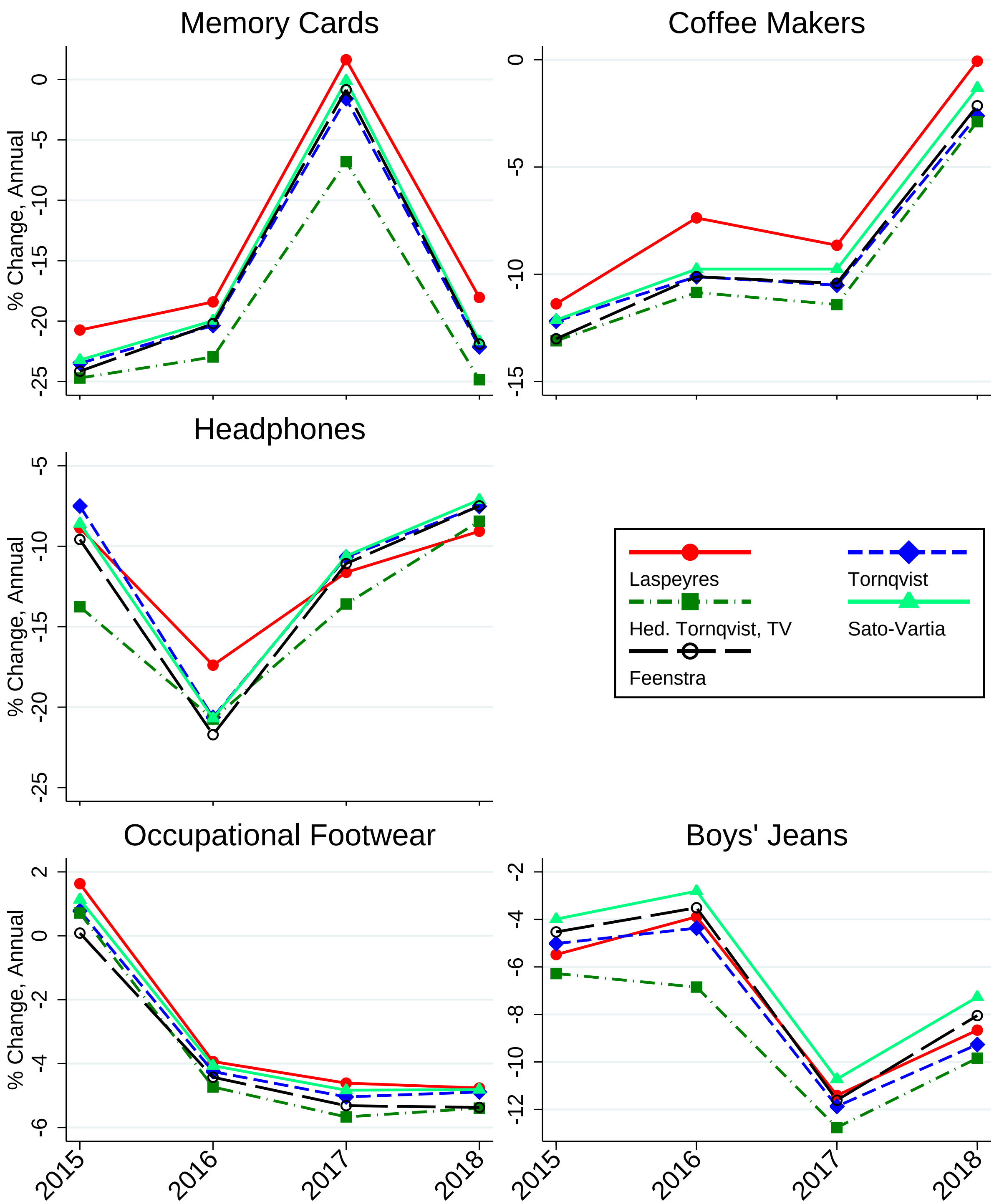
Notes: Values are percent change on an annual basis, aggregated from chained quarterly indices. For the characteristics-based nests, we assign items to groups based on shared observable characteristics. The p-hat based nests are based on the decile of predicted prices from unweighted hedonic log-level 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 7: Comparison of Main Price Index Specifications



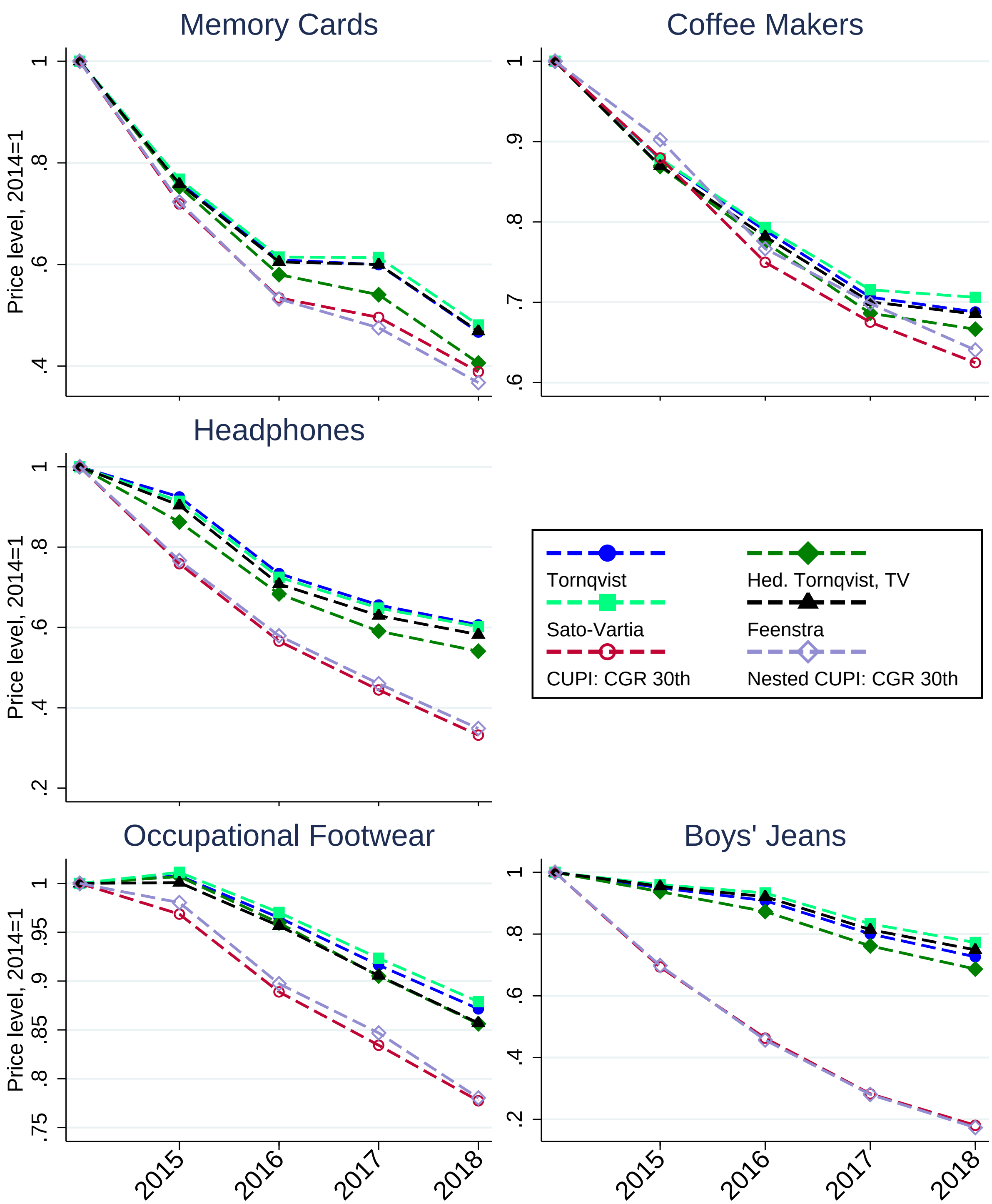
Notes: Values are percent change on an annual basis, aggregated from chained quarterly indices. The hedonic time-varying unobservables model is estimated over log price differences using WLS and with weights that are average quantity-shares in adjacent periods. Data comes from the NPD Group. The Nested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

Figure 8: Main Price Index Specifications, Without CUPI



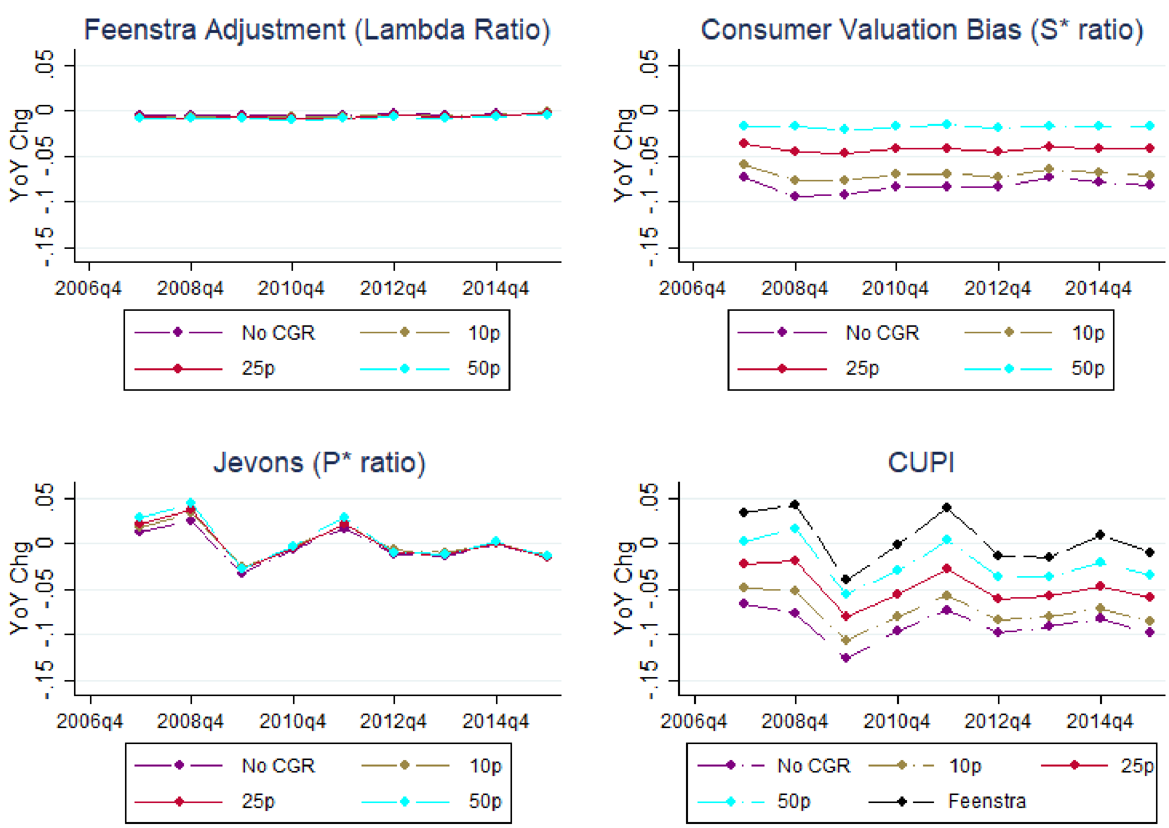
Notes: Values are percent change on an annual basis, aggregated from chained quarterly indices. The Laspeyres series reports a geometric mean Laspeyres index. The hedonic time-varying unobservables model is estimated over log price differences using WLS and with weights that are average quantity-shares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

Figure 9: Main Price Index Specifications: Cumulative Price Level Changes



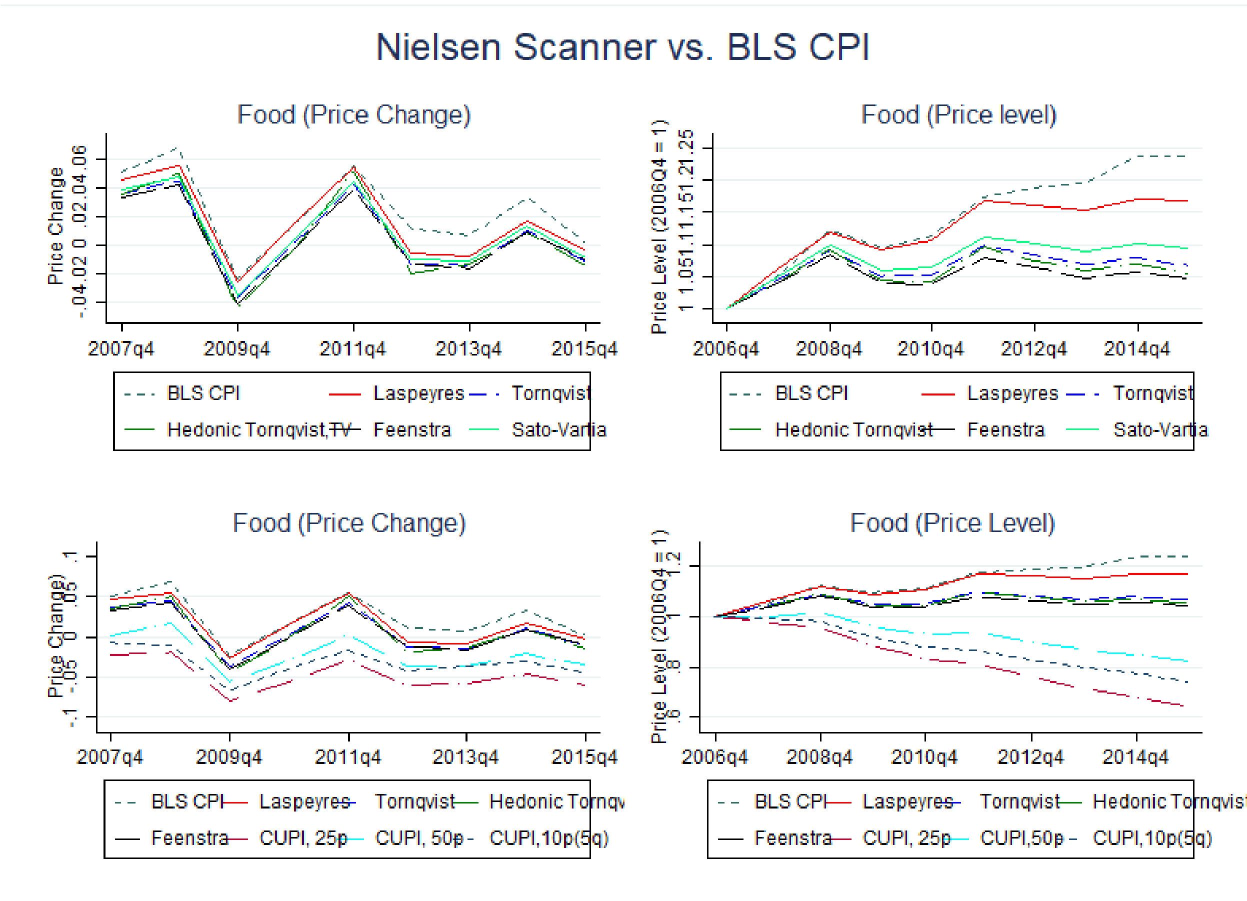
Notes: Values are cumulative changes relative to the 2014 price level, with 2014 price level set to 1. The hedonic time-varying unobservables model is estimated over log price differences using WLS and with weights that are average quantity-shares in adjacent periods. Data comes from the NPD Group.

Figure 10: Sensitivity of CUPI to CGR, Nielsen Food



Notes: Values are annual changes from cumulative chained quarterly indices. Figures use Nielsen Retail Scanner data for food product groups.

Figure 11: Main Price Index Specifications: Price Changes and Levels, Nielsen food



Notes: Price changes are are annual changes from cumulative chained quarterly indices. Price levels reflect cumulative changes relative to the 2006 price level, with 2006 price level set to 1. Figures use Nielsen Retail Scanner data for food product groups.

Appendix

A Hedonic Imputation Indices

A.1 Hedonic Estimation: Levels vs. Difference and Weighted vs. Unweighted Results

We estimate hedonic models in both log-levels and log-differences. We also consider weighted and unweighted approaches. Figure D.1 presents results from these alternative estimation approaches for the five product groups we have explored in the NPD data. The log-level specifications, whether weighted or unweighted, yield more erratic patterns than the log-difference specifications. The log first-difference results are similar whether weighted or unweighted. Importantly, the log first-difference specification in this figure controls only for time-invariant unobservable characteristics. Our preferred TV approach, which we illustrate in Figure 7 of the main text, also controls for time-varying valuations of unobservable product characteristics.

The log-level specifications are sensitive to omitted unobservable characteristics. To illustrate this point clearly, Figure D.2 presents an enhanced version of Figure 3 that shows the sensitivity of the levels specification to intentionally omitted key observable characteristics. Unlike the TV approach, the log-levels specification is very sensitive to omitting these observable characteristics.

We report goodness of fit statistics for the alternative specifications in Table D.1. 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. The EP methods (EP1 is first differences and EP2 is the TV approach) yield much higher R^2s for the price relatives, especially for the weighted specifications.

In the main text, we focus on weighted hedonic specifications. As we have noted, the time dummy method inherently calls for a weighted specification, as the estimation weights determine the type of price index produced. The hedonic imputation specifications we consider also use weighted specifications in the hedonic estimation procedure. Using weights promotes consistency with the the time dummy results. Moreover, in the context of item-level transactions data, there are additional reasons to prefer weighted estimation of hedonic models as in Bajari et al. (2021). First, item-level data is generally inclusive of products with a wide range of availability. The objective is to obtain the quarter-by-quarter mapping between prices and characteristics. With scanner data, the sample will typically include goods that are not widely geographically available or that few consumers actually purchase, partly because they have recently entered or are about to exit the marketplace. These low-quantity goods will have an outsized influence on unweighted hedonic estimates, and may therefore lead to hedonic price indices that do not reflect the environment faced by the representative consumer. Intuitively, weighted regression coefficients should be interpreted as the implicit prices of characteristics that consumers actually purchased (see Silver (2003) and Diewert (2002) for motivation of using weighted hedonic specifications along these lines).

We use quantity-share weights in our hedonic specifications, but we have found broadly similar results using market-share weights. De Haan (2008) advocates for quantity weights to be used in estimation. He notes that in the context of scanner data in particular, we do not observe prices but average unit values. Given that consumers purchase items from different stores, at different times during the aggregated periods over which average unit values are calculated, and perhaps with different bargaining power, it is likely that these unit values are likely to be measured with heteroskedastic errors across different items. Importantly, the variance of unit values is inversely proportional to the square root of the number of units sold, rather than the total value sold. Quantity weights are also frequently used in the trade literature, which similarly often depends on unit values of imports or exports (e.g., Broda and Weinstein, 2006).

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 method presented in the main text take the same approach.

B Using the Nielsen Data

B.1 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. For the latter, we use Personal Consumption Expenditures (PCE) data from the Bureau of Economic Analysis for nominal sales comparisons. 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.

Figure D.5 presents comparisons of nominal expenditures for the broad food and nonfood categories. 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.[44]

These patterns are consistent with the discussion in the main text that the Nielsen data's coverage of nonfood items has deteriorated over time. Figure D.6 provides more guidance on this point by showing for detailed categories the Scanner data the ratio of the growth in nominal sales for the Nielsen Scanner for the period 2008:1 to 2015:4 relative to the growth in nominal sales for the PCE over the same period. The upper panel shows results for food categories and the lower panel for nonfood categories.[45] The categories from left to right

[44]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.

[45]The categories are more aggregated than Nielsen product groups. They reflect a concordance provided

in each panel are ranked by expenditure shares. Many of the food product categories have ratios close to one. In contrast, the nonfood categories have ratios that are much more variable and also typically below one.[46]

Figure D.7 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.2 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.8 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), vs. a 5-quarter horizon (i.e., computing percentiles for sales pooled over quarters t and $t-4$).[47] 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.9 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 us by BLS between PCE categories and Nielsen product modules.

[46]The results presented by detailed category are for the Nielsen Scanner data, which is the primary focus of our analysis. In unreported results, we find similar patterns for the Nielsen Consumer Panel data.

[47]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.

To facilitate comparison of our results to Redding and Weinstein (2020), who report pooled results for food and nonfood product groups, Figure D.10 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.11, 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.[48]

Figure D.12 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.[49] 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 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.13 displays for the nonfood product groups the analogous plots to Figure 11, 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.[50]

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

[48]We note that we do not impose a CGR in computing the other price indices shown in Figure D.10. 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 10 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.

[49]Figure D.12 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.

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

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.3 Machine Learning and Hedonics

This appendix summarizes our procedure for incorporating machine learning into hedonic estimation. Our companion paper, Cafarella et al. (2021), provides further details.

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 "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 (1) and (3). 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.

Several features of our methodology merit particular discussion. First, to convert text-based product descriptions into numerical characteristic representations, we use a hybrid feature encoding architecture that allows the system to incorporate "pre-trained" word embeddings (numerical representations) trained from an external corpus of text as well as specifically trained or "text-tailored" embeddings trained specifically on the product descriptions in the Nielsen Kilts Retail Scanner Data set. Second, our architecture does not predict prices or price changes directly, but rather predicts a set of probabilities that the price or price change lies in each of a set of price or price-change bins that partition the observed range. Third, the ML system minimizes the weighted cross-entropy loss function for the products' true price and price change distributions in the hedonic estimation.[51] Both steps are weighted using products' unit sales (quantities) shares in a product-group quarter. Fourth, because of the noise in the estimated probabilities, it may not be optimal to calculate price predictions as the simple probability-weighted expected price. We use a receiver operating characteristic (ROC) curve procedure to determine the optimal number of bins to include in the price prediction.

Cafarella et al. (2021) explores the ML procedure's performance as measured by the prediction "near accuracy" across every product group-quarter. We define the model's near

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

accuracy as the proportion of products for which it assigns the highest probability to the correct or an adjacent bin. The median in-sample near accuracy for food price change bins is well above 80%. The out-of-sample near accuracy for the median product group-quarter is nearly 60% for the food product groups. In other words, the median-performing model predicts the the correct bin or an adjacent bin more than 80% of the time. We view these model performances as remarkable: in the median product group-quarter, the system is able to closely predict a product's price change over half the time based on the short, nonstandard product descriptions.

C Examining the Behavior of the Exact CES Price Indices with Time-Varying Product Appeal

In this appendix, we examine analytically and via simulation evidence 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. We begin in section C.1 by examining 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.

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.2, 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 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:

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

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}:

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

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:

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

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.

C.2 Simulation Evidence on the Behavior of the CES Exact Price Indices

C.2.1 Simulation Model Environment

We base our simulations on the general equilibrium environment of Hottman et al. (2016).[52] 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.

[52] 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.

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_{ut}^{U} units of product u at time t, A_{ut}, is given by:

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

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 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.[53] 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}.[54] 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

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

[54] 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.

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.

C.2.2 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.14 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.1, 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.15 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.14, 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.

Product Upgrading and Downgrading via Turnover

Figure D.16 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.[55] 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.16 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.16 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.14, 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.17 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

[55] 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.

Section C.2.1. 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 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.4, 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.16, the horizontal axis of Figure D.17 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.16 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.17 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.17 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.18 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.[56]

The unit expenditure function in Figure D.18 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.[57] 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.

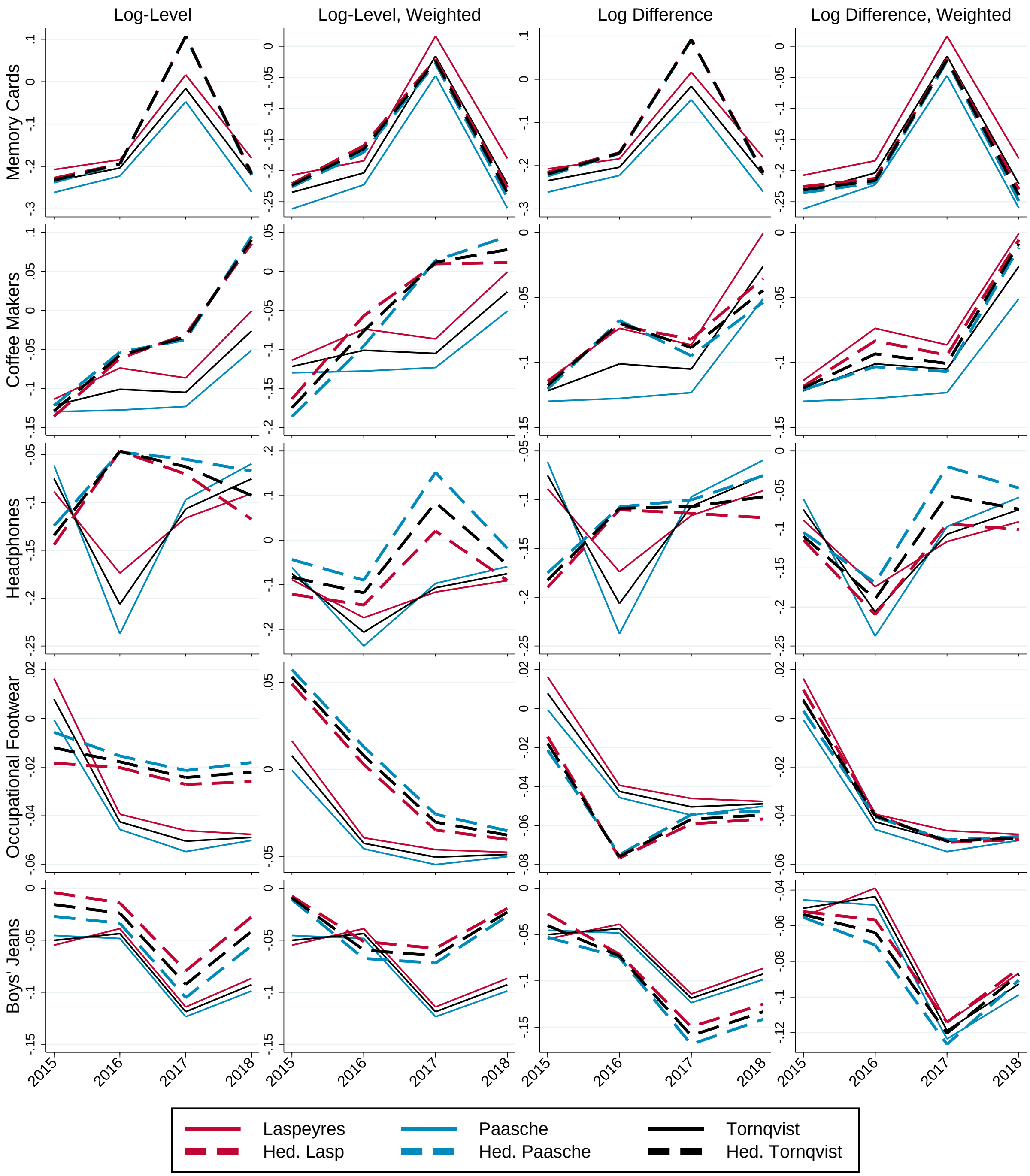
The key messages from Figure D.18 are similar to those from Figure D.17. The theoretical CUPI is significantly biased in the presence of this market friction, while the Sato-Vartia index is approximately unbiased. Imposing a CGR helps move the CUPI closer to the true unit expenditure function. Again, 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.

[56] 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.

[57] 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.

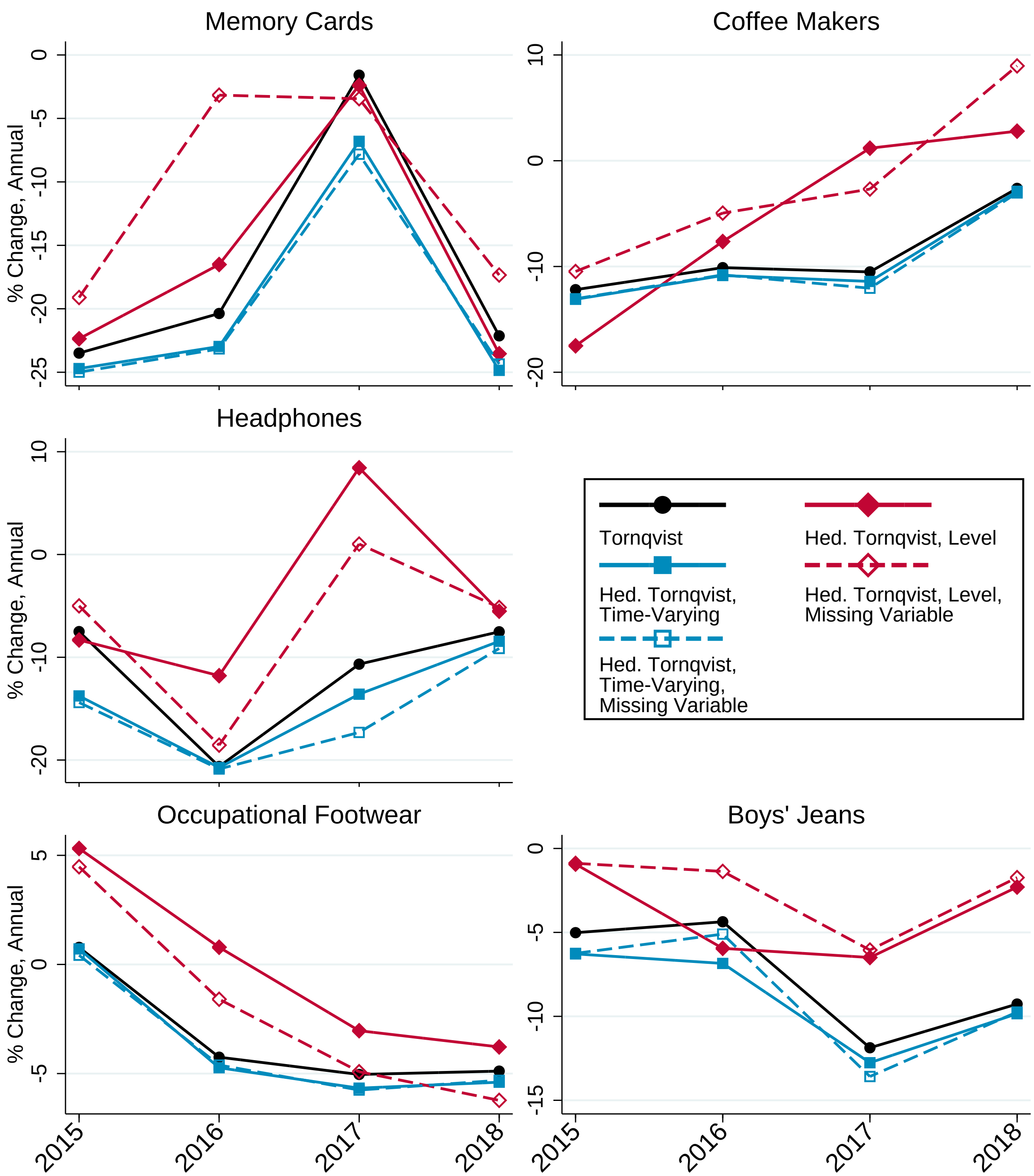
D Appendix Tables and Figures

Figure D.1: Alternative Hedonic Estimation Strategies
NPD Data, Annual % Changes



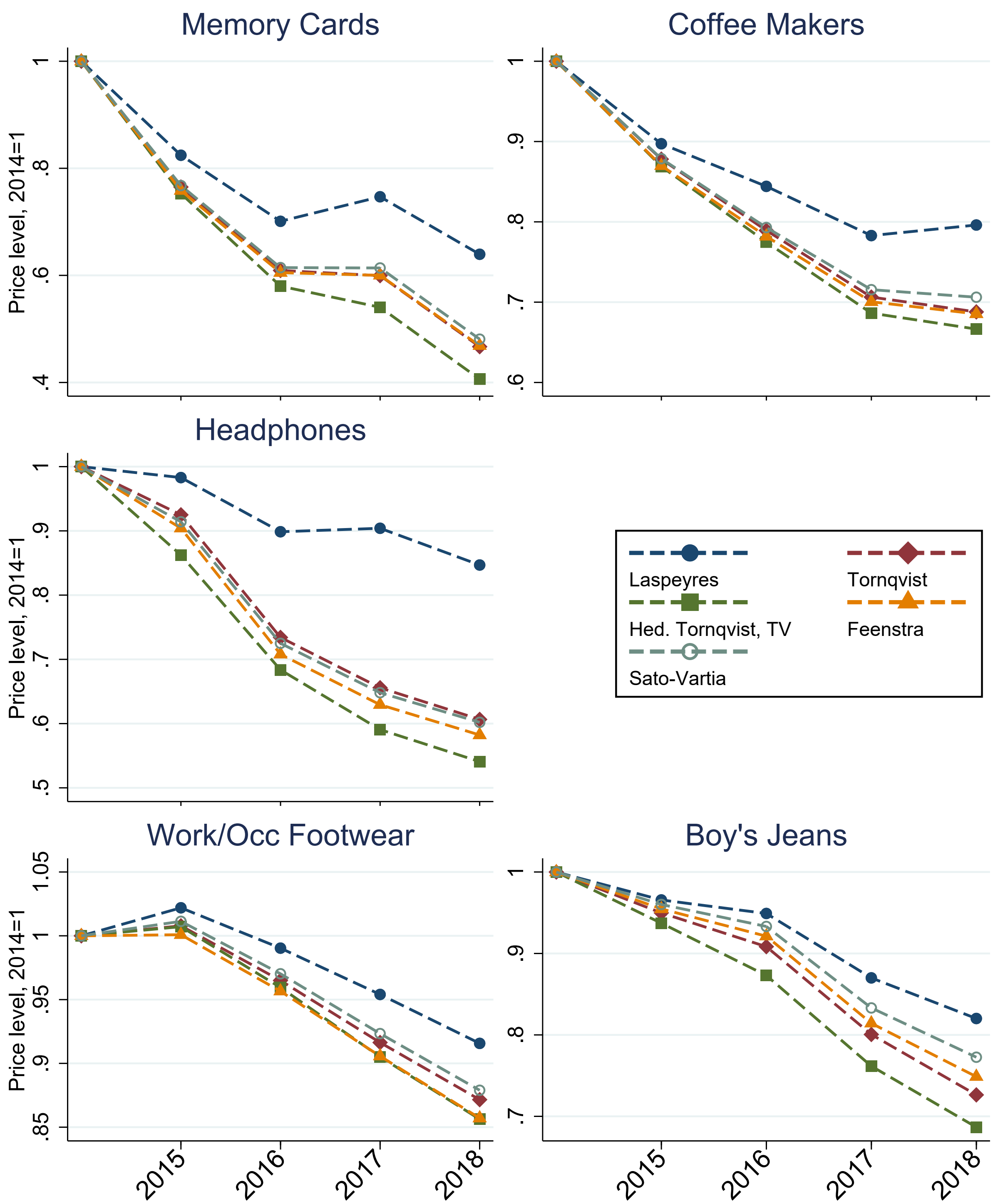
Notes: Values are annual changes from cumulative chained quarterly indices. Data comes from the NPD Group.

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



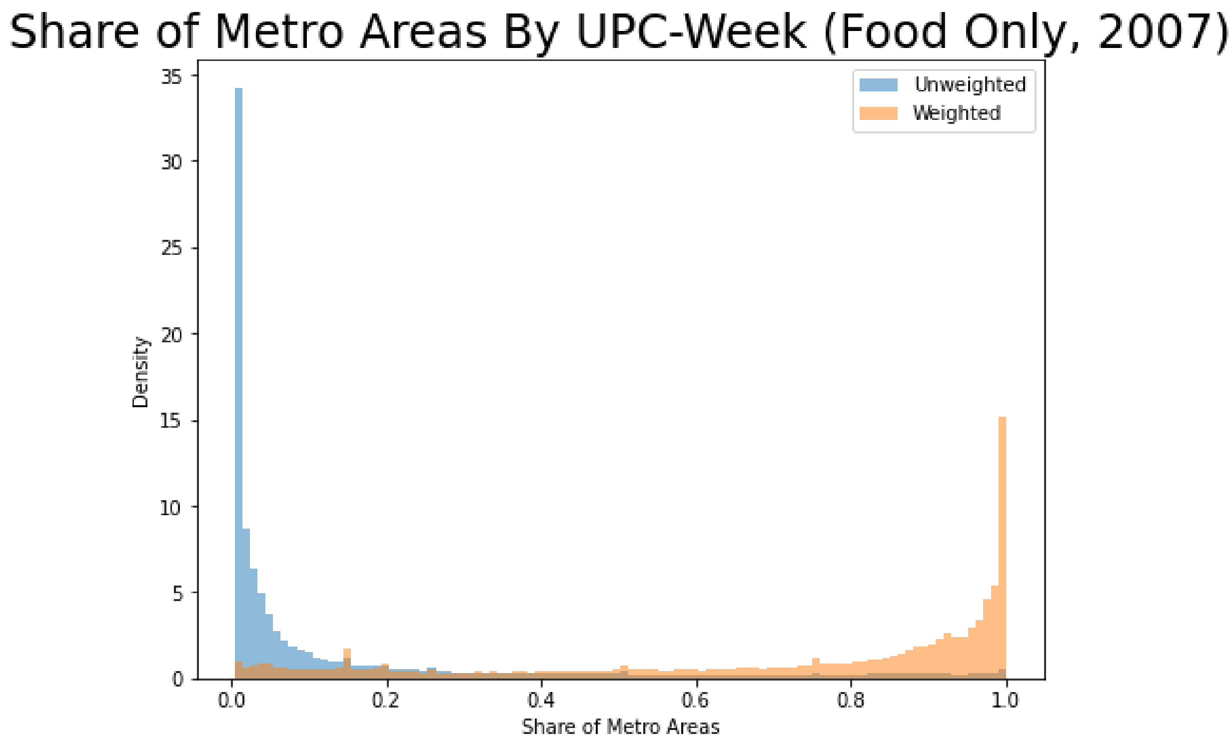
Notes: Values are log differences on an annual basis, aggregated from chained quarterly indices. Data comes from the NPD Group.

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



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.

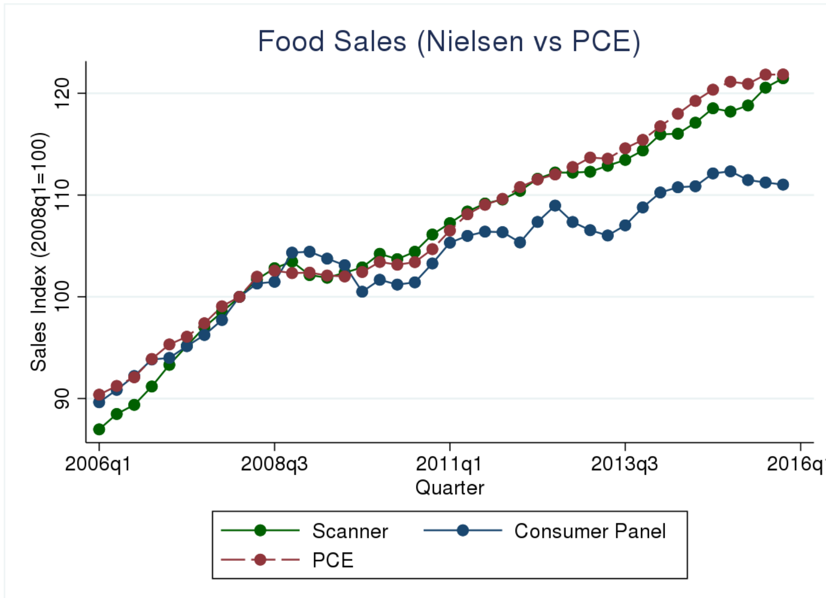
Figure D.4: Sales-weighted and Unweighted Distributions of Market Penetration of Items in Nielsen Data, Food



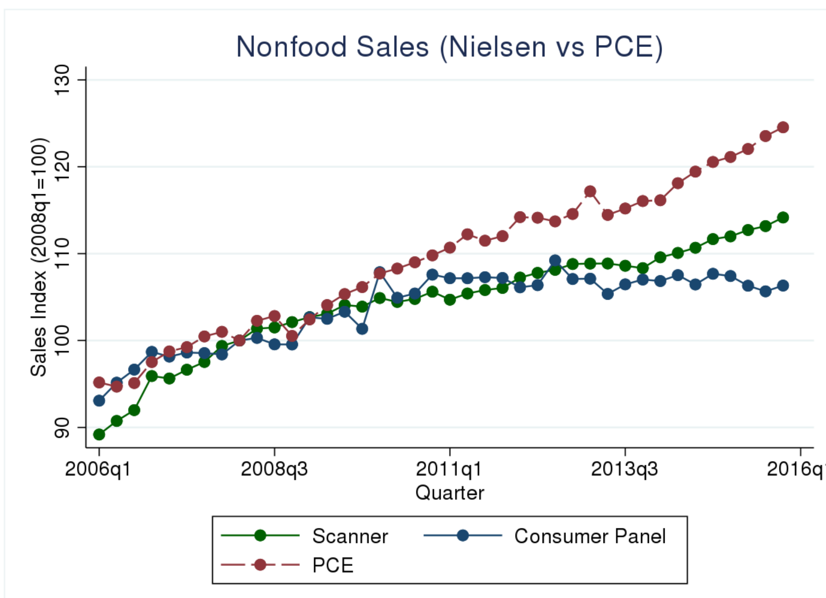
Notes: All UPC items at a weekly frequency are used from 2006-2015. Unweighted shows the market penetration at the metro area of the unweighted pooled distribution. Sales-weighted shows the equivalent using sales weights. Figure uses Nielsen Retail Scanner data for food product groups.

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

(a) Food



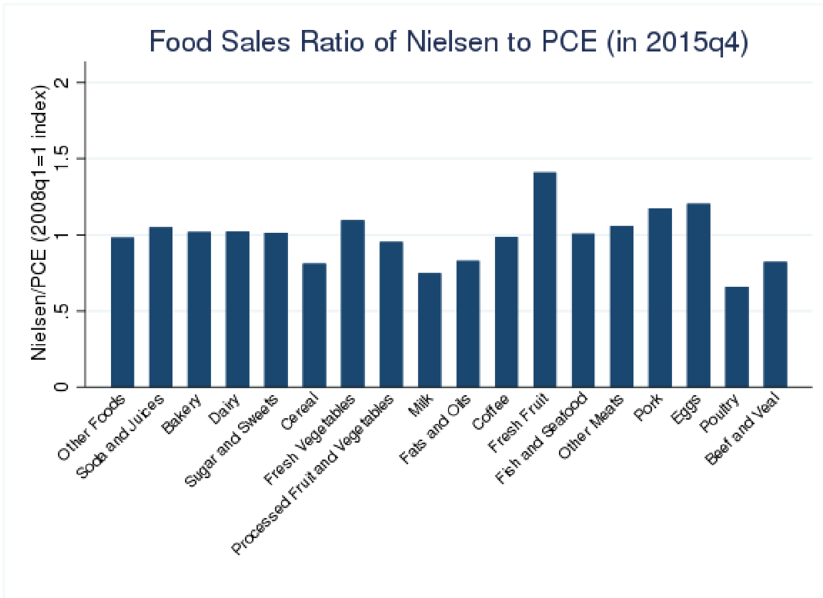
(b) Nonfood



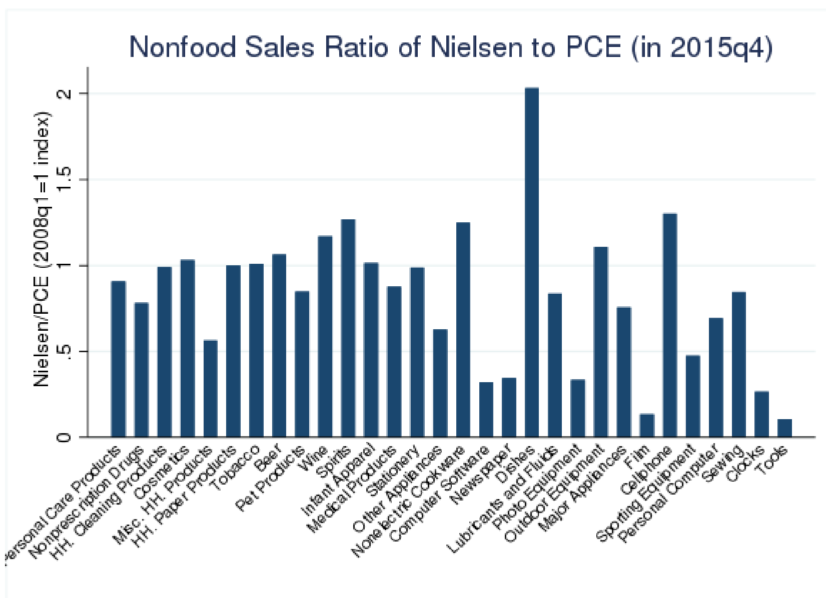
Notes: Figures uses Nielsen Scanner and Consumer Panel data for Food and Nonfood (aggregated) product groups. PCE is personal consumption expenditures (nominal) from BEA. All series indexed to 1 in 2008:q1.

Figure D.6: PCE vs Nielsen Sales for Scanner, By Category within Food and Nonfood

(a) Food



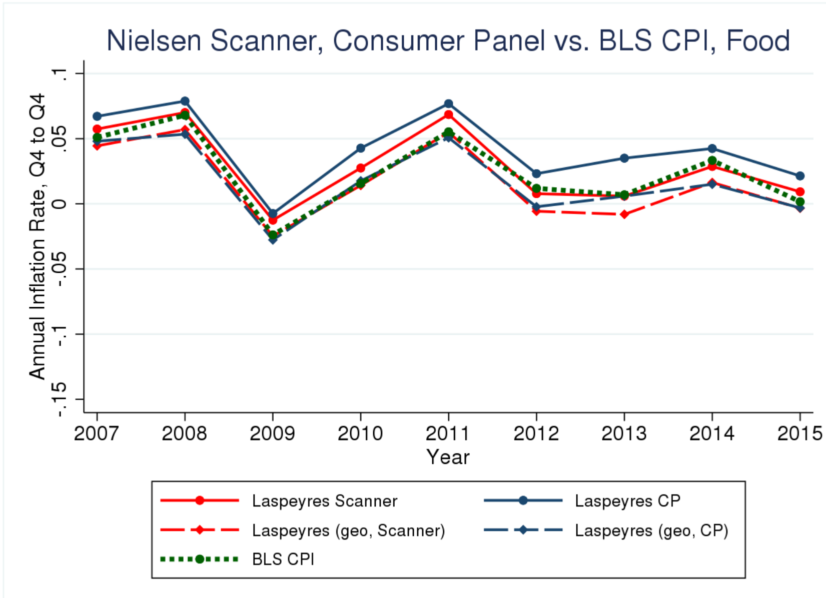
(b) Nonfood



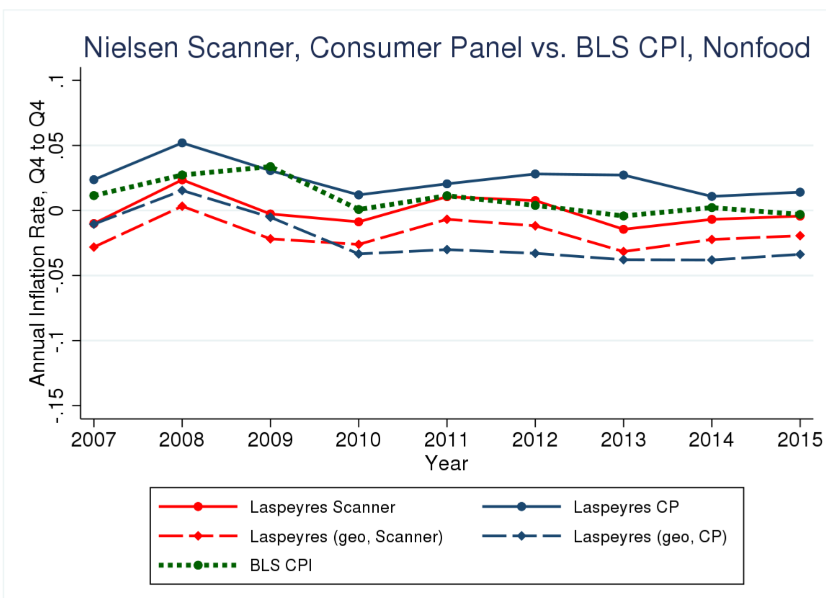
Notes: Figures uses Nielsen Scanner and Consumer Panel data for Food and Nonfood (aggregated) product groups. PCE is personal consumption expenditures (nominal) from BEA. All series indexed to 1 in 2008:q1.

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

(a) Food

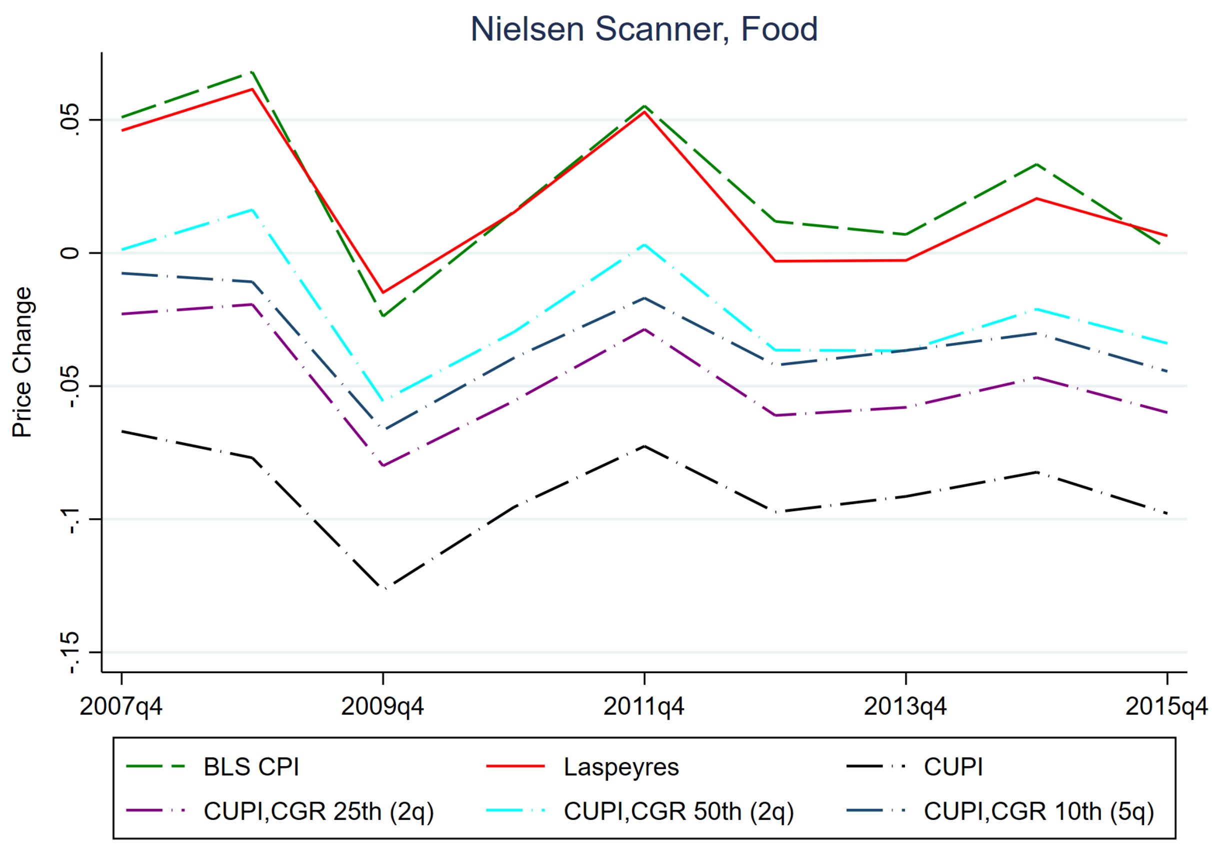


(b) Nonfood



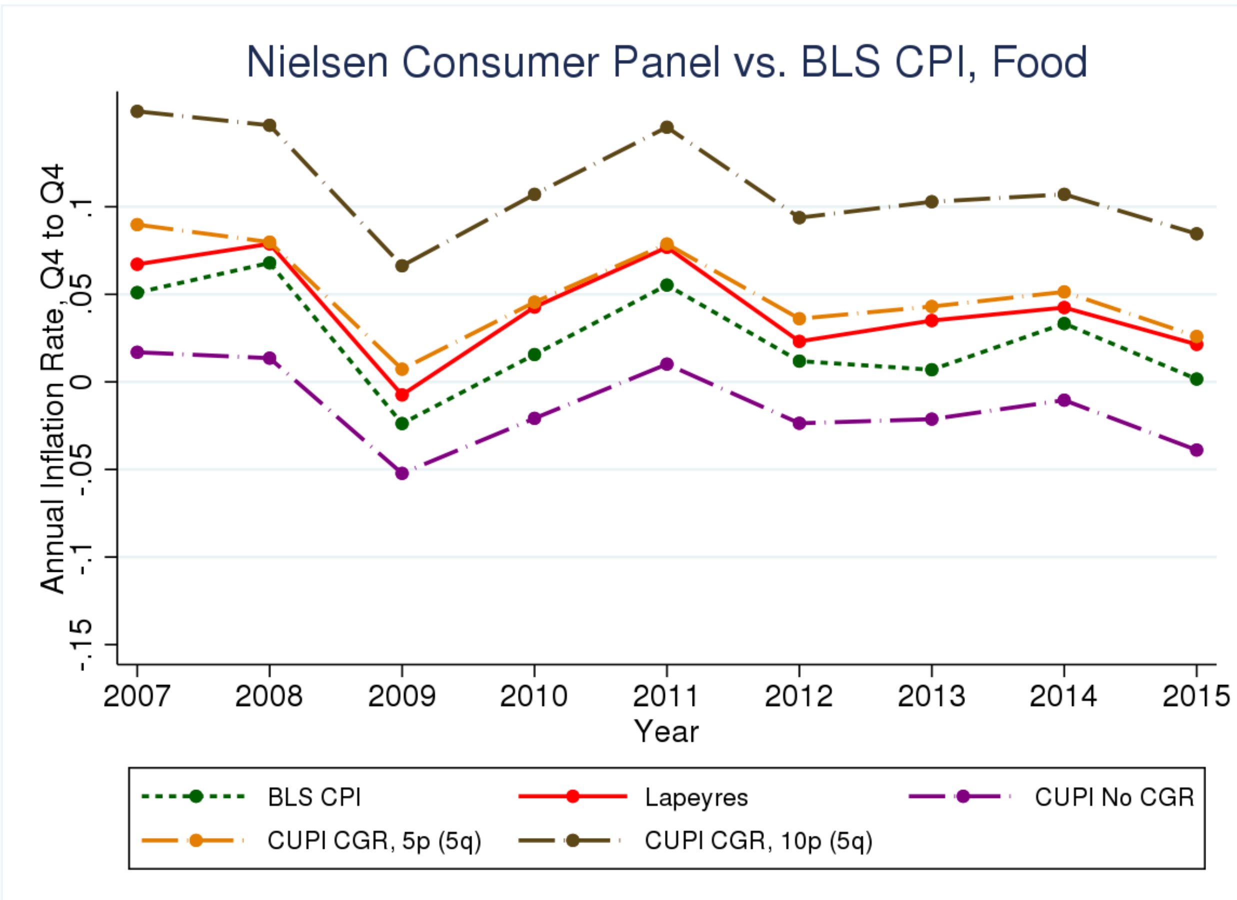
Notes: Figures uses Nielsen Scanner and Consumer Panel data for Food and Nonfood (aggregated) product groups. BLS CPI computed by BLS staff.

Figure D.8: Common Goods Rules–2-quarter vs 5-quarter Horizons



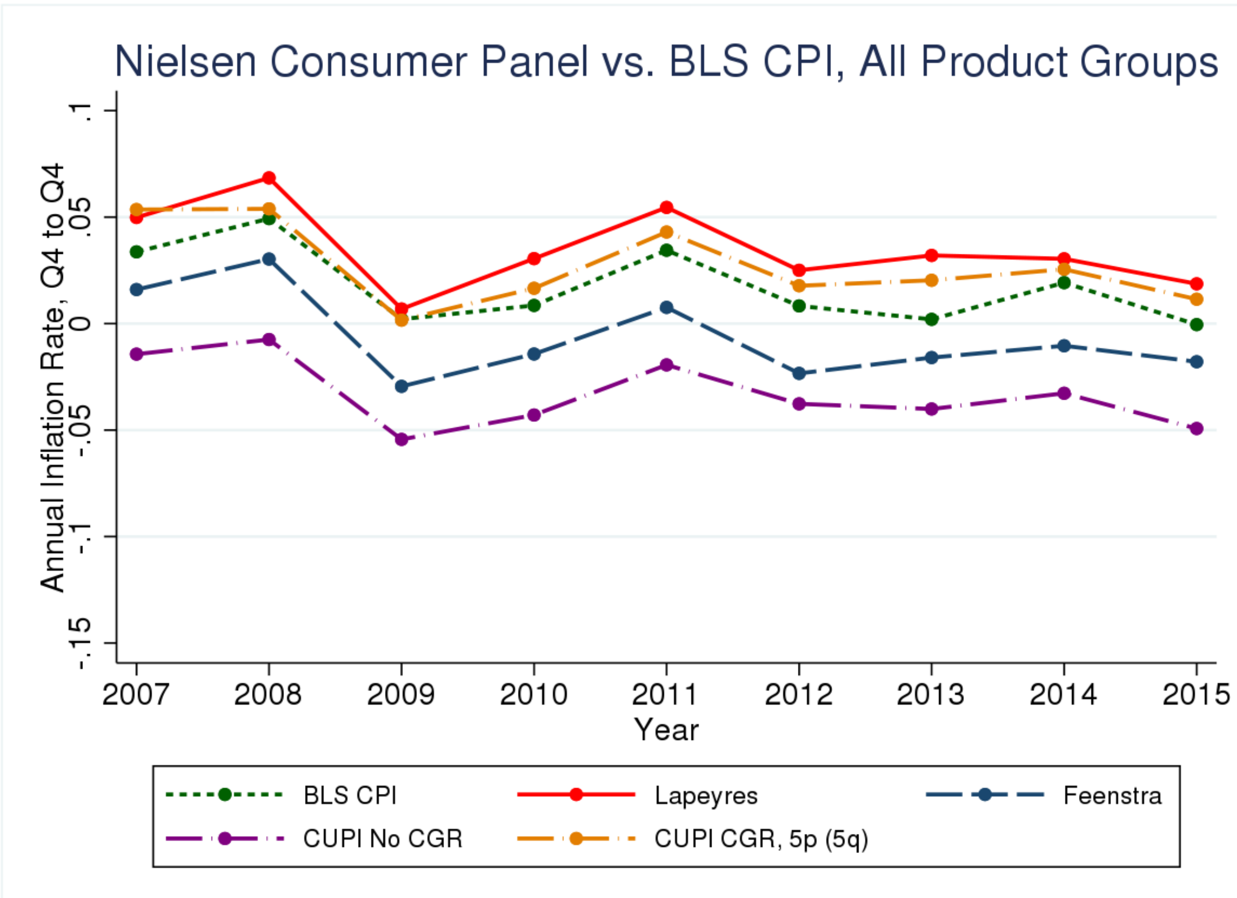
Notes: Figure uses Nielsen Scanner data for food. The $2q$ CUPI computes CGR percentile thresholds using sales pooled over a two quarter horizon (t and $t-1$). The $5q$ CUPI computes CGR percentile thresholds using sales pooled over a 5 quarter horizon (current and prior 4 quarters). Laspeyres is arithmetic.

Figure D.9: Common Goods Rules – Nielsen Consumer Panel



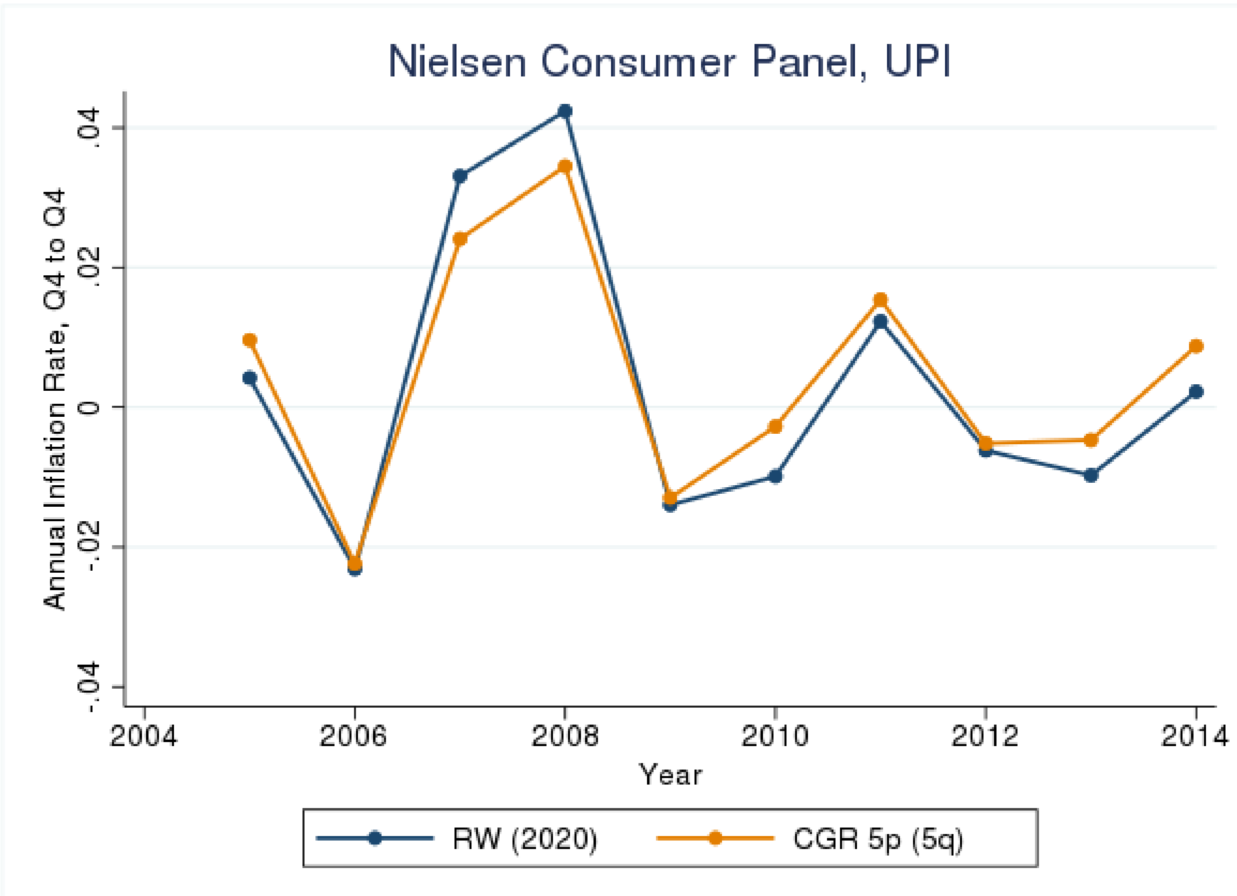
Notes: Figure uses Nielsen Consumer Panel data for food. The $5q$ CUPI computes CGR percentile thresholds using sales pooled over a five quarter horizon (t and $t-1$). (current and prior 4 quarters). Laspeyres is arithmetic.

Figure D.10: Common Goods Rules – Nielsen Consumer Panel



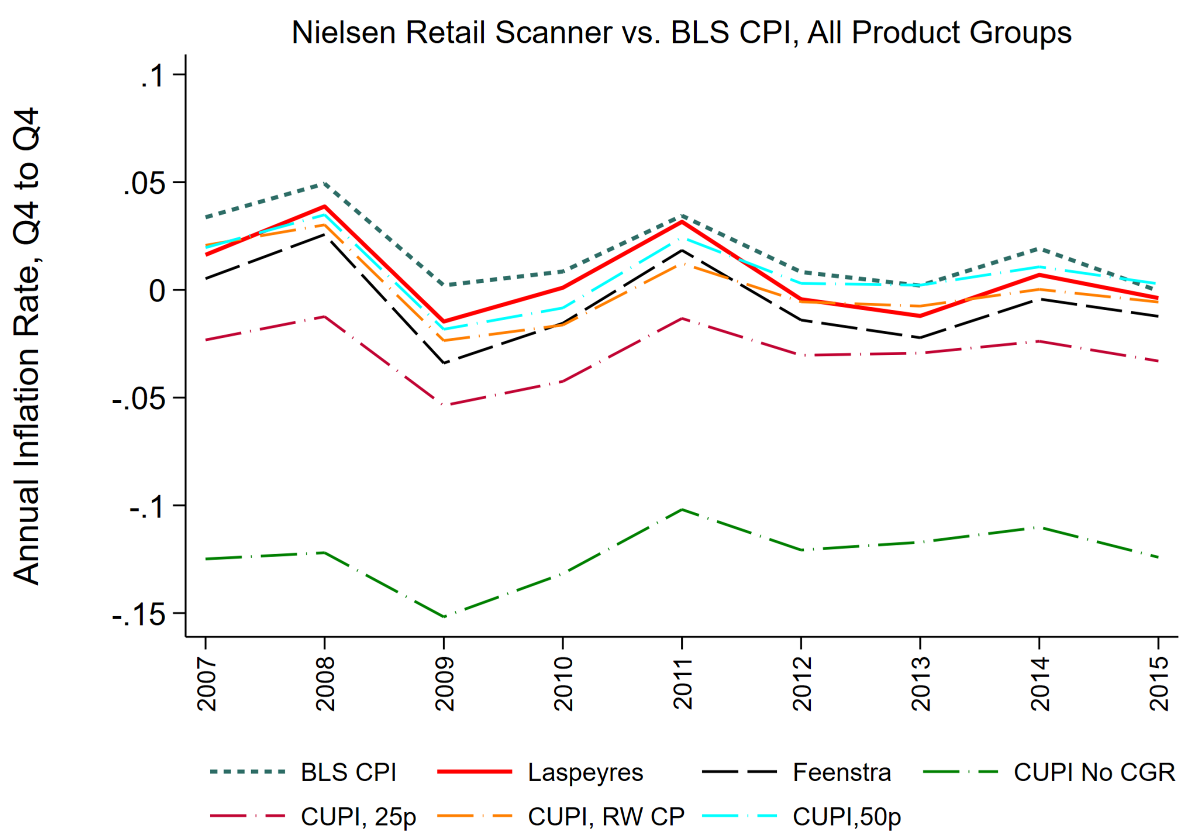
Notes: Figure uses Nielsen Consumer Panel data for food and nonfood product groups. The series "CUPI CGR RW" uses a 5th-percentile sales cutoff for the common goods rule. Percentile computed from sales pooled over 5 quarter horizon (current and prior 4 quarters). Laspeyres is arithmetic.

Figure D.11: Replication of Redding and Weinstein (2020) with Nielsen Consumer Panel



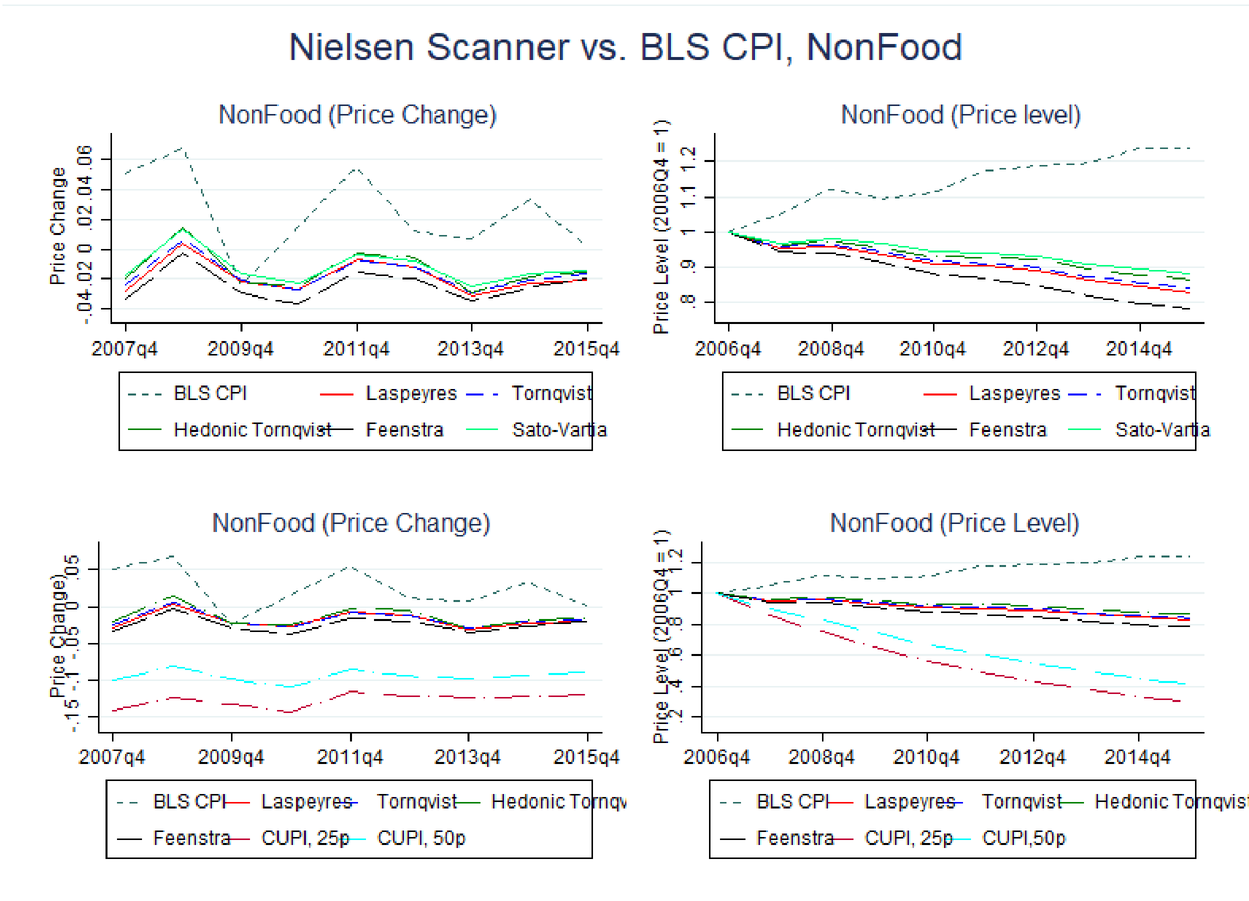
Notes: Figure uses Nielsen Consumer Panel for food and nonfood product groups. The indices are YoY for Q4. The series RW(2020) uses the same CGR duration rule as in Redding and Weinstein (2020). The series 5p(5q) use percentiles based on sales pooled over 5 quarter horizon (current and prior 4 quarters.

Figure D.12: Common Goods Rules – Nielsen Scanner Panel



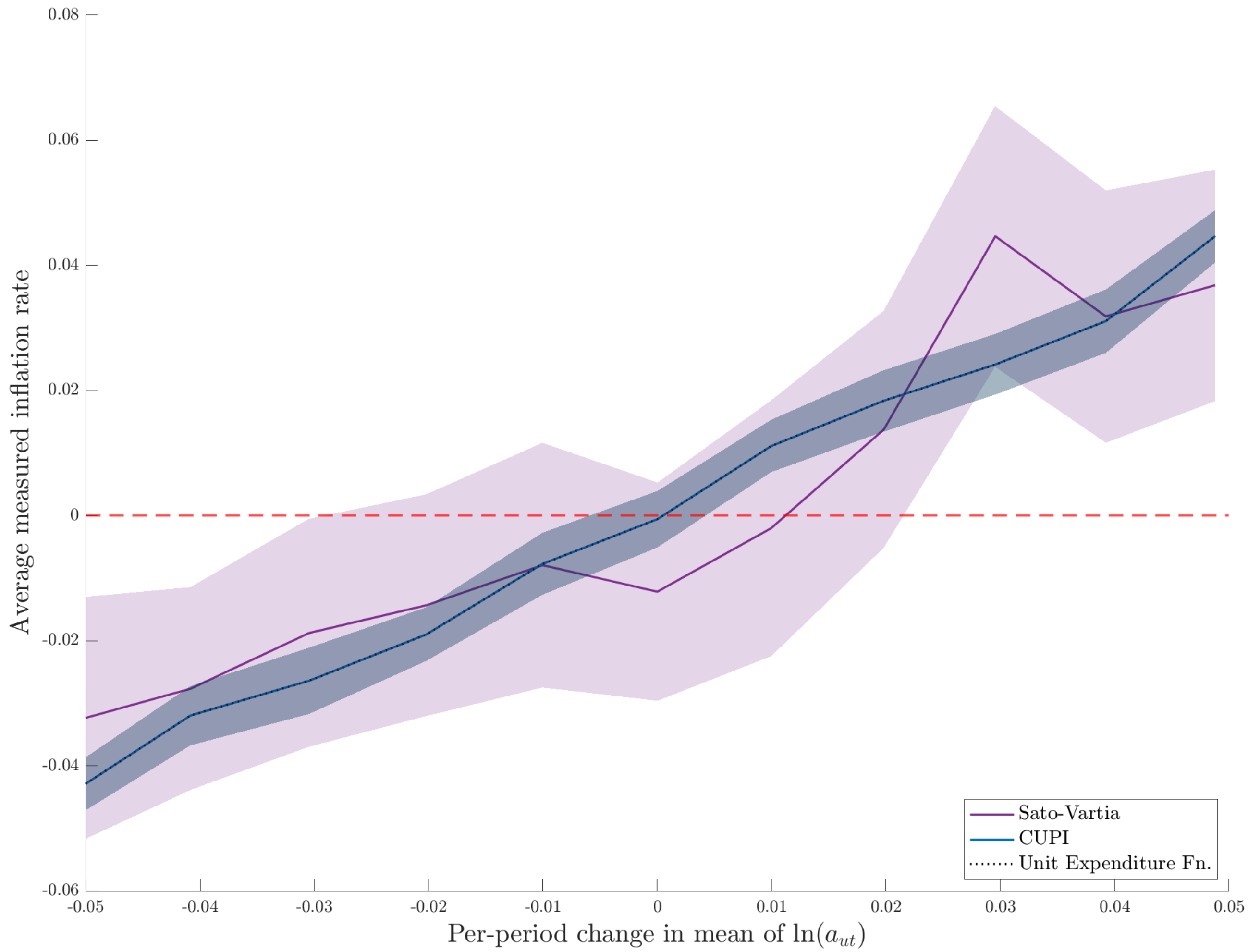
Notes: Figure uses Nielsen Retail Scanner data for food and nonfood product groups. The "CUPI, 25p" and "CUPI, 50p" series use 25th- and 50th-percentile cutoffs for the common goods rule, respectively. The series "CUPI, RW CP" uses the CGR 5th percentile threshold from the consumer Panel data for the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

Figure D.13: Main Price Index Specifications: Price Changes and Levels, Nielsen Nonfood



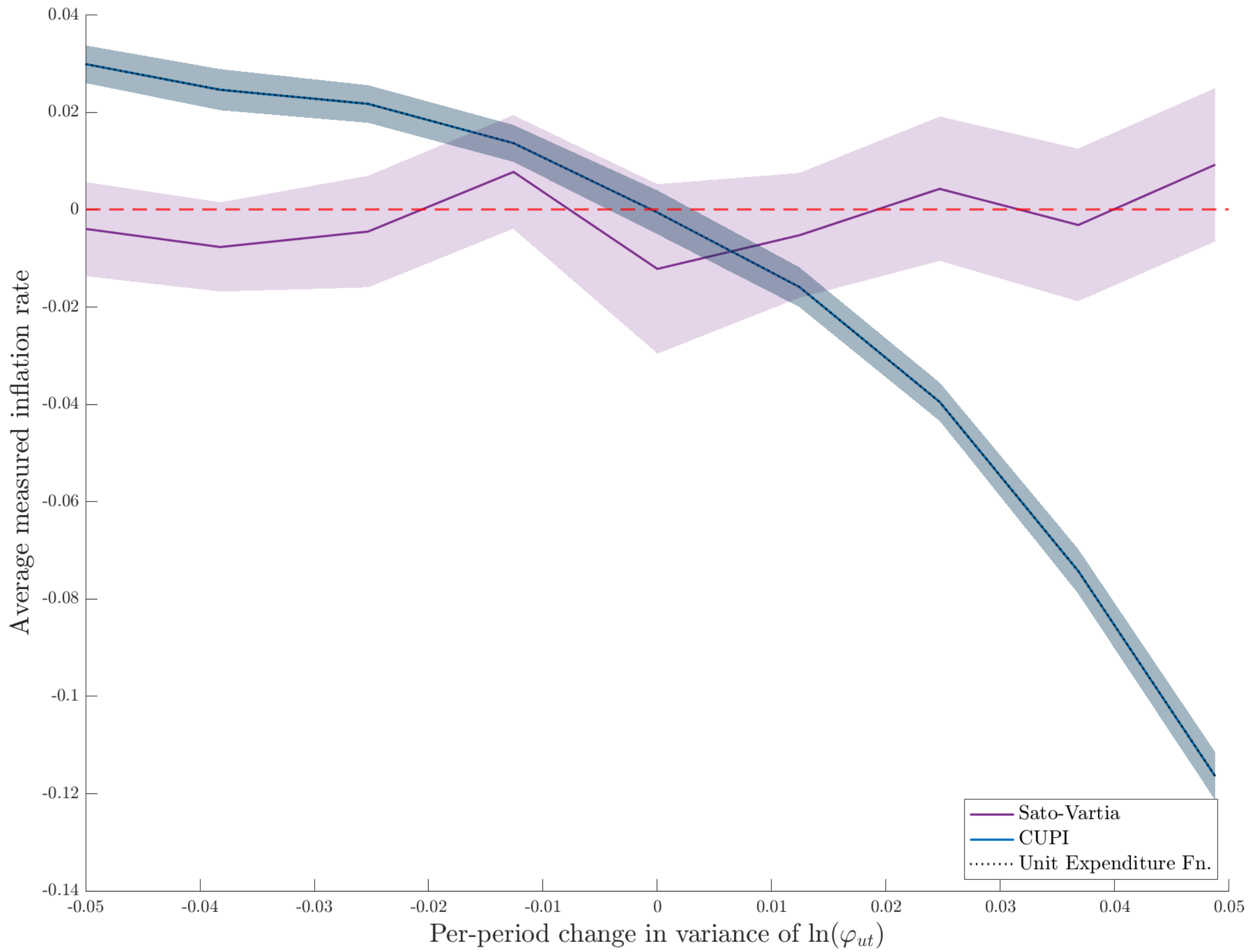
Notes: Price changes are are annual changes from cumulative chained quarterly indices. Price levels reflect cumulative changes relative to the 2006 price level, with 2006 price level set to 1. Figure uses Nielsen Retail Scanner data for nonfood product groups. Laspeyres is geometric.

Figure D.14: Simulated CES Exact Price Indices with Trends in Cost Shifters



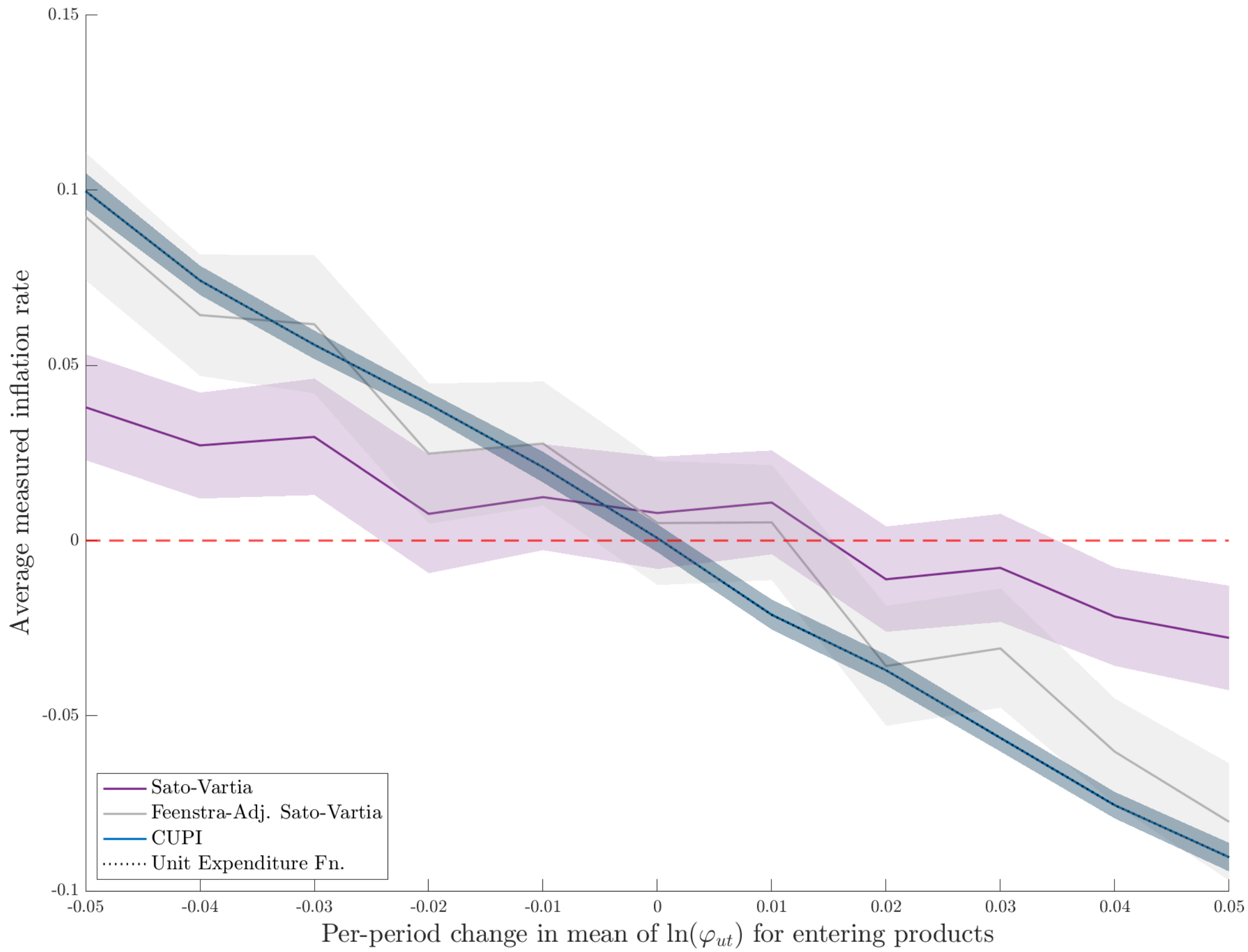
Notes: The figure displays inflation as measured by various CES exact price indices from Monte Carlo simulations of the general equilibrium environment of Hottman et al. (2016). See Section C.2 for simulation details. The horizontal axis displays different average growth rates for the products' marginal 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.15: Simulated CES Exact Price Indices with Trends in Dispersion of Product Appeal



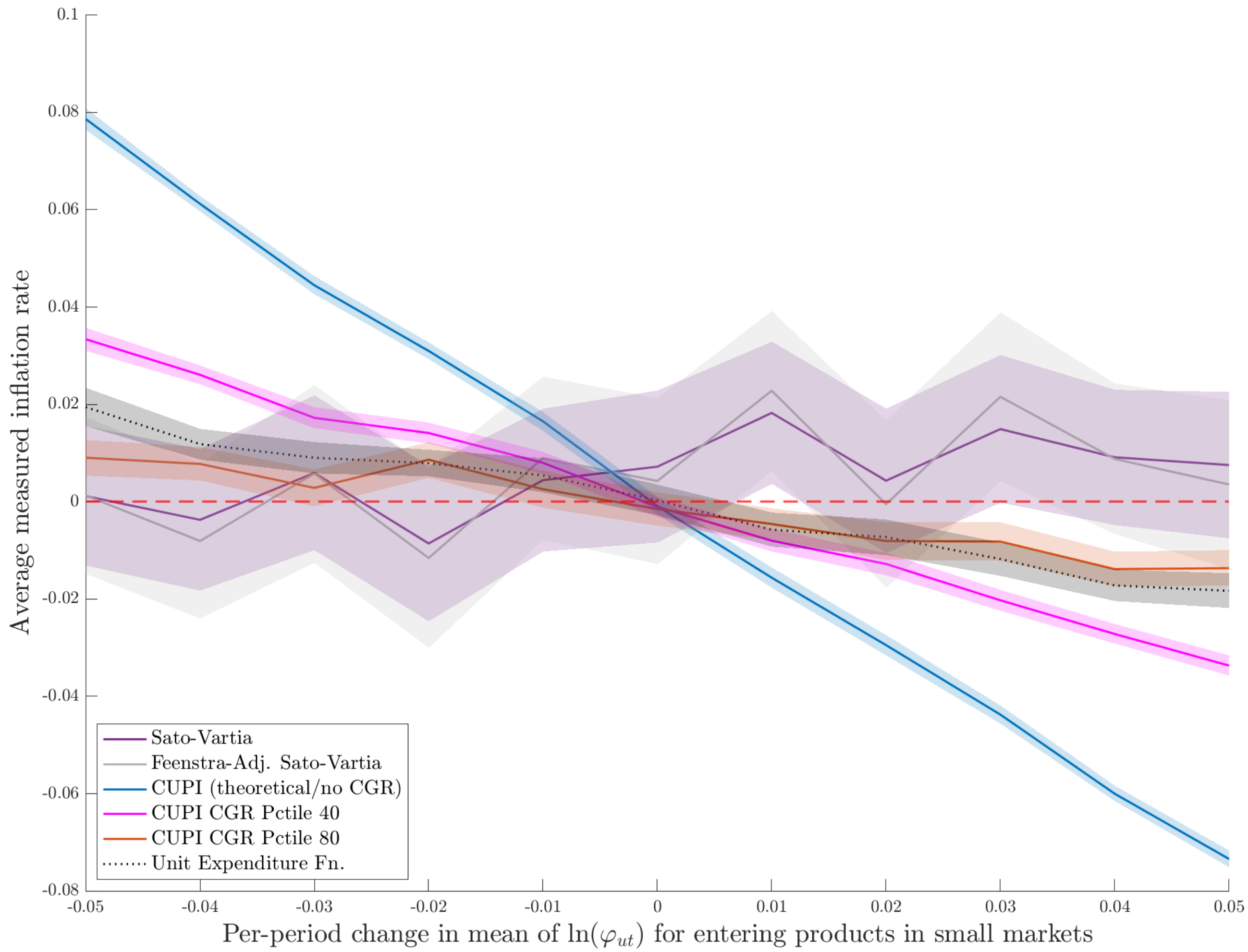
Notes: The figure displays inflation as measured by various CES exact price indices from Monte Carlo simulations of the general equilibrium environment of Hottman et al. (2016). See Section C.2 for simulation details. The horizontal axis displays different average growth rates for the variance of the 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.16: Simulated CES Exact Price Indices with Product Turnover



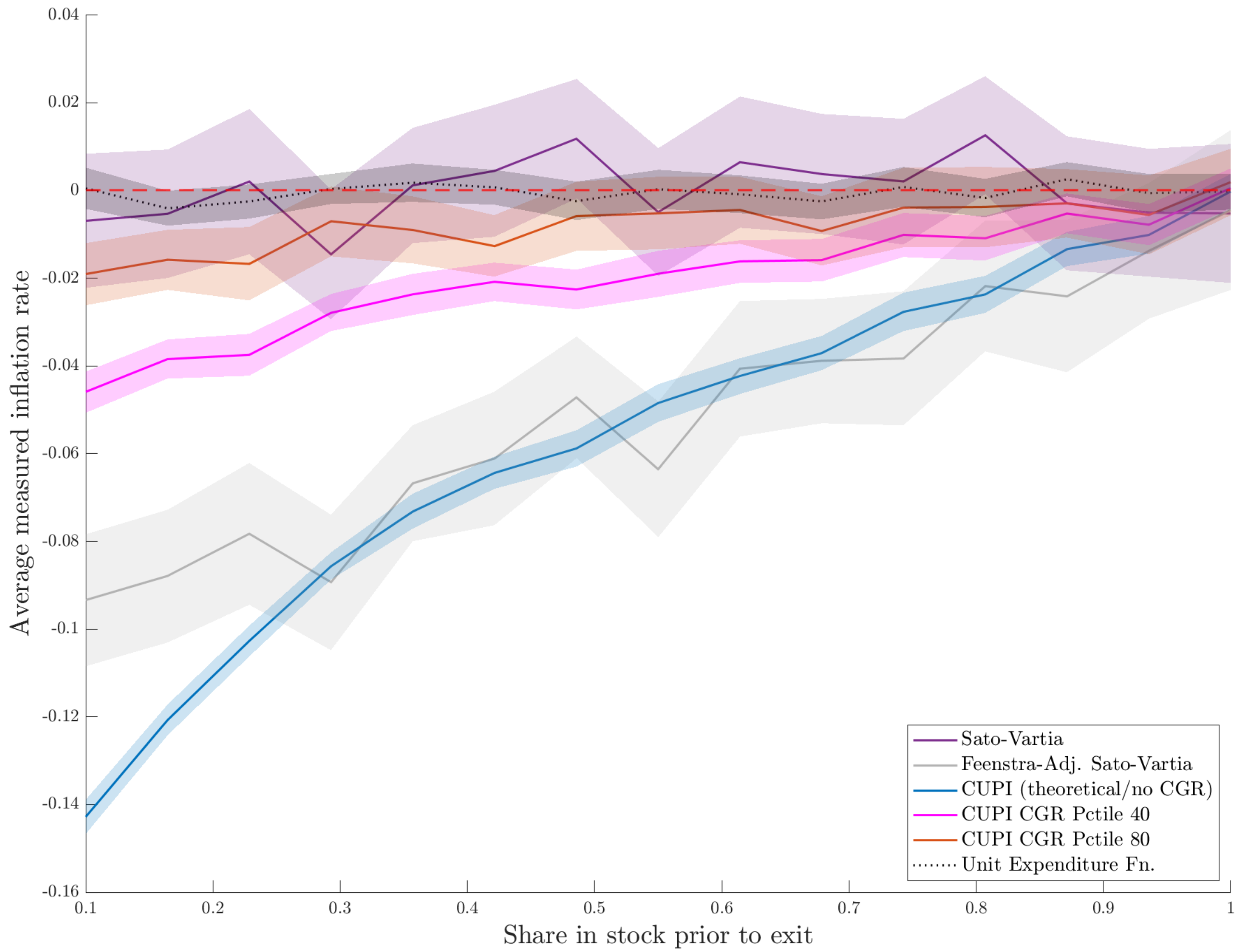
Notes: The figure displays inflation as measured by various CES exact price indices from Monte Carlo simulations of the general equilibrium environment of Hottman et al. (2016). The simulations feature product turnover, with equal numbers of products entering and exiting the market each period. Each product spends five periods in the market. See Section C.2 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.17: 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.2 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.18: 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.2 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.

Table D.1: Hedonic Models: Goodness of Fit

R^2 for:	Log Price Level		Log Price Relative				
Model:	Log-Level		Log-Level		EP1		EP2
Coffee Makers	0.69	0.62	0.09	0.05	0.14	0.20	0.22
Headphones	0.20	0.89	0.04	0.24	0.11	0.43	0.47
Memory Cards	0.65	0.71	0.02	0.05	0.03	0.09	0.13
Work/Occ Footwear	0.58	0.73	0.08	0.10	0.21	0.37	0.39
Boy's Jeans	0.34	0.72	0.08	0.22	0.13	0.43	0.47
Weighted:	n	y	n	y	n	y	y

Average quarterly R^2 for hedonic regression models. For cases where the outcome variable (price level or price relative) does not match the LHS variable from the hedonic model, we report the R^2 from a regression of transformed predicted values from the hedonic model on actual values. For example, the price-relative R^2 for the log-level model is the R^2 from a regression of price relatives constructed from a log-level hedonic model on actual price relatives. Weights used in regressions are consistent in hedonic estimation and construction of R^2 measures. For the log-level model, weights are the quantity shares in the current period. For the log-difference and time-varying unobservables model, weights are the average quantity shares in the current and lagged periods. The time-varying unobservable model includes lagged residuals from a log-level hedonic regression.