

Supplementary Appendix

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

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

American Economic Review

June 2026

A Hedonic Estimation Details: Additional Results with NPD Data

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

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

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

For the time dummy method, we specify the hedonic regression equation (13) using the same vector of characteristics Z_k in each pair of adjacent periods. Occasionally, new features are introduced to the data. In pairs of adjacent periods entirely prior to the introduction of a new characteristic, it will be omitted from the regression because of collinearity with the intercept term. In pairs of adjacent periods in which the new feature is absent during period $t-1$ and present during period t , the feature will be included in the estimated regression. Symmetric arguments apply for characteristics that exit.

Intuitively, the period- t fixed effect δ_t reflects the difference in average price of a “generic” good between $t - 1$ and t because the contributions of all of the product characteristics have been partialled out. The hedonic time dummy specification includes goods entering in period t and exiting after period $t-1$ through its use of the Tornqvist weights, which are

average market shares between the two periods. Nonetheless, a limitation of the time dummy method relative to the EP-TV approach is that the former does not account for unobservable product characteristics. Another issue emphasized by Pakes (2003) and Diewert, Heravi and Silver (2008) is that this method imposes constant coefficients on characteristics in adjacent periods, a restriction that is often rejected by the data.

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

The time dummy method suggests a notable hedonic adjustment for coffee makers relative to the matched-model Tornqvist index, but for other products the difference is modest or is positive rather than negative. Our finding of limited quality adjustment using the time dummy method is broadly consistent with the discussion in Erickson and Pakes (2011). As they emphasize, traditional hedonic approaches cannot account for the changing valuations of unobservable product characteristics, and in particular, how those changing valuations interact with product turnover. For example, if entering goods have desirable unobserved characteristics and correspondingly high prices, then the time dummy method may erroneously suggest a higher index value relative to the matched-model Tornqvist.¹

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

Diewert (2019) favors using expenditure-weights in hedonic estimation based on the argument that this provides more weight on items with more economic importance. He also notes that this approach facilitates comparisons of the time dummy and full imputation hedonic approaches. He acknowledges, however, that his preference for expenditure weighting is based more on index-number issues than econometric issues (footnote 14, p. 6 of Diewert, 2019).² We report sensitivity of results using expenditure-weights in the regression-based EP-TV results for NPD in Table D.3.

We present a large number of statistics including comparisons with related demand-based approaches to quality adjustment in Table D.3. Given this, we focus on the results that incorporate chain drift using the GEKS-Lite procedure of Section III.E. Such results reflect the many different issues including chain drift relevant for comparing alternatives. We

¹For headphones, the matched-model Tornqvist is notably lower in 2016 compared to the Hedonic Tornqvist using the time dummy method. This is a year when the share-weighted average price per item increases substantially. This pattern is consistent with entering goods having higher prices than existing goods. The time dummy method still yields a negative price change in that year, but not as negative as the standard Tornqvist. The hedonic Tornqvist EP-TV method yields a more negative price index than the standard Tornqvist.

²Gorajek (2022) highlights that using expenditure weights yields potential issues with consistency of the estimation given that the expenditure shares are direct functions of the dependent variable (prices). He proposes an alternative transformation of the dependent variable to address these issues.

report means and standard deviations of annual chained price indices as well as correlations. We included for comparison purposes the matched-model Tornqvist and Sato-Vartia indices, the Feenstra index, and the hedonic Tornqvist using the EP-TV approach estimated using both quantity weights and expenditure weights.

We find that using the expenditure-weighted and quantity-weighted approaches yield broadly similar results. For four of the five product groups, the quantity-weighted and expenditure-weighted EP-TV price indices yield lower rates of annual price inflation than the matched-model Tornqvist. The exception is headphones, for which the expenditure-weighted EP-TV mean is slightly above the matched-model Tornqvist. We also report and compare the hedonic Tornqvist indices' differences versus the matched-model indices with the differences between the Sato-Vartia and Feenstra indices. For the latter, all five product groups exhibit a lower rate of inflation for Feenstra than Sato-Vartia. The patterns of these differences are more similar to the Tornqvist and Hedonic Tornqvist, EP-TV using the quantity weighting. As discussed in the main text, we regard the demand-based quality approaches as a relevant benchmark to compare with the hedonic-based indices.³

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

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

B Using the NielsenIQ Data

B.1 NielsenIQ Data Preparation

The NielsenIQ Retail Scanner (RMS) data consists of more than 100 billion unique observations at the week-store-UPC level. We first aggregate the weekly by store data to the monthly frequency according to the NRF calendar and then aggregate the monthly data to quarterly. Following procedures used by Hottman, Redding and Weinstein (2016) and Red-

³One note of caution about making such comparisons is that in principle, the Feenstra (1994) adjustment to the Sato-Vartia index captures pure love of variety effects in addition to quality improvement via entry and exit. The hedonic indices do not feature a direct role for gains from increasing product variety, although if increasing variety brings about lower prices or higher quality, it will affect the hedonic price indices indirectly.

ding and Weinstein (2020), we drop outliers from the weekly by store by UPC data before aggregating to the data to the monthly by UPC frequency. Specifically, we drop observations with prices above 3 times or below one-third the module-level median for each UPC in a given month. We also drop product-month observations with quantities sold that are more than 24 times that product’s median quantity sold per month. One feature of barcoded products is that goods of different sizes and packaging have different barcodes, even if the product contained in the packaging is the same. To ensure comparability between prices, we follow Hottman, Redding and Weinstein (2016) and normalize UPC prices to the same units (e.g., ounces), utilizing the size and packaging information provided by NielsenIQ. Consistent with the literature, we winsorize price and sales changes at the top and bottom 1%. We also winsorize prices at the 99th percentile at the product module by quarter level. Tables D.1 and D.2 show our classification of the product groups in the NielsenIQ Retail Scanner data into Food and Nonfood categories, respectively. Our normalization of units carries over to our measurement of quantities (e.g., all quantities within a product module are measured in consistent units such as ounces). In this appendix, we also use the NielsenIQ Consumer Panel and apply the same procedures from Hottman, Redding and Weinstein (2016) and Redding and Weinstein (2020) to prepare the data. In the next subsection we discuss differences between the Retail Scanner and Consumer Panel data.

B.2 Retail Scanner versus Consumer Panel

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

The Consumer Panel has distinct limitations. First, it is a challenge for participants to accurately record item-level prices from their transactions – the protocol NielsenIQ uses is to use the item-level price from the Retail Scanner data for all purchases from participating retail partners in the Retail Scanner data. Second, the Consumer Panel sample is highly nonrepresentative prior to the application of sample weights: it is dominated by individuals that are older than 55 and are White, Non-Hispanic. For example, in the 2016 Consumer Panel the share of panelists that are 55+ is 56% and are White, Non-Hispanic is 81%. In comparison, Census population statistics for that year show the share of household heads that are 55+ is 42% and are White, Non-Hispanic is 72%. While the projection factors make adjustments to make the weighted sample statistics representative, the projection factors are highly right skewed with a skewness statistic of 1.55. To highlight the potential issues here, the CPS monthly sample (in March 2016) even in raw form is much more representative with the share of household heads that are 55+ at 42% and are White, Non-Hispanic at 71%. Moreover, the CPS sample weights (using household weights) used to make the CPS

representative exhibit essentially no skewness: skewness statistic is -0.05. With the highly skewed projection factors in the Consumer Panel, a small number of observations can have very large influence.

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

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

B.3 Comparisons of the NielsenIQ Data to Official Statistics

This section provides a comparison of both the NielsenIQ Retail Scanner and Consumer Panel with official statistics. Figure D.6 presents comparisons of nominal expenditures for the food and nonfood categories of the NielsenIQ data compared to the Bureau of Economic Analysis (BEA) personal consumption expenditure (PCE) data.⁴ As we reported in the main text, for food, we find nominal sales indices for the NielsenIQ Retail Scanner data tracks the PCE quite closely. The NielsenIQ Consumer Panel tracks the PCE reasonably well through 2012, but it rises less rapidly than either the NielsenIQ Retail Scanner or PCE thereafter. The poorer performance of the Consumer Panel likely reflects the limitations discussed above.

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

Figure D.7 presents the relationship between the BLS CPI and corresponding Laspeyres indices from the NielsenIQ Retail Scanner and Consumer Panel data sets.⁵ We show both arithmetic and geometric Laspeyres. The CPI is a two-stage index with a geometric un-weighted index at the MSA level and arithmetic Laspeyres to the national level. For food, both the NielsenIQ Retail Scanner and Consumer Panel Laspeyres indices are highly correlated with the CPI. In terms of inflation levels, however, the NielsenIQ Retail Scanner more closely matches the CPI. The BLS CPI for Food averages 2.5% over our sample period, the arithmetic Laspeyres for the Retail Scanner averages 3.1% and the arithmetic Laspeyres for

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

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

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

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

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

This section presents sensitivity results to alternative common good rule approaches for both the NielsenIQ Retail Scanner and NielsenIQ Consumer Panel data sets. We consider alternative specifications of the CGR using market thresholds using percentiles of sales pooled over a two-quarter and a five-quarter horizon (that is, the current and prior 4 quarters). For the latter, a common good is defined in this context if it is present in periods t and $t - 4$. Using a duration component in the CGR puts more weight on goods present for the longer horizon, yielding greater comparability with the duration-based CGR used by Redding and Weinstein (2020).

To start, we use the 2-quarter horizon. Figure D.8 shows the results for the aggregated food categories of the CUPI and its components using various CGRs defined by different sales-based percentiles using the Retail Scanner data. In this figure, all series are based on log differences so the decomposition is exact. The Feenstra adjustment and S^* ratio terms have again been scaled by $\frac{1}{\sigma-1}$ for each constituent product group so that the components sum to the CUPI. The Feenstra adjustment (Lambda ratio) and Jevons index (P^* ratio) components of the CUPI show very little sensitivity to the alternative CGRs. Indeed, the plots for the different values are nearly indistinguishable. In contrast, the S^* ratio is very sensitive to the CGR in the NielsenIQ data, which leads directly to sensitivity in the CUPI. The baseline CUPI without a CGR percentile threshold has average four-quarter price inflation about 10 log points below the Feenstra. Using a 50th percentile for the CGR yields a price index that is substantially closer to the Feenstra index.

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

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

To facilitate comparison of our results to Redding and Weinstein (2020), who report pooled results for food and nonfood product groups, Figure D.13 shows various price indices calculated using all product groups in the NielsenIQ Consumer Panel data. The results are broadly consistent with Redding and Weinstein (2020). However, importantly our analysis focuses on chained quarterly annual indices while Redding and Weinstein (2020) focus on year-over-year indices for fourth quarters of each year. In Figure D.14, we show we can closely mimic their results for the CUPI using a market share common goods rule at the 5th percentile if we calculate a year over year (YOY) price index instead of the chained quarterly price indices that have been the focus of this paper. As we have noted in the preceding discussion, the use of a YOY index imparts a duration-based component to the CGR in addition to the expenditure share-based thresholds.⁶

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

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

B.5 Discussing the National Market Assumption for the CUPI

Our empirical implementation of the CUPI relies on the assumption of a unified national market. Redding and Weinstein (2016) note that “Our CES price index is not superlative, because it does not approximate any continuous and differentiable utility function.” Although Redding and Weinstein (2024) show how to aggregate the CUPI in a theoretically consistent manner, the fact that it is not superlative means we cannot rely on the standard “approximate consistency in aggregation” property of superlative indices (Diewert, 1978)

⁶We note that we do not impose a CGR in computing the other price indices shown in Figure D.13. In contrast, Redding and Weinstein (2020) apply the same common goods rule for all of the price indices they display. In unreported analysis, we have found that the Sato-Vartia and Feenstra are not very sensitive to the CGR. This inference is also evident in Figure D.8, which shows that the Lambda ratio is not sensitive to the CGR for NielsenIQ Food data. Because our objective is to compare demand-based indices with the hedonic indices, we aim to treat entry and exit symmetrically across these indices.

⁷Figure D.15 also displays the Bureau of Labor Statistics’ Consumer Price Index for all of the product groups included in the NielsenIQ data as a point of reference.

in its empirical implementation. Deviations from the aggregation structure assumed in the model can have potentially important effects on the measured index. Relatedly, because the CUPÍ includes unweighted geometric mean terms, any failure of this national market assumption has the potential to affect the CUPÍ more than other indices.

The literature on spatial variation in brands (see, e.g., Bronnenberg, Dahr and Dube (2007) and Bronnenberg, Dahr and Dube (2009)) highlights that there are important roles for local, regional and national brands (they focus on consumer packaged goods like those covered by NielsenIQ). Moreover, even for national brands, this analysis shows a slow roll-out of new items geographically. These patterns raise questions about applying a national market based CES price index for most items. In Appendix D.2, we show that the CUPÍ can be badly biased when the national market assumption fails, suggesting these patterns may be important for understanding the empirical behavior of the CUPÍ.

C Additional Details of Machine Learning Algorithm

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

Step 1: Convert text to embeddings. The product descriptions from the NielsenIQ data provided by the Kilts Center for Marketing at the University of Chicago are generally not coded to be human-intelligible. For instance, two product descriptions for soft drinks are: ZR DT LN/LM CF NBP CT and NATURAL R CL NB 12P. A product description for toilet paper is: DR W 1P 308S TT 6PK. (Brand is included in the text fed into the algorithm, though we have suppressed brand name in reporting these examples.)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

D Additional Evidence on the Behavior of the CUPI

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

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

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

approach we assume that price tiers (for example, low-end versus high-end coffee makers) define the substitution structure.

The nested approach requires estimation of elasticities of substitution for products within the same nest and across nests. We follow the approach of Hottman, Redding and Weinstein (2016) to estimate within- and between-nest elasticities for each product group. The within-group estimation uses a modified Feenstra (1994) estimator that double-differences market shares and prices with respect to time and a time-varying nest-level mean.⁸ The between-nest estimator of the elasticity of substitution uses an instrumental variable (IV) approach building on Hottman, Redding and Weinstein (2016).⁹

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

In principle, these within-nest vs. between-nest elasticity estimates could produce significantly different results for the Feenstra index and the CUPI, but in our application the differences are modest. Figure D.9 plots nested versions of the CUPI using our two nesting strategies alongside un-nested versions of the CUPI and Feenstra index. Both versions of the CUPI are implemented using a 30th-percentile CGR, applied at the within-nest level in the nested version.¹⁰ The alternative nesting approaches yield similar results, with the nested CUPI tending to show slightly less deflation than the un-nested (or “flat”) CUPI. In unreported results, we find that the relationship between the nested and flat CUPIs is robust to using alternative CGR cutoffs.

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

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

In this section, we present numerical simulation evidence on the taste shock bias highlighted by Redding and Weinstein (2020). We build on the core insight of Redding and Weinstein (2020) that the sign of the taste shock bias will depend on the sign of the correlation between Sato-Vartia expenditure weights and taste shocks. We start with some simple examples that

⁸The identifying assumption of the Feenstra (1994) estimator is that supply and demand shocks are orthogonal when sales growth and price growth are differenced with respect to a time-varying mean. The Hottman, Redding and Weinstein (2016) assumption is arguably more natural, as differencing with respect to a within-nest mean more plausibly identifies orthogonal supply and demand shocks.

⁹We follow Hottman, Redding and Weinstein (2016) by specifying the between-group relationship between the nest-level price index and expenditure share. The former is endogenous, and Hottman, Redding and Weinstein (2016) overcome this by using variation in the nest-level price index caused by changes in within-nest expenditure share dispersion. We innovate on the procedure of Hottman, Redding and Weinstein (2016) by using the S^* ratio (i.e., changes in common goods expenditure share dispersion) from the within-nest CUPI as the instrument, which removes changes in expenditure-share dispersion induced by product turnover. The identifying assumption is that within-nest demand shocks are uncorrelated with between-nest demand shocks. This innovation integrates the insights of Hottman, Redding and Weinstein (2016) with those of Redding and Weinstein (2020).

¹⁰Nests are weighted by the number of products to ensure differential product group sizes. This ensures that the effective weight assigned to each good is equal to that in the non-nested CUPI, exclusive of the effect of changes to the sample through the different common goods rules.

illustrate cases with positive, negative and no bias. In all of these examples and simulations the CUPPI is the exact price index – with no need for a CGR. This analysis provides useful context for exploring frictions that may underlie the need for a CGR.

D.2.1 Conditional (Non-Stochastic) Examples

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

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

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

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

The consumer’s true cost of living declines from period 1 to period 2 because the dispersion of the product appeal parameters rises. The Sato-Vartia index does not measure this change in the cost of living correctly, because it does not capture changes in the distribution of product appeal. The Sato-Vartia index’s upward bias in this example is consistent with the argument in Redding and Weinstein (2020) that the Sato-Vartia index will tend to be

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

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

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

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

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

Example 3: In our third example, product A again has appeal parameter 2.0 in period one, and product B has appeal parameter 0.5. In period two, the appeal parameters switch places, so that product A has appeal parameter 0.5 and product B has appeal parameter

2.0. In this example, the Sato-Vartia index correctly measures that the consumer’s true cost of living does not change, despite the changing product appeal parameters. In this example, the dispersion of the product appeal parameters did not change across periods. Moreover, the taste-shock bias is zero consistent with the zero correlation between the Sato-Vartia weights and the log appeal shocks.

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

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

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

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

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

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

We note that the realized value of the taste-shock bias will generally be non-zero in these simulations due to the appeal shocks; the CUP’s ability to capture these shocks’ effects

on the unit expenditure function is an important strength even when the expected value of the taste-shock bias is zero. Put differently, the CUPI will capture volatility in the unit expenditure function that is not captured by the Sato-Vartia price index.

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

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

$$\begin{aligned}
 (D.1) \quad & \ln p_{kt} = \mu t + \varepsilon_{kt} \\
 (D.2) \quad & \ln \varphi_{kt} = \nu_{kt} \\
 (D.3) \quad & \begin{bmatrix} \varepsilon_{kt} \\ \nu_{kt} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \chi_p^2 & \rho_{p\varphi} \\ \rho_{p\varphi} & \chi_\varphi^2 \end{bmatrix} \right).
 \end{aligned}$$

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

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

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

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

D.2.2.2 Baseline Simulations with AR(1) Prices and Appeal Figure D.18 shows that the same pattern for the taste-shock bias extends to the case in which the log prices and log appeal parameters follow $AR(1)$ processes. Consumers again have CES preferences

with elasticity of substitution $\sigma = 5$. The data generating process for prices and appeal parameters is now given by

$$(D.4) \quad \ln p_{kt} = \eta \ln p_{kt-1} + \varepsilon_{kt}$$

$$(D.5) \quad \ln \varphi_{kt} = \eta \ln \varphi_{kt-1} + \nu_{kt}$$

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

We consider different rates of persistence in the $AR(1)$ processes η and different degrees of correlation between price and appeal draws $\rho_{p\varphi}$ along the x-axis.¹¹ We set $\chi_p = \chi_\varphi = 0.2$ across all of the simulations, which again feature 1,000 individual products. This set of simulations features 36 time periods, as opposed to the 12 in the baseline simulations, to reduce the volatility introduced by the persistence in the $AR(1)$ processes. We run 50 simulations for each combination of parameters η and $\rho_{p\varphi}$.

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

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

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

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

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

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

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

¹¹To keep the number of parameters to be displayed manageable, we do not consider non-zero intercepts in the process for log prices, and we restrict the degree of persistence in the processes for prices and appeal parameters to be equal. We draw log prices and appeal parameters from their implied unconditional stationary distribution in the first period of each simulation.

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

The results in Figure D.20 reinforce the illustrative non-stochastic examples in section D.2.1: the Sato-Vartia index has an unconditional upward bias when the dispersion of product appeal shocks is rising over time but an unconditional downward bias when the dispersion of product appeal shocks is falling. The CUPI, in contrast, accurately captures these trends’ effects on the consumer’s cost of living in the absence of other frictions.¹²

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

D.2.3.2 Unit Root in Product Appeal This set of simulations explores the taste-shock bias in a baseline setting with no frictions but with a unit root in the data generating process for product appeal parameters. The simulation environment is the same as in Section D.2.2.1, but the data generating process for price and appeal parameters is now

$$\begin{aligned}
 (D.8) \quad & \ln p_{kt} = \varepsilon_{kt} \\
 (D.9) \quad & \ln \varphi_{kt} = \ln \varphi_{kt-1} + \nu_{kt} \\
 (D.10) \quad & \begin{bmatrix} \varepsilon_{kt} \\ \nu_{kt} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \chi_p^2 & 0 \\ 0 & \chi_\varphi^2 \end{bmatrix} \right).
 \end{aligned}$$

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

Figure D.23 displays the correlations of the Sato-Vartia weights and the log appeal shocks

¹²In unreported results, we find that a similar pattern holds in the presence of time-varying correlations between the price and appeal draws. The Sato-Vartia index has an upward unconditional bias relative to the unit expenditure function when this correlation is falling over time and a downward unconditional bias when this correlation is rising over time.

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

D.2.4 Empirical Evidence

We examine the empirical relevance of rising dispersion in product appeal in Figure D.10. The figure plots the measured dispersion (standard deviation) in normalized product appeal for each of the NPD product groups. We find evidence of rising relative product appeal dispersion for memory cards and headphones. This increase is more pronounced without a common goods rule. For other goods, we observe a nonmonotonic change in dispersion even when the S^* ratio is highly negative (see, e.g., boys' jeans in Figure 5). In other words, rising dispersion of product appeal does not obviously drive the CUPi's ubiquitous finding of rapid deflation without using a common goods rule.

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

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

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

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

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

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

$$(D.11) \quad \ln p_{kt}^{\text{large}} = \varepsilon_{kt}^{\text{large}}$$

$$(D.12) \quad \ln \varphi_{kt}^{\text{large}} = \nu_{kt}^{\text{large}}$$

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

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

$$(D.14) \quad \ln p_{kt}^{\text{small}} = \varepsilon_{kt}^{\text{small}}$$

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

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

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

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

The assumptions of market segmentation in these simulations are meant to evoke the spatial variation in brands documented in Bronnenberg, Dahr and Dube (2007) and Bronnenberg, Dahr and Dube (2009).

Figure D.24 shows that the CUPi is downward biased when the entering products in the small market have rising appeal over time. The intuition for the bias in the CUPi is that its P^* and S^* ratio terms are unweighted geometric means. The theoretical CUPi therefore assigns the price movements in the small markets, driven by product turnover, equal impor-

tance to the price movements in the large market. Although that equal weighting scheme would be theoretically justified in a unified marketplace under CES preferences with appeal shocks, it implicitly overweights the small markets in the segmented market environment. A CGR reduces the bias in the theoretical CUPI by reallocating products from the unweighted geometric mean terms to the lambda ratio term, which is weighted.

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

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

$$(D.17) \quad \ln p_{kt} = \varepsilon_{kt}$$

$$(D.18) \quad \ln \varphi_{kt} = \nu_{kt}$$

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

The simulations feature a stylized version of stock-outs, or a “clearance rack,” in which product sales are rationed in the period before they exit the marketplace. Products’ expenditure shares in the period prior to exit are set equal to a given fraction of their level in a frictionless CES demand system. We assume that consumers optimally reallocate their demands toward the unconstrained products in response to the rationing.¹³ We set $\chi_p = \chi_\varphi = 0.2$ in these simulations, and consider different “in-stock proportions” for shortly exiting goods along the x-axis of the figure.

The unit expenditure function in Figure D.25 shows that stock-outs do not have a meaningful effect on the consumer’s cost of living, because there is no trend in the average appeal of goods over time. Likewise, the Sato-Vartia index is not meaningfully affected, because goods that exit from periods $t-1$ to t are excluded. Yet stock-outs introduce a substantial bias to both the CUPI and the Feenstra index. The intuition for the bias is subtle. Stock-

¹³The assumption that consumers have homothetic CES preferences makes it straightforward to calculate their re-optimized demands in the presence of rationing; consumers reallocate their expenditure to each of the non-rationed goods in proportion to their unconstrained demands had there been no rationing. The unit expenditure function under rationing can then be computed as the ratio of indirect utilities provided by a unit of expenditure between periods.

outs lower expenditure shares on goods just prior to their exit from the marketplace, with the expenditure reallocated to unconstrained goods. Stock-outs therefore raise the dispersion of expenditure shares on continuing goods relative to the frictionless case, leading to a negative log S^* ratio. The Feenstra adjustment to the Sato-Vartia index is also negative, because new goods enter the market un-rationed, allowing consumers to buy whatever quantities they please; prior to exit, quantities are constrained below consumers' desired levels. The expenditure share on exiting products is therefore lower than the expenditure share on entering products. The Feenstra index and the CUPI are therefore both downward-biased, with the bias being larger for the CUPI.

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

D.3 Analytical Characterization of the Taste Shock Bias

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

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

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

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

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

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

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

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

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

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

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

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

The unconditional expected value of the taste shock bias is therefore

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

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

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

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

We can write the Sato-Vartia weights as

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

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

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

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

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

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

$$(D.27) \quad \omega(p_{kt}, p_{kt-1}, \varphi_{kt}, \varphi_{kt-1}) = \omega(p_{kt-1}, p_{kt}, \varphi_{kt-1}, \varphi_{kt}),$$

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

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

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

Likewise, we can write

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

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

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

it is clear that the definitions in equations (D.28) and (D.29) must be equal.

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

References

- Bajari, Patrick, Zhihao Cen, Victor Chernozhukov, Manoj Manukonda, Jin Wang, Ramon Huerta, Junbo Li, Ling Leng, George George Monokroussos, Suhas Vijaykumar, and Shan Wan. 2021. “Hedonic Prices and Quality Adjusted Price Indices Powered by AI.” CENMAP Working Paper 04-21; forthcoming, *Journal of Econometrics*.
- Blundell, Richard, and Stephen Bond. 1998. “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models.” *Journal of Econometrics*, 87(1): 115–143.
- Broda, Christian, and David Weinstein. 2010. “Product Creation and Destruction: Evidence and Price Implications.” *American Economic Review*, 100(3): 691–723.
- Bronnenberg, Bart, Sanjay Dahr, and Jean-Pierre Dube. 2007. “Consumer Packaged Goods in the US: National Brands, Local Branding.” *Journal of Marketing Research*, 44(1): 4–13.
- Bronnenberg, Bart, Sanjay Dahr, and Jean-Pierre Dube. 2009. “Brand History, Geography, and the Persistence of Brand Shares.” *Journal of Political Economy*, 117(1): 87–115.
- Cafarella, Michael, Gabriel Ehrlich, Tian Gao, John C Haltiwanger, Matthew D Shapiro, and Laura Zhao. 2023. “Using Machine Learning to Construct Hedonic Price Indices.” NBER Working Paper 31315.
- Diewert, Erwin. 1978. “Superlative Index Numbers and Consistency in Aggregation.” *Econometrica*, 46(4): 883–900.
- Diewert, Erwin. 2019. “Quality Adjustment and Hedonics: A Unified Approach.” University of British Columbia Vancouver School of Economics Discussion Paper 19-01.
- Diewert, Erwin, Saeed Heravi, and Mick Silver. 2008. “Hedonic Imputation versus Hedonic Time Dummy Indexes.” NBER Working Paper 14018.
- Einav, Liran, Ephraim Leibtag, and Aviv Nevo. 2010. “Recording Discrepancies in Nielsen Homescan data: Are They Present and Do They Matter?” *Quantitative Marketing and Economics*, 8: 207–239.
- Erickson, Tim, and Ariel Pakes. 2011. “An Experimental Component Index for the CPI: From Annual Computer Data to Monthly Data on Other Goods.” *American Economic Review*, 101(5): 1707–1738.
- Feenstra, Robert C. 1994. “New Product Varieties and the Measurement of International Prices.” *The American Economic Review*, 84(1): 157–177.
- Gorajek, Adam. 2022. “Generalizing the Stochastic Approach to Price Indices.” *Review of Income and Wealth*, 70(1): 1–22.

- Hochreiter, Sepp, and Jürgen Schmidhuber.** 1997. “Long Short-Term Memory.” *Neural Comput.*, 9(8): 1735–1780.
- Hottman, Colin J, Stephen J Redding, and David E Weinstein.** 2016. “Quantifying the sources of firm heterogeneity.” *The Quarterly Journal of Economics*, 131(3): 1291–1364.
- Ivancic, Lorraine, Erwin Diewert, and Kevin J. Fox.** 2011. “Scanner Data, Time Aggregation and the Construction of Price Indexes.” *Journal of Econometrics*, 161(1): 24–35.
- Kingma, Diederik P, and Jimmy Ba.** 2014. “Adam: A method for stochastic optimization.” *arXiv preprint arXiv:1412.6980*.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean.** 2013. “Efficient estimation of word representations in vector space.” *arXiv preprint arXiv:1301.3781*.
- Pakes, Ariel.** 2003. “A Reconsideration of Hedonic Price Indexes With an Application to PC’s.” *American Economic Review*, 93(5): 1578–1596.
- PyTorch Contributors.** 2022. “PyTorch.”
- Redding, Stephen J., and David E. Weinstein.** 2016. “Measuring Aggregate Price Indexes with Demand Shocks: Theory and Evidence for CES Preferences.” NBER Working Paper 22479.
- Redding, Stephen J., and David E. Weinstein.** 2020. “Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences.” *The Quarterly Journal of Economics*, 135(1): 503–560.
- Redding, Stephen J., and David E. Weinstein.** 2024. “Accounting for Trade Patterns.” *Journal of International Economics*, 150: 103910.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

	Memory Cards	Coffeemakers	Headphones	Boys' Jeans	Occupational Footwear
Tornqvist (Chained)	-16.89	-8.86	-11.58	-7.63	-3.35
Tornqvist (GEKS-Lite)	-15.41	-6.64	-11.55	-5.55	-2.31
Tornqvist (Rolling-Year GEKS)	-15.57	-6.69	-9.12	-4.77	-1.96

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

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

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

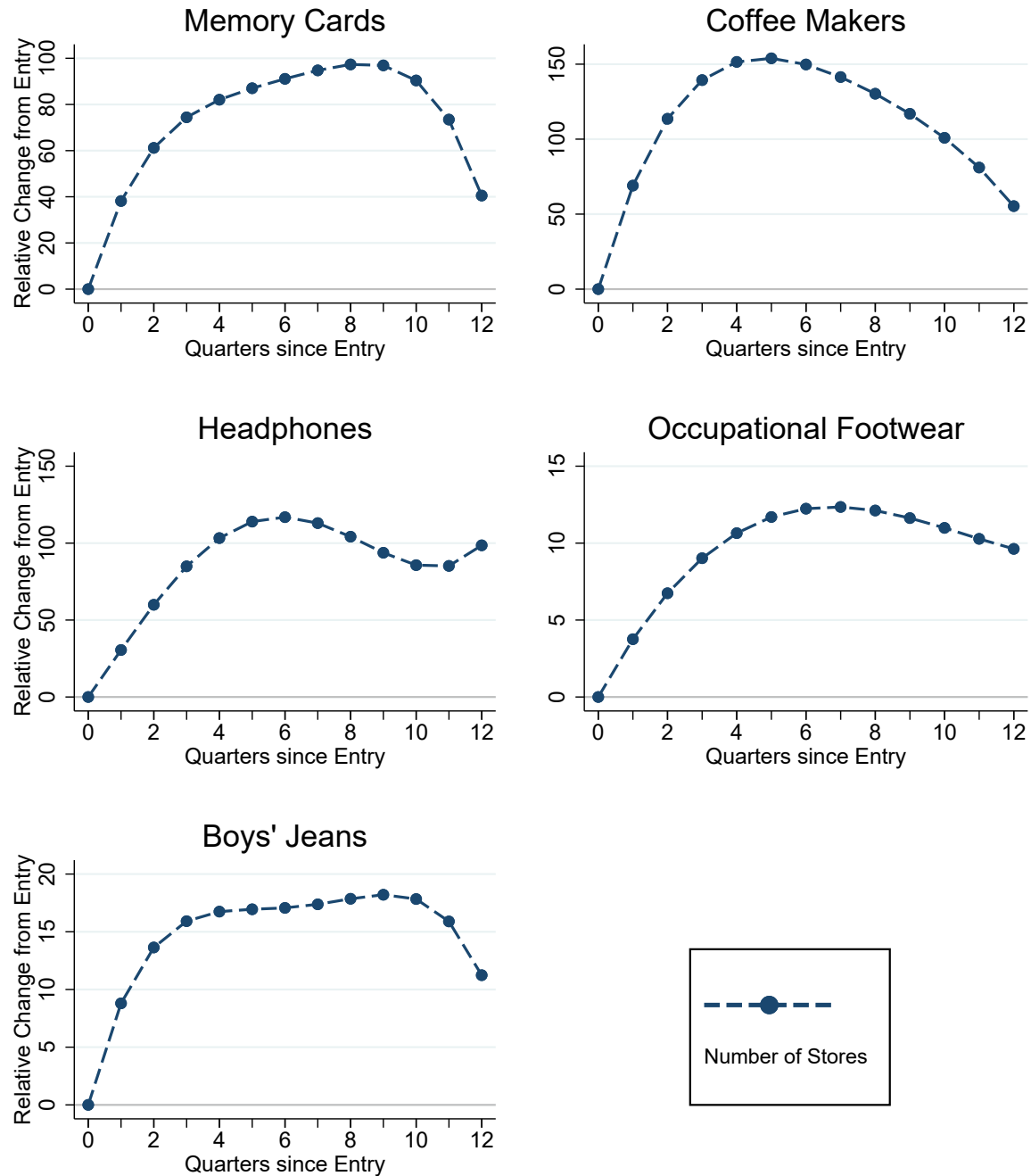
Notes: Estimated elasticities of substitution for nested CES models. Standard errors in parentheses. Data come from NPD Group. Within-nest elasticities are estimated using the methodology of Feenstra (1994) and Redding and Weinstein (2020). Across-nest elasticities are estimated using the nested CES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

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

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

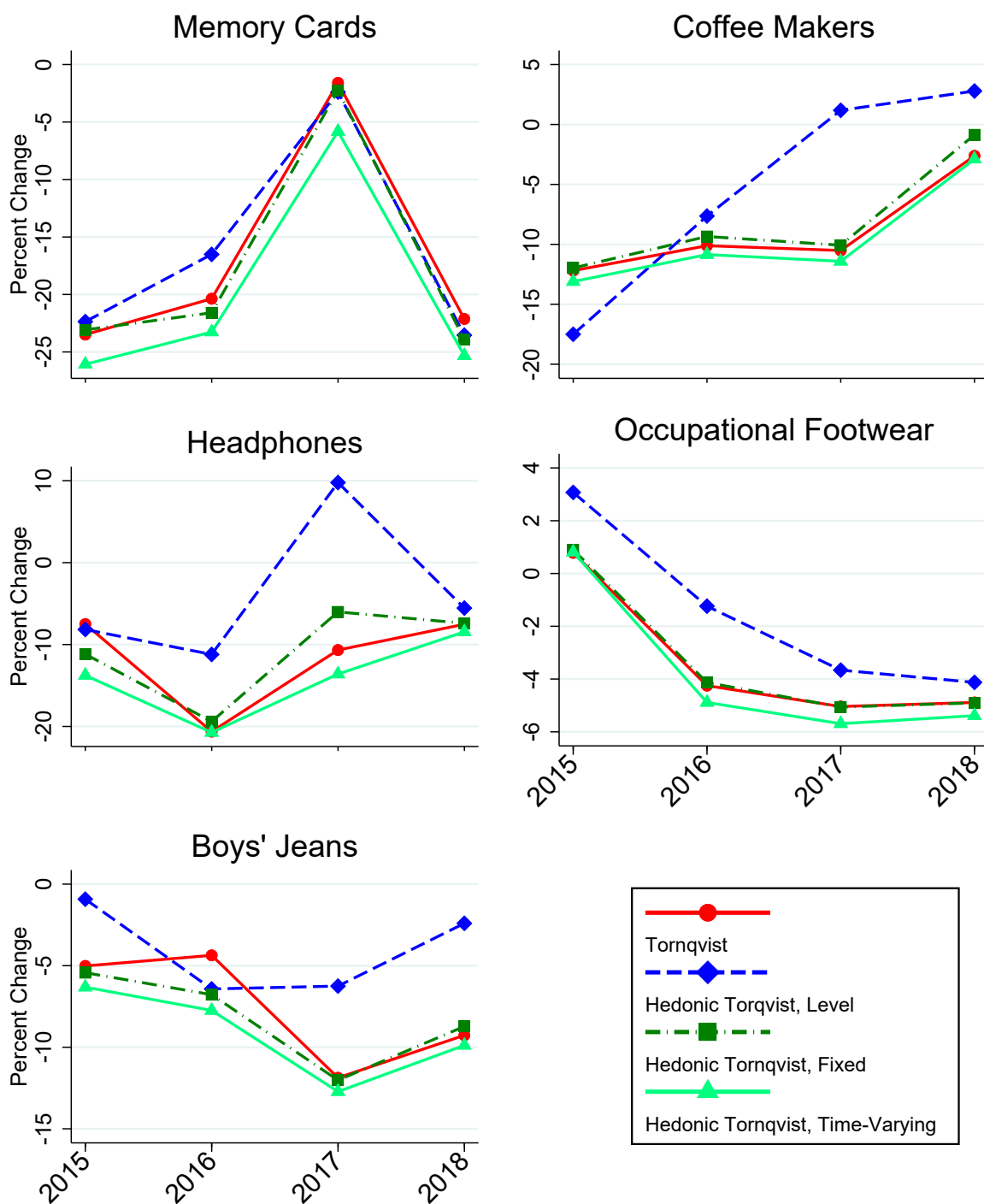
Notes: Regression of current relative quality on lagged relative quality from CUPi model. Standard errors in parentheses. Blundell-Bond allows for item-level fixed effects. Data come from NPD Group.

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



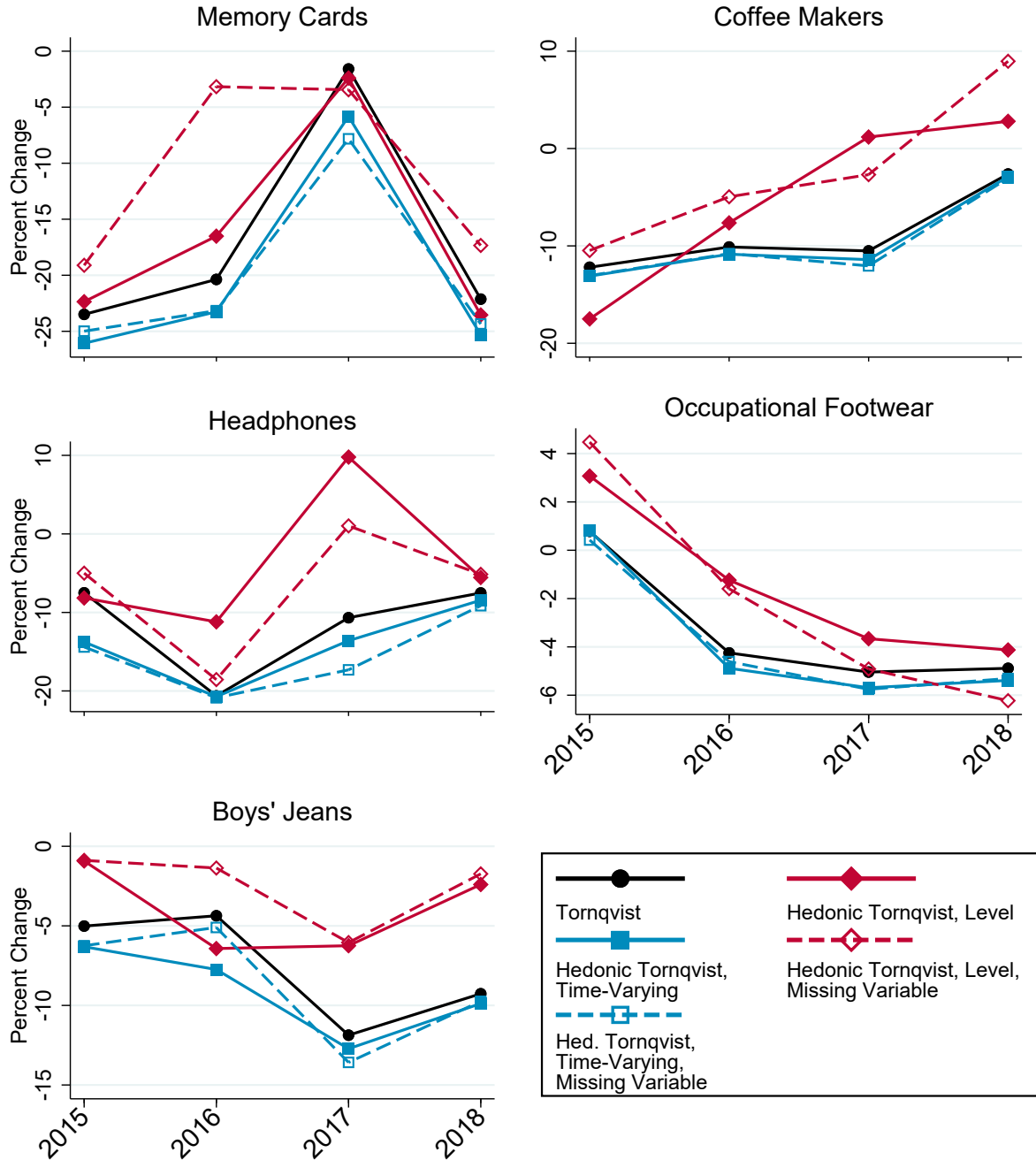
Notes: Number of stores relative to their value in the period of their initial entry. Entry occurs in period 0. All series are smoothed with a quartic spline.

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



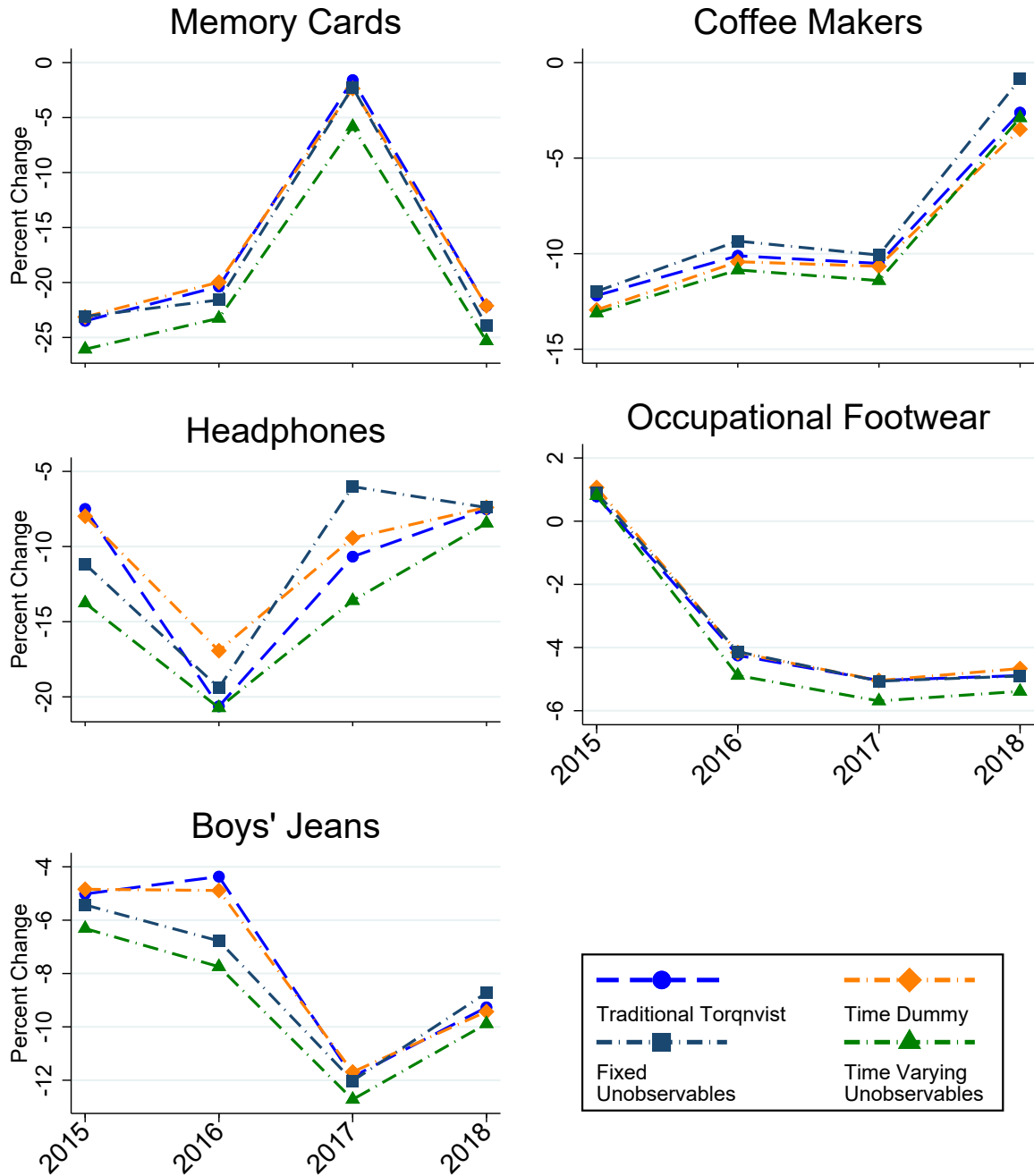
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices.

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



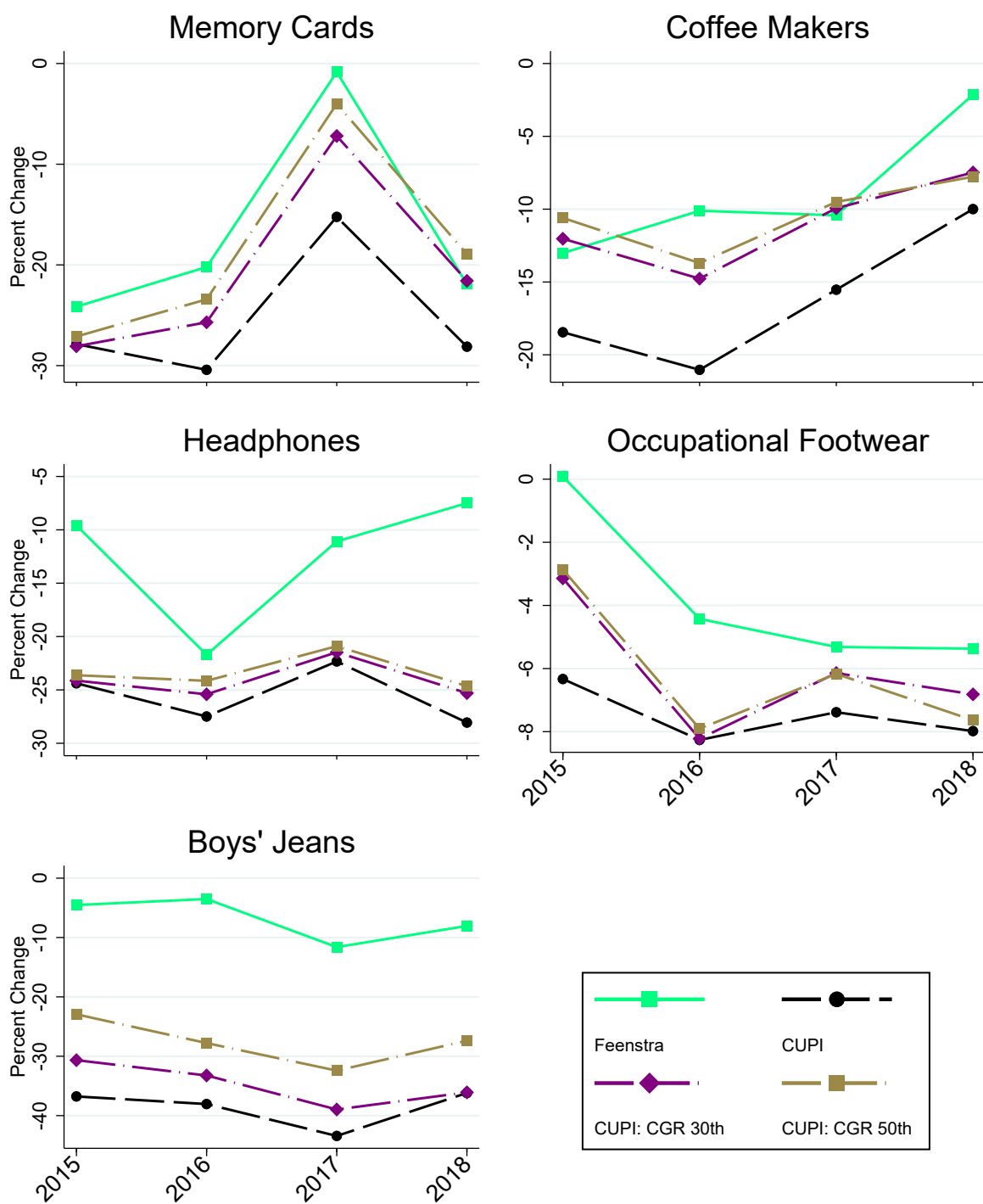
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices.

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



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

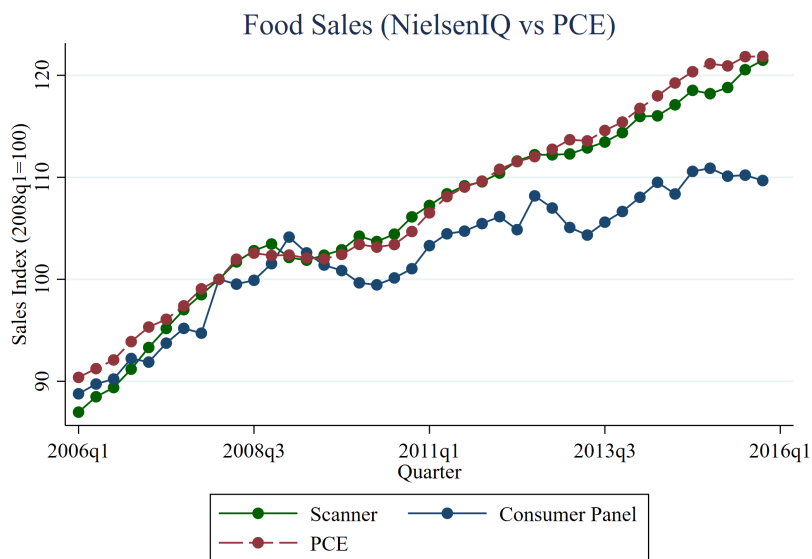
Figure D.5: CUPI, Sensitivity to Alternative Common Goods Rules (CGRs): NPD Data



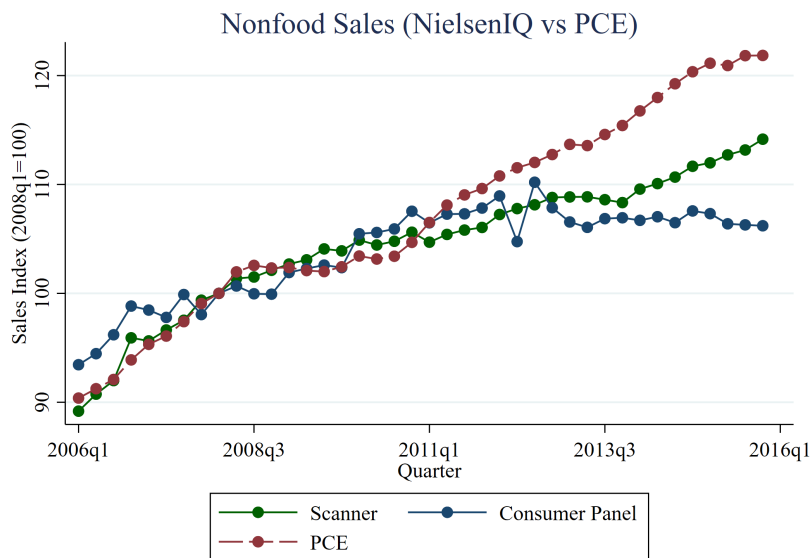
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly price indices. “CUPI: CGR Xth” excludes from the group of common goods those products with market shares below the Xth percentile in both periods. The Feenstra index is included for reference.

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

(a) Food

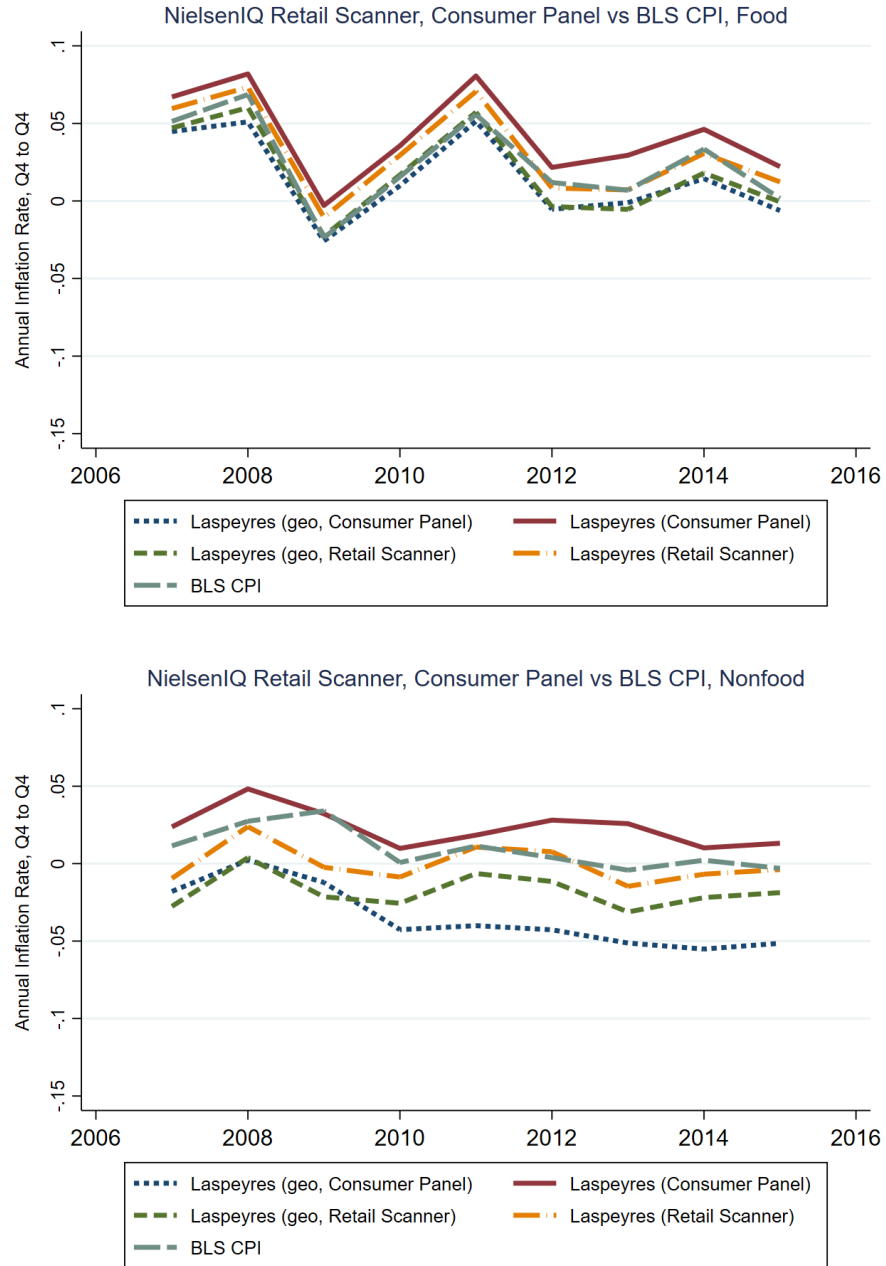


(b) Nonfood



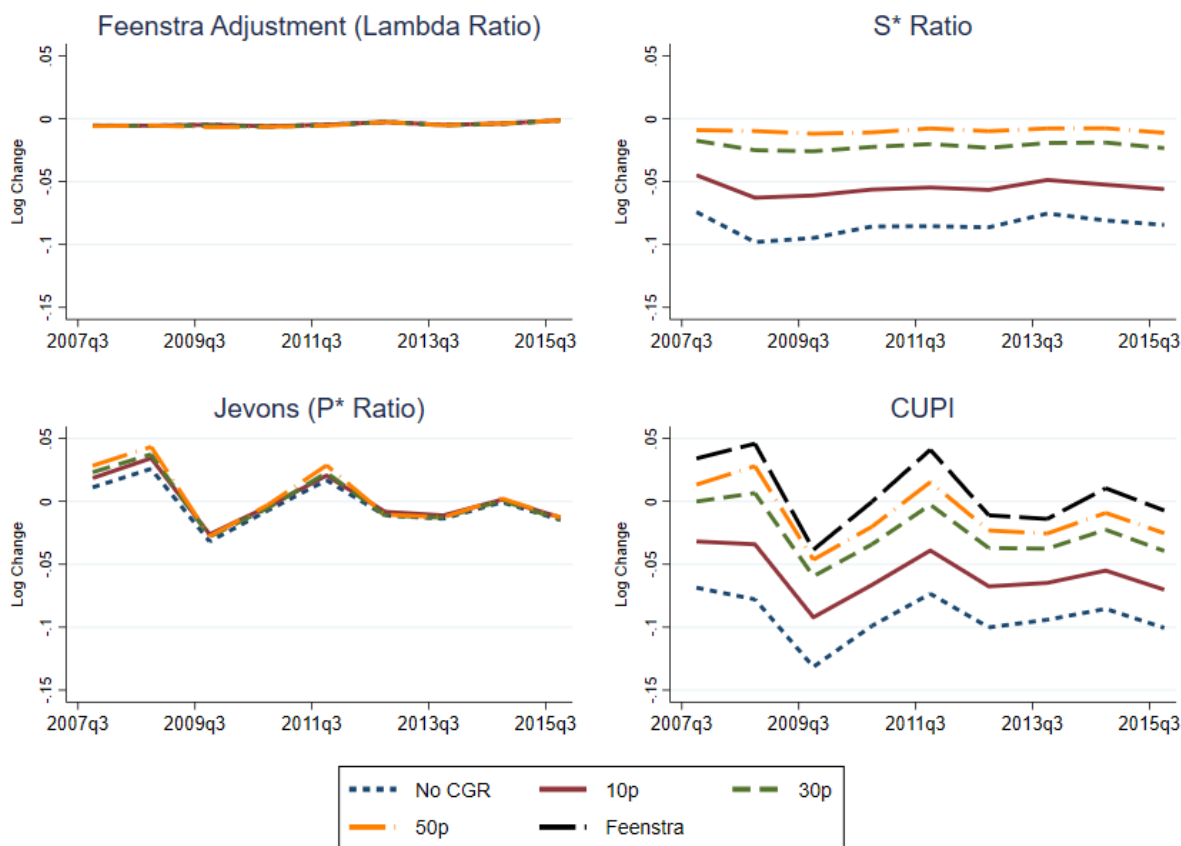
Notes: NielsenIQ values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. Figures uses NielsenIQ Retail Scanner and Consumer Panel data for Food and Nonfood (aggregated) product groups. PCE is personal consumption expenditures (nominal) from BEA. All series indexed to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

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



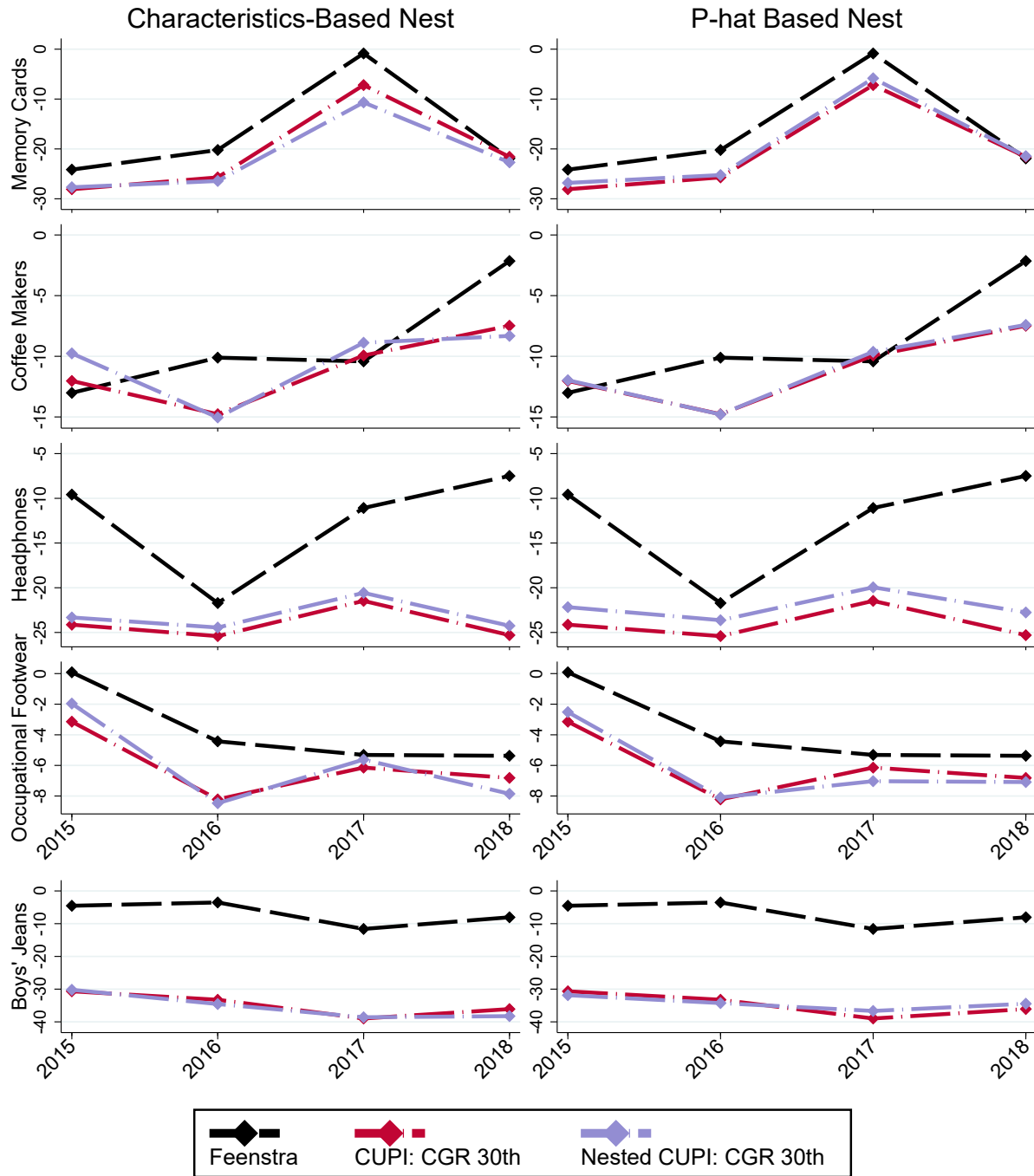
Notes: NielsenIQ values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. The panels use NielsenIQ Retail Scanner and Consumer Panel data for Food and Nonfood (aggregated) product groups. The BLS CPI was computed by BLS staff.

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



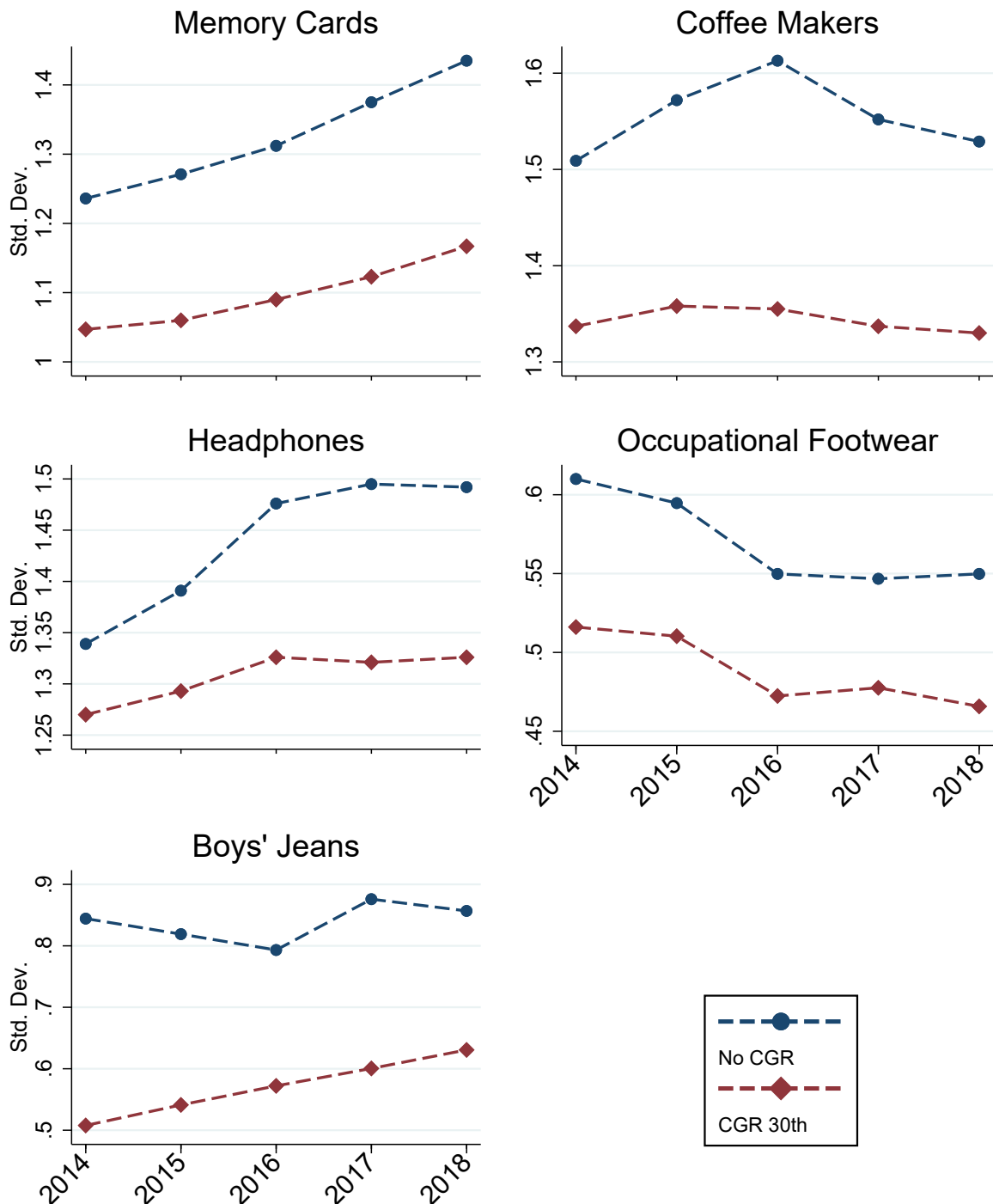
Notes: The figure shows NielsenIQ Retail Scanner data for food product groups. Each plot shows log changes from the fourth quarter of the previous year to the fourth quarter of the labeled year. The values are cumulative changes from chained quarterly indices. The Feenstra adjustment and S^* ratio panels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

Figure D.9: Nested CUPI: Characteristics- and P-Hat- Based Nests
Percent Changes: NPD Data



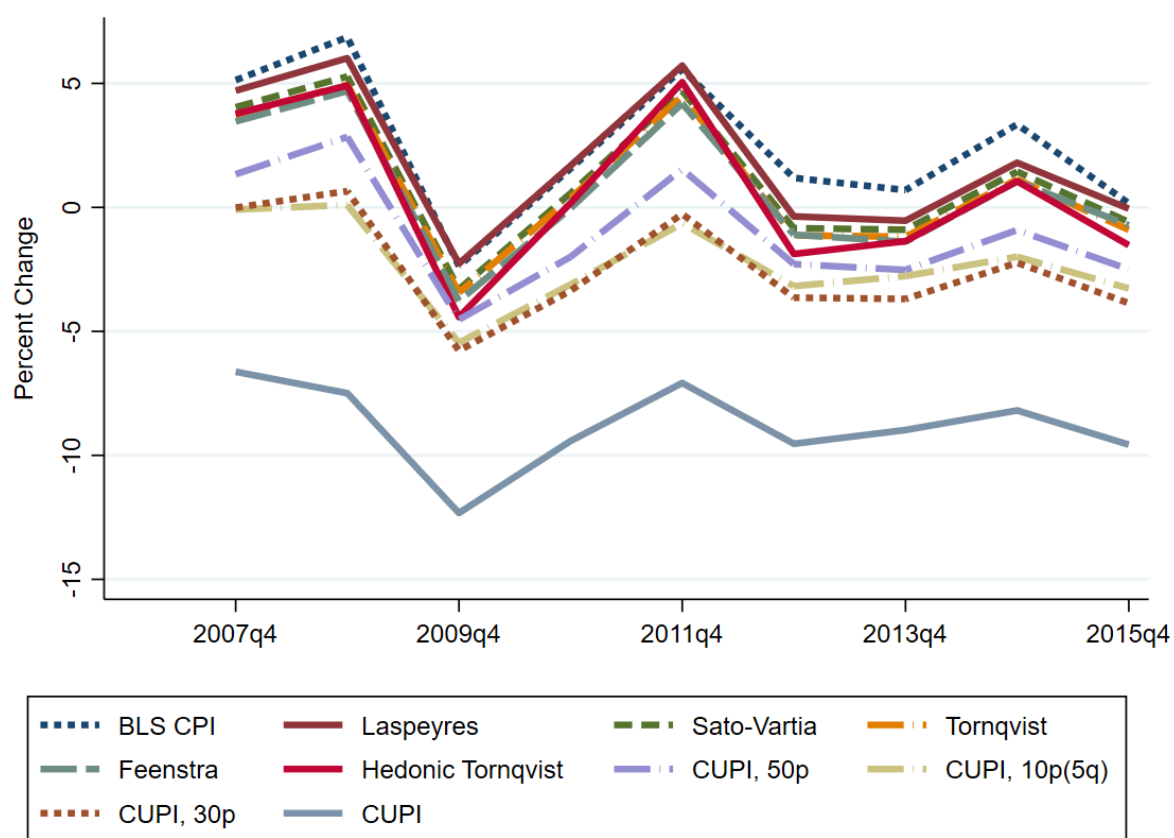
Notes: Values are percent change on a q4-to-q4 basis, aggregated from chained quarterly indices. Laspeyres is the geometric Laspeyres. For the characteristics-based nests, we assign items to groups based on shared observable characteristics. The p-hat based nests are based on the decile of predicted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

Figure D.10: Dispersion of Relative Product Appeal: NPD Data



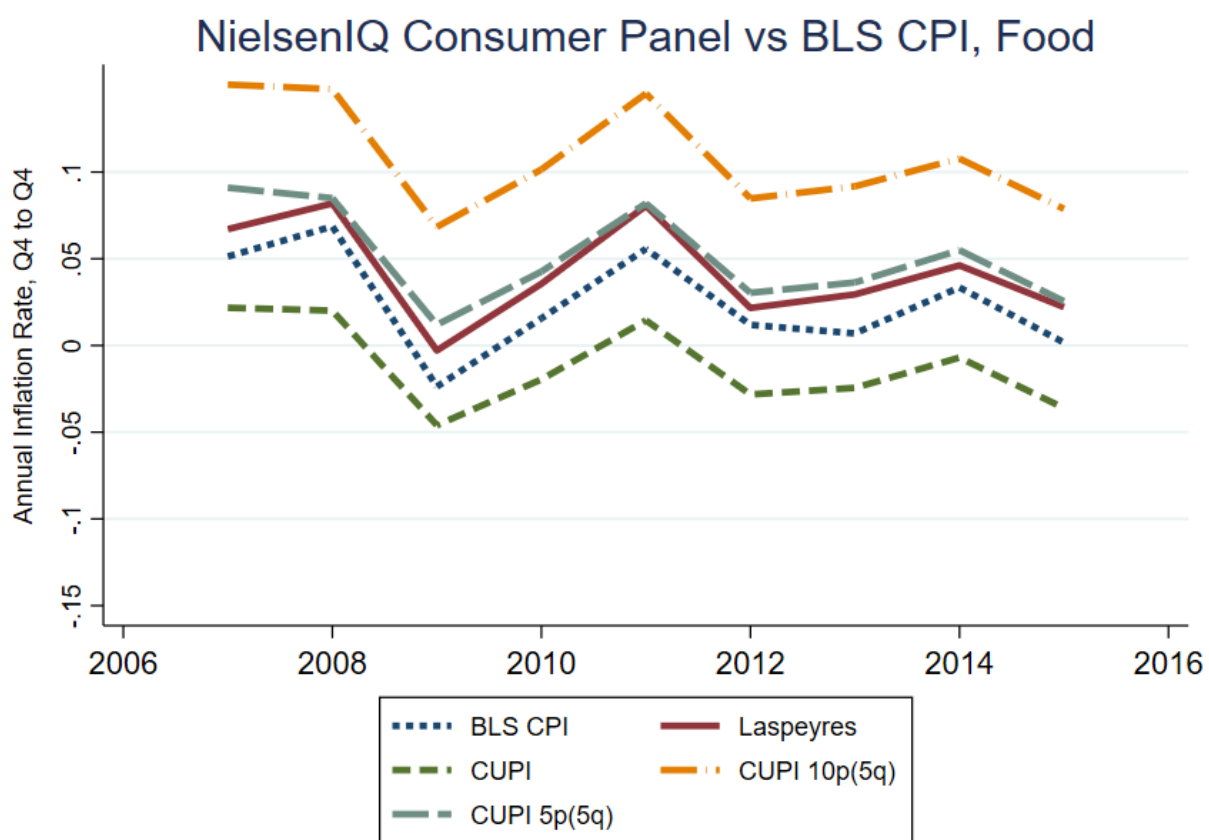
Notes: Values are annual averages of quarterly dispersion (standard deviation) in normalized log relative product appeal for common goods. Reported are the annual averages without imposing a common goods rule and also those with a 30th percentile common goods rule.

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



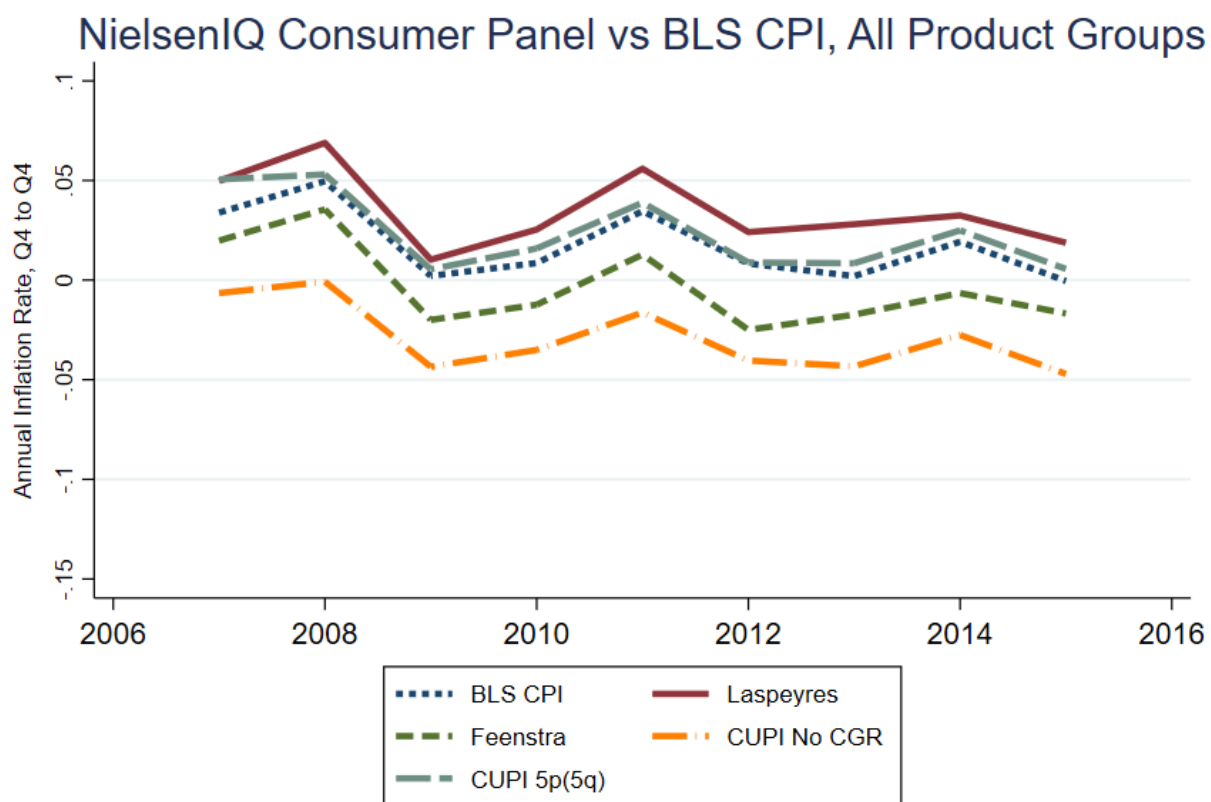
Notes: Figure uses NielsenIQ Retail Scanner data for food. The 2q CUPI computes CGR percentile thresholds using sales pooled over a two quarter horizon (t and $t - 1$). The 5q CUPI computes CGR percentile thresholds using sales pooled over a 5 quarter horizon (current and prior 4 quarters). Laspeyres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

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



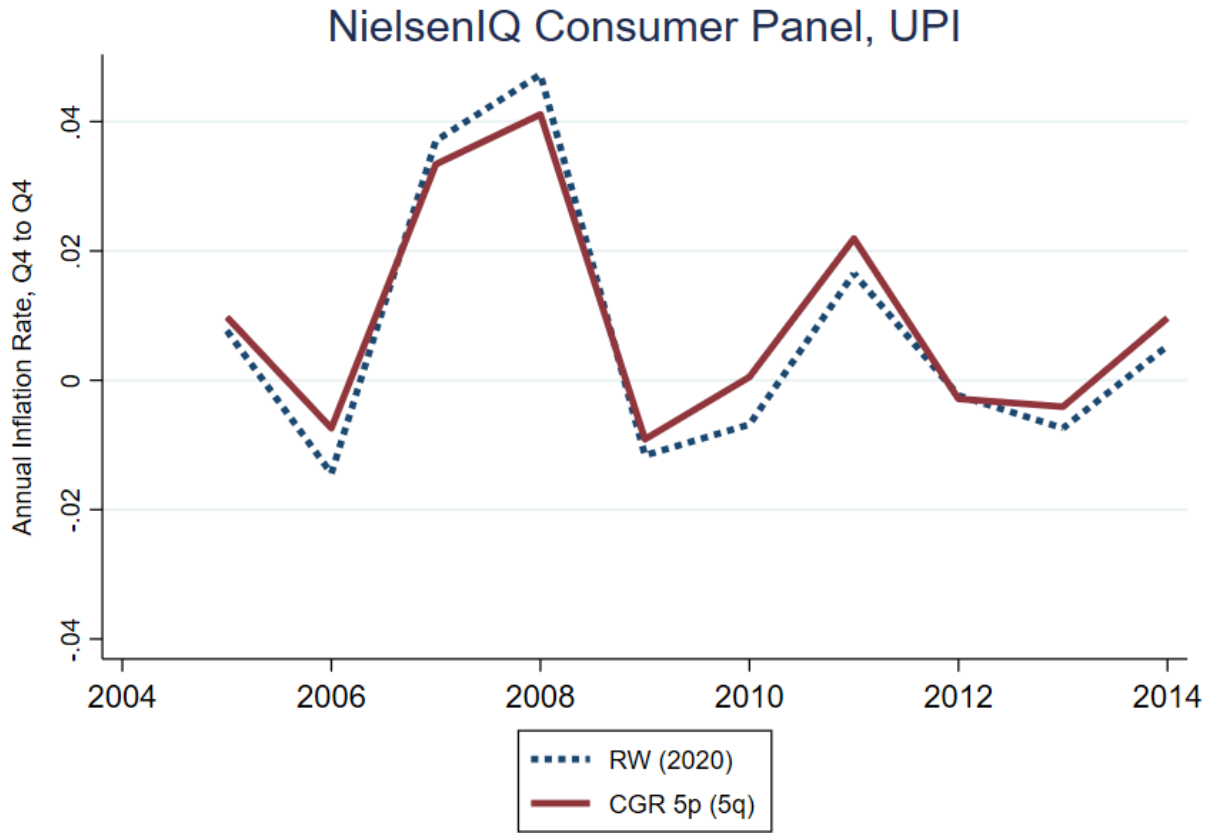
Notes: Figure uses NielsenIQ Consumer Panel data for food. The 5q CUPI computes CGR percentile thresholds using sales pooled over a five quarter horizon (t and $t-1$) (current and prior 4 quarters). Laspeyres is arithmetic.

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



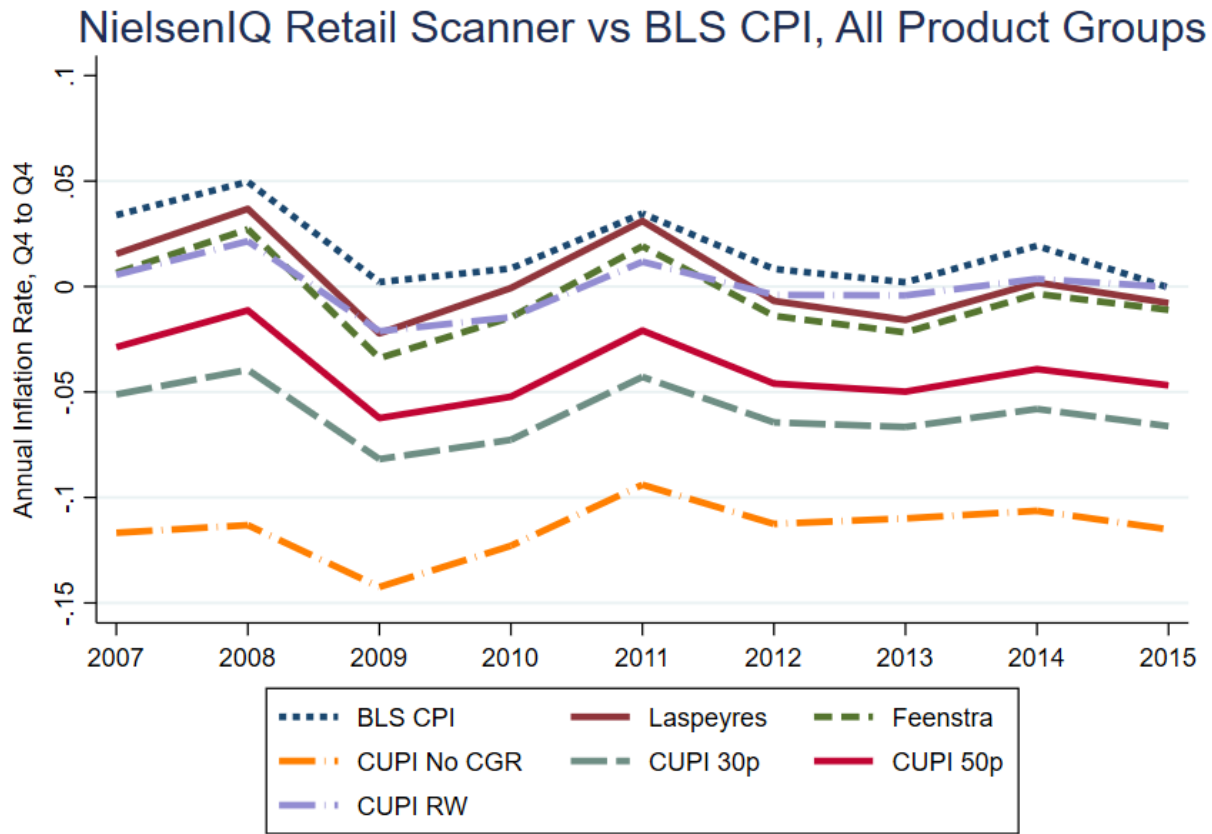
Notes: Figure uses NielsenIQ Consumer Panel data for food and nonfood product groups. The series “CUPI CGR RW” uses a 5th-percentile sales cutoff for the common goods rule. Percentile computed from sales pooled over 5 quarter horizon (current and prior 4 quarters). Laspeyres is arithmetic.

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



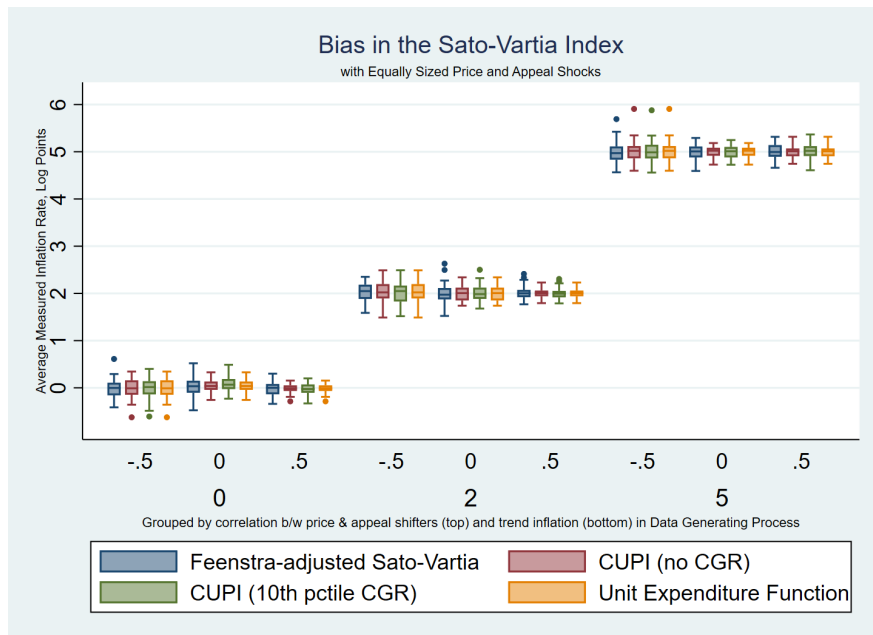
Notes: Figure uses NielsenIQ Consumer Panel for food and nonfood product groups. The indices are YoY for Q4. The series RW(2020) uses the same CGR duration rule as in Redding and Weinstein (2020). The series 5p(5q) use percentiles based on sales pooled over 5 quarter horizon (current and prior 4 quarters).

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



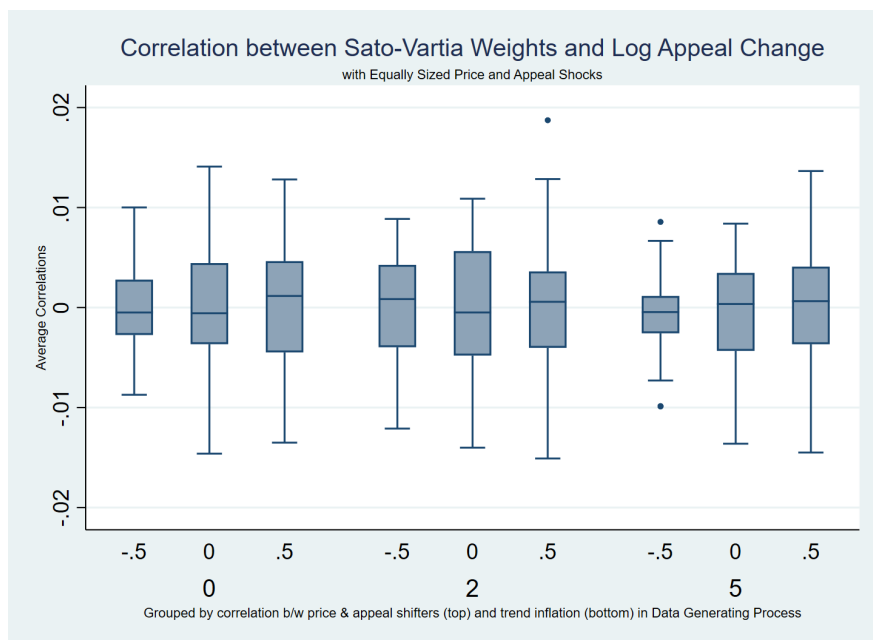
Notes: Figure uses NielsenIQ Retail Scanner data for food and nonfood product groups. The “CUPI, 25p” and “CUPI, 50p” series use 25th- and 50th-percentile cutoffs for the common goods rule, respectively. The series “CUPI, RW CP” uses the CGR 5th percentile threshold from the consumer Panel data for the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

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



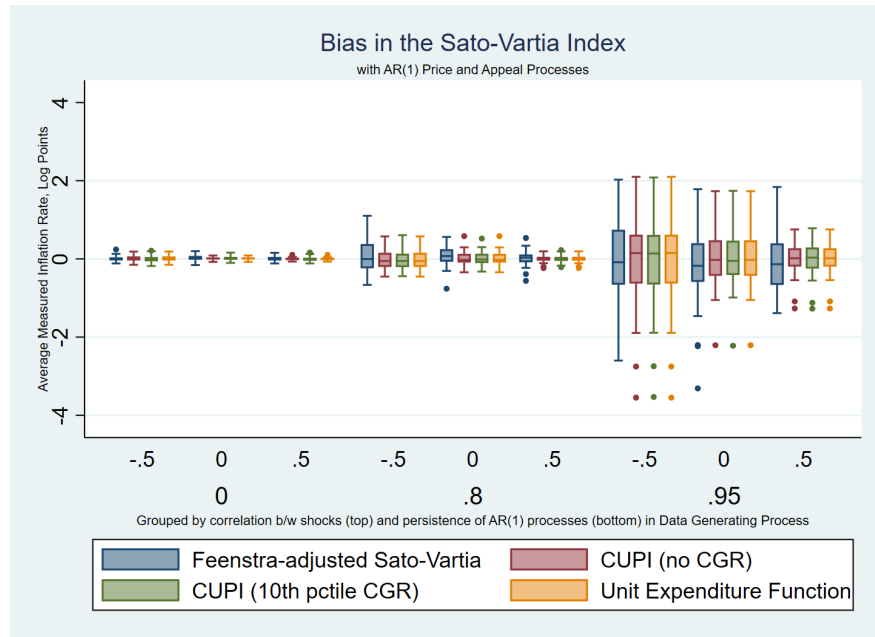
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.2.1. The 0, 2, and 5 along the horizontal axis specify the average trend rates of log price growth in log points, while the -0.5 , 0, and 0.5 along the horizontal axis specify the correlations between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

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



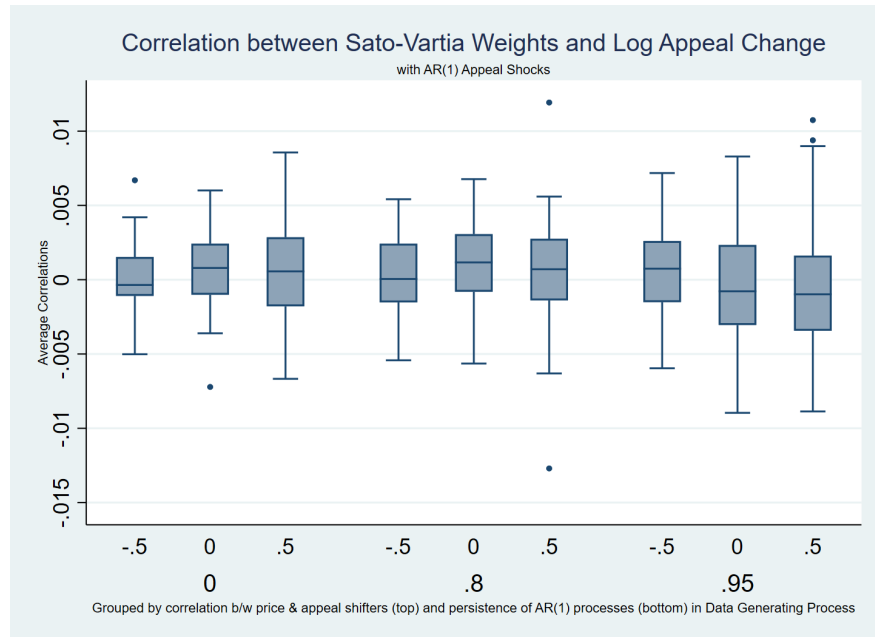
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.2.1. The 0, 2, and 5 along the horizontal axis specify the average trend rates of log price growth in log points, while the -0.5 , 0, and 0.5 along the horizontal axis specify the correlations between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

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



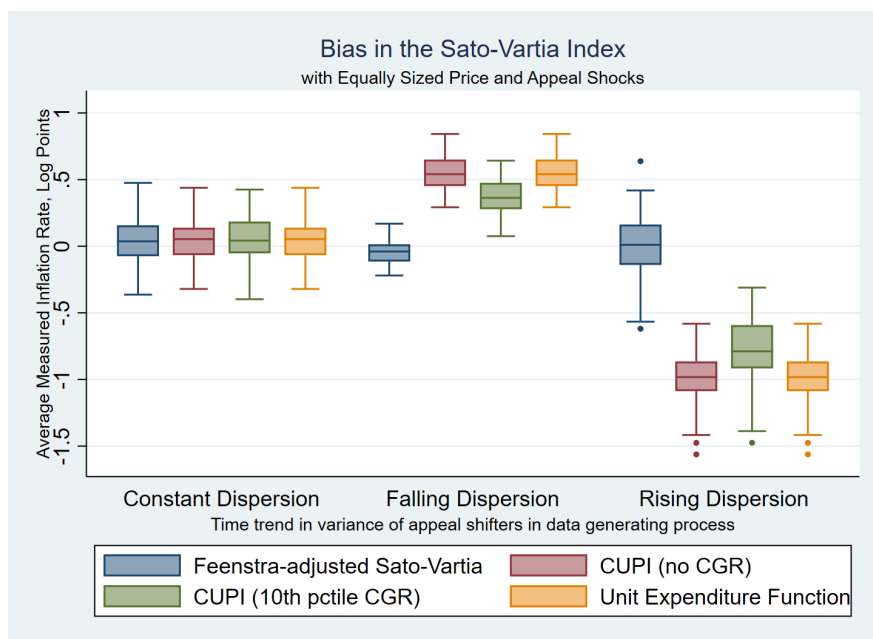
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.2.2. The 0, 0.8, and 0.85 along the horizontal axis specify the autocorrelations of log prices and log appeal parameters, while the -0.5 , 0, and 0.5 along the horizontal axis specify the correlations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

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



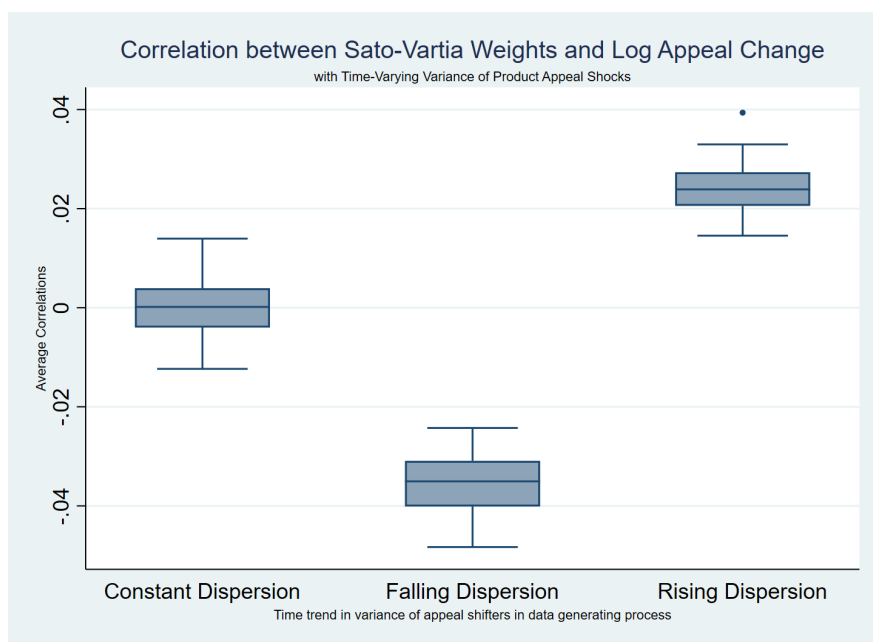
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.2.2. The 0, 0.8, and 0.85 along the horizontal axis specify the autocorrelations of log prices and log appeal parameters, while the -0.5 , 0, and 0.5 along the horizontal axis specify the correlations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

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



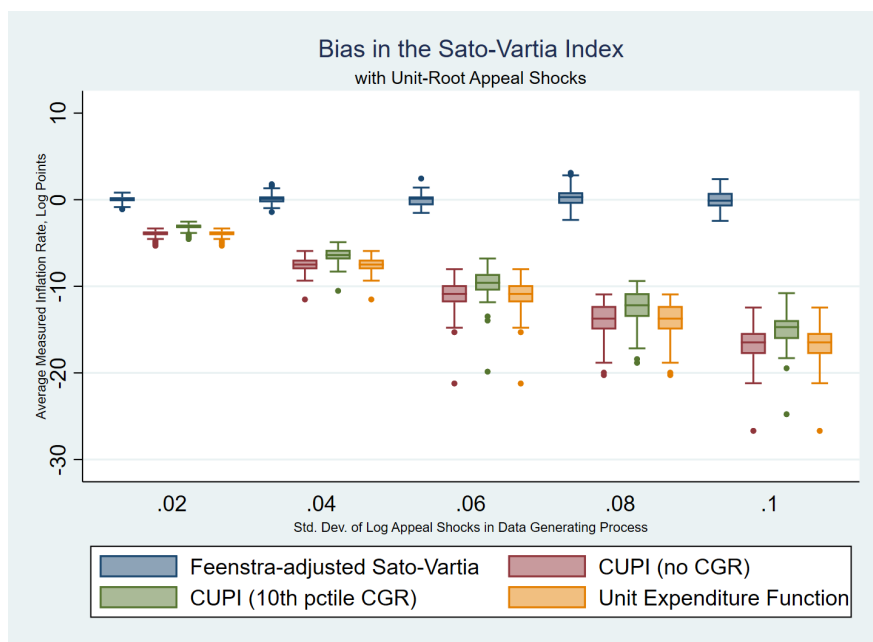
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.3.1. The variance of log product appeal draws is constant over time in the simulations labeled “Constant Dispersion,” falling over time in the simulations labeled “Falling Dispersion,” and rising over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

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



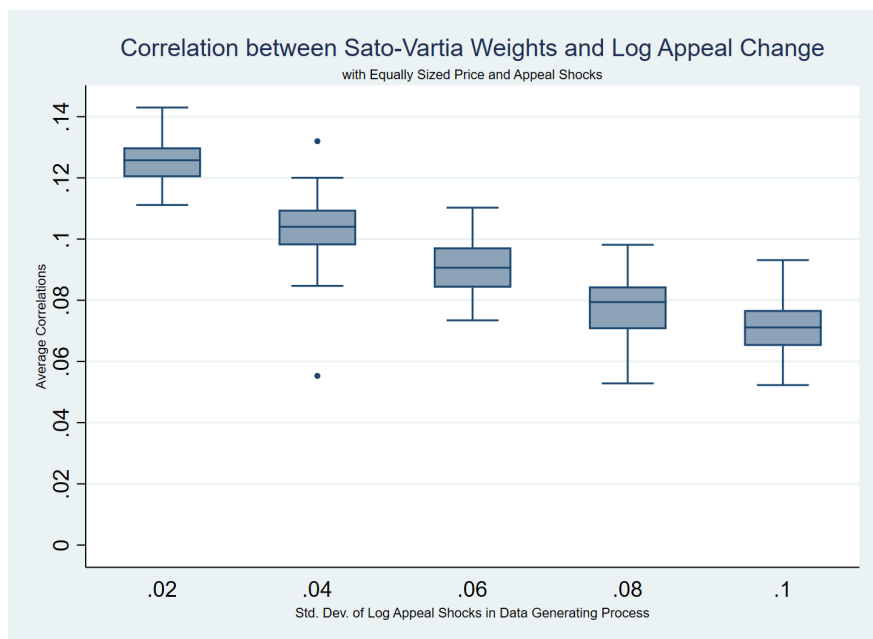
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.3.1. The variance of log product appeal draws is constant over time in the simulations labeled “Constant Dispersion,” falling over time in the simulations labeled “Falling Dispersion,” and rising over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

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



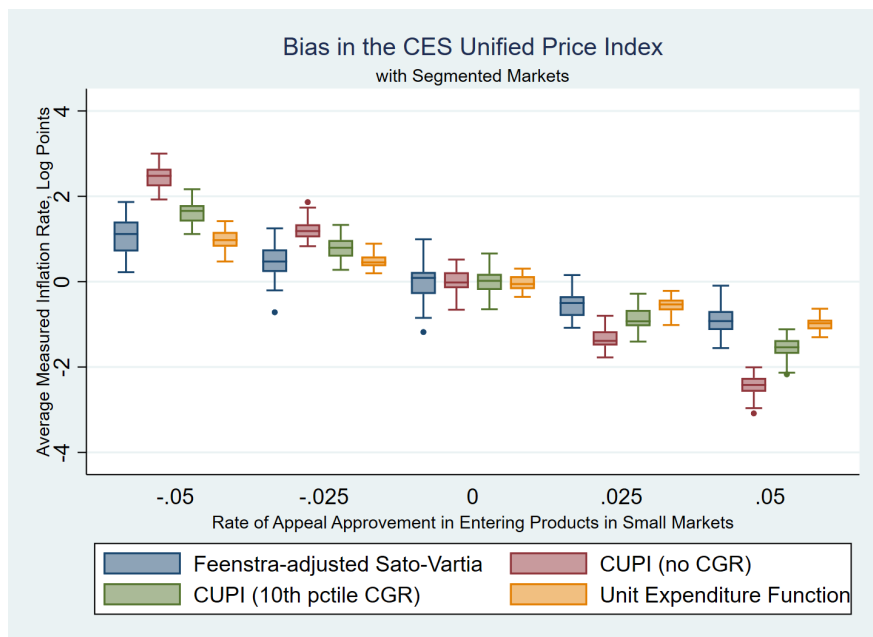
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.3.2. The horizontal axis displays the standard deviations of log appeal shocks. The vertical axis displays the average log inflation rate across simulation periods.

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



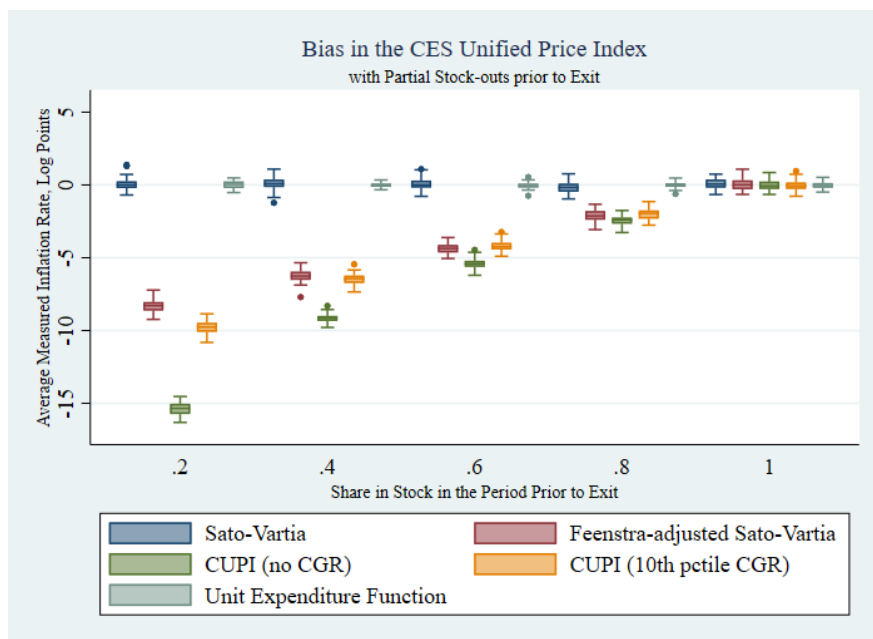
Notes: The figure displays measured inflation from the Monte Carlo simulations described in section D.2.3.2. The horizontal axis displays the standard deviations of log appeal shocks. The vertical axis displays the average correlation between the Sato-Vartia weights and the log appeal shocks across simulation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

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



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

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



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